

ECOLE DOCTORALE SCIENCES POUR L'INGÉNIEUR : MÉCANIQUE, PHYSIQUE, MICRO ET
NANOÉLECTRONIQUE

Thèse présentée pour obtenir le titre universitaire de Docteur d'Aix-Marseille Université

Discipline : Sciences pour l'Ingénieur
Spécialité : Acoustique

Oumar Issiaka TRAORE

Méthodologie de traitement et d'analyse de signaux expérimentaux
d'Émission Acoustique : Application au comportement d'un élément
combustible en situation accidentelle

Soutenue publiquement le 15 janvier 2018 devant le jury composé de :

J. MOYSAN	Professeur, Aix-Marseille Université / LMA	<i>Président</i>
N. GODIN	Maître de Conférences (HDR) , INSA Lyon / MATEIS	<i>Rapporteur</i>
E. SALINELLI	Professeur, Università del Piemonte Orientale, Novara (Italie)	<i>Rapporteur</i>
L. SIMON	Professeur, Le Mans Université / LAUM	<i>Examineur</i>
Ph. VIEU	Professeur, Université Toulouse III Paul Sabatier / IMT	<i>Examineur</i>
K. BENHENNI	Maître de Conférences (HDR), Université P. Mendès-France, Grenoble / LJK	<i>Examineur</i>
N. FAVRETTO-CRISTINI	Chargée de Recherche (HDR) / LMA	<i>Directrice de thèse</i>
S. VIGUIER-PLA	Maître de Conférences (HDR), Université de Perpignan via Domitia / IMT	<i>Co-directrice de thèse</i>
L. PANTERA	Ingénieur de Recherche (PhD), CEA-Cadarache / LPRE (Encadrant)	<i>Membre invité</i>

Numéro national de thèse/suffixe local : 2018AIXM0011/002ED353

Résumé

L'accident d'injection de réactivité, ou Reactivity Initiated Accident (RIA) en anglais, est un accident dit de type grave, caractérisé par une insertion de réactivité dans le cœur d'un réacteur nucléaire. Il induit une excursion de puissance quasi-instantanée avec un dépôt significatif d'énergie dans les crayons de combustibles, soumettant ainsi ceux-ci à une situation de stress extrême. Dans le cadre de campagnes expérimentales visant à étudier le comportement du combustible dans de telles situations, le Commissariat français à l'Energie Atomique (CEA), en collaboration avec l'Institut de Radioprotection et de Sécurité Nucléaire (IRSN), exploite un réacteur de type piscine destiné à reproduire sur un crayon de combustible irradié, prélevé dans une centrale en activité, une situation de stress représentative d'un RIA. Le dispositif d'essais contenant ce crayon est équipé de plusieurs capteurs, dont deux capteurs piézoélectriques permettant de recueillir des signaux d'émission acoustique traduisant le comportement du combustible pendant les essais. L'objectif de cette thèse est de contribuer à l'amélioration du processus de dépouillement des essais RIA, via le traitement des signaux d'émission acoustique acquis grâce à ces deux capteurs, en tenant compte des contraintes d'hostilité et d'accès intrinsèques à l'environnement nucléaire. Il s'agit notamment d'identifier clairement les mécanismes physiques pouvant intervenir au cours des essais à travers leur signature acoustique.

Dans un premier temps, est étudié au travers de calculs analytiques et des simulation numériques conduites au moyen d'une méthode d'éléments finis spectraux, l'impact du dispositif d'essais sur la propagation des ondes. Si une fonction de transfert globale, ouvrant la voie à l'utilisation de méthodes de déconvolution du signal, n'a pu être dégagée ici, une fréquence de résonance du dispositif, pouvant être filtrée ou pénalisée afin d'améliorer l'interprétabilité des signaux, a été cependant identifiée. Par ailleurs, le dispositif d'essais semble n'avoir que très peu d'impact sur les signaux dans la gamme des très basses fréquences.

En second lieu, diverses techniques de traitement du signal (soustraction spectrale, analyse spectrale singulière, ondelettes. . .) ont été expérimentées, afin de proposer des outils permettant de traiter différents types de bruit susceptibles de polluer les signaux reçus. La soustraction spectrale s'avère être la méthode la plus robuste aux changements de nature du bruit, avec un fort potentiel d'amélioration du rapport signal-à-bruit. Par ailleurs, si les paramètres d'émission acoustique liés à la forme d'onde du signal source sont bien restaurés, ceux liés au contenu fréquentiel ou à l'activité acoustique sont plus difficiles à recouvrer.

Enfin, après estimation des signaux sources associés aux salves détectées, des méthodes d'analyse de données multivariées et d'analyse de données fonctionnelles ont été appliquées, afin de proposer un algorithme de classification statistique permettant de mieux comprendre la phénoménologie des accidents de type RIA et d'identifier les mécanismes physiques pouvant survenir. Selon l'approche (multivariée ou fonctionnelle), les algorithmes obtenus permettent de reconnaître le mécanisme associé à une salve dans près de 80% des cas, ce pourcentage pouvant monter jusqu'à 99% pour les mécanismes les mieux représentés dans l'échantillon de salves. Les mécanismes associés à la rupture de la gaine du combustible ou à la chute des barres de contrôle permettant de mettre fin aux essais sont notamment identifiés de manière très satisfaisante. Par ailleurs, l'étude comparative des deux approches permet de conclure que, dans le contexte qui est le nôtre, les méthodes classiques d'analyse multivariée sont mieux adaptées que les méthodes fonctionnelles. Cependant, les résultats de l'approche fonctionnelle

étant également très corrects et prometteurs, elle constitue une très bonne alternative à l'approche multivariée classique. Elle permet notamment d'éliminer les étapes de construction et de choix des paramètres EA qui constituent les principales barrières à la création de modèles statistiques robustes dans plusieurs contextes de contrôle par émission acoustique.

Mots clés : Émission Acoustique, Soustraction spectrale, Analyse de données fonctionnelles, Analyse de données multivariées, Data mining, Clustering, Environnement nucléaire, Modélisation numérique.

Abstract

Reactivity Initiated Accident (RIA) is a nuclear reactor accident that involves an unexpected and very fast increase in fission rate and reactor power due to the ejection of a control rod. The power increase may damage the fuel clad and the fuel pellets of the reactor. The French Alternative Energies and Atomic Energy Commission (CEA) in collaboration with the Institute for Radiation Protection and Nuclear Safety (IRSN) operate a pool-type reactor dedicated to fuel behavior study in RIA conditions. During RIA experiments, the fuel rod to test is included in a test device equipped with two piezoelectric sensors allowing to register acoustic emission about the fuel behavior during the experiments. The objective of the thesis is to contribute to the improvement of the monitoring process of the experiments in such a complex environment, by using the acoustic emission technique. In particular, we want to identify the physical mechanisms occurring during the experiments through their acoustic signatures.

Firstly, analytical derivations and numerical simulations using the spectral finite element method have been performed in order to evaluate the impact of the wave travel path in the test device on the recorded signals. It has been shown that the transfer function of the test device is strongly dependent on both source and receiver locations, which precludes signal processing deconvolution. Furthermore, a resonant frequency has been identified and it has been shown that the geometry and the configuration of the test device may not influence the wave propagation in the low frequency range.

Secondly, signal processing methods (spectral subtraction, singular spectrum analysis, wavelets, . . .) have been explored in order to propose different denoising strategies according to the type of noise observed during the experiments. If we consider only the global SNR improvement ratio, the spectral subtraction method is the most robust to changes in the stochastic behavior of noise. The analysis of the impact of the different selected denoising methods on some classically used AE parameters shows that, whatever the type of noise and the denoising method, AE parameters associated with the waveform are better restored than those associated with the acoustic activity and the frequency content.

Finally, after the estimation of the source signal associated with the events detected during each experiment, classical multivariate and functional data analysis tools are used in order to create a machine learning algorithm dedicated to contribute to a better understanding of the phenomenology of RIA accidents. According to the method (multivariate or functional), the obtained algorithms allow to identify up to 80% of the mechanisms. This result can be about 99% for well represented mechanisms of the sample. Furthermore, the comparison of the results of the multivariate and the functional methods leads to the conclusion that the first ones are better indicated to process the kind of data of interest in this study. However, the results of the functional approach being good, this innovative method could constitute an excellent alternative to classical multivariate approaches.

Keywords: Acoustic emission, Spectral subtraction, Functional data analysis, Multivariate data analysis, Data mining, Clustering, Nuclear environment, Numerical modeling.

Remerciements

Cette thèse a été effectuée au laboratoire de mécanique et d'acoustique (LMA) de Marseille, dans le cadre d'une collaboration avec le Commissariat à l'Energie Atomique (CEA) de Cadarache et l'Institut de Mathématiques de Toulouse (IMT). Que toutes les personnes ayant participé à ma formation, de l'école fondamentale de Missira (cours F) au LMA de Marseille, trouvent ici l'expression de ma profonde gratitude.

Je remercie chaleureusement Mme Nathalie GODIN et M. Ernesto SALINELLI d'avoir accepté d'être rapporteurs de mon travail, ainsi que tous les membres du jury.

Un grand merci à mes deux directrices de thèse Nathalie et Sylvie pour la rigueur scientifique et l'accompagnement de qualité dont j'ai bénéficié tout au long de cette aventure. A Laurent, auprès de qui j'ai énormément progressé ces dernières années. A Philippe pour son accompagnement depuis de nombreuses années maintenant. A Paul, pour ses nombreux conseils très utiles. A Fabien, dont le travail de stage a contribué à la richesse de cette thèse. A Jean-Pascal et Christophe, dont les efforts ont permis de réaliser cette thèse dans de très bonnes conditions matérielles. Jean-Pascal, malheureusement il n'y aura pas de théorème du chanoine kir, le sujet s'y prêtait mal. Mais bon, on pourra toujours donner ce nom à l'un des modules de l'atelier de dépouillement.

Je remercie également le personnel administratif des trois laboratoires (LMA, CEA/LPRE et IMT) m'ayant accueilli pendant ces trois années. Un grand merci à tous mes collègues doctorants Bence, Olivier, Harold, Mathilde (ma tata), Bachir, Alexis..., et à tous ceux qui ont contribué à rendre le déroulement de cette thèse si agréable.

Une pensée pour mes grandes familles Toulousaine (Pep, Sab-girl, Abdou, Sidiki,...) et Marseillaise (ma Bouramiche, Sory, Adama, Aly, Togo...). A Maimourchou, Mimi, Babiri et Sanou pour avoir été là, malgré la distance.

A ma *Wôlôfô Déni*, pour sa constance dans l'engagement et son soutien moral de tous les instants.

Enfin, une pensée pour mes parents, ainsi que toute la grande famille TRAORE, merci pour tout !

Table des matières

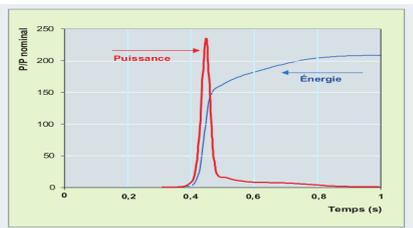
Introduction générale	1
1. Émission Acoustique : principes et utilisation en contexte nucléaire	6
1.1. Généralités	7
1.2. Mise en œuvre du contrôle par EA	7
1.2.1. Phase 1 : Détection d'ondes et prétraitement du signal reçu	8
1.2.1.1. Détection des ondes EA	8
1.2.1.2. Prétraitement du signal EA reçu	9
1.2.2. Phase 2 : Analyse et interprétation des salves	12
1.2.2.1. Localisation du mécanisme physique	12
1.2.2.2. Identification de mécanismes physiques	12
1.3. Utilisation de l'EA en contexte nucléaire	14
1.3.1. Surveillance du circuit primaire et de la cuve sous pression	15
1.3.1.1. Détection de fuites	15
1.3.1.2. Estimation de la gravité de la fuite	15
1.3.1.3. Détection d'ébullition du fluide caloporteur	16
1.3.2. Surveillance du comportement du combustible nucléaire	16
1.3.2.1. Comportement du combustible	16
1.3.2.2. Comportement de la gaine	17
1.4. Conclusion	17
2. De la source aux capteurs : physique des ondes associée à la structure contrôlée	19
2.1. Introduction	20
2.2. Simplification du dispositif d'essais	21
2.3. Étude analytique de la propagation des ondes au sein du dispositif d'essais simplifié	23
2.3.1. Matrice des fréquences	24
2.3.2. Cas des modes de torsion	26
2.4. Modélisation numérique de la propagation des ondes au sein du dispositif d'essais simplifié	27
2.4.1. Généralités sur la méthode des éléments finis spectraux	27
2.4.2. Mise en œuvre de la modélisation numérique	28
2.4.2.1. Création du maillage	28
2.4.2.2. Calculs numériques	28
2.4.3. Résultats	29
2.4.3.1. Fréquence de résonance du dispositif d'essais	29
2.4.3.2. Influence de la position de la source et du capteur sur la fonction de transfert du dispositif d'essais	29
2.4.3.3. Analyse fine des fonctions de transfert selon différentes gammes de fréquences	31
2.4.3.4. Impact d'un changement de fluide caloporteur	31
2.5. Conclusion	33

3. Estimation du signal source et détection de salves	34
3.1. Introduction	35
3.2. Types de bruit associés à l'environnement d'un RIA	35
3.3. Principes de quelques méthodes de détection de salves et de traitement du bruit	37
3.3.1. Filtrage par sélection de fréquences	38
3.3.2. Transformée de Fourier à court terme et Soustraction spectrale	38
3.3.2.1. Transformée de Fourier à court terme	38
3.3.2.2. Soustraction spectrale	39
3.3.3. Analyse spectrale singulière	40
3.3.3.1. Principe de la SSA	40
3.3.3.2. Détection de salves par SSA	43
3.3.3.3. Débruitage par SSA	44
3.3.4. Analyse par ondelettes	46
3.3.4.1. Détection de salves par ondelettes	46
3.3.4.2. Débruitage par ondelettes	46
3.4. Étude comparative des performances des méthodes	47
3.4.1. Détection de salves	48
3.4.2. Débruitage	50
3.4.2.1. Potentiel d'amélioration du rapport signal-à-bruit	50
3.4.2.2. Impact du débruitage sur les paramètres EA	52
3.4.2.3. Amélioration du débruitage par utilisation conjointe des méthodes	53
3.5. Conclusion	55
4. Classification et interprétation de données EA	56
4.1. Introduction	58
4.2. Prétraitement des salves	61
4.2.1. Cas des paramètres classiques	61
4.2.1.1. Correction des biais inhérents à la détection des salves	62
4.2.1.2. Réduction de la taille de la matrice de données par suppression des multicollinéarités	64
4.2.2. Prétraitement des courbes dans le cas d'une approche fonctionnelle	66
4.2.2.1. Reconstruction de la forme fonctionnelle et lissage des courbes	67
4.2.2.2. Traitement des différences de support et des décalages horizontaux	68
4.3. Analyse exploratoire : de l'ACP classique à l'ACP fonctionnelle	70
4.3.1. Qu'est-ce qui change ?	70
4.3.2. Analyse des résultats de l'ACP classique et de l'ACP fonctionnelle	71
4.4. Classification non supervisée et identification des mécanismes	74
4.4.1. Exploitation des informations <i>a priori</i> et critère statistique de comparaison des classifications	74
4.4.2. Approche classique	75
4.4.3. Approche fonctionnelle	76
4.4.3.1. Approches et méthodes de clustering envisageables	76
4.4.3.2. Résultats obtenus	77
4.4.4. Recoupement des résultats	81
4.5. Classification supervisée	82
4.5.1. Construction des jeux d'entraînement et test	83
4.5.2. Approche classique	83
4.5.2.1. Régression logistique multinomiale	83
4.5.2.2. Analyse discriminante linéaire	84
4.5.2.3. Les méthodes du type plus-proches-voisins	85
4.5.2.4. Arbre de classification (CART)	86

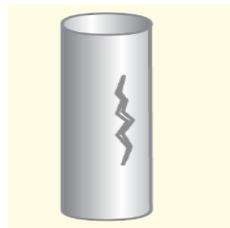
4.5.2.5. Bagging d'arbres de classification	87
4.5.2.6. Boosting d'arbres de classification	88
4.5.2.7. Forêts aléatoires	88
4.5.2.8. Choix du modèle final	89
4.5.3. Approche fonctionnelle	90
4.5.3.1. Stratégie exclusivement non paramétrique	90
4.5.3.2. Stratégie exclusivement basée sur la réduction de dimension	91
4.5.3.3. Choix du modèle final	92
4.5.4. Comparaison des résultats des approches classique et fonctionnelle et prise en charge de salves inconnues	92
4.5.4.1. Comparaison des performances des méthodes classe par classe	92
4.5.4.2. Prise en charge de salves inconnues	94
4.6. Conclusion	96
Conclusion générale et Perspectives	101
Références	102
Liste des Figures	114
Liste des Tableaux	117
A. Annexes	118
A.1. Liste non exhaustive de paramètres EA utilisés dans la littérature	118
B.1. Résolution de l'équation d'onde en coordonnées cylindriques par la méthode de séparation des variables	121
B.2. Conditions aux limites relatives à la résolution de l'équation d'onde dans le cas du dispositif simplifié	122
B.3. Coefficients de la matrice des fréquences	123
B.4. Brève présentation des fonctions de Bessel	124
B.4.0.1. Fonctions de Bessel de première espèce	125
B.4.0.2. Fonctions de Bessel de deuxième espèce	125
C.1. Impact des méthodes de débruitage sur des paramètres EA choisis	126
D.1. Résultats détaillés de la mise en œuvre de l'ACP	130
D.1.1. Considérations pratiques de mise en œuvre de l'ACP fonctionnelle	130
D.1.2. Résultats obtenus	130
D.1.2.1. Cas classique	130
D.1.2.2. Cas fonctionnel	131
D.2. Classification non supervisée	131
D.2.1. Informations complémentaires sur la mise en œuvre des méthodes	131
D.2.2. Description des classes obtenues	132
D.3. Classification supervisée	137
D.3.1. Décomposition de la variance d'une matrice de données	137
D.3.2. Notions de validation croisée	138

Introduction générale

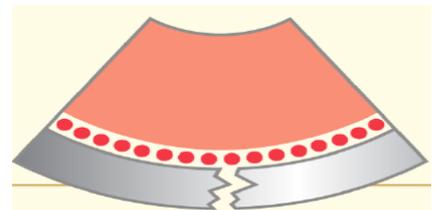
L'accident d'injection de réactivité, ou «**Reactivity-Initiated Accident**» (RIA) en anglais, est un accident dit de type grave, caractérisé par une insertion de réactivité dans le cœur d'un réacteur nucléaire. Il se définit par une cinématique très rapide conduisant à une excursion de puissance quasi-instantanée, avec un dépôt significatif d'énergie dans les crayons de combustible, soumettant ainsi ceux-ci à une situation de stress extrême. Dans une première phase, le RIA s'étend sur une dizaine de millisecondes et se caractérise par une forte interaction mécanique entre le combustible et la gaine. En cas de rupture de la gaine, lors de cette phase initiale, une dispersion de combustible, sous forme de petits fragments, peut conduire à la vaporisation brutale de l'eau entourant le crayon, avec risque d'explosion vapeur et de dégradation des crayons combustibles voisins (Figure 0.1). La réaction en chaîne s'arrête alors d'elle-même par échauffement du combustible et effet Doppler. Dans sa seconde phase, d'une durée de quelques secondes à une dizaine de secondes, l'accident est gouverné par l'évolution thermique du système, conduisant à un fort accroissement de la température de la gaine, avec le risque d'atteindre la crise d'ébullition dans le canal d'eau qui l'entoure, et au maintien d'une forte pression interne du crayon pouvant, là encore, affecter l'intégrité de la gaine [1].



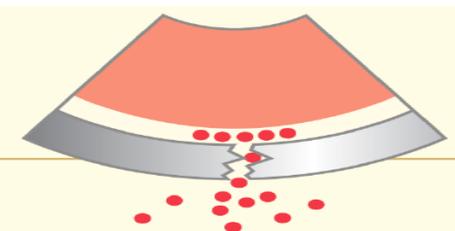
(a) Évolution de la puissance et de l'énergie injectée lors d'un accident de réactivité en fonction du temps



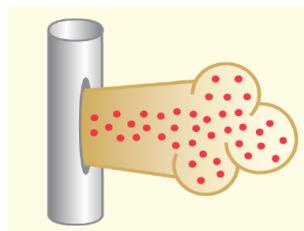
(b) Fissuration de gaine



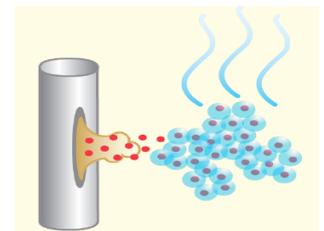
(c) Rupture de gaine sans éjection de combustible



(d) Rupture de gaine avec éjection de combustible



(e) Rupture de gaine avec éjection de combustible et de gaz de fission



(f) Rupture de gaine avec éjection de combustible et de gaz de fission accompagnée d'une crise d'ébullition

Figure 0.1. – Illustration de l'impact d'un RIA sur le couple gaine/combustible (©[1, 2])

L'accident RIA le plus grave de l'histoire survint à Tchernobyl (Ukraine) en 1986. Il eut comme conséquence le craquage de l'eau des circuits de refroidissement, conduisant par la suite à une explosion et à la libération d'importantes quantités d'éléments radioactifs dans l'atmosphère. La très forte

irradiation et la contamination de l'environnement provoquèrent alors de nombreux décès et maladies survenus immédiatement ou à plus long terme. Cet événement fut alors un déclic pour un début et/ou une intensification des programmes de recherche nucléaire visant à mettre en place ou à valider les critères de sûreté relatifs à ce type d'accident. Dès le début des années 1990, d'importants programmes de recherche furent initiés en France, au Japon et en Russie [3]. En France, la majeure partie de ces programmes sont réalisés au centre du Commissariat à l'Énergie Atomique et aux Énergies Alternatives (CEA) de Cadarache. Ils permettent d'étudier, selon divers scénarios, le comportement du combustible, mais également les interactions entre celui-ci et son environnement immédiat. Ainsi, ont été menés des essais portant sur les accidents de réactivité des réacteurs rapides refroidis au sodium sur des crayons uniques. Les fortes interactions mécanique pastille-gaine se produisant sans échauffement significatif de la gaine ont notamment été étudiées dans ce cadre. Des conditions représentatives d'un échauffement de la gaine sous une forte pression feront prochainement l'objet d'une nouvelle campagne d'essais dans le cadre d'une large coopération internationale.

Le suivi ou «**monitoring**» de tels essais est absolument crucial ; il constitue un enjeu technique majeur. En effet, l'environnement associé à un réacteur nucléaire se caractérise par son hostilité et sa difficulté d'accès, rendant presque impossible l'utilisation de toute méthode de surveillance intrusive. Par conséquent, sont privilégiées les méthodes de surveillance non intrusive dont on retrouve un très large panel dans la littérature [4–10]. Elles sont basées sur différents principes et leur application peut parfois nécessiter une haute technicité quant à la préparation et la réalisation des essais expérimentaux, ainsi que l'analyse des résultats. Parmi celles potentiellement adaptées au contexte nucléaire, les techniques (actives et passives) de contrôle par ondes acoustiques sont particulièrement attractives, puisque faciles à adapter aux contraintes d'hostilité et d'accès intrinsèques à l'environnement. Ces méthodes sont utilisées par exemple dans le suivi de la compaction du combustible [11] et font l'objet d'intenses efforts de recherche, notamment en prévision de la surveillance des réacteurs dits de quatrième génération [12, 13]. Elles sont également employées pour le suivi d'accidents simulés de type RIA.

Les essais réalisés au centre CEA Cadarache, et destinés à reproduire les situations accidentelles représentatives d'un accident RIA, se font sur un réacteur expérimental de type piscine (Figure 0.2). Ce réacteur est constitué d'un cœur à base de combustible fabriqué à partir d'oxyde d'uranium (UO_2), conçu pour résister aux variations rapides de puissance lors des essais. Le cœur est équipé, en son centre, d'une cavité verticale permettant de recevoir le dispositif d'essai contenant l'échantillon de combustible à tester. Ce dispositif est équipé de plusieurs capteurs (pression, déplacement, débit-mètre...), dont deux capteurs piézoélectriques permettant de suivre les essais et d'étudier le comportement du combustible pendant les transitoires de puissance. Distants de 2 mètres, ces deux capteurs piézoélectriques sont situés respectivement au-dessus et en-dessous du crayon de combustible (Figure 0.2). Composés de cristaux de niobate de lithium, ils ont été fabriqués pour résister à des températures pouvant atteindre 600 degrés Celsius. Associés à un préamplificateur 40dB, ils permettent d'acquérir, avec une fréquence d'échantillonnage de $2.5\mu s$, les signaux d'**émission acoustique** (EA) générés par différents mécanismes sources, engendrés au cours des essais RIA et potentiellement liés au comportement du combustible (Figure 0.1).

On voit alors tout l'intérêt et toute l'importance d'identifier clairement ces mécanismes sources inconnus, au travers de leur signature acoustique. C'est précisément l'objectif-cible du travail de thèse présenté ici. En particulier, nous cherchons à apporter des réponses aux questions suivantes : **des mécanismes physiques d'intérêt ont-ils eu lieu pendant l'essai RIA ? Si oui, à quels moments de l'essai ? De quels mécanismes s'agit-il ?**

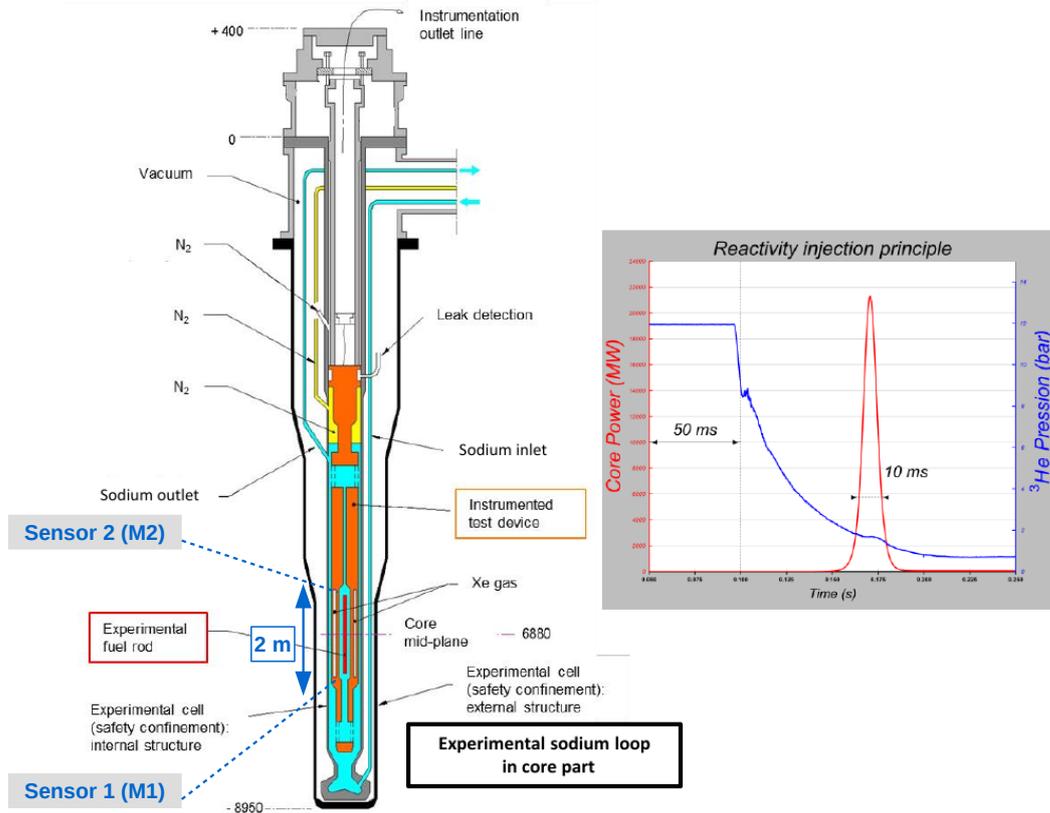


Figure 0.2. – Schéma illustratif du dispositif d'essais dans lequel sont positionnés deux capteurs d'émission acoustique (sensors 1 et 2).

La technique EA est certes une méthode acoustique passive efficace pour détecter et suivre un processus évolutif au sein d'une structure, mais il faut garder à l'esprit deux points importants. D'une part, l'émissivité des sources dépend généralement des propriétés de ladite structure et des conditions de sollicitation. Elle est donc propre à la configuration étudiée. Par conséquent, seule la méthodologie employée pour l'analyse des signaux EA (et non les résultats de cette analyse) peut être généralisée à d'autres configurations. D'autre part, les limitations de la technique EA sont essentiellement liées à l'analyse des données recueillies. Outre leur grande quantité, c'est surtout l'essence même de ces données qui est problématique. En effet, elles sont souvent composées d'un très grand nombre de salves (ou événements) résultant de la propagation, au sein de la structure étudiée, des ondes acoustiques générées par divers mécanismes physiques sources. Or, le signal essai x résultant des acquisitions réalisées au-cours d'un essai est le résultat d'une composition de plusieurs phénomènes. Outre le signal source s associé aux potentiels mécanismes physiques d'intérêt (les salves) survenant à des moments précis de l'essai, x est affecté par diverses sources de "pollution". Parmi les plus importantes, on peut citer le bruit environnemental n , la fonction de transfert h_2 du système d'acquisition, et l'impact du chemin de propagation des ondes (caractérisé ci-après par la fonction de transfert h_1). Ce dernier est conditionné par la position du capteur par rapport à la source et par les propriétés de la structure. D'un point de vue mathématique, chaque signal essai peut ainsi s'exprimer sous la forme :

$$x(t) = h_1 * h_2 * s(t) + n(t), \quad t \in \{1, \dots, N\} \quad (0.1)$$

où $*$ désigne l'opérateur de convolution et t le vecteur temps.

En théorie, l'identification des mécanismes sources suppose que l'on ait estimé convenablement le signal source s . Cela implique donc que l'on ait pu au préalable isoler le bruit n , et que l'on ait pu également préciser le rôle des différentes fonctions de transfert h_1 et h_2 , ce qui en pratique est loin d'être trivial. Dans la littérature, il n'est pas rare de constater l'absence de "dépollution" des signaux EA car le niveau de pollution est considéré suffisamment bas dans certaines configurations pour ne pas compromettre les analyses de localisation et d'identification des mécanismes associés aux salves. Cependant, cette hypothèse ne peut s'appliquer en contexte RIA. En effet, d'une part, la complexité du dispositif d'essai contenant le crayon de combustible contraint fortement la propagation d'ondes. D'autre part, l'environnement d'un réacteur nucléaire est très bruyé et les sources potentielles de bruit assez nombreuses. Enfin, le protocole expérimental varie d'un essai RIA à l'autre, ce qui conduit à des changements de nature et/ou de niveau du bruit. En conséquence, en contexte RIA on se doit de mettre en œuvre une méthodologie efficace de **débruitage du signal essai reçu** et de **détection** des salves, avant d'**identifier** aussi précisément que possible la nature physique de ces mécanismes. **Ces étapes constituent le fil conducteur de notre démarche.**

Dans le premier chapitre, après avoir rappelé les généralités sur l'EA, nous passerons en revue le processus de mise en œuvre du contrôle par EA. Il s'agira, pour la première phase de détection d'ondes et de prétraitement du signal essai, ainsi que pour la seconde phase d'analyse et d'interprétation des salves, de rappeler les étapes-clés de ce processus. Ensuite, nous procéderons à un état de l'art sur l'utilisation de l'EA en contexte nucléaire.

Le second chapitre sera consacré à l'étude de la physique des ondes associée à l'environnement des essais RIA. Une approche basée sur des calculs analytiques, consolidés par une modélisation numérique et validée par les signaux expérimentaux, sera proposée afin d'évaluer l'impact h_1 du chemin de propagation des ondes sur le signal essai x . Nous étudierons notamment le comportement de h_1 selon divers scénarios, avant d'analyser les conséquences qu'aurait un changement de fluide caloporteur sur la propagation des ondes.

Dans le troisième chapitre, nous procéderons à une étude comparative de divers outils pour le débruitage du signal essai reçu, ainsi que la détection des salves d'intérêt. Dans le cadre du débruitage, on s'intéressera à des méthodes allant du simple filtrage par sélection de fréquences à des méthodes plus élaborées comme le seuillage par ondelettes. Au-delà de l'objectif classique d'amélioration du rapport signal-à-bruit, «signal-to-noise ratio» (SNR) en anglais, on s'attachera également à évaluer les capacités de ces méthodes à restaurer fidèlement le signal source de manière à ne pas biaiser les paramètres EA permettant de caractériser les salves.

En ce qui concerne la détection des mécanismes, la méthode la plus couramment utilisée est celle du seuillage. Elle consiste à supposer que l'avènement d'un potentiel mécanisme d'intérêt conduit à une augmentation plutôt significative de l'énergie du signal EA. Ainsi, tout dépassement de l'amplitude de référence choisie est considéré comme associé à un potentiel mécanisme d'intérêt. Malheureusement, dans le cas des essais RIA, on ne dispose que d'un *a priori* faible sur les mécanismes attendus. Aussi l'existence de mécanismes physiques très faiblement énergétiques n'entraînant pas d'augmentation substantielle de l'énergie du signal n'est pas à exclure. Cela rend une application directe du seuillage inefficace dans certains cas, conduisant à la nécessité de développer ou d'expérimenter des stratégies de détection plus sophistiquées.

Enfin, la question de l'identification de la nature physique des mécanismes d'intérêt constitue un long processus statistique faisant appel à des outils d'analyse exploratoire, de classification non supervisée et de classification supervisée. La littérature regorge d'exemples d'applications de la plupart des méthodes «classiques», mais n'insiste que très rarement sur la nécessité de définir une stratégie de sélection d'une méthode optimale selon le contexte du contrôle par EA. Dans le dernier chapitre,

nous aborderons donc la question de l'identification de la nature physique des mécanismes sources sous cet angle, en expérimentant une démarche permettant de choisir les méthodes optimales ainsi que les paramétrages associés. Ce chapitre s'inscrira dans une logique d'étude comparative entre des méthodes, que nous appellerons classiques et qui sont couramment utilisées dans la littérature, et des méthodes plus récentes basées sur la statistique fonctionnelle. Il s'agira notamment de positionner ces méthodes fonctionnelles innovantes, par rapport aux méthodes classiques, l'objectif-cible demeurant l'attribution de signatures acoustiques aux mécanismes physiques détectés.

1

■ Émission Acoustique : principes et utilisation en contexte nucléaire

Sommaire

1.1. Généralités	7
1.2. Mise en œuvre du contrôle par EA	7
1.2.1. Phase 1 : Détection d'ondes et prétraitement du signal reçu	8
1.2.1.1. Détection des ondes EA	8
1.2.1.2. Prétraitement du signal EA reçu	9
1.2.2. Phase 2 : Analyse et interprétation des salves	12
1.2.2.1. Localisation du mécanisme physique	12
1.2.2.2. Identification de mécanismes physiques	12
1.3. Utilisation de l'EA en contexte nucléaire	14
1.3.1. Surveillance du circuit primaire et de la cuve sous pression	15
1.3.1.1. Détection de fuites	15
1.3.1.2. Estimation de la gravité de la fuite	15
1.3.1.3. Détection d'ébullition du fluide caloporteur	16
1.3.2. Surveillance du comportement du combustible nucléaire	16
1.3.2.1. Comportement du combustible	16
1.3.2.2. Comportement de la gaine	17
1.4. Conclusion	17

1.1. Généralités

Les travaux de Lord Rayleigh exposés dans "The Theory of Sound" en 1870 et la découverte de la piézoélectricité par les frères Curie en 1880 constituent deux événements majeurs à la base des méthodes de contrôle par les ondes [6]. Quant au contrôle par acoustique passive ou EA, sa paternité est attribuée au professeur allemand J. Kaiser, avec la découverte de l'effet Kaiser en 1950 [4]. L'EA est un phénomène de libération d'énergie sous forme d'ondes élastiques transitoires résultant de micro-déplacements locaux internes à un matériau lorsque celui-ci est soumis à des sollicitations d'origines mécaniques, thermiques ou chimiques (norme AFNOR NF EN1330-9) [14]. La relaxation rapide de l'énergie peut résulter par exemple de déformations plastiques, de ruptures dues à des micro-inclusions, de mouvements inter et intra-granulaires associés à la naissance ou à la croissance de fissures, et de micro-contacts dus à l'interaction entre le matériau et son voisinage [15]. L'effet Kaiser traduit l'absence d'EA lorsque le matériau est porté à une sollicitation inférieure à celle précédemment appliquée.

Au fil des années, le contrôle par EA a énormément gagné en popularité et continue de faire l'objet d'un intérêt croissant dans plusieurs domaines. On y a massivement recours dans le suivi du fonctionnement de structures dans l'industrie nucléaire, la pétrochimie et le génie civil [16–19]. Il existe d'autres applications, mais dans une moindre mesure, dans l'industrie pharmaceutique, notamment pour le suivi de la fabrication des comprimés [20, 21].

Selon les caractéristiques du signal associé, on distingue deux principaux types d'EA : l'émission continue et l'émission discrète (cf. Figure 1.1) [22]. L'émission continue se caractérise par un grand nombre de signaux aléatoires de faible amplitude qui se superposent les uns aux autres et qui s'ajoutent au bruit de fond. Elle correspond à des mécanismes de type déformations plastiques, frottements ou fuites [23]. L'émission discrète ou pulsée, quant à elle, correspond à une suite d'événements de faible durée, d'énergie importante et bien séparés les uns des autres [24]. Elle traduit généralement la présence de mécanismes de type fissuration, naissance et croissance de ruptures et inclusions. On notera que certains travaux évoquent également l'EA dite mixte qui résulte d'une combinaison des deux précédents types d'EA [25]. Elle pourrait, par exemple, être liée à des phénomènes de ruptures accompagnés de fuite.

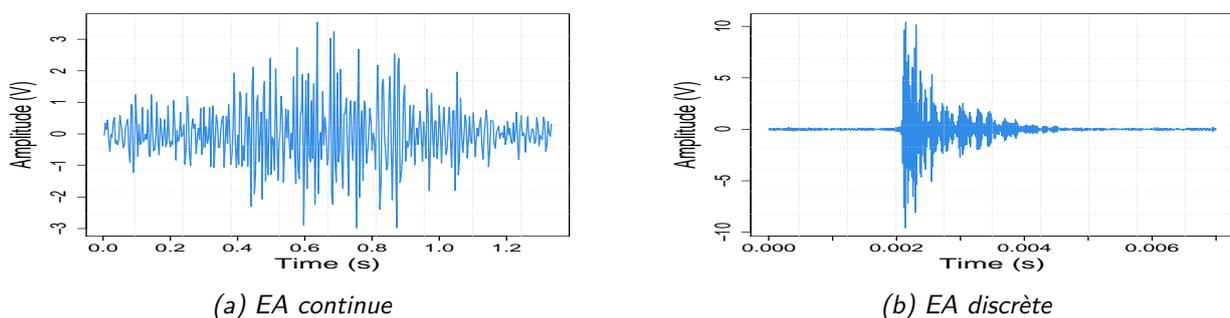


Figure 1.1. – Illustration des deux principaux types d'EA.

1.2. Mise en œuvre du contrôle par EA

Nous nous proposons de passer en revue les différentes phases du processus de contrôle par EA, en mettant en exergue les défis techniques associés et en présentant certaines solutions proposées dans

la littérature.

1.2.1. Phase 1 : Détection d'ondes et prétraitement du signal reçu

1.2.1.1. Détection des ondes EA

L'étape de détection des ondes est d'une importance capitale dans le processus de contrôle par EA. Comme illustré en Figure 1.2 et mentionné en Introduction, le signal reçu et visualisé par l'expérimentateur peut être considéré comme la convolution et/ou la somme de plusieurs signaux. Afin de s'assurer d'être en mesure de collecter les informations correspondant aux mécanismes physiques d'intérêt, la stratégie de détection doit être pensée à la lumière d'études préalables sur le milieu ou la structure à contrôler et la nature des ondes susceptibles de s'y propager. La mise en œuvre de cette stratégie se traduit ensuite par le choix et le paramétrage des éléments constitutifs de la chaîne d'acquisition composée de capteur(s) EA, de préamplificateur(s) et/ou amplificateur(s), et d'un système de visionnage et/ou de stockage de données.

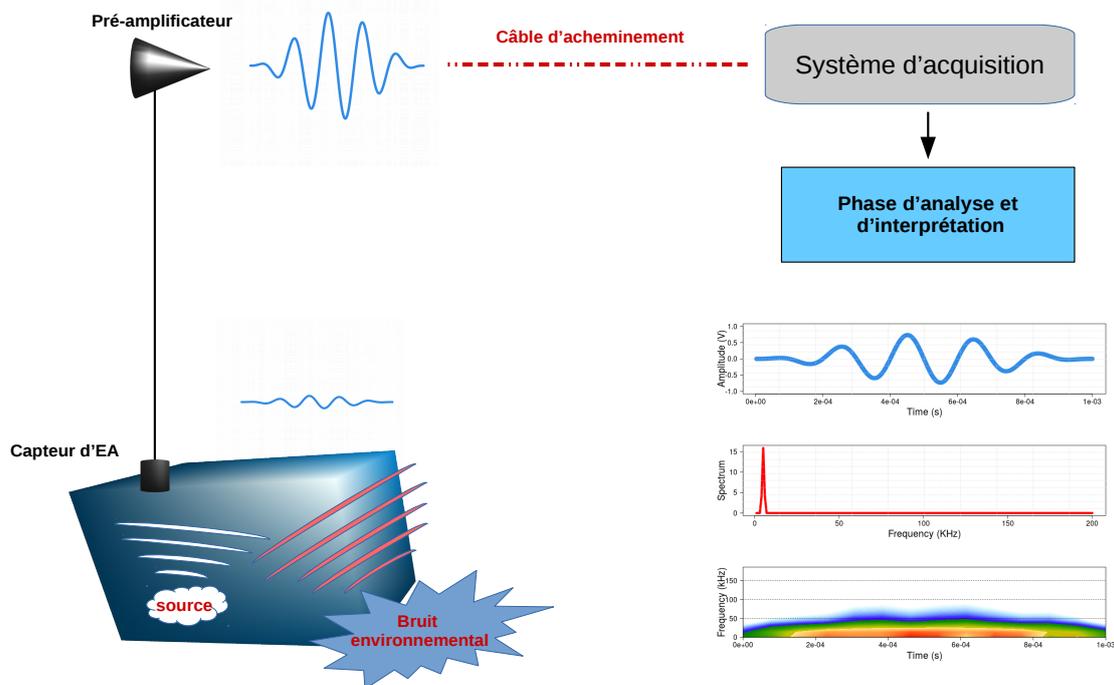


Figure 1.2. – Illustration du processus de contrôle par EA, de la source à l'analyse du signal EA.

a. Choix du capteur EA

Le capteur EA, généralement constitué de matériaux piézoélectriques, traduit le mouvement mécanique issu de la propagation des ondes en un signal électrique [26, 27]. Les fréquences auxquelles le capteur présente les plus fortes sensibilités sont dites fréquences de résonance. En général, il est nécessaire de les traiter avant interprétation des données. Pour ce faire, on peut les filtrer lors de l'acquisition ou les traiter pendant l'étape de prétraitement des signaux.

Un capteur EA peut être résonant ou large bande. Le choix d'un capteur se fait toujours sur la base d'un compromis entre bande passante et sensibilité, et en tenant compte de l'environnement dans

lequel apparaît l'EA. Un capteur résonant présente l'avantage d'avoir une forte sensibilité sur une plage fréquentielle relativement étroite, centrée autour de la fréquence de résonance. Cependant, compte-tenu de cette bande passante, il ne convient que pour des études où les phénomènes observés ont des spectres relativement étroits et similaires. Contrairement au capteur résonant, la bande passante relativement large du capteur large bande permet son utilisation dans les cas où les phénomènes peuvent avoir des fréquences très différentes. Cependant, leur faible sensibilité limite la qualité des signaux captés.

b. Étalonnage du capteur

L'étalonnage permet la traçabilité et la vérification des performances du capteur. Les résultats sont ensuite utilisés pour la quantification des mesures et la comparaison objective des résultats entre expérimentateurs. Différentes normes internationales (ISO), issues notamment des travaux du National Institute of Standards and Technology (USA), définissent les procédures applicables pour l'étalonnage des capteurs. Selon les objectifs de l'étalonnage et les moyens disponibles, les méthodes proposées dans la littérature peuvent être regroupées en trois types : absolues, relatives et de vérification sur site [28, 29].

Les méthodes absolues ont pour objectif de mesurer la sensibilité du capteur sur une échelle absolue. Pour cela, on crée une onde de surface dont les caractéristiques temporelles sont connues à l'emplacement du capteur. La mesure de la réponse du capteur permet ensuite de calculer sa fonction de transfert en phase et en module. Les méthodes relatives sont celles utilisées le plus souvent, notamment pour leur simplicité et leur reproductibilité. Contrairement aux méthodes absolues, elles ne permettent pas de déterminer la fonction de transfert du capteur. Leur utilisation a pour but de vérifier la stabilité de la réponse impulsionnelle du capteur. Elles permettent également de comparer différents capteurs et de procéder à des contrôles périodiques d'efficacité. Enfin, les méthodes de vérification sur site ont pour but de contrôler sur le site le bon fonctionnement de chaque voie d'analyse, une fois leur mise en place terminée. Elles permettent également de comparer les différentes voies d'un système multicanal pour s'assurer qu'elles présentent toutes des réponses analogues. Lors d'essais de longue durée (essais de fatigue ou de surveillance d'installations en service), elles doivent permettre de vérifier que la sensibilité de détection n'évolue pas dans le temps ou, le cas échéant, permettre d'évaluer l'importance de l'évolution.

c. Couplage du capteur à la structure contrôlée

Le couplage permet d'assurer la transmission des ondes EA entre structure et capteur à l'aide d'un milieu approprié, appelé couplant. La qualité et la stabilité du couplage sont essentielles pour une bonne détection des ondes EA, et selon l'objectif visé et le matériau à contrôler, on utilisera préférentiellement le couplage fluide avec pression de maintien, le couplage solide ou le couplage avec guide d'onde interposé [30].

d. Amplification des signaux

Comme illustré en Figure 1.2, les signaux générés et propagés ont une amplitude très faible. En pratique, pour les détecter, on utilise une association capteur-amplificateur et/ou préamplificateur. Ce dernier effectue simultanément l'amplification et l'adaptation en basse impédance des signaux [11]. A la sortie du préamplificateur, le signal est alors acheminé par câble coaxial pour stockage et analyse.

1.2.1.2. Prétraitement du signal EA reçu

Les ondes mécaniques reçues par le capteur sont converties en signaux électriques, amplifiés par le préamplificateur puis transmis au système d'acquisition qui stocke ces signaux. Ces signaux contiennent l'information transmise par la source et les éventuelles sources de pollution que sont le chemin de propagation, le système d'acquisition et le bruit environnemental. Dans les cas où le milieu de propagation

est relativement simple et bien connu, et l'environnement expérimental bien maîtrisé, l'utilisation d'un simple filtre de type passe-bande à la sortie de l'amplificateur suffit pour nettoyer le signal. Dans les cas plus complexes, un recours à des méthodes de traitement du signal plus élaborées est nécessaire.

Le prétraitement des signaux bruts reçus consiste à identifier (ou détecter) les signaux (ou salves) potentiellement associés à des mécanismes physiques d'intérêt. Pour ce faire, la démarche la plus courante, utilisée notamment dans les logiciels commerciaux, est basée sur une stratégie de seuillage. Il s'agit d'utiliser un seuil légèrement au-dessus du niveau du bruit, ainsi que des fenêtres temporelles glissantes au nombre de trois (Figure 1.3) :

- la **PDT** (Peak Definition Time) permet de déterminer le pic de forte amplitude,
- la **HDT** (Hit Definiton Time) consiste à fixer une durée maximale pour chaque salve,
- la **HLT** (Hit Lockout Time) permet de fixer un temps de réarmement après la détection d'une salve

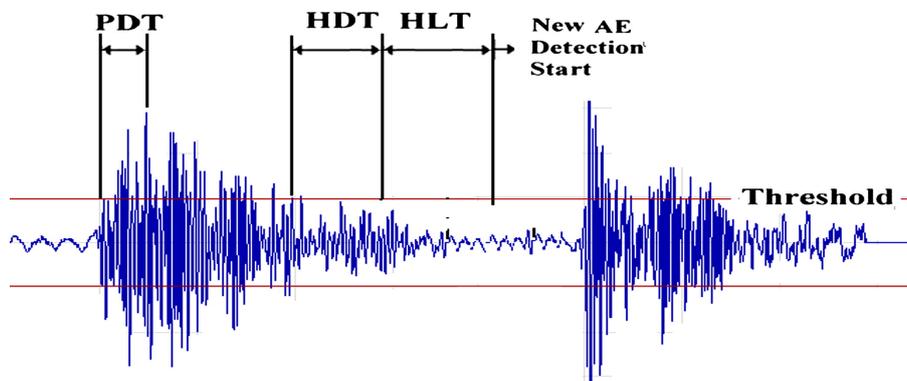


Figure 1.3. – Illustration de la stratégie de détection des salves par seuillage (©[31])

Ces paramètres d'acquisition pilotent la définition des salves EA. S'ils sont mal choisis, ils peuvent entraîner la troncature d'une salve avant sa fin, ou l'assimilation d'une succession de salves à une seule. Par ailleurs, l'application de cette stratégie sans traitement préalable du signal semble peu prudente, notamment dans les cas où le milieu de propagation et son environnement sont relativement complexes. Par exemple, lorsque le bruit environnemental est important, il devient nécessaire d'augmenter le seuil d'acquisition des signaux, avec comme conséquence une perte de sensibilité. Cette perte de sensibilité peut alors conduire à la non détection de certaines salves d'intérêt et à un manque de précision en terme de localisation de la source [32]. Comme mentionné précédemment, une démarche plus rigoureuse serait alors d'estimer d'abord le signal source, en supprimant toutes les composantes du signal reçu autres que celles associées à de potentiels mécanismes physiques d'intérêt.

a. Impact du bruit environnemental n

Traiter le bruit environnemental n nécessite souvent de faire appel à des méthodes de traitement du signal plutôt avancées. Selon les caractéristiques du bruit, stochastique ou déterministe, stationnaire ou non stationnaire, large bande ou à bande étroite, plusieurs méthodes de débruitage existent. BARRAT, BORODIN & KUZMIN [33] en proposent une classification en trois groupes (cf. Tableau 1.1). Les sources du bruit pouvant être diverses et variées, ces méthodes peuvent être utilisées de manière complémentaire. Selon les cas, il convient d'identifier tous les types de bruits en présence et de procéder à un traitement par étapes.

Tableau 1.1. – Classification des méthodes de traitement du bruit selon BARAT, BORODIN & KUZMIN [33]

Type de bruit	Source potentielle	Méthodes de traitement du bruit	Difficulté de mise en œuvre
— Basse fréquence — Haute fréquence	— Bruit électrique — Bruit électromagnétique	— Filtre à réponse impulsionnelle — Filtrage passe-bande	— Faible
— Bruit blanc	— Composantes électroniques — Signaux acoustiques	— Méthodes basées sur les transformées en ondelette	— Moyenne
— Impulsionnel — Non stationnaire	— Frottements — Bruits hydrodynamiques — Phénomènes de cavitation	— Analyse fine du signal sur une longue durée, puis choix d'une stratégie	— Élevée

b. Impact du chemin de propagation des ondes (Fonction de transfert h_1)

D'après l'équation (0.1) (cf. Introduction), si on souhaite estimer le signal source essai s , il est nécessaire de déconvoluer le signal essai reçu x par la fonction de transfert h_1 de la structure testée. Cette fonction peut être vue comme la réponse de la structure à une sollicitation impulsionnelle. En d'autres termes, c'est la fonction de Green associée à la configuration étudiée. Il est particulièrement important d'évaluer la fonction de transfert h_1 et, de fait, d'explorer dans quelle mesure le comportement des ondes est affecté par les propriétés géométriques et physiques de la structure, ainsi que par la localisation de la source et celle des capteurs. Cette opération, loin d'être triviale, se conduit généralement au travers d'approches analytiques, expérimentales ou numériques, selon la complexité de la configuration étudiée.

Les propriétés physiques et géométriques de la structure vont impacter non seulement la propagation, mais également la nature même des ondes. On sait que les ondes (de compression) P peuvent se propager quel que soit le milieu (fluide ou solide), alors que les ondes (de cisaillement) S n'évoluent qu'au sein de matériaux solides. Ces ondes dites de volume se réfléchissent et se transmettent, en interagissant entre elles, au niveau des discontinuités d'impédance (ou interfaces). Elles transportent avec elles plus ou moins d'énergie en fonction du contraste d'impédance entre les milieux, placés de part et d'autre de l'interface. Elles ont aussi une vitesse et un contenu fréquentiel propres dépendant des propriétés physiques des milieux traversés. Selon la géométrie de la structure, et sous certaines conditions, ces ondes peuvent même se combiner et donner naissance à des ondes de plaque (par exemple, les ondes de Lamb au sein d'une structure d'épaisseur de l'ordre de la longueur d'onde), des ondes de surface (par exemple, l'onde de Rayleigh dans une structure où l'épaisseur est supérieure à la longueur d'onde) ou des ondes de torsion/flexion (dans le cas d'une structure cylindrique). Ces ondes guidées ont des vitesses propres, différentes de celles des ondes P et S . Si la propagation est modale, outre les modes fondamentaux qui transportent une large partie de l'énergie mise en jeu, peuvent apparaître également des modes d'ordre élevé à certaines fréquences (dites de coupure). Il n'est pas question ici de dresser une liste exhaustive des ondes guidées et de leurs propriétés. Pour cela, le lecteur pourra se référer à certains ouvrages essentiels sur la question [34–37].

c. Impact du capteur (Fonction de transfert h_2)

Pour corriger l'impact du capteur, il "suffit" de déconvoluer le signal reçu par la courbe d'étalonnage de celui-ci. Nous noterons cependant que dans certains cas (mauvaise qualité des courbes d'étalonnage, par exemple), cette opération de déconvolution est difficile. L'expertise de l'expérimentateur peut alors aider à identifier les gammes de fréquences du signal reçu qui sont susceptibles d'être for-

tement impactées par le système d'acquisition. C'est le cas, par exemple, de la gamme de fréquences autour de la fréquence de résonance du capteur. Par la suite, des méthodes de traitement du signal du type filtrage passe-bande peuvent minimiser ces impacts.

1.2.2. Phase 2 : Analyse et interprétation des salves

1.2.2.1. Localisation du mécanisme physique

Selon le nombre de capteurs et la qualité du signal source estimé, plusieurs approches peuvent être envisagées pour la localisation du mécanisme physique associé à une salve donnée. Dans le cas où l'expérimentateur ne disposerait que d'un seul capteur, des méthodes comme celles utilisant les différences de vitesse de propagation selon la nature de l'onde (Figure 1.4) ou les courbes géodésiques peuvent être envisagées. Dans le premier cas, le signal exploité doit permettre une identification claire du pic correspondant à chaque type d'onde [38]. Dans le cas des courbes géodésiques, l'hypothèse de base est qu'une onde prend toujours le chemin de propagation minimisant la perte d'énergie [39]. Connaissant la géométrie et la configuration de la structure testée, des outils basés sur la théorie des graphes permettent d'estimer ce chemin. Il suffit ensuite de le remonter pour localiser la source.

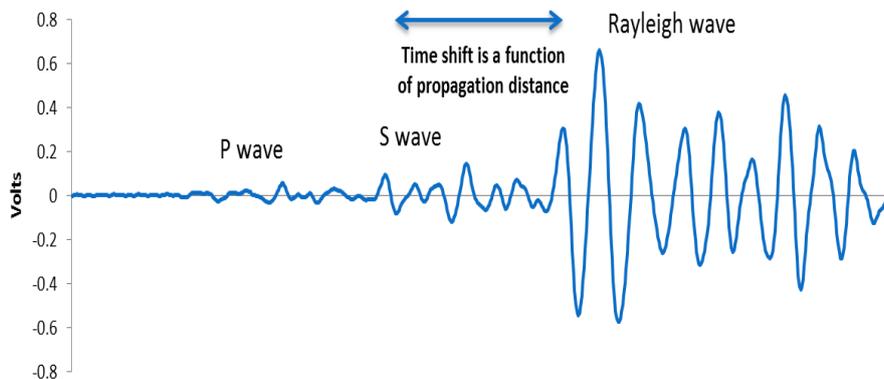


Figure 1.4. – Illustration d'un processus de localisation de source par utilisation des différences de temps d'arrivée selon le type d'onde (©<http://muravin.com/>)

Lorsque le nombre de capteurs est supérieur ou égal à deux, outre les méthodes géodésiques, d'autres techniques comme celles basées sur l'atténuation de l'énergie sont envisageables. Néanmoins, dans la littérature, les méthodes les plus communément utilisées sont celles basées sur l'estimation des temps d'arrivée (ou décalages). Parmi ces dernières, la fonction d'intercorrélacion est l'outil auquel les auteurs ont le plus souvent recours, notamment lorsque le nombre de capteurs utilisés est très faible. Nous noterons cependant que cette fonction peut s'avérer inefficace dans certaines situations, telles qu'en présence de plusieurs sources [40] ou encore en cas de mauvaise estimation du signal source. Dans ces cas, un prétraitement des signaux est nécessaire. Cela conduit à une généralisation de l'approche d'estimation du décalage par intercorrélacion (Generalized Cross Correlation), en pré-filtrant les signaux, afin d'en pénaliser certaines composantes [41]. Selon les cas, diverses méthodes d'intercorrélacion généralisée peuvent être envisagées. Par exemple, en cas de pollution importante du signal, l'utilisation de la fonction de cohérence ou de la fonction de phase est réputée efficace [42].

1.2.2.2. Identification de mécanismes physiques

L'objectif de l'étape d'identification est d'associer chaque salve à un mécanisme physique source. Cette démarche d'identification repose sur l'hypothèse fondamentale que **chaque salve enregistrée possède des propriétés liées aux caractéristiques de la source l'ayant engendrée** [43].

En général, la forme d'onde d'une salve est complexe ; son utilisation directe pour identifier le mécanisme physique associé est donc délicate. En pratique, on privilégie plutôt des approches permettant de caractériser chaque salve de manière à attribuer une signature acoustique au mécanisme physique correspondant. Ces méthodes de caractérisation s'appuient sur la forme temporelle, le spectre fréquentiel ou encore une décomposition du signal (coefficients d'ondelettes par exemple). L'approche la plus utilisée dans la littérature consiste à calculer des descripteurs quantitatifs à partir du signal temporel et de son spectre (Figure 1.5). Parmi ceux-ci, nous pouvons par exemple citer l'amplitude, le temps de montée, le nombre de coups ou encore la fréquence moyenne du signal [22, 24, 44]. Nous fournissons en Annexe A.1, une liste plus complète (mais non exhaustive) de ces paramètres. Par ailleurs, notons que l'expertise de l'expérimentateur est une ressource cruciale pour exploiter convenablement chaque signature acoustique, afin de l'associer ou non de manière certaine à un mécanisme physique précis.

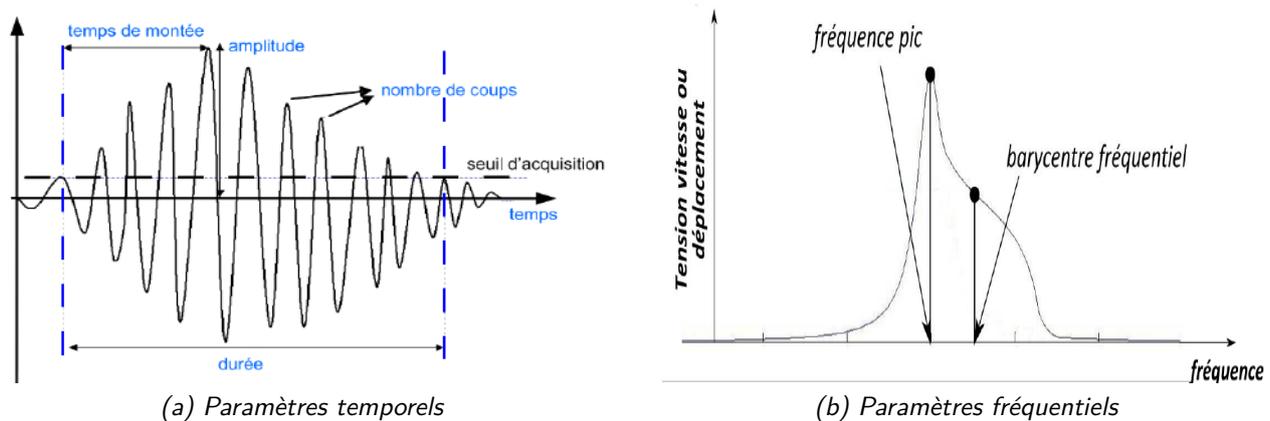


Figure 1.5. – Exemples de paramètres EA quantitatifs calculés à partir du signal temporel et son spectre ([45] and [43]).

Lorsque le nombre de salves détectées au cours d'un contrôle devient très important, ce qui est souvent le cas, un recours aux méthodes statistiques est nécessaire. Depuis le début des années 1980, la littérature consacrée à l'analyse des données EA s'est intéressée à la plupart des méthodes d'analyse exploratoire, de clustering et de classification supervisée. Il est difficile de faire un état de l'art exhaustif ici. Nous adopterons donc une démarche permettant d'aborder la question centrale du choix de variables (paramètres) discriminantes, avant d'établir une hiérarchie entre les méthodes les plus courantes dans la littérature.

a. Choix des variables caractéristiques

Selon la nature des phénomènes d'intérêt, les variables n'ont pas la même capacité de discrimination [46–51]. Il peut également arriver que plusieurs variables, bien que discriminantes chacune, expliquent exactement le même phénomène. La question du temps de calcul doit également être prise en compte, notamment dans le cas de problématiques de contrôle continu. Toutes ces raisons font que pour une classification réussie, il convient de consacrer beaucoup de temps à la recherche de variables pertinentes. Dans certains cas, la seule expertise de l'expérimentateur est suffisante [52–54]. Cependant, l'approche méthodologiquement reproductible est d'avoir recours à des méthodes similaires aux techniques de choix de modèle optimal en régression ou méthodes transformationnelles, basées sur des techniques de statistique exploratoire [52, 55, 56].

