Faculté Polytechnique de Mons

Service de Génie Mécanique-Pôle Risques

Académie universitaire Wallonie-Bruxelles

THÈSE

présentée pour l'obtention du grade de
Docteur en Sciences de l'Ingénieur,
le 10 juillet 2008

par

Bovic KILUNDU Y'Ebondo

Détection et diagnostic vibratoire de dégradations mécaniques : apport du *data mining*

Composition du jury :

Prof.	**O.**	**VERLINDEN**	:	Faculté Polytechnique de Mons	Président
Prof.	**B.**	**GOSSELIN**	:	Faculté Polytechnique de Mons	Secrétaire
Prof.	**P.**	**DEHOMBREUX**	:	Faculté Polytechnique de Mons	Promoteur
Dr	**A.**	**DERAEMAEKER**	:	Université Libre de Bruxelles	Membre
Prof.	**J.-P.**	**DRON**	:	Université de Reims	Membre
Prof.	**E.**	**FILIPPI**	:	Faculté Polytechnique de Mons	Membre
Prof.	**P.**	**FORTEMPS**	:	Faculté Polytechnique de Mons	Membre
Prof.	**K.**	**HARRI**	:	Ecole Royale Militaire	Membre
Dr	**P.**	**MACK**	:	PEPITe sa, Liège	Membre

Juillet 2008

A tous ceux qui me sont chers

A ma famille

Résumé

Cette thèse de doctorat s'est fixée comme objectif général d'explorer les possibilités offertes par les méthodes de *data mining* pour l'optimisation des processus de détection et de diagnostic vibratoire de dégradations mécaniques.

Après avoir circonscrit le contexte de l'étude et présenté les outils de base utilisés en détection et diagnostic vibratoire, la thèse aborde la question du prétraitement des signaux vibratoires en vue de l'extraction des indicateurs, et celle du traitement des ensembles de données constituées d'indicateurs dans le but d'en sélectionner les plus pertinents. Ensuite une description des méthodes d'apprentissage est présentée tout en soulignant l'importance de la prise en compte de coûts de classification.

Cette thèse propose, pour différents besoins d'analyse, des démarches essentiellement basées sur l'analyse en composantes principales (ACP) : l'analyse spectrale singulière pour le débruitage et le filtrage des signaux, la combinaison de l'ACP et des poids synaptiques d'un réseau de neurones pour la sélection des indicateurs pertinents et la construction des résidus ACP pour la détection précoce des dégradations et pour la construction de variables monotones croissantes utilisables dans le cadre d'un processus de pronostic.

Ces travaux démontrent l'applicabilité des techniques d'apprentissage et leur complémentarité à l'analyse vibratoire traditionnelle dans un contexte industriel. Les différentes démarches développées sont validées expérimentalement dans deux domaines techniques :

- la surveillance des machines tournantes : diagnostic de défauts multiples et détection précoce de défauts de roulement ;
- la surveillance de l'usure d'outils en usinage : identification « qualitative » de l'état de l'outil, détection d'anomalie de fonctionnement et suivi de l'évolution temporelle de l'usure.

Mots-clé : Maintenance intelligente,Détection de défauts, Diagnostic vibratoire, Data mining, Analyse en composantes principales, Analyse spectrale singulière, Défauts de roulements, Usure d'outils .

Remerciements

Au terme de cette recherche doctorale, je tiens à exprimer ma sincère gratitude envers mon promoteur de thèse, le Professeur Pierre Dehombreux, pour sa constante disponibilité, ses précieux conseils et sa bonne humeur dont j'ai joui tout au long de ces quatre années passées au sein du laboratoire de génie mécanique de la Faculté Polytechnique de Mons (FPMs).

Je remercie de tout coeur le Professeur Filippi pour ses avis très utiles sur les aspects concernant l'usinage et l'usure d'outils.

Mes remerciements au Professeur Édouard Mulapi de l'Université de Kinshasa pour m'avoir assuré son encadrement lors de mes séjours locaux en RD Congo.

Mes remerciements s'adressent aussi à tous les membres du laboratoire de génie mécanique de la FPMs, avec qui il a été très agréable de collaborer.

Toute ma reconnaissance aux membres du Laboratoire de Mécanique Appliquée (LMA) de l'Université de Reims, particulièrement au Professeur Dron pour sa précieuse collaboration qui m'a permis l'accès au banc de test des roulements. Je remercie également M. Xavier Chiementin qui a su trouver, malgré ses occupations, du temps pour conduire avec moi les tests sur les roulements.

Je voudrai exprimer ma reconnaissance aux dirigeants de Colinet pour nous avoir permis de procéder à des essais d'usinage sur leur machine.

A M. Mack et à la société PEPITE, un grand merci pour leur collaboration en mettant à notre disposition une licence gratuite du logiciel PEPITo.

Je ne m'en irai pas oublier ces personnes sans l'aide et l'amitié desquelles, le temps passé, loin de ma famille, à préparer cette thèse aurait été dur à vivre. Je pense à Jean Tshimanga, François Kitenge, Guy, Seguy, Janny, Mans,...

A Guillaume, François et Poussy Th., toute ma reconnaissance pour les relectures de ce rapport.

Enfin, que la Coopération Technique Belge trouve ici toute ma reconnaissance pour avoir

rendu possible cette entreprise en m'accordant une bourse de doctorat.

Table des matières

1 Introduction **1**
1.1 Contexte de l'étude . . . 1
1.2 Objectifs de la thèse . . . 5
1.3 Organisation du document . . . 6
Références . . . 10

2 Les fondamentaux **13**
2.1 Introduction . . . 13
2.2 La surveillance vibratoire . . . 14
2.2.1 Outils de détection . . . 15
2.2.2 Outils de diagnostic . . . 19
2.3 La détection et le diagnostic . . . 21
2.3.1 Les méthodes qualitatives . . . 21
2.3.2 Les méthodes quantitatives . . . 22
2.4 Le pronostic . . . 23
2.4.1 Les techniques basées sur le modèle . . . 24
2.4.2 Les techniques basées sur les probabilités (approche fiabiliste) . . . 25
2.4.3 Les techniques basées sur l'apprentissage . . . 25
2.5 Conclusion . . . 25
Références . . . 27

3 Prétraitement de signaux vibratoires et extraction d'indicateurs **31**
3.1 Introduction . . . 31
3.2 Représentation des données temporelles . . . 32
3.3 Représentations basées sur la transformation . . . 32
3.3.1 La décomposition en ondelettes . . . 33
3.3.2 Les représentations temps-fréquence . . . 37

3.4 Représentation continue dans le domaine temporel 38
3.5 Représentations basées sur les modèles 42
3.6 Fusion des données 43
3.7 Réduction dimensionnelle de l'espace d'indicateurs 44
3.7.1 L'analyse en composantes principales (ACP) 45
3.7.2 Normalisation 46
3.7.3 Saturations, communalités et variance expliquée 47
3.7.4 Choix de la dimension effective 47
3.7.5 Les rotations en analyse en composantes principales 48
3.8 Application : usure d'outils en tournage 49
3.9 Conclusion 54
Références 58

4 *Data mining*, détection et diagnostic 61
4.1 Définitions 61
4.1.1 *Data mining* et Extraction des Connaissances à partir de Données (ECD) 61
4.1.2 Le *machine learning* ou apprentissage automatique 63
4.1.3 Instances et Attributs 63
4.2 Les techniques du *data mining* 64
4.3 La classification 65
4.3.1 Les arbres de décision 66
4.3.2 Les règles de classification 68
4.3.3 Les modèles statistiques 68
4.3.4 Les réseaux bayésiens 70
4.3.5 Les réseaux neuronaux artificiels 72
4.3.6 Les machines à vecteurs de support 79
4.3.7 L'inférence floue 85
4.3.8 Les K plus proches voisins 88
4.4 Des instances non étiquetées 90
4.5 Sélection des indicateurs pertinents 92
4.5.1 Les méthodes du filtre 93
4.5.2 Les méthodes d'enveloppe 95
4.6 Une approche hybride de sélection des indicateurs pertinents 96
4.7 Mesure des performances d'un système de diagnostic 102
4.7.1 La validation croisée 105
4.7.2 Illustration de l'évaluation des performances 106

4.8 Approches en diagnostic . . . 107
4.8.1 Application . . . 109
4.8.2 Synthèse . . . 113
4.9 Prise en compte des coûts de non détection et de fausses alarmes . . . 114
4.10 *Dynamicité* . . . 117
4.10.1 Structuration des données. Cube de données de surveillance vibratoire . 118
4.10.2 Classification des problèmes de surveillance en vue du *data mining* . . 119
4.10.3 Les problèmes statiques . . . 120
4.10.4 Les problèmes dynamiques . . . 121
4.10.5 Méthode de transformation par dérivation . . . 122
4.10.6 Méthode de transformation par moyenne mobile . . . 122
4.11 Conclusion . . . 123
Références . . . 124

5 Détection précoce et identification de défauts des roulements 131
5.1 Introduction . . . 131
5.2 Manifestations vibratoires de défauts de roulements . . . 132
5.3 Détection d'enveloppe . . . 135
5.4 Induction des défauts . . . 135
5.5 Évolution du niveau fréquentiel et des indicateurs temporels avec la dégradation 136
5.6 Prétraitement des signaux . . . 139
5.7 Détection précoce . . . 143
5.7.1 Détection de défaut par analyse en composantes principales . . . 143
5.7.2 Application à la détection de défaut de roulement . . . 148
5.8 Diagnostic . . . 154
5.9 Conclusion . . . 157
Références . . . 158

6 Surveillance d'usure d'outils en usinage 161
6.1 Introduction . . . 161
6.2 Mécanismes et modes d'endommagements des outils . . . 162
6.2.1 Usure adhésive . . . 163
6.2.2 Usure abrasive . . . 165
6.2.3 Usure par diffusion . . . 165
6.3 Critères associés à l'usure de l'outil . . . 166
6.4 Les méthodes de surveillance d'usure d'outils . . . 167

6.4.1 Les méthodes directes 167
6.4.2 Les méthodes indirectes 168
6.5 Modèle pour l'évaluation de l'usure d'outils 170
6.6 Surveillance d'outils par signaux vibratoires 172
6.7 Classification d'états d'outils : application 172
6.7.1 Sélection de variables pertinentes 173
6.7.2 Résultats et discussion 174
6.7.3 Extension de l'étude par l'utilisation de tous les canaux 176
6.7.4 Approche de l'étude avec les indicateurs scalaires 178
6.7.5 Conclusion de l'étude 179
6.8 Suivi de l'évolution temporelle de la dégradation d'un outil de coupe 180
6.8.1 Définition du symptôme vibratoire pour l'usure 181
6.8.2 Application aux données expérimentales 185
6.8.3 Synthèse 185
6.8.4 Conclusion 188
Références 190

7 Conclusions et perspectives **193**

Annexes **199**

A Essais expérimentaux sur banc d'essai de machines tournantes **199**

B Essais expérimentaux pour la surveillance d'usure d'outils en tournage **207**

C Essais expérimentaux pour l'évolution temporelle de l'usure en fraisage **211**

D Essais expérimentaux sur banc de détection de défauts des roulements **215**

Chapitre 1

Introduction

1.1 Contexte de l'étude

Les équipements techniques sont soumis à des contraintes de plus en plus exigeantes relativement à leur sûreté de fonctionnement (fiabilité, maintenabilité, disponibilité, sécurité), leur performance (productivité, qualité) et leur coût global mesuré tout au long de leur cycle de vie [1].

Dans ce contexte, l'optimisation des opérations de maintenance apparaît comme un levier d'action efficace : la détection des signes précurseurs d'une défaillance, le diagnostic sûr et rapide d'une panne, la prédiction de la durée de vie résiduelle d'un équipement figurent à présent parmi les préoccupations communes de la communauté scientifique et industrielle.

Trois approches peuvent être envisagées pour rencontrer ces objectifs (figure 1.1) : la modélisation physique de dégradations, les méthodes fiabilistes et les méthodes basées sur la mesure (*condition monitoring*).

Le concept actuel de maintenance intelligente ambitionne l'intégration de toutes ces sources de connaissance afin d'améliorer la gestion des équipements techniques, et par conséquent la prédiction des défaillances. La maintenance intelligente vise aussi à prendre en compte, dans le processus décisionnel, des données relatives aux états passé et présent de l'équipement et de son environnement en vue de prédire son évolution.

Ces trois approches reprises par la figure 1.1 présentent de manière synthétique les thématiques de recherche, en rapport avec la maintenance et la fiabilité, exploitées dans le service de génie mécanique de la Faculté Polytechnique de Mons. La thèse de M. Basile [2] a proposé des modèles fiabilistes pour l'optimisation de la maintenance d'équipements soumis à des sol-

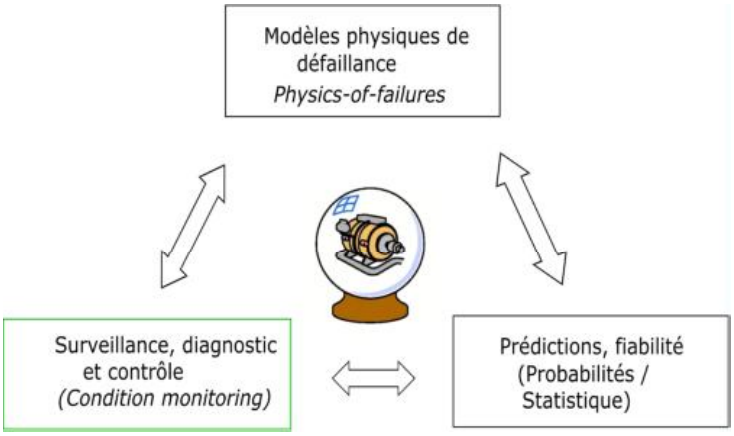


FIG. 1.1 – Méthodes de prédiction des défaillances

licitations variables et a jeté des ponts entre les approches fiabilistes et les modèles physiques de défaillance. La thèse de M. Letot initiée cette année va s'intéresser aux modèles physiques de dégradations pour les relier aux données de mesures (*condition monitoring*) et aux modèles fiabilistes.

Notre thèse se positionne dans l'axe *condition monitoring*. Elle explore les voies d'une exploitation de données vibratoires relatives aux états passé et présent de l'équipement dans le but de la détection et du diagnostic. Les données vibratoires sont les plus utilisées dans la surveillance des machines tournantes à cause du caractère localisé de leur mesure, de leur sensibilité et de leur capacité à prédire assez précocement le développement des défauts [3, 4]. Comparées aux autres techniques, elles sont, selon [5], les seules à donner la mesure la plus pertinente de l'état d'une machine tournante.

Le développement actuel des technologies de mesure et les progrès réalisés dans les moyens de transmission et de stockage permettent de mesurer, en continu ou périodiquement, presque tous les paramètres de fonctionnement d'un processus/équipement et de les archiver. Le profit que l'on peut tirer de ces gisements de données stockées peut être énorme si les schémas de *condition monitoring* intelligent émergent par rapport aux anciennes stratégies. En effet, il s'impose un urgent besoin d'ajuster les moyens de compréhension et d'analyse aux capacités de génération et de stockage de données. Ces schémas sont appelés à mettre en œuvre l'analyse intelligente de données, particulièrement les méthodes d'apprentissage, pour pouvoir aboutir à leur transformation en informations utiles à la prise de décision, et poser ainsi le socle des activités de détection, de diagnostic et de pronostic.

Ces besoins nous placent directement dans le champ de l'Extraction des Connaissances à

partir des Données (ECD), communément appelée *data mining*, qui ouvre un domaine offrant théories et outils nécessaires pour assister l'homme dans la tâche d'extraction de l'information utile de ces vastes ensembles de données électroniques [6, 7]. Le *data mining* s'appuie sur l'informatique en utilisant les outils traditionnels de la statistique et les nouveaux outils de *machine learning*. Dans le cadre d'aide à la maintenance, cette exploitation des données, passées et présentes, doit pouvoir permettre d'une part la découverte d'indicateurs pertinents pour la détection, la localisation et l'identification des défauts, et d'autre part mener à la construction de systèmes qui procèdent automatiquement à ces tâches.

Le cycle des activités de *condition monitoring* et leur interfaçage avec les autres méthodes de prédiction tel que résumé par la figure 1.2, fait ressortir trois étapes principales : la détection, le diagnostic et le pronostic. Ces étapes sont précédées par la phase de prétraitement des signaux. La même logique se retrouve dans les standards MIMOSA OSA-CBM*, repris sous la norme ISO 13374-1, qui établissent de manière satisfaisante un cadre de développement d'un système de *condition monitoring* répondant aux exigences actuelles de distributivité, de transmission et d'interchangeabilité des informations (figure 1.3).

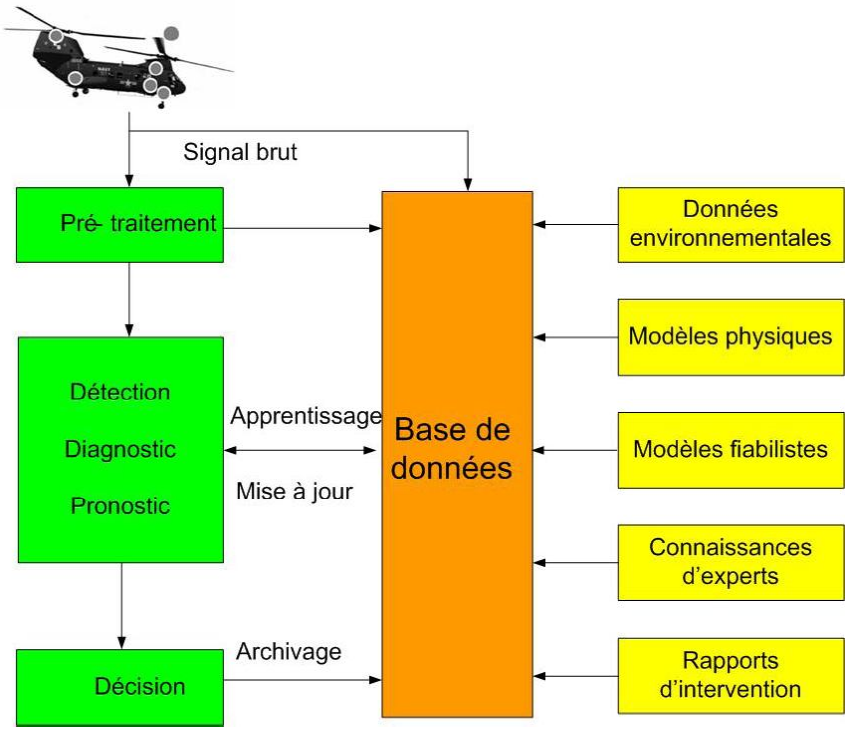


FIG. 1.2 – Cycle du *condition monitoring*

*Open Systems Architecture for Condition-based Maintenance

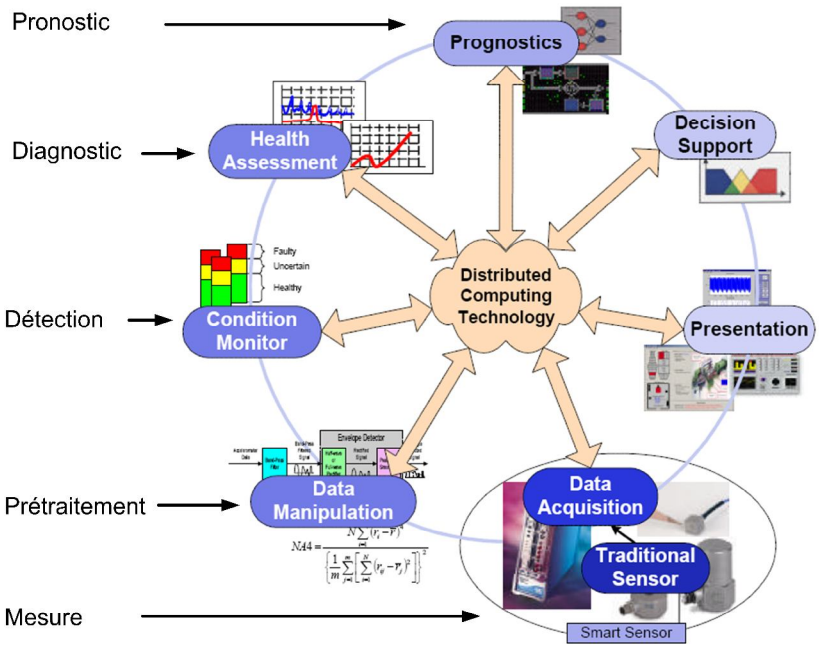


FIG. 1.3 – Flux de données dans le modèle OSA-CBM Mimosa

Dans le schéma de la norme ISO 13374-1, des travaux de recherche ont été réalisés en relation avec chaque couche du modèle [8, 9]. Cependant des efforts restent à consentir pour l'amélioration des performances des couches allant de la manipulation des données au pronostic. Cette amélioration s'entend en termes de précision, de temps de traitement et de déploiement industriel des algorithmes utilisés [10].

En ce qui concerne la représentation des données issues des mesures, c'est-à-dire l'extraction d'indicateurs pertinents, cette étape est fortement dépendante du problème à analyser et des défauts à surveiller [11]. Il est connu qu'un défaut donné induit des changements dans le signal vibratoire relevé sur l'équipement, mais la question qui demeure est celle de savoir comment trouver cette information dans un état le moins brouillé possible.

La détection et l'identification des défauts ont été abordées dans les domaines des machines tournantes [12], du contrôle de processus [13] et de la surveillance des outils de coupe en usinage [14, 15]. Il existe différentes méthodes pour le diagnostic et le pronostic de l'état des équipements ou des processus. D'une manière générale, ces approches peuvent être groupées en deux classes : celles basées sur la modélisation de la dynamique de l'équipement ou du processus et celles basées sur le retour d'expérience avec des données historiques. Ce travail de thèse s'inscrit dans cette seconde.

Comment choisir la meilleure technique de diagnostic ? Quelle confiance accorder aux ré-

sultats obtenus ? Voilà tant de questions qui méritent une attention particulière.

1.2 Objectifs de la thèse

Dans ce contexte général de l'exploitation des mesures vibratoires en vue de la détection et du diagnostic, cette thèse poursuit les objectifs suivants :

1. développer une démarche de prétraitement des signaux en vue du débruitage et adresser la question de la définition des indicateurs pertinents, y compris dans le cas où la dimension temporelle ne peut être négligée,
2. contribuer au développement des approches de détection précoce de dégradations mécaniques,
3. évaluer différents systèmes de diagnostic basés sur l'apprentissage. Les méthodes utilisées sont les réseaux de neurones, les réseaux bayésiens, les K plus proches voisins, les machines à vecteurs de support, les arbres de décision et l'inférence floue.

Les contributions principales de ce travail se situent au niveau méthodologique dans le processus de la manipulation de données afin d'améliorer les résultats au niveau de la détection et du diagnostic par les méthodes du *data mining*.

Le cadre méthodologique préconisé pour l'analyse des signaux vibratoires en vue de la détection et du diagnostic de dégradations mécaniques est rappelé sur une feuille intercalaire (reprise en pages 8 et 9). Les quatre campagnes expérimentales menées au cours de cette thèse ont contribué à des illustrations des méthodes exposées. Un tableau synthétique facilitera la localisation des résultats pour chacune d'entre elles, disséminés tout au long du texte pour les besoins de l'exposé des méthodes.

La première application expérimentale concerne l'identification de défauts multiples de machines tournantes. Les vibrations sont relevées sur un banc d'essai où plusieurs défauts sont induits simultanément. Ces données permettront d'appliquer la méthode de sélection de variables élaborée au cours de nos travaux et d'évaluer les capacités d'identification en cas d'occurrence simultanée de défauts de natures différentes.

La seconde application concerne la détection précoce de défauts superficiels de roulements. Les données sont traitées de manière adéquate et des systèmes intelligents sont construits de façon à permettre non seulement une détection automatique du début de dégradation, mais aussi une identification des phases d'évolution du défaut. L'approche est comparée aux méthodes

traditionnelles.

La troisième application s'intéresse à la reconnaissance d'usure d'outils par les vibrations. Dans ce domaine aucune connaissance méthodologique n'est clairement identifiée pour la détermination des gammes de fréquence intéressantes. En plus, l'application industrielle peut générer des signaux vibratoires non-stationnaires, ce qui nécessite l'adaptation du processus de traitement du signal afin de définir des indicateurs vibratoires.

Enfin, la dernière étude expérimentale a été appliquée au suivi de l'usure d'outil en fraisage. Un outil est suivi jusqu'à la mort pour démontrer la nécessité et la possibilité de construire un indicateur qui reflète l'évolution temporelle de l'usure.

Ces différentes applications ont posé chacune un problème différent : diagnostic de défauts combinés, exploitation des signaux instationnaires, détection précoce et évolution temporelle de dégradations. La sélection des indicateurs pertinents et la réduction dimensionnelle ont constitué des questions communes à toutes ces applications. Les annexes A à D détaillent les expérimentations que nous avons conduites pour chaque application.

Au cours de ces travaux de thèse, nous avons développé une librairie de fonctions sous MATLAB que nous avons intégrée à une interface d'acquisition programmée sous LABVIEW afin de permettre le calcul de certains indicateurs en cours d'acquisition. Cette librairie MATLAB est principalement destinée au prétraitement des signaux, à la manipulation des données pour l'extraction des indicateurs pertinents et à la détection précoce des avaries. Pour l'évaluation des systèmes de diagnostic, et donc la classification, nous avons principalement fait usage des outils disponibles sous les environnements *WEKA* et *PEPITo* [16, 17].

1.3 Organisation du document

Le chapitre 2 présente une synthèse des méthodes de détection, de diagnostic et de pronostic vibratoires.

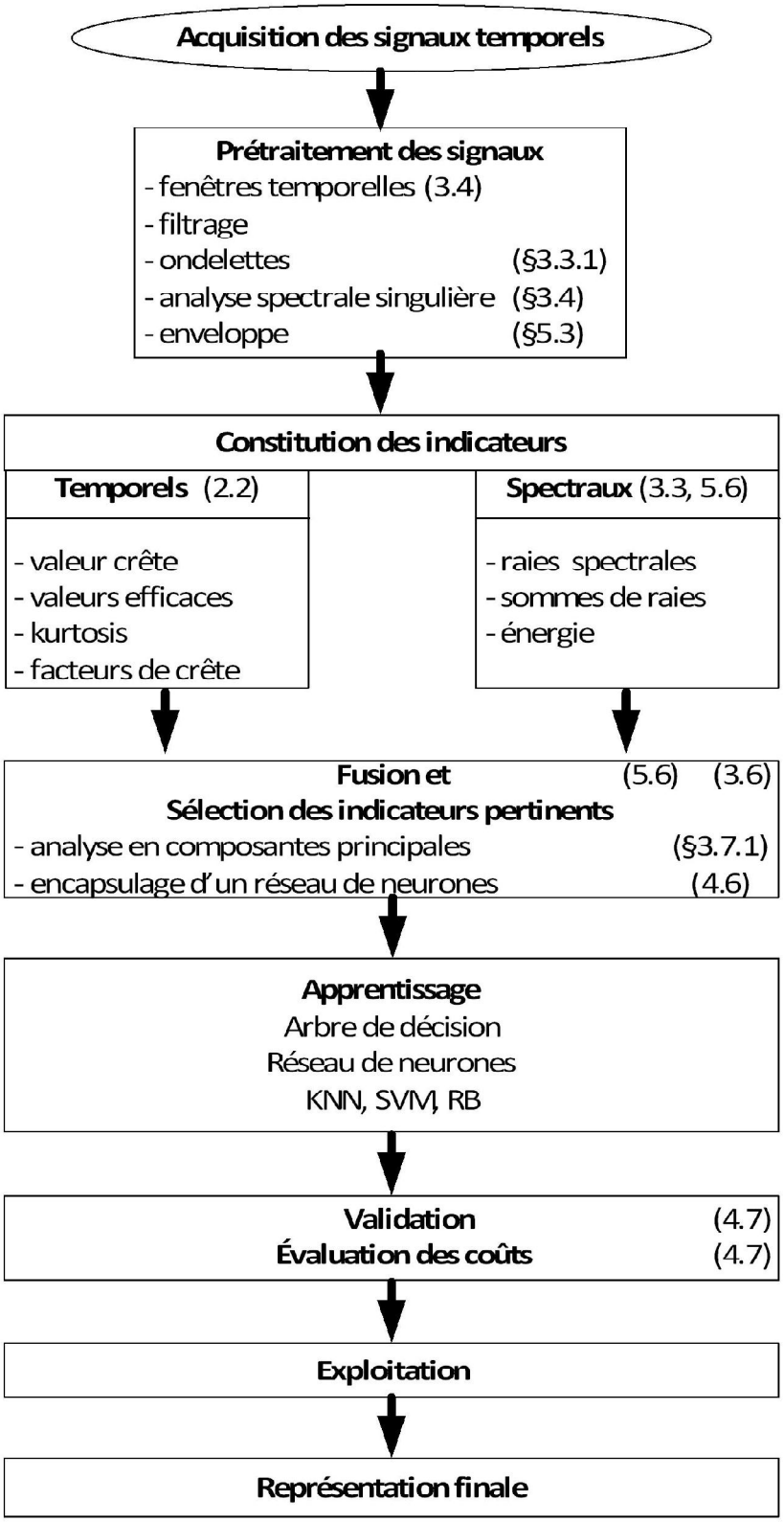
Dans le chapitre 3, nous analysons le problème de l'extraction des indicateurs. Différentes approches d'extractions d'indicateurs y sont présentées. **Nous présentons une démarche de prétraitement des signaux instationnaires en vue du débruitage et de la capture des phénomènes se produisant localement sur l'échelle du temps. L'analyse spectrale singulière en tant que moyen de débruitage et de décomposition des signaux est introduite aux côtés des méthodes traditionnelles d'analyse en ondelettes et de filtrage.** Ensuite les questions de fusion de données et de réduction dimensionnelle sont traitées. Une étude expérimentale est enfin

présentée pour illustrer les notions introduites dans le chapitre. Elle concerne une application dans le domaine de la surveillance d'usure d'outils en usinage.

Le chapitre 4 vient, par la suite, présenter en profondeur les techniques du data mining appliquées à la détection et au diagnostic. Nous adressons la question de sélection des indicateurs pertinents et **proposons une méthode de sélection qui est basée sur l'analyse en composantes principales (ACP) et les poids synaptiques d'un réseau de neurones. La méthode procède en enveloppant un classifieur de type réseau de neurones dans le processus de prétraitement afin de sélectionner les variables qui présentent un grand contenu informationnel vis-à-vis des défauts à diagnostiquer**. La validité de la méthode de sélection et les performances de différentes méthodes de diagnostic par apprentissage sont évaluées sur des données d'une étude expérimentale. Cette étude est basée sur un banc d'essai de machines tournantes et analyse le problème du diagnostic en cas de défauts multiples à plusieurs sévérités.

Les chapitres 5 et 6 présentent des cas complets d'applications concernant les machines tournantes et la surveillance d'usure d'outils respectivement. Dans le chapitre 5, nous discutons le problème de détection précoce et d'identification des avaries sur des roulements. L'ACP est revisitée, cette fois comme moyen de détection. **Nous l'utilisons pour modéliser la structure des données représentant le fonctionnement normal du système. Sa combinaison avec les arbres de décision fournit un bon moyen de détection**. Au chapitre 6, l'usure en tournage est analysée. L'exemple d'application entamé au chapitre 3 est repris pour illustrer les méthodologies du *data mining* et démontrer à travers les résultats la consistance de la démarche proposée passant par un prétraitement par analyse spectrale singulière pour extraire les indicateurs. Enfin, une étude de la surveillance d'usure en fraisage est présentée pour illustrer la manipulation des données à dépendance temporelle.

Méthodologie générale



Essais expérimentaux

	Défauts rotor	Défauts roulement	Usure en tournage Colinet	Usure en fraisage Sirris
Objectifs	Diagnostic de défauts multiples (4 types, 3 ou 5 niveaux par type, 3 vitesses de rotation)	Détection précoce de défauts de roulement	Identification « qualitative » de l'état de l'outil (5 classes) Détection d'anomalies de fonctionnement	Suivi de l'évolution temporelle de l'usure Identification d'un indicateur permettant une étude de pronostic
Annexes et pages	Annexe A, page 199	Annexe D, page 215	Annexe B, page 207	Annexe C, page 211
Nature des mesures **Taille de l'ensemble de données**	- 4 accélérations radiales pour chaque combinaison de défauts - 162 conditions différentes - échantillonnage à 5120 Hz - durée de chaque enregistrement = 20 s	- 3 accélérations dans des directions orthogonales - 353 enregistrements sous conditions variables - échantillonnage à 50 kHz - durée de chaque enregistrement de 0,2s - succession temporelle des signaux	- 3 accélérations orthogonales sur tourelle porte-outil - 43 conditions différentes - échantillonnage à 51,2 kHz - durée de chaque enregistrement ≈ 20s - signaux non-stationnaires	- 1 accélération sur pièce - 117 conditions différentes - échantillonnage à 25kHz - durée de chaque enregistrement de 10 s - succession temporelle des signaux jusqu'à défaillance
Prétraitement	Filtrage passe-bande	Détection d'enveloppe avec 3 préfiltrages distincts : - filtrage passe-bande - analyse en ondelettes - analyse spectrale singulière	Analyse spectrale singulière Filtrage passe-bande	Filtrage passe-bande
Indicateurs	- valeur efficace - kurtosis - skewness dans 4 bandes de fréquence	Somme des raies spectrales Résidus par ACP	Somme des raies spectrales	Moyennes courantes Résidus par ACP
Méthodes utilisées	Réseaux de neurones Arbre de décision K plus proches voisins (KNN) Réseaux bayésiens Logique floue	Réseaux de neurones Arbre de décision K plus proches voisins (KNN) Réseaux bayésiens Machines à vecteurs de support	Réseaux de neurones Arbre de décision K plus proches voisins (KNN) Réseaux bayésiens	Réseau de neurones dynamique
Présentation des résultats	4.3.4 4.3.6 4.3.7 5.7 4.6 4.8.1	5.8	3.8 6.7	6.8

Références

[1] B.S. Dhillon : *Engineering Maintenance : A modern approach.* CRC Press LLC, New York, 2002.

[2] O. Basile : *Prise en compte de l'incertitude dans les modèles fiabilistes de maintenance industrielle. Extension aux sollicitations variables*. Thèse de doctorat, Faculté Polytechnique de Mons, 2007.

[3] M. Serridge : Ten crucial concepts behind trustworthy fault detection in machine condition monitoring. *In Proceedings of the 1st International machinery monitoring and diagnostics conference and exhibit, Las Vegas, NV, 722-727*, 1989.

[4] D.Chang-Ching Lin et B.Hsu-Pin Wang : Performance analysis of rotating machinery using enhanced cerebellar model articulation (e-cemac) neural networks. *Computers Ind.Eng.*, 30(2):227–242, 1996.

[5] M. Angels : Choosing accelerometers for machinery health monitoring. *Sound and vibration*, pages 20–24, 1990.

[6] U. Fayyad, G. Piatetsky-Shapiro et P. Smyth : From data mining to knowledge discovery in databases. *American Association of Artificial Intelligence*, pages 37–54, 1996.

[7] D.J. Bertholds, M. & Hand, éditeur. *Intelligent Data Analysis*. Springer, 2003.

[8] H.R. DePold et F.D. Gass : The application of expert systems and neural networks to gas turbine prognostics and diagnostics. *Journal of Engineering for Gas Turbines and Power*, 121 :4:607–612, 1999.

[9] Andrew K.S. Jardine, Daming Lin et Dragan Banjevic : A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20(7):1483–1510, 2006.

[10] J. Lee, J. Ni, D. Djurdjanovic, H. Qiu et H. Liao : Intelligent prognostics tools and e-maintenance. *Computers in Industry*, 57:476–489, 2006.

[11] G. Vachtsevanos, F.L. Lewis, M. Roemer, A. Hess et B. Wu : *Intelligent fault diagnosis and prognosis for engineering systems*. John Wiley & Sons, 2006.

[12] P.J Lu, M.C. Zhang, T.C. Hsu et J. Zhang : An evaluation of engine faults diagnostics using artificial neural networks. *Journal of Engineering for Gas Turbines and Power*, 12 :2:340–346, 2001.

[13] A. Devillez, M. Sayed-Mouchaweh et P. Billaudel : A process monitoring module based on fuzzy logic and pattern recognition. *International Journal of Approximate Reasoning*, 37:43–70, 2004.

[14] Dimla E. Dimla SNR : Sensor signals for tool-wear monitoring in metal cutting operations–a review of methods. *International journal of Machine Tools and Manufacture*, 40:1073–1098, 2000.

[15] X. LI, S. DONG et P. K. VENUVINOD : Hybrid learning for tool wear monitoring. *International Journal of Advanced Manuacturing Technology*, 16:303–307, 2000.

[16] Ian H.WITTEN et Eibe FRANK : *Data mining : Practical machine learning tools and techniques*. Morgan Kaufmann, San Francisco, 2nd édition, 2005.

[17] *PEPITo Data Mining Software UserGuide v1.5 (c)*, 2006.

Chapitre 2

Les fondamentaux de la détection et du diagnostic vibratoire

Le but de ce chapitre est de présenter, en préambule à nos apports, les bases de la détection et du diagnostic vibratoire.

2.1 Introduction

Le *condition monitoring* est un outil de gestion préventive des équipements techniques basé essentiellement sur la mesure. Les techniques mises en œuvre sont de divers types. Nous pouvons citer, à titre d'exemples :

- la surveillance vibratoire ;
- la surveillance des grandeurs électriques ;
- la thermographie infrarouge ;
- l'analyse des propriétés des huiles et de leur concentration en particules d'usure ;
- les émissions ultrasoniques ;
- les contrôles non destructifs.

Les méthodes du *condition monitoring* peuvent être vues comme des approches qui permettent la détermination du moment optimum auquel programmer une intervention de maintenance [1]. On prélève sur un équipement, de manière continue (*on-line*) ou périodique (*off-line*), des mesures d'indicateurs pertinents de l'évolution de la dégradation de la machine. Si l'on dispose d'une signature caractéristique de l'état de fonctionnement nominal, on peut décider si les mesures acquises traduisent un changement significatif de comportement : c'est la phase

de **détection**. La finalité d'une surveillance est l'émission d'alertes à partir d'événements. Ces alertes sont censées provoquer une réaction humaine ou automatique qui vise à localiser (isoler) et identifier la cause de la déviation ainsi que d'en préciser la gravité par un processus de **diagnostic** [2]. Sur base de l'état courant de l'équipement ou du processus, une prédiction de la vie résiduelle peut être effectuée par un processus de **pronostic**.

Nous allons d'abord circonscrire les outils de détection et de diagnostic tels que vus traditionnellement en maintenance vibratoire, avant de présenter une taxinomie de la détection et du diagnostic automatisés.

2.2 La surveillance vibratoire

Les vibrations mécaniques d'une machine contiennent des informations extrêmement riches quant aux défauts de fonctionnement qui les induisent. Leur analyse, temporelle ou fréquentielle, permet d'effectuer une détection précoce des avaries et, éventuellement, un diagnostic précis des défectuosités rencontrées. La caractérisation du comportement vibratoire d'un équipement s'avère ainsi un moyen privilégié pour la mise en œuvre d'un programme de maintenance prédictive.

L'analyse vibratoire appliquée à la maintenance préventive procède en plusieurs étapes :

1. la mesure d'une grandeur cinématique en des points définis à proximité des organes critiques de la machine ;
2. la détermination des caractéristiques du signal relevé sur les machines d'un parc donné, et dont la nature (valeur globale, spectre fréquentiel, évolution temporelle, ...) varie en fonction du degré d'analyse (de la « simple » surveillance au diagnostic le plus précis) ;
3. l'enregistrement et la gestion informatique des informations recueillies sur site, afin de permettre le suivi de leur évolution historique et d'enrichir la base de données nécessaire à l'expertise ;
4. la reconnaissance des avaries proprement dites et une assistance à la prise de décision quant aux actions à entreprendre (mesures complémentaires, planification d'une opération de maintenance, ...)

La maintenance d'une machine nécessite une connaissance intime de son fonctionnement et des éléments qui la composent. Des études théoriques permettent d'associer une signature vibratoire particulière à un défaut d'un organe déterminé : par exemple, il est possible de dé-

terminer la composante fréquentielle caractéristique d'un écaillage sur la bague intérieure d'un roulement [3–5].

Pour chaque type de défaut à détecter ou à diagnostiquer, il faut trouver la grandeur physique la plus pertinente et lui appliquer la procédure de traitement de signal la plus appropriée.

Les méthodes de caractérisation des signaux vibratoires seront de niveaux de complexité fort différents selon que l'on souhaite détecter l'apparition des défauts ou diagnostiquer les problèmes détectés. Dans ce dernier cas, il s'agit de mettre en œuvre un ensemble de techniques d'analyse des signaux vibratoires permettant la reconnaissance de la signature de défauts particuliers. Pour être précis, un diagnostic doit pouvoir s'appuyer sur des critères discriminants. Ceux-ci ne sont pas toujours disponibles : par exemple, l'apparition d'une fréquence (harmonique 2) peut être causée par plusieurs phénomènes.

Dans ce qui suit nous rappelons les outils [6] correspondants, classés par ordre croissant de complexité, en fournissant les références bibliographiques qui les décrivent.

2.2.1 Outils de détection

En surveillance vibratoire, les outils les plus simples pour la détection sont basés sur les indicateurs extraits du domaine temporel. Dans le domaine temporel, les indicateurs sont généralement définis sur base des moments statistiques [7].

Si la série temporelle de la vibration est regardée comme une variable aléatoire X, la répartition de la probabilité est définie comme

$$F(x) = P(X < x) \tag{2.1}$$

la fonction de densité de probabilité est

$$f(x) = \frac{dF(x)}{dx} \tag{2.2}$$

l'espérance mathématique d'une fonction $g(x)$ est l'intégrale

$$E\{g(x)\} = \int_{-\infty}^{+\infty} g(x)f(x)dx \tag{2.3}$$

On définit la fonction caractéristique de la variable aléatoire comme la transformée de Fourier de sa fonction de densité de probabilité $f(x)$ moyennant un changement de signe [7]. Elle

est donnée par :

$$\Phi(\omega) = \int_{-\infty}^{+\infty} e^{j\omega x} f(x) dx = E\{e^{j\omega x}\} \tag{2.4}$$

cette fonction est aussi appelée *fonction de génération des moments statistiques*. En effet, les moments sont les coefficients des puissances de $j\omega$ dans son développement en série de Taylor

$$E\{e^{j\omega x}\} = E\left\{1 + (j\omega)x + \frac{(j\omega)^2 x^2}{2!} + \frac{(j\omega)^3 x^3}{3!} + \cdots\right\} \tag{2.5}$$

soit encore

$$E\{e^{j\omega x}\} = 1 + (j\omega)E\{x\} + \frac{(j\omega)^2}{2!}E\{x^2\} + \frac{(j\omega)^3}{3!}E\{x^3\} + \cdots \tag{2.6}$$

Les moments d'ordre n sont

$$E\{x^n\} = m_n \tag{2.7}$$

On pourra aisément remarquer que les moments sont des dérivées de la fonction caractéristique par rapport à $j\omega$ et calculées au point $\omega = 0$

$$m_n = \frac{d^n \Phi(0)}{d(j\omega)^n} \tag{2.8}$$

La transformation consistera à représenter le signal temporel dans l'espace d'indicateurs basés sur les moments statistiques dont les plus utilisés sont

la valeur efficace centrée (RMS) : c'est un indicateur temporel qui mesure le contenu énergétique dans un signal vibratoire. Pour une série temporelle $\{x_i\}$ de longueur n, la valeur efficace s'exprime par

$$RMS = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{2.9}$$

où $\bar{x}$ est la valeur moyenne de $\{x_i\}$. La valeur RMS est la racine carrée du deuxième moment statistique m_2.

Les valeurs efficaces calculées dans des bandes fréquentielles données peuvent servir d'indicateurs particulièrement simples à utiliser [8] : on considère généralement un en-

semble de quatre à dix bandes de fréquences pour leur définition.

le coefficient d'asymétrie ou facteur de vrillage (*skewness*) : Il caractérise le degré de dissymétrie d'une distribution autour de sa moyenne. Il est calculé par

$$Skew = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{RMS}\right)^3 \tag{2.10}$$

le coefficient d'applatissement ou *kurtosis* : il est un indicateur de la proportion d'échantillons qui s'écartent faiblement de la valeur moyenne comparée à ceux qui s'en écartent largement. Il caractérise l'allure aplatie d'une distribution par rapport à la distribution normale. Il est calculé par

$$Kurt = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{RMS}\right)^4 - 3 \tag{2.11}$$

La présence d'un terme de puissance 4 donne un poids considérable aux amplitudes élevées. Le *kurtosis* permet de détecter l'apparition de chocs et de suivre l'évolution des défauts induisant des forces impulsionnelles périodiques.

facteur de crête : le facteur de crête est un des indicateurs spécifiques qui accentuent l'apparition de chocs dans les vibrations. Les pics dans le signal temporel correspondront à un accroissement du facteur de crête. Cet indicateur, comme le *kurtosis*, est utilisé pour la détection des anomalies qui se traduisent par des chocs liés par exemple au contact de surfaces dégradées [9, 10]. Dès l'apparition du défaut, l'augmentation des chocs induit la croissance du niveau de crête alors que la valeur efficace évolue de manière moins marquée. Il en résulte une augmentation significative du facteur de crête. Le facteur de crête est donné par

$$FC = \frac{\max(x_i)}{RMS} \tag{2.12}$$

Afin de mieux capturer le contenu informationnel des signaux de vibration, il est toujours conseillé, bien qu'étant dans le domaine temporel, d'extraire les indicateurs suivant des plages fréquentielles judicieusement choisies. Dans cet objectif, les filtres analogiques et digitaux fourniront un traitement adéquat au signal.

Pour que ces indicateurs puissent être comparés, il faut veiller à ce que les échantillons temporels desquels ils sont déduits soient filtrés dans des conditions rigoureusement identiques ; il

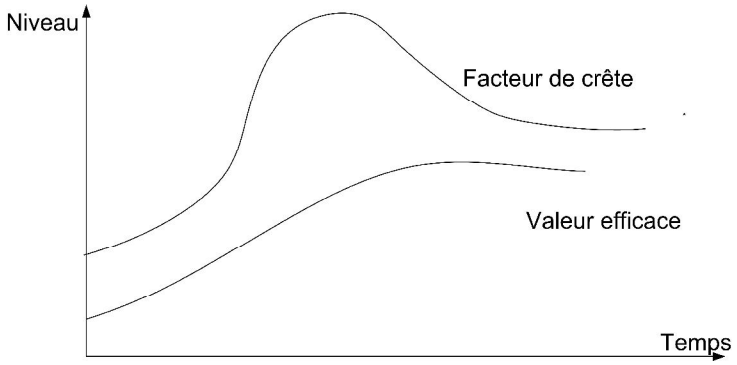


FIG. 2.1 – Evolution du facteur de crête

a en effet été constaté que les caractéristiques d'un filtre influencent fortement la valeur des niveaux extrêmes. De plus, la comparaison de valeurs de crête ne peut être logiquement effectuée que sur des échantillons de même durée, si le signal comporte une composante aléatoire. Un facteur crête supérieur à 6 est caractéristique de l'apparition de forces impulsionnelles [11].

De plus, il est possible de préconditionner le signal temporel afin de s'affranchir de problèmes tels que la variation de vitesse de rotation, la résonance du capteur, les effets de masque induits par l'excitation de nombreuses résonances ou par le fonctionnement d'autres organes de la même machine, ou le bruit de fond. Dron, Bolaers et Rasolofondraibe [11] proposent une méthode de débruitage des signaux applicable aux défauts de roulements.

En plus des indicateurs calculés dans le domaine temporel, il existe d'autres outils de détection qui sont

- la valeur d'un indicateur caractéristique du contenu « hautes fréquences » du signal. On souhaite s'affranchir (ou bénéficier ?) des résonances structurelles et utiliser la résonance de l'accéléromètre pour amplifier, par exemple, les composantes fréquentielles du signal entre 25 et 35 kHz ; ces méthodes sont l'apanage de fournisseurs qui commercialisent des matériels spécifiques exploitant des méthodes brevetées. La méthode SPM («*Shock Pulse Method*», [12]) est basée sur ce principe.
- l'examen d'un spectre fournissant la répartition des niveaux des vibrations dans des bandes de fréquence (tiers d'octave) relativement étendues, pour une réduction rapide des données [9] ;
- les méthodes basées sur une modélisation paramétrique du signal (modèles autorégressifs) [13]

2.2.2 Outils de diagnostic

En analyse vibratoire traditionnelle appliquée à la maintenance, les outils de diagnostic utilisés sont :

la transformée de Fourier discrète (DFT)

La transformée de Fourier fournit les composantes fréquentielles caractéristiques [10] d'un signal filtré et échantillonné. Cette opération est généralement effectuée en temps réel sur des analyseurs spécialisés, mais elle peut être réalisée après l'enregistrement des signaux temporels sur le système informatique utilisé pour la gestion des informations.

Pour un signal temporel $x(t)$ décrit par une série discrète de n valeurs x_i exprimées aux instants t_i, la DFT est définie par la série des coefficients complexes :

$$X_k = \frac{1}{n} \sum_{i=1}^{n} x_i e^{-j2\pi ki/n} \tag{2.13}$$

Cette relation est établie sur une fenêtre de n valeurs échantillonnées produisant ainsi des raies fréquentielles avec une résolution spectrale de f_{ech}/n, où f_{ech} est la fréquence d'échantillonnage.

l'analyse cepstrale

Suite à la présence d'harmoniques et de raies latérales, les spectres fréquentiels présentent eux-mêmes une certaine périodicité ; l'analyse cepstrale facilite la détection des multiplicités fréquentielles par des transformées de Fourier successives.

l'analyse d'enveloppe

Les défauts impulsionnels excitent périodiquement les fréquences propres structurelles ; les résonances sont donc modulées en amplitude à la fréquence caractéristique du défaut. En démodulant une de ces résonances, on peut retrouver un signal caractéristique du défaut. En pratique, le signal est filtré autour d'une bande caractéristique d'une fréquence de résonance, ce qui a le mérite d'éliminer les composantes des autres sources de vibration. Ce signal est ensuite redressé et filtré pour éliminer ses composantes hautes fréquences.

le suivi d'ordre (*order tracking*)

C'est une méthode d'analyse fréquentielle, qui exprime les phénomènes en fonction de la vitesse de rotation *order* plutôt qu'en fonction de fréquences absolues (Hz). Le suivi d'ordre est particulièrement apprécié dans les applications où la vitesse de rotation varie. Pour être mise en œuvre, les signaux vibratoires doivent être échantillonnés non plus à intervalle de temps régulier mais en fonction d'un incrément constant d'angle de rotation d'axe. Alternativement, le signal peut être rééchantillonné pour autant que l'on dispose de l'évolution temporelle de l'angle de rotation ou de de la vitesse de rotation [14, 15]. L'analyse d'ordre facilite l'identification des composantes liées à la fréquence de rotation.

l'analyse modale

Plutôt que de s'intéresser à une caractérisation locale fournie par des mesures effectuées en quelques points particuliers de la machine, cette technique vise à caractériser le comportement dynamique de la structure dans son ensemble en fournissant, par exemple, ses modes propres de déformation. Une confrontation des caractéristiques vibratoires relevées expérimentalement en fonctionnement avec celles prédites par un modèle numérique ou identifiées dans l'état nominal permet une localisation des défauts structuraux. Cette technique est développée pour les structures mais assez peu exploitée pour les machines.

l'analyse temps-fréquence

Dans certains cas [16], les défauts ne se manifestent que pendant un laps de temps très court relativement au cycle de fonctionnement de la machine et ont un caractère transitoire ou instationnaire prononcé. De tels cas nécessitent des méthodes qui permettent une analyse à la fois dans le domaine temporel et dans le domaine fréquentiel. Les méthodes d'analyse temps-fréquence sont très nombreuses [17]. Les méthodes classiques (Wigner-Ville, Choi-Williams,...) fournissent une interprétation directe temps-fréquence alors que d'autres comme les ondelettes [18] ne permettent pas une interprétation directe dans le domaine fréquentiel [10, 16] ;

Les moyens alloués à la maintenance fixent le temps de collecte et d'exploitation des données [19] : il va sans dire que les méthodes les plus avancées nécessitent des études expérimentales et numériques qui ne se justifient que dans les cas économiques les plus critiques.

Les outils qui viennent d'être présentés pour la détection et le diagnostic vibratoires, nécessitent une forte expertise humaine. Il est pourtant possible de tirer profit de données et décisions

de détection et de diagnostic qui sont déjà archivées. Il faut pour cela mettre en œuvre des méthodes automatiques basées sur l'apprentissage, la classification et la reconnaissance des formes [20].

Nous allons, dans les sections qui suivent, présenter les différentes méthodes de détection et de diagnostic automatiques.

2.3 La détection et le diagnostic

La détection de défauts est basée sur la notion de déviation par rapport à l'état normal de fonctionnement. Il s'agit de disposer d'indicateurs sensibles dont les valeurs varient significativement dès les premières phases de l'apparition du défaut ; une alerte doit être générée dès qu'ils dépassent des amplitudes repères.

Le diagnostic doit pouvoir permettre la localisation et l'identification précises, et sans fausses alarmes, des défauts imminents ou naissants. Les algorithmes utilisés pour le diagnostic sont soumis à la contrainte de minimisation de faux positifs (fausses alarmes) et faux négatifs (non détection) tout en réduisant le délai entre la naissance et la détection/isolation d'un défaut [21].

La référence [22] présente une taxinomie assez complète des méthodes de diagnostic à partir de la forme requise pour le traitement des données du système. Cette classification distingue les techniques quantitatives des techniques qualitatives.

2.3.1 Les méthodes qualitatives

La stratégie employée est basée sur la notion de causalité dans le comportement du système. Les plus utilisées de ces méthodes sont les arbres de causes et la méthode AMDEC (Analyse des Modes des Défaillances, de leurs Effets et de leur Criticité). La méthode de l'arbre des causes permet d'identifier toutes les combinaisons possibles d'événements qui entraînent la réalisation d'un événement unique indésirable. Elle utilise une chaîne qui remonte jusqu'à un événement primaire considéré comme cause de la déviation observée dans le comportement du système [23]. L'AMDEC est une technique déductive et qualitative avec laquelle les effets (conséquences) des causes de défaillances des composants élémentaires sont systématiquement identifiés [24].

Cependant une sérieuse limitation à l'utilisation de ces méthodes est la génération d'un grand nombre d'hypothèses, ce qui rend le processus de décision incertain. Ceci est dû aux

ambiguïtés qualitatives qu'impliquent ces méthodes.

A ces deux méthodes qualitatives viennent s'ajouter deux autres qui utilisent les informations provenant d'un retour d'expérience :

- les méthodes basées sur des règles : c'est le cas des systèmes experts. L'historique de l'équipement est enregistré sous forme de couple (antécédent, conséquence) lié à une défaillance connue. Ceci exige une correspondance explicite des symptômes connus aux causes primaires ;
- l'analyse qualitative des tendances : elle utilise la présence d'information sur la tendance d'évolution dans une mesure donnée. Elle présente deux étapes fondamentales : l'identification des tendances dans les mesures et l'interprétation des tendances en termes de scénarios de défaillances. La procédure d'identification doit donc être robuste vis-à-vis des variations du signal dues aux bruits [22].

2.3.2 Les méthodes quantitatives

Ces méthodes peuvent être classées en méthodes internes et méthodes externes.

Dans les méthodes internes, la source de connaissance est la compréhension profonde du système. Ces méthodes reposent sur les relations mathématiques qui existent entre les variables, et que l'on essaie de modéliser de la façon la plus proche possible du processus réel. Les modèles sont développés soit en utilisant les principes fondamentaux qui régissent la physique du système soit en utilisant les données d'entrée et de sortie du processus. Ces modèles peuvent être dynamiques, statiques, linéaires ou non linéaires. Parmi ces méthodes, nous citerons la méthode du modèle, les méthodes par identification de paramètres, les méthodes par estimation du vecteur d'état et les méthodes par modélisation de signatures [24]. Dans ces méthodes, la détection et l'identification sont souvent effectuées sur base de ce qu'on appelle le résidu. Le résidu est un indicateur de défauts basé sur la différence entre les mesures relevées sur l'équipement et les résultats de calcul à partir d'équations du modèle.

Quand il est difficile de construire un modèle ou quand aucun modèle n'est disponible, les techniques basées sur les modèles du système sont inapplicables et on met en oeuvre des techniques dites externes. La seule connaissance repose alors sur l'expertise humaine ou automatisée confortée par un solide retour d'expérience. Les plus connues de ces méthodes sont les techniques de reconnaissance des formes, les techniques statistiques et les techniques basées sur l'intelligence artificielle. Ces dernières constituent des outils privilégiés pour la catégorisation de signaux issus du système à surveiller en fonctionnement, c'est-à-dire l'affectation de la

catégorie de défauts à ces signaux.

Dans l'utilisation des techniques basées sur l'intelligence artificielle, on peut faire usage des approches de classification ou des approches de modélisation. Le diagnostic est principalement intéressé par les approches de classification. On peut également définir un résidu qui témoigne de l'écart par rapport au fonctionnement normal. Dans ce cas un processus de diagnostic peut se construire sur base d'une approche modélisation afin de pouvoir détecter une déviation par rapport à la condition normale de fonctionnement.

D'une manière générale, ces méthodes font usage des techniques d'apprentissage automatiques qui font l'objet de ce travail de thèse. Les plus utilisées sont les réseaux de neurones. Dans le cas où les données seraient non étiquetées, c'est-à-dire dont les classes de défaut ne sont pas fournies, des processus d'apprentissage non supervisé ou de clusterisation sont appliqués. Les cartes de Kohonen [25] et la clusterisation floue sont deux méthodes très utiles à cet effet [26–29].

2.4 Le pronostic

Le pronostic ajoute au résultat du diagnostic une estimation de la durée de vie résiduelle d'un équipement ; il tente de prédire l'évolution ultérieure d'un état de dégradation. Le pronostic procède d'une analyse des tendances des performances des équipements en fonction des *limites technologiques connues* afin d'analyser et de pouvoir corriger les problèmes avant l'apparition de la défaillance [3, 30]. Il reste un des challenges dans la gestion de l'état des équipements.

Les capacités de pronostic permettent de tirer complètement profit de la durée de vie d'un composant/système en estimant les délais avant défaillance, ce qui permet l'organisation la plus rationnelle des activités de maintenance.

Les principales questions de recherche dans le domaine du pronostic identifiées dans [21] portent sur l'estimation de l'évolution du dommage physique :

- en l'absence d'une mesure directe du dommage physique, comment peut-on inférer son évolution en fonction du temps ? Est-il possible d'estimer le dommage à partir des quantités mesurables telles que les vibrations, la température, etc ?
- quelles méthodes faut-il pour exploiter les mesures disponibles aux capteurs et générer dans un sens probabiliste l'évolution du système jusqu'à la défaillance ?
- comment fixer les limites de confiance ou d'incertitude associées à la prédiction ?

Comme pour le diagnostic, on peut distinguer différents types de pronostics : les pronostics

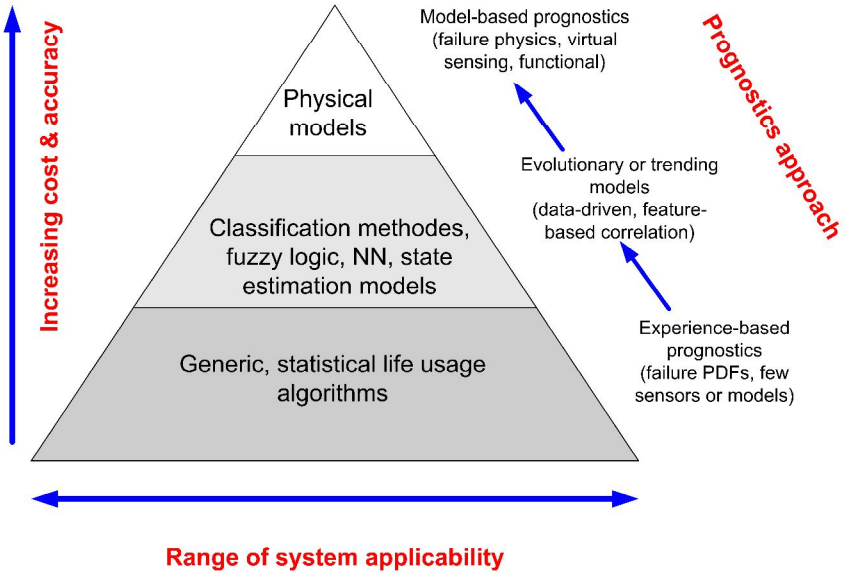


FIG. 2.2 – Approches techniques du pronostic (selon [21])

basés sur les modèles, les pronostics basés sur les données et les pronostics hybrides [31]. La figure 2.2 présente la vision de Vachtsevanos *et al.* [21] sur les approches possibles en fonction de l'applicabilité à divers systèmes et leur coût relatif d'implémentation.

2.4.1 Les techniques basées sur le modèle

Le pronostic peut se baser sur un modèle mathématique qui reçoit en entrée des paramètres de fonctionnement actuels du système pour délivrer l'estimation de son état à un instant futur. Cette catégorie comprend les techniques qui utilisent un modèle dynamique du processus à prédire. On y distingue les modèles basés sur la physique de dégradation (loi de fatigue, de corrosion,...), les modèles ARMA, les filtres de Kalman, etc.

Si l'on sait arriver à implémenter un modèle basé sur la dégradation physique, il peut être possible de calculer le dommage d'un composant critique en fonction des conditions de fonctionnement, et ainsi évaluer les effets cumulés au cours de la vie du composant. Il n'est cependant pas toujours aisé de construire un modèle mathématique précis basé sur la physique de la dégradation de l'équipement, plusieurs modes de défaillance étant concurrents et corrélés. Aussi recourt-on à des retours d'expérience pour la construction des modèles, généralement par une identification de paramètres. On suppose alors que le système obéit à un modèle dynamique donné et on utilise des données d'observation pour l'identification de paramètres du modèle.

2.4.2 Les techniques basées sur les probabilités (approche fiabiliste)

La détermination d'un modèle complet en terme d'équations différentielles peut être impossible voire non pertinente dans certains cas. Les historiques des défaillances pour une classe donnée d'équipements peuvent être analysés statistiquement. Le pronostic basé sur la modélisation des lois de fiabilité utilise les données de temps de défaillance pour initier les actions de maintenance en terme de fiabilité. Les modèles fiabilistes permettent d'envisager aussi une optimisation des opérations de maintenance vis-à-vis de critères opérationnels (disponibilité) ou économiques. Dans notre service, la thèse de doctorat de M. Basile [32] et le projet de recherche OPTIMAIN s'y sont intéressés.

2.4.3 Les techniques basées sur l'apprentissage

Si l'on dispose d'un historique de données de mesure jusqu'à la défaillance, on peut faire usage d'approximations pour produire directement l'information de prédiction nécessaire en terme des données. Dans ce cas, on n'a pas besoin de construire un modèle comme défini dans les sections précédentes. Ces méthodes d'approximation vont utiliser les techniques d'apprentissage. On peut citer les réseaux de neurones dynamiques et la logique floue qui présentent l'avantage de résoudre de manière satisfaisante le problème de prédiction. Les réseaux de neurones de type *GRNN* (*Generalized Regression Neural Network*) peuvent en même temps adresser la question des limites de confiance (incertitude).

2.5 Conclusion

Dans ce chapitre, nous avons établi une taxinomie générale des approches de diagnostic et de pronostic dont font partie les méthodes d'apprentissage, objet de ce travail. Les méthodes d'analyse vibratoire ont été présentées, principalement pour les machines tournantes. La surveillance vibratoire d'usure d'outil, qui est aussi étudiée dans ce travail de thèse, ne bénéficie pas d'une méthodologie aussi fournie que celle des machines tournantes.

Même quand elle existe (cas des machines tournantes), cette méthodologie nécessite une forte expertise humaine, qui est généralement chère. L'idée d'utiliser les méthodes d'apprentissage part du besoin de tirer profit de données et décisions de détection et de diagnostic déjà archivées.

Les systèmes de diagnostic et de pronostic reposent sur les données de mesures effectuées

sur les équipements. Les signaux vibratoires ne peuvent guère être utilisés à l'état brut. L'analyse nécessite un filtrage en des bandes de fréquence spécifiques et une synthétique représentation par une extraction d'indicateurs pertinents. Ces aspects sont abordés le chapitre suivant.

Références

[1] B.S. Dhillon : *Engineering Maintenance : A modern approach*. CRC Press LLC, New York, 2002.

[2] C. Pachaud et A. Boulanger : *Analyse vibratoire en maintenance : Surveillance et diagnostic des machines*. Dunod, Paris, seconde édition, 2003.

[3] R. Huang, L. Xi, X. Li, C. R. Liu, H. Qui et J. Lee : Residual life predictions for ball bearing based on self-organizing map and back propagation neural network methods. *Mechanical Systems and Signal Processing*, 21:193–2007, 2007.

[4] H. Qiu, Jay Lee, J. Lin et G. Yu : Robust performance degradation assessment methods for enhanced rolling element bearing prognostics. *Advances Engineering Informatics*, 17:127–140, 2003.

[5] J. Qiu, C. Zhang, B.B. Seth et S.Y. Liang : Damage mechanics approach for bearing lifetime prognostics. *Mechanical Systems and Signal Processing*, 16(5):817–829, 2002.

[6] A.W. Smyth, S.F. Masri, T.K. Caughey et N.F. Hunter : Surveillance of mechanical systems on the basis of vibration signature analysis. *Journal of Applied Mechanics*, 67: 540–551, 2000.

[7] A. Papoulis : *Probability, Random Variables and Stochastic Processes*. McGrawHill, 1965.

[8] T. Toutountzakis et D. Mba : Observations of acoustic measurement methods for the detection of defects in rolling element bearings. *Trybology International*, pages 469–480, 2003.

[9] Z. Kiral et H. Karagulle : Simulation and analysis of vibration signals generated by rolling element bearing with defects. *Tribology International*, 36(9):667–678, 2003.

[10] F.K. Choy, S. Huang, J.J. Zakrajsek, R.F. Handschum et D.P. Townsend : Comparaison of interpolation methods as applied to time synchronous averaging. Rapport technique, NASA Glen Research Center, Report tm-1999-209086 arl-tr-1960,, 1999.

[11] J. P. Dron, F. Bolaers et L. Rasolofondraibe : Optimisation de la détection de defauts de roulements par débruitage des signaux par soustraction spectrale. *Mécanique et Industries*, 4(3):213–219, 2003.

[12] Sweden SPM Instrument, Strängnäs, éditeur. *A complete set of tools for evaluating the running condition of operating rooling element bearings*. 1994.

[13] C. K. Mechefske et J. Mathew : Fault detection and diagnosis in low speed rolling element bearings part ii : The use of nearest neighbour classification. *Mechanical Systems and Signal Processing*, 1992:309–316, 6(4).

[14] K.M. Bossley, R.J. McKendrick, C. J. Harris et C. Mercer : Hybrid computed order tracking. *Mechanical Systems and Signal Processing*, 13(4):627–641, 1999.

[15] P. Dehombreux, D. Bain et O. Basile : Détection des défauts instationnaires dans les machines tournantes par suivi d'ordre implicite. *In receuil de conférences et communications du colloque international MAO 2001, p.A10/1-A10/9*, 2001.

