



**HAL**  
open science

# Suivi dynamique de composantes modulées : application à la surveillance automatique de défauts dans les éoliennes

Timothée Gerber

## ► To cite this version:

Timothée Gerber. Suivi dynamique de composantes modulées : application à la surveillance automatique de défauts dans les éoliennes. Traitement du signal et de l'image [eess.SP]. Université Grenoble Alpes, 2015. Français. NNT : 2015GREAT142 . tel-01313323v2

**HAL Id: tel-01313323**

**<https://theses.hal.science/tel-01313323v2>**

Submitted on 20 Dec 2016

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

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

## THÈSE

pour obtenir le grade de

## DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Signal, Image, Parole, Télécommunications (SIPT)**

Arrêté ministériel : 7 août 2006

Présentée par

**Timothée GERBER**

Thèse dirigée par **Nadine MARTIN** et  
codirigée par **Corinne MAILHES**

préparée au sein du  
**laboratoire Grenoble, Images, Parole, Signal, Automatique  
(GIPSA-lab)**  
dans l'école doctorale d'Électronique, Électrotechnique,  
Automatique et Traitement du Signal (EEATS)

## Suivi dynamique de composantes modulées Application à la surveillance automatique de défauts dans les éoliennes

Thèse soutenue publiquement le **30 novembre 2015**,  
devant le jury composé de :

**Sylvain MARCHAND**

Professeur, Université de la Rochelle, Président

**Guillaume GELLE**

Professeur, Université de Reims, Rapporteur

**Pascal LARZABAL**

Professeur, ENS Cachan, Rapporteur

**Guy d'URSO**

Ingénieur de recherche, EDF, Examineur

**Nadine MARTIN**

Directrice de recherche CNRS, GIPSA-lab, Directrice de thèse

**Corinne MAILHES**

Professeur INPT, Université de Toulouse, Co-directrice de thèse





# Remerciements

L'achèvement d'une thèse n'est pas une épreuve facile. Quand vient le temps de la rédaction, surgit alors le doute accompagné de nombreuses questions : Est-ce que ces pages seront lues attentivement ? Par combien de personnes ? Est-ce que quelques passionnés s'amuseront à chercher la référence cachée dans la page 42 ? Est-ce qu'il ne suffirait pas d'écrire uniquement les remerciements car soyons honnêtes, vous ne comptiez pas lire le reste, si ? Sans réponse, il faut tout de même poursuivre pour arriver au but. Ne pas abandonner. Ne pas baisser les bras. Jusqu'au jour où la rédaction finit par se terminer. Ce moment marque alors la fin d'une belle et longue aventure incluant de nombreux protagonistes. Pour leur implication de près ou de loin à mes travaux, je souhaite tous les remercier.

Et je démarre par le jury qui a évalué mes travaux de thèse. Merci au président Sylvain Marchand, aux rapporteurs Guillaume Gelle et Pascal Larzabal, et à Guy d'Urso pour leurs retours constructifs.

Je continue par celles sans qui cette thèse n'aurait jamais eu lieu, à savoir mes deux directrices de thèse Nadine Martin et Corinne Mailhes. Durant ces trois années partagées ensemble, vous m'avez guidé et vous m'avez fait découvrir le monde de la recherche. Merci pour vos nombreux retours et conseils qui m'ont permis de faire de grands bons en avant. Au passage, je tiens à te rassurer Corinne : malgré la distance, tu as su être très présente. Merci pour tout.

Ces travaux ont été riches en échanges scientifiques. Merci aux partenaires des projets KAStrion et SUPREME qui m'ont permis de voir ce qu'une collaboration européenne peut donner. Merci également aux membres de l'équipe SAIGA avec qui j'ai pu discuter. En particulier, je remercie tous les collègues avec qui j'ai régulièrement échangé : Pierre, Georgia, Gildas, Zhongyang, Guanghan, Thanh Trung, Alexis et Pascal. Sans oublier Marcin avec qui j'ai partagé quelques vodkas polonaises et de nombreuses problématiques.

Une pensée également à tous les collègues enseignants. Vous m'avez fait confiance et je vous en remercie.

De manière plus générale, un grand merci au personnel du GIPSA-lab qui travaille dans l'ombre et qui fait économiser une énergie folle quand vient le moment de s'occuper de l'administratif, du financier, du support informatique ou de tout autre service. Sans vous, la recherche ne peut pas se faire dans de bonnes conditions. Mention spéciale à ceux qui me sont régulièrement venus en aide : Lucia, Nadine B., Anne et Mikaël.

Les remerciements suivants sont plus personnels et s'adressent à celles et ceux qui ont fait de ce séjour au GIPSA-lab une belle aventure humaine. Sans ordre particulier, merci à Pascal, Arnaud, Céline, Cindy, Aude, Robin, Quentin, Raphaël, l'autre Tim, Guillaume, Miguel, Alexis, Taia, Florian, Lucas, Romain, Edouard, Manu, Rémy et tous les autres que j'oublie... Merci à vous tous pour ces fous rires et ces beaux moments partagés ensemble.

Bravo à toute la famille Gerber qui me supporte depuis tant d'années. Vous avez toujours répondu présents pour me soutenir et m'encourager. C'est grâce à vous que j'en suis là aujourd'hui. Merci pour tout.

Il me reste à remercier ceux qui, d'une manière ou d'une autre, m'ont donné l'occasion de penser à autre chose durant ces trois années. Se reconnaitrons tous ceux avec qui j'ai partagé un moment de vie : une rando en forêt ou en montagne, une sortie ski, un après-midi jeu ou bricolage, un voyage, une bière, une répétition ou une discussion pour parler de la vie. Je compte sur vous pour vivre plein d'autres belles aventures.

Enfin, la meilleure pour la fin ! Ne soyez pas jaloux car vous n'auriez de toute façon pas pu rivaliser avec son sourire. Un énorme merci à celle dont la patience, le soutien et l'amour m'ont été indispensables durant toutes les étapes de cette thèse.

# Table des matières

<b>Introduction</b>	<b>1</b>
<b>1 Surveillance automatique</b>	<b>5</b>
1.1 Contexte de recherche . . . . .	5
1.2 Analyse d'un signal . . . . .	9
1.3 Analyse d'une séquence de signaux . . . . .	18
1.4 Synthèse du chapitre . . . . .	22
<b>2 Identification automatique des harmoniques et des modulations</b>	<b>25</b>
2.1 Contexte . . . . .	26
2.2 Identification des séries harmoniques . . . . .	28
2.3 Identification des bandes de modulation . . . . .	36
2.4 Caractérisation des structures spectrales . . . . .	43
2.5 Applications . . . . .	46
2.6 Synthèse du chapitre . . . . .	51
<b>3 Suivi temps-fréquence des structures spectrales</b>	<b>53</b>
3.1 Contexte . . . . .	54
3.2 Méthode de suivi . . . . .	59
3.3 Suivi de structures spectrales . . . . .	67
3.4 Applications . . . . .	73
3.5 Synthèse du chapitre . . . . .	80
<b>4 Diagnostic du système surveillé</b>	<b>81</b>
4.1 Génération de tendances . . . . .	82
4.2 Détection de l'évolution d'un système . . . . .	94

---

4.3 Synthèse du chapitre . . . . .	109
<b>Conclusions et perspectives</b>	<b>111</b>
<b>A Les différentes stratégies de maintenance</b>	<b>115</b>
A.1 Maintenance corrective . . . . .	115
A.2 Maintenance préventive . . . . .	116
A.3 Comparaison des stratégies . . . . .	117
<b>B AStrion</b>	<b>121</b>
B.1 AStrion, un analyseur de spectre automatique . . . . .	121
B.2 Les modules d'AStrion . . . . .	125
<b>C Banc d'essai éolien</b>	<b>133</b>
C.1 Description du banc d'essai . . . . .	133
C.2 Les séquences de signaux du banc . . . . .	135
<b>Notations</b>	<b>137</b>
<b>Table des sigles et acronymes</b>	<b>141</b>
<b>Table des figures</b>	<b>143</b>
<b>Liste des tableaux</b>	<b>147</b>
<b>Bibliographie</b>	<b>149</b>

# Introduction

## Présentation du contexte

La surveillance est une opération qui consiste à vérifier le bon fonctionnement d'un système tout au long de sa durée d'utilisation. Elle présente un intérêt économique majeur. En effet, l'identification précoce d'un défaut permet de planifier et de préparer les opérations de maintenance alors que le système est encore opérationnel. Le temps d'arrêt est alors limité au minimum nécessaire pour l'entretien ou la réparation. De plus, les pièces du système ne sont remplacées que lorsque le changement est nécessaire, ce qui permet de les utiliser jusqu'au bout de leur durée de vie. La réduction du temps d'arrêt et du nombre d'interventions sur le système est une source importante d'économies.

La surveillance peut être réalisée de manière manuelle ; une personne vient de temps à autre vérifier l'état du système. Cependant, l'accès au système ou à une partie du système peut s'avérer long et ou difficile. C'est particulièrement le cas des éoliennes dont la construction ne peut se faire qu'à une certaine distance des habitations humaines. Les champs éoliens sont souvent isolés, ce qui est d'autant plus vrai pour les fermes éoliennes construites en pleine mer. L'accès se fait alors par bateau et dépend fortement des conditions climatiques. La surveillance manuelle devient coûteuse et difficile à planifier. Le développement d'une méthode de surveillance automatique est une perspective intéressante pour ces systèmes difficiles d'accès.

La forme minimale de la surveillance automatique consiste à effectuer des télémessures sur le système surveillé. Cela signifie que le système est instrumenté par un jeu de capteurs et qu'un module permet l'acquisition des mesures. Cette étape de collecte peut être complétée par un traitement des données mesurées, voire éventuellement une étape de diagnostic automatique, ou d'aide au diagnostic. Dans tous les cas, les données (ou les résultats des traitements) sont envoyées vers un centre de contrôle.

Les données collectées par le système de surveillance peuvent prendre plusieurs formes. Il peut s'agir de données scalaires ou de signaux, acquis plus ou moins régulièrement. S'il est possible de réaliser un diagnostic du système à partir de données scalaires [Lebranchu 2015], nous ne nous y intéresserons pas dans le cadre de cette thèse et nous nous concentrerons sur les signaux, porteurs de davantage d'informations sur le système surveillé. La richesse des informations contenues dans un signal dépend de la complexité du système. Plus le système surveillé comporte de pièces différentes, plus le nombre de phénomènes physiques y prenant place sera grand. Là où une valeur scalaire aura tendance à intégrer l'ensemble des phénomènes physiques, un signal pourra être traité pour séparer ces phénomènes et les interpréter individuellement. La manifestation de ces phénomènes au sein d'un signal peut prendre plusieurs formes, qui peuvent être très simples ou présenter des structures plus complexes comme des harmoniques ou des modulations.

L'acquisition des données collecte une ou plusieurs séquences de signaux. Ces séquences grandissent et contiennent de plus en plus de signaux au fur et à mesure de la surveillance. L'objectif de cette thèse est de proposer une méthode de traitement de ces séquences afin d'extraire un maximum d'information sur le système surveillé pour en faciliter le diagnostic. Cette méthode doit répondre à plusieurs problématiques différentes, qui seront abordées les unes à la suite des autres dans ce manuscrit.

Premièrement, il s'agit de répondre au problème d'extraction de toute l'information utile contenue dans chacun des signaux d'une séquence. Cette étape d'analyse individuelle est rendue complexe par les multiples formes que peuvent prendre les manifestations des phénomènes physiques du système surveillé. En effet, ces phénomènes peuvent faire apparaître des harmoniques ou des modulations au sein des signaux recueillis.

Deuxièmement, une fois extraites, les informations similaires d'un signal à l'autre doivent être associées. En particulier, la manifestation d'un même phénomène physique doit être suivi d'un signal à l'autre. Cette étape de suivi va alors permettre la comparaison de signaux successifs. Les informations ne seront plus vues comme locales pour chaque signal, mais de manière globale à l'échelle de toute la séquence.

Enfin, la dernière problématique qui va nous intéresser porte sur l'utilisation des informations globales de la séquence afin d'identifier l'état de santé du système surveillé. Ce bilan de santé doit servir d'aide au diagnostic, voire permettre le diagnostic automatique du système. La finalité de cette étape doit être la prise de décisions à distance des opérations de maintenance à effectuer.

À ces trois problématiques viennent s'ajouter quelques contraintes supplémentaires liées au contexte de ces travaux de thèse. Ces derniers prennent place au cœur de deux projets européens visant à intégrer et à commercialiser la méthodologie proposée dans un système de surveillance. En conséquence, des problématiques liées au transfert et à l'intégration des traitements dans un système embarqué viennent ajouter des contraintes supplémentaires.

## Présentation des chapitres

Cette thèse se compose de quatre chapitres. Le premier chapitre introduit le contexte de recherche. Nous présenterons les problématiques liées à la surveillance automatique d'un système, ainsi que les contraintes supplémentaires fixées par les projets européens [*KAStrion*] et [*SUPREME*]. Ce chapitre sera également l'occasion de présenter des travaux antérieurs à cette thèse, sur lesquels nous nous sommes basés pour proposer une méthode de surveillance automatique.

Le deuxième chapitre traite de la recherche des structures spectrales dans un spectre. En particulier, nous présenterons en détail une méthode permettant d'identifier toutes les séries harmoniques présentes dans un ensemble de pics estimés. Cette méthode est réutilisée sur des sous-ensembles de pics afin d'identifier également les bandes de modulation présentes

dans le spectre. Nous caractériserons toutes les structures spectrales identifiées et fournirons à chacune d'elles une *carte d'identité spectrale*.

Le troisième chapitre parle de la problématique du suivi des structures spectrales au cours d'une séquence de signaux. Nous présenterons une méthode de la littérature permettant de suivre des pics au cours du temps. Cette méthode, développée pour traiter des signaux de paroles, sera dans un premier temps adaptée au contexte de surveillance automatique. Dans un second temps, elle sera adaptée au suivi des séries harmoniques et des bandes de modulation.

Le quatrième chapitre est à la fois plus applicatif et plus exploratoire. Nous détaillerons la création d'indicateurs sur l'état de santé du système surveillé et présenterons de nombreuses études de cas sur des signaux issus de la surveillance de véritables systèmes mécaniques. Nous proposerons une méthode permettant de trier tous ces indicateurs pour faire ressortir en priorité ceux qui signalent l'apparition et l'évolution d'un défaut sur le système.



# Surveillance automatique

---

## Sommaire

<b>1.1</b>	<b>Contexte de recherche</b>	<b>5</b>
1.1.1	KAStrion et SUPREME, deux projets européens	6
1.1.2	État de l'art sur la surveillance automatique	6
<b>1.2</b>	<b>Analyse d'un signal</b>	<b>9</b>
1.2.1	Analyse spectrale	9
1.2.2	Contenu spectral d'un signal	11
1.2.3	Détection des pics	16
1.2.4	Identification des séries harmoniques et des bandes de modulation	18
<b>1.3</b>	<b>Analyse d'une séquence de signaux</b>	<b>18</b>
1.3.1	Suivi temporel des structures spectrales	19
1.3.2	Diagnostic du système surveillé	20
<b>1.4</b>	<b>Synthèse du chapitre</b>	<b>22</b>

---

La surveillance automatique consiste à fournir, sans intervention humaine, des indicateurs sur l'état de santé du système surveillé. Pour y parvenir, il est nécessaire d'acquérir des informations durant les phases de fonctionnement du système. Ces informations sont récoltées sous la forme de signaux représentant des mesures de grandeurs pertinentes du système. La séquence de signaux doit ensuite être traitée afin d'extraire les indicateurs permettant de conclure sur la santé du système.

Dans ce premier chapitre, nous verrons tout d'abord quel a été le contexte de recherche durant cette thèse. Ensuite, nous expliquerons comment la recherche d'une méthode de surveillance automatique se subdivise en plusieurs problématiques, que sont l'extraction du contenu spectral d'un signal, le suivi du contenu au cours du temps et la caractérisation de l'évolution du contenu pour faciliter le diagnostic.

## 1.1 Contexte de recherche

Cette thèse est fortement liée à deux projets de recherche européens. Ce contexte particulier a entraîné un lot de problématiques et de contraintes que nous allons détailler. Ensuite, un état de l'art nous montrera qu'il reste des défis à relever pour proposer une méthode de surveillance automatique répondant à toutes les contraintes fixées.

### 1.1.1 KAStrion et SUPREME, deux projets européens

Cette thèse s’est déroulée dans le cadre des projets européens KAStrion [*KAStrion*] et SUPREME [*SUPREME*]. Ces derniers ont pour objectif le développement de nouvelles méthodes de surveillance automatique d’un système mécanique. Les systèmes considérés par ces deux projets sont différents. Il s’agit d’éoliennes dans KAStrion, alors que SUPREME se concentre sur des machines à papier.

Si les applications sont différentes, les attentes et les contraintes sont quant à elles très similaires. Les cahiers des charges des deux projets visent à équiper une ou plusieurs machines tournantes avec un système de surveillance automatique, communément appelé CMS (Condition Monitoring System, en anglais). Ces systèmes s’occupent d’acquérir des signaux vibratoires et électriques sur les machines tournantes pendant leur fonctionnement. L’acquisition peut être régulière, ou dépendre de certaines conditions, comme la vitesse du vent ou la vitesse de rotation de la machine. Dans tous les cas, le traitement des signaux acquis doit permettre d’évaluer l’état de santé de la machine afin d’en proposer le diagnostic.

Les travaux de thèse présentés ici se sont concentrés sur le traitement des signaux acquis. Ils ont donc pour objectif de proposer une nouvelle méthode de surveillance automatique d’un système mécanique, répondant aux contraintes des deux projets. Parmi ces contraintes, nous pouvons citer :

- la généralité des traitements, devant fournir des résultats sur des systèmes très différents ;
- la nécessité de travailler « en-ligne », les signaux arrivant les uns après les autres au cours du temps ;
- la puissance de calcul et la mémoire limitées, les traitements devant s’effectuer sur un système embarqué.

Il est important de noter également que si ces travaux de thèse ont été fortement guidés par les applications de ces deux projets, les méthodologies proposées ont été développées d’une manière générique, avec l’idée de pouvoir être réutilisées dans d’autres contextes.

### 1.1.2 État de l’art sur la surveillance automatique

La surveillance de systèmes est un sujet déjà largement abordé dans la littérature et qui intéresse beaucoup les industriels. En effet, la surveillance d’un système permet de mettre en place une maintenance prévisionnelle. Ce type de maintenance permet d’anticiper une panne et de planifier à l’avance les opérations de maintenance. De plus, les différentes parties du système sont exploitées jusqu’à la fin de leur durée de vie. Finalement, le système est opérationnel plus longtemps et les coûts liés à la maintenance sont réduits. Un système de surveillance présente donc un intérêt économique majeur (voir annexe A).

Pour remplir efficacement son rôle, un système de surveillance automatique doit fournir un ou plusieurs indicateurs sur l’état de santé du système. Ces indicateurs sont utilisés pour décider si oui ou non, il est nécessaire de déclencher les opérations de maintenance. Ils doivent donc présenter les propriétés suivantes :

- permettre la détection précoce d'un défaut, afin d'anticiper au maximum la panne ;
- caractériser la sévérité du défaut avec précision pour planifier au mieux la maintenance ;
- être robuste et ne pas déclencher de fausses alarmes, entraînant des opérations de maintenance inutiles.

Le calcul d'indicateurs peut se faire de nombreuses manières, comme le suggère la littérature abondante sur le sujet [Jardine 2006, Heng 2009, García Márquez 2012]. Toutes ces méthodes se classent en deux catégories, que sont les méthodes de calcul d'indicateurs spécifiques au système surveillé et les méthodes de calcul d'indicateurs génériques, dérivés directement des données sans aucune information supplémentaire.

### 1.1.2.1 Indicateurs spécifiques au système

Le calcul des indicateurs spécifiques au système nécessite d'avoir des informations sur la conception du système surveillé, comme le nombre de billes dans un roulement ou le nombre de dents d'un engrenage [Gong 2013, He 2013, Rauber 2015, Zhang 2011, Kim 2013]. Cette connaissance *a priori* de la cinématique du système permet de calculer les fréquences caractéristiques et donc de spécialiser les indicateurs. En effet, chaque indicateur est construit pour isoler et surveiller une sous-partie du système. Cela signifie qu'en cas de détection de défaut, l'indicateur ayant déclenché l'alarme permet d'identifier immédiatement la pièce en cause, voire le type de défaut.

La force d'identification de ce type d'indicateurs est un avantage majeur dans un CMS. Cependant, les contraintes l'accompagnant sont fortes : il est nécessaire de connaître *a priori* la cinématique du système surveillé. Cela signifie que ces CMS peuvent être installés uniquement sur des systèmes dont les informations de conception sont connues. De plus, la re-configuration de ces CMS est nécessaire pour chaque nouveau système à surveiller. Ainsi, l'exclusion de certains systèmes et le long processus de mise en œuvre pour les autres compliquent l'utilisation de ces indicateurs spécifiques.

### 1.1.2.2 Indicateurs génériques, dérivés des données

Afin d'être plus génériques et de permettre leur installation sur n'importe quel système, certains CMS utilisent des indicateurs calculés uniquement à partir des données mesurées, sans ajouter d'information *a priori*. Ces indicateurs sont de différents types et sont plus ou moins élaborés. Le plus simple d'entre eux se contente de regarder l'énergie contenue dans le signal et la compare à des seuils de référence [ISO10816-21 2015]. D'autres utilisent des méthodes statistiques sur les échantillons du signal, ou sur des caractéristiques simples du signal comme son amplitude crête à crête ou ses premiers moments (moyenne, variance, coefficient d'asymétrie et coefficient d'aplatissement) [Večeř 2005]. Ces indicateurs sont très simples et très rapides à calculer. Cependant, en cas de défaut, ils ne permettent pas de définir quelle partie du système est en cause. Ils ne sont pas non plus capables de dire si un ou plusieurs éléments sont défectueux. Enfin, ils ne permettent pas non plus une détection

précoce du ou des défauts.

Il est également possible de dériver des indicateurs plus élaborés à partir des données. C'est par exemple le cas dans [Hang 2014] où une décomposition modale empirique permet d'extraire des caractéristiques du signal, avant qu'un classifieur n'utilise ces caractéristiques pour diagnostiquer le système. Seulement, cette méthode réalise un diagnostic à partir d'un signal unique, sans considérer l'évolution du défaut dans le temps. Dans [Court 2013], les auteurs exposent le concept de décomposition temps-échelle intrinsèque, créant ainsi un indicateur robuste face aux non-stationnarités du système surveillé. Cependant, l'indicateur prend la forme d'une carte temps-fréquence qui doit être manuellement inspectée. Dans le même registre, la méthode proposée dans [Sun 2014] permet d'isoler les signatures temporelles des défauts, mais ces dernières doivent être manuellement inspectées ou traitées plus finement pour faire le diagnostic du système. Dans le cas où des variables d'état du système sont disponibles, [Yin 2014a, Yin 2014b] propose de générer un résidu qui est alors automatiquement comparé à un seuil pour déterminer si un défaut est présent ou non. Cette décision binaire ne donne pas d'indications sur la partie du système en cause. La même idée est utilisée dans [Ding 2009] où plusieurs résidus sont générés pour identifier précisément la partie défectueuse du système. Cela se fait au prix d'un algorithme très gourmand en capacités de calcul. De plus, ces méthodes demandent de nombreux capteurs pour obtenir les variables d'état du système.

La principale force des indicateurs dérivés des données est leur généricité : quel que soit le système, le calcul des indicateurs est identique. Cette généricité simplifie l'installation d'un CMS, qui ne nécessite alors aucune configuration. Cependant, ces indicateurs ne permettent pas toujours le diagnostic automatique du système ou l'identification de la partie du système en cause. Quand ils le permettent, c'est au prix de grands temps de calcul.

### 1.1.2.3 Position du problème

L'objectif de cette thèse est de combiner les avantages des indicateurs spécifiques avec ceux des indicateurs génériques, afin de répondre aux cahiers des charges fixés par les deux projets. Nous allons donc proposer une méthode permettant de créer des indicateurs reflétant l'état de santé de chacune des parties du système, tout en étant construits sans aucun *a priori* sur sa cinématique. Ainsi, ces indicateurs serviront de base à un CMS générique, capable de surveiller l'état de n'importe quel système, sans nécessiter de configuration.

De plus, il faut que cette méthode fonctionne en ligne, c'est-à-dire qu'à chaque nouveau signal acquis, les indicateurs sur l'état de santé du système doivent être mis à jour. Et ce, sans être trop gourmand en temps de calcul ou en mémoire. Cette démarche soulève plusieurs interrogations, que nous allons détailler dans la suite du chapitre :

- Comment extraire efficacement toute l'information utile d'un signal ?
- Comment comparer les informations extraites entre deux ou plusieurs signaux ?
- Comment utiliser ces informations pour diagnostiquer le système et prendre une décision sur la nécessité de déclencher les opérations de maintenance ?

## 1.2 Analyse d'un signal

La première étape de la méthode de surveillance automatique est l'analyse individuelle de chacun des signaux. Un signal est un vecteur porteur d'information. Ce vecteur peut prendre plusieurs formes, quelles soient analogiques ou numériques. Si l'analyse ou le traitement d'un signal peut être fait de manière analogique, le niveau technologique et les outils dont nous disposons aujourd'hui rendent le traitement numérique des signaux bien plus simple, rapide et efficace. Ainsi, un signal analogique (comme un signal vibratoire ou un signal de parole) sera tout d'abord numérisé avant analyse.

Soit  $s(t)$  un signal continu, dépendant du temps  $t$ . Ce signal est observé sur une durée finie, représentée par l'intervalle  $[t_0 ; t_0 + \tau]$  où  $t_0$  correspond à la date où l'on commence à observer le signal et  $\tau$  la durée pendant laquelle il est observé. La numérisation du signal  $s(t)$  implique son échantillonnage, ou autrement dit, sa discrétisation. Le signal discret est noté  $s[k]$  et on a alors  $s[k] = s(t_0 + \frac{k}{f_e})$  avec  $f_e$  la fréquence d'échantillonnage du signal et  $k$  le pas de discrétisation  $k \in \{0, 1, \dots, \lfloor \frac{\tau}{f_e} \rfloor\}$ ,  $\lfloor \cdot \rfloor$  dénotant la fonction partie entière.

Les échantillons du signal  $s[k]$  contiennent une certaine quantité d'information que nous souhaitons extraire. Pour mener à bien l'extraction de tout le contenu d'un signal et ce, quel que soit le type de signal, nous nous sommes focalisés sur le domaine fréquentiel, fortement adapté à la description des signaux.

### 1.2.1 Analyse spectrale

L'analyse spectrale recouvre l'ensemble des techniques permettant de passer du domaine temporel au domaine fréquentiel. Ce passage permet de passer d'une représentation des variations du signal en fonction du temps à la représentation des proportions du signal appartenant à telle ou telle bande fréquentielle. Cette dernière représentation correspond à la densité spectrale du signal, plus communément appelée *spectre*. La définition de la densité spectrale, que nous noterons  $\gamma_s(f)$ , dépend de la nature du signal [Priestley 1981] Les 4 cas suivants sont à distinguer :

- le signal  $s(t)$  est déterministe à énergie finie. Cela signifie que sa puissance instantanée  $|s(t)|^2$  est intégrable sur  $\mathbb{R}$ . Dans ce cas, on calcule la densité spectrale d'énergie de la manière suivante :

$$\gamma_s(f) = |S(f)|^2, \quad (1.1)$$

avec  $S(f)$  la transformée de Fourier de  $s(t)$  se calculant ainsi :

$$S(f) \triangleq \int_{-\infty}^{+\infty} s(t)e^{-2\pi i f t} dt. \quad (1.2)$$

- le signal  $s(t)$  est déterministe, périodique et à puissance moyenne finie. Cela signifie que sa puissance instantanée est intégrable sur la période  $T_0$  du signal. Comme le signal est

périodique de période  $T_0 = \frac{1}{f_0}$ , il se décompose en série de Fourier :

$$s(t) = \sum_{k=-\infty}^{+\infty} c_k e^{2i\pi k f_0 t}. \quad (1.3)$$

La Densité Spectrale de Puissance (DSP) d'un tel signal est un « spectre de raies » aux fréquences  $k f_0$  :

$$\gamma_s(f) = \sum_{k=-\infty}^{+\infty} |c_k|^2 \delta(f - k f_0), \quad (1.4)$$

$\delta(\cdot)$  notant la distribution de Dirac.

- le signal  $s(t)$  est déterministe, non périodique et à puissance moyenne finie. Cette définition signifie que la puissance moyenne  $P_{\text{moy}}$  du signal sur un intervalle  $T$  a une limite finie lorsque  $T$  tend vers l'infini. Autrement dit,

$$P_{\text{moy}} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} |s(t)|^2 dt \text{ existe.} \quad (1.5)$$

Dans ce cas, la DSP du signal se calcule ainsi :

$$\gamma_s(f) = \lim_{T \rightarrow \infty} \frac{|S_T(f)|^2}{T}, \quad (1.6)$$

avec  $S_T(f)$  la transformée de Fourier du signal  $s(t)$  restreint à l'intervalle  $[-\frac{T}{2}, \frac{T}{2}]$  :

$$S_T(f) = \int_{-\frac{T}{2}}^{+\frac{T}{2}} s(t) e^{-2\pi i f t} dt. \quad (1.7)$$

- le signal  $s(t)$  est aléatoire, stationnaire, ergodique et à puissance moyenne finie. Dans ce cas, le signal  $s(t)$  est une réalisation du processus aléatoire  $S(t)$ . La DSP se calcule alors de la manière suivante :

$$\gamma_s(f) = \lim_{T \rightarrow \infty} \mathbb{E} \left[ \frac{|S_T(f)|^2}{T} \right], \quad (1.8)$$

où  $\mathbb{E}[\cdot]$  représente l'espérance mathématique et  $S_T(f)$  représente les transformées de Fourier des réalisations du processus aléatoire  $S(t)$  restreintes à l'intervalle  $[-\frac{T}{2}, \frac{T}{2}]$ . Il est également possible, d'après le théorème de Wiener-Khintchine, de calculer la DSP en prenant la transformée de Fourier de l'autocorrélation du signal.

Dans la pratique, les signaux utilisés dans ce manuscrit sont bruités et comportent donc une part d'aléatoire. Nous nous retrouvons donc dans le dernier des cas présentés. Cependant, les signaux sont acquis pendant un temps fini. Nous n'avons donc accès qu'à une réalisation de taille finie du signal. En conséquence, le calcul de la densité spectrale n'est pas possible ; seule une estimation du spectre est envisageable.

### 1.2.1.1 Choix de la méthode d'estimation du spectre

Les méthodes d'analyse spectrale sont classées en deux grandes catégories : les méthodes paramétriques et les méthodes non paramétriques. Les méthodes paramétriques, souvent appelées méthodes à « haute résolution », s'appuient sur des modèles physiques ou mathématiques. Elles sont particulièrement efficaces pour estimer un spectre, ou une partie de spectre, avec des signaux courts, tout en ayant une très bonne résolution. Les méthodes non paramétriques, quant à elles, ne s'appuient sur aucun modèle pour estimer le spectre d'un signal. En contrepartie, la résolution de l'analyse dépendra de la longueur du signal : plus le signal sera long, meilleure sera la résolution.

Pour que l'analyse spectrale soit la plus générique possible et puisse s'appliquer à tout type de signaux, il est nécessaire d'utiliser une technique ne nécessitant aucun *a priori* sur les signaux ou sur un modèle à utiliser. C'est pourquoi les méthodes paramétriques ne seront pas considérées ici. Ce choix se justifie d'autant plus que les capacités de stockage actuelles permettent d'enregistrer des signaux longs, de plusieurs centaines de milliers de points, voire plus. Les méthodes paramétriques sont peu adaptées aux signaux longs, car elles reposent bien souvent sur des inversions de matrices dont la taille dépend de la longueur du signal et du nombre de modes qu'il contient. *A contrario*, les méthodes non paramétriques seront très performantes sur des signaux longs et suffiront à obtenir une haute résolution spectrale.

L'analyse spectrale sera donc réalisée à partir de méthodes non paramétriques. Plus précisément, nous utiliserons des estimateurs basés sur la Transformée de Fourier Discrète (TFD), dont la définition est rappelée ci-dessous.

**Définition 1.1:** Transformée de Fourier Discrète (TFD)

La *TFD* est un outil de traitement du signal numérique, permettant d'estimer une densité spectrale discrète à partir d'un signal lui aussi discret et échantillonné sur une fenêtre temporelle finie de  $N$  points.

$$S[k] \triangleq \sum_{n=0}^{N-1} s[n] \cdot e^{-2i\pi k \frac{n}{N}} \quad \text{pour} \quad 0 \leq k < N.$$

Dans ce manuscrit, les signaux que nous considérerons sont réels. Cela signifie que l'ensemble des échantillons sont réels ( $\forall k, s[k] \in \mathbb{R}$ ). Or une propriété intéressante de la *TFD* stipule que le spectre d'un signal réel est pair. Les fréquences négatives et les fréquences positives portent les mêmes informations. Par souci de simplification de la représentation des spectres estimés, nous représenterons uniquement les fréquences positives.

## 1.2.2 Contenu spectral d'un signal

Nous allons à présent définir les différentes structures spectrales que l'on retrouve au sein d'un signal et qui sont porteuses d'information. L'élément le plus simple que l'on retrouve

dans un spectre est ce que nous appellerons un « pic spectral ». Un ensemble de pics respectant certaines règles peut alors devenir une structure spectrale plus complexe, comme une série harmonique ou une bande de modulation. Ces structures sont présentes dans les spectres des signaux issus de la surveillance de système et permettent entre autres d'évaluer l'état de dégradation de certaines pièces.

### 1.2.2.1 Les pics spectraux

La transformée de Fourier décompose un signal en une somme infinie de fonctions trigonométriques de toutes fréquences. Si le signal  $s(t)$  à transformer est sinusoïdal, sa transformée de Fourier est alors triviale et prend la forme d'une distribution de Dirac. Il en va de même pour sa densité spectrale. Cependant, le passage au numérique entraîne la discrétisation du signal sur un support borné, ce qui équivaut à la multiplication du signal par une fenêtre temporelle. L'estimation de la densité spectrale par la TFD transforme cette multiplication en une convolution, dont les opérandes sont la distribution de Dirac (transformée de Fourier de la sinusoïde) et la fenêtre spectrale (transformée de Fourier de la fenêtre temporelle). C'est la forme de cette fenêtre spectrale qui donne son nom à ce que nous appellerons un « pic spectral ». Un pic est donc la signature spectrale associée à une sinusoïde temporelle à une résolution donnée (dépendant de la taille et de la forme de la fenêtre temporelle).

Soit le signal sinusoïdal  $s[k] = \sin(2\pi f_0 k)$ , de fréquence  $f_0 \in \mathbb{R}^*$ . Ce signal discret et l'estimation de son spectre sont illustrés sur la figure 1.1. L'estimation du spectre prend la forme d'un pic spectral, dont le maximum se situe à la fréquence  $f_0$ .

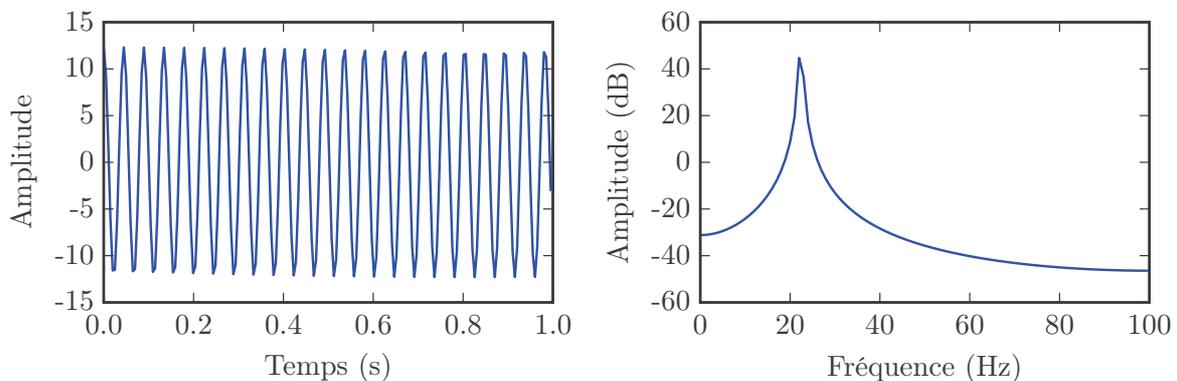


FIGURE 1.1 – Une sinusoïde dans le domaine temporel à la fréquence  $f_0 = 22,4$  Hz (à gauche) est représentée par un « pic spectral » dans le domaine fréquentiel à la fréquence  $f_0$  (à droite).

### 1.2.2.2 Les séries harmoniques

Une série harmonique est un ensemble de pics spectraux dont les fréquences  $f_i$  sont liées par une relation mathématique. Toutes les fréquences  $f_i$  sont des multiples d'une fréquence fondamentale  $f_0$ , tel que  $f_i = r \times f_0$ , avec  $r \in \mathbb{N}^*$  le rang de l'harmonique dans la série.

Typiquement, tous les signaux périodiques sont décomposables en séries de Fourier et sont donc la somme d'une sinusoïde de fréquence  $f_0$  et de ses harmoniques. La figure 1.2 représente un « signal carré », qui est un exemple de signal périodique. Sa décomposition en série de Fourier nous apprend que sont présents dans son spectre uniquement les harmoniques dont le rang  $r$  est impair et que l'amplitude des harmoniques décroît en  $\frac{1}{r}$ .

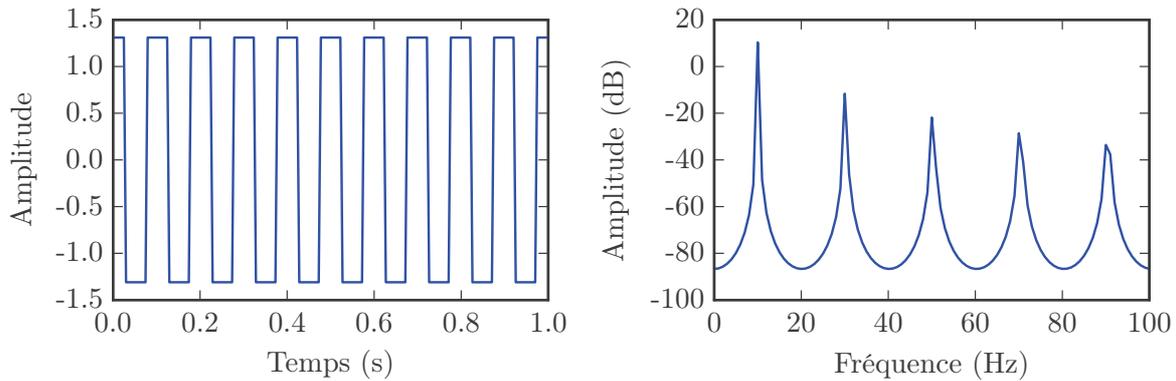


FIGURE 1.2 – Un signal « carré » dans le domaine temporel (à gauche) est représenté dans le domaine fréquentiel (à droite) par une infinité de pics spectraux, formant une série harmonique.

### 1.2.2.3 Les bandes de modulation

La modulation correspond à la transformation d'un signal porteur  $s_p(t)$  par un signal modulant  $m(t)$ . De manière générique, le signal porteur peut s'écrire sous la forme

$$s_p(t) = A(t) \cos(\Phi(t)),$$

avec  $A(t)$  l'amplitude instantanée de la porteuse et  $\Phi(t)$  sa phase instantanée. Si le signal modulant vient transformer l'amplitude du signal, on parlera de modulation d'amplitude. Sinon, si le modulant transforme la phase instantanée, on parle alors de modulation angulaire.

Dans le cas de la modulation d'amplitude, l'amplitude dépend du signal modulant :

$$A(t) = Km(t),$$

avec une constante  $K \in \mathbb{R}^*$ . La phase instantanée, quant à elle, peut s'écrire sous la forme

$$\Phi(t) = 2\pi f_p t + \varphi_0,$$

avec  $f_p$  la fréquence de la porteuse et  $\varphi_0$  la phase initiale lorsque  $t = 0$ . Pour simplifier la suite des explications, la phase initiale sera considérée comme nulle ( $\varphi_0 = 0$ ). Le cas simple d'une modulation d'amplitude par un modulant sinusoïdal de fréquence  $f_m$  est représenté sur la figure 1.3. L'estimation du spectre fait apparaître deux pics aux fréquences  $f_p \pm f_m$  en plus

de la porteuse. En effet, si  $m(t) = 1 + A_m \cos(2\pi f_m t)$ , alors

$$\begin{aligned} s_p(t) &= K (1 + A_m \cos(2\pi f_m t)) \cos(2\pi f_p t) \\ &= K \cos(2\pi f_p t) + \frac{K A_m}{2} \cos(2\pi (f_p - f_m)t) + \frac{K A_m}{2} \cos(2\pi (f_p + f_m)t). \end{aligned} \quad (1.9)$$

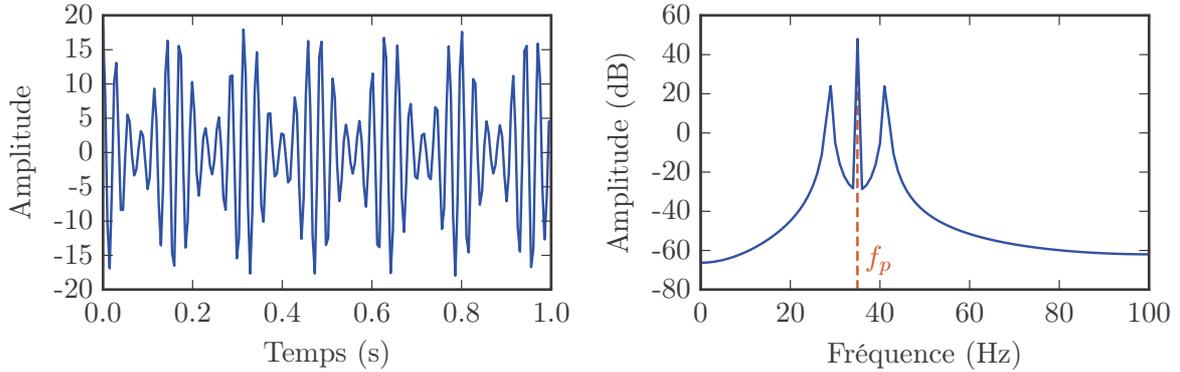


FIGURE 1.3 – La modulation en amplitude illustrée dans le domaine temporel à gauche se traduit dans le domaine fréquentiel (à droite) par deux pics également espacés de  $f_m = 6,3$  Hz autour de la fréquence porteuse  $f_p = 35$  Hz.

Dans le cas de la modulation angulaire, l'amplitude du signal porteur est constante  $A(t) = A_p$ . La phase instantanée quant à elle peut être modifiée de deux manières différentes. On parle de modulation de phase lorsque la phase du signal est modulée :

$$s_p(t) = A_p \cos(2\pi f_p t + \beta m(t)). \quad (1.10)$$

Dans le cas où la modulation est réalisée sur la fréquence instantanée, dérivée de la phase instantanée, on parle alors de modulation de fréquence :

$$f_i(t) = \frac{1}{2\pi} \frac{d\Phi(t)}{dt} = f_p + f_\Delta m(t),$$

avec  $f_\Delta$  représentant la déviation maximale en fréquence par rapport à la fréquence porteuse  $f_p$ . Les structures spectrales engendrées par ces deux modulations angulaires sont identiques, du moins dans le cas de signaux modulateurs périodiques. En effet, si on considère une modulation en fréquence, le signal modulé s'écrit alors

$$\begin{aligned} s_p(t) &= A_p \cos\left(2\pi \int_0^t f_i(\tau) d\tau\right) \\ &= A_p \cos\left(2\pi \int_0^t f_p + f_\Delta m(\tau) d\tau\right) \\ &= A_p \cos\left(2\pi f_p t + 2\pi f_\Delta \int_0^t m(\tau) d\tau\right). \end{aligned} \quad (1.11)$$

Les équations (1.10) et (1.11) nous montrent que les modulations de phase et de fréquence ont une formulation similaire. Moduler en phase revient à dériver le signal modulateur et à moduler

en fréquence. Si le signal modulant est sinusoïdal, sa dérivée sera également sinusoïdale : la modulation de phase et la modulation de fréquence sont alors identiques à un facteur près. C'est également le cas pour les signaux modulateurs périodiques, qui peuvent se décomposer en séries de Fourier et dont la dérivée sera alors identique à un facteur près.

La figure 1.4 montre le cas particulier d'une modulation angulaire par un modulant sinusoïdal,  $m(t) = \cos(2\pi f_m t)$ . Le signal peut alors s'écrire sous la forme

$$\begin{aligned} s_p(t) &= A_p \cos(2\pi f_p t + \beta \sin(2\pi f_m t)) \\ &= A_p [\cos(2\pi f_p t) \cos(\beta \sin(2\pi f_m t)) - \sin(2\pi f_p t) \sin(\beta \sin(2\pi f_m t))] \\ &= A_p \sum_{n=-\infty}^{+\infty} J_n(\beta) \cos(2\pi(f_p + n f_m)t), \end{aligned} \quad (1.12)$$

avec  $J_n(\cdot)$  la fonction de Bessel de première espèce. La dernière transition s'effectue par décomposition en séries de Fourier des fonctions  $\cos(\beta \sin(2\pi f_m t))$  et  $\sin(\beta \sin(2\pi f_m t))$ . Ce résultat nous apprend que la modulation angulaire fait apparaître une infinité de pics spectraux, à des fréquences régulièrement espacées autour de la fréquence porteuse :  $\forall n, f = f_p \pm n \times f_m$ . Cependant, l'amplitude de ces fréquences décroît rapidement avec le rang  $n$  des bandes de modulation.

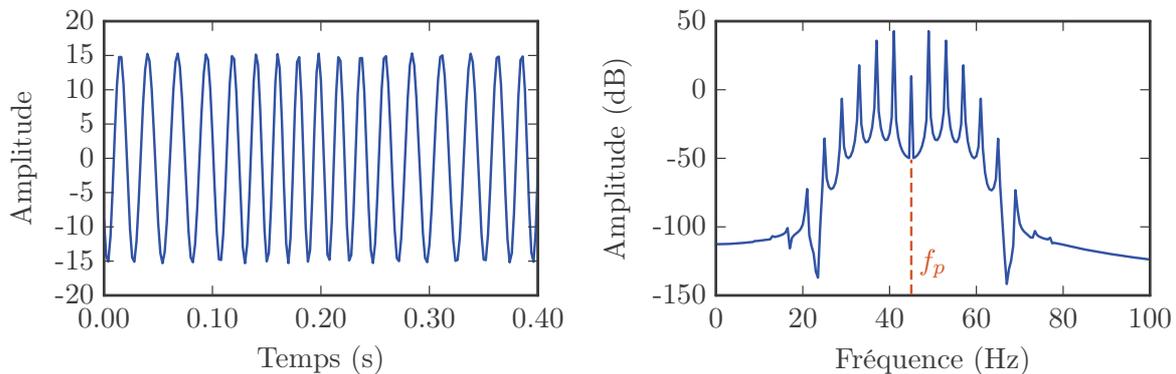


FIGURE 1.4 – La modulation angulaire se traduit dans le domaine fréquentiel (à droite) par une infinité de pics symétriquement espacés autour d'une fréquence porteuse, ici,  $f_p = 45$  Hz.

Enfin, il est important de noter qu'un signal peut être modulé à la fois en amplitude et en angle. L'un n'empêche pas l'autre.

#### 1.2.2.4 L'importance des structures spectrales dans la modélisation d'un système mécanique

Les structures spectrales présentées précédemment sont abondamment présentes dans la littérature sur la détection de défauts mécaniques. En effet, les mouvements dans les machines tournantes sont quasiment circulaires. Les imperfections lors de la conception des pièces, leur usure ou un défaut naissant ont pour conséquence l'apparition d'impulsions ou de modula-

tions, prenant la forme de séries harmoniques ou de bandes latérales dans la densité spectrale des signaux vibratoires [Mark 1978, Randall 1982, McFadden 1986]. De plus, la cinématique de la machine permet de calculer les fréquences caractéristiques où apparaîtront ces structures. L'identification et l'analyse des structures spectrales permet donc de caractériser l'état du système à surveiller. Il est donc primordial de les identifier dans la densité spectrale des signaux recueillis.

La figure 1.5 nous montre un zoom sur une densité spectrale de puissance estimée d'un signal vibratoire, acquis sur un banc de test simulant le fonctionnement d'une éolienne. Le nombre de pics, de séries harmoniques et de bandes de modulation présents dans le spectre est considérable. De plus, la détection de tous ces éléments spectraux est rendue complexe par la présence d'un bruit important.

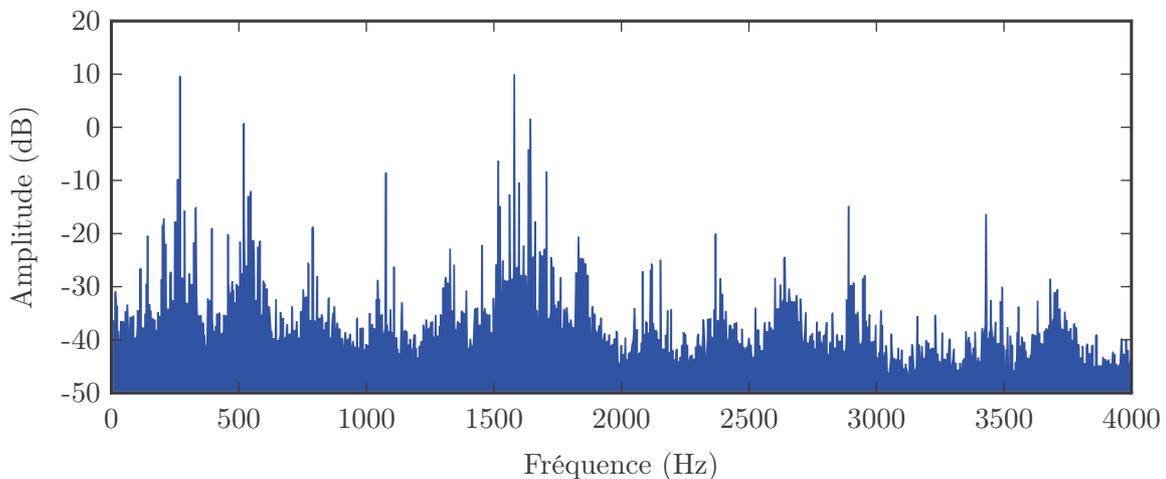


FIGURE 1.5 – Estimation de la densité spectrale de puissance d'un signal vibratoire, acquis sur un banc de test simulant le fonctionnement d'une éolienne. L'estimation a été réalisée par la méthode du périodogramme, sur un signal d'un million de points, échantillonné à 39062,5 Hz.

Suite à tous ces rappels sur l'analyse spectrale et l'importance de l'identification des structures dans le spectre, symptômes d'un défaut sur le système surveillé, nous allons maintenant nous focaliser sur l'approche proposée. Cette dernière se base sur des travaux antérieurs et respecte au mieux les contraintes exposées dans la section 1.1.1.

### 1.2.3 Détection des pics

Les travaux de thèse de Matthieu Durnerin [Durnerin 1999] se sont focalisés sur la détection de pics dans le spectre. Comme il n'existe aucune méthode d'estimation du spectre qui soit optimale pour tous les types de signaux qu'on peut rencontrer, l'idée a été de confronter plusieurs méthodes d'estimation différentes et de se servir des avantages de chacune des méthodes.

Pour chaque estimateur spectral, une première identification des pics est réalisée en ex-

trayant l'ensemble des maxima de l'estimation. Il est clair que certains de ces maxima ne correspondent pas à des pics, mais sont en réalité du bruit d'estimation. C'est pourquoi un test statistique permet alors de décider pour chaque maximum s'il s'agit d'un pic ou simplement d'un pic de bruit. Enfin, l'estimation spectrale des pics détectés comme tels est comparée à la fenêtre spectrale de l'estimateur sur-échantillonné avec pour objectifs de 1) classer le pic dans des catégories : fréquence pure, bande étroite ou encore bruit ; et de 2) améliorer l'estimation de la fréquence du pic en minimisant l'erreur quadratique entre le pic et la fenêtre spectrale sur-échantillonnée.

Une fois l'identification des pics réalisée par chaque estimateur, une opération de fusion synthétise les résultats obtenus et crée une carte d'identité spectrale pour chacun des pics détectés. Ainsi, la stratégie proposée par Matthieu Durnerin permet d'identifier dans un signal  $s[k]$  quelconque, sans aucun *a priori*, l'ensemble  $\mathcal{P}$  des pics spectraux présents dans le spectre :

$$\mathcal{P} = \{p^i\}_{i \in \{1, \dots, \#\mathcal{P}\}}, \quad (1.13)$$

où  $p^i$  représente le  $i^{\text{ème}}$  pic de l'ensemble  $\mathcal{P}$  contenant un total de  $\#\mathcal{P}$  pics. Chaque pic  $p^i$  possède sa propre « carte d'identité spectrale » sur laquelle figurent les caractéristiques estimées du pic. Parmi les caractéristiques qui nous serviront par la suite se trouvent son amplitude  $a^i$ , sa fréquence  $f^i$  et l'intervalle d'incertitude  $\varepsilon^i$  sur l'estimation de sa fréquence.

Il est important de noter qu'une hypothèse forte est faite lors de la détection des pics. Les signaux analysés doivent être stationnaires, sans quoi les estimateurs à base de transformée de Fourier ne seront pas efficaces. Un module de pré-analyse permet de s'assurer de la stationnarité des données avant de les traiter. D'autres tests sont également effectués, afin de s'assurer qu'il n'y a pas eu de saturation au moment de l'acquisition ou qu'un filtre anti-repliement a bien été utilisé. Enfin, quelques informations sur le signal sont calculées, comme sa périodi-

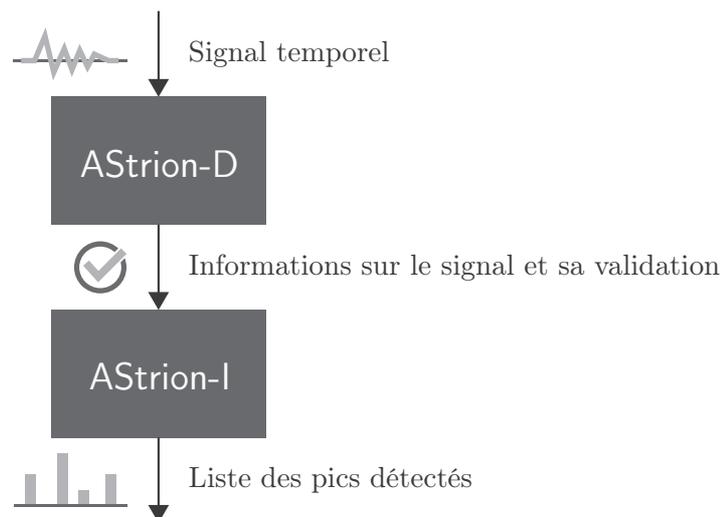


FIGURE 1.6 – L'analyseur de spectre automatique AStrion, au début des travaux de thèse, en octobre 2012.

cité ou son Rapport Signal sur Bruit (RSB) et servent à choisir les meilleurs estimateurs à confronter pour l'identification des pics.

Ces travaux, antérieurs à cette thèse, forment les deux premiers modules AStrion-D et AStrion-I de la méthode globale de surveillance automatique, nommée AStrion. La figure 1.6 illustre l'état du projet AStrion au début des travaux présentés dans ce manuscrit et les sections B.2.2 et B.2.4 de l'annexe B détaillent le fonctionnement de ces deux modules.

### 1.2.4 Identification des séries harmoniques et des bandes de modulation

L'ensemble  $\mathcal{P}$  contient la liste de tous les pics détectés à partir des multiples estimations du spectre. Si cette liste correspond bien à l'ensemble de l'information spectrale qu'il est possible d'extraire du signal, elle n'en reste pas moins très brute. En effet, parmi cette liste se trouvent des sous-ensembles de pics formant des structures spectrales plus complexes. Ces sous-ensembles de pics ont une origine commune et les considérer dans leur globalité donne plus de sens à cette information que de considérer chaque pic indépendamment. Nous proposons donc d'identifier les séries harmoniques et les bandes de modulation parmi l'ensemble  $\mathcal{P}$  des pics détectés.

Une série harmonique  $h^i$  est un ensemble de pics spectraux dont les fréquences sont liées entre elles par une relation de multiplicité. D'une manière similaire, les fréquences des pics dans une bande de modulation  $m^j$  sont également liées par une relation mathématique. Cependant, même si les relations mathématiques sont clairement définies, le problème d'identification des séries harmoniques et des bandes de modulation reste un problème complexe pour les raisons suivantes :

- les fréquences des pics sont estimées, il y a donc une incertitude quant à leur position fréquentielle exacte ;
- l'ensemble  $\mathcal{P}$  est très dense, le nombre de pics détectés est important.

Des méthodes d'identification des structures spectrales existent et sont expliquées dans la section 2.1.2. Cependant, leurs limitations, à savoir la non-prise en compte de ces deux difficultés et la non-exhaustivité de la recherche des structures, nous ont invités à proposer, dans le chapitre 2, une nouvelle méthode d'identification des séries harmoniques et des bandes de modulation. L'ensemble  $\mathcal{H}$  contient alors la liste de toutes les séries  $h^i$  détectées. De même, l'ensemble  $\mathcal{M}$  contient l'ensemble des bandes de modulation  $m^i$  identifiées.

