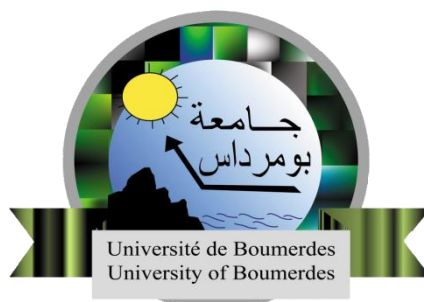


République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITE M'HAMED BOUGARA-BOUMERDES



Faculté des Sciences de l'Ingénieur

Mémoire de Master

Présenté par :

- Mr ROUIDJALI Mohammed Amine
- Mr CHAIMI Abdelkrim

En vue de l'obtention du diplôme de **Master** en
Génie Electrique
Option : Automatique

Thème :

**Détection des défauts basés sur l'Analyse
en Composantes Principales
Application à un Four électrique, à un
Groupe électrogène et à une Chaudière**

Président

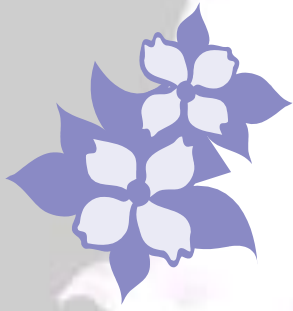
Rapporteurs

Mr IKHLEF Boualem MAA

UMBB

Examineurs

Promotion Juin 2017



Remerciements

C'est avec humilité et gratitude que nous reconnaissons ce que nous devons :

Nous commencerons par remercier et rendre grâce à Dieu tout puissant pour nous avoir donné le courage et la volonté de mener à bon terme ce travail.

Nos remerciements vont en particulier à Mr Ikhllef Boualem, notre promoteur qui n'a jamais ménagé son temps et ses efforts pour nous prodiguer ses conseils.

À Mr Menad Tadriss, notre encadreur pour son aide inestimable et ses précieux conseils.

Notre grand respect à nos professeurs, transmetteurs du savoir, qui nous ont aidé à soutenir notre volonté de réaliser ce mémoire, ne serait ce, que par un encouragement, nous leurs disons mille merci.

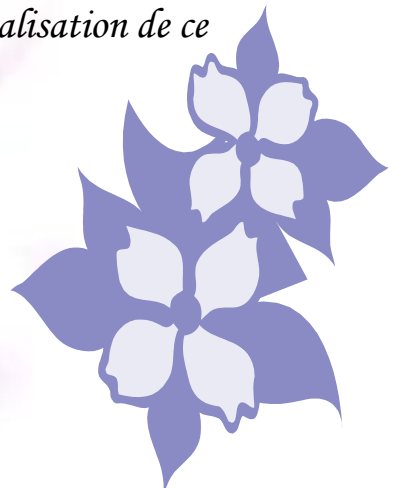
Tous nos remerciements à l'ensemble du personnel de la raffinerie GRD Labelle pour leurs collaborations et leur accueil chaleureux, en particulier Mr Adel, Farouk Aboubakar, et tout le groupe de l'atelier électrique.

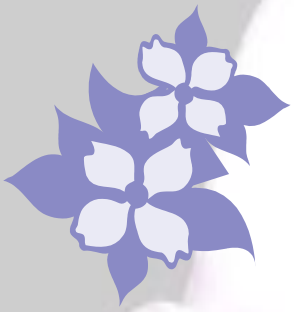
À nos amis qui nous ont été d'un grand secours moral durant tout ce stage qu'ils trouvent ici nos vifs remerciements

Aux membres du jury, pour avoir bien voulu accepter de juger ce travail et lui apporter les corrections nécessaires.

Enfin nos remerciements vont à tous ceux qui dans l'anonymat et dans la grandeur de leur modestie ont contribué de loin ou de près à la réalisation de ce travail.

Pour vous tous merci beaucoup





Dédicaces

Pour que ma réussite soit complète je la partage avec tous les personnes que j'aime, je dédie ce modeste travail à :

À la lumière de ma vie, mes très chers parents pour leur amour et pour le courage et la volonté qu'ils m'ont inculqué.

À ma seule très chère sœur Siham pour sa patience, son courage, son amour, et le soutien qu'elle m'a donnée.

À tout la famille ROUIDJALI et la famille GEURBI.

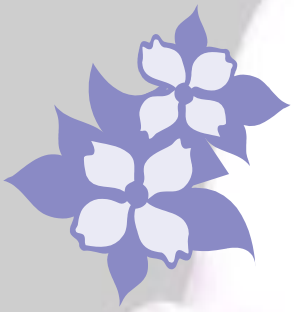
À mon binôme Abdelkrim ainsi que toute sa famille.

Et à tous mes très chers amis (es), et particulièrement Aribi Yacine et à mon ami d'enfance Bouamra Sofiane.

À tous mes collègues de la promotion 2017



Mohammed Amine



Dédicaces

Pour que ma réussite soit complète je la partage avec toutes les personnes que j'aime, je dédie ce modeste travail à :

À la mémoire de Mon père qui nous a quitté trop tôt dans l'âge.

À la lumière de ma vie, ma très chère maman pour l'amour, le courage et la volonté qu'elle m'a inculqué.

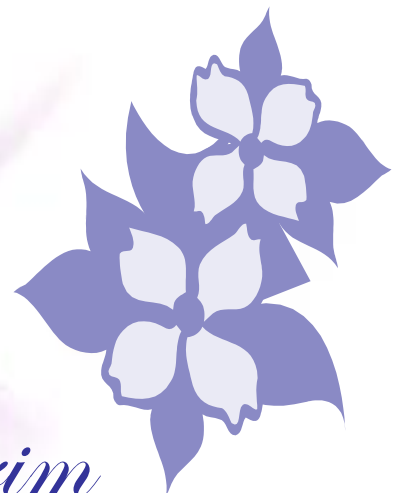
À mes très chères sœurs Mounira, Salima et Asma pour leur patience, eur courage, leur amour, et leur soutien qu'elles m'ont donnée.

À tout la famille CHAIMI et la famille SADAOUI.

À mon binôme Mohammed Amine ainsi que toute sa famille.

Et à tous mes très chers amis (es), et particulièrement Aribi Yacine et à mes deux amis d'enfance Rquidjali Rachid et Djellal Amir.

À tous mes collègues de la promotion 2017



Abdelkrim

ملخص

تحليل المكونات الرئيسية (ACP) هي تقنية معروفة وتستخدم على نطاق واسع الأبعاد من البيانات في الكشف عن العيوب في العمليات الصناعية. في حالتنا نستعملها في ثلاث أنظمة صناعية تتمثل في : فرن دوار, مولد كهربائي و غلاية صناعية للكشف عن مختلف العيوب في هذه الأنظمة. لقد أظهرنا في نتيجة عملنا، كيف يمكننا استخدام هذه التقنية لتمثيل البيانات متعددة الأبعاد و كشف العيوب في هذه البيانات. تظهر نتائج الاختبار التي تم الحصول عليها من فعالية الخيارات المعتمدة استخدام نهج حسن اختيار وسيلة واعدة لأنظمة دعم التشخيص.

الكلمات المفتاحية : تحليل المكونات الرئيسية - عطب - كشف العيوب - انذارات كاذبة

Résumé

L'analyse en composantes principales (ACP) est une technique de dimensionnalité des données bien connue qui est largement utilisée dans la détection des défauts des processus industriels. Dans notre cas, nous l'avons utilisée pour faire de la détection des défauts dans les trois systèmes industriels : un Four électrique, un Groupe électrogène et une Chaudière. Nous avons démontré dans la suite de notre travail, comment on peut se servir de cette technique pour représenter les données multidimensionnelles et détecter les défauts dans ces données. Les résultats de test obtenus montrent l'efficacité du choix adopté. L'utilisation de la démarche ainsi retenue, constitue une voie prometteuse pour les systèmes d'aide au diagnostic.

Mots clés : Analyse en composantes principales – défaut – détection des défauts – Fausses alarmes.

Abstract

Principal Component Analysis (PCA) is a well-known technique of data dimensionality that is widely used in the detection fault in industrial processes. In our case, we used it to make detection of fault in the three industrial systems: an rotary kiln, a generator and a boiler. We have demonstrated in the rest of our work how this technique can be used to represent multidimensional data and to detect fault in these data. The test results obtained show the effectiveness of the choice adopted. The use of this approach is a promising avenue for diagnostic support systems.

Key words : principal component analysis – fault – faults detection - false alarms.

Sommaire

Introduction générale	1
Chapitre 1 : Présentation de la raffinerie du sucre GRD Labelle et état de l'art de diagnostic	
1.1 Introduction.....	2
1.2 Présentation de la raffinerie du sucre GRD Labelle.....	2
1.2.1 Situation géographique	3
1.2.2 Process	3
1.2.3 Alimentation de l'usine.....	3
1.3 Etat de l'art de diagnostic.....	3
1.3.1 Détection des défauts et diagnostic	3
1.3.1.1 Quelques Définitions.....	3
1.3.1.2 Les fonctions de surveillances.....	7
1.3.1.3 Classification des défauts	8
1.3.1.4 Les types des défauts.....	9
1.3.1.5 Caractéristiques des défauts	9
1.3.1.6 Caractéristiques souhaitables d'un système de diagnostic de défaut.....	11
1.3.2 Classification des méthodes de diagnostic.....	12
1.3.2.1 Méthodes de détection des défauts basées sur un modèle.....	12
1.3.2.2.1 Méthode quantitatives basée sur modèle de défaut	12
1.3.2.2.2 Méthodes qualitatives basées sur modèle de défaut	13
1.3.2.2 Approches basées sur des données de diagnostic de défaut.....	14
1.3.2.2.1 Méthodes basées sur des données qualitatives	14
1.3.2.2.2 Méthodes basées sur des données quantitatives	15
1.4 Conclusion.....	17
Chapitre 2 : Analyse en composantes principales ACP	
2.1 Introduction	18
2.2 Principe de l'ACP.....	18
2.3 Sélection de la dimensions du modèle	23
2.3.1 Règle de Kaiser – Guttman	23

2.3.2	La méthode Scree plot	23
2.3.3	Pourcentage cumulé de la variance (PCV)	24
2.3.4	La méthode bâtons brisés (BB).....	25
2.3.5	Analyse parallèle (AP)	25
2.3.6	Validation croisée (VC)	26
2.3.7	Méthode de bootstrap.....	27
2.4	Les inconvénients principaux de l'APC	27
2.5	ACP réursive (ACPR) et fenêtre mobile ACP (FMACP)	28
2.6	ACP robuste	29
2.7	L'ACP à Noyau	30
2.8	Détection des défauts basée sur l'ACP	31
2.9	Conclusion.....	33

Chapitre 3 : Description des systèmes et l'application de l'ACP

3.1	Introduction	34
3.2	Présentation des systèmes utilisés	34
3.2.1	Groupe électrogène SDMO GT1900	34
3.2.1.1	Définition	34
3.2.1.2	Utilisation	35
3.2.1.3	Fonctionnement.....	35
3.2.1.4	Rendement.....	35
3.2.1.5	Equipements d'automatisme	35
3.2.2	Four électrique	36
3.2.2.1	Définition	36
3.2.2.2	Fonctionnement.....	36
3.2.3	Chaudière	37
3.2.3.1	Définition	37
3.2.3.2	Chaudière à tubes d'eau	38
3.2.3.3	Fonctionnement de chaudière.....	38
3.2.3.4	Brûleur.....	40
3.3	Application de l'ACP sur les trios systèmes	40
3.4	Détermination du nombre de composants retenus (CP).....	41
3.5	Conclusion.....	42

Chapitre 4 : Résultats et discussions

4.1	Introduction	43
4.2	Test de L'ACP sur les trois systèmes à l'état de fonctionnement normal (sans défaut) ..	43
4.3	Résultat de la simulation des fausses alarmes	47
4.3.1	fausses alarmes de four électrique	47
4.3.2	fausses alarmes de la chaudière	49
4.3.3	fausses alarmes de Groupe électrogène	51
4.4	Test de L'ACP sur les trois systèmes à l'état de fonctionnement anormal (avec défaut) .	52
4.5	Conclusion.....	55
	Conclusion générale	56

Liste des figures

Figure 1.2 : Les différents types de diagnostic.....	7
Figure 1.3 : Défauts additif	8
Figure 1.4 : Défauts multiplicatifs.....	8
Figure 1.5 : Caractéristiques de défauts abrupts	9
Figure 1.6 : Caractéristiques de défauts intermittents	10
Figure 1.7 : Caractéristiques de Défauts graduels	10
Figure 1.8 : Les Méthodes de Diagnostic.....	17
Figure 2.1 : Illustration de l'ACP dans les données 2-D.	19
Figure 3.1 : Groupe électrogène SDMO GT1900	34
Figure 3.2 : Four rotatif	36
Figure 3.3 : Chaudière à tubes d'eau	37
Figure 3.4 : Fonctionnement d'une Chaudière à tubes d'eau	40
Figure 4.1 : Surveillance de Four électrique sans défaut basée sur ACP.....	43
Figure 4.2 : Surveillance de la chaudière sans défaut basée sur ACP.....	44
Figure 4.3 : Surveillance de Groupe sans défaut basée sur ACP.....	45
Figure 4.4 : La diminution des fausses alarmes de four électrique pour T ² de Hotteling.	48
Figure 4.5 : La diminution des fausses alarmes de four électrique pour Qstat.	48
Figure 4.6 : La diminution des fausses alarmes de la chaudière Pour T ² de Hotteling	49
Figure 4.7 : La diminution des fausses alarmes de la chaudière pour Qstat.	50
Figure 4.8 : La diminution des fausses alarmes de groupe pour T ² de Hotteling.....	51
Figure 4.9 : La diminution des fausses alarmes de groupe pour Qstat.....	51
Figure 4.10 : Surveillance de Four électrique avec un défaut en temps réel basée sur ACP. ...	52
Figure 4.11 : Surveillance de la chaudière avec défaut basée sur ACP.	53
Figure 4.12 : Surveillance de la chaudière avec défaut basée sur ACP.	53
Figure 4.13 : Surveillance de groupe avec défaut basée sur ACP.....	54
Figure 4.14 : Surveillance de groupe avec défaut basée sur ACP.....	54

Liste des Tableaux

Tableau 3.1 : Nombre de composants retenus (CP) de chaque système.	41
Tableau 4.1 : Limites de contrôle de Four électrique pour différent niveau de confiance.	44
Tableau 4.2 : Limites de contrôle de la chaudière pour différent niveau de confiance.	45
Tableau 4.3 : Limites de contrôle de groupe pour différent niveau de confiance.	46
Tableau 4.4 : Fausses alarmes de T^2 de Hotelling.	46
Tableau 4.5 : Fausses alarmes de Q_{stat}	46

Abréviations

ACP	: Analyse des composantes principales.
ACI	: Analyse en composantes indépendantes.
ACPR	: Analyse des composantes principales récursives.
ASCP	: Analyse statique des composantes principales.
ADCP	: Analyse Dynamique des Composants Principaux.
AP	: Analyse parallèle.
ACPN	: Analyse des composantes principales Noyau.
BB	: Bâtons brisés.
CQA	: Critère de qualité de l'ajustement.
CP	: Composantes principales (ou composants retenus).
SPE	: Erreur quadratique d'estimation (Squared Prediction Error).
EDCM	: Ensemble de données de construction de modèles.
FMACP	: Fenêtre mobile d'analyse des composantes principales.
LVP	: Logarithmique des valeurs propres.
MCP	: Moindres carrés partiels.
MSV	: Machines support vectorielles.
PCV	: Pourcentage cumulé de la variance.
SVD	: Décomposition en valeurs singulières.
VC	: Validation croisée.

Symboles

T^2 : La statistique du T-square d'Hotelling.

Q : Q-statistique (Erreur de prévision quadratique).

P : Matrice de chargement (Matrice de projection orthogonale).

a : Nombre de composantes retenues par modèle ACP.

X : La matrice des données originales.

P_a : Matrice de charge principale.

δ_u : L'écart type de la variable aléatoire u .

δ_{uv} : La covariance des variables aléatoires u et v .

δ_X : La matrice de covariance.

ρ_{uv} : La corrélation entre les variables aléatoires u et v .

ρ_X : La matrice de corrélation.

Z_α : La valeur de la distribution normale normalisée au niveau de confiance $1-\alpha$

Σ : La matrice des valeurs singulières.

I_m : La matrice d'identité.

Introduction générale

Dans les processus industriels, la majorité des pannes causant une baisse de la production est généralement due à des anomalies dans le procédé qui n'ont pas été corrigées à temps. En plus des contraintes de la qualité des produits et celles économiques, il y a plusieurs autres intérêts afin d'assurer un fonctionnement normal pour ces processus.

La détection, ou même l'anticipation d'une défaillance au début de son apparition peut éviter de grands dommages et catastrophes. Par conséquent, la détection et le diagnostic des défaillances de processus représentent un intérêt capital.

La complexité des systèmes et les exigences croissantes en termes de performance et de fiabilité nécessitent le développement des techniques de diagnostic systématique. Etant donné un système, de nombreux aspects de diagnostic peuvent être traités. Parmi ces aspects, on a la détection des défauts, elle revient simplement à partir d'observation et de connaissance sur le système à déterminer si ce dernier se comporte normalement ou non.

Les méthodes de diagnostic des systèmes industriels sont globalement distinguées en deux grandes familles:

- Les méthodes basées sur une modélisation des systèmes ou des signaux, que nous dénommerons « diagnostic quantitatif ».
- Les méthodes basées sur l'intelligence artificielle que nous appellerons « diagnostic qualitatif ».

Parmi les méthodes de détection qui sont utilisées ces dernières années, on trouve l'ACP (Analyse en Composantes Principales), tel que cette dernière est employée pour modéliser les comportements du système en fonctionnement normal et les défauts sont alors détectés en comparant le comportement observé avec celui de l'état normal ainsi qu'elle permet d'exploiter toutes les relations linéaires qui peuvent exister entre les différentes variables.

Ce mémoire est composé de quatre chapitres organisés comme suit :

Le premier chapitre est consacré à la définition et à la classification des défauts ainsi que des généralités sur le diagnostic et la détection des défauts.

Le deuxième chapitre sera consacré à la définition de la méthode de l'Analyse en Composantes Principales (ACP) ainsi que leur principe d'utilisation pour la détermination des composantes principales et la détection des défauts.

Le troisième chapitre aborde la présentation de l'entreprise GRD labelle et la description des systèmes utilisés pour l'application de notre méthode.

Le quatrième chapitre, réservé à la partie expérimentale. Nous mettons en œuvre les différents algorithmes de la méthode adoptée au chapitre 02 sur un groupe électrogène, une chaudière et une cimenterie présentés dans le chapitre 03.

Nous terminons notre travail par une conclusion générale.

Chapitre I :

*Présentation de la raffinerie du sucre GRD
Labelle et état de l'art de diagnostic*

1.1 Introduction

La complexité croissante des systèmes industriels, leur utilisation de plus en plus fréquente rendent nécessaire le diagnostic des défauts pouvant affecter le fonctionnement normal de ces mêmes systèmes.

Le diagnostic d'un système matériel désigne toute méthode permettant de déterminer si ce système est affecté par un défaut quelconque et de discriminer l'origine du défaut. Le diagnostic des systèmes s'effectue grâce à des algorithmes de détection utilisant les informations relevées par observation, contrôles et tests.

Il y a plusieurs manières de détecter des défauts et ils peuvent être divisés en deux classes : à base de modèle et à base des données. Pour la base de modèle, on doit extraire un modèle mathématique pour décrire le processus, où la base des données sont basées sur la disponibilité des données de système.

Le but principal de ce chapitre est d'obtenir une idée générale sur les défauts et de leurs classifications, ainsi que d'obtenir suffisamment de connaissances et d'informations pour pouvoir sélectionner et utiliser la meilleure approche ou modèle de détection et de diagnostic des défauts.

1.2 Présentation de la raffinerie du sucre GRD Labelle

La raffinerie de sucre GRD Labelle SPA est issue de la joint-venture entre la Société Labelle et le groupe coopératif français Cristal-Union (4ème producteur de sucre de betteraves et d'alcool en Europe ; 1.7 Millions Tonnes de sucre, 5.5 millions d'hls d'alcool ; 11500 Personnes dont 9300 agriculteurs ; 2 Milliards d'Euros de CA).

Le site est réparti en plusieurs zones depuis l'approvisionnement en sucre brut acheminé par bateau et stocké dans un silo de 60.000 T de capacité, le sucre est ensuite raffiné dans le bâtiment process, le produit fini est stocké dans un silo dôme d'une capacité de 30 milles tonnes. Une unité de traitement d'eau et une chaudière d'une capacité de 40 T/H.

La raffinerie de sucre d'OULED MOUSSA est entrée en production depuis le mois de Décembre 2015, avec une production en sucre blanc de 1000T/jour issu du raffinage du sucre de canne, en provenance principalement du Brésil ou d'Afrique subsaharienne, avec une masse salariale d'environ 400 personnes, une extension pour 2000 Tonnes/jours est prévue d'ici 2018.

1.2.1 Situation géographique

La raffinerie de sucre est située à OULED MOUSSA dans la wilaya de Boumerdes, à 35 Km à l'est d'Alger.

1.2.2 Process

Le process de raffinage du sucre consiste en une remise en solution des sucres bruts de cannes, une purification calcocarbonique du sirop suivie d'une décoloration puis d'une cristallisation.

Le sucre blanc cristallisé s'obtient enfin par centrifugation puis séchage. La conservation du sucre raffiné nécessite des conditions d'hygrométrie et de température spécifiques à ce produit compte tenu de sa pureté et de ses qualités à préserver.

Sur le plan énergétique, le sucre requiert de la vapeur produite par une chaudière à gaz. Le site produira 1.000 T. /J avec une capacité de 360.000 T/an. En premier lieu, le sucre sera commercialisé et vendu en sacs de 50 kgs, en big-bag de 1 t. et en camion-citerne, container vrac.

