No d'ordre : 2687



## THÈSE

#### présentée par

#### Noureddine ELHOR

pour obtenir le grade de Docteur de l'Université des Sciences et Technologies de Lille

Spécialité : Productique, Automatique et Informatique Industrielle

## Suivi de fonctionnement d'une éolienne par réseaux de neurones.

Soutenue le 17 Janvier 2000

C. VASSEUR	Professeur à l'USTL	Président de Jury
L. RADOUANE	Professeur à l'Université de Fès	Rapporteur
S. THIRIA	Professeur à l'Université de Versaille Saint-Quentin	Rapporteur
M. OULADSINE	MDC Habilité à l'UPJV	Examinateur
A. FAURE	Professeur à l'Université du Havre	Examinateur
A. EL HAJJAJI	MDC à l'UPJV	Examinateur
J. G. POSTAIRE	Professeur à l'USTL	Examinateur
J. M. GRAVE	Directeur du Département TS à NORELEC	Invité
D. HAMAD	Professeur à l'UPJV	Directeur de thèse

Thèse préparée au Laboratoire d'Automatique  $I^3D$ 





## Remerciements

L'ensemble des travaux présentés dans ce mémoire a été effectué au laboratoire d'Automatique I3D (ex. Centre d'Automatique de Lille) de l'Université des Sciences et Technologies de Lille.

J'exprime ma grande gratitude à Monsieur Christien VASSEUR, Professeur à l'USTL, qui m'a fait l'honneur de présider le jury de ma thèse.

C'est en grande partie grâce à Monsieur Denis HAMAD, Professeur à l'UPJV et mon Directeur de Recherche, que cette étude a abouti. Ses conseils, ses constants encouragements, sa grande disponibilité et son amitié m'ont permis de mener à bien ce travail. Qu'il trouve en ces quelques mots l'expression de mes remerciements les plus sincères.

Que Monsieur Jack-Gérard POSTAIRE, Professeur à l'USTL, trouve ici ma profonde reconnaissance pour l'aide et les encouragements qu'il m'a apporté durant mes travaux de thèse.

Ma profonde gratitude va également à Madame Sylvie THIRIA, Professeur à l'Université de Versaille-Saint Quentin, et à Monsieur Larbi RADOUANE, Professeur à l'Université des Sciences de Fès, pour l'intérêt qu'ils ont bien voulu porter à ce travail en acceptant de l'examiner et d'en être rapporteurs.

Je tiens à remercier Monsieur Alain FAURE, Professeur à l'Université de Le Havre, Monsieur Ahmed ELHAJJAJI, Maître de Conférence à l'UPJV, et Monsieur Mustapha OULADSINE, Maître de Conférence Habilité à l'UPJV, d'avoir accepté d'examiner ma thèse.

Je remercie également, l'ensemble du personnel du Laboratoire d'Automatique I3D pour l'amitié et le soutien qu'ils m'ont apportés durant le déroulement de la thèse.

A tous ceux là, je leur rends tout simplement hommage.

## Table des matières

1	Intr	oducti	ion géné	rale	8
2	Inst	rumer	itation d	'une éolienne pilote	12
	2.1	Introd	uction .		12
	2.2	Descri	ption de l	l'éolienne	13
	2.3	Princi	paux défa	uts et modes de défaillance d'une machine tournante .	14
		2.3.1	Principa	ux défauts	14
		2.3.2	Etude d	es modes de défaillance du système éolien	17
	2.4	Princi	paux capt	teurs installés	18
		2.4.1	Capteur	s installés sur le multiplicateur	19
			2.4.1.1	Accélération	19
			2.4.1.2	Température d'huile	19
		2.4.2	Capteur	s installés sur la génératrice	20
			2.4.2.1	Accélération	20
			2.4.2.2	Courant	20
		2.4.3	Capteur	s installés sur l'arbre pales-rotor	20
			2.4.3.1	Vitesse	20
			2.4.3.2	Référence de position	20
			2.4.3.3	Pas des pâles	20
		2.4.4	Capteur	s installés dans la nacelle	21
			2.4.4.1	Microphone	21
			2.4.4.2	Position angulaire de la nacelle	21
			2.4.4.3	Température d'ambiance	21

	2.5	Acqui	sition des signaux	21
		2.5.1	Acquisition avec un enregistreur de type DAT	22
		2.5.2	Description de la carte d'acquisition	30
		2.5.3	Décodage des signaux	31
	2.6	Procé	dure d'acquisition événementielle	32
		2.6.1	Principe	32
		2.6.2	Mise en oeuvre du système d'acquisition événementiel	33
		2.6.3	Acquisition événementielle avec la carte DT3001	33
	2.7	Conclu	usion	35
3	Ana	alyse sj	pectrale et extraction des attributs	36
	3.1	Introd	uction	36
	3.2	Transf	formée de Fourier	37
	3.3	Spectr	ogramme d'un signal simulé	40
	3.4	Spectr	ogramme de démarrage de l'éolienne	43
	3.5	Spectr	ogrammes du fonctionnement normal de l'éolienne	44
	3.6	Fréquences théoriques caractérisant le fonctionnement normal de l'éoli-		
		enne		48
		3.6.1	Rotor-Pales	49
		3.6.2	Multiplicateur	49
		3.6.3	Génératrice	51
	3.7	Extrac	ction d'attributs par réduction des spectrogrammes	52
	3.8	Conclu	usion	53
4	Rés	eaux d	le neurones autoassociateurs	54
	4.1	Introd	uction	54
	4.2	Réseau	u autoassociateur linéaire à trois couches	58
		4.2.1	Architecture	58
		4.2.2	Equations de fonctionnement	58
		4.2.3	Apprentissage par la méthode d'analyse en composantes prin-	
			cipales	61

	4.2.4	Critère d'erreur	61		
		4.2.4.1 Calcul des biais de la couche cachée	62		
		Rôle du biais :	62		
		4.2.4.2 Calcul des poids des connexions couche cachée - couche			
		de sortie	63		
		4.2.4.3 Calcul des poids des connexions couche entrée-couche			
		cachée	64		
	4.2.5	Exemple d'application	66		
4.3	Réseau	autoassociateur non linéaire à 3 couches	70		
	4.3.1	Architecture	70		
	4.3.2	Equations de fonctionnement	70		
	4.3.3	Apprentissage par la méthode de rétropropagation du gradient			
		de l'erreur	71		
	4.3.4	Gradient du critère d'erreur partielle	72		
	4.3.5	Adaptation des poids des connexions couche cachée-couche de			
		sortie	72		
	4.3.6	Adaptation des poids des connexions couche d'entrée-couche			
		cachée	73		
	4.3.7	Comparaison entre les réseaux autoassociateurs linéaires et			
		non linéaires	74		
	4.3.8	Exemple d'application	75		
4.4	Réseau	autoassociateur non linéaire à cinq couches	77		
	4.4.1	Architecture	77		
	4.4.2	Equations de fonctionnement	78		
	4.4.3	Apprentissage par rétropropagation du gradient de l'erreur	79		
		4.4.3.1 Gradient du critère d'erreur partielle	79		
	4.4.4	Optimisation de l'architecture du réseau	80		
	4.4.5	Exemple d'application	81		
		4.4.5.1 Optimisation de l'architecture	81		
		4.4.5.2 Simulation d'une dérive de fonctionnement	84		

	4.5	Concl	usion	87
5	Suiv	vi de f	onctionnement de l'éolienne par RNA	88
	5.1	Introd	luction	88
	5.2	Const	itution de la base d'apprentissage et de généralisation	90
		5.2.1	Réduction des spectrogrammes des signaux de vibrations de	
			l'accéléromètre installé sur le multiplicateur en position axiale	90
		5.2.2	Réduction des spectrogrammes des signaux de vibrations issus	
			de l'accéléromètre installé sur le multiplicateur en position	
			radiale	103
		5.2.3	Réduction des spectrogrammes des signaux de vibrations issus	
			de l'accéléromètre installé sur la génératrice	109
	5.3	Suivi o	de fonctionnement du multiplicateur par réseaux autoassociateurs	3115
		5.3.1	Autoassociateur linéaire à trois couches	115
		5.3.2	Autoassociateur non linéaire à trois couches	122
		5.3.3	Autoassociateur non linéaire à cinq couches	125
	5.4	Suivi	de fonctionnement de la génératrice par réseaux autoassociateurs	s129
		5.4.1	Autoassociateur linéaire à trois couches	129
		5.4.2	Autoassociateur non linéaire à trois couches	131
		5.4.3	Autoassociateur non linéaire à cinq couches	133
5.5 Simulation d'un		Simula	ation d'une dérive de fonctionnement du système éolien	135
		5.5.1	Constitution de la base de test	135
		5.5.2	Apparition d'une composante parasite dont la fréquence se	
			situe dans une bande étroite	137
		5.5.3	Apparition d'une composante parasite dont la fréquence se	
			situe dans une bande large	142
	5.6	Conclu	usion $\ldots$	147
6	Rés	eaux d	le Neurones Modulaires	148
	6.1	Introd	luction	148
	6.2	Archit	ecture et équations de fonctionnement	150

	6.2.1 Architecture			
	6.2.2	Equations de fonctionnement	151	
6.3	Algorithme d'apprentissage			
	6.3.1	Adaptation des poids des modules experts	156	
	6.3.2	Adaptation des poids du module contrôleur	158	
6.4	3.4 Optimisation de l'architecture du réseau modulaire			
6.5	Exemples d'application		160	
	6.5.1	Exemple 1 : Deux classes en forme de croissant.	160	
	6.5.2	Exemple 2 : Deux classes en Ou-Exclusif	170	
6.6	Application à la surveillance du multiplicateur de l'éolienne 178			
6.7	Conclu	sion	184	
a .			105	
Conclu	Conclusion générale et perspectives 185			

#### Annexes

#### Bibliographie

202

189

## Chapitre 1

### Introduction générale

Dans le développement durable, l'énergie éolienne occupe une place de choix du fait que c'est une source d'énergie renouvelable, non polluante et exigeant peu de maintenance. De plus, le coût du KWatt/H, qui a été le principal handicap au développement de cette énergie, n'a cessé de diminuer rendant cette solution une alternative intéressante.

Ces dernières années, de nombreux pays ont mis en place des programmes nationaux de développement éoliens avec plus de 30000 éoliennes installées dans le monde entier [Dév94]. La figure 1.1 résume l'évolution mondiale de la "puissance installée" entre 1994 et 2000 et montre que celle-ci est en constante évolution.

En France, le parc éolien a été pendant longtemps presque inexistant. Récemment, cette énergie a pris de plus en plus d'importance avec le contexte écologique actuel et on assiste à la naissance d'une filière industrielle française.

Toutefois, pour assurer une bonne exploitation et une large diffusion de l'énergie éolienne, il est nécessaire de bien maîtriser le coût de production et d'optimiser le rendement qui restent largement liés aux conditions climatiques et à la disponibilité des machines. A cette fin, une maintenance préventive et prédictive basée sur un système de suivi de fonctionnement "*monitoring system*" s'avère nécessaire.

Le travail effectué dans le cadre de la préparation de cette thèse s'inscrit dans



FIG. 1.1 : Evolution mondiale de la puissance installée entre 1994 et 2000.

un programme de recherche et développement sur l'optimisation du fonctionnement de l'éolienne de Malo les Bains (Région Nord-Pas de Calais). L'objectif de ce programme est de suivre le fonctionnement d'une centrale éolienne afin de mettre en place une politique efficace de maintenance préventive et prédictive.

Pour analyser le comportement d'une éolienne, il convient, à partir de mesures et de traitements appropriés, d'être capable de caractériser son fonctionnement. Afin de disposer de données pertinentes, des campagnes de mesures ont été effectuées pour obtenir, grâce à une instrumentation poussée de l'éolienne, des informations caractéristiques sur son état et ses modes de fonctionnement.

Le but est de chercher à différencier des modes de fonctionnement de l'éolienne selon qu'ils correspondent à un fonctionnement normal, à un dysfonctionnement, voire à une situation de panne. Généralement, les automaticiens utilisent des modèles de connaissance afin de comparer le fonctionnement réel au comportement de ces modèles soumis aux mêmes sollicitations.

De nombreux systèmes de suivi de fonctionnement en temps réel ont été implan-

tés avec succès dans différents domaines industriels tels que la supervision des systèmes mécaniques [Zwi95], des centrales nucléaires [PMA94], dans l'industrie chimique [WMA<sup>+</sup>89] et l'industrie automobile [ROG97] afin d'améliorer la sûreté et la fiabilité de leur fonctionnement.

Ces systèmes de suivi et de diagnostic nécessitent généralement la connaissance d'un modèle physique entrées-sorties du processus à surveiller. Or, dans le cas des éoliennes, il serait difficile, sinon irréaliste, de tenter de modéliser le fonctionnement sous la forme d'un modèle classique de connaissance. En effet, une éolienne est un ensemble constitué de sous-systèmes très différents dont les comportements sont liés aux conditions atmosphériques locales et instantanées, très difficiles à connaître avec précision. On pense en particulier au rotor dont les pales sont sollicitées par le vent qui est souvent très perturbé et à la nacelle qui doit suivre les variations d'orientation du vent. On peut aussi citer la diversité de conception des éoliennes rendant le développement d'un système de suivi générique très délicat à réaliser. De ce fait, les méthodes basées sur la connaissance d'un modèle entrées-sorties ne peuvent pas être exploitées [CAS 94].

Pour pallier cette difficulté méthodologique qui interdit une modélisation classique, nous avons soumis les informations délivrées par les capteurs installés sur l'éolienne à une analyse fine afin d'identifier des modes de fonctionnement normaux. Pour atteindre cet objectif de surveillance en ligne, nous avons mis en oeuvre des architectures neuronales spécifiques qui permettent de délimiter ces modes de fonctionnement normaux et de cerner les origines des pannes et avaries en suivant le comportement de l'éolienne. Au cours d'une phase d'apprentissage, la synthèse multidimensionnelle de ces données permettra de mettre en évidence les différents modes de fonctionnement, chacun d'eux ayant ses caractéristiques propres.

En cours d'exploitation, il s'agirait de déterminer, en continu, dans quel mode de fonctionnement se trouve l'éolienne. Ces informations seront exploitées pour assurer une meilleure disponibilité de la machine grâce à une stratégie de maintenance prédictive. En effet, toute dérive d'un mode de fonctionnement normal vers un mode correspondant à un état anormal archivé au cours de l'apprentissage, ou inconnu, permettra soit d'alerter l'opérateur qui entreprendra l'action corrective nécessaire, soit de provoquer une mise en état de sauvegarde de la machine, sans perdre les objectifs de production.

Le contenu de ce travail est structuré en six chapitres :

- Le chapitre 2, décrit l'installation dédiée à la collecte des informations concernant l'état de fonctionnement de l'éolienne.
- Le chapitre 3, décrit la procédure d'extraction des attributs pertinents caractérisant son fonctionnement obtenue par une analyse spectrale des signaux issus des différents capteurs installés.
- Dans le chapitre 4, une étude théorique portant sur les méthodes de visualisation et de reconstruction des données multidimensionnelles est développée grâce à des réseaux de neurones autoassociateurs.
- Le chapitre 5 concerne le suivi de fonctionnement de l'éolienne par analyse vibratoire. Des réseaux autoassociateurs sont employé à la fois pour la visualisation plane et pour la reconstruction multidimensionnelle des signaux.
- Dans le dernier chapitre, nous développons une architecture neuronale dite modulaire bien adaptés aux problèmes de classification supervisée. Ce type d'architecture est utilisé afin de décider si le point de fonctionnement actuel est situé à l'intérieur d'un domaine de fonctionnement normal ou s'il dérive vers un mode de fonctionnement anormal.

## Chapitre 2

## Instrumentation d'une éolienne pilote

#### 2.1 Introduction

La compréhension du comportement des générateurs éoliens est essentielle pour la mise en oeuvre d'un système de suivi de fonctionnement en vue de leurs maintenances. Dans une première phase, une instrumentation poussée de l'éolienne pilote a été réalisée suivie d'un système de collecte de données. Le choix des capteurs a été guidé par une étude AMDEC (Analyse des Modes de Défaillances, de leurs Effets et de leurs Criticités). Comme il n'est pas possible de connaître les défauts typiques d'une éolienne, il nous a semblé naturel de surveiller les modes de fonctionnement normal de la machine. Pour cela, il est nécessaire de bien définir auparavant le domaine de fonctionnement normal. Dans ce sens, nous considérons les signaux acquis lors de nos campagnes de mesure comme étant normaux vu que la machine est assez récente.

Ce chapitre est organisé de la façon suivante : la section 2 décrit le système éolien et la section 3 résume les différents défauts typiques des composants d'une machine tournante. Une instrumentation de l'éolienne conforme aux conclusions d'une étude AMDEC (Analyse des Modes de Défaillance et de leur Effet et de leur Criticité) est exposée dans la section 4. Enfin, les sections 5 et 6 sont consacrées à la description du système d'acquisition des signaux et à la procédure de sélection des événements à enregistrer.

#### 2.2 Description de l'éolienne

Une éolienne est un système mécanique de production d'énergie électrique. L'éolienne, objet de notre étude, est une machine à trois pales, de diamètre 30 mètres et de puissance 300KW. Elle est constituée d'une nacelle supportée par un mât de 21,7 mètres de hauteur. Cette nacelle comporte un arbre lent qui supporte les trois pales, un multiplicateur de vitesse et une génératrice asynchrone. Une vue globale des principaux éléments de l'éolienne est présentée figure 2.1.



FIG. 2.1 : Vue générale des éléments constituant une éolienne.

Le rotor à trois pales assure le démarrage de la machine. La régulation par pas fixe de l'orientation des pales se fait par l'intermédiaire d'une palette située dans le même plan que le rotor, donc normale à la direction du vent. Pour éviter que la machine ne soit endommagée lorsque le vent atteint des vitesses importantes, cette palette place les pales dans le lit du vent (mise en position drapeau), provoquant ainsi l'arrêt de l'éolienne. D'une manière générale, le système éolien présente l'inconvénient de ne pas fonctionner pour des vents faibles, inférieurs à 5 m/s, ni pour des vents très forts supérieurs à 18 m/s. Son exploitation n'est donc efficace que dans les zones géographiques à vents moyens réguliers et à durées significatives.

Les structures de l'éolienne sont sujettes à des conditions extrêmes pouvant produire des arrêts intempestifs, voire des défaillances conduisant à son endommagement. Il est donc important d'étudier et d'analyser les principaux défauts de ses composants afin de prévoir d'une manière précoce tout dysfonctionnement éventuel.

## 2.3 Principaux défauts et modes de défaillance d'une machine tournante

#### 2.3.1 Principaux défauts

Il est impossible de connaître les défauts et les modes de défaillances d'une éolienne en pleine production sur laquelle, pour des raisons de sécurité, on ne peut pas provoquer des pannes. En effet, l'étude des défauts et leurs impacts ne peut s'effectuer que sur des machines en essai dans un laboratoire ou par simulation. Pour cela, dans la suite, nous présentons les principaux défauts typiques des machines tournantes bien connus dans la littérature [BP95].

• Au niveau du rotor : En pratique, il est impossible d'obtenir une concentricité parfaite des centres de gravité de chaque élément constitutif d'un rotor. Cette non-concentricité provenant, généralement de défauts d'usinage, d'assemblage et de montage, conduit à l'application de forces centrifuges qui déforment le rotor. Ce déséquilibre va donc induire une vibration, dans le plan radial, dont le spectre présente une composante d'amplitude prépondérante et de fréquence fondamentale correspondant à la fréquence de rotation du rotor (cf. figure 2.2(a)). Un autre type de défauts dit défaut d'alignement ou de flexion, est l'une des principales causes de réduction de la durée de vie du rotor. Ce défaut génère des efforts importants entraînant la dégradation rapide du système d'accouplement rotormultiplicateur (cf. figure 2.2 (b)). Ceci se manifeste le plus souvent par la présence, dans la direction radiale, de composantes d'ordre 2,3 ou 4 de la fréquence de rotation avec des amplitudes supérieures à celle de la composante d'ordre 1.



FIG. 2.2 : Image vibratoire théorique : (a) d'un balourd, (b) d'un défaut d'alignement.

D'autres défauts typiques du rotor peuvent trouver leurs origines dans la rugosité de la surface, le gel, les fissures, etc... Ces défauts peuvent contribuer à une augmentation significative du bruit et provoquer ainsi une dégradation de la qualité de l'énergie produite.

• Au niveau du multiplicateur de vitesse : L'inconvénient de la conception d'une éolienne asynchrone tournant à une vitesse constante est la haute charge mécanique que supportent ses composants, en particulier le multiplicateur de vitesse. Le multiplicateur est souvent considéré comme l'élément le plus fragile et le plus coûteux pour des machines dont la puissance est de l'ordre du Megawatt. Selon les rapports de maintenance, le multiplicateur de vitesse doit être remplacé tous les cinq à dix ans, ce qui influe fortement sur le prix des KWh produits [NEG<sup>+</sup>96].

Les défauts au niveau d'un multiplicateur sont généralement dus à l'usure des roulements et des arbres de transmission ainsi qu'à la détérioration des dents des engrenages. Ces défauts peuvent être détectés par une analyse des vibrations mécaniques, ou par une mesure de la température d'huile.

Considérons un engrenage d'un multiplicateur composé de deux couronnes dentées, possédant  $n_1$  et  $n_2$  dents et tournant aux vitesses  $v_1$  et  $v_2$ , c'est à dire aux fréquences de rotation  $f_1$  et  $f_2$ . L'engrènement se fait au rythme d'engagement des dents selon une fréquence multipliée par le nombre de dents :

$$f_{eng} = n_1 f_1 = n_2 f_2.$$

Si la denture est correcte et si aucun phénomène parasite ne vient perturber l'engrènement, le spectre vibratoire est constitué de composantes dont les fréquences correspondent à la fréquence d'engrènement  $f_{eng}$  et à ses harmoniques. Cependant, si l'une des deux roues possède une dent détériorée, il se produit un choc périodique à la fréquence de rotation de la roue considérée. Le spectre correspondant comporte, non seulement la fréquence d'engrènement, mais aussi une pigne de raies dont le pas correspond à cette fréquence de rotation. Les figures 2.3 (a) et 2.3 (b) donnent respectivement l'image vibratoire théorique d'un engrenage sain et d'un engrenage présentant une dent détériorée.

• Au niveau de la génératrice : Les déséquilibres électriques sont les types de défauts les plus fréquemment rencontrés sur les génératrices asynchrones. De même, ces défauts peuvent être détectés par une analyse spectrale de ses vibrations.

Afin d'assurer une bonne protection de l'éolienne contre ces différents défauts, il est souhaitable de concevoir un système de suivi de son état de fonctionnement qui prévient de toute dérive ou incident. La mise en oeuvre d'un tel système exige l'installation de certains capteurs pour constituer une base de connaissances sur le



FIG. 2.3 : Image vibratoire théorique : (a) d'un engrenage sain, (b) d'un engrenage présentant une dent détériorée.

fonctionnement normal du système. Pour décider du choix des composants sensibles à surveiller ainsi que les capteurs à installer, une étude AMDEC (Analyse de Modes de Défaillance, de leur Effet et leur Criticité) s'avère nécessaire.

#### 2.3.2 Etude des modes de défaillance du système éolien

L'analyse des modes de défaillance du système éolien peut être guidée par une approche structurée de type AMDEC qui a fait ses preuves dans l'automobile, l'aéronautique, le nucléaire, etc...

Cette approche a pour objet d'évaluer les risques de défaillance des éléments constituant le système éolien et leurs conséquences. Elle permet ainsi de révéler les composants sensibles à surveiller.

La méthodologie AMDEC consiste à :

- recenser toutes les défaillances dont les conséquences affectent le fonctionnement de l'éolienne,
- prévoir l'occurrence de la défaillance de chaque composant de l'éolienne,

- déterminer la gravité des effets de chacune des défaillances recensées,
- trouver les moyens de détecter les défaillances,
- rechercher et décrire les conditions susceptibles d'engendrer ces défaillances.

En d'autres termes, il s'agit de répondre aux cinq questions suivantes :

- 1. quel composant, dans une éolienne, risque de ne pas fonctionner correctement ?
- 2. que se passera-t-il en cas de défaillance?
- 3. quelle est la probabilité pour que cette défaillance se manifeste à court, moyen ou long terme?
- 4. est-ce grave?
- 5. quelle sont les conditions susceptibles d'engendrer ces défaillances?

L'instrumentation de l'éolienne pilote dépend des résultats de l'analyse de ces réponses.

#### 2.4 Principaux capteurs installés

Notre étude AMDEC nous a révélé que le système de suivi de fonctionnement et de diagnostic d'une éolienne nécessite la surveillance des éléments en rotation. Plus précisément, il s'agit de surveiller les vibrations du multiplicateur et de la génératrice, la vitesse de rotation du rotor, les déplacements de la nacelle, la température d'huile du multiplicateur etc... Pour cela, différents types de capteurs doivent être installés sur les principaux composants de l'éolienne. Leurs choix et leurs emplacements sont d'une importance capitale. En effet, les signaux issus de ces capteurs doivent traduire les changements inhérents de l'état de fonctionnement de la machine.

Pour surveiller les différents éléments sensibles de l'éolienne, nous avons installé les capteurs suivants [NEG<sup>+</sup>96] :

• Deux accéléromètres, l'un en position axiale parallèle à l'axe du rotor et l'autre en position radiale perpendiculaire à cet axe, mesurant les vibrations sur le multiplicateur de vitesse.

- Un accéléromètre en position axiale qui peut être déplacé en position radiale, mesurant les vibrations de la génératrice.
- Un codeur incrémental, couplé à l'arbre lent, mesurant la vitesse de rotation du rotor.
- Un capteur de déplacement mesurant le pas des pâles.
- Une pince ampèremétrique mesurant le courant délivré par la génératrice.
- Deux pinces mesurant les tensions sur les deux phases du courant triphasé.
- Un microphone à l'intérieur de la nacelle mesurant le bruit généré par l'éolienne.
- Deux capteurs mesurant la température de l'huile du multiplicateur et la température ambiante à l'intérieur de la nacelle.
- Trois capteurs de référence mesurant la position angulaire de la nacelle.

#### 2.4.1 Capteurs installés sur le multiplicateur

#### 2.4.1.1 Accélération

Deux accéléromètres en position axiale et radiale du multiplicateur génèrent des données reflétant le comportement vibratoire du multiplicateur. Ce type de capteur est en effet très coûteux pour permettre une instrumentation complète et permanente de chaque composant de l'éolienne. On pourra, éventuellement le déplacer au cours de l'étude sur d'autres composants.

#### 2.4.1.2 Température d'huile

Ce capteur est une sonde de température industrielle de type Pt100 plongée dans l'huile du multiplicateur.

#### 2.4.2 Capteurs installés sur la génératrice

#### 2.4.2.1 Accélération

Un accéléromètre installé dans la même direction que l'arbre de la génératrice permet de mesurer les vibrations de cette dernière.

#### 2.4.2.2 Courant

Outre la mesure des vibrations de la génératrice asynchrone triphasée, nous effectuons la mesure du courant sur l'une de ses phases lors de son couplage avec le réseau EDF. Cette mesure, effectuée par une pince ampèremétrique, servira à détecter des défauts électriques sur cette phase ou des défauts mécaniques par analyse du spectre du courant de sortie.

#### 2.4.3 Capteurs installés sur l'arbre pales-rotor

#### 2.4.3.1 Vitesse

La vitesse du rotor est mesurée par un codeur incrémental délivrant 50 impulsions par tour, couplé à l'arbre lent par un galet de 50 cm de développement. Le codeur délivre trois signaux destinés respectivement au comptage d'impulsions, à la détermination du sens de rotation et à la mise à zéro des compteurs à chaque tour d'axe. L'échelle de mesure, définie par le rapport entre l'angle de rotation du rotor et le nombre d'impulsions, est fonction du diamètre de l'axe de ce rotor.

#### 2.4.3.2 Référence de position

La référence de position du rotor est donnée par un capteur inductif placé en parallèle avec un autre capteur fixé sur la machine.

#### 2.4.3.3 Pas des pâles

Le déplacement par pas des pâles est commandé par un vérin hydraulique. Un potentiomètre linéaire fournit la position de ce vérin.

#### 2.4.4 Capteurs installés dans la nacelle

#### 2.4.4.1 Microphone

Un microphone installé à l'intérieur de la nacelle permet de connaître l'intensité du bruit régnant dans la machine.

#### 2.4.4.2 Position angulaire de la nacelle

Le système d'orientation de la nacelle comprend une roue dentée de 90 dents ce qui nous donne un pas de 4 degrés. Deux capteurs sont placés devant cette roue, légèrement décalés l'un par rapport à l'autre de façon à fournir des impulsions décalées dans le temps. Cela nous permet de déterminer le sens de rotation de la nacelle. Un troisième capteur indique la référence de position de la nacelle. Il est donc physiquement lié à la nacelle et est activé lors de son passage devant un index métallique fixé sur la paroi intérieure du mât.

#### 2.4.4.3 Température d'ambiance

La température de l'huile est une valeur relative qui doit être mise en relation avec la température ambiante pour avoir une réelle signification. C'est pourquoi un capteur de température de l'air ambiant est installé dans la nacelle pour servir de référence.

Les signaux issus des différents capteurs sont enregistrés grâce à un système d'acquisition et peuvent être visualisés en ligne sur l'écran d'un ordinateur installé au pied du mât de l'éolienne.

#### 2.5 Acquisition des signaux

L'acquisition des informations relatives au comportement de l'éolienne est une étape importante de l'étude. Elle a pour but de disposer d'un ensemble de signaux caractérisant les différents modes de fonctionnement de l'éolienne. Pour ce faire, deux systèmes d'acquisition ont été employé : un enregistreur DAT (Digital Audio Tape) et une carte d'acquisition.

Les signaux à enregistrer peuvent être répertoriés en deux catégories : les signaux analogiques et les signaux tout ou rien. Ces derniers évoluant très lentement peuvent donc être échantillonner à une fréquence réduite.

Le choix de la fréquence d'échantillonnage des signaux analogiques implique, en pratique, une connaissance a priori de leurs étendues spectrales. Pour ce faire, une campagne de mesures a été menée à l'aide d'un enregistreur DAT de fréquence d'échantillonnage  $f_e$  égale à 48KHz. Une analyse spectrale des signaux, va nous permettre d'avoir une idée sur leurs étendues spectrales et par suite, de déterminer les fréquences maximales d'échantillonnage sur chacune des entrées de la carte d'acquisition.

#### 2.5.1 Acquisition avec un enregistreur de type DAT

L'enregistreur DAT ayant pour fréquence d'acquisition  $f_e = 48$ KHz a été choisi pour étudier l'étendue des spectres des signaux issus des différents capteurs. Les enregistrements ont été réalisés le 02/07/96, alors que l'éolienne était connectée au réseau EDF. La vitesse du vent était relativement faible, de l'ordre de 5m/s, et variable en direction, entre Sud-Est et Sud-Sud-Est.