[16] H. Oehlmann, D. Brie, M. Tomczak et A. Richard : A method for analysing gearbox faults using time-frequency representations. *Mechanical Systems and Signal Processing*, 11(4):529–545, 1997.

[17] J. K. Hammond et P. R White : The analysis of non-stationary signals using time-frequency methods. *Journal of Sound and Vibration*, 190(3):419–447, 1996.

[18] S. Mallat et W.L. Hwang : Singularity detection and processing with wavelets. *IEEE Transactions on Information Theory*, 38:617–643, 1992.

[19] G. Waeyenbergh et L. Pintelon : A framework for maintenance concept development. *International Journal of Production Economics*, 77:299–313, 2002.

[20] A.C. McCormick et A.K. Nandi : Classification of the rotating machine condition using artificial neural networks. *In Proceedings of IMeschE :PartC*, volume 211(6), pages 439–450, 1999.

[21] G. Vachtsevanos, F.L. Lewis, M. Roemer, A. Hess et B. Wu : *Intelligent fault diagnosis and prognosis for engineering systems*. John Wiley & Sons, 2006.

[22] V. Venkatsubramanian, R Rengaswamy, S.N. Kavuri et K. Yen : A review of process fault detection and diagnosis :process history based methods. *Computers and Chemical Engineering*, 27:327–346, 2003.

[23] S.A. Lapp et G.A. Powers : Computer-aided synthesis of fault-trees. *IEEE Trans.Reliability*, 37:2–13, 1977.

[24] G. Zwingelstein : *Diagnostic des défaillances :Théorie et pratique pour les systèmes industriels*. Traités des Nouvelles Technologie, série Diagnostic et Maitenance,Hermès, 1995.

[25] James C. Bezdek, Eric Chen-Kuo Tsao et Nikhil R. Pal : Fuzzy Kohonen clustering networks. *In Proc. IEEE Int. Conf. on Fuzzy Systems 1992*, pages 1035–1043, San Diego, 1992.

[26] E. Trauwert L. Kaufman et P. Rousseeuw : . fuzzy clustering algorithms based on the maximum likehood principle. *Fuzzy Sets and Systems computers*, 42:213–227, 1991.

[27] P.K. Simpson : Fuzzy min-max classification with neural networks. *J. Knowledge Engineering*, 4:1–9, 1991.

[28] Pekka TEPPOLA, Satu-Pia MUJUNEN et Pentti MINKKINEN : Adaptative fuzzy c-means clustering in process monitoring. *Chemometrics and intellligent laboratory ssystems*, 45:23–38, 1999.

[29] E. TRAUWERT, L. KAUFMAN et P. ROUSSEEUW : Fuzzy clustering algorithms based on the maximum likehood principle. *Fuzzy Sets and Systems*, 42:213–227, 1991.

[30] Mark A. SCHWABACHER : A survey of data-driven prognostics. *AIAA*, 2005-7002, 2005.

[31] J. LEE, J. NI, D. DJURDJANOVIC, H. QIU et H. LIAO : Intelligent prognostics tools and e-maintenance. *Computers in Industry*, 57:476–489, 2006.

[32] O. BASILE : *Prise en compte de l'incertitude dans les modèles fiabilistes de maintenance industrielle. Extension aux sollicitations variables.* Thèse de doctorat, Faculté Polytechnique de Mons, 2007.

Chapitre 3

Prétraitement de signaux vibratoires et extraction d'indicateurs

Pluralitas non est ponenda sine neccesitate.
(Le rasoir d'Occam)

Ce chapitre présente différentes méthodes de construction d'indicateurs de surveillance et de diagnostic de l'état d'un équipement sur base de mesures vibratoires. Le problème général de la représentation des données temporelles sous une forme exploitable par les algorithmes d'apprentissage y est abordé.

3.1 Introduction

Le succès et la qualité d'un système de diagnostic ou de pronostic repose dans la qualité des variables utilisées comme indicateurs. Ces variables doivent avoir un contenu informationnel suffisant pour pouvoir discriminer les différents scénarios de fonctionnement de l'équipement. Les données de mesures effectuées sur les équipements sont rarement utilisables dans leur forme brute. Par conséquent, les indicateurs sont extraits des signaux bruts mesurés par un processus de prétraitement qui peut s'effectuer dans le domaine temporel, dans le domaine fréquentiel ou dans les deux à la fois. Ce processus peut être vu comme une représentation de ces signaux dans un espace à N dimensions où la discrimination est meilleure.

3.2 Représentation des données temporelles

Par données temporelles, nous convenons d'entendre aussi bien les signaux vibratoires bruts relevés que les séries temporelles constituées des indicateurs extraits des signaux successivement mesurés.

La motivation du besoin de représentation des séries temporelles sous une forme autre que la forme originale tient du souci de représenter les caractéristiques de la série dans une forme plus concise. Un bénéfice additionnel qui doit en découler est la compression avec comme corollaire la rapidité du processus d'extraction de connaissances, le débruitage et la mise en évidence des caractéristiques importantes des données. Le problème de la représentation est étroitement lié à celui de l'extraction des indicateurs.

En *condition monitoring*, l'état d'un équipement est observé à travers un ensemble de variables d'état qui évoluent avec le temps. Ces variables sont appelées indicateurs ou attributs. Dans le cas particulier de la surveillance vibratoire, les indicateurs sont extraits des signaux de capteurs. Ces signaux sont par essence des séries temporelles. Chaque signal engendrera un vecteur d'indicateurs. Ces indicateurs, rangés dans l'ordre temporel d'enregistrement des signaux dont ils sont issus, forment une série temporelle multivariée sur laquelle les techniques de *data mining* devront être appliquées.

Le problème de la représentation devra être adressé à deux niveaux : au niveau de la série temporelle de base pour l'extraction des indicateurs et au niveau de la série temporelle multivariée faite de vecteurs d'indicateurs.

Dans les sections suivantes, nous présentons une taxinomie de la représentation des données temporelles inspirée de [1] et de [2]. On peut distinguer trois modes de représentation :

- la représentation basée sur la transformation (transformée de Fourier, transformée en ondelettes, analyse temps-fréquence) ;
- la représentation continue dans le domaine temporel (filtrage, analyse spectrale singulière) ;
- la représentation basée sur les modèles génératifs (ARMA,...).

3.3 Représentations basées sur la transformation

L'idée est de transformer le signal temporel initial afin de le représenter par les coordonnées d'un point dans un espace d'un autre domaine ou du même domaine temporel. On citera prin-

cipalement la représentation dans l'espace composé d'indicateurs statistiques issus du signal temporel et la représentation dans une base de fonctions. Pour ce dernier type, la transformée de Fourier discrète (DFT) et la transformée en ondelettes discrète (DWT) sont des exemples.

Dans le cas de la DFT, la base est composée d'exponentielles complexes qui représentent les fonctions sinus et cosinus. Pour la DWT, différentes bases de décomposition sont proposées : les ondelettes de Haar, de Daubechies, de Symlet et de Coiflets sont les plus utilisées.

On peut choisir pour la DFT, de n'utiliser que quelques coefficients de la décomposition. Si, par exemple, on utilise seulement les k premiers coefficients, cela revient à ne garder que les composantes basses fréquences de la série temporelle. Le signal est alors représenté par un point dans un espace de dimension k, ce qui correspond à une compression de données. Celle-ci sera d'une très faible erreur de reconstruction si réellement la série ne contient que des composantes basse fréquence.

Au lieu de choisir les k premiers coefficients des transformées, on peut utiliser les k plus grands coefficients. En DFT, on parle de fréquences dominantes.

3.3.1 La décomposition en ondelettes

L'analyse en ondelettes [3, 4] a été introduite en 1984 par Morlet pour l'analyse multirésolution d'un signal par des processus de dilatation et translation afin d'extraire des indicateurs temps-fréquence.

La transformée en ondelettes procède comme la transformée de Fourier par une décomposition du signal dans une base de fonctions. La transformée de Fourier utilise une base composée de fonctions harmoniques alors que la transformée en ondelettes utilise une base de fonctions oscillatoires définies localement et appelées ondelettes. Chaque ondelette est localisée en une position donnée sur l'axe de temps, et son caractère local tient du fait qu'elle décroît jusqu'à zéro quand on s'éloigne de son centre.

Transformée en ondelettes continue

Considérons un signal $x(t)$, sa transformée en ondelettes $W_x(\alpha,\beta)$ consiste à projeter le signal $x(t)$ sur une famille d'ondelettes déduites d'une fonction élémentaire (ondelette mère) par translation et dilatation :

$$W_x(\alpha,\beta) = \int_{-\infty}^{+\infty} x(t)\Psi^*_{\alpha,\beta}(t)dt \tag{3.1}$$

où

$$\Psi_{\alpha,\beta}(t) = \frac{1}{\sqrt{|\alpha|}}\Psi\Big(\frac{t-\beta}{\alpha}\Big) \tag{3.2}$$

(*) dénote le conjugué complexe, α et β étant respectivement les paramètres de dilatation et de translation.

La variable α correspond à un facteur d'échelle, en ce sens que $|\alpha| < 1$ dilate l'ondelette et $|\alpha| > 1$ la comprime. Par définition, la transformée en ondelettes est une représentation temps-échelle plutôt que temps-fréquence. La figure 3.1 illustre une analyse en ondelettes.

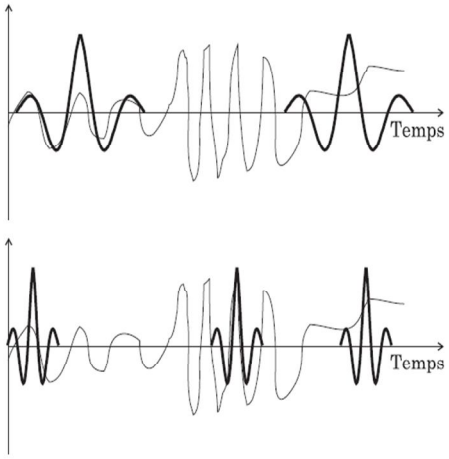


FIG. 3.1 – Illustration d'une analyse en ondelettes

Transformée en ondelettes discrète

Calculer les coefficients de la transformée en ondelettes à toutes les positions possibles aboutit à une quantité importante de données. C'est ainsi qu'on préfère le faire juste pour un ensemble de positions et d'échelles. Une solution intéressante est de choisir des échelles et positions dyadiques, c'est-à-dire basées sur les puissances de 2 ($\alpha = 2^j$ et $\beta = k\alpha$, $j,k \in Z$). Dans ce cas, l'ondelette est

$$\Psi_{j,k} = 2^{-j/2}\Psi(2^{-j}t - k) \tag{3.3}$$

Et la transformée discrète est donnée par

$$c_{j,k} = \int x(t)\Psi_{j,k}(t)dt \tag{3.4}$$

où les $c_{j,k}$ sont les coefficients d'ondelettes.

Une approche efficace pour implémenter cette décomposition en utilisant des filtres a été développée par Mallat [5]. L'étape de base de cette décomposition est illustrée sur la figure 3.2 où le signal est convolué avec un filtre passe-bas L et un filtre passe-haut H pour donner deux vecteurs de coefficients cA_1 et cD_1. Le vecteur cA_1 représente les coefficients relatifs à l'approximation et cD_1, ceux relatifs au détail. En effet, l'information de base d'un signal se trouve dans les basses fréquences. Le symbole $\downarrow 2$ représente un sous-échantillonnage. Le signal original S sera donné par

$$S = A_1 + D_1 \tag{3.5}$$

où A_1 est le signal d'approximation reconstruit à partir des coefficients cA_1, et D_1, le signal de détail reconstruit à partir de cD_1.

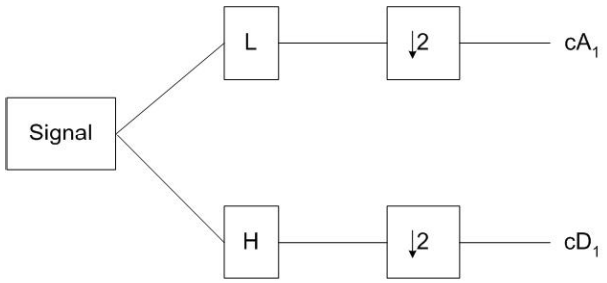


FIG. 3.2 – Décomposition de niveau 1 d'un signal en ondelettes

On peut procéder à une décomposition à plusieurs niveaux où chaque approximation à un niveau est décomposée en une autre approximation et un détail, on parle alors d'arbre de décomposition en ondelettes (voir figure 3.3). Dans ce cas, pour une décomposition de niveau J, on peut écrire :

$$S = A_J + \sum_{j=1}^{J} D_j(t) \tag{3.6}$$

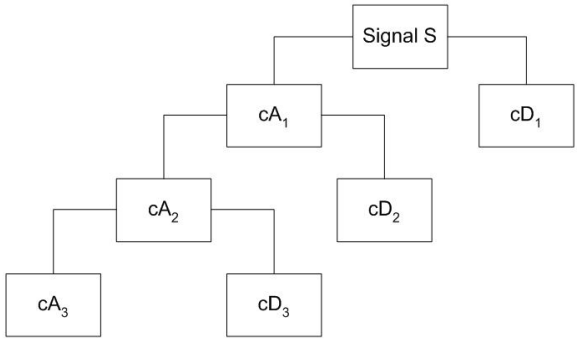


FIG. 3.3 – Décomposition de niveau 3 d'un signal en ondelettes

Analyse en paquets d'ondelettes

La décomposition en paquets d'ondelettes est une généralisation de la décomposition en ondelettes. Dans cette dernière, le passage d'un niveau de décomposition à un autre se fait par la décomposition du signal d'approximation alors que pour l'analyse en paquets, la décomposition concerne aussi bien le signal de détail que celui d'approximation.

Extraction d'indicateurs d'une analyse en ondelettes

Plusieurs manières d'extraire des indicateurs d'une analyse en ondelettes peuvent être imaginées. Une première famille de méthodes d'extraction part des signaux reconstruits pour appliquer toutes les méthodes définies dans le domaine fréquentiel et dans le domaine temporel. La décomposition en ondelettes suivie de la reconstruction est utilisée ici comme un filtrage. L'on peut imaginer des moyens de reconstruction après une modification de certains coefficients. La plupart des coefficients d'ondelettes sont très petits et peuvent ainsi être négligés avec un très faible risque d'erreur dans le signal. Ceci peut amener à l'obtention d'un ensemble réduit de coefficients qu'on peut utiliser comme indicateurs. On utilise un seuil pour déclasser les coefficients faibles. Si les coefficients d'ondelettes sont notés $c_{j,k}$, on aura pour un seuil fixé à θ :

$$A(c_{j,k}) = \begin{cases} c_{j,k} & c_{j,k} > \theta \\ 0 & c_{j,k} \leq \theta \end{cases} \tag{3.7}$$

où $A(c_{j,k})$ représente l'opération de seuillage. Ce procédé sur les coefficients d'un détail avant reconstruction est un processus de débruitage.

L'analyse en ondelettes suivie d'une synthèse est à classer parmi les représentations conti-

nues dans le domaine temporel.

La seconde famille de méthodes d'extraction d'indicateurs est basée sur les coefficients de la décomposition. Ce que l'on peut faire des coefficients d'une analyse en ondelettes n'est pas limité. On peut par exemple construire les indicateurs en procédant à la moyenne des coefficients d'ondelettes [6] ou en utilisant leur maximum au niveau de chaque nœud de décomposition. On peut aussi définir comme indicateurs les énergies correspondant à chaque nœud [7]. L'énergie est définie comme la somme des carrés des coefficients d'ondelettes.

3.3.2 Les représentations temps-fréquence

Il n'est pas rare que les signaux vibratoires relevés sur les équipements mécaniques soient instationnaires. Un signal instationnaire est un signal dont la structure statistique (moyenne, variance, corrélation,...) ou le contenu fréquentiel changent avec le temps . Dans ces conditions les méthodes qui permettent d'extraire l'information de l'évolution de la fréquence en fonction du temps s'imposent.

Les méthodes temps-fréquence peuvent être groupées en deux classes [8] :

- la première classe adopte une représentation linéaire. On reste attaché à une décomposition du signal vis-à-vis d'un ensemble de fonctions orthogonales élémentaires : spectre évolutionnaire de la transformée de Fourier à fenêtre glissante, représentation temps-échelle des ondelettes,...
- la seconde classe adopte une représentation quadratique. On cherche à repartir l'énergie totale du signal dans le domaine temps-fréquence : on définit une densité conjointe d'énergie $d_x(t, f)$ telle que

$$E_x = \int_{-\infty}^{+\infty} |x(t)|^2 \, dt = \int_{-\infty}^{+\infty} |X(f)|^2 \, df = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} d_x(t, f) dt \, df \tag{3.8}$$

  Les transformées de Wigner-Ville ou de Choï-Williams en sont les prototypes les plus fameux.

Ces méthodes peuvent être utilisées pour la construction des indicateurs de détection ou de diagnostic. Par exemple, la présence de variations des fréquences avec le temps peut être reconnue par la détection et l'estimation d'une ligne sur l'image de représentation temps-fréquence. En reconnaissance de formes, ce type de problème est résolu en utilisant la transformée de Hough dont les détails peuvent être trouvés dans la référence [9]. La figure 3.4 illustre la question de compromis entre résolution temporelle et résolution fréquentielle dans les méthodes

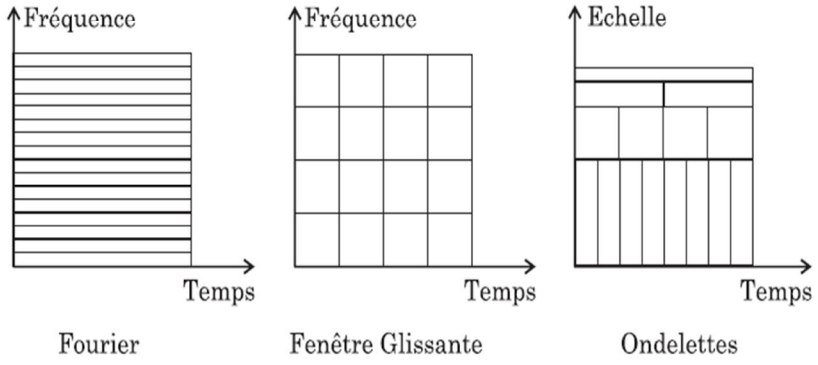


FIG. 3.4 – Compromis entre résolution temporelle et résolution fréquentielle

temps-fréquence.

3.4 Représentation continue dans le domaine temporel

En surveillance vibratoire, le signal temporel vibratoire original ne permet généralement pas l'identification directe et aisée des similarités ou des changements dans ses motifs.

La représentation continue dans le domaine temporel peut concerner l'entièreté ou une partie du signal. Dans le premier cas, elle s'effectue au moyen d'un filtrage ou d'un autre processus qui décompose le signal puis le reconstruit dans le domaine temporel, et dans le second, elle fait usage de la segmentation du signal, à l'aide de fenêtres glissantes pondérées ou non, dans le but d'améliorer la découverte des particularités dans la série. Keogh [10] a proposé de projeter la série dans un espace où ses coordonnées sont les moyennes des valeurs sur des segments consécutifs de longueurs fixées. Cette façon de procéder a été appelée *Piece-wise Aggregate Approximation* (PAA) ou moyenne segmentée. Si la fenêtre glissante est d'une longueur très faible, cette projection peut encore être considérée comme une représentation continue.

Les traitements qui consistent en une représentation continue dans le domaine temporel sont généralement utiles pour des fins de débruitage.

Dans ce travail, une méthode similaire à la moyenne segmentée a été utilisée pour la représentation des signaux dans l'application concernant la détection et l'identification des défauts de roulements. Dans l'espace de projection utilisé, les coordonnées sont les valeurs efficaces calculées sur une fenêtre temporelle glissante de forme exponentielle négative. La valeur efficace calculée pour chaque fenêtre est assignée au point courant qui correspond au début de la fenêtre. La fenêtre exponentielle a pour effet d'atténuer sensiblement les contributions des

points éloignés du point courant. Le signal obtenu peut être considéré à un facteur près comme un signal enveloppe du signal d'origine.

Nous avons porté un intérêt particulier à deux processus de représentation continue qui procèdent par une décomposition et une reconstruction. Il s'agit de la transformation en ondelettes et de l'analyse spectrale singulière. La transformation en ondelettes a été détaillée dans la section précédente concernant les méthodes de représentation basées sur la transformation. Nous présentons, ici, l'analyse spectrale singulière et les différents moyens d'en extraire des indicateurs pour le besoin du diagnostic.

L'analyse spectrale singulière (ASS)

L'analyse spectrale singulière est une technique non paramétrique d'analyse de séries temporelles. Elle permet la décomposition d'un signal en signaux indépendants, mais dont la somme donne le signal d'origine.

Le point de départ de l'ASS est la construction d'une matrice dite « de trajectoire » qui est composée de vecteurs de dimension m extraits de la série temporelle.

Soit $\{x_t | t = 1, ..., n\}$ une série temporelle de longueur n. La séquence des vecteurs $\tilde{\mathbf{x}}_t$ de dimension m est obtenue en traversant, avec recouvrement, la série de données à l'aide d'une fenêtre glissante de longueur m. On obtient ainsi

$$\tilde{\mathbf{x}}_t = (x_t, x_{t+1}, ..., x_{t+m-1}) \tag{3.9}$$

Les vecteurs $\tilde{\mathbf{x}}_t$ sont indexés par $t = 1, ..., n'$, où $n' = n - m + 1$. Ils vont intervenir dans la constitutions des colonnes de la matrice de trajectoire $\mathbf{T} = \left[\tilde{\mathbf{x}}_1^T, \tilde{\mathbf{x}}_2^T, \cdots, \tilde{\mathbf{x}}_{n'}^T\right]$,

$$\mathbf{T} = \begin{pmatrix} x_1 & x_2 & \cdots & x_{n-m+1} \\ x_2 & x_3 & \cdots & x_{n-m+2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m-1} & x_m & \cdots & x_{n-1} \\ x_m & x_{m+1} & \cdots & x_n \end{pmatrix} \tag{3.10}$$

La matrice construite est telle que les éléments des diagonales montantes sont identiques. Le choix de la longueur de la fenêtre est un compromis entre deux considérations : la quantité d'information extraite et le degré de confiance statistique en cette information. Le premier aspect

requiert une fenêtre aussi large que possible alors que le second aspect exige plusieurs répétitions de fenêtre, c'est-à-dire un rapport n/m aussi grand que possible [11, 12]. Wang *et al.* [13], qui ont utilisé pour la première l'ASS en diagnostic des machines tournantes, ont discuté de l'effet de la longueur de la fenêtre glissante et du recouvrement lors de son application.

En analyse vibratoire, le choix de la longueur de la fenêtre peut se voir imposer une borne inférieure en fonction de la résolution fréquentielle désirée. Si l'on désire, par exemple, pouvoir observer des phénomènes se produisant au minimum à une fréquence f [14], la longueur minimale de la fenêtre doit être telle que

$$m \geq \frac{f_{\text{ech}}}{f} \tag{3.11}$$

où f_{ech} est la fréquence d'échantillonnage du signal.

On procède, ensuite, à une projection de la série temporelle dans l'espace des vecteurs propres de la matrice **S**, de covariance de **T**, et dont les éléments peuvent être calculés directement à partir des données par :

$$s_{ij} = \frac{1}{n - |i - j|} \sum_{t=1}^{n-|i-j|} x_t x_{t+|i-j|} \quad i = 1...n,\ j = 1...m \tag{3.12}$$

Cette projection donne les composantes principales a^k.

$$a_t^k = \sum_{j=1}^{n} x_{t+j-1} \rho_j^k \tag{3.13}$$

où ρ_j^k représente les composantes des vecteurs propres de la matrice **S** des covariance de **T**.

On peut alors reconstruire la partie de la série temporelle associée à un seul ou à plusieurs vecteurs propres en combinant les composantes principales correspondantes :

$$y_t^k = \frac{1}{M_t} \sum_{k \in D} \sum_{j=L_t}^{U_t} a_{t-j+1}^k \rho_j^k \tag{3.14}$$

où D représente les indices des composantes principales choisies pour la reconstruction. La valeur du facteur de normalisation M_t, et celles de L_t et U_t sont déterminées par l'algorithme dit

de « moyennage diagonal » [15] :

$$(\frac{1}{M_t}, L_t, U_t) = \begin{cases} (t, 1, t), & 1 \leq t \leq m-1 \\ (m, 1, m), & m \leq t \leq n' \\ (n-t+1, t-n+m, m), & n'+1 \leq t \leq n \end{cases} \tag{3.15}$$

On peut envisager plusieurs manières d'extraire des vecteurs d'indicateurs à partir d'une analyse spectrale singulière.

Analyse spectrale singulière combinée à un filtrage

On applique l'ASS sur les signaux mesurés, et les signaux reconstruits peuvent être filtrés dans des gammes de fréquences bien choisies. Les méthodes du domaine fréquentiel et/ou du domaine temporel peuvent ensuite appliquées.

Analyse spectrale singulière en cascade. Semi-décomposition en arbre

Au lieu d'effectuer une ASS suivie d'un filtrage, nous proposons une autre possibilité qui consiste en une ASS en cascade. Cette approche est similaire à la décomposition en ondelettes. Partant d'un signal brut, on construit deux signaux, l'un approchant le signal de départ et l'autre -le résidu- représentant ce que l'on peut qualifier de bruit. Soit, par exemple, X(t) le signal brut de départ, la première décomposition s'écrit

$$X(t) \approx RC_1(t) + RE_1(t) \tag{3.16}$$

où $RC_1(t)$ est l'approximation reconstruite du signal et $RE_1(t)$ le résidu.

On procède ensuite de manière récurrente à la décomposition de signaux approchés reconstruits jusqu'à un niveau J de décomposition. Il vient alors que le signal brut, pour une décomposition d'ordre J, s'écrit

$$X(t) \approx RC_J(t) + \sum_{j=1}^{J} RE_j(t) \tag{3.17}$$

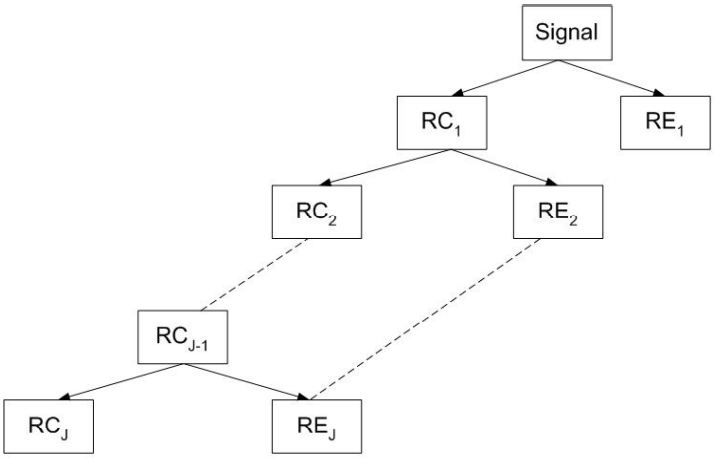


FIG. 3.5 – Analyse spectrale singulière en cascade. Semi-décomposition en arbre

Les indicateurs peuvent alors être définis en se basant sur les niveaux énergétiques dans chaque composante. La somme de raies spectrales peut être utilisée comme mesure de niveau énergétique. Les indicateurs scalaires statistiques définis ci-haut peuvent être utilisés pour compléter la carte de représentation du signal.

Analyse spectrale singulière en cascade. Décomposition en arbre complet

Plutôt que de procéder à la décomposition des seuls signaux approchés, on peut également décomposer tous les signaux résiduels jusqu'au niveau J. Ce qui donnerait une décomposition du signal en des bandes de fréquence de largeur constante, alors que la cascade de semi-décomposition en arbre donne une largeur de bande de plus en plus fine quand j passe de 1 à J.

3.5 Représentations basées sur les modèles

On peut supposer que la série est engendrée par un modèle génératif. Le modèle est ajusté à la série et les paramètres du modèle sont utilisés comme indicateurs. Certains travaux ont utilisé les modèles cachés de Markov (HMM) [16] et les modèles ARMA [17]. Cette thèse n'insistera pas sur ces modes de représentation. Le lecteur désireux d'approfondir ces aspects pourra consulter [18] en plus des références citées ci-dessus.

3.6 Fusion des données

Les données à fournir au système de diagnostic ne doivent pas nécessairement être du même type ou provenir d'une même source. La combinaison de sources aide généralement à l'amélioration de la qualité de l'information à tirer des données. On parle de fusion des données. Lucien Wald la définit comme « un cadre formel dans lequel s'expriment les moyens et techniques permettant l'alliance des données provenant de sources diverses » [19]. La fusion peut s'effectuer à trois niveaux différents :

- niveau 1 : fusion directe des données de capteurs (figure 3.6) ;
- niveau 2 : fusion après extraction des indicateurs (figure 3.7) ;
- niveau 3 : fusion des décisions.

À son niveau le plus bas (niveau 1), la fusion est utilisée pour combiner les informations provenant d'un réseau de capteurs afin de valider les signaux et construire les indicateurs. Un exemple de ce type de fusion est la combinaison d'un signal de vitesse et d'un signal vibratoire pour effectuer un moyennage synchrone.

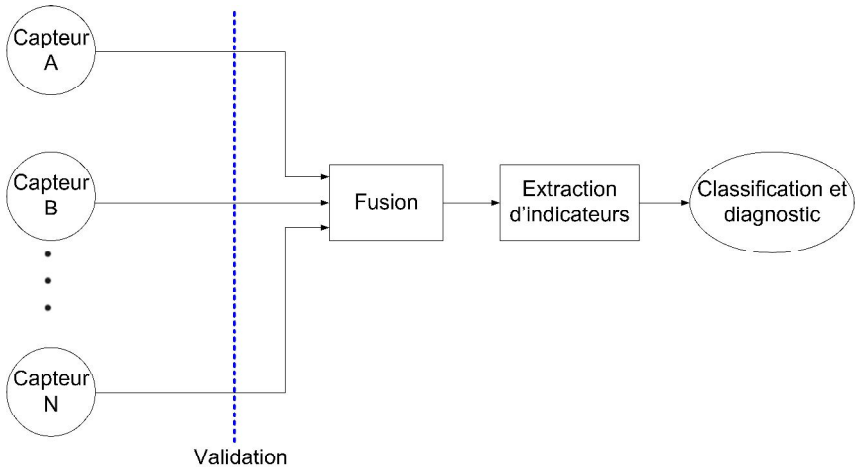


FIG. 3.6 – Fusion de niveau 1

Au niveau 2, la fusion combine les indicateurs d'une manière intelligente pour obtenir la meilleure information possible pour le diagnostic. On peut, par exemple, combiner un indicateur vibratoire et la concentration des particules dans le lubrifiant pour le suivi d'un palier.

Au niveau le plus élevé (niveau 3), on procède à la fusion des connaissances pour incorporer une information basée sur l'expérience telle que les taux de défaillance ou un modèle physique avec une information basée sur le signal.

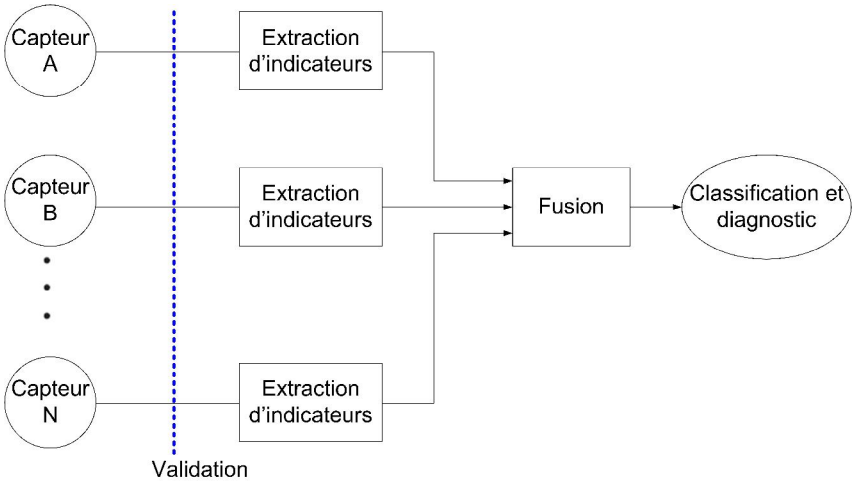


FIG. 3.7 – Fusion de niveau 2

Il importe, cependant, de noter qu'un danger possible de la fusion est d'aboutir à une plus faible capacité de discrimination. En effet, un mélange de bons indicateurs avec des variables de très pauvre contenu informationnel ne peut qu'affecter le résultat final.

Dans les applications de surveillance vibratoire, il est fréquent de fusionner des informations provenant de capteurs différents. La fusion la plus utilisée est celle de niveau 2, où l'on procède à une concaténation des indicateurs calculés sur différents signaux pour constituer un vecteur d'indicateurs de diagnostic. Les différents signaux peuvent aussi provenir d'un même signal de départ après quelques prétraitements tels que le filtrage, l'ASS ou l'analyse en ondelettes. Il peut arriver que l'on aboutisse à des vecteurs d'indicateurs de grande dimension. Dans ce cas, une réduction dimensionnelle est conseillée afin de rendre la phase d'apprentissage aisée.

Des plus amples informations sur la fusion de données peuvent être obtenues dans [20].

3.7 Réduction dimensionnelle de l'espace d'indicateurs

C'est seulement quand le nombre d'observations, M, est largement supérieur au nombre d'indicateurs, N, que l'on peut espérer avoir une bonne capacité de généralisation pour le classifieur. Le rapport M/N est appelé rapport de dimensionnalité.

Le choix de ce rapport a été étudié par plusieurs auteurs. Nous pouvons retenir qu'en général un rapport de dimensionnalité au moins égal à 3 est recommandé [21–24].

Deux procédures de réduction dimensionnelle sont souvent utilisées : l'analyse en composantes principales et la sélection des indicateurs vibratoires pertinents.

3.7.1 L'analyse en composantes principales (ACP)

L'ACP est une méthode de transformation des données qui consiste à trouver une base orthogonale de l'espace des indicateurs, dont les dimensions sont déterminées par les directions suivant lesquelles les données changent rapidement. L'ACP permet donc d'abord de passer d'un ensemble de vecteurs $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_M$ repris dans la matrice $\mathbf{X}_{M\times N}$ à une matrice $\mathbf{Y}_{M\times N}$ composée de vecteurs $\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_M$. Les composantes des vecteurs $\mathbf{x}_i$ sont les variables d'origine et celles des vecteurs $\mathbf{y}_i$, les facteurs ou scores factoriels. Il importe que les nouvelles variables ne présentent aucun degré de redondance dans l'information qu'elles véhiculent. Il s'agira ensuite de ne retenir pour les vecteurs $\mathbf{y}_i$ que les éléments qui répondront à un critère informationnel.

L'ACP procède souvent à cette transformation de la façon la plus simple possible, c'est-à-dire linéairement. Les facteurs sont, en effet, construits comme des combinaisons linéaires des variables et réciproquement. Dans ce contexte linéaire, la condition de non-redondance de la liste de facteurs s'exprime alors par la condition de non-corrélation des facteurs entre eux. L'idée de base dans une ACP est qu'une grande information correspond à une grande variance. Il faudra transformer $\mathbf{X}$ en $\mathbf{Y} = \mathbf{X}.\mathbf{A}$, en choisissant $\mathbf{A}$ de telle manière que $\mathbf{Y}$ ait la plus grande variance possible. La matrice $\mathbf{A}_{N\times N}$ contient les coefficients des combinaisons transformant les variables de départ en facteurs. Ses colonnes sont les vecteurs propres de Σ, cette dernière étant la matrice de covariance de $\mathbf{X}$ [25].

On montre que la direction de variance maximum est parallèle au vecteur propre correspondant à la plus grande valeur propre de la matrice Σ. Les vecteurs propres de Σ peuvent donc servir à la construction de nouveaux axes vus comme une rotation du système original. La variance liée au premier axe sera la plus grande, elle sera suivie de celle du second axe et ainsi de suite.

En pratique la matrice Σ n'est pas connue, on utilise alors la matrice de covariance $\mathbf{S}$ calculée sur l'échantillon.

$$\mathbf{S} = \frac{1}{M-1}\mathbf{X}^T\mathbf{X} \tag{3.18}$$

Les vecteurs $\mathbf{y}_i$ sont caractérisés par le fait que la plus grande partie du contenu informationnel est stockée dans les quelques premières composantes seulement. On peut donc ne considérer qu'un nombre relativement réduit de facteurs pour décrire les données. Supposons que l'on n'ait gardé que P facteurs sur les N, la matrice de transformation serait alors $\mathbf{E}_{N\times P}$ au lieu de $\mathbf{A}_{N\times N}$.

Il vient que

$$\mathbf{Y} = \mathbf{X}.\mathbf{E} \tag{3.19}$$

3.7.2 Normalisation

Le processus d'apprentissage peut être plus efficace si certains prétraitements sont effectués sur les entrées et sur les cibles (sorties). La normalisation est un traitement qui permet d'éviter que les indicateurs qui ont de grandes valeurs numériques n'aient d'influence plus grande que les autres. La normalisation s'impose encore davantage si les indicateurs n'ont pas tous la même unité physique.

Un autre avantage de la normalisation se trouve dans l'allègement de la tâche d'apprentissage qui se traduit par une réduction sensible du temps d'apprentissage. Deux manières de normaliser les données sont possibles :

1. **la normalisation proprement dite** : elle est linéaire.
   Les entrées et les sorties sont toutes remises dans la plage [0,1] ou [-1, 1]. Le vecteur normalisé $\mathbf{x}_n$ est obtenu du vecteur original $\mathbf{x}$ par

$$\mathbf{x}_n = \frac{[\mathbf{x} - \min(\mathbf{x})]}{[\max(\mathbf{x}) - \min(\mathbf{x})]} \tag{3.20}$$

   ou par

$$\mathbf{x}_n = 2\frac{[\mathbf{x} - \min(\mathbf{x})]}{[\max(\mathbf{x}) - \min(\mathbf{x})]} - 1 \tag{3.21}$$

   $\min(\mathbf{x})$ et $\max(\mathbf{x})$ sont respectivement le plus petit et le plus grand élément de $\mathbf{x}$

2. **la standardisation :**
   Cette approche agit sur l'ensemble d'apprentissage de telle manière qu'après le traitement, la moyenne soit nulle et l'écart-type unitaire. Le vecteur standardisé $\mathbf{x}_n$ est obtenu du vecteur original $\mathbf{x}$ par

$$\mathbf{x}_n = \frac{[\mathbf{x} - moy(\mathbf{x})]}{std(\mathbf{x})} \tag{3.22}$$

   où $moy(\mathbf{x})$ est la moyenne de $\mathbf{x}$ et $std(\mathbf{x})$, son écart-type.

Dans l'application de l'ACP, la normalisation est un préalable. Cela s'impose davantage si les données de départ ont des unités de mesure différentes.

3.7.3 Saturations, communalités et variance expliquée

La relation entre les variables brutes ou standardisées et les facteurs est explicitée dans une matrice **L**, dite de structure ou des saturations (*factor matrix*) dont les composantes sont notées $L_{j\alpha}$ (*factor loadings*). Les $L_{j\alpha}$ sont les corrélations entre les variables brutes ou standardisées et les facteurs.

La matrice des saturations **L** a les propriétés suivantes :

1. la somme des carrés de toutes les saturations pour une variable donnée vaut 1. Cette somme est appelée communalité.

$$h_j^2 = \sum_{\alpha=1}^{N} L_{j\alpha}^2 = 1 \qquad (j = 1, ..., N)$$

   Comme souvent, l'objectif de l'ACP est la condensation de l'information, on ne conservera pas l'ensemble des N facteurs, mais seulement les P premiers ($P < N$), ceux dont la variance λ_α est la plus grande. De cette façon, les communalités sont alors inférieures à l'unité. De fait, h_j^2 mesure la proportion de variance de la variable j expliquée par les P premiers facteurs, et constitue une mesure locale de la qualité d'une ACP, variable par variable ;

2. la somme des carrés de toutes les saturations pour un facteur donné vaut λ_α, sa variance.

$$\sum_{j=1}^{N} L_{j\alpha}^2 = \lambda_\alpha \qquad (\alpha = 1, ..., P \text{ ou } N)$$

3.7.4 Choix de la dimension effective

Le choix du nombre de facteurs à retenir est un compromis à adopter entre les objectifs opposés de « conservation de l'information » et « de simplification de l'information ». Il existe plusieurs façons de déterminer le nombre de composantes requis pour garder une représentation adéquate des données. Nous en citerons quelques unes :

1. proportion expliquée de la variance par les P premiers facteurs : cette méthode propose de sommer les variances des P premières composantes principales et de diviser cette

somme par la somme de toutes les variances. En général, les composantes qui donnent une proportion d'au moins 75% sont suffisantes pour représenter les données [25] :

$$\frac{\sum_{\alpha=1}^{P} \lambda_\alpha}{\sum_{\alpha=1}^{N} \lambda_\alpha} \geq 0,75 \tag{3.23}$$

2. proportion expliquée de la variance par le facteur α : on sélectionne les composantes principales qui expliquent un certain pourcentage de la variance totale, par exemple 2% au minimum :

$$frac_var(\alpha) = \frac{\lambda_\alpha}{\sum_{\alpha=1}^{N} \lambda_\alpha} \geq seuil \tag{3.24}$$

3. moyenne des variances : la direction α sera considérée sans importance si $\lambda_\alpha < \bar{\lambda}$, où $\bar{\lambda} = \frac{1}{N} \sum_{\alpha=1}^{N} \lambda_\alpha$ est la moyenne des variances ou simplement ce que serait la variance si toutes les directions étaient également importantes. C'est le critère de Guttman-Kaiser.

4. *scree test* : on utilise, ici, le graphique des λ_α en fonction de α. Là où la courbe devient plate, c'est là que la variation devient aléatoire, donc sans structure. On ne gardera alors que les valeurs propres apparaissant avant le coude (critère de Cattell).

3.7.5 Les rotations en analyse en composantes principales

Les facteurs sont des abstractions mathématiques et ne possèdent aucune signification intuitive. Leur interprétation n'est pas souvent aisée, mais elle peut être sérieusement facilitée si les saturations sont soit grandes (c'est-à-dire proches de 1), soit proches de zéro. Dans ce cas on saurait à quelle(s) variable(s) correspondrait principalement un facteur.

Pour ce faire, on peut procéder à une nouvelle transformation par rotation des facteurs pour avoir chacune des variables associée à un seul facteur. Plusieurs algorithmes ont été développés pour l'obtention de la matrice de rotation, les plus utilisés portent les noms de *varimax*, *quartimax*, ou *equamax* [26, 27].

L'ACP, par ses différentes qualités, s'impose donc dans beaucoup de cas comme prétraitement des données destinées à l'alimentation d'un système de diagnostic. Elle permet aussi l'analyse des contributions des indicateurs sur les performances de classification.

3.8 Application : usure d'outils en tournage

Pour illustrer l'exploitation de l'analyse spectrale singulière, nous présentons un exemple qui concerne la surveillance de l'usure d'outils en tournage [28, 29]. Les détails sur l'essai expérimental sont présentés dans l'annexe B.

L'étude ne considère, dans un premier temps, que le signal relevé suivant la direction tangentielle, donc parallèle à l'effort de coupe pour la simple raison que l'accélération mesurée y a la plus grande intensité.

Prétraitement et extraction d'indicateurs par analyse spectrale singulière

Les signaux ayant fait l'objet de cette étude sont très longs et représentent des cycles de coupe d'une durée moyenne de 20 secondes. L'échantillonnage ayant été réalisé à 51,2 kHz, la longueur moyenne des signaux est de 10^6 points. La figure 3.8 illustre un signal vibratoire typique et démontre l'effet de l'usinage conique, tandis que la figure 3.9 montre que l'essentiel de l'information se trouve contenu dans les composantes fréquentielles en-dessous de 6000 Hz. Ceci suggère un sous-échantillonnage par 4 du signal, mais avant cela le signal est filtré passe-bas afin de prévenir l'aliasing. Cette diminution du nombre d'échantillons est effectuée essentiellement dans le but de réduire le temps de calcul lors des traitements subséquents.

Pour tenir compte de la non-stationnarité du processus de coupe et afin de contourner la difficulté liée au temps de calcul lorsque l'on manipule de longs signaux, une technique similaire à l'ASS locale a été utilisée [30]. Elle a consisté à procéder localement à l'ASS sur des segments de signal de longueur w. Ces fenêtres de longueur w sont définies sans recouvrement sur le signal d'origine. La longueur de la fenêtre glissante de base pour l'ASS est m alors que la longueur du signal original est n, ces longueurs étant telles que $m < w < n$. Le signal reconstruit est obtenu à partir des portions reconstruites localement. La figure 3.10 synthétise tout le processus de traitement avant l'alimentation du classifieur.

Extraction d'indicateurs

Pour l'ASS, le signal original est d'abord découpé en 50 segments pour permettre d'effectuer l'ASS sur des portions de signal de longueur environ égale à 5000 points. Ce qui a l'avantage de rendre la matrice de trajectoire facile à stocker lors des calculs. Ayant choisi de pouvoir isoler des phénomènes se produisant au moins à 250 Hz, la fenêtre d'analyse ASS doit avoir

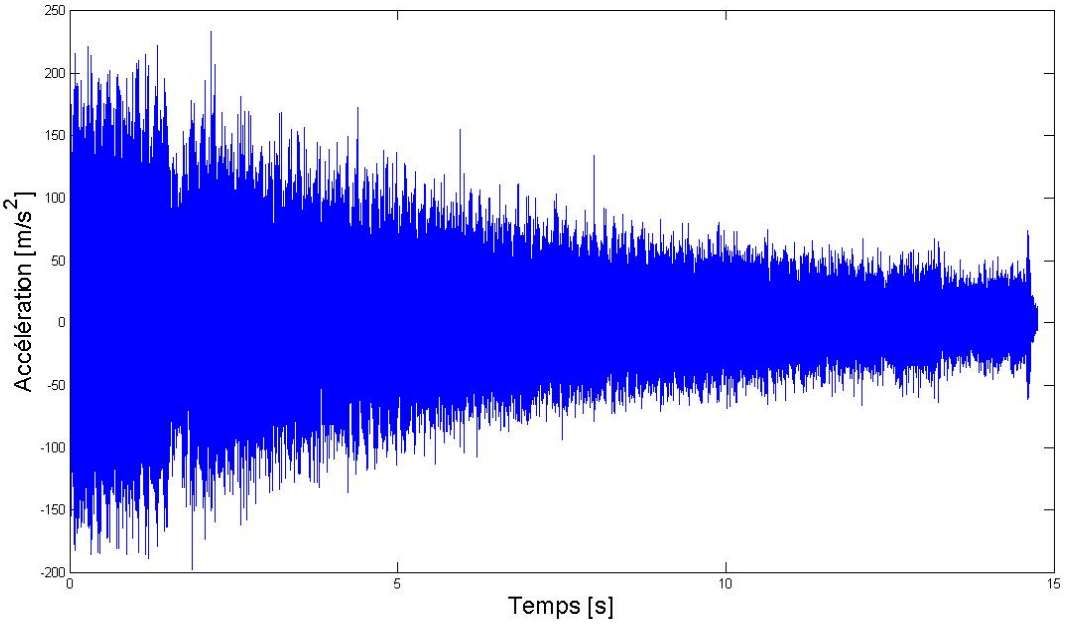


FIG. 3.8 – Signal typique enregistré lors d'un cycle de coupe

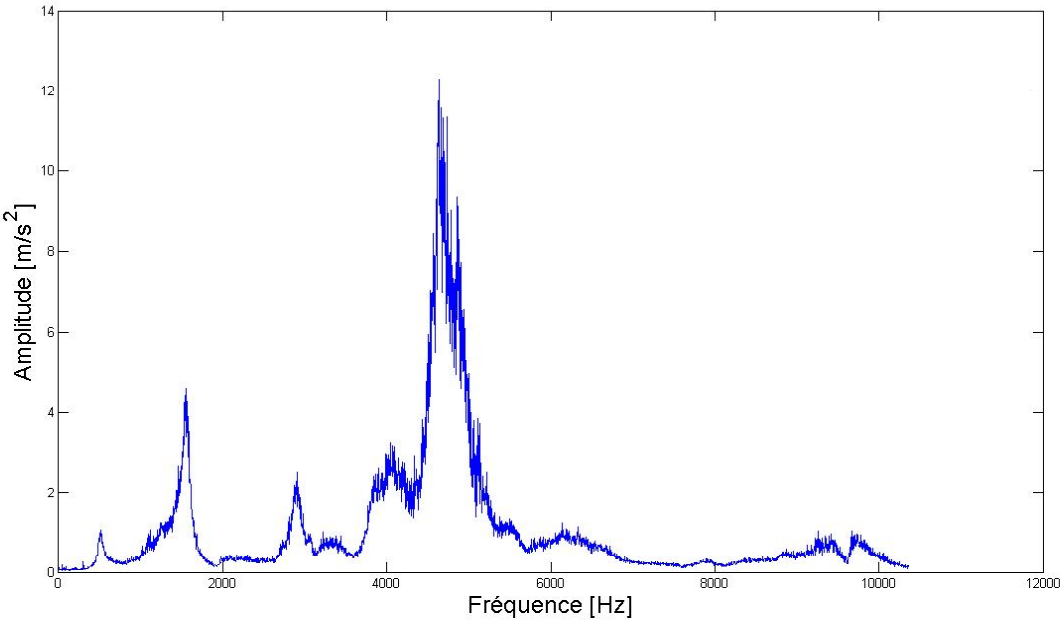


FIG. 3.9 – Spectre d'un signal typique

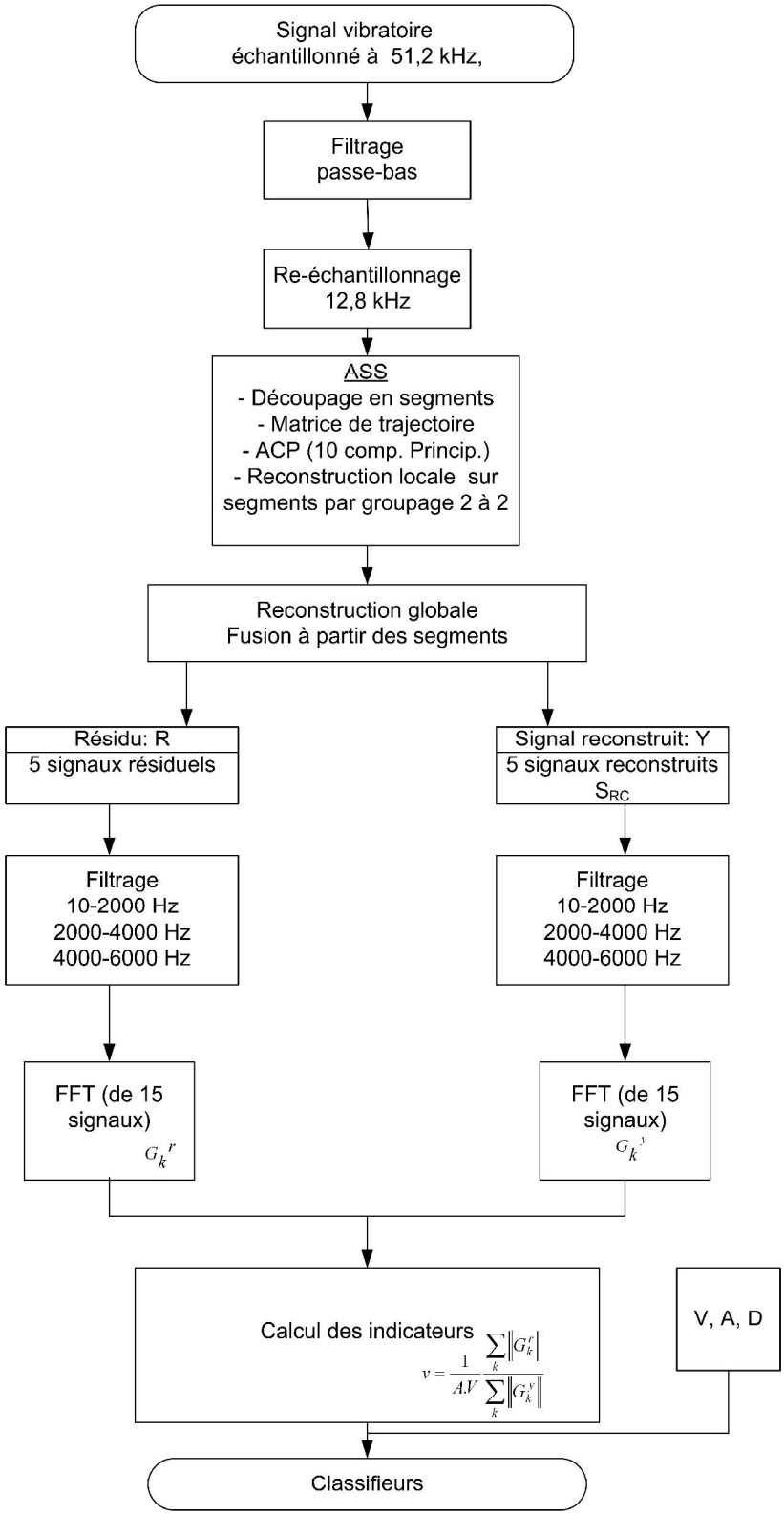


FIG. 3.10 – Processus de traitement de données

une durée de 4 ms, soit 50 points à la fréquence d'échantillonnage de 12500 Hz (après sous-échantillonnage).

On peut remarquer sur la figure 3.11, qu'un *scree test* permet de retenir les 10 premières valeurs propres. La reconstruction sera basée sur seulement les 10 premières composantes principales. Et ces composantes seront groupées 2 par 2 du fait que la contribution moyenne à la variance totale est presque la même pour 2 valeurs propres successives. Nous obtenons 5 signaux reconstruits par signal original.

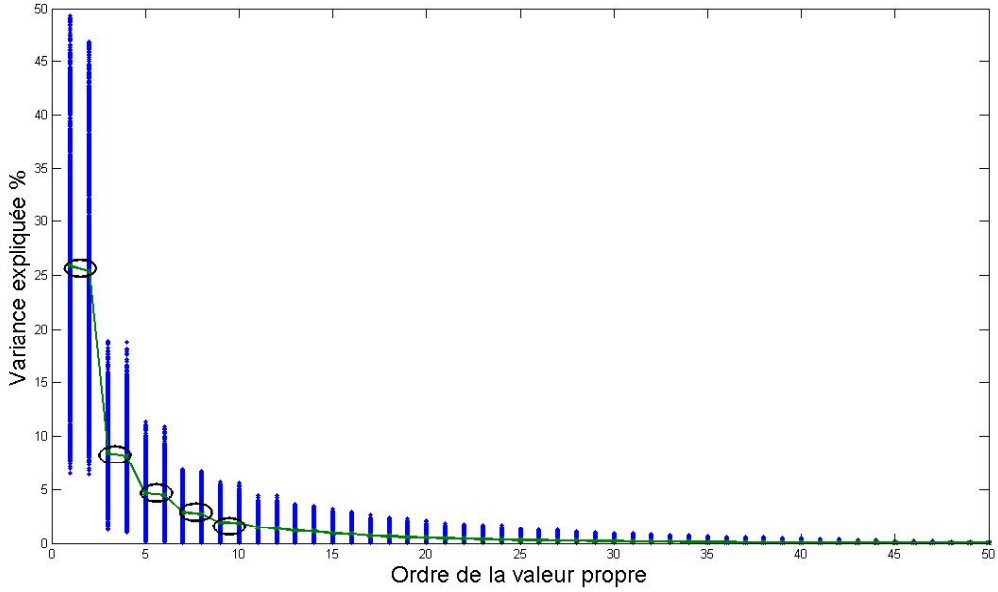


FIG. 3.11 – Contribution moyenne à la variance totale pour chaque valeur propre

Pour chaque signal reconstruit, noté S_{RC}, on évalue aussi son résidu c'est-à-dire sa différence au signal d'origine. L'objectif est de définir des indicateurs à partir de la notion de rapport signal/bruit, en considérant que le résidu représenterait le bruit si toute l'information utile du signal d'origine était contenue dans la composante reconstruite. Les figures 3.12 et 3.13 illustrent, pour un enregistrement, le temporel et le spectre fréquentiel des signaux reconstruits et de leurs résidus pour le cas de S_{RC1} et S_{RC5}. La reconstruction de S_{RC1} est basée sur les deux premières des 10 composantes principales, alors que celle de S_{RC5} est basée sur les deux dernières.

Par la suite, chaque signal reconstruit est filtré, à l'aide d'un filtre de Butterworth, dans les 3 gammes de fréquence suivantes 10-2000 Hz, 2000-4000 Hz et 4000-6000 Hz . Cette phase de traitement est proposée afin de construire des indicateurs selon les gammes de fréquences qui

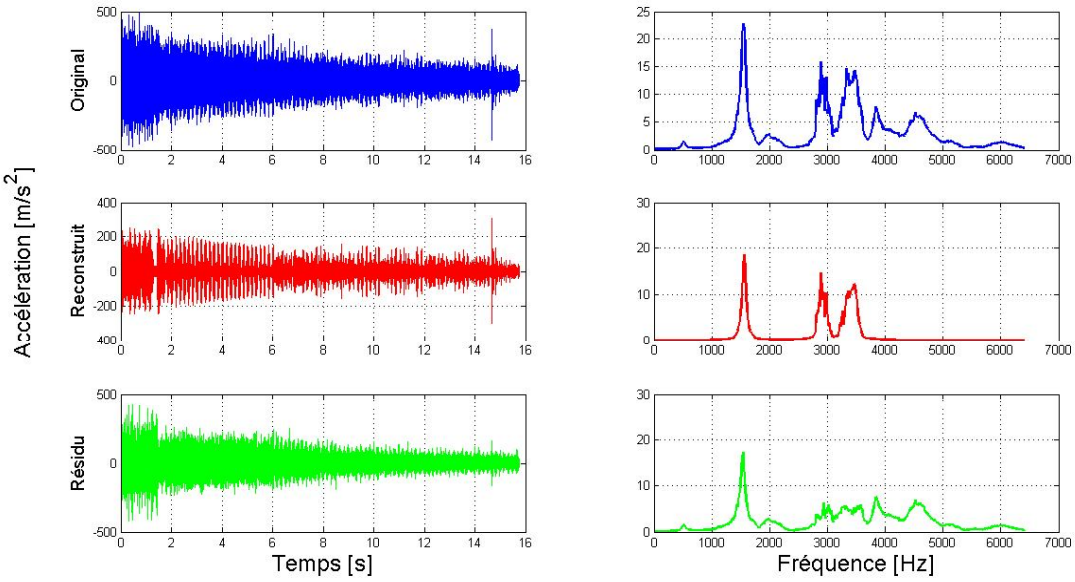


FIG. 3.12 – Signaux et spectres relatifs à la composante S_{RC1} d'un signal typique

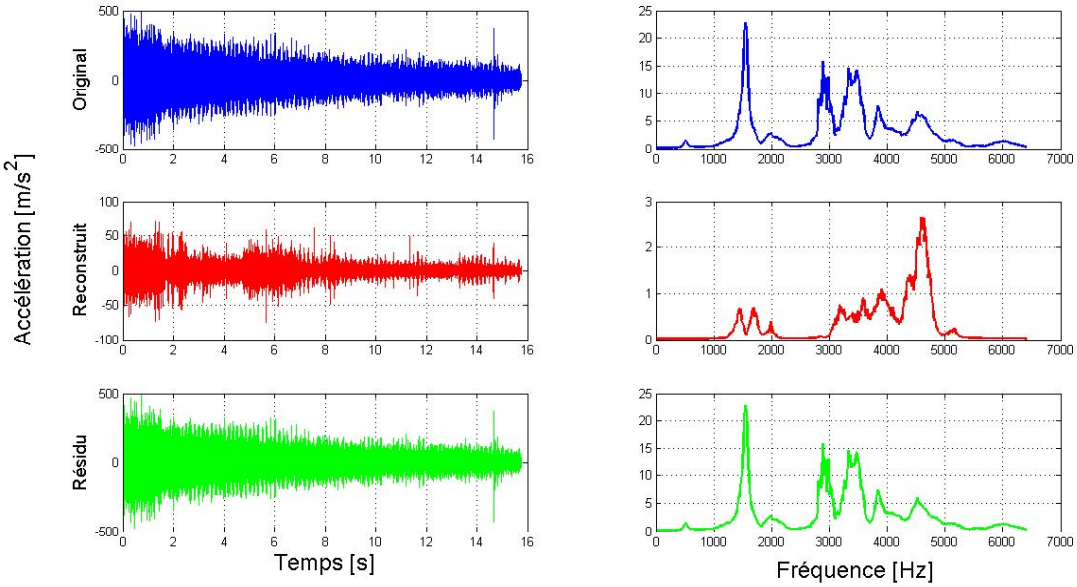


FIG. 3.13 – Signaux et spectres relatifs à la composante S_{RC5} d'un signal typique

s'imposent par la configuration même du contenu spectral des signaux traités (voir par exemple la figure 3.9).

Dans chaque gamme de fréquence, les indicateurs v extraits dans le but de la classification sont calculés de la manière suivante : si $\{y_n\}$ représente la série du signal reconstruit et $\{r_n\}$ celle du résidu correspondant ($n = 1 \cdots N$)

$$v = B\frac{\sum_k \left\|G_k^r\right\|}{\sum_k \left\|G_k^y\right\|} \tag{3.25}$$

où G_k^r et G_k^y sont respectivement les coefficients des transformées de Fourier de $\{r_n\}$ et $\{y_n\}$. B est un facteur qui permet une normalisation par l'avance A et la vitesse de coupe V, et il est donné par

$$B = \frac{1}{A.V} \tag{3.26}$$

Ce qui aboutit à 15 variables pour chaque signal brut relevé sur le porte-outil (3 bandes fréquentielles pour chacune des 5 reconstructions du signal). Les indicateurs extraits v seront notés RC. Parmi ces 15 variables, certaines semblent suivre fidèlement l'évolution de l'état de l'outil et l'apparition de certains phénomènes tels que le bourrage copeau ou un passage à vide de l'outil. Sur la figure 3.14 qui représente les variables RC1 4000-6000 Hz et RC5 4000-6000 Hz normalisées pour les 43 essais, on remarquera que RC5 dans la gamme 4000-6000 Hz traduit l'évolution de l'usure de l'outil alors que les pics de RC1 dans la même gamme sont liés à un passage à vide. En effet, la classe N est caractérisée par de faibles valeurs de RC5 4000-6000 Hz tandis que U2 correspond à ses grandes valeurs. La figure 3.15 représente un histogramme conditionnel des distributions de l'indicateur RC5 4000-6000 Hz étant données les classes.

Pour tenir compte d'une éventuelle dépendance de l'usure vis-à-vis des paramètres de coupe, à ces 15 variables sont ensuite ajoutés l'avance, la vitesse de coupe et le diamètre du manchon.

Les capacités de discrimination des indicateurs extraits par cette procédure seront comparées à celles des indicateurs classiques et peu coûteux tels que la valeur efficace (chapitre 6).

3.9 Conclusion

Ce chapitre a présenté les méthodes d'extraction des indicateurs. Les moyens de réduction dimensionnelle de l'espace d'indicateurs y ont été discutés.

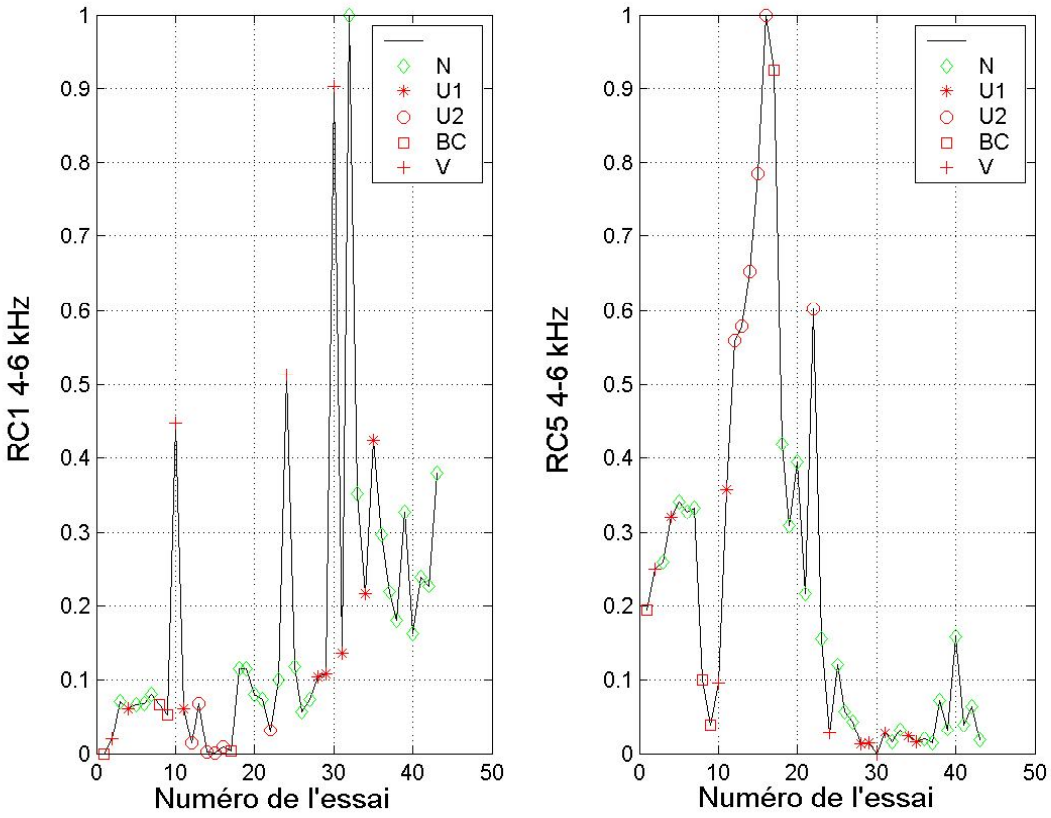


FIG. 3.14 – Évolution des variables RC1 et RC5 4-6 kHz en fonction du numéro de l'essai.

