

**REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE**  
**MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE**  
**SCIENTIFIQUE**  
**UNIVERSITE KASDI MERBAH - OUARGLA**  
**FACULTÉ DES NOUVELLES TECHNOLOGIES DE L'INFORMATION ET DE**  
**TÉLÉCOMMUNICATIONS**  
**DÉPARTEMENT D'ÉLECTRONIQUE ET DES COMMUNICATIONS**



**Filière : Automatique**

Spécialité : AUTOMATIQUE ET SYSTEMES

Mémoire présenté pour l'obtention du diplôme de MASTER

**Présenté Par:**

**CHIBANI Messaoud & TELLI Lokmane**

**Thème :**

***DIAGNOSTIC DES DEFAUTS POUR LA SURVEILLANCE  
DES PROCESSUS NON LINEAIRE PAR KARNAEL PCA***

Soutenu publiquement

Le :15/06/2023

Devant le jury :

Mr	BENSID Khaled	MCB	Président	UKM Ouargla
Mr	CHEBBARA Fouad	Professeur	Examineur	UKM Ouargla
Mm	KARA Fouzia	MAA	Encadreur	UKM Ouargla

**Année Universitaire: 2022/2023**



## *Remerciements*

*Nous tenons tout d'abord à exprimer notre profonde gratitude à Dieu Tout-Puissant qui nous a accordé la force, le courage et la patience pour mener à bien cette tâche avec succès. L'exécution de ce message n'aurait pas été possible sans la contribution de nombreuses personnes, et nous souhaitons exprimer notre reconnaissance et notre profonde gratitude envers elles. Nous tenons à exprimer notre gratitude et notre reconnaissance à la professeure **Kara Fouzia**, qui a accepté de superviser et de guider ce travail scientifique et technique, et qui l'a surveillé attentivement jusqu'à la fin. Nous la remercions également pour sa patience et sa disponibilité, notamment pour ses précieux conseils qui ont contribué à orienter nos idées. Nous souhaitons également exprimer notre gratitude au professeur docteur **Chebbara Fouad** qui a accepté de présider le comité de discussion de ce travail, ainsi qu'au professeur docteur **Bensidkhaled** qui a accepté de participer à la discussion.*

*Nous sommes reconnaissants envers toutes les personnes qui ont contribué, soutenu et encouragé ce travail, et nous leur souhaitons à tous le meilleur et beaucoup de succès dans leur vie future.*



## الملخص

تحليل المكونات الرئيسية (ACP) هو أداة قوية جدًا لاستخراج المعلومات المفيدة عندما يكون لدينا كمية كبيرة من البيانات لمعالجتها وتفسيرها. في السنوات الأخيرة، أدى ظهور برامج أكثر كفاءة وسهولة في الاستخدام إلى جعل هذا الشكل من تحليل البيانات متاحًا لجميع الباحثين العلميين، وليس فقط المتخصصين.

الغرض من مشكلة تشخيص الأخطاء في التحكم غير الخطي في العمليات هو تطوير طرق تحليلية جديدة. من المهم ملاحظة أن معظم الأنظمة في الواقع غير خطية، في حين أن تقنية (ACP) خطية. وبالتالي، فإن (KACP) هو امتداد لتحليل (ACP) الخطي، باستخدام النواة للكشف عن الفشل في العمليات غير الخطية. ويمثل تحليل هذه العمليات تحديًا كبيرًا بسبب طبيعتها غير الخطية وما يرتبط بها من تعقيد.

وتعتبر نواة (ACP) هي الأداة الرئيسية لتحقيق هذا الهدف، حيث تقدم نهجًا قويًا لتحليل البيانات غير الخطية. تم تطوير طرق محسنة لتعزيز القدرة التشخيصية وزيادة دقة التحكم في العمليات غير الخطية. تكمن أهمية هذه الأطروحة في عرض نتائج التجارب والدراسات التجريبية التي تظهر فعالية وكفاءة الطرق المقترحة لتشخيص الإخفاقات والتحكم في العمليات غير الخطية. تهدف هذه الورقة إلى المساهمة في تطوير مجال رصد العمليات غير الخطية وتشخيصها، وكذلك تحسين الكفاءة والاستدامة في هذا المجال.

**الكلمات المفتاحية:** تحليل المكون الرئيسي ذات النواة , تشخيص الاعطاب , الانظمة الخطية , الانظمة الغير الخطية

## Résumé

L'analyse en composantes principales (ACP) est un outil très puissant pour extraire de l'information utile lorsque nous avons une grande quantité de données à traiter et à interpréter. Au cours des dernières années, l'émergence de programmes de plus en plus efficaces et conviviaux a rendu cette forme d'analyse des données accessible à tous les chercheurs scientifiques, pas seulement aux spécialistes.

Le but de la question du diagnostic des erreurs dans le contrôle non linéaire des processus est de développer de nouvelles méthodes analytiques. Il est important de noter que la plupart des systèmes sont en fait non linéaires, tandis que la technologie ACP est linéaire. Ainsi, Kernel ACP est une extension de l'analyse linéaire ACP, utilisant les grains pour détecter les défaillances dans les processus non linéaires. L'analyse de ces processus est un défi majeur en raison de leur nature non linéaire et de la complexité connexe.

Le noyau de l'ACP est le principal outil pour atteindre cet objectif, offrant une approche robuste de l'analyse des données non linéaires. Des méthodes améliorées ont été développées pour améliorer la capacité de diagnostic et accroître la précision du contrôle non linéaire des processus. L'importance de cette thèse est de présenter les résultats d'expériences et d'études expérimentales montrant l'efficacité et l'efficacité des méthodes proposées pour diagnostiquer les échecs et contrôler les processus non linéaires. Ce document vise à contribuer au développement du domaine du suivi et du diagnostic des processus non linéaires, ainsi qu'à améliorer l'efficacité et la durabilité dans ce domaine.

**Mots-clés** : ACP non-linéaire, ACP à noyau, Détection et localisation de défauts, Systèmes dynamiques.

## **Abstract**

The principal component analysis (PCA) is a very powerful tool for extracting useful information when we have a large amount of data to process and interpret. In recent years, the emergence of increasingly efficient and user-friendly programs has made this form of data analysis accessible to all scientific researchers, not just specialists.

The purpose of the problem of error diagnosis in non-linear process control is to develop new analytical methods. It is important to note that most systems are non-linear, while PCA technology is linear. Thus, Kernel PCA is an extension of linear PCA analysis, using grains to detect failures in non-linear processes. Analysis of these processes is a major challenge due to their non-linear nature and associated complexity.

The core of the PCA is the main tool to achieve this goal, offering a robust approach to non-linear data analysis. Improved methods have been developed to improve diagnostic capability and increase the accuracy of non-linear process control. The importance of this thesis is to present the results of experiments and experimental studies showing the effectiveness and efficiency of the proposed methods for diagnosing failures and controlling non-linear processes. This document aims to contribute to the development of the field of monitoring and diagnosis of non-linear processes, as well as to improve

***Key-words:*** Nonlinear PCA, Kernel PCA, Fault detection and diagnosis, Dynamic system.

# Table des matières

## Table des matières

<b>Introduction générale.....</b>	<b>1</b>
<b>Chapitre 1 Généralités sur le diagnostic.....</b>	<b>3</b>
1.1 Introduction .....	3
1.2 Qu'est-ce qu'un diagnostic.....	3
1.3 Terminologie et définitions .....	4
1.4 La supervision.....	5
1.5 La surveillance .....	5
1.5.1 La détection .....	5
1.5.2 La localisation .....	5
1.5.3 L'identification.....	5
1.6 Typologie de défauts .....	6
1.6.1 Défaut actionneur .....	6
1.6.2 Défaut capteur .....	6
1.6.3 Défauts composants.....	6
1.7 Caractérisation de défauts.....	6
1.7.1 Défaut graduel .....	7
1.7.2 Défaut intermittent .....	7
1.7.3 Défaut abrupt.....	7
1.8 Classification des méthodes de diagnostic .....	7
1.8.1 Méthode à base de modèles.....	7
1.8.2 Méthodes sans modèles .....	8
1.9 Principe de diagnostic de défaut .....	8
1.9.1 Redondance physique ou matérielle.....	8
1.9.2 Redondance analytique .....	9
1.10 Conclusion .....	9
<b>Chapitre 2 Analyse en Composantes Principales .....</b>	<b>10</b>
2.1 Introduction .....	10
2.2 Principe de base de la méthode ACP.....	10
2.2.1 Principe.....	11

## Table des matières

2.2.2	Identification du modèle ACP .....	13
2.2.3	Pourcentage cumulé de la variance totale (PCV).....	14
2.2.4	Critère de validation croisée(PRESS) .....	14
2.3	Exemple de simulation .....	15
2.4	ACP non linéaire (NLPCA).....	19
2.4.1	Principe de l'ACP non linéaire.....	19
2.4.2	ACP non linéaire à noyau (kernel PCA) .....	20
2.4.3	Algorithme de l'ACP à noyau .....	21
2.5	Conclusion .....	22
<b>Chapitre 3</b>	<b>Diagnostic de défauts .....</b>	<b>23</b>
3.1	Introduction .....	23
3.2	Détection de défauts .....	23
3.2.1	Indice SPE (cas linéaire) .....	24
3.2.2	Indice SPE (cas non linéaire) .....	25
3.2.3	Indice $T^2$ (cas linéaire) .....	25
3.2.4	Indice $T^2$ (cas non linéaire) .....	27
3.2.5	Phase de modélisation .....	27
3.2.6	Phase de test .....	28
3.3	Localisation de défauts .....	28
3.3.1	La reconstruction des variables .....	28
3.4	Exemple de simulation .....	32
3.5	Conclusion .....	38
<b>Chapitre 4</b>	<b>Application.....</b>	<b>39</b>
4.1	Description du TE.....	39
4.2	Modélisation de la détection de défauts du processus TE .....	39
4.3	Résultats de la simulation .....	43
4.4	Conclusion .....	47
	<b>Conclusion générale .....</b>	<b>48</b>
	<b>Bibliographie.....</b>	<b>50</b>

## Liste des figures

### Liste des figures

Figure 1.1-Structure générale d'un système de diagnostic.....	3
Figure 1.2-Différent type de défaut s'agissant sur un système. ....	4
Figure 1.3-Evolution temporelle des différents types de défauts.....	7
Figure 1.4-Principe de diagnostic avec modèle. ....	8
Figure 2.1-Principe de la méthode ACP.....	11
Figure 2.2-Evolution des différentes variables. ....	16
Figure 2.3-Évolution des composantes principales.....	17
Figure 2.4-Evolution des mesures et leurs estimations. ....	18
Figure 2.5-Évolution de la norme PCV et le Critère PRESS. ....	18
Figure 2.6-La méthode ACP. ....	19
Figure 2.7-L'idée de base de L'ACP à noyau. ....	20
Figure 3.1-Localisation basée sur le principe de reconstruction. ....	32
Figure 3.2-Evolution de SPEd en défaut. ....	33
Figure 3.3-Evolution SPE sans défaut (bloc 3). ....	33
Figure 3.4-L'évolution de SPE dans le bloc (1-5) de défaut.....	34
Figure 3.5-L'indice SPE et T2 en utilisant le modèle ACP statique.....	35
Figure 3.6-L'indice SPE et T2 en utilisant le modèle ACP statique.....	36
Figure 3.7-Evolution de l'indice SPE et T2 sans défaut. ....	36
Figure 3.8-L'évolution de la variable x1 sans et avec défaut.....	37
Figure 3.9-Evolution de SPE avec défaut. ....	38
Figure 4.1-Processus de Tennessee Eastman. ....	41
Figure 4.2-Evolution des indices de détection SPE, T2 et Phi sans défaut.....	43
Figure 4.3-Evolution de mesures 'training sans défaut 'Noir' -et testing avec défaut 'Rouge' du processus TE le défaut IDV1.....	44



## Liste des figures

Figure 4.4-Evolution des indices de détection SPE, T2 et Phi pour le défaut IDV1. ....	44
Figure 4.5-Evolution de mesures 'training sans défaut 'Noir', testing avec défaut 'Rouge' du processus TE pour le défaut IDV2. ....	45
Figure 4.6-Evolution des indices de détection SPE, T2 et Phi pour le défaut IDV2. ....	45
Figure 4.7-Evolution de mesures 'training sans défaut 'Noir' et testing avec défaut 'Rouge' du processus TE pour le défaut IDV7. ....	45
Figure 4.8-Evolution des indices de détection SPE, T2 et Phi pour le défaut IDV7. ....	46
Figure 4.9-Evolution de mesures 'training sans défaut 'Noir' et testing avec défaut 'Rouge' du processus TE pour le défaut IDV20. ....	46
Figure 4.10-Evolution des indices de détection SPE, T2 et Phi pour le défaut IDV20. ....	47

# Introduction générale

## Introduction générale

La surveillance des processus est une pratique cruciale dans l'industrie pour garantir la qualité de la production. En effet, la détection précoce des défauts dans les processus peut permettre d'éviter les pertes financières, les déchets de production et les problèmes de sécurité. C'est pourquoi les entreprises investissent souvent dans des systèmes de surveillance des processus pour détecter les anomalies et les défauts en temps réel.

Cependant, la surveillance des processus peut être difficile dans les processus non linéaires. Les processus non linéaires sont caractérisés par des relations complexes entre les variables d'entrée et de sortie, ce qui les rend plus difficiles à surveiller et à contrôler. Les méthodes de surveillance traditionnelles, telles que le contrôle statistique de processus (SPC), peuvent ne pas être suffisamment précises pour détecter les défauts dans les processus non linéaires.[1]

L'analyse en composantes principales (ACP) est une méthode statistique couramment utilisée pour la surveillance des processus. La ACP permet de réduire la dimensionnalité des données et de détecter les défauts dans les processus en analysant les variations des données. Cependant, la ACP traditionnelle peut ne pas être adaptée aux processus non linéaires, car elle ne prend pas en compte les relations non linéaires entre les variables d'entrée et de sortie.[2]

Le KernelACP est une extension de la ACP traditionnelle qui permet de surveiller les processus non linéaires. Cette méthode utilise des fonctions de noyau pour transformer les données en un espace de grande dimension, dans lequel les défauts peuvent être plus facilement détectés. La méthode PCA à noyau peut également être utilisée pour détecter les défauts dans des processus multivariés, où plusieurs variables sont surveillées simultanément[3].

Ce mémoire est réparti en quatre chapitres et la simulation est faite en se basant sur des programmes formulés en MATLAB.

Dans ce premier chapitre, nous avons rappelé quelques généralités sur diagnostic. Dans ce chapitre, nous allons explorer les principes fondamentaux du diagnostic, sa définition et son importance dans différents domaines. Nous aborderons également les différentes étapes du processus de diagnostic, de la collecte des données à l'interprétation des résultats.

Le deuxième chapitre est consacré à l'Analyse en Composantes Principales (ACP). L'ACP est une technique statistique utilisée pour réduire la dimensionnalité des données et

## Introduction générale

identifier les variables les plus significatives. Nous expliquerons les concepts de base de l'ACP à noyau, ses applications et sa mise en œuvre pratique.

Le chapitre 3 traite spécifiquement du diagnostic de défauts. Nous explorerons les différentes méthodes et techniques utilisées pour détecter, localiser et identifier les défauts dans un système ou un processus donné. Nous aborderons des sujets tels que l'analyse de signaux, les techniques de surveillance et les modèles de détection de défauts.

Le dernier chapitre est consacré à l'application pratique des concepts et des méthodes de diagnostic. Dans la première partie de ce chapitre, nous décrirons le processus ou le système spécifique sur lequel nous travaillons, appelé TE. Dans la deuxième partie, nous aborderons la modélisation de la détection de défauts du processus TE, en mettant l'accent sur la simulation hors ligne et en ligne. Nous expliquerons comment ces simulations peuvent être utilisées pour prévoir et détecter les défauts potentiels dans le processus TE, afin de mettre en place des mesures correctives appropriées.

En résumé, le KernalACP est une méthode de surveillance des processus non linéaires qui utilise la PCA traditionnelle en combinant des fonctions de noyau pour transformer les données en un espace de grande dimension. Cela permet de détecter les défauts plus facilement dans les processus industriels et peut être utilisé pour surveiller des processus multivariés

## Chapitre 1 Généralités sur le diagnostic

### 1.1 Introduction

Depuis la Renaissance européenne, la quête d'une production de meilleure qualité a entraîné une utilisation optimale des moyens de production et des installations industrielles pour assurer leur fonctionnement en toute sécurité, en évitant tout vice ou défaut. Pour atteindre cet objectif, des observateurs et des superviseurs internationaux ont mis en place un système de diagnostic visant à surveiller les mécanismes en cas de défaut ou de dysfonctionnement pendant le fonctionnement de la machine. Cette approche utilise des méthodes mathématiques et analytiques, qui font l'objet de recherches scientifiques, afin de garantir la sécurité de l'automatisation industrielle.

### 1.2 Qu'est-cequediagnostic

Le diagnostic est un procédé logique basé sur des étapes précises visant à identifier l'origine des défaillances ou à anticiper le défaut avant qu'il ne se produise.

Généralement, les systèmes industriels contiennent trois éléments de base : les actionneurs, les capteurs et les composants. Le défaut peut se produire dans chacune de ces parties. La détermination du type, de l'amplitude et du moment d'un défaut comprend les trois étapes suivantes : la détection, la localisation et l'identification des défauts[4].

