



UNIVERSITE BADJI MOKHTAR ANNABA

FACULTE DE L'INGENIEUR

DEPARTEMENT DE L'ELECTRONIQUE

Mémoire

PRESENTÉ EN VUE DE L'OBTENTION DU DIPLOME DE **Magister**

Modélisation et diagnostic de défauts par l'approche MSPCA

Auteur :

Ouassim BARA

Superviseur :

Dr. Mohamed Tarek KHADIR

6 juillet 2009

Remerciements

Je tiens à présenter ma considérable gratitude à tout ceux qui m'ont aider avec le travail présenté dans ce mémoire. Tout d'abord, j'aimerais remercié le Docteur Khadir Med Tarek, mon superviseur, pour sa constante disponibilité, son assistance et ses précieux conseils.

Ce travail a été fait au sein de l'université Badji-Mokhtar Annaba.

Mes remerciements vont au Professeur Doghmane Noureddine, le directeur de la formation pour avoir accepté d'examiner ce travail et d'assurer la présidence du jury.

Je tiens à remercier le Professeur Djeghaba Messaoud, pour son aide, sa disponibilité, ses judicieux conseils. Ses remarques m'ont été d'une grande utilité dans l'avancement de mes travaux.

Je remercie messieurs les Docteurs Ramdani Messaoud et Moussaoui Abdelkrim d'avoir accepté d'examiner ce travail.

Je ne saurais oublié de présenter mes grand remerciements à mes très cher parents pour leurs encouragements tout au long de mon parcours.

Abstract

Fault detection is becoming increasingly important for process monitoring due to the higher demand for better performance, gain, as well as for increased safety and reliability of dynamic systems. In the general frame of process surveillance, the well known Principal Component Analysis (PCA) approach is chosen. PCA has been often selected due to its simplicity and ability to capture the linear relations between the stationary process variables. However the method showed limitations when dealing with industrial data that generally, present non linear and multiscale features implying events whose behavior change over time and frequency. The approach proposed in this study rests on the modeling using nonlinear PCA coupled with Artificial Neural Networks (ANNs) to extract the nonlinear inter-correlation between variables and on the wavelet analysis to decompose each sensor signal in a set of coefficients at different scales. The contribution of each variable for each scale is then collected in separated matrices and a nonlinear PCA model is constructed for each matrix. The proposed approach is applied to fault detection of pollution parameters affecting the region of Annaba in Algeria. The performance of the proposed approach is then illustrated and compared with those of classic PCA and Multiscale PCA(MSPCA) .

Résumé

La détection de défauts est devenue de plus en plus importante pour la surveillance des processus à cause de la demande croissante de la haute performance, du gain, ainsi que de la nécessité accrue de la sécurité et de la fiabilité des systèmes dynamiques. Dans le cadre général de la surveillance des systèmes, l'approche connue de l'Analyse en Composante Principale (ACP) a été choisie. L'ACP a été souvent utilisée pour sa simplicité et sa capacité à capturer les relations linéaires entre les variables du processus à l'état stationnaire. Cependant cette méthode montre des limitations à traiter les données industrielles qui généralement présentent un caractère non linéaire multi-échelle, impliquant des événements dont le comportement change dans le domaine temporel et fréquentiel. L'approche proposée dans ce travail s'appuie sur la modélisation par ACP non linéaire à l'aide des réseaux de neurones artificiels (RNAs), pour extraire l'inter-corrélation non linéaire entre les variables et de l'analyse en ondelette pour décomposer chaque signal de capteur en signaux d'approximation et de détails à différentes échelles. Les contributions de chaque variable pour chaque échelle sont alors collectées dans des matrices séparées. Un modèle ACP non linéaire est construit pour chaque matrice pour extraire la corrélation non linéaire entre les variables présentes dans chaque échelle. L'approche proposée est appliquée pour la détection de défauts de paramètres de pollutions affectant la région d'Annaba. La performance de l'approche proposée est alors illustrée et comparée avec celle de l'ACP linéaire multi-échelles et de l'ACP classique.

Table des matières

Remerciements	i
Abstract	ii
Résumé	iii
Table des figures	vii
Liste des symboles	x
1 Introduction	1
1.1 Motivations	1
1.2 Etat de l'art	2
1.2.1 Approche de diagnostic de défauts à base de modèle	2
1.2.2 Approche de diagnostic de défaut à base de données historiques	6
1.3 Objectif du Mémoire	10
1.4 Organisation du Mémoire	11
2 Analyse en ondelette	12
2.1 Introduction à la transformée en ondelette	12
2.2 Analyse multi-résolution	13
2.3 Propriétés et applications de la DWT	19
2.4 Surveillance par analyse multi-résolution	20
2.5 Combinaison des ondelettes avec l'ACP	21
3 Analyse en composante principale	22
3.1 Introduction	22
3.2 Identification du modèle ACP	24
3.2.1 Décomposition en valeur singulière	24

3.2.2	Détermination de la structure du modèle	26
3.3	Détéction du défaut	27
3.3.1	Statistique SPE et statistique T^2 de Hotelling :	27
3.3.2	Statistique SWE	29
3.4	Localisation du défaut	30
3.4.1	Localisation par calcul des contributions	30
3.4.2	Localisation par structuration des résidus	30
3.4.3	Localisation utilisant un banc de modèles	30
4	ACP non-linéaire	32
4.1	Introduction	32
4.2	ACP non linéaire	33
4.2.1	Approche neuronale de l'ACPNL	34
4.2.2	Apprentissage du réseau	35
4.2.3	Exemple illustratif	37
4.3	Détection de défaut par ACPNL	38
5	Analyse en composantes principales multi-échelle	40
5.1	Introduction	40
5.2	MSPCA non-linéaire Méthodologie	41
5.3	Procédure détaillé de l'MSPCA non linéaire :	42
6	Application de l'MSPCA aux paramètres de pollutions de l'air	44
6.1	Introduction	44
6.2	Pollution atmosphérique	45
6.2.1	Polluants primaires	45
6.2.2	Polluants secondaires	46
6.3	Phénomène de pollution	46
6.4	Prétraitement des données	47
6.5	Application de l'MSPCA	47
6.5.1	Choix des paramètres MSPCA	48
6.5.2	Détection et localisation de défauts	49
6.6	conclusion	55

7 Conclusion et Perspective	57
7.1 Conclusion	57
7.2 Perspective	58
A Apprentissage du réseau	59
A.1 Algorithme séquentiel	59
A.2 Algorithme parallèle	60
Bibliographie	60

Table des figures

1.1	Classification des méthodes de diagnostic du défaut	2
1.2	Redondance analytique et matériel	3
1.3	diagnostic de défaut à base de modèle	5
2.1	Analyse en ondelette	13
2.2	Décomposition avec l'ondelette Haar	16
2.3	Approximative décorrélation due à la transformée en ondelette dyadique	17
2.4	Décomposition et reconstruction multiéchelle	18
3.1	Analyse en composante principale (PCA)	23
4.1	comparaison entre l'ACP non-linéaire avec l'ACP	33
4.2	réseau à 5 couche pour l'extraction d'une seule composante principale	34
4.3	Mesure avec la C.P non linéaire pour un réseau à 5 couche[3-4-1-4-3]	38
4.4	Mesure et estimation	38
4.5	Erreur de reconstruction	38
4.6	Mesure et reconstruction avec défaut sur x_1	39
4.7	Erreur de reconstruction de la variable x_1	39
5.1	MSPCA non-linéaire Méthodologie	42
6.1	Illustration du cycle de Chapman	46
6.2	Donnée de pollution de l'air de la région d'Annaba	48
6.3	SPE de l'ACP dynamique avec simulation du défaut sur la variable 2 entre 100 et 400	49
6.4	Détection de défaut par l'MSPCA linéaire	50
6.5	Mesure et estimation par le modèle MSPCA linéaire	51
6.6	Mesure et estimation par le modèle MSPCA non linéaire	51
6.7	Modèle ACP non-linéaire	52

6.8	SPE : MSPCA (défaut sur la variable x_1	53
6.9	Statistique T^2 MSPCA(défaut sur la variable x_1)	53
6.10	SPE : MSPCA non linéaire (défaut sur la variable x_1	53
6.11	Graph de contribution pour detecter la cause de la déviation	53
6.12	SPE : MSPCA (défaut sur la variable x_2)	54
6.13	SWE : MSPCA (défaut sur la variable x_2)	54
6.14	MSPCA nonlinéaire (défaut sur x_2)	54
6.15	Graph de contribution pour detecter la cause de la déviation	54
6.16	SPE : MSPCA (défaut sur la variable x_3)	55
6.17	SWE : MSPCA (défaut sur la variable x_3)	55
6.18	SPE : MSPCA non linéaire (défaut sur variable x_3)	55
6.19	SWE : MSPCA (défaut sur la variable x_3)	55
6.20	SPE : MSPCA (défaut sur la variable x_4)	56
6.21	SWE : MSPCA (défaut sur la variable x_4)	56
6.22	SPE : MSPCA non linéaire(défaut sur la variable x_4)	56
6.23	SWE : MSPCA non linéaire (défaut sur la variable x_4)	56
A.1	Extraction séquentielle des composantes principales non linéaire	59
A.2	Extraction simultanée des q composantes principales non linéaires	60

Liste des symboles

α	Niveau de confiance
\hat{C}	Matrice représentant le modèle ACP
λ_i	ième valeur propre de Σ
$\Phi(t)$	Fonction d'échelle
$\Psi(t)$	Fonction d'ondelette mère
Σ	Matrice de covariance de X
$a_{m,k}$	Coefficients d'approximation de l'échelle m à l'instant k
<i>ACP</i>	Analyse en Composante Principale
$d_{m,k}$	Coefficients de détail de l'échelle m à l'instant k
db_1, db_2, \dots	Ondelette Daubechies 1, Ondelette Daubechies 2,...
<i>DWT</i>	Discret Wavelet Transform (Transformée en ondelette discrète)
G, g	Coefficients de la fonction ondelette (Filtre passe haut)
H, h	Coefficients de la fonction échelle (Filtre passe bas)
<i>MRA</i>	Multi-résolution Analysis (Analyse Multi-résolution)
<i>MSPCA</i>	Multi-scale principal component analysis
n	Nombre d'échantillon
$P \in \mathbb{R}^{p \times p}$	Matrice des composantes principale
p	Nombre de variables originales
<i>PLS</i>	Partial Least Square (Moindre Carrée Partielle)
<i>PSSE</i>	Predicted Some Squares of error
<i>SPE</i>	Erreur Quadratique d'Estimation (Squared Prediction Error)

T Scores(Matrice des valeurs des composantes principale)

T^2 Statistique de Hotelling

X Matrice des données

Chapitre 1

Introduction

Sommaire

1.1 Motivations	1
1.2 Etat de l'art	2
1.2.1 Approche de diagnostic de défauts à base de modèle	2
1.2.2 Approche de diagnostic de défaut à base de données historiques	6
1.3 Objectif du Mémoire	10
1.4 Organisation du Mémoire	11

1.1 Motivations

Les méthodes de détection et de la localisation de défauts sont devenues de plus en plus importantes pour la surveillance et la supervision des processus à cause de la demande croissante en haute performance, ainsi que de la nécessité accrue de la sécurité et de la fiabilité des systèmes dynamiques. Les processus réel sont généralement difficile à modéliser, et par conséquent à surveiller, à cause du nombre important de variables et de la complexité des relations qui lient ces dernières. L'Analyse en Composantes Principales (ACP) est l'une des méthodes statistique multi-variables la plus réponde. Bien qu'elle permette une modélisation du système d'une manière à capturer les corrélations linéaire entre ses variables, elle n'arrive pas cependant, à modéliser d'une manière efficace les variables issues des processus non stationnaire, c'est à dire des processus dont le modèle est variant dans le temps. L'utilisation de l'ACP se trouve aussi inapproprié dans le cas de la modélisation de signaux de capteurs qui sont non-linéaire, ce qui est généralement le cas dans la pratique. La plupart des méthodes existantes représentent les mesures de données sur une résolution fixe ou sur une seule échelle (échelle d'échantillonnage). Cette représentation uni-échelle est meilleure pour les données contenant des contributions avec la même localisation dans le domaine temporel et fréquentiel. Comme la plupart des processus réel sont par nature multi-échelle, ceci nous a amené à présenter l'analyse en composante principale multi-échelle non-linéaire (MSPCA non linéaire) introduite pour la première fois par

Bakshi[4]. En effet, les données peuvent être mesurées à différents taux d'échantillonnage, ou par exemple contenir des segments de données perdues. La méthode MSPCA non linéaire est intéressante du fait qu'elle extrait simultanément, l'inter-corrélation non-linéaire à travers les capteurs (ACP non-linéaire (Kramer [33])) et l'auto-corrélation à l'intérieur de chaque capteur (Analyse en ondelette[49]).

1.2 Etat de l'art

Ils existent une littérature abondante sur la diagnostique de défauts, partant des méthodes analytique jusqu'à l'intelligence artificielle, pour arriver enfin aux différentes approches statistique. Partant d'une perspective de modélisation, ils existent des méthodes qui nécessitent une modélisation précise des processus. Il existe aussi des méthodes qui ne supposent aucune forme d'information à partir de modèle et compte seulement sur les données historiques des processus. Les méthodes de diagnostique de défaut étudiées dans [56],[58] et [57] peuvent être classées en deux catégories, les méthodes à base de modèle et à base de données historique (data driven), en se basant sur une connaissance à priori du processus la hiérarchie de l'approche de diagnostic de défaut est montré sur la Figure.1.1.

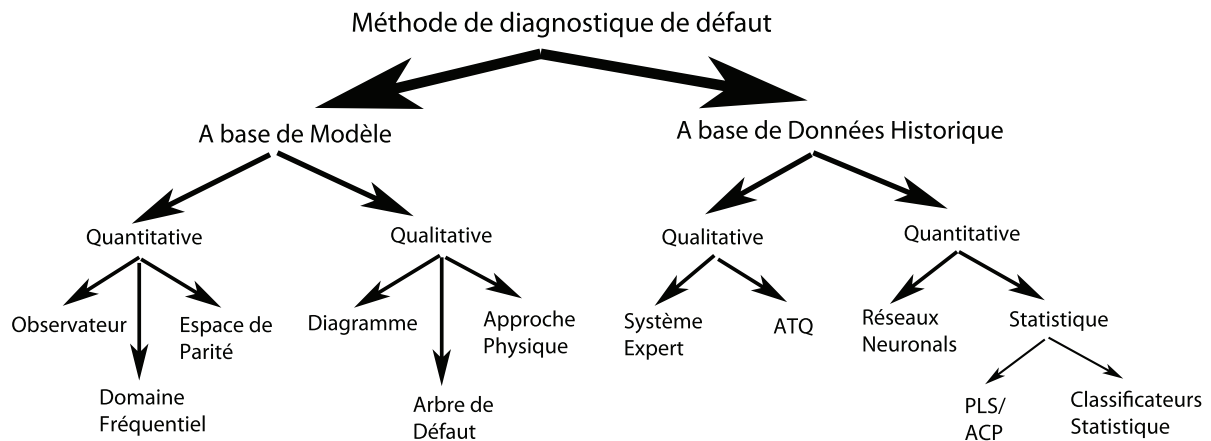


FIGURE 1.1 – Classification des méthodes de diagnostic du défaut

1.2.1 Approche de diagnostic de défauts à base de modèle

Le diagnostic de défaut à base de modèle peut être considéré comme une procédure de surveillance, de détection et de localisation du défaut dans un système, en comparant les mesures du système disponible avec une information à priori présentée par le modèle mathématique du

système.

Les procédures qui utilisent l'information du modèle pour générer des signaux additionnel, à être comparé avec les quantités de mesure original sont connue sous le nom d' approches de redondance analytique, elles sont considérées plus fiable et plus rentable que les approches alternatives connues sous redondance matériel, dans le sens que quand une mesure est disponible à partir d'un système dynamique sous contrôle , un modèle mathématique du système est utilisé pour générer les estimées d'autre mesures, par ce moyen on développe la redondance dans une forme analytique. la figure 1.2 illustre le concept de redondance analytique et matériel.

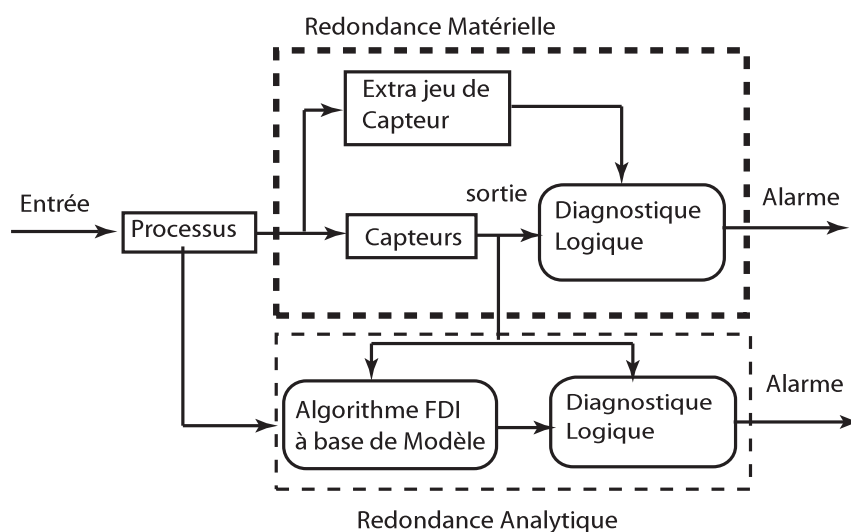


FIGURE 1.2 – Redondance analytique et matériel

D'après les terminologies générale acceptées, le diagnostic de défaut consiste en ses différentes taches :

- **Détection du défaut** : prendre une décision binaire, que ca soit pour déclarer que tous va bien ou bien que non.
- **Localisation du défaut** : pour déterminer la localisation du défaut, e.g. quelle capteur ou actionneur est en défaut.
- **Evaluation du défaut** : estimer la taille et le type ou la nature du défaut.

L'importance relative de ces trois taches est subjectivement évidente, cependant la localisation de défaut est aussi importante que la détection. L'évaluation du défaut, d'autre pars, ne peut être importante si aucune action de configuration n'est entreprise. D'ou l'abréviation FDI (Fault Detection and Isolation) détection et localisation de défaut.

1.2.1.1 Méthode quantitative à base de modèle :

La plupart des travaux sur les approches à base de modèle quantitatifs ont été basés sur l'utilisation des entrées- sorties générales et des modèles d'espace d'état pour générer des résidus.

Ces approches peuvent être classées par rapport au type d'observateurs, espace de parité et les méthodes du domaine fréquentiel.

1. **Approche à base de filtre ou d'observateur** : L'idée principale derrière l'observateur ou les approches à base de filtre est d'estimer les sorties du système à partir des mesures (ou un jeu de données) en utilisant soit les observateurs dans un environnement déterministe ou bien les filtres statistiques (e.g. Filtre de Kalman) dans un environnement stochastique. Ainsi, l'erreur d'estimation de sorties (pondéré) ou bien (innovation dans le cas stochastique) sont utilisés comme résidus. Selon les circonstances, on peut utiliser soit l'approche linéaire, non-linéaire, d'ordre réduite ou non, à observateur adaptatif ou fixe (ou filtre de Kalman)
2. **Approche d'espace de parité** : L'idée principale est de vérifier la parité (consistance) du modèle avec les sorties des capteurs (mesures) et les entrées connues du processus. L'idée de cette approche est de réarranger la structure du modèle afin d'obtenir la meilleure localisation de défaut possible. Les relations de parité dynamique ont été introduite par Willsky[1]. La redondance nous donne une liberté pour établir les équations qui génèrent les résidus pour que la localisation de défaut soit achevé. La localisation de défaut requiert l'habilité à générer les vecteurs résidus qui sont orthogonaux l'un par rapport à l'autre pour différents défauts. Gertler et al.([18],[20]) ont suggérés l'approche equation de parité orthogonal dans la conception des jeu de résidus structurés. Chow et Wilsky[13] ont proposés une procédure pour générer les équations de parités à partir de la représentation de l'espace d'état d'un système dynamique.
3. **Approche de domaine fréquentiel** : Les résidus peuvent aussi être générés dans le domaine fréquentiel via la factorisation de la fonction de transfert du système à surveiller. Une étude comprehensive à été détaillée par Ding et frank [61] dans laquelle, le problème de FDI à été formulé d'une manière méthodique et résolu par des techniques de factorisation.

Remarque 1 : La structure générale des systèmes de diagnostics de défauts à base de modèle est représenté sur la figure 1.3.

Le défaut est premièrement détecté en mettant (fixe ou variable) un seuil à la quantité résiduel générée par la différence entre les mesures réels et leurs estimées en utilisant un modèle mathématique. Ensuite viens l'étape de prise de décision, qui consiste en un simple test de seuil sur les valeurs instantanées ou bien sur la moyenne glissante des résidus, ou encore comparer les résidus à un jeu de signatures connue pour appartenir à des défauts simple.

Remarque 2 : Les techniques de génération de résidus pour le diagnostic à base de modèle est concentré sur le problème de génération des résidus puisque l'étape de prise de décision reste une étape facile à réaliser.

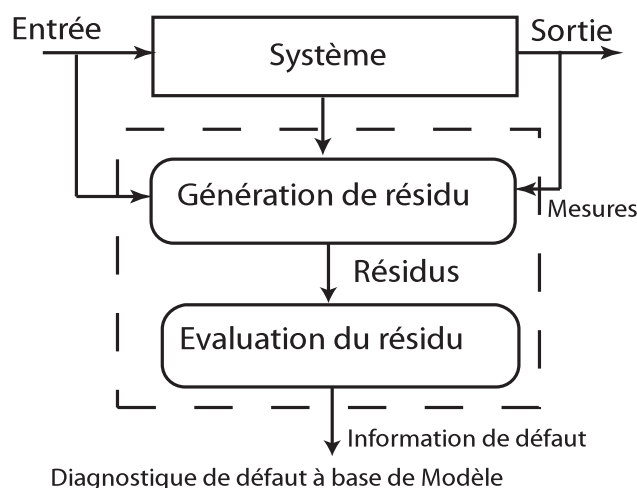


FIGURE 1.3 – diagnostic de défaut à base de modèle

On distingue les techniques de génération de résidus à base d'observateur, d'espace de parité et finalement celle qui utilisent les méthodes du domaine fréquentiel. Nous n'allons pas détailler la génération des résidus avec ces approches, puisque notre travail se base essentiellement sur l'approche de diagnostic à base de données historique du processus, et la technique d'évaluation des résidus se fera principalement à l'aide de l'analyse en composante principale.

1.2.1.2 Méthode qualitative à base de modèle

En se basant sur différentes formes des connaissances qualitatives utilisées dans le diagnostic de défauts, les approches à base de modèle qualitatifs peuvent être classées en : Diagramme, Arbre de défaut (fault tree) et les méthodes physiques qualitatifs.

1. **Approche à modèle causale utilisant les diagrammes** : Les relations cause à effet ou modèles peuvent être représentées dans la forme de diagrammes signés (SDG). Un diagramme est un graphe avec des arcs directionnel entre les noeuds et le SDG est un graphe dont lequel les arcs possèdent un signe positive ou négatif attaché à eux. Les arcs dirigés partent des noeuds *cause* pour arriver au noeud *effet*. Les SDGs sont une manière efficace pour représenter les modèles qualitatifs graphiquement et ils ont été la forme la plus largement utilisée des connaissances pour le diagnostic de défaut. Iri et al [22] ont été les premiers à utiliser les SDG pour le diagnostic de défaut.
2. **Approche à arbres de défaut** : les arbres de défauts sont utilisés pour analyser la fiabilité et la sécurité. Cette approche a été développée en premier par les Laboratoire Bell Telephone en 1961. Pour plus de de détail sur cette approche voir [23]et [48].
3. **Approches physique qualitative** : Les connaissances physiques qualitatives dans le diagnostic de défauts ont été principalement représentée de deux manière. La première est d'extraire des équations qualitatives à partir des équations différentiels connue en

tant qu'équation de confluence. Un travail considérable à été fait dans ce domaine de modélisation qualitative des systèmes et la représentation des connaissances causales [45]. Les autres approches dans la physique qualitative est la dérivation du comportement qualitative à partir des équations différentiel ordinaires (ODEs) Kuipers [2].