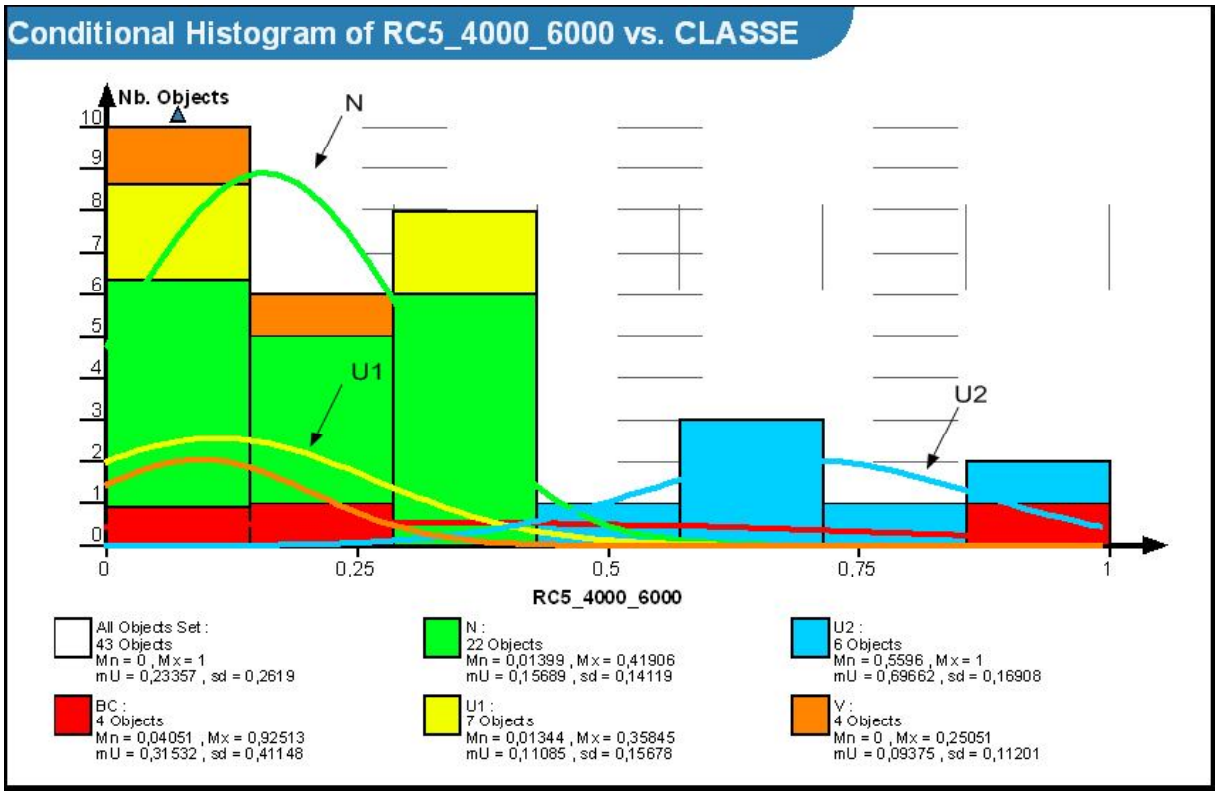


FIG. 3.15 – histogramme conditionnel de RC5 4000-6000Hz vs. classes (graphique obtenu avec le logiciel de data mining PEPITo™

Une méthode d'extraction d'indicateurs à base de l'analyse spectrale singulière a été proposée et appliquée aux données de surveillance d'usure d'outils en tournage. **L'originalité de cette méthode se situe dans l'application sous forme pseudo-locale de l'ASS, ce qui permet la capture des singularités des signaux vu leur non-stationnarité et leur longueur, et conduit à la définition d'indicateurs de bonne qualité discriminatoire.**

Le chapitre suivant introduit le *data mining* et présente les techniques d'apprentissage de manière approfondie.

Références

[1] Cláudia Antunes : Temporal data mining : an overview. *In Workshop on Temporal Data Mining (KDD2001)*, San Francisco,September 2001.

[2] F. Moerchen : *Time Series Knowledge Mining*. Thèse de doctorat, Philipps-Universitat Marburg, 2006.

[3] I. Daubechies : *Ten lecture on Wavelets*. Philadephia, Pa, 1992.

[4] Y. Meyer : *Wavelets and Operators*. Cambridge University, 1992.

[5] S.G. Mallat : A theory for multiresolution signal decomposition : the wavelet representation. *IEEE Pattern Anal. and Machine Intell.,*, 11(7):674–693, 1989.

[6] Issam Abu-Mahfouz : Drilling wear detection and classification using vibration signals and artificial neural network. *International Journal of Machine Tools & Manufacture*, 43:707–720, 2003.

[7] H. Ocak, Loparo K.A. et F.M. Discenzo : Online tracking of bearing wear using wavelet packet decomposition and probabilistic modeling : A method for bearing prognostics. *Journal of Sound and Vibration*, 302:951–961, 2007.

[8] J. K. Hammond et P. R White : The analysis of non-stationary signals using time-frequency methods. *Journal of Sound and Vibration*, 190(3):419–447, 1996.

[9] S. Barbarossa : Analysis of multicomponent lfm signals by a combined wigner-hough transform. *IEEE Transaction on Signal Processing*, 43(6), 1995.

[10] E. Keogh, K. Chakrabarti, M.J. Pazzani et S. Mehrotra : Dimensionality reduction for fast similarity search in large time series databases. *Knowledge and Information Systems*, 3(3):263–286, 2001.

[11] M. Ghi et all : Advanced spectral methods for climatic time series. *Reviews of geophysics*, 40:1.1–1.41, 2002.

[12] D.S. Broomhead et G. King : Extracting qualitative dynamics from experimental data. *Physica D*, 20:217–236, 1986.

[13] W. J. Wang, J. Chen, X. K. Wu et Z. T. Wu : The application of some non-linear methods in rotating machinery fault diagnosis. *Mechanical Systems and Signal Processing*, 15(4): 697–705, 2001.

[14] Randal R.B. : *Frequency analysis*. Brüel & Kjaer, 1987.

[15] N. Golyandina, V. Nekrutkin et A. Zhigljavsky : *Analysis of Time Series Structure-SSA and related techniques*. Chapman & Hall / CRC, 2001.

[16] P. Smyth : Clustering sequences with hidden markov models. *Advances in Neural Information Processing Systems*, 9:648, 1997.

[17] P. Sebastiani et M. Ramoni : Clustering continuous time series. *In Proceedings of the 18th International Conference on Machine Learning (ICML'01), p.497-504, Morgan Kaufmann*, 2001.

[18] Lawrence R. Rabinar : A tutorial on hidden markov models and selected applications in speech recognition. *In Proc. IEEE 77(2)*, pages 257–286, 1989.

[19] L. Wald : Some terms of reference in data fusion. *IEEE Transactions on Geosciences and Remote Sensing*, 37(3):1190–1193, 1999.

[20] David Hall et James Llinas, éditeurs. *Handbook of multisensor data fusion*. CRC Press LLC, 2001.

[21] J.P. Marquès de Sá : *Applied Statistics using SPSS, STATISTICA and MATLAB*. Springer, Berlin, 2003.

[22] K. Fukunaga : *Introduction to Statistical Pattern Recognition*. Academic Press, 1990.

[23] K. Fukunaga et RR Hayes : Effect of sample size in classifier performance. *IEEE Tr.Patt.Anal.Mach.Intel.*, 11:1087–1101, 1989.

[24] AK Jain et B Chandrasekaran : *Handbook of statistics, 2*, chapitre Dimensionality and sample size considerations in pattern recognition., pages 835–855. North Holland Pub.Co., 1989.

[25] D.J. Bertholds, M. & Hand, éditeur. *Intelligent Data Analysis*. Springer, 2003.

[26] H. H. Harman : *Modern Factor Analysis*. University of Chicago Press, 3rd édition, 1976.

[27] D. N. Lawley et A. E. Maxwell : *Factor Analysis as a Statistical Method*. American Elsevier Publishing, 2nd édition, 1971.

[28] Bovic Kilundu et Pierre Dehombreux : Analyse spectrale singulière des signaux vibratoires et machine learning pour la surveillance d'usure d'outils. *Mécanique et Industries*, 9(1):1–8, 2008.

[29] Bovic Kilundu et Pierre Dehombreux : Data mining techniques and singular spectrum analysis for tool condition monitoring in metal cutting. *International Review of Mechanical Engineering*, 2(1):49–55, 2008.

[30] P. Yiou, D. Sornette et M. Ghil : Data-adaptive wavelets and multi-scale singular-spectrum analysis. *Physica D*, 142:254–290, 2000.

Chapitre 4

Le *data mining* : définitions, méthodes et applications à la détection et au diagnostic

Le chapitre précédent a discuté de l'extraction d'indicateurs à partir des signaux bruts. Dans un programme de surveillance vibratoire, ces indicateurs sont généralement stockés dans une base de données à laquelle peut accéder un système de détection et de diagnostic. Le présent chapitre pose les bases du data mining *en le situant dans le contexte général de l'extraction des connaissances à partir des données. Nous présentons ici les différentes méthodes de* machine learning *utilisables pour le besoin de la détection et du diagnostic, discutons de la question cruciale de sélection des indicateurs pertinents et abordons le problème de l'évaluation des performances d'un système de diagnostic.*

4.1 Définitions

4.1.1 *Data mining* et Extraction des Connaissances à partir de Données (ECD)

Une des premières définition de l'ECD, et la plus citée d'ailleurs, est donnée dans [1] où les auteurs la définissent comme « un processus non trivial d'identification de structures inconnues, valides et potentiellement exploitables dans les bases de données ». De ce point de vue, l'ECD est présentée comme un processus général dont le *data mining* n'est qu'une étape parmi cinq [2] :

1. l'acquisition des données : accès aux données stockées sous une forme structurée (base

de données, fichiers tabulaires) ou non structurée (texte, image, etc.). Cette phase vise à cibler l'espace des données qui va être exploré. On définit une zone de prospection et on en rapatrie les données potentiellement utiles à l'aide de requêtes ;

2. le prétraitement des données : au cours de cette phase, les données sont mises sous la forme qui se prête le mieux à l'exploitation. La plupart des techniques ne traitent que des tableaux de données rangées sous forme de matrice appelés fichiers plats. Dans ces opérations de préparation, il peut s'agir de :
   - sélectionner les attributs ;
   - traiter les données manquantes ou aberrantes ;
   - transformer les attributs ;
   - construire des agrégats d'attributs ;
   - traiter des données complexes non structurées ;
3. la fouille de données : on utilise des techniques issues de la statistique ou de l'apprentissage automatique (*machine learning*) pour la visualisation, la classification, la structuration ou la prédiction. C'est cette phase qui portait le nom de *data mining* ;
4. la validation des connaissances extraites : les modèles extraits sont évalués et validés avant de les soumettre à l'épreuve de la réalité. Il s'agira souvent d'estimer le taux d'erreur du modèle à travers une matrice de confusion.
5. le déploiement du modèle de connaissances en vue d'une utilisation effective ;

Cette démarche est aussi reprise dans les recommandations du consortium CRISP-DM* [3].

Aujourd'hui, plusieurs tentatives de redéfinition sont apparues pour mieux préciser le domaine mais aucune ne s'est réellement imposée. Le terme *data mining* est maintenant utilisé pour désigner toutes les étapes du processus d'extraction des connaissances. Par conséquent, nous adoptons la définition suivante du terme *data mining* :

le *data mining* est le processus d'extraction d'informations ou de structures inavouées dans une collection de données. Ce processus a pour but d'extraire de la connaissance qui soit interprétable et directement exploitable dans un processus automatique d'analyse ou de décision. La sélection, le prétraitement et la transformation des attributs sont inclus dans le processus [4] [5].

*Cross-Industry Standard Process for Data Mining

4.1.2 Le *machine learning* ou apprentissage automatique

Selon Herbert Simon [6], l'apprentissage fait référence à tout changement dans le comportement d'un système, changement qui lui permet de mieux effectuer la seconde fois une tâche quand celle-ci est répétée. Dès lors, l'apprentissage sera exprimé en terme d'évaluation des performances futures ou capacité de généralisation. Généralement, l'apprentissage demande une certaine quantité de données que l'on regroupe dans un ensemble « d'exemples d'apprentissage », et qui sont des données empiriques utilisées pour apprendre un concept donné. Il est souvent utile de disposer d'autres ensembles de données pour tester la validité du comportement appris par un système intelligent. Ces ensembles sont appelés ensembles d'exemples de test ou de généralisation. Ce sont surtout les capacités de généralisation d'un système qui déterminent son efficacité en pratique.

Les algorithmes d'apprentissage peuvent être catégorisés en apprentissage supervisé, dans lequel on dispose d'un comportement de référence précis que l'on désire faire apprendre au système, et en apprentissage non supervisé, caractérisé par des données d'apprentissage non-étiquetées.

Quand on se questionne sur la façon dont les données d'entrée doivent être présentées au système intelligent pour son apprentissage ou sa validation, on arrive à l'introduction de deux notions clés du *data mining* : les instances et les attributs.

4.1.3 Instances et Attributs

Le *data mining* concerne en général des données présentées sous forme matricielle. Si cela n'est pas le cas, des requêtes spécifiques doivent extraire des bases de données des tableaux à partir desquelles une recherche d'information peut être effectuée. Ces tableaux ou fichiers plats (*flat files*) reprennent un nombre N d'exemples, chacun étant décrit à l'aide d'un nombre P de variables. Dans le langage du *data mining*, les exemples sont appelés instances ou objets et les variables sont des attributs, propriétés ou descripteurs. Le tableau 4.1 illustre un fichier plat d'un problème de diagnostic vibratoire à défauts combinés. Les quatre dernières colonnes sont les étiquettes des classes. Les lignes de données représentent les instances tandis que les colonnes de données sont les valeurs d'attributs.

Les instances peuvent être indépendantes ou non. Le cas d'instances indépendantes est le plus courant et pour lequel beaucoup de méthodes d'analyse ont été développées. Cependant, les problèmes à instances dépendantes ne sont pas dépourvus d'intérêt. Il peut s'agir des cas où

	RMS glob z1	Kurt glob z1	Skew glob z1	···	B	F	R	A
1	0,79951155	3,13808421	-0,170534315	···	niveau 0	niveau 0	niveau 0	niveau 0
2	0,85598188	3,00522092	-0,111246596	···	niveau 0	niveau 0	niveau 0	niveau 0
3	0,84074742	2,94367781	-0,143816563	···	niveau 0	niveau 0	niveau 0	niveau 0
4	0,80265959	3,15104948	-0,16544746	···	niveau 0	niveau 0	niveau 0	niveau 0
5	0,97708765	6,41736015	0,052527882	···	niveau 0	niveau 0	niveau 1	niveau 0
6	0,85578123	5,00187263	0,072355619	···	?	?	?	?
⋮	⋮	⋮	⋱	⋮	⋮	⋮	⋮	⋮

TAB. 4.1 – Exemple d'un fichier plat d'un problème en diagnostic vibratoire

la prise en compte de l'historique des données s'imposent, mais aussi des cas de données séquentielles ou de séries temporelles. Une distinction devra être faite entre les données statiques et les données dynamiques. Les premières sont strictement indépendantes entre elles tandis que les dernières sont marquées par une dépendance chronologique entre elles.

On distingue les attributs numériques ou continus des attributs nominaux ou discrets. Les attributs numériques mesurent des nombres, réels ou entiers alors que les attributs nominaux prennent leurs valeurs dans un ensemble fini de possibilités préspécifiées. Pour le besoin de traitement, on peut être amené à discrétiser un attribut numérique. Le principe utilisé consiste à repartir les valeurs de l'attribut continu dans différents intervalles dont chacun représentera une valeur discrète. On peut, par exemple, utiliser les percentiles pour la discrétisation d'un attribut numérique. Illustrons-le avec l'attribut RMS glob z1 du tableau 4.1. Cet attribut, qui est continu, peut être discrétisé pour être représenté par l'ensemble de 4 intervalles suivants :
' $]-\infty\ \ 0,8848]$', ' $]0,8848\ \ 1,3087]$', ' $]1,3087\ \ 2,6978]$', ' $]2,6978\ \ +\infty]$'

4.2 Les techniques du *data mining*

Le *data mining* essaie d'extraire de données des informations qui n'ont pas été explicitement stockées. Une manière de catégoriser les problèmes résolus par les techniques de *data mining* distingue deux classes de problèmes : la prédiction et la découverte de connaissances [7]. La découverte de connaissance constitue généralement une étape préalable à la prédiction.

a) La prédiction : les problèmes de prédiction comprennent principalement des problèmes de classification et de régression. Les séries temporelles, qui sont aussi classées dans le groupe de problèmes de prédiction, peuvent être considérées comme un cas spécial de régression, parfois de classification, dans lequel les mesures sont relevées au cours du temps pour les mêmes variables. Elles peuvent bien représenter la situation de données

de surveillance vibratoire ou de condition monitoring en général quand l'on s'intéresse aux tendances.

b) la découverte de connaissances : dans cette classe, on compte la détection des déviations, la clusterisation, les règles associatives, la visualisation, etc.

4.3 La classification

Le processus de classification consiste à affecter une classe à des objets à l'aide d'un modèle entraîné sur un autre ensemble d'objets.

Convenons de noter par $X = \{X_1, X_2, \ldots, X_N\}$ l'ensemble des attributs et par $\mathbf{x}$ une affectation des valeurs à chaque attribut dans X , où $\mathbf{x} = (x_1, x_2, \ldots, x_N)$. Soit $C = \{C_1, C_2, \ldots, C_K\}$ l'ensemble des classes. Dans un processus d'apprentissage supervisé, on dénotera une instance d'apprentissage par $E = (\mathbf{x}, C_i)$. Un classifieur est une fonction f qui associe l'instance $\mathbf{x}$ à une classe C_i, c'est-à-dire $f(\mathbf{x}) = C_i$. La tâche de l'algorithme d'apprentissage sera de construire une approximation h de la fonction f, de façon que h prédise de manière satisfaisante l'affectation d'un objet inconnu $\mathbf{y}$ à une classe C_j donnée, au moyen de la relation $h(\mathbf{y}) = C_j$.

Pour l'illustrer avec l'exemple du tableau 4.1, l'ensemble des attributs sera

$$X = \{\text{RMS glob z1},\ \text{Kurt glob z1},\ \text{Skew glob z1}, \cdots\}$$

Si l'on aborde le problème défaut par défaut, on aura des cas multiclasse pour chacun des défauts. Pour le défaut B (balourd), par exemple, l'ensemble des classes sera

$$C_B = \{\textit{niveau}\ 0,\ \textit{niveau}\ 1,\ \textit{niveau}\ 2, \cdots\}$$

Le processus de classification cherchera, par exemple, à affecter une classe en rapport avec le défaut B à l'instance n°6.

$$\mathbf{x}_6 = (0,85578123,\ 5,00187263,\ 0,072355619, \cdots)$$

On pourra construire un classifieur pour chacun des défauts. Le classifieur h_B pour le défaut B donnerait par exemple $h_B(\mathbf{x}_6) = \textit{niveau}\ 0$.

Dans les sous-sections qui suivent, nous décrivons quelques techniques de classification utilisables pour les tâches de diagnostic.

4.3.1 Les arbres de décision

Un arbre de décision est une représentation hiérarchique utilisée pour déterminer la classification d'un objet en testant les valeurs de certains de ses attributs. Dans un arbre de décision, les nœuds terminaux sont des nœuds de décision ou de classification, on les appelle « feuilles ». Les nœuds intermédiaires sont des nœuds de test sur les propriétés des objets. Le processus de construction des arbres de décision est en général récursif.

En diagnostic [8, 9], deux questions se posent pour la construction de la structure de l'arbre de décision : quel indicateur choisir, et quelle valeur de l'indicateur constituera le seuil de décision pour la segmentation au niveau d'un nœud de test ? La stratégie est de sélectionner, à chaque étape du processus, une propriété sur base du gain d'information, appelé pureté, qui est calculé sur les exemples d'apprentissage. La notion de pureté induit juste le fait que les ensembles devraient ne contenir que de données très similaires, idéalement d'une même classe. A la place de la pureté, on calcule souvent une mesure d'impureté qui est l'entropie statistique de Shannon donnée par :

$$-\sum_i P_i . \log_2 \ P_i \tag{4.1}$$

où P_i est la proportion d'exemples concernés par une propriété ou une valeur particulière de cette propriété. La propriété qui présentera le meilleur gain, c'est-à-dire une entropie minimum, sera choisie comme racine de l'arbre ou nœud de test. Le processus sera ainsi exécuté de manière hiérarchique jusqu'aux noeuds terminaux, c'est-à-dire les noeuds qui sont tels que tous les objets leur associés se trouvent dans une même classe. L'algorithme le plus utilisé dans la construction des arbres de décision est le C4.5 [7, 10]. Il est exploitable dans les environnements utilisés. La figure 4.1 illustre une partie d'un arbre de décision.

Les arbres de décision aboutissent à de bons résultats dans la pratique. Ils possèdent l'avantage d'être compréhensibles par tout utilisateur (si la taille de l'arbre produit est raisonnable) et d'avoir une traduction immédiate en terme de règles de décision.

Élagage des arbres de décision

On peut pousser la construction d'un arbre de décision jusqu'à l'obtention d'une feuille par instance de l'ensemble d'apprentissage. Dans ce cas, l'arbre serait très complexe et surentraîné, avec naturellement une capacité très faible de généralisation sur de nouvelles données.

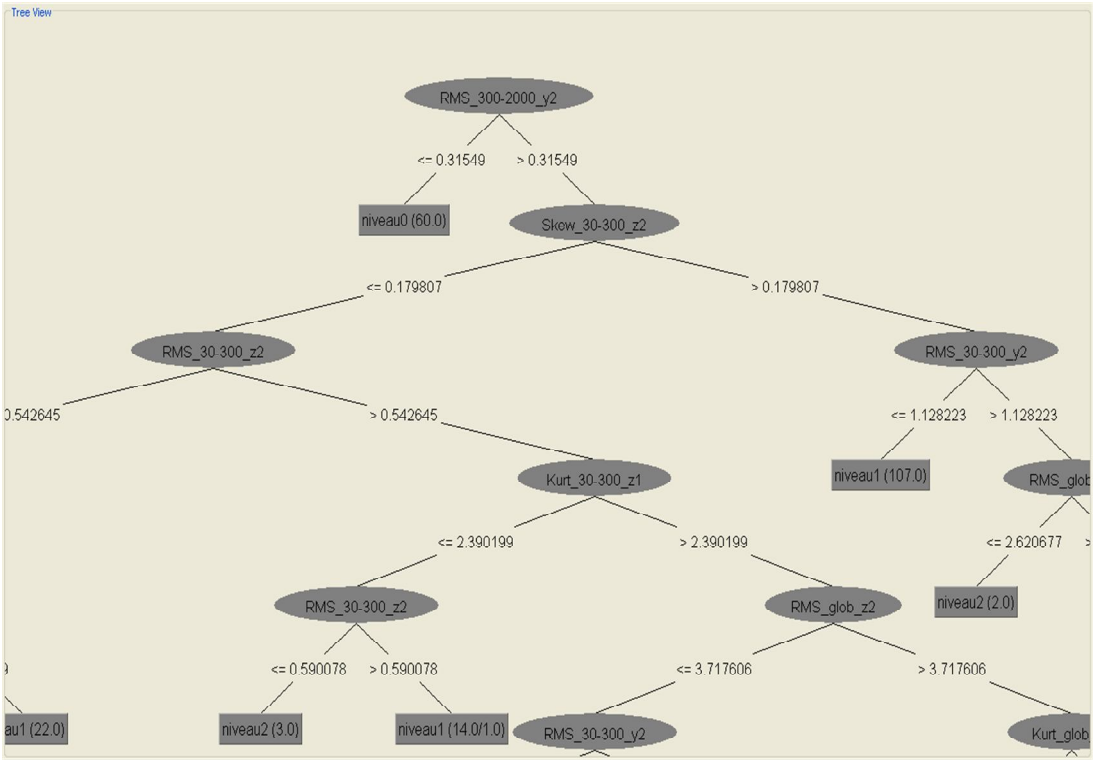


FIG. 4.1 – Illustration d'un arbre de décision (généré avec WEKA)

D'intuition, il appert que l'objectif lors de l'apprentissage doit être d'obtenir un arbre de dimension minimale ayant les meilleures performances. Il faut donc savoir limiter la segmentation de l'arbre pour ne pas apprendre les spécificités de l'ensemble d'apprentissage. Pour cela, on procède à l'élagage pour supprimer les branches qui sont supposées apporter plutôt une spécificité du fichier d'apprentissage qu'une réelle relation dans les données, relation pouvant et devant être généralisée.

On distingue deux stratégies d'élagage : le pré-élagage et le post-élagage [11]. Par le pré-élagage, on décide pendant la croissance de l'arbre quand stopper le développement d'une branche ou sous-arbre. On peut, par exemple, arrêter la segmentation lorsque l'effectif est trop faible ou lorsque le sommet atteint une pureté suffisante.

Le pré-élagage peut sembler être une perspective attrayante en ce sens qu'elle aboutit à une réduction d'opérations lors de la construction de l'arbre. Mais le post-élagage, qui consiste à élaguer l'arbre après son développement complet, offre quelques avantages réels. Par exemple, il y a des cas où deux attributs considérés individuellement ont chacun une faible valeur informationnelle alors que combinés, ils constituent de puissantes variables de prédiction.

4.3.2 Les règles de classification

Souvent il est possible de décrire la structure d'un ensemble de données en utilisant une seule propriété (attribut). Des règles rudimentaires inférées de cette manière peuvent être parfois plus précises que tout autre modèle complexe. Par conséquent, il est conseillé, dans la fouille de données, de penser d'abord aux solutions simples. Voici un pseudocode pour l'inférence de règles simples de type 1R (*one rule*) [11] :

```
Pour chaque attribut
  Pour chaque valeur de cet attribut, construire une règle comme suit:
     compter combien de fois chaque classe apparaît
     trouver la classe la plus fréquente
     construire la règle qui affecte cette classe à cette valeur de l'attribut.
  Calculer le taux d'erreur des règles.
Choisir les règles avec les taux d'erreur les plus faibles.
```

4.3.3 Les modèles statistiques

Une règle 1R utilise un seul attribut comme base de décision. Une autre technique simple est d'utiliser plusieurs attributs en considérant que chacun d'eux contribue de manière indépendante et avec égale importance à l'élaboration de la décision. Certes, ceci ne traduit pas la réalité parce que les attributs ne sont pas toujours également importants et indépendants, des redondances peuvent exister. Cependant, cette façon de procéder aboutit à un schéma très simple et aux résultats satisfaisants en pratique.

Son fondement dérive du théorème de Bayes qui permet de calculer une probabilité a posteriori $P(E_i|\mathbf{x})$ d'un événement E_i sachant que $\mathbf{x}$ s'est produit à partir d'une probabilité a priori $P(E_i)$ de production de l'événement E_i [12].

$$P(E_i|\mathbf{x}) = \frac{P(\mathbf{x}|E_i).P(E_i)}{\sum_j P(\mathbf{x}|E_j).P(E_j)} \tag{4.2}$$

ou plus simplement

$$P(E|\mathbf{x}) = P(\mathbf{x}|E).\frac{P(E)}{P(\mathbf{x})} \tag{4.3}$$

Considérons une instance $\mathbf{x} = (x_1, x_2, ...x_N)$ à laquelle on veut affecter une étiquette parmi

K classes $C_1, C_2, ...C_K$. La classe la plus probable qui lui sera assignée est celle qui maximise $P(C_k|\mathbf{x})$,

$$P(C_k|\mathbf{x}) = P(\mathbf{x}|C_k).\frac{P(C_k)}{P(\mathbf{x})} \tag{4.4}$$

où la probabilité $P(\mathbf{x})$ est supposée constante en vertu de l'équiprobabilité des objets.

Donc il faudra chercher la classe qui maximise $P(\mathbf{x}|C_k).P(C_k)$ pour $k = 1...K$. La probabilité $P(C_k)$ est une probabilité a priori qui sera déduite de l'échantillon par la relation

$$P(C_k) = \frac{\text{Nombre d'objets dans la classe } \mathrm{C_k}}{\text{Taille de l'échantillon}} = \frac{\text{Taille}(\mathrm{C_k})}{\text{Taille(Echantillon)}}$$

La probabilité $P(\mathbf{x}|C_k)$ sera approximée en considérant l'hypothèse d'indépendance des attributs,

$$P(\mathbf{x}|C_k) = \Pi_i P(X_i = x_i|C_k) \quad i = 1...N \tag{4.5}$$

où $P(X_i = x_i|C_k)$ est estimée de la façon suivante :

– si la variable est nominale :

$$P(X_i = x_i|C_k) = \frac{\text{Taille}(X_i = x_i \; de \; C_k)}{\text{Taille}(C_k)} \tag{4.6}$$

– si la variable est continue, on fait une discrétisation et on utilise une distribution gaussienne.

Cette méthode qui suppose une indépendance complète des variables est appelée classification bayésienne naïve.

Quand une des probabilités $P(X_i = x_i|C_k)$ est nulle, elle opposera un veto sur les autres probabilités $P(X_i|C_k)$ quelles que soient leurs grandeurs, et ainsi annuler la probabilité $P(C_k|\mathbf{x})$. Pour éviter cela, un ajustement mineur est fait dans le calcul des probabilités $P(X_i = x_i|C_k)$ à partir des fréquences. On utilise l'estimateur de Laplace pour initialiser toutes les probabilités à des valeurs non nulles :

$$P(X_i = x_i|C_k) = \frac{\text{Taille}(X_i = x_i \; de \; C_k) + \mu.p_i}{\text{Taille}(C_k) + \mu} \tag{4.7}$$

où μ est une constante de faible valeur et p_i la probabilité a priori que l'attribut X_i prenne la valeur x_i.

4.3.4 Les réseaux bayésiens

Comme le diagnostic n'est pas toujours une certitude mais plutôt un faisceau de présomptions convergentes, il est intéressant d'aborder le problème de la reconnaissance de défaut sous un angle probabiliste [13, 14]. C'est d'ailleurs ce que font les arbres de décision et les classifieurs bayésiens naïfs en représentant l'appartenance à une classe comme une probabilité conditionnelle. Mais les classifieurs bayésiens naïfs ne peuvent modéliser que des distributions de probabilité simples alors les arbres de décision peuvent représenter - ou mieux approcher- des distributions arbitraires. L'inconvénient des arbres de décision est qu'ils fragmentent l'ensemble d'apprentissage en de pièces tellement minuscules qu'on aboutit à des estimations de probabilité moins fiables. Le meilleur compromis est obtenu avec des représentations graphiques appelées « réseaux bayésiens ». Ce sont des outils très adaptés à la prise de décision sous incertitude.

Un réseau bayésien est un classifieur basé sur les probabilités conditionnelles. A la différence du modèle bayésien naïf, le réseau bayésien admet certaines dépendances. C'est un graphe direct acyclique de dépendance où les noeuds représentent des variables aléatoires. Certaines variables sont liées par des notions de parenté (parent, descendant) et les variables qui ne sont pas liées sont indépendantes.

Les réseaux bayésiens reposent sur l'hypothèse que la valeur d'un noeud ne dépend que des parents directs de celui-ci et non de ses ancêtres, c'est-à-dire :

$$P(\text{noeud}|\text{ancêtres}) = P(\text{noeud}|\text{parents}) \tag{4.8}$$

Cette propriété est appelée en statistique « indépendance conditionnelle » et a pour conséquence que le calcul de la probabilité jointe de N attributs x_i se décompose dans le produit :

$$P(x_1, x_2, ..., x_n) = \Pi_{i=1}^{n} P(x_i|x_{i-1}, ..., x_1) = \Pi_{i=1}^{n} P(x_i|parents\ de\ x_i) \tag{4.9}$$

Considérons l'exemple du diagnostic du balourd (défaut B dans le tableau 4.1). Le nombre d'instance par classe et les probabilités a priori de chaque classe sont donnés dans le tableau 4.2. L'ensemble de données contient au total 648 instances. La figure 4.2 montre le réseau bayésien construit pour le défaut de balourd. Nous reprenons, pour fixer les idées, la distribution de probabilités de l'indicateur RMS glob z2 dans le tableau 4.3. Cet indicateur n'admet qu'un seul parent qui est B. La variable RMS glob z2 est discrétisée et représentée dans les intervalles suivants :

1. $I_1 = -\infty <$RMS glob z2$\leq 0,7263$,
2. $I_2 = 0,7263 <$RMS glob z2$\leq 2,5967$,
3. $I_3 = 2,5967 <$RMS glob z2$\leq 4,3013$,
4. $I_4 = 4,3013 <$RMS glob z2$\leq 5,2426$,
5. $I_5 = 5,2426 \leq$ RMS glob z2$< +\infty$,

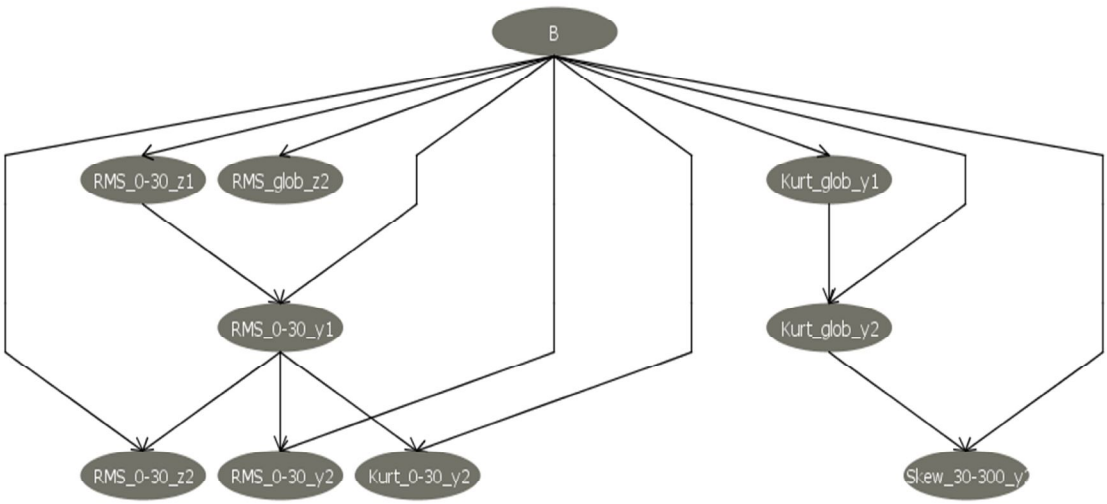


FIG. 4.2 – Réseau bayésien pour le diagnostic du balourd (2 parents au maximum) (obtenu sous l'environnement WEKA)

Balourd B	Effectif	Probabilité a priori P(Ci) de la classe
niveau 0	144	0,222
niveau 1	144	0,222
niveau 2	96	0,148
niveau 3	132	0,204
niveau 4	132	0,204

TAB. 4.2 – Probabilités a priori des classes du défaut de balourd

	RMS glob z2				
Balourd B	I_1	I_2	I_3	I_4	I_5
niveau 0	0,003413	0,569966	0,365188	0,003413	0,058020
niveau 1	0,003413	0,624573	0,283276	0,003413	0,085324
niveau 2	0,005076	0,979695	0,005076	0,005076	0,005076
niveau 3	0,092937	0,635688	0,211896	0,055762	0,003717
niveau 4	0,003717	0,732342	0,204461	0,055762	0,003717

TAB. 4.3 – Table de distribution de probabilités pour le noeuds RMS glob z2

4.3.5 Les réseaux neuronaux artificiels

Le réseau de neurones formels

Le neurone artificiel, encore appelé neurode, noeud ou unité neuronale, est une modélisation mathématique du neurone biologique. Il en reprend les grands principes de fonctionnement, notamment la sommation des entrées. Les réseaux de neurones artificiels sont inspirés du système nerveux. Ils sont constitués de neurones interconnectés entre eux.

Chaque neurone reçoit en entrée des informations venant de plusieurs neurones à travers les connexions synaptiques. Chacune de ces entrées est pondérée par une valeur w, dite poids synaptique. Les valeurs ainsi pondérées sont ensuite sommées avant de passer en argument à une fonction f appelée fonction d'activation ou fonction de transfert. Les valeurs des poids synaptiques sont déterminées à la suite d'un processus d'apprentissage au cours duquel le neurone ajuste ses forces de connexion afin d'obtenir une sortie désirée.

Le neurone simple

Par neurone simple nous entendons un neurone auquel n'est transmis en entrée qu'une seule valeur scalaire p [15]. Cette valeur est multipliée dans la connexion par le poids w, pour former le produit wp qui est aussi un scalaire. Le neurone peut avoir ou non un biais b (figure 4.3). S'il y a un biais, celui-ci sera sommé à la valeur pondérée wp avant de passer en argument au niveau de la fonction d'activation f. L'argument de f est $n = wp + b$ et la valeur a que retourne la fonction f est la sortie du neurone. On aura donc

$$a=f(wp+b) \tag{4.10}$$

Le biais b peut être vu comme un décalage de la fonction f. Il peut tout aussi bien être appréhendé comme un poids synaptique qui s'applique à une entrée unitaire. Les valeurs w et

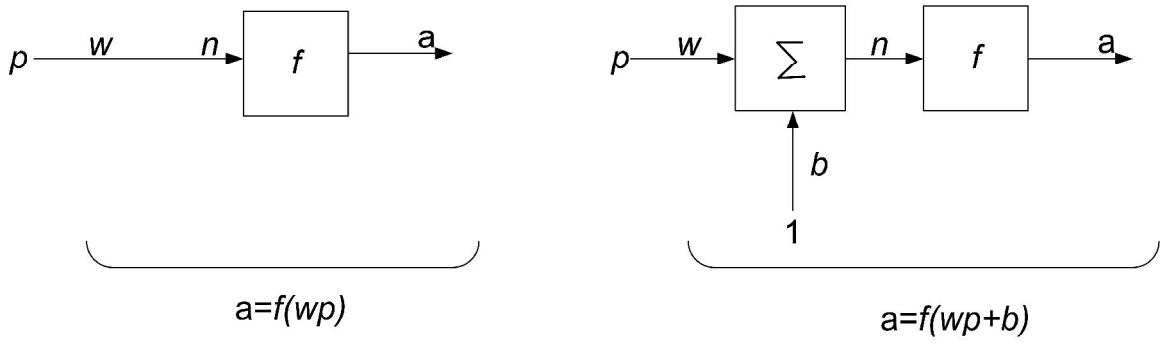


FIG. 4.3 – Le neurone formel simple sans biais (à gauche) et avec biais (à droite)

b sont des paramètres scalaires ajustables de ce neurone simple. L'idée centrale est de faire présenter au neurone un comportement désiré par le seul ajustement des paramètres w et b. On peut ainsi donc entraîner un réseau de plusieurs neurones interconnectés à exécuter une tâche particulière en ajustant ces paramètres ou en laissant le réseau les ajuster seul.

Les fonctions d'activation

Les fonctions d'activation sont de nature très variées : elle peuvent être déterministes, continues, discontinues ou aléatoires [4]. Les fonctions les plus utilisées sont :

1. la fonction à seuil de Heaviside (figure 4.4) ;
2. la fonction sigmoïde de type Log (figure 4.5) ;
3. la fonction sigmoïde de type tangente hyperbolique (figure 4.5) ;
4. la fonction linéaire sans saturation ;
5. la fonction linéaire à seuil ou saturation.

Le neurone à entrée vectorielle

Contrairement au neurone simple présenté précédemment, l'entrée du neurone peut être un vecteur de dimension R (figure 4.6).

$$\mathbf{p} = (p_1, p_2, \cdots, p_R) \tag{4.11}$$

Chaque entrée individuelle p_i est multipliée par le poids $w_{1,i}$, l'indice 1 dans le poids indiquant le neurone auquel sont transmises ces entrées. La somme des valeurs pondérées est le produit

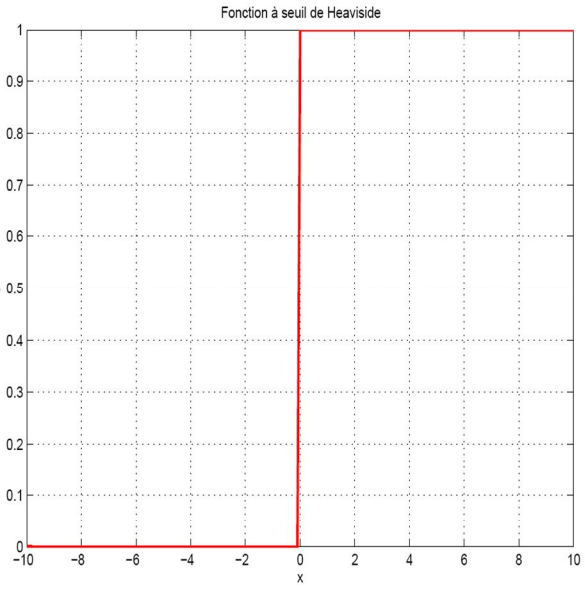


FIG. 4.4 – Fonction à seuil de Heaviside

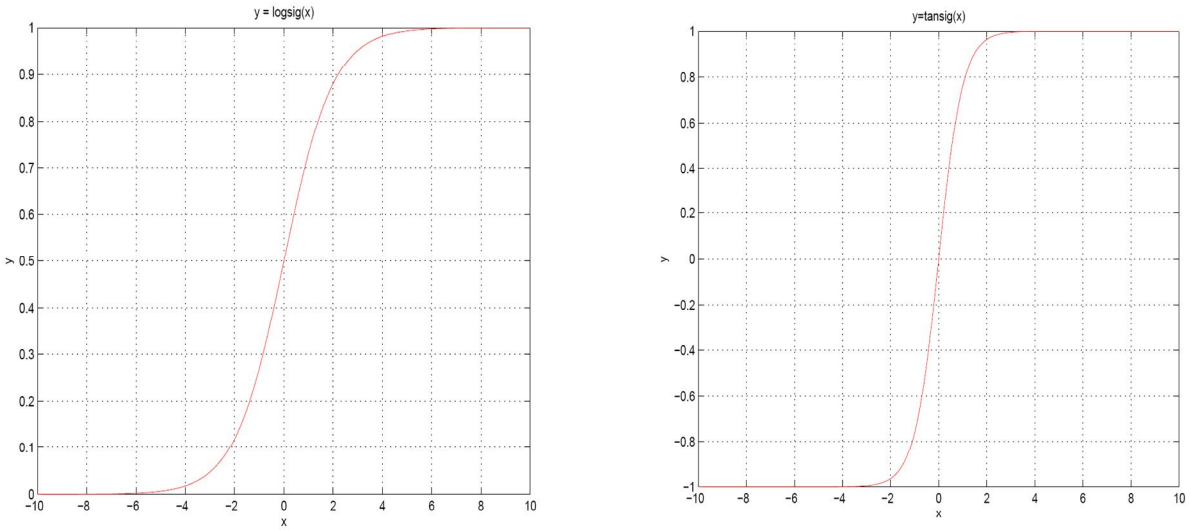


FIG. 4.5 – Fonctions sigmoïde log et tan

scalaire des vecteurs **w** et **p**.

Avec **w** donné par

$$\mathbf{w} = (w_{1,1}, w_{1,2}, \cdots, w_{1,R}) \tag{4.12}$$

L'argument n de la fonction d'activation f est

$$n = w_{1,1}p_1 + w_{1,2}p_2 + \cdots + w_{1,R}p_R + b \tag{4.13}$$

ou vectoriellement

$$n = \mathbf{w}.\mathbf{p} + b \tag{4.14}$$

et la sortie du neurone a sera évidemment calculée par $a = f(n)$.

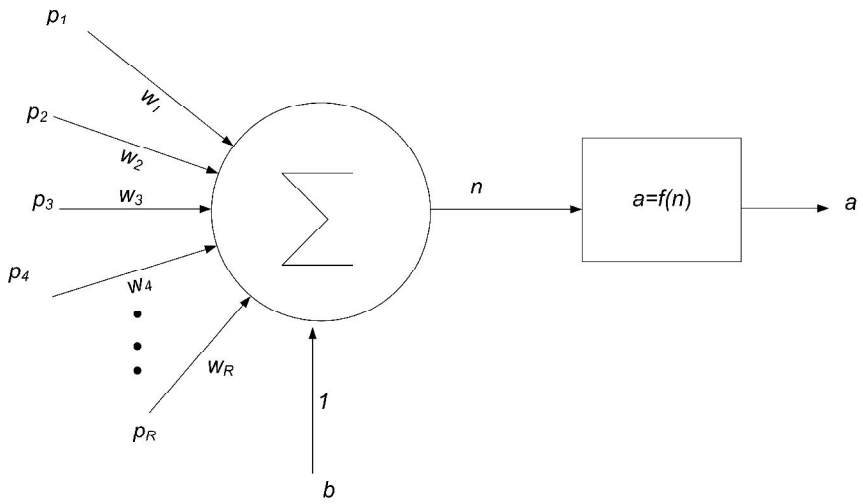


FIG. 4.6 – Neurone à entrée vectorielle

Dans ce qui suit, les deux topologies principales de réseaux de neurones seront brièvement présentées : les réseaux de perceptrons multi-couches (*Multi-layer Perceptron* ou MLP) et les réseaux de fonctions à base radiale (*Radial Basis Functions*[†] ou RBF)

[†]Une fonction à base radiale est une fonction dont la valeur ne dépend que de la distance du point auquel elle est calculée à un point **c** appelé centre : $\Phi(\mathbf{x}) = \|\mathbf{x} - \mathbf{c}\|$

En diagnostic, les réseaux de neurones sont très utilisés. Environ la moitié de publications sur l'utilisation de la classification en diagnostic concerne les réseaux de neurones [16]. Les moyens et outils de leurs apprentissage et implémentation sont devenus facilement accessibles. Les environnements de calcul scientifique tels que MATLAB et SCILAB ont des boîtes à outils qui permettent de construire presque toutes les typologies de réseaux de neurones. Les réseaux neuronaux sont aussi implémentés dans les outils plus orientés *data mining*, comme PEPITo et WEKA que nous avons utilisés.

Les réseaux de perceptrons multi-couches (MLP)

En généralisant l'architecture par combinaison de plusieurs neurones dans une couche et en considérant plusieurs couches disposées les unes après les autres, nous obtenons ce qu'on appelle réseau de neurones. Une couche sera donc caractérisée par le nombre S de ses neurones, un vecteur biais $\mathbf{b}$ de dimension S, une matrice $\mathbf{W}$ des poids synaptiques et une série de S fonctions d'activation identiques ou non pour chaque neurone. La matrice $\mathbf{W}$ est une matrice de dimension $S \times R$, R étant le nombre d'entrées pour la couche considérée ou la dimension du vecteur d'entrée $\mathbf{p}$.

$$\mathbf{W} = \begin{bmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,R} \\ w_{2,1} & w_{2,2} & \cdots & w_{2,R} \\ \vdots & \vdots & & \vdots \\ w_{S,1} & w_{S,2} & \cdots & w_{S,R} \end{bmatrix} \tag{4.15}$$

Le premier indice (indice de ligne) des éléments de la matrice $\mathbf{W}$ indique le neurone auquel le synapse aboutit, et l'indice de colonne l'origine du synapse. Ainsi, par exemple, $w_{1,2}$ est la force du signal provenant du second élément d'entrée et destiné au premier neurone. Il faudra faire une distinction entre la matrice de poids connectés aux entrées et les matrices de poids de connexion entre couches. Pour cela convenons d'appeler $\mathbf{IW}$ (*Input Weights*) la matrice de poids d'entrée et $\mathbf{LW}_{j,i}$ (*Layer Weights*), la matrice de connexion entre la couche j et la couche i. La couche de départ est i et celle de destination j.

La structure d'un réseau MLP est montré à la figure 4.7. Le réseau utilise tous les indicateurs comme entrées et calcule une sortie pour chaque défaut. Comme la prédiction d'un défaut est une probabilité, la sortie doit être comprise entre 0 et 1, ce qui est aisément assuré par l'utilisation d'une fonction d'activation sigmoïde dans la couche de sortie.

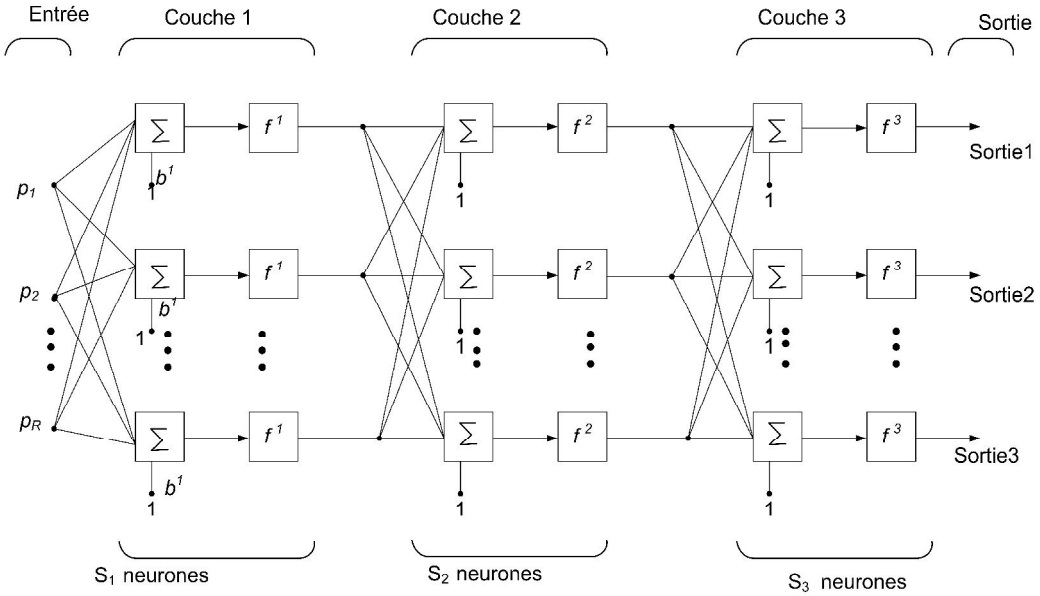


FIG. 4.7 – Réseau de neurones multicouche

La figure 4.7 illustre un réseau de neurones à trois couches et à S_i neurones dans la couche i. Dans cet exemple, toutes les fonctions d'activation d'une couche sont identiques. La première couche du réseau projette les données à l'aide d'une somme pondérée de valeurs d'entrée et un biais. La sortie $\mathbf{a}^i$ d'une couche i est un vecteur dont les composantes sont

$$\mathbf{a}^i = (a_1^i, a_2^i, \cdots, a_{S_i}^i) \tag{4.16}$$

Les réseaux MLP sont très adaptés aux problèmes où l'on dispose de suffisamment de données et où les classes ont des frontières complexes. Des exemples d'utilisation des réseaux de neurones en diagnostic peuvent être trouvées dans les références [17–21]. Dans certaines applications, on utilise un ensemble de réseaux dont chacun est destiné à l'identification d'un seul défaut. L'avantage de cette approche est que ces réseaux sont de taille réduite. Mais dans l'ensemble, le nombre de paramètres peut rester supérieur à celui d'un seul réseau pour tous les défauts.

D'une manière générale, la sortie **a** du réseau de neurones est comparée à une sortie attendue **t** appelée « cible ». L'écart entre la sortie obtenue et la sortie attendue est représentée par une fonction d'erreur dont la définition dépend du type de problème. L'erreur est fonction des valeurs des poids synaptiques et des biais. La fonction d'erreur la plus utilisée est la moyenne des

carrés de l'erreur (*mse* : *mean squared error*). Le rôle du processus d'apprentissage est d'ajuster les poids de manière à avoir une erreur minimale.

Si l'on dispose de M exemples de vecteurs d'entrée ou instances $\mathbf{x}_q$ ($q = 1 \cdots M$) auxquels on associe M vecteurs de sortie attendus $\mathbf{t}_q$ ($q = 1 \cdots M$), la fonction d'erreur mse est définie comme suit :

$$mse(\mathbf{W}) = \frac{1}{M} \sum_{q=1}^{M} (\mathbf{t}_q - \mathbf{a}_q)^2 \tag{4.17}$$

Les réseaux de fonctions à base radiale

Le plus grand problème avec les réseaux MLP est leur incapacité à l'extrapolation. Or en diagnostic les ensembles de données ne sont pas toujours complets. D'où le besoin d'un système de diagnostic qui opère au-delà du domaine couvert par les données d'apprentissage. Les réseaux de fonctions à base radiale ne permettent pas non plus d'opérer au-delà du domaine d'apprentissage, mais ils ne fournissent pas une sortie erronée. En effet, chaque fonction radiale représente une zone donnée du domaine d'entrée. Lorqu'une nouvelle entrée (instance) n'est pas couverte par l'ensemble d'apprentissage, le réseau donne une sortie nulle (rejet). Ce qui évite au moins de prendre une fausse décision. Il y a ensuite lieu d'augmenter le réseau par un apprentissage *on-line* en ajoutant une nouvelle fonction chaque fois qu'une donnée non encore couverte est présentée.

Les neurones d'un réseau RBF opèrent en calculant les distances euclidiennes entre les entrées et les centres qui sont des prototypes (représentants) de régions données du domaine d'entrée. Les centres sont déterminés par l'algorithme de k-moyennes (*k-means*). Cet algorithme procède d'abord par la fixation au hasard de k points comme centres, puis par le regroupement des points de l'ensemble de données autour de chacun des centres en calculant les distances euclidiennes. Ensuite, la procédure calcule les nouveaux centres des groupes et affine le regroupement de manière itérative.

Dans le réseau RBF, les distances entre les entrées et les centres sont passées en argument à la fonction d'activation. Avec un nombre H de neurones cachés et une sortie $\mathbf{y}$ de dimension p, comme illustré à la figure 4.8, l'entrée $\mathbf{x}$ est transformée suivant la rélation de l'équation 4.19

$$G_k(\mathbf{x}) = G_k\left(\|\mathbf{x} - \mathbf{x_k}\|\right) \qquad k = 1, ...H \tag{4.18}$$

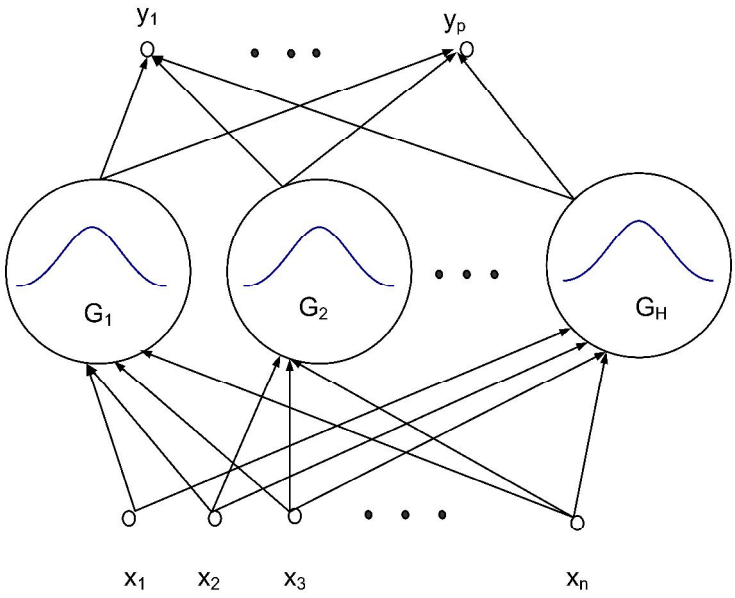


FIG. 4.8 – Architecture générale d'un réseau RBF

$$y_i = \sum_{k=1}^{H} w_{ik} G_k(\mathbf{x}) + w_{i0} \qquad i = 1, ..., p \tag{4.19}$$

où w_{ik} représente les poids.

La fonction d'activation est une fonction à base radiale, généralement de forme gaussienne. Les centres $\mathbf{x_k}$ et les variances des fonctions sont déterminés a priori.

$$G_k(\mathbf{x}) = \exp\left(-\frac{\|\mathbf{x} - \mu_k\|^2}{2\sigma_k^2}\right) \tag{4.20}$$

4.3.6 Les machines à vecteurs de support

Les machines à vecteurs de support sont une méthode de classification binaire utilisant une séparation linéaire. Si les données sont linéairement séparables, il peut exister plusieurs possibilités de séparation de données en deux classes (figure 4.9 et 4.10). Mais l'on préférera un classifieur linéaire qui maximise la marge, celle-ci étant sa distance du point le plus proche. Un tel classifieur linéaire est appelé hyperplan optimal de séparation.

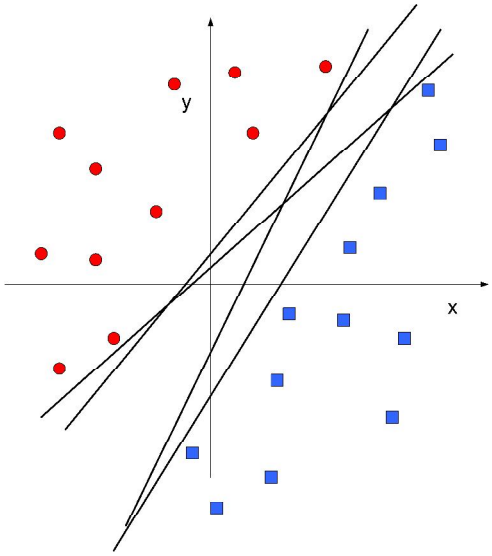


FIG. 4.9 – Hyperplans de séparation

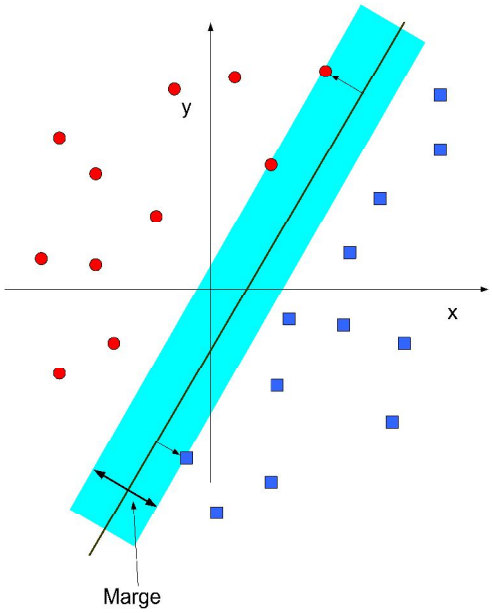


FIG. 4.10 – Maximisation de la marge

Maximisation de la marge

Soit D un ensemble de vecteurs d'apprentissage appartenant à deux classes linéairement séparables tel que

$$D = \{(\mathbf{x}_i, C_i), ..., (\mathbf{x}_m, C_m)\} \tag{4.21}$$

Avec $\mathbf{x} \in \Re^n$ et $C = \{-1, 1\}$, l'ensemble de classes. L'hyperplan de séparation est

$$\mathbf{w}.\mathbf{x} + b = 0 \tag{4.22}$$

où $\mathbf{w}$ représente un vecteur normal à l'hyperplan, $|b| \,/\, ||\mathbf{w}||$ la distance de l'hyperplan à l'origine, $\mathbf{w}.\mathbf{x}$ dénote le produit scalaire des vecteurs $\mathbf{w}$ et $\mathbf{x}$. $||\mathbf{w}||$ est la norme euclidienne.

Les vecteurs de données seront séparés de manière optimale par l'hyperplan si la séparation se fait sans erreur et si la marge est maximale. Supposons que tous les vecteurs satisfont

$$\mathbf{w}.\mathbf{x}_i + b \;\geq\; 1 \quad pour\;\; C_i = +1 \tag{4.23}$$

$$\mathbf{w}.\mathbf{x}_i + b \;\leq\; -1 \quad pour\;\; C_i = -1 \tag{4.24}$$

Ces deux inégalités peuvent être combinées en une seule :

$$C_i\,(\mathbf{w}.\mathbf{x}_i + b) \geq 1 \;\; \forall i \tag{4.25}$$

La distance de l'hyperplan $H_1 : \mathbf{w}.\mathbf{x} + b \geq 1$ à l'origine vaut $|b - 1| \,/\, ||\mathbf{w}||$ et celle de l'hyperplan $H_2 : \mathbf{w}.\mathbf{x} + b \leq -1$ à l'origine vaut $|b + 1| \,/\, ||\mathbf{w}||$. Donc la marge qui est la distance entre les deux hyperplans sera donnée par

$$\rho(\mathbf{w}, b) = \frac{2}{||\mathbf{w}||} \tag{4.26}$$

Maximiser cette marge revient à minimiser $||\mathbf{w}||$. Comme il est aisé de minimiser $||\mathbf{w}||^2$, le problème devient celui de la minimisation de

$$\phi(\mathbf{w}) = \frac{1}{2}\,||\mathbf{w}||^2 \tag{4.27}$$

La solution à la minimisation de 4.27 sous les contraintes de l'équation 4.25 est donnée par

l'optimisation du Lagrangien primal [22] suivant

$$L_P(\mathbf{w}, b, \alpha) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^{m} \alpha_i C_i (\mathbf{w}.\mathbf{x}_i + b - 1) \tag{4.28}$$

où les $\alpha_i \geq$ sont les multiplicateurs de Lagrange. Il sera question de minimiser le Lagrangien par rapport à $\mathbf{w}$ et b, et de le maximiser par rapport à $\alpha \geq 0$. La dualité classique du Lagrangien permet de transformer le problème primal en un problème dual facile à résoudre. On impose une stationnarité aux dérivées de L_P par rapport $\mathbf{w}$ et à b :

$$\frac{\partial L_P(\mathbf{w}, \mathbf{b}, \alpha)}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^{m} \alpha_i C_i \mathbf{x}_i = \mathbf{0} \tag{4.29}$$

$$\frac{\partial L_P(\mathbf{w}, \mathbf{b}, \alpha)}{\partial b} = \sum_{i=1}^{m} \alpha_i C_i = 0 \tag{4.30}$$

ou encore

$$\mathbf{w} = \sum_{i=1}^{m} \alpha_i C_i \mathbf{x}_i \tag{4.31}$$

$$0 = \sum_{i=1}^{m} \alpha_i C_i \tag{4.32}$$

Substituant dans 4.28, on obtient

$$L_D(\mathbf{w}, \mathbf{b}, \alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j} C_i C_j \alpha_i \alpha_j \mathbf{x}_i \mathbf{x}_j \tag{4.33}$$

Le problème dual se pose alors comme suit

$$\begin{cases} max_\alpha L_D \\ \sum_{i=1}^{m} \alpha_i C_i = 0 \\ \alpha_i \geq 0, \quad \forall i \end{cases} \tag{4.34}$$

Si les α_i^* sont solutions de 4.34, on montre que seuls les points qui satisfont $C_i (\mathbf{w}.\mathbf{x}_i + b) = 1$ correspondront à des valeurs non nulles de multiplicateurs de Lagrange α_i^*. Ils sont appelés vecteurs de support. Si les données sont linéairement séparables, les vecteurs de support se trouvent sur la marge et il peut en exister seulement un nombre très restreint. Par conséquent,

l'hyperplan optimal n'est déterminé que par un nombre réduit de données d'apprentissage.

Tout ce qui vient d'être développé a supposé que les données étaient linéairement séparables. Dans le cas où, elles ne le sont pas, on peut faire usage des fonctions noyaux (*kernel functions*), qui transforme les données en un espace à grande dimension où la séparation linéaire est possible. Cette transformation se fait par une relation de la forme $k(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i).\Phi(\mathbf{x}_j)$. Il existe plusieurs types de noyaux [23] dont les plus connus sont :

– le noyau linéaire :

$$k(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i.\mathbf{x}_j \tag{4.35}$$

– le noyau polynomial :

$$k(\mathbf{x}_i, \mathbf{x}_j) = \left(c + \mathbf{x}_i.\mathbf{x}_j\right)^p \tag{4.36}$$

– le noyau gaussien :

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(\frac{\left\|\mathbf{x}_i - \mathbf{x}_j\right\|}{\sigma}\right) \tag{4.37}$$

Illustrons l'action d'un noyau polynomiale de degré 2 agissant dans un espace d'indicateurs à 2 dimensions. Soient deux instances $\mathbf{x} = (x_1, x_2)$ et $\mathbf{y} = (y_1, y_2)$, on aura

$$\begin{aligned} k(\mathbf{x}, \mathbf{y}) &= (\mathbf{x}.\mathbf{y})^2 \\ &= ((x_1, x_2).(y_1, y_2))^2 \\ &= (x_1y_1 + x_2y_2)^2 \\ &= x_1^2y_1^2 + x_2^2y_2^2 + 2x_1x_2y_1y_2 \\ &= \left(x_1^2, x_2^2, \sqrt{2}x_1x_2\right).\left(y_1^2, y_2^2, \sqrt{2}y_1y_2\right) = \Phi(\mathbf{x}).\Phi(\mathbf{y}) \end{aligned}$$

ce qui donne une représentation dans un espace à 3 dimensions.

D'une manière générale, un noyau polynomial de degré d projette les données d'un espace de dimension n vers un autre de dimension C_d^{n+d-1}. Il en ressort donc qu'utiliser un noyau puissant dans le but d'obtenir à tout prix une séparation linéaire aboutit à un surapprentissage. La figure 4.11 montre la dégradation des taux de reconnaissance des défauts par un classifieur

de type machine à vecteurs de support, appliqué au problème du banc d'essai présenté en annexe A. Cette dégradation est représentée en fonction du degré du *kernel* polynomial utilisé.

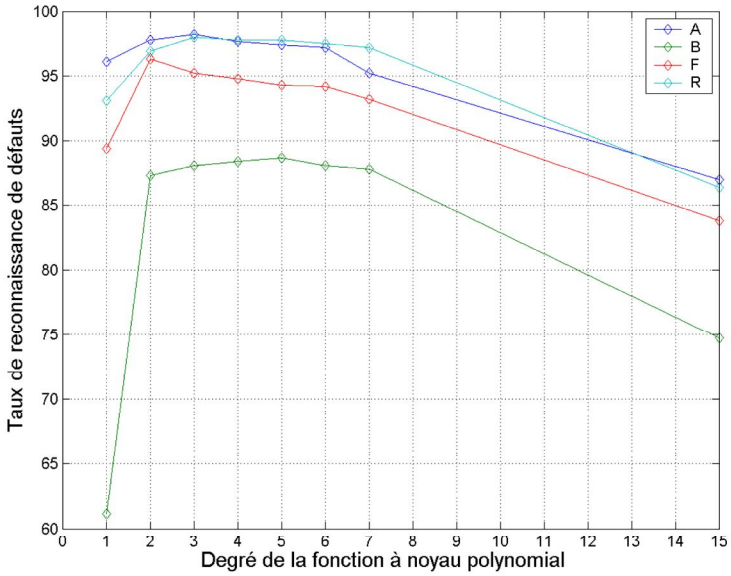


FIG. 4.11 – Evolution des taux de reconnaissance en fonction du degré du noyau

En outre, en pratique il est incongru de chercher à séparer exactement les données du moment que ces dernières peuvent être entachées de bruit et ne sont disponibles qu'en nombre fini. On permettra donc que les contraintes sur la marge soient violées en utilisant les variables « ressorts »

$$\begin{cases} \min \frac{1}{2} \|\mathbf{w}\|^2 + K \sum_{i=1}^{m} \zeta_i^2 \\ C_i \left(\mathbf{w}.\mathbf{x}_i + b\right) \geq 1 - \zeta_i \end{cases} \tag{4.38}$$

En passant au problème dual, on établit que K est une borne supérieure aux coefficients de Lagrange [24]. La borne K agit comme un paramètre de régularisation. Il permet le contrôle du classifieur et doit être choisi de manière à refléter le bruit sur les données. Ce sera un compromis entre la valeur de la marge et les erreurs de classification. Plus grand est K moins large sera la marge. Le choix de K pourra se faire par validation croisée.

On pourra trouver des exemples d'utilisation des machines à vecteurs de support en diagnostic dans les références [25–31]

4.3.7 L'inférence floue

La logique floue est, avec les réseaux bayésiens, une méthode de décision en environnement incertain. Il peut être intéressant d'aborder la question du diagnostic de défauts à l'aide d'un système d'inférence floue. La logique floue est basée sur la théorie des ensembles flous développée par Lofti Zadeh dans les références [32–36].

Le concept de la logique floue vient du constat que la variable booléenne, qui ne peut prendre que deux valeurs (vrai ou faux) est mal adaptée à la représentation de la plupart des phénomènes courants. Considérons l'exemple d'une valeur de seuil qui sépare 2 groupes de niveaux vibratoires : bon et sévère. Sur la figure 4.12, la valeur 20 est le seuil de passage de l'état bon à l'état sévère. Cette façon de définir un seuil est rigide, car une valeur de 19,8 sera considérée comme bonne alors que celle de 20, 1 comme sévère. La notion de seuil est donc une notion floue qu'il conviendra de représenter à l'aide de ce qu'on appelle un degré d'appartenance (*degree-of-membership*) à l'état décrit et qui est vu comme une probabilité d'appartenir à l'ensemble considéré. Sur l'exemple de la figure 4.12, la valeur 21 sera considérée comme bonne à 27% et comme sévère à 73%. Le problème sera donc modélisé en représentant chaque niveau par une fonction d'appartenance.