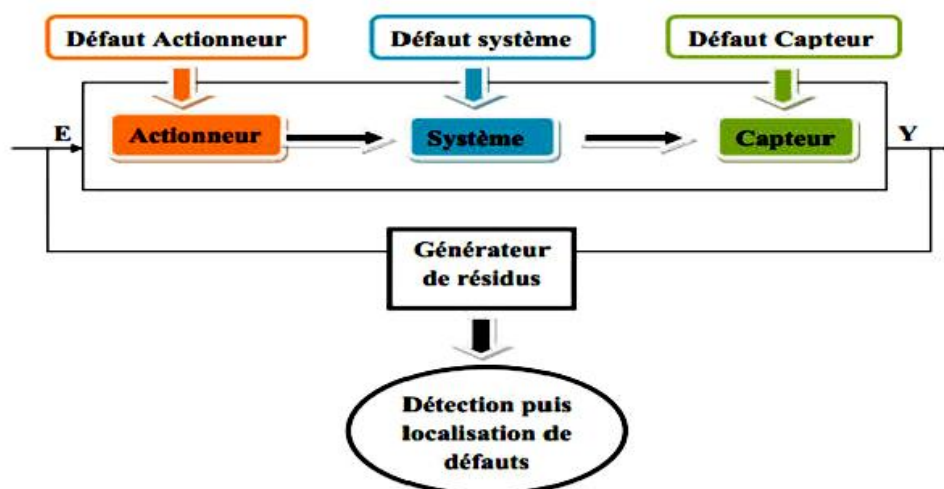


Figure 1.1-Structure générale d'un système de diagnostic.

### 1.3 Terminologie et définitions

Avant d'aller plus loin, nous présentons quelques terminologies et concepts de base relatifs au domaine du diagnostic des défauts. Les définitions les plus importantes sont les suivantes :

**1. Défaillance** : Il s'agit d'une interruption permanente de la capacité d'un système à remplir une fonction requise dans des conditions d'opération spécifiées.

**2. Panne** : Une panne est une erreur, un état de dysfonctionnement ou une interruption permanente du système, généralement causée par l'apparition d'un défaut.

**3. Défaut** : Il est défini comme une déviation d'une variable observée ou d'un paramètre calculé par rapport à sa valeur fixée, ce qui correspond à une déviation du comportement normal du système ou de son unité de commande. Les défauts peuvent survenir dans les capteurs, les actionneurs ou les processus eux-mêmes.

Les défauts peuvent être classés par ordre croissant d'importance du niveau de dégradation.

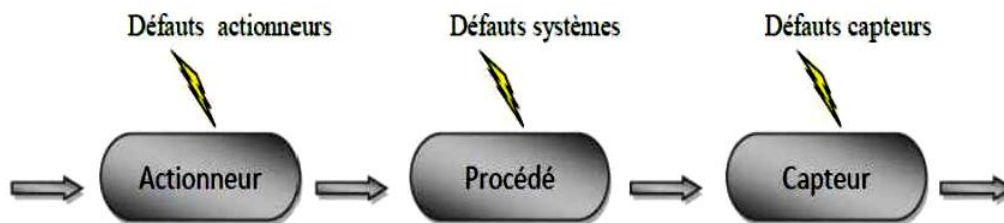


Figure 1.2-Différent type de défaut s'agissant sur un système.

- défautnaissant.
- défautpassager.
- défautpermanent.
- défautcatastrophique.

### Résidu

Un résidu, communément appelé indicateur de défaut ou variable d'écart, est un signal conçu comme un indicateur d'anomalies comportementales ou fonctionnelles. Il exprime l'incohérence entre les informations disponibles et celles théoriques fournies par un modèle censé décrire correctement le processus [5].

### **1.4 La supervision**

La supervision, qui intègre plusieurs activités telles que la surveillance, l'analyse et le pilotage, permet de rapporter les fonctionnements normaux et anormaux des systèmes, afin de fournir à l'opérateur des critères suffisants pour la prise de décisions [6]. La supervision correspond à l'action de surveiller en continu le fonctionnement du système, afin de prendre des décisions adéquates lorsque le système s'écarte de l'objectif désiré. Le système de surveillance observe en continu l'évolution de l'équipement à travers des données quantifiables.

### **1.5 La surveillance**

La supervision correspond à l'action de surveiller le fonctionnement du système afin de prendre des décisions adéquates lorsque le système s'écarte de l'objectif désiré. Le système de surveillance observe en continu l'évolution de l'équipement à travers des données quantifiables.

#### **1.5.1 La détection**

Le premier niveau du diagnostic consiste à prendre une décision binaire : soit le système fonctionne correctement, soit un défaut s'est produit. Le résultat de la procédure de détection est une alarme qui indique que le fonctionnement réel du système ne correspond plus au modèle de fonctionnement sain. L'objectif est de détecter les défauts le plus tôt possible et sans erreur (fausse alarme ou non-détection) [7].

#### **1.5.2 La localisation**

La deuxième étape du processus de diagnostic est déclenchée par une procédure de détection et vise à déterminer de manière plus approfondie les composants défectueux tels que les capteurs, les actionneurs, les processus ou les unités de commande.

#### **1.5.3 L'identification**

L'identification d'un défaut est le fait d'estimer l'amplitude et l'évolution temporelle du défaut afin d'expliquer au mieux le comportement du système. Cette partie d'identification du défaut est la dernière phase de la procédure de diagnostic[4].

### **1.6 Typologie de défauts**

Les défauts se produisent à divers endroits du système, ce qui a conduit à la classification des types de défauts en fonction de leur localisation ou de leur origine [8].

#### **1.6.1 Défautactionneur**

Les actionneurs constituent la partie de l'application qui fournit une représentation concrète du type de commande ou de signal émis par la partie de contrôle. Les dysfonctionnements des actionneurs se manifestent généralement du côté opérationnel du système, lorsqu'un mouvement contraire aux consignes du système se produit (défaut ou panne). Ces erreurs peuvent entraîner des déclenchements de sécurité dans le système[6].

#### **1.6.2 Défautcapteur**

Les défauts au niveau des actionneurs affectent la partie opérative du système en altérant le signal d'entrée. Ils se traduisent par une perte totale ou partielle d'un actionneur. Une perte totale survient lorsque l'actionneur reste coincé dans une position, ce qui l'empêche de fonctionner. Les défauts partiels des actionneurs se manifestent de manière similaire au fonctionnement normal, mais avec une certaine dégradation de leur action sur le système [8].

#### **1.6.3 Défauts composants**

Les défauts des composants sont des défaillances qui affectent directement les composants du système. Ils ne peuvent être classés ni comme des défauts d'actionneurs, ni comme des défauts de capteurs. Ces défaillances sont causées par des altérations de la structure ou des paramètres du modèle, et peuvent entraîner une instabilité du système[6].

### **1.7 Caractérisation de défauts**

Les défauts sont caractérisés par leur comportement dans le temps. En fait, ils sont classés selon leurs évolutions temporelles. Ils peuvent être brusques, intermittents ou graduels, et être de nature additive ou multiplicative [9].

## Chapitre1 Généralités sur le diagnostic

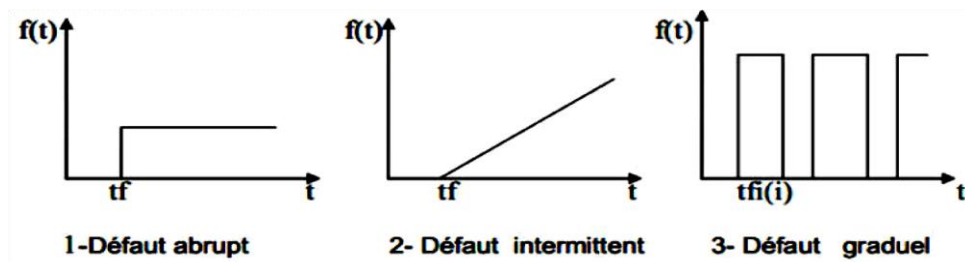


Figure 1.3-Evolution temporelle des différents types de défauts.

### 1.7.1 Défaut graduel

Un défaut graduel est présenté par la figure 1.3. Il conduit à une divergence lente et continue des variables concernées. C'est un défaut difficile à détecter.

### 1.7.2 Défaut intermittent

Il représente une particularité des défauts précédents avec des départs et retours successifs aux valeurs normales, figure 1.3. Il est souvent provoqué par des faux contacts[10].

### 1.7.3 Défaut abrupt

Il est caractérisé par une discontinuité brutale de l'évolution temporelle d'une variable du procédé. Il provoque un écart dans l'évolution de la variable en question par rapport à l'évolution normale.

## 1.8 Classification des méthodes de diagnostic

Dans le milieu industriel, il existe plusieurs méthodes de diagnostic des défauts, qui reposent généralement sur la comparaison entre les données de fonctionnement observées et les connaissances acquises sur le comportement normal et défaillant du système. Les deux principales méthodes utilisées sont les méthodes à base de modèle et les méthodes sans modèle [5].

### 1.8.1 Méthode à base de modèles

Les méthodes de diagnostic à base de modèle consistent à comparer le comportement du système réel avec celui d'un modèle qualitatif et/ou quantitatif établi. Toute différence entre ces comportements est considérée comme une défaillance, comme illustré dans la figure 1.4. Pour appliquer ces méthodes, il est nécessaire de posséder des connaissances



## Chapitre 1 Généralités sur le diagnostic

approfondies sur le procédé à diagnostiquer, sous la forme d'un modèle représentatif qui fournit des grandeurs caractéristiques du procédé. Ces grandeurs sont ensuite constamment comparées aux grandeurs réelles obtenues du procédé en fonctionnement. La nature du modèle utilisé, qu'il soit qualitatif et/ou quantitatif, dépend du type de procédé à diagnostiquer [11].

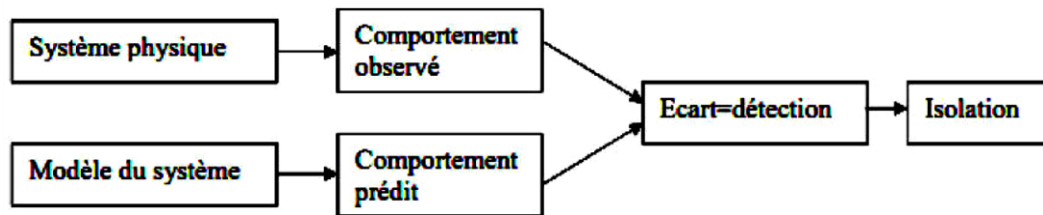


Figure 1.4-Principe de diagnostic avec modèle.

**Modèle qualitatif** : Un modèle de système décrivant le comportement avec des relations entre des variables et des paramètres du système en termes heuristiques tels que des causalités ou des règles [12].

**Modèle quantitatif** : Un modèle de système qui décrit le comportement en utilisant des relations entre les variables et les paramètres du système, exprimées sous forme analytique telle que des équations différentielles ou aux différences. [12]

### 1.8.2 Méthodes sans modèles

Dans certaines applications industrielles, la création d'un modèle mathématique précis peut être difficile, voire impossible, en raison de la complexité des phénomènes impliqués ou des nombreuses reconfigurations qui surviennent dans le processus de production. Dans ces situations, il est nécessaire de recourir à des méthodes de diagnostic qui ne nécessitent pas une connaissance approfondie du procédé.

## 1.9 Principes de diagnostic de défaut

### 1.9.1 Redondance physique ou matérielle

La redondance physique consiste à utiliser plusieurs actionneurs, capteurs, processeurs et logiciels pour mesurer et/ou contrôler une variable particulière. Un principe de vote est appliqué sur les valeurs redondantes pour décider si une faute est présente ou non. Cette approche entraîne un coût important en instrumentation, mais s'avère extrêmement fiable et simple à mettre en œuvre. Elle est principalement utilisée dans des systèmes à hauts risques tels que les centrales nucléaires ou les avions.

## **Chapitre1Généralités sur le diagnostic**

Le diagnostic utilisant la redondance physique se limite à la surveillance des capteurs redondants présents sur une installation. Avec cette technique unique, il ne sera pas possible de détecter les pannes survenant sur des éléments non redondants [11].

### **1.9.2 Redondance analytique**

Son rôle est d'exploiter les contraintes liant les différentes variables du système. Ces dernières peuvent s'exprimer sous la forme de relations analytiques liant les variables connues (relation entrée/sortie ou sortie/sortie). Le diagnostic utilisant la redondance analytique consiste à vérifier si les relations analytiques sont égales à zéro en utilisant les mesures prélevées en ligne sur le système. Le concept de redondance analytique repose sur l'utilisation d'un modèle mathématique du système à surveiller[12].

### **1.10 Conclusion**

Ce chapitre nous a permis de mieux comprendre l'importance du diagnostic, ainsi que l'utilité de la surveillance et du diagnostic dans l'industrie. Nous avons également examiné les principales fonctions de la procédure de surveillance et de diagnostic, ainsi que les différentes techniques de diagnostic disponibles. Dans le prochain chapitre, on va concentrer sur les méthodes de diagnostic basées sur l'analyse de données, en particulier la méthode d'analyse en composantes principales. On va détailler le principe de modélisation des systèmes à l'aide de ces méthodes.

## **Chapitre 2 Analyse en Composantes Principales**

### **2.1 Introduction**

L'Analyse en Composantes Principales (ACP) est une méthode de traitement de données qui permet de réduire la dimensionnalité d'un ensemble de données multidimensionnelles en les projetant sur un espace de variables latentes appelées "composantes principales".

Le but de l'ACP est de décrire les relations complexes entre les variables originales en les résumant en un nombre plus restreint de variables synthétiques. Les composantes principales sont calculées de telle sorte qu'elles représentent la plus grande partie de la variance des données originales[13].

Le Kernel PCA (analyse en composantes principales à noyau) permet d'exploiter les relations potentiellement non linéaires entre les variables. Cette méthode repose sur une transformation non linéaire des données à l'aide de fonctions noyaux, qui sont généralement non linéaires. Cette transformation effectue un changement de base permettant de projeter les données de l'espace d'entrée dans un nouvel espace où les relations entre les variables deviennent linéaires. Cela facilite la modélisation en appliquant l'ACP linéaire, ce qui permet de passer d'un système initialement non linéaire à un autre système linéaire. L'espace de représentation obtenu a une dimension plus élevée que l'espace de départ. L'ACP à noyau repose sur un formalisme mathématique intéressant qui s'appuie sur l'algèbre linéaire, et peut être considérée comme une généralisation naturelle de l'ACP linéaire.

### **2.2 Principe de base de la méthode ACP**

L'Analyse en Composantes Principales (ACP) est une technique descriptive utilisée pour explorer les similitudes entre les individus et les relations entre les variables d'un système étudié. Elle est considérée comme un algorithme classique de réduction de dimension, en général non supervisé et linéaire. L'objectif principal de l'ACP est de remplacer un ensemble de variables inter corrélées par un nombre réduit de variables de l'essentiel de

## Chapitre 2 Analyse en Composantes Principales

l'information. Ces nouvelles variables, appelées les composantes principales, sont des combinaisons linéaires des variables initiales.

L'ACP vise à étudier les structures de liaisons linéaires entre les variables qui définissent le système étudié. Lorsque ces variables sont fortement corrélées, cela signifie que les informations contenues dans celles-ci sont redondantes. L'ACP consiste donc à remplacer cette famille de variables par de nouvelles variables de variance maximale, non corrélées deux à deux, qui sont des combinaisons linéaires des variables d'origine. Ces nouvelles variables définissent des plans factoriels qui servent de base pour représenter graphiquement les variables initiales.

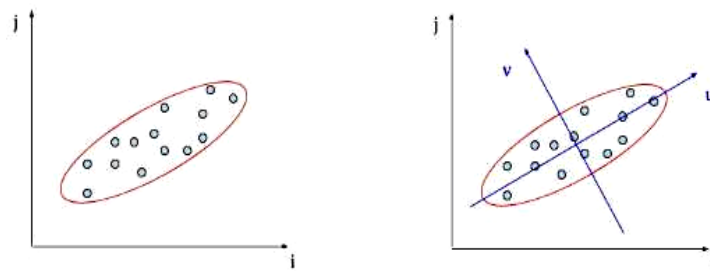


Figure 2.1-Principe de la méthode ACP.

### 2.2.1 Principe

L'objectif de l'ACP est d'identifier les relations linéaires qui existent entre les différentes variables du système.

Pour commencer l'identification du modèle ACP, on construit une matrice qui regroupe l'ensemble des données disponibles, sans faire de distinction entre les entrées et les sorties du système.

On considère la matrice de données  $X^b \in \mathfrak{R}^{N \times m}$  contenant  $n$  échantillons de  $m$  variables duprocessus recueillis dans des conditions de fonctionnement normales (NOC). Afin de procéder à la PCA, la matrice  $X$  doit être normalisée avec les vecteurs de paramètres d'échelle de moyenne et de variance. L'ACP détermine une projection linéaire optimale de la matrice de données  $X$  [6] :

$$T = XP^T \quad (2.1)$$

$$\hat{X} = TP^T \quad (2.2)$$

## Chapitre 2 Analyse en Composantes Principales

Avec  $T \in \mathfrak{R}^{N \times m}$  étant la matrice des composantes principales et la matrice  $P \in \mathfrak{R}^{m \times m}$  contient les vecteurs principaux qui sont les vecteurs propres associés aux valeurs propres de la matrice de covariance  $\Sigma$  de  $X$

$$\Sigma = \frac{1}{N-1} X^T X = P \Lambda P^T \quad (2.3)$$

$$P P^T = P^T P = I_m \quad (2.4)$$

$$\Sigma = [p_1 p_{m-l}] \begin{bmatrix} \Lambda_m & 0 \\ 0 & \Lambda_{m-l} \end{bmatrix} \begin{bmatrix} P_m^T \\ P_{m-l}^T \end{bmatrix} \quad (2.5)$$

où  $\Lambda$  est une matrice diagonale contenant les valeurs propres dans sa diagonale, classées par ordre décroissant.

Une étape essentielle de la modélisation ACP consiste à déterminer le nombre de composants ( $l < m$ ) à conserver, en utilisant différents critères possibles.