Influence de la nature des variables

Comme souligné ci-dessus, le signal EA peut être caractérisé grâce à des variables temporelles, fréquentielles, temps-fréquence ou encore par des variables dont l'interprétation physique n'est pas im-

médiate (par exemple, les coefficients d'un modèle Auto-Régressif). Dans tous les cas, il convient de s'assurer que, parmi les variables choisies au départ, figurent celles à même de retranscrire l'information nécessaire à l'identification des mécanismes physiques d'intérêt. Par exemple, en ce qui concerne les mécanismes de déformation d'alliages d'aluminium, en s'appuyant sur un échantillon très hétérogène (déformation plastique, fracture, éjection de gaz, bruits électroniques, cassé de mine...), et sur une trentaine de variables aussi bien temporelles que fréquentielles, HAY *et al.* [57] concluent que le contenu fréquentiel est prépondérant pour une classification réussie.

Influence du nombre de variables

L'influence du nombre de variables est assez peu évoquée dans la littérature, mais de manière générale, on constate que le meilleur algorithme de classification nécessite souvent au plus une dizaine de variables explicatives [57–59]. Le plus souvent, on peut définir un nombre maximal de variables au-delà duquel il n'y a plus d'amélioration de la qualité de la classification. Il semblerait qu'en matière de choix des variables caractéristiques, il n'existe pas de méthode standard. Outre le recours très fréquent au savoir-faire de l'expérimentateur, les méthodes les plus utilisées sont l'optimisation des variances intra et inter classes, l'ACP et l'analyse de la matrice de corrélation des individus.

b. Méthodes de reconnaissance de formes

Depuis le début des années 1980, plusieurs auteurs ont procédé à des études comparatives visant à identifier les meilleurs outils en termes d'analyse exploratoire, de clustering et de classification non supervisée [46, 47, 53, 57]. Jusqu'à la fin des années 1990, les points de vue sont plutôt convergents sur les méthodes à adopter. Ainsi, l'ACP s'impose de manière unanime comme LA méthode d'analyse exploratoire. Pour le clustering, les méthodes les plus appréciées sont la CAH et les k-means. Par ailleurs, LA méthode de classification supervisée par excellence est celle des k-plus-proches voisins (kNN).

Plus récemment, l'essor de nouvelles méthodes de classification, dont les réseaux de neurones, n'a bouleversé que très peu cette hiérarchie. En effet, l'ACP, la CAH et les k-means restent toujours les méthodes les plus appréciées en matière d'analyse exploratoire et de clustering [60–63]. Par ailleurs, les réseaux de neurones s'imposent de plus en plus en matière de classification supervisée. Ils sont également très utilisés comme outil de clustering avec des résultats très satisfaisants [59, 64, 65]. Cependant, on notera que les résultats issus de l'utilisation des réseaux de neurones, comme outils de classification supervisée, ne permettent pas d'établir une hiérarchie claire entre ceux-ci et la méthode des kNN [66, 67]. Leur utilisation croissante s'explique donc essentiellement par un effet de mode.

1.3. Utilisation de l'EA en contexte nucléaire

L'EA, comme la plupart des techniques de CND, intervient dans toutes les phases du processus de production industrielle. On retrouve dans la littérature des applications allant du suivi de la fabrication des produits ou matériaux à la surveillance après mise en service. L'industrie nucléaire fait partie des premières à s'être intéressée à l'important potentiel de l'EA [23], et ce, dès le début des années 1960. La documentation relative aux applications de l'EA dans ce domaine est abondante et concerne essentiellement la surveillance du circuit primaire [68–70] et de la cuve sous pression [16–18]. On retrouve également quelques rares travaux sur la surveillance du comportement du combustible [11, 71]. Les objectifs et le niveau de précision demandés à la technique varient selon la composante à contrôler.

1.3.1. Surveillance du circuit primaire et de la cuve sous pression

Les premiers articles traitant de la surveillance de la cuve sous pression et du circuit primaire par des techniques EA remontent à la fin des années 1970 [17]. Les premières années, plusieurs facteurs, parmi lesquels la mauvaise compréhension des bases physiques de l'EA et la pression commerciale, ont conduit à des travaux manquant parfois de rigueur scientifique [16]. Certains articles datant de cette époque sont donc à considérer avec réserve.

1.3.1.1. Détection de fuites

Le problème de la détection des fuites et fissures est d'une importance capitale en contexte nucléaire. En effet, une perte de réfrigérant peut avoir des conséquences allant jusqu'à la fonte du cœur du réacteur. De manière classique, les fuites sont détectées par analyse des changements d'amplitude et de contenu fréquentiel du signal. Plusieurs travaux permettent de dégager les paramètres ayant un impact sur le contenu fréquentiel et l'énergie du signal en cas de fuite [69, 70, 72, 73].

- Pour une taille de fissure donnée, l'EA augmente avec la température. Mais dans le cas des réacteurs à eau sous pression, les fluctuations de température n'affectent pas de manière significative les paramètres du signal EA [70].
- Un choix optimal de la bande passante exige une comparaison des contenus fréquentiels correspondant à un bruit de fonctionnement et à une fuite. En général, les fréquences correspondantes à une fuite sont plus élevées que celles du bruit de fonctionnement [69, 72]. Dans le cas de la fissuration par corrosion sous tension par exemple, le signal correspondant au bruit connaît une forte atténuation pour les fréquences supérieures à 200kHz [69]. Pour ces fréquences supérieures à 200kHz , la géométrie de la source n'affecte pas de manière significative l'énergie du signal.
- Le signal est affecté par la taille de la fuite. En général, à température constante, l'amplitude du signal augmente avec la taille de la fuite. Pour une taille de fissure entraînant une fuite de débit, KUPPERMAN, CLAYTOR & GROENWALD [69] arrivent sous des conditions très précises à établir une relation liant le débit à l'amplitude du signal.
- D'autres paramètres, comme la présence d'un guide d'onde, d'interférence électrique ou l'isolation, peuvent également avoir un impact sur le signal EA. Par exemple, une mauvaise isolation des circuits peut avoir comme conséquence une augmentation de l'amplitude du signal.

1.3.1.2. Estimation de la gravité de la fuite

L'estimation de la gravité des fuites est difficile à cause de la dépendance entre les paramètres EA et plusieurs facteurs, dont la géométrie de la fissure, sa taille et dans une moindre mesure, la température. KUPPERMAN, CLAYTOR & GROENWALD [69] proposent l'idée d'utiliser la courbe d'évolution de la valeur efficace du signal (RMS) en fonction du temps pour déterminer la gravité des fuites. Elle repose sur l'axiome qu'une diminution de la variance de cette valeur indique une augmentation de la taille de la fissure, et donc de la gravité de la fuite.

Actuellement, la prévention des accidents graves liés à la rupture du circuit primaire et provoqués par des fuites, s'effectue au travers du critère dit de fuite avant rupture (Leak Before Break) [74]. Cette méthode repose sur l'axiome qu'il existe une relation de cause à effet entre la défaillance du circuit et les caractéristiques de la fuite. Ainsi, des estimateurs statistiques basés sur un certain nombre de mesures, avec recours ou non à l'EA, permettent de lancer l'alerte avant que la fuite n'atteigne une taille critique fixée [75–79].

1.3.1.3. Détection d'ébullition du fluide caloporteur

La détection de phénomènes de changement d'état du fluide caloporteur occupe également une place importante en sûreté nucléaire. Dans leurs travaux parus en 2011, CARMÍ *et al.* [80] s'intéressent à l'application de l'EA comme outil de caractérisation du processus d'ébullition de l'eau. En s'appuyant sur plusieurs travaux précédents [81–85], les auteurs établissent des relations entre certaines caractéristiques du signal EA (forme d'onde et contenu fréquentiel) et différents mécanismes d'ébullition identifiés par Nukiyama [86].

De manière générale, on note que les fréquences caractérisant des signaux issus du processus d'ébullition sont supérieures à 50 *kHz*. De plus, il est possible d'établir une relation entre la taille des bulles et les fréquences observées. Ainsi, des bulles de l'ordre de 6 μm se caractérisent par des fréquences d'environ 500 *kHz*. Ces fréquences ne sont que de l'ordre de 300 *kHz* lorsque la taille des bulles atteint 10 μm . De manière plus spécifique :

- Pendant la phase dite de convection naturelle, le nombre de coups du signal est constant. La transition entre cette phase et la phase d'ébullition nucléée se caractérise par une augmentation d'amplitude.
- La phase d'ébullition nucléée se traduit par une fluctuation de l'activité acoustique en termes de nombres de coups et d'amplitude du signal, ce qui correspond à un signal avec des salves sporadiques. Le spectre devient alors moins large que pendant la phase de convection naturelle, sans qu'on ne puisse dégager de fréquences caractéristiques du signal.
- La transition vers la phase de coalescence des bulles se traduit par une nouvelle augmentation de l'amplitude du signal. Les salves sporadiques deviennent donc plus énergétiques et on distingue dans ce dernier cas une fréquence caractéristique autour de 300 *kHz*.

On notera que les auteurs s'intéressent également aux phénomènes physiques liés à ces différentes phases, ainsi qu'à l'impact de l'environnement. Pour conclure, bien que les résultats de ces différents travaux semblent tout à fait satisfaisants, la problématique de l'automatisation de la détection des phénomènes d'ébullition, à travers notamment des algorithmes statistiques s'appuyant sur le signal EA, reste encore à résoudre avant de pouvoir envisager une mise œuvre en centrale nucléaire.

1.3.2. Surveillance du comportement du combustible nucléaire

En matière de suivi de fabrication et contrôle du combustible, l'utilisation de l'EA n'en est qu'à ses débuts. L'utilisation de la méthode pour le contrôle du combustible dans les réacteurs en activité n'est quasiment pas évoquée dans les travaux existants. Cela s'explique notamment par le nombre très important (plusieurs centaines) de crayons et la complexité des assemblages regroupant ces crayons. La méthode intervient essentiellement lors du suivi de fabrication du combustible. Il s'agit ici d'utiliser l'EA pour mieux comprendre les mécanismes de compaction et le comportement de la gaine. Ainsi, en établissant une corrélation entre l'EA et un mécanisme donné, l'expérimentateur est capable de piloter le pilotage du processus de fabrication.

1.3.2.1. Comportement du combustible

Dans ses travaux de thèse, HÉGRON [24] étudie l'EA pendant la compaction de particules fragiles d' UO_2 . En s'appuyant sur les travaux de WARING, RUBINSTEIN & HOWARD [20], EISSA [21] & KERBOUL [87], l'auteur rappelle que la compaction de particules de l'ordre de centaines de micromètres se divise en 3 étapes :

1. Un régime de densification sous faibles contraintes pendant laquelle l'activité acoustique demeure très faible. Cette phase correspond essentiellement à des mécanismes de réarrangement par glissement et à des fragmentations de granulés peu cohésifs ;

2. Un régime de compaction intermédiaire où le nombre de coups, ainsi que l'amplitude des signaux augmentent. Cette augmentation est d'autant plus rapide que les granulés sont cohésifs ou que leur diamètre moyen est grand ;
3. Un régime de consolidation correspondant à la fois à des phénomènes de réarrangement et fragmentation. Pendant cette phase, l'amplitude des signaux diminue fortement à partir d'une contrainte appliquée qui dépend, entre autres, de la cohésion et de la taille des granulés.

Quelles que soient les caractéristiques du matériau granulaire compacté, l'évolution de l'EA est similaire. Seules les contraintes délimitant les différents régimes cités évoluent avec la nature, la cohésion, la forme et la taille des particules [20, 88, 89]. Dans le cas particulier de l' UO_2 , la fragmentation des particules se caractérise par un signal dont l'enveloppe a un temps de montée très court et une décroissance quasi-exponentielle de l'amplitude [11]. Par contre, aucune signature acoustique indiscutable n'a pu être attribuée aux phénomènes mécaniques de type réarrangement granulaire, frottements... Enfin, pour caractériser les phénomènes liés au comportement du combustible, il semble que les paramètres liés à l'enveloppe du signal soient les plus pertinents [11]. L'utilisation du contenu fréquentiel donne des résultats peu fiables car le milieu étudié évolue fortement au cours de l'essai de compaction, passant d'un milieu granulaire lâche en début d'essai à un milieu poreux consolidé en fin d'essai. La propagation des ondes s'en trouve donc affectée.

1.3.2.2. Comportement de la gaine

Généralement en alliage de zirconium (ou matériaux dérivés), la gaine est considérée comme la première barrière de confinement. A ce titre, sa fiabilité est essentielle pour la sûreté des installations nucléaires. Elle peut être soumise à des situations de stress extrême, notamment lors d'un RIA ou d'un accident de perte de réfrigérant primaire (Loss Of Coolant) [1, 90].

La recherche consacrée à la caractérisation et à l'étude du comportement de la gaine est très dynamique depuis la fin des années 1970 [91–99]. Cependant, la majeure partie des travaux s'appuie sur des méthodes autres que l'EA. A ce jour, très peu d'informations sont disponibles sur la signature acoustique des phénomènes liés au comportement de la gaine [100–104]. Par exemple, PERES *et al.* [105] étudient l'oxydation sous haute température du zircaloy-4 en combinant l'EA et la thermogravimétrie. Deux mélanges de gaz ont été utilisés pour réaliser deux séries tests permettant d'étudier le comportement de l'alliage. Ainsi, pour chaque type de gaz, les auteurs ont observé le comportement de la gaine d'une dimension de $4.8 \times 4.6 \times 0.5 \text{ mm}^3$ pendant une phase isotherme de 5h à 10h, à une température de 900°C , puis pendant la phase de refroidissement.

1.4. Conclusion

Dans ce chapitre, après quelques rappels généraux sur l'EA et ses principes, nous avons exposé le processus complexe de la mise en œuvre d'un contrôle par EA. D'un point de vue méthodologique, ce processus peut être divisé en deux phases. Dans un premier temps, il s'agit d'acquérir et de préparer les données EA. Dans un second temps est abordée la phase d'analyse et d'interprétation de ces données, consistant, selon les objectifs du contrôle, à identifier les signaux d'intérêt, à localiser leurs sources, à leur attribuer des mécanismes physiques et à évaluer la gravité de ceux-ci.

Pendant la première phase, chaque étape, du choix des capteurs EA au prétraitement des signaux acquis, doit être pensée à la lumière de contraintes propres au contexte de l'étude. Le signal EA, issu du processus d'acquisition et contenant l'information transmise par la source, est affecté par les propriétés physiques, géométriques et la configuration de la structure contrôlée, par le système d'acquisition et par le bruit environnemental. Nous avons vu que la suppression ou la pénalisation des

biais introduits par ces facteurs constituent des défis techniques et méthodologiques importants. Si l'état de l'art sur l'application de l'EA en contexte nucléaire permet de mettre en exergue la nécessité de mieux traiter ces facteurs polluants, peu de travaux semble s'attacher à proposer une démarche méthodologique rigoureuse. Par conséquent, des questions de fond subsistent toujours. Par exemple, en mettant en œuvre une stratégie visant à identifier et supprimer l'impact de chaque facteur polluant, peut-on obtenir un estimateur du signal source à même de mieux répondre aux attentes du contrôle ? Toute tentative de réponse à cette question ne peut qu'en soulever d'autres, notamment en termes de détection et d'identification de mécanismes sources.

Nous avons rappelé que l'identification des signaux d'intérêt repose en général sur une simple stratégie de seuillage. Cette stratégie peut s'avérer inefficace, notamment lorsque le bruit résiduel issu de la phase d'estimation du signal source est non négligeable. Dans un contexte nucléaire, quels outils peut-on alors envisager en complément d'un simple seuillage ? Quelles seraient les conséquences de l'utilisation de ces outils sur le reste du processus, notamment en terme de localisation du signal source ?

Quant à l'identification des mécanismes physiques associés aux signaux d'intérêt, elle repose généralement sur l'interprétation de paramètres EA calculés, avec un recours fréquent à des outils d'apprentissage statistique. L'efficacité de ces méthodes dépendant fortement des paramètres EA utilisés, l'utilisation de méthodes de sélection de variables doit donc être un réflexe lors de cette phase. Cependant, dans les cas où il y a une diversité importante de mécanismes sources à détecter, la sélection de variables peut s'avérer délicate. Peut-on envisager des méthodes alternatives d'identification de sources permettant de contourner la sélection de variables discriminantes ? Si oui, que valent-elles face aux approches classiques ? Quelles conséquences a la qualité de l'estimation du signal source sur les différents résultats ?

Dans la suite de nos travaux, notre ambition sera de proposer une démarche méthodologique permettant d'apporter des réponses à ces différentes questions dans le cas particulier des essais RIA.

2 ■ De la source aux capteurs : physique des ondes associée à la structure contrôlée

Le travail présenté dans ce chapitre a fait l'objet d'un article soumis dans IEEE TNS [106] et d'une communication [107]

Sommaire

2.1. Introduction	20
2.2. Simplification du dispositif d'essais	21
2.3. Étude analytique de la propagation des ondes au sein du dispositif d'essais simplifié	23
2.3.1. Matrice des fréquences	24
2.3.2. Cas des modes de torsion	26
2.4. Modélisation numérique de la propagation des ondes au sein du dispositif d'essais simplifié	27
2.4.1. Généralités sur la méthode des éléments finis spectraux	27
2.4.2. Mise en œuvre de la modélisation numérique	28
2.4.2.1. Création du maillage	28
2.4.2.2. Calculs numériques	28
2.4.3. Résultats	29
2.4.3.1. Fréquence de résonance du dispositif d'essais	29
2.4.3.2. Influence de la position de la source et du capteur sur la fonction de transfert du dispositif d'essais	29
2.4.3.3. Analyse fine des fonctions de transfert selon différentes gammes de fréquences	31
2.4.3.4. Impact d'un changement de fluide caloporteur	31
2.5. Conclusion	33

2.1. Introduction

L'équation (0.1) de l'introduction générale met en exergue la nécessité d'une estimation du signal source essai. Pour ce faire, de manière générale, il convient d'évaluer l'impact de la structure sur la propagation des ondes et donc sur le signal source reçu, c'est-à-dire estimer la fonction de transfert h_1 . De plus, une telle démarche permettrait de mieux appréhender la physique des ondes associée au contexte du contrôle, contribuant ainsi à la mise en place d'une stratégie efficace d'acquisition des signaux.

Dans le cas des essais RIA, cela est d'autant plus prégnant que pour la quasi-totalité des 14 essais réalisés à ce jour, l'analyse du bruit de fonctionnement du réacteur révèle la présence d'une composante fortement énergétique autour de 180 kHz (Figure 2.1). Ayant été enregistrée avant le début de l'essai, et donc avant le pic de puissance représentatif de l'accident RIA, cette composante ne peut pas correspondre à un mécanisme physique d'intérêt. Cependant, elle perdure pendant tout l'essai et est assez énergétique pour masquer l'information relative à d'éventuels mécanismes physiques d'intérêt. Il devient alors crucial, dans l'optique d'une caractérisation efficiente de ceux-ci, d'en identifier la source, afin d'éviter si possible sa présence pour les prochains essais ou d'utiliser des outils de traitement du signal pour la supprimer. Sachant qu'elle n'est pas liée à l'impact du système d'acquisition h_2 , l'équation (0.1) suggère que cette composante ne peut avoir que deux sources potentielles : le bruit environnemental n ou l'impact du dispositif d'essai h_1 . Afin de lever l'indétermination on va chercher à évaluer h_1 dans ce chapitre.

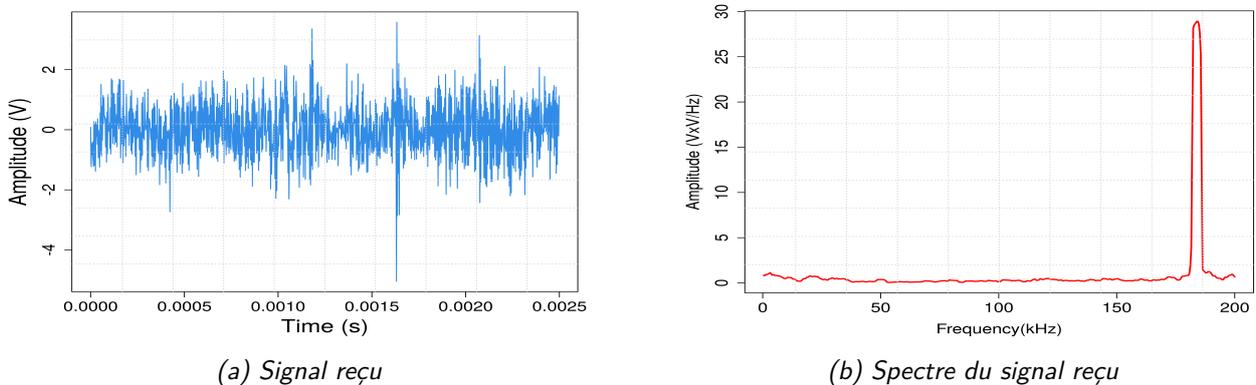


Figure 2.1. – Exemple de bruit de fonctionnement enregistré avant le début d'un essai RIA réalisé en 1997. Signal enregistré par le capteur 2.

Pour ce faire, trois principales approches peuvent être envisagées : expérimentale, analytique et numérique.

L'approche expérimentale, consistant à étudier la physique des ondes à partir, par exemple, de mesures d'acoustique active réalisées selon différents scénarios de propagation, est difficilement envisageable ici. En effet, en raison de l'environnement hostile, de la géométrie complexe du dispositif d'essai ou encore du besoin de contrôler à la fois la source et les capteurs, il est extrêmement fastidieux d'effectuer des mesures réalistes et en nombre suffisant pour une évaluation correcte de h_1 .

Lorsque la géométrie de la structure est équivalente ou peut être assimilée à un cas où on sait résoudre analytiquement l'équation d'ondes, l'approche analytique permet quant à elle d'obtenir l'équation des fréquences associée à celle-ci, et donc d'identifier les fréquences de résonance de la structure.

Enfin, plusieurs auteurs se sont intéressés au potentiel des outils numériques comme support à la compréhension de la physique des ondes dans une problématique d'EA. Des travaux portant aussi bien sur

la modélisation des sources que sur celui de la propagation d'ondes existent [43, 108] et montrent la capacité de l'outil numérique à simuler la propagation dans la gamme de fréquences associée à l'EA. De plus, ces méthodes permettent de réaliser des analyses fines, y compris pour des structures à géométrie complexe.

Dans ce chapitre, nous nous proposons d'expérimenter de manière complémentaire les approches analytiques et numériques, afin de cerner la physique des ondes associée aux essais RIA. Outre la détermination des possibles fréquences de résonance du dispositif d'essais, nous étudions l'impact de la localisation source/récepteur sur la fonction de transfert h_1 , avant d'analyser les conséquences d'un changement de fluide caloporteur. Par ailleurs, notons qu'une démarche qui se voudrait strictement rigoureuse en tenant compte du contexte RIA dans toute sa complexité serait extrêmement fastidieuse, sans pour autant être forcément concluante. Nous commençons donc par simplifier le contexte de l'étude en nous appuyant sur des considérations physiques et géométriques.

2.2. Simplification du dispositif d'essais

Dans ce qui suit, nous nous focalisons sur une région du dispositif d'essais (cf. Figure 0.2), longue de 3 mètres et incluant les positions des capteurs et celle du combustible testé. Ce choix est suffisant pour étudier la propagation d'ondes en tenant compte de toutes les positions potentielles des mécanismes sources d'intérêt. Dans cette région, la structure du dispositif correspond approximativement à un cylindre plus ou moins finement stratifié et composé de huit types de matériaux fluides (Sodium et Xénon) et solides (Uranium, Zircaloy, Molybdène, Niobium, Zirconium, Aluminium) (Figure 2.2). La présence d'une lame de Xénon, un gaz neutrophage ayant une impédance très largement inférieure à celles des matériaux environnants (Tableau 2.1), permet de simplifier grandement le dispositif (Figure 2.3). En effet, si on considère une onde incidente se propageant depuis le cœur du dispositif vers les couches externes, le contraste d'impédance entre les couches de Zircaloy et Xénon conduit à une transmission de l'onde quasi-nulle au-delà de la couche de Zircaloy, et par conséquent à une propagation d'ondes confinée dans un milieu stratifié à 4 couches.

Par ailleurs, la démarche d'étude complémentaire entre les approches analytique et numérique nous conduit à considérer deux modèles de niveaux de complexité différents :

- Le premier modèle, présenté en Figure 2.3a, correspond à un dispositif simplifié d'une longueur de $3m$ et d'un diamètre $17.2mm$. Il tient compte uniquement de la simplification relative à la barrière naturelle que constitue la lame de Xénon, le reste du dispositif étant décrit dans toute sa complexité. C'est ce modèle qui sera considéré plus loin pour les calculs numériques.
- Le second modèle, présenté en Figure 2.3b, correspond à un dispositif simplifié de mêmes dimensions que le précédent, mais cette fois-ci constitué uniquement de trois matériaux (Uranium, Zircaloy et Sodium). En effet, du fait des fréquences et des longueurs d'onde mises en jeu (respectivement autour de $200kHz$ et de l'ordre du centimètre), il semble raisonnable de négliger dans les calculs analytiques la présence de la gaine du combustible de Zircaloy, d'épaisseur ($0.625mm$) très inférieure à celle de la couche d'Uranium ($4.125mm$). Par ailleurs, le coefficient de transmission à l'interface Uranium/Inox étant très proche de l'unité, il ne semble pas déraisonnable de remplacer la couche d'Inox, positionnée à la base du dispositif, par une couche d'Uranium. Toutes ces simplifications permettent de développer aisément les calculs analytiques présentés ci-après.

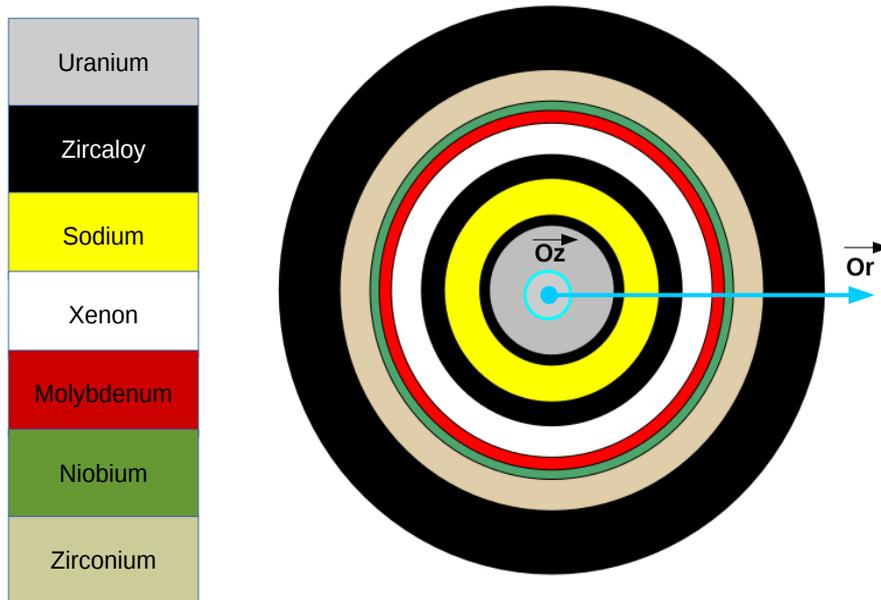
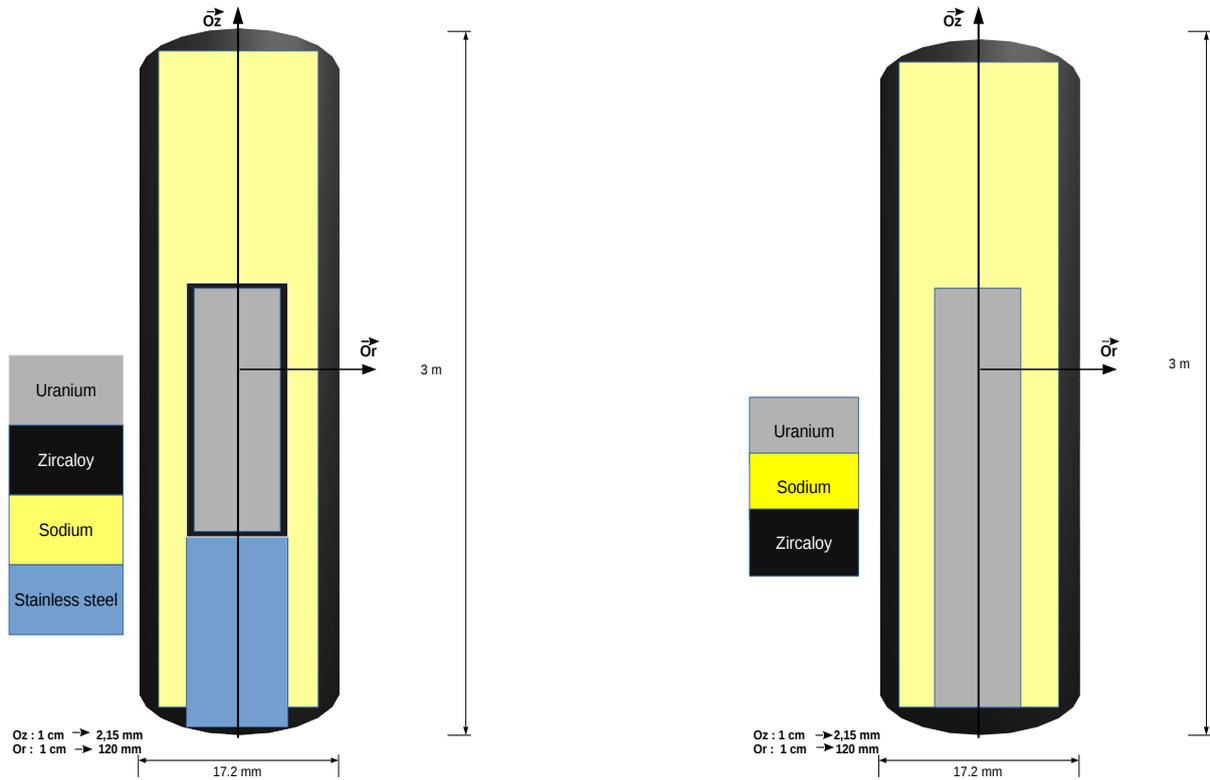


Figure 2.2. – Coupe radiale du dispositif d'essais (non simplifié) ($z = 0, \theta = 0$) illustrant les différents types de matériaux (ou strates) le constituant. Il est à noter que le nombre et l'ordre des strates peuvent changer selon la position sur l'axe Oz (cf. Figure 2.3a).

Tableau 2.1. – Propriétés physiques (densités, vitesses de propagation des ondes, et impédances) des différentes strates constituant le dispositif d'essais simplifié.

Matériau	ρ (kg/m^3)	c_P (m/s)	c_S (m/s)	Z_P	Z_S
Uranium	11000	3370	2020	$3,70 \cdot 10^7$	$2,22 \cdot 10^7$
Zircaloy	6500	4720	2360	$3,07 \cdot 10^7$	$1,53 \cdot 10^7$
Inox	7900	5460	3070	$4,31 \cdot 10^7$	$2,42 \cdot 10^7$
Sodium	968	2300	-	$2,23 \cdot 10^6$	-
Eau	1000	1436	-	$1,44 \cdot 10^6$	-
Xénon	6	1090	-	6540	-
Molybdène	10190	6679	3525	$6,81 \cdot 10^7$	$3,59 \cdot 10^7$
Niobium	8400	5068	2092	$4,26 \cdot 10^7$	$1,76 \cdot 10^7$
Zirconium	6400	4650	2300	$2,98 \cdot 10^7$	$1,47 \cdot 10^7$



(a) Dispositif simplifié considéré dans les calculs numériques

(b) Dispositif simplifié considéré dans les calculs analytiques

Figure 2.3. – Coupe longitudinale du dispositif d'essais simplifié, de longueur 3m et de diamètre 17.2mm.

2.3. Étude analytique de la propagation des ondes au sein du dispositif d'essais simplifié

Le ratio longueur/épaisseur du dispositif d'essais étant de 174, nous considérons dans cette partie que le dispositif d'essais simplifié représenté par la Figure 2.3b peut être approché par un cylindre stratifié infini (Uranium/Sodium/Zircaloy) de diamètre 17.2mm, renfermant en son sein un noyau rigide semi-infini suivant l'axe Oz (correspondant au combustible d'Uranium). La propagation des ondes dans les cylindres est une problématique largement étudiée dans la littérature. Ici, nous nous appuyons sur des résultats connus pour étudier le cas particulier associé au dispositif d'essais. Le lecteur pourra se reporter, par exemple, aux références [35, 109–111] pour de plus amples détails.

Soit le vecteur déplacement \vec{u} exprimé comme la somme d'un potentiel vectoriel $\vec{\psi}$ et d'un potentiel scalaire ϕ :

$$\vec{u} = \text{grad } \phi + \text{rot } \vec{\psi} \quad , \quad (2.1)$$

les deux satisfaisant l'équation d'ondes :

$$-\frac{1}{c^2} \frac{\partial^2 \xi}{\partial t^2} + \Delta \xi = 0 \quad (2.2)$$

où c est la vitesse de propagation des ondes de volume P ou S associées respectivement aux potentiels scalaire et vectoriel.

La résolution de l'équation (2.2) en coordonnées cylindriques, en utilisant la méthode de séparation des variables (cf. Annexe B.1), conduit aux expressions générales suivantes du potentiel scalaire ϕ et des composantes ψ_r , ψ_θ et ψ_z du potentiel vectoriel $\vec{\psi}$:

$$\begin{aligned}\phi(r, \theta, z; t) &= \phi_r(r) \cos(n\theta) e^{i(kz-\omega t)} \\ \psi_r(r, \theta, z; t) &= \psi_{rr}(r) \cos(n\theta) e^{i(kz-\omega t)} \\ \psi_\theta(r, \theta, z; t) &= \psi_{\theta r}(r) \sin(n\theta) e^{i(kz-\omega t)} \\ \psi_z(r, \theta, z; t) &= \psi_{zr}(r) \cos(n\theta) e^{i(kz-\omega t)}\end{aligned}\quad (2.3)$$

avec

$$\begin{aligned}\phi_r &= A_1 Z_n(\alpha r) + A_2 W_n(\alpha r) \\ \psi_{zr} &= D_1 Z_n(\beta r) + D_2 W_n(\beta r) \\ \psi_{rr} &= B_1 Z_{(n+1)}(\beta r) + B_2 W_{(n+1)}(\beta r) \\ \psi_{\theta r} &= -\psi_{rr}\end{aligned}\quad (2.4)$$

où $k = \omega/c$ est le nombre d'onde associé au mode de propagation, $\alpha^2 = k^2 - k_p^2$ et $\beta^2 = k^2 - k_s^2$, les indices p et s correspondant aux ondes P et S présentes dans le milieu solide. Par ailleurs, A_1 , A_2 , B_1 , B_2 , D_1 et D_2 sont des constantes inconnues, et Z_n et W_n les fonctions de Bessel du première et deuxième espèce d'ordre n .

La relation (2.1) permet alors d'écrire les composantes du vecteur déplacement \vec{u} en fonction de celles des potentiels scalaire et vectoriel :

$$\begin{aligned}u_r &= \left(\frac{\partial \phi_r}{\partial r} + \frac{n}{r} \psi_{zr} + i k \psi_{rr} \right) \cos(n\theta) e^{i(kz-\omega t)} \\ u_\theta &= \left(-\frac{n}{r} \phi_r + i k \psi_{rr} - \frac{\partial \psi_{zr}}{\partial r} \right) \sin(n\theta) e^{i(kz-\omega t)} \\ u_z &= \left(i k \phi_r - \frac{\partial \psi_{rr}}{\partial r} - \frac{n+1}{r} \psi_{rr} \right) \cos(n\theta) e^{i(kz-\omega t)}\end{aligned}\quad (2.5)$$

2.3.1. Matrice des fréquences

Le dispositif d'essais simplifié ayant été assimilé à un cylindre infini avec un noyau rigide semi-infini d'axe Oz (Figure 2.3b), nous avons [112, 113] :

$$\psi_r(r, \theta, z; t) = \psi_\theta(r, \theta, z; t) = 0 \quad (2.6)$$

Ainsi, l'expression (2.5) du vecteur déplacement \vec{u} devient :

$$\begin{aligned}u_r &= \left(\frac{\partial \phi_r}{\partial r} + \frac{n}{r} \psi_{zr} \right) \cos(n\theta) e^{i(kz-\omega t)} \\ u_\theta &= \left(-\frac{n}{r} \phi_r - \frac{\partial \psi_{zr}}{\partial r} \right) \sin(n\theta) e^{i(kz-\omega t)} \\ u_z &= i k \phi_r \cos(n\theta) e^{i(kz-\omega t)}\end{aligned}\quad (2.7)$$

Le noyau d'Uranium contenant l'origine de l'axe Oz , les fonctions de Bessel associées à la solution de l'équation différentielle de Bessel dans ce milieu se résument à celles du premier ordre. Notons par 1 l'indice de ce milieu. En combinant les expressions (2.4) et (2.7), nous obtenons alors :

$$\begin{aligned}
 u_r^1 &= \left[A_{11}\alpha_1 Z_n'(\alpha_1 r) + \frac{n}{r} D_{11} Z_n(\beta_1 r) \right] \cos(n\theta) e^{i(k_1 z - \omega t)} \\
 u_\theta^1 &= \left[-\frac{n}{r} A_{11} Z_n(\alpha_1 r) - D_{11} \beta_1 Z_n'(\beta_1 r) \right] \sin(n\theta) e^{i(k_1 z - \omega t)} \\
 u_z^1 &= i k_1 A_{11} Z_n(\alpha_1 r) \cos(n\theta) e^{i(k_1 z - \omega t)}
 \end{aligned} \tag{2.8}$$

où $Z_n' = \frac{\partial Z_n}{\partial r}$.

En supposant que le fluide caloporteur (en l'occurrence ici, le Sodium) est un fluide parfait, et en notant par 2 l'indice de ce milieu, seules se propagent les ondes de compression P et le déplacement est alors lié au potentiel scalaire ϕ_2 par :

$$u_r^2 = -\vec{grad} \phi_2, \tag{2.9}$$

où

$$\phi_2 = [E_{21} Z_n(\alpha_2 r) + E_{22} W_n(\alpha_2 r)] \cos(n\theta) e^{i(k_2 z - \omega t)} \tag{2.10}$$

avec E_{21} et E_{22} des constantes inconnues.

Enfin, si nous notons par 3 l'indice de la couche externe de Zircaloy, en combinant (2.4) et (2.7), nous avons :

$$\begin{aligned}
 u_r^3 &= \left[A_{31}\alpha_3 Z_n'(\alpha_3 r) + A_{32}\alpha_3 W_n'(\alpha_3 r) + \frac{n}{r} D_{31} Z_n(\beta_3 r) + \frac{n}{r} D_{32} W_n(\beta_3 r) \right] \cos(n\theta) e^{i(k_3 z - \omega t)} \\
 u_\theta^3 &= \left[-\frac{n}{r} A_{31} Z_n(\alpha_3 r) - \frac{n}{r} A_{32} W_n(\alpha_3 r) - D_{31} \beta_3 Z_n'(\beta_3 r) - D_{32} \beta_3 W_n'(\beta_3 r) \right] \sin(n\theta) e^{i(k_3 z - \omega t)} \\
 u_z^3 &= [i k_3 A_{31} Z_n(\alpha_3 r) + i k_3 A_{32} W_n(\alpha_3 r)] \cos(n\theta) e^{i(k_3 z - \omega t)}
 \end{aligned} \tag{2.11}$$

Rappelons que les conditions de continuité à une interface fluide/solide s'expriment par :

$$p^{fluide} = \sigma_{rr}^{solide} ; u_r^{fluide} = u_r^{solide} \text{ et } \sigma_{r\theta}^{solide} = 0 \tag{2.12}$$

et la condition de surface libre à une interface solide/vide par :

$$\sigma_{rr} = 0 \text{ et } \sigma_{r\theta} = 0 . \tag{2.13}$$

L'application des conditions (2.12) aux interfaces Uranium/Sodium et Sodium/Zircaloy et de la condition (2.13) à l'interface Zircaloy/Vide (cf. Annexe B.2) conduit à un système de huit équations à huit inconnues $A_{11}, D_{11}, E_{21}, E_{22}, A_{31}, A_{32}, D_{31}, D_{32}$ correspondant aux amplitudes des déplacements dans les différents milieux, tel que :

$$[\mathbf{M}] \vec{x} = 0 \tag{2.14}$$

où \mathbf{M} est une matrice de taille 8×8 dont les coefficients sont explicités en Annexe B.3 et \vec{x} un vecteur tel que :

$$\vec{x} = (A_{11}, D_{11}, E_{21}, E_{22}, A_{31}, A_{32}, D_{31}, D_{32})$$

Annuler le déterminant de \mathbf{M} conduit à exprimer l'équation des fréquences associée au dispositif. En fonction des différentes valeurs de n , la résolution de cette équation permet d'identifier les différents modes de propagation associés au dispositif d'essais, chaque mode ayant des propriétés particulières [35, 109, 110].

2.3.2. Cas des modes de torsion

Les modes de torsion correspondent aux solutions de l'équation des fréquences lorsque $n = 0$. Une des caractéristiques principales de ces modes est que le vecteur déplacement \vec{u} se résume alors uniquement à sa composante u_θ [35], le potentiel scalaire ϕ ainsi que les composantes ψ_r et ψ_θ du potentiel vectoriel $\vec{\psi}$ étant alors nuls. La matrice \mathbf{M} associée au dispositif d'essais s'écrit alors :

$$\mathbf{M} = \begin{pmatrix} a_{11} & 0 & a_{13} & a_{14} & 0 & 0 & 0 & 0 \\ a_{21} & 0 & a_{23} & a_{24} & 0 & 0 & 0 & 0 \\ 0 & a_{32} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & a_{43} & a_{44} & a_{45} & a_{46} & a_{47} & a_{48} \\ 0 & 0 & a_{53} & a_{54} & a_{55} & a_{56} & a_{57} & a_{58} \\ 0 & 0 & 0 & 0 & a_{65} & a_{66} & a_{67} & a_{68} \\ 0 & 0 & 0 & 0 & a_{75} & a_{76} & a_{77} & a_{78} \\ 0 & 0 & 0 & 0 & a_{85} & a_{86} & a_{67} & a_{68} \end{pmatrix} \quad (2.15)$$

Le déterminant de \mathbf{M} est nul si le coefficient a_{32} ou son cofacteur s'annule. Le coefficient a_{32} étant obtenu à partir de la condition de nullité de $\sigma_{r\theta}$ à l'interface Uranium/Sodium, en tenant compte du fait que si $n = 0$, $u_r^1 = u_z^1 = 0$, l'expression (2.8) et la loi d'Hooke permettent d'écrire :

$$\sigma_{r\theta} = \mu_1 \left[r_1 \frac{\partial}{\partial r} \left(\frac{u_\theta^1}{r} \right) \right] = D_{11} \mu_1 \beta_1 \left(\frac{1}{r_1} Z_0'(\beta_1 r) - \beta_1 Z_0''(\beta_1 r) \right) = 0 \quad (2.16)$$

où $Z_n'' = \frac{\partial^2 Z_n}{\partial r^2}$. On a donc

$$a_{32} = Z_0'(\beta_1 r_1) - \beta_1 r_1 Z_0''(\beta_1 r_1) = 0 \quad (2.17)$$

De l'expression de l'équation différentielle de Bessel (cf. Annexe B.4), on obtient

$$Z_0''(\beta_1 r_1) = \left(\frac{n^2}{(\beta_1 r_1)^2} - 1 \right) Z_0(\beta_1 r_1) - \frac{1}{\beta_1 r_1} Z_0'(\beta_1 r_1).$$

Sachant que $Z_0'(\beta_1 r_1) = -Z_1(\beta_1 r_1)$, nous concluons que :

$$\begin{aligned} Z_0'(\beta_1 r_1) - \beta_1 r_1 Z_0''(\beta_1 r_1) &= 0 \\ \iff (\beta_1 r_1) Z_0(\beta_1 r_1) &= 2Z_1(\beta_1 r_1) \end{aligned} \quad (2.18)$$

Cette dernière équation admet une infinité de solutions données par les zéros de la fonction de Bessel d'ordre 2. Cependant, elle ne tient pas compte du caractère semi-infini du cylindre d'Uranium.

ZEMANEK [110] et MASON [35, Page. 141-149] se sont intéressés à l'impact qu'une interface en $z = a$ (donc un cylindre semi-infini) aurait sur l'équation (2.18). Ils ont conclu que celle-ci provoquerait une forte concentration d'énergie. Pour certaines fréquences, cette concentration d'énergie provoquerait alors une forte amplification du déplacement, donc de l'amplitude du signal reçu. Ce phénomène, appelé «end resonance», permet alors de simplifier l'équation (2.18) comme suit :

$$(\beta_1 r_1) Z_0(\beta_1 r_1) = Z_1(\beta_1 r_1). \quad (2.19)$$

Connaissant la vitesse de propagation des ondes S dans l'Uranium, les racines de l'équation (2.19) permettent d'exprimer la fréquence associée à la première racine comme suit [114] :

$$f_0 = \frac{1.841c_s\sqrt{2}}{2\pi r_1} = \frac{1.841 \times 2020 \times \sqrt{2}}{2 \times 0.00475\pi} = 176.22 \text{ kHz} \quad (2.20)$$

Cette dernière valeur suggère que la fréquence fortement énergétique observée en préambule de la plupart des essais RIA (cf. Figure 2.1) est une fréquence de résonance du dispositif d'essais. Elle n'est en tout cas pas associée à un bruit environnemental.

2.4. Modélisation numérique de la propagation des ondes au sein du dispositif d'essais simplifié

Nous nous proposons d'étudier numériquement, au moyen de la méthode des éléments finis spectraux, la propagation des ondes au sein du dispositif d'essais. L'objectif ici est de tenter d'évaluer la fonction de transfert h_1 de la structure et l'impact des positions de la source et des capteurs sur les signaux reçus. L'influence de la nature du fluide caloporteur est également abordée. Nous considérons ici le dispositif simplifié présenté en Figure 2.3a.

2.4.1. Généralités sur la méthode des éléments finis spectraux

Nous procédons ici à un bref rappel sur la méthode des éléments finis spectraux, ou Spectral-Element Method (SEM), permettant ainsi de la positionner par rapport à d'autres approches plus classiques comme, par exemple, la méthode des différences finies ou la méthode des éléments finis. Depuis une vingtaine d'années, la SEM, méthode full-wave initialement introduite en mécanique des fluides, est multi-validée et appliquée avec succès pour la propagation des ondes et pour des problèmes d'imagerie dans des structures 3D complexes, notamment en géophysique [115], en acoustique sous-marine [116] et en CND [117]. Alors que pour résoudre le problème de propagation d'ondes, la méthode de différences finies s'appuie sur une discrétisation, par différences finies entre les points de grille voisins, des opérateurs différentiels de la formulation forte de l'équation de l'élastodynamique (ou équation des ondes) :

$$\rho \frac{\partial^2 u_i}{\partial t^2} = \partial_j \sigma_{ji} + f_i \quad (2.21)$$

où σ_{ij} représente les composantes du tenseur des contraintes et f_i les composantes du champ de force volumique (*i.e.* la source), la SEM est quant à elle fondée sur une formulation variationnelle (formulation faible) de cette équation :

$$\int_{\Omega} \left(\rho \frac{\partial^2 u}{\partial t^2} \cdot w \right) d\mathbf{x} + \int_{\Omega} \bar{\sigma} : \nabla w d\mathbf{x} = \int_{\Omega} (f \cdot w) d\mathbf{x} + \int_{\Gamma} (\tau \cdot w) d\gamma(\mathbf{x}) \quad (2.22)$$

où :

$$\int_{\Gamma} (\tau \cdot w) d\gamma(\mathbf{x}) = \int_{\Gamma_{int}} (\tau \cdot w) d\gamma(\mathbf{x}) + \int_{\Gamma_{abs}^{ext}} (\tau \cdot w) d\gamma(\mathbf{x})$$

et w est une fonction test quelconque appartenant à l'espace de Sobolev $H_0^1(\Omega)$ telle que :

$$H_0^1(\mathbf{x}) = \{w(\mathbf{x}) \in H^1(\mathbf{x}) : \Omega \rightarrow \mathbb{R}^n ; w(\mathbf{x}) = \vec{0} \text{ sur } \Gamma_g^{int}\} \quad (2.23)$$

La SEM combine la flexibilité d'une méthode d'éléments finis avec la précision d'une méthode pseudo-spectrale globale, ce qui la rend très attractive [118]. Le champ d'onde est représenté en

termes de fonctions interpolantes de Lagrange d'ordre élevé, et les intégrales sont calculées à partir de la quadrature de Gauss-Lobatto-Legendre, ce qui permet d'obtenir une matrice de masse parfaitement diagonale conduisant ainsi à un schéma numérique explicite en temps. Comme pour toute méthode d'éléments finis, la SEM repose sur un découpage de l'espace étudié selon un maillage en quadrangles (2D) ou hexaèdres (3D) qui s'adapte à toutes les discontinuités principales présentes, permettant ainsi le traitement de configurations aux géométries complexes.

Pour plus de détails concernant la SEM, le lecteur pourra se reporter, entre autres, aux travaux de thèse de KOMATITSCH [119].

2.4.2. Mise en œuvre de la modélisation numérique

2.4.2.1. Création du maillage

En raison de la géométrie cylindrique du dispositif d'essais simplifié et donc de la présence d'un axe de révolution (Oz), nous avons choisi de réaliser les simulations sur la base d'un modèle 2D axi-symétrique [120]. Cela est équivalent, pour les positions de source situées sur l'axe Oz , à faire une simulation 3D. Le maillage du modèle a été effectué avec le logiciel libre GMSH [121, 122]. Ce logiciel est bâti autour de quatre modules permettant de définir la géométrie du milieu étudié (module geometry), de générer le maillage selon des critères choisis (module mesh), de créer une interface avec un outil externe (module solver) et de gérer le maillage après création (module post-processing). Dans notre cas, les deux premiers modules ont suffi à la réalisation du maillage.

Typiquement, la taille d'éléments qui est requise pour générer un maillage correct est de l'ordre de la longueur d'onde λ_c associée à la plus petite des vitesses de propagation d'ondes et à la fréquence maximale considérée :

$$\lambda_c = \frac{\min(\text{vitesse})}{\max(\text{fréquence})} \quad (2.24)$$

Dans le cas du dispositif d'essais considéré ici, la plus petite vitesse de propagation observée est celle des ondes P dans le fluide caloporteur (2300 m/s pour le Sodium et 1436 m/s pour l'Eau), et la fréquence maximale des signaux attendue devrait se trouver autour de 200 kHz. En conséquence, la taille du plus petit élément du maillage devrait être au moins égale à $\lambda_c = 7\text{mm}$ dans le cas du fluide caloporteur eau (respectivement, 11 mm pour le sodium). Compte-tenu de l'épaisseur des strates des matériaux composant le dispositif d'essais, nous avons cependant choisi de considérer $\lambda_c = 1\text{mm}$. La fréquence maximale admissible est dans ce cas de 2.3 MHz pour une simulation avec un fluide caloporteur Sodium (respectivement, de 1.436 MHz pour l'Eau), ce qui couvre très largement les plages fréquentielles d'intérêt dans ce travail.

2.4.2.2. Calculs numériques

La version 2D axi-symétrique [120] du logiciel open-source SPEC-FEM, qui implémente la méthode des éléments spectraux, a été utilisée pour conduire les simulations numériques de la propagation des ondes dans le dispositif d'essais simplifié. Nous avons considéré une source ponctuelle avec un comportement de type Dirac en temps. En pratique, cela correspond à une gaussienne très étroite dont la réponse spectrale est plate sur une très large plage de fréquences, bien plus grande que la fréquence maximale considérée ici. Le choix de ce type de signal source est motivé par notre volonté de générer les fonctions de transfert associées à différentes combinaisons de positions source/capteur, et d'analyser ainsi l'impact d'un changement de position de source (respectivement, de capteur) sur le signal

reçu.

SPECFEM étant un code numérique dans le domaine temporel, il est nécessaire d'utiliser un schéma en temps pour calculer les solutions du problème posé. Ici, nous avons choisi un schéma de Newmark. En plus d'un échantillonnage spatial adéquat, il est aussi nécessaire d'avoir un échantillonnage temporel qui puisse assurer la stabilité des simulations numériques. La valeur du pas de temps Δt est alors déterminée par la condition CFL. Pour un schéma de Newmark, la condition CFL doit être inférieure à 0.5. En conséquence, nous avons choisi un pas de temps $\Delta t = 2 \cdot 10^{-9}s$, donnant ainsi une CFL de 0.3, ce qui assure la stabilité de nos simulations numériques.

2.4.3. Résultats

2.4.3.1. Fréquence de résonance du dispositif d'essais

La fonction de transfert du dispositif d'essais a tout d'abord été calculée numériquement dans le cas d'une source ponctuelle placée sur la gaine du combustible, le capteur étant positionné en $M2$ (cf. Figure 0.2). Son spectre normalisé met clairement en évidence, avec un léger décalage vers les hautes fréquences, la composante très fortement énergétique autour de $180kHz$ (Figure 2.4), déjà observée sur le spectre normalisé du bruit de fonctionnement, et confirmée par les calculs analytiques (cf. section 2.3.2). Cette composante est donc bien une fréquence de résonance du dispositif d'essais.

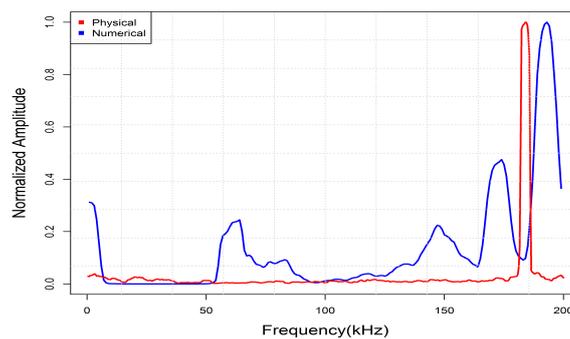


Figure 2.4. – *Superposition des spectres normalisés du signal correspondant au bruit de fonctionnement du réacteur (—) et de la fonction de transfert du dispositif d'essais simplifié (—) pour une source ponctuelle placée sur la gaine du combustible.*

2.4.3.2. Influence de la position de la source et du capteur sur la fonction de transfert du dispositif d'essais

Dans un second temps, le calcul numérique des signaux reçus associés à différentes combinaisons source/capteur nous a permis d'analyser les impacts d'un changement de position de source (respectivement, de capteur) sur ces signaux, et donc sur la fonction de transfert du dispositif. Plus spécifiquement, la Figure 2.5 illustre l'impact de la position de la source (respectivement, au sein du combustible, de sa gaine et au sein du canal sodium) sur le spectre de la fonction de transfert du dispositif, le capteur M1 étant fixe. L'impact de la position d'un même capteur peut, quant à lui, être observé sur la Figure 2.6.

On voit alors clairement qu'aucune fonction de transfert globale "universelle" ne peut être dégagée. En effet, les spectres associés ont des contenus fréquentiels très différents selon la position du binôme source/capteur. En particulier, pour un capteur donné, faire varier la position de la source d'une

strate du dispositif à l'autre revient à privilégier ou négliger certains phénomènes de propagation, et à faire varier la matrice des fréquences, aboutissant ainsi à l'émergence de différentes fréquences énergétiques dans le spectre. Par ailleurs, on note bien sur la Figure 2.6 que la position et le couplage des capteurs sur le dispositif ont un effet important sur le comportement des signaux reçus. Pour les futurs essais, il semble utile de s'assurer d'un couplage identique pour tous les capteurs. En conséquence, pour toutes ces raisons, il n'est pas envisageable ici d'avoir recours aux méthodes de déconvolution pour s'affranchir de l'impact de la structure et de la configuration sur le signal reçu x (cf. Équation (0.2)), si la position de la source n'est pas connue au préalable.

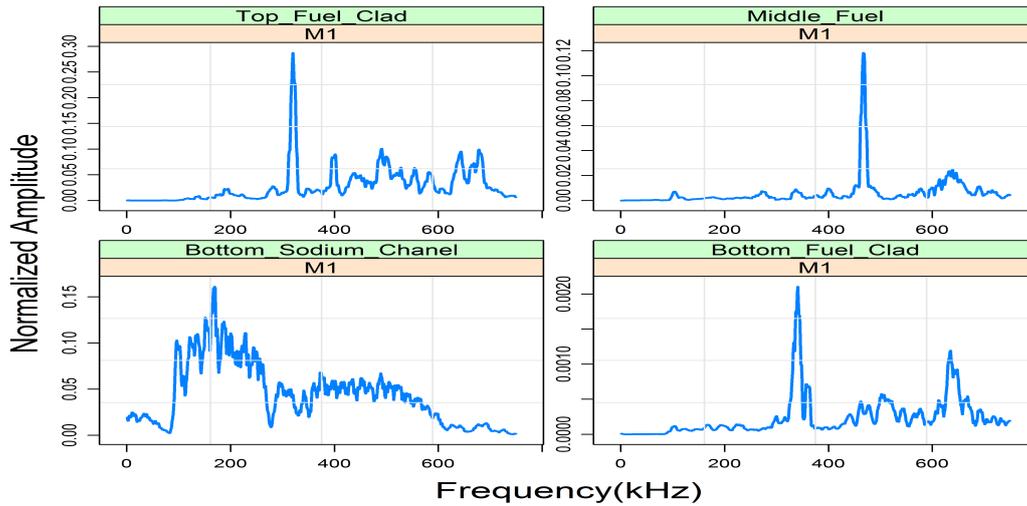


Figure 2.5. – Illustration de l'impact d'un changement de position de la source sur le spectre de la fonction de transfert du dispositif, le capteur M1 étant fixé.

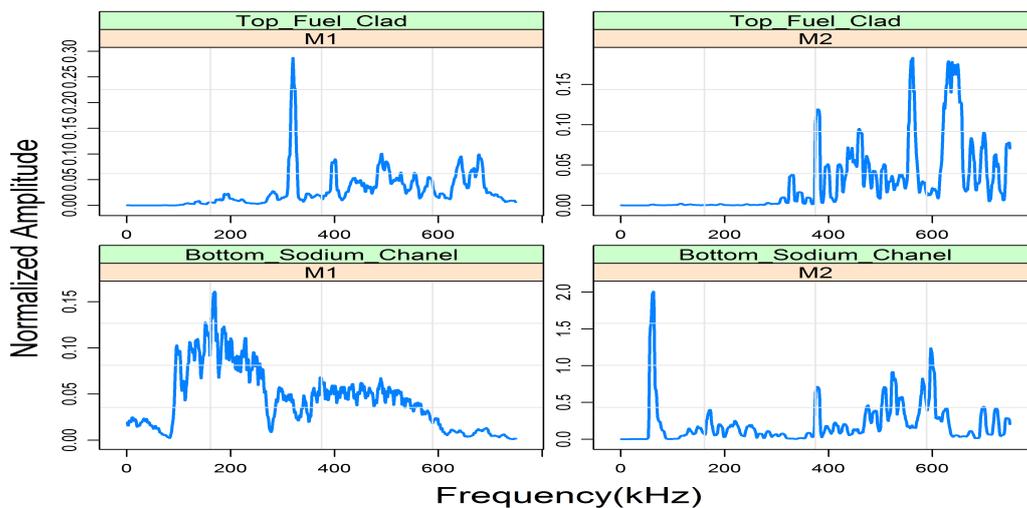


Figure 2.6. – Illustration de l'impact d'un changement de position du capteur sur le spectre de la fonction de transfert du dispositif, la source étant fixe (haut : au sommet de la gaine du combustible, bas : en partie basse du canal sodium)

2.4.3.3. Analyse fine des fonctions de transfert selon différentes gammes de fréquences

Considérons maintenant les fonctions de transfert correspondant à diverses positions de la source sur la gaine du combustible et à diverses positions du capteur (M1 ou M2). L'intérêt de localiser la source sur la gaine est fortement lié à l'existence possible, lors d'un RIA, de mécanismes physiques d'importance majeure tels que les ruptures de gaine.

L'analyse des spectres associés aux fonctions de transfert dans la gamme de fréquences $[0, 200]$ kHz met en évidence deux résultats intéressants (Figure 2.7). D'une part, les spectres, quasi-constants jusqu'à 50 kHz, indiquent l'absence d'impact significatif du dispositif d'essais dans cette gamme de fréquences. En conséquence, les composantes spectrales des signaux générés par les mécanismes sources localisés sur la gaine du combustible (par exemple, les ruptures de gaine), et enregistrés par les capteurs de fréquence de résonance 25 kHz, ne seront vraisemblablement pas affectées par le dispositif. Par ailleurs, la composante spectrale autour de 180 kHz est encore observée, et ce quelle que soit la position de du capteur.

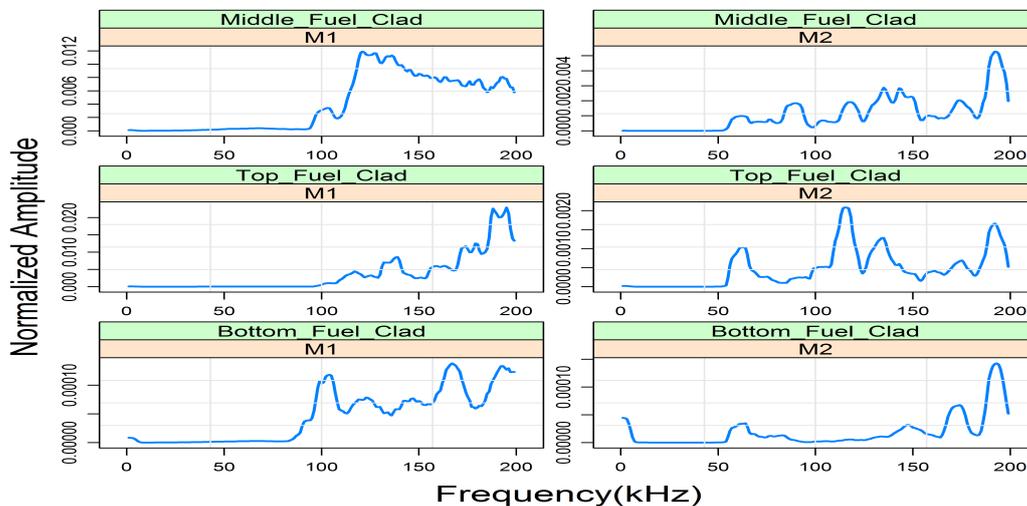


Figure 2.7. – Zoom, dans la gamme de fréquences $[0, 200]$ kHz, sur les spectres des fonctions de transfert pour différentes positions de source dans la strate correspondant à la gaine du combustible et pour deux positions du capteur (M1 et M2).

