

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

Université Mouloud Mammeri De Tizi-Ouzou



Faculté De Génie Electrique Et D'informatique
DEPARTEMENT D'AUTOMATIQUE

**Mémoire de Fin d'Etudes
de MASTER ACADEMIQUE**
Spécialité : **Automatique et informatique industrielle**

Présenté par
**Amimar Tarek
Akbal Mohamed Amine**

Mémoire dirigé par **Mr Djennoune Said**

Thème

**Détection et localisation de défauts par
l'analyse en composantes principales :
Application à un système hydraulique**

Mémoire soutenu publiquement le 08 juillet 2018 devant le jury composé de :

Mr Hammouche Kamel

Professeur, UMMTO, Président

Mr Maidi Ahmed

Professeur, UMMTO, Examineur

Mr Mellah Rabah

Professeur, UMMTO, Rapporteur

Remerciements

Nous tiennons à exprimer nos sincères remerciements à Monsieur Djennoune Saïd, professeur à l'UMMTO, pour nous avoir proposé ce sujet, pour son soutien permanent et les nombreux conseils et informations apportées, mais surtout pour la confiance qu'il nous a accordés.

Nos remerciements vont également à Mr Hammouche Kamel, professeur à l'UMMTO pour nous avoir fait l'honneur d'accepter la présidence du jury de notre soutenance.

Nous adressons nos vifs remerciements à Mr Mellah Rabah, professeur à l'UMMTO pour nous avoir fait l'honneur d'examiner ce travail, ainsi que Mr Maida Ahmed, professeur à l'UMMTO, pour avoir accepté d'examiner ce travail.

Nos sincères gratitude vont aussi à tous les enseignants du département d'automatique qui ont contribué à notre formation.

Enfin, que toute personne ayant contribué de près ou de loin à la réalisation de ce travail, trouve ici l'expression de nos sincères remerciements.

Merci !

Mohemed Amine et Tarek

Dédicaces

Je dédie ce modeste travail à :

Ceux qui sont les êtres les plus chers pour moi, Mon père et ma mère
que je remercie infiniment pour Tout ce qu'ils ont fait pour moi durant toutes ces années
d'études et pour leurs réconforts.

Mes très chères sœurs ; en particulier les jumelles Asma et chaima.

Mes très chers frères A mes neveux et nièces

A mon binôme Amine.

Tous mes amies.

Tous ceux que j'aime que je n'ai pas cité.

Tous mes camarades de la promotion en 2ème année master D'Automatique.

Tarek

Je dédie ce modeste travail à :

Ceux qui sont les êtres les plus chers pour moi, à mes chers parents
que je remercie infiniment pour Tout ce qu'ils ont fait pour moi durant toutes ces années
d'études et pour leurs réconforts.

Ma très chère sœur et ses enfants.

Mes très chers frères

A mon binôme Tarek .

Tous mes amies.

Tous ceux que j'aime que je n'ai pas cité.

Tous mes camarades de la promotion en 2ème année master D'Automatique.

Amine

Résumé

L'analyse en composantes principales (ACP) est un outil extrêmement puissant pour la synthèse de l'information qui est très utile quand il y a une grande quantité de données quantitatives à traiter et interpréter. L'apparition au cours des dernières années de logiciel de plus en plus efficace et facile à utiliser rend maintenant ce type de l'analyse de données accessible à tous les chercheurs scientifiques, et pas seulement aux spécialistes seuls. C'est pourquoi nous proposons ici de présenter le principe et l'intérêt du PCA d'un exemple simple, celui d'une analyse de trois-cuves basée sur trois variables.

Mots-clés : Diagnostic, ACP, détection et localisation, contribution, reconstruction.

Abstract

Principal Component Analysis (PCA) is an extremely powerful tool for synthesizing information that is very useful when there is a large amount of quantitative data to process and interpret. The appearance in recent years of increasingly efficient and easy-to-use software now makes this type of data analysis accessible to all scientific researchers, and no longer to specialists alone. This is why we propose here to present the principle and the interest of the PCA from a simple example that of a three-tanks analysis based on three variables.

Keywords : diagnosis, PCA, fault detection and localisation, contribution, reconstruction.

Notations

ACP : Analyse en composantes principales

$X \in \mathfrak{R}^{n \times m}$: Matrice de données

$x_i \in \mathfrak{R}^m$: Vecteur d'observation (individu) de mesures des variables de X

$x_{ci} \in \mathfrak{R}^m$: Vecteur des coordonnées centrées de l'individu x_i

$x^j \in \mathfrak{R}^n$:

$G \in \mathfrak{R}^m$: Centre de gravité

$Y \in \mathfrak{R}^{n \times m}$: Matrice de données centrées

$R \in \mathfrak{R}^{n \times m}$: Matrice de données normalisées

$D \in \mathfrak{R}^n$: Matrice des poids

$\Sigma \in \mathfrak{R}^{m \times m}$: Matrice de corrélation de X

$V \in \mathfrak{R}^{m \times m}$: Matrice de variance de X

$T \in \mathfrak{R}^{n \times m}$: Matrice de composantes principales

$t_i \in \mathfrak{R}^m$: Vecteur d'observation de composantes principales de X

w_j : Facteur principal

$P \in \mathfrak{R}^{m \times m}$: Matrice des vecteurs propres

p^j : Vecteur propre

$\Lambda \in \mathfrak{R}^{m \times m}$: Matrice des valeurs propres

λ_j : Valeur propre

M : Métrique

SP : Espace principale

SR : Espace résiduel

\hat{C} : Matrice de projection sur SP

\tilde{C} : Matrice de projection sur SR

\hat{X} : Partie modélisée

\tilde{X} : Partie résiduel

n : Nombre d'individus

m : Nombre de variables

ℓ : Nombre de composantes à retenir

Table des matières

Introduction générale	10
1 Diagnostic	13
1.1 Objectifs du diagnostic	14
1.2 Surveillance des processus	15
1.2.1 Objectif de la surveillance du processus	15
1.2.2 Principe de la surveillance	15
1.3 Supervision des processus	15
1.3.1 Les étapes de la supervision	16
1.4 Des concepts et de la terminologie	16
1.5 Différentes structures de défauts	18
1.5.1 Nature des défauts	18
1.5.1.1 Défauts multiplicatifs	18
1.5.1.2 Défauts additifs	19
1.5.2 Types de défauts	19
1.5.2.1 Défauts actionneurs	19
1.5.2.2 Défauts capteurs	19
1.5.2.3 Défauts systèmes	20
1.6 Principe et classification des méthodes de diagnostic de défauts	20
1.6.1 Principe du diagnostic des défauts	20
1.6.1.1 Redondance Analytique	21
1.6.1.2 Redondance matérielle	22
1.6.2 Présentation des méthodes de diagnostic	22
1.6.2.1 Méthodes sans modèles	23
1.6.3 Méthodes avec modèles	27
1.6.3.1 Approches qualitatives	27
1.6.3.2 Approches quantitatives	28
1.7 Conclusion	28
2 Analyse en Composantes Principales	30
2.1 Introduction	30

2.2	Avantages et inconvénients de l'ACP	31
2.2.1	Avantages	31
2.2.2	Inconvénients	32
2.3	Principe de l'analyse en composantes principales	32
2.4	Algorithme général	32
2.5	Préparation des données	33
2.5.1	Tableau des données	33
2.5.2	Centrage et pondération des données	33
2.6	Représentation du nuage de points	36
2.7	Analyse du nuage des individus	36
2.7.1	Choix de la distance	36
2.7.2	Méthode de projection	38
2.7.2.1	L'utilité de la projection	39
2.7.3	Projections orthogonales	39
2.7.4	Matrice de variance-covariance	39
2.7.5	Matrice de corrélation	41
2.7.5.1	Étude de la matrice de corrélation	41
2.7.6	Moments d'inertie	41
2.7.6.1	Inertie du nuage des individus	42
2.7.6.2	Inertie du nuage par rapport à un axe	42
2.7.7	Éléments principaux	44
2.7.7.1	Axes principaux	44
2.7.7.2	Facteurs principaux	46
2.7.7.3	Composantes principales	47
2.8	Interprétation des résultats d'une ACP	47
2.8.1	Étude des variables	47
2.8.1.1	Corrélation entre les CPs et les variables d'origine	48
2.8.1.2	Représentation des variables dans les nouveaux axes	48
2.8.1.3	Études des liaisons entre les variables	49
2.8.2	Étude des individus	49
2.8.2.1	Représentation des individus dans les nouveaux axes	50
2.8.2.2	Qualité de représentation des individus	50
2.8.2.3	Interprétation des nouveaux axes en fonction des individus	51
2.9	Analyse en composantes principales normées	52
2.10	Conclusion	52
3	Diagnostic par l'ACP	54
3.1	Identification du modèle ACP	55
3.2	Choix de la dimension	58

3.2.1	Pourcentage Cumulé de la Variance (<i>PCV</i>)	58
3.2.2	Critère de Kaiser	59
3.2.3	Variance de l'Erreur de Reconstruction(<i>VER</i>)	60
3.2.3.1	Approche itérative de reconstruction	60
3.2.4	Validation croisée	63
3.3	Détection de défauts	64
3.3.1	L'erreur de prédiction SPE	65
3.3.2	Statistique T^2 de Hotelling	66
3.4	Localisation de défauts	67
3.4.1	Localisation par calcul des contributions	67
3.5	Conclusion	67
4	Application à un système hydraulique	69
4.1	Avant-propos	69
4.2	Objectifs de l'ACP dans notre application	69
4.3	Procédé	70
4.3.1	Description du procédé	70
4.3.2	Modèle	71
4.3.3	Génération des données	73
4.3.4	Simulation du modèle	73
4.4	Réduction de la dimension	73
4.4.1	Tableau du données	74
4.4.2	Normalisation	75
4.4.3	Matrice de corrélation	77
4.4.4	Les vecteurs et les valeurs propres	78
4.4.5	Cordonnées factorielles des individus	79
4.4.6	Choix de nombre d'axe	81
4.5	Interprétation des axes	82
4.5.1	Étude des variables	82
4.5.2	Étude des individus	84
4.6	Modélisation par l'ACP	87
4.7	Détection et localisation	89
4.8	Conclusion	96
	Conclusion générale	97

Table des figures

1.1	Différents types affectant un système physique	19
1.2	Étapes de diagnostic.	21
1.3	Classification des méthodes de diagnostic.	23
2.1	Cercle de corrélation	49
3.1	Nombre de valeurs propres	60
4.1	Système à trois bacs.	70
4.2	Les niveaux dans les trois bacs	73
4.3	Les composantes principales	81
4.4	Graphique de valeurs propres	82
4.5	Projection des variables sur le plan factoriel (1-2)	84
4.6	Projection des individus sur les axes factoriels 1-2	85
4.7	mesures des qualités de représentation des variables sur les axes factoriels .	86
4.8	Contribution des variables à l'axe 2	86
4.9	Contribution des variables à l'axe 3	87
4.10	L'évolution des erreurs d'estimations des mesures sans défauts	89
4.11	L'évolution des mesures affectés par F1	90
4.12	L'évolution des mesures affectés par F2	90
4.13	L'évolution des mesures affectés par F3	91
4.14	L'évolution de SPE en absence de défaut	93
4.15	L'évolution de SPE en présence de F1	93
4.16	L'évolution de SPE en présence de F2	94
4.17	L'évolution de SPE en présence de F3	94
4.18	Contributions à l'indice SPE en présence de F1	95
4.19	Contributions à l'indice SPE en présence de F2	95
4.20	Contributions à l'indice SPE en présence de F3	96

Liste des tableaux

- 4.1 Les variables du système 74
- 4.2 Tableau des données 75
- 4.3 Données centrées 76
- 4.4 Données réduits 77
- 4.5 Matrice de corrélation 78
- 4.6 Les vecteurs et les valeurs propres 79
- 4.7 Valeurs propres des axes et leurs contributions 79
- 4.8 Les composantes principales 80
- 4.9 Corrélation variables-facteurs 83
- 4.10 Les coordonnées factorielles des variables 83
- 4.11 Données reconstruites 88
- 4.12 L'effet de F1 sur les résidus 91
- 4.13 L'effet de F2 sur les résidus 92
- 4.14 L'effet de F3 sur les résidus 92

Introduction générale

Depuis la naissance de l'industrie, les procédés industriels présentent toujours des défaillances intrinsèques (propres ou matérielles) et des défaillances extrinsèques (dues à l'environnement)[1]. Pour-cela, il est absolument important de diagnostiquer et gérer la maintenance des systèmes tout en ayant une expertise dans la maintenance, vu l'importance de savoir mettre l'accent sur la règle de priorité à traiter les défaillances puisque certaines défaillances ont moins d'importance en termes d'effets et de couts. Alors avoir une connaissance sur le système à diagnostiquer est indispensable afin de choisir la méthode idéale pour diagnostiquer le processus dans le but de lui assurer la sureté des matériels et celle des personnels, et cela grâce au diagnostic qui consiste à détecter une ou plusieurs défaillances, à localiser son origine et à déterminer ses causes, sachant que ces fonctions apportent des corrections au système afin qu'il puisse retourner à son fonctionnement normal.

L'objectif d'apporter une solution efficace et assurer au processus industriels un fonctionnement nominal dans le cadre de respecter les contraintes économiques et celles de la qualité des produits, le diagnostic prouve un grand intérêt au monde industriel que celui de la recherche scientifique. Par conséquent l'avancée technologique rend les systèmes industriels de plus en plus complexes, comme l'automatisation et plusieurs d'autres facteurs qui optent pour leur complexité que ce soit dans leurs incertitudes, l'imprédictibilité et leurs dynamiques. Cela rend les décisions difficiles[2]. Pour les processus complexes ou ceux de types de boites noires, leur modélisation mathématique est difficile même dans le cas où la réalisation de tels modèles est possible, les approches analytiques présentent une description insuffisante du fonctionnement du processus. Pour la résolution de ce problème, les méthodes de l'analyse statistique des données ont un rôle majeur pour la visualisation de l'évolution en temps réel d'un processus afin de détecter les changements susceptibles qui affectent ses performances.

La masse des données et la dimension des variables que présentent les systèmes sont tellement grandes, ce qui crée l'inefficacité en terme d'isolation de défaut. Pour cela, la réduction de la dimension de l'espace des variables en utilisant les méthodes de projection, comme l'analyse en composantes principales (ACP) peut révéler des information cachées

mieux interprétables[3].

L'objectif de la réduction de données pour une analyse statistique multivariée est de rendre la visualisation plus claire tout en gardant le principal d'informations, afin de pouvoir constater et interpréter un dérèglement. La plus part des méthodes qui réduisent la dimension d'un espace multivarié optent pour l'utilisation des variables latentes (cachées), dans le but de comprendre un intérêt caractéristique qui ne peut pas être mesuré directement. Ces variables latentes expriment la combinaison linéaire des variables originelles. Étant donné, l'analyse en composantes principales est une méthode révélatrice d'informations latentes et particulièrement adaptée à l'analyse des données corrélées.

L'ACP a une place importante pour la surveillance et le diagnostic des défauts, et cela grâce à sa capacité de créer des corrélations linéaires entre les variables du processus malgré l'absence d'un modèle d'entrée/ sortie propre au système. Cette approche représente un outil de modélisation du comportement des processus, plus particulièrement en fonctionnement normale, les défauts alors sont détectés en comparant le comportement observé à celui donné par le modèle ACP. Plus précisément, la phase de détection est liée à une étape génératrice de résidus, ou plutôt d'indices de détection basée sur une estimation paramétrique[4].

L'objectif de notre travail est d'implémenter sur ordinateur la méthodes ACP linéaires pour la détection et la localisation de défauts d'un processus physique. Nous appliquons en suite à un système hydraulique constitué de trois cuves en cascade.

L'organisation de notre mémoire est divisé en quatre chapitre d'une manière suivante :

Dans le premier chapitre, nous exposerons les différents concepts liés au diagnostic définis dans la littérature, et employés donc dans notre mémoire. Nous présentons aussi les théories de la surveillance et de la supervision et du diagnostic. Par la suite, en aborderons les différentes méthodes de diagnostic des défaillances développées par les différentes communautés de recherche.

Le deuxième chapitre sera consacré aux rappels mathématiques de l'ACP linéaire et son principe.

Le chapitre trois est consacré à l'identification du modèle ACP qui repose sur le calcul des valeurs et des vecteurs propres, et la détermination du nombre de composantes principales à retenir dans le modèle ACP.

Une fois le modèle ACP identifié, alors la phase de prise de décision pour la détection et la localisation de défauts fera l'objet des méthodes consacrées à la détection et la localisation de défauts par une analyse en composantes principales . Nous avons deux indices pour la détection de défauts par ACP, qui sont la statistique T^2 de Hotelling calculée à partir des premières composantes principales, et l'erreur quadratique d'estimation (SPE).

Pour illustrer notre travail, et prouver une légère maîtrise que se soit dans le diagnostic, la construction du modèle ACP, et la détection de défauts par l'ACP, le chapitre quatre est consacré à la concrétisation de toute l'étude que nous avons mené sur une application (système hydraulique).

Dans ce chapitre dans une première partie, la description et la modélisation du système hydraulique constitué de trois cuves en cascade sont exposées. Les défauts que nous avons considérés sont des fuites qui peuvent surgir dans les liaisons entre les trois cuves. Nous présentons ensuite les résultats de simulation obtenus par l'ACP concernant la détection et la localisation des défauts considérés.

Chapitre 1

Diagnostic

Introduction

Les travaux de recherche sur le diagnostic ont mobilisé durant ces dernières années une large communauté de chercheurs. Ils ont été un sujet important dès l'apparition de l'automatique moderne, basé sur le calcul numérique[5].

Aujourd'hui, en raison des contraintes économiques, de compétitivité accentuée par la mondialisation d'une part et de la sûreté de fonctionnement d'autre part, le diagnostic a connu une évolution très importante surtout pour les processus chimiques, nucléaires ...etc.

Cette évolution a permis le développement de plusieurs méthodes, donnant une solution aux problèmes de la détection des défauts pour assurer la sûreté de fonctionnement, et prendre une décision permettant d'améliorer la conduite du système, ou d'empêcher une dérive pouvant mener à des échecs de production. Le souci de la protection de l'environnement est aussi une raison supplémentaire de l'importance du diagnostic de ce type de procédé. Une procédure de diagnostic doit donc apporter des réponses aux questions suivantes en ce qui concerne ses objectifs, ses principes de mise en œuvre et ses critères d'évaluation :

- Objectifs : Que veut-on contrôler ? Quels types de défauts doit-on détecter ?
- Principes : Quel est le principe de diagnostic à mettre en œuvre ? Quelle est la procédure à suivre ?
- Critères : Quelles sont les performances souhaitées ? Quels sont les indices d'évaluation de ces performances ?

Cependant, la conception d'une procédure de diagnostic doit prendre en compte des aspects tels que le comportement du système et la sensibilité aux erreurs de mesure ainsi que les contraintes d'ordre économique et de mise en œuvre pratique. Par conséquent et vue l'importance du problème de diagnostic et des difficultés rencontrées, plusieurs approches ont été développées au cours de ces dernières années. Elles couvrent une large

variété de techniques telles que les approches sans modèles. Ces approches exigent un grand volume des données. Ces méthodes peuvent être de type quantitatif qui sont basées sur le traitement de données ou bien de type qualitatifs, exprimées à base de connaissance.

Nous proposons dans ce chapitre de faire un état de l'art sur le diagnostic des défauts dans les systèmes lorsque un raisonnement sur les comportements dynamiques du procédé n'est pas utile.

L'intérêt se portera essentiellement sur le problème de détection des défauts et leur localisation par la méthode d'analyse en composantes principales. Cette approche représente un outil de modélisation du comportement de processus plus particulièrement en fonctionnement normal, tel que, les défauts sont alors détectés en comparant le comportement observé à celui donné par le modèle ACP. Plus précisément, la phase de détection de défaut liée à une étape génératrice de résidus ou plutôt d'indice de détection basés sur l'estimation paramétrique[4]. Ainsi, la localisation en composantes principales comporte trois approches qui sont : l'approche par structuration des résidus qui s'inspire des méthodes de diagnostic utilisant la redondance analytique, les approches utilisant des bancs de modèles, et l'approche reposant sur le principe de calcul des contributions à l'indice de détection (SPE). Cette dernière est la méthodes la plus efficace parmi les autres citées, elle consiste à calculer la contribution des différentes variables à l'indice de détection. La variable ayant la plus grande contribution à la statistique de détection calculée à l'instant de détection est la variable incriminée. Donc il est nécessaire dans un premier temps, de donner les différents concepts, notions et outils cités dans la littérature et qui sont utilisés en diagnostic de défauts, puisque un bon diagnostic demande une bonne compréhension de ces notions. Le principe de base sera alors donné ainsi que les différentes méthodes proposées dans ce domaine.