## 1.3 Analyse d'une séquence de signaux

Jusqu'à présent, nous nous sommes intéressés à la problématique de l'extraction du contenu spectral d'un unique signal. Cependant, dans un système de surveillance, on n'observe pas un unique signal, mais plusieurs signaux acquis successivement, de façon périodique ou non, tout au long de la surveillance du système. Nous utiliserons la notation indicielle pour identifier

les signaux. Ainsi, le signal  $s_k$  sera le  $k^{\text{ème}}$  signal acquis. La séquence  $\mathcal{S}$  contiendra l'ensemble des  $n$  signaux acquis jusqu'à présent :  $\mathcal{S} = \{s_1, \dots, s_n\}$ .

Chaque signal  $s_k \in \mathcal{S}$  est acquis à partir d'une date  $t_k$  et pendant une durée  $\tau_k$ . La séquence de signaux peut être :

- régulière, les dates d'acquisition sont prédéfinies ( $t_k = t_1 + k\Delta t$ , avec  $\Delta t$  une durée arbitrairement choisie) ;
- continue, le signal suivant démarre au moment où le précédent termine ( $t_{k+1} = t_k + \tau_k$ ) ;
- continue avec recouvrement, le signal suivant démarre alors que l'acquisition du précédent est encore en cours ( $t_{k+1} < t_k + \tau_k$ ) ;
- irrégulière, il n'y a aucun lien entre les dates d'acquisition.

La seconde étape de la surveillance automatique consiste à analyser la séquence de signaux dans sa globalité. Nous cherchons à suivre un même contenu spectral sur l'ensemble de la séquence. Autrement dit, nous cherchons à identifier la trajectoire qu'un pic ou qu'une structure spectrale va suivre de signal en signal.

### 1.3.1 Suivi temporel des structures spectrales

Lors de l'analyse de la séquence  $\mathcal{S}$  de signaux, chaque signal  $s_k \in \mathcal{S}$  est analysé individuellement. Le contenu spectral de chaque signal est extrait et enregistré dans les trois ensembles  $\mathcal{P}_k$ ,  $\mathcal{H}_k$  et  $\mathcal{M}_k$ . L'ensemble  $\mathcal{P}_k$  contient la liste des pics détectés dans le spectre de  $s_k$ . L'ensemble  $\mathcal{H}_k$  contient la liste des séries harmoniques identifiées parmi l'ensemble  $\mathcal{P}_k$  des pics détectés. De même,  $\mathcal{M}_k$  contient la liste des bandes de modulation identifiées parmi l'ensemble  $\mathcal{P}_k$ .

Pour comprendre l'évolution du système surveillé, il est important de suivre chaque structure spectrale pour voir comment elle évolue avec le temps. C'est cette évolution qui va nous permettre de caractériser l'apparition d'un défaut ou l'usure du système. La figure 1.7 illustre la problématique. Elle montre l'évolution d'un spectre fictif au cours du temps. Pour pouvoir comparer ces spectres, il est utile d'introduire la notion de trajectoire. Une trajectoire permet de suivre une structure spectrale au cours du temps. Il s'agit donc d'un nouvel ensemble qui possède au maximum une structure spectrale par signal présent dans la séquence. Le rôle de la trajectoire est de suivre cette structure spectrale au travers des différents signaux, afin de permettre la visualisation de son évolution.

Les structures spectrales que nous avons sont de trois types différents. Nous aurons donc trois types de trajectoires : les trajectoires de pics, les trajectoires de séries harmoniques et les trajectoires de bandes de modulation. L'ensemble  $\mathcal{T}_{\mathcal{P}}$  contient la liste de toutes les trajectoires de pics  $t_{\mathcal{P}}^i$ . Cet ensemble n'est plus attaché à un unique signal, mais à la séquence  $\mathcal{S}$  dans sa globalité. Il en va de même pour les ensembles  $\mathcal{T}_{\mathcal{H}}$  et  $\mathcal{T}_{\mathcal{M}}$ , regroupant les trajectoires de séries harmoniques  $t_{\mathcal{H}}^i$  et les trajectoires de bandes de modulation  $t_{\mathcal{M}}^i$ , respectivement.

La figure 1.8 reprend l'illustration du problème de la figure 1.7. Elle y ajoute une proposition de trajectoires : en orange, une trajectoire de pics et en bleu, une trajectoire harmonique.

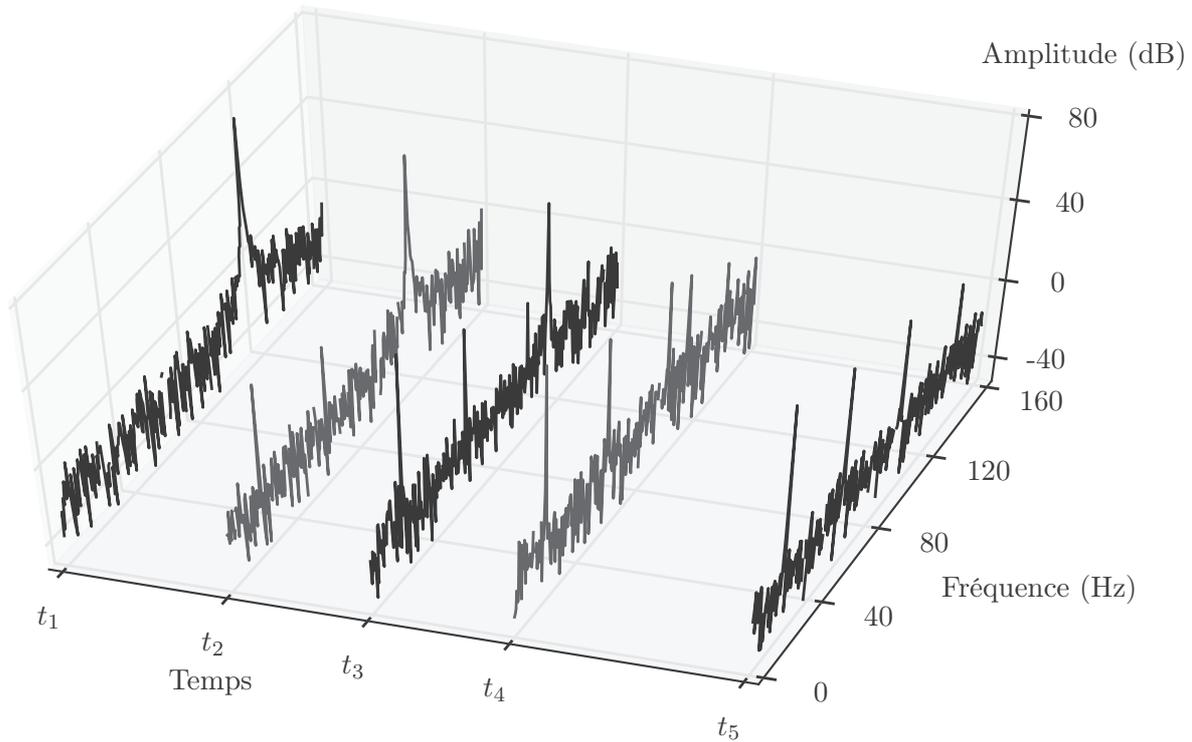


FIGURE 1.7 – Évolution d'un spectre fictif au cours du temps. Les signaux sont acquis de manière irrégulière.

Cependant, une série harmonique étant un ensemble de pics, la trajectoire harmonique peut être également vue comme une combinaison de plusieurs trajectoires de pics. Cette imbrication entre les structures ajoute une difficulté supplémentaire lors de l'identification des trajectoires.

Le suivi du contenu spectral et l'identification des trajectoires sont détaillés dans le chapitre 3. Les principales difficultés d'identification des trajectoires sont dues au fait que :

- il y a trois types de structures différentes à suivre ;
- les structures spectrales sont imbriquées les unes dans les autres ;
- le nombre de trajectoires varie suite aux naissances et aux disparitions de pics au cours des différentes analyses successives.

### 1.3.2 Diagnostic du système surveillé

L'identification des trajectoires permet de suivre le contenu spectral des signaux au cours du temps. Cependant, afin d'arriver au diagnostic du système, il faut réussir à caractériser son état de santé. Cette étape passe par l'établissement d'indicateurs, permettant de rendre compte de l'évolution des structures spectrales dans le spectre. La figure 1.9 illustre la problématique sur un exemple fictif. Le pic à 250 Hz est entouré de modulations. Lorsque le pseudo-système était sain, les modulations étaient quasiment inexistantes alors qu'en pré-

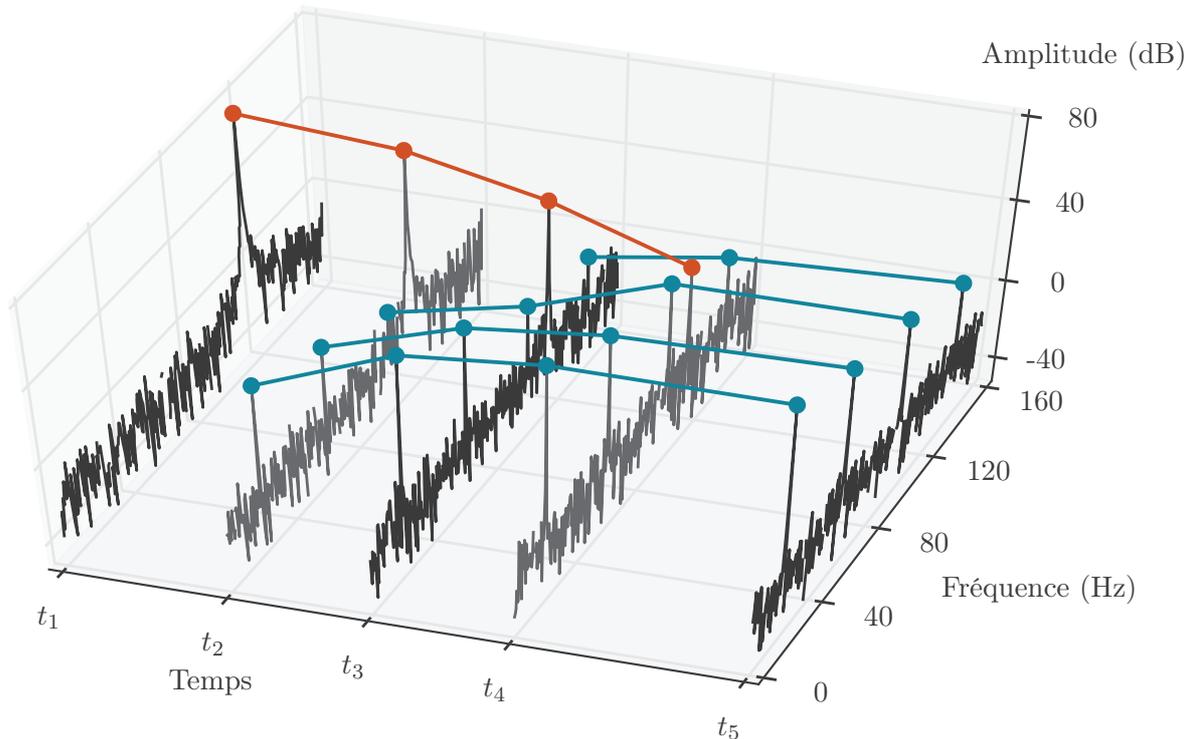


FIGURE 1.8 – Illustration du principe de trajectoire spectrale : en orange, une trajectoire de pics et en bleu, une trajectoire harmonique.

sence du défaut, de nombreuses bandes sont présentes. Cette évolution doit être caractérisée sous la forme d'un ou plusieurs indicateurs afin de permettre l'évaluation automatique de la sévérité du défaut.

Si chaque trajectoire spectrale débouche sur la création d'un ou plusieurs indicateurs afin de surveiller un point particulier du système, la surveillance automatique mène alors à la génération de nombreux indicateurs. Ces derniers doivent alors être évalués afin de faire ressortir en priorité ceux qui portent en eux une information sur l'évolution du système et son éventuelle dégradation.

Afin d'évaluer la quantité d'information portée par un indicateur, l'hypothèse suivante est faite : un indicateur n'évoluant pas ou peu portera peu d'information sur la dégradation du système alors qu'un défaut sera marqué par l'évolution drastique d'un ou plusieurs indicateurs. Cela nous mène alors à devoir caractériser l'évolution des indicateurs dans le temps.

L'étape de diagnostic sera détaillée dans le chapitre 4. Les principaux défis à relever viennent du fait que :

- les trajectoires sont de natures différentes, il faut donc générer des indicateurs différents en fonction de la structure spectrale suivie ;
- les trajectoires naissent et meurent, les indicateurs ne sont alors pas toujours continus en temps ;

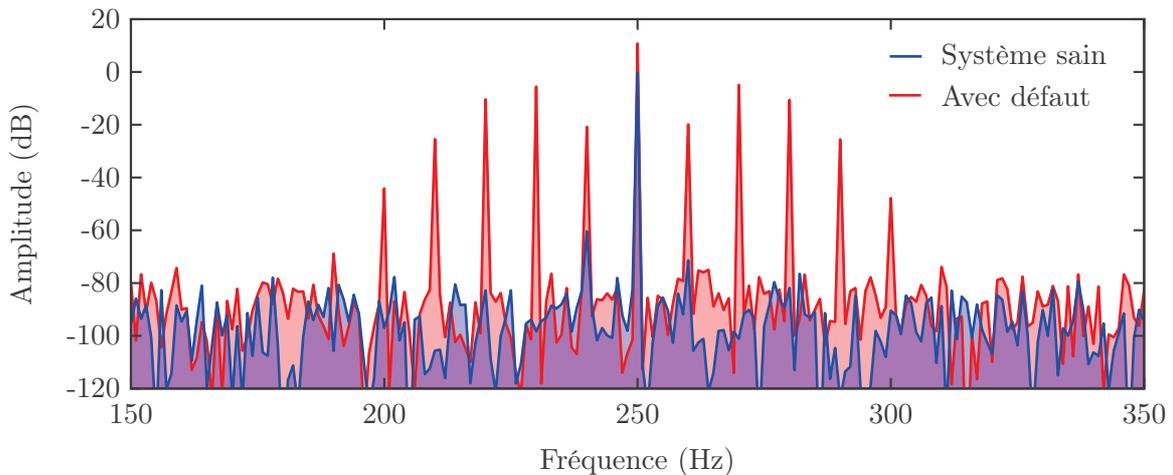


FIGURE 1.9 – Comparaison de deux spectres : le spectre en bleu a été acquis sur le système sain ; il sert de référence. Le développement des modulations sur le spectre rouge nous indique la présence d'un défaut.

- le nombre d'indicateurs est important, il faut alors identifier ceux qui sont porteurs d'une information sur l'évolution du système surveillé.

## 1.4 Synthèse du chapitre

L'objectif des travaux développés au cours de cette thèse est de proposer une méthode automatique de surveillance d'un système mécanique. Cette méthode doit être à la fois générique, capable de fonctionner en-ligne et ne pas nécessiter des ressources processeurs et mémoires trop importantes. Elle sera intégrée dans un CMS, enregistrant des signaux régulièrement (ou non). La méthode doit donc répondre à la problématique suivante : comment transformer la séquence  $\mathcal{S} = \{s_1, \dots, s_n\}$  de signaux en un diagnostic du système surveillé ?

Ce problème est découpé en plusieurs étapes. Les deux étapes principales consistent à 1) analyser un signal pour en extraire toute l'information contenue, et 2) confronter cette information aux réalisations précédentes afin de mettre à jour le bilan de santé du système surveillé.

L'analyse d'un signal se base sur des travaux antérieurs réalisés au laboratoire GIPSA-lab, notamment l'analyseur de spectre automatique AStrion. Ce dernier identifie l'ensemble  $\mathcal{P}_k$  des pics spectraux présents dans un signal  $s_k$ . Afin de compléter cette lecture de spectre, il est nécessaire d'identifier les structures spectrales présentes dans cette liste de pics. Le chapitre 2 répondra donc à la question de l'identification des séries harmoniques et des bandes de modulation au sein d'un ensemble de pics.

Ensuite, la phase de comparaison des données se subdivise en deux problématiques. D'une part, le chapitre 3 répondra au problème de suivi des structures spectrales et de création des

---

trajectoires. D'autre part, la génération d'indicateurs sur l'état de santé du système et leur utilisation pour le diagnostic seront abordés dans le chapitre 4.



# Identification automatique des séries harmoniques et des bandes de modulation

---

## Sommaire

---

<b>2.1</b>	<b>Contexte . . . . .</b>	<b>26</b>
2.1.1	Position du problème . . . . .	26
2.1.2	État de l'art sur l'identification des structures spectrales . . . . .	27
<b>2.2</b>	<b>Identification des séries harmoniques . . . . .</b>	<b>28</b>
2.2.1	Recherche par intersection d'intervalles . . . . .	28
2.2.2	Mise à jour des paramètres de recherche . . . . .	31
2.2.3	Critères d'arrêt . . . . .	32
2.2.4	Validation d'une série harmonique . . . . .	33
2.2.5	Cas de l'absence du fondamental . . . . .	34
2.2.6	Post-traitement : suppression des sous-séries . . . . .	35
2.2.7	Synthèse sur la recherche d'harmoniques . . . . .	36
<b>2.3</b>	<b>Identification des bandes de modulation . . . . .</b>	<b>36</b>
2.3.1	Recherche unilatérale de modulations . . . . .	38
2.3.2	Fusion des bandes latérales . . . . .	40
<b>2.4</b>	<b>Caractérisation des structures spectrales . . . . .</b>	<b>43</b>
2.4.1	Carte d'identité d'une série harmonique . . . . .	43
2.4.2	Carte d'identité d'une bande de modulation . . . . .	45
<b>2.5</b>	<b>Applications . . . . .</b>	<b>46</b>
2.5.1	Choix des paramètres de recherche . . . . .	46
2.5.2	Simulation d'un défaut . . . . .	48
2.5.3	Surveillance d'un système mécanique . . . . .	49
<b>2.6</b>	<b>Synthèse du chapitre . . . . .</b>	<b>51</b>

---

Les structures spectrales, comme les séries harmoniques ou les bandes de modulation, sont porteuses d'information. C'est par exemple le cas en télécommunications, où le signal utile est modulé/démodulé afin d'être transmis dans une bande fréquentielle adaptée au canal. C'est également le cas en mécanique vibratoire, où la plupart des modèles de vibrations des machines tournantes sont basés sur des séries harmoniques et des modulations. Il est donc

primordial d'identifier la présence de telles structures dans la densité spectrale d'un signal, afin d'en extraire toute l'information utile.

Dans ce chapitre, nous ferons dans un premier temps un état de l'art des méthodes de recherche de séries harmoniques et de bandes de modulation. Dans un second temps, nous proposerons une méthode d'identification des séries harmoniques parmi un ensemble de pics estimés. Cette méthode sera réutilisée afin d'identifier, dans le même ensemble, les bandes de modulation présentes au sein du signal. Ensuite, nous caractériserons toutes les structures spectrales estimées afin de pouvoir plus facilement les comparer. Enfin, nous validerons l'approche proposée en l'appliquant sur des signaux simulés et réels.

## 2.1 Contexte

L'analyse d'un signal consiste à en extraire toute l'information utile. Dans le cas de la surveillance d'un système mécanique, nous nous intéresserons aux informations prenant la forme de pics et de structures de pics dans la densité spectrale du signal. L'identification automatisée de ces structures est une opération primordiale dans le cadre de la surveillance automatique.

### 2.1.1 Position du problème

Comme évoqué dans la section 1.2.3, la détection des pics au sein de la densité spectrale d'un signal est un problème qui a déjà été abordé au [GIPSA-lab](#)<sup>1</sup>, en collaboration avec l'[IRIT](#)<sup>2</sup>, au cours de la thèse de Matthieu Durnerin. Ce dernier propose de confronter plusieurs estimateurs du spectre et de fusionner les résultats, afin d'avoir une méthode applicable à tout type de signal, sans nécessiter de paramétrer quoi que ce soit. Les annexes B.2.2 et B.2.4 détaillent le fonctionnement de cette méthode, qui permet de détecter l'ensemble  $\mathcal{P}$  des pics présents dans le spectre d'un signal  $s$ .

Chacun des pics  $p^i \in \mathcal{P}$  possède une *carte d'identité spectrale* synthétisant les caractéristiques estimées du pic. Le tableau 2.1 donne la liste restreinte des caractéristiques utilisées dans la suite du manuscrit. En particulier, on y trouve la fréquence estimée  $f^i$ . D'après le principe d'incertitude [[Heisenberg 1927](#)], la véritable position fréquentielle d'un pic  $p^i$  ne peut être calculée que si l'on possède un signal d'une durée infinie. Ce ne sera jamais le cas dans un contexte d'estimation. Cependant, nous savons que la véritable fréquence du pic se situe au sein d'un intervalle d'indétermination de largeur  $\varepsilon^i$ , centré sur la fréquence estimée  $f^i$ .

La méthode de détection nous renvoie un ensemble  $\mathcal{P}$  ordonné. Les pics  $p^i$  sont classés selon leur fréquence  $f^i$ , par ordre croissant. De plus, l'ensemble  $\mathcal{P}$  est très dense, c'est-à-dire que le nombre de pics détectés, noté  $\#\mathcal{P}$ , est en général très grand (plusieurs milliers de pics). Cela vient du fait que les seuils de détection sont volontairement très bas, juste au-dessus

---

1. Grenoble, Images, Parole, Signal et Automatique  
2. Institut de Recherche en Informatique de Toulouse

Caractéristique du pic $p^i$	Notation
Amplitude du pic	$a^i$
Fréquence du pic	$f^i$
Intervalle d'indétermination sur la fréquence	$\varepsilon^i$

Tableau 2.1 – Carte d'identité spectrale restreinte d'un pic  $p^i$ .

de la variance du bruit. Cela garantit la détection de toutes les informations pertinentes qui pourraient mener à la détection précoce d'un défaut. Cependant, cela se fait au détriment de fausses détections.

$$\mathcal{P} = \{p^1, p^2, \dots, p^{\#\mathcal{P}} \mid \forall (i, j) \in \{1, \dots, \#\mathcal{P}\}^2, i < j \Rightarrow f^i < f^j\}. \quad (2.1)$$

L'ensemble  $\mathcal{P}$  contient tous les pics détectés à partir du spectre. Cette liste représente l'ensemble des informations extractibles à partir du signal. Cependant, ces informations sont sous une forme brute. En effet, dans le cas où un pic appartient à une structure spectrale, il est plus intéressant de considérer l'ensemble des pics formant la structure plutôt que de considérer chaque pic indépendamment. La problématique à résoudre est donc l'identification, au sein d'un ensemble  $\mathcal{P}$ , des sous-ensembles de pics formant des séries harmoniques et des bandes de modulation. Cette identification est d'autant plus complexe que les paramètres des pics sont estimés et que nous ne souhaitons pas fournir d'informations *a priori* sur l'emplacement ou le nombre de séries à détecter.

### 2.1.2 État de l'art sur l'identification des structures spectrales

La recherche de séries harmoniques au sein de la densité spectrale d'un signal est une problématique très présente dans la littérature. En effet, les harmoniques se retrouvent dans plusieurs types de signaux issus de nombreux domaines d'applications. C'est par exemple le cas des signaux de parole pour lesquels la voix présente naturellement une structure harmonique. C'est également le cas de nombreux enregistrements musicaux où certains instruments produisent des notes prenant la forme d'une série harmonique, dont la fréquence fondamentale dépend de la hauteur de la note. Retrouver la hauteur de la note ou de la voix représente en général la première étape d'un traitement automatique de ce type de signaux.

Les méthodes identifiant la série harmonique principale sont nombreuses. Dans [Schroeder 1968], l'auteur cherche à identifier la fréquence fondamentale du signal, même s'il n'y a pas de pic spectral à cette fréquence. Pour cela, il utilise un histogramme des fréquences présentes dans le signal, ainsi que leurs sous-multiples. L'histogramme fait alors apparaître le fondamental comme étant la fréquence ayant cumulé le plus d'harmoniques et de sous-harmoniques. Dans [Martin 1982, Wang 2000], la corrélation de la densité spectrale avec un peigne de raies est utilisée afin de trouver la fréquence fondamentale. Seule une série harmonique est donc identifiée, à partir d'un modèle imposant l'amplitude des harmoniques en fonction de leur

rang. Une corrélation modifiée est également utilisée dans [Redortier 1987] pour détecter cette fois-ci l'ensemble des séries harmoniques du spectre. Cependant, leur validation est sujette à l'expertise d'un utilisateur averti. Dans [Doval 1991], les auteurs s'intéressent à la recherche, parmi un ensemble de pics, de la série harmonique prédominante. La méthode du maximum de vraisemblance leur permet d'identifier cette unique série. Dans [Maher 1994], la méthode permet de détecter plusieurs séries harmoniques en s'aidant d'informations temporelles. Cependant, le nombre de séries à détecter doit être fourni comme paramètre. Enfin, les méthodes proposées dans [Zahorian 2008, Fujimoto 2012] ne permettent pas d'identifier la série principale en entier. Dans [Zahorian 2008], la densité spectrale est comparée à un modèle de série, avec un nombre d'harmoniques limité sous l'hypothèse que les premiers harmoniques portent quasiment toute l'énergie de la série. Dans [Fujimoto 2012], seule la fréquence fondamentale et les harmoniques de rang 2 et 3 sont détectés à partir d'un filtrage particulière.

Les méthodes présentées ici ne répondent pas complètement au cahier des charges fixé dans le contexte de cette thèse. En effet, si certaines méthodes identifient bien des séries harmoniques au sein d'un ensemble de pics, le nombre de séries détectées est limité ou doit alors être renseigné. De plus, ces méthodes ne prennent pas en compte l'indétermination liée à l'estimation de la position fréquentielle des pics. Ces raisons nous ont poussés à proposer une nouvelle méthode détaillée dans la section suivante.

## 2.2 Identification des séries harmoniques

La méthode proposée dans cette section cherche à identifier l'ensemble  $\mathcal{H}$  des séries harmoniques, que nous noterons  $h^j$ . Ces séries se trouvent dans l'ensemble  $\mathcal{P}$  des pics  $p^i$ , dont la position fréquentielle exacte est comprise dans un intervalle d'indétermination d'amplitude  $\varepsilon^i$  centré sur la fréquence estimée  $f^i$ . Chaque série harmonique  $h^j$  est identifiée comme étant un sous-ensemble de  $\mathcal{P}$  dont les pics ont une fréquence multiple d'une fréquence fondamentale donnée. L'indétermination quant à la fréquence exacte des pics se propage et la fréquence fondamentale possède elle aussi un intervalle d'indétermination.

La présentation de la méthode est progressive. Elle commence par la présentation du point clef, qui est la recherche d'harmoniques par intersection d'intervalles. Elle continue par la proposition de raffinements, permettant de répondre à des problématiques de précision et de performance. Une synthèse est finalement présentée, schématisant l'algorithme dans sa globalité et récapitulant l'ensemble des paramètres de la méthode proposée.

### 2.2.1 Recherche par intersection d'intervalles

La méthode proposée de recherche de séries harmoniques est exhaustive. Cela signifie que quel que soit le nombre de séries présentes, la méthode tente de toutes les identifier. Cela implique que chaque pic doit être considéré comme le fondamental potentiel d'une série harmonique. La méthode fonctionne de manière itérative et considère chaque pic  $p^i \in \mathcal{P}$  à

tour de rôle comme étant le fondamental d'une hypothétique série. La méthode démarre par le pic de plus petite fréquence, à savoir  $p^1$ , et termine par celui ayant la fréquence la plus haute,  $p^{\#\mathcal{P}}$ . En pratique, la recherche s'arrête auparavant comme cela sera détaillé dans la section 2.2.4.

Supposons que la recherche d'harmoniques a été réalisée pour les pics  $p^1$  à  $p^{i-1}$ . Soit le pic  $p^i$  le point de départ d'une série potentielle dont la fréquence fondamentale est initialisée à la fréquence estimée du pic. De même, l'amplitude de l'intervalle d'indétermination quant à l'exactitude de la fréquence fondamentale est initialisée à celle du pic.

La recherche des harmoniques du pic  $p^i$  consiste à chercher le sous-ensemble  $\mathcal{P}^i \subset \mathcal{P}$  des pics ayant une fréquence non nulle multiple de  $f^i$ . Les fréquences étant estimées, la relation stricte de multiplicité ne peut être vérifiée. Nous allons alors nous servir de l'indétermination quant à la position fréquentielle exacte afin de modéliser un pic non plus comme un élément ponctuel à la position  $f^i$  sur l'échelle des fréquences, mais comme un intervalle d'amplitude  $\varepsilon^i$  centré sur  $f^i$  :

$$\left[ f^i - \frac{\varepsilon^i}{2}, f^i + \frac{\varepsilon^i}{2} \right]. \quad (2.2)$$

La relation de multiplicité change alors elle aussi de forme. Les multiples de la fréquence  $f^i$  sont les intervalles

$$\left[ r \left( f^i - \frac{\varepsilon^i}{2} \right), r \left( f^i + \frac{\varepsilon^i}{2} \right) \right], \text{ avec } r \in \mathbb{N}^* \text{ le rang du multiple.} \quad (2.3)$$

La recherche des harmoniques se fait alors par intersection d'intervalles. En effet, montrer que la fréquence  $f^k$  est multiple de rang  $r \in \mathbb{N}^*$  de  $f^i$  débute par calculer l'intervalle  $[f_{\min}, f_{\max}]$  résultat de l'intersection d'intervalles selon :

$$[f_{\min}, f_{\max}] \triangleq \left[ r \left( f^i - \frac{\varepsilon^i}{2} \right), r \left( f^i + \frac{\varepsilon^i}{2} \right) \right] \cap \left[ f^k - \frac{\varepsilon^k}{2}, f^k + \frac{\varepsilon^k}{2} \right]. \quad (2.4)$$

Si cette intersection d'intervalles est non nulle, c'est-à-dire si

$$[f_{\min}, f_{\max}] \neq \emptyset, \quad (2.5)$$

alors la fréquence  $f^k$  est considérée comme multiple de la fréquence  $f^i$ .

Tous les pics  $p^k \in \mathcal{P}$  vérifiant (2.5) pour  $r \in \mathbb{N}^*$  fixé sont des candidats au titre d'harmonique de rang  $r$  du pic  $p^i$ . Trois cas différents peuvent se présenter. Ils sont illustrés sur la figure 2.1 et détaillés dans la suite :

- (a) Il n'y a aucun candidat et dans ce cas, la série n'a pas d'harmonique de rang  $r$ .
- (b) Un seul pic est candidat ; il est alors automatiquement élu et ajouté à l'ensemble  $\mathcal{P}^i$ .
- (c) Plusieurs pics sont candidats ; il est nécessaire d'en élire un pour l'ajouter à l'ensemble  $\mathcal{P}^i$ . L'élection se fait selon la règle suivante : le pic  $p^k$  choisi est celui dont la fréquence estimée est la plus proche du modèle de l'harmonique de rang  $r$ . Il s'agit donc du pic

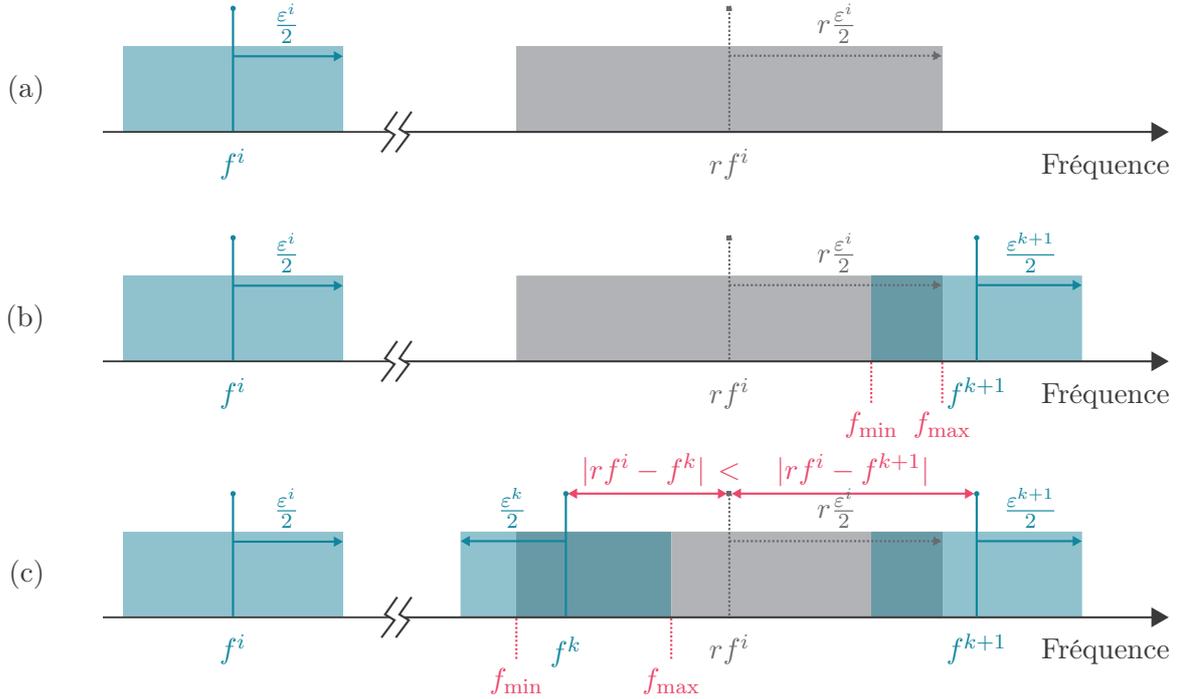


FIGURE 2.1 – Recherche de l'harmonique de rang  $r$  par intersection d'intervalles. Les trois cas possibles sont représentés : (a) aucun candidat n'est présent dans l'intervalle de recherche, (b) un pic est candidat et (c) deux pics sont candidats; celui le plus proche du modèle est élu. Les bornes  $f_{\min}$  et  $f_{\max}$  marquent l'intersection entre le modèle de l'harmonique de rang  $r$  et le pic élu.

dont la fréquence vérifie

$$f^k = \underset{f^l \text{ vérifiant (2.5)}}{\operatorname{argmin}} |r f^i - f^l|. \quad (2.6)$$

Il est intéressant de noter que pour être le plus générique possible, la recherche d'harmonique n'intègre qu'un critère sur la fréquence. Aucun autre paramètre des pics, comme l'amplitude, n'est utilisé. En effet, utiliser un autre paramètre spécialiserait le modèle d'une série harmonique et la méthode perdrait en généralité.

L'exhaustivité de la recherche des harmoniques signifie que pour un pic  $p^i$  donné, nous vérifions parmi tous les pics ayant une fréquence supérieure à celle de  $p^i$  s'ils en sont des harmoniques ou non. Comme l'ensemble  $\mathcal{P}$  des pics est trié par fréquence croissante, cela signifie que pour  $p^i$ , l'ensemble des pics  $\{p^{i+1}, p^{i+2}, \dots, p^{\#\mathcal{P}}\}$  est testé, ce qui représente  $\#\mathcal{P} - i$  intersection d'intervalles. Au total, la recherche des harmoniques pour tous les pics présents entraîne  $\frac{\#\mathcal{P}(\#\mathcal{P}-1)}{2}$  intersections d'intervalles.

Pour être le plus efficace possible, l'implémentation de la méthode de recherche des harmoniques de  $p^i$  est réalisée de la manière suivante : les intersections sont calculées par ordre de fréquence croissante, commençant donc par le pic  $p^{i+1}$  et terminant par  $p^{\#\mathcal{P}}$ . Une intersection d'intervalles nécessite 3 étapes de calcul. Pour vérifier si le pic  $p^k$ , de fréquence  $f^k$  avec

$k \in \{i + 1, \dots, \#\mathcal{P}\}$ , appartient à la série de fréquence fondamentale  $f^i$ , il faut :

1. Calculer le rang possible de l'harmonique :  $r = \left\lfloor \frac{f^k}{f^i} \right\rfloor$ , avec  $\lfloor \cdot \rfloor$  l'arrondi à l'entier le plus proche ;
2. Calculer l'intersection  $\left[ f^k - \frac{\varepsilon^k}{2}, f^k + \frac{\varepsilon^k}{2} \right] \cap \left[ r \left( f^i - \frac{\varepsilon^i}{2} \right), r \left( f^i + \frac{\varepsilon^i}{2} \right) \right]$  et vérifier qu'elle est non nulle avant de continuer ;
3. Si le pic précédent, à savoir  $p^{k-1}$  a déjà été proposé comme harmonique de rang  $r$ , vérifier laquelle des fréquences  $f^{k-1}$  ou  $f^k$  est la plus proche du modèle  $r f^i$  et élire le pic le plus proche.

Finalement, nous obtenons pour chaque pic  $p^i \in \mathcal{P}$ , un sous-ensemble  $\mathcal{P}^i \subset \mathcal{P}$  des pics dont la fréquence est multiple de  $f^i$ . Cependant, cette méthode de recherche par intersection d'intervalles comporte une faille. En effet, plus le rang de l'harmonique que l'on recherche est élevé, plus l'intervalle à intersecter est grand. Le risque d'associer un pic quelconque, qui n'est pas réellement un harmonique de  $f^i$ , augmente donc avec le rang  $r$ . Une mise à jour des paramètres de recherche, présentée dans la section suivante, permet d'éviter ce phénomène.

### 2.2.2 Mise à jour des paramètres de recherche

La recherche des harmoniques du pic  $p^i$  se fait par intersection d'intervalles. Cette méthode est basée sur l'utilisation de deux paramètres du pic : la fréquence  $f^i$  et l'amplitude  $\varepsilon^i$  de l'intervalle d'indétermination. Cependant, plus le rang  $r$  de l'harmonique recherché est grand, plus l'intervalle à intersecter devient grand. Il faut donc mettre à jour les paramètres servant à la recherche afin d'éviter l'agrandissement disproportionné de l'intervalle à intersecter. Plutôt que de modifier directement les valeurs des caractéristiques estimées du pic  $p^i$ , nous proposons d'introduire deux nouvelles notations qui nous serviront de paramètres de recherche : la fréquence fondamentale  $\nu$  de la série recherchée et l'amplitude  $\epsilon$  de l'intervalle d'indétermination quant à l'exactitude de la fréquence fondamentale. Ces deux paramètres de recherche sont initialisés à partir des caractéristiques du pic  $p^i$ , fondamental potentiel de la série recherchée :

$$\nu = f^i, \quad \epsilon = \varepsilon^i. \quad (2.7)$$

La recherche des harmoniques du pic  $p^i$  est réalisée de manière itérative, en commençant par rechercher les harmoniques de rangs faibles. Chaque fois qu'un harmonique est trouvé, cela signifie que (2.6) est vérifiée. L'intersection  $[f_{\min}, f_{\max}]$  entre le pic élu et le modèle harmonique est non nulle. Cette intersection est utilisée pour mettre à jour les paramètres de recherche, à savoir la fréquence fondamentale  $\nu$  de la série ainsi que l'amplitude  $\epsilon$  de son intervalle d'indétermination. Les paramètres de la série harmonique potentielle sont mis à jour de la manière suivante :

$$\nu = \frac{f_{\max} - f_{\min}}{2r}, \quad \epsilon = \frac{f_{\min} + f_{\max}}{2r}. \quad (2.8)$$

Le processus de mise à jour est illustré sur la figure 2.2. À chaque fois qu'un pic est

identifié comme étant un harmonique, les paramètres de recherche sont affinés, ce qui empêche l'accroissement linéaire de l'intervalle du modèle harmonique avec le rang  $r$ . Cependant, si le dernier harmonique trouvé est de rang  $r$  et qu'aucun harmonique n'est trouvé aux rangs  $r+1, r+2, \dots, r+r_\emptyset$ , où  $r_\emptyset$  représente le nombre de rangs consécutifs pour lesquels aucun candidat n'a été trouvé, l'intervalle de recherche finira tout de même par croître. Cela signifie qu'il faut prévoir un critère d'arrêt pour empêcher l'intervalle de recherche de croître indéfiniment.

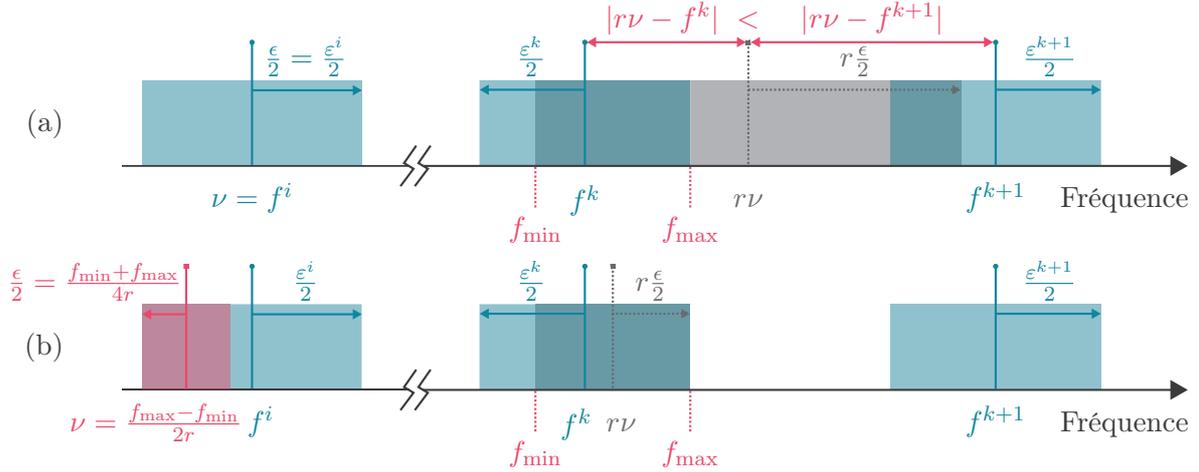


FIGURE 2.2 – Mise à jour des paramètres de la série harmonique, en deux étapes : (a) identification de l'harmonique de rang  $r$ , puis (b) mise à jour de la fréquence fondamentale  $\nu$  et de l'amplitude  $\epsilon$  de l'intervalle d'indétermination.

### 2.2.3 Critères d'arrêt

La méthode d'identification des harmoniques présentée jusqu'ici cherche pour un pic  $p^i$  donné, tous les harmoniques possibles parmi les pics restants de plus haute fréquence. Le fait de faire cette recherche de manière itérative en commençant par les pics dont la fréquence est proche du pic donné, permet d'associer en priorité les harmoniques de rangs faibles. Il est possible que dans une série, certains rangs soient absents. Cela ne gêne en rien la recherche de l'harmonique de rang suivant. Cependant, si le nombre  $r_\emptyset$  de rangs consécutifs absents devient grand, l'intervalle de recherche n'est plus mis à jour et commence à devenir très grand. Cela signifie également que la série est certainement terminée et qu'elle ne comportera plus d'autres harmoniques. Il n'est alors pas nécessaire de chercher les harmoniques suivants.

Un critère d'arrêt est donc mis en place. Il prend la forme d'une condition sur le nombre maximal de rangs consécutifs pouvant être absents. Si ce nombre, noté  $r_\emptyset^{\max}$ , est atteint, la recherche des harmoniques suivants est arrêtée. La valeur de ce paramètre est fixée et expliquée dans la section 2.5.1. Ce critère a un objectif double :

1. Éviter l'accroissement trop important de l'intervalle de recherche d'harmoniques et garantir que les résultats obtenus sont cohérents ;
2. Diminuer la complexité de l'algorithme de recherche et réduire le temps de calcul.

D'autres critères peuvent être imaginés et intégrés ici. Il est par exemple possible de stopper la recherche si l'amplitude des harmoniques identifiés n'est pas strictement décroissante ou ne suit pas un modèle prédéfini. Ce type de critère d'arrêt induit des *a priori* sur le type de séries harmoniques attendu et permettrait de spécialiser l'algorithme pour trouver des séries particulières. Cependant, dans notre démarche de réalisation d'un système de surveillance générique, de tels *a priori* ne sont pas souhaitables.

#### 2.2.4 Validation d'une série harmonique

La méthode proposée permet de rechercher les harmoniques potentiels d'un pic  $p^i$  donné. Ces harmoniques forment un sous-ensemble  $\mathcal{P}^i \subset \mathcal{P}$ . Il reste à déterminer si ce sous-ensemble forme une série harmonique. Pour cela, il faut des critères précis afin de définir ce qu'est une série harmonique. Dans la mesure où nous souhaitons que la recherche de séries soit la plus générique possible, nous n'allons définir qu'un unique critère. Il faut que le nombre de pics soit supérieur à un seuil fixé :

$$\#\mathcal{P}^i > \#\mathcal{P}^{\min}. \quad (2.9)$$

La discussion sur le choix de la valeur du seuil se trouve dans la section 2.5.1. Cependant, il est d'ores et déjà possible de remarquer que la valeur minimale de ce seuil est 2. En effet, s'il n'y a que le fondamental dans l'ensemble  $\mathcal{P}^i$ , on ne peut pas parler de série harmonique à proprement parler. Il faut nécessairement au moins un harmonique avant de commencer à parler de série. De plus, ce seuil nous donne un renseignement supplémentaire permettant de réduire le temps de calcul pour la recherche d'harmoniques. En effet, s'il faut un nombre minimal de pics pour être considéré comme une série, cela signifie que les pics ayant une fréquence supérieure à la fréquence maximale ( $f^{\#\mathcal{P}}$ ) divisée par ce seuil ne formeront jamais une série harmonique. Il est donc possible d'arrêter les itérations au pic  $p^i$  lorsque :

$$f^i > \frac{f^{\#\mathcal{P}}}{\#\mathcal{P}^{\min}}. \quad (2.10)$$

Cette étape de validation est intentionnellement très légère, permettant à de nombreuses séries d'être détectées, quelle que soit leur forme. Les séries étant des conséquences d'un défaut naissant ou d'une usure dans le système à surveiller, leurs détections précoces permettent d'avoir très tôt des informations sur l'état du système. Dans un autre contexte, des critères supplémentaires peuvent être ajoutés durant cette étape de validation afin de ne valider que les séries suivant un modèle harmonique prédéfini.

Si la série harmonique est validée au cours de cette étape, elle est alors étiquetée  $h^j$ , avec l'exposant  $j$  égal au nombre de séries validées jusqu'à présent. Les exposants  $i$  et  $j$  sont différents, car tous les pics  $p^i \in \mathcal{P}$  ne servent pas de point de départ à une série harmonique. La fréquence fondamentale  $\nu$  et l'intervalle d'indétermination  $\epsilon$  prennent également l'exposant  $j$  et deviennent  $\nu^j$  et  $\epsilon^j$ , caractéristiques de la nouvelle série détectée. La série est ensuite ajoutée à l'ensemble  $\mathcal{H}$  contenant toutes les séries harmoniques identifiées jusqu'à présent. La recherche de séries commençant par les basses fréquences, l'ensemble  $\mathcal{H}$  est donc ordonné par

fréquence fondamentale croissante.

### 2.2.5 Cas de l'absence du fondamental

La méthode d'identification des harmoniques présentée jusqu'ici autorise tous les rangs  $r \in 2, 3, \dots$  de la série à être absents. Cependant, la méthode n'autorise pas l'absence du fondamental car elle s'en sert pour initialiser les paramètres de recherche ( $\nu$  et  $\epsilon$ ). Dans le cas où le fondamental n'est pas présent dans la densité spectrale du signal, ou qu'il n'a pas été détecté en tant que pic, l'identification de la série harmonique n'est donc pas possible en l'état. Pour contrevenir à cette limitation, nous considérerons qu'un pic, en plus d'être le fondamental potentiel d'une série, peut également être l'harmonique potentiel de rang 2 d'une série. En plus de la recherche d'une série où le pic  $p^i$  serait le fondamental, nous allons également chercher une série dont la fréquence fondamentale et l'intervalle d'indétermination seraient

$$\nu = \frac{f^i}{2}, \quad \epsilon = \frac{\varepsilon^i}{2}. \quad (2.11)$$

Nous nous retrouvons donc à chercher deux sous-ensembles de  $\mathcal{P}$ . Le sous-ensemble  $\mathcal{P}_1^i$ , correspondant au cas où le pic  $p^i$  est considéré comme le fondamental de la série et le sous-ensemble  $\mathcal{P}_2^i$  où le pic  $p^i$  est considéré comme l'harmonique de rang 2 de la série.

Il apparait clairement que prendre en compte la possible absence du fondamental double la quantité de recherches à réaliser. Ce doublement prend deux facettes différentes :

1. Le temps de calcul est doublé, ce qui n'est pas souhaitable ;
2. Le nombre de séries identifiées risque de doubler également.

Ce deuxième point est d'autant plus gênant que les harmoniques de  $f^i$  sont également des harmoniques de  $\frac{f^i}{2}$  et que nécessairement  $\mathcal{P}_1^i \subset \mathcal{P}_2^i$ . Le doublement des résultats n'est donc pas très pertinent dans la mesure où les séries sont imbriquées les unes dans les autres.

Pour diminuer le temps de calcul et ne pas obtenir deux séries imbriquées, l'idée est de procéder à la double recherche en une seule et unique passe et de prendre une décision *a posteriori* pour ne garder qu'une série, en décidant quelle sera sa fréquence fondamentale. En effet, le test consiste à vérifier si le pic ayant servi à l'initialisation de la recherche des harmoniques est le fondamental de la série, ou s'il s'agit de l'harmonique de rang 2. La recherche des harmoniques du pic  $p^i$  se déroule donc de la manière suivante :

1. Initialisation des paramètres de recherche :  $\nu = \frac{f^i}{2}$  et  $\epsilon = \frac{\varepsilon^i}{2}$  ;
2. Recherche de l'ensemble  $\mathcal{P}_2^i$  des pics harmoniques de  $\nu$  (suivant les méthodes proposées dans les sections 2.2.1 à 2.2.3) ;
3. Extraction des harmoniques de rangs impairs de  $\mathcal{P}_2^i$  pour former l'ensemble des pics  $\mathcal{P}_1^i$  ;
4. Validation des ensembles  $\mathcal{P}_2^i$  et  $\mathcal{P}_1^i$  en tant que séries harmoniques (critère de la section 2.2.4) ;
5. Choix de l'ensemble à retenir et enregistrement de la série harmonique.

Le choix de l'ensemble à retenir est réalisé de la manière suivante : si seulement un ensemble est validé, c'est lui qui est sélectionné. Dans le cas où les deux ensembles sont validés, on regarde si le nombre de pics dans l'ensemble  $\mathcal{P}_2^i$  est significativement plus important que le nombre de pics dans  $\mathcal{P}_1^i$ . Si c'est le cas, alors on sélectionne l'ensemble  $\mathcal{P}_2^i$ , sinon on sélectionne  $\mathcal{P}_1^i$ . Le ratio suivant doit donc franchir le seuil  $q^{\min}$  pour que l'ensemble  $\mathcal{P}_2^i$  soit conservé :

$$\frac{\#\mathcal{P}_2^i - \#\mathcal{P}_1^i}{\#\mathcal{P}_1^i} > q^{\min}. \quad (2.12)$$

Cette modification de la méthode de recherche des harmoniques permet finalement d'autoriser tous les rangs  $r \in \mathbb{N}^*$  à être absents, sans gêner la recherche de la série et sans augmenter le temps de calcul. Elle ajoute simplement un test, dont le paramétrage est discuté dans la section 2.5.1.

### 2.2.6 Post-traitement : suppression des sous-séries

Une fois que la recherche et la validation des séries harmoniques sont terminées, nous avons un ensemble  $\mathcal{H}$  contenant l'ensemble des séries harmoniques identifiées. Cependant, parmi cet ensemble, de nombreuses séries harmoniques ne portent pas d'informations dans la mesure où il s'agit de sous-séries harmoniques. Ce concept découle de la nature même d'une série harmonique. En effet, si l'on se place dans un cadre théorique et qu'on prend une série harmonique de fréquence fondamentale  $\nu$ , possédant une infinité d'harmoniques, les fréquences des harmoniques sont tous les multiples de  $\nu$ , à savoir  $r\nu$  avec  $r \in \mathbb{N}^*$ . Si nous sélectionnons dans cet ensemble de fréquences uniquement les fréquences dont le rang est pair  $r = 2k$  avec  $k \in \mathbb{N}^*$ , nous obtenons l'ensemble des fréquences  $2k\nu$ . Cet ensemble est équivalent à une série harmonique de fréquence fondamentale  $2\nu$ , incluse dans la série de fréquence  $\nu$ . On parle alors de sous-série. De la même manière, il est également possible de prendre une fréquence sur trois, une sur quatre, ou une sur n'importe quel nombre entier non nul pour obtenir une série harmonique de fréquence fondamentale  $3\nu$ ,  $4\nu$ , etc.

Afin de supprimer les sous-séries de l'ensemble  $\mathcal{H}$ , les séries harmoniques sont comparées une à une. En particulier, on regarde si l'ensemble des pics d'une série est inclus dans celui d'une autre série : si la série  $h^k \subset h^j$ , alors la série  $h^k$  est retirée de l'ensemble  $\mathcal{H}$ . Comme seule une série de plus grande fréquence peut être incluse dans une série de fréquence plus faible, l'inverse étant impossible, il faut procéder à  $\frac{\#\mathcal{H}(\#\mathcal{H}-1)}{2}$  comparaisons d'ensembles.

Les modèles de vibrations des systèmes mécaniques font état de relation de multiplicité entre les différents composants de la machine. Les fréquences fondamentales des séries harmoniques sont donc toutes liées entre elles et sont en général multiples d'une unique fréquence. Cela signifie que des séries harmoniques peuvent être des sous-séries d'une plus grande série. Cette plus grande série mélange alors les informations provenant de différentes parties du système. Dans ce cas-là, il peut être intéressant de conserver les sous-séries. Deux stratégies différentes peuvent être appliquées pour les conserver :

1. Ne pas appliquer le post-traitement supprimant les sous-séries et conserver l'ensemble

des séries harmoniques. Certaines d'entre elles ne porteront peut-être pas beaucoup d'information, mais toutes les séries souhaitées seront présentes ;

2. Utiliser une liste de fréquences caractéristiques et modifier le post-traitement afin de ne pas supprimer les sous-séries ayant une fréquence fondamentale dans cette liste. Ceci revient à injecter des informations *a priori*, ce qui est contraire à nos souhaits premiers. Cependant, si ces informations sont absentes, la méthode fonctionne toujours grâce à la stratégie précédente. La présence des informations *a priori* permet simplement de raffiner les résultats ; elles ne sont pas nécessaires pour les obtenir.

### 2.2.7 Synthèse sur la recherche d'harmoniques

La méthode proposée pour l'identification des séries harmoniques repose sur plusieurs points clefs. Le premier est la prise en compte de l'indétermination sur les fréquences exactes des pics, impliquant alors une recherche basée sur l'intersection d'intervalles. Pour éviter aux intervalles de recherche de grandir linéairement avec le rang des harmoniques recherchés, deux étapes supplémentaires ont été ajoutées, à savoir la mise à jour des paramètres de recherche et la prise en compte d'un critère d'arrêt. Une fois les harmoniques détectés, il faut également valider la série en s'assurant qu'elle vérifie les critères préétablis. Cette étape de validation est rendue légèrement plus complexe par l'ajout d'un choix à faire sur le rang du pic initial de la série. L'ajout de ce choix n'est pas anodin car il permet alors au fondamental de la série d'être absent, tout en rendant possible la détection de la série avec la bonne fréquence fondamentale. Enfin, une étape finale permet de supprimer les sous-séries détectées, afin de fournir un ensemble  $\mathcal{H}$  ne contenant que des séries pertinentes.