2.4.3.4. Impact d'un changement de fluide caloporteur

Pour évaluer l'impact d'un changement de fluide caloporteur sur la propagation des ondes, examinons la distribution de la pression au sein du dispositif d'essais lorsque la source est placée au centre de la couche de combustible et pour différents temps de propagation. La Figure 2.8 se focalise plus particulièrement sur le domaine de la Figure 2.3a, constitué d'une partie du support Inox et d'une partie du couple combustible (Uranium)/gaine (Zircaloy) (zone 1), du fluide caloporteur (zone 2), et de la couche extérieure de Zircaloy (zone 3).

Quelle que soit la nature du fluide caloporteur (eau ou sodium liquide), trois principales zones d'activité acoustique peuvent être distinguées.

Dans la zone 1, où se trouve la source, la distribution de la contrainte normale semble assez indépendante de la nature du fluide caloporteur, ce qui semble logique au vu du fort contraste d'impédance entre les milieux des zones 1 et 2 (cf. Tableau 2.1). Par ailleurs, aucune discontinuité notable entre les régions associées au support Inox et au combustible (Uranium)/gaine (Zircaloy) n'est observée,

confirmant ainsi la pertinence de nos hypothèses de simplification formulées dans les calculs analytiques qui assimilaient ces deux régions à une même région homogène équivalente à de l'Uranium. Dans la zone 2, l'effet de la nature du fluide caloporteur est notable, notamment aux temps courts de propagation (cf. instantanés à $t = 36\mu s$ de la Figure 2.8) pour lesquels les fronts d'onde, associés notamment aux champs transmis aux interfaces entre les zones 1-2 et 2-3, sont de manière logique plus rapides et plus développés dans le cas de la boucle sodium. Pour des temps plus longs (cf. instantanés à $t = 100\mu s$ de la Figure 2.8), pour lesquels le régime de propagation modale s'est établi dans les zones 2 et 3, on observe encore une fois des motifs assez identiques quel que soit le fluide caloporteur avec, en toute logique, des longueurs d'onde spatiales plus grandes dans le cas de la boucle sodium. En examinant de plus près les signaux reçus au capteur, et plus particulièrement leur spectre (Figure 2.9), on s'aperçoit alors que le changement de fluide caloporteur, d'une boucle eau sous pression vers une boucle sodium, induit une augmentation de l'amplitude du signal reçu et un léger décalage du contenu fréquentiel vers les hautes fréquences.

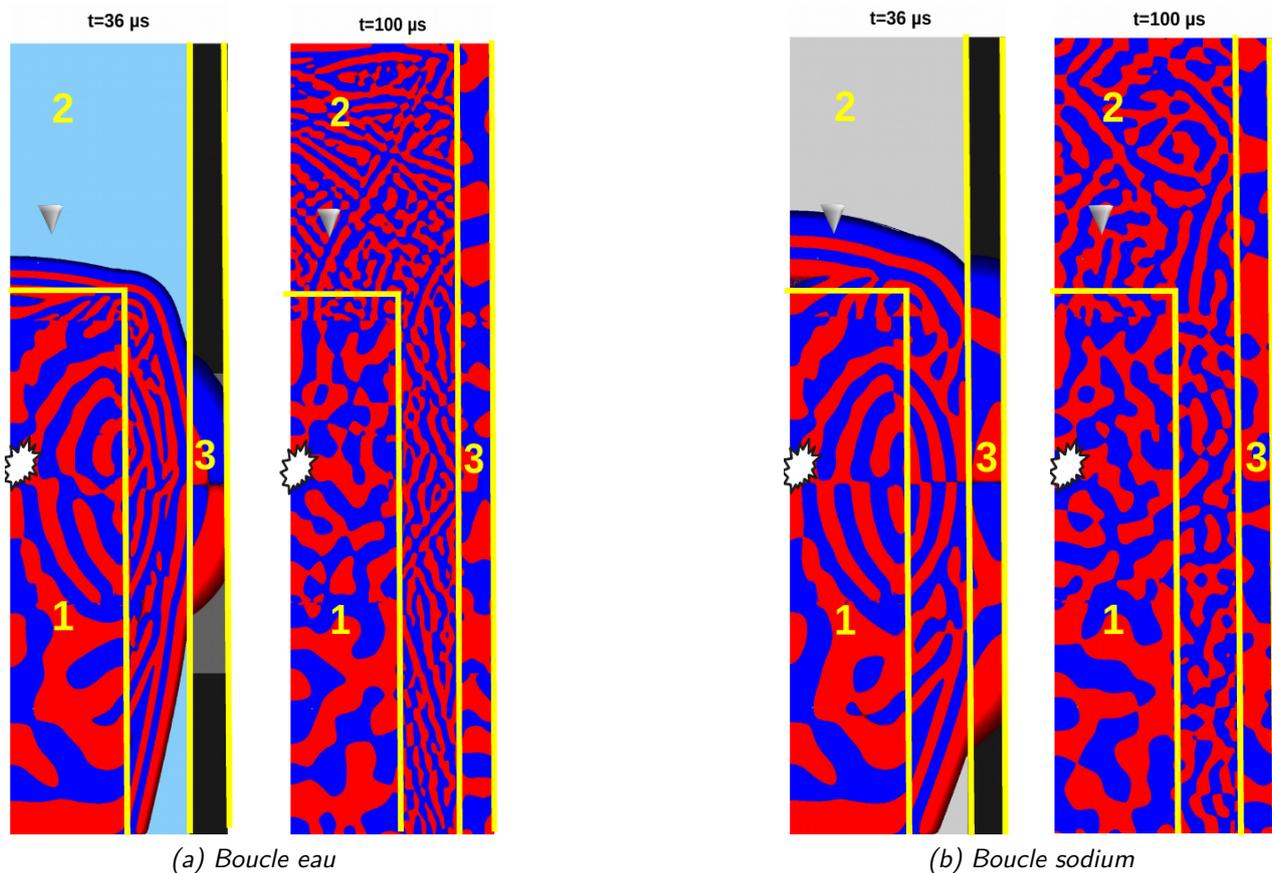


Figure 2.8. – Instantanés de la distribution pression/contrainte normale à différents temps au cœur du dispositif d'essais pour deux types de fluide caloporteur. La zone 1 correspond au support Inox et au couple combustible (Uranium)/gaine (Zircaloy), tandis que la zone 2 (respectivement, la zone 3) représente le fluide caloporteur (respectivement, la couche externe de Zircaloy). La position de la source (respectivement, du capteur) est représentée par une étoile blanche (respectivement, un triangle gris).

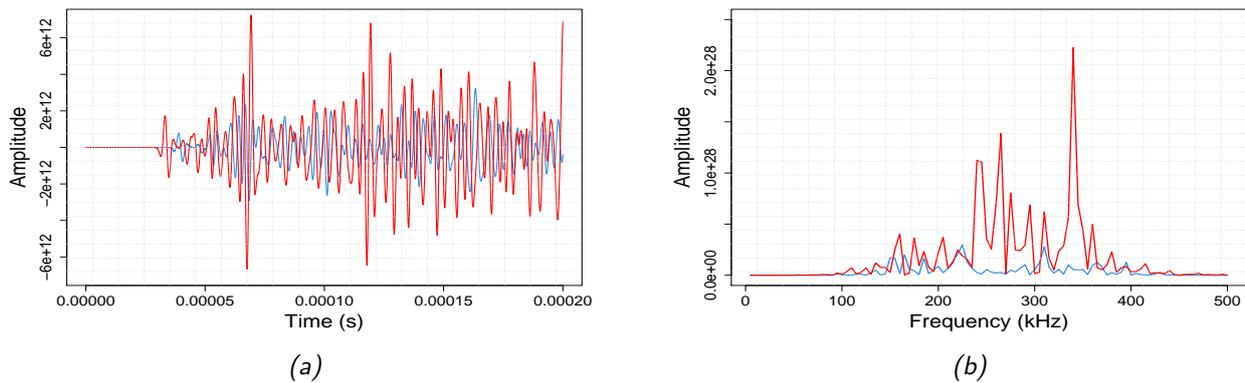


Figure 2.9. – *Signaux simulés (a) et spectres associés (b) obtenus dans le cas où la source est localisée dans le combustible (cf. Figure 2.8) et le récepteur dans le fluide caloporteur eau (—) ou sodium (—).*

2.5. Conclusion

L'objectif principal de ce chapitre était d'évaluer l'impact h_1 du chemin de propagation des ondes au sein du dispositif d'essais sur le signal reçu x , afin d'orienter le choix des outils de traitement du signal dans l'optique d'une estimation du signal source s . Pour ce faire, nous avons choisi d'expérimenter de manière complémentaire une approche analytique et une approche numérique.

La combinaison de ces deux approches a fortement contribué à la compréhension de la physique des ondes associée au contexte de notre étude. D'un point de vue analytique, malgré des simplifications significatives du dispositif d'essais néanmoins validées par la suite par le résultat des simulations numériques, une fréquence de résonance de l'ordre de 180kHz observée sur les signaux physiques reçus lors des essais RIA a pu être retrouvée. Sur le plan numérique, nous avons mis en évidence une forte dépendance de la fonction de transfert du dispositif d'essais par rapport aux positions relatives de la source et du capteur. Ainsi, dès que la position d'un mécanisme physique est incertaine, un traitement de l'impact de h_1 par déconvolution est à exclure. Cependant, il est tout à fait envisageable de filtrer ou pénaliser les composantes du signal reçu qui sont associées à la gamme de fréquences autour de 180kHz , correspondant à une fréquence de résonance du dispositif. Par ailleurs, les résultats de simulations numériques suggèrent que jusqu'à 50kHz , la géométrie et la configuration du dispositif ne biaisent pas les caractéristiques du signal source. Enfin, un changement de fluide caloporteur, d'une boucle eau sous pression vers une boucle sodium, induit une augmentation de l'amplitude du signal reçu et un léger décalage du contenu fréquentiel vers les hautes fréquences.

3 ■ Estimation du signal source et détection de salves

Le travail présenté dans ce chapitre a fait l'objet de deux publications [123, 124] et deux communications [125, 126]

Sommaire

3.1. Introduction	35
3.2. Types de bruit associés à l'environnement d'un RIA	35
3.3. Principes de quelques méthodes de détection de salves et de traitement du bruit	37
3.3.1. Filtrage par sélection de fréquences	38
3.3.2. Transformée de Fourier à court terme et Soustraction spectrale	38
3.3.2.1. Transformée de Fourier à court terme	38
3.3.2.2. Soustraction spectrale	39
3.3.3. Analyse spectrale singulière	40
3.3.3.1. Principe de la SSA	40
3.3.3.2. Détection de salves par SSA	43
3.3.3.3. Débruitage par SSA	44
3.3.4. Analyse par ondelettes	46
3.3.4.1. Détection de salves par ondelettes	46
3.3.4.2. Débruitage par ondelettes	46
3.4. Étude comparative des performances des méthodes	47
3.4.1. Détection de salves	48
3.4.2. Débruitage	50
3.4.2.1. Potentiel d'amélioration du rapport signal-à-bruit	50
3.4.2.2. Impact du débruitage sur les paramètres EA	52
3.4.2.3. Amélioration du débruitage par utilisation conjointe des méthodes	53
3.5. Conclusion	55

3.1. Introduction

Les travaux décrits au chapitre précédent ont permis de mettre en exergue l'impact significatif du dispositif d'essais sur la propagation des ondes, au travers notamment de l'identification de la fréquence de résonance observée sur les signaux physiques. D'après la relation (0.1), et après avoir appliqué les outils appropriés pour supprimer les composantes de x associées aux fonctions de transfert h_1 et h_2 , la quête de l'estimateur \hat{s} de s se résumerait alors au traitement du bruit environnemental n . C'est l'un des principaux objectifs de ce chapitre. Par ailleurs, que s ait été estimé ou pas, la détection des salves est une étape indispensable dans la mise en œuvre du contrôle par EA. Cette question occupera également une place centrale dans ce chapitre.

Par souci de clarté, dans ce qui suit (sauf mention contraire), nous négligerons h_1 et h_2 et considérerons ainsi la problématique résumée par la relation :

$$x(t) = s(t) + n(t) \quad t \in \{1, \dots, N\} \quad (3.1)$$

en lieu et place de l'expression (0.1). Nous adopterons une démarche permettant non seulement de répondre aux problématiques de traitement de n et de détection des salves, mais également d'anticiper le reste du processus, c'est-à-dire l'interprétation des salves détectées. Outre l'objectif classique d'amélioration du SNR, il s'agira ainsi de s'assurer que la méthode de débruitage choisie ne dénature pas les paramètres permettant de caractériser et comparer les salves. Quelques paramètres EA classiques comme le temps de montée, l'énergie absolue, le nombre de coups ou encore la fréquence moyenne, seront considérés, ainsi que les paramètres fonctionnels comme l'enveloppe et le spectre des salves.

En contrôle par EA en général, et dans le contexte des essais RIA en particulier, il est fréquent de disposer d'échantillons de bruit de bonne qualité, ce qui permet une étude du comportement stochastique du bruit et de son évolution dans le temps. Les résultats de cette caractérisation sont cruciaux pour le choix des outils permettant de répondre aux objectifs de ce chapitre. Dans un premier temps, nous procéderons donc à une analyse du bruit environnemental.

Ensuite, nous présenterons diverses méthodes de détection de salves et de traitement du bruit environnemental. Certaines méthodes, déjà éprouvées et très largement utilisées dans la littérature, ne seront présentées que très succinctement. D'autres, moins évoquées mais disposant d'un fort potentiel, seront présentées de manière plus détaillée. Dans tous les cas, on s'attachera à rappeler, s'il y en a, les hypothèses de base sous-jacentes à l'application de chaque méthode, avant d'insister sur des considérations pratiques relatives à sa mise en œuvre.

Enfin, nous procéderons à une étude comparative des performances des méthodes de débruitage présentées. Leur potentiel en terme d'amélioration du SNR est d'abord étudié, puis nous analyserons leur impact sur des paramètres EA choisis, avant d'expérimenter diverses combinaisons afin de pallier aux limites de chaque méthode.

3.2. Types de bruit associés à l'environnement d'un RIA

Après analyse des bruits de fonctionnement survenus lors des essais réalisés depuis 1993 et analyse du contexte global des RIA, nous considérerons quatre types de bruit réellement observés ou potentiellement observables. Les deux premiers sont le bruit blanc gaussien (Figure 3.1a) et le bruit électrique (Figure 3.1b) qui sont des types de bruit plutôt classiques dans le contexte nucléaire. Le troisième,

que nous appellerons bruit mixte (Figure 3.1c), comprend une composante très basse fréquence caractérisée par une ondulation et une composante haute fréquence associée à la résonance du dispositif d'essai. Ce choix, que le lecteur pourrait à juste titre considérer comme la réintroduction du problème que nous avons résolu au chapitre précédent, permet en fait d'explorer la possibilité d'un traitement de h_1 , sans avoir recours aux outils utilisés au chapitre précédent, ce qui allégerait la démarche d'estimation de s . Enfin, nous considérerons un quatrième type de bruit correspondant à la somme des trois premiers (Figure 3.1d).

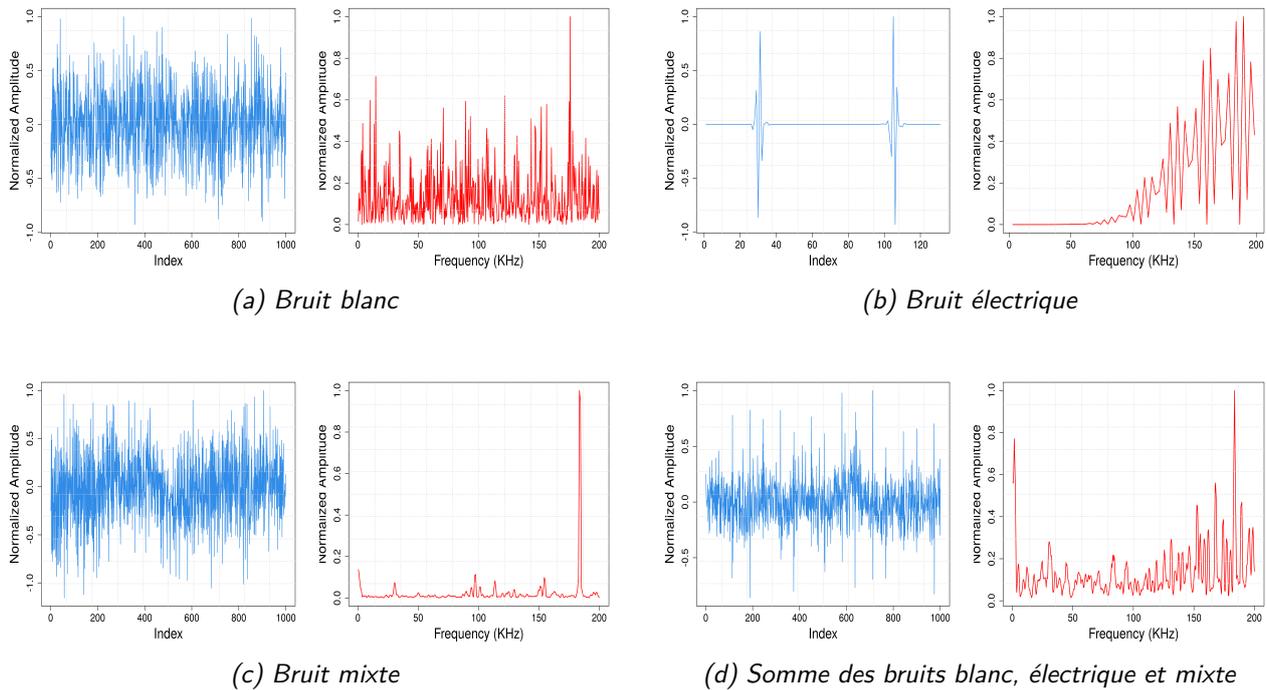
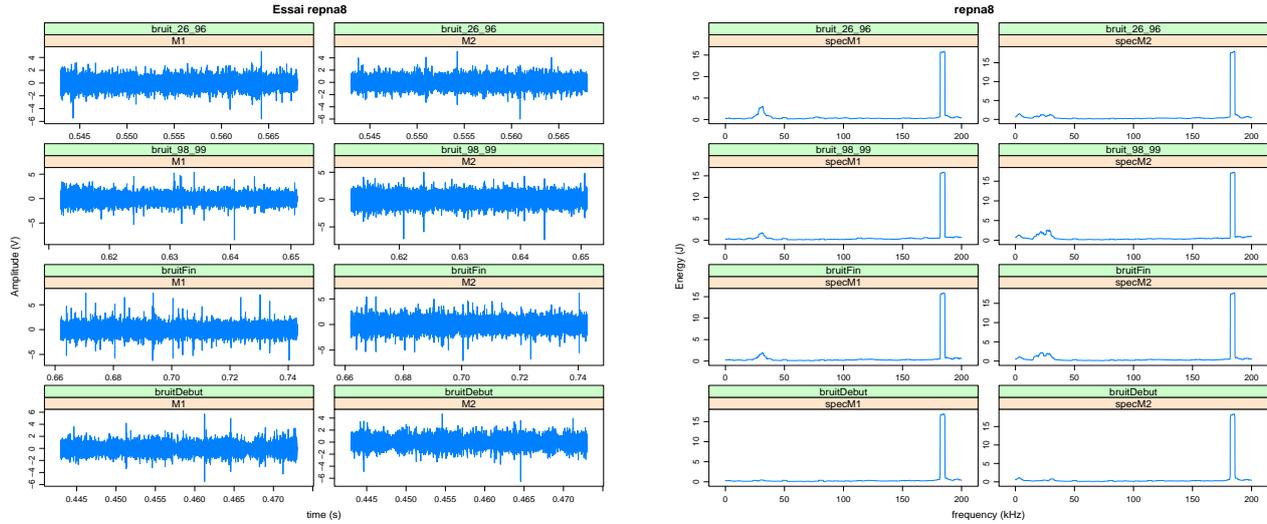


Figure 3.1. – Types de bruit observés ou potentiellement observables dans un environnement RIA. Signaux temporels (—) et spectres associés (—)

Une fois la nature physique du bruit identifiée, une seconde étape tout aussi cruciale pour son traitement efficace est l'analyse de son comportement stochastique. Il s'agit notamment de juger de sa (non-)stationnarité pendant l'essai RIA afin d'orienter le choix de la méthode de débruitage. La durée très courte d'un RIA et l'analyse de divers segments de bruit prélevés à différents moments lors d'un essai réalisé en 1997 laissent supposer que le bruit est stationnaire (Figure 3.2). Cette observation qualitative est confortée par les résultats du test de Dickey-Fuller [127] présentés dans le Tableau 3.1. En effet, les p – *value* étant largement inférieures au seuil critique choisi de 5%, on rejette l'hypothèse nulle de non-stationnarité du bruit. Cet état de fait étant général pour les essais étudiés, nous concluons à la lumière de ce travail préliminaire que, dans le cas des essais RIA, les méthodes de traitement du bruit qu'il convient d'expérimenter en priorité sont celles capables d'appréhender des bruits stationnaires. A noter que leurs caractéristiques physiques sont cependant susceptibles d'évoluer d'un essai à l'autre.



(a) Segments de bruit prélevés à différents moments d'un essai

(b) Spectres associés

Figure 3.2. – Exemple illustratif de l'évolution temporelle et fréquentielle du bruit environnemental pendant un essai RIA.

Tableau 3.1. – Test de non-stationnarité de Dickey-Fuller pour des segments de bruit prélevés à divers moments d'un essai RIA.

Bruits	Capteur	Stat de Fisher	P-value
Début essai	M1	-22.96	0.01
	M2	-19.50	0.01
Fin essai	M1	-29.61	0.01
	M2	-29.27	0.01
99-100	M1	-12.80	0.01
	M2	-10.98	0.01
98-99	M1	-24.49	0.01
	M2	-19.29	0.01

3.3. Principes de quelques méthodes de détection de salves et de traitement du bruit

Dans ce qui suit, lorsqu'il sera nécessaire d'illustrer la mise en œuvre des méthodes présentées, nous utiliserons le signal source simulé composé de deux salves et présenté en Figure 3.3. La première salve, harmonique de 5 kHz que nous noterons s^1 , peut être assimilée à un mécanisme source de type fuite. Quant à la seconde, obtenue à partir de l'expression

$$s^2(t) = \frac{a}{1 + e^{-b(t-\mu_1)}} e^{-c(t-\mu_2)} \sum_{i=1}^n A_i \cos(2\pi f_i t) \quad , \quad (3.2)$$

proposée par WOTZKA [128] et notée s^2 , elle peut être associée à un mécanisme de type rupture.

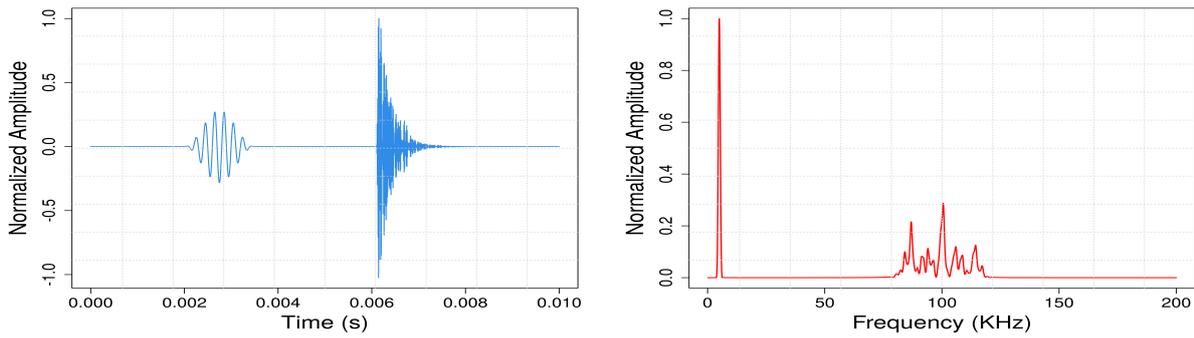


Figure 3.3. – Signal source simulé (—) et spectre associé (—)

3.3.1. Filtrage par sélection de fréquences

En raison de leur simplicité de mise en œuvre et de leur grande efficacité dans le cas du traitement de bruits à bande étroite, les méthodes de filtrage par sélection de fréquences font partie des plus utilisées dans la littérature. Selon la bande de fréquences qu'ils permettent de supprimer ou garder, on distingue les filtres passe-bas, passe-haut, passe-bande et stop-bande [129, chap. 5]. Théoriquement, le filtre idéal est celui ayant un gain constant sur sa bande passante et un gain nul dans la bande coupée, avec une phase linéaire. Cependant, ce genre de filtre est physiquement impossible à réaliser. En pratique, on utilise plutôt un filtre causal caractérisé par sa fonction de transfert exprimée comme suit :

$$H(z) = \frac{\sum_{k=0}^M b_k z^{-k}}{1 + \sum_{k=1}^N a_k z^{-k}} \quad , \quad z \in \mathbb{C} \quad (3.3)$$

Lorsque $N = 0$, le filtre est à Réponse Impulsionnelle Finie (RIF), sinon il est à Réponse Impulsionnelle Infinie (RII).

La mise en œuvre de ce type de filtre nécessite deux étapes distinctes : l'approximation et la réalisation.

L'approximation consiste à trouver les paramètres a_k et b_k permettant de minimiser la distance entre le filtre idéal et $H(z)$. Cela implique le choix d'une norme appropriée (généralement, la norme de Tchebychev ou la norme \mathcal{L}^2) et de la fonction de transfert $H(z)$. Dans le cas des RIF, plusieurs types de fonctions de transfert sont envisageables, les plus utilisées étant les fonctions rectangulaire, de hanning, de hamming et de blackman. En ce qui concerne les RII, les fonctions les plus courantes sont celles de Butterworth et Tchebychev. En pratique, pendant la phase d'approximation, outre le choix du type de filtre et de la fonction de transfert, il faut également spécifier l'ordre N du filtre (cf. Eq.(3.3)). L'étape de réalisation consiste à utiliser la fonction de transfert $H(z)$ afin d'obtenir une estimation du signal source $s(t)$ à partir du signal pollué $x(t)$. Le choix d'une méthode d'implémentation repose sur la prise en compte de divers facteurs, comme la complexité d'implémentation et la sensibilité (cf. [130], chap.5 and 8).

3.3.2. Transformée de Fourier à court terme et Soustraction spectrale

3.3.2.1. Transformée de Fourier à court terme

Une stratégie de détection de mécanismes sources basée sur le seuillage n'utilise que l'information relative au contenu temporel du signal. Afin d'améliorer la détection des salves, l'utilisation de méthodes temps-fréquence est une alternative naturelle. Parmi ces méthodes, la transformée de Fourier

à court terme ou à fenêtre glissante, ou la transformée de Gabor, ou encore le spectrogramme occupe une place de choix. Il s'agit d'utiliser une fonction fenêtre de taille (ou durée) constante pour étudier le contenu fréquentiel du signal segment par segment :

$$STFT(n, m) = \sum_{k=0}^{N-1} x(k)w(n-k)e^{-j2\pi km/N}, \quad (3.4)$$

où w est la fonction de fenêtrage de taille fixe β (en général, la fenêtre de Hanning). Il est alors facile de visualiser les instants d'apparition des différentes composantes fréquentielles (Figure 3.4).

D'un point de vue pratique, le paramètre le plus important, lors de la mise en œuvre d'une transformée de Fourier à court terme, est β . Son choix détermine la résolution temps-fréquence de la transformée. Une diminution de β (respectivement, une augmentation) entraîne une diminution de la résolution en fréquence (respectivement, en temps) [131].

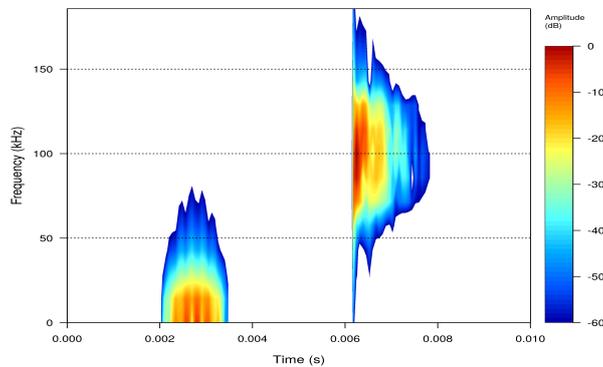


Figure 3.4. – Transformée de Fourier à court terme du signal source simulé présenté en Figure 3.3

3.3.2.2. Soustraction spectrale

La Soustraction Spectrale (SS), dont la paternité est attribuée à BOLL [132], est une méthode de traitement du bruit réputée très efficace dans le cas de bruit stationnaire ou faiblement évolutif, comme celui des RIA. Depuis la fin des années 70, elle a suscité un grand intérêt et fait l'objet de nombreuses applications en traitement du signal numérique, notamment pour la restauration de la parole [133, 134]. De manière intuitive, la SS peut être considérée comme une méthode qui, à travers un ou des estimateurs du bruit, permet de restaurer un signal source en s'appuyant sur une comparaison de densités spectrales en module ou en énergie. Ici, nous présentons brièvement le principe de la méthode et les principales difficultés liées à sa mise en œuvre.

Si on suppose que x est de longueur finie, donc d'énergie finie, l'équation (3.1) est équivalente à :

$$X(f) = S(f) + N(f) \quad f \in \left[-\frac{f_e}{2}, \frac{f_e}{2}\right] \quad (3.5)$$

où X , S et N sont respectivement les transformées de Fourier de x , s et n , et f_e est la fréquence d'échantillonnage. Soit \hat{N} un estimateur du spectre du bruit. Pour une fréquence donnée, la SS consiste à estimer la transformée de Fourier du signal source par :

$$\hat{S}(f) = \left[|X(f)| - \alpha|\hat{N}(f)|\right] e^{i\theta_x(f)} \quad (3.6)$$

où $\theta_x(f)$ est la phase du signal reçu pour la fréquence f .

La principale difficulté liée à l'application des méthodes de type SS est la gestion des distorsions de forme d'onde induites par les erreurs d'estimation dues aux variations aléatoires du bruit. La source de distorsion la plus gênante est celle liée aux estimations négatives ou trop petites de l'amplitude du signal source pour certaines gammes de fréquences [135]. Pour faire face à ce problème, plusieurs approches sont proposées par Boll. Celle que nous avons choisie de mettre en œuvre est la méthode dite de «Half-Wave Rectification» que nous traduirons par méthode de troncature. L'idée de cette technique est de remplacer $\hat{S}(f)$ par 0 à chaque fois que $\alpha|\hat{N}(f)| > |X(f)|$, où le paramètre α , appelé paramètre d'«over-subtraction», permet d'amplifier l'énergie associée au bruit. L'équation (3.6) s'écrit alors :

$$\hat{S}(f) = E(f)e^{i\theta_x(f)} \quad (3.7)$$

avec :

$$E(f) = \begin{cases} |X(f)| - \alpha|\hat{N}(f)| & \text{si } |X(f)| > \alpha|\hat{N}(f)| \\ 0 & \text{sinon} \end{cases} \quad (3.8)$$

La méthode de troncature permet d'atténuer les distorsions de forme d'onde liées aux estimations négatives de l'amplitude du signal. Elle entraîne cependant la perte de toute information relative au signal source pour les fréquences où $E(f) = 0$.

3.3.3. Analyse spectrale singulière

L'analyse spectrale singulière, ou «Singular Spectrum Analysis» (SSA), fait l'objet d'un intérêt de plus en plus croissant depuis le début des années 1990 [136, 137]. Elle a récemment été appliquée avec succès dans plusieurs domaines comme la géophysique [138, 139] et l'économie [140, 141]. Contrairement aux autres méthodes dont il est question dans ce chapitre, la SSA ne nécessite aucune hypothèse forte quant au comportement stochastique du signal source ou du bruit [142]. En utilisant une décomposition du signal en une somme de composantes indépendantes, elle permet d'effectuer diverses tâches telles que l'extraction de composantes spécifiques à partir d'un signal complexe (bruit, tendance, saisonnalité ...), la détection de changements de structure et l'imputation de valeurs manquantes. Hormis quelques rares travaux où la méthode est présentée comme potentiellement efficace pour le traitement de signaux EA [33], la SSA est très peu évoquée dans la littérature. Nous la présenterons donc de manière plutôt détaillée en proposant quelques outils permettant de répondre à des exigences spécifiques aux types de signaux dont il est question dans ce travail.

3.3.3.1. Principe de la SSA

Comme mentionné ci-dessus, la SSA consiste, en deux phases de deux étapes chacune, à décomposer le signal d'intérêt en plusieurs composantes indépendantes.

Phase 1 : Décomposition

La première étape de la phase de décomposition consiste, pour une taille de fenêtre choisie L , à transformer le signal unidimensionnel x en un signal multidimensionnel (X_1, \dots, X_K) , où $X_i = (x_i, \dots, x_{i+L-1})^T \in \mathbb{R}^L$, avec $K = N - L + 1$. Les vecteurs X_1, \dots, X_K forment les colonnes de la **matrice trajectoire** \mathbf{X} de x telle que :

$$\mathbf{X} = [X_1, \dots, X_K] = (x_{ij})_{i,j=1}^{L,K} \quad .$$

Notons que pour une mise en œuvre efficace de la SSA, il est nécessaire de porter une attention particulière au choix de L ($2 \leq L \leq N$). Ce choix doit tenir compte des caractéristiques du signal analysé et des composantes que l'on souhaite isoler. De manière générale, L doit être assez large pour que les

colonnes de \mathbf{X} soient informatives quant au comportement stochastique du signal, tout en n'étant pas supérieure à $N/2$ [143].

Pendant la seconde étape de cette phase, une **décomposition en valeur singulière** (SVD) de \mathbf{X} est réalisée. Notons par $\lambda_1, \dots, \lambda_L$ les valeurs propres de la matrice $\mathbf{X}\mathbf{X}^T$ telles que ($\lambda_1 \geq \lambda_2 \geq \dots \lambda_L \geq 0$), et par U_1, \dots, U_L le système orthonormal correspondant aux vecteurs propres associés. Soit $d = \max\{i \text{ tel que } \lambda_i > 0\}$ et posons $V_i = \mathbf{X}^T U_i / \sqrt{\lambda_i}$. La SVD de \mathbf{X} est définie par une somme bi-orthogonale de matrices élémentaires $\mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^T$ de rang 1 telle que :

$$\mathbf{X} = \mathbf{X}_1 + \dots + \mathbf{X}_d \quad .$$

A l'issue de cette phase, l'analyse des résultats de la SVD, notamment à travers les vecteurs propres (Figure 3.5), permet d'avoir une idée de la complexité du signal analysé, du nombre de composantes indépendantes le constituant, ainsi que de leurs poids.

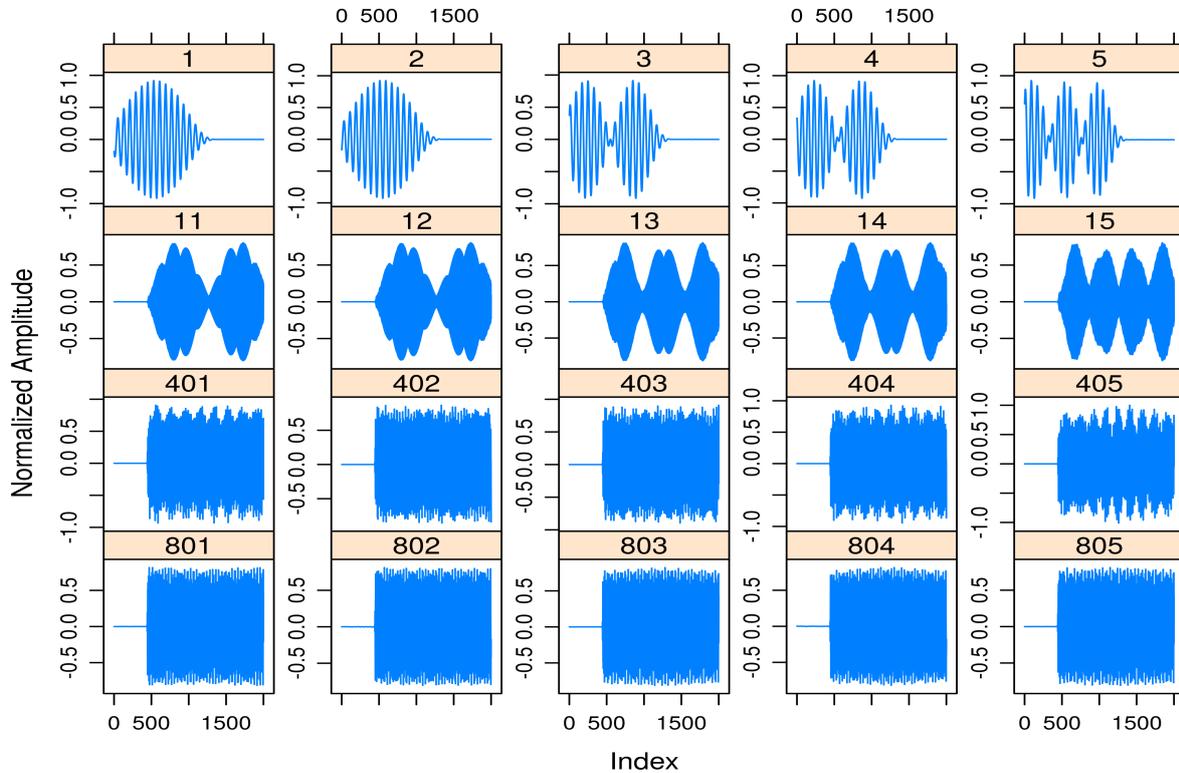


Figure 3.5. – Quelques vecteurs propres U_i choisis à l'issue de la SVD de la matrice trajectoire du signal source simulé présenté en Figure 3.3. ($L = 2000$)

Phase 2 : Reconstruction

La phase de reconstruction consiste, dans une première étape appelée **regroupement des triplets**, à procéder à un regroupement des matrices \mathbf{X}_i en m ($1 \leq m \leq d$) groupes tel que :

$$\mathbf{X} = \mathbf{X}_{I_1} + \dots + \mathbf{X}_{I_m} \quad ,$$

où chaque groupe $\mathbf{X}_I = \mathbf{X}_{i_1} + \dots + \mathbf{X}_{i_p}$ résulte de la somme des matrices élémentaires du groupe I tel que $I = \{i_1, \dots, i_p\}$.

A ce stade de la mise en œuvre de la SSA, l'enjeu principal est de s'assurer, autant que possible, que les matrices \mathbf{X}_I sont associées à des composantes du signal analysé qui sont différentes et d'intérêt. Pour ce faire, en plus de l'analyse des vecteurs propres ou de leurs spectres, d'autres outils comme la matrice de corrélation des vecteurs propres peuvent être envisagés. Par exemple, dans le cas du signal d'illustration présenté en Figure 3.3, la matrice de corrélation des vecteurs propres permet de conclure que les deux composantes ne sont pas corrélées (Figure 3.6). De plus, on peut observer que $m = 2$ et $I_1 = \{i_1, \dots, i_9\}$.

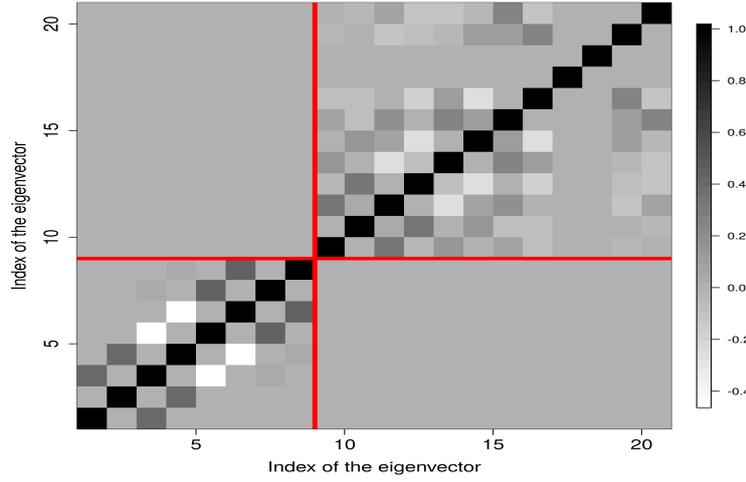


Figure 3.6. – Matrice de corrélation des 20 premiers vecteurs propres obtenus à la suite de la SVD de la matrice trajectoire du signal source simulé présenté en Figure 3.3. ($L = 2000$)

La dernière étape de la SSA permet de transformer chaque matrice obtenue à la suite de la phase de regroupement des triplets propres en un nouveau signal de même taille que le signal reçu x . Cette étape appelée **Hankelization** consiste à approcher chaque matrice \mathbf{X}_{I_k} par la matrice de Hankel la plus proche $\tilde{\mathbf{X}}_{I_k}$:

$$\tilde{\mathbf{X}}_{I_k} = \mathcal{H}\mathbf{X}_{I_k}$$

où \mathcal{H} est appelé opérateur de Hankel. On obtient ainsi, pour chaque $k \in \{1, \dots, m\}$, un signal $\tilde{x}_t^{(k)}$ correspondant à la matrice $\tilde{\mathbf{X}}_{I_k}$ tel que :

$$x_t = \sum_{k=1}^m \tilde{x}_t^{(k)} \quad .$$

Enfin, il convient de s'assurer que les signaux reconstruits $\tilde{x}_t^{(k)}$ répondent bien aux objectifs de la décomposition, en s'assurant de leur séparabilité. Ce dernier critère peut être évalué en calculant les coefficients de corrélation, ou «weighted correlation» (w-correlation), entre les $\tilde{x}_t^{(k)}$. Pour deux signaux reconstruits $x^{(1)}$ et $x^{(2)}$, la w-correlation est définie par :

$$\rho_{12}^{(w)} = \frac{\langle x^{(1)}, x^{(2)} \rangle_w}{\|x^{(1)}\|_w \|x^{(2)}\|_w} \quad (3.9)$$

Une valeur de $|\rho_{12}^{(w)}|$ tendant vers zéro indique que $x^{(1)}$ et $x^{(2)}$ sont bien séparés.

3.3.3.2. Détection de salves par SSA

La détection de salves par SSA repose sur l'hypothèse que les segments du signal x associés au bruit sont gouvernés par une relation de récurrence linéaire (linear recurrent formulae) [142, Chap.3 et 5]. Toute violation de cette propriété constitue alors un changement de structure et donc l'avènement d'un nouveau mécanisme physique. D'un point de vue pratique, l'outil mathématique utilisé pour détecter les changements de structure est la matrice d'hétérogénéité (H-matrix) dont les coefficients $g(x^{(1)}, x^{(2)})$, appelés indices d'hétérogénéité (H-index), sont définis par :

$$g(x^{(1)}, x^{(2)}) = \frac{\sum_{l=1}^{T-L+1} \text{dist}^2(X_l^{(2)}, \mathcal{L}_r^{(1)})}{\sum_{l=1}^{T-L+1} \|X_l^{(2)}\|^2} \quad (3.10)$$

où, $\mathcal{L}_r^{(1)}$ est l'espace vectoriel engendré par l'ensemble $\{U_{i_1}, \dots, U_{i_r}\}$ de vecteurs propres choisis à la suite d'une SVD de la matrice trajectoire de $x^{(1)}$. $X_1^{(2)}, \dots, X_{T-L+1}^{(2)}$ sont les colonnes de la matrice trajectoire de $x^{(2)}$, et $\text{dist}(X, \mathcal{L})$ est la distance euclidienne entre le vecteur $X \in \mathbb{R}^L$ et l'espace vectoriel $\mathcal{L} \subset \mathbb{R}^L$. Ces coefficients traduisent la dissemblance entre un segment de base fixé $x^{(1)}$ de longueur $B > L$ et un segment test $x^{(2)}$ de longueur $T > L$.

Considérons le segment $x_{i,j}$ de x tel que :

$$x_{i,j} = (x_i, \dots, x_j) \quad 1 \leq i < j \leq N.$$

Le coefficient g_{ij} de la H-matrix $\mathbf{G} = \mathbf{G}_{B,T}$, de $N - B + 1$ lignes et $N - T + 1$ colonnes, est :

$$g_{ij} = g(x_{i,i+B-1}; x_{j,j+T-1})$$

Chaque ligne de la H-matrix permet d'observer à quels moments d'un essai l'information fixée par le segment de base $x_{i,i+B-1}$ réapparaît. Par exemple, dans le cas du signal source représenté en Figure 3.3, la H-matrix met clairement en exergue les composantes basses et hautes fréquences (Figure 3.7). Par ailleurs, pour les positions de segments de base correspondant à la composante basses fréquences (respectivement, hautes fréquences), les H-index g_{ij} associés à l'intersection avec la composante hautes fréquences (respectivement, basses fréquences) sont presque égaux à 1. Cela permet de conclure que si l'on se base sur le critère de l'hétérogénéité, ces deux composantes correspondent à des mécanismes physiques différents.

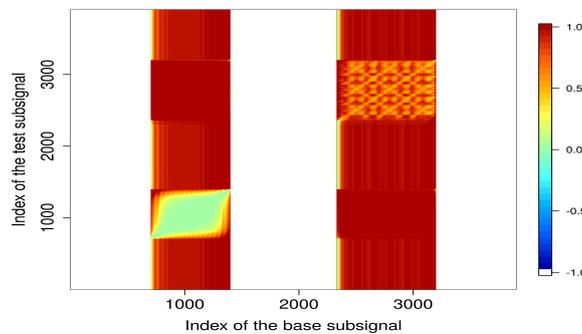


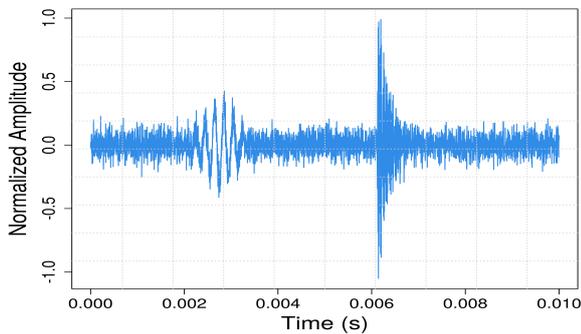
Figure 3.7. – H-matrix du signal source simulé présenté en Figure 3.3 : $B = T = 100$, $L = 50$, $\mathcal{L}_r^{(1)} = \text{vect}(U_1^1, U_1^2)$. Notons que, dans ce graphique, les coefficients indéfinis de la H-matrix ont été remplacés par -1 et sont représentés en blanc.

3.3.3.3. Débruitage par SSA

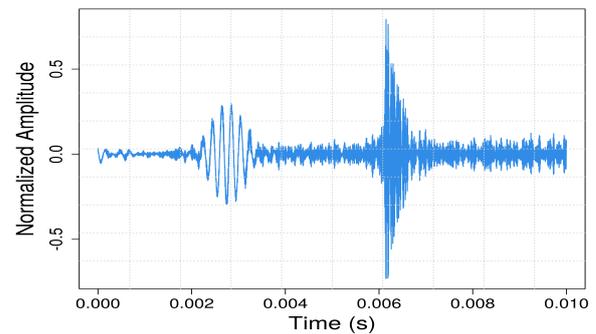
Selon la nature du bruit, différentes stratégies de débruitage basées sur la SSA peuvent être envisagées. Lorsque le bruit est à bande étroite, une SSA directe avec une stratégie de regroupement des vecteurs propres, visant à isoler ceux appartenant à la même bande fréquentielle que le bruit, conduit à d'excellents résultats. Dans ce cas, la SSA peut d'ailleurs être considérée comme équivalente à un simple filtre passe-bande. En ce qui concerne les bruits plus complexes, des stratégies plus élaborées doivent être mises en œuvre. Dans le cas des bruits large-bande par exemple, l'un des plus difficiles à traiter par SSA, une détérioration de la séparabilité entre le signal source s et le bruit n est observée. Il convient alors de définir des stratégies de débruitage plus élaborées. Ici, nous en expérimentons deux : une première basée sur la comparaison des contributions des vecteurs propres, et une seconde combinant la H-matrix et la SSA.

a. Regroupement par comparaison des contributions

Lorsque le SNR est élevé, en notant que les valeurs propres obtenues à la suite de la SVD de \mathbf{X} sont classées par ordre de contribution décroissante, il est naturel de supposer que les premières d'entre elles sont associées au signal source s (Figure 3.8). Dans ce cas, l'estimation de s reposerait sur la détermination du seuil à partir duquel le reste des vecteurs propres est considéré comme associé à n . Pour ce faire, plusieurs outils, parmi lesquels l'ébouilissement des valeurs propres ou la matrice des w -corrélations, sont envisageables. Dans le cas d'un bruit gaussien, cette stratégie permet ainsi d'obtenir une légère amélioration du SNR (Figure 3.8).



(a) Signal pollué par un bruit gaussien



(b) Signal traité obtenu à la suite d'une application de la méthode de comparaison des contributions

Figure 3.8. – Exemple illustratif de l'estimation du signal source par comparaison des contributions de vecteurs propres. Cas d'une pollution du signal présenté en Figure 3.3 par un bruit blanc gaussien ($SNR=10dB$). ($L=2000$)

b. Débruitage par combinaison de la H-matrix et de la SSA

Afin d'obtenir une meilleure qualité de débruitage dans les cas où une SSA directe serait insuffisante pour séparer convenablement s et n , nous proposons une démarche en deux étapes.

La première consiste à créer une fonction globale de détection de changements de structure. Fixons un segment test $x_{j,j+T-1}$, avec $j \in \{1, \dots, N - T + 1\}$. La fonction de détection de changements de structure associée à la colonne j de la H-matrix est définie par :

$$D^j(n - B + 1) = g(x_{n-B+1,n}, x_{j,j+T-1}) \quad B \leq n \leq N$$

où $g(x_{n-B+1,n}, x_{j,j+T-1})$ est la H-index entre le segment de base $x_{n-B+1,n}$ et le segment test $x_{j,j+T-1}$. L'analyse de la H-matrix permet de déterminer les colonnes j correspondant à d'éventuels mécanismes physiques (Figure 3.9a), et ainsi de définir la fonction globale de détection de changements de structure D (Figure 3.9b) telle que :

$$D = \sum_j D^j$$

L'application de D à x supprime tous les segments associés à du bruit pur (Figure 3.9c).

La seconde étape consiste à affiner l'estimateur obtenu par application d'une SSA directe, conduisant à une amélioration substantielle de la qualité du débruitage (Figure 3.10a). Cependant, l'analyse du spectrogramme du signal source estimé permet de constater que des artefacts persistent (Figure 3.10b). Ceux-ci s'expliquent par un manque d'adaptabilité de la méthode, traduit par l'impossibilité de garder une composante pour un segment de signal donné tout en la supprimant pour d'autres. Ainsi, dans ce cas précis, la SSA atteint ses limites.

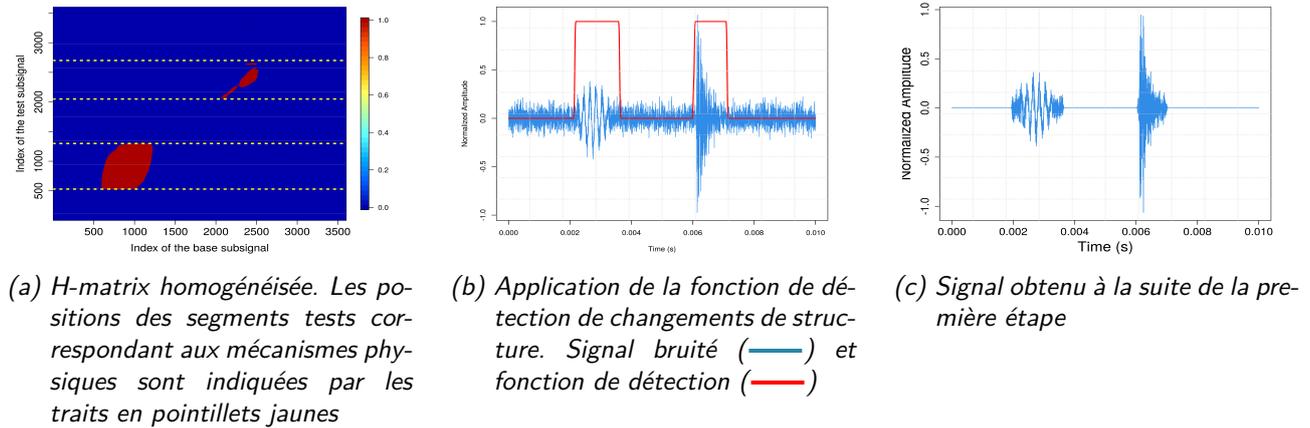


Figure 3.9. – Illustration de la première étape de la stratégie de débruitage en deux étapes.

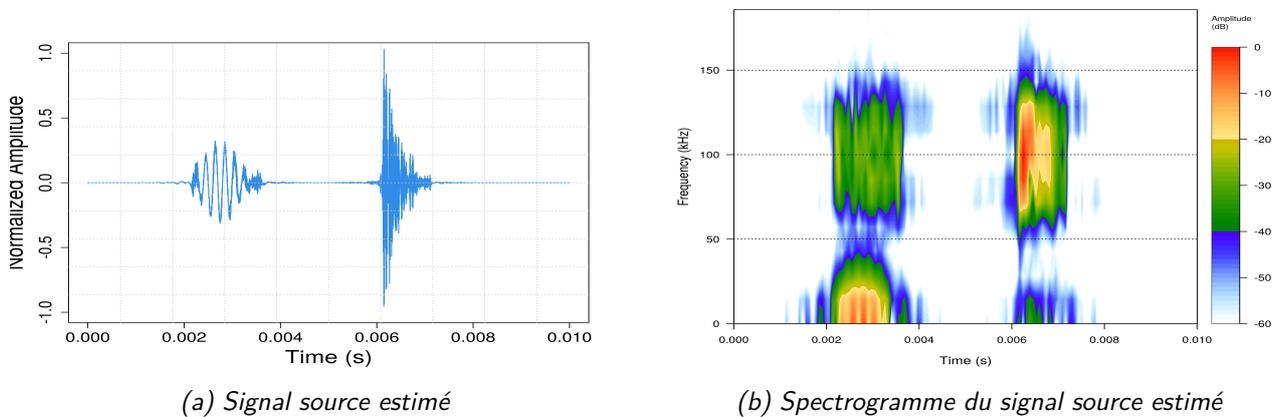


Figure 3.10. – Résultat de l'amélioration de la SSA classique par application d'une stratégie de débruitage en deux étapes, consistant à supprimer les segments de bruit pur avant d'appliquer la SSA.

3.3.4. Analyse par ondelettes

Contrairement à la STFT qui impose le choix d'une fonction de fenêtrage de durée fixe, la transformée en ondelettes permet de procéder à une décomposition adaptative du signal en temps et en fréquence. Au-delà de son utilisation naturelle comme outil d'analyse de la structure de signaux ou d'images, les nombreuses applications de ce puissant outil ont révolutionné le traitement du signal et des images au cours des deux dernières décennies [144]. La méthode étant très présente dans la littérature, nous nous limiterons à une présentation très succincte des outils dont on fera usage par la suite. Le lecteur pourra se reporter aux ouvrages de HUBBARD [145] & MALLAT [146] pour des rappels historiques ou mathématiques.

3.3.4.1. Détection de salves par ondelettes

Une ondelette ψ est une fonction oscillante de moyenne nulle possédant un certain degré de régularité et dont le support est fini. La transformée en ondelettes d'un signal x est une décomposition multi-échelles permettant de représenter et d'analyser x en fonction du temps et de l'échelle. Pour une ondelette mère donnée $\psi(t)$, la transformée en ondelettes continue, ou «continuous wavelet transform» (CWT), du signal x se définit par :

$$CWT(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} \psi\left(\frac{t-b}{a}\right) x(t) dt$$

où a est le paramètre d'échelle ou de dilatation associé à la fréquence et b le coefficient de translation associé à la position de l'ondelette. $CWT(a, b)$ est la corrélation entre le signal $x(t)$ et l'ondelette $\psi_{a,b}(t)$. La transformée en ondelettes discrète, ou «Discret wavelet transform» (DWT), est une version échantillonnée du CWT calculée en remplaçant $a, b \in \mathbb{R}$ par une grille discrète, telle que :

$$a = 2^{-j}, \quad b = k \cdot 2^{-j}, \quad j, k \in \mathbb{Z}$$

En pratique, deux éléments essentiels à prendre en compte lors de la mise œuvre d'une transformée en ondelettes sont le choix de l'ondelette mère et celui du paramètre d'échelle (ou niveau de résolution). L'ondelette mère doit être sélectionnée de manière à caractériser convenablement le signal, en évaluant par exemple son niveau de corrélation avec ce dernier. Quant au choix du niveau de résolution, il dépend de la complexité de x : plus celle-ci est élevée, plus il est nécessaire d'utiliser un nombre croissant d'échelles.

3.3.4.2. Débruitage par ondelettes

La plupart des méthodes de débruitage par ondelettes proposées dans la littérature sont basées sur une approche de seuillage. Celle-ci s'appuie sur la propriété d'orthogonalité des ondelettes afin de décomposer le signal reçu x à l'échelle j et la position k tel que [144] :

$$x_{jk} = s_{jk} + n_{jk} \quad ,$$

où s est le signal source et n un bruit aléatoire. Sachant que le signal source est déterministe, on suppose alors qu'il est résumé par un nombre réduit de coefficients d'ondelettes d'amplitudes relativement fortes. Ainsi, estimer s revient à définir une stratégie de seuillage permettant de garder uniquement ces coefficients [144, 147].

Parmi les méthodes de seuillage les plus utilisées, on peut citer le seuillage dur, ou «hard thresholding» :

$$\delta_{\lambda}(x_{jk}) = \begin{cases} 0, & \text{if } |x_{jk}| \leq \lambda \\ x_{jk}, & \text{otherwise} \end{cases} \quad (3.11)$$

et le seuillage doux, ou «soft thresholding» :

$$\delta_{\lambda}(x_{jk}) = \begin{cases} 0, & \text{if } |x_{jk}| \leq \lambda \\ x_{jk} - \lambda, & \text{if } |x_{jk}| > \lambda \\ x_{jk} + \lambda, & \text{if } |x_{jk}| < -\lambda \end{cases} \quad (3.12)$$

La fonction de seuillage λ dépend en général de l'échelle j de la décomposition en ondelettes ($\lambda = \lambda(j)$) et peut être définie selon plusieurs approches. On peut notamment citer la méthode dite du «SureShrink» proposée par DONOHO & JOHNSTONE [148] et la méthode du «NeighBlock» proposée par CAI & SILVERMAN [149].

En pratique, la mise en œuvre d'un débruitage par ondelettes nécessite un choix judicieux de l'ondelette mère et du niveau de décomposition J du signal. L'ondelette mère doit être sélectionnée selon la même règle que celle exposée ci-dessus. Quant au choix du niveau de décomposition, il est plus délicat. Il dépend de plusieurs facteurs dont le plus important est le comportement du bruit.

3.4. Étude comparative des performances des méthodes

Il s'agit ici de comparer les performances des différentes méthodes présentées précédemment sur des cas tests maîtrisés. Nous avons choisi de commencer par la question de la détection des salves au lieu de celle du traitement du bruit. Ce choix s'explique par le fait que dans la démarche de contrôle par EA, il peut arriver que l'expérimentateur décide de ne pas traiter le bruit ou que la méthode de traitement envisagée soit inefficace. Par ailleurs, traiter le bruit équivaut à une tentative d'amélioration du SNR et donc à se ramener à un contexte où celui-ci est élevé. Pour rester général et envisager tous les cas de figure, il suffit de traiter la problématique de détection des salves dans le cas d'un SNR faible et dans celui d'un SNR élevé.

Considérons les six salves présentées en Figure 3.11 issues de différents essais RIA. La comparaison de leurs forme d'onde et contenu spectral montre qu'elles sont *a priori* associées à des mécanismes physiques différents. En les concaténant, nous créons artificiellement un signal source essai s composé de ces divers types de mécanismes physiques (Figure 3.12). En ajoutant à s les différents types de bruits présentés précédemment, nous reproduisons ainsi des signaux essais x représentatifs de ceux obtenus à la suite d'un RIA. Ce choix, plutôt que celui de considérer les signaux reçus réels, s'explique par le fait qu'en plus de la résolution du problème de débruitage, nous avons également pour objectif de quantifier l'impact des méthodes choisies sur les paramètres EA classiques. Il faut donc partir d'un signal source s connu qu'il s'agira de restaurer le plus fidèlement possible.