1.2.2 Approche de diagnostic de défaut à base de données historiques

1.2.2.1 Méthodes quantitative à base de données historiques

Les méthodes qui extraient les informations peuvent être largement classifiées en deux approches : non statistique ou statistique. Les réseaux de neurones sont une classe importante des classificateurs non statistiques. L'analyse en composante principale (ACP) /les moindres carrés partiel(PLS) et les classificateurs à base de modèle statistique forment une composante importante des méthodes d'extraction des caractéristique statistique.

1. **Approche statistique multi-variables** : Sachant que nous utiliserons dans notre travail l'approche statistique multi-variable, celle-ci sera développée en détail, on présentera ainsi, un aperçu sur les différents travaux qui ont été effectués jusqu'ici et leurs évolution.

Le contrôle des processus statistique multi-variables (MSPC) a été appliqué avec succès pour la surveillance et la détection de défauts. L'application de l'MSPC aux différents processus est significativement dépendant de la qualité des données du processus et de la capacité des différentes techniques et méthode à extraire l'information comme l'ACP et ses extensions.

Le contrôle des processus statistique multi-variables (MSPC) comme l'ACP et le PLS (Partial Least Square) ont été appliqué avec succès pour la surveillance et la détection de défauts. Une synthèse de l'utilisation de l'ACP et PLS dans le contrôle des processus et le diagnostic de défaut a été donnée par MacGregor et al [26][24]. L'analyse en composante principale (ACP) à été largement utilisée dans le contrôle statistique des processus. Dunia et al [41]. Qin et Li [52] ont utilisées l'ACP pour la détection de défauts de capteurs, l'identification et la reconstruction. Dunia et Qin[39] ont regardé l'ACP d'un point de vue géométrique et ont présenté une méthodologie qui analyse le sous espace résiduel pour la détection de défauts. Les meilleures applications de l'ACP ont été limitées à sa simplicité et fiabilité à capturer les relations linéaire entre les variables du processus à l'état stationnaire, c'est à dire que le modèle construit sera invariant dans le temps, d'ou l'idée de mettre à jours le modèle ACP d'une manière récursive. L'ACP adaptative proposé par Dayal et McGregor[6] met à jour les paramètres du modèle en appliquant une fenêtre glissante exponentiel pour ajuster le modèle aux nouvelles conditions, néanmoins il faut faire attention à ce que le modèle ne s'adapte pas aux défauts et aux perturbations du processus, aussi s'il n'y a pas assez d'excitation des données l'information du processus sera perdu puisque les données précédentes seront écartées.(Rosen et Lennox[7]).

Considérant les limitations de l'ACP conventionnel d'autres extensions ont été développées chacune en se basant sur différents aspects du processus dans le but est de modéliser le processus pratique d'une manière plus efficace, Nomikos et MacGregor[36] ont étendu l'utilisation des méthodes de projection multi-variables au processus en paquet (Batch processes) en utilisant l'ACP multiple (multiway pca). L'ACP hiérarchique ou multi-bloque permet une modélisation et une interprétation plus facile d'une grande matrice en la décomposant en petite matrice ou bloque[47]. L'ACP non linéaire s'étend à extraire les relations non linéaires parmi les variables du processus et peut capturer plus de variance avec un minimum de dimension en comparaison avec l'approche ACP classique ([9],[33]). Toutes les méthodes citées jusqu'ici opèrent sur des données collectées sur une seule échelle, cependant les données provenant de la plus part des processus pratiques sont par nature multi-échelle et ceci due aux différentes contributions d'événements se produisant à différentes échelles. bakshi[4] :

- Événements se produisant à différents endroits et avec différentes localisations dans le domaine temporel et fréquentiel.
- Processus stochastique dont l'énergie ou le spectre de puissance change avec le temps ou la fréquence.
- Variables mesurées à différents taux d'échantillonnages ou contenant des données perdus.

L'analyse multirésolution permet une décomposition temps échelles, ce qui augmente par conséquent la sensibilité de la surveillance, en la rendant plus sensible à détecter les petits changements dans des données présentant de grands changements. Parfois l'information temps échelle est primordial pour l'interprétation de la cause physique du défaut. Ces dernières années, il y a eu un intérêt considérable de combiner l'ACP conventionnel avec la transformée en ondelette (Bakshi [4]; Kosanovich et Piovoso [27]; Misra et al [31]; Rosen et Lennox [7]). Dans le but d'améliorer les performances de la surveillance et de détecter différents événements à différentes échelles, un nombre important d'approches de contrôle statistique des processus multi-variables multi-échelle ont été proposées. Ils combinent un modèle de projection multi-variables avec l'analyse multi-résolution. Ces méthodes de surveillance des processus basés sur l'ACP utilisent l'analyse en ondelette pour transformer le signal du domaine temporel dans le domaine temps-fréquence. Piovoso [27] et Bakshi [4] ont été parmi les premiers à proposer la méthode de l'ACP multi-échelle (MSPCA), qui construit des modèles ACP pour chaque échelle d'ondelette. Les échelles où des événements significatifs se produisent seront reconstruites pour obtenir le modèle ACP global. D.Wang [10] a proposé une approche robuste de l'ACP multi-échelle en introduisant le principe d'incertitude de la distribution du bruit du signal. Misra et al. [31] ont démontrés l'efficacité de l'approche MSPCA par rapport à celle de l'ACP classique.

Rossen et Lennox [7] ont introduit le principe de l'ACP adaptative à l'approche multi-échelle et ont démontré son efficacité par rapport à l'approche ACP classique et ont noté aussi, qu'en décomposant les données, le problème du changement des conditions du processus est partiellement résolu. Ding et al [11] ont utilisées des mesures de concentration de gaz et ont appliqué un algorithme pour la détection en ligne de la dérive des capteurs appelé Algorithme Dynamique et Adaptative pour la Compensation de la Dérive (ADDC). L'algorithme est basé sur la combinaison de l'ACP et de l'analyse en ondelette.

Toutes les approches qu'on a vu jusqu'à maintenant traite seulement des données linéaire, sachant que la plus part des données pratique sont non-linéaire, des études récentes ont été proposé Shao et al.[43] qui utilisent un algorithme d'optimisation en ligne des entrées du réseau de neurone (Input Training Network), les coefficients d'ondelette sont utilisés comme entree de l'algorithme, des limites de contrôle non parametrique sont appliqué sur les graphe de contrôle pour la détection de défauts. Choi et Morris [53] ont proposé une approche multi-échelle non linéaire qui utilise le principe de l'ACP noyau (kernel PCA). Nous présenterons dans ce travail, pour la modélisation non linéaire l'approche ACP non linéaire à l'aide des réseaux de neurones.

L'approche ACPNL Multi-échelle proposée dans ce Mémoire est une variante de celle de Bakshi [4]. Notamment par rapport à :

- La localisation de la variable en défaut.
- Introduction de la modélisation par ACP non linéaire à l'aide des réseaux de neurones (kramer [33]).

2. **Approche à classifieur statistique :** Le diagnostic de défaut est essentiellement un problème de classification, ainsi il peut être modelé dans une structure de reconnaissance de modèle statistique classique. Il peut aussi être considéré comme un problème qui combine, à travers le temps, les estimations instantanées du classifieur en utilisant la connaissance à propos des propriétés statistique du mode en défaut du système.
3. **Approche à réseaux de neurones :** Un intérêt considérable a été montré dans la littérature à l'application des réseaux de neurone dans le diagnostic de défaut. En général, les réseaux de neurone qui ont été utilisées pour le diagnostic de faut peuvent être classifié à travers deux dimension : l'architecture du réseau et la stratégie d'apprentissage comme l'apprentissage supervisé et non supervisé. La stratégie la plus connue dans l'apprentissage supervisé des réseaux de neurones a été l'algorithme de rétro-propagation. Différentes architectures de réseaux ont été utilisées pour le problème du diagnostic de défaut. Par exemple Bakshi et Stephanopoulos proposèrent les Wavenet un réseau de neurone hiérarchique multi-résolution, Scholz [32] à développer aussi une approche d'ACP non linéaire hiérarchique.

1.2.2.2 Méthode qualitative à base de modèle

On distingue parmi les méthodes majeures pour extraire les informations historique qualitative, les système expert et l'analyse à tendance qualitative.

1. **Approche par système expert :** Un système expert est généralement un système très spécialisé qui résout des problèmes dans un domaine restreint d'expertise, les méthodes par système expert semblent particulièrement adaptées lorsque les connaissances sur le système sont de type heuristique. C'est à dire lorsque les informations utilisées (qualitative ou quantitative) permettent l'utilisation de règles imbriquées issues de la connaissance et de l'expérience de l'expert, come par exemple des liens de causes à effets. Le principe de base d'un système expert s'articule autour de deux éléments principaux, une base de connaissance et un moteur d'inférence. Cette approche est particulièrement intéressante pour le diagnostic car elle permet de manipuler un grand nombre de données non homogènes et indépendantes du contexte tout en rendant compte du raisonnement suivi.
2. **Analyse à tendance qualitative :** L'analyse à tendance et la prédiction sont d'importante composante pour la surveillance et le contrôle des processus. La modélisation par tendance peut être utilisée pour expliquer différents évènements qui se produisent dans le processus, pour la prédiction de défaut de fonctionnement et pour prédire les états futurs. Cheung et Stephanopoulos [25] ont construit un cadre conventionnel pour la représentation des processus à tendance.

1.3 Objectif du Mémoire

Le but de ce Mémoire est de développer une méthode de contrôle statistique multi-variables qui permet d'améliorer la vitesse de détection des événements anormaux à différentes échelles ainsi que la réduction du nombre de fausses alarmes. Plusieurs approches de ce type, qui combinent une projection multi-variables linéaire du modèle avec l'analyse multi-résolution ont été proposées dans la littérature. Notre travail se base sur une approche non-linéaire multi-échelles qui utilise les réseaux de neurones auto-associatif avec l'analyse en ondelette pour examiner les performances des processus multi-variables multi-échelle dans le domaine de la surveillance et en particulier dans la détection et la localisation du défaut dans les processus non-linéaire. Les questions qui sont adressées dans ce Mémoire sont :

- Comment améliorer la détection et la localisation du défaut dans les processus réel ?
- Comment déterminer les meilleures relations qui lie les différentes variables du système ?
- Quelles est la meilleure méthode qui offre les meilleures performances dans le domaine de la surveillance(FDI :Fault Detection and Isolation) ?

l'analyse en composantes principales non linéaire multi-échelle possède l'avantage d'extraire les corrélation linéaire et non-linéaire à partir des données du processus pour obtenir une représentation réduite des données, avec l'aide de l'analyse en ondelette pour améliorer la sensibilité de détection en décomposant le signal original sur différentes échelles. Les coefficients d'ondelettes resultants sont approximativement décorrélés, par conséquent la formulation de l'MSPCA non-linéaire lui donne la possibilité de travailler avec des données qui sont typiquement non-stationnaire et qui represente l'effet cumulatif de plusieurs evenements, tous se produisant à différentes échelle.

La validité et la crédibilité des données fournies par les différents capteurs issues d'une ou plusieurs stations de surveillances de la qualité de l'air, représente la tache principale de l'approche qu'on a développer dans ce travail. La modification de l'information réel que dois fournir un capteur influe considérablement sur les décisions prisent par l'opérateur ou par le système de traitement de l'information, notamment dans les choix pris pour la réduction des émissions.

1.4 Organisation du Mémoire

Ce travail est organisé comme suit :

Le chapitre 1 concerne l'introduction. Il présente les motivations avec un aperçu sur les différents travaux reportés dans la littérature et enfin l'objectif du mémoire avec une vue globale sur les chapitres présentés.

Le deuxième chapitre présente l'Analyse en ondelette à travers l'introduction de la transformée en ondelette et de l'analyse multi-résolution. L'introduction de l'analyse en ondelette est importante puisqu'elle sera utilisée dans la procédure de l'Analyse en composante principale multi-échelle non linéaire que nous allons introduire plus tard.

Le principe de l'Analyse en Composante Principale sera introduit dans le chapitre 3, ainsi que la modélisation par ACP, qui consiste à estimer les paramètres du modèle par le calcul des valeurs et vecteurs propres de la matrice de corrélation des données, et la détermination de la structure du modèle par la détermination du nombre de composantes à retenir par le principe du pourcentage de la variance cumulé.

Une extension de l'Analyse en composante principale pour le cas non linéaire est introduite dans le chapitre 4 à l'aide des réseaux de neurones auto-associatif à 5 couches (AAN). Un exemple illustratif sera présenté pour valider la modélisation par ACPNL. Un exemple pour la détection de défaut sera aussi présenté.

Le chapitre 5 introduit le principe de l'Analyse en composante principale multi-échelle (MSPCA). La méthodologie et la procédure détaillée de l'MSPCA non linéaire seront présentées.

Le chapitre 6 sera consacré à l'application de l'MSPCA dans le cadre de la détection et localisation de défaut sur des paramètres de pollution de l'air de la région d'Annaba avec une comparaison des différentes approches, ACP conventionnel, MSPCA et MSPCA non linéaire. Beaucoup d'activités humaine produisent des polluants primaires comme les oxydes d'azotes (NO_2 et NO), le dioxyde de soufre et les composés organique volatiles(COV) qui forment, dans la basse atmosphère, par des réactions chimiques ou photochimiques, des polluants secondaire comme l'ozone(O_3). Un certain nombre de ces polluants sont susceptibles de poser des problèmes pour la santé humaine et les systèmes écologiques. L'étude présentée concerne quatre polluants qui sont le NO_x et les oxydes d'azotes NO_2 et NO avec l'ozone.

Les résultats obtenus sont discutés avec une conclusions générale et des perspectives.

Chapitre 2

Analyse en ondelette

Sommaire

2.1	Introduction à la transformée en ondelette	12
2.2	Analyse multi-résolution	13
2.3	Propriétés et applications de la DWT	19
2.4	Surveillance par analyse multi-résolution	20
2.5	Combinaison des ondelettes avec l'ACP	21

2.1 Introduction à la transformée en ondelette

Pour faire face à l'inconvénient majeur de la transformée de Fourier à savoir de traiter seulement les signaux stationnaires étant donné qu'on perd complètement l'information en temps après avoir transformé notre signal dans le domaine fréquentiel. La transformée en ondelette est apparue en apportant une projection du signal du domaine temporel au plan temps-échelle. Comme la transformée de Fourier, elle projette le signal sur des fonctions de base, mais cette fois-ci le signal sera projeté sur une fonction ondelette mère $\Psi_{s,u}(t)$ échelonnée et décalée dans le temps. Cette approche permettra d'introduire la notion de multi-résolution. Une famille de fonctions d'ondelettes est ainsi créée avec la même forme que l'ondelette mère mais avec des tailles et des localisations différentes. La famille d'ondelette peut être définie comme suit :

$$\Psi(t) = \frac{1}{\sqrt{|s|}} \Psi\left(\frac{t-u}{s}\right) \quad (2.1)$$

s et u sont respectivement les paramètres de dilatation et de translation. Dans le cas discret, ces paramètres sont données comme suit $s = 2^m$ et $u = 2^m k$, $(m, k) \in \mathbb{Z}^2$. Une dilatation et translation dite dyadique est opérées. Cependant on peut ne pas se restreindre à ces valeurs. On utilisera dans notre travail une transformée en ondelette dyadique.

2.2 Analyse multi-résolution

La transformée en ondelette nous donne le moyen d'analyser un signal dans un nombre de niveau de différentes résolutions d'une manière hiérarchique connue en tant que Analyse Multi-résolution (Multi Resolution Analysis(MRA)). Ainsi, le signal mesuré en série temporel et qui contient différentes formes physiques ou perturbations, e.g. changement soudain aux niveaux les plus fin et des changements lent dans les grande échelles, peut être mieux représenté en le considérons comme une combinaison de plusieurs signaux à différents niveau de résolution.

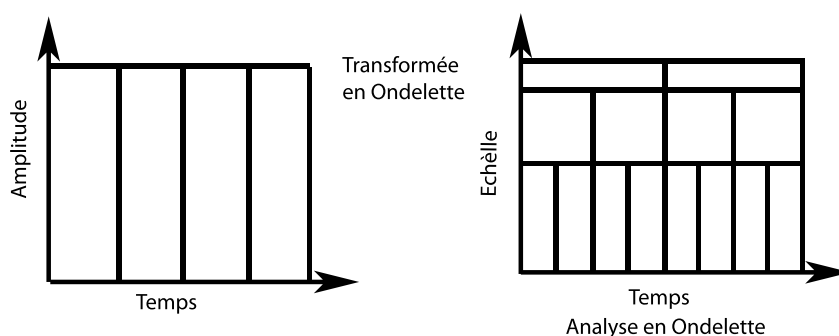


FIGURE 2.1 – Analyse en ondelette

L'AMR est conçu pour donner

- une bonne résolution en temps et une faible résolution en fréquence aux hautes fréquences.
- une bonne résolution en fréquence et une faible en temps aux basses fréquence.

Il faut savoir que les hautes fréquences vont correspondre aux petites échelles (low scales) et les basses fréquences aux plus grandes échelles (high scales).

L'analyse en ondelette permet l'utilisation de long intervalles temporel où l'on recherche l'information basse fréquence plus précisément, et des région plus courte où on recherche l'information haute fréquence (Voir Figure 2.1). Cette approche a un sens particulièrement quand le signal a des composantes à haute fréquence pour des courtes durées et des composantes de basse fréquences pour de longues durées. Les signaux réels sont souvent de ce type.

Il s'avère que le choix de l'échelle et de la position sur la base de puissance de deux (dyadique) permettra une analyse plus efficace ([4], [17]). On obtient une telle analyse avec la DWT(Discret Wavelet Analysis). Une façon efficace d'implémenter ce schéma en utilisant les bands de filtre a été développer par Mallat[49] dans sa théorie de l'analyse multi-résolution, qui di que n'importe quel signal peut être décomposé en le projetant sur des fonctions échelon et des fonctions d'ondelette. L'analyse multi-resolution utilise des algorithmes de décomposition et de reconstruction pour obtenir une représentation multi-échelle du signal par une series de filtrages avec des opérations de sous-échantillonnages. L'algorithme récursif pour la décomposition en ondelette

est aussi connu sous le nom d'algorithme pyramide ou bien transformée en ondelette rapide. $\Phi_{mk}(t)$ et $\Psi_{mk}(t)$ sont respectivement les versions translatés et dilaté des ondelettes mère de la fonction d'échelle (scaling function) $\Phi(t)$ et de la fonction ondelette $\Psi(t)$. m et k sont respectivement, les paramètres de dilatation et de translation. Le paramètre de translation détermine la localisation de l'ondelette dans le domaine temporel, tandis que le paramètre de dilatation détermine la localisation de celle-ci dans le domaine fréquentiel. Ces choix de s et de u permettent aux ondelettes d'être orthonormales [17], mais cause un sous-échantillonnage (reduction ou décimation dyadique) dans le nombre des coefficients de l'ondelette d'une échelle à une autre.

Remarque La discrétisation dyadique permet l'utilisation d'ondelette orthonormale. L'orthonormalité des fonctions de base si elle est satisfaite permet d'avoir des coefficients d'ondelette approximativement dé-corrélés, mais introduit un retard dans le calcul pour le cas de la surveillance en ligne. Si le paramètre de translation est discretisé uniformément, alors il n'y a pas sous échantillonnage des coefficients. Cela permet d'avoir une décomposition en ligne sans retard mais dans ce cas les ondelettes sont non orthonormales, et les coefficients sont autocorrélés.

$$\Phi_{mk}(t) = 2^{-\frac{m}{2}} \Phi(2^{-m}t - k) \quad (2.2)$$

$$\Psi_{mk}(t) = 2^{-\frac{m}{2}} \Psi(2^{-m}t - k) \quad (2.3)$$

La projection sur la fonction d'échelle (scaling function), i.e. le produit du signal avec $\Phi_{mk}(t)$ est connue comme le coefficient échelon ou bien coefficient d'approximation a_{mk} . Les coefficients d'approximation ne sont que des version lissés du signal original et le degré de lissage augmente avec l'échelle, par contre la projection sur la fonction d'ondelette, i.e. convoluer le signal avec $\Psi_{mk}(t)$ est connue comme étant les coefficients d'ondelette d_{mk} . Les coefficients d'ondelette aussi connue sous le nom de détails du signal perdue en passant d'une échelle à une autre plus grande.

$$a_{m,k} = \langle f, \Phi_{mk} \rangle = \int f(t) \Phi_{mk}(t) dt \quad (2.4)$$

$$d_{m,k} = \langle f, \Psi_{mk} \rangle = \int f(t) \Psi_{mk}(t) dt \quad (2.5)$$

On peut écrire les équations(2.4) et(2.5) sous la forme discrète suivante :

$$a(m) = \sum_{k=-\infty}^{k=\infty} x(k)h(2m - k) \quad (2.6)$$

$$d(m) = \sum_{k=-\infty}^{k=\infty} x(k)g(2m - k) \quad (2.7)$$

h et g représente respectivement les coefficients de la fonction d'échelle (filtre pass bas) et les coefficients de la fonction ondelette (filtre pass haut).

On peut aussi noter que n'importe quel signal peut être décomposé en ces contributions de plusieurs échelles comme une somme pondérée de fonctions de bases orthonormal discrétisés dyadiquement.

$$x(t) = \sum_{m=1}^L \sum_{k=1}^N d_{mk} \psi_{mk}(t) + \sum_{k=1}^N a_{Lk} \phi_{Lk} \quad (2.8)$$

Un algorithme récursif rapide, qui représente la projection du signal sur les fonctions de bases correspondante connue comme une opération de filtrage, donne un calcul plus efficace pour la décomposition en ondelettes et la reconstruction. Il y a trois composantes principales impliqués dans la transformée en ondelette rapide, le filtre passe bas H , le filtre passe haut G , et la décimation dyadique.

Le résultat est un jeu de coefficients qui décrivent les détails du signal et un autre jeu décrivant l'approximation. La décomposition peut être exécutée jusqu'à un nombre désiré d'échelles en appliquant récursivement le filtre passe haut et passe bas aux coefficients d'approximations du niveau précédent. De cette manière, le vecteur du signal original est lissé et divisé à travers le filtre H , et les coefficients d'approximation résultants sont encore lissés et divisés à travers le filtre H . L'application successive du filtre passe bas résulte en une version plus lisse du signal original. En même temps que le signal est lissé, chaque itération extrait l'information haute fréquence des données.

Les différentes étapes pour appliquer la DWT pour un niveau de décomposition $L = 3$ sont décrites sur la Figure 2.4(a).