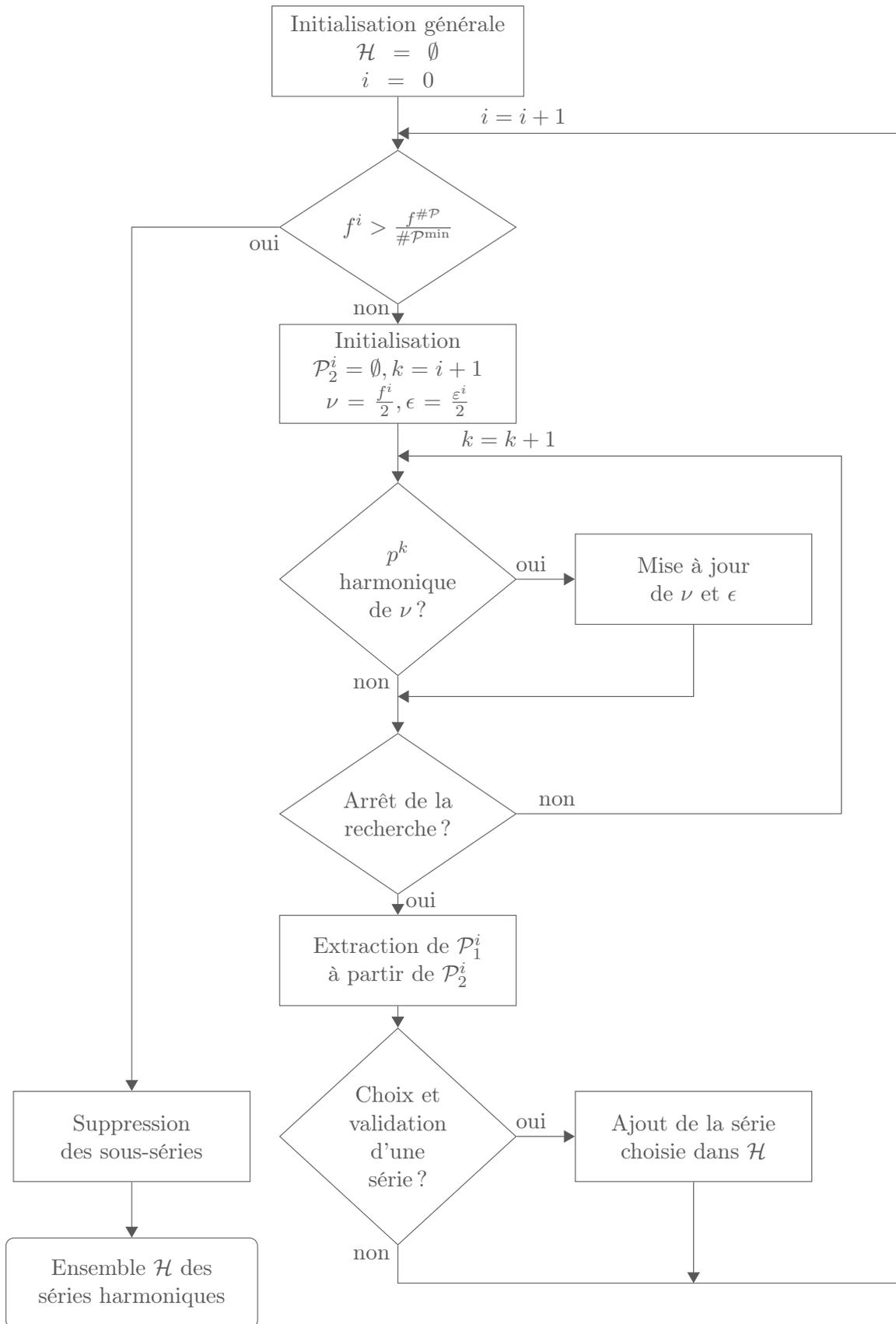
Toutes les étapes que nous venons de voir en détail sont représentées sur la figure 2.3. Ce schéma synthétise l'algorithme de recherche et montre bien comment les différentes étapes s'enchaînent. Pour finir, la méthode repose sur un certain nombre de paramètres, résumés dans le tableau 2.4. Le choix des paramètres est expliqué dans la section 2.5.1.

L'identification des séries harmoniques ne résout qu'à moitié la détection des structures spectrales au sein de l'ensemble  $\mathcal{P}$  des pics détectés. Il reste à présent à identifier les bandes de modulation.

## 2.3 Identification des bandes de modulation

Comme nous l'avons vu au cours de l'introduction sur les structures présentes dans la densité spectrale d'un signal (section 1.2.2.3), les modulations se présentent dans le spectre sous la forme de pics spectraux régulièrement espacés autour d'une fréquence porteuse  $f_p$ . Un modèle générique nous montre que les pics sont situés aux fréquences  $f^r$  telles que

$$\forall r \in \mathbb{Z}^*, f^r = f_p + r f_m, \quad (2.13)$$

FIGURE 2.3 – Algorithme de recherche de l'ensemble  $\mathcal{H}$  des séries harmoniques.

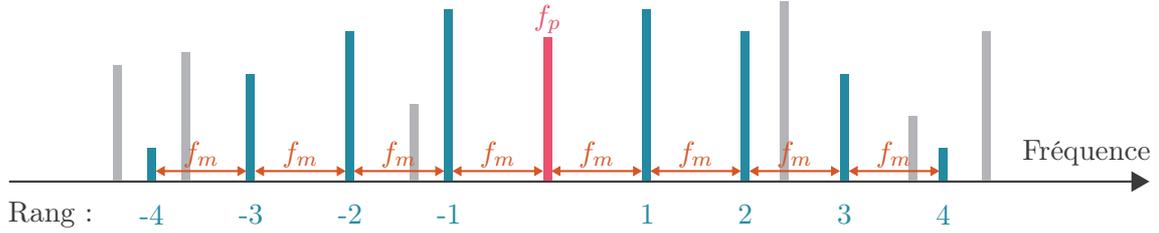


FIGURE 2.4 – Illustration d’une bande de modulation : les pics en bleu, régulièrement espacés d’une distance  $f_m$  autour de la porteuse  $f_p$  (en rose), forment une modulation bilatérale. Le rang du pic dans la bande est inscrit en dessous. Les pics en gris ne font pas partie de la modulation.

avec  $f_m$  la fréquence de modulation et  $r$  le rang du pic dans la bande. Ce modèle, illustré sur la figure 2.4, n’est pas loin du modèle des séries harmoniques dans la mesure où un facteur de multiplicité est également présent. Pour cette raison, nous proposons de réutiliser l’algorithme de recherche des séries harmoniques, présenté précédemment, afin de construire une méthode d’identification des bandes de modulation. En effet, une simple translation des fréquences permet de transformer la recherche des bandes à droite de la fréquence porteuse en recherche de séries harmoniques. Une transformation similaire permet d’identifier de la même manière les bandes à gauche de la fréquence porteuse. Une étape finale est alors nécessaire afin de fusionner les séries détectées de chacun des côtés et reconstruire les modulations dans leur globalité.

### 2.3.1 Recherche unilatérale de modulations

La recherche de modulations s’effectue autour d’une fréquence porteuse  $f_p$ . Si la méthode se veut exhaustive, elle doit considérer chaque pic  $p^i \in \mathcal{P}$  comme potentiellement porteur de modulations. La recherche est donc une fois encore itérative et nous allons à nouveau considérer qu’elle a été effectuée pour les pics de  $p^1$  à  $p^{i-1}$ . La suite des explications se place alors dans le cas général où nous considérons  $p^i$  comme le potentiel porteur de modulations. Ainsi, la fréquence porteuse est égale à la fréquence de  $p^i$  :

$$f_p = f^i. \quad (2.14)$$

Si théoriquement les modulations autour de  $f_p$  peuvent être infinies, en pratique, l’amplitude des pics diminue lorsque la valeur absolue du rang  $r$  augmente. C’est pourquoi il n’est pas forcément pertinent de chercher des modulations *loin* de la fréquence porteuse. En conséquence, nous introduisons le paramètre de recherche  $f_{\Delta}^{\max}$  qui est la déviation maximale en fréquence où la méthode ira chercher des modulations. Ce nouveau paramètre permet de définir deux autres nouvelles notations, à savoir les ensembles  $\mathcal{D}^p \subset \mathcal{P}$  et  $\mathcal{G}^p \subset \mathcal{P}$ . Ils dénotent respectivement l’ensemble des pics à droite et à gauche de la fréquence porteuse, mais compris

dans la déviation maximale. Autrement dit,

$$\begin{aligned}\mathcal{D}^p &\triangleq \{p^k \in \mathcal{P} \mid 0 < f^k - f_p < f_{\Delta}^{\max}\}, \\ \mathcal{G}^p &\triangleq \{p^k \in \mathcal{P} \mid 0 < f_p - f^k < f_{\Delta}^{\max}\}.\end{aligned}\tag{2.15}$$

La recherche de modulation est unilatérale, c'est-à-dire que nous allons chercher les modulations uniquement d'un côté de la fréquence porteuse. Concrètement, cela signifie que les modulations vont être identifiées au sein de l'ensemble  $\mathcal{D}^p$ , puis de l'ensemble  $\mathcal{G}^p$  de manière indépendante.

### 2.3.1.1 Modulations à droite de la fréquence porteuse

Afin d'identifier les modulations présentes dans l'ensemble  $\mathcal{D}^p$ , nous appliquons simplement une translation sur les fréquences des pics. Chaque pic possède alors une nouvelle fréquence  $\tilde{f}^k$ , calculée ainsi

$$\tilde{f}^k = f^k - f_p.\tag{2.16}$$

Or, d'après la définition de  $\mathcal{D}^p$ , cette translation implique que

$$\forall p^k \in \mathcal{D}^p, \tilde{f}^k \in [0, f_{\Delta}^{\max}].\tag{2.17}$$

L'équation (2.13) du modèle de modulations peut alors être réécrite

$$\forall r \in \mathbb{Z}^*, \tilde{f}^r = r f_m.\tag{2.18}$$

Si on restreint le modèle aux rangs positifs uniquement, c'est-à-dire que  $r \in \mathbb{N}^*$ , on se retrouve dans une situation identique au problème de recherche de séries harmoniques au sein de l'ensemble restreint  $\mathcal{D}^p$ .

Pour rappel, les résultats de l'identification des séries harmoniques prennent la forme d'un ensemble  $\mathcal{H}$ , dans lequel chaque série  $h^j$  est caractérisée par sa fréquence fondamentale  $\nu^j$ , l'indétermination  $\epsilon^j$  et l'ensemble  $\mathcal{P}^j$  des pics appartenant à la série. De plus, l'ensemble  $\mathcal{H}$  est ordonné par fréquence fondamentale croissante.

En appliquant l'algorithme de recherche de séries harmoniques au problème d'identification des modulations parmi l'ensemble  $\mathcal{D}^p$  de pics aux fréquences translattées, nous obtenons alors l'ensemble  $\mathcal{M}_{\mathcal{D}}^i$  des bandes de modulation  $m_{\mathcal{D}}^j$ . Chaque bande  $m_{\mathcal{D}}^j$  est une modulation unilatérale à droite de la porteuse, prenant la forme d'une série harmonique dans l'espace des fréquences transformées. L'exposant  $i$  dans  $\mathcal{M}_{\mathcal{D}}^i$  correspond à l'exposant du pic ayant servi de fréquence porteuse, alors que l'exposant  $j$  des bandes de modulation  $m_{\mathcal{D}}^j$  est l'indice de la bande dans l'ensemble  $\mathcal{M}_{\mathcal{D}}^i$ . De plus, chaque bande  $m_{\mathcal{D}}^j$  est caractérisée par sa fréquence de modulation  $f_{\mathcal{D}}^j$ , l'amplitude  $\epsilon_{\mathcal{D}}^j$  de l'intervalle d'indétermination quant à la valeur exacte de la fréquence de modulation et l'ensemble  $\mathcal{P}_{\mathcal{D}}^j$  des pics appartenant à la modulation.

La mise en place de la recherche des modulations autour de la fréquence  $f_p$  est illustrée

par un exemple factice sur la figure 2.5 (a). L'identification des modulations à droite de la porteuse est représentée sur la figure 2.5 (b).

### 2.3.1.2 Modulations à gauche de la fréquence porteuse

D'une manière très similaire à l'identification des modulations à droite de la fréquence porteuse, une transformation des fréquences nous permet d'identifier les modulations à gauche de la porteuse en rendant le problème identique à une recherche de séries harmoniques. Chaque pic possède alors une nouvelle fréquence  $\tilde{f}^k$ , calculée ainsi

$$\tilde{f}^k = -(f^k + f_p). \quad (2.19)$$

Cela implique, d'après la définition de  $\mathcal{G}^p$ , que

$$\forall p^k \in \mathcal{G}^p, \tilde{f}^k \in [0, f_{\Delta}^{\max}]. \quad (2.20)$$

Le fait de prendre l'opposée de la fréquence nous oblige à réordonner l'ensemble  $\mathcal{G}^p$ , pour que les pics soient ordonnés par fréquence transformée croissante. C'est une des conditions nécessaires pour pouvoir utiliser l'algorithme de recherche des séries harmoniques. D'ailleurs, une fois utilisé, nous obtenons l'ensemble  $\mathcal{M}_{\mathcal{G}}^i$  des bandes de modulation  $m_{\mathcal{G}}^j$ . Chaque bande  $m_{\mathcal{G}}^j$  est caractérisée par sa fréquence de modulation  $f_{\mathcal{G}}^j$ , l'amplitude  $\epsilon_{\mathcal{G}}^j$  de l'intervalle d'indétermination quant à la valeur exacte de la fréquence de modulation et l'ensemble  $\mathcal{P}_{\mathcal{G}}^j$  des pics appartenant à la modulation. Cette étape de l'identification des modulations est représentée sur la figure 2.5 (c).

### 2.3.2 Fusion des bandes latérales

La recherche unilatérale nous retourne deux ensembles de modulations,  $\mathcal{M}_{\mathcal{G}}^i$  à gauche et  $\mathcal{M}_{\mathcal{D}}^i$  à droite de la fréquence porteuse  $f_p$ . Afin de reconstruire des modulations bilatérales, il est nécessaire d'identifier dans ces ensembles les fréquences de modulation identiques. Pour cela, nous considérons encore une fois les fréquences de modulation comme des intervalles, et l'égalité de deux fréquences est vérifiée lorsque l'intersection des intervalles est non nulle :

$$f_{\mathcal{G}}^j = f_{\mathcal{D}}^k \Leftrightarrow \left[ f_{\mathcal{G}}^j - \frac{\epsilon_{\mathcal{G}}^j}{2}, f_{\mathcal{G}}^j + \frac{\epsilon_{\mathcal{G}}^j}{2} \right] \cap \left[ f_{\mathcal{D}}^k - \frac{\epsilon_{\mathcal{D}}^k}{2}, f_{\mathcal{D}}^k + \frac{\epsilon_{\mathcal{D}}^k}{2} \right] \neq \emptyset. \quad (2.21)$$

Cependant, vérifier l'égalité précédente n'est pas suffisant pour reconstruire une modulation bilatérale. En effet, il est possible qu'une fréquence de modulation  $f_{\mathcal{G}}^j$  prise à gauche de la porteuse soit égale à plusieurs fréquences de modulation à droite de la porteuse, et inversement. Il faut donc non seulement vérifier l'égalité, mais considérer un critère supplémentaire permettant de départager les éventuels multiples candidats. Ce critère a été choisi comme étant la minimisation de la distance entre les fréquences. Ainsi, pour la fréquence  $f_{\mathcal{G}}^j$

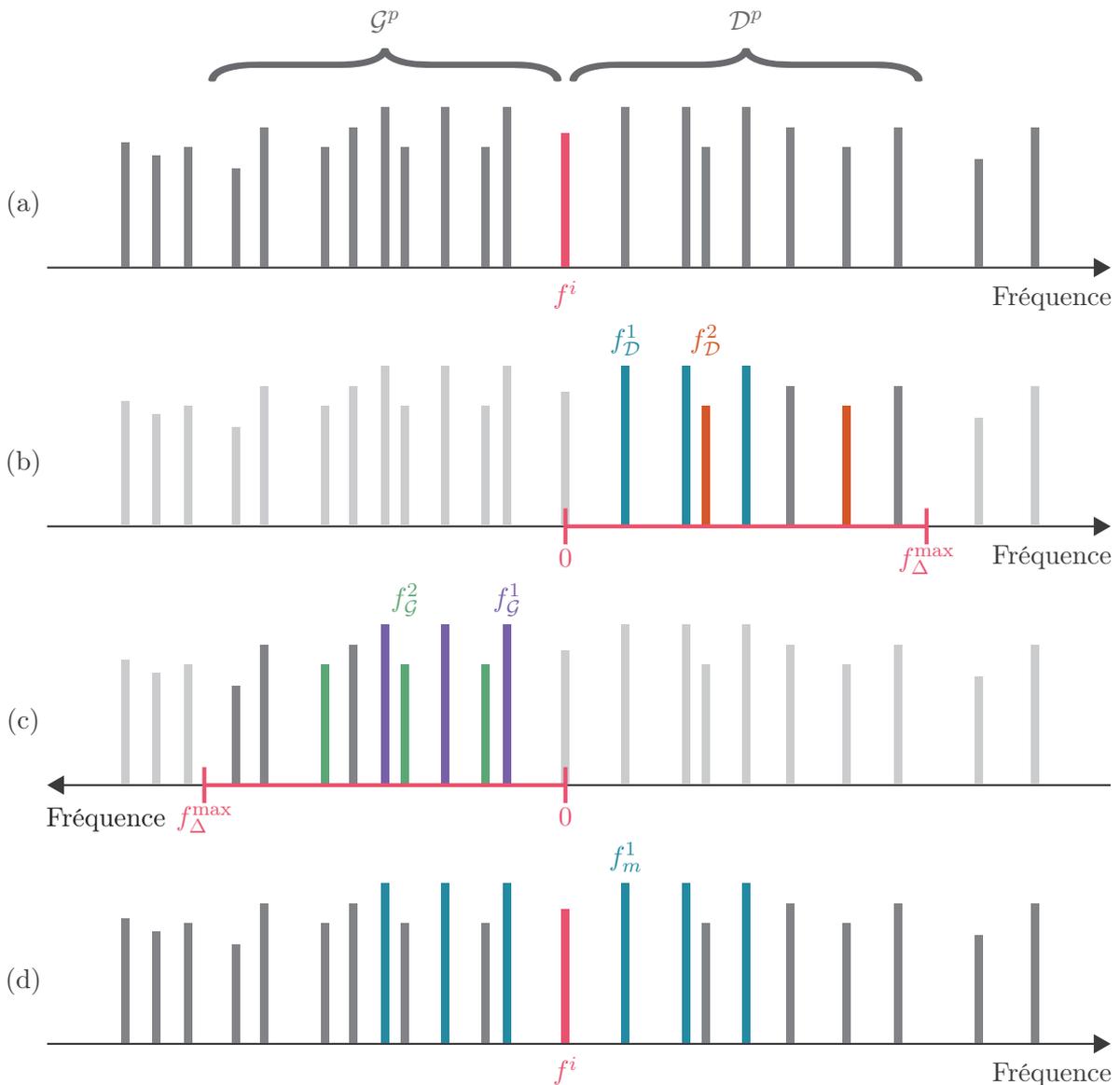


FIGURE 2.5 – Illustration des 4 étapes de la méthode de recherche de bandes de modulation : (a) détermination des ensembles  $\mathcal{G}^p$  et  $\mathcal{D}^p$  en fonction de la fréquence porteuse  $f_p = f^i$  et de la déviation maximale  $f_{\Delta}^{\max}$ , (b) recherche des modulations dans la bande latérale  $\mathcal{D}^p$  : 2 bandes de modulation sont identifiées (en bleu et en orange), (c) recherche des modulations dans la bande latérale  $\mathcal{G}^p$  : 2 bandes de modulation sont identifiées (en orange et en violet) et (d) fusion des bandes latérales ayant la même fréquence : une seule bande de modulation est finalement détectée ( $f_{\mathcal{G}}^1 = f_{\mathcal{D}}^1 \Rightarrow f_m^1$ ).

à gauche, la fréquence à droite choisie sera celle vérifiant

$$f_{\mathcal{D}}^k = \underset{f_{\mathcal{D}}^l \text{ vérifiant (2.21)}}{\operatorname{argmin}} \left| f_{\mathcal{G}}^j - f_{\mathcal{D}}^l \right|. \quad (2.22)$$

Inversement, pour une fréquence  $f_{\mathcal{D}}^k$  à droite, la fréquence choisie à gauche doit vérifier

$$f_{\mathcal{G}}^j = \underset{f_{\mathcal{G}}^l \text{ vérifiant (2.21)}}{\operatorname{argmin}} \left| f_{\mathcal{D}}^k - f_{\mathcal{G}}^l \right|. \quad (2.23)$$

La reconstruction d'une modulation bilatérale à partir des modulations unilatérales  $m_{\mathcal{G}}^j$  et  $m_{\mathcal{D}}^k$  n'est donc effective qu'à partir du moment où (2.22) appliquée à  $f_{\mathcal{G}}^j$  renvoie  $f_{\mathcal{D}}^k$  et où (2.23) appliquée à  $f_{\mathcal{D}}^k$  renvoie  $f_{\mathcal{G}}^j$ .

Afin de procéder à la recherche des couples de modulations unilatérales à fusionner pour reconstruire des modulations bilatérales, nous proposons l'algorithme ci-après. Les 3 étapes garantissent le respect simultané de (2.22) et (2.23) :

1. Calcul exhaustif des distances entre fréquences et création d'un ensemble  $\mathcal{C}$  de combinaisons possibles :
  - $\forall j \in \{1, \dots, \#\mathcal{M}_{\mathcal{G}}^i\}, \forall k \in \{1, \dots, \#\mathcal{M}_{\mathcal{D}}^i\}, d_{jk} = \left| f_{\mathcal{G}}^j - f_{\mathcal{D}}^k \right|$ ;
  - Si  $d_{jk} < \frac{\epsilon_{\mathcal{G}}^j + \epsilon_{\mathcal{D}}^k}{2}$  (équivalent à (2.21)), ajout du triplet  $(d_{jk}, j, k)$  dans l'ensemble  $\mathcal{C}$ ;
2. Tri par distance croissante de l'ensemble  $\mathcal{C}$ ;
3. Parcours itératif de l'ensemble  $\mathcal{C}$  et fusion des bandes unilatérales vérifiant (2.22) et (2.23) :
  - Récupération du triplet  $(d_{ij}, j, k)$  courant et fusion des modulations  $m_{\mathcal{G}}^j$  et  $m_{\mathcal{D}}^k$ ;
  - Suppression de tous les triplets de  $\mathcal{C}$  contenant soit l'indice  $j$ , soit l'indice  $k$ .

La fusion de deux modulations unilatérales,  $m_{\mathcal{G}}^j$  et  $m_{\mathcal{D}}^k$ , en une modulation bilatérale  $m^l$  se passe de la manière suivante. La fréquence porteuse  $f_p = f^i$  ne change pas, elle est la même pour les deux. L'ensemble des pics est fusionné  $\mathcal{P}^l = \mathcal{P}_{\mathcal{G}}^j \cup \mathcal{P}_{\mathcal{D}}^k$ . L'indétermination sur la fréquence de modulation est prise comme le maximum des deux indéterminations  $\epsilon^l = \max(\epsilon_{\mathcal{G}}^j, \epsilon_{\mathcal{D}}^k)$ . La fréquence de modulation est calculée à partir des rangs et des fréquences maximales dans chacune des bandes unilatérales. Soit  $r_{\mathcal{G}}^{\max}$  le rang du pic de fréquence maximale  $f_{\mathcal{G}}^{\max}$  à gauche. De même, on a  $r_{\mathcal{D}}^{\max}$  et  $f_{\mathcal{D}}^{\max}$  à droite. Alors la fréquence de modulation est calculée comme étant égale à

$$f_m^l = \frac{1}{2} \left( \frac{f_{\mathcal{G}}^{\max}}{r_{\mathcal{G}}^{\max}} + \frac{f_{\mathcal{D}}^{\max}}{r_{\mathcal{D}}^{\max}} \right). \quad (2.24)$$

Ainsi, cette étape de fusion, représentée sur la figure 2.5 (d), nous retourne l'ensemble  $\mathcal{M}^i$  des modulations  $m^j$  bilatérales dont les caractéristiques sont la fréquence porteuse  $f_p = f^i$ , la fréquence de modulation  $f_m^j$ , l'indétermination sur la fréquence de modulation  $\epsilon^j$  et l'ensemble  $\mathcal{P}^j$  des pics appartenant à la modulation.

En fonction du contexte, les modulations des ensembles  $\mathcal{M}_{\mathcal{G}}^i$  et  $\mathcal{M}_{\mathcal{D}}^i$  qui n'ont pas trouvé leur pendant de l'autre côté de la porteuse peuvent être soit conservées, soit supprimées.

En effet, dans certains domaines comme celui des télécommunications, des bandes latérales uniques sont tout à fait possibles alors qu'elles n'auront aucun sens physique dans d'autres domaines.

Finalement, après avoir considéré chacun des pics de l'ensemble  $\mathcal{P}$  comme le potentiel porteur de modulations, nous obtenons l'ensemble  $\mathcal{M}$  comprenant toutes les modulations identifiées dans le spectre :

$$\mathcal{M} = \bigcup_{i \in \{1, \dots, \#\mathcal{P}\}} \mathcal{M}^i. \quad (2.25)$$

## 2.4 Caractérisation des structures spectrales

L'identification des séries harmoniques et des bandes de modulation est une étape importante dans la lecture du spectre d'un signal. En vue de réaliser un système de diagnostic automatique, la seule présence des séries ne suffit pas. Il faut aussi caractériser les structures spectrales. Les caractéristiques calculées forment alors la *carte d'identité spectrale* de la structure considérée.

### 2.4.1 Carte d'identité d'une série harmonique

La carte d'identité spectrale d'une série harmonique  $h^i \in \mathcal{H}$  comporte déjà 2 caractéristiques que nous avons utilisées pour son identification : la fréquence fondamentale  $\nu^i$  et l'amplitude  $\epsilon^i$  de l'intervalle d'indétermination. Nous connaissons l'ensemble  $\mathcal{P}^i$  des pics formant la série. C'est à partir de cet ensemble que nous allons dériver d'autres caractéristiques.

#### La taille

La taille d'une série peut se mesurer de deux manières différentes. La première consiste à simplement compter le nombre d'harmoniques dans la série :

$$N_{\text{pics}}^i = \#\mathcal{P}^i \quad (2.26)$$

La seconde consiste à considérer le rang du dernier harmonique de la série :

$$r_{\text{max}}^i = \left\lfloor \frac{f_{\text{max}}^i}{\nu^i} \right\rfloor, \text{ avec } f_{\text{max}}^i = \operatorname{argmax}_{p^k \in \mathcal{P}^i} f^k. \quad (2.27)$$

### La densité

La densité est simplement le ratio entre les deux tailles que nous avons définies précédemment :

$$D^i = \frac{N_{\text{pics}}^i}{r_{\text{max}}^i}. \quad (2.28)$$

La densité varie entre 0 et 1 et représente le taux de « remplissage » de la série. Si tous les harmoniques sont présents ou presque, la densité sera proche de 1. *A contrario*, si de nombreux rangs sont absents, la densité sera plus proche de 0.

### L'énergie

L'énergie correspond au cumul de l'énergie des pics appartenant à la série :

$$E^i = \sum_{p^k \in \mathcal{P}^i} (a^k)^2, \quad (2.29)$$

où  $a^k$  représente l'amplitude du pic  $p^k$ .

### Le taux de distorsion harmonique

Le THD<sup>3</sup> est un indicateur classiquement utilisé pour mesurer la qualité d'une chaîne de traitements. C'est particulièrement le cas en audio, où le THD sert à mesurer la distorsion engendrée par l'amplification du signal [Slone 2001]. Il se calcule ainsi

$$\text{THD}^i = \frac{\sqrt{\sum_{p^k \in \mathcal{P}^i - \{p^1\}} (a^k)^2}}{a^1}, \quad (2.30)$$

où le pic  $p^1$  exclu de la somme au numérateur correspond au pic représentant le fondamental de la série, dont l'amplitude  $a^1$  se retrouve au dénominateur. Cette définition demande donc à ce que le fondamental soit présent. Dans le cas où il ne l'est pas, le calcul du THD n'est pas possible.

### La richesse

La richesse d'une série est calculée comme le rapport entre sa taille, en termes de rang, et la taille maximale qu'elle pourrait prendre dans l'ensemble  $\mathcal{P}$ . En effet, si on considère  $f_{\text{max}}$  comme étant la plus grande fréquence qu'un pic contenu dans  $\mathcal{P}$  pourrait prendre (typiquement  $\frac{f_e}{2}$ , avec  $f_e$  la fréquence d'échantillonnage), alors la série de fondamentale  $\nu^i$  pourrait

---

3. Taux de distorsion harmonique (*Total Harmonic Distorsion*, en anglais)

contenir  $N_{\max, \text{pics}}^i = \left\lfloor \frac{f_{\max}}{\nu^i} \right\rfloor$  pics. Le calcul de la richesse s'effectue alors ainsi :

$$R^i = \frac{r_{\max}^i}{N_{\max, \text{pics}}^i}. \quad (2.31)$$

La richesse varie entre 0 et 1. Une valeur proche de 1 signifiera que l'ensemble des pics pouvant être présents dans le spectre sont effectivement présents dans la série.

### Carte d'identité spectrale

Finalement, l'ensemble des caractéristiques présentes sur la carte d'identité spectrale d'une série harmonique est résumé au sein du tableau 2.2.

Caractéristique de la série harmonique $h^i$	Notation
Fréquence fondamentale	$\nu^i$
Intervalle d'indétermination sur la fréquence fondamentale	$\epsilon^i$
Nombre de pics	$N_{\text{pics}}^i$
Énergie	$E^i$
Rang du dernier harmonique	$r_{\max}^i$
Densité	$D^i$
Taux de distorsion harmonique	THD <sup><i>i</i></sup>
Richesse	$R^i$

Tableau 2.2 – Carte d'identité spectrale d'une série harmonique  $h^i$ .

#### 2.4.2 Carte d'identité d'une bande de modulation

Les bandes de modulation unilatérales sont identifiées comme si elles étaient des séries harmoniques. Elles peuvent donc se caractériser de la même manière que les séries, en considérant non plus  $\mathcal{P}$  comme l'ensemble de référence, mais  $\mathcal{G}^p$  ou  $\mathcal{D}^p$  avec leurs fréquences transformées. La liste des caractéristiques d'une bande unilatérale de modulation est donc la même que dans le tableau 2.2.

Quant aux bandes bilatérales, elles sont issues de la fusion de deux bandes unilatérales. Il est donc possible de caractériser une bande bilatérale en concaténant simplement les informations provenant de la bande à gauche et de la bande à droite de la fréquence porteuse. Cependant, certaines caractéristiques sont tout de même mises en commun, lorsqu'il est plus intéressant ne n'avoir qu'une seule valeur pour caractériser toute la série. C'est par exemple le cas de la fréquence de modulation, dont la mise en commun est réalisée via (2.24). De même, l'énergie de la modulation est calculée sur l'ensemble des pics dans la bande. Nous obtenons alors la carte d'identité spectrale résumée par le tableau 2.3.

Caractéristique de la bande de modulation $m^i$	À gauche	À droite
Fréquence porteuse	$f_p$	
Fréquence de modulation	$f_m^i$	
Intervalle d'indétermination sur la fréquence de modulation	$\epsilon^i$	
Nombre de pics	$N_{\text{pics}}^i$	
Énergie	$E^i$	
Rang de la dernière modulation	$r_{\max, \mathcal{G}}^i$	$r_{\max, \mathcal{D}}^i$
Densité	$D_{\mathcal{G}}^i$	$D_{\mathcal{D}}^i$
Taux de distorsion harmonique	$\text{THD}_{\mathcal{G}}^i$	$\text{THD}_{\mathcal{D}}^i$
Richesse	$R_{\mathcal{G}}^i$	$R_{\mathcal{D}}^i$

Tableau 2.3 – Carte d'identité spectrale d'une bande de modulation bilatérale  $m^i$ .

## 2.5 Applications

La méthode d'identification des structures spectrales est appliquée sur des signaux simulés, ainsi que sur des signaux réels issus de la surveillance de système. Cette démarche est une étape de validation de l'approche proposée. Avant cela, il est cependant nécessaire de paramétrer la méthode. Nous allons donc commencer par expliquer le choix des paramètres de recherche et commenter les conséquences de ces choix.

### 2.5.1 Choix des paramètres de recherche

Le tableau 2.4 liste l'ensemble des paramètres utilisés pour l'identification des séries harmoniques. Il donne également les valeurs par défaut que nous avons choisies pour ces paramètres.

Pour parler de séries harmoniques, le nombre minimal de pics dans une série doit être de 2 : le fondamental et un de ses harmoniques. Cependant, la valeur de ce paramètre a été fixée à 3. Ce choix découle de la trop forte densité de pics dans l'ensemble  $\mathcal{P}$ . En effet, la méthode de détection des pics fonctionne sans *a priori* et détecte des pics dont l'amplitude est très faible, à peine plus forte que la variance du bruit, ce qui garantit que l'ensemble des informations que la densité spectrale peut nous apporter est détecté. Cependant, cette force de détection conduit également à un ensemble  $\mathcal{P}$  très dense et la probabilité d'associer un harmonique à un pic isolé devient grande. Ainsi, le fait de considérer une série valide uniquement si elle possède au moins 3 pics diminue le nombre de fausses détections de séries, tout en continuant à garantir la détection précoce des petites séries harmoniques.

Les choix du nombre maximal de rangs consécutifs absents et du quotient minimal si le fondamental est absent ont été déterminés expérimentalement sur de nombreux signaux, de types très différents : signaux électriques, vibratoires ou encore des signaux musicaux. Les valeurs retenues sont celles qui, subjectivement parlant, donnaient les résultats les plus

pertinents sur l'ensemble des cas.

Paramètre	Notation	Valeur par défaut
Nombre minimal de pics dans une série	$\#\mathcal{P}^{\min}$	3
Nombre maximal de rangs consécutifs absents	$r_{\emptyset}^{\max}$	8
Quotient minimal si le fondamental est absent	$q^{\min}$	30 %

Tableau 2.4 – Liste des paramètres utilisés pour l'identification des séries harmoniques.

Le tableau 2.5 liste l'ensemble des paramètres utilisés pour l'identification des bandes de modulation. Dans la mesure où cette identification réutilise la détection des séries harmoniques, il est normal d'y retrouver des paramètres communs.

La recherche des modulations est très gourmande en temps de calcul. En effet, pour chaque fréquence porteuse, il faut utiliser deux fois la méthode de recherche d'harmoniques et fusionner leurs résultats. L'identification exhaustive de toutes les séries dans l'ensemble du spectre est donc très longue et coûteuse en termes de ressources. Nous avons donc fait le choix de limiter l'ensemble  $\mathcal{P}^p$  des pics potentiellement porteurs de modulations. Ce sous-ensemble de  $\mathcal{P}$  est construit à partir des ensembles de pics appartenant aux séries harmoniques préalablement identifiés. Ainsi, chaque fréquence porteuse potentielle fait donc partie d'au moins une série harmonique. La fréquence fondamentale de cette série (ou la fréquence fondamentale la plus grande si le pic appartient à plusieurs séries harmoniques simultanément) est utilisée pour fixer la déviation maximale  $f_{\Delta}^{\max}$  par rapport à la fréquence porteuse.

La présence de bandes unilatérales dans les signaux électriques ou vibratoires issus de la surveillance d'un système est très peu probable. Pour cette raison, nous ne conserverons que les modulations bilatérales. Nous relâchons alors un peu la contrainte sur le nombre de pics minimal lors d'une recherche unilatérale, en ne demandant que la présence d'au minimum 2 pics. Ce relâchement s'explique entre autres par la présence de l'opération de fusion des bandes unilatérales, qui vient diminuer les probabilités de fausses détections. Cependant, cette valeur de 2 pics de chaque côté d'une porteuse signifie aussi qu'il ne sera pas possible de détecter une modulation d'amplitude parfaitement sinusoïdale, qui produirait un unique pic de chaque côté de la porteuse. Ce cas de figure est très peu probable dans des systèmes mécaniques complexes. Pour cette raison, nous ne prendrons pas le risque de baisser le seuil à 1 et d'engendrer de trop nombreuses fausses détections.

Paramètre	Notation	Valeur par défaut
Ensemble des pics potentiellement porteurs	$\mathcal{P}^p$	$\cup_{h^l \in \mathcal{H}} \mathcal{P}^l$
Déviaton maximale en fréquence de $f_p$	$f_{\Delta}^{\max}$	dépend de $f_p$
Nombre minimal de pics dans une bande unilatérale	$\#\mathcal{P}^{\min}$	2
Nombre maximal de rangs consécutifs absents	$r_{\emptyset}^{\max}$	8
Quotient minimal si le fondamental est absent	$q^{\min}$	30 %

Tableau 2.5 – Liste des paramètres utilisés pour l'identification des bandes de modulation.

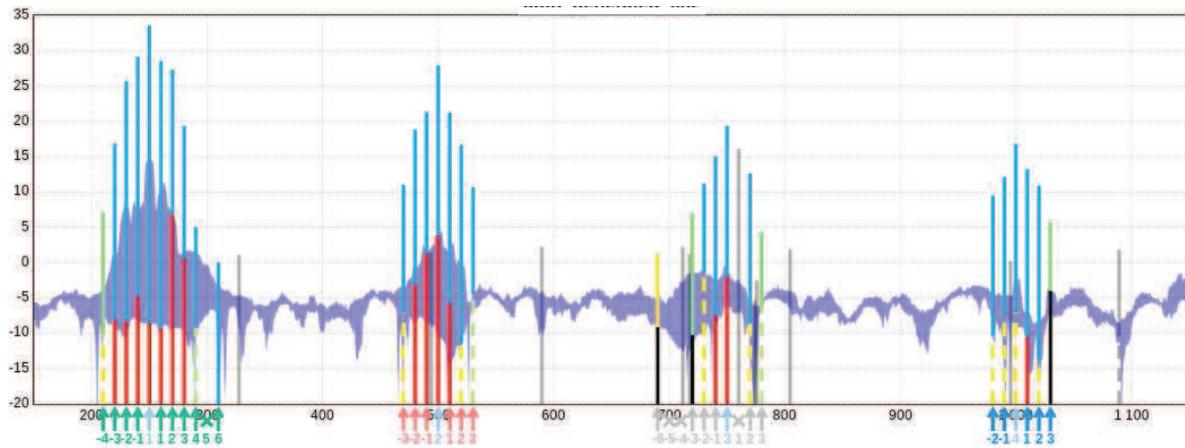


FIGURE 2.6 – Zoom sur l'interprétation du spectre d'un défaut simulé : une série harmonique est détectée à la fréquence fondamentale de 250 Hz ; tous les pics de la série sont modulés à une fréquence de 10 Hz.

### 2.5.2 Simulation d'un défaut

Le premier exemple d'application de la méthode est réalisé sur un signal artificiel, que nous noterons  $s_a[k]$ . Il a été construit afin de simuler l'engrènement de deux engrenages. Dans le spectre d'un signal vibratoire, l'engrènement prend la forme d'une série harmonique dont la fréquence fondamentale dépend du nombre de dents de chacun des engrenages et de la vitesse de rotation de l'engrenage en entrée. Dans  $s_a[k]$ , la fréquence de cette série a été fixée à 250 Hz et le nombre d'harmoniques à 4 (fondamental inclus). L'un des engrenages a été modélisé comme ayant un défaut, entraînant l'apparition de modulations autour des 4 pics de la série d'engrènement. La fréquence de ces modulations a été fixée à 10 Hz. Nous souhaitons donc retrouver cette série harmonique, ainsi que les 4 bandes de modulation associées, noyées dans un bruit blanc Gaussien ( $RSB = 0$  dB).

Le signal  $s_a[k]$ , échantillonné à 10 kHz, dure une seconde. L'ensemble  $\mathcal{P}$  compte 66 pics détectés dans la densité spectrale du signal. La méthode proposée dans ce chapitre permet d'identifier correctement la série harmonique à 250 Hz, ainsi que les 4 modulations de porteuses 250, 500, 750 et 999,94 Hz et de fréquences de modulation 9,99, 10,01, 10,03 et 9,95 Hz, respectivement. L'identification de ces structures est présentée sur la figure 2.6. On y voit de nombreux pics, représentés par des lignes verticales bicolores. Les deux couleurs de chaque pic codent des informations que nous n'utiliserons pas dans ce manuscrit. Par contre, les flèches colorées sous l'axe des fréquences permettent d'identifier les structures spectrales : les pics appartenant à une même structure spectrale sont pointés par des flèches de couleur identique. De plus, sous la flèche peut se lire le rang du pic dans la structure. Si une croix apparaît en lieu et place d'une flèche, cela signifie qu'il n'y a pas d'harmonique ou de modulation pour ce rang-là.

Cependant, l'ensemble  $\mathcal{H}$  contient non pas une, mais 3 séries harmoniques. Les deux séries supplémentaires ont des fréquences fondamentales à 239,99 et 260,01 Hz respectivement. Elles

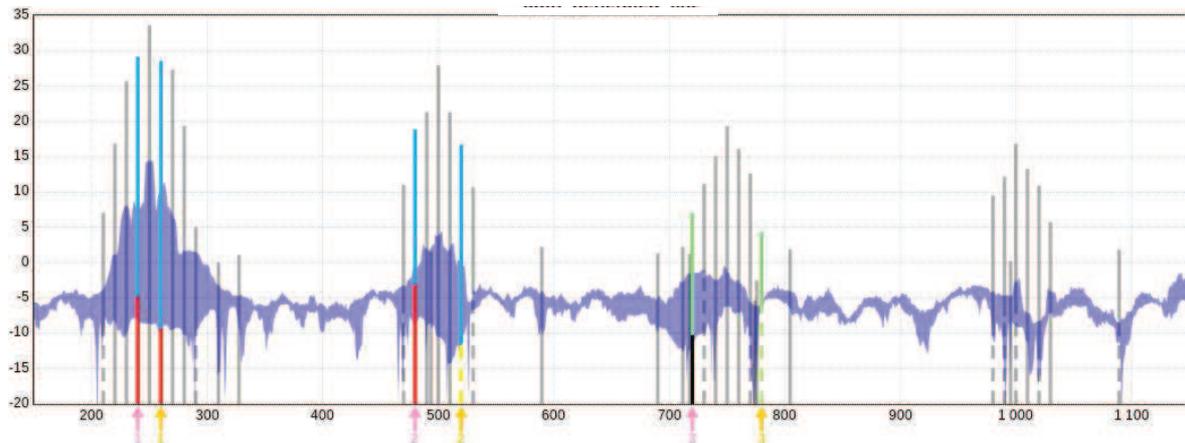


FIGURE 2.7 – Détection de deux séries harmoniques parmi les modulations de la série d'intérêt du défaut simulé.

sont représentées sur la figure 2.7. Ces deux séries proviennent d'une limitation de la méthode, fortement liée aux modèles harmoniques et de modulations. En effet, si tous les pics d'une série de fondamentale  $\nu$  sont modulés par une fréquence  $f_m$ , il est possible d'identifier dans les bandes des modulations des séries harmoniques de fréquences fondamentales  $\nu \pm r f_m$ , avec  $r \in \mathbb{N}^*$ . Le nombre de séries qu'il sera possible d'identifier et leur longueur dépend du nombre de pics dans chacune des modulations.

De manière similaire, l'ensemble  $\mathcal{M}$  contient 6 bandes de modulation. Les deux bandes de modulation supplémentaires ont pour fréquences porteuses 239,99 et 260,01 Hz et pour fréquences de modulation 9,95 et 9,99 Hz, respectivement. Elles proviennent donc de l'identification des modulations dans les 2 séries harmoniques supplémentaires.

### 2.5.3 Surveillance d'un système mécanique

L'exemple qui va suivre est issu de la surveillance du banc de test simulant le fonctionnement d'une éolienne. Une description du banc est disponible dans l'annexe C. Les signaux considérés sont des signaux vibratoires, échantillonnés à 39062,5 Hz, sur une durée de 150 secondes. Il y a trois signaux, pris par le même accéléromètre à trois dates différentes. Ces signaux permettent de suivre l'évolution d'un défaut sur le roulement principal qui a été contraint, afin de provoquer son usure mécanique. Le premier signal sert de référence, quand le roulement vient juste d'être mis en service. Le second signal est pris pendant la formation du défaut et le dernier juste avant le démontage du roulement pour inspection visuelle et confirmation du défaut par des experts en mécanique du CETIM<sup>4</sup>.

Les signaux sont longs et la fréquence d'échantillonnage est grande. Cela permet aux estimateurs de la densité spectrale d'avoir une résolution spectrale très fine. Le nombre de pics détectés pour ces trois signaux est par conséquent important : 10171, 8646 et 9289.

4. Centre Technique des Industries Mécaniques

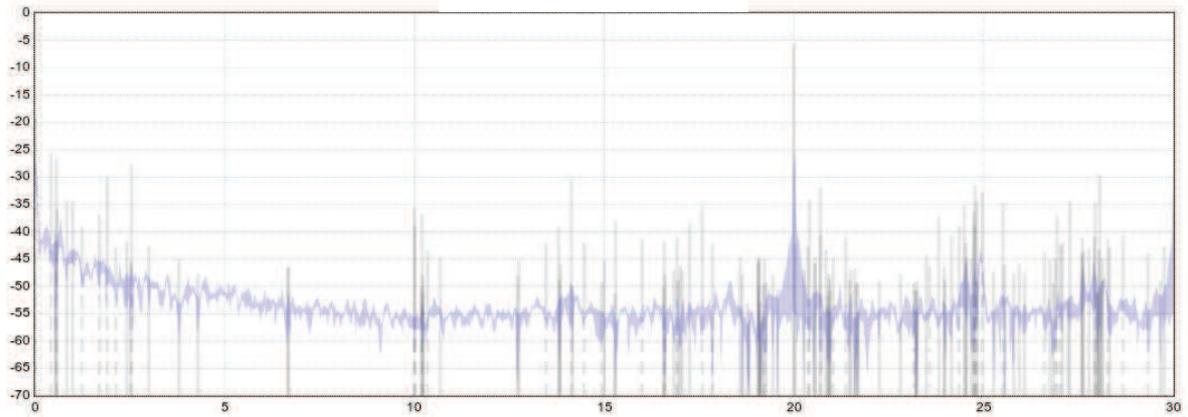


FIGURE 2.8 – Zoom sur le spectre vibratoire du banc d'essai éolien, avant le défaut. La série harmonique à 3,45 Hz n'est pas présente.

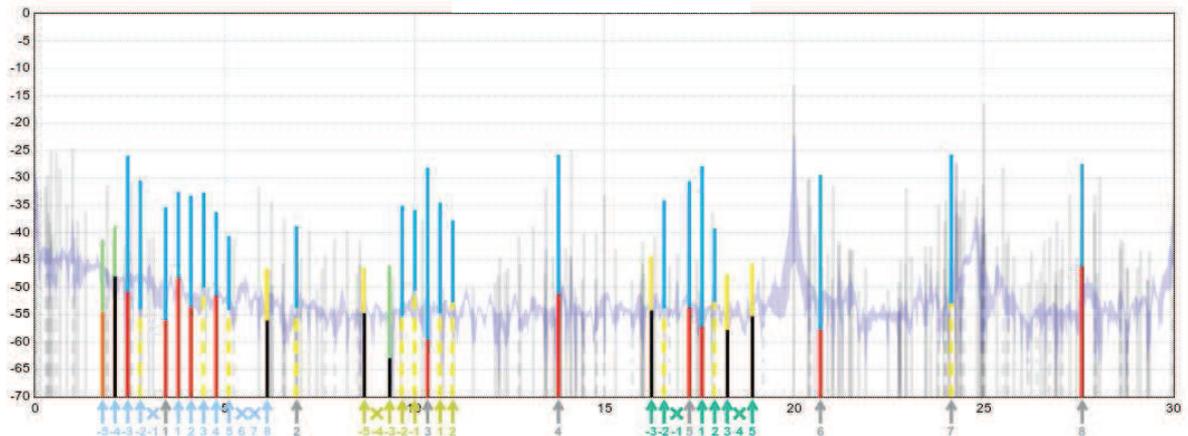


FIGURE 2.9 – Zoom sur le spectre vibratoire du banc d'essai éolien, pendant la formation du défaut. La série harmonique à 3,45 Hz est présente, avec des modulations de 0,33 Hz autour de chacun des harmoniques.

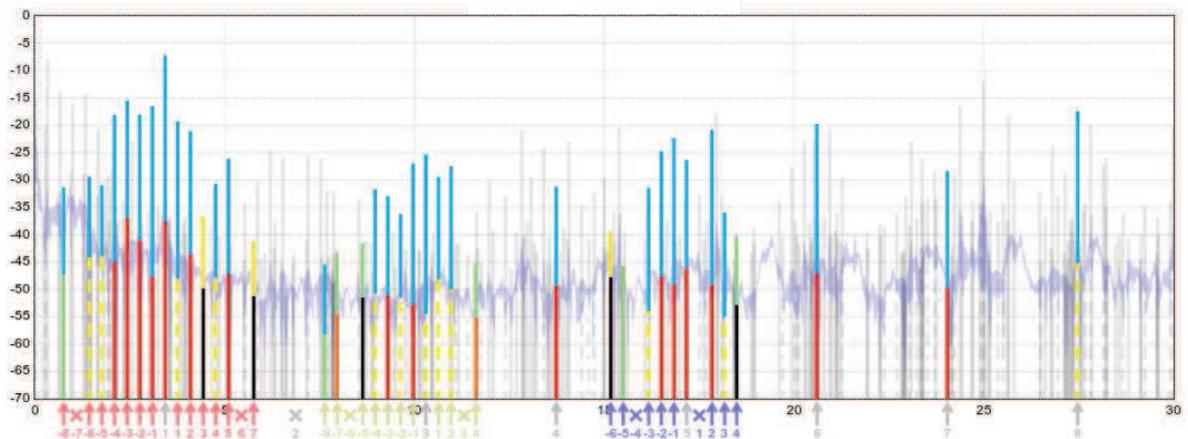


FIGURE 2.10 – Zoom sur le spectre vibratoire du banc d'essai éolien, au plus fort du défaut. Les amplitudes des pics dans la série harmonique et dans les modulations ont augmenté.

L'application de la méthode d'identification des structures spectrales permet de détecter 714 séries harmoniques et 5827 bandes de modulation dans le premier ensemble de pics, 448 et 4571 dans le second et 541 et 6104 dans le troisième. Ces nombres sont conséquents, mais sont le reflet d'une volonté de proposer une méthode sans aucun *a priori*, et qui ne souhaite pas passer à côté d'une information essentielle dans le spectre.

Nous allons nous concentrer sur une bande fréquentielle restreinte entre 0 et 30 Hz, et plus particulièrement sur une série harmonique à 3,45 Hz, liée au passage des billes sur la bague externe du roulement principal. Les figures 2.8, 2.9 et 2.10 nous montrent la bande 0–30 Hz pour chacun des signaux. Dans un premier temps, avant l'apparition du défaut, le spectre ne présente pas de série harmonique à 3,45 Hz. Cette dernière apparaît avec le défaut. Elle est d'ailleurs modulée par une fréquence de 0,333 Hz, liée à la vitesse de rotation de l'arbre. Les modulations sont représentées uniquement pour le fondamental et les harmoniques de rang 3 et 5, afin d'éviter de surcharger les figures. Ces structures spectrales se retrouvent dans le dernier spectre présenté, au plus fort du défaut, juste avant le démontage. Si les structures sont globalement les mêmes, on notera tout de même un nombre de modulations légèrement plus important et surtout, une augmentation des amplitudes.

Il est important de noter ici que la méthode proposée dans ce chapitre permet de caractériser le spectre d'un signal. Une mesure est faite sur l'état du système. Cependant, le diagnostic du système à partir d'une unique description de spectre serait très délicat à faire. L'évolution au cours du temps fera ressortir les informations importantes et facilitera le diagnostic.

## 2.6 Synthèse du chapitre

La problématique résolue au cours de ce chapitre consiste à analyser un signal pour en extraire toute l'information disponible. La méthode proposée devait être le plus générique possible, afin de s'appliquer à tous types de signal, quel que soit le domaine d'application. Nous avons donc choisi de procéder en deux étapes principales. Premièrement, le contenu brut de la densité spectrale du signal est extrait à partir de travaux antérieurs réalisés au GIPSA-lab et à l'IRIT. Ce contenu prend la forme d'une liste  $\mathcal{P}$  de tous les pics détectés au sein du spectre. Cependant, ces pics s'organisent pour la plupart sous la forme de structures dans la densité spectrale, ce qui ne se reflète pas dans l'ensemble  $\mathcal{P}$ . La deuxième étape consiste donc à identifier les sous-ensembles de  $\mathcal{P}$ , formant des structures spectrales afin de redonner tout son sens au contenu extrait.

L'identification des séries harmoniques et l'identification des bandes de modulation sont réalisées à partir d'une méthode commune, reposant sur l'intersection d'intervalles fréquentiels. Cette méthode permet de prendre en compte l'indétermination sur les positions fréquentielles exactes des pics, conséquence de leur estimation. La recherche de séries est exhaustive, permettant de retrouver l'ensemble  $\mathcal{H}$  de toutes les séries harmoniques du spectre, même celles dont le fondamental est absent. Pour des soucis de temps de calcul, la recherche de l'ensemble  $\mathcal{M}$  des modulations ne peut pas être exhaustive. Cependant, elle est réalisée de manière pragmatique, aux fréquences les plus intéressantes du spectre.

L'ensemble des travaux présentés au cours de ce chapitre a fait l'objet d'une publication [Gerber 2013] et a donné lieu à un dépôt de brevet [Martin 2013]. De plus, dans le cadre des projets européens [*KAStrion*] et [*SUPREME*], la méthodologie présentée a été transférée à un industriel partenaire, afin de l'intégrer dans leur nouveau CMS.

Si la lecture du spectre d'un signal est une étape importante pour le diagnostic du système, elle ne fait pas tout. Le suivi dans le temps des structures spectrales permettra de visualiser l'évolution du système et son éventuelle dégradation.

# Suivi temps-fréquence des structures spectrales

---

## Sommaire

<b>3.1</b>	<b>Contexte</b> . . . . .	<b>54</b>
3.1.1	Carte temps-fréquence . . . . .	54
3.1.2	État de l'art des méthodes de suivi de structures spectrales . . . . .	57
3.1.3	Choix de la méthode de suivi . . . . .	58
<b>3.2</b>	<b>Méthode de suivi</b> . . . . .	<b>59</b>
3.2.1	Méthode de McAulay & Quatieri . . . . .	59
3.2.2	Adaptations et améliorations de la méthode . . . . .	61
3.2.3	Paramétrisation de la méthode de suivi . . . . .	66
<b>3.3</b>	<b>Suivi de structures spectrales</b> . . . . .	<b>67</b>
3.3.1	Suivi de séries harmoniques . . . . .	68
3.3.2	Suivi de bandes de modulation . . . . .	70
3.3.3	Stratégie de suivi . . . . .	70
3.3.4	Complexité calculatoire de la méthode de suivi proposée . . . . .	72
<b>3.4</b>	<b>Applications</b> . . . . .	<b>73</b>
3.4.1	Simulation d'un défaut . . . . .	73
3.4.2	Surveillance d'un système mécanique . . . . .	77
<b>3.5</b>	<b>Synthèse du chapitre</b> . . . . .	<b>80</b>

---

L'analyse d'un signal et l'identification des structures spectrales qu'il contient permet de décrire l'état du système au moment de l'acquisition du signal. Cette analyse peut suffire à un expert afin d'établir un diagnostic du système. Cependant, pour conforter ce diagnostic, il est intéressant de comparer ce signal à un signal de référence de l'état sain du système ou à des signaux acquis antérieurement afin de visualiser l'évolution de la santé du système. Cette étape de confrontation est d'autant plus nécessaire lorsque l'on souhaite se passer de l'étape de diagnostic manuel, dans le but d'automatiser la surveillance.

L'objectif est d'associer les structures spectrales identiques entre les signaux lors de leurs confrontations. Par identique, nous entendons qu'une structure spectrale est la conséquence d'un phénomène physique, et que chacun des signaux capture une réalisation de ce phénomène à la date d'acquisition du signal. L'association d'une structure spectrale sur l'ensemble des signaux correspond à ce que nous appellerons le suivi temporel de la structure. Ce suivi sera

réalisé à partir des fréquences caractéristiques des structures, c'est pourquoi nous parlerons dans la suite de suivi temps-fréquence.

Au cours de ce chapitre, nous commencerons par faire un état de l'art des méthodes de suivi temps-fréquence de structures spectrales afin de sélectionner celle répondant au mieux aux exigences fixées dans cette thèse. Nous présenterons ensuite les améliorations et modifications proposées dans le but d'adapter complètement la méthode sélectionnée à notre contexte. Le suivi des différents types de structures sera alors réalisé dans un ordre permettant d'éviter la redondance, tout en couvrant l'ensemble du contenu spectral des signaux. Enfin, l'utilisation de la méthode proposée sur des séquences de signaux réels permettra de valider cette approche.

### 3.1 Contexte

Un système de surveillance acquiert de nombreux signaux au cours du temps. Ces acquisitions peuvent être réalisées de manière périodique, ou bien de façon tout à fait irrégulière. Dans tous les cas, on obtient une séquence de signaux  $\mathcal{S} = \{s_1, \dots, s_n\}$ , où  $n$  correspond au nombre total de signaux acquis jusqu'à présent. Chaque signal  $s_k \in \mathcal{S}$  est acquis à partir de la date  $t_k$ , pendant une durée  $\tau_k$ . La séquence est ordonnée par date d'acquisition croissante, ce qui signifie que

$$\forall k \in \{1, \dots, n-1\}, t_{k+1} > t_k. \quad (3.1)$$

L'analyse individuelle de chacun des signaux de la séquence  $\mathcal{S}$ , conduite suivant les méthodes présentées au chapitre précédent, va donner une carte temps-fréquence que nous présenterons dans la section suivante. Ensuite, nous nous intéresserons de manière précise au problème de suivi de structures spectrales en présentant un état de l'art sur le sujet dans la section 3.1.2.

