



République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur
et de la Recherche Scientifique
Université 20 Août 1955 - Skikda
Faculté de Technologie
Département de génie électrique



Mémoire présentée pour l'obtention du diplôme de Master

Filière : Electronique
Spécialité : Instrumentation

Présenté par :
LAKEHAL AYMEN

Thème :

**Sciences des données et diagnostic prédictif des
systèmes de mesure dans l'industrie 4.0**

Membres du Jury :

Pr. CHAKOUR Chouaib	Encadrant	Université 20 Août 1955 - Skikda
Pr. LACHOURI Abderezzak	Président	Université 20 Août 1955 - Skikda
Dr. KARED Saber	Examineur	Université 20 Août 1955 - Skikda
Dr. TARFA Rahma	Examineur	Université 20 Août 1955 - Skikda

Année universitaire : 2025/2026

Remerciements

Avant tout développement de ce travail, il me tient particulièrement à cœur d'exprimer ma profonde gratitude envers toutes les personnes qui, de près ou de loin, ont contribué à la réalisation de ce mémoire.

Je tiens tout d'abord à adresser mes remerciements les plus sincères à mon encadrant, Professeur Chakour Chouaib, pour sa disponibilité constante, ses conseils avisés, sa patience et son accompagnement tout au long de ce travail. Sa rigueur scientifique, ses orientations pertinentes et ses remarques constructives ont grandement contribué à l'aboutissement de ce mémoire.

Mes remerciements s'adressent également aux membres du jury, qui me font l'honneur d'évaluer ce travail. Je les remercie pour l'intérêt qu'ils ont bien voulu porter à mon mémoire, ainsi que pour le temps et l'attention consacrés à sa lecture et à son évaluation.

Je souhaite également exprimer toute ma reconnaissance à mes collègues et amis, pour leur soutien moral, leurs encouragements sincères et les précieux moments de partage qui ont jalonné ce parcours.

Enfin, mes remerciements les plus profonds et les plus chaleureux vont à ma famille, et tout particulièrement à mes parents, pour leur soutien indéfectible, leurs encouragements constants et les sacrifices qu'ils ont consentis afin de me permettre d'arriver jusqu'ici. Qu'ils trouvent ici le témoignage de ma plus profonde reconnaissance.

Que toutes les personnes ayant contribué, de près ou de loin, à l'élaboration de ce travail, trouvent ici l'expression de ma sincère et respectueuse gratitude.

Dédicace

Je dédie ce modeste travail à ma chère famille, source inépuisable de patience et d'encouragement,

À mes parents, pour leur amour inconditionnel, leurs sacrifices et leur soutien indéfectible tout au long de mon parcours,

À ma sœur, pour sa présence et ses encouragements dans les moments de doute,

À toute ma famille, pour avoir toujours cru en moi et m'avoir donné la force de persévérer,

À tous mes amis, pour leur soutien précieux, leurs encouragements constants et les moments partagés qui ont adouci les difficultés de ce parcours,

À tous ceux qui, de près ou de loin, ont contribué à la réussite de ce travail.

ملخص

في ظل الثورة الصناعية الرابعة (Industry 4.0) ، أدى التزايد الكبير في عدد الحساسات وما يرافقه من عدم يقين متأصل في القياسات الصناعية إلى جعل التشخيص التنبؤي للأعطال أكثر تعقيداً. تتناول هذه المذكرة موضوع تشخيص الأنظمة الصناعية باستخدام تحليل المكونات الأساسية (ACP) ، وهي طريقة إحصائية معروفة تُستخدم في نمذجة العمليات ذات الارتباط القوي بين متغيراتها. وبعد دراسة الأسس النظرية لتحليل المكونات الأساسية الكلاسيكي وحدوده عند التعامل مع بيانات غير دقيقة، يقترح هذا العمل توسيع هذه الطريقة لتشمل البيانات من نوع المجال (interval data) من خلال مقارنة القياسات بعين الاعتبار بشكل أكثر واقعية. تم تطبيق هذه المقاربة بعد ذلك على عمليتي كشف وتحديد موقع الأعطال، باستخدام مؤشر SPE من جهة، ومبدأ إعادة البناء (reconstruction) للمتغيرات من جهة أخرى، بعد توسيعهما ليلتسبا مع الإطار الرمزي (symbolic). وقد تم التحقق من أداء الطريقة المقترحة من خلال تطبيقها على حالة صناعية حقيقية، تتمثل في مجموعة بيانات مرجعية (benchmark) خاصة بتوربين غازي. وأظهرت النتائج أن مقاربة SPCA تتيح كشفاً موثقاً وتحديداً دقيقاً لمواقع الأعطال، حتى في ظل وجود عدم يقين في القياسات، مما يؤكد أهميتها في مراقبة الأنظمة الصناعية الحديثة وتشخيصها بشكل تنبؤي .

الكلمات المفتاحية : تحليل المكونات الأساسية (ACP) ، SPCA ، تحليل المكونات الأساسية بالمجال، التشخيص التنبؤي، كشف وتحديد موقع الأعطال، الثورة الصناعية الرابعة .

Abstract

In the context of Industry 4.0, the increasing number of sensors and the inherent uncertainty of industrial measurements make predictive fault diagnosis increasingly challenging. This thesis addresses the diagnosis of industrial systems using Principal Component Analysis (PCA), a well-established statistical method for modeling highly correlated processes. After reviewing the theoretical foundations of classical PCA and its limitations when dealing with imprecise data, this work proposes an extension to interval-valued data through the SPCA (Symbolic Principal Component Analysis) approach, allowing for a more realistic representation of measurement uncertainty. This approach is then applied to fault detection and isolation, using respectively the SPE index and the variable reconstruction principle, both extended to the symbolic framework. The performance of the proposed method is validated through an application to a real industrial case : a benchmark dataset from a gas turbine. The results show that the SPCA approach enables reliable fault detection and accurate fault isolation, even under measurement uncertainty, confirming its relevance for the monitoring and predictive diagnosis of modern industrial systems.

Keywords : PCA, SPCA, Interval PCA, predictive diagnosis, fault detection and isolation, Industry 4.0.

Résumé

Dans le contexte de l'industrie 4.0, la multiplication des capteurs et l'incertitude inhérente aux mesures industrielles rendent le diagnostic prédictif des défauts de plus en plus complexe. Ce mémoire porte sur le diagnostic des systèmes industriels par Analyse en Composantes Principales (ACP), une méthode statistique reconnue pour la modélisation de procédés fortement corrélés. Après une étude des fondements théoriques de l'ACP classique et de ses limites face aux données imprécises, ce travail propose son extension aux données de type intervalle à travers l'approche SPCA (*Symbolic Principal Component Analysis*), permettant une prise en compte plus réaliste des incertitudes de mesure. Cette approche est ensuite exploitée pour la détection et la localisation de défauts, à l'aide respectivement de l'indice SPE et du principe de reconstruction des variables, étendus au cadre symbolique. Les performances de la méthode proposée sont validées à travers une application sur un cas industriel réel : un benchmark de données issues d'une turbine à gaz. Les résultats obtenus montrent que l'approche SPCA permet une détection fiable et une localisation précise des défauts de capteurs, même en présence d'incertitudes de mesure, confirmant son intérêt pour la surveillance et le diagnostic prédictif des systèmes industriels modernes.

Mots-clés : ACP, SPCA, ACP par intervalle, diagnostic prédictif, détection et localisation de défauts, industrie 4.0.

Table des matières

1	Généralités sur le diagnostic	13
1.1	Introduction	13
1.2	Objectifs de la supervision	13
1.3	Terminologie	14
1.4	Système de surveillance et/ou de diagnostic	15
1.5	Typologie des défauts	16
1.6	Caractérisation des défauts	17
1.7	Les étapes du diagnostic	17
1.8	Méthodes de diagnostic	18
1.9	Critères de performance d'un système de diagnostic	20
1.10	Conclusion	21
2	L'Analyse en Composantes Principales pour la modélisation statistique des données	22
2.1	Introduction	22
2.2	Analyse en Composantes Principales	22
2.2.1	Principe de la méthode ACP	23
2.3	Nombre de composantes principales	25
2.3.1	Pourcentage cumulé de la variance totale (PCV)	26
2.4	Exemple de simulation	27
2.5	Le problème des grandeurs d'influence et de l'imprécision des mesures	30
2.6	Analyse de Données Symboliques	31
2.6.1	Données de type intervalle	32
2.6.2	Formalisation matricielle des données de type intervalle	33
2.6.3	Erreur de mesure et construction de l'intervalle	33
2.6.4	Arithmétique et statistique d'intervalle	34
2.6.5	Normalisation des données de type intervalle	35
2.7	ACP de type intervalle	35
2.7.1	ACP Symbolique (SPCA)	36

2.8	Résultats de simulation	37
2.9	Conclusion	40
3	Diagnostic de défauts par approche statistique multivariée (ACP)	42
3.1	Introduction	42
3.2	Détection de défaut	42
3.2.1	Indice SPE	43
3.2.2	Indice SPE (cas des données intervalle)	43
3.3	Localisation de défaut basée sur le principe de reconstruction	44
3.3.1	Principe de reconstruction de variable	44
3.4	Simulation	47
3.4.1	Cas de données classiques (mono-valued Data)	47
3.4.2	Cas de données imprécises de type intervalle	51
3.5	Conclusion	54
4	Application sur un benchmark industriel : turbine à gaz	55
4.1	Introduction	55
4.2	Systèmes de production d'énergie	55
4.3	Turbine à gaz	57
4.3.1	Classification des turbines à gaz	57
4.3.2	Composants d'une turbine à gaz	58
4.4	Centrale à turbine à gaz	59
4.4.1	Principe de fonctionnement	59
4.5	Modélisation et diagnostic par la méthode ACP Symbolique	60
4.6	Conclusion	69

Table des figures

1.1	La supervision et la surveillance	14
1.2	La cause principale d'une panne	15
1.3	Typologie de défauts	16
1.4	Les étapes du diagnostic.	18
1.5	Méthodes de diagnostic.	20
2.1	L'idée de base de la méthode ACP	23
2.2	Décomposition du vecteur de mesures	26
2.3	Évolution des différentes variables	27
2.4	Évolution des variables normalisées	28
2.5	Évolution des composantes principales	29
2.6	Évolution des mesures et de leurs estimations	29
2.7	Schéma du système de mesure	30
2.8	Impact de la qualité des données sur la qualité de la connaissance	31
2.9	Processus classique de traitement des données	31
2.10	Processus d'analyse de données symboliques pour les données imprécises	31
2.11	Évolution des variables de type intervalle	38
2.12	Évolution de X_{max} et de son estimation avec la méthode ACP classique	38
2.13	Évolution des différentes mesures (X_{min} , X_{max}) et de leurs estimations par la méthode SPCA	39
2.14	Évolution de X_{new} et de son estimation	40
3.1	Localisation basée sur le principe de reconstruction	47
3.2	Évolution de x_1 sans et avec défaut	48
3.3	Évolution de SPE avec son seuil, avec et sans défaut	48
3.4	l'évolution de x_1 sans et avec défaut et sa variable reconstruite z_1	49
3.5	Ensemble des SPE en l'absence de défaut	49
3.6	Ensemble des SPE en présence d'un défaut sur x_1	50
3.7	Ensemble des SPE en présence d'un défaut sur x_2	50

3.8	Évolution de x_1 sans et avec défaut	51
3.9	Évolution de SPE sans et avec défaut	51
3.10	Évolution de x_1 sans et avec défaut, et sa variable reconstruite z_1	52
3.11	SPE structuré en l'absence de défaut	53
3.12	SPE structuré en présence de défaut sur x_1	53
3.13	SPE structuré en présence de défaut sur x_2	54
4.1	Une centrale de production d'énergie	56
4.2	Différents systèmes de production d'énergie	56
4.3	Turbine à gaz	57
4.4	Classification des turbines à gaz	58
4.5	Principe de fonctionnement d'une turbine à gaz	59
4.6	Évolution des mesures des 11 capteurs	61
4.7	l'évolution des différentes variable après filtrage	62
4.8	Évolution des mesures de type intervalle	63
4.9	Évolution des mesures et leur estimation	64
4.10	Évolution de x_4 sans et avec défaut	65
4.11	Évolution de SPE sans et avec défaut	65
4.12	SPE structuré en l'absence de défaut	66
4.13	SPE structuré en présence de défaut sur x_4	67
4.14	SPE structuré en présence de défaut sur x_5	68

Liste des tableaux

2.1	Capacité de l'ACP	40
4.1	Description des variables de la base de données	60

Introduction générale

Au cours des dernières décennies, les installations industrielles ont connu une évolution technologique majeure, marquée par une complexification croissante des systèmes de production. Cette transformation s'inscrit aujourd'hui dans le cadre plus large de l'industrie 4.0, caractérisée par l'intégration massive de capteurs, de systèmes d'acquisition de données et d'outils d'analyse intelligents au sein des procédés industriels. Cette évolution s'est traduite par une multiplication considérable du nombre de capteurs installés, permettant de mesurer en continu différentes grandeurs physiques et de fournir, à chaque instant, une image représentative de l'état de fonctionnement du système. Dans ce contexte, le diagnostic prédictif de défauts occupe une place centrale : anticiper les pannes et corriger à temps les défaillances permet de préserver la productivité des installations et d'éviter des arrêts de production coûteux. Tout dérèglement non détecté à temps peut en effet engendrer des conséquences économiques importantes, dans un contexte où la qualité, la sécurité et le rendement des processus industriels sont devenus des exigences incontournables. Le diagnostic de défauts représente ainsi un enjeu stratégique majeur pour les installations industrielles modernes.

Le diagnostic industriel se compose de plusieurs étapes complémentaires, qui permettent de détecter, localiser et identifier les défauts affectant un système. Parmi les méthodes et techniques utilisées à cette fin figure l'analyse en composantes principales (ACP). L'objectif principal de cette méthode est de transformer les variables d'un système en un nombre restreint de nouvelles variables, appelées composantes principales, définies comme des combinaisons linéaires des variables d'origine. Cette réduction de dimension n'est possible qu'en présence de redondance entre les variables initiales, redondance que l'ACP exploite précisément pour construire un modèle représentatif du fonctionnement normal du système.

Cependant, la majorité des travaux existants en ACP reposent sur l'hypothèse de variables monovaluées, c'est-à-dire de mesures ponctuelles supposées exactes. Or, dans un contexte industriel réel, en particulier celui de l'industrie 4.0 où les données massives issues de capteurs hétérogènes sont exploitées à grande échelle, les mesures sont souvent imprécises, bruitées ou entachées d'incertitudes de natures diverses. L'extension de l'ACP aux données de type intervalle permet de mieux prendre en compte cette imprécision, en représentant chaque observation non plus par une valeur unique, mais par une plage de valeurs possibles, offrant ainsi une modélisation plus

fidèle et plus robuste des systèmes industriels réels.

C'est dans cette perspective que s'inscrit ce mémoire, consacré à l'étude d'une nouvelle méthode d'extension de l'ACP aux données de type intervalle, en vue du diagnostic prédictif de défauts dans un contexte d'industrie 4.0. Il est structuré en quatre chapitres, organisés comme suit :

Le premier chapitre présente des généralités sur le diagnostic et la surveillance des systèmes industriels, ainsi que les principales méthodes associées.

Le deuxième chapitre est consacré à la modélisation par ACP classique et par ACP symbolique (données de type intervalle) ; une approche d'extension de l'ACP aux données intervalles y est présentée et étudiée en détail.

Le troisième chapitre traite de la détection et de la localisation de défauts, à l'aide de l'ACP classique et de l'ACP symbolique.

Le quatrième chapitre présente une application de ces techniques au diagnostic prédictif de défauts sur un cas industriel réel, à travers un benchmark de turbine à gaz.

Généralités sur le diagnostic

1.1 Introduction

Depuis les années 1970, le diagnostic des défauts connaît un intérêt croissant, aussi bien dans l'industrie que dans la recherche scientifique. Il vise à concevoir des systèmes capables d'assurer une maintenance intelligente, voire prédictive, à travers des générateurs de résidus robustes permettant de détecter et localiser les dysfonctionnements.

La complexité croissante des systèmes de commande, liée à la fois à leur nature et à la nécessité de maîtriser l'ensemble des paramètres et perturbations, a favorisé l'essor de la sûreté de fonctionnement. Il devient ainsi indispensable de surveiller les organes de conduite (actionneurs, capteurs, instrumentation) afin de détecter et d'isoler rapidement toute anomalie.

Cette démarche, connue sous le nom de FDI (Fault Detection and Isolation), consiste à identifier l'apparition d'un défaut et à en déterminer l'origine le plus rapidement possible. Elle a fait l'objet de nombreux travaux de recherche proposant des approches variées, adaptées à la diversité des applications industrielles.

Ce chapitre présente les notions fondamentales du diagnostic des systèmes, les différents types de défauts ainsi que les principales méthodes de diagnostic utilisées.

1.2 Objectifs de la supervision

La supervision correspond à l'intégration des fonctions de surveillance, de diagnostic et de contrôle de la performance et de l'exécution des opérations au sein d'un système industriel. Elle permet d'assurer le suivi du comportement du système en temps réel et de garantir son bon fonctionnement [1].

Le processus de supervision peut être réalisé par un opérateur humain ou par un système automatisé. Il repose sur la collecte et l'analyse des données issues du procédé, telles que les mesures, les alarmes et les indicateurs de performance. Lorsqu'une anomalie est détectée, des actions correctives sont mises en œuvre afin de rétablir le système dans un état normal de

fonctionnement.

L'intégration d'un système de supervision dans les systèmes automatisés de production vise principalement à [1] :

- optimiser les coûts et maximiser la production ;
- améliorer la qualité et la rentabilité des produits ;
- assurer la disponibilité des services et une meilleure utilisation des ressources ;
- anticiper les variations du système et prévenir les risques d'incidents ou d'accidents.

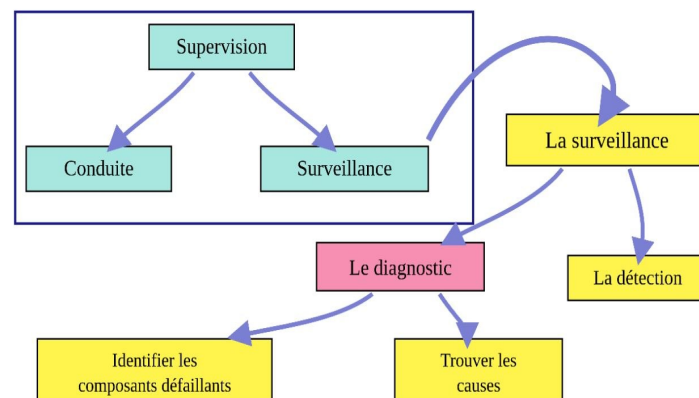


FIGURE 1.1 – La supervision et la surveillance

1.3 Terminologie

Dans le domaine du diagnostic des systèmes industriels, plusieurs notions fondamentales sont utilisées pour décrire l'état, le comportement et les dysfonctionnements d'un système. Les principales définitions sont présentées ci-dessous.

Système Un système est un ensemble d'éléments interconnectés et organisés, interagissant entre eux afin de réaliser une fonction ou un objectif déterminé.

Défaut Un défaut est défini comme tout écart entre la caractéristique observée d'un système et la caractéristique nominale ou attendue. Cet écart est théoriquement nul en fonctionnement normal. Les défauts peuvent apparaître au niveau des capteurs, des actionneurs ou du processus lui-même [2].

Défaillance Une défaillance correspond à l'altération ou à la perte de la capacité d'un système à accomplir une ou plusieurs fonctions requises conformément aux performances définies dans les spécifications techniques. Elle se traduit par un dysfonctionnement du système conduisant à un comportement jugé inacceptable du point de vue des performances.

L'apparition d'une défaillance implique généralement l'existence d'un défaut, puisqu'un écart est observé entre le comportement réel du système et son comportement théorique. En revanche, la présence d'un défaut n'entraîne pas nécessairement une défaillance, dans la mesure où le dispositif peut continuer à assurer correctement sa fonction principale malgré cet écart [2].

Panne Une panne est l'incapacité d'un système à assurer sa fonction requise. Elle résulte généralement d'une défaillance non corrigée ou aggravée au cours du temps [3].

On peut ainsi résumer la progression des états de dégradation comme suit :



FIGURE 1.2 – La cause principale d'une panne

Fonctionnement normal d'un système Un système est dit en fonctionnement normal lorsque ses variables caractéristiques (variables d'état, variables d'entrée, variables de sortie et paramètres du système) évoluent au voisinage de leurs valeurs nominales, sans écarts significatifs [2].

1.4 Système de surveillance et/ou de diagnostic

Le système de surveillance et/ou de diagnostic, connu sous l'appellation *Fault Detection and Isolation* (FDI), repose sur un ensemble de fonctions essentielles permettant d'analyser l'état d'un procédé et de détecter d'éventuelles anomalies. Il est généralement structuré autour de trois étapes principales : la détection, la localisation et l'identification des défauts [4].

La **détection** consiste à mettre en évidence la présence d'un comportement anormal au sein du système. Elle repose sur la comparaison entre le comportement réel du procédé et un comportement de référence. Ce dernier peut être obtenu à partir d'un modèle analytique ou numérique. Toutefois, les écarts observés peuvent être dus non seulement à des défauts, mais également à des perturbations externes ou à des incertitudes de modélisation. Ainsi, la détection vise à distinguer les véritables défauts des simples variations tolérables du système, afin de générer des alarmes fiables et pertinentes [4].

La **localisation**, également appelée isolation des défauts, intervient une fois l'anomalie détectée. Elle a pour objectif d'identifier précisément le composant, le sous-système ou la zone du procédé à l'origine du dysfonctionnement. Cette étape s'appuie sur l'analyse des relations entre les variables du système ainsi que sur l'interprétation des symptômes observés. Elle peut s'avérer complexe, notamment dans le cas où plusieurs défauts interagissent simultanément ou lorsqu'un seul défaut engendre plusieurs effets visibles [4].

L'**identification** constitue la dernière étape du processus de diagnostic. Elle consiste à caractériser le défaut détecté en déterminant sa nature, son intensité ainsi que son évolution. Cette étape permet de quantifier le niveau de dégradation du système et fournit des informations essentielles pour la prise de décision, notamment en matière de maintenance, de reconfiguration ou d'adaptation des stratégies de commande.

1.5 Typologie des défauts

Un défaut correspond à un écart non toléré entre le comportement réel d'un système et son comportement attendu ou nominal. Ces anomalies peuvent apparaître à différents niveaux d'un système automatisé et sont généralement classées en trois catégories principales : les défauts capteurs, les défauts actionneurs et les défauts composants (ou système), comme illustré dans la figure 1.3.

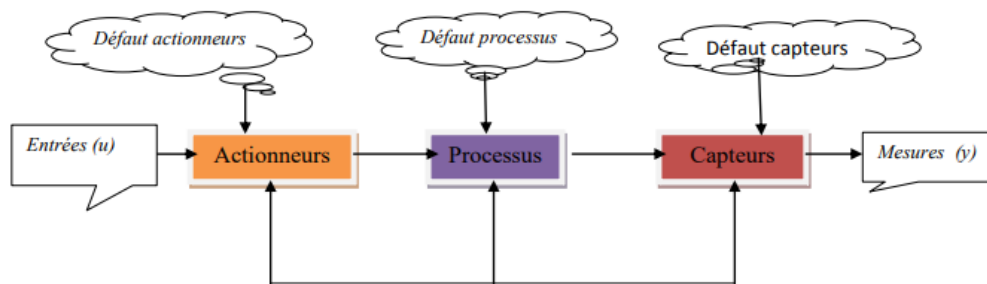


FIGURE 1.3 – Typologie de défauts

Défauts capteurs

Un capteur est un dispositif qui convertit une grandeur physique issue de l'environnement en un signal exploitable par le système de commande. Il joue le rôle d'interface entre le processus physique et l'unité de traitement de l'information [5].