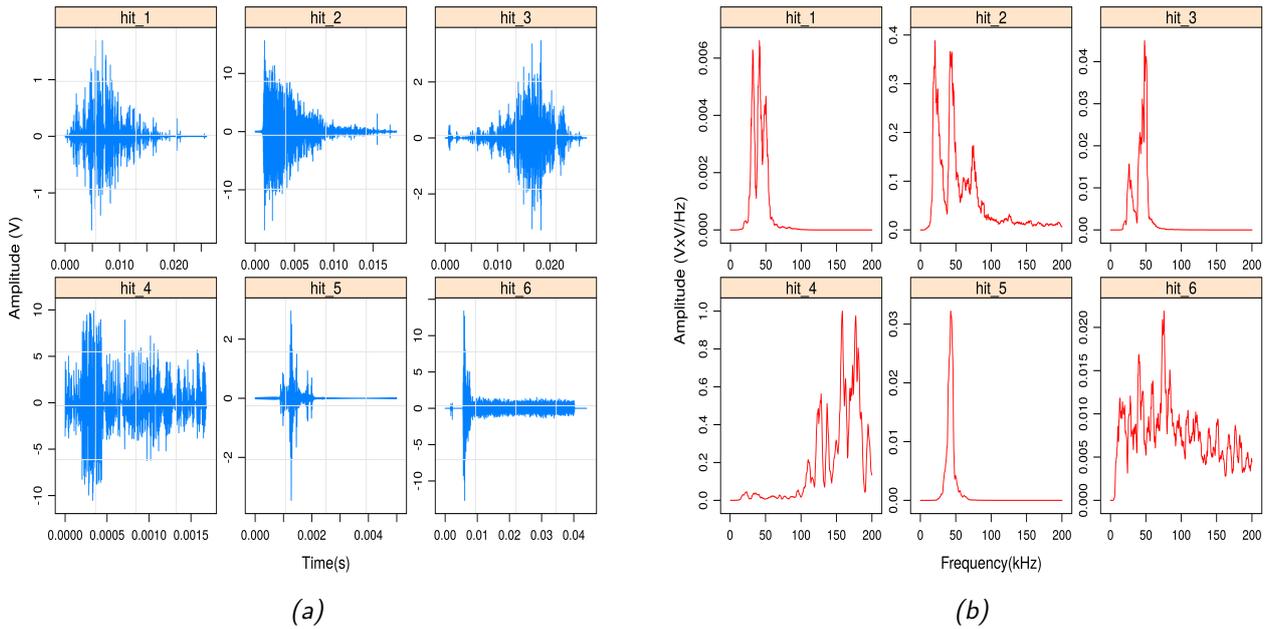


Figure 3.11. – Salves choisies (a) et spectres associés (b)

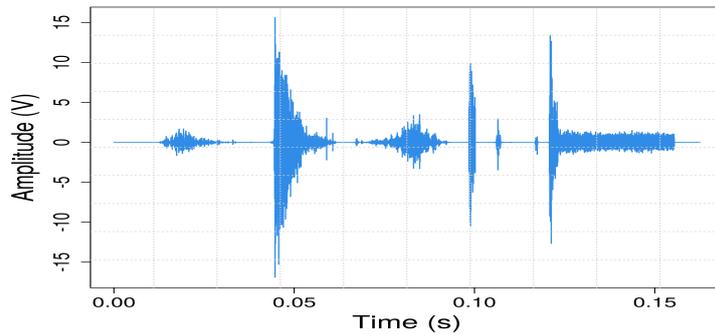


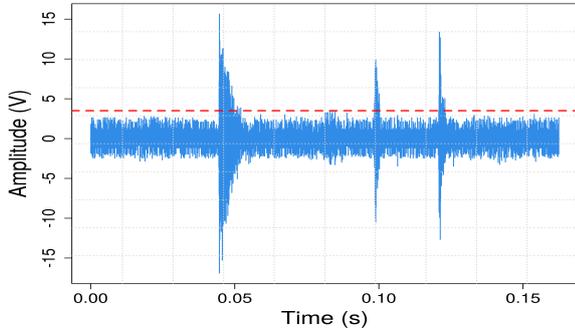
Figure 3.12. – Signal source s résultant de la concaténation des six salves présentées en Figure 3.11a

3.4.1. Détection de salves

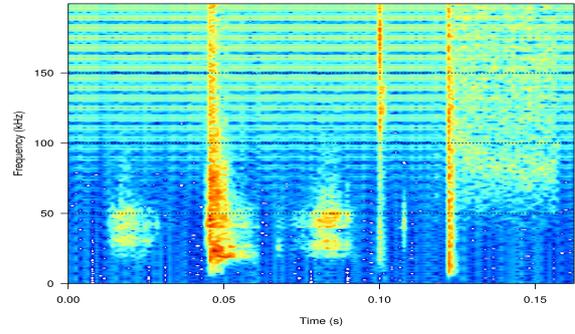
Comme nous le soulignons précédemment, la méthode de seuillage est inadéquate pour la détection des salves faiblement énergétiques. Cela est illustré en Figure 3.13a par la non détectabilité des salves 1, 3 et 5. Il est alors nécessaire d'utiliser des méthodes adaptatives comme celles présentées ci-dessus.

Dans un contexte où le SNR est relativement élevé, quels que soient la méthode choisie et le type de bruit, toutes les salves, y compris celles faiblement énergétiques, sont relativement bien détectées (Figures 3.13b-3.13d). D'un point de vue purement comparatif, si la CWT est évidemment préférable à la STFT pour une analyse fine de la structure du signal, on notera que la SSA serait plutôt un outil complémentaire à l'une ou l'autre de ces deux méthodes. En effet, la construction même de la H -matrix, à travers le choix de l'espace vectoriel \mathcal{L} , permet d'une part, de choisir les composantes du signal que l'on souhaite mettre en exergue, et d'autre part, de quantifier le niveau de corrélation (en

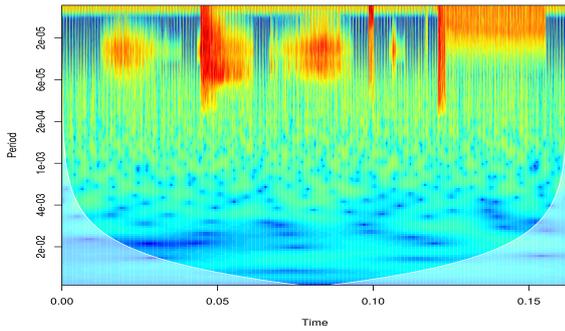
termes d'hétérogénéité) de ces composantes avec les différentes salves [123].



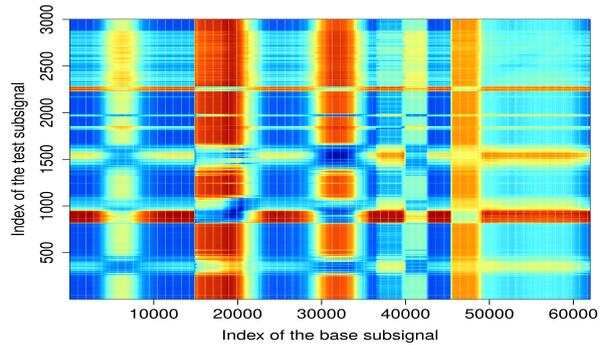
(a) Signal source s (Figure 3.12) pollué par un bruit électrique (Figure 3.1b) (—). Seuil de détection des salves fixé à 3.5 Volts (- - -)



(b) Détection des salves par STFT avec une fenêtre de taille 512



(c) Détection des salves par CWT avec une ondelette de Morlet

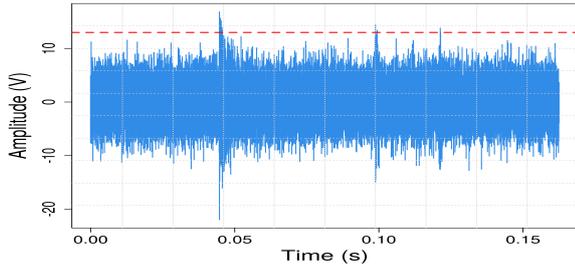


(d) Détection des salves par la H-matrix. $B = 300$, $T = 3000$, $L = 100$ et $\mathcal{L}_r^{(1)} = \text{vect}(U_1^1, \dots, U_1^{30})$

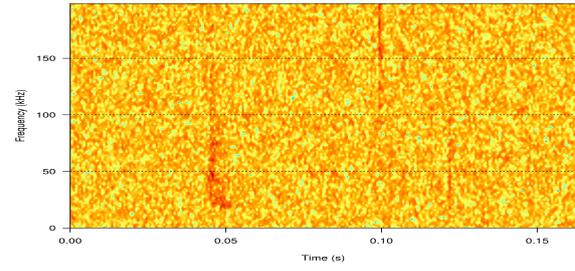
Figure 3.13. – Illustration de la détection des salves au moyen des différentes méthodes, dans le cas d'un signal source pollué par un bruit électrique et dans un contexte de fort SNR (+10dB).

Lorsque le SNR est très faible, les méthodes expérimentées ici comme alternatives au seuillage peuvent également s'avérer inefficaces (Figures 3.14b-3.14d). En effet, la détectabilité des salves devient alors très sensible à la nature du bruit. L'une des situations les plus complexes à gérer correspond à celle où le SNR est faible et le bruit très large bande ; ni une analyse fine de la structure de x par utilisation de la CWT, ni une SSA ne permettent alors d'obtenir des résultats réellement satisfaisants (Figure 3.14). Par ailleurs, lorsque l'énergie du bruit est associée à des bandes de fréquences particulières, comme dans le cas d'un bruit électrique ou du bruit mixte, la meilleure méthode d'analyse de structure reste la CWT.

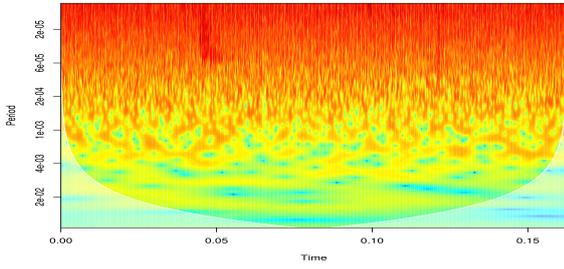
D'un point de vue pratique, notons que l'automatisation de la détection de salves par STFT et CWT peut se faire en ayant recours au même type de démarche que celui proposé pour la détection des changements de structure par SSA (cf. page 44).



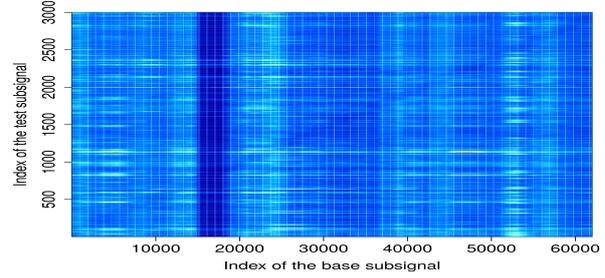
(a) Signal source s (Figure 3.12) pollué par un bruit blanc gaussien (Figure 3.1a) (—). Seuil de détection des salves fixé à 13 Volts (- - -)



(b) Détection des salves par STFT avec une fenêtre de taille 256



(c) Détection des salves par CWT avec une ondelette de Morlet



(d) Détection des salves par la H-matrix. $B = 300$, $T = 3000$, $L = 100$ et $\mathcal{L}_r^{(1)} = \text{vect} \{U_i^1, \max(|FT(U_i^1)|) \in 40 - 60kHz\}$

Figure 3.14. – Illustration de la difficulté de détection des salves au moyen de différentes méthodes, dans le cas d'un signal pollué par un bruit large bande et dans un contexte de faible SNR (-10dB).

3.4.2. Débruitage

Les résultats des travaux sur la détection des salves finissent de prouver, s'il en était besoin, qu'une démarche rigoureuse, notamment en contexte de faible SNR, exige de traiter d'abord le bruit environnemental avant toute autre analyse du signal essai x . Ici, nous comparons les différentes méthodes de débruitage présentées préalablement dans ce chapitre.

3.4.2.1. Potentiel d'amélioration du rapport signal-à-bruit

Les ratios d'amélioration du SNR, exprimés par :

$$ratio = \frac{(SNR_{output} - SNR_{input})}{(SNR_{input})} \quad (3.13)$$

et présentés dans le Tableau 3.2, permettent de conclure que d'un point de vue global la SS et la SSA sont plus robustes aux changements de nature du bruit que les ondelettes, avec un potentiel d'amélioration du SNR plus élevé pour la SS. Ce résultat s'explique facilement en revenant à des considérations plus théoriques. En effet, contrairement aux deux autres méthodes, la SS utilise l'hypothèse physique de disponibilité d'un estimateur \hat{n} du bruit. En seuillage par ondelettes, le bruit est supposé aléatoire alors qu'une bonne séparabilité entre le bruit et le signal source est requise pour la SSA. Ainsi, lorsque $\hat{n} \simeq n$ comme c'est le cas ici, la SS donne des résultats robustes et améliore considérablement le SNR. Cependant, on note qu'à la suite d'une SS il reste toujours un bruit résiduel, et cela indépendamment

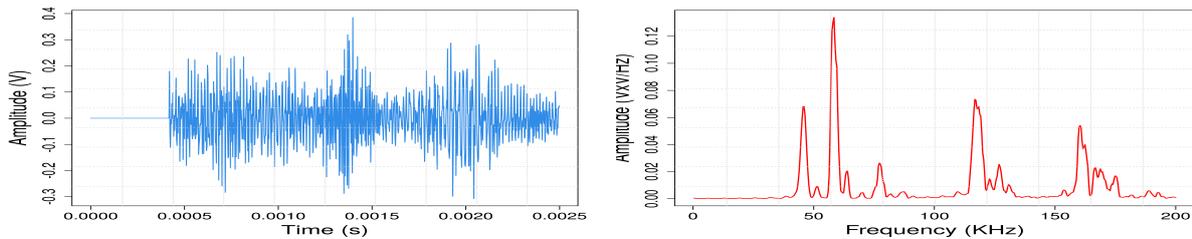
de la nature physique du bruit (Figure 3.15a).

Comme on pouvait s’y attendre, dans le cas d’un bruit blanc, les ondelettes présentent le meilleur potentiel d’amélioration du SNR. En effet, le spectre d’un bruit blanc étant très large bande, une décomposition en ondelettes de celui-ci répartirait quasi-équitablement l’énergie du bruit entre les différents coefficients d’ondelettes, facilitant de ce fait le seuillage. Ainsi, si on prend comme unique critère de choix le potentiel d’amélioration du SNR, le seuillage par ondelettes doit être préféré aux autres approches pour le traitement de bruits gaussiens. Dès que certaines composantes du bruit sont déterministes, comme c’est le cas pour le bruit électrique ou le bruit mixte, le seuillage par ondelettes peut s’avérer totalement inefficace (Figure 3.16) et garder intactes les composantes déterministes du bruit (Figure 3.15b). Dans ces cas, il faut par conséquent avoir recours à des outils de prétraitement du signal reçu avant application d’un seuillage par ondelettes.

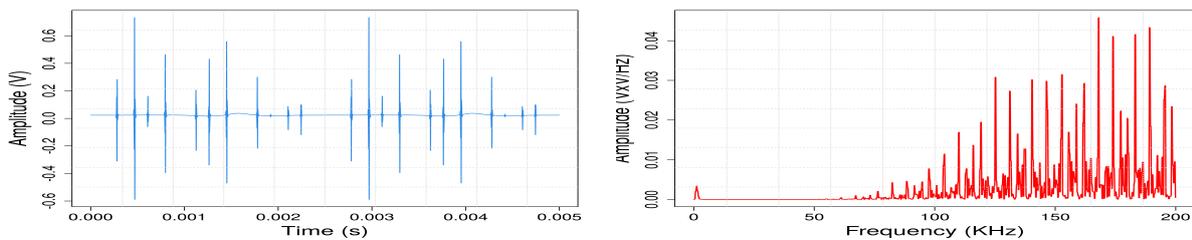
Enfin, la SSA présente un potentiel d’amélioration du SNR moins important que les autres méthodes. Cependant, comme nous le montrerons dans la suite, une fois les éventuels problèmes de séparabilité entre s et n résolus, la SSA permet malgré tout d’obtenir d’excellents résultats.

Tableau 3.2. – Ratio d’amélioration du SNR en fonction de la méthode de débruitage utilisée.

	SS	Ondelettes	SSA
Bruit gaussien	3.52	9.31	3.42
Bruit électrique	5.73	1.62	4
Bruit mixte	6.68	3.42	3.50
Somme des bruits	4.63	4.49	3.36

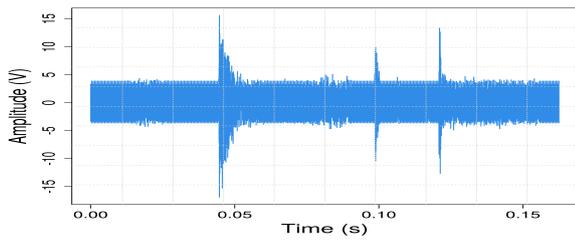


(a) Bruit résiduel (—) et son spectre (—) obtenus à la suite d’un débruitage par SS dans le cas du signal source (Figure 3.12) pollué par un bruit blanc gaussien (cf. Figure 3.1a)

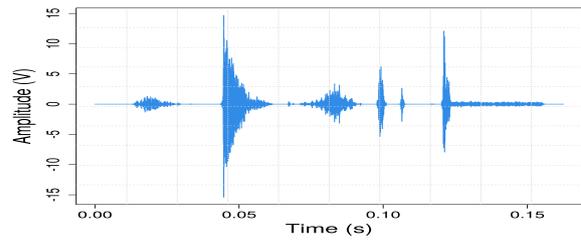


(b) Bruit résiduel (—) et son spectre (—) obtenus à la suite d’un débruitage par ondelettes dans le cas du signal source (Figure 3.12) pollué par la somme des bruits (cf. Figure 3.1d)

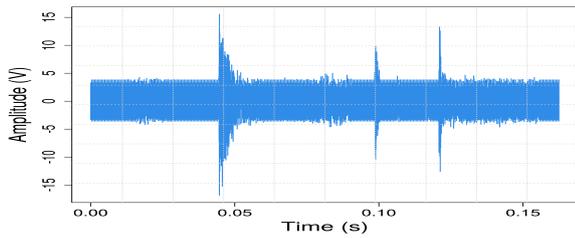
Figure 3.15. – Illustration des bruits résiduels à la suite de débruitage par SS et par ondelettes.



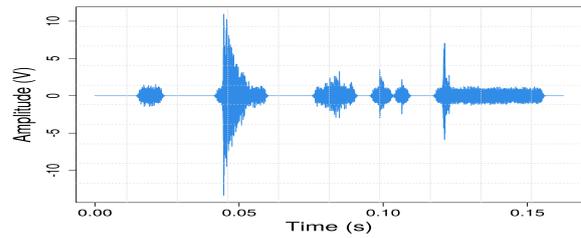
(a) Signal x constitué du signal source (Figure 3.12) et du bruit électrique (Figure 3.1b). SNR=5dB



(b) Résultat du débruitage de x par SS denoising avec une fenêtre d'overlapping de taille 512 et un paramètre d'over-subtraction $\alpha = 4$



(c) Résultat du débruitage par ondelettes en utilisant la méthode du soft thresholding et une fonction de seuillage sureShrink



(d) Résultat du débruitage par une SSA en deux étapes. Première étape : $B = 300$, $T = 2000$, $L = 100$ and $\mathcal{L}_r^{(1)} = \text{span}(U_1^1, \dots, U_1^{10})$. Seconde étape : $L = 500$

Figure 3.16. – Débruitage au moyen des différentes méthodes, dans le cas du signal source (cf. Figure 3.12) pollué par un bruit électrique (cf. Figure 3.1b).

3.4.2.2. Impact du débruitage sur les paramètres EA

Les analyses réalisées jusqu'ici permettent, selon le type de bruit, d'établir une hiérarchie entre les méthodes choisies en se basant sur le critère classique de l'amélioration du SNR. Cependant, dans un processus de contrôle par EA, il est également souhaitable de s'assurer que la méthode de traitement choisie *in fine* ne biaise pas les caractéristiques des salves pour une interprétation ultérieure pertinente. Pour ce faire, nous étudions l'impact des méthodes de débruitage au travers d'une comparaison de quelques paramètres EA associés aux six salves du signal source s (Tableau 3.3) avec ceux obtenus après restauration de s (Annexe C.1).

Tableau 3.3. – Quelques paramètres EA classiques pour les six salves présentées en Figure (3.11a)

	hit_1	hit_2	hit_3	hit_4	hit_5	hit_6
Envelopes correlation	1.00	1.00	1.00	1.00	1.00	1.00
Kullback-Leiber distance	0.00	0.00	0.00	0.00	0.00	0.00
Absolute energy (MV)	12.05	1458.37	59.30	4224.40	31.68	81.15
Amplitude (V)	4.55	23.88	10.82	19.89	9.39	19.43
Count	1135.00	842.00	1186.00	250.00	204.00	3841.00
Mean frequency (kHz)	43.65	46.75	43.90	147.93	40.80	87.29
Rise Time (ms)	6.84	1.20	18.31	0.34	1.27	5.82
Count to pic	306.00	51.00	797.00	58.00	51.00	129.00

A la suite du débruitage, on constate que les valeurs estimées des paramètres de forme d'onde classiques comme l'enveloppe, le temps de montée et l'amplitude sont relativement proches de celles des signaux sources. Par contre, les valeurs estimées des paramètres liés à l'activité acoustique et au

contenu fréquentiel comme le nombre de coups, la fréquence moyenne et le nombre de coups au pic, sont fortement biaisées. Quels que soient le type de bruit et la méthode de débruitage choisie, les paramètres EA associés à la forme d'onde sont donc moins affectés par le débruitage que ceux liés à l'activité acoustique et au contenu fréquentiel. Cependant, ce résultat doit être nuancé par l'analyse des distances de Kullback-Leiber entre les spectres des salves du signal source s et ceux du signal débruité. En effet, pour une méthode donnée, lorsque les caractéristiques du bruit permettent une application optimale, ces distances sont relativement faibles. Il est donc possible de restaurer la forme globale des spectres. Par ailleurs, une comparaison des méthodes basées sur leur potentiel de restauration des paramètres EA choisis ne permet pas d'établir de hiérarchie claire entre les méthodes. Ainsi, le potentiel d'amélioration du SNR reste l'outil le plus efficace pour orienter le choix d'une méthode de traitement du bruit.

3.4.2.3. Amélioration du débruitage par utilisation conjointe des méthodes

Dans certains cas, l'analyse du bruit résiduel permet de mettre clairement en exergue un manque d'efficacité de la méthode de débruitage choisie (cf. Figure 3.15). Il convient alors de changer de méthode ou d'avoir recours à des outils supplémentaires pour faire face aux limites observées. Nous nous proposons ici d'étudier différentes combinaisons des méthodes présentées précédemment, afin d'obtenir à terme une estimation correcte du signal source quelle que soit la méthode choisie initialement,

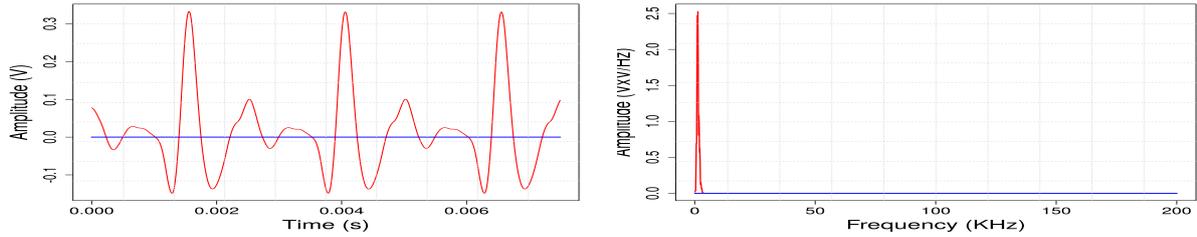
Considérons le cas où la méthode initiale est le seuillage par ondelettes. Lorsque le bruit résiduel est à bande étroite, plusieurs des outils présentés précédemment peuvent être utilisés pour améliorer le débruitage. L'approche la plus simple est d'avoir recours à une sélection de fréquences afin de supprimer la composante déterministe du bruit avant application du seuillage par ondelettes (Figure 3.17a). Dans la même optique, on peut également utiliser une transformée en ondelettes ou une décomposition SS pour isoler cette composante. Dans le cas d'un bruit large bande, comme le bruit électrique, l'utilisation d'une SS en amont du seuillage par ondelettes conduit à d'excellents résultats.

Si la méthode de base choisie est la SS et si $\hat{n} \simeq n$, une simple augmentation du paramètre d'over-subtraction α est suffisante pour supprimer le bruit résiduel. Malheureusement, un tel choix peut conduire à une forte pénalisation de l'énergie du signal source et peut parfois fortement biaiser les paramètres EA. Dans ce cas, une approche plus judicieuse consiste à avoir recours à un seuillage par ondelettes après la SS (Figure 3.17b).

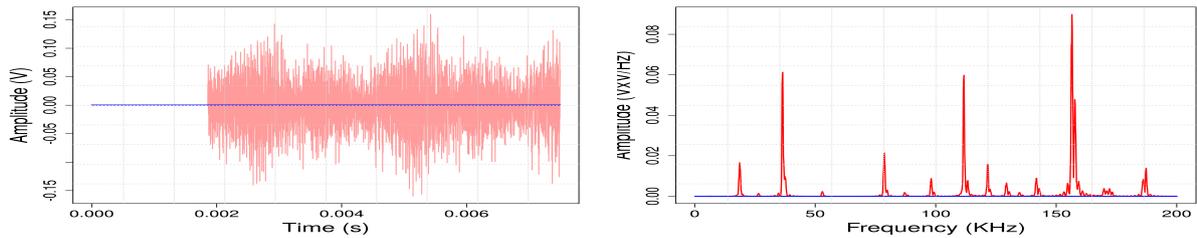
Comme soulignée précédemment (cf. page.44) et illustrée en Figure 3.10b, la principale limitation de la SSA réside dans l'introduction d'artefacts lorsque la séparabilité entre le signal source à estimer et le bruit à supprimer n'est pas bonne, ce qui est souvent le cas pour des bruits large bande. Pour pallier à cette limitation, nous avons alors expérimenté une stratégie en deux étapes. Ici, nous proposons d'améliorer cette méthode en combinant la SSA et la SS.

Cette nouvelle méthode est également composée de deux étapes. La première consiste à garder telle quelle l'étape initiale de l'approche précédente en utilisant la fonction de détection de changements de structure pour identifier les segments du signal bruité, potentiellement associés à des mécanismes physiques. Dans la seconde étape, la décomposition en SSA du signal pré-traité est remplacée par une SS. La Figure 3.18 permet de juger de la pertinence d'une telle approche.

3. Estimation du signal source et détection de salves

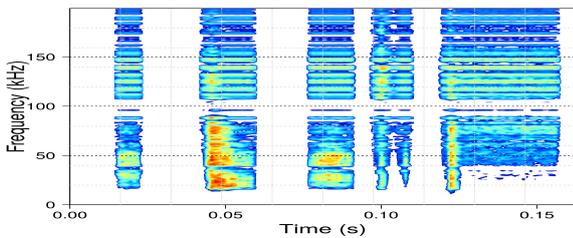


(a) Traitement du bruit résiduel, à la suite d'un débruitage par ondelettes, en utilisant un filtre passe-bande dans le cas du signal source illustré sur la Figure 3.12 et pollué par un bruit mixte (cf. Figure 3.1c). Résultat avant (—) et après (—) traitement

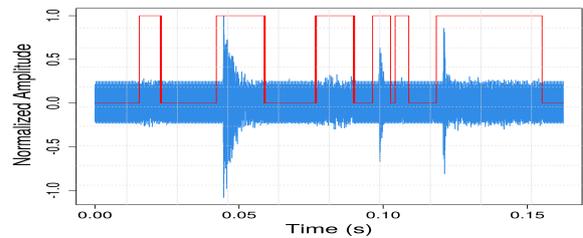


(b) Traitement du bruit résiduel à la suite d'un débruitage par SS, en utilisant un seuillage par ondelettes, dans le cas du signal source illustré sur la Figure 3.12 et pollué par la somme des bruits (cf. Figure 3.1d). Résultat avant (—) et après (—) traitement

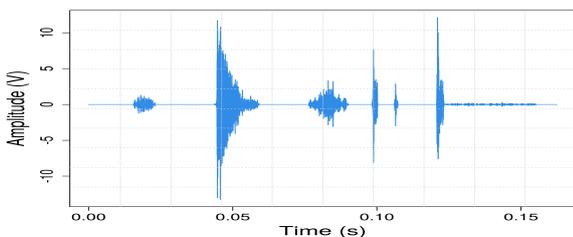
Figure 3.17. – Amélioration de débruitage, par diminution du bruit résiduel, en utilisant une combinaison des méthodes présentées.



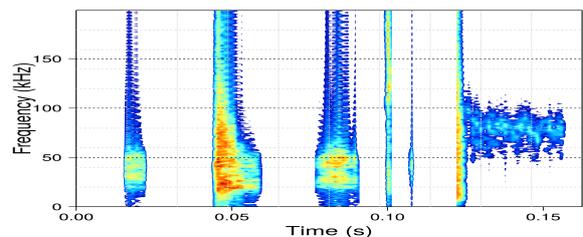
(a) Spectrogramme du signal débruité par utilisation de la SSA



(b) Détection de salves par H-matrix (Etape 1). Signal x à traiter (—) et fonction de détection de changements de structures (—)



(c) Signal débruité (à comparer à celui de la Figure 3.16d) après application de la SS à la suite de l'étape 1



(d) Spectrogramme du signal débruité

Figure 3.18. – Amélioration du débruitage par utilisation de la SSA combinée à la SS.

3.5. Conclusion

L'objectif de ce chapitre était, d'une part, de traiter la problématique de la suppression du bruit environnemental, et d'autre part, d'explorer les performances de méthodes de détection de salves capables de pallier aux limites du seuillage classique. Nous avons pu établir que le bruit environnemental d'un essai RIA est stationnaire. Cependant, sa nature physique étant susceptible de changer d'un essai à l'autre, il est difficile de dégager une méthode adaptée à toutes les situations et applicable pour tous les essais. Notre démarche a donc été de procéder à une étude comparative de plusieurs méthodes de traitement du bruit et d'analyser les résultats selon divers critères.

Lorsqu'on s'intéresse au critère classique d'amélioration du SNR, la SS apparaît de manière générale comme la méthode la plus robuste aux changements de nature du bruit. Cependant, dans certains cas, d'autres approches peuvent lui être préférées. Si le bruit est à bande étroite (respectivement, blanc gaussien), on préférera un simple filtre passe-bande (respectivement, un seuillage par ondelettes).

Dans le but d'obtenir un bruit résiduel quasi-nul quelle que la soit la méthode de débruitage choisie, nous avons étudié les apports de diverses combinaisons des méthodes. Nous concluons qu'il est toujours possible d'atteindre ce résultat, et cela indépendamment de la méthode choisie initialement et du type de bruit.

L'analyse de l'impact des méthodes expérimentées sur les paramètres EA ne permet pas d'établir de hiérarchie entre celles-ci. Cependant, quelles que soient la méthode de débruitage choisie et la nature du bruit, il est préférable (lorsque cela est pertinent) d'interpréter les salves à partir des paramètres liés à la forme d'onde plutôt que ceux liés au contenu fréquentiel. En effet, il est extrêmement difficile de recouvrer avec efficacité les paramètres traduisant le contenu fréquentiel et l'activité acoustique, contrairement à ceux liés à la forme d'onde.

En ce qui concerne la détection des salves, les limitations annoncées de la méthode de seuillage classique ont été mises en évidence, y compris en contexte de fort SNR. Dans ce cas, les ondelettes continues apparaissent comme la méthode de détection la plus efficace, et cela quel que soit le type de bruit. Lorsque le SNR est faible, la détectabilité des salves dépend fortement du type de bruit. Pour certains types de bruit, comme les bruits très large bande par exemple, aucune des approches envisagées ici n'aboutit à des résultats satisfaisants.

4 ■ Classification et interprétation de données EA

Le travail présenté dans ce chapitre a fait l'objet d'un chapitre de livre [150]

Sommaire

4.1. Introduction	58
4.2. Prétraitement des salves	61
4.2.1. Cas des paramètres classiques	61
4.2.1.1. Correction des biais inhérents à la détection des salves	62
4.2.1.2. Réduction de la taille de la matrice de données par suppression des multicollinéarités	64
4.2.2. Prétraitement des courbes dans le cas d'une approche fonctionnelle	66
4.2.2.1. Reconstruction de la forme fonctionnelle et lissage des courbes	67
4.2.2.2. Traitement des différences de support et des décalages horizontaux	68
4.3. Analyse exploratoire : de l'ACP classique à l'ACP fonctionnelle	70
4.3.1. Qu'est-ce qui change ?	70
4.3.2. Analyse des résultats de l'ACP classique et de l'ACP fonctionnelle	71
4.4. Classification non supervisée et identification des mécanismes	74
4.4.1. Exploitation des informations <i>a priori</i> et critère statistique de comparaison des classifications	74
4.4.2. Approche classique	75
4.4.3. Approche fonctionnelle	76
4.4.3.1. Approches et méthodes de clustering envisageables	76
4.4.3.2. Résultats obtenus	77
4.4.4. Recoupement des résultats	81
4.5. Classification supervisée	82
4.5.1. Construction des jeux d'entraînement et test	83
4.5.2. Approche classique	83
4.5.2.1. Régression logistique multinomiale	83
4.5.2.2. Analyse discriminante linéaire	84
4.5.2.3. Les méthodes du type plus-proches-voisins	85
4.5.2.4. Arbre de classification (CART)	86
4.5.2.5. Bagging d'arbres de classification	87
4.5.2.6. Boosting d'arbres de classification	88
4.5.2.7. Forêts aléatoires	88
4.5.2.8. Choix du modèle final	89

4.5.3.	Approche fonctionnelle	90
4.5.3.1.	Stratégie exclusivement non paramétrique	90
4.5.3.2.	Stratégie exclusivement basée sur la réduction de dimension	91
4.5.3.3.	Choix du modèle final	92
4.5.4.	Comparaison des résultats des approches classique et fonctionnelle et prise en charge de salves inconnues	92
4.5.4.1.	Comparaison des performances des méthodes classe par classe	92
4.5.4.2.	Prise en charge de salves inconnues	94
4.6.	Conclusion	96

4.1. Introduction

Avant d'aborder les problématiques de classification et d'interprétation des salves dont il sera essentiellement question dans ce chapitre, un préalable indispensable est la prise en charge du signal x associé à chaque essai, afin d'en extraire les salves s_i . Pour ce faire, en nous appuyant sur les travaux présentés dans les chapitres 2 et 3, diverses stratégies peuvent être dégagées. Nous commençons le chapitre par en proposer une par défaut, qui sera celle appliquée à toutes les données utilisées dans ce chapitre, sauf mention contraire.

La démarche que nous adoptons est de commencer le processus d'analyse des signaux par un débruitage par soustraction spectrale (cf. chapitre 3). L'application de cette méthode sur le signal essai reçu x (Figure 4.1) montre qu'au-delà des performances annoncées, la SS permet également d'éliminer l'impact du dispositif d'essais (cf. chapitre 2). D'un point de vue mathématique, nous pouvons donc supposer que l'utilisation de la méthode de soustraction spectrale comme outil de débruitage permet de passer de l'expression (0.1) du signal essai à l'expression

$$x(t) = h_2 * s(t), \quad t \in \{1, \dots, N\} \quad , \quad (4.1)$$

où les impacts du bruit environnemental et du dispositif d'essais ont été corrigés.

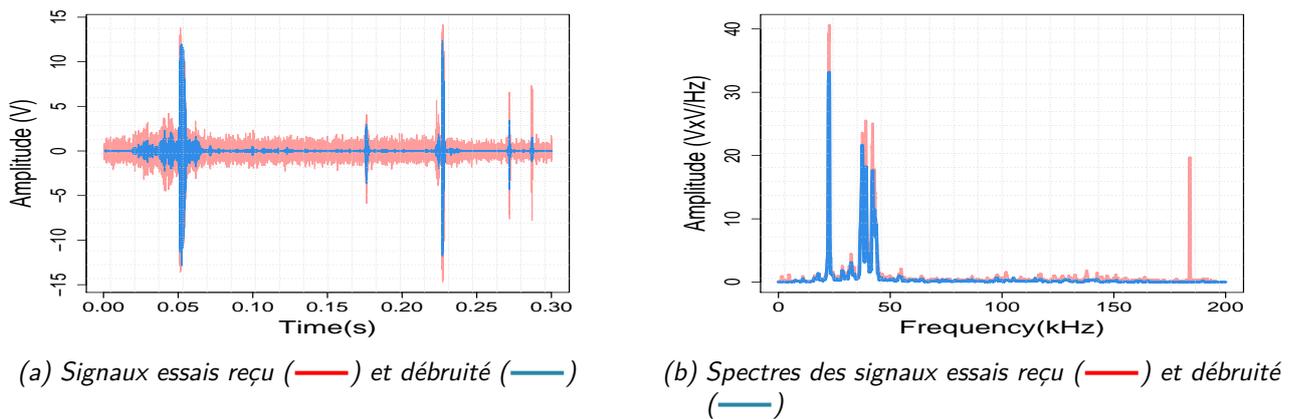


Figure 4.1. – Illustration du traitement du bruit de fonctionnement du signal essai reçu au capteur M1. Cas d'un essai rupture réalisé en 1998.

A la suite de ce débruitage et avant d'aborder la question de la fonction de transfert h_2 du système d'acquisition, nous proposons de procéder à la détection des salves (Figure 4.2). Ce choix, plutôt qu'un traitement de h_2 avant la détection des salves, s'explique par le fait que l'impact du système d'acquisition est susceptible d'évoluer selon le contenu fréquentiel de la salve. Par ailleurs, on notera que la stratégie de détection des salves utilisée ici est basée sur un seuillage par variance mobile. Cette méthode a notamment la spécificité d'être plus sensible qu'un seuillage s'appuyant sur l'évolution de l'énergie du signal (cf. chapitre 1).

L'application de cette stratégie permet, pour les quatorze essais réalisés à ce jour, de disposer d'un échantillon de 168 salves de la forme :

$$x_i(t) = h_2 * s_i(t), \quad t \in \{1, \dots, N\} \quad (4.2)$$

où $i \in \{1, \dots, 168\}$ est le numéro de la salve.

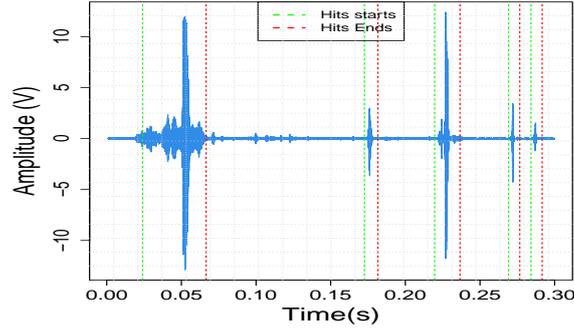


Figure 4.2. – Illustration de la mise en œuvre de l'algorithme de détection des salves.

Notons que des informations habituellement cruciales pour une compréhension fine de h_2 , comme les courbes d'étalonnage des capteurs, n'ont pas pu être exploités dans le cadre de ces travaux. Par ailleurs, l'hostilité de l'environnement nucléaire nécessite l'utilisation de nouveaux capteurs à chaque essai. Malgré des caractéristiques similaires, ces changements peuvent engendrer des évolutions de h_2 .

Afin d'étudier et comprendre l'impact de h_2 sur les signaux, nous avons donc adopté une approche exploratoire consistant à analyser le dictionnaire des salves obtenu à la suite de la détection. Celui-ci a permis de mettre en évidence qu'une fréquence de résonance des capteurs se situant autour de 25kHz était de nature à biaiser l'interprétation des spectres dans certains cas. Pour traiter cet impact, nous proposons d'utiliser une fonction de pénalisation, au lieu d'un filtre passe-bande dont l'utilisation entraînerait la perte de toute information relative à cette gamme de fréquences. En effet, sachant que dans notre cas certains mécanismes sources émettent également autour de 25kHz , cette stratégie permet d'atténuer l'effet d'échelle introduit par le capteur, tout en préservant l'information relative au mécanisme source.

Soit $X_i(f)$ la transformée de Fourier d'un signal salve $x_i(t)$ de l'échantillon, $X_i(f)$ est de la forme :

$$X_i(f) = |X_i(f)| e^{j\theta_{x_i}(f)} \quad (4.3)$$

où $j^2 = -1$ et $\theta_{x_i}(f)$ est la phase de la salve pour la fréquence f . Le traitement que nous proposons consiste donc tout d'abord à vérifier si

$$\arg \max_f (|X_i(f)|) \in [25 - \epsilon, 25 + \epsilon] \quad \text{avec } \epsilon \in \mathbb{R}_+.$$

Dans le cas positif, nous supposons alors que le spectre de $x_i(t)$ est susceptible d'être corrompu par la résonance du capteur. Nous définissons ainsi l'estimateur $\hat{s}_i(t)$ du signal source associé à la i -ème salve, telle que la transformée de Fourier $\hat{S}_i(f)$ de $\hat{s}_i(t)$ s'écrive :

$$\hat{S}_i(f) = |\hat{S}_i(f)| e^{j\theta_{x_i}(f)}, \quad (4.4)$$

où

$$|\hat{S}_i(f)| = \begin{cases} \alpha |X_i(f)| & \text{si } f \in [25 - \epsilon, 25 + \epsilon] \\ |X_i(f)| & \text{sinon} \end{cases} \quad (4.5)$$

avec $\alpha \in (0, 1)$ le paramètre de pénalisation de l'énergie. Le résultat de cette stratégie permettant de pénaliser l'impact du système d'acquisition sur les salves est illustré sur la Figure 4.3.

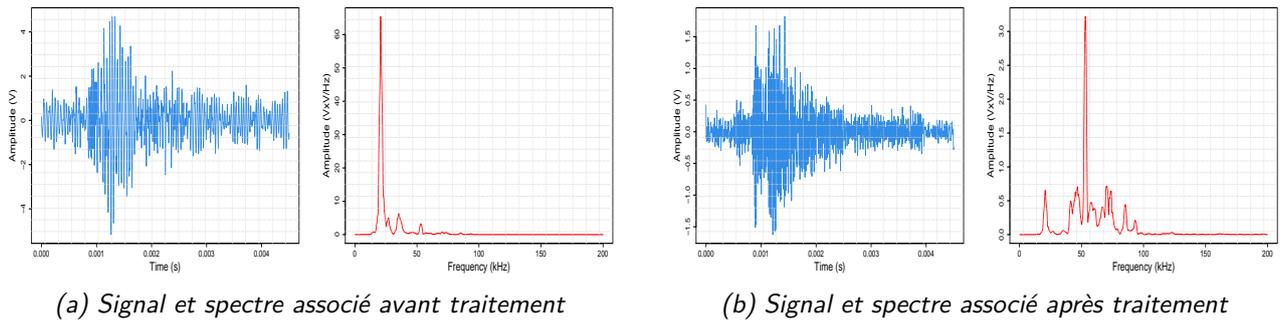


Figure 4.3. – Illustration de la mise en œuvre de la méthode de pénalisation de l'impact du système d'acquisition dans le cas de la salve 128. Signal reçu au capteur M1.

La démarche présentée ci-dessus permet donc de disposer de 168 salves (par capteur) correspondant à divers mécanismes physiques liés ou non au comportement du couple combustible/gaine. En général, l'identification du mécanisme lié à certaines salves, comme la rupture de la gaine du combustible ou à la chute des barres de contrôle permettant de mettre fin aux essais, est immédiate. Cependant, ceux-ci ne représentent qu'une faible proportion de l'échantillon de 168 salves. D'autres mécanismes comme ceux précurseurs ou résiduels d'une rupture de gaine restent à identifier, s'ils existent. Par ailleurs, toute stratégie d'identification des mécanismes basée uniquement sur l'expertise de l'expérimentateur serait difficilement pérennisable et sujette à une subjectivité pouvant conduire à des erreurs d'identification plus ou moins importantes selon l'expérimentateur. Il est donc naturel de privilégier la mise en place d'une démarche visant à proposer un algorithme permettant, d'une part, d'identifier automatiquement des mécanismes déjà observés, et d'autre part, d'alerter les expérimentateurs quant à l'avènement d'un nouveau type de mécanisme physique.

Le point de départ du processus visant à créer un tel outil est la caractérisation des salves. Comme nous l'avons souligné dans l'état de l'art (cf. page 12), cette caractérisation peut reposer sur diverses approches. La plus courante dans la littérature, que nous appellerons par la suite «**approche classique**», consiste à s'appuyer sur des paramètres EA quantitatifs obtenus à partir du signal temporel et de son spectre (cf. Annexe A.1). D'un point de vue statistique, une seconde approche, que nous appellerons «**approche fonctionnelle**», consisterait à considérer chaque salve $\chi_i = \{\chi_i(t); t \in \mathcal{T}_i\}$ comme la réalisation d'une variable aléatoire fonctionnelle \mathcal{X} . Depuis le milieu des années 2000, cette seconde approche suscite un engouement de plus en plus croissant dans la communauté statistique. On retrouve notamment des applications en météorologie, en médecine, en chimie ou encore en linguistique. Le lecteur pourra se reporter aux travaux de ULLAH & FINCH [151] pour un état de l'art sur les applications de l'analyse de données fonctionnelles. Au-delà de l'objectif de mise en place d'un algorithme de classification statistique des signaux évoqué ci-dessus, un autre objectif de ce chapitre sera également de positionner cette approche fonctionnelle innovante par rapport à l'approche classique.

Dans les deux cas, la démarche statistique est plutôt standard et commence par un prétraitement des salves, afin de garantir la fiabilité, et si possible, l'utilité de l'information exploitée par la suite. En second lieu, il s'agit de réaliser une analyse exploratoire de l'échantillon de 168 salves pour en comprendre la structure et identifier d'éventuelles anomalies. Ensuite, vient la phase de classification non supervisée (ou clustering), visant à répartir les 168 salves en divers groupes de mécanismes, dont il faudra identifier la nature physique en s'appuyant notamment sur de l'expertise physique. Enfin, la dernière phase consiste à automatiser la reconnaissance des mécanismes. Ces différentes étapes constituent le fil conducteur de cet ultime chapitre.

Le travail présenté dans ce chapitre est basé pour partie sur les travaux [152–154] et plus particulièrement sur le travail de stage de LLOBELL [155] co-encadré par O.I. Traoré.

4.2. Prétraitement des salves

La Figure 4.4 représente un échantillon de 12 salves choisies parmi les 168 détectées au cours des essais menés depuis 1993. Nous avons vu que ces salves sont issues d'un processus de détection où beaucoup de précautions ont été prises, afin d'assurer les meilleurs résultats possibles. Cependant, dans un contexte comme celui des essais RIA où le nombre et la nature des salves (en termes de forme d'onde et de contenu fréquentiel) sont importants, il est difficile de détecter de manière optimale toutes les salves avec des paramètres de détection globaux. Par ailleurs, quelle que soit l'approche envisagée pour la caractérisation des salves, il faut s'assurer que les valeurs des paramètres EA exploités traduisent fidèlement les informations physiques qui leur sont associées. Dans cette partie, selon l'approche (classique ou fonctionnelle), nous proposons deux processus de prétraitement applicable par défaut à toutes les salves, de manière à améliorer la fiabilité des paramètres EA.

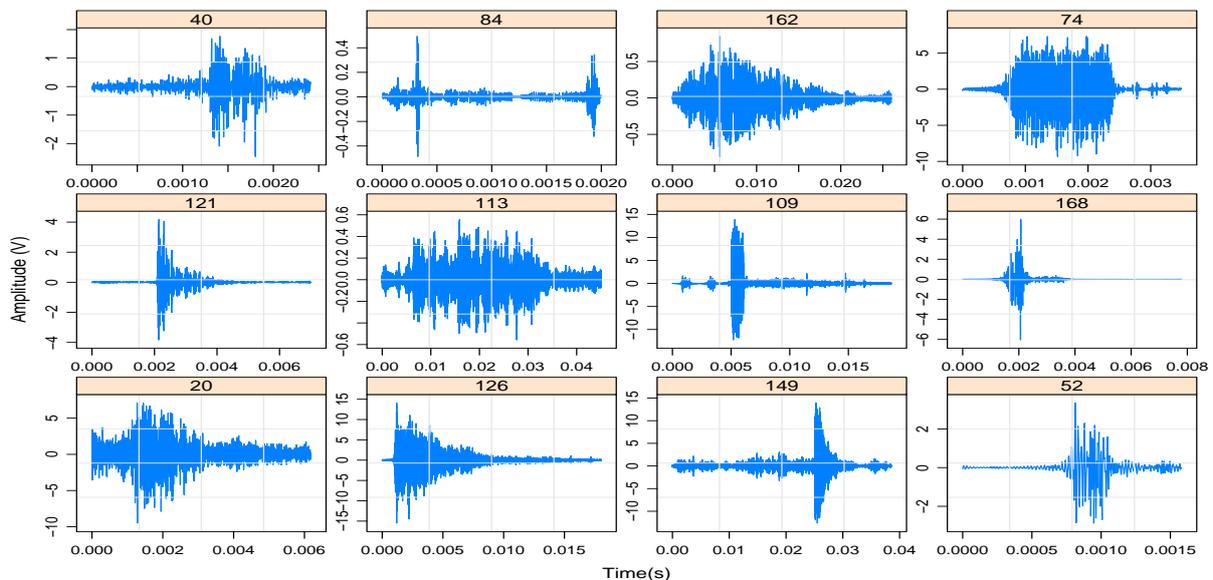


Figure 4.4. – Échantillon de 12 salves (reçues au capteur M1) choisies parmi les 168 détectées au cours des essais menés depuis 1993.

Le lecteur notera que dans le reste de ce chapitre, nous ne ferons pas de distinction entre, d'une part, les termes paramètres EA, prédicteurs et variables, et d'autre part, les termes salve, individu et observation, sauf dans le cas spécifique de la variable réponse. Par ailleurs, nous attirons l'attention sur le fait que **les notations utilisées en introduction seront réutilisées pour désigner d'autres objets mathématiques**. Cependant, le contexte permettra à chaque fois de lever toute ambiguïté.

4.2.1. Cas des paramètres classiques

Dans le cas classique, il existe un très grand nombre de paramètres EA (cf. Annexe A.1). Il est évident que, selon le contexte, seul un nombre limité de ces paramètres est nécessaire pour l'identification du mécanisme associé à une salve. Pour notre part, nous mènerons nos travaux à partir des 23 variables

non surlignées en bleu dans les Tableaux A.1 à A.3 de l'Annexe A.1. Parmi ceux-ci figurent des paramètres relatifs à la forme d'onde de la salve (temps de montée, temps de chute, amplitude...), à son contenu fréquentiel (fréquence moyenne, fréquence maximale, fréquence à la montée) et à l'activité acoustique (nombre de coups, nombre de coups au pic...). Par ailleurs, nous considérerons également des paramètres n'ayant pas d'interprétation physique immédiate (skewness, kurtosis...).

4.2.1.1. Correction des biais inhérents à la détection des salves

Lorsqu'on observe les salves présentées en Figure 4.4, il est clair qu'une application stricte de la définition de certains paramètres, comme le temps de montée ou du temps de chute, conduirait à des valeurs biaisées. Cela s'explique par une mauvaise estimation du temps de début et/ou de fin de la salve, ou par une position plutôt "incertaine" de l'amplitude maximale de salve. Afin d'obtenir des variables mieux interprétables d'un point de vue physique, nous proposons d'appliquer par défaut le processus de prétraitement en trois étapes présenté ci-dessous.

a. Correction d'une éventuelle mauvaise détection du début ou de la fin de la salve

Soit \mathcal{Z}_i le signal analytique associé à la salve χ_i tel que :

$$\mathcal{Z}_i(t) = \chi_i(t) + j\mathcal{Y}_i(t) = \mathcal{E}_i(t)e^{j\phi(t)} \quad ,$$

où $\mathcal{Y}_i(t)$ est la transformée de Hilbert de $\chi_i(t)$:

$$\mathcal{Y}_i(t) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{\chi_i(\tau)}{t - \tau} d\tau,$$

$\phi(t)$ est la phase instantanée et $\mathcal{E}_i(t)$ l'enveloppe de $\chi_i(t)$ telle que :

$$\mathcal{E}_i(t) = \sqrt{\chi_i^2(t) + \mathcal{Y}_i^2(t)}. \quad (4.6)$$

En nous inspirant du seuil flottant proposé par HUGUET [44], notons par I_i l'intervalle défini par :

$$I_i = \{t \in \mathcal{T}_i, \mathcal{E}_i(t) > \alpha_1 \max(\mathcal{E}_i(t))\}, \quad \alpha_1 \in (0, 1) \quad . \quad (4.7)$$

Si $I_i = \mathcal{T}_i$, la salve est laissée telle quelle, sans aucune modification. Autrement, si $I_i \neq \mathcal{T}_i$, une opération de correction est réalisée, en substituant le support \mathcal{T}_i de χ_i par I_i , tel que $\chi_i^* = \{\chi_i(t), t \in I_i\}$. Si le support I_i de χ_i^* est continu, comme dans le cas de la salve 121, le processus s'arrête et χ_i est remplacé par χ_i^* (Figure 4.5).

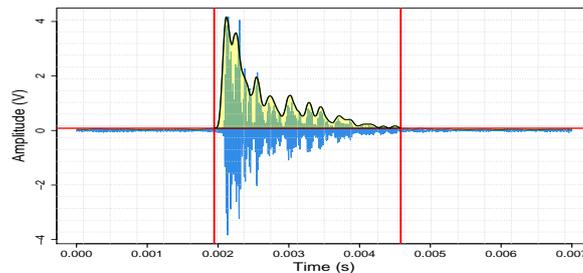


Figure 4.5. – Illustration de la correction des biais de détection dans le cas de la salve 121. Le support I_i de χ_i^* est indiqué par les lignes verticales en rouge. $\alpha = 0.002$

b. Suppression d'éventuels maxima locaux

Lorsque le support I_i de χ_i^* est discontinu, comme dans le cas de la salve 84, l'opération précédente ne suffit plus (Figure 4.6). En effet, I_i peut alors s'écrire comme l'union d'au moins deux intervalles disjoints et non connexes :

$$I_i = \bigcup_{j=1}^J I_{ij}, \quad J \geq 2 \text{ tel que } \forall j \neq j' \in \{1, \dots, J\} \text{ on ait } I_{ij} \cap I_{ij'} = \emptyset \quad (4.8)$$

Dans ce cas, nous choisissons de ne garder qu'un seul intervalle I_i^* à support continu tel que :

$$I_i^* = \{I_{ij}, \max(\mathcal{E}_i(t)) \in I_{ij}\} \text{ avec } j \in \{1, \dots, J\} \quad (4.9)$$

Cette dernière opération permet alors de faire un zoom uniquement autour de l'amplitude maximale du signal en négligeant les maxima locaux.

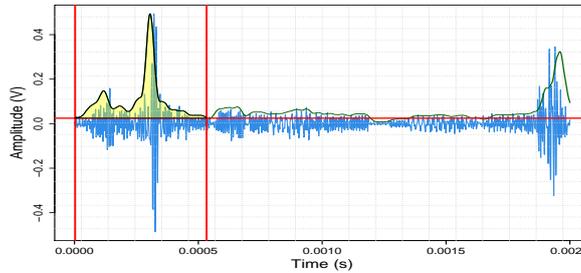


Figure 4.6. – Illustration de la correction des biais de détection dans le cas de la salve 84. Le support I_i^* de χ_i^* est indiqué par les lignes verticales en rouge. $\alpha_1 = 0.12$

c. Prise en compte d'un éventuel temps de plateau

Une position incertaine de l'amplitude maximale du signal correspond en général à un signal plutôt entretenu, comme dans le cas de la salve 74 (Figure 4.7). Une fois un éventuel zoom réalisé grâce au processus présenté ci-dessus, nous introduisons ici une notion de temps de plateau consistant à décider s'il faut recalculer ou non les temps de montée et de chute. Notons I_i^{**} le support d'un éventuel plateau, tel que :

$$I_i^{**} = \{t \in I_i^*, \mathcal{E}_i(t) \geq (1 - \alpha_2) \max(\mathcal{E}_i(t))\}, \quad \alpha_2 \in (0, 1) \quad (4.10)$$

Notons par $\Delta_{I_i^{**}}$ (respectivement, $\Delta_{I_i^*}$) la longueur de l'intervalle I_i^{**} (respectivement, I_i^*). Plus le rapport $\frac{\Delta_{I_i^{**}}}{\Delta_{I_i^*}}$ est grand, plus le temps de plateau est important. Il suffit alors de choisir un seuil *threshold* tel que les temps de montée et de chute soient alors recalculés si $\frac{\Delta_{I_i^{**}}}{\Delta_{I_i^*}} \geq \text{threshold}$. On redéfinit alors ces deux paramètres comme suit :

$$\text{rise Time} = \inf(I_i^{**}) - \inf(I_i^*) \quad , \quad \text{fall Time} = \sup(I_i^*) - \sup(I_i^{**}) \quad (4.11)$$

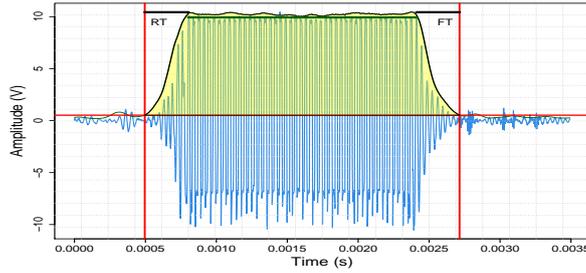


Figure 4.7. – Illustration de la correction des temps de montée et de chute dans le cas d'une salve entretenue (74). Les temps de montée et de chute recalculés sont notés respectivement par RT (rise time) et FT (fall time). $\alpha_2 = 0.05$

Dans la suite, quelle que soit l'approche, nous supposons que ce processus de prétraitement a été appliqué à toutes les salves. Dans le cas de l'échantillon de salves présenté en Figure 4.4, le résultat de ce processus est présenté en Figure 4.8.

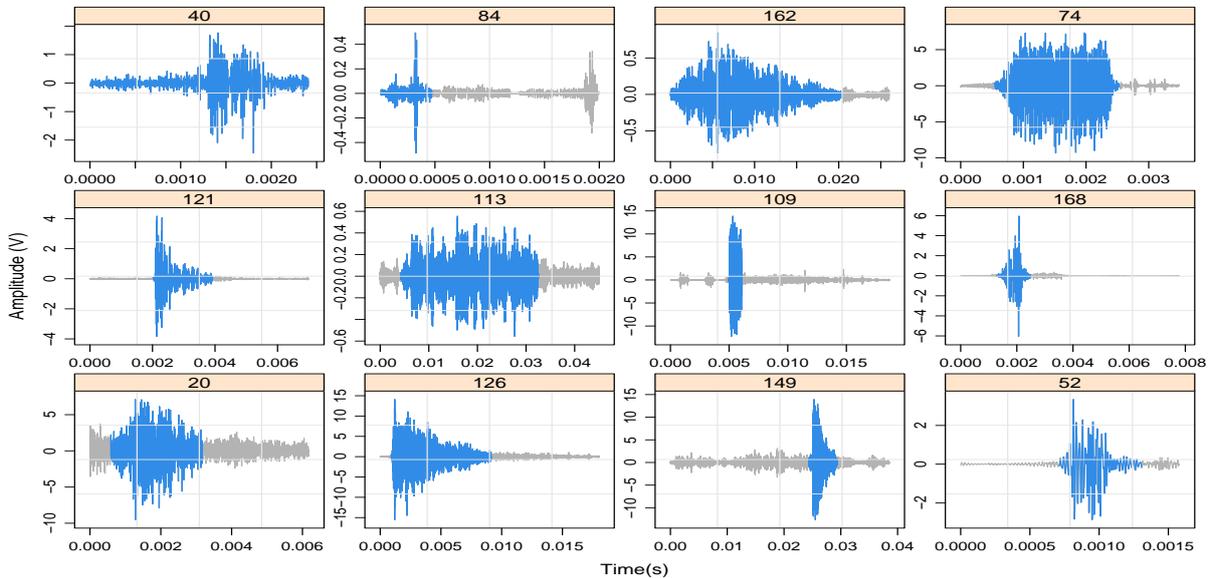


Figure 4.8. – Résultat du processus de prétraitement pour les salves présentées en Figure 4.4. Les salves pré-traitées sont en bleu.

4.2.1.2. Réduction de la taille de la matrice de données par suppression des multicollinéarités

La problématique de sélection de paramètres EA pertinents ne saurait s'appuyer uniquement sur l'expertise de l'expérimentateur, comme nous l'avons fait ici en choisissant 23 paramètres (Annexe A.1). En effet, en plus de s'assurer que les paramètres calculés sont fiables, il convient d'enlever, autant que possible, les variables superflues. Celles-ci peuvent être de deux types : les variables non pertinentes, choisies à tort à la suite de l'expertise, et les variables «redondantes». S'il est difficile d'identifier le premier type de paramètres sans variable réponse, et donc sans un modèle supervisé, plusieurs outils comme la matrice de corrélation peuvent être envisagés dans le second cas. Pour notre part, nous proposons d'utiliser un processus plus général basé sur des modèles linéaires multiples, visant à enlever toutes les variables fortement dépendantes des autres (Algorithme 1) [156].

Considérons tout d'abord des modèles de régression linéaire expliquant chaque variable par les autres et fixons un seuil de qualité du modèle $T \in (0, 1)$. Au début, nous avons donc autant de modèles que de variables. Le processus consiste à supprimer, tant qu'elles existent, toutes les variables qu'on peut expliquer par les autres à l'aide d'un modèle linéaire. Cela se traduit par une élimination de la variable correspondant au modèle ayant le R^2 ajusté (R_{aj}^2) le plus élevé, si ce $R_{aj}^2 \geq T$. Le processus est réitéré avec les variables restantes, jusqu'à ce qu'aucune ne s'associe à un modèle ayant un R_{aj}^2 supérieur ou égal à T .

Dans notre cas d'étude, pour une valeur du seuil $T = 0.85$, 10 variables parmi les 23 initiales ont été trouvées comme étant "redondantes". Les 13 variables restantes ont alors été utilisées pour classifier les 168 salves (Tableau 4.1). Notons que pour ce seuil, le résultat est identique pour les deux capteurs M1 et M2.

Algorithm 1 Sélection des variables les moins dépendantes

Require: $T \in (0, 1)$

Soit p le nombre de variables

Soient x^1, \dots, x^p les variables

$a \leftarrow 1$

while $a \geq T$ **do**

Calculer tous les $R_{aj}^{2(j)} = R_{aj}^2$ des modèles $x_i^j = \sum_{\substack{l=1 \\ l \neq j}}^p \alpha_l x_i^l + \epsilon_i, \quad j = 1, \dots, p$

$a \leftarrow \max_j (R_{aj}^{2(j)})$

$k \leftarrow \arg \max_j (R_{aj}^{2(j)})$

if $a \geq T$ **then**

supprimer x^k

$p \leftarrow p - 1$

réindexer les x^j de 1 à p .

end if

end while

Tableau 4.1. – Variables restantes après élimination des variables redondantes. Ces variables sont celles utilisées pour la classification non supervisée (clustering) des salves.