Les systèmes flous peuvent être utilisés pour établir une relation entre un ensemble de données d'entrée et un ensemble de données de sortie. Pour traiter un problème par la logique floue, il convient d'abord de modéliser les entrées du système à l'aide des fonctions dites d'appartenance. On appelle souvent « fuzzyfication » cette quantification « floue » des variables. Ce processus se fait itérativement. On peut identifier plusieurs domaines d'évolution pour les variables. Dans ce cas on construira une fonction d'appartenance pour chaque domaine, ces fonctions pouvant se chevaucher (exemple : bon-moyen-sévère qui donne lieu à 3 fonctions).

L'étape qui suit la fuzzyfication est la liaison des entrées et des sorties à l'aide des règles dites floues de la forme « Si Condition 1 et/ou Condition 2...et/ou Condition N alors Conséquence ». Selon l'état de l'entrée, chacune de ces conditions est satisfaite avec un degré d'appartenance. Il faudra composer ces différents degrés d'appartenance pour en attribuer un à la condition globale. Il existe plusieurs lois de composition, mais la plus utilisée impose de prendre le degré d'appartenance maximum parmi les conditions d'entrée si celles-ci sont liées par une logique « ou », et le minimum dans le cas de la logique « et » [37].

Pour l'implication, la conséquence est construite par un processus de troncature de la fonction d'appartenance de la sortie au niveau du degré d'appartenance de la condition globale.

Il peut arriver que pour un état donné des entrées, plusieurs règles soient validées simulta-

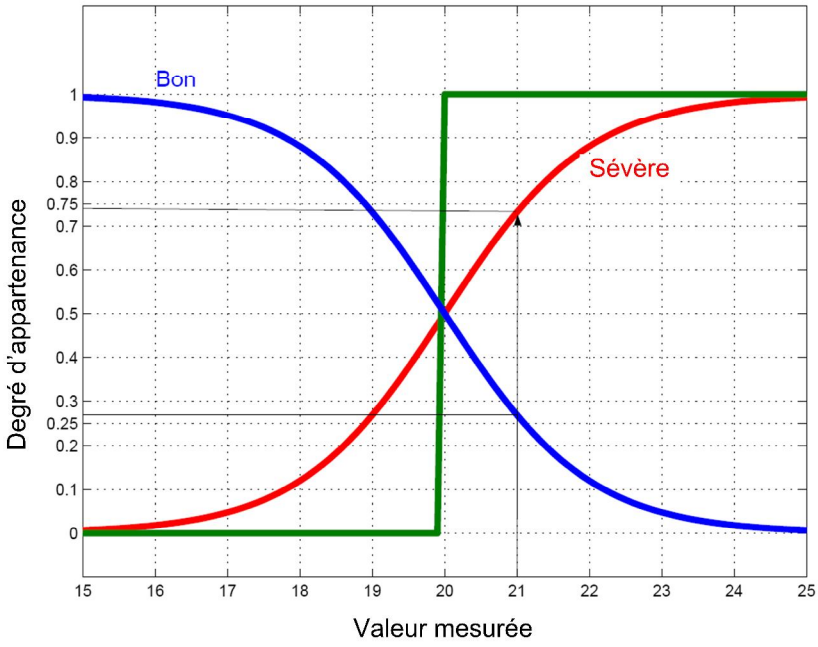


FIG. 4.12 – Exemple de degré d'appartenance

nément et fournissent des consignes différentes pour les sorties. Il faut donc disposer d'une méthode de composition ou agrégation des règles pour obtenir la valeur finale des sorties. L'agrégation est un processus par lequel les ensembles flous qui représentent les sorties de chaque règle sont combinés dans un seul ensemble flou. Elle intervient une seule fois pour chaque variable de sortie, juste avant l'étape de défuzzyfication. L'entrée du processus d'agrégation est une liste de fonctions de sortie tronquées par le processus d'implication de chaque règle. La sortie de l'agrégation est un ensemble flou pour chaque variable de sortie. Soient, par exemple, 3 règles dont la variable de sortie prend respectivement les valeurs linguistiques faible, moyen et élevé. Si après troncature en phase d'implication, on passe à l'agrégation des règles en utilisant la méthode du maximum l'ensemble flou de la sortie se présentera comme illustré à la figure 4.13

Pour la défuzzyfication, la méthode des centroïdes est la plus performante. Le centre de gravité fixe la valeur que prend la sortie. Les fonctions d'appartenance peuvent être quelconques mais avec l'obligation de ne prendre que des valeurs comprises entre 0 et 1.

Un système flou peut aussi être construit à partir des données par un processus d'apprentissage. C'est au cours de cet apprentissage que la fuzzyfication et la construction des règles s'effectuent [38]. Si l'on dispose déjà des règles d'expert pour le problème, on peut les intégrer au modèle. Ce processus d'adaptation des paramètres est très similaire à celui rencontré dans le

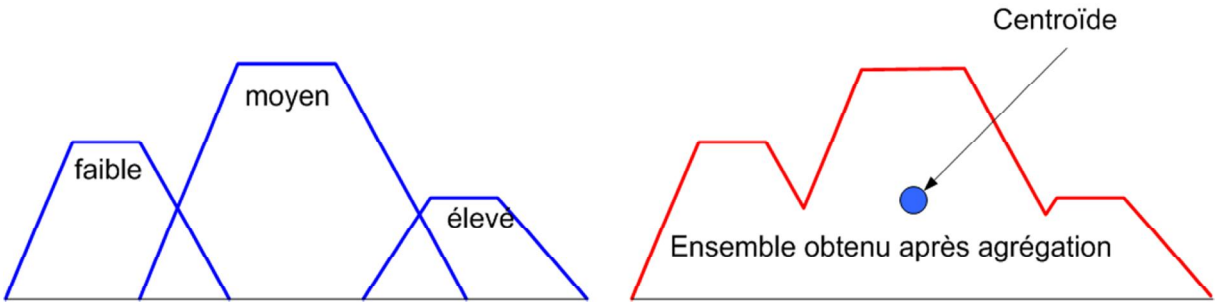


FIG. 4.13 – Agrégation

cas des réseaux de neurones et on utilise pour l'apprentissage l'algorithme de rétropropagation, seul ou en combinaison avec des méthodes de moindres carrés [39, 40]. Le type de système flou qui permet une adaptation des paramètres est le modèle de Sugeno [41]. Il est appelé *Adaptative Neuro-Fuzzy Inference System* (ANFIS), système adaptatif d'inférence floue. Le modèle de Sugeno est soumis à la contrainte que les fonctions d'appartenance pour une variable de sortie ne peuvent être que de deux sortes : soit linéaires soit constantes, les fonctions de degré supérieure n'améliorant pas les résultats. Les règles floues seront donc de la forme :

Si entrée1=x et/ou entrée2=y alors sortie est z=ax+by+c

on parle alors d'un modèle de Sugeno de premier ordre. Le modèle devient d'ordre zéro si a=b=0 c'est-à-dire le niveau de sortie z est constant.

Pour N entrées, x_i, avec m_i fonctions d'appartenance, $i = 1, ..., N$, et une sortie, y, le nombre total r de règles est $r = \prod_{i=1}^{N} m_i$.

L'algorithme d'apprentissage procède à la segmentation des entrées et des sorties par un processus de clusterisation. Chaque cluster correspond à une règle floue. Les différents clusters ne sont évidemment pas disjoints, ce qui explique le chevauchement des fonctions d'appartenance.

Les figures 4.14 et 4.15 illustrent le modèle ANFIS et les fonctions d'appartenance (de type gaussien) pour le défaut de frottement correspondant à l'étude dont sont issues les données du tableau 4.1 (Annexe A). La figure 4.16 montre la réponse du système à des données de test, pour lesquelles il a présenté 90, 7% de probabilité de reconnaissance. La sortie étant continue, l'affectation des classes peut se faire en choisissant la classe la plus proche. Les classes pour le défaut sont codé en binaire : 1 pour la présence du défaut et 0 pour son absence. Le test est effectué sur 162 exemples alors le système a été entraîné sur 486 exemples. Le modèle génère des règles floues de la forme :

Si (RMS glob z1 est *in1mf1*) et (RMS glob z2 est *in2mf1*) et (RMS glob y2 est *in3mf1*) alors (Frottement est *out1mf1*)

où *in1mf1* représente la première fonction d'appartenance de l'entrée 1 (*inmf*≡*input membership function*)

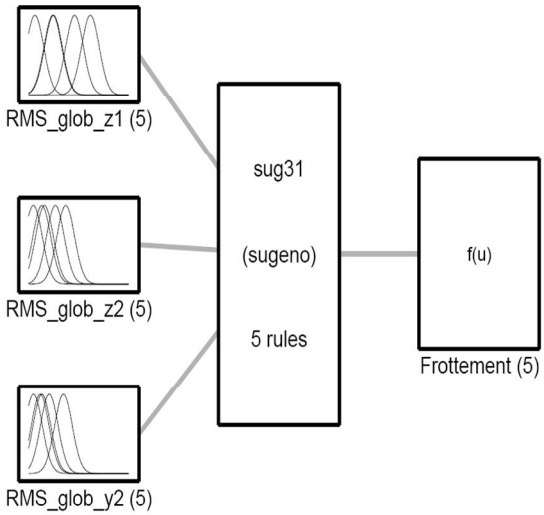


FIG. 4.14 – Modèle flou pour le diagnostic d'un défaut de frottement

4.3.8 Les K plus proches voisins

La méthode de classsification par les K plus proches voisins (*K Nearest Neighbors*) est une méthode basée sur les instances. Le classifieur procède à une mémorisation des instances d'apprentissage, et use de la notion de distances pour assigner une classe à un nouvel objet (instance).

La distance la plus utilisée est la distance euclidienne qui est définie par

$$d = \sqrt{\sum_{\mathrm{i}=1}^{n} (x_{1\mathrm{i}} - x_{2\mathrm{i}})^2} \tag{4.39}$$

pour deux objets $\mathbf{x}_1 = (x_{1\mathrm{i}}|i = 1 \cdots n)$ et $\mathbf{x}_2 = (x_{2\mathrm{i}}|i = 1 \cdots n)$

Puisque l'objectif est de comparer des distances, il n'est pas nécessaire d'effectuer la racine carrée, les sommes de carrés peuvent être directement comparées. Il existe d'autres définitions

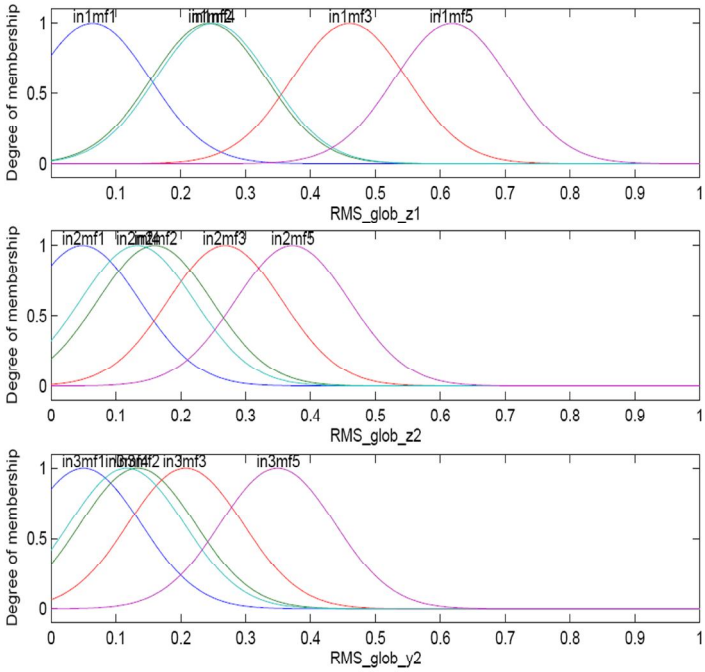


FIG. 4.15 – Fonction d'appartenance pour le défaut de frottement (les indicateurs sont normalisés)

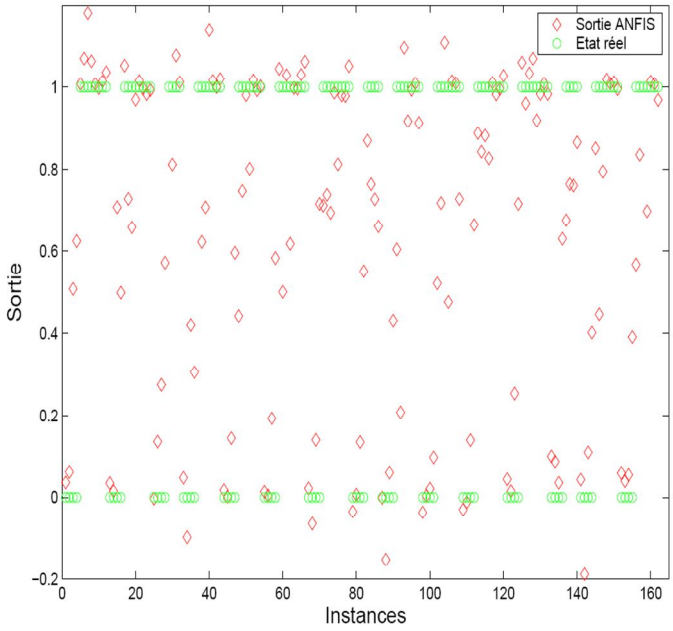


FIG. 4.16 – Classification fournie par l'*ANFIS* pour le frottement

de la distance, mais la distance euclidienne reste le meilleur compromis :

– distance de Manhattan ou *city-block* :

$$d = \sum_{\mathrm{i}=1}^{n} |x_{1\mathrm{i}} - x_{2\mathrm{i}}|$$

– p-distance :

$$d = \sqrt[p]{\sum_{\mathrm{i}=1}^{n} |x_{1\mathrm{i}} - x_{2\mathrm{i}}|^p}$$

Comme les attributs sont généralement mesurés sous différentes échelles, ils doivent être normalisés pour que certains d'entre-eux ne viennent masquer les autres.

Quand les attributs sont nominaux, la distance est simplement calculée en tenant compte des définitions suivantes :

$$\begin{cases} |x_{1\mathrm{i}} - x_{2\mathrm{i}}| = 1 \quad si \quad x_{1\mathrm{i}} \neq x_{2\mathrm{i}} \\ \\ |x_{1\mathrm{i}} - x_{2\mathrm{i}}| = 0 \quad si \quad x_{1\mathrm{i}} = x_{2\mathrm{i}} \end{cases} \tag{4.40}$$

Pour affecter une classe à un nouvel objet, on considère les k objets qui lui sont les plus proches et on choisit la classe la plus fréquente parmi ces voisins. Une démarche plus judicieuse de décision consiste à pondérer les classes du voisinage avec l'inverse de la distance à l'exemple à classer. Ainsi la classification se fait sur base des poids et non sur base de la simple majorité.

Le choix de la meilleure valeur de k dépend de la structure de données sur lesquelles on travaille. Généralement une grande valeur k réduit l'effet du bruit sur la classification, mais elle conduit à des frontières moins nettes entre classes. On pourrait utiliser la validation croisée pour choisir k. Une heuristique fréquemment utilisée est de choisir k égal au nombre d'attribut plus 1.

4.4 Des instances non étiquetées

En condition monitoring, l'identification des défauts est essentiellement un problème de classification. Il faut associer à un vecteur d'indicateurs noté $\mathbf{x}$ une classe C_i par la relation $h(\mathbf{x}) = C_i$, où h est le classifieur. Il a été indiqué plus haut que la construction du classifieur demandait la fourniture d'un ensemble de couple $E_i = (\mathbf{x}, C_i)$ qui sont des exemples d'appren-

tissage. Mais il peut arriver que l'on se trouve en présence d'instances non étiquetées. Assigner des classes à ces instances relève du domaine de l'apprentissage non supervisé.

L'intérêt de l'étude des possibilités de construction d'un classifieur à partir des données non étiquetées est réel dans le cas de la surveillance vibratoire. Dans l'environnement industriel, deux raisons l'expliquent :

1. il est fréquent que les données soient archivées sans information supplémentaire nécessaire à leur future analyse ;

2. les informations nécessaires à l'étiquetage peuvent avoir été archivées dans d'autres fichiers de gestion des activités de maintenance, sans lien direct avec les données de surveillance. Dans ce cas vu le possible gros volume de données, imaginer un étiquetage ne peut qu'être onéreux.

Dans une première approche, on pourrait penser à l'utilisation des méthodes de clusterisation pour une découverte non supervisée des classes dans le jeu de données. Cependant, dans le contexte de l'identification des défauts, cette démarche serait vide de tout intérêt si l'on arrivait pas à associer à chaque cluster identifié un état de fonctionnement donné de l'équipement. La littérature sur le *machine learning* n'est pas très abondante à ce sujet, mais l'approche introduite par Blum et Mitchell [42] a le mérite d'avoir inspiré quelques travaux dont l'idée était de combiner un petit nombre d'exemples étiquetés et des exemples non étiquetés pour la construction d'un classifieur. Cette approche peut être exploitée dans le traitement des données industrielles de *condition monitoring*.

En effet, la démarche peut être de construire, avec les quelques données étiquetées, un classifieur que l'on applique sur les exemples sans étiquettes afin de leur en assigner une pour, au final, construire de manière supervisée un classifieur définitif sur tout l'ensemble des données. Une autre possibilité est de procéder à une clusterisation sur tous les exemples d'apprentissage et d'assigner aux clusters les classes suggérées par les exemples étiquetés ayant participé à l'apprentissage. Du fait de la possibilité de rencontrer plus d'un exemple étiqueté dans un cluster, nous proposons une approche de clusterisation floue au bout de laquelle la compétition des degrés d'appartenance au cluster des exemples initialement étiquetées servira à l'étiquetage des classes.

4.5 Sélection des indicateurs pertinents

Il a été prouvé expérimentalement qu'avec un nombre élevé d'indicateurs, les performances de classification peuvent se trouver sensiblement dégradées [43]. Les capacités de généralisation d'un classifieur opérant dans un espace d'indicateurs de grande dimension sont moins bonnes par rapport à celles d'un classifieur opérant dans un espace de faible dimension. En outre, l'utilisation d'indicateurs non pertinents peut sensiblement entamer les performances d'un système de classification. A cause de ces effets néfastes sur la plupart des schémas d'apprentissage et sur les capacités de généralisation, il est commun de procéder à une sélection de variables avant un processus de classification. Obtenir la meilleure classification possible avec le minimum d'indicateurs mène à des résultats très compacts et facilement interprétables [44].

La sélection des indicateurs pertinents nécessite les ingrédients suivants :

- un critère d'évaluation des indicateurs pour comparer les sous-ensembles d'indicateurs ;
- une procédure d'exploration de l'espace d'indicateurs ;
- un critère d'arrêt de la recherche.

On distingue essentiellement deux groupes de méthodes d'évaluation des indicateurs [11, 45, 46]. Le premier reprend les méthodes qui se basent seulement sur les données sans tenir compte du processus d'apprentissage, ce sont les méthodes du filtre. Elles consistent à hiérarchiser les variables en examinant un quelconque paramètre. Il peut s'agir de la corrélation avec les classes, d'un critère informationnel ou des paramètres d'un modèle linéaire. Les méthodes du deuxième groupe mettent en œuvre l'évaluation des sous-ensembles en faisant usage d'un algorithme de classification enveloppé dans le processus de sélection.

Pour l'exploration de l'espace d'indicateurs, une recherche exhaustive des sous-ensembles est évidemment à proscrire. On peut partir de tout l'ensemble d'indicateurs et procéder par élimination un à un en analysant le critère d'évaluation (*backward*), ou partir de rien et ajouter un à un les indicateurs (*forward*). Ces méthodes pas-à-pas de recherche nécessiteront un critère d'arrêt par stopper la recherche. Le critère d'arrêt peut être basé sur une statistique ou une simple prédétermination de la taille de l'ensemble d'indicateurs à retenir.

Les algorithmes génétiques sont aussi utilisés pour la recherche des sous-ensembles d'indicateurs [47–49].

4.5.1 Les méthodes du filtre

Ces méthodes de sélection ont lieu avant que le processus d'apprentissage ne soit mis en œuvre. Elles sont indépendantes du classifieur à utiliser par la suite. Présentons quelques méthodes de sélection à base de filtres.

Critère de corrélation

Le coefficient de corrélation de Pearson, entre les indicateurs et les classes de défauts, est utilisé comme critère de sélection [50]. L'exploration pas-à-pas vérifiera un critère de redondance avant d'ajouter un indicateur au sous-ensemble en sélection. Ce critère de corrélation suppose une dépendance linéaire entre les classes et les indicateurs.

Critère du gain d'information

On peut définir une mesure d'information U pour la prédiction d'une classe de telle manière qu'une grande valeur de U représente un niveau élevé d'incertitude [51]. Étant donnée une fonction d'incertitude et les probabilités a priori $P(C_k)$ des classes, le gain d'information $GI(x)$ correspondant à un indicateur x, est défini comme la différence entre l'incertitude a priori $\sum_k U\,[P(C_k)]$ et l'incertitude a posteriori :

$$GI(x) = \sum_k U\,[P(C_k)] - E\left(\sum_k U\,[P(C_k|x)]\right) \tag{4.41}$$

On sélectionnera l'indicateur x_i au lieu de x_j si

$$GI(x_i) > GI(x_j)$$

Donc un indicateur est sélectionné s'il réduit l'incertitude. La fonction d'incertitude la plus utilisée est l'entropie de Shannon. C'est cette logique qui est utilisée si un arbre de décision est utilisé comme moyen de sélection.

Analyse en composantes principales

L'ACP peut aussi être vue comme une méthode de sélection d'indicateurs à travers sa phase de réduction dimensionnelle [52]. Mais l'inconvénient de cette approche c'est qu'elle ne rencontre pas le besoin d'interprétation facile que poursuit la sélection de variables [53, 54]. Ceci

est dû au fait que l'on procède dans l'espace des composantes principales et non dans celui des variables originales.

Une méthode proposée dans [55] considère pour chaque composante principale, l'indicateur original qui correspond au plus grand coefficient. Cette méthode est très intuitive et moins coûteux en calcul, mais le fait de considérer chaque axe séparément risque d'aboutir à des indicateurs redondants. Cohen *et al.* [56] ont proposé une approche qui part des vecteurs composés des coefficients de chaque indicateur dans la matrice de transformation, puis procède à leur clusterisation en un nombre de classe égal au nombre d'indicateurs à sélectionner. Dans chaque cluster, l'indicateur correspondant au vecteur le plus proche du centre est retenu.

Analyse discriminante

L'utilisation de l'analyse discriminante comme moyen de sélection des indicateurs se base sur la notion des matrices de covariance intraclasse **W** et interclasse **B** [52, 57].

Considérons M observations reparties entre K classes de défauts, chaque observation étant décrite par N indicateurs. Chaque classe est caractérisée par son effectif n_k et son centre de gravité $\mathbf{g}_k$. Les matrices de covariance des classes sont notées $\mathbf{W}_k$. On définit

– la variance intraclasse par

$$\mathbf{W} = \frac{1}{M} \sum_k n_k \mathbf{W}_k \tag{4.42}$$

– et la variance interclasse par

$$\mathbf{B} = \frac{1}{M} \sum_k n_k (\mathbf{g}_k - \mathbf{g})(\mathbf{g}_k - \mathbf{g})^T \tag{4.43}$$

où $\mathbf{g}$ est le centre de gravité de tout l'ensemble des observations.

La variance globale sera donnée par

$$\mathbf{V} = \mathbf{W} + \mathbf{B} \tag{4.44}$$

Les indicateurs les plus discriminant devront maximiser **B** et minimiser **W**. La sélection des indicateurs se fait pas-à-pas en observant un critère basé sur le lambda de Wilks Λ. Celui-ci est

donné par

$$\Lambda = \frac{\det(\mathbf{W})}{\det(\mathbf{V})} \tag{4.45}$$

Dans la procédure pas-à-pas, on évaluera la variation du lambda de Wilks quand on passe du sous-ensemble à j variables à celui à $j+1$ variables. Le statistique de test est

$$F = \frac{M-K-j}{K-1}\left(\frac{\Lambda_j}{\Lambda_{j+1}} - 1\right) \tag{4.46}$$

Cette statistique suit une loi de Fisher dont les paramètres sont $(K-1, M-K-j)$. Il suffira de se fixer un degré de confiance et de considérer la valeur critique de F comme critère d'arrêt [58].

En diagnostic vibratoire, la littérature fait état de l'utilisation de l'analyse discriminante principalement comme moyen de classification plutôt que de sélection des indicateurs [57, 59–62]. Dans d'autres domaines, les références abondent faisant usage de l'analyse discriminante pour la sélection [63] et pour la classification [64, 65].

4.5.2 Les méthodes d'enveloppe

Les méthodes d'enveloppe utilisent un classifieur pour procéder à une sélection. Le critère d'évaluation est alors le taux de classification [66]. La plus simple méthode utilise un classifieur 1R (*one rule*) et ne retient que les indicateurs qui auront donné une règle à taux de classification « satisfaisant ». Il peut aussi arriver que le critère d'arrêt soit un nombre d'indicateurs fixé a priori.

Les arbres de décision ont aussi été utilisés pour la sélection de variables [67, 68]. L'idée est de ne retenir que les indicateurs qui sont utilisés dans la construction de l'arbre. Il a aussi été montré que généralement les quelques indicateurs situés près de la racine de l'arbre peuvent suffire pour une meilleure discrimination [11]. La référence [69] fournit un cas d'utilisation de ce type de sélection en diagnostic vibratoire.

La littérature rapporte l'utilisation des réseaux de neurones comme moyen de sélection des variables dans des domaines divers [70–72]. Leray *et al.* [70] ont distingué trois classes de méthodes qui mettent en œuvre les réseaux de neurones :

- les méthodes d'ordre 0, qui utilise juste les paramètres (poids) du réseau ;
- les méthodes d'ordre 1, qui utilisent les dérivées premières des paramètres par rapport aux variables ;

– les méthodes d'ordre 2 qui font usage des dérivées secondes.

Dans le domaine du diagnostic, Matsuura [73] a procédé à une sélection d'indicateurs vibratoires en analysant la sensibilité de la sortie d'un réseau de neurones aux variations des variables d'entrée.

4.6 Une approche hybride de sélection des indicateurs pertinents

Il est souvent fait usage de l'analyse en composantes principales pour réduire la dimension de l'espace d'indicateurs et éliminer toute redondance dans les données.

Dans le contexte de l'utilisation d'un classifieur de type réseau de neurones pour le diagnostic vibratoire, il est possible de choisir, par l'examen des poids synaptiques, les facteurs qui contribuent de manière déterminante à la construction de la sortie du réseau. Il peut alors être intéressant d'examiner ces facteurs pour déterminer les indicateurs originaux significatifs portés par leurs axes. C'est ainsi qu'au cours de nos travaux concernant la surveillance vibratoire des machines tournantes et de l'usure d'outils en usinage, nous avons élaboré une approche de sélection d'indicateurs basée sur l'analyse en composantes principales et les poids synaptiques d'un réseau de neurones [74].

Notre méthode de sélection d'indicateurs procède à la fois par un filtre constitué de l'ACP et par encapsulation d'un classifieur à réseau de neurones. Le processus, qui est représenté à la figure 4.17, considère d'une part le fait que lors de l'apprentissage, le réseau de neurones attribue aux différents facteurs, à l'entrée, des poids qui correspondent dans une certaine mesure à leurs contributions à la réponse du réseau de neurones, et d'autre part, que l'importance d'un facteur est liée à sa contribution à la variance totale des données.

Il s'agira donc dans un premier temps de compresser les données par ACP, puis de procéder à une classification dans l'espace des facteurs. Les facteurs à forte pondération sont retenus pour en déduire, par examen des saturations, les variables d'origine qu'ils portent.

Nous avons postulé que les variables d'origine pertinentes étaient portées par les facteurs dont les poids synaptiques à l'entrée du réseau étaient caractérisés par un bon regroupement autour d'une moyenne élevée. Ce qui a suggéré la définition d'un nombre pour quantifier l'importance du facteur en divisant la moyenne par l'écart-type :

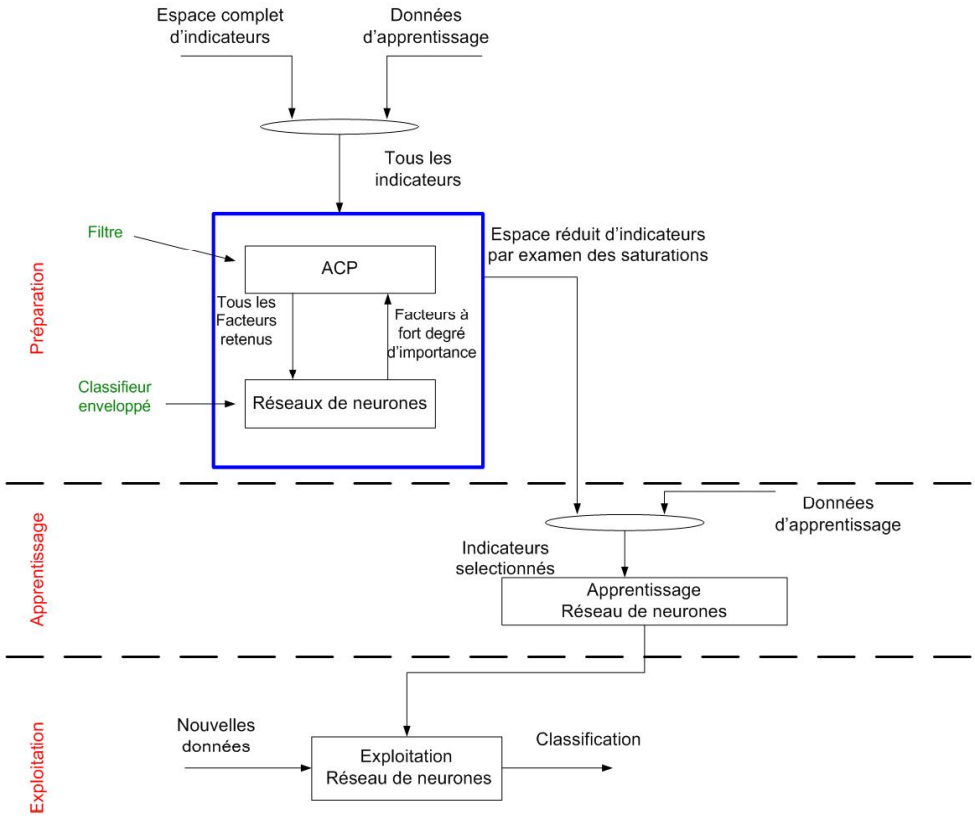


FIG. 4.17 – Processus de sélection de variables pertinentes par encapsulage d'un réseau de neurones

$$d_i = \frac{mean\,[abs(IW_i)]}{std\,[abs(IW_i)]} \tag{4.47}$$

où

- *mean*, *std* et *abs* représentent respectivement le calcul de la moyenne, de l'écart-type et de la valeur absolue ;
- IW_i, le vecteur des poids synaptiques associés au facteur i ;
- d_i, le degré d'importance du facteur i ;

Comme il ne peut être fixé de seuil sur le degré d'importance d_i, le choix est automatisé par clusterisation (*K-means*) en 2 classes. Pour les saturations, un minimum de 0,75 a été fixé pour retenir les variables.

Application expérimentale sur un banc d'essai

Nous nous proposons d'appliquer cette approche de sélection sur les données d'un banc d'essai et d'en comparer les résultats à ceux de certaines approches classiques de sélection des indicateurs pertinents.

Considérons l'étude expérimentale effectuée sur des données d'un banc d'essai où ont été induits quatre types de défauts : le balourd, le frottement, l'ovalisation de la bague externe d'un roulement et le délignage (voir annexe A pour les détails sur l'essai). Les signaux vibratoires sont relevés à 5120 Hz au niveau des deux paliers suivant deux directions orthogonales. La figure 4.18 montre une représentation schématique du banc.

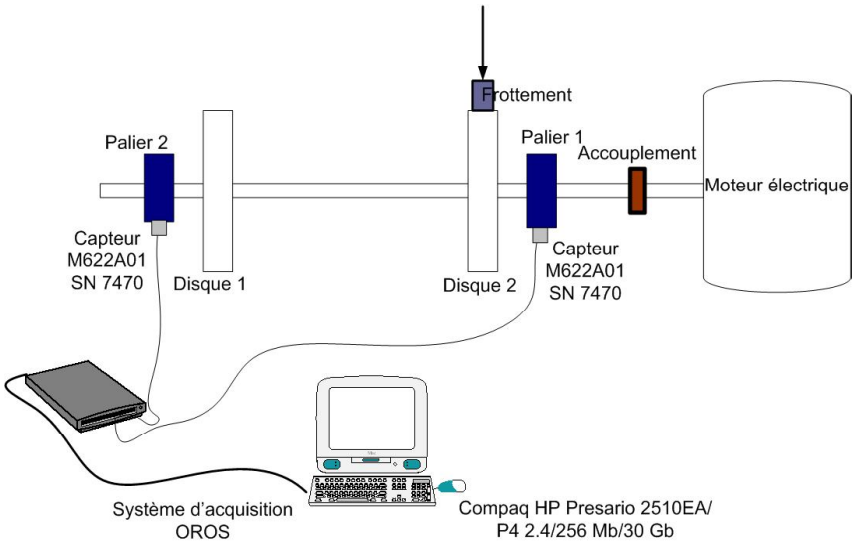


FIG. 4.18 – Représentation schématique du banc d'essai

Nous utilisons des indicateurs statistiques du domaine temporel. Leur désignation est réalisée en combinant les 3 colonnes du tableau 4.4 (par exemple : RMS 0-30y2). Ce qui conduit à $3 \times 4 \times 4 = 48$ variables.

Indicateur	**Gamme fréquentielle (Hz)**	**Point de mesure**
RMS	0-30	z1
Kurt	30-300	y1
Skew	300-2000	z2
	glob (0-2000)	y2

TAB. 4.4 – Nomenclature des variables construites

Fusion de données

Dans cet exemple, on a dû effectuer une double opération de fusion de niveau 2 : fusion d'informations provenant des gammes de fréquences différentes et fusion d'informations issues des différents points et directions de mesure. La fusion est entreprise par une simple opération d'alignement ou de concaténation pour former les vecteurs d'indicateurs. C'est ce qui a conduit à 48 variables. La base de données est constituée de 648 exemples.

Réduction dimensionnelle et sélection des variables pertinentes

Les vecteurs d'indicateurs sont de dimension 48. Pour réduire la dimension du vecteur d'entrée, une analyse en composantes principales est effectuée de manière à produire des données non corrélées entre elles. Ensuite il n'est retenu que les directions principales qui contribuent le plus à la variation dans l'ensemble des données. Le choix de la dimension finale à retenir a été déterminé par un *scree test* (figures 4.19 et 4.20). Ainsi, le vecteur d'entrée se trouve-t-il réduit à la dimension 13 (13 facteurs).

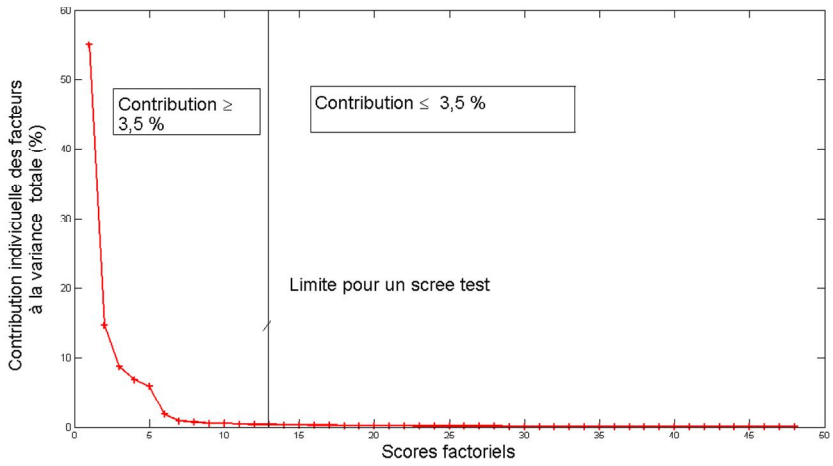


FIG. 4.19 – Scree test et Critère de la contribution individuelle à la variance totale

Une première analyse sur la matrice des saturations des indicateurs du domaine temporel conduit à relever que les facteurs 1, 2, 3, 4 et 5 présentent les plus fortes corrélations avec les variables de départ tandis que les autres facteurs sont les moins corrélés avec les variables initiales. Les graphiques de la figure 4.21 représentent la corrélation de chaque facteur avec les variables d'origine.

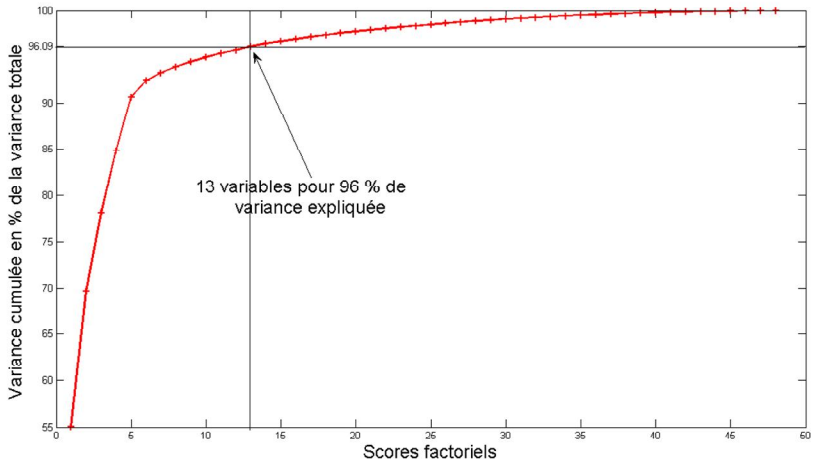


FIG. 4.20 – Critère de la proportion cumulée de la variance

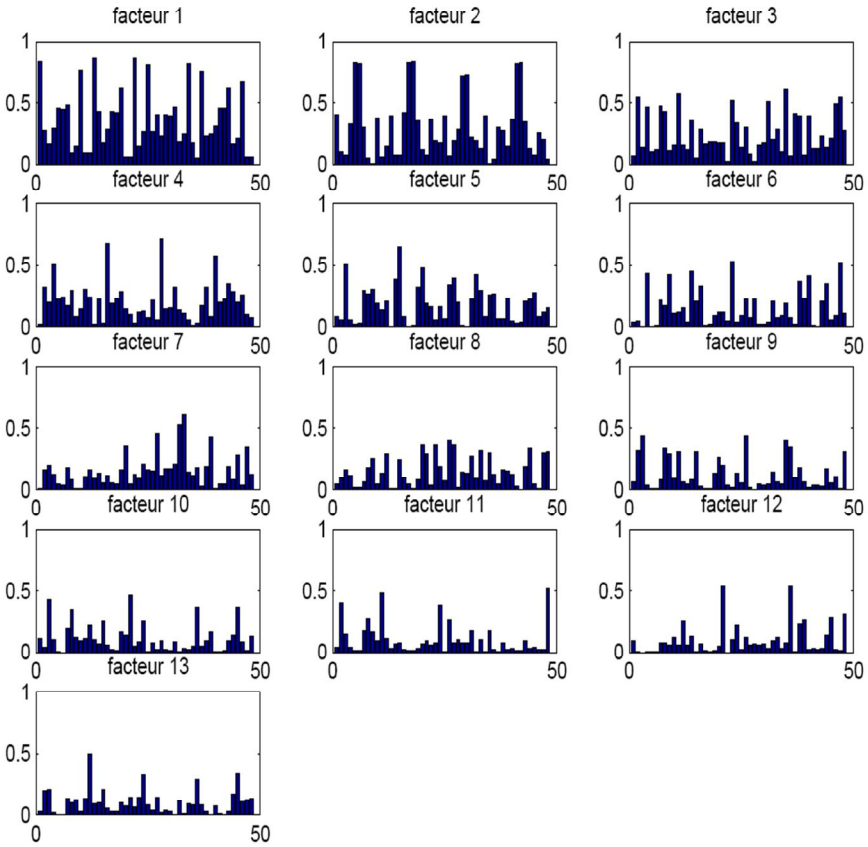


FIG. 4.21 – Valeurs absolues des saturations. En abscisse sont repris les indicateurs temporels et en ordonnées les valeurs absolues des corrélations de ces indicateurs avec le facteur considéré

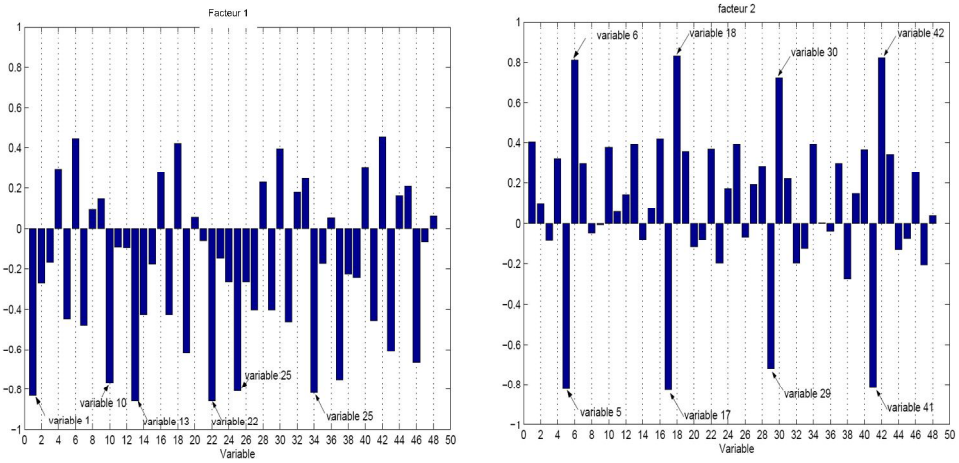


FIG. 4.22 – Saturations pour les facteurs 1 et 2

En procédant par encapsulation d'un seul classifieur pour tous les défauts, nous avons retenu les deux premiers facteurs. Les saturations des indicateurs portés par ces axes ont montré que les plus pertinents étaient (figure 4.22) :

- la valeur efficace du signal global (toutes les directions)
- la valeur efficace du signal haute fréquence (300-2000 Hz, toutes les directions)
- le *kurtosis* en basse fréquence (0-30 Hz, toutes les directions)
- et le *skewness* en basse fréquence (0-30 Hz, toutes les directions)

L'approche par encapsulation d'un classifieur pour chacun des défauts a conduit à retenir les variables reprises dans le tableau 4.5 pour la combinaison ACP-poids synaptiques. Les tableaux 4.6 et 4.7 fournissent les indicateurs sélectionnés avec l'analyse discriminante (95% de confiance) et l'arbre de décision (trois variables situées à la racine).

Les réseaux de neurones MLP utilisés pour la sélection ont une structure avec une couche cachée dont le nombre de nœuds est déterminé par essai et erreur. Ainsi, pour le cas où un seul réseau est utilisé pour tous les défauts, on a une couche cachée à 80 neurones et pour la sélection défaut par défaut, des réseaux avec une couche à 25 neurones ont été utilisés.

L'examen de ces résultats de sélection montre que certains indicateurs sont communs à ces processus de sélection. On rappelera que le problème de sélection est fort dépendant de la méthode et du critère utilisés tout comme de l'ordre de la recherche (*forward ou backward*).

Pour le défaut de balourd l'indicateur RMS 0-30 y1 est sélectionné par l'analyse discriminante et par notre approche. Nous nous souviendrons cependant que pour l'arbre de décision,

nous n'avions retenu que les trois premiers indicateurs. Mais l'analyse de l'arbre complet a montré que RMS 0-30 y1 a aussi participé à sa construction. L'indicateur RMS 0-30 y2 a été retenu par l'arbre de décision et par notre approche. Le *kurtosis* Kurt 0-30 y2 été sélectionné par toutes les trois méthodes. Si il est évident que le balourd influence la valeur efficace dans la gamme de fréquence 0-1$\times f_{rot}$, la sélection du *kurtosis* dans la même gamme comme indicateur décisif pour le balourd semble être un apport non négligeable des méthodes d'apprentissage par rapport au diagnostic vibratoire traditionnel.

Concernant le défaut de frottement, ces méthodes de sélection ont en commun avec notre approche soit l'indicateur RMS glob z1 soit RMS glob z2. Ceci traduit le fait que l'influence du frottement ne peut être localisé dans une gamme précise. Pour le roulement, l'indicateur RMS 300-2000 y2 est sélectionné par les 3 méthodes, les indicateurs RMS 30-300 z1, RMS glob y2 et Skew 30-300 z1 par notre méthode et l'analyse discriminante, tandis que l'analyse discriminante et l'arbre de décision se recoupe sur RMS 30-300 y2 . Le choix de ces indicateurs restent compatible avec la manifestation des défauts de roulement dans les gammes hautes fréquences. Pour l'alignement aussi, les trois méthodes se recoupent sur certains indicateurs.

Cette analyse nous permet de confirmer les conclusions auxquelles sont arrivés Guyon et Elisseeff [44] selon lesquelles des variables au départ moins importantes peuvent gagner en pertinence si elles sont combinées. Ceci, parce que chacune des méthodes a pu retenir des indicateurs non repris par les autres.

Dans la suite, nous allons comparer la pertinence des variables ainsi sélectionnées sur plusieurs schémas de classification.

Zio *et al.* [47, 75] ont travaillé sur le même ensemble de données en utilisant les algorithmes génétiques et les plus proches voisins pour la sélection des indicateurs pertinents. Les résultats auxquels cette équipe a abouti sont comparables aux nôtres. En effet, le rapport sur ces travaux [76] fait état des ensembles d'indicateurs qui se recoupent avec ceux obtenus dans nos travaux. Pour le balourd, ils ont retenus entre autres les indicateurs RMS 0-30 y1 et RMS 0-30 z2. Pour le frottement, les indicateurs RMS glob z1, RMS glob y1 et RMS glob y2.

4.7 Mesure des performances d'un système de diagnostic

Comment évaluer la performance et l'efficacité d'un système de diagnostic ?

Pour répondre à cette question, il convient de définir certaines métriques qui tiennent compte des aspects techniques et économiques d'une mauvaise décision de la part d'un système de

Défaut	Indicateurs sélectionnés
Balourd	RMS 0-30 Hz-y1
	RMS 0-30 Hz-y2
	Kurt 0-30 Hz-y2
Frottement	RMS glob-z1
	RMS glob-z2
	RMS glob-y2
Roulement	RMS 30-300 Hz-z1
	RMS 300-2000 Hz-y2
	RMS glob-y2
	Kurt 0-30 Hz-y1
	Kurt glob-z2
	Skew 30-300-z1
Alignement	RMS 30-300 Hz-y2
	RMS 300-2000 Hz-z2
	Skew 30-300-z2

TAB. 4.5 – Variables pertinentes-Sélection par ACP et réseau de neurones

diagnostic.

Dans la communauté de *machine learning*, l'évaluation d'un classifieur passe principalement par le nombre de faux positifs et de faux négatifs. Mais en diagnostic, il est important de compléter ces métriques avec le délai de temps qui s'écoule entre le début d'un défaut et sa détection.

Les faux négatifs présentent un risque majeur sur l'état de l'équipement, parce que des défauts dont on a raté la détection peuvent conduire à une défaillance catastrophique pouvant aboutir en des pertes en vies humaines ou en une perte totale de l'équipement. Les faux positifs, ou fausses alarmes, peuvent amener à des pertes de production pour arrêts inutiles et conduire à une perte de confiance de l'opérateur vis-à-vis du système de diagnostic.

La performance de détection ou de diagnostic de défauts se mesure à travers une matrice de confusion qui se présente comme montré dans le tableau 4.8 pour un cas de diagnostic mono-défaut où l'on a deux classes F_1 et F_0, représentant respectivement la présence et l'absence du défaut.

En machine learning, on définit les indicateurs de performance suivants :

- Vrais positifs : $VP = a$
- Faux positifs : $FP = b$
- Taux d'erreur : $TE = \frac{c+b}{a+b+c+d}$
- Sensibilité ou rappel. C'est le taux de VP : $TVP = \frac{a}{a+c}$
- Précision : $Précision = \frac{a}{a+b}$

Défaut	Indicateurs sélectionnés
Balourd	RMS 0-30 Hz-y1
	RMS 30-300 Hz-y1
	Kurt 0-30 Hz-y2
	Kurt glob-y1
	Kurt 300-2000 Hz-z1
	Skew 0-30 Hz-y2
Frottement	RMS glob-z1
	RMS 30-300-z2
	Kurt 300-2000 Hz-y2
	Skew glob-z2
	Skew 30-300 Hz-z2
Roulement	RMS 30-300 Hz-z1
	RMS 30-300 Hz-z2
	RMS 30-300 Hz-y2
	RMS 300-2000 Hz-y1
	RMS 300-2000 Hz-y2
	RMS glob-y1
	RMS glob-y2
	Kurt 300-2000 Hz-y1
	Skew 30-300-z1
Alignement	RMS 30-300 Hz-y2
	RMS 30-300 Hz-z1
	RMS glob-y1
	RMS glob-y2
	RMS 300-2000 Hz-y2
	Skew 30-300-z2
	Skew 30-300-z1
	Kurt 30-300-y2
	Kurt glob-z2

TAB. 4.6 – Variables pertinentes- Sélection par analyse discriminante

Défaut	Indicateurs sélectionnés
Balourd	RMS 0-30 Hz-y2
	Kurt 0-30 Hz-z2
	Kurt 0-30 Hz-y2
Frottement	RMS glob-z2
	RMS 30-300-y1
	Kurt 300-2000-y2
Roulement	RMS 300-2000 Hz-y2
	RMS 300-2000 Hz-z1
	RMS 30-300 Hz-y2
Alignement	RMS 300-2000 Hz-z2
	Skew 30-300 Hz-z2
	Kurt glob-z2

TAB. 4.7 – Variables pertinentes-Sélection par arbre de décision

	Classe prédite		
	F_1	F_0	Total
F_1	a Nombre de défauts détectés	c Nombre de défauts non détectés	a+c Nombre total de défauts
F_0	b Nombre de fausses alarmes	d Nombre de rejets corrects	b+d Nombre total de cas sans défaut
Total	a+b Nombre total d'alarmes	c+d Nombre total de non-alarmes	a+b+c+d Nombre total de cas

TAB. 4.8 – Matrice de confusion

- Taux de FP : $TFP = \frac{b}{b+d}$
- Spécificité=1-Taux de FP : $Spécificité = \frac{d}{b+d}$

En diagnostic, on utilisera plutôt les métriques suivantes :

- la probabilité d'identification, pour un défaut donné, comme le rapport du nombre de défauts détectés sur le nombre total de cas potentiels à défauts :

$$PI = \frac{a}{a+c} \tag{4.48}$$

  La PI correspond à la sensibilité.
- la probabilité de fausses alarmes comme la proportion de tous les cas sans défauts qui ont déclenché une alarme

$$PFA = \frac{b}{b+d} \tag{4.49}$$

  La PFA correspond à aux taux de FP
- le taux de classification ou précision

$$TC = \frac{a+d}{a+b+c+d} \tag{4.50}$$

4.7.1 La validation croisée

D'une manière générale, l'évaluation d'un système de classification se fait sur un ensemble dit de test. Les proportions conseillées sont de 1 à 2 ou de 3 à 4 pour les ensembles de test et d'apprentissage respectivement.

Si l'on ne dispose pas d'un nombre élevé d'exemples, on peut procéder par la méthode de validation croisée à k blocs (*k-fold cross validation*) [77]. Cet algorithme consiste à créer, dans l'ensemble de données D, k sous-ensembles D_i disjoints tels que leur union donne l'ensemble

de données. Chaque bloc D_i est ensuite utilisé comme ensemble de test, l'apprentissage ayant été effectué sur le complément à D de D_i. A la fin du processus l'erreur de généralisation est estimée en faisant la moyenne des erreurs obtenues pour chaque bloc. On utilise généralement k=10.

4.7.2 Illustration de l'évaluation des performances

Pour illustrer l'utilisation des méthodes du *data mining* au diagnostic de défauts mécaniques, reprenons l'exemple de l'étude sur banc d'essai présentée à l'annexe A. Les quatre défauts concernés par cette étude ont chacun plusieurs modes ou niveaux d'expression :

- le balourd : 5 modes représentés par B_0, B_1, B_2, B_3 et B_4. Les modes B_1 et B_3 sont caractérisés par un balourd faible tandis les modes B_2 et B_4 sont relatifs à un balourd sévère, B_0 étant un mode de fonctionnement sans balourd. Les balourds B_1 et B_2 sont induits sur le disque 1 tandis que B_3 et B_4 le sont sur le disque 2 ;
- le frottement : 3 modes représentés par F_0, F_1 et F_2 ;
- le défaut du palier à roulement : 3 modes désignés par R_0, R_1 et R_2 ;
- et le défaut d'alignement : 3 modes désignés par A_0, A_1 et A_2 ;

Pour les défauts F, R et A, les modes 1 et 2 sont respectivement relatifs au niveau faible et au niveau sévère du défaut. Le mode 0 représentent l'absence du défaut considéré.

Présentons comment se construit une matrice de confusion et comment se calculent les probabilités d'identification et le taux de classification. Appliquons, en guise d'exemple, un classifieur au défaut de balourd pour identifier les différents modes.

Considérons les indicateurs définis dans le domaine temporel sur base de moments statistiques comme présenté sous la section 4.6. Utilisons un arbre de décision pour l'identification de différents modes du défaut de balourd. L'ensemble des indicateurs est utilisé sans sélection préalable. L'apprentissage et le test sont réalisés par validation croisée.

On aboutit à un arbre de très grande taille (97 branches, 49 feuilles). Le taux de classification est de 83,18%. La matrice de confusion est réprise dans le tableau 4.9

On peut ainsi calculer la probabilité d'identification d'un mode de fonctionnement donné. Prenons, par exemple, le mode à balourd B_1, on aura

$$PI(B_1) = \frac{117}{144} = 81,25\%$$

classified as	⇒	B_0	B_1	B_2	B_3	B_4	Total		PI
	B_0	136	4	0	4	0	144	→	94,44%
	B_1	3	117	3	11	10	144	→	81,25%
	B_2	0	5	72	1	18	96	→	75,00%
	B_3	5	15	1	108	3	132	→	81,82%
	B_4	0	8	9	9	106	132	→	80,30%
Total		144	149	85	133	137	648		

TAB. 4.9 – Matrice de confusion et probabilités d'identification

Et le taux de classification global sera

$$TC = \frac{136 + 117 + 72 + 108 + 106}{648} = 83,18\%$$

A l'examen des probabilités d'identification de modes de fonctionnement, on se rend compte que l'état libre de balourd B_0 se démarque bien des autres modes. Les 8 cas B_0 non reconnus sont classés comme B_1 ou B_3, ce qui est normal du fait que ces 2 modes sont caractérisés par un balourd faible. Ils sont donc moins distants de l'état sans balourd.

Remarquons aussi qu'une importante part de l'erreur de reconnaissance est provoquée par la symétrie des modes B_1 et B_2 et celle des modes B_3 et B_4. Sur les 109 objets mal reconnus, 53 sont dus à cette symétrie. Donc la sévérité du balourd est bien identifiée, le problème se trouve au niveau de la localisation. Si la localisation du balourd sur un des disques n'était pas importante, les probabilités d'identification deviendraient : $PI(B_0) = 94,4\%$, $PI(B_{\text{faible}} = B_1 \cup B_3) = 90,9\%$ et $PI(B_{\text{sevère}} = B_2 \cup B_4) = 89,9\%$.

4.8 Les différentes approches dans l'application du *data mining* au diagnostic

En aide au diagnostic vibratoire, les techniques de *data mining* peuvent servir à la résolution du problème suivant :
« étant donné l'état de fonctionnement d'une machine, représenté par un vecteur dans l'espace des indicateurs vibratoires, trouver l'image de ce vecteur -si elle existe- pour chacune des relations établies avec des ensembles représentant chacun un type de défaut ». Les éléments de ces ensembles peuvent être modélisés en terme d'existence ou non du défaut considéré. Alors dans ce cas on a un ensemble à deux éléments, 1 et 0, oui et non. Si, par contre, les éléments de

ces ensembles sont modélisés par des niveaux de sévérité du défaut, l'ensemble peut contenir n éléments représentant chacun un niveau de défaut (figure 4.23).

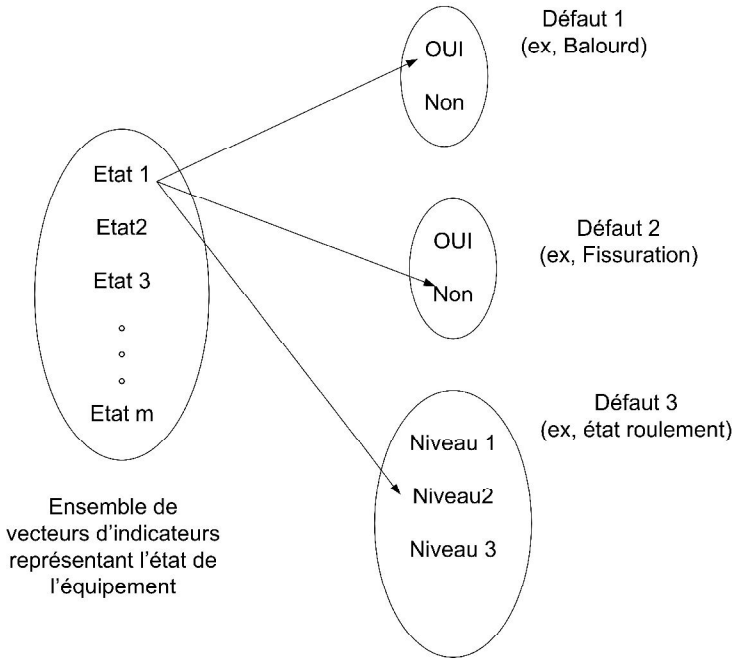


FIG. 4.23 – Problème de diagnostic

Il s'agit bien ici d'un problème de classification, chaque relation correspond à un classifieur. Et généralement un apprentissage supervisé peut efficacement s'appliquer aux problèmes de maintenance, car grâce à un retour d'expérience on peut présenter au classifieur un ensemble d'exemples étiquetés.

Selon la modélisation que l'on aura faite des données, il pourra être possible de procéder à une identification combinée ou séparée de différents défauts. La figure 4.23 correspond à une représentation où les défauts sont diagnostiqués séparément. Ainsi, de tels problèmes peuvent être abordés de trois manières :

1. en construisant un seul système de diagnostic multidéfaut multiclasse (**MuD-MuC**) : chaque niveau d'un défaut est considéré comme une classe. Les classes ne sont mutuellement exclusives que si elles représentent les modes d'un même défaut. C'est à ce niveau qu'une compétition des prédictions est effectuée pour prédire un seul de différents modes. Nous avons implémentée cette approche avec un réseau de neurones à 14 sorties [74], avec compétition par groupe de défaut (voir la figure 4.24). L'inconvénient de cette approche, si l'on utilise une réseau de neurones MLP, c'est d'avoir un nombre élevé de paramètres.

Pour ce cas, le réseau a exigé 80 noeuds dans la couche cachée (par essai et erreur !).

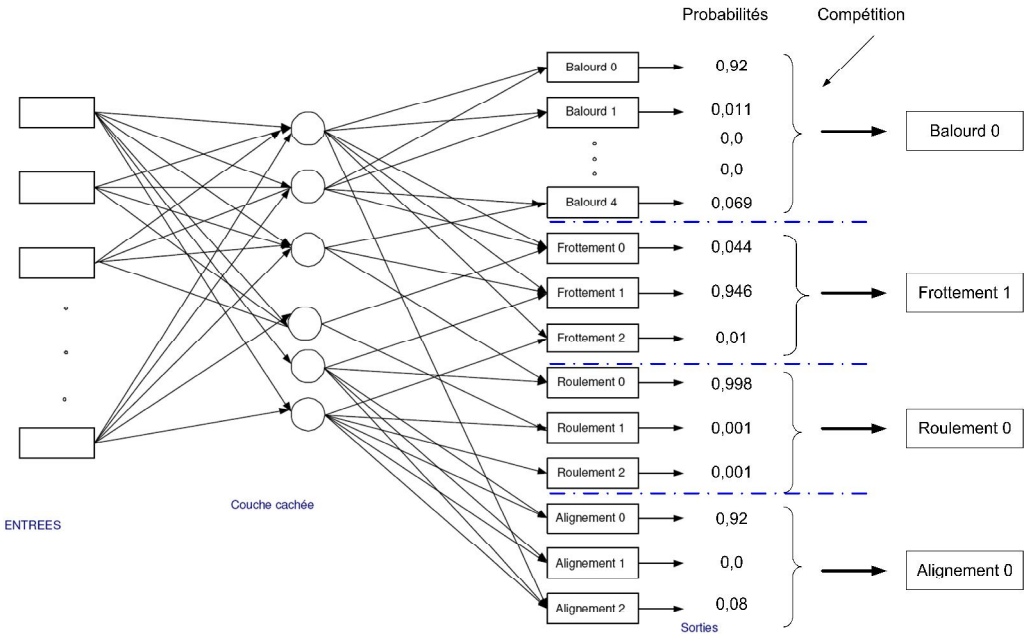


FIG. 4.24 – Architecture du réseau de neurones

2. en utilisant plusieurs systèmes monodéfaut multiclasse (**MoD-MuC**) : Les classes sont représentées par les modes du défaut considéré ;
3. en utilisant plusieurs systèmes monodéfaut biclasse (**MoD-BiC**) : on distingue juste l'existence et la non existence d'un type de défaut. Les classifieurs sont bi-classe. Cette approche peut se justifier dans les cas où en pratique, la distinction des modes d'expression d'un défaut ne présente aucun intérêt ;
4. en construisant un seul système de diagnostic multidéfaut biclasse (**MuD-BiC**)

Dans ce qui suit, nous allons maintenant illustrer le processus de diagnostic et comparer différents classifieurs, principalement avec l'approche **MoD-BiC** sur le problème du banc d'essai. L'apprentissage et l'évaluation des performances des classifieurs sont réalisés sur un ensemble de 648 cas par validation croisée à 10 blocs.

4.8.1 Application

L'identification de différents défauts est fait séparément en ne considérant que deux états pour chacun des défauts : présence ou absence du défaut. Nous nous en allons présenter les

résultats obtenus à l'aide de quelques méthodes de classification dont les arbres de décision et les réseaux bayésiens. La sélection de variables pertinentes est effectuée de trois façons :

- en utilisant un arbre de décision en amont, et seules les 3 variables à la racine de l'arbre sont retenues. L'ensemble de variables obtenu est noté « ER 1 » ;
- en appliquant l'analyse discriminante. L'ensemble de variables obtenu est noté « ER 2 » ;
- en appliquant la méthode que nous avons développée sur base de l'ACP et des poids synaptiques. On obtient un ensemble noté « ER 3 ».

On pourra se référer à la sous-section 4.6 pour la constitution de ces espaces réduits. L'ensemble complet sera noté « EC ».

Les réseaux de neurones de classification sont construits avec une seule couche cachée dont le nombre de neurones est choisi égale à la moitié de la somme du nombre de classes et du nombres d'indicateurs, selon la proposition empirique reprise dans [11]. Ce nombre de neurones peut ensuite être ajusté si le taux obtenu n'est pas satisfaisant, comparé aux autres classifieurs. Les fonctions d'activation sont des fonctions log-sigmoïdes. L'apprentissage des réseaux de neurones est réalisé par paquets (*batch training*).

Le balourd

Un arbre de décision construit sur les données sélectionnées fournit le modèle de la figure 4.25 dont l'évaluation donne 98,61% de succès. L'analyse de l'arbre de décision du défaut de balourd révèle deux faits essentiels :

1. la classification repose essentiellement sur les valeurs efficaces et *kurtosis* dans la gamme $1 \times f_{\text{rot}}$ (0-30 Hz), f_{rot} représente la fréquence de rotation ;
2. la présence du balourd semble être liée aux grandes valeurs de la valeur efficace de l'accélération tandis qu'un effet inverse est observé avec les valeurs du *kurtosis*.

En faisant usage de règles de classification, nous avons abouti à un taux de classification de 95%, avec l'avantage que l'expression de la règle est très simplifiée : *(RMS 0-30 y2 > 0,078869)* $\Rightarrow$ *B = oui.*

Les réseaux bayésiens ont été aussi expérimentés comme classifieurs. Nous avons considéré un réseau dans lequel un noeud pouvait avoir au maximum 2 parents.

Le tableau 4.10 synthétise les résultats obtenus sur l'espace global d'indicateurs et sur l'espace réduit aux seuls indicateurs jugés pertinents.

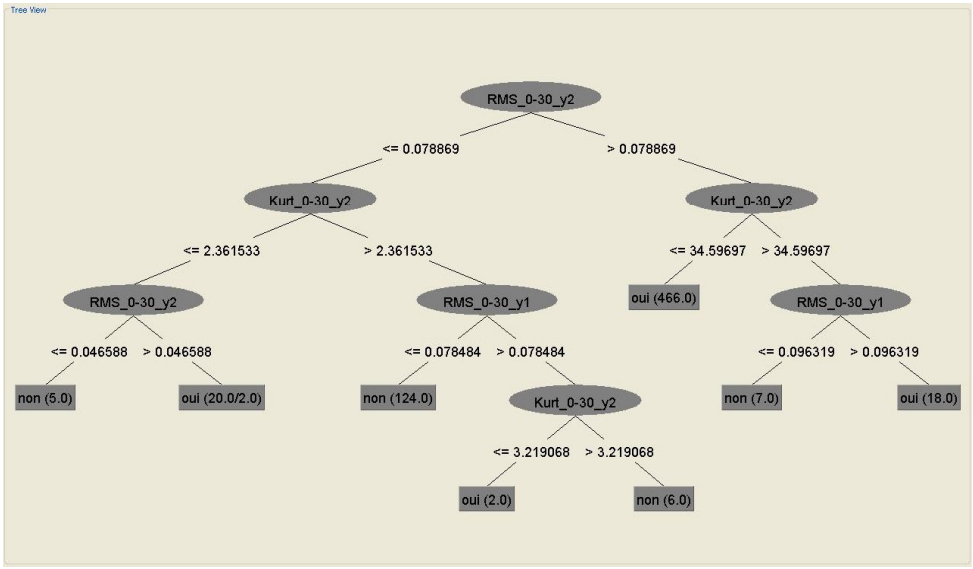


FIG. 4.25 – Arbre de décision relatif au balourd sur les données d'apprentissage (construit avec le logiciel Weka)

Méthode	Taux de classification			
	EC	ER 1	ER 2	ER 3
Arbre de décision	97,7%	98,0%	97,7%	98,6%
Règle de classification (conjonctive)	95,2%	95,5%	91,2%	95,5%
Réseau Bayésien (2 parents au plus par noeud)	98,8%	98,0%	98,8%	98,5%
K-NN (1 voisin)	98,5%	98,5%	98,5%	99,2%
RN (MLP avec 1 couche cachée)	99,2%	98,2%	98,5%	98,9%
RN (sans couche cachée)	99,2%	98,1%	98,3%	98,6%

TAB. 4.10 – Résultats de reconnaissance du défaut de balourd

Le frottement

En appliquant les mêmes classifieurs à la reconnaissance du frottement, on aboutit aux résultats du tableau 4.11

Méthode	Taux de classification			
	EC	ER 1	ER 2	ER 3
Arbre de décision	97,5%	94,4%	93,5%	97,2 %
Règle de classification (conjonctive)	90,1%	90,1%	75,3%	90,9%
Réseau Bayésien (2 parents au plus par noeud)	94,8%	94,9%	86,9%	91,7%
K-NN (1 voisin)	97,2%	97,1%	97,2%	98,8%
RN (MLP avec 1 couche cachée)	97,5%	92,6%	93,5%	96,3%
RN (sans couche cachée)	97,5%	93,7%	91,0%	90,1%

TAB. 4.11 – Résultats de reconnaissance du défaut de frottement

La règle s'énonce comme suit :
RMS glob z2 $\leq$ 1,407605 $\Rightarrow$F=non

Ce qu'on peut retenir de la structure de cette règle et de l'arbre de décision, c'est que le défaut de frottement influence seulement le niveau RMS global dans la direction Z.