1.2.3 Alimentation de l'usine

L'usine possède trois sources d'énergie électrique qui sont asservies pour assurer la continuité de service dans le site. Pour une bonne gestion d'énergie électrique, le site possède un turboalternateur produit 6250 KVA qui nous permet d'avoir une autosuffisance d'énergie. Un groupe électrogène 1800 KVA une autre source d'énergie qui sert à être le secours en cas d'absence de GTA. SONALGAZ 30 KV est la troisième source d'énergie électrique qu'on peut utiliser pour assurer les besoins de site.

1.3 Etat de l'art de diagnostic

1.3.1 Détection des défauts et diagnostic

1.3.1.1 Quelques Définitions

Pour mieux étayer nos propos, nous introduisons quelques définitions utilisées dans le domaine du diagnostic, ce qui facilite la compréhension et la comparaison des différentes approches et méthodes utilisées dans ce domaine.

- **Fonctionnement normal d'un système** : Un système est dit dans un état de fonctionnement normal lorsque les variables le caractérisant (variables d'état, variables de sortie, variables d'entrée, paramètres du système) demeurent au voisinage de leurs valeurs nominales. Le système est dit défaillant dans le cas contraire.

- **Défaillance** : C'est la cause d'une anomalie.
- **Défaut** : Un défaut est un état dans le système. C'est une déviation en dehors d'un intervalle acceptable, d'une variable observée ou d'un paramètre associé au procédé. C'est-à-dire, un défaut est un processus anormal ou symptôme.
- **Erreur**: C'est un écart entre une valeur mesurée ou calculée d'une variable de sortie et sa valeur vraie ou théoriquement correcte.
- **Localisation d'un défaut** : A l'issue de la détection d'un défaut, il s'agit de déterminer le ou les éléments à l'origine du défaut.
- **Perturbation**: Il s'agit d'une entrée inconnue et incontrôlée agissant sur un système.
- **Symptôme**: Changement d'une quantité observable par rapport au comportement normal [1].
- **Résiduel**: un indicateur de défaut, basé sur l'écart entre les mesures et les calculs basés sur l'équation du modèle.
- **Intégrité**: L'intégrité d'un système est la capacité de détecter les défauts dans son propre fonctionnement et d'informer un opérateur humain.
- **Suivi**: La tâche réelle et continue afin de déterminer les conditions d'un système physique en enregistrant des informations, en reconnaissant et en indiquant des anomalies dans le comportement et le processus de surveillance.
- **Panne** : est un dysfonctionnement voire une défaillance autrement dit un arrêt de fonctionnement momentané et accidentel, d'une partie ou de tous les composants d'un système matériel.
- **Diagnostic de défaut**: C'est la détermination du type, la taille, la localisation de la panne, y compris l'isolation et l'identification des défauts d'un processus. Il comprend trois étapes successives :
 - ❖ la détection du défaut.
 - ❖ l'isolation du défaut.
 - ❖ l'identification du défaut.
- **Détection de défaut** : La détection d'un défaut consiste à décider si le système se trouve ou non dans un état de fonctionnement normal et aussi le temps de ce défaut[1].

- **Isolation des défauts:** Il s'agit de la détermination du type, de l'emplacement et de l'heure de détection du défaut suite à l'isolement des défauts.
- **Identification des défauts:** Il s'agit de la détermination des variables de processus responsables du défaut et donc de l'information permettant de diagnostiquer le défaut. Il aide à se concentrer sur les sous-systèmes où le défaut s'est produit dans une grande usine [1].
- **DéTECTABILITÉ des défauts :** Un défaut est détectable si, et seulement si, il a un effet sur au moins une des sorties du système.
- **Echec:** Un échec est une interruption permanente d'un système, la capacité à effectuer une nécessaire fonction dans des conditions de fonctionnement spécifiées [2].
- **Protection automatique:** Dans le cas d'un état de processus dangereux, la surveillance la fonction déclenche automatiquement une contraction appropriée. Habituellement, le processus est commandé à un état de sécurité, qui est souvent une coupure d'urgence.
- **Fiabilité :** La fiabilité est la capacité d'un composant, un procédé ou un système pour fonctionner correctement sous condition énoncée dans une portée donnée, au cours d'une période donnée de temps.
- **Sécurité:** La capacité d'un système de ne pas causer de danger pour les personnes ou l'équipement ou l'environnement.
- **Disponibilité:** Probabilité qu'un système ou l'équipement fonctionnent de façon satisfaisante et efficace au toute période de temps.
- **Observation :** Est une information obtenue à partir du comportement ou du fonctionnement réel du processus.
- **Signe :** est le caractère distinctif d'un état comportemental anormal. Il est caractérisé par un ensemble d'observations en provenance d'un système physique et révèle la présence d'un défaut.
- **La reprise:** Englobe la fonction de recouvrement et de décision, celle-ci détermine un état accessible pour le retour au nouveau fonctionnement normal et les différentes actions correctives modifiant la configuration de procédé et de la commande.

- **Une mesure** : est une observation 'élémentaire du fait qu'elle reflète une et une seule grandeur physique. Elle est représentée par une variable dont le contenu est l'image d'une grandeur physique. Son obtention s'effectue par l'intermédiaire d'un système physique (capteur). Ainsi, les mesures peuvent être de différentes natures :
 - ✓ **Qualitatives** : représentées par un ensemble fini de modalités dont chacune est associée à un symbole.
 - ✓ **Quantitatives** : dont les valeurs appartiennent à l'ensemble des réels.

1.3.1.2 Les fonctions de surveillances

Un système de surveillance reprend généralement à trois fonctions qui sont: la détection, le diagnostic et la reprise.

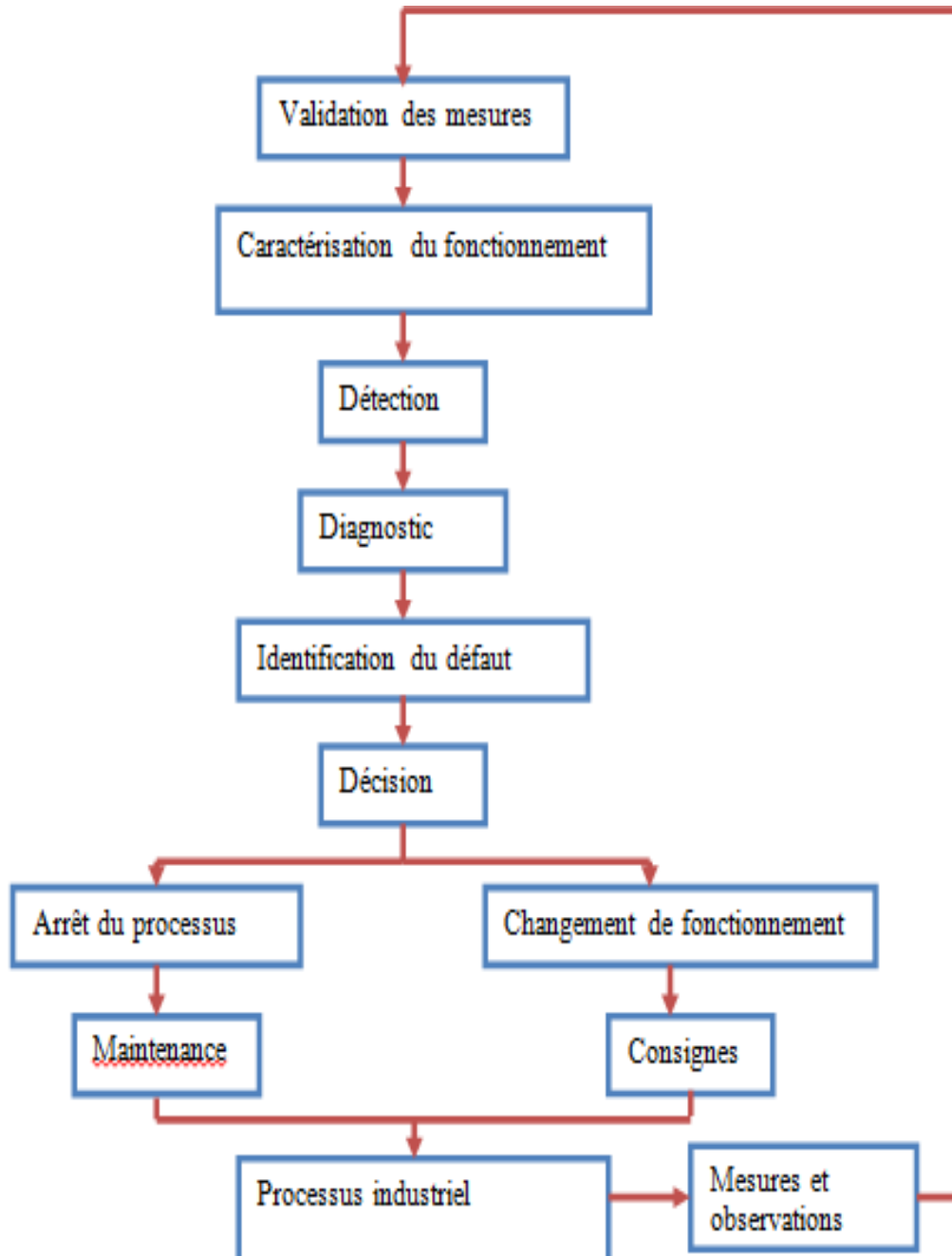


Figure1.2: les différents types de diagnostic

1.3.1.3 Classification des défauts

Une classification des défauts à partir de leurs évolutions temporelles est définie comme :

- **Défaut additif:** Ce sont des entrées connues agissant sur les installations qui sont normalement nulles et lorsqu'elles sont présentes peuvent provoquer une variation de la sortie de la centrale indépendamment des entrées connues [4].

Il peut être décrit mathématiquement comme:

$$S = E + \Delta E$$

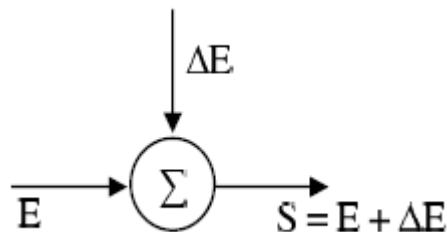


Figure 1.3 : Défaut additif

- **Défaut multiplicatif :** il affecte un système à travers la multiplication de l'entrée E du système par un défaut δx .

Un défaut δx affecte une variable x du système. L'équation de sortie $S = Ex$ est alors modifiée et en présence du défaut δx , elle s'écrit $S = E(x + \delta x)$.

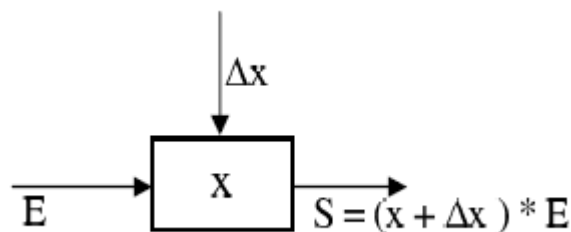


Figure 1.4 : Défaut multiplicatif

1.3.1.4 Les types des défauts

De manière générale, un système industriel est composé de trois parties :

- ✓ les capteurs.
- ✓ les actionneurs.
- ✓ le processus.

Les défauts peuvent survenir sur chacune de ces trois parties.

- **Défaut capteur :** ce type des défauts est la cause d'une mauvaise image de l'état physique du système. Un défaut capteur partiel produit un signal avec plus ou moins d'adéquation avec la valeur vraie de la variable à mesurer. Ceci peut se traduire par une réduction de la valeur affichée par rapport à la valeur vraie. Un défaut capteur total produit une valeur qui n'est pas en rapport avec la grandeur à mesurer [4].
- **Défaut actionneurs:** ces défauts agissent au niveau de la partie opérative et détériorent le signal d'entrée du système. Ils représentent une perte totale (défaillance) ou partielle d'un actionneur agissant sur le système. Les défauts actionneurs partiels sont des actionneurs réagissant de manière similaire au régime nominal mais en partie seulement, c'est-à-dire avec une certaine dégradation dans leur action sur le système.
- **Défaut composant ou « système » :** ce type des défauts provient du système lui-même, bien souvent les défauts n'appartenant pas à un défaut capteur ou actionneur sont classés de manière arbitraire dans cette catégorie. En pratique, ceci revient à considérer une modification des caractéristiques du système proprement dit, et ce sont les types les plus fréquemment rencontrés dans la famille de défaut à traiter [2].

1.3.1.5 Caractéristiques des défauts

Les défauts peuvent être caractérisés d'une part, selon leur comportement dans le temps et d'autre part selon leur modélisation mathématique.

- **Défauts abrupts :** la caractéristique principale de ce type de défauts est la discontinuité dans l'évolution temporelle de la variable.

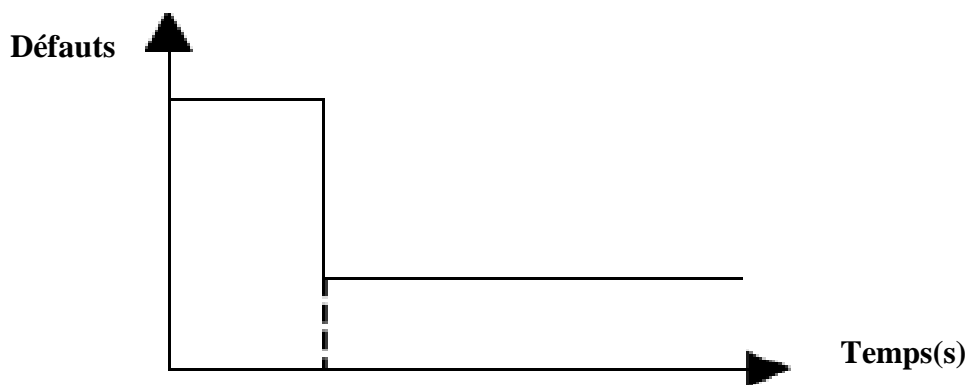


Figure 1.5 : Caractéristiques de défauts abrupts

- **Défauts intermittents** : il s'agit d'un type de défauts caractéristiques de faux contacts ou de pannes intermittentes de capteurs. C'est un cas particulier de défaut brutal sur un capteur avec perte aléatoire de signal.

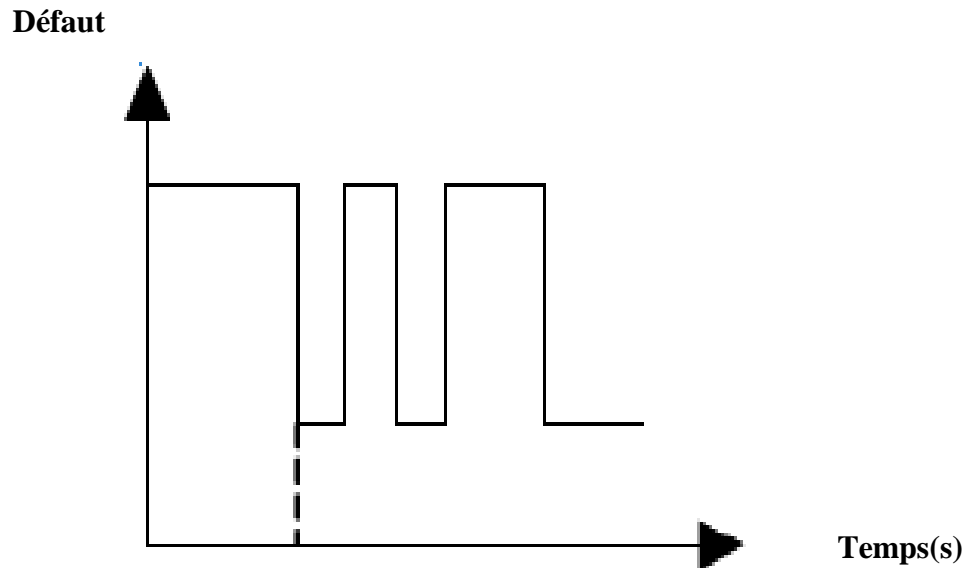


Figure 1.6 : Caractéristiques de défaut intermittent

- **Défaut graduel** : ce type de défauts est caractéristique d'un encrassement ou d'une dérive dans les paramètres caractéristiques du procédé. ils sont plus difficiles à détecter.

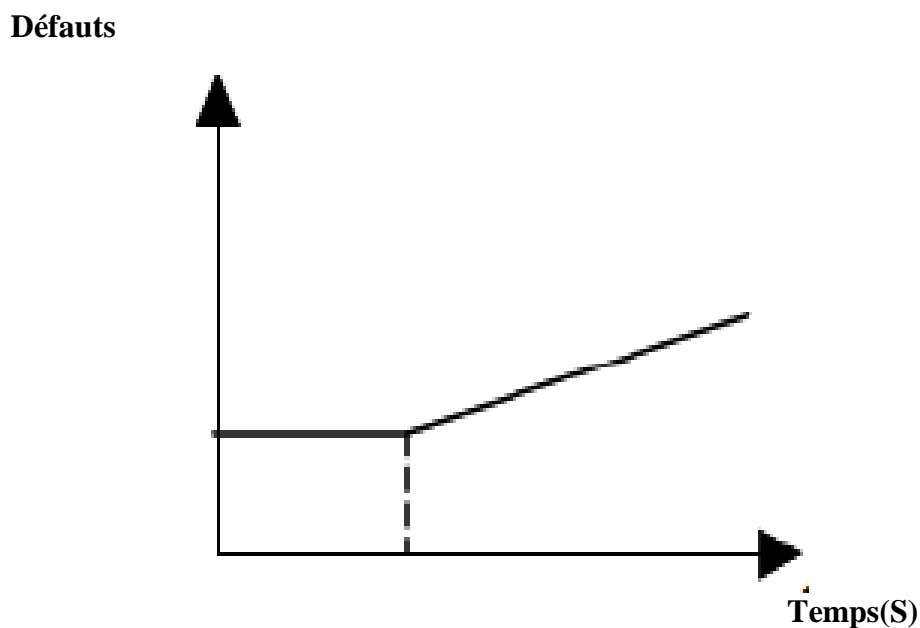


Figure 1.7: Caractéristique de Défaut graduel

1.3.1.6 Caractéristiques souhaitables d'un système de diagnostic de défaut

Il est utile pour un système de détection de défaut et de diagnostics d'avoir un ensemble de caractéristiques pour être reconnu comme une méthodologie efficace. Ces ensembles de caractères sont connus comme étant de caractéristiques souhaitables. Ils sont utilisés pour comparer les différentes approches de diagnostic. Il existe plusieurs caractéristiques qui sont considérées dans les détections des défauts et diagnostic, ces caractéristiques sont :

- **Détection rapide et diagnostic :** En général, il est souhaitable de réduire le délai entre l'heure de début du défaut et son temps d'apparition dans la surveillance du processus. En outre, le délai entre le temps d'apparition et le temps de détection du défaut dû à la décision doit être minimisé. Le système de détection des défauts et diagnostic devrait réagir le plus rapidement possible pour détecter une défaillance ou un dysfonctionnement et doit identifier rapidement les racines de l'anomalie [6].
- **Robustesse:** C'est la capacité du système à résister aux conditions environnementales comme la poussière ou l'humidité. Une robustesse en fonction des mesures bruit, perturbations du système, et modélisation est un attribut souhaitable d'un système de diagnostic destiné à des implémentations.
- **Adaptabilité :** En général, le fonctionnement des processus change en fonction de la condition de fonctionnement qui est modifiée en raison de perturbations et de conditions environnementales (y compris les variations des quantités de production). Ainsi, le système de diagnostic doit être adaptable aux changements.
- **Estimation de l'erreur de classification:** Mesures d'erreur serait utiles pour projeter la confiance et fiabilité des décisions de diagnostic prises par le système. Cette caractéristique de système de diagnostic plus recommandé.
- **Identification multiple des défauts :** Il s'agit de la capacité d'un système de diagnostic à identifier et à classer correctement les multiples défauts qui peuvent coexister dans le processus. L'identification de multiples défauts est une tâche importante et difficile puisque dans les systèmes non linéaires les interactions seraient synergiques et les détections des défauts et diagnostic peuvent modéliser l'effet combiné des défauts [6].

- **Exigences de stockage et de calcul:** La solution rapide et en temps réel requiert des algorithmes et une implémentation qui sont moins complexes en termes de calcul, mais qui peuvent nécessiter des besoins de stockage élevés. Souvent, un système de détection des défauts et diagnostic avec la capacité à équilibrer sur ces exigences est fortement nécessaire. [3]
- **Exigences de modélisation:** Pour un déploiement rapide et facile de classificateurs de diagnostic en temps réel, l'effort de modélisation doit être aussi minime que possible.

1.3.2 Classification des méthodes de diagnostic

1.3.2.1 Méthode de détection des défauts basées sur un modèle

Les méthodes basées sur des modèles reposent sur le concept de redondance analytique. L'essence de ce concept est la comparaison des sorties réelles du système surveillé avec les sorties obtenues à partir d'un modèle mathématique analytique (redondant, c'est-à-dire non physique). Elle comporte deux étapes: La production résiduelle et l'évaluation résiduelle. Cette approche suppose que la structure et les paramètres du modèle sont précisément connus [3].

1.3.2.2.1 Méthode quantitative basée sur modèle de défaut

Les approches basées sur des modèles quantitatifs ont été basées sur l'utilisation d'entrée-sortie générale et les modèles d'espace d'état pour générer des résidus. Ces approches peuvent être classées :

- **Observateur ou basé sur filtre :** L'idée de l'observateur ou des approches basées sur filtre est d'estimer les états ou les sorties du système des mesures. La liberté dans la conception de l'observateur peut être employée pour augmenter les résiduels pour l'isolement. L'approche géométrique et le contrôle adaptatif sont combinés avec succès avec les techniques basées sur observateur de détection et de diagnostic de défaut. [4]
- **Approches de l'espace de parité :** l'approche par espace de parité se base sur un test de cohérence entre le modèle du système (représentation d'état) et les mesures. Elle consiste à réécrire les équations d'état et de mesure afin de ne garder que des variables connues (commandes et sorties). Le principe est d'exprimer une mesure en fonction des autres par une relation connue.
- **Approches de domaine de fréquence :** Des résiduels sont également produits dans les domaines de fréquence par l'intermédiaire de la factorisation de la fonction de transfert du système surveillé.