Paramètre	Unité	Définition
Temps de montée (riseTime)	s	Temps qui s'écoule entre le déclenchement de la salve et l'amplitude maximale
Amplitude	V	Plus forte amplitude observée
Temps de décroissance (fall-Time)	s	Temps qui s'écoule entre l'amplitude maximale et la fin de la salve
Energie (energy)	J	Intégrale de la tension du signal carré
shapeRatio	-	Rapport de l'aire sous la courbe du signal positif par son aire sous la courbe s'il formait un rectangle
Puissance (absoluteEnergy)	J/s	Intégrale de la tension du signal carré divisée par la durée de la salve
Fréquence à la décroissance (fallFreq)	kHz	Rapport entre le nombre de coups à la décroissance et la durée de décroissance
Centroïde du spectre (spec-Cen)	kHz	Fréquence associée au centre de gravité de l'aire sous la courbe du spectre
shapeRatioSpec	-	Rapport de l'aire sous la courbe du spectre par son aire sous la courbe s'il formait un rectangle
skewness	-	Coefficient d'asymétrie
kurt	-	kurtosis normalisé : kurtosis -3 : Coefficient d'aplatissement
statAD	-	Statistique du test d'Anderson-Darling
coefVar	-	Coefficient de variation sur le signal positif

4.2.2. Prétraitement des courbes dans le cas d'une approche fonctionnelle

Puisque dans le cas de l'approche classique tous les paramètres EA calculés à partir de la salve ne sont pas utiles, on en déduit que pour une approche fonctionnelle toute l'information transmise par la courbe salve n'est pas pertinente. Par ailleurs, afin de faciliter la comparaison des courbes correspondant aux différentes salves, il est souvent souhaitable de leur assurer une certaine régularité. En pratique, cela s'obtient en ayant recours à des méthodes de lissage. Cependant, le lissage d'une salve reviendrait à lui appliquer un filtre de type passe-bas, avec le risque de provoquer la perte de toute information relative à certaines composantes hautes fréquences. Nous faisons donc l'hypothèse que l'information contenue dans une salve est entièrement résumée par celles contenues dans l'enveloppe et le spectre du signal correspondant :

$$\mathcal{I}(\chi_i) = \mathcal{I}(\mathcal{E}_i) + \mathcal{I}(\mathcal{S}_i) \quad . \quad (4.12)$$

Ce choix se justifie, d'une part, par le fait que le lissage est une opération tout à fait classique pour l'enveloppe ou le spectre d'un signal. D'autre part, l'enveloppe d'une salve résume l'essentiel des paramètres EA classiques relatifs à la forme d'onde, alors que le spectre permet de retrouver ceux relatifs au contenu fréquentiel.

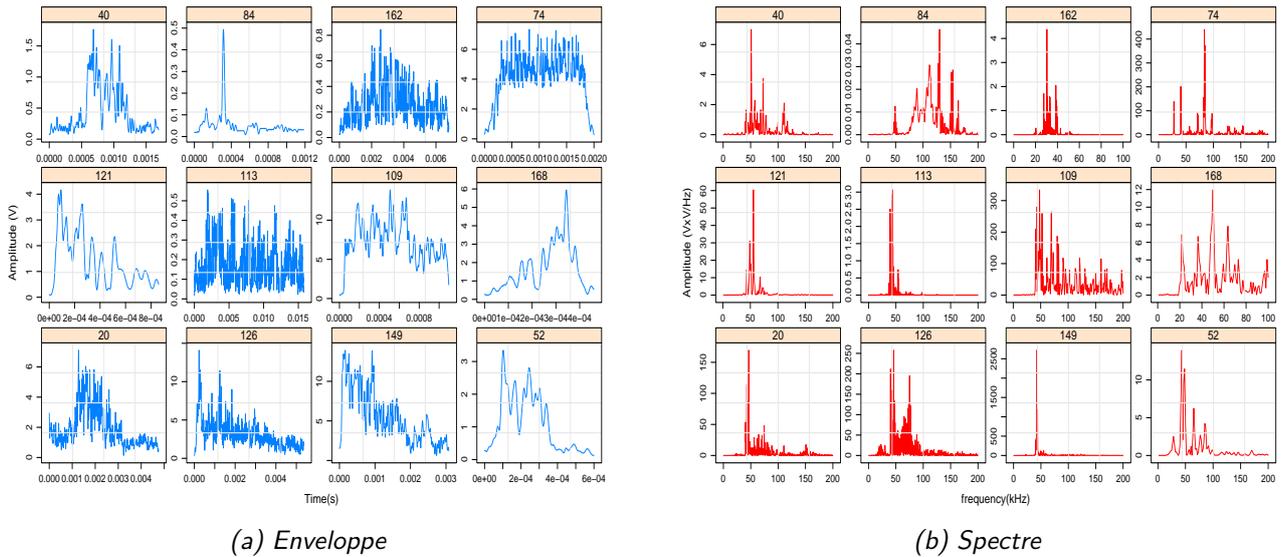


Figure 4.9. – Enveloppe et spectre associés aux 12 salves présentées en Figure 4.8

De manière formelle, une variable aléatoire $\mathcal{X} = \{\mathcal{X}(t); t \in T\}$ est dite fonctionnelle lorsqu'elle prend ses valeurs dans un espace de dimension infinie (fonctionnel). Une observation

$$\chi_i = \{\chi_i(t); t \in \mathcal{T}_i\}$$

de \mathcal{X} est dite donnée fonctionnelle [157, Chap.1]. En pratique, il est évident qu'un tel état de fait n'est pas envisageable, une donnée fonctionnelle ne pouvant être qu'un vecteur de très grande dimension. Par exemple, la i -ème enveloppe \mathcal{E}_i correspond à n_i couples $(t_{ij}, \mathcal{E}_{ij})$, $j = 0, \dots, n_i$, où \mathcal{E}_{ij} est l'information associée à l'enveloppe i au temps t_{ij} . Il convient alors à partir des observations $(t_{ij}, \mathcal{E}_{ij})$ de reconstruire $\mathcal{E}_i = \{\mathcal{E}_i(t); t \in \mathcal{T}_i\}$.

Par ailleurs, les salves χ_i sont susceptibles d'être de durée \mathcal{T}_i différente, mais également d'être acquises lors d'essais pour lesquels le système d'acquisition n'échantillonne pas à la même fréquence f_i . Il faut donc comparer des courbes (enveloppes ou spectres) pouvant avoir des supports différents. En général, ce type d'opération étant non envisageable, il convient d'uniformiser les supports.

Enfin, pour ce qui concerne les enveloppes, en plus du problème de différences de supports, il est nécessaire de traiter celui lié à un éventuel décalage horizontal introduit par les aléas de la détection des salves. En effet, lorsqu'on observe par exemple les enveloppes 126 et 168 (Figure 4.9a), il apparaît clairement que dans le cas de l'enveloppe de la salve 126, l'essentiel de l'information se trouve en début de salve, alors que pour l'enveloppe de la salve 168, c'est le contraire. Il convient donc de les recaler de manière à gommer ce décalage horizontal. Dans le cas d'une métrique de type euclidienne par exemple, sans une opération de recalage, la comparaison des enveloppes des salves 126 et 168 donnerait un résultat aberrant.

Pour résumer, après avoir appliqué par défaut le processus présenté en Section 4.2.1.1, le prétraitement des données fonctionnelles consistera à reconstruire la forme fonctionnelle des courbes, à les lisser, à apporter des solutions aux problèmes de différences de supports et à ceux relatifs aux décalages horizontaux. Notons que toutes ces problématiques sont largement évoquées dans l'ouvrage de référence de RAMSAY & SILVERMAN [158, chap 3-7].

4.2.2.1. Reconstruction de la forme fonctionnelle et lissage des courbes

En général, les mêmes outils pouvant être utilisés pour la reconstruction de la forme fonctionnelle des courbes et pour leur lissage, il s'agit de s'appuyer sur les observations $(t_{ij}, \mathcal{E}_{ij})$ (respectivement, $(f_{ij}, \mathcal{S}_{ij})$) pour créer un modèle permettant d'estimer $\mathcal{E}_i(t)$, $\forall t \in \mathcal{T}_i = [0, t_{n_i}]$ (respectivement, $\mathcal{S}_i(f)$, $\forall f \in [0, f_i/2]$). Par ailleurs, le besoin de lissage introduit une seconde notion de bruit qui, cette fois, n'est pas liée à l'environnement des essais, mais traduit l'information transmise par la source et non pertinente pour son identification. Nous partons donc de modèles du type :

$$\mathcal{E}_{ij} = \mathcal{E}_i(t_j) = \mathcal{E}_i^*(t_j) + \epsilon_i(t_j), \quad \text{avec } t_j \in \{0, \dots, t_{n_i}\} \quad , \quad (4.13)$$

$$\mathcal{S}_{ij} = \mathcal{S}_i(f_j) = \mathcal{S}_i^*(f_j) + \xi_i(f_j), \quad \text{avec } f_j \in \{0, \dots, f_i/2\} \quad , \quad (4.14)$$

avec l'objectif d'estimer $\mathcal{E}_i^*(t)$ et $\mathcal{S}_i^*(f)$ tels que :

$$\mathcal{E}_i(t) = \mathcal{E}_i^*(t) + \epsilon_i(t), \quad \text{avec } t \in \mathcal{T}_i \quad , \quad (4.15)$$

$$\mathcal{S}_i(f) = \mathcal{S}_i^*(f) + \xi_i(f), \quad \text{avec } f \in [0, f_i/2] \quad . \quad (4.16)$$

Pour ce faire, on utilise des fonctions de base ϕ_{ik} et ψ_{ik} telles que :

$$\mathcal{E}_i^*(t) = \sum_{k=1}^{K_i} c_{ik} \phi_{ik}(t), \quad \text{avec } t \in \mathcal{T}_i \quad , \quad (4.17)$$

$$\mathcal{S}_i^*(f) = \sum_{k=1}^{L_i} d_{ik} \psi_{ik}(f), \quad \text{avec } f \in [0, f_i/2] \quad , \quad (4.18)$$

où K_i (respectivement, L_i) est le nombre de fonctions de base ϕ_{ik} (respectivement, ψ_{ik}) choisies pour l'enveloppe \mathcal{E}_i (respectivement, le spectre \mathcal{S}_i). $c_i = (c_{i1}, \dots, c_{iK_i})^T$ et $d_i = (d_{i1}, \dots, d_{iL_i})^T$ sont des vecteurs de coefficients associés aux différentes fonctions de base. En général, ces vecteurs sont

estimés en utilisant la méthode classique de la minimisation de l'erreur quadratique moyenne.

En pratique, selon la nature des courbes plusieurs types de fonctions de base peuvent être envisagés (exponentielles, ondelettes, splines...). Dans le cas de courbes non périodiques, comme les enveloppes et les spectres considérés ici, les fonctions de bases les plus courantes sont les splines, le niveau de lissage de la variable fonctionnelle $\mathcal{E}_i^*(t)$ (respectivement, $\mathcal{S}_i^*(f)$) étant décroissant lorsque le nombre K_i (respectivement, L_i) et l'ordre des splines augmentent. La figure 4.10 illustre la mise en œuvre du lissage par spline dans le cas de l'enveloppe de la salve 149 et du spectre de la salve 126. On y observe notamment qu'un nombre plutôt faible de fonctions splines est suffisant pour reconstruire convenablement les courbes. Cependant, on note qu'un tel choix entraînerait une forte diminution de l'énergie des enveloppes et spectres, ce qui constituerait l'introduction d'un biais d'interprétation. Une solution simple à ce problème consiste à multiplier chaque courbe lissée par un coefficient de manière à réajuster son énergie.

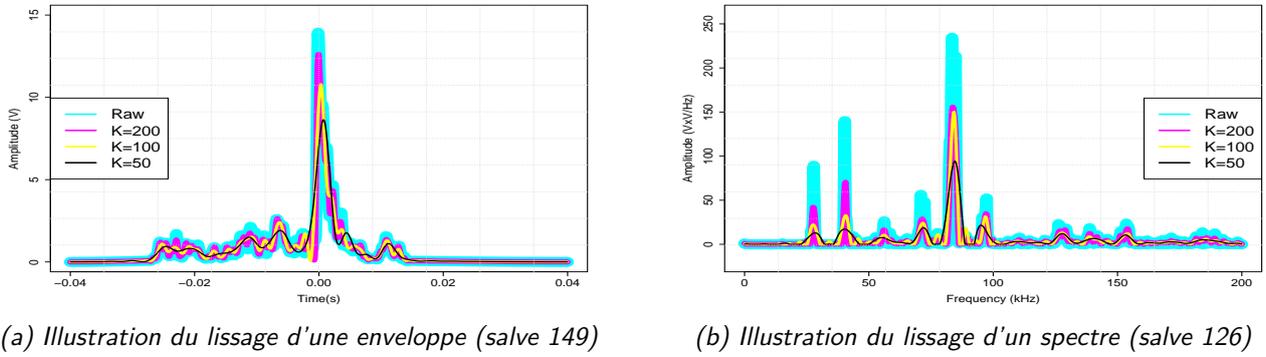


Figure 4.10. – Illustration du processus de lissage des enveloppes et spectres par des fonctions splines cubiques.

4.2.2.2. Traitement des différences de support et des décalages horizontaux

Afin d'assurer le même support à tous les spectres \mathcal{S}_i , l'approche la plus naturelle est de sur-échantillonner toutes les salves χ_i ayant une fréquence d'échantillonnage inférieure à la fréquence d'échantillonnage maximale observée. Le support commun à tous les spectres serait alors $[0, f/2]$, avec :

$$f = \max(f_i) \quad , \quad (4.19)$$

Comme souligné précédemment, dans le cas des enveloppes, en plus du problème de différences de supports, il convient également de traiter celui relatif aux décalages horizontaux. Pour ce faire, nous proposons de traiter chaque enveloppe \mathcal{E}_i en utilisant les transformations ci-dessous.

Soit $\tilde{\mathcal{E}}_i(t)$, $t \in [-(2t_{n_i} - \delta_i), 2t_{n_i} - \delta_i]$, tel que :

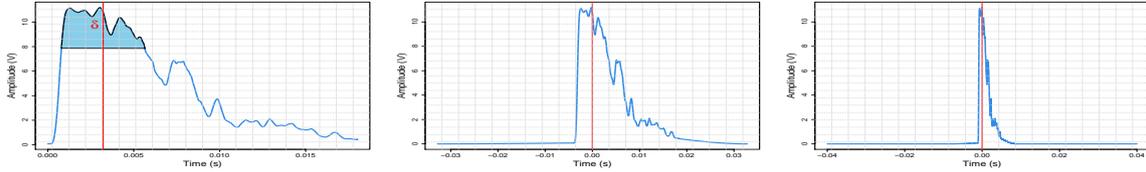
$$\tilde{\mathcal{E}}_i(t) = \begin{cases} \mathcal{E}_i(t + \delta_i) & \text{si } t \in [0, t_{n_i}] \\ 0 & \text{sinon} \end{cases} \quad . \quad (4.20)$$

avec $\delta_i = \text{median}\{t, \mathcal{E}_i(t) \geq \alpha \max(\mathcal{E}_i(t))\}$, avec $\alpha \in (0, 1)$. Ainsi, $\forall i$, on a $\tilde{\mathcal{E}}_i(0) = \mathcal{E}_i(\delta_i)$. Cela revient à recaler toutes les enveloppes de manière à avoir ce qu'on considère comme leur pic au même point.

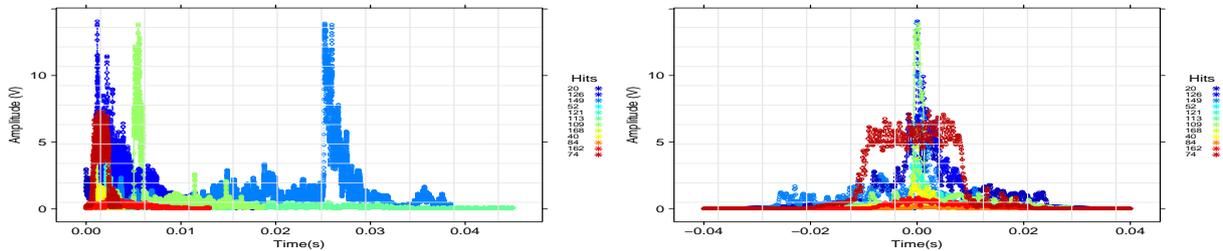
A présent, soient $D \in \mathbb{R}_+^*$ et $D_i = \frac{2t_{n_i} - \delta_i}{D}$. Posons :

$$\tilde{\tilde{\mathcal{E}}}_i(t) = \tilde{\mathcal{E}}_i\left(\frac{t}{2D_i}\right) \quad \text{avec } t \in [-(2t_{n_i} - \delta_i), 2t_{n_i} - \delta_i] \quad , \quad (4.21)$$

Cette dernière opération assure le même support $[-D/2, D/2]$ à toutes les enveloppes. On notera que $\forall i$, si $D > t_{n_i}$ l'opération de renormalisation du support est équivalente à une contraction de l'enveloppe, sinon, à une dilatation (Figure 4.11).

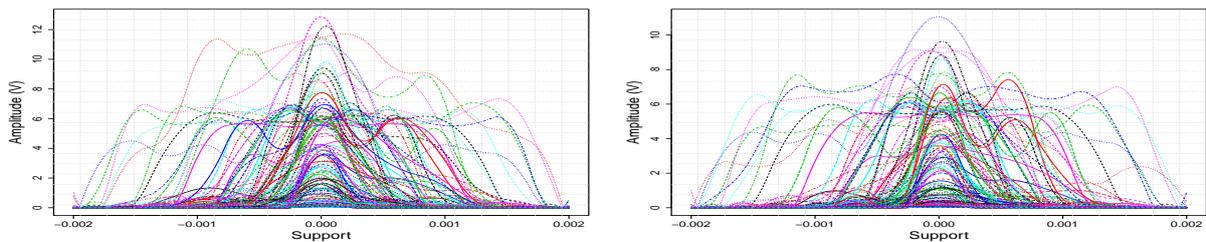


(a) Présentation détaillée du processus de recalage et de contraction de l'enveloppe pour l'enveloppe de la salve 126

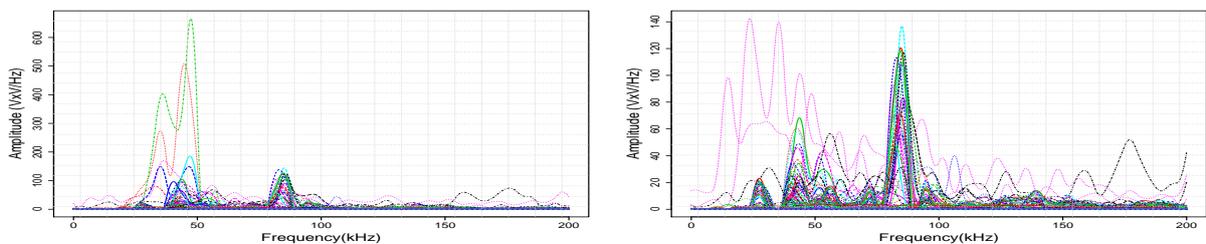


(b) Illustration du processus de recalage et de contraction/dilatation des enveloppes dans le cas des 12 salves présentées en Figure 4.4

Figure 4.11. – Illustration du traitement des différences de supports et des décalages horizontaux dans le cas des enveloppes des salves.



(a) Spline d'ordre 3 et $K=15$



(b) Spline d'ordre 3 et $K=50$

Figure 4.12. – Échantillon de 168 courbes obtenues à la suite du processus de prétraitement des salves reçues aux capteur M1 (droite) et au capteur M2 (gauche).

4.3. Analyse exploratoire : de l'ACP classique à l'ACP fonctionnelle

Nous proposons ici une brève présentation conceptuelle de l'ACP dans le cas des approches classique et fonctionnelle, avec l'objectif de permettre au lecteur de situer les deux ACP l'une par rapport à l'autre.

Dans la suite, pour le cas classique, nous noterons X la matrice de données de taille $N \times p$, \bar{X} la matrice de données centrée et Σ la matrice de variance-covariance de taille $p \times p$:

$$X = \begin{pmatrix} x_1^1 & x_1^2 & \dots & x_1^p \\ x_2^1 & x_2^2 & \dots & x_2^p \\ \vdots & \vdots & \dots & \vdots \\ x_N^1 & x_N^2 & \dots & x_N^p \end{pmatrix} \quad \bar{X} = \begin{pmatrix} x_1^1 - \bar{x}^1 & x_1^2 - \bar{x}^2 & \dots & x_1^p - \bar{x}^p \\ x_2^1 - \bar{x}^1 & x_2^2 - \bar{x}^2 & \dots & x_2^p - \bar{x}^p \\ \vdots & \vdots & \dots & \vdots \\ x_N^1 - \bar{x}^1 & x_N^2 - \bar{x}^2 & \dots & x_N^p - \bar{x}^p \end{pmatrix} \quad (4.22)$$

$$\Sigma = {}^t \bar{X} \bar{X} = \begin{pmatrix} \sigma(x^1) & \sigma(x^1, x^2) & \dots & \sigma(x^1, x^p) \\ \sigma(x^1, x^2) & \sigma(x^2) & \dots & \sigma(x^2, x^p) \\ \vdots & \vdots & \dots & \vdots \\ \sigma(x^1, x^p) & \sigma(x^2, x^p) & \dots & \sigma(x^p) \end{pmatrix} \quad (4.23)$$

Notons que les expressions ci-dessus sous-entendent que la métrique par défaut est la distance euclidienne, et que la matrice de poids par défaut est celle attribuant un poids uniforme à toutes les salves.

4.3.1. Qu'est-ce qui change ?

Rappelons que l'objectif principal d'une ACP est de procéder à une réduction de dimension pour des matrices de taille importante ($p \geq 3$), de manière à faciliter la visualisation et la compréhension de la structure des données, en s'imposant comme contrainte de conserver au maximum leur inertie initiale. Cela revient à représenter les individus dans une base orthonormale $v^1, \dots, v^p \in \mathbb{R}^p$ telle que le score c_i^k associé à l'individu i s'écrit :

$$c_i^k = {}^t v^k \tilde{x}_i = \sum_{j=1}^p v_j^k \tilde{x}_i^j, \quad (4.24)$$

où $\tilde{x}_i = x_i - \mu \in \mathbb{R}^p$ avec $\mu = {}^t(\bar{x}^1, \dots, \bar{x}^p)$, les vecteurs $c^1, \dots, c^p \in \mathbb{R}^n$, appelés composantes principales, étant tels que :

$$\text{pour } k < k', \quad \sigma^2(c^k) < \sigma^2(c^{k'}).$$

La résolution du problème d'optimisation visant à trouver les c^k , et donc les composantes principales, ne sera pas rappelée ici. Nous invitons le lecteur à consulter par exemple [159, Chap.1 et 3]. Nous noterons simplement que les v^k sont les vecteurs propres de la matrice de variance-covariance Σ de X et que les $\sigma^2(c^k)$ sont ses valeurs propres.

Dans le cas fonctionnel, il s'agit de représenter les individus dans une base orthonormale $\{f_k\}_{k \geq 1}$, où le score c_i^k associé à l'individu i s'écrit :

$$c_i^k = \int_{-D/2}^{D/2} f_k(t) \bar{\chi}_i(t) dt. \quad (4.25)$$

où $[-\frac{D}{2}, \frac{D}{2}]$ est le support de $\bar{\chi}_i(t)$, avec $\bar{\chi}_i(t) = \chi_i(t) - \mu(t)$ et $\mu(t) = \frac{1}{N} \sum_{k=1}^N \chi_k(t)$.

Si nous notons

$$\mathcal{V}(s, t) = \frac{1}{N-1} \sum_{k=1}^N (\chi_k(t) - \mu(t))(\chi_k(s) - \mu(s)),$$

l'opérateur de variance-covariance associé à l'échantillon de salves χ_i , les fonctions de pondération $f_k(t)$ sont alors les fonctions propres de \mathcal{V} .

L'une des principales nuances que nous avons notée entre les deux types d'ACP est le nombre maximum de vecteurs de pondération v^k (respectivement, de fonctions de pondération f_k). En effet, dans le cas classique, puisque nous ne disposons que de p variables, le nombre maximum de vecteurs de pondération est égal à p . Pour l'approche fonctionnelle, puisque nous sommes en dimension infinie, on serait donc tenté de conclure qu'il y a une infinité de fonctions de pondération. Cependant, il est important de ne pas perdre de vue que les observations χ_i pour les différentes salves sont indépendantes. Le rang de l'opérateur \mathcal{V} est donc de $N - 1$ et on a donc au maximum $N - 1$ valeur propres non nulles de \mathcal{V} , ce qui fait $N - 1$ fonctions de pondération potentielles.

Pour résumer, en pratique, l'ACP classique repose sur la recherche des vecteurs propres de la matrice de variance-covariance Σ des paramètres EA, alors qu'en ACP fonctionnelle, il s'agit de déterminer les fonctions propres de l'opérateur de variance-covariance \mathcal{V} associé à l'échantillon de courbes χ_i avec $i \in \{1, \dots, N\}$.

4.3.2. Analyse des résultats de l'ACP classique et de l'ACP fonctionnelle

Quelques résultats de mise en œuvre de l'ACP dans les cas classique et fonctionnel sont présentés sur les Figures 4.13 à 4.16, ainsi que dans l'Annexe D.1.

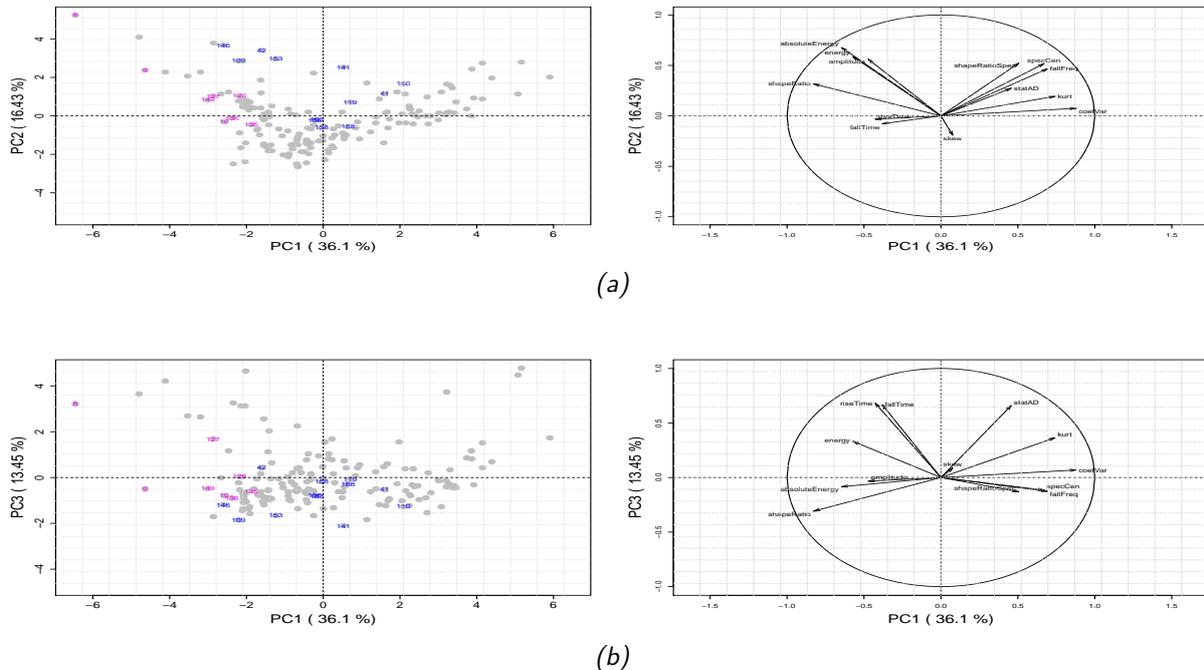


Figure 4.13. – Représentations des individus (salves) et variables (paramètres EA) dans les plans (1,2) et (1,3) à la suite de l'ACP classique (capteur M1).

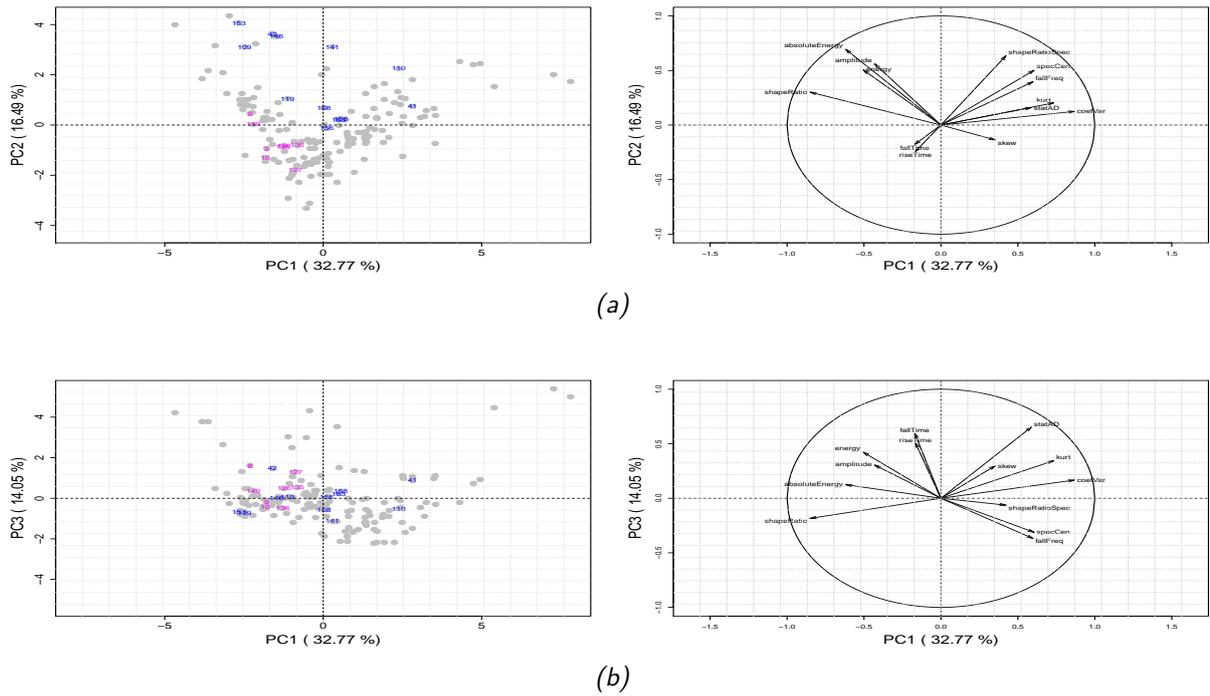


Figure 4.14. – ACP classique : Représentations des individus (salves) et variables (paramètres EA) dans les plans (1,2) et (1,3) à la suite de l'ACP classique (capteur M2)

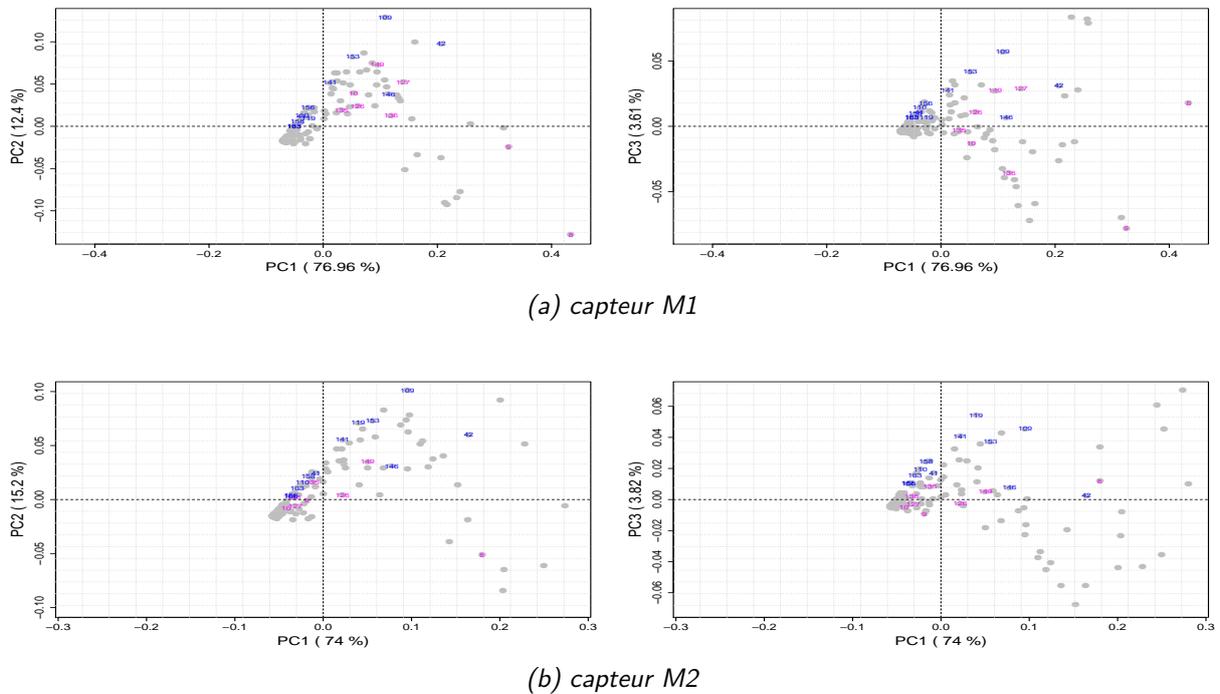


Figure 4.15. – Représentations des individus (salves) dans les plans (1,2) et (1,3) à la suite de l'ACP fonctionnelle sur les enveloppes, pour les capteurs M1 et M2.

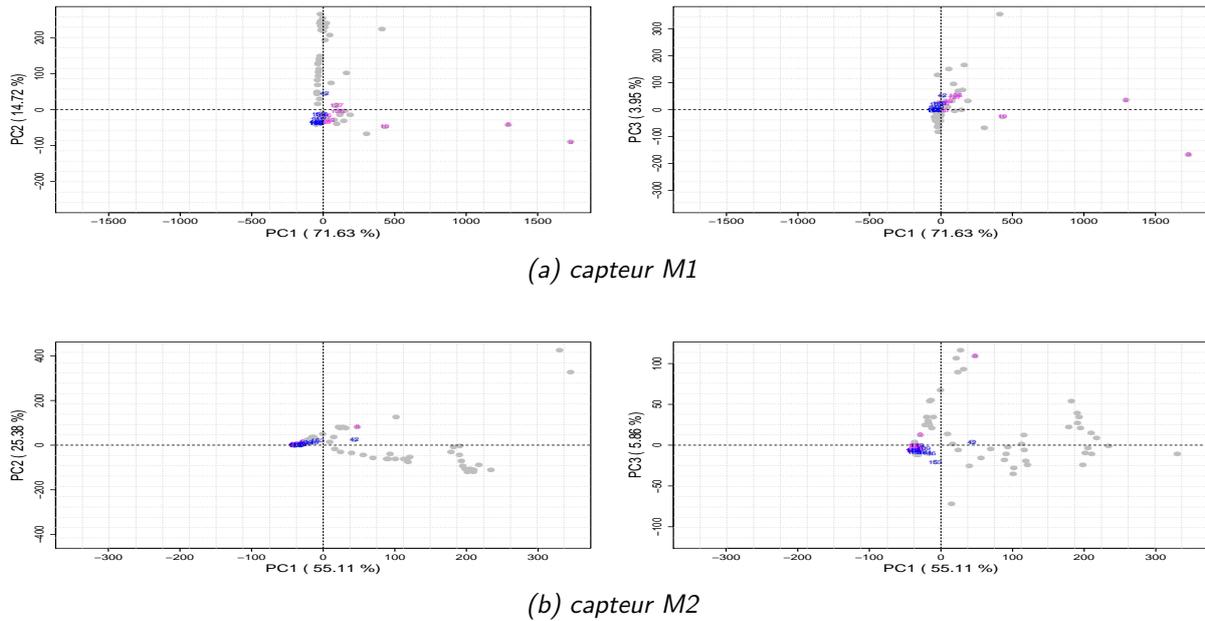


Figure 4.16. – Représentations des individus (salves) dans les plans (1,2) et (1,3) à la suite de l'ACP fonctionnelle sur les spectres, pour les capteurs M1 et M2.

Nous constatons tout d'abord que quelle que soit l'approche choisie (classique ou fonctionnelle), les représentations graphiques des salves dans les plans (1,2) et (1,3) révèlent des résultats quasi-identiques pour les deux capteurs. Ceci permet de soutenir l'hypothèse que l'information contenue dans les deux échantillons de salves (capteurs M1 et M2) est sensiblement la même. Puisque nous ne prévoyons pas d'analyse conjointe des signaux enregistrés aux deux capteurs, dans la suite du chapitre, sauf mention contraire, nous utiliserons les signaux enregistrés uniquement au capteur M1.

Par ailleurs, la structure des nuages de points, notamment dans le cas fonctionnel, révèle l'existence de valeurs extrêmes dont la trop forte contribution à la constitution des axes pénalise grandement la lisibilité des graphiques (cf. Figures 4.15 et 4.16). Enfin, l'analyse des deux premières harmoniques dans le cas des spectres nous permet d'observer que les variabilités entre les salves sont maximales autour de zéro (Figure 4.17). Rappelons que cela constituait l'objectif du travail de recalage des enveloppes (*i.e.* concentrer l'essentiel de l'information autour d'un point de référence choisi). Ce résultat constitue donc une source de validation de la stratégie de prétraitement des enveloppes.

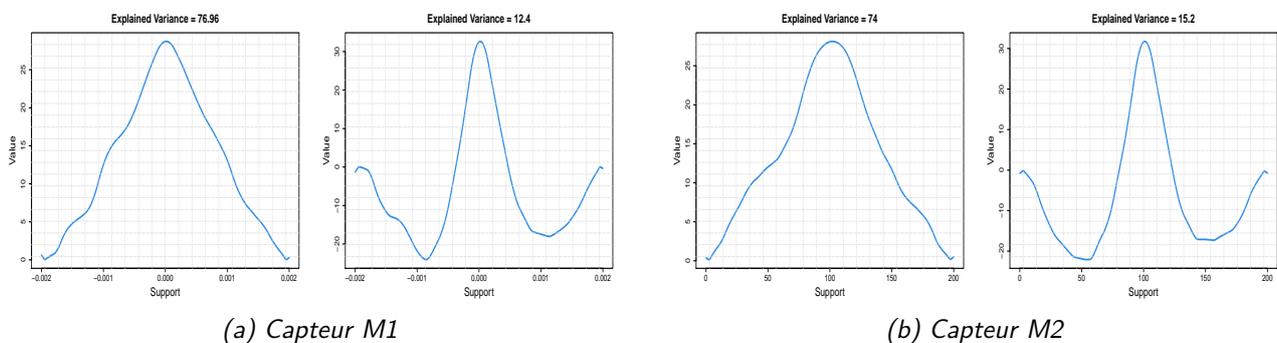


Figure 4.17. – Représentation des deux premières harmoniques (fonctions propres) à la suite de la mise en œuvre de l'ACP fonctionnelle.

4.4. Classification non supervisée et identification des mécanismes

L'objectif principal de cette partie est de créer des groupes homogènes de salves auxquels nous tenterons d'attribuer des mécanismes sources par la suite. Pour ce faire, dans un premier temps, nous expérimentons de manière séparée les approches classiques et fonctionnelles. Ensuite, nous procéderons à un recoupement des classes obtenues, avant de nous appuyer sur de l'expertise physique pour affiner et finaliser les classes.

4.4.1. Exploitation des informations *a priori* et critère statistique de comparaison des classifications

En introduisant ce chapitre, nous avons fait remarquer que la rupture de gaine du combustible et la chute des barres de commande sont deux mécanismes physiques dont l'identification est immédiate. Cela constitue une information *a priori* sur laquelle nous nous appuyons pour comparer les classes obtenues via les différentes méthodes de clustering que nous expérimentons ici. Nous introduisons un critère de validation de classes basé sur la notion de "perte". A chaque fois qu'une rupture ou une chute des barres n'est pas classifiée avec les autres, elle constitue "une perte". Par exemple, nous avons recensé 8 événements de type rupture de gaine. Si pour une méthode, seules 5 ruptures sont dans la même classe, nous concluons alors qu'il y a 3 pertes.

Par ailleurs, la littérature propose un grand nombre de méthodes de mesure de l'homogénéité des classes issues d'un clustering [160]. Dans notre cas, nous utiliserons le critère de la silhouette correspondant, pour chaque observation, à la différence entre sa distance au centre de gravité du groupe dans lequel il est classé et sa distance au centre de gravité du groupe le plus proche. De manière plus formelle, en prenant comme référence le cas de l'approche classique, notons par a_i la distance moyenne entre une salve i et les salves appartenant au même groupe \mathcal{G}_k . On a alors :

$$a_i = \frac{1}{N_k - 1} \sum_{\substack{i' \in \mathcal{G}_k \\ i' \neq i}} d(x_i, x_{i'}). \quad (4.26)$$

Notons par $\mathcal{D}(x_i, \mathcal{G}_{k'})$ la distance moyenne entre x_i et un groupe quelconque $\mathcal{G}_{k'}$ différent de \mathcal{G}_k . On a alors :

$$\mathcal{D}(x_i, \mathcal{G}_{k'}) = \frac{1}{N_{k'}} \sum_{i' \in \mathcal{G}_{k'}} d(x_i, x_{i'}). \quad (4.27)$$

Soit b_i la plus petite distance moyenne telle que :

$$b_i = \min_{k' \neq k} \mathcal{D}(x_i, \mathcal{G}_{k'}). \quad (4.28)$$

Le groupe $\mathcal{G}_{k'}$ pour lequel on obtient la distance moyenne minimale b_i est donc celui auquel on devrait affecter x_i , s'il fallait lui trouver une autre affectation. La largeur de silhouette s_i associée à x_i , comprise entre -1 et 1, se définit alors par :

$$s_i = \frac{b_i - a_i}{\max(a_i, b_i)}. \quad (4.29)$$

Une largeur de silhouette s_i proche de 1 signifie que le point a été affecté à la bonne classe.

A présent notons par \mathfrak{S}_k la largeur silhouette moyenne (encore nommée silhouette moyenne) pour le k -ème groupe. L'indice général \mathcal{C} permettant de mesurer la qualité du clustering est défini par :

$$\mathcal{C} = \frac{1}{K} \sum_{k=1}^K \mathfrak{S}_k = \frac{1}{K} \sum_{k=1}^K \frac{1}{N_k} \sum_{i \in \mathcal{G}_k} \mathfrak{s}_i. \quad (4.30)$$

Dans la suite du travail de classification non supervisée, nous accorderons toujours plus de poids au critère physique du nombre de pertes qu'à celui de la silhouette moyenne. Ce choix se justifie par le fait que les matrices de données ne sont pas les mêmes selon les approches, ce qui est susceptible d'avoir une influence significative sur le critère statistique de la silhouette.

Enfin, notons qu'en ce qui concerne le choix du nombre de groupes K , les résultats d'analyse des essais réalisés jusqu'ici [153, 154] permettent de supposer qu'il existe au moins 6 types de salves. Afin de comparer aisément les résultats obtenus selon les deux approches, nous choisirons $K = 6$ par défaut, quitte à procéder à un affinement manuel par la suite.

4.4.2. Approche classique

L'état de l'art réalisé en préambule de nos travaux (cf. Section 1.2.2.2) a permis de mettre en évidence que dans le cas classique, les méthodes de clustering les plus appréciées par les auteurs étaient les k -means et la CAH. En plus de ces deux méthodes, nous avons choisi d'expérimenter également les k -médoids et la HCPC. Dans le Tableau 4.2, nous rappelons de manière très succincte les principes de ces méthodes. Le lecteur pourra se référer à FRIEDMAN, HASTIE & TIBSHIRANI [161] pour plus de détails.

Tableau 4.2. – Méthodes de classification non supervisée testées

Méthode	Définition
k -means	Algorithme qui part de k centres initiaux, chaque individu étant affecté au centre le plus proche. Puis, à chaque itération les k centres sont redéfinis en prenant le centre de gravité de chaque groupe, jusqu'à stabilisation.
k -médoids	Même principe que les k -means, sauf que les centres affectés à chaque étape sont les individus les plus centraux de chaque classe.
CAH	Agrégation des individus les plus proches au fur et à mesure, ce qui donne un dendrogramme pouvant être coupé en autant de classes que voulu.
HCPC	Méthode CAH appliquée sur les composantes principales.

La comparaison des résultats issus de ces 4 méthodes, en utilisant la matrice de paramètres EA réduite obtenue par suppression des multicollinéarités (cf. Algorithme 1), permet de retenir la HCPC comme meilleure approche de classification non supervisée (Tableau 4.3). Le lecteur notera que la hiérarchie établie ici, ainsi que les résultats présentés, dépendent à la fois des données (les salves) et des paramètres EA. Il ne faut notamment pas perdre de vue que la réduction de dimension par suppression des multicollinéarités ne constitue en aucun cas une sélection de variables au sens statistique du terme. En effet, en expérimentant par exemple diverses combinaisons des variables de la matrice réduite, on pourrait arriver à un clustering avec moins de pertes et une silhouette moyenne plus élevée.

Tableau 4.3. – Tableau récapitulatif des résultats obtenus suite à la mise en œuvre des 4 méthodes de clustering choisies dans le cas classique. Signalons que CDB est l’abréviation de Chute Des Barres.

	Nombre de classes	Pertes de Ruptures	Pertes de CDB	Silhouette moyenne
<i>k</i> -means	6	2	4	0.20
<i>k</i> -médoides	6	3	7	0.18
CAH	6	3	3	0.18
HCPC	6	1	3	0.36

4.4.3. Approche fonctionnelle

4.4.3.1. Approches et méthodes de clustering envisageables

Une fois la forme fonctionnelle des données reconstruite (Section 4.2), la littérature permet de dégager deux principales approches pour leur exploitation statistique. La plus populaire consiste à passer du problème en dimension infinie intrinsèque aux données fonctionnelles à un problème en dimension finie où on pourrait avoir recours aux outils classiques d’analyse de données multivariées [158, 162]. Quant à la seconde approche, elle repose sur des méthodes non paramétriques utilisant des (semi-)métriques permettant de travailler directement avec les courbes [157].

Dans le cas des signaux d’intérêt ici, en ne perdant pas de vue que deux courbes (l’enveloppe et le spectre) sont associées à chaque salve, plusieurs stratégies d’analyse sont envisageables. La première, exclusivement basée sur la réduction de dimension, consiste à créer une matrice de données regroupant les variables obtenues à la suite d’une réduction de dimension appliquée, d’une part, sur les spectres, et d’autre part, sur les enveloppes.

La seconde stratégie exclusivement non paramétrique repose sur l’utilisation de (semi-)métriques permettant de comparer deux salves χ_i et $\chi_{i'}$ à travers une combinaison des résultats de comparaison de leurs enveloppes (\mathcal{E}_i et $\mathcal{E}_{i'}$) et de leurs spectres (\mathcal{S}_i et $\mathcal{S}_{i'}$), telle que :

$$d(\chi_i(t), \chi_{i'}(t)) = \alpha \frac{d_1(\mathcal{E}_i(t), \mathcal{E}_{i'}(t))}{\beta_1} + (1 - \alpha) \frac{d_2(\mathcal{S}_i(f), \mathcal{S}_{i'}(f))}{\beta_2} \quad (4.31)$$

où $\alpha \in [0, 1]$ permet alors de définir le poids accordé à chaque type de courbe. Par exemple, $\alpha = 0$ (respectivement, $\alpha = 1$), signifie une analyse uniquement basée sur les spectres (respectivement, sur les enveloppes). Quant aux paramètres β_k , ils sont dédiés à la correction d’éventuels effets d’échelle entre les (semi-)métriques d_1 et d_2 . En nous inspirant de la normalisation des matrices de données en ACP, une approche de normalisation par défaut est de choisir

$$\beta_k = \sqrt{\text{var}(d_k)} \quad . \quad (4.32)$$

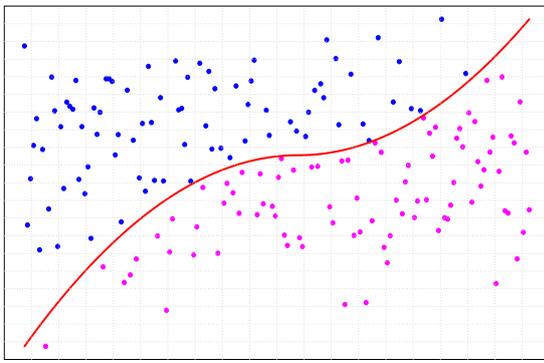
On aurait ainsi par exemple

$$\text{var}(d_1) = \frac{\sum_{i=1}^N \sum_{i'=1}^N (d_1(\mathcal{E}_i(t), \mathcal{E}_{i'}(t)) - \bar{d}_1)^2}{2N^2} \quad \text{avec} \quad \bar{d}_1 = \frac{\sum_{i=1}^N \sum_{i'=1}^N d_1(\mathcal{E}_i(t), \mathcal{E}_{i'}(t))}{2N^2} \quad . \quad (4.33)$$

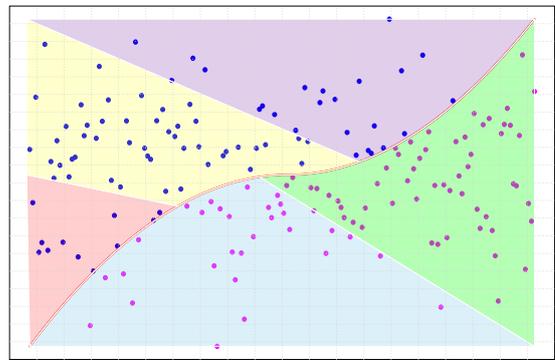
Enfin, une troisième stratégie utilise de manière conjointe l’approche de réduction de dimension et l’approche non paramétrique (Figure 4.18). Il s’agit alors de réaliser une première classification à partir de l’information relative au contenu spectral (les spectres), pour ensuite procéder à un affinement des résultats obtenus en s’appuyant sur l’information relative à la forme d’onde (les enveloppes). D’un point de vue physique, nous justifions l’utilisation dans cet ordre des spectres et enveloppes, par les deux axiomes suivants :

1. deux signaux associés au même mécanisme source ne peuvent avoir des spectres sensiblement différents ;
2. si deux mécanismes sources différents ont le même spectre, leurs enveloppes permettent nécessairement de les dissocier.

Notons que pour cette dernière stratégie, à la lumière des travaux réalisés au chapitre 2, la première hypothèse pourrait ne pas être vraie, si nous ne considérons pas les salves issues des deux capteurs de manière séparée. Une fois cette éventuelle limitation levée en faisant une distinction claire entre les salves selon le capteur associé, le premier axiome et l'expression (4.12) permettent alors de justifier le second.



(a) Etape 1 : utilisation des spectres pour obtenir une première classification



(b) Etape 2 : Affinement des classes issues de la première étape en s'appuyant sur les enveloppes

Figure 4.18. – Schéma illustratif de l'approche de classification non supervisée en deux étapes

Une fois l'approche de classification non supervisée choisie, une autre problématique est de choisir les méthodes de clustering. L'état de l'art réalisé par ULLAH & FINCH [151] permet de conclure que la CAH est la méthode la plus appréciée pour la classification non supervisée des données fonctionnelles. Nous l'adopterons dans la suite de cette partie, à la différence près que les résultats obtenus seront à chaque fois améliorés à l'aide de k -means.

4.4.3.2. Résultats obtenus

Lorsque la stratégie adoptée s'appuie sur une réduction de dimension, l'ACP fonctionnelle s'impose alors comme l'outil de réduction de dimension le plus naturel. Cependant, dans le cas des spectres, il nous est apparu pertinent d'explorer également le potentiel d'une Analyse Factorielle des Correspondances (AFC). En effet, dès lors que l'on revient au fondement intuitif d'un spectre, qui est d'étudier la répartition de l'énergie totale du signal selon différentes gammes de fréquences, une AFC devient envisageable. Dans notre cas, le Tableau 4.4 permet d'ailleurs d'observer que cette approche est plus prometteuse que l'ACP fonctionnelle.

Tableau 4.4. – Tableau récapitulatif des résultats du processus de choix de l’approche d’analyse dans le cas des spectres. Signalons que CDB est l’abréviation de Chute Des Barres. Notons que l’absence du premier axe s’explique par le fait qu’il correspond à une variation commune à toutes les salves, et donc non discriminante.

	Axes choisis	Nombre de classes	Pertes de Ruptures	Pertes de CDB	Silhouette moyenne
ACP fonctionnelle	2,3,4	3	0	6	0.8
AFC	1,2	3	0	0	0.65

Le Tableau 4.5 résume les résultats d’une répartition des 168 salves en 6 classes selon les 3 approches. On constate que la méthode mixte combinant l’approche de réduction de dimension et l’approche non paramétrique est celle permettant d’aboutir aux meilleurs résultats. Cependant, il convient de garder à l’esprit qu’il est difficile d’optimiser les performances des méthodes dans une problématique de classification non supervisée. Cette hiérarchie est donc susceptible d’évoluer en présence d’une variable réponse.

Tableau 4.5. – Tableau comparatif des résultats des différentes approches de classification non supervisée par CAH dans le cas fonctionnel. Les détails de mise en œuvre des méthodes sont présentés dans le Tableau D.1 en Annexe D.2.1

	Nombre de classes	Pertes de Ruptures	Pertes de CDB	Silhouette moyenne
Réduction de dimension	6	4	6	0.5
non paramétrique Fonc.	6	3	6	0.56
Mixte	6	0	5	0.4

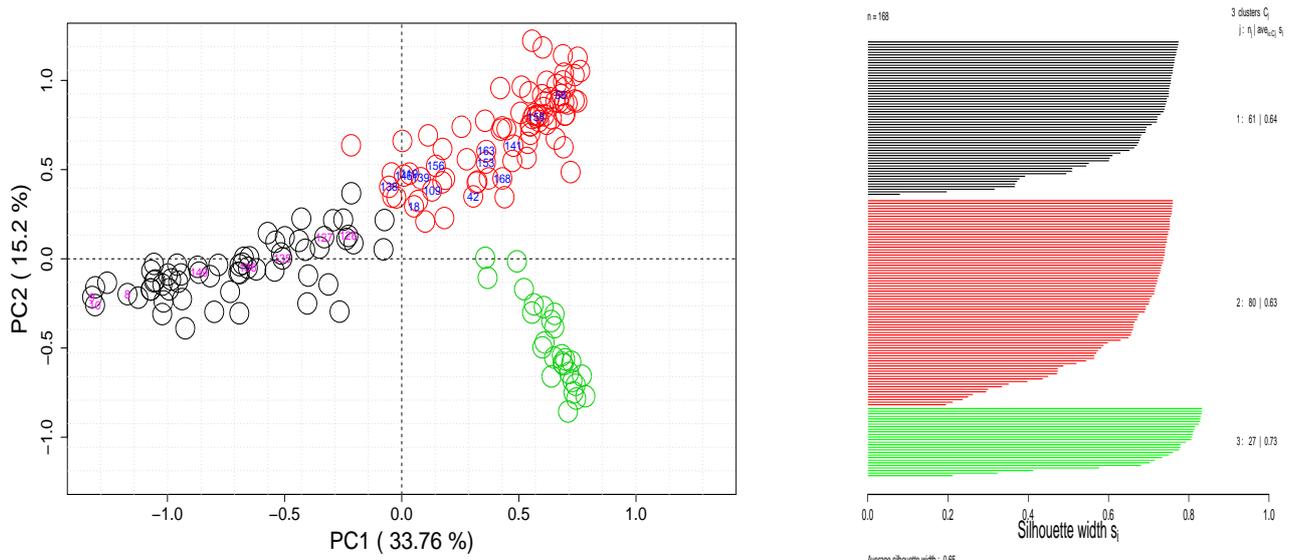
Détails de mise en œuvre du clustering par approche mixte

Étape 1 : Classification des spectres par réduction de dimension

La mise en œuvre d’une CAH, consolidée par la suite en ayant recours à la méthode des k-means, permet de réaliser une première séparation des salves en 3 groupes. La Figure 4.19 permet d’observer que les trois classes obtenues ici sont très homogènes.

Étape 2 : Affinement des classes par application de l’approche non paramétrique sur les enveloppes

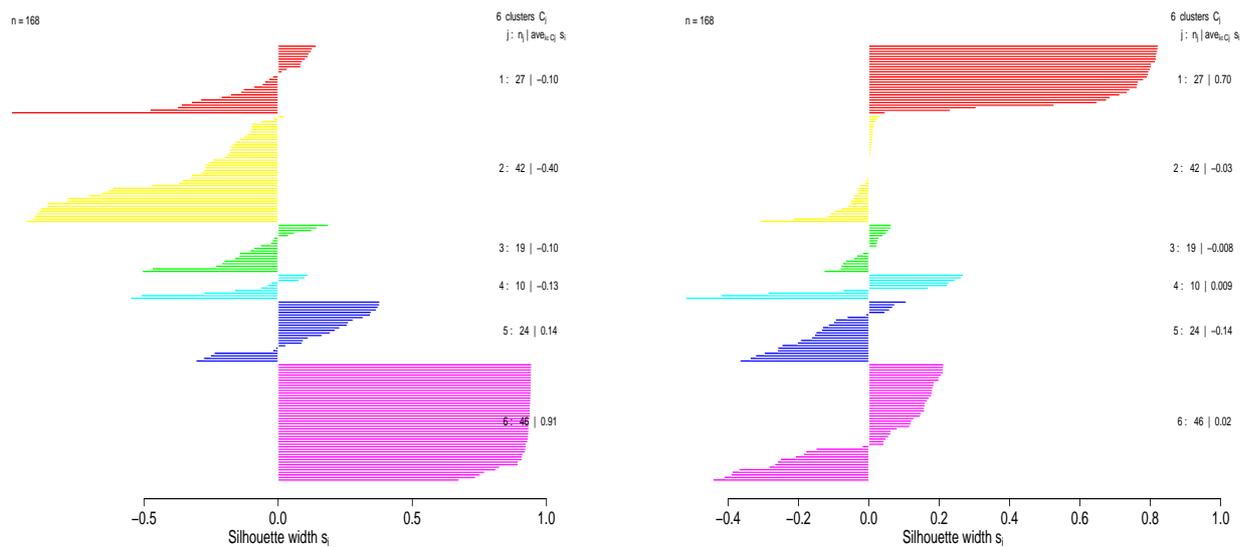
L’analyse des salves appartenant à la classe 3 (en vert sur la Figure 4.19) permet de constater qu’elle est très homogène, en plus d’être très bien séparée des deux autres classes. Nous la garderons donc telle quelle. En ce qui concerne les classes 1 et 2 (respectivement, en noir et rouge), nous procédons ici à leur affinement en nous appuyant, comme à l’étape précédente, sur une CAH, mais cette fois sur les enveloppes. Ainsi, la classe 1 est affinée en deux sous-classes et la classe 2 en trois sous-classes. A l’analyse des graphiques des silhouettes (Figure 4.20), nous observons qu’aucun des deux types de courbes (enveloppes ou spectres) ne permet d’obtenir une répartition homogène des salves en 6 classes. On constate par exemple que les spectres (respectivement, les enveloppes) des salves appartenant à la classe 1 (respectivement, à la classe 6) sont caractéristiques uniquement de ces classes, alors que des salves appartenant à d’autres classes peuvent avoir sensiblement les mêmes types d’enveloppes (respectivement, de spectres). Par ailleurs, lorsqu’on compare par exemple les classes 2 et 3 sur les Figures 4.22 et 4.21, on conclut que l’énergie des salves est un paramètre non négligeable. Enfin, on constate également que cette stratégie semble avoir du mal à isoler convenablement les chutes de barres (Tableau 4.6).



(a) Représentation des individus dans le plan (1,2) à la suite de la mise en œuvre de l'AFC. Les numéros des salves associées à une rupture de la gaine sont en rouge et ceux associés à une chute de barres en bleu.

(b) Graphique des silhouettes des salves par groupe

Figure 4.19. – Résultats de la première étape de mise en œuvre de la stratégie de classification combinant les approches de réduction de dimension et non paramétrique.



(a) Silhouettes associées aux enveloppes

(b) Silhouettes associées aux spectres

Figure 4.20. – Silhouettes obtenues à la suite de la seconde étape de la stratégie de classification combinant les approches de réduction de dimension et non paramétrique.