Les signaux enregistrés avec le DAT sont ceux issus des cinq capteurs suivants :

- 1. Microphone,
- 2. Accéléromètre placé en position axiale sur le multiplicateur,
- 3. Accéléromètre placé en position radiale sur le multiplicateur à 90 degré par rapport au précédent,
- 4. Accéléromètre placé en position axiale sur la génératrice,

5. Pince ampèremétrique placée sur l'une des trois phases de sortie de la génératrice.

Les représentations temporelles et spectrales des signaux issus des cinq capteurs sont données dans les figures 2.4 à 2.13.

Un examen attentif des différents signaux et une analyse préliminaire de leurs spectres montrent que ceux-ci n'ont pas les mêmes étendues. Par exemple, dans le cas des signaux issus du microphone, la plus haute fréquence utile contenue dans le spectre se trouve à moins de 3KHz (cf. figure 2.5). Par contre, dans le cas des signaux issus des accéléromètres du multiplicateur, le spectre vibratoire s'étend jusqu'à 6KHz (cf. figure 2.7). Le spectre le plus riche est celui de la génératrice puisqu'il s'étend jusqu'à 11KHz.



FIG. 2.4 : Signal acoustique issu du microphone, la fréquence d'échantillonnage  $f_e$ =48KHz.



FIG. 2.5: Spectre d'amplitude du microphone calculé sur une fenêtre de 32768 échantillons.



FIG. 2.6 : Signal d'accélération du multiplicateur issu de l'accéléromètre axial, la fréquence d'échantillonnage  $f_e$ =48KHz.



**FIG. 2.7 :** Spectre d'amplitude du signal d'accélération du multiplicateur issu de l'accéléromètre axial calculé en utilisant une fenêtre de 32768 échantillons.



FIG. 2.8 : Signal d'accélération du multiplicateur issu de l'accéléromètre radial, la fréquence d'échantillonnage  $f_e$ =48KHz.



**FIG. 2.9 :** Spectre d'amplitude du signal d'accélération du multiplicateur issu de l'accéléromètre radial calculé en utilisant une fenêtre de 32768 échantillons.



FIG. 2.10 : Signal d'accélération de la génératrice issu de l'accéléromètre axial, la fréquence d'échantillonnage  $f_e$ =48KHz.



**FIG. 2.11 :** Spectre d'amplitude du signal d'accélération de la génératrice issu de l'accéléromètre axial calculé en utilisant une fenêtre de 32768 échantillons.



FIG. 2.12 : Signal courant mesuré sur une phase de la génératrice, la fréquence d'échantillonnage  $f_e$ =48KHz.



**FIG. 2.13 :** Spectre d'amplitude du signal courant calculé sur une fenêtre de 32768 échantillons.

Le théorème d'échantillonnage de Shannon nous renseigne sur la valeur minimale de la fréquence d'échantillonnage d'un signal analogique. Celle-ci doit être supérieure ou égale à deux fois la fréquence maximale utile contenue dans le signal. Le non respect de cette condition aboutit au repliement du spectre du signal échantillonné, phénomène bien connu dans la littérature de traitement du signal [Max97].

Les valeurs minimales des fréquences d'échantillonnage des signaux issus des différents capteurs sont regroupées dans le tableau de la figure 2.14.

Capteurs analogiques	Fréquence Max. utile	Fréquence Min. d'échantillonnage
Les signaux rapides		
1. Micro	3 KHz	≥ 8 KHz
2. Accéléromètre axial du	8 KHz	$\geq 21~{ m KHz}$
multiplicateur		
3. Accéléromètre radial	8 KHz	≥ 21 KHz
du multiplicateur		
4. Accéléromètre de la	11 KHz	$\geq 30 \text{ KHz}$
génératrice		
5. Pince Ampèremétrique	1 KHz	≥ 2.7 KHz
Les signaux lents	· · · · · · · · · · · · · · · · · · ·	
6. Température d'huile	0.1 Hz	$\geq 0.27 \text{ Hz}$
7. Température ambiante	0.1 Hz	$\geq 0.27~{ m Hz}$
8. Pas de l'hélice	10 Hz <sup>1</sup>	≥ 27 Hz

FIG. 2.14 : Fréquences maximales d'échantillonnage pour les capteurs analogiques.

<sup>&</sup>lt;sup>1</sup>A priori, nous n'avons pas d'informations suffisantes pour calculer la fréquence maximale de façon sûre.

Capteurs tout ou rien	Fréquence Max. utile	Fréquence Min. d'échantillonnage
9. Codeur vitesse de rotor	$400 \text{ imp/s}^1$	$\geq 1 \ \mathrm{KHz^3}$
10. Inductif de référence	$3 \text{ imp/s}^1$	$\geq 1 \ \mathrm{KHz^3}$
de position du rotor		
11. Inductif de rotation	inconnue	$\geq 1 \ \mathrm{KHz^3}$
de la nacelle 1 <sup>2</sup>		
12. Inductif de rotation	inconnue	$\geq 1 \ \mathrm{KHz^3}$
de la nacelle 2 <sup>2</sup>		
13. Inductif de référence	inconnue	$\geq 1 \ \mathrm{KHz^3}$
de position de la nacelle <sup>2</sup>		

Dans le tableau de la figure 2.15 nous avons regroupé les fréquences d'échantillonnage des signaux issus des différents capteurs tout ou rien :

FIG. 2.15 : Fréquences maximales d'échantillonnage pour les capteurs tout ou rien.

Par conséquent, il nous paraît naturel de stocker les différents signaux à des fréquences d'échantillonnage distinctes. Pour ce faire nous avons employé une carte d'acquisition multi-entrées. Chaque entrée peut être échantillonnée à une fréquence propre.

#### 2.5.2 Description de la carte d'acquisition

Le choix des caractéristiques de la carte d'acquisition dépend de la nature des signaux à enregistrer. Pour ce faire, nous avons effectué une analyse spectrale préliminaire des signaux enregistrés par le DAT qui nous a permis de déterminer les

<sup>&</sup>lt;sup>2</sup>Les signaux issus de ces deux capteurs doivent être échantillonnés en permanence. En effet, la position de la nacelle est calculée par incrément/décrément d'un capteur à chaque impulsion des capteurs 11 et 12. Chaque impulsion manquée entraîne une erreur de  $\pm 4^{\circ}$  sur la position de la nacelle, qui ne pourra être corrigée qu'au passage à la position de référence (capteur 13).

<sup>&</sup>lt;sup>3</sup>Quelque soit la carte d'acquisition, toutes les entrées TTL sont échantillonnées au même instant, et donc à la même cadence. La fréquence indiquée ici est celle de l'entrée TTL la plus rapide (capteur 9).

fréquences maximales contenues dans ces spectres et par suite les fréquences minimales d'échantillonnage illustrées dans les deux tableaux (cf. figures 2.14 et 2.15). Ces fréquences d'échantillonnage sont obtenues en respectant la condition de Shannon. Conformément aux deux tableaux résumant les fréquences minimales d'échantillonnage des différents signaux, nous avons retenu le modèle de carte DT3001 qui présente les caractéristiques suivantes :

- 16 entrées d'acquisition analogiques et 8 entrées logiques en technologie TTL : Les entrées analogiques peuvent être configurées comme 16 entrées indépendantes ou 8 entrées différentielles. Ce dernier mode est recommandé pour les environnements bruités car il élimine le mode commun. Le gain pour chacune des entrées analogiques peut être adapté à des signaux d'amplitudes de 50mV à 10V.
- La fréquence d'acquisition peut-être ajustée pour chacune des voies permettant ainsi d'éviter un sur-échantillonnage des signaux à variations lentes.
- L'échantillonnage continu de toutes les entrées à une cadence maximale de 250.000 échantillons par seconde.
- L'acquisition se déclenche lorsqu'un seuil particulier est atteint sur une des entrées analogiques.

#### 2.5.3 Décodage des signaux

Une caractéristique intéressante de la carte DT3001 est qu'elle permet d'échantillonner les signaux d'entrées à des fréquences différentes. En effet, les signaux enregistrés via la carte d'acquisition se présentent sous la forme de trames. Chaque trame est composée d'un nombre de valeurs égal au nombre d'entrées câblées de la carte d'acquisition. Ainsi, l'acquisition en continu de huit entrées est réalisée comme suit :

Puisque la carte a une bande passante, notée  $B_P$  de 250KHz, les signaux sont tous échantillonnés à la fréquence  $\frac{B_P}{N_c}$  KHz, où  $N_c$  désigne le nombre d'entrées. Dans le cas des signaux présents sur les entrées 1 et 2 (cf. figure 2.16), on a  $T_1 = T_2$ . Il



FIG. 2.16 : Cas 1 : Tous les signaux sont échantillonnés à la même fréquence.

est cependant possible d'acquérir plusieurs fois le même signal au sein d'une même trame. Dans ce cas, tous les signaux sont échantillonnés à  $\frac{B_P}{N_c}$ , à l'exception du signal présent sur l'entrée 1 qui l'est à une fréquence double, puisqu'il est présent deux fois dans la même trame ( $T_1 = T_2/2$ ).



FIG. 2.17 : Cas 2 : Les signaux sont échantillonnés à des fréquences différentes.

#### 2.6 Procédure d'acquisition événementielle

Pour recueillir toutes les informations nécessaires à l'étude, une procédure d'acquisition événementielle a été mise au point. Elle permet de contrôler les acquisitions des signaux de manière autonome.

#### 2.6.1 Principe

Le suivi de fonctionnement d'une éolienne nécessite l'élaboration d'une base de données représentant les différents modes de fonctionnement de cette machine. Le volume des informations recueillies par les capteurs est considérable. Par conséquent, il n'est pas possible de tout stocker de manière continue, et il est nécessaire de mettre en oeuvre une procédure de sélection lors de la collecte des données.

De plus, le traitement, même hors-ligne, de cette masse de données sera fastidieux. Il est donc nécessaire de pré-sélectionner les données que nous souhaitons enregistrer, éliminant ainsi les informations jugées non intéressantes.

Afin de réaliser cette pré-sélection, nous avons choisi de préciser quelles sont les informations à enregistrer.

#### 2.6.2 Mise en oeuvre du système d'acquisition événementiel

L'éolienne est conçue pour fonctionner sous des régimes de vent, des températures et des puissances différentes. Pour avoir une base de données suffisamment riche, il est nécessaire de recenser ces différents modes de fonctionnement. Pour cela, nous avons adopté une procédure d'enregistrement basée sur l'apparition de certains événements tels que :

- Raccordement de l'éolienne sur le réseau EDF,
- Régime du vent,
- Date et heure d'enregistrement,
- Conditions thermiques dues au lever et au coucher du soleil.

Un avantage très important de l'approche adoptée est la possibilité d'affiner au fur et à mesure notre connaissance sur les modes de fonctionnement du système éolien. Ceci permet d'obtenir progressivement une base de données représentative de la plupart des modes de fonctionnement de la machine.

#### 2.6.3 Acquisition événementielle avec la carte DT3001

Dans le cadre d'une étude de faisabilité et vue l'importance des flux d'informations enregistrées, nous nous sommes d'abord intéressés à l'étude des signaux de vibrations issus des capteurs d'accélération installés sur le multiplicateur en position axiale et radiale et sur la génératrice en position axiale.

Dans l'intérêt d'obtenir une base de données qui reflète le mieux possible l'ensemble des modes de fonctionnement de l'éolienne, nous avons effectué plusieurs campagnes de mesure grâce à la carte d'acquisition DT3001 et la procédure d'acquisition événementielle. Tous les enregistrements ont été réalisés après couplage de l'éolienne avec le réseau EDF. Afin de ne pas avoir des enregistrements très longs représentant le même mode de fonctionnement, la durée d'enregistrement a été fixée à 15 secondes. Ensuite, l'acquisition est suspendue pendant quelques minutes.

#### 2.7 Conclusion

Dans ce chapitre nous avons tout d'abord décrit les différents composants de l'éolienne de Malo les Bains et les capteurs qui ont été installés. Comme il n'est pas possible de créer des défauts sur cette machine et pour avoir une idée plus précise sur son comportement en cas de panne, nous avons présenté certains défauts typiques connus dans la littérature des machines tournantes.

Pour choisir une carte d'acquisition, nous avons d'abord employé un DAT pour enregistrer les signaux issus des différents capteurs et effectué une analyse fréquentielle qui nous a permis de préciser les fréquences d'échantillonnage de chacun des signaux vibratoires.

Pour constituer une base de données représentant au mieux le fonctionnement du processus à surveiller, nous avons opté pour une acquisition conditionnée à l'apparition de certains événements. La base de données ainsi obtenue devrait être suffisamment riche pour décrire de façon efficace le fonctionnement normal de la machine. Dans la suite nous avons procédé à une analyse spectrale des signaux vibratoires dans le but d'extraire les attributs caractérisant le fonctionnement normal de l'éolienne.

## Chapitre 3

# Analyse spectrale et extraction des attributs

#### 3.1 Introduction

L'éolienne, située dans un environnement hostile, est sujette à différents types de vibrations internes inhérentes au fonctionnement de ses composants tournants et externes dues aux conditions climatiques. Comme toute machine tournante, les signaux générés par les divers sous-ensembles de l'éolienne sont le fruit d'un mélange complexe de nombreux signaux sinusoïdaux. Dans ce sens, l'analyse spectrale des signaux issus des différents accéléromètres installés sur l'éolienne nous semble la voie naturelle pour étudier et surveiller l'évolution de son état de fonctionnement [Zwi95].

Cependant, l'interprétation des spectres de l'éolienne est d'autant plus précise et pertinente que l'on dispose d'un maximum d'informations concernant les fréquences théoriques des divers éléments à surveiller. En effet, toute anomalie affectant la machine (Balourd, défaut de roulement, engrènement défectueux,...) se traduit par l'apparition et le déplacement de certaines fréquences et de leurs harmoniques correspondantes [BP95].

Dans la suite, pour caractériser le fonctionnement de la machine, nous procédons
à une analyse spectrale des signaux générés par les différents accéléromètres installés sur les composants sensibles. Nous présentons, tout d'abord, cette analyse sur un signal simulé puis sur les signaux réels provenant des accéléromètres. Ensuite, nous déterminons les fréquences théoriques propres à certains composants de la machine. Enfin, nous effectuons une réduction des spectrogrammes dans le but d'établir des gabarits de fonctionnements normaux.

## 3.2 Transformée de Fourier

La transformée de Fourier est une technique d'une utilisation courante dans le diagnostic des machines tournantes. Rappelons brièvement la définition de cette transformée.

Soit un signal x(t), fonction continue du temps et supposé de largeur de bande finie. Son spectre, décrit par la transformation de Fourier continue [Max97], est :

$$X(f) = \int_{-\infty}^{+\infty} x(t) e^{-j2\pi f t} dt .$$
 (3.1)

X(f) est généralement de nature complexe et par conséquent de la forme :

$$X(f) = \Re(f) + j\Im(f) . \tag{3.2}$$

où  $\Re(f)$  est la partie réelle et  $\Im(f)$  est la partie imaginaire du spectre complexe X(f).

Aussi, on caractérise généralement X(f) par son module :

$$|X(f)|^{2} = \Re^{2}(f) + \Im^{2}(f) , \qquad (3.3)$$

et sa phase :

$$\phi(f) = \arctan(\frac{\Im(f)}{\Re(f)}) . \tag{3.4}$$

Parmi applications de la transformée de Fourier, on trouve le calcul de la densité spectrale d'énergie (DSP). Pour des signaux réels à énergie finie, la DSP notée par S(f) est définie par :

$$S(f) = |X(f)|^2$$
 (3.5)

On peut également définir l'énergie moyenne dans une bande de fréquence B par :

$$E_B = \frac{1}{B} \int_{f-B/2}^{f+B/2} S(f) df . \qquad (3.6)$$

La transformée de Fourier X(f), obtenue par l'équation (3.1), nécessite un signal à durée infinie. Or, le calcul des spectres caractérisant le fonctionnement d'un processus industriel s'opère sur des signaux de durée finie. Pour cela on fait appel à la transformée de Fourier discrète, X(m), donnée par l'équation :

$$X(m) = TFD([x(n)]) = T \sum_{n=0}^{N-1} x(n) e^{-2j\pi n m/N} , \qquad m = 0, 1, \dots, N. \quad (3.7)$$

où x(n),  $0 \le n \le N$  est le signal temporel échantillonné. T est un facteur de normalisation arbitraire. En général, on choisit  $T = T_e$ ,  $T_e$  est la période d'échantillonnage.

En fait, l'observation du signal sur une durée finie revient à multiplier ce signal par une fenêtre rectangulaire. Cette fenêtre présente l'inconvénient de déformer le spectre du signal et on lui préfère la fenêtre de Hanning. Le tableau de la figure 3.1 donne les équations temporelles et fréquentielles de ces deux fenêtres. Les figure 3.2 et 3.2 montrent l'allure temporelle et fréquentielle des deux fenêtres : rectangulaire et de Hanning.

Fenêtre	Temporelle $F_{T_0}(t)$ et Fréquentielle $F_{T_0}(f)$		
Rectangulaire	$\Box_{T_{0}}(t) = \begin{cases} 1  pour   t  \le T_{0} \end{cases}$		
	0  ailleurs		
	$\Pi_{T_0}(f) = 2T_0\left(\frac{\sin(2\pi fT_0)}{2\pi fT_0}\right)$		
Hanning	$b_{-}(t) = \int 0.54 + 0.46 * \cos(\frac{2\pi t}{T_0})  pour \  t  \le T_0$		
	$\begin{pmatrix} n_{I_0}(t) = \\ 0 & ailleurs \end{pmatrix}$		
	$H_{T_0}(f) = 0.54\Pi_{T_0}(f) + 0.23\Pi_{T_0}(f + \frac{1}{2T_0}) + 0.23\Pi_{T_0}(f - \frac{1}{2T_0})$		

FIG. 3.1 : Fenêtres temporelles et fréquentielles rectangulaires et de Hanning.

La transformée de Fourier de la fenêtre de Hanning, contrairement à celle de la fenêtre rectangulaire, présente très peu d'oscillations. Son premier lobe est plus large

et son amplitude plus réduite à l'origine. Par conséquent, la fenêtre de Hanning est mieux adaptée à l'analyse spectrale des signaux.



FIG. 3.2 : Fenêtres temporelles rectangulaire et de Hanning.



FIG. 3.3 : Réponses fréquentielles des fenêtres : rectangulaire et de Hanning.

En pratique, la TFD est devenue une méthode d'analyse spectrale très utilisée après la mise au point de l'algorithme de transformation de Fourier Rapide (TFR), plus connu sous son abréviation anglaise de FFT (Fast Fourier Transform). La version la plus répandue de cet algorithme calcule une TFD sur un nombre de points qui est une puissance de 2 :  $N = 2^{\gamma}$ ,  $\gamma$  est un entier.

## 3.3 Spectrogramme d'un signal simulé

Soit x(t) un signal composé de 3 sinusoïdes, de fréquences et d'amplitudes différentes, et d'un bruit aléatoire gaussien de moyenne nulle et de variance unité :

$$x(t) = x_d(t) + x_b(t) , (3.8)$$

avec

а.

$$\begin{cases} x_d(t) = A\sin(2\pi f_1 t) + B\sin(2\pi f_2 t) + C\sin(2\pi f_3 t) ,\\ x_b(t) \sim \mathcal{N}(0, 1) . \end{cases}$$
(3.9)

où A = 2, B = 1.6 et C = 2.3 sont les amplitudes des 3 sinusoïdes. Les fréquences fondamentales sont données par :  $f_1 = 550$ Hz,  $f_2 = 660$ Hz et  $f_3 = 850$ Hz. La fréquence d'échantillonnage respectant la condition de Shannon a été fixée à 3KHz.

En fait, x(t) simule un signal de vibrations issu d'un accéléromètre installé sur un système composé de deux arbres de transmission en rotation les uns par rapport aux autres. Les fréquences  $f_1$ ,  $f_2$  et  $f_3$  représentent respectivement les fréquences de rotation de arbre lent, arbre lent/arbre rapide et arbre rapide.

La figure 3.3 montre l'allure du signal x(t) qui, à première vue, paraît difficile à exploiter. La figure 3.3 illustre le spectre d'amplitude de ce signal calculé en utilisant une fenêtre de Hanning de taille N = 1024 échantillons. Il est aisé de remarquer la présence des 3 pics correspondant aux 3 fréquences  $f_1$ ,  $f_2$  et  $f_3$ .



FIG. 3.4 : Allure du signal temporel simulé.



FIG. 3.5 : Spectre d'amplitude du signal simulé calculé sur une fenêtre de 1024 échantillons, la fréquence d'échantillonnage  $f_e = 3$ KHz.

Le spectre tel qu'il est calculé est trop riche en information pour être utilisé directement. En effet, seuls les 3 pics correspondant aux 3 fréquences données ci-dessus qui nous intéressent.

Le suivi de ces fréquences nous renseigne sur l'état du processus. Une manière efficace de suivre cette évolution est de calculer le spectre du signal sur plusieurs fenêtres successives non chevauchées. La visualisation de ces spectres en fonction du temps dite aussi spectrogramme, permet d'avoir une vision dynamique sur l'évolution des certaines fréquences contenues dans le spectre. Il s'agit donc d'une représentation de l'amplitude du spectre en fonction de la fréquence et du temps.

La figure 3.4 montre le spectrogramme du signal simulé. On observe la présence de 3 pics de fréquences plus au moins stables au cours du temps.



FIG. 3.6 : Spectrogramme d'amplitude du signal simulé calculé sur 15 fenêtres successives. La taille de chaque fenêtre est N = 1024 échantillons.

## 3.4 Spectrogramme de démarrage de l'éolienne

Nous avons enregistré le signal de vibration de la génératrice lors d'un démarrage de l'éolienne jusqu'au couplage avec le réseau EDF. Lors du démarrage, la vitesse de l'arbre rapide croît de zéro jusqu'à la vitesse de 51 tr/mn. Comme cet arbre est relié au multiplicateur par l'intermédiaire d'un ensemble d'engrenages en rotation les uns par rapport aux autres, qui est lui-même connecté à la génératrice. Toutes les fréquences liées à des phénomènes purement mécaniques doivent, par conséquent, augmenter en respectant des rapports de proportionnalités entre elles. Ce phénomène dû au démarrage est mis en évidence sur le spectrogramme en fausse couleur de la figure 3.4. Dans ce type de représentation, la couleur correspond à l'amplitude dans les spectres. Le bleu correspond aux faibles amplitudes alors que le vert correspond aux amplitudes maximales.

On note qu'effectivement, entre l'instant t=0 mn et t=5 mn, la majorité des fréquences se déplacent avec un rapport constant entre elles. Les fréquences se stabilisent une fois la connexion avec le réseau est établit.



FIG. 3.7 : Spectrogramme du signal de vibration de la génératrice lors du démarrage de l'éolienne obtenu en utilisant une fenêtre de Hanning de  $2^{15} = 32768$  échantillons, la fréquence d'échantillonnage  $f_e$ =31KHz.

# 3.5 Spectrogrammes du fonctionnement normal de l'éolienne

Dans ce paragraphe, nous proposons :

- de vérifier le choix de la fréquence d'échantillonnage utilisée par la carte d'acquisition DT3001,
- d'étudier l'évolution des fréquences contenues dans les spectrogrammes des signaux issus des accéléromètres installés sur le multiplicateur et la génératrice.

Pour cela, nous visualisons les spectrogrammes correspondants aux signaux enregistrés, tout d'abord par le DAT puis en utilisant la carte d'acquisition DT3001. Les figures 3.5 à 3.5 illustrent les spectrogrammes des 3 signaux de vibrations issus des trois accéléromètres en utilisant la carte d'acquisition DT3001 et un enregistreur DAT.

A partir de l'analyse de ces spectrogrammes, on peut remarquer que, pour chaque capteur, la majorité des fréquences existent et d'une manière régulière dans les deux cas. Ainsi, on peut conclure que le choix de la fréquence d'échantillonnage dans le cas des signaux enregistrés par la carte DT3001 respecte la condition de Shannon.



FIG. 3.8 : Spectrogramme du signal de vibration issu de l'accéléromètre axial sur le multiplicateur enregistré par le DAT obtenu en utilisant une fenêtre de Hanning de  $2^{15} = 32768$ échantillons, la fréquence d'échantillonnage  $f_e$ =48KHz.



FIG. 3.9 : Spectrogramme du signal de vibration issu de l'accéléromètre axial sur le multiplicateur enregistré par la carte DT3001 obtenu en utilisant une fenêtre de Hanning de  $2^{15} = 32768$ échantillons, la fréquence d'échantillonnage  $f_e=31$ KHz.



FIG. 3.10 : Spectrogramme du signal de vibration issu de l'accéléromètre radial sur le multiplicateur enregistré par le DAT obtenu en utilisant une fenêtre de Hanning de  $2^{15} = 32768$ échantillons, la fréquence d'échantillonnage  $f_e$ =48KHz.



FIG. 3.11 : Spectrogramme du signal de vibration issu de l'accéléromètre radial sur le multiplicateur enregistré par la carte DT3001 obtenu en utilisant une fenêtre de Hanning de  $2^{15} = 32768$ échantillons, la fréquence d'échantillonnage  $f_e=31$ KHz.

• Pour la génératrice :



FIG. 3.12 : Spectrogramme du signal de vibration de la génératrice enregistré par le DAT obtenu en utilisant une fenêtre de Hanning de  $2^{15} = 32768$  échantillons, la fréquence d'échantillonnage  $f_e$ =48KHz.



FIG. 3.13 : Spectrogramme du signal de vibration de la génératrice enregistré par la carte DT3001 obtenu en utilisant une fenêtre de Hanning de  $2^{15} = 32768$  échantillons, la fréquence d'échantillonnage  $f_e=31$ KHz.

## 3.6Fréquences théoriques caractérisant le fonctionnement normal de l'éolienne

Pour interpréter les spectrogrammes, il est nécessaire de déterminer les fréquences théoriques correspondant aux différents composants en rotation constituant l'éolienne.

Les fréquences théoriques de vibrations sont calculées à partir des formules suivantes :

• Vibration due à des mouvements de rotation des arbres :

$$f_{rot}(\text{Hz}) = \text{Vitesse de rotation en } (\text{Tr/mn})/60,$$
 (3.10)

• Vibration due à des mouvements d'engrenages :

 $f_{eng}(\text{Hz}) = \text{Vitesse de rotation en (Tr/mn)*Nombre de dents/60},$ (3.11)

• Vibration due aux mouvements des roulements à billes :

Il existe quatre fréquences caractéristiques :

1. La fréquence de passage des billes sur la bague externe du roulement notée  $f_{be}$ ,

$$f_{be} = \frac{f_{rot} * N_b}{2} \left( 1 - \frac{d}{D} \cos \alpha \right), \qquad (3.12)$$

où  $\begin{cases} f_{rot}, \text{ la fréquence de rotation de la bague interne} \\ (\text{la bague externe étant supposée fixe}), \\ N_b, \text{ le nombre de billes,} \\ D, \text{ le diamètre du roulement,} \\ d, \text{ le diamètre des billes,} \\ c_c \text{ l'angle de contact} \end{cases}$ 

- $\alpha$ , l'angle de contact.
- 2. La fréquence de passage des billes sur la bague interne du roulement notée  $f_{bi},$

$$f_{bi} = \frac{f_{rot} * N_b}{2} \left( 1 + \frac{d}{D} \cos \alpha \right), \qquad (3.13)$$

3. La fréquence de rotation des billes sur eux-mêmes notée  $f_{br}$ ,

$$f_{br} = \frac{f_{rot} * D}{2d} \left( 1 - \left(\frac{d}{D}\cos\alpha\right)^2 \right), \qquad (3.14)$$

4. La fréquence de passage d'un défaut de cage notée  $f_c$ ,

$$f_c = \frac{f_{rot}}{2} \left( 1 - \frac{d}{D} \cos \alpha \right), \qquad (3.15)$$

Nous allons déterminer par la suite, les fréquences théoriques des éléments à surveiller.

#### 3.6.1 Rotor-Pales

Les deux fréquences induites par la rotation du rotor sont la fréquence de sa propre rotation et cette même fréquence multipliée par le nombre de pales (cf. le tableau de la figure 3.14).

Origine de la vibration	Vitesse de rotation (Tr/mn)	Fréquence (Hz)
Rotation du rotor	51	0.85
Rotation des pales	153	2.55

FIG. 3.14 : Fréquences dues à la rotation de la structure pales-rotor.

#### 3.6.2 Multiplicateur

Le multiplicateur possède 4 arbres en rotation. L'arbre lent est relié au rotor de l'éolienne. Les arbres intermédiaires 1 et 2 sont internes au multiplicateur. L'arbre rapide est relié à la génératrice. Les engrenages du multiplicateur sont des engrenages droits. Les roulements sont des roulements à rouleaux droits ou coniques. Les 3 tableaux des 3 figures 3.15, 3.16 et 3.17 résument les fréquences de rotation des engrenages et leurs harmoniques d'ordre supérieur ainsi que les fréquences de rotation des roulements à billes du multiplicateur de vitesse.

Origine de la vibration	Nombre	Vitesse de rotation	Fréquence en (Hz)
	de dents	(Tr/mn)	
Arbre lent	-	51	0.85
Arbre lent/Arbre inter 1	47	_	40.3
Arbre inter 1	14	172	2.87
Arbre inter 1/Arbre inter 2	49	-	141
Arbre inter 2	20	420	7
Arbre inter 2/Arbre rapide	78	-	550
Arbre rapide	22	1500	25

FIG. 3.15: Fréquences dues à la rotation des arbres et des engrenages du multiplicateur.

Fréquence fondamentale	Harmonique	Harmonique	Harmonique	Harmonique
en (Hz)	d'ordre 1	d'ordre 2	d'ordre 3	d'ordre 4
	en (Hz)	en (Hz)	en (Hz)	en (Hz)
0.85	1.7	2.5	3.4	4.25
40.3	80.6	121	161	201
2.9	5.8	8.7	11.6	14.5
141	282	423	564	705
7	14	21	28	35
550	1100	1650	2200	2750
25	50	75	100	125

**FIG. 3.16 :** Harmoniques d'ordre 1, 2, 3 et 4 dues à la rotation des arbres de transmission et des engrenages du multiplicateur.

Arbre lent	$N_b = 20$	d = 60mm	D = 420mm	$\alpha = 0^{\circ}$
	$f_{be} = 7.4$	$f_{bi} = 9.7$	$f_{br} = 2.9$	$f_{c} = 0.4$
Arbre inter 1	$N_b = 20$	d = 40mm	D = 280mm	$\alpha = 0^{\circ}$
	$f_{be} = 25.6$	$f_{bi} = 34.3$	$f_{br} = 10.3$	$f_{c} = 1.3$
Arbre inter 2	$N_b = 20$	d = 30mm	D = 210mm	$\alpha = 0^{\circ}$
	$f_{be} = 86.6$	$f_{bi} = 115.4$	$f_{br} = 34.6$	$f_{c} = 4.3$
Arbre rapide	$N_b = 20$	d = 20mm	D = 140mm	$\alpha = 0^{\circ}$
	$f_{be} = 255.4$	$f_{bi} = 340.6$	$f_{br} = 102.1$	$f_{c} = 12.8$

FIG. 3.17 : Fréquences dues aux roulements du multiplicateur.