#### 3.1.1 Carte temps-fréquence

Chaque signal  $s_k \in \mathcal{S}$  est analysé individuellement. Cela signifie que le contenu spectral de chaque signal  $s_k$  est extrait. Il prend la forme d'une liste de pics spectraux  $\mathcal{P}_k$ , à partir de laquelle nous récupérons la liste des séries harmoniques  $\mathcal{H}_k$  et la liste des bandes de modulation  $\mathcal{M}_k$ , grâce à la méthode exposée au chapitre 2. L'indice  $k$  ajouté aux ensembles  $\mathcal{P}_k$ ,  $\mathcal{H}_k$  et  $\mathcal{M}_k$  fait référence à l'indice du signal considéré dans la séquence  $\mathcal{S}$ . Cet indice se propage également aux éléments des ensembles et à leurs différentes caractéristiques. Ainsi, l'ensemble  $\mathcal{P}_k$  est défini comme

$$\mathcal{P}_k = \{p_k^i\}_{i \in \{1, \dots, \#\mathcal{P}_k\}}, \quad (3.2)$$

où  $p_k^i$  représente le  $i^{\text{ème}}$  pic détecté dans le spectre du signal  $s_k$ . La fréquence du pic  $p_k^i$  se note alors  $f_k^i$ . Pour rappel, l'ensemble  $\mathcal{P}_k$  est ordonné par fréquence croissante :

$$\forall i \in \{1, \dots, \#\mathcal{P}_k - 1\}, f_k^{i+1} > f_k^i. \quad (3.3)$$

De la même manière, les ensembles  $\mathcal{H}_k$  et  $\mathcal{M}_k$  sont définis comme

$$\mathcal{H}_k = \{h_k^i\}_{i \in \{1, \dots, \#\mathcal{H}_k\}}, \quad (3.4)$$

$$\mathcal{M}_k = \{m_k^i\}_{i \in \{1, \dots, \#\mathcal{M}_k\}}, \quad (3.5)$$

avec  $h_k^i$  et  $m_k^i$  les  $i^{\text{èmes}}$  série harmonique et bande de modulation, respectivement, identifiées parmi l'ensemble de pics  $\mathcal{P}_k$ , lui-même extrait du signal  $s_k$ . L'ensemble des séries harmoniques  $\mathcal{H}_k$  est ordonné par fréquence fondamentale croissante, alors que l'ensemble  $\mathcal{M}_k$  est ordonné par fréquence de modulation croissante.

Chaque pic  $p_k^i$  de la séquence de signaux peut être représenté sur une carte temps-fréquence comme un élément ponctuel au temps  $t_k$  et à la fréquence  $f_k^i$ . Un exemple factice de carte temps-fréquence est donné figure 3.1. Volontairement très simple, cette carte permet également de bien comprendre les notations utilisées dans ce chapitre. Afin d'illustrer également les structures spectrales sur la carte, les pics appartenant à une même structure sont dessinés avec une combinaison de forme et de couleur identique, exception faite des cercles gris, représentant les pics n'appartenant à aucune structure spectrale.

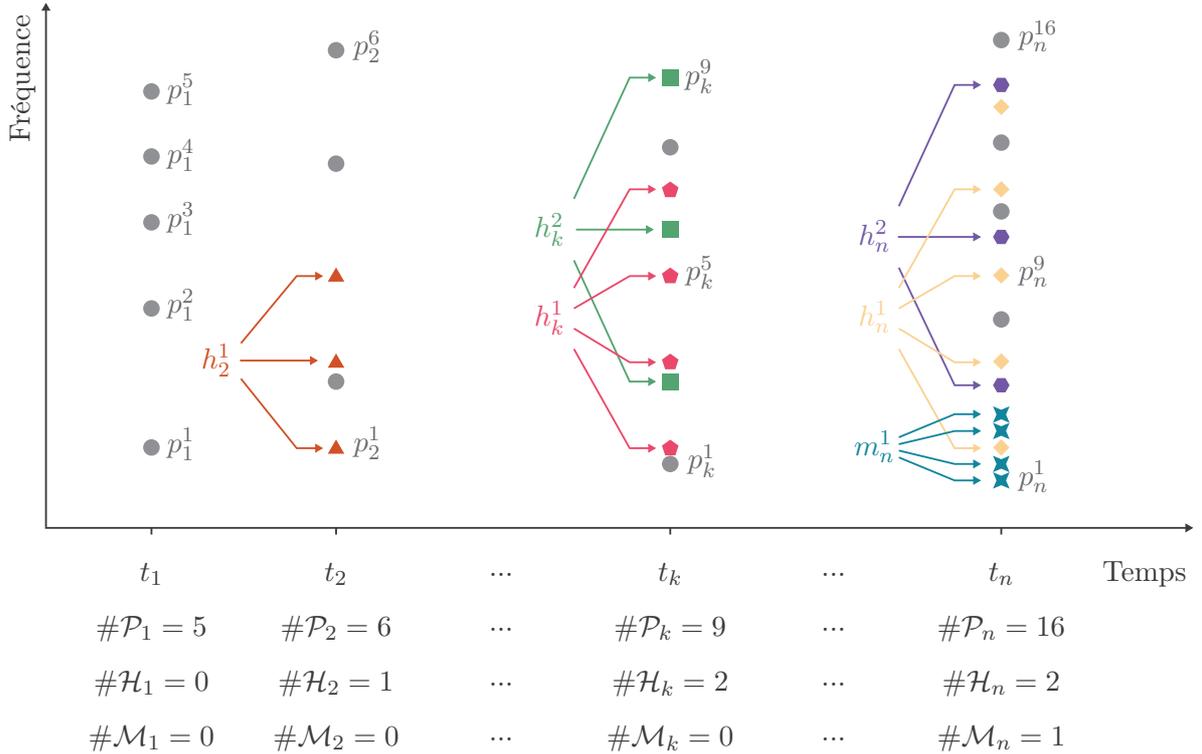


FIGURE 3.1 – Un exemple de carte temps-fréquence résumant l'information extraite à partir des signaux de la séquence  $\mathcal{S} = \{s_1, \dots, s_n\}$ . Chaque pic détecté est représenté par la combinaison d'une forme et d'une couleur. Les pics appartenant à la même structure spectrale sont dessinés avec la même combinaison de forme et de couleur, alors que les disques gris représentent les pics n'appartenant à aucune structure spectrale.

La carte temps-fréquence associée à une séquence  $\mathcal{S}$  de signaux est une manière de syn-

thétiser l'ensemble des informations extraites individuellement à partir de chacun des signaux de la séquence. Cette synthèse n'est cependant pas suffisante pour effectuer la surveillance et le diagnostic automatique d'un système. En effet, elle demande une étude minutieuse afin d'évaluer l'évolution du système et de conclure sur son éventuelle dégradation.

Afin de suivre l'évolution du système, nous proposons d'établir un lien entre les analyses des différents signaux de la séquence. Ce lien prend la forme de trajectoires spectrales, permettant de suivre un pic ou une structure spectrale d'une date d'acquisition à la suivante. La figure 3.2 illustre ce principe et propose trois trajectoires pour une carte temps-fréquence semblable à celle de la figure 3.1. On y voit une trajectoire de pics  $t_{\mathcal{P}}^1$  (en rose), démarrante à l'instant  $t_1$  et se terminant à l'instant  $t_4$ , ainsi que deux trajectoires de séries harmoniques  $t_{\mathcal{H}}^1$  (en orange) et  $t_{\mathcal{H}}^2$  (en vert). Ces dernières se terminent toutes les deux à l'instant  $t_4$ , et démarrent aux instants  $t_2$  et  $t_3$ , respectivement.

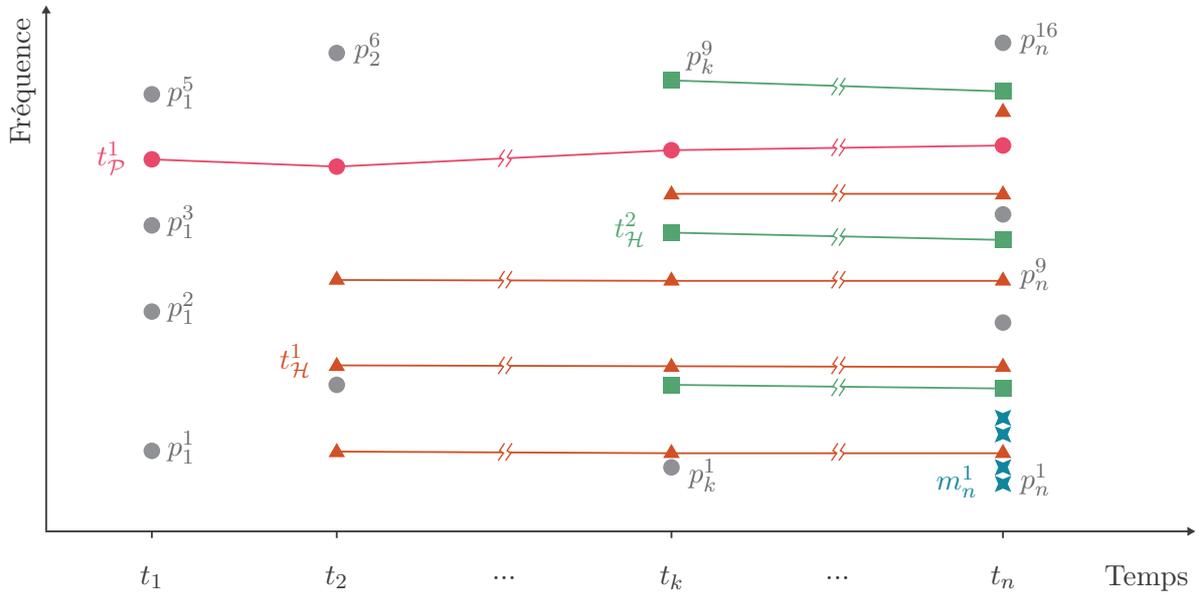


FIGURE 3.2 – Trois trajectoires représentées sur la carte temps-fréquence de la figure 3.1 : en rose, la trajectoire de pics  $t_{\mathcal{P}}^1$  ; en orange et en vert, les trajectoires de séries harmoniques  $t_{\mathcal{H}}^1$  et  $t_{\mathcal{H}}^2$ , respectivement.

Cette première illustration du principe de trajectoires fait apparaître un des points clés que la méthode de suivi devra être capable de gérer. Nous avons trois types de structures spectrales à suivre. Nous devons donc identifier trois types de trajectoires : l'ensemble des trajectoires de pics  $\mathcal{T}_{\mathcal{P}}$ , l'ensemble des trajectoires de séries harmoniques  $\mathcal{T}_{\mathcal{H}}$  et l'ensemble des trajectoires de modulations  $\mathcal{T}_{\mathcal{M}}$ .

En plus de cela, les contraintes fixées dans cette thèse continuent à s'appliquer. La méthode de suivi doit donc être le plus générique possible et doit nécessiter le moins de ressources possibles, surtout en termes de temps de calcul. Nous garderons ces trois critères à l'esprit lors de l'état de l'art sur les méthodes de suivi, afin de choisir la plus adaptée aux contraintes.

### 3.1.2 État de l'art des méthodes de suivi de structures spectrales

Il existe déjà de nombreuses méthodes de suivi de composantes spectrales au cours du temps dans la littérature. L'une des plus utilisées est celle développée par McAulay et Quatieri [McAulay 1986]. Il s'agit d'une méthode simple et intuitive pour réaliser le suivi de pics spectraux. Elle est basée sur l'utilisation de la fréquence des pics : un pic  $p_k^i$  de fréquence  $f_k^i$  à un instant  $t_k$  cherche à s'associer avec un pic  $p_{k+1}^j$  à l'instant  $t_{k+1}$ , dont la fréquence doit être suffisamment proche de  $f_k^i$ , suivant un critère qui sera précisé plus loin dans ce chapitre. Si un tel pic existe, la trajectoire se crée ou se prolonge. Il peut arriver qu'un conflit se déclare si deux pics à l'instant  $t_k$  se disputent le même pic à l'instant  $t_{k+1}$ . Dans la méthode originale, McAulay et Quatieri proposent dans ce cas de chercher quelle serait la meilleure association du point de vue du pic à l'instant  $t_{k+1}$ . Cependant, des variantes existent et règlent le conflit différemment. Dans [Smith 1987], le pic le plus proche en fréquence gagne le conflit, alors que dans [Yamahata 2009], un second critère basé sur l'amplitude vient départager les candidats multiples.

D'autres variantes de la méthode existent. Les modifications ne portent plus sur la gestion des conflits, mais sur le placement de l'intervalle de recherche. En effet, dans [Serra 1997, Serra 1990, Lagrange 2003, Pendharkar 2012], les différents auteurs proposent de prédire la fréquence du pic, en fonction de l'historique de la trajectoire. Ils placent alors l'intervalle de recherche autour de la fréquence prédite au lieu de la fréquence du dernier pic dans la trajectoire. Cette manière de procéder améliore le suivi de pics dont la fréquence varie rapidement, comme dans un enregistrement audio. Cependant, les méthodes de prédictions demandent à ce que les signaux de la séquence soient régulièrement espacés en temps, ce qui n'est pas nécessairement le cas pour la surveillance d'un système.

Dans [Chen 2006], la méthode originale est appliquée, mais elle est complétée par un traitement des trajectoires identifiées. Seules les trajectoires répondant à des critères précis en termes de fréquence ou d'énergie sont conservées. Finalement, la méthode de McAulay et Quatieri répond bien au critère de généricité, car elle est utilisable dans de nombreux contextes (traitement audio, séparation de sources ou classification de sons), ainsi qu'au critère de performance, car la méthode est simple et son implémentation a une faible complexité.

D'autres méthodes permettent également le suivi temps-fréquence de pics spectraux. Dans [Depalle 1993, Quinn 2001, Doval 1993], une méthode de suivi basée sur un modèle de Markov caché est présentée. Pour retrouver la séquence des états cachés, l'algorithme de Viterbi [Viterbi 1967] est utilisé. Ce dernier a une forte complexité, ce qui rend ces méthodes applicables uniquement à un faible nombre de signaux, à une petite bande spectrale, ou à un nombre limité de trajectoires. Dans [Wolcin 1980], l'auteur souhaite détecter une trajectoire noyée dans un bruit blanc Gaussien. Il propose d'utiliser une méthode basée sur un maximum *a posteriori* pour réaliser la détection de cette unique trajectoire. Dans le cas où plusieurs trajectoires seraient présentes, il est éventuellement possible d'itérer la méthode : chaque trajectoire identifiée est soustraite ou annulée dans la TFD du signal avant de procéder à une nouvelle détection. Cependant, la détection de plusieurs trajectoires ne fonctionne que si les trajectoires ne se chevauchent pas sur tout l'intervalle d'observation. Une autre

méthode [Mazet 2013], basée sur des contraintes de parcimonie, permet d'identifier les lignes spectrales en spectroscopie. Cependant, la méthode repose sur « l'algorithme hongrois » proposé dans [Kuhn 1955] pour identifier correctement les trajectoires ; la complexité de cet algorithme d'optimisation combinatoire est en  $\mathcal{O}(n^3)$ , ce qui engendre de longs temps de calcul pour des séquences constituées d'un nombre important de signaux, contenant chacun de nombreux pics.

Il existe aussi des méthodes de suivi de séries harmoniques. Elles sont très utilisées pour suivre la fréquence fondamentale de la mélodie dans un enregistrement audio, qui présente naturellement une structure harmonique. Dans [Maher 1994], un peigne de raies spectrales est utilisé pour estimer et suivre la série harmonique représentant la mélodie. Une extension de la méthode permet de suivre deux séries harmoniques en même temps, mais uniquement si leurs trajectoires ne se chevauchent pas. Dans [Duan 2014], le nombre de séries harmoniques suivies peut être bien plus grand. Les trajectoires harmoniques des différents sons polyphoniques sont reconstituées à l'aide d'une méthode reposant sur une classification contrainte assurant la cohérence du timbre dans la trajectoire. Cependant, le nombre de séries à suivre est une information nécessaire à la méthode pour fonctionner. Les auteurs de [Le Coz 2013] proposent d'effectuer d'abord l'identification des trajectoires de pics, avant de reconstruire les trajectoires harmoniques en groupant les trajectoires de pics appropriées entre elles. Le regroupement s'effectue suivant une distance utilisant les informations d'amplitudes et de fréquences, introduisant par la même occasion un modèle *a priori* sur les séries harmoniques.

### 3.1.3 Choix de la méthode de suivi

La méthode de suivi temps-fréquence que nous souhaitons utiliser doit idéalement vérifier les trois critères suivants :

1. pouvoir suivre tous les types de structures spectrales, à savoir les pics, les séries harmoniques et les bandes de modulation ;
2. être générique, la méthode ne doit nécessiter aucune information *a priori* ;
3. avoir de bonnes performances, en particulier en termes de temps de calcul.

Cependant, à notre connaissance, aucune des méthodes présentées précédemment ne vérifie les trois critères simultanément. En conséquence, nous proposons d'améliorer la méthode répondant le mieux aux exigences fixées. Nous avons donc choisi d'utiliser la méthode de McAulay et Quatieri. En effet, cette dernière est très simple et s'exécute rapidement. Il est donc possible de l'utiliser sur des grandes quantités de données. De plus, elle peut fonctionner « en ligne », c'est-à-dire qu'il est possible de compléter les trajectoires déjà identifiées jusque-là, lors de l'acquisition d'un nouveau signal. Enfin, elle ne nécessite aucune information *a priori* sur les données et le modèle sur lequel elle repose ne fait que peu d'hypothèses. Cependant, cette méthode ne permet de suivre que des pics spectraux. Elle n'est pas adaptée au suivi des séries harmoniques ou des bandes de modulation. Nous verrons, après avoir présenté la méthode en détail, une stratégie permettant de suivre également ces structures.

## 3.2 Méthode de suivi

La méthode de McAulay et Quatieri est présentée en détail dans la section suivante. Les notations utilisées ne sont pas celles de l'article original [McAulay 1986], mais celles propres à ce manuscrit. Ensuite, les améliorations proposées de la méthode seront expliquées dans la section 3.2.2.

### 3.2.1 Méthode de McAulay & Quatieri

Dans leur article [McAulay 1986], McAulay et Quatieri se basent sur l'idée d'une modélisation sinusoïdale du son [Lagrange 2004] : un signal de parole est exprimé comme la somme de sinusoïdes dont les amplitudes, les fréquences et les phases varient au cours du temps. Ils proposent une technique d'analyse permettant d'identifier ces multiples sinusoïdes et de suivre leurs variations au cours du temps. Le suivi d'une sinusoïde forme alors un *partiel* ; autrement dit, une trajectoire de pics. Il est possible, à partir de la liste de tous les partiels, de reconstruire le signal d'origine grâce à une opération de synthèse basée sur la somme de tous les partiels. Cette modélisation sinusoïdale du signal de parole ouvre alors des perspectives comme la modification de la hauteur du son ou le codage bas débit d'un signal de parole.

Pour passer à une représentation sous la forme de partiels, l'analyse d'un signal de parole se déroule en trois étapes principales :

1. le découpage du signal en plusieurs segments courts (quelques millisecondes) ;
2. l'analyse de chaque segment pour extraire les pics spectraux ;
3. le suivi des pics de segments en segments.

La première étape découpe un signal audio  $s$  en une séquence de signaux  $\mathcal{S} = \{s_1, \dots, s_n\}$ . La longueur des segments est courte, à peine quelques millisecondes. L'opération de découpage et le choix de la longueur des segments sont motivés par le fait qu'un signal audio est très variable au cours du temps, c'est-à-dire qu'il est non stationnaire. Or la stationnarité d'un signal est nécessaire afin d'estimer sa densité spectrale et d'extraire l'information contenue dans le signal. En choisissant une longueur de segment suffisamment petite, le signal peut alors être considéré comme localement stationnaire sur la durée du segment.

La courte longueur des segments engendre également un second point clef de la méthode. Les segments étant consécutifs, le temps écoulé entre  $t_k$  et  $t_{k+1}$  est très faible. Cela signifie que le signal audio  $s$  n'a eu que très peu de temps pour évoluer entre les segments  $s_k$  et  $s_{k+1}$ . Les différences entre les spectres de ces deux segments sont donc elles aussi très faibles. Autrement dit, un pic  $p_k^i$  à la fréquence  $f_k^i$  à l'instant  $t_k$  aura une forte probabilité d'être présent autour de la même valeur de fréquence à l'instant  $t_{k+1}$ . Il y a donc une *faible variabilité des fréquences* entre deux segments consécutifs et c'est sur cette hypothèse qu'est construite la méthode de suivi. En particulier, la méthode définit un intervalle de recherche  $\Delta f$  qui limite l'excursion possible d'une fréquence entre deux instants consécutifs. Le choix de cet intervalle est basé, entre autres, sur la longueur des segments.

Le suivi des pics est réalisé de manière itérative à la fois en temps et en fréquence. En effet, le suivi est séquentiel, ce qui signifie que pour chaque paire de segments consécutifs  $s_k$  et  $s_{k+1}$ , les pics  $p_k^i \in \mathcal{P}_k$  (détectés dans le spectre de  $s_k$ ) sont associés aux pics  $p_{k+1}^j \in \mathcal{P}_{k+1}$  (détectés dans le spectre de  $s_{k+1}$ ). Ce processus est itéré pour tous les indices  $k \in \{1, \dots, n-1\}$ , avec  $n$  le nombre de signaux dans la séquence  $\mathcal{S}$ . De plus, le suivi est réalisé pic par pic, en commençant par le pic ayant la fréquence la plus basse.

Supposons que le suivi a été réalisé jusqu'au temps  $t_k$  et jusqu'au pic  $p_k^{i-1}$ . Il s'agit maintenant d'associer le pic  $p_k^i$  avec un pic de l'ensemble  $\mathcal{P}_{k+1}$ , à l'instant  $t_{k+1}$ . L'opération de suivi se déroule en deux étapes, détaillées ci-dessous et illustrées figure 3.3 :

1. On cherche un candidat à associer à  $p_k^i$  parmi les pics de  $\mathcal{P}_{k+1}$  dont la fréquence se situe dans l'intervalle  $\left[ f_k^i - \frac{\Delta f}{2}, f_k^i + \frac{\Delta f}{2} \right]$ . Trois cas peuvent se produire :
  - aucun candidat n'est présent dans l'intervalle de recherche : la trajectoire du pic  $p_k^i$  se meurt (figure 3.3 (a)) ;
  - un seul et unique pic  $p_{k+1}^j$  est présent dans l'intervalle de recherche : il est sélectionné comme candidat à l'association (figure 3.3 (b) et (d)) ;
  - plusieurs pics sont présents dans l'intervalle de recherche : le pic dont la fréquence est la plus proche de  $f_k^i$  est élu comme le meilleur candidat (figure 3.3 (c)).

Cela signifie que si le pic  $p_{k+1}^j$  est le meilleur candidat pour  $p_k^i$ , les inégalités suivantes sont vérifiées :

$$\left| f_k^i - f_{k+1}^j \right| < \frac{\Delta f}{2}, \quad (3.6)$$

et

$$\forall l \in \{1, \dots, \#\mathcal{P}_{k+1}\} \text{ tel que } l \neq j, \left| f_k^i - f_{k+1}^j \right| < \left| f_k^i - f_{k+1}^l \right|. \quad (3.7)$$

Chacun des ensembles  $\mathcal{P}_k$  et  $\mathcal{P}_{k+1}$  est ordonné par fréquence croissante, ce qui signifie que (3.7) peut se vérifier à l'aide de deux inégalités uniquement :

$$\left| f_k^i - f_{k+1}^j \right| < \left| f_k^i - f_{k+1}^{j-1} \right| \text{ et } \left| f_k^i - f_{k+1}^j \right| < \left| f_k^i - f_{k+1}^{j+1} \right|. \quad (3.8)$$

2. Le meilleur candidat élu à l'étape 1 doit vérifier la condition de compatibilité inverse, c'est-à-dire que la fréquence du pic élu doit être plus proche de la fréquence  $f_k^i$  que de la fréquence de n'importe quel autre pic de l'ensemble  $\mathcal{P}_k$ . Cela permet de vérifier que le pic élu n'a pas une meilleure association à faire avec un autre pic à l'instant  $t_k$ . Dans le cas où le meilleur candidat du pic  $p_k^i$  est le pic  $p_{k+1}^j$ , la condition s'écrit sous la forme de l'inégalité suivante<sup>1</sup> :

$$\delta a < \delta b, \text{ avec } \delta a = \left| f_k^i - f_{k+1}^j \right| \text{ et } \delta b = \left| f_k^{i+1} - f_{k+1}^j \right|. \quad (3.9)$$

Lorsque la condition de compatibilité inverse est vérifiée, le meilleur candidat est ajouté à la trajectoire du pic  $p_k^i$  (figure 3.3 (b)). Dans le cas contraire, il reste deux possibi-

---

1. L'équation (3.9) peut paraître insuffisante pour vérifier que le pic  $p_k^i$  est rétrospectivement le meilleur candidat du pic  $p_{k+1}^j$ , surtout si on la compare à (3.8). Cependant, nous sommes dans un processus de recherche itératif, ce qui implique que certaines inégalités ont déjà été testées aux itérations précédentes. L'inégalité (3.9) est donc bien suffisante pour vérifier la compatibilité inverse.

lités. Si le meilleur candidat  $p_{k+1}^j$  ne vérifie pas la condition de compatibilité inverse mais que le pic de fréquence inférieure, à savoir  $p_{k+1}^{j-1}$ , est présent dans l'intervalle de recherche, l'association est immédiatement réalisée avec  $p_{k+1}^{j-1}$  (figure 3.3 (c)). Autrement, la trajectoire du pic  $p_k^i$  meurt (figure 3.3 (d)).

À la fin des deux étapes, si des pics ont été associés, ils sont retirés des listes de pics pour ne plus être considérés aux itérations suivantes. Dans tous les cas, on passe alors à l'itération suivante consistant à suivre le pic  $p_k^{i+1}$ .

Une fois que tous les pics de l'ensemble  $\mathcal{P}_k$  ont vu leur trajectoire se prolonger ou mourir, il peut rester des pics à l'instant  $t_{k+1}$  qui n'ont pas été associés. Chacun des pics  $p_{k+1}^j$  non associés démarre alors une nouvelle trajectoire, dans laquelle il est le premier et unique pic à l'instant  $t^{k+1}$ . Cela signifie que tous les pics de la carte temps-fréquence appartiennent à une trajectoire.

### 3.2.2 Adaptations et améliorations de la méthode

Afin d'utiliser la méthode de suivi de pics de McAulay et Quatieri dans un contexte de surveillance, 4 adaptations de la méthode originale sont proposées. Elles modifient l'hypothèse à respecter et la paramétrisation de la méthode, en plus d'ajouter une fonctionnalité et un post-traitement.

#### 3.2.2.1 Adaptation aux signaux issus de la surveillance de système

La méthode de McAulay et Quatieri a été développée dans le but de suivre des pics sur les segments réguliers et contigus d'un signal de parole. Ce contexte particulier permet de vérifier l'hypothèse de faible variabilité des fréquences entre deux instants consécutifs. Cependant, le contexte de la surveillance de système est bien différent. L'acquisition des signaux n'est en général pas continue, pour éviter l'explosion de la taille de la base de données. De plus, les acquisitions ne sont pas nécessairement régulières dans le temps. Ainsi, l'état du système surveillé peut énormément changer entre deux acquisitions de signaux. Cet éventuel changement risque de poser problème, en particulier dans le cas de machines tournantes, pour lesquelles les fréquences caractéristiques dépendent de la vitesse de rotation en entrée du système.

Nous allons donc faire une nouvelle hypothèse, pour être sûr que l'hypothèse de faible variabilité des fréquences soit respectée et que nous puissions utiliser la méthode de suivi temps-fréquence. L'hypothèse est que l'acquisition des signaux est faite à *état machine constant*, c'est-à-dire que les paramètres opérationnels sont identiques entre toutes les acquisitions. Dans le cas d'une machine tournante par exemple, cela signifie que l'acquisition des signaux doit toujours être faite à la même vitesse de rotation. Cette contrainte étant relativement forte, il n'est pas toujours possible de la respecter. Heureusement, il existe des méthodes de ré-échantillonnage angulaire [Bonnardot 2005, Wang 2014, Firla 2014], qui agissent comme un pré-traitement des signaux. Les fréquences sont alors normalisées par la vitesse de rotation, ce

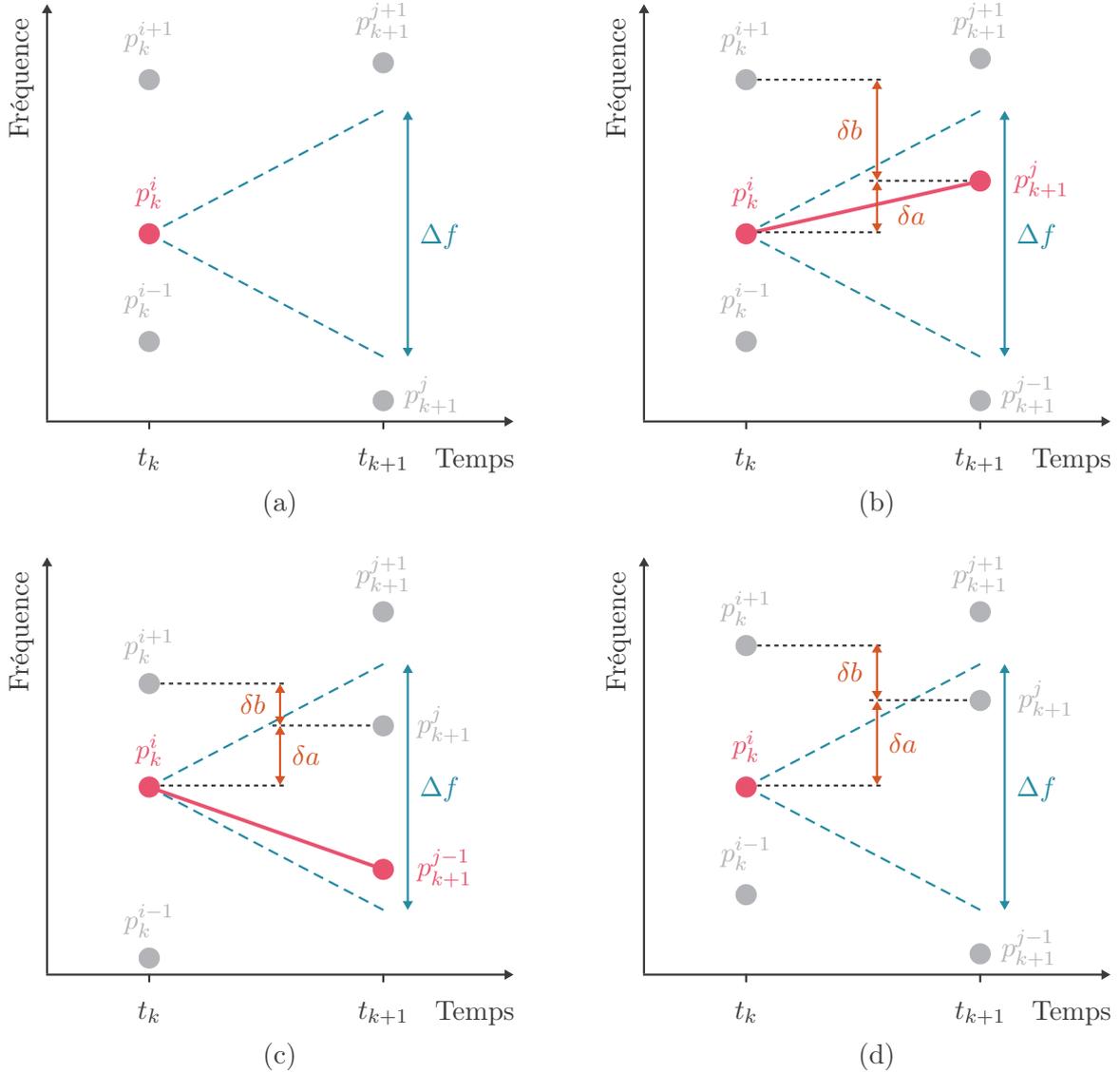


FIGURE 3.3 – Suivi du pic  $p_k^i$  (en rose) : (a) aucun pic n'est présent dans l'intervalle de recherche  $\Delta f$  (en bleu), la trajectoire meurt ; (b) le seul pic présent dans l'intervalle de recherche vérifie la condition de compatibilité inverse  $\delta a < \delta b$  (en orange), le pic  $p_{k+1}^j$  est alors ajouté à la trajectoire du pic  $p_k^i$  ; (c) le pic le plus proche ne vérifie pas la condition de compatibilité inverse, l'association est alors faite avec le pic de fréquence immédiatement inférieure,  $p_{k+1}^{j-1}$ , également présent dans l'intervalle de recherche ; (d) le seul pic présent dans l'intervalle de recherche ne vérifie pas la condition de compatibilité inverse, la trajectoire meurt.

qui permet de vérifier l'hypothèse de faible variabilité des fréquences. Pour plus d'informations sur le ré-échantillonnage angulaire, se référer à l'annexe B.2.1.

### 3.2.2.2 Choix de l'intervalle de recherche $\Delta f$

La méthode de suivi de McAulay et Quatieri possède un unique paramètre qu'il faut définir, à savoir la largeur de l'intervalle de recherche  $\Delta f$ . Dans la méthode originale,  $\Delta f$  est une constante fixée à l'avance. Le choix de la constante est basé sur la variation maximale autorisée d'une fréquence entre deux segments. Indirectement, cela signifie que ce choix dépend de la longueur des segments utilisée lors de la découpe du signal de parole et d'un modèle de voix. Dans l'idée de proposer une méthode de suivi la plus générique possible, nous avons décidé de ne pas utiliser une constante mais de baser le choix de cette valeur sur les données.

Lors de l'ajout du signal  $s_{k+1}$  au suivi déjà réalisé des signaux  $s_1$  à  $s_k$ , les pics à l'instant  $t_k$  sont associés aux pics à l'instant  $t_{k+1}$ . Or, dans méthodologie proposée, la fréquence  $f_k^i$  de chaque pic  $p_k^i$  est estimée. Il existe un intervalle d'indétermination d'amplitude  $\varepsilon_k^i$  quant à la position fréquentielle exacte du pic. La taille de cet intervalle dépend généralement de la longueur des signaux et de la méthode d'estimation utilisée lors de la détection des pics. Or, comme expliqué dans l'annexe B.2.4, la détection des pics est réalisée à partir de plusieurs estimateurs différents. En conséquence, deux pics identifiés dans le même signal peuvent avoir deux intervalles d'indétermination différents, en fonction des méthodes ayant permis leur détection. C'est encore plus vrai si l'on compare un pic à l'instant  $t_k$  et un autre à l'instant  $t_{k+1}$ , dans la mesure où les longueurs des signaux  $s_k$  et  $s_{k+1}$  peuvent être différentes.

Pour prendre correctement en compte les intervalles d'indétermination lors de l'association des pics  $p_k^i$  aux pics  $p_{k+1}^j$ , il faudrait procéder comme dans la section 2.2.1 en réalisant des intersections d'intervalles entre les instants  $t_k$  et  $t_{k+1}$ . Cependant, ce changement entraînerait des coûts calculatoires plus importants et nous éloignerait de la simplicité de la méthode originale. Pour cette raison, le choix a été fait de ne pas prendre en compte tous les intervalles d'indétermination. Nous proposons de fixer l'intervalle de recherche  $\Delta f$  à

$$\max_{p_k^i \in \{\mathcal{P}_k \cup \mathcal{P}_{k+1}\}} \varepsilon_k^i, \quad (3.10)$$

c'est-à-dire la valeur du plus grand intervalle d'indétermination pour l'association des pics  $p_{k+1}^j$  aux pics  $p_k^i$ .

Un tel choix pour l'intervalle de recherche  $\Delta f$  n'autorise pas les fréquences à évoluer au cours de la séquence de signaux. Il faut que les fréquences soient identiques entre deux signaux, aux erreurs d'estimation près, afin de former des trajectoires. Cette restriction forte n'est pas en adéquation à un contexte de surveillance de système mécanique. En effet, les fréquences caractéristiques de certains éléments mécaniques comme les roulements ne sont pas entièrement déterministes [Randall 2011]. Des variations de fréquence apparaissent au cours du temps, dues à divers phénomènes physiques comme le glissement. Il est donc nécessaire d'élargir l'intervalle de recherche afin de permettre aux trajectoires de suivre ces variations.

Nous proposons donc de modifier (3.10) en ajoutant un facteur multiplicatif  $k_\Delta$  :

$$\Delta f = k_\Delta \times \max_{p_l^i \in \{\mathcal{P}_k \cup \mathcal{P}_{k+1}\}} \varepsilon_l^i. \quad (3.11)$$

Le choix de la valeur du facteur multiplicatif dépend des conditions d'application du suivi. Si l'hypothèse d'*état machine constant* est bien respectée, il est possible de lui attribuer une petite valeur. Les expériences qui seront présentées dans la section 4.1.2 respectent bien cette hypothèse et montrent que les variations de fréquence nécessitent au minimum que  $k_\Delta = 2$  pour suivre toutes les trajectoires identifiant un défaut. Cependant, dans le cas où l'hypothèse d'*état machine constant* n'est pas entièrement respectée, il est possible d'augmenter la valeur de  $k_\Delta$  proportionnellement aux variations des conditions de fonctionnement du système afin de suivre les variations des fréquences caractéristiques des composants mécaniques. Il n'est cependant pas conseillé d'utiliser une trop grande valeur pour  $k_\Delta$  car cela nuit à la qualité du suivi et à la stabilité fréquentielle des trajectoires.

La valeur de  $\Delta f$  peut évoluer lors des différentes itérations de l'algorithme de suivi. En effet, sa valeur est recalculée à chaque nouveau pas temporel et ce calcul est basé sur les caractéristiques des données en entrée de l'algorithme de suivi.

### 3.2.2.3 État sommeil d'une trajectoire

La méthode originale de suivi n'est pas robuste face à l'absence de détection d'un pic à un instant donné. En effet, si aucun candidat n'est présent à l'instant  $t_{k+1}$  pour continuer une trajectoire terminant à l'instant  $t_k$ , la trajectoire meurt immédiatement. Et ce même si des pics qui auraient pu être des candidats à la même trajectoire sont présents aux instants  $t_{k+2}, t_{k+3}$ , etc.

Pour pallier ce phénomène d'interruption de trajectoires, nous proposons d'introduire l'*état sommeil*, comme décrit dans [Serra 1997]. Une trajectoire peut s'endormir durant un ou plusieurs instants si elle ne trouve pas de pics pour la compléter durant ces instants-là. Dès qu'un pic est finalement associé à la trajectoire endormie, cette dernière se réveille. Une trajectoire peut s'endormir et se réveiller autant de fois que nécessaire. Cependant, si la trajectoire est endormie durant une trop longue période, elle sera finalement considérée comme morte. Nous avons fixé à 2 le nombre maximal d'instants consécutifs pendant lesquels une trajectoire peut être endormie.

Le concept d'état sommeil est illustré sur la figure 3.4. Il s'agit d'un exemple factice de carte temps-fréquence sur laquelle on peut voir le suivi des pics réalisé jusqu'à l'instant  $t_k$ . Parmi les 6 trajectoires identifiées, les trajectoires  $t_{\mathcal{P}}^3$  et  $t_{\mathcal{P}}^6$  ne sont jamais en sommeil et sont représentées en trait plein. La trajectoire  $t_{\mathcal{P}}^1$  s'est endormie durant l'instant  $t_{k-2}$ ; la transition entre les instants  $t_{k-3}$  et  $t_{k-1}$  est représentée en pointillé pour signaler cette interruption momentanée. Les trajectoires  $t_{\mathcal{P}}^2$  et  $t_{\mathcal{P}}^4$  sont toujours en sommeil à l'instant  $t_k$ , alors que la trajectoire  $t_{\mathcal{P}}^5$  quitte finalement son état sommeil au même instant pour mourir car elle a été endormie trop longtemps. La mort de la trajectoire est symbolisée par une croix.

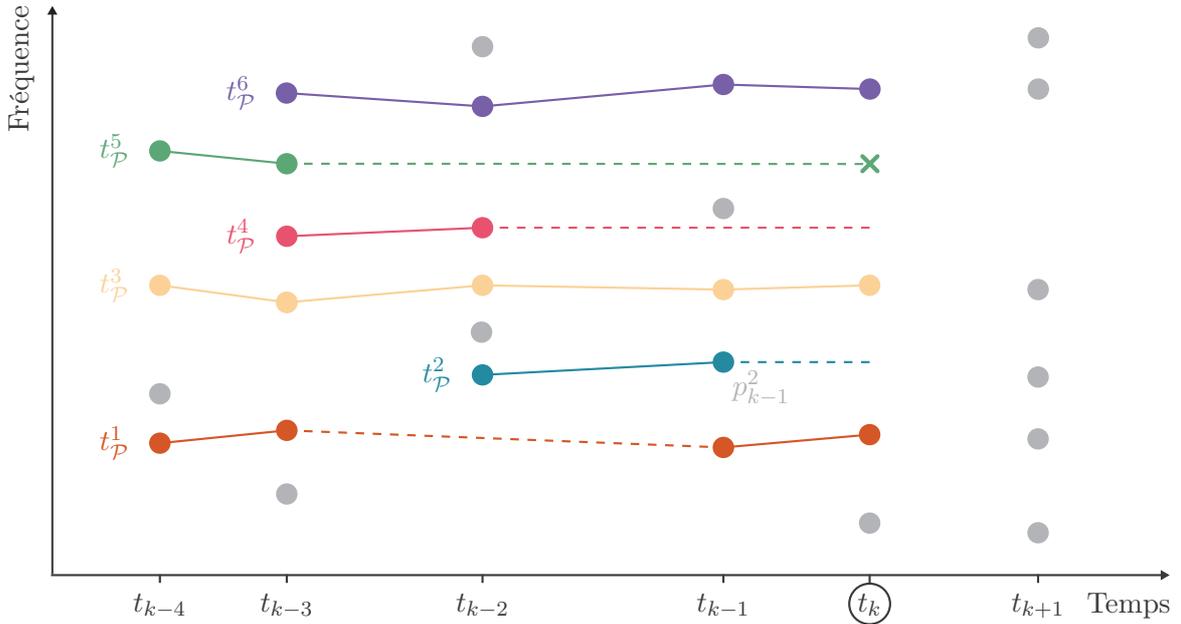


FIGURE 3.4 – Illustration du principe d'état sommeil d'une trajectoire de pics. Le suivi est réalisé jusqu'à l'instant  $t_k$ . Les pics sont représentés par des cercles, les trajectoires par des lignes en trait plein et l'état sommeil est symbolisé par des lignes en pointillés. La trajectoire  $t_{\mathcal{P}}^5$  est définitivement morte. Sa fin est représentée par une croix.

Pour réaliser le suivi d'une trajectoire en sommeil, nous nous servons du dernier pic de la trajectoire en conservant ses caractéristiques à l'identique. Pour illustrer ce point, prenons l'exemple de la trajectoire  $t_{\mathcal{P}}^2$  sur la figure 3.4. Lors de l'association à l'instant  $t_{k+1}$ , le dernier pic de la trajectoire  $t_{\mathcal{P}}^2$  (à savoir  $p_{k-1}^2$  de fréquence  $f_{k-1}^2$ ) est ajouté temporairement à l'ensemble  $\mathcal{P}_k$  à la position permettant de conserver l'ordre de l'ensemble (ordonné par fréquence croissante). Le suivi est alors réalisé comme expliqué dans la section 3.2.1 entre l'ensemble  $\mathcal{P}_k$  temporairement modifié et l'ensemble  $\mathcal{P}_{k+1}$ .

Le choix d'utiliser les caractéristiques du dernier pic de la trajectoire pour le suivi des trajectoires en sommeil permet à la méthode de rester aussi générique que possible. En effet, il aurait également été possible d'introduire un modèle pour la trajectoire et d'estimer pendant l'état sommeil la position du pic recherché. Cette démarche, en plus d'introduire des *a priori*, aurait également été plus coûteuse en termes de temps de calcul. De plus, le choix du modèle n'aurait pas été trivial ; le temps séparant deux signaux consécutifs dans la séquence  $\mathcal{S}$  n'étant pas fixe, nous aurions pu nous retrouver dans des cas où la différence de temps entre deux signaux est supérieure à la longueur de la trajectoire déjà suivie.

### 3.2.2.4 Suppression des trajectoires trop courtes

Dans la méthode originale proposée par McAulay et Quatieri, tous les pics appartiennent à une trajectoire. En effet, lors du suivi entre les instants  $t_k$  et  $t_{k+1}$ , tous les pics de l'instant

$t_{k+1}$  qui n'ont pas été associés démarrent une nouvelle trajectoire. Cette manière de faire engendre de nombreuses trajectoires, dont certaines ne contiennent qu'un unique pic. Ces dernières trajectoires ne sont pas porteuses d'information sur la globalité de la séquence de signaux, elles sont donc peu intéressantes. Pour éviter la création de trajectoires inutiles, nous proposons de ne faire naître une trajectoire que lorsque 2 pics n'appartenant pas encore à une trajectoire s'associent l'un à l'autre. Ainsi, chaque trajectoire contient au minimum 2 pics.

Comme expliqué dans la section 2.5.1 du chapitre 2, la densité des ensembles de pics est forte. La probabilité d'associer un pic de l'instant  $t_{k+1}$  à un pic isolé à l'instant  $t_k$  est suffisamment grande pour créer des trajectoires de deux pics ne reflétant pas réellement l'évolution du spectre des signaux. La probabilité que 3 pics forment, par pur hasard également, une trajectoire est quant à elle beaucoup plus faible. Ce propos a été vérifié empiriquement, en réalisant une simulation de Monte Carlo [Metropolis 1949]. 1000 séquences de 20 signaux ont été générées ; tous les signaux contiennent une unique sinusoïde de fréquence 20 Hz, à laquelle vient s'ajouter un bruit blanc Gaussien avec un RSB de 0 dB. Les signaux sont échantillonnés à 500 Hz et durent tous 20 secondes. Le nombre moyen de pics identifiés dans chaque signal est de 41 pics. À partir des données, l'intervalle de recherche pour le suivi a été calculé selon (3.11) et vaut  $\Delta f = 0.8$  Hz.

La figure 3.5 nous donne la répartition du nombre de trajectoires en fonction du nombre de pics contenus dans les trajectoires pour l'ensemble des 1000 séquences. Les 1000 trajectoires de 20 pics correspondent aux trajectoires attendues, celles suivant la fréquence de 20 Hz. Toutes les autres trajectoires sont donc des trajectoires identifiées au sein de pics représentant du bruit. Parmi elles, la proportion de trajectoires ne possédant que 2 pics est largement supérieure à celle des trajectoires contenant 3 pics ou plus. Ce constat nous amène à proposer un post-traitement afin d'éliminer un maximum de trajectoires inintéressantes, sans prendre le risque d'éliminer une trajectoire porteuse d'information. Le post-traitement consiste donc à éliminer les trajectoires trop courtes, c'est-à-dire celles ne comportant que 2 pics.

La suppression des trajectoires trop courtes est effectuée à chaque itération temporelle de la méthode de suivi. Ainsi, nous proposons de supprimer toutes les trajectoires formées par 2 pics uniquement, mais seulement si cette trajectoire est définitivement morte et n'est pas en état sommeil. En effet, si la trajectoire est encore en sommeil, il y a des chances qu'elle se réveille, s'agrandisse et qu'elle comporte à terme plus de 2 pics.

### 3.2.3 Paramétrisation de la méthode de suivi

Cette section synthétise le paramétrage de la méthode de suivi. Tous les paramètres à fixer ne dépendant pas des données sont récapitulés dans le tableau 3.1. Les valeurs par défaut ont été choisies de manière à adapter au mieux la méthode de suivi aux différentes applications mécaniques testées au cours de la thèse. Certaines de ces applications sont présentées dans la section 3.4.

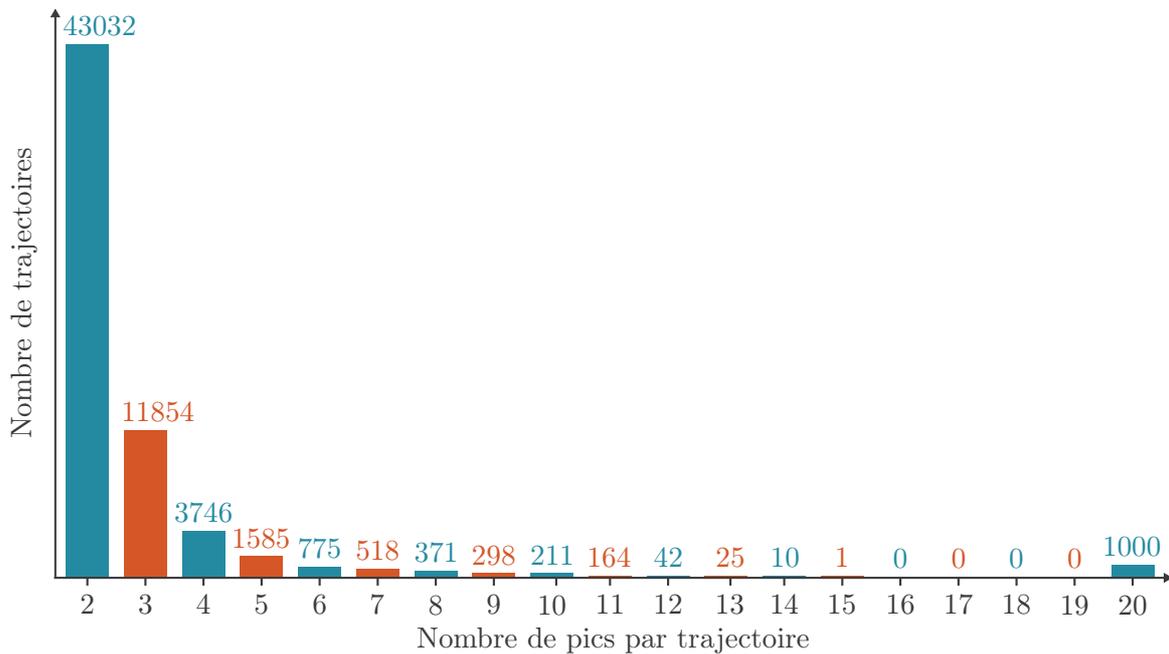


FIGURE 3.5 – Répartition du nombre de trajectoires en fonction du nombre de pics par trajectoire pour 1000 séquences de 20 signaux, contenant chacun une sinusoïde noyée dans un bruit blanc Gaussien.

Paramètre	Valeur par défaut
Facteur multiplicatif $k_{\Delta}$ de l'intervalle de recherche	10
Nombre maximal d'instantanés consécutifs	2
Nombre minimal de pic dans une trajectoire <sup>2</sup>	3

Tableau 3.1 – Liste des paramètres utilisés pour le suivi des structures spectrales.

### 3.3 Suivi de structures spectrales

Les modifications précédentes améliorent le suivi temps-fréquence des pics. Cependant, elles ne permettent pas encore de suivre les structures spectrales comme les harmoniques ou les modulations. Nous allons à présent voir comment adapter la méthode proposée au suivi des séries harmoniques dans un premier temps, puis au suivi des bandes de modulation dans un second temps.

2. Des trajectoires plus courtes existent mais elles sont en sursis. Tant qu'elles sont actives ou en sommeil, elles sont conservées. Cependant, si elles meurent sans s'être prolongées, elles sont alors supprimées à ce moment-là.

### 3.3.1 Suivi de séries harmoniques

La carte temps-fréquence synthétise toutes les informations spectrales de la séquence de signaux  $\mathcal{S}$ . En particulier, pour chaque signal  $s_k \in \mathcal{S}$ , la méthode d'identification des séries harmoniques, présentée au chapitre 2, nous donne l'ensemble  $\mathcal{H}_k$  des séries harmoniques identifiées parmi l'ensemble  $\mathcal{P}_k$  des pics détectés dans le spectre du signal. Chaque série harmonique  $h_k^i$  a été caractérisée en particulier par une fréquence fondamentale estimée. Les ensembles  $\mathcal{H}_k$  sont d'ailleurs ordonnés par fréquence fondamentale croissante. Cette analogie entre les ensembles  $\mathcal{H}_k$  et les ensembles  $\mathcal{P}_k$  n'est pas anodine et permet de réutiliser l'algorithme de suivi des pics décrit dans la section 3.2.1 directement sur les séries harmoniques en l'appliquant non plus sur la fréquence  $f_k^i$  des pics  $p_k^i$ , mais sur la fréquence fondamentale  $\nu_k^i$  des séries  $h_k^i$ . Les séries harmoniques suivies forment alors des trajectoires de séries harmoniques que nous noterons  $t_{\mathcal{H}}^j$ , où  $j$  représente l'indice de la trajectoire dans l'ensemble  $\mathcal{T}_{\mathcal{H}}$  rassemblant toutes les trajectoires harmoniques de la séquence  $\mathcal{S}$ .

Le suivi des séries harmoniques permet de réaliser un suivi structurel du contenu spectral de la séquence. En effet, chaque trajectoire  $t_{\mathcal{H}}^j$  identifiée est une liste de séries harmoniques. Or chaque série harmonique est elle-même une liste de pics classés par rang dans la série. Il est donc possible de créer des trajectoires de pics simplement à l'aide du rang des pics au sein des séries contenues dans une même trajectoire  $t_{\mathcal{H}}^j$ . Ce principe de suivi structurel est illustré sur la figure 3.6, en même temps que les 3 étapes du suivi de séries harmoniques :

1. Passage de la carte temps-fréquence, où l'ensemble du contenu spectral de la séquence  $\mathcal{S}$  est représenté, à une *carte temps-fréquence fondamentale* où seules les fréquences fondamentales  $\nu_k^i$  des séries harmoniques  $h_k^i$  sont représentées.
2. Suivi des fréquences fondamentales pour former les trajectoires harmoniques  $t_{\mathcal{H}}^j$ . Le suivi est réalisé grâce à la méthode décrite dans la section 3.2, appliquée sur les fréquences fondamentales des séries.
3. Retour à la carte temps-fréquence, sur laquelle apparaissent les trajectoires de pics générées directement à partir du rang des pics dans les séries harmoniques des trajectoires de séries harmoniques. Si le pic de rang  $r$  n'est pas présent à tous les instants de la trajectoire  $t_{\mathcal{H}}^j$ , la trajectoire de pics qui découle du rang  $r$  sera mise en sommeil durant les moments d'absence. Différents cas de figure sont illustrés par les trajectoires  $t_{\mathcal{P}}^1$ ,  $t_{\mathcal{P}}^5$  et  $t_{\mathcal{P}}^1$  sur la figure 3.6 (d). Cette dernière étape de suivi structurel des pics est facultative.

L'exemple de suivi de la figure 3.6 est volontairement simple afin de bien illustrer les différentes étapes du suivi. Cependant, il faut garder à l'esprit que la méthodologie proposée permet d'identifier un grand nombre de séries harmoniques, sans *a priori* sur les séries ou sur leur nombre. De plus, cette approche structurelle du suivi est robuste. En effet, le suivi des séries harmoniques est possible, même si des pics sont absents dans les séries et même si le fondamental est absent.

Le suivi structurel permet de générer, à partir de la trajectoire harmonique, les trajectoires des pics inclus dans les séries suivies (en fonction de leur rang). Réaliser à la fois le suivi des pics, puis le suivi des séries harmoniques va donc générer de la redondance au niveau des trajectoires de pics. Ce point sera discuté dans la section 3.3.3, après la présentation du suivi

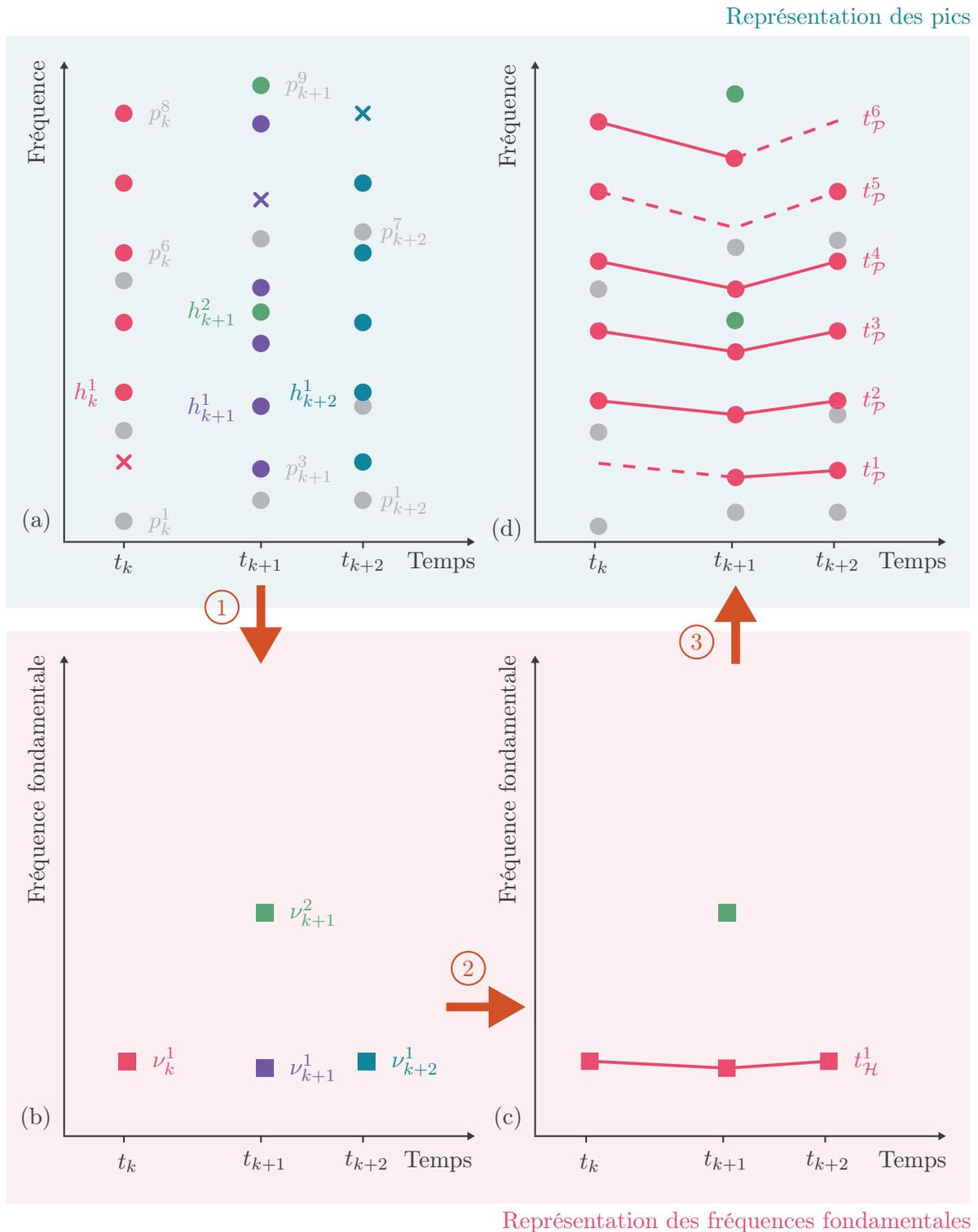


FIGURE 3.6 – Illustration des étapes du suivi de séries harmoniques. Les cartes (a) et (d) sont des cartes temps-fréquence où les disques représentent des pics et les croix l’absence d’un pic à un rang donné dans une série harmonique, alors que les cartes (b) et (c) sont des cartes temps-fréquence fondamentale. Le changement de domaine (1) permet de réaliser le suivi des fréquences fondamentales (2), avant de revenir dans le domaine initial (3) pour visualiser les effets du suivi structurel.

des bandes de modulation.