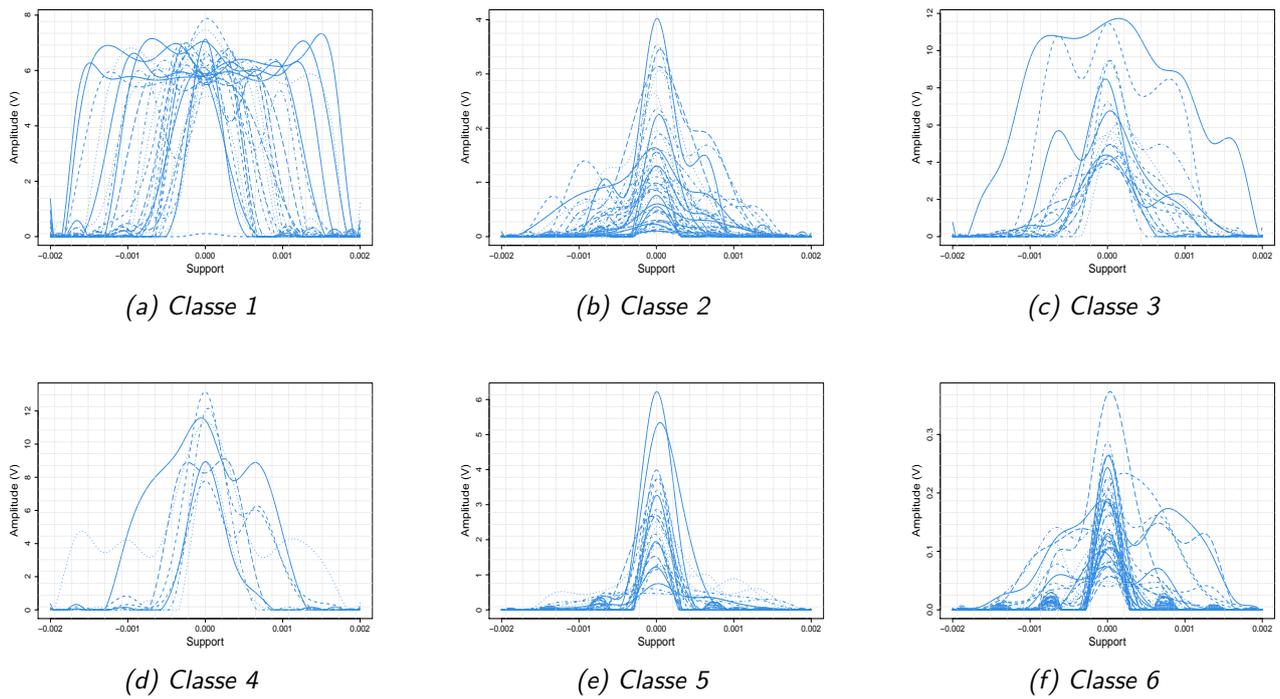


Figure 4.21. – Superposition des enveloppes selon la classe

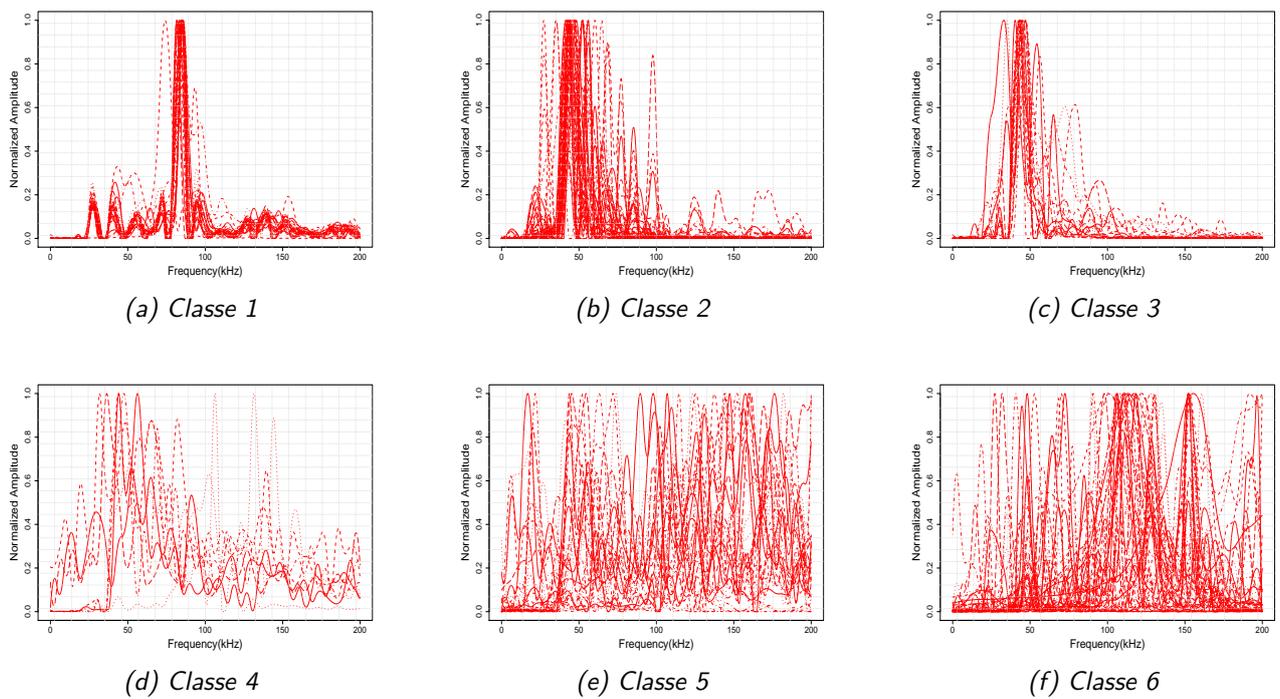


Figure 4.22. – Superposition des spectres normalisés selon la classe.

Tableau 4.6. – Tableau récapitulatif des résultats du processus de choix de l'approche d'analyse dans le cas des spectres

	Nombre de classes	Ruptures	CDB	Silhouette moyenne
Clustering fonctionnel	6	0	5	0.11

4.4.4. Recoupement des résultats

Afin de situer les deux approches l'une par rapport à l'autre, considérons la Figure 4.23 obtenue à la suite d'une AFC de la table de confusion des deux clusterings (Tableau 4.7). Nous observons que la classe B de l'approche fonctionnelle correspond à la classe 2 de l'approche classique. On remarque également que la classe F de l'approche fonctionnelle (respectivement, la classe 4 de l'approche classique) est divisée entre les classes 4 et 5 de l'approche classique (respectivement, C et F de l'approche fonctionnelle). Les résultats de classification obtenus selon l'approche sont donc significativement différents, d'où la pertinence d'un recoupement.

A la suite de ce recoupement et de son affinement par expertise acoustique, nous obtenons finalement une répartition des salves en huit classes. Les résultats d'analyse détaillés de ces classes sont présentés en Annexe D.2.2. Il en ressort notamment qu'en plus des classes déjà identifiées que sont les ruptures de la gaine du combustible (\mathcal{G}_3) et les chutes de barre de contrôle (\mathcal{G}_4), deux classes d'effectifs relativement importants ne correspondent pas à des mécanismes d'intérêt. En effet, les spectres large bande des salves de la classe \mathcal{G}_7 et leur forme d'onde impulsionnelle suggèrent que ce sont des bruits électroniques. Quant aux salves de la classe \mathcal{G}_5 , leur très faible énergie, leur spectre très large bande et surtout leur forme d'onde "aléatoire" permettent de les associer à du bruit résiduel au travail d'estimation du signal source.

Par ailleurs, la quasi-totalité des mécanismes de la classe \mathcal{G}_6 appartiennent à des essais où une rupture de la gaine du combustible a effectivement été observée. Il semble que ce soient des mécanismes précurseurs ou résiduels d'une rupture de la gaine.

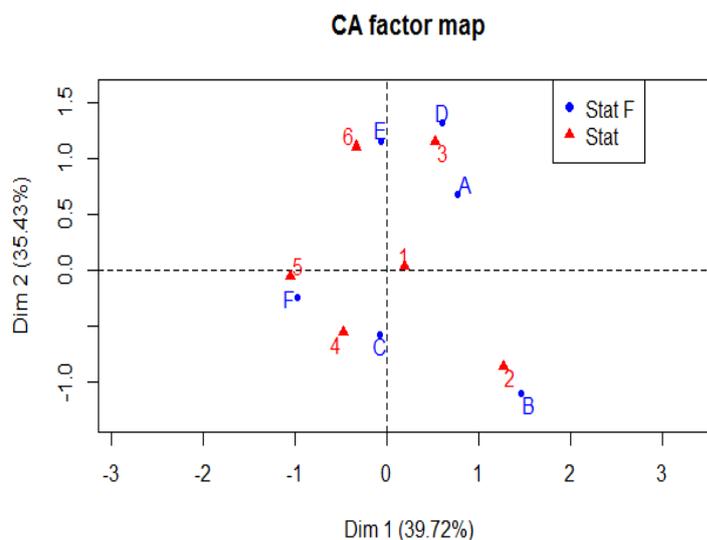
**Figure 4.23.** – Comparaison entre les classes de la classification non supervisée et de la statistique fonctionnelle (Plan 1/2 de l'AFC du Tableau 4.7)

Tableau 4.7. – Table de confusion entre les classes de la classification non supervisée et de la statistique fonctionnelle

	Stat	1	2	3	4	5	6
Stat F	A	3	6	17	1	0	0
	B	0	19	0	1	0	0
	C	5	6	2	25	0	0
	D	1	0	7	0	0	0
	E	0	0	12	0	5	7
	F	1	0	0	18	30	2

Tableau 4.8. – Classes obtenues grâce à l'expertise acoustique et au recouplement avec la statistique fonctionnelle

Classe	Identification de la classe	Effectif
\mathcal{G}_1	Signaux entretenus à faible amplitude	7
\mathcal{G}_2	Signaux entretenus à forte amplitude	26
\mathcal{G}_3	Ruptures	12
\mathcal{G}_4	Chutes de barres	13
\mathcal{G}_5	Bruits non électroniques	28
\mathcal{G}_6	Mécanismes précurseurs ou résiduels d'une rupture	30
\mathcal{G}_7	Bruits électroniques	44
\mathcal{G}_8	Salves ressemblant aux chutes de barres mais qui n'en sont pas ?	6

4.5. Classification supervisée

Le travail de clustering et d'identification des mécanismes nous permet de disposer de $N = 166$ observations (x_i, g_i) (respectivement, (χ_i, g_i) pour le cas fonctionnel) correspondant à N réalisations "indépendantes" de $(\mathbf{x}, \mathcal{G})$ (respectivement, $(\mathcal{X}, \mathcal{G})$) à valeurs dans $\mathbb{R}^p \times \mathcal{G}$ (respectivement, $E \times \mathcal{G}$), avec E l'ensemble des fonctions possibles χ_i et $\mathcal{G} = \{\mathcal{G}_1, \dots, \mathcal{G}_8\}$. Dans cette section, quelles que soient l'approche et la méthode adoptées, l'objectif est de trouver la fonction de prédiction G à valeurs dans \mathcal{G} permettant d'associer une salve donnée à un groupe $g \in \mathcal{G}$, et donc d'identifier le mécanisme source correspondant. Dans le cas classique, lorsqu'une salve caractérisée par le vecteur de paramètres EA $x \in \mathbb{R}^p$ appartient à la classe $g \in \mathcal{G}$, on a :

$$G(x) = g \quad g \in \mathcal{G}. \tag{4.34}$$

En pratique, cela revient à trouver l'estimateur \hat{G} de G à travers la minimisation de l'espérance de la fonction de perte [161, Chap.2]. Le classificateur dit de Bayes s'écrit alors :

$$\hat{G}(x) = \arg \min_{g \in \mathcal{G}} (1 - p_g(x)) = \arg \max_{g \in \mathcal{G}} p_g(x), \tag{4.35}$$

où

$$p_g(x) = \mathbb{P}(\mathcal{G} = g / \mathbf{x} = x). \tag{4.36}$$

Ainsi :

$$\hat{G}(x) = \mathcal{G}_k \quad \text{si} \quad \mathbb{P}(\mathcal{G} = \mathcal{G}_k / \mathbf{x} = x) = \arg \max_{g \in \mathcal{G}} p_g(x). \tag{4.37}$$

4.5.1. Construction des jeux d'entraînement et test

Lors de la mise en place de modèles de classification supervisée, il est de coutume d'allouer 75% de l'échantillon de données à la construction d'un jeu d'entraînement servant à calibrer le modèle, et d'utiliser le reste sous forme de jeu test pour en évaluer les performances. De manière classique, cela se fait en procédant simplement à un tirage aléatoire de 75% des observations. Dans notre cas, l'effectif très faible de certains groupes (\mathcal{G}_1 et \mathcal{G}_8) rend une telle démarche imprudente. En effet, on risque de n'avoir aucune salve appartenant à ces groupes dans l'échantillon test.

Nous avons donc opté pour une stratégie visant à construire ces deux jeux de données en conservant autant que possible les proportions de chaque classe. Pour ce faire, nous affectons, de manière aléatoire, 1/4 de l'effectif de chaque groupe au jeu test. Si le total n'est pas divisible par 4, le reste de la division euclidienne est alors attribué au jeu d'entraînement (Tableau 4.9). Ce découpage permet d'avoir 127 salves (soit 76.5%) dans le jeu d'entraînement et 39 salves (soit 23.5%) dans le jeu test, sélectionnées au hasard. Dans la suite, nous nous appuyerons sur 10 jeux d'entraînement et test construits grâce à ce processus afin de calibrer et tester les modèles proposés.

Tableau 4.9. – Nombre de salves dans chaque jeu de données

	\mathcal{G}_1	\mathcal{G}_2	\mathcal{G}_3	\mathcal{G}_4	\mathcal{G}_5	\mathcal{G}_6	\mathcal{G}_7	\mathcal{G}_8
jeu d'entraînement	6	20	9	10	21	23	33	5
jeu test	1	6	3	3	7	7	11	1

4.5.2. Approche classique

L'état de l'art réalisé au chapitre 1 nous a permis de passer en revue les différentes méthodes de classification supervisée utilisées dans le cas classique. Il en ressort notamment que les méthodes les plus appréciées des auteurs sont la méthode des kNN et la méthode des réseaux de neurones.

Nous expérimentons ici six méthodes, aussi bien paramétriques que non paramétriques. Elles sont ensuite optimisées sur les jeux d'entraînement, puis leurs performances comparées sur les jeux tests, afin de dégager une méthode applicable par défaut aux données d'intérêt ici. Notons que les résultats de l'application de la méthode des réseaux de neurone s'étant avérée largement en-deçà de ceux des autres méthodes, nous avons donc choisi de ne pas les présenter ici. Nous expliquons cette contre performance notamment par la taille très limitée de l'échantillon de salves.

Par ailleurs, quelle que soit la méthode, nous partons d'une matrice de données de 20 prédicteurs obtenus à la suite de l'application de l'algorithme 1 avec un seuil à 0.999. Les variables duration, count2Fall et meanAmplitude ne seront pas prises en compte.

4.5.2.1. Régression logistique multinomiale

La régression logistique multinomiale est une méthode paramétrique reposant sur l'hypothèse qu'en choisissant une modalité (une classe) de référence $\mathcal{G}_{ref} \in \mathcal{G}$, il existe une relation linéaire entre les quantités $\ln \left(\frac{p_{\mathcal{G}_k(x)}}{p_{\mathcal{G}_{ref}(x)}} \right)$ avec ($k \neq ref$). Il s'agit d'estimer, pour chaque classe \mathcal{G}_k avec ($k \neq ref$), un vecteur de paramètres $\beta^k \in \mathbb{R}^{p+1}$ (avec p le nombre de paramètres EA), tel que :

$$LOGIT_{\mathcal{G}_k}(x_i) = \ln \left(\frac{p_{\mathcal{G}_k(x_i)}}{p_{\mathcal{G}_{ref}(x_i)}} \right) = \beta_0^k + \sum_{j=1}^p \beta_j^k x_i^j. \quad (4.38)$$

Sachant que $\sum_{k=1}^K p_{\mathcal{G}_k}(x_i) = 1$, on a alors :

$$p_{\mathcal{G}_{ref}}(x_i) = \frac{1}{1 + \sum_{\substack{k \in \mathcal{G} \\ k \neq ref}} e^{LOGIT_{\mathcal{G}_k}(x_i)}}, \quad (4.39)$$

et

$$p_{\mathcal{G}_{k,k \neq ref}}(x_i) = \frac{e^{LOGIT_{\mathcal{G}_k}(x_i)}}{1 + \sum_{\substack{k \in \mathcal{G} \\ k \neq ref}} e^{LOGIT_{\mathcal{G}_k}(x_i)}}. \quad (4.40)$$

Les paramètres β^k sont estimés par la méthode du maximum de la vraisemblance (algorithme de Newton-Raphson). L'expression (4.37) permet alors de prédire le groupe \mathcal{G}_k associé à chaque salve.

Afin d'optimiser les performances du modèle, nous avons procédé à des sélections de variables de type "pas à pas" (Forward, Backward et mixte) [163], par minimisation du critère AIC [164]. La sélection de type Backward garde uniquement les variables du Tableau 4.10.

Tableau 4.10. – Variables obtenues par la sélection de variables de type Backward. A noter que les sélections de variables de type Backward et mixte aboutissent aux mêmes variables. La méthode Forward conduit à des modèles de moins bonne qualité.

shapeRatioSpec	maxFreqSpec	riseTime	fallTime	skew	ecTy	statAD
----------------	-------------	----------	----------	------	------	--------

4.5.2.2. Analyse discriminante linéaire

L'analyse discriminante est une méthode paramétrique s'appuyant sur une hypothèse de normalité de la distribution du vecteur aléatoire \mathbf{x} des paramètres EA. La fonction de densité f_g de la loi de \mathbf{x} conditionnellement à \mathcal{G} s'exprime alors par :

$$f_g(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} \sqrt{\det \Sigma_g}} \exp \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_g)^T \Sigma_g^{-1} (\mathbf{x} - \boldsymbol{\mu}_g) \right], \quad (4.41)$$

où Σ_g est la matrice de variance-covariance du groupe g et $\boldsymbol{\mu}_g$ son centre de gravité.

Si on suppose que les matrices de variance-covariance des groupes sont égales,

$$\Sigma_{\mathcal{G}_1} = \dots = \Sigma_{\mathcal{G}_K} = \mathbf{W}, \quad (4.42)$$

où \mathbf{W} est la matrice de variance-covariance intra-groupe globale, on parle d'analyse discriminante linéaire. C'est l'approche que nous adopterons ici.

L'expression de la probabilité *a posteriori* $p_g(x)$ vérifie :

$$p_g(x) = \frac{\pi_g f_g(x)}{\sum_{k=1}^K \pi_k f_k(x)}, \quad (4.43)$$

où $\pi_g = \mathbb{P}(\mathcal{G} = g)$ est la probabilité *a priori* d'appartenance au groupe $g \in \mathcal{G}$ et K le nombre total de groupes.

Moyennant l'hypothèse de normalité, maximiser $p_g(x)$ revient à maximiser :

$$\ln(\pi_g) + \boldsymbol{\mu}_g \mathbf{W}^{-1} \mathbf{x}^T - \frac{1}{2} \boldsymbol{\mu}_g \mathbf{W}^{-1} \boldsymbol{\mu}_g^T, \quad (4.44)$$

l'estimation $\hat{p}_g(x)$ de $p_g(x)$ étant obtenue en estimant μ_g et Σ par maximisation de la vraisemblance. Les estimateurs obtenus correspondent alors à ceux présentés dans l'Annexe D.3.1. Quant à π_g , elle est estimée par $\frac{N_g}{N}$.

En revenant à des considérations plus géométriques, il s'agit d'attribuer chaque observation x au groupe pour lequel la quantité $(x - \mu_g)^T \Sigma^{-1} (x - \mu_g)$ est minimale. On minimise donc les distances par rapport aux centres de gravité par la métrique de Mahalanobis (modulo la probabilité *a priori* π_g) [161, Chap.4] :

$$\hat{G}(x) = \arg \min_{g \in \mathcal{G}} \left[d_W^2(x, g) - 2 \ln \pi_g \right] = \arg \min_{g \in \mathcal{G}} \left((x - \mu_g)^T \mathbf{W}^{-1} (x - \mu_g) - 2 \ln \pi_g \right). \quad (4.45)$$

Notons que la méthode est assez robuste par rapport à l'hypothèse de normalité de la distribution de x . Cependant, elle est gênée par la violation de l'hypothèse d'homoscédasticité (4.42) et sensible à une forte dimensionnalité et/ou une corrélation des variables explicatives. Il est donc souhaitable de procéder à une réduction de dimension de la matrice de données avant d'appliquer cette méthode. Dans notre cas, nous avons commencé par effectuer des sélections de variables de type pas à pas (Tableau 4.11).

Tableau 4.11. – Variables obtenues par la sélection de variables de type Forward. A noter que la sélection Mixte permet d'obtenir les mêmes résultats que la Forward, et que la Backward aboutit à de moins bons résultats.

specCen	shapeRatioSpec	maxFreqSpec	power	amplitude	count2Rise
---------	----------------	-------------	-------	-----------	------------

4.5.2.3. Les méthodes du type plus-proches-voisins

Les méthodes du type plus-proches-voisins dont la plus célèbre est la méthode des k plus-proches-voisins, ou «**k-nearest neighbours**» en anglais, sont des méthodes relativement simples, non paramétriques et non structurées, réputées parmi les méthodes de classification supervisée les plus efficaces [161, Chap.13], y compris pour les problématiques en EA (cf. Section 1.2.2.2 du chapitre 1). Il s'agit d'affecter un individu à la classe de son voisin le plus proche (en général, au sens la distance euclidienne sur les variables centrées réduites), ou à effectuer un vote avec celles de ses voisins $k > 1$ [163, 165]. Notons $p_{g,k}$ la fonction qui à une salve x affecte un score d'appartenance à g en tenant compte de ses k plus-proches-voisins. Dans le cas le plus simple, pour un nombre de plus-proches-voisins $k > 1$, $p_{g,k}(x)$ est estimée par :

$$\hat{p}_{g,k}(x) = \frac{k_g(x)}{k}, \quad g \in \mathcal{G}, \quad (4.46)$$

où $k_g(x)$ est le nombre de salves appartenant à la classe g parmi les k plus-proches-voisins de x .

L'optimisation de cette approche consiste à choisir le nombre k de plus-proches-voisins de manière à minimiser l'erreur de prédiction. Pour ce faire, nous avons utilisé la méthode de validation croisée de type «leave-one-out» (cf. Annexe D.3.2), ce qui conduit à $k = 3$ (Figure 4.24).

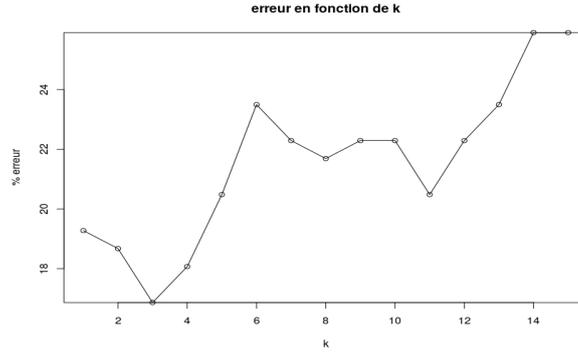


Figure 4.24. – Taux d'erreur en fonction du nombre k de plus-proches-voisins

4.5.2.4. Arbre de classification (CART)

Le modèle de classification par arbre, **Classification and Regression Tree** (CART) en anglais, est une méthode non paramétrique, dont on doit l'essor aux travaux de BREIMAN *et al.* [166]. Elle permet d'estimer $p_g(x)$ à travers un arbre obtenu par partitionnement récursif et dyadique de l'espace des paramètres EA. La classe $\hat{G}(x)$ d'une nouvelle salve x est donnée par :

$$\hat{G}(x) = \mathcal{G}_{F(t)} \quad \text{si } x \in R_t, \quad (4.47)$$

où R_t désigne l'ensemble des données salves appartenant au nœud t .

La fonction $F : \{1, \dots, M\} \rightarrow \{1, \dots, K\}$ permet d'associer chaque nœud de l'arbre à une classe,

$$F(m) = \arg \max_{k \in \{1, \dots, K\}} \hat{p}_{t,k}, \quad (4.48)$$

où $\hat{p}_{t,k}$ est la fréquence de la classe k dans le nœud t

$$\hat{p}_{t,k} = \frac{1}{\text{Card}(R_t)} \sum_{i \in R_t} \mathbb{1}_{G(x_i)=\mathcal{G}_k}. \quad (4.49)$$

L'expression (4.47) résume la structure de l'arbre de classification, ainsi que les principaux enjeux relatifs à sa mise en œuvre.

R_1 est appelée racine de l'arbre et contient toutes les salves de l'échantillon. A chaque nœud de l'arbre, nous disposons donc d'un vecteur $\hat{p}_t = (\hat{p}_{t,1}, \dots, \hat{p}_{t,K})$ qui devrait permettre de séparer au mieux les classes. Il faut donc en mesurer la qualité. En général, cette opération appelée mesure de l'impureté repose sur la minimisation de l'indice de Gini I_G , dont l'objectif est de mesurer le taux d'erreur si la classe associée au nœud R_t était choisie de manière aléatoire en fonction de \hat{p}_t . On a :

$$I_G(\hat{p}_t) = \sum_{k=1}^K \hat{p}_{t,k}(1 - \hat{p}_{t,k}). \quad (4.50)$$

La question pratique qui se pose à présent est la définition d'une stratégie de partitionnement optimale à partir de la racine de l'arbre. Pour ce faire, l'une des méthodes les plus utilisées est l'algorithme glouton (greedy algorithm) [167]. La première étape de cet algorithme consiste à partitionner l'échantillon de salves en deux groupes, donnant naissance à un nœud (R_2) fils gauche et un nœud (R_3) fils droit. Pour un paramètre EA x^j et un seuil $s \in \mathbb{R}$, le partitionnement s'écrit :

$$\begin{aligned} R_2(j, s) &= \{x = (x^1, \dots, x^p)^T \in \mathbb{R}^p \text{ tel que } x^j \leq s\} \\ R_3(j, s) &= \{x = (x^1, \dots, x^p)^T \in \mathbb{R}^p \text{ tel que } x^j > s\} \end{aligned}$$

Le couple optimal (j, s) correspond à celui minimisant la somme des indices de Gini (normalisés) associés à $R_2(j, s)$ et $R_3(j, s)$.

Cette opération est réitérée en partitionnant les nouveaux nœuds obtenus à chaque étape, jusqu'à l'atteinte d'un critère d'arrêt. Pour un nœud donné, la croissance de l'arbre s'arrête lorsqu'il est parfaitement homogène ou lorsqu'il n'existe plus de partition admissible ou, pour éviter un découpage inutilement fin, si le nombre d'observations qu'il contient est inférieur à une valeur à choisir en général entre 1 et 5. Ce nœud prend alors le nom de nœud final ou feuille.

La Figure 4.25 en donne un exemple, avec la construction de l'arbre CART sur les données considérées dans le travail présenté ici. Cet arbre présente un problème de surapprentissage, notamment à cause de l'absence d'élagage. On constate également que \mathcal{G}_4 n'est jamais représentée par une feuille. Nous nous tournons donc vers des méthodes plus adaptées à la prédiction en multipliant les arbres CART.

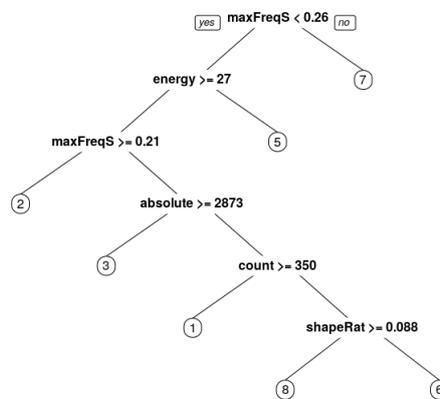


Figure 4.25. – Arbre de décision du jeu de données

4.5.2.5. Bagging d'arbres de classification

La méthode du "bootstrap aggregating" (bagging) est une méthode d'agrégation visant à améliorer les performances d'un modèle de prédiction en s'appuyant sur une série d'échantillons de même taille obtenus (par bootstrap) à partir de l'échantillon de départ [168]. Cela permet notamment de stabiliser le modèle à travers une réduction de variance. Dans le cas du bagging d'arbres de classification, si on note par B le nombre d'échantillons obtenus par bootstrap, il s'agit de construire B arbres de classification et de prédire ensuite la classe d'appartenance d'une nouvelle salve en procédant à un vote à la majorité.

Il existe deux méthodes populaires pour construire les échantillons bootstrap dans ce cadre. La première consiste à tirer avec remise N individus, puis à recommencer B fois. Dans le second cas, on tire sans remise $l < N$ individus, B fois. Pour ce dernier cas, il est de coutume de tirer 63.20% des individus pour déterminer l , ce qui correspond à la moyenne d'individus différents tirés par arbre avec la première méthode. Le Tableau 4.12 indique la moyenne, le minimum et l'écart-type des pourcentages de bonnes prédictions sur les 10 jeux tests en utilisant ces deux approches. Les résultats issus de la méthode 2 étant meilleurs, nous gardons donc cette approche.

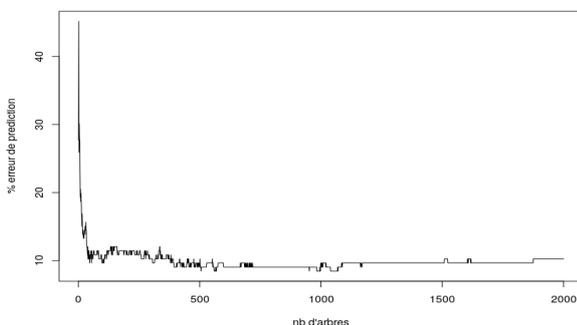
Tableau 4.12. – Moyenne, minimum et écart-type des pourcentages de bonnes prédictions des 2 approches du modèle bagging

Méthode	Moyenne	Minimum	Écart-type
1	82.56	71.79	6
2	86.41	79.49	6

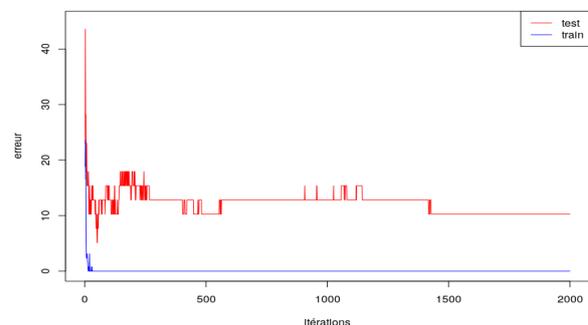
4.5.2.6. Boosting d'arbres de classification

Le boosting d'arbres de classification est une méthode qui, comme le bagging, consiste à s'appuyer sur un grand nombre d'arbres pour améliorer les performances de la méthode CART classique. Dans le cas du boosting, on s'appuie sur une approche itérative dont chaque étape consiste à construire un nouvel arbre à partir du précédent, tout en apprenant des limites de celui-ci. Cet apprentissage est réalisé par réajustement (augmentation) du poids des individus mal classés lors de la construction du nouveau [169, 170]. Le modèle final est obtenu par combinaison linéaire de tous les arbres construits. Cela permet notamment d'éviter les problèmes de sur-apprentissage.

Dans notre cas, en utilisant la méthode de validation croisée de type leave-one-out (Figure 4.26a) pour identifier les individus mal classés, nous constatons qu'il y a une certaine stabilisation autour de 800 arbres. Mais puisqu'il y a stabilisation, nous avons intérêt à prendre beaucoup d'arbres pour avoir plus de votes et ne pas avoir de résultats trop différents à chaque fois que nous relançons le processus. Choisissons donc 2000 arbres. La stabilisation peut paraître étonnante mais elle s'explique en réalité par le fait que l'erreur arrive très vite à 0 sur les individus concernés par la création du modèle (Figure 4.26b). L'algorithme n'apprend donc plus à chaque itération.



(a) Erreur en fonction du nombre d'arbres



(b) Erreur en fonction du nombre d'arbres sur un jeu d'entraînement et un jeu test

Figure 4.26. – Courbes d'évolution de l'erreur de prédiction en fonction du nombre d'arbres.

4.5.2.7. Forêts aléatoires

La méthode des forêts aléatoires peut être comprise comme une généralisation de la méthode du bagging d'arbre de classification (cf. Page 87) consistant à rajouter une couche d'aléa dans la construction des arbres. Pour chaque arbre, à chaque nœud, la sélection de la variable de coupure (cf. Page 86) ne se fait plus sur l'ensemble des variables, mais sur un sous-ensemble correspondant à un nombre fixé de variables tirées au hasard. Le choix de ce nombre constitue alors un paramètre supplémentaire à optimiser.

Les arbres étant construits à partir d'échantillons obtenus par bootstrap, ils n'utilisent pas toutes les observations. Pour une observation $(x_i, G(x_i))$ de l'échantillon d'apprentissage, notons par I_B l'en-

semble des arbres de la forêt qui ne contiennent pas $(x_i, G(x_i))$ dans leur échantillon bootstrap. On dit que l'observation est Out Of Bag (OOB) pour ces arbres. La prévision $\hat{G}(x_i)$ de la forêt sur $G(x_i)$ s'obtient alors en faisant voter l'ensemble des arbres pour lesquels il est OOB. L'erreur de prédiction qui s'écrit alors :

$$\frac{1}{N} \sum_{i=1}^N \mathbb{1}_{\hat{G}(x_i) \neq G(x_i)} \quad (4.51)$$

La méthode des forêts aléatoires utilise cette procédure OOB pour optimiser la valeur du nombre de variables à tirer au hasard à chaque nœud, mais également pour faire une sélection de variables. Comme dans le cas d'une validation croisée, la valeur optimale du nombre de variables est celle qui minimise l'erreur de prédiction. Quant à la procédure de sélection de variables, elle est un peu plus complexe.

Notons par OOB_k l'échantillon OOB du k -ième arbre, c'est-à-dire l'ensemble des observations qui ne figurent pas dans le k -ième échantillon bootstrap. Notons par E_{OOB_k} l'erreur de prédiction du k -ième arbre sur cet échantillon s'écrit :

$$E_{OOB_k} = \frac{1}{\text{card}(OOB_k)} \sum_{i \in OOB_k} \mathbb{1}_{\hat{G}(x_i) \neq G(x_i)} \quad (4.52)$$

Notons par OOB_k^j l'échantillon OOB_k dans lequel les valeurs de la j -ème variable ont été permutées aléatoirement et par $E_{OOB_k^j}$ l'erreur de prédiction de l'arbre associé. Plus la variable j est prépondérante, plus cette permutation dégrade la capacité prédictive de l'arbre, ce qui se traduit par une augmentation de la différence $E_{OOB_k^j} - E_{OOB_k}$. L'importance de la variable j sur la forêt entière est alors mesurée en moyennant ces différences d'erreur sur tous les arbres :

$$\text{Imp}(x^j) = \frac{1}{B} \sum_{i=1}^B (E_{OOB_k^j} - E_{OOB_k}) \quad (4.53)$$

où B est le nombre d'arbres construits.

Les résultats de l'application de la méthode des forêts aléatoires, selon la méthode de bootstrap choisie, sont présentés dans les Tableaux 4.13 et 4.14.

Tableau 4.13. – Variables conservées

specCen	shapeRatioSpec	maxFreqSpec	riseTime	amplitude	count	energy	ecTy
---------	----------------	-------------	----------	-----------	-------	--------	------

Tableau 4.14. – Moyenne, minimum et écart-type des pourcentages de bonnes prédictions suivant le palier de la sélection de variables avec la méthode 1 de bootstrap. $mtry$ est le nombre de variables tirées à chaque nœud.

Méthode de Bootstrap	moyenne	minimum	écart-type	$mtry$ optimal
1	88.72	79.49	4.9	2
2	88.97	82.05	4.4	2

4.5.2.8. Choix du modèle final

Le Tableau 4.15 indique la moyenne, le minimum et l'écart-type des pourcentages de bonnes prédictions de chaque modèle sur les 10 jeux de données, ainsi que le nombre de variables utilisées. Nous remarquons que le modèle forêts aléatoires et le boosting ont la meilleure moyenne ainsi que

le meilleur minimum. Notons que les forêts aléatoires font partie des modèles conduisant aux plus faibles variations. Pour effectuer un choix entre les deux modèles, nous nous appuyons sur le fait que les forêts aléatoires utilisent un faible nombre de variables, ce qui permet de gagner du temps de calcul dans le cas de la prise en compte d'un nouveau signal. C'est donc le modèle que nous suggérons d'utiliser par défaut dans le cas de l'approche classique.

Tableau 4.15. – Comparaison des modèles par moyenne, minimum et écart-type des pourcentages de bonnes prédictions

Modèle optimisé	% bonnes prédictions			nombre de variables utilisées
	moyenne	minimum	écart-type	
Régression logistique multinomiale	83.33	71.79	6.3	7
Analyse discriminante linéaire	87.18	82.05	3.6	6
k plus-proches-voisins (knn)	81.54	74.36	5.6	20
Bagging	86.41	79.49	6	20
Boosting	88.97	82.05	5	20
Forêts aléatoires	88.97	82.05	4.4	8

4.5.3. Approche fonctionnelle

Nous avons choisi d'expérimenter deux approches de classification supervisée basées sur l'approche fonctionnelle. La première, exclusivement non-paramétrique, s'appuie sur des (semi-)métriques FERRATY & VIEU [157, 171]. Quant à la seconde, elle est exclusivement basée sur la réduction de dimension, permettant de se ramener à une matrice de données ouvrant la voie à l'expérimentation de toutes les méthodes présentées dans le cas classique.

Notons qu'une troisième stratégie serait d'utiliser de manière complémentaire les approches non paramétriques basées sur les (semi-)métriques et les approches de réduction de dimension. Cela consiste à considérer un modèle avec des variables fonctionnelles et classiques [172]. Cependant, il semblerait que ce type de modèle nécessite «toujours» une régularisation consistant à procéder à une réduction de dimension (par représentation dans une base particulière, ACP fonctionnelle...) des variables fonctionnelles (voir les codes sources de FEBRERO-BANDE, FUENTE *et al.* [173]). Nous ne la retiendrons pas ici.

4.5.3.1. Stratégie exclusivement non paramétrique

Dans le cas fonctionnel, la probabilité *a posteriori* d'appartenance d'une courbe χ au groupe g (4.36) s'écrit :

$$p_g(\chi) = \mathbb{P}(\mathcal{G} = g / \{\mathcal{X}(t) = \chi(t), t \in \mathcal{T}\}), \quad g \in \mathcal{G}, \quad (4.54)$$

où $\{\chi(t), t \in \mathcal{T}\}$ est une réalisation de la variable aléatoire fonctionnelle \mathcal{X} .

La méthode d'estimation de $p_g(x)$ la plus utilisée dans la littérature est celle proposée par FERRATY & VIEU [171, 174] utilisant des estimateurs à noyau. On a :

$$\hat{p}_{g,h}(\chi) = \frac{\sum_{i, y_i=g}^N K(h^{-1}d(\chi, \chi_i))}{\sum_{i=1}^N K(h^{-1}d(\chi, \chi_i))}, \quad (4.55)$$

où K est le noyau asymétrique et $h \in \mathbb{R}_+$ est le bandwidth (ou paramètre de lissage) [157, Chap.4].

L'idée intuitive derrière l'expression (4.55) est d'estimer $p_{g,h}(\chi)$ en utilisant uniquement les courbes appartenant à la classe g , tout en s'assurant que ces courbes ne sont pas trop différentes de χ grâce à un choix judicieux de h tel que $d(\chi, \chi_i) \leq h$, ce qui, philosophiquement, revient à utiliser une méthode du type kNN. FERRATY & VIEU [157, Chap.8] proposent d'ailleurs une réadaptation de l'estimateur (4.55) comme suit :

$$\hat{p}_{g,k}(\chi) = \frac{\sum_{i, y_i=g}^N K(h_k^{-1}d(\chi, \chi_i))}{\sum_{i=1}^N K(h_k^{-1}d(\chi, \chi_i))}, \quad (4.56)$$

où $h_k \in \mathbb{R}_+$ est un paramètre associé au nombre de plus-proches-voisins k par :

$$\text{card}\{i : d(\chi, \chi_i) < h_k\} = k.$$

En pratique, si le choix du type de noyau n'a en général pas d'incidence significative sur les performances du modèle, ceux du nombre k de plus-proches-voisins et de la semi-métrique d sont cruciaux. Pour l'optimisation de la valeur de k , comme dans le cas de la méthode des kNN classique, on peut avoir recours à la validation croisée.

Les résultats d'application de cette méthode fonctionnelle des kNN sont présentés dans le Tableau 4.16. Les résultats des méthodes basées uniquement sur les spectres et de celle utilisant la semi-métrique mixte (4.31) étant sensiblement identiques, on conclut que pour cette approche, l'apport des courbes enveloppes est très limitée.

Tableau 4.16. – Cas de l'approche fonctionnelle non paramétrique : comparaison des modèles par moyenne, minimum et écart-type des pourcentages de bonnes prédictions. La métrique utilisée est la métrique \mathcal{L}_2

Modèle optimisé	% bonnes prédictions			nombre de variables utilisées
	moyenne	minimum	écart-type	
enveloppes	52.56	43.6	6.3	1
spectres	75.13	66.68	4.1	1
mixte	73.33	71.79	6.3	2

4.5.3.2. Stratégie exclusivement basée sur la réduction de dimension

Comme souligné précédemment, dans le cas d'une stratégie exclusivement basée sur la réduction de dimension, puisqu'on se ramène à une matrice de données classique, toutes les méthodes présentées dans le cas classique deviennent alors applicables.

Les résultats présentés dans le Tableau 4.17 sont obtenus à partir d'une matrice de données de 16 variables, dont 10 correspondent aux scores associés aux 10 premières composantes principales d'une ACP fonctionnelle sur les enveloppes, les 6 autres correspondant aux 6 premières composantes d'une AFC sur les spectres.

Tableau 4.17. – Cas de l'approche fonctionnelle par réduction de dimension : comparaison des modèles par moyenne, minimum et écart-type des pourcentages de bonnes prédictions

Modèle optimisé	% bonnes prédictions			nombre de variables utilisées
	moyenne	minimum	écart-type	
LDA	73.59	64.10	4.3	5
kNN	59.23	46.15	7.6	16
Bagging	78.97	71.79	3.6	16
Boosting	78.72	71.79	4.1	16
Forêts aléatoires	82.82	79.49	2.7	4

4.5.3.3. Choix du modèle final

La comparaison des résultats présentés dans les Tableaux 4.16 et 4.17 permet de conclure que dans le cadre de l'approche fonctionnelle également, la méthode des forêts aléatoires est celle qu'il faudrait appliquer par défaut. D'autre part, les résultats de la sélection de variables réalisée lors de l'optimisation de cette méthode révèle que sur les quatre variables utilisées par le modèle final, une seule est liée à l'enveloppe des salves. Cela rejoint les conclusions de la méthode des kNN fonctionnelle.

4.5.4. Comparaison des résultats des approches classique et fonctionnelle et prise en charge de salves inconnues

La méthode des forêts aléatoires apparaît donc comme celle à appliquer par défaut quelle que soit l'approche choisie. Par ailleurs, la conclusion immédiate d'une comparaison des deux approches est que de manière globale, pour les données considérées ici, l'approche fonctionnelle est moins efficace que l'approche classique. Cependant, d'un point de vue statistique, la capacité prédictive relativement élevée (plus de 80%) des deux modèles indique que les deux approches conviennent pour le type de données EA considérées dans nos études.

Dans la suite de cette dernière partie du chapitre, nous proposons de comparer les performances des approches selon la classe d'intérêt, d'analyser leur robustesse à un changement de la stratégie de prétraitement des données et enfin, d'aborder la question de la prise en charge d'une salve inconnue.

4.5.4.1. Comparaison des performances des méthodes classe par classe

Les Tableaux 4.18 et 4.19 résument les capacités prédictives des modèles optimaux, selon l'approche choisie, à travers les indicateurs *rappel* et *précision* [175] :

- le *rappel* indique la capacité du modèle à retrouver la classe. Nous pouvons le formuler avec la phrase suivante : Un événement de la classe \mathcal{G}_k est bien prédit dans *rappel*% des cas.
- la *précision* indique la capacité du modèle à prédire la classe avec justesse. Nous pouvons la formuler avec la phrase suivante : Un événement prédit en classe \mathcal{G}_k est bien classé dans *précision*% des cas.

Au-delà de la conclusion prévisible de prédominance de l'approche classique, on constate que les performances des modèles sont variables selon la classe d'appartenance de la salve. On remarque par exemple que leurs capacités prédictives sont relativement médiocres dans le cas particulier de la classe \mathcal{G}_8 . Cela s'explique notamment par l'effectif relativement faible des mécanismes appartenant à cette classe.

Une conséquence pratique de ce dernier résultat est qu'il faudrait prévoir des «**warnings**» attirant l'attention de l'expérimentateur sur les limites de l'algorithme pour les classes à faible effectif.

Tableau 4.18. – *Rappels des modèles*

	\mathcal{G}_1	\mathcal{G}_2	\mathcal{G}_3	\mathcal{G}_4	\mathcal{G}_5	\mathcal{G}_6	\mathcal{G}_7	\mathcal{G}_8
F. Al. Classique	50.00	100.00	80.00	73.33	91.43	84.29	100.00	30.00
F. Al. Fonctionnelle	10.00	100.00	63.33	90	84.29	74.29	91.82	20.00

Tableau 4.19. – *Précisions des modèles*

	\mathcal{G}_1	\mathcal{G}_2	\mathcal{G}_3	\mathcal{G}_4	\mathcal{G}_5	\mathcal{G}_6	\mathcal{G}_7	\mathcal{G}_8
F. Al. Classique	100.00	98.36	92.31	100.00	86.49	83.10	92.44	25.00
F. Al. Fonctionnelle	25.00	100.00	86.36	90.00	72.84	72.22	87.07	40.00

a. Robustesse des modèles à un changement de prétraitement ou de jeu de données

Une question pratique pertinente à se poser sur les modèles proposés ici est : les algorithmes resteraient-ils toujours performants si on change le jeu de données ou lorsque l'expérimentateur ne procède à aucun prétraitement des signaux ?

Pour répondre à cette interrogation, considérons les Tableaux 4.20 et 4.21 présentant les performances des modèles de forêts aléatoires classique et fonctionnel en cas de non prétraitement des salves et/ou de changement de capteur. En ne perdant pas de vue que les modèles sont calibrés sur les données du capteur M1, auxquels on a appliqué la stratégie de prétraitement par défaut proposée en début de chapitre, la première conclusion qui se dégage est que les modèles sont sensibles au type de prétraitement subi par les signaux et à un changement de capteur.

Dans le cas classique (Tableau 4.20), lorsqu'on garde exactement la même procédure de prétraitement des salves et qu'on applique l'algorithme aux données du capteur M2, puisque la capacité prédictive du modèle baisse de plus de 15%, on conclut que le biais résiduel laissé par la stratégie d'estimation du signal source est non négligeable. Nous expliquons cela, d'une part, par l'impossibilité de disposer d'une fonction de transfert permettant une prise en charge plus efficace de l'impact du dispositif (cf. conclusions du chapitre 2), et d'autre part, par le fait que pour certaines salves, l'atténuation du signal n'est pas la même d'un capteur à l'autre. Cependant, les performances de l'algorithme restent relativement correctes (plus de 70% de bonnes prédictions), on peut donc soutenir l'hypothèse que la stratégie de prétraitement proposée est tout à fait pertinente. De plus, on constate qu'il y a une forte dégradation des performances du modèle sans prétraitement (seulement 57% de bonnes prédictions).

En ce qui concerne l'approche fonctionnelle, au vu des conclusions du chapitre 3, les résultats présentés dans le Tableau 4.21 sont surprenants et incohérents. En effet, nous avons conclu que de manière globale, la forme d'onde (donc l'enveloppe) et le spectre étaient relativement bien recouverts à la suite du débruitage. On s'attendrait donc à ce que les performances du modèle basé sur l'approche fonctionnelle restent plutôt stables quel que soit le capteur. Le résultat tout à fait révélateur du Tableau 4.22 prouve que cette contre performance des approches fonctionnelles est essentiellement due au fait que le modèle de forêts aléatoires utilisé s'appuie sur des variables latentes.

Tableau 4.20. – Performance du modèle classique de forêts aléatoires sur les données des deux capteurs et ayant subi ou non un débruitage. Le modèle a été construit via les données du capteur M1 débruitées par soustraction spectrale (SS). Le capteur M2 est l'autre capteur du réacteur expérimental.

Capteur	Prétraitement	% de bien classés
M2	Oui	70
M1	aucun	70
M2	aucun	57

Tableau 4.21. – Performance du modèle fonctionnel de forêts aléatoires sur les données des deux capteurs et ayant subi ou non un débruitage. Le modèle a été construit via les données du capteur M1 débruitées par soustraction spectrale (SS). Le capteur M2 est l'autre capteur du réacteur expérimental.

Capteur	Prétraitement	% de bien classés
M2	Oui	9
M1	aucun	26
M2	aucun	27

Tableau 4.22. – Performance du modèle kNN fonctionnel sur les données des deux capteurs et ayant subi ou non un débruitage. Le modèle a été construit via les données du capteur M1 débruitées par soustraction spectrale (SS). Le capteur M2 est l'autre capteur du réacteur expérimental.

Capteur	Prétraitement	% de bien classés
M2	Oui	66
M1	aucun	69
M2	aucun	65

4.5.4.2. Prise en charge de salves inconnues

Au-delà de la reconnaissance de mécanismes déjà observés lors des essais précédents et appartenant donc aux classes \mathcal{G}_1 à \mathcal{G}_8 , une dernière problématique est de définir une stratégie permettant de décider si une salve est trop différente, et donc d'alerter les expérimentateurs quant à l'avènement d'un nouveau type de mécanisme.

Pour ce faire, nous avons choisi, en plus du résultat d'exécution de l'algorithme qui permet d'attribuer chaque nouvelle observation x à l'une des classes \mathcal{G}_1 à \mathcal{G}_8 , de réaliser un test visant à "calculer" la distance entre x et toutes les classes existantes. Pour un seuil fixé s , deux cas de figure peuvent alors se présenter. S'il existe $k^* \in \bar{\mathcal{G}}$ tel que :

$$d(x, \mathcal{G}_{k^*}) = \min_{k \in \bar{\mathcal{G}}} d(x, \mathcal{G}_k) < s ;$$

la salve est considérée comme assez proche des mécanismes déjà observés pour qu'aucune alerte ne soit donnée. Dans le cas contraire, l'expérimentateur est alerté et il lui revient d'attribuer x à la classe prédite par l'algorithme ou de créer une nouvelle classe, ce qui consiste à décider qu'un mécanisme inconnu vient de survenir.

La méthode d'analyse discriminante ayant conduit à des résultats probants (cf. Section 4.5.2.2 et Tableau 4.15), nous avons choisi d'utiliser la distance de Mahalanobis comme métrique :

$$d(x, \mathcal{G}_k) = (x - \mu_k)^T W^{-1} (x - \mu_k) .$$

μ_k est le centre de gravité de la classe \mathcal{G}_k et W la matrice d'inertie intra-classe.

a. Optimisation de l'algorithme de test

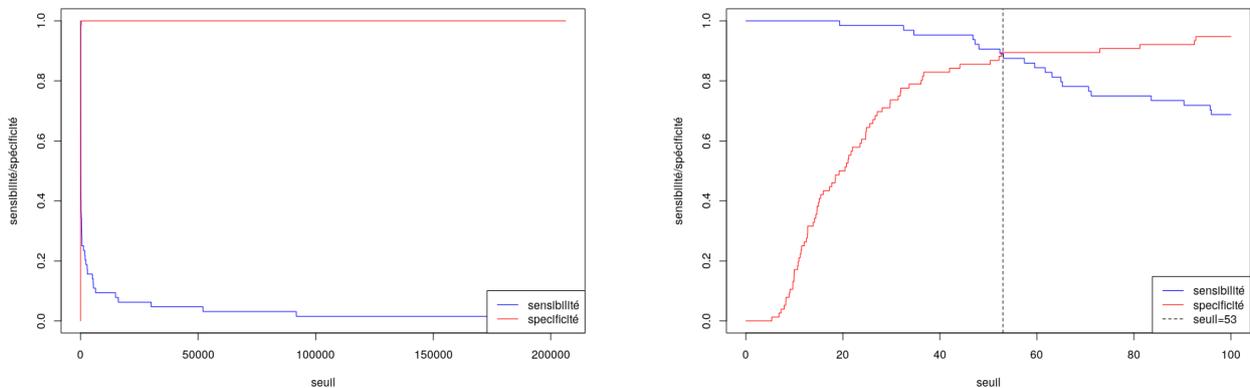
Optimiser les performances de l'algorithme revient à trouver un seuil s permettant d'alerter l'expérimentateur uniquement lorsque cela est nécessaire, c'est-à-dire à l'avènement effectif d'un nouveau type de mécanisme. Pour ce faire, nous nous appuyerons sur un compromis entre la sensibilité et la spécificité de l'algorithme de test :

- La sensibilité est une fonction décroissante du seuil correspondant à la probabilité d'alerter "à raison" l'expérimentateur sur l'avènement d'un nouveau mécanisme. Lorsque le seuil est nul, quelle que soit la nature du mécanisme survenant, l'expérimentateur est alerté, la sensibilité vaut alors 1.
- La spécificité est une fonction croissante du seuil traduisant la probabilité de ne pas alerter l'expérimentateur "à raison" lorsque le mécanisme appartient déjà à l'une des classes existantes. Pour un seuil nul, puisqu'il y a alerte dans tous les cas, la spécificité est alors également nulle.

Afin de valider l'algorithme et donc définir un seuil optimal, nous avons utilisé un échantillon de 206 salves constitué de salves simulés (cf. expression (3.2), chapitre 3), de salves acquises dans des conditions autres que celle d'un RIA [71] et des salves ayant subi un prétraitement autre que celui par défaut proposé en introduction de ce chapitre. Disposant d'une variable réponse (alerte ou pas alerte) pour chacune de ces salves, nous avons constitué un échantillon d'entraînement de 140 salves et un échantillon test de 66 salves. Une maximisation du compromis sensibilité/spécificité (Figure 4.27) conduit à utiliser uniquement les variables présentées dans le Tableau 4.23 et à seuil optimal de 53.

Tableau 4.23. – Variables utilisées pour construire la distance de Mahalanobis

specCen	shapeRatioSpec	maxSpec	maxFreqSpec	riseTime	power
amplitude	count2Rise	count	meanFrequency	energy	skew
kurt	ecTy	riseFreq	fallFreq	shapeRatio	coefVar



(a) Sensibilité/spécificité en fonction du seuil

(b) Zoom de la Figure sur le seuil allant de 0 à 100

Figure 4.27. – Évolution de la spécificité et de la sensibilité en fonction du seuil.

b. Performances de l'algorithme de test

En considérant les 166 salves utilisées lors de la classification supervisée (cf. Section 4.5), seules 4 conduisent à une fausse alerte de l'algorithme, ce qui est très encourageant (Figure 4.28). Notons que parmi ces 4 salves, les sales 8 et 9 qui avaient été identifiées comme valeurs extrêmes lors de l'analyse exploratoire font partie des erreurs (Figures 4.13 à 4.16). Par ailleurs, l'analyse des performances de l'algorithme avec le seuil optimal de 53 sur le jeu test (Tableau 4.24) conduit à une sensibilité de 86.67%, une spécificité de 86.11% et une performance globale de 86.36%.

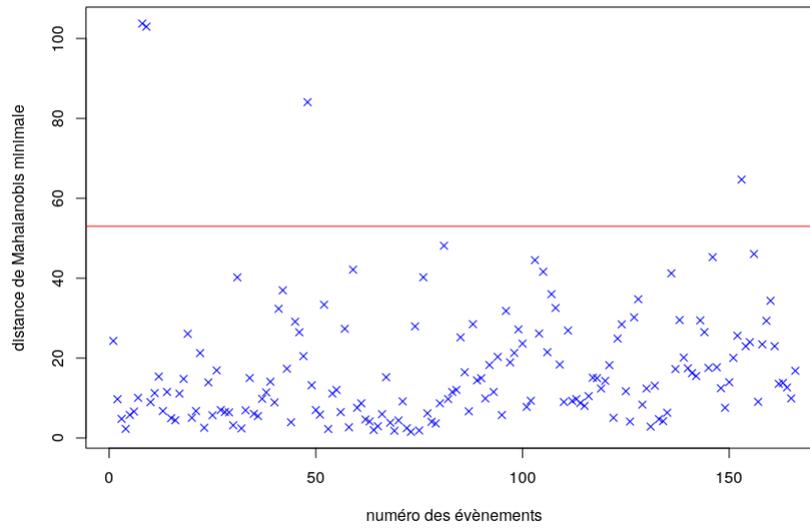


Figure 4.28. – Minimum des distances aux centres de gravité des classes pour les événements qui ont construit le modèle

Tableau 4.24. – Table de confusion de l'algorithme avec le seuil valant 53

Observations \ Prédictions	accepté	rejeté
	accepté	31
rejeté	4	26

4.6. Conclusion

Ce chapitre peut être considéré comme celui dans lequel nous apportons des réponses concrètes aux problématiques posées en introduction de ce travail, les précédents chapitres étant plutôt consacrés à des travaux préliminaires. Dans un premier temps, nous avons proposé une démarche de prétraitement applicable par défaut aux signaux essais. Il s'agissait d'illustrer de quelle manière les travaux présentés dans les chapitres 2 et 3 pouvaient être exploités dans le cadre d'une démarche visant à estimer le signal source. Ainsi, une démarche par défaut permettant de traiter le bruit environnemental, d'atténuer l'impact du dispositif d'essai, de détecter les salves et de pénaliser l'impact des capteurs a été proposée.

Malgré les précautions prises dans la mise en œuvre de cette démarche, nous avons mis en exergue diverses limitations intrinsèques à tout processus de détection des salves basé sur des paramètres glo-

baux. Une série de traitements applicables par défaut à toutes les salves a alors été proposée. Celle-ci permet notamment de réajuster la forme d'onde des salves de manière à éviter, autant que possible, de biaiser le calcul des paramètres EA.

Une fois les salves détectées et pré-traitées, afin de mettre en place un algorithme de classification supervisée permettant d'associer chaque salve à un mécanisme physique, nous avons expérimenté deux approches statistiques pour lesquelles la démarche standard constituée des phases d'analyse exploratoire, de clustering et de classification supervisée a été conduite. La première approche, classiquement utilisée en EA, s'appuie sur une matrice de données construite à partir de paramètres EA calculés sur le signal temporel, son spectre ou sur d'autres transformations. Quant à la seconde, dite approche fonctionnelle, elle utilise des variables aléatoires fonctionnelles (des courbes). Dans notre cas, nous avons choisi de travailler sur l'enveloppe et le spectre associés aux salves. Cette seconde approche nécessite un travail de prétraitement supplémentaire des salves visant à reconstruire la forme fonctionnelle des courbes à partir des observations discrètes et à gérer les différences de support et/ou les décalages horizontaux entre les courbes, afin d'assurer une meilleure comparabilité.

Lors des phases d'analyse exploratoire et de clustering, les deux approches ont été utilisées de manière complémentaire afin de réaliser la découpe la plus pertinente possible de l'échantillon de salves en groupes de mécanismes physiques de diverses natures. Ainsi, il a été conclu que les salves observées lors des essais précédents se répartissent entre huit types de mécanismes. Parmi ceux-ci, nous pouvons citer la rupture de la gaine du combustible, la chute des barres de contrôle ou encore des mécanismes précurseurs ou résiduels d'une rupture de la gaine. D'autre part, deux classes ont été attribuées à du bruit électronique et non électronique. Elles ne correspondent donc pas à des mécanismes d'intérêt. Les tailles relativement importantes de ces deux classes, malgré le prétraitement des signaux, témoignent d'une sensibilité significative de la technique EA au bruit dans notre contexte.

En ce qui concerne la dernière phase consacrée à la classification supervisée, nous avons opté pour une étude comparative des approches classiques et fonctionnelles. Pour chacune d'elles, diverses stratégies et méthodes ont été présentées et mises en œuvre, afin d'identifier, pour chaque approche, la méthode qui conduit aux meilleurs résultats. Ainsi, il a été établi que quelle que soit l'approche, la méthode des forêts aléatoires est celle qu'il faudrait appliquer par défaut. Si l'approche fonctionnelle apparaît comme moins efficace que l'approche classique, avec des taux de bonnes prédictions de plus de 80%, les deux approches conviennent pour la classification des données considérées dans notre contexte. Par ailleurs, les capacités prédictives de l'algorithme ne sont pas les mêmes selon les classes. On constate notamment que pour les classes de très faible effectif, les résultats de prédiction de l'algorithme sont relativement mauvais.

Enfin, la robustesse de l'algorithme a été testée grâce à des salves provenant d'un autre capteur et/ou pré-traités avec une autre procédure. Dans tous les cas, on observe une baisse des performances de l'algorithme. Une stratégie visant à mettre en place un algorithme permettant de tester si une nouvelle salve correspond à un mécanisme déjà observé ou non a également été proposée. Les analyses de performances de celui-ci montrent une capacité globale de bonnes prédictions de plus de 85%.

Conclusion générale et Perspectives

L'objectif-cible de ce travail était d'évaluer la possibilité d'identifier clairement les mécanismes sources liés au comportement du combustible lors des essais RIA, au travers de leur signature acoustique. D'un point de vue statistique, cet objectif pourrait être reformulé sous la forme : est-il possible d'identifier de manière satisfaisante, c'est-à-dire avec un taux de mauvaise prédiction suffisamment faible (fixé par l'expérimentateur), les mécanismes associés au comportement du combustible ? L'une des principales difficultés liées à la réponse de cette question réside dans le fait que l'environnement nucléaire se caractérise par une grande complexité rendant les signaux reçus particulièrement difficiles à interpréter, sans un traitement préalable permettant d'estimer les signaux sources associés aux potentiels mécanismes physiques d'intérêt. A l'issue du travail présenté ici, et en particulier celui décrit dans le [chapitre 4](#), nous pouvons répondre par l'affirmative à cette question, mais cela suppose un certain nombre de préalables.

La première source de pollution du signal essai reçu à laquelle nous nous sommes intéressés est l'impact du chemin de propagation des ondes au sein du dispositif ([chapitre 2](#)). Nous avons choisi d'expérimenter, de manière complémentaire, une approche analytique et une approche numérique. La combinaison de ces deux approches a fortement contribué à la compréhension de la physique des ondes associée au contexte de notre étude. D'un point de vue analytique, malgré des simplifications significatives du dispositif d'essais néanmoins validées par la suite par le résultat des simulations numériques, une fréquence de résonance de l'ordre de 180kHz observée sur les signaux physiques reçus lors des essais RIA a pu être retrouvée. Sur le plan numérique, nous avons mis en évidence une forte dépendance de la fonction de transfert du dispositif d'essais par rapport aux positions relatives de la source et du capteur. Ainsi, dès que la position d'un mécanisme physique est incertaine, un traitement de l'impact du dispositif par déconvolution est à exclure. Cependant, il est tout à fait envisageable de filtrer ou pénaliser les composantes du signal reçu qui sont associées à la gamme de fréquences autour de 180kHz , correspondant à la fréquence de résonance du dispositif. Par ailleurs, les résultats de simulations numériques suggèrent que jusqu'à 50kHz , la géométrie et la configuration du dispositif ne biaisent pas les caractéristiques du signal source. Enfin, un changement de fluide caloporteur, d'une boucle eau sous pression vers une boucle sodium, induit une augmentation de l'amplitude du signal reçu et un léger décalage du contenu fréquentiel vers les hautes fréquences.