## 3.6.3 Génératrice

Les fréquences vibratoires induites par la génératrice sont d'une part la vitesse de rotation et d'autre part les fréquences induites par les roulements mis en rotation à l'intérieur de la génératrice (cf. tableau de la figure 3.18)

Origine de la vibration	Vitesse de rotation (Tr/mn)	Fréquence (Hz)
Rotation de la génératrice	1500	25

FIG. 3.18 : Fréquence dues à la rotation de la génératrice.

La confrontation entre les fréquences théoriques obtenues par les équations (3.10) à (3.15) et les fréquences retrouvées au sein des différentes spectrogrammes est effectuée plus tard dans le chapitre 5.

## 3.7 Extraction d'attributs par réduction des spectrogrammes

L'extraction des attributs consiste à rechercher les paramètres caractéristiques correspondant aux fonctionnements normal et anormal de l'éolienne. Comme il est très difficile, voire impossible, d'obtenir des situations représentant un dysfonctionnement, nous allons nous focaliser uniquement sur l'étude des spectrogrammes du fonctionnement normal.

Le choix de la méthode d'extraction d'attributs est une étape cruciale de l'étude. En effet, l'efficacité d'un système de suivi de fonctionnement dépend de la méthode utilisée. Nous proposons une technique de réduction des spectrogrammes basées sur le calcul de l'énergie par bandes de fréquences. En fait, toute dérive de fonctionnement de la machine se manifeste dans le spectrogramme par la naissance d'une ou plusieurs composante affectant ainsi l'énergie contenue dans une ou plusieurs bandes. Par conséquent, le suivi de fonctionnement de l'élément à surveiller revient à suivre l'énergie contenue dans chacune des bandes. La démarche consiste à :

- 1. Rechercher les principaux fréquences contenues dans les spectrogrammes représentant les fréquences de rotation de la machine,
- 2. Déterminer les intervalles de fluctuations des différentes fréquences,
- Décomposer le spectrogramme en plusieurs bandes de fréquences. Des bandes étroites autour des fréquences à surveiller et des bandes larges constituées entre les bandes étroites.
- 4. Calculer l'énergie contenue dans chacune des bandes.

## 3.8 Conclusion

Dans ce chapitre, nous avons présenté un ensemble de spectrogrammes relatifs aux deux enregistrements. L'analyse visuelle de ces derniers montrent qu'il existe plusieurs fréquences en commun pour les 3 signaux avec des amplitudes comparables. Ces fréquences fluctuent légèrement autour des valeurs théoriques représentant les fréquences de rotation des éléments de la machine.

Nous avons montré grâce aux spectrogrammes de démarrage de la génératrice que suivre le fonctionnement de l'éolienne revient à suivre ces différentes fréquences qui se déplacent avec un rapport constant entre elles.

Nous avons proposé une méthode d'extraction d'attributs basées sur le calcul de l'énergie par bande. Le choix de la taille de chacune des bandes est discuté dans le chapitre 5. Donc, suivre le fonctionnement d'un élément à surveiller revient à suivre l'énergie contenue dans chacune des bandes des spectrogrammes de vibrations.

Dans le suite, plusieurs architectures neuronales seront adaptées pour le suivi de fonctionnement de l'éolienne.

## Chapitre 4

## Réseaux de neurones autoassociateurs

## 4.1 Introduction

L'intérêt suscité par les réseaux de neurones est motivé par l'analogie biologique qu'ils comportent et qui les distingue des principes de calcul des ordinateurs classiques.

En effet, le cerveau est un système de traitement d'information très complexe, non linéaire et parallèle. Il a la capacité d'organiser et de combiner les réponses de ses neurones pour résoudre certaines tâches complexes, telles que les problèmes d'apprentissage, de reconnaissance de formes et de contrôle.

Les réseaux de neurones sont souvent appelés réseaux connexionnistes. Ce sont des processeurs distribués massivement parallèles composés d'unités de calcul appelées neurones qui sont fortement interconnectées. Un neurone effectue une sommation, pondérée par les poids des connexions, de ses entrées suivie d'une non linéarité souvent de type sigmoïde ou tangente hyperbolique. Les résultats du calcul du neurone sont transmis à l'ensemble des neurones qui lui sont connectés. Par analogie avec le cerveau humain, les réseaux acquièrent les connaissances grâce à un processus d'apprentissage sur un ensemble d'exemples. L'algorithme d'apprentissage permet de renforcer ou diminuer les différents poids des connexions du réseau pour atteindre un objectif désiré. Les connaissances acquises par le réseau sont essentiellement stockées dans les poids de ses connexions.

Le Perceptron multicouche est le plus employé des réseaux de neurones artificiels. Ce réseau associe aux entrées les sorties désirées, et de ce fait est appelé réseau associateur. L'analyse de ce réseau a été effectuée par différents auteurs qui ont montré ses liens avec la méthode statistique d'analyse discriminante non linéaire [Thi89] [LD96] [Mao94] [WL90].

Lorsque les sorties désirées sont les entrées mêmes, le réseau est dit autoassociateur. Pour ce type de réseau, une attention particulière est portée aux sorties des neurones de la couche cachée qui sont en général au nombre inférieur à celui des entrées et le réseau réalise, alors, une réduction de la dimension. En particulier, lorsque le nombre des sorties est égal à deux, le réseau effectue une projection plane non linéaire. Comme tout système non linéaire, il est difficile de comprendre le mécanisme d'apprentissage des réseaux autoassociateurs. Par contre en linéarisant ses équations de fonctionnement, on peut donner une signification aux différents poids ou paramètres du réseau.

Nous proposons d'exploiter à la fois ses propriétés de projection pour visualiser sur un écran graphique la position courante du fonctionnement de l'éolienne mais aussi ses propriétés d'autoassociation pour délimiter le domaine du fonctionnement normal de l'éolienne et vérifier son état courant de fonctionnement.

Un réseau associateur à trois couches est un approximateur universel. Autrement dit, on peut approximer toute fonction en choisissant convenablement le nombre des neurones cachées. Or le réseau autoassociateur est constitué d'une juxtaposition de deux réseaux à trois couches. Pour un nombre de neurones sur la couche cachée centrale donné, on peut espérer reconstruire toute projection sur cette couche quitte à voir la  $2^{\grave{e}me}$  et la  $3^{\grave{e}me}$  couche cachées de tailles suffisantes [Kra91].

Plusieurs applications ont été proposées, on peut citer :

- Au niveau de l'exploitation de la couche cachée :
  - Compression d'images [CMZ87],
  - Projection pour la visualisation, Projection plane pour la classification interactive [Mao94] [Dao93] [Bet98],
  - Réduction de la dimension pour rechercher des relations entre les variables d'entrées du réseau [Kra91].
- Au niveau de l'exploitation de la couche de sortie :
  - Diagnostic des moteurs [PMD+95], diagnostic de défaillance des capteurs [BFC98].
  - Vérification de la parole [LGS94] [BFG95], vérification des billets de banque [FGP96] [TO95].

Dans ce chapitre, nous nous intéressons au réseau autoassociateur pour exploiter à la fois ses propriétés de projection afin de visualiser sur un écran graphique la position courante du fonctionnement de l'éolienne et ses propriétés d'autoassociation pour délimiter le domaine du fonctionnement normal de l'éolienne et vérifier son état courant de fonctionnement.

Nous présentons tout d'abord deux types d'architectures, un réseau autoassociateur à 3 couches puis à 5 couches. Nous étudierons comment déterminer les poids du réseau, premièrement, par une décomposition en valeurs et vecteurs propres de la matrice des observations (ACP) [BK88], et deuxièmement, par l'algorithme de rétropropagation du gradient stochastique. Ensuite, quelques propriétés du réseau portant sur l'utilisation ou non des fonctions de transfert non linéaires seront présentées. Nous nous intéresserons également à l'optimisation du nombres de neurones dans les couches cachées par critère informationnel. Enfin, ces architectures seront appliquées pour le suivi du fonctionnement normal d'une éolienne dans le chapitre suivant.

Le chapitre est organisé de la façon suivante :

Dans la section 2, nous présentons l'architecture, les équations de fonctionnement d'un réseau autoassociateur linéaire à trois couches dont les matrices du poids et du biais sont déterminées par une analyse en composantes principales. Les sections 3 et 4 décrivent respectivement deux autoassociateur non linéaires à 3 et 5 couches. L'apprentissage est effectué en utilisant l'algorithme de rétropropagation du gradient de l'erreur. Ensuite, dans la section 5, l'optimisation du nombre de neurones dans la couche centrale de l'autoassociateur à 5 couches est assuré par les critères informationnels d'Akaike [Aka72].

Au fur et à mesure du développement du chapitre, nous appliquons les différentes architectures sur un exemple artificiel simulant le fonctionnement d'un système de deux arbres de transmission en rotation les uns par rapport aux autres.

## 4.2 Réseau autoassociateur linéaire à trois couches

Dans ce paragraphe nous présentons l'architecture, les équations de fonctionnement et le critère d'erreur du réseau. Pour mettre en évidence le lien avec la méthode d'analyse en composantes principales, nous reprenons l'analyse effectuée par Bourlard et Kamp [Bou].

#### 4.2.1 Architecture

On considère un réseau de neurones autoassociateur dont l'architecture est constituée de 3 couches (cf. figure 4.1) :

- la couche d'entrée est composée de (N + 1) éléments. A chaque entrée correspond une variable dont la valeur est une composante d'un vecteur dit vecteur observation. Les composantes du vecteur observation sont transmises aux neurones de la couche cachée à travers des connexions pondérées. Notons que la première entrée reçoit toujours la valeur 1 et les poids de ses connexions sont appelés des biais.
- la couche cachée comporte M neurones, chacun effectue un traitement linéaire consistant en une sommation pondérée des composantes de l'observation présentée à l'entrée du réseau. Le résultat du traitement est transmis aux neurones de la couche de sortie.
- la couche de sortie comprend N neurones. Chacun d'eux effectue une sommation pondérée des informations reçues à ses entrées.

Soulignons que les éléments de la couche d'entrée du réseau n'effectuent aucun traitement. De ce fait, ils ne sont pas considérés comme étant des neurones et sont représentés par des petits cercles sur la figure 4.1.

#### 4.2.2 Equations de fonctionnement

Soit  $x_n$ , la  $n^{ime}$  composante du vecteur observation appliquée à la  $n^{ime}$  entrée du réseau,  $n = 1, \ldots, N$ ,



**FIG. 4.1** : Réseau autoassociateur linéaire à trois couches.

- $e_m$ , l'entrée totale du  $m^{eme}$  neurone caché,  $m = 1, \ldots, M$ ,
- $h_m$ , la sortie du  $m^{eme}$  neurone caché,  $m = 1, \ldots, M$ ,
- $s_n$ , l'entrée totale du  $n^{ime}$  neurone de sortie,  $n = 1, \ldots, N$ ,
- $y_n$ , la sortie du  $n^{eme}$  neurone de sortie,  $n = 1, \ldots, N$ ,
- $a_{mn}$ , le poids de la connexion entre le  $m^{\grave{e}me}$  neurone caché et la  $n^{\grave{e}me}$  entrée,
- b<sub>nm</sub>, le poids de la connexion entre le n<sup>ème</sup> neurone de sortie et le m<sup>ème</sup> neurone caché,
- $\vartheta_m$ , le biais de la couche d'entrée associé au  $m^{\grave{e}me}$  neurone caché,  $m = 1, \ldots, M$ ,
- $\theta_n$ , le biais de la couche cachée associé au  $n^{eme}$  neurone de sortie,  $n = 1, \ldots, N$ ,
- $f(\cdot)$ , est la fonction d'activation associée aux neurones de la couche cachée et aux neurones de la couche de sortie. Ici  $f(\cdot)$  est la fonction identité.

La figure 4.2 illustre les notations utilisées dans les différentes équations de fonctionnement du réseau.

Equations de la couche cachée :

$$e_m = \sum_{n=1}^N a_{mn} x_n + \vartheta_m , \qquad m = 1, \ldots, M , \qquad (4.1)$$



FIG. 4.2 : Notations des poids et des variables du réseau autoassociateur linéaire.

$$h_m = f(e_m) = e_m , \qquad m = 1, \dots, M .$$
 (4.2)

Equations de la couche de sortie :

. .

$$s_n = \sum_{m=1}^{M} b_{nm} h_m + \theta_n , \qquad n = 1, \dots, N ,$$
 (4.3)

$$y_n = f(s_n) = s_n , \qquad n = 1, \dots, N .$$
 (4.4)

Pour un vecteur observation  $\boldsymbol{x}$  présenté à l'entrée du réseau, les équations de la couche cachée et de la couche de sortie s'écrivent sous la forme vectorielle :

$$\boldsymbol{e} = A\boldsymbol{x} + \boldsymbol{\vartheta} , \qquad (4.5)$$

$$\boldsymbol{h} = \boldsymbol{e} \;, \tag{4.6}$$

où  $\boldsymbol{h} = (h_1, h_2, \ldots, h_M)^T$  est le vecteur de sortie des neurones cachés.

$$\boldsymbol{s} = \boldsymbol{B}\boldsymbol{h} + \boldsymbol{\theta} \;, \tag{4.7}$$

$$\boldsymbol{y} = \boldsymbol{s} \ . \tag{4.8}$$

Le réseau reçoit en entrée les composantes des vecteurs observations, les traite et les résultats en sortie sont comparés aux valeurs présentées en entrée. Les poids et les biais du réseau, constituant ses paramètres libres, sont modifiés par un mécanisme d'apprentissage de telle sorte que le réseau établit une fonction identité entre les entrées et les sorties.

## 4.2.3 Apprentissage par la méthode d'analyse en composantes principales

Considérons un ensemble de P vecteurs  $\{x_1, x_2, \ldots, x_P\}$  constituant la base d'apprentissage du réseau. Le but de l'apprentissage est d'adapter les poids du réseau de telle sorte que celui-ci fournit en sortie un vecteur y, aussi proche que possible du vecteur observation x, présenté à son entrée. Le réseau réalise ainsi la fonction identité.

#### 4.2.4 Critère d'erreur

Soit J l'erreur quadratique entre le vecteur observation  $\boldsymbol{x}$  et le vecteur de sortie  $\boldsymbol{y}$  correspondant :

$$J = \frac{1}{2} ||\boldsymbol{x} - \boldsymbol{y}||^2.$$
 (4.9)

Les poids optimum du réseau sont donc ceux qui minimisent ce critère.

Le problème de l'apprentissage devient un problème de minimisation d'une erreur quadratique dont la solution peut être obtenue, soit par la méthode d'analyse en composantes principales que nous décrivons par la suite, soit par l'algorithme de rétropropagation du gradient de l'erreur.

Soient  $\mathbf{X} = [\mathbf{x}_1 | \mathbf{x}_2 | \dots | \mathbf{x}_P]$  la matrice des observations de dimension N \* P, formée par les P vecteurs  $\mathbf{x}$  de la base d'apprentissage,  $\mathbf{H} = [\mathbf{h}_1 | \mathbf{h}_2 | \dots | \mathbf{h}_P]$  la matrice des réponses des P neurones de la couche cachée correspondant aux vecteurs observations, de dimension M \* P et  $\mathbf{Y} = [\mathbf{y}_1 | \mathbf{y}_2 | \dots | \mathbf{y}_P]$  la matrice des P réponses des neurones de la couche de sortie du réseau, de dimension N \* P.

Les équations de fonctionnement du réseau (4.5) à (4.8) deviennent :

$$\boldsymbol{E} = A\boldsymbol{X} + \boldsymbol{\vartheta}\boldsymbol{\nu}^{T}, \tag{4.10}$$

$$\boldsymbol{H} = \boldsymbol{E} , \qquad (4.11)$$

$$\boldsymbol{S} = \boldsymbol{B}\boldsymbol{H} + \boldsymbol{\theta}\boldsymbol{\nu}^{T},\tag{4.12}$$

$$\boldsymbol{Y} = \boldsymbol{S} , \qquad (4.13)$$

où  $\nu$  est un vecteur de dimension P dont les éléments sont tous égaux à 1. Le critère d'erreur totale  $J_T$  calculé sur toute la base d'apprentissage peut être évalué par l'équation suivante :

$$J_T(A, B, \boldsymbol{\vartheta}, \boldsymbol{\theta}) = \frac{1}{2} \|\boldsymbol{X} - \boldsymbol{Y}\|^2.$$
(4.14)

#### 4.2.4.1 Calcul des biais de la couche cachée

Utilisant l'équation (4.14), le critère d'erreur des moindres carrés  $J_T$  peut être réécrit de la façon suivante :

$$J_T = \frac{1}{2} \| \boldsymbol{X} - \boldsymbol{B}\boldsymbol{H} - \boldsymbol{\theta}\boldsymbol{\nu}^T \|^2 = \frac{1}{2} \left( \boldsymbol{X} - \boldsymbol{B}\boldsymbol{H} - \boldsymbol{\theta}\boldsymbol{\nu}^T \right) \left( \boldsymbol{X} - \boldsymbol{B}\boldsymbol{H} - \boldsymbol{\theta}\boldsymbol{\nu}^T \right)^T. \quad (4.15)$$

Le calcul de la sensibilité du critère  $J_T$  par rapport à  $\boldsymbol{\theta}$  donne :

$$\frac{\partial J_T}{\partial \boldsymbol{\theta}} = (\boldsymbol{X} - B\boldsymbol{H}) (-\nu) , \qquad (4.16)$$

Comme  $\nu^T \nu = P$ , on obtient :

$$\frac{\partial J_T}{\partial \boldsymbol{\theta}} = P \boldsymbol{\theta} - (\boldsymbol{X} - B \boldsymbol{H}) \, \nu = 0 \; . \tag{4.17}$$

Donc, la valeur  $\boldsymbol{\theta}$  minimisant ce critère est :

$$\hat{\boldsymbol{\theta}} = \frac{1}{P} \left( \boldsymbol{X} - B\boldsymbol{H} \right) \boldsymbol{\nu} . \tag{4.18}$$

Rôle du biais : En substituant l'équation (4.18) dans (4.15), on obtient :

$$J_T = \frac{1}{2} \| \boldsymbol{X} - \frac{\boldsymbol{X} \nu \nu^T}{P} - B\boldsymbol{H} + \frac{B\boldsymbol{H} \nu \nu^T}{P} \|^2, \qquad (4.19)$$

Posons  $\mathbf{X'} = \mathbf{X}(I_P - \nu\nu^T/P)$  et  $\mathbf{H'} = \mathbf{H}(I_P - \nu\nu^T/P)$ , on obtient ainsi l'équation suivante :

$$J(A, B, \boldsymbol{\vartheta}, \boldsymbol{\theta}) = \frac{1}{2} \|\boldsymbol{X'} - B\boldsymbol{H'}\|^2, \qquad (4.20)$$

où  $\nu\nu^T$  est une matrice dont tous les éléments sont égaux à 1 alors que  $\nu^T\nu$  est un

scalaire :  $\nu^T \nu = P$ .

Par conséquent le rôle du vecteur biais de la couche cachée  $\theta$  est de ramener le problème d'optimisation quadratique sur l'ensemble des observations non centrées à un problème d'apprentissage avec des observations centrées, donc de moyennes nulles.

#### 4.2.4.2 Calcul des poids des connexions couche cachée - couche de sortie

La couche cachée étant une couche de réduction de la dimension, la matrice des poids couche cachée-couche de sortie B est d'ordre M < N.

L'expression (4.20) montre que le produit BH' minimisant le critère  $J_T$  est la meilleure approximation d'ordre M de X' au sens de la norme Euclidienne [BK88].

L'estimation des poids  $b_{nm}$  de la matrice B est simplement obtenue par une décomposition de la matrice de covariance des observations en valeurs - vecteurs propres [BK88], soit :

$$\boldsymbol{X'} = U_N \Sigma_N V_N^T, \tag{4.21}$$

où  $U_N(\text{resp. } V_N)$  une matrice de passage de dimension N \* N (resp. P \* N) formée des vecteurs propres normalisés de  $\mathbf{X'X'}^T$  (resp.  $\mathbf{X'}^T\mathbf{X'}$ ) associés aux valeurs propres ordonnées par valeurs décroissantes :  $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$ .

 $\Sigma_N = diag[\sigma_1, \sigma_2, \cdots, \sigma_N]$  est une matrice diagonale d'éléments  $\sigma_n = \sqrt{\lambda_n}$ .

Pour simplifier, nous supposons que X est une matrice de rang plein (i.e. que toutes les valeurs propres sont non nulles).

Selon [GL83, Ste73], la meilleure approximation d'ordre M de X' est obtenue en gardant les M plus grandes valeurs propres, soit :

$$\hat{B}\hat{H}' = U_M \Sigma_M V_M^T, \qquad (4.22)$$

avec  $\Sigma_M = diag[\sigma_1, \sigma_2, \cdots, \sigma_M]$  où les deux matrices de passage  $U_M$  et  $V_M$  sont constituées des M premiers vecteurs propres de  $U_N$  et  $V_N$ .

Par conséquent, les matrices  $\hat{H}'$  et  $\hat{B}$  sont déduites de l'équation (4.22) par :

$$\hat{\boldsymbol{H}}' = Q \Sigma_M V_M^T, \tag{4.23}$$

$$\hat{B} = U_M Q^{-1}. \tag{4.24}$$

où Q est une matrice M \* M orthogonale non singulière arbitraire vérifiant  $Q^T Q = I_M$ . En particulier, si  $Q = I_M$ , les poids de la couche cachée sont les M premiers vecteurs propres de la matrice des observations X'.

**Remarque :** Les expressions optimales des poids et biais de la couche cachée couche de sortie dans les équations (4.18) et (4.24) ont été déduites par décomposition en valeurs - vecteurs propres de la matrice des observations X', indépendamment des réponses des neurones cachés du réseau et par conséquent, indépendamment des fonctions d'activation utilisées dans les neurones de la couche cachée. Cependant, dans le cas où les neurones de la couche de sortie sont non linéaires, les résultats précédents ne sont plus valables.

#### 4.2.4.3 Calcul des poids des connexions couche entrée-couche cachée

En multipliant les deux membres de l'équation (4.11) par le terme  $(I - \nu \nu^T / P)$  qui permet de centrer les deux matrices X et H, nous devons ainsi résoudre l'équation des deux variables A et  $\vartheta$  suivante :

$$\boldsymbol{H}\left(I_{P}-\nu\nu^{T}/P\right)=A\boldsymbol{X}\left(I_{P}-\nu\nu^{T}/P\right)+\boldsymbol{\vartheta}\nu^{T}\left(I_{P}-\nu\nu^{T}/P\right) . \tag{4.25}$$

où encore :

$$\boldsymbol{H'} = A\boldsymbol{X'} + \boldsymbol{\vartheta}\nu^T \left( I_P - \nu\nu^T / P \right) \ . \tag{4.26}$$

Comme  $\nu^T \nu = P$ , le second terme du membre de droite de l'équation disparaît. On peut donc obtenir AX' indépendamment du choix de  $\vartheta$ :

$$A\mathbf{X}' = \mathbf{H}' \ . \tag{4.27}$$

Si on prend en compte l'expression de H' donnée dans l'équation (4.23) on obtient :

$$A\mathbf{X}' = Q\Sigma_M V_M^T. \tag{4.28}$$

Comme la matrice de passage  $U_M$  vérifie l'équation  $U_M^T U_M = I_M$ , et en prenant en compte l'équation (4.21), on peut voir que :

$$Q\Sigma_M V_M^T = QU_M U_M^T \Sigma_M V_M^T = QU_M \mathbf{X}' , \qquad (4.29)$$

Soit alors :

$$A\mathbf{X}' = QU_M \mathbf{X}' \ . \tag{4.30}$$

Finalement, la matrice des poids optimale couche d'entrée - couche cachée minimisant le critère d'erreur totale est donnée par :

$$\hat{A} = Q U_M^T . \tag{4.31}$$

#### **Remarque** :

A partir de ce résultat, on peut déterminer dans le cas d'une fonction d'activation linéaire, la valeur optimale du vecteur biais de la couche cachée  $\boldsymbol{\theta}$ . Pour cela, il suffit d'intégrer dans l'équation (4.18) les résultats des équations (4.11), (4.24) et (4.31). On obtient ainsi l'équation suivante :

$$\hat{\boldsymbol{\theta}} = (I_M - U_M U_M^T) \mu_{\boldsymbol{X}} - U_M Q^{-1} \hat{\boldsymbol{\vartheta}} , \qquad (4.32)$$

où  $\mu_{\mathbf{X}} = \mathbf{X}\nu/P$  est le vecteur moyenne des observations de dimension N.

Ainsi, pour un choix arbitraire du vecteur de biais d'entré  $\hat{\vartheta}$ , le vecteur de biais de la couche cachée  $\hat{\theta}$  est ajusté selon l'équation (4.32) qui assure une égalité entre le vecteur moyen des observations et le vecteur moyen des réponses du réseau pour chaque observation présentée (expliquer!).

#### Récapitulatif du calcul direct des poids du réseau :

Après une décomposition en valeurs - vecteurs propres de la matrice centrée X', les équations suivantes donnent les solutions optimales pour les différentes matrices A, B,  $\vartheta$  et  $\theta$  du réseau autoassociateur linéaire.

- On calcule  $\mathbf{X'} = \mathbf{X}(I_P \nu \nu^T / P),$
- On déduit  $U_M$  et  $V_M$  par une décomposition en valeurs-vecteurs propres :  $\mathbf{X'} = U_M \Sigma_M V_M^T$ ,
- Le choix de Q est arbitraire, on peut prendre  $Q = I_M$ .
- $\hat{A} = QU_M^T = U_M^T$ ,
- $\hat{B} = U_M Q^{-1} = U_M,$
- $\hat{\boldsymbol{\vartheta}}$  est arbitraire. En particulier, on peut prendre le vecteur nul.
- $\hat{\boldsymbol{\theta}} = (I_M U_M U_M^T) \mu_{\boldsymbol{X}} U_M Q^{-1} \hat{\boldsymbol{\theta}} = (I_M U_M U_M^T) \mu_{\boldsymbol{X}}.$

La méthode de calcul des poids décrite ci-dessus est une méthode bien connue en algèbre linéaire [BK88]. Elle consiste à déterminer les vecteurs poids à partir d'une procédure de calcul des valeurs - vecteurs propres de la matrice des données d'apprentissage. Toutefois, la mise en oeuvre de cette méthode hors-ligne nécessite des hypothèses de travail telles que la linéarité des fonctions d'activation et la non singularité de certaines matrices. Or, en général, on ne possède aucune connaissance sur la structure des observations ni sur les propriétés de l'ensemble d'apprentissage. Dans ces conditions, l'algorithme de rétropropagation du gradient est mieux adapté, car il peut être mis en oeuvre sans aucune hypothèse particulière sur les données ni sur la nature des fonctions d'activations.

Dans la suite du travail, on intègre les biais dans les matrices des poids.

### 4.2.5 Exemple d'application

Reprenons l'exemple du signal simulé décrit dans le chapitre précédent représentant un système de deux arbres de transmission en rotation les uns par rapport aux autres. Le signal est composé d'un mélange de 3 sinusoïdes et d'un bruit gaussien de moyenne nulle et d'écart type égale à 1. La fréquence d'échantillonnage a été fixée à 3KHz. Pour constituer la base d'apprentissage et de test du réseau, nous avons réduit le spectrogramme du signal illustré dans la figure 3.4 en 5 bandes : trois bandes étroites centrées sur les fréquences à surveiller séparées par deux bandes larges. Le choix de la largeur des bandes étroites est fonction de la résolution du spectrogramme. Dans notre cas, la résolution est égale à  $\mathcal{R} = f_e/N = 3000/1024 \simeq 3$ Hz. Une étude des fluctuations des fréquences principales  $f_i$ , i = 1, 2 et 3 au cours du temps nous a révélé que ces fréquences prennent des valeurs comprises entre  $f_i + \Delta f_i$  et  $f_i - \Delta f_i$ , avec  $\Delta f_i = \mathcal{R}$ . Donc, pour assurer la présence des pics à surveiller au sein des bandes étroites, nous avons fixé la taille d'une bande étroite à  $2 * \Delta f_i$ . La figure 4.11 montre l'allure d'un spectre réduit.



FIG. 4.3 : Allure d'un spectre réduit en 5 bande du signal simulé.

A partir du spectrogramme réduit à 5 bandes selon la démarche ci-dessus, nous générons un ensemble d'apprentissage et de test de 290 spectres chacune. Autrement dit, nous disposons d'une matrice X de données de 5x290 éléments.

Nous considérons un réseau de neurones autoassociateur linéaire d'architecture 5x2x5. L'algorithme du calcul direct des poids du réseau, basé sur une analyse en composantes principales, permet d'obtenir respectivement les valeurs propres et les vecteurs propres consignés dans le tableau de la figure 4.12. Notons que le plan principal correspondant aux deux valeurs propres maximales explique 78.3% de l'information totale des données.

Valeurs propres	Quantité d'informations expliquée
$1.9 * 10^2$	45.9%
$1.3 * 10^2$	32.4%

FIG. 4.4 : Les 2 premières valeurs propres et le pourcentage de la variance totale des observations expliquée par chaque valeur propre.

Une fois la matrice de poids et de biais déterminées, nous nous intéressons aux sorties des deux neurones de la couche cachée centrale définissant les axes du plan de projection. Nous présentons successivement les observations de la base d'apprentissage et de test à l'entrée du réseau. La figure 4.5 montre que les projections des deux bases se superposent et se regroupent à l'intérieur d'un domaine qui peut être vu comme un domaine de fonctionnement normal du système à surveiller.

Les erreurs de reconstruction par bande sont données sur la figure 4.6. On remarque que l'erreur au niveau des bandes étroites est faible. Par contre, l'erreur au niveau des bandes larges est voisin de 1, ce qui signifie que le réseau fournit en sortie la valeur moyenne de l'énergie dans ces bandes.



FIG. 4.5 : La réponse des sorties de la couche centrale d'un autoassociateur linéaire.



FIG. 4.6 : Les erreurs de reconstruction par bande obtenues au niveau des sorties de l'autoassociateur linéaire à 3 couches.

## 4.3 Réseau autoassociateur non linéaire à 3 couches

#### 4.3.1 Architecture

L'architecture du réseau autoassociateur non linéaire est similaire à celle du réseau linéaire du paragraphe précédent. Les fonctions d'activation des neurones sont par contre non linéaires.



FIG. 4.7 : Réseau autoassociateur non linéaire à trois couches.