Pour illustrer le principe de la décomposition en ondelette on définit le signal suivant :

$$x(t) = \begin{cases} w(t), & 0 \leq t < 100 \\ w(t) + 3, & 100 \leq t \leq 150 \\ w(t), & t > 150 \end{cases} \quad (2.9)$$

Il est présenté sur la Figure 2.2, où $w(t)$ est un bruit blanc avec une moyenne zéro et une variance unité. Dans ce cas, le signal est décomposé jusqu'à une échelle $L = 3$. Cette décomposition est basée sur l'ondelette Haar et sa fonction d'échelle correspondante qui sont respectivement définis par

$$\psi(t) = \begin{cases} 1, & 0 \leq t < \frac{1}{2} \\ -1, & \frac{1}{2} \leq t < 1 \\ 0, & \text{ailleurs} \end{cases} \quad (2.10)$$

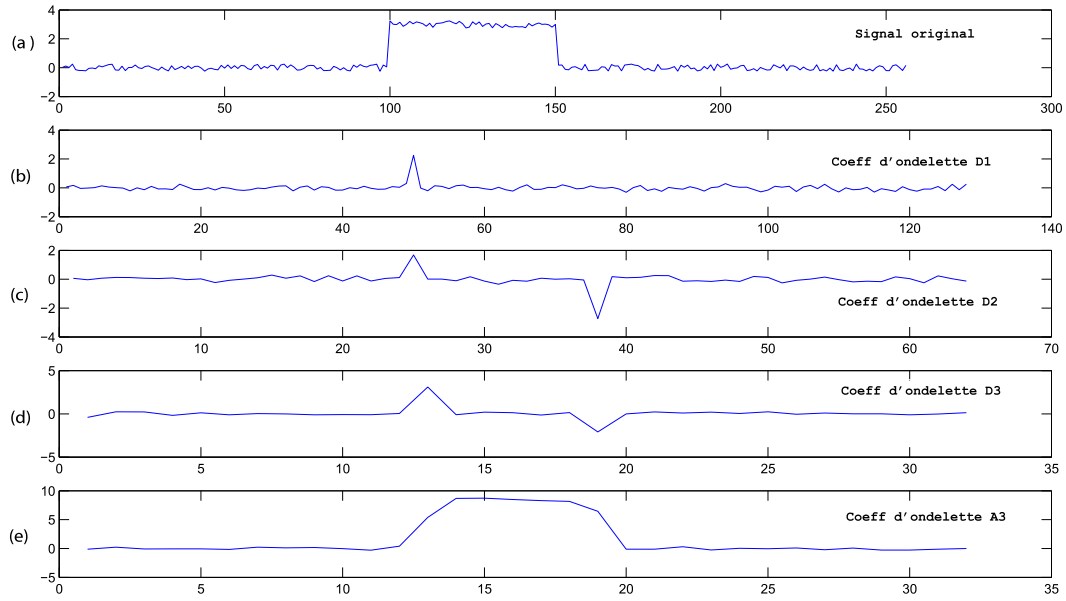


FIGURE 2.2 – Décomposition avec l'ondelette Haar

$$\phi(t) = \begin{cases} 1, & 0 \leq t < 1 \\ 0, & \text{ailleurs} \end{cases} \quad (2.11)$$

Le contenu basse fréquence du signal est présenté sur un jeu de coefficients de la fonction d'échelle, comme décrit sur la Figure 2.2(e). Le nombre de coefficients de la fonction ondelette et de la fonction d'échelle diminue aux grandes échelles due à la discrétisation dyadique des paramètres de dilatation et de translation.

Il faut savoir que le nombre total de coefficients dans le domaine d'ondelette est le même que le nombre d'échantillon dans le signal original. Pour notre exemple, les coefficients du filtre de la fonction d'échelle et de la fonction ondelette sont respectivement, $H = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$ et $G = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$. Ainsi les coefficients à différentes échelles peuvent être obtenus comme suit :

$$a_m = H a_{m-1} \quad d_m = G a_{m-1} \quad (2.12)$$

avec d_m le vecteur des coefficients d'ondelette à l'échelle m (Figure 2.2(b),(c),(d), pour $m=1,2,3$), et a_m est le vecteur des coefficients de la fonction d'échelle (scaling function). Les données originales sont considérées comme étant les coefficients d'ondelette à l'échelle la plus fine c-à-d, $x = a_0$.

L'Equation 2.12 peut aussi être représentée en terme du vecteur des données originaux, x , comme :

$$a_m = H_m x, \quad d_m = G_m x \quad (2.13)$$

Où H_m est obtenu en appliquant le filtre $H(m)$ fois et G est obtenu en appliquant le filtre $H(m-1)$ fois et le filtre G une seule fois. Les données de départ peuvent être exactement reconstruites à partir des coefficients d'ondelettes pour chaque échelles, d_m pour $m = 1, 2, \dots, L$, et des coefficients de la fonction d'ondelette a_L . (Voir Figure 2.4)

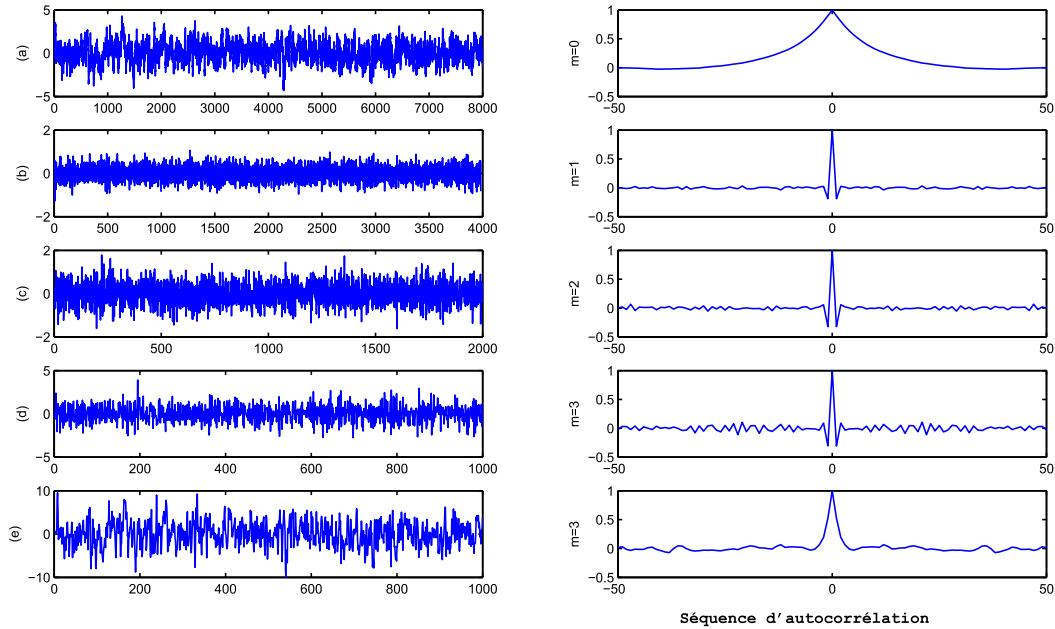


FIGURE 2.3 – Approximative décorrelation due à la transformée en ondelette dyadique

Une propriété importante des ondelettes est que malgré qu'elles ne sont pas connues pour être la forme exacte de fonctions propres ou bien de composantes principales d'opérateur quelconque, elles sont l'approximation en fonction propre d'une large variété d'opérateurs [44]. Par conséquent, les coefficients d'ondelettes de la plupart des processus stochastique sont approximativement décorrelés.

La capacité des ondelette à décorreléer les données est décrite sur la Figure 2.3. Le signal original, $x_k = 0.9x_{k-1} + 0.5w_k$ simule un processus stochastique stationnaire, ou w_k est un bruit blanc de moyenne zéro et de variance unité. Sa fonction d'autocorrélation est montré sur la Fig.2.3(a). Les coefficients d'ondelette ainsi que les fonction d'auto-corrélation correspondante sont montrés sur Fig.2.3(b), (c), et (d). Ils indiquent que les coefficients sont approximativement décorrelés. Le dernier signal échelonné dans la Fig.2.3(e) contient de la corrélation résiduelle, qui peut être réduite en décomposant à des échelles plus grandes. La variance des coefficients d'ondelette à chaque échelle varie en fonction du spectre de puissance du signal original. Ainsi la variance des coefficients d'ondelette et le dernier signal échelonné sur la Fig.2.3(b), (c), (d), (e), augmente avec la diminution de la fréquence.

Les coefficients d'ondelette contiennent toujours l'information portée par le signal original. Pour chaque filtre passe haut et passe bas, il existe leurs filtres de reconstruction cor-

respondant, voir sur la Figure 2.4(b) \hat{G} et \hat{H} . En utilisant ces filtres pour la reconstruction, avec le sur-échantillonnage correspondant et en sachant que les filtres sont biorthogonaux, ceci résultera en une reconstruction parfaite du signal original uni-échelle. Alternativement il est possible de reconstruire séparément, au domaine temporel, les composantes correspondantes à chaque échelle d'après l'algorithme décrit sur l'équation 2.13 et comme on peut le voir sur la Figure 2.4(b). Après seulement avoir opéré une reconstruction de chaque coefficient au domaine temporel, le signal original pour l'exemple de la Figure 2.4 est donné par l'égalité suivante : $X = A_3 + D_3 + D_2 + D_1$.

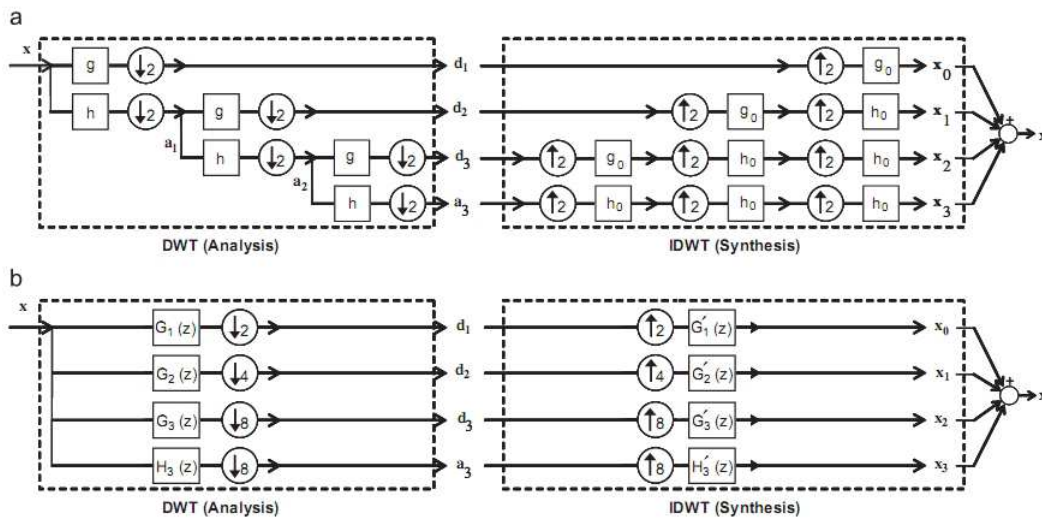


FIGURE 2.4 – Décomposition et reconstruction multi-échelle

La Figure 2.4(a) présente clairement la décomposition multi-échelle (Transformée en Ondelette Discrète : DWT) sur deux niveaux du signal x ainsi que l'étape de reconstruction (Transformée en Ondelette Discrète Inverse : IDWT), Figure 2.4(b)

L'algorithme de la transformée en ondelette discrète est généralement restreint à une utilisation hors-ligne [5]. Due au fait que les filtres d'ondelette sont de nature non-causale puisqu'il requiert des mesures de données futures pour calculer les coefficients d'ondelette récentes [5], un retard est introduit dans le calcul. Une autre raison pour le restreindre à une utilisation hors-ligne est la discrétisation dyadique des paramètres de l'ondelette, qui requiert un signal de longueur dyadique pour la décomposition en ondelette. Pour des raisons de surveillances et de détection des capteurs en dérive, une approche de transformée en ondelette discrète en ligne est étendue en se référant à Bakshi[5], cet algorithme est résumé comme suit :

1. Décomposer les données de mesurées à l'intérieur d'une fenêtre d'une longueur dyadique en utilisant l'ondelette de Haar.
2. Retenir seulement les point de données des coefficients d'approximation pour un usage

ultérieur

3. Quand de nouvelles mesures sont disponibles, déplacer la fenêtre dans le temps pour inclure les mesures les plus récentes pendant que la même taille de la fenêtre dyadique est maintenu.

Un choix inhérent pour ce type d'utilisation désigne l'ondelette Haar puisque son filtre correspondant est de taille 2. Quand la fenêtre glissante est de taille dyadique, l'ondelette Haar est adoptée pour effectuer une décomposition en ondelette avec aucun retard.

2.3 Propriétés et applications de la DWT

Nous donnerons dans cette section un aperçu sur les propriétés les plus importantes des ondelettes ainsi que leurs applications dans différents domaines. Une transformation en ondelette convertit des données du domaine d'origine au domaine d'ondelette, en projetant ces dernières sur des fonctions de bases orthonormées générées par la translation et la dilatation d'une fonction d'ondelette père et une fonction mère. Les propriétés des ondelettes sont décrites comme suit :

1. Complexité de calcul : La transformée en ondelette rapide basée sur l'algorithme pyramidal de Mallat [49] à besoin seulement de $O(N)$ multiplications.
2. Moment de fuite (Vanishing moment) : Chaque ondelette a un certain nombre de moments zéro ou moment de fuite égaux au nombre de coefficients. Par exemple db_1 (ondelette Haar) a un moment de fuite, db_2 (Daubechies 2) possède 2 moments de fuites. Le moment de fuite se rapporte à la capacité de l'ondelette de représenter le comportement polynomial ou l'information dans un signal. Par exemple db_1 avec un moment encode des polynômes d'un coefficient. db_2 encode des polynômes de deux coefficients.
3. Support compact : Chaque fonction de base d'ondelette est supportée sur un intervalle fini. Par exemple, le support de la fonction Haar, comme on l'a vu dans l'exemple précédent est $[0, 1]$; le support de l'ondelette db_2 est $[0, 3]$. Un support compact garantit la localisation de l'ondelette. En d'autres termes, le traitement d'une région des données avec l'ondelette n'affecte pas le reste des données.
4. Coefficients décorrélés : Un autre aspect important des ondelettes est leur capacité à réduire la corrélation temporelle d'une manière que la corrélation des coefficients d'ondelettes soit moins importante que la corrélation du processus temporel. Ceci a été démontré sur la Section 2.2 (Voir Figure 2.3). Par conséquent, la transformée en ondelette pourrait être utilisée pour réduire la complexité du processus dans le domaine temporel en un processus plus simple dans le domaine de l'ondelette.

Les applications des ondelettes sont innombrables. Elles couvrent une variété immense d'applications d'engineering, parmi lesquelles nous citons :

- Le débruitage (Denoising) : la procédure de débruitage se base sur trois étape importante à savoir :
 1. Décomposition : Choix de l'ondelette w et du niveau de décomposition L
 2. Seuillage des coefficients de détails : Pour chaque niveau de décomposition $1 \rightarrow L$, sélectionner le seuil en appliquant le seuillage hard ou le seuillage soft.
 3. reconstruction : Calculer la reconstruction par ondelette, en utilisant le coefficient d'approximation original avec les coefficients de détails modifiés.
- La compression de données (Data compression) : La notion derrière la compression est basée sur le concept que le signal original peut être approximé avec précision en utilisant peu de coefficients. La procédure de compression est quasiment similaire à celle du débruitage à la seule différence du choix du seuil des coefficients d'ondelette.
- Estimation de fonction (regression, Densité)
- Exploration de données(Data mining)
- Traitement d'image : e.g extraction de contour.

et bien d'autres application qu'on peut pas tous citer.

2.4 Surveillance par analyse multi-résolution

Pour des raisons de surveillances et de détection, les ondelettes peuvent être utilisés comme nous venons de le voir pour décomposer le signal en différents signaux temps-échelles avec diminution du niveau de détail ou de résolution. Un signal est filtré avec un filtre passe haut et un filtre passe bas, respectivement. Le résultat est un jeu de coefficients qui décrit les détails du signal et un autre jeu qui décrit l'approximation. La décomposition peut être exécutée jusqu'à un nombre désiré d'échelles en appliquant récursivement les filtres passe haut et passe bas aux coefficients d'approximation du niveau précédent. Quand un filtre à été appliqué, le résultat est sous-échantillonné. Ce qui veut dire que le nombre total de coefficient dans le domaine de l'ondelette est le même que le nombre d'échantillon dans le domaine temporel. Les coefficients contiennent toujours toute l'information porté par le signal original.

Les techniques de contrôle statistique des processus uni-variable peuvent être utilisées pour localiser les échelles dont lesquelles une perturbation est détectée. Les perturbations rapides sont détectées dans les petites échelles alors que les perturbations lentes sont détectées dans les grandes échelles. La façon dont la perturbation apparait à travers les échelles peut révéler des informations concernant les caractéristiques du défaut. Un échelon apparaitra clairement sur toutes les échelles, alors qu'une rampe apparaîtra sur les échelles inférieures. Faut savoir qu'une augmentation de la sensibilité mène vers plus de détections, de plus les évènements qui se produisent dans différentes échelles doivent être évalué sur chaque échelle même celle qui

proviennent d'un événement unique. Tout ceci avec le fait que toutes les variables doivent être contrôlées sur différentes échelles, appelle à utiliser les méthodes multi-variables qui sont non seulement capable de réduire la dimension du problème mais aussi modéliser les relations entre les variables. Un choix naturel est l'ACP.

2.5 Combinaison des ondelettes avec l'ACP

Au lieu d'appliquer l'ACP sur les données dans le domaine temporel, Kosanovich et Piovoso[27] ont appliqué l'ACP sur les coefficients d'ondelettes pour la surveillance des processus. Bakshi[4] a développé la formulation de l'MSPCA et a démontré son efficacité pour la surveillance des processus multivariés. L'idée principale qui permet d'utiliser conjointement l'ACP et l'ondelette découle de l'idée que dans l'ACP, la corrélation entre les capteurs est utilisée pour transformer l'espace multivariés vers un sous-espace qui préserve le maximum de variance de l'espace original. Cependant, l'ACP n'arrive pas à traiter la corrélation à l'intérieur de chaque capteur le long du temps. En d'autres termes elle n'utilise pas l'information appartenant à la fréquence ou aux caractéristiques de l'échelle de chaque capteur. Les Ondelette, d'autre part, arrive à capturer la corrélation à l'intérieur de chaque capteur. Plus clairement, la modélisation par ACP permet d'extraire l'intercorrélation entre les variables alors que l'analyse en ondelette permet d'extraire l'autocorrélation présente dans chaque capteur. Celle-ci arrive approximativement, grâce aux caractéristiques de l'ondelette (Voir Fig.2.3) à décorrélérer les coefficients de l'ondelette de la plupart des processus stochastiques. Ainsi, les Ondelette et l'ACP représentent pour l'analyse des données multivariés deux approches complémentaires.

Le chapitre suivant présentera l'approche de l'Analyse en Composante Principale.

Chapitre 3

Analyse en composante principale

Sommaire

3.1	Introduction	22
3.2	Identification du modèle ACP	24
3.2.1	Décomposition en valeur singulière	24
3.2.2	Détermination de la structure du modèle	26
3.3	Détéction du défaut	27
3.3.1	Statistique SPE et statistique T^2 de Hotelling :	27
3.3.2	Statistique SWE	29
3.4	Localisation du défaut	30
3.4.1	Localisation par calcul des contributions	30
3.4.2	Localisation par structuration des résidus	30
3.4.3	Localisation utilisant un banc de modèles	30

3.1 Introduction

La plupart des processus réels présentent des données corrélées, cette corrélation est utilisée par l'Analyse en Composante Principale (ACP) dans ce travail pour un but de modélisation du système à surveiller. La comparaison entre le comportement réel du système et le comportement attendu donné par le modèle fournit une quantité appelée résidu qui va servir à déterminer si le système est dans un état défaillant ou non. L'ACP utilise aussi la corrélation entre les variables pour transformer l'espace multi-variable en un autre sous-espace qui préserve le maximum de variance des données de départ dans un nombre minimum de dimension. Ceci est illustré pour trois dimension dans la Figure 3.1. L'ACP transforme le vecteur $x = (x_1, x_2, \dots, x_p)^T$ de dimension p en un vecteur généralement de dimension plus réduite $y = (y_1, y_2, \dots, y_q)^T$, où p est le nombre de variables et q est le nombre de composantes choisis. La transformation linéaire

de l'ACP est donné par la matrice P de dimension $p \times p$ tel que :

$$T = P.X \quad (3.1)$$

Chaque ligne du vecteur de la matrice y contient les valeurs (scores) d'une nouvelle variable y_j connue sous le nom de composante principale (PC). La composante PC j , donnée par la nouvelle variable $y_j = (y_{j1}, y_{j2}, \dots, y_{jn})$ est une combinaison de toutes les variables originales $x_i = (x_{i1}, x_{i2}, \dots, x_{in})$, pondérée par les éléments du vecteur $v_j = (v_{j1}, v_{j2}, \dots, v_{jp})$ de la transformation correspondante.

$$y_j = \sum_{i=1}^p v_{ji}x_i = v_{j1}x_1 + v_{j2}x_2 + \dots + v_{jp}x_p \quad (3.2)$$

n est le nombre d'échantillon et p est le nombre de variables originales. Les poids v_{ji} (souvent nommé vecteur de chargements (loadings) nous donne la contribution de toutes les variables originales x_i à la j^{eme} composante. Géométriquement, l'ACP est équivalent à une rotation de l'espace original des données. Les nouveaux axes sont les composantes principales (Pcs). Le vecteur v_j donne la direction de la j^{eme} composantes principales (PC j) dans l'espace des données originales. La première composante, PC1, représenté par la variable y_1 , est dans la direction de la plus haute variance. La seconde composante, PC2, est la direction qui maximises le reste de la variance dans le sous-espace orthogonale complemenaire à la première composante. La première et la seconde composante, ensemble, explique le plan bi-dimensionnelle de la plus haute variance. Ceci peut être naturellement étendu pour obtenir les q composantes principales. Le vecteur colonne $y = (y_1, y_2, \dots, y_q)^T$ contient les nouvelles q coordonnées dans l'espace des composantes principales de la matrice de données x

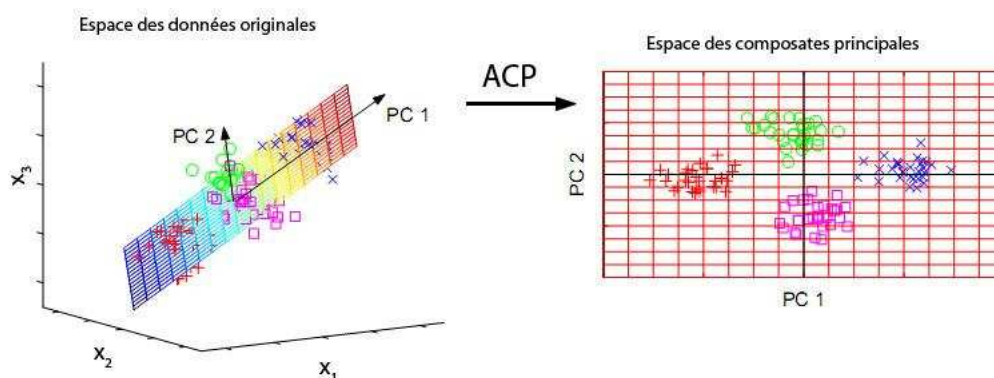


FIGURE 3.1 – Analyse en composante principale (PCA)

La Figure 3.1 illustre la transformation de l'ACP qui transforme un grand nombre de variables à un nombre plus réduit nommé composante principale(PCs). Trois échantillon de variables sont projetées sur l'espace de composantes à deux dimensions qui maintient la plus grande

variance dans les données. Cette visualisation bi-dimensionnelles des échantillons nous permet de faire des conclusions qualitatives sur la séparabilité de nos quatre conditions d'expérimentations. L'espace des composantes explique un sous-espace linéaire, où les données s'étalent sur, ou près de ce dernier, de l'espace original de dimension plus grande. L'ACP opère simplement une rotation de l'espace des données originales de telle sorte que les composantes principales (Pcs) sont les axes des nouvelles coordonnées du système.

3.2 Identification du modèle ACP

3.2.1 Décomposition en valeur singulière

soit $X \in \mathbb{R}^{p \times n}$ est une matrice de p variables avec chacune n échantillon qui peut être décomposer comme suit :

$$X = U\Lambda V^T = \lambda_1 u_1 v_1^T + \lambda_2 u_2 v_2^T + \cdots + \lambda_p u_p v_p^T \quad (3.3)$$

où U contient les vecteurs singulier gauche $(u_i, i = 1, 2, \dots, p)$, V contient les vecteurs singulier droit $(v_i, i = 1, 2, \dots, p)$, et λ_i sont les valeurs singulière.

$U_{(n \times p)} = [u_1 u_2 \cdots u_p]$, $V_{(p \times p)} = [v_1 v_2 \cdots v_p]$ sont des matrices orthonormale.