Dans le [chapitre 3](#), nous avons, d'une part, traité la problématique de la suppression du bruit environnemental, et d'autre part, exploré les performances de méthodes de détection de salves capables de pallier les limites d'un seuillage classique. Il s'agissait donc, au-delà du traitement de la source de pollution qu'est le bruit environnemental, d'expérimenter diverses stratégies et méthodes permettant d'isoler les segments du signal essai associés à de potentiels mécanismes d'intérêts (les salves). Nous avons pu établir que le bruit environnemental d'un essai RIA est stationnaire. Cependant, sa nature physique étant susceptible de changer d'un essai à l'autre, il est difficile de dégager une méthode adaptée à toutes les situations et applicable indépendamment de l'essai. Notre démarche a donc été de procéder à une étude comparative de plusieurs méthodes de traitement du bruit et d'analyser les résultats selon divers critères. Lorsqu'on s'intéresse au critère classique d'amélioration du signal-à-

bruit (SNR), la soustraction spectrale apparaît de manière générale comme la méthode la plus robuste aux changements de nature du bruit. Cependant, dans certains cas, d'autres approches peuvent lui être préférées. Si le bruit est à bande étroite (respectivement, blanc gaussien), on préférera un simple filtre passe-bande (respectivement, un seuillage par ondelettes). Dans le but d'obtenir un bruit résiduel quasi-nul quelle que soit la méthode de débruitage choisie, nous avons étudié les apports de diverses combinaisons des méthodes étudiées. Nous concluons qu'il est toujours possible d'atteindre ce résultat, et cela indépendamment de la méthode choisie initialement et du type de bruit. L'analyse de l'impact des méthodes expérimentées sur les paramètres EA ne permet pas d'établir de hiérarchie entre celles-ci. Cependant, quelles que soient la méthode de débruitage choisie et la nature du bruit, il est préférable (lorsque cela est pertinent) d'interpréter les salves à partir des paramètres liés à la forme d'onde plutôt que de ceux liés au contenu fréquentiel. En effet, il est extrêmement difficile de recouvrer avec efficacité les paramètres traduisant le contenu fréquentiel et l'activité acoustique, contrairement à ceux liés à la forme d'onde. En ce qui concerne la détection des salves, les limitations de la méthode de seuillage classique ont été mises en évidence, y compris en contexte de fort SNR. Dans ce cas, les ondelettes continues apparaissent comme la méthode de détection la plus efficace, et cela quel que soit le type de bruit. Lorsque le SNR est faible, la détectabilité des salves dépend fortement du type de bruit. Pour certains types de bruit, comme les bruits très large bande par exemple, aucune des approches envisagées dans ce travail n'aboutit à des résultats satisfaisants.

Une démarche de prétraitement applicable par défaut aux signaux essais, et illustrant de quelle manière les travaux présentés dans les chapitres 2 et 3 pouvaient être exploités dans le cadre d'une démarche visant à estimer le signal source, a été proposée dans le [chapitre 4](#). Au-delà de l'estimation du signal source, nous proposons notamment des traitements permettant de réajuster la forme d'onde des salves de manière à éviter, autant que possible, de biaiser le calcul des paramètres EA. Deux approches statistiques permettant de mettre en place un algorithme de classification supervisée afin d'associer chaque salve pré-traitée à un mécanisme physique ont ensuite été proposées. La première approche, classiquement utilisée en EA, s'appuie sur une matrice de données construite à partir de paramètres EA calculés sur le signal temporel, son spectre ou sur d'autres transformations. Quant à la seconde, dite approche fonctionnelle, elle vise à faire l'économie de la phase de sélection et de calcul des paramètres EA du cas classique, en s'appuyant directement sur des courbes (l'enveloppe et le spectre) obtenues à partir du signal salve. Pendant les phases préliminaires à la mise en place de l'algorithme que sont l'analyse exploratoire et le clustering, nous avons choisi d'utiliser les deux approches de manière complémentaire afin d'assurer la découpe la plus pertinente possible de l'échantillon de salves en groupes de mécanismes physiques de diverses natures. Ainsi, nous avons établi que les salves observées lors des essais précédents se répartissent entre huit types de mécanismes. Parmi ceux-ci, nous pouvons citer la rupture de la gaine du combustible, la chute des barres de contrôle ou encore des mécanismes précurseurs ou résiduels d'une rupture de la gaine. Par ailleurs, des classes de tailles relativement importantes ont été attribuées à du bruit électronique et du bruit non électronique, ce qui témoigne d'une sensibilité significative de la technique EA au bruit dans notre contexte. Lors de la dernière phase consacrée à la classification supervisée, une étude comparative des approches classiques et fonctionnelles, à diverses stratégies et méthodes, permet de conclure que la méthode des forêts aléatoires est celle qu'il faudrait appliquer par défaut, quelle que soit l'approche choisie. Si l'approche fonctionnelle apparaît moins efficace que l'approche classique, avec des taux de bonnes prédictions de plus de 80%, les deux approches conviennent cependant pour la classification des données considérées dans notre étude.

Perspectives

La perspective immédiate de ce travail est son incorporation dans la plateforme de dépouillement des essais de sûreté RIA mise en place par le Laboratoire de Préparation et Réalisation des Essais

(LPRE) du CEA. Il permettra notamment d'améliorer la qualité des signaux EA et de contribuer à l'identification des mécanismes sources survenant lors des essais. Nous notons que le travail présenté ici a été entièrement réalisé avec des logiciels libres (R, PostgreSQL, GMSH et SPEC-FEM). Il a permis de dégager une méthodologie de traitement des données EA spécifique à l'environnement nucléaire complexe des essais RIA. Si les résultats présentés ici ne sont pas directement applicables, la méthodologie peut l'être. Certaines des fonctions développées ici seraient utiles dans le cadre d'autres travaux de recherche autour de l'analyse de données EA.

Par ailleurs, certains des résultats obtenus dans ce travail de thèse, notamment dans le [chapitre 2](#) permettent de formuler des recommandations quant au choix des capteurs EA et à leur positionnement, pour une meilleure interprétabilité des données. Par exemple, nous avons établi que les fonctions de transfert dépendent à la fois de la position de la source et de celui du capteur. Cette double dépendance pourrait être réduite à une simple dépendance à la position de la source. En effet, il serait intéressant d'optimiser les positions des deux capteurs de manière à avoir des fonctions de transfert sensiblement égales pour une position de source fixée. Cette étude pourrait être conduite soit à travers des simulations numériques de la propagation d'ondes, soit au moyen d'expérimentations sur une maquette de la structure étudiée.

Il est bien connu que la capacité prédictive d'une méthode de classification supervisée est dépendante des données d'entrée. La hiérarchie établie ici, entre l'approche classique et l'approche fonctionnelle ne saurait donc, en aucun cas, se généraliser à tous les types de données EA. En général, la capacité prédictive d'un modèle peut dépendre de la taille de l'échantillon de données, du nombre de variables explicatives ou encore des corrélations qui existent entre celles-ci. Une validation définitive de la pertinence de l'approche fonctionnelle expérimentée dans ce travail, nécessiterait de l'expérimenter dans d'autres contextes. Pour ce faire, d'un point de vue statistique, une démarche classique est la mise en œuvre de cette approche sur des données EA de référence pour lesquelles les capacités prédictives de modèles basées sur l'approche classique sont connues. N'ayant pas pu disposer de telles données dans le cadre de nos travaux, la mise en place d'une plateforme collaborative permettant le partage de données de référence serait souhaitable.

La baisse des performances de l'algorithme de classification supervisée à la suite d'un changement de capteur met en exergue la nécessité d'un approfondissement du travail de prétraitement des signaux, et donc d'estimation du signal source. Nous suggérerions par exemple de définir un coefficient de réajustement de l'énergie du signal en cas de changement de capteur. Cela permettrait notamment de corriger l'effet de la différence d'atténuation du signal selon le capteur.

Notons que l'algorithme de classification supervisée proposé dans ce travail s'appuie uniquement sur le capteur M1, les données du capteur M2 ayant été utilisées uniquement à titre de données de validation. D'autres pistes, qui permettraient notamment d'améliorer la robustesse de l'algorithme, seraient de fusionner les données des deux capteurs pour augmenter la taille de l'échantillon de salves ou de procéder à une analyse conjointe des salves issues des deux capteurs. Dans ce dernier cas, à chaque salve, donc à chaque ligne de la matrice de données, seraient associés les paramètres EA (classiques ou fonctionnels) correspondant aux capteurs M1 et M2. De plus, il serait très pertinent de rajouter des paramètres associés au comportement conjoint des deux salves comme le décalage en temps, l'intercorrélation, le coefficient de corrélation...

Enfin, notons qu'il est tout à fait possible d'envisager des modèles regroupant à la fois les paramètres EA classiques et les courbes spectres et enveloppes, ce qui revient à combiner l'approche classique et l'approche fonctionnelle. La question qui reste alors posée est de savoir si la mise en place de ce type de modèles, *a priori* plus complexes que ceux proposés dans ce travail, conduit à un gain significatif

en termes de capacité prédictive.

Références

- [1] J. ESTRADE. *Les réacteurs nucléaires expérimentaux*. Tech. Report. Commissariat à l'Énergie Atomique et aux Énergies Alternatives, 2012.
- [2] P. RUDLING *et al.* “Nuclear Fuel Behaviour under RIA Conditions”. In : *Advanced Nuclear Technology International* (2016).
- [3] L. O. JERNKVIST & A. R. MASSIH. *Nuclear Fuel Behavior Under Reactivity-initiated Accident (RIA) Conditions*. Tech. Report. Nuclear Energy Agency, 2010.
- [4] T. AASTROEM. “From fifteen to two hundred NDT-methods in fifty years”. In : *17th World Conference on Nondestructive testing*. 2008, p. 25–28.
- [5] A. S. GUERRA. “Sensory metrology in the environment of visual quality control”. Thèse de doctorat. Université de Savoie, 2008.
- [6] C. HELLIER. *Handbook of nondestructive evaluation*. McGraw-Hill, 2001.
- [7] D. LOVEJOY. “The history and basis of the magnetic particle testing method”. In : *Magnetic Particle Inspection*. Springer, 1993, p. 1–12.
- [8] H. ZAIDI. “Méthodologies pour la modélisation des couches fines et du déplacement en contrôle non destructif par courants de Foucault : application aux capteurs souples”. Thèse de doctorat. Université Paris Sud-Paris XI, 2012.
- [9] M. BRELAZ. “Intérêt et limites de la thermographie infrarouge en tant qu'outil diagnostique en aviculture”. Thèse de doctorat. Université de Toulouse, 2011.
- [10] G. CORNELOUP & C. GUEUDRÉ. *Le contrôle non destructif destructif et la contrôlabilité des matériaux et des structures*. Collection METIS LyonTech. PPUR, 2016.
- [11] L. HÉGRON, P. SORNAY & N. FAVRETTO-CRISTINI. “Compaction of a bed of fragmentable particles and associated acoustic emission”. In : *IEEE Transactions on Nuclear Science* 61.4 (2014), p. 2175–2181.
- [12] F. BAQUE *et al.* “Generation IV nuclear reactors-R&D program to improve sodium-cooled systems inspection”. In : *1st International Conference on Advancements in Nuclear Instrumentation Measurement Methods and their Applications (ANIMMA)*. IEEE. 2009, p. 1–10.
- [13] F. BAQUE *et al.* “Non destructive examination of immersed structures within liquid sodium”. In : *2nd International Conference on Advancements in Nuclear Instrumentation Measurement Methods and their Applications (ANIMMA)*. IEEE. 2011, p. 1–5.
- [14] AFNOR. *NF EN 1330-9, Essais non-destructifs- Terminologie -Partie 9 : Termes utilisés en contrôle par émission acoustique*. AFNOR, 2009.
- [15] G. POSSA. *Introduction to acoustic emission*. Tech. Report. National Aeronautics & Space Administration (NASA), 1983.
- [16] C. B. SCRUBY & H. N. WADLEY. “An assessment of acoustic emission for nuclear pressure vessel monitoring”. In : *Progress in Nuclear Energy* 11.3 (1983), p. 275–297.
- [17] S. P. YING. “The use of acoustic emission for assessing the integrity of a nuclear reactor pressure vessel”. In : *NDT International* 12.4 (1979), p. 175–179.

- [18] A. SINCLAIR, D. C. CONNORS & C. L. FORMBY. “Acoustic emission analysis during fatigue crack growth in steel”. In : *Materials Science and Engineering* 28.2 (1977), p. 263–273.
- [19] A. BEHNIA, H. K. CHAI & T. SHIOTANI. “Advanced structural health monitoring of concrete structures with the aid of acoustic emission”. In : *Construction and Building Materials* 65 (2014), p. 282–302.
- [20] M. J. WARING, M. H. RUBINSTEIN & J. R. HOWARD. “Acoustic emission of pharmaceutical materials during compression”. In : *International Journal of Pharmaceutics* 36.1 (1987), p. 29–36.
- [21] S. EISSA. “Condition monitoring of pharmaceutical powder compression during tableting using acoustic emission”. Thèse de doctorat. Brunel University School of Engineering & Design, 2003.
- [22] A. G. BEATTIE. “Acoustic emission, principles and instrumentation”. In : *Journal of Acoustic Emission* 2.12 (1983), p. 95–128.
- [23] P. RUNOW. “The use of acoustic emission methods as aids to the structural integrity assessment of nuclear power plants”. In : *International Journal of Pressure Vessels and Piping* 21.3 (1985), p. 157–207.
- [24] L. HÉGRON. “Suivi par émission acoustique de la compaction de particules fragiles d’ UO₂”. Thèse de doctorat. Aix-Marseille Université, 2014.
- [25] M. KHARRAT *et al.* “A signal processing approach for enhanced acoustic emission data analysis in high activity systems : application to organic matrix composites”. In : *Mechanical Systems and Signal Processing* 70 (2016), p. 1038–1055.
- [26] V. A. ARNAU. *Piezoelectric Transducers and Applications*. Springer, 2008.
- [27] G. GAUTSCHI. *Piezoelectric sensorics : force, strain, pressure, acceleration and acoustic emission sensors, materials and amplifiers*. Springer Science & Business Media, 2002.
- [28] ASTM. *Standard test method for primary calibration Acoustic Emission sensors*. ASTM international, 2012.
- [29] S. DIA *et al.* “Primary Calibration of Acoustic Emission Sensors by the Method of Reciprocity, Theoretical and Experimental Considerations”. In : *30th European Conference on Acoustic Emission Testing & 7th International Conference on Acoustic Emission University of Granada*. 2012, p. 12–15.
- [30] J. ROGET. *Essais non destructifs : L’émission acoustique : Mise en œuvre et applications*. AFNOR; CETIM, 1988.
- [31] C. WANG, Y. ZHANG & A. MA. “Investigation into the fatigue damage process of rubberized concrete and plain concrete by AE analysis”. In : *Journal of Materials in Civil Engineering* 23.7 (2010), p. 953–960.
- [32] D. OZEVIN *et al.* “Damage assessment of gearbox operating in high noisy environment using waveform streaming approach”. In : *Journal of Acoustic Emission* (2007), p. 355–363.
- [33] V. BARAT, Y. BORODIN & A. KUZMIN. “Intelligent Acoustic Emission signal filtering methods”. In : *Journal of Acoustic Emission* 28 (2010), p. 109–119.
- [34] I. A. VIKTOROV. *Rayleigh and Lamb waves : physical theory and applications*. Plenum Press, 1970.
- [35] W. P. MASON. *Physical Acoustics : Principles and Methods*. Academic Press, 1971.
- [36] B.A. AULD. *Acoustic fields and waves in solids*. 2^e éd. T. 1. Krieger, 1973.
- [37] J. ACHENBACH. *Wave propagation in elastic solids*. Elsevier, 2012.
- [38] M. WEVERS. “One sensor linear location of acoustic emission events using plate wave theories”. In : *Materials Science and Engineering : A* 265.1 (1999), p. 254–261.

- [39] G. PRASANNA, M. R. BHAT & C. R. L. MURTHY. “Acoustic emission source location on an arbitrary surface by geodesic curve evolution”. In : *Journal of Acoustic Emission* 25 (2007), p. 224–230.
- [40] T. KOSEL, I. GRABEC & F. KOSEL. “Intelligent location of two simultaneously active acoustic emission sources”. In : *Aerospace Science and Technology* 9.1 (2005), p. 45–53.
- [41] C. H. KNAPP & G. C. CARTER. “The generalized correlation method for estimation of time delay”. In : *IEEE Transactions on Acoustics, Speech and Signal Processing* 24.4 (1976), p. 320–327.
- [42] J. L. LACOUME. “Mesure de retard. De l’acoustique sous-marine à la tomographie acoustique océanique”. In : *GRETSI, Saint Martin d’Hères* (2007).
- [43] T. LE GALL. “Simulation of acoustic emission : Assisting in identification of acoustic signature of damage mechanisms”. Thèse de doctorat. Université de Lyon, 2016.
- [44] S. HUGUET. “Application de classificateurs aux données d’émission acoustique : identification de la signature acoustique des mécanismes d’endommagement dans les composites à matrice polymère”. Thèse de doctorat. INSA de Lyon, 2002.
- [45] D. SCIDA *et al.* “Effet de l’humidité sur le comportement mécanique des composites à fibres de lin”. In : *17èmes Journées Nationales sur les Composites (JNC17)*. 2011, p. 186.
- [46] W. Y. ROGER *et al.* “Classification of acoustic emission signals generated during welding”. In : *Journal of Acoustic Emission* 4 (1985).
- [47] M. OHTSU & K. ONO. “Pattern recognition analysis of magnetomechanical acoustic emission signals”. In : *Journal of Acoustic Emission* 3.4 (1984), p. 69–79.
- [48] S. KEYVAN & R. G. PICKARD. “Feature extraction of metal impact acoustic signals for pattern classification by neural networks”. In : *Journal of Acoustic Emission* 15.1-4 (1997), p. 79–87.
- [49] R. B. MELTON. “Classification of NDE waveforms with Autoregressive models”. In : *Review of Progress in Quantitative Nondestructive Evaluation*. Springer, 1983, p. 1117–1126.
- [50] M. OHTSU & K. ONO. “Pattern recognition analysis of acoustic emission from unidirectional carbon fiber-epoxy composites by using autoregressive modeling”. In : *Journal of Acoustic Emission* 6 (1987), p. 61–71.
- [51] M. KARMOUCHE, S. RIFFI & M. LAMOTTE. “Sur la modélisation des signaux non-stationnaires”. In : *Journal de Physique IV* 4.C5 (1994), p. C5–1357.
- [52] K. J. SIDDIQUI *et al.* “Optical waveform feature selection using a pseudo-similarity method”. In : *Journal of Acoustic Emission* 9.1 (1990), p. 9–16.
- [53] R. BELCHAMBER, D. BETTERIDGE & Y. T. CHOW. “Evaluation of pattern recognition analysis of acoustic emission from stressed polymers and composites”. In : *Journal of Acoustic Emission* 4.4 (1985), p. 71–83.
- [54] C. R. L. MURTHY, B. DATTAAGURU & A. K. RAO. “Application of pattern recognition concepts to acoustic emission signals analysis”. In : *Journal of Acoustic Emission* 6.1 (1987), p. 19–28.
- [55] I. FOROUTAN & J. SKLANSKY. “Feature selection for automatic classification of non-gaussian data”. In : *IEEE Transactions on Systems, Man and Cybernetics* 17.2 (1987), p. 187–198.
- [56] V. BARAT & A. L. ALYAKRITSKIY. “Automated method for statistical processing of Acoustic Emission testing data”. In : *Journal of Acoustic Emission* 26 (2008), p. 140–149.
- [57] D. R. HAY *et al.* “Classification of acoustic emission signals from deformation mechanisms in aluminum alloys”. In : *Journal of Acoustic Emission* 3 (1984), p. 118–129.
- [58] A. VINOGRADOV. “Classification of acoustic emissions in metallic glasses”. In : *Journal of Acoustic Emission* 17.1-2 (1999), p. 1–13.

- [59] D. KOUROUSSIS *et al.* “Unsupervised pattern recognition of acoustic emission from full scale testing of a wind turbine blade”. In : *Journal of Acoustic Emission* 18 (2000), p. 217.
- [60] A. CHICHIBU, T. KIKUCHI & T. KISHISHITA. “Principal components analysis of AE waveform parameters for investigating an instability of geotechnical structures”. In : *Journal of Acoustic Emission* 11.4 (1993), S47–S56.
- [61] A. A. ANASTASSOPOULOS & T. P. PHILIPPIDIS. “Clustering methodology for the evaluation of acoustic emission from composites”. In : *Journal of Acoustic Emission* 13.1-2 (1995), p. 11–22.
- [62] T. YAN *et al.* “Classification of acoustic emission signatures using a self-organization neural network”. In : *Journal of Acoustic Emission* 17.1-2 (1999), p. 49–59.
- [63] A. VINOGRADOV & A. YU. “Principles of statistical and spectral analysis of acoustic emission and their application to plastic deformation of metallic glasses”. In : *Journal of Acoustic Emission* 16.1-4 (1998), S158–S169.
- [64] S. KEYVAN & J. NAGARAJ. “Pattern recognition of acoustic signatures using ART2 : A neural network”. In : *Journal of Acoustic Emission* 14.2 (1996), p. 97–102.
- [65] T. P. PHILIPPIDIS, V. N. NIKOLAIDIS & J. G. KOLAXIS. “Unsupervised pattern recognition techniques for the prediction of composite failure”. In : *Journal of Acoustic Emission* 17.1-2 (1999), p. 69–81.
- [66] M. MUSTAFA *et al.* “Comparison between KNN and ANN classification in brain balancing application via spectrogram image”. In : *Journal of Computer Science & Computational Mathematics* 2.4 (2012).
- [67] A. MOOSAVIAN *et al.* “Comparison of two classifiers ; K-nearest neighbor and artificial neural network, for fault diagnosis on a main engine journal-bearing”. In : *Shock and Vibration* 20.2 (2013), p. 263–272.
- [68] Q. AI *et al.* “Acoustic emission of fatigue crack in pressure pipe under cyclic pressure”. In : *Nuclear Engineering and Design* 240.10 (2010), p. 3616–3620.
- [69] D. S. KUPPERMAN, T. N. CLAYTOR & R. GROENWALD. “Acoustic leak detection for reactor coolant systems”. In : *Nuclear Engineering and Design* 86.1 (1985), p. 13–20.
- [70] D. S. KUPPERMAN *et al.* “Analysis of acoustic leak detection for light water reactor piping”. In : *Periodic Inspection of Pressurized Components, I Mech E Conference Publications*. T. 1989. 1982, p. 45.
- [71] N. FAVRETTO-CRISTINI, L. HÉGRON & P. SORNAY. “Identification of the fragmentation of brittle particles during compaction process by the acoustic emission technique”. In : *Ultrasonics* 67 (2016), p. 178–189.
- [72] R. K. SINGH & A. R. RAO. “Steam leak detection in advance reactors via acoustics method”. In : *Nuclear Engineering and Design* 241.7 (2011), p. 2448–2454.
- [73] R. K. MILLER *et al.* “A reference standard for the development of acoustic emission pipeline leak detection techniques”. In : *NDT & E International* 32.1 (1999), p. 1–8.
- [74] R. BOURGA *et al.* “Leak-Before-Break : Global perspectives and procedures”. In : *International Journal of Pressure Vessels and Piping* (2015).
- [75] D. D. PAUL *et al.* *Evaluation and refinement of leak-rate estimation models*. Tech. Report. Nuclear Regulatory Commission, Washington, DC (USA). Div. of Engineering, Battelle Columbus Labs., OH (USA), 1991.
- [76] D. NORRIS *et al.* *PICEP : pipe crack evaluation program*. Tech. Report. Electric Power Research Inst., Palo Alto, CA (USA), 1984.

- [77] S. RAHMAN *et al.* “A computer model for probabilistic leak-rate analysis of nuclear piping and piping welds”. In : *International Journal of Pressure Vessels and Piping* 70.3 (1997), p. 209–221.
- [78] R. P. COLLIER. *Two-phase flow through intergranular stress corrosion cracks and resulting acoustic emission*. Electric Power Research Institute, 1984.
- [79] J. H. PARK *et al.* “Estimation of leak rate through circumferential cracks in pipes in nuclear power plants”. In : *Nuclear Engineering and Technology* 48.4 (2015), p. 1264–1272.
- [80] R. CARMI *et al.* “Detection of transient zones during water boiling by acoustic emission”. In : *Journal Acoustic Emission* 29 (2011), p. 89–97.
- [81] S. HUSIN & D. MBA. “Acoustic emission of a single bubble activities”. In : *Proceedings of the World Congress on Engineering 2010 Vol. II*. 2010.
- [82] T. LEIGHTON. *The acoustic bubble*. Academic Press, 2012.
- [83] R. MANASSEH *et al.* “Sound emission on bubble coalescence : imaging, acoustic and numerical experiments”. In : *16th Australasian Fluid Mechanics Conference (AFMC)*. School of Engineering, The University of Queensland. 2007, p. 167–173.
- [84] C. HAN & P. GRIFFITH. “The mechanism of heat transfer in nucleate pool boiling—Part I : Bubble initiation, growth and departure”. In : *International Journal of Heat and Mass Transfer* 8.6 (1965), p. 887–904.
- [85] T. DIVOUX *et al.* “Acoustic emission associated with the bursting of a gas bubble at the free surface of a non-Newtonian fluid”. In : *Physical Review E* 77.5 (2008), p. 056310.
- [86] S. NUKIYAMA. “The maximum and minimum values of the heat Q transmitted from metal to boiling water under atmospheric pressure”. In : *International Journal of Heat and Mass Transfer* 9.12 (1966), p. 1419–1433.
- [87] G. KERBOUL. “Etude de l’endommagement des produits céramiques crus par émission acoustique”. Thèse de doctorat. INSA de Lyon, 1992.
- [88] F. FERNANDES, A. I. SYAHRIAL & J. R. VALDES. “Monitoring the oedometric compression of sands with acoustic emissions”. In : *ASTM Geotechnical Testing Journal* 33.5 (2010), p. 410–415.
- [89] T. KATHRINA & R. D. RAWLINGS. “Acoustic studies of the effect of SiC particle reinforcement on the compaction of alumina powder”. In : *Journal of the European Ceramic Society* 17.9 (1997), p. 1157–1166.
- [90] M. HE. “Caractérisation du comportement à rupture des alliages de zirconium de la gaine du crayon combustible des centrales nucléaires dans la phase post-trempe d’un APRP (Accident de Perte de Réfrigérant Primaire)”. Thèse de doctorat. Ecole Nationale Supérieure des Mines de Paris, 2012.
- [91] S. ARSÈNE. “Effet de la microstructure et de la température sur la transition ductile-fragile des zircaloy hydrures”. Thèse de doctorat. *École Centrale Paris*, 1997.
- [92] M. BALOURDET *et al.* “The PROMETRA programme : Assessment of mechanical properties of Zircaloy-4 fuel cladding during an RIA”. In : *15th International Conference on Structural Mechanics in Reactor Technology, Korean Nuclear Society, Seoul, Korea*. 1999, p. 485–492.
- [93] R. R. BIEDERMAN *et al.* *Study of Zircaloy-4 : steam oxidation reaction kinetics. Final report*. Tech. Report. Worcester Polytechnic Inst., Mass.(USA). Higgins Lab., 1978.
- [94] M. BILLONE *et al.* *Cladding embrittlement during postulated loss-of-coolant accidents*. Tech. Report. Argonne National Laboratory (ANL), 2008.

- [95] L. R. BUNNELL, J. L. BATES & G. B. MELLINGER. “Some high-temperature properties of Zircaloy-oxygen alloys”. In : *Journal of Nuclear Materials* 116.2 (1983), p. 219–232.
- [96] V. BUSSEY *et al.* “Mécanismes d’endommagement de la couche d’oxyde des gaines de crayons de combustible en situation accidentelle de type RIA”. In : *Matériaux & Techniques* 99.1 (2011), p. 127–130.
- [97] M. LE SAUX. “Comportement et rupture de gaines en zircaloy-4 détendu vierges, hydrurées ou irradiées en situation accidentelle de type RIA”. Thèse de doctorat. École Nationale Supérieure des Mines de Paris, 2008.
- [98] A. STERN. “Comportements métallurgique et mécanique des matériaux de gainage du combustible REP oxydes à haute température”. Thèse de doctorat. École Nationale Supérieure des Mines de Paris, 2007.
- [99] P. YVON *et al.* “Development of new techniques to assess the mechanical behaviour of Zircaloy-4 during an RIA”. In : *Fuel behaviour under transient and LOCA conditions* (2002), p. 111.
- [100] C. R. HEIPLE & S. H. CARPENTER. “Acoustic emission produced by deformation of metals and alloys-A review.” In : *Journal of Acoustic Emission* 6 (1987), p. 177–204.
- [101] C. R. HEIPLE & S. H. CARPENTER. “Acoustic emission produced by deformation of metals and alloys. II”. In : *Journal of Acoustic Emission* 6 (1987), p. 215–237.
- [102] A. VINOGRADOV, A. DANUYK & E. POMPONI. “Mechanisms of Plastic Deformation and Acoustic Emission in ZK60 Mg Alloy”. In : *Proc. of 30th Euro Conf. on Acoustic Emission Testing & 7th Int. Conf on Acoustic Emission*. 2012.
- [103] P. K. CHAURASIA *et al.* “Tensile behaviour and acoustic emission in zirconium alloys after thermal aging and sodium exposure”. In : *Journal of Nuclear Materials* 322.2 (2003), p. 217–227.
- [104] S. A. NIKULIN *et al.* “The mechanism and kinetics of the fuel cladding failure during loading after high-temperature oxydation”. In : *Journal of Nuclear Materials* 452.1 (2014), p. 102–109.
- [105] V. PERES *et al.* “In situ high temperature oxydation analysis of Zircaloy-4 using acoustic emission coupled with thermogravimetry”. In : *Journal of Nuclear Materials* 461 (2015), p. 365–375.
- [106] O. I. TRAORE *et al.* “Impact of the Test Device on Acoustic Emission Signals from Nuclear Safety Experiments : Contribution of Wave Propagation Modeling to Signal Processing”. In : *Submitted in IEEE Transactions on Nuclear Science* (2017).
- [107] O. I. TRAORE *et al.* “Impact of the Test Device on the Behavior of the Acoustic Emission Signals : Contribution of the Numerical Modeling to Signal Processing”. In : *5th International Conference on Advancements in Nuclear Instrumentation Measurement Methods and their Applications (ANIMMA)*. IEEE. 2017.
- [108] M. OHTSU & K. ONO. “The generalized theory and source representations of acoustic emission”. In : *Journal of Acoustic Emission* 5.4 (1986), p. 124–133.
- [109] D. C. GAZIS. “Three-dimensional investigation of the propagation of waves in hollow circular cylinders. I. Analytical foundation”. In : *The Journal of the Acoustical Society of America* 31.5 (1959), p. 568–573.
- [110] J. Jr. ZEMANEK. “An experimental and theoretical investigation of elastic wave propagation in a cylinder”. In : *The Journal of the Acoustical Society of America* 51.1B (1972), p. 265–283.
- [111] J. N. BARSHINGER & Joseph L. ROSE. “Guided wave propagation in an elastic hollow cylinder coated with a viscoelastic material”. In : *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 51.11 (2004), p. 1547–1556.

- [112] R. D. DOOLITTLE & H. ÜBERALL. “Sound scattering by elastic cylindrical shells”. In : *The Journal of the Acoustical Society of America* 39.2 (1966), p. 272–275.
- [113] D. C. GAZIS. “Exact Analysis of the Plane-Strain Vibrations of Thick-Walled Hollow Cylinders”. In : *The Journal of the Acoustical Society of America* 30.8 (1958), p. 786–794.
- [114] M. ONOE. *Tables of Modified Quotients of Bessel Functions of the First Kind for Real and Imaginary Arguments*. Columbia University Press, 1958.
- [115] D. KOMATITSCH & J. P. VILOTTE. “The spectral element method : An efficient tool to simulate the seismic response of 2D and 3D geological structures”. In : *Bulletin of the seismological society of America* 88.2 (1998), p. 368–392.
- [116] P. CRISTINI & D. KOMATITSCH. “Some illustrative examples of the use of a spectral-element method in ocean acoustics”. In : *The Journal of the Acoustical Society of America* 131.3 (2012), EL229–EL235.
- [117] D. KOMATITSCH *et al.* “Elastic surface waves in crystals—Part 2 : Cross-check of two full-wave numerical modeling methods”. In : *Ultrasonics* 51.8 (2011), p. 878–889.
- [118] J. TROMP, D. KOMATITSCH & Q. LIU. “Spectral-element and adjoint methods in seismology”. In : *Communications in Computational Physics* 3.1 (2008), p. 1–32.
- [119] D. KOMATITSCH. “Spectral and spectral-element methods for the 2D and 3D elastodynamics equations in heterogeneous media”. Thèse de doctorat. Institut de Physique du Globe de Paris, 1997.
- [120] A. BOTTERO *et al.* “An axisymmetric time-domain spectral-element method for full-wave simulations : Application to ocean acoustics”. In : *The Journal of the Acoustical Society of America* 140.5 (2016), p. 3520–3530.
- [121] C. GEUZAINÉ & J. F. REMACLE. “Gmsh : A 3-D finite element mesh generator with built-in pre-and post-processing facilities”. In : *International Journal for Numerical Methods in Engineering* 79.11 (2009), p. 1309–1331.
- [122] C. GEUZAINÉ & J. F. REMACLE. *Gmsh Reference Manual*. 2.11.0. 2015. URL : <http://www.geuz.org/gmsh>.
- [123] O. I. TRAORE *et al.* “Structure analysis and denoising using Singular Spectrum Analysis : application to acoustic emission signals from nuclear safety experiments”. In : *Measurement* 104 (2017), p. 78–88. ISSN : 0263-2241.
- [124] O. I. TRAORE *et al.* “Which methods and strategies to cope with noise complexity for an effective interpretation of acoustic emission signals in noisy nuclear environment?” In : *Acta Acustica united with Acustica* (2017).
- [125] O. I. TRAORÉ *et al.* “Emission acoustique et traitement du bruit : cas de signaux expérimentaux en contexte nucléaire”. In : *13e Congrès Français d’Acoustique*. 2016.
- [126] O. I. TRAORE *et al.* “Time series analysis by singular spectrum methods : Application to the processing of acoustic emission signals”. In : *16th Anglo-French Physical Acoustics Conference (AFPAC)*. 2017.
- [127] B. SALANIÉ. “Guide pratique des séries non-stationnaires”. In : *Économie & Prévision* 137.1 (1999), p. 119–141.
- [128] D. WOTZKA. “Mathematical model and regression analysis of acoustic emission signals generated by partial discharges”. In : *Applied and Computational Mathematics* 3.5 (2014), p. 225–230.
- [129] J. G. PROAKIS & D. G. MANOLAKIS. *Digital Signal Processing : Principles, Algorithms and Applications*. Sous la dir. de Pearson Prentice HALL. 4^e éd. 2007.

- [130] T. W. PARKS & C. S. BURRUS. *Digital filter design*. Wiley-Interscience, 1987.
- [131] S. SRIRAM *et al.* “Signal denoising techniques for partial discharge measurements”. In : *IEEE Transactions on Dielectrics and Electrical Insulation* 12.6 (2005), p. 1182–1191.
- [132] S. F. BOLL. “Suppression of acoustic noise in speech using spectral subtraction”. In : *IEEE Transactions on Acoustics, Speech and Signal Processing* 27.2 (1979), p. 113–120.
- [133] M. BEROUTI, R. SCHWARTZ & J. MAKHOUL. “Enhancement of speech corrupted by acoustic noise”. In : *International Conference on Acoustics, Speech, and Signal Processing*. T. 4. IEEE. 1979, p. 208–211.
- [134] A. CHOKKARAPU, S. C. UPPALAPATI & A. CHINTAKUNTLA. “Implementation of spectral subtraction noise suppressor using DSP processor”. In : *International Journal of Computer Science and Telecommunications* 4.3 (2013), p. 29–33.
- [135] S. V. VASEGHI. *Advanced digital signal processing and noise reduction*. John Wiley & Sons, 2008.
- [136] D. S. BROOMHEAD *et al.* “Singular system analysis with application to dynamical systems”. In : *Chaos, Noise and Fractals* . (1987), p. 15–27.
- [137] H. HASSANI. “Singular spectrum analysis : methodology and comparison”. In : *Journal of Data Science* 5.2 (2007), p. 239–257.
- [138] M. GHIL *et al.* “Advanced spectral methods for climatic time series”. In : *Reviews of Geophysics* 40.1 (2002).
- [139] A. SERITA *et al.* “Principal component analysis and singular spectrum analysis of ULF geomagnetic data associated with earthquakes”. In : *Natural Hazards and Earth System Science* 5.5 (2005), p. 685–689.
- [140] H. HASSANI & A. ZHIGLJAVSKY. “Singular spectrum analysis : methodology and application to economics data”. In : *Journal of Systems Science and Complexity* 22.3 (2009), p. 372–394.
- [141] H. HASSANI, A. S. SOOFI & A. A. ZHIGLJAVSKY. “Predicting daily exchange rate with singular spectrum analysis”. In : *Nonlinear Analysis : Real World Applications* 11.3 (2010), p. 2023–2034.
- [142] N. GOLYANDINA, V. NEKRUTKIN & A.A. ZHIGLJAVSKY. *Analysis of time series structure : SSA and related techniques*. CRC Press, 2001.
- [143] H. HASSANI, R. MAHMOUDVAND & M. ZOKAEI. “Separability and window length in singular spectrum analysis”. In : *Comptes-rendus Mathématiques l’Académie des Sciences* 349.17 (2011), p. 987–990.
- [144] R. COHEN. “Signal denoising using wavelets”. In : *Project Report, Department of Electrical Engineering Technion, Israel Institute of Technology, Haifa* (2012).
- [145] B. B. HUBBARD. *The World According to Wavelets. The Story of a Mathematical Technique in the Making*. Universities Press, 2005.
- [146] S. MALLAT. *A wavelet tour of signal processing*. Academic Press, 1999.
- [147] D. B. PERCIVAL & A. T. WALDEN. *Wavelet methods for time series analysis*. T. 4. Cambridge University Press, 2006.
- [148] D. L. DONOHO & I. M. JOHNSTONE. “Ideal spatial adaptation by wavelet shrinkage”. In : *Biometrika* (1994), p. 425–455.
- [149] T. T. CAI & B. W. SILVERMAN. “Incorporating information on neighbouring coefficients into wavelet estimation”. In : *Sankhyā : The Indian Journal of Statistics, Series B* (2001), p. 127–148.

- [150] O. I. TRAORE *et al.* “Contribution of functional approach to the classification and the identification of acoustic emission source mechanisms”. In : *Functional Statistics and Related Fields*. Springer, 2017, p. 251–259.
- [151] S. ULLAH & C. F. FINCH. “Applications of functional data analysis : A systematic review”. In : *BMC Medical Research Methodology* 13.1 (2013), p. 43.
- [152] V. LEFRANÇOIS. “Caractérisation d’événements à partir de mesures expérimentales”. Mém.de mast. UTC, 2010.
- [153] O.I. TRAORE. “Caractérisation et classification de signaux issus de capteurs de mesure”. Mém.de mast. Universités Toulouse I & III, 2014.
- [154] L. PANTERA & O. I. TRAORE. “Reproducible data processing research for the CABRI RIA experiments Acoustic Emission signal analysis”. In : *4th International Conference on Advancements in Nuclear Instrumentation Measurement Methods and their Applications (ANIMMA)*. IEEE. 2015, p. 1–8.
- [155] F. LLOBELL. “Classification de signaux d’émission acoustique en contexte expérimental d’accident nucléaire”. Mém.de mast. Université de Nantes, 2017.
- [156] J. M. AZAÏS & J. M. BARDET. *Le Modèle Linéaire par l'exemple Régression, Analyse de la Variance et Plans d'Expériences Illustrations numériques avec les logiciels R, SAS et Splus*. Dunod, 2006, p. 326.
- [157] F. FERRATY & P. VIEU. *Nonparametric functional data analysis : theory and practice*. Springer Science & Business Media, 2006.
- [158] J. O. RAMSAY & B. W. SILVERMAN. *Applied Functional Data Analysis : Methods and Case Studies*. New York, NY : Springer New York, 2002.
- [159] I. JOLLIFFE. *Principal component analysis*. Wiley Online Library, 2002.
- [160] B. DESGRAUPES. “Clustering indices”. In : *University of Paris Ouest-Lab Modal’X* 1 (2013), p. 34.
- [161] J. FRIEDMAN, T. HASTIE & R. TIBSHIRANI. *The elements of statistical learning*. T. 1. Springer Series in Statistics, Berlin, 2001.
- [162] J. RAMSAY & B. W. SILVERMAN. *Functional Data Analysis*. Springer Science & Business Media, 2005.
- [163] W. N. VENABLES & B. D. RIPLEY. *Modern applied statistics with S-PLUS*. Springer Science & Business Media, 2013.
- [164] R. LANCELOT & M. LESNOFF. “Sélection de modèles avec l’AIC et critères d’information dérivés”. In : *CIRAD, Montpellier* (2005).
- [165] C. M. BISHOP. *Pattern recognition and machine learning*. Springer, 2006.
- [166] L. BREIMAN *et al.* *Classification and regression trees*. CRC Press, 1984.
- [167] H. A. CHIPMAN, E. I. GEORGE & R. E. MCCULLOCH. “Bayesian CART model search”. In : *Journal of the American Statistical Association* 93.443 (1998), p. 935–948.
- [168] L. BREIMAN. “Bagging predictors”. In : *Machine Learning* 24.2 (1996), p. 123–140.
- [169] Y. FREUND, R. SCHAPIRE & N. ABE. “A short introduction to boosting”. In : *Journal-Japanese Society For Artificial Intelligence* 14.771-780 (1999), p. 1612.
- [170] J. ZHU *et al.* “Multi-class adaboost”. In : *Statistics and its Interface* 2.3 (2009), p. 349–360.
- [171] F. FERRATY & P. VIEU. “Curves discrimination : a nonparametric functional approach”. In : *Computational Statistics & Data Analysis* 44.1 (2003), p. 161–173.

-
- [172] A. GOIA. “A functional linear model for time series prediction with exogenous variables”. In : *Statistics & Probability Letters* 82.5 (2012), p. 1005–1011.
- [173] M. FEBRERO-BANDE, M. O. de la FUENTE *et al.* “Statistical computing in functional data analysis : The R package `fda.usc`”. In : *Journal of Statistical Software* 51.4 (2012), p. 1–28.
- [174] F. FERRATY & P. VIEU. “The functional nonparametric model and application to spectrometric data”. In : *Computational Statistics* 17.4 (2002), p. 545–564.
- [175] R. RAKOTOMALALA. “Pratique de la Régression Logistique-Régression Logistique Binaire et Polytomique”. In : *Université Lumière Lyon 2* (2009).
- [176] Y. ANDO. “Autocorrelation-based features for speech representation”. In : *Acta Acustica united with Acustica* 101.1 (2015), p. 145–154.

Liste des Figures

0.1. Illustration de l'impact d'un RIA sur le couple gaine/combustible (©[1, 2])	1
0.2. Schéma illustratif du dispositif d'essais dans lequel sont positionnés deux capteurs d'émission acoustique (sensors 1 et 2).	3
1.1. Illustration des deux principaux types d'EA.	7
1.2. Illustration du processus de contrôle par EA, de la source à l'analyse du signal EA.	8
1.3. Illustration de la stratégie de détection des salves par seuillage (©[31])	10
1.4. Illustration d'un processus de localisation de source par utilisation des différences de temps d'arrivée selon le type d'onde (© http://muravin.com/)	12
1.5. Exemples de paramètres EA quantitatifs calculés à partir du signal temporel et son spectre (©[45] and [43]).	13
2.1. Exemple de bruit de fonctionnement enregistré avant le début d'un essai RIA réalisé en 1997. Signal enregistré par le capteur 2.	20
2.2. Coupe radiale du dispositif d'essais (non simplifié) ($z = 0, \theta = 0$) illustrant les différents types de matériaux (ou strates) le constituant. Il est à noter que le nombre et l'ordre des strates peuvent changer selon la position sur l'axe Oz (cf. Figure 2.3a).	22
2.3. Coupe longitudinale du dispositif d'essais simplifié, de longueur $3m$ et de diamètre $17.2mm$	23
2.4. Superposition des spectres normalisés du signal correspondant au bruit de fonctionnement du réacteur (—) et de la fonction de transfert du dispositif d'essais simplifié (—) pour une source ponctuelle placée sur la gaine du combustible.	29
2.5. Illustration de l'impact d'un changement de position de la source sur le spectre de la fonction de transfert du dispositif, le capteur M1 étant fixé.	30
2.6. Illustration de l'impact d'un changement de position du capteur sur le spectre de la fonction de transfert du dispositif, la source étant fixe (haut : au sommet de la gaine du combustible, bas : en partie basse du canal sodium)	30
2.7. Zoom, dans la gamme de fréquences $[0, 200] kHz$, sur les spectres des fonctions de transfert pour différentes positions de source dans la strate correspondant à la gaine du combustible et pour deux positions du capteur (M1 et M2).	31
2.8. Instantanés de la distribution pression/contrainte normale à différents temps au cœur du dispositif d'essais pour deux types de fluide caloporteur. La zone 1 correspond au support Inox et au couple combustible (Uranium)/gaine (Zircaloy), tandis que la zone 2 (respectivement, la zone 3) représente le fluide caloporteur (respectivement, la couche externe de Zircaloy). La position de la source (respectivement, du capteur) est représentée par une étoile blanche (respectivement, un triangle gris).	32
2.9. Signaux simulés (a) et spectres associés (b) obtenus dans le cas où la source est localisée dans le combustible (cf. Figure 2.8) et le récepteur dans le fluide caloporteur eau (—) ou sodium (—).	33
3.1. Types de bruit observés ou potentiellement observables dans un environnement RIA. Signaux temporels (—) et spectres associés (—)	36

3.2. Exemple illustratif de l'évolution temporelle et fréquentielle du bruit environnemental pendant un essai RIA.	37
3.3. Signal source simulé (—) et spectre associé (—)	38
3.4. Transformée de Fourier à court terme du signal source simulé présenté en Figure 3.3	39
3.5. Quelques vecteurs propres U_i choisis à l'issue de la SVD de la matrice trajectoire du signal source simulé présenté en Figure 3.3. ($L = 2000$)	41
3.6. Matrice de corrélation des 20 premiers vecteurs propres obtenus à la suite de la SVD de la matrice trajectoire du signal source simulé présenté en Figure 3.3. ($L = 2000$)	42
3.7. H-matrix du signal source simulé présenté en Figure 3.3 : $B = T = 100$, $L = 50$, $\mathcal{L}_r^{(1)} = \text{vect}(U_1^1, U_1^2)$. Notons que, dans ce graphique, les coefficients indéfinis de la H-matrix ont été remplacés par -1 et sont représentés en blanc.	43
3.8. Exemple illustratif de l'estimation du signal source par comparaison des contributions de vecteurs propres. Cas d'une pollution du signal présenté en Figure 3.3 par un bruit blanc gaussien (SNR=10dB). ($L=2000$)	44
3.9. Illustration de la première étape de la stratégie de débruitage en deux étapes.	45
3.10. Résultat de l'amélioration de la SSA classique par application d'une stratégie de débruitage en deux étapes, consistant à supprimer les segments de bruit pur avant d'appliquer la SSA.	45
3.11. Salves choisies (a) et spectres associés (b)	48
3.12. Signal source s résultant de la concaténation des six salves présentées en Figure 3.11a	48
3.13. Illustration de la détection des salves au moyen des différentes méthodes, dans le cas d'un signal source pollué par un bruit électrique et dans un contexte de fort SNR (+10dB).	49
3.14. Illustration de la difficulté de détection des salves au moyen de différentes méthodes, dans le cas d'un signal pollué par un bruit large bande et dans un contexte de faible SNR (-10dB).	50
3.15. Illustration des bruits résiduels à la suite de débruitage par SS et par ondelettes.	51
3.16. Débruitage au moyen des différentes méthodes, dans le cas du signal source (cf. Figure 3.12) pollué par un bruit électrique (cf. Figure 3.1b).	52
3.17. Amélioration de débruitage, par diminution du bruit résiduel, en utilisant une combinaison des méthodes présentées.	54
3.18. Amélioration du débruitage par utilisation de la SSA combinée à la SS.	54
4.1. Illustration du traitement du bruit de fonctionnement du signal essai reçu au capteur M1. Cas d'un essai rupture réalisé en 1998.	58
4.2. Illustration de la mise en œuvre de l'algorithme de détection des salves.	59
4.3. Illustration de la mise en œuvre de la méthode de pénalisation de l'impact du système d'acquisition dans le cas de la salve 128. Signal reçu au capteur M1.	60
4.4. Échantillon de 12 salves (reçues au capteur M1) choisies parmi les 168 détectées au cours des essais menés depuis 1993.	61
4.5. Illustration de la correction des biais de détection dans le cas de la salve 121. Le support I_i de χ_i^* est indiqué par les lignes verticales en rouge. $\alpha = 0.002$	62
4.6. Illustration de la correction des biais de détection dans le cas de la salve 84. Le support I_i^* de χ_i^* est indiqué par les lignes verticales en rouge. $\alpha_1 = 0.12$	63
4.7. Illustration de la correction des temps de montée et de chute dans le cas d'une salve entretenue (74). Les temps de montée et de chute recalculés sont notés respectivement par RT (rise time) et FT (fall time). $\alpha_2 = 0.05$	64
4.8. Résultat du processus de prétraitement pour les salves présentées en Figure 4.4. Les salves pré-traitées sont en bleue.	64
4.9. Enveloppe et spectre associés aux 12 salves présentées en Figure 4.8	66

4.10. Illustration du processus de lissage des enveloppes et spectres par des fonctions splines cubiques.	68
4.11. Illustration du traitement des différences de supports et des décalages horizontaux dans le cas des enveloppes des salves.	69
4.12. Échantillon de 168 courbes obtenues à la suite du processus de prétraitement des salves reçues aux capteur M1 (droite) et au capteur M2 (gauche).	69
4.13. Représentations des individus (salves) et variables (paramètres EA) dans les plans (1,2) et (1,3) à la suite de l'ACP classique (capteur M1).	71
4.14. ACP classique : Représentations des individus (salves) et variables (paramètres EA) dans les plans (1,2) et (1,3) à la suite de l'ACP classique (capteur M2)	72
4.15. Représentations des individus (salves) dans les plans (1,2) et (1,3) à la suite de l'ACP fonctionnelle sur les enveloppes, pour les capteurs M1 et M2.	72
4.16. Représentations des individus (salves) dans les plans (1,2) et (1,3) à la suite de l'ACP fonctionnelle sur les spectres, pour les capteurs M1 et M2.	73
4.17. Représentation des deux premières harmoniques (fonctions propres) à la suite de la mise en œuvre de l'ACP fonctionnelle.	73
4.18. Schéma illustratif de l'approche de classification non supervisée en deux étapes	77
4.19. Résultats de la première étape de mise en œuvre de la stratégie de classification combinant les approches de réduction de dimension et non paramétrique.	79
4.20. Silhouettes obtenues à la suite de la seconde étape de la stratégie de classification combinant les approches de réduction de dimension et non paramétrique.	79
4.21. Superposition des enveloppes selon la classe	80
4.22. Superposition des spectres normalisés selon la classe.	80
4.23. Comparaison entre les classes de la classification non supervisée et de la statistique fonctionnelle (Plan 1/2 de l'AFC du Tableau 4.7)	81
4.24. Taux d'erreur en fonction du nombre k de plus-proches-voisins	86
4.25. Arbre de décision du jeu de données	87
4.26. Courbes d'évolution de l'erreur de prédiction en fonction du nombre d'arbres.	88
4.27. Évolution de la spécificité et de la sensibilité en fonction du seuil.	95
4.28. Minimum des distances aux centres de gravité des classes pour les événements qui ont construit le modèle	96
A.1. Illustration du calcul de quelques paramètres EA obtenus à partir de la fonction d'auto-corrélation.	120
D.1. Éboulis de valeurs propres. ACP cas classique.	130
D.2. Éboulis de valeurs propres. ACP cas fonctionnel	131
D.3. Centre de gravité du groupe 1 avec intervalles de confiance	132
D.4. Centre de gravité du groupe 2 avec intervalles de confiance	132
D.5. Centre de gravité du groupe 3 avec intervalles de confiance	133
D.6. Centre de gravité du groupe 4 avec intervalles de confiance	133
D.7. Centre de gravité du groupe 5 avec intervalles de confiance	133
D.8. Centre de gravité du groupe 6 avec intervalles de confiance	134
D.9. Centre de gravité du groupe 7 avec intervalles de confiance	134
D.10. Centre de gravité du groupe 8 avec intervalles de confiance	134
D.11. Représentations des superpositions d'enveloppes selon la classe	136
D.12. Représentations des superpositions de spectres normalisés selon la classe	136

Liste des Tableaux

1.1. Classification des méthodes de traitement du bruit selon BARAT, BORODIN & KUZMIN [33]	11
2.1. Propriétés physiques (densités, vitesses de propagation des ondes, et impédances) des différentes strates constituant le dispositif d'essais simplifié.	22
3.1. Test de non-stationnarité de Dickey-Fuller pour des segments de bruit prélevés à divers moments d'un essai RIA.	37
3.2. Ratio d'amélioration du SNR en fonction de la méthode de débruitage utilisée.	51
3.3. Quelques paramètres EA classiques pour les six salves présentées en Figure (3.11a)	52
4.1. Variables restantes après élimination des variables redondantes. Ces variables sont celles utilisées pour la classification non supervisée (clustering) des salves.	65
4.2. Méthodes de classification non supervisée testées	75
4.3. Tableau récapitulatif des résultats obtenus suite à la mise en œuvre des 4 méthodes de clustering choisies dans le cas classique. Signalons que CDB est l'abréviation de Chute Des Barres.	76
4.4. Tableau récapitulatif des résultats du processus de choix de l'approche d'analyse dans le cas des spectres. Signalons que CDB est l'abréviation de Chute Des Barres. Notons que l'absence du premier axe s'explique par le fait qu'il correspond à une variation commune à toutes les salves, et donc non discriminante.	78
4.5. Tableau comparatif des résultats des différentes approches de classification non supervisée par CAH dans le cas fonctionnel. Les détails de mise en œuvre des méthodes sont présentés dans le Tableau D.1 en Annexe D.2.1	78
4.6. Tableau récapitulatif des résultats du processus de choix de l'approche d'analyse dans le cas des spectres	81
4.7. Table de confusion entre les classes de la classification non supervisée et de la statistique fonctionnelle	82
4.8. Classes obtenues grâce à l'expertise acoustique et au recoupement avec la statistique fonctionnelle	82
4.9. Nombre de salves dans chaque jeu de données	83
4.10. Variables obtenues par la sélection de variables de type Backward. A noter que les sélections de variables de type Backward et mixte aboutissent aux mêmes variables. La méthode Forward conduit à des modèles de moins bonne qualité.	84
4.11. Variables obtenues par la sélection de variables de type Forward. A noter que la sélection Mixte permet d'obtenir les mêmes résultats que la Forward, et que la Backward aboutit à de moins bons résultats.	85
4.12. Moyenne, minimum et écart-type des pourcentages de bonnes prédictions des 2 approches du modèle bagging	88
4.13. Variables conservées	89
4.14. Moyenne, minimum et écart-type des pourcentages de bonnes prédictions suivant le palier de la sélection de variables avec la méthode 1 de bootstrap. <i>mtry</i> est le nombre de variables tirées à chaque nœud.	89

4.15. Comparaison des modèles par moyenne, minimum et écart-type des pourcentages de bonnes prédictions	90
4.16. Cas de l'approche fonctionnelle non paramétrique : comparaison des modèles par moyenne, minimum et écart-type des pourcentages de bonnes prédictions. La métrique utilisée est la métrique \mathcal{L}_2	91
4.17. Cas de l'approche fonctionnelle par réduction de dimension : comparaison des modèles par moyenne, minimum et écart-type des pourcentages de bonnes prédictions	92
4.18. Rappels des modèles	93
4.19. Précisions des modèles	93
4.20. Performance du modèle classique de forêts aléatoires sur les données des deux capteurs et ayant subi ou non un débruitage. Le modèle a été construit via les données du capteur M1 débruitées par soustraction spectrale (SS). Le capteur M2 est l'autre capteur du réacteur expérimental.	94
4.21. Performance du modèle fonctionnel de forêts aléatoires sur les données des deux capteurs et ayant subi ou non un débruitage. Le modèle a été construit via les données du capteur M1 débruitées par soustraction spectrale (SS). Le capteur M2 est l'autre capteur du réacteur expérimental.	94
4.22. Performance du modèle kNN fonctionnel sur les données des deux capteurs et ayant subi ou non un débruitage. Le modèle a été construit via les données du capteur M1 débruitées par soustraction spectrale (SS). Le capteur M2 est l'autre capteur du réacteur expérimental.	94
4.23. Variables utilisées pour construire la distance de Mahalanobis	95
4.24. Table de confusion de l'algorithme avec le seuil valant 53	96
A.1. Paramètres EA liés au signal temporel	118
A.2. Paramètres EA liés au spectre du signal	119
A.3. Paramètres EA à connotation statistique	119
C.1. Quelques paramètres EA classiques associés aux six salves présentées en Figure (3.11a) recouverts à la suite d'un débruitage par SS dans le cas d'un bruit gaussien.	126
C.2. Quelques paramètres EA classiques associés aux six salves présentées en Figure (3.11a) recouverts à la suite d'un débruitage utilisant un seuillage par ondelettes dans le cas d'un bruit gaussien.	126
C.3. Quelques paramètres EA classiques associés aux six salves présentées en Figure (3.11a) recouverts à la suite d'un débruitage par SSA dans le cas d'un bruit gaussien.	127
C.4. Quelques paramètres EA classiques associés aux six salves présentées en Figure (3.11a) recouverts à la suite d'un débruitage par SS dans le cas d'un bruit électrique.	127
C.5. Quelques paramètres EA classiques associés aux six salves présentées en Figure (3.11a) recouverts à la suite d'un débruitage utilisant un seuillage par ondelettes dans le cas d'un bruit électrique.	127
C.6. Quelques paramètres EA classiques associés aux six salves présentées en Figure (3.11a) recouverts à la suite d'un débruitage par SSA dans le cas d'un bruit électrique.	127
C.7. Quelques paramètres EA classiques associés aux six salves présentées en Figure (3.11a) recouverts à la suite d'un débruitage par SS dans le cas d'un bruit mixte.	128
C.8. Quelques paramètres EA classiques associés aux six salves présentées en Figure (3.11a) recouverts à la suite d'un débruitage utilisant un seuillage par ondelettes dans le cas d'un bruit mixte.	128
C.9. Quelques paramètres EA classiques associés aux six salves présentées en Figure (3.11a) recouverts à la suite d'un débruitage par SSA dans le cas d'un bruit mixte.	128
C.10. Quelques paramètres EA classiques associés aux six salves présentées en Figure (3.11a) recouverts à la suite d'un débruitage par SS dans le cas de la somme des bruits.	128

C.11. Quelques paramètres EA classiques associés aux six salves présentées en Figure (3.11a) recouverts à la suite d'un débruitage utilisant un seuillage par ondelettes dans le cas de la somme des bruits.	129
C.12. Quelques paramètres EA classiques associés aux six salves présentées en Figure (3.11a) recouverts à la suite d'un débruitage par SSA dans le cas de la somme des bruits.	129
D.1. Informations complémentaires relatives à la mise en œuvre du clustering selon les trois stratégies expérimentées dans le cas fonctionnel.	131
D.2. Effectifs par classe	132
D.3. Caractérisation des classes . + : significativement plus grand que la moyenne. – : significativement plus petit que la moyenne	135
D.4. Caractérisation des classes . + : significativement plus grand que la moyenne. – : significativement plus petit que la moyenne	135

Annexe A.

Annexes

A.1. Liste non exhaustive de paramètres EA utilisés dans la littérature

Les paramètres non utilisés dans le cadre de nos travaux sont surlignés en bleu.

Paramètre	Unité	Définition
Temps de montée (riseTime)	<i>s</i>	Temps qui s'écoule entre le déclenchement de la salve et l'amplitude maximale
Taux de passage par zéro	-	Nombre de fois que le signal passe par zéro
Temps de décroissance (fall-Time)	<i>s</i>	Temps qui s'écoule entre l'amplitude maximale et la fin de la salve
Durée (duration)	<i>s</i>	Temps qui s'écoule entre le déclenchement et la fin de la salve
Energie (energy)	<i>J</i>	Intégrale de la tension du signal carré
Puissance (absoluteEnergy)	<i>J/s</i>	Intégrale de la tension du signal carré divisée par la durée de la salve
Amplitude	<i>V</i>	Plus forte amplitude observée
Angle de montée	°	Angle formé par l'axe $y = 0$ et la droite reliant l'origine à l'Amplitude [43]
Amplitude moyenne (meanAmplitude)	<i>V</i>	Amplitude moyenne observée
Nombre de coups (count)	-	Nombre de franchissements de seuil par la salve sur toute sa durée
Nombre de coups à la montée (count2Rise)	-	Nombre de coups pendant la montée
Nombre de coups à la décroissance (count2Fall)	-	Nombre de coups pendant la descente
shapeRatio	-	Rapport de l'aire sous la courbe du signal positif par son aire sous la courbe s'il formait un rectangle

Tableau A.1. – Paramètres EA liés au signal temporel

Paramètre	Unité	Définition
Fréquence moyenne (mean-Frequency)	kHz	Fréquence moyenne observée
Étendue spectrale	kHz	Écart-type de la distribution fréquentielle
Fréquence à la montée (rise-Freq)	kHz	Rapport entre le nombre de coups au pic et le temps de montée
Fréquence à la décroissance (fallFreq)	kHz	Rapport entre le nombre de coups à la décroissance et la durée de décroissance
Coefficient de dissymétrie	-	Distribution des fréquences par rapport au barycentre
shapeRatioSpec	-	Rapport de l'aire sous la courbe du spectre par son aire sous la courbe s'il formait un rectangle
Centroïde du spectre (spec-Cen)	kHz	Fréquence associée au centre de gravité de l'aire sous la courbe du spectre
Maximum du spectre (maxSpec)	$V \times V/kHz$	Amplitude maximale du spectre
Fréquence du maximum (maxFreqSpec)	kHz	Fréquence associée au maximum d'amplitude du spectre

Tableau A.2. – Paramètres EA liés au spectre du signal

Paramètre	Définition
skewness	coefficient d'asymétrie
kurt	kurtosis normalisé : kurtosis -3 : Coefficient d'aplatissement
ecTy	Écart-type sur le signal positif
coefVar	Coefficient de variation sur le signal positif
statAD	Statistique du test d'Anderson-Darling (*)

Tableau A.3. – Paramètres EA à connotation statistique

Notons qu'on peut également extraire des paramètres EA de la fonction d'auto-corrélation de la salve [176].