1.3.2.2 Méthodes qualitatives basées sur modèle de défaut

Des modèles qualitatifs permettent à des conclusions d'être tirées au sujet de l'état d'un système avec la connaissance inachevée ou incertaine du processus physique. Ils sont basés sur de diverses formes de la connaissance qualitative utilisées dans le diagnostic de défaut, des approches basées sur modèle qualitatifs peuvent être classifiés comme suit:

- **Modèle causal des approches utilisant les diagrammes :** Un graphique dirigé signé ou un digraphe signé (GDS), comme modèle qualitatif, représente effectivement et graphiquement un système de processus. des relations ou les modèles de Cause-effet peuvent être représentés sous forme de digraphes signés. GDS peut être obtenu à partir des équations algébriques différentielles pour le processus.

La question des arcs conditionnels dans GDS est abordée, également l'idée de GDS a été prolongée pour inclure des modèles de cinq-gamme au lieu du modèle habituel de trois-gamme utilisé dans le GDS standard. La dynamique partielle de système, les informations statistiques sur la panne d'équipement, et les digraphes représentent le réseau de propagation d'échec pour identifier l'emplacement de défaut sont employés. Des méthodes basées sur les règles utilisant GDS ont été employées pour le diagnostic de défaut [4].

- **Approche par arbre de défaillance :** Des arborescences de défaillances sont employées en analysant la fiabilité du système et la sécurité. Les approches d'arborescence de défaillances sont la technique hiérarchisée d'analyse qui décrit les relations entre les événements de base (échecs, erreurs humaines, etc.), les états intermédiaires (modes opérationnels, états environnementaux, etc.) et les événements supérieurs (incidents, accidents). Les relations sont modélisées dans une structure comme un arbre avec la porte logique ET/OU. L'arborescence de défaillances numériques peut alors être analysée les ensembles minimaux de coupe (combinaison des événements et des conditions de base qui causent l'événement supérieur) et être employée pour déterminer la probabilité de l'événement supérieur.
- **Approches physiques qualitatives :** Les modèles physiques détaillés sont basés sur la connaissance détaillée des relations et des caractéristiques physiques de tous les composants dans un système. Les modèles détaillés peuvent simuler les états opérationnels normaux et défectueux du système. La connaissance physique qualitative dans le diagnostic de défaut a été représentée dans principalement deux approches, la première approche est de dériver des équations qualitatives des équations nommées comme équations de confluence. L'autre approche dans la physique qualitative est la dérivation du comportement qualitatif des équations ordinaires. Ces comportements qualitatifs pour différents échecs peuvent être employés comme source de la connaissance [4].

Les avantages principaux des modèles qualitatifs pour la détection et le diagnostic de défaut sont :

- ❖ Bien adaptés pour les environnements riches en données et les processus non critiques.
- ❖ Ces méthodes sont simples pour se développer et s'appliquer.
- ❖ Leur raisonnement est transparent, et ils fournissent la capacité de raisonner même sous l'incertitude.
- ❖ Ils possèdent la capacité de fournir des explications pour les diagnostics suggérés parce que la méthode se fonde sur des relations de cause-effet.
- ❖ Quelques méthodes fournissent la capacité d'exécuter Détection des défauts et diagnostic sans connaissance précise du système et d'exiger des valeurs numériques pour des entrées et des paramètres [4].

1.3.2.2 Approches basées sur des données de diagnostic de défaut

Il y a différentes façons dont les données peuvent être transformées et présentées comme a priori, la connaissance d'un système de diagnostic. Ceci est connu comme l'extraction de caractéristiques. Ce procédé d'extraction peut être qualitatif ou quantitative.

1.3.2.2.1 Méthodes basées sur des données qualitatives

Il existe deux méthodes importantes qui emploient l'extraction de caractéristiques qualitatives :

- **Système expert** : Un système expert est le système qui permet de résoudre les problèmes dans un étroit domaine d'expertise. Les éléments essentiels de ce développement de système sont: l'acquisition de connaissances, le choix de la représentation des connaissances, le codage des connaissances dans une base de connaissances, le développement de procédures d'inférence pour le raisonnement diagnostic et le développement d'interfaces d'entrée / sortie.
Ses avantages dans le développement de la résolution de problèmes diagnostiques sont: la facilité de développement, le raisonnement transparent, la capacité de raisonner dans l'incertitude et la capacité de fournir des explications pour les solutions fournies. [5]
- **Analyse qualitative des tendances** : L'analyse des tendances et la prévision sont des composantes importantes de la surveillance des processus et de la supervision.
La modélisation des tendances peut être utilisée pour expliquer les différents événements importants qui se produisent dans le processus, diagnostiquer le dysfonctionnement et prédire les états futurs.

Du point de vue de la procédure, afin d'obtenir une tendance de signal trop peu sensible aux variations momentanées dues au bruit, une certaine sorte de filtrage doit être employée. Par exemple, les représentations de séries chronologiques supposent, a priori, certains comportements lorsqu'ils sont identifiés en utilisant un comportement de processus connu. Alternativement, on peut simplement utiliser un filtre (tel qu'un filtre auto-régressif) avec des coefficients de filtre choisis à priori (en précisant le degré de lissage requis). Les deux types de filtres souffrent du fait qu'ils ne peuvent pas bien distinguer entre une instabilité transitoire et une véritable instabilité. [5]

1.3.2.2 Méthodes basées sur des données quantitatives

Les approches quantitatives formulent essentiellement la résolution de problème diagnostic comme un problème de reconnaissance de formes. Le but de la reconnaissance des formes est la classification des points de données, en général, des classes prédéterminées. Les méthodes statistiques utilisent la connaissance des distributions de classes a priori pour effectuer la classification.

Les méthodes qui extraient des informations quantitatives peuvent être classées en général comme des méthodes non statistiques ou statistiques. Les réseaux neuronaux sont une classe importante de classificateurs non statistiques. L'analyse des composantes principales (ACP), les moindres carrés partiels (MCP) et les classificateurs de modèles statistiques constituent une composante majeure des méthodes d'extraction de caractéristiques statistiques. [4,5]

- **Approches statistiques multi variées :** les techniques statistiques multi variées sont des outils puissants capables de compresser les données et en réduisant sa dimensionnalité de telle sorte que les informations essentielles sont conservé en plus facile à analyser que le vaste ensemble de données d'origine; et ils sont aussi capables de gérer le bruit et la corrélation pour extraire des informations vraies efficace .Les applications réussies de méthodes statistiques multivariées au diagnostic de défauts telles que l'analyse de composantes principales (ACP) et les moindres carrés partiels (MCP) ont été largement rapportées dans la littérature. L'analyse des composantes principales est l'une des méthodes les plus étudiées pour détecter les défauts des systèmes [5].
- ❖ **L'analyse factorielle :** est une méthode statistique utilisée pour décrire la variabilité entre les variables observées, variables corrélées en termes d'un nombre potentiellement inférieur de variables inobservées appelées facteurs.
L'objectif de l'analyse factorielle est d'expliquer le résultat de variables dans la matrice de données en utilisant moins de variables, les soi-disant facteurs.
- ❖ **L'Analyse en composantes principales (ACP):** transforme un certain nombre de variables éventuellement corrélées dans un ensemble de données dans un plus petit nombre de pseudo non corrélées ou variables latentes.

Cela se fait par une décomposition bilinéaire de la matrice de covariance de l'ensemble de données. Les variables non corrélées (orthogonales) obtenues sont appelées les composantes principales et elles représentent les axes obtenus par rotation du système de coordonnées d'origine suivant la direction de la variance maximale.

Les principales hypothèses de cette méthode sont que les données suivent une distribution gaussienne et que tous les échantillons sont indépendants l'un de l'autre. Puisque l'approche ACP est adoptée dans ce travail, une discussion complète sur les différentes approches de l'analyse des composantes principales est fournie dans le deuxième chapitre.

- ❖ **Moindre carrés partiels (MCP) :** est une réduction dimensionnelle ainsi qu'une régression technique qui trouve un nouvel ensemble de variables latentes qui maximisent la covariance entre la matrice de données d'entrée et la matrice de données de sortie. L'objectif principal ici est de rapprocher et en formes dimensionnelles réduites, ainsi que pour modéliser une relation linéaire entre eux.
- ❖ **Analyse en composantes indépendantes (ACI) :** est une statistique et informatique technique pour révéler des facteurs cachés qui sous-tendent des ensembles de variables aléatoires des mesures ou des signaux.
- **Approches classificatrices statistiques:** le diagnostic de défaut est essentiellement un problème de classification et peut donc être jeté dans un cadre statistique classique de reconnaissance de formes. Le diagnostic de défaut peut être considéré comme un problème de combinaison, au cours du temps, des estimations instantanées du classificateur en utilisant la connaissance des propriétés statistiques des modes de défaillance du système. [4]
- **Réseau Neural :** Les réseaux de neurones ne nécessitent pas de connaissances spécifiques de structure du processus. Ils contiennent de nombreux paramètres, mais ces paramètres ne conviennent généralement pas pour l'interprétation physique du système modélisé. [4,5]

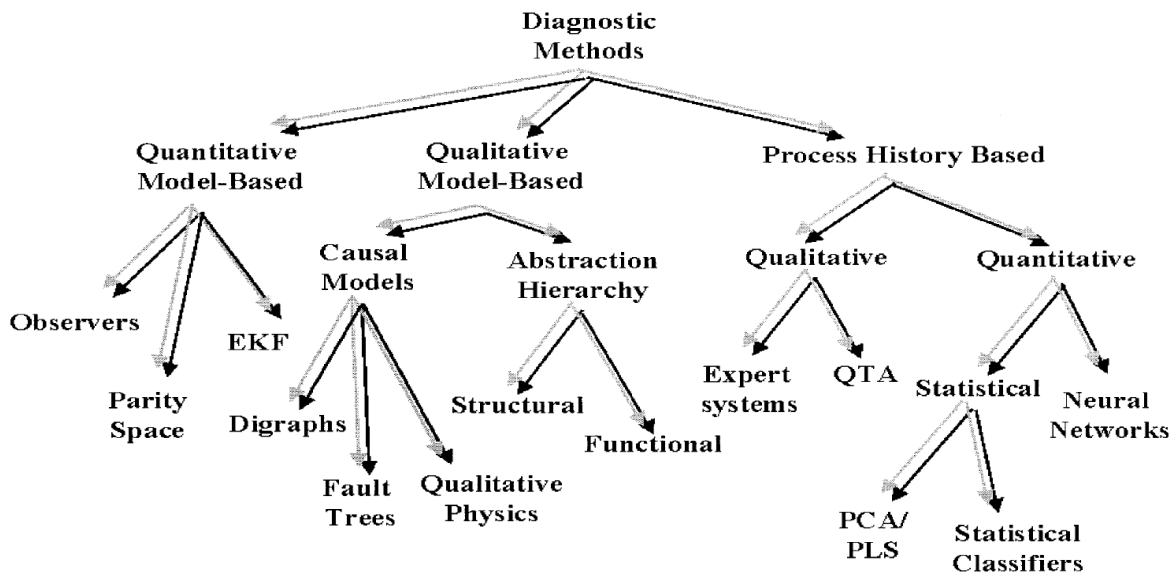


Figure 1.8 : Les Méthodes de Diagnostic

1.4 Conclusion

L'objectif du diagnostic est de détecter rapidement les divers défauts existant sur un procédé (défaut d'instrumentation, de paramètres, structurel) pour éviter la dégradation de ses performances et augmenter la sécurité des opérateurs et de l'environnement. Dans ce chapitre nous avons fourni une présentation succincte des différentes techniques de diagnostic classiquement utilisées. Une méthode de diagnostic quantitative a suscité notre attention elle est destinée à détecter les défauts par Analyse en Composante Principale (ACP).

Chapitre II :

Analyse en composantes principales ACP

2.1 Introduction

L'analyse en composantes principales est la plus utilisée des analyses multi variées. C'est aussi la plus ancienne. Ses fondements datent de 1901 quand Karl Pearson en a conçu les idées essentielles à deux dimensions. Ce chapitre est consacré à l'exploration des différentes approches de l'ACP et à leur applicabilité dans le domaine de la détection et du diagnostic des défauts.

2.2 Principe de l'ACP

Analyse en composantes principales ou fonctions orthogonales empiriques est une méthode statistique multi-variée pour l'analyse de données qui vise à ré-exprimer une donnée multidimensionnelle en un ensemble des composantes orthogonales (non corrélées) en projetant l'ensemble de données original à un nouvel axe (changement de base) visant à extraire les informations importantes de la base de données d'origine [7].

Ce raisonnement nous permet de définir l'objectif principal de l'ACP comme suit: Transformer un ensemble de variables dépendantes en un ensemble de variables non corrélées, appelées composantes principales, qui sont l'ordre, de sorte que les premiers composants contiennent la plupart des informations dans le jeu de données d'origine. Étant donné qu'un changement de base est appliqué, l'ACP suppose clairement une relation linéaire entre les données originales.

Les principaux objectifs de l'analyse des composantes principales sont les suivants [8]:

- Réduire la dimensionnalité des données, c'est-à-dire représenter les données avec moins de nombre de variables aléatoires.
- Suppression de dépendances de données, c'est-à-dire transformation des données en un ensemble de variables non corrélées.

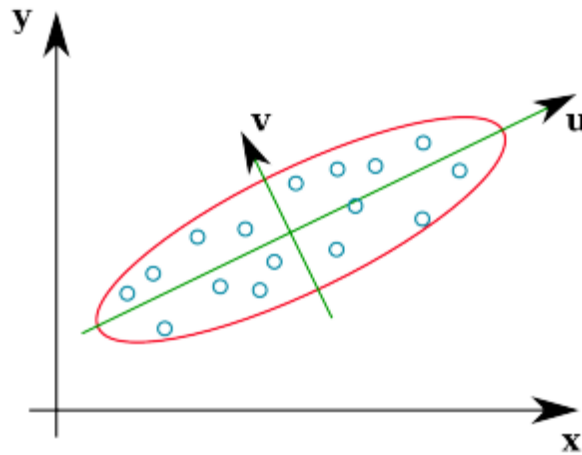


Figure 2.1 : Illustration de l'ACP dans les données 2-D.

La figure 2.1. représente une illustration pour l'ACP avec des données bidimensionnelles. x et y représentent les données dépendantes à l'origine, alors que $\{u, v\}$ représentent les composantes principales. Sur cette figure, le nouveau système de coordonnées peut mieux décrire les données.

Les coordonnées v sont très proches de zéro et peuvent être négligées sans perte d'information. De plus, on peut supposer que les coordonnées v ne sont pas nulles uniquement à cause du bruit aléatoire et des incertitudes affectant les mesures.

Supposons un ensemble de données $X \in \mathbb{R}^{n \times m}$ où m est le nombre de variables (capteurs) et n le nombre d'observations (instances).

Le problème de l'ACP peut être considéré comme le problème de trouver une transformation appropriée $P \in \mathbb{R}^{m \times m}$ lorsqu'il est appliqué à X donne un nouvel ensemble de données $T \in \mathbb{R}^{n \times m}$ c'est-à-dire :

$$T = X.P \quad (2.1)$$

La première chose que nous cherchons dans T est d'avoir un ensemble de variables non corrélées, qui peuvent être adressées via la corrélation ou les matrices de covariance du nouvel ensemble de données T .

La covariance de deux variables aléatoires $u, v \in \mathbb{R}^n$ est définie comme suit:

$$Cov(u, v) = \delta_{uv} = E((u - \mu_u)(v - \mu_v)) \quad (2.2)$$

Où E désigne l'espérance et μ désigne la moyenne. La covariance peut être calculée comme suit :

$$\delta_{uv} = \frac{\sum_{i=1}^n (u_i - \mu_u)(v_i - \mu_v)}{n} \quad (2.3)$$

La covariance telle que représentée dans (2.3) est un estimateur biaisé. Alors que l'estimateur non biaisé de la covariance [9.10]:

$$\delta_{uv} = \frac{\sum_{i=1}^n (u_i - \mu_u)(v_i - \mu_v)}{n-1} \quad (2.4)$$

La corrélation est alors définie comme suit:

$$\rho_{uv} = \frac{\delta_{uv}}{\delta_u \cdot \delta_v} \quad (2.5)$$

Pourvu que $\delta_u^2 = \delta_{uu}$ soit la variance de la variable aléatoire u . Pour une matrice de données :

$$X = [x_1; x_2; \dots; x_m] \quad (2.6)$$

x_i désigne un type de mesure (colonne) de X . La matrice covariance et la matrice de corrélation sont définies comme dans (2.7) et (2.8) respectivement :

$$\delta_X = \begin{bmatrix} \delta_{x_1}^2 & \delta_{x_1 x_2} & \dots & \delta_{x_1 x_m} \\ \delta_{x_2 x_1} & \delta_{x_2}^2 & \dots & \delta_{x_2 x_m} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{x_m x_1} & \delta_{x_m x_2} & \dots & \delta_{x_m}^2 \end{bmatrix} \quad (2.7)$$

$$\rho_X = \begin{bmatrix} 1 & \rho_{x_1 x_2} & \dots & \rho_{x_1 x_m} \\ \delta_{x_2 x_1} & 1 & \dots & \rho_{x_2 x_m} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{x_m x_1} & \rho_{x_m x_2} & \dots & 1 \end{bmatrix} \quad (2.8)$$

Retour au problème de l'ACP, forcer T à avoir des colonnes non corrélées est mathématiquement équivalent à diagonaliser la matrice de covariance (corrélations) de T . Avant de traiter ce problème, l'ensemble de données d'origine doit être normalisé, c'est-à-dire :

- Chaque type de mesure a une moyenne nulle; ceci simplifie le calcul de la matrice de covariance (corrélations) à:

$$\delta_x = \frac{X_0^T \cdot X_0}{n-1} \quad (2.9)$$

Avec :

$$X_0 = [x_1 - \mu_{x_1}; x_2 - \mu_{x_2}; \dots; x_m - \mu_{x_m}] \quad (2.10)$$

- En cas d'utilisation de la matrice de covariance, chaque type de mesure doit avoir une variance unitaire pour unifier les unités des différentes mesures. Nous définissons l'ensemble de données normalisé :

$$X_n = \left[\frac{x_1 - \mu_{x_1}}{\delta_{x_1}}, \frac{x_2 - \mu_{x_2}}{\delta_{x_2}}, \dots, \frac{x_m - \mu_{x_m}}{\delta_{x_m}} \right] \quad (2.11)$$

La transformation prévue est changée en:

$$T_n = X_n \cdot P \quad (2.12)$$

Pour que T_n soit normalisé, P doit être orthonormé. La matrice de covariance de T_n est alors:

$$\delta_{T_n} = \frac{T_n^T \cdot T_n}{n-1} = P^T \frac{X_n^T \cdot X_n}{n-1} P \quad (2.13)$$

Ainsi:

$$\delta_{T_n} = P^T \cdot Cov(X_n) \cdot P \quad (2.14)$$

En notant que la matrice de covariance (corrélation) est symétrique et en prenant l'hypothèse que P est orthonormale, on peut écrire:

$$Cov(X_n) = P \cdot \delta_{T_n} \cdot P^T \quad (2.15)$$

L'équation (2.15) transforme le problème de l'ACP en un problème plus simple consistant à trouver une factorisation appropriée pour la matrice δ_{X_n} qui prend la forme de l'équation (2.15) avec P une matrice orthonormale $m \times m$ et δ_{T_n} une matrice diagonale. Il est intéressant de noter que la forme de l'équation (2.15) peut être simplement trouvée en appliquant la décomposition en valeur singulière SVD. (Voir l'annexe)

T_n est appelée la matrice scores ou caractéristiques, tandis que P s'appelle la matrice de charges ou la matrice des CP.

Les colonnes de la matrice de chargement sont les composantes principales qui sont les vecteurs propres de la matrice δ_{X_n} , ce dernier peut être prouvé sur la base de l'équation de la structure d'Eigen de la matrice de covariance:

$$\delta_{X_n} \cdot p_i = \lambda_i \cdot p_i \quad (2.16)$$

L'assemblage de ces équations conduit à:

$$\delta_{X_n} \cdot [p_1 \ p_2 \ \dots \ p_m] = \begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_2 \end{bmatrix} \cdot [p_1 \ p_2 \ \dots \ p_m] \quad (2.17)$$

Lorsque les vecteurs propres peuvent être sélectionnés pour être orthogonaux, l'équation (2.17) peut être agencée comme :

$$\begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_2 \end{bmatrix} = [p_1 \ p_2 \ \dots \ p_m]^T \cdot \delta_{X_N} \cdot [p_1 \ p_2 \ \dots \ p_m] \quad (2.18)$$

Le vecteur propre associé à la plus grande valeur propre décrit la direction de la plus grande variance et ainsi de suite. De plus, on peut montrer que la valeur propre λ_i représente la variance dans la direction du vecteur principal correspondant p_i .

Les colonnes de P (les vecteurs principaux) sont ordonnées selon l'ordre décroissant des valeurs singulières de δ_{X_N} . Dans de nombreux cas, en raison de la redondance entre les variables, moins de composants sont suffisants pour représenter les données.

Ainsi, en utilisant $a < m$ des composants, on peut obtenir des scores a -dimensionnels par la relation suivante [9]:

$$\hat{T} = X_n \cdot P_a \quad (2.19)$$

Où P_a contient seulement les premières a colonnes de P . a est appelée la dimension du modèle ACP. En fait, le nombre de composants retenues est le facteur le plus critique qui affecte la sensibilité de tout modèle ACP.

En supposant que la projection orthogonale comprimera et séparera l'information et le bruit, la surestimation du nombre de composants retenus augmentera le caractère aléatoire du modèle, ce qui affectera la qualité de la détection et augmentera les fausses alarmes. De l'autre côté, sous-estimer le nombre de composants entraîne une perte d'information [11,12].

En été, la construction d'un modèle ACP signifie:

- Détermination la matrice des charges P .
- Détermination du nombre de composants à retenir.
- Calcul des caractéristiques (scores).

Alors qu'un modèle ACP ressemblerait (2.20), avec E est la matrice d'erreur.

$$X_n = T_n \cdot P_a^T + E \quad (2.20)$$

La matrice de chargement et les scores peuvent être calculés en utilisant SVD (Décomposition de valeur singulière).