Le défaut de roulement

Le tableau 4.12 reprend les scores de classification obtenus pour ce défaut : La règle apprise

Méthode	Taux de classification			
	EC	ER 1	ER 2	ER 3
Arbre de décision	95,2%	93,1%	96,8%	96,8%
Règle de classification (conjonctive)	86,9%	86,9%	86,9%	86,9%
Réseau Bayésien (2 parents au plus par noeud)	94,6%	88,9%	91,0%	93,5%
K-NN (1 voisin)	99,5%	97,8%	99,2%	99,2%
RN (MLP avec 1 couche cachée)	98,9%	86,0%	97,7%	94,1%
RN (sans couche cachée)	97,8%	88,9%	97,5%	86,4%

TAB. 4.12 – Résultats de reconnaissance du défaut du roulement

est : *(RMS 300-2000 y2* $\leq$ *0,334924)* $\Rightarrow$ *R = non*. Nous pouvons dire que le défaut de roulement a élevé le niveau de vibration dans la gamme de fréquence au delà de $10\text{x}f_{\text{rot}}$.

Le délignage

Le tableau 4.13 reprend les taux de reconnaissance du défaut d'alignement. La règle apprise

Méthode	Taux de classification			
	EC	ER 1	ER 2	ER 3
Arbre de décision	93,9%	96,2%	92,4%	92,9%
Règle de classification (conjonctive)	69,3%	68,4%	69,9%	68,4%
Réseau Bayésien (2 parents au plus par noeud)	91,1%	84,9%	84,1%	89,2%
K-NN (1 voisin)	98,2%	86,1%	97,2%	98,6%
RN (MLP avec 1 couche cachée)	97,5%	82,4%	91,0%	93,2%
RN (sans couche cachée)	94,0%	77,9%	89,7%	75,4%

TAB. 4.13 – Résultats de reconnaissance du défaut d'alignement

est : *RMS 30-300 y2* $\leq$ *0,770447* $\Rightarrow$ *A = non*.

4.8.2 Synthèse

Les résultats obtenus avec cette application montrent que les réseaux de neurones sont dans la plupart parmi les meilleurs systèmes de diagnostic. Cependant leur inconvénient reste leur opacité. A côté d'eux les arbres de décision qui, à l'heure actuelle, ne sont presque pas utilisés en diagnostic, présentent des taux de reconnaissance satisfaisants et offrent une possibilité de traduire la décision en règles compréhensibles. Ils peuvent être utilisés dans un processus de définition des seuils de surveillance d'équipements si l'on dispose d'un historique bien documenté. Les plus proches voisins présentent aussi de bonnes capacités de diagnostic. Il faut, cependant, noter que les plus proches voisins sont très influençables par la qualité des indicateurs.

Il nous semble utile, à ce stade, de dire un mot sur les taux de reconnaissance. Ce qui peut être considéré comme taux satisfaisant dépend du domaine d'application, donc de la criticité. En aéronautique, les critères seront bien sûr très contraignants. En établissant un système de diagnostic, on doit se fixer des objectifs en termes de proportion détectable des défauts surveillés [51]. Pour notre part, dans les applications considérés, nous avons estimé qu'une probabilité de 90% de reconnaître un défaut pouvait être considérée comme satisfaisant.

L'étude qui a été menée dans cette section a permis de mettre en évidence et de comparer les capacités de discriminations des indicateurs sélectionnés au moyen de différentes méthodes. A l'examen des taux de classification obtenus avec divers classifieurs, on peut remarquer que la méthode de sélection que nous avons proposée trouve sa place, en diagnostic vibratoire, à côté des méthodes classiques tant les résultats sont comparables. L'utilisation de cette méthode de sélection en amont d'autres classifieurs a semblé donner des résultats acceptables bien qu'à priori il ne puisse être garanti que des indicateurs sélectionnés avec un réseau de neurones restent pertinents pour d'autres classifieurs.

Les faibles nombres de neurones (3 à 5) dans la couche cachée des réseaux de neurones de classification a suggéré de tester des réseaux sans couche cachée. Par rapport aux indicateurs sélectionnés par la méthode proposée dans cette thèse, ces réseaux de neurones sans couche cachée ont accusé une forte dégradation des capacités de reconnaissance de tous les défauts sauf pour le balourd. Appliqués sur l'ensemble complet des indicateurs, ces réseaux de neurones n'ont pas beaucoup détérioré les taux de reconnaissance. Une simple analyse des poids synaptiques, à l'entrée du réseau de neurones sans couche cachée alimenté avec les indicateurs d'origine, a montré la possibilité de sélection des indicateurs. La figure 4.26 illustre, pour le cas du balourd, cette possibilité par ce processus.

L'évaluation des performances d'un système de diagnostic au travers des éléments fournis

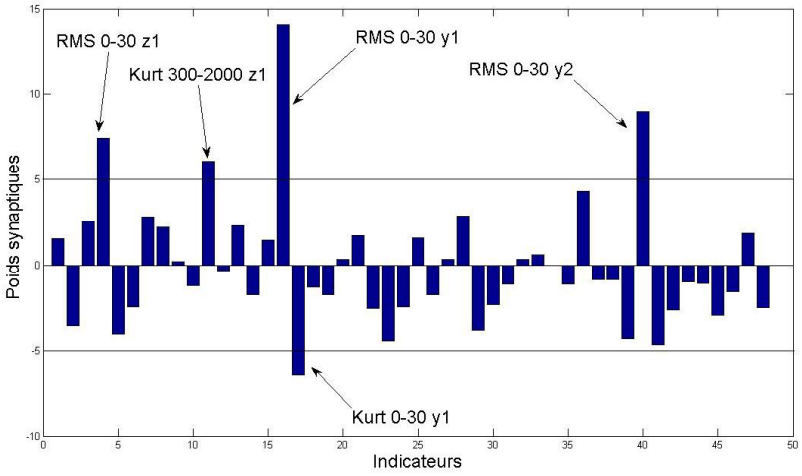


FIG. 4.26 – Poids synaptiques à l'entrée du réseau de neurones sans couche cachée pour le diagnostic du balourd

par la matrice de confusion (taux de classification) suppose que toutes les erreurs de classification présentent le même coût en terme de risque. Or, un faux positif ne présente pas les mêmes risques qu'un faux négatif. Il faut donc tenir compte des coûts d'erreurs de classification, la section suivante en discute.

4.9 Prise en compte des coûts de non détection et de fausses alarmes

Lorsqu'un système de diagnostic est mis en place, il est légitime de se poser des questions sur sa crédibilité et sa fiabilité. En effet, l'émission de fausses alarmes tout comme la non détection d'avaries représentent un coût pour un processus de production industriel. Ce coût comprend différentes composantes dont la perte de production, la main d'œuvre et le matériel (pièce de rechange). Dans le cas du *condition monitoring*, le temps de diagnostic est sensiblement réduit et l'intervention est ciblée à des composants bien identifiés. Par contre dans le cas d'une intervention corrective consécutive à la non détection, en plus du coût lié au temps nécessaire au diagnostic, un coût relatif à d'éventuels dommages collatéraux peut s'ajouter (exemple d'un problème de délignage qui entraîne un balourd et une avarie des paliers).

Une bonne métrique d'évaluation des performances devra tenir compte des coûts de mauvaise classification. Il est alors logique de tenir compte, dans la modélisation du système de

diagnostic, des facteurs de coût pour corriger le taux d'erreur ou de succès attendu du système [78]. Il y a deux approches pour intégrer le coût dans un système de diagnostic : au niveau de la classification ou au niveau de l'apprentissage.

La formulation de la première approche postule qu'une instance aura une prédiction d'une classe qui représente une espérance de coût la plus faible, cette espérance étant calculée sur base de probabilités conditionnelles de chaque classe étant donnée l'instance. Les coûts liés à un problème de classification seront représentés dans une matrice de coûts **E**. L'élément E_{ij} de la matrice de coûts est le coût de prédiction de la classe C_i alors que la classe réelle est C_j. La prédiction optimale pour un vecteur d'indicateurs **x** est donc la classe C_i qui minimalise

$$L(\mathbf{x}, i) = \sum_{j=1}^{m} P(C_j|\mathbf{x}).E_{ij} \tag{4.51}$$

Cette formulation suppose que le processus de classification fournit les probabilités a posteriori $P(C_j|\mathbf{x})$; ce qui est le cas de la plupart des schémas d'apprentissage. Pour les algorithmes qui ne fournissent pas ces probabilités, on utilise la seconde approche qui consiste à pondérer les données d'entrée à l'aide des coûts. La démarche est de modifier les proportions des classes dans l'ensemble d'apprentissage au moyen des duplications. Ainsi, en forçant le système à minimaliser le nombre d'erreurs lors de l'apprentissage, on aboutit à un structure de décision biaisée par rapport à une classe dont on aura pénalisé les erreurs. Les machines à vecteurs de support sont dans ce cas.

Quant à la construction de la matrice de coûts, une analyse AMDEC peut fournir une échelle de criticité des défauts sous étude. A partir de cette échelle, on peut définir les coûts de mauvaise classification comme étant proportionnelle à la distance entre la criticité du défaut réel et celle du défaut prédit. Si la prédiction donne un défaut de criticité élevée par rapport à la réalité, la distance est négative. Dans une telle situation, on prend l'opposé de la valeur mais avec un coefficient de réduction $m < 1$. Dans la référence [79], les auteurs ont proposé de normaliser les distances par la somme des criticités R, puis de multiplier le résultat par un facteur d'échelle S pour obtenir le coût :

$$C_{ij} = \begin{cases} \frac{\mathrm{d_{ij}}}{\sum_i R}.S, & \text{pour } \mathrm{d_{ij}} > 0 \\ -m.\frac{\mathrm{d_{ij}}}{\sum_i R}.S, & \text{pour } \mathrm{d_{ij}} \leq 0 \end{cases} \tag{4.52}$$

où d_{ij} représente la distance entre les criticités. La multiplication de la matrice de coûts par un facteur quelconque n'a pas d'influence sur les résultats de la classification.

Construire la matrice de coûts par la voie des criticités n'est pas la seule façon de procéder, on peut procéder à une affectation d'indices de coût sur base d'une analyse des conséquences des défauts. Considérons pour illustration les résultats obtenus sur ces données du banc d'essai pour la reconnaissance du balourd. Prenons une approche **MoD-MuC** avec un arbre de décision pour comparer les résultats avec ceux obtenus sans prise en compte du coût.

Les matrices de coûts de chaque classifieurs sont construites en considérant les indices de coût suivants :

- meilleur diagnostic : indice 0 ;
- fausse alarme alors que le défaut est absent : 1 ;
- bonne prédiction d'un mode de défaut avec erreur de localisation (exemple balourd de même niveau détecté sur un disque alors qu'il se trouve sur un autre) : 1 ;
- prédiction du défaut sur un mode à niveau de sévérité élevé par rapport à la réalité : 0,5 ;
- non détection : 3 ;
- prédiction du défaut sur un mode à faible sévérité par rapport à la réalité : 2.

On aboutit à la matrice de coûts reprise dans le tableau 4.14 pour le balourd

↗	B0	B1	B2	B3	B4
B0	0	1	1	1	1
B1	3	0	1	0,5	1
B2	3	2	0	2	0,5
B3	3	0,5	1	0	1
B4	3	2	0,5	2	0

TAB. 4.14 – Matrice de coût du défaut de balourd

Les résultats de reconnaissance obtenus avec l'ensemble des indicateurs sont représentés dans le tableau 4.15

classifié comme	⇒	B_0	B_1	B_2	B_3	B_4	Total		*PI*
	B_0	140	1	0	3	0	144	→	97,22%
	B_1	2	108	9	11	14	144	→	75,00%
	B_2	0	5	70	1	20	96	→	72,92%
	B_3	4	15	3	108	2	132	→	81,82%
	B_4	0	7	15	5	105	132	→	79,55%
Total		146	136	97	128	141	648		

TAB. 4.15 – Matrice de confusion et probabilités d'identification avec prise en compte des coûts

Le taux global de reconnaissance est de 81,94% ($\frac{(140+108+70+108+105)}{648}$). Le tableau 4.16 compare les résultats obtenus avec et sans considération du coût. On remarque que les probabilités d'identification et le taux de classification sont plus ou moins dégradées dans ce cas (sauf pour B_0). Notons qu'une autre affectation des coûts modifiera les probabilités d'identification.

PI Classe de défaut	Sans coûts	Avec coûts
$PI(B_0)$	94,44	97,22
$PI(B_1)$	81,25	75,00
$PI(B_2)$	75,00	72,92
$PI(B_3)$	81,82	81,82
$PI(B_4)$	80,30	79,55
TC	83,18	81,94

TAB. 4.16 – Comparaison des résultats obtenus avec et sans prise en compte des coûts

Synthèse

La prise en compte de coûts d'erreur est une approche qui permet de construire un système de diagnostic dont la fiabilité est plus ou moins améliorée que si l'on tient pas en compte les risque de fausses alarmes et de non-identification. Généralement, à paramètres identiques, les classifieurs entraînés avec la notion de coût aboutissent à des taux faibles à l'entraînement. Il faut donc ajuster les paramètres pour améliorer le taux de succès à l'apprentissage et s'assurer ainsi une plus grande fiabilité en exploitation.

4.10 *Dynamicité* des données et formulation générale des problèmes de *condition monitoring*

La majorité des travaux identifiés dans la bibliographie relative à l'utilisation des techniques de *data mining* en diagnostic ont en commun d'avoir abordé le problème sous l'angle statique, c'est-à-dire sans considérer l'évolution des données avec le temps. Intégrer la mesure dans son contexte de dépendance temporelle vis-à-vis des autres mesures qui la précédent, peut améliorer la valeur informationnelle de la mesure considérée.

Avant d'exposer quelques méthodes utilisables pour la construction des indicateurs qui capturent la dépendance temporelle, nous nous allons d'abord présenter la structuration de données vibratoires de condition monitoring pour leur exploitation avec les outils de *data mining*.

4.10.1 Structuration des données. Cube de données de surveillance vibratoire

Le stockage de données orientées décision suscite la problématique de leur structuration en vue d'une exploration aisée. Il est certes vrai que dans le but de la maintenance conditionnelle, une historisation des données est indispensable. Conserver les données passées et les exploiter adéquatement permet une meilleure appréhension du présent et une anticipation du futur.

Nous avons montré dans le chapitre précèdent qu'à la suite du processus d'extraction des indicateurs sur les signaux de mesure, l'historisation aboutit à un série temporelle multivariée pour un équipement donné.

Soit $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ un vecteur de variable d'indicateurs extraits du signal relevé à la date t, éventuellement complétés d'autres informations opératoires ou logistiques telles que la charge, la vitesse, etc. Une série de ces mesures effectuées durant la vie de la machine peut être représentée comme une matrice $\mathbf{X} = \left[\mathbf{x}^T(t_1), \mathbf{x}^T(t_2), \ldots, \mathbf{x}^T(t_p)\right]$ d'ordre $n \times p$, p étant le nombre de mesures relevées. La matrice ainsi construite représente la série temporelle multivariée, chaque colonne étant un vecteur d'indicateurs calculés pour une date donnée. Les p vecteurs ne sont pas nécessairement également espacés dans le temps.

Considérons maintenant la surveillance de m machines semblables supposées opérant dans des conditions similaires et ayant des modes de défaillances similaires. On construit pour chaque machine k une matrice de conditions $\mathbf{X}_k$. Il peut se faire que chaque machine n'eût pas fait l'objet du même nombre de prises de mesure. Pour traiter ces données manquantes, nous nous sommes proposé l'utilisation de la notion de barycentre. Supposons, en effet, pour un équipement k que l'on n'ait pas relevé la mesure à la date t_i. Pour ramener toutes les matrices $\mathbf{X}$ à la même dimension, on construira le vecteur manquant à l'équipement k par l'opération suivante : $\mathbf{x}_k(t_i) = \mathbf{x}_k(t_{i-1})(t_{i+1} - t_i) + \mathbf{x}_k(t_{i+1})(t_i - t_{i-1})$. Ceci revient à supposer une évolution linéaire des indicateurs entre les deux instants. Dans cette logique, si au lieu d'une seule mesure mais plusieurs mesures manquent entre deux dates, on peut les reconstruire en suivant toujours le schéma linéaire pour autant que le nombre de mesures absentes ne dépasse 4, valeur au-delà de laquelle un rejet pur et simple de l'équipement est proposé pour l'étude.

Au final, on aboutit à un cube de données à 3 dimensions dans ce cas, chaque dimension correspondant à un axe (figure 4.27) :

– un axe de temps ou axe dynamique, référencé par i : on dispose de p valeurs temporelles enregistrées à différentes dates ;

– le second est l'axe des paramètres ou des variables (attributs), référencé par j : pour chaque équipement on mesure n paramètres pour suivre l'état du matériel ;
– et suivant le troisième axe, on retrouve m machines semblables pour lesquelles n variables ont été mesurées à p dates différentes. C'est l'axe des objets, il est référencé par k.

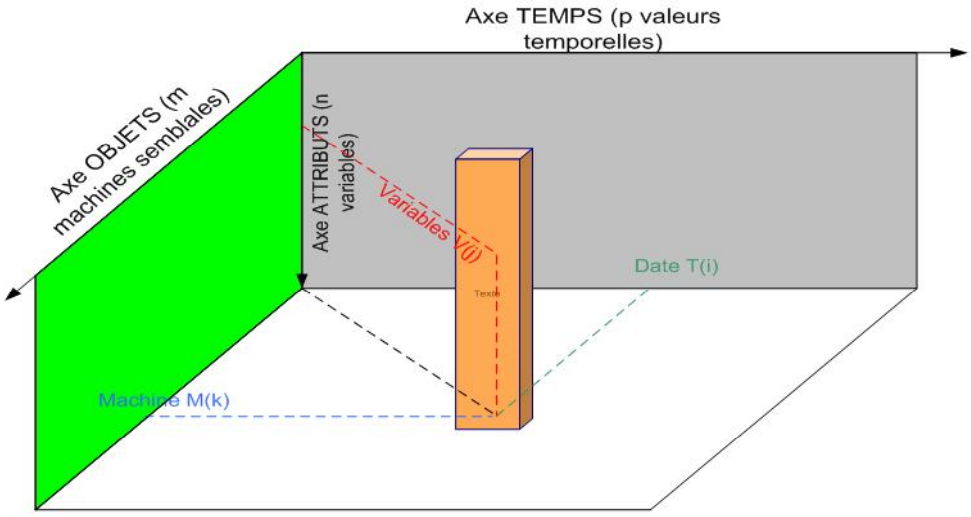


FIG. 4.27 – Représentation et structuration de données de surveillance

En d'autres termes, on a m séries temporelles de p vecteurs de dimension n qu'il faudra traiter pour une bonne extraction d'informations. Les applications des techniques de *machine learning* en identification de défauts n'ont en général concerné que des données se trouvant dans le plan objets-attributs, c'est-à-dire des données relevées sur plusieurs machines semblables, presque à une même date, ou des données provenant de plusieurs essais réalisés sur une même machine sans prise en compte de l'évolution dans le temps de l'état de la machine. Or, dans la pratique il n'y a que très peu de problèmes qui se posent en termes statiques. D'où toute l'importance de prise en compte de la dynamicité des données de surveillance le besoin de l'application des méthodes de *data mining*.

Nous allons tout d'abord procéder à une classification des problèmes rencontrées en condition monitoring en fonction de la disposition des données par rapport aux trois axes ci-haut définis.

4.10.2 Classification des problèmes de surveillance en vue du *data mining*

Les problèmes seront catégorisés en problèmes dynamiques et problèmes statiques, ce qui nous permettra de distinguer trois classes distinctes de problèmes.

4.10.3 Les problèmes statiques

Cette catégorie de problèmes est de loin la plus rencontrée en *data mining*. La plupart de méthodes et techniques de *data mining* ont été développées pour ce type de problèmes.

Problèmes de classe 1

Cette classe représente les problèmes idéaux de laboratoire. Les indicateurs sont extraits soit des mesures prises sur plusieurs machines semblables à la même date, soit de plusieurs mesures relevées sur une même machine dans des conditions plus ou moins précises sans considération de la dépendance temporelle. Ces problèmes qui se prêtent très facilement à l'application du *data mining* (classification ou règles associatives), seront représentés par une tranche du cube parallèle au plan objets-attributs.

Problèmes de classe 2

Le traitement de cette classe de problèmes va situer la discussion au niveau des signaux bruts relevés sur les équipements en les considérant dans leur représentation continue dans le domaine temporel (figure 4.28). Il s'agit d'analyser des séries temporelles ou signaux relevées à une date donnée sur plusieurs machines semblables. Il peut tout aussi bien s'agir de signaux relevés sur une même machine à des dates successives différentes, mais sans prise en compte de l'évolution de la machine en fonction du temps. Les données relatives à cette classe de problèmes sont situées dans le plan temps-objets, mais la dépendance par rapport à l'axe dynamique sera supprimée par une transformation du fait de la durée relativement très courte (quelques secondes) d'un signal temporel par rapport aux périodes de surveillance d'usage en condition monitoring.

Dans leur traitement, les problèmes de cette classe se rabattent immédiatement à la classe 1. Chaque signal est représenté, après un processus de transformation, par un vecteur d'indicateurs. L'extraction qui se fait dans le domaine temporel ou dans le domaine fréquentiel, peut être suivie au besoin d'un processus de compression ou réduction dimensionnelle. Diverses techniques telles que l'analyse en composantes principales sont utilisées. Les problèmes de surveillance vibratoire se place dans cette catégorie avant de retomber dans la classe 1 après le processus d'extraction de données.

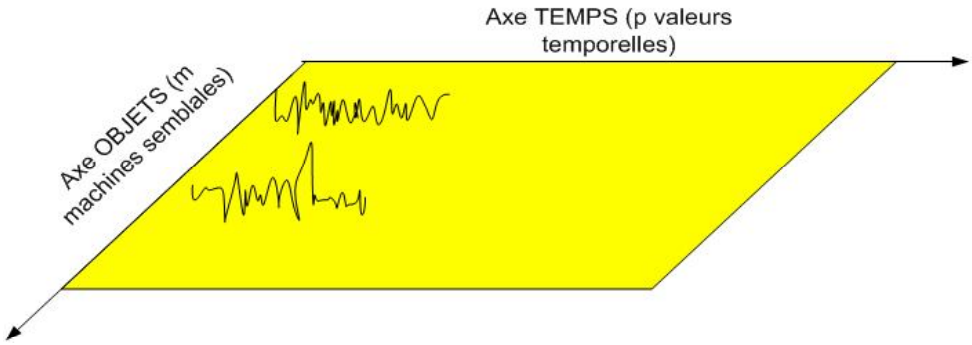


FIG. 4.28 – Problème de Classe 2

4.10.4 Les problèmes dynamiques

Ces problèmes constituent la **classe 3**. Dans les problèmes dits « dynamiques », l'évolution temporelle de données ne peut être négligée. Nous nous situons dans un cas général du cube de données avec un axe dynamique. Les objets à analyser ne peuvent plus être considérés de manière individuelle et indépendante. Chaque objet dépend tout au moins de l'objet et de l'action de maintenance qui le précèdent immédiatement sur l'axe de temps. De façon générale, on peut supposer que tout vecteur présente une dépendance, vis-à-vis de r vecteurs précédents, de la forme

$$\mathbf{x}_k(t_i) = f\left\{\mathbf{x}_k(t_{i-1}), \mathbf{x}_k(t_{i-2}), \ldots, \mathbf{x}_k(t_{i-r})\right\}$$

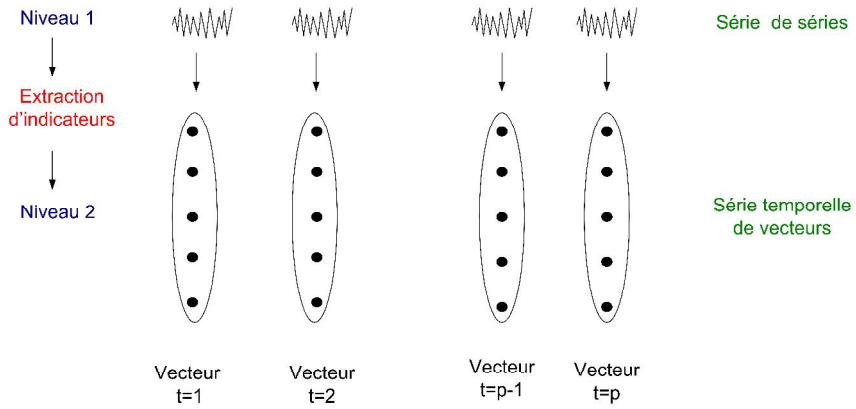


FIG. 4.29 – Représentation d'une série temporelle de vecteurs d'indicateurs

La prise en compte de cette dépendance temporelle permet une compréhension fiable de la mesure et de sa signification à travers les notions de croissance et de taux de croissance. Dans

la modélisation de ces données en vue du *data mining*, il s'avérera un besoin réel d'élaborer un traitement qui rende possible la capture de cette dépendance temporelle par de nouvelles variables qui, elles, doivent être indépendantes afin d'utiliser les techniques traditionnelles de *data mining* qui ont été développées pour des données statiques. Ces nouvelles variables pourront provenir d'une projection des données dans un nouvel espace dont la dimension peut être plus élevée que celle de l'espace de départ. Nous proposons deux moyens pour la construction des indicateurs qui capturent la dépendance temporelle.

4.10.5 Méthode de transformation par dérivation

Afin de permettre la représentation des données d'un cube de données dynamiques, relatives à une famille d'équipements, sous un format de fichiers plats pour une exploitation par *data mining*, nous proposons une approche qui utilise les taux de croissance des indicateurs. Cette procédure construit de nouvelles variables indépendantes du temps et capables de capturer la dépendance temporelle présente dans les variables de départ. Par exemple, on peut utiliser deux pentes successives et le niveau actuel d'un indicateur. Si $\mathbf{x}(t_i)$ est le vecteur d'indicateurs calculés sur le signal enregistré à la date t_i, les nouvelles variables déduites de l'indicateur $x_j(t_i)$ par le calcul des pentes seront

$$y_{j1} = \frac{x_j(t_{j-1}) - x_j(t_{j-2})}{t_{j-1} - t_{j-2}} \tag{4.53}$$

$$y_{j2} = \frac{x_j(t_j) - x_j(t_{j-1})}{t_j - t_{j-1}} \tag{4.54}$$

$$y_{j3} = x_j \tag{4.55}$$

Ainsi, un indicateur donne naissance à trois autres. Le nombre d'indicateurs est triplé par cette transformation.

4.10.6 Méthode de transformation par moyenne mobile

Cette procédure de transformation permet la construction des indicateurs qui sont des moyennes des L indicateurs calculés à L dates qui précèdent la date considérée. On peut choisir de pondérer la moyenne pour atténuer la contribution des données très éloignées du point courant. On

parle d'une moyenne mobile d'ordre L. Pour un vecteur d'indicateurs $\mathbf{x}(t_i)$, on aura

$$\mathbf{y}(t_i) = \frac{1}{L} \sum_{l=0}^{L} b_l \mathbf{x}(t_i - l) \tag{4.56}$$

où b_l représente le poids. La pondération peut être exponentielle, par exemple.

On peut partir la notion de moyenne mobile pour définir de nouvelles variables à partir de la moyenne courante d'un indicateur. La moyenne courante $\mathbf{y}(t_i)$ d'un vecteur d'indicateurs $\mathbf{x}$ à un temps t_i est calculée de la manière suivante :

$$\mathbf{y}(t_i) = \frac{1}{i} \sum_{j=1}^{i} \mathbf{x}(t_j) \tag{4.57}$$

4.11 Conclusion

Ce chapitre avait pour objectif de présenter les méthodes de *data mining* que l'on peut utiliser en diagnostic. L'importance de la sélection d'indicateurs a été soulignée pour l'allégement de l'apprentissage comme pour l'amélioration des capacités de généralisation.

Différentes méthodes de *data mining* ont été appliquées à une étude expérimentale, et à travers cette étude nous avons montré l'efficacité de la procédure de sélection d'indicateurs que nous avons développée en la comparant à des méthodes classiques dont l'analyse discriminante et les arbres de décision. La prise en compte des coûts dans la classification a été introduite et appliquée à un classifieur dont les performances ont été comparées à celles obtenus sans considération des coûts.

Les capacités d'identification auxquelles ont abouti ces applications permettent de justifier l'utilisation des méthodes de *data mining* pour le besoin de diagnostic. En s'appuyant sur la définition des axes dans la représentation du cube de données, des classes de problème de diagnostic vibratoire ont été identifiées en fonction de la structuration des données.

En surveillance vibratoire, le besoin d'une capacité de détection précoce des avaries mécaniques n'est pas à démontrer. Le chapitre suivant adresse le problème de détection précoce de dégradation d'un élément de machine dont les défaillances représentent une grande part de causes d'arrêt des machines : il s'agit des roulements.

Références

[1] U. Fayyad, G. Piatetsky-Shapiro et P. Smyth : From data mining to knowledge discovery in databases. *American Association of Artificial Intelligence*, pages 37–54, 1996.

[2] D. A. Zighed et R.Rakotomalala : Extraction de connaissances à partir de données (ecd). *In Traité Informatique (H3744)*. Techniques de l'Ingénieur.

[3] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer et R. Wirth : Crisp-dm 1.0 step-by-step data mining guide. Rapport technique, SPSS, 2000. URL www.crisp-dm.org.

[4] D. Hand, H. Manilla et P. Smyth : *Principles of Data Mining*. MIT Press, 2001.

[5] F. Moerchen : *Time Series Knowledge Mining*. Thèse de doctorat, Philipps-Universitat Marburg, 2006.

[6] H.A. Simon : How should machine learn. *In* R. S. Michalski, J. G. Carbonell et T. M. Mitchell, éditeurs : *Machine Learning : An Artificial Intelligence Approach*, volume I. Tioga, Palo Alto, CA,., 1983.

[7] Zhengxin Chen : *Computational intelligence for decision support*. CRC Press LLC, New York, 2000.

[8] V. Sugumaran et K.I. Ramachandran : Automatic rule learning using decision tree for fuzzy classifier in fault diagnosis of roller bearing. *Mechanical Systems and Signal Processing*, 21(5):2237–2247, 2007.

[9] Yubao Chen : Impending failure detection for a discrete process. *Mechanical Systems and Signal Processing*, 7(2):121–132, 1993.

[10] Ross Quinlan : *C4.5 : Programs for Machine Learning*. Morgan Kaufmann Publishers, San Mateo, CA, 1993.

[11] Ian H.Witten et Eibe Frank : *Data mining : Practical machine learning tools and techniques*. Morgan Kaufmann, San Francisco, 2nd édition, 2005.

[12] H. Liang et Y. Yan : Learning naive bayes for conditional probability estimation. *In Proceedings of the Canadian Al-2006 Conference*, 2006.

[13] Eduardo Gilabert et Aitor Arnaiz : Intelligent automation systems for predictive maintenance : A case study. *Robotics and Computer-Integrated Manufacturing*, 22:543–549, 2006.

[14] G.Weidl, A.L. Madsen et S. Israelson : Application of object-oriented bayesian networks for condition monitoring,root cause analysis and decision support on operation of complex continuousprocesses. *Computers and Chemical Engineering*, 29:1996–2009, 2005.

[15] H. Demuth, M. Beale et M. Hagan : *Neural Network Toolbox for use with Matlab.* The MathWorks, Natick, USA, 2007.

[16] Rolf Iserman et P. Ballé : Trends in the application of model-based fault detection and diagnosis in technical process. *In 13th IFAC World Congrss, volume N, pages 1-12, San Francisco, CA, USA*, 1996.

[17] D.Chang-Ching Lin et B.Hsu-Pin Wang : Performance analysis of rotating machinery using enhanced cerebellar model articulation (e-cemac) neural networks. *Computers Ind.Eng.*, 30(2):227–242, 1996.

[18] A.C. McCormick et A.K. Nandi : Classification of the rotating machine condition using artificial neural networks. *In Proceedings of IMeschE :PartC*, volume 211(6), pages 439–450, 1999.

[19] Erkki Jantunen : *Indirect multisignal monitoring and diagnosis of drill wear.* Thèse de doctorat, Helsinki University of technology, Janvier 2006.

[20] M.A. Essawy, S. Diwakar, S. Zein-Sabato et M. Bodruzzaman : Helicopter transmission fault diagnostics using neuro-fuzzy techniques. *Intelligent Engineering Systems Through Artificial Neural Networks*, 7:661–666, 1997.

[21] F. Filippetti, G. Franceschini et C. Tassoni : Neural networks aided on-line diagnostics of induction motor rotor faults. *IEEE Transactions on Industry Applications*, 31 : 4:892–899, 1995.

[22] M. Minoux : *Mathematical Programming : Theory and Algorithms.* John Wiley and Sons, 1986.

[23] N Cristianini et J. Shawe-Taylor : *An Introduction to Support Vector Machines.* Cambridge University Press, 2000.

[24] F. Girosi : An equivalence between sparse approximation and support vector machines. *A.I. Memo, MIT Artificial Intelligence Laboratory*, 1997.

[25] L. B. Jack et A. K. Nandi : Fault detection using support vector machines and artificial neural networks, augmented by genetic algorithms. *Mechanical Systems and Signal Processing*, 16 :2-3:373–390, 2002.

[26] B Samanta : Gear fault detection using artificial neural networks and support vector machines with genetic algorithms. *Mechanical Systems and Signal Processing*, 18 :3:625–644, 2004.

[27] Dongfeng Shi et Nabil N. Gindy : Tool wear predictive model based on least squares support vector machines. *Mechanical Systems and Signal Processing*, doi :10.1016/j.ymssp.2006.07.016, 2006.

[28] Ming GE, R. DU, Guicai ZHANG et Yangsheng XU : Fault diagnosis using support vector machine with an application in sheet metal stamping operations. *Mechanical Systems and Signal Processing*, 18(1):143–159, 2004.

[29] J. SUN, M. RAHMAN, Y. S. WONG et G. S. HONG : Multiclassification of tool wear with support vector machine by manufacturing loss consideration. *International Journal of Machine Tools and Manufacture*, 44(11):1179–1187, 2004.

[30] Alfonso Rojas andAsoke K. NANDI : Practical scheme for fast detection and classification of rolling-element bearing faults using support vector machines. *Mechanical Systems and Signal Processing*, 20(7):1523–1536, 2006.

[31] Achmad WIDODO et Bo-Suk YANG : Support vector machine in machine condition monitoring and fault diagnosis. *Mechanical Systems and Signal Processing*, 21(6):2560–2574, 2007.

[32] L.A. ZADEH : Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[33] L.A. ZADEH : Outline of a new approach to the analysis of complex systems and decision-processes. *IEEE Transactions on Systems, Man, and Cybernetics*, 3(1):28–44, 1973.

[34] L.A. ZADEH : The concept of a linguistic variable and its application to approximatereasoning, parts 1, 2, and 3. *Information Sciences*, pages 8 :199–249, 8 :301–357, 9 :43–80, 1975.

[35] L.A. ZADEH : Fuzzy logic. *Computer*, 1(4):83–93, 1988.

[36] L.A. ZADEH : Knowledge representation in fuzzy logic. *IEEE Transactions on Knowledge and Data Engineering*, 1:89–100, 1989.

[37] James J. BUCKLEY et Yoichi HAYASHI : Fuzzy neural networks : A survey. *Fuzzy Sets and Systems*, 66:1–13, 1994.

[38] Shigeo ABE et Ming-Shong LAN : A method for fuzzy rules extraction directly from numerical data and its application to pattern classification. *IEEE Trans. Fuzzy Systems*, 3 (1):18–28, février 1995.

[39] Javadpour R. & KNAPP : A fuzzy neural network approach to machine condition monitoring. *Computers and Industrial Engineering*, 45:323–330, 2003.

[40] H. ISHIBUCHI, H. OKADA et H. TANAKA : Interpolation of fuzzy of fuzzy if-then rules by neural networks. *In Proc. 2nd Int. Conf. on Fuzzy Logic and Neural Networks (IIZUKA'92)*, pages 337–340, Iizuka, 1992.

[41] M. SUGENO : *Industrial applications of fuzzy control.* Elsevier Science Pub.Co., 1985.

[42] A. BLUM et T. MITCHELL : Combining labeled and unlabeled data with co-training. *In Proceedings of the 11th Annual Conference on Computational Learning Theory*, 1998.

[43] J.P. Marquès de Sá : *Applied Statistics using SPSS, STATISTICA and MATLAB*. Springer, Berlin, 2003.

[44] I. Guyon et A. Elisseeff : An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3:1157–1182, 2003.

[45] M. Kudo et J. Sklansky : Comparison of algorithms that select features for pattern classifiers. *Pattern Recognition*, 33(1):25–41, 2000.

[46] M. A. Hall et G. Holmes : Benchmarking attribute selection techniques for discrete class data mining. *IEEE Transactions on knowledge and data engineering*, 15(3), 2003.

[47] E. Zio, P. Baraldi et G. Gola : Feature-based classifier ensembles for diagnosing multiple faults in rotating machinery. *Applied Soft Computing*, doi :10.1016/j.asoc.2007.10.005, 2007.

[48] J. Yang et V. Honavar : Feature subset selection using a genetic algorithm. *IEEE Intell. Syst. Appl.*, 13(2):44–49, 1998.

[49] L.B. Jack et A.K. Nandi : Genetic algorithms for feature selection in machine condition-monitoring with vibration signals. *IEE Proceedings of Vision, Image and Signal Processing*, 147(3):205–212, 2000.

[50] Mark A. Hall : *Correlation-based feature selection for machine learning*. Thèse de doctorat, University of Waikato, 1999.

[51] G. Vachtsevanos, F.L. Lewis, M. Roemer, A. Hess et B. Wu : *Intelligent fault diagnosis and prognosis for engineering systems*. John Wiley & Sons, 2006.

[52] E. Diday, J. Lemaire, J. Pougiet et F. Testu : *Eléments d'analyse de données*. Dunod, 1982.

[53] W.J Krzanowski : Selection of variables to preserve multivariate data structure using principal component analysis. *Applied Statistics- Journal of the Royal Statistical Society Series C*, 36:22–33, 1987.

[54] W.J Krzanowski : A stopping rule for structure- preserving variable selection. *Statistics and Computing*, 6:51–56, 1996.

[55] I. Joliffe : *Principal component analysis*. Springer-Verlag, New York, 1991.

[56] I. Cohen, Q. Tian, X. S. Zhou et T. S. Huang : Feature selection using principal feature analysis. *Proceedings of the 15th international conference on Multimedia*, pages 301–304, 2007.

[57] Cheong Hee Park et Moonhwi Lee : On applying linear discriminant analysis for multi-labeled problems. *Pattern Recognition Letters*, 29:878–887, 2008.

[58] D.J. Bertholds, M. & Hand, éditeur. *Intelligent Data Analysis*. Springer, 2003.

[59] R. Avila, D. Pat Banerjee He et E. Bechhoefer : Discriminant analysis for helicopter rotor vibration prognostics. *IEEE Aerospace Conference*, page 14, 2006.

[60] Christopher Brooks, Karl Iagnemma et Steven Dubowsky : Vibration-based terrain analysis for mobile robots. *In Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, pages 3426–3431, Barcelona, Spain, April 2005.

[61] Weihua Li, Kang Ding, Tielin Shi, et Guanglan Liao : Gear fault classification using kernel discriminant analysis. *Key Engineering Materials*, 321-323:1556–1559, 2006.

[62] Li Weihua, Shi Tie-lin et Yang Shu-zi : An approach for mechanical fault classification based on generalized discriminant analysis. *Frontiers of Mechanical Engineering in China*, 1(3):292–298, 2006.

[63] Hongwen Zheng et Yanxia Zhang : Feature selection for high-dimensional data in astronomy. *Advances in Space Research*, 41:1960–1964, 2008.

[64] Feng Tang et Hai Tao : Fast linear discriminant analysis using binary bases. *Pattern Recognition Letters*, 28:2209–2218, 2007.

[65] C. Pizarro et *al.* : Vinegar classification based on feature extraction and selection from headspace solid-phase microextraction/gas chromatography volatile analyses : A feasibility study. *analytica chimica acta*, 608:38–47, 2008.

[66] Ron Kohavi et George H. John : Wrappers for feature subset selection. *Artificial Intelligence*, 97:273–324, 1997.

[67] C. Cardie : Using decision trees to improve case-based learning. *Proceedings of the Tenth National Conference on Machine Learning*, pages 25–32, 1993.

[68] M. Hall : A decision tree-based attribute weighting filter for naive bayes. *Knowledge-Based Systems*, 20:120–126, 2007.

[69] V. Sugumaran, V. Muralidharan et K.I. Ramachandran : Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing. *Mechanical Systems and Signal Processing*, 21:930–942, 2007.

[70] P. Leray et P. Gallinari : Feature selection with neural networks. *Behaviormetrika*, 26, January 1999. URL citeseer.ist.psu.edu/leray98feature.html.

[71] Keming Wang : Neural network approach to vibration feature selection and multiple fault detection for mechanical systems. *First International Conference on Innovative Computing, Information and Control*, 3:431–434, 2006.

[72] T. Cibas, F. Soulié, P. Gallinari et S. Raudys : Variable selection with neural networks. *Neurocomputing*, 12:223–248, 1996.

[73] T. Matsuura : An application of neural network for selecting feature parameters in machinery diagnosis. *Journal of Materials Processing Technology*, 157-158:203–207, 2004.

[74] Bovic Kilundu et Pierre Dehombreux : A feature selection approach based on principal component analysis and synaptic weights for mechanical faults diagnosis. *In Advances in Safety and Reliability*, volume 1, pages 1041–1046. ESREL, Taylor & Francis Group, 2005.

[75] E Zio, P Baraldi et G Gola : Ensemble feature selection for diagnosing multiple faults in rotating machinery. *Proc IMechE, Part O : Journal of Risk and Reliability*, 221(1):29–41, 2008.

[76] Enrico Zio et Giulio Gola : Genetic algorithm and fknn classification applied to multiple fault diagnosis. 2005.

[77] R. Kohavi : A study of cross-validation and bootstrap for accuracy estimation and model selection. *In Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, pages 1137–1143. Morgan Kaufmann, 1995.

[78] Charles Elkan : The foundations of cost-sensitive learning. *In IJCAI*, pages 973–978, 2001. URL citeseer.ist.psu.edu/elkan01foundations.html.

[79] M.J. Ashby et W.J. Scheuren : Intelligent maintenance advisor for turbine engines. *Proceedings of 2000 IEEE Aerospace Conference*, 6:211–219, 2000.

Chapitre 5

Détection précoce et identification de défauts des roulements

Dans ce chapitre, nous nous intéressons à la question de la détection précoce des avaries des roulements. Nous y proposons une procédure de construction des variables de détection à l'aide d'un modèle ACP qui génère un résidu témoignant de l'écart par rapport au fonctionnement normal. Les signaux sont préalablement traités pour l'extraction d'un signal d'enveloppe. Trois méthodes de prétraitement sont comparées : l'ASS, l'analyse en ondelettes et le simple filtrage passe-bande.

5.1 Introduction

La dégradation des paliers peut être très dommageable dans certaines situations. Son caractère progressif pose la question de la détermination du moment propice où il faut procéder à leur remplacement, au prix d'un arrêt des machines. Si l'on sait détecter, dans leur phase initiale, les avaries d'un roulement et identifier toutes les phases de leur évolution, on peut estimer la courbe de fiabilité du roulement, estimer à chaque instant sa durée de vie résiduelle, réaliser un pronostic et répondre à la préoccupation évoquée ci-dessus dans le cadre d'une maintenance optimalisée. Ce processus d'estimation de durée de vie sera basée sur le *condition monitoring*, en parallèle ou en complément des méthodes fiabilistes [1].

Dans le cadre de ce travail, nous avons conduit avec la précieuse collaboration de l'URCA*, une étude expérimentale utilisant les mesures vibratoires pour surveiller le défaut de la bague in-

*Université de Reims Champagne Ardenne

térieure d'un roulement à rouleaux cylindriques. L'étude vise à mettre en évidence la possibilité de détection précoce et d'identification de différentes sévérités du défaut.

La justification de l'utilisation des vibrations repose dans le processus même de dégradation du roulement. La fatigue démarre avec l'initiation de microfissures dans le matériau de la bague. Au cours du service, celles-ci se propagent pour, éventuellement, atteindre la surface de la bague en en délogeant des particules de métal [2]. Les piqûres laissées vont être le siège d'impacts lors du passage des éléments roulants causant ainsi l'excitation de fréquences propres [3–5].

Selon les standards industriels, la taille fatale d'une piqûre est fixée à 6,25 mm^2 (0,01 in^2) [1, 6, 7]. Mais, comme nous l'établirons au cours de cette étude, cette taille reste de loin très grande par rapport aux défauts détectables des paliers à roulements. Lorsqu'un défaut de la taille fatale est détecté, l'arrêt doit se produire en situation d'urgence, ce qui risque d'entraîner des perturbations coûteuses. Cela justifie l'importance de pouvoir prédire le taux d'accroissement des défauts et la vie résiduelle des roulements dans une approche pronostic en plus de l'approche diagnostic. L'approche pronostic ne sera pas adressée dans cette étude, elle nécessite la détemination d'un ou de plusieurs indicateurs liés au développement du *pitting* et, ensuite, un ensemble d'essais « *Run-to-Failure* » conduits dans des conditions variées représentatives des sollicitations réelles.

5.2 Manifestations vibratoires de défauts de roulements

En dépit des nombreux efforts qui sont consentis aussi bien par le fabricant que par l'utilisateur, la durée de vie effective d'un roulement à billes ou à rouleaux ne peut être prédite qu'avec une incertitude relativement importante due aux approximations dans la caractérisation des efforts à transmettre et à l'estimation des conditions de lubrification. Une politique de maintenance, systématique ou préventive conditionnelle, permet d'éviter leur destruction, toujours dommageable pour la machine [8].

Certaines études ont conduit à l'explication du mécanisme de la génération de vibrations et de bruit dans les roulements. Les roulements agissent comme source de vibration ou de bruit soit du fait de leur raideur (compliance) variable, soit de par la présence de défauts. Les roulements à billes avec des charges radiales produisent des vibrations, même s'ils sont parfaits géométriquement. Ceci est dû à l'utilisation d'un nombre limité d'éléments roulants pour supporter les charges. Le nombre d'éléments roulants et leur position dans la zone de charge varient selon la rotation du palier, ce qui conduit à une variation périodique de la raideur totale

du palier. Cette variation de raideur produit des vibrations qui sont connues sous le nom d'excitations sous raideur variable. Le changement de direction des forces de contact appliquées par les éléments roulants peut provoquer des vibrations de flexion ou d'ovalisation des pistes de roulement, même si celles-ci sont géométriquement parfaites .

La présence de défauts provoque une augmentation significative du niveau de vibration. On peut distinguer parmi les défauts de palier :

- les défauts distribués ou répartis qui comprennent la rugosité , l'ondulation de la surface, les désalignements des pistes, et les variations de la taille des éléments roulants. Les caractéristiques des surfaces sont considérées en fonction de leur longueur d'onde comparée avec la largeur du contact de Hertz des éléments roulants avec leurs pistes.
  Les défauts répartis sont produits par les erreurs de fabrication, par une installation incorrecte, ou par usure abrasive. La variation de la force de contact entre les éléments roulants et leurs supports, due aux défauts répartis, résulte en une augmentation du niveau de vibration. L'étude de la réponse vibratoire due à cette catégorie de défauts est très importante en vue de l'inspection de qualité et de la surveillance de machine.
- les défauts locaux incluent les fissures, les piqûres et les écaillages dans les surfaces de roulement. Le mode dominant de la défaillance de roulements est l'écaillage des bagues ou des éléments roulants. Une fissure de fatigue est initiée en dessous de la surface et se propage vers la surface jusqu'à la libération d' un morceau de métal et à l'apparition d'une piqûre ou d'un écaillage. La dégradation par fatigue peut être accélérée par surcharge ou par choc pendant le fonctionnement ou l'installation. Quand un défaut local sur un élément agit avec un autre élément voisin, le changement brusque des contraintes de contact à l'interface produit une impulsion de très courte durée. Cette impulsion provoque la vibration et du bruit qui peuvent être surveillés.

La dégradation des roulements se traduit généralement par une détérioration de leur résistance à la fatigue superficielle : des défauts de surface (écaillages, fissures, ...) apparaissent sur les éléments roulants et les bagues intérieures et extérieures. Ils vont engendrer, lors de la rotation, des chocs périodiques dont la fréquence d'apparition peut être prédite à partir d'une analyse cinématique faisant intervenir les vitesses de rotation des bagues interne et externe, le nombre d'éléments roulants, et les grandeurs géométriques (diamètres, angle de contact) caractéristiques.

Ainsi, pour un roulement à rouleaux à contact oblique, tel que représenté sur la figure 5.1, on peut définir les fréquences caractéristiques suivantes, à partir des fréquences de rotation f_i et

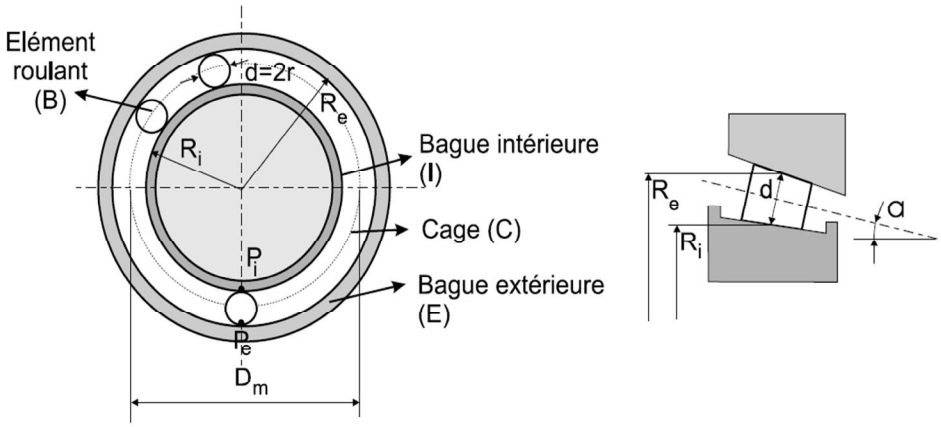


FIG. 5.1 – Eléments d'un roulement à rouleaux coniques.

f_e des bagues intérieure et extérieure, respectivement :

– de rotation de la cage, de défaut de cage :

$$f_c = f_i \frac{1}{2}\left(1 - \frac{d\cos\alpha}{D_m}\right) + f_e \frac{1}{2}\left(1 + \frac{d\cos\alpha}{D_m}\right) \tag{5.1}$$

– de rotation d'un des Z éléments roulants :

$$f_b = \frac{1}{2}\frac{D_m}{d\cos\alpha}\left(1 - \frac{d\cos\alpha}{D_m}\right)\left(1 + \frac{d\cos\alpha}{D_m}\right)|f_e - f_i| \tag{5.2}$$

– de défaut d'un des éléments roulants : $2f_b$, ou aléatoire
– de défaut de bague intérieure : $Z|f_i - f_c|$
– de défaut de bague extérieure : $Z|f_e - f_c|$

En pratique, on constatera dans la majorité des cas que les fréquences caractéristiques des défauts de roulement se situeront, pour une fréquence f_i de rotation de l'arbre intérieur, dans un intervalle s'étendant de 0,4 à 10 f_i.

Tandon et Choudhury [9] présentent une synthèse relativement exhaustive des méthodes applicables à la surveillance et au diagnostic de roulements.

5.3 Détection d'enveloppe

Les signaux vibratoires relevés sur des roulements dégradés contiennent des chocs répétitifs excitant les résonances à haute fréquence. Une analyse directe en fréquence ne donne pas toujours accès à l'information intéressante lorsque le contenu énergétique du signal est, par suite de ces résonances, situé dans ces hautes fréquences. Ces fréquences de répétition sont cependant facile à mettre en évidence dans le signal d'enveloppe.

La démarche classique consiste à d'abord filtrer le signal dans les régions de résonance [10, 11]. Du signal filtré qui ne doit plus reprendre que les impacts répétitifs, on extrait l'enveloppe. Les méthodes usuelles procèdent par détection de pics ou par transformée de Hilbert. Nous proposons, pour un meilleur lissage, d'utiliser un détecteur de valeur efficace qui calcule celle-ci à différentes positions temporelles du signal. Alors une fenêtre glissante de forme et longueur adéquates est utilisée. La figure 5.2 illustre la procédure de détection d'enveloppe.

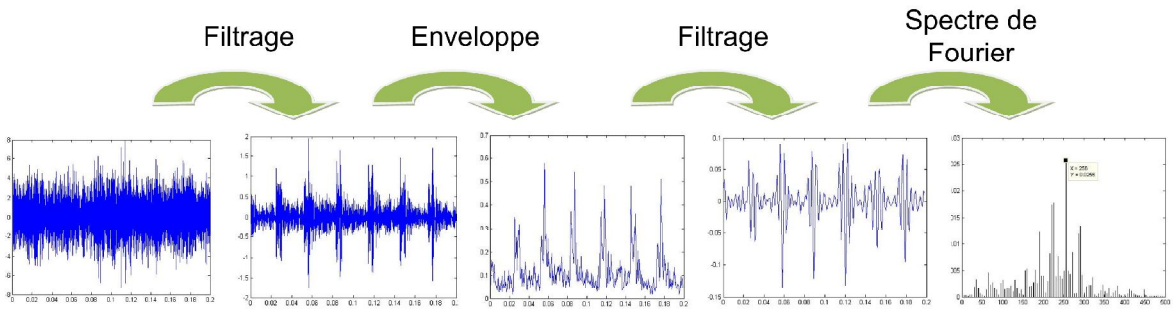


FIG. 5.2 – Exemple de prétraitement pour détection de défaut de roulement

5.4 Etude expérimental : induction des défauts et acquisition des signaux vibratoires

Les tests expérimentaux ont été effectués sur un banc d'essai de détection de défauts de roulements où les signaux sont échantillonnés à 50 kHz. Les essais sont réalisés à trois vitesses de rotation et charges différentes. L'ensemble de données étudiées contient 353 enregistrements vibratoires suivant trois directions orthogonales. Le banc d'essai et les détails sur les tests sont fournis dans l'annexe D. Les cas étudiés sont repris dans le tableau 5.1.

		Taille défaut	Désignation	Effectifs par direction de mesure
1.	Roulement sain	0 mm^2	Niveau 0	8 signaux
2.	Défaut très faible	0,196 mm^2	Niveau 1	8 signaux
3.	Défaut faible	≈0,250 mm^2	Niveau 2	176 signaux
4.	Défaut avancé	≈1,5 mm^2	Niveau 3	71 signaux
5.	Défaut très avancé	3 mm^2	Niveau 4	90 signaux

TAB. 5.1 – Tailles de défauts de roulements étudiés

5.5 Évolution du niveau fréquentiel et des indicateurs temporels avec la dégradation

Observons, dans le spectre fréquentiel, l'évolution du niveau de la fréquence BPFI[†] avec la dégradation. Nous choisissons pour la comparaison des enregistrements effectués à 2000 tr/min et avec une charge de 7000 N. Les figures 5.3 à 5.7 montrent que le pic à la fréquence BPFI connaît une croissance avec la dégradation, mais ces niveaux restent noyés par rapport aux autres raies dans le spectre (voir figure 5.9).

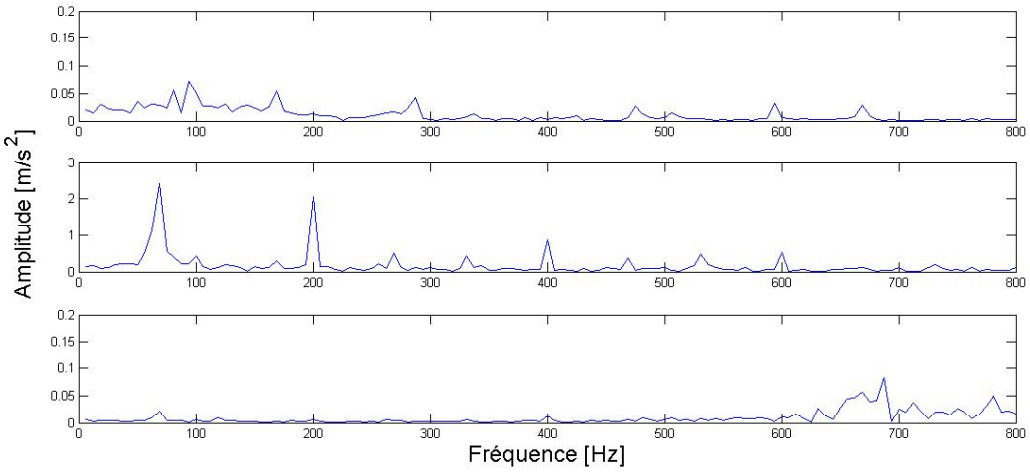


FIG. 5.3 – Spectres du roulement sans défaut (de haut en bas : direction H, A et V)

Nous procédons au même examen dans le domaine temporel où le signal est filtré dans les gammes de fréquences <1000, 1000-3000, 3000-5000, 5000-10000 et >10000 Hz. Ces bandes fréquentielles sont suggérées par la configuration spectrale des signaux typiques. Les indica-

[†]Fréquence de passage des éléments roulants sur la bague intérieure

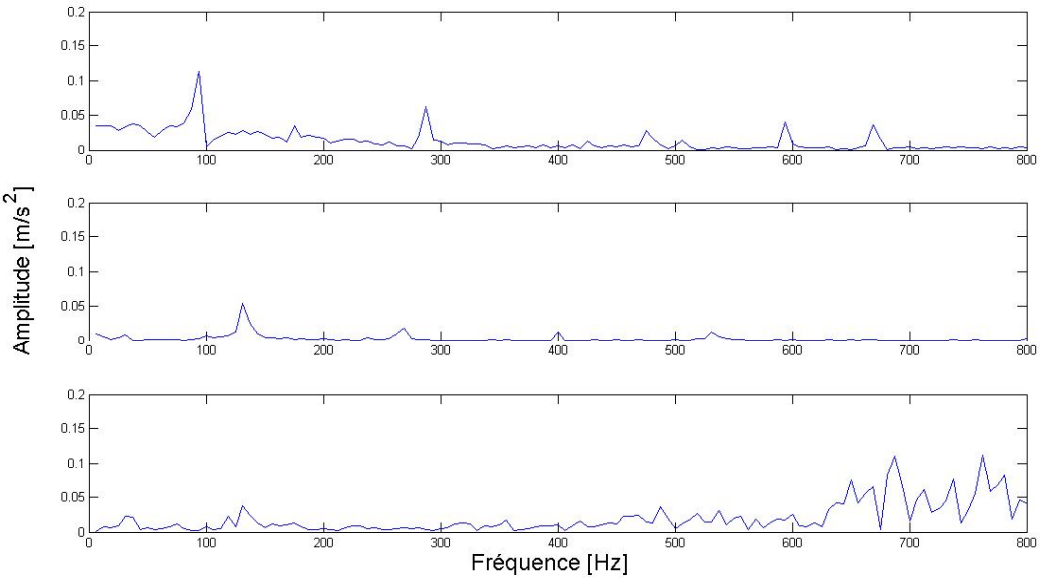


FIG. 5.4 – Spectres du roulement avec défaut de niveau 1 (de haut en bas : direction H, A et V)

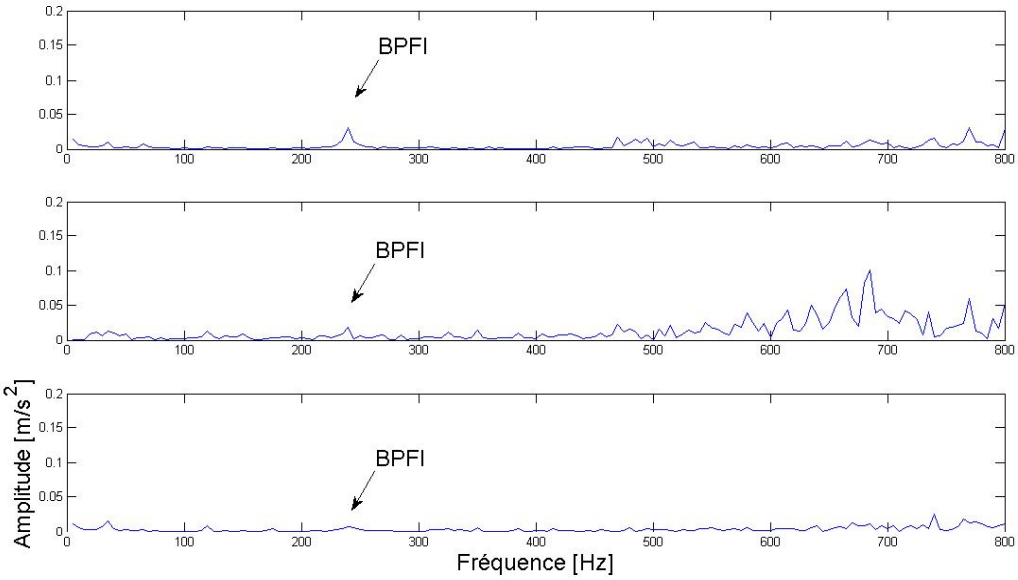


FIG. 5.5 – Spectres du roulement avec défaut de niveau 2 (de haut en bas : direction H, A et V)

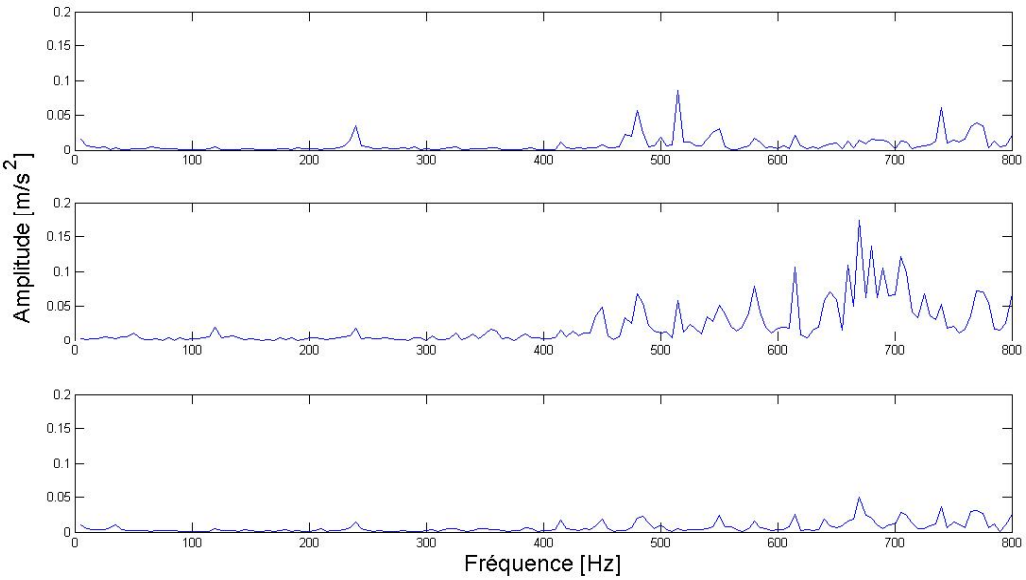


FIG. 5.6 – Spectres du roulement avec défaut de niveau 3 (de haut en bas : direction H, A et V)

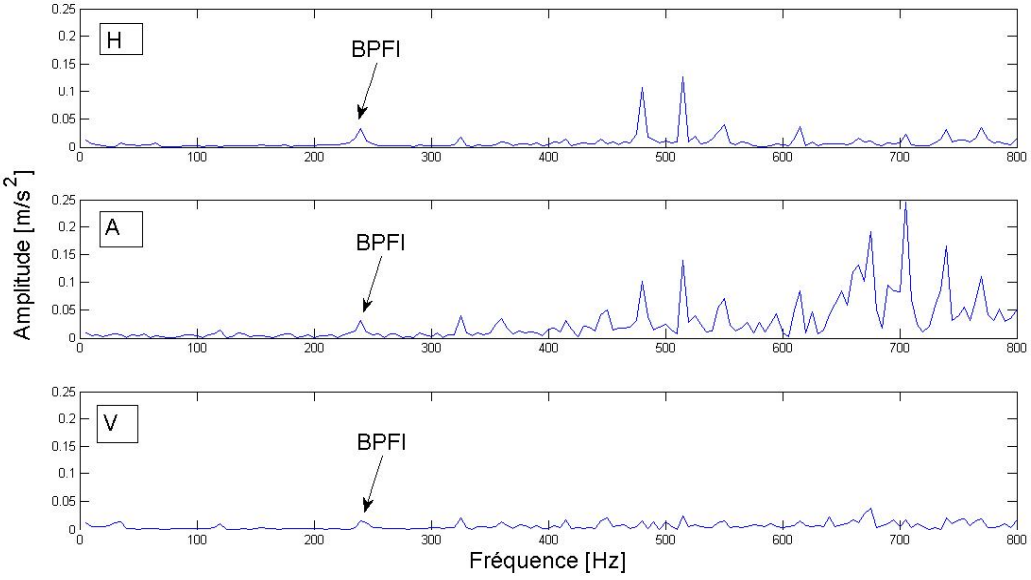


FIG. 5.7 – Spectres du roulement avec défaut de niveau 4 (de haut en bas : direction H, A et V)

teurs scalaires sont calculés puis normalisés avec le carré de la vitesse et la charge. Il en ressort que la valeur efficace dans certaines gammes de fréquences suit l'évolution de la dégradation : celle définie sur la gamme 3-5 kHz de la direction de mesure H illustre ce fait sur la figure 5.8.

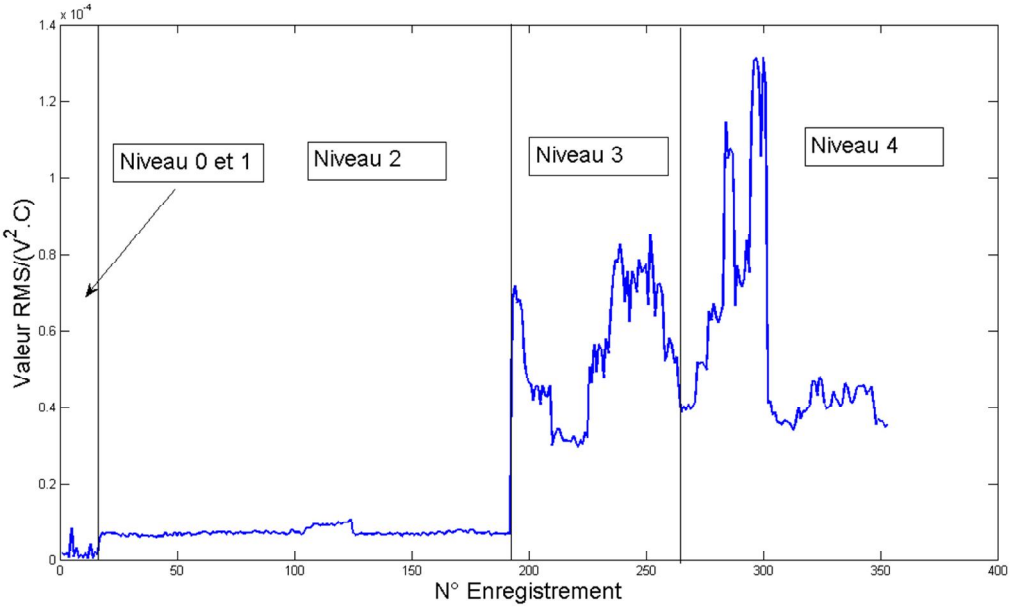


FIG. 5.8 – Evolution de la valeur efficace normalisée par rapport au carré de la vitesse (V) et à la charge (C)

Malgré cette tendance à suivre l'évolution de la dégradation, la valeur efficace a l'inconvénient de pouvoir être influencée par d'autres défauts qui peuvent se manifester dans la même bande de fréquence. La détection d'enveloppe permet de s'affranchir de cette limitation.