### 3.3.2 Suivi de bandes de modulation

Le suivi des bandes de modulation semble plus complexe à appréhender que le suivi des pics ou des séries harmoniques. En effet, contrairement aux structures précédentes, une bande de modulation  $m_k^i$  ne possède pas une fréquence caractéristique, mais deux : la fréquence porteuse  $f_p$  et la fréquence de modulation  $f_{m,k}^i$ . Plutôt que d’imaginer une méthode de suivi prenant conjointement ces deux paramètres en compte, nous pouvons nous servir du constat suivant : suivre une fréquence porteuse au cours de la séquence de signaux  $\mathcal{S}$  nous donnera une trajectoire de pics. Cela signifie que si les trajectoires de pics sont d’ores et déjà identifiées (soit par le suivi des pics, soit grâce au suivi structurel des séries harmoniques), les fréquences porteuses sont elles aussi déjà suivies. Il devient alors possible de réutiliser la méthode de suivi présentée dans la section 3.2 pour suivre les modulations autour d’une trajectoire de pics, en utilisant la fréquence de modulation  $f_{m,k}^i$  des bandes de modulation  $m_k^i$  à la place de la fréquence  $f_k^i$  des pics  $p_k^i$ .

La méthode de suivi des bandes de modulation est donc une méthode itérative. Une itération de la méthode est illustrée sur la figure 3.7, avec un exemple volontairement simple. Les itérations se font sur l’ensemble  $\mathcal{T}_{\mathcal{P}}$  des trajectoires de pics, qu’il faut préalablement avoir identifié dans la séquence de signaux  $\mathcal{S}$ . Pour chaque trajectoire de pics, il faut identifier les bandes de modulation ayant pour pic porteur un pic de la trajectoire. Sur l’exemple de la figure 3.7, la trajectoire de pics que nous considérons est  $t_{\mathcal{P}}^1$ . Elle contient les pics  $p_k^8$ ,  $p_{k+1}^5$  et  $p_{k+2}^7$ , qui sont porteurs de 4 bandes de modulation  $m_k^1$ ,  $m_k^2$ ,  $m_{k+1}^3$  et  $m_{k+2}^4$ . Les fréquences de ces modulations sont affichées sur une *carte temps-fréquence de modulation* représentée figure 3.7 (b). Elles sont ensuite suivies pour former des trajectoires de modulations  $t_{\mathcal{M}}^j$ , grâce à l’algorithme de suivi de la section 3.2 (figure 3.7 (c)). De la même manière que pour les séries harmoniques, les trajectoires de modulations permettent un suivi structurel du contenu spectral. Il est donc possible de créer des trajectoires de pics à partir du rang des pics au sein des bandes de modulation. Les trajectoires de pics créées peuvent être partiellement endormies lorsqu’un rang particulier n’est pas présent à un instant donné. Cette dernière étape de suivi structurel, représentée figure 3.7 (d), est elle aussi facultative.

### 3.3.3 Stratégie de suivi

Nous venons de voir comment suivre les trois types de structures spectrales que nous identifions dans le spectre d’un signal, à savoir les pics spectraux, les séries harmoniques et les bandes de modulation. Il est donc possible d’identifier les ensembles de trajectoires  $\mathcal{T}_{\mathcal{P}}$ ,  $\mathcal{T}_{\mathcal{H}}$  et  $\mathcal{T}_{\mathcal{M}}$  en utilisant les trois méthodes successivement. Cependant, nous avons vu que le suivi des harmoniques et des modulations permet également, de par leur structure, de suivre certains pics du spectre. Cela signifie qu’appliquer successivement les trois méthodes de suivi pour chacune des structures entraîne une certaine redondance dans les résultats. Autrement dit, les performances du suivi des trois types de structures ne sont pas optimales et du temps

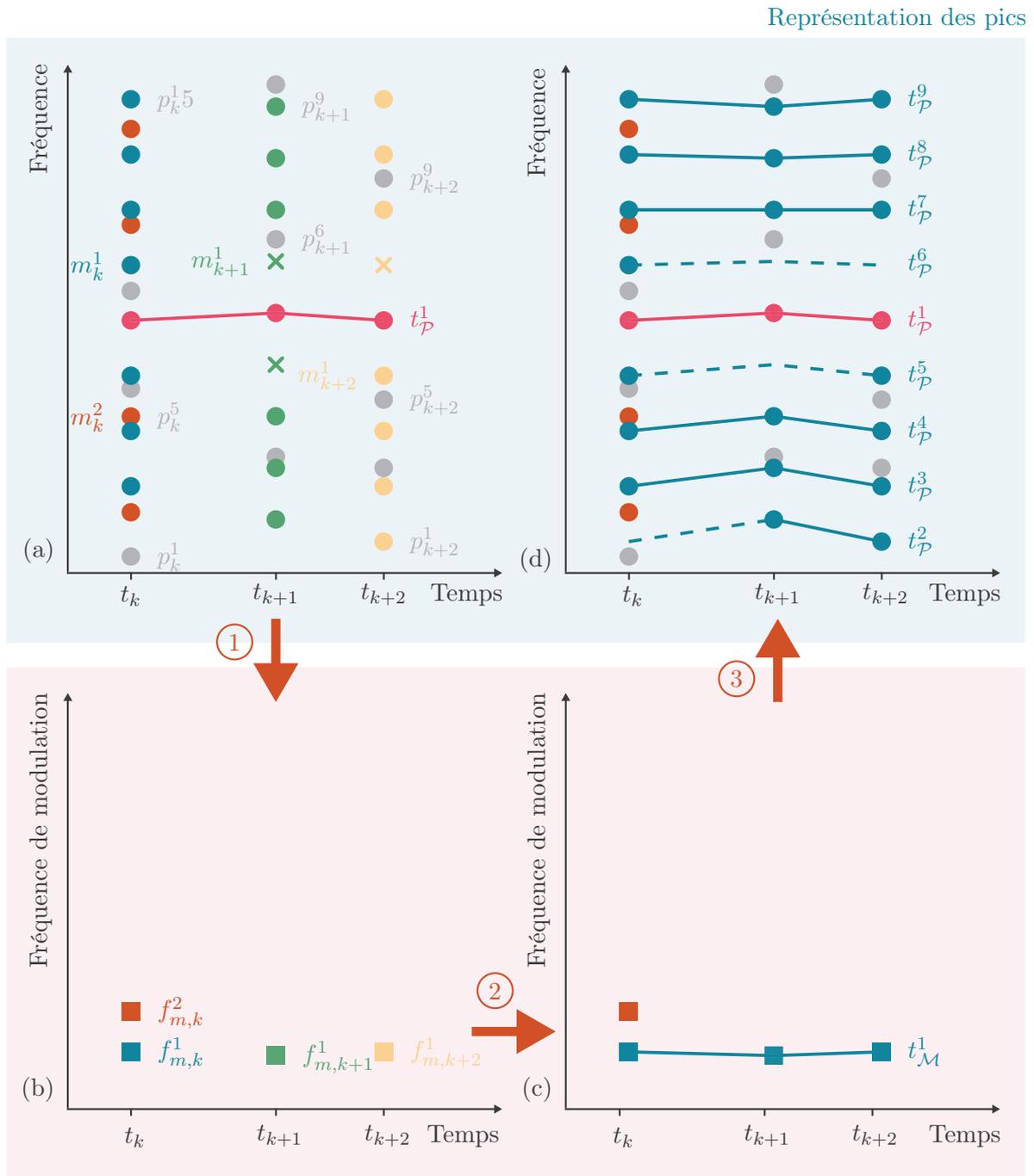


FIGURE 3.7 – Illustration des étapes du suivi de bandes de modulation. Les cartes (a) et (d) sont des cartes temps-fréquence où les disques représentent des pics et les croix l'absence d'un pic à un rang donné dans une bande de modulation, alors que les cartes (b) et (c) sont des cartes temps-fréquence de modulation. Le changement de domaine (1) permet de réaliser le suivi des fréquences de modulation (2), avant de revenir dans le domaine initial (3) pour visualiser les effets du suivi structurel.

de calcul est gaspillé à identifier des trajectoires déjà disponibles. Cette perte de performance va à l'encontre du cahier des charges fixé au cours de cette thèse.

Pour limiter la redondance des résultats, nous proposons une stratégie de suivi. Elle est basée, entre autres, sur la paramétrisation choisie de la méthode d'identification des bandes de modulation, discutée dans la section 2.5.1. En effet, pour des raisons de temps de calcul encore une fois, les bandes de modulation ne sont recherchées qu'autour des pics appartenant au moins à une série harmonique. Autrement dit, les fréquences porteuses des modulations des ensembles  $\mathcal{M}_k$  appartiennent forcément à au moins une série harmonique. Elles ne peuvent pas appartenir à des pics isolés dans le spectre. Ce constat permet d'organiser la stratégie de suivi, en utilisant l'approche structurelle du suivi des harmoniques comme base au suivi des modulations.

Nous proposons donc, dans l'ordre, de suivre :

1. les séries harmoniques grâce à leur fréquence fondamentale. La structure des séries permet de générer facilement les trajectoires de pics appartenant aux séries suivies ;
2. les bandes de modulation, le long des trajectoires de pics, à l'aide de leur fréquence de modulation. De même, la structure des modulations permet de générer les trajectoires des pics leur appartenant ;
3. les pics restants, à savoir les pics qui n'ont pas été automatiquement suivis lors du suivi des séries harmoniques ou du suivi des bandes de modulation.

### 3.3.4 Complexité calculatoire de la méthode de suivi proposée

La méthode de suivi a été choisie pour sa simplicité, reflet d'une faible complexité algorithmique et donc de temps de calcul faibles. En effet, la méthode présentée dans la section 3.2.1 peut être implémentée de manière très efficace, notamment grâce au fait que les ensembles à associer sont déjà triés par fréquence croissante. L'itération courante de l'algorithme peut alors se baser sur les indices de l'itération précédente.

Par exemple, si le pic  $p_k^i$  a été associé au pic  $p_{k+1}^j$ , c'est que l'association du pic  $p_k^{i+1}$  ne pouvait pas se faire avec  $p_{k+1}^j$ . Cette possibilité a été testée pour vérifier la condition de compatibilité inverse de l'association précédente. Il ne reste alors au pic  $p_k^{i+1}$  que la possibilité de s'associer avec un pic dont la fréquence est supérieure à celle de  $p_{k+1}^j$ , c'est-à-dire avec un pic du sous-ensemble  $\{p_{k+1}^{j+1}, \dots, p_{k+1}^{\#\mathcal{P}_{k+1}}\} \subset \mathcal{P}_{k+1}$ . Il n'est cependant pas nécessaire de tester tous les pics dans ce sous-ensemble. Si la fréquence du pic  $p_{k+1}^{j+2}$  n'est pas dans l'intervalle de recherche, autrement dit si  $f_{k+1}^{j+2} > f_k^{i+1} + \frac{\Delta f}{2}$ , alors tous les pics suivants ( $p_{k+1}^{j+3}, p_{k+1}^{j+4}, \dots$ ) seront également en dehors de l'intervalle de recherche.

La progression dans chacune des listes  $\mathcal{P}_k$  et  $\mathcal{P}_{k+1}$  est croissante. Si l'on atteint le dernier élément dans l'une des deux listes, la recherche d'associations à réaliser est alors terminée. Il est donc possible qu'un des deux ensembles ne soit parcouru entièrement. Pour chaque élément des ensembles, le nombre de comparaisons de fréquences à réaliser est faible, de l'ordre de une

à trois comparaisons par élément. Cela signifie que la complexité de l'association des ensembles  $\mathcal{P}_k$  et  $\mathcal{P}_{k+1}$  est de l'ordre de  $\mathcal{O}(\#\mathcal{P}_k + \#\mathcal{P}_{k+1})$ . En faisant l'hypothèse que le nombre de pics dans tous les signaux d'une même séquence est à peu près identique, nous avons finalement un nombre de comparaisons de fréquence de l'ordre de  $\mathcal{O}(\#\mathcal{P})$ , où  $\#\mathcal{P}$  correspond au nombre moyen de pics détectés dans un signal. Finalement, la complexité moyenne pour le suivi des pics dans une séquence de  $n$  signaux est de l'ordre de

$$\mathcal{O}(n\#\mathcal{P}). \quad (3.12)$$

Le nombre de comparaisons de fréquences dépend donc linéairement du nombre de signaux dans la séquence et du nombre de pics détectés dans chaque signal. Cette linéarité face à la quantité de données nous permet d'utiliser cet algorithme sur de longues séquences, sans souci de performances.

La méthode de suivi pour les séries harmoniques et les modulations étant identique à celle des pics, la complexité est elle aussi identique. La complexité globale de la stratégie de suivi proposée est de l'ordre de

$$\mathcal{O}(n(\#\mathcal{H} + \#\mathcal{M} + \#\mathcal{P}_{\text{restants}})), \quad (3.13)$$

où  $\#\mathcal{H}$  est le nombre moyen de séries harmoniques dans les signaux de la séquence,  $\#\mathcal{M}$  le nombre moyen de bandes de modulation et  $\#\mathcal{P}_{\text{restants}}$  le nombre moyen de pics restants après le suivi structurel des pics dans les séries harmoniques et dans les modulations.

## 3.4 Applications

Afin de valider la stratégie proposée, nous allons appliquer le suivi des structures spectrales sur des signaux simulés et des signaux réels issus de la surveillance de système. En particulier, nous vérifierons que nous arrivons bien à suivre les éléments du spectre correspondant à la présence et l'évolution d'un défaut, tout en respectant les contraintes fixées dans cette thèse.

### 3.4.1 Simulation d'un défaut

Le premier exemple d'application de la méthodologie de suivi est réalisé sur une séquence  $\mathcal{S}$  de 100 signaux artificiels. L'un des signaux de cette séquence a déjà été utilisé dans la section 2.5.2 pour tester la recherche d'harmoniques et de bandes de modulation.

#### 3.4.1.1 Génération de la séquence de signaux

La séquence de signaux  $\mathcal{S}$  simule la dégradation de l'engrènement de deux engrenages. L'engrènement prend la forme d'une série harmonique dans le spectre des signaux. Le fondamental de cette série dépend du nombre de dents des engrenages et de la vitesse de rotation

en entrée du système. Pour les 100 signaux de la séquence, la fréquence de la série a été fixée à 250 Hz et compte 4 harmoniques (fondamental inclus). Cette série harmonique est donc très stable, seule l'amplitude des pics présents dans la série augmente avec le temps. Afin de complexifier l'expérience, le fondamental de la série harmonique a été artificiellement retiré des signaux  $s_{89}$  et  $s_{93}$ .

La présence d'un défaut sur l'un des engrenages a été modélisée par l'apparition de modulations autour des 4 pics de la série d'engrènement. L'apparition des modulations est progressive : tout d'abord, une bande apparaît autour du fondamental de la série d'engrènement. L'énergie de cette bande augmente au fur et à mesure des signaux et de nouvelles bandes apparaissent pour simuler l'évolution de la dégradation. En parallèle, une bande commence à apparaître autour de l'harmonique de rang 2 de la série d'engrènement. Cette seconde modulation va également gagner en énergie avec le temps et voir naître de nouvelles bandes. Il en va de même pour les harmoniques de rang 3 et 4, qui, successivement, vont voir apparaître une bande de modulation qui grandira avec le temps. La fréquence de ces modulations a été fixée à 10 Hz. Le nombre de bandes et l'amplitude des pics dans les bandes augmentent au fur et à mesure des 100 signaux de la séquence.

Afin de nous rapprocher des conditions réelles, un bruit blanc Gaussien avec un RSB de 0 dB a été ajouté à chaque signal, échantillonné à 10 kHz sur une durée de 1 seconde. Il n'est pas correct de parler de dates d'acquisition  $t_k$  des signaux alors qu'ils sont générés artificiellement. Cependant, les dates associées à ces signaux ont été choisies de manière à être régulières :

$$\forall k \in \{2, \dots, 99\}, t_k - t_{k-1} = t_{k+1} - t_k. \quad (3.14)$$

Le temps entre deux signaux a été fixé à 1 jour<sup>2</sup>.

L'objectif est donc de réussir à établir les 5 trajectoires importantes dans la séquence : la trajectoire harmonique suivant la série d'engrènement et les 4 trajectoires de modulations suivant la modulation à 10 Hz autour des 4 pics de la série d'engrènement. Les signaux passent alors par toutes les étapes de traitement : validation des signaux, détection des pics et identification des séries harmoniques et de bandes de modulation. Ces étapes servent à établir la carte temps-fréquence de la séquence, qui servira d'entrée à la stratégie de suivi.

### 3.4.1.2 Suivi de la séquence simulée

La visualisation de la carte temps-fréquence complète de la séquence n'est malheureusement pas envisageable tant le nombre d'informations à afficher est important. Cependant, nous proposons deux zooms sur cette carte (figure 3.9), qui sont complétés par les courbes de la figure 3.8 affichant le nombre de pics, de série harmoniques et de bandes de modulation identifiés pour chaque signal de la séquence. Le nombre de pics a tendance à augmenter avec

---

2. Ce choix de régularité n'a aucune influence sur le suivi. En effet, les dates des signaux n'entrent pas en compte dans l'algorithme de suivi. Les résultats auraient donc été identiques avec des dates non régulières. Le choix de régularité pour cette séquence de signaux simulés est donc uniquement un choix pratique, car toutes les dates disposent du même espace sur la carte temps-fréquence, ce qui la rend plus facile à lire.

le temps, ce qui était attendu car le nombre de pics dans les modulations augmente au fur et à mesure des signaux. La variance sur le nombre de pics détectés est le reflet de fausses détections dues aux seuils de détection volontairement bas pour être certain d'extraire toute l'information spectrale des signaux (voir annexe B.2.4). La série harmonique que nous souhaitons suivre est détectée sur les 100 signaux. Cependant, d'autres séries sont également détectées, certaines parmi les pics correspondant au bruit, d'autres parmi les bandes de modulation comme expliqué dans la section 2.5.2. Le nombre de bandes de modulation est de 0 au départ car aucune modulation n'est effectivement présente. Il devrait être de 4 à la fin, une fois que les 4 modulations d'intérêt sont bien formées. Les 4 modulations que nous souhaitons suivre sont bien détectées dès que possible, cependant le nombre de bandes détectées est plus grand à cause de la recherche de modulations autour des harmoniques des séries supplémentaires (détectées entre autres aux fréquences 240 et 260 Hz ; voir section 2.5.2 pour plus d'informations).

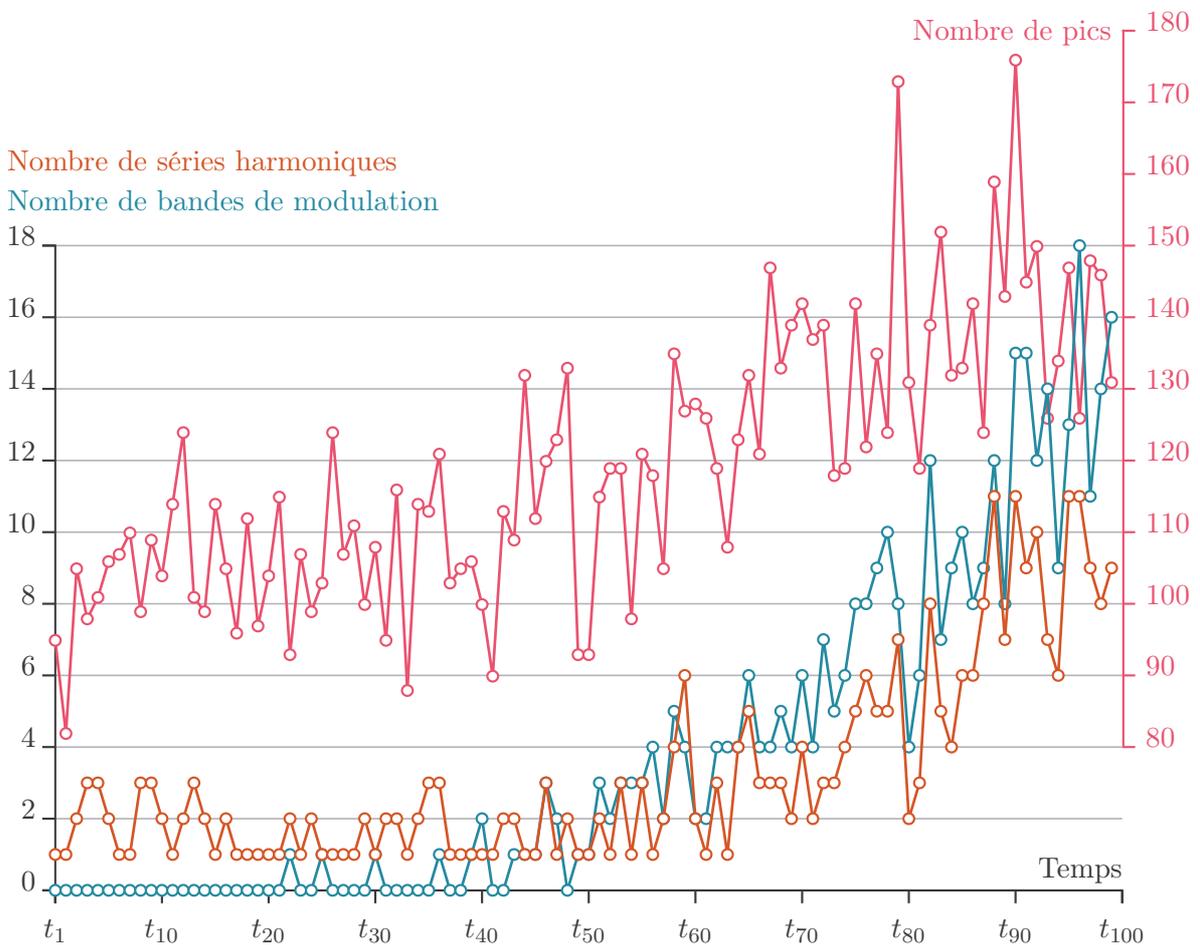


FIGURE 3.8 – Nombre de structures spectrales à suivre dans la séquence artificielle de signaux. Les séries harmoniques et les modulations partagent la même échelle à gauche, alors que le nombre de pics est indiqué sur l'échelle à droite.

L'application de la stratégie de suivi sur l'ensemble des structures spectrales identifiées

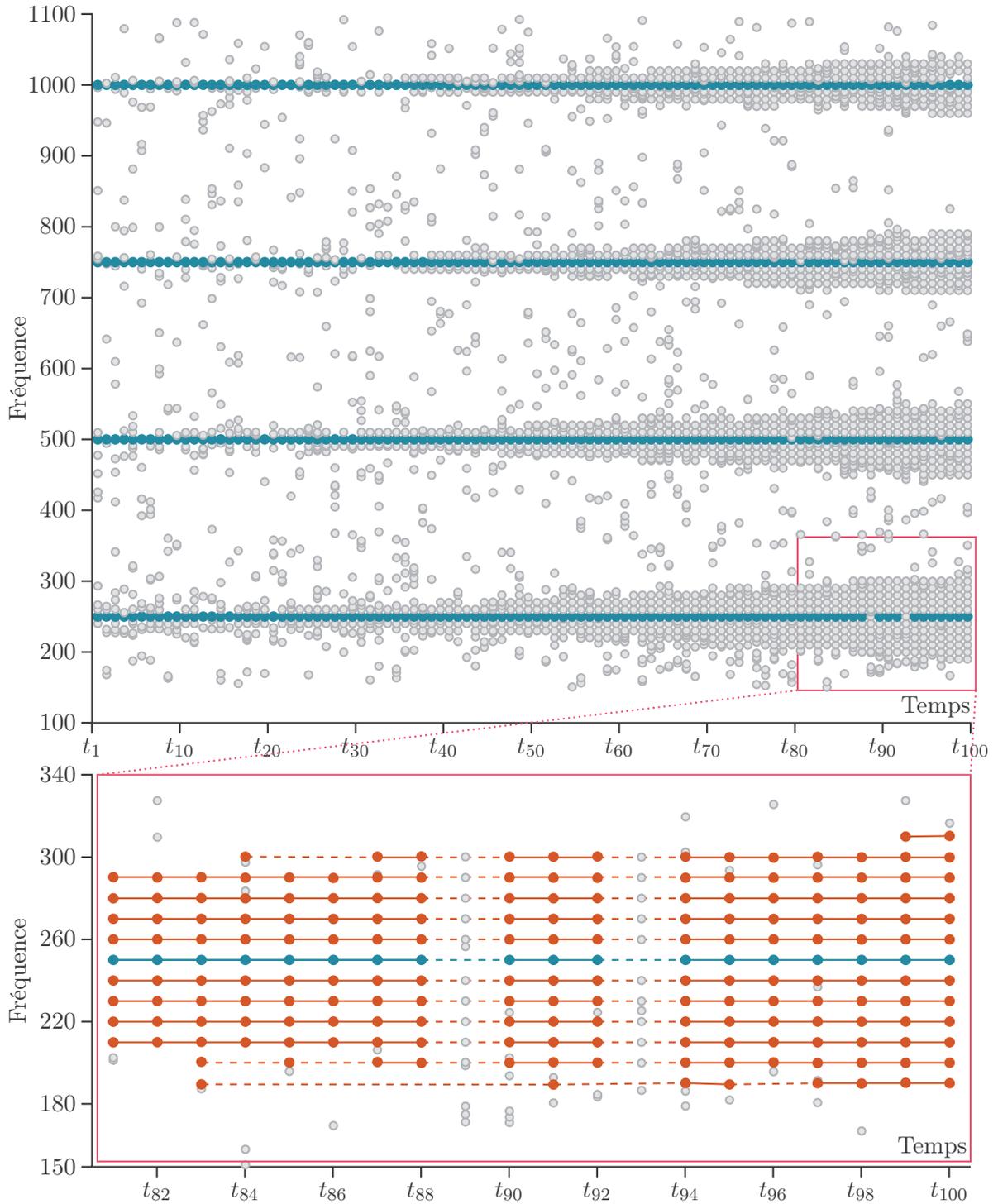


FIGURE 3.9 – Zooms sur la carte temps-fréquence de la séquence du défaut simulé. En haut, un zoom large (100 à 1100 Hz) faisant apparaître la trajectoire de la série d’engrènement à 250 Hz (en bleu). En bas, un zoom plus serré en temps ( $t_{81}$  à  $t_{100}$ ) et en fréquence (150 à 340 Hz), mettant en avant le suivi structurel des modulations (en orange).

dans la séquence nous donne au total 11 trajectoires de séries harmoniques, 19 trajectoires de modulations et 42 trajectoires de pics. Les 5 trajectoires que nous souhaitons suivre font partie de celles identifiées. Nous obtenons les résultats escomptés sous la forme de 5 trajectoires complètes, c'est-à-dire que chaque trajectoire comprend toutes les structures spectrales qu'elle devait contenir.

La figure 3.9 présente deux zooms sur la carte temps-fréquence de la séquence du défaut d'engrènement simulé. Le premier, en haut, est assez large afin de visualiser la trajectoire de la série d'engrènement dans sa globalité. Il est à noter que l'absence du fondamental aux instants  $t_{89}$  et  $t_{93}$  n'a pas gêné l'établissement de la trajectoire. On distingue bien, même si leurs trajectoires ne sont pas marquées par soucis de lisibilité, les modulations autour des 4 harmoniques. Les modulations prennent la forme de rangées de pics, régulièrement espacées autour des harmoniques de la série d'engrènement. Le nombre de bandes dans les modulations augmente avec le temps. Le second zoom, au bas de la figure 3.9, est plus serré à la fois en temps et en fréquence et permet d'observer le suivi structurel des pics suite à l'établissement de la trajectoire harmonique et de la trajectoire de modulations. Les cercles bleus représentent les pics dans la trajectoire du fondamental de la série d'engrènement. L'absence du fondamental aux instants  $t_{89}$  et  $t_{93}$  se traduit par une mise en sommeil de la trajectoire, symbolisée par un trait en pointillés. Cette absence n'a pas gêné le suivi de la série harmonique. Cependant, elle se répercute sur le suivi des modulations (en orange). En effet, la méthode d'identification des bandes de modulation présentée dans le chapitre 2 a besoin d'un pic porteur pour démarrer la recherche des modulations. Lorsque le pic est absent, les modulations ne sont pas détectées et l'état sommeil entre alors en jeu pour assurer malgré tout la continuité de la trajectoire de modulations.

### 3.4.2 Surveillance d'un système mécanique

Le second exemple d'application de la stratégie de suivi est réalisé sur les signaux vibratoires issus de la surveillance d'un système mécanique. Il s'agit du même système que dans la section 2.5.3, à savoir un banc de test simulant le fonctionnement d'une éolienne. Le descriptif détaillé du banc et de la séquence de signaux issus du banc sont donnés dans l'annexe C.

Le test consiste à dégrader le roulement principal du banc d'essai afin d'identifier le moment où un défaut va apparaître et de visualiser son évolution. Cette dégradation est forcée grâce à un module permettant d'appliquer des charges axiales et radiales sur le roulement, durant des plages de fonctionnement non stationnaires inspirées par des profils de vent. Le roulement principal est surveillé par 3 accéléromètres. Les résultats présentés dans la suite sont identiques sur les 3 accéléromètres, c'est pourquoi nous nous contenterons de les présenter pour le capteur dans la direction  $+y$ . Le test a été conduit jusqu'à ce que le système ne soit plus en état de fonctionner suite à la dégradation trop importante du roulement principal. À la fin du test, le roulement a été démonté et inspecté par des experts en mécanique.

Les signaux vibratoires sont acquis pendant une plage stationnaire du fonctionnement du banc. Les conditions de fonctionnement sont les mêmes entre toutes les acquisitions :

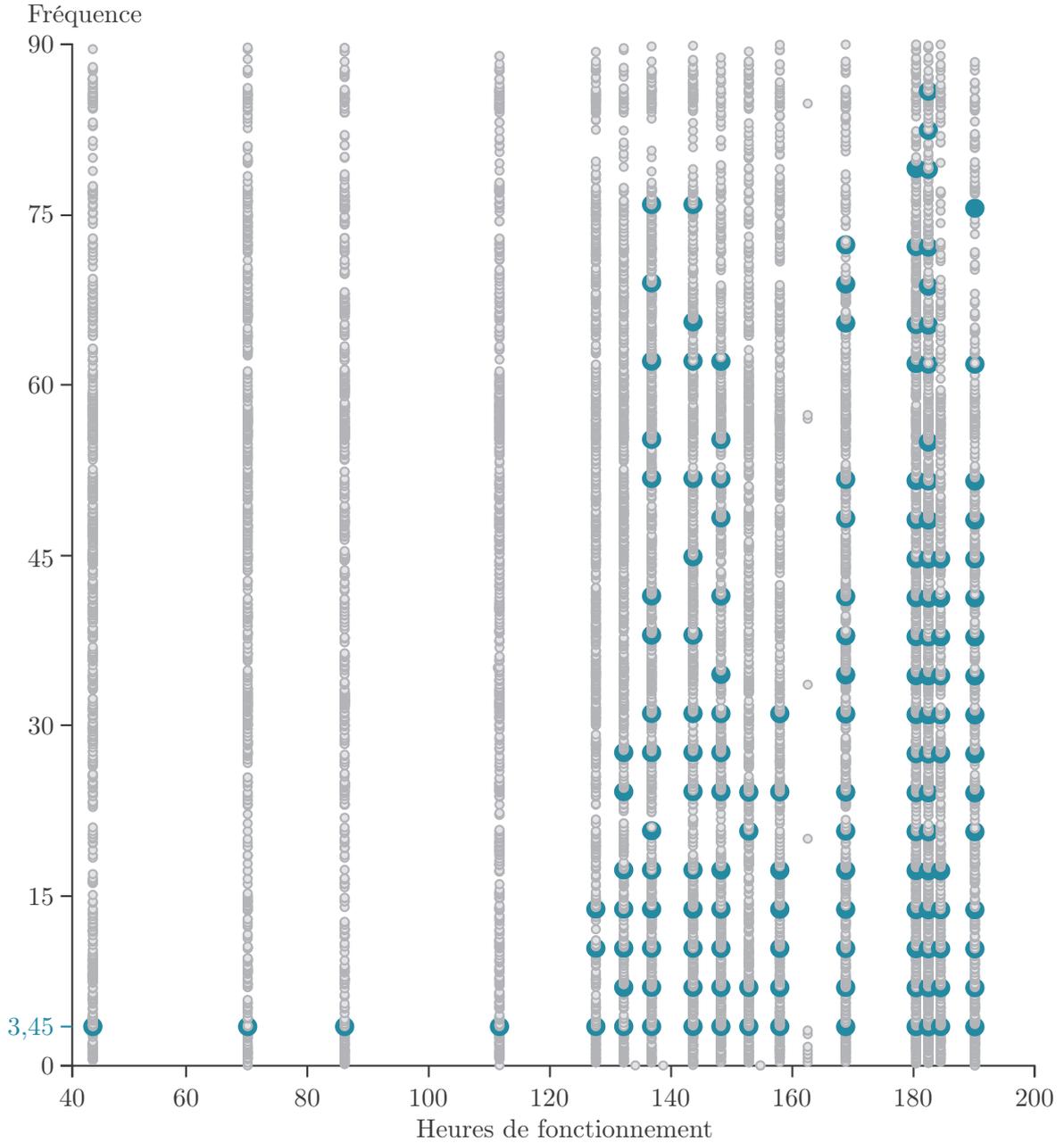


FIGURE 3.10 – Zoom sur la carte temps-fréquence des signaux réels enregistrant la dégradation du roulement principal. La trajectoire de pics à 3,45 Hz se transforme en trajectoire harmonique au bout de 129 heures de fonctionnement. Il s’agit là du premier symptôme de l’écaillage de la bague externe.



FIGURE 3.11 – Photo de la bague externe du roulement principal à la fin du test d'usure. Un écaillage est présent sur toute la circonférence de la bague.

l'hypothèse d'état machine constant est donc vérifiée, ce qui permet de réaliser les opérations de suivi. Chaque signal est échantillonné à 39 062,5 Hz et dure 150 secondes, ce qui permet une résolution spectrale très fine. Les 17 signaux sont acquis à différents moments de la vie du roulement, que nous compterons en heures de fonctionnement. L'analyse individuelle des signaux nous donne en moyenne un ensemble de 19 000 pics, 600 séries harmoniques et 12 000 bandes de modulation par signal. Toutes ces structures se transforment, après l'application de la stratégie de suivi, en 828 trajectoires harmoniques et 9 373 trajectoires de modulations, auxquelles viennent s'ajouter 12 406 trajectoires de pics identifiées parmi les pics restants.

Parmi l'ensemble des trajectoires suivies, trois d'entre elles méritent une attention toute particulière. Elles sont présentes sur la figure 3.10. Dès les premières heures de fonctionnement, une trajectoire de pics à la fréquence de 3,45 Hz est suivie. Cette trajectoire évolue pour se transformer en trajectoire harmonique au bout de 129 heures de fonctionnement environ. Le nombre d'harmoniques évolue alors aux fils des heures de fonctionnement. La troisième trajectoire intéressante est une trajectoire de modulations. Elle n'a pas été représentée sur la figure 3.10 pour des raisons de visibilité. Cependant, la modulation de 0,333 Hz apparaît autour du fondamental à 3,45 Hz au bout de 134 heures de fonctionnement. Tout comme la série harmonique, le nombre de bandes augmente tout au long de la trajectoire.

Les fréquences 3,45 Hz et 0,333 Hz correspondent à la fréquence caractéristique de la bague externe du roulement principal et à la vitesse de rotation de l'arbre dans le roulement, respectivement. La transformation de la trajectoire de pics en trajectoire harmonique et l'apparition de la trajectoire de modulations sont donc les premiers symptômes visibles dans le spectre des signaux vibratoires d'un défaut naissant sur la bague externe du roulement. Le démantèlement du roulement à la fin de l'expérience et son analyse par des experts en mécanique confirment le défaut sur la bague externe. La figure 3.11 montre l'état de cette bague et son usure prononcée à la fin du test. Cette expérience confirme la capacité de la méthodologie proposée jusqu'ici à donner des informations sur l'état de santé du système.

Il est également possible de voir sur la figure 3.10 qu'autour de 163 heures de fonctionnement, la série harmonique n'est pas détectée, et par conséquent, n'est pas non plus intégrée à la trajectoire. Ce signal est en réalité hautement non stationnaire à cause d'une impulsion forte dans les vibrations (45 g) certainement due à un détachement de matière sur la bague externe du roulement. La non-stationnarité empêche alors le bon déroulement de l'analyse spectrale. Cependant, même si la série n'est pas détectée pour ce signal-là, la trajectoire n'est pas stoppée grâce à l'état sommeil permettant d'assurer la continuité du suivi.

La détection des trajectoires permet un premier diagnostic du système. Nous verrons, dans le chapitre 4, comment utiliser ces trajectoires pour visualiser l'évolution de l'état de santé du système. Ce prochain chapitre sera également l'occasion de présenter de nouvelles applications de la méthodologie proposée sur signaux réels.

### 3.5 Synthèse du chapitre

Au cours de ce chapitre, nous avons répondu à la problématique du suivi des structures spectrales au sein d'une séquence de signaux. La stratégie proposée devait répondre à plusieurs contraintes. Elle devait être le plus générique possible, tout en ayant de très bonnes performances en termes de temps de calcul. Elle devait également être capable de suivre trois types de structures, à savoir les pics spectraux, les séries harmoniques et les bandes de modulation.

La méthode de suivi se base sur les travaux de McAulay et Quatieri, qui proposaient une méthode de suivi des pics dans les signaux de parole. La méthode a été adaptée et améliorée afin de la rendre compatible avec notre contexte de surveillance. Elle a ensuite été utilisée astucieusement afin de suivre les séries harmoniques et les bandes de modulation. Enfin, une stratégie pour le suivi permet de limiter la redondance des opérations, tout en garantissant la détection de l'ensemble  $\mathcal{T}_P$  des trajectoires de pics, de l'ensemble  $\mathcal{T}_H$  des trajectoires harmoniques et de l'ensemble  $\mathcal{T}_M$  des trajectoires de modulations.

Tout comme les travaux du chapitre 2, les travaux présentés au cours de ce chapitre ont fait l'objet d'un transfert industriel auprès d'un partenaire des projets européens [*KAStrion*] et [*SUPREME*]. Cette méthode est donc intégrée au sein d'un CMS, sous la forme d'un module nommé AStrion-T (voir annexe B.2.8). De plus, ces résultats ont été publiés dans une revue internationale [*Gerber 2015*]

Les trajectoires permettent de suivre le contenu spectral de signal en signal. Leurs apparitions sont de bons indicateurs signalant un changement de l'état du système. Cependant, la simple détection des trajectoires n'est pas suffisante pour diagnostiquer le système. Il est nécessaire de transformer ces trajectoires en tendances permettant alors de visualiser l'évolution de la santé du système au cours du temps. C'est l'utilisation de ces tendances qui permettra d'arriver au diagnostic automatique.

# Diagnostic du système surveillé

---

## Sommaire

---

<b>4.1</b>	<b>Génération de tendances</b>	<b>82</b>
4.1.1	Génération à partir de la carte d'identité spectrale	82
4.1.2	Applications	83
4.1.3	Conclusions	92
<b>4.2</b>	<b>Détection de l'évolution d'un système</b>	<b>94</b>
4.2.1	Contexte	94
4.2.2	Description des tests statistiques	96
4.2.3	Utilisation des tests statistiques	103
4.2.4	Combinaison des tests	106
<b>4.3</b>	<b>Synthèse du chapitre</b>	<b>109</b>

---

La méthodologie proposée jusque-là permet de caractériser le contenu spectral des signaux d'une séquence. Ce contenu est alors suivi au cours du temps pour former des trajectoires spectrales. L'apparition d'une nouvelle trajectoire est le reflet d'un changement dans le spectre des signaux et peut être interprétée comme le premier symptôme d'un défaut. Cependant, un tel indice n'est pas suffisant pour établir un diagnostic du système ou pour suivre l'évolution du défaut.

Les trajectoires spectrales contiennent une grande quantité d'information. En effet, elles sont constituées d'éléments spectraux qui ont été individuellement caractérisés. Ces caractéristiques vont servir à dériver des **tendances** qui seront le reflet de l'évolution de la santé du système. Les variations des tendances seront particulièrement intéressantes pour le diagnostic. Une tendance stable ne variant pas ou peu laisse penser que le système surveillé ne varie pas non plus, autrement dit qu'il ne se dégrade pas. En revanche, la croissance ou la décroissance d'une tendance, un changement de valeur moyenne ou de variance sont autant de phénomènes qui indiquent une évolution du système. Ce sont les tendances non stables qui intéresseront davantage l'utilisateur du système de surveillance et permettront de réaliser le diagnostic automatique.

Au cours de ce chapitre, nous allons voir dans un premier temps comment transformer les trajectoires spectrales, détectées dans une séquence de signaux, en **tendances** permettant de suivre l'évolution de l'état de santé du système. Plusieurs applications sur signaux réels viendront illustrer la génération des tendances. Ensuite, nous nous intéresserons à l'identification

automatique des tendances porteuses d'information sur l'évolution du système. En particulier, nous proposerons une méthode basée sur des tests statistiques de non-stationnarité.

## 4.1 Génération de tendances

Une trajectoire spectrale contient une grande quantité d'information. Brutes, ces informations sont délicates à analyser. Néanmoins, en séparant toutes ces informations et en les affichant sous une forme différente, il devient alors possible de surveiller simplement l'évolution du système.

### 4.1.1 Génération à partir de la carte d'identité spectrale

Les trajectoires permettent de suivre une structure spectrale à travers une séquence de signaux. Il existe trois types de trajectoires permettant le suivi de trois types de structures spectrales : les pics, les séries harmoniques et les bandes de modulation. Chacune de ces structures a été caractérisée à un moment ou un autre de la chaîne de traitement d'une séquence de signaux. Chaque structure, en fonction de son type, possède une carte d'identité spectrale listant toutes les caractéristiques qui lui sont propres. Pour rappel, les caractéristiques principales<sup>1</sup> d'un pic sont données au chapitre 2, dans le tableau 2.1 (page 27). De même, les tableaux 2.2 et 2.3 (pages 45 et 46) récapitulent les caractéristiques présentes sur les cartes d'identité des séries harmoniques et des bandes de modulation, respectivement.

Les tendances sont générées à partir des caractéristiques présentes sur les cartes d'identité spectrale. Dans le cas d'une trajectoire harmonique, nous générerons alors 8 tendances correspondant aux 8 caractéristiques données dans le tableau 2.2. Les tendances héritent du support temporel de la trajectoire qui leur a donné naissance. Si une trajectoire démarre à l'instant  $t_k$  et se termine à l'instant  $t_{k+m}$ , avec  $t_k < t_{k+m}$ , alors les tendances qui en découlent démarreront à l'instant  $t_k$  et se termineront à l'instant  $t_{k+m}$ . Si la trajectoire est en sommeil à un instant donné, les tendances seront considérées elles aussi en sommeil à cet instant-là. Pour résumer, une tendance permet de visualiser une caractéristique de la structure spectrale suivie dans le temps sous la forme d'une trajectoire.

La figure 4.1 illustre le principe de génération des tendances sur un exemple très simple. La carte temps-fréquence, en haut de la figure, nous montre une trajectoire harmonique démarrant à l'instant  $t_1$  et se terminant à l'instant  $t_4$ . Chacune des séries harmoniques dans la structure possède une carte d'identité spectrale, volontairement simplifiée pour l'exemple. On trouve deux caractéristiques sur les cartes d'identité, à savoir le nombre de pics  $N_{\text{pics}}$  dans la série et l'énergie  $E$  de la série. En utilisant ces caractéristiques aux instants où la trajectoire est active, nous pouvons construire les deux courbes (en bas de la figure) que nous appelons **tendances**. Elles permettent en un coup d'œil de visualiser l'évolution des caractéristiques de la série harmonique suivie.

1. Une liste complète des caractéristiques d'un pic spectral est donnée dans l'annexe B, section B.2.4.

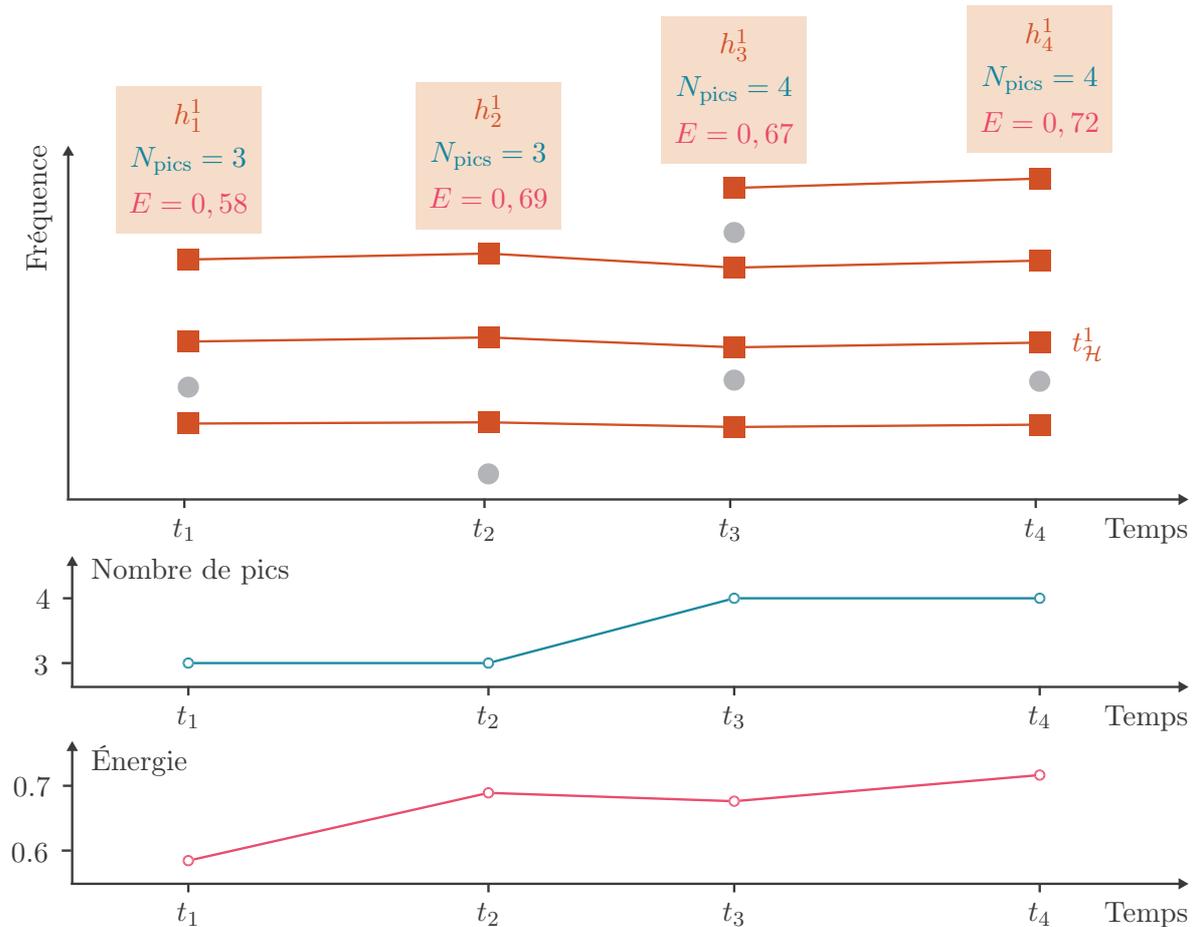


FIGURE 4.1 – Illustration du principe de génération des tendances. La trajectoire harmonique représentée sur la carte temps-fréquence (en haut) donne naissance à deux tendances : le nombre de pics dans la série (au milieu) et l'énergie de la série (en bas).

Le nombre total de tendances générées pour une séquence de signaux est grand. Pour le calculer, il suffit de compter le nombre de trajectoires de chaque type. Soit  $\#\mathcal{T}_{\mathcal{P}}$  le nombre de trajectoires de pics,  $\#\mathcal{T}_{\mathcal{H}}$  le nombre de trajectoires harmoniques et  $\#\mathcal{T}_{\mathcal{M}}$  le nombre de trajectoires de modulations. Il faut ensuite multiplier  $\#\mathcal{T}_{\mathcal{P}}$  par le nombre de caractéristiques retenues sur la carte d'identité des pics, à savoir 3. De même, il faut multiplier  $\#\mathcal{T}_{\mathcal{H}}$  par 8 et  $\#\mathcal{T}_{\mathcal{M}}$  par 13. La somme des trois nombres coefficientés correspond au nombre total de tendances générées. En général, il n'est pas rare d'avoir pour des signaux réels dont le spectre est riche, des milliers voire des dizaines de milliers de tendances.

#### 4.1.2 Applications

Cette section va présenter quelques tendances qu'il est possible de générer en utilisant la méthode proposée dans ce manuscrit. Tout d'abord, nous allons compléter les résultats obtenus jusqu'à présent pour les applications présentées dans les sections 2.5 et 3.4. Ensuite,

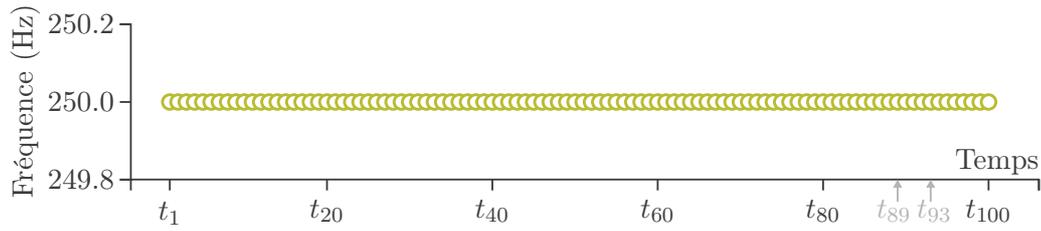


FIGURE 4.2 – Fréquence de la trajectoire harmonique correspondant à l’engrènement du défaut simulé.

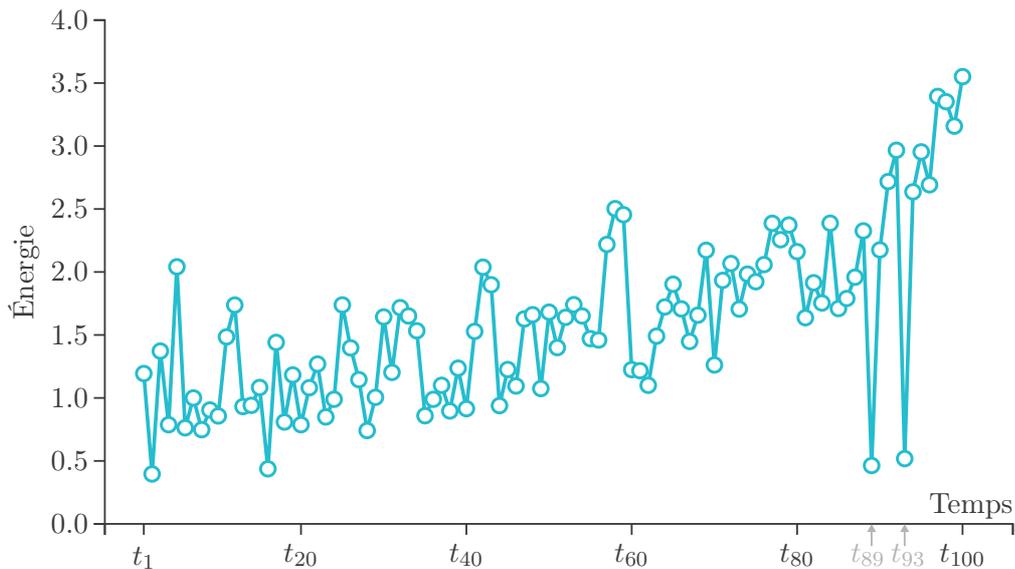


FIGURE 4.3 – Énergie de la trajectoire harmonique correspondant à l’engrènement du défaut simulé.

nous visualiserons des tendances générées à partir d’une séquence de signaux mesurés sur une éolienne en action.

#### 4.1.2.1 Simulation d’un défaut

La séquence de signaux simulant un défaut d’engrènement a été présentée dans les sections 2.5.2 et 3.4.1. L’analyse de cette séquence nous donne plusieurs trajectoires spectrales, dont 5 d’entre elles étaient particulièrement attendues. Il s’agit de la trajectoire harmonique à 250 Hz représentant la signature spectrale du phénomène d’engrènement et des 4 trajectoires de modulations à 10 Hz autour des 4 harmoniques de la série d’engrènement. Les fréquences porteuses des modulations sont donc 250, 500, 750 et 1000 Hz.

Les figures 4.2 et 4.3 représentent deux tendances générées à partir de la trajectoire harmonique correspondant au phénomène d’engrènement. La première tendance, illustrée sur la figure 4.2, nous permet de suivre l’évolution de la fréquence fondamentale de la série au cours de la séquence de signaux. Cette tendance est parfaitement stable et ne présente aucune varia-

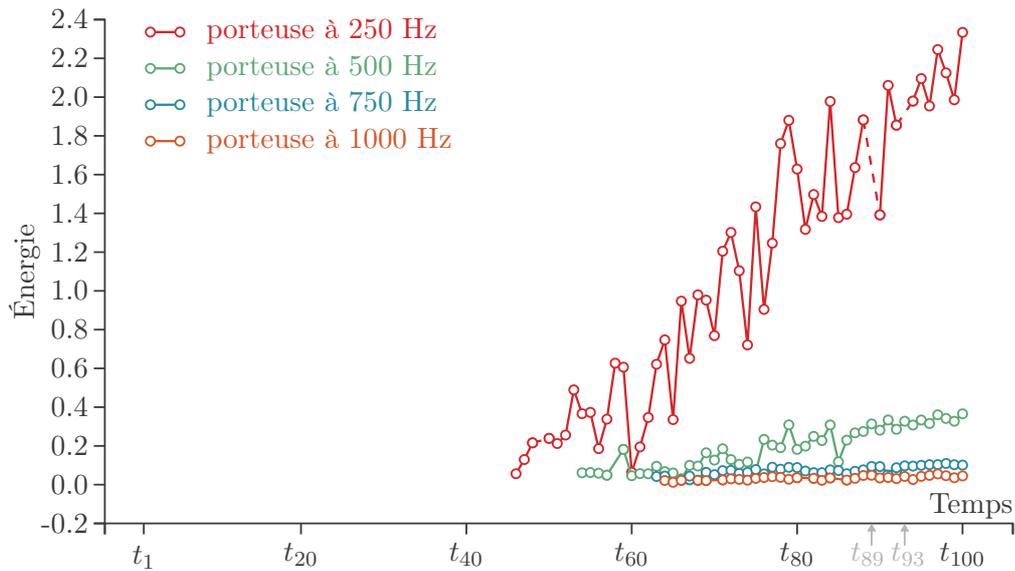


FIGURE 4.4 – Énergie des 4 trajectoires de modulations situées autour des harmoniques de la série d'engrènement.

tion : la fréquence fondamentale de la série est de 250 Hz à partir de l'instant  $t_1$  et elle gardera cette valeur jusqu'à l'instant  $t_{100}$ . Cette tendance nous permet de vérifier que la fréquence fondamentale est correctement détectée tout au long de la séquence, et ce, malgré l'absence du fondamental aux instants  $t_{89}$  et  $t_{93}$ . Cependant, la stabilité de cette tendance ne donne aucun indice sur la présence et l'aggravation du défaut.