1.1 Objectifs du diagnostic

Dans l'industrie, la compétitivité s'appuie sur plusieurs critères de performance, de robustesse et de qualité. Cependant même si ces critères sont réunis les défaillances des systèmes industrielles se font fréquentes et peuvent résulter soit d'une panne partielle ou totale qui peut présenter des pertes économiques considérables, sans parler de l'environnement et des dangers pour les personnes, les salariés de l'entreprise en premier lieu. Pour éviter des conséquences irréversibles, des méthodes de supervision et de surveillance ont été mise en place et qui reposent toutes sur le diagnostic des systèmes[7].

Le diagnostic industriel dont la vocation première est de détecter et localiser une défaillance des matériels propres aux systèmes industriels, permet de détecter de façon précoce une dégradation[1]. Plus généralement, le diagnostic apporte une contribution à

la détection et la localisation d'un mauvais fonctionnement.

1.2 Surveillance des processus

La tendance actuelle d'économie d'énergie et de l'environnement pousse à une optimisation de la plupart des processus industriels. Cela engendre naturellement une complexification notable des systèmes, ce qui accroît par conséquent le risque de dysfonctionnement. De ce fait, la surveillance des systèmes industriels est un domaine actuellement plein essor.

1.2.1 Objectif de la surveillance du processus

L'objectif de la surveillance c'est de permettre au processus industriels une longue durée d'activité sans présenter des interruptions à cause de la panne par exemple, pour les systèmes électrique les surtension et autres tensions transitoires élevées qui sollicitent beaucoup d'équipements et peuvent provoquer des pannes[8]. Les tensions transitoires apparaissant au cours du fonctionnement d'un système électrique sont influencées par le réseau d'interconnexion, les câbles, le fonctionnement du disjoncteur, ..., etc.

1.2.2 Principe de la surveillance

La thématique de la surveillance est née avec l'industrie et depuis toujours plusieurs recherches sont mises en place afin de développer cette théorie. Pour la surveillance des systèmes industriels, il est indispensable de connaître toutes les informations relatives au système et aux types de défaut(s) qui peuvent se manifester sur le système[9]. En effet, deux catégories de surveillance existent : celles qui sont basées sur un modèle qui décrit le comportement et la dynamique du système, et celles que seules les données acquises sur le processus permettant de caractériser son mode de fonctionnement. Vu l'avancée technologique et le développement des systèmes industriels, les processus devient de plus en plus complexe ; alors la robustesse de la surveillance doit être indispensable devant l'incertitude et les erreurs qui entachent les modèles et les données.

1.3 Supervision des processus

La supervision joue un rôle primordial dans le contexte économiques de productivité et de flexibilité, elle se développe en parallèle avec l'avancée technologique, elle est née du besoin de surveiller des processus industriels. Initialement, se composait d'un tableau mural comportant des boutons et des voyants permettant la vision des opérations des systèmes industrielles.

La fonction de la supervision est d'offrir d'information sur l'état des systèmes en temps réel pour pouvoir bénéficier de prendre des décisions et assurer le bon fonctionnement du système et maintenir ces critères de performances (qualité, sûreté, etc).

L'objectif principale de la supervision est d'attirer l'attention de l'opérateur dès l'apparition d'un ou plusieurs évènements susceptibles d'affecter le bon fonctionnement du système et cela par des signaux d'alarmes[11].

En effet, vu le développement des systèmes industriels et leurs complexité, les alarmes ont une importance cruciale et non négligeable pour assurer la sécurité de l'environnement industriel et assurer le fonctionnement nominal des ses systèmes.

1.3.1 Les étapes de la supervision

Dans un processus industriel l'exécution tâche ou d'une opération par l'homme ou par la machine est commandée et surveillée par le système de supervision qui comporte plusieurs outils et de méthodes de conduire des systèmes industriels tant en fonctionnement normale qu'en présence de défaillance ou d'une perturbation.

Plusieurs propositions sur les étapes de supervision sont données par la littérature, par exemple Breuker et Van De Velse (1994)[12] suggère que les étapes de supervision se succèdent en une chaîne de planification, prédiction, surveillance, diagnostic, et affectation des tâches infaisables du processus et la manipulation des dispositifs. Sauf que certaines tâches sont infaisables en ligne. Alors, Acosta et Al (2001)[13] ont utilisé une représentation plus détaillée et plus globale, comporte huit étapes qui sont : la surveillance, le diagnostic, l'évaluation d'état, le mode de fonctionnement, le pronostic, la planification, interface HM et la validation des données.

Même si il y a plusieurs propositions, généralement la supervision prend trois étapes principales qui sont :

- Étape de surveillance : traite les données recueillies en ligne afin d'obtenir l'état de fonctionnement du processus.
- Étape de diagnostic : le diagnostic consiste à estimer les causes des anomalies afin de pouvoir prendre des mesures de correction sur le système.
- Étape de configuration : cette étape consiste à prendre les bonnes décisions pour corriger les défaillances.

1.4 Des concepts et de la terminologie

Il existe plusieurs leviers permettant une optimisation optimale des systèmes industriels et qui assurent la disponibilité de l'outil de production à savoir que la disponibilité est une des composants de la sûreté de fonctionnement qui peut se décliner en termes de

fiabilité, disponibilité, maintenabilité et sécurité, c'est la tâche majeure pour un ingénieur.

Zwingelstein a défini ces caractéristiques comme suit :

- Fiabilité : c'est l'aptitude d'un système à accomplir sa mission dans des conditions données d'utilisation.
- Disponibilité : c'est l'aptitude d'un système à fonctionner lorsqu'on le sollicite.
- Maintenabilité : c'est l'aptitude d'un système à être entretenu ou remis en marche.
- La sûreté : c'est l'aptitude d'un système à respecter l'utilisateur et son environnement.

Afin de développer notre propos, nous présentons d'autres terminologies et concepts de base dans le domaine du diagnostic. La plus part de ces définitions ont été discutées au sein du SAFEPROCESS Technical Committee of International Federation of Automatic Control IFAC[14]

- Fonctionnement normal : un système est en fonctionnement normal lorsque ses états, ses entrées et ses sorties sont très proches de leur valeur nominale.

Typiquement, il y a fonctionnement normal lorsqu'il n'y a pas d'apparition de défauts.

- Erreur : elle est définie comme l'écart entre une valeur mesurée ou estimée d'une variable et une valeur donnée par un modèle et qui est théoriquement correcte.
- Défaut : un défaut est généralement défini comme une déviation d'une variable observée ou d'un paramètre calculé par rapport à sa valeur fixée dans les caractéristiques attendues du processus lui-même, des capteurs, des actionneurs ou de tout autre équipement.
- Détérioration : ce terme est utilisé pour définir une perte de performance altérant partiellement le fonctionnement du système.
- Défaillance : une défaillance correspond à un défaut ayant un effet néfaste sur le fonctionnement du système. Ce terme est utilisé lorsqu'un dispositif devient inopérant ou il accomplit difficilement sa fonction.
- Panne : est une interruption permanente de la capacité du système à réaliser sa fonction requise. Elle est au-delà de la défaillance car elle implique l'arrêt total du système.
- Surveillance : c'est une tâche continue, réalisée en temps réel, de détermination de l'état d'un système physique qui consiste en l'enregistrement des informations ainsi la reconnaissance et l'indication des anomalies du comportement.
- Seuil : c'est la valeur limite de l'écart d'un résidu avec le zéro, ainsi s'il est dépassé, un défaut est déclaré comme détecté.
- Résidu ou indicateur de défaut : il exprime l'incohérence entre les informations disponibles et les informations théoriques fournies par un modèle. C'est la définition qui a été retenue par les instances nationales et internationales de normalisation (AFNOR, CEI).
- Diagnostic : le diagnostic est l'identification de la cause probable de la (ou des)

défaillance(s) à l'aide d'une inspection, d'un contrôle ou d'un test.

- Supervision : c'est la surveillance d'un système physique et la prise de décisions appropriées en vue de maintenir son opération lors de l'apparition de défauts.
- Détection de défaut : c'est la détermination de la présence de défaut et de l'instant de son occurrence.
- Isolation de défaut : c'est la détermination de la taille et du comportement temporel d'un défaut.
- Identification de défaut : c'est la détermination de taille et du comportement temporel d'un défaut.

Il est important de savoir distinguer entre symptôme et signe. Le travail effectué par [15] et [16] marque cette différence en donnant des définitions précises.

Signe : un défaut est perceptible au travers d'un signe, caractère manifestant un état comportemental anormal. Un signe est caractérisé par un ensemble d'observation en provenance d'un système physique et est révélateur de la présence d'un défaut.

Symptôme : le pendant de signe dans le cas d'une défaillance correspond au terme symptôme, définit comme suit : caractère distinctif d'un état fonctionnel anormal.

La distinction entre les termes symptôme et signe s'apparente donc à la différence entre notions de défaut et de défaillance. Un symptôme révélateur d'une défaillance est nécessairement un signe, alors qu'un défaut se manifeste au travers de signes qui ne sont pas nécessairement des symptômes[17].

1.5 Différentes structures de défauts

Généralement, un défaut est caractérisé par une déviation de fonctionnement normal d'un système qui est circonvenue soit par les signaux de commande ou les signaux de mesures. Les défauts affectant un système sont d'évolutions, de natures et de types différents.

1.5.1 Nature des défauts

Les défauts peuvent être classés en défauts de nature multiplicative ou nature additive, selon leurs effets sur les performances du système.

1.5.1.1 Défauts multiplicatifs

Les défauts au niveau de la dynamique du procédé sont modélisées par des défauts multiplicatifs. il correspondent aux modifications paramétriques du modèle représentant

le système. Ces derniers induisent des changements sur la corrélation du signal de sortie du système, ainsi que des changements dans la dynamique du système.

1.5.1.2 Défauts additifs

Ces défauts sont modélisés sous forme de termes additifs dans le modèle du système. Ils influent sur état ou sur sa sortie. Cette modélisation est habituellement attribuée aux défauts de capteurs et d'actionneurs[18].

1.5.2 Types de défauts

Le concept de défaut est important dans les opérations de surveillance pour la conduite et la maintenance des processus industriels. ils peuvent être purement aléatoires (défauts cataleptiques), ou progressifs dans le cas du vieillissement ou de l'usure (défauts évolutifs).

Comme étant présenté dans la (1.1) , trois types de défauts peuvent toucher les différents éléments d'un système :

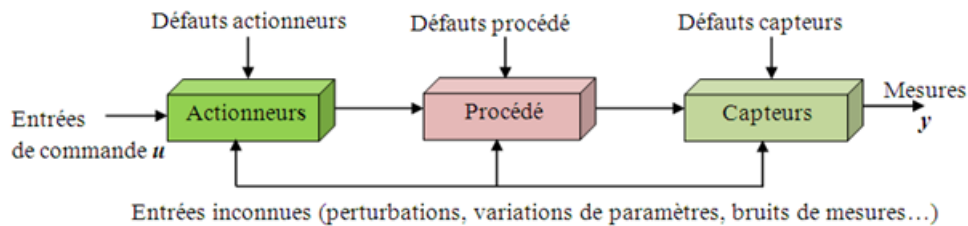


FIGURE 1.1 – Différents types affectant un système physique

1.5.2.1 Défauts actionneurs

Les défauts actionneurs agissent au niveau de la partie opérative et détériorent le signal d'entrée du système. ils représentent une perte totale (défaillance) ou partielle d'un actionneur agissant sur le système. un exemple de perte totale d'un actionneur est un système par le biais de cet actionneur. les défauts actionneurs partiels sont des défauts actionneurs réagissant de manière similaire au régime nominal mais en partie seulement, c'est-à-dire une certaine dégradation dans leur action sur le système.

1.5.2.2 Défauts capteurs

Ce type de défaut est la cause d'une mauvaise image de l'état physique du système. Un défaut capteur partiel produit un signal avec plus ou moins d'adéquation avec la valeur vraie de la variable à mesurer. Ceci peut se traduire par une réduction de la valeur affichée par rapport à la vraie valeur , ou la présence d'un biais ou de bruit accru empêchant une

bonne lecture. Un défaut capteur total produit une valeur qui n'est pas en rapport avec la grandeur à mesurer.

1.5.2.3 Défauts systèmes

Ce type de défaut provient du système lui-même en réduisant la capacité des composants du système et cela est due par des modifications de la structure ou des paramètres internes du modèle, bien souvent les défauts n'appartenant pas à un défaut à la catégorie des défauts capteur ou actionneur sont classés de manière arbitraire dans cette catégorie. Néanmoins, un défaut composant résulte de la classe ou de l'altération d'un composant du système réduisant les capacités de celui-ci à effectuer une tâche[19].

1.6 Principe et classification des méthodes de diagnostic de défauts

Dans le domaine des Sciences de l'ingénieur, le terme diagnostic désigne l'analyse des mauvais fonctionnements et des défaillances d'un système afin d'en déterminer la nature et la cause. En effet, le diagnostic est le processus d'évaluation d'un état de fonctionnement donné. Cet état est comparé avec un état de référence, si la différence est non nulle alors il s'agit d'évaluation de dérive de fonctionnement. Le diagnostic intègre différentes étapes dont la première consiste à détecter cet état de fonctionnement, une fois que l'anomalie est détectée, on passe à l'évaluation des causes de son occurrence qui consiste à identifier, analyser et localiser ces causes, et finalement la décision d'action pour le modifier[20].

Dans l'étude qui suit, il sera question de présenter le principe de diagnostic et les différentes méthodes de détection et d'isolation des défauts.

1.6.1 Principe du diagnostic des défauts

Le diagnostic est l'identification de la cause probable de la (ou des) défaillance (s) à l'aide d'un raisonnement logique fondé sur un ensemble d'information provenant d'une inspection, d'un contrôle ou d'un test.

Cette définition très courte et concise résume les deux tâches essentielles en diagnostic :

- Observer les symptômes de la défaillance.
- Identifier la cause de la défaillance à l'aide d'un raisonnement logique fondé sur des observations.

La procédure de diagnostic de défaillances et de dégradations susceptibles d'affecter les différentes entités d'un processus industriel s'articule autour des étapes suivantes :

- L'extraction des informations à partir de mesures appropriées ou d'observations.

- La mise en œuvre d'une méthode de diagnostic.
- La prise de décision[20].

Ces étapes de diagnostic sont résumées dans le schéma suivant :

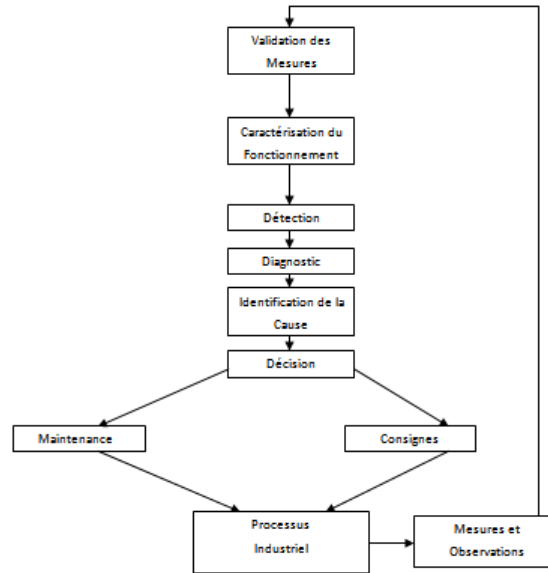


FIGURE 1.2 – Étapes de diagnostic.

Les méthodes du diagnostic sont des méthodes modernes, elles permettent d'éliminer la redondance matérielle pour le diagnostic des systèmes industrielles. Ces méthodes sont comme suit :

- Méthodes quantitative
- Méthode qualitative

1.6.1.1 Redondance Analytique

Par la condition d'observabilité, peut être exprimée en terme de capteurs, en plus de l'état des paramètres du modèle. Cette expression est connue comme étant la matrice dérivé d'observabilité dynamique, la multiplication de cette matrice avec une matrice nulle à gauche, appelé matrice de projection ou matrice de parité, elle va donner des relations de redondances analytique non linéaires. Ces relations nous permettent de générer des résidus non linéaire[21]. Ces résidus sont proches de zéros si le comportement du système est mobilisé correctement. Quand un défaut se produit, le résidu dévié considérablement de zéro (l'écart est beaucoup plus grande que l'écart causé par les incertitudes du modèle et les bruits de mesures), dans ce cas, on dit qu'il y a détection. Les seuils doivent être choisi afin de réduire des problèmes des fausses alarmes[22].

A noté que cette méthode de redondance analytique ne peut être mise en œuvre si

l'on dispose d'un modèle statique ou dynamique, linéaire ou non linéaire, déterministe ou stochastique, du systèmes ou de sous système, reliant les entrées et les sorties mesurées.

1.6.1.2 Redondance matérielle

Le but principale pour valider une mesure consiste à doubler, tripler, . . . , multiplier les chaines de mesures. Cette redondance permet la localisation et la détection des capteurs défaillants, malgré la simplicité de la méthode et sa fiabilité. Néanmoins, elle présente un inconvénient qui réside sur le surcout de l'installation et une diminution du temps moyenne de bon fonctionnement[23].

Le choix du nombre de capteur à placer offre différentes solutions sur la surveillance du processus. Le choix de placer deux capteurs (redondance matérielle doubler) ne permet que la détection d'une panne simple mais elle ne permet pas de localiser le capteur qui est en panne. Or que l'utilisation de trois capteurs ou plus (système triplex), permet la détection et la localisation de la panne, en cascade, un détecteur et un voteur, sachant que le role de voteur est de déterminer le capteur qui est en panne, cela en analysant la dispersion des trois mesures ou plus et la prise en compte des caractéristiques statiques précédentes du système[24].

1.6.2 Présentation des méthodes de diagnostic

La diversité des approches qui ont été développées pour le diagnostic des systèmes dynamiques semble être le résultat de contextes différents. Ces contextes sont associés à la nature des applications visées et aux caractéristiques propres du cahier des charges qui en résultent. ainsi, la nature des informations disponibles sur le système ou le type de défaut à détecter conduisent à la mise en œuvre de stratégies spécifiques. Dans ce cadre, les travaux sur la détection et l'isolation de défauts et ceux dédiés au diagnostic montrent une grande variété à la fois dans les points de vues et dans les méthodes utilisées[25, 26].

Ainsi, elles sont principalement basées sur deux approches que nous pouvons classer en méthodes avec ou sans modèles (1.3).

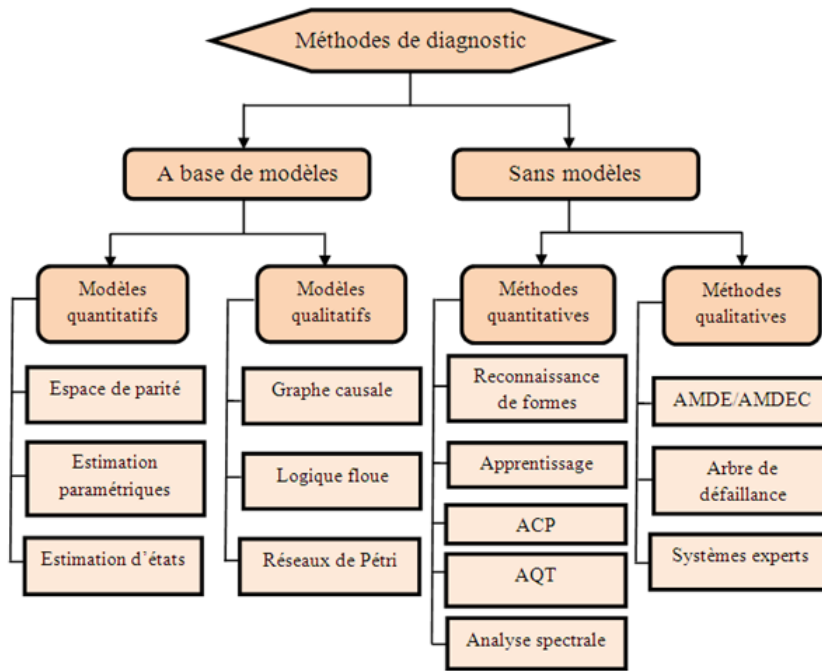


FIGURE 1.3 – Classification des méthodes de diagnostic.