$\Lambda_{(p \times p)} = \text{diag}[\lambda_1, \lambda_2, \dots, \lambda_p]$ est une matrice diagonale non négatif réel. Ses éléments diagonales sont les valeurs singulière $\lambda_i, i = 1, 2, \dots, p$ qui sont les valeurs absolue de la racine carrée des valeurs propres de la matrice $X^T X$.

Les colonnes de U , $u_i, i = 1, 2, \dots, p$ sont les vecteurs propres normalisés de XX^T .

$$XX^T U = U\Lambda^2 = [u_1 u_2 \cdots u_p] \cdot \text{diag}[\lambda_1^2, \lambda_2^2, \dots, \lambda_p^2] = [\lambda_1^2 u_1, \lambda_2^2 u_2 \cdots \lambda_p^2 u_p].$$

Les colonnes de V , $v_i, i = 1, 2, \dots, p$ sont les vecteurs propres normalisés de la matrice $X^T X$.

$$X^T X V = V\Lambda^2 = [v_1 v_2 \cdots v_p] \cdot \text{diag}[\lambda_1^2, \lambda_2^2, \dots, \lambda_p^2] = [\lambda_1^2 v_1, \lambda_2^2 v_2 \cdots \lambda_p^2 v_p].$$

Remarque 2 : La matrice de chargement de l'équation 3.1 peut être obtenu à partir de l'équation 3.3 si on prend $P = V$. et $T = U\Lambda$.

Pour comparer les variables de différentes amplitudes et de différentes variabilités les données sont centrées et réduites. La matrice X permettra d'obtenir à partir de la matrice de corrélation les vecteur et valeur propre, respectivement la matrice P et λ_i . La matrice des valeurs de composantes principales T (Scores) est donnée par l'expression suivante :

$$\hat{T} = X P_q \in \mathbb{R}^{q \times n} \quad (3.4)$$

La matrice P_q contient seulement les q vecteurs propres (On verra dans la section suivante comment choisir le nombre q de composantes qui détermine la structure du modèle ACP) obtenu à partir du calcul des valeurs et vecteurs propres de la matrice de corrélation Σ tel que Σ est défini par l'équation suivante :

$$\begin{aligned}\Sigma &= \frac{X^T X}{N-1} \\ \Sigma &= P \Lambda P\end{aligned}\tag{3.5}$$

La contribution de l'erreur dans la matrice de données peut être réduite en éliminant les composantes principales (vecteurs propres) de la matrice de chargement correspondantes aux plus petites valeurs propres et en reconstruisant la matrice des données. La matrice estimée sera :

$$\begin{aligned}\hat{X} &= T_q P_q^T \\ \hat{X} &= X P_q P_q^T \in S_p\end{aligned}\tag{3.6}$$

\hat{X} Représente la projection des données sur le sous-espace des composantes principales S_p . Le modèle ACP sera défini par la matrice suivante :

$$\hat{C} = P_q P_q^T\tag{3.7}$$

L'erreur d'estimation sera donnée comme suit :

$$\begin{aligned}E_r &= X - \hat{X} \\ E_r &= X(I - P_q P_q^T) \in S_r\end{aligned}\tag{3.8}$$

Le modèle ACP peut être établi en se basant sur les données historique quand le processus est à l'état stationnaire. Une nouvelle observation multi-variable x_{new} peut être projetée sur le sous-espace des composantes principales défini par la matrice de chargement pour obtenir le vecteur des valeurs des composantes principales Scores ($t_{new} = P_q^T x_{new}$), et sur l'espace résiduel ($e_{new} = x_{new} - \hat{x}_{new}$), $\hat{x}_{new} = P_q t_{new}$, ou P_q représente les premières q colonnes de la matrice de chargement qui définit la projection des composantes principales. Une interprétation géométrique détaillée de l'ACP est donnée dans la section .

Deux mesures statistiques importantes, la statistique T^2 et la SPE (Squared Prediction Error, aussi connue comme la statistique Q), sont utilisées pour la surveillance des processus. On reviendra pour définir ces statistiques dans la section 3.3. Les limites de contrôle de ces deux mesures statistiques sont calculées sur la base des données historique. Les nouvelles observations sont alors vérifiées par rapport aux limites de contrôles pour détecter les événements anormaux. Au lieu de surveiller plusieurs variables en même temps, seulement deux variables sont nécessaire pour la surveillance statistique par ACP. Une alarme sera enclenchée si par hasard les variables

surveillées dépassent les limites de contrôles.

3.2.2 Détermination de la structure du modèle

Déterminer la structure du modèle revient à choisir le nombre adéquat de composantes principales. Si peu de composantes principales sont choisies que celles requises, un modèle pauvre en résultera avec une représentation incomplète du processus. Au contraire si beaucoup de composantes principales sont retenues, le modèle sera sur-paramétré et pourra inclure les composantes correspondantes aux plus petites valeurs propres et qui sont porteuses de bruit.

Ils existent plusieurs méthodes concernant la détermination du nombre de composantes principales, parmi elles le pourcentage cumulé de la variance totale (PCV), critère de validation croisée, critère de la moyenne des valeurs propres qui prend en considération que les composantes pour lesquelles la valeur propre est supérieure à la moyenne arithmétique de toutes les valeurs propres, et le critère de la variance non reconstruite.

3.2.2.1 Pourcentage cumulé de la variance totale (PCV)

Le PCV est la mesure du pourcentage de la variance capturée par les l_k composantes principales :

$$PCV(q) = 100 \left(\frac{\sum_{i=1}^q \lambda_i}{\sum_{i=1}^p \lambda_i} \right) \% \quad (3.9)$$

avec p le nombre de variables et λ_i est la i^{eme} valeur propre de la matrice de corrélation Σ ou de la matrice de corrélation Σ de la matrice de données X . On peut choisir le $PCV=90\%$. e.g. Le critère de PCV est quelque peu arbitraire et a tendance à sélectionner un grand nombre de composantes principales.

3.2.2.2 Moyenne des valeurs propres

Cette méthode consiste à ne prendre en considération que les composantes pour lesquelles la valeur propre est supérieure à la moyenne arithmétique de toutes les valeurs propres. En particulier, si on travaille sur les données centrées réduites, cela revient à négliger les composantes dont la variance est inférieure à l'unité ($\frac{1}{p}traces(\Sigma) = 1$). Dans le cas du modèle ACP calculé à partir de la matrice de covariance Σ , la moyenne arithmétique des valeurs propres est donnée par $\frac{1}{p}traces(\Sigma)$.

3.2.2.3 Critère de validation croisé

L'objectif de l'analyse en composant principal est de construire un modèle qui doit être employé pour le diagnostic d'un système, le critère suggéré pour identifier le nombre approprié de composantes est la validation croisée. La validation croisée est une technique très populaire et simple mais comportant de nombreux calculs. Cette procédure de validation croisée est basée

sur la minimisation de la quantité PSSE. Cette quantité se présente comme la somme des carrés des erreurs entre les données observées et celles prédites ou estimées par le modèle obtenu à partir d'un jeu d'identification différent.

$$PRESS(q) = \frac{1}{Np} \sum_{k=1}^N \sum_{i=1}^p (\hat{x}^{(q)}(k) - x_i(k))^2 \quad (3.10)$$

N étant la taille du jeu de validation.

3.2.2.4 Nombre de composantes pour une meilleure reconstruction

Le principe de reconstruction consiste à estimer une des variables du vecteur $x(k)$ à un instant donné, notée $x_i(k)$, en utilisant toutes les autres variables $x_j(k)$ au même instant à partir du modèle ACP déjà obtenu. Il existe trois approches différentes de reconstruction qui aboutissent exactement à la même solution.

3.3 Détection du défaut

Le principe de détection de défaut est de modéliser le comportement normal du processus en utilisant dans notre cas l'ACP, après de détecter le défaut en mettant en référence le comportement observé contre celui du modèle. La détection du défaut consiste après l'évaluation du résidu, à identifier l'échantillon qui dépasse une certaine limite de contrôle, il est intéressant de faire suivre l'étape de détection par la localisation de la variable à l'origine du défaut, on présentera dans la sous-Section 3.4 les graphes de localisation du capteur en défaut.

3.3.1 Statistique SPE et statistique T^2 de Hotelling :

Généralement on distingue dans les processus deux anomalies :

- Pour le premier type les relations de base entre les variables des processus ne changent pas, mais plusieurs variables subissent des changements. Pour ce type d'anomalie le SPE demeure petit, mais les valeurs des composantes principales (Scores T), se déplace en dehors de la région dont le système a été développé.
- Pour le second type d'anomalie, les relations entre les variables du processus changent, ce qui résultera en un changement du SPE, parce que les nouvelles mesures ne peuvent plus être expliquées par le modèle ACP.

La statistique de Hotelling T^2 est donnée par l'équation

$$T^2 = \hat{t}^T \Lambda_k^{-1} \hat{t} \quad (3.11)$$

$$T^2 = \sum_{i=1}^q \frac{t_i^2}{\Lambda_i^2}$$

avec t_i , ($i = 1, 2, \dots, q$), est le i^{eme} composant du vecteur $t = P^T x$, s_{ti}^2 est la variance de t_i , qui est aussi la i^{eme} valeur propre de la matrice de corrélation Σ et q biensure est le nombre de composantes principales. La statistique T^2 de hotelling donne une mesure de la distance d'un échantillon à la moyenne du processus à l'intérieur du plan défini par les composantes principales. La statistique T^2 peut seulement détecter les variations à l'intérieur du sous-espace des composantes principales qui sont plus grandes que celles qui peuvent être expliquées par le modèle ACP. Quand un évènement se produit qui n'était pas présent dans le modèle ACP, la nouvelle observation se déplacera en dehors de l'espace des CPs. Ce type d'évènement peut être détecté par la statistique Q qui mesure les variations dans l'espace résiduel. La statistique Q ou SPE est défini comme suit :

$$Q = e^T e = (x - \hat{x})^T (x - \hat{x}) \quad (3.12)$$

où $e = (I - P_q P_q^T)x$ est le résidu de l'échantillon x . Ceci représente la distance carrée de chaque observation perpendiculaire au sous-espace des CPs.

Remarque3 : Les deux statistiques T^2 et Q sont complémentaire l'un et l'autres et ensemble elles peuvent mesurer les variations dans l'espace de mesures entier.

Détermination de la limite de controle pour la statistique T^2

La quantité T^2 suit une distribution du Chi-2(χ^2) avec q degrés de liberté où $\Lambda_q = diag(\lambda_1, \lambda_2, \dots, \lambda_q)$ est une matrice diagonale qui contient les q plus grandes valeurs propres de la matrice de corrélation. Le seuil approprié pour la statistique T^2 pour un seuil de confiance α peut être déterminé.

La statistique $\frac{N(N-q)}{q(N+1)(N-1)} T^2$ suit une distribution de Fisher $F_{q,(N-q)}$. La limite supérieure pour un seuil de confiance α est obtenue à partir de $F_{q,(N-q),\alpha}$, soit :

$$\chi_{q,\alpha}^2 = \frac{q(N+1)(N-1)}{(N(N-q))} F_{q,(N-q),\alpha} \quad (3.13)$$

Le processus est supposé en défaut (déviation anormale), à l'instant k , si :

$$T^2(k) > \chi_{q,\alpha}^2 \quad (3.14)$$

Détermination de la limite de contrôle pour la statistique Q :

Le processus est considéré en fonctionnement anormal à l'instant k si :

$$Q(k) > \delta_\alpha^2 \quad (3.15)$$

Soit x une observation d'une population multi-normale $N(0, \Sigma)$ et λ_i , $i = 1, 2, \dots, p$ sont les valeurs propres de la matrice Σ , alors l'approximation de la limite de contrôle pour la statistique

Q tel que $Q = e^T e$ est donnée par :

$$\delta_\alpha^2 = \theta_1 \left[\frac{c_\alpha \sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]^{\frac{1}{h_0}} \quad (3.16)$$

Soit $\theta_i = \sum_{p=j}^{j=q+1} \lambda_j^i$ pour $i = 1, 2, 3$ et λ_j est la j^{eme} valeur propre de la matrice Σ .

$$\text{avec } h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2} \text{ et } c_\alpha = \frac{\left[\left(\frac{\|e\|^2}{\theta_1} \right) - 1 - \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]}{\sqrt{2\theta_2 h_0^2}}.$$

On peut d'après Box [15] écrire la limite de détection sous la forme suivante :

$$\delta_\alpha^2 = g \chi_{h,\alpha}^2 \quad (3.17)$$

avec $g = \frac{\theta_2}{\theta_1}$ et $h = \frac{\theta_2^2}{\theta_1}$. Nomikos et McGregor ont démontré que les deux approximations données par les équations (3.16) et (3.17) sont équivalente.

3.3.2 Statistique SWE

L'indice de détection SWE (Squared Weighted Error), plus sensible au défaut, peut être défini comme étant le SPE pondéré par l'inverse de la variance des dernières composantes principales, et n'est en fait que la statistique T^2 appliquée aux dernières composantes. Le SWE peut être défini par les deux équations équivalentes suivantes :

$$\begin{aligned} SWE &= \tilde{t}^T \Lambda_{p-q}^{-1} \tilde{t} \\ SWE &= \sum_{i=q+1}^p \frac{t_i^2}{\Lambda_{ti}^2} \end{aligned} \quad (3.18)$$

où $\Lambda_{p-k} = \text{diag}(\lambda_{q+1}, \dots, \lambda_p)$ est une matrice diagonale contenant les $(p - q)$ dernières valeurs propres de la matrice de corrélation Σ .

le SWE suit une distribution du Chi-2 avec $(p - q)$ degrés de liberté :

$$SWE = \tilde{t}^T \Lambda_{p-q}^{-1} \tilde{t}. \quad (3.19)$$

Cette statistique suit une loi du $\chi_{(p-q),\alpha}^2$. Le processus est supposé en défaut si :

$$SWE > \chi_{(p-q),\alpha}^2 \quad (3.20)$$

Remarque : Il est à noter qu'aucune supposition sur la distribution des variables du processus n'est imposée, par contre on suppose que les résidus $\tilde{t}(k) = \tilde{P}^T x(k)$ sont des bruits blancs gaussiens.

3.4 Localisation du défaut

Quand une observation se déplace en dehors des limites de contrôles, on suppose qu'un évènement inhabituel s'est produit sur le processus ou bien qu'un défaut de fonctionnement sur un équipement a eu lieu. Il est par conséquent impératif d'identifier la ou les variables, mises en causes, plusieurs approches pour la localisation du défaut ont été développées dans la littérature, on note la localisation par structuration des résidus[18], localisation par approche utilisant les banc de modèles[37], la localisation par ACP partielles[62], localisation par élimination[8], localisation par reconstruction et localisation par calcul des contributions. On commencera par expliquer cette dernière approche sachant qu'elles sera utilisée dans ce travail.

3.4.1 Localisation par calcul des contributions

Les graphes de contributions comme sont nom l'indique détermine la contribution de chaque variables par rapport à l'indice de détection, la variable ayant la plus forte contribution est considérée comme la variable en défaut. Cette approche a été largement utilisée dans l'ACP[43]. Dans le cas de l'indice de detection SPE , la contribution $cont_j^{SPE}(k)$ de la j^{eme} variable à l'instant k est définie par l'équation suivante :

$$cont_j^{SPE}(k) = (e_j(k))^2 = (x_j(k) - \hat{x}_j(k))^2 \quad (3.21)$$

où $x_j(k)$ est j^{eme} élément du vecteur de mesure x . Localisation par structuration des résidus, localisation utilisant un banc de modèles, localisation par ACP partielle, localisation par élimination, localisation basé sur le principe de reconstruction, localisation par ACPNL partielle.

3.4.2 Localisation par structuration des résidus

Cette approche s'inspire des méthodes de diagnostic utilisant la redondance analytique. L'évaluation de ces résidus permet de générer un vecteur de cohérence ou de signature de défauts représenté sous forme binaire suite à un test logique des résidus par rapport à leurs seuils respectifs. Ces approches de localisation comparent la signature expérimentale et un ensemble de signatures théoriques également binaires. La localisation peut reposer soit sur le calcul de distance entre la signature expérimentale et les différentes signatures théoriques, soit sur une projection de la signature expérimentale sur les différentes signatures théoriques. Ainsi on cherche à transformer les résidus primaires en résidus secondaires ayant les propriétés de localisation recherchées.

3.4.3 Localisation utilisant un banc de modèles

La localisation utilisant un banc de modèles dans le cas de l'ACP, exploite le même principe utilisé avec des observateurs de type GOS (Generalized Observer Scheme) où le i^{eme} observateur est piloté par toutes les entrées (sorties), sauf la i^{eme} et toutes les sorties (entrées). La sortie de

cet observateur est donc sensible aux défauts de toutes les entrées (sorties) sauf ceux de la i^{eme} . Ainsi on retrouve trois approches

3.4.3.1 Localisation par ACP partielles

Une autre façon de structurer des résidus est d'utiliser ce que l'on appelle l'ACP partielle [62]. L'ACP offre une autre possibilité qui consiste à utiliser des modèles partiels. On sous-entend par ACP partielle une ACP effectuée sur un vecteur de données réduit où quelques variables sont écartées par rapport au vecteur original. Donc, les résidus vont être sensibles uniquement aux défauts associés aux variables qui forment le vecteur réduit. Les défauts associés aux variables éliminées n'affecteront pas les résidus correspondants. L'idée de l'ACP partielle a été décrite en premier par Gertler [19]. La procédure consiste à structurer les indices de détection en calculant les ACP partielles ainsi que les seuils de détection des indices correspondants.

3.4.3.2 Localisation par élimination

Cette approche a été proposée par Stork et al. [8]. La philosophie derrière cette méthode est similaire à celle de Dunia et al. [41]. Dans le cas où un défaut est détecté avec la statistique SPE, un ensemble de modèles ACP est généré à partir du modèle global en éliminant à chaque fois une variable de l'ensemble des variables à surveiller.

3.4.3.3 Localisation basée sur le principe de reconstruction

Cette méthode basée sur le principe de reconstruction [41] consiste à suspecter qu'un capteur est défaillant et à reconstruire la valeur de sa mesure en se basant sur le modèle ACP déjà calculé et les mesures des autres capteurs. La procédure est répétée pour chaque capteur. La localisation est effectuée par comparaison de l'indice de détection avant et après reconstruction.

Chapitre 4

ACP non-linéaire

Sommaire

4.1	Introduction	32
4.2	ACP non linéaire	33
4.2.1	Approche neuronale de l'ACPNL	34
4.2.2	Apprentissage du réseau	35
4.2.3	Exemple illustratif	37
4.3	Détection de défaut par ACPNL	38

4.1 Introduction

La complexité et la dimensionnalité élevée des processus réel et les limites des approches à base de modèle a facilité l'introduction des approches qui utilise l'analyse de données, entre autre l'approche ACP que nous avons vu dans le chapitre 3. Cependant, l'ACP est une méthode de projection linéaire, ainsi seules les dépendances linéaire entre les variables peuvent être détectées. Comme la plupart des données réel sont non-linéaire plusieurs extensions de l'ACP ont été proposées, ACPNL basée sur les courbes principales Hastie[54]. Cette approche peut seulement offrir un tracé non-linéaire des variables du processus aux composantes principales non-linéaire, mais ne peut pas offrir un chemin inverse utilisé pour produire la valeur prédite d'un nouvel échantillon. Donc cette méthode ne peut être utilisée pour la surveillance. En combinant l'algorithme des courbes principales et le réseau de neurone auto-assocative, Dong et McAvoy[9] proposèrent un algorithme ACPNL. les limites de leur approche est que l'algorithme des courbes principales suppose que les fonction non-linéaire peuvent être approximées en une combinaison linéaire de plusieurs fonctions univariable. Cette projection par conséquent sera restreinte à identifier des structure de variables limités. Tan et Mavrovouniotis[46] ont proposé une ACPNL en utilisant un réseau de neurones à trois couches(IT-net) et dont l'apprentissage est effectué en minimisant à la fois les poids et les entrées du réseau. Schölkopf et al. [3] ont introduit une approche non neuronale connue comme l'ACP noyau et qui utilise des fonctions

noyau non-linéaire pour construire le modèle ACPNL. Kramer[33] a été le premier à proposer l'ACP non linéaire en utilisant les réseaux de neurones auto-associatif à 5 couches(AAN).

Dans ce chapitre nous allons présenter la généralisation de l'ACP linéaire dans le cas non linéaire, dans un premier temps, le principe de l'analyse en composantes principale est présenté, Il faut noter que dans notre travail on s'intéresse pas à la structure du réseau étudié. A la fin de ce chapitre nous présenterons un exemple non linéaire pour illustrer le principe de l'ACPNL.

4.2 ACP non linéaire

L'ACP non-linéaire est une extension de l'ACP linéaire, tandis que l'ACP identifie les corrélations linéaires entre les variables, l'ACP non-linéaire permet d'extraire à la fois les corrélations linéaires (second ordre) et non-linéaires (ordre supérieur); ceci est effectué en projetant les variables du processus sur des courbes et des surfaces au lieu de les projeter sur des lignes et des plans (Voir Figure 4.1), tout en minimisant l'erreur quadratique moyenne $E \left\{ \|X - \hat{X}\| \right\}$. La matrice X est en terme de q composantes principales non-linéaire, avec $q \ll p$.

$$X = \Phi_{gen}(T) + E \quad (4.1)$$

$$T = \Phi_{extr}(X) \quad (4.2)$$

La matrice T représente la matrice des valeurs de composantes principales non-linéaire(scores), Φ_{gen} est une fonction non linéaire équivalente à la matrice de projection P dans l'ACP linéaire, E est la matrice résiduel.

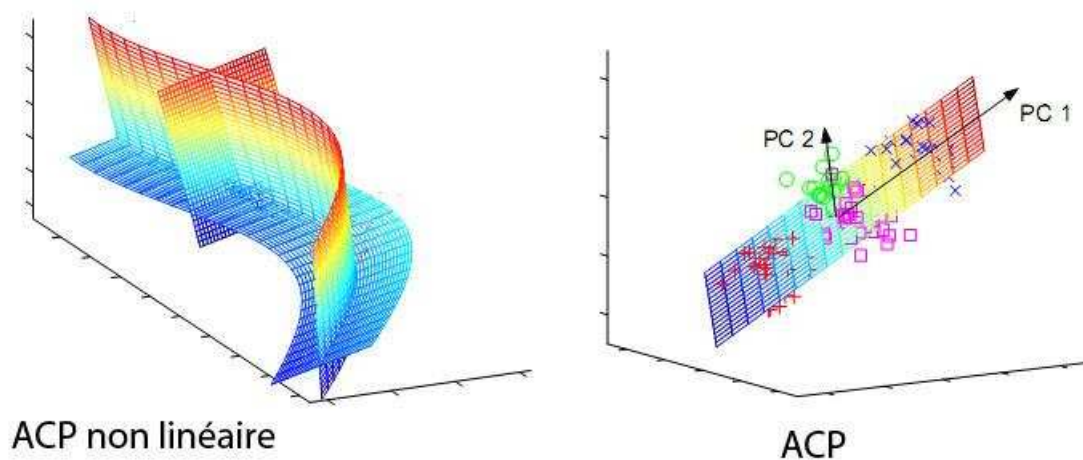


FIGURE 4.1 – comparaison entre l'ACP non-linéaire avec l'ACP

4.2.1 Approche neuronale de l'ACPNL

L'ACP nonlinéaire proposé par Kramer [33] est basé sur le Pércéptron Multi-Couches(PMC) avec une topologie auto-associative constitué de 5 couches avec une couche d'etraglement qui execute une projection identité. Le réseau en réalité est composé de deux partie : la première represente la fonction d'extraction non-linéaire $\Phi_{extr} : \mathcal{X} \rightarrow \mathcal{Z}$ qui permet d'exécuter une projection non-linéaire de X sur la première couche cachée pour obtenir la matrice T (scores)valeurs des composantes principales non linéaire, représenté par z_1 sur la Figure 4.2 alors que le seconde partie represente la fonction non-linéaire inverse, de génération ou fonction de reconstruction, $\Phi_{gen} : \mathcal{Z} \rightarrow \hat{\mathcal{X}}$, qui permet à partir de la matrice T de reconstruire les données originels.