2.3 Sélection de la démontions du modèle

La sélection du nombre de composants retenus joue un rôle majeur dans la qualité du modèle ACP. De nombreuses recherches ont été réalisées afin de déterminer le nombre optimal de composantes principales. Cependant, une méthode pour estimer le nombre de composants retenus qui est garanti pour culminer le meilleur nombre de CP encore hors de portée; même si un nombre considérable de méthodes sont déjà présentées dans la littérature. Certaines des méthodes largement utilisées sont présentées dans cette section.

2.3.1 Règle de Kaiser – Guttman

Le critère dit de Kaiser-Guttman a été trouvé au début de 1954 et est encore la méthode la plus utilisée pour l'estimation du nombre de composants dans l'ACP et l'analyse factorielle. Le critère de Kaiser, également connu sous le nom de k_1 , consiste à choisir les composantes correspondant à ces valeurs propres plus grandes qu'une (valeur propre moyenne).

La première enquête sur la méthode conduit au résultat que le critère k_1 fournie une bande inférieure pour le nombre de composants retenus alors que d'autres études montrent que le critère tend à surestimer le nombre de composants retenus.

De plus, la sélection des composantes avec des valeurs singulières supérieures à une rend le critère si discutable, car une composante de valeur singulière égale à 1.01 est considérée comme significative et informative et doit donc être conservée alors qu'une composante de valeur singulière égale à 0.99 n'est pas significative [13].

2.3.2 La méthode Scree plot

Le test scree est une méthode consistant à tracer les valeurs propres de la matrice de covariance / corrélation en ordre décroissant puis à explorer le graphe résultant pour déterminer le point où la dernière goutte a lieu et le graphe commence à lisse. Le raisonnement derrière cette approche peut être vu comme si le point du coude divisant les composantes importantes (majeures) des composantes insignifiantes (mineures).

Ce test est simple à appliquer, mais il peut être difficile à interpréter en raison du fait qu'une méthode graphique sans règle systématique peut s'avérer très subjective. En outre, le graphe lui-même peut être trompeur en raison de l'ambiguïté du coude due au sloop progressif du graphe ou à l'existence de plus d'un coude.

Habituellement, un graphe cumulatif de la proportion des valeurs propres est utilisé en parallèle avec le test de l'échelon pour augmenter la confiance dans la sélection faite par la méthode. Le test peut être réalisé en fonction du test logarithmique des valeurs propres (LVP),

ce test étend le test de scree en traçant les logarithmes des valeurs propres ($\log(\lambda_i)$) au lieu des valeurs propres. Cette approche peut accroître l'interprétation de l'intrigue. [8,14]

2.3.3 Pourcentage cumulé de la variance (PCV)

Il est bien connu que la variance est une bonne mesure pour l'importance d'une dimension principale donnée et pour la quantité d'information qui pourrait exister dans cette dimension par rapport aux autres. Ainsi, le maintien d'un certain nombre de composantes qui correspond à un certain pourcentage de la variance totale est raisonnable.

La quantité de variance préservée par a -dimensions sur un total de m types de mesure disponibles peut être calculée sur la base des valeurs propres de la matrice de covariance / corrélation comme suit:

$$PCV \% = \frac{\sum_{i=1}^a \lambda_i}{\sum_{i=1}^m \lambda_i} \times 100 \quad (2.22)$$

Même si le PCV est l'une des méthodes les plus utilisées, un problème majeur émerge du fait que le pourcentage de la variance cumulée doit être fixé à l'avance. Habituellement, un pourcentage compris entre 80 et 85% est utilisé dans un certain contexte tandis que certaines références encouragent l'utilisation de pourcentages supérieurs à 85%, mais la sélection optimale qui garantit la meilleure représentation des données est dépendante des données et toujours hors de portée.

La valeur optimale du PCV qui représente le mieux les données peut varier largement en fonction de la quantité de bruit présent dans la mesure et de la redondance entre les différentes variables. En raison de l'absence d'une règle claire, le PCV s'avère subjectif et le fait que le pourcentage cumulatif augmente avec l'augmentation du nombre de composantes rend la méthode ambiguë.

En outre, alors que nous voulons conserver autant que possible de la variance, nous voulons conserver autant de composants principaux que possible et de garder le bruit hors de l'interférence. Par conséquent, la décision consiste à négocier le montant de la variance conservée et le nombre de composantes retenues.

2.3.4 La méthode bâtons brisés (BB)

Le raisonnement derrière cette méthode est basé sur la considération de la variance partagée par les axes principaux à intégrer dans un bâton d'une longueur unitaire. Si nous avons un bâton de longueur unitaire, brisés aléatoirement à m segments, on peut montrer que la longueur attendue du $K^{\text{ème}}$ segment le plus long serait :

$$E(\mathcal{L}k) = \frac{1}{m} \sum_{i=k}^m \frac{1}{i} \quad (2.23)$$

Si l'on considère que le modèle ACP divise la variance aléatoirement sur toutes les m -dimensions, la fraction de la variance expliquée par chaque dimension serait la même que les longueurs relatives obtenues en décomposant les bâtons de façon aléatoire en m pièces [15].

Il serait alors inutile de conserver les composantes qui expliquent la variance inférieure ou égale à celle prédite par le modèle de bâtons brisés. BB peut être transporté de deux façons après avoir mis les valeurs propres et les prédictions BB dans l'ordre décroissant:

- Comparer les valeurs propres individuelles avec les prédictions individuelles des BB.
- Comparer les valeurs propres cumulées avec les prévisions cumulatives de BB.

Certaines études ont montré que le modèle de bâton brisé a tendance à sous-estimer le nombre de composants retenus, et un problème majeur avec ce modèle est son indépendance par rapport à la taille de l'échantillon [16]. Néanmoins, le modèle BB est facile à calculer et il a montré une bonne performance avec une précision acceptable, du moins lorsque les données étudiées sont fortement corrélées.

2.3.5 Analyse parallèle (AP)

L'analyse en parallèle est une adaptation basée sur l'échantillon pour la règle $k1$ basée sur la population, ce qui implique un processus de simulation Monte Carlo. La méthodologie est basée sur la comparaison des valeurs propres de la matrice de covariance / corrélation de l'ensemble de données d'origine avec celles d'un ensemble de données échantillonnées au hasard à partir d'une population normalement répartie.

Un composant est considéré comme et ont significatif une fois que la valeur propre associée est plus grande que celle générée à partir de l'échantillon aléatoire. Basé sur le fait qu'une méthode basée sur l'échantillonnage aléatoire peut présenter une grande variabilité, il est habituellement nécessaire de répéter la génération des pseudo-valeurs propres pour k -temps.

En ayant un ensemble de pseudo-valeurs propres ordonnées pour chaque expérience, on pourrait augmenter l'estimation en prenant soit la valeur propre moyenne pour chaque composante, soit en évaluant la distribution empirique des valeurs propres associées à une composante donnée pour déterminer une limite supérieure de confiance pour chaque pseudo-valeur propre avec un centile présélectionné.

L'algorithme d'AP est résumé dans les lignes suivantes :

- ✓ Calculez les valeurs propres de la matrice de covariance: $\lambda = \{\lambda_1, \lambda_2, \dots, \lambda_m\}$.
- ✓ Générer k -ensembles de données aléatoires de même taille que X ($Y_i \in \mathbb{R}^{n \times m}, i = 1, 2, \dots, k$).
- ✓ Calculez les valeurs propres Y_i et stockez-les dans $\Xi \in \mathbb{R}^{k \times m}$ (en rangée). Avec ξ_{ij} est la $j^{\text{ème}}$ pseudo-valeur propre pour l'expérience $i^{\text{ème}}$. [24]

$$\Xi = \begin{bmatrix} \xi_{11} & \xi_{12} & \dots & \xi_{1m} \\ \xi_{21} & \xi_{22} & \dots & \xi_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \xi_{k1} & \xi_{k2} & \dots & \xi_{km} \end{bmatrix} \quad (2.25)$$

Effectuez soit:

- Calculer la valeur propre moyenne: $\hat{\Xi} = [\bar{\xi}_1 \ \bar{\xi}_2 \ \dots \ \bar{\xi}_m]$.
- Calculer le α -percentile pour chaque colonne: $\hat{\Xi} = [\xi_{1\alpha} \ \xi_{2\alpha} \ \dots \ \xi_{m\alpha}]$. Avec α est habituellement choisi entre 0,95 et 0,99 sert à fixer la tendance de la méthode à surestimer le nombre de composants.
- ✓ Conservez les composants pour lesquels: $\Xi_i > \lambda_i$. [24]

2.3.6 Validation croisée (VC)

La validation croisée est une méthode statistique utilisée pour comparer les algorithmes d'apprentissage. VC travaille sur la base du principe de la division de l'ensemble disponible de données en deux parties, des ensembles de formation et d'essais. L'ensemble d'apprentissage est utilisé pour estimer un modèle tandis que l'ensemble de tests est utilisé pour valider ce modèle. Un critère de qualité de l'ajustement (CQA) doit être sélectionné afin d'évaluer le modèle puisque l'étape de validation d'un modèle ACP consistera à sélectionner le nombre de composants à conserver hors de m les valeurs possibles. Fondamentalement, un modèle ACP doit être construit pour $i = 1, 2, \dots, m$ et celui qui satisfait le mieux à la sélection de CQA est sélectionné [14].

2.3.7 Méthode de bootstrap

Bootstrap peut être utilisée d'une manière similaire à VC, ce qui, en prenant une formation et des ensembles de tests. Néanmoins, le Bootstrap est une méthode utilisée pour évaluer la performance du modèle. Fondamentalement, cette méthode est basée sur le tirage aléatoire n des échantillons de l'ensemble de données avec remplacement afin de former l'échantillon bootstrap. L'échantillon bootstrap est ensuite utilisé pour former le modèle, qui doit être testé sur les instances non sélectionnées par le processus de dessin. En raison de l'échantillonnage aléatoire, la méthode bootstrap présente une grande variabilité.

2.4 Les inconvénients principaux de l'ACP

Lorsqu'on arrive à appliquer l'ACP sur un ensemble de données pré-acquises pour générer un modèle, la détermination du nombre exact de composants à conserver pourrait être le problème le plus important.

Comme on l'a vu plus haut, la surestimation du nombre de composantes signifie que nous avons contaminé les informations extraites par l'ajout d'une dimension bruyante avec approximativement aucune information utile, ceci cas une quantité excessive de fausses alarmes. De l'autre côté, sous-estimer le nombre de composants signifie un manque d'informations car une ou plusieurs dimensions importantes sont négligées.

Si un défaut qui affecte en grande partie les dimensions négligées se produit, il sera totalement contenu dans ces dimensions et donc indétectable, ou il ne peut être détecté à moins qu'il ne grandisse suffisamment pour affecter les autres dimensions. De toute évidence, cela impose, dans le meilleur des cas, un retard dans la détection de défaut, ce qui peut être désastreux pour le système.

L'analyse des composantes principales suppose des relations linéaires entre les variables, cette hypothèse peut être trompeuse. Les différentes variables mesurées peuvent présenter n'importe quel type de relations non linéaires (quadratique, cubique, etc.), l'existence d'une telle relation pose le problème d'extraire des dimensions inutiles et non informatives. De plus, le fait que l'ACP soit une méthode non paramétrique sert de force et de faiblesse simultanément. L'absence de paramètres rend le processus d'ajustement des résultats indésirables en modifiant certains paramètres inutiles.

Le troisième problème qui existe avec l'ACP se pose lors de la surveillance du système. Il a été recommandé que les données soient normalisées avant d'appliquer le modèle ACP. Cependant, lorsqu'une nouvelle instance de données est reçue, l'utilisation des moyens et des écarts de l'ensemble de tests peut être inappropriée.

Par exemple, des parties choisies au hasard des données ont habituellement des moyennes et des variances différentes, c'est-à-dire que les moyennes et les variances varient avec le temps. Les données avec des moyennes et des variances fixes, quel que soit l'intervalle de temps choisi, sont appelées données stationnaires. La stationnarité d'un ensemble de données est tout à fait impossible à atteindre dans les processus réels.

En outre, l'ACP telle que présentée précédemment ne prend pas en considération l'existence de relations entre observations à différents moments ou l'autocorrélation dans les séries temporelles. L'autocorrélation augmente souvent lorsque les mesures dans une série temporelle (type de mesure) ne sont pas indépendantes, cela se produit souvent en raison du comportement dynamique du processus et des capteurs.

2.5 ACP réursive (ACPR) et fenêtre mobile ACP (FMACP)

Les cartes des contrôles ASCP et ADCP sont toutes deux incapables de faire face à la non-stationnarité présente dans la plupart des données industrielles. Si l'un de ces modèles est appliqué à des données provenant d'un processus non stationnaire, des problèmes peuvent survenir lorsque la structure moyenne et / ou de covariance du modèle devient mal spécifiée car elles sont estimées à l'aide d'observations d'une période de temps avec peu de similitude avec celle surveillée. L'idée derrière la fenêtre réursive et la fenêtre mobile se rapproche de limiter l'effet de la non stationnarité en limitant l'effet des données anciennes.

Pour les deux approches, lorsqu'un nouvel instant de données est mesuré, il est évalué selon le modèle existant. Si l'indicateur de défaut pour ces nouvelles données dépasse les limites de contrôle, c'est-à-dire un défaut, le modèle reste inchangé. Cependant, lorsque les nouvelles données sont sous contrôle, les nouvelles données sont ajoutées à l'ensemble de données et un nouveau modèle est évalué.

Lorsque les données entrantes sont jugées saines et ajoutées à l'ensemble de données de construction de modèle (EDCM), les deux approches sont différentes:

- Dans l'ACPR (Analyse des composantes principales réursives), toutes les données historiques sont stockées dans l'ensemble de données de construction du modèle. Sauf pour les nouvelles données, l'EDCM est pondéré en utilisant un facteur d'oubli $0 < \eta < 1$ et le modèle est mis à jour.

Néanmoins, le fait de porter la méthode de cette manière a pour résultat un temps de calcul élevé, une augmentation du besoin de capacité de stockage afin de conserver toutes les données anciennes.

En pratique, la mise à jour n'est pas effectuée à l'aide de la matrice de données complète, mais plutôt une pondération est effectuée pour mettre à jour uniquement la moyenne et la matrice de covariance. Une méthode proposée pour traiter le temps de calcul est de réduire le nombre de mises à jour pour le modèle.

- Contrairement à ACPR, FMAP met à jour le modèle ACP avec chaque instant de données entrantes (en bonne santé) tout en limitant les observations utilisées dans les estimations à celles qui tombent dans une fenêtre de temps spécifiée. À chaque nouvelle observation, l'EDCM a la même taille.

Ceci est réalisé en excluant la plus ancienne observation dans l'EDCM et en incluant la nouvelle observation. A deux instants de temps successifs, l'EDCM aura la forme dans les équations (2.31) et (2.32) respectivement :

$$X_t = [x_{t-w+1}, x_{t-w+2}, \dots, x_t]^T \quad (2.31)$$

$$X_{t+1} = [x_{t-w+2}, x_{t-w+3}, \dots, x_{t+1}]^T \quad (2.32)$$

Avec w la longueur de la fenêtre.

Le principal défi dans FMAP est le choix de la taille de la fenêtre. Une estimation approximative basée sur la convergence de la distribution χ^2 à la distribution F pour la surveillance basée sur T^2 Conduire au résultat: la taille de la fenêtre nécessaire pour estimer correctement la T^2 -statistique doit être au moins dix fois le nombre de variables.

De plus, pour réduire l'effet de la non stationnarité sur l'ensemble du processus de surveillance, l'ensemble de données d'étalonnage utilisé pour générer le modèle au premier instant du processus de surveillance doit être soigneusement sélectionné, c'est-à-dire que l'ensemble de données d'étalonnage doit être stationnaire [25].

2.6 ACP robuste

L'ACP robuste est proposée comme une modification pour la méthode d'ACP classique, l'ACP conventionnelle souffre d'observations grossièrement corrompues. Un certain nombre d'approches différentes existent pour l'ACP robuste, y compris une version idéalisée de l'ACP robuste, qui vise à récupérer une matrice de rang inférieur L_0 à partir de mesures hautement corrompues.

$$M = L_0 + S_0 \quad (2.33)$$

Cette décomposition se traduit par des matrices de faible rang et de faible densité. Contrairement au petit terme de bruit E dans l'ACP conventionnelle, les entrées dans S_0 peuvent avoir une grandeur arbitrairement grande, et leur support est supposé être clairsemé mais inconnu [21].

2.7 L'ACP à Noyau

Comme il a été mentionné précédemment, l'ACP est une technique linéaire qui est utilisée pour comprimer, filtrer et extraire des informations d'un ensemble de données de grande dimension en fonction des relations linéaires existantes entre les différents types de mesures. Cependant, dans la pratique, les données industrielles ont une quantité excessive de relations non linéaires entre les différentes variables.

Puisque l'ACP conventionnelle ne peut être appliquée efficacement à moins que la linéarité soit supposée, les observations peuvent toujours être mappées à un espace de dimension supérieure, appelé espace des fonctionnalités (\mathcal{F}), où les nouvelles observations varient linéairement.

ACPN utilise des machines de support vectoriel (MSV) pour trouver une correspondance non linéaire appropriée de l'espace d'entrée à l'espace des fonctionnalités grâce à une fonction de noyau simple.

Ensuite, on applique une ACP linéaire sur les données afin de la déplacer de l'espace caractéristiques à l'espace ACPN.

Sur le plan conceptuel, l'ACPN est réalisé selon les étapes suivantes:

- Une fonction de noyau $\Phi(X)$ est utilisée pour mapper les données de l'espace d'entrée à l'espace des entités.
- ACP linéaire est effectuée dans l'espace de fonctionnalité donnant une représentation basée sur l'espace ACPN de dimension inférieure.
- Pour reconstruire les données originales, la représentation spatiale du noyau doit être projetée dans l'espace des entités, puis la représentation de l'espace des entités est projetée dans l'espace d'entrée.

Néanmoins, la procédure décrite ci-dessus prend du temps. Par conséquent, aucune des étapes précédentes ne doit être effectuée. La procédure de calcul est réalisée directement à partir de l'espace d'entrée vers l'espace ACPN en utilisant un ensemble de fonctions de noyau $\sum_i a_i k(x, x_i)$. [19]

2.8 Détection des défauts basée sur l'ACP

Afin de suivre un processus multi-varié, des tableaux de contrôle ont été développés sur la base de modèles ACP. Ayant un modèle ACP basé sur des données historiques recueillies dans l'état sain du processus (seule la variation de la cause du hasard existe), le comportement futur peut être référencé par rapport à ce modèle en contrôle. De nouvelles observations multi-variées peuvent être projetées sur le plan défini par le chargement ACP.

Quelle que soit l'approche ACP utilisée pour générer les scores et les chargements, une surveillance en ligne peut être fournie à l'aide de la statistique T^2 d'Hotelling et de la statistique Q , cette dernière est habituellement appelée aussi SPE (Squared Prediction Error). A chaque instance d'échantillonnage, un nouveau point de données est acquis et sauvegardé dans le vecteur ligne x_i et ces données entrantes sont utilisées pour générer un indicateur de défaut basé sur le modèle ACP.

Pour le seuillage constant, un ensemble d'apprentissage est utilisé pour former le modèle, cette formation consiste à générer un nombre relativement important de points de contrôle et à évaluer leur distribution empirique afin de déterminer des limites de contrôle constantes.

Les statistiques T^2 et Q sont définies comme suit:

$$T_t^2 = (x_t - \mu) \cdot P_a \cdot \Sigma_a^{-1} \cdot P_a^T \cdot (x_t - \mu)^T \quad (2.34)$$

$$Q_i = (x_t - \mu) \cdot (I_m - P_a \cdot P_a^T) \cdot (x_t - \mu)^T = \|x_t - \mu\|^2 \quad (2.35)$$

Lorsque Σ_a est la matrice diagonale des plus grandes a valeurs singulières de la matrice de covariance / corrélation, I_m est la matrice d'identité $m \times m$, et P_a est la matrice des charges principales. La statistique Q est la distance orthogonale quadratique à l'espace des Composantes Principales. Les deux statistiques sont basées sur les symptômes générés qui sont toujours positifs et habituellement proches de zéro.

Par conséquent, l'utilisation de ces statistiques conduit à une carte de contrôle où seule une limite supérieure de contrôle est nécessaire. Dans le cadre d'une surveillance de processus, l'utilisation de la statistique T^2 de Hotelling basée uniquement sur les premiers composants n'est pas suffisante.

Etant donné que le nombre de composants peut seulement être estimé, cette méthode ne détectera que si la variation des propriétés mesurées par les premiers CP dans le processus est supérieure à ce qui peut être expliqué par la cause fortuite prédéfinie uniquement.

Si un nouveau type de spécification qui n'était pas présent dans les données utilisées pour développer le modèle ACP apparaît, alors de nouveaux CP apparaîtront et le processus de suivi pour les nouvelles observations sera trompeur [20].

Ce problème est résolu par la Q -statistique où l'erreur de prédiction est fortement affectée par les variations dans les derniers composants $m-a$.

Si nous supposons qu'il n'y a pas d'autocorrélation et de normalité multi-variée pour les scores, les limites de contrôle supérieures peuvent être calculées comme suit:

- Pour la surveillance T^2 :

$$UCL_{T^2}(\alpha) = \frac{a(n^2-1)}{n(n-1)} \cdot F_\alpha(a, n-a) \quad (2.36)$$

Avec $F_\alpha(a, n-a)$ est le $1-\alpha$ percentile de la F -distribution avec a et $n-a$ degrés de liberté et n est le nombre d'observations dans l'ensemble d'entraînement. Si le nombre d'observations est suffisamment grand, l'UCL peut être trouvée à partir de la distribution chi-carré avec a degrés de liberté en utilisant la formule suivante :

$$UCL_{T^2} = x_\alpha(a) \quad (2.37)$$

- En utilisant la statistique Q , basée sur la somme des erreurs de prédiction, on utilise la formule suivante :

$$UCL_Q(\alpha) = \theta_1 \cdot \left(1 + Z_\alpha \cdot \frac{\sqrt{2\theta_2 h_0^2}}{\theta_1} + \frac{\theta_2}{\theta_1^2} \cdot h_0(1-h_0) \right)^{1/h_0} \quad (2.38)$$

Avec :

$$h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2} \quad (2.39)$$

$$\theta_i = \sum_{j=a+1}^m \lambda_j^i \quad (2.40)$$

Et λ_j est la $j^{\text{ème}}$ plus grande valeur propre de la matrice de covariance / corrélation. Et Z_α représente le percentile $1-\alpha$ de la distribution normale standard. [23]

Il existe d'autres méthodes pour calculer le seuil de la Q -statistique. Néanmoins, la méthode présentée par l'équation (2.38) est plus rapide à calculer et donc plus appropriée pour la surveillance en temps réel [22].