5.6 Prétraitement des signaux

Dans la phase initiale du défaut de roulement, il est difficile de détecter une singularité dans le domaine temporel et dans le domaine fréquentiel comme l'indique la figure 5.9 pour un enregistrement typique au début du test.

Du bruit noie le contenu utile à la détection de l'avarie sur le roulement. Nous procédons par l'extraction d'un signal d'enveloppe après un préfiltrage. Trois méthodes de préfiltrage sont testées et comparées :

1. **filtrage passe-bande** : un filtre de Butterworth est utilisé parce que ce type de filtre pré-

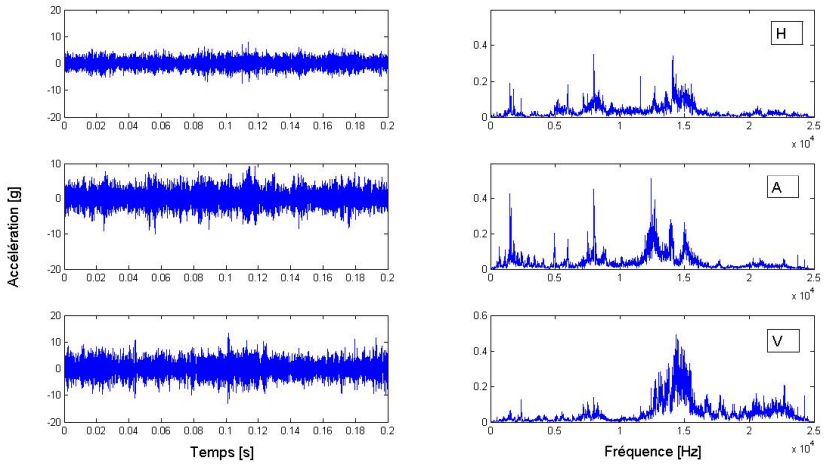


FIG. 5.9 – Signaux temporels et spectres fréquentiels typiques au début du test (Directions radiales horizontale (H) et verticale(V), direction axiale (A))

sente un gain presque constant sur la bande passante. L'ordre 5 pour le filtre présente un meilleur compromis entre le besoin d'atténuation rapide hors bande passante et le déphasage qui doit être réduit. En effet, des ordres élevés du filtre de Butterworth présentent une dérive non négligeable de phase. Les bandes fréquentielles 0-1000, 1000-3000, 3000-5000, 5000-10000 et >10000 Hz sont retenues pour avoir 5 signaux respectivement notés Y_0, Y_1,Y_2,Y_3 et Y_4 (figure 5.10) ;

2. **analyse spectrale singulière** : nous utilisons l'ASS comme moyen de filtrage. Elle est appliquée par décomposition-reconstruction en cascade (arbre complet de niveau 2). Les fenêtres d'analyse sont choisies longues de 50 points par commodité de calcul. L'analyse par *scree test* de l'évolution des valeurs propres sur les signaux a permis de baser les reconstructions sur les 15 premières composantes principales. Toutes les composantes retenues sont regroupées pour ne reconstruire qu'un seul signal à chaque niveau. On aboutit à 4 signaux notés YY, YR, RY et RR (figure 5.11) ;

3. **décomposition en ondelettes** : une décomposition en 5 niveaux est effectuée avec l'ondelette de Daubechies d'ordre 4. Le choix de cette ondelette est justifié par ses propriétés de régularité et d'asymétrie qui semble adaptées aux impulsions caractéristiques des défauts du roulement. On aboutit à 6 signaux : 5 signaux de détails notés $D1$ à $D5$ et un signal d'approximation $A5$ (figure 5.12).

Ensuite, pour chaque bande, un signal d'enveloppe est extrait par la procédure décrite ci-

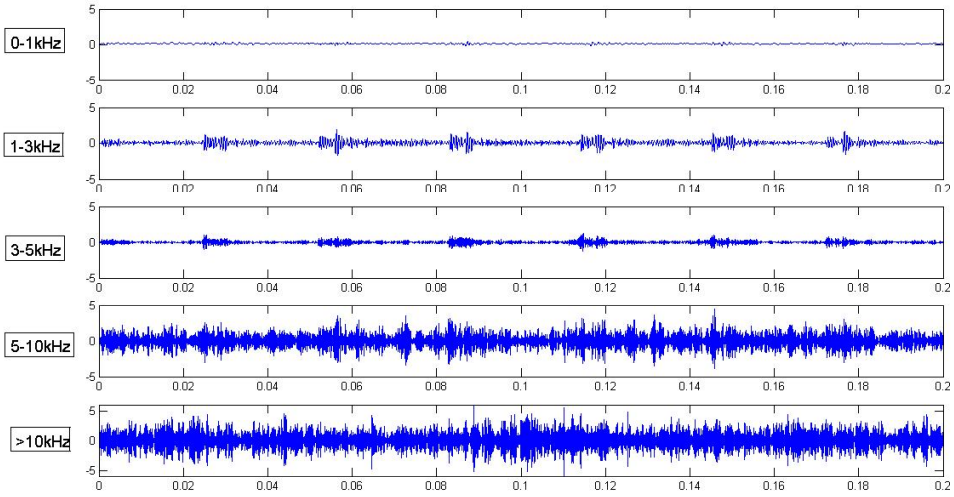


FIG. 5.10 – Préfiltrage par filtrage

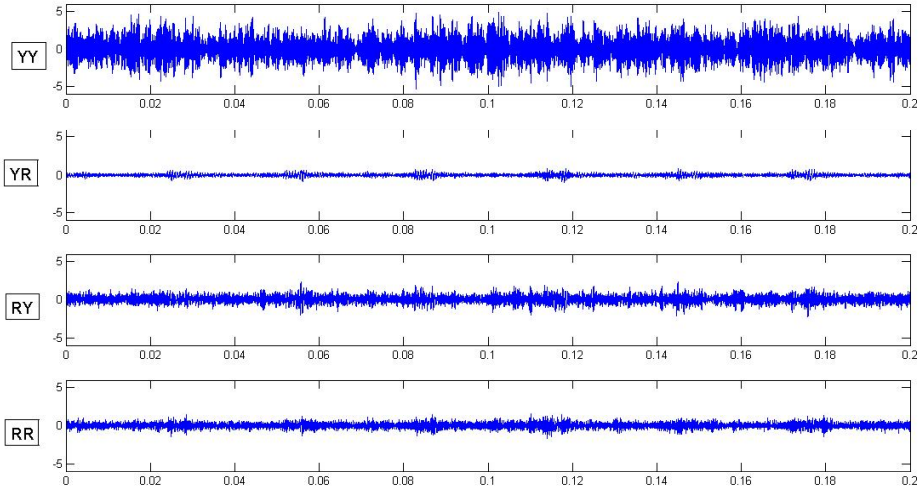


FIG. 5.11 – Préfiltrage par analyse spectrale singulière

après. Le signal est scanné à l'aide d'une fenêtre glissante de pondération exponentielle décroissante, avec un recouvrement de 50 %. La valeur efficace calculée pour chaque fenêtre est assignée au point courant qui correspond au début de la fenêtre. La fenêtre exponentielle a pour effet d'atténuer sensiblement les contributions des points éloignés du point courant. Les impul-

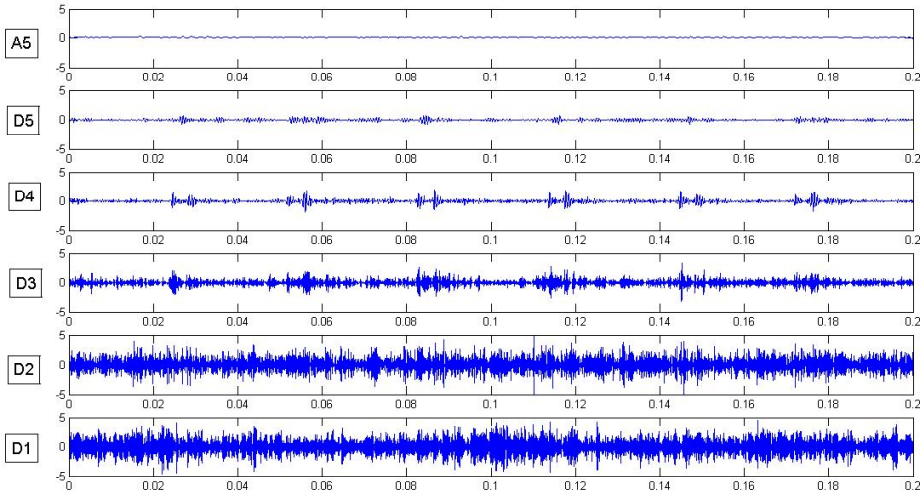


FIG. 5.12 – Préfiltrage par analyse en ondelettes

sions individuelles du signal durent environs 20 ms. Ayant choisi de les couvrir avec 20 points de valeurs efficaces, la longueur de la fenêtre glissante a été fixée à 2 ms pour un recouvrement de 50%. Le signal d'enveloppe est à un facteur près obtenu à partir des valeurs efficaces (figure 5.13).

Le signal d'enveloppe sera ensuite filtré en basses fréquences autour de la fréquence caractéristique du roulement avant de le représenter dans le domaine spectral. La figure 5.14 schématise le traitement de données pour l'extraction d'indicateurs. Les figures 5.15 à 5.17 illustrent le traitement par filtrage d'un signal relevé dans la direction H pour une vitesse de rotation de 2000 tr/min, roulement sans défaut.

Cette procédure est appliquée aux signaux relevés suivant les trois directions H, A, et V. Les indicateurs pour chaque signal préfiltré, dans chaque direction, sont construits par la somme des raies spectrales. Par exemple, pour la décomposition en ondelettes, on a 3 indicateurs pour $A5$ et pour chacun des signaux $D1$ à $D5$. Les trois indicateurs sont extraits des trois directions de mesure. La fusion de niveau 2 donne 3×6 indicateurs.

Ces ensembles d'indicateurs définis par les sommes des raies spectrales sont complétés avec la charge [N] et normalisés par le carré de la vitesse.

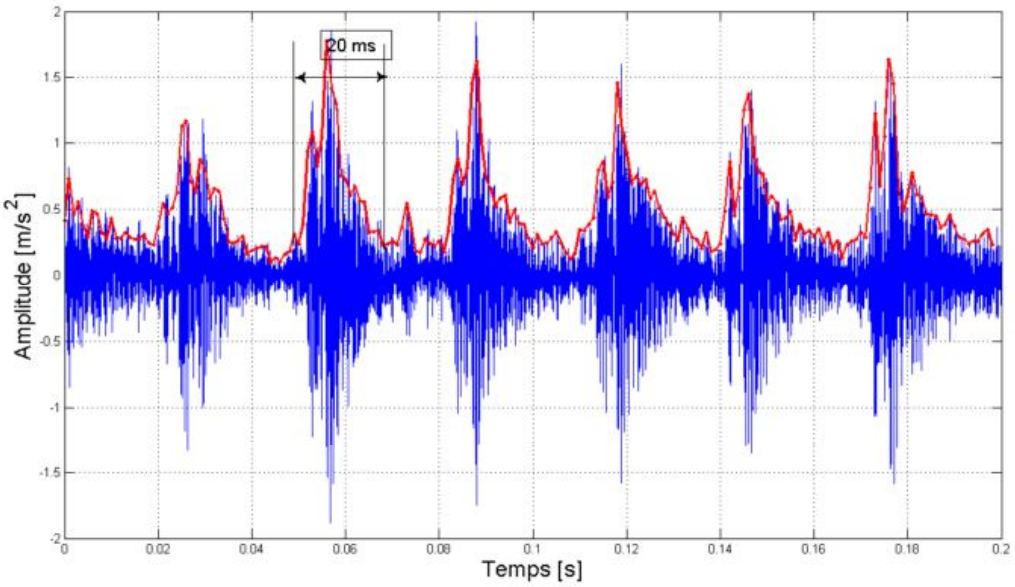


FIG. 5.13 – Signal d'enveloppe

5.7 Détection précoce

L'objectif est d'arriver à détecter un défaut dans sa phase initiale, tel que présumé pour les niveaux 1 et 2 des essais menés sur roulements. Voulant tirer profit des informations de chacun des signaux relevés dans les 3 directions, on se situe dans un cadre d'une analyse multivariée. Nous proposons une application de l'analyse en composantes principales.

5.7.1 Détection de défaut par analyse en composantes principales

L'ACP comme moyen de détection de défauts a déjà été étudiée de manière intensive. Mais son application n'a principalement concerné que le domaine de la surveillance des processus chimique, où le nombre de capteurs utilisés est généralement important [12, 13]. On lui trouve aussi des applications pour la détection des défauts de capteurs dans le même domaine des processus chimique [14].

En détection de dégradations mécaniques, cette technique ne s'est pas faite assez de place probablement à cause du fait que le nombre de variables à surveiller n'est généralement pas grand. Baydar *et al.* [15] l'ont utilisée pour la détection précoce des défauts d'engrenages hélicoïdaux.

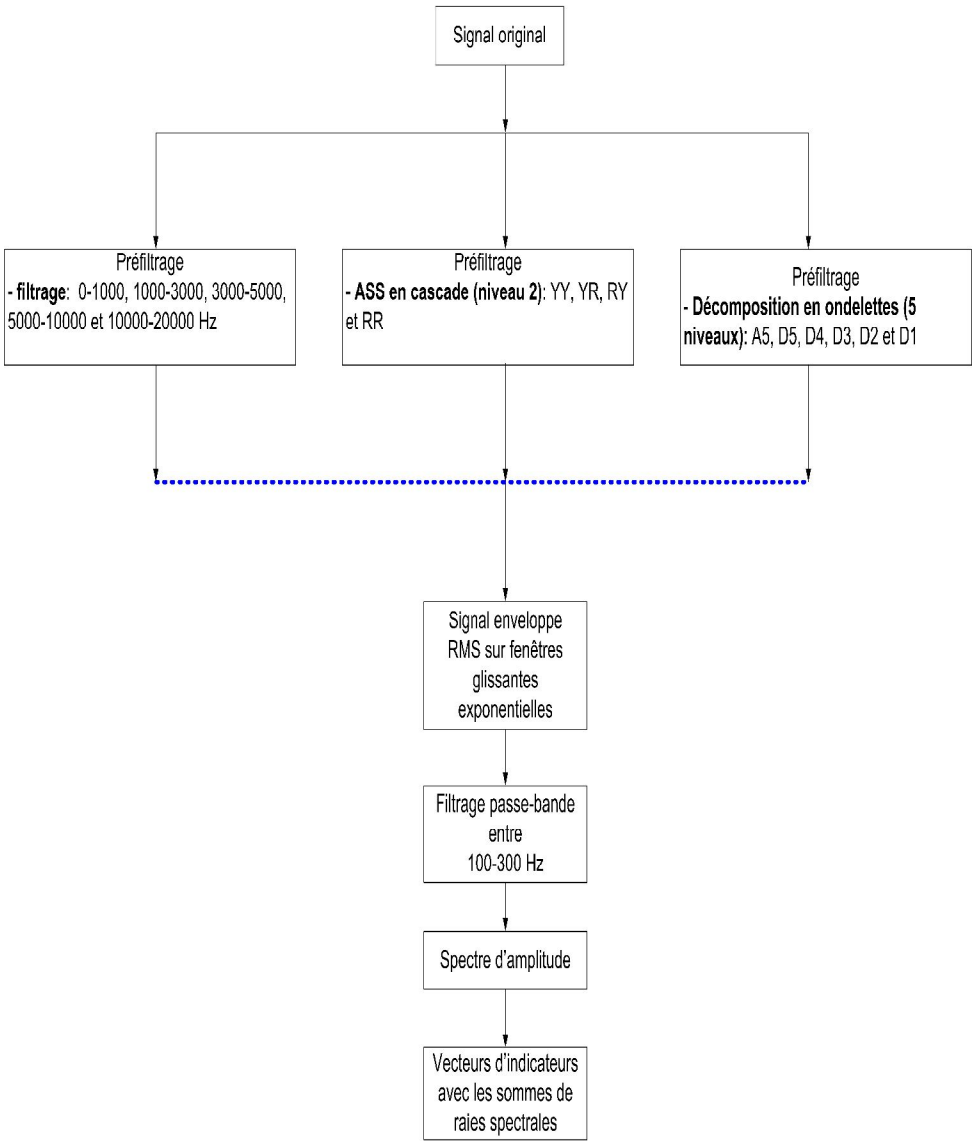


FIG. 5.14 – Organigramme de traitement de données

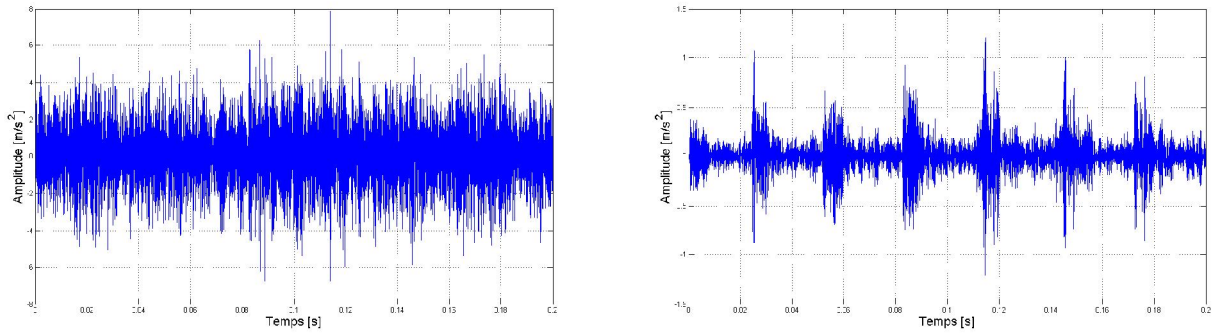


FIG. 5.15 – Signal original (à gauche) et son filtrage dans la gamme 3-5 kHz (à droite)

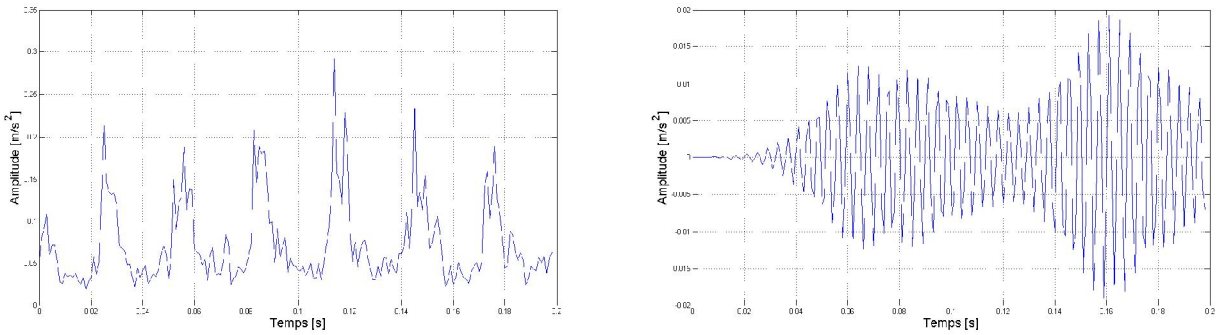


FIG. 5.16 – Signal enveloppe (à gauche) et son filtrage autour de la fréquence BPFI (à droite)

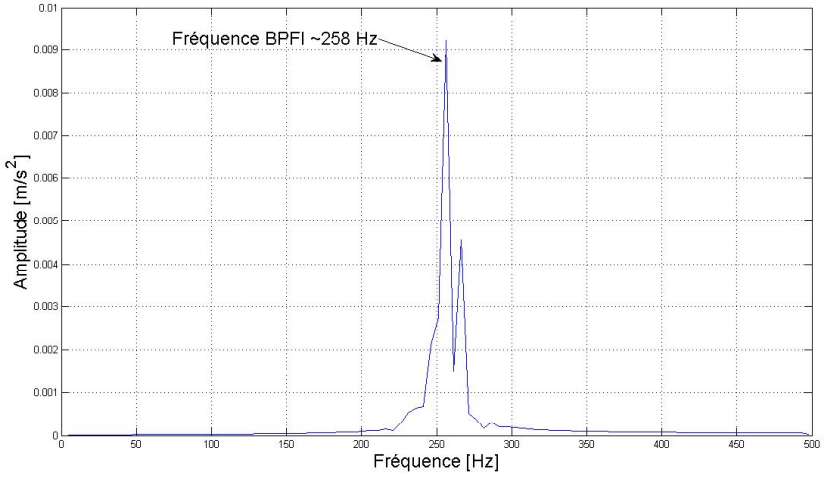


FIG. 5.17 – Spectre du signal enveloppe filtré autour de la fréquence BPFI

Pourtant, il peut être très intéressant de faire usage de cette technique de détection pour les défauts mécaniques en combinaison avec des méthodes de *data mining*. En effet, l'idée de procéder à la construction d'un maximum d'indicateurs vibratoires et d'exploiter l'historique des équipements, nous place dans un contexte où cette technique d'analyse multivariée peut être utile. Le concept de base de l'utilisation de l'ACP pour la détection est résumé ci-après.

Considérons un ensemble de données représentatif d'un fonctionnement normal et construisons à partir de celui-ci une matrice de transformation en composantes principales. Tout nouveau vecteur de données sera projeté dans l'espace des composantes principales avant de subir une transformation inverse vers l'espace de départ. On évaluera ensuite l'écart entre le vecteur de départ et sa transformation inverse.

Soit, par exemple, une matrice $\mathbf{X}$ d'indicateurs (normalisés) représentant plusieurs états observés de fonctionnement normal, que nous convenons de désigner *base normale*. Sa transformation en composantes principales se fera à l'aide d'une matrice de transformation $\mathbf{A}$ de la façon suivante (voir sous-section 3.7.1) :

$$\mathbf{Y} = \mathbf{X}.\mathbf{A} \tag{5.3}$$

En ne retenant que les composantes principales à forte contribution, on obtient la matrice $\mathbf{Y}'$ au lieu de $\mathbf{Y}$. De même, la considération des seuls vecteurs propres correspondant aux axes retenus conduit à une matrice de transformation $\mathbf{E}$ à la place de $\mathbf{A}$. Ce qui peut s'écrire

$$\mathbf{Y}' = \mathbf{X}.\mathbf{E} \tag{5.4}$$

L'estimation $\mathbf{X}^*$ de $\mathbf{X}$ par transformation inverse se fera de la manière suivante :

$$\mathbf{X}^* = \mathbf{Y}'.\mathbf{E}^T \tag{5.5}$$

Il vient que l'on peut évaluer l'écart entre $\mathbf{X}$ et sa reconstruction par la matrice de résidus

$$\begin{aligned} \mathbf{R} = \mathbf{X} - \mathbf{X}^* &= \mathbf{X} - \mathbf{Y}'.\mathbf{E}^T \\ &= \mathbf{X} - \mathbf{X}.\mathbf{E}.\mathbf{E}^T \\ &= \mathbf{X}\left[\mathbf{I} - \mathbf{E}\mathbf{E}^T\right] \end{aligned} \tag{5.6}$$

Les vecteurs constituant la matrice des résidus sont notés $\mathbf{r}$. Pour un vecteur d'indicateurs

x relatif au vecteur **r**, la déviation par rapport à l'état normal est calculée comme le carré de la norme de $\mathbf{r}_{1\times N}$:

$$R(\mathbf{x}) = \mathbf{r}_{1\times N}.\mathbf{r}^{T}_{N\times 1} \tag{5.7}$$

Dans la pratique de l'utilisation de l'ACP pour la détection, ce résidu constitue ce qu'on appelle « statistique Q » ou erreur quadratique d'estimation SPE (*Squared Predicted Error*) [16]. Les figures 5.18 et 5.19 montrent la procédure de construction du résidu.

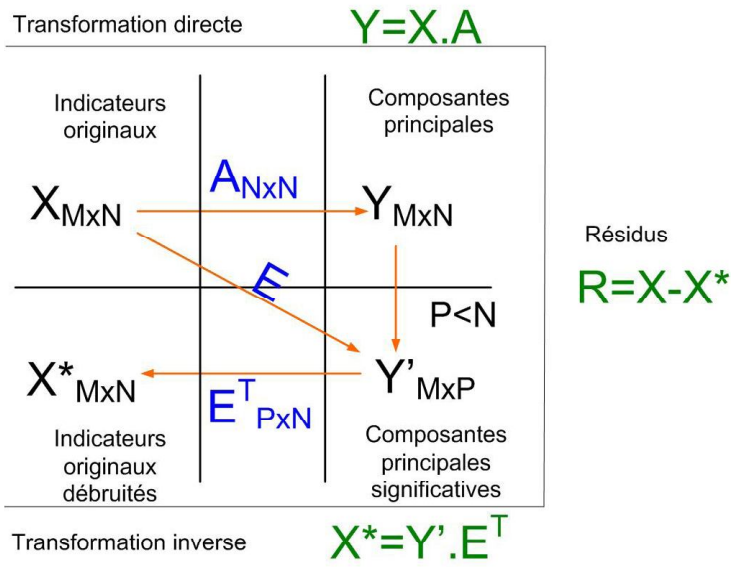


FIG. 5.18 – Procédure de construction du résidu

La question du nombre des composantes principales à utiliser a été abondamment discutée dans la littérature. Les méthodes usuelles basées sur la contribution à la variance totale restent valables. L'utilisation de plus de composantes qu'il n'en faut, conduit à des fausses alarmes alors qu'un nombre insuffisant de composantes présente un risque de non détection. A ce sujet, Tamura et Tsujita [17] ont discuté de la sensibilité de la capacité de détection vis-à-vis du nombre de composantes retenues. La méthode la plus sure et mathématiquement justifiée, utilise la validation croisée en surveillant l'évolution d'un nombre appelé PRESS (Prediction Residual Sum of Squares). L'inconvénient de cette méthode est qu'elle est lourde en calcul. Dunia et Qin [14] ont proposé une méthode basée sur l'erreur sur la variance reconstruite. Cependant, il doit être souligné qu'une méthode simple comme celle de Cattell (*Scree test*) est très efficace. Son seul inconvénient est la difficulté d'être automatisé.

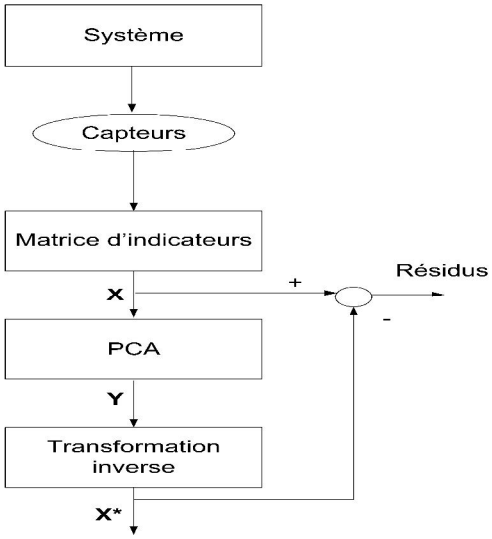


FIG. 5.19 – Détection par analyse en composantes principales

Pour de nouveaux états de fonctionnement dont les données sont stockées dans $\mathbf{X}_1$, on procédera par les étapes suivantes :

1. normaliser $\mathbf{X}_1$ par standardisation
2. calculer la matrice des résidus à partir de la matrice $\mathbf{E}$ déduite des données de fonctionnement normal

$$\mathbf{R} = \mathbf{X}_1 \left[\mathbf{I} - \mathbf{E}\mathbf{E}^T \right] \tag{5.8}$$

Généralement, la détection de défaut par ACP se fait à l'aide des tests statistiques sur la valeur SPE. Nous proposons l'utilisation de l'apprentissage par un arbre de décision pour la détection. Le seuil pouvant être ainsi fixé, l'est par apprentissage et non par des tests statistiques. Ce qui est un réel avantage.

5.7.2 Application à la détection de défaut de roulement

Le nombre de capteurs et les procédés d'extraction des indicateurs utilisés dans cette étude conduisent à une bonne quantité d'indicateurs, selon le préfiltrage adopté pour l'analyse d'enveloppe. Le besoin de détection précoce du défaut de bague intérieure et celui de l'exploitation combinée de tous ces indicateurs a suggéré la modélisation de l'état normal de fonctionnement au moyen de l'ACP.

Pour appliquer cette méthode au cas du roulement, nous avons choisi une base normale constituée des signaux correspondant au roulement sain (au total 8 signaux mesurés dans chaque direction). La procédure de détection est testée pour chaque ensemble de données correspondant aux différents prétraitements. Le nombre de composantes à retenir se fait par *scree test* comme l'illustre la figure 5.20 pour un prétraitement par ASS.

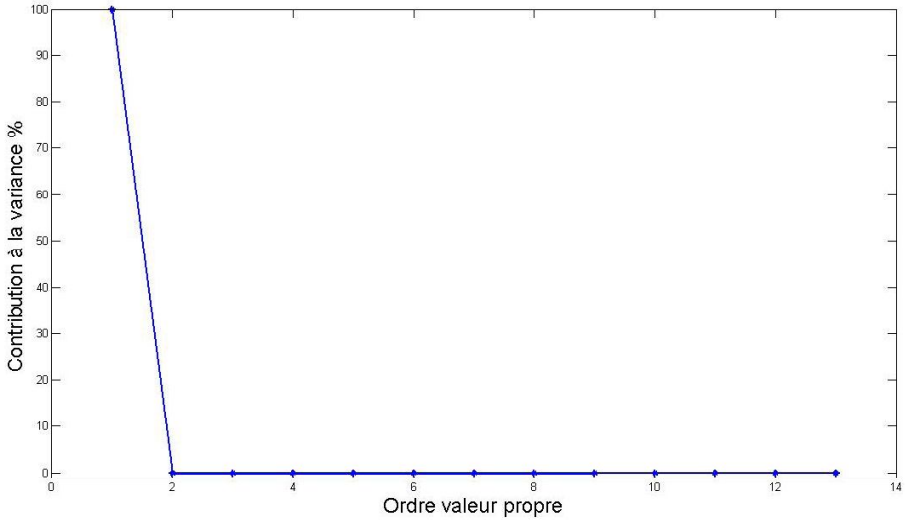


FIG. 5.20 – *Scree test* pour les cas de prétraitement par ASS

La figure 5.21 montre schématiquement comment les résidus sont calculés avec ou sans fusion des indicateurs extraits du domaine fréquentiel après détection d'enveloppe.

Rappelons que chaque prétraitement à conduit à plusieurs signaux desquels un signal d'enveloppe a été extrait. Les résidus calculés sur base de ces différents signaux individuellement n'ont pas fourni une capacité satisfaisante de détection précoce de la dégradation. D'où l'idée de fusionner les vecteurs de sommes des raies pour chaque préfiltrage. C'est ainsi qu'ensuite on obtient des vecteurs de résidus, notés *R-DWT*, *R-FILT* et *R-SSA*.

L'observation des figures 5.22 à 5.24, obtenues avec les résidus issus des indicateurs fusionnés, montre que l'on peut relativement bien détecter visuellement le passage du niveau de dégradation 1 au niveau 2. La détection du passage de l'état sain à la première phase de dégradation (niveau 1) n'est pas très évidente. Il se pose tout de même la question de l'utilité pratique de la détection du passage $0 \rightarrow 1$, puisque le fonctionnement sous l'état 1 a tout d'un fonctionnement normal (sain). En plus, l'état 2 qui est facilement détectable n'est en soi qu'une dégradation de très faible ampleur. Nous ne nous intéresserons donc qu'à la détection du pas-

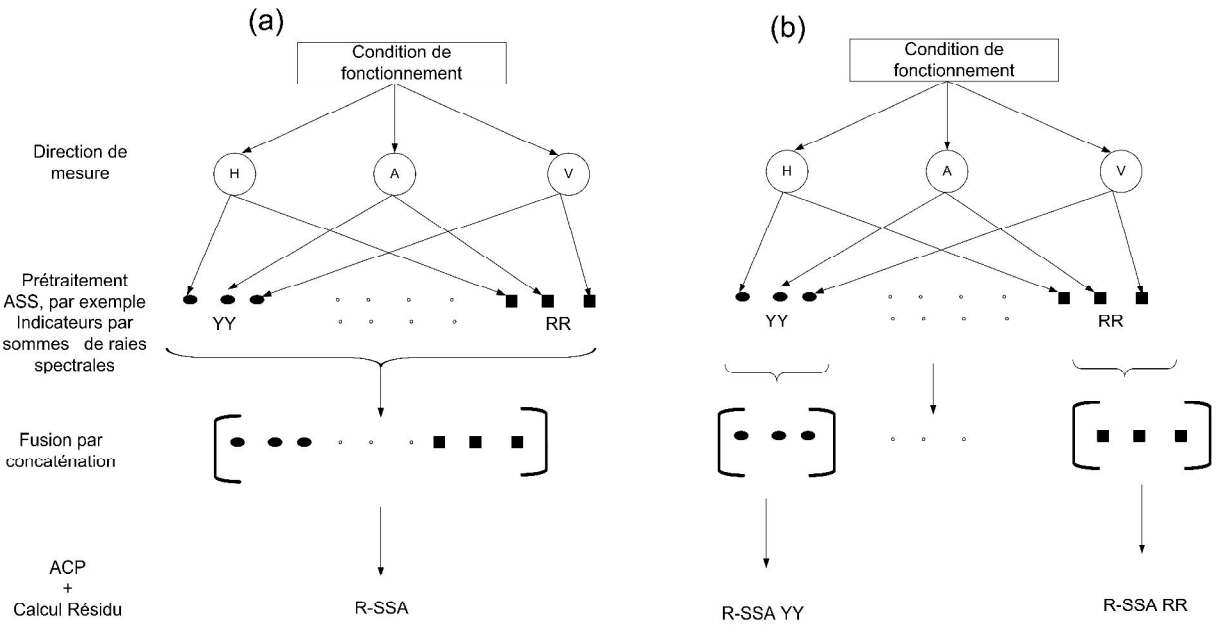


FIG. 5.21 – Processus de calcul du résidu avec fusion (à (illustration avec un prétraitement ASS)

sage 1 → 2.

L'examen visuel des courbes ne permet pas de décider laquelle des méthodes de traitement permet une meilleure détection de ce passage, ces courbes étant représentées à des échelles très différentes. Ceci suggère l'automatisation de cette tâche au moyen d'un système d'apprentissage. Le meilleur choix sera celui d'un système transparent dont le résultat de détection peut se traduire en règles compréhensibles, ce qui est le cas des arbres de décision [18]. Un système d'apprentissage rend aussi possible l'exploitation simultanée de l'information contenue dans plusieurs variables.

On n'exigera pas des variables de détection précoce des capacités de reconnaissance de toutes les phases de dégradation, la détection précise du début de la dégradation étant suffisante.

Nous construisons d'abord un arbre de décision pour chacune des trois variables *R-DWT*, *R-FILT* et *R-SSA*, puis un arbre avec leur fusion. Les matrices de confusion sont représentées dans les tableaux 5.2 à 5.5. Comme il est question de détection précoce de défaut, le système doit pouvoir établir une séparation nette entre les deux groupes d'instances : niveaux 0 et 1 (sains) et niveaux 2 à 4 (dégradés). Hors diagonale de la matrice de confusion, on a les objets non convenablement classés.

A l'examen de ces résultats, on remarque que les capacités de détection obtenus avec les différents prétraitements sont comparables et que la possibilité de détecter le défaut dans sa

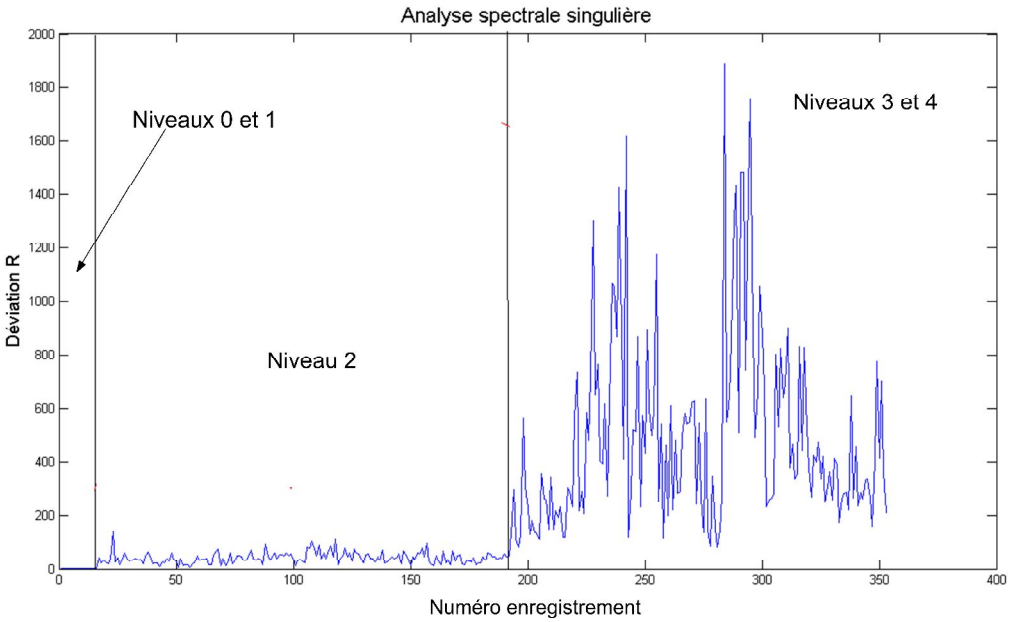


FIG. 5.22 – Détection sur base d'analyse spectrale singulière avec fusion de différents résidus

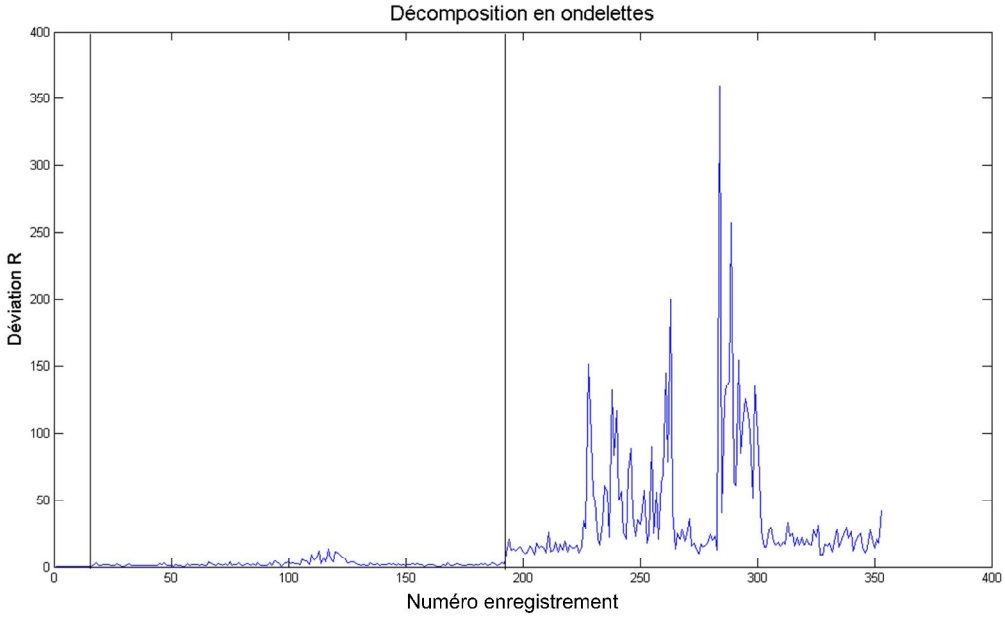


FIG. 5.23 – Détection sur base d'une décomposition en ondelettes (variables fusionnées)

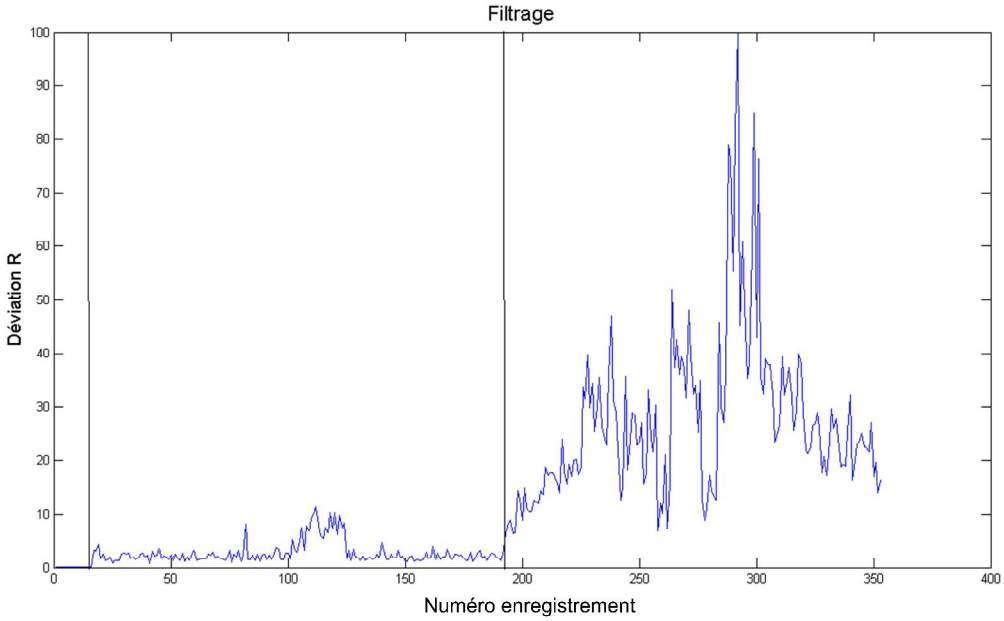


FIG. 5.24 – Détection sur base d'un filtrage (variables fusionnées)

phase initiale est établie. Ceci rejoint la conclusion tirée sur base de l'inspection visuelle des courbes.

↗	sains	dégradés
sains	14	2
dégradés	2	335

TAB. 5.2 – *R-DWT*

↗	sains	dégradés
sains	15	1
dégradés	0	337

TAB. 5.3 – *R-FILT*

↗	sains	dégradés
sains	15	1
dégradés	0	337

TAB. 5.4 – *R-SSA*

↗	sains	dégradés
sains	15	1
dégradés	0	337

TAB. 5.5 – Fusion des trois résidus

Les figures 5.25 et 5.26 illustrent les arbres relatifs à l'exploitation des 3 variables de résidu. Le regroupement de tous les niveaux dégradés en une classe permet d'obtenir un arbre compact san modification des seuils de décision. L'analyse de ces arbres montre que

1. la discrimination entre l'état sain (niveaux 0 et 1) et les états dégradés est déterminé par l'indicateur *R-FILT*
2. *R-DWT* permet de distinguer les niveaux 0 et 1 entre eux

Il vient qu'on peut fixer le seuil de surveillance par la règle suivante :

Si *R-FILT* $\leq 0,055102$ alors «*Etat Roulement=Sain*»

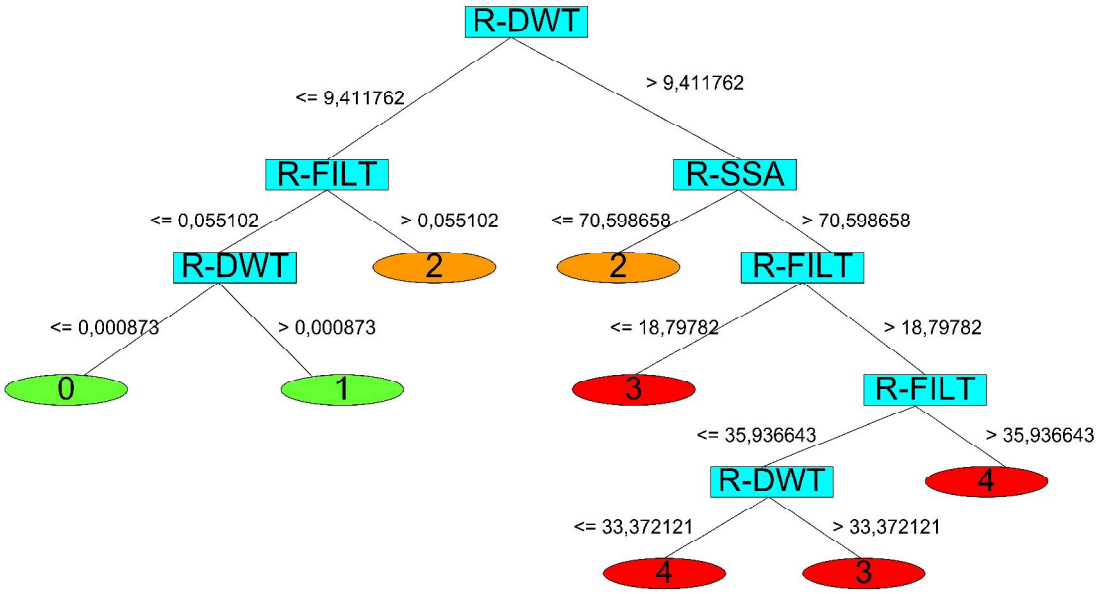


FIG. 5.25 – Arbre

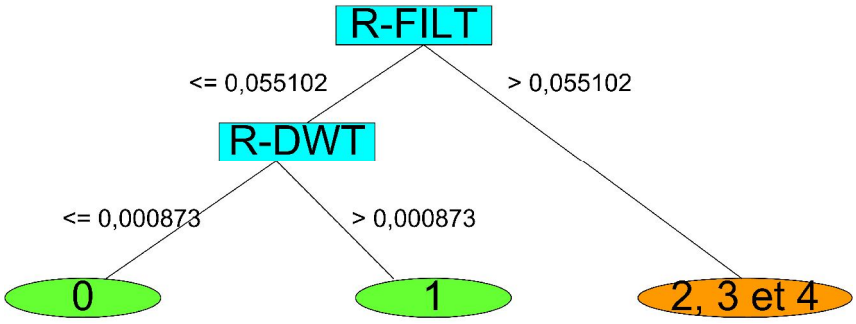


FIG. 5.26 – Arbre

Synthèse

Dans ce qui a précédé, nous avons montré qu'il était possible de construire un système de détection sur base d'un arbre de décision qui fixe un seuil sur les indicateurs construits par ACP. En exploitation, le modèle peut être déployé *on-line* tant le temps de calcul reste raisonnable. Les étapes successives de traitement que subit un nouvel enregistrement, suivant les trois directions, seraient les suivantes en utilisant le prétraitement par filtrage :

1. Extraire le signal d'enveloppe : filtrage en plusieurs bandes → enveloppe → filtrage en zone BPFI ;
2. calculer les spectres des signaux obtenus ;
3. calculer les sommes de raies spectrales pour former le vecteur d'indicateurs (3 indicateurs correspondant aux 3 directions de mesure) ;
4. fusionner les vecteurs d'indicateurs de différentes bandes de filtrage ;
5. introduire le vecteur obtenu par fusion dans le modèle ACP lui correspondant pour obtenir les résidus ;
6. fournir à l'arbre de décision les résidus fusionnés.

Nous avons insisté sur l'avantage d'utiliser l'analyse d'enveloppe par rapport aux indicateurs scalaires ou au suivi de l'accroissement du niveau spectral.

Enfin, nous aimerons souligner le mérite de cette approche pour la détection précoce vue la taille très faible du défaut détecté. En effet, l'analyse de la littérature à ce sujet a révélé l'utilisation des tailles de défauts détectables plus grandes pour différentes méthodes de détection [1, 19–21].

5.8 Diagnostic

La détectabilité de la dégradation du roulement à rouleaux cylindriques dans sa phase initiale a été établie. L'étude menée dans cette section va ambitionner la mise en évidence des capacités d'identification des différents niveaux de sévérité du défaut de la bague intérieure.

L'évaluation des performances du diagnostic va tenir compte de la notion du coût de mauvaise classification (fausses alarmes et non-détection). La matrice de coûts sera construite sur base des criticités des défauts. Nous avons rangé les différents niveaux sur une échelle de criticité allant de 0 à 1, 1 pour la dégradation avancée et 0 pour le roulement sain (Tableau 5.6).

Niveau 0	Niveau 1	Niveau 2	Niveau 3	Niveau 4
0	0,1	0,3	0,8	1

Tab. 5.6 – Criticité des défauts

Faisant usage de la démarche proposée par Ashby *et al.* [22] pour la construction de la matrice de coûts, nous avons utilisé comme coefficient de réduction $m = 0,8$, mais sans procéder

à la normalisation et à la multiplication par S[‡]. Nous avons abouti à la matrice de coûts reprise sous le tableau 5.7.

	Prédiction				
	Niveau 0	Niveau 1	Niveau 2	Niveau 3	Niveau 4
Niveau 0	0	0,08	0,24	0,64	0,8
Niveau 1	0,1	0	0,16	0,56	0,72
Niveau 2	0,3	0,2	0	0,4	0,56
Niveau 3	0,8	0,7	0,5	0	0,16
Niveau 4	1	0,9	0,7	0,2	0

TAB. 5.7 – Matrice des coûts

Nous avons testé les différents ensembles d'indicateurs sur les algorithmes de classification suivants :

- arbre de décision : noté J48 ;
- réseau de neurones : MLP (1 couche cachée, le nombre de neurones dans la couche cachée est égale à la moitié de la somme du nombre de classes et du nombre d'attributs). Nous avons systématiquement prévu une couche cachée pour le réseau de neurones. Le choix du nombre de neurones caché comme la moitié de la somme du nombre de classes et du nombre d'attribut, est empiriquement démontré comme un bon choix de départ [23]. Si ce choix n'est pas satisfaisant, il suffira de l'améliorer par essai et erreur.
- machine à vecteurs de support : SVM (noyau polynomial de degré 2, coefficient de régularisation 10) ;
- plus proches voisins : K-NN (le nombre qui suit indique le nombre de proches voisins. Par exemple, K-NN 2 est relatif à 2 voisins proches). Il sied toutefois de souligner ici que l'utilisation d'un nombre pairs de voisins peut conduire à des indéterminations pour le choix de la majorité. Mais ce risque peut être évité par l'usage des pondérations lors des décisions.
- réseaux bayésiens : RB (RB 1 : 1 seul parent, RB 2 : 2 parents au maximum).

Les indicateurs sont composés des sommes de raies spectrales augmentées de la charge et de la vitesse. Ce qui aboutit aux nombres suivants d'indicateurs :

L'ensemble de données contient 353 exemples. Une expérimentation intensive est lancée avec 10 cycles de validation croisée à 10 blocs chacun. Le coût retenu est la moyenne des coûts non nuls obtenus sur tous les cycles de validation croisée. Ceci permet de tirer des conclusions

[‡]détaillé à la section 4.9

Prétraitement	Indicateurs non fusionnées	Indicateurs non fusionnées
DWT	5	20
ASS	5	14
FILT	5	17

TAB. 5.8 – Nombre de variables

de comparaison statistiquement valables entre les résultats obtenus de différents classifieurs. Les taux de classification, avec leurs écart-types, sont représentés dans le tableau 5.9

Données	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
DWT-A5	90.73±4.32	77.73±5.50	72.56±6.12	90.54±4.27	85.52±4.08	86.52±4.78	85.24±4.23	88.46±5.80	88.95±5.22
DWT-D1	94.28±4.39	80.72±5.10	76.84±4.93	80.18±6.38	82.36±4.89	82.05±5.43	81.85±5.05	86.43±5.31	87.19±5.37
DWT-D2	92.35±3.93	83.70±5.71	77.90±5.41	83.72±5.65	84.16±5.17	84.91±6.33	84.45±5.50	88.10±5.91	87.48±5.71
DWT-D3	95.24±2.98	91.95±4.65	90.84±4.83	96.20±2.71	94.79±3.23	94.82±3.56	93.94±3.41	93.93±4.25	94.53±4.25
DWT-D4	97.28±2.92	87.93±5.03	81.90±4.69	92.84±3.89	90.66±4.00	92.41±3.89	91.34±4.09	91.07±4.95	90.62±4.69
DWT-D5	95.80±3.27	90.86±4.68	77.58±5.36	96.09±2.58	93.01±3.04	94.45±2.83	94.17±2.89	90.78±5.29	92.68±4.77
DWT-Fusion	95.89±3.60	96.85±2.77	96.74±3.02	97.99±2.06	97.37±2.24	97.73±2.24	97.28±2.51	95.75±3.73	96.96±3.21
FILT-Fusion	94.91±3.55	99.04±1.35	96.95±2.71	98.67±1.86	97.96±2.10	98.30±1.98	97.48±2.01	94.19±3.55	96.66±3.19
FILT-Y0	95.73±2.83	82.10±5.42	62.38±4.51	88.87±4.18	85.81±5.11	87.39±5.16	85.81±4.64	90.96±4.72	91.88±5.06
FILT-Y1	95.61±3.68	85.77±5.49	80.47±4.82	95.39±2.91	93.04±3.73	93.54±3.48	91.31±3.88	90.28±5.14	92.92±4.59
FILT-Y2	96.06±3.15	92.52±4.11	89.72±4.22	95.98±3.42	94.68±3.54	94.99±3.26	93.91±3.39	94.96±3.43	94.81±4.13
FILT-Y3	95.67±3.14	93.66±3.65	89.32±4.18	95.16±3.30	94.17±3.18	94.71±3.26	93.92±3.47	92.86±3.81	93.21±4.18
FILT-Y4	95.01±3.61	83.15±5.08	79.10±4.17	83.46±4.94	83.80±5.40	84.94±5.40	84.00±5.28	92.15±5.20	93.49±4.54
SSA-Fusion	95.01±3.73	91.55±4.47	90.54±4.00	95.23±3.68	94.21±3.53	94.84±3.51	93.84±3.46	93.86±4.81	93.76±4.62
SSA-rr	94.98±3.92	81.67±5.28	78.84±4.09	83.65±5.63	82.41±5.60	83.01±5.64	81.76±5.67	88.68±5.19	87.24±5.57
SSA-ry	93.57±4.49	82.53±5.31	79.13±4.35	83.79±5.85	82.90±5.46	86.01±4.82	84.85±5.00	91.85±5.28	91.59±5.57
SSA-yr	94.87±3.37	83.72±4.71	81.45±5.21	86.49±4.94	84.11±4.77	87.26±5.29	87.31±4.71	91.91±5.38	92.09±4.79
SSA-yy	92.89±4.29	87.93±5.31	81.16±5.54	88.83±4.98	87.68±5.01	87.17±5.42	86.49±4.83	90.54±6.08	90.45±5.60

(1) J48 ;(2) MLP ;(3) SVM ; (4) K-NN 1 ; (5) K-NN 2 ;(6) K-NN 3 ; (7) K-NN 4 ; (8) RB1 ; (9) RB2

TAB. 5.9 – Taux de classification et écart-type

Cependant, comme nous venions de le souligner, le taux de classification seul ne suffit pas pour l'évaluation de la performance du système de diagnostic. Il faut tenir compte du coût des erreurs de classification pour le choix de la méthode à implémenter pour cette étude.

Les coûts pour les différents systèmes de diagnostic sur les différents espaces d'indicateurs sont repris dans le tableau 5.10.

On remarque que le meilleur coût, comme le meilleur taux de classification, est obtenu avec le réseaux de neurones (MLP) sur les données *FILT-Fusion.*

Ayant retenu comme système de diagnostic le classifieur *MLP/FILT-Fusion*, nous allons pouvoir nous assurer de sa robustesse par examen du coût et du taux de classification en augmentant le nombre de cycles de validation croisée. Le tableau 5.11 montre que le taux de classification ne varie presque pas avec le nombre de cycles (0,14% de variation). Malgré sa variation qui atteint 26%, le coût total obtenu avec *MLP/FILT-Fusion* reste le plus faible par rapport aux autres classifieurs.

	J48	MLP	SVM	KNN 1	KNN 2	KNN 3	KNN 4	RB1	RB2
DWT-A5	19,32	25,46	42,94	9,6	16,18	17,32	18,28	15,02	12,66
DWT-D1	5,8	14,82	18,5	17,9	13,2	13,44	14,66	10,72	9,9
DWT-D2	10,64	12,22	19,38	14,88	13,02	14,96	14,38	19,38	11,12
DWT-D3	4,98	5,32	8,42	3,68	4,88	4,68	5,98	6,2	4,7
DWT-D4	2,02	7,88	18,52	6,94	8,74	7,86	9,2	6,06	5,2
DWT-D5	6,16	10,22	29	3,2	8,96	7,84	8,74	9,32	7,24
DWT-Fusion	4,32	2,9	2,4	1,64	1,76	1,8	1,86	3,38	2,18
FILT-Fusion	4,6	0,2	2,04	0,9	2,1	1,54	3,08	5,56	2,04
FILT-Y0	6,4	22,46	66,14	13,36	15,78	14,56	18,72	9,66	8,56
FILT-Y1	4,58	8,16	18,88	3,72	5,92	5,5	9,8	9,24	3,84
FILT-Y2	3,96	5,2	11,16	4,02	5,86	5	6,78	4,36	5,7
FILT-Y3	4,04	5,2	11,66	5,32	6,92	6,08	7,82	7,22	7,04
FILT-Y4	5,02	11,42	14,76	14,5	12,3	12,02	12,84	5,36	5,28
SSA-Fusion	3,42	6,7	6,2	2,82	3,48	3,22	3,36	4,52	3,6
SSA-rr	3,84	11,6	15,08	11,54	11,2	11,8	11,66	8,8	10,18
SSA-ry	6,42	12,44	17,28	15	13,1	10,74	10,44	7,84	8,6
SSA-yr	4,2	10,2	13,56	8,22	10,08	8,94	9,12	4,84	5,82
SSA-yy	9,42	8	21,28	11,28	12,9	13,02	13,04	8,8	9,74

TAB. 5.10 – Coûts de classification pour différents classifieurs et ensembles d'indicateurs

Nombre de cycles de validation croisée à 10 blocs			
	10	20	30
Coût	0,26	0,328	0,312
Taux de classification	99,04	98,90	98,95

TAB. 5.11 – Sensibilité du coût et du taux de classification à la variation du nombre de cycles

5.9 Conclusion

Ce chapitre a montré la possibilité de détection très précoce d'une avarie sur la bague intérieure d'un roulement. La démarche du traitement de signal utilise l'extraction d'un signal enveloppe par valeur efficace d'une fenêtre exponentielle glissante. Le signal est d'abord préfiltré au moyen de trois méthodes : l'analyse spectrale singulière, le filtrage et l'analyse en ondelettes. La méthode de détection proposée et basée sur un traitement adéquat du signal et sur la combinaison de l'ACP et d'un arbre de décision a montré son efficacité.

En ce qui concerne l'identification, le réseau de neurones s'est imposé comme meilleur classifieur. L'ensemble d'indicateurs construit par fusion des données issues du prétraitement par filtrage a été retenu comme meilleur en terme de coût et de taux de classification, suivi de l'ensemble d'indicateurs construit par fusion de données issues du prétraitement par analyse en ondelettes.

Références

[1] Y. Li, C. Zhang, T. Kurfess, S. Danyluk et S. Liang : Adaptative prognostics for rolling element bearing condition. *Mechanical Systems and Signal Processing*, 13(1):103–113, 1999.

[2] N. Gebraeel, M. Lawley, C. R. Liu et Parmeshwaran : Residual life predictions from vibration-based degradation signals : A neural network approach. *IEEE Transactions on Industrial Electronics*, 51:694–700, 2004.

[3] J. P. Dron, F. Bolaers et L. Rasolofondraibe : Optimisation de la détection de defauts de roulements par débruitage des signaux par soustraction spectrale. *Mécanique et Industries*, 4(3):213–219, 2003.

[4] J. P. Dron, F. Bolaers et l. Rasolofondraibe : Improvement of the sensitivity of the scalar indicators (crest factor, kurtosis) using a de-noising method by spectral subtraction : application to the detection of defects in ball bearings. *Journal of Sound and Vibration*, 270(1-2):61–73, 2004.

[5] X. Chiementin, F. Bolaers, L. Rasolofondraibe et J.-P. Dron : Localization and quantification of vibratory sources : Application to the predictive maintenance of rolling bearings. *Journal of Sound and Vibration*, doi :10.1016/j.jsv.2008.02.029, 2008.

[6] J. Qiu, C. Zhang, B.B. Seth et S.Y. Liang : Damage mechanics approach for bearing lifetime prognostics. *Mechanical Systems and Signal Processing*, 16(5):817–829, 2002.

[7] FAG : Rolling bearing damage. recognition of damage and bearing inspection. Publ. No. WL 82 102/2 EA.

[8] C. Pachaud et A. Boulanger : *Analyse vibratoire en maintenance : Surveillance et diagnostic des machines*. Dunod, Paris, seconde édition, 2003.

[9] N. Tandon et A. Choudhury : A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings. *Tribology International*, 32(8):469–480, 1999.

[10] H.A. Mohl : Rolling bearing localised defect detection through vibration envelope analysis. *Sweden, SKF Engineering Research Center BV*, 2000.

[11] Randal R.B. : *Frequency analysis*. Brüel & Kjaer, 1987.

[12] J.F. MacGregor, T. Kourti et P. Nomikos : Analysis, monitoring and fault diagnosis of industrial processes using multivariate statistical projection methodes. *In Proceedings of IFAC Congress*, volume M, pages 145–150, San Francisco, 1996.

[13] H. Mohamed-Faouzi : *Détection et Localisation de Défauts par Analyse en Composantes Principales*. Thèse de doctorat, Institut National Polytechnique de Lorraine, 2003.

[14] R. Dunia et S.J. Qin : Joint diagnosis of process and sensor faults using principal component analysis. *Control Engineering Practice*, 6:457–469, 1998.

[15] N. Baydar, Q. Chen, A. Ball et U. Kruger : Detection of incipient tooth defect in helical gears using multivariate statistics. *Mechanical Systems and Signal Processing*, 15(2):303–321, 2001.

[16] P. Ralston, G. DePuy et J.H. Graham : Computer-based monitoring and fault diagnosis : a chemical process case study. *ISA transactions*, 40:85–98, 2001.

[17] Masayuki Tamura et Shinsuke Tsujita : A study on the number of principal components and sensitivity of fault detection using pca. *Computers and Chemical Engineering*, 31: 1035–1046, 2007.

[18] V. Sugumaran et K.I. Ramachandran : Automatic rule learning using decision tree for fuzzy classifier in fault diagnosis of roller bearing. *Mechanical Systems and Signal Processing*, 21(5):2237–2247, 2007.

[19] S. Al-Dossary, R.I Raja Hamzah et D. Mba : Observations of changes in acoustic emission waveform for varying seeded defect sizes in a rolling element bearing. *Applied Acoustics*, doi :10.1016/j.apacoust.2008.01.005, 2008.

[20] Abdullah M. Al-Ghamda et David Mba : A comparative experimental study on the use of acoustic emission and vibration analysis for bearing defect identification and estimation of defect size. *Mechanical Systems and Signal Processing*, 20:1537–1571, 2006.

[21] A. Choudhury et N. Tandon : Application of acoustic emission technique for the detection of defects in rolling element bearings. *Tribology International*, 33:39–45, 2000.

[22] M.J. Ashby et W.J. Scheuren : Intelligent maintenance advisor for turbine engines. *Proceedings of 2000 IEEE Aerospace Conference*, 6:211–219, 2000.

[23] Ian H.Witten et Eibe Frank : *Data mining : Practical machine learning tools and techniques*. Morgan Kaufmann, San Francisco, 2nd édition, 2005.

Chapitre 6

Surveillance d'usure d'outils en usinage

Ce chapitre étudie les possibilités de détection et d'identification de l'usure d'outils en usinage. Après avoir présenté diverses méthodes de surveillance d'usure d'outils, nous illustrons les méthodologies d'analyse intelligente de données par deux études expérimentales. La première étude est réalisée en tournage et démontre la possibilité de suivre l'usure. La seconde, réalisée en fraisage, montre la possibilité de construire un indicateur qui capture l'évolution temporelle de l'usure.

6.1 Introduction

La minimisation des coûts se présente comme une exigence incontournable dans tous les secteurs de production. En usinage, une des composantes critiques du coût est intimement liée à l'état de l'outil et à sa durée de vie.

L'usure d'outil est un phénomène inévitable puisque inhérent à tout processus de coupe vu les niveaux de contraintes, de frottement et de température auxquels l'outil est soumis. Cependant, usiner avec un outil usé peut non seulement affecter les tolérances, mais aussi la machine-outil. L'usure de l'outil mène à une vibration indésirable qui entame l'état de surface et la précision dimensionnelle de la pièce. Dans le plus mauvais des cas, on peut même assister à une rupture d'outil lors de son engagement dans la pièce [1].

La capacité de statuer sur l'état de l'outil et de prédire sa vie résiduelle s'impose donc, d'une part pour une gestion rationnelle du temps de production et, d'autre part pour garantir la qualité des pièces usinées en terme de respect des tolérances et d'état de surface. Il y a donc un réel besoin de méthodes permettant la détection de l'apparition de l'usure et le suivi de son évolution

dans le cadre d'une politique « juste à temps » de changement d'outil [2, 3].

Le développement d'un système de surveillance passe par deux principales, après le choix de la méthode de surveillance, donc du type de capteur à utiliser [4]. Il s'agit de la procédure d'extraction des indicateurs des signaux bruts mesurés, et de la classification et du développement d'une stratégie de décision [5].

Ce chapitre va, dans un premier temps, être consacré à la description de mécanismes d'usure d'outils, de méthodes de surveillance de l'usure et des critères associés à l'usure. Par la suite, deux études appliquées de surveillance vibratoire d'outils seront menées en tournage et en fraisage.

6.2 Mécanismes et modes d'endommagements des outils

Les outils de coupe sont soumis à des contraintes superficielles très élevées, à des températures élevées, au glissement du copeau sur la surface de coupe et au glissement de la surface de dépouille principale sur la surface usinée. Ces conditions induisent l'usure de l'outil.

Quelle que soit la technique d'usinage employée, un outil comprend (figure 6.1) :

- une ou plusieurs zones tranchantes que l'on peut décrire par des dièdres élémentaires définis par des faces A_α (face de dépouille) et A_γ (face de coupe) et une arête d'intersection ;
- des zones de guidage du copeau soit pour faciliter son évacuation, soit pour le stocker provisoirement ;
- des zones assurant la résistance mécanique de l'outil.

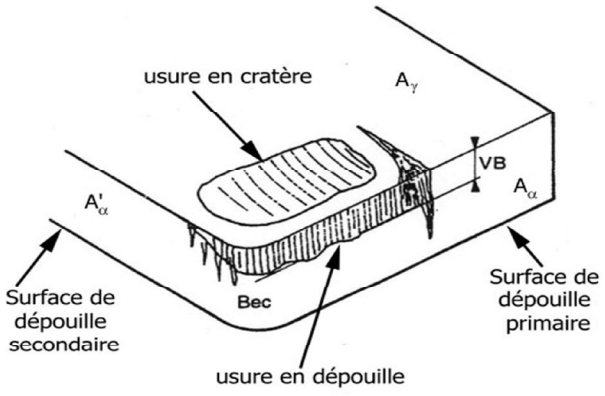


FIG. 6.1 – Usure de l'outil

La formation du copeau implique le triple frottement sec de deux corps pressés l'un sur

l'autre : le couple copeau-outil (face de coupe A_γ), le couple pièce-outil (face de dépouille A_α) et le couple pièce-outil (face A'_α). Les conditions de température et de confinement dans lesquelles se produisent ces trois types de contact sont différentes (tableau 6.1).