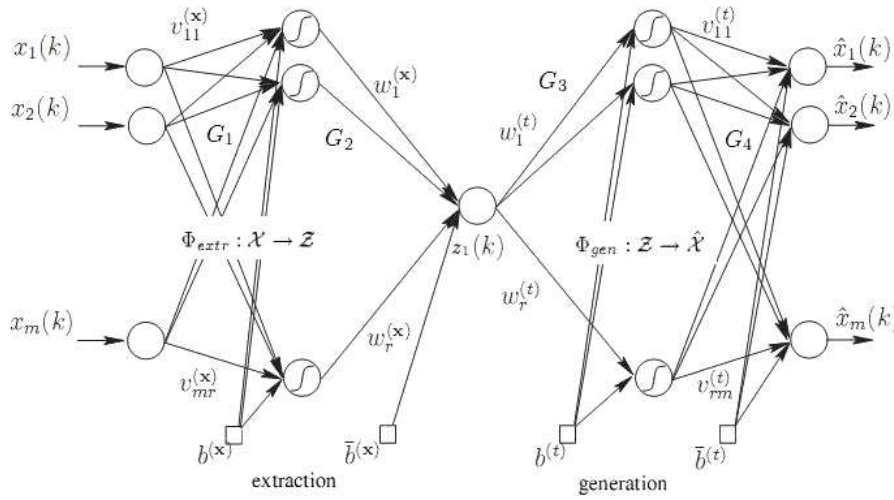


FIGURE 4.2 – réseau à 5 couche pour l'extraction d'une seule composante principale

La Figure 4.2 nous montre la sortie du réseau \hat{x} qui est calculée pour être égale à l'entrée x en minimisant l'erreur quadratique. Les échantillons x sont compressé(projetées) sur le nombre de composantes principales représenté par un seul neurone dans cet exemple z ceci est montré sur la partie extraction, la partie inverse de génération reconstruit \hat{x} à partir de z . L'échantillon \hat{x} est généralement une représentation de x réduite en bruit. La seconde et la quatrième couche cachées avec chacune leurs unité non-linéaire, permettent au réseau d'exécuter des projections non-linéaires. Le réseau peut être étendu à extraire plus d'une composante en utilisant des noeuds additionnels dans la couche d'etraglement.

Pour effectuer l'ACPNL, le réseau contient dans la partie extraction trois couches cachées. Une fonction de transfert G_1 réalise une projection du vecteur x , vecteur d'entrée de dimension m , vers la première couche cachée(couche de codage), représentée par $h_j^{(x)}$ ($j = 1, \dots, r$), un vecteur colonne de dimension r ou r représente le nombre de neurones dans la première couche cachée :

$$h_j^{(x)} = G_1 \left(\left(V^{(x)} x + b^{(x)} \right) \right) \quad (4.3)$$

$V^{(x)}$ est la matrice des poids de dimension $(r \times m)$, $b^{(x)}$ un vecteur contenant les r paramètres

de biais. La deuxième fonction de transfert G_2 projette les données de la première couche cachée vers la couche d'étranglement "bottleneck layer" contenant un seul neurone, qui représente la composante principale non linéaire z . La fonction de transfert G_1 est généralement non linéaire (on utilise la fonction tangente hyperbolique ou la fonction sigmoïde), tandis que la fonction G_2 représente la fonction identité ($G_2(x) = x$) :

$$z = G_2 \left(w^{(x)} h^{(x)} + \bar{b}^{(x)} \right) = \sum_{j=1}^r w_j^{(x)} h_j^{(x)} + \bar{b}^{(x)} \quad (4.4)$$

Ensuite, la fonction de transfert G_3 , qui est une fonction non linéaire, projette les données à partir de z vers la dernière couche cachée (couche de décodage) $h_j^{(t)}$ ($j = 1; \dots, r$) et r représente le nombre de neurone dans la troisième couche cachée :

$$h_j^{(t)} = G_3 \left(\left(w^{(t)} t + b^{(t)} \right)_j \right) = G_3 \left(w_j^{(t)} t + b_j^{(t)} \right) \quad (4.5)$$

la dernière fonction de transfert G_4 est la fonction identité et projette les données à partir de $h^{(t)}$ vers \hat{x} . Le vecteur de sortie de dimension m :

$$\hat{x}_i = G_4 \left(\left(V^{(t)} h^{(t)} + \bar{b}^{(t)} \right)_i \right) = \sum_{j=1}^r v_{ji}^{(t)} h_j^{(t)} + \bar{b}_i^{(t)} \quad (4.6)$$

La fonction coût $E = \|x - \hat{x}\|^2$ est minimisée pour trouver les valeurs optimales de $V^{(x)}$, $b^{(x)}$, $w^{(x)}$, $\bar{b}^{(x)}$, $w^{(t)}$, $b^{(t)}$, $V^{(t)}$ et $\bar{b}^{(t)}$.

Une fois la structure générale définie, il reste à déterminer la structure précise du modèle neuronal (architecture du réseau). Pour cela, il faut déterminer le nombre de couches cachées nécessaire et le nombre de neurones dans chaque couche cachée. Pour la détermination du nombre de couches cachées, Cybenko [14] a montré que toute fonction continue peut être approximée par un réseau de neurones à trois couches utilisant une fonction d'activation sigmoïdale pour les neurones de la couche cachée et une fonction d'activation linéaire pour les neurones de la couche de sortie. Le nombre de neurones dans la couche cachée est généralement déterminé en effectuant une validation croisée sur un jeu de validation. Il existe des algorithmes permettant de construire itérativement la couche cachée, le contrôle de la croissance du réseau est effectué par une validation croisée. D'autres méthodes utilisent exactement un chemin inverse des précédentes, partant d'un réseau avec un grand nombre de neurones dans la couche cachée ; les connexions jugées inutiles sont éliminées progressivement jusqu'à l'obtention d'une structure satisfaisante.

4.2.2 Apprentissage du réseau

La plupart des algorithmes d'apprentissage des réseaux de neurones sont des algorithmes d'optimisation, ils cherchent à minimiser, par des méthodes d'optimisation non linéaire, une fonction de coût. Cette optimisation se fait de manière itérative, en modifiant les poids en fonc-

tion du gradient de la fonction de coût (qui, dans le cas d'un apprentissage supervisé, constitue une mesure de l'écart entre les réponses réelles du réseau et ses réponses désirées). Les poids sont initialisés aléatoirement avant l'apprentissage, puis modifiés itérativement jusqu'à obtention d'un compromis satisfaisant entre la précision de l'approximation sur l'ensemble d'apprentissage et la précision de l'approximation sur l'ensemble de validation. On note $E(k)$ l'erreur quadratique obtenue à partir des erreurs obtenues sur les m neurones de sortie en présence de l'exemple k :

$$E(k) = \frac{1}{2} \sum_{i=1}^m (x_i(k) - \hat{x}_i(k))^2 \quad (4.7)$$

La minimisation de E_k par rapport aux poids $v_{ji}^{(t)}$ est donnée par :

$$\frac{\partial E(k)}{\partial v_{ji}^{(t)}} = \frac{\partial E(k)}{\partial \hat{x}_i(k)} \frac{\partial \hat{x}_i(k)}{\partial v_{ji}^{(t)}} \quad (4.8)$$

$$\frac{\partial E(k)}{\partial \hat{x}_i(k)} = -(x_i(k) - \hat{x}_i(k)) \text{ et } \frac{\partial \hat{x}_i(k)}{\partial v_{ji}^{(t)}} = h_j^{(t)}.$$

Le gradient de $E(k)$ calculé sur la couche de sortie est donné par :

$$\frac{\partial E(k)}{\partial v_{ji}^{(t)}} = -(x_i(k) - \hat{x}_i(k)) h_j^{(t)} \quad (4.9)$$

Ainsi les modifications des poids sont données par :

$$\Delta v_{ji}^{(t)} = \mu (x_i(k) - \hat{x}_i(k)) h_j^{(t)} \quad (4.10)$$

De même pour les couches intermédiaire, le gradient de $E(k)$ par rapport à $w_j^{(t)}$ est donné par :

$$\frac{\partial E(k)}{\partial w_j^{(t)}} = \frac{\partial E(k)}{\partial \hat{x}_i(k)} \frac{\partial \hat{x}_i(k)}{\partial h_j^{(t)}} \frac{\partial h_j^{(t)}}{\partial w_j^{(t)}} \quad (4.11)$$

$$\frac{\partial E(k)}{\partial w_j^{(t)}} = -G_3 (w_j^{(t)} t + b_j^{(t)}) t \sum_{i=1}^m (x_i(k) - \hat{x}_i(k)) v_{ji}^{(t)} \quad (4.12)$$

Les modifications de $w_{ji}^{(t)}$ sont donnés par :

$$\Delta w_j(t) = \mu G_3 (w_j^{(t)} t + b_j^{(t)}) t \sum_{i=1}^m (x_i(k) - \hat{x}_i(k)) v_{ji}^{(t)} \quad (4.13)$$

Les même opération de dérivation et de calcul des modifications seront appliquées pour les poids $w_j^{(x)}$ et $v_{ji}^{(x)}$, ce qui nous permet d'écrire :

$$\frac{\partial E(k)}{\partial w_j^{(x)}} = \frac{\partial E(k)}{\partial t} \frac{\partial t}{\partial w_j^{(x)}} \quad (4.14)$$

On peut écrire $\frac{\partial t}{\partial w_j^{(x)}} = h_j^{(x)}$ d'où l'expression du gradient de $E(k)$ par rapport à $w_j^{(x)}$:

$$\frac{\partial E(k)}{\partial w_j^{(x)}} = -h_j^{(x)} \sum_{j=1}^r \dot{G}_3(w_j^{(t)}t + b^{(t)}) w_j^{(t)} \sum_{i=1}^m (x_i(k) - \hat{x}_i(k)) v_{ji}^{(t)} \quad (4.15)$$

Finalement, le gradient de $E(k)$ par rapport aux poids $v_{ij}^{(x)}$ est donné par :

$$\frac{\partial E(k)}{\partial v_{ij}^{(x)}} = \frac{\partial E(k)}{\partial t} \frac{\partial t}{h_j^{(t)}} \frac{h_j^{(t)}}{\partial v_{ij}^{(x)}} \quad (4.16)$$

$$\frac{\partial E(k)}{\partial v_{ij}^{(x)}} = -\dot{G}_1\left(\sum_{i=1}^m v_{ij}^{(x)} x_i + b_j^{(x)}\right) x_i \sum_{p=1}^r w_p^{(x)} \sum_{j=1}^r \dot{G}_3(w_j^{(t)}t + b^{(t)}) w_j^{(t)} \sum_{i=1}^m (x_i(k) - \hat{x}_i(k)) v_{ji}^{(t)}.$$

Ainsi, on a obtenu les gradients de $E(k)$ par rapport aux différents poids du réseau et la règle de modification de ces poids est donnée par $\Delta w = -\mu \frac{\partial E}{\partial w}$ où μ est le pas d'adaptation.

Il est à noter que l'extraction des composantes principales peut se faire de deux manière, les algorithmes sont données en Annexe. La première consiste à extraire les composantes séquentiellement en ayant qu'un seule neurone dans la couche du milieu "Bottleneck Layer" (ACPNL séquentielle Figure A.1). La seconde consiste à extraire les q composantes désirées simultanément en insérant q neurones dans la couche du milieu (ACPNL parallèle Fig.A.2).

L'ACPNL parallèle nécessite, avant l'apprentissage du réseau, la détermination du nombre de composantes non linéaire à retenir q . Tandis que dans le cas de l'ACPNL séquentielle le réseau est entraîné initialement avec une seule composante principale non linéaire, on doit soustraire le résultat obtenu de l'ensemble des données de départ et l'opération d'extraction d'une deuxième composante non linéaire est effectuée sur les résidus obtenus. Cette procédure peut être répétée jusqu'à ce que le nombre de composantes voulu soit atteint ou l'erreur d'estimation inférieure à un certain seuil choisi à priori.

4.2.3 Exemple illustratif

L'exemple choisi est donné par les équations suivantes :

$$\begin{aligned} x_1 &= a^2 + 0.3 \sin(2\pi a) + E_1. \\ x_2 &= a + E_2. \\ x_3 &= a^3 + a + 1 + E_3. \end{aligned} \quad (4.17)$$

où $a \in [-1,1]$ est une variable aléatoire uniforme et E_i un bruit aléatoire uniformément distribué dans l'intervalle $[-0.1,0.1]$.

Les mesures en bleu La Figure 4.3 représente les échantillons des variables x_1, x_2, x_3 , après l'apprentissage du réseau on obtient les échantillons en rouge qui représentent les valeurs de la seule composante principale non linéaire issue du neurone de la couche d'entrelacement.

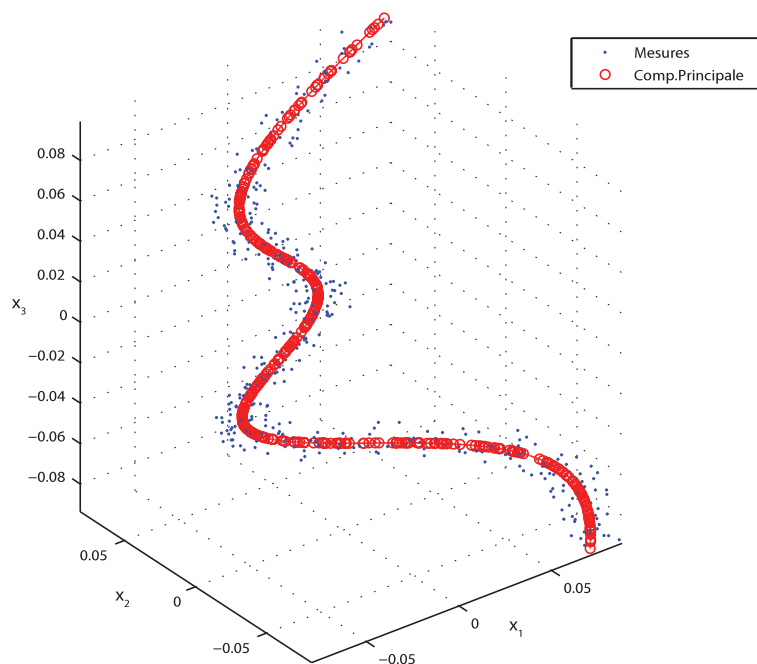


FIGURE 4.3 – Mesure avec la C.P non linéaire pour un réseau à 5 couche[3-4-1-4-3]

4.3 Détection de défaut par ACPNL

Nous utiliserons la statistique de contrôle SPE et SWE pour la détection du défaut, que nous avons introduit pour le cas de l'ACP classique ainsi que de la procédure de localisation de la variables en défaut.

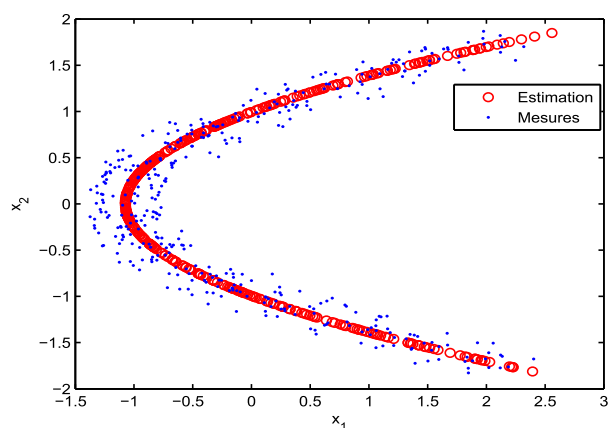


FIGURE 4.4 – Mesure et estimation

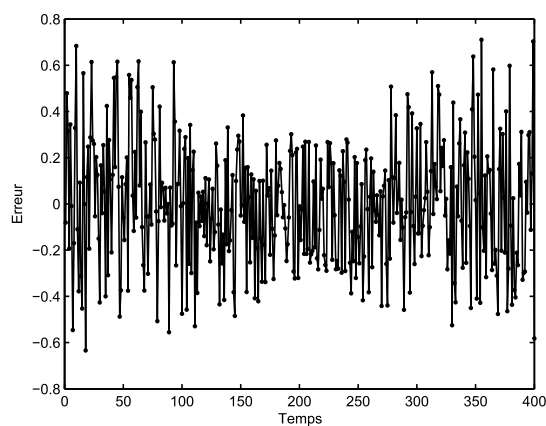


FIGURE 4.5 – Erreur de reconstruction

Une simulation d'un système non linéaire du 2^{ème} ordre est présenté pour illustrer l'application de la méthode de détection de défaut en utilisant l'ACPNL. Dans cet exemple toutes les données de l'apprentissage et de simulation de défaut sont produites par les fonctions décritent par l'équation 4.18. La Figure 4.4 et Figure 4.5, représentent respectivement l'estimation des nouvelles données sans défaut, ainsi que l'erreur de reconstruction de la variable x_1 .

Les E_i sont des bruit aléatoire uniformément distribué dans l'intervalle $[-0.1,0.1]$, a une variables uniformément distribuée entre $[-1,1]$. Le nombre de composante principale à retenir ne peut être qu'une seule composante ($q = 1$) pour cet exemple, après l'apprentissage du réseau dont la structure est $[3-4-1-4-3]$, le modèle ACPNL construit sera utilisé pour estimer les données avec une seule composante.

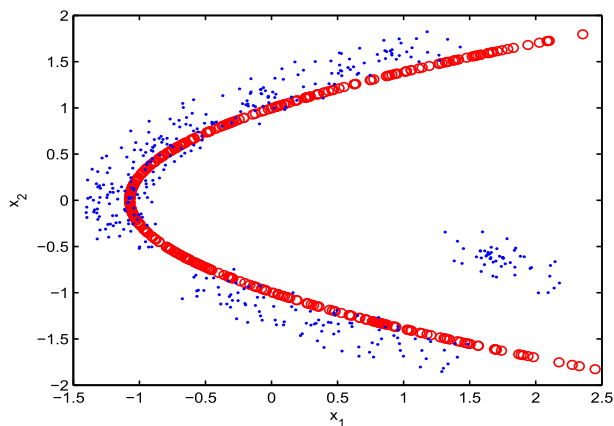


FIGURE 4.6 – Mesure et reconstruction avec défaut sur x_1

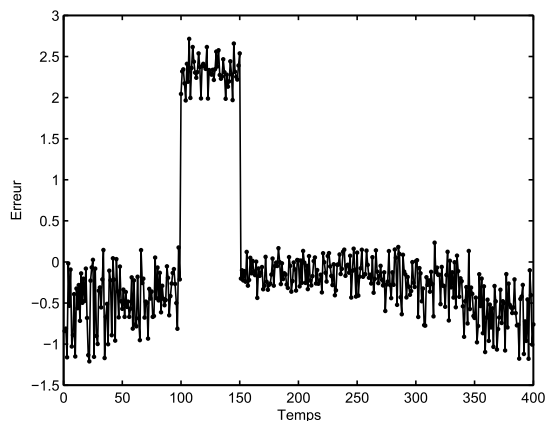


FIGURE 4.7 – Erreur de reconstruction de la variable x_1

$$\begin{aligned} x_1 &= a^2 + E1 \\ x_2 &= a + E2 \end{aligned} \tag{4.18}$$

Dans la suite nous allons simuler un biais d'une amplitude égale à 1 sur la première variable entre les instant 100 et 150. L'estimation des nouvelle données fournies par le modèle NLPCA est montrée sur la Figure 4.6. Les mesures avec défaut représentés en bleu et l'estimation des nouvelle données par le modèle ACPNL en rouge. La Figure 4.7 représente l'erreur de reconstruction de la variables x_1 .

On remarque qu'à partir de ces dernières Figures, la reconstruction de la variable x_1 est très satisfaisante et l'amplitude du défaut peut être estimée en calculant l'erreur de reconstruction (Figure 4.7).

Chapitre 5

Analyse en composantes principales multi-échelle

Sommaire

5.1	Introduction	40
5.2	MSPCA non-linéaire Méthodologie	41
5.3	Procédure détaillé de l'MSPCA non linéaire :	42

5.1 Introduction

L'ACP classique s'effectue sur une échelle fixe, l'échelle de la fréquence d'échantillonnage, et représente une bonne méthode pour la détection du changement sur une seule échelle. Cependant, les données dans la plus part des processus pratique sont de nature multi-échelle[4] et ceci est du aux :

1. Evénements se produisant dans différents endroit et avec différentes localisation dans le temps et la fréquence.
2. Processus stochastique non-stationnaire avec des changements temporel et fréquentiel dans leurs spectres de puissance.
3. variables mesurées à différents taux d'échantillonnages ou bien présentant des données perdus.

On note généralement que les composantes du bruit sont principalement localisées dans les hautes bandes de fréquences, tandis que la dynamique du processus est généralement localisée dans les bandes basse fréquence. Par conséquent, une technique multi-échelle comme les ondelettes devient un choix naturel pour séparer ces deux quantités.

Beaucoup de méthode pour le Control Statistique du processus (CSP) assume les mesures non-corrélées, alors qu'en pratique les mesures autocorrélés sont extrêmement communes. Une

approche commune pour le CSP des mesures auto-corrélées est de les décorrelées en l'adaptant à un modèle à série temporel et surveiller l'erreur résiduel.

Cependant, cette approche n'est souvent pas pratique pour une utilisation industrielle, particulièrement pour les processus multi-variables, due à leur haute dimensionnalité. Pour le CSP multi-variable, la dynamique est capturée par un modèle à serie temporel linéaire en utilisant par exemple l'ACP dynamique ou bien mieux encore l'analyse en composantes principales multi-échelle, comme mentioné auparavant dans la section 2.2, on peut dire que les coefficients d'ondelettes de la plus part des processus stochastique sont approximativement décorrelés.

Ainsi l'approche MSPCA extrait simultanément l'inter-corrélation à travers les capteurs (Approche ACP) et l'auto-corrélation à l'intérieur de chaque capteurs (Approche ondelette). En utilisant les ondelettes, chaque signal de capteur est décomposé en signaux d'approximations et de détails à différentes échelles. Les contributions de chaque échelle sont collectées dans des matrices séparées, et un modèle ACP est construit pour extraire la corrélation à chaque échelle. La nature multi-échelle de l'MSPCA la rend plus approprié pour travailler avec des données qui sont non-stationnaire et qui présentent des effet de plusieurs phénomène, chacuns opérant à différentes échelles.

Comme la plupart des données pratique sont non-linéaire, des études récentes ont été proposé R.Shao et al.[43], Choi et Morris[53]. On utilisera l'ACP non-linéaire à base de réseaux de neurones [33], pour la partie modélisation, présenté dans le Chapitre 4.

5.2 MSPCA non-linéaire Méthodologie

La méthodologie de l'MSPCA non-linéaire est montrée dans la Figure 5.1 suivante :

Comme expliqué dans la Figure 5.1, on considère une matrice X ($n \times m$) avec m variables et n échantillons. chaque colonne est décomposée en appliquant la DWT(Discret Wavelet Transform) comme expliqué dans la Section 2.2, la même transformée en ondelette avec le même niveau de décomposition L est appliqué à chaque m variables. Les signaux d'approximations A_L des m décomposition sont collectés dans des matrices de taille $m \times n/2^L$ (comme montré sur la matrice A_L dans la Figure5.1). De même, les coefficients de détails(D_1 à D_L) de chacunes des m décompositions sont collectées dans L matrices(matrice D_L jusqu'à D_1 pour chaque L niveau), avec la taille de la matrice qui varie de $m \times n/2^i, i = 1, 2, \dots, L$. Ainsi on aura formé $L + 1$ matrices, chacunes représentée à différentes échelles. On construit ensuite le modèle ACP non linéaire(voir le chapitre 4) pour chaque $L + 1$ matrices afin d'extraire les corrélations non-linéaires entre les capteurs. On presentera dans la sous-Section suivante la procédure détaillé de l'MSPCA non linéaire.