### 4.3.2 Equations de fonctionnement

Rappelons que  $\boldsymbol{x} = (x_1, \ldots, x_N)^T$  représente une observation présentée à l'entrée du réseau et que  $\boldsymbol{y} = (y_1, \ldots, y_N)^T$  est la sortie correspondante. Soit  $\boldsymbol{h} = (h_1, \ldots, h_M)^T$  la sortie des neurones de la couche cachée. Les équations de propagation en avant du réseau sont données par les deux équations suivantes :

$$e_m = \sum_{n=1}^{N} a_{mn} x_n , \qquad (4.33)$$

$$h_m = f(e_m), \qquad m = 1, \dots, M.$$
 (4.34)

$$s_n = \sum_{m=1}^M b_{nm} h_m,$$
 (4.35)

$$y_n = f(s_n), \qquad n = 1, \dots, N.$$
 (4.36)



FIG. 4.8 : Notations des poids et des variables du réseau autoassociateur non linéaire.

## 4.3.3 Apprentissage par la méthode de rétropropagation du gradient de l'erreur

Le principe de l'apprentissage par la méthode de rétropropagation du gradient de l'erreur est de présenter les vecteurs observations de manière aléatoire et répétitive à l'entrée du réseau. On calcule les sorties du réseau et on les compare aux sorties désirées qui sont, dans notre cas, égales aux entrées du réseau. L'erreur entre les sorties désirées et les sorties calculées par le réseau est utilisée pour adapter les différents poids des connexions.

Parmi les variantes de l'algorithme de rétropropagation du gradient, deux sont les plus utilisées. Le premier algorithme consiste à minimiser le critère d'erreur totale calculé après une présentation de toute la base d'apprentissage. Le second, que nous allons détailler dans le paragraphe suivant, minimise le critère d'erreur partielle calculé après chaque présentation d'un nouveau vecteur observation de la base d'apprentissage.

## 4.3.4 Gradient du critère d'erreur partielle

Le critère d'erreur partielle J calculé après chaque présentation d'une observation  $\boldsymbol{x}$  est :

$$J = \frac{1}{2} \|\boldsymbol{x} - \boldsymbol{y}\|^2 = \frac{1}{2} \sum_{n=1}^{N} (x_n - y_n)^2, \qquad (4.37)$$

En substituant  $y_n$  donnée par l'équation (4.36) dans l'équation (4.37) et en utilisant les équations (4.33)-(4.35) on obtient :

$$J = \frac{1}{2} \left[ \sum_{n=1}^{N} \left( x_n - f\left( \sum_{m=1}^{M} b_{nm} f\left( \sum_{n=1}^{N} a_{mn} x_n \right) \right) \right)^2 \right]$$
(4.38)

## 4.3.5 Adaptation des poids des connexions couche cachéecouche de sortie

La sensibilité du critère d'erreur par rapport aux poids des connexions couche cachée-couche de sortie du réseau s'écrit :

$$\frac{\partial J}{\partial b_{nm}} = \frac{\partial J}{\partial y_n} * \frac{\partial y_n}{\partial s_n} * \frac{\partial s_n}{\partial b_{nm}} , \qquad (4.39)$$

D'après l'équation (4.37), on a :

$$\frac{\partial J}{\partial y_n} = -\left(x_n - y_n\right) \quad , \tag{4.40}$$

L'équation (4.36) donne :

$$\frac{\partial y_n}{\partial s_n} = f'(s_n) \quad , \tag{4.41}$$

Ensuite, d'après l'équation (4.35), on obtient :

$$\frac{\partial s_n}{\partial b_{nm}} = h_m \quad , \tag{4.42}$$

La sensibilité du critère d'erreur par rapport aux poids  $b_{nm}$  est donc donnée par :

$$\frac{\partial J}{\partial b_{nm}} = -\left(x_n - y_n\right) f'\left(s_n\right) h_m , \qquad (4.43)$$
On en déduit, la règle d'adaptation des poids du réseau :

$$b_{nm}(t+1) = b_{nm}(t) - \eta(t) \frac{\partial J}{\partial b_{nm}} , \qquad (4.44)$$

Par conséquent, à l'itération t + 1, l'équation permettant l'ajustement des poids des connexions couche cachée-couche de sortie après une présentation d'un vecteur observation  $\boldsymbol{x}$  peut être formulée comme suit :

$$b_{nm}(t+1) = b_{nm}(t) + \eta(t) \left( x_n(t) - y_n(t) \right) f'(s_n(t)) h_m(t)$$
(4.45)

## 4.3.6 Adaptation des poids des connexions couche d'entréecouche cachée

De la même manière, on calcule la sensibilité du critère d'erreur par rapport aux poids des connexions couche d'entrée-couche cachée :

$$\frac{\partial J}{\partial a_{mn}} = \frac{\partial J}{\partial y_n} * \frac{\partial y_n}{\partial s_n} * \frac{\partial s_n}{\partial a_{mn}} , \qquad (4.46)$$

$$\frac{\partial J}{\partial a_{mn}} = -\sum_{n=1}^{N} \left( x_n - y_n \right) * f'(s_n) \left( \sum_{m=1}^{M} b_{nm} \frac{\partial h_m}{\partial e_m} * \frac{\partial e_m}{\partial a_{mn}} \right) , \qquad (4.47)$$

$$\frac{\partial J}{\partial a_{mn}} = -f'(e_m)x_n \left(\sum_{n=1}^N \left(x_n - y_n\right)f'(s_n)b_{nm}\right) \quad (4.48)$$

On en déduit la loi d'adaptation des poids :

$$a_{mn}(t+1) = a_{mn}(t) + \eta(t)f'(e_m)x_n\left(\sum_{n=1}^N (x_n - y_n)f'(s_n)b_{nm}\right)$$
(4.49)

Dans le cas où la fonction d'activation est une sigmoïde, on a :

$$\frac{\partial y_n}{\partial s_n} = f'(s_n) = \left(\frac{1}{1+e^{-s_n}}\right)' = y_n(1-y_n). \tag{4.50}$$

Dans le cas de fonctions d'activation de forme tangente hyperbolique, on a :

$$\frac{\partial y_n}{\partial s_n} = f'(s_n) = \left(\frac{1 - e^{-s_n}}{1 + e^{-s_n}}\right)' = 1 - y_n^2.$$
(4.51)

## 4.3.7 Comparaison entre les réseaux autoassociateurs linéaires et non linéaires

Après avoir présenté les réseaux autoassociateurs linéaires et non linéaires, la question qui se pose est qu'apporte la non linéarité dans le problème d'autoassociation?

Définissons ce que signifie la notion d'autoassociation [BFG95].

**Définition 4.1.** Soit  $\epsilon$  un réel positif, le vecteur observation x de dimension N est dit  $\epsilon$ -autoassocié s'il vérifie la relation suivante :

$$\frac{\|\boldsymbol{x} - \boldsymbol{y}\|_{\infty}}{\|\boldsymbol{y}\|_{\infty}} < \epsilon \tag{4.52}$$

où la norme infinie du vecteur x est donnée par  $||x||_{\infty} = \sup_{n=1,\dots,N} (x_n)$ , y étant la sortie correspondante du réseau.

Gori et al. ont établit le théorème suivant :

**Théorème 4.1.** Considérons un autoassociateur non linéaire à L couches avec des neurones de sortie linéaires. Si un vecteur observation est  $\epsilon$ -autoassocié alors

$$\forall \lambda \ge \lambda_{\epsilon} = \frac{(1+\epsilon) \|W^{(L-1)}\|_{\infty}}{\|\boldsymbol{x}\|_{\infty}}$$
(4.53)

le vecteur observation  $\lambda x$  n'est pas  $\epsilon$ -autoassocié.

Par conséquent, l'autoassociateur non linéaire, grâce à la présence de fonctions d'activations non linéaires sur ses couches cachées, autoassocie les groupements ou classes naturellement présents dans l'espace d'observation. Notons que, dans le cas d'un autoassociateur linéaire, la propriété précédente n'est plus valable. En d'autres termes, si une observation  $\boldsymbol{x}$  est  $\epsilon$ -autoassocié alors  $\lambda \boldsymbol{x}$  est  $\epsilon$ -autoassocié quelque soit le choix de  $\lambda$ .

Une implication pratique de ce résultat est que l'autoassociateur non linéaire se révèle être mieux adapté aux applications dans lesquelles nous avons besoin d'effectuer un groupement dans l'espace des observations.

Lastrucci et al. ont utilisé cette propriété pour la vérification de la parole [LGS94] où un réseau autoassociateur a été utilisé pour modéliser le signal de parole. Frosini et al. ont appliqué les réseaux autoassociateurs pour la détection de faux billets [FGP96].

#### 4.3.8 Exemple d'application

Reprenons le même exemple décrit dans la section précédente, et considérons un réseau autoassociateur non linéaire à 3 couches d'architecture 5x2x5. La fonction d'activation des neurones cachés est de type tangente hyperbolique et celle des neurones de sorties est linéaire. La base d'apprentissage X est normalisée entre -1 et 1 et l'algorithme de rétropropagation du gradient est utilisé avec un pas d'adaptation égale à 0.01. On appelle itération la présentation de toutes les observations de la base d'apprentissage.

La figure 4.9 illustre la projection sur le plan défini par les sorties des neurones de la couche cachée, des ensembles d'apprentissage et de test qui se superposent et se situent à l'intérieur d'un carré de côté 2 défini par les limites de la fonction d'activation tangente hyperbolique. Les erreurs de reconstruction par bande sont données sur la figure 4.10, on peut remarquer que les bandes étroites sont bien reconstruites. Par contre, au niveau des bandes larges, les erreurs reste toujours non négligeable.



FIG. 4.9 : La réponse des sorties de la couche centrale d'un autoassociateur non linéaire à 3 couches.



FIG. 4.10 : Les erreurs de reconstruction par bande obtenues au niveau des sorties de l'autoassociateur non linéaire à 3 couches.

## 4.4 Réseau autoassociateur non linéaire à cinq couches

#### 4.4.1 Architecture

L'architecture du réseau autoassociateur comprend cinq couches (cf. figure 4.11), soit une couche d'entrée et une couche de sortie comportant chacune N neurones et trois couches cachées ayant respectivement  $H_1$ ,  $H_2$  et  $H_3$  neurones. La fonction d'activation des neurones de la couche cachée centrale et de la couche de sortie est linéaire, par contre la fonction d'activation des neurones de la première et la troisième couches cachées sont de type tangente hyperbolique.

L'observation présentée à l'entrée du réseau est aussi présentée comme sortie désirée. De ce fait, le réseau doit fournir en sortie une estimation des observations présentées à son entrée. La première et la troisième couche cachée ont souvent le même nombre de neurones.



FIG. 4.11 : Réseau de neurones autoassociateur à trois couches cachées.

#### 4.4.2 Equations de fonctionnement

A la présentation d'une observation x à l'entrée du réseau, les équations des sorties des neurones de la  $l^{\grave{e}me}$  couches en fonction des neurones de la  $(l-1)^{\grave{e}me}$ couche sont données par (cf. figure 4.12) :

$$y_j = f\left(\sum_{i=1}^{n_{(l-1)}} w_{ji}^{(l)} y_i^{(l-1)}\right) , \qquad l = 1, \dots, 4,$$
(4.54)

où

- $y_n^{(0)} = x_n, n = 1, ..., N$ , représentent les composantes du vecteur observation présentée à l'entrée du réseau,
- $y_n^{(4)}$ ,  $n = 1, \ldots, N$ , représentent les composantes de l'estimation  $\hat{x}$  obtenue à la sortie du réseau,
- $n_{(l-1)}$  désigne le nombre de neurones sur la couche (l-1),
- $f(\cdot)$  est la fonction d'activation des neurones.



**FIG. 4.12** : Connexion du  $i^{ieme}$  neurone de la couche (l-1) vers le  $j^{ieme}$  neurone de la couche (l).

Comme les fonctions d'activation des neurones sont des fonctions sigmoïdes, les observations doivent être normalisées pour se situer à l'intérieur d'un hypercube de côté unité avant d'être exploitées pour l'apprentissage.

## 4.4.3 Apprentissage par rétropropagation du gradient de l'erreur

Pour apprendre une tâche à un réseau, on doit ajuster au fur et à mesure les poids des connexions de chaque neurone afin de minimiser la différence entre la sortie désirée et la sortie calculée par le réseau. Cette minimisation est obtenue par un calcul de la dérivée du critère d'erreur par rapport aux poids du réseau. La méthode la plus utilisée est la rétropropagation du gradient de l'erreur.

#### 4.4.3.1 Gradient du critère d'erreur partielle

Soit J(W) le critère d'erreur partielle défini comme le carré entre la sortie désirée  $\boldsymbol{x}$  qui est aussi l'entrée du réseau et la sortie estimée par ce réseau  $\boldsymbol{y}$ :

$$J(W) = \frac{1}{2} ||\boldsymbol{x}(t) - \boldsymbol{y}(t)||^2 = \frac{1}{2} \sum_{n=1}^{N} (x_n - y_n)^2, \qquad (4.55)$$

La règle d'adaptation des poids  $w_{ji}^{(l)}$  consiste à calculer la sensibilité du critère d'erreur J(W) par rapport aux poids du réseau :

$$w_{ji}^{l}(t+1) = w_{ji}^{(l)}(t) - \eta \frac{\partial J(W)}{\partial w_{ji}^{(l)}}, \qquad (4.56)$$

Le terme  $\frac{\partial J(W)}{\partial w_{ji}^{(l)}}$  est déduit des équations (4.54) à (4.55) par la règle de rétropropagation du gradient [RHW86].

A partir des équations (4.54) à (4.56) on déduit, l'équation d'adaptation des poids dite règle delta généralisée :

$$w_{ji}^{(l)}(t+1) = w_{ji}^{(l)}(t) + \eta(t) \ \delta_j^{(l)}(t) \ y_j^{(l-1)}(t), \tag{4.57}$$

• Pour les poids aboutissant à la couche de sortie l = 4:

$$\delta_n^{(4)}(t) = (x_n(t) - y_n(t)) f'(y_n^{(4)}), \qquad n = 1, \dots, N, \qquad (4.58)$$

• Pour les poids aboutissant à une couche cachée l, l < 4:

$$\delta_n^{(l)}(t) = f'\left(y_n^{(4)}\right) \sum_{k=1}^{n_{(l-1)}} \delta_k^{(l+1)}(t) \ w_{kj}^{(l+1)}(t). \tag{4.59}$$

#### Algorithme d'apprentissage :

- 1. Initialiser les poids du réseau aléatoirement à des valeurs voisines de 0. Fixer le coefficient d'apprentissage  $\eta$  et le nombre maximal d'itérations,
- 2. Choisir de façon aléatoire une observation  $\boldsymbol{x}$  de la base d'apprentissage et la présenter au réseau,
- Propagation en avant : Calculer les différentes sorties du réseau données par l'équation (4.54),
- 4. Actualiser les poids du réseau ( les équations (4.57) à (4.57),
- Incrémenter t (t=t+1) et aller en 2 jusqu'à ce que l'erreur est faible ou la valeur maximale des itérations est atteinte.

#### 4.4.4 Optimisation de l'architecture du réseau

Dans un Perceptron multicouche autoassociateur, le nombre des neurones de la couche d'entrée est égal au nombre des neurones de la couche de sortie qui est aussi égal à la dimension N de l'espace d'observation. Comme le nombre des neurones sur la couche de projection est peut être fixé à deux ou à trois, seuls les nombres des neurones de la première et la troisième couches cachées sont inconnus. Pour simplifier le problème on impose que les deux couches ont le même nombre H de neurones et nous proposons d'optimiser ce nombre. L'une des méthodes permettant de choisir le nombre H de neurones des couches cachées consiste à utiliser des fonctions exprimant un compromis entre le nombre total des poids du réseau et le logarithme de l'erreur moyenne de reconstitution des observations. Kramer a utilisé deux fonctions proposées par Akaike dans le cadre de l'identification des séries temporelles [Kra91] : l'erreur de prédiction finale (FPE pour Final Prediction Error) :

$$FPE(H) = \bar{E}\left(\frac{1+\frac{N_w}{NP}}{1-\frac{N_w}{NP}}\right),\tag{4.60}$$

et le critère informationnel d'Akaike (AIC pour Akaike Information Criterion) [Aka72] :

$$AIC(H) = \ln(\bar{E}) + 2\frac{N_w}{NP}, \qquad (4.61)$$

où :

- $\ln(\cdot)$  est le logarithme naturel,
- $\overline{E}$  est la fonction erreur moyenne définie par :

$$\bar{E} = \frac{1}{ND} \sum_{n=1}^{N} \sum_{d=1}^{D} \left( x_{nd} - \hat{x}_{nd} \right)^2 .$$
(4.62)

- $N_p$  représente le nombre total des poids du réseau.
- N et P sont respectivement la dimension et le cardinal de la base d'apprentissage.

Nous avons utilisé ces fonctions pour optimiser l'architecture du réseau autoassociatif. Le principe consiste à faire varier le nombre de neurones sur la première et la troisième couche cachée et à calculer les valeurs de AIC et FPE pour ce nombre de neurones. Le nombre de neurones minimisant ces deux fonctions est retenu [Bet 95], [Ham 95].

#### 4.4.5 Exemple d'application

#### 4.4.5.1 Optimisation de l'architecture

Nous considérons un réseau autoassociateur à 5 couches. Rappelons, la fonction d'activation des neurones de la couche centrale et ceux de la couche de sortie est linéaire, par contre la fonction de la première et la troisième couche cachée est non linéaire et de type tangente hyperbolique. L'algorithme de rétropropagation du gradient est utilisé avec un pas d'adaptation égal à 0.01. L'optimisation de l'architecture du réseau est effectuée en faisant varier le nombre H des neurones cachés et en calculant les critères AIC et FPE pour chaque nombre. Le nombre optimum est celui qui correspond à la fois au minimum du critère AIC et au maximum du critère FPE. Les figures 4.13(a) et 4.13(b) montrent l'évolution des critères FPE et AIC en fonction du nombre de neurones H. On remarque que, dans les deux cas, l'optimum est obtenu pour H = 4.

Les figures 4.14 et 4.15 montrent respectivement la projection de la base d'apprentissage et de test et les erreurs de reconstruction par bande obtenues en utilisant un réseau autoassociateur non linéaire à 5 couches avec 4 neurones dans la couche centrale.



**FIG. 4.13 :** L'évolution des critères AIC (figure (a)) et FPE (figure (b)) en fonction du nombre de neurones H.



FIG. 4.14 : La réponse des sorties de la couche centrale d'un autoassociateur non linéaire à 5 couches.



FIG. 4.15 : Les erreurs de reconstruction par bande obtenues au niveau des sorties de l'autoassociateur non linéaire à 5 couches.

#### 4.4.5.2 Simulation d'une dérive de fonctionnement

Pour mettre en évidence l'intérêt de notre approche de suivi de fonctionnement par visualisation plane et reconstruction par bande, nous allons étudié les réponses du réseau autoassociateurs non linéaires à 5 couches lors de l'apparition d'une composante perturbatrice simulant ainsi une dérive du fonctionnement du système de deux engrenages. Nous allons considéré deux cas de figures : La composante perturbatrice est situé au sein d'une bande étroite numéro 1 (resp. large numéro 2) à la fréquence  $f_E$  (resp.  $f_L$ ).

Les amplitudes et les pourcentages d'énergie apportée par la composante perturbatrice dans les bandes étroite et large sont résumés dans le tableau de la figure 4.16. Nous avons crée deux bases de test constituées chacune de 290 observations.

	$f_E = 550 \text{Hz}$	$f_L = 650 \mathrm{Hz}$
$A_{per}$	$0.5 * A_1$	$1 * A_2$
$E_{per}$	$16\% * E_1$	$10\% * E_2$

FIG. 4.16 : Pourcentage d'énergie apportée par la composante perturbatrice au sein d'une bande étroite et large.

Dans le cas où la composante perturbatrice est situé au niveau de la bande étroite : on constate que l'ensemble des projections de la base de tests notées o sont décalés par rapport aux projections des observations de la base d'apprentissage (cf. figure 4.17). On peut noter aussi une augmentation considérable, au niveau des erreurs de reconstruction par bande, lors de la présentation de la base de test (cf. figure 4.18).

Dans le cas où la composante perturbatrice est situé au niveau de la bande large : les projections des deux bases se superposent (cf. figure 4.19), par contre, au niveau des erreurs de reconstruction, on peut noter la présence d'une légère augmentation au niveau de la bande numéro 2 (cf. figure 4.20).



FIG. 4.17 : La réponse des sorties de la couche centrale d'un autoassociateur non linéaire à 5 couches.



FIG. 4.18 : Les erreurs de reconstruction par bande obtenues au niveau des sorties de l'autoassociateur non linéaire à 5 couches.



FIG. 4.19 : La réponse des sorties de la couche centrale d'un autoassociateur non linéaire à 5 couches.



FIG. 4.20 : Les erreurs de reconstruction par bande obtenues au niveau des sorties de l'autoassociateur non linéaire à 5 couches.

### 4.5 Conclusion

Nous avons présenté trois réseaux autoassociateurs pour le suivi de fonctionnement par projection plane et par reconstruction des entrées. Le premier est un autoassociateur linéaire dont l'apprentissage est direct et déduit des méthodes de l'algèbre linéaire. Ce réseau réalise une analyse en composantes principales. Le second autoassociateur est non linéaire et est à trois couches. Son apprentissage est effectué par la méthode de rétropropagation de l'erreur. Ce type de réseaux apporte peu d'améliorations par rapport à l'autoassociateur linéaire. Le réseau autoassociateur à cinq couches présente, par contre, de nettes améliorations en projection et reconstruction par rapport aux deux premiers.

Les trois réseaux ont été appliqués sur un signal simulé pour la vérification du fonctionnement normal et pour la détection d'un défaut pouvant apparaître sur une bande étroite ou large du spectre du signal. Les projections obtenues par le réseau à cinq couches différencient bien la classe de fonctionnement normal de la classe de défaut. De plus les reconstructions obtenues par ce réseau permettent de localiser le défaut et quantifier sa gravité.

Les résultats obtenus par visualisation plane offrent un suivi interactif de l'état de fonctionnement du système à surveiller qui a l'avantage de montrer sur un écran graphique l'état courant. Cet état manifesté par un point sur l'écran peut se situer dans le domaine normal ou dériver vers une situation anormale. Par contre, l'inconvénient de la projection plane est que deux points loin dans l'espace multidimensionnel peuvent être représentés par deux points voisins sur le plan de projection. De plus la méthode ne nous renseigne pas sur la gravité du défaut ni sur sa localisation dans la bande de fréquence dans laquelle il apparaît. Pour compléter notre démarche, il est important d'évaluer les erreurs de reconstruction sur chaque bande.

Le chapitre suivant traite le cas d'une application réelle de suivi de fonctionnement d'une éolienne.

# Chapitre 5

# Suivi de fonctionnement de l'éolienne par réseaux de neurones autoassociateurs

### 5.1 Introduction

L'éolienne dispose d'un système de commande qui oriente la nacelle et les pales dans la direction du vent afin d'optimiser l'énergie produite. De plus, un dispositif de sécurité permet à l'éolienne de produire une énergie de qualité satisfaisante tout en la protégeant des conditions climatiques extrêmes. Ce dispositif compare la vitesse du vent à des seuils limites. Pour un vent faible inférieur à 5 m/s, l'automate découple l'éolienne du réseau électrique alors que pour un vent fort supérieur à 18 m/s l'éolienne est mise en arrêt en position drapeau.

Cependant, ce dispositif de sécurité, basé sur la simple comparaison des signaux des capteurs à des valeurs limites, est loin d'être satisfaisant pour détecter des pannes naissantes ou évolutives. Pour cela, nous proposons d'analyser, au moyen de techniques sophistiquées de traitement de signal, les vibrations des différents composants afin de détecter toute dégradation pouvant affecter le fonctionnement normal. En particulier, nous nous sommes intéressés aux deux composants les plus

#### CHAPITRE 5. SUIVI DE FONCTIONNEMENT DE L'ÉOLIENNE PAR RNA89

sensibles qui sont le multiplicateur et la génératrice. Pour cela, nous avons installé deux accéléromètres sur le multiplicateur et un autre sur la génératrice.

Comme il est difficile, voire dangereux, de créer une panne sur un système réel en pleine production, notre approche de suivi de fonctionnement est basée sur la modélisation de son état normal.

Pour caractériser ce fonctionnement normal, une base de données de signaux de vibrations enregistrés dans différentes conditions climatiques a été constituée. Suite à l'analyse de cette base, nous avons obtenu un ensemble de spectrogrammes représentant le fonctionnement normal. Cependant, ces spectrogrammes sont trop riches en information pour être directement exploités par un système de suivi automatique. Pour cela, nous avons repéré les fréquences d'amplitudes les plus significatives et défini des bandes d'incertitude autour de celles-ci.

La surveillance automatique de l'éolienne revient alors à suivre l'évolution des énergies dans chacune de ces bandes de fréquences. Pour ce faire, plusieurs architectures des réseaux autoassociateurs ont été employé pour mémoriser les différents spectres correspondant à différentes situations de fonctionnement normal de l'éolienne. Nous proposons d'exploiter à la fois :

- ses propriétés de projection pour visualiser sur un écran graphique la position courante du fonctionnement de l'éolienne [BEBH98] [NEG<sup>+</sup>99],
- ses propriétés d'autoassociation pour délimiter le domaine du fonctionnement normal de l'éolienne et vérifier son état courant de fonctionnement. En fait, la comparaison entre les valeurs des entrées et des sorties du réseau permet de calculer les erreurs de reconstruction afin de vérifier si le spectre correspond ou non à un fonctionnement normal [EBPH], [EBH99], [HB95] et [BEHP98].

Pour mettre en évidence l'intérêt de cette approche, nous avons procédé à des expériences en simulant des situations de dégradations par génération artificielle de certaines fréquences qui sont ajoutées aux signaux réels.

# 5.2 Constitution de la base d'apprentissage et de généralisation

Nous allons appliquer différentes architectures neuronales sur les trois signaux de vibrations issus des trois accéléromètres installés sur le multiplicateur de vitesse en positions axiale et radiale et sur la génératrice en position axiale.

Pour les trois capteurs, nous considérons deux bases de données. La première base d'apprentissage est constituée de 24 enregistrements de 15 secondes chacun, soit environ 470Mo de données à traiter, sur une période de 3 jours. La seconde base, dite base de test ou de généralisation, est constituée de 24 enregistrements de 15 secondes chacun, étalés sur une période de 15 jours. Ces enregistrements sont collectés en phase d'exploitation de l'éolienne, c'est à dire lorsqu'elle est connectée au réseau EDF. En effet, la procédure d'acquisition événementielle déclenche l'enregistrement automatique des données, pendant une durée de 15 secondes, lorsque l'éolienne fournit un courant dans le réseau électrique EDF.

Nous supposons que ces enregistrements sont représentatifs au fonctionnement normal de la machine.

# 5.2.1 Réduction des spectrogrammes des signaux de vibrations de l'accéléromètre installé sur le multiplicateur en position axiale

Nous considérons les signaux de vibrations issus de l'accéléromètre installé sur le multiplicateur en position axiale. Rappelons que la fréquence d'échantillonnage  $f_e$  a été fixée à 31KHz.

Pour déterminer la densité spectrale d'énergie du signal, nous avons utilisé une fenêtre de Hanning de taille  $N = 2^{15} = 32768$  échantillons, ce qui revient à calculer

le spectre sur une durée un peu plus d'une seconde (cf. figure 5.1). La résolution fréquentielle est donnée par :

$$\mathcal{R} = f_e/N \simeq 1 \mathrm{Hz}$$
 .

Par conséquent les fréquences qui sont inférieures à la résolution ne seront pas décelées.



**FIG. 5.1 :** Densité spectrale d'énergie obtenue en utilisant une fenêtre de Hanning comprenant  $2^{15}$  échantillons, la fréquence d'échantillonnage est de 31KHz.

Comme le spectre est trop riche en information, nous avons effectué un "zoom" sur 4 zones de fréquences :

- [0 500]Hz,
- [500 1000]Hz,
- [1000 1500]Hz et
- [1500 2400]Hz.

Les densités spectrales d'énergie et les spectrogrammes correspondants à ses 4 zones sont illustrés sur les figures 5.2 à 5.9.



FIG. 5.2 : Densité spectrale d'énergie, zone 1 ([0-500]Hz).



FIG. 5.3 : Spectrogramme d'énergie du signal, zone 1 ([0-500]Hz).



FIG. 5.4 : Densité spectrale d'énergie, zone 2 ([500-1000]Hz).



FIG. 5.5 : Spectrogramme d'énergie du signal, zone 2 ([500-1000]Hz).



FIG. 5.6 : Densité spectrale d'énergie, zone 3 ([1000-1500]Hz).



FIG. 5.7 : Spectrogramme d'énergie du signal, bande 3 ([1000-1500]Hz).



FIG. 5.8 : Densité spectrale d'énergie, zone 4 ([1500-2400]Hz).



FIG. 5.9 : Spectrogramme d'énergie du signal, zone 4 ([1500-2400]Hz).

Cette représentation en spectrogramme montre clairement que les fréquences des principaux pics fluctuent légèrement autour des fréquences théoriques. On peut remarquer aussi que les amplitudes varient énormément au cours du temps. Cette variation peut être liée aux conditions climatiques qui peuvent changer à chaque instant.

Une analyse fine de chaque zone nous a permis de retrouver la majorité des fréquences de rotation des éléments mécaniques constituant la machine déterminées à partir des équations théoriques (cf. Rapport de fin phase 2).

Les 4 zones contiennent plusieurs fréquences de rotation, les principales fréquences sont :

- La fréquence de rotation de la génératrice à 25Hz et son harmonique à 100Hz,
- La fréquence de rotation à 40Hz de l'engrenage reliant l'arbre lent à l'arbre intermédiaire 1 du multiplicateur,
- La fréquence de rotation à 555Hz de l'engrenage reliant l'arbre intermédiaire 2 à l'arbre rapide du multiplicateur et ses harmoniques à 1105Hz et 2210Hz,
- Il faut noter que la fréquence de rotation à 141Hz de l'engrenage reliant l'arbre intermédiaire 1 à l'arbre intermédiaire 2 du multiplicateur ainsi que ses harmoniques à 282Hz et 423Hz existent dans le spectrogramme avec des amplitudes très faibles.
- On trouve aussi des fréquences dues à la rotation de certains engrenages et des roulements à billes de la machine non répertoriées dans la liste qui nous a été fournie par le constructeur, nous pouvons citer :
  - 210Hz,
  - 300Hz,
  - 680Hz,
  - 850Hz et
  - 2360Hz.

Le tableau de la figure 5.10 résume les 11 principales fréquences retrouvées dans le spectre dont les amplitudes sont supérieures à un seuil qui dépend de l'élément à surveiller. Dans le cas du capteur axial du multiplicateur, le seuil est fixé à  $5 * 10^{-3}$ . Elles représentent les fréquences de rotation des différents éléments engendrant des vibrations de fortes amplitudes.