L'énergie de la trajectoire harmonique (figure 4.3) est plus intéressante pour le diagnostic. En effet, malgré la variance de cette tendance, nous pouvons observer une croissance au cours du temps. Il y a donc une évolution de la tendance, reflet d'un changement dans le système. Deux points vers la fin de la courbe présentent une énergie faible et pourraient être vus comme des valeurs aberrantes. Il n'en est rien car ces deux baisses brutales sont situées aux instants  $t_{89}$  et  $t_{93}$  ; ces faibles valeurs correspondent donc à la suppression artificielle du fondamental qui porte une part importante de l'énergie de la série.

Les figures 4.4 et 4.5 présentent des tendances générées à partir des 4 trajectoires de modulations autour des harmoniques de la série d'engrènement. Même si l'échelle de la figure 4.4 ne permet pas d'apprécier correctement l'évolution de l'énergie dans les trajectoires dont la porteuse est à 750 et à 1000 Hz, l'énergie des 4 modulations augmentent bien au cours de la séquence de signaux. De même, les tendances représentant le nombre de pics dans les modulations (figure 4.5) croissent au cours du temps.

Finalement, plusieurs tendances présentent un signe d'évolution temporelle. Ces évolutions sont le reflet de la présence d'un défaut et de l'augmentation de sa sévérité avec le temps. Le fait que plusieurs tendances évoluent simultanément apporte de la redondance et donne davantage de poids au diagnostic d'un défaut.

Il est à noter également que la trajectoire de modulations dont la porteuse est à 250 Hz est

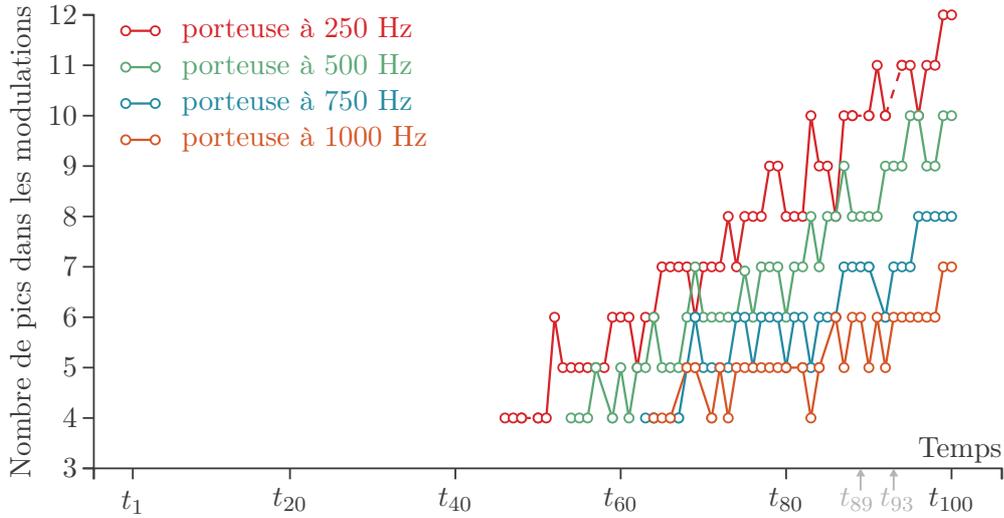


FIGURE 4.5 – Nombre de pics présents dans les 4 trajectoires de modulations.

en sommeil aux instants  $t_{89}$  et  $t_{93}$  à cause de l'absence de la fréquence porteuse. Les tendances générées à partir de cette trajectoire héritent de son support temporel et sont alors elles aussi en sommeil aux instants concernés.

#### 4.1.2.2 Banc d'essai éolien — forte dégradation du roulement principal

Les trois exemples d'application qui vont suivre sont réalisés sur trois séquences de signaux provenant du banc de test éolien. Chaque séquence correspond à la dégradation du roulement principal du banc de test. Ce roulement est remplacé par un neuf à chaque nouvel essai. Les trois tests sont arrêtés à des degrés d'usure différents du roulement principal. La description complète du banc d'essai et des signaux recueillis est disponible dans l'annexe C.

Nous allons commencer par la séquence de 17 signaux correspondant à la dégradation du roulement principal la plus forte. Il s'agit du cas le plus facile à détecter, dans la mesure où la signature du défaut dans le spectre vibratoire est importante. Cette séquence a déjà été utilisée dans les sections 2.5.3 et 3.4.2, afin d'illustrer l'identification des séries harmoniques et des bandes de modulation, ainsi que le suivi des structures spectrales.

Pour rappel, la méthode identifie un grand nombre de trajectoires dans cette séquence de signaux : 828 trajectoires harmoniques, 9 373 trajectoires de modulations et 12 406 trajectoires parmi les pics restants. La transformation de ces trajectoires va alors générer la création de 165 691 tendances. Bien évidemment, la visualisation manuelle et exhaustive de toutes ces tendances n'est pas imaginable. Nous discuterons de ce point dans la section 4.2.

En attendant, nous allons nous concentrer sur deux tendances générées à partir de la trajectoire harmonique à 3,45 Hz. Cette trajectoire particulière, comme expliqué dans la section 3.4.2, correspond à la signature spectrale du passage des billes du roulement sur la bague externe. L'énergie contenue dans cette signature est représentée sur la figure 4.6. Cette ten-

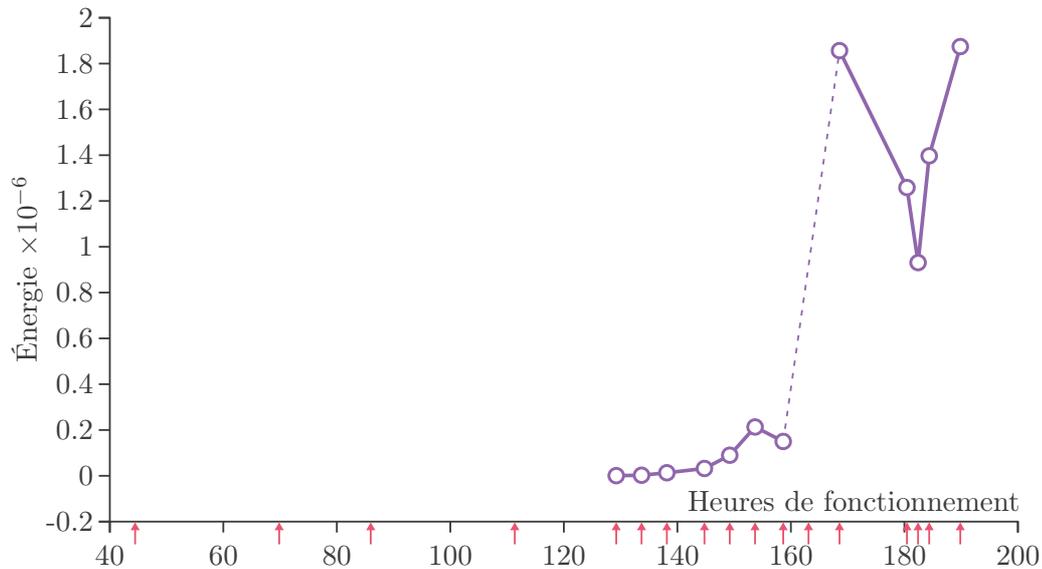


FIGURE 4.6 – Énergie vibratoire de la bague externe du roulement principal du banc d’essai lors de sa forte dégradation. Les petites flèches roses sur l’axe des abscisses correspondent aux instants où les 17 signaux de la séquence ont été acquis.

dance démarre à 129 heures de fonctionnement, date à laquelle la trajectoire du pic à 3,45 Hz se transforme en trajectoire harmonique. L’énergie de la série augmente d’abord doucement, jusqu’à 163 heures de fonctionnement. À cette date, le signal acquis est hautement non stationnaire à cause d’une forte impulsion dans les vibrations (45 g). Les experts en mécanique du [CETIM](#) font l’hypothèse que l’impulsion correspond à un détachement de matière sur la bague externe. Dans tous les cas, la non-stationnarité empêche l’analyse du signal. L’état sommeil permet au suivi de la trajectoire de continuer malgré tout. Le test est poursuivi jusqu’à ce que le banc ne puisse plus tourner à cause de la trop forte dégradation du roulement. Durant la fin de vie du roulement, l’énergie vibratoire dans la signature de la bague externe est forte. Le démontage du roulement et son inspection à la fin du test montre un roulement écaillé sur tout le périmètre de la bague externe.

En parallèle de l’augmentation de l’énergie vibratoire de la bague externe, la fréquence fondamentale de la série harmonique décroît faiblement. Ce phénomène, illustré par la tendance sur la figure 4.7, est une conséquence de l’écaillage de la bague externe. À partir du moment où le roulement commence à se détériorer, le passage des billes n’est plus parfait et du glissement peut entrer en jeu. C’est ce glissement qui explique la faible diminution de la fréquence fondamentale (de l’ordre de 0,02 Hz).

#### 4.1.2.3 Banc d’essai éolien — dégradation moyenne du roulement principal

Nous allons à présent analyser une séquence de 12 signaux provenant d’une seconde dégradation du roulement principal du banc d’essai éolien. La dégradation moyenne du roulement a nécessité un peu plus de 200 heures de fonctionnement. Contrairement à la forte dégradation,

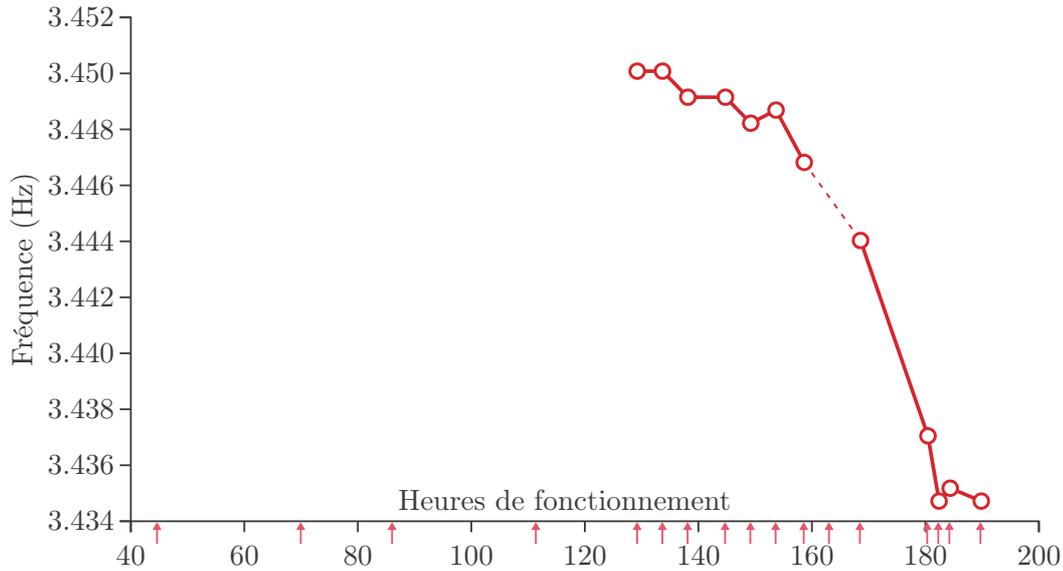


FIGURE 4.7 – Fréquence fondamentale de la série harmonique correspondant à la bague externe du roulement principal du banc d’essai lors de sa forte dégradation. La décroissance s’explique par l’apparition de glissement dû aux premiers symptômes de l’écaillage de la bague externe.

le test a été arrêté avant que l’usure du roulement empêche le fonctionnement du banc. Cet arrêt prématuré a permis, après démontage du roulement, de visualiser un défaut en cours de formation. L’usure a conduit à l’écaillage de la bague externe : trois écailles de 2 mm de largeur se sont formées. La profondeur maximale de l’écaillage est de 210  $\mu\text{m}$ .

La séquence de 12 signaux a été enregistrée par le capteur dans la direction  $-z$ . L’application de la méthode proposée dans cette thèse conduit à la génération d’environ 180 000 tendances. L’une d’entre elles mérite une attention particulière, car elle permet de visualiser l’évolution du défaut. Il s’agit de la tendance d’énergie, dérivée à partir de la trajectoire harmonique de fréquence fondamentale 2,72 Hz. Cette fréquence correspond à nouveau à la fréquence caractéristique de la bague externe du roulement. En effet, le modèle de roulement utilisé pour l’essai de dégradation moyenne est différent du modèle utilisé pour la forte dégradation. Les modèles ayant des géométries différentes, les fréquences caractéristiques ne sont pas les mêmes.

L’énergie vibratoire de la bague externe est présentée sur la figure 4.8. La tendance démarre dès le premier signal acquis au bout de 20 heures de fonctionnement. La série harmonique correspondant à la bague externe du roulement est présente durant toute la séquence, à part pour le signal vers 96 heures de fonctionnement. L’état sommeil permet de continuer le suivi malgré cette non-détection. L’énergie dans la série est à la fois stable et très faible jusqu’à 193 heures de fonctionnement. Ensuite, l’écaillage de la bague externe se traduit par une augmentation de l’énergie vibratoire dans la série harmonique de fréquence fondamentale 2,72 Hz.



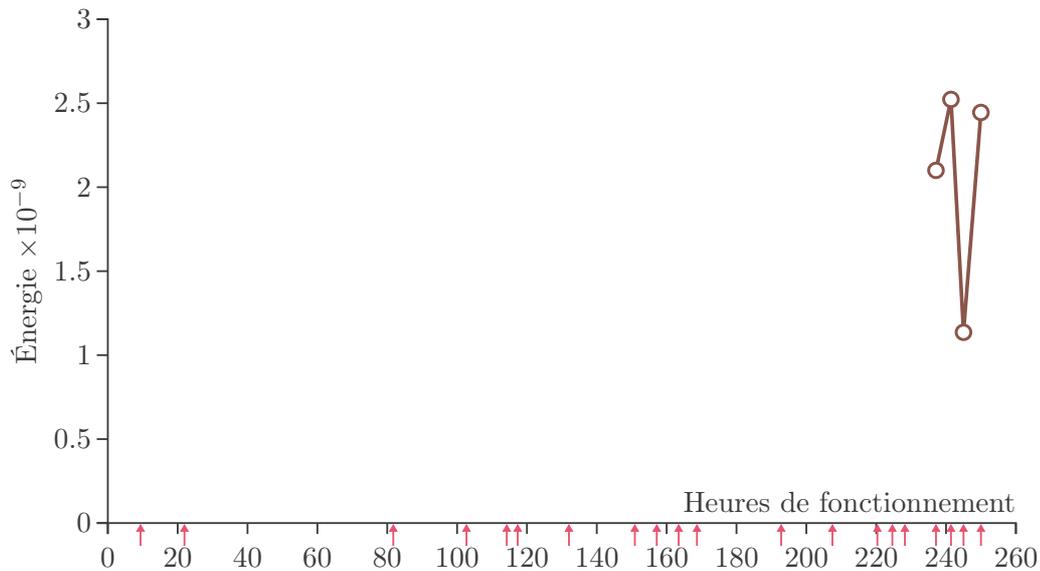


FIGURE 4.9 – Énergie vibratoire de la bague externe du roulement principal du banc d’essai lors de sa faible dégradation.

Comparée aux autres tests d’usure, l’énergie vibratoire de la bague externe est ici 1000 fois plus faible à la fin du test.

L’utilisation des séquences de signaux acquises sur le banc d’essai permet de tester la méthode proposée sur des signaux réels. Cependant, l’environnement du système surveillé est contrôlé. En particulier, lors des phases d’acquisition, les conditions de fonctionnement sont parfaitement maîtrisées et identiques afin de faciliter l’analyse des signaux. Il est à présent temps d’essayer la méthodologie proposée sur des signaux réels, acquis en conditions réelles.

#### 4.1.2.5 Surveillance d’une éolienne

La séquence de signaux que nous allons étudier à présent est issue de la surveillance d’une véritable éolienne en production dans un parc éolien du sud de la France. L’acquisition d’une telle séquence a été réalisée dans le cadre du projet européen [*KAStrion*], grâce au *CMS* développé par l’entreprise polonaise *EC System*, partenaire du projet. L’éolienne surveillée a été mise à disposition par la société française *VALEMO*, elle aussi partenaire du projet. Des photos de l’éolienne et du système de surveillance sont disponibles sur la figure 4.10.

L’éolienne surveillée mesure 70 m de haut et possède des pâles de 40 m de long. Elle est capable de fonctionner dans une plage de vent allant de 3 à 25 m/s et de générer jusqu’à 2 MW. La boîte de vitesse servant de multiplicateur est composée de trois trains : un train planétaire suivi de deux trains simples à axe parallèle. De nombreux capteurs vibratoires et électriques ont été placés dans la nacelle de l’éolienne. Nous nous intéresserons uniquement à la séquence de signaux enregistrés par l’accéléromètre placé sur la partie avant de la génératrice, là où l’arbre rapide entre dans la génératrice.

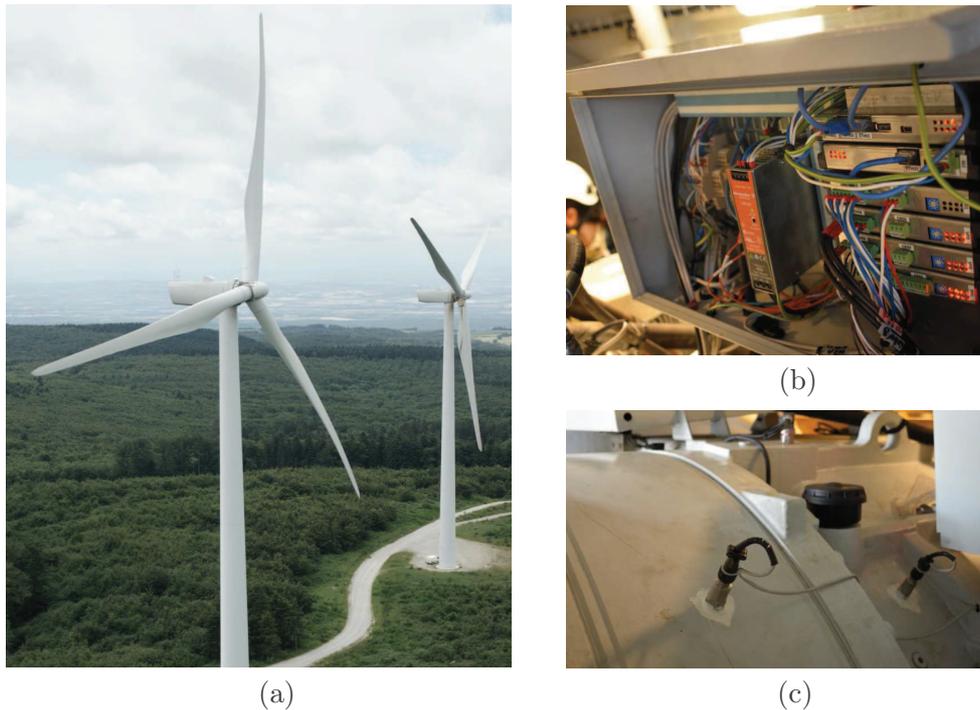


FIGURE 4.10 – Photos de (a) l'éolienne surveillée par (b) le système d'acquisition enregistrant les signaux issus des accéléromètres, en (c), posés sur le roulement principal et sur le multiplicateur.

Sur l'année 2014 (du 10 février au 31 décembre), 165 signaux ont été acquis avec une fréquence d'échantillonnage de 25 kHz pendant une durée de 10 secondes. Les conditions de fonctionnement pour chacune des acquisitions sont similaires. En l'occurrence, la vitesse de rotation de l'arbre rapide pour les 165 signaux varie entre 1780 et 1795 RPM. Si la variation des conditions de fonctionnement est faible, elle n'en reste pas moins gênante pour la méthode de suivi utilisée. Les signaux sont donc ré-échantillonnés dans le domaine angulaire, afin d'effacer ces variations. Le signal de vitesse utilisé pour le ré-échantillonnage est mesuré sur l'arbre rapide, en entrée de la génératrice. Cela signifie que la fréquence de rotation de l'arbre rapide qui est d'environ 30 Hz ( $1780/60 \approx 30$ ) deviendra l'ordre 1. À l'instar des fréquences pour le domaine temporel, les ordres quantifient la répétition d'un phénomène dans le domaine angulaire. Des informations supplémentaires sur le ré-échantillonnage angulaire sont disponibles dans [Firla 2016] et dans la section B.2.1 de l'annexe B.

Dans cette séquence de signaux réels, nous allons nous intéresser plus particulièrement à la signature spectrale de l'arbre rapide. Cette signature prend la forme d'une série harmonique dont l'ordre fondamental est 1. Le suivi de cet ordre est présenté sur la figure 4.11. Il a la particularité d'être réalisé par deux trajectoires différentes. La première trajectoire démarre début février et se termine le 9 juin. La seconde trajectoire démarre au moment où la première se termine et continue jusqu'à fin décembre. Si le suivi est réalisé par deux trajectoires, c'est parce que la première trajectoire a *décroché*. En effet, si on regarde le zoom autour du 9 juin, on peut voir que l'ordre de la première trajectoire (en rouge) augmente légèrement jusqu'à 1,0011

ordre. Le 9 juin, deux séries harmoniques sont présentes, dont les ordres sont très proches (0,9963 et 1,0011). La détection de la série harmonique de l'arbre rapide à l'instant suivant est plus proche de l'ordre 0,9963. C'est donc avec lui que l'association est faite pendant le suivi (voir l'équation (3.7) dans la section 3.2.1). La première trajectoire ne trouve pas d'autre candidat pour se poursuivre, ni à cet instant-là, ni au suivant. Elle va donc se terminer.

Les deux trajectoires permettent de suivre la série harmonique de l'arbre rapide au cours de la séquence de signaux. La superposition des deux trajectoires à la date du 9 juin est un artefact de la méthode de suivi. Il est cependant possible d'utiliser les deux trajectoires pour poursuivre le diagnostic de l'éolienne. L'ordre de l'arbre rapide, contrairement à ce que la très petite échelle de la figure 4.11 laisse penser, est très stable.

La figure 4.12 affiche quant à elle l'énergie des séries harmoniques contenue dans nos deux trajectoires. Un changement important intervient sur cette courbe, puisque la valeur moyenne et la variance de l'énergie changent complètement début novembre. Ce changement intervient juste après deux longues périodes sans signal (du 21 juin au 31 juillet et du 5 septembre au 30 octobre), périodes pendant lesquelles l'éolienne était en maintenance. Parmi les opérations de maintenance effectuées, un réalignement du multiplicateur et de la génératrice a été fait. La forte baisse du niveau d'énergie vibratoire suite au réalignement semble indiquer qu'un désalignement était présent pendant la première partie de 2014. Malheureusement, les données avant 2014 ne sont pas disponibles et il est donc impossible de voir l'installation progressive de ce désalignement.

### 4.1.3 Conclusions

L'analyse des différentes séquences de signaux permet de tirer les conclusions suivantes quant aux résultats de la méthode proposée :

- le nombre de tendances générées est important (plusieurs centaines de milliers). La visualisation exhaustive de toutes les tendances n'est pas envisageable ;
- les tendances intéressantes pour le diagnostic du système surveillé sont les tendances présentant une évolution, un changement ;
- l'évolution d'une tendance peut prendre plusieurs formes : croissance, décroissance, changement de valeur moyenne et, ou de variance.

Au vu de ces conclusions, il est nécessaire de proposer une méthode permettant de trier automatiquement les tendances afin de les séparer en deux classes distinctes : les tendances stables signifiant que l'état de santé du système ne change pas ou que les tendances ne sont pas porteuses de l'information du défaut, et les tendances qui évoluent et reflètent la fatigue et l'usure du système. Un tel classifieur est proposé dans la section suivante.

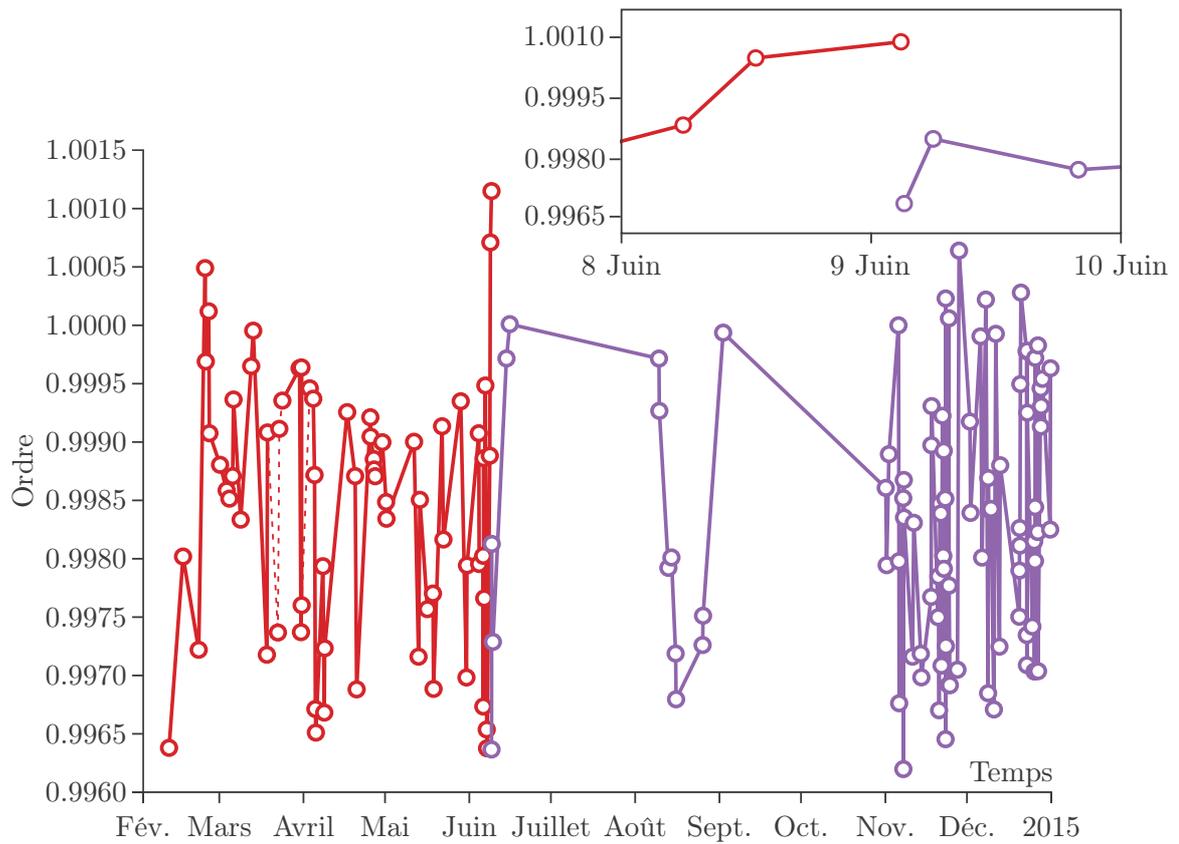


FIGURE 4.11 – Évolution de l'ordre de la série harmonique correspondant à l'arbre rapide de l'éolienne surveillée. En haut, un zoom permet de visualiser le *décrochage* du suivi.

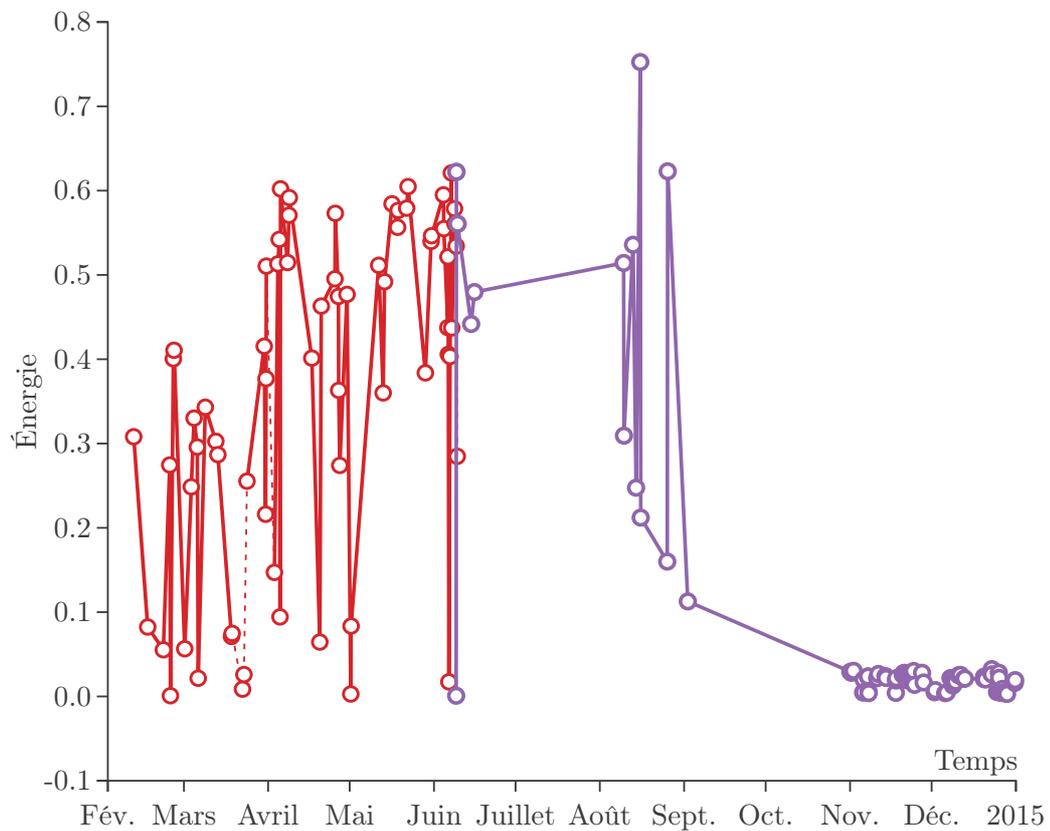


FIGURE 4.12 – Évolution de l'énergie de la série harmonique correspondant à l'arbre rapide de l'éolienne surveillée.

## 4.2 Détection de l'évolution d'un système

L'évolution d'un système sous surveillance se caractérise par la présence d'un changement dans au moins une des tendances générées à partir de la séquence de signaux acquis. Nous allons donc nous intéresser à l'identification des tendances présentant une évolution. Cette identification est complexe pour plusieurs raisons qui sont détaillées dans la section 4.2.1. Nous proposerons alors une méthode d'identification basée sur plusieurs tests statistiques issus du domaine de l'économétrie<sup>2</sup>. Un vote entre les différents tests permettra finalement de prendre une décision binaire pour chacune des tendances.

### 4.2.1 Contexte

La détection d'un changement dans une tendance est un problème déjà abordé dans la littérature. Le vocabulaire est légèrement différent de celui employé dans ce manuscrit : ce que nous appelons tendance est par définition une série temporelle dans laquelle nous cherchons à détecter un changement, appelé anomalie ou nouveauté. Dans une revue sur la détection d'anomalie [Chandola 2009], de nombreuses méthodes très différentes sont présentées. Par exemple, dans [Liu 2013], les séries temporelles sont modélisées par leur forme dans un sous-ensemble dépendant du caractère linéaire ou non de la série. Un écart par rapport à la forme apprise pendant la phase d'entraînement sera considéré comme une anomalie et déclenchera une alarme. La modélisation des séries pendant la phase d'apprentissage est réalisée sur environ 2 000 points, ce qui nécessite d'avoir un long historique pour chaque série temporelle. Des historiques plus courts entraîneraient des modèles incomplets déclenchant alors de nombreuses fausses alarmes. Dans [Ma 2003], un apprentissage est également nécessaire. Cet apprentissage est dynamique dans la mesure où de nouveaux points peuvent s'ajouter à l'historique, tout comme les plus vieux points peuvent être oubliés. Cependant, pour initialiser l'apprentissage, il est nécessaire d'avoir un historique de l'ordre de 200 à 400 points. La détection de nouveauté est réalisée par une régression à vecteur de supports.

Les séries temporelles des méthodes présentées dans la revue [Chandola 2009] sont la plupart du temps construites en interrogeant régulièrement un capteur, comme un thermomètre ou un encodeur. Il n'y a pas besoin de traiter les mesures effectuées pour estimer les éléments contenus dans ces mesures. En conséquence, les séries sont considérées comme *complètes*, c'est-à-dire qu'il n'y a pas de valeur manquante dans la série temporelle ; elle n'est pas lacunaire. La régularité et la complétude ne font pas partie des qualités des tendances générées dans la section 4.1.2. Au contraire, les points irrégulièrement espacés et l'état sommeil des tendances sont deux difficultés qu'il faut prendre en compte lors de la recherche d'anomalies. Ainsi, dans un contexte de surveillance, l'identification de nouveauté dans une tendance est rendue complexe par plusieurs facteurs synthétisés ici :

- l'évolution de la tendance peut prendre plusieurs formes différentes. Les exemples donnés dans la section 4.1.2 font état de croissance, de décroissance ou de modification de la

---

2. L'économétrie est un domaine mélangeant science économique et statistiques afin d'établir et de vérifier des modèles économiques.

valeur moyenne et de la variance ;

- les tendances générées sont hétérogènes. Elles ont des unités différentes, voire pas d'unité du tout. L'ordre de grandeur des variations peut être très différent d'une tendance à l'autre ;
- les tendances ont des longueurs différentes. L'identification de nouveauté doit fonctionner aussi bien sur des tendances d'une dizaine de points que sur des tendances beaucoup plus longues contenant des centaines de points ;
- les tendances héritent du support temporel de la trajectoire qui leur a donné naissance et peuvent donc également être en sommeil à certains instants, ce qui en fait des séries lacunaires.

À ces difficultés, il ne faut pas oublier non plus le cahier des charges fixé dans cette thèse, qui comprend notamment des contraintes de généricité et de vitesse d'exécution. Ce contexte particulièrement contraint nous a encouragé à proposer une méthode basée sur l'hypothèse suivante : une tendance ne présentant pas d'évolution est *stable*, autrement dit elle est stationnaire. *A contrario*, un changement dans une tendance impliquera nécessairement une non-stationnarité. En effet, la définition stricte de la stationnarité stipule que la distribution conjointe de probabilité d'un processus stationnaire ne dépend pas du temps. Un processus aléatoire  $\{X_t\}$  est donc considéré comme strictement stationnaire si :

$$\forall k, \forall \tau, F_X(X_1, \dots, X_k) = F_X(X_{1+\tau}, \dots, X_{k+\tau}), \quad (4.1)$$

où  $F_X(\cdot)$  représente la fonction de répartition de la distribution jointe du processus  $\{X_t\}$ . Une série temporelle stationnaire aura donc des propriétés statistiques identiques à chaque instant ; il n'y a pas d'*évolution* des propriétés de la série au cours du temps.

La définition stricte de la stationnarité est rarement utilisée en pratique, car l'estimation de la loi de probabilité d'une distribution jointe est difficile à réaliser. Elle est généralement remplacée par une définition plus simple appelée stationnarité au sens large, ou encore, stationnarité à l'ordre 2. Un processus  $\{X_t\}$  est stationnaire au sens large s'il vérifie les 2 conditions suivantes :

$$\forall k, \mathbb{E}[X_k] = \mu, \quad (4.2)$$

$$\forall k, \forall \tau, R_X(k, k + \tau) = \mathbb{E}[X_k X_{k+\tau}^*] = f(\tau), \quad (4.3)$$

où  $\mathbb{E}[\cdot]$  représente l'espérance mathématique,  $R_X$  l'autocorrélation de  $X$  et  $f(\tau)$  une fonction dépendant uniquement de  $\tau$ . La condition (4.2) stipule que l'espérance est constante au cours du temps et prend une valeur constante  $\mu \in \mathbb{R}$ . La condition (4.3) stipule quant à elle que l'autocorrélation entre  $X_k$  et  $X_{k+\tau}$  ne dépend que du décalage  $\tau$ , c'est-à-dire que pour un même décalage, quel que soit l'instant  $k$  choisi, la valeur de l'autocorrélation sera identique. Ces conditions demandent donc à ce que les moments d'ordre 1 et 2 du processus n'évoluent pas au cours du temps.

Pour être stationnaire, une tendance devra vérifier les conditions (4.2) et (4.3). Si l'une des conditions n'est pas respectée, la tendance est alors non stationnaire. Cela signifie qu'il y a plusieurs possibilités de montrer qu'une tendance est non stationnaire. L'identification de

la non-stationnarité d'une série n'est pas simple et ressemble beaucoup aux problématiques rencontrées par les statisticiens en économétrie ou en finance. En effet, les changements ne sont pas nécessairement francs ou brutaux, mais prennent souvent la forme d'une lente évolution qu'on cherche à prédire. La similarité des problématiques nous a encouragé à essayer certaines méthodes utilisées en économétrie. Nous proposons une méthode basée sur l'utilisation de plusieurs tests statistiques ; chaque test jouant sur une des possibilités de montrer la stationnarité ou la non-stationnarité de la tendance. La multiplicité des tests doit permettre d'identifier les différentes formes de changement qui peuvent apparaître dans une tendance. La description des tests utilisés est faite dans la section suivante.

## 4.2.2 Description des tests statistiques

Les tests statistiques que nous avons choisi d'utiliser pour caractériser la présence d'un changement dans une tendance sont au nombre de six. Les deux premiers, que nous présenterons en section 4.2.2.1, sont basés sur la comparaison des coefficients obtenus dans le cadre d'une régression linéaire. Les trois tests suivants sont classiquement utilisés en économétrie et testent la présence d'une racine unitaire dans la tendance qui impliquerait sa non-stationnarité. Ce principe sera expliqué dans la section 4.2.2.2. Enfin, le dernier test est également issu du domaine de l'économétrie. Il cherche à vérifier de manière directe l'hypothèse de stationnarité. Il sera présenté dans la section 4.2.2.3.

### 4.2.2.1 Tests basés sur une régression linéaire

Les tests de Student et de Chow que nous allons présenter dans cette section sont tous deux réalisés à la suite d'une régression linéaire. Ils font donc l'hypothèse qu'une tendance  $x_t$  peut se modéliser comme une fonction affine du temps. Les termes de la tendance pourraient donc s'écrire sous la forme :

$$x_i = \mu + \delta t_i + \varepsilon_i, \quad (4.4)$$

avec  $(\mu, \delta) \in \mathbb{R}^2$  et  $\varepsilon_i$  un bruit blanc gaussien qui représente l'erreur résiduelle entre le modèle affine et les valeurs réelles de  $x_i$ .

#### Test de la pente nulle

Pour être stationnaire, la condition (4.2) stipule que la moyenne d'une tendance doit être constante au cours du temps. Si la tendance est modélisée par une fonction affine comme proposée dans (4.4), la moyenne ne pourra être constante que dans le cas où le coefficient directeur de la droite est nulle ( $\delta = 0$ ). Nous proposons donc de calculer les coefficients de la régression linéaire entre les valeurs de la tendance et son échelle temporelle et de vérifier, grâce à un test de Student [Student 1908], que la pente de la régression est nulle.

En effet, le modèle de régression linéaire (4.4) inclut l'hypothèse d'erreurs de modèle  $\varepsilon_i$

indépendantes et identiquement distribuées selon une loi normale de moyenne nulle et de variance  $\sigma^2$ , ce que nous notons :  $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ . Ainsi, les variables  $x_i$  suivent aussi une loi normale et l'estimateur du maximum de vraisemblance de la pente s'écrit :

$$\hat{\delta} = \frac{\sum (t_i - \bar{t}) X_i}{\sum (t_i - \bar{t})^2}, \quad (4.5)$$

avec  $\bar{t} = \frac{1}{n} \sum t_i$  et  $n$  le nombre de points dans la série temporelle. Cet estimateur étant une fonction linéaire des variables aléatoires  $X_i$ , il suit une loi normale. On peut montrer qu'il est biaisé avec une variance :

$$\sigma_{\hat{\delta}}^2 = \frac{\sigma^2}{\sum (t_i - \bar{t})^2}. \quad (4.6)$$

Nous obtenons ainsi que

$$\frac{\hat{\delta} - \delta}{\sigma_{\hat{\delta}}} \sim \mathcal{N}(0, 1). \quad (4.7)$$

Un estimateur de  $\sigma_{\hat{\delta}}^2$  s'écrit, en utilisant l'estimateur du maximum de vraisemblance de  $\sigma^2$  (non biaisé) :

$$s_{\hat{\delta}}^2 = \frac{1}{n-2} \frac{\sum (X_i - \hat{X}_i)^2}{\sum (t_i - \bar{t})^2}, \quad (4.8)$$

avec  $\hat{X}_i = \hat{\mu} + \hat{\delta} t_i$ ,  $\hat{\mu}$  étant l'estimateur du maximum de vraisemblance de  $\mu$ . On peut montrer que

$$\frac{\hat{\delta} - \delta}{s_{\hat{\delta}}} \sim T_{n-2}, \quad (4.9)$$

où  $T_{n-2}$  désigne une variable aléatoire de Student avec  $n-2$  degrés de liberté. Pour tester les deux hypothèses :

$$\mathcal{H}_0 : \delta = 0,$$

$$\mathcal{H}_1 : \delta \neq 0,$$

on utilise alors la statistique de Student suivante :

$$S = \frac{\hat{\delta}}{s_{\hat{\delta}}} \quad (4.10)$$

pour décider de l'acceptation ou du rejet de  $\mathcal{H}_0$ .

En résumé, le test de la pente nulle permet de décider entre les deux hypothèses suivantes :

- **hypothèse nulle** : le coefficient de la régression linéaire appliquée à la série temporelle est nul ;
- **hypothèse alternative** : le coefficient de la régression linéaire appliquée à la série temporelle est non nul.

Interprétée plus largement, l'hypothèse nulle est une condition nécessaire à la stationnarité de la tendance, alors que son rejet représente forcément un cas de tendance non stationnaire.

## Test de Chow

Le test de Chow [Chow 1960] détermine si les coefficients de la régression linéaire de deux séries sont égaux. S'il est appliqué sur deux sous-ensembles d'une même série temporelle, il permet alors de déterminer s'il y a une *cassure* de pente significative au sein de la série. Si une telle cassure est effectivement présente, la série est nécessairement non stationnaire.

La séparation de la série en deux sous-ensembles à comparer est un choix qu'il faut réaliser : à partir des  $n$  points de la série totale, il faut former deux ensembles de  $n_1$  et  $n_2$  points, avec  $n_1 < n$  et  $n_2 < n$ . Le nombre de possibilités est grand et de nombreuses stratégies peuvent être imaginées. Dans la suite du manuscrit, nous proposons d'utiliser la stratégie suivante : la série est simplement coupée en deux avec les  $\frac{n}{2}$  premiers points qui formeront le premier sous-ensemble et les  $\frac{n}{2}$  derniers le second. Tous les points de la série sont utilisés et cette stratégie permet alors de vérifier que le début de la série est identique à sa fin. Elle a l'avantage de prendre en compte toute l'information qu'apporte la série. Cependant, quand  $n$  devient grand, l'influence d'un point sur le calcul des coefficients de la régression diminue. Une évolution de la tendance pourrait alors passer inaperçue, ce qui n'est pas souhaitable. Quand la série est longue, il serait préférable d'utiliser une autre stratégie, ne prenant par exemple en compte que les derniers points de la série (les plus récents) pour essayer de détecter la cassure de pente.

Dans les expérimentations qui suivent, la base de données que nous utiliserons pour essayer les différents tests statistiques contient uniquement des tendances courtes (voir section 4.2.3.2). La stratégie proposée permet alors d'exploiter au mieux toute l'information des tendances.

Le calcul du test de Chow nécessite d'effectuer trois régressions : deux d'entre elles sont réalisées sur chacun des sous-ensembles et la troisième est réalisée en utilisant l'ensemble des points compris dans les deux sous-ensembles. La statistique du test de Chow peut s'écrire :

$$F = \frac{\frac{S_u - (S_1 + S_2)}{k}}{\frac{S_1 + S_2}{n_1 + n_2 - 2k}}, \quad (4.11)$$

avec  $k$  le nombre de paramètres à estimer par régression (2 dans notre cas) et  $S_1$ ,  $S_2$  et  $S_u$  la somme du carré des résidus estimés pour, respectivement, le premier sous-ensemble de points, le second sous-ensemble de points et l'union des deux sous-ensembles. Ces résidus doivent être blancs et gaussiens. La statistique  $F$  suit alors une loi de Fisher-Snedecor [Snedecor 1991], car elle est construite comme étant égale au ratio de deux variables indépendantes suivant une loi du  $\chi^2$ .

Les hypothèses du test sont les suivantes :

- **hypothèse nulle** : les coefficients des régressions linéaires des deux sous-séries sont identiques ;
- **hypothèse alternative** : les coefficients des régressions linéaires des deux sous-séries sont différents.

L'hypothèse nulle peut être interprétée comme une condition nécessaire à la stationnarité, alors que l'hypothèse alternative entraîne nécessairement la non-stationnarité de la tendance.

#### 4.2.2.2 Tests basés sur la présence d'une racine unitaire

Les tests que nous allons décrire à présent cherchent à déterminer si la série temporelle possède une racine unitaire, parfois appelée marche aléatoire. La racine unitaire correspond au cas particulier où la modélisation de la série par un modèle autorégressif possède une racine sur le cercle unité. Prenons l'exemple d'une série  $x_t$  que nous allons modéliser par un modèle autorégressif d'ordre 1 :

$$x_t = \rho x_{t-1} + \varepsilon_t, \quad (4.12)$$

avec  $\varepsilon_t$  un bruit blanc gaussien de variance  $\sigma^2$ . Si  $\rho = 1$ , alors  $z = 1$  est une racine de l'équation caractéristique  $z - \rho = 0$ . Le processus dispose alors d'une racine unitaire, incompatible avec l'hypothèse de stationnarité de la série. En effet, si  $\rho = 1$ , alors

$$x_t = x_0 + \sum_{i=1}^t \varepsilon_i, \quad (4.13)$$

et donc

$$\text{Var}(x_t) = \sum_{i=1}^t \text{Var}(\varepsilon_i) = t\sigma^2. \quad (4.14)$$

La variance de la série dépend du temps, ce qui contredit la définition de stationnarité d'une série temporelle. Une racine unitaire dans une série est donc synonyme de non-stationnarité.

Toujours dans le cas d'un modèle autorégressif, la série sera stationnaire si toutes les racines se trouvent à l'intérieur du cercle unité, c'est-à-dire si  $|\rho| < 1$ . Dans le cas contraire, si  $|\rho| > 1$ , on se retrouve dans le cas explosif, qui à l'instar d'une racine unitaire est lui aussi non stationnaire.

#### Test de Dickey-Fuller

Dickey et Fuller proposent dans [Dickey 1979] un test qui porte leur nom et que nous abrègerons **DF**. Ce test cherche à vérifier la présence d'une racine unitaire au sein d'une série temporelle. Pour cela, ils proposent de modéliser une série par un modèle autorégressif d'ordre 1 :

$$x_t = \rho x_{t-1} + \varepsilon_t, \quad (4.15)$$

avec  $\rho \in \mathbb{R}$  et  $\varepsilon_t$  un bruit blanc et gaussien. La régression du modèle est obtenue en soustrayant  $x_{t-1}$  aux deux membres de (4.15) :

$$\Delta x_t = (\rho - 1)x_{t-1} + \varepsilon_t, \quad (4.16)$$

où  $\Delta$  est l'opérateur des différences premières. Le test se déroule alors en deux étapes : la première consiste à réaliser l'estimation de la valeur du paramètre  $\rho$  ; la seconde consiste à tester si la pente de la régression (4.16) est nulle ou pas, en se basant sur le test de Student détaillé dans la section 4.2.2.1. Toutefois, la statistique du test dérivée de la régression (4.16) ne suit pas une distribution de Student, mais une autre distribution dont la caractérisation représente une contribution majeure de Dickey et Fuller [Dickey 1979].

Il existe deux autres versions de ce test qui complètent le modèle (4.16), que nous appellerons par la suite modèle sans terme déterministe. La seconde version du test ajoute une constante  $\mu \in \mathbb{R}$  au modèle :

$$\Delta x_t = (\rho - 1)x_{t-1} + \mu + \varepsilon_t. \quad (4.17)$$

La troisième version du test ajoute, en plus de la constante, une dérive temporelle linéaire déterministe  $\delta t$  (avec  $\delta \in \mathbb{R}$ ). Le modèle de la troisième version du test s'écrit alors :

$$\Delta x_t = (\rho - 1)x_{t-1} + \mu + \delta t + \varepsilon_t. \quad (4.18)$$

La statistique est différente pour les trois versions du test. Le choix du modèle va alors avoir un impact sur la puissance du test, c'est-à-dire sa capacité à rejeter l'hypothèse nulle quand elle est fautive. Nous ne considérerons par la suite que le modèle (4.17), à savoir celui avec une constante. Les considérations qui nous ont amenés à faire ce choix sont les suivantes :

- le modèle sans terme déterministe est très restrictif. En effet, dans le cas où la série serait stationnaire ( $|\rho| < 1$ ), elle tendrait plus ou moins rapidement vers 0 (en fonction de la valeur de  $\rho$ ). Or la plupart des séries générées possèdent une moyenne  $\mu > 0$ .
- le modèle avec dérive temporelle cherche à vérifier si la racine unitaire se trouve dans les variations autour de la tendance. Cependant, la simple présence de la tendance est déjà une source de non-stationnarité et devrait être détectée par un autre test, celui de la pente nulle décrit précédemment.

En résumé, les hypothèses du test sont les suivantes :

- **hypothèse nulle** : une racine unitaire est présente dans la série temporelle ;
- **hypothèse alternative** : la série temporelle est stationnaire au sens large ( $|\rho| < 1$ ).

Le test DF ne fait pas partie des six tests que nous allons utiliser par la suite. En effet, une condition de bon fonctionnement du test est le respect de l'hypothèse de blancheur du bruit. Si le bruit est auto-corrélé, la blancheur n'est plus respectée. Pour être sûr de respecter cette hypothèse, des modifications du test DF ont été proposées dans la littérature. Une première possibilité consiste à modifier la régression du test pour assurer la blancheur du bruit. Cette proposition porte le nom de test augmenté de Dickey-Fuller. La deuxième possibilité consiste à modifier la statistique du test comme le proposent Phillips et Perron. Ces deux modifications sont présentées ci-après.

### ADF — Dickey-Fuller augmenté

Le test augmenté de Dickey-Fuller [MacKinnon 1994, Greene 2003], abrégé **ADF**, propose une nouvelle régression permettant de s'assurer de l'indépendance du bruit dans le modèle. Elle consiste à remplacer le modèle autorégressif d'ordre 1 par un modèle ARIMA (de l'anglais *AutoRegressive Integrated Moving Average*). L'ordre de la partie autorégressive reste à 1, celui de la partie intégrée ira jusqu'à l'ordre  $d$  et la partie à moyenne mobile ne sera pas considérée (ordre 0). Pour la version sans dérive temporelle, la régression devient alors :

$$\Delta x_t = (\rho - 1)x_{t-1} + \left( \sum_{i=1}^d \alpha_i \Delta x_{t-i} \right) + \varepsilon_t. \quad (4.19)$$

Tout comme pour le test **DF**, les deux autres versions du test s'obtiennent en ajoutant une constante  $\mu$  à la régression, ou la somme d'une constante et d'une dérive :  $\mu + \delta t$ . Parmi les trois versions du test, nous utiliserons toujours, pour les considérations déjà énoncées, la version avec constante.

Ce modèle introduit un nouveau paramètre : l'ordre  $d$ . Il est nécessaire d'estimer la valeur de  $d$  avant de procéder à l'estimation de la valeur de  $\rho$ . Pour cela, la littérature suggère d'utiliser le critère d'information d'Akaike [Akaike 1974]. Une fois l'ordre fixé, le test **ADF** se poursuit comme un test **DF** classique.

Les hypothèses du test **ADF** sont les mêmes que celles du test **DF** :

- **hypothèse nulle** : une racine unitaire est présente dans la série temporelle ;
- **hypothèse alternative** : la série temporelle est stationnaire au sens large ( $|\rho| < 1$ ).

Tout comme pour le test **DF** le bruit  $\varepsilon_t$  dans le modèle doit être blanc et décorrélé. Le respect de cette hypothèse est rendu possible dans un nombre de cas plus grand pour le test **ADF** que pour le test **DF** grâce au modèle intégré d'ordre  $d$ .

### PP — Phillips–Perron

Le test PP, pour Phillips–Perron [Phillips 1988, Davidson 2004], cherche lui aussi à déterminer si une racine unitaire est présente dans la série temporelle. Il est basé sur le modèle autorégressif d'ordre 1 décrit par (4.16). Tout comme le test **ADF**, le test **PP** s'intéresse au fait que les termes de la série peuvent être davantage corrélés que ce que le modèle (4.16) autorise. Plutôt que d'introduire un nouveau modèle avec un paramètre supplémentaire comme pour le test **ADF**, le test **PP** garde le modèle (4.16) et propose une modification non paramétrique de la statistique qui rend le test robuste à l'autocorrélation qui pourrait être présente dans le terme d'erreur  $\varepsilon_t$ .

Trois versions du test sont construites d'après les modèles (4.16), (4.17) et (4.18). Il est à noter que si les équations des modèles sont similaires, les termes d'erreur  $\varepsilon_t$  peuvent ici être autocorrélés. Pour les mêmes raisons que le test **DF**, nous n'utiliserons que le modèle (4.17) avec constante.

Les hypothèses du test **PP** sont les mêmes que celles du test **DF** :

- **hypothèse nulle** : une racine unitaire est présente dans la série temporelle ;
- **hypothèse alternative** : la série temporelle est stationnaire au sens large ( $|\rho| < 1$ ).

### DF-GLS — Dickey-Fuller — Generalized Least Square

Le test **DF-GLS**, pour *Dickey Fuller — Generalized Least Square* [Elliott 1996], diffère d'un test **ADF** classique à cause de l'utilisation d'un pré-traitement qui consiste à retirer la dérive temporelle et la moyenne de la série avant de lui appliquer un test **ADF**. L'estimation de la dérive temporelle et de la moyenne est réalisée à l'aide de la méthode des moindres carrés généralisés (*Generalized Least Square* en anglais).

L'intérêt du test **DF-GLS** est d'être plus puissant qu'un test **ADF** classique. Autrement dit, la probabilité de rejeter l'hypothèse nulle quand elle est fautive sera plus grande avec le test **DF-GLS** qu'avec le test **ADF**.

Les hypothèses du test **DF-GLS** sont les mêmes que celles du test **ADF** :

- **hypothèse nulle** : une racine unitaire est présente dans la série temporelle ;
- **hypothèse alternative** : la série temporelle est stationnaire au sens large ( $|\rho| < 1$ ).

#### 4.2.2.3 KPSS — Test de l'hypothèse de stationnarité

Le test de Kwiatkowski–Phillips–Schmidt–Shin (**KPSS**) [Shin 1992] est un test statistique ayant pour hypothèse nulle la stationnarité de la série temporelle. Il a été construit en réponse aux tests plus classiques de stationnarité qui ont pour hypothèse nulle la présence d'une racine unitaire. Selon les auteurs du test **KPSS**, les tests **ADF** et **PP** manquent de puissance et l'hypothèse nulle est rarement rejetée alors qu'elle devrait l'être.

Deux modèles de série temporelle sont proposés dans [Shin 1992]. Le premier, le plus simple, et celui auquel nous allons nous intéresser, s'écrit sous la forme :

$$x_t = r_t + \varepsilon_t, \quad (4.20)$$

avec  $\varepsilon_t$  une erreur stationnaire (typiquement, un bruit blanc) et  $r_t = r_{t-1} + u_t$  une racine unitaire avec  $u_t$  un bruit centré indépendant et identiquement distribué de variance  $\sigma_u^2$ . Pour qu'une série suivant ce modèle soit stationnaire, il faut que la variance  $\sigma_u^2$  soit nulle, ce qui transformerait le terme  $r_t$  en une constante égale à  $r_0$ . Les auteurs proposent alors une statistique et la distribution asymptotique associée avant de tester l'hypothèse  $\sigma_u^2 = 0$ .

La seconde forme du modèle d'une série temporelle fait intervenir une dérive temporelle dans la série. Le modèle peut s'écrire :

$$x_t = \delta t + r_t + \varepsilon_t. \quad (4.21)$$

Cependant, de manière similaire au test **DF**, la présence d'une dérive temporelle est déjà un signe de non-stationnarité. Vérifier uniquement si les variations autour de cette tendance sont stationnaires ne nous aidera pas à conclure de manière générale sur la série. C'est pourquoi nous nous intéresserons uniquement au modèle (4.20).

Les hypothèses du test **KPSS** sont inversées par rapport à ce qu'on a eu l'occasion de voir pour les tests vérifiant la présence d'une racine unitaire :

- **hypothèse nulle** : la série est stationnaire au sens large ( $\sigma_u^2 = 0$ ) ;
- **hypothèse alternative** : la série n'est pas stationnaire ( $\sigma_u^2 > 0$ ).

### 4.2.3 Utilisation des tests statistiques

Les tests statistiques ont été choisis dans le but de déterminer si les tendances présentent une évolution ou non. Il s'agit de prendre une décision binaire sur la classe de la tendance. La manière dont est prise cette décision pour chaque test est expliquée dans la section suivante. Nous verrons ensuite les résultats de ces tests sur une base de données expertisée.