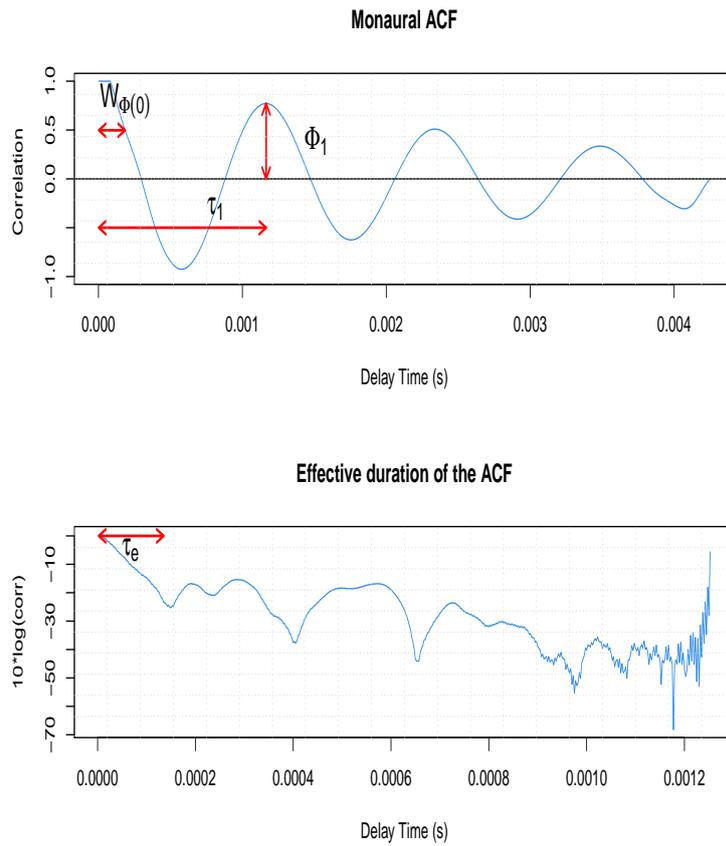


Figure A.1. – Illustration du calcul de quelques paramètres EA obtenus à partir de la fonction d'auto-corrélation.

Annexe B.

B.1. Résolution de l'équation d'onde en coordonnées cylindriques par la méthode de séparation des variables

Une approche classique de résolution de l'équation d'onde en coordonnées cylindriques est la méthode de séparation de variables. Pour une équation en ζ , elle consiste à trouver les solutions de la forme :

$$\zeta(t, r, \theta, z) = \zeta_t(t)\zeta_r(r)\zeta_\theta(\theta)\zeta_z(z) \quad (\text{B.1})$$

Sachant que $\Delta\zeta = \frac{\partial^2\zeta}{\partial r^2} + \frac{1}{r}\frac{\partial\zeta}{\partial r} + \frac{1}{r^2}\frac{\partial^2\zeta}{\partial\theta^2} + \frac{\partial^2\zeta}{\partial z^2}$, l'équation d'onde en coordonnées cylindriques s'écrit :

$$-\frac{1}{c^2}\frac{\partial^2\zeta}{\partial t^2} + \frac{\partial^2\zeta}{\partial r^2} + \frac{1}{r}\frac{\partial\zeta}{\partial r} + \frac{1}{r^2}\frac{\partial^2\zeta}{\partial\theta^2} + \frac{\partial^2\zeta}{\partial z^2} = 0 \quad (\text{B.2})$$

où c est la vitesse de propagation des ondes dans le milieu considéré. En remplaçant ζ par sa décomposition (B.1), l'équation (B.2) est équivalente à :

$$\frac{1}{\zeta_r}\left(\frac{\partial^2\zeta_r}{\partial r^2} + \frac{1}{r}\frac{\partial\zeta_r}{\partial r}\right) + \frac{1}{\zeta_\theta}\frac{1}{r^2}\frac{\partial^2\zeta_\theta}{\partial\theta^2} + \frac{1}{\zeta_z}\frac{\partial^2\zeta}{\partial z^2} - \frac{1}{\zeta_t}\frac{1}{c^2}\frac{\partial^2\zeta}{\partial t^2} = 0 \quad (\text{B.3})$$

Alors, chacun des termes de l'équation (B.3) dépendant des composantes $\zeta_t, \zeta_r, \zeta_\theta$ et ζ_z doivent être égaux à une constante. Ainsi, ζ_t satisfait :

$$\frac{\partial^2\zeta_t}{\partial t^2} + k^2 c^2 \zeta_t = 0 \quad \Leftrightarrow \quad \frac{\partial^2\zeta_t}{\partial t^2} + \omega^2 \zeta_t = 0 \quad (\text{B.4})$$

où $k = \frac{\omega}{c}$. Une solution de l'équation (B.4) est $\zeta_t(t) = e^{-i\omega t}$.

De même, ζ_z satisfait :

$$\frac{\partial^2\zeta_z}{\partial z^2} - k_z^2 \zeta_z = 0 \quad (\text{B.5})$$

dont les solutions de la forme $\zeta_z = e^{ik_z z}$.

De la même manière, pour l'équation en ζ_θ on peut poser :

$$\frac{\partial^2 \zeta_\theta}{\partial \theta^2} + n^2 \zeta_\theta = 0 \quad (\text{B.6})$$

dont les solutions sont de la forme $\zeta_\theta = \cos(n\theta)$.

En remplaçant ζ_t , ζ_z et ζ_θ par leurs valeurs, l'équation (B.3) devient :

$$\frac{\partial^2 \zeta_r}{\partial r^2} + \frac{1}{r} \frac{\partial \zeta_r}{\partial r} + \left(\alpha^2 - \frac{n^2}{r^2} \right) \zeta_r = 0 \quad (\text{B.7})$$

où $\alpha^2 = k^2 - k_z^2$. L'équation (B.7) est une équation de Bessel du premier ordre dont la solution générale s'exprime sous forme de fonctions de Bessel si αr est réel ou de fonctions de Bessel modifiées si αr est imaginaire.

$$\zeta_r(r) = A_1 Z_n(\alpha r) + A_2 W_n(\alpha r) \quad (\text{B.8})$$

Ainsi, une solution de l'équation d'onde en coordonnées cylindriques est de la forme :

$$\zeta = [A_1 Z_n(\alpha r) + A_2 W_n(\alpha r)] \cos n\theta e^{i(k_z z - \omega t)} \quad (\text{B.9})$$

B.2. Conditions aux limites relatives à la résolution de l'équation d'onde dans le cas du dispositif simplifié

La loi d'Hooke généralisée s'écrit :

$$\sigma_{rr} = \lambda \Delta \phi + 2\mu \epsilon_{rr}, \quad \sigma_{r\theta} = 2\mu \epsilon_{r\theta} \quad (\text{B.10})$$

L'application des conditions de continuité aux interfaces (Uranium/Sodium, Sodium/Zircaloy) et de la condition de surface libre à l'interface (Zircaloy/vide) nécessite le calcul des contraintes σ_{rr} et $\sigma_{r\theta}$ pour les milieux solides (Uranium et Zircaloy). Puisque

$$\epsilon_{rr} = \frac{\partial u_r}{\partial r}, \quad \epsilon_{r\theta} = \frac{1}{2} \left[r \frac{\partial}{\partial r} \left(\frac{u_\theta}{r} \right) + \frac{1}{r} \frac{\partial u_r}{\partial \theta} \right] \quad (\text{B.11})$$

nous avons alors :

$$\begin{aligned} \sigma_{rr} = & [A_1 (-\lambda(\alpha^2 + k^2) Z_n(\alpha r) + 2\mu \alpha^2 Z_n''(\alpha r)) + A_2 (-\lambda(\alpha^2 + k^2) W_n(\alpha r) + 2\mu \alpha^2 W_n''(\alpha r))] \\ & + D_1 \left(\frac{n}{r} \alpha Z_n'(\beta r) - \frac{n}{r^2} Z_n(\beta r) \right) + D_2 \left(\frac{n}{r} \alpha W_n'(\beta r) - \frac{n}{r^2} W_n(\beta r) \right) \cos(n\theta) \end{aligned} \quad (\text{B.12})$$

et

$$\begin{aligned} \sigma_{r\theta} = & \left[A_1 \left[2n \mu \left(\frac{1}{r^2} Z_n(\alpha r) - Z_n'(\alpha r) \right) \right] + A_2 \left[2n \mu \left(\frac{1}{r^2} W_n(\alpha r) - W_n'(\alpha r) \right) \right] \right] \\ & + D_1 [\beta^2 (Z_n(\beta r) - 2Z_n''(\beta r))] + D_2 [\beta^2 (W_n(\beta r) - 2W_n''(\beta r))] \sin(n\theta) \end{aligned} \quad (\text{B.13})$$

Notons que dans le cas du noyau d'Uranium, puisque $W_n = 0$, les équations (B.13) et (B.14) se simplifient (cf. Equation (2.8)).

a. Conditions aux limites à l'interface Uranium/Sodium

Rappelons que l'indice du noyau d'Uranium est 1 et que celui du fluide caloporteur Sodium est 2. Sachant que la pression dans le Sodium est liée au potentiel scalaire par $p = \rho_2 \frac{\partial^2 \phi_2}{\partial t^2}$, les conditions de continuité (2.12) s'écrivent :

$$A_{11} \left(-\lambda_1 (\alpha_1^2 + k_1^2) Z_n(\alpha_1 r_1) + 2\mu_1 \alpha_1^2 Z_n''(\alpha_1 r_1) \right) + D_{11} \left(\frac{n}{r_1} \alpha_1 Z_n'(\beta_1 r_1) - \frac{n}{r_1^2} Z_n(\beta_1 r_1) \right) + E_{21} (\rho_2 \omega Z_n(\alpha_2 r_1)) + E_{22} (\rho_2 \omega W_n(\alpha_2 r_1)) = 0 \quad (\text{B.14})$$

$$A_{11} \alpha_1 Z_n'(\alpha_1 r_1) + D_{11} \frac{n}{r_1} Z_n(\beta_1 r_1) + E_{21} \alpha_2 Z_n'(\alpha_2 r_1) + E_{22} \alpha_2 W_n'(\alpha_2 r_1) = 0 \quad (\text{B.15})$$

$$A_{11} 2n \mu_1 \left(\frac{1}{r_1^2} Z_n(\alpha_1 r_1) - Z_n'(\alpha_1 r_1) \right) + D_{11} \beta_1^2 \left(Z_n(\beta_1 r_1) - 2Z_n''(\beta_1 r_1) \right) = 0 \quad (\text{B.16})$$

b. Conditions aux limites à l'interface Sodium/Zircaloy

L'indice de la couche externe de Zircaloy étant 3, les conditions de continuité (2.12) s'écrivent :

$$E_{21} \rho_2 \omega Z_n(\alpha_2 r_2) + E_{22} \rho_2 \omega W_n(\alpha_2 r_2) + A_{31} \left(-\lambda_3 (\alpha_3^2 + k_3^2) Z_n(\alpha_3 r_2) + 2\mu_3 \alpha_3^2 Z_n''(\alpha_3 r_2) \right) + A_{32} \left(-\lambda_3 (\alpha_3^2 + k_3^2) W_n(\alpha_3 r_2) + 2\mu_3 \alpha_3^2 W_n''(\alpha_3 r_2) \right) + D_{31} \left(\frac{n}{r_2} \alpha_3 Z_n'(\beta_3 r_2) - \frac{n}{r_2^2} Z_n(\beta_3 r_2) \right) + D_{32} \left(\frac{n}{r_2} \alpha_3 W_n'(\beta_3 r_2) - \frac{n}{r_2^2} W_n(\beta_3 r_2) \right) = 0 \quad (\text{B.17})$$

$$E_{21} \alpha_2 Z_n'(\alpha_2 r_2) + E_{22} \alpha_2 W_n'(\alpha_2 r_2) + A_{31} \alpha_3 Z_n'(\alpha_3 r_2) + A_{32} \alpha_3 W_n'(\alpha_3 r_2) + D_{31} \frac{n}{r} Z_n(\beta_3 r_2) + D_{32} \frac{n}{r} W_n(\beta_3 r_2) = 0 \quad (\text{B.18})$$

$$A_{31} 2n \mu_3 \left(\frac{1}{r_2^2} Z_n(\alpha_3 r_2) - Z_n'(\alpha_3 r_2) \right) + A_{32} 2n \mu_3 \left(\frac{1}{r_2^2} W_n(\alpha_3 r_2) - W_n'(\alpha_3 r_2) \right) + D_{31} \beta_3^2 \left(Z_n(\beta_3 r_2) - 2Z_n''(\beta_3 r_2) \right) + D_{32} \beta_3^2 \left(W_n(\beta_3 r_2) - 2W_n''(\beta_3 r_2) \right) = 0 \quad (\text{B.19})$$

c. Condition de surface libre à l'interface Zircaloy/vide

Les conditions (2.13) s'écrivent :

$$A_{31} \left(-\lambda_3 (\alpha_3^2 + k_3^2) Z_n(\alpha_3 r_3) + 2\mu_3 \alpha_3^2 Z_n''(\alpha_3 r_3) \right) + A_{32} \left(-\lambda_3 (\alpha_3^2 + k_3^2) W_n(\alpha_3 r_3) + 2\mu_3 \alpha_3^2 W_n''(\alpha_3 r_3) \right) + D_{31} \left(\frac{n}{r_3} \alpha_3 Z_n'(\beta_3 r_3) - \frac{n}{r_3^2} Z_n(\beta_3 r_3) \right) + D_{32} \left(\frac{n}{r_3} \alpha_3 W_n'(\beta_3 r_3) - \frac{n}{r_3^2} W_n(\beta_3 r_3) \right) = 0 \quad (\text{B.20})$$

$$A_{31} 2n \mu_3 \left(\frac{1}{r_3^2} Z_n(\alpha_3 r_3) - Z_n'(\alpha_3 r_3) \right) + A_{32} 2n \mu_3 \left(\frac{1}{r_3^2} W_n(\alpha_3 r_3) - W_n'(\alpha_3 r_3) \right) + D_{31} \beta_3^2 \left(Z_n(\beta_3 r_3) - 2Z_n''(\beta_3 r_3) \right) + D_{32} \beta_3^2 \left(W_n(\beta_3 r_3) - 2W_n''(\beta_3 r_3) \right) = 0 \quad (\text{B.21})$$

B.3. Coefficients de la matrice des fréquences

$$\begin{aligned}
a_{11} &= -\lambda_1(\alpha_1^2 + k_1^2)Z_n(\alpha_1 r_1) + 2\mu_1 \alpha_1^2 Z_n''(\alpha_1 r_1) \\
a_{12} &= \frac{n}{r_1} \alpha_1 Z_n'(\beta_1 r_1) - \frac{n}{r_1^2} Z_n(\beta_1 r_1) \\
a_{13} &= \rho_2 \omega Z_n(\alpha_2 r_1) \\
a_{14} &= \rho_2 \omega W_n(\alpha_2 r_1) \\
a_{15} &= a_{16} = a_{17} = a_{18} = 0 \\
a_{21} &= \alpha_1 Z_n'(\alpha_1 r_1) \\
a_{22} &= \frac{n}{r_1} Z_n(\beta_1 r_1) \\
a_{23} &= \alpha_2 Z_n'(\alpha_2 r_1) \\
a_{24} &= \alpha_2 W_n'(\alpha_2 r_1) \\
a_{25} &= a_{26} = a_{27} = a_{28} = 0 \\
a_{31} &= 2n \mu_1 \left(\frac{1}{r_1^2} Z_n(\alpha_1 r) - Z_n'(\alpha_1 r_1) \right) \\
a_{32} &= \beta_1^2 (Z_n(\beta_1 r_1) - 2Z_n''(\beta_1 r_1)) \\
a_{33} &= a_{34} = a_{35} = a_{36} = a_{37} = a_{38} = 0 \\
a_{41} &= a_{42} = 0 \\
a_{43} &= \rho_2 \omega Z_n(\alpha_2 r_2) \\
a_{44} &= \rho_2 \omega W_n(\alpha_2 r_2) \\
a_{45} &= -\lambda_3(\alpha_3^2 + k_3^2)Z_n(\alpha_3 r_2) + 2\mu_3 \alpha_3^2 Z_n''(\alpha_3 r_2) \\
a_{46} &= -\lambda_3(\alpha_3^2 + k_3^2)W_n(\alpha_3 r_2) + 2\mu_3 \alpha_3^2 W_n''(\alpha_3 r_2) \\
a_{47} &= \frac{n}{r_2} \alpha_3 Z_n'(\beta_3 r_2) - \frac{n}{r_2^2} Z_n(\beta_3 r_2) \\
a_{48} &= \frac{n}{r_2} \alpha_3 W_n'(\beta_3 r_2) - \frac{n}{r_2^2} W_n(\beta_3 r_2)
\end{aligned}$$

$$\begin{aligned}
a_{51} &= a_{52} = 0 \\
a_{53} &= \alpha_2 Z_n'(\alpha_2 r_2) \\
a_{54} &= \alpha_2 W_n'(\alpha_2 r_2) \\
a_{55} &= \alpha_3 Z_n'(\alpha_3 r_2) \\
a_{56} &= \alpha_3 W_n'(\alpha_3 r_2) \\
a_{57} &= \frac{n}{r} Z_n(\beta_3 r_2) \\
a_{58} &= \frac{n}{r} W_n(\beta_3 r_2) \\
a_{61} &= a_{62} = a_{63} = a_{64} = 0 \\
a_{65} &= 2n \mu_3 \left(\frac{1}{r_2^2} Z_n(\alpha_3 r_2) - Z_n'(\alpha_3 r_2) \right) \\
a_{66} &= 2n \mu_3 \left(\frac{1}{r_2^2} W_n(\alpha_3 r_2) - W_n'(\alpha_3 r_2) \right) \\
a_{67} &= \beta_3^2 (Z_n(\beta_3 r_2) - 2Z_n''(\beta_3 r_2)) \\
a_{68} &= \beta_3^2 (W_n(\beta_3 r_2) - 2W_n''(\beta_3 r_2)) \\
a_{71} &= a_{72} = a_{73} = a_{74} = 0 \\
a_{75} &= (-\lambda_3(\alpha_3^2 + k_3^2)Z_n(\alpha_3 r_3) + 2\mu_3 \alpha_3^2 Z_n''(\alpha_3 r_3)) \\
a_{76} &= (-\lambda_3(\alpha_3^2 + k_3^2)W_n(\alpha_3 r_3) + 2\mu_3 \alpha_3^2 W_n''(\alpha_3 r_3)) \\
a_{77} &= \left(\frac{n}{r_3} \alpha_3 Z_n'(\beta_3 r_3) - \frac{n}{r_3^2} Z_n(\beta_3 r_3) \right) \\
a_{78} &= \left(\frac{n}{r_3} \alpha_3 W_n'(\beta_3 r_3) - \frac{n}{r_3^2} W_n(\beta_3 r_3) \right) \\
a_{81} &= a_{82} = a_{83} = a_{84} = 0 \\
a_{85} &= 2n \mu_3 \left(\frac{1}{r_3^2} Z_n(\alpha_3 r_3) - Z_n'(\alpha_3 r_3) \right) \\
a_{86} &= 2n \mu_3 \left(\frac{1}{r_3^2} W_n(\alpha_3 r_3) - W_n'(\alpha_3 r_3) \right) \\
a_{87} &= \beta_3^2 (Z_n(\beta_3 r_3) - 2Z_n''(\beta_3 r_3)) \\
a_{88} &= \beta_3^2 (W_n(\beta_3 r_3) - 2W_n''(\beta_3 r_3))
\end{aligned}$$

B.4. Brève présentation des fonctions de Bessel

Les fonctions de Bessel sont particulièrement importantes lorsqu'on étudie la propagation d'ondes dans un milieu à géométrie cylindrique ou sphérique. En effet, dans ces cas, la résolution de l'équation d'ondes par la méthode très fréquemment utilisée de la séparation des variables demande une résolution de l'équation dite de Bessel. Introduite par le mathématicien Daniel Bernoulli, puis formalisée par Friedrich Bessel, cette équation est de la forme :

$$r^2 \frac{\partial^2 y}{\partial r^2} + r \frac{\partial y}{\partial r} + (r^2 - n^2)y = 0 \quad (\text{B.22})$$

Les solutions y_n de cette équation sont appelées fonctions de Bessel, où le paramètre n pouvant être un nombre complexe est appelé ordre de la fonction. En géométrie cylindrique, n est entier, la fonction de Bessel prend alors le nom de fonction cylindrique ou harmonique cylindrique dans certaines références de la littérature.

L'équation de Bessel étant une équation différentielle du second ordre, elle admet deux solutions

linéairement indépendantes appelées fonctions de première et deuxième espèce. Selon le contexte, plusieurs formulations peuvent être utilisées pour exprimer ces solutions. On distingue les fonctions de Bessel, les fonctions de Bessel modifiées, les fonctions de Hankel, les fonctions sphériques de Bessel et les fonctions sphériques de Hankel.

B.4.0.1. Fonctions de Bessel de première espèce

Les fonctions de Bessel de première espèce $J_n(x)$ se caractérisent essentiellement par leur comportement au voisinage de l'axe (Oz) du cylindre (c'est-à-dire lorsque le rayon est nul). Elles sont finies lorsque $r = 0$ pour les valeurs entières ou positives de n et elles divergent lorsque r tend vers zéro pour les valeurs négatives de n . Au voisinage de l'axe (Oz) du cylindre, les fonctions de Bessel de première espèce s'expriment comme suit :

$$J_n(r) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k! \Gamma(k+n+1)} \left(\frac{r}{2}\right)^{2k+n} \quad (\text{B.23})$$

où Γ est une fonction de gamma. Lorsque n entier, les fonctions de Bessel sont des fonctions entières. Dans le cas contraire, elles sont multivaluées avec des singularités en zéro.

B.4.0.2. Fonctions de Bessel de deuxième espèce

Les fonctions de Bessel de deuxième espèce notées Y_n sont des solutions de l'équation de Bessel présentant une singularité en $r = 0$ et multivaluées. Dans les cas où n est entier, elle s'écrivent comme comme suit :

$$Y_n(r) = \lim_{\alpha \rightarrow n} Y_\alpha(r) \quad (\text{B.24})$$

où α est un réel et :

$$Y_\alpha(r) = \frac{J_\alpha(r) \cos(\alpha\pi) - J_{-\alpha}(r)}{\sin(\alpha\pi)} \quad (\text{B.25})$$

Annexe C.

C.1. Impact des méthodes de débruitage sur des paramètres EA choisis

Tableau C.1. – *Quelques paramètres EA classiques associés aux six salves présentées en Figure (3.11a) recouverts à la suite d'un débruitage par SS dans le cas d'un bruit gaussien.*

	hit_1	hit_2	hit_3	hit_4	hit_5	hit_6
Envelopes correlation	0.59	0.99	0.96	0.85	0.94	0.97
Kullback-Leiber distance	8.46	0.44	2.26	1.17	1.69	0.88
Absolute energy (MV)	4.25	696.55	15.37	1439.10	19.29	51.80
Amplitude (V)	-7.39	21.35	3.58	18.79	5.93	18.68
Count	3317.00	1843.00	3136.00	250.00	608.00	5687.00
Mean frequency (kHz)	127.58	102.33	116.07	147.93	121.60	129.24
Rise Time (ms)	8.24	1.25	18.31	0.32	1.27	5.82
Count to pic	1071.00	148.00	2064.00	53.00	157.00	788.00

Tableau C.2. – *Quelques paramètres EA classiques associés aux six salves présentées en Figure (3.11a) recouverts à la suite d'un débruitage utilisant un seuillage par ondelettes dans le cas d'un bruit gaussien.*

	hit_1	hit_2	hit_3	hit_4	hit_5	hit_6
Envelopes correlation	0.83	0.99	0.89	0.99	0.87	0.94
Kullback-Leiber distance	1.43	0.15	1.01	0.14	0.46	0.53
Absolute energy (MV)	0.03	585.47	2.40	1787.42	5.24	44.15
Amplitude (V)	-10.95	22.77	3.46	17.23	4.69	19.66
Count	128.00	475.00	384.00	235.00	34.00	333.00
Mean frequency (kHz)	4.92	26.37	14.21	139.05	6.80	7.57
Rise Time (ms)	5.64	1.20	16.62	0.34	1.27	6.04
Count to pic	44.00	14.00	190.00	54.00	5.00	56.00

Tableau C.3. – Quelques paramètres EA classiques associés aux six salves présentées en Figure (3.11a) recouverts à la suite d'un débruitage par SSA dans le cas d'un bruit gaussien.

	hit_1	hit_2	hit_3	hit_4	hit_5	hit_6
Envelopes correlation	0.93	1.00	0.94	0.71	0.96	0.98
Kullback-Leiber distance	2.42	1.30	0.78	3.05	2.36	4.23
Absolute energy (MV)	19.73	1103.87	73.60	873.82	59.61	91.26
Amplitude (V)	2.75	20.88	10.01	14.16	7.46	17.76
Count	2775.00	1443.00	2585.00	259.00	543.00	4740.00
Mean frequency (kHz)	106.73	80.12	95.68	153.25	108.60	107.72
Rise Time (ms)	5.64	2.21	18.31	0.40	1.29	6.04
Count to pic	605.00	182.00	1774.00	61.00	144.00	678.00

Tableau C.4. – Quelques paramètres EA classiques associés aux six salves présentées en Figure (3.11a) recouverts à la suite d'un débruitage par SS dans le cas d'un bruit électrique.

	hit_1	hit_2	hit_3	hit_4	hit_5	hit_6
Envelopes correlation	1.00	1.00	1.00	0.72	1.00	0.99
Kullback-Leiber distance	0.17	0.27	0.11	1.05	0.06	0.76
Absolute energy (MV)	9.52	1210.93	51.50	1021.68	26.15	89.67
Amplitude (V)	2.36	23.34	10.53	15.92	8.42	21.65
Count	2098.00	885.00	1599.00	242.00	505.00	5004.00
Mean frequency (kHz)	80.69	49.14	59.18	143.20	101.00	113.72
Rise Time (ms)	5.64	1.20	16.62	0.72	1.27	5.82
Count to pic	289.00	92.00	942.00	108.00	63.00	952.00

Tableau C.5. – Quelques paramètres EA classiques associés aux six salves présentées en Figure (3.11a) recouverts à la suite d'un débruitage utilisant un seuillage par ondelettes dans le cas d'un bruit électrique.

	hit_1	hit_2	hit_3	hit_4	hit_5	hit_6
Envelopes correlation	0.89	1.00	0.99	0.99	0.73	1.00
Kullback-Leiber distance	14.20	0.17	6.32	0.03	11.16	1.47
Absolute energy (MV)	122.04	1538.93	166.45	4136.92	141.59	270.79
Amplitude (V)	12.62	23.85	13.99	19.84	13.04	22.46
Count	1571.00	971.00	1515.00	250.00	309.00	4455.00
Mean frequency (kHz)	60.42	53.91	56.07	147.93	61.80	101.24
Rise Time (ms)	4.43	1.20	17.43	0.34	1.22	5.82
Count to pic	242.00	64.00	976.00	58.00	69.00	439.00

Tableau C.6. – Quelques paramètres EA classiques associés aux six salves présentées en Figure (3.11a) recouverts à la suite d'un débruitage par SSA dans le cas d'un bruit électrique.

	hit_1	hit_2	hit_3	hit_4	hit_5	hit_6
Envelopes correlation	0.94	1.00	0.99	0.73	0.94	0.99
Kullback-Leiber distance	5.52	1.13	1.33	3.27	3.89	3.78
Absolute energy (MV)	12.23	1004.00	57.65	285.99	36.46	67.12
Amplitude (V)	3.70	20.78	10.23	10.95	6.76	16.93
Count	2936.00	1618.00	2716.00	222.00	581.00	5235.00
Mean frequency (kHz)	112.92	89.84	100.53	131.36	116.20	118.97
Rise Time (ms)	8.58	1.20	18.31	0.26	1.27	6.04
Count to pic	969.00	121.00	1807.00	34.00	146.00	707.00

Tableau C.7. – Quelques paramètres EA classiques associés aux six salves présentées en Figure (3.11a) recouverts à la suite d'un débruitage par SS dans le cas d'un bruit mixte.

	hit_1	hit_2	hit_3	hit_4	hit_5	hit_6
Envelopes correlation	0.92	1.00	0.98	0.84	0.96	0.98
Kullback-Leiber distance	2.19	0.35	0.95	1.17	0.33	0.82
Absolute energy (MV)	0.70	886.31	14.32	1471.88	15.86	55.40
Amplitude (V)	-7.23	21.97	5.53	18.34	6.97	19.76
Count	2542.00	1199.00	2188.00	244.00	428.00	5131.00
Mean frequency (kHz)	97.77	66.57	80.98	144.38	85.60	116.61
Rise Time (ms)	6.42	1.20	18.31	0.30	1.27	6.04
Count to pic	525.00	141.00	1506.00	45.00	105.00	685.00

Tableau C.8. – Quelques paramètres EA classiques associés aux six salves présentées en Figure (3.11a) recouverts à la suite d'un débruitage utilisant un seuillage par ondelettes dans le cas d'un bruit mixte.

	hit_1	hit_2	hit_3	hit_4	hit_5	hit_6
Envelopes correlation	0.04	0.98	0.69	0.99	0.80	0.92
Kullback-Leiber distance	25.93	0.23	14.68	0.16	8.29	1.46
Absolute energy (MV)	4.24	613.75	6.80	1979.55	9.07	50.51
Amplitude (V)	-7.75	22.47	5.37	18.29	5.98	19.63
Count	85.00	396.00	244.00	232.00	16.00	283.00
Mean frequency (kHz)	3.27	21.99	9.03	137.28	3.20	6.43
Rise Time (ms)	6.51	1.20	18.31	0.34	1.27	6.04
Count to pic	37.00	11.00	180.00	54.00	5.00	56.00

Tableau C.9. – Quelques paramètres EA classiques associés aux six salves présentées en Figure (3.11a) recouverts à la suite d'un débruitage par SSA dans le cas d'un bruit mixte.

	hit_1	hit_2	hit_3	hit_4	hit_5	hit_6
Envelopes correlation	0.95	0.99	0.98	0.83	0.93	0.97
Kullback-Leiber distance	3.70	1.62	1.40	2.76	2.48	4.95
Absolute energy (MV)	12.75	915.26	56.67	572.11	38.11	68.01
Amplitude (V)	2.96	20.06	9.14	14.44	7.72	16.48
Count	2886.00	1638.00	2749.00	246.00	568.00	5287.00
Mean frequency (kHz)	111.00	90.95	101.75	145.56	113.60	120.15
Rise Time (ms)	4.76	1.22	18.31	0.26	1.27	6.19
Count to pic	537.00	115.00	1837.00	37.00	139.00	748.00

Tableau C.10. – Quelques paramètres EA classiques associés aux six salves présentées en Figure (3.11a) recouverts à la suite d'un débruitage par SS dans le cas de la somme des bruits.

	hit_1	hit_2	hit_3	hit_4	hit_5	hit_6
Envelopes correlation	0.68	1.00	0.95	0.58	0.89	0.97
Kullback-Leiber distance	6.91	0.41	1.26	1.06	2.69	0.87
Absolute energy (MV)	1.25	756.56	13.39	717.59	3.41	41.96
Amplitude (V)	-8.67	22.08	4.17	12.68	-3.69	18.96
Count	3305.00	1617.00	2958.00	254.00	606.00	5567.00
Mean frequency (kHz)	127.12	89.78	109.48	150.30	121.20	126.52
Rise Time (ms)	4.89	1.20	18.31	0.72	1.27	6.04
Count to pic	629.00	128.00	1987.00	115.00	125.00	816.00

Tableau C.11. – Quelques paramètres EA classiques associés aux six salves présentées en Figure (3.11a) recouvrés à la suite d'un débruitage utilisant un seuillage par ondelettes dans le cas de la somme des bruits.

	hit_1	hit_2	hit_3	hit_4	hit_5	hit_6
Envelopes correlation	0.44	0.99	0.89	0.99	0.86	0.94
Kullback-Leiber distance	14.39	0.13	2.78	0.12	1.96	0.49
Absolute energy (MV)	1.42	645.63	4.15	1961.79	6.93	50.05
Amplitude (V)	-2.69	22.71	4.89	18.22	5.45	20.32
Count	511.00	603.00	726.00	240.00	110.00	934.00
Mean frequency (kHz)	19.65	33.48	26.87	142.01	22.00	21.23
Rise Time (ms)	12.97	1.20	18.31	0.34	1.27	6.04
Count to pic	292.00	31.00	492.00	56.00	34.00	145.00

Tableau C.12. – Quelques paramètres EA classiques associés aux six salves présentées en Figure (3.11a) recouvrés à la suite d'un débruitage par SSA dans le cas de la somme des bruits.

	hit_1	hit_2	hit_3	hit_4	hit_5	hit_6
Envelopes correlation	0.94	0.99	0.96	0.75	0.92	0.89
Kullback-Leiber distance	8.72	1.63	2.99	1.95	6.27	4.85
Absolute energy (MV)	18.24	1025.53	68.00	820.34	51.70	53.53
Amplitude (V)	3.37	20.48	9.54	13.35	7.16	17.01
Count	3418.00	1941.00	3371.00	262.00	681.00	5493.00
Mean frequency (kHz)	131.46	107.77	124.77	155.03	136.20	124.83
Rise Time (ms)	6.78	1.20	16.62	0.22	1.27	6.04
Count to pic	928.00	116.00	2070.00	39.00	164.00	826.00

Annexe D.

D.1. Résultats détaillés de la mise en œuvre de l'ACP

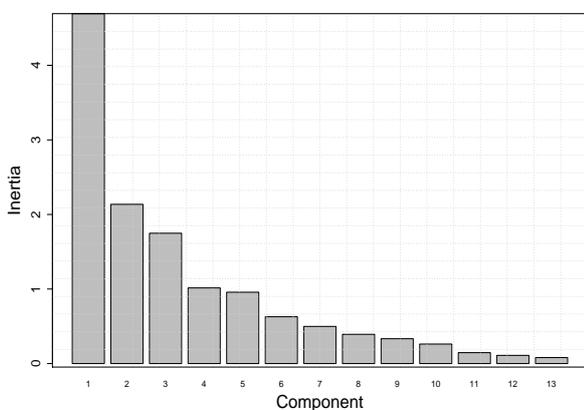
D.1.1. Considérations pratiques de mise en œuvre de l'ACP fonctionnelle

Une approche simple de mise en œuvre de l'ACP dans le cas fonctionnel est de se ramener à "un cas classique" en remplaçant les variables fonctionnelles par des variables discrètes sur une grille d'observation très fine. Si on note par n le nombre de points de discrétisation et par \mathbf{V} la matrice de variance-covariance de taille $n \times n$ de l'échantillon de courbes discrétisées, on est alors ramené à un problème classique de recherche de valeurs et vecteurs propres de la matrice \mathbf{V} , tel que :

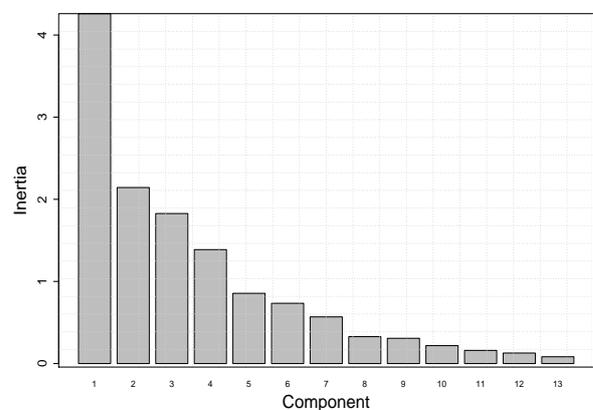
$$\mathbf{V}v = \lambda v \quad (\text{D.1})$$

D.1.2. Résultats obtenus

D.1.2.1. Cas classique



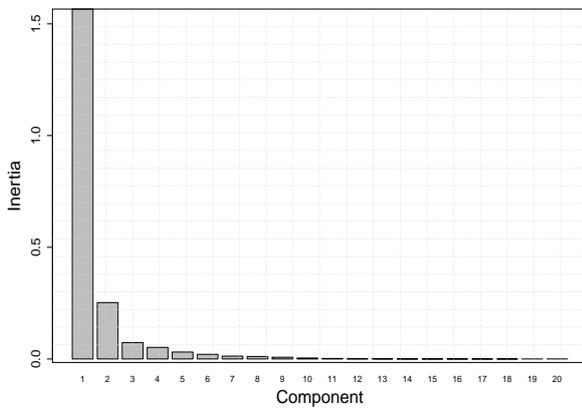
(a) Capteur M1



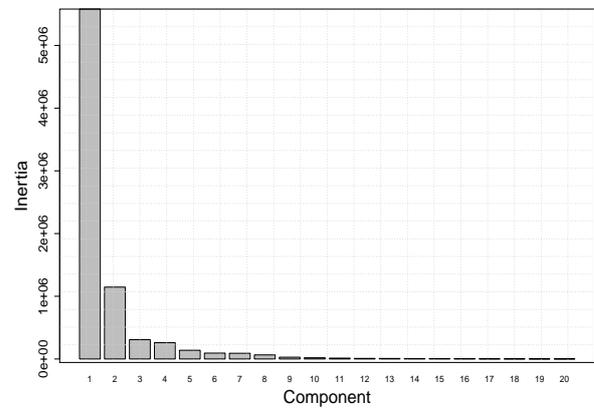
(b) Capteur M2

Figure D.1. – Éboulis de valeurs propres. ACP cas classique.

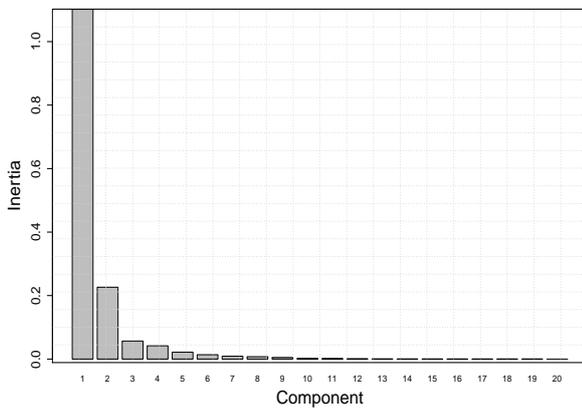
D.1.2.2. Cas fonctionnel



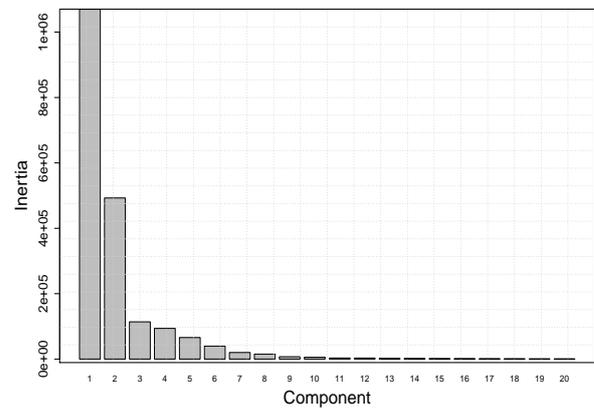
(a) Enveloppes capteur M1



(b) Spectres capteur M1



(c) Enveloppes capteur M2



(d) Spectres capteur M2

Figure D.2. – Éboulis de valeurs propres. ACP cas fonctionnel

D.2. Classification non supervisée

D.2.1. Informations complémentaires sur la mise en œuvre des méthodes

Tableau D.1. – Informations complémentaires relatives à la mise en œuvre du clustering selon les trois stratégies expérimentées dans le cas fonctionnel.

Libellé	Information complémentaire
CAH	le critère d'agrégation choisi est celui de Ward
Courbes enveloppes	la métrique de comparaison choisie est la métrique \mathcal{L}_2
Courbes spectres	la métrique de comparaison choisie est la métrique \mathcal{L}_2
Approche exclusivement fonctionnelle	$\alpha = 0.5$
Approche de réduction de dimension	la matrice de données choisie a 5 variables, dont 3 obtenues à la suite d'une ACP fonctionnelle sur les enveloppes et 2 à la suite d'une AFC sur les spectres

D.2.2. Description des classes obtenues

Tableau D.2. – Effectifs par classe

Classe	1	2	3	4	5	6	7	8
Effectif	7	26	12	13	28	30	44	6

Afin d'analyser les huit classes de salves obtenues (cf. Section 4.4.3.2), nous représentons ici les barplots correspondant au centre groupe après avoir centré et réduit la matrice de données (Figures D.3 à D.10). Nous testons également si la moyenne est égale à 0 avec le test de Student au seuil 5% de chaque coté (supérieur ou inférieur à 0) avec la correction de Bonferroni. Nous noterons cependant que ces tests sont donnés uniquement à titre indicatif. En effet, toutes les variables ne vérifient pas l'hypothèse de normalité. Les Tables D.3 et D.4 résument ces résultats. Lorsqu'il y a un "+" (respectivement "-"), la moyenne est significativement plus grande (respectivement plus petite) que la moyenne de l'échantillon. L'absence de signe indique que la différence de la moyenne par rapport à 0 n'est pas significative.

Par ailleurs, nous représentons sur les Figures D.11 et D.12 les superpositions d'enveloppes et spectres selon les classes. Nous noterons que pour des raisons de représentation, mises en évidence à la suite de l'ACP fonctionnelle, les spectres ont été normalisés.

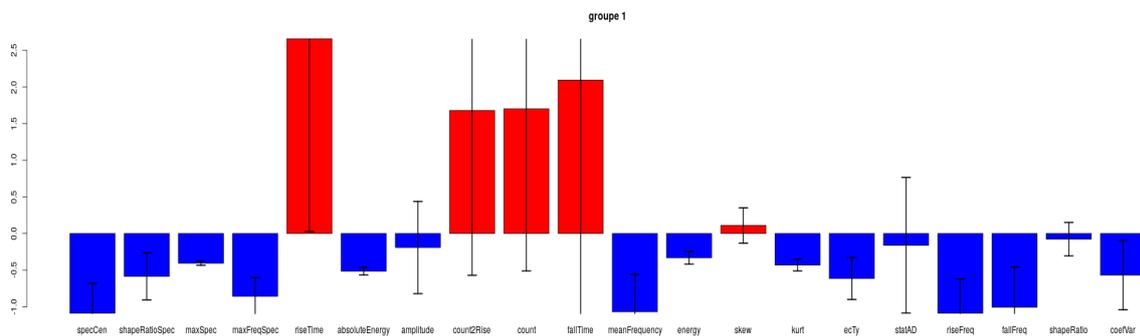


Figure D.3. – Centre de gravité du groupe 1 avec intervalles de confiance

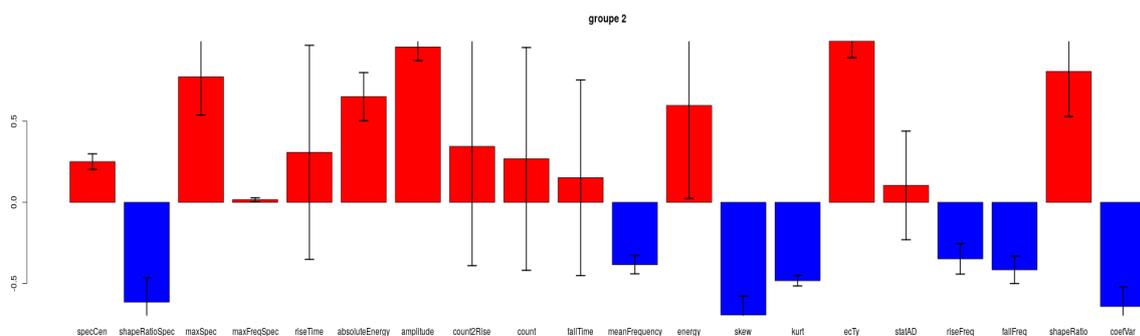


Figure D.4. – Centre de gravité du groupe 2 avec intervalles de confiance

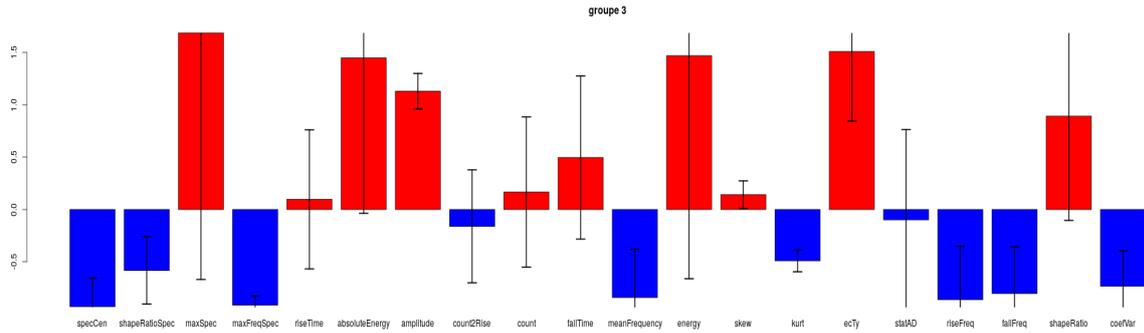


Figure D.5. – Centre de gravité du groupe 3 avec intervalles de confiance

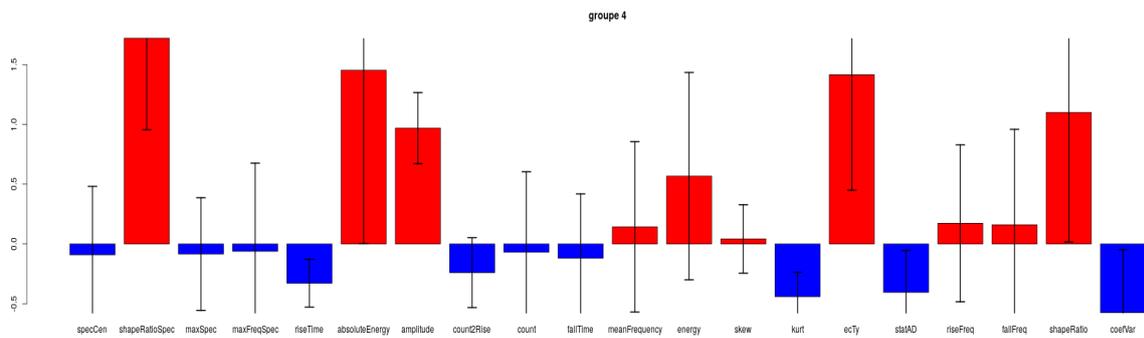


Figure D.6. – Centre de gravité du groupe 4 avec intervalles de confiance

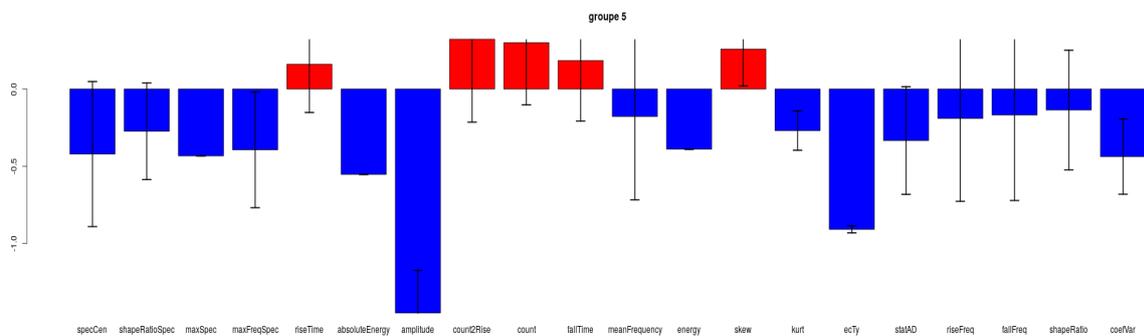


Figure D.7. – Centre de gravité du groupe 5 avec intervalles de confiance

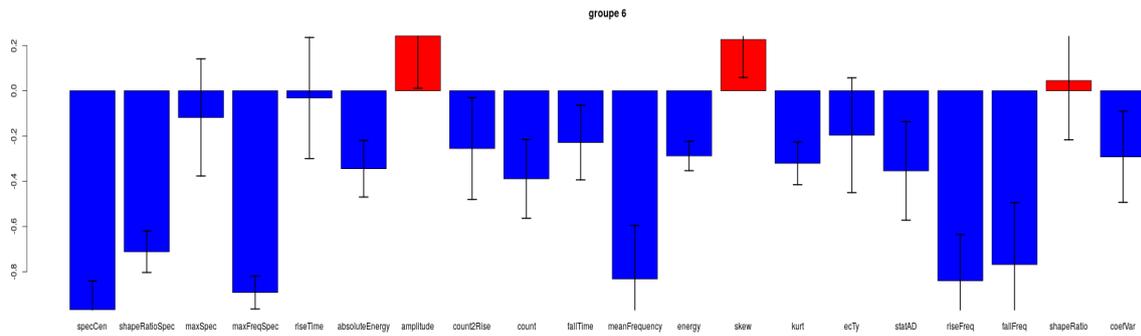


Figure D.8. – Centre de gravité du groupe 6 avec intervalles de confiance

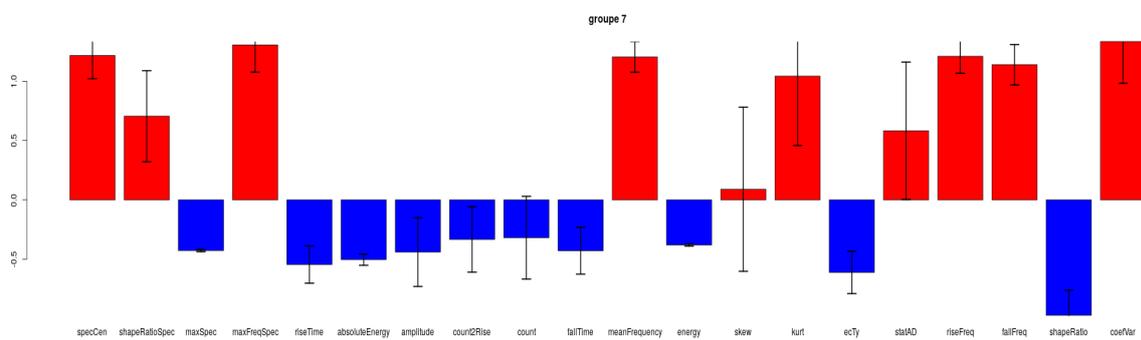


Figure D.9. – Centre de gravité du groupe 7 avec intervalles de confiance

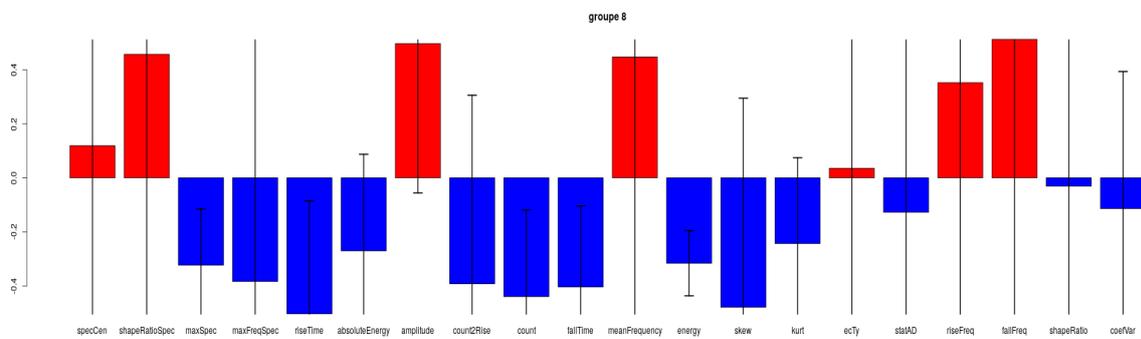


Figure D.10. – Centre de gravité du groupe 8 avec intervalles de confiance

Tableau D.3. – Caractérisation des classes . + : significativement plus grand que la moyenne. - : significativement plus petit que la moyenne

	specCen	shapeRatioSpec	maxSpec	maxFreqSpec	riseTime	power	amplitude	count2Rise	count	fallTime
groupe 1	-	-	-	-	+	-				
groupe 2	+	-	+	+		+	+			
groupe 3	-	-		-			+			
groupe 4		+			-		+			
groupe 5			-	-		-	-			
groupe 6	-	-		-		-	+	-	-	-
groupe 7	+	+	-	+	-	-	-	-	-	-
groupe 8			-		-				-	-

Tableau D.4. – Caractérisation des classes . + : significativement plus grand que la moyenne. - : significativement plus petit que la moyenne

	meanFrequency	energy	skew	kurt	ecTy	statAD	riseFreq	fallFreq	shapeRatio	coefVar
groupe 1	-	-		-	-		-	-		-
groupe 2	-	+	-	-	+		-	-	+	-
groupe 3	-		+	-	+		-	-		-
groupe 4				-	+	-			+	-
groupe 5		-	+	-	-					-
groupe 6	-	-	+	-		-	-	-		-
groupe 7	+	-		+	-		+	+	-	+
groupe 8		-								

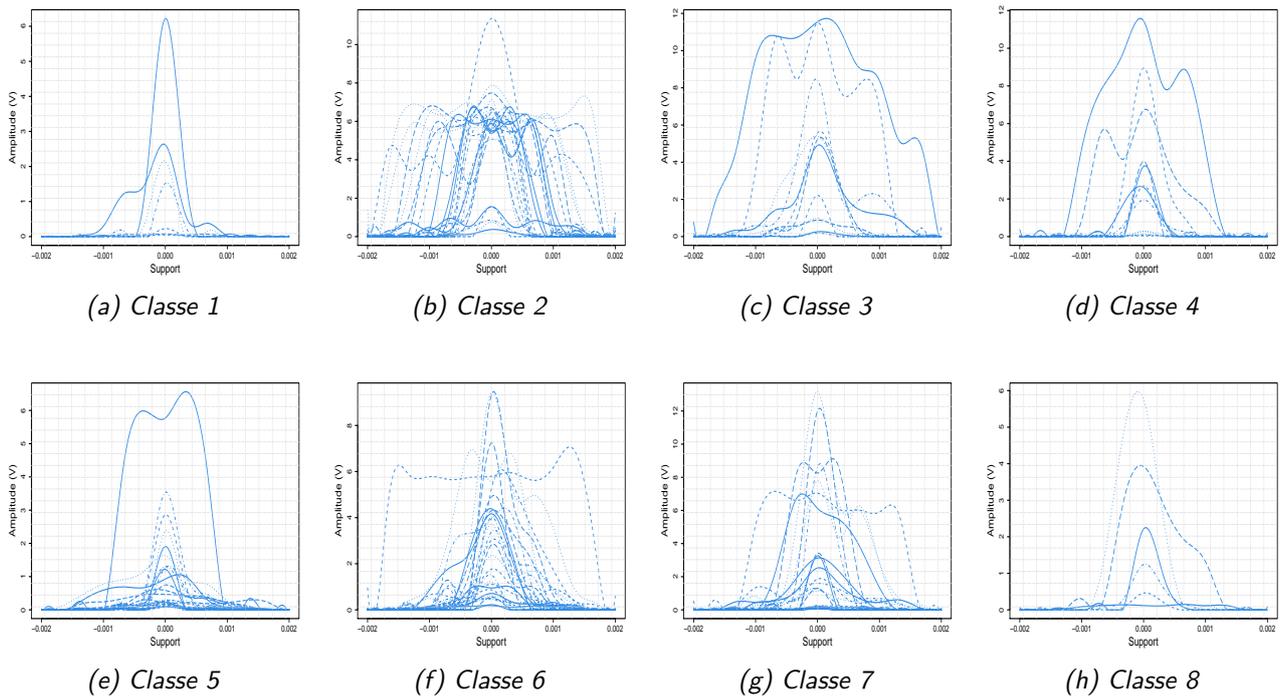


Figure D.11. – Représentations des superpositions d'enveloppes selon la classe

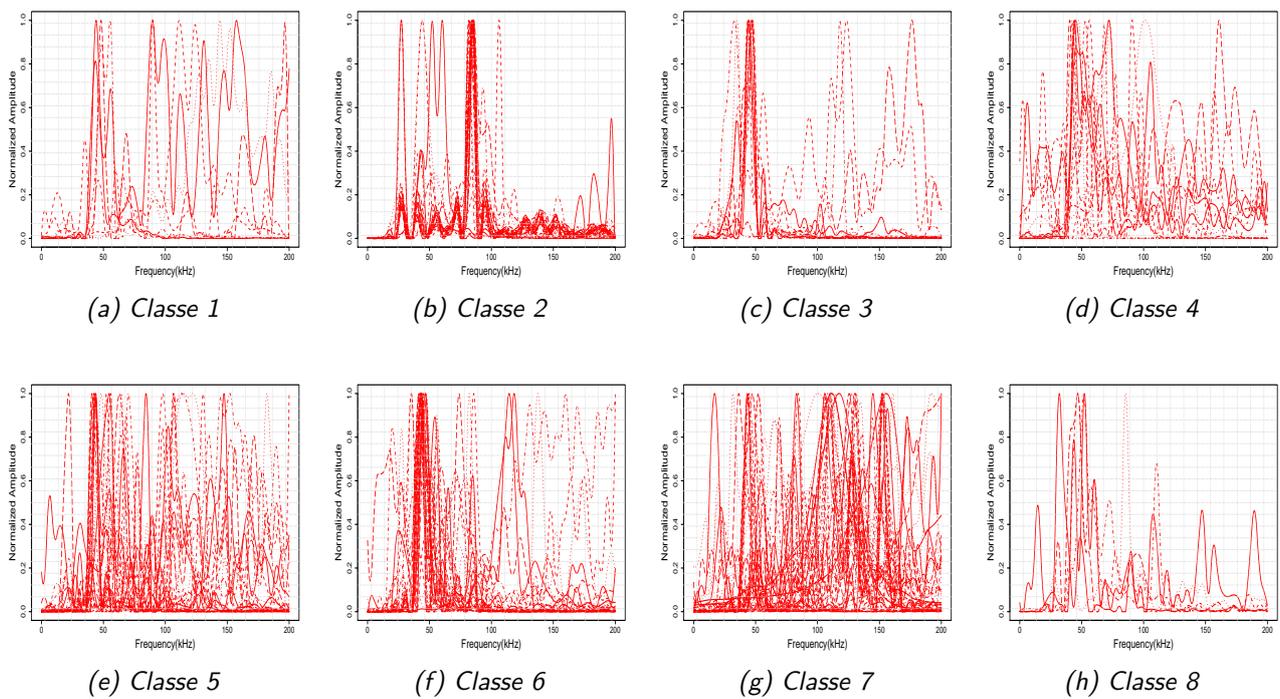


Figure D.12. – Représentations des superpositions de spectres normalisés selon la classe

D.3. Classification supervisée

D.3.1. Décomposition de la variance d'une matrice de données

Considérons le cas classique et la matrice de données de taille $N \times p$ présentée dans la section 4.3, sans perte de généralité et pour simplifier les notations, nous pouvons supposer que les N salves (individus) sont numérotées de telle sorte que les premiers individus soient ceux du groupe \mathcal{G}_1 , puis ceux du groupe \mathcal{G}_2 et ainsi de suite jusqu'au groupe \mathcal{G}_K . On a alors :

$$X = \begin{pmatrix} X(1) \\ X(2) \\ \vdots \\ X(K) \end{pmatrix} \quad \bar{X} = \begin{pmatrix} \bar{X}(1) \\ \bar{X}(2) \\ \vdots \\ \bar{X}(K) \end{pmatrix} \quad (\text{D.2})$$

où $X(k)$ désigne la sous-matrice des données associée au groupe \mathcal{G}_k .

Pour chacune des K matrices, le centre de gravité du groupe \mathcal{G}_k est défini par :

$$\mu_{\mathcal{G}_k} = \frac{1}{N_k} \sum_{i \in \mathcal{G}_k} x_i \quad (\text{D.3})$$

où N_k est le nombre de salves appartenant à la classe \mathcal{G}_k et $x_i = (x_i^1, x_i^2, \dots, x_i^p)^T \in \mathbb{R}^p$ est la ligne de X décrivant le i -ème individu.

La matrice de variance-covariance de taille $p \times p$ du groupe \mathcal{G}_k est définie par :

$$\Sigma_{\mathcal{G}_k} = \frac{1}{N_k} \sum_{i \in \mathcal{G}_k} (x_i - \mu_{\mathcal{G}_k})(x_i - \mu_{\mathcal{G}_k})^T \quad (\text{D.4})$$

La matrice de variance-covariance intra-groupes est définie par :

$$W = \frac{1}{N} \sum_{k=1}^K N_k \Sigma_{\mathcal{G}_k} \quad (\text{D.5})$$

La matrice de variance-covariance inter-groupes est définie par :

$$B = \frac{1}{N} \sum_{k=1}^K N_k (\mu_{\mathcal{G}_k} - \mu)(\mu_{\mathcal{G}_k} - \mu)^T \quad (\text{D.6})$$

où $\mu = \frac{1}{N} \sum_{i=1}^N x_i$. On a :

$$\Sigma = W + B \quad (\text{D.7})$$

où $\Sigma = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)(x_i - \mu)^T$

Notons que l'analyse discriminante est une ACP dont les composantes principales sont des combinaisons des variables de départ qui maximisent la variance inter-groupes, ou de manière équivalente, minimisent la variance intra-groupes (alors que dans l'ACP standard elles optimisent la variance totale).

Enfin, les prédicteurs de l'analyse discriminante décisionnelle sont basés sur la variance intra-groupes (cf. Section 4.5.2.2)

D.3.2. Notions de validation croisée

Réputée parmi les méthodes les plus simples pour l'estimation de l'erreur de prédiction [161, Chap.7], la K-Fold Cross-Validation consiste à diviser l'échantillon en $K \leq N$ sous-échantillons de tailles N_k sensiblement égales et à utiliser successivement chacune d'elles comme jeu de test d'un modèle créé à partir des $K - 1$ autres. L'erreur de prédiction associée au k -ième sous-échantillon s'écrit alors :

$$Error_k = \frac{1}{N_k} \sum_{i \in E_k} \mathbb{1}_{\hat{G}(x_i) \neq G(x_i)}, \quad (\text{D.8})$$

où E_k désigne l'ensemble des salves appartenant au k -ième sous-échantillon et $\hat{G}(x_i)$ l'estimation de la classe d'appartenance de l'individu x_i grâce au modèle créé à partir des $K - 1$ autres sous-échantillons. On a alors l'erreur globale du modèle de validation croisée :

$$CV = \sum_{k=1}^K \frac{N_k}{N} Error_k \quad (\text{D.9})$$

Lorsque $K = N$, on parle de validation croisée de type «leave-one-out».