Habituellement, si la normalité des scores n'est pas garantie, il est préférable de définir l'UCL de l'une ou l'autre statistique en utilisant la fonction de densité de probabilité empirique. Dans la littérature, des modèles hybrides sont également utilisés.

La sortie de l'analyse des composantes principales, qui est un ensemble de données inférieures et non corrélées, pourrait être utilisée comme entrée pour un deuxième étage de réseaux neuronaux ou logique floue.

Cette combinaison a été trouvée pour présenter une bonne précision de prédiction et la dimensionnalité réduite de l'espace d'entrée entraîne une diminution de la complexité du réseau neuronal et du temps d'apprentissage.

2.9 Conclusion

Dans ce chapitre, l'ACP est une approche statistique multi-variée pour la détection de défauts dans des processus complexes. La base mathématique des modèles ACP a été établie ainsi que la sélection des paramètres du modèle, tels que le nombre de CP.

De plus, différentes approches ACP proposées pour faire face aux problèmes d'autocorrélation, de non-stationnement et de linéarité ont été présentées. Les fondements de la détection des défauts basés sur ACP ont été présentés à l'aide de Q statistique et T^2 Hotelling.

Chapitre III :

*Descriptions des systèmes et application de
l'ACP*

3.1 Introduction

Dans ce chapitre on tient compte des descriptions des systèmes utilisés pour l'application de notre méthode « ACP » telle que cette méthode est définie dans le deuxième chapitre.

Pour l'application de l'ACP on a choisi un groupe électrogène SDMO GT1900 une chaudière à tube d'eau et un four électrique qui appartient à une cimenterie. Dans les matrices des variables X_1, X_2, X_3 . on a pris des échantillons pour les trois systèmes dans les états de fonctionnement normaux et anormaux.

3.2 Présentation des systèmes utilisés

3.2.1 Groupe électrogène SDMO GT1900

3.2.1.1 Définition

Un groupe électrogène est un dispositif autonome capable de produire de l'électricité. La plupart des groupes sont constitués d'un moteur thermique qui actionne un alternateur. Leur taille et leur poids peuvent varier de quelques kilogrammes à plusieurs dizaines de tonnes. La puissance d'un groupe électrogène s'exprime en VA (voltampère), kVA (kilovolt ampère) ou MVA (méga voltampère) selon la puissance. Les unités les plus puissantes sont mues par des turbines à gaz ou de gros moteurs Diesel.

Notre groupe SDMO T1900 est composé de trois parties, un moteur type MITSUBISHI S16R-PTA, un alternateur LSA 51.2S55 et un transformateur élévateur de tension 0.5KV-30KV.

La puissance unitaire du groupe électrogène assignée sur site est 1 888 KVA, de fréquence 50 Hz, facteur de puissance 0.8, et de tension 30 KV.



Figure 3.1: Groupe électrogène SDMO GT1900

3.2.1.2 Utilisation

Les groupes électrogènes sont utilisés soit dans les zones où le réseau de distribution électrique ne dessert pas, soit pour pallier une éventuelle coupure d'alimentation électrique. Dans le deuxième cas, ils sont alors souvent utilisés en complément d'une alimentation sans interruption constituée d'une batterie d'accumulateurs qui alimente un onduleur.

Ces dispositifs sont généralement utilisés dans des situations où l'interruption de l'alimentation électrique entraîne des conséquences graves ou des pertes financières, par exemple dans les hôpitaux, l'industrie y compris l'industrie agro-alimentaire, les aéroports, les centres informatiques, les pompiers pour les interventions, etc.

3.2.1.3 Fonctionnement

Ils fonctionnent à partir de tous les carburants. Les plus fréquents sont l'essence, le gazole, le gaz naturel, le GPL, les biocarburants et pour les plus puissants le fioul lourd. Le groupe peut être mis en fonctionnement de différentes manières : manuellement, électriquement ou grâce à l'air comprimé, selon la puissance.

3.2.1.4 Rendement

Le rendement des groupes électrogènes croît avec leur puissance, mais reste limité au maximum que permet le cycle de Carnot, duquel doivent être soustraites les pertes mécaniques et électriques dans l'alternateur et la transmission. En particulier, les groupes de puissance modérée se caractérisent par un médiocre rendement et une consommation élevée.

3.2.1.5 Equipements d'automatisme

Le groupe est commandé par un automate programmable APM802 de 18 entrées et 18 sorties. L'APM802 permet une conduite conviviale, ergonomique et autonome de groupe électrogène et de centrale d'énergie. (Voir l'annexe)

3.2.2 Four électrique

Le cœur vital d'une usine de ciment est le four rotatif qui sert à produire le clinker.

3.2.2.1 Définition

Dispositif d'incinération utilisé pour les déchets solides, pâteux et liquides qui sont introduits dans un cylindre en rotation, avec un flux d'air circulant, soit dans le même sens que les déchets, soit à contre courant. Une chambre de postcombustion permet de parfaire la combustion et de traiter les imbrûlés.

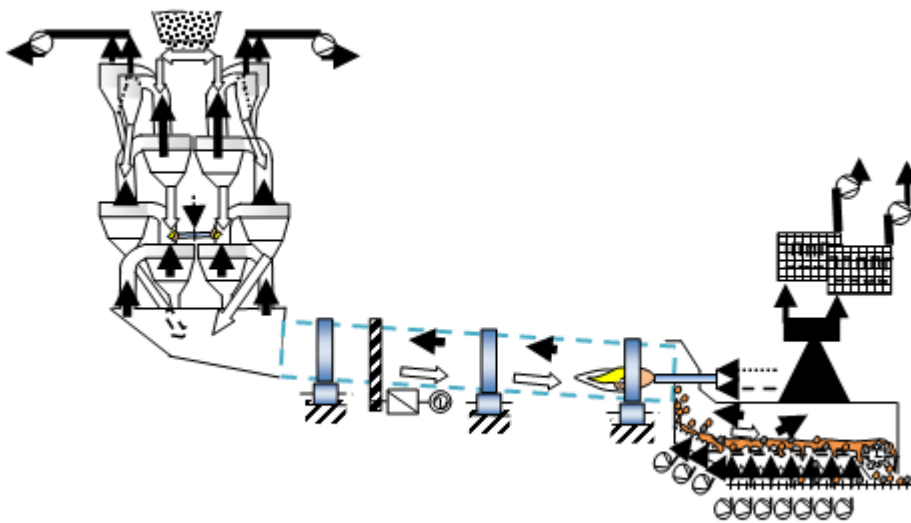


Figure 3.2: Four rotatif

3.2.2.2 Fonctionnement

Il s'agit d'un matériau dense lourd noir pierreux qui est ensuite mélangé à certains additifs pour fabriquer le ciment. Le clinker est obtenu grâce à un four rotatif de 80 mètres et à 5 mètres de diamètre. Deux moteurs à induction de 250 kW font tourner le four à une vitesse variable de 2 r.p.m par un entraînement électrique.

Le flux de matière alimente le four après avoir été préchauffé à 900 ° C dans la tour de préchauffage. Cela constitue quatre étages de cyclones qui sont montés dans deux positions verticales parallèles. Chacun d'entre eux est alimenté séparément par un chargeur de matériel de craie sèche qui a une capacité de 150 tonnes / heure.

La matière première elle-même est broyée en utilisant un mélange à trois éléments, le calcaire, la marne et le fer avec des proportions adéquates. Un pyroclone de brûleur de tour,

qui utilise du gaz naturel comme combustible de combustion, est monté à la fin du four rotatif.

Le matériau de sortie du four est alimenté dans un four après-poste appelé le refroidisseur. Il se compose de nombreux fans qui soufflent le flux de matière se déplaçant sur une grille mobile afin de refroidir moins à de 100 ° C et obtenir le clinker prêt final pour une prochaine utilisation.

3.2.3 Chaudière

3.2.3.1 Définition

La chaudière est un appareil destiné à extraire l'énergie calorifique du combustible et, alors elle a pour rôle de transformer l'eau d'alimentation en vapeur, à haute pression et à une température plus élevée en vue d'alimenter la turbine. La chaudière est de type pression (sous-pression) à circulation naturelle.



Figure 3.3 : Chaudière à tubes d'eau

3.2.3.2 Chaudière à tubes d'eau

Une chaudière à tubes d'eau ou chaudière multitubulaire est une chaudière dans laquelle l'eau circule à l'intérieur de tubes chauffés par des gaz.

3.2.3.3 Fonctionnement de chaudière

La chaudière proposée est un générateur à tubes d'eau, à un réservoir et à circulation naturelle, réalisée suivant la technique ERK. Sa disposition est du type compact en ligne à harpes, chambre de rayonnement horizontale équipée d'un brûleur à gaz naturel installé en façade.

Les tubes de retour extérieurs aux écrans du foyer non chauffés de gros diamètre sont disposés aux quatre coins de la chaudière. Ils en constituent l'ossature et assurent le support âgé de la chaudière et du réservoir, la rendant ainsi autoportante.

Les quatre tubes d'angle verticaux sont reliés entre eux par des tubes horizontaux qui jouent le rôle de distributeurs ou de collecteurs assurant l'alimentation et la circulation dans les surfaces évaporisatrices.

Les collecteurs supérieurs sont doublés de tubes de dégagement vapeur reliés directement au réservoir vapeur. Ce dispositif assure une première séparation de l'émulsion eau/vapeur à l'extérieur du réservoir. Cette séparation primaire est complétée par une séparation secondaire assurée dans le réservoir vapeur par un système de chicanes.

La structure tubulaire soudée ainsi constituée est en équilibre thermique constant, ce qui supprime les contraintes provoquées par les dilatations différentielles. De plus, les sections largement calculées des tubes de retour assurent une excellente circulation naturelle. Le taux de circulation de ce système est 3 à 4 fois plus élevé que dans une chaudière classique (sans tube d'angle), et pour ce cas particulier il est de l'ordre de 25.

Les surfaces évaporisatrices du générateur sont constituées :

- d'une part, des tubes écrans disposés entre les distributeurs et collecteurs et formant des parois étanches grâce à des ailettes rapportées et soudées des 2 côtés sur toute la longueur des tubes. Les étanchéités entre les différents écrans (avant, latéraux, arrière) sont réalisées par des peignes en acier, soudés sur les panneaux de tubes.
- d'autre part, de faisceaux de vaporisation par convection disposés verticalement dans un carneau horizontal dans le prolongement de la chambre foyer.

Ces faisceaux de convection sont réalisés en tubes verticaux lisses (faisceau "coup de feu") et à ailettes circulaires crantées enroulées et soudées en continu électriquement sur les tubes (évaporateurs), reliés aux tubes d'écran par soudure et sont disposés perpendiculairement au flux des gaz qui circulent horizontalement. Les surchauffeurs sont du type pendentifs, réalisés en tubes à ailettes pleines pour le SHT, tubes à ailettes dito ci-dessus pour le SBT, avec collecteurs horizontaux.

La chaudière est prévue pour être exploitée en mode Surveillance Permanente selon NFE 32020. Ce mode d'exploitation nécessite la présence d'un opérateur près de la chaudière au moment du démarrage et en cas d'arrêt sur incident ; il n'y a pas de séquences programmées pour les démarrages et les arrêts. Lorsque le générateur est en service, la régulation et les automatismes sont assurés à partir d'une armoire BMS, pour les fonctions d'automatismes et de régulation.

- La régulation de niveau chaudière proposée est du type à 3 éléments.
- La régulation de charge est du type positionneuse numérique avec correction d'O₂ sur les vannes à gaz naturel et le registre air comburant.
- La régulation de désurchauffe est du type à 2 influences. Elle s'effectue par injection et est pilotée par la mesure de la température vapeur sortie surchauffeur haute température avec une anticipation par la mesure de température vapeur après désurchauffe.
- Un dispositif de régulation de mise à l'air libre à 1 élément permet :
 - ✓ d'évacuer le débit de vapeur lors des phases de démarrage tant que les caractéristiques ne sont pas correctes
 - ✓ d'évacuer l'excédent de vapeur lorsque les besoins de vapeur de l'usine sont inférieurs au minimum technique de la chaudière.
- Les opérations de mise en conditionnement de la chaudière sont manuelles et nécessitent l'intervention de l'opérateur sur la vanne manuelle de départ vapeur.
- Les pompes alimentaires (exclues de notre FOURNITURE) sont supposées dédiées à la chaudière uniquement.

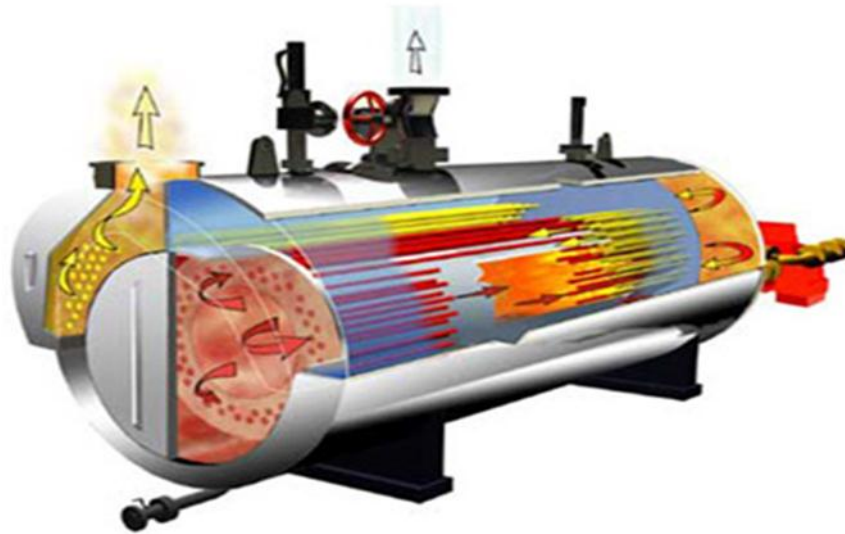


Figure 3.4 : Fonctionnement de la Chaudière à tubes d'eau

3.2.3.4 Brûleur

L'équipement de chauffe est composé d'1 brûleur avec rampe en acier pour la combustion exclusive de Gaz Naturel.

3.3 Application de l'ACP sur les trios systèmes

Afin de tester la fiabilité de la méthode ACP on l'a appliqué sur trois systèmes réels. Les capteurs de ces systèmes donnent les valeurs des paramètres de la matrice des données X . Pour cela on doit passer par les étapes suivantes :

- ❖ Normaliser la matrice des données X .
- ❖ Appliquer de l'ACP : détermination de la matrice des charges P par SVD.
- ❖ Choisir le nombre de composantes principales par PCV%.
- ❖ Détecter les défauts avec Q statistique et T^2 d'Hotelling.
- ❖ Tracer UCL (limites de confiance).
- ❖ Diminuer les fausses alarmes.

Pour le

➤ Groupe électrogène SDMO GT1900

On a appliqué notre méthode sur des mesures définies les paramètres du Groupe électrogène SDMO GT1900 dans les matrices des données X_1 . On a pris un échantillon de 2000 observations dans les états de fonctionnement normaux et anormaux.

Pour la

➤ **Chaudière**

Les paramètres du Chaudière dans la matrice des données X_2 . On a pris un échantillon de 3284 observations dans les états de fonctionnement normaux et anormaux.

Pour le

➤ **Four électrique rotatif**

Les paramètres d'un four électrique rotatif dans les matrices des données X_3 . On a pris un échantillon de 65526 observations dans les états de fonctionnement normaux et anormaux.

3.4 Détermination du nombre de composants principale (retenus) CP

Il existe plusieurs méthodes applicables sur l'ensemble de données afin de choisir la meilleure et le nombre de composants les plus descriptifs pour séparer avec succès les parties informatives du signal du bruit. Nous avons choisit la plus utilisé et la plus rapide c'est la méthode PCV (Pourcentage Cumulé de la Variance).

Tableau 3.1 : Nombre de composants retenus (CP) de chaque système.

Les systèmes utilisés	Nombre de composants retenus (CP)		
	Ci=75%	Ci=85%	Ci=95%
Groupe électrogène SDMO GT1900	A=2	A=4	A=5
Chaudière	A=10	A=13	A=16
Four électrique	A=9	A=11	A=14

Tableau 3.1. Montre le pourcentage d'information **Ci** et les composants retenus **A** de chaque système.

3.5 Conclusion

Dans ce chapitre, une application de l'ACP pour la détection des défauts dans un ensemble de données industrielles multi-variées a été réalisée .L'objectif de ce chapitre décrit les systèmes utilisés : un groupe électrogène SDMO GT1900, une chaudière et un four électrique.

Chapitre IV :

Résultats et discussions

4.1 Introduction

On va voir dans ce chapitre les résultats de l'application de l'ACP pour la détection des défauts et on a illustré les différents graphes obtenus lors de l'exécution des programmes en Matlab pour les différents algorithmes de cette méthode.

On a appliqué notre méthode sur des mesures définies les paramètres de trois systèmes présentés dans le chapitre 3 et on va essayer de diminuer quelques fausses alarmes dans l'état de fonctionnement normal.

4.2 Test de L'ACP sur les trois systèmes a l'état de fonctionnement normal (sans défaut)

➤ Test 1 : sur le four électrique

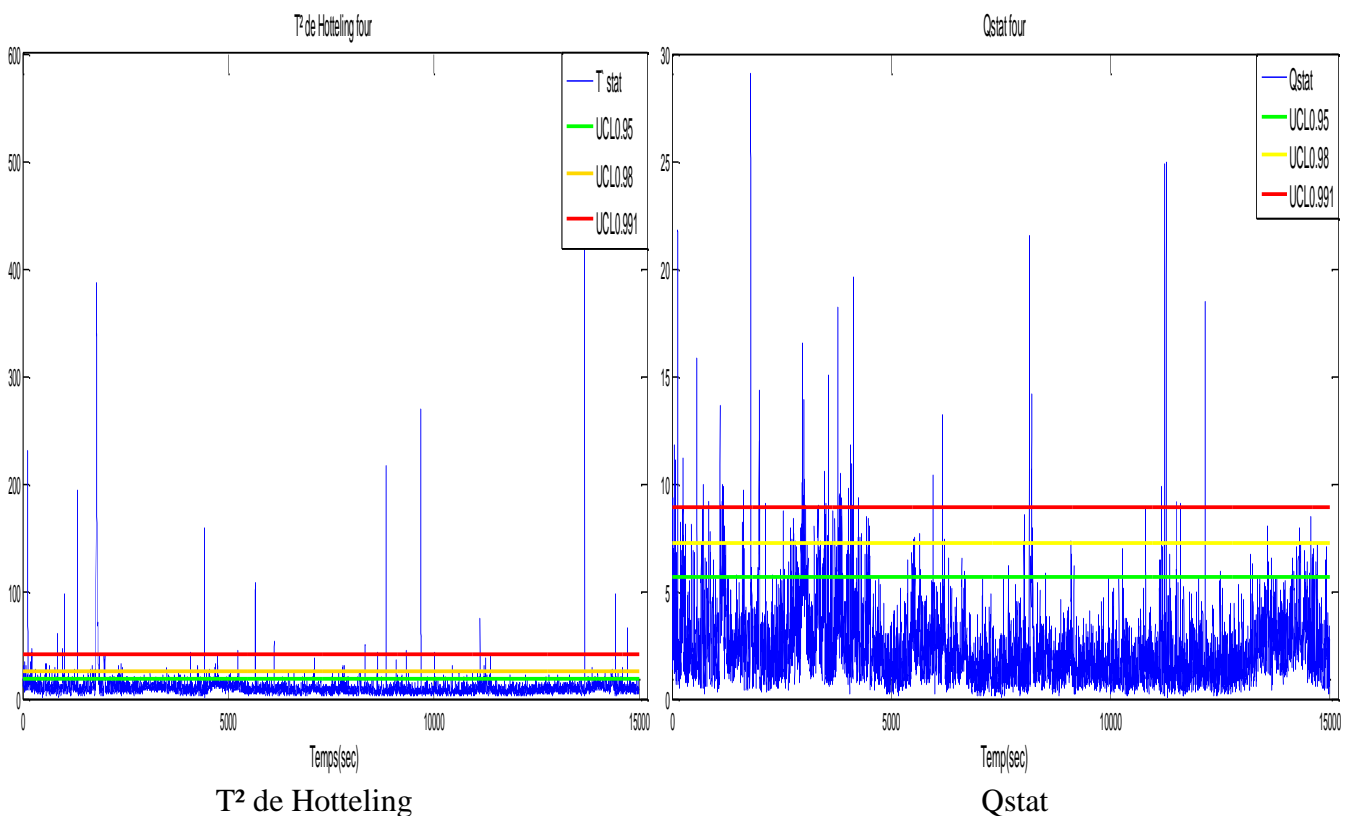
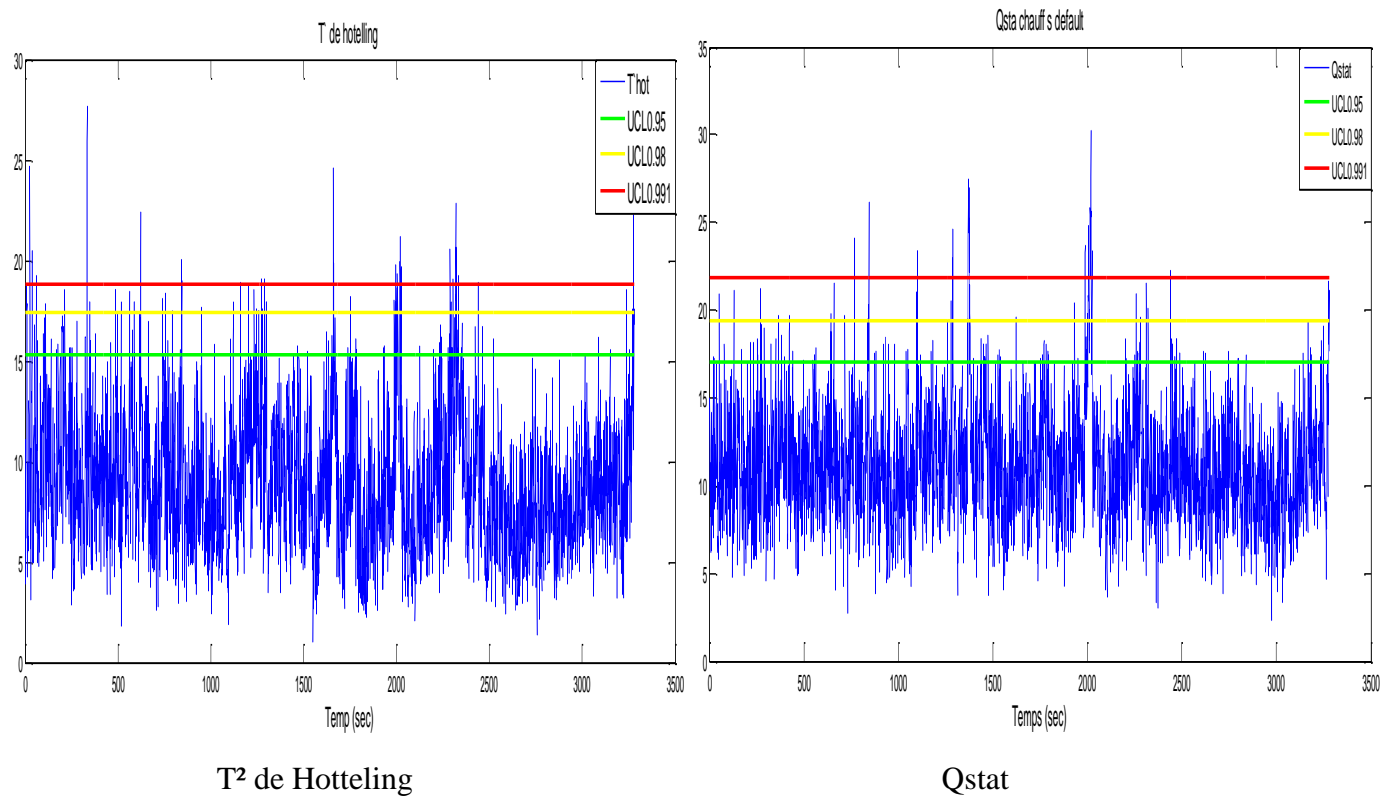


Figure 4.1 Surveillance de Four électrique sans défaut basée sur ACP.