Un défaut de capteur se manifeste lorsque la mesure fournie ne reflète plus correctement la réalité du phénomène observé. Cette dégradation de l'information peut entraîner une mauvaise estimation de l'état du système. Dans les systèmes en boucle fermée, cela conduit à une génération de commandes inadaptées, ce qui réduit la précision et la performance globale du contrôle.

Défauts actionneurs

Un actionneur est un élément de la partie opérative chargé de transformer l'énergie reçue en une action physique, telle qu'un mouvement ou une variation d'état. À titre d'exemple, un moteur électrique convertit l'énergie électrique en énergie mécanique [5].

Les défauts actionneurs affectent la partie opérative du système. Ils altèrent les signaux de commande et perturbent l'exécution correcte des actions sur le système [6].

Défauts composants (défauts système)

Les défauts composants correspondent aux anomalies qui touchent directement les éléments internes du système. Ils ne peuvent être associés ni aux défauts capteurs ni aux défauts actionneurs.

Ce type de défaillance résulte généralement d'une modification des paramètres internes du système, entraînant une dégradation de son fonctionnement global [5].

1.6 Caractérisation des défauts

Les défauts peuvent être caractérisés selon leur évolution dans le temps. Cette classification repose sur leur comportement temporel et permet de distinguer plusieurs catégories, notamment les défauts brusques, intermittents, à dérive lente, ainsi que les défauts additifs et multiplicatifs [7].

Défaut brusque : Un défaut brusque se caractérise par une apparition soudaine et une variation instantanée de son amplitude. Il provoque une modification immédiate du comportement du système et est généralement causé par une rupture ou une défaillance brutale d'un composant [8].

Défaut intermittent : Un défaut intermittent apparaît et disparaît de manière irrégulière au cours du temps. Ce type de défaut est souvent lié à des mauvais contacts électriques, à des vibrations mécaniques ou à des conditions de fonctionnement particulières. Sa nature aléatoire rend sa détection et sa localisation particulièrement difficiles [9].

Défaut à dérive lente : Un défaut à dérive lente, également appelé défaut progressif, se manifeste par une augmentation graduelle de son amplitude au cours du temps. Il est généralement associé à un phénomène de vieillissement, d'usure ou de dégradation progressive d'un composant. En raison de son évolution lente, il est souvent difficile à distinguer des variations normales ou des incertitudes de mesure.

Défaut additif : Un défaut additif agit en ajoutant un terme supplémentaire à une variable du système, telle qu'une mesure ou une entrée. Il est fréquemment utilisé pour modéliser les défauts de capteurs ou les perturbations externes.

Défaut multiplicatif : Un défaut multiplicatif modifie le gain ou un paramètre du système en multipliant une variable par un facteur. Il est généralement associé à une variation des paramètres du procédé ou à une perte d'efficacité d'un actionneur.

1.7 Les étapes du diagnostic

Le diagnostic d'un système repose sur une succession d'étapes permettant d'analyser son comportement, de détecter d'éventuelles anomalies, de les localiser, puis de caractériser les défauts afin de prendre les décisions appropriées.

Acquisition des données : Cette phase consiste à recueillir les informations nécessaires au suivi du système. Elle comprend notamment le prétraitement des signaux, leur conditionnement, ainsi que la vérification de leur cohérence, afin d'assurer la qualité des mesures utilisées pour les étapes suivantes.

Construction des indicateurs de défaut : À partir des données collectées, des variables caractéristiques (résidus, indices ou signatures) sont élaborées afin de mettre en évidence d'éventuels écarts par rapport au fonctionnement normal du système.

Détection : Cette étape vise à déterminer si le comportement du système s'écarte de son fonctionnement nominal. Elle repose sur la comparaison entre les mesures observées et un modèle de référence (ou des seuils prédéfinis). Lorsqu'un écart significatif est constaté, une alerte est générée.

Localisation : Une fois l'anomalie détectée, cette étape permet de déterminer l'origine du défaut au sein du système, qu'il s'agisse d'un capteur, d'un actionneur, d'un organe de commande ou du procédé lui-même.

Identification : Cette étape consiste à caractériser le défaut localisé, notamment en déterminant son type (brusque, graduel, intermittent, etc.), son amplitude et son instant d'apparition, afin d'évaluer sa gravité et son impact sur le système.

Prise de décision : Cette dernière phase consiste à définir les actions appropriées pour maintenir ou rétablir le bon fonctionnement du système, telles que la reconfiguration de la commande, la planification d'une maintenance ou l'intervention d'un opérateur.

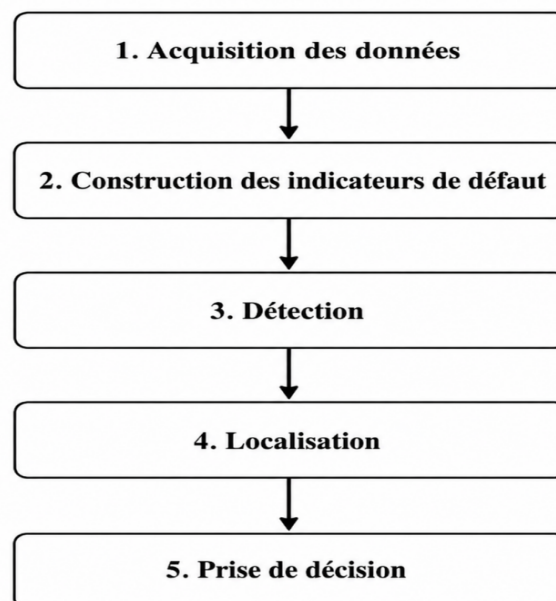


FIGURE 1.4 – Les étapes du diagnostic.

1.8 Méthodes de diagnostic

Une bonne compréhension de la dynamique des processus constitue un élément essentiel pour assurer une interprétation fiable du fonctionnement d'un système industriel. Pour cela, il est

nécessaire de disposer d'un modèle représentant le comportement normal du procédé, afin de pouvoir détecter les éventuelles anomalies par comparaison avec le comportement observé.

Plusieurs approches ont été développées dans le domaine du diagnostic des systèmes. De manière générale, ces méthodes peuvent être regroupées en trois grandes catégories : les approches analytiques, les approches à base de connaissances et les approches à base d'analyse de données [10].

Approches analytiques : Les approches analytiques reposent sur une modélisation mathématique du système étudié. Le procédé est décrit à l'aide d'équations issues des lois physiques, chimiques ou mécaniques, selon la nature du système considéré [11]. La détection des défauts s'effectue alors par génération de résidus, obtenus notamment par des méthodes à base d'observateurs, d'espace de parité ou d'estimation paramétrique, qui traduisent l'écart entre le comportement réel du système et celui prédit par le modèle. Ces approches offrent l'avantage d'une bonne précision et d'une interprétation physique directe des défauts détectés ; toutefois, leur mise en œuvre nécessite une connaissance approfondie du système et devient difficile, voire impossible, lorsque le procédé est complexe, non linéaire ou mal connu.

Approches à base de connaissances : Dans certaines applications industrielles, l'élaboration d'un modèle analytique précis peut s'avérer difficile en raison de la complexité structurelle du système ou de son comportement dynamique. Afin de surmonter cette difficulté, des approches à base de connaissances ont été développées.

Ces méthodes exploitent l'expérience de l'expert ainsi que des règles de raisonnement (systèmes experts, arbres de défaillance, graphes causaux) pour représenter le comportement du système et orienter le diagnostic. Elles offrent une plus grande flexibilité face aux incertitudes et aux perturbations, puisqu'elles ne dépendent pas directement des équations physiques du procédé. Toutefois, leur efficacité reste fortement liée à la qualité et à l'exhaustivité des connaissances disponibles sur le système, et leur élaboration peut s'avérer longue et coûteuse [11].

Approches à base d'analyse de données : Les approches à base de données s'appuient sur les informations fournies par les capteurs afin d'analyser le comportement du système. Les données acquises durant le fonctionnement du procédé sont enregistrées puis exploitées pour construire des modèles de surveillance et de diagnostic, sans nécessiter de connaissance explicite des relations physiques régissant le système.

Ces méthodes sont particulièrement adaptées aux systèmes complexes pour lesquels la modélisation analytique est difficile, voire impossible. Elles reposent généralement sur l'analyse statistique et les techniques d'apprentissage automatique appliquées à un historique de données représentatif du fonctionnement normal et, le cas échéant, défaillant du système. Parmi les méthodes les plus utilisées, l'Analyse en Composantes Principales (ACP) occupe une place importante pour l'extraction des caractéristiques représentatives du comportement du processus et la réduction de la dimensionnalité des données [11]. D'autres techniques, telles que la régression aux moindres carrés partiels (PLS), les réseaux de neurones artificiels ou les méthodes d'apprentissage automatique, sont également largement employées selon la nature du problème

traité.

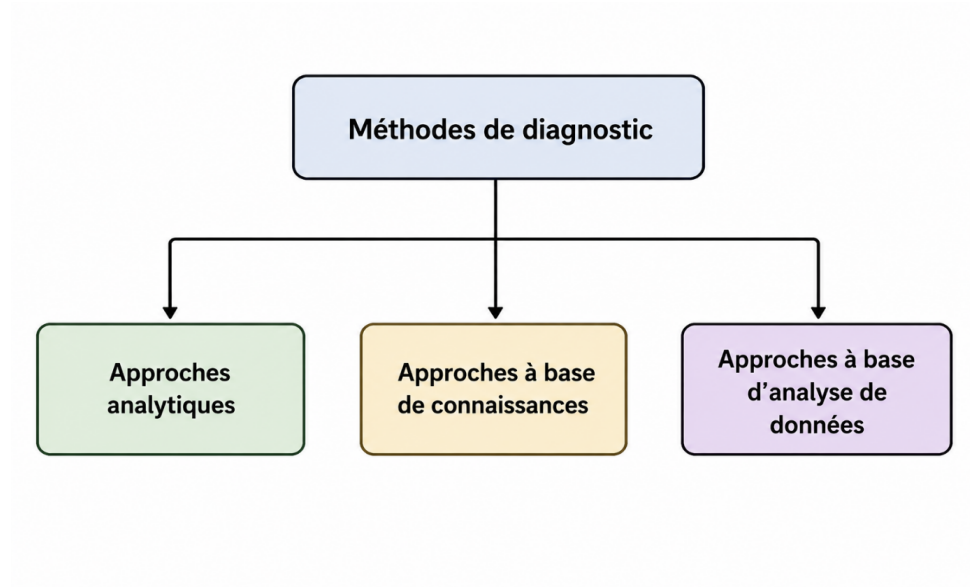


FIGURE 1.5 – Méthodes de diagnostic.

1.9 Critères de performance d'un système de diagnostic

Dans la littérature, plusieurs critères ont été proposés afin d'évaluer les performances d'un système de diagnostic [12]. Une méthode de diagnostic efficace doit répondre aux exigences suivantes :

- **Rapidité de détection** : le système de diagnostic doit être capable de détecter un défaut le plus rapidement possible après son apparition, afin de limiter sa propagation et ses conséquences sur le fonctionnement du système. Un compromis doit néanmoins être trouvé entre rapidité de détection et robustesse, une détection trop rapide pouvant favoriser les fausses alarmes.
- **Sensibilité** : la méthode doit être suffisamment sensible pour détecter des défauts de faible amplitude, y compris à un stade précoce de leur évolution, avant qu'ils n'entraînent des dégradations importantes ou des pannes critiques.
- **Robustesse** : le système de diagnostic doit être capable de distinguer les effets d'un défaut de ceux dus aux bruits de mesure, aux perturbations externes ou aux incertitudes de modélisation, afin de limiter les fausses alarmes et les non-détections.
- **Isolabilité (localisation)** : la méthode doit permettre de localiser avec précision l'élément défaillant (capteur, actionneur, organe de commande ou procédé), et non seulement de signaler la présence d'une anomalie.
- **Identifiabilité de défauts multiples** : le système doit, dans la mesure du possible, être capable de détecter et de distinguer plusieurs défauts survenant simultanément ou de

manière successive.

- **Facilité d'implémentation et fonctionnement en temps réel** : la méthode doit pouvoir être traduite sous forme algorithmique simple, avec un temps de calcul compatible avec une surveillance en temps réel ou quasi temps réel, afin d'être intégrée dans des systèmes de supervision automatisés.

1.10 Conclusion

Dans ce chapitre, les concepts fondamentaux du diagnostic des systèmes ont été présentés. Nous avons tout d'abord introduit les différents types de défauts ainsi que leurs principales caractéristiques, avant de détailler les étapes constitutives d'un processus de diagnostic, de l'acquisition des données jusqu'à la prise de décision. Les trois grandes familles d'approches utilisées pour le diagnostic des systèmes, à savoir les méthodes analytiques, les approches à base de connaissances et celles fondées sur l'analyse de données, ont ensuite été présentées et comparées, en mettant en évidence leurs principes, leurs avantages ainsi que leurs limites respectives. Enfin, les principaux critères permettant d'évaluer les performances d'un système de diagnostic ont été énoncés.

L'ensemble de ces notions constitue une base théorique essentielle pour la compréhension des techniques modernes de surveillance et de détection des défauts. Elles serviront de fondement au chapitre suivant, dans lequel la méthode d'Analyse en Composantes Principales sera présentée en détail et appliquée aux problèmes de prédiction de défauts.

L'Analyse en Composantes Principales pour la modélisation statistique des données

2.1 Introduction

Il existe de nombreuses méthodes de diagnostic permettant d'éprouver la cohérence des mesures. Elles peuvent être classées en deux catégories, selon qu'elles exploitent ou non la connaissance d'un modèle a priori du système.

Les méthodes dites sans modèle relèvent des techniques de traitement des données. L'Analyse en Composantes Principales (ACP), bien qu'initialement destinée à la compression de données, constitue également un outil intéressant pour détecter et localiser des erreurs de mesure ainsi que des dysfonctionnements de procédés [14].

Elle a été introduite en 1901 par Karl Pearson, puis formalisée mathématiquement dans les années 1930 par Harold Hotelling. Dans les années 1960, le statisticien français Jean-Paul Benzécri a poursuivi le développement de cette méthode en s'appuyant sur des représentations géométriques et graphiques, dans le cadre de ce qui deviendra l'analyse des données à la française. Il s'agit à l'origine d'une approche descriptive, qui ne repose pas sur des hypothèses probabilistes.

À partir des années 1970, l'ACP a commencé à être utilisée pour modéliser le fonctionnement normal des processus industriels. À partir des données disponibles en fonctionnement sain, elle permet de construire un modèle de référence qui sert ensuite à la surveillance des variables importantes du système.

Aujourd'hui, l'ACP est utilisée dans de nombreux domaines : biologie, économie, sciences sociales, traitement d'images, ainsi que surveillance et diagnostic des défauts.

2.2 Analyse en Composantes Principales

L'Analyse en Composantes Principales (ACP) est une méthode statistique descriptive utilisée pour mieux comprendre les relations entre les variables ainsi que les similarités entre les

individus d'un ensemble de données. Elle appartient aux techniques de réduction de dimension et fonctionne généralement de manière non supervisée. Son principe consiste à transformer un grand nombre de variables souvent corrélées en un nombre plus réduit de nouvelles variables synthétiques, tout en conservant l'essentiel de l'information contenue dans les données initiales. Ces nouvelles variables sont appelées composantes principales et résultent de combinaisons linéaires des variables d'origine.

L'objectif principal de l'ACP est d'analyser les liens linéaires existant entre les variables d'un système. Lorsque plusieurs variables sont fortement corrélées, cela signifie qu'elles apportent des informations similaires, donc redondantes. L'ACP permet alors de simplifier l'étude du système en construisant de nouvelles variables, non corrélées entre elles, capables de résumer au mieux la variabilité totale des données.

Les composantes principales ainsi obtenues servent également à construire des plans factoriels, qui facilitent la représentation graphique et l'interprétation des données initiales. Ces représentations permettent de visualiser plus aisément les relations entre les variables ainsi que d'identifier d'éventuelles structures ou comportements particuliers au sein du système de données étudié.

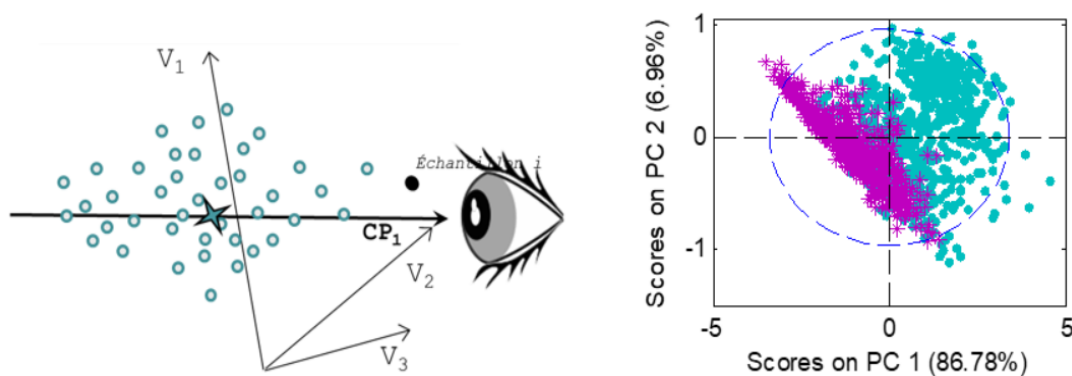


FIGURE 2.1 – L'idée de base de la méthode ACP

2.2.1 Principe de la méthode ACP

En pratique, pour modéliser un processus à l'aide de l'analyse en composantes principales, on commence par collecter des mesures des différentes variables du processus lorsqu'il fonctionne normalement. Toutes ces données sont ensuite regroupées dans une matrice X^b .

Soit m le nombre de variables mesurées et N le nombre d'observations enregistrées pour chaque variable :

$$X^b \in \mathbb{R}^{N \times m}$$

donnée par :

$$X^b = \begin{pmatrix} x_1(1) & x_2(1) & \cdots & x_m(1) \\ x_1(2) & x_2(2) & \cdots & x_m(2) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(N) & x_2(N) & \cdots & x_m(N) \end{pmatrix} \quad (2.1)$$

où $x_1(1)$ représente la mesure de la première variable au premier instant de mesure.

Avant d'appliquer l'ACP, il est nécessaire d'effectuer un prétraitement afin d'éviter que les résultats ne dépendent des unités utilisées. Ce prétraitement consiste à centrer et réduire les variables.

Chaque colonne X_j de la nouvelle matrice centrée réduite est donnée par :

$$X_j = \frac{X_j^b - M_j}{\sigma_j} \quad (2.2)$$

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

$$M_j = \frac{1}{N} \sum_{k=1}^N X_j^b(k) \quad (2.3)$$

et σ_j est son écart-type, avec σ_j^2 la variance estimée à l'aide de l'équation :

$$\sigma_j^2 = \frac{1}{N-1} \sum_{k=1}^N (X_j^b(k) - M_j)^2 \quad (2.4)$$

La nouvelle matrice des données normalisées (centrées réduites) est :

$$X = [X_1 \quad \cdots \quad X_m]$$

Après normalisation de la matrice, on calcule la matrice de corrélation de X^b , qui correspond, du fait de la normalisation préalable des données, à la matrice de covariance de X . Cette matrice met en évidence les relations linéaires existant entre les variables et est donnée par :

$$\Sigma = \frac{1}{N-1} X^T X \quad (2.5)$$

La matrice de corrélation étant obtenue, on peut à présent passer à l'identification du modèle ACP. Cette étape consiste à diagonaliser la matrice Σ , c'est-à-dire à la décomposer pour obtenir ses valeurs propres et ses vecteurs propres associés. Ces vecteurs propres définissent les nouvelles directions principales sur lesquelles les données seront projetées afin de réduire leur dimension :

$$\Sigma = P \Lambda P^T = \sum_{j=1}^m \lambda_j P_j P_j^T \quad (2.6)$$

avec :

$$PP^T = P^T P = I_m \quad (2.7)$$

tel que :

$$P = [P_1, P_2, \dots, P_m] \in \mathbb{R}^{m \times m}$$

est la matrice des vecteurs propres, et :

$$\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_m) \in \mathbb{R}^{m \times m}$$

est la matrice diagonale des valeurs propres, classées par ordre décroissant :

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$$

La projection d'un vecteur d'observations $x(k) \in \mathbb{R}^m$ sur les vecteurs propres donne le vecteur des composantes principales (CPs) correspondant, exprimé par l'équation suivante :

$$t(k) = P^T x(k) \quad (2.8)$$

Pour reconstruire une estimation du vecteur de mesures $x(k)$ à partir de son vecteur de composantes principales $t(k)$, il suffit d'utiliser la relation suivante :

$$x(k) = P t(k) = \sum_{j=1}^m t_j(k) P_j \quad (2.9)$$

soit :

$$x(k) = \hat{x}(k) + \tilde{x}(k) \quad (2.10)$$

$$x(k) = \hat{P} \hat{t}(k) + \tilde{P} \tilde{t}(k) \quad (2.11)$$

Le vecteur de mesures $x(k)$ peut ainsi être décomposé en deux parties : un vecteur $\hat{x}(k)$, qui représente les mesures estimées (appelé projection sur le sous-espace principal), et un vecteur $\tilde{x}(k)$, qui représente les erreurs ou écarts résiduels (appelé projection sur le sous-espace résiduel).

2.3 Nombre de composantes principales

Le choix du nombre de composantes principales (CP) à conserver constitue l'étape la plus importante lors de la construction d'un modèle par ACP. En effet, l'intérêt de la méthode est de représenter les données de manière plus simple, en les projetant dans un espace réduit, tout en conservant le maximum d'information.

Les vecteurs propres indiquent les directions dans lesquelles la variance des données est la plus grande ; ce sont les axes factoriels du sous-espace retenu. En projetant les données sur ces axes, on obtient les composantes principales t_j , dont la variance correspond directement à la

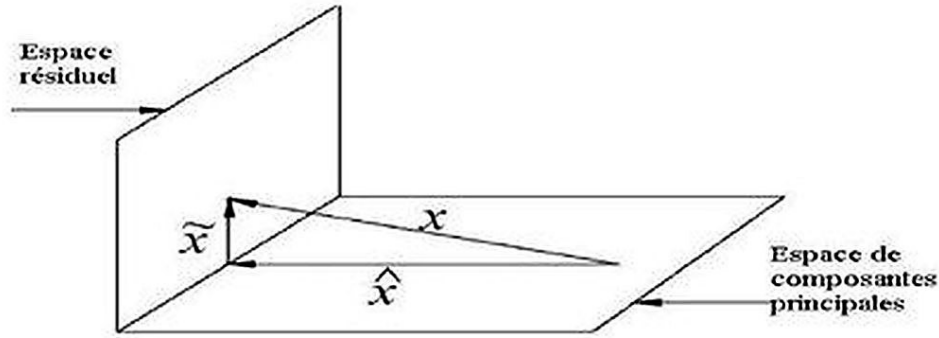


FIGURE 2.2 – Décomposition du vecteur de mesures

valeur propre associée.