	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$	<i>f</i> <sub>7</sub>	$f_8$	$f_9$	$f_{10}$	$f_{11}$
Fréquences	25	40	100	210	300	555	680	850	1105	2210	2360
en (Hz)											

FIG. 5.10 : Fréquences à surveiller.

La méthode de segmentation consiste à diviser la densité spectrale d'énergie en un ensemble de bandes. Des bandes étroites autour des 11 principaux pics à surveiller et des bandes plus au moins larges constituées entre ces bandes étroites. Cela revient à surveiller 20 bandes de fréquences.

La question qui se pose est comment choisir la largeur de ces bandes?

Pour choisir la largeur de chacune des bandes, nous avons étudié les fluctuations autour des fréquences des principaux pics.

En fait, la génératrice produisant de l'énergie électrique est une machine dite asynchrone. Cette technologie présente les caractéristiques de vitesse suivantes :

Lorsque le vent est égal à 5 m/s, la force déployée sur le rotor est tout juste suffisante pour compenser les frottements de l'ensemble "multiplicateur - génératrice". La vitesse du rotor est d'environ 51 tr/mn (vitesse de l'arbre lent du multiplicateur). La vitesse de l'arbre rapide du multiplicateur et donc la vitesse de l'arbre de la génératrice est égale exactement à 1500 tr/mn. Aucune puissance n'est transmise en aval de la génératrice. Le courant de sortie produit par la machine électrique est nul. Lorsque le vent dépasse 5 m/s, la machine commence à produire de l'énergie. Cela se traduit par une augmentation des vitesses de la chaîne de transmission mécanique pales-multiplicateur-génératrice. La puissance fournie est une fonction linéaire de la variation de vitesse du rotor, appelée glissement et noté g (cf. figure 5.11).

$$g = V_r / V_s - 1$$

où  $V_r$  est la vitesse de rotation du rotor de la génératrice et  $V_s$ =1500 tr/mn est la vitesse de synchronisme de la génératrice.

- Lorsque le vent atteint 14 m/s environ, la vitesse de l'arbre de la génératrice atteint 1515 tr/mn et le glissement vaut 1%. La puissance fournie est nominale et égale à 300kW.
- La machine reste connectée si pendant une fraction de seconde, le vent passe sous le seuil de 5 m/s. En conséquence, le glissement devient négatif et peut atteindre -0.13%, la vitesse de l'arbre de la génératrice décroît en deçà de 1500 tr/mn. Comme la machine est réversible, elle devient moteur et consomme de l'énergie électrique.
- En régime de vent supérieur à 14 m/s, la machine régule en permanence la puissance fournie par modification de l'angle d'orientation des pales. Une rafale peut provoquer instantanément un dépassement de la vitesse nominale du rotor de façon à provoquer en quelques fractions de seconde un glissement pouvant atteindre 1.3% ou 1.4%.

La figure 5.11 montre l'évolution de l'énergie produite lorsque la vitesse de rotation de la génératrice varie entre 1498 tr/mn et 1520 tr/mn.

Si on considère  $f_i$  la fréquence de rotation d'un élément de la machine lorsque la génératrice tourne à la vitesse de synchronisme. Lorsque la vitesse de rotation de la génératrice atteint des valeurs comprises entre 1498 tr/mn et 1520 tr/mn, alors la fréquence  $f_i$  vérifie :

$$\frac{1498}{1500}f_i \le f_i \le \frac{1520}{1500}f_i \; .$$

Le tableau de la figure 5.12 résume les limites inférieures (resp. supérieures) notées





FIG. 5.11 : Puissance fournie en fonction de la vitesse de rotation du rotor de la génératrice.

 $f_j$  (resp.  $f_k)$  des principales fréquences à surveiller.

$f_i$ (Hz)	$f_j = 0.9987 * f_i (\text{Hz})$	$f_k = 1.0133 * f_i (\text{Hz})$
25	25	25
40	40	41
100	100	101
210	210	213
300	300	304
550	549	557
680	679	689
850	849	861
1100	1099	1115
2200	2197	2229
2360	2357	2391

FIG. 5.12 : Valeurs prises par les principales fréquences de rotation lorsque la vitesse de rotation de la génératrice varie entre 1498 tr/mn et 1520 tr/mn.

Les figures 5.13, 5.14 et 5.15 montrent les fluctuations autour des fréquences 555, 680 et 850. On remarque que les fréquences des 3 principaux pics varient respectivement dans les bandes de fréquences [549-557]Hz, [679-689]Hz et [849-861]Hz. Ceci confirme les résultats théoriques obtenus par l'étude du glissement.

Pour assurer la présence des pics à surveiller au sein d'une bande étroite, nous avons fixé respectivement les limites inférieures et supérieures des bandes de fréquences à  $f_i - \mathcal{R}$  et  $f_j + \mathcal{R}$ , avec  $\mathcal{R}$  est la résolution fréquentielle.



FIG. 5.13 : Fluctuations de la fréquence du pic principal autour de la fréquence 555Hz.



FIG. 5.14 : Fluctuations de la fréquence du pic principal autour de la fréquence 680Hz.



FIG. 5.15 : Fluctuations de la fréquence du pic principal autour de la fréquence 850Hz.

Après avoir déterminé, la taille des bandes de fréquences à surveiller, nous avons calculé l'énergie contenue dans chacune de ces bandes. L'allure du spectre réduit est donnée par la figure 5.16. Donc, chaque spectre réduit représente un point dans un espace de 20 dimensions. L'ensemble des spectres réduits constitue un nuage de points représentant le fonctionnement normal de l'élément à surveiller. Dans notre cas, nous avons généré un ensemble d'apprentissage de 432 points dans un espace de dimension 20. Autrement dit, nous disposons d'une matrice de 20x432 éléments.



FIG. 5.16 : Spectre réduit en 20 bandes.

# 5.2.2 Réduction des spectrogrammes des signaux de vibrations issus de l'accéléromètre installé sur le multiplicateur en position radiale

La densité spectrale d'énergie du signal d'accélération radiale du multiplicateur est représentée sur la figure 5.17. Elle a été obtenue en utilisant une fenêtre de Hanning de taille  $2^{15}$  échantillons.



FIG. 5.17 : Densité spectrale d'énergie du signal d'accélération radiale du multiplicateur obtenue en utilisant une fenêtre de Hanning comprenant  $2^{15}$  échantillons, la fréquence d'échantillonnage est de 31KHz.

Les figures 5.18 à 5.25 montrent les densités spectrales d'énergie et les spectrogrammes correspondant aux quatre zones définies dans le paragraphe précédent.



FIG. 5.18 : Densité spectrale d'énergie, zone 1 ([0-500]Hz).



FIG. 5.19 : Spectrogramme d'énergie du signal, zone 1 ([0-500]Hz).



FIG. 5.20 : Densité spectrale d'énergie, zone 2 ([500-1000]Hz).



FIG. 5.21 : Spectrogramme d'énergie du signal, zone 2 ([500-1000]Hz).



FIG. 5.20 : Densité spectrale d'énergie, zone 2 ([500-1000]Hz).



FIG. 5.21 : Spectrogramme d'énergie du signal, zone 2 ([500-1000]Hz).



FIG. 5.22 : Densité spectrale d'énergie, zone 3 ([1000-1500]Hz).



FIG. 5.23 : Spectrogramme d'énergie du signal, zone 3 ([1000-1500]Hz).



FIG. 5.24 : Densité spectrale d'énergie, zone 4 ([1500-2400]Hz).



FIG. 5.25 : Spectrogramme d'énergie du signal, zone 4 ([1500-2400]Hz).
A partir d'une analyse fine des différents spectrogrammes, nous avons sélectionné les 11 fréquences à surveiller dont les amplitudes dépassent le seuil fixé à  $1*10^{-3}$ . Cela revient, donc, à surveiller 20 bandes dont les tailles sont déterminées en fonction du glissement de la machine. Le tableau de la figure 5.26 résume les limites inférieures et supérieures des principales fréquences du spectrogramme.

$f_i$ (Hz)	$f_j = 0.9987 * f_i \; (\text{Hz})$	$f_k = 1.0133 * f_i$ (Hz)
100	100	101
280	280	284
300	300	304
480	479	486
550	549	557
680	679	689
850	849	861
1100	1099	1115
2000	1997	2027
2200	2197	2229
2360	2357	2391

FIG. 5.26 : Valeurs prises par les principales fréquences de rotation lorsque la vitesse de rotation de la génératrice varie entre 1498 tr/mn et 1520 tr/mn.

## 5.2.3 Réduction des spectrogrammes des signaux de vibrations issus de l'accéléromètre installé sur la génératrice

La densité spectrale d'énergie du signal d'accélération axiale de la génératrice a été calculée en utilisant une fenêtre de Hanning comprenant 2<sup>15</sup> échantillons pour une fréquence d'échantillonnage de 31KHz (cf. figure 5.27).

Les figures 5.28 à 5.35 montrent les densités spectrales d'énergie et les spectrogrammes correspondant aux 4 zones. On peut remarquer que les fréquences à surveiller fluctuent autour des fréquences théoriques alors que les amplitudes varient de manière significative.



**FIG. 5.27 :** Densité spectrale d'énergie du signal d'accélération axiale de la génératrice obtenue en utilisant une fenêtre de Hanning comprenant 2<sup>15</sup> échantillons, la fréquence d'échan-tillonnage est de 31KHz.



FIG. 5.28 : Densité spectrale d'énergie, zone 1 ([0-500]Hz).



FIG. 5.29 : Spectrogramme d'énergie du signal, zone 1 ([0-500]Hz).



FIG. 5.30 : Densité spectrale d'énergie, zone 2 ([500-1000]Hz).



FIG. 5.31 : Spectrogramme d'énergie du signal, zone 2 ([500-1000]Hz).



FIG. 5.32 : Densité spectrale d'énergie, zone 3 ([1000-1500]Hz).



FIG. 5.33 : Spectrogramme d'énergie du signal, zone 3 ([1000-1500]Hz).



FIG. 5.34 : Densité spectrale d'énergie, zone 4 ([1500-2400]Hz).



FIG. 5.35 : Spectrogramme d'énergie du signal, zone 4 ([1500-2400]Hz).

Etant donné que le spectrogramme de vibrations de la génératrice est plus riche en fréquence que celui du multiplicateur, nous avons été amené à surveiller 18 fréquences dont les amplitudes sont supérieures en moyenne au seuil fixé à  $1 * 10^{-3}$ . Ce qui revient à segmenter notre spectre en 34 bandes dont 18 bandes étroites autour des principaux pics et 16 bandes plus au moins larges constituées entre les bandes étroites.

$f_i$ (Hz)	$f_j = 0.9987 * f_i \; (\text{Hz})$	$f_k = 1.0133 * f_i \; (\text{Hz})$
15	15	15
100	100	101
200	200	203
300	300	304
550	549	557
680	679	689
850	849	861
900	899	912
1100	1099	1115
1200	1198	1216
1500	1498	1520
1700	1698	1723
1900	1898	1925
2000	1997	2027
2100	2097	2128
2200	2197	2229
2300	2297	2331
2360	2357	2391

**FIG. 5.36 :** Valeurs prises par les principales fréquences de rotation lorsque la vitesse de rotation de la génératrice varie entre 1498 tr/mn et 1520 tr/mn.

## 5.3 Suivi de fonctionnement du multiplicateur par réseaux autoassociateurs

L'ensemble des spectres réduits sont utilisés pour caractériser le fonctionnement normal de l'éolienne au moyen d'un réseau de neurones autoassociateur. Les entrées et les sorties du réseau sont les valeurs des énergies contenues dans chacune des bandes. Le réseau, durant sa phase d'apprentissage, mémorise dans ses poids synaptiques, l'état normal de la machine décrit par les spectres réduits. Lors de la phase de test, on vérifie si l'état courant est normal ou non.

Dans ce paragraphe, nous présentons les résultats obtenus par différentes architectures de réseaux autoassociateurs. Tout d'abord, nous étudions les résultats d'un autoassociateur linéaire à trois couches. Ensuite, nous comparons les résultats des projections et des erreurs de reconstruction de ce réseau avec ceux obtenues par un autoassociateur non linéaire à trois couches, puis par un réseau autoassociateur à cinq couches. Notons que cette étude porte sur les signaux de vibrations issus des deux accéléromètres installés sur le multiplicateur en positions axiale et radiale.

#### 5.3.1 Autoassociateur linéaire à trois couches

Nous considérons un réseau de neurones autoassociateur linéaire d'architecture 20x2x20. L'algorithme du calcul direct des poids du réseau, basé sur la méthode d'analyse en composantes principales, permet d'obtenir respectivement les valeurs propres et les pourcentages de la variance totale de l'information expliquée par chacune des valeurs propres consignées dans les deux figures 5.37 et 5.38 respectivement pour les signaux issus de l'accéléromètre en position axiale et radiale. Notons que le plan principal correspondant aux deux valeurs propres maximales explique 74% de la variance totale de l'ensemble d'apprentissage pour le signal issu de l'accéléromètre en position axiale et 62% pour le signal issu de l'accéléromètre en position radiale.

Une fois les matrices de poids et de biais déterminés, nous nous intéressons aux sorties des deux neurones de la couche centrale définissant les axes de projection.

Valeurs propres	Quantité d'informations expliquée
$5.22 * 10^4$	65%
$7.21 * 10^3$	9%

**FIG. 5.37 : Accéléromètre axial :**Pourcentage de la variance totale des observations expliqué par chaque valeur propre.

Valeurs propres	Quantité d'informations expliquée	
$3.85 * 10^4$	51.4%	
$8 * 10^3$	10.6%	

**FIG. 5.38 : Accéléromètre radial :** Pourcentage de la variance totale des observations expliqué par chaque valeur propre.

Nous présentons successivement les observations de la base d'apprentissage à l'entrée du réseau. Les figures 5.39 et 5.41 montrent les projections de la base d'apprentissage pour les deux capteurs.

Les erreurs de reconstruction par bande sont données sur les figures 5.40 et 5.42. On remarque que l'erreur au niveau des bandes étroites est faible. Par contre, l'erreur au niveau des bandes larges est voisine de 1, ce qui signifie que le réseau effectue un moyennage au niveau de ces bandes.

On peut aussi constater que, au niveau du capteur axial, les deux bandes numérotées 12 et 14, qui sont centrées respectivement sur les deux pics principaux du spectre autour des fréquences  $f_7 = 680$ Hz et  $f_8 = 850$ Hz, sont bien reconstruites. De même, au niveau du capteur radial, les bandes numérotées 10 et 12, centrées autour des mêmes pics, sont également bien reconstruites. En effet, un réseau autoassociateur linéaire projete sur le plan défini par les deux sorties des deux neurones de la couche centrale du réseau autoassociateur et reconstruit à sa sortie l'ensemble des observations de dimension 20 en utilisant les deux plus grandes valeurs propres expliquant le maximum d'information.

#### CHAPITRE 5. SUIVI DE FONCTIONNEMENT DE L'ÉOLIENNE PAR RNA117

Les figures 5.43 et 5.45 montrent les projections des bases d'apprentissage et de test sur le plan défini précédemment. On peut remarquer que les projections des bases d'apprentissage et de test se superposent. Les erreurs de reconstruction par bandes lors des phases d'apprentissage et de test, obtenues par le réseau autoassociateur linéaire, pour les deux capteurs, sont illustrées dans les figures 5.44 et 5.46 respectivement.



FIG. 5.39 : Accéléromètre axial : Le plan de projection obtenu par les sorties de la couche centrale d'un autoassociateur linéaire à 3 couches.



FIG. 5.40 : Accéléromètre axial : Les erreurs de reconstruction par bande obtenues au niveau des sorties de l'autoassociateur linéaire à 3 couches.



FIG. 5.41 : Accéléromètre radial : Le plan de projection obtenu par les sorties de la couche centrale d'un autoassociateur linéaire à 3 couches.



FIG. 5.42 : Accéléromètre radial : Les erreurs de reconstruction par bande obtenues au niveau des sorties de l'autoassociateur linéaire à 3 couches.



**FIG. 5.43 : Accéléromètre axial :** Le plan de projection obtenu par les sorties de la couche centrale de l'autoassociateur linéaire à 3 couches.



FIG. 5.44 : Accéléromètre axial : Les erreurs de reconstruction par bande obtenues au niveau des sorties de l'autoassociateur linéaire à 3 couches.



**FIG. 5.45 : Accéléromètre radial :** Le plan de projection obtenu par les sorties de la couche centrale de l'autoassociateur linéaire à 3 couches.



FIG. 5.46 : Accéléromètre radial : Les erreurs de reconstruction par bande obtenues au niveau des sorties de l'autoassociateur linéaire à 3 couches.

#### 5.3.2 Autoassociateur non linéaire à trois couches

Dans la suite et pour ne pas alourdir l'étude, nous présentons uniquement les résultats obtenus en utilisant le signal de vibration issu de l'accéléromètre en position axiale.

Nous considérons un réseau autoassociateur non linéaire à trois couches d'architecture 20x2x20. La fonction d'activation des neurones cachés est de type tangente hyperbolique et celle des neurones de sorties est linéaire. Les bases d'apprentissage et de test ont été normalisées entre -1 et +1. Le pas d'adaptation de l'algorithme de rétropropagation est fixé à 0.01. Les figures 5.47 et 5.48 montrent l'évolution de la moyenne et de l'écart type de l'erreur au cours de la phase d'apprentissage et de la phase de test. On peut remarquer que les valeurs de l'erreur moyenne et de l'écart type relatives à la base d'apprentissage se stabilisent à 4% et 3% respectivement, tandis que pour la base de test, elles sont respectivement égales à 8.3% et 4.2%.

La figure 5.49 illustre la projection sur le plan défini par les sorties des neurones de la couche cachée, pour les ensembles d'apprentissage et de test. On remarque que les projections des points de la base de test se situent dans le même domaine que ceux de la base d'apprentissage. Les erreurs de reconstruction par bandes sont représentées sur la figure 5.49. On peut constater facilement que ces erreurs sont nettement moins importantes que celles fournies en sorties de l'autoassociateur linéaire (cf. figure 5.40).



FIG. 5.47 : La moyenne et l'écart type de l'erreur en fonction du nombre d'itérations calculés sur l'ensemble d'apprentissage pour un autoassociateur non linéaire à 3 couches.



**FIG. 5.48 :** La moyenne et l'écart type en fonction du nombre d'itérations de l'erreur calculés sur l'ensemble de test pour un autoassociateur non linéaire à 3 couches.



**FIG. 5.49 :** Plan de projection obtenu par les sorties de la couche centrale de l'autoassociateur non linéaire à 3 couches.



FIG. 5.50 : Erreurs de reconstruction calculées sur l'ensemble d'apprentissage et de test à la sortie de l'autoassociateur non linéaire à 3 couches.

#### 5.3.3 Autoassociateur non linéaire à cinq couches

Rappelons que l'architecture de ce réseau comprend N neurones dans la couche d'entrée et dans celle de sortie, puisque le réseau réalise la fonction identité. Les trois couches cachées contiennent respectivement  $H_1$ ,  $H_2$  et  $H_3$  neurones. Notons que, afin d'avoir une projection sur le plan de l'ensemble des observations de la base d'apprentissage et de test, le nombre de neurones de la couche centrale est  $H_2 = 2$ .

Si le nombre de neurones des couches d'entrée et de sortie sont parfaitement connus et correspondent à la dimension de l'espace d'observation, les nombres de neurones  $H_1$  et  $H_3$  de la première et de la troisième couches cachées sont, quant à eux, inconnus et délicats à choisir. En effet, un nombre important de neurones sur ces couches cachées risque de détériorer les performances du réseau au niveau de la généralisation, alors qu'un nombre réduit de neurones sur ces couches risque de fournir un modèle de représentation trop simple.

Nous proposons d'optimiser le nombre des neurones sur la première et la troisième couches cachées grâce à la théorie de l'information. Pour ce faire, nous utilisons le critère d'Akaike et le critère FPE décrits dans le chapitre précédent. On peut noter que,  $H_1 = H_3 = H$  puisque le réseau est considéré symétrique par rapport à la couche centrale de projection.

L'optimisation de l'architecture du réseau est effectuée en faisant varier le nombre H des neurones cachés et en calculant les critères AIC et FPE pour chaque nombre. Le nombre optimum est celui qui correspond à la fois au minimum du critère AIC et au maximum du critère FPE. Les figures 5.51(a) et 5.51(b) montrent l'évolution des critères FPE et AIC en fonction du nombre de neurones H. On remarque que, dans les deux cas, l'optimum est obtenu pour H = 5.

Nous utilisons la même base d'apprentissage pour entraîner un réseau autoassociateur non linéaire d'architecture 20x5x2x5x20. Rappelons que la fonction d'acti-



FIG. 5.51 : L'évolution des critères AIC (figure (a)) et FPE (figure (b)) en fonction du nombre de neurones H.

vation de la première et la troisième couches cachées est non linéaire et est de type tangente hyperbolique, tandis que celle des neurones de la couche centrale et de la couche de sortie est linéaire. Les figures 5.52 et 5.53 montrent les évolutions de la moyenne et de l'écart type de l'erreur commise pendant les phases d'apprentissage et de test après chaque itération. Les figures 5.54 et 5.55 illustrent la projection sur le plan de l'ensemble d'apprentissage et de généralisation et les erreurs de reconstruction en sortie par bande.

On peut constater facilement que ces erreurs sont nettement moins importantes que celles fournies en sortie de l'autoassociateur linéaire.



FIG. 5.52 : La moyenne et l'écart type de l'erreur calculée sur l'ensemble d'apprentissage pour un autoassociateur non linéaire à 5 couches.



FIG. 5.53 : La moyenne et l'écart type de l'erreur calculée sur l'ensemble de test pour un autoassociateur non linéaire à 5 couches.



**FIG. 5.54 :** Plan de projection obtenu par les sorties de la couche centrale de l'autoassociateur non linéaire à 5 couches.



FIG. 5.55 : Erreurs de reconstruction par bande obtenues au niveau des sorties de l'autoassociateur non linéaire à 5 couches.

## 5.4 Suivi de fonctionnement de la génératrice par réseaux autoassociateurs

#### 5.4.1 Autoassociateur linéaire à trois couches

Nous utilisons un réseau autoassociateur linéaire d'architecture 34x2x34. Les résultats obtenus par une décomposition en valeurs singulières de la matrice des observations sont illustrés sur la figure 5.56. Le plan principal correspondant aux deux valeurs propres maximales explique environ 77.3% de la variance totale des observations.

Valeurs propres	Quantité d'informations expliquée	
$4.51 * 10^4$	70.4%	
$4.14 * 10^3$	6.9%	

FIG. 5.56 : Les 2 premières valeurs propres et le pourcentage de la variance totale des observations expliquée par chaque valeur propre.

Une fois les matrices des poids et de biais déterminées, nous nous intéressons aux sorties des deux neurones de la couche centrale définissant les axes de projection. Les figures 5.57 et 5.58 montrent les projections des bases d'apprentissage et de test ainsi que les erreurs de reconstruction par bande. On remarque que l'erreur au niveau de la bande numéro 25 est très faible, ce qui signifie que cette bande a été bien reconstruite. Les bandes numérotées 4, 9 et 27 présentent également des erreurs relativement faibles. Par contre, les bandes 7, 16 et 27 ont des erreurs proches de 1, ce qui révèle que le réseau fourni en sortie l'énergie moyenne de la bande.



FIG. 5.57 : Plan de projection obtenu par les sorties de la couche centrale de l'autoassociateur linéaire à 3 couches.



FIG. 5.58 : Erreurs de reconstruction par bande obtenues au niveau des sorties de l'autoassociateur linéaire à 3 couches.

#### 5.4.2 Autoassociateur non linéaire à trois couches

Nous considérons un réseau autoassociateur non linéaire à 3 couches d'architecture 34x2x34 dont la fonction d'activation des neurones cachés est de type tangente hyperbolique et celle des neurones de sorties est linéaire. Les observations des bases d'apprentissage et de test, constituées chacune de 432 observations, ont été normalisées entre -1 et +1. Après plusieurs essais, le pas d'adaptation de l'algorithme de rétropropagation est fixé à 0.01. Les poids du réseau sont fixés après 2000 itérations.

Ensuite, nous présentons séquentiellement les observations de la base d'apprentissage et de test. La figure 5.59 illustre la projection sur le plan défini par les sorties des neurones de la couche cachée pour les ensembles d'apprentissage et de test. On remarque que ces projections se superposent et se situent à l'intérieur du carré [-1 1]x[-1 1]. L'emploi d'un autoassociateur à 5 couches avec une couche de projection à fonction d'activation linéaire permettra d'écarter cette limitation.

Les erreurs de reconstruction par bandes sont représentées sur la figure 5.60. On peut remarquer que ces erreurs sont nettement inférieures par rapport à celles obtenues avec un autoassociateur linéaire.



FIG. 5.59 : Plan de projection obtenu par les sorties de la couche centrale de l'autoassociateur non linéaire à 3 couches.



**FIG. 5.60 :** Erreurs de reconstruction calculées sur l'ensemble d'apprentissage et de test calculées à la sortie d'un autoassociateur non linéaire à 3 couches.

#### 5.4.3 Autoassociateur non linéaire à cinq couches

Nous utilisons la même base d'apprentissage pour entraîner un réseau autoassociateur non linéaire d'architecture 34x5x2x5x34. En effet, l'étude du critère informationnel d'Akaike a montré que le nombre optimal des neurones dans la première et la troisième couches cachées est égal à 5. Nous rappelons que la fonction d'activation de ces deux dernières couches est non linéaire et de type tangente hyperbolique, tandis que celle des neurones de la couche centrale et de la couche de sortie est linéaire.

La figure 5.61 illustre la projection sur le plan des ensembles d'apprentissage et de test alors que la figure 5.62 montre les erreurs de reconstruction par bandes. On peut constater que les projections des observations de la base de test se situent dans le même domaine de fonctionnement normal défini par les projections des observations de la base d'apprentissage. Les erreurs de reconstruction sont nettement moins importantes que celles fournies en sortie de l'autoassociateur linéaire, mais elles sont par contre comparables à celles obtenues en sortie de l'autoassociateur non linéaire à 3 couches.

Pour mettre en évidence l'intérêt de notre approche de suivi de fonctionnement par réseau de neurones, nous allons étudier les réponses du réseau autoassociateur non linéaire à cinq couches lors de l'apparition d'une anomalie provoquant une dérive de fonctionnement du système éolien.



**FIG. 5.61 :** Plan de projection obtenue par les sorties de la couche centrale de l'autoassociateur non linéaire à 5 couches.



FIG. 5.62 : Erreurs de reconstruction par bande obtenues au niveau des sorties de l'autoassociateur non linéaire à 5 couches.

# 5.5 Simulation d'une dérive de fonctionnement du système éolien

Pour mettre en évidence l'intérêt de notre approche de suivi de fonctionnement, nous avons étudié les réponses des réseaux de neurones autoassociateurs linéaires et non linéaires lors de l'apparition d'une ou plusieurs anomalies provoquant ainsi une dérive de fonctionnement du système éolien.

Rappelons qu'il est impossible, pour des raisons de sécurité, de provoquer des défauts dans le cas d'une éolienne en état de production. Pour cela, nous nous sommes basés sur des études théoriques de certains défauts concernant des machines tournantes ainsi que leurs images vibratoires.

En général, la plupart des défauts affectant les machines tournantes se manifestent dans le spectre :

- soit, par la naissance dans la direction radiale ou axiale d'une composante d'amplitude prépondérante et de fréquence fondamentale correspondant à la fréquence de rotation de l'élément tournant mis en cause,
- soit, par le déplacement d'une fréquence existante, etc...

#### 5.5.1 Constitution de la base de test

La constitution d'une base de test simulant une dérive de fonctionnement d'un élément tournant de l'éolienne est effectuée de la manière suivante :

- On calcule le spectrogramme d'énergie des signaux de vibrations considérés sur plusieurs fenêtres consécutives de taille égale à 2<sup>15</sup> échantillons,
- 2. Chaque spectre est divisé en plusieurs bandes, 20 bandes dans le cas des signaux issus des accéléromètres installés sur le multiplicateur et 34 bandes dans

le cas du signal issu de l'accéléromètre installé sur la génératrice,

- 3. Pour simuler l'apparition d'une composante parasite dont la fréquence se situe dans une bande étroite ou large, on ajoute à la bande considérée un pic d'amplitude proportionnelle à l'amplitude du plus grand pic dans cette bande. De plus, la bande qui contient le premier harmonique de la fréquence perturbatrice voit son énergie augmenter.
- 4. Ensuite, on calcule l'énergie contenue dans chacune de ces bandes.

Nous considérons les notations suivantes :

- $f_E$  (resp.  $f_L$ ) est la fréquence fondamentale de la composante perturbatrice apparaissant dans une bande étroite (resp. large),
- $E_n$  est l'énergie contenue dans la bande numéro n,
- $A_n$  est l'amplitude maximale de la bande numéro n,
- $A_{per}$  est l'amplitude de la composante perturbatrice,
- $E_{per}$  est le pourcentage d'énergie apportée par la composante perturbatrice.

## 5.5.2 Apparition d'une composante parasite dont la fréquence se situe dans une bande étroite

Pour cela, nous avons perturbé la bande numéro 12 par la présence d'une composante à la fréquence 685Hz, d'amplitude proportionnelle à l'amplitude maximale de la bande.

Nous avons élaboré, selon les étapes décrites ci-dessus, trois bases de test qui correspondent respectivement à la simulation d'une faible, moyenne et forte perturbation. Les 3 tableaux des 3 figures 5.63, 5.64 et 5.65 résument les amplitudes et les pourcentages d'énergie apportée par la composante perturbatrice et son premier harmonique.

	$f_E = 685 \text{Hz}$	$2*f_E = 1370 \mathrm{Hz}$
$A_{per}$	$0.5 * A_{12}$	$1 * A_{17}$
$E_{per}$	$12\% * E_{12}$	$11\% * E_{17}$

FIG. 5.63 : Base 1 : Pourcentage d'énergie apportée par la composante perturbatrice et son premier harmonique.

	$f_E = 685 \text{Hz}$	$2*f_E = 1370 \mathrm{Hz}$
Aper	$1 * A_{12}$	$2 * A_{17}$
Eper	$44\% * E_{12}$	$37\% * E_{17}$

FIG. 5.64 : Base 2 : Pourcentage d'énergie apportée par la composante perturbatrice et son premier harmonique.

	$f_E = 685 \text{Hz}$	$2*f_E = 1370 \mathrm{Hz}$
$A_{per}$	$2 * A_{12}$	$4 * A_{17}$
$E_{per}$	$175\% * E_{12}$	$150\% * E_{17}$

FIG. 5.65 : Base 3 : Pourcentage d'énergie apportée par la composante perturbatrice et son premier harmonique.

Les figures 5.66, 5.68 et 5.70 montrent la projection de la base d'apprentissage et des trois bases de test sur le plan défini par les sorties des deux neurones de la couche centrale du réseau. Les figures 5.67, 5.69 et 5.71 montrent les erreurs de reconstruction de la base d'apprentissage et des trois bases de test obtenues à la sortie d'un réseau autoassociateur non linéaire à 5 couches.