1.6.2.1 Méthodes sans modèles

Dans certains cas, il s'avère difficile ou parfois impossible de trouver le bon modèle mathématique pour un système réel, à cause des nombreuses configurations intervenant dans le processus de production ou de la complexité des phénomènes mise en jeu. La solution dans ce cas est le recours à des méthodes qui ne nécessitent aucune connaissance approfondie du système.

En effet, on trouve :

- Les approches qualitatives qui sont basées sur le traitement de données.
- Les approches qualitatives à base de connaissance.

1.6.2.1.1 Approches quantitatives

Les méthodes de traitement de données consistent en l'exploitation d'une base de connaissance symbolique. La seule information disponible, dans ce cas, est sous la forme de données historiques qui correspondent aux différents modes de fonctionnement du système ou bien à l'aide d'un traitement en ligne des signaux issus de capteurs.

Reconnaissance de formes :

L'objectif est la reconnaissance d'une forme parmi différentes possibilités à partir d'observation bruitées. Il existe plusieurs approches : la reconnaissance de la forme et la reconnaissance de formes de type numérique. Cette forme est représentée par un ensemble de paramètres ou attributs caractéristiques qui peuvent être numériques ou symboliques. Un

prototype est défini par des valeurs précises de l'ensemble des paramètres caractéristiques de la défaillance. Une classe est définie par un ensemble possible de valeurs des attributs et elle est représentée par son prototype. La reconnaissance d'une forme parmi M classes est la décision d'affecter l'objet à une classe particulière. Le diagnostic consiste à décider de quel prototype cette forme est la plus proche en fonction d'un critère d'évaluation comme les distances euclidiennes (classes circulaires) ou les distances Mahalanobis (classes elliptique) par exemple[27].

Apprentissage :

Les méthodes basées sur la démarche d'apprentissage comme les réseaux de neurones RN et les méthodes machine learning sont utilisées pour le diagnostic. Les réseaux de neurones sont souvent employés pour le diagnostic des systèmes présentant des dynamiques complexes, pour lesquels il est difficile de connaître les situations de défauts. Ils sont utilisés à différentes étapes : tout d'abord lors de la phase d'apprentissage, à partir des données issues du procédé en comportement normal et en opérations sous défauts, une classification des états qualitatifs ainsi que l'identification de symptômes sont effectuées.

Cet apprentissage permet alors, sur la base de l'optimisation d'un critère, de reproduire le comportement du système à modéliser. Il peut être effectué de deux manières :

- Apprentissage supervisé : le réseau de neurones utilise les données d'entrée et de sortie du système à modéliser.
- Apprentissage non supervisé : le réseau de neurones n'utilise que les données d'entrée du système et l'apprentissage s'effectue par comparaison entre plusieurs cas d'exemples.

Une fois que l'apprentissage est achevé, il peut être, ensuite, utilisé pour la généralisation. Il s'agit de la phase de test où de nouveaux exemples sont présentés au réseau de neurones pour juger de sa capacité à prédire les comportements du système modélisé. lors d'un problème de diagnostic, leur emploi peut alors se faire à trois niveaux :

- comme modèle du système à surveiller en état normal et générer un résidu d'erreur entre les observations et les prédictions.
- comme système d'évaluation de résidus pour le diagnostic.
- comme système de détection en une seule étape[28].

Analyse en Composantes Principales (ACP) :

C'est un outil statistique largement utilisé pour l'analyse de données collectées sur des systèmes en cours de fonctionnement afin de surveiller leur comportement. L'analyse en composantes principales est une technique numérique bien connue dans le domaine du traitement de données pour réduire la dimension de l'espace de représentation d'un système. L'ACP est essentiellement basée sur la mise en évidence de relations linéaires entre les

variables et elle présente un caractère d'optimalité au sens d'un critère portant sur l'erreur quadratique d'estimation en valeur moyenne[29]. Rappelons que l'approche classique de l'ACP utilise un calcul préliminaire de la moyenne des données et de leur matrice de covariance, la moyenne et la variance sont sensibles à la présence de valeur aberrantes, et les résultats obtenus s'avèrent souvent inexploitable car trop biaisés par l'influence de ces valeurs aberrantes. Par la suite, l'un des inconvénients majeurs de l'approche ACP résulte de l'utilisation de technique d'estimation par moindres carrées, techniques qui échouent souvent à s'affranchir des biais de mesure accidentels, ce qui est malheureusement assez fréquent sur le plan pratique.

Analyse Qualitative de Tendance (AQT) :

La représentation et l'analyse qualitative de signaux à travers l'information des dérivées de premier et du deuxième ordre représentent l'une des principales technique pour le diagnostic et la supervision basées sur des données et sur l'abstraction de la dynamique continue. L'analyse qualitative de tendance des signaux est basée, en général, sur la segmentation temporelle des signaux en unités, appelées épisode, en fonction de leur pente et de leur convexité et d'une classification de ces épisodes. D'après Colomer et al. les besoins des systèmes de supervision qui motivent l'utilisation des méthodes d'analyse qualitative de tendances sont l'incertitude, l'incomplétude et l'hétérogénéité des données. Ces besoins demandent un raisonnement qualitatif et un raisonnement temporel qui est nécessaire non seulement sur des informations instantanées mais aussi sur le comportement historique[30].

Analyse spectrale :

C'est la première méthode à investiguer pour détecter des harmoniques dans un signal. Elle s'avère intéressante pour l'analyse des signaux qui montrent des oscillations avec des périodes longues comme les débits ou bien les pressions. La plus part des mesures possèdent un spectre typique de fréquence sous certaines conditions normales, tout déviation des caractéristiques fréquentielles d'un signal est liée à une anomalie.

L'utilisation d'une procédure de décision permet de détecter et de localiser l'élément défaillant du système. Parmi les outils de décision appliqués sur un échantillon de mesures, on peut citer : le test empirique de franchissement de seuil, le test de variance ou le test de moyenne[31].

Ces méthodes de traitement de données sont des approches envisageables lorsqu'un raisonnement sur les comportements dynamique du procédé n'est pas utile. Les approches de classification de données qui utilisent la reconnaissance de formes sont basées sur l'analyse des données issues des signaux reliés entre eux pour la distinction des différents modes de fonctionnement et certains modes défaillants. Tout ces méthode sont profondément dépendantes d'un grand volume des données, ce qui les limitent en général à la détection.

Le diagnostic dépend donc de la représentation et de la discrimination de tous les modes de défaut. Dans le cadre du suivi de régions transitoires, les méthodes d'AQT sont les plus utilisées.

1.6.2.1.2 Approches qualitatives

Ces méthodes sont fondées sur la connaissance, elle sont envisageables quand l'obtention d'un modèle analytique du procédé s'avère difficile, et lorsque la plus part des mesures sont indisponible. Elles sont basées sur une connaissance associative dépendante du système et sur une connaissance a priori des défauts et de leur effets.

Analyse des Modes de Défaillances et de leurs Effets et leur Criticité (AMDEC) :

C'est une technique spécifique de sureté de fonctionnement, l'AMDEC est avant tout une méthode d'analyse des systèmes s'appuyant sur un raisonnement inductif causes conséquences, pour l'étude des causes, des effets des défauts puis de leur criticité. Historiquement, la méthode initiale est appelée Analyse des Modes de Défaillances et de leurs Effets (AMDE) qui constitue une analyse préventive de la sureté de fonctionnement au sens de la fiabilité, la disponibilité, la maintenabilité et la sécurité. Une fois complétée par une analyse de criticité, qui permet d'extraire les modes de défaillance les plus critique, l'AMDE devient l'AMDEC.

Cette dernière consiste à estimer la criticité des modes de défaillances à partir de trois critères : gravité, fréquence d'apparition et probabilité de non détection. Le principe de la prévention repose sur le recensement systématique et l'évaluation des risques potentiels d'erreurs susceptibles de se produire à toutes les phases de réalisation d'un système.

Arbre de causes à effets ou arbre de défaillances :

contrairement à l'analyse des modes de défaillances, l'arbre de défaillances est une méthode déductive qui procède un raisonnement logique rigoureux[1]. Elle permet de savoir comment un système peut être indisponible. Il s'agit de représenter les différents évènements et leurs liaisons par l'intermédiaire d'opérateurs logiques ET ou OU selon que le défaut se produit lorsque les évènement se réalisent ensemble ou séparément. L'arbre de défaillances est une représentation graphique qui représente une démarche d'analyse quantitative et qualitative d'évènements. Elle est construite en recherchant l'ensemble des évènements élémentaires, ou les combinaisons d'évènements, qui conduisent à un évènement non désiré. L'objectif est de suivre une logique déductive en partant de cet évènement indésirable pour déterminer de manière exhaustive l'ensemble de ses causes jusqu'aux plus élémentaires. Ce type d'analyse permet, dans le domaine de la maintenance, d'améliorer la conception, de faire un diagnostic rapide et de prévoir une meilleure

logistique. Pour établir cet arbre, il est souhaitable de s'aider de l'analyse des modes de pannes et de défaillances décrits précédemment en AMDEC.

Systemes experts

Un système expert est une application informatique capable de prendre des décisions ou de résoudre des problèmes dans un domaine précis. C'est un programme faisant appel à l'intelligence artificielle. Une base de données est conçue à partir du savoir d'un expert avant d'être intégrée à un programme. Le diagnostic par système expert se base sur l'expérience disponible sur le système pour construire une table de correspondance permettant d'associer efficacement les observations aux diagnostics correspondants. L'expérience peut être fournie par un opérateur humain, dans ce cas, la connaissance humaine doit être traduite en langage informatique ou par un enregistrement éventuellement annoté des précédentes exécutions du système, ou un algorithme d'apprentissage automatique soit être utilisé.

Les principaux inconvénients de cette méthode sont :

- l'acquisition de l'expertise.
- la taille du système expert.
- la non robustesse.

1.6.3 Méthodes avec modèles

Les méthodes à base de modèles se reposent sur les modèles comportementaux explicites du système soumis au diagnostic.

On trouve deux approches, qualitatives et quantitatives.

1.6.3.1 Approches qualitatives

La communauté d'Intelligence Artificielle (IA) a proposée des approches qui utilisent une connaissance profonde des composants du système basée sur une théorie logique du raisonnement.

Les méthodes les plus utilisées par cette communauté sont :

Graphes causaux :

Ils déterminent les composants défectueux qui peuvent expliquer le fonctionnement anormal observé[32].

Logique floue :

La logique floue est une théorie mathématique, qui tient compte des incertitudes et permet une fusion des informations[33].

Réseaux de Pétri :

Le diagnostic des Systèmes à Evénements Discrets SED est basé sur l'utilisation de modèles discrets tels que les réseaux de Pétri et les automates d'états finis, ...etc. Ces approches sont fondées sur une représentation purement discrète du procédé dues à la nature discrète des variables de ses composants[34].

1.6.3.2 Approches quantitatives

Ces approches sont basées sur la modélisation et le contrôle des systèmes industriels avec un modèle dynamique quantitatif. Ces approches appartiennent à la communauté Fault Détection and Isolation (FDI).

Les méthodes principales les plus utilisées sont :

Estimation paramétrique :

Cette approche permet d'analyser l'influence des défauts sur les paramètres structuraux du modèle du système[36].

Estimation d'états :

Ces méthodes sont basées sur l'utilisation des observateurs d'état. Elles s'appuient sur une bonne connaissance du modèle et de ses paramètres, elle consistent à estimer les variables d'état par un observateur pour reconstruire l'information[37].

Espace de parité :

les méthodes basées sur les relations de parité, s'appuient sur l'élaboration de signaux permettant de tester la cohérence des mesures par rapport à leurs valeurs calculées à l'aide d'un modèle c'est à dire vérifier la parité du modèle du procédé avec les sorties mesurées et des entrées connues du modèle. La conception de l'espace de parité est basée sur le développement des expression analytiques[35].

1.7 Conclusion

Dans ce chapitre, nous avons présenté d'une manière générale un résumé de l'état de l'art sur les principes de base du diagnostic de défauts. En effet nous avons rappelé quelques définitions et nous avons exposé un panorama assez complet des principales méthodes de diagnostic sans modèles. Les techniques décrites ici sont les plus connues et les plus couramment utilisées.

Deux catégories de méthodes existent : les méthodes à base de modèles et celles sans modèles. Le choix d'une de ces méthodes dépend de la connaissance du système, de la présence de capteurs ou de modèles qui permettent de suivre l'état réel du système. Pour améliorer la procédure de diagnostic et réduire les coûts liés aux défaillances, il est nécessaire d'étudier et d'analyser le plus précisément possible le comportement réel du système. Nous nous intéresserons dans notre travail aux approches de diagnostic sans modèle qui se basent sur l'Analyse en composantes principales (ACP) pour étudier des processus auquel le raisonnement sur les comportements dynamiques du procédé n'est pas utile. Cette méthode est essentiellement basée sur la mise en évidence de relations linéaires entre les variables et elle présente un caractère d'optimalité au sens d'un critère portant sur l'erreur quadratique d'estimation de valeur moyenne. Les approches quantitatives qu'on a citées sont profondément dépendantes d'un grand volume des données, ce qui les limitent en générales à la détection, et les approches qualitative sont fondées la connaissance.

Chapitre 2

Analyse en Composantes Principales

2.1 Introduction

Grâce aux progrès de l'informatique, en particulier en matière de puissance de calcul et de capacité de stockage des données, les moyennes mises en œuvre pour la caractérisation des données donnent naissance à des bases de grandes dimensions, de ce fait, le besoin de réduire la dimension est née, plusieurs méthodes ont vues le jour basées sur la statistique classique, qui s'intéresse aux valeurs prises par quelques variables pour un nombre d'individus réduit, parmi lesquels on cite l'analyse en composantes principales qui a fait l'objet de beaucoup d'investigation, on peut citer pour cela les travaux notamment de Dauxois et Pousse(1976), Dauxois et al (1982), Besse et Ramsey (1986), Bouhadou et al (1987), Nagai et Taniguchi (1987). Cette méthode fait partie du groupe de méthodes descriptives multidimensionnelles, appelées méthodes factorielles.

L'ACP peut se présente comme une technique de projection la plus répondue dans la littérature, qui permet de fournir un outil de visualisation des données, et aussi explorer les liaisons entre variables et ressemblance entre individus.

L'idée à la base de l'analyse en composantes principales est de présenter dans un espace de dimension réduit un grand jeu de données décrits par un nombre important de variables quantitative, dont l'analyse par les techniques statistiques classiques était initialement irréalisable et la visualisation est impraticable. Cette transformation des variables est basée sur la recherche des directions de plus grandes variances dans les données, et projeté les données sur ces axes afin de rendre compte de la variance observée sur les données initiale, tout en se limitant à un nombre réduit de composantes. Cette réduction n'est possible que si les variables initiales ne sont pas indépendantes, et ont des coefficients de corrélation non nuls. Les composantes ainsi obtenue sont des combinaisons linéaire des variables initiales non corrélé les unes au autres.

La représentation approchée du nuage des individus dans le nouveau espace après

l'enlèvement de toute information inconséquente et redondante, peut être accompagné d'une perte d'information engendrée par la réduction de la dimension, l'appréciation de cette perte est l'un des points les plus délicats de l'ACP, la qualité de représentation d'une variable sur un axe factoriel est exprimée par le coefficient de corrélation linéaire entre la variable initial et la nouvel variable.

Dans ce chapitre nous donnerons quelques notions de base concernant la technique d'analyse des données(ACP) afin de résumer les données, qui seront utilisées par la suite dans le chapitre suivant tout en laissant le soin au lecteur désireux d'en connaître d'avantage de consulter les ouvrages dédié dont nous donnerons les références [3, 6, 38, 39, 40].

2.2 Avantages et inconvénients de l'ACP

2.2.1 Avantages

- Simplicité mathématique : L'ACP est une méthode factorielle car la réduction du nombre des caractères ne se fait pas par une simple sélection de certains d'entre eux, mais par la construction de nouveaux caractères synthétiques obtenus en combinant les caractères initiaux au moyen des "facteurs". Cependant, il s'agit seulement de combinaisons linéaires. Les seuls véritables outils mathématiques utilisés dans l'ACP sont le calcul des valeurs/vecteurs propres d'une matrice, et les changements de base. Sur le plan mathématique, l'ACP est donc une méthode simple à mettre en oeuvre.
- Simplicité des résultats : Grâce aux graphiques qu'elle fournit, l'Analyse en Composantes Principales permet d'appréhender une grande partie de ses résultats d'un simple coup d'oeil.
- Puissance : L'ACP a beau être simple, elle n'en est pas moins puissante. Elle offre, en quelques opérations seulement, un résumé et une vue complète des relations existant entre les variables quantitatives d'une population d'étude, résultats qui n'auraient pas pu être obtenus autrement, ou bien uniquement au prix de manipulations fastidieuses.
- Flexibilité : L'ACP est une méthode très souple, puisqu'elle s'applique sur un ensemble de données de contenu et de taille quelconques, pour peu qu'il s'agisse de données quantitatives organisées sous forme individus/variables. Cette souplesse d'utilisation se traduit surtout par la diversité des applications de l'ACP, qui touche tous les domaines, comme exposé dans la partie précédente.

2.2.2 Inconvénients

- En tant que méthode d'analyse de données, l'ACP n'a pas réellement d'inconvénients en soi. Elle s'applique simplement sur des cas précis et pour générer un type de résultat particulier. Ça n'aurait donc aucun sens de dire que c'est un inconvénient de l'ACP qu'elle ne s'applique pas en dehors de ce contexte. De même, étant donné qu'il s'agit avant tout d'une technique de résumé de données, la perte d'information forcément engendrée n'est pas un inconvénient, mais plutôt une condition d'obtention du résultat, même si elle occulte parfois des caractéristiques pourtant représentatives dans certains cas particuliers.

2.3 Principe de l'analyse en composantes principales

L'ACP cherche une représentation des n individus dans un sous-espace \mathfrak{R}^ℓ de \mathfrak{R}^m de dimension (ℓ), tel que le sous espace \mathfrak{R}^ℓ doit respecter deux conditions, la première, doit être ajuster le mieux possible au nuage des individus (la somme des carrées des distances des individus à \mathfrak{R}^ℓ doit être minimal), et la deuxième il doit avoir une inertie (dispersion du nuage projeté) maximale.

Alors l'ACP consiste à remplacer une famille de variables par de nouvelles variables (indices synthétiques) de variance maximale qui résument au mieux l'information latente dans l'ensemble des données. Autrement dit, l'ACP cherche à définir ℓ nouvelles variables non corrélés deux à deux qui sont en combinaison linéaire des p variables initiales, qui feront perdre le moins d'information possible.

Les nouvelles variables appelées composantes principales, définissent des plans factoriels qui servent de base à une représentation graphique plane des variables originales. L'interprétation des résultats se restreint généralement aux deux premières plans factoriels qui, comme nous le verrons dans la suite contiennent l'essentiel de l'information, sous réserve que ceux-ci, expliquent la majeure partie de la variance du nuage des variables initiales.

En pratique, pour un tableau de données comportant m colonnes, on pourra définir m indices synthétiques appelés scores ou valeurs des composantes principales d'importance décroissante et non corrélés.

2.4 Algorithme général

- Acquisition des données
- Normalisation : centrage et réduction
- Calcul de la matrice de covariance ou corrélation
- Décomposition

- Sélection du nombre de composantes principales
- Reconstruction des données à partir des composantes principales

2.5 Préparation des données

2.5.1 Tableau des données

L'information permettant de traduire le comportement d'un système tenu lors du fonctionnement normal est donnée par les mesures des variables à surveiller de celui-ci.

Considérons n observations repérées par un indice ($i=1 \dots\dots n$) sur lesquelles on mesure p variables repérées par un indice ($j=1 \dots\dots m$), ces observations seront disposées dans un tableau de données X à n lignes et m colonnes.

$$X = \begin{bmatrix} x_1^1 & \cdots & x_1^j & \cdots & x_1^m \\ \vdots & & \vdots & & \vdots \\ x_i^1 & \cdots & x_i^j & \cdots & x_i^m \\ \vdots & & \vdots & & \vdots \\ x_n^1 & \cdots & x_n^j & \cdots & x_n^m \end{bmatrix} \quad X \in \mathfrak{R}^{n \times m} \quad (2.1)$$

où x_i^j mesure la valeur de la variable j pour l'individu i .