La Figure (4.1) montre les statistiques Q et T^2 pour le four électrique. Et d'après les résultats illustrés on remarque qu'il n'y a pas une détection de défaut car les données prélevées sont en état normal (sans défaut) et il y a des dépassements sur les limites de détection qui sont considérés comme des fausses alarmes.

Tableau 4.1 Limites de contrôle de Four électrique pour différents niveaux de confiance.

La limite de confiance	Utilisant Distribution de Fisher	
	T ² de Hotelling	Qstat
95%	18.7388	5.6664
98%	24.5945	7.2051
99.1%	40.5232	8.9182

➤ **Test 2 : sur la chaudière****Figure 4.2** Surveillance de la chaudière sans défaut basée sur ACP.

La Figure (4.2) montre les statistiques Q et T^2 pour la chaudière. Et d'après les résultats illustrés on remarque qu'il n'y a pas une détection de défaut car les données prélevées sont en état normal (sans défaut) et il y a des dépassements sur les limites de détection qui sont considérés comme des fausses alarmes.

Tableau 4.2 Limites de contrôle de la chaudière pour différents niveaux de confiance.

La limite de confiance	Utilisant Distribution de Fisher	
	T ² de Hotteling	Qstat
95%	16.5008	11.1417
98%	18.6221	12.8941
99.1%	20.3198	14.7031

➤ **Test 3:** sur le groupe électrogène

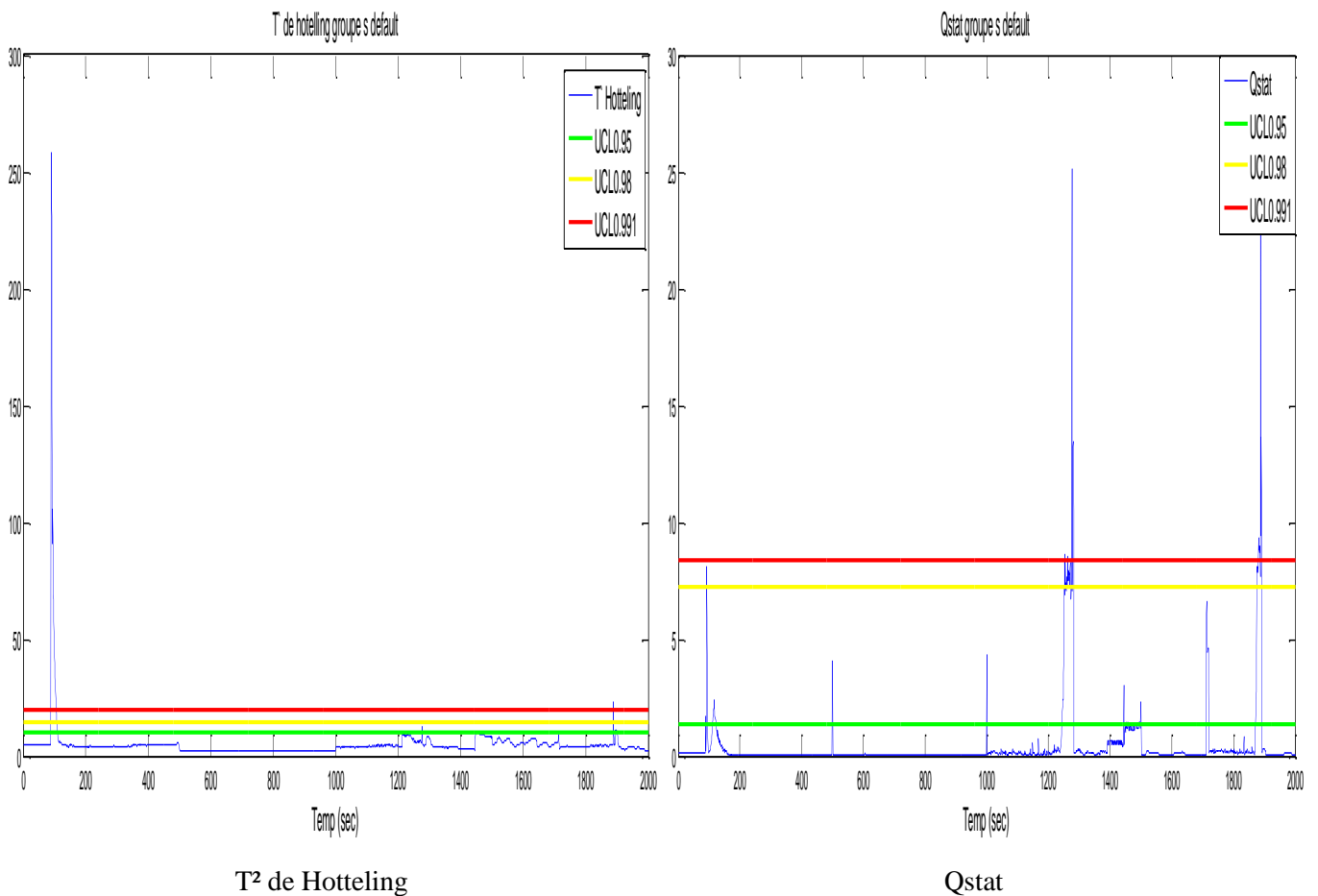


Figure 4.3 Surveillance de Groupe sans défaut basée sur ACP

La figure (4.3) montre les statistiques Q et T^2 pour la groupe. Et d'après les résultats illustrés on remarque qu'il n'y a pas une détection de défaut car les données prélevées sont en état normal (sans défaut) et il y a des dépassements sur les limites de détection qui sont considérés comme des fausses alarmes. Mais les résultats ne sont pas clairs parce qu'il y a un manque d'informations à cause de l'arrêt de groupe dans les périodes de temps.

Tableau 4.3 Limites de contrôle de groupe pour différent niveau de confiance.

La limite de confiance	Utilisant Distribution de Fisher	
	T ² de Hotteling	Qstat
95%	8.3725	1.3903
98%	9.9283	7.2369
99.1%	19.7608	8.3972

Tableaux(4.1),(4.2),(4.3). On remarque que la valeur limite de contrôle augmente quand on augmente le pourcentage de confiance **Ci**.

Les Figures (4.1), (4.2), (4.3), montrent les statistiques *Q* et *T²* pour les trois systèmes utilisés. D'après les résultats illustrés on remarque qu'il n'y a pas une détection de défaut car les données prélevées sont en état normal (sans défaut) et d'après les résultats configurant dans les graphes précédents il y a toujours des fausses alarmes pour les différentes combinaisons des composantes principales, on les rattrape par une algorithmme qui sert a calculé la somme des seuils qui dépasse la limite de contrôle a un pourcentage choisie et les tableaux suivants montrent les différents pourcentages.

Tableau 4.4 Fausses alarmes de T² de Hotteling.

Temps % / Systèmes	Etat initial	Ts=3sec → 30%	Ts=5sec → 50%	Ts=8sec → 80%
Groupe électrogène	S= 99	S=71	S=58	S=46
Chaudière	S=341	S=33	S=14	S=4
Four	S=7629	S=4066	S=3018	S=2089

Tableau 4.5 Fausses alarmes de Qstat.

Temps % / Systèmes	Etat initial	Ts=3sec → 30%	Ts=5sec → 50%	Ts=8sec → 80%
Groupe électrogène	S= 18	S=6	S=1	S=0
Chaudière	S=30	S=0	S=0	S=0
Four	S=135	S=29	S=18	S=13

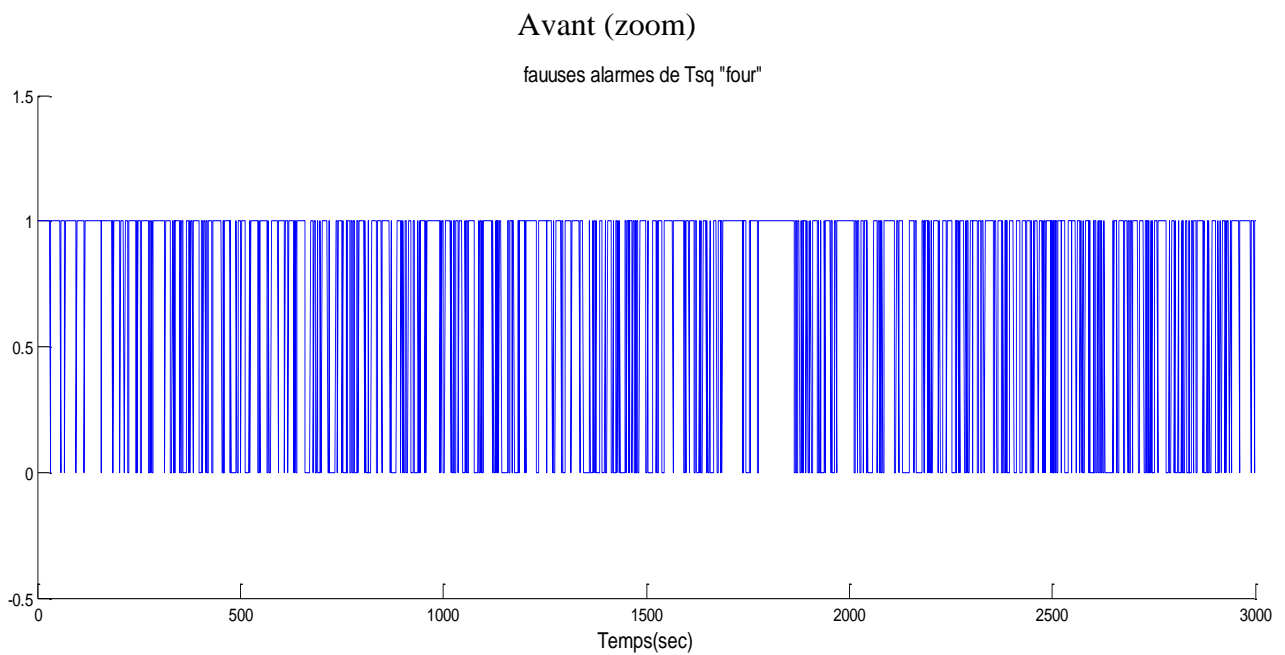
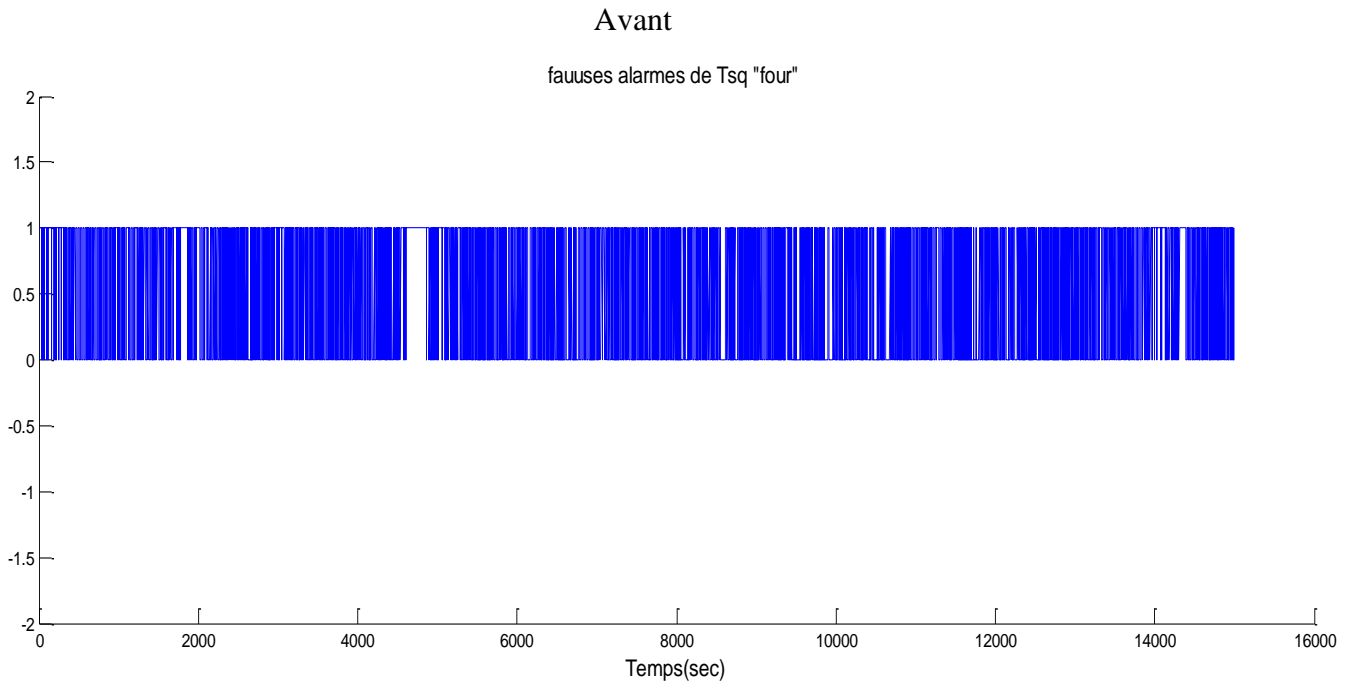
Où **S** est la somme des seuils qui dépasse la limite de contrôle.

Les résultats configurant dans les tableaux (4.4) et (4.5) nous permettent de dire qu'il y a toujours des fausses alarmes pour les différentes combinaisons des composantes principales et pour les pourcentages de risque 1%, 2% et 5%.

4.3 Résultats de la simulation des fausses alarmes

4.3.1 fausses alarmes de four électrique

➤ T^2 de Hotelling :



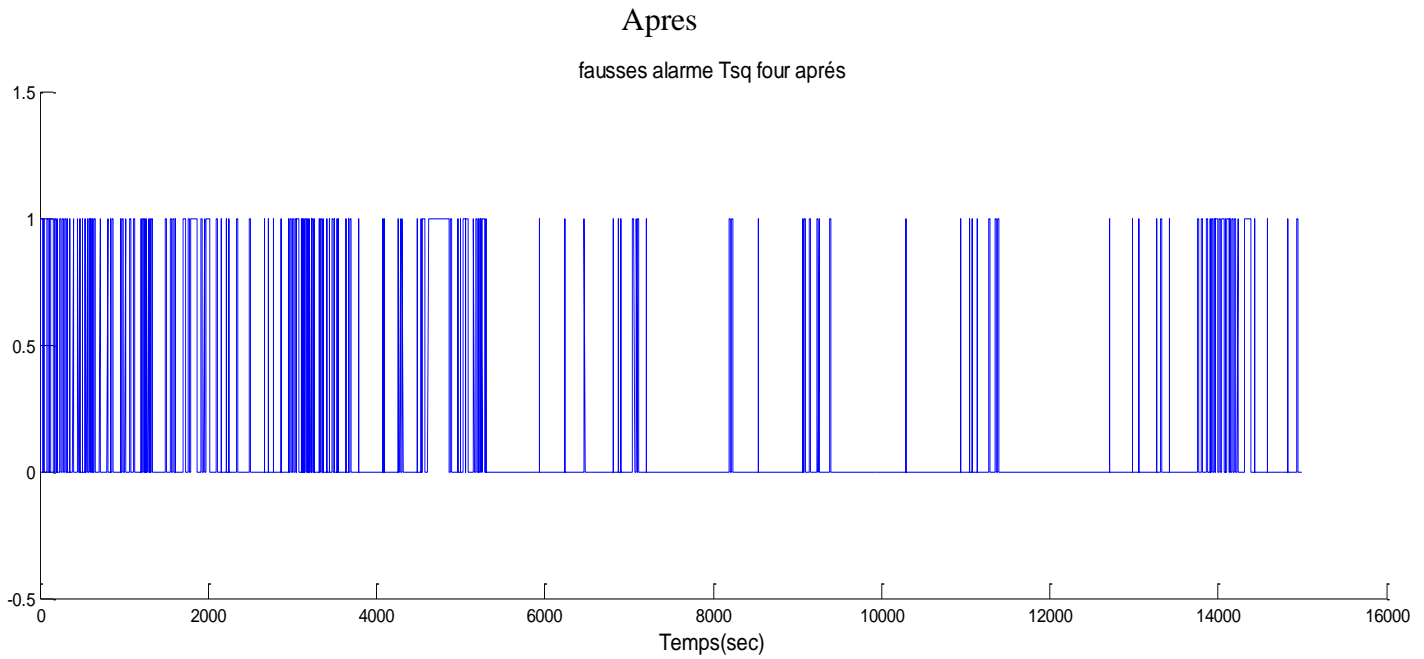


Figure 4.4 La diminution des fausses alarmes de four électrique pour T^2 de Hotteling.

➤ Qstat

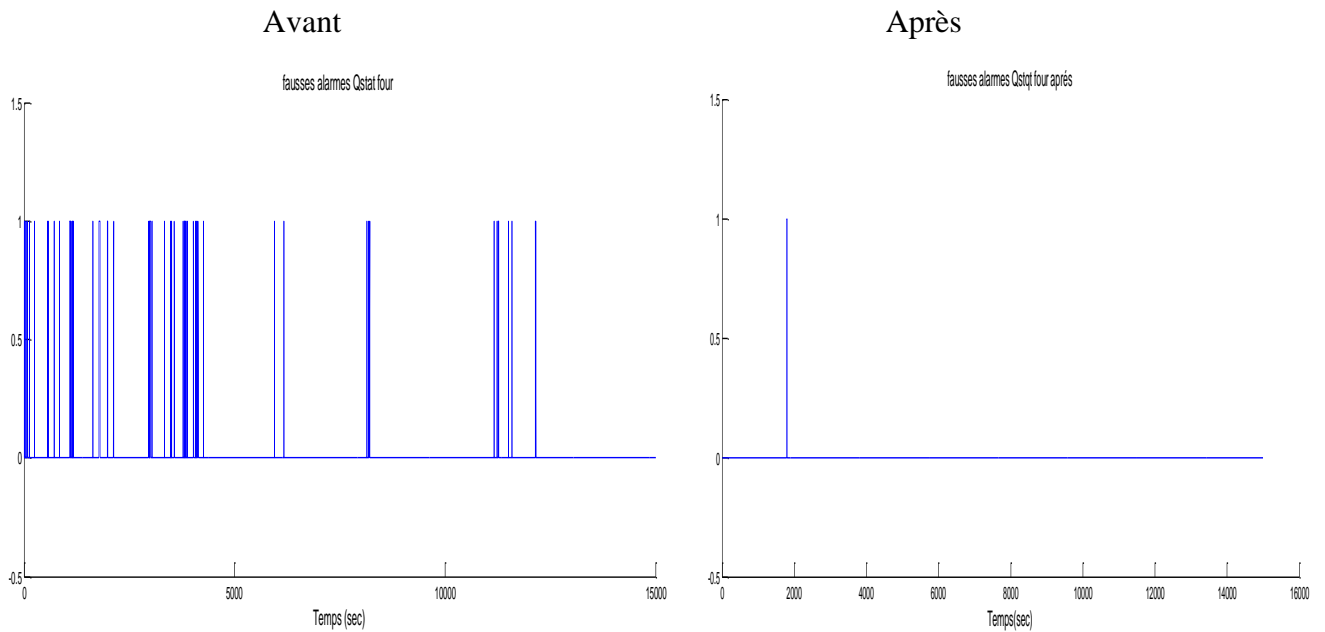


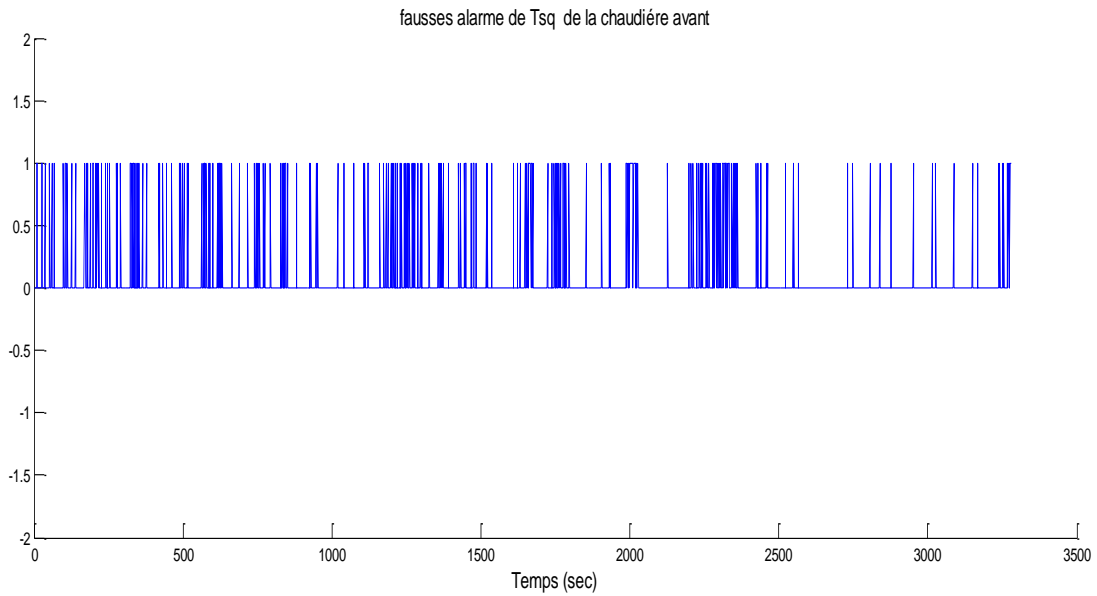
Figure 4.5 La diminution des fausses alarmes de four électrique pour Qstat.