Lorsqu'on analyse le plan de projection, on remarque que l'ensemble de test dans le cas d'une faible perturbation ne permet pas d'indiquer la présence d'une anomalie ou une dérive de fonctionnement. En revanche, au niveau des erreurs de reconstruction, on peut remarquer déjà un écart important entre l'erreur d'apprentissage et de test au niveau de la bande perturbée numéro 12.

Dans le cas d'une moyenne ou forte perturbation, une dérive de fonctionnement est visible aussi bien au niveau de la projection que lors de la reconstruction.



FIG. 5.66 : Base 1 : Plan de projection obtenue par les sorties de la couche centrale de l'autoassociateur non linéaire à 5 couches.



FIG. 5.67 : Base 1 : Erreurs de reconstruction par bande obtenues au niveau des sorties de l'autoassociateur non linéaire à 5 couches.



FIG. 5.68 : Base 2 : Plan de projection obtenue par les sorties de la couche centrale de l'autoassociateur non linéaire à 5 couches.



FIG. 5.69 : Base 2 : Erreurs de reconstruction par bande obtenues au niveau des sorties de l'autoassociateur non linéaire à 5 couches.



FIG. 5.70 : Base 3 : Plan de projection obtenue par les sorties de la couche centrale de l'autoassociateur non linéaire à 5 couches.



FIG. 5.71 : Base 3 : Erreurs de reconstruction par bande obtenues au niveau des sorties de l'autoassociateur non linéaire à 5 couches.

### 5.5.3 Apparition d'une composante parasite dont la fréquence se situe dans une bande large

Pour générer artificiellement un défaut dans une bande large, nous avons perturbé la bande numéro 13 par la présence à la fréquence 750Hz d'un pic d'amplitude proportionnelle à l'amplitude principale de la bande. De même, nous avons aussi généré trois bases de test qui correspondent respectivement à la simulation d'une faible, moyenne et forte perturbation. Les amplitudes et les pourcentages d'énergie apportée par la composante perturbatrice et son premier harmonique sont résumés dans les 3 tableaux des 3 figures ci-dessous 5.72, 5.73 et 5.74.

	$f_L = 750 \text{Hz}$	$2*f_L = 1500 \mathrm{Hz}$
$A_{per}$	$0.5 * A_{13}$	$1 * A_{17}$
$E_{per}$	$9\% * E_{13}$	$1\% * E_{17}$

FIG. 5.72 : Base 1 : Pourcentage d'énergie apportée par la composante perturbatrice et son premier harmonique.

	$f_L = 750 \mathrm{Hz}$	$2*f_L = 1500 \mathrm{Hz}$
$A_{per}$	$1 * A_{13}$	$2 * A_{17}$
$E_{per}$	$48\% * E_{13}$	$5\% * E_{17}$

FIG. 5.73 : Base 2 : Pourcentage d'énergie apportée par la composante perturbatrice et son premier harmonique.

	$f_L = 750 \mathrm{Hz}$	$2 * f_L = 1500 \mathrm{Hz}$
$A_{per}$	$2 * A_{13}$	$4 * A_{17}$
$E_{per}$	$200\% * E_{13}$	$20\% * E_{17}$

FIG. 5.74 : Base 3 : Pourcentage d'énergie apportée par la composante perturbatrice et son premier harmonique.
Les figures 5.75, 5.77 et 5.79 montrent la projection de la base d'apprentissage et des trois bases de test sur le plan défini par les sorties des deux neurones de la couche centrale du réseau. Les figures 5.76, 5.78 et 5.80 montrent les erreurs de reconstruction de la base d'apprentissage et des trois bases de test obtenues à la sortie d'un réseau autoassociateur non linéaire à 5 couches. Les mêmes conclusions sont obtenues : Les erreurs de reconstruction sont plus sensibles aux dérives de fonctionnement que la projection plane.



FIG. 5.75 : Base 1 : Plan de projection obtenue par les sorties de la couche centrale de l'autoassociateur non linéaire à 5 couches.



FIG. 5.76 : Base 1 : Erreurs de reconstruction par bande obtenues au niveau des sorties de l'autoassociateur non linéaire à 5 couches.



**FIG. 5.77 : Base 2 :** Plan de projection obtenue par les sorties de la couche centrale de l'autoassociateur non linéaire à 5 couches.



FIG. 5.78 : Base 2 : Erreurs de reconstruction par bande obtenues au niveau des sorties de l'autoassociateur non linéaire à 5 couches.



FIG. 5.79 : Base 3 : Plan de projection obtenue par les sorties de la couche centrale de l'autoassociateur non linéaire à 5 couches.



FIG. 5.80 : Base 3 : Erreurs de reconstruction par bande obtenues au niveau des sorties de l'autoassociateur non linéaire à 5 couches.

### 5.6 Conclusion

Dans un premier temps, nous avons utilisé un autoassociateur à 3 couches avec des neurones cachés linéaires, cela revient à une ACP (Analyse en composante principale). L'inconvénient de cette approche est que seules les bandes de fréquences comprenant les deux plus grands pics du spectre sont bien reconstruites. Pour les autres bandes, le réseau fournit en sortie l'énergie moyenne et non l'énergie réelle contenue dans chacune des bandes.

L'emploi d'une architecture neuronale à trois couches avec des neurones cachés non linéaires réduit nettement les erreurs de reconstruction pour l'ensemble des bandes de fréquences. Il faut noter que les erreurs de reconstruction des bandes étroites sont inférieures à celles des bandes larges. Néanmoins, la non linéarité des neurones cachés réduit le champs de projection à un domaine fermé [-1 +1]x[-1 +1]. Pour éviter ce problème, nous avons utilisé un autoassociateur à cinq couches avec des neurones linéaires dans la couche centrale et non linéaires dans la première et la troisième couche cachée.

L'architecture de ce réseau a été optimisée grâce aux critères informationnels d'Akaike (AIC et FPE). Notons que les erreurs de reconstruction avec un autoassociateur à trois ou cinq couches sont comparables. Les résultats obtenus montrent que l'autoassociateur non linéaire à cinq couches est le mieux adapté pour le suivi de fonctionnement des éléments sensibles d'une éolienne.

La simulation d'une dérive de fonctionnement de l'éolienne a été réalisée par l'ajout artificiel d'une composante perturbatrice dans une bande étroite ou large. Dans les deux cas, il est possible de percevoir le défaut par projection et par les erreurs de reconstruction. Toutefois, l'utilisation de la projection reste moins sensible que la reconstruction.

# Chapitre 6

# Réseaux de neurones modulaires appliqués à la surveillance d'une éolienne

### 6.1 Introduction

Les réseaux de neurones artificielles de type Perceptrons multicouches sont largement utilisés pour résoudre des tâches complexes dans de nombreux domaines des sciences de l'ingénieur. Dans ce type d'architecture, les neurones sont organisés en couches : une couche d'entrée, une ou plusieurs couches cachées et une couche de sortie. Les neurones d'une couche sont reliés aux neurones de la couche suivante par des connexions pondérées. Grâce à la faculté d'apprentissage par l'exemple, le réseau multicouche adapte ses poids pour réaliser tout type de fonctions d'un espace à N dimensions dans un espace à K dimensions. La règle d'apprentissage la plus utilisée est celle de la rétropropagation du gradient de l'erreur.

Récemment, Jacobs a mis en évidence les défauts d'un tel réseau à architecture multicouches utilisé souvent comme une boîte noire et a proposé des réseaux modulaires d'architectures plus flexibles pouvant intégrer des connaissances a priori dans leurs conceptions [JJ94a]. L'idée de base est de résoudre une tâche de classification complexe en la décomposant en sous-tâches plus simples facilement réalisables par des modules appropriés dits experts qui, grâce à un module de contrôle, coopèrent pour aboutir à la décision finale. Ce type de réseau modulaire associe à la fois des notions statistiques et biologiques et est de plus en plus utilisé dans différents domaines tels que la commande d'un bras d'un robot [JJNH91], le traitement du signal [Hay96], etc.

Plus récemment, des réseaux modulaires et hiérarchiques à apprentissage par l'algorithme statistique EM (Estimation Maximisation) ont été développés par Jordan et al. [JJ94b]. Le lecteur intéressé par une étude plus complète des réseaux modulaires et hiérarchiques peut se référer aux travaux [JJ94b] [WR94].

Dans ce chapitre, nous décrivons un réseau modulaire à apprentissage par la méthode du maximum de vraisemblance pour la classification supervisée et en particulier à la surveillance du fonctionnement des éléments sensibles d'une éolienne.

Les sections 2 et 3 décrivent l'architecture du réseau modulaire et expliquent comment lors de la phase d'apprentissage, les modules experts apprennent les soustâches et combinent les réponses des experts pour aboutir à la décision finale. Dans la section 4, on montre comment déterminer l'architecture optimale pour le réseau modulaire. La section 5 illustre sur deux exemples générés artificiellement, les performances du réseau modulaire alors que la section 6 montre l'application du réseau dans un problème de surveillance du fonctionnement d'une éolienne. Nous considérons que l'éolienne admet deux modes de fonctionnements ou classes : normal ou défaillant. D'une manière générale, le problème de diagnostic avec M modes de fonctionnements peut être considéré comme M(M-1)/2 problèmes de discriminations à deux classes.

### 6.2 Architecture et équations de fonctionnement

Dans ce paragraphe nous présentons l'architecture et les équations de fonctionnement d'un réseau modulaire pour la discrimination entre deux classes  $C_1$  et  $C_2$ .

#### 6.2.1 Architecture

Considérons la configuration d'un réseau modulaire représenté sur la figure 6.1. Sa structure contient trois blocs :

- Le premier bloc est constitué de K modules nommés experts à apprentissages supervisés. Chacun de ces experts est constitué d'un simple neurone à fonction d'activation sigmoïdale et admet N entrées qui reçoivent les N composantes du vecteur observation.
- Le second bloc est composé d'un module, appelé contrôleur ou superviseur, dont l'apprentissage, contrairement aux experts, est non supervisé. Le module contrôleur admet N entrées qui reçoivent les composantes du vecteur observation et K neurones de sortie dont les fonctions d'activation sont de type sigmoïde généralisée ou "Softmax".
- Le troisième bloc est constitué d'un module appelé module de sortie. Il est constitué d'un seul neurone ayant K entrées qui reçoivent les valeurs des sorties des experts. Les poids des connexions de ce neurone sont les K sorties du contrôleur. Le neurone fournit en sortie le produit scalaire entre le vecteur de sortie du contrôleur et son vecteur d'entrée qui est en fait le vecteur de sortie des experts.

La figure 6.2 montre l'architecture d'un module expert constitué d'un simple neurone. Le module contrôleur comporte K neurones, K est aussi le nombre de modules experts utilisés (cf. figures 6.3 et 6.4). Les fonctions d'activation g des neurones sont de type "Softmax".



FIG. 6.1 : Schéma bloc du réseau modulaire. Les sorties des modules experts sont pondérées par les sorties du module contrôleur.

#### 6.2.2 Equations de fonctionnement

Considérons les notations suivantes :

- $\boldsymbol{x} = (x_1, x_2, \dots, x_N)^T$  est le vecteur observation présenté à l'entrée du réseau modulaire, il est appliqué simultanément aux entrées de tous les experts et à celles du contrôleur,
- $\boldsymbol{w}_k = (w_{k1}, w_{k2}, \ldots, w_{kN})$  est le vecteur de poids du  $k^{\check{e}me}$  expert,
- $\boldsymbol{W} = (\boldsymbol{w}_1, \ \boldsymbol{w}_2, \ \ldots, \ \boldsymbol{w}_K)$  est la matrice de poids de l'ensemble des K experts,
- $v_k$  est l'entrée totale du  $k^{\grave{e}me}$  expert,
- $y_k$  est la sortie du  $k^{ime}$  expert,



 ${\bf FIG.}~{\bf 6.2}$  : Architecture et notations des variables pour un module expert.



FIG. 6.3 : Architecture du module contrôleur.



FIG. 6.4: Notations des variables pour un neurone du module contrôleur.

- $u_k$  est l'entrée totale du  $k^{ime}$  neurone du contrôleur;  $\boldsymbol{u} = (u_1, u_2, \ldots, u_K)^T$ ,
- $\pi_k$  est la sortie du  $k^{ime}$  neurone du contrôleur;  $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_K)^T$ ,
- $a_k$  est le vecteur poids des connexions entrées-sortie du  $k^{\grave{e}me}$  neurone du contrôleur,
- $\boldsymbol{A} = (\boldsymbol{a}_1, \ \boldsymbol{a}_2, \ \ldots, \ \boldsymbol{a}_K)$  est la matrice de poids du contrôleur,
- d est la sortie désirée du réseau modulaire,
- y est la sortie estimée par le réseau modulaire.

#### Equations de fonctionnement des experts :

L'entrée totale du  $k^{ime}$  expert est donnée par :

$$v_k = \boldsymbol{w}_k^T \boldsymbol{x}, \qquad k = 1, 2, \dots, K$$
 (6.1)

Pour une observation  $\boldsymbol{x}$  présentée à l'entrée du réseau, chacun des experts fournit en sortie :

$$y_k = f(v_k), \qquad k = 1, 2, \dots, K$$
, (6.2)

où  $f(\cdot)$  est une fonction d'activation de type sigmoïde définie par :

$$f(v_k) = \frac{1}{1 + \exp(-v_k)} .$$
 (6.3)

#### Equations de fonctionnement du contrôleur :

L'entrée totale du  $k^{ime}$  neurone du contrôleur est la suivante :

$$u_k = \boldsymbol{a}_k^T \boldsymbol{x}, \qquad k = 1, 2, \dots, K.$$
(6.4)

La sortie du  $k^{\grave{e}me}$  neurone du contrôleur est :

$$\pi_k = g(u_k), \qquad k = 1, 2, \dots, K.$$
 (6.5)

où  $g(\cdot)$  est une fonction d'activation de type non linéaire dite *softmax* [Bri89], définie comme une généralisation multi-entrées de la fonction sigmoïde :

$$\pi_k = g(u_k) = \frac{\exp(u_k)}{\sum_{j=1}^K \exp(u_j)} .$$
(6.6)

#### Equation de la sortie du réseau modulaire :

La sortie estimée par le réseau modulaire est donnée par :

$$y = \sum_{k=1}^{K} \pi_k y_k \ . \tag{6.7}$$

### 6.3 Algorithme d'apprentissage

Le réseau modulaire est utilisé pour discriminer entre deux classes  $C_1$  et  $C_2$ . Les observations de la classe  $C_1$  sont codées par d = 1 et ceux de la classe  $C_2$  sont codées par d = 0. Or un vecteur observation  $\boldsymbol{x}$  appartient soit à la classe  $C_1$  soit à la classe  $C_2$ . La sortie  $y_k$  d'un expert est comprise entre 0 et 1 et elle mesure en quelque sorte le degré d'appartenance de  $\boldsymbol{x}$  ou plus précisément la probabilité d'appartenance de  $\boldsymbol{x}$  [Bis95]. Si on suppose que  $y_k$  est la probabilité d'appartenance de  $\boldsymbol{x}$  à la classe  $C_1$ ,  $1-y_k$  sera alors la probabilité d'appartenance de  $\boldsymbol{x}$  à la classe  $C_2$ .

La probabilité que le vecteur observation  $\boldsymbol{x}$  appartient à la classe  $C_1$  (d=1) est égale à la sortie du  $k^{\grave{e}me}$  expert :

$$P(d = 1 | \boldsymbol{x}, k) = P(C_1 | \boldsymbol{x}, k) = y_k .$$
(6.8)

La probabilité que le vecteur observation  $\boldsymbol{x}$  appartient à la classe  $C_2$  est :

$$P(d = 0 | \boldsymbol{x}, k) = P(C_2 | \boldsymbol{x}, k) = 1 - y_k .$$
(6.9)

En combinant ces deux équations en une seule, on peut définir la distribution de probabilité conditionnelle de la sortie désirée connaissant le vecteur observation  $\boldsymbol{x}$ , et sachant que le  $k^{\grave{e}me}$  expert est sélectionné comme suit :

$$\mathcal{P}\left(d \mid \boldsymbol{x}, k\right) = y_{k}^{d} \left(1 - y_{k}\right)^{(1-d)} \quad . \tag{6.10}$$

Notons que  $\mathcal{P}(d \mid \boldsymbol{x}, k)$  est la distribution de probabilité de Bernoulli qui est un cas particulier de la distribution de probabilité binomiale.

D'après l'équation (6.7), la distribution de probabilité conditionnelle de la sortie désirée du réseau modulaire est un mélange, c'est à dire une combinaison linéaire de K différentes probabilités conditionnelles correspondant aux K différents experts. Nous avons alors :

$$\mathcal{P}(d \mid \boldsymbol{x}) = \sum_{k=1}^{K} \pi_k \mathcal{P}(d \mid \boldsymbol{x}, k) \quad .$$
(6.11)

où  $\pi_k = P(k \mid \boldsymbol{x})$  peut être vue comme une probabilité a priori que le  $k^{eme}$  expert est sélectionné connaissant le vecteur d'entrée  $\boldsymbol{x}$ . Par conséquent, les probabilités a priori  $\pi_k, k = 1, 2, \ldots, K$  sont contraintes à satisfaire les deux conditions nécessaires :

$$\begin{cases} 0 \le \pi_k \le 1 , \quad k = 1, 2, \dots, K , \\ \text{et} \\ \sum_{k=1}^{K} \pi_k = 1 . \end{cases}$$
(6.12)

En substituant l'équation (6.10) dans l'équation (6.11), la distribution de probabilité conditionnelle peut être réécrite sous la forme :

$$\mathcal{P}(d \mid \boldsymbol{x}) = \sum_{k=1}^{K} \pi_k \left( y_k^d \left( 1 - y_k \right)^{(1-d)} \right) .$$
(6.13)

Le but de l'apprentissage est d'entraîner le réseau modulaire de la figure 6.1 afin de modéliser la distribution de probabilité de l'ensemble de la base d'apprentissage caractérisé par les couples  $(\boldsymbol{x}, d)$ . Pour ce faire, la distribution de probabilité conditionnelle  $\mathcal{P}(d \mid \boldsymbol{x})$  peut être vue comme une fonction de vraisemblance, où l'ensemble des poids synaptiques des experts  $\boldsymbol{W}$  et du contrôleur  $\boldsymbol{A}$  représentent les paramètres libres du modèle probabiliste.

Nous pouvons ainsi définir la fonction de log-vraisemblance comme suit :

$$\mathcal{L} = \ln \left( \mathcal{P}(d \mid \boldsymbol{x}) \right) = \ln \left( \sum_{k=1}^{K} \pi_k \mathcal{P}\left( d \mid \boldsymbol{x}, k \right) \right) .$$
(6.14)

En substituant l'équation (6.13) dans l'équation (6.14), on obtient :

$$\mathcal{L} = \ln\left(\sum_{k=1}^{K} \pi_k y_k^d \left(1 - y_k\right)^{(1-d)}\right) \ . \tag{6.15}$$

Il est généralement plus commode de minimiser l'opposée de cette fonction, soit :

$$J = -\ln\left(\sum_{k=1}^{K} \pi_k y_k^d \left(1 - y_k\right)^{(1-d)}\right)$$
(6.16)

où J est le critère d'erreur à minimiser.

Pour mettre en évidence la dépendance du critère par rapport aux poids synaptiques, nous remplaçons les équations (6.1), (6.2), (6.4) et (6.5) dans l'équation (6.16). Le critère d'erreur peut être réécrit sous la forme :

$$J(\boldsymbol{W}, \boldsymbol{A}) = \ln \left( \sum_{k=1}^{K} g\left( \boldsymbol{a}_{k}^{T} \boldsymbol{x} \right) f\left( \boldsymbol{w}_{k}^{T} \boldsymbol{x} \right)^{d} \left( 1 - f\left( \boldsymbol{w}_{k}^{T} \boldsymbol{x} \right) \right)^{(1-d)} \right)$$
(6.17)

Dans la suite nous calculons la sensibilité du critère J par rapport aux différents poids synaptiques du réseau modulaire.

#### 6.3.1 Adaptation des poids des modules experts

La sensibilité du critère d'erreur J, équation (6.17), par rapport au vecteur poids synaptiques du  $k^{\grave{e}me}$  expert  $w_k$  est définie par le gradient :

$$\frac{\partial J}{\partial \boldsymbol{w}_k} = \frac{\partial J}{\partial y_k} \frac{\partial y_k}{\partial v_k} \frac{\partial v_k}{\partial \boldsymbol{w}_k} , \qquad k = 1, 2, \dots, K .$$
 (6.18)

D'après les équations (6.1) et (6.2), la dérivée partielle de la sortie  $y_k$  par rapport au vecteur poids  $v_k$  donne :

$$\frac{\partial y_k}{\partial v_k} = y_k \left(1 - y_k\right) \ . \tag{6.19}$$

et la dérivée partielle de  $v_k$  par rapport à  $\boldsymbol{w}_k$  est :

$$\frac{\partial v_k}{\partial \boldsymbol{w}_k} = x \ . \tag{6.20}$$

D'après l'équation (6.16), en calculant la dérivée partielle de J par rapport à la sortie  $y_k$ , nous obtenons après simplification :

$$\frac{\partial J}{\partial y_k} = -\frac{(d-y_k)}{y_k (1-y_k)} \frac{\pi_k y_k^d (1-y_k)^{(1-d)}}{\sum_{j=1}^K \pi_j y_j^d (1-y_j)^{(1-d)}} , \qquad k = 1, 2, \dots, K .$$
(6.21)

On peut définir la probabilité a posteriori notée  $h_k$  associée à la sortie du  $k^{eme}$ expert :

$$h_{k} = \frac{\pi_{k} y_{k}^{d} \left(1 - y_{k}\right)^{(1-d)}}{\sum_{j=1}^{K} \pi_{j} y_{j}^{d} \left(1 - y_{j}\right)^{(1-d)}} , \qquad k = 1, 2, \dots, K.$$
(6.22)

qui doit satisfaire les deux conditions nécessaires suivantes :

$$\begin{cases} 0 \le h_k \le 1 , \quad k = 1, 2, \dots, K , \\ \text{et} \\ \sum_{k=1}^{K} h_k = 1 . \end{cases}$$
(6.23)

En substituant l'équation (6.22) dans l'équation (6.21), l'expression de  $\left(\frac{\partial J}{\partial y_k}\right)$  s'écrit :

$$\frac{\partial J}{\partial y_k} = -\frac{(d-y_k)}{y_k \left(1-y_k\right)} h_k , \qquad k = 1, 2, \dots, K.$$
(6.24)

Soit  $e_k$  l'erreur commise à la sortie de  $k^{\grave{e}me}$  expert donnée par :

$$e_k = d - y_k . ag{6.25}$$

L'équation (6.24) peut s'écrire en tenant compte de l'équation (6.25) sous la forme :

$$\frac{\partial J}{\partial y_k} = -\frac{h_k}{y_k \left(1 - y_k\right)} e_k , \qquad k = 1, 2, \dots, K.$$
(6.26)

En substituant les équations (6.19), (6.20) et (6.26) dans l'équation (6.18), nous obtenons :

$$\frac{\partial J}{\partial \boldsymbol{w}_k} = -h_k e_k \boldsymbol{x} , \qquad k = 1, 2, \dots, K.$$
(6.27)

Le vecteur poids  $w_k$  à l'itération (t+1) est adapté selon la règle du gradient suivante :

$$\boldsymbol{w}_k(t+1) = \boldsymbol{w}_k(t) - \eta \left(\frac{\partial J}{\partial \boldsymbol{w}_k}\right), \qquad k = 1, 2, \dots, K.$$
 (6.28)

où  $\eta$  est le pas d'adaptation.

En substituant l'équation (6.27) dans l'équation (6.28), on obtient la règle d'adaptation des poids synaptiques des différents modules experts :

$$\boldsymbol{w}_{k}(t+1) = \boldsymbol{w}_{k}(t) + \eta h_{k} e_{k} \boldsymbol{x} , \qquad k = 1, 2, \dots, K$$

$$(6.29)$$

Cette équation montre que, pendant le processus d'apprentissage, à chaque présentation d'un vecteur observation  $\boldsymbol{x}$ , les poids synaptiques du  $k^{\grave{e}me}$  expert sont ajustés afin de réduire l'erreur  $e_k$ , entre la sortie  $y_k$  et la valeur désirée d. La probabilité a posteriori  $h_k$  joue le rôle de terme de pondération.

#### 6.3.2 Adaptation des poids du module contrôleur

La valeur de sortie de chaque neurone contrôleur pondère la sortie d'un expert et la somme fournit la sortie du module de sortie qui est la sortie finale du réseau modulaire. Autrement dit, le contrôleur a pour rôle d'attribuer à chaque expert une probabilité a priori d'être sélectionné connaissant le vecteur observation présenté à l'entrée du réseau modulaire.

En substituant l'équation (6.6) dans l'équation (6.16), le critère d'erreur à minimiser peut être réécrit sous la forme suivante :

$$J = \ln\left(\sum_{k=1}^{K} \exp(u_k) y_k^d \left(1 - y_k\right)^{(1-d)}\right) + \ln\left(\sum_{j=1}^{K} \exp(u_j)\right) .$$
 (6.30)

La sensibilité du critère d'erreur J par rapport au vecteur poids synaptiques  $a_k$ est défini par :

$$\frac{\partial J}{\partial \boldsymbol{a}_k} = \frac{\partial J}{\partial u_k} \frac{\partial u_k}{\boldsymbol{a}_k} , \qquad k = 1, 2, \dots, K.$$
(6.31)

La dérivée partielle de J donnée par (6.31) par rapport à l'entrée totale  $u_k$  du  $k^{\grave{e}me}$  neurone du module contrôleur est donnée après simplification par :

$$\frac{\partial J}{\partial u_k} = \pi_k - h_k , \qquad k = 1, 2, \dots, K.$$
(6.32)

D'après l'équation (6.4), la dérivée partielle de la somme pondérée  $u_k$  par rapport au vecteur  $a_k$  est :

$$\frac{\partial u_k}{\partial \boldsymbol{a}_k} = \boldsymbol{x} , \qquad k = 1, 2, \dots, K.$$
 (6.33)

En substituant les équations (6.32) et (6.33) dans (6.31), nous obtenons :

$$\frac{\partial J}{\partial \boldsymbol{a}_k} = (\pi_k - h_k) \boldsymbol{x} , \qquad k = 1, 2, \dots, K.$$
(6.34)

L'équation d'adaptation des poids synaptiques pour le module contrôleur est :

$$\boldsymbol{a}_{k}(t+1) = \boldsymbol{a}_{k}(t) - \eta \left( \pi_{k}(t) - h_{k}(t) \right) \boldsymbol{x} , \qquad k = 1, 2, \dots, K$$
(6.35)

On peut remarquer que, l'adaptation des poids synaptiques du réseau contrôleur dépend de la différence entre la probabilité a priori et celle a posteriori. La fin de l'apprentissage a lieu donc lorsque les deux probabilités sont égales.

# 6.4 Optimisation de l'architecture du réseau modulaire

Nous proposons d'optimiser le nombre des experts dans le réseau modulaire. En fait, pour un problème de classification donné, un nombre élevé d'experts permet de créer des surfaces de décision artificiellement complexes conduisant à une dégradation des performances du réseau au niveau de la généralisation. Par contre, un nombre réduit d'experts risque de fournir une décision très simple ne représentant pas suffisamment la vraie nature du problème. Le nombre optimum peut être déterminé grâce à l'emploi du critère AIC d'Akaike défini par [Aka72] [KA97] :

$$AIC(K) = -\ln(\mathcal{L}) + 2\left(\frac{N_{\boldsymbol{W}} + N_{\boldsymbol{A}}}{N * P}\right), \qquad (6.36)$$

où :

- $\ln(\cdot)$  est le logarithme naturel,
- L'représente la fonction log-vraisemblance,
- N et P représentent respectivement la dimension et la taille de l'ensemble d'apprentissage,
- $N_{W}$  et  $N_{A}$  représentent respectivement le nombre des poids des experts et du contrôleur tels que :  $N_{W} = K * N$  et  $N_{A} = K * N$ .

Le principe consiste à faire varier le nombre K des experts du réseau modulaire et à calculer les valeurs du critère AIC. Le nombre K minimisant cette fonction est celui qui sera retenu.

Nous avons présenté l'architecture et la stratégie d'apprentissage d'un réseau modulaire. Pour des raisons pédagogiques, nous avons considéré des experts identiques d'architectures très simples réduites à de neurones formels. D'une manière générale, les réseaux modulaires peuvent faire intégrer des experts multicouches et d'architectures différentes. La stratégie d'apprentissage a été la règle de rétropropagation du gradient de l'erreur pour minimiser l'opposé de la fonction logvraisemblance. Cette règle assez simple nous a semblé satisfaisante. D'autres règles plus sophistiquées pourraient être utilisées [WR94].

Dans la suite, nous analysons tout d'abord le fonctionnement du réseau modulaire sur deux exemples classiques de données générées artificiellement. Ensuite, nous appliquons le réseau à la surveillance d'un système aérogénérateur et testons ses performances sur le système en situation de fonctionnement normal et défaillant.

### 6.5 Exemples d'application

Nous allons présenter tout d'abord deux exemples bien connu en classification générés artificiellement. Le premier traite le cas de deux classes en forme de croissant et le second est celui du ou-exclusif. Ces exemples ont été choisis de façon à illustrer l'aptitude du réseau modulaire à classer des exemples non linéairement séparables. Ensuite, nous appliquons cette architecture modulaire à la surveillance du multiplicateur d'une éolienne. Le problème consiste à discriminer entre le fonctionnement normal et défaillant du multiplicateur dans le but de décider si son point de fonctionnement actuel se trouve dans une zone de fonctionnement normal ou dans une zone de défaillance.

#### 6.5.1 Exemple 1 : Deux classes en forme de croissant.

Nous avons utilisé une base de données constituée de 300 observations représentant la base d'apprentissage (150 observations) et la base de test (150 observations) réparties en deux classes, notées  $C_1$  et  $C_2$ , en forme de deux croissants dans un espace à deux dimensions. Chacune des classes de la base d'apprentissage est donc constituée de 75 observations.

Les équations des croissants sont définies par :

$$\begin{cases} x_1 = \alpha_1 \cos(\theta) \cos(\phi) + \beta_1, \\ x_2 = \alpha_2 \sin(\theta) \cos(\phi) + \beta_2, \end{cases}$$
(6.37)

où

- $\theta$  est une variable aléatoire normale de moyenne  $\mu_{\theta}$  et d'écart-type  $\sigma_{\theta}$ ,
- $\beta_1$  et  $\beta_2$  sont aussi des variables normales de moyennes  $\mu_{\beta_1}$ ,  $\mu_{\beta_2}$  et d'écartstypes  $\sigma_{\beta_1}$ ,  $\sigma_{\beta_2}$  respectivement.
- $\phi$ ,  $\alpha_1$  et  $\alpha_2$  sont constants.