Chaque variable x^j est représentée par un vecteur colonne de X correspondant, qui rassemble les n valeurs qu'elle prend sur les n individus.

$$x^j = \begin{bmatrix} x_1^j \\ \vdots \\ x_i^j \\ \vdots \\ x_n^j \end{bmatrix} \quad (2.2)$$

De même, chaque individu i étant décrit dans le tableau par une ligne de p variables.

$$x_i^T = \left[x_i^1 \quad \cdots \quad x_i^j \quad \cdots \quad x_i^m \right] \quad (2.3)$$

2.5.2 Centrage et pondération des données

L'origine du repère de représentation des individus R^m n'est pas satisfaisant lorsque les coordonnées des points du nuage sont grandes le nuage est éloigné de cette origine, de

ce fait il est nécessaire d'appliqué une translation de l'origine de tel sort de confondre l'origine avec le centre de gravité du nuage G , donc de soustraire à chaque variable sa moyenne.

Soit, en appelant p_i le poids affecté à l'individu i . Pour l'ACP on choisit de donner la même importance $\left(\frac{1}{n}\right)$ à tout les individus $\left(\sum_i p_i = 1\right)$, ces poids sont regroupés dans une matrice diagonale D de taille n .

$$D = \begin{bmatrix} p_1 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & p_i & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & p_n \end{bmatrix} \quad (2.4)$$

Le centre de gravité G du nuage des individus est le point dont les cordonnés sont les valeurs moyennes des variables.

On a, $G = X^T D 1_n$, où 1_n est un vecteur de \mathfrak{R}^n , dont tout les composantes sont égale à 1 :

$$G = \begin{bmatrix} x_1^1 & \cdots & x_i^1 & \cdots & x_n^1 \\ \vdots & & \vdots & & \vdots \\ x_1^j & \cdots & x_i^j & \cdots & x_n^j \\ \vdots & & \vdots & & \vdots \\ x_1^m & \cdots & x_i^m & \cdots & x_n^m \end{bmatrix} \begin{bmatrix} p_1 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & p_i & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & p_n \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$G = \begin{bmatrix} x_1^1 p_1 & \cdots & x_i^1 p_i & \cdots & x_n^1 p_n \\ \vdots & & \vdots & & \vdots \\ x_1^j p_1 & \cdots & x_i^j p_i & \cdots & x_n^j p_n \\ \vdots & & \vdots & & \vdots \\ x_1^m p_1 & \cdots & x_i^m p_i & \cdots & x_n^m p_n \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$G = \begin{bmatrix} x_1^1 p_{1+} & \cdots & +x_i^1 p_{i+} & \cdots & +x_n^1 p_n \\ \vdots & & \vdots & & \vdots \\ x_1^j p_{1+} & \cdots & +x_i^j p_{i+} & \cdots & +x_n^j p_n \\ \vdots & & \vdots & & \vdots \\ x_1^m p_{1+} & \cdots & +x_i^m p_{i+} & \cdots & +x_n^m p_n \end{bmatrix}$$

Avec

$$\bar{x}^j = \sum_i p_i x_i^j \quad (2.5)$$

Donc :

$$G = \begin{bmatrix} \bar{x}^1 \\ \vdots \\ \bar{x}^j \\ \vdots \\ \bar{x}^m \end{bmatrix} \quad (2.6)$$

Prendre G comme origine revient alors à travailler sur le tableau des données centrées Y tel que :

$$y_i^j = x_i^j - \bar{x}^j \quad (2.7)$$

On a :

$$Y = X - 1_m G^T \quad (2.8)$$

$$Y = \begin{bmatrix} x_1^1 - \bar{x}^1 & \dots & x_1^j - \bar{x}^j & \dots & x_1^m - \bar{x}^m \\ \vdots & & \vdots & & \vdots \\ x_i^1 - \bar{x}^1 & \dots & x_i^j - \bar{x}^j & \dots & x_i^m - \bar{x}^m \\ \vdots & & \vdots & & \vdots \\ x_n^1 - \bar{x}^1 & \dots & x_n^j - \bar{x}^j & \dots & x_n^m - \bar{x}^m \end{bmatrix}$$

Le vecteur des coordonnées centrées de l'individu x_i est :

$$x_{ci} = \begin{bmatrix} y_i^1 \\ \vdots \\ y_i^j \\ \vdots \\ y_i^m \end{bmatrix} \quad (2.9)$$

Afin de rendre les résultats indépendant des unités lorsque les mesures ne sont pas homogènes, il est indispensable de les standardiser pour leur donner la même importance.

La réduction consiste généralement à mettre toutes les variables sur un pied d'égalité (la même variabilité) en les divisant par l'écart type S des variables, de ce fait on aura la même influence des variables dans les calculs des distances entre individus. Donc les variables initiales qui sont très dispersées elle vont pas prendre le pas sur les autres dans le choix des axes qui présentent le maximum de variance.

Ainsi lorsque les variables sont centrées réduites, la variance de chaque variable vaut 1, le nuage alors située sur une sphère de rayon 1, dans ce cas on parle alors d'ACP normée.

Le tableau des données normalisées à pour terme général :

$$R = \begin{bmatrix} \frac{x_1^1 - \bar{x}^1}{S_1} & \dots & \frac{x_1^j - \bar{x}^j}{S_j} & \dots & \frac{x_1^m - \bar{x}^m}{S_m} \\ \vdots & & \vdots & & \vdots \\ \frac{x_i^1 - \bar{x}^1}{S_1} & \dots & \frac{x_i^j - \bar{x}^j}{S_j} & \dots & \frac{x_i^m - \bar{x}^m}{S_m} \\ \vdots & & \vdots & & \vdots \\ \frac{x_n^1 - \bar{x}^1}{S_1} & \dots & \frac{x_n^j - \bar{x}^j}{S_j} & \dots & \frac{x_n^m - \bar{x}^m}{S_m} \end{bmatrix} \quad (2.10)$$

2.6 Représentation du nuage de points

Les données que comporte le tableau génèrent un nuage de points qui est une représentation graphique dans un repère qu'on peut exploiter selon notre besoin d'étude.

Pour cela deux ongles d'études se présentent, la première est pour l'étude des variables, cette étude s'intéresse au liaison entre variables et cela par la voie du calcul du coefficient de corrélation, sachant qu'une variable est une colonne du tableau qui contient un ensemble de mesures (n composantes), alors le nuage des variables est représenté dans un repère à n dimensions noté (R^n), et chaque dimension correspond à un individu. La deuxième étude s'intéresse à l'étude des individus afin d'étudier la ressemblance entre individus par voie de calcul de distance, à savoir que l'individu représente une ligne du tableau, et cette ligne contient un ensemble de variables, alors un individus c'est un paquet de m valeurs numériques qui se représentent dans un espace à m dimensions, qu'on note (R^m), et dans cette espace chaque dimension correspond à une variable.

Dans notre étude on va s'intéresser :

- Au nuage centré des n points individus
- Au nuage centré-réduit des n points individus

2.7 Analyse du nuage des individus

2.7.1 Choix de la distance

Pour évaluer la ressemblance entre individus on mesure la différence de leurs positions dans l'espace. Deux individus se rassemblent d'autant leur position est proche, la distance utilisée par l'ACP dans l'espace des individus est la distance euclidienne classique.

On suppose que les axes définis par les variables homogène constituent une base orthogonale, la distance entre deux points ce calcule par la formule de Pythagore, le carré

de la distance est la somme des carrés des différences des coordonnées :

$$d^2(x_i, x_{i'}) = \sum_{j=1}^m (x_i^j - x_{i'}^j)^2 \quad (2.11)$$

à cette distance on associe un produit scalaire entre deux vecteurs :

$$\langle \vec{ox}_i, \vec{ox}_{i'} \rangle = \sum_{j=1}^m x_i^j x_{i'}^j = x_i^T x_{i'} \quad (2.12)$$

La norme d'un vecteur :

$$\|\vec{ox}_i\|^2 = \sum_{j=1}^m x_i^{j2} = x_i^T x_i \quad (2.13)$$

Lorsque on veut privilégier une variable par rapport à l'autre on accorde un poids $m_j > 0$ à chaque variable j .

Définition : soit M une matrice symétrique de taille p définie positive qui renferme les différents poids m_j

La fonction suivante $d_M : \mathfrak{R}^m \times \mathfrak{R}^m \rightarrow \mathfrak{R}^+$ définit une métrique

$$M = \text{diag}(m_j)$$

La distance entre individus x_i et $x_{i'}$ est définie par la forme quadratique :

$$d_M^2(x_i, x_{i'}) = (x_i - x_{i'})^T M (x_i - x_{i'}) = \|x_i - x_{i'}\|_M^2 \quad (2.14)$$

Cette distance est appelée distance de Mahalanobis lorsque $M = V^{-1}$, où V est la matrice de variance-covariance des données.

La métrique définie ci-dessus dérive du produit scalaire :

$$\langle x_i, x_{i'} \rangle_M = x_i^T M x_{i'} \quad (2.15)$$

Les deux individus sont orthogonaux si $\langle x_i, x_{i'} \rangle_M = 0$.

Le choix de la métrique M qui dépend de l'utilisateur à une influence fondamentale sur le résultat de l'analyse, lorsque les mesures ne sont pas homogènes la métrique utilisée est suivant la métrique réduite qui consiste à prendre M diagonale des inverses des variances des variables, cette métrique permet de s'affranchir de l'unité de mesure des variables et donner à chaque variable le même importance dans le calcul de la distance.

$$M = D_{\frac{1}{S^2}} = \begin{bmatrix} \frac{1}{S_1^2} & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & \frac{1}{S_j^2} & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & \frac{1}{S_m^2} \end{bmatrix} \quad (2.16)$$

L'utilisation de cette métrique sur le tableau de données X revient à travailler avec la métrique euclidienne sur le tableau transformé $X D_{\frac{1}{S}}$, en effet :

$$d_M^2(x_i, x_{i'}) = (x_i - x_{i'})^T D_{\frac{1}{S}}^T D_{\frac{1}{S}} (x_i - x_{i'})$$

$$d_M^2(x_i, x_{i'}) = \left(D_{\frac{1}{S}} x_i - D_{\frac{1}{S}} x_{i'} \right)^T \left(D_{\frac{1}{S}} x_i - D_{\frac{1}{S}} x_{i'} \right) \quad (2.17)$$

Travailler avec la métrique $D_{\frac{1}{S^2}}$ sur le tableau Y est équivalent à travailler avec la métrique euclidienne I sur le tableau centré réduit R.

Le tableau de données centré réduit R ce calcul matricielle-ment par :

$$R = Y D_{\frac{1}{S}} \quad (2.18)$$

Nous considérons dans la suite que les variables son centrées et possède le même poids ($m_j = 1$).

2.7.2 Méthode de projection

L'existence des phénomènes complexes que se soit physique ou naturel offrent des difficultés majeures pour la visualisation et l'interprétation d'un ensemble d'observation qui décrivent le phénomène étudié, et cela à cause du nombre élevé des données que produisent ces phénomènes complexes. Pour cela les méthodes d'analyse des données offrent des solutions pour pouvoir comprendre et bien interprété la masse de données qu'offre ces phénomènes, et cela en fosant appelle à la méthode de projection puisque cette dernière offre une aide à l'humain pour la compréhension de ses nombreuses données vu que l'objectif de la projection est de compresser les données afin d'extraire un résumé d'information que contiennent ces données. Plusieurs techniques existent, les plus connues sont l'analyse en composantes principales (ACP), l'analyse canonique (AC), l'analyse factorielle (AF), la régression sur composantes principales (RCP), et la projection dans les structures latentes

(cacher) (PSL), également connue sous la dénomination des moindres carrés partielle.

2.7.2.1 L'utilité de la projection

Le risque d'avoir des informations manquantes ou erronées est la conséquence d'une mauvaise réduction de données en supprimant certaines variables importantes. Alors la bonne visualisation des données tout en conservant l'important d'informations afin d'avoir une très bonne interprétation est l'objectif de la réduction pour une analyse statistique multivariées.

Les méthodes utilisées pour réduire la dimension focalisent leurs attentions sur les variables latentes. Ces variables cachées sont des variables virtuelles construites afin de comprendre un intérêt caractéristique qui ne peut être mesuré, tout en sachant que ces variables ne sont pas observables et elles ont un petit impacte sur les variables mesurées ou originales, dans la majorité du temps elles sont définies comme une combinaison linéaire des variables initiales.

L'analyse en composantes principales est une solution pour une réduction tout en gardant les informations pertinentes du processus à étudier, elle est adaptée à des données corrélées qui sont projetées sur des sous espaces de dimension réduite afin de garder l'essentiel d'informations qui décrivent le processus. Donc l'ACP c'est une méthode qui réduit la dimension de l'espace des variables du processus pour trouver des vecteurs latents qui décrivent des directions principales non corrélés.

2.7.3 Projections orthogonales

L'image du nuage de l'espace original dans le sous espace de dimension réduite avec une perte minimum d'informations est obtenue par la projection orthogonale, en effet la distance entre un vecteur et leur projections est minimale lorsque cette projection est orthogonal.

Soit z un vecteur et S un sous espace, le vecteur qui minimise la distance de z aux vecteurs de S est le projeté orthogonal de z sur S .

$$\|z - \hat{z}\| = \underset{v \in S}{\text{Min}} \|z - v\| \quad (2.19)$$

\hat{z} et v sont respectivement le projeté et le vecteur de S .

2.7.4 Matrice de variance-covariance

Pour déterminer le sous espace, il est important de voir les liens entre les variables, pour le faire nous devons déterminer une matrice qui traduit la variabilité du nuage.

La matrice de variance-covariance (ou simplement matrice de covariance) est une matrice carré qui caractérise les interactions (linéaire) entre m variables aléatoires x^1, \dots ,

x^m . La covariance mesure la liaison linéaire qui peut exister entre un couple de variables statistiques ou un couples de variables aléatoires quantitatives.

Après le centrage des données, soit $\bar{x}^j = 0 \forall j \in \{1, \dots, m\}$. L'importance de cette hypothèse est de simplifie les calculs.

Ayant n individus décrit par leur m caractères peuvent être mis sous forme d'une matrice défini par (n) lignes et (m) colonnes : la ligne décrit un individu et chaque colonne correspond à une variable. Rappelant que cette matrice est noté par X .

On définit la matrice de variance V comme suit :

$$V = \frac{1}{n} Y^T Y \tag{2.20}$$

on peut écrire V sous forme suivante :

$$V = X^T D X - G G^T = Y^T D Y$$

En effet, de l'équation (2.8) on a :

$$V = Y^T D Y = (X - 1_n G^T)^T D (X - 1_n G^T)$$

$$V = (X^T - G 1_n^T) D (X - 1_n G^T)$$

$$V = X^T D X - X^T D 1_n G^T - G 1_n^T D X + G 1_n^T D 1_n G^T$$

$$V = X^T D X - G G^T - G G^T + G G^T$$

$$V = X^T D X - G G^T \tag{2.21}$$

Avec

$$X^T D X = \sum_{i=1}^n p_i x_i x_i^T \tag{2.22}$$

$$V = \begin{bmatrix} var(x^1) & \cdots & cov(x^1, x^j) & \cdots & cov(x^1, x^m) \\ \vdots & \ddots & \vdots & & \vdots \\ cov(x^j, x^1) & \cdots & var(x^j) & \cdots & cov(x^j, x^m) \\ \vdots & & \vdots & \ddots & \vdots \\ cov(x^m, x^1) & \cdots & cov(x^m, x^j) & \cdots & var(x^m) \end{bmatrix}$$

Citons quelques propriété de la matrice de variance-covariance :

— La matrice V est symétrique, c'est à dire $cov(x, y) = cov(y, x)$.

— Ses valeurs propres sont positives ou nulles.

La mesure de liaison linéaire entre deux variables normalisées est donnée par la corrélation linéaire.

2.7.5 Matrice de corrélation

L'étude de la matrice de corrélation permet de repérer immédiatement les caractères fortement corrélés et les caractères qui ne sont pas du tout corrélés.

Soit Σ la matrice de corrélation définit comme suit :

$$\Sigma = R^T D R = D_{\frac{1}{s}} V D_{\frac{1}{s}} \tag{2.23}$$

$$\Sigma = \begin{bmatrix} 1 & \cdots & r(x^1, x^j) & \cdots & r(x^1, x^m) \\ \vdots & \ddots & \vdots & & \vdots \\ r(x^j, x^1) & \cdots & 1 & \cdots & r(x^j, x^m) \\ \vdots & & \vdots & \ddots & \vdots \\ r(x^m, x^1) & \cdots & r(x^m, x^j) & \cdots & 1 \end{bmatrix}$$

elle est symétrique puisque $r(x^j, x^{j'}) = r(x^{j'}, x^j)$.

Pour les caractères centrés réduits, la matrice de variance est égale avec celle de corrélation :

Σ résume la structure des dépendances linéaires entre les m variables.

Comme V, la matrice de corrélation est une matrice symétrique positive.

2.7.5.1 Étude de la matrice de corrélation

Le calcul des valeurs et des vecteurs propres de la matrice de corrélation fournit toutes les informations qu'on cherche.

Soit $\lambda_j \in \{1, \dots, m\}$ les m valeurs propres de Σ et leur m vecteurs propres $\vec{V}_j \in \{1, \dots, m\}$, vu que la matrice de corrélation est symétrique alors ses valeurs propres sont réelle et positives, et ses vecteurs ont des coordonnées réelles. L'existence d'une ou plusieurs valeurs propres nulles, signifie que les variables ne sont pas linéairement indépendantes les unes au autres.

2.7.6 Moments d'inertie

L'inertie mesure la dispersion autour du barycentre, c'est une variance multidimensionnelle.

2.7.6.1 Inertie du nuage des individus

L'inertie totale est définie comme la somme pondérée des carrés des distances de chaque individu au centre de gravité.

On note I_G le moment d'inertie du nuage des individus par rapport au centre de gravité G , dans l'espace initial, on a donc :

$$I_G = p \sum_{i=1}^n d_M^2(x_i, G) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m a_j (x_i^j - \bar{x}^j)^2 = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m a_j (y_i^j)^2$$

$$I_G = \frac{1}{n} \sum_{i=1}^n x_{ci}^T M x_{ci} \quad (2.24)$$

sous forme matricielle, l'inertie totale est donnée par :

$$I_G = \text{Tr}(MV) \quad (2.25)$$

Si on utilise la métrique réduite $M = D_{\frac{1}{s^2}}$, l'inertie I_G est la trace de la matrice de corrélation :

$$I_G = \text{Tr} \left(D_{\frac{1}{s^2}} V \right) = \text{Tr} \left(D_{\frac{1}{s}} V D_{\frac{1}{s}} \right) = \text{Tr}(R) = p \quad (2.26)$$

Dans le cas de la métrique usuelle $M = I$, on montre facilement que l'inertie du nuage est la trace de la matrice de variance covariance :

$$I_G = \text{Tr}(V) \quad (2.27)$$

Cette inertie I_G traduit la position du nuage dans l'espace par rapport à G . lorsque elle est grande cela signifie que le nuage est très dispersé, tandis que la concentration du nuage autour de son centre de gravité implique des valeurs d'inerties faibles.

2.7.6.2 Inertie du nuage par rapport à un axe

La mesure de la dispersion donne l'information concernant la direction des axes d'allongement du nuage qui contribue dans la construction des axes principaux qui portent le maximum d'informations.

Considérons un axe Δ de l'espace des individus dirigé par un vecteur u unitaire, les projections orthogonales des individus sur cette axe forment une nouvelle variable t^j combinaison linéaire des variables initiales.

$$t_i^j = \langle u_j, x_i \rangle_M = x_i^T M u_j \quad (2.28)$$

Posant :

$$w = Mu_j \quad (2.29)$$

Donc :

$$t^j = \begin{bmatrix} t_1^j \\ \vdots \\ t_i^j \\ \vdots \\ t_n^j \end{bmatrix} = XM u_j = Xw = \sum_{j=1}^P x^j w_j \quad (2.30)$$

L'ensemble de ces variables forment un sous espace vectoriel de \mathfrak{R}^m de dimension égale ou inférieure à m .