Le premier vecteur propre p_1 , associé à la plus grande valeur propre λ_1 , représente la direction la plus informative. À l'inverse, le vecteur propre p_m , lié à la plus petite valeur propre λ_m , correspond à la direction contenant le moins d'information [13].

Pour réduire la dimension des données, on conserve les premières composantes principales, celles qui présentent la plus grande variance, car ce sont elles qui contiennent le plus d'information.

L'espace réduit ainsi obtenu, de dimension l (avec $l \leq m$), est donc construit à partir des l vecteurs propres correspondant aux l plus grandes valeurs propres de la matrice de covariance Σ . Ces vecteurs représentent les directions les plus pertinentes pour décrire la structure des données.

2.3.1 Pourcentage cumulé de la variance totale (PCV)

Cette méthode de sélection du nombre de composantes principales repose sur le fait que chaque composante représente une part de la variance totale des données. Les valeurs propres de la matrice de corrélation Σ indiquent la contribution de chaque composante à cette variance et permettent ainsi de déterminer le nombre de composantes à retenir. En pratique, on retient généralement un nombre l de composantes suffisant pour atteindre un pourcentage cumulé de variance compris entre 90% et 95%, afin de conserver l'essentiel de l'information tout en réduisant la dimension des données [13].

Le pourcentage cumulé de variance expliquée par les l premières composantes principales est donné par :

$$PCV(l) = 100 \times \left(\frac{\sum_{i=1}^l \lambda_i}{\sum_{i=1}^m \lambda_i} \right) \quad (2.12)$$

Le nombre de composantes l est alors choisi comme la plus petite valeur telle que $PCV(l)$ dépasse le seuil fixé (généralement 90% à 95%).

2.4 Exemple de simulation

Pour mieux illustrer la méthode, nous présentons un exemple de simulation dans lequel les différentes étapes précédentes de construction du modèle ACP sont appliquées.

Nous disposons de 4 variables représentant un système statique, décrites par les équations suivantes :

$$\begin{cases} x_1(t) = u_1(t) + \varepsilon_1 \\ x_2(t) = u_1(t) + \varepsilon_2 \\ x_3(t) = x_1(t) + x_2(t) + 0.115 \cdot \varepsilon_3 \\ x_4(t) = x_1(t) + x_3(t) + 0.15 \cdot \varepsilon_4 \end{cases} \quad (2.13)$$

où :

- les ε_i sont des bruits blancs ;
- u_1 est un signal sous forme de créneaux, de taille 1000 mesures.

La figure 2.3 représente l'évolution de ces données.

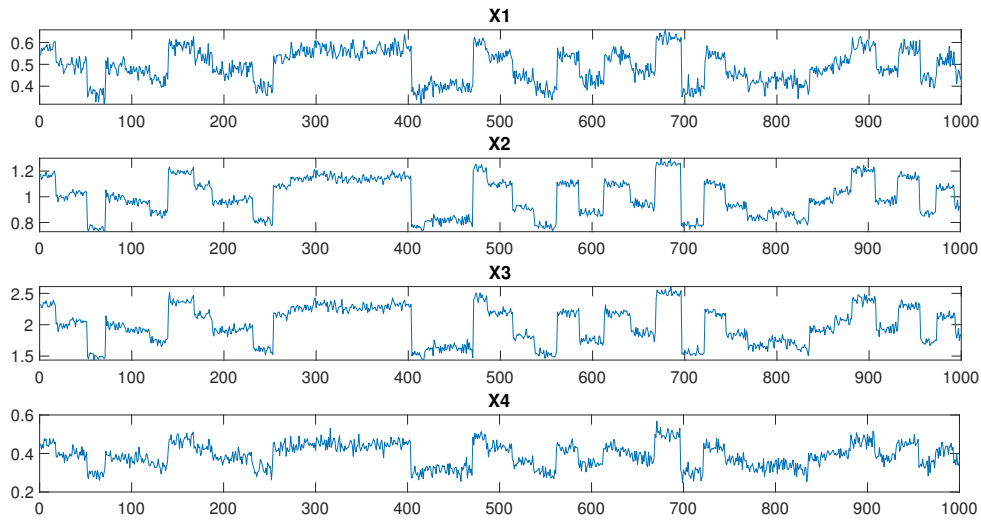


FIGURE 2.3 – Évolution des différentes variables

Lors de la normalisation, le vecteur des moyennes ainsi que le vecteur des écarts-types des variables sont donnés par :

$$\text{Mean} = \begin{pmatrix} 0.4922 & 1.0022 & 1.9939 & 0.3916 \end{pmatrix}, \quad \sigma = \begin{pmatrix} 0.0745 & 0.1447 & 1.2910 & 0.0611 \end{pmatrix}$$

La matrice de corrélation obtenue est :

$$\Sigma = \begin{pmatrix} 1.0000 & 0.9653 & 0.9630 & 0.9089 \\ 0.9653 & 1.0000 & 0.9977 & 0.9423 \\ 0.9630 & 0.9977 & 1.0000 & 0.9398 \\ 0.9089 & 0.9423 & 0.9398 & 1.0000 \end{pmatrix}$$

Après diagonalisation, les valeurs propres et les vecteurs propres associés, classés par ordre décroissant, sont donnés par :

$$\Lambda = \text{diag}(3.8591, 0.0963, 0.0423, 0.0023)$$

$$P = \begin{pmatrix} 0.4972 & 0.5100 & 0.7014 & -0.0252 \\ 0.5061 & 0.1490 & -0.4410 & 0.7261 \\ 0.5055 & 0.1577 & -0.4977 & -0.6869 \\ 0.4910 & -0.8324 & 0.2566 & -0.0157 \end{pmatrix}$$

L'évolution des différentes variables après normalisation est présentée dans la figure 2.4.

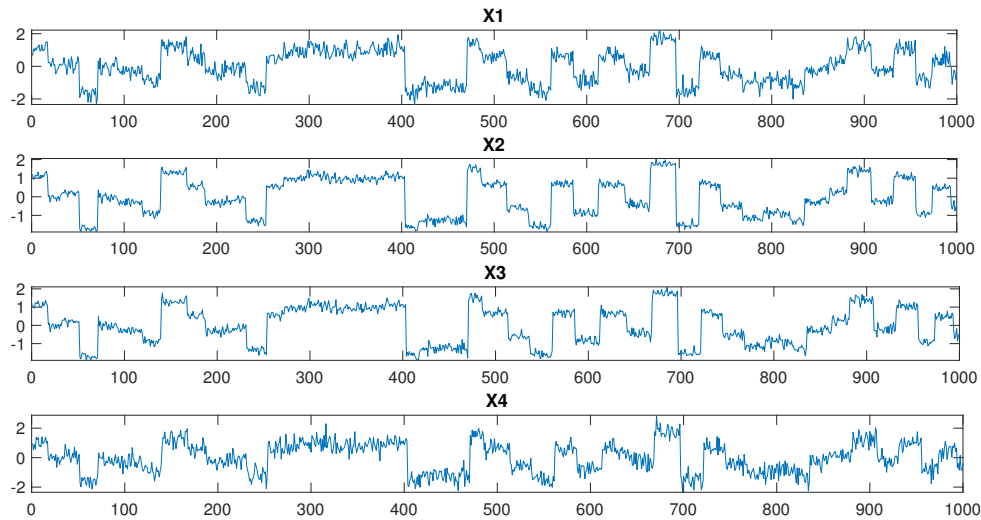


FIGURE 2.4 – Évolution des variables normalisées

Les composantes principales T sont obtenues par projection de la matrice des données centrées réduites X (normalisées) sur la matrice des vecteurs propres P . Elles sont représentées sur la figure 2.5.

On observe que la première composante principale concentre l'essentiel de l'information utile, tandis que les composantes suivantes traduisent principalement le bruit de mesure. Le critère du pourcentage cumulé de la variance (PCV) a été appliqué pour confirmer ce choix : il montre que l'essentiel de l'information pertinente est porté par ce premier axe ($l = 1$), retenu comme composante principale du modèle.

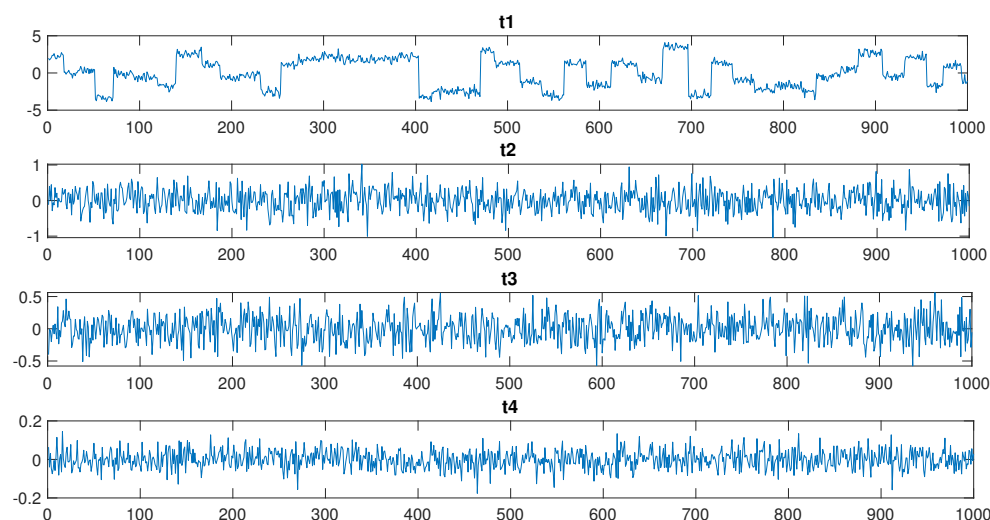


FIGURE 2.5 – Évolution des composantes principales

La figure 2.6 illustre la variabilité des données ainsi que leur estimation obtenue à partir de ce modèle.

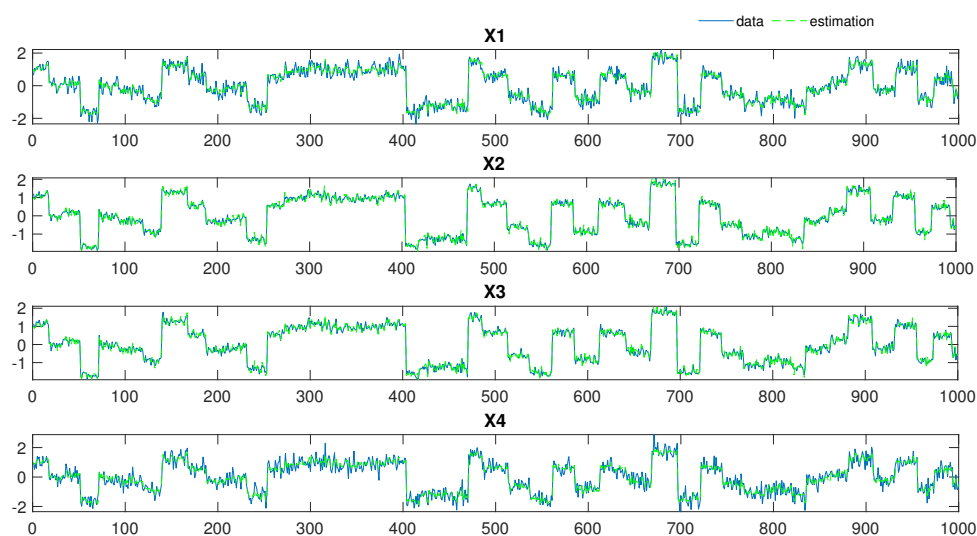


FIGURE 2.6 – Évolution des mesures et de leurs estimations

D'après la figure 2.6, on remarque que les estimations reproduisent fidèlement les données initiales : l'erreur d'estimation sur l'ensemble des variables reste quasiment nulle. Cela signifie que le modèle ACP, construit avec une seule composante principale, fournit une estimation satisfaisante des différentes mesures.

2.5 Le problème des grandeurs d'influence et de l'imprécision des mesures

Dans tout processus d'analyse de données, l'acquisition des observations constitue une étape préalable et déterminante pour garantir la fiabilité des résultats obtenus. Une grandeur physique m que l'on cherche à caractériser n'est jamais directement accessible : elle est perçue au travers d'un système de mesure (ou capteur) qui délivre une valeur mesurée Y . Cette valeur résulte d'une combinaison entre la grandeur physique m et un ensemble de grandeurs d'influence, inhérentes au dispositif de mesure ou aux conditions environnementales dans lesquelles s'effectue l'acquisition, soit :

$$Y = f(m, e_1, e_2, e_3) \quad (2.1)$$

où e_1, e_2, e_3 désignent des grandeurs de perturbation.

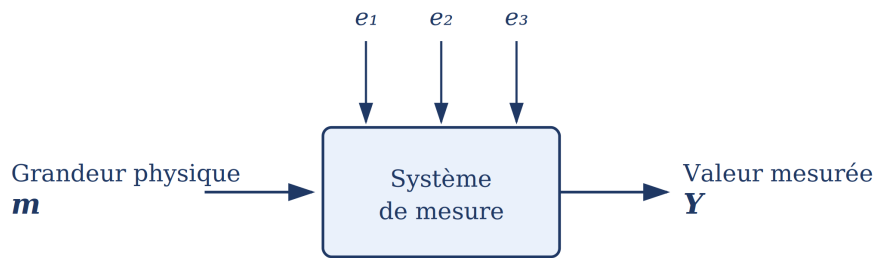


FIGURE 2.7 – Schéma du système de mesure

Cette relation met en évidence que toute mesure est, par nature, entachée d'une certaine imprécision, si bien que la valeur mesurée peut s'écrire sous la forme :

$$y = y^* \pm \Delta y \quad (2.2)$$

où y est la valeur délivrée par le capteur, y^* est la valeur vraie de la grandeur physique (non observable), et Δy représente l'imprécision liée au capteur (donnée par le constructeur).

Par conséquent, les données collectées ne peuvent représenter parfaitement la réalité physique du système observé. La présence d'erreurs de mesure, de bruit ou de perturbations est susceptible d'altérer la qualité et la fiabilité des données acquises. Or, la qualité des données recueillies sur le fonctionnement du système conditionne directement la qualité de la connaissance (du modèle) obtenue sur ce système.

En effet, un modèle ne peut restituer fidèlement le comportement réel d'un système si les données d'entrée sont elles-mêmes entachées de biais ou d'imprécision, le modèle héritant alors de ce biais. Cette problématique soulève une question essentielle : comment prendre en compte ces données imprécises de manière rigoureuse, plutôt que de simplement les ignorer ?

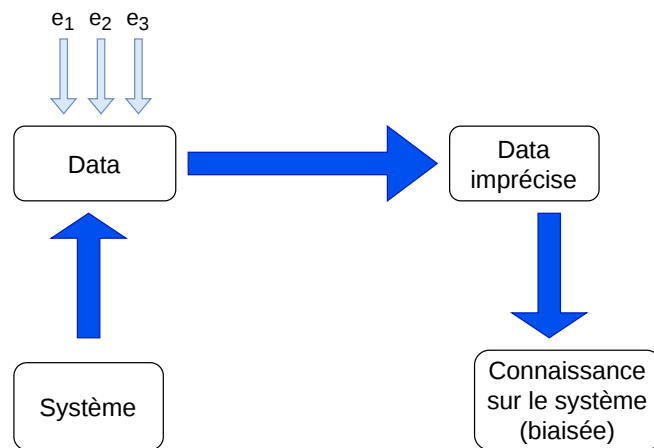


FIGURE 2.8 – Impact de la qualité des données sur la qualité de la connaissance

Face à cette problématique, deux approches méthodologiques peuvent être envisagées pour traiter l'imprécision des données. La première consiste à appliquer un prétraitement (filtrage) visant à atténuer l'effet des grandeurs d'influence, afin d'obtenir des données directement exploitables par les algorithmes de traitement classiques.

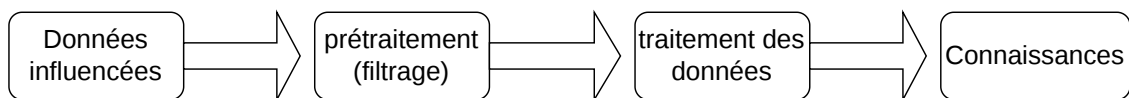


FIGURE 2.9 – Processus classique de traitement des données

Cette approche présente toutefois une limite majeure : le filtrage opère nécessairement une élimination partielle de l'information, ce qui peut conduire à un appauvrissement de la connaissance finalement extraite du système.

La seconde approche, plus récente, consiste non pas à éliminer l'imprécision, mais à la représenter explicitement au sein même du modèle de données. C'est précisément l'objet de l'Analyse de Données Symboliques (Symbolic Data Analysis, SDA).

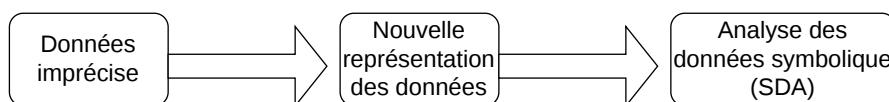


FIGURE 2.10 – Processus d'analyse de données symboliques pour les données imprécises

2.6 Analyse de Données Symboliques

L'analyse des données symboliques (ADS) constitue une extension de la statistique classique, introduite par Diday, dont l'objectif fondamental est de prendre en compte la complexité et l'imprécision inhérentes aux données réelles. Contrairement à l'approche statistique classique,

où chaque unité statistique est décrite par une valeur unique et précise, l'analyse des données symboliques associe à chaque observation un objet symbolique, pouvant prendre la forme d'un intervalle, d'un ensemble de valeurs ou d'une distribution de probabilité. Cette représentation enrichie permet de modéliser explicitement la variabilité naturelle des phénomènes observés ainsi que les incertitudes de mesure, sans perte d'information, offrant ainsi une description plus fidèle du comportement réel des systèmes étudiés [20, 21].

Parmi les différents types de données symboliques, les données de type intervalle sont particulièrement adaptées à la représentation de l'imprécision inhérente aux mesures et aux observations : elles permettent de capturer directement les variations et les incertitudes associées à chaque mesure. Ce type de données rend cependant leur traitement complexe pour les méthodes analytiques classiques telles que l'Analyse en Composantes Principales (ACP), qui, bien que puissante pour la réduction de dimensionnalité, se heurte à des limitations significatives lorsqu'elle est appliquée à des données imprécises : elle suppose en effet que chaque variable est décrite par une valeur ponctuelle unique, et ne permet donc pas de représenter la structure interne d'une donnée de type intervalle.

Pour répondre à ces limitations, l'ACP symbolique se présente comme une extension naturelle et nécessaire de l'ACP classique, spécifiquement adaptée à la complexité des données symboliques. Elle offre un cadre permettant de représenter et d'analyser les données de manière plus riche et flexible, en intégrant pleinement les aspects d'incertitude, tout en préservant la structure et la richesse de l'information contenue dans les données de type intervalle.

L'adoption de tels paradigmes de représentation et de techniques d'analyse adaptées, tels que l'ACP symbolique, permet ainsi de réduire l'incertitude et d'améliorer la fiabilité des résultats obtenus, conduisant à des conclusions plus robustes et à une meilleure prise de décision.

2.6.1 Données de type intervalle

La majorité des domaines scientifiques et industriels reposent aujourd'hui sur des données décrites par des variables mono-valuées, c'est-à-dire que chaque observation d'une variable, pour un individu donné, est résumée par une valeur unique. Ce mode de représentation s'avère toutefois insuffisant face à la complexité des systèmes réels. Les évolutions récentes des systèmes de bases de données ont permis de stocker de nouveaux types de données, tels que les intervalles, introduisant explicitement la notion d'imprécision ou de variation. Ces nouvelles structures ont ouvert la voie à de nombreuses applications capables de manipuler des objets décrits de manière plus fidèle et plus proche de la réalité.

Le résultat d'une mesure ou d'une observation peut ainsi être de type [18] :

- **Mono-valué** : le résultat d'une mesure ou d'une observation de la variable est une valeur unique. C'est le cas, par exemple, de l'âge ou de la taille d'une personne, ou encore des résultats d'un examen.
- **Intervalle** : le résultat d'une mesure ou d'une observation de la variable est un ensemble

continu de valeurs, borné par une valeur minimale et une valeur maximale, comme c'est le cas pour une température ou un taux cumulé. Tous les points compris entre ces deux bornes appartiennent à l'intervalle.

2.6.2 Formalisation matricielle des données de type intervalle

Soit $X \in \mathbb{R}^{n \times m}$ une matrice de données regroupant n observations de m variables, où $x_j(k)$ désigne la $k^{\text{ième}}$ mesure de la $j^{\text{ième}}$ variable. Chaque mesure étant entachée d'une incertitude dont la valeur exacte est inconnue, il est plus approprié de représenter cette observation par un intervalle réel plutôt que par une valeur unique.

On suppose que cette incertitude est bornée et peut être décrite par un intervalle de la forme $[\underline{x}_j(k), \bar{x}_j(k)]$, où $\underline{x}_j(k)$ représente la borne inférieure (**LB**, pour *Lower Bound*) et $\bar{x}_j(k)$ la borne supérieure (**UB**, pour *Upper Bound*) de la $k^{\text{ième}}$ observation de la $j^{\text{ième}}$ variable. Cette représentation par intervalle permet de tenir compte des incertitudes de mesure et de fournir des bornes garanties sur les valeurs estimées.

Ainsi, la $k^{\text{ième}}$ observation de la $j^{\text{ième}}$ variable est notée $[x_j(k)]$, et la matrice globale d'intervalles $[X]$, regroupant l'ensemble des observations pour les différentes variables de description, est formalisée par [19] :

$$[X] = \begin{pmatrix} [\underline{x}_1(1), \bar{x}_1(1)] & \dots & [\underline{x}_m(1), \bar{x}_m(1)] \\ \vdots & \ddots & \vdots \\ [\underline{x}_1(n), \bar{x}_1(n)] & \dots & [\underline{x}_m(n), \bar{x}_m(n)] \end{pmatrix} \quad (2.14)$$

2.6.3 Erreur de mesure et construction de l'intervalle

Toute mesure physique est entachée d'une erreur, définie comme l'écart entre la valeur mesurée et la valeur vraie (mais inconnue) de la grandeur d'intérêt. Cette erreur est donnée par :

$$\delta x_j(k) = x_j^c(k) - x_j^*(k) \quad (2.15)$$

où $x_j^c(k)$ est la valeur mesurée (le centre de l'intervalle) et $x_j^*(k)$ la valeur vraie, non observable, de la grandeur.

Le fabricant de l'instrument de mesure doit garantir une borne supérieure $\varepsilon_j(k)$ sur cette erreur, telle que $|\delta x_j(k)| \leq \varepsilon_j(k)$; en l'absence d'une telle garantie, aucune précision ne peut être assurée et l'instrument devient, en pratique, inexploitable.