La matrice de transformation  $P \in \mathfrak{R}^{m \times l}$  est obtenue en sélectionnant  $l$  vecteurs propres correspondant aux  $l$  plus grandes valeurs propres. La matrice  $P_l$  transforme l'espace des variables mesurées en un espace de dimensions réduites  $\hat{X}$ .

$$\hat{X} = X P_l P_l^T = X C_l \quad (2.6)$$

L'espace résiduel, également noté  $E$ , est recouvert par la matrice générée en choisissant les derniers  $(m-l)$  vecteurs propres ou colonnes de  $P$ .

$$E = X - \hat{X} + X(l - C_l) \quad (2.7)$$

A chaque instant  $k$ , le vecteur de mesure est donné par :

$$X(k) = [x_1(k) x_2(k) \dots x_m(k)]^T \quad (2.8)$$

où son estimation est donnée par :

$$\hat{X}(k) = C_l X(k) \quad (2.9)$$

et l'erreur d'estimation peut être exprimée comme suit :

$$e(k) = X(k) - \hat{X}(k) \quad (2.10)$$

Dans le cadre du diagnostic, le modèle PCA obtenu décrit le comportement normal du processus en utilisant des projections dans l'espace principal et résiduel. Ensuite, les événements inhabituels sont détectés en comparant le comportement observé avec ce modèle à l'aide de diverses techniques de contrôle. Les défauts sont plus clairement identifiés par les

## Chapitre 2 Analyse en Composantes Principales

résidus, qui devraient être nuls en l'absence de défauts et différents de zéro en présence de défauts.

### 2.2.2 Identification du modèle ACP

En pratique, lors de la modélisation d'un processus à l'aide de l'analyse en composantes principales, les mesures des variables du processus représentant tous les modes de fonctionnement normal sont recueillies dans une matrice  $X^b$ . Cette matrice  $X^b$ , de dimension  $X^b \in \mathfrak{R}^{N \times m}$ , est construite en prenant en compte  $m$  variables et  $N$  observations pour chaque variable.

$X^b \in \mathfrak{R}^{N \times m}$  est donnée par :

$$X^b = \begin{pmatrix} x_1(1) & x_2(1) & \dots & x_m(1) \\ x_1(2) & x_2(2) & \dots & x_m(2) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(N) & x_2(N) & \dots & x_m(N) \end{pmatrix} \quad (2.11)$$

où  $x_1(1)$  représente la mesure de la première variable au premier instant de mesure. Au préalable, afin de rendre les résultats indépendants des unités utilisées pour chaque variable, un prétraitement indispensable consiste à centrer et réduire les variables. Chaque colonne  $X_j$  de la nouvelle matrice centrée est donnée par :

$$X_j = \frac{x_j^b - M_j}{\sigma_j} \quad (2.12)$$

où  $X_j^b$  est la  $j^{\text{ième}}$  colonne de la matrice  $X^b$  et  $M_j$  est sa moyenne donnée par :

$$M_j = \frac{1}{N} \sum_{k=1}^N x_j(k) \quad (2.13)$$

et  $\sigma_j^2$  est sa variance qui sera estimée en utilisant l'équation :

$$\sigma_j^2 = \frac{1}{N} \sum_{k=1}^N (x_j(k) - M_j)^2 \quad (2.14)$$

La nouvelle matrice des données normalisées est notée :

$$X = [X_1 \ \dots \ X_m] \quad (2.15)$$

## Chapitre 2 Analyse en Composantes Principales

La matrice de corrélation est donnée par:

$$\Sigma = \frac{1}{N-1} X^T X \quad (2.16)$$

En général, pour identifier un modèle, il est nécessaire de sélectionner une classe de modèle, de définir une structure fixe, puis d'estimer les paramètres du modèle avant de le valider. En ce qui concerne l'analyse en composantes principales (ACP), l'estimation des paramètres est relativement simple car elle consiste essentiellement à calculer les valeurs et les vecteurs propres. Toutefois, le choix de la structure est plus complexe, comme nous le verrons plus tard.

### 2.2.3 Pourcentage cumulé de la variance totale (PCV)

Cette méthode de sélection d'un nombre de composantes se base sur l'hypothèse que chaque composante principale représente en fait une partie de la variance totale des mesures du processus en question. Les mesures de variance représentent les valeurs propres de la matrice de corrélation  $\Sigma$ , et elles sont utilisées dans la sélection d'un nombre de composantes principales. Généralement, on essaie de retenir un nombre de composantes principales ayant un pourcentage cumulé de variance maximale compris entre 90 % et 95 % de la variance totale des données [14].

$$PCV = 100 \left[ \sum_{i=1}^m \frac{\lambda_i}{\sum_{i=1}^m \lambda_i} \right] \quad (2.17)$$

### 2.2.4 Critère de validation croisée (PRESS)

L'idée fondamentale de cette méthode consiste à estimer les mesures  $x_i(\mathbf{k})$  d'un jeu de données de validation à l'aide d'un modèle calculé à partir d'un jeu de données d'identification. On évalue ensuite la comparaison entre les valeurs mesurées et leurs estimations obtenues à partir d'un jeu d'identification différent. Le nombre optimal de composantes principales retenues est déterminé lorsque la moyenne de l'estimation globale des mesures  $x_i(k)$  cesse de s'améliorer de manière significative avec l'ajout d'une composante supplémentaire. La procédure de validation croisée repose sur un critère appelé PRESS, qui vise à minimiser la somme des carrés des erreurs d'estimation [15].

$$PRESS(l) = \frac{1}{Nm} \sum_{k=1}^N \sum_{i=1}^m (\hat{x}_i^{(l)}(k) - x_i(k))^2 \quad (2.18)$$

## Chapitre 2 Analyse en Composantes Principales

En utilisant l'acronyme ACP pour l'Analyse en Composantes Principales, nous pouvons estimer la valeur de  $x_i$  en utilisant un modèle ACP composé de  $l$  composantes principales (PCs), représentée par  $\hat{x}_i^{(l)}$ .

L'algorithme simplifié pour déterminer le nombre de composantes principales (ACPs), noté  $l$ , est comme suit [16] :

Divisez les données en un jeu d'identification et un jeu de validation.

Effectuez une ACP avec  $l$  composantes sur le jeu d'identification, où  $l$  varie de 1 à  $m$ .

Calculez les critères correspondants sur le jeu de validation, PRESS(1) à PRESS( $m$ ).

La  $l^{\text{ème}}$  composante principale, pour laquelle le minimum de PRESS est obtenu, détermine le nombre de composantes principales à retenir.

### 2.3 Exemple de simulation

Les différentes étapes de l'élaboration d'un modèle ACP sont appliquées sur un exemple afin de mieux comprendre les étapes précitées. Nous disposons de 05 variables représentant un système statique et qui sont décrites par les équations suivantes :

$$\begin{cases} x_1 = \sin(\omega * t) + \varepsilon \\ x_2 = \cos(\omega * t) + \varepsilon \\ x_3 = x_1 + x_2 + \varepsilon \\ x_4 = x_1 + x_3 + \varepsilon \\ x_5 = x_2 + x_3 + \varepsilon \end{cases} \quad (2.19)$$

Où, les  $\varepsilon_i$  sont des bruits blancs,  $x_1$  et  $x_2$  sont les signaux en forme sinusoïdale avec une taille de 1000 mesures. La visualisation de la variation de ces données est représentée dans la Figure (2.2).



## Chapitre 2 Analyse en Composantes Principales

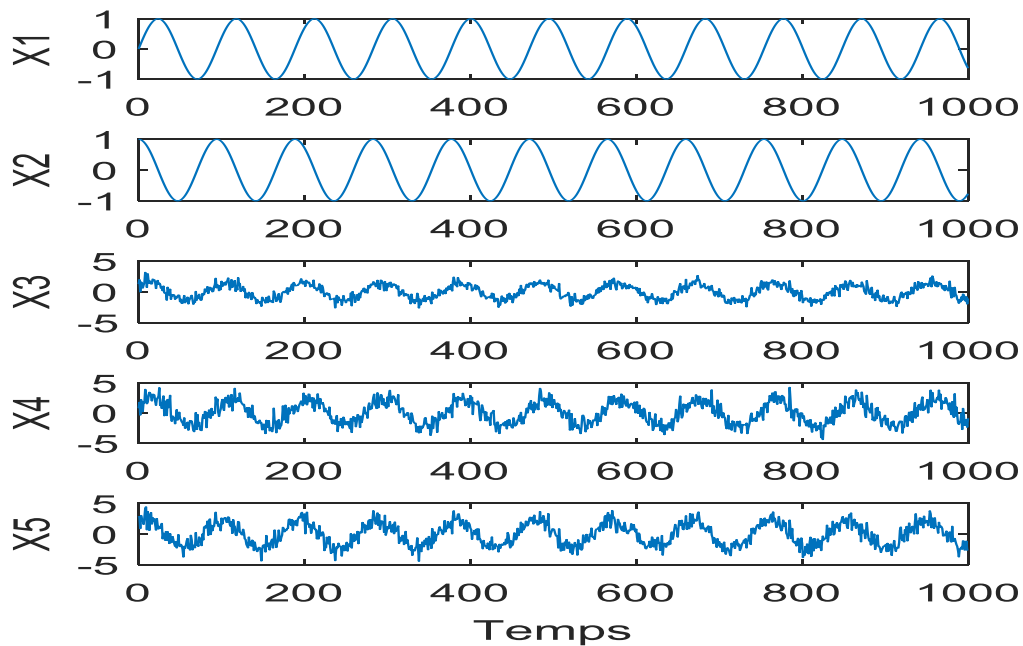


Figure 2.2-Evolution des différentes variables.

Lors de la normalisation le vecteur des moyennes ainsi que le vecteur des variances des variables sont donnés par :

$$m_x = [ 0.0011 \quad -0.0041 \quad -0.0190 \quad 0.0085 \quad -0.0011 ]$$

Ensuite, la matrice de covariance est

$$M_c = \begin{pmatrix} 1.0000 & 0.0021 & 0.6306 & 0.7754 & 0.3983 \\ 0.0021 & 1.0000 & 0.6398 & 0.4193 & 0.8032 \\ 0.6306 & 0.6398 & 1.0000 & 0.8765 & 0.8956 \\ 0.7754 & 0.4193 & 0.8765 & 1.0000 & 0.7318 \\ 0.3983 & 0.8032 & 0.8956 & 0.7318 & 1.0000 \end{pmatrix}$$

Après diagonalisation, les valeurs propres  $V$  et les vecteurs propres correspondant  $P$  sont données comme suit :

$$V = \begin{pmatrix} 0.05670000 \\ 0.10880000 \\ 0.12320000 \\ 0.16010000 \\ 0.35513000 \end{pmatrix}$$

$$P = \begin{pmatrix} -0.0679 & 0.1089 & 0.6508 & -0.6603 & 0.3522 \\ 0.0858 & -0.4106 & 0.5304 & 0.6404 & 0.3641 \\ 0.7663 & 0.2854 & -0.2525 & -0.0046 & 0.5172 \\ -0.2400 & -0.6467 & -0.4604 & -0.2770 & 0.4852 \\ -0.5858 & 0.5656 & -0.1393 & 0.2777 & 0.4904 \end{pmatrix}$$

## Chapitre 2 Analyse en Composantes Principales

Les composantes principales  $T$  en faisant la projection de la base de données matricielle  $X$  normalisée sur la matrice des vecteurs propres, sont visualisées dans Figure (2.3)

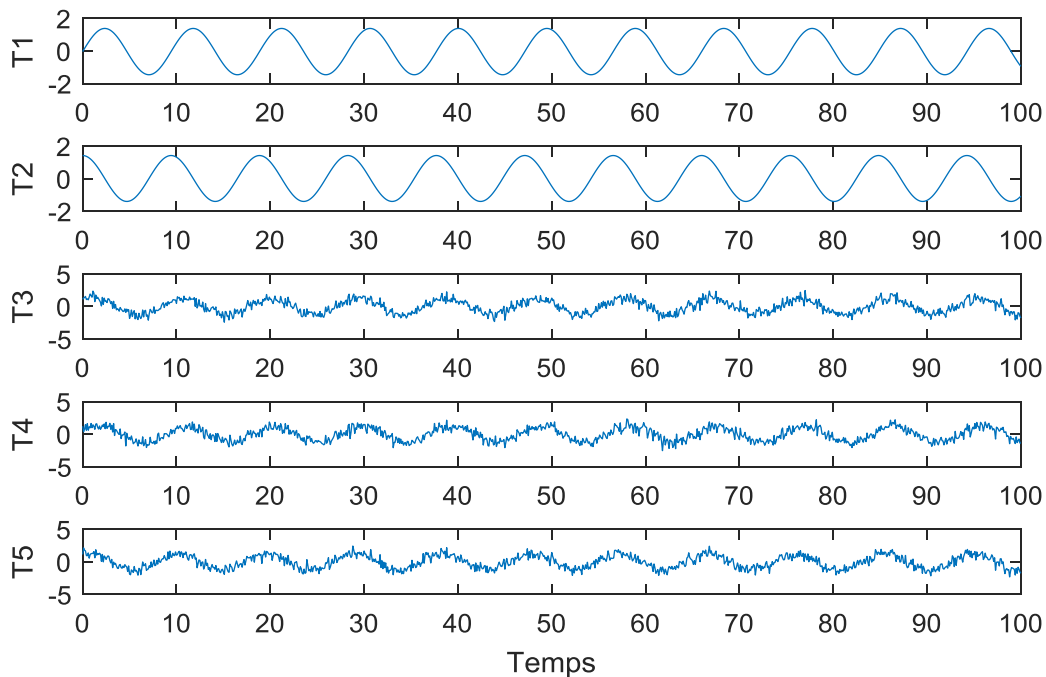


Figure 2.3-Évolution des composantes principales.

Notez que les deux premières composantes représentent de l'information et que les autres dernières composantes représentent les bruits de mesures. Afin de déterminer le nombre de composants nécessaire dans le modèle ACP nous avons utilisé le critère PCV. L'information la plus pertinente qui représente les composantes principales ( $x_1, x_2$ ) est bien illustré sur les deux premières composantes ( $l=2$ ). En effet, la variation des données et leurs estimations en utilisant deux composantes principales dans le modèle est bien illustrée sur Figure (2.4).

On peut remarquer sur la figure que les estimations sont conformes aux données initiales ainsi que les erreurs d'estimation de toutes les variables sont presque nulles. Ceci montre que le modèle ACP donne une estimation assez correcte des différentes mesures en utilisant deux composantes principales.

## Chapitre 2 Analyse en Composantes Principales

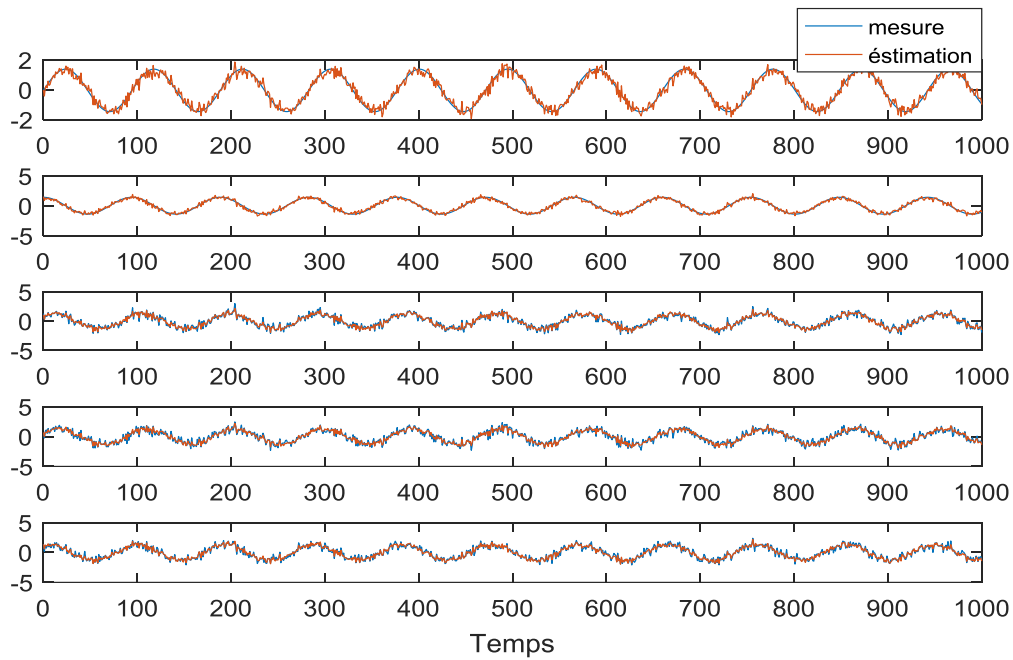


Figure 2.4-Evolution des mesures et leurs estimations.

Afin de déterminer le nombre de composants nécessaire dans le modèle ACP nous avons utilisé le critère PCV et PRESS.

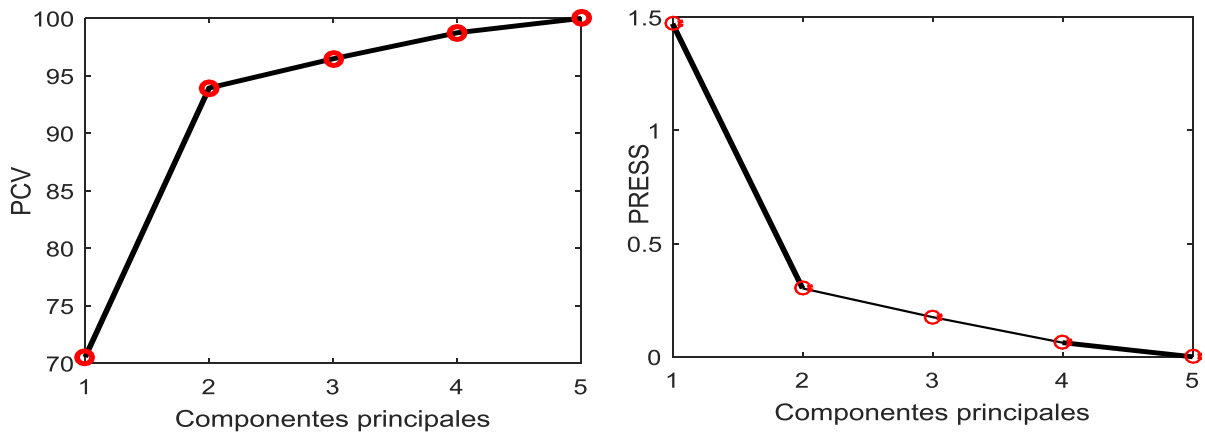


Figure 2.5-Évolution de la norme PCV et le Critère PRESS.