Une importante décision pratique dans la méthodologie de l'MSPCA est la profondeur de l'échelle pour la décomposition en ondelette. Idéalement, la profondeur devra être choisie pour fournir le maximum de séparation entre les composantes stochastiques et les caractéristiques

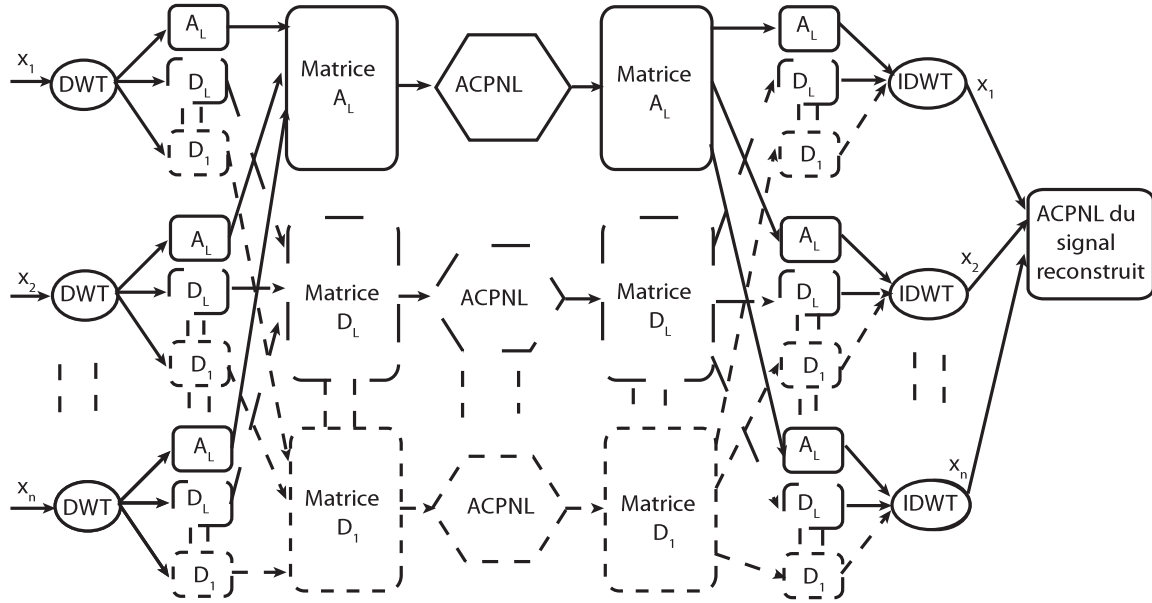


FIGURE 5.1 – MSPCA non-linéaire Méthodologie

principale du signal. Si la profondeur est trop petite, alors le dernier signal échelonné aura un taux de bruit significatif qui sera retenu dans le résultat de l'MSPCA, mais si la profondeur est trop grande, la matrice des coefficients aux grandes échelles n'aura que quelques lignes dues à l'échantillonnage dyadique, et ceci affectera aussi la précision de l'ACP à cette échelle.

5.3 Procédure détaillée de l'MSPCA non linéaire :

Afin d'obtenir le modèle MSPCA non-linéaire les étapes suivantes sont nécessaires :

1. Obtenir les données de référence quand le processus est en fonctionnement normal. Pour chaque variable dans la matrice des données de référence, calculer la décomposition en ondelette et obtenir les coefficients d'ondelette de référence.
2. Pour chaque échelle, mettre ensemble les coefficients d'ondelette de référence de toutes les variables, appliquer l'ACP non linéaire pour avoir le modèle ACP non linéaire de référence (sauvegarder la structure du réseau de neurone obtenu), fixer les seuils pour les statistiques de contrôles, répéter cette procédure pour chaque échelle.
3. Définir les scénarios de reconstruction en se basant sur le nombre de niveau de décomposition.

Pour chaque scénario de reconstruction, assigner l'échelle significative sélectionnée avec le coefficient d'ondelette de référence correspondant et les échelles non significatives avec zéro (hard thresholding). Reconstruire le signal à partir des coefficients seuillés, sélectionnés pour chaque variable. Mettre le signal reconstruit de toutes les variables ensemble et

appliquer l'ACP non linéaire pour avoir le modèle ACP non linéaire de référence (incluant la structure du réseau de neurone, ainsi que les limites de contrôle) pour ce scénario de reconstruction. Répéter la même procédure pour tous les autres scénarios de reconstruction. Les nouvelles données sont décomposées et projetées sur les modèle correspondant à chaque échelle. Si le résidu résultant pour chaque échelle dépasse la limite de confiance, l'échelle est dite significative et est utilisée pour reconstruire le signal, ce dernier sera utilisé pour la surveillance du processus uni-échelle par ACP non linéaire.

4. Localiser la variable en défaut par calcul de contribution du signal reconstruit.

Concernant la procédure de l'analyse en composante principale multi-échelle linéaire, le principe de fonctionnement reste le même, mais dans la modélisation on utilisera l'ACP classique et on n'oubliera pas de sauvegarder pour chaque échelle les variances et les matrices de chargement P , la matrice des valeurs de composante principale T et les limites de détection du processus en fonctionnement normal.

Bien que nous n'allons pas utiliser dans notre travail la procédure de surveillance en ligne, il est intéressant de voir comment celle-ci est effectuée pour le cas de l'MSPCA classique.

Surveillance en ligne du processus

1. Déterminer la taille de la fenêtre glissante de taille dyadique, w . Générer une fenêtre de données avec w échantillon à partir des données temps réel en déplaçant la fenêtre de temps. Pour chaque variable dans la fenêtre de données, calculer les coefficients d'ondelette.
2. Pour chaque échelle, calculer les graphes de contrôle T^2 et Q en se basant sur le modèle ACP de référence.
3. Comparer la T^2 et Q avec les limites de contrôle dans I. (2), retenir le coefficient d'ondelette qui violent les limites de contrôles et assigner ceux qui à l'intérieur de la limite à zéro.
4. Reconstruire le signal dans la fenêtre glissante variable par variables.
5. Puisque seulement les échantillons les plus récents (les dernières données sur la fenêtre glissante) représente un intérêt, déterminer le scénario de reconstruction en se basant sur la dernière T^2 et Q de chaque échelle et calculer la limite de détection pour le dernier signal reconstruit. Calculer T^2 et Q du dernier signal reconstruit en se basant sur le modèle ACP de référence.
6. Si T^2 ou Q du dernière signal reconstruit dépasse les limites de détection : alarme.

Chapitre 6

Application de l'MSPCA aux paramètres de pollutions de l'air

Sommaire

6.1	Introduction	44
6.2	Pollution atmosphérique	45
6.2.1	Polluants primaires	45
6.2.2	Polluants secondaires	46
6.3	Phénomène de pollution	46
6.4	Prétraitement des données	47
6.5	Application de l'MSPCA	47
6.5.1	Choix des paramètres MSPCA	48
6.5.2	Détection et localisation de défauts	49
6.6	conclusion	55

6.1 Introduction

Maintenir une qualité de l'air qui suis des normes spécifiques reviens à établir une stratégie de surveillance, à travers des stations qui permettent de fournir des données pour informer la population et les services public en cas de dépassement des limites fixés. La validité et la crédibilité des données fournies par les différents capteurs issue d'une ou de différentes stations représente la tache principale de la procédure de l'MSPCA non linéaire, entre autre améliorer la procédure de détection et de localisation du défaut. La modification de l'information réel que dois fournir un capteur influe considérablement sur les décisions present par l'opérateur ou par le système de traitement de l'information, notamment dans les choix pris pour la réduction des émissions.

Dans ce chapitre nous allons voir une application de l'MSPCA non linéaire pour la détection

et la localisation de défaut de capteur sur des paramètres de pollution qui servent à surveiller la qualité de l'air de la région d'Annaba, nous comparerons cette approche avec l'approche MSPCA classique, et celle de l'ACP classique .

6.2 Pollution atmosphérique

L'air est une ressource fondamentale et en apparence inépuisable. La qualité de l'air demeure un sujet préoccupant. Les émissions de contaminants atmosphériques provenant des activités industrielles comme les oxydes d'azotes(NO_2 et NO), du transport(NO_x , COV , CO) et de l'utilisation de combustibles fossiles à des fins énergétiques sont concentrées majoritairement en milieu urbain et sont susceptibles d'affecter la santé de la population et le système écologique. L'air en réalité n'est pas si pur. C'est en fait un mélange essentiellement composé d'oxygène (21%) et d'azote (78%). Mais l'air contient d'autres gaz, certes dans une bien moindre proportion, ainsi que des particules liquides ou solides en suspension, qu'on appelle aérosols. Certains gaz ont un rôle important, voire indispensable. L'ozone(O_3) par exemple, s'il est néfaste à basse altitude, l'ozone stratosphérique, situé à 25km d'altitude environ, filtre les ultraviolets les plus nocifs du rayonnement solaire. Le gaz carbonique, par effet de serre, assure une température moyenne sur Terre de 15 °C. Sans l'effet de serre naturel, elle serait de -18°C!

Plutôt que de pureté, mieux vaut parler de qualité de l'air, en regard des effets bénéfiques ou nuisibles sur les êtres vivants.

La pollution atmosphérique (ou de l'air) c'est l'altération de la composition chimique de l'air, par l'augmentation de la concentration de certaines espèces (chimique ou solide), cet altération durera une certaine période de temps pour créer des effets toxique, et qui aura aussi des effets nocif sur l'écosystème et la santé.

La pollution est dites locale (à l'échelle régional tel que : les pics d'ozone, le smog), globale à l'échelle planétaire (exemple l'effet de serre, Destruction de la couche d'ozone).

6.2.1 Polluants primaires

Les polluants primaires, sont ceux qui sont émis directement dans l'air, ces émissions sont catégorisés en deux types :

- Biogénique : liées a l'activité naturel tel que : l'activité photosynthétique, sels de mer, volcaniques etc.
- Anthropique : liées a l'activité humaine tel que : les processus industriel, transport, production de l'énergie etc.

On peut citer comme exemple de polluant primaires l'oxyde de soufre qui est émis par l'opération de combustion de charbon.

6.2.2 Polluants secondaires

Les polluants secondaire résultent des réactions photochimique, dans l'air, dans lesquelles les polluants primaires jouent un rôle très important, un exemple de ce genre de polluant est l'ozone(O_3), ce dernier résultent de la photochimie des oxydes de l'azote(NO_x). On s'intéressera dans notre application à ce dernier type de polluant. On expliquera dans la section suivante le Phénomène de photochimie qui lie l'ozone aux dioxydes d'azote.

6.3 Phénomène de pollution

La pollution photochimique correspond aux polluants formé par le biais du rayonnement solaire à partir des espèce chimique présentes dans l'atmosphère. Comme les paramètres météorologique varient en fonction des saison, les phénomènes de photochimie sont plus spécifiques à une période de l'année. On rencontre d'avantage ce type de polluant en été et tout particulièrement les jours chauds et ensoleillés, avec un vent faible. L'ozone est l'un de ces polluants. Il met en jeu les oxydes d'azotes grace au cycle de chapman (Fig. 6.1) Mais l'ozone est aussi produit grace à d'autres reactions chimiques faisant intervenir les Composantes Organiques Volatils (COV), qui jouent le rôle de précurseurs. De plus, si le vent est faible, l'ozone crée stagne au-dessus des sources d'émission, ce qui conduit à un pic de pollution. Le maximum est atteint en fin d'après midi. L'évolution des concentrations d'ozone se déroule essentiellement en deux phases qui dépendent de l'éclairement solaire. Durant la première phase de 6h à 16h environ, le soleil réchauffe progressivement le sol créant des structures turbulentes qui dispersent les oxydes d'azote. De plus, la dégradation des COV engendre des radicaux tel que HO_2 qui favorisent par la suite le cycle de formation $NO + HO_2 \rightarrow NO_2 + OH$, ainsi la destruction de l'ozone par $NO + O_3 \rightarrow NO_2 + O_2$ est moins sollicitée. Ce phénomène crée donc une hausse de l'ozone durant cette période. Ensuite, pendant la baisse de l'éclairement solaire et la nuit, l'ozone se retrouve consommé par les radicaux et n'est plus produit.

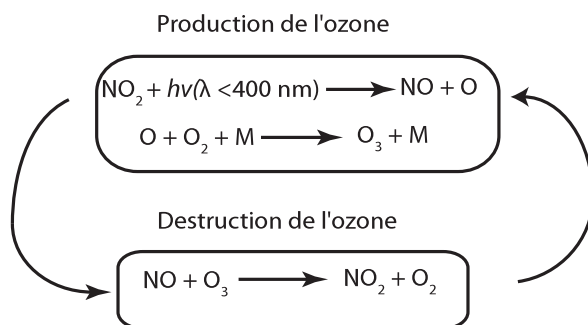


FIGURE 6.1 – Illustration du cycle de Chapman

A l'inverse, les concentrations de NO et NO_2 sont fortes aux environs de 6H, diminuent dans l'après midi puis augmentent pendant la nuit. La stabilité thermique matinale bloquant les espèces en basse atmosphère, crée de fortes concentrations de polluants dans la matinée.

Puis le réchauffement du sol engendre de la turbulence permettant de disperser les polluants, ce qui fait baisser les concentrations. Ces dernières remontent ensuite en soirée pour les mêmes raisons : les structures convectives disparaissent, la couche limite atmosphérique diminue et bloque à nouveau les espèces.

Le nombre important de capteur et l'éventuel présence de corrélation entre les signaux de ces derniers, rend la tâche de modélisation sous forme de modèle entrée/sortie une tâche difficile d'où la nécessité de remédier aux méthodes de contrôle statistique multi-variables qui permettent de modéliser le comportement de l'ensemble des variables. Les paramètres à surveiller dans notre application sont : O_3 , NO_2 , NO , NO_x , elle sont lié comme on l'a vu auparavant par le cycle de chapman(Fig. 6.1). Les concentrations des polluants sont données en $\mu g/m^3$.

6.4 Prétraitement des données

Les données provenant des stations de mesures peuvent être altérées par des pannes dans la chaîne d'instrumentation, maintenance de capteurs, ou bien par des mesures qui ne sont pas acceptées par le système (changement très rapide), il devient important pour les traitements ultérieurs de trouver un remplacement pour les valeurs manquantes par simple interpolation linéaire par exemple ou en éliminant l'ensemble des données correspondant à une période où les valeurs manquantes présente une série de mesures sur plusieurs jours ou sur plusieurs variables. Il faut savoir que sans un prétraitement approprié des données, la construction d'un modèle robuste du processus pour la surveillance de ce dernier reste problématique et sans intérêt.

6.5 Application de l'MSPCA

Dans cette Section, l'analyse en composantes principales multi-échelle linéaire et non linéaire, ainsi que l'approche classique de l'ACP seront appliqués à des données représentant des paramètres de pollution de l'air de la région d'Annaba. L'approche de détection et de la localisation du défauts sera effectuée sur des capteurs mesurant les paramètres suivant : NO_2 , NO , NO_x et O_3 . Les relations qui lient ces variables sont décritent par les réactions citées auparavant (Voir Figure 6.1). On peut constater que l'on considère 4 variables dans notre application x_1, x_2, x_3, x_4 correspondant respectivement, au NO_2, NO, NO_x, O_3 (Voir Figure 6.2).

Pour effectuer une comparaison entre les trois approches de surveillance, ACP conventionnelle, MSPCA, et MSPCA non linéaire, nous avons identifier les trois modèles sur le même jeu de données. c'est à dire qu'une partie des données représentées sur la Figure 6.2 servira à la phase de modélisation, l'autre partie sera utilisée pour la phase de validation, entre autre, la détection et la localisation du défaut.

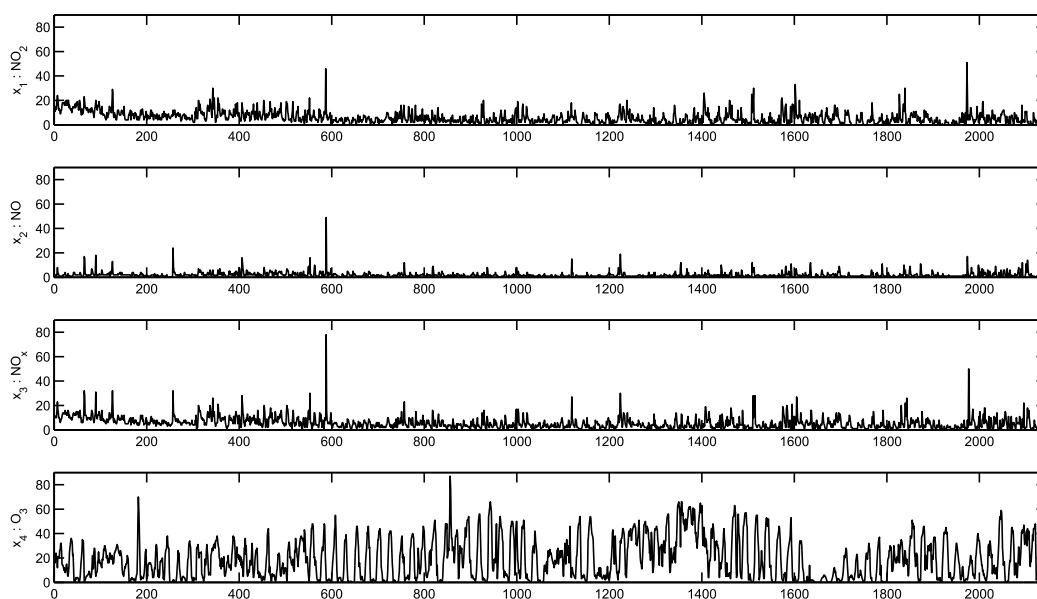


FIGURE 6.2 – Donnée de pollution de l'air de la région d'Annaba

6.5.1 Choix des paramètres MSPCA

Avant de commencer la phase de modélisation, il est important de définir quelques paramètres nécessaires pour le développement de l'approche MSPCA en générale.

La fonction d'ondelette w et le niveau de décomposition L doivent être déterminés convenablement sachant que ces deux facteurs vont influencer considérablement sur la séparation des caractéristiques stochastiques des caractéristiques déterministes du signal original. Une variété de fonction d'ondelette peuvent être utilisées. L'ondelette Daubechies 4 a été choisie pour ces bonnes propriétés [17]. On note que les ondelettes de daubechies (DaubN) possèdent différents niveaux de lissage qui sont déterminées par le moment de fuite (vanishing moment N). Ainsi on pourrait faire correspondre le degré de lissage de l'ondelette au lissage du signal à être analysé. Ces propriétés permettent une reconstruction précise et efficace du signal original, une facilité de calcul, et le stockage d'un grand nombre d'information dans un nombre réduit de coefficients. La décision d'utiliser les ondelettes Daubechies a été aussi influencée par différents chercheurs [43], [10], [50]. Une fois qu'une ondelette spécifique a été choisie, la paire de filtre passe bas et filtre passe haut sera déduite, ou alternativement.

Concernant le choix du niveau de décomposition L , il sera basé sur le nombre d'échantillons des données et des caractéristique du bruit. Idéalement, la profondeur de l'échelle devrait être choisie pour fournir un maximum de séparation entre les composantes stochastiques et déterministe du signal. Si la profondeur est trop petite, alors le dernier signal échelonné contiendra un taux de bruit important, qui sera retenu dans le résultat de l'MSPCA, mais si la profondeur est

trop grande, les coefficients aux plus grandes échelles seront limités en nombre d'échantillons due au sous échantillonnage dyadique, et ceci affectera la précision de l'ACP à cette échelle. Pour cette application, on veillera à ne pas décomposer au delà de 32 coefficients pour la plus grande échelle, c'est-à-dire prendre $L = \log_2 n - 5$ [4]. Un niveau de décomposition $L = 4$ est choisi pour cette application.

6.5.2 Détection et localisation de défauts

Pour effectuer une comparaison entre les différentes approches de surveillance, ACP, ACP dynamique, MSPCA, et MSPCA non linéaire, nous avons identifié chaque modèles sur le même jeu de données.

6.5.2.1 Défaut sur la variable x_2 (NO) avec 20% de variations

Sachant qu'on ne dispose pas de données de capteur de pollution montrant un état de fonctionnement défaillant, le défaut sera simulé sur l'une des variables. Les graphes de détection des différentes approches permettront de confirmer la présence ou non d'un défaut. Un défaut affectant la variable x_2 est simulé sous forme de fenêtre qui commence à l'échantillon 100 jusqu'à l'échantillon 400 et d'une amplitude qui s'élève à environ 20% de la plage de variation de x_2 . Les limites de détection en trait discontinu et en trait continu représentent respectivement les limites de contrôle à un niveau de confiance à 95% et 99%.

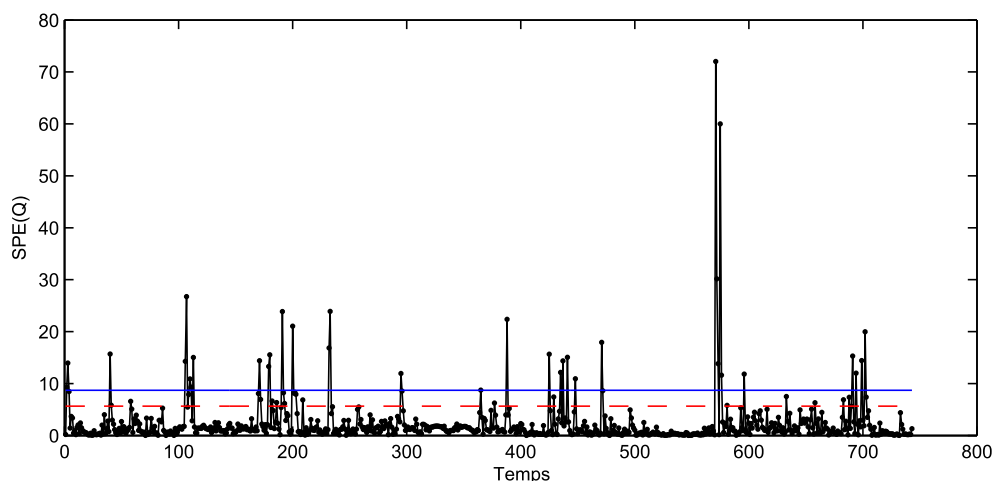


FIGURE 6.3 – SPE de l'ACP dynamique avec simulation du défaut sur la variable 2 entre 100 et 400

La Figure 6.3 présente l'évolution du SPE, pour le cas de la modélisation par ACP dynamique, en présence de défaut. Il est clair qu'on ne détecte rien d'inhabituel sauf la présence de bruit. La Figure qui présente le SPE du modèle ACP conventionnelle n'a pas été représenté étant donné qu'on obtient presque le même résultat.

Après avoir décomposé chacun des signaux de capteurs définis sur les colonnes de la matrice d'apprentissage X avec l'ondelette $DB4$ au niveau de décomposition correspondant, $L=4$, nous procédons à collecter les coefficients d'ondelette résultants, approximation (representant la composante basse fréquence du signal), détails (composantes hautes fréquences). Comme expliqué dans la section précédente (Section 5.3). Pour chaque échelle on calcule l'ACP des coefficients d'approximation et des coefficients détails. Le nombre de composantes principales pour le modèle ACP est déterminé pour chaque échelle par le principe du pourcentage cumulé de la variance (PCV) expliqué dans le chapitre 3. On oublie pas de sauvegarder le modèle ACP avec les limites de détection pour les données d'apprentissage (fonctionnement normal).