Le rayon de l'intervalle est directement lié à cette borne d'erreur, tandis que son centre correspond à la valeur mesurée $x_j^c(k)$. L'intervalle est alors construit, sous la forme dite *centre-rayon*, par :

$$[x_j(k)] = [x_j^c(k) - \varepsilon_j(k), x_j^c(k) + \varepsilon_j(k)] \quad (2.16)$$

avec :

$$\underline{x}_j(k) = x_j^c(k) - \varepsilon_j(k) \quad \text{et} \quad \bar{x}_j(k) = x_j^c(k) + \varepsilon_j(k) \quad (2.17)$$

Le choix de la largeur de l'intervalle, c'est-à-dire de la borne d'erreur $\varepsilon_j(k)$, dépend directement de l'incertitude associée à l'instrument de mesure, et constitue une étape essentielle pour la construction de données de type intervalle représentatives de la réalité physique du système. En effet, ni la mesure ni son estimation ne sont jamais parfaites : la valeur vraie $x_j^*(k)$ peut différer de la valeur mesurée $x_j^c(k)$. Ce que l'on peut néanmoins garantir, à partir de la borne d'erreur fournie par le constructeur, est que la valeur vraie (inconnue) appartient nécessairement à l'intervalle ainsi construit :

$$x_j^*(k) \in [\underline{x}_j(k), \bar{x}_j(k)] \quad (2.18)$$

2.6.4 Arithmétique et statistique d'intervalle

À la suite des développements de l'analyse d'intervalles, les opérations arithmétiques et statistiques fondamentales ont été généralisées afin de prendre en compte les données à valeur d'intervalle. Les principales opérations sont présentées ci-dessous.

- **Objet à valeur intervalle** : Un objet à valeur intervalle est un sous-ensemble fermé et connexe de $\mathbb{R}$, défini par :

$$[x] = [\underline{x}, \bar{x}], \quad \underline{x} \leq \bar{x} \quad (2.19)$$

- **Somme** : La somme de deux intervalles $[x]$ et $[y]$ est obtenue en additionnant leurs bornes respectives :

$$[x] + [y] = [\underline{x} + \underline{y}, \bar{x} + \bar{y}] \quad (2.20)$$

- **Différence** : La différence entre les intervalles $[x]$ et $[y]$ est donnée par :

$$[x] - [y] = [\underline{x} - \bar{y}, \bar{x} - \underline{y}] \quad (2.21)$$

- **Produit** : Le produit des intervalles $[x]$ et $[y]$ est l'intervalle $[\min S, \max S]$, où :

$$S = \{\underline{x}\underline{y}, \underline{x}\bar{y}, \bar{x}\underline{y}, \bar{x}\bar{y}\} \quad (2.22)$$

est l'ensemble des produits des bornes.

- **Moyenne** : La moyenne d'un vecteur d'intervalles $[x_i]$ est définie par :

$$[m] = \frac{1}{n} \sum_{i=1}^n [x_i] = \left[\frac{1}{n} \sum_{i=1}^n \underline{x}_i, \frac{1}{n} \sum_{i=1}^n \bar{x}_i \right] \quad (2.23)$$

- **Centre de l'intervalle** : Le centre (ou milieu) x^c de l'intervalle $[x]$ est donné par :

$$x^c = \frac{\underline{x} + \bar{x}}{2} \quad (2.24)$$

— **Rayon de l'intervalle** : Le rayon x^r de l'intervalle $[x]$ est donné par :

$$x^r = \frac{\bar{x} - x}{2} \quad (2.25)$$

— **Distance** : La distance entre deux intervalles $[x]$ et $[y]$ est mesurée par la somme des écarts absolus entre leurs centres et leurs rayons respectifs :

$$d([x], [y]) = |x^c - y^c| + |x^r - y^r| \quad (2.26)$$

Voici votre texte corrigé. Attention : erreur mathématique dans la formule de la variance (2.27) — il manque le carré, l'expression telle qu'écrite calcule une moyenne nulle par définition et non une variance. latex

2.6.5 Normalisation des données de type intervalle

La normalisation est une étape essentielle avant le traitement des données, afin d'obtenir un résultat objectif et indépendant des unités ou des échelles de mesure des variables. Il existe plusieurs façons de normaliser des données de type intervalle ; nous avons choisi de nous appuyer sur la dispersion des centres des intervalles pour normaliser ce type de données.

La moyenne et la variance des centres des intervalles sont données par :

$$m_j = \frac{1}{n} \sum_{k=1}^n \frac{x_j(k) + \bar{x}_j(k)}{2} \quad (2.26)$$

$$\sigma_j^2 = \frac{1}{n} \sum_{k=1}^n \left(\frac{x_j(k) + \bar{x}_j(k)}{2} - m_j \right)^2 \quad (2.27)$$

et l'intervalle normalisé correspondant est donné par :

$$[x_j(k)]_{\text{norm}} = \left[\frac{x_j(k) - m_j}{\sigma_j}, \frac{\bar{x}_j(k) - m_j}{\sigma_j} \right] \quad (2.28)$$

où $\sigma_j = \sqrt{\sigma_j^2}$ est l'écart-type des centres des intervalles de la $j^{\text{ième}}$ variable.

2.7 ACP de type intervalle

L'Analyse en Composantes Principales (ACP) est depuis longtemps utilisée comme outil d'analyse exploratoire des données et de construction de modèles prédictifs. Ces dernières années ont vu l'émergence continue de données volumineuses, issues de diverses industries informatisées, dont le traitement se heurte à un problème fondamental : l'imprécision inhérente aux mesures qui composent ces données. Afin de discriminer un défaut d'une simple erreur

de modélisation, il devient nécessaire de définir une stratégie de diagnostic capable de prendre en compte des incertitudes susceptibles de fluctuer dans le temps et d'affecter n'importe quel paramètre du modèle, en les représentant sous forme de variables bornées.

Cette stratégie repose sur l'Analyse en Composantes Principales de type intervalle, qui constitue une généralisation de l'ACP classique aux données de type intervalle. Plutôt que de considérer chaque observation comme une valeur exacte, cette approche associe à chaque variable un intervalle de variation, ce qui permet de mieux représenter l'imprécision des données.

Plusieurs approches généralisant l'ACP aux données de type intervalle ont été proposées dans la littérature, parmi lesquelles :

- la méthode ACP des centres (*Center Points PCA*, CPCA);
- la méthode ACP à information complète (*Complete Information PCA*, CIPCA);
- la méthode ACP symbolique (*Symbolic PCA*, SPCA).

C'est cette dernière méthode, l'ACP Symbolique, que nous avons choisi d'étudier dans ce chapitre.

2.7.1 ACP Symbolique (SPCA)

Une méthode améliorée, fondée sur la covariance symbolique, a été introduite par Le-Rademacher et Billard en 2012; elle améliore encore les performances des méthodes précédentes d'ACP pour les données à valeurs d'intervalle. L'idée consiste à remplacer le calcul classique de la covariance par une nouvelle covariance, dite *symbolique*, qui constitue l'estimateur du maximum de vraisemblance de la variance globale d'une variable à valeurs d'intervalle. Ainsi, l'espace des composantes principales obtenu à partir de la matrice de covariance symbolique rend correctement compte de la variation totale des données à valeurs d'intervalle [19].

Soit $[X]$ la matrice de données intervalles définie en (2.14), et W_j un point de $[X_j]$, où $[X_j]$ ($j = 1, \dots, m$) désigne la j -ième variable intervalle de $[X]$. Les formules de la moyenne et de la variance empiriques de W_j sont respectivement données par :

$$\bar{W}_j = \sum_{k=1}^n \frac{\underline{x}_j(k) + \bar{x}_j(k)}{2n} \quad (2.29)$$

$$\text{Var}(j, j) = \sum_{k=1}^n \frac{\underline{x}_j^2(k) + \underline{x}_j(k)\bar{x}_j(k) + \bar{x}_j^2(k)}{3n} - \bar{W}_j^2 \quad (2.30)$$

La covariance empirique entre W_j et $W_{j'}$ est donnée par :

$$\begin{aligned} Cov(j, j') = \frac{1}{6n} \sum_{k=1}^n \left[2 (\underline{x}_j(k) - \bar{W}_j) (\underline{x}_{j'}(k) - \bar{W}_{j'}) \right. \\ + (\underline{x}_j(k) - \bar{W}_j) (\bar{x}_{j'}(k) - \bar{W}_{j'}) \\ + (\bar{x}_j(k) - \bar{W}_j) (\underline{x}_{j'}(k) - \bar{W}_{j'}) \\ \left. + 2 (\bar{x}_j(k) - \bar{W}_j) (\bar{x}_{j'}(k) - \bar{W}_{j'}) \right] \end{aligned} \quad (2.31)$$

Après le calcul de la matrice de covariance symbolique, la procédure suit les mêmes étapes que celle de l'ACP classique :

- diagonalisation de la matrice de covariance symbolique ;
- calcul des valeurs et vecteurs propres, triés par ordre décroissant ;
- calcul du pourcentage cumulé de variance (PCV) et détermination du nombre de composantes principales à retenir ;
- construction du modèle SPCA.

2.8 Résultats de simulation

Dans cet exemple, nous utilisons le même système statique que celui défini par l'équation (2.13), la normalisation étant réalisée comme dans l'équation (2.28), mais en utilisant cette fois la moyenne et la variance empiriques symboliques données par les équations (2.29) et (2.30).

Pour simuler l'incertitude de mesure, nous avons construit la matrice d'intervalles définie par l'équation (2.14), en ajoutant à chaque variable une imprécision ($\pm\Delta$) correspondant à 10% de sa plage de variation. Les nouvelles données de type intervalle ainsi obtenues sont représentées sur la figure ci-dessous.

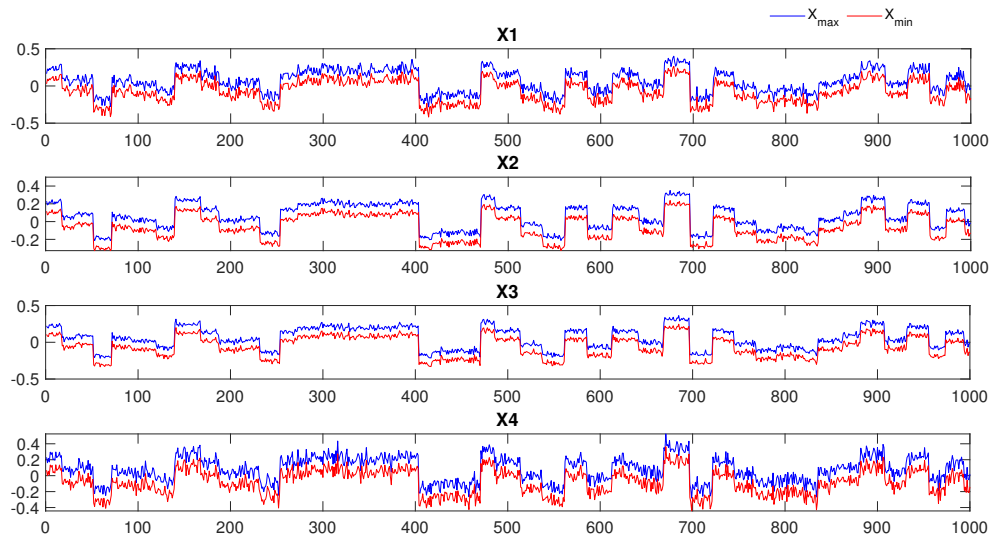


FIGURE 2.11 – Évolution des variables de type intervalle

Modélisation des données intervalles par ACP classique

Afin de montrer la limitation de l'ACP classique face aux données de type intervalle, nous avons appliqué l'ACP classique à ces données en construisant le modèle uniquement à partir des bornes inférieures X_{min} , puis en évaluant la capacité de ce modèle à estimer les bornes supérieures X_{max} .

La figure 2.12 montre une mauvaise estimation de la borne supérieure X_{max} , ce qui confirme que l'ACP classique n'est pas capable de traiter correctement des données incertaines de type intervalle.

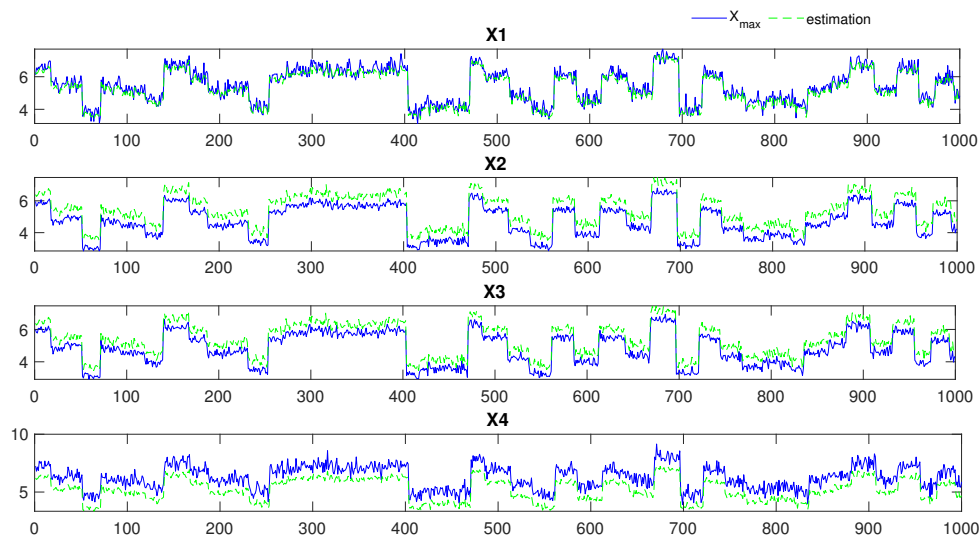


FIGURE 2.12 – Évolution de X_{max} et de son estimation avec la méthode ACP classique

Modélisation des données intervalles par la méthode SPCA

Dans le cadre de la SPCA, le modèle est construit en intégrant simultanément les deux bornes de l'intervalle, X_{min} et X_{max} , à travers la moyenne symbolique $\bar{W}_j$, la variance symbolique σ_j^2 et la covariance symbolique $Cov(j, j')$.

La figure 2.13 montre que les estimations obtenues sont identiques aux données d'origine, ce qui signifie que le modèle obtenu à l'aide de la méthode SPCA permet de fournir les meilleures estimations possibles pour les différentes bornes des données de type intervalle.

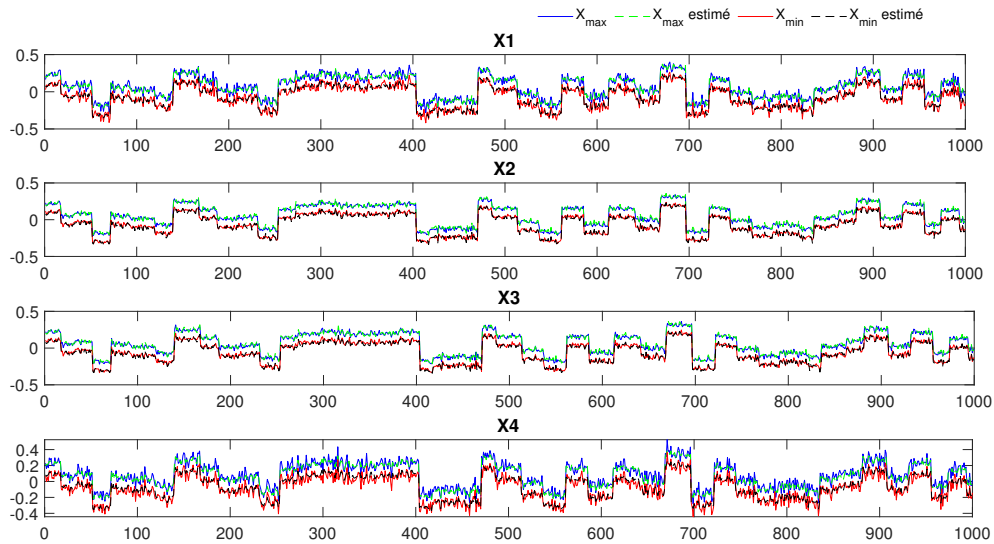
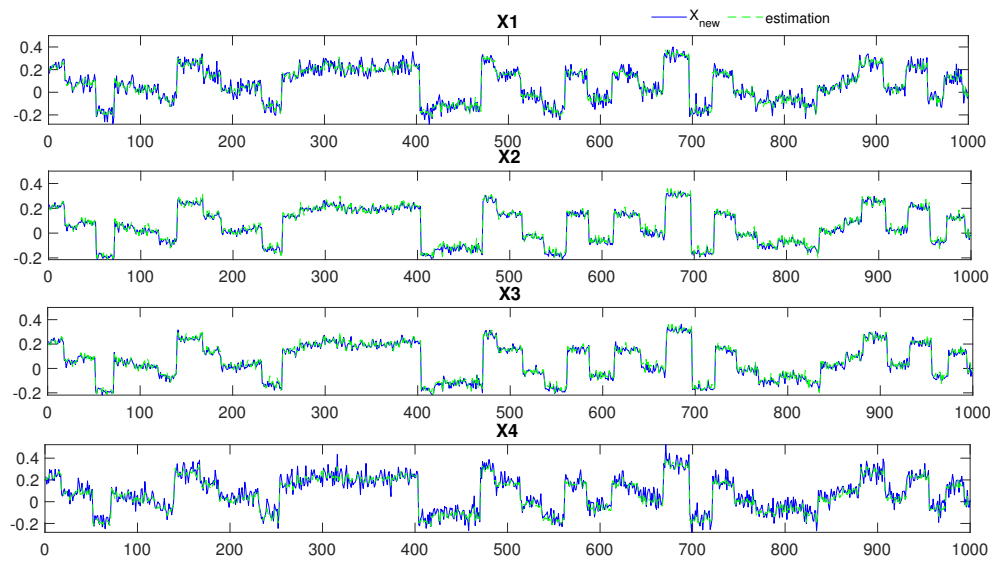


FIGURE 2.13 – Évolution des différentes mesures (X_{min} , X_{max}) et de leurs estimations par la méthode SPCA

Après avoir montré la capacité du modèle SPCA à estimer correctement les deux bornes X_{min} et X_{max} de l'intervalle, il reste à vérifier si ce modèle est également capable d'estimer toute valeur intermédiaire appartenant à l'intervalle $[X_{min}, X_{max}]$, et non uniquement ses extrémités. Pour cela, une nouvelle variable X_{new} est générée aléatoirement à l'intérieur de l'intervalle, pour chaque observation k et chaque variable j , puis estimée à l'aide du modèle SPCA.

FIGURE 2.14 – Évolution de X_{new} et de son estimation

La figure 2.14 montre que les estimations obtenues sont identiques aux données d'origine, ce qui confirme que le modèle SPCA fournit également les meilleures estimations possibles pour des valeurs comprises entre les bornes de l'intervalle. Ce résultat confirme que le modèle SPCA, construit en intégrant simultanément les deux bornes X_{min} et X_{max} , est capable d'estimer correctement toute valeur intermédiaire de l'intervalle $[X_{min}, X_{max}]$, et non uniquement ses bornes.

Capacité de l'ACP

	Données mono-valuées	Données intervalle (multi-valuées)
ACP classique	Oui	Non
ACP intervalle (SPCA)	Oui	Oui

TABLE 2.1 – Capacité de l'ACP

Le tableau 2.1 résume la capacité de chaque approche à traiter les différents types de données. On note que l'ACP classique, conçue uniquement pour des données mono-valuées, n'est pas capable de prendre en charge des données de type intervalle. En revanche, le modèle SPCA est capable de traiter aussi bien les données mono-valuées que les données intervalles (multi-valuées).

2.9 Conclusion

Dans ce chapitre, nous avons présenté le principe de modélisation par l'Analyse en Composantes Principales (ACP), ainsi que ses limites lorsqu'elle est appliquée à des données symboliques de type intervalle. Face à ces limites, nous avons introduit une approche plus adaptée,

l'ACP Symbolique (SPCA), permettant de modéliser efficacement ce type de données en intégrant simultanément les bornes inférieure et supérieure de chaque intervalle. Les résultats de simulation obtenus ont confirmé la supériorité de la SPCA par rapport à l'ACP classique pour la représentation de données imprécises.

Dans le chapitre suivant, nous nous intéresserons au rôle de l'ACP et de l'ACP Symbolique dans le diagnostic des défauts, en abordant plus particulièrement les étapes de détection et de localisation.

Diagnostic de défauts par approche statistique multivariée (ACP)

3.1 Introduction

Les systèmes industriels modernes sont soumis à des exigences croissantes en matière de fiabilité, de sécurité et de disponibilité. Toute anomalie non détectée peut entraîner des arrêts de production non planifiés, voire des accidents graves. La surveillance et le diagnostic des défauts, ou FDI (Fault Detection and Isolation), consistent à surveiller en continu l'état d'un système afin de détecter, localiser et caractériser toute déviation par rapport à un comportement de référence.

Parmi les approches existantes, les méthodes fondées sur des modèles se distinguent par leur rigueur et leur interprétabilité : elles comparent le comportement réel du système à celui prédit par un modèle mathématique, la différence entre ces deux signaux, appelée résidu, constituant l'information fondamentale de la décision de diagnostic. Toutefois, l'obtention d'un modèle physique précis reste souvent difficile, voire impossible, pour les procédés industriels complexes.

C'est dans ce contexte que s'inscrit l'approche traitée dans ce chapitre, fondée sur l'analyse en composantes principales (ACP), présentée au chapitre 2. Grâce à sa simplicité, l'ACP constitue un outil particulièrement adapté à la modélisation de procédés complexes en vue de la détection et de la localisation des défauts.

Dans ce chapitre, nous traiterons la détection et la localisation des défauts par ACP, en utilisant l'indice SPE pour la détection et le principe de reconstruction pour la localisation. Les procédures conventionnelles seront adaptées au cas des données de type intervalle.

3.2 Détection de défaut

Plusieurs indicateurs sont utilisés pour la détection des défaillances par ACP ; dans ce travail, nous avons retenu l'erreur quadratique de prédiction (SPE).

3.2.1 Indice SPE

L'indice SPE (Squared Prediction Error) permet de détecter les défauts dans le sous-espace résiduel. Il est défini par l'équation suivante :

$$SPE(k) = \sum_{j=1}^m (x_j(k) - \hat{x}_j(k))^2 = e(k)^T e(k) \quad (3.1)$$

où $x_j(k)$ désigne la valeur mesurée de la variable j à l'instant k , $\hat{x}_j(k)$ son estimation obtenue par projection sur le modèle ACP, et $e(k) = x(k) - \hat{x}(k)$ le vecteur des résidus.

L'indice SPE est un indicateur global qui additionne les résidus quadratiques sans pondération par leur variance, contrairement à l'indice T^2 qui, lui, normalise chaque composante par l'inverse de sa variance. Cette absence de pondération n'est pas neutre : les résidus associés aux directions de forte variance résiduelle traduisent principalement des erreurs de modélisation de l'ACP (nombre de composantes principales mal estimé, non-linéarités non captées, etc.), alors que les résidus de faible variance reflètent plutôt la violation des relations de redondance linéaire entre variables, c'est-à-dire les corrélations structurelles du processus en fonctionnement normal.

Pour l'indice SPE, on considère qu'un processus fonctionne de manière anormale, ou qu'un défaut est présent à l'instant k , lorsque la condition suivante est vérifiée [16] :

$$SPE(k) > \delta_\alpha^2 \quad (3.2)$$

où δ_α^2 est le seuil de détection de $SPE(k)$, calculé à partir des grandeurs $\theta_i = \sum_{j=\ell+1}^m \lambda_j^i$, pour $i = 1, 2, 3$, avec λ_j la $j^{\text{ème}}$ valeur propre de la matrice de covariance Σ . Ce seuil est approximé par [17] :

$$\delta_\alpha^2 = \theta_1 \left[\frac{C_\alpha \sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]^{\frac{1}{h_0}} \quad (3.3)$$

où $h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2}$ et C_α est la valeur limite correspondant au seuil de confiance $(1 - \alpha)$ d'une distribution normale.