L'inertie du nuage par rapport au point p le long de l'axe Δ est l'inertie des point t_i^j :

$$I_P(u_j) = \sum_i p_i \|t_i^j - p\|^2 \quad (2.31)$$

On considère le point p est l'origine, on a :

$$I(u_j) = \sum_i p_i \|t_i^j\|^2 \quad (2.32)$$

D'après l'équation (2.28) :

$$\|t_i^j\|^2 = (x_i^T w_j)^2$$

$$\|t_i^j\|^2 = (x_i^T w_j)^T (x_i^T w_j)$$

$$\|t_i^j\|^2 = w_j^T x_i x_i^T w_j \quad (2.33)$$

On remplaçons (2.33) dans (2.32), on trouve :

$$I(u_j) = \sum_i p_i w_j^T x_i x_i^T w_j \quad (2.34)$$

On utilise la symétrie du produit scalaire, on déduit :

$$I(u_j) = w_j^T \left(\sum_i p_i x_i x_i^T \right) w_j$$

Comme les variables son supposes centrées, a partir de (2.21) et (2.22) on a :

$$I(u_j) = w_j^T V w_j \quad (2.35)$$

De l'équation (2.29), on a :

$$(w_j)^T = (M u_j)^T = u_j^T M$$

donc :

$$I(u_j) = u_j^T M V M u_j \quad (2.36)$$

La matrice $\Gamma = M V M$ est appelée matrice d'inertie

2.7.7 Éléments principaux

2.7.7.1 Axes principaux

Le premier axe correspond à la droite qui restituera la plus grande partie de la dispersion du nuage après la projection c-à-d celui qui minimise les distances entre les points e_i et leur projections pour avoir l'image la moins déformée du nuage, selon le théorème de Pythagore :

$$\|x_i - G\|^2 = \|x_i - t_i^j\|^2 + \|t_i^j - G\|^2$$

cela revient à maximiser l'inertie du nuage projeté.

On cherche alors la droite Δ_1 de \mathbb{R}^m qui passe par le centre de gravité G et de dispersion maximal.

Soit u_1 un vecteur porté par cette droite, d'après ce qui précède (2.35) :

$$I(u_1) = w_1^T V w_1$$

Le problème d'optimisation s'écrit :

$$\max_{\|u_1\|=1} w_1^T V w_1$$

Donc :

$$I(\Delta_1) = \frac{u_1^T M V M u_1}{u_1^T M u_1} \quad (2.37)$$

Cette fonction est maximale lorsque sa dérivé par rapport à u_1 est nulle.

Soit $I(\Delta_1) = \lambda_1$, nous devons chercher les extrémum de L définie par :

$$L = u_1^T M V M u_1 - \lambda_1 u_1^T M u_1 \quad (2.38)$$

ça dérivé est donnée par :

$$\frac{dL}{du_1} = 2(MVMu_1 - \lambda_1 Mu_1) = 0 \quad (2.39)$$

Le vecteur u_1 cherché est caractérisé par :

$$VMu_1 = \lambda_1 u_1 \quad (2.40)$$

donc u_1 est le vecteur propre de VM associé à la valeur propre λ_1 .

Le premier axe pour lequel le nuage des individus a l'inertie maximum a comme vecteur directeur unitaire le premier vecteur propre associé à la plus grande valeur propre de la matrice VM.

Lorsque l'inertie présenté par le premier axe n'est pas satisfaisant, on cherche ensuite un deuxième axe orthogonal au premier et d'inertie maximum, soit Δ_2 le deuxième axe passant par G et d'un vecteur directeur unitaire u_2 .

D'après (2.37), l'inertie du nuage par rapport à cette axe est :

$$I(\Delta_2) = \frac{u_2^T MVMu_2}{u_2^T Mu_2} \quad (2.41)$$

sous la contraintes $u_2^T u_1 = 0$.

En appliquant les même calcul précédente, on trouve que u_2 est le vecteur propre de VM correspondant à la deuxième plus grande valeur propre.

Les deux premiers axes forment un sous espace de dimension 2 qui porte le maximum d'inertie.

La matrice VM étant symétrique possède des vecteurs propres orthogonaux deux à deux et normée constituons une base orthonormée, ses valeurs propre sont toutes positives, dans ce cas les axes principaux sont tous orthogonaux.

D'une façon générale, pour un tableau de données comportant p mesures, on pourra définir p axes d'importance décroissante et non corrélées qui correspond au dimension intrinsèque des données, le seconde axe contiendra une information non redondante par rapport au premier et ainsi de suite.

On appelle axes principaux d'inertie les axes de direction des vecteur propres de VM normés.

Considérons le vecteur propre u de la matrice VM associé à la valeur propre λ .

L'inertie expliqué par u_j est calculé par :

$$I(u_j) = u_j^T V u_j = \lambda u_j^2 = \lambda \quad (2.42)$$

chaque valeur propre mesure la porte de variance expliquée par l'axe de direction u_j .

Contribution des axes à l'inertie totale

L'indicateur utilisé pour représenter la capacité d'un axe à présenter le nuage est le taux d'inertie du sous espace, qui représente le pourcentage d'inertie totale expliquée par ce espace.

On peut montrer que l'inertie projeté est additive lorsqu'il se décompose sur des axes orthogonaux, en utilisant le théorème de Huyegens, la décomposition de l'inertie totale du nuage des individus est :

$$I_G = I_{\Delta_1} + I_{\Delta_2} + \dots + I_{\Delta_2} = \lambda_1 + \lambda_2 + \dots + \lambda_P \quad (2.43)$$

Nous pouvons donc écrire :

La contribution relative de l'axe Δ_j à l'inertie totale du nuage des individus est égale :

$$\frac{I_{\Delta_k}}{I_G} = \frac{\lambda_j}{\lambda_1 + \lambda_2 + \dots + \lambda_P} \quad (2.44)$$

Ainsi, le pourcentage d'inertie expliqué par le plan engendré par les deux premier axes Δ_1 et Δ_2 est égale à :

$$\frac{I_{\Delta_1 \oplus \Delta_2}}{I_G} = \frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2 + \dots + \lambda_P} \quad (2.45)$$

Ces pourcentage d'inertie informe sur la variabilité du nuage des individus porté par ces sous espaces qui contribué dans le choix du nombre d'axe principaux qui restituer une part non négligeable de l'inertie totale, souvent on se contente de gardé les premier axe qui présent un pourcentage d'inertie proche de 1.

2.7.7.2 Facteurs principaux

Pour chaque axe de direction u_j est associe la forme linéaire w_j coordonnées M-orthogonale sur l'axe défini par u_j .

Comme $u_j \in (\mathfrak{R}^m)$ et $w_j = M u_j$, donc w_j est un élément de $(\mathfrak{R}^m)^*$ (dual de l'espace des individus) qui définit une combinaison linéaire des variables x_1, x_2, \dots, x_m .

Puisque u_j est un vecteur propre de VM , M -normé à 1, il suit que son associé le facteur principal w_j est le vecteur propre de MV , M^{-1} normé, en effet :

$$VMu = \lambda u \implies MVMu = \lambda Mu$$

Soit :

$$MVw = \lambda w \quad (2.46)$$

2.7.7.3 Composantes principales

Les composantes principales ce sont les nouvelles variables t^j dans le sous espace défini par les nouveaux axes principaux.

Pour obtenir t_i^j la i ème composante coordonnée de l'individu x_i sur l'axe Δ_j , on projette orthogonalement le vecteur $\overrightarrow{Gx_i}$ sur cet axe et on obtient :

$$t_i^j = \left\langle \overrightarrow{Gx_i}, \overrightarrow{u_j} \right\rangle = w_j^T x_i \quad (2.47)$$

La variable t^j est le vecteur qui renferme les composantes des projections de tout les individus sur l'axe Δ_j .

Ces composantes de variance maximale sont centrées de même que les variables de départ et non corrélées.

2.8 Interprétation des résultats d'une ACP

L'analyse en composantes principales est utilisée dans le but de la réduction du dimension de l'espace des individus afin de faciliter l'analyse, mais aussi pour avoir une représentation graphique pour visualiser les relations entre variables et faire apparaitre les approximations entre individus, pour apprécier la perte d'informations et évaluer la qualité de la représentation, l'indicateur globale basé sur l'inertie expliqué par les axes factorielles doit être complété par d'autres considérations, en effet il peut y'avoir des individus mal présentés malgré que les axes retenue présentent une part suffisamment grande de l'inertie totale du nuage.

Avant de conclure, pour nous assurer de la pertinence de la représentation nous effectuons deux étude dans l'ACP :

- L'étude des individus.
- L'étude des variables.

2.8.1 Étude des variables

Il est souvent utile d'étudier les contributions des variables à la formation des axes par la mesure de la corrélation entre les variables d'origine x^j , cette étude peut être utilisée dans l'interprétation du graphe d'individus, cela veut dire que si par exemple une corrélation entre une variable et l'axe factoriel est négative les individus qui prennent de faibles valeurs pour la variable prennent de fortes valeurs sur l'axe factoriel.

2.8.1.1 Corrélacion entre les CPs et les variables d'origine

Lorsque la qualité de représentation est bonne, les approximations observées des variables lue sur le graphique retrace bien leur proximité réelle, par contre si cette représentation est mal, la lecture direct des proximités sur le graphique peut donc s'avérer erroné due aux projections.

De même, on définit la qualité de représentation de la variable x^j sur l'axe représenté par la variable t^j par :

$$r(x^j, t^j) = \frac{\text{covariance}(x^j, t^j)}{\sqrt{\text{variance}(t^j)}\sqrt{\text{variance}(x^j)}} \quad (2.48)$$

Si les variables sont centré réduit, r est le coefficient de corrélation.

Les variables qui forment un angle nul avec l'axe sont corrélées, tandis que celles qui s'oppose sont corrélées négativement et enfin celles qui forme un angle droit ne sont pas corrélées

$r_{jj} = 1 \implies$ corrélation positive.

$r_{jj} = -1 \implies$ corrélation négative.

$r_{jj} = 0 \implies$ corrélation nulle.

Pour étudier les relations entre les variables il ne faut interpréter qu'a partir les variables les mieux représentées, pour cela en représente les variables sur les nouveaux axes.

2.8.1.2 Représentation des variables dans les nouveaux axes

Les cordonnées des variables sur les axes est leur coefficients de corrélations avec chaque axe, comme c'est coefficients variant entre -1 et 1, les variables seront représentées comme des points dans un cercle de rayon 1 qu'on appel cercle de corrélation.

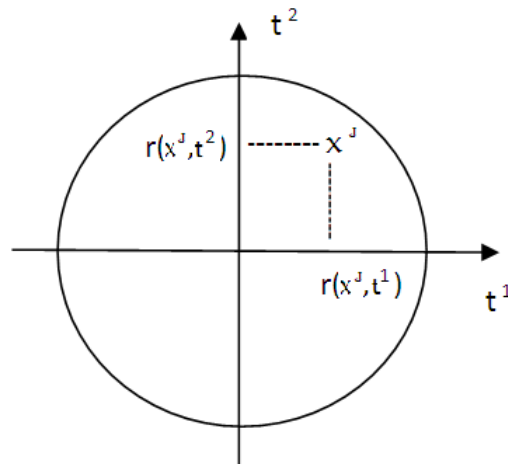


FIGURE 2.1 – Cercle de corrélation

Chaque variable x^j est représentée par l'abscisse $r(x^j, t^1)$ et l'ordonnée $r(x^j, t^2)$.

2.8.1.3 Études des liaisons entre les variables

Le graphe du cercle de corrélations va permettre une lecture direct sur les relations qui existent entre les variables en terme de corrélation si c'est variables sont bien représentées, une variable d'autant mieux représentée sur un axe qu'elle est proche du borde du cercle des corrélations, d'autant plus mal représenté qu'elle est proche de l'origine.

Lorsque deux variables sont proches et bien représentées donc elles sont corrélées positivement, ceci signifie que les deux variables varient dans le même sens, par contre si deux variables sont proches du cercle mais s'opposent sur un axe sont corrélées négativement, et lorsque l'une augment l'autre ce diminue, ainsi que deux variables orthogonales sont non corrélées.

Il faut garder à l'esprit que cette mesure de corrélation permet de détecter les liaisons linéaire, donc il peut y'avoir que deux variables aient un coefficient de corrélation nul ou très faibles si leur liaison n'est pas linéaire.

2.8.2 Étude des individus

La représentation des individus dans le sous espace principale permet d'évaluer les ressemblances observées entre les points du nuage des individus, et conclure à une proximité de point de vue des valeurs prises par les variables. Lorsque les projections des individus sur les nouveaux axes sont éloignées il y'aura pas de doute concernant la qualité de cette représentation, par contre la présence de regroupement est suspecté, dans ce cas nous regarderons en premier lieu la qualité de cette représentation.

2.8.2.1 Représentation des individus dans les nouveaux axes

Le nuage de point d'individus nous informe sur la ressemblance entre les individus, pour avoir une représentation des individus dans les nouveaux axes, il est indispensable de faire une projection orthogonale du vecteur \vec{Gx}_i sur le nouveau axe Δ_j cela en calculons les naturelles coordonner des individus dans les nouveaux axes.

Les cordonnés de cette individu dans le sous espace est obtenue par une transformation linéaire qui associe pour chaque individus de l'espace m un vecteur t_i dans le nouveaux sous espace

$$t_i = P^T x_i \quad (2.49)$$

tel que t_i est le vecteur des cordonnées de l'individu x_i dans le sous espace.

Soulignons que P est la matrice de projection, ces vecteurs sont les vecteurs propres de la matrice VM qui constitue une base orthonormée d'un sous espace \mathfrak{R}^m .

2.8.2.2 Qualité de représentation des individus

La bonne représentation du nuage de points d'individus aide à garder une bonne partie d'informations lors de la projections des points sur un axe ou sur un plan.

Alors pour une bonne représentation de l'individu x_i sur un axe (ou plan, ou sous espace), l'obligation que l'angle formé entre le vecteur \vec{Gx}_i et l'axe (ou plan, ou sous espace) soit petit doit être vérifié. Le passage par le calcul du cosinus de cette angle ou bien le cosinus carré (pour éviter tout problèmes de signe qui peuvent falsifier les résultats) est la solution pour bien représenté l'individu x_i . En effet, en appuyant sur le théorème de Pythagore, la possibilité de montrer que le \cos^2 d'un vecteur avec un plan engendré par deux vecteurs orthogonaux est égal à la somme des cosinus des angles du vecteur avec chacun des deux vecteurs orthogonaux qui engendrent un plan sera assurer, et cette propriété se généralise à l'angle d'un vecteur avec n'importe lequel sous espace de dimension ℓ . Pour bien savoir si x_i est bien représenté, on doit trouver le \cos^2 de l'angle entre le vecteur \vec{Gx}_i et l'axe (ou le plan, ou le sous espace) proche de 1. Aussi on dit que deux individu sont proche dans l'espace si ils sont bien représentés en projection sur un axe (ou plan, ou sous espace) et qui ont des projections proches.

On définit le \cos^2 de l'angle entre \vec{Gx}_i et un axe Δ_j de vecteur directeur unitaire u_j comme suit :

$$\cos^2(\alpha_{ij}) = \frac{\langle \vec{Gx}_i, u_j \rangle^2}{\|\vec{Gx}_i\|^2} = \frac{u_j^T x_{ci} x_{ci}^T u_j}{x_{ci}^T x_{ci}}$$

$$\cos^2(\alpha_{ij}) = \frac{\left[\sum_{j=1}^m (x_i^j - \bar{x}_j) u_j \right]^2}{\sum_{j=1}^m (x_i^j - \bar{x}_j)^2} \quad (2.50)$$

Le \cos^2 de l'angle $\alpha_{ijj'}$ entre $\overrightarrow{Gx_i}$ et le plan engendré par deux axes Δ_j et $\Delta_{j'}$ se calcul en appuyant sur le théorème de Pythagore comme suit :

$$\cos^2(\alpha_{ijj'}) = \cos^2(\alpha_{ij}) + \cos^2(\alpha_{ij'}) \quad (2.51)$$

Le calcul de la qualité de représentation d'un individu x_i se fait en calculant le \cos^2 de l'angle $\overrightarrow{Gx_i}$ avec le sous espace et cela c'est l'étude du pourcentage d'inertie.

2.8.2.3 Interprétation des nouveaux axes en fonction des individus

L'inertie due à un individu x_i se fait en calculant l'inertie $I(\Delta_j)$ portée par l'axe Δ_j .

Sachant que, l'axe qui passe par G présente moins de déformation du nuage et une perte minime d'information en projetant le nuage des individus sur cette axe puisque Δ_j est proche de l'ensemble des points du nuage.

Contribution absolue d'un individu à un axe

L'individu dont sa projection sur un axe est éloignée par rapport au centre de gravité du nuage d'individu valorise cette axe en information.

Aussi si les projections des points d'individus sur un axe sont proche du centre de gravité créent un affaiblissement d'information sur cette axe. Ces contributions nous aide à interpréter :

Étant donné, soit $I(\Delta_j)$ qui est égal à $\frac{1}{n} \sum_{i=1}^n d^2(t_i^j, G)$, la contribution absolue de x_i à cette inertie est donnée par :

$$Ca(x_i/\Delta_j) = \frac{1}{n} d^2(t_i^j, G) \quad (2.52)$$

tel que t_i^j est la projection de x_i sur cet axe.

Contribution relative d'un individu à un axe

La contribution relative c'est un outil pour interpréter les axes principaux avec les individus.

Pour chaque individu x_i sa contribution relative à l'inertie qui est portée par cette axe est définit comme suit :

$$Cr(x_i/\Delta_j) = \frac{\frac{1}{n}d^2(t_j^j, G)}{I(\Delta_j)} = \frac{\frac{1}{n}\langle \overrightarrow{Gx_i}, u_j \rangle^2}{\lambda_J}$$

$$Cr(x_i/\Delta_j) = \frac{\frac{1}{n}u_j^T x_i x_i^T u_j}{\lambda_J} \quad (2.53)$$

notons que : $\sum_{i=1}^n Cr(x_i/\Delta_j) = 1$.

2.9 Analyse en composantes principales normées

Dans les paragraphes précédents, nous avons développés les principes de l'analyse en composantes principales d'une façon générale sans distinguer entre l'ACP simple et l'ACP normée. La différence entre ces deux approches réside dans le choix de la distance à considérer. si le choix porte sur la distance euclidien classique on prend dans ce cas la métrique usuelle $M=I$, les variables seront traitées de façon symétrique, et l'analyse sera focalisée sur la matrice de variance-covariance, dans ce cas les axes principaux et les axes factoriels correspond sans distinction entre les deux au vecteur propre de V .

Par contre l'ACP normée se distingue par l'utilisation de la métrique $M = D_{\frac{1}{s^2}}$, et l'analyse mène à l'exploitation de la matrice de corrélation Σ .

En pratique on travaillera sur des données centré-réduit (tableau R), donc on utilisera la métrique $M = I$.

Dans ce cas les facteurs principaux seront donc les vecteurs propres de la matrice de corrélation :

$$\Sigma w = \lambda w \quad (2.54)$$

2.10 Conclusion

Dans ce chapitre, nous avons présenté le principe de l'analyse en composantes principales linéaires. L'idée principale de l'ACP est de réduire la dimension du tableau des données en gardant le mieux possible l'information qui représente le tableau des données. Afin d'aboutir à cette fin, cette méthode permet d'étudier la corrélation entre les variables et identifie leurs dépendances entre elles pour obtenir une représentation compacte de ces dernières. Cette réduction ne sera possible que si les variables initiales ne sont pas indépendantes et qui ont des coefficients de corrélations non nuls. Ces variables initiales sont transformées après la projection en nouvelles variables, appelées composantes principales. Elles sont obtenues par combinaison linéaire des précédentes variables et elles sont

ordonnées et non corrélées entre elles.

L'analyse en composantes principales cherche à identifier les vecteurs propres ortho-normaux et les valeurs propres de la matrice de variance ou de corrélation des variables originales. Les vecteurs propres sont utilisés pour construire les composantes principales, et les valeurs propres sont les variances des CPs.