#### 4.2.3.1 Décision binaire sur la stationnarité de la tendance

Un test statistique se compose généralement de deux hypothèses  $\mathcal{H}_0$  et  $\mathcal{H}_1$  qu'on souhaite tester et d'une statistique de test qui permet de décider de l'acceptation ou du rejet de l'hypothèse nulle. Ces éléments ont été précisés dans la description de chacun des tests de la section précédente. L'acceptation ou le rejet de l'hypothèse nulle s'accompagne de deux risques d'erreur [Saporta 2011] :

- le risque de 1<sup>ère</sup> espèce, égal à la probabilité de rejeter  $\mathcal{H}_0$  alors que  $\mathcal{H}_0$  est vraie :

$$p = P[\text{rejeter } \mathcal{H}_0 \mid \mathcal{H}_0 \text{ vraie}] ; \quad (4.22)$$

- le risque de 2<sup>ème</sup> espèce, égal à la probabilité d'accepter  $\mathcal{H}_0$  alors que  $\mathcal{H}_0$  est fautive :

$$\beta = P[\text{accepter } \mathcal{H}_0 \mid \mathcal{H}_0 \text{ fautive}] . \quad (4.23)$$

Pour être complet, un résultat de test doit comporter les valeurs de  $p$  et de  $\beta$  (ou  $\pi = 1 - \beta$  qui représente la puissance du test). Toutefois, pour les tests présentés dans la section précédente, l'hypothèse alternative  $\mathcal{H}_1$  n'est jamais complètement spécifiée et ne permet pas le calcul de  $\beta$ . C'est pourquoi nous avons fixé un seuil de risque de 1<sup>ère</sup> espèce à 5% (valeur classiquement utilisée). Si la valeur de  $p$  est inférieure à ce seuil, cela signifie que le risque de 2<sup>ème</sup> espèce  $\beta$  augmente et donc la puissance du test diminue. Dans ce cas, l'hypothèse nulle  $\mathcal{H}_0$  est rejetée. Dans le cas contraire, elle est acceptée. L'ensemble des décisions prises pour chaque test est résumé dans le tableau 4.1.

Il est important de rappeler que ces décisions sont prises en considérant un seuil de risque d'erreur de 1<sup>ère</sup> espèce de 5%. C'est pourquoi nous avons choisi de ne pas utiliser un seul test

Méthode	$\mathcal{H}_0$ $p \geq 5\%$	$\mathcal{H}_1$ $p < 5\%$
Pente nulle	condition nécessaire à la stationnarité respectée	<b>non stationnaire</b>
Chow	condition nécessaire à la stationnarité respectée	<b>non stationnaire</b>
ADF	<b>non stationnaire</b>	stationnaire au sens large
DF-GLS	<b>non stationnaire</b>	stationnaire au sens large
PP	<b>non stationnaire</b>	stationnaire au sens large
KPSS	stationnaire au sens large	<b>non stationnaire</b>

Tableau 4.1 – Décisions prises pour chaque test statistique.

statistique mais de fusionner les résultats de ces différents tests. La combinaison des tests sera proposée dans la section 4.2.4.

Toutefois, vouloir fusionner ces tests nécessite que les décisions d'acceptation ou de rejet de l'hypothèse nulle  $\mathcal{H}_0$  s'expriment de la même façon pour chaque test. Or le tableau 4.1 met bien en évidence que si tous les tests comportent la décision « non stationnaire », la décision contraire varie entre « stationnaire au sens large » et « condition nécessaire à la stationnarité respectée ». Dans la mesure où la conclusion que nous souhaitons détecter est « non stationnaire », nous allons étiqueter la décision contraire « stationnaire » dans le sens où aucune non-stationnarité n'a été détectée par le test.

Ainsi, chaque test nous permet de classer les tendances en deux groupes : les tendances non stationnaires et celles qui ne sont pas vues comme telles par le test. En réalité, un troisième cas est possible : les tests peuvent échouer si la tendance considérée est trop courte. C'est particulièrement le cas des tests PP, ADF et DF-GLS pour lesquels les valeurs critiques de la statistique n'ont pas été calculées pour des tendances de quelques points seulement. Lorsque l'un de ces tests échoue, aucune décision n'est prise et le résultat du test est alors indéfini.

Tous les tests présentés sont appliqués sur un vecteur  $X$  de valeurs réelles :

$$X = [x_1, \dots, x_m], \quad (4.24)$$

où  $\forall i \in \{1, \dots, m\}, x_i \in \mathbb{R}$ . Dans le cas où une tendance est en sommeil pendant un ou plusieurs instants, cela reviendrait à ne pas définir certaines valeurs du vecteur. Les tests ne sont pas conçus pour prendre ce type de cas en compte. Nous allons alors réduire toutes les tendances avant de leur faire passer les tests statistiques. Par réduction, nous entendons simplement le fait de supprimer dans chaque tendance les instants où elle est en sommeil. Ainsi, nous récupérerons un vecteur, certes plus court que le vecteur original, mais pour lequel toutes les valeurs sont définies.

### 4.2.3.2 Base de données expertisée

Afin de vérifier l'efficacité des tests statistiques, nous avons expertisé une base de données de tendances. Chaque tendance a été étiquetée à la main comme étant stationnaire ou porteuse d'un changement. Cette base de données a été construite à partir d'une sélection des tendances générées et analysées dans la section 4.1.2. Les tendances proviennent donc de plusieurs séquences différentes ; la base de données mélange les tendances provenant à la fois du défaut simulé, des défauts sur le banc de test éolien et des signaux réels de l'éolienne surveillée. La base contient une sélection de 500 tendances, dont 83 sont étiquetées comme faisant état d'un changement.

Chacun des 6 tests statistiques est utilisé sur l'ensemble des 500 tendances expertisées. Si le test juge qu'une tendance est non stationnaire alors qu'elle a été étiquetée comme faisant état d'un changement, une bonne détection est faite ( $TP$  – *True Positive*). Au contraire, si la tendance étiquetée est jugée stationnaire par le test, nous sommes en présence d'un faux négatif ( $FN$  – *False Negative*). De manière symétrique, nous avons également les fausses détections ( $FP$  – *False Positive* ; tendance jugée non stationnaire mais étiquetée stable) et les bonnes non-détections ( $TN$  – *True Negative* ; tendance jugée stationnaire et étiquetée stable). Le nombre  $UN$  (*Undefined*) correspond au nombre de fois où le test a échoué et où le résultat est indéfini.

Les résultats sont complétés par 3 mesures : le pourcentage de bonnes détections, le pourcentage de fausses alarmes et enfin une distance par rapport à un classifieur idéal. Le pourcentage de bonnes détections ( $TPR$  – *True Positive Rate*) est calculé comme le ratio entre le nombre de bonnes détections  $TP$  et le nombre  $P$  de tendances étiquetées manuellement comme porteuses d'une évolution ( $P = 83$  dans la base utilisée) :

$$TPR = 100 \frac{TP}{P}. \quad (4.25)$$

Le pourcentage de fausses alarmes ( $FPR$  – *False Positive Rate*) est calculé comme le ratio entre le nombre de fausses détections  $FP$  et le nombre  $N$  de tendances étiquetées comme stationnaires ( $N = 417$  dans la base utilisée) :

$$FPR = 100 \frac{FP}{N}. \quad (4.26)$$

Enfin, une distance permet de comparer le test au classifieur idéal :

$$d = \left(1 - \frac{TP}{P}\right)^2 + \left(\frac{FP}{N}\right)^2. \quad (4.27)$$

Les résultats obtenus par les 6 tests sont présentés dans le tableau 4.2. Dans chaque colonne, les valeurs en gras correspondent aux meilleurs résultats obtenus relatifs à cette colonne.

Deux des tests se détachent parmi les 6 : il s'agit du test de Chow qui a le taux de fausses

Méthode	TP	TN	FP	FN	UN	% bonnes détections	% fausses alarmes	Distance
KPSS	40	374	43	43	<b>0</b>	48.19	10.31	0.279
ADF	73	278	107	7	35	87.95	25.65	0.080
DF-GLS	74	191	194	6	35	89.15	46.52	0.228
PP	59	381	33	24	3	71.08	7.91	0.089
Chow	29	<b>385</b>	<b>32</b>	54	<b>0</b>	34.93	<b>7.67</b>	0.429
Pente nulle	<b>81</b>	352	65	<b>2</b>	<b>0</b>	<b>97.59</b>	15.58	<b>0.024</b>
Expertise	83	417						

Tableau 4.2 – Résultats de la classification pour chaque test statistique.

alarmes le plus bas (7,67%), mais qui ne détecte que 29 des tendances changeantes sur les 83. À l'inverse, le test de la pente nulle permet de retrouver quasiment toutes les tendances changeantes (81 sur 83), mais réalise de nombreuses fausses alarmes (65, soit 15,5%). Nous proposons à présent de fusionner les résultats des tests afin de tirer parti des avantages de chacun des tests statistiques.

#### 4.2.4 Combinaison des tests

La fusion des résultats des différents tests pour une tendance est réalisée sous la forme d'un vote. Si le nombre de votants pour la non-stationnarité est supérieur ou égal au seuil fixé, la tendance est considérée comme non stationnaire. Dans le cas contraire, la tendance est considérée comme non porteuse d'une information de défaut. Cela signifie que les résultats indéfinis pour les tendances trop courtes sont considérés comme de l'abstention.

Toutefois, il appartient de déterminer si les 6 tests sont pertinents et doivent tous être utilisés ou si seulement un sous-ensemble de ces tests nous permet de conclure efficacement. De plus, la valeur du seuil à atteindre lors du vote doit être fixée. C'est pourquoi nous proposons de tester toutes les combinaisons de vote possibles. Parmi l'ensemble des 6 tests, nous allons donc considérer tous les sous-ensembles que nous pouvons former. Pour chaque sous-ensemble, nous allons faire varier le seuil à atteindre pour décider de la non-stationnarité. Ce seuil peut prendre toutes les valeurs de 1 au nombre de tests présents dans le sous-ensemble. Le nombre de combinaisons des sous-ensembles et des valeurs du seuil peut se calculer ainsi :

$$\binom{6}{1} \times 1 + \binom{6}{2} \times 2 + \binom{6}{3} \times 3 + \binom{6}{4} \times 4 + \binom{6}{5} \times 5 + \binom{6}{6} \times 6 = 192, \quad (4.28)$$

où le coefficient  $\binom{6}{k} = \frac{6!}{k!(6-k)!}$  correspond au nombre de sous-ensembles de  $k$  tests qu'il est possible de faire et où les coefficients de pondération correspondent au nombre de seuils. Pour chacune des combinaisons possibles, nous comptons les bonnes détections ( $TP$ ), les bonnes non-détections ( $TN$ ), les fausses alarmes ( $FP$ ) et les faux négatifs ( $FN$ ). Le résumé des résultats pour les 20 meilleures méthodes est donné dans le tableau 4.3. Les meilleures

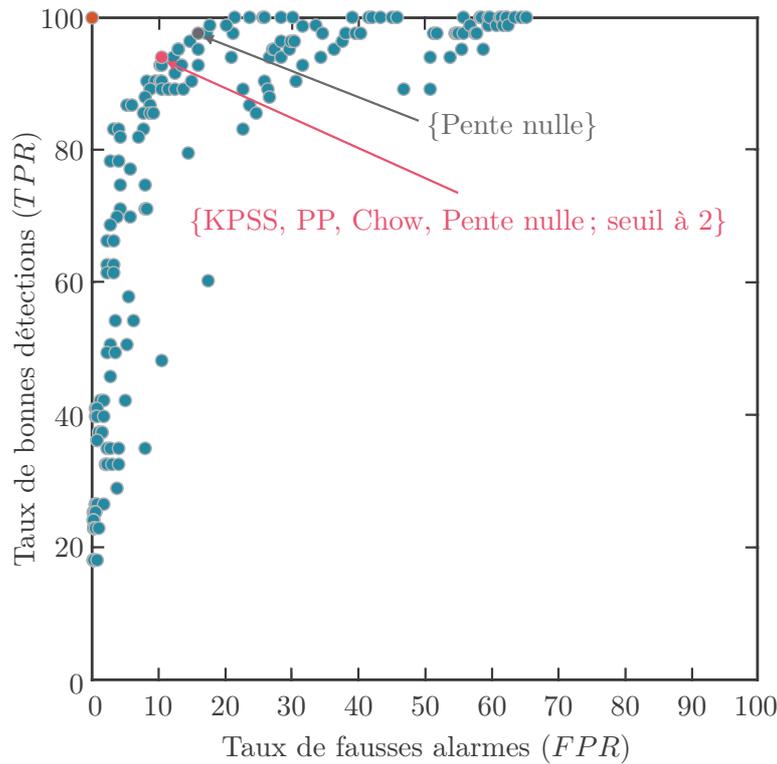


FIGURE 4.13 – Répartition des combinaisons de tests (cercles bleus) sur le plan défini par les taux de bonnes détections et de fausses alarmes. Le point en haut à gauche, en orange, correspond au classifieur idéal. Le point en gris représente le meilleur test seul (relativement à la distance  $d$  définie), alors que le point en rose représente la meilleure combinaison de tests.

méthodes sont considérées comme celles minimisant la distance  $d$  au classifieur idéal (voir (4.27)). La figure 4.13 permet d'ailleurs d'apprécier la répartition de toutes les combinaisons sur le plan défini par le taux de fausses alarmes et le taux de bonnes détections. Sur ce plan, le point situé en haut à gauche (en orange) correspond au classifieur idéal. Les points bleus correspondent aux différentes combinaisons de fusion des tests.

La fusion des tests permet d'arriver à des classifieurs qui sont plus proches du classifieur idéal que ne l'était chaque test statistique pris indépendamment. Le meilleur résultat obtenu est celui correspondant à la fusion des 4 tests suivants : {KPSS, PP, Chow, Pente nulle} avec un seuil fixé à 2. Il permet de détecter 78 des 83 tendances changeantes. La figure 4.14 permet d'observer l'allure d'une sélection de trois tendances que cette combinaison de tests a classées comme non stationnaires. Effectivement, on observe une croissance dans les tendances qui a été correctement détectée par les tests. Trois tendances classées comme stationnaires sont également présentées sur la figure 4.14. La conclusion « stationnaire » doit être ici interprétée comme le fait qu'aucune non-stationnarité n'a été détectée par les tests.

La figure 4.13 nous indique qu'il est possible de créer des combinaisons permettant de détecter toutes les tendances changeantes ( $TPR = 100\%$ ), mais cela est fait au prix d'un taux de fausses alarmes très élevé ( $FPR > 20\%$ ). Sachant que le nombre de tendances générées pour

Méthode	Seuil	TP	TN	FP	FN	% bonnes dét.	% fausses alarmes	Distance
KPSS, PP, Chow, Pente nulle	2	78	374	43	5	93.97	10.31	<b>0.0142</b>
KPSS, ADF, DF-GLS, PP, Pente nulle	3	77	375	42	6	92.77	10.07	0.0153
KPSS, ADF, Pente nulle	2	77	374	43	6	92.77	10.31	0.0158
KPSS, PP, Pente nulle	2	75	383	34	8	90.36	8.15	0.0159
ADF, PP, Pente nulle	2	75	379	38	8	90.36	9.11	0.0175
KPSS, ADF, DF-GLS, PP, Chow, Pente nulle	3	78	367	50	5	93.97	11.99	0.0180
DF-GLS, PP, Pente nulle	2	79	364	53	4	95.18	12.70	0.0184
KPSS, ADF, Chow, Pente nulle	2	79	364	53	4	95.18	12.70	0.0184
ADF, DF-GLS, PP, Pente nulle	3	74	382	35	9	89.15	8.39	0.0188
ADF, Chow, Pente nulle	2	75	376	41	8	90.36	9.83	0.0189
ADF, DF-GLS, PP, Chow, Pente nulle	3	75	375	42	8	90.36	10.07	0.0194
KPSS, ADF, PP, Chow, Pente nulle	3	72	396	21	11	86.74	5.03	0.0201
ADF, Pente nulle	2	73	385	32	10	87.95	7.67	0.0204
KPSS, DF-GLS, PP, Chow, Pente nulle	3	72	392	25	11	86.74	5.99	0.0211
ADF, PP, Chow, Pente nulle	2	76	367	50	7	91.56	11.99	0.0214
KPSS, ADF, PP, Pente nulle	2	80	357	60	3	96.38	14.38	0.0220
KPSS, ADF, DF-GLS, Chow, Pente nulle	3	74	374	43	9	89.15	10.31	0.0223
KPSS, DF-GLS, Pente nulle	2	77	362	55	6	92.77	13.18	0.0226
KPSS, ADF, PP, Chow	2	74	371	46	9	89.15	11.03	0.0239
DF-GLS, Pente nulle	2	74	370	47	9	89.15	11.27	0.0244

Tableau 4.3 – Résultats de la classification pour les 20 meilleures combinaisons de tests statistiques et de seuils.

une séquence peut être très grand (des centaines de milliers), un fort taux de fausses alarmes va générer des milliers de faux positifs. Il peut alors être intéressant d'utiliser une combinaison ayant un taux de fausses alarmes proches de 0. Cependant, ces combinaisons ont un taux de bonnes détections très bas. Il faut alors espérer que le défaut du système transparaisse dans plusieurs tendances et que l'une d'entre elles au moins soit détectée comme non stationnaire.

La méthode de fusion proposée ici est très simple et très rapide à mettre en œuvre. Elle présente cependant un désavantage. En effet, lors de la prise de décision de chaque test, la valeur  $p$  permet de trancher sur la classe de la tendance. Quand la valeur  $p$  est loin du seuil fixé à 5%, la décision prise est claire et on peut lui accorder une confiance relativement importante. Au contraire, quand la valeur  $p$  est proche du seuil, la confiance accordée à la décision est plus faible. Il est fort dommage de perdre cette information sur la confiance accordée à la décision lors de la classification de la tendance. C'est pourquoi, il pourrait être intéressant de faire une méta-analyse sur nos tests. Une méta-analyse, comme expliquée dans [Whitlock 2005] et [Owen 2009], consiste à fusionner les résultats de plusieurs tests indépendants cherchant tous à vérifier la même hypothèse. Les valeurs  $p$  de chacun des tests sont utilisées afin d'obtenir une valeur  $p$  globale. La décision binaire est donc prise au dernier moment et utilise toutes les nuances de confiance que les tests précédents peuvent fournir.

### 4.3 Synthèse du chapitre

Au cours de ce dernier chapitre, nous avons vu de nombreux exemples d'application de la méthode proposée pour la surveillance de systèmes mécaniques. La majorité des applications a été faite sur des signaux réels et a permis de montrer que la méthodologie proposée est capable de générer des tendances reflétant l'état de santé du système surveillé. Ces tendances font partie d'une longue liste qu'il est nécessaire de classer pour faire un pas en avant vers le diagnostic automatique du système.

L'analyse des tendances est un problème complexe, du fait de leur hétérogénéité et de leur support temporel souvent très court. Nous avons alors proposé une méthode de classification basée sur l'utilisation conjointe de plusieurs tests statistiques visant à déterminer si la tendance est non stationnaire. La non-stationnarité est ici recherchée car elle est le reflet d'une évolution dans la tendance, c'est-à-dire une évolution de la santé du système.

La partie applicative de ce chapitre a donné lieu à deux publications dans des conférences internationales portant sur la surveillance de système [Gerber 2014, Li 2015], ainsi qu'à une soumission dans un journal [Firla 2015]. La seconde partie du chapitre est plus exploratoire et mérite qu'on s'intéresse aux pistes d'améliorations proposées afin d'arriver à un diagnostic automatique du système surveillé.

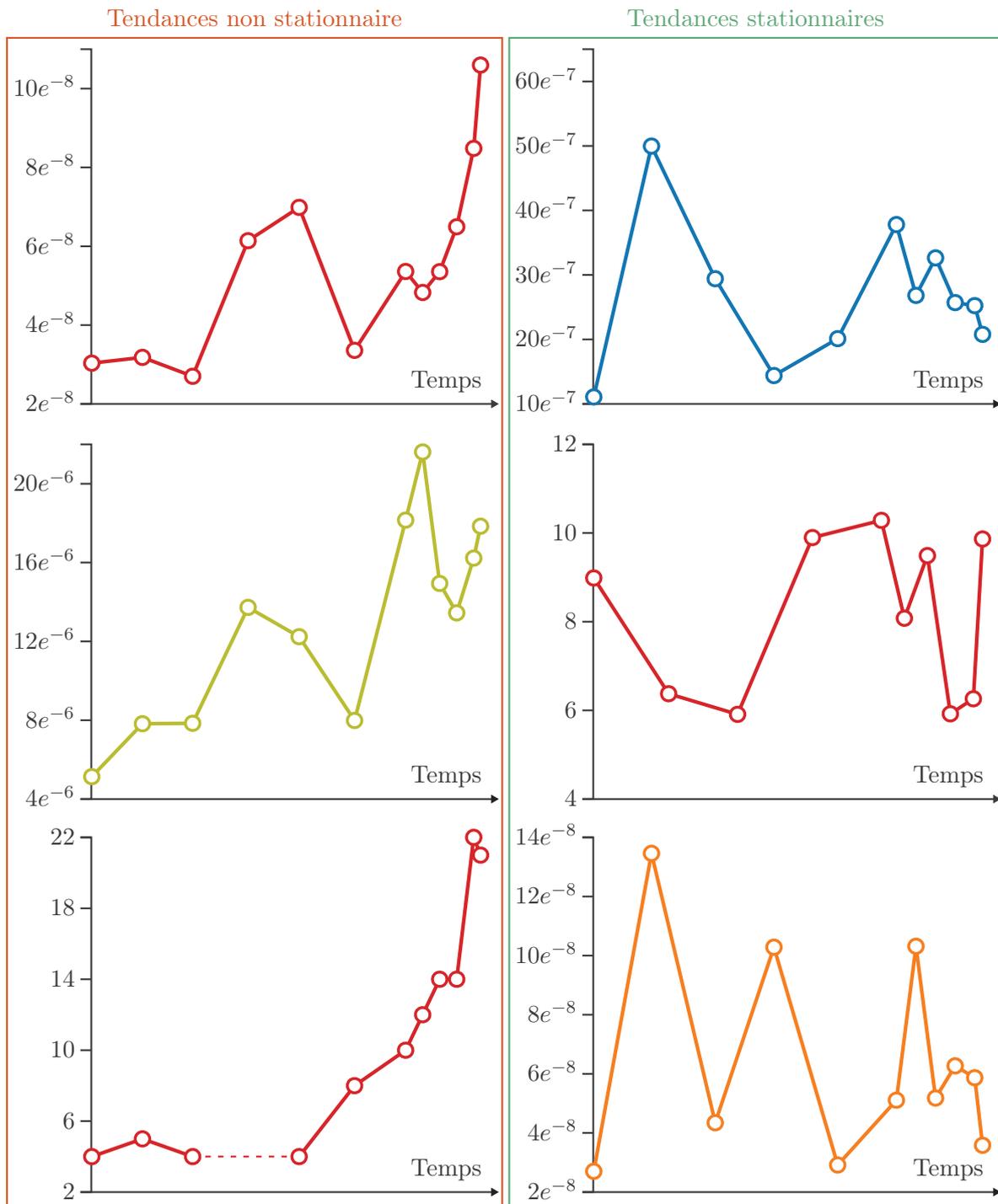


FIGURE 4.14 – Allures de certaines tendances classées par la combinaison de test {KPSS, PP, Chow, Pente nulle}. Les tendances dans la colonne de gauche ont été classées comme non stationnaires alors que celles de la colonne de droite ont été classées comme stationnaires.

# Conclusions et perspectives

Le principal objectif de cette thèse était de participer au développement d'une méthode de surveillance automatique. Le contexte de ces travaux était particulièrement bien défini dans la mesure où la méthode devait se baser sur des travaux antérieurs et devait répondre à toutes les exigences de deux projets européens, [*KAStrion*] et [*SUPREME*]. Parmi les contraintes les plus fortes, nous retiendrons en particulier la nécessité de développer une méthode générique, capable de s'appliquer à n'importe quel système. Elle devait également avoir une complexité algorithmique la plus faible possible afin d'effectuer tous les traitements sur un système embarqué. En dernier lieu, la méthode devait fonctionner « en ligne », c'est-à-dire être capable de prendre un signal nouvellement acquis et ajouté à la séquence de signaux déjà traitée.

La découverte en détails du contexte de ces travaux de thèse a été notre point de départ. En particulier, les travaux antérieurs sur l'analyse spectrale d'un signal ont été expliqués. Ils ont donné naissance à un analyseur de spectre automatique, nommé AStrion, capable d'estimer, sans aucun *a priori* ou paramétrage, l'ensemble des pics présents dans la densité spectrale d'un signal. Cet outil permet d'analyser un signal avec une granularité très fine. L'obtention d'un diagnostic du système surveillé à partir de cette description fine des signaux a posé trois problématiques principales : comment interpréter l'ensemble des pics identifiés dans la densité spectrale d'un signal ? Comment suivre ces informations spectrales de signal en signal ? Comment réaliser le diagnostic du système à partir de ce suivi ?

Une fois le contexte mis en place, nous nous sommes attaqués à la problématique de l'interprétation du spectre d'un signal. Cette interprétation consiste à lire le spectre et à déterminer les différents éléments qui le constituent. La détection des pics réalisée grâce aux travaux antérieurs propose une granularité trop fine et ne permet pas d'identifier les structures complexes présentes dans le spectre. Ces structures prennent deux formes différentes, à savoir la forme d'une série harmonique ou de bandes de modulation. Nous avons alors proposé une méthode capable de reconstituer les séries harmoniques à partir des pics détectés, en prenant en compte les erreurs d'estimation quant à la position fréquentielle exacte des pics. Cette méthode d'identification des séries harmoniques a ensuite été réutilisée sur des sous-ensembles de pics spécialement définis afin de procéder à l'identification des bandes de modulation. Une caractérisation des structures ainsi identifiées a également été proposée. L'ensemble de ces traitements forme un module qui a été ajouté au cœur d'AStrion et a subi un transfert industriel.

L'analyse spectrale et l'interprétation du spectre permettent de décrire l'état du système à la date d'acquisition du signal analysé. Il est intéressant de comparer cette description à celle d'un signal pour lequel le système est sain, ou à des signaux acquis antérieurement afin de visualiser l'évolution de la santé du système. Cette comparaison demande à associer les structures spectrales identiques d'un signal à l'autre. Pour cela, nous avons proposé des améliorations à une méthode de suivi de pics dans le cadre des signaux de paroles afin de l'adapter au contexte de la surveillance automatique. La méthode a également été adaptée

afin de permettre le suivi des séries harmoniques et des bandes de modulation. Ce suivi est basé sur l'utilisation des fréquences caractéristiques de chacune des structures et permet de créer des trajectoires spectrales qui nous autorisent à décrire la séquence de signaux dans sa globalité. Le suivi a également donné naissance à un module qui a été ajouté à l'analyseur de spectre AStrion et qui a été transféré à un partenaire industriel.

La naissance ou la mort d'une trajectoire est en général un premier symptôme de l'évolution de la santé du système. Afin d'obtenir une visualisation plus précise de cette évolution, nous avons proposé de dériver des tendances à partir des caractéristiques des structures spectrales suivies dans une trajectoire. Des études de cas sur des signaux réels issus de la surveillance de système mécanique ont permis de valider la pertinence de l'approche proposée en démontrant son potentiel de diagnostic. Cependant, le nombre de tendances calculées est grand (plusieurs dizaines, voire centaines de milliers) et un tri est nécessaire afin de ne faire ressortir que celles qui sont liées à une évolution du système surveillé. Un travail exploratoire a débouché sur la proposition d'une méthode de classification basée sur la fusion de plusieurs tests statistiques. Chaque test cherche à déterminer si la tendance est porteuse d'une évolution, ou si au contraire, elle est stable.

Ces derniers travaux sont très exploratoires et méritent d'être approfondis. En effet, la méthode de fusion des tests utilisée est très simple. Elle demande à chaque test statistique de prendre une décision binaire avant de procéder à un vote pour prendre une décision finale. Cependant, cette démarche ne prend pas en compte le degré de confiance que chaque test accorde à sa décision. Il serait intéressant d'utiliser d'autres outils prenant ce degré de confiance en compte pour procéder à la fusion des tests. Nous pouvons citer par exemple les fonctions de croyance ou la meta-analyse, c'est-à-dire l'utilisation conjointe de la valeur  $p$  de chaque test, comme piste d'exploration.

Toujours sur le thème de la fusion, une autre perspective consisterait à utiliser conjointement les diagnostics provenant de différents capteurs. En effet, il n'est pas rare lors de la surveillance d'un système mécanique d'avoir plusieurs capteurs du même type. Par exemple, trois accéléromètres dans trois directions différentes ou encore trois capteurs de courant ou de tension (par exemple, un par phase). Chaque capteur acquiert alors sa propre séquence de signaux et les tendances générées par ces différents capteurs se rapportent aux mêmes phénomènes physiques. La fusion des résultats des différents capteurs permettrait d'accorder un poids supplémentaire au diagnostic réalisé, ce qui simplifierait la prise de décision. De plus, cette fusion pourrait rendre la surveillance robuste face à un dysfonctionnement de l'un des capteurs.

Enfin, la méthode de suivi est basée sur une hypothèse relativement forte : l'état machine constant qui demande à ce que les paramètres opérationnels soient identiques d'une acquisition de signal à l'autre. Sur des systèmes comme les éoliennes, cette condition est difficile à respecter, dans la mesure où le vent en entrée du système est très changeant. La vitesse de rotation est donc fortement variable. Heureusement, il est possible via un ré-échantillonnage angulaire de normaliser cette vitesse d'entrée et de fixer les fréquences caractéristiques des structures spectrales présentes dans les signaux. Ainsi, le suivi temps-fréquence des structures est possible. Cependant, la vitesse de rotation n'est pas l'unique paramètre opérationnel d'une

---

éolienne ; la puissance générée dépend également de la charge du système. En fonction de la charge, les niveaux vibratoires changent également, ce qui peut se traduire par des variations de l'énergie au sein d'une trajectoire qui rendent plus difficile la détection précoce d'un défaut. Il serait intéressant de caractériser précisément l'influence de la charge sur le niveau vibratoire afin de proposer une normalisation qui, à l'instar du ré-échantillonnage angulaire pour les fréquences, harmoniserait l'énergie vibratoire au sein des tendances.



# Les différentes stratégies de maintenance

---

## Sommaire

<b>A.1 Maintenance corrective</b> . . . . .	<b>115</b>
A.1.1 Maintenance palliative . . . . .	116
A.1.2 Maintenance curative . . . . .	116
<b>A.2 Maintenance préventive</b> . . . . .	<b>116</b>
A.2.1 Maintenance systématique . . . . .	116
A.2.2 Maintenance conditionnelle . . . . .	117
A.2.3 Maintenance prévisionnelle . . . . .	117
<b>A.3 Comparaison des stratégies</b> . . . . .	<b>117</b>

---

Un système est en général construit dans un but précis. On attend de lui qu'il accomplisse une ou plusieurs tâches. Cependant, une panne peut survenir, empêchant le bon déroulement des opérations. L'ensemble des actions mises en œuvre pour rétablir ou maintenir le fonctionnement du système est désigné par le terme de maintenance. Il existe différentes stratégies de maintenance, qui se regroupent en deux grandes classes [Randall 2011, pages 2-3]. Il y a d'une part la *maintenance corrective* qui consiste à intervenir sur le système à partir du moment où ce dernier est en panne. D'autre part, il y a la *maintenance préventive* dont l'objectif est d'anticiper la panne et d'intervenir sur le système suffisamment tôt pour éviter l'arrêt inopiné du système [Monchy 2010, chapitre 2]. Ces deux classes se divisent à nouveau en plusieurs stratégies, comme expliqué dans les Sections A.1 et A.2.

## A.1 Maintenance corrective

La maintenance corrective ne s'anticipe pas. Elle consiste à intervenir uniquement à partir du moment où le système est défaillant et ne fonctionne plus comme il le devrait. Le temps d'arrêt du système peut alors être long, en fonction de la gravité du défaut, de la disponibilité d'une équipe de maintenance, de la disponibilité des pièces de rechange, etc. C'est d'ailleurs la longueur de ce temps d'arrêt, entraînant une perte de production, qui est responsable du coût élevé de ce type de maintenance. La maintenance corrective se subdivise en deux stratégies : la palliative et la curative.

### **A.1.1 Maintenance palliative**

La maintenance palliative consiste à dépanner le système et à rétablir l'une ou l'ensemble de ses fonctions au plus vite. Il s'agit donc d'une remise en état de marche provisoire et une telle opération doit être suivie d'une opération de maintenance curative, plus pérenne. Ce type d'intervention permet essentiellement de réduire le temps d'arrêt du système.

### **A.1.2 Maintenance curative**

La maintenance curative consiste à rétablir le système dans l'état précédant la panne, afin que le système puisse à nouveau accomplir l'ensemble des tâches pour lequel il est prévu. Cette réparation est plus durable que celle réalisée durant une maintenance palliative. Cette stratégie de maintenance s'accompagne souvent d'un stock de pièces de rechange afin de réduire le temps d'arrêt du système en cas de panne. En effet, il n'est alors pas nécessaire de commander la pièce et d'attendre qu'elle arrive pour effectuer le remplacement. Cependant, la gestion de ce stock engendre un coût supplémentaire.

## **A.2 Maintenance préventive**

La maintenance préventive consiste à anticiper et à intervenir sur le système avant qu'une panne ne se déclare. Ainsi, le temps d'arrêt de la machine correspond strictement au temps des opérations de maintenance. Il n'y a pas de temps d'arrêt supplémentaire, lié par exemple à la commande d'une pièce de rechange ou à l'attente de conditions favorables pour atteindre le système. Ce type de maintenance est donc extrêmement intéressant pour des systèmes difficilement accessibles, comme dans le cas des éoliennes en pleine mer. En effet, ces éoliennes ne sont accessibles que lorsque les conditions météorologiques le permettent.

La maintenance préventive se subdivise en trois stratégies : la maintenance systématique, la maintenance conditionnelle et la maintenance prévisionnelle.

### **A.2.1 Maintenance systématique**

La maintenance systématique consiste à effectuer certaines opérations de manière régulière, quel que soit l'état du système. Il peut s'agir d'un simple changement de lubrifiant ou du remplacement complet d'une partie du système. La période des opérations de maintenance est établie en fonction de la durée de vie des différentes parties du système. Les durées de vie sont en général fournies par le ou les constructeurs.

Ce type de maintenance permet de planifier les opérations longuement à l'avance et diminue fortement les taux de pannes sur le système. Cependant, cette stratégie coûte relativement chère dans la mesure où le remplacement des pièces est réalisé systématiquement, même si les

pièces sont encore en très bon état et pourraient endurer des milliers d'heures de fonctionnement supplémentaires. De plus, les interventions régulières sur le système peuvent introduire des défauts qui ne seraient pas apparus autrement.

### A.2.2 Maintenance conditionnelle

La maintenance conditionnelle cherche à ne réaliser les opérations de maintenance que lorsque c'est nécessaire. Pour cela, un relevé ou un dispositif de mesure doit permettre de juger de l'état du système. Lorsque l'état est jugé suffisamment critique, une décision volontaire déclenche les opérations de maintenance. Ainsi, les équipements fonctionnent quasiment jusqu'en fin de vie. Pour être efficace et éviter les pannes, l'opération de contrôle du système doit être au moins régulière, sinon continue.

### A.2.3 Maintenance prévisionnelle

La maintenance prévisionnelle est similaire à la maintenance conditionnelle, dans le sens où les opérations de maintenance ne sont déclenchées que lorsque cela est nécessaire. La principale distinction vient de l'instant où la décision est prise. En effet, la maintenance prévisionnelle consiste à surveiller le système en continu. Dès les premiers signes d'usure, une estimation de la durée de vie résiduelle permet de planifier au mieux les opérations de maintenance. Ainsi, la fréquence des interventions est grandement diminuée, en plus d'une diminution de la fréquence des pannes. Cette stratégie de maintenance est cependant plus délicate à mettre en place : il faut pouvoir identifier les signes précurseurs d'un défaut et estimer la durée de vie résiduelle des composants.

## A.3 Comparaison des stratégies

Le tableau A.1 résume les forces et les faiblesses de toutes les stratégies de maintenance. Les symboles € et ⊕ dénotent respectivement un coût et une période de temps. Plus le nombre de symboles est important, plus le coût ou le temps est important. Les performances des stratégies sont jugées comparativement les unes aux autres. Pour chaque critère, autrement dit pour chaque ligne du tableau, les stratégies les plus intéressantes sont en vert, alors que les moins intéressantes sont en rouge. Un niveau intermédiaire, en orange, vient compléter cette classification.

### Coût de la surveillance

Les maintenances correctives et la maintenance systématique n'ont pas besoin d'un système de surveillance, au contraire des maintenances conditionnelle et prévisionnelle. Ces dernières demandent à investir dans des outils permettant de collecter des données sur le fonc-

	Maintenance				
	Corrective		Préventive		
	Palliative	Curative	Systématique	Conditionnelle	Prévisionnelle
Coût de la surveillance	/	/	/	€	€€
Planification de la maintenance	/	/	⊕⊕⊕	⊕	⊕⊕
Temps d'arrêt du système	⊕⊕⊕⊕	⊕⊕⊕	⊕⊕	⊕	⊕
Coût total de la maintenance	€€€€€€	€€€€€	€€€€	€€€	€€

Tableau A.1 – Avantages et inconvénients des différentes stratégies de maintenance.

tionnement du système. De plus, la maintenance prévisionnelle a besoin de ressources pour traiter et analyser ces données, ce qui engendre un coût supplémentaire.

### Planification de la maintenance

La surveillance du système permet d'anticiper une panne et de planifier les opérations de maintenance. La maintenance conditionnelle consiste à prendre la décision lorsque la partie du système est en fin de vie, ce qui ne laisse que peu de temps pour anticiper les opérations de maintenance. *A contrario*, les traitements plus poussés des données dans le cas d'une maintenance prévisionnelle laissent une plus grande marge de manœuvre. Cependant, ces délais ne sont rien comparativement à une stratégie préventive systématique, qui permet de planifier des mois, voire des années à l'avance, les opérations de maintenance.

### Temps d'arrêt du système

Le temps d'arrêt du système est court dans le cas d'une maintenance prévisionnelle. En effet, la maintenance n'est réalisée que lorsqu'elle est nécessaire, c'est-à-dire en cas de défaut naissant. De plus, l'apparition et l'évolution du défaut étant surveillées, la maintenance se planifie longuement à l'avance, permettant d'organiser les ressources matérielles et humaines en amont. La maintenance conditionnelle ne laisse pas autant de marge de manœuvre ; la période

pour planifier la maintenance est plus courte, ce qui peut éventuellement poser des problèmes d'organisation. Dans tous les cas, ces deux stratégies de maintenance ont des temps d'arrêt du système bien inférieurs à ceux des méthodes correctives. Ces dernières ne permettent pas d'anticiper l'arrivée d'une panne et il faut alors compter le temps pour identifier le problème, pour préparer et pour effectuer les opérations de maintenance. Enfin, la maintenance systématique doit prémunir le système contre les arrêts inopinés. Cependant, elle entraîne des temps d'arrêt réguliers pour changer les pièces du système, alors que ce n'est pas toujours nécessaire.

### Coût total de la maintenance

Le coût total de la maintenance est la conséquence de la combinaison des critères précédents. Les stratégies de maintenance préventive s'en sortent bien mieux que les stratégies correctives. En effet, elles permettent d'éviter les pannes et d'anticiper les opérations de maintenance. Ainsi, elles réduisent le temps d'arrêt du système, synonyme de pertes de production.

En conclusion, le tableau [A.1](#) permet en un coup d'œil de se rendre compte de l'intérêt des stratégies de maintenance préventive, en particulier des stratégies de maintenance conditionnelle et prévisionnelle. Si ces dernières demandent quelques efforts pour être mises en place, elles permettent d'anticiper les pannes, évitant ainsi les interruptions inopinées du système, et permettent également d'utiliser les différentes parties du système pendant toute leur durée de vie.



# AStrion

## Sommaire

<b>B.1</b>	<b>AStrion, un analyseur de spectre automatique . . . . .</b>	<b>121</b>
<b>B.2</b>	<b>Les modules d’AStrion . . . . .</b>	<b>125</b>
B.2.1	AStrion-A : Angular resampling . . . . .	125
B.2.2	AStrion-D : Data validation . . . . .	126
B.2.3	AStrion-C : Cropping . . . . .	126
B.2.4	AStrion-I : peak Identification . . . . .	127
B.2.5	AStrion-H : Harmonic and sideband identification . . . . .	128
B.2.6	AStrion-K : Kinematics . . . . .	129
B.2.7	AStrion-M : deModulation . . . . .	130
B.2.8	AStrion-T : time-frequency Tracking . . . . .	130
B.2.9	AStrion-S : Surveillance . . . . .	131

## B.1 AStrion, un analyseur de spectre automatique

AStrion est un analyseur de spectre automatique. Il a été construit dans l’optique d’être très simple à utiliser et de ne nécessiter aucun paramétrage de la part de l’utilisateur. Le paramétrage est réalisé par la méthode elle-même en se basant uniquement sur le signal fourni. Cette absence de configuration préalable est la principale force d’AStrion, apportant simplicité et généricité. Le même outil permet très simplement l’analyse automatique de n’importe quel type de signal.

AStrion a été développé selon une architecture modulaire. Les modules sont indépendants et correspondent chacun à une étape de la méthode d’analyse. Chaque module répond donc à une problématique différente et c’est l’enchaînement des modules qui permet d’arriver à une analyse globale du signal. La section B.2 liste tous les modules et détaille leur rôle dans la méthode d’analyse.

Tous les modules n’ont pas la même importance dans la chaîne de traitement. Certains sont primordiaux et représentent le cœur de l’analyse spectrale, alors que d’autres, optionnels, ajoutent des pré- et post-traitements. La sélection des modules se fait en fonction du contexte et des informations que l’utilisateur souhaite extraire du signal.

Les modules d’AStrion se regroupent autour de deux objectifs. Il y a d’une part les modules permettant l’analyse d’un unique signal et d’autre part, les modules travaillant sur une séquence complète de signaux.

L’analyse d’un unique signal est organisée sous la forme d’une chaîne de traitements qui cherche à en extraire le maximum d’information. Cette chaîne est illustrée figure B.1 et elle permet de :

- transformer le signal dans le domaine angulaire (AStrion-A) ;
- vérifier que l’acquisition du signal a été faite correctement (AStrion-D) ;
- sélectionner un passage stationnaire du signal (AStrion-C) ;
- détecter les pics dans le spectre du signal (AStrion-I) ;
- identifier les structures spectrales parmi les pics détectés (AStrion-H) ;
- associer les structures spectrales aux différentes parties du système étudié (AStrion-K) ;
- démoduler les bandes de modulation (AStrion-M).

Sur la figure B.1, les modules sont représentés par trois couleurs différentes. En gris foncé, nous retrouvons les modules AStrion-D (section B.2.2) et AStrion-I (section B.2.4) qui existaient déjà au début de la thèse et qui ont servi de base aux travaux détaillés dans ce manuscrit. Les modules en gris clair sont des modules développés par d’autres membres de l’équipe SAIGA du laboratoire GIPSA-lab au cours de la thèse. Enfin, les modules en bleu sur les figures B.1 et B.2 représentent les contributions de cette thèse. Le détail de leur fonctionnement respectif fait l’objet des chapitres 2, 3 et 4.

De plus, la figure B.1 montre également que les deux modules AStrion-A et AStrion-K nécessitent des informations supplémentaires pour remplir leur rôle respectif. En effet, si ces informations sont disponibles, elles viennent enrichir les résultats fournis par la chaîne de traitement. La non-disponibilité de ces informations complémentaires n’est en revanche pas bloquante et n’empêche en rien l’analyse des signaux. Ces deux modules sont donc facultatifs, mais viennent enrichir les résultats dans le cas où des informations supplémentaires sont disponibles.

Les modules travaillant sur la séquence complète de signaux se servent des informations extraites par la chaîne de traitements AStrion-ADCIHKM, utilisée sur chaque signal  $s_k$  de la séquence  $\mathcal{S} = \{s_1, \dots, s_n\}$ . Leur objectif est de décrire l’évolution du spectre des signaux au cours de la séquence. Cette démarche est illustrée sur la figure B.2 et comprend deux étapes principales :

- l’identification des trajectoires que suivent les structures spectrales (AStrion-T) ;
- la génération d’indicateurs de santé du système et leur utilisation pour le diagnostic du système (AStrion-S).

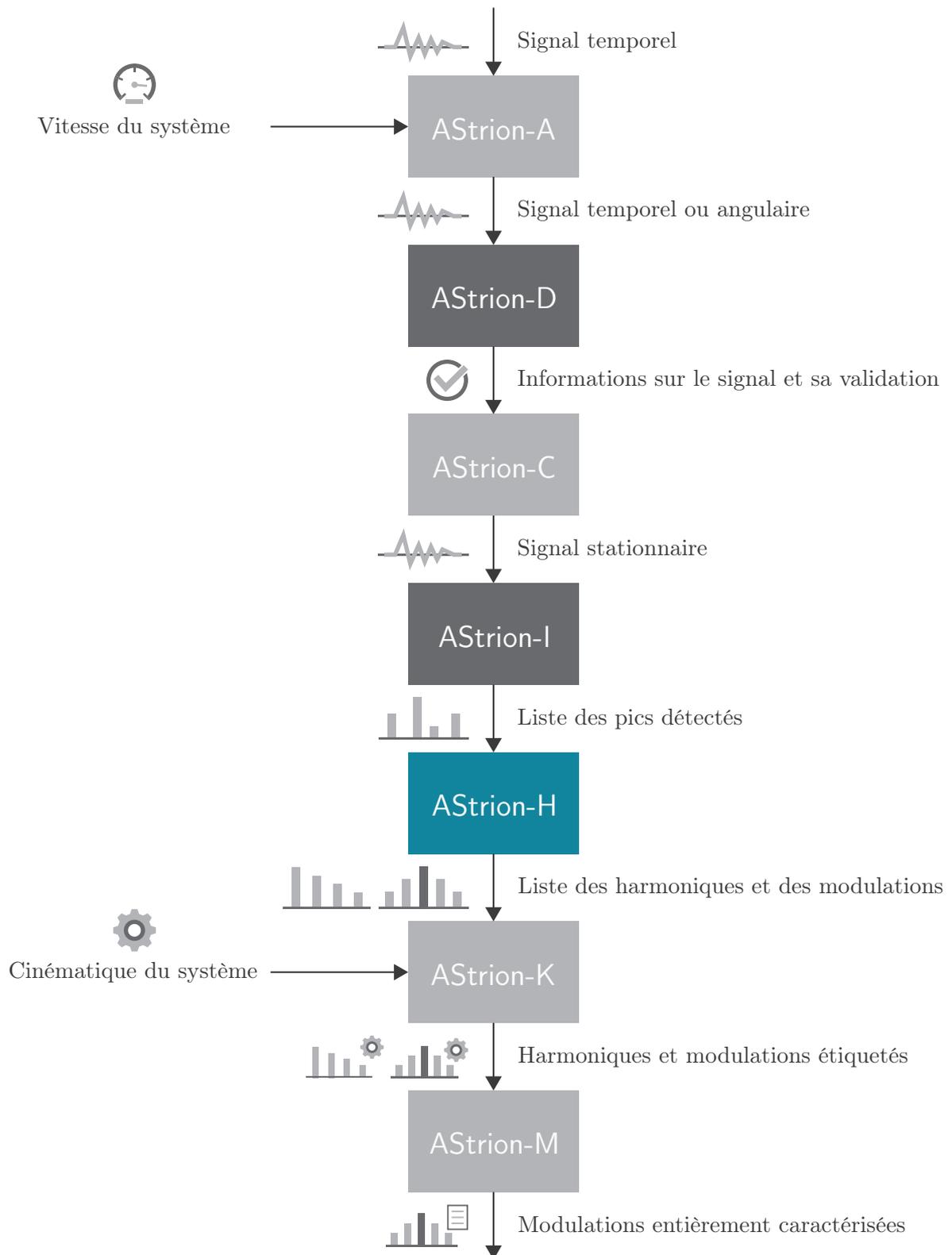


FIGURE B.1 – La chaîne de traitements AStrion-ADCIHKM permet de valider le signal, de le rendre stationnaire, d'en extraire les structures spectrales et de caractériser ces dernières.

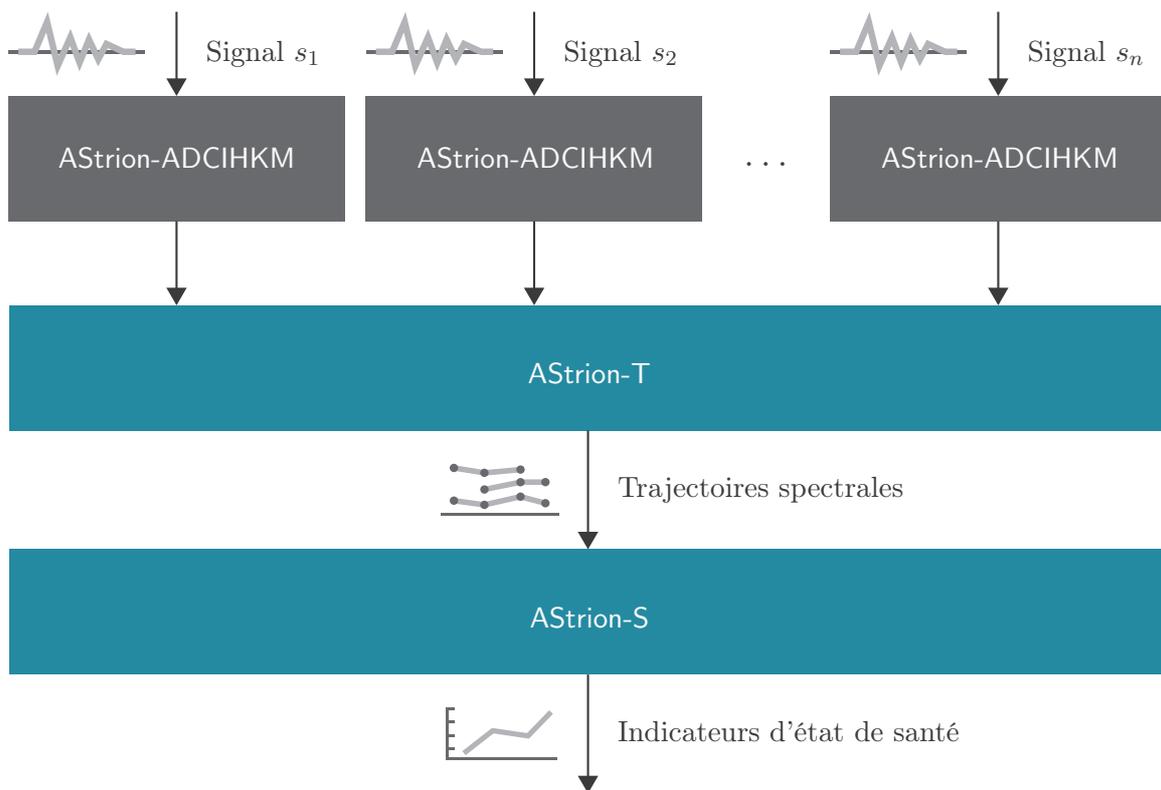


FIGURE B.2 – AStrion-T et AStrion-S permettent d'agrégier les résultats obtenus individuellement sur chacun des signaux afin d'analyser la séquence dans sa globalité.

## B.2 Les modules d'AStrion

### B.2.1 AStrion-A : Angular resampling

Le ré-échantillonnage angulaire consiste à passer un signal du domaine temporel au domaine angulaire. La différence entre ces deux domaines vient de la manière dont sont échantillonnés les signaux : dans le domaine temporel, le pas de temps est constant entre deux échantillons, alors que dans le domaine angulaire c'est le pas en angle de rotation qui est constant.

La transformée de Fourier d'un signal temporel nous amène dans le domaine fréquentiel. Dans le cas d'un signal angulaire, la transformée de Fourier nous amène dans le domaine des ordres. Ce domaine est particulièrement apprécié des mécaniciens, car il est indépendant de la vitesse de rotation en entrée du système. En effet, lorsque la vitesse de rotation d'un système évolue avec le temps, les fréquences dans le spectre évoluent également. Les ordres quant à eux, n'évoluent pas et resteront aux mêmes endroits dans le spectre en ordre.

La principale force du ré-échantillonnage angulaire est sa capacité à rendre stationnaires les composantes dépendantes de la vitesse de rotation. Pour cela, il est nécessaire d'acquérir également le signal de vitesse de rotation en même temps que le signal à analyser. Ainsi, les composantes du signal temporel évoluant avec la vitesse de rotation seront rendues stationnaires dans le signal angulaire [Firla 2014].

Il est important de noter que les composantes stationnaires ne dépendant pas de la vitesse de rotation dans le signal temporel seront transformées également et deviendront des composantes non-stationnaires du signal angulaire.

Le module AStrion-A est optionnel car il nécessite des informations en plus du signal. Les entrées et les sorties du module sont illustrées sur la figure B.3 et sont détaillées dans la suite.

#### Entrées :

- un signal temporel ;
- le signal de vitesse de rotation, synchrone avec le signal temporel.

#### Sortie :

- le signal angulaire.



FIGURE B.3 – Les entrées et sorties d'AStrion-A.

### B.2.2 AStrion-D : Data validation

Le module de validation des données a un objectif double : vérifier que l'acquisition du signal a été bien réalisée et extraire quelques informations sur le signal afin de permettre le paramétrage des modules suivants dans la chaîne de traitements.

La validation du signal est réalisée par quatre tests. Le premier test s'assure qu'il n'y a pas de saturation dans le signal. Le second vérifie qu'un filtre anti-repliement a correctement été utilisé durant l'acquisition. Enfin, les deux derniers tests vérifient la stationnarité du signal [Martin 2009]. Ces tests sont complétés par une recherche de périodicité au sein du signal et l'estimation du RSB<sup>1</sup> [Martin 2010].

Les entrées et les sorties du module AStrion-D sont illustrées sur la figure B.4 et sont détaillées dans la suite.

#### Entrée :

- un signal (angulaire ou temporel).

#### Sorties :

- les tests sur la validité du signal ;
- les informations sur le signal.



FIGURE B.4 – Les entrées et sorties d'AStrion-D.

### B.2.3 AStrion-C : Cropping

La détection des pics dans le spectre demande à ce que le signal soit stationnaire. En effet, les non-stationnarités entraînent l'étalement des pics, ce qui complique la détection. Ainsi, si le module de validation des données nous indique que le signal est fortement non-stationnaire, l'utilisation du signal en l'état n'est pas recommandée.

Ce module cherche alors à découper le signal à la fois en temps et en fréquence (ou en ordre) afin de sélectionner la plus grande plage de signal stationnaire possible. Ainsi, la suite de l'analyse ne s'effectuera que sur la plage de signal sélectionnée, qui est alors suffisamment stationnaire pour permettre la détection des pics spectraux. Bien évidemment, si le signal en entrée est stationnaire, il ne sera pas découpé et sera utilisé dans sa globalité.

Les entrées et les sorties du module AStrion-C sont illustrées sur la figure B.5 et sont détaillées dans la suite.

---

1. Rapport Signal sur Bruit

**Entrées :**

- un signal (angulaire ou temporel) ;
- les informations sur la non-stationnarité du signal.

**Sortie :**

- la sélection d'une plage stationnaire du signal.



FIGURE B.5 – Les entrées et sorties d'AStrion-C.

**B.2.4 AStrion-I : peak Identification**

L'identification des pics dans le spectre est le cœur d'AStrion et de l'analyse du signal. C'est au cours de cette étape que toute l'information est extraite. Les modules précédents peuvent être vus comme des pré-traitements afin de vérifier que cette étape se déroule correctement et les modules suivants comme des raffinements de l'information extraite ici.

L'identification des pics est basée sur la thèse de Matthieu Durnerin [Durnerin 1999] et sur les travaux qui en ont découlé [Mailhes 2006]. La stratégie d'analyse ne nécessite aucun *a priori* sur les signaux. Elle est basée sur l'utilisation et la confrontation de plusieurs estimateurs du spectre afin de tirer avantage de leurs forces respectives.

La méthode identifie l'ensemble  $\mathcal{P}$  des pics présents dans la densité spectrale du signal. Chaque pic  $p^i \in \mathcal{P}$  possède une « carte d'identité spectrale », concentrant toutes les informations estimées du pic :

- son amplitude  $a^i$  ;
- sa fréquence  $f^i$  ;
- l'amplitude  $\varepsilon^i$  de l'intervalle d'indétermination quant à la position fréquentielle exacte ;
- sa classe (sinusoïde pure, bande étroite ou bruit) ;
- sa probabilité de fausse détection ;
- son RSB local ;
- son RSB émergent ;
- sa bande à -3 dB ;
- sa stabilité ;
- son erreur quadratique à -3 dB ;
- son erreur quadratique totale ;

- le nombre d’estimateurs ayant permis sa détection ;
- le choix du meilleur estimateur pour détecter ce pic.

Les entrées et les sorties du module AStrion-I sont illustrées sur la figure B.6 et sont détaillées dans la suite.

**Entrées :**

- le signal (angulaire ou temporel) ;
- les informations sur le signal.

**Sortie :**

- la liste des pics identifiés à partir du spectre.



FIGURE B.6 – Les entrées et sorties d’AStrion-I.