3.2.2 Indice SPE (cas des données intervalle)

Comme dans le cas de l'ACP classique, la détection de défauts repose sur le calcul de l'indice SPE. Toutefois, dans le cadre de l'ACP symbolique (SPCA), deux indices SPE sont calculés : l'un pour la borne inférieure et l'autre pour la borne supérieure de l'intervalle :

$$SPE_{min}(k) = \sum_{j=1}^m (x_j(k) - \hat{x}_j(k))^2 \quad (3.4)$$

$$SPE_{max}(k) = \sum_{j=1}^m (\bar{x}_j(k) - \hat{\hat{x}}_j(k))^2 \quad (3.5)$$

où $\underline{x}_j(k)$ et $\bar{x}_j(k)$ désignent respectivement les bornes inférieure et supérieure de la variable j à l'instant k , et $\hat{\underline{x}}_j(k)$, $\hat{\bar{x}}_j(k)$ leurs estimations obtenues par projection sur le modèle SPCA.

Un défaut est détecté lorsque l'un des deux indices dépasse le seuil δ^2 , ce dernier étant calculé selon la même formule que dans le cas de l'ACP classique (équation (3.3)), en utilisant les valeurs propres de la matrice de covariance symbolique.

3.3 Localisation de défaut basée sur le principe de reconstruction

Après la détection d'un défaut, il est nécessaire d'identifier la variable qui en est à l'origine : c'est la phase de localisation. Plusieurs méthodes de localisation de défauts ont été développées dans la littérature. Parmi celles-ci, la méthode basée sur le principe de reconstruction est l'une des plus efficaces.

3.3.1 Principe de reconstruction de variable

La méthode de reconstruction suppose que chaque variable entachée d'un défaut peut être reconstruite, notée z_i , en se basant sur le modèle ACP et sur les mesures des autres variables x_j ($j \neq i$), à l'exclusion de la variable en défaut. Nous illustrons dans ce qui suit le déroulement de l'opération de reconstruction. Le modèle ACP est exprimé par la relation suivante :

$$C = PP^T = \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{pmatrix} \quad (3.6)$$

L'estimation des mesures est alors donnée par :

$$\hat{X} = \begin{pmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \hat{x}_3 \end{pmatrix} = Cx = C \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \quad (3.7)$$

Soit :

$$\hat{x}_1 = c_{11}x_1 + c_{12}x_2 + c_{13}x_3 \quad (3.8)$$

$$\hat{x}_2 = c_{21}x_1 + c_{22}x_2 + c_{23}x_3 \quad (3.9)$$

$$\hat{x}_3 = c_{31}x_1 + c_{32}x_2 + c_{33}x_3 \quad (3.10)$$

En se basant sur le principe de reconstruction, la $i^{\text{ème}}$ variable est calculée à partir de toutes

les autres variables, en utilisant le modèle ACP, à l'exclusion de la $i^{\text{ème}}$ variable elle-même :

$$z_1 = f(x_2, x_3) \quad (3.11)$$

$$z_2 = f(x_1, x_3) \quad (3.12)$$

$$z_3 = f(x_1, x_2) \quad (3.13)$$

En supposant $\hat{x}_i = x_i$, les équations précédentes deviennent :

$$x_1 = c_{11}x_1 + c_{12}x_2 + c_{13}x_3 \quad (3.14)$$

$$x_2 = c_{21}x_1 + c_{22}x_2 + c_{23}x_3 \quad (3.15)$$

$$x_3 = c_{31}x_1 + c_{32}x_2 + c_{33}x_3 \quad (3.16)$$

d'où l'on tire :

$$x_1(1 - c_{11}) = c_{12}x_2 + c_{13}x_3$$

$$x_2(1 - c_{22}) = c_{21}x_1 + c_{23}x_3$$

$$x_3(1 - c_{33}) = c_{31}x_1 + c_{32}x_2$$

soit :

$$x_1 = \frac{c_{12}x_2 + c_{13}x_3}{1 - c_{11}} \quad (3.17)$$

$$x_2 = \frac{c_{21}x_1 + c_{23}x_3}{1 - c_{22}} \quad (3.18)$$

$$x_3 = \frac{c_{31}x_1 + c_{32}x_2}{1 - c_{33}} \quad (3.19)$$

En posant $z_i = \hat{x}_i = x_i$, on obtient :

$$z_1 = \frac{c_{12}x_2 + c_{13}x_3}{1 - c_{11}} \quad (3.20)$$

$$z_2 = \frac{c_{21}x_1 + c_{23}x_3}{1 - c_{22}} \quad (3.21)$$

$$z_3 = \frac{c_{31}x_1 + c_{32}x_2}{1 - c_{33}} \quad (3.22)$$

La solution z_i vérifie ainsi :

$$z_1(1 - c_{11}) = \begin{bmatrix} 0 & c_{12} & c_{13} \end{bmatrix} x(k) \quad (3.23)$$

$$z_2(1 - c_{22}) = \begin{bmatrix} c_{21} & 0 & c_{23} \end{bmatrix} x(k) \quad (3.24)$$

$$z_3(1 - c_{33}) = \begin{bmatrix} c_{31} & c_{32} & 0 \end{bmatrix} x(k) \quad (3.25)$$

d'où :

$$z_1 = \frac{\begin{bmatrix} 0 & c_{12} & c_{13} \end{bmatrix}}{1 - c_{11}} x(k), \quad z_2 = \frac{\begin{bmatrix} c_{21} & 0 & c_{23} \end{bmatrix}}{1 - c_{22}} x(k), \quad z_3 = \frac{\begin{bmatrix} c_{31} & 0 & c_{33} \end{bmatrix}}{1 - c_{33}} x(k) \quad (3.26)$$

La forme générale de la variable reconstruite z_i s'écrit donc :

$$z_i(1 - c_{ii}) = \begin{bmatrix} c_{i-} & 0 & c_{i+} \end{bmatrix} x(k) \quad (3.27)$$

et la $i^{\text{ème}}$ composante reconstruite du vecteur de mesure $x(k)$ est donnée par :

$$z_i = \frac{\begin{bmatrix} c_{i-} & 0 & c_{i+} \end{bmatrix}}{1 - c_{ii}} x(k) \quad (3.28)$$

Pour le cas des données de type intervalle, le même principe de reconstruction est étendu au cadre symbolique en l'appliquant séparément aux bornes inférieure et supérieure de l'intervalle. Pour chaque variable i , la variable reconstruite z_i est estimée à partir de toutes les autres variables et du modèle SPCA, à l'exclusion de la $i^{\text{ème}}$ variable, ce qui donne :

$$z_i^{\min}(k) = \frac{\begin{bmatrix} c_{i-} & 0 & c_{i+} \end{bmatrix}}{1 - c_{ii}} \underline{x}(k) \quad (3.29)$$

$$z_i^{\max}(k) = \frac{\begin{bmatrix} c_{i-} & 0 & c_{i+} \end{bmatrix}}{1 - c_{ii}} \bar{x}(k) \quad (3.30)$$

La méthode de localisation basée sur le principe de reconstruction, développée par Dunia et al[19], exploite la capacité de l'ACP à reconstruire chaque variable à partir des autres variables du système. Le principe, illustré dans la Figure 2.5, consiste à créer plusieurs blocs de données dans lesquels chaque variable suspecte est systématiquement remplacée par sa reconstruction.

Le processus de localisation suit les étapes suivantes :

1. **Calcul des reconstructions de référence.** À partir des données de fonctionnement X , on calcule les reconstructions $z_1, z_2, \dots, z_m$ de l'ensemble des variables à l'aide du modèle ACP. Ces reconstructions représentent les valeurs estimées de chaque variable, obtenues uniquement à partir des autres variables du système, en exploitant les corrélations linéaires capturées par le modèle.
2. **Génération des blocs de données modifiés.** Pour identifier la variable défectueuse, m blocs de données modifiés sont construits : dans le $i^{\text{ème}}$ bloc, la variable x_i est remplacée par sa reconstruction z_i , tandis que les $m - 1$ autres variables demeurent inchangées.
3. **Calcul de la signature de défaut.** Chaque bloc de données modifié est ensuite traité par le modèle ACP afin de calculer l'indice SPE correspondant. Cette procédure génère une signature de défaut composée de m valeurs : $SPE_1, SPE_2, \dots, SPE_m$.
4. **Isolation du défaut par analyse de la signature.** L'analyse de cette signature permet d'isoler la variable en défaut : si le remplacement de la variable j par sa reconstruction

z_j fait chuter significativement l'indice SPE en dessous du seuil de détection δ_α^2 , cela indique que cette variable est la source du défaut. En effet, la substitution d'une variable défectueuse par sa valeur reconstruite « saine » élimine l'effet de l'anomalie sur l'indicateur global, alors que la reconstruction d'une variable saine à partir de variables corrélées à une variable défaillante ne permet pas de restaurer l'indice sous le seuil.

Cette méthode est particulièrement efficace car elle exploite la redondance analytique présente dans les données corrélées, ce qui permet une isolation précise des défauts. Dans le cas des données de type intervalle, ce même principe de comparaison avant/après reconstruction s'applique séparément aux indices SPE_{min} et SPE_{max} [14].

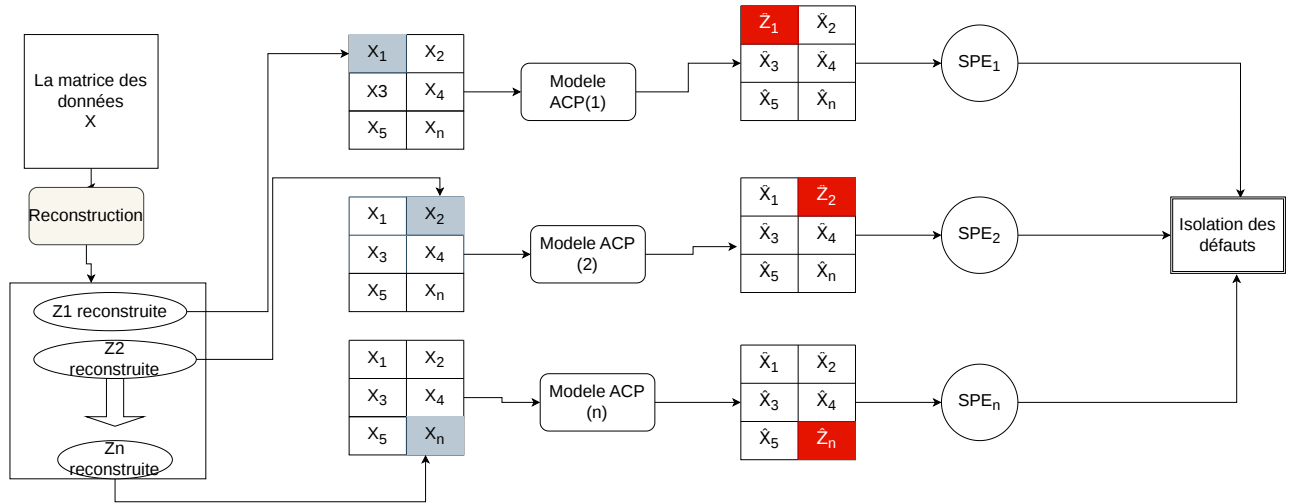
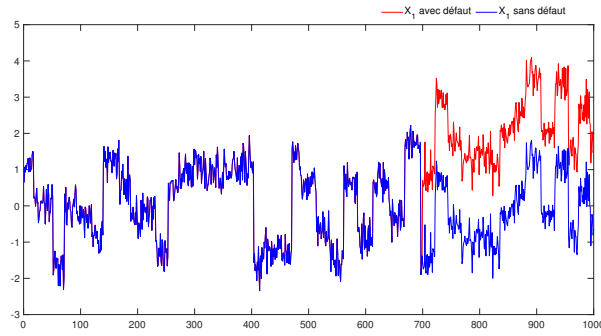


FIGURE 3.1 – Localisation basée sur le principe de reconstruction

3.4 Simulation

3.4.1 Cas de données classiques (mono-valued Data)

Afin d'illustrer la procédure de détection et de localisation, nous reprenons le même exemple utilisé dans le chapitre précédent. Un défaut de type biais, d'amplitude égale à 30% de la plage de variation de la variable x_1 , est simulé entre les instants $k = 700$ et $k = 1000$. L'évolution de la variable x_1 , avec et sans défaut, est représentée sur la figure 3.2.

FIGURE 3.2 – Évolution de x_1 sans et avec défaut

L'indice SPE est ensuite calculé dans le cas de l'ACP classique ; la figure 3.3 présente son évolution, avec son seuil de détection, en l'absence puis en présence du défaut.

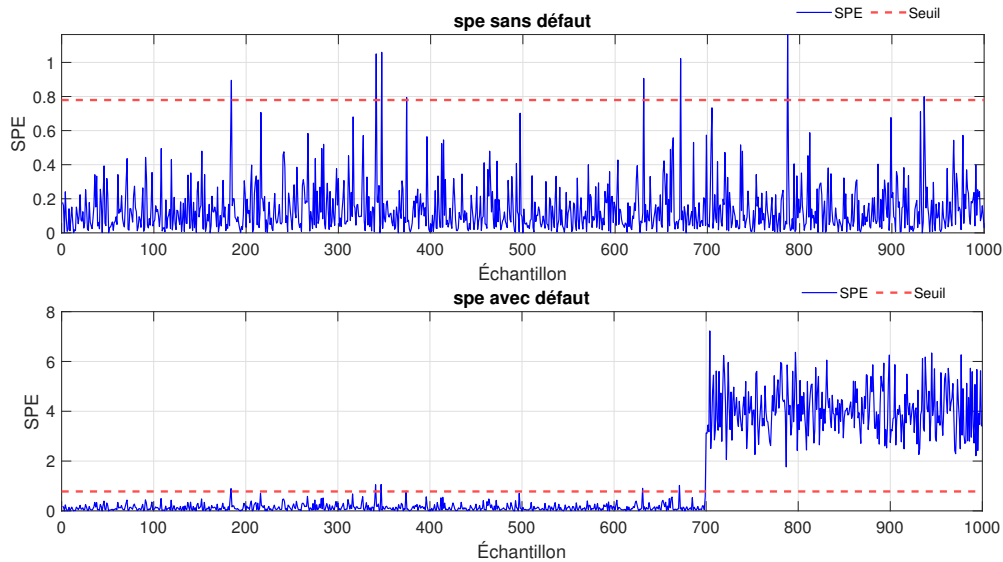


FIGURE 3.3 – Évolution de SPE avec son seuil, avec et sans défaut

Localisation des défauts

Une fois le défaut détecté, l'étape suivante consiste à isoler la variable fautive (dans notre cas, x_1), à l'aide de la méthode de reconstruction des variables z_i présentée précédemment dans ce chapitre (section 3.3.1). La variable x_1 étant en défaut, sa reconstruction z_1 est calculée à partir de l'ensemble des autres variables x_i ($i \neq 1$), c'est-à-dire à partir de variables saines. Ainsi, z_1 constitue une estimation de l'évolution naturelle de x_1 , indépendante de l'erreur affectant cette dernière. En revanche, les variables reconstruites z_i (pour $i \neq 1$) sont calculées en utilisant x_1 parmi leurs variables d'entrée : elles se trouvent donc entachées par le défaut de x_1 . Il en résulte que l'ensemble des indices SPE_i associés aux blocs de données dans lesquels x_1 intervient reste sensible à ce défaut.

La figure 3.4 présente l'évolution de x_1 , avec et sans défaut, ainsi que sa variable reconstruite z_1 .

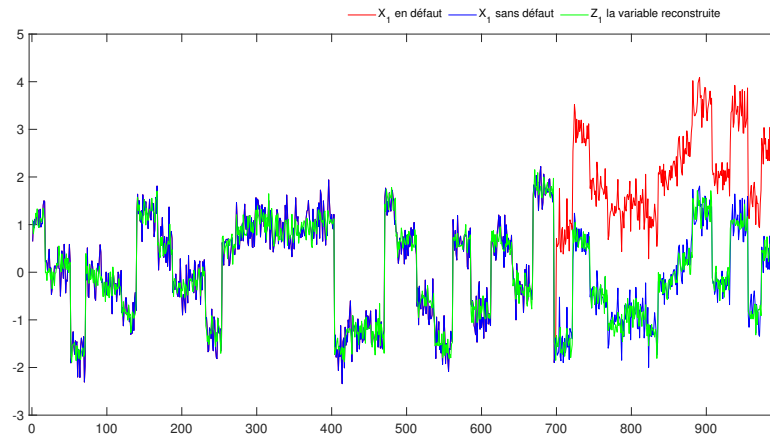


FIGURE 3.4 – l'évolution de x_1 sans et avec défaut et sa variable reconstruite z_1

Les figures 3.5, 3.6 et 3.7 présentent l'ensemble des indices SPE calculés pour les différents blocs de données, respectivement en l'absence de défaut, en présence d'un défaut sur x_1 , et en présence d'un défaut sur x_2 .

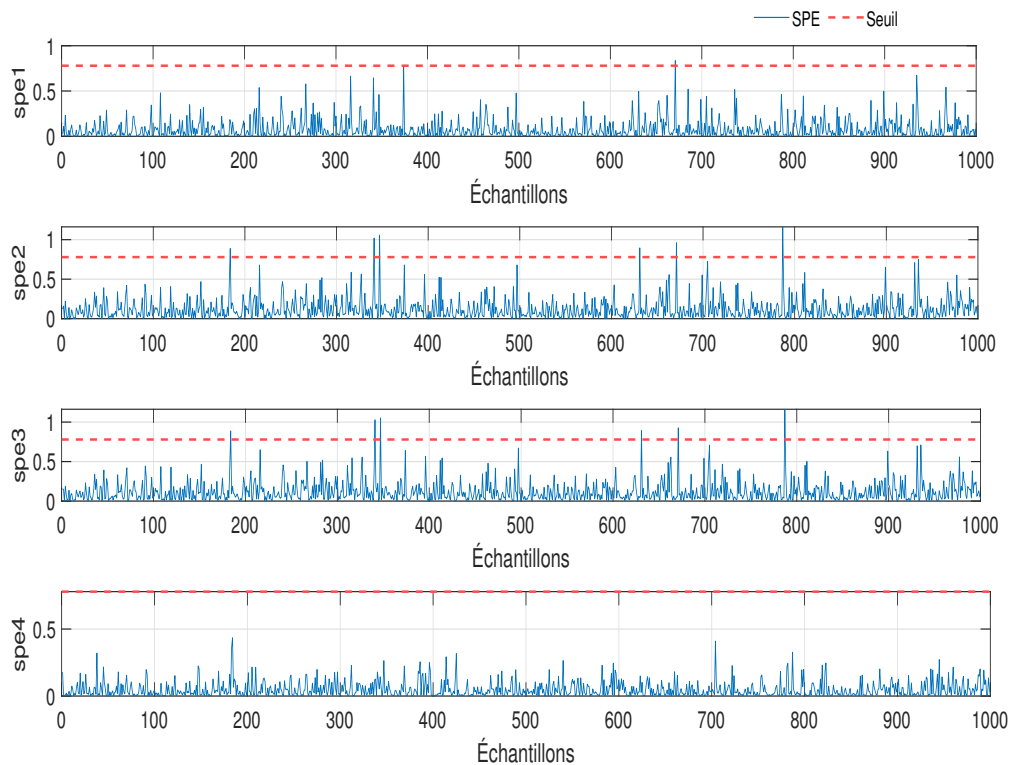


FIGURE 3.5 – Ensemble des SPE en l'absence de défaut

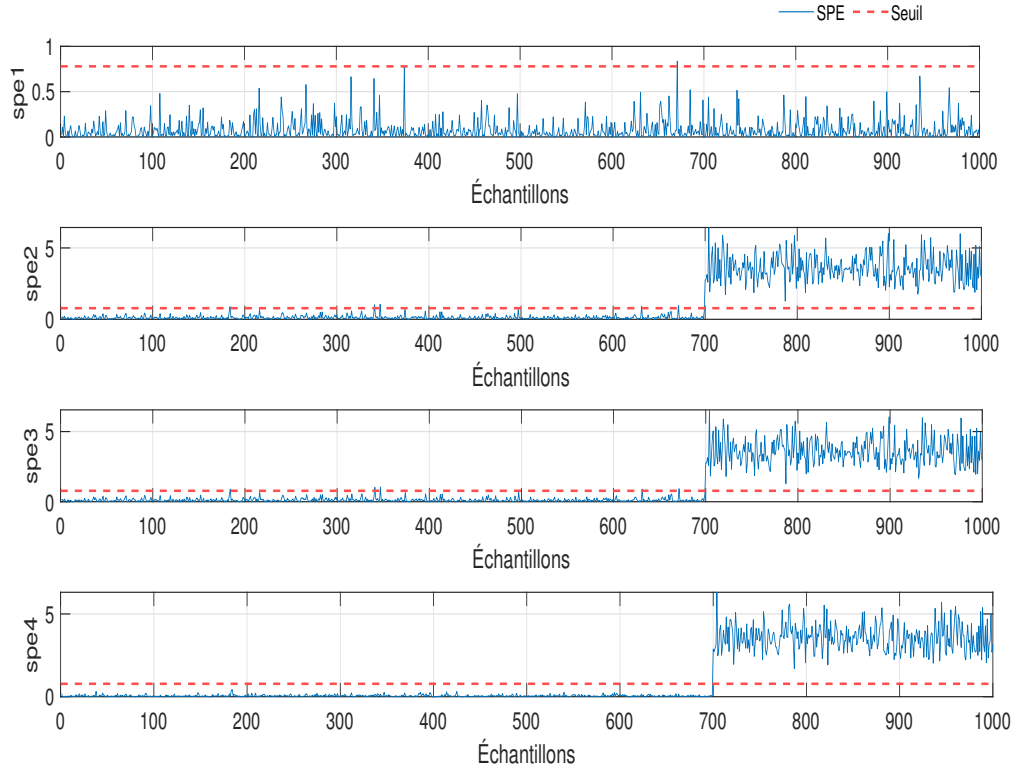


FIGURE 3.6 – Ensemble des SPE en présence d'un défaut sur x_1

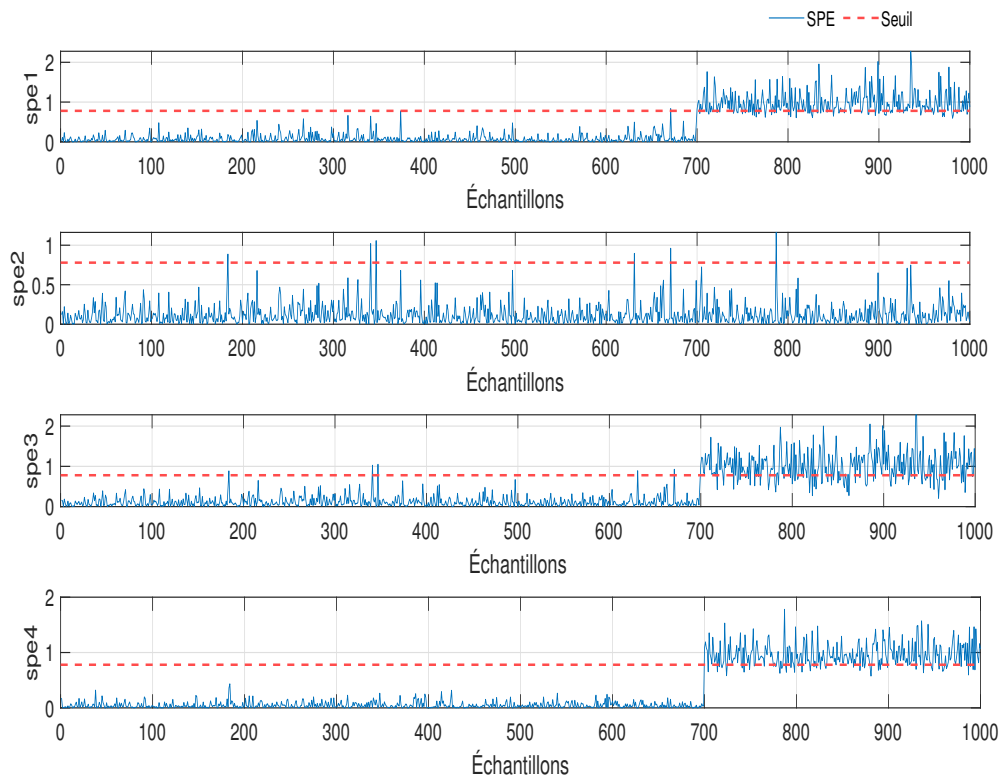


FIGURE 3.7 – Ensemble des SPE en présence d'un défaut sur x_2

3.4.2 Cas de données imprécises de type intervalle

Afin d'évaluer les performances du modèle SPCA en détection et localisation de défauts, nous reprenons le même exemple de simulation que celui utilisé pour l'ACP classique, à savoir $n = 1000$ observations et $m = 4$ variables, avec une incertitude $\delta = 10\%$ de la plage de variation de chaque variable.