	Température	Confinement
A_γ	haute	sous l'effet l'effort de coupe
A_α	modérée	semi-confinement
A'_α	modérée	sans

TAB. 6.1 – Condition des contacts copeau/outil/pièce

Lorsqu'on examine des outils de compositions différentes (acier rapide, carbures, cermets, céramiques,...), on remarque des faciès d'usure très variés. Certains sont cassés (rupture d'arête), d'autres sont écaillés. La plupart ont une géométrie qui a évolué quasi continûment (outils usés). Mis à part les ruines d'outils par écaillages et micro-écaillages, on peut observer que chacun de ces couples donne lieu plus ou moins rapidement à trois faciès d'endommagement typiques (figures 6.1 et 6.2) :

1. le creusement d'un cratère sur la face A_γ , dont la profondeur maximale est couramment désignée par KT ; la résistance de l'outil diminuant lorsque KT augmente ;
2. une plage d'usure frontale plane sur la face A_α en dépouille principale, dont la hauteur moyenne est couramment désignée par le symbole VB. Quand VB croît, l'arête de coupe recule et les cotes s'écartent de la valeur visée ;
3. le creusement d'un sillon sur la face A_α en limite de la zone de coupe et d'une multitude de petits sillons sur la face A'_α. L'état de la surface usinée est d'autant plus détérioré que ces sillons sont plus développés.

Trois modes d'endommagement contribuent à ces différents faciès d'usure : usure adhésive, usure abrasive et usure par diffusion. Leur efficacité relative dépend de la température de l'interface et par conséquent de la vitesse de coupe.

6.2.1 Usure adhésive

La surface interne du copeau qui frotte sur l'outil est vierge, exempte de toute pollution. Il en dévient de même de la surface de l'outil nettoyée par le copeau. Les conditions sont donc favorables pour réaliser des soudures par friction après écrasement des aspérités des deux antagonistes, à moins qu'un film lubrifiant ne vienne s'interposer. Compte tenu des pressions de contact et des vitesses relatives effectives, du confinement de la zone de coupe, l'éventualité

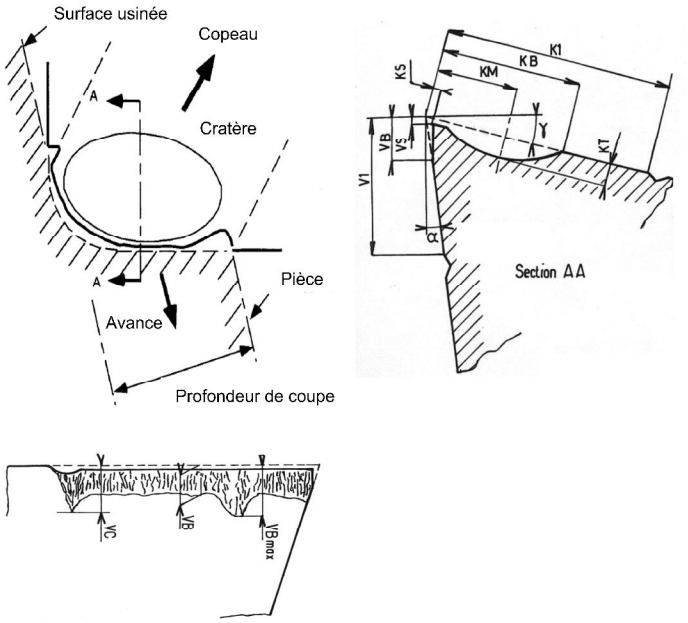


FIG. 6.2 – Usure typique d'un outil au carbure

d'une lubrification externe hydrodynamique est très faible. Dans ce cas, les aspérités des deux antagonistes se soudent, mais le défilement du copeau impose de rompre en continu les jonctions qui se forment.

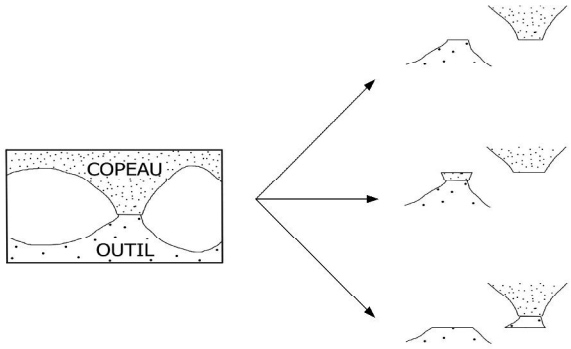


FIG. 6.3 – Cisaillement d'une jonction due à une adhésion copeau-outil

Trois situations peuvent être rencontrées (figure 6.3) :

– la plupart des soudures qui se constituent n'ont aucune résistance mécanique ; il y a rupture de la soudure à l'interface copeau-outil où elle s'est créée sans arrachement de matière. On rencontre ce cas favorable lorsqu'il est possible de provoquer une pollution des surfaces en contact. Des matériaux tels que le laiton, les aciers, etc. peuvent être élaborés

de telle manière qu'ils contiennent des particules métalliques de métaux à bas point de fusion (plomb, bismuth, etc.) qui jouent un rôle de poison des soudures de contact ; les sulfures peuvent avoir une efficacité équivalente ;

– rupture de la jonction côté du matériau le plus doux c'est-à-dire le copeau ; ces fragments très écrouis se déposent sur l'outil et peuvent rapidement constitué un dépôt macroscopique. Cela provoque une évolution de la zone de coupe (perte d'acuité de l'arête) avec, éventuellement création d'une arête rapportée. En règle générale, pour un couple outil-matériau usiné, le dépôt a un volume qui s'accroît puis décroît lorsque la vitesse de coupe augmente . Le mécanisme de l'adhésion peut être actif à toute vitesse de coupe, mais ne conduit à des dépôts macroscopiques qu'aux vitesses basses (ex. brochage) ou moyennes. Très souvent, des morceaux des dépôts macroscopiques sont évacués, ce qui peut entraîner un régime vibratoire : l'outil se cambre, puis se relaxe. On observe une spectaculaire altération de la qualité des surfaces usinées et un endommagement mécanique de l'outil soumis à des surcharges importantes ;
– rupture de la jonction côté outil ; il y a entraînement progressif de particules de ce dernier par le copeau.

6.2.2 Usure abrasive

L'usure abrasive résulte de l'arrachement de micro-particules des outils de coupe par l'impact de constituants ayant une dureté équivalente ou supérieure à la leur. Ces débris peuvent provenir du matériau usiné, de grains de l'outil arrachés par rupture des microsoudures, de particules du matériau usiné oxydées à l'air ou de morceaux d'arêtes rapportées évacuées, etc...

6.2.3 Usure par diffusion

Les températures très élevées atteintes au contact copeau-outil, peuvent provoquer des phénomènes de diffusion entre les surfaces non polluées de l'outil et de la matière usinée. Ils ont un effet notable si certains éléments chimiques présents dans un antagoniste (l'outil ou le copeau) ont une forte solubilité et une faible concentration dans le deuxième antagoniste ce qui est fréquemment le cas notamment lorsque l'on usine des matériaux ferreux avec des outils au carbure de tungstène fritté dont les grains sont liés par du cobalt ; on constate dans cette situation une diffusion du cobalt de l'outil dans les copeaux et une diffusion de fer dans l'outil qui se substitue au cobalt.

Il se superpose souvent à l'usure par diffusion une pollution chimique et une abrasion mécanique de la surface polluée notamment dans le cas des outils revêtus de couches minces (<10μm) de matériaux céramiques condensés sur leur surface (carbure de titane, nitrure de titane, alumine, etc.). Ces couches servent notamment d'efficaces barrières anti-diffusion qui sont, en raison de leurs faibles épaisseurs, très sensibles à l'abrasivité des copeaux. Dès qu'elles sont percées, le substrat sous-jacent est soumis à des phénomènes de diffusion.

D'autres modes d'endommagement existent : usure par fatigue thermique, usure par fluage, rupture brutale, rupture par fatigue mécanique, oxydation, etc. [6].

6.3 Critères associés à l'usure de l'outil

Pour pouvoir décider de l'état usé ou non de l'outil, des critères quantitatifs doivent être établis. Deux approches sont possibles :

- on peut choisir de lier ces critères à la dégradation dimensionnelle des pièces et à la qualité de leur surface. Un état d'usure est alors déclaré si l'on s'éloigne d'une certaine zone de tolérance pour les dimensions et l'état de surface. Cette approche nécessite de mesurer les dimensions et l'état de surface de chaque pièce usinée ;
- la seconde approche définit des critères liés à l'usure mesurée directement sur l'outil.

On préfère souvent la seconde approche à la première. Le critère le plus utilisé est basé sur la mesure de l'usure en dépouille (VB), parce qu'elle est le type d'usure qui influence le plus la qualité de la surface générée. Une mesure standard de la durée de vie d'outil de coupe, telle que proposée par les normes NF E66-505 et ISO 3685, est le temps nécessaire au développement d'une usure moyenne en dépouille VB égale à 300 μm quand l'usure est relativement uniforme et facile à mesurer, donc principalement causée par abrasion. Par contre, lorsque l'usure présente un profil irrégulier, l'usure moyenne en dépouille n'est plus un critère de durée de vie de l'outil, on utilise alors VBmax=600 μm (figure 6.2). Cependant, d'autres sources prétendent que fixer l'usure maximale à un VB donné peut conduire à de graves mécomptes économiques, car des plaquettes d'ébauche d'épaisseur et de dimensions d'arêtes importantes supportent des VB allant jusqu'à 1 mm tout en restant dans un régime de fonctionnement admissible [7].

Contrairement à l'usure en dépouille, celle en cratère n'influence pas la finition de la surface. Elle peut, cependant, sous certaines conditions thermiques, résulter en une défaillance brutale de l'outil. Une profondeur de cratère KT allant de 0,05 mm à 0,1 mm est considérée comme un critère de fin de vie de l'outil (figure 6.2).

6.4 Les méthodes de surveillance d'usure d'outils

Les méthodes utilisées pour la surveillance d'usure d'outils sont généralement classées en deux groupes : les méthodes directes et les méthodes indirectes.

On distinguera, dans ces deux groupes, les méthodes *on-line*, c'est-à-dire réalisées pendant le processus de coupe, des méthodes *off-line*, où l'on interrompt le processus pour procéder au contrôle.

6.4.1 Les méthodes directes

Dans les méthodes directes, l'état de l'arête de coupe ou d'un élément en contact avec la surface usée est directement mesurée. Les techniques les plus utilisées sont :

Capteurs de proximité

L'usure de l'outil est estimée en mesurant la variation de la distance entre l'arête de l'outil et la pièce. Cette mesure se fait à l'aide de palpeurs électriques ou pneumatiques. La technique se heurte à une limitation due à la dilatation thermique de l'outil ou aux vibrations de la pièce et la flexion de l'outil causée par l'effort de coupe.

Radioactivité

On implante une petite quantité de matière radioactive sur la face de l'outil. Pendant la coupe, les particules d'usure de l'outil sont transférées au copeau et par une mesure de la radioactivité de ce dernier, le niveau d'usure est évalué [8].

Résistance électrique

La résistance à la jonction pièce/outil est évaluée à l'aide d'un voltmètre et ses valeurs sont corrélées à l'usure de l'outil. Cette technique peut être appliquée en cours de processus d'usinage, mais des questions se posent sur l'efficacité de son application (résistances en série,...).

Capteurs optiques et vision artificielle

Ces capteurs reposent sur le fait que les surfaces usées sont plus réfléchissantes que les surfaces non usées. Cette technique est utilisée pour évaluer la forme ou la position de l'arête de coupe [9].

Beaucoup de recherches se sont penchées sur les méthodes directes basées sur la vision artificielle, laser ou vidéo [9, 10]. D'une manière générale, les méthodes directes ne se sont pas avérées économiquement et techniquement consistantes pour être appliquées industriellement en cours de processus d'usinage. Par exemple, du point de vue technologique la variation de niveau d'illumination rend difficile l'implémentation des méthodes basées sur la vision artificielle. Le palpage de l'outil, qui est une méthode directe, ne peut en aucun cas être imaginé appliqué *on-line*. La présence du fluide de coupe complique aussi l'utilisation des méthodes directes en cours de processus.

6.4.2 Les méthodes indirectes

Ces méthodes, contrairement aux méthodes directes, procèdent par une évaluation de l'usure sur base des paramètres mesurées pendant le processus de coupe : les efforts de coupe, l'émission acoustique ou les vibrations. Les méthodes indirectes sont en général des méthodes *on-line*.

Efforts de coupe et puissance absorbée

L'utilisation des efforts de coupe comme indicateur de l'usure d'outil est une pratique courante [11]. En effet, il est établi que les efforts de coupe croissent avec l'usure de l'outil. Cela est dû au changement des paramètres de coupe. Un dynamomètre est monté sur le porte-outil afin de mesurer les efforts de coupe suivant des directions orthogonales.

Cependant, les dynamomètres présentent l'inconvénient d'être chers ($\approx$ 50 k€) et d'avoir une gamme de fréquence limitée : une fréquence naturelle f_0 (pour la partie piézoélectrique) d'environ 3 kHz, et une fréquence de résonance f_r autour de 1 kHz quand ils sont rigidement fixés [12].

La puissance électrique absorbée est souvent une mesure assez fidèle de la dégradation de l'état de l'outil. Son évolution est une image de l'évolution des efforts de coupe [2].

Vibrations

Les vibrations sont de loin le type de signal le plus utilisé dans le *condition monitoring* des machines tournantes. Mais elles n'ont pas réussi à faire l'unanimité dans le domaine de l'usure d'outils de coupe, cela étant principalement dû aux autres sources « parasites » de vibrations. Il est pourtant évident qu'usiner avec un outil usé fait accroître les fluctuations des efforts sur l'outil ; ce qui, naturellement, l'excite en vibration. Les recherches qui ont utilisé les vibrations

dans le but de la surveillance d'usure d'outils sont toutes arrivées à la conclusion qu'avec un procédé approprié d'extraction d'indicateurs pertinents du signal vibratoire, on arrive à des capacités de détection satisfaisantes [13–15].

Émission acoustique et vibration ultrasonique

Pendant l'usinage, la pièce est le siège de déformations plastiques. L'émission acoustique est définie comme l'énergie élastique transitoire libérée dans les zones de déformation [16]. D'autres sources d'émission acoustique sont les transformations de phases et les mécanismes de fissuration [17].

En perçage avec des outils de très faibles diamètres, où la surveillance par efforts de coupe et courant du moteur n'était pas applicable vu leurs niveaux très faibles, l'utilisation des émissions acoustiques s'est trouvée justifiée [18].

Autres mesures indirectes

En dehors de ces principaux signaux utilisés pour la surveillance indirecte de l'usure d'outils, d'autres types de mesures ont été testés dans divers processus d'usinage. Ces signaux ont été souvent utilisés en fusion avec d'autres. Ghosh *et al.* [1] ont combiné les mesures de courant et de tension du moteur avec les efforts de coupe et les vibrations en fraisage. Dans [19], Axinte et Gindy, qui ont travaillé sur le brochage, ont en plus des émissions acoustiques, des vibrations et des efforts de coupe fait usage de la pression dans le circuit hydraulique de la machine. Dans [20] et [21], seul le courant consommé par le moteur a constitué le signal de surveillance. La figure 6.4 synthétise la procédure de mise en œuvre des méthodes indirectes.

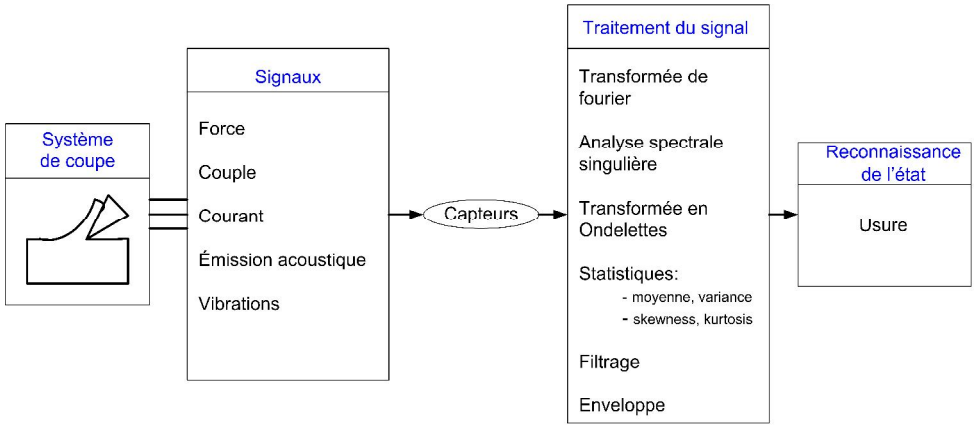


FIG. 6.4 – Méthodes indirectes de surveillance d'usure

Les méthodes indirectes peuvent être économiquement acceptables, mais comme dans tout problème de surveillance, le plus grand défi reste la définition des indicateurs qui puissent suivre fidèlement le phénomène surveillé, traduire et signaler suffisamment tôt les signes d'une avarie.

Les signaux relevés dans les méthodes indirectes contiennent, en plus de l'information sur l'usure de l'outil, d'autres contributions dues aux conditions de coupe et à la dynamique de la machine-outil. Ces contributions parasites sont plus marquées avec les signaux vibratoires et acoustiques, cas des phénomènes d'instabilités [22]. Elles présentent une difficulté en plus du fait que les signaux de surveillance peuvent être non-stationnaires, et que les relations entre les indicateurs extraits des signaux mesurés et l'état de l'outil ne sont pas linéaires [23]. Il est, dès lors, utile de procéder, par un traitement de signal adéquat, à une séparation des composantes du signal afin d'isoler celles directement corrélées à l'usure. Dimla [24] a essayé de caractériser l'impact des conditions de coupe (vitesse de coupe, avance et profondeur de coupe) sur les signaux d'efforts de coupe et vibratoires relevés dans un *monitoring* d'outils. Ses travaux ont insisté sur la nécessité, pour un système de surveillance d'outils, d'être capable de distinguer les variations dans le signal dues aux conditions de coupe de celles dues à l'usure.

6.5 Modèle pour l'évaluation de l'usure d'outils

En vue de l'évaluation de la durée de vie d'un outil de coupe, il est essentiel de disposer d'un modèle d'estimation de l'usure de l'outil. Le modèle de Taylor, qui est historiquement le premier modèle établi pour l'estimation de la durée de vie de l'outil, n'est pas optimal parce qu'il demande un grand nombre de tests pour l'estimation de ses coefficients. Ces dernières années ont vu se développer des modèles non linéaires d'estimation directe de l'usure et de son taux de variation avec le temps [6]. De manière générale, ces modèles se présentent sous la forme d'équations d'état non linéaires qui associent un vecteur $\mathbf{x}$, caractéristique des conditions de fonctionnement, à un vecteur $\mathbf{w}$ reprenant les différents types d'usure (VB, KT, VN,...). Le vecteur $\mathbf{x}$ reprend, en plus de la vitesse de coupe, de l'avance et de la profondeur de coupe, les paramètres géométriques de l'outil, la température et les contraintes sur l'outil. Le vecteur $\dot{\mathbf{w}}$ représente les taux d'usure.

$$\mathbf{w} = \mathcal{W}(\mathbf{x}) \tag{6.1}$$

$$\dot{\mathbf{w}} = \dot{\mathcal{W}}(\mathbf{x}) \tag{6.2}$$

Les opérateurs non linéaires $\mathcal{W}$ et $\dot{\mathcal{W}}$ lient l'usure et son taux à $\mathbf{x}$. Dans l'objectif de la surveillance de l'usure d'outil par des méthodes indirectes, $\mathbf{x}$ est augmenté de composantes issues d'une ou plusieurs des grandeurs suivantes : efforts de coupe, vibrations et émissions acoustiques. Ces grandeurs sont traitées dans le domaine temporel et/ou dans le domaine fréquentiel. En plus, on fait l'hypothèse d'une corrélation entre le signal relevé et l'usure.

Il existe des relations analytiques entre les formes de l'usure et les paramètres géométriques de l'outil, la température et la contrainte sur l'outil. Ces relations présentent en général un intérêt plutôt scientifique que pratique, vu les difficultés qui entourent la détermination des paramètres qu'elles font intervenir. Par exemple, les références [6] et [25] ont proposé, pour une usure combinée en dépouille par abrasion/adhérence et par diffusion, une relation globale de la forme

$$\frac{dw}{dt} = \frac{A}{H}\sigma_n V_s + B\ \exp\Big(\frac{-E}{RT_f}\Big) \tag{6.3}$$

avec H, V_s et σ_n respectivement la dureté de l'outil, la vitesse de glissement et la contrainte normale entre la face de dépouille de l'outil et la pièce. E, R et T_f sont l'énergie d'activation du processus, la constante des gaz parfaits et la température de coupe dans la zone de dépouille de l'outil.

La construction de l'opérateur non linéaire intervenant dans les relations 6.1 et 6.2 pose le problème du choix de la méthode de classification et du développement de la stratégie de décision. Cet opérateur peut provenir d'un modèle analytique [25], d'une régression non linéaire ou d'un modèle d'apprentissage (*machine learning*). Cette dernière possibilité est porteuse d'espoir étant donnée les possibilités offertes par certaines techniques d'apprentissage pour représenter des relations impossibles à modéliser analytiquement.

Les réseaux neuronaux ont été très utilisés seuls [14] ou en combinaison avec les modèles flous [2, 26, 27] ou les algorithmes génétiques [28]. Les arbres de décision et les réseaux bayésiens ont aussi fait l'objet d'une prospection dans le but de l'évaluation de l'état d'usure des outils de coupe [29].

6.6 Surveillance d'outils par signaux vibratoires

Beaucoup d'études tendent à démontrer l'incapacité d'un seul type de signal pour diagnostiquer l'état de l'outil. Ceci explique la tendance à la fusion d'indicateurs provenant de capteurs différents [1, 16, 19, 23, 26]. Des récents travaux ont montré que malgré la fusion, le défi demeure encore, dans le traitement du signal, de réaliser l'isolation de la contribution due à l'usure [12, 30] de celles dues à la cinématique de la machine, par exemple.

Dans cette thèse, nous nous sommes penchés sur les possibilités de tirer le meilleur parti du signal vibratoire dans le cadre de la surveillance d'usure d'outil.

Pour ce qui concerne l'extraction d'indicateurs à partir des signaux vibratoires, la littérature relève des approches très variées. Abu-Mahfouz [14] a extrait, pour la détection et la classification de l'état de l'outil en perçage à l'aide des réseaux de neurones, des indicateurs en procédant à la moyenne des coefficients d'une décomposition en ondelettes des vibrations. Obikawa *et al.* [31] ont aussi fait usage des coefficients d'ondelettes des signaux d'efforts de coupe.

Dans sa revue des méthodes indirectes de *monitoring* d'outils, Dimla [17, 32] cite plusieurs travaux qui ont analysé la corrélation entre l'état de l'outil et les vibrations. El-Wardany *et al.* [15] ont utilisé, dans le domaine temporel, le rapport entre le *kurtosis* et la valeur moyenne du signal vibratoire. Pour leur part, Jiang *et al* [33] ont montré que l'énergie dans certaines bandes de fréquence judicieusement choisies était sensible à l'évolution de l'état de l'outil. Diverses méthodes simples d'extraction d'indicateurs sont aussi rapportées comme ayant fourni de bons résultats, c'est par exemple le cas des valeurs moyennes dans des bandes de fréquence données [13]. Shi et Gindy [34] ont par contre trouvé une faible corrélation entre la valeur moyenne et l'écart type d'un signal avec l'état de l'outil. Cependant ils ont plutôt montré l'utilité d'une analyse en composantes principales pour l'extraction d'indicateurs quand deux capteurs sont utilisés suivant des directions perpendiculaires.

6.7 Classification d'états d'outils : application

A la section 3.8, nous avons introduit une étude sur la surveillance de l'usure en tournage (voir annexe B pour les détails). Nous nous sommes limité à illustrer la procédure d'extraction d'indicateurs. Nous n'avons pas justifié pourquoi l'analyse spectrale singulière (ASS) a été préférée comme traitement du signal. En effet, l'usinage conique du manchon induit une instationnarité dans les signaux. Il a donc fallu une méthode qui capture les singularités locales du

signal. Cependant dans la suite de cette étude, nous analyserons les capacités de reconnaissance d'usure d'un système basé sur des indicateurs du domaine temporel. Une discussion sur les résultats comparera les indicateurs définis à la suite de l'ASS de ceux définis directement dans le domaine temporel.

Nous poursuivons donc l'étude entamée pour l'identification automatique de l'état de l'outil de coupe à l'aide de méthodes de *machine learning*. Rappelons que l'usure de l'outil n'est pas caractérisée quantitativement, mais qualitativement comme indiqué dans le tableau 6.2 que nous reprenons ici.

Etat	Symbole	Effectif
Plaquette neuve ou comme neuve	N	22
Faible usure	U1	7
Usure assez avancée	U2	6
Bourrage copeau	BC	4
Passage à vide	V	4

TAB. 6.2 – Etat de l'outil et du processus de coupe pour les 43 essais

Quatre méthodes de classification ont été testées sur ce jeu de données :

1. réseaux bayésiens (RB) : 2 parents au maximum ;
2. k plus proches voisins (KNN) : le nombre de voisins a été fixé à 2. En effet, ce nombre est le plus faible des nombres optimaux et présente donc l'avantage de demander moins d'opérations de calcul (figure 6.5). L'utilisation de la pondération lors du choix de la classe préserve de l'indétermination due au nombre pair de voisins ;
3. arbres de décision (AdD) ;
4. réseaux de neurones (RN) : une couche cachée dont le nombre de neurones est déterminée par essai en démarrant avec un nombre égal à la moyenne du nombre de classes et du nombre de variables.

6.7.1 Sélection de variables pertinentes

Avant d'alimenter les classifieurs, les données sont réduites dans l'intervalle [0,1] par une normalisation linéaire, puis une sélection de variables pertinentes pour le problème est effectuée par enveloppe d'un classifieur à réseau de neurones. Cette sélection n'a retenu que 4 des 18 variables de départ :

– la variable RC1 4000-6000 Hz ;

– la variable RC3 2000-4000 Hz ;
– la variable RC5 4000-6000 Hz ;
– l'avance A.

Rappelons qu'une sélection d'indicateurs pertinents par un type classifieur ne garantit pas leur efficacité avec d'autres types classifieurs.

Comme le jeu de données n'est pas large, et afin d'améliorer le processus de classification, les ensembles d'apprentissage et de test ont été construits sur base de l'algorithme de validation croisée à 10 blocs (*10-fold cross validation*).

6.7.2 Résultats et discussion

Le processus de classification a été mené dans l'environnement «*Weka KnowledgeFlow*» [35]. Le tableau 6.3 donne les résultats globaux de classification. On y reprend le taux de classification sur l'ensemble complet des variables à côté de celui obtenu sur l'espace réduit ne contenant que les variables jugées pertinentes à l'issue de la sélection. La figure 6.6 illustre l'arbre de décision construit sur ces variables sélectionnées.

Nous avons observé une amélioration effective des résultats de classification dans le cas de l'espace réduit. Un autre bénéfice est la réduction du temps de calcul pour le processus de classification qui ne concerne plus que 4 variables au lieu de 18.

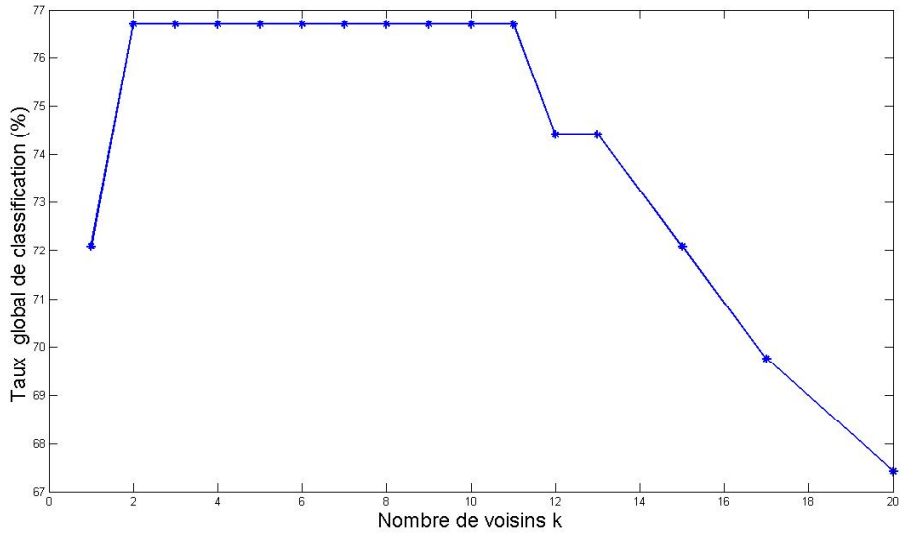


FIG. 6.5 – Evolution du taux de classification avec le nombre de voisins

Méthode	Taux de Classification	
	Espace complet	Espace réduit
Réseaux Bayésiens	53,5	55,8
KNN	67,4	76,7
Arbres de Décisions	62,8	76,7
Réseaux de Neurones	72,9	76,7

TAB. 6.3 – Taux de reconnaissance de l'état du processus (en %)

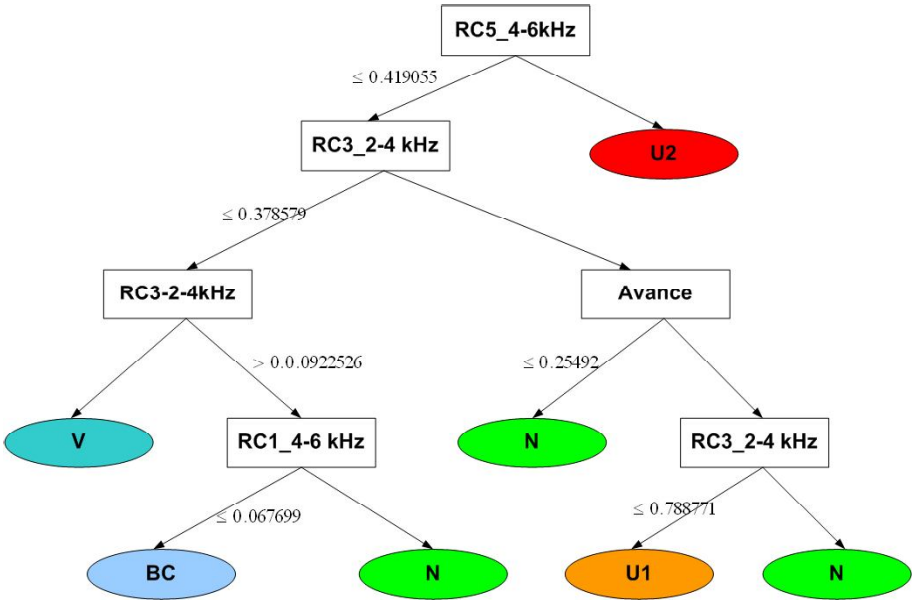


FIG. 6.6 – Arbre de décision

L'examen des détails du tableau 6.4, qui reprend les probabilités d'identification (*PI*) de chaque classe sur l'espace réduit, révèle de très bons résultats de reconnaissance de l'état de l'outil pour presque toutes les méthodes sauf pour les réseaux bayésiens qui ont mal reconnu beaucoup d'instances U1 en les classifiant comme N. Les *PI* des classes relatives à l'état du processus (BC et V) sont très faibles. Particulièrement le passage à vide a été classifié comme N. Ceci peut s'expliquer par le faible niveau de contraintes qui est également rencontré dans le cas d'un outil neuf. En ce qui concerne la reconnaissance du bourrage copeau, les matrices de confusion relatives aux différents classifieurs ont révélé que BC a été classifié comme N ou comme U2. Une explication à cela est évidente du fait que la condition BC ne caractérise pas l'état de l'outil. En effet, un bourrage copeau peut s'être produit lors d'une coupe avec un outil neuf ou usé. Cette observation a révélé la robustesse des indicateurs définis à l'issue de l'ASS vis-à-vis du bruit provenant du fouettement de copeau lors du bourrage.

	N	U1	U2	BC	V
Réseaux Bayésiens	72.7	42.9	66.7	25	0
KNN	90.9	85.7	83.3	50	0
Arbres de Décisions	72.7	100	100	50	50
Réseaux de Neurones	95.5	100	83.3	0	0

TAB. 6.4 – Probabilités d'identification (PI) de l'état de l'outil (en %)

6.7.3 Extension de l'étude par l'utilisation de tous les canaux

Les résultats qui viennent d'être présentés ont été obtenus avec les seules vibrations relevées dans la direction *Z* de l'effort de coupe. La figure 6.7 rappelle l'orientation des capteurs.

Analysons maintenant la contribution des autres directions de mesure à la reconnaissance de l'état de l'outil. Nous utilisons les mêmes classifieurs que dans l'étude précédente. Les résultats de classification (*PI* et *TC*) sont représentés dans les tableaux 6.5 à 6.7 pour les ensembles de données issus de différentes direction de mesure et pour leur fusion. Le tableau 6.8 synthétise les résultats obtenus avec les espaces réduits d'indicateurs.

	N		U1		U2		BC		V		Taux	
	Complet	Réduit	Complet	Réduit	Complet	Réduit	Complet	Réduit	Complet	Réduit	Complet	Réduit
RB	81,8	77,3	0	57,1	66,7	50	50	0	0	25	55,8	58,1
KNN	90,9	86,4	71,4	57,1	66,7	66,7	0	0	0	0	67,4	62,8
AdD	77,3	90,9	57,1	71,4	50	33,3	0	0	0	0	55,8	62,8
RN	90,9	77,3	71,4	57,1	66,7	100	0	0	50	25	72,1	65,11

TAB. 6.5 – Résultats de classification avec l'ensemble de données du canal X

FIG. 6.7 – Orientation des capteurs

	N		U1		U2		BC		V		Taux	
	Complet	Réduit	Complet	Réduit	Complet	Réduit	Complet	Réduit	Complet	Réduit	Complet	Réduit
RB	81,8	90,9	0	0	66,7	83,3	50	50	0	0	55,8	62,8
KNN	86,4	72,7	71,4	28,6	66,7	83,3	0	50	0	0	65,1	58,1
AdD	77,3	86,4	71,4	28,6	66,7	83,3	25	50	25	0	65,1	65,1
RN	72,7	77,3	71,4	0	66,7	83,3	0	50	25	0	60,5	55,8

TAB. 6.6 – Résultats de classification avec l'ensemble de données du canal Y

	N		U1		U2		BC		V		Taux	
	Complet	Réduit	Complet	Réduit	Complet	Réduit	Complet	Réduit	Complet	Réduit	Complet	Réduit
RB	81,8	68,2	28,6	42,9	50	66,7	25	25	25	0	58,1	53,5
KNN	90,9	77,3	85,7	100	83,3	83,3	0	75	0	25	72,1	76,7
AdD	77,3	81,8	42,9	42,9	100	100	0	0	25	50	62,8	67,4
RN	95,5	81,8	85,7	57,1	66,7	83,3	0	25	25	25	74,4	67,4

TAB. 6.7 – Résultats de classification avec l'ensemble de données fusionnées (X, Y et Z)

A l'observation des résultats issus des différents modèles avec les différents ensembles de données, on peut tirer la conclusion qu'effectivement la direction Z contient la plus riche information sur l'usure par rapport aux autres directions. Remarquons que la fusion n'a pas amélioré la classification. En effet, nous l'avions déjà souligné, une fusion avec des variables de faible contenu informationnel ne peut avoir qu'un effet négatif sur la classification.

Si l'on tient compte du risque de mauvaise classification c'est-à-dire du coût, on conviendra que c'est dans la reconnaissance des états U1 et U2 qu'il faudra admettre le moins d'erreurs. Et de ce fait, le choix d'utiliser un arbre de décision avec les données du canal Z s'imposera.

Classe	Classifieur	X	Y	Z	Fusion
N	RB	77,3	90,9	72,7	68,2
N	KNN	86,4	72,7	90,9	77,3
N	AdD	90,9	86,4	72,7	81,8
N	RN	77,3	77,3	95,5	81,8
U1	RB	57,1	0	42,9	42,9
U1	KNN	57,1	28,6	85,7	100
U1	AdD	71,4	28,6	100	42,9
U1	RN	57,1	0	100	57,1
U2	RB	50	83,3	66,7	66,7
U2	KNN	66,7	83,3	83,3	83,3
U2	AdD	33,3	83,3	100	100
U2	RN	100	83,3	83,3	83,3
BC	RB	0	50	25	25
BC	KNN	0	50	50	75
BC	AdD	0	50	50	0
BC	RN	0	50	0	25
V	RB	25	0	0	0
V	KNN	0	0	0	25
V	AdD	0	0	50	50
V	RN	25	0	0	25

TAB. 6.8 – Synthèse de résultats pour les espaces réduits

6.7.4 Approche de l'étude avec les indicateurs scalaires

Nous avons introduit l'analyse en composantes principales comme moyen de détection, parfois de diagnostic (section 5.7). L'ACP sera utilisée dans cette étude pour construire un indicateur de diagnostic. L'objectif est d'explorer les indicateurs statistiques du domaine temporel.

Comme le signal est caractérisé par une non-stationnarité du fait de l'évolution conique de la coupe, l'idée de le balayer avec des fenêtres glissantes doit demeurer. Des fenêtres exponentielles sont utilisées pour extraire des statistiques sur des segments de signal. Il s'agit de la valeur efficace, du *kurtosis*, du *skewness* et du facteur de crête. Les valeurs ainsi extraites vont constituer des vecteurs de variables. Ensuite 12 des 22 enregistrements avec outil neuf sont utilisés comme *base normale* afin de construire le modèle ACP. Les résidus calculés sur ce modèle vont constituer le vecteur d'indicateurs.

Une investigation menée sur les différents canaux de mesure, les différentes gammes de fréquence et sur les variables, a montré que le résidu issu de la concaténation de la valeur efficace et de l'inverse du facteur de crête du canal Z (gamme 4-6 kHz) donnait la meilleure ségrégation des classes. Son évolution est représentée sur la figure 6.8. La figure 6.9 montre que la procédure de construction du résidu n'est pas très affectée par le nombre d'essais choisis

dans la base normale.

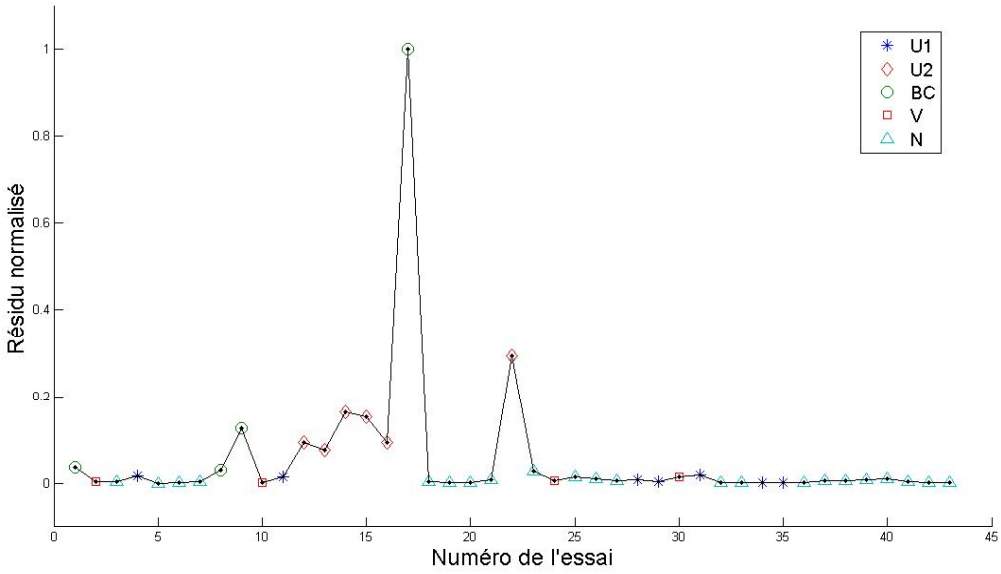


FIG. 6.8 – Évolution de l'indicateur en fonction du numéro de l'essai

Le tableau 6.9 confirme la conclusion à laquelle on peut s'attendre en observant la figure 6.8, c'est-à-dire seules les classes U2 et N peuvent bien se discriminer. On remarquera encore que l'arbre de décision se comporte mieux que les autres classifieurs. L'analyse des matrices de confusion a encore révélé que l'état U1 a été classé comme N, V comme N et BC comme U2.

	N	U1	U2	BC	V	Taux global
RB	95,5	0	100	0	0	62,8
KNN	100	0	83,3	25	0	65,11
AdD	90,9	0	100	75	0	67,4
RN	86,4	28,6	83,3	0	25	62,8

TAB. 6.9 – Résultats de reconnaissance avec l'indicateur Résidu ACP

6.7.5 Conclusion de l'étude

Cette étude ne s'est pas seulement intéressée à l'état de l'outil, mais aussi à d'autres dysfonctionnements du processus tels que le bourrage copeau ou un passage à vide. Les résultats obtenus sont encourageants pour l'utilisation des techniques de *machine learning* dans le do-

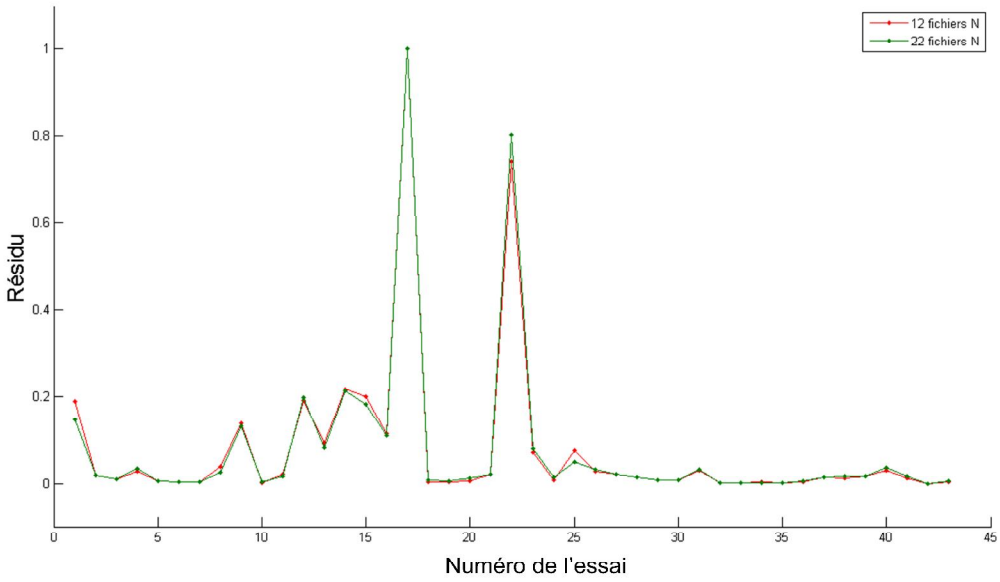


FIG. 6.9 – Comparaison des résidus issus des bases à 12 et à 22 enregistrements à outil neuf

maine de la surveillance d'usure d'outils. On devra, néanmoins, rappeler l'exigence de tenir compte du temps de traitement du signal si l'on veut implémenter un système *in situ*.

L'analyse a mis en évidence deux aspects importants : la forte richesse en informations de composantes hautes fréquences des signaux vibratoires et la robustesse des indicateurs définis qui prouve la possibilité de s'affranchir du bruit inutile par la combinaison de l'ASS et d'un filtrage passe-bande.

L'approche du problème par l'ACP avec des variables à très faible coût de calcul, a montré ses capacités de ne reconnaître l'usure que dans son stade avancé alors que l'ASS a permis de mieux détecter les états U1, donc le début de l'usure.

6.8 Suivi de l'évolution temporelle de la dégradation d'un outil de coupe

Cette seconde application en surveillance d'outils d'usinage adresse principalement le problème de la prise en compte de la structure temporelle des données de surveillance vibratoire. Le dispositif expérimental et les conditions des essais sont présentés à l'annexe C.

La représentation de l'usure à l'aide des modèles non linéaires a proposé la construction

de deux opérateurs (équations 6.1 et 6.2), l'un corrélant l'usure et l'autre son taux au vecteur d'indicateurs. Modéliser le taux d'usure tient du besoin d'évaluer l'évolution de l'usure avec le temps. Mais dans une approche *machine learning*, il est possible de ne construire qu'un seul modèle qui tienne compte de cet aspect. Pour permettre la capture de l'évolution temporelle de l'usure, nous proposerons d'abord l'introduction de la *dynamicité* dans la construction des indicateurs, puis l'utilisation de modèles d'apprentissage de type réseaux de neurones dynamiques.

Dans la section 4.10.2 de ce rapport, nous avons présenté deux méthodes pour la définition de variables qui capturent la dynamicité des données. Nous utiliserons dans cette étude la méthode basée sur les moyennes courantes.

6.8.1 Définition du symptôme vibratoire pour l'usure

A cette étape, nous trouvons utile de rappeler que l'extraction d'indicateurs est un processus qui dépend fortement du problème sous analyse. Les indicateurs retenus doivent être les plus riches en contenu informationnel et les moins coûteux en calcul.

Le problème que nous allons analyser diffère de la précédente étude de surveillance d'usure d'outil dans la structure des signaux enregistrés. En effet, ici le processus de coupe n'est pas conique comme dans le premier cas. Ce qui n'oblige pas un moyen de détection des singularités locales. En plus, dans la présente expérience, les données sont enregistrées dans leur chronologie de façon continue jusqu'à la dégradation totale d'un seul outil.

La figure 6.10 représente un spectre typique des signaux relevés pendant le suivi de l'évolution de l'usure. La configuration du contenu vibratoire va suggérer de mener l'étude dans les 5 bandes de fréquences suivantes en plus du signal global : ≤1, 1-2, 2-4, 4-7 et 7-12 kHz.

Nous proposons l'exploration des indicateurs simples du domaine temporel pour le suivi de l'usure. Ces indicateurs seront calculés dans ces différentes bandes fréquentielles.

De la figure 6.11, on sait voir que le spectre des signaux enregistrés pour le suivi de l'usure est principalement dominé par les harmoniques de la fréquence des dents de la fraise ($2 \times 14,7$ Hz, la fréquence de rotation étant de 14,7 Hz soit 885 tr/min). Du fait que l'initiation de l'usure se manifeste par de microchocs et que son accroissement induit une croissance du niveau vibratoire général, nous nous sommes proposé d'utiliser comme indicateurs le facteur de crête A_{fc} et la valeur efficace A_{rms} dans les gammes de fréquences considérées (figure 6.8.1).

Les variables considérées, A_{rms} et A_{fc} dans différentes gammes fréquentielles, sont représentées en fonction de la durée de coupe. Le temps mort entre deux passes étant assimilé ici au temps d'usinage, seul l'arrêt de coupe dû à l'opération de contrôle de l'outil n'est pas compté.

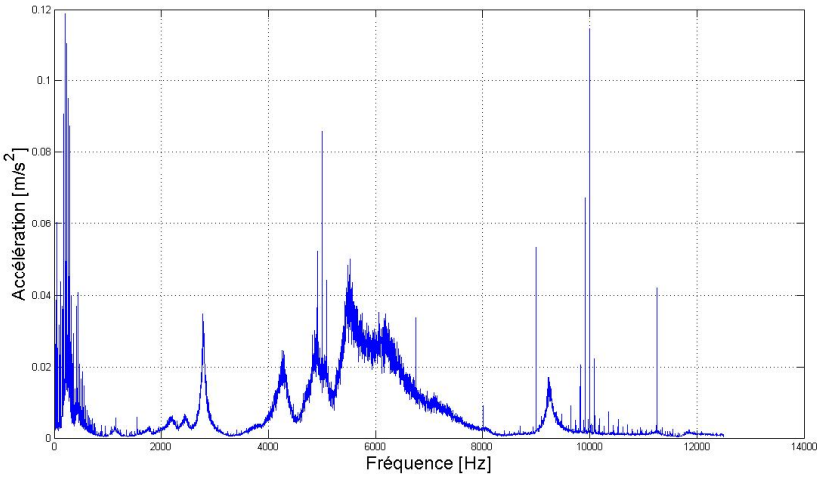


FIG. 6.10 – Spectre fréquentiel d'un signal typique de vibration

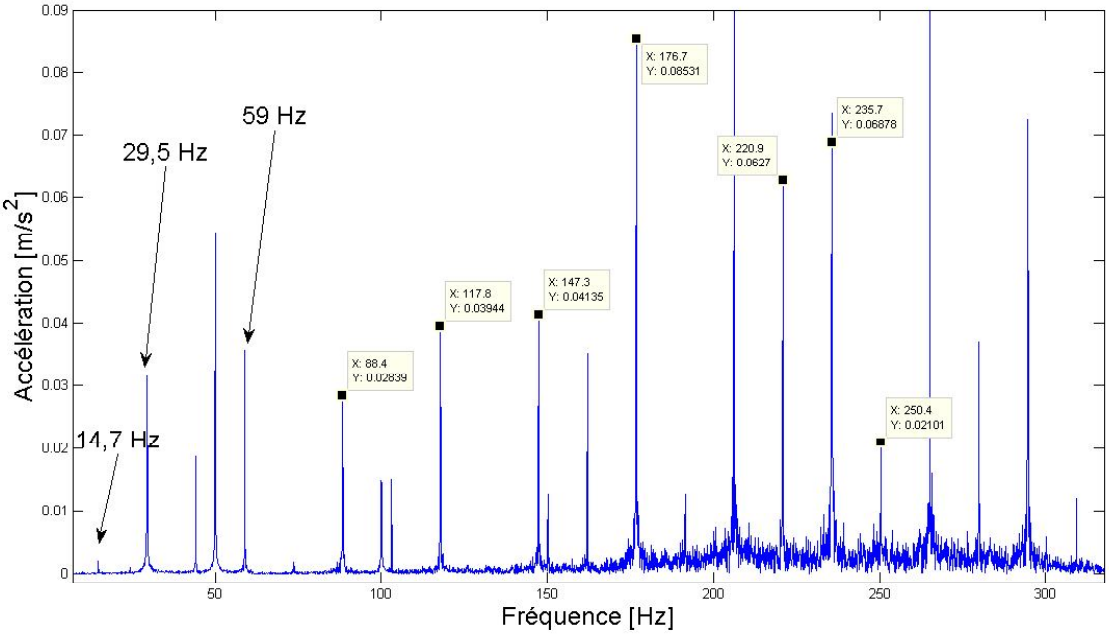


FIG. 6.11 – Zoom sur la zone basse fréquence du spectre

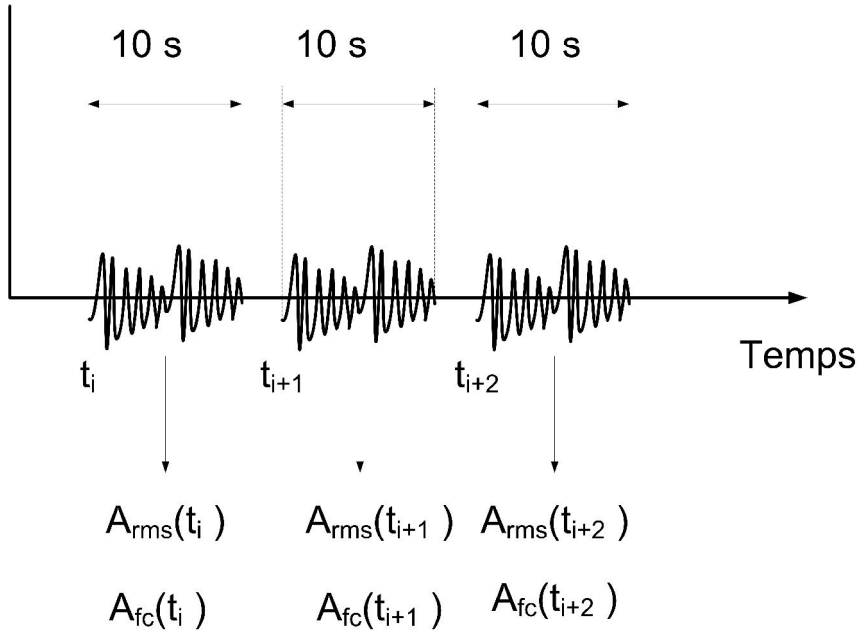


FIG. 6.12 – Extraction des indicateurs des signaux relevés

Pour tenir compte de la dépendance temporelle entre les N différentes mesures successives, de nouvelles variables ont été définies pour introduire « une mémoire temporelle » du passé de l'outil. Pour la valeur efficace, on calcule les moyennes courantes des valeurs A_{rms} dejà relevées. Pour le facteur de crête, la nouvelle variable est calculée à partir de la dernière plus grande valeur de pic enregistrée jusqu'à l'instant t_i. Donc si $x(t_\text{i})$ est le signal relevé à l'instant t_i (avec $t_\text{i} > t_{\text{i}-1}$, $i = 1...N$), $A_{\text{rms}}(t_\text{i})$ sa valeur efficace et $A_{\text{fc}}(t_\text{i})$ son facteur de crête, on détermine les nouvelles variables comme suit

$$Z_{\text{rms}}(t_\text{i}) = \frac{1}{i}\sum_{k=1}^{i} A_{\text{rms}}(t_\text{k}) \tag{6.4}$$

$$Z_{\text{fc}}(t_\text{i}) = \frac{\max_{k=1}^{i}\left(A_{\text{rms}}(t_\text{k}).A_{\text{fc}}(t_\text{k})\right)}{Z_{\text{rms}}(t_\text{i})} \tag{6.5}$$

En analysant l'évolution de la variable Z_{rms} dans les bandes de fréquence choisies (figure 6.13), on remarque une évolution croissante, donc cohérente avec l'avancement de l'usure dans toutes les gammes de fréquences sauf en basses fréquences (0-1 kHz) où cette variable accuse d'abord une forte chute avant de croître.

La figure 6.14 montre que les valeurs de Z_{fc} présentent des sauts qui sont probablement dus à la contribution des signaux correspondant à des chocs liés à une microrupture sur l'outil. Cette

nouvelle forme du tranchant de l'outil va provoquer un accroissement des forces d'excitation, donc une augmentation du niveau vibratoire.

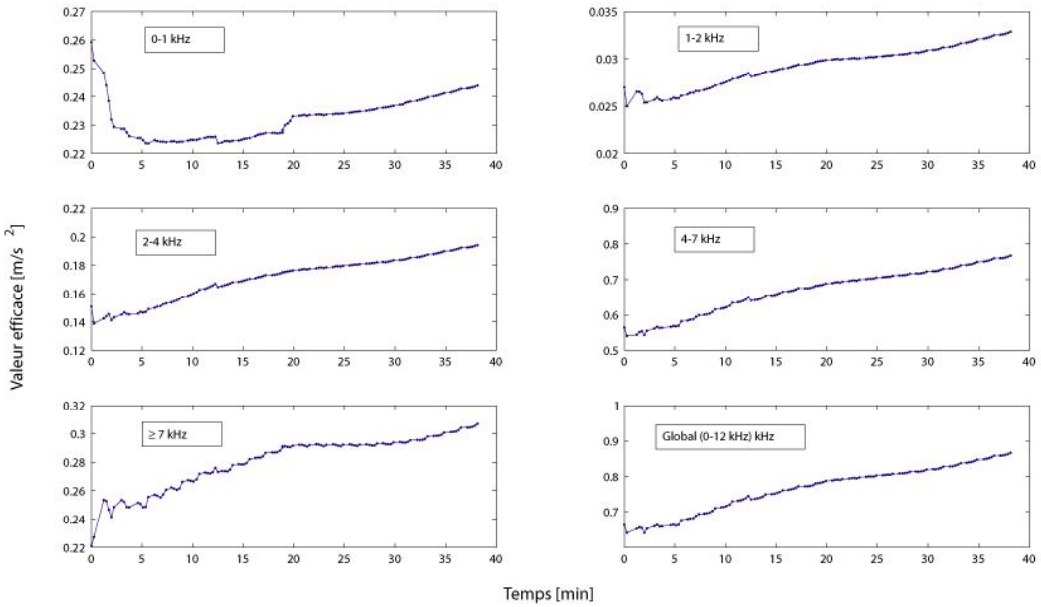


FIG. 6.13 – Évolution en fonction du temps des niveaux Z_{rms} par gamme de fréquence

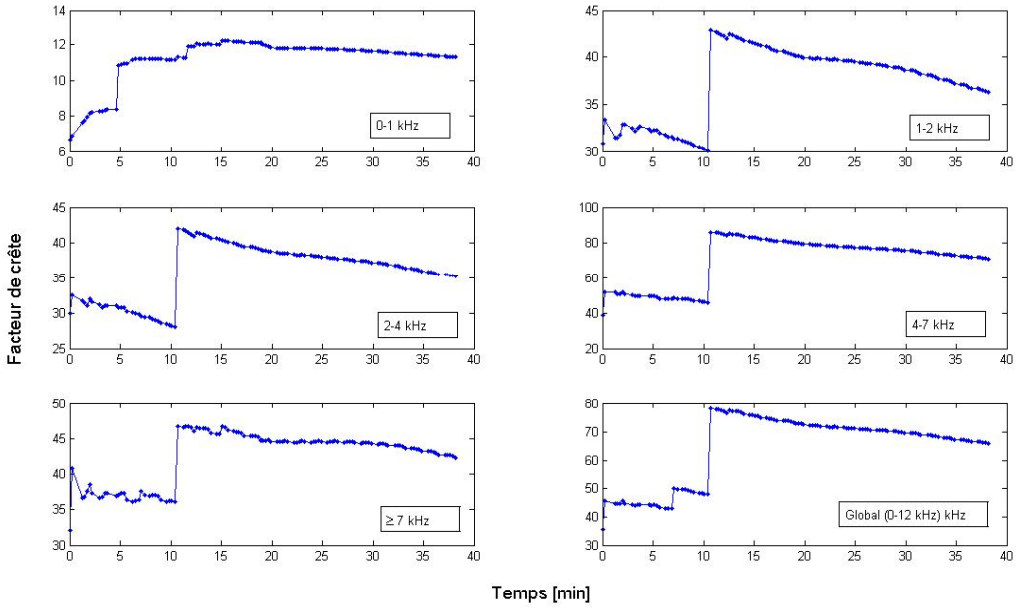


FIG. 6.14 – Évolution en fonction du temps de Z_{fc} par gamme de fréquence

En considérant les différentes gammes fréquentielles et en ajoutant les indicateurs calculés sur le signal global, nous aboutissons à 12 indicateurs pour chacun des 117 enregistrements. Le suivi de l'évolution de chaque indicateur n'étant pas très économique, la méthode basée sur l'extraction d'un résidu sur base de l'ACP semble être le meilleur choix (cf. section 5.7).

6.8.2 Application aux données expérimentales

La matrice de données est composée des variables Z_{rms} et Z_{fc}. Pour appliquer l'approche ACP aux données de l'usinage en fraisage, il va falloir d'abord choisir parmi toutes les données enregistrées celles qui peuvent être considérées comme représentant l'état normal de l'outil (base normale). Pour cela, la démarche suivante est adoptée :

- procéder une clusterisation en deux classes des données : N pour l'état normal et U pour l'état usé ;
- construire les résidus par ACP en prenant comme base normale la classe N, puis observer l'évolution de la valeur des résidus ;
- si au passage de la frontière entre les deux classes, cette valeur varie de manière abrupte, alors ajuster la frontière en avant ou en arrière jusqu'à avoir une allure continue pour le résidu. Le rejet de la variation abrupte permet de s'assurer le caractère progressif de l'usure.

La clusterisation en deux classes par la méthode de k-moyennes a montré que le premier cluster allait de l'enregistrement 1 à l'enregistrement 30 (après 10,4 minutes de coupe). En fixant la frontière de la base normale à 30 on aboutit à une variation abrupte de l'évolution des résidus, alors que si l'on bouge la frontière d'un pas, c'est-à-dire à 31 (10,7 minutes de coupe), on a un passage continu entre les deux classes. La figure 6.15 illustre cet effet.

La variable ainsi déduite de l'ACP et représentée à la figure 6.16 affiche un comportement qui suit fidèlement l'évolution temporelle de l'usure. On peut y distinguer trois zones avec des taux d'usure différents.

6.8.3 Synthèse

Cette étude vient de mettre en évidence un moyen de construction d'une variable qui suit l'évolution temporelle d'une dégradation. Elle suscite beaucoup d'intérêt pour l'extension des travaux en vue du pronostic (estimation de la durée de vie résiduelle). En effet, des méthodes d'apprentissage pourront être utilisées pour estimer la courbe de vie d'un équipement. Des sys-

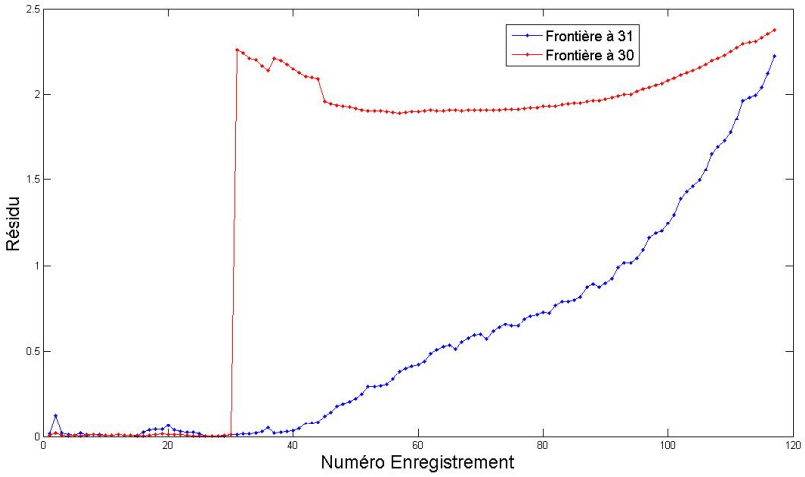


FIG. 6.15 – Évolution du résidu pour 2 frontières différents

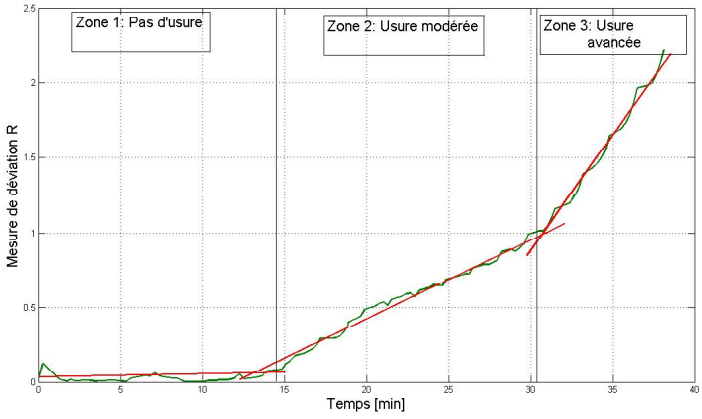


FIG. 6.16 – Evolution du taux d'usure

tèmes tels que les réseaux de neurones dynamiques (figure 6.17) peuvent être indiqués dans ce cas. L'intérêt de cette démarche est de pouvoir identifier une loi monotone croissante qui correspond à une dégradation progressive. Un telle loi présente beaucoup d'intérêt pour le pronostic.

Les réseaux de neurones dynamiques

Les réseaux de neurones qui ont été utilisés dans tous les exemples traités jusque là sont de type statique. Il ne tiennent pas compte de la dimension temporelle dans leur structure. Ce type de réseau s'avère inutilisable dès que l'objectif devient celui d'apprendre à partir des séquences temporelles de données. Si $\mathbf{x}(t)$ représente le vecteur d'indicateurs extraits du signal relevé à l'instant t, l'entrée du réseau de neurones peut être précédée d'une couche à délai qui stocke les vecteurs d'indicateurs $\mathbf{x}(t-d)\ \ d = 1, ...D$, où D est le délai maximum. L'entrée du réseau est donc composé de $D+1$ vecteurs. Ce type de réseau de neurones dont la structure fait intervenir la dynamique temporelle, sous forme de délai, dans la couche d'entrée d'un réseau statique à couches unidirectionnels est appelé *Focused Time-Delay Neural Network* (FTDNN) [36]. Il faut choisir un délai raisonnable de manière à avoir une taille acceptable du réseau de neurones.

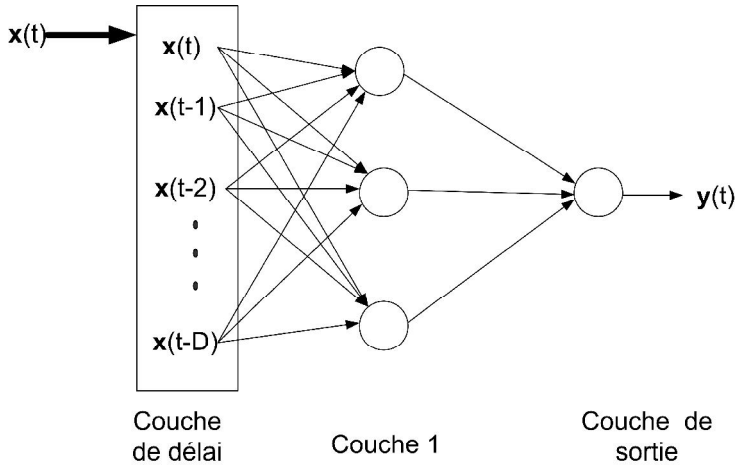


FIG. 6.17 – Réseau de neurone FTDNN

Ce type de réseau peut fournir la prédiction de l'évolution temporelle de la variable définie ci-haut. Nous avons utilisé un réseau avec un délai global de 6 et une couche cachée de 3 neurones. Les fonctions d'activation pour la couche cachée sont de type tangente sigmoïde alors que le neurone de sortie est linéaire. Le choix du délai global de 6 s'est fait de manière arbitraire mais avec le soucis de limiter la taille du réseau. Le nombre de neurones cachés est déterminé par essai. Sur la figure 6.18, nous illustrons la prédiction de l'évolution temporelle de

l'indicateur vibratoire d'usure. La figure montre aussi l'effet, sur la sortie, du pas de temps du vecteur d'entrée. Le pas de temps entre les enregistrements expérimentaux est de 0,3 minutes. Pour une meilleure prédiction, il est nécessaire que le pas de temps soit proche du pas des données d'apprentissage.