Chapitre 3

Diagnostic par l'ACP

Introduction

Les progrès permanents réalisés dans le domaine de diagnostic ont abouti à une amélioration de surveillance et de conduite des processus.

La mise en œuvre d'une procédure de diagnostic approprié au fonctionnement du capteur est indispensable avant toute utilisation des mesures pour déceler d'éventuelles anomalies de fonctionnement des capteurs et assurer les missions du système. Cependant, il existe de nombreuses méthodes de diagnostic et qui peuvent être classer en deux catégories selon qu'elles utilisent ou non un modèle à priori. Les méthodes dites sans modèles relèvent des techniques de traitement des données, regroupent un ensemble de méthodes basée sur le concept de redondance analytique existant entre les différentes variables du processus. Parmi lesquelles, les méthodes reposées sur l'analyse en composantes principales ont reçu une attention particulière et ont été largement utilisées pour la surveillance des processus industriels.

L'analyse en composantes principales, si elle a été initialement destiné à la compression de données, mais aussi un outil intéressant pour la détection et la localisation des erreurs de mesures et des dysfonctionnements des procédés (Dunia,1996) . Le principe de cette approche est de transformer un ensemble de données ou de mesures des capteurs obtenues lors de fonctionnement normale en un nombre restreint de nouvelles variables qui préservent le maximum de la variabilité des données mesurées sur le système, ensuite utilisées pour la modélisation, l'estimation d'état, et le diagnostic de fonctionnement de système.

L'identification du modèle consiste en premier lieu à décomposer la matrice des corrélations en valeurs et vecteurs propres pour estimer ses paramètres, ensuite à déterminer la structure du modèle par le choix du nombre de composantes à retenir. Vu l'impact de la sélection des CPs sur la capacité de la détection et localisation des défauts affectant les données, pour cela plusieurs critères et règles ont été proposés dans la littérature qui aide

au choix du nombre qui permet d'élaborer un modèle optimal.

Le modèle sera considéré dans la suite comme une référence pour la détection des variables aberrantes basées sur l'approche de reconstruction qui estime les variables en fonction des autres variables, les défauts sont alors détectés en comparant le comportement observé avec celui donné par le modèle. Selon les défauts à détecter simples ou multiples sur les variables redondantes ou non, Tamura et Tsujita [42] ont proposés une nouvelle méthode qui cherche non pas un modèle ACP unique mais plutôt différents modèles ACP chacun sensible aux défauts à détecter.

La qualité de la reconstruction mesurée par l'erreur quadratique de prédiction, sert à détecter la présence ou non des défauts qui se traduit par le franchissement du seuil de détection. La certitude de l'apparition des défauts mène ensuite à la recherche de la cause des défauts qui est réalisée par le calcul des contributions de chaque variable à la valeur de l'erreur pour localiser là où les variables qui sont en cause.

3.1 Identification du modèle ACP

L'analyse en composantes principales, en particulier dans le domaine du diagnostique est utilisée comme une méthode de modélisation des systèmes à surveiller, basé principalement sur l'analyse statistique pour la mise en évidence des corrélations linéaires significatives entre les variables du système sans formuler d'une façon explicite le modèle du système. L'objectif est de fournir un modèle mathématique réduit peut être utilisé pour prédire un comportement observé du processus, ce modèle comprend un certain nombre de paramètres dont les valeurs correspondent aux vecteurs propres obtenus par la diagonalisation de la matrice des corrélations comportant des mesures des variables du système, recueillies par des capteurs lors du fonctionnement normale, ensuite à définir la structure du modèle qui consiste à chercher le nombre optimum de vecteurs propres à maintenir dans le modèle et qui permet de présenter le système avec le moins possible perte d'informations.

L'identification du modèle ACP débute par la construction d'une matrice contenant l'ensemble des données disponibles sans distinction entre les entrées et les sorties du système [43].

Au départ les n mesures des m variables utilisées sont collectées dans une matrice de données X , les variables du système sont généralement exprimées avec des unités de mesures et des échelles différents dont la nécessité d'effectuer un prétraitement qui consiste à normaliser les données avant toute analyse, par la suite, l'analyse est localisée sur la matrice de corrélation Σ de X

Soit x_i le vecteur de moyenne nulle et de variance unité contenant les m variables (mesures ou commande) observées sur le système à l'instant (i) :

$$x_i = \left[x_i^1 \quad \dots \quad x_i^j \quad \dots \quad x_i^m \right]$$

La matrice X comportant les observations recueillies sur le système au différentes instants est définie par :

$$X = \left[x^1 \quad \dots \quad x^j \quad \dots \quad x^m \right] \in \mathfrak{R}^{n \times m} \quad (3.1)$$

Un modèle ACP est défini globalement par la matrice des premier vecteurs propres de la matrice de corrélation qui expriment le lien entre les différentes variables donnée par (2.23), une décomposition spectrale de cette dernière permet d'écrire :

$$\Sigma = P \Lambda P^T \quad (3.2)$$

avec

$$P = \left[p^1 \quad \dots \quad p^j \quad \dots \quad p^m \right] \in \mathfrak{R}^{m \times m} \text{ et } P^T P = P^T P = I_m$$

où Λ est une matrice diagonale dont les éléments sont les valeurs propres λ mis dans l'ordre décroissant et P la matrice des vecteurs propres associées aux valeurs propres λ de la matrice de corrélation Σ dont chaque vecteur p^j représente une direction de l'espace des données porteuses d'une inertie équivalente à λ associée à cette direction.

[43]

La matrice des vecteurs propres p^j obtenue par l'équation (3.2) appelée matrice de projection représente la matrice de passage vers le nouveau sous espace, la relation linéaire entre la matrice des données X et des composantes T est définie comme suit :

$$X = T P^T \quad (3.3)$$

avec

$$T = \left[t^1 \quad \dots \quad t^j \quad \dots \quad t^m \right] \in \mathfrak{R}^{n \times m}$$

L'ACP peut être également utilisée comme outil de diagnostic, l'application de cette méthode pour la détection des défauts nécessite après le choix du nombre de CPs porteuses d'informations significatives et suffisantes pour expliquer la variabilité d'un processus à travers sa base de données X , d'effectuer une projection orthogonale des données sur le sous espace vectoriel engendré par ses composantes, cette réduction entraîne une perte d'informations mesurées par l'erreur entre la représentation exacte et celle approcher par la projection inverse, en effet, cette erreur représente la quantité résiduelle \tilde{X} projection des données sur le sous espace engendré par les $m - \ell$ derniers vecteurs propres les moins

significative.

La matrice des observations X peut être restituée dans le sous espace réduit qui est donnée par son estimation comme suit :

$$\hat{X} = \sum_{j=1}^{\ell} t^j p^{jT} = \hat{T}_{\ell} \hat{P}_{\ell}^T \quad (3.4)$$

comme $\hat{T}_{\ell} = X \hat{P}_{\ell}$, alors :

$$\hat{X} = X \hat{P}_{\ell} \hat{P}_{\ell}^T = X \hat{C} \in SP \quad (3.5)$$

Par conséquent, l'erreur d'approximation a pour valeurs l'expression suivante :

$$E = X - \hat{X}$$

qui correspond à :

$$\tilde{X} = X - \hat{X} = X (1 - \hat{C}) = X \tilde{C} \in SR \quad (3.6)$$

avec

$$\hat{C} = \hat{P}_{\ell} \hat{P}_{\ell}^T \text{ et } \tilde{C} = \tilde{P} \tilde{P}^T$$

et

$$\hat{P}_{\ell} = \begin{bmatrix} p_1 & \dots & p_{\ell} \end{bmatrix} \text{ et } \tilde{P} = \begin{bmatrix} p_{\ell+1} & \dots & p_m \end{bmatrix}$$

D'une façon générale, si on ne conserve que ℓ premières CPs déterminés par les critères de choix discuté par la suite, on peut décomposer la matrice des données en deux matrices \hat{X} et \tilde{X} définies par (3.5), (3.6) comme suite[45] :

$$X = \hat{X} + \tilde{X}$$

par conséquent, la matrice des vecteurs propres et composantes principales se partitionne comme suite :

$$P = \left[\hat{P}_{\ell} \mid \tilde{P}_{m-\ell} \right], T = \left[\hat{T}_{\ell} \mid \tilde{T}_{m-\ell} \right] \quad (3.7)$$

où, \hat{X} et \tilde{X} sont respectivement les parties modélisées et résiduelles

Avec :

$$\tilde{X}^T \hat{X} = 0$$

\hat{P}_{ℓ} et \tilde{P} , représentent les premiers et ($\ell \ll m$) vecteurs propres qui forment respectivement le sous espace vectoriel réduit appelé sous espace principal (SP), et le sous-espace

résiduel (SR) orthogonal avec (SP).

\hat{C} et \tilde{C} représentent les matrices de projection respectivement sur le (SP) et (SR) qui constituent le modèle ACP.

.

Pour obtenir notre modèle, il nous reste à déterminer le nombre ℓ de composantes principales à conserver.

3.2 Choix de la dimension

L'étape la plus délicate dans l'identification d'un modèle par l'ACP est la détermination de la structure de ce dernier, de rang inférieur qui maintient l'essentielle des données à analyser, sous la contrainte d'avoir le moins possible de déformation.

Un diagnostic basé sur une ACP pour la construction d'un modèle adéquat au système étudié repose toujours sur une étape primordiale qui porte sur le choix optimal de nombre de CPs à retenir, et qui a un impact significatif sur chaque étape de détection et de localisation, un choix non judicieux revient soit à perdre des informations essentielles contenues dans l'espace de départ en projetant certaines variables dans le sous-espace des résidus (SR) si un nombre de composantes réduit est retenu ce qui provoque des erreurs de modélisation dans le modèle des résidus donc il peut y avoir des générations de fausses alarmes, ou bien si par contre on abuse dans le choix, un nombre inutile de composantes qui correspondent aux valeurs propres les plus faibles par rapport à celles qui véhiculent réellement l'information du système et qui sont porteuses de bruits seront retenues dans le modèle en projetant certaines variables dans le sous-espace principal, donc il y a un risque de non-détection des défauts.

La question qui se pose alors et qui a été largement débattue dans la littérature concerne le choix du nombre de composantes principales qui doit être retenu. Plusieurs critères sont mis à disposition pour déterminer le nombre de valeurs propres significatives. Parmi lesquelles on cite : le Pourcentage Cumulé de la Variance (PCV), la Variance Non Reconstituée (VNR) et la validation croisée (PRESS) disponibles dans la littérature[45]. Selon le critère utilisé, cette sélection permet d'éliminer les informations non-pertinentes et redondantes donc de réduire la dimension de l'espace de départ et rendre l'ensemble des données plus représentatif du problème.

3.2.1 Pourcentage Cumulé de la Variance (PCV)

Cette méthode est basée sur la contribution des axes à l'inertie totale, comme les composantes principales portent l'information concernant la variance des mesures effectuées sur les variables du système et qui est quantifiée par les valeurs propres, le pourcentage d'inertie expliqué par les ℓ CPs retenues, peut être utilisé comme un critère PCV pour

déterminer la structure optimale d'un modèle ACP[46].

Le pourcentage de variance expliqué par les ℓ premières CPs retenues s'écrit comme suit :

$$PCV(\ell) = 100 \left(\frac{\sum_{j=1}^{\ell} \lambda_j}{\sum_{j=1}^m \lambda_j} \right) \% \quad (3.8)$$

La sélection du nombre des composantes à retenir est sélectionné par le plus petit nombre des CPs qui représente une valeur d'inertie proche de 1, tel que le pourcentage qu'on veut conserver (PCV) est compris entre 90% et 95% ou 99% de la variance totale, les composantes sont choisies donc successivement dans un ordre d'inertie décroissant jusqu'à ce que ce pourcentage soit atteint ou dépassé[47].

Cependant, le choix de la dimension du modèle ACP basée sur une telle approche est très subjective. Le problème est que nous cherchons un modèle tout en retenant le maximum possible de variance mais en même temps avec un minimum possible de CPs.

3.2.2 Critère de Kaiser

La règle de Kaiser consiste à ne conserver que les variables (CPs principales) de variances exprimées par les valeurs propres supérieures à la moyenne de l'inertie totale du nuage I_G (variance moyenne des variables initiales).

$$\lambda_i > \frac{I_G}{p}$$

En ACP normée on s'intéresse qu'aux axes associées à des valeurs propres supérieures à 1.

comme $I_G = p$ (2.26), donc :

$$\lambda_j > 1 \quad (3.9)$$

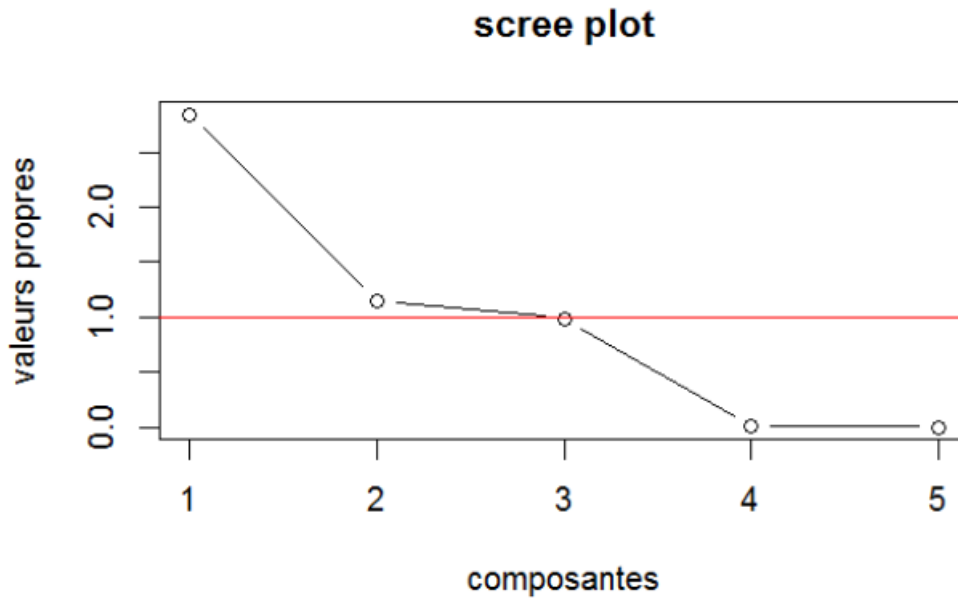


FIGURE 3.1 – Nombre de valeurs propres

3.2.3 Variance de l'Erreur de Reconstruction (VER)

Cette méthode qui a été proposée par Qin et Dunia[48] est basée essentiellement sur la reconstruction, qui consiste à exploiter les relations de redondance qui existent entre les variables du modèle ACP pour estimer certaines variables[47].

Le choix du nombre ℓ du composantes principales est obtenu par la minimisation de la variance de l'erreur de cette reconstruction qui dépend principalement de composantes retenues. Dans un premier lieu, nous allons montrer comment reconstruire une variable et enfin nous présenterons le critère à minimiser pour la sélection du nombre de composantes.

3.2.3.1 Approche itérative de reconstruction

C'est une méthode qui consiste à estimer une variable à partir les autres mesures du modèle ACP. Cette procédure est réalisée plusieurs fois en remplaçant la variable pour la quelle on a estimé sa valeur par son estimé, et en la ré-estime de nouveau jusqu'à ce que la variable estimée converge vers la valeur z_i [49].

soit x_i le vecteur de mesure de X à l'instant (i)

$$x_i = \left[x_i^1 \quad \dots \quad x_i^j \quad \dots \quad x_i^m \right]^T$$

et

$$\hat{C}_\ell = \left[c^1 \quad \dots \quad c^j \quad \dots \quad c^\ell \right]$$

avec

$$c^j = \left[c_{-i}^T \quad c_i^j \quad c_{+i}^T \right]^T$$

où les indices +i et -i désignent, respectivement les vecteurs formés par les (i-1) premiers et les (m-i) derniers éléments du vecteur c^j j ème colonne de \hat{C}_ℓ .

La j ème composante de x_i a été reconstruite par (Wise et Gallager, 1996).

Calcul de la premier estimation :

D'après (3.3), on a :

$$t_i^j = (p^j)^T x_i \quad (3.10)$$

Donc :

$$t_i^j = p_1^j x_i^1 + p_2^j x_i^2 + \dots + p_i^j x_i^j + \dots + p_m^j x_i^m$$

On sépare la variable à reconstruire du reste des variables, on obtient :

$$t_i^j = p_i^j x_i^j + \begin{bmatrix} p_1^j & \dots & 0 & \dots & p_m^j \end{bmatrix} x_i$$

l'estimation de x_i^j est $\hat{x}_i^j = t_i^j p_i^j$, donnée par :

$$t_i^j p_i^j = p_i^j p_i^j x_i^j + \begin{bmatrix} p_1^j p_i^j & \dots & 0 & \dots & p_m^j p_i^j \end{bmatrix} x_i$$

$$\hat{x}_i^j = c_i^j x_i^j + \begin{bmatrix} c_1^j & \dots & 0 & \dots & c_m^j \end{bmatrix} x_i$$

$$\hat{x}_i^j = c_i^j x_i^j + \begin{bmatrix} c_{-i}^T & 0 & c_{+i}^T \end{bmatrix} x_i$$

on remplace x_i^j par $z_i^{(iter)}$ et \hat{x}_i^j par $z_i^{(iter-1)}$ dans l'expression, il vient l'algorithme suivante :

$$z_i^{(iter-1)} = c_i^j z_i^{(iter)} + \begin{bmatrix} c_{-i}^T & 0 & c_{+i}^T \end{bmatrix} x_i$$

avec $z_i^{(0)} = x_i^j$

où $z_i^{(iter-1)}$ est la valeur estimée par l'ACP à l'itération précédente, et $z_i^{(iter)}$ est la nouvelle valeur de la mesure estimée par le modèle à partir du vecteur $\hat{x}_i^{(iter-1)}$ qui représente le vecteur x_i dont la j ème valeur à été reconstruite à l'itération précédente, à la convergence de l'algorithme, la solution z_i vérifie :

$$z_i (1 - c_i^j) = \begin{bmatrix} c_{-i}^T & 0 & c_{+i}^T \end{bmatrix} x_i$$

ainsi, z_i sera donnée par :

$$z_i = \frac{\begin{bmatrix} c_{-i}^T & 0 & c_{+i}^T \end{bmatrix} x_i}{1 - c_i^j} \quad (3.11)$$

avec $c_i^j < 1$.

Selon la valeur de c_i^j on distingue :

- Si $c_i^j = 1$, donc la j ème variable ne peut pas être reconstruit à partir des autres variables puisque ne sont pas corrélées (la variable z_j est indépendant des autres).

- Si $c_i^j \neq 1$, la j ème variable est en fonction des autres, donc peut être reconstruit par les autres variables.

La condition nécessaire de reconstruction d'une variable par le modèle est :

$$\tilde{\xi}_j \neq 0 \quad (3.12)$$

Principe de la variance non reconstruite

La valeur d'une variable estimée en utilisant toutes les autres variables est différente de la variable initial, il y a toujours une partie de la variance des mesures qui ne peuvent pas être reconstruites à partir des autres mesures et qu'on appelle l'erreur de reconstruction.[49]

L'erreur de reconstruction de la j ème composante de x_i est donnée par :

$$\begin{aligned} x_i^j - \hat{x}_i^j &= \xi_j^T (x_i - \hat{x}_i) = \left(\xi_j^T - \frac{\begin{bmatrix} c_{-i}^T & 0 & c_{+i}^T \end{bmatrix}}{1 - c_i^j} \right) x_i \\ \xi_j^T (x_i - \hat{x}_i) &= \left(\frac{\begin{bmatrix} 0 & 1 - c_i^j & 0 \end{bmatrix} - \begin{bmatrix} c_{-i}^T & 0 & c_{+i}^T \end{bmatrix}}{1 - c_i^j} \right) x_i \\ \xi_i^T (x_i - \hat{x}_i) &= \left(\frac{\begin{bmatrix} -c_{-i}^T & 1 - c_i^j & -c_{+i}^T \end{bmatrix}}{1 - c_i^j} \right) x_i = \frac{(I - \hat{C}) \xi_j}{1 - c_i^j} x_i \\ \xi_j^T (x_i - \hat{x}_i) &= \frac{\tilde{\xi}_j^T x_i}{\tilde{\xi}_j^T \tilde{\xi}_j} \end{aligned} \quad (3.13)$$

où $\tilde{\xi}_j = (I - \hat{C}_\ell) \xi_j$ et $\tilde{\xi}_j^T \tilde{\xi}_j = (1 - c_i^j)$ avec $\xi_j = [0 \dots 1 \dots 0]^T$

La variance de l'erreur de reconstruction de la j ème composante de $x(k)$ est donnée par :

$$\rho_J(\ell) = var \xi_j^T (x_i - \hat{x}_i) = \frac{\tilde{\xi}_j^T \Sigma \tilde{\xi}_j}{(\tilde{\xi}_j^T \tilde{\xi}_j)^2} \quad (3.14)$$

Cette variance représente la variance de l'amplitude estimée du défaut. Pour choisir la dimension du modèle ACP, l'idée est de minimiser cette variance puisque les données utilisées pour la modélisation sont supposées être sans défauts, et cela doit être effectué pour tout les variables.