Détection de défauts

Un défaut de type biais, d'amplitude égale à 30% de la plage de variation de la variable x_1 , est simulé à partir de l'instant $k = 700$. La figure 3.8 montre l'évolution des bornes X_{min} et X_{max} de la variable x_1 , avec et sans défaut.

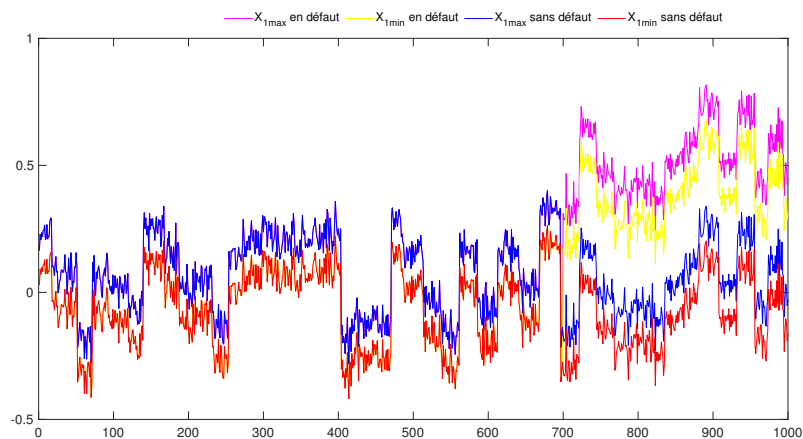


FIGURE 3.8 – Évolution de x_1 sans et avec défaut

La figure 3.9 présente l'évolution de l'indice de détection SPE, sans et avec défaut sur x_1 .

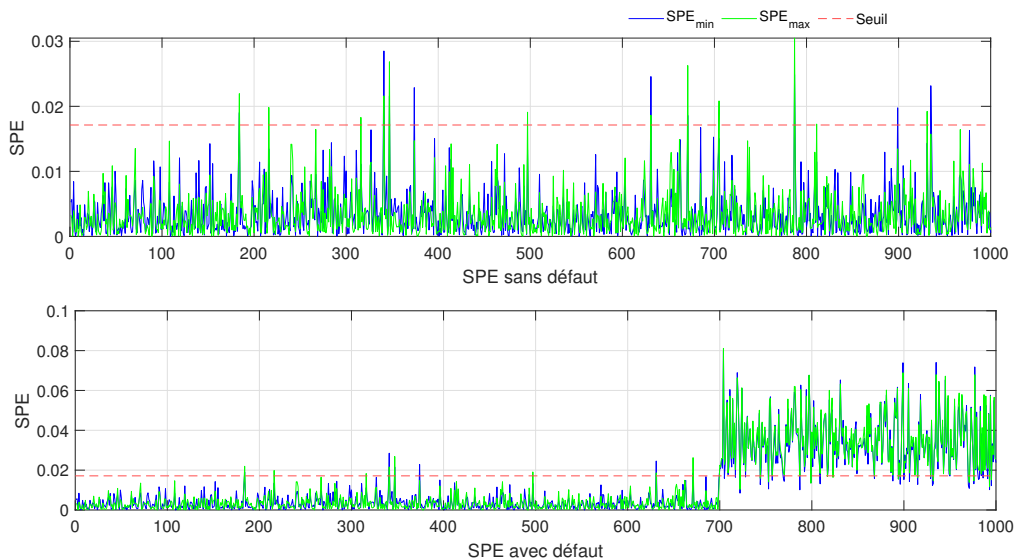


FIGURE 3.9 – Évolution de SPE sans et avec défaut

D'après la figure 3.9, les indices SPE_{min} et SPE_{max} restent inférieurs au seuil de détection

δ^2 en l'absence de défaut, ce qui confirme le bon fonctionnement du modèle SPCA. À partir de l'instant $k = 700$, les deux indices dépassent nettement le seuil, confirmant ainsi la détection du défaut par la méthode SPCA.

Localisation de défaut

Pour localiser la variable en défaut, la procédure de reconstruction décrite précédemment est appliquée à l'ensemble des $m = 4$ variables. L'évolution de la variable x_1 , sans et avec défaut, ainsi que sa reconstruction z_1 par la méthode SPCA, est illustrée sur la figure 3.10.

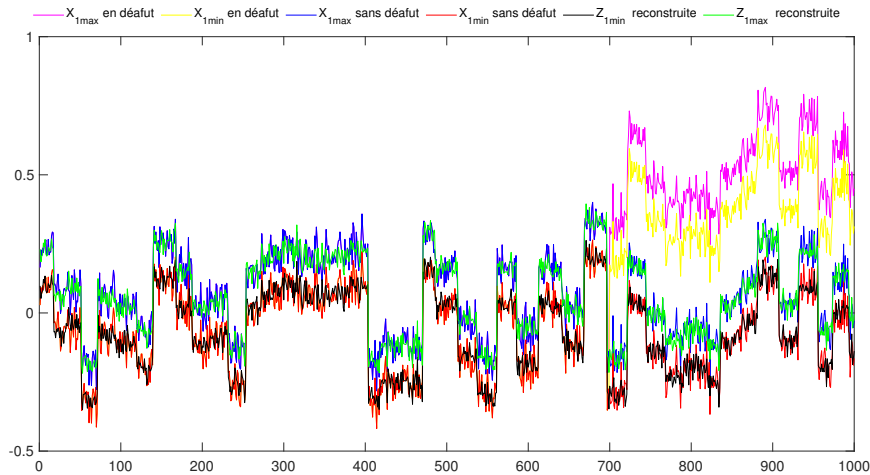


FIGURE 3.10 – Évolution de x_1 sans et avec défaut, et sa variable reconstruite z_1

Les résultats de la localisation de défaut obtenus avec la méthode SPCA sont illustrés sur les figures 3.11 et 3.12.

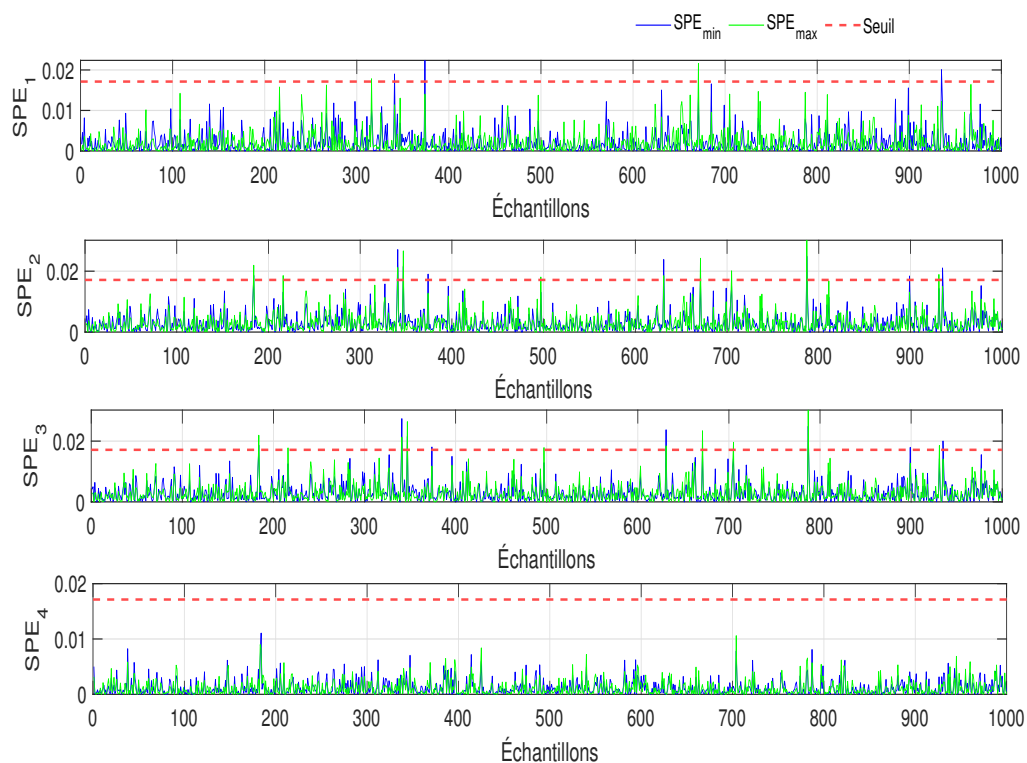


FIGURE 3.11 – SPE structuré en l'absence de défaut

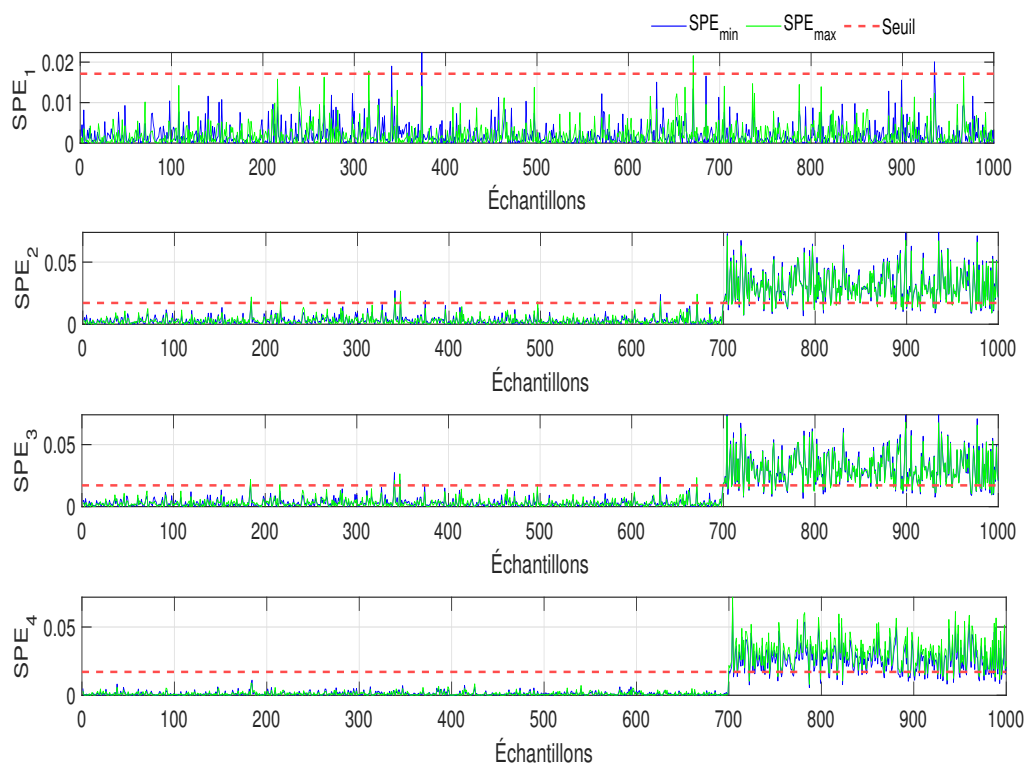
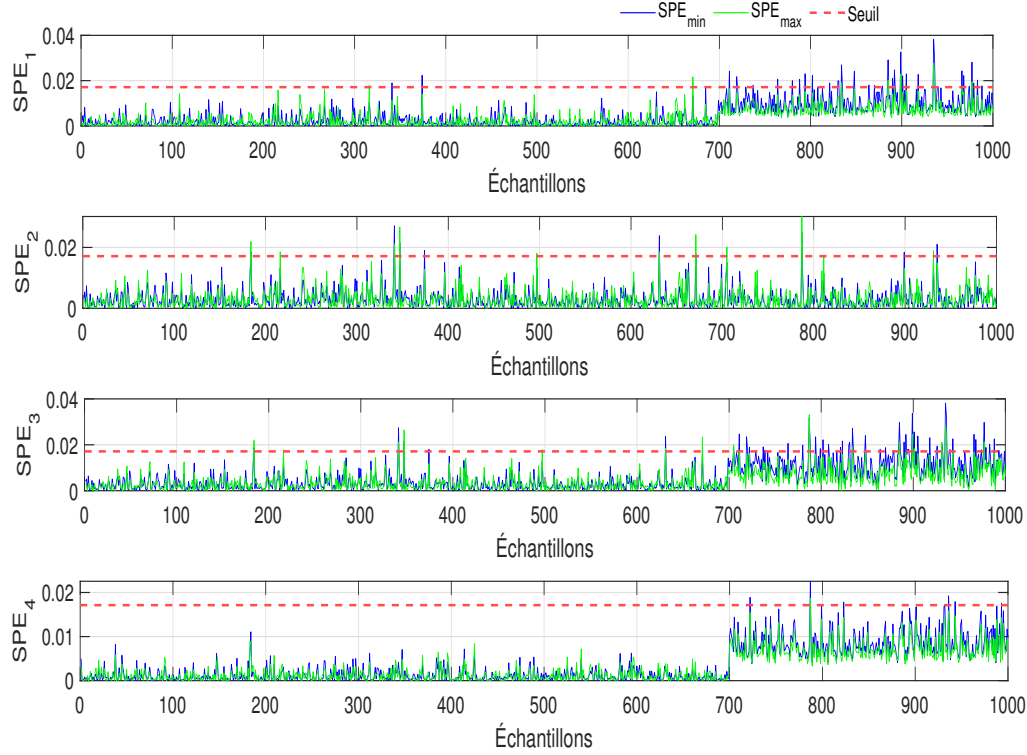


FIGURE 3.12 – SPE structuré en présence de défaut sur x_1

FIGURE 3.13 – SPE structuré en présence de défaut sur x_2

Les figures 3.12 et 3.13 montrent que seul l'indice SPE_1 (respectivement SPE_2), correspondant à la reconstruction de la variable x_1 (respectivement x_2), reste inférieur au seuil de détection après l'instant $k = 700$, alors que les indices $SPE_2, \dots, SPE_4$ (respectivement SPE_1, SPE_3, SPE_4) demeurent supérieurs au seuil. Cette signature confirme que la variable en défaut a été localisée avec succès par l'approche de reconstruction basée sur le modèle SPCA.

3.5 Conclusion

Dans ce chapitre, nous avons présenté la détection et la localisation de défauts pour des données ponctuelles et de type intervalle, respectivement à l'aide de l'ACP classique et de l'ACP symbolique. Les résultats de simulation ont montré l'efficacité de ces deux approches pour la détection et l'isolation des défauts.

Le chapitre suivant est consacré à un cas d'application sur un benchmark.

Application sur un benchmark industriel : turbine à gaz

4.1 Introduction

Le développement économique et technologique à l'échelle mondiale est intrinsèquement lié à la disponibilité et à la gestion de l'énergie électrique. Les systèmes de production d'énergie ont pour fonction principale de transformer des sources d'énergie primaire — fossiles, nucléaires ou renouvelables — en une forme d'énergie utile et transportable, principalement l'électricité. Dans un contexte d'augmentation continue de la demande énergétique mondiale et de normes environnementales de plus en plus strictes, l'optimisation, la décarbonation et la diversification des systèmes de production revêtent une importance stratégique majeure pour les nations et les industries énergétiques. Dans le cadre de la gestion des centrales électriques, la continuité de service et la sécurité des installations constituent des priorités absolues pour les exploitants : un arrêt non planifié, causé par une défaillance soudaine du système, peut engendrer des pertes financières considérables et compromettre la stabilité du réseau électrique interconnecté. C'est dans ce contexte que le diagnostic industriel et la surveillance d'état apparaissent comme des outils essentiels pour évaluer l'état des équipements et optimiser leurs performances [23].

4.2 Systèmes de production d'énergie

Définition. Un système de production d'énergie désigne un ensemble d'équipements et d'infrastructures technologiques permettant de convertir une source d'énergie primaire en une forme d'énergie finale utilisable par les consommateurs. Ces sources primaires peuvent être le gaz naturel, l'énergie éolienne ou l'énergie hydraulique, tandis que la forme d'énergie finale peut être l'électricité, la chaleur ou un carburant.



FIGURE 4.1 – Une centrale de production d'énergie

Dans le domaine de l'énergie, les technologies de génération d'électricité se classent principalement en trois grandes familles, selon la nature de leur source primaire :

- **Les centrales thermiques conventionnelles** utilisent des combustibles fossiles (charbon, gaz naturel, fioul) ou l'énergie nucléaire pour produire de l'électricité à partir de cycles thermodynamiques tels que le cycle de Joule-Brayton, permettant une production efficace et contrôlée.
- **Les systèmes à énergies renouvelables**, tels que les parcs éoliens ou les centrales solaires, offrent l'avantage d'une faible empreinte carbone, mais restent sujets à des fluctuations de production liées à la variabilité de leur source d'énergie.
- **Les systèmes hybrides et de cogénération** combinent plusieurs sources d'énergie ou valorisent les pertes thermiques, dans le but d'optimiser le rendement global de l'installation en tirant parti de l'ensemble des sources disponibles.

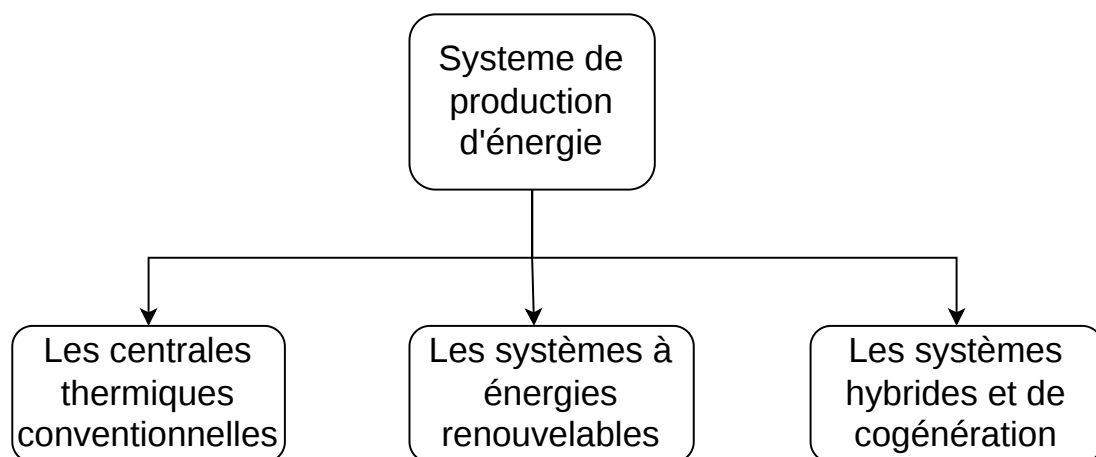


FIGURE 4.2 – Différents systèmes de production d'énergie

Parmi ces technologies, la turbine à gaz occupe une place particulière au sein des systèmes de production d'énergie.

4.3 Turbine à gaz

Une turbine à gaz, également désignée sous les appellations de turbine à combustion ou turbine à gaz de combustion, est une machine tournante thermodynamique appartenant à la famille des moteurs à combustion interne, dont le rôle est de produire de l'énergie mécanique sous forme de rotation d'un arbre. Cette énergie mécanique provient de l'énergie cinétique des gaz issus de la combustion d'un hydrocarbure (fioul, gaz combustible, etc.), qui subissent une détente dans la turbine. Le comburant, le plus souvent de l'air ambiant, est généralement comprimé avant son admission dans la chambre de combustion ; cette compression est assurée par un compresseur rotatif entraîné par le même arbre que la turbine [23].

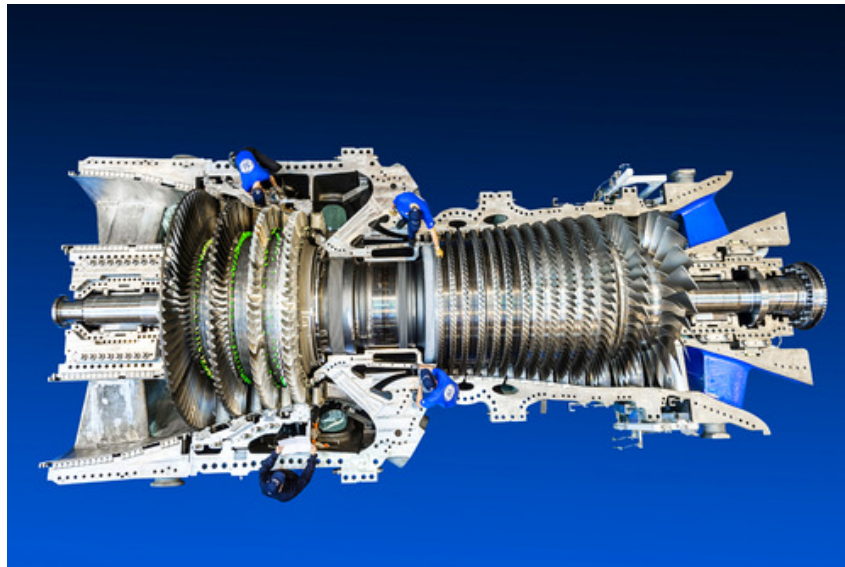


FIGURE 4.3 – Turbine à gaz

4.3.1 Classification des turbines à gaz

Les turbines à gaz peuvent être classées selon plusieurs critères :

1. Classification selon le cycle thermodynamique

- **Cycle simple** : ces machines présentent l'avantage d'être simples, peu coûteuses et faciles à installer et à exploiter. Leur rendement reste cependant généralement inférieur à celui des moteurs diesel ou des centrales équipées de turbines à vapeur, en raison de pertes à l'échappement significatives.
- **Cycle de régénération** : ce cycle améliore l'efficacité globale du système grâce à la récupération de chaleur, utilisée pour préchauffer l'air comprimé avant son admission dans la chambre de combustion. À puissance délivrée égale, la quantité de carburant nécessaire est ainsi réduite, ce qui améliore le rendement global.
- **Cycle combiné** : dans ce cycle, la chaleur contenue dans les gaz d'échappement de la turbine à gaz — caractérisés par un débit élevé et des températures supérieures à

450 °C — est valorisée pour produire de l'énergie supplémentaire. Dans un contexte industriel, la vapeur ainsi produite peut être utilisée dans un cycle vapeur, l'eau sous pression pour le chauffage, et l'air chaud pour des applications de séchage, notamment dans les industries du bois et de la céramique.