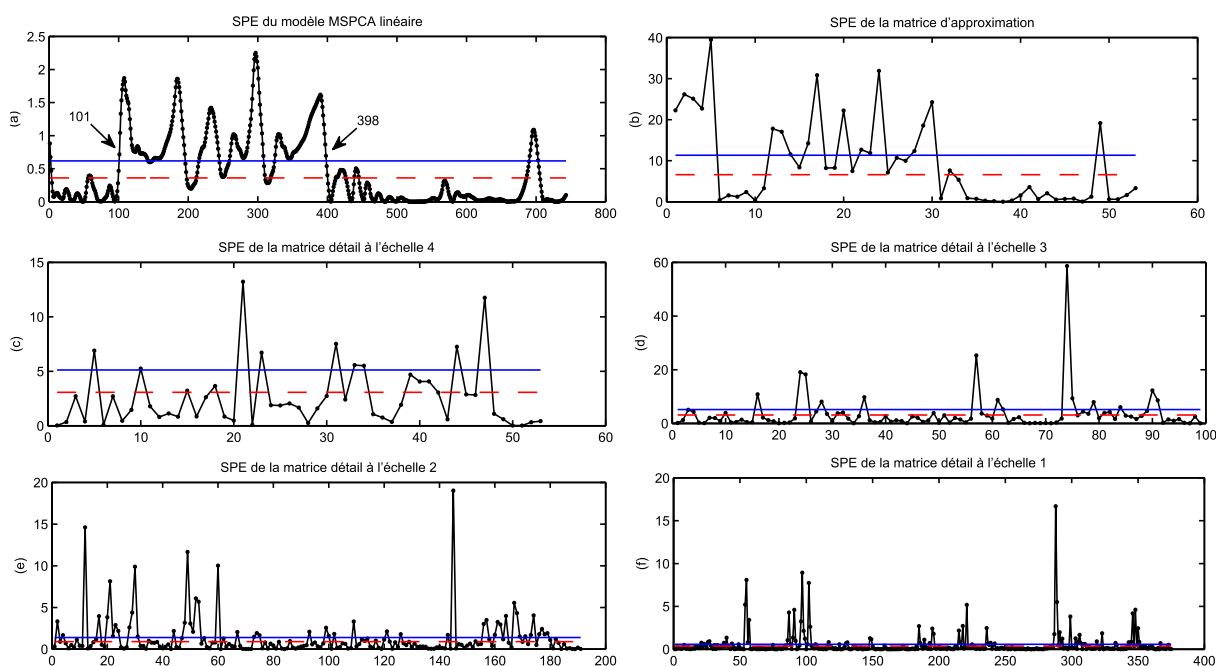


FIGURE 6.4 – Détection de défaut par l'MSPCA linéaire

La Figure 6.5 présente les mesures observées et les mesures estimées du polluants O_3 . Les estimations sont données par le modèle MSPCA linéaire. Les données estimées sont obtenues en projetant les nouvelles données sur le modèle construit avec les données d'apprentissage. La procédure de l'MSPCA détermine pour chaque échelle, un modèle ACP, par le calcul du nombre de composantes principales à retenir. la construction du modèle MSPCA final est réalisé à partir de la reconstruction des coefficients d'ondelette correspondantes à chaque échelle pour obtenir le signal uni-échelle dans le domaine temporel.

Dans la phase de simulation du fonctionnement anormal, seulement les échelles dont le SPE dépassent les limites de contrôle prédéfinis dans la phase de modélisation seront utilisées pour la reconstruction. Dans notre application elles seront assignées à zéro, parce qu'après avoir essayer différents scénarios de reconstruction, tout en sachant que les limites de détection sont dépassé sur les 4 échelle, le SPE du signal reconstruit étais entaché de bruit et ne pouvait pas détecter

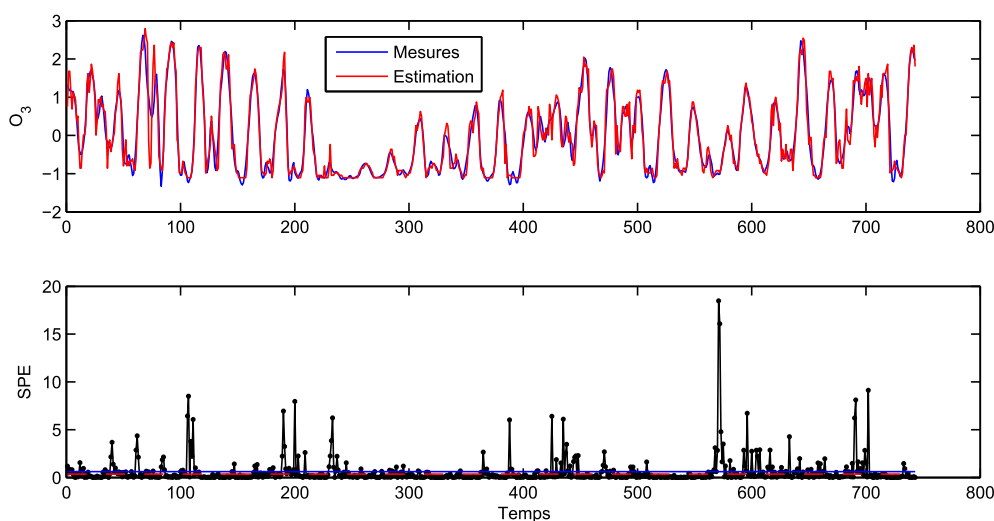


FIGURE 6.5 – Mesure et estimation par le modèle MSPCA linéaire

le défaut.

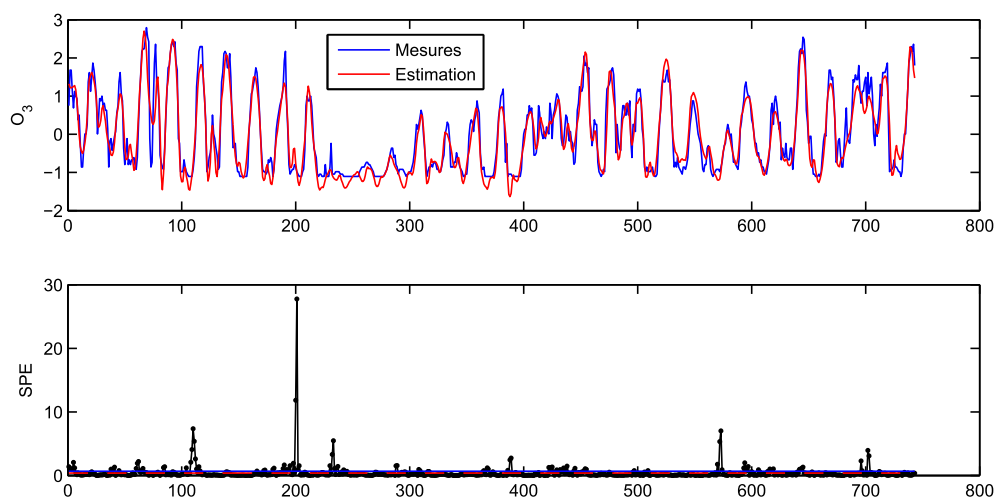


FIGURE 6.6 – Mesure et estimation par le modèle MSPCA non linéaire

La même procédure sera utilisée pour la détection et la localisation de défaut sur l'MSPCA non linéaire. Il faut noter que le nombre de composantes principales choisi par le critère du pourcentage cumulé de la variance(PCV) et qui est égal à deux lors de la modélisation par MSPCA linéaire pouvait capturer 94,16% de la variabilité des données.

La Figure 6.4((b)(c)(d)(e)(f)) présente le SPE des différents niveaux de décompositions des variables à surveiller (Approximation et détails). La Figure 6.4(a), représente le SPE du modèle MSPCA linéaire. Le défaut simulé sur la variable x_2 pour la fenêtre [100 400] est bien détecté.

Dorénavant, les figures qui présenteront le SPE des différentes échelles ne seront plus représentées puisqu'ils seront assignés à zéro.

Pour pouvoir comparer les deux approches, la même procédure sera appliquée pour l'analyse en composante principale non linéaire multi-échelle avec les mêmes paramètres de décomposition (niveau de décomposition $L=4$, et l'ondelette daubechies de niveau 4). L'ACP non-linéaire cette fois, est appliquée sur les coefficients d'ondelette à chaque niveau, pour extraire la corrélation non-linéaire entre les différents capteurs. La modélisation par ACPNL suppose la détermination de la structure du modèle. Dans notre application, nous avons choisi deux composante principale non-linéaire, permettant de capturer 96,5 % de la variabilité des données. Comme on l'a vu sur le chapitre 4 le réseau de neurone proposé par Kramer [33] est composé de 5 couche. La structure suivante [4 6 2 6 4] est choisie. Elle est composée de 4 variables qui représente les paramètres de pollutions en entrée avec les mêmes variables estimées en sortie avec 6 neurones pour chacune des deux couches cachées.

Remarque : Pour cette application l'analyse en composante principale non-linéaire multi-échelle, à montré des résultats quasiment similaire à l'MSPCA classique pour la détection de défaut. Ceci, peut être expliqué par le peu de non linéarité présent sur les paramètres de pollutions et la présence toujours de dépendance temporelle sur les coefficients d'ondelette.

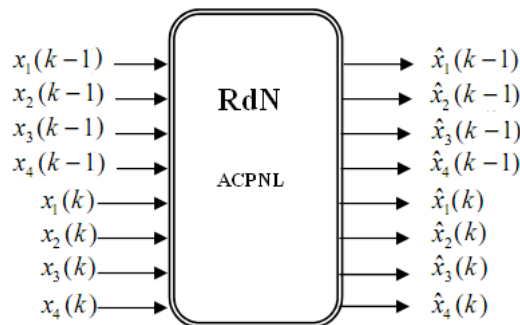


FIGURE 6.7 – Modèle ACP non-linéaire

Pour cette raison nous avons décidé d'ajouter des entrées décalées, rendant la structure du réseau dynamique, et capable ainsi, de mieux détecter les relations qui lient les variables du système. La structure du réseau est élargit à une topologie [8 8 3 8 8]. En plus des 4 entrées à l'instant (k), vient s'ajouter 4 autres mesures à l'instant ($k-1$). (Voir Figure 6.7).

6.5.2.2 Défaut sur la variable x_1 (NO_2) avec 20% de variations

La Figure 6.8 et 6.9 représentent respectivement, le SPE et la statistique T^2 appliqué au modèle MSPCA linéaire pour un défaut affectant la variable x_1 . La statistique T^2 n'a rien détecté d'inhabituelle. Ceci peut être expliqué par le fait que la base statistique n'est plus vérifiée, qui stipule que les mesures suivent une loi normale. On utilisera pour le reste des simulations la statistique SWE pour le cas de l'MSPCA linéaire. La statistique SWE, comme expliqué aupa-

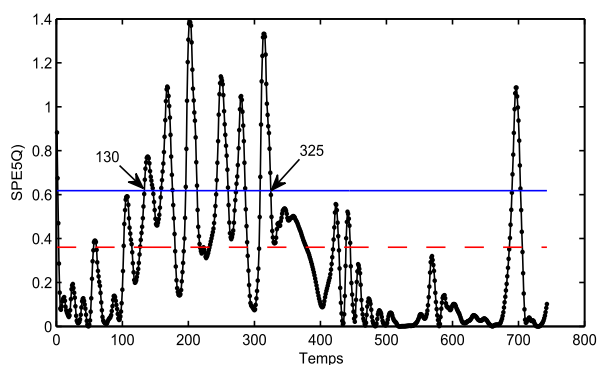


FIGURE 6.8 – SPE : MSPCA (défaut sur la variable x_1)

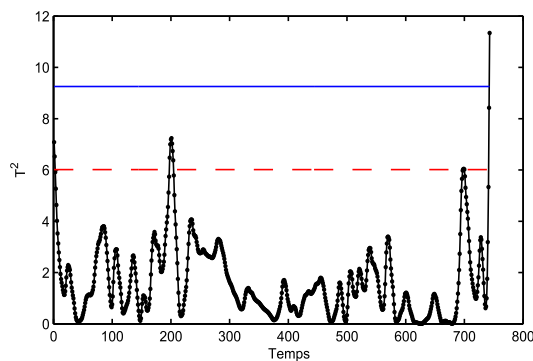


FIGURE 6.9 – Statistique T^2 MSPCA(défaut sur la variable x_1)

avant, est plus sensible au défaut puisqu'elle se base sur les dernières composantes principales, et qu'elle n'impose aucune supposition sur les mesures des variables.

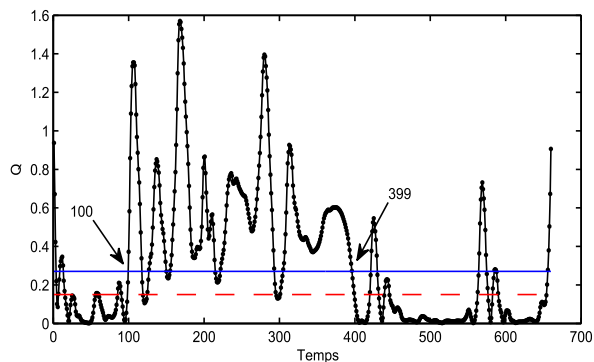


FIGURE 6.10 – SPE : MSPCA non linéaire (défaut sur la variable x_1)

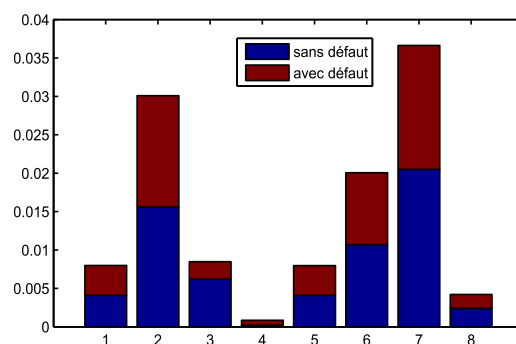


FIGURE 6.11 – Graph de contribution pour détecter la cause de la déviation

La Figure 6.10 présente le SPE du modèle MSPCA non linéaire. Il est clair que l'instant d'apparition du défaut et du retour au fonctionnement normal sont bien meilleurs que ceux de l'MSPCA classique. Néanmoins on remarque que le graph de contribution a failli de localiser la variable en défaut x_1 .

6.5.2.3 Défaut sur la variable x_2 (NO) avec 10% de variations

La première simulation a été sur les variables x_2 avec une amplitude de 20% de la plage de variation de cette dernière. On simulera cette fois un défaut d'une amplitude de 10% de la plage de variation de la variable x_2 .

On remarque que pour cette faible déviation de la variable x_2 , que le défaut est mieux détecté avec le modèle mspca non linéaire (Figure 6.14) avec moins de fausses alarmes que sur le modèle mspca classique (Figure 6.12). Bien que ce dernier arrive à détecter le défaut à son instant d'apparition, on remarque par la suite que l'erreur quadratique fluctue entre la limite de contrôle provoquant plusieurs fausses alarmes.

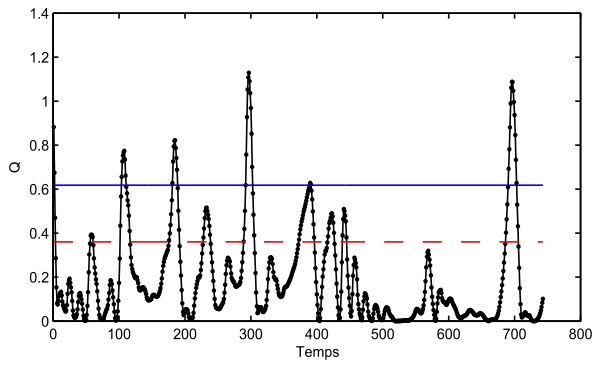


FIGURE 6.12 – SPE : MSPCA (défaut sur la variable x_2)

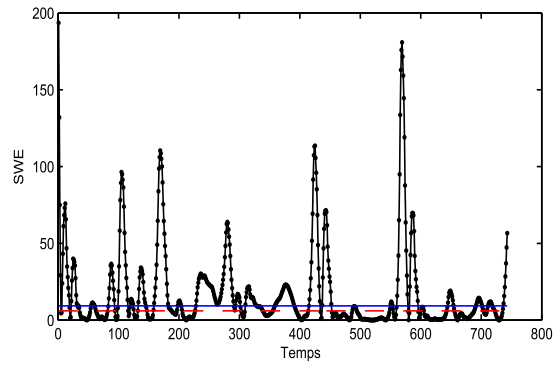


FIGURE 6.13 – SWE : MSPCA (défaut sur la variable x_2)

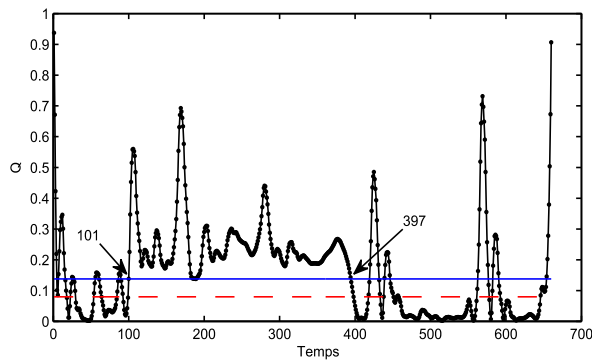


FIGURE 6.14 – MSPCA nonlinéaire (défaut sur x_2)

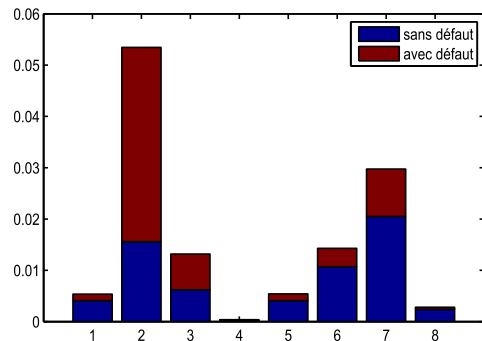


FIGURE 6.15 – Graph de contribution pour detecter la cause de la déviation

Après qu'une déviation du processus est identifiée, l'étape suivante est la détermination de la cause. La Figure 6.15 présente le calcul de contribution calculé à l'instant 100 du SPE pour localiser la variable en défaut. Il est clair que la variable x_2 présente la plus large contribution ce qui montre que la variable en défaut a été bien localisée. Il est intéressant aussi d'observer que non seulement la variable en défaut a augmenté mais aussi les autres variables. Ce scénario a été discuté par R.Dunia [41] et a été identifié comme étant du à la nature du SPE.

6.5.2.4 Défaut sur la variable x_3 (NO) avec 10% de variations

Les Figures ci-dessous (Figure 6.16, Figure 6.17) représentent, respectivement le SPE et le SWE du modèle MSPCA pour un défaut simulé entre les instant [150, 450] sur la variable x_3 .

On remarque que la statistique Q n'a pas réussi à détecter le défaut sur l'intervalle [150, 450]. Par contre, dans la statistique SWE le défaut est détecté avec moins de fausses alarmes mais la précision de détection avec l'instant de retour au fonctionnement normal reste non concluante.

La Figure 6.18, et la Figure 6.19 représentent respectivement, le SPE et le graphe de contribution du modèle MSPCA non linéaire pour le défaut simulé entre [150, 450]. On peut remar-

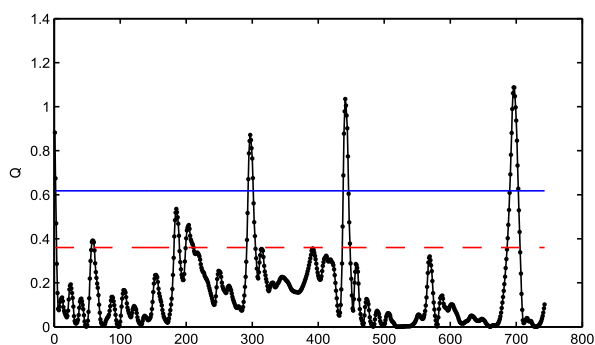
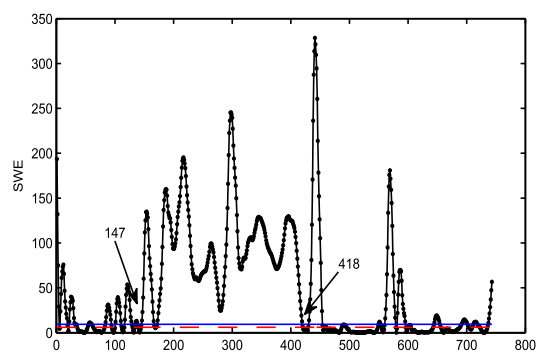
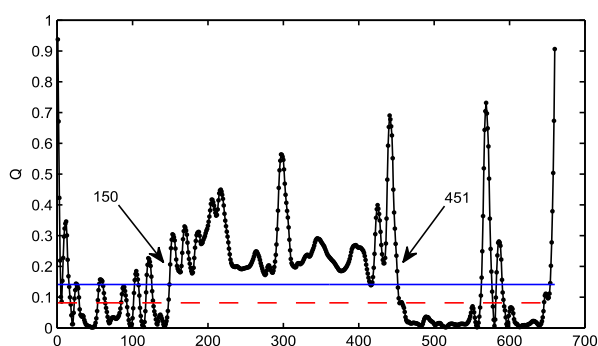
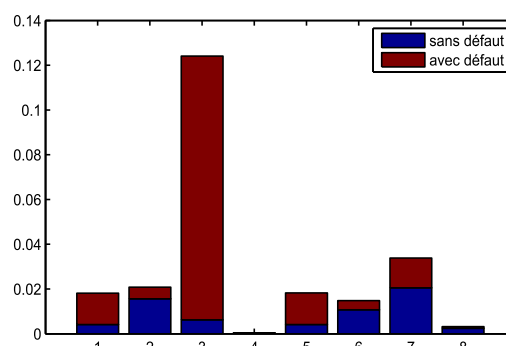
FIGURE 6.16 – SPE : MSPCA (défaut sur la variable x_3)FIGURE 6.17 – SWE : MSPCA (défaut sur la variable x_3)FIGURE 6.18 – SPE : MSPCA non linéaire (défaut sur variable x_3)

FIGURE 6.19 – Graph de contribution pour détecter la cause de la déviation

qué comme dans le cas discuté pour la variable x_2 que le SPE donne de bon résultats en ce qui concerne la détection de l'instant d'apparition et de retour au fonctionnement normale du défaut. Le graph de contribution nous montre encore une fois que la variable x_3 possède la plus grande contribution. Ceci nous indique que c'est bien la variable x_3 qui est incriminée.

6.5.2.5 Défaut sur la variable x_4 (O_3) avec 10% de variations

Les Figures ci-dessus démontrent encore une fois la sensibilité de l'MSPCA non linéaire à détecter le défaut. Le graph du SPE (Figure 6.20) et même le SWE (Figure 6.21) n'indique rien d'inhabituelle à part la présence de bruit. Le graphe de contribution calculé à l'instant 150 (instant de détection du défaut) incrimine la variable x_4 , ce qui montre que c'était bien la capteur porteur de défaut.

6.6 conclusion

Ce chapitre a été consacré à l'application de l'analyse en composante principale multi-échelle non linéaire (NLMSPCA) pour la détection et la localisation de défaut de capteurs d'un réseau

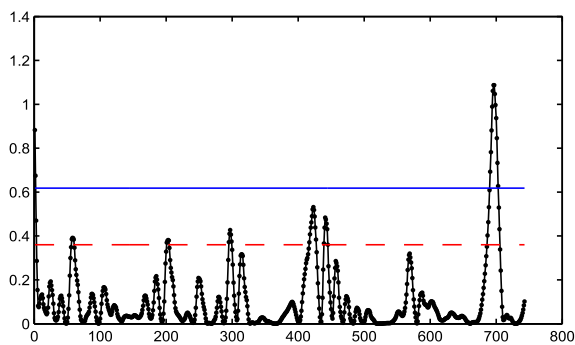


FIGURE 6.20 – SPE : MSPCA (défaut sur la variable x_4)

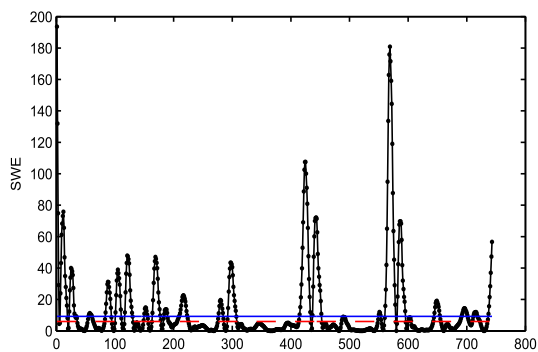


FIGURE 6.21 – SWE : MSPCA (défaut sur la variable x_4)

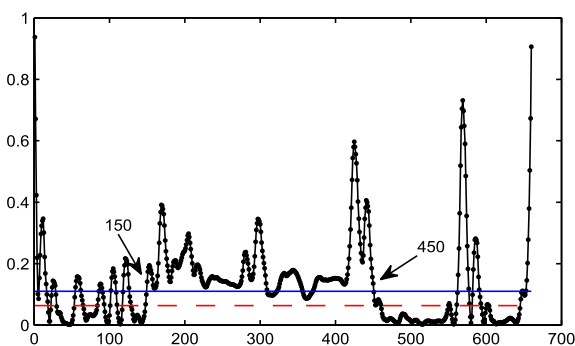


FIGURE 6.22 – SPE : MSPCA non linéaire(défaut sur la variable x_4)

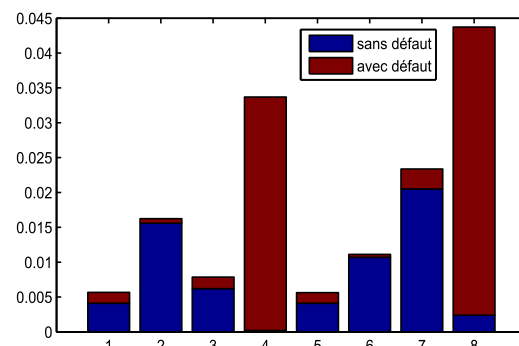


FIGURE 6.23 – Graph de contribution pour detecter la cause de la déviation

de surveillance de la qualité de l'air.