La valeur de  $x_i$  est estimée à l'aide d'un modèle ACP composé d'un nombre spécifié de composantes principales (ACPs). Grâce à l'algorithme simplifié mentionné précédemment, il est possible de déterminer le nombre approprié de composantes principales. Cette estimation permet d'utiliser efficacement les outils de l'ACP pour mieux comprendre les relations et les variations entre les variables, et ainsi obtenir un modèle plus précis et performant. La mise en œuvre de l'algorithme implique la division des données en un jeu d'identification et un jeu de

## Chapitre 2 Analyse en Composantes Principales

validation, suivi de la réalisation de l'ACP avec différentes configurations de composantes principales sur le jeu d'identification. Les mesures PRESS sont calculées pour évaluer les performances du modèle sur le jeu de validation. Enfin, le nombre optimal de composantes principales est choisi en identifiant celle qui présente la plus faible valeur de la mesure PRESS, assurant ainsi un équilibre entre la complexité et les performances prédictives du modèle.

### 2.4 ACP non linéaire (NLPCA)

L'ACP linéaire est un outil de modélisation des relations linéaires entre un ensemble de variables qui représentent le comportement d'un processus étudié. Son objectif est de réduire la dimension des variables en exploitant les corrélations potentielles entre les variables initiales. Cependant, la plupart des processus réels sont dynamiques et non linéaires, ce qui rend l'application de l'ACP classique incapable de trouver une représentation compacte de ces données. Il devient difficile d'exploiter la corrélation non linéaire entre les variables pour réduire leur dimension. Pour surmonter cette difficulté, plusieurs méthodes d'analyse en composantes principales non linéaires (ACPNL) ont été proposées afin de résoudre ce problème.

#### 2.4.1 Principe de l'ACP non linéaire

L'ACP non linéaire est une extension de l'ACP linéaire qui vise à identifier à la fois les relations linéaires et non linéaires entre les variables d'un processus. Contrairement à l'ACP linéaire, qui se concentre uniquement sur les relations linéaires, l'ACP non linéaire permet une généralisation plus complète. Pour ce faire, les données du processus sont projetées sur des courbes ou des surfaces plutôt que sur des droites ou des plans [17].

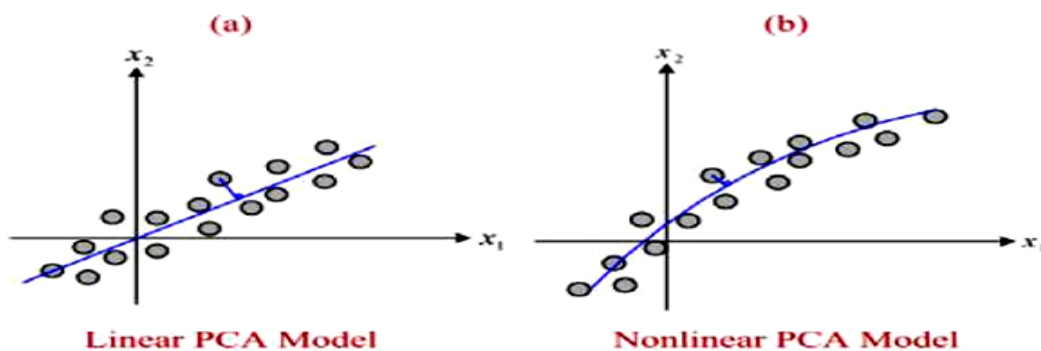


Figure 2.6-La méthode ACP.

## Chapitre 2 Analyse en Composantes Principales

Afin de faciliter la compréhension du problème et d'établir une connexion avec le modèle linéaire, nous allons examiner l'approche de l'analyse en composantes principales non linéaire utilisant les méthodes à noyau.

### 2.4.2 ACP non linéaire à noyau (kernel PCA)

Récemment, l'attention portée à l'Analyse en Composantes Principales (ACP) pour résoudre des problèmes d'apprentissage a été ravivée grâce à l'introduction d'une version non linéaire de cette méthode. La technique de l'ACP à noyau offre la possibilité d'exploiter des relations potentiellement non linéaires entre les variables. Cette approche implique de projeter les observations dans un espace de dimensions supérieures, puis d'appliquer une ACP classique sur les observations transformées dans cet espace. Cette méthode repose sur une transformation non linéaire des données à l'aide de fonctions noyaux généralement non linéaires.

L'idée de la version non linéaire de l'ACP consiste à envoyer préalablement les données  $X_i$  par une application  $f : X \rightarrow H$  (appelée Feature map) dans un espace linéaire de grande dimension  $H$  muni d'un produit scalaire [6].

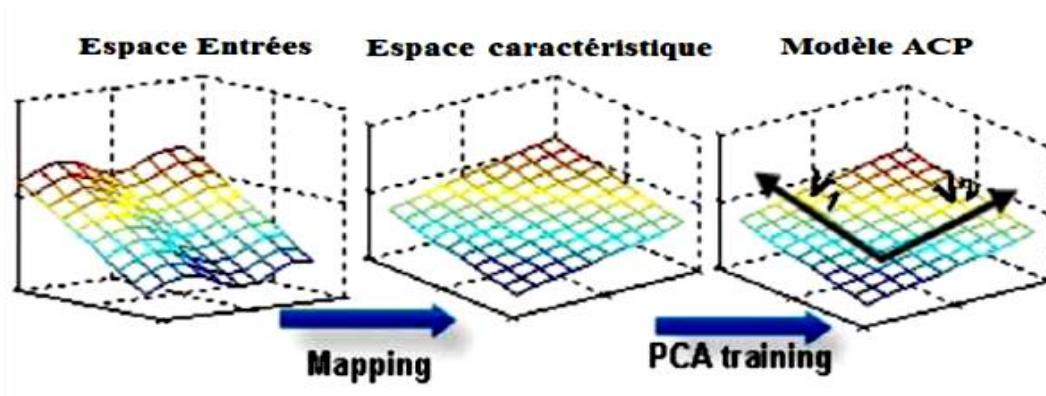


Figure 2.7-L'idée de base de l'ACP à noyau.

Soit  $k(X_i, X_j), (i, j) \in \{1, \dots, n\}$  un noyau de Mercer. Ce noyau peut être utilisé pour calculer une matrice  $K$  définie positive. Il permet de reformuler une projection en terme de produit scalaire :

$$k(X_i, X_j) = \langle \phi(X_i), \phi(X_j) \rangle \quad (2.20)$$

Dans les méthodes utilisant des noyaux, ces derniers peuvent être considérés comme une mesure de similarité non linéaire. Supposons un ensemble non-vide  $X$  et un noyau positif

## Chapitre 2 Analyse en Composantes Principales

$K$ , qui représente une mesure des fonctionnalités non linéaire du système  $x_1, x_2, \dots, x_n \in X$ , définissant un sous-espace de  $R^m$ . Les données  $X$  sont ensuite transformées dans un espace fonctionnel appelé espace de caractéristique (featurespace), où l'ACP classique est appliqué. Soit  $\Phi$  la transformation non-linéaire de l'espace des observations  $X$  à l'espace fonctionnel  $H$  qui, à chaque  $x_i$  lui fait correspondre son image  $\phi(x_i)$ .

L'ACP à noyau calcule les composantes principales non linéaires des données transformées  $\phi(x_1), \phi(x_2), \dots, \phi(x_n)$ .

L'algorithme de l'ACP est reformulé donc en termes de produit scalaire des données dans l'espace des caractéristiques. Les données d'apprentissage sont alors disposées comme suit

$X = [\phi_1, \phi_2, \dots, \phi_n]^T$ . La matrice de covariance dans l'espace de caractéristique  $H$  est donc :

$$XX^T = \sum_{i=1}^n \phi_i \phi_i^T \quad (2.21)$$

On définit alors  $K$ , comme suit :

$$k = XX^T = \begin{pmatrix} \phi_1^T \phi_1 & \dots & \phi_1^T \phi_n \\ \vdots & \ddots & \vdots \\ \phi_n^T \phi_1 & \dots & \phi_n^T \phi_n \end{pmatrix} = \begin{pmatrix} k(x_1, x_1) & \dots & k(x_1, x_n) \\ \vdots & \ddots & \vdots \\ k(x_n, x_1) & \dots & k(x_n, x_n) \end{pmatrix} \quad (2.22)$$

Il y a plusieurs fonctions noyaux. Parmi eux : le noyau gaussien, le noyau polynomial, et le noyau sigmoïde.

Le noyau polynomial :

$$k(x_i, x_j) = (\gamma \langle x_i, x_j \rangle + a)^d \quad (2.23)$$

Le noyau sigmoïde

$$k(x_i, x_j) = \tanh(\gamma \langle x_i, x_j \rangle + b) \quad (2.24)$$

Le noyau gaussien :

$$k(x_i, x_j) = \exp \left[ \frac{\left( (x_i x_j)^T (x_i x_j) \right)}{c} \right] \quad (2.25)$$

### 2.4.3 Algorithme de l'ACP à noyau

1. La matrice  $X$  qui englobe les différentes variables du système dans le bon fonctionnement.

## Chapitre 2 Analyse en Composantes Principales

2. Calcul de la matrice K :  $k(X_i, X_j) = \langle \phi(X_i), \phi(X_j) \rangle$ .
3. Centrage de la matrice de Gram (on retire la moyenne des données projetées dans l'espace des caractéristiques) :  $G = k - \mathbf{1}_n k - k \mathbf{1}_n - \mathbf{1}_n k \mathbf{1}_n$
4. Diagonalisation de la matrice de Gram 'G' ("valeurs/vecteurs" propres).
5. Calcul des projections sur les composantes principales  $B_k = \sum_{i=1}^n \alpha_i^k k(x_i x_j)$ .
6. Calculer l'erreur quadratique entre les images de données dans H et leurs estimations

### 2.5 Conclusion

Dans ce chapitre, nous avons présenté l'analyse en composantes principales (ACP) comme une méthode très intéressante pour mettre en évidence les corrélations linéaires entre les variables d'un processus. L'ACP permet de sélectionner les directions de l'espace qui contiennent le maximum d'informations en maximisant la variance des projections.

Cependant, la plupart des processus industriels sont non linéaires et dynamiques, ce qui rend l'ACP traditionnelle inadaptée pour ce type de systèmes. C'est pourquoi nous avons opté pour l'utilisation d'un noyau fonctionnel, qui permet de générer des données non linéaires dans un autre espace linéaire.

## **Chapitre 3 Diagnostic de défauts**

### **3.1 Introduction**

La procédure de diagnostic à base de modèle est une méthode utilisée pour détecter les défaillances dans un système en comparant les observations réelles avec les prévisions fournies par un modèle. Cette méthode se compose de deux phases: la phase de détection et la phase de localisation.

La phase de détection est la première étape dans la procédure de diagnostic de défauts. Cette étape implique la génération des résidus, qui sont des écarts entre les observations réelles et les prévisions fournies par le modèle. En temps normal, les résidus devraient être idéalement nuls, mais en pratique, ils sont différents de zéro en raison de diverses perturbations.

L'indice de détection de défauts le plus couramment utilisé dans les techniques d'ACP (Analyse en Composantes Principales) est l'indice SPE (Squared Prediction Error). Cet indice mesure la déviation des résidus par rapport à leur moyenne et assure la détection des défaillances dans l'espace résiduel en utilisant toutes les variables à surveiller.

La phase de localisation est la deuxième étape de la procédure de diagnostic de défauts et consiste à identifier l'emplacement exact de la défaillance dans le système. Cette étape est généralement réalisée en utilisant des techniques de classification, telles que les arbres de décision, les réseaux de neurones, ou les méthodes d'apprentissage automatique[18].

### **3.2 Détection de défauts**

L'indice SPE (pour Squared Prediction Error) est un indicateur de détection couramment utilisé dans l'analyse en composantes principales (ACP) pour la détection des anomalies ou des défaillances dans les données.

L'indice SPE mesure la distance entre chaque observation et la projection de celle-ci sur un sous-espace défini par les composantes principales. Plus précisément, l'indice SPE est



la somme des carrés des distances entre chaque observation et sa projection sur le sous-espace des composantes principales.

Lorsque l'indice SPE est élevé pour une observation donnée, cela peut indiquer qu'elle est éloignée de la distribution normale des données et peut être considérée comme une anomalie ou une défaillance. Cette observation peut nécessiter une enquête plus approfondie pour déterminer la cause de cette défaillance ou pour identifier toute autre source de variabilité dans les données.

Il convient de noter que d'autres indices de détection peuvent également être utilisés en conjonction avec l'indice SPE pour améliorer la détection des défaillances dans les données[19].

### **3.2.1 Indice SPE (cas linéaire)**

La statistique SPE (Squared Prediction Error), également appelée Q, est une mesure couramment utilisée pour détecter les conditions anormales dans un modèle statistique. Elle est calculée à l'aide de l'équation [16], qui n'a pas été fournie dans votre demande.

Le calcul de la statistique SPE se fait généralement en utilisant un modèle de régression ou un modèle prédictif. Voici une explication générale de la façon dont la statistique SPE est calculée :

1. Tout d'abord, un modèle statistique est ajusté aux données normales ou de référence.
2. Ensuite, les données sont utilisées pour prédire les valeurs correspondantes à l'aide du modèle ajusté. Les résidus, c'est-à-dire les différences entre les valeurs prédites et les valeurs réelles, sont calculés.
3. Les résidus sont ensuite utilisés pour calculer la statistique SPE. Cette mesure est généralement obtenue en sommant les carrés des résidus.

La statistique SPE est souvent utilisée dans les applications de détection de l'anomalie et de surveillance de processus. Elle permet de quantifier la différence entre les données observées et les valeurs prédites par le modèle. Lorsque la statistique SPE dépasse un seuil prédéfini, cela peut indiquer la présence d'une condition anormale ou d'un comportement inhabituel.

## Chapitre 3 Diagnostic de défauts

Il est important de noter que l'équation [19], nécessaire pour le calcul précis de la statistique SPE, n'a pas été fournie dans votre demande. Sans cette équation spécifique, il n'est pas possible de donner une explication plus détaillée du calcul de la statistique SPE dans votre cas particulier.

Appelée aussi  $Q$  (Squared Prediction Error) qui est donnée par l'équation [17]

$$SPE = e e^T = \sum_{i=1}^n (x - \hat{x})^2 \quad (3.1)$$

avec,  $e$  représente le vecteur résiduel de la reconstruction de données, et  $P$  la matrice des vecteurs propres.

### 3.2.2 Indice SPE (cas non linéaire)

L'ACP à noyau a apporté beaucoup en tant qu'outil de surveillance des systèmes non linéaires. En effet, l'idée de base consiste à projeter les données de l'espace d'entrée, où leurs relations sont non linéaires, sur un autre espace où elles sont distribuées de manière linéaire. Cet espace est appelé espace caractéristique. Ainsi, les techniques de détection de défauts utilisées avec l'ACP linéaire peuvent être appliquées dans cet nouvel espace. L'indice de SPE est défini comme la norme du vecteur résiduel dans l'espace caractéristique et est formulé comme suit :

$$SPE = \|\phi(x) - \hat{\phi}(x)\|^2 \quad (3.2)$$

Généralement, le processus est considéré en fonctionnement normal si :

$$SPE < \delta_\alpha^2 \quad (3.3)$$

Avec  $\delta_\alpha^2$  représente le seuil de détection du SPE

### 3.2.3 Indice $T^2$ (cas linéaire)

La statistique  $T^2$  (aussi appelée T au carré) est couramment utilisée dans l'analyse en composantes principales (ACP) pour détecter les conditions anormales. Elle est appliquée aux premières composantes principales obtenues lors de l'analyse en composantes principales.

L'ACP est une technique statistique utilisée pour réduire la dimensionnalité des données en les projetant dans un nouvel espace, appelé espace des composantes principales. Les

### Chapitre 3 Diagnostic de défauts

premières composantes principales capturent généralement la plus grande part de la variance des données.

Pour calculer la statistique  $T^2$ , voici une explication générale :

1. Les données normales ou de référence sont utilisées pour ajuster un modèle d'ACP.
2. Les données observées sont ensuite projetées dans l'espace des composantes principales à l'aide du modèle d'ACP ajusté. Cela permet d'obtenir les valeurs des premières composantes principales pour chaque observation.
3. La statistique  $T^2$  est ensuite calculée en utilisant les valeurs des premières composantes principales. Elle est généralement basée sur la distance multivariée entre chaque observation et la moyenne des observations de référence dans l'espace des composantes principales.

Lorsque la statistique  $T^2$  dépasse un seuil prédéfini, cela peut indiquer la présence d'une condition anormale ou d'un comportement inhabituel. Une valeur élevée de la statistique  $T^2$  suggère une plus grande différence par rapport à la normale dans l'espace des composantes principales.

Cependant, vous avez mentionné l'équation [21], qui n'a pas été fournie dans votre demande. Sans cette équation spécifique, il n'est pas possible de donner une explication plus détaillée sur le calcul de la statistique  $T^2$  dans votre cas particulier.

$$T^2(k) = \hat{t}^T(k) \Lambda_\ell^{-1} \hat{t}(k) \quad (3.4)$$

où  $\Lambda_\ell = \text{diag}\{\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_\ell\}$  est une matrice diagonale contenant les plus grandes valeurs propres de la matrice de covariance.