2. Classification selon le mode de construction

- **Turbine mono-arbre** : l'ensemble du système est entraîné par un seul arbre, mis en rotation par un moteur de lancement jusqu'à une vitesse donnée, puis maintenu par la turbine haute pression. Ce type de configuration est privilégié pour les applications à charge constante, notamment l'entraînement de générateurs électriques.
- **Turbine bi-arbre** : une seconde turbine, mécaniquement indépendante, est ajoutée en aval sur un arbre séparé. Cette configuration offre une grande souplesse d'exploitation et convient aux applications de forte puissance nécessitant l'entraînement d'appareils à charge variable (pompes, compresseurs, etc.).

3. Classification selon le mode de fonctionnement : on distingue deux types de turbines, la turbine à action et la turbine à réaction.

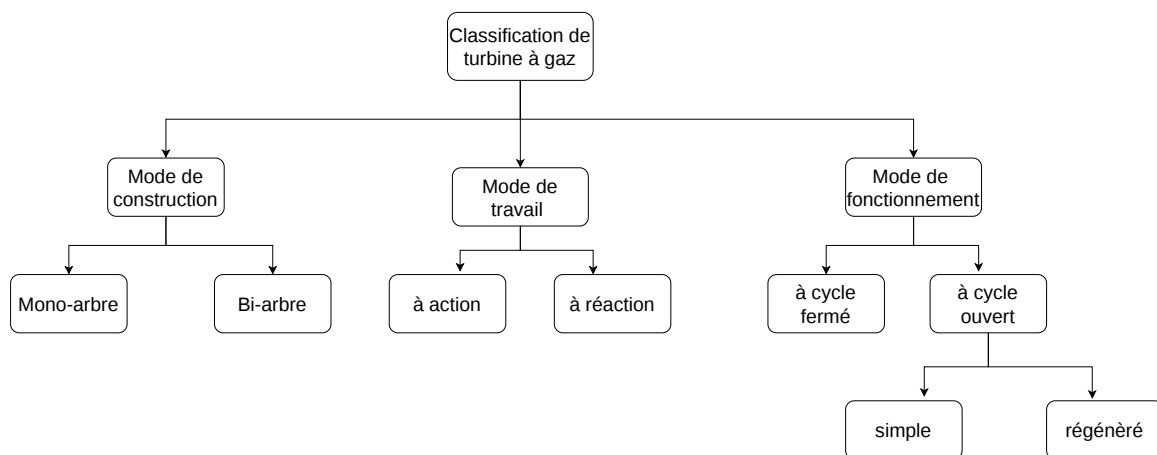


FIGURE 4.4 – Classification des turbines à gaz

4.3.2 Composants d'une turbine à gaz

Une turbine à gaz est composée des éléments suivants :

- **Compresseur axial** : comprime l'air atmosphérique jusqu'à une pression très élevée.
- **Chambre de combustion** : l'air comprimé issu du compresseur y est mélangé au combustible.
- **Détente dans la turbine** : les gaz chauds et sous pression se détendent dans la turbine, entraînant l'arbre rotatif et produisant ainsi de l'énergie mécanique.
- **Production d'énergie** : l'arbre de la turbine entraîne un générateur, produisant de l'énergie électrique.

- **Émission des gaz d'échappement** : les gaz brûlés, chargés de polluants, sont évacués vers l'atmosphère.

Dans ce travail, nous nous intéressons plus particulièrement aux centrales à turbine à gaz (*Gas Turbine Power Plants*).

4.4 Centrale à turbine à gaz

Une centrale à turbine à gaz (*Gas Turbine Power Plant*) désigne une installation de production d'électricité dans laquelle une turbine à gaz convertit l'énergie chimique du combustible en énergie mécanique, puis en énergie électrique via un alternateur [22].

4.4.1 Principe de fonctionnement

Le fonctionnement d'une turbine à gaz repose sur le cycle de Brayton, constitué de quatre transformations thermodynamiques principales : la compression de l'air, l'apport de chaleur par combustion à pression quasi constante, la détente des gaz dans la turbine, et le rejet des gaz d'échappement. Ce cycle assure la conversion de l'énergie chimique du combustible en énergie mécanique, puis en énergie électrique lorsque la turbine entraîne un alternateur.

L'air atmosphérique est d'abord aspiré à travers le système d'admission, puis comprimé au sein du compresseur, ce qui entraîne une augmentation significative de sa pression et de sa température. Il est ensuite acheminé vers la chambre de combustion, où il est mélangé au combustible ; l'allumage du mélange produit des gaz de combustion à très haute température et haute pression. Ces gaz chauds subissent une détente dans la turbine : une fraction de l'énergie thermique est ainsi convertie en énergie mécanique, provoquant la rotation du rotor. Une partie de cette puissance mécanique est utilisée pour entraîner le compresseur, tandis que le reste est transmis à un alternateur pour produire de l'énergie électrique [22].

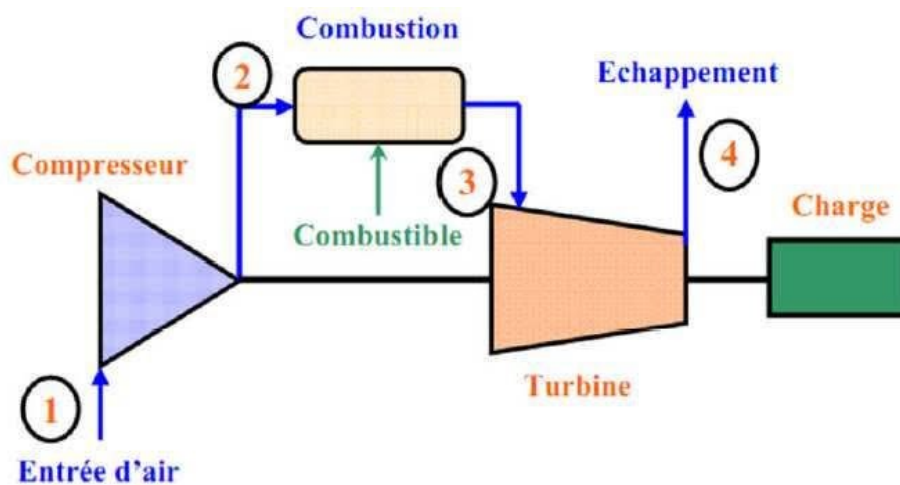


FIGURE 4.5 – Principe de fonctionnement d'une turbine à gaz

4.5 Modélisation et diagnostic par la méthode ACP Symbolique

Dans cette étude, les algorithmes PCA et SPCA sont appliqués à un jeu de données benchmark issu d'une turbine à gaz industrielle installée dans une usine en Turquie. Ce jeu de données de référence (*benchmark dataset*), disponible publiquement, a été utilisé à cet effet [24].

L'ensemble de données a été collecté à partir de onze capteurs différents : la température ambiante (AT), la pression ambiante (AP), l'humidité ambiante (AH), la pression différentielle du filtre à air (AFDP), la pression d'échappement de la turbine (GTEP), la température d'entrée de la turbine (TIT), la température de sortie de la turbine (TAT), la pression de refoulement du compresseur (CDP), la puissance produite (TEY), ainsi que les émissions de CO et de NO_x. Ces variables ont été mesurées en continu, en fonctionnement normal, sous forme de moyennes horaires sur une année, soit un total de 7627 échantillons par variable.

TABLE 4.1 – Description des variables de la base de données

Notation	Variable	Abréviation	Unité
X_1	Température ambiante	AT	°C
X_2	Pression ambiante	AP	mbar
X_3	Humidité ambiante	AH	%
X_4	Pression différentielle du filtre à air	AFDP	mbar
X_5	Pression des gaz d'échappement de la turbine	GTEP	mbar
X_6	Température d'entrée de la turbine	TIT	°C
X_7	Température de sortie de la turbine	TAT	°C
X_8	Pression de refoulement du compresseur	CDP	mbar
X_9	Énergie produite par la turbine	TEY	MWh
X_{10}	Monoxyde de carbone	CO	mg/m ³
X_{11}	Oxydes d'azote	NO _x	mg/m ³

L'évolution des différentes variables est représentée sur la figure 4.6.

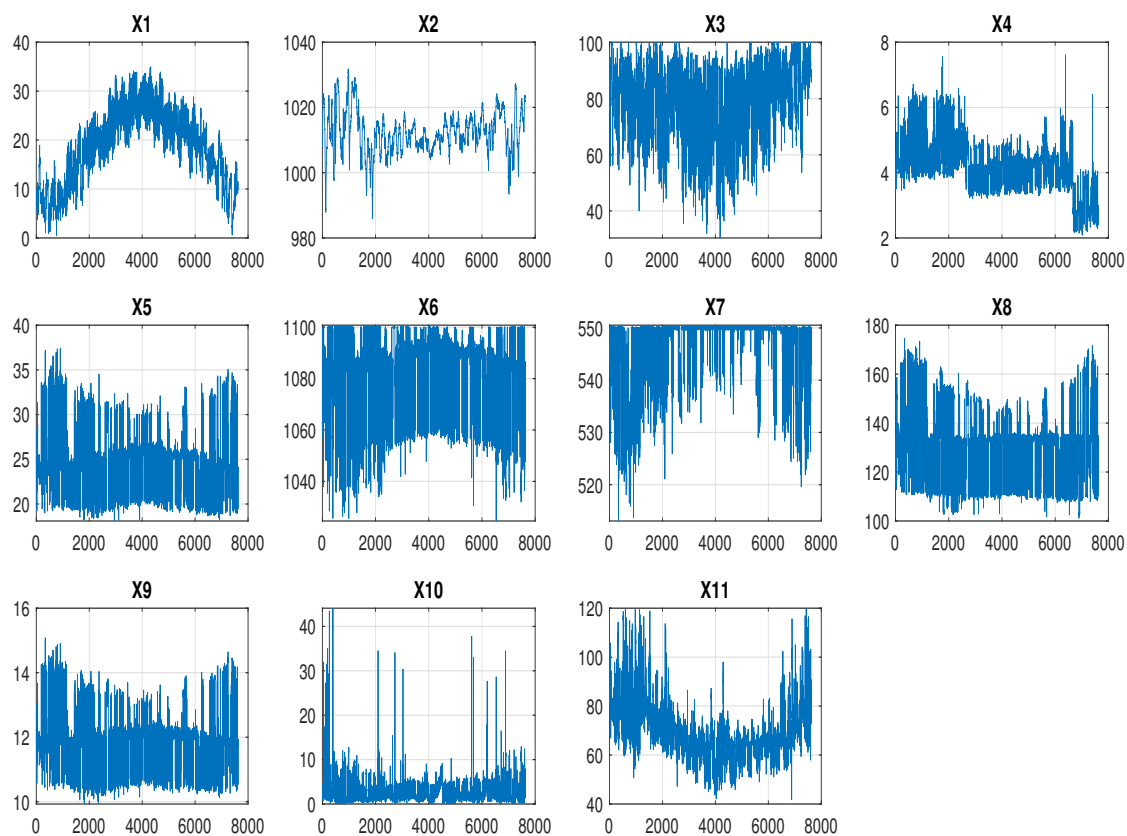


FIGURE 4.6 – Évolution des mesures des 11 capteurs

l'évolution des différentes variable après filtrage sur la figure 4.7

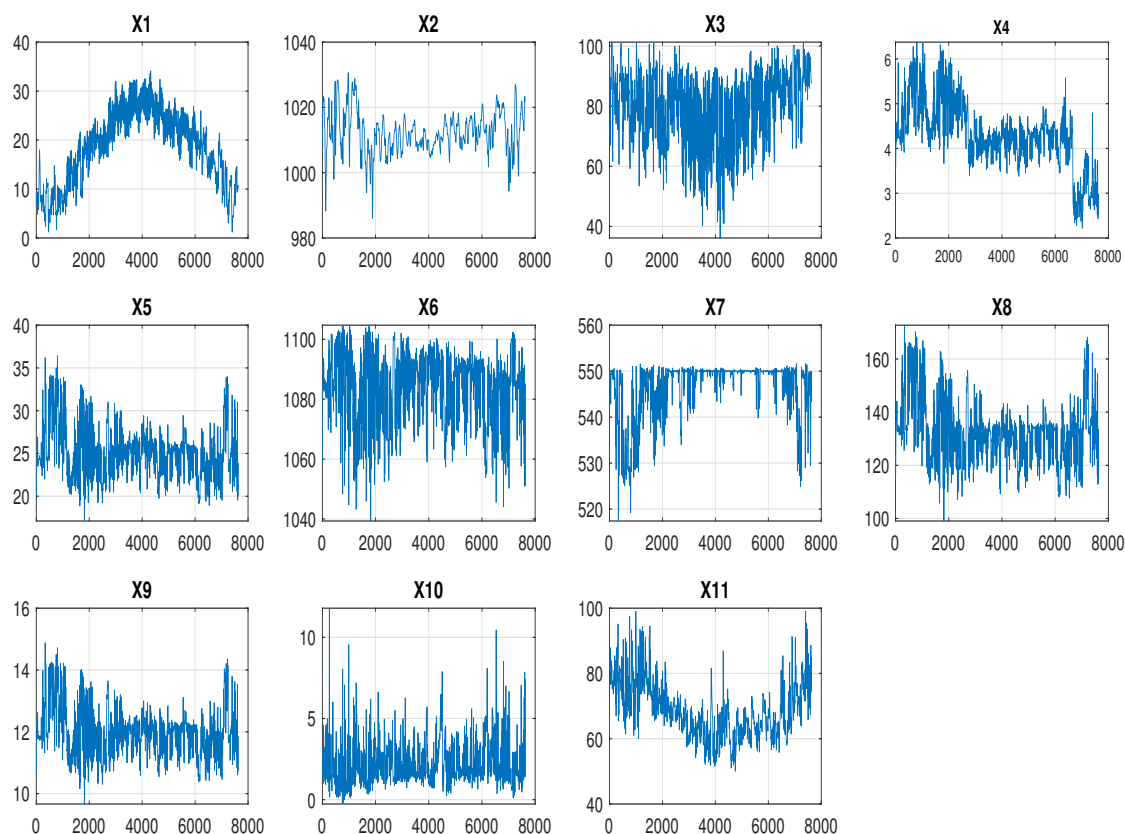


FIGURE 4.7 – l'évolution des différentes variable après filtrage

Afin de simuler l'incertitude de mesure, une imprécision ($\pm\Delta$) égale à 10% de la plage de variation a été ajoutée aux mesures de chaque capteur surveillé. Les données de type intervalle ainsi obtenues sont illustrées sur la figure 4.8.

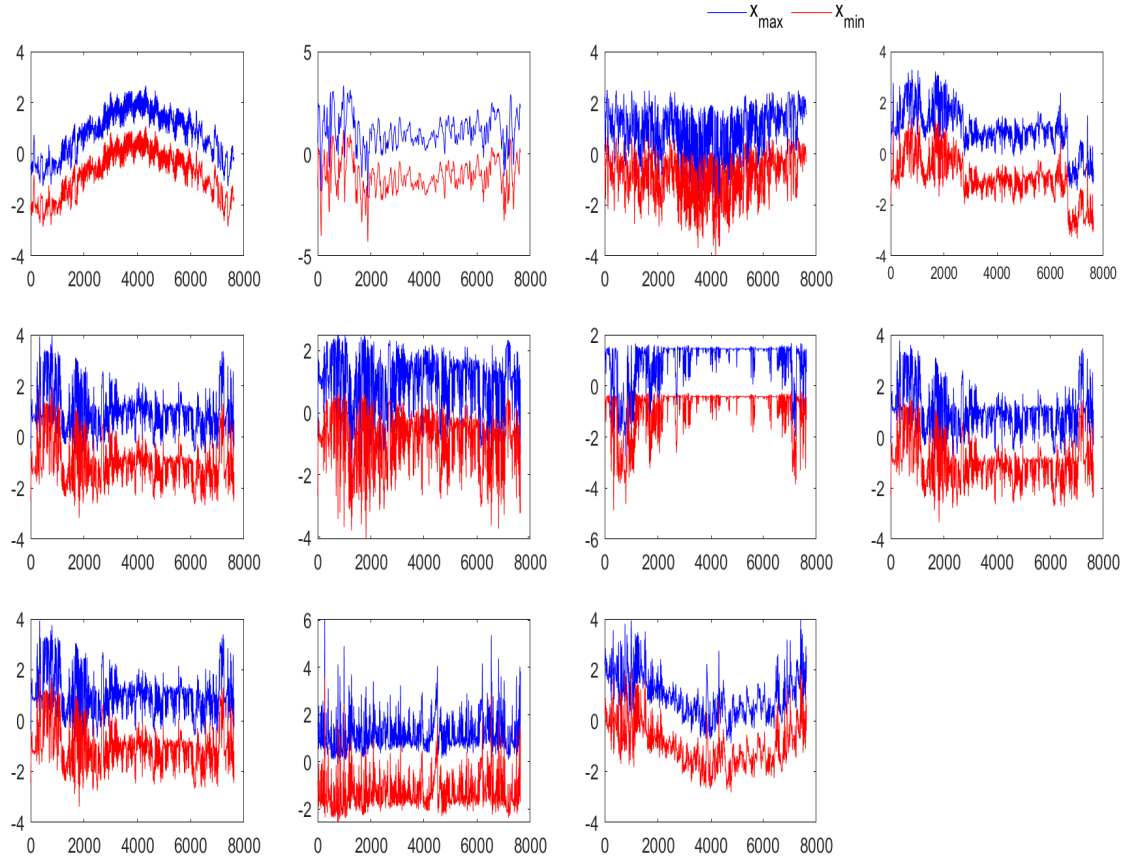


FIGURE 4.8 – Évolution des mesures de type intervalle

Après normalisation, l'ACP symbolique est utilisée pour modéliser ces données de type intervalle. Les mesures et leurs estimations obtenues avec cette méthode sont illustrées sur la figure 4.9. On observe que les estimations des données intervalles restent très proches des données d'origine, ce qui confirme la bonne représentativité du modèle SPCA construit.

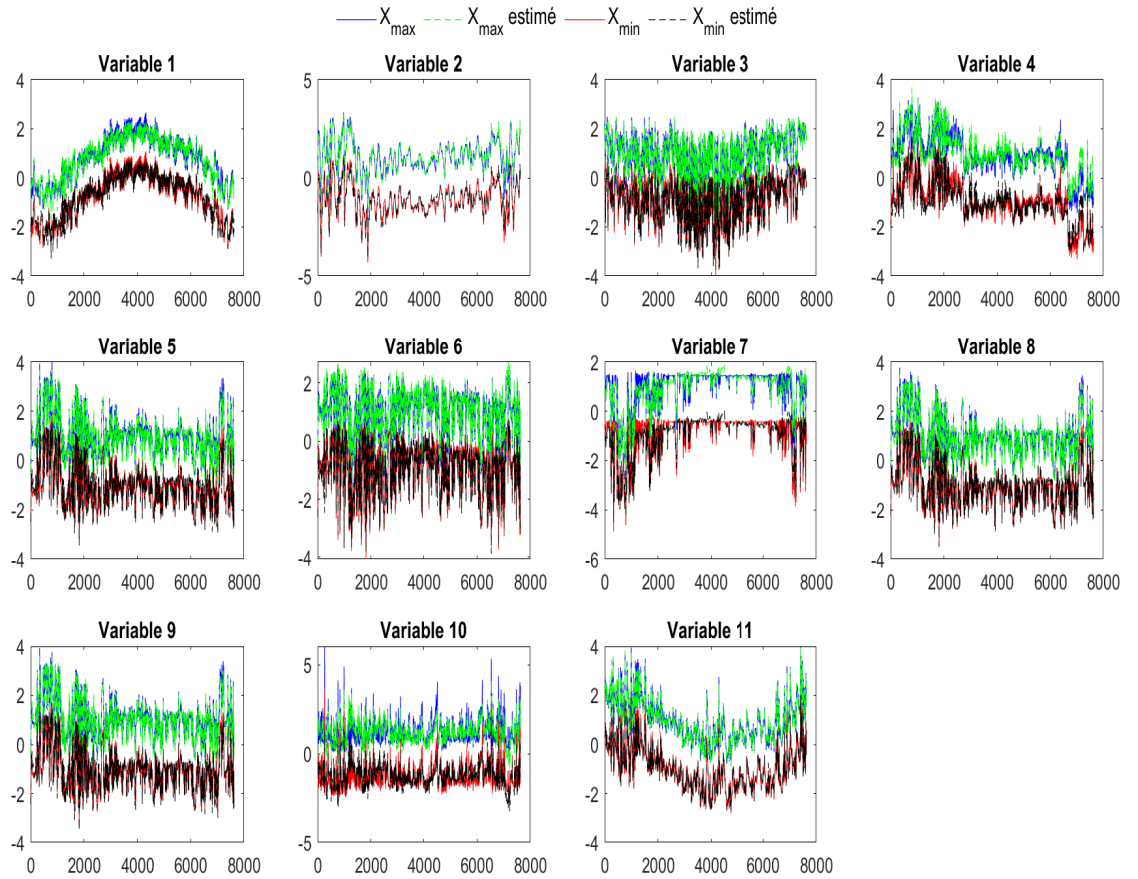


FIGURE 4.9 – Évolution des mesures et leur estimation

Diagnostic de défaut par ACP Symbolique

Une fois le modèle ACP déterminé, la phase de détection et de localisation des défauts peut être engagée. Un défaut de type biais, d'amplitude égale à 30% de la plage de variation de la variable x_4 , est simulé à partir de l'échantillon $k = 5000$. La figure 4.10 montre l'évolution des bornes $X_{\min}$ et $X_{\max}$ de la variable x_4 , avec et sans défaut.

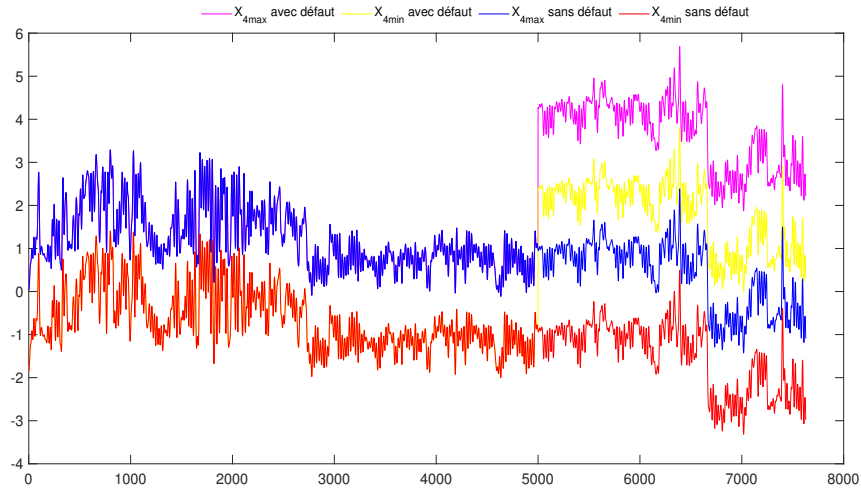


FIGURE 4.10 – Évolution de x_4 sans et avec défaut

La figure 4.11 présente l'évolution de l'indice de détection SPE, sans et avec défaut, sur x_4 . On observe que les indices SPE_{min} et SPE_{max} restent inférieurs au seuil de détection δ^2 en l'absence de défaut, confirmant le bon fonctionnement du modèle SPCA. À partir de l'instant $k = 5000$, les deux indices dépassent nettement le seuil, ce qui confirme la détection du défaut par la méthode SPCA.