Nous avons choisi de modéliser 4 variables représentant les paramètres de pollutions, à savoir, le dioxyde d'azote(NO_2), le monoxyde d'azote(NO), le NO_x , et l'ozone(O_3). Ces paramètres de pollution ont été choisis pour leurs relation données par le cycle de Chapman(Figure 6.1). Dans l'application présentée ci-dessus, l'MSPCA classique et l'MSPCA non linéaire, que nous avons introduit à l'aide du réseau de neurone de Kramer [33], sont principalement comparés. La modélisation par ces deux méthodes nécessite la détermination à priori, de certains paramètres, qui sont la fonction d'ondelette, le niveau de décomposition avec la détermination du nombre de composante principale. Une fois ces paramètres fixés, la comparaison de ces deux méthodes pour la détection et la localisation de défaut est effectuée. Il est montré sur cette application que l'MSPCA non linéaire donne de bien meilleurs résultats que l'approche MSPCA classique.

Chapitre 7

Conclusion et Perspective

7.1 Conclusion

L'objectif de notre travail a été de trouver une méthode de contrôle statistique multi-variables qui permettra d'améliorer la vitesse de détection des événements anormaux à différentes échelles ainsi que la réduction du nombre de fausses alarmes. Pour cela nous avons présenté l'approche MSPCA non linéaire. Afin de comprendre cette approche on a introduit sur le premier chapitre l'analyse en ondelette à travers l'analyse multi-résolution qui permet d'analyser le signal sur différentes résolutions en le décomposant en coefficients d'approximation et coefficients de détails à différentes échelles. Comme toutes les variables doivent être contrôlées à différents niveaux, ceci nous a amené à utiliser les méthodes multi-variables qui sont non seulement capable de réduire la dimensionnalité du problème mais aussi de modéliser les relations entre les variables. L'ACP a été choisi pour sa simplicité et fiabilité à capturer les relations linéaires entre les variables du processus à l'état stationnaire. La complémentarité de l'approche ACP, ondelette a été montré pour la première fois par Bakshi [5]. La discrétisation dyadique du paramètre de translation permet l'utilisation d'ondelette orthonormale, et approximativement décorréler les mesures autocorrélées. L'ACP par contre permet d'extraire l'intercorrélation linéaire entre les capteurs. Nous avons introduit l'ACP non linéaire (à l'aide des réseaux de neurones auto-associatif à 5 couche, approche proposé par Kramer [33]) dans l'approche MSPCA puisque la plupart des processus réel sont de nature non linéaire. Ceci étendra l'extraction de la corrélation non linéaire entre les variables. L'application des paramètres de pollution aux différentes approche montre les faiblesse des méthode ACP dynamique et ACP conventionnel pour la détection de défaut. L'introduction de l'approche MSPCA nécessite la détermination du type d'ondelette, niveau de décomposition L et du nombre de composantes principales à retenir. Les ondelette daubechies ont été choisis comme ondelette mère pour leurs propriétés d'orthonormalité, et le haut nombre de moment de fuite. Ces propriétés permettent une reconstruction précise et efficace du signal original, une facilité de calcul, et le stockage d'un grand nombre d'information dans un nombre réduit de coefficients. La décision d'utiliser

les ondelettes Daubechies a été aussi influencée par différents chercheurs [50],[43].

L'application de l'MSPCA non linéaire à montré des résultats quasiment similaire à l'approche MSPCA de Bakshi, ceci peut être expliqué par le peu de non linéarité présent sur les paramètres de pollutions. Afin d'améliorer les performance de détection, nous avons décidé de rendre le modèle MSPCA non linéaire dynamique en ajoutant des entrée décallées. Bien que la décomposition en ondelette permette d'approximativement décorrélérer l'autocorrélation présente dans chaque capteur, l'utilisation de donnée dynamique avec l'ACP non linéaire à permis une meilleurs Détection et localisation du défaut lors des simulations. Il est à noté que la surveillance des processus en se basant seulement sur l'ACP des coefficients d'ondelette et des coefficients échelon ne permettra pas une détection continue et rapide d'une déviation, et pourra créer des fausses alarmes après que le processus soit retourné au fonctionnement normal. Les graphes de contrôle pour le signal reconstruit, améliore la vitesse de détection des évènements anormaux et élimine les fausses alarmes après que le processus soit retourné à l'état normal.[4]

L'application de l'MSPCA non linéaire dynamique aux signaux de capteurs d'une station de surveillance de la qualité de l'air dans la région d'Annaba a démontré clairement les améliorations qu'on peut atteindre en terme de sensibilité par rapport à l'approche MSPCA classique, pour la détection des disfonctionnement anormaux, avec moins de fausse alarmes, et en identifiant la source potentiel du défaut à l'aide des graphes de contribution.

7.2 Perspective

Dans notre application nous nous somme intéressé à des défauts de capteurs de type fenêtre, ce type de défaut est considéré comme une dérive de capteur et pourra être détecté dans les coefficients d'approximation ou bien les hautes échelles, une reconstruction des coefficients au domaine temporelle à permis de détecter ce type de défaut. Par contre avec l'approche qu'on a développée on ne pourra pas détecter les évènements à hautes fréquence puisque nous avons assigné toutes les échelles à zéro sachant qu'on les a considérées comme du bruit. Le dépassement des limites de contrôle sur toutes les échelles nous a ammené à ne pas tenir compte des échelles de détails. Pour cette raison l'une des priorités dans nos futurs travaux, est de prévoir d'utiliser le filtrage par ondelette afin de supprimer le bruit présent sur les coefficients de détails des différentes échelles ce qui permettra d'avoir une reconstruction moins bruitée, et permettra de détecter tout évènement haute fréquence. La modification de l'approche MSPCA non-linéaire pour une utilisation en ligne avec l'utilisation de limites de contrôle non paramétrique basé sur le principe de Kernel density estimation sera aussi une voie intéressante.

Annexe A

Apprentissage du réseau

A.1 Algorithme séquentiel

1. Projeter les données X dans un espace de dimension 1 des composantes principales en utilisant une structure de réseau à cinq couches.
2. Apprendre le réseau et estimer \hat{X} par projection inverse.
3. Calculer l'erreur d'estimation $E = X - \hat{X}$. Repeter les étapes 1 et 2 en prenant la matrice E à la place de X jusqu'à ce qu'un pourcentage de la variance de X soit capturé.

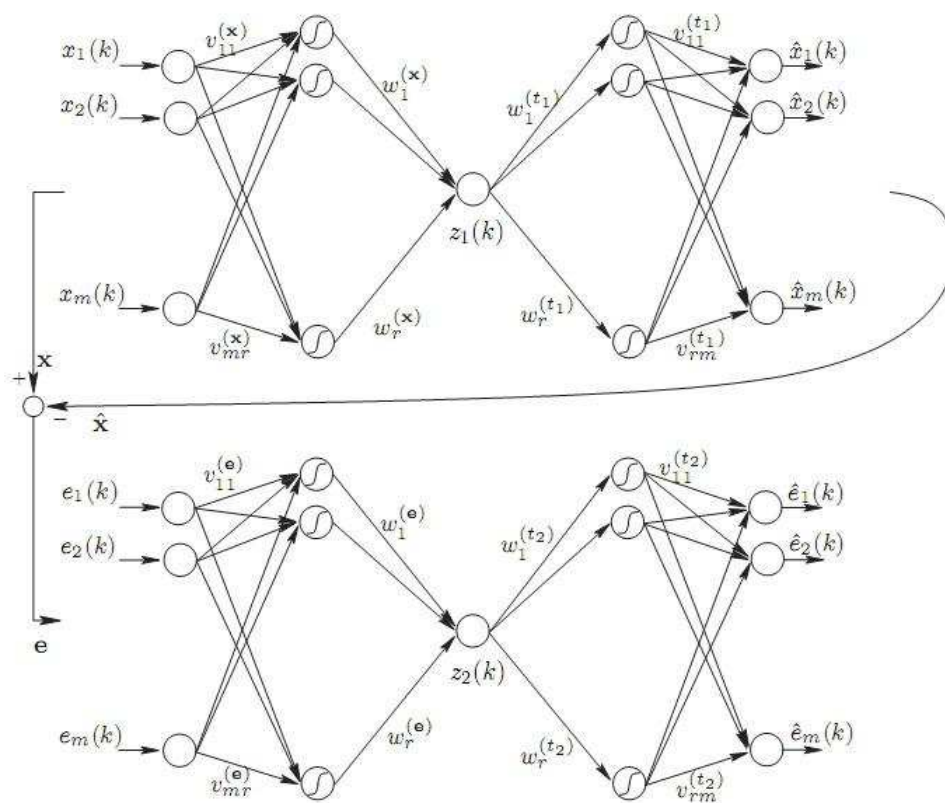


FIGURE A.1 – Extraction séquentielle des composantes principales non linéaire

A.2 Algorithme parallèle

1. Projeter les données X dans un espace de composante parincipales de dimension q en utilisant une structure de réseau à cinq couches.
2. Apprendre le réseau et estimer les données \hat{X} par projection inverse.

Les deux algorithmes ont leurs avantages et inconvénients. L'avantage principal de l'algorithme parallèle est qu'il est plus rapide à apprendre en le comparant à l'algorithme séquentiel : pour q composantes, un seul réseau de neurone est nécessaire dans le cas parallèle, alors qu'il faut q réseaux dans le cas séquentiel. L'inconvénient de l'algorithme parallèle, réside dans la première étape. Cette étape suppose que le nombre de composantes principales doit être connu à priori. $z_1(k), z_2(k), z_q(k)$

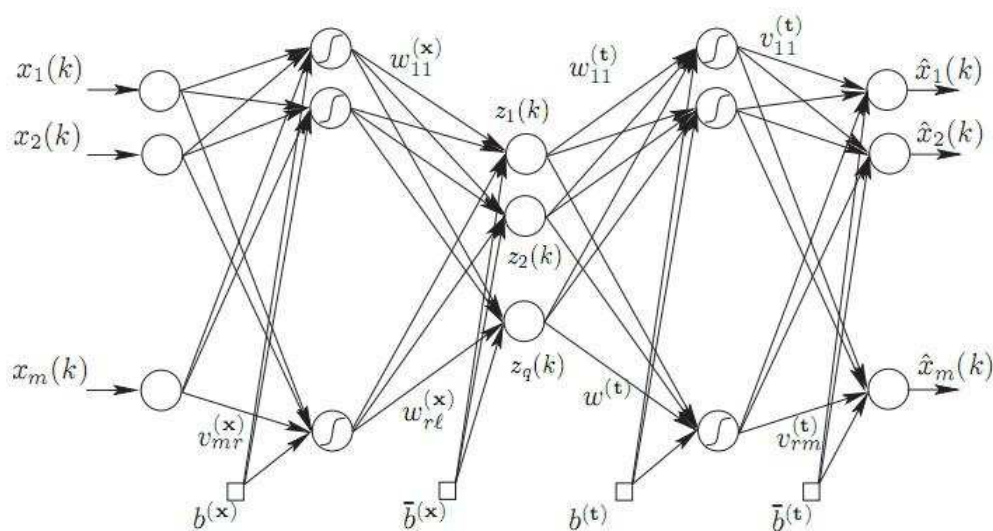


FIGURE A.2 – Extraction simultanée des q composantes principales non linéaires

Bibliographie

- [1] Willsky. A.S. A survey of design methods for failure detection in dynamic systems. *Automatica*, 12(6) :601–611, 1976.
- [2] Kuipers. B. Qualitative simulation. *Artificiel Intelligence*, 29(3) :289–338, 1986.
- [3] Schölkopf B., Smola A.J., and Müller. K.-R. Nonlinear component analysis as a kernel eigenvalue problem. *Neural computation*, 10 :1299–1319, 1998.
- [4] Bakshi B.R. Multiscale pca with application to multivariate statistical process monitoring. *AIChE Journal*, 44(7) :1596–1610, 1998.
- [5] Bakshi B.R. Multiscale analysis and modeling using wavelets. *Journal of chemometrics*, 13 :415–434, 1999.
- [6] Dayal B.S and MacGregor J.F. Recursive exponentially weighted pls and its applications to adaptive control and prediction. *Journal of Process Control*, 7(3) :169–179, 1997.
- [7] Rosen C. and Lennox. J. A. Multivariate and multiscale monitoring of wastewater treatment operation. *Water Science and Technology*, 35 :3402–3410, 2001.
- [8] Stork C.L., Veltkamp D.J, and Kowalski. B.R. Identification od multiple sensor disturbances during process monitoring. *Analytical Chemistry*, 69(24) :5031–5036, 1997.
- [9] Dong D. and McAvoy T.J. Nonlinear principal component analysis–based on principal cureves and neural networks. *Computer Chemical Engineering*, 20(1) :65–78, 1996.
- [10] Wang D. and romagnoli. J.A. Robust multiscale principal component analysis with applications to process monitoring. *Journal of Process Control*, 15 :869–882, 2005.
- [11] Hui Ding, Liu Jun-hua, and Shen. Zhong-ru. Drift reduction of gas sensor by wavelet and principal component analysis. *Sensors and actuators*, 96 :354–363, 2003.
- [12] Lee D.S., Park J.Moon, and Vanrolleghem. Peter A. Adaptive multiscale principal component analysis for on-line monitoring of a sequencing batch reactor. *Journal of Biotechnology*, 116 :195–210, 2005.

- [13] Chow E.Y. and Willsky A.S. Analytical redundancy and the design of robust failure detection systems. *IEEE Trans.on automat.Contr*, 7 :603–614, 1984.
- [14] Cybenko G. Approximation by superposition of sigmoidal function. *Mathematics of control Signal and Systems*, 2 :303–314, 1989.
- [15] Box G.E.P. Some theorems on quadratique forms applied in the study of analysis of variance problems : Effect of inequality of variance in one-way classification. *The Anals of Mathematical Statistics*, 25 :290–302, 1954.
- [16] Aradhye H.B., Bakshi B.R., Strauss Ramon A., and Davis James F. Multiscale spc using wavelets-theoretical analysis and properties. *AIChE Journal*, 49 :939–958, 2003.
- [17] Daubechies I. Orthonormal bases of compactly supported wavelets. *Communications of Pure and Applied Mathematics XII*, 909, 1988.
- [18] Gertler J. Fault detection and isolation using parity relations. *Control Engineering Practice*, 5(5) :653–661, 1997.
- [19] Gertler J., Weihua L., Yunbing H., and McAvoy T.J. Isolation enhanced principal component analysis. In *3rd IFAC workshop on Online Fault Detection and Supervision in the Chemical Process Industries,Lyon, June 4-5 France*, 1998.
- [20] Gertler J., Luoand Anderson Q., K., and Fang X.W. Diagnosis of plant failures using orthogonal parity equations. In *"IFAC world congress ,Tallin"*, 1990.
- [21] Lennox J. and Rosen. C. Adaptive multiscale principal components analysis for on-line monitoring of wastewater treatment. *Water Science and Technology*, 45 :N° 4–5 pp 227–235, 2002.
- [22] Shiozaki J., Matsuyama H., O'shima E., and IRI ; M. An algorithm for diagnosis of system failures in chemical process. *Computer Chemical Engineering*, 3(1-4) :489–493, 1979.
- [23] Fussel J.B. Fault tree analysis-state of the art. *IEEE Trans. On reliability*, 23(1) :51–53, 1974.
- [24] MacGregor J.F and Kourti. T. Statistical process control of multivariate processes. *Control Engineering Practice*, 3 :403–414, 1995.
- [25] Cheung J.T. and Stephanopoulos. G. Representation of process trends part i.a formal representation framework. *Computer Chemical Engineering*, 14(4-5) :495–510, 1990.
- [26] Kresta J.V., MacGregor J. F, and MARLIN. T. E. Multivariate statistical monitoring of process operating perormence. *Can. J. of Chem.Eng*, 69(1) :35, 1991.

- [27] Kosanovich K.A. and Piovoso. M.J. Pca for wavelet transformed process data for monitoring. *Intelligent Journal of Data Analysis*, 1(2) :85–99, 1997.
- [28] Harkat M., Mourot G., and Ragot. J. An improved pca scheme for fdi : Application to an air quality monitoring network. *Journal of Process Control*, 2006.
- [29] Harkat M., Mourot G., and Ragot. J. Sensor fault detection and isolation of an air quality monitoring network using non linear pca. In *16th IFAC World Congress, Prague*, 2006.
- [30] Harkat M., Djelal S., Doghmane N., and Benouareth M. Sensor fault detection, isolation and reconstruction using non linear pca. In *IFAC symposium on fault detection, Supervision and Safety for Techincal Process, Pekin, Chine*, 2007.
- [31] Misra M., Henry Yue H., joe Qin S., and Ling. Cheng. Multivariate process monitoring and fault diagnosis by multiscale pca. *Computer Chemical Engineering*, 26 :1281–1293, 2002.
- [32] Scholz. M. *Nonlinear PCA based on neural networks*. PhD thesis, Dep of Computer science, Humboldt-University Berlin, 2006.
- [33] Kramer. Mark A. Nonlinear principal component analysis using autoassociative neural networks. *AIChE Journal*, 37 :233–243, 1991.
- [34] Nounou M.N. and Bakshi. M.N. On-line multiscale filtering of random and gross error without process models. *AIChE Journal*, 45 :1041–1055, 1999.
- [35] Miller P., Swanson R.E., and Heckler. C.F. Contribution plots :the missing link in multivariate quality control. *37th Annual Fall Conference.ASQC, Rochester, NY*, 1993.
- [36] Nomikos P. and MacGregor. J.F. Monitoring batch processes using multiway principal component analysis. *AIChE Journal*, 40(8) :1361, 1994.
- [37] Frank P.M. Fault diagnosis in dynamic system using analytical and knowledge based redundancy -a survey ans some new results. *Automatica*, 26(3) :459–474, 1990.
- [38] Dunia R. and Qin S.J. Joint diagnosis of process and sensor faults using principal component analysis. *Control Engineering Practice*, 6 :457–469, 1998.
- [39] Dunia R. and Qin S.J. A subspace approach to multi-dimensional fault identification and reconstruction. *AIChE Journal*, 44 :1813–1831, 1998.
- [40] Dunia R. and Qin S.J. A unified geometric approach to process and sensor fault identification and reconstruction :the unidimensional fault case. *Computer Chemical Engineering*, 22 :927–943, 1998.
- [41] Dunia R., Qin S.J., Edgar T.F, and McAvoy T.J. Identification of faulty sensor using principal component analysis. *AIChE Journal*, 42(10) :2797–2812, 1996.

- [42] Dunia R., Qin S.J, Edgar T.F, and McAvoy T.J. Use of principal component analysis for sensor fault identification. *Computer and Chemical Engineering*, 20 :pp S713–S718, 1996.
- [43] Shao R., Jia F., Martin E.B., and Morris. A.J. Wavelets and non-linear principal components analysis for process monitoring. *Control Engineering Practice*, 7 :865–879, 1999.
- [44] Dijkerman R.W. and Majumdar R.R. Wavelet representations of stochastic process and multi-resolution stochastic models. *IEEE Trans.Signal processing*, 42 :1640, 1994.
- [45] Soylemez S. and Seider. S. A new technique for precedence-ordering chemical process equation sets. *AICHE Journal*, 19(5) :934–942, 1973.
- [46] Tan S. and Mavrovouniotis. M.L. Reduction data dimensionality through optimizing neural network inputs. *AICHE Journal*, 41(6) :1471–1480, 1995.
- [47] Wold S., Kettaneh N., and Tjessem. K. Hierarchical multiblock pls and pc model for easier model interpretation and as an alternative to variable selection. *Journal Chemometrics*, 10 :463, 1996.
- [48] Lapp S.A. and Powers. G.A. Computer-aided synthesis of fault trees. *IEEE Trans.on reliability*, 26(1) :2–13, 1977.
- [49] Mallat. S.G. A theory for multiresolution signal decomposition :the wavelet representation. *IEEE Transactions on Pattern Analysis and machine intelligence*, 11(7) :674–693, 1989.
- [50] Fourie S.H and Vaal de P. Advanced process monitoring using an on-line non-linear multiscale principal component analysis methodology. *Computer and Chemical Engineering*, 24 :755–760, 2000.
- [51] Qin S.J and Dunia. R. Determining of the number of principal components for best reconstruction. *Journal of Process control*, 10 :245–250, 2000.
- [52] Qin S.J. and Li. W. Detection ,identification, and reconstruction of faulty sensors with maximized sensitivity. *AICHE Journal*, 45(9) :1963–1976, 1999.
- [53] Choi S.W, Morris J., and Lee In-Beum. Nonlinear multiscale modeling for fault detection and identification. *Chemical Engineering Science*, 63 :2252–2266, 2008.
- [54] Hastie T. and Stuetzle. W. Principal curves. *Journal of the Americain Statistical Association*, 84 :502–516, 1989.
- [55] Kruger U., Antory D., Hahn J., Irwin G.W., and McCullough. G. Introduction of non-linearity measure for principal component models. *Computer and Chemical Engineering*, 29 :2355–2362, 2005.

- [56] Venkatasubramanian V., Rengaswamy R., Yin K., and Kavuri. S.N. A review of process fault detection and diagnosis part i : Qunatitative model-based methods. *Computer Chemical Engineering*, 27 :293–311, 2003.
- [57] Venkatasubramanian V., Rengaswamy R., Yin K., and Kavuri. S.N. A review of process fault detection and diagnosis part iii :process history based methods. *Computer Chemical Engineering*, 27 :327–346, 2003.
- [58] Venkatasubramanian V., Rengaswamy R., Yin K., and Kavuri. S.N. A review of process fault detection and diagnosis part ii :qualitative models and search strategies. *Computer Chemical Engineering*, 27 :313–326, 2003.
- [59] Ku W., Storer R.H., and Georgakis C. Disturbance detection and isolation by dynamic principal component analysis. *Chem.Intell.Lab.Syst.*, 30 :179, 1991.
- [60] Lin W., Qian Yu, and Li. Xiuxi. Nonlinear dynamic principal component analysis for on-line process monitoring and diagnosis. *Computer and Chemical Engineering*, 24 :423–429, 2000.
- [61] Ding X. and Frank P.M. Fault detection via factorization approach. *Syst.Contr.Lett.*, 14(5) :431–436, 1990.
- [62] Huang Y. and Gertler. J. Fault isolation by partial pca and partial nlpca. *IFAC'99, 14th Triennial world congress, Beijing, P.R. China*, pages 545–550, 1999.
- [63] Huang Y., Gertler J., and McAvoy T. Sensor and actuator fault isolation by structured partial pca with non-linear extention. *Journal of Process Control*, 10 :459–469, 2000.