Le processus est en fonctionnement normal, à un instant donné, si :

$$T^2(k) \leq \tau^2 = \chi^2_{\alpha}(\ell) \quad (3.5)$$

Lorsqu'on utilise la statistique  $T^2$  dans l'analyse en composantes principales (ACP) pour détecter les défauts, il est possible d'établir un intervalle de confiance pour déterminer les seuils de détection. L'intervalle de confiance est généralement exprimé avec un niveau de confiance de  $(1 - \alpha) \times 100\%$ , où  $\alpha$  représente le niveau de signification.

La statistique  $T^2$  suit une distribution de Fisher avec les degrés de liberté  $F_\ell$ ,  $(N-\ell)$ , où  $\ell$  représente le nombre de composantes principales utilisées et  $N$  est la taille de l'échantillon.

## Chapitre 3 Diagnostic de défauts

Pour obtenir la limite supérieure pour un seuil de confiance  $\alpha$ , on utilise la fonction de distribution de Fisher, notée  $F_{\ell, (N-\ell), \alpha}$ . Cette limite supérieure, notée  $X_{\ell, \alpha}^2$ , est calculée à l'aide de l'équation suivante :

$$X_{\ell, \alpha}^2 = \frac{\ell(N+1)(N-1)}{N(N-\ell)} F_{\ell, (N-\ell), \alpha} \quad (3.6)$$

Une fois les défauts détectés à l'aide de la statistique  $T^2$ , il est nécessaire d'identifier les variables responsables de ces défauts. Cette opération est appelée localisation de défauts. Il existe de multiples techniques capables de déterminer les variables en défaut, notamment celles mentionnées dans la référence [6]. Cependant, la référence [6] n'est pas spécifiée dans votre demande, donc je ne peux pas fournir des informations précises sur ces techniques en particulier. Les techniques de localisation de défauts peuvent varier en fonction du contexte spécifique de l'analyse et des données disponibles.

### 3.2.4 Indice $T^2$ (cas non linéaire)

L'indice  $T^2$  de Hotelling est calculé dans l'espace de fonction comme  $T^2 = t^T \Lambda^{-1} t$ , où la matrice  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_l)$  contient les variances des composantes principales  $t_i$  dans l'espace des caractéristiques. A partir de l'équation des composantes principales dans l'espace des caractéristiques, l'indice de détection  $T^2$  est calculé en utilisant la fonction noyau comme suit [6]:

$$T^2(x) = k(x)^T P \Lambda^{-2} P^T k(x) = k(x)^T D k(x) \quad (3.7)$$

avec  $D = P \Lambda^{-2} P^T$ .

Le seuil de l'indice de détection de défauts non linéaires  $T^2$  est donné ainsi,

$$\tau^2 = g^{T^2} X_{\alpha}^2(h^{T^2}) \quad (3.8)$$

Avec, un intervalle de confiance  $(1 - \alpha) \times 100\%$ ,  $g^{T^2} = \frac{1}{n-1}$ , et  $h^{T^2} = l$ .

### 3.2.5 Phase de modélisation

1. Acquisition la base de données qui représente le fonctionnement du système.
2. Normalisation : centrage et réduction.
3. Déterminer sa matrice de covariance  $\Sigma$ .
4. Déterminer les valeurs propres et les vecteurs propres la matrice de covariance

5. Déterminer les composants principaux à partir le modèle.
6. Déterminer les résidus  $e_i(k)$ .

### 3.2.6 Phase de test

7. Acquérir une nouvelle base de données de test.
8. Calculer l'indice SPE.
9. Si  $SPE > seuil$ , le système est défaillant ; sinon le système fonctionne proprement. [22]

## 3.3 Localisation de défauts

La localisation de défauts fait référence à l'identification de la zone où se produit un dysfonctionnement ou une défaillance dans un système ou un processus. La localisation de défauts peut être utilisée dans une variété de domaines, tels que l'électronique, l'informatique, les réseaux, l'ingénierie mécanique, etc.

En général, il existe deux approches principales pour la localisation de défauts :

1. L'approche basée sur les tests : Cette approche consiste à effectuer des tests sur le système ou le processus pour identifier la source du dysfonctionnement. Par exemple, dans le cas d'un circuit électronique, des tests peuvent être effectués pour mesurer la tension, la fréquence, la résistance, etc. dans différentes parties du circuit afin d'identifier la zone où se trouve le défaut.
2. L'approche basée sur les données : Cette approche utilise des données existantes pour identifier la zone où se produit le dysfonctionnement. Par exemple, dans le cas d'un réseau informatique, des données telles que les journaux de serveur, les enregistrements de trafic, etc. peuvent être analysées pour identifier la zone où se produit le dysfonctionnement.

Dans les deux approches, l'utilisation de techniques avancées telles que l'apprentissage automatique et l'analyse de données peut aider à accélérer le processus de localisation de défauts en fournissant des informations précises et rapides sur la zone où se produit le dysfonctionnement.

### 3.3.1 La reconstruction des variables

La méthode de reconstruction des variables défectueuses en utilisant l'Analyse en Composantes Principales (ACP) comprend les étapes suivantes :

### Chapitre 3 Diagnostic de défauts

1. Collecte des données : Rassemblez les données pour les variables mesurées et les variables défectueuses.
2. Exclusion de la variable défectueuse : Identifiez la variable défectueuse et excluez-la de l'ensemble de données.
3. Prétraitement des données : Effectuez des étapes de prétraitement comme la normalisation ou la standardisation des variables mesurées.
4. Calcul des composantes principales : Appliquez l'ACP aux variables mesurées restantes pour réduire la dimensionnalité des données.
5. Reconstruction de la variable défectueuse : Utilisez les composantes principales pour reconstruire la variable défectueuse en effectuant une régression linéaire entre les composantes principales et la variable défectueuse.
6. Validation de la reconstruction : Vérifiez la qualité de la reconstruction en comparant les valeurs prédites avec les valeurs réelles ou en utilisant des mesures d'évaluation appropriées.

Il est important de considérer les limitations de cette méthode, telles que l'hypothèse de linéarité de la relation entre les variables mesurées et la variable défectueuse. Une analyse approfondie est recommandée pour évaluer la pertinence de cette méthode dans un contexte spécifique.

$$C = \hat{P}\hat{P}^T = \begin{bmatrix} C_{11} & \cdots & C_{1m} \\ \vdots & \ddots & \vdots \\ C_{1m} & \cdots & C_{nm} \end{bmatrix} \quad (3.9)$$

L'estimation des mesures est ainsi,

$$X = \begin{Bmatrix} \widehat{X}_1 \\ \vdots \\ \widehat{X}_n \end{Bmatrix} = Cx = C \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \quad (3.10)$$

Alors :

$$\hat{x}_1 = C_{11}x_n + C_{12}x_n + C_{13}x_n \quad (3.11)$$

$$\hat{x}_2 = C_{21}x_n + C_{22}x_n + C_{23}x_n \quad (3.12)$$

$$\hat{x}_3 = C_{31}x_n + C_{32}x_n + C_{33}x_n \quad (3.13)$$

### Chapitre 3 Diagnostic de défauts

En se basant sur le principe de la reconstruction, la  $i^{ieme}$  variable est calculée à partir toutes les autres variables et le modèle ACP sauf le  $i^{ieme}$  variable.

$$\hat{x}_1 = z_1 = f(x_1, x_{13}) \quad (3.14)$$

$$\hat{x}_2 = z_2 = f(x_1, x_3) \quad (3.15)$$

$$\hat{x}_3 = z_3 = f(x_1, x_2) \quad (3.16)$$

Si on suppose :  $\hat{x}_i = x_i$

$$x_1 = C_{11}x_1 + C_{12}x_2 + C_{13}x_3 \quad (3.17)$$

$$x_2 = C_{21}x_1 + C_{22}x_2 + C_{23}x_3 \quad (3.18)$$

$$x_3 = C_{31}x_1 + C_{32}x_2 + C_{33}x_3 \quad (3.19)$$

On trouve :

$$x_1(1 - C_{11}x_1) = C_{12}x_2 + C_{13}x_3 \quad (3.20)$$

$$x_2(1 - C_{22}x_2) = C_{21}x_1 + C_{23}x_3 \quad (3.21)$$

$$x_3(1 - C_{33}x_3) = C_{31}x_1 + C_{32}x_2 \quad (3.22)$$

Donc :

$$x_1 = \frac{C_{12}x_2 + C_{13}x_3}{(1 - C_{11}x_1)} \quad (3.23)$$

$$x_2 = \frac{C_{21}x_1 + C_{23}x_3}{(1 - C_{22}x_2)} \quad (3.24)$$

$$x_3 = \frac{C_{31}x_1 + C_{32}x_2}{(1 - C_{33}x_3)} \quad (3.25)$$

Nous mettons  $\hat{x}_i = x_i = z_i$ , nous obtenons alors :

$$z_1 = \frac{C_{12}x_2 + C_{13}x_3}{(1 - C_{11}x_1)} \quad (3.26)$$

$$z_2 = \frac{C_{21}x_1 + C_{23}x_3}{(1 - C_{22}x_2)} \quad (3.27)$$

$$z_3 = \frac{C_{31}x_1 + C_{32}x_2}{(1 - C_{33}x_3)} \quad (3.28)$$

### Chapitre 3 Diagnostic de défauts

La solution de  $z_i$  vérifie :

$$z_1(1 - C_{11}) = [0 \ C_{12} \ C_{13}] \quad (3.29)$$

$$z_2(1 - C_{22}) = [C_{21} \ 0 \ C_{23}] \quad (3.30)$$

$$z_3(1 - C_{33}) = [C_{31} \ C_{32} \ 0] \quad (3.31)$$

Donc :

$$z_1 = \frac{[0 \ C_{12} \ C_{13}]}{1 - C_{11}} x(k) \quad (3.32)$$

$$z_2 = \frac{[C_{21} \ 0 \ C_{23}]}{1 - C_{22}} x(k) \quad (3.33)$$

$$z_3 = \frac{[C_{31} \ C_{32} \ 0]}{1 - C_{33}} x(k) \quad (3.34)$$

Après avoir effectué la reconstruction de la  $i^{\text{ème}}$  variable on remplace la  $i^{\text{ème}}$  variable  $x_i$  par  $z_i$  dans le vecteur de mesure dans le  $i^{\text{ème}}$  bloc de données [23].

La reconstruction de la mauvaise variable annule l'effet d'échec (avant et après la reconstruction de tous les blocs de variable pour toutes les données), de sorte que l'indice de bloc ne dépasse pas son seuil de détection.

Quant aux autres, on remarque qu'ils ont franchi le seuil de détection pour chacun d'eux [24].



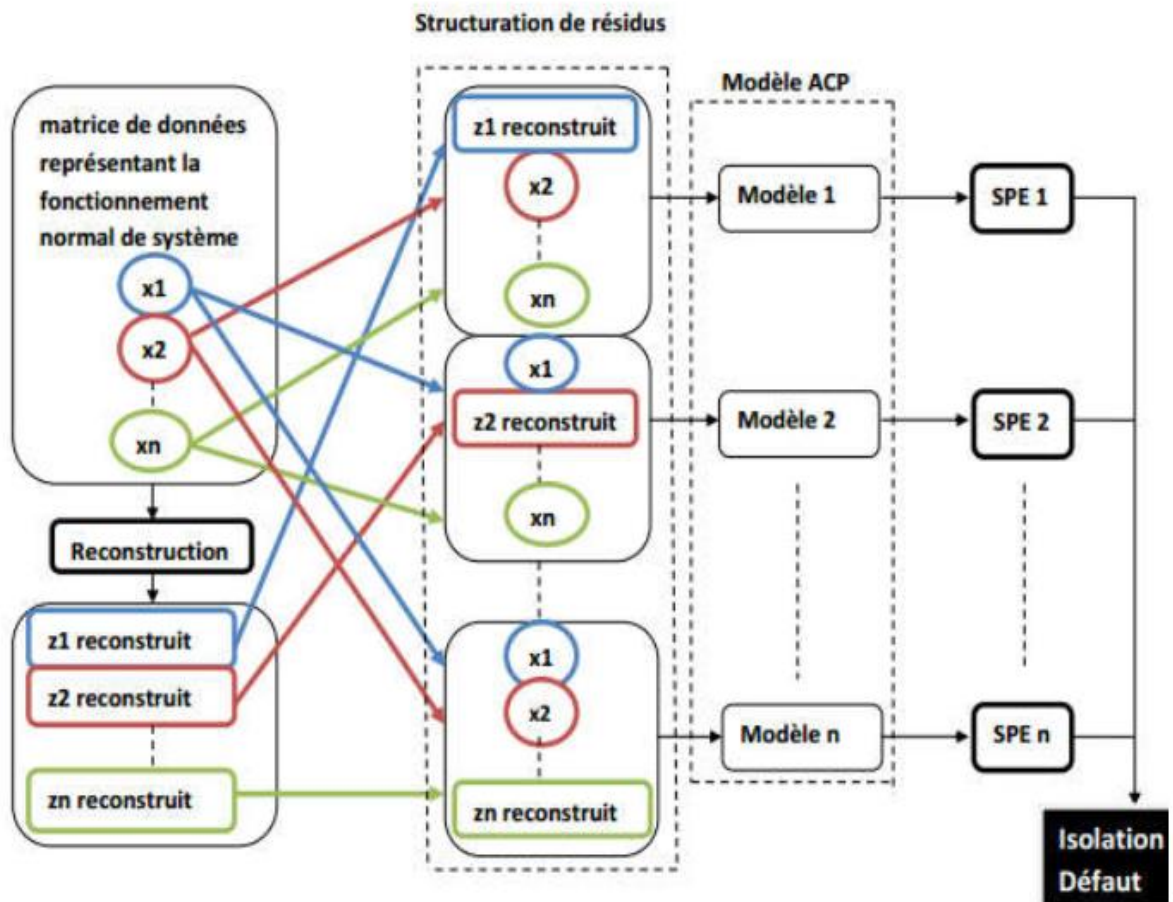


Figure 3.1-Localisation basée sur le principe de reconstruction.

### 3.4 Exemple de simulation

#### Cas linéaire

Pour étudier l'impact des défauts sur la génération de résidus et les composantes principales, une simulation de défaut est effectuée sur les données, plus précisément sur la variable  $x_3$ , à partir de l'instant 700 jusqu'à 1000. La figure suivante illustre l'évolution de la variable en présence du défaut.

### Chapitre 3 Diagnostic de défauts

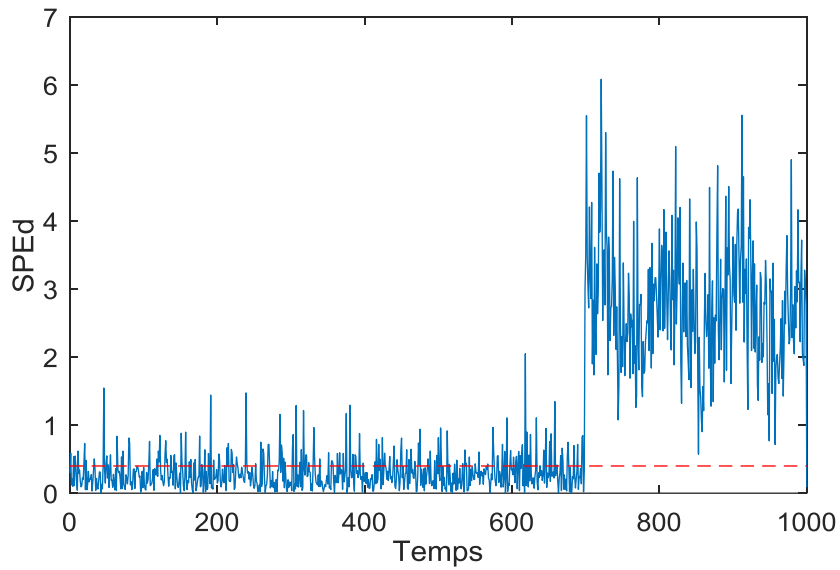


Figure 3.2-Evolution de SPEd en défaut.

Dans l'instant même du défaut, la figure suivante montre l'évolution de l'indice de détection.

Ce dernier dépasse sa valeur du seuil, ce qui reflète la détection du défaut. Dans le cas anormal du processus, le défaut est détecté par l'indicateur SPE représentant les variations des données dans l'espace résiduel, Comme le montre la figure suivante:

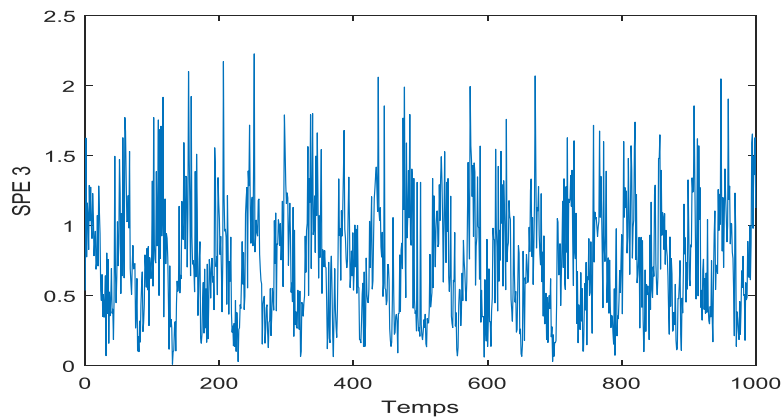


Figure 3.3-Evolution SPE sans défaut (bloc 3).