Les valeurs prises par ces paramètres sont consignées dans le tableau de la figure 6.5. Les observations de la classe  $C_1$  sont codées par d = 1 et celles de la classe  $C_2$  sont codées par d = 0. La figure 6.6 illustre la répartition dans le plan de l'ensemble des observations des bases d'apprentissage et de test.

	Classe $C_1$	Classe $C_2$			
Nombre de points	150	150			
$\phi$	0	0			
$(\alpha_1 \ \alpha_2)$	(55)	(55)			
$(\mu_{ heta} \sigma_{ heta})$	(08)	(180 40)			
$(\mu_{\beta_1} \sigma_{\beta_1})$	(01)	(-1  0.5)			
$(\mu_{eta_2} \sigma_{eta_2})$	(01)	$(-5 \ 0.5)$			

FIG. 6.5 : Caractéristiques statistiques des deux croissants représentant les deux classes  $C_1$  et  $C_2$  de la base de données.



**FIG. 6.6** : Représentations graphiques de deux classes  $C_1$  et  $C_2$  en forme de croissant représentant l'ensemble des observations de la base d'apprentissage et de la base de test.

Nous avons utilisé un réseau modulaire comprenant plusieurs experts. Chaque expert est constitué d'un simple Perceptron.

Lors de l'apprentissage, les observations ont été tirées aléatoirement sans remise de la base et présentées au réseau. La présentation complète de toute la base est appelée itération. Le nombre total des itérations a été fixé à 500. Le pas d'adaptation  $\eta$  est un paramètre très délicat à fixer. Après plusieurs essais, nous avons choisi  $\eta = 0.01$ .

A la fin de l'apprentissage, les poids du réseau sont fixés. Pour évaluer les performances du réseau en apprentissage et surtout en test, nous présentons séquentiellement les observations de chacune des bases à l'entrée du réseau qui fournit en sortie des valeurs comprises entre 0 et 1. Nous considérons qu'une observation appartient à la classe  $C_1$  si la sortie est supérieure à 0.5 sinon elle appartient à la classe  $C_2$ .

Pour optimiser l'architecture du réseau, nous présentons à nouveau séquentiellement les observations de la base d'apprentissage au réseau. Nous faisons varier le nombre des experts de 2 à 5. Pour chaque nombre nous avons entraîné un réseau modulaire et calculé le critère AIC(k), k étant le nombre des experts. Les résultats résumés dans le tableau de la figure 6.7 montrent que la valeur minimale du critère AIC est obtenue en utilisant trois experts. On note dans la suite :

- $\bar{E}$  et  $\sigma_E$  la moyenne et l'écart-type de l'erreur, et
- $N_{obs}$  le nombre des observations mal classées.

Nombre des Experts	2	3	4	5
Critère AIC	3.99	2.61	2.64	2.67
$ar{E}$ (Apprentissage %)	2.35	1.52	1.56	1.58
$ar{E}$ (Test %)	2.47	1.69	1.72	1.71
$\sigma_E$ (Apprentissage %)	6.22	4.09	4.10	4.25
$\sigma_E$ (Test %)	6.46	4.89	4.87	4.84
$N_{obs}$ (Apprentissage)	0	0	0	0
$N_{obs}$ (Test)	1	0	0	0

**FIG. 6.7 :** Résultats de la discrimination entre deux croissants obtenus en utilisant un réseau modulaire.

Les figures 6.8 et 6.9 montrent l'allure de la moyenne de l'erreur et de son écarttype calculés lors de la phase d'apprentissage et de test en utilisant un réseau modulaire à trois experts. A partir de l'itération 250, la moyenne de l'erreur décroît très lentement pour atteindre la valeur 1.5% (cf. le tableau de la figure 6.7). On peut constater que l'écart-type croît au début de l'apprentissage puis décroît pour se stabiliser à la valeur 4.1%.

Une fois l'apprentissage terminé, on présente les deux bases et on calcule les sorties estimées par le réseau modulaire pour chaque observation. Les figures 6.10 et 6.11 donnent, pour chaque observation présentée au réseau, la sortie estimée et sa valeur désirée correspondante. Dans le cas des observations des deux bases, pour la classe  $C_1$ , aucun pic ne dépasse la frontière de décision qui est fixée à 0.5. Donc aucune observation de la classe  $C_1$  n'a été affectée à la classe  $C_2$ . De même, aucune observation de la classe  $C_2$  n'a été affectée à la classe  $C_1$ . Par conséquent, l'emploi d'une architecture modulaire à trois experts a permis de séparer les deux classes  $C_1$ et  $C_2$  sans aucune erreur de classification.



FIG. 6.8 : Allure de la moyenne de l'erreur lors de la phase d'apprentissage et de test obtenue en utilisant un réseau modulaire à trois experts.



**FIG. 6.9 :** Allure de l'écart type lors de la phase d'apprentissage et de test obtenue en utilisant un réseau modulaire à trois experts.



FIG. 6.10 : Sorties estimées par le réseau modulaire à trois experts lors de la présentation séquentielle de la base d'apprentissage. Les 75 premières observations appartiennent à la classe  $C_1$  et les 75 suivantes à la classe  $C_2$ .



FIG. 6.11 : Sorties estimées par le réseau modulaire à trois experts lors de la présentation séquentielle de la base de test. Les 75 premières observations appartiennent à la classe  $C_1$  et les 75 suivantes à la classe  $C_2$ .

Maintenant, on peut se poser la question, comment les différents experts partagent entre eux l'ensemble des observations constituant la base d'apprentissage?

Pour répondre à cette question, nous proposons de calculer les probabilités a posteriori  $h_k$ ,  $k = 1, 2, \dots, K$ , estimées à la sortie des experts pour chacune des observations présentées à l'entrée du réseau modulaire. En fait, l'expert qui présente la plus grande probabilité a posteriori est celui le plus probable d'être sélectionné par le contrôleur pour classer l'observation présentée à son entrée.

Les probabilités a posteriori dans le cas de notre réseau modulaire à trois experts sont représentées dans les trois figures 6.12, 6.13 et 6.14. On peut remarquer que pour chaque expert, la probabilité a posteriori est maximale pour certaines observations et minimale pour d'autres observations. On note :

- $h_1$ , la probabilité a posteriori de l'expert 1, est maximale pour certaines observations de la classe  $C_2$  et minimale pour les autres observations,
- $h_2$ , la probabilité a posteriori de l'expert 2, est maximale pour certaines observations de la classe  $C_1$  et minimale pour les autres observations,
- h<sub>3</sub>, la probabilité a posteriori de l'expert 3, est maximale pour les observations telles que h<sub>1</sub> et h<sub>2</sub> sont minimales. Par contre, elle est minimale dans le cas où h<sub>1</sub> et h<sub>2</sub> sont maximales.

En d'autres termes, le rôle du contrôleur est de sélectionner :

- l'expert 1 lorsque certaines observations appartenant à la classe  $C_2$  sont présentées à l'entrée du réseau modulaire,
- l'expert 2 lorsque certaines observations provenant de la classe  $C_1$  sont présentées à l'entrée du réseau modulaire,

• l'expert 3 lorsque certaines observations appartenant à la classe  $C_1$  et à la classe  $C_2$  sont présentées à l'entrée du réseau modulaire.

Par conséquent, on peut conclure que les experts 1 et 2 sont spécialisés respectivement dans une bonne partie des observations de la classe  $C_2$  et  $C_1$ . Par contre, l'expert 3 est spécialisé dans le reste des observations des deux classes  $C_1$  et  $C_2$ .



FIG. 6.12 : Probabilités a posteriori de l'experts 1, dans le cas d'un réseau modulaire à trois experts. Les 75 premières observations appartiennent à la classe  $C_1$  et les 75 suivantes à la classe  $C_2$ .



FIG. 6.13 : Probabilités a posteriori de l'expert 2, dans le cas d'un réseau modulaire à trois experts. Les 75 premières observations appartiennent à la classe  $C_1$  et les 75 suivantes à la classe  $C_2$ .



FIG. 6.14 : Probabilités a posteriori de l'expert 3, dans le cas d'un réseau modulaire à trois experts. Les 75 premières observations appartiennent à la classe  $C_1$  et les 75 suivantes à la classe  $C_2$ .

#### 6.5.2 Exemple 2 : Deux classes en Ou-Exclusif.

Considérons un ensemble d'apprentissage composé de 600 observations réparties en 4 groupes de points  $G_1$ ,  $G_2$ ,  $G_3$  et  $G_4$  dans un espace à deux dimensions. Les groupes sont constitués respectivement de 200, 100, 200 et 100 points générés à partir de lois normales. Les caractéristiques statistiques à savoir les vecteurs moyennes et les matrices de covariances sont résumées dans les deux tableaux des figures 6.16 et 6.17. On suppose que les groupes de points  $G_1$  et  $G_2$  appartiennent à la même classe notée  $C_1$  et que  $G_3$  et  $G_4$  appartiennent à la même classe notée  $C_2$ . C'est le problème du Ou-Exclusif. On code la classe  $C_1$  par d = 1 et la classe  $C_2$  par d = 0.

Nous avons aussi généré un ensemble de généralisation constitué de 600 observations dans les mêmes conditions que celles de l'ensemble d'apprentissage. La figure 6.15 illustre la répartition dans le plan de l'ensemble des observations de la base de données.



FIG. 6.15 : Deux classes  $C_1$  et  $C_2$  en Ou-Exclusif.

	Classe $C_1$					
	Gaussienne $G_1$	Gaussienne $G_2$				
Nombre de points	200	100				
Vecteur moyenne	$\left(\begin{array}{c} 2.0\\ 2.0\end{array}\right)$	$\left(\begin{array}{c} -2.0\\ -2.0 \end{array}\right)$				
Matrice de covariance	$\left(\begin{array}{rrr} 0.70 & -0.08 \\ -0.08 & 0.66 \end{array}\right)$	$\left(\begin{array}{rrr} 0.78 & -0.09 \\ -0.09 & 0.77 \end{array}\right)$				

FIG. 6.16 : Caractéristiques statistiques des deux gaussiennes de la classe  $C_1$ .

	Classe $C_2$					
	Gaussienne $G_3$	Gaussienne $G_3$				
Nombre de points	200	100				
Vecteur moyenne	$\left(\begin{array}{c} -2.0\\ 2.0 \end{array}\right)$	$\left(\begin{array}{c} 2.0\\ -2.0\end{array}\right)$				
Matrice de covariance	$\left(\begin{array}{cc} 0.68 & -0.03 \\ -0.03 & 0.61 \end{array}\right)$	$\left(\begin{array}{cc} 0.63 & 0.04 \\ 0.04 & 0.55 \end{array}\right)$				

FIG.	6.17	:	Caractéristiques	statistiques	des d	deux	gaussiennes	de	la c	lasse	$C_2$	•
------	------	---	------------------	--------------	-------	------	-------------	----	------	-------	-------	---

Pour entraîner les différentes architectures du réseau modulaire, nous avons choisi un pas d'adaptation égal à 0.01 et un nombre d'itérations égal à 500.

Nombre des Experts	2	3	4	5
Critère AIC	24.1	22.56	21.61	22.50
$\bar{E}$ (Apprentissage %)	2.29	2.22	2.10	2.24
$\bar{E}$ (Test %)	2.32	2.40	2.36	2.36
$\sigma_E$ (Apprentissage %)	10.40	9.81	9.61	9.81
$\sigma_E \text{ (Test \%)}$	10.85	10.70	10.73	10.53
$N_{obs}$ (Apprentissage)	9	7	7	7
$N_{obs}$ (Test)	11	9	9	9

Les résultats obtenus sont consignés dans le tableau de la figure 6.18. Ils montrent que la valeur minimale du critère AIC est obtenue en utilisant quatre experts.

FIG. 6.18 : Résultats de la discrimination des deux classes en XOR obtenus en utilisant des architectures modulaires.

Une fois l'apprentissage terminé, on présente les deux bases au réseau. On calcule les sorties estimées par le réseau modulaire pour chaque observation. Les figures 6.19 et 6.20 donnent, pour chaque observation présentée au réseau, la sortie estimée et sa valeur désirée correspondante. Le réseau modulaire à quatre experts permet de séparer les deux classes  $C_1$  et  $C_2$  en effectuant 7 erreurs de classification dans le cas de la base d'apprentissage et 9 erreurs de classification dans le cas de la base de test.



FIG. 6.19 : Sorties estimées par le réseau modulaire à 4 experts lors de la présentation séquentielle de la base d'apprentissage. Les 300 premières observations appartiennent à la classe  $C_1$  et les 300 suivantes à la classe  $C_2$ .



FIG. 6.20 : Sorties estimées par le réseau modulaire à 4 experts lors de la présentation séquentielle de la base de généralisation. Les 300 premières observations appartiennent à la classe  $C_1$  et les 300 suivantes à la classe  $C_2$ .

Les probabilités a posteriori des observations estimées à la sortie de chaque expert dans le cas d'une architecture modulaire à 4 experts sont données dans la figure 6.22. On peut remarquer que chaque expert admet est une probabilité a posteriori maximale dans une région de l'espace des observations. On note que :

- $h_1$  est maximale pour les observations issues du groupement  $G_1$  de la classe  $C_1$  et minimale pour les autres observations,
- $h_2$  est maximale pour les observations issues du groupement  $G_3$  de la classe  $C_2$  et minimale pour les autres observations,
- $h_3$  est maximale pour les observations issues du groupement  $G_4$  de la classe  $C_2$  et minimale pour les autres observations,
- $h_4$  est maximale pour les observations issues du groupement  $G_2$  de la classe  $C_1$  et minimale pour les autres observations.

En d'autres termes, le rôle du contrôleur est de sélectionner :

- l'expert 1 lorsqu'une observation appartenant au groupement  $G_1$  de la classe  $C_1$  est présentée à l'entrée du réseau modulaire,
- l'expert 2 lorsqu'une observation appartenant au groupement  $G_3$  de la classe  $C_2$  est présentée à l'entrée du réseau modulaire,
- l'expert 3 lorsqu'une observation appartenant au groupement  $G_4$  de la classe  $C_2$  est présentée à l'entrée du réseau modulaire,
- l'expert 4 lorsqu'une observation appartenant au groupement  $G_2$  de la classe  $C_1$  est présentée à l'entrée du réseau modulaire.

Maintenant, analysons la répartition des observations dans le cas où le nombre d'experts est égal à 3. Les probabilités a posteriori estimées à la sortie de chaque expert sont illustrées dans la figure 6.21. On peut remarquer globalement que chacun des trois experts est sélectionné par le contrôleur pour classer un des trois groupements  $G_2$ ,  $G_3$  et  $G_4$ . La discrimination des observations du groupement  $G_1$ est réalisée grâce à une coopération entre les deux experts 2 et 3. Etudions la répartition des observations dans le cas d'un réseau modulaire à 5 experts (cf. figure 6.23). A partir de l'analyse des probabilités a posteriori des différents experts, on constate que les quatre experts respectivement 1, 2, 3 et 4 ont des probabilités a posteriori maximales respectivement pour les groupements  $G_4, G_1$ ,  $G_2$  et  $G_3$ . Par contre, la probabilité a posteriori de l'expert 5 reste toujours égale à 0, par conséquent, il ne participe pas dans la partition de l'espace des observations.

On peut conclure que le choix d'un réseau modulaire avec quatre experts est le plus adapté pour résoudre ce problème, ce qui est en accord avec le résultat obtenu à l'aide du critère informationnel d'Akaike AIC.



**FIG. 6.21 :** Probabilités a posteriori dans le cas d'un réseau modulaire à 3 experts. Les 150 premières observations appartiennent à la classe  $C_1$  et les 150 suivantes à la classe  $C_2$ .



FIG. 6.22 : Probabilités a posteriori dans le cas d'un réseau modulaire à 4 experts. Les 150 premières observations appartiennent à la classe  $C_1$  et les 150 suivantes à la classe  $C_2$ .



FIG. 6.23 : Probabilités a posteriori dans le cas d'un réseau modulaire à 5 experts. Les 150 premières observations appartiennent à la classe  $C_1$  et les 150 suivantes à la classe  $C_2$ .

# 6.6 Application à la surveillance du multiplicateur de l'éolienne

Tout d'abord, nous précisons que cette étude va être effectuée uniquement sur la base de données extraites à partir des signaux issus de l'accéléromètre axial installé sur le multiplicateur. Nous constituons une base d'apprentissage de 400 observations dans un espace de dimension 20 réparties en deux classes notées  $C_1$  et  $C_2$ . La classe  $C_1$  comporte 200 observations représentant le fonctionnement normal du multiplicateur de l'éolienne et la classe  $C_2$  comprend 200 observations simulant une dérive de fonctionnement due à l'apparition d'une perturbation dont l'amplitude est égale à l'amplitude maximale dans la bande numéro 12 (cf. tableau de la figure 5.64).

Pour tester les performances de notre approche, nous avons généré trois bases de test de deux classes de 200 observations chacune. Pour les 3 bases de test, la classe  $C_1$  représente le fonctionnement normal du multiplicateur de vitesse, tandis que la classe  $C_2$  simule :

- Pour la base de test 1 : une dérive de fonctionnement obtenue dans les mêmes conditions que la base d'apprentissage (cf. tableau de la figure 5.64).
- Pour la base de test 2 : une dérive de fonctionnement dont l'amplitude de la composante perturbatrice est égale à la moitié de l'amplitude maximale dans la bande numéro 12 (cf. tableau de la figure 5.63),
- Pour la base de test 3 : une dérive de fonctionnement dont l'amplitude de la composante perturbatrice est égale à deux fois l'amplitude maximale dans la bande numéro 12 (cf. tableau de la figure 5.65).

Notre objectif est de pouvoir décider si le point de fonctionnement actuel de la machine est situé dans un domaine de fonctionnement normal ou défaillant. Pour ce faire, Nous avons entraîné plusieurs fois un réseau modulaire avec différentes architectures afin de discriminer entre les deux modes de fonctionnement représentés par les deux classes  $C_1$  et  $C_2$ . Rappelons que le pas d'adaptation a été fixé à 0.01 et le nombre d'itérations est égal à 500.
D'après le tableau de la figure 6.24, le nombre optimal des experts, celui qui minimise le critère AIC, est égal à 2. Donc, dans ce cas, l'architecture la mieux adaptée sera composée de 2 experts.

Nombre des Experts	2	3	4	5
Critère AIC	9.16	11.79	14.98	15.29
$\bar{E}$ (Apprentissage %)	2.15	2.75	3.47	3.54
$\bar{E}$ (Test %)	8.13	8.89	10.11	10.21
$\sigma_E$ (Apprentissage %)	4.31	4.95	5.62	5.67
$\sigma_E$ (Test %)	17.07	16.56	17.48	17.50
$N_{obs}$ (Apprentissage)	0	0	0	0
$N_{obs}$ (Test)	19	19	22	24

**FIG. 6.24 :** Résultats de la discrimination entre les deux classes  $C_1$  et  $C_2$  obtenus en utilisant des architectures modulaires.

Les probabilités a posteriori estimées à la sortie des deux experts sont illustrées dans la figure 6.25. On peut constater que chaque expert est spécialisé dans un mode de fonctionnement.

Une fois l'apprentissage terminé, on fige les poids et on présente séquentiellement les observations de la base d'apprentissage et celles des trois bases de test au réseau. On calcule les sorties estimées par le réseau modulaire pour chaque observation. Les figures 6.26 à 6.29 donnent, pour chaque observation présentée au réseau, la sortie estimée et sa valeur désirée correspondante.



FIG. 6.25 : Probabilités a posteriori dans le cas d'un réseau modulaire à 2 experts. Les 200 premières observations appartiennent à la classe  $C_1$  et les 200 suivantes à la classe  $C_2$ .



FIG. 6.26 : Sorties estimées par le réseau modulaire à 2 experts lors de la présentation séquentielle de la base d'apprentissage. Les 200 premières observations appartiennent à la classe  $C_1$  et les 200 suivantes à la classe  $C_2$ .



FIG. 6.27 : Sorties estimées par le réseau modulaire à 2 experts lors de la présentation séquentielle de la base de test 1. Les 200 premières observations appartiennent à la classe  $C_1$  et les 200 suivantes à la classe  $C_2$ .



FIG. 6.28 : Sorties estimées par le réseau modulaire à 2 experts lors de la présentation séquentielle de la base de test 2. Les 200 premières observations appartiennent à la classe  $C_1$  et les 200 suivantes à la classe  $C_2$ .



FIG. 6.29 : Sorties estimées par le réseau modulaire à 2 experts lors de la présentation séquentielle de la base de test 3. Les 200 premières observations appartiennent à la classe  $C_1$  et les 200 suivantes à la classe  $C_2$ .

Les matrices de confusion sont données dans le tableau de la figure 6.30. On remarque que dans le cas de la base de test 2 simulant une faible perturbation, le réseau modulaire à 2 experts partage les deux classes en effectuant un nombre important d'erreurs de classification. Ce nombre décroît lorsque l'amplitude de la composante perturbatrice augmente (cas des deux bases 1 et 3). En fait, d'après les projections planes obtenues dans les figures 5.66, 5.68 et 5.70 du chapitre précédent, on note un chevauchement très important dans le cas de la base de test 2. Ce chevauchement est moins important pour les deux bases de test 1 et 3.

:	Apprentissage			Tes	st 1			Tes	t 2		Tes	st 3
	$C_1$	$C_2$		$C_1$	$C_2$			$C_1$	$C_2$		$C_1$	$C_2$
$\hat{C}_1$	200	0	$\hat{C}_1$	196	4		$\hat{C}_1$	196	4	$\hat{C}_1$	196	4
$\hat{C}_2$	0	200	$\hat{C}_2$	15	185		$\hat{C}_2$	187	13	$\hat{C}_2$	0	200

FIG. 6.30 : Les matrices de confusion dans le cas de la base d'apprentissage et des trois bases de test.

Lors de la phase d'exploitation, chaque observation représente environ une seconde de fonctionnement de la machine. Donc, pour chaque observation présentée à l'entrée, le réseau modulaire doit fournir en sortie une réponse indiquant la classe d'appartenance de cette observation. Le suivi de fonctionnement par le réseau conduit à 3 cas de figures :

- si la réponse du réseau est toujours voisine de 1, le point de fonctionnement de la machine se trouve dans un domaine de fonctionnement normal,
- si la réponse du réseau oscille entre des valeurs tantôt voisines de 1 et tantôt voisines de 0, cela indique la présence d'une composante perturbatrice,
- si la réponse du réseau est constamment proche de 0, la machine présente un défaut permanent.

### 6.7 Conclusion

Dans ce chapitre, nous avons présenté un réseau de neurones d'architecture modulaire. L'idée de base est de décomposer une tâche complexe en un ensemble de tâches plus simples facilement réalisables par des modules dits experts. Le résultat final est un mélange de plusieurs solutions pondérées par un module contrôleur qui active le module expert approprié connaissant l'observation présentée à l'entrée du réseau modulaire.

Nous avons décrit la stratégie d'apprentissage basée sur la vraisemblance d'un mélange de distributions de probabilités conditionnelles de Bernoulli qui est bien adaptée aux problèmes de classification supervisée. Les poids synaptiques du réseau sont actualisés grâce à la règle du gradient appliquée au critère d'erreur faisant intervenir la fonction du log-vraisemblance. L'optimisation de l'architecture du réseau est assurée à l'aide du critère informationnel d'Akaike.

Les performances de l'application d'une telle architecture ont été validées sur deux exemples simulés bien connus dans la littérature de la classification.

Pour suivre le fonctionnement de l'éolienne, un réseau modulaire a été entraîné pour discriminer entre ses deux modes de fonctionnement : normal et défaillant. L'apprentissage du réseau est effectué en mode hors-ligne. En phase d'exploitation, le réseau fournit, en temps réel, une réponse indiquant que le point de fonctionnement actuel est situé soit dans le domaine de fonctionnement normal soit dans un domaine de défaillance.

### Conclusion générale et perspectives

Nous avons présenté une approche de suivi de fonctionnement d'une éolienne par analyse des signaux de vibrations observés sur des fenêtres temporelles glissantes. En particulier, nous sommes intéressés par la surveillance du multiplicateur et de la génératrice éléments sensibles sur lesquels nous avons installé des capteurs de vibrations. En fait, notre souci est de développer une méthode de diagnostic suffisamment générale pour être applicable à toute machine tournante.

Puisqu'il est difficile voire dangereux de créer des défauts sur une machine en pleine production, les signaux que nous avons collectés représentent uniquement le fonctionnement normal. Notre démarche a consisté alors à caractériser ce fonctionnement par des spectrogrammes d'énergie calculés sous différentes conditions d'exploitation, à savoir différentes vitesses du vent sur une durée couvrant plusieurs jours. Ces spectres ont été réduits à un nombre limité de bandes d'énergies réalisant ainsi des gabarits. Les spectres réduits ont constitué les entrées et les sorties de réseaux de neurones autoassociateurs. Ce type de réseaux assure une double fonction : une projection non linéaire plane obtenue sur sa couche centrale et une estimation du spectre de fonctionnement normal en sortie. La projection non linéaire offre une visualisation plane des données alors que les comparaisons entre les sorties estimées et réelles génèrent des résidus permettant de vérifier si le spectre courant a dévié ou non du fonctionnement normal mémorisé par le réseau.

Pour valider notre démarche, nous avons testé les performances du réseau sur des données en fonctionnement normal acquis dans des situations différentes de celles de la base d'apprentissage et sur des situations correspondant à des défauts simulés type balourd. Ce type de défauts se manifeste par l'apparition d'une fréquence et ses harmoniques dans le spectrogramme. Dans les deux situations, les résultats obtenus par le réseau ont été quantifiés et se sont révélés satisfaisants.

Nous avons exploité, par ailleurs, un réseau modulaire à apprentissage supervisé pour la discrimination entre deux situations : le fonctionnement normal et la présence d'un défaut type balourd d'une amplitude minimale fixée. Les performances du réseau ont été testées sur des défauts d'amplitudes différentes.

Notre avons validé notre approche sur une éolienne asynchrone à vitesse constante par analyse des signaux de vibrations issus du multiplicateur et de la génératrice. Reste à surveiller d'autres parties de l'éolienne comme par exemple la température, le niveau d'huile et les signaux sonores du micro etc.

Nous sommes conscient que la phase de caractérisation du fonctionnement reste le problème le plus important à résoudre. Nous avons employé les techniques d'analyse spectrale utilisant la transformée de Fourier rapide. Nous justifions cette démarche par le fait que l'éolienne est une machine tournante et il est important de connaître et vérifier la présence de certaines fréquences caractéristiques de certains éléments de la machine. Pour ce faire, nous avons eu recours à l'observation : observation des signaux du multiplicateur et de la génératrice et leurs spectres pour bien définir les fréquences à surveiller et mieux délimiter l'étendu des bandes d'énergie. Bien sûr, d'autres techniques temporelles ou statistiques peuvent aussi être employées voire combinées.

Une nouvelle technologie performante d'éoliennes à vitesses variables a été récemment développée. Il serait intéressant d'appliquer notre démarche pour surveiller le fonctionnement de ces machines. A notre sens, l'analyse spectrale doit être reprise afin d'en extraire certaines régularités pour caractériser le fonctionnement. Par contre, la partie décision par réseaux de neurones est suffisamment générale pour être directement exploitée et ne devrait poser aucun problème particulier.

Souvent les méthodes de diagnostic sont appliquées sur des bancs d'essais dans des conditions de laboratoire contrôlées. Notre démarche est d'autant plus intéressante que nous surveillons une machine réelle en pleine production.