Figure (4.4) et figure (4.5) D'après les résultats obtenus, on observe qu'il y'a toujours des fausses alarmes dans les différents pourcentages de risque pour le Qstat et T^2 Hotteling. On a essayé de diminué les fausses alarmes au maximum.

4.3.2 fausses alarmes de la chaudière

➤ T² de Hotteling :

Avant



Après

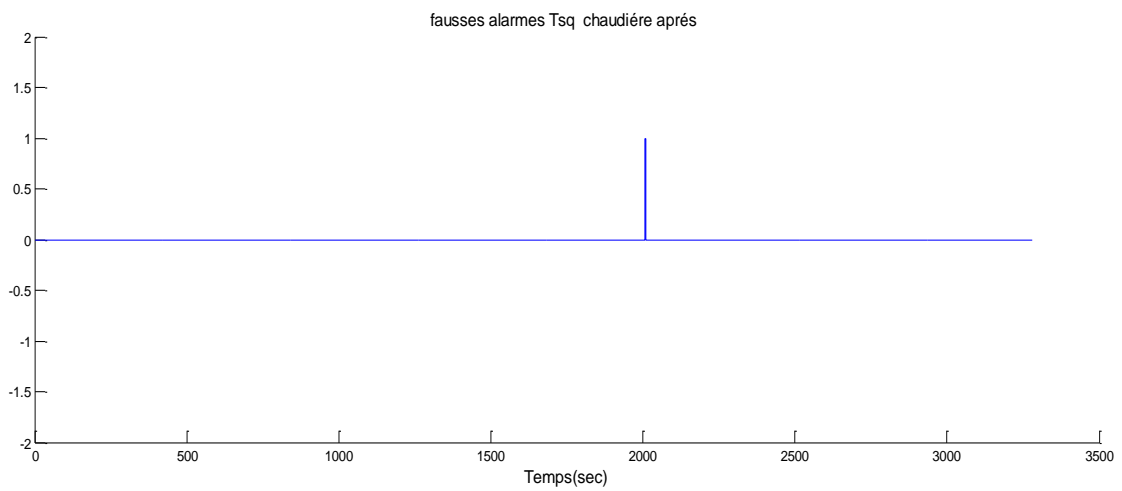


Figure 4.6 La diminution des fausses alarmes de la chaudière Pour T² de Hotteling

➤ Qstat :

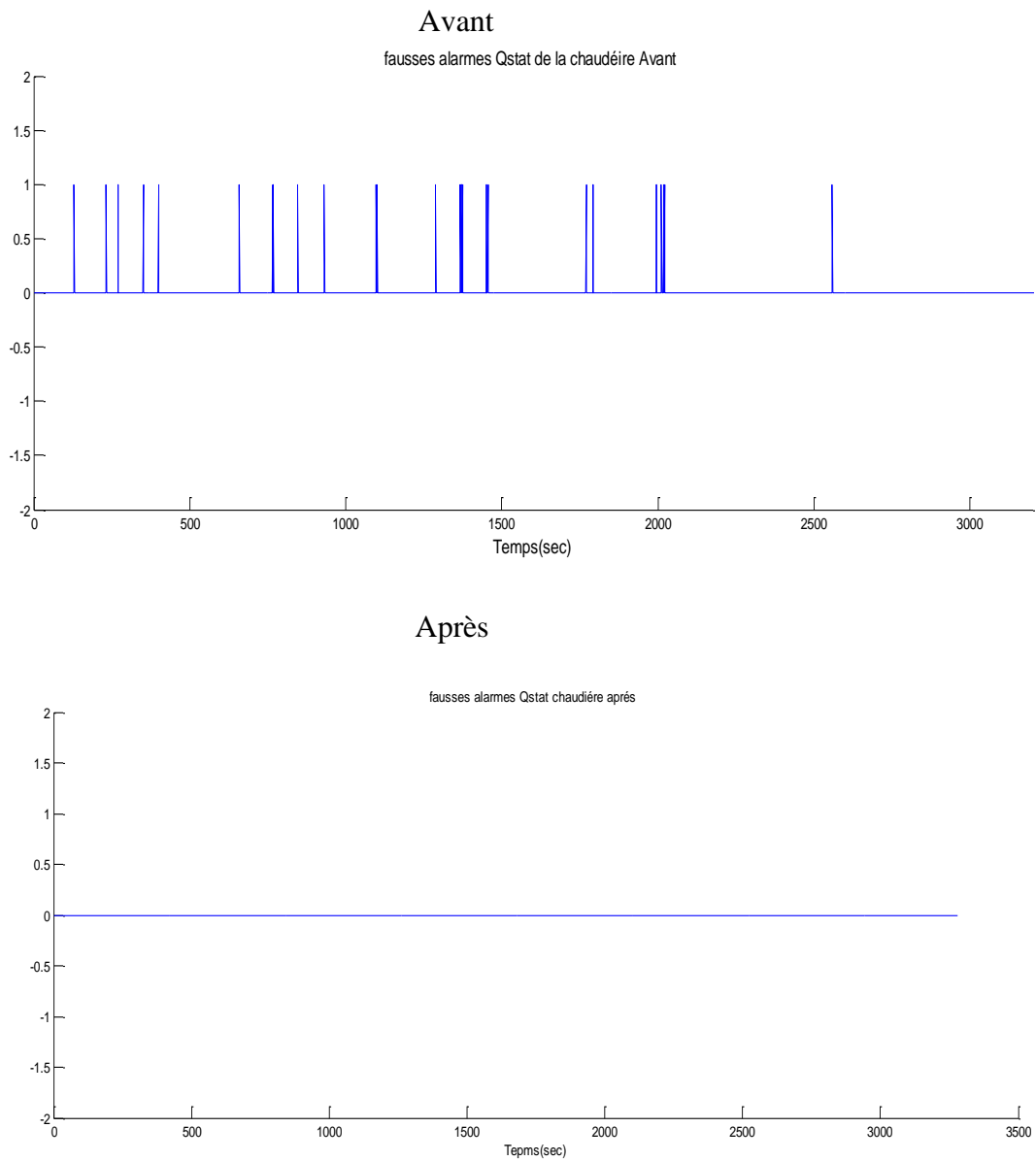


Figure 4.7 La diminution des fausses alarmes de la chaudière pour Qstat.

Figure (4.6) et figure (4.7) D'après les résultats obtenus, on observe qu'il y 'a toujours des fausses alarmes dans les différents pourcentages de risque pour le Qstat et T² Hotelling. On a essayé de les diminués, dans ce cas il n'y a pas des fausses alarmes.

4.3.3 fausses alarmes de Groupe électrogène

➤ T² de Hotteling :

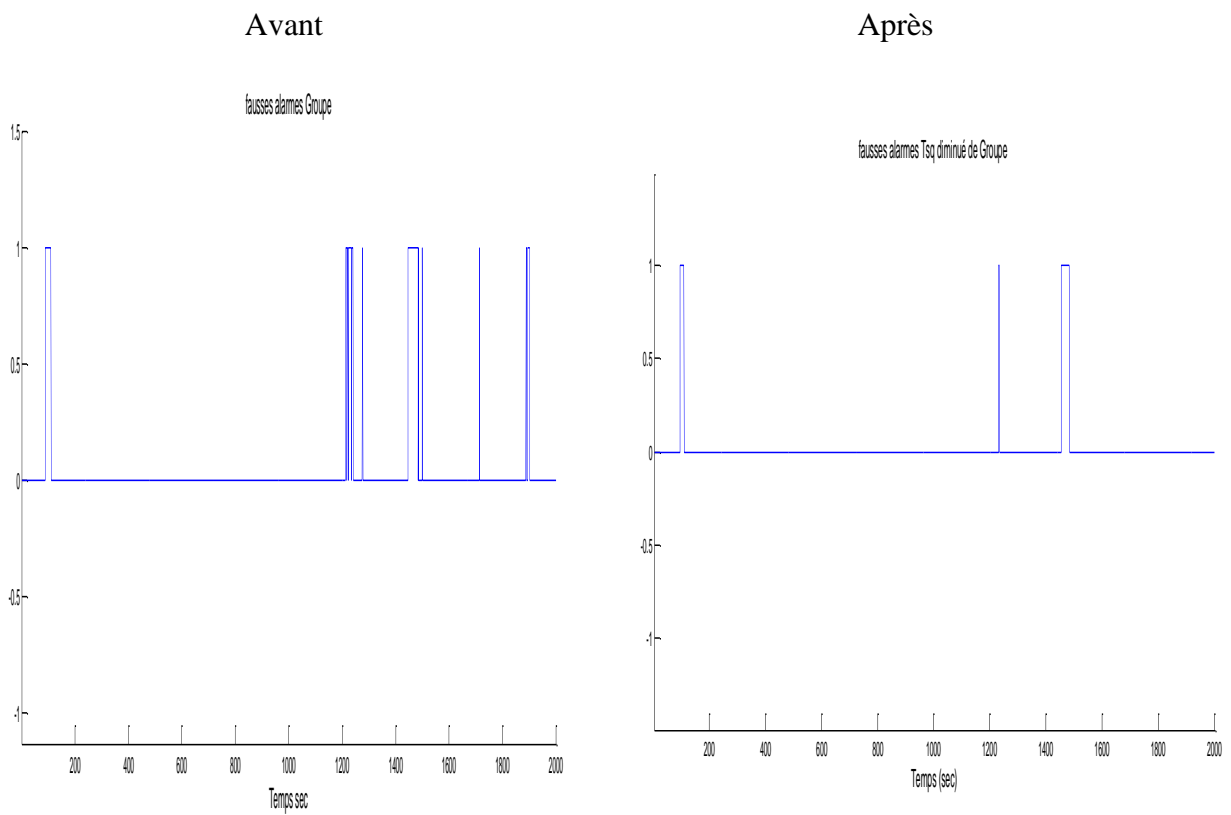


Figure 4.8 La diminution des fausses alarmes de groupe pour T² de Hotteling.

➤ Qstat :

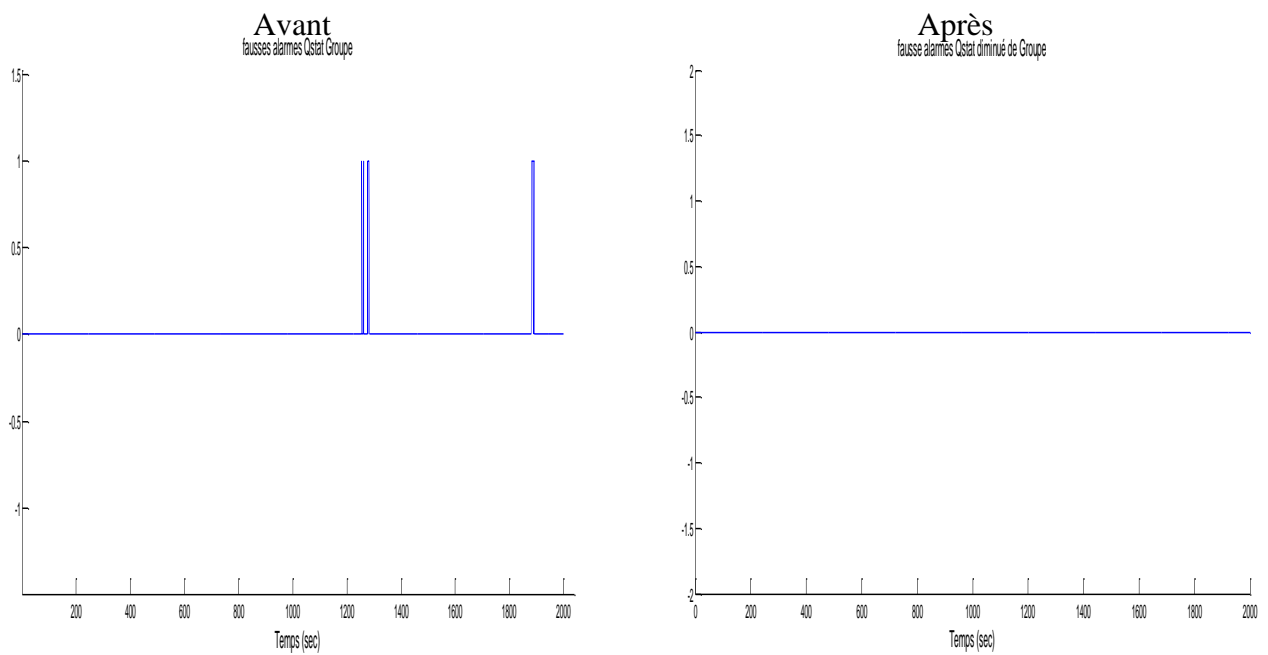


Figure 4.9 La diminution des fausses alarmes de groupe pour Qstat

Figure (4.8) et figure (4.9) D'après les résultats obtenus, on remarque qu'il y'a toujours des fausses alarmes dans les différents seuils de détection sauf que la figure (4.9) pour Qstat.

4.4 Test de L'ACP sur les trois systèmes a l'état de fonctionnement anormal (avec défaut)

➤ **Test 1** : sur le four électrique (cas défaut réel)

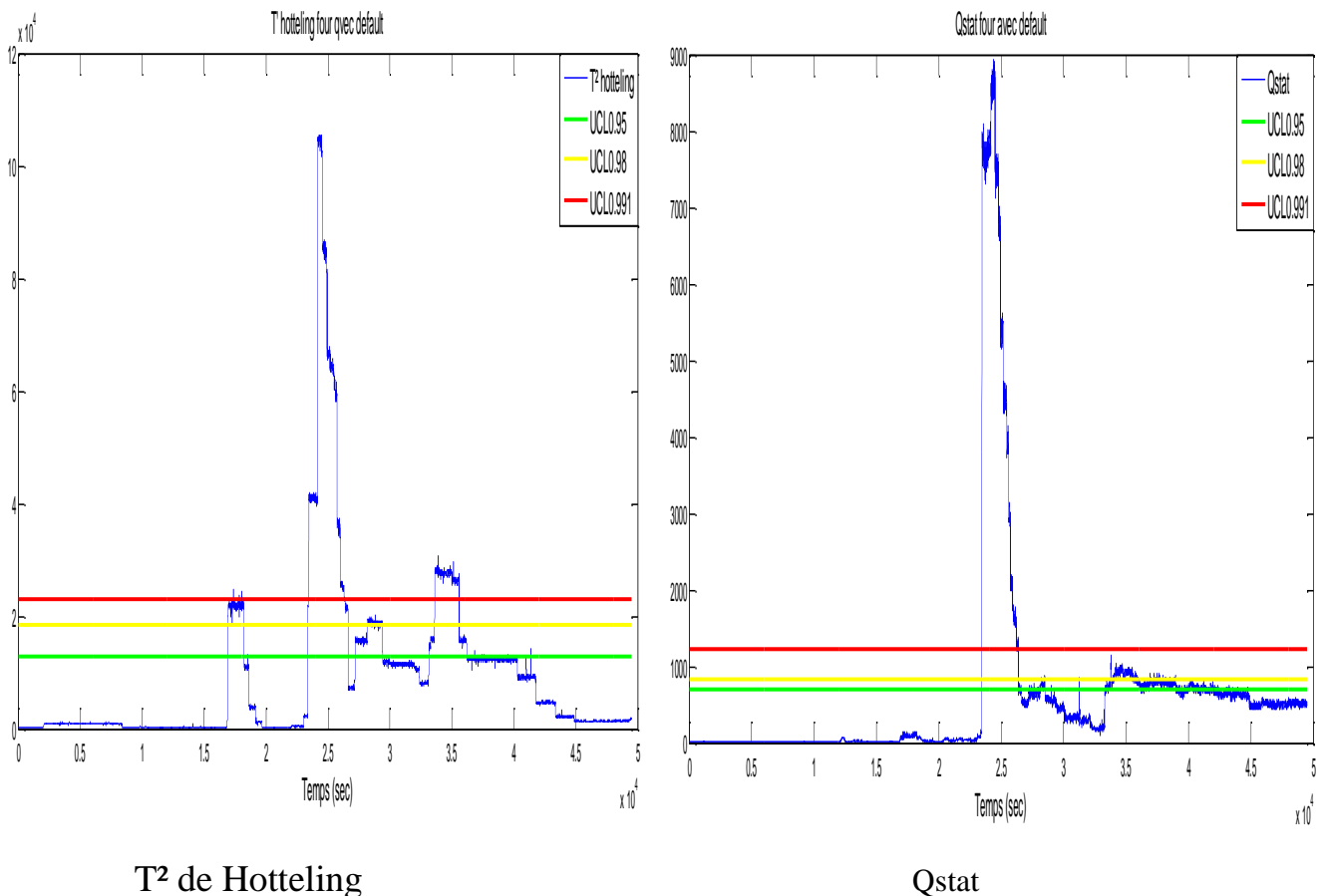


Figure 4.10 Surveillance de Four électrique avec un défaut en temps réel basée sur ACP.

D'après les résultats illustrés on remarque qu'il y a une détection de défaut dans un cas où les données sont réelles et on voit bien dans la figure (4.10) que les statistiques Q et T^2 réagissent lors de l'apparition de ce défaut.

➤ Test 2 : sur la chaudière

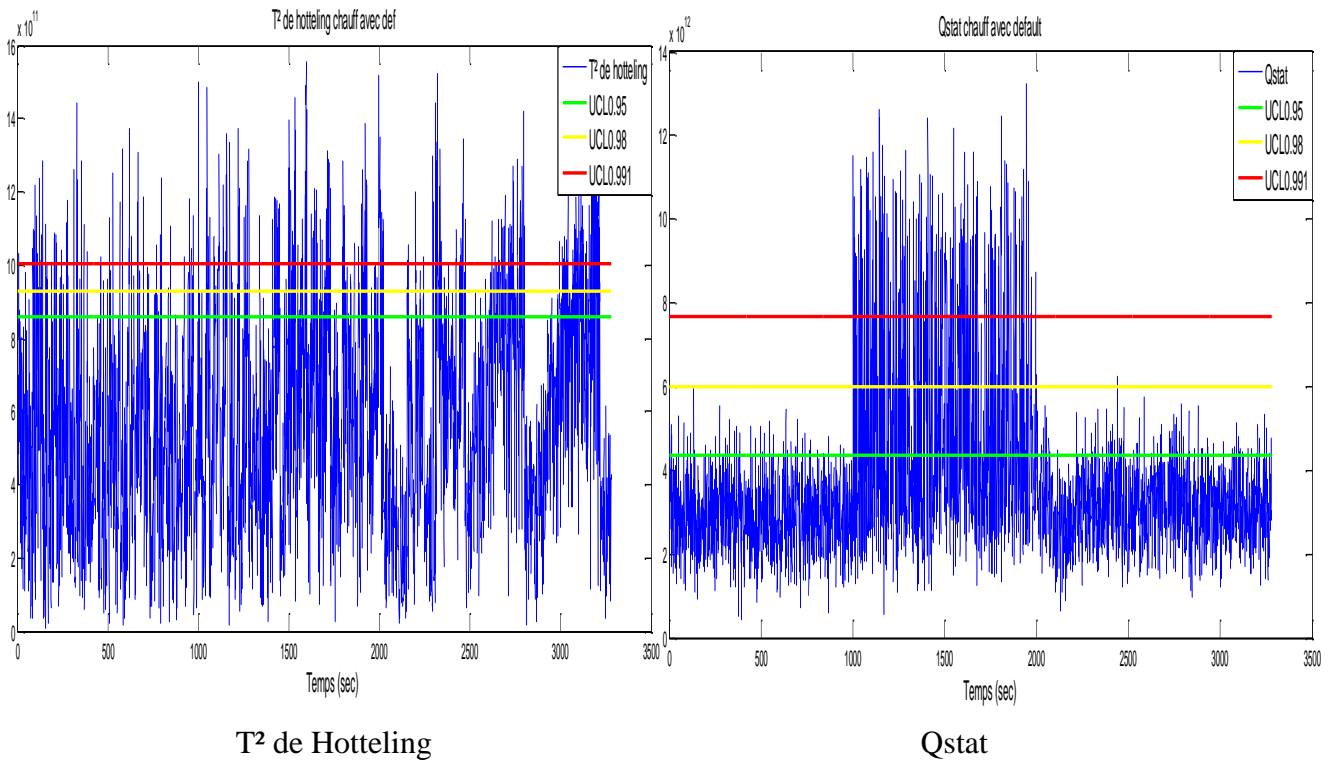


Figure 4.11 Surveillance de la chaudière avec défaut basée sur ACP.