### B.2.5 AStrion-H : Harmonic and sideband identification

*Ce module fait partie des contributions de cette thèse. De ce fait, le chapitre 2 détaille entièrement son fonctionnement.*

La liste des pics représente l’ensemble de l’information extraite du signal, sous une forme relativement brute. En effet, les pics sont les éléments de base qu’on retrouve dans le spectre. Plusieurs pics peuvent être liés entre eux afin de former des structures plus complexes, comme les séries harmoniques ou les bandes de modulation. Le rôle de ce module est de parcourir la liste des pics afin d’identifier les structures spectrales qui s’y cachent [Gerber 2013, Martin 2013].

Un second objectif du module est de caractériser les structures spectrales détectées. Ainsi, les séries harmoniques et les bandes de modulation ont également droit à une « carte d’identité spectrale ». La carte d’identité d’une série harmonique contient plusieurs informations comme la fréquence fondamentale, le nombre d’harmoniques ou l’énergie de la série. La liste complète des caractéristiques est donnée dans le tableau 2.2, au chapitre 2, à la page 45 du manuscrit.

D’une manière similaire, la carte d’identité d’une bande de modulation est entièrement décrite dans le tableau 2.3, au chapitre 2, à la page 46 du manuscrit. Elle contient entre autres la fréquence porteuse, la fréquence de modulation, le nombre de bandes, l’énergie dans les bandes, etc.

Les entrées et les sorties du module AStrion-H sont illustrées sur la figure B.7 et sont détaillées dans la suite.

**Entrée :**

- la liste des pics détectés dans le spectre.

**Sorties :**

- la liste des séries harmoniques ;
- la liste des bandes de modulation.



FIGURE B.7 – Les entrées et sorties d'AStrion-H.

**B.2.6 AStrion-K : Kinematics**

Le rôle de ce module est de faire le lien entre les structures spectrales extraites du signal et la cinématique du système. Le contenu spectral d'un signal est la conséquence de la structure du système surveillé et de son mode de fonctionnement. Dans le cas de la surveillance d'un système mécanique, des modèles permettent d'estimer les fréquences présentes dans le spectre en fonction de la cinématique du système. Il est alors possible de comparer ces fréquences caractéristiques aux fréquences des structures spectrales estimées, afin d'étiqueter les structures en lien avec une partie du système [Firla 2014].

Le module AStrion-K est optionnel car il nécessite des informations *a priori* sur le système surveillé. Les entrées et les sorties du module sont illustrées sur la figure B.8 et sont détaillées dans la suite.

**Entrées :**

- la liste des séries harmoniques ;
- la liste des bandes de modulation ;
- les informations sur la cinématique du système sous surveillance.

**Sorties :**

- la liste étiquetée des séries harmoniques ;
- la liste étiquetée des bandes de modulation.

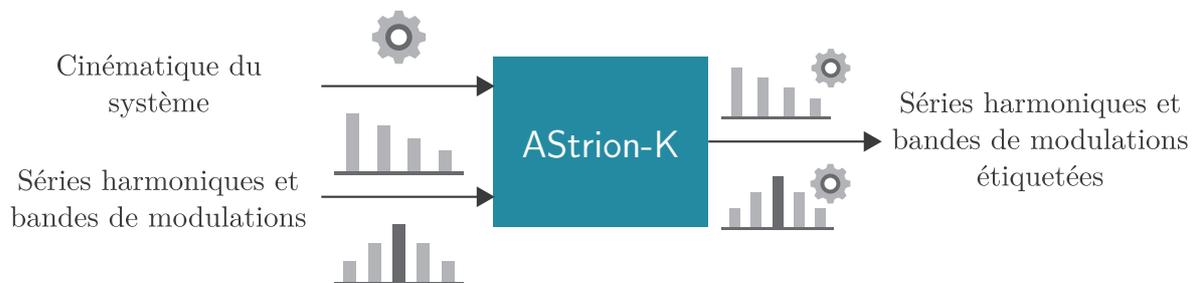


FIGURE B.8 – Les entrées et sorties d'AStrion-K.

### B.2.7 AStrion-M : deModulation

Les bandes de modulation sont d'ores et déjà caractérisées en sortie d'AStrion-H. Cependant, ces caractéristiques ne suffisent pas à décrire finement les modulations. Pour arriver à une description plus poussée, il est nécessaire de démoduler le signal. C'est le rôle de ce module.

La démodulation est faite pour chacune des bandes détectées. Les informations déjà disponibles dans la carte d'identité spectrale des bandes permet de construire un filtre qui isole la bande à démoduler [Firla 2014]. Les informations extraites suite à la démodulation sont ajoutées à la carte d'identité spectrale. Il s'agit par exemple de l'indice de modulation en amplitude, l'indice de modulation en fréquence, etc.

Les entrées et les sorties du module AStrion-M sont illustrées sur la figure B.9 et sont détaillées dans la suite.

#### Entrées :

- la liste des bandes de modulation ;
- le signal.

#### Sortie :

- la liste des bandes de modulation, caractérisées plus finement.



FIGURE B.9 – Les entrées et sorties d'AStrion-M.

### B.2.8 AStrion-T : time-frequency Tracking

*Ce module fait partie des contributions de cette thèse. De ce fait, le chapitre 3 détaille entièrement son fonctionnement.*

Ce module permet de synthétiser les informations extraites individuellement pour chaque signal de la séquence. En effet, les modules précédents ont extrait et caractérisé les structures spectrales de chaque signal. AStrion-T permet de rassembler ces structures dans des trajectoires spectrales. Une trajectoire permet de suivre une structure spectrale dans le temps, d'un signal au suivant.

Les entrées et les sorties du module AStrion-T sont illustrées sur la figure B.10 et sont détaillées dans la suite.

**Entrées :**

- les listes de pics ;
- les listes de séries harmoniques ;
- les listes de bandes de modulation.

**Sortie :**

- la liste des trajectoires spectrales.

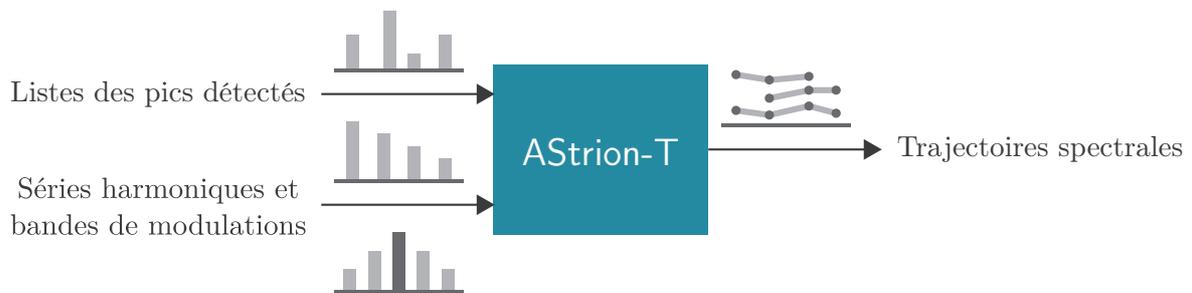


FIGURE B.10 – Les entrées et sorties d'AStrion-T.

**B.2.9 AStrion-S : Surveillance**

*Ce module fait partie des contributions de cette thèse. De ce fait, le chapitre 4 détaille entièrement son fonctionnement.*

Ce dernier module utilise les trajectoires spectrales pour générer des indicateurs sur l'état de santé du système. Ces indicateurs sont analysés afin de proposer un diagnostic du système surveillé.

Les entrées et les sorties du module AStrion-S sont illustrées sur la figure B.11 et sont détaillées dans la suite.

**Entrée :**

- les trajectoires spectrales.

**Sorties :**

- la liste des indicateurs de santé du système ;
- le diagnostic du système.

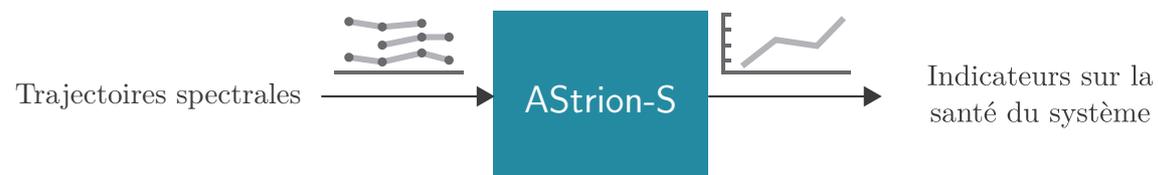


FIGURE B.11 – Les entrées et sorties d'AStrion-S.



# Banc d'essai éolien

---

## Sommaire

<b>C.1</b>	<b>Description du banc d'essai</b>	<b>133</b>
<b>C.2</b>	<b>Les séquences de signaux du banc</b>	<b>135</b>

---

Dans le cadre du projet européen [*KAStrion*], un banc d'essai simulant le fonctionnement d'une éolienne a été construit au *CETIM*, partenaire du projet. Le but du banc d'essai est d'observer la dégradation des éléments mécaniques le composant. Une description générale du banc est faite dans la section *C.1*, juste avant de présenter les signaux acquis sur ce banc dans la section *C.2*.

## C.1 Description du banc d'essai

Le banc d'essai a été construit en suivant le modèle cinématique d'une éolienne. À la place des pâles, un moteur entraîne la rotation de l'arbre principal à une vitesse d'environ 20 RPM<sup>1</sup>. Ensuite, un multiplicateur augmente la vitesse de rotation d'un facteur 100 : 1, ce qui permet de faire fonctionner la génératrice autour des 2000 RPM. Le multiplicateur est constitué de deux trains : un train planétaire (ou encore appelé train épicycloïdal) et d'un train simple à axe parallèle. La puissance du banc est de 10 kW, ce qui correspond à une puissance réduite par rapport aux puissances que l'on trouve généralement dans les éoliennes (1 à 3 MW).

Le rôle du banc d'essai est d'étudier l'usure des pièces mécaniques le composant, à savoir :

- le roulement principal à l'entrée du système (entre le moteur et le multiplicateur) ;
- le roulement à la sortie du système (entre le multiplicateur et la génératrice) ;
- le train planétaire ;
- le train simple à axe parallèle.

Pour être sûr qu'une pièce se dégradera pendant que les autres resteront en bon état, le banc est équipé de deux unités permettant d'appliquer des forces axiales et ou radiales aux niveaux des deux roulements principaux. Il est donc possible d'utiliser une unique pièce du système et d'isoler les symptômes que provoque cette usure.

---

1. Rotation Par Minute

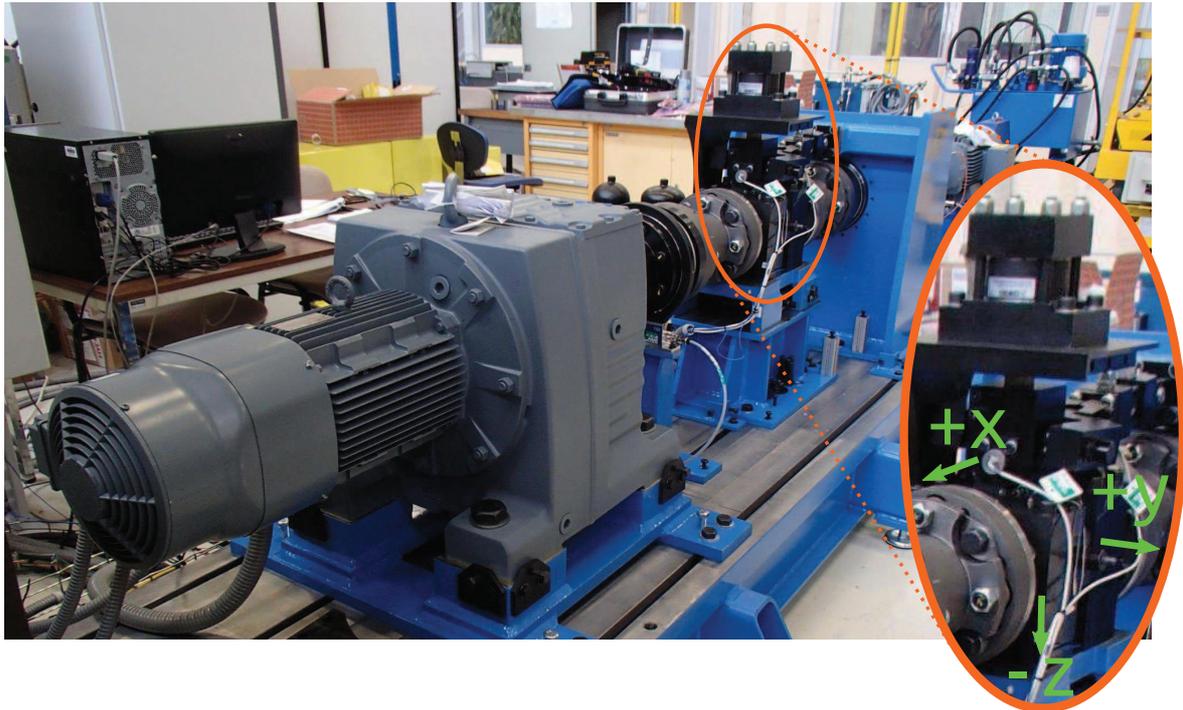


FIGURE C.1 – Photo du banc de test simulant le fonctionnement d'une éolienne. Au premier plan, un moteur remplace la combinaison du vent et des pâles. L'ellipse orange entoure l'unité de charge du roulement principal. Les flèches vertes représentent les directions des trois accéléromètres surveillant le roulement.

Le banc de test tourne sous deux modes de fonctionnement : le mode de détérioration et le mode de mesure. Pendant le mode de détérioration, les conditions de fonctionnement sont non stationnaires. En particulier, le moteur est programmé pour tourner selon un profil de vent fourni par la société VALEMO également partenaire du projet. Ce profil de vent a été mesuré sur une véritable éolienne. Ainsi, les contraintes subies par le banc sont similaires aux contraintes réelles rencontrées sur le terrain. *A contrario*, les conditions de fonctionnement pendant le mode de mesure sont stationnaires. De plus, ces conditions sont identiques d'une acquisition à l'autre, ce qui facilite la comparaison des signaux. L'hypothèse d'état machine constant (introduite dans la section 3.2.2.1) est donc vérifiée.

Le banc d'essai est une plateforme expérimentale surveillée par de nombreux capteurs : accéléromètres, thermocouples, couplemètres, ampèremètres et voltmètres. Cette surveillance exhaustive permet d'essayer de nombreuses méthodes de diagnostic, pas uniquement celle présentée dans cette thèse. Par exemple, dans [Cablea 2014], les courants et les tensions sont utilisés pour faire un diagnostic électrique du banc.

Les applications de la méthodologie présentée dans ce manuscrit se basent uniquement sur les signaux vibratoires issus des accéléromètres posés sur le roulement principal. Un accéléromètre est placé dans la direction axiale ( $+x$ ), alors que les deux autres sont en direction radiale ( $+y$  et  $-z$ ). Le positionnement des capteurs est illustré sur la figure C.1.

## C.2 Les séquences de signaux du banc

Les signaux sont acquis alors que le banc de test fonctionne en mode de mesure. Cela signifie que les conditions opérationnelles sont stationnaires et sont les mêmes entre toutes les acquisitions. Les signaux sont acquis à différents instants de la vie du roulement, que nous compterons en heures de fonctionnement à partir de la mise en place du nouveau roulement. La durée des acquisitions est de 150 secondes. La fréquence d'échantillonnage a été choisie la plus haute possible, afin de ne pas perdre d'éventuelles informations en haute fréquence. Elle est de 39062,5 Hz, ce qui nous donne finalement des signaux de presque 6 millions de points.

À la date d'écriture de ces lignes, tous les essais de dégradation des éléments du banc de tests ne sont pas terminés. Nous ne parlerons, dans ce manuscrit, que de la dégradation du roulement principal. Trois roulements différents ont été usés jusqu'à trois stades différents :

- jusqu'à l'apparition du défaut : aux premiers symptômes le test est arrêté ;
- jusqu'à la dégradation moyenne du roulement : le test est arrêté quelques temps après les premiers symptômes ;
- jusqu'à la dégradation totale du roulement : le test est arrêté lorsque le banc ne fonctionne plus à cause de l'état avancé de la dégradation.

Nous nous retrouvons alors avec 9 séquences de signaux, issues des trois roulements dégradés, surveillés par trois accéléromètres. L'étude des 9 séquences nous montre qu'il y a une grande redondance entre les capteurs : les résultats de la méthodologie présentée dans cette thèse sont similaires d'un capteur à l'autre pour un même défaut. Il faut cependant nuancer ce propos, car sur les 9 séquences de signaux, la méthode ne parvient à détecter le défaut que 8 fois. La séquence n'ayant pas abouti à la détection du défaut est celle concernant le défaut le plus léger, vue par l'accéléromètre dans la direction  $-z$ . L'analyse manuelle des signaux, réalisée également par un expert au CETIM, confirme l'absence des symptômes dans les vibrations enregistrées par ce capteur. L'hypothèse la plus probable qui peut expliquer cette absence est le fait que la direction  $-z$  correspond à la direction d'application de la force radiale pour accélérer l'usure (60 kN). Cette force aurait tendance à contraindre davantage les billes du roulement, ce qui atténuerait les vibrations dans cette direction. Le défaut venant tout juste d'apparaître, ses symptômes sont faibles et l'atténuation provoquée par la force d'usure suffit à effacer leur trace.

Les résultats entre capteurs étant similaires, ils ne sont pas détaillés pour les 9 séquences de signaux. Seule une séquence pour chaque défaut est présentée dans le manuscrit :

- la séquence de l'accéléromètre  $+x$  sur la faible dégradation est utilisée dans la section 4.1.2.4 ;
- la séquence de l'accéléromètre  $-z$  sur la dégradation moyenne est utilisée dans la section 4.1.2.3 ;
- la séquence de l'accéléromètre  $+y$  sur la forte dégradation est utilisée dans les sections 2.5.3, 3.4.2 et 4.1.2.2.



# Notations

## Ensembles

Les ensembles sont représentés par des lettres capitales calligraphiées. Les éléments de ces ensembles sont représentés par des lettres minuscules. Soient  $\mathcal{P}$  et  $\mathcal{H}$  deux ensembles et  $p$  un élément de  $\mathcal{P}$ . On note :

- $p \in \mathcal{P}$  l'appartenance de l'élément  $p$  à l'ensemble  $\mathcal{P}$ ,
- $\mathcal{P} \setminus \mathcal{H}$  la soustraction de l'ensemble  $\mathcal{H}$  à l'ensemble  $\mathcal{P}$ ,
- $\mathcal{P} \subset \mathcal{H}$  l'inclusion de l'ensemble  $\mathcal{P}$  dans l'ensemble  $\mathcal{H}$ ,
- $\#\mathcal{P}$  le cardinal de l'ensemble  $\mathcal{P}$ .

Les ensembles usuels sont notés :

- $\mathbb{N}$  pour l'ensemble des entiers naturels,
- $\mathbb{Z}$  pour l'ensemble des entiers relatifs,
- $\mathbb{R}$  pour l'ensemble des réels.

L'ajout d'une étoile en exposant signifie que l'élément nul est retiré de l'ensemble :

$$\mathbb{N}^* = \mathbb{N} \setminus \{0\}.$$

## Symboles

$s(t)$	Signal à temps continu, $t \in T \subseteq \mathbb{R}$
$s[k]$	Signal discrétisé, $k \in K \subseteq \mathbb{Z}$
$\lfloor x \rfloor$	Partie entière de $x$
$\lceil x \rceil$	Arrondi à l'entier le plus proche de $x$
$x \triangleq \dots$	Définition de $x$
$f(x) \in \mathcal{O}(g(x))$	Comparaison asymptotique : pour $x \rightarrow \infty$ , à partir d'un certain rang on a $ f(x)  < k \cdot  g(x) $ avec $k \in \mathbb{R}$ tel que $k > 0$

## Indices et exposants

Sauf contre-indication, les indices représentent l'indice du signal considéré dans la séquence de signaux. Soient un ensemble  $\mathcal{P}_n$  et la séquence de signaux  $\mathcal{S} = \{s_1, \dots, s_N\}$ . L'ensemble  $\mathcal{P}_n$  est donc lié au  $n^{\text{ème}}$  signal de la séquence  $\mathcal{S}$ , à savoir  $s_n$ . La séquence  $\mathcal{S}$  étant chronologiquement ordonnée, les indices portent en eux une notion de temporalité.

Les exposants, quant à eux, marquent l'indice de l'élément dans l'ensemble. Soient l'élément  $p^i$  et l'ensemble  $\mathcal{P} = \{p^j\}_{j \in \{1, \dots, \#\mathcal{P}\}}$ . L'élément  $p^i$  est donc le  $i^{\text{ème}}$  élément de l'ensemble  $\mathcal{P}$ . Les ensembles sont le plus souvent ordonnés par fréquence croissante. Le cas échéant, les exposants portent en eux la notion de fréquence et l'élément  $p^{i+1}$  aura alors une fréquence supérieure à l'élément  $p^i$ .

Les indices et les exposants peuvent être combinés. Ainsi, l'élément  $p_n^i$  sera le  $i^{\text{ème}}$  élément de l'ensemble  $\mathcal{P}_n = \{p_n^j\}_{j \in \{1, \dots, \#\mathcal{P}_n\}}$ ; l'ensemble  $\mathcal{P}_n$  étant issu du signal  $s_n$ , le  $n^{\text{ème}}$  signal de la séquence  $\mathcal{S}$ .

## Variables utilisées

### Générales

$\mathcal{S}$	Séquence de signaux
$s_n$	Le $n^{\text{ème}}$ signal de la séquence
$t_n$	Date de démarrage de l'acquisition du signal $s_n$
$\tau_n$	Durée du signal $s_n$ (en secondes)
$f_e$	Fréquence d'échantillonnage

### Structures spectrales

$\mathcal{P}$	Ensemble de pics spectraux
$\mathcal{P}_n$	Ensemble des pics spectraux du $n^{\text{ème}}$ signal de la séquence
$p_n^i$	Le $i^{\text{ème}}$ pic de l'ensemble $\mathcal{P}_n$
$f_n^i$	Fréquence du pic $p_n^i$
$\varepsilon_n^i$	Amplitude de l'intervalle d'indétermination sur la position fréquentielle exacte de $f_n^i$
$\mathcal{H}$	Ensemble de séries harmoniques
$\mathcal{H}_n$	Ensemble des séries harmoniques du $n^{\text{ème}}$ signal de la séquence
$h_n^i$	La $i^{\text{ème}}$ série harmonique de l'ensemble $\mathcal{H}_n$
$\nu_n^i$	Fréquence fondamentale de la série harmonique $h_n^i$
$\epsilon_n^i$	Amplitude de l'intervalle d'indétermination sur la position fréquentielle exacte de $\nu_n^i$
$\mathcal{M}$	Ensemble de bandes de modulation
$\mathcal{M}_n$	Ensemble des bandes de modulation du $n^{\text{ème}}$ signal de la séquence
$m_n^i$	La $i^{\text{ème}}$ modulation de l'ensemble $\mathcal{M}_n$
$f_p$	Fréquence porteuse
$f_{m,n}^i$	Fréquence de modulation de la bande de modulation $m_n^i$
$f_{\Delta}^{\max}$	Déviations maximales en fréquence où rechercher les modulations

$\mathcal{D}^p$	Ensemble des pics à droite de la fréquence porteuse $f_p$ compris dans la déviation maximale $f_{\Delta}^{\max}$
$\mathcal{G}^p$	Ensemble des pics à gauche de la fréquence porteuse $f_p$ compris dans la déviation maximale $f_{\Delta}^{\max}$
$\#\mathcal{P}^{\min}$	Nombre minimal de pics dans une structure spectrale
$r_{\emptyset}^{\max}$	Nombre maximal de rangs consécutifs manquants autorisé
$q^{\min}$	Ratio de choix entre une structure spectrale commençant au rang 1 ou au rang 2
$N_{\text{pics}}$	Nombre de pics dans une structure spectrale
$r_{\max}$	Rang du dernier élément dans le structure spectrale
$D$	Densité d'une structure spectrale
$E$	Énergie d'une structure spectrale
THD	Taux de distorsion harmonique d'une structure spectrale
$R$	Richesse d'une structure spectrale

### Trajectoires spectrales

$\mathcal{T}_{\mathcal{P}}$	Ensemble des trajectoires de pics spectraux
$t_{\mathcal{P}}^i$	La $i^{\text{ème}}$ trajectoire de pics de l'ensemble $\mathcal{T}_{\mathcal{P}}$
$\mathcal{T}_{\mathcal{H}}$	Ensemble des trajectoires de séries harmoniques
$t_{\mathcal{H}}^i$	La $i^{\text{ème}}$ trajectoire harmonique de l'ensemble $\mathcal{T}_{\mathcal{H}}$
$\mathcal{T}_{\mathcal{M}}$	Ensemble des trajectoires de bandes de modulation
$t_{\mathcal{M}}^i$	La $i^{\text{ème}}$ trajectoire de modulations de l'ensemble $\mathcal{T}_{\mathcal{M}}$
$\Delta f$	Intervalle de recherche pour le suivi temps-fréquence
$k_{\Delta}$	Facteur multiplicatif pour le calcul de l'intervalle de recherche

### Séries temporelles et tests statistiques

$x_t$	Valeur de la série temporelle à l'instant $t$
$\varepsilon_t$	Résidu entre la valeur de la série temporelle et un modèle à l'instant $t$
$\mu$	Valeur moyenne ou constante d'une série temporelle
$\sigma_x^2$	Variance d'une série temporelle $x_t$
$\delta$	Dérive temporelle
$\Delta$	Opérateur des différences premières
$\rho$	Coefficient d'un modèle autorégressif d'ordre 1
$\mathcal{H}_0$	Hypothèse nulle
$\mathcal{H}_1$	Hypothèse alternative
$p$	Valeur $p$ : probabilité d'obtenir une valeur au moins aussi grande de la statistique si on part du principe que $\mathcal{H}_0$ est vraie



# Table des sigles et acronymes

<b>ADF</b>	Augmented Dickey-Fuller
<b>CETIM</b>	CEntre Technique des Industries Mécaniques
<b>CMS</b>	Condition Monitoring System
<b>DF</b>	Dickey-Fuller
<b>DF-GLS</b>	Dickey-Fuller — Generalized Least Square
<b>DSE</b>	Densité Spectrale d'Énergie
<b>DSP</b>	Densité Spectrale de Puissance
<b>GIPSA-lab</b>	Grenoble, Images, Parole, Signal et Automatique
<b>IRIT</b>	Institut de Recherche en Informatique de Toulouse
<b>KPSS</b>	Kwiatkowski–Phillips–Schmidt–Shin
<b>PP</b>	Phillips–Perron
<b>RPM</b>	Rotation Par Minute
<b>RSB</b>	Rapport Signal sur Bruit
<b>SAIGA</b>	Signal et Automatique pour la surveillance, le diaGnostic et la biomécAnique
<b>TFD</b>	Transformée de Fourier Discrète
<b>THD</b>	Taux de distortion harmonique ( <i>Total Harmonic Distorsion</i> , en anglais)



# Table des figures

1.1	Un pic spectral . . . . .	12
1.2	Une série harmonique . . . . .	13
1.3	Une modulation en amplitude . . . . .	14
1.4	Une modulation angulaire . . . . .	15
1.5	Exemple de densité spectrale de puissance d'un signal vibratoire . . . . .	16
1.6	AStrion-DI . . . . .	17
1.7	Évolution d'un spectre au cours du temps . . . . .	20
1.8	Illustration du principe de trajectoires . . . . .	21
1.9	Comparaison de deux spectres . . . . .	22
2.1	Recherche de l'harmonique de rang $r$ par intersection d'intervalles . . . . .	30
2.2	Mise à jour des paramètres de recherche de la série harmonique . . . . .	32
2.3	Algorithme de recherche de l'ensemble $\mathcal{H}$ des séries harmoniques. . . . .	37
2.4	Modèle d'une modulation bilatérale . . . . .	38
2.5	Recherche des modulations . . . . .	41
2.6	Interprétation du spectre d'un défaut simulé . . . . .	48
2.7	Détection de séries harmoniques au sein des modulations . . . . .	49
2.8	Application sur signal réel : avant le défaut . . . . .	50
2.9	Application sur signal réel : pendant la formation du défaut . . . . .	50
2.10	Application sur signal réel : le défaut est bien formé . . . . .	50
3.1	Un exemple de carte temps-fréquence . . . . .	55
3.2	Un exemple de trajectoires . . . . .	56
3.3	Suivi d'un pic spectral . . . . .	62
3.4	Illustration du principe d'état sommeil d'une trajectoire . . . . .	65

3.5	Répartition du nombre de trajectoires en fonction de leur taille . . . . .	67
3.6	Illustration des étapes du suivi de séries harmoniques . . . . .	69
3.7	Illustration des étapes de suivi de bandes de modulation . . . . .	71
3.8	Nombre de structures spectrales dans la séquence artificielle . . . . .	75
3.9	Carte temps-fréquence de la séquence du défaut d'engrènement simulé . . . . .	76
3.10	Zoom sur la carte temps-fréquence du défaut de roulement . . . . .	78
3.11	Photo du défaut sur le roulement principal . . . . .	79
4.1	Illustration de la génération de tendances . . . . .	83
4.2	Défaut simulé — Fréquence de la trajectoire harmonique . . . . .	84
4.3	Défaut simulé — Énergie de la trajectoire harmonique . . . . .	84
4.4	Défaut simulé — Énergie des trajectoires de modulations . . . . .	85
4.5	Défaut simulé — Nombre de pics des trajectoires de modulations . . . . .	86
4.6	Forte dégradation du roulement principal — Énergie vibratoire de la bague externe . . . . .	87
4.7	Forte dégradation du roulement principal — Fréquence fondamentale de la série harmonique . . . . .	88
4.8	Dégradation moyenne du roulement principal — Énergie vibratoire de la bague externe . . . . .	89
4.9	Faible dégradation du roulement principal — Énergie vibratoire de la bague externe . . . . .	90
4.10	Photos d'une éolienne et de son système de surveillance . . . . .	91
4.11	Éolienne — Ordre de l'arbre rapide . . . . .	93
4.12	Éolienne — Énergie vibratoire de l'arbre rapide . . . . .	93
4.13	Répartition des combinaisons de tests sur le plan défini par les taux de bonnes détections et de fausses alarmes . . . . .	107
4.14	Allures des tendances classées comme stationnaires et non stationnaires . . . . .	110
B.1	AStrion-ADCIHKM . . . . .	123
B.2	AStrion-ADCIHKMTS . . . . .	124

---

B.3	AStrion-A . . . . .	125
B.4	AStrion-D . . . . .	126
B.5	AStrion-C . . . . .	127
B.6	AStrion-I . . . . .	128
B.7	AStrion-H . . . . .	129
B.8	AStrion-K . . . . .	129
B.9	AStrion-M . . . . .	130
B.10	AStrion-T . . . . .	131
B.11	AStrion-S . . . . .	131
C.1	Photo du banc de test simulant le fonctionnement d'une éolienne . . . . .	134



# Liste des tableaux

2.1	Carte d'identité spectrale restreinte d'un pic $p^i$ . . . . .	27
2.2	Carte d'identité spectrale d'une série harmonique $h^i$ . . . . .	45
2.3	Carte d'identité spectrale d'une bande de modulation bilatérale $m^i$ . . . . .	46
2.4	Liste des paramètres utilisés pour l'identification des séries harmoniques. . . . .	47
2.5	Liste des paramètres utilisés pour l'identification des bandes de modulation. . . . .	47
3.1	Liste des paramètres utilisés pour le suivi des structures spectrales. . . . .	67
4.1	Décisions prises pour chaque test statistique. . . . .	104
4.2	Résultats de la classification pour chaque test statistique. . . . .	106
4.3	Résultats de la classification pour les 20 meilleures combinaisons de tests statistiques et de seuils. . . . .	108
A.1	Avantages et inconvénients des différentes stratégies de maintenance. . . . .	118



# Bibliographie

- [Akaike 1974] AKAIKE, H (1974). “A new look at the statistical model identification”. *Automatic Control, IEEE Transactions on* 19.6, p. 716–723 (cité page 101).
- [Asiminoael 2007] ASIMINOAEL, Lucian, Frede BLAABJERG et Steffan HANSEN (2007). “Detection is key - Harmonic detection methods for active power filter applications”. *IEEE Industry Applications Magazine* 13.4, p. 22–33.
- [Bonnardot 2005] BONNARDOT, F., M. EL BADAoui, R. B. RANDALL, J. DANIÈRE et F. GUILLET (2005). “Use of the acceleration signal of a gearbox in order to perform angular resampling (with limited speed fluctuation)”. *Mechanical Systems and Signal Processing* 19.4, p. 766–785 (cité page 61).
- [Cablea 2014] CABLEA, G, P GRANJON et C BÉRENGUER (2014). “Method for computing efficient electrical indicators for offshore wind turbine monitoring”. *Insight - Non-Destructive Testing and Condition Monitoring* 56.8, p. 443–448 (cité page 134).
- [Cai 2009] CAI, Tao, Shanxu DUAN et Bangyin LIU (2009). “A robust parametric method for power harmonic estimation based on M-estimators”. *2009 IEEE 6th International Power Electronics and Motion Control Conference, IPEMC '09* 43.1, p. 2501–2506.
- [Chandola 2009] CHANDOLA, Varun, Arindam BANERJEE et Vipin KUMAR (2009). “Anomaly detection”. *ACM Computing Surveys* 41.3, p. 1–58 (cité page 94).
- [Chen 2006] CHEN, Zhixin et Robert C MAHER (2006). “Semi-automatic classification of bird vocalizations using spectral peak tracks”. *The Journal of the Acoustical Society of America* 120.5, p. 2974–2984 (cité page 57).
- [Chow 1960] CHOW, Gregory C. (1960). “Tests of Equality Between Sets of Coefficients in Two Linear Regressions”. *Econometrica* 28.3, p. 591–605 (cité page 98).
- [Court 2013] COURT, Richard, Peter J. TAVNER, Christian LITTLE et Wenxian YANG (2013). “Data-driven technique for interpreting wind turbine condition monitoring signals”. *IET Renewable Power Generation* May 2013, p. 151–159 (cité page 8).
- [Davidson 2004] DAVIDSON, Russell et James G MACKINNON (2004). *Econometric theory and methods*. T. 5. Oxford University Press New York (cité page 101).
- [Depalle 1993] DEPALLE, P., G. GARCIA et X. RODET (1993). “Tracking of partials for additive sound synthesis using hidden Markov models”. *IEEE International Conference on Acoustics, Speech, and Signal Processing*. T. 1, p. 225–228 (cité page 57).
- [Dickey 1979] DICKEY, David A et Wayne A FULLER (1979). “Distribution of the estimators for autoregressive time series with a unit root”. *Journal of the American statistical association* 74.366a, p. 427–431 (cité pages 99, 100).

- [Ding 2009] DING, S. X., P. ZHANG, A. NAIK, E. L. DING et B. HUANG (2009). “Subspace method aided data-driven design of fault detection and isolation systems”. *Journal of Process Control* 19.9, p. 1496–1510 (cité page 8).
- [Doval 1991] DOVAL, B. et X. RODET (1991). “Estimation of fundamental frequency of musical sound signals”. *IEEE International Conference on Acoustics, Speech, and Signal Processing*. T. 5, p. 3657–3660 (cité page 28).
- [Doval 1993] DOVAL, B. et X. RODET (1993). “Fundamental frequency estimation and tracking using maximum likelihood harmonic matching and HMMs”. *IEEE International Conference on Acoustics, Speech, and Signal Processing*. T. 1, p. 221–224 (cité page 57).
- [Duan 2014] DUAN, Zhiyao, Jinyu HAN et Bryan PARDO (2014). “Multi-pitch streaming of harmonic sound mixtures”. *IEEE Transactions on Audio, Speech and Language Processing* 22.1, p. 138–150 (cité page 58).
- [Durnerin 1999] DURNERIN, Matthieu (1999). “Une stratégie pour l’interprétation en analyse spectrale. Détection et caractérisation des composantes d’un spectre.” Thèse de doct. Institut National Polytechnique de Grenoble (cité pages 16, 127).
- [Elliott 1996] ELLIOTT, Graham, Thomas J ROTHENBERG et James H STOCK (1996). “Efficient tests for an autoregressive unit root”. *Econometrica* 64.4, p. 813–836 (cité page 102).
- [Firla 2016] FIRLA, Marcin (2016). “Automatic Signal Processing for Wind Turbine Condition Monitoring. Time-Frequency Cropping, Kinematic Association, and All-Sideband Demodulation.” Thèse. Univ. Grenoble Alpes, Grenoble, France (cité page 91).
- [Firla 2014] FIRLA, Marcin, Zhong-Yang LI, Nadine MARTIN et Tomasz BARSZCZ (2014). “Automatic and Full-band Demodulation for Fault Detection. Validation on a Wind Turbine Test Bench”. *4th International Conference on Condition Monitoring of Machinery in Non-Stationary Operations (CMMN0’2014)*. Lyon, France (cité pages 61, 125, 129, 130).
- [Firla 2015] FIRLA, Marcin, Timothée GERBER, Pascal BELLEMAIN et Nadine MARTIN (2015). “Automatic Method for Spectral Pattern Association With Characteristic Frequencies”. *Diagnostyka* 16.4, p. 77–84 (cité page 109).
- [Fujimoto 2012] FUJIMOTO, K., N. HAMADA et W. KASPRZAK (2012). “Estimation and tracking of fundamental, 2nd and 3d harmonic frequencies for spectrogram normalization in speech recognition”. *Bulletin of the Polish Academy of Sciences : Technical Sciences* 60.1, p. 71–81 (cité page 28).
- [García Márquez 2012] GARCÍA MÁRQUEZ, Fausto Pedro, Andrew Mark TOBIAS, Jesús María PINAR PÉREZ et Mayorkinos PAPAELIAS (2012). “Condition monitoring of wind turbines : Techniques and methods”. *Renewable Energy* 46, p. 169–178 (cité page 7).

- [Gelman 2000] GELMAN, L.M., D.a. KRIPAK, V.V. FEDOROV et L.N. UDOVENKO (2000). “Condition Monitoring Diagnosis Methods of Helicopter Units”. *Mechanical Systems and Signal Processing* 14.4, p. 613–624.
- [Gerber 2013] GERBER, Timothée, Nadine MARTIN et Corinne MAILHES (2013). “Identification of harmonics and sidebands in a finite set of spectral components”. *The Tenth International Conference on Condition Monitoring and Machinery Failure Prevention Technologies (CM & MFPT 2013)*. Kraków, Poland (cité pages 52, 128).
- [Gerber 2014] GERBER, Timothée, Nadine MARTIN et Corinne MAILHES (2014). “Monitoring based on time-frequency tracking of estimated harmonic series and modulation sidebands”. *4th International Conference on Condition Monitoring of Machinery in Non-Stationary Operations (CMMN0’2014)*. Lyon, France (cité page 109).
- [Gerber 2015] GERBER, Timothée, Nadine MARTIN et Corinne MAILHES (2015). “Time-Frequency Tracking of Spectral Structures Estimated by a Data-Driven Method”. *Industrial Electronics, IEEE Transactions on* 62.10, p. 6616–6626 (cité page 80).
- [Gong 2013] GONG, Xiang et Wei QIAO (2013). “Bearing Fault Diagnosis for Direct-Drive Wind Turbines via Current Demodulated Signals”. *IEEE Transactions on Industrial Electronics* 60.8, p. 3419–3428 (cité page 7).
- [Greene 2003] GREENE, William H (2003). *Econometric analysis*. Pearson Education India (cité page 101).
- [Hang 2014] HANG, Jun, Jianzhong ZHANG et Ming CHENG (2014). “Fault diagnosis of wind turbine based on multi-sensors information fusion technology”. *IET Renewable Power Generation* 8.3, p. 289–298 (cité page 8).
- [He 2013] HE, David, Ruoyu LI et Junda ZHU (2013). “Plastic Bearing Fault Diagnosis Based on a Two-Step Data Mining Approach”. *Industrial Electronics, IEEE Transactions on* 60.8, p. 3429–3440 (cité page 7).
- [Heisenberg 1927] HEISENBERG, W. (1927). “Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik”. German. *Zeitschrift für Physik* 43.3-4, p. 172–198 (cité page 26).
- [Heng 2009] HENG, Aiwina, Sheng ZHANG, Andy C C TAN et Joseph MATHEW (2009). “Rotating machinery prognostics : State of the art, challenges and opportunities”. *Mechanical Systems and Signal Processing* 23.3, p. 724–739 (cité page 7).
- [ISO10816-21 2015] ISO10816-21 (2015). *Vibrations mécaniques - Évaluation des vibrations des machines par mesurages sur les parties non tournantes - Partie 21 : Turbines éoliennes à axe horizontal avec multiplicateur* (cité page 7).
- [Jardine 2006] JARDINE, a. K S, Daming LIN et Dragan BANJEVIC (2006). “A review on machinery diagnostics and prognostics implementing condition-based maintenance”. *Mechanical Systems and Signal Processing* 20.7, p. 1483–1510 (cité page 7).
- [KAstrion] KAstrion. URL : <http://www.gipsa-lab.fr/projet/KASTRION/> (visité le 08/06/2015) (cité pages 2, 6, 52, 80, 90, 111, 133).

- [Kim 2013] KIM, Yong Hwa, Young Woo YOUN, Don Ha HWANG, Jong Ho SUN et Dong Sik KANG (2013). “High-resolution parameter estimation method to identify broken rotor bar faults in induction motors”. *IEEE Transactions on Industrial Electronics* 60.9, p. 4103–4117 (cité page 7).
- [Kuhn 1955] KUHN, H.W. (1955). “The Hungarian method of solving the assignment problem”. *Naval Res. Logistics Quart* 2.1-2, p. 83–97 (cité page 58).
- [Lagrange 2004] LAGRANGE, Mathieu (2004). “Modélisation sinusoïdale des sons polyphoniques”. Thèse de doct. Université de Bordeaux 1 (cité page 59).
- [Lagrange 2003] LAGRANGE, Mathieu, Sylvain MARCHAND, Martin RASPAUD et J-B RAULT (2003). “Enhanced partial tracking using linear prediction”. *Proceedings of the Digital Audio Effects (DAFx03) Conference*. United Kingdom, p. 141–146 (cité page 57).
- [Le Coz 2013] LE COZ, Maxime, Julien PINQUIER, Régine ANDRÉ-OBRECHT et J MAUCLAIR (2013). “Audio indexing including frequency tracking of simultaneous multiple sources in speech and music”. *Content-Based Multimedia Indexing (CBMI), 2013, 11th International Workshop on*. June. Veszprém, Hungary, p. 23–28 (cité page 58).
- [Lebranchu 2015] LEBRANCHU, Alexis, Sylvie CHARBONNIER, Christophe BERENGUER et Frédéric PREVOST (2015). “Using SCADA Data for Fault Detection in Wind Turbines : Local Internal Model Versus Distance to a Wind Farm Reference”. *Advances in Condition Monitoring of Machinery in Non-Stationary Operations SE - 17*. Sous la dir. de Fakher CHAARI, Radoslaw ZIMROZ, Walter BARTELMUS et Mohamed HADDAR. T. 4. Applied Condition Monitoring. Springer International Publishing, p. 225–236 (cité page 1).
- [Li 2015] LI, Z-Y, T GERBER, M FIRLA, P BELLEMAIN, N MARTIN et C MAILHES (2015). “Astrion strategy : from acquisition to diagnosis. Application to wind turbine monitoring”. *Insight - Non-Destructive Testing and Condition Monitoring* 57.8, p. 442–447 (cité page 109).
- [Liu 2013] LIU, Bin, Haifeng CHEN, Abhishek SHARMA, Guofei JIANG et Hui XIONG (2013). “Modeling heterogeneous time series dynamics to profile big sensor data in complex physical systems”. *Proceedings - 2013 IEEE International Conference on Big Data, Big Data 2013*, p. 631–638 (cité page 94).
- [Ma 2003] MA, Junshui et Simon PERKINS (2003). “Online novelty detection on temporal sequences”. *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '03*, p. 613–618 (cité page 94).
- [MacKinnon 1994] MACKINNON, James G (1994). “Approximate asymptotic distribution functions for unit-root and cointegration tests”. *Journal of Business & Economic Statistics* 12.2, p. 167–176 (cité page 101).
- [Mahanty 2004] MAHANTY, Rn et Pbd GUPTA (2004). “Application of RBF neural network to fault classification and location in transmission lines”. *IEE Proceedings - Generation, Transmission and Distribution* 151.2, p. 201–212.

- [Maher 1994] MAHER, Robert C. (1994). “Fundamental frequency estimation of musical signals using a two-way mismatch procedure”. *The Journal of the Acoustical Society of America* 95, p. 2254 (cité pages 28, 58).
- [Mailhes 2006] MAILHES, Corinne, Nadine MARTIN, Kheira SAHLI et Gérard LEJEUNE (2006). “A spectral identity card”. *EUSIPCO, European Signal Processing Conference*. Florence, Italy (cité page 127).
- [Mark 1978] MARK, WD (1978). “Analysis of the vibratory excitation of gear systems : basic theory”. *The Journal of the Acoustical Society of America*, p. 1409–1430 (cité page 16).
- [Martin 2009] MARTIN, Nadine et Corinne MAILHES (2009). “A non-stationary index resulting from time and frequency domains”. *Sixth International Conference on Condition Monitoring and Machinery Failure Prevention Technologies (CM & MFPT 2009)*. Dublin, Ireland (cité page 126).
- [Martin 2010] MARTIN, Nadine et Corinne MAILHES (2010). “About periodicity and signal to noise ratio - The strength of the autocorrelation function.” *Seventh International Conference on Condition Monitoring and Machinery Failure Prevention Technologies (CM 2010 & MFPT 2010)*. Stratford, United Kingdom (cité page 126).
- [Martin 2013] MARTIN, Nadine, Corinne MAILHES et Timothée GERBER (2013). *Anomaly detection system*. Patent no. FR N°13/53860. URL : <http://hal.archives-ouvertes.fr/hal-00848297> (cité pages 52, 128).
- [Martin 1982] MARTIN, P. (1982). “Comparison of pitch detection by cepstrum and spectral comb analysis”. *ICASSP '82. IEEE International Conference on Acoustics, Speech, and Signal Processing* 7, p. 180–183 (cité page 27).
- [Mazet 2013] MAZET, Vincent, Charles SOUSSEN et El-Hadi DJERMOUNE (2013). “Décomposition de spectres en motifs paramétriques par approximation parcimonieuse”. *XXIVe Colloque GRETSI Traitement du Signal & des Images*, p. 1–4 (cité page 58).
- [McAulay 1986] MCAULAY, R. et T. QUATIERI (1986). “Speech analysis/Synthesis based on a sinusoidal representation”. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 34.4, p. 744–754 (cité pages 57, 59).
- [McFadden 1986] MCFADDEN, P D (1986). “Detecting Fatigue Cracks in Gears by Amplitude and Phase Demodulation of the Meshing Vibration”. *Journal of vibration, acoustics, stress, and reliability in design* 108.2, p. 165–170 (cité page 16).
- [Metropolis 1949] METROPOLIS, Nicholas et S. ULAM (1949). “The Monte Carlo Method”. *Journal of the American Statistical Association* 44.247, p. 335–341 (cité page 66).
- [Monchy 2010] MONCHY, François et Jean-Pierre VERNIER (2010). *Maintenance : Méthodes et organisations*. 3ème édit. Dunod (cité page 115).
- [Owen 2009] OWEN, Art B. (2009). “Karl Pearson’s meta-analysis revisited”. *Annals of Statistics* 37.6 B, p. 3867–3892. arXiv : 0911.3531 (cité page 109).

- [Pendharkar 2012] PENDHARKAR, Chinmay (2012). “Auralization of road vehicles using spectral modeling synthesis”. Master thesis. Chalmers University of Technology (cité page 57).
- [Phillips 1988] PHILLIPS, Peter C B et Pierre PERRON (1988). “Testing for a unit root in time series regression”. *Biometrika* 75.2, p. 335–346 (cité page 101).
- [Priestley 1981] PRIESTLEY, Maurice Bertram (1981). “Spectral analysis and time series” (cité page 9).
- [Quinn 2001] QUINN, B G et E J HANNAN (2001). *The estimation and tracking of frequency*. T. 9. Cambridge University Press (cité page 57).
- [Randall 1982] RANDALL, R. B. (1982). “A New Method of Modeling Gear Faults”. *Journal of Mechanical Design* 104.2, p. 259 (cité page 16).
- [Randall 2011] RANDALL, Robert B. et Jérôme ANTONI (2011). “Rolling element bearing diagnostics - A tutorial”. *Mechanical Systems and Signal Processing* 25.2, p. 485–520 (cité page 63).
- [Randall 2011] RANDALL, Robert Bond (2011). *Vibration-based Condition Monitoring*. John Wiley & Sons (cité page 115).
- [Rauber 2015] RAUBER, Thomas W., Francisco DE ASSIS BOLDT et Flavio Miguel VAREJAO (2015). “Heterogeneous Feature Models and Feature Selection Applied to Bearing Fault Diagnosis”. *IEEE Transactions on Industrial Electronics* 62.1, p. 637–646 (cité page 7).
- [Redortier 1987] REDORTIER, Bernard (1987). “Détection des harmoniques dans un spectre de vibrations par des méthodes de traitement d’images”. *Traitement du Signal* 6, p. 269–276 (cité page 28).
- [Saporta 2011] SAPORTA, Gilbert (2011). *Probabilités, analyse des données et statistique*. Editions Technip (cité page 103).
- [Schroeder 1968] SCHROEDER, M R (1968). “Period histogram and product spectrum : new methods for fundamental-frequency measurement.” *The Journal of the Acoustical Society of America* 43.4, p. 829–834 (cité page 27).
- [Serra 1990] SERRA, Xavier (1990). “A sound decomposition system based on a deterministic plus residual model”. *The Journal of the Acoustical Society of America* 87.S1, S97 (cité page 57).
- [Serra 1997] SERRA, Xavier (1997). *Musical Sound Modeling with Sinusoids plus Noise*. Swets & Zeitlinger Publishers, p. 1–25 (cité pages 57, 64).
- [Shin 1992] SHIN, Yongcheol, Denis KWIATKOWSKI, Peter SCHMIDT et Peter C. B. PHILLIPS (1992). “Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root : How Sure Are We That Economic Time Series Are Nonstationary ?” *Journal of Econometrics* 54, p. 159–178 (cité page 102).

- [Slone 2001] SLONE, G Randy (2001). *The audiophile's project sourcebook*. McGraw-Hill/TAB Electronics (cité page 44).
- [Smith 1987] SMITH, Julius O et Xavier SERRA (1987). *PARSHL : An Analysis / Synthesis Program for Non-Harmonic Sounds Based on a Sinusoidal Representation*, p. 1–22 (cité page 57).
- [Snedecor 1991] SNEDECOR, George W et William Gemmell COCHRAN (1991). *Statistical Methods*. Wiley, p. 503 (cité page 98).
- [Student 1908] STUDENT (1908). “The probable error of a mean”. *Biometrika* 6.1, p. 1–25 (cité page 96).
- [Sun 2014] SUN, Hailiang, Yanyang ZI et Zhengjia HE (2014). “Wind turbine fault detection using multiwavelet denoising with the data-driven block threshold”. *Applied Acoustics* 77, p. 122–129 (cité page 8).
- [SUPREME] SUPREME. URL : <http://www.supreme-fof.eu/> (visité le 08/06/2015) (cité pages 2, 6, 52, 80, 111).
- [Tong 2007] TONG, Liqing, Zhaoming QIAN, Naixing KUANG, Xue LINGXIAO et Fang Zheng PENG (2007). “Optimal design of synchronous reference frame harmonic detection method”. *PESC Record - IEEE Annual Power Electronics Specialists Conference* 1, p. 1071–1076.
- [Večeř 2005] VEČEŘ, P, M KREIDL et R ŠMÍD (2005). “Condition Indicators for Gearbox Condition Monitoring Systems”. *Acta Polytechnica* 45.6, p. 35–43 (cité page 7).
- [Viterbi 1967] VITERBI, a. (1967). “Error bounds for convolutional codes and an asymptotically optimum decoding algorithm”. *IEEE Transactions on Information Theory* 13.2, p. 260–269 (cité page 57).
- [Wang 2000] WANG, Chao Wang Chao et S. SENEFF (2000). “Robust pitch tracking for prosodic modeling in telephone speech”. *IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings*. T. 3, p. 1343–1346 (cité page 27).
- [Wang 2014] WANG, Tianyang, Ming LIANG, Jianyong LI et Weidong CHENG (2014). “Rolling element bearing fault diagnosis via fault characteristic order (FCO) analysis”. *Mechanical Systems and Signal Processing* 45.1, p. 139–153 (cité page 61).
- [Whitlock 2005] WHITLOCK, M. C. (2005). “Combining probability from independent tests : The weighted Z-method is superior to Fisher's approach”. *Journal of Evolutionary Biology* 18.5, p. 1368–1373 (cité page 109).
- [Wolcín 1980] WOLCÍN, Joseph J. (1980). “Maximum a posteriori estimation of narrow-band signal parameters”. *The Journal of the Acoustical Society of America* 68.1, p. 174 (cité page 57).
- [Yamahata 2009] YAMAHATA, Shoko, Mitsuharu MATSUMOTO et Shuji HASHIMOTO (2009). “A blind separation of monaural sound based on peak tracking of frequency spec-

- tra”. *Proceedings - 2009 International Conference on Information Management and Engineering, ICIME 2009*, p. 305–311 (cité page 57).
- [Yin 2014a] YIN, Shen, Steven X. DING, Xiaochen XIE et Hao LUO (2014a). “A review on basic data-driven approaches for industrial process monitoring”. *IEEE Transactions on Industrial Electronics* 61.11, p. 6414–6428 (cité page 8).
- [Yin 2014b] YIN, Shen, Guang WANG et Hamid Reza KARIMI (2014b). “Data-driven design of robust fault detection system for wind turbines”. *Mechatronics* 24.4, p. 298–306 (cité page 8).
- [Zahorian 2008] ZAHORIAN, Stephen a et Hongbing HU (2008). “A spectral/temporal method for robust fundamental frequency tracking.” *The Journal of the Acoustical Society of America* 123.6, p. 4559–4571 (cité page 28).
- [Zeytinoglu 1995] ZEYTINOGLU, Mehmet et Kon Max WONG (1995). “Detection of harmonic sets”. *IEEE Transactions on Signal Processing* 43.11, p. 2618–2630.
- [Zhang 2011] ZHANG, Bin, Chris SCONYERS, Carl BYINGTON, Romano PATRICK, Marcos E. ORCHARD et George VACHTSEVANOS (2011). “A probabilistic fault detection approach : Application to bearing fault detection”. *IEEE Transactions on Industrial Electronics* 58.5, p. 2011–2018 (cité page 7).

---

**Résumé** — La surveillance automatique consiste à vérifier le bon fonctionnement d'un système tout au long de sa durée d'utilisation et ce, sans intervention humaine. Elle permet de mettre en place une stratégie de maintenance prévisionnelle qui présente un intérêt économique majeur, en particulier dans le cas de systèmes isolés comme les éoliennes construites en pleine mer. La surveillance automatique se base sur l'acquisition plus ou moins régulière de signaux pendant le fonctionnement du système surveillé. L'analyse de ces signaux doit permettre d'établir un diagnostic et de prendre une décision sur le déclenchement des opérations de maintenance. Dans cette thèse, nous proposons une méthode d'analyse générique permettant de s'adapter à n'importe quel système surveillé. La méthode se déroule en plusieurs étapes. Premièrement, chaque signal est analysé individuellement pour en extraire son contenu spectral, c'est-à-dire identifier les pics spectraux, les séries harmoniques et les bandes de modulation présents dans sa densité spectrale. Ensuite, ce contenu spectral est suivi au cours du temps pour former des trajectoires sur l'ensemble de la séquence de signaux acquis. Ces trajectoires permettent de générer des tendances qui sont le reflet de la santé du système. Enfin, les tendances sont analysées pour identifier un changement au cœur du système qui serait synonyme d'usure ou de défaut naissant. Cette méthodologie est validée sur de nombreux signaux réels provenant de la surveillance de différents systèmes mécaniques.

**Mots clés** : surveillance automatique, système mécanique, détection de défaut, série harmonique, bandes de modulation, suivi temps-fréquence, diagnostic.

---

**Abstract** — The automatic monitoring consists in verifying without any human intervention that a system is operating well. The monitoring allows to use a predictive maintenance strategy, which is economically interesting, especially in the case of isolated systems like offshore wind turbines. The automatic monitoring is based on signals acquired more or less regularly while the monitored system is operating. The analysis of these signals should be sufficient to diagnose the system and to decide whether or not the maintenance operations should be done. In this thesis, we propose a generic analysis method able to adapt itself to any monitored system. This method is composed by several steps. First, each signal is analyzed individually in order to extract its spectral content, that is to identify the spectral peaks, the harmonic series and the modulation sidebands presents in the signal spectrum. Then, the spectral content is tracked through time to construct spectral trajectories in the sequence of acquired signal. These trajectories are used to generate trends which indicate the state of the system health. Finally, the trends are analyzed to identify a change in the system response which would indicate some wear or a fault in its early stage. This analysis method is validated on real world signals acquired on different mechanical systems.

**Keywords** : monitoring, mechanical system, fault detection, harmonic series, modulation sidebands, time-frequency tracking, diagnosis.

---