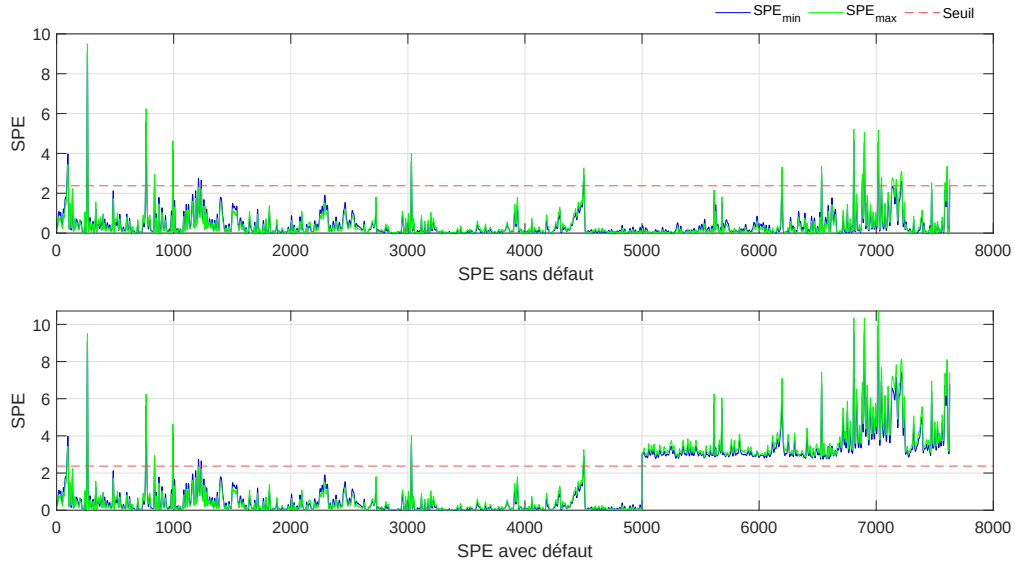


FIGURE 4.11 – Évolution de SPE sans et avec défaut

Localisation de défaut

Après la détection du défaut, il est nécessaire de localiser et d'isoler la variable fautive. La procédure de localisation, fondée sur le principe de reconstruction, a été appliquée à l'ensemble des variables ; les résultats obtenus sont présentés sur les figures suivantes.

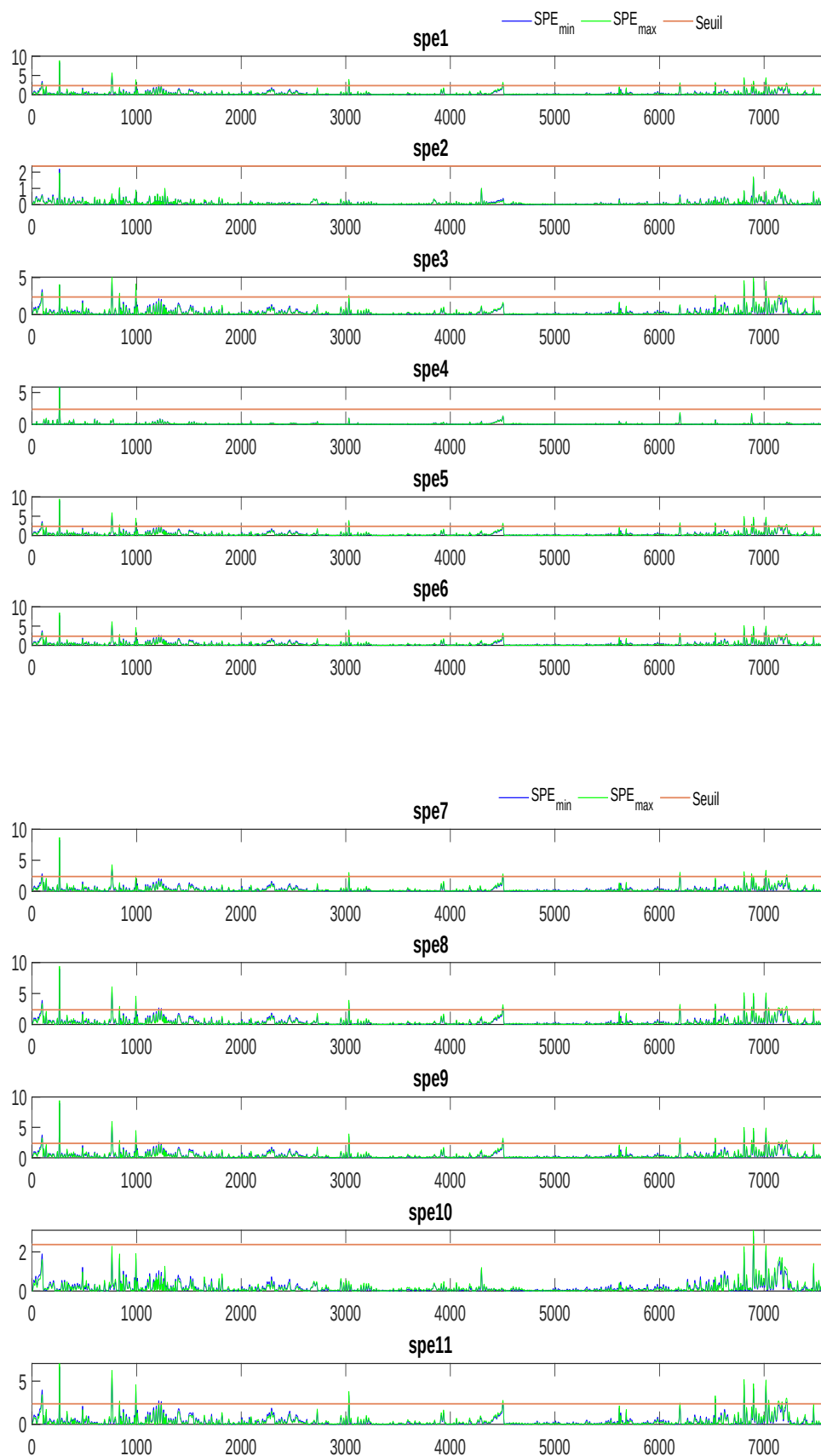


FIGURE 4.12 – SPE structuré en l'absence de défaut

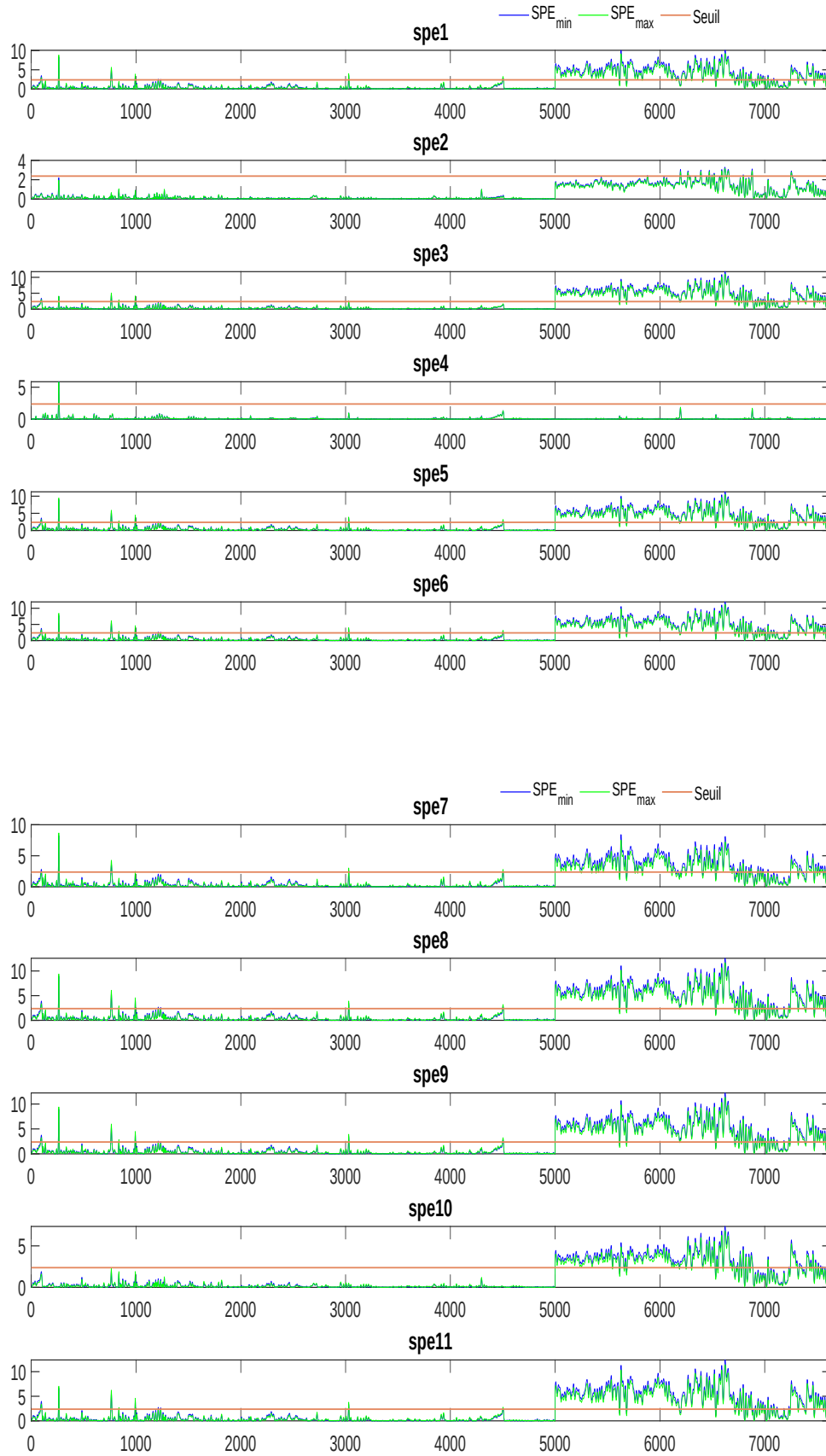


FIGURE 4.13 – SPE structuré en présence de défaut sur x_4

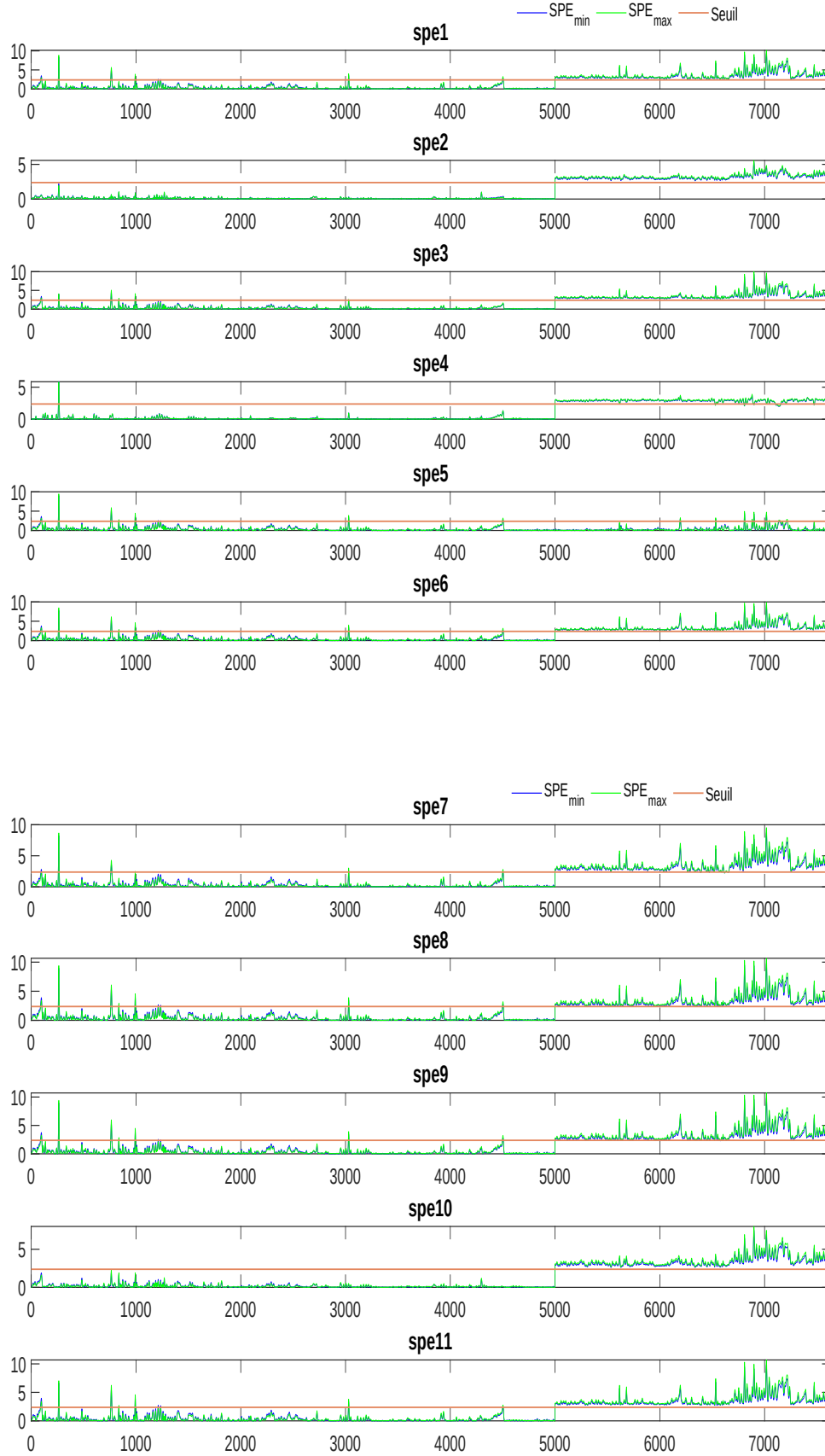


FIGURE 4.14 – SPE structuré en présence de défaut sur x_5

D'après les figures 4.12, 4.13 et 4.14, l'analyse de l'ensemble des indices SPE calculés pour les différents blocs, c'est-à-dire la signature de défaut, permet d'affirmer que la variable en défaut a été correctement identifiée par l'approche de reconstruction, confirmant ainsi la capacité de la méthode SPCA à localiser précisément l'origine d'une anomalie sur ce cas industriel réel.

4.6 Conclusion

Dans ce chapitre, nous avons présenté les étapes de modélisation et de diagnostic de défauts des capteurs par la technique de l'ACP par intervalle (SPCA), appliquée aux données de plusieurs capteurs d'une turbine à gaz. Les résultats obtenus ont montré que l'ACP par intervalle est capable de modéliser fidèlement le comportement d'un système industriel incertain, tout en assurant une détection rapide et une isolation précise des défauts affectant les capteurs, même en présence d'incertitudes de mesure. Contrairement au cas de l'ACP classique, qui suppose des mesures ponctuelles exactes, l'approche symbolique permet ainsi de traiter des données plus réalistes, représentatives des conditions réelles d'exploitation industrielle. Ce chapitre nous a ainsi permis de valider, sur un cas d'étude pratique et à partir d'un jeu de données benchmark, l'approche de diagnostic basée sur l'ACP développée dans ce travail.

Conclusion générale

L'analyse en composantes principales (ACP) est un outil statistique largement utilisé pour l'analyse de données, et exploité depuis longtemps dans de nombreux domaines de l'ingénierie. Dans ce travail, nous nous sommes intéressés au diagnostic des systèmes industriels à travers l'ACP et son extension aux données de type intervalle, dans une perspective de diagnostic prédictif adaptée aux exigences de l'industrie 4.0.

Nous avons d'abord présenté les généralités sur le diagnostic industriel, ainsi que les fondements théoriques de l'ACP classique, en abordant notamment la problématique de l'influence des observations sur le modèle. Nous avons ensuite montré les limites de cette approche lorsque les données sont entachées d'imprécisions, ce qui nous a conduits à explorer l'analyse de données symboliques, et plus particulièrement les variables de type intervalle. Dans ce cadre, nous avons développé et présenté l'approche SPCA (*Symbolic Principal Component Analysis*), qui étend le principe de l'ACP classique aux données de type intervalle, permettant une meilleure prise en compte de l'incertitude inhérente aux données réelles issues de capteurs industriels.

Nous avons ensuite abordé les méthodes de détection et de localisation de défauts fondées sur le principe de l'ACP, aussi bien pour les données classiques que pour les données de type intervalle, en nous appuyant respectivement sur l'indice SPE et sur le principe de reconstruction des variables. Enfin, le dernier chapitre a été consacré à l'application de cette étude à un cas industriel réel, à travers un benchmark de turbine à gaz, ce qui a permis de valider expérimentalement la pertinence et l'efficacité de l'approche SPCA pour la détection et l'isolation de défauts sur des données de capteurs incertaines.

Les résultats obtenus confirment que l'extension de l'ACP aux données de type intervalle constitue une réponse pertinente aux enjeux actuels du diagnostic industriel, dans un contexte où la multiplication des capteurs et l'imprécision inhérente aux mesures rendent les approches classiques insuffisantes. Cette approche s'inscrit ainsi pleinement dans la dynamique de l'industrie 4.0, où la surveillance et le diagnostic prédictif des équipements, à partir de données massives et incertaines, deviennent des leviers essentiels pour améliorer la fiabilité, la sécurité et la disponibilité des installations industrielles.

Comme perspectives à ce travail, il serait intéressant d'étendre l'approche SPCA à d'autres types de données symboliques (histogrammes, données catégorielles), d'explorer son couplage

avec des techniques d'intelligence artificielle pour la maintenance prédictive, ou encore de valider la méthode sur d'autres systèmes industriels complexes afin d'en confirmer la généralité.

Bibliographie

- [1] W. Bekiri, B. Sid, C. Chakour (2019). *Diagnostic de Fonctionnement par Analyse en Composantes Principales Multi-Echelles*. Thèse de doctorat.
- [2] Z. Sabeh (2006). *Diagnostic à base de modèle : application à un moteur diesel suralimenté à injection directe*. Thèse de doctorat, Institut National Polytechnique de Lorraine - INPL.
- [3] M. Dreibine (2014). *Diagnostic des systèmes par analyse en composantes principales*. Thèse de doctorat.
- [4] O. K. Ghezzal, O. E. Tarmoune, C. Chakour. *Modélisation et Diagnostic des Systèmes Incertains par Analyse en Composantes Principales*. Mémoire de master, Université Kasdi Merbah Ouargla.
- [5] R. T. Abdelatif (2019). *Diagnostic des systèmes non linéaires par les méthodes statistiques à base de fonctions à noyaux*. Thèse de doctorat.
- [6] M. Baligh (2012). *Analyse de données multi-variées et surveillance des processus industriels par analyse en composantes principales*. Thèse de doctorat.
- [7] A. Mahamoud (2010). *Observation et diagnostic de processus industriels à modèle non linéaire : application aux machines électriques*. Thèse de doctorat, Ecole Centrale de Nantes (ECN).
- [8] C. Chakour, M. F. Harkat, M. Djeghaba (2015). "Neuronal principal component analysis for nonlinear time-varying processes monitoring". *IFAC-PapersOnLine*, vol. 48, no. 21, p. 1408-1413.
- [9] D. Fragkoulis (2008). *Détection et localisation des défauts provenant des capteurs et des actionneurs : application sur un système non linéaire*. Thèse de doctorat, Université Paul Sabatier - Toulouse III.
- [10] K. Arar, S. Tadjine (2007). *Modélisation et Diagnostic des systèmes par l'Analyse en Composantes Principales Multi-Echelle (MSPCA)*. Faculté des Sciences de l'Ingénieur, Département d'Electronique, Annaba.

- [11] C. Chakour (2016). *Diagnostic et surveillance des procédés industriels et de leur environnement sur la base de l'analyse de données*. Thèse de doctorat, Université Badji Mokhtar Annaba. bibitemdun R. Dunia and S. J. Qin, "Subspace Approach to Multidimensional Fault Identification and Reconstruction," *AIChE Journal*, vol. 44, no. 8, pp. 1813–1831, 1998.
- [12] A. Philippot (2006). *Contribution au diagnostic décentralisé des systèmes à événements discrets : application aux systèmes manufacturiers*. Thèse de doctorat, Université de Reims Champagne Ardenne.
- [13] Auteur Anonyme. Comment j'ai réussi à prouver mon existence , thèse , 28 janvier 1986.
- [14] Mohamed-Faouzi Harkat, Noureddine Doghmane Modélisation et diagnostic des systèmes dynamiques par analyse en composantes principales multi-échelle (MSPCA). article 2007
- [15] Méthodes des Composantes Principales dans R : Guide Pratique, ACP - Analyse en Composantes Principales avec R : L'Essentiel ,15/10/2017 .[http ://www.sthda.com/french/articles](http://www.sthda.com/french/articles)
- [16] Harkat Mohamed-Faouzi Détection et Localisation de Défauts par Analyse en Composantes Principales, These Doctorat de l'Institut National Polytechnique de Lorraine le 30 Juin 2003
- [17] Jackson J. E., Mudholkar G. S. (1979). Control procedures for residuals associated with principal component analysis.
- [18] P. CAZES, A. CHOUAKRIA, E. DIDAY, Y. SCHEKTMAN : Extension de l'analyse en composantes principales à des données de type intervalle *Revue de statistique appliquée*, tome 45, no 3 (1997), p. 5-24
- [19] Tarek Ait-Izem " Diagnosis of Uncertain Systems using Principal Component Analysis". Thèse de doctorat. 2018. Université Badji Mokhtar Annaba .
- [20] E. Diday, "Introduction à l'approche symbolique en Analyse des Données", *Premières Journées Symbolique-Numérique*, Université Paris IX Dauphine, Paris, France, 1987.
- [21] L. Billard, E. Diday, *Symbolic Data Analysis : Conceptual Statistics and Data Mining*, John Wiley & Sons, Chichester, United Kingdom, 2006.
- [22] (Gas Turbine Engineering Handbook, 2021 ; Gas Turbine Theory, 2017).
- [23] Boyce, M. P. (2021). *Gas Turbine Engineering Handbook* (5th ed.). Elsevier.
- [24] H. Kaya, P. Tüfekci, and S. F. Gürgen, Gas Turbine CO and NOx Emission Data Set, UCI Machine Learning Repository, 2019. Available : [https ://archive.ics.uci.edu/dataset/551/gas+turbine+co+and+nox+emission+data+set](https://archive.ics.uci.edu/dataset/551/gas+turbine+co+and+nox+emission+data+set)