# Annexes

# Table des figures

1.1	Evolution mondiale de la puissance installée entre 1994 et 2000	9
2.1	Vue générale des éléments constituant une éolienne	13
2.2	Image vibratoire théorique : (a) d'un balourd, (b) d'un défaut d'alignement	15
2.3	Image vibratoire théorique : (a) d'un engrenage sain, (b) d'un engrenage	
	présentant une dent détériorée	17
2.4	Signal acoustique issu du microphone, la fréquence d'échantillonnage $f_e$ =48KHz.	24
2.5	Spectre d'amplitude du microphone calculé sur une fenêtre de 32768 échan-	
	tillons	24
2.6	Signal d'accélération du multiplicateur issu de l'accéléromètre axial, la fréquence	
	d'échantillonnage $f_e$ =48KHz	25
2.7	Spectre d'amplitude du signal d'accélération du multiplicateur issu de l'ac-	
	céléromètre axial calculé en utilisant une fenêtre de 32768 échantillons	25
2.8	Signal d'accélération du multiplicateur issu de l'accéléromètre radial, la fréquence	
	d'échantillonnage $f_e$ =48KHz	26
2.9	Spectre d'amplitude du signal d'accélération du multiplicateur issu de l'ac-	
	céléromètre radial calculé en utilisant une fenêtre de 32768 échantillons.	26
2.10	Signal d'accélération de la génératrice issu de l'accéléromètre axial, la fréquence	
	d'échantillonnage $f_e$ =48KHz	27
2.11	Spectre d'amplitude du signal d'accélération de la génératrice issu de l'ac-	
	céléromètre axial calculé en utilisant une fenêtre de 32768 échantillons	27
2.12	Signal courant mesuré sur une phase de la génératrice, la fréquence d'échan-	
	tillonnage $f_e$ =48KHz	28

2.13	Spectre d'amplitude du signal courant calculé sur une fenêtre de 32768 échan-	
	tillons	28
2.14	Fréquences maximales d'échantillonnage pour les capteurs analogiques	29
2.15	Fréquences maximales d'échantillonnage pour les capteurs tout ou rien	30
2.16	Cas 1 : Tous les signaux sont échantillonnés à la même fréquence.	32
2.17	Cas 2 : Les signaux sont échantillonnés à des fréquences différentes	32
3.1	Fenêtres temporelles et fréquentielles rectangulaires et de Hanning	38
3.2	Fenêtres temporelles rectangulaire et de Hanning	39
3.3	Réponses fréquentielles des fenêtres : rectangulaire et de Hanning	39
3.4	Allure du signal temporel simulé	41
3.5	Spectre d'amplitude du signal simulé calculé sur une fenêtre de $1024$ échan-	
	tillons, la fréquence d'échantillonnage $f_e=3$ KHz	41
3.6	Spectrogramme d'amplitude du signal simulé calculé sur 15 fenêtres succes-	
	sives. La taille de chaque fenêtre est $N=1024$ échantillons	42
3.7	Spectrogramme du signal de vibration de la génératrice lors du démarrage	
	de l'éolienne obtenu en utilisant une fenêtre de Hanning de $2^{15}=32768$	
	échantillons, la fréquence d'échantillonnage $f_e=$ 31KHz	43
3.8	Spectrogramme du signal de vibration issu de l'accéléromètre axial sur le mul-	
	tiplicateur enregistré par le DAT obtenu en utilisant une fenêtre de Hanning	
	de $2^{15}=32768$ échantillons, la fréquence d'échantillonnage $f_e=$ 48KHz	45
3.9	Spectrogramme du signal de vibration issu de l'accéléromètre axial sur le	
	multiplicateur enregistré par la carte DT3001 obtenu en utilisant une fenêtre	
	de Hanning de $2^{15}~=~32768$ échantillons, la fréquence d'échantillonnage	
	$f_e=31$ KHz	45
3.10	Spectrogramme du signal de vibration issu de l'accéléromètre radial sur le	
	multiplicateur enregistré par le DAT obtenu en utilisant une fenêtre de Han-	
	ning de $2^{15} = 32768$ échantillons, la fréquence d'échantillonnage $f_e$ =48KHz.	46

3.11	Spectrogramme du signal de vibration issu de l'accéléromètre radial sur le	
	multiplicateur enregistré par la carte DT3001 obtenu en utilisant une fenêtre	
	de Hanning de $2^{15}=32768$ échantillons, la fréquence d'échantillonnage	
	$f_e=31$ KHz	46
3.12	Spectrogramme du signal de vibration de la génératrice enregistré par le DAT	
	obtenu en utilisant une fenêtre de Hanning de $2^{15}=32768$ échantillons, la	
	fréquence d'échantillonnage $f_e$ =48KHz	47
3.13	Spectrogramme du signal de vibration de la génératrice enregistré par la	
	carte DT3001 obtenu en utilisant une fenêtre de Hanning de $2^{15}=32768$	
	échantillons, la fréquence d'échantillonnage $f_e=$ 31KHz	47
3.14	Fréquences dues à la rotation de la structure pales-rotor.	49
3.15	Fréquences dues à la rotation des arbres et des engrenages du multiplicateur.	50
3.16	Harmoniques d'ordre 1, 2, 3 et 4 dues à la rotation des arbres de transmission	
	et des engrenages du multiplicateur	50
3.17	Fréquences dues aux roulements du multiplicateur	51
3.18	Fréquence dues à la rotation de la génératrice	51
4.1	Réseau autoassociateur linéaire à trois couches	59
4.2	Notations des poids et des variables du réseau autoassociateur linéaire	60
4.3	Allure d'un spectre réduit en 5 bande du signal simulé	67
4.4	Les 2 premières valeurs propres et le pourcentage de la variance totale des	
	observations expliquée par chaque valeur propre	68
4.5	La réponse des sorties de la couche centrale d'un autoassociateur linéaire	69
4.6	Les erreurs de reconstruction par bande obtenues au niveau des sorties de	
	l'autoassociateur linéaire à 3 couches	69
4.7	Réseau autoassociateur non linéaire à trois couches	70
4.8	Notations des poids et des variables du réseau autoassociateur non linéaire.	71
4.9	La réponse des sorties de la couche centrale d'un autoassociateur non linéaire	
	à 3 couches	76
4.10	Les erreurs de reconstruction par bande obtenues au niveau des sorties de	
	l'autoassociateur non linéaire à 3 couches.	76

4.11	Réseau de neurones autoassociateur à trois couches cachées	77
4.12	Connexion du $i^{\grave{e}me}$ neurone de la couche $(l-1)$ vers le $j^{\grave{e}me}$ neurone de la	
	couche ( <i>l</i> )	78
4.13	L'évolution des critères AIC (figure (a)) et FPE (figure (b)) en fonction du	
	nombre de neurones $H$	82
4.14	La réponse des sorties de la couche centrale d'un autoassociateur non linéaire	
	à 5 couches	83
4.15	Les erreurs de reconstruction par bande obtenues au niveau des sorties de	
	l'autoassociateur non linéaire à 5 couches	83
4.16	Pourcentage d'énergie apportée par la composante perturbatrice au sein	
	d'une bande étroite et large	84
4.17	La réponse des sorties de la couche centrale d'un autoassociateur non linéaire	
	à 5 couches	85
4.18	Les erreurs de reconstruction par bande obtenues au niveau des sorties de	
	l'autoassociateur non linéaire à 5 couches	85
4.19	La réponse des sorties de la couche centrale d'un autoassociateur non linéaire	
	à 5 couches	86
4.20	Les erreurs de reconstruction par bande obtenues au niveau des sorties de	
	l'autoassociateur non linéaire à 5 couches	86
5.1	Densité spectrale d'énergie obtenue en utilisant une fenêtre de Hanning com-	
	prenant $2^{15}$ échantillons, la fréquence d'échantillonnage est de $31$ KHz	91
5.2	Densité spectrale d'énergie, zone 1 ([0-500]Hz)	92
5.3	Spectrogramme d'énergie du signal, zone 1 ([0-500]Hz)	92
5.4	Densité spectrale d'énergie, zone 2 ([500-1000]Hz)	93
5.5	Spectrogramme d'énergie du signal, zone 2 ([500-1000]Hz)	93
5.6	Densité spectrale d'énergie, zone 3 ([1000-1500]Hz)	94
5.7	Spectrogramme d'énergie du signal, bande 3 ([1000-1500]Hz)	94
5.8	Densité spectrale d'énergie, zone 4 ([1500-2400]Hz)	95
5.9	Spectrogramme d'énergie du signal, zone 4 ([1500-2400]Hz)	95
5.10	Fréquences à surveiller.	97

5.11	Puissance fournie en fonction de la vitesse de rotation du rotor de la génératrice. 99
5.12	Valeurs prises par les principales fréquences de rotation lorsque la vitesse de
	rotation de la génératrice varie entre 1498 tr/mn et 1520 tr/mn $100$
5.13	Fluctuations de la fréquence du pic principal autour de la fréquence $555 \mathrm{Hz.}$ . $101$
5.14	Fluctuations de la fréquence du pic principal autour de la fréquence $680 {\rm Hz.}$ . $101$
5.15	Fluctuations de la fréquence du pic principal autour de la fréquence $850 {\rm Hz.}$ . $101$
5.16	Spectre réduit en 20 bandes
5.17	Densité spectrale d'énergie du signal d'accélération radiale du multiplicateur
	obtenue en utilisant une fenêtre de Hanning comprenant $2^{15}$ échantillons, la
	fréquence d'échantillonnage est de 31KHz
5.18	Densité spectrale d'énergie, zone 1 ([0-500]Hz)
5.19	Spectrogramme d'énergie du signal, zone 1 ([0-500]Hz)
5.20	Densité spectrale d'énergie, zone 2 ([500-1000]Hz)
5.21	Spectrogramme d'énergie du signal, zone 2 ([500-1000]Hz)
5.22	Densité spectrale d'énergie, zone 3 ([1000-1500]Hz)
5.23	Spectrogramme d'énergie du signal, zone 3 ([1000-1500]Hz)
5.24	Densité spectrale d'énergie, zone 4 ([1500-2400]Hz)
5.25	Spectrogramme d'énergie du signal, zone 4 ([1500-2400]Hz)
5.26	Valeurs prises par les principales fréquences de rotation lorsque la vitesse de
	rotation de la génératrice varie entre 1498 tr/mn et 1520 tr/mn $108$
5.27	Densité spectrale d'énergie du signal d'accélération axiale de la génératrice
	obtenue en utilisant une fenêtre de Hanning comprenant $2^{15}$ échantillons, la
	fréquence d'échantillonnage est de 31KHz
5.28	Densité spectrale d'énergie, zone 1 ([0-500]Hz)
5.29	Spectrogramme d'énergie du signal, zone 1 ([0-500]Hz)
5.30	Densité spectrale d'énergie, zone 2 ([500-1000]Hz)
5.31	Spectrogramme d'énergie du signal, zone 2 ([500-1000]Hz)
5.32	Densité spectrale d'énergie, zone 3 ([1000-1500]Hz)
5.33	Spectrogramme d'énergie du signal, zone 3 ([1000-1500]Hz)
5.34	Densité spectrale d'énergie, zone 4 ([1500-2400]Hz)

5.35	Spectrogramme d'énergie du signal, zone 4 ([1500-2400]Hz)
5.36	Valeurs prises par les principales fréquences de rotation lorsque la vitesse de
	rotation de la génératrice varie entre 1498 tr/mn et 1520 tr/mn
5.37	Accéléromètre axial : Pourcentage de la variance totale des observations
	expliqué par chaque valeur propre
5.38	Accéléromètre radial : Pourcentage de la variance totale des observations
	expliqué par chaque valeur propre
5.39	Accéléromètre axial : Le plan de projection obtenu par les sorties de la
	couche centrale d'un autoassociateur linéaire à 3 couches. $\ldots$ $\ldots$ $\ldots$ $118$
5.40	Accéléromètre axial : Les erreurs de reconstruction par bande obtenues au
	niveau des sorties de l'autoassociateur linéaire à 3 couches
5.41	Accéléromètre radial : Le plan de projection obtenu par les sorties de la
	couche centrale d'un autoassociateur linéaire à 3 couches
5.42	Accéléromètre radial : Les erreurs de reconstruction par bande obtenues
	au niveau des sorties de l'autoassociateur linéaire à 3 couches
5.43	Accéléromètre axial : Le plan de projection obtenu par les sorties de la
	couche centrale de l'autoassociateur linéaire à 3 couches
5.44	Accéléromètre axial : Les erreurs de reconstruction par bande obtenues au
	niveau des sorties de l'autoassociateur linéaire à 3 couches
5.45	Accéléromètre radial : Le plan de projection obtenu par les sorties de la
	couche centrale de l'autoassociateur linéaire à 3 couches
5.46	Accéléromètre radial : Les erreurs de reconstruction par bande obtenues
	au niveau des sorties de l'autoassociateur linéaire à 3 couches. $\ldots$ $\ldots$ $121$
5.47	La moyenne et l'écart type de l'erreur en fonction du nombre d'itérations
	calculés sur l'ensemble d'apprentissage pour un autoassociateur non linéaire
	à 3 couches
5.48	La moyenne et l'écart type en fonction du nombre d'itérations de l'erreur
	calculés sur l'ensemble de test pour un autoassociateur non linéaire à 3 couches. $123$
5.49	Plan de projection obtenu par les sorties de la couche centrale de l'autoasso-
	ciateur non linéaire à 3 couches

5.50	Erreurs de reconstruction calculées sur l'ensemble d'apprentissage et de test
	à la sortie de l'autoassociateur non linéaire à 3 couches. $\ldots$ $\ldots$ $\ldots$ $124$
5.51	L'évolution des critères AIC (figure (a)) et FPE (figure (b)) en fonction du
	nombre de neurones $H$
5.52	La moyenne et l'écart type de l'erreur calculée sur l'ensemble d'apprentissage
	pour un autoassociateur non linéaire à 5 couches
5.53	La moyenne et l'écart type de l'erreur calculée sur l'ensemble de test pour un
	autoassociateur non linéaire à 5 couches
5.54	Plan de projection obtenu par les sorties de la couche centrale de l'autoasso-
	ciateur non linéaire à 5 couches
5.55	Erreurs de reconstruction par bande obtenues au niveau des sorties de l'au-
	toassociateur non linéaire à 5 couches
5.56	Les 2 premières valeurs propres et le pourcentage de la variance totale des
	observations expliquée par chaque valeur propre
5.57	Plan de projection obtenu par les sorties de la couche centrale de l'autoasso-
	ciateur linéaire à 3 couches
5.58	Erreurs de reconstruction par bande obtenues au niveau des sorties de l'au-
	toassociateur linéaire à 3 couches
5.59	Plan de projection obtenu par les sorties de la couche centrale de l'autoasso-
	ciateur non linéaire à 3 couches
5.60	Erreurs de reconstruction calculées sur l'ensemble d'apprentissage et de test
	calculées à la sortie d'un autoassociateur non linéaire à 3 couches. $\ldots$ $132$
5.61	Plan de projection obtenue par les sorties de la couche centrale de l'autoas-
	sociateur non linéaire à 5 couches
5.62	Erreurs de reconstruction par bande obtenues au niveau des sorties de l'au-
	toassociateur non linéaire à 5 couches
5.63	Base 1 : Pourcentage d'énergie apportée par la composante perturbatrice et
	son premier harmonique
5.64	Base 2 : Pourcentage d'énergie apportée par la composante perturbatrice et
	son premier harmonique

.

5.65	Base 3 : Pourcentage d'énergie apportée par la composante perturbatrice et
	son premier harmonique
5.66	Base 1 : Plan de projection obtenue par les sorties de la couche centrale de
	l'autoassociateur non linéaire à 5 couches
5.67	Base 1 :Erreurs de reconstruction par bande obtenues au niveau des sorties
	de l'autoassociateur non linéaire à 5 couches
5.68	Base 2 : Plan de projection obtenue par les sorties de la couche centrale de
	l'autoassociateur non linéaire à 5 couches
5.69	Base 2 :Erreurs de reconstruction par bande obtenues au niveau des sorties
	de l'autoassociateur non linéaire à 5 couches
5.70	Base 3 : Plan de projection obtenue par les sorties de la couche centrale de
	l'autoassociateur non linéaire à 5 couches
5.71	Base 3 :Erreurs de reconstruction par bande obtenues au niveau des sorties
	de l'autoassociateur non linéaire à 5 couches
5.72	Base 1 : Pourcentage d'énergie apportée par la composante perturbatrice et
	son premier harmonique
5.73	Base 2 : Pourcentage d'énergie apportée par la composante perturbatrice et
	son premier harmonique
5.74	Base 3 : Pourcentage d'énergie apportée par la composante perturbatrice et
	son premier harmonique
5.75	Base 1 : Plan de projection obtenue par les sorties de la couche centrale de
	l'autoassociateur non linéaire à 5 couches
5.76	Base 1 :Erreurs de reconstruction par bande obtenues au niveau des sorties
	de l'autoassociateur non linéaire à 5 couches
5.77	Base 2 : Plan de projection obtenue par les sorties de la couche centrale de
	l'autoassociateur non linéaire à 5 couches
5.78	Base 2 :Erreurs de reconstruction par bande obtenues au niveau des sorties
	de l'autoassociateur non linéaire à 5 couches
5.79	Base 3 : Plan de projection obtenue par les sorties de la couche centrale de
	l'autoassociateur non linéaire à 5 couches.

5.80	Base 3 :Erreurs de reconstruction par bande obtenues au niveau des sorties	
	de l'autoassociateur non linéaire à 5 couches	16
6.1	Schéma bloc du réseau modulaire. Les sorties des modules experts sont	
	pondérées par les sorties du module contrôleur	51
6.2	Architecture et notations des variables pour un module expert	52
6.3	Architecture du module contrôleur	52
6.4	Notations des variables pour un neurone du module contrôleur	52
6.5	Caractéristiques statistiques des deux croissants représentant les deux classes	
	$C_1$ et $C_2$ de la base de données. $\ldots$	31
6.6	Représentations graphiques de deux classes $C_1$ et $C_2$ en forme de croissant	
	représentant l'ensemble des observations de la base d'apprentissage et de la	
	base de test	52
6.7	Résultats de la discrimination entre deux croissants obtenus en utilisant un	
	réseau modulaire	;3
6.8	Allure de la moyenne de l'erreur lors de la phase d'apprentissage et de test	
	obtenue en utilisant un réseau modulaire à trois experts	5
6.9	Allure de l'écart type lors de la phase d'apprentissage et de test obtenue en	
	utilisant un réseau modulaire à trois experts	5
6.10	Sorties estimées par le réseau modulaire à trois experts lors de la présenta-	
	tion séquentielle de la base d'apprentissage. Les 75 premières observations	
	appartiennent à la classe $C_1$ et les 75 suivantes à la classe $C_2$	6
6.11	Sorties estimées par le réseau modulaire à trois experts lors de la présentation	
	séquentielle de la base de test. Les $75$ premières observations appartiennent	
	à la classe $C_1$ et les 75 suivantes à la classe $C_2$	6
6.12	Probabilités a posteriori de l'experts 1, dans le cas d'un réseau modulaire à	
	trois experts. Les 75 premières observations appartiennent à la classe $C_1$ et	
	les 75 suivantes à la classe $C_2$	8
6.13	Probabilités a posteriori de l'expert 2, dans le cas d'un réseau modulaire à	
	trois experts. Les 75 premières observations appartiennent à la classe $C_1$ et	
	les 75 suivantes à la classe $C_2$	9

6.14	Probabilités a posteriori de l'expert 3, dans le cas d'un réseau modulaire à
	trois experts. Les 75 premières observations appartiennent à la classe $C_1$ et
	les 75 suivantes à la classe $C_2$
6.15	Deux classes $C_1$ et $C_2$ en Ou-Exclusif
6.16	Caractéristiques statistiques des deux gaussiennes de la classe $C_1$
6.17	Caractéristiques statistiques des deux gaussiennes de la classe $C_2$
6.18	Résultats de la discrimination des deux classes en XOR obtenus en utilisant
	des architectures modulaires
6.19	Sorties estimées par le réseau modulaire à 4 experts lors de la présentation
	séquentielle de la base d'apprentissage. Les 300 premières observations ap-
	partiennent à la classe $C_1$ et les 300 suivantes à la classe $C_2$
6.20	Sorties estimées par le réseau modulaire à 4 experts lors de la présentation
	séquentielle de la base de généralisation. Les 300 premières observations ap-
	partiennent à la classe $C_1$ et les 300 suivantes à la classe $C_2$
6.21	Probabilités a posteriori dans le cas d'un réseau modulaire à 3 experts. Les
	$150$ premières observations appartiennent à la classe $C_1$ et les $150$ suivantes
	à la classe $C_2$
6.22	Probabilités a posteriori dans le cas d'un réseau modulaire à 4 experts. Les
	$150$ premières observations appartiennent à la classe $C_1$ et les $150$ suivantes
	à la classe $C_2$
6.23	Probabilités a posteriori dans le cas d'un réseau modulaire à 5 experts. Les
	$150$ premières observations appartiennent à la classe $C_1$ et les $150$ suivantes
	à la classe $C_2$
6.24	Résultats de la discrimination entre les deux classes $C_1$ et $C_2$ obtenus en
	utilisant des architectures modulaires
6.25	Probabilités a posteriori dans le cas d'un réseau modulaire à 2 experts. Les
	$200$ premières observations appartiennent à la classe $C_1$ et les $200$ suivantes
	à la classe $C_2$

6.26	Sorties estimées par le réseau modulaire à 2 experts lors de la présentation
	séquentielle de la base d'apprentissage. Les 200 premières observations ap-
	partiennent à la classe $C_1$ et les 200 suivantes à la classe $C_2$
6.27	Sorties estimées par le réseau modulaire à 2 experts lors de la présentation
	séquentielle de la base de test 1. Les 200 premières observations appartiennent
	à la classe $C_1$ et les 200 suivantes à la classe $C_2$
6.28	Sorties estimées par le réseau modulaire à 2 experts lors de la présentation
	séquentielle de la base de test 2. Les 200 premières observations appartiennent
	à la classe $C_1$ et les 200 suivantes à la classe $C_2$
6.29	Sorties estimées par le réseau modulaire à 2 experts lors de la présentation
	séquentielle de la base de test 3. Les 200 premières observations appartiennent
	à la classe $C_1$ et les 200 suivantes à la classe $C_2$
6.30	Les matrices de confusion dans le cas de la base d'apprentissage et des trois
	bases de test

### **Publications** personnelles

#### Publications internationales avec actes

- N. ELHOR, D. HAMAD ET J-G. POSTAIRE. Classification Supervisée par réseau de Neurones Modulaire, Les Journées Nationales d'Automatique, Electronique, Traitement de Signal et leurs Applications, pages 392-397, Marrakech, MAROC, 24-25 Octobre, 1996.
- N. ELHOR, R. BERTRAND ET D. HAMAD. Bernoulli Mixture of Experts for Supervised Pattern Recognition, In Proc. of the Int. Conf. on Artificial Neural Networks and Genetic Algorithms (ICANNGA'97), pages 44-48, Norwich, ENGLAND, April 2-4, 1997.
- E. VAN HECKE, N. ELHOR ET D. HAMAD. Caractérisation Granulométrique de produits Alimentaires par Analyse d'Images, Colloque International sur le Traitement d'Images et les Systèmes de Vision Artificielle (TISVA'98), pages 111-119, Oujda, MAROC, 28-29 Avril, 1998.
- R. BERTRAND, N. ELHOR ET M. BETROUNI AND D. HAMAD. On-line Supervision system for Wind Energy Converters, In Proc. of the 4nd Int. Conf. on Engineering Applications of Neural Networks (EANN'98), pages 269-272, GIBRALTAR, June 10-12, 1998.
- R. BERTRAND, N. ELHOR ET D. HAMAD AND J.-G. POSTAIRE. Supervision of wind energy converters by non linear mapping neural network, In Proc. of the Int. ICSC/IFAC Symposium on Neural Computation (NC'98), pages 329-334, Vienna, AUSTRIA, September 23-25, 1998.

N. ELHOR, R. BERTRAND ET D. HAMAD. Autoassociators for Supervision of Wind Energy Converters, Proc. of the Int. Conf. on World Manufacturing Congress (WMC'99), pages 427-431, Durham, ENGLAND, September 27-30, 1999.

#### Article dans un journal international

N. ELHOR, R. BERTRAND ET J.-G. POSTAIRE AND D. HAMAD. Neural Network for Wind Turbine Supervision, e&i, pages 123-127, juin 99.

### Rapport Région Nord-Pas de Calais

- NORELEC, EED, ETS GROS, UNIV. LITTORAL ET USTL IUT-A GMP. Optimisation du fonctionnement d'éoliennes par analyse de contraintes de maintenance, Rapport de fin phase 1 du programme de Recherche et Développement sur l'énergie éolienne, Décembre 1996.
- NORELEC, EED, ETS GROS, UNIV. LITTORAL ET USTL IUT-A GMP. Optimisation du fonctionnement d'éoliennes par analyse de contraintes de maintenance, Rapport de la phase 2 du programme de Recherche et Développement sur l'énergie éolienne, Décembre 1997.
- NORELEC, EED, ETS GROS, UNIV. LITTORAL ET USTL IUT-A GMP. Optimisation du fonctionnement d'éoliennes par analyse de contraintes de maintenance, Rapport de fin phase 2 du programme de Recherche et Développement sur l'énergie éolienne, Décembre 1998.
- NORELEC, EED, ETS GROS, UNIV. LITTORAL ET USTL IUT-A GMP. Optimisation du fonctionnement d'éoliennes par analyse de contraintes de maintenance, Rapport de la phase 3 du programme de Recherche et Développement sur l'énergie éolienne, Novembre 1999.

# Bibliographie

- [Aka72] H. Akaike. Information theory and an extension of the maximum likelihood principle. In Proc. of the 2nd int. Symposium on Information Theory, pages 267–281, 1972.
- [BEBH98] R. Bertrand, N. Elhor, M. Betrouni, et D. Hamad. On-line Supervision system for Wind Energy Converters. In Proc. of the 4nd Int. Conf. on Engineering Applications of Neural Networks (EANN'98), pages 269-272, GIBRALTAR, June 10-12, 1998.
- [BEHP98] R. Bertrand, N. Elhor, D. Hamad, et J.-G. Postaire. Supervision of wind energy converters by non linear mapping neural network. In Proc. of the Int. ICSC/IFAC Symposium on Neural Computation (NC'98), pages 329-334, Vienna, AUSTRIA, September 23-25, 1998.
- [Bet98] M. Betrouni. Projection de données multidimensionnelles par réseaux de neurones. PhD thesis, Université des Sciences et Technologies de Lille, 1998.
- [BFC98] T.J. Boehme, I. Fletcher, et C.S. Cox. Sensor Failure and Signal Reconstruction using Autoassociative Neural Networks. In Proc of the Int. ICSC/IFAC Symposium on Neural Computation (NC'98), pages 220-225, Vienna, Austria, September 23-25 1998.
- [BFG95] M. Bianchini, P. Frasconi, et M. Gori. Learning in Multilayered Networks Used as Associators. volume volume 6, No. 2, pages 512-515, March 1995.

- [Bis95] C.M. Bishop. Neural Networks for Pattern Recognition. Clarendon Press, Oxford, 1995.
- [BK88] H. Bourlard et Y. Kamp. Auto-Association by Multilayer Perceptrons and Singular Value Decomposition. Biological Cybernetics, volume 59, pages 291-294, 1988.
- [BP95] A. Boulenger et C. Pachaud. Surveillance des machines par analyse des vibrations. Du dépistage au diagnostic. AFNOR, Paris, 1995.
- [Bri89] J. S. Bridle. Probabilistic Interpretations of Feedforward Classification Network Outputs, with relationships to statistical Pattern Recognition. NATO ASI Series, Neuro-Computing : Algorithms, Architectures and Applications, vol. F68 :227-236, 1989.
- [CMZ87] G. Cottrell, P. Munro, et D. Zipser. Learning internal representations from gray-scale images : An example of extensional programming. In In Ninth Annual Conf. of the Cognitive Science Society, pages 462–473, Lawrence, Erlbaum, Hillsdale 1987.
- [Dao93] M. Daoudi. Classification interactive multidimensionnelle par les réseaux de neurones et la morphologie mathématique. PhD thesis, Université des Sciences et Technologies de Lille, 23 Novembre 1993.
- [Dév94] Espace Eolien Développement. Enérgie éolienne dans le monde. 12
  Décembre 1994.
- [EBH99] N. Elhor, R. Bertrand, et D. Hamad. Autoassociators for Supervision of Wind Energy Converters. In Proc. of the Int. Conf. on World Manufacturing Congress (WMC'99), pages 427-431, Durham, England, September 27-30, 1999.
- [EBPH] N. Elhor, R. Bertrand, J.-G. Postaire, et D. Hamad. Neural Network for Wind Turbine Supervision. ei, volume 5, No. 1, pages 123-127.
- [FGP96] A. Frosini, M. Gori, et P. Priami. A Neural Network-Based Model

for Paper Currency Recognition and Verification. volume volume7, No. 6, pages 1482-1490, November 1996.

- [GL83] G.H. Golub et C.F. Van Loan. Matrix Computations. North Oxford Academic, Oxford, 1983.
- [Hay96] S. Haykin. Neural Networks expand SP's horizons. Advenced Algorithms for Signal Processing simultaneously account for nonlinearity, nonstationarity and non-Gaussianity. IEEE Signal Processing Magazine, pages 24–49, March 1996.
- [HB95] D. Hamad et M. Betrouni. Artificial Networks for Non Linear Projection and Exploratory Data Analysis. In Proc. of the Int. Conf. on Artificial Neural Networks and Genetic Algorithms (ICANNGA'95), pages 164–167, Alès, FRANCE, Avril 18-21, 1995.
- [JJ94a] R. A. Jacobs et M. I. Jordan. A Competitive Modular Connectionist Architecture. In In Advances in Neural Information Processing Systems 3, pages 776–773. CA Morgan Kauffmann, San Mateo 1994.
- [JJ94b] M. I. Jordan et R. A. Jacobs. Hierarchical Mixtures of Experts and the EM Algorithm. Neural Computation, vol. 6 :189-214, 1994.
- [JJNH91] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, et G. E. Hinton. Adaptive Mixtures of Local Experts. Neural Computation, vol. 3, pages 79-87, 1991.
- [KA97] H. H. Kim et Y. Anzai. A Method for Task Allocation in Modular Neural Network with an Information Criterion. In Proc. of the Int. Conf. on Artificial Neural Networks and Genetic Algorithms (ICANNGA'97), pages 501-504, Norwich, ENGLAND, April 2-4, 1997.
- [Kra91] M.A. Kramer. Nonlinear Principal Component Analysis Using Autoassociative Neural Networks. AIChE Journal, volume 37, No. 2, pages 233-243, February 1991.
- [LD96] R. Lengellé et T. Denoeux. Training MLPs by Layer Using an ob-

jective Function for Internal Representations. Neural Networks, volume 9, No. 1, pages 83-97, 1996.

- [LGS94] L. Lastrucci, M. Gori, et G. Soda. Neural Autoassociators for Phoneme-Based Speaker Verification. In Proc. of the int. Symposium Workshop on Automatic Speaker Recognition, Identification and Verification, pages 189–192, April 5-7, Martigny, Suisse, 1994.
- [Mao94] J. Mao. Design and Analysis of Neural Network for Pattern Recognition. PhD thesis, Michigan State University, 1994.
- [Max97] J. Max. *Traitement du signal*. Hermes, Paris, 1997.
- [NEG+96] NORELEC, EED, Ets GROS, Univ. Littoral, et USTL IUT-A GMP. Optimisation du fonctionnement d'éoliennes par analyse de contraintes de maintenance. Rapport de fin phase 1 du programme de Recherche et Développement sur l'énergie éolienne, Décembre 1996.
- [NEG+99] NORELEC, EED, Ets GROS, Univ. Littoral, et USTL IUT-A GMP. Optimisation du fonctionnement d'éoliennes par analyse de contraintes de maintenance. Rapport de la phase 3 du programme de Recherche et Développement sur l'énergie éolienne, Novembre 1999.
- [PMA94] A. G. Parlos, J. Mathusami, et A. F. Atiya. Incipient Fault Detection and Identification in Process Systems using Accelerated Neural Network Learning. Nuclear Technology, vol. 1 :145–159, February 1994.
- [PMD<sup>+</sup>95] T. Petsche, A. Marcantonio, C. Darken, S.J. Hanson, G.M. Kuhn, et I. Santoso. A Neural Network Autoassociator for Induction Motor Failure Prediction. NIPS, MIT Press, volume 8, pages 924-930, 1995.
- [RHW86] D. E. Rumelhart, G. E. Hinton, et R. J. Williams. Learning Representations by Back-Propagating Errors. Nature, volume 332, pages 533-536, 1986.

- [ROG97] M. Rychetsky, S. Ortmann, et M. Glesner. Application of hierarchical mixture of experts networks to engine knock detection. In EUFIT'97, pages 1714–1718, Aachen, Germany, 1997.
- [Ste73] G.W. Stewart. Introduction to Matrix Computations. Academic Press, New York, 1973.
- [Thi89] S. Thiria. L'apprentissage supervisé dans les modèles connexionnistes. PhD thesis, Université RENE DESCARTES, Paris V, 1989.
- [TO95] F. Takeda et S. Omatu. High Speed Paper Currency Recognition by Neural Networks. *IEEE Trans. on Neural Network*, volume 6, No. 1, pages 73-77, March 1995.
- [WL90] A.R. Webb et D. Lowe. The Optimised Internal Representation of Multilayer CLassifier Networks Performs Nonlinear Discriminant Analysis. Neural Networks, volume 3, pages 367-375, 1990.
- [WMA<sup>+</sup>89] K. Watanabe, I. Matsuura, M. Abe, M. Kubota, et D. M. Himmelblau. Incipient Fault Diagnosis of Chimical Processes via Artificial Neural Networks. AIChE Journal, vol. 105(no. 11) :1803-1812, 1989.
- [WR94] S. R. Waterhouse et A. J. Robinson. Classification using Hierarchical Mixtures of Experts. In IEEE Int. Conf. on Neural Networks, volume vol. 1, pages 177–186, Orlando, Floride, USA 1994.
- [Zwi95] G. Zwingelstein. *Diagnostic des défaillances*. Hermes, Paris, 1995.