Le critère à minimisé donc est présenté sous forme d'une somme pondérer de tout les variances :

$$VNR(\ell) = \sum_{j=1}^m \frac{\rho_j(\ell)}{\xi_j^T \Sigma \xi_j} = \sum_{j=1}^m \frac{\tilde{\xi}_j^T \Sigma \tilde{\xi}_j}{\left(\tilde{\xi}_j^T \Sigma \tilde{\xi}_j\right)^2 (\xi_j^T \Sigma \xi_j)} \quad (3.15)$$

La pondération par $\xi_j^T \Sigma \xi_j$ qui représente la variance originelle de la j ème variable, est utilisé afin d'éviter les problèmes des unités des variances non reconstruites et qui veut 1 dans le cas où les variables sont initialement réduites.

En se basant sur ce critère le nombre optimal des composantes principales est celui qui offre une meilleure reconstruction.

$$\ell_{op} = \min \{VNR(\ell)\} \quad (3.16)$$

3.2.4 Validation croisée

La validation croisé est un critère statistique très utile pour le choix du nombre de composantes optimal pour un modèle ACP.

L'idée général de cette procédure de validation est basée sur la minimisation de la somme des carrés des erreurs de prédiction entre les données observées et celles estimées par le modèle obtenue à partir d'un jeu d'identification différent, on mesure la moyenne des distances des prédictions à leur observation pour des différentes valeurs de ℓ croissantes, selon Wold (1978) et Eastment et Krzanowski (1982) [50] le nombre des CPs optimal à retenir est déterminé par la ℓ ième composante pour laquelle le minimum de cette moyenne connu sous le nom de PRESS n'est plus amélioré par l'ajout de CPs supplémentaires :

$$PRESS(\ell) = \frac{1}{nm} \sum_{k=1}^n \sum_{j=1}^m (\hat{x}_i^\ell(k) - x_i(k))^2 \quad (3.17)$$

où $\hat{x}_i^\ell(k)$ représente la prédiction de $x_i(k)$ qui correspond à la k ème mesure de la i ème variable, en utilisant un modèle ACP constitué de ℓ CPs.

Ainsi, le modèle optimal est alors constitué de ℓ ième composantes nécessaires pour une prédiction adéquate.

Cependant, pour un nombre d'observation très important et suite à un développement par la théorie des perturbation, un développement de Taylor a permis de montrer que la fonction PRESS est très proche de la somme des valeurs propres du (SR) et peut être approximée comme suit [47] :

$$PRESS(\ell) \approx \frac{1}{m} \sum_{j=\ell+1}^m \lambda_j \quad (3.18)$$

En effet cette quantité représente la variance du vecteur résiduel :

$$\text{PRESS}(\ell) \approx \text{var}(\tilde{x}) = \text{E}(\|\tilde{x}\|^2) = \text{var}(\tilde{t}) = \text{E}(\|\tilde{t}\|^2) = \text{trace}(\tilde{\Lambda})$$

Par ailleurs, la sélection du nombre de CPs par le critère PRESS n'est pas très objective par rapport aux restes des critères à cause du nombre important d'observations considérées[51].

3.3 Détection de défauts

Les défauts se définissent comme des anomalies, c'est à dire des conditions anormales qui diminuent ou suppriment l'aptitude du système à remplir ses fonctions et dégradent la qualité des produits fabriqués. L'apparition du défaut s'accompagne par un changement dans la corrélation des variables du système.

Conservant une observation x_i sur le système à l'instant (i) composé de la mesure sans défaut x_i^* perturbé par un bruit blanc ν_i identiquement distribué et un défaut sur la j ème variable d'amplitude \tilde{h}_i défini de la manière suivante :

$$x_i = x_i^* + \nu_i + \xi_j \tilde{h}_i \quad (3.19)$$

Pour constater l'influence du défaut sur les données en projet l'observation sur les différents sous espaces (principales et résiduel), sa projection sur le SP est donnée par :

$$\hat{t}_i = \hat{p}^T x_i$$

$$\hat{t}_i = \hat{p}^T (x_i^* + \nu_i + \xi_j \tilde{h}_i) \quad (3.20)$$

sa projection dans SR est :

$$\tilde{t}_i = \tilde{p}^T x_i$$

$$\tilde{t}_i = \tilde{p}^T (x_i^* + \nu_i + \xi_j \tilde{h}_i)$$

$$\tilde{t}_i = \tilde{p}^T \nu_i + \tilde{p}^T \xi_j \tilde{h}_i \quad (3.21)$$

On remarque que dans le sous espace résiduel la projection de l'observation affecté se manifeste comme un résidu composé de bruit et de défauts puisque la mesure sans défauts se projette dans l'espace principale, contrairement à la projection dans SP, il reste l'influence des données. Un défaut perturbe donc à la fois les projections des observations dans les deux sous espaces, de ce fait, les projections des données à tester sur les deux sous espaces

peuvent être utilisées pour la mise en évidence d'un comportement anormale.

Selon l'amplitude de défaut, pour détecter un défaut dans l'espace principale, il faut que son amplitude soit de grande valeur pour qu'il influence sur la dispersion dans SP, par contre dans le sous espace réduit, la dispersion des projections est de faible valeur par rapport à SP, donc les défauts de faibles amplitude peuvent être détecter

Pour détecter un défaut en s'appuyant sur l'ACP, deux méthodes statistique sont habituellement considérés qui repose sur l'analyse statistique des données collectés au moyenne d'une partis des axes principaux qui sont la statistique de Hotteling calculé à partir des premières composantes et l'erreur quadratique d'estimation .

- La statistique T^2 de Hotteling donne une mesure de la distance d'un échantillon à la moyenne des processus à l'intérieur des plans définit par les composantes principales retenues.
- L'erreur de prédiction SPE (Squared Prediction Error), calculées à partir d'une loi du χ^2 , sous la condition d'une distribution normale multivariée respectivement pur X et \tilde{X} . Dans le cas où ces deux indicateurs sont utilisés conjointement, il est possible d'utilisé un seul indicateur qui est l'indice combiné proposé par Raich and Cinar [52] et Yue and Qin [53].

3.3.1 L'erreur de prédiction SPE

La statistique SPE, consiste à la détection des défauts dans un espace résiduel qui provoquent une déformation dans la corrélation entre les variables. En basant sur les erreurs, la SPE tient à surveiller toutes les variables à observées, pour cela en estimant à partir du modèle ACP les mesures délivrées par les capteurs pour les comparer avec ce dernier, l'amplitude de cette erreur d'estimation est affectée par la présence de défaut(s) dans une mesure qui sera révélée par un dépassement de seuil qu'on nomme seuil de confiance.

L'indicateur de détection SPE est donné par le vecteur des erreurs d'estimation comme suit :

$$SPE(k) = e(k)^T e(k) \quad (3.22)$$

Le système étudié est considéré en fonctionnement normal à l'instant k si :

$$SPE(k) < \delta^2 \quad (3.23)$$

où δ^2 est le seuil de confiance connu sous le nom Q-statistique qui est développé par Jackson et Mudholkar[54], donnée par :

$$\delta^2 = \theta_1 \left[\frac{Ca\sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right] \frac{1}{h_0} \quad (3.24)$$

avec

$$h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2} \quad (3.25)$$

et

$$Ca = \theta_1 \left[\frac{\left(\frac{\|e\|^2}{\theta_1} \right)^{h_0} - 1 - \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2}}{\sqrt{2\theta_2 h_0^2}} \right] \quad (3.26)$$

Sachant que $\theta_i = \sum_{j=\ell+1}^m \lambda_j^i$ pour $i \in \{1, 2, 3\}$ et λ_j est la jème valeur propre de la matrice de corrélation et Ca représente la limite du seuil de confiance $(1 - a)$.

à noter qu'il y a deux conditions à respecter pour arriver à la détection des défauts :

- Le vecteur des mesures suit une distribution normale multivariée.
- Un franchissement de seuil est effectuée.

3.3.2 Statistique T^2 de Hotelling

On appelle statistique de Hotelling, l'indicateur qui permet de détecter les anomalies présentées dans l'espace principale engendrées par les ℓ premières composantes principales par la mesure des variations pondérées par les valeurs propre des projections des observations sur cette espace [45] :

$$T^2(k) = \hat{t}(k)^T \hat{\Lambda}^{-1} \hat{t}(k) \quad (3.27)$$

La présence de défaut à l'instant k est traduit par un dépassement d'une limite qu'on appelle le seuil de détection de T^2 , lorsque le nombre d'observations n est considérable, ce seuil peut être approximé par une distribution de chi-2 (χ^2) avec ℓ degrés de liberté pour un seuil de confiance α qu'on note $\chi_{\ell, \alpha}^2$.

Dans le cas contraire, si le nombre d'observations n est réduit, la limite supérieure appropriée pour la statistique T^2 pour un seuil de confiance α est déterminée par :

$$\frac{\ell(n^2 - 1)}{n(n - \ell)} F_{\ell, n - \ell, \alpha} \quad (3.28)$$

où $F_{\ell, n-\ell, \alpha}$ est la distribution de Fisher avec ℓ et $n - \ell$ degré de liberté.

Cependant, l'utilisation de cette statistique dans la détection n'est pas très performant lorsque il s'agit des défauts de faibles amplitudes[45], car ces dernier n'ont pas d'influence sur la mesure de l'erreur d'estimation vu leurs faibles valeurs devant les variations significatives des données présentées dans le sous espace principale (SP) ainsi comme les variables de faible variance sont exclus dans la sélection des composantes principales à retenir dans l'espace (SP), de ce fait le défaut qui entache ces variables n'est pas détectable par T^2 par ce que ces variables elles même sont en dehors des limites de contrôle.

3.4 Localisation de défauts

La localisation consiste à trouver où le défaut est survenu. Plusieurs méthodes ont été développées dans la littérature. Pour l'analyse en composantes principales, il existe trois approches qui sont : l'approche par structuration des résidus, les approches utilisant des bancs de modèle et l'approche reposant sur le principe de calcul des contributions.

Notre travail est basé sur l'approche de calcul des contributions, et c'est une méthode qui largement utilisée dans le diagnostic par ACP.

3.4.1 Localisation par calcul des contributions

Le principe de contribution s'appuie généralement sur la quantification de la part de chaque variable dans le calcul d'un indice de détection donné. Cette méthode est basée sur l'idée que les variables avec les plus grandes contributions à l'indice de détection de défaut sont les plus susceptibles d'être porteuse de défaut, elle consiste à déterminer la contribution de chaque variable à l'indice de défaut utilisé.

Cette méthode est déterminée par (Miller, 1993), elle exprime la contribution d'une variable x^j du vecteur de mesure x_i à l'indice SPE comme étant le carré de son propre résidu :

$$Const_j^{SPE} = (e_j(k))^2 = (x_i - \hat{x}_i) = \tilde{x}_i^2 \quad (3.29)$$

3.5 Conclusion

Ce chapitre est consacré à définir l'utilisation de l'ACP dans le domaine du diagnostic et à la présentation des techniques de détections et de localisations de défauts. Un modèle ACP est alors défini par la matrice des premiers vecteurs propres de la matrice de corrélations des données, cette matrice permet à la fois de définir la projection permettant d'avoir les composantes principales et la projection inverse permettant d'estimer les données originales .

Pour l'identification du modèle ACP, il est indispensable de déterminer le nombre de composantes principales à garder dans le modèle ACP, afin de déterminer ce nombre plusieurs critères de sélection ont été présentés.

Chapitre 4

Application à un système hydraulique

4.1 Avant-propos

Il apparait que certains facteurs (ou variables) de mesure ne sont tout simplement pas les bonnes :

Soit qu'ils sont des fonctions affines d'une autre variable (donc colinéaires).

Soit qu'ils ne représentent pas grand chose.

Et surtout que l'axe 'vrai' de la mesure peut être entre les trois axes de trois variables de mesure.

Il est donc souhaitable de changer les axes et donc de changer le type de variables de mesure. La plupart du temps il n'est guère évident en application d'ingénieur (ou en physique appliquée) de changer d'outil de mesure et même de grandeur de mesure.

L'aspect mathématique du 'changement de représentation' de l'analyse par composantes principales est bien pratique. Mais l'abord mathématique est difficile, l'emploi de matrice est trop compliqué et l'interprétation hasardeuse. Une fois tous ces écueils passé, l'ACP est une méthode qui a une bonne réflexion sur la représentation des phénomènes physiques et sur sa façon de mesurer.

4.2 Objectifs de l'ACP dans notre application

L'ACP dans notre travail est utilisée pour réduire la dimension du système afin d'élaborer un modèle ACP optimal qu'on va utiliser par la suite pour la détection et la localisation des défauts.

4.3 Procédé

4.3.1 Description du procédé

Le procédé à étudier est un système hydraulique de trois bacs communiquant comme le montre le schéma de la figure. Le premier bac (B1) est alimenté par un débit d'entrée noté $q_e(t)$, ce débit est fourni par une vanne contrôlée (VC). Les trois bacs de section effective S_i communiquent entre eux à travers des orifices de diamètre d_i de section Φ_i et de coefficient de débit α_d . Deux sorties de communication en bacs diamètre d_{si} et de section Φ_{si} soit réalisées. On note de plus $x_i(t)$ la hauteur du liquide dans chaque bac i , et $p_i(t)$ une perturbation de débit qui peut affecter chacun de ces bacs. On a naturellement : $x_1(t) \geq x_2(t) \geq x_3(t)$. La hauteur maximale pour les trois bacs est de 50 cm. En pratique, les sections des orifices d'écoulement correspondent aux diamètre suivants : $d_1 = 7.2\text{mm}$, $d_2 = 5.4\text{mm}$, $d_3 = 3.5\text{mm}$. En outre, les sections effectives des bacs sont identiques ($S_1 = S_2 = S_3 = 1.3 \times 10^{-3}\text{mm}^2$) et le coefficient de débit est donné par : $\alpha_d = 0.72 \text{ sec}^{-0.5}$. Le débit d'équilibre est choisi égal à $q_e(t) = q_{e0} = 100\text{L/h}$. Le débit maximum réalisable est de 140 L/h. On peut noter que q_{ec} et (S_1, S_2, S_3) sont les facteurs qui jouent sur le temps de réponse du système.

Les grandeur qui décrivent ce système sont l'entrée $q_e(t)$ qui est le débit d'entrée, la sortie mesurée $y(t) = x_3(t)$ qui est la hauteur de remplissage du troisième bac et les trois variables d'état $x_1(t)$, $x_2(t)$, $x_3(t)$ qui sont les hauteurs de remplissage des trois bacs.

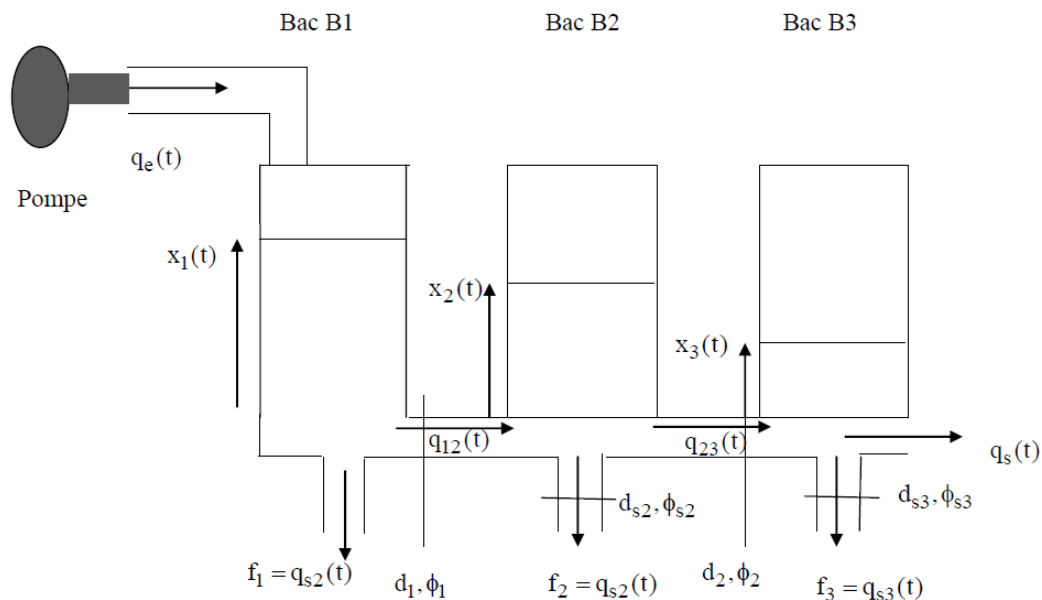


FIGURE 4.1 – Système à trois bacs.

4.3.2 Modèle

Pour obtenir le modèle mathématique, on établit le bilan des débits entrants et sortants pour chacun des bacs.

On aboutit alors au modèle d'état non linéaire décrit par les équations d'état suivantes :

$$S_1 \frac{dx_1(t)}{dt} = q_e(t) - q_{12}(t) - q_{s1}(t)$$

$$S_2 \frac{dx_2(t)}{dt} = q_{12}(t) - q_{s2}(t) - q_{23}(t)$$

$$S_3 \frac{dx_3(t)}{dt} = q_{23}(t) - q_{s3}(t) - q_s(t)$$

La relation de Bernoulli permet d'exprimer le débit sortant d'un bac i vers un bac j en fonction des différences des hauteurs d'eau en amont et en aval, c'est-à-dire :

$$q_{ij} = a_{ij} \sqrt{\Delta x}$$

Où les coefficients a_{ij} sont donnés en fonction du coefficient du débit moyen α_d et de la section Φ par la relation suivante :

$$a_{ij} = \alpha_d \Phi \sqrt{2g}$$

Tenant compte de ces relations, en définissant par $x(t) = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}^T$ le vecteur d'état, par $u(t) = q_e(t)$ l'entrée de commande et par $y_i(t) = x_i(t)$, $i = 1, 2, 3$, les trois sorties, le système s'écrit alors

$$\dot{x}(t) = f(x(t), u(t))$$

$$y(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \\ y_3(t) \end{bmatrix} = \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix}$$

avec

$$f(x(t), u(t)) = \begin{bmatrix} \frac{1}{S_1}(u(t) - a_{12}\sqrt{x_1(t) - x_2(t)} - f_1(t)) \\ \frac{1}{S_2}(a_{12}\sqrt{x_1(t) - x_2(t)} - a_{23}\sqrt{x_2(t) - x_3(t)} - f_2(t)) \\ \frac{1}{S_3}(a_{23}\sqrt{x_2(t) - x_3(t)} - f_3(t) - q_s(t)) \end{bmatrix}$$

Les débits de fuite $f_2 = q_{s1} = a_{s1}\sqrt{x_1}$, $f_2 = q_{s2} = a_{s2}\sqrt{x_2}$, $f_3 = q_{s3} = a_{s3}\sqrt{x_3}$ seront considérés comme des défauts du système inattendus. Si $q_{si} = 0$ ($i = 1, 2, 3$) alors il n'y a pas de défaut par contre si $q_{si} \neq 0$ alors le système est défectueux.