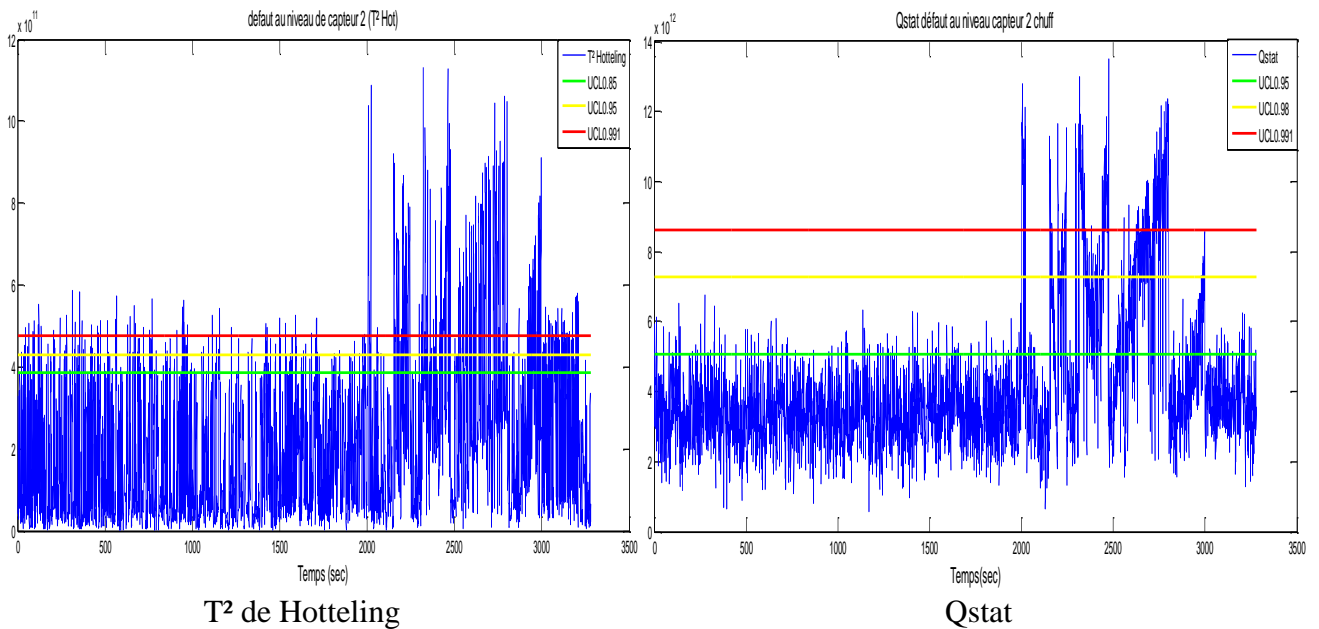


Figure 4.12 Surveillance de la chaudière avec défaut basée sur ACP.

Figure (4.11) Création de défaut au niveau des capteurs 3,4 et 5 entre (1000 sec et 2000 sec) et, Figure (4.12) création de défaut au niveau de capteur 2 entre (2000 sec et 3000 sec). Les statistiques Q et T^2 réagissent lors de l'apparition de ces défauts.

➤ **Test 3** : sur le groupe électrogène SDMO GT1900 :

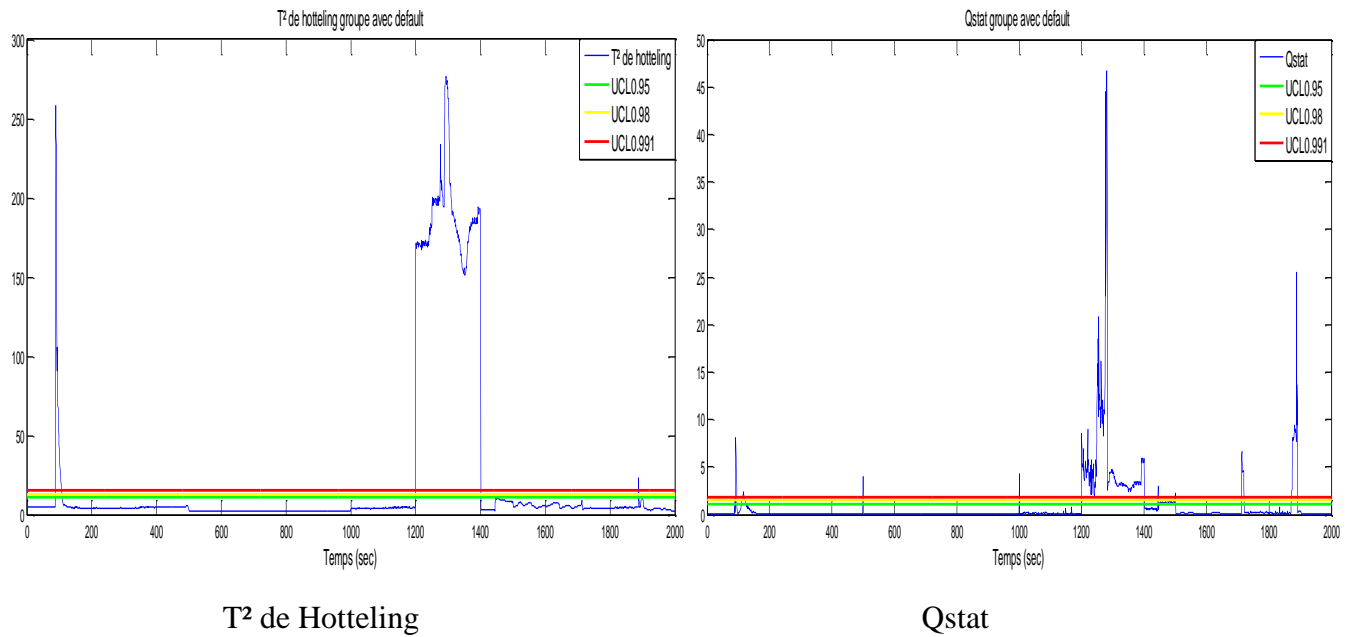


Figure 4.13 Surveillance de groupe avec défaut basée sur ACP.

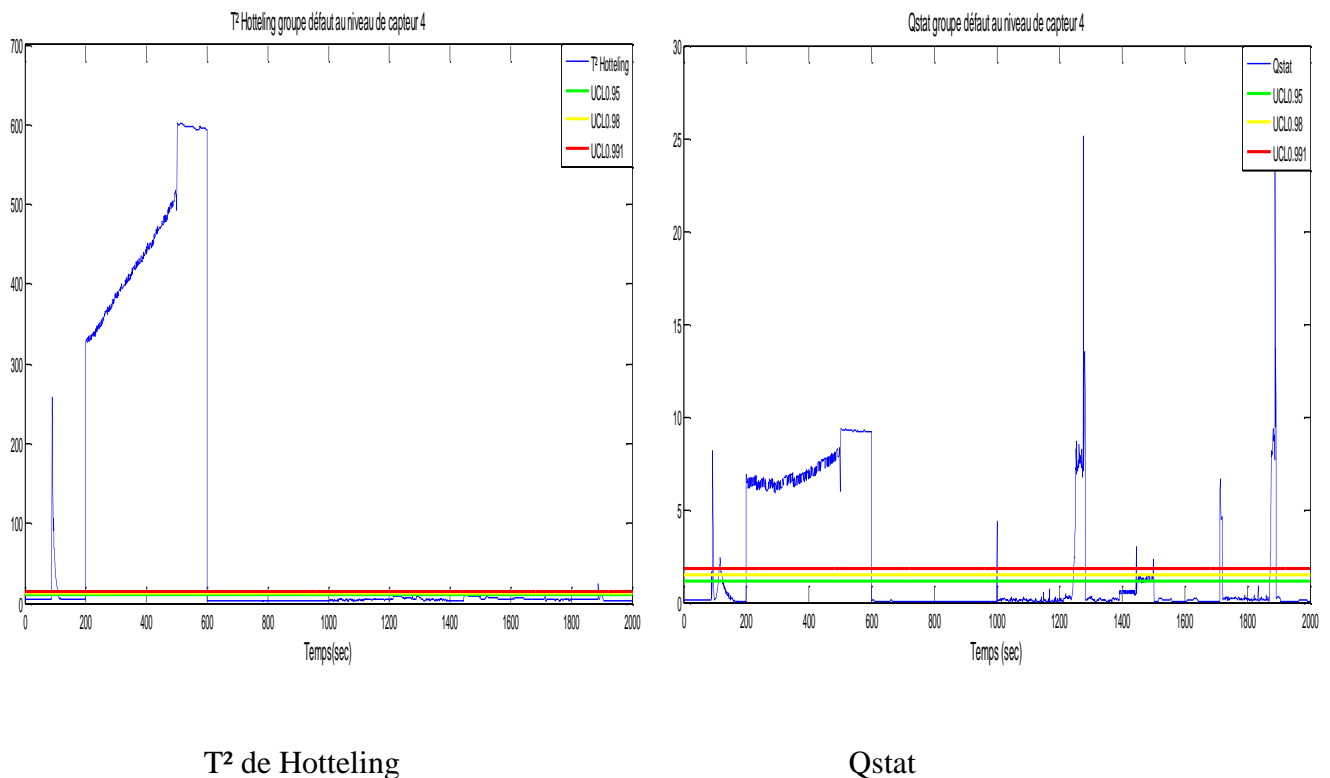


Figure 4.14 Surveillance de groupe avec défaut basée sur ACP.

Figure (4.13) Création de défaut au niveau des capteurs 3,4 et 5 entre (1200 sec et 1400 sec) et la figure (4.14) Création de défaut au niveau de capteur 2 entre (200 sec et 600 sec) et les statistiques Q et T^2 réagissent lors de l'apparition de ce défaut.

Dans les deux cas précédents le défaut a été crié, c-t-d ce n'est pas un défaut réel, nous avons modifié quelque composants comme c'est dans le premier cas de groupe, on a multiplié 200 valeurs entre (1200 et 1400) des capteurs 3,4 et 5 par 1.5, ensuite une deuxième fois on a multiplié 300 valeurs entre (200 et 600) de 4ième capteur par 1.7, et dans le 2ème cas de la chaudière on a multiplié aussi 1000 valeurs entre (1000 et 2000) des capteurs 3,4 et 5 par 1.9 et 1000 entre (2000 et 3000) de 2ième capteur par 2, on remarquent que le défaut été détecté clairement et les figures (4.11),(4.12),(4.13),(4.14) montrent comment ces modifications ont influé.

4.5 Conclusion

Dans ce chapitre, nous avons présenté les différents résultats obtenus lors de l'application de l'ACP développée dans les chapitres 2 et 3 sur un four électrique, une chaudière et un groupe électrogène contenant des variables aux les états de fonctionnement normal et anormal.

En premier lieu, l'obtention des composantes principales est effectuée. Cette dernière est basée sur le calcul de la matrice de corrélation ainsi que les vecteurs et les valeurs propres.

L'application de pourcentage de la variance cumulée (PCV) est imposée pour le choix du nombre de composantes principales à retenir. Une fois que les composantes principales sont déterminées on applique les statistiques de Hotelling et Qstat pour la détection des défauts.

Les statistiques de Hotelling et Qstat nécessitent un seuil de détection bien déterminé car le mauvais choix de ce seuil indique des fausses alarmes dans l'état de fonctionnement normal.

Conclusion générale

L'objectif du diagnostic est de détecter rapidement les divers défauts existant sur un procédé (défaut d'instrumentation, des paramètres, et de structure) pour éviter la dégradation de ses performances et augmenter la sécurité des opérateurs et de l'environnement.

Notre but a donc été d'utiliser l'analyse en composantes principales (ACP) pour la détection des divers défauts pour les systèmes industriels.

Dans ce travail, l'ACP est utilisée comme une technique de modélisation des relations entre les différentes variables du groupe électrogène, de la chaudière, et du four électrique. L'estimation des paramètres du modèle ACP est effectuée par estimation des valeurs et des vecteurs propres de la matrice des corrélations des données. Cependant, pour la détermination de la structure du modèle, il faut déterminer le nombre de composantes à retenir dans ce modèle (nombre de vecteurs propres). Pour cette raison, plusieurs critères de sélection du nombre de composantes ont été présentés.

Une fois le modèle identifié, la procédure de détection de défauts peut être effectuée.

Dans notre travail on a utilisé deux approches pour la détection.

- **La statistique Q** : qui est un test global qui cumule les erreurs de modélisation présentes sur chaque résidu.
- **T² d'Hotelling** : Les valeurs T² d'Hotelling représentent une mesure de la variation de chaque échantillon dans le modèle. Cette statistique est définie par la somme des scores au carré normalisée.

Nous avons utilisé dans notre projet une méthode qui utilise un seuil fixe. Cette méthode consiste à détecter l'instant pour lequel les grandeurs surveillées sortent de la zone limitée par des seuils autour de l'objectif (valeur de la grandeur au sein du modèle en fonctionnement normal).

L'étude menée est exploratrice et nous nous sommes donc limités à la détection et ceci sur un certain nombre de mesures. Ce qui se dégage de ce travail est que la détermination du modèle ACP est l'étape la plus délicate de cette méthode, parce qu'elle dépend de la précision de la détection.

Bibliographie

Bibliographie

- [1] “*Model-Based Fault Diagnosis in Dynamic Systems Using Identification Techniques*”, Ph.D. Thèse ,Silvio Simani, Université de Ferrare.Italy, Département d'ingénierie, 1999.
- [2] “*Fault-diagnosis system–An introduction for fault detection to fault tolerance*”, R. Isermann, Springer, Heidelberg, 2006.
- [3] V. Venkatasubramanian, R. Rengaswamy, K. Yin, Surya N. Kavuri. *A review of Process Fault Detection and Diagnosis part 1: Quantitative model-based methods*, Computers et Chemical Engineering, vol. 27, pp.293-311,2003.
- [4] Alkan Alkaya, *Novel Data Driven-Based Fault Detection for Electro Mechanical and Process Control Systems*, Ph.D. Thèse, Çukurova University, Institut des sciences naturelles et appliquées Département de l'ingénierie électrique et électronique, 2012
- [5] V. Venkatasubramanian, R. Rengaswamy, Surya N. Kavuri, K. Yin, *A Review of Process Fault Detection and Diagnosis, Process History part III: Process history based methods*, Computers & Chemical Engineering, vol. 27, pp.327-346, 2003.
- [6] E. Sobhani-Tehrani, K. Khorasani, “*Fault diagnosis of nonlinear systems using a hybrid approach*”, Springer, 2009.
- [7] Herve Abdi and Lynne J. Williams, “*Principal component analysis*”, John Wiley & Sons, Inc. WIREs Comp Stat, vol. 2, pp. 433–459, 2010.
- [8] I.T. Jolliffe, “*Principal component analysis*”, Springer, New York, 2002.
- [9] Dawen Liang, “*Maximum Likelihood Estimator for Variance is Biased: Proof*”, L'université de Carnegie Mellon.
- [10] Jonathon Shlens, “*A Tutorial on Principal Component Analysis*”, Laboratoire de neurobiologie des systèmes, Institut Salk pour les études biologiques, 2005.
- [11] Aly A. Farag, Shireen Elhabian, “*A tutorial on principle component analysis*”, Université de Louisville, CVIP Lab, 2009
- [12] 36-350, Data Mining, “*The Truth about Principal Components and Factor Analysis*”, September 2009.

- [13] Donald. A. Jackson, "Stopping rules in principal component analysis: a comparison of heuristical and statistical approaches", *Société écologique d'Amérique*, vol. 74 (8), pp. 2204-2214, 1993.
- [13]I. T. Jolliffe, "Discarding Variables in a Principal Component Analysis. I: Artificial Data", *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol. 21 (2), pp. 160-173, 1972.
- [14] William R. Zwick and Wayne F. Velicer, "Comparison of Five Rules for Determining the Number of Components to Retain", *Psychological Bulletin*, Vol. 99 (3), 452-442, 1986.
- [15] P. LEGENDRE, L. LEGENDRE, "Numerical ecology", 2nd edition, Elsevier Science B.V, Amsterdam, pp. 409-410, 1998
- [16] Richard Cangelosi, Alain Goriely, "Component retention in principal component analysis with application to cDNA microarray data", *Biology Direct*, vol. 2 (2), 2007.
- [17] Wayne F. Velicer, "Determining the number of components from the matrix of partial correlations", *PSYCHOMETRIKA*, vol. 41 (3), pp. 321-327, 1976.
- [18] Brian P. O'Connor, "SPSS and SAS Programs for Determining the Number of Components Using Parallel Analysis and Velicer's MAP Test", *Behavior Research Methods, Instruments, & Computers*, vol. 32, pp. 396-402, 2000.
- [19] Sami Romdhani, Shaogang Gong, Alexandra Psarrou, "A Multi-View Nonlinear Active Shape Model Using Kernel PCA", *BMVC*, pp. 483-492, 1999.
- [20] J.F. MacGregor, T. Kourtl, "Statistical process control of multivariate processes", *Control Eng. Practice*, vol. 3 (3), pp. 403-414, 1995.
- [21] Emmanuel. J. Candés, Xiaodong Li, Yi Ma, John Wright, "Robust Principal Component Analysis", *Journal of the ACM*, vol. 58 (11), 2011.
- [22] Nomikos. P., MacGregor. J, "Multivariate SPC Charts for Monitoring Batch Processes". *Technometrics*, vol. 37, pp. 41-59, 1995.
- [23] Jackson. J, Mudholkar. G, "Control Procedures for Residuals Associated with Principal Component Analysis". *Technometrics*, vol. 21 (3), pp. 341-349, 1979
- [24] John L. Horn, 'A rationale and test for the number of factors in factor analysis', *PSYCHOMETRIKA*, vol. 30 (2), pp. 179-185, 1965

[25] Chiang. L, Russell. E, Braatz. R, “*Fault Detection and Diagnosis in Industrial Systems*”, Springer-Verlag, London, UK, 2001.

[26] Radu. Cascaval, “*Eigenvalues, Singular Value Decomposition*”,
Département de mathématiques, Université du Colorado.

Annexes

Annexes

1. Décomposition en valeurs singulières (SVD)

Une décomposition en valeur singulière d'une matrice $M \in \mathbb{C}^{n \times m}$ avec $\text{rank}(M) = r$ et $m \leq n$ est une factorisation :

$$M = U \cdot \Sigma \cdot V^*$$

avec U une matrice unitaire $m \times m$ sur \mathbb{C} , Σ une matrice $m \times n$ dont les coefficients diagonaux sont des réels positifs ou nuls et tous les autres sont nuls, et V^* est la matrice adjointe à V , matrice unitaire $n \times n$ sur K . On appelle cette factorisation la *décomposition en valeurs singulières* de M .

- Les entrées diagonales de Σ sont appelées valeurs singulières de M .
- Les colonnes de U sont appelées vecteurs singuliers de gauche de M .
- Les colonnes de V sont appelées vecteurs de droite de M .

Il existe de nombreux algorithmes dans la littérature pour calculer les matrices de décomposition en valeurs singulières pour une matrice rectangulaire donnée. Le calcul de SVD repose essentiellement sur le calcul des valeurs propres et des vecteurs propres, qui peut être effectué soit par des algorithmes numériques standard, soit par l'utilisation d'algorithmes plus sophistiqués selon la structure de la matrice.

Le calcul de la SVD de la matrice M est, au moins en théorie, équivalent au calcul des valeurs propres et des vecteurs propres pour la matrice $A^* \cdot A$. La décomposition en valeurs singulières a trouvé un large éventail d'applications dans la science moderne, telles que la compression d'image, l'analyse de réseau et l'exploration de données (reconnaissance du visage et de l'écriture manuscrite, analyse des composants Principal) [26].

2. Les entrées et les sorties de l'APM802 de Groupe électrogène SDMO GT1900

2.1 Les entrées module de base

- ✓ Marche à distance (BLACKSTART).
- ✓ Défaut niveau bas eau.
- ✓ Disjoncteur alternateur fermé.
- ✓ Disjoncteur alternateur ouvert.
- ✓ Disjoncteur alternateur disjoncté.
- ✓ Disjoncteur secours fermé.
- ✓ Disjoncteur secours ouvert.
- ✓ Disjoncteur secours disjoncté.
- ✓ Alarme chargeur batterie de démarrage.
- ✓ Défaut disjonction circuit mesure.
- ✓ Absence tension réseau.
- ✓ Alarme d'isolement.
- ✓ Défaut bobinage aux.
- ✓ Disjoncteur secours débrosche.
- ✓ Disjoncteur normal ferme.
- ✓ Disjoncteur normal ouvert.
- ✓ Disjoncteur normal disjoncté.
- ✓ Arrêt d'urgence.

2.2 Les sorties module de base

- ✓ Ordre de fermeture disjoncteur alternateur.
- ✓ Ordre d'ouverture disjoncteur alternateur.
- ✓ Groupe disponible en automatique.
- ✓ Synthèse alarme.
- ✓ Synthèse défaut.
- ✓ Marche groupe.
- ✓ Commande préchauffage eau.
- ✓ Commande 029-09 démarreur N°1.
- ✓ Commande arrêt groupe.
- ✓ Commande pré-excitation.
- ✓ Commande excitation.
- ✓ Ordre de fermeture disjoncteur secours.
- ✓ Ordre d'ouverture disjoncteur secours.
- ✓ Commande avertisseur sonore.
- ✓ Automatisation couplage normal / secours.
- ✓ Ordre de fermeture disjoncteur normal.
- ✓ Ordre de fermeture disjoncteur normal.
- ✓ Watch dog (chien de garde).