### Chapitre 3 Diagnostic de défauts

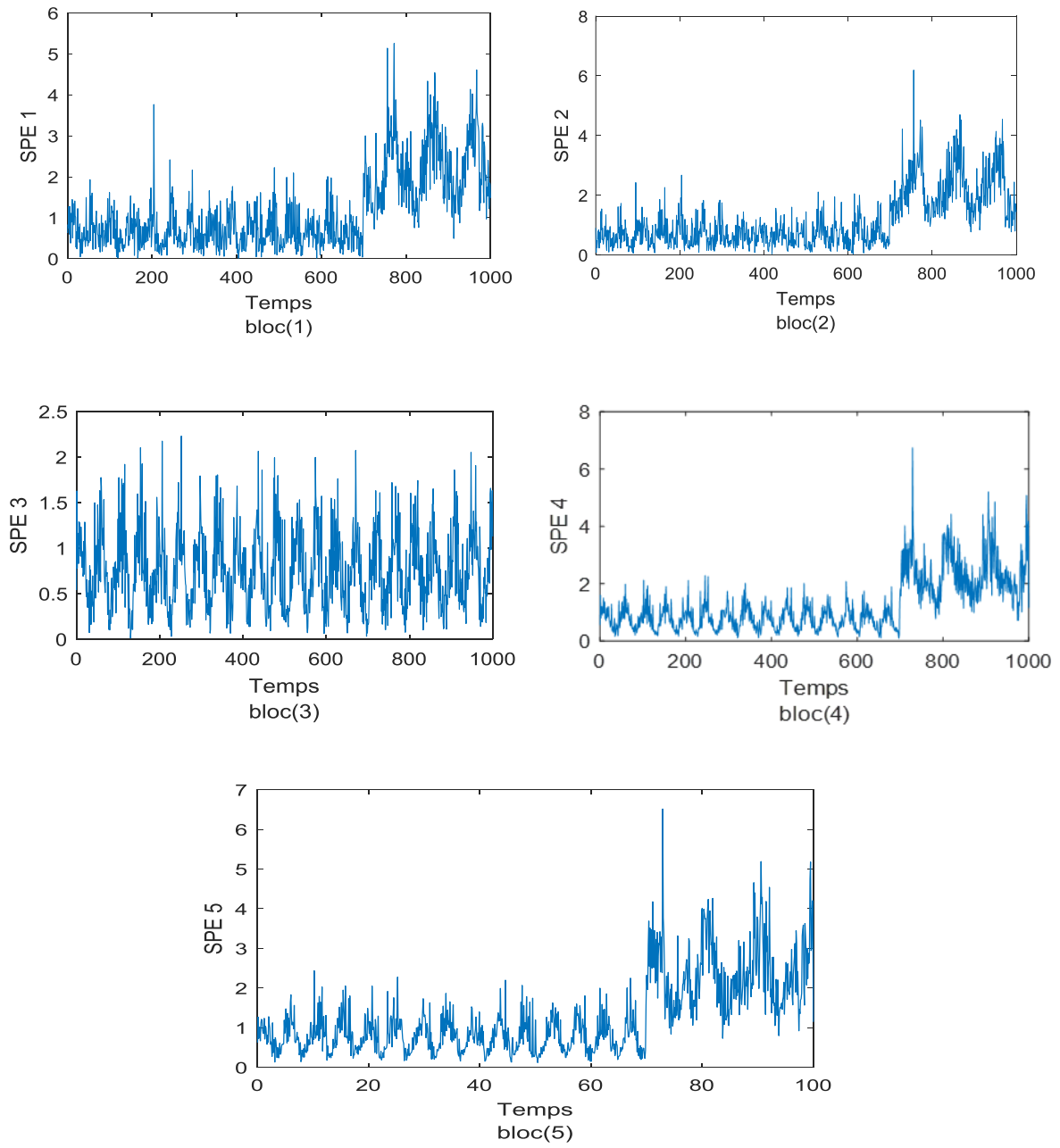


Figure 3.4-L'évolution de SPE dans le bloc (1-5) de défaut.

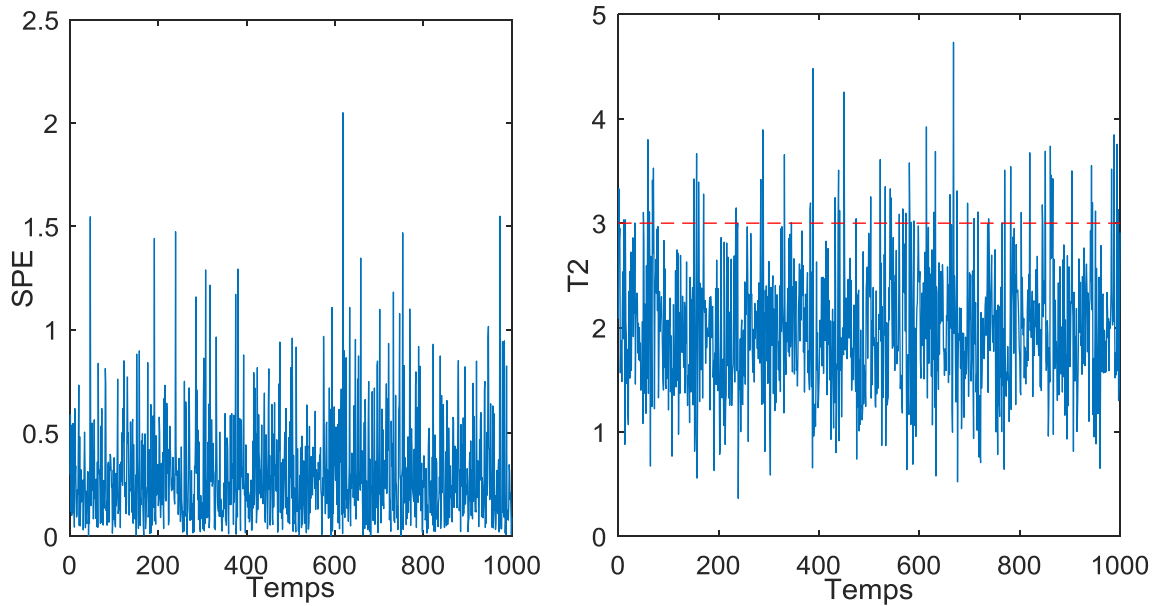


Figure 3.5-L'indice SPE et T2 en utilisant le modèle ACP statique.

#### Cas non linéaire

Dans cet exemple nous avons appliqué la méthode KACP pour la détection et la localisation de défauts d'un système statique représenté par les trois (03) variables modélisées par la formulation suivante :

$$\begin{cases} x1 = ecg(t) + \varepsilon \\ x2 = sig(t) + \varepsilon \\ x3 = \cos(t)^2 + \varepsilon \end{cases}$$

L'évolution des mesures des variables  $x_i(k)$ ,  $i = 1, \dots, 3$  est illustrée sur la figure (3.6)

### Chapitre 3 Diagnostic de défauts

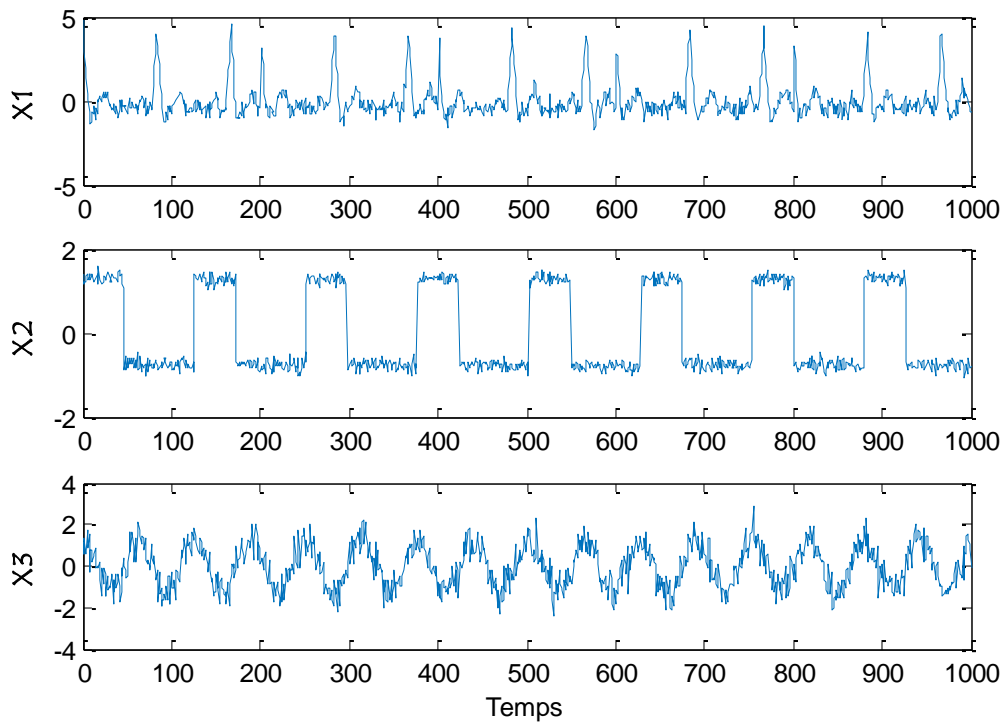


Figure 3.6- Evolution des différentes variables.

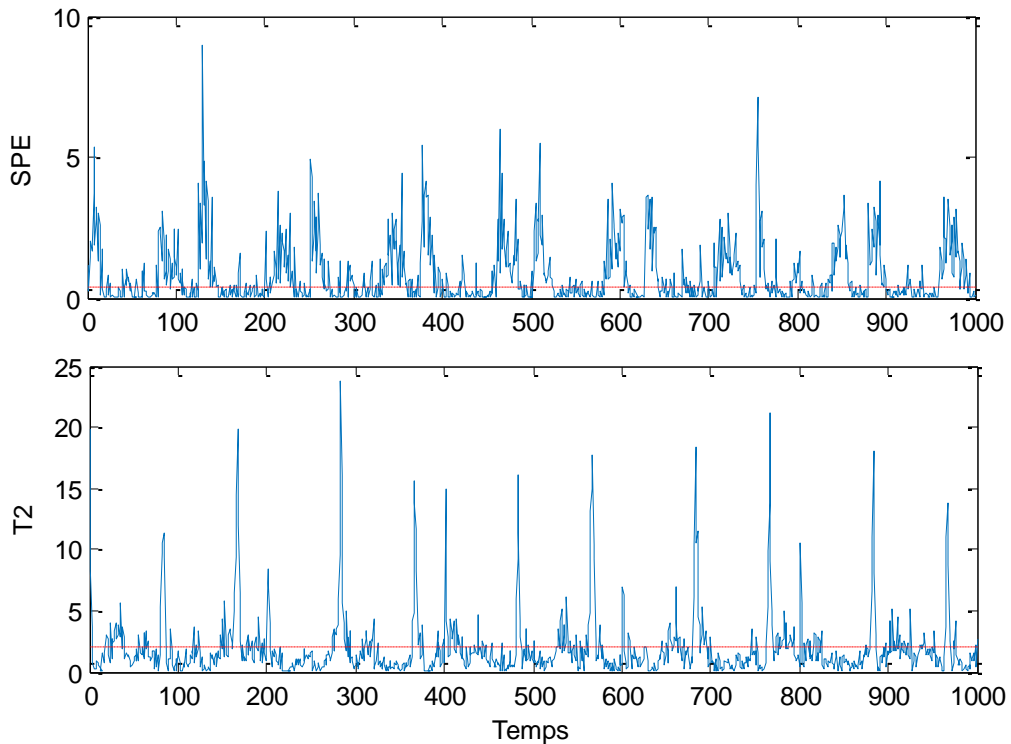


Figure 3.7-Evolution de l'indice SPE et T2 sans défaut.

### Chapitre 3 Diagnostic de défauts

Une simulation de défaut est réalisée sur les données afin d'étudier l'effet des défauts sur la génération de résidus et les composantes principales, spécifiquement sur la variable  $x_1$ , entre les instants 700 et 1000. La figure ci-dessous représente l'évolution de la variable en présence du défaut.

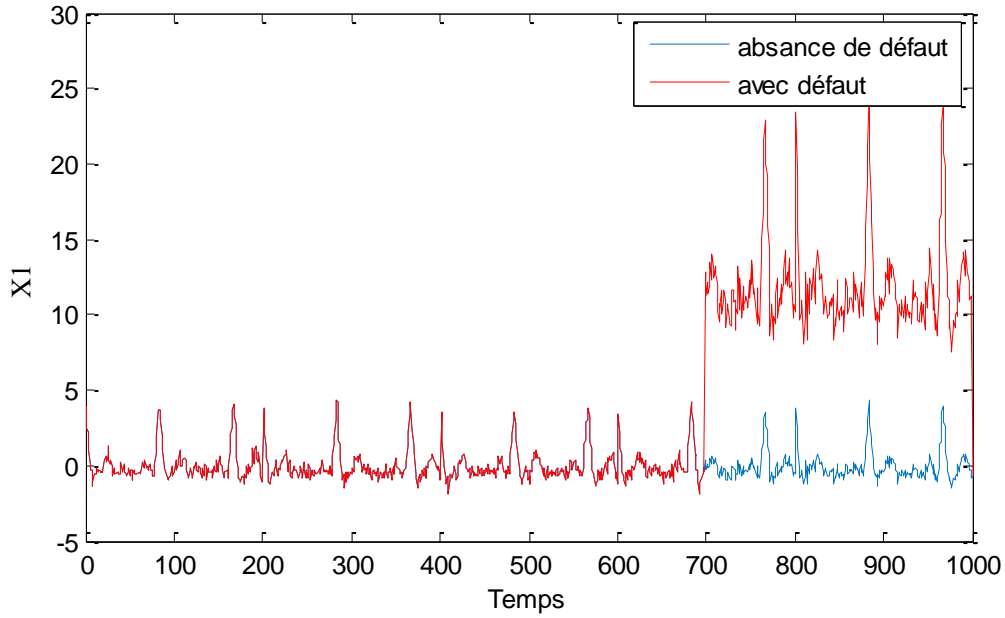


Figure 3.8-L'évolution de la variable  $x_1$  sans et avec défaut.

Dans l'instant même du défaut, la figure suivante montre l'évolution de l'indice de détection.

Ce dernier dépasse sa valeur du seuil, ce qui reflète la détection du défaut. Dans le cas anormal du processus, le défaut est détecté par l'indicateur SPE représentant les variations des données dans l'espace résiduel, Comme le montre la figure suivante :

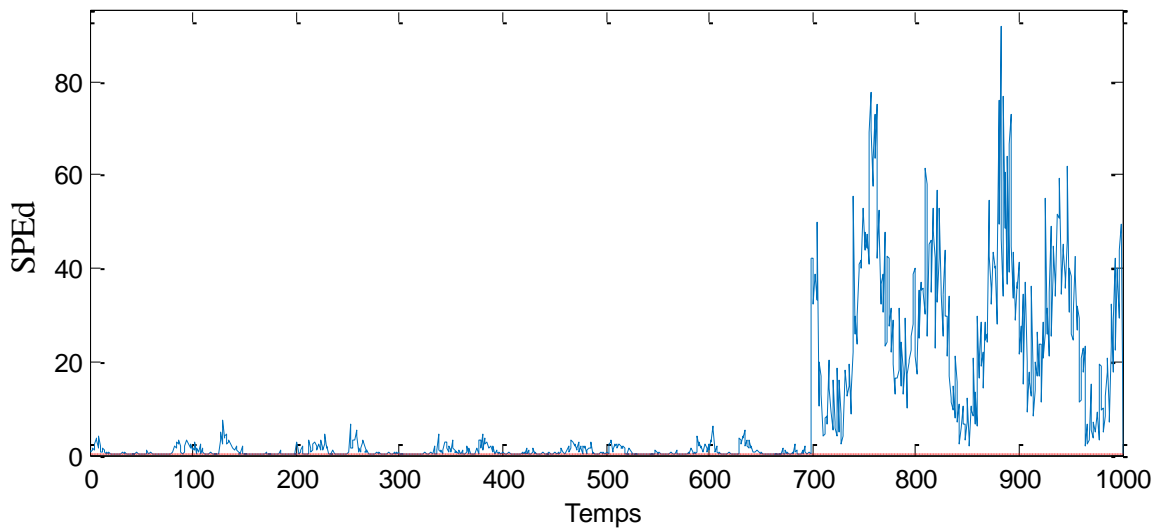


Figure 3.9-Evolution de SPE avec défaut.

### 3.5 Conclusion

Dans ce chapitre, nous avons présenté les démarches à suivre dans l'approche de l'Analyse en Composantes Principales (ACP) et ACP à noyau pour la détection et la localisation des défauts.

Le but du diagnostic est de détecter rapidement les différents défauts présents dans un processus afin d'éviter la dégradation de ses performances et d'augmenter la sécurité des opérateurs et de l'environnement. Le choix des méthodes de diagnostic de défaillance du modèle dépend principalement de la qualité de la mathématique disponible du système, ainsi que de la connaissance du système et de sa structure. De plus, la qualité atteignable de l'isolement des défauts dépend de la manière décisive du nombre de mesures disponibles. Parmi ces méthodes, nous avons choisi la méthode statistique, et l'une de ses techniques (PCA, KPCA) sera détaillée.

## **Chapitre 4 Application**

### **4.1 Description du TE**

L'opération Tennessee Eastman (TE) est un problème intéressant et complexe de contrôle des processus industriels. Elle a été proposée pour la première fois par Downs et Vogel en 1993. Les problèmes liés à cette opération incluent le contrôle multivariable des processus, le contrôle non linéaire, le diagnostic et la surveillance, l'éducation, entre autres.

L'opération TE représente un simulateur de processus qui imite un processus réel. Le processus comprend six unités principales, telles qu'un réacteur exothermique, un condenseur de produit, un séparateur liquide-vapeur, un compresseur de recyclage et un stripper de produit. La feuille de débit du processus est donnée dans **la Figure (4.1)**(non fournie).

Elle se compose de 41 variables mesurées (**Tableau 4.1**) et de 11 variables manipulées (**Tableau 4.2**). Les 41 variables mesurées comprennent 22 variables de processus continues et 19 variables de composition.