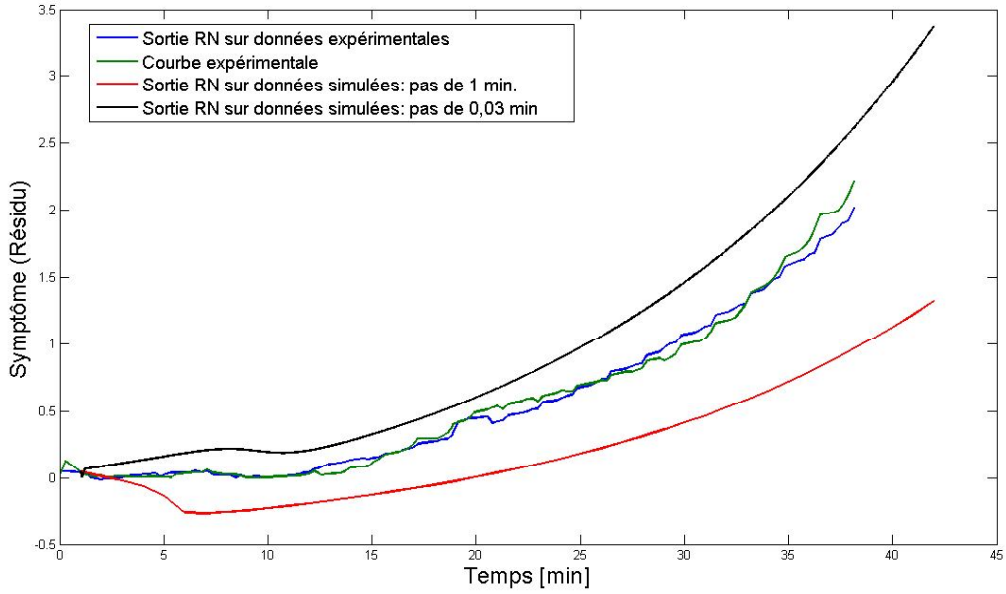


FIG. 6.18 – Prédiction de la valeur du résidu par réseau de neurones

6.8.4 Conclusion

Ce chapitre s'est focalisé sur la surveillance de l'usure d'outils en usinage à l'aide de mesures vibratoires. La manifestation vibratoire de l'usure est souvent mélangée à d'autres contributions vibratoires dues par exemple à la cinématique de la machine-outil. Les études expérimentales menées dans le cadre de cette thèse ont montré qu'un bon traitement du signal vibratoire peut permettre d'isoler les composantes qui contiennent l'information sur l'usure.

La première étude, qui était caractérisée par des signaux non-stationnaires, a exigé un moyen de traitement qui puisse capturer les singularités locales. Elle a proposé une méthode de traitement du signal, l'ASS, et a défini à travers celle-ci des variables qui ont été suffisamment fidèles à l'évolution de l'état de l'outil et robustes vis-à-vis du bruit. Par la suite, l'étude a été étendue suivant deux axes : d'abord pour montrer qu'une seule direction de mesure de vibrations suffisait pour la surveillance de l'usure, ensuite pour établir que les indicateurs moins coûteux du

domaine temporel ne permettaient pas d'identifier l'état d'usure modérée.

Le seconde étude expérimentale a proposé la construction de variables qui suivent l'évolution temporelle d'une dégradation, en l'occurrence l'usure d'outils. Elle ouvre une perspective pour des travaux en pronostic vibratoire.

Références

[1] N. Ghosh et al. : Estimation of tool wear during cnc milling using neural network-based sensor fusion. *Mechanical Systems and Signal Processing*, 21:466–479, 2007.

[2] Erkki Jantunen : *Indirect multisignal monitoring and diagnosis of drill wear*. Thèse de doctorat, Helsinki University of technology, Janvier 2006.

[3] C.S. Lee et D.A. Dornfeld : Design and implementation of sensor-based tool-wear monitoring systems. *Mechanical Systems and Signal Processing*, 10(4):328–347, 1996.

[4] E. Riviére-Lorphévre, E. Filippi et P. Dehombreux : Forces, vibrations and roughness prediction in milling using dynamic simulation. *In Seventh national congress on theoretical and applied mechanics 2006, FPMs, Mons*, 2006.

[5] H.V. Ravindra, Y.G. Srinisvasa et R. Krishnamurthy : Modeling of tool wear based on cutting forces in turning. *Wear*, 169:25–32, 1993.

[6] T. Childs et al. : *Metal machining. Theory and applications*. Arnold, London, 2000.

[7] ENSAM, éditeur. *Usinage par enlèvement de copeaux. De la technologie aux applications industrielles*. Eyrolles, 2005.

[8] N.H. Cook et K. Subriamanian : Micro-isotope tool wear sensor. *Annals of CIRP*, 27(1): 73–78, 1978.

[9] S. Kurada et C. Bradley : A review of machine vision sensors for tool condition monitoring. *Computers in Industry*, 34:55–72, 1997.

[10] S. Kurada et C. Bradley : A machine vision system for tool wear assessment. *Trybology International*, 30(4):295–304, 1997.

[11] M. Balazinski, E. Czogala, K. Jemielniak et J Leski : Tool condition monitoring using artificial intelligence methods. *Eng. App. of Artificial Intelligence*, 15:75–80, 2002.

[12] C. Scheffer et Heyns P.S. : An industrial tool wear monitoring system for interrupted turning. *Mechanical Systems and Signal Processing*, 18:1219–1242, 2004.

[13] X. Li, S. Dong et P. K. Venuvinod : Hybrid learning for tool wear monitoring. *International Journal of Advanced Manuacturing Technology*, 16:303–307, 2000.

[14] Issam Abu-Mahfouz : Drilling wear detection and classification using vibration signals and artificial neural network. *International Journal of Machine Tools & Manufacture*, 43:707–720, 2003.

[15] T.I. El-Wardany, D. Gao et M.A. Elbestawi : Tool condition monitoring in drilling using vibration signature analysis. *International Journal of Machine Tools and Manufacture*, 36(6)(6):687–711, 1996.

[16] R. E. Haber, J. E. Jiménez, C. R. Peres et J. R. Alique : An investigation of tool-wear monitoring in a high-speed machining process. *Sensors and Actuators A*, 116:539–545, 2004.

[17] D.E. Dimla Sr. et P.M. Lister : On-line metal cutting tool condition monitoring i : force and vibration analyses. *International Journal of Machine Tools and Manufacture*, 40:739–768, 2000.

[18] W. König, K. Kutzner et U. Schehl : Tool monitoring of small drills with acoustic emission. *International Journal of Machine Tools and Manufacture*, 32(4):487–493, 1992.

[19] A. Dragoş Axinte et Nabil Gindy : Tool condition monitoring in broaching. *Wear*, 254: 370–382, 2003.

[20] Xiaoli Li, Shiu Kit Tso et Jun Wang : Real-time condition monitoring using wavelet transforme and fuzzy techniques. *IEEE Transanctions on systems, man and cybernetics-Part C : Applications and Reviews*, 30(3):352–357, 2000.

[21] Faleh A. Al-Sulaiman, M. Abdul Baseer et Anwar K. Sheikh : Use of electrical power for online monitoring of tool condition. *J. Material Processing Technology*, 166:364–371, 2005.

[22] E. Riviére-Lorphévre, E. Filippi et P. Dehombreux. : Chatter prediction using dynamic simulation. *International review of mechanical engineering*, 1:78–86, 2007.

[23] G. O'Donnel, P. Young, K. Kelly et G. Byrne : Towards the improvement of tool condition monitoring systems in the manufacturing environment. *Journal of Materials Processing Technology*, 119:133–139, 2001.

[24] Dimla E. Dimla Snr : The impact of cutting conditions on cutting forces and vibration signals in turning with plane face geometry inserts. *J. Material Processing Technology*, 155-156:1708–1715, 2004.

[25] X. Luo, K. Cheng, R. Holt et X. Liu : Modeling flank wear of carbide tool insert in metal cutting. *Wear*, 259:1235–1240, 2005.

[26] R.J. Kuo et P.H. Cohen : Intelligent tool wear estimation system through artificial neural networks and fuzzy modeling. *Artificial Intelligence in Engineering*, 12:229–242, 1998.

[27] Andrzej Sokolowski : On some aspects of fuzzylogic application in machine monitoring and diagnostics. *Engineering Applications of Artificial Intelligence*, 17:429–437, 2004.

[28] S. Achiche, M. Balazinski, L. Baron et K. Jemielniak : Tool wear monitoring using genetically-generated fuzzy knowledge bases. *Engineering Applications of Artificial Intelligence*, 15(3-4):303–314, 2002.

[29] Bovic KILUNDU et Pierre DEHOMBREUX : Analyse spectrale singulière des signaux vibratoires et machine learning pour la surveillance d'usure d'outils. *Mécanique et Industries*, En presse, 2008.

[30] D.R. SALGADO et F.J. ALONSO : Tool wear detection in turning operations using singular spectrum analysis. *Journal of Materials Processing Technology*, 171:451–458, 2006.

[31] T. OBIKAWA et J. SHINOZUKA : Monitoring of flank wear of coated tools in high speed machining with a neural network art2. *International Journal of Machine Tools and Manufacture*, 44:1311–1318, 2004.

[32] Dimla E. Dimla SNR : Sensor signals for tool-wear monitoring in metal cutting operations–a review of methods. *International journal of Machine Tools and Manufacture*, 40:1073–1098, 2000.

[33] C.Y. JIANG, Y.Z. ZHANG et H.J. XU : In-process monitoring of tool wear stage by the frequency band-energy method. *Annals of the CIRP (1987)*, 36 (1):45–48, 1987.

[34] Dongfeng SHI et Nabil N. GINDY : Tool wear predictive model based on least squares support vector machines. *Mechanical Systems and Signal Processing*, doi :10.1016/j.ymssp.2006.07.016, 2006.

[35] Ian H.WITTEN et Eibe FRANK : *Data mining : Practical machine learning tools and techniques*. Morgan Kaufmann, San Francisco, 2nd édition, 2005.

[36] D. HAND, H. MANILLA et P. SMYTH : *Principles of Data Mining*. MIT Press, 2001.

Chapitre 7

Conclusions et perspectives

Cette thèse s'est fixée comme objectif général d'explorer les possibilités offertes par les méthodes du *data mining* aux processus de détection et de diagnostic vibratoire de dégradations mécaniques. Nous avons, dans un premier temps, circonscrit les notions fondamentales de la détection et du diagnostic vibratoire avant d'aborder la question du prétraitement des signaux. Ayant établi que la détection et le diagnostic sont d'autant plus précis que la qualité des données sur lesquelles ils se basent est meilleure, les objectifs spécifiques de cette thèse s'inscrivaient dans la contribution à l'élaboration des processus adéquats de prétraitement du signal vibratoire mesuré en vue de l'extraction des indicateurs. Deux domaines techniques importants ont été utilisés pour l'application des démarches développées : la surveillance des machines tournantes et celle de l'usure d'outils en usinage. Les apports de notre travail se situent à la fois sur le plan méthodologique et sur le plan expérimental.

Sur le plan expérimental, l'étendue des observations a permis de montrer la particularité de chaque problème rencontré dans le cadre de la détection et du diagnostic vibratoire. Déjà, dans le domaine des machines tournantes le besoin d'identifier les défauts combinés implique une démarche différente de celui de la détection précoce d'une avarie. La surveillance de l'usure d'outils a permis de proposer des solutions satisfaisantes pour l'exploitation des signaux bruités et/ou instationnaires.

Ces différents besoins de traitement du signal et d'analyse des données nous ont amenés au développement d'une librairie de fonctions sous MATLAB. Cette librairie a été intégrée avec une interface d'acquisition réalisée sous LABVIEW, permettant quelques prétraitements et l'extraction de certains indicateurs directement en cours d'acquisition.

Les principales contributions méthodologiques de cette thèse se sont basées sur l'analyse en

composantes principales. Celle-ci a été utilisée dans trois procédures différentes :

- au niveau du traitement du signal brut par l'analyse spectrale singulière,
- au niveau de la définition des indicateurs de détection,
- et au niveau de la procédure de sélection des indicateurs pertinents.

Nous avons proposé une démarche pour le traitement des signaux instationnaires dans le but de permettre l'extraction des indicateurs à forte richesse informationnelle. L'analyse spectrale singulière s'appuie effectivement sur l'ACP dans le but de débruiter le signal. Celui-ci est projeté dans l'espace des composantes principales de sa matrice de trajectoire, puis reconstruit en regroupant ses composantes suivant un critère informationnel basé sur la contribution à la variance totale. Cette démarche a été présentée dans le chapitre 3. Elle a été appliquée à l'étude de signaux de surveillance d'usure d'outils en tournage où la nature même du processus introduisait une légère instationnarité dans le signal. Une particularité dans l'utilisation de l'ASS au cours de cette étude a été le caractère pseudo-local de son application. Ceci s'est justifié par le besoin d'obtenir des matrices de trajectoire de taille raisonnable. Le processus de décomposition-reconstruction s'est ainsi effectué sur des segments de signal pour, *in fine*, procéder à une concaténation de segments de signal reconstruits afin d'obtenir la reconstruction globale. Les résultats auxquels cette analyse a abouti, ont montré, au chapitre 6, que l'ASS permettait la reconnaissance de l'état de l'outil au-delà des phénomènes qui pouvaient bruiter les signaux, en l'occurrence le bourrage copeau. Comparés aux indicateurs construits dans le domaine temporel, mais toujours par balayage du signal à l'aide d'une fenêtre temporelle, les indicateurs issus de l'ASS ont été les seuls à permettre la distinction entre l'usure modérée et l'usure avancée.

Par contre, appliquée au traitement des signaux destinés à la détection des défauts de roulements, l'ASS ne s'est pas avérée supérieure aux autres méthodes utilisées connues : le filtrage simple et l'analyse en ondelettes. Cette étude, présentée au chapitre 5, a plutôt montré que le filtrage comme prétraitement au cours de la détection d'enveloppe menait à la construction d'indicateurs plus riches que l'ASS et l'analyse en ondelettes. Ces conclusions ont découlé d'une vaste expérimentation, par validation croisée, de différents types de classifieurs sur différents ensembles d'indicateurs construits sur base de ces trois prétraitements. Le choix du meilleur ensemble de données et du meilleur classifieur a été effectué en analysant les coûts globaux et les taux de classification.

L'importance de la réduction dimensionnelle et de la sélection des variables an été soulignée aux chapitres 3 et 4. L'ACP est un moyen qui permet de procéder à la réduction dimensionnelle.

Au lieu de laisser les classifieurs opérer dans l'espace des composantes principales, nous avons préféré utiliser cette étape pour la sélection des indicateurs pertinents. La démarche que nous avons proposée a consisté à encapsuler un classifieur de type réseau de neurones dans un processus de sélection. En effet le réseau de neurones est nourri avec les facteurs issus de l'ACP, et l'observation des poids synaptiques à l'entrée du réseau permet de retenir les facteurs jugés plus significatifs. Par la suite, l'analyse des saturations de ces facteurs mène à un ensemble de variables pertinentes dans l'espace d'origine. Cette approche de sélection a été utilisée dans les applications expérimentales présentées dans ce travail. Au chapitre 4, nous l'avons comparée à la sélection basée sur un arbre de décision et sur l'analyse discriminante.

Les différentes méthodes de *data mining* qui sont utilisées pour le besoin de diagnostic ont été présentées au chapitre 4. Leur application aux différentes études expérimentales a permis de discuter la conduite du processus d'apprentissage, en prenant en compte la valeur des coûts d'erreur en cours d'apprentissage. Nous avons aussi eu l'occasion de nous former un jugement de valeur sur chaque méthode vis-à-vis des problèmes traités. D'une manière générale, les réseaux de neurones et les arbres de décision se sont avérés être de bons systèmes pour le besoin de la détection et du diagnostic. Les arbres de décision ne sont presque pas repris dans la littérature de la communauté du diagnostic. Et pourtant, ils peuvent offrir de bons résultats en termes de règles de détection ou d'identification, et surtout dans l'établissement des seuils de surveillance.

La question de détection précoce des avaries des équipements mécaniques pose le problème de la construction d'indicateurs de surveillance qui soient sensibles à la manifestation de l'anomalie dès sa phase initiale. L'idée de tirer profit d'une situation multivariables (plusieurs variables issues d'un ou de plusieurs capteurs) a encore suggéré l'utilisation de l'ACP. Nous avons ainsi proposé la combinaison des arbres de décision et d'une procédure de construction des variables de détection à l'aide d'un modèle ACP qui génère un résidu témoignant de l'écart par rapport au fonctionnement normal. Avec la construction du résidu, l'analyse est réduite à une seule variable.

Dans certaines situations, la frontière entre le fonctionnement normal et les autres états de fonctionnement n'est pas explicite. C'est à ce niveau que nous avons proposé une clusterisation pour fixer la frontière. Cette approche a été utilisée avec succès dans une application de détection précoce des défauts de bague interne d'un roulement à rouleaux cylindriques (chapitre 5). Appliquée à une surveillance d'usure d'outils en fraisage au chapitre 6, la démarche a encore montré son efficacité en permettant la distinction de différentes phases de dégradation de l'outil

par le phénomène d'usure.

Au cours de ce travail, nous avons également abordé le problème de données à dépendance temporelle. Ceci nous a conduit à proposer un moyen de construction des variables qui capturent cette dépendance dans leur définition. Nous avons proposé deux démarches de construction de telles variables : basée sur les pentes et basée sur les moyennes courantes. La seconde démarche a été appliquée aux données d'un essai *run-to-failure* dans le but du suivi de l'usure de l'outil. Le comportement de la variable ainsi construite, avec l'évolution de l'usure, permet d'envisager avec confiance des perspectives dans le domaine du pronostic à partir des mesures.

En effet, dans le but du pronostic les méthodes basées sur la mesure et les techniques d'apprentissage devront susciter de plus en plus d'intérêt aux côtés des méthodes fiabilistes et des modèles physiques de dégradations. Elles sont les plus précises, et pas nécessairement chères du fait des progrès réalisés dans le domaine de la mesure. Ces méthodes témoignent de l'état réel du système alors que les modèles physiques de dégradation peuvent contenir des incertitudes dues aux hypothèses inhérentes à la modélisation. L'estimation de l'évolution du dommage physique n'est pas plus précise avec un modèle qu'avec une mesure. En outre, les méthodes fiabilistes ne sont pas toujours moins coûteuses puisqu'elles nécessitent tout de même un effort considérable pour la récolte et la validation des données.

Le futur de la maintenance se trouve dans l'intégration de ces trois méthodologies. Les modèles physiques de dégradation et les modèles fiabilistes sont des moyens de prévenir le développement de défauts. Cependant, compte tenu des incertitudes de ces modèles, il peut arriver que les défauts se développent plus tôt que ne le prédisent ces modèles. Dans ces situations, les méthodes basées sur les mesures doivent venir en complément pour mettre à jour ces modèles et tenir compte du fait que le défaut a déjà été initié.

Annexes

Annexe A

Essais expérimentaux sur banc d'essai de machines tournantes

Objectifs de l'étude et dispositif expérimental

Cette étude avait pour objectif l'identification des défauts combinés sur une machine tournante à l'aide des méthodes d'apprentissage. Les données de cette expérimentation ont aussi été utilisées pour tester la méthode de sélection des variables pertinentes.

Le dispositif expérimental sur lequel a porté cette étude est composé d'une ligne d'arbre comprenant un moteur à vitesse variable accouplé élastiquement à un arbre portant deux disques (figure A.1). Aux deux disques peuvent être fixées des masses pour créer des balourds de niveaux différents.

Le défaut de frottement est induit à l'aide d'un levier qui applique un galet contre un des deux disques. En faisant varier la force de contact galet-disque, différents niveaux de frottement peuvent être générés.

Il est aussi possible d'induire des défauts d'alignement et de roulements à billes. Le premier type de défaut est généré en déplaçant latéralement un des paliers tandis que le deuxième type (roulement) est produit par ovalisation de la bague externe du roulement à l'aide d'une vis de pression.

Chacun de ces quatre défauts induits a plusieurs niveaux d'expression et tous les défauts peuvent être induits simultanément.

Le comportement du dispositif est analysé pour 3 vitesses de rotation différentes obtenues à l'aide d'un variateur de vitesse contrôlant le moteur. Les fréquences de rotation retenues sont :

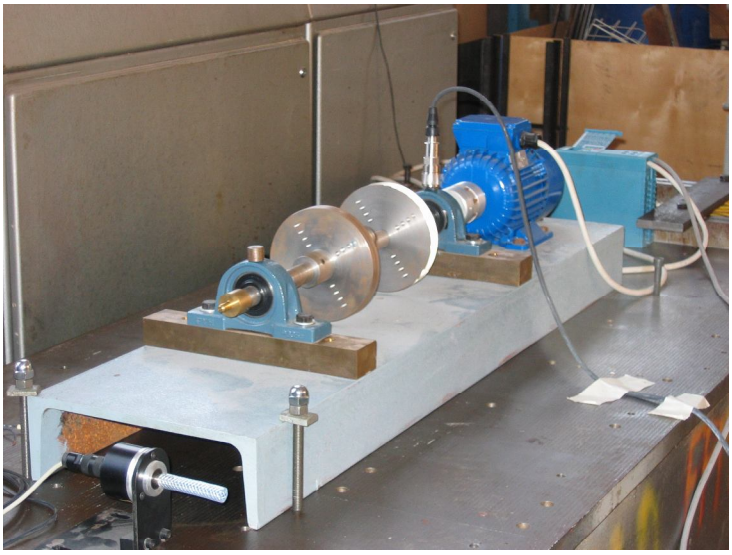


FIG. A.1 – Le banc d'essai

25 Hz, 22,5 Hz et 20 Hz. L'acquisition des signaux vibratoires est effectuée à 5120 Hz d'échantillonnage à l'aide d'un système d'acquisition multi-canaux. Chaque enregistrement dure 20 secondes. L'étude a concerné 162 conditions opératoires ; chacune d'entre elle est représentée par 4 signaux vibratoires relevés.

Le système d'acquisition utilisé comprend :

- un PC portable Compaq HP Presario 2510EA/P4 2.4/256 MB/30 GB
- une carte d'acquisition, un analyseur FFT en temps réel 4x20 kHz OROS
- un accéléromètre industriel (PCB piezotronics) M 622A01 (SN 7470) de sensibilité 10,3 mV/(m/s^2)

Procédure des essais

Après avoir imposé un défaut et mis en marche le banc d'essai, les mesures sont relevées sur chaque palier suivant deux directions radiales perpendiculaires (Z : verticale et Y : horizontale). Les points de mesure sont notés de la manière suivante :

- Z1 : direction verticale sur le palier 1
- Y1 : direction horizontale sur le palier 1
- Z2 : direction verticale sur le palier 2
- Y2 : direction horizontale sur le palier 2

Une condition de fonctionnement est complètement caractérisée par la vitesse de rotation, les défauts simulés et leurs niveaux respectifs.

Codification des défauts et des fichiers

Pour faciliter le traitement, un système de codification a été adopté. Il comporte toute l'information devant caractériser complètement la condition simulée. Le tableau A.1 reprend tous les défauts qui font l'objet de la campagne complète d'essais.

Défaut	Niveau	Signification	Désignation
Balourd	0	Machine supposée exempte de défaut de balourd	B0
	1	Masse de 20 g sur le premier disque	B1
	2	Masse de 40 g sur le premier disque	B2
	3	Masse de 20 g sur le second disque	B3
	4	Masse de 40 g sur le second disque	B4
Frottement	0	Machine supposée sans défaut de frottement	F0
	1	Force normale modérée au point de contact	F1
	2	Force normale élevée au point de contact	F2
Défaut de	0	Machine supposée sans ce défaut	R0
roulement	1	Vis de pression modérément serrée	R1
(ovalisation)	2	Vis de pression suffisamment serrée	R2
Désalignement	0	Pas de défaut d'alignement	A0
	1	Faible désalignement	A1
	2	Désalignement relativement grand	A2

TAB. A.1 – Défauts simulés

La codification comprend aussi une information sur le niveau de fréquence de rotation et la gamme de fréquences de mesure. Les fréquences de rotation du moteur sont désignées par les chiffres 1, 2 et 3

- fréquence 1 : 25 Hz
- fréquence 2 : 20 Hz
- fréquence 3 : 22.5Hz

Ainsi, par exemple, un fichier noté H1B1F2R0A0 désignera un enregistrement à la fréquence de rotation de 25 Hz sur une condition de balourd1, de frottement2 et sans défaut de roulement ni d'alignement.

Indicateurs

Les indicateurs vibratoires sont calculés à partir d'un signal accélérométrique x_k traité dans le domaine temporel. Il s'agit de la valeur efficace RMS, du *kurtosis* Kurt et du *skewness* Skew.

Pour une condition de fonctionnement donnée, ces indicateurs sont extraits sur des signaux relevés en 4 points de mesure différents (2 points par palier, direction verticale Z et horizontale Y) et dans les gammes de fréquence suivantes :

- $1 \times f_{\text{rot}}$: 0-30 Hz ;
- $2 - 10 \times f_{\text{rot}}$: 30-300 Hz ;
- $11 \times f_{\text{rot}}$ et plus : 300-2000 Hz, f_{rot} étant la fréquence de rotation ;
- et 0-2000 Hz

Les défauts traités expérimentalement sur le banc d'essais

Toutes les conditions de fonctionnement qui ont été traitées sur le banc d'essais sont reprises ci-dessous.

VITESSE 1		VITESSE 2		VITESSE 3	
1	H1B0F0R0A0	**1**	H2B0F0R0A0	**1**	H3B0F0R0A0
2	H1B0F0R1A0	**2**	H2B0F0R1A0	**2**	H3B0F0R1A0
3	H1B0F0R2A1	**3**	H2B0F0R2A1	**3**	H3B0F0R2A1
4	H1B0F0R2A2	**4**	H2B0F0R2A2	**4**	H3B0F0R2A2
5	H1B0F1R0A0	**5**	H2B0F1R0A0	**5**	H3B0F1R0A0
6	H1B0F1R1A0	**6**	H2B0F1R1A0	**6**	H3B0F1R1A0
7	H1B0F1R2A1	**7**	H2B0F1R2A1	**7**	H3B0F1R2A1
8	H1B0F1R2A2	**8**	H2B0F1R2A2	**8**	H3B0F1R2A2
9	H1B0F2R0A0	**9**	H2B0F2R0A0	**9**	H3B0F2R0A0
10	H1B0F2R1A0	**10**	H2B0F2R1A0	**10**	H3B0F2R1A0
11	H1B0F2R2A1	**11**	H2B0F2R2A1	**11**	H3B0F2R2A1
12	H1B0F2R2A2	**12**	H2B0F2R2A2	**12**	H3B0F2R2A2
13	H1B1F0R0A0	**13**	H2B1F0R0A0	**13**	H3B1F0R0A0
14	H1B1F0R1A0	**14**	H2B1F0R1A0	**14**	H3B1F0R1A0
15	H1B1F0R2A1	**15**	H2B1F0R2A1	**15**	H3B1F0R2A1
16	H1B1F0R2A2	**16**	H2B1F0R2A2	**16**	H3B1F0R2A2
17	H1B1F1R0A0	**17**	H2B1F1R0A0	**17**	H3B1F1R0A0
18	H1B1F1R1A0	**18**	H2B1F1R1A0	**18**	H3B1F1R1A0
19	H1B1F1R2A1	**19**	H2B1F1R2A1	**19**	H3B1F1R2A1

VITESSE 1		VITESSE 2		VITESSE 3	
20	H1B1F1R2A2	**20**	H2B1F1R2A2	**20**	H3B1F1R2A2
21	H1B1F2R0A0	**21**	H2B1F2R0A0	**21**	H3B1F2R0A0
22	H1B1F2R1A0	**22**	H2B1F2R1A0	**22**	H3B1F2R1A0
23	H1B1F2R2A1	**23**	H2B1F2R2A1	**23**	H3B1F2R2A1
24	H1B1F2R2A2	**24**	H2B1F2R2A2	**24**	H3B1F2R2A2
25	H1B2F0R0A0	**25**	H2B2F0R0A0	**25**	H3B2F0R0A0
26	H1B2F0R1A0	**26**	H2B2F0R1A0	**26**	H3B2F0R1A0
27	H1B2F0R2A1	**27**	H2B2F0R2A1	**27**	H3B2F0R2A1
28	H1B2F0R2A2	**28**	H2B2F0R2A2	**28**	H3B2F0R2A2
29	H1B2F1R0A0	**29**	H2B2F1R0A0	**29**	H3B2F1R0A0
30	H1B2F1R1A0	**30**	H2B2F1R1A0	**30**	H3B2F1R1A0
31	H1B2F1R2A1	**31**	H2B2F1R2A1	**31**	H3B2F1R2A1
32	H1B2F1R2A2	**32**	H2B2F1R2A2	**32**	H3B2F1R2A2
33	H1B3F0R0A0	**33**	H2B3F0R0A0	**33**	H3B3F0R0A0
34	H1B3F0R1A0	**34**	H2B3F0R1A0	**34**	H3B3F0R1A0
35	H1B3F0R1A1	**35**	H2B3F0R1A1	**35**	H3B3F0R1A1
36	H1B3F0R2A2	**36**	H2B3F0R2A2	**36**	H3B3F0R2A2
37	H1B3F1R0A0	**37**	H2B3F1R0A0	**37**	H3B3F1R0A0
38	H1B3F1R1A0	**38**	H2B3F1R1A0	**38**	H3B3F1R1A0
39	H1B3F1R2A1	**39**	H2B3F1R2A1	**39**	H3B3F1R2A1
40	H1B3F1R2A2	**40**	H2B3F1R2A2	**40**	H3B3F1R2A2
41	H1B3F2R1A0	**41**	H2B3F2R1A0	**41**	H3B3F2R1A0
42	H1B3F2R1A1	**42**	H2B3F2R1A1	**42**	H3B3F2R1A1
43	H1B3F2R2A2	**43**	H2B3F2R2A2	**43**	H3B3F2R2A2
44	H1B4F0R0A0	**44**	H2B4F0R0A0	**44**	H3B4F0R0A0

VITESSE 1		VITESSE 2		VITESSE 3	
45	H1B4F0R1A0	**45**	H2B4F0R1A0	**45**	H3B4F0R1A0
46	H1B4F0R1A1	**46**	H2B4F0R1A1	**46**	H3B4F0R1A1
47	H1B4F0R2A2	**47**	H2B4F0R2A2	**47**	H3B4F0R2A2
48	H1B4F1R0A0	**48**	H2B4F1R0A0	**48**	H3B4F1R0A0
49	H1B4F1R1A0	**49**	H2B4F1R1A0	**49**	H3B4F1R1A0
50	H1B4F1R1A1	**50**	H2B4F1R1A1	**50**	H3B4F1R1A1
51	H1B4F1R2A2	**51**	H2B4F1R2A2	**51**	H3B4F1R2A2
52	H1B4F2R1A0	**52**	H2B4F2R1A0	**52**	H3B4F2R1A0
53	H1B4F2R1A1	**53**	H2B4F2R1A1	**53**	H3B4F2R1A1
54	H1B4F2R2A2	**54**	H2B4F2R2A2	**54**	H3B4F2R2A2

Quelques spectres

Les figures A.2 et A.3 illustrent des spectres pour quelques conditions de fonctionnement.

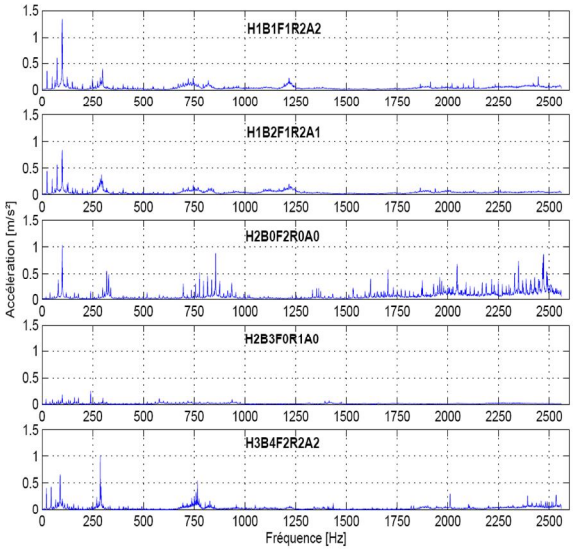


FIG. A.2 – Quelques spectres types

Synthèse des résultats

Sélection des indicateurs

Une sélection d'indicateurs effectuée en une fois pour tous les défauts a abouti au tableau A.2 qui représente les différentes variables fortement corrélées avec les scores factoriels. Le réseau de neurones encapsulé pour la sélection à une architecture 13-80-14 :

- 13 entrées (facteurs) ;
- 80 neurones pour la couches cachée ;
- 14 neurones de sortie pour les 14 classes de défauts (5 pour le balourd, 3 le frottement, 3 pour le roulement et 3 pour l'alignement)

L'approche par encapsulation d'un classifieur pour chacun des défauts a conduit à retenir les variables reprises dans le tableau A.3. Les couches cachés ont chaque fois 25 neurones ici.

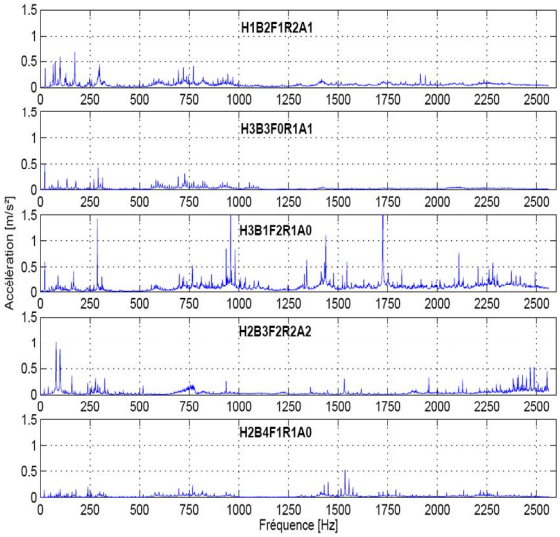


FIG. A.3 – Quelques spectres types

Variable ID	Indicateurs sélectionnés
1	RMS glob-z1
13	RMS glob-y1
25	RMS glob-z2
37	RMS glob-y2
10	RMS 300-2000 Hz-z1
22	RMS 300-2000 Hz-y1
34	RMS 300-2000 Hz-z2
46	RMS 300-2000 Hz-y2
5	Kurt 0-30 Hz-z1
17	Kurt 0-30 Hz-y1
29	Kurt 0-30 Hz-z2
41	Kurt 0-30 Hz-y2
6	Skew 0-30-z1
18	Skew 0-30-y1
30	Skew 0-30-z2
42	Skew 0-30-y2

TAB. A.2 – Variables à fortes saturations

Défaut	Variable ID	Indicateurs sélectionnés
Balourd	16	RMS 0-30 Hz-y1
	40	RMS 0-30 Hz-y2
	41	Kurt 0-30 Hz-y2
Frottement	1	RMS glob-z1
	25	RMS glob-z2
	37	RMS glob-y2
Roulement	7	RMS 30-300 Hz-z1
	46	RMS 300-2000 Hz-y2
	37	RMS glob-y2
	17	Kurt 0-30 Hz-y1
	26	Kurt glob-z2
	9	Skew 30-300-z1
Alignement	34	RMS 30-300 Hz-y2
	43	RMS 300-2000 Hz-z2
	33	Skew 30-300-z2

TAB. A.3 – Variables pertinentes

Taux de reconnaissance

Les résultats obtenus avec différentes méthodes de classification sur les ensembles d'indicateurs sélectionnés sont repris dans le tableau A.4 (approche **MoD-BiC**).

Défaut	Méthodes					
	RN	KNN	RB	AdD	Règles	FL
B	98,92	99,23	97,99	98,61	95,52	99,38
F	96,3	98,77	91,67	97,22	90,28	90,74
R	94,14	99,23	93,52	96,75	86,88	90,12
A	93,21	98,61	89,2	92,9	63,37	80,25

TAB. A.4 – Synthèse des résultats d'identification des défauts

Annexe B

Essais expérimentaux pour la surveillance d'usure d'outils en tournage

0bjectifs et dispostif expérimental

En vue d'évaluer les capacités de détection de l'usure d'outils à l'aide d'une approche *machine learning*, nous avons effectué une étude sur des machines à fileter les manchons de raccord de tubes utilisés en forage pétrolier (figure B.1). Trois accéléromètres monoaxiaux sont fixés à la tourelle du côté opposé au porte-outil (figure B.2) :

- le premier est orienté suivant la direction X, parallèle au porte-outil et à l'axe du manchon à usiner ;
- le second suivant la direction Y orthogonale à X ;
- et le troisième suivant la direction Z perpendiculaire au plan de la tourelle et parallèle à l'effort de coupe.

Pour le dégrossissage comme pour le filetage, l'usinage est exécuté en deux temps, d'abord d'une première extrémité vers le centre puis de l'autre vers le centre après retournement (rendu possible sur cette machine spéciale). Lors du dégrossissage, l'usinage est conique : la quantité de matière enlevée n'est pas la même en toute position de l'outil suivant l'axe du manchon. Cette quantité de matière enlevée est dégressive de l'extrémité du manchon vers le milieu suivant une conicité de 6,25% (figure B.3). Ceci introduit une légère instationnarité dans les signaux relevés. Les manchons usinés sont en acier de nuance API L80.

Les signaux vibratoires relevés à l'aide d'un système d'acquisition multi-canaux ont les caractéristiques suivantes :

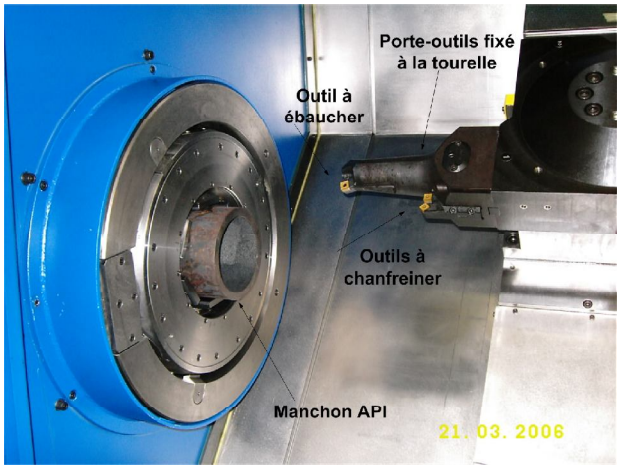


FIG. B.1 – Machine à fileter (avec la permission de M.P.Colinet sprl)

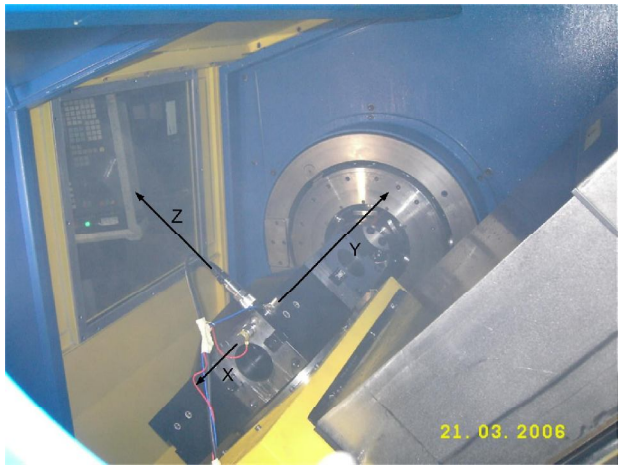


FIG. B.2 – Orientation des capteurs

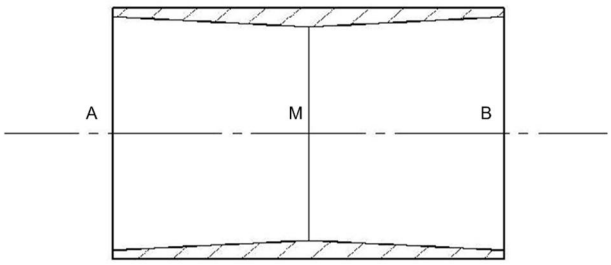


FIG. B.3 – Coupe du manchon

– fréquence d'échantillonnage : 51,2 kHz, pour une gamme d'analyse de 0-20 kHz ;
– taille échantillon : variable en fonction de la durée de l'opération d'usinage (en moyenne 20 secondes) ;

Un signal caractéristique des enregistrements est représenté à la figure B.4. Cette campagne

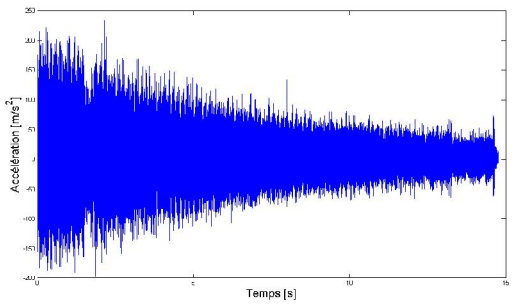


FIG. B.4 – Signal caractéristique enregistré lors d'un cycle de coupe

de mesures a été effectuée à l'occasion des essais de mise en marche de la machine. L'étude a été menée sur un ensemble de 43 enregistrements (de 3 signaux chacun) représentant principalement 5 états de fonctionnement dont trois caractérisent l'état de l'outil (N,U1,U2) et deux le processus (BC,V) comme repris dans le tableau B.1.

Le tableau B.2 reprend les conditions expérimentales pour les 43 essais, rangés dans leur ordre de succession chronologique.

Etat	Symbole	Effectif
Plaquette neuve ou comme neuve	N	22
Faible usure	U1	7
Usure assez avancée	U2	6
Bourrage copeau	BC	4
Passage à vide	V	4

TAB. B.1 – Etat de l'outil et du processus de coupe pour les 43 essais

Exp.	Condition	A (mm/tr)	V (m/min)	D (pouces)	Exp.	Condition	A (mm/tr)	V (m/min)	D (pouces)
1	BC	0.8	200	7	23	N	1	250	7
2	V	0.8	200	7	24	V	1	250	7
3	N	1	200	7	25	N	1	250	7
4	U1	1	200	7	26	N	1	200	7
5	N	1	200	7	27	N	1	200	7
6	N	1	200	7	28	U1	1	200	7
7	N	1	200	7	29	U1	1	200	7
8	BC	1	200	7	30	V	0.6	200	7
9	BC	1	200	7	31	U1	1.2	200	7
10	V	1	190	7	32	N	1.2	200	7
11	U1	1	190	7	33	N	1.2	200	7
12	U2	1	220	7	34	U1	1.2	200	7
13	U2	1	220	7	35	U1	1.2	200	7
14	U2	1	220	7	36	N	1.1	180	5.5
15	U2	1	220	7	37	N	1.1	180	5.5
16	U2	1	220	7	38	N	1.1	180	5.5
17	BC	1.2	220	7	39	N	1.1	180	5.5
18	N	1.2	220	7	40	N	1.1	180	5.5
19	N	1.2	220	7	41	N	1.1	180	5.5
20	N	1.2	220	7	42	N	1.1	180	5.5
21	N	1.2	220	7	43	N	1.1	180	5.5
22	U2	1	250	7					

TAB. B.2 – Données de l'étude : Condition, Avance (A), Vitesse de coupe (V) et Diamètre du manchon(D).

Les indicateurs

Les indicateurs construits sont des rapports des sommes de raies spectrales des signaux reconstruits par analyse spectrale singulière et de leurs résidus correspondants après filtrage passe-bande (cf. section 3.8). Par exemple, la variable RC3 2000-4000 Hz est issus du signal

reconstruit S_{RC3} et de son résidu, filtrés dans la gamme 2000-4000 Hz. Les vecteurs d'indicateurs sont de dimension 18. Quinze indicateurs proviennent de l'analyse spectrale singulière et 3 représentent les conditions d'usinage (vitesse de coupe, avance et diamètre de la pièce).

Après processus de sélection des indicateurs pertinents, on retient 4 des 18 variables de départ :

- la variable RC1 4000-6000 Hz ;
- la variable RC3 2000-4000 Hz ;
- la variable RC5 4000-6000 Hz ;
- l'avance A.

Synthèse des résultats

Les résultats de reconnaissance de l'état de l'outil auxquels a abouti cette étude sont repris dans le tableau B.3

	N	U1	U2	BC	V
Réseaux Bayésiens	72.7	42.9	66.7	25	0
KNN	90.9	85.7	83.3	50	0
Arbres de Décisions	72.7	100	100	50	50
Réseaux de Neurones	95.5	100	83.3	0	0

TAB. B.3 – Probabilités d'identification (PI) de l'état de l'outil (en %)

Annexe C

Essais expérimentaux pour l'évolution temporelle de l'usure en fraisage

Objectifs de l'étude et dispositif expérimental

L'objectif poursuivi par cette étude expérimentale est de définir un indicateur qui témoigne de l'évolution de la dégradation de l'outil avec le temps. Les essais sont réalisés jusqu'à la mort de l'outil.

L'expérience est effectuée sur une fraiseuse à commande numérique pour un seul outil, un seul matériau et des conditions de coupe constantes. La fraiseuse est représentée à la figure C.1.

- Vitesse de coupe : V=875 tr/min
- Profondeur de coupe radiale : ae=1 mm
- Profondeur de coupe (axiale) : ap=4 mm
- Avance : s=70 mm/min

L'expérience conduite a consisté en 117 enregistrements de 10 secondes chacun relevés successivement à 25 kHz d'échantillonnage pendant une suite de passes jusqu'à l'usure complète de l'outil à VB≈300 μm. L'accéléromètre est fixé sur la pièce. Le contrôle a été effectué 2 fois au microscope : à VB≈200 μm et à la mort de l'outil (voir figure C.3). Les signaux temporels caractéristiques de l'outil en états bon et usé sont représentés à la figure C.2.

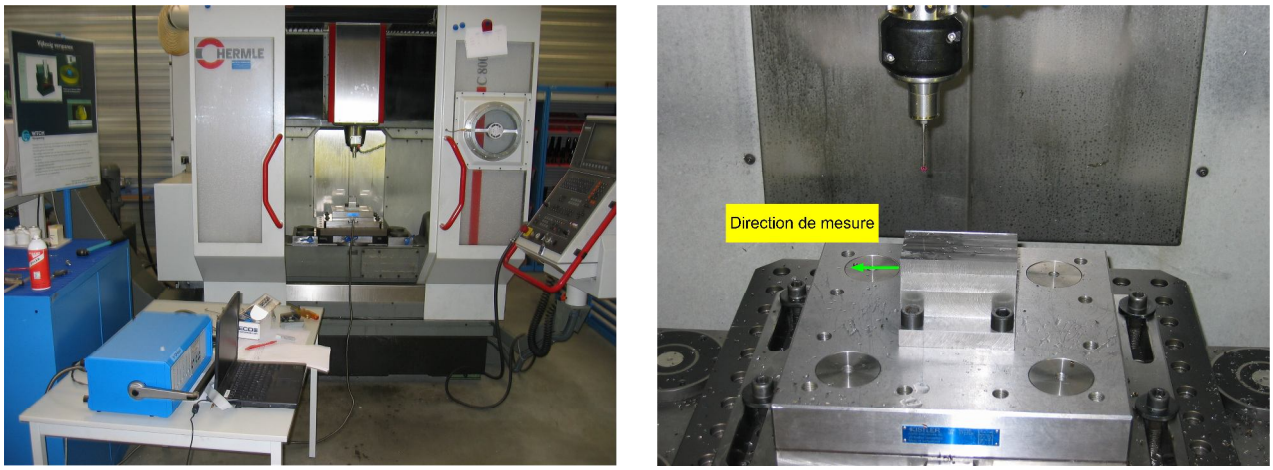


FIG. C.1 – Fraiseuse numérique ayant servi à l'étude

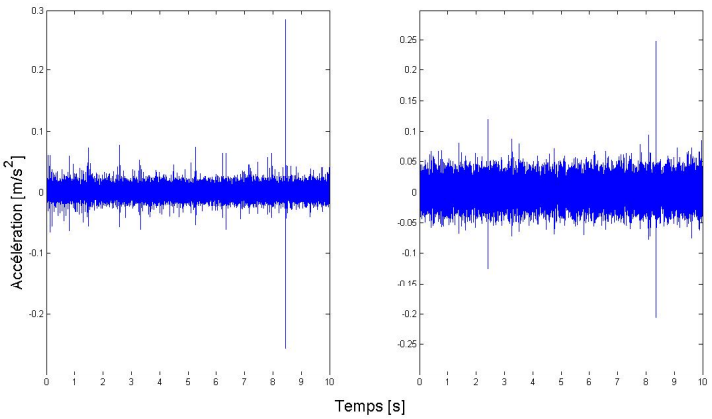


FIG. C.2 – Signaux temporels typiques de l'outil bon (à gauche) et usé (à droite)

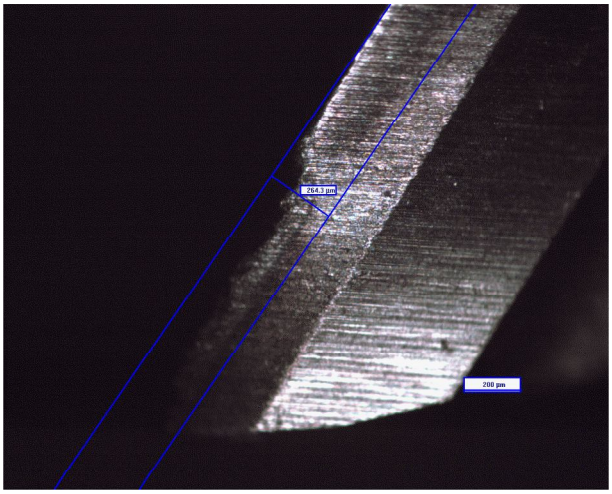


FIG. C.3 – Usure observée au microscope

Les indicateurs

Les indicateurs sont définis, dans différentes bandes de fréquence, à partir des moyennes courantes des valeurs efficaces et facteurs de crête pour introduire une mémoire temporelle du passé de l'outil. Ces indicateurs sont ensuite traités par ACP pour construire un résidu qui traduit la dégradation de l'outil.

Résultats

Le résultat auquel cette étude a abouti est la construction d'un indicateur qui traduit l'évolution temporelle de la dégradation. L'intérêt de cette démarche est de pouvoir identifier une loi monotone croissante qui correspond à une dégradation progressive. Un telle loi présente beaucoup d'intérêt pour le pronostic. Son évolution temporelle est montré sur la figure C.4.

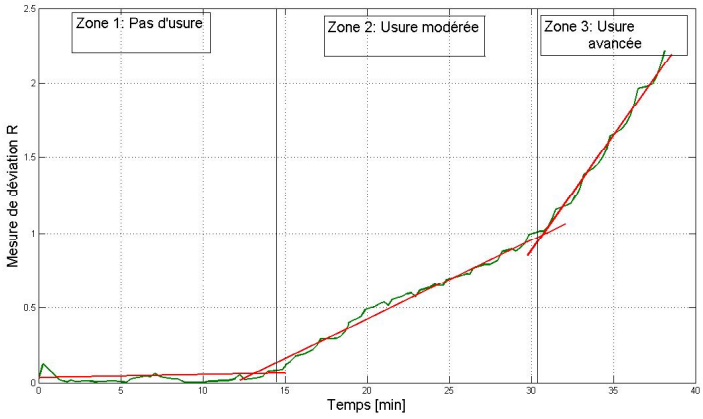


FIG. C.4 – Evolution du taux d'usure

Annexe D

Essais expérimentaux sur banc de détection de défauts des roulements

Cet annexe présente la conduite des expériences et le dispositif expérimental utilisé pour le test des roulements à l'Université de Reims.

Objectifs et dispositif expérimental

Cette étude a visé la mise en œuvre d'une procédure de détection précoce de défauts de roulements.

Le banc d'essai est constitué d'un arbre reposant sur deux paliers logés dans un carter et entraîné par un moteur électrique à vitesse variable. Le roulement à tester est monté du côté accouplement. C'est un roulement NU206 de FAG, à rouleaux cylindriques. Une charge radiale est appliquée à l'aide d'un vérin hydraulique du côté opposé au roulement sous test. Les figures D.1 à D.3 illustrent le banc de test.

Les défauts de tailles différentes sont réalisées sur les bagues internes de la manière suivante :

- le défaut très faible (niveau 1) est réalisé avec une seule pointe au stylo de gravure électrique ;
- le défaut faible (niveau 2) est réalisé sur un duromètre avec une pointe Vickers ;
- les autres niveaux sont réalisés avec un stylo de gravure électrique montré à la figure D.4 (voir tableau D.1).

Trois vitesses (1000, 1500 et 2000 tr/min) et trois charges (2000, 5000 et 7000 N) ont été

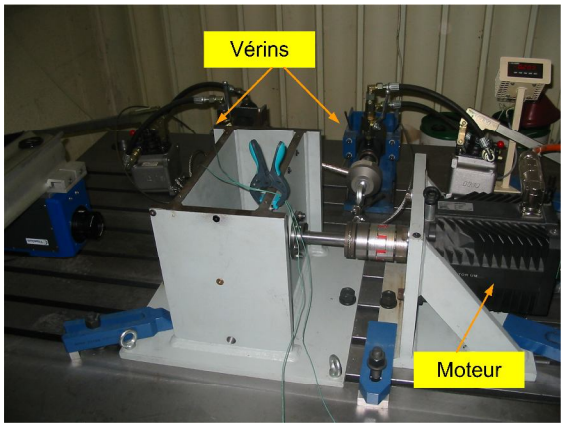


FIG. D.1 – Banc de test des roulements à l'Université de Reims

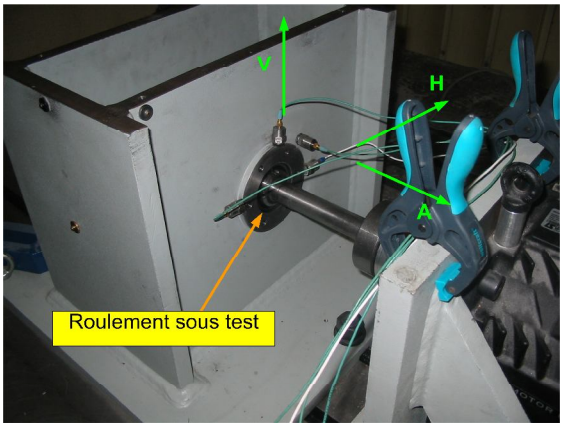


FIG. D.2 – Emplacement des accéléromètres à l'endroit du roulement sous test sur le banc

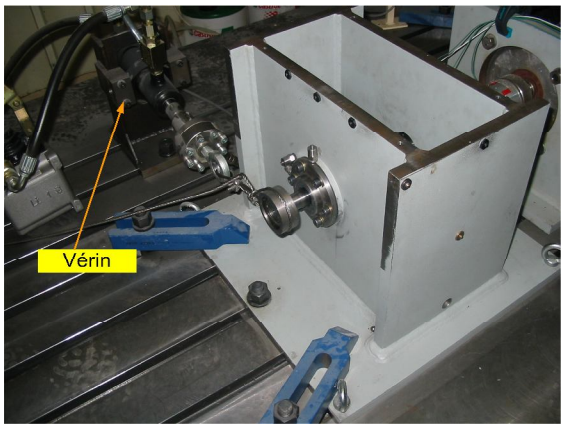


FIG. D.3 – Banc chargé au moyen d'un vérin hydraulique

FIG. D.4 – Réalisation des points de défaut sur la bague intérieur du roulement

		Taille défaut	Désignation	Effectifs par direction de mesure
1.	Roulement sain	0 mm^2	Niveau 0	8 signaux
2.	Défaut très faible	0,196 mm^2	Niveau 1	8 signaux
3.	Défaut faible	≈0,250 mm^2	Niveau 2	176 signaux
4.	Défaut avancé	≈1,5 mm^2	Niveau 3	71 signaux
5.	Défaut très avancé	3 mm^2	Niveau 4	90 signaux

TAB. D.1 – Tailles de défauts de roulements étudiés

utilisées. Les expériences combinent les différentes valeurs de vitesse et de charge.

Conduite de l'expérience

Le roulement avec une taille de défaut donnée, est laissé en fonctionnement pendant un certain temps au cours duquel les signaux vibratoires sont relevés automatiquement toutes les minutes (figure D.6). L'acquisition des signaux vibratoires est effectuée au moyen du système d'acquisition *CompactDaq* de *National Instruments* sur lequel est monté le module 9233 pour le conditionnement du signal (figure D.5). L'interface d'acquisition et de traitement est programmé sous *Labview*. On utilise 3 accéléromètres piézoélectriques à électronique intégrée dont les sensibilités sont reprises dans le tableau D.2 et la disposition représentée sur la figure D.2. Les signaux sont échantillonnés à 50 kHz. Les tests sont conduits à différentes vitesses de rotation et sous des charges différentes.

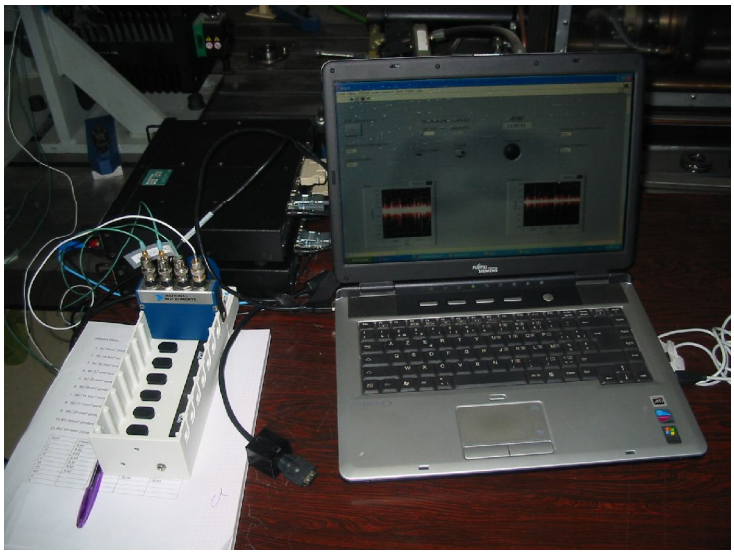


FIG. D.5 – Système d'acquisition

Capteur	Sensibilité [mV/g]	Direction
2125	9,52	Radiale horizontale (H)
2164	9,98	Axiale (A)
1698	10	Radiale verticale (V)

TAB. D.2 – Capteurs

Les fréquences caractéristiques du roulement pour les différentes vitesses de test sont données dans le tableau D.3. Ces fréquences peuvent se calculer à l'aide des relations suivantes :

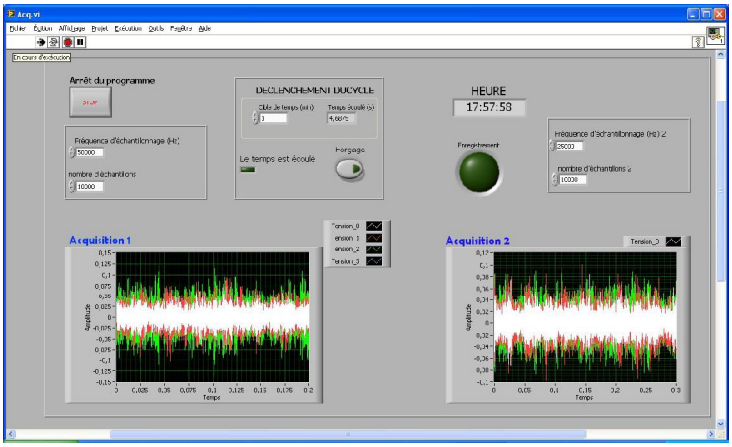


FIG. D.6 – Interface d'acquisition

– Fréquence de la bague extérieure (BPFO) :

$$f(\text{Hz}) = \frac{n}{2} f_r \left(1 - \frac{BD}{PD}\cos\beta\right) \tag{D.1}$$

– Fréquence de la bague intérieure (BPFI) :

$$f(\text{Hz}) = \frac{n}{2} f_r \left(1 + \frac{BD}{PD}\cos\beta\right) \tag{D.2}$$

– Fréquence de défaut des corps roulants :

$$f(\text{Hz}) = \frac{BD}{PD} f_r \left[1 - \left(\frac{BD}{PD}\cos\beta\right)^2\right] \tag{D.3}$$

où n est le nombre d'éléments roulants, f_r est la fréquence relative de rotation entre les bagues intérieure et extérieure, BD est le diamètre des éléments roulants, PD est le diamètre primitif et β l'angle de contact.

Vitesse [tr/min]	2000	1500	1000
Cage	13,4	10,1	6,7
Corps roulants	165,8	124,3	82,9
Bague extérieure (BPFO)	174,7	131,0	87,4
Bague intérieure (BPFI)	258,6	194,0	129,3

TAB. D.3 – Fréquences caractéristiques en Hz du roulement NU206E

Extraction des indicateurs

Les signaux mesurés sont d'abord traités afin d'extraire le signal d'enveloppe. Trois modes de prétraitement sont comparés : le filtrage passe-bande, l'analyse en ondelettes et l'analyse spectrale singulière. Ensuite, on calcule la somme de raies spectrales du spectre d'enveloppe. Ces sommes de raies, pour différentes composantes des signaux issus du traitement, sont fusionnées avant de passer dans un modèle ACP pour produire un résidu (figure D.7).

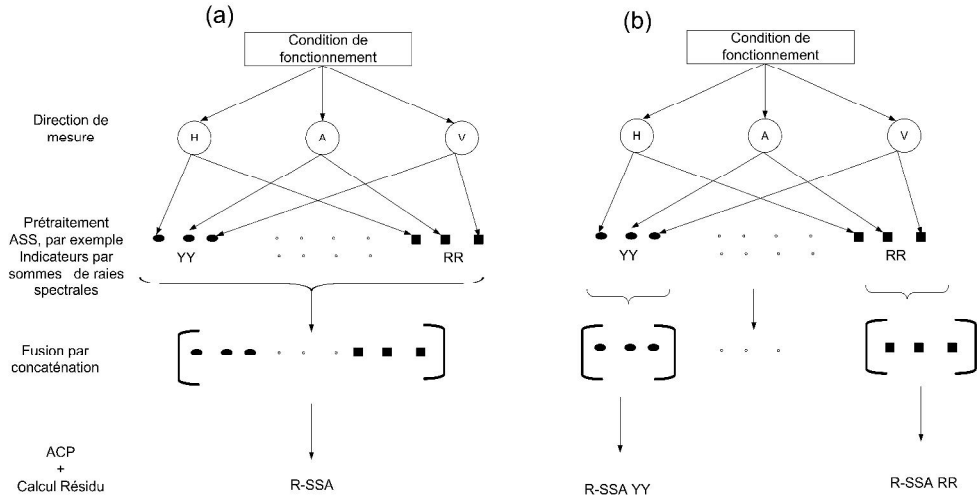


FIG. D.7 – Processus de calcul du résidu avec fusion (à (illustration avec un prétraitement ASS)

La figure D.8 illustre une phase du traitement.

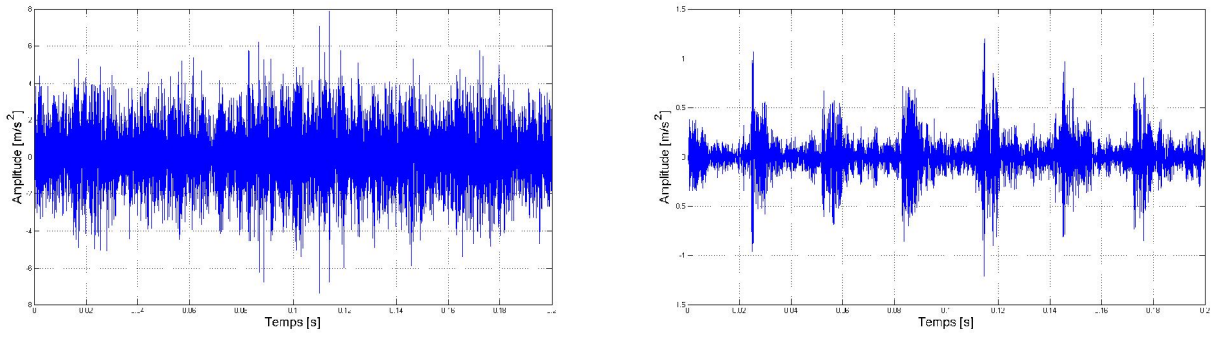


FIG. D.8 – Signal original (à gauche) et son filtrage dans la gamme 3-5 kHz (à droite)

Résultats

Détection précoce

Les résidus calculés servent d'indicateurs de détection et de diagnostic. Pour la détection, les arbres de décision sont construites afin de définir des seuils d'alerte correspondant aux résidus (figure D.9).

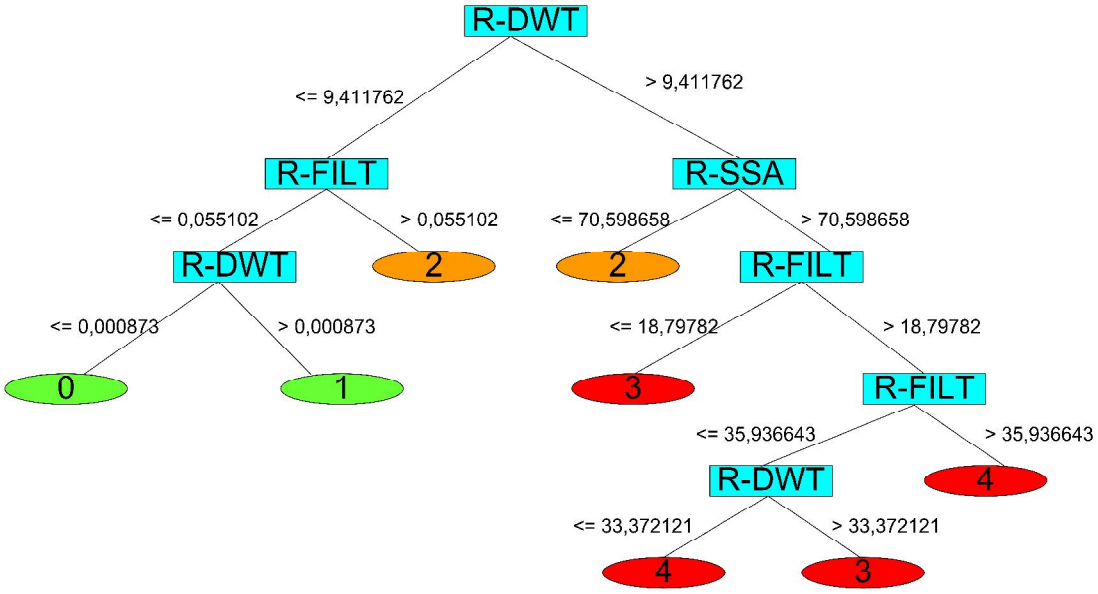


FIG. D.9 – Arbre

Identification

L'identification des niveaux de sévérité des défaut est effectuée au moyen de différents types de classifieurs, avec prise en compte des coûts. Les indicateurs sont les sommes de raies spectrales. Nous avons testé les différents ensembles d'indicateurs sur les algorithmes de classification suivants :

- Arbre de décision : noté J48 ;
- Réseau de neurones : MLP (3 couches, le nombre de neurones dans la couche cachée est égale à la moitié de la somme du nombre de classes et du nombre d'attributs) ;
- Machine à vecteurs de support : SVM (noyau polynomial de degré 2, coefficient de régularisation 10) ;

– Plus proches voisins : K-NN (le nombre qui suit indique le nombre de proches voisins. Par exemple, K-NN 2 est relatif à 2 voisins proches)
– Réseaux bayésiens : RB (RB 1 : 1 seul parent, RB 2 : 2 parents au maximum).

Le meilleur ensemble de données est constitué de la fusion des sommes de raies spectrales issues du prétraitement par filtrage. Le meilleur classifieur est le réseau de neurones. Il présente un taux de clasification de 99,04% pour un coût de 0,2 alors que le réseau bayésiens a un coût de 5,56 pour 94,19% de succès (tableau D.4). Le réseau de neurones a une seule couche cachée dont le nombre de neurones est égale à la moyenne entre le nombre de variables et le nombre de classes.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Taux	94.91	99.04	96.95	98.67	97.96	98.30	97.48	94.19	96.66
Coût	4,6	0,2	2,04	0,9	2,1	1,54	3,08	5,56	2,04

Key :(1) J48 ;(2) MLP ;(3) SVM ; (4) K-NN 1 ; (5) K-NN 2 ;(6) K-NN 3 ; (7) K-NN 4 ; (8) RB1 ; (9) RB2

TAB. D.4 – Taux de classification et coût