L'opération TE dispose de 22 mesures de variables de processus continues, de 12 variables manipulées et de 19 mesures de variables de composition échantillonnées moins fréquemment. Un total de 52 variables est utilisé pour la détection des anomalies dans cette étude. Un ensemble de 21 anomalies est introduit dans le processus. Toutes ces 21 anomalies sont répertoriées dans le (**Tableau 4.3**).

Dans l'ensemble, l'opération TE sert de banc d'essai complexe et complet pour évaluer et développer des stratégies de contrôle, des algorithmes de détection d'anomalies et d'autres techniques de contrôle des processus. Elle permet aux chercheurs d'étudier et de résoudre divers problèmes de contrôle et de diagnostic qui se posent dans les processus

### **4.2 Modélisation de la détection de défauts du processus TE**

Un ensemble de 22 ensembles de données a été généré à partir du simulateur de processus TE, un ensemble pour la condition de fonctionnement normal appelé "Training Data" et 21 ensembles pour les données de test, chaque ensemble contenant les données de 1024 échantillons. La défaillance dans chaque ensemble de données de test est introduite à partir de l'échantillon 224



## Chapitre4 Application

Tableau 4.1-Variables mesurées du processus TE.

Variables	Description	variables	Description
$Xmeas_1$	Un flux(flux 1)	$Xmeas_{23}$	Composition A (flux 6)
$Xmeas_2$	D alimentation(flux 1)	$Xmeas_{24}$	Composition B (flux 6)
$Xmeas_3$	E alimentation (flux 1)	$Xmeas_{25}$	Composition C (flux 6)
$Xmeas_4$	Total des aliments (volet 4)	$Xmeas_{26}$	Composition D (flux 6)
$Xmeas_5$	Flux de recyclage (flux 8)	$Xmeas_{27}$	Composition E (flux 6)
$Xmeas_6$	Débit d'alimentation du réacteur (flux 6)	$Xmeas_{28}$	Composition F (flux 6)
$Xmeas_7$	Pression du réacteur	$Xmeas_{29}$	Composition A (flux 9)
$Xmeas_8$	Niveau du réacteur	$Xmeas_{30}$	Composition B (flux 9)
$Xmeas_9$	Température du réacteur	$Xmeas_{31}$	Composition C (flux 9)
$Xmeas_{10}$	Taux de purge (flux 9)	$Xmeas_{32}$	Composition D (flux 9)
$Xmeas_{11}$	Température du séparateur de produit	$Xmeas_{33}$	Composition E (flux 9)
$Xmeas_{12}$	Niveau séparateur produit	$Xmeas_{34}$	Composition F (flux 9)
$Xmeas_{13}$	Pression du séparateur de produit	$Xmeas_{35}$	Composition G (flux 9)
$Xmeas_{14}$	Sous-flux du séparateur de produit (flux 10)	$Xmeas_{36}$	Composition H (flux 9)
$Xmeas_{15}$	Niveau stripteur	$Xmeas_{37}$	Composition D (flux 11)
$Xmeas_{16}$	Pression de décapant	$Xmeas_{38}$	Composition E (flux 11)
$Xmeas_{17}$	Sous-débit du dénudeur (flux 11)	$Xmeas_{39}$	Composition F (flux 11)
$Xmeas_{18}$	Température du décapant	$Xmeas_{40}$	Composition G (flux 11)
$Xmeas_{19}$	Flux de décapant	$Xmeas_{41}$	Composition H (flux 11)
$Xmeas_{20}$	Travaux de compression		
$Xmeas_{21}$	Température de sortie d'eau de refroidissement du réacteur		
$Xmeas_{22}$	Température de sortie d'eau de refroidissement du séparateur		

## Chapitre4 Application

Tableau 4.2-Variables manipulées du processus TE.

Variables	Description
$XMV1_{42}$	Débit d'alimentation de D (flux 2)
$XMV2_{43}$	Débit d'alimentation de A (flux 3)
$XMV3_{44}$	Débit d'alimentation de E (flux 1)
$XMV4_{45}$	Débit d'alimentation total (flux 4)
$XMV5_{46}$	Vanne de recyclage du compresseur
$XMV6_{47}$	Vanne de purge (flux 9)
$XMV7_{48}$	Débit de liquide du pot de séparation (flux 10)
$XMV8_{49}$	Débit de produit liquide du stripper (flux 11)
$XMV9_{50}$	Vanne du flux du stripper
$XMV10_{51}$	Débit d'eau de refroidissement du réacteur
$XMV11_{52}$	Débit d'eau de refroidissement du condenseur

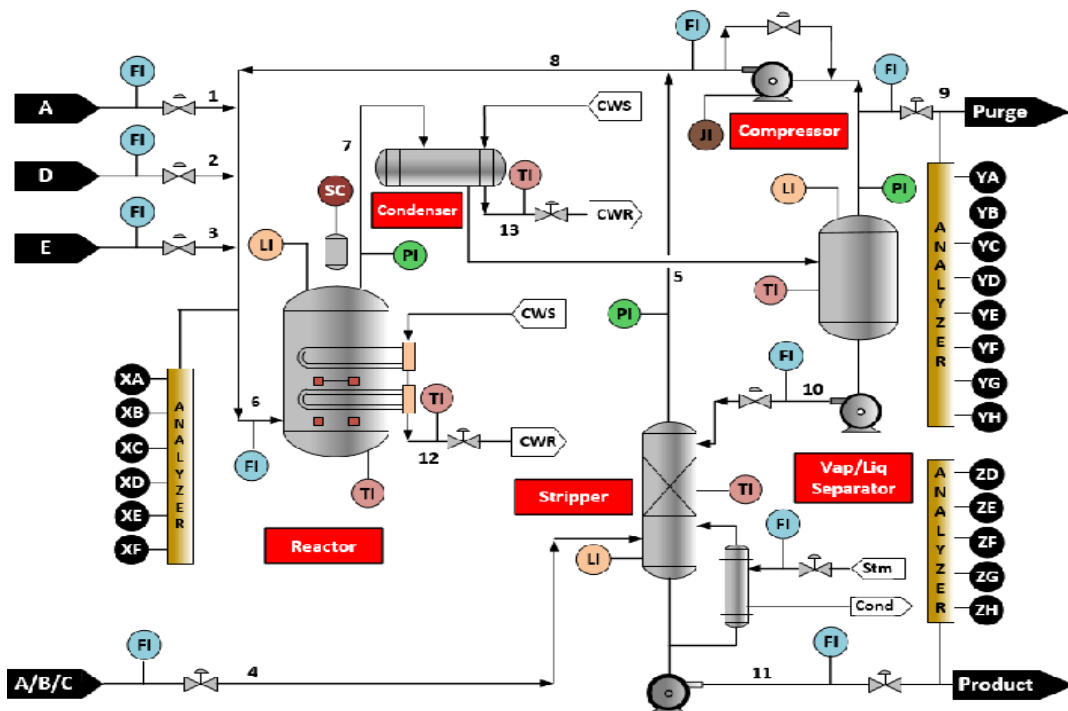


Figure 4.1-Processus de Tennessee Eastman.

## Chapitre4 Application

Tableau 4.3-Résumé des défauts du processus TE

<b>Défaut</b>	<b>Variable de processus</b>	<b>Type</b>
<i>IDV</i> <sub>1</sub>	Ratio d'alimentation A/C, composition de B constante (flux 4)	Étape
<i>IDV</i> <sub>2</sub>	Composition de B, ratio d'alimentation A/C constant (flux 4)	Étape
<i>IDV</i> <sub>3</sub>	Température d'alimentation de D (flux 2)	Étape
<i>IDV</i> <sub>4</sub>	Température d'entrée de l'eau de refroidissement du réacteur	Étape
<i>IDV</i> <sub>5</sub>	Température d'entrée de l'eau de refroidissement du condenseur	Étape
<i>IDV</i> <sub>6</sub>	Perte d'alimentation de A (flux 1)	Étape
<i>IDV</i> <sub>7</sub>	Perte de pression de l'en-tête de C - disponibilité réduite (flux 4)	Étape
<i>IDV</i> <sub>8</sub>	Compositions d'alimentation A, B et C (flux 4)	Étape
<i>IDV</i> <sub>9</sub>	Température d'alimentation de D (flux 2)	Variation aléatoire
<i>IDV</i> <sub>10</sub>	Température d'alimentation de C (flux 4)	Variation aléatoire
<i>IDV</i> <sub>11</sub>	Température d'entrée de l'eau de refroidissement du réacteur	Variation aléatoire
<i>IDV</i> <sub>12</sub>	Température d'entrée de l'eau de refroidissement du condenseur	Variation aléatoire
<i>IDV</i> <sub>13</sub>	Cinétique de réaction	Décalage lent
<i>IDV</i> <sub>14</sub>	Vanne de l'eau de refroidissement du réacteur	Inconnu
<i>IDV</i> <sub>15</sub>	Vanne de l'eau de refroidissement du condenseur	Inconnu
<i>IDV</i> <sub>16</sub>	Inconnu	Inconnu
<i>IDV</i> <sub>17</sub>	Inconnu	Inconnu
<i>IDV</i> <sub>18</sub>	Inconnu	Inconnu
<i>IDV</i> <sub>19</sub>	Inconnu	Inconnu
<i>IDV</i> <sub>20</sub>	Inconnu	Inconnu
<i>IDV</i> <sub>21</sub>	Position de la vanne constante (flux 4)	Position constante

### 4.3 Résultats de la simulation

Afin d'illustrer ces variables de manière appropriée, nous présenterons certaines d'entre elles dans la figure 4.2. Cette figure montre comment différentes mesures évoluent sur une plage de fonctionnement de  $t = 1024$  secondes.

Nous avons commencé par préciser notre choix de perturbation, puis nous avons lu deux matrices :  $xtr$  (matrice des données du système en bon fonctionnement) et  $xstst$  (matrice défectueuse). Nous avons ensuite calculé la moyenne et l'écart-type des données pour les centrer.

Par la suite, nous avons calculé la matrice de covariance des données d'entraînement afin d'obtenir les valeurs et vecteurs propres correspondants. Les valeurs propres ont été triées par ordre décroissant, et nous avons déterminé le nombre optimal de composantes principales à retenir, qui constitueront la nouvelle matrice saine réduite.

#### Cas 1\_ Sans défaut

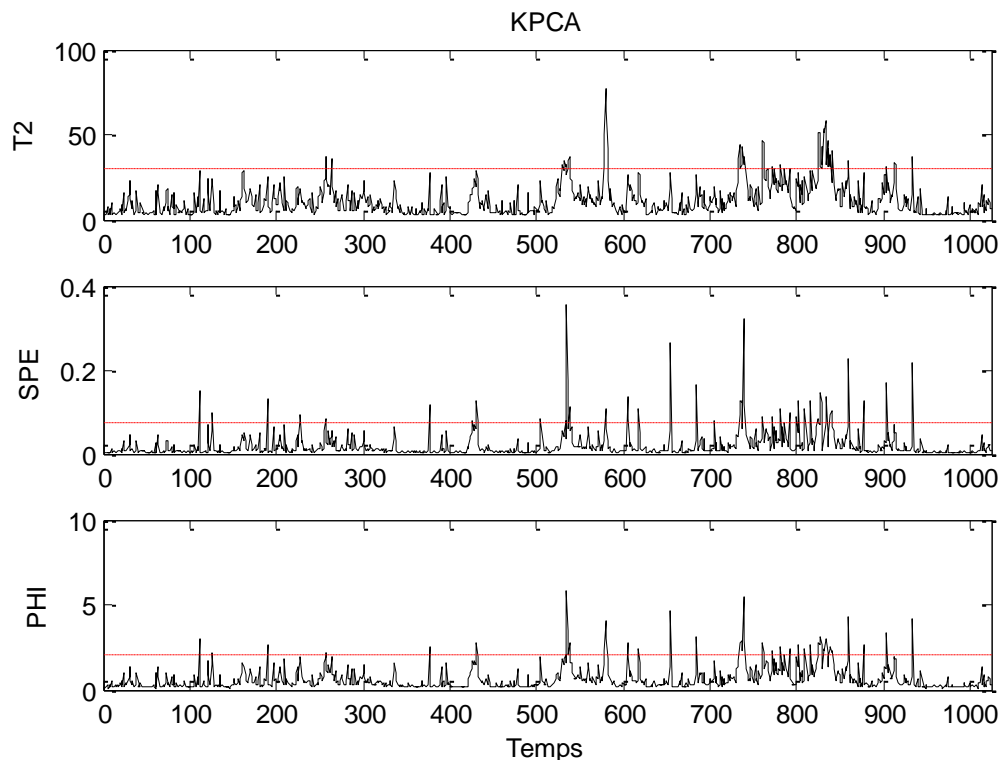


Figure 4.2-Evolution des indices de détection SPE, T2 et Phi sans défaut.

## Chapitre4 Application

Dans cette étude, nous utilisons les mesures de 41 variables du processus de Traitement de l'Eau (TE) pour créer une matrice de données. Ces variables sont répertoriées dans le tableau 4.2. Nous appliquons ensuite la méthode d'analyse en composantes principales à noyaux (KPCA) pour modéliser et diagnostiquer les défauts de ces variables. L'ACP à noyaux est utilisée pour les données avec une structure non linéaire.

À titre d'exemple, nous avons sélectionné quatre défauts spécifiques, et les courbes correspondantes des trois indices  $T^2$ , SPE et Phi sont présentées ci-dessous :

### *Cas 2\_ Avec défaut*

#### Défaut IDV1

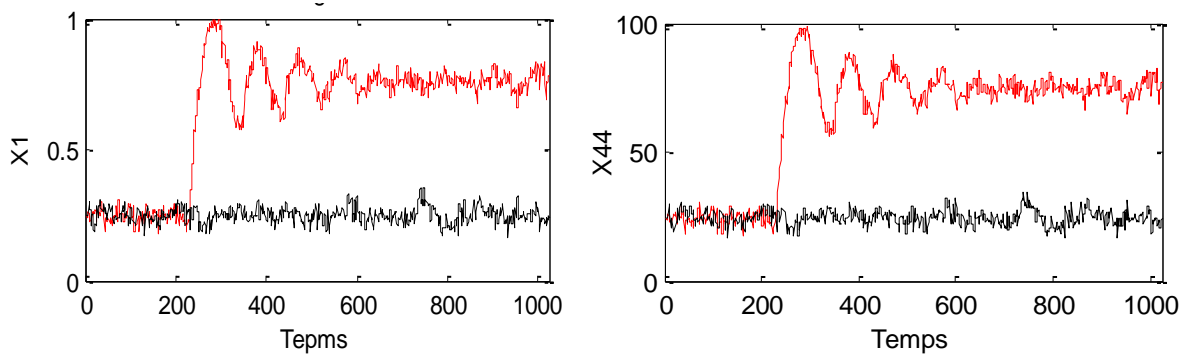


Figure 4.3-Evolution de mesures 'training sans défaut' 'Noir' -et testing avec défaut 'Rouge' du processus TE le défaut IDV1.

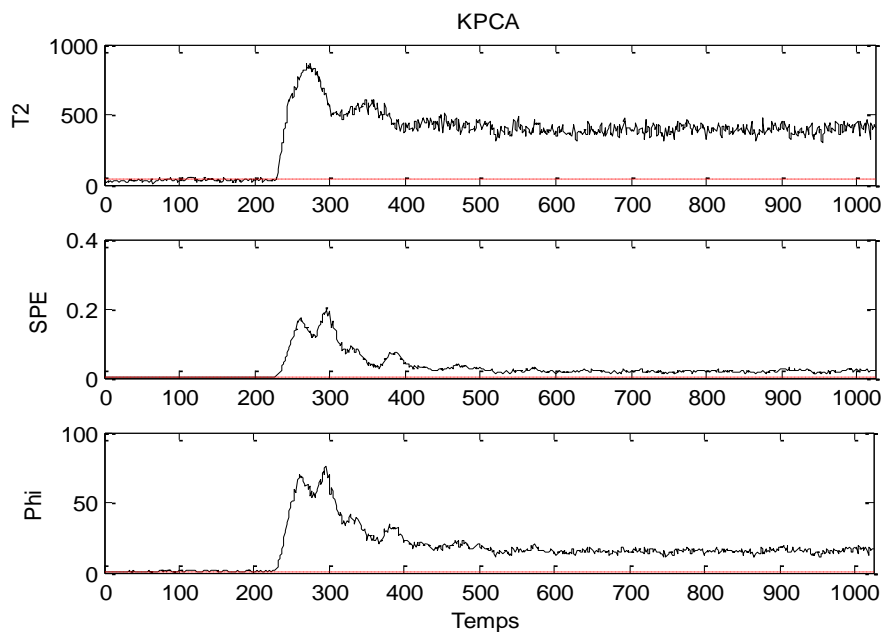


Figure 4.4-Evolution des indices de détection SPE,  $T^2$  et Phi pour le défaut IDV1.

### Défaut IDV2

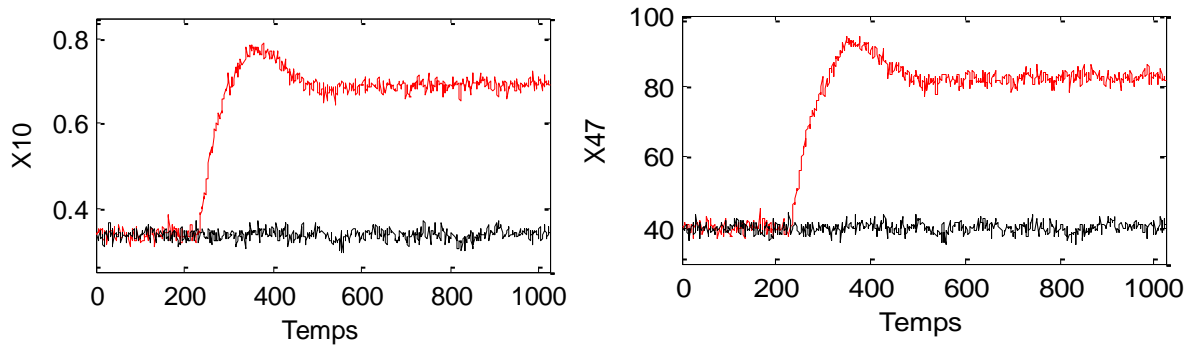


Figure 4.5-Evolution de mesures 'training sans défaut 'Noir' -et testing avec défaut 'Rouge' du processus TE pour le défaut IDV2.

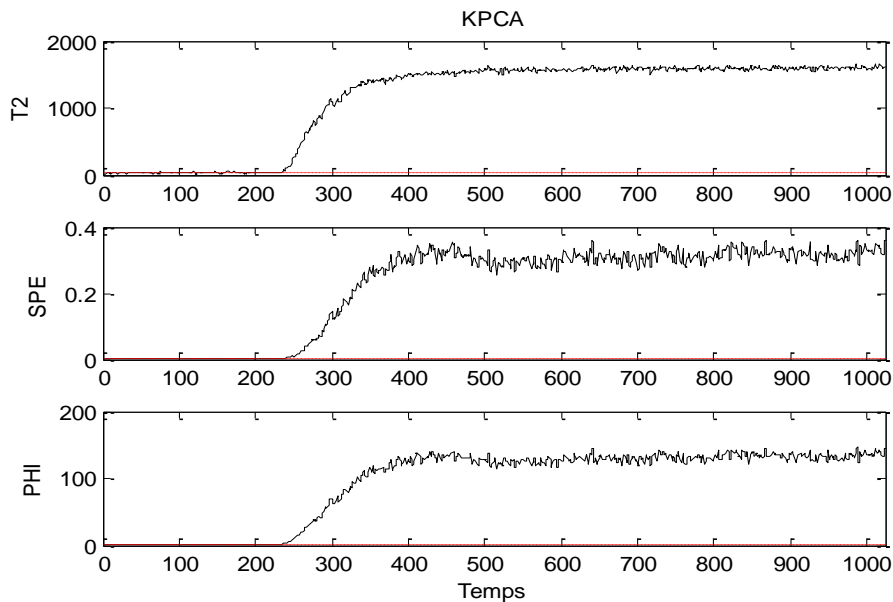


Figure 4.6-Evolution des indices de détection SPE, T2 et Phi pour le défaut IDV2.

### Défaut IDV7

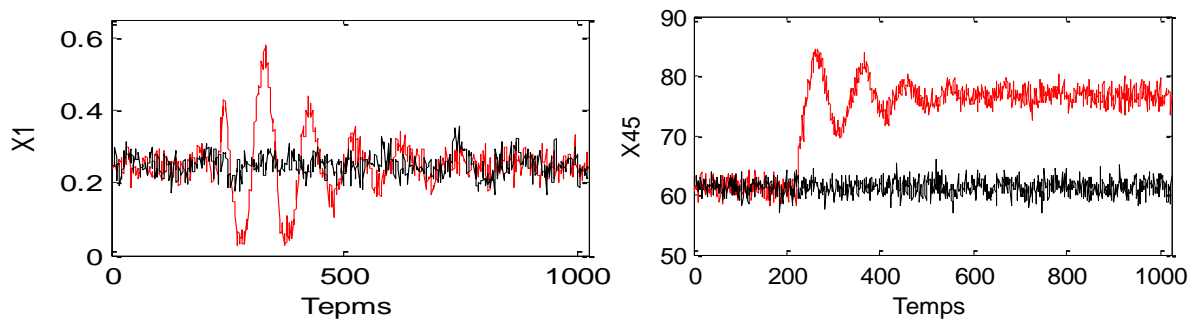


Figure 4.7-Evolution de mesures 'training sans défaut 'Noir' et testing avec défaut 'Rouge' du processus TE pour le défaut IDV7.

## Chapitre4 Application

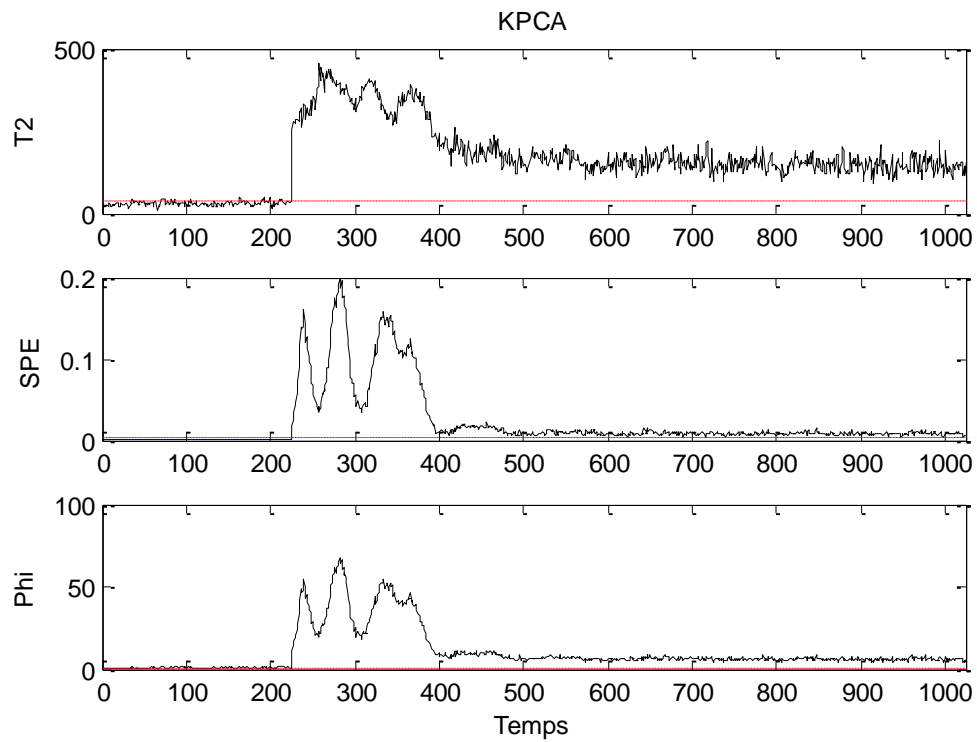


Figure 4.8-Evolution des indices de détection SPE, T2 et Phi pour le défaut IDV7.

### Défaut IDV20

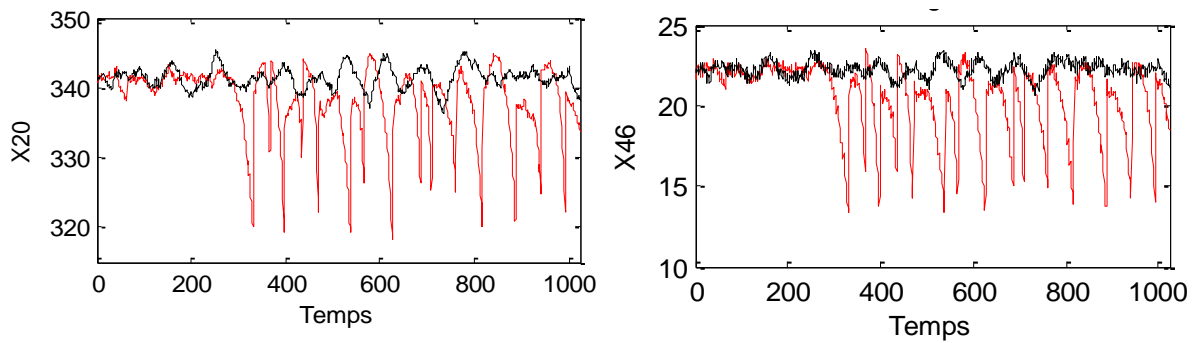


Figure 4.9-Evolution de mesures 'training sans défaut' 'Noir' et testing avec défaut 'Rouge' du processus TE pour le défaut IDV20.

## Chapitre4 Application

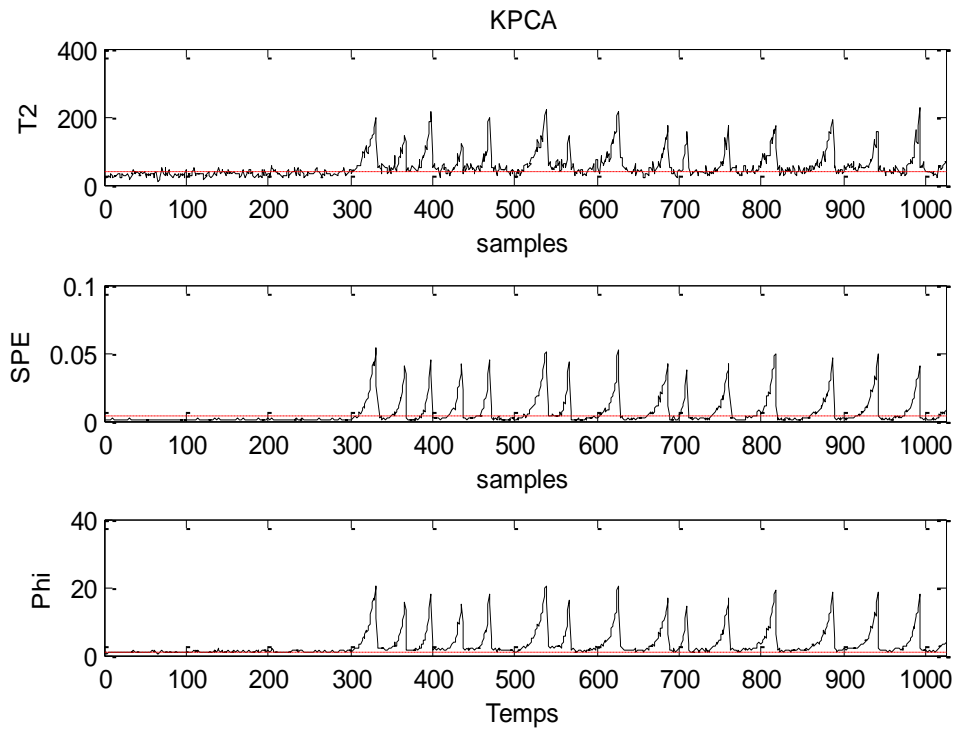


Figure 4.10-Evolution des indices de détection SPE, T2 et Phi pour le défaut IDV20.

Les résultats de simulation démontrent clairement que le modèle KPCA présente de meilleures performances que le modèle PCA.

### 4.4 Conclusion

Ce chapitre examine l'utilisation de l'analyse en composantes principales (ACP) non linéaire dynamique qui est appliquée sur le simulateur du processus chimique Tennessee Eastman. Les résultats obtenus sont présentés dans ce chapitre. L'avantage principal de la méthode ACP à noyau par rapport aux autres extensions de l'ACP non linéaire, reste la simplicité de la représentation des données non linéaires dans un linéaire espace différent, même si cela peut entraîner des temps de calcul importants lors de l'exploitation des données dans le nouvel espace linéaire.



## Conclusion générale

### Conclusion générale

L'utilisation de l'analyse de données multi-variées pour l'identification des états de fonctionnement d'un processus, en particulier dans les systèmes où il est difficile de construire un modèle complet. Les méthodes statistiques sont préférées pour le diagnostic des défauts de processus par rapport aux méthodes basées sur un modèle mathématique. L'analyse des mesures des variables de processus permet d'identifier les causes des défauts, car les états de défaillance sont directement liés aux variables.

L'analyse en composantes principales (ACP) est présentée comme une méthode statistique utilisée pour la surveillance des systèmes. Elle met en évidence les corrélations linéaires entre les variables du processus sans formuler explicitement un modèle entrées/sorties du système.

L'ACP cherche à identifier les vecteurs propres et les valeurs propres de la matrice de corrélation des variables de départ. Cependant, il est nécessaire de déterminer le nombre de composantes principales à retenir pour obtenir la structure du modèle.

Étant donné que la plupart des processus industriels sont non linéaires et dynamiques, l'ACP linéaire n'est pas adaptée à ce type de systèmes et peut entraîner de mauvaises interprétations.

L'ACP à noyau (Kernel PCA) est présentée comme une généralisation de l'ACP linéaire, plus appropriée pour extraire des caractéristiques non linéaires des données à modéliser. L'approche Kernel PCA est utilisée pour la détection et la localisation des défauts. Elle permet de représenter les données non linéaires dans un autre espace linéaire, mais cela peut entraîner une explosion de la base de données et des temps de calcul importants.

L'analyse en Composantes Principales (ACP) et de l'ACP à noyau dans leur version statique, en soulignant que la plupart des processus industriels sont dynamiques et que leurs comportements et caractéristiques statistiques changent dans le temps. Cela est dû à l'incertitude du système et aux variations des conditions de fonctionnement. Les approches proposées pour résoudre ce problème incluent l'ACP à noyau dynamique. Cependant, ces techniques sont confrontées à la complexité de calcul élevée. Une nouvelle approche est proposée comme alternative pour s'adapter à ce problème. L'intérêt de l'ACP à noyau et de son extension dans le domaine dynamique est souligné pour la modélisation et la surveillance des processus industriels.

## Conclusion générale

Le dernier chapitre du texte concerne l'application des algorithmes développés sur un simulateur d'un processus chimique spécifique (TE).

L'objectif futur mentionné est d'adapter la procédure de localisation proposée en utilisant l'ACP linéaire, basée sur la reconstruction à partir de l'espace des caractéristiques, vers l'espace de la pré-image. L'optimisation des fonctions noyaux utilisées pour l'ACP à noyau est également étudiée, car elle a un impact sur la sensibilité de détection des défauts. La normalisation de la matrice noyau dans l'espace des caractéristiques est également considérée comme une question importante.

## Bibliographie

## Bibliographie

- [1]. Jelali, M., & Ben Hmida, F. (2013). Nonlinear process monitoring and fault diagnosis: State-of-the-art review and future trends. *Computers & Chemical Engineering*
- [2]. Montgomery, D. C. (2012). *Introduction to statistical quality control*. John Wiley & Sons.
- [3]. Jolliffe, I. T. (2002). *Principal component analysis*. Wiley Online Library.
- [4]. BEKIRI Walid et SID Bouchra “Diagnostic de Fonctionnement par Analyse en Composantes Principales Multi-Echelles ”.Mémoire master, 2019.Université Kasdi Merbah –Oaurgla
- [5]. Heng, A., Hong, G. S., & Xi, Z. (2018). Fault diagnosis of industrial systems using data-driven methods: A review and a new hierarchical multi-class support vector machine. *IEEE Transactions on Industrial Electronics*, 66(1), 151-162.
- [6]. Chakour Chouaib, “ Diagnostic et surveillance des procédés industriels et de leur environnement sur la base de l’analyse de données.” Thèse de doctorat, 2015/2016, Université Badji Mokhtar Annaba.
- [7]. D. Bensoussan, A. Ben-Israel, and M. Goursat, "Fault Detection and Diagnosis in Engineering Systems", CRC Press, 2018.
- [8]. ROUIDJALI Mohammed Amine et CHAIMI Abdelkrim“ Détection des défauts basés sur l’Analyse en Composantes Principales Application à un Four électrique, à un Groupeélectrogène et à une Chaudière”. Mémoire master Juin 2017. UNIVERSITE M’HAMED BOUGARA-BOUMERDES
- [9]. Bergoug Aicha“ Détection et Localisation de Défauts par Analyse en Composantes Principale (ACP) ”Mémoire master, 2017. Université Kasdi MerbahOuargla.
- [10]. KHENFER RIAD “ Détection et isolation de défauts combinant des méthodes à base de données appliquées aux systèmes électro-énergétiques” Thèse de doctorat 05 Mars 2015. UNIVERSITE FERHAT ABBES – SETIF
- [11]. NAIT SLIMANI Boukhalfa, “Synthèse d’observateurs non linéaires : Application au diagnostic de défauts ” Mémoire master, 2011. Université Mouloud Mammeri de TiziOuzou
- [12]. Dimitrios Fragkoulis. “ Détection et localisation des défauts provenant des capteurs et des actionneurs : application sur un système non linéaire”. Thèse de doctorat 13 novembre 2008 Automatique / Robotique. Université Paul Sabatier - Toulouse, Français.
- [13]. Abdi, H., & Williams, L. J. (2010). *Principal component analysis*. Wiley Interdisciplinary Reviews: Computational Statistics, 2(4), 433-459.

## Bibliographie

- [14]. Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20(1), 141-151.
- [15]. Hinton, G. E. et Kruskal, J. B.(2005) .Cross-validation and model selection in principal component analysis .*Journal of Classification*, Volume 22, Issue 2, pp. 255-282.
- [16]. Smith, J., & Johnson, A. (2011).Determining the Number of Principal Components in Principal Component Analysis: A Comparative Study.*Journal of Multivariate Analysis*, Volume 102, Issue 9, Pages 1542-1557.
- [17]. *Harkat Mohamed-Faouzi* Détection et Localisation de Défauts par Analyse en Composantes Principales, Thèse Doctorat de l'Institut National Polytechnique de Lorraine. le 30 Juin 2003.
- [18]. Ding, S. X., Zhang, Y., & Huang, B. (2020). *Model-based fault diagnosis and control system design*. Springer.
- [19]. Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*.
- [20]. Vardeman, S. B. et Jobe, J. M. (2013). *Statistical Quality Assurance Methods for Engineers*. John Wiley & Sons.
- [21]. Yvon T, Mohamed F H , Gilles M , Jose R , Détection et localisation de défautsmultiples par analyse en composantes principales Bordeaux, France May 2009.
- [22]. Jolliffe, I. T. (2002). *Principal Component Analysis*. Wiley Online Library.
- [23]. Mohamed-Faouzi Harkat, Nouredine Doghmane Modélisation et diagnostic des systèmes dynamiques par analyse en composantes principales multi-echelle (MSPCA), article 2007
- [24]. Anissa BEN AICHA, “Détection et localisation de d'défauts par Analyse en Composantes Principales certaine ou incertaine de type intervalle ”. Thèse de doctorat 20septembre 2011, Génie Electrique. L'Ecole Nationale d'Ingénieurs de Monastir