On considère l'entrée constante d'équilibre $u(t) = u_0 = 54\text{L/h} = 1.5 \times 10^{-5} \text{m}^3/\text{s}$. Les valeurs des niveaux à l'équilibre sont égales à ($x_{10} = 0.2948\text{m}$, $x_{20} = 0.2814\text{m}$, $x_{30} = 0.2392\text{m}$, $y_0 = 0.2392\text{m}$) Le modèle d'état linéaire sans défaut autour du point d'équilibre (x_{10} , x_{20} , x_{30} , y_0) est donné par :

$$\frac{\delta x(t)}{dt} = A\delta x(t) + B\delta u(t)$$

$$\delta y(t) = C\delta x(t)$$

Où $\delta u(t) = u(t) - u_0$, $\delta x(t) = x(t) - x_0$, $\delta y(t) = y(t) - y_0$, représentent les petites variations de l'entrée, des variables d'état et de la sortie autour de valeurs d'équilibre.

Les matrices A, B et C sont données par :

$$A = \begin{bmatrix} -0.4139 & 0,4139 & 0 \\ 0,4139 & -0.5685 & 0.1367 \\ 0 & 0.1367 & -0.1608 \end{bmatrix}; B = \begin{bmatrix} 769.2308 \\ 0 \\ 0 \end{bmatrix}; C = I_3$$

Le modèle linéaire avec les défauts s'écrit :

$$\frac{\delta x(t)}{dt} = A\delta x(t) + B\delta u(t) + B_s f(t)$$

$$\delta y(t) = C\delta x(t)$$

$f(t) = \begin{bmatrix} f_1(t) \\ f_2(t) \\ f_3(t) \end{bmatrix}$: défauts système correspondant aux fuites des trois bacs. La matrice de distribution des défauts est donnée par :

$$B_f = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

4.3.3 Génération des données

Système sans défaut : Entrée échelon de valeur variation sur l'entrée égale à $\delta u(t) = 10^{-5} \text{m}^3/\text{s}$.

Système avec défaut : prendre les défauts d'amplitude $F = 1.0 \cdot 10^{-3} \text{m}^3/\text{s}$.

4.3.4 Simulation du modèle

Une fois le modèle linéaire du système établie, nous pouvons aborder l'aspect lié à la simulation de celui-ci en utilisant MATLAB. Cela permet par la suite de mettre en évidence le comportement du système dans le cas sein et dans le cas où le système est défaillant.

Chaque niveau est mesuré sur une plage de fonctionnement $t=700$ seconde. La figure ci dessous montre l'évolution de ces trois niveaux.

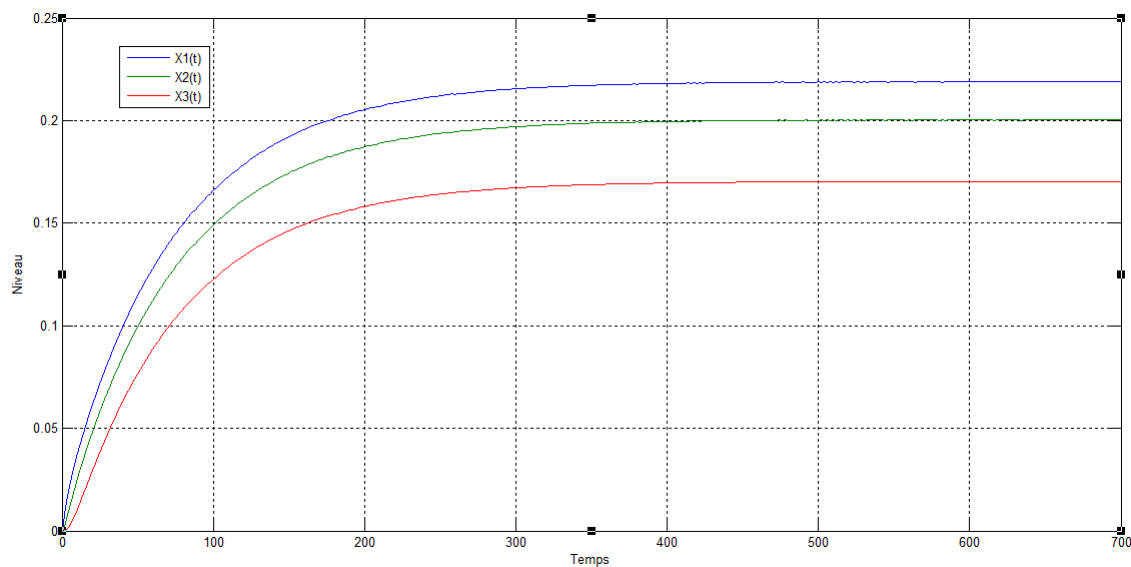


FIGURE 4.2 – Les niveaux dans les trois bacs

4.4 Réduction de la dimension

Afin de réduire la dimension de notre système, il faut passer par les étapes suivantes :

- Collecter les différentes mesures.
- Normalisation.
- Calcul de la matrice de corrélation.
- Calcul des valeurs propres et vecteurs propres.
- Choix de la dimension des données.
- Déduire les coordonnées factorielles dans le nouveau sous-espace.
- Choix de nombre d'axe.

4.4.1 Tableau du données

Le système à étudier contient des données qui sont présentées dans un tableau récapitulant les mesures du niveau dans les trois bacs et l'alimentation du système. La matrice X comporte donc quatre variables qui sont représentées dans le tableau suivant :

Variabes	Capteurs
X1	Le niveau du liquide dans bac 1
X2	Le niveau du liquide dans bac 2
X3	Le niveau du liquide dans bac 3
X4	Débit d'alimentation ($q_e(t)$)

TABLE 4.1 – Les variables du système

Plus précisément nous utilisons que trois variables seulement, à noter que la quatrième variable représente la commande dont sa valeur est constante (10^{-5}). L'ACP ne sait pas traiter les variables qui présentent des valeurs identiques pour chaque observation, ceci nous permet d'éliminer cette quatrième variable.

Nous allons échantillonner chaque signal de mesure par une période d'échantillonnage choisie égale à deux secondes ($\Delta t = 2s$) afin de générer un ensemble de données de 351 observations pour chacune des trois variables et de récupérer les valeurs dans un tableau à 351 ligne et 3 colonnes. présenter dans le tableau (4.2).

N°	X1	X2	X3	N°	X1	X2	X3	N°	X1	X2	X3	N°	X1	X2	X3	N°	X1	X2	X3
1	.0000	.0000	.0000	40	.1476	.1312	.1062	79	.1941	.1766	.1481	118	.2104	.1921	.1627	157	.2161	.1974	.1677
2	.0114	.0036	.0003	41	.1494	.1332	.1079	80	.1949	.1771	.1487	119	.2107	.1922	.1629	158	.2160	.1977	.1677
3	.0193	.0094	.0018	42	.1513	.1349	.1096	81	.1955	.1777	.1493	120	.2108	.1927	.1630	159	.2162	.1976	.1678
4	.0260	.0151	.0042	43	.1531	.1368	.1112	82	.1961	.1783	.1499	121	.2112	.1925	.1633	160	.2162	.1978	.1679
5	.0320	.0204	.0072	44	.1549	.1384	.1128	83	.1968	.1789	.1504	122	.2111	.1932	.1634	161	.2163	.1979	.1680
6	.0376	.0256	.0107	45	.1566	.1401	.1143	84	.1974	.1795	.1510	123	.2115	.1931	.1636	162	.2164	.1979	.1680
7	.0429	.0304	.0143	46	.1583	.1417	.1158	85	.1980	.1799	.1515	124	.2117	.1934	.1638	163	.2165	.1979	.1681
8	.0479	.0351	.0181	47	.1599	.1432	.1173	86	.1984	.1807	.1520	125	.2119	.1936	.1640	164	.2165	.1980	.1681
9	.0526	.0396	.0219	48	.1614	.1448	.1187	87	.1991	.1811	.1525	126	.2121	.1937	.1642	165	.2167	.1980	.1682
10	.0572	.0441	.0257	49	.1632	.1461	.1201	88	.1996	.1817	.1530	127	.2123	.1939	.1643	166	.2166	.1981	.1683
11	.0616	.0483	.0294	50	.1644	.1479	.1214	89	.2001	.1822	.1534	128	.2125	.1940	.1645	167	.2167	.1982	.1683
12	.0659	.0524	.0331	51	.1661	.1490	.1228	90	.2007	.1826	.1539	129	.2126	.1943	.1646	168	.2167	.1983	.1684
13	.0700	.0564	.0368	52	.1673	.1507	.1240	91	.2011	.1832	.1543	130	.2129	.1942	.1648	169	.2168	.1983	.1684
14	.0740	.0602	.0403	53	.1688	.1519	.1253	92	.2017	.1835	.1548	131	.2128	.1947	.1649	170	.2169	.1984	.1685
15	.0779	.0640	.0438	54	.1701	.1532	.1265	93	.2020	.1841	.1551	132	.2132	.1946	.1651	171	.2169	.1984	.1685
16	.0818	.0676	.0472	55	.1714	.1544	.1277	94	.2027	.1843	.1556	133	.2132	.1949	.1652	172	.2170	.1984	.1686
17	.0854	.0712	.0504	56	.1727	.1556	.1288	95	.2028	.1851	.1559	134	.2134	.1951	.1654	173	.2171	.1985	.1686
18	.0890	.0747	.0536	57	.1739	.1569	.1299	96	.2034	.1853	.1564	135	.2136	.1952	.1655	174	.2171	.1986	.1687
19	.0925	.0781	.0568	58	.1753	.1579	.1310	97	.2038	.1858	.1567	136	.2137	.1953	.1656	175	.2172	.1985	.1687
20	.0959	.0813	.0598	59	.1763	.1593	.1321	98	.2042	.1861	.1571	137	.2139	.1954	.1658	176	.2171	.1987	.1687
21	.0992	.0845	.0628	60	.1775	.1602	.1331	99	.2046	.1865	.1575	138	.2140	.1955	.1659	177	.2172	.1988	.1688
22	.1024	.0876	.0657	61	.1786	.1614	.1341	100	.2050	.1869	.1578	139	.2142	.1957	.1660	178	.2173	.1988	.1688
23	.1055	.0907	.0685	62	.1797	.1624	.1351	101	.2054	.1872	.1581	140	.2142	.1959	.1661	179	.2173	.1988	.1689
24	.1086	.0936	.0712	63	.1808	.1634	.1360	102	.2057	.1876	.1585	141	.2144	.1960	.1662	180	.2174	.1988	.1689
25	.1115	.0965	.0739	64	.1817	.1644	.1369	103	.2062	.1877	.1588	142	.2145	.1961	.1663	181	.2175	.1988	.1690
26	.1144	.0993	.0765	65	.1829	.1653	.1379	104	.2063	.1885	.1591	143	.2146	.1962	.1664	182	.2175	.1989	.1690
27	.1172	.1020	.0790	66	.1837	.1664	.1387	105	.2069	.1885	.1594	144	.2148	.1963	.1666	183	.2176	.1989	.1691
28	.1199	.1046	.0814	67	.1848	.1671	.1396	106	.2070	.1890	.1597	145	.2149	.1964	.1667	184	.2176	.1990	.1691
29	.1226	.1072	.0838	68	.1855	.1683	.1404	107	.2074	.1892	.1600	146	.2150	.1964	.1668	185	.2176	.1990	.1691
30	.1252	.1097	.0861	69	.1865	.1689	.1412	108	.2077	.1895	.1603	147	.2150	.1967	.1668	186	.2176	.1991	.1691
31	.1278	.1120	.0884	70	.1873	.1699	.1420	109	.2080	.1898	.1605	148	.2153	.1966	.1670	187	.2177	.1991	.1692
32	.1301	.1146	.0906	71	.1882	.1707	.1427	110	.2083	.1901	.1608	149	.2152	.1970	.1670	188	.2177	.1992	.1692
33	.1326	.1168	.0927	72	.1891	.1714	.1435	111	.2086	.1904	.1610	150	.2154	.1969	.1671	189	.2177	.1992	.1692
34	.1348	.1191	.0948	73	.1898	.1722	.1442	112	.2090	.1905	.1613	151	.2155	.1970	.1672	190	.2178	.1992	.1693
35	.1371	.1212	.0968	74	.1907	.1729	.1449	113	.2091	.1910	.1615	152	.2155	.1971	.1673	191	.2178	.1993	.1693
36	.1393	.1233	.0988	75	.1913	.1738	.1456	114	.2094	.1911	.1618	153	.2157	.1972	.1674	192	.2179	.1992	.1693
37	.1414	.1254	.1007	76	.1923	.1742	.1463	115	.2097	.1914	.1620	154	.2158	.1972	.1675	193	.2178	.1994	.1693
38	.1436	.1273	.1026	77	.1927	.1753	.1469	116	.2099	.1917	.1622	155	.2159	.1973	.1675	194	.2179	.1993	.1694
39	.1455	.1294	.1044	78	.1937	.1757	.1475	117	.2102	.1918	.1624	156	.2159	.1974	.1676	195	.2179	.1994	.1694

N°	X1	X2	X3	N°	X1	X2	X3	N°	X1	X2	X3	N°	X1	X2	X3
196	.2179	.1994	.1694	235	.2187	.2000	.1700	274	.2187	.2004	.1702	313	.2189	.2004	.1703
197	.2180	.1994	.1695	236	.2186	.2001	.1700	275	.2189	.2002	.1702	314	.2189	.2003	.1703
198	.2180	.1994	.1695	237	.2188	.1999	.1701	276	.2188	.2003	.1702	315	.2189	.2003	.1703
199	.2181	.1994	.1695	238	.2185	.2002	.1700	277	.2188	.2003	.1702	316	.2190	.2003	.1703
200	.2180	.1995	.1695	239	.2188	.1999	.1701	278	.2189	.2002	.1703	317	.2189	.2004	.1703
201	.2181	.1995	.1696	240	.2185	.2003	.1700	279	.2188	.2003	.1702	318	.2191	.2002	.1703
202	.2181	.1995	.1696	241	.2187	.2001	.1701	280	.2190	.2002	.1703	319	.2188	.2005	.1703
203	.2181	.1996	.1696	242	.2187	.2001	.1701	281	.2188	.2004	.1702	320	.2190	.2003	.1703
204	.2181	.1997	.1696	243	.2187	.2001	.1701	282	.2190	.2001	.1703	321	.2189	.2004	.1703
205	.2182	.1996	.1696	244	.2187	.2001	.1701	283	.2187	.2005	.1702	322	.2189	.2004	.1703
206	.2182	.1996	.1696	245	.2187	.2002	.1701	284	.2189	.2002	.1703	323	.2189	.2003	.1703
207	.2182	.1997	.1697	246	.2188	.2000	.1701	285	.2188	.2003	.1703	324	.2189	.2004	.1703
208	.2183	.1996	.1697	247	.2186	.2003	.1701	286	.2189	.2003	.1703	325	.2189	.2003	.1703
209	.2182	.1997	.1697	248	.2187	.2001	.1701	287	.2189	.2003	.1703	326	.2190	.2003	.1703
210	.2184	.1996	.1697	249	.2187	.2002	.1701	288	.2189	.2003	.1703	327	.2190	.2003	.1703
211	.2182	.1998	.1697	250	.2187	.2002	.1701	289	.2189	.2002	.1703	328	.2189	.2004	.1703
212	.2184	.1997	.1698	251	.2188	.2001	.1702	290	.2188	.2004	.1703	329	.2189	.2004	.1703
213	.2182	.1999	.1697	252	.2187	.2002	.1701	291	.2190	.2001	.1703	330	.2190	.2003	.1703
214	.2183	.1998	.1698	253	.2189	.2000	.1702	292	.2187	.2005	.1702	331	.2189	.2004	.1703
215	.2183	.1998	.1698	254	.2187	.2003	.1701	293	.2190	.2002	.1703	332	.2190	.2003	.1703
216	.2184	.1998	.1698	255	.2189	.2000	.1702	294	.2188	.2004	.1703	333	.2189	.2004	.1703
217	.2184	.1998	.1698	256	.2186	.2004	.1701	295	.2189	.2004	.1703	334	.2190	.2003	.1703
218	.2184	.1999	.1698	257	.2188	.2001	.1702	296	.2189	.2003	.1703	335	.2189	.2005	.1703
219	.2185	.1997	.1699	258	.2187	.2003	.1702	297	.2189	.2003	.1703	336	.2191	.2002	.1704
220	.2183	.2000	.1699	259	.2188	.2002	.1702	298	.2189	.2003	.1703	337	.2188	.2005	.1703
221	.2185	.1999	.1699	260	.2188	.2002	.1702	299	.2189	.2003	.1703	338	.2189	.2004	.1703
222	.2184	.1999	.1699	261	.2188	.2002	.1702	300	.2190	.2002	.1703	339	.2189	.2004	.1703
223	.2185	.1999	.1699	262	.2189	.2001	.1702	301	.2188	.2004	.1703	340	.2189	.2004	.1703
224	.2185	.1999	.1699	263	.2187	.2003	.1702	302	.2189	.2003	.1703	341	.2190	.2004	.1703
225	.2185	.2000	.1699	264	.2190	.2000	.1702	303	.2189	.2003	.1703	342	.2190	.2003	.1703
226	.2186	.1998	.1700	265	.2187	.2004	.1702	304	.2189	.2004	.1703	343	.2190	.2003	.1703
227	.2185	.2000	.1699	266	.2189	.2001	.1702	305	.2189	.2003	.1703	344	.2190	.2003	.1703
228	.2186	.1998	.1700	267	.2187	.2004	.1702	306	.2189	.2004	.1703	345	.2190	.2003	.1704
229	.2185	.2001	.1700	268	.2188	.2003	.1702	307	.2190	.2002	.1703	346	.2189	.2005	.1703
230	.2186	.2000	.1700	269	.2188	.2002	.1702	308	.2188	.2005	.1703	347	.2190	.2004	.1703
231	.2185	.2001	.1700	270	.2188	.2003	.1702	309	.2191	.2001	.1703	348	.2189	.2004	.1703
232	.2186	.2000	.1700	271	.2189	.2002	.1702	310	.2188	.2005	.1703	349	.2189	.2004	.1703
233	.2186	.2000	.1700	272	.2188	.2003	.1702	311	.2189	.2003	.1703	350	.2190	.2004	.1703
234	.2186	.2001	.1700	273	.2189	.2001	.1703	312	.2189	.2004	.1703	351	.2189	.2004	.1703

TABLE 4.2 – Tableau des données

4.4.2 Normalisation

L'opération de centrage consiste à enlever la moyenne à chaque variable. Les nouvelles coordonnées du point du nuage sont obtenus à partir de MATLAB conformément aux relations (2.6), (2.8) et sont illustré sur le tableau (4.3).

N°	X1	X2	X3	N°	X1	X2	X3	N°	X1	X2	X3	N°	X1	X2	X3	N°	X1	X2	X3
1	-0.1970	-0.1792	-0.1508	40	-0.0494	-0.0480	-0.0446	79	-0.0029	-0.0026	-0.0027	118	0.0134	0.0129	0.0119	157	0.0191	0.0182	0.0169
2	-0.1856	-0.1756	-0.1505	41	-0.0476	-0.0460	-0.0429	80	-0.0021	-0.0021	-0.0021	119	0.0137	0.0130	0.0121	158	0.0190	0.0185	0.0169
3	-0.1777	-0.1698	-0.1490	42	-0.0457	-0.0443	-0.0412	81	-0.0015	-0.0015	-0.0015	120	0.0138	0.0135	0.0122	159	0.0192	0.0184	0.0170
4	-0.1710	-0.1641	-0.1466	43	-0.0439	-0.0424	-0.0396	82	-0.0009	-0.0009	-0.0009	121	0.0142	0.0133	0.0125	160	0.0192	0.0186	0.0171
5	-0.1650	-0.1588	-0.1436	44	-0.0421	-0.0408	-0.0380	83	-0.0002	-0.0003	-0.0004	122	0.0141	0.0140	0.0126	161	0.0193	0.0187	0.0172
6	-0.1594	-0.1536	-0.1401	45	-0.0404	-0.0391	-0.0365	84	0.0004	0.0003	0.0002	123	0.0145	0.0139	0.0128	162	0.0194	0.0187	0.0172
7	-0.1541	-0.1488	-0.1365	46	-0.0387	-0.0375	-0.0350	85	0.0010	0.0007	0.0007	124	0.0147	0.0142	0.0130	163	0.0195	0.0187	0.0173
8	-0.1491	-0.1441	-0.1327	47	-0.0371	-0.0360	-0.0335	86	0.0014	0.0015	0.0012	125	0.0149	0.0144	0.0132	164	0.0195	0.0188	0.0173
9	-0.1444	-0.1396	-0.1289	48	-0.0356	-0.0344	-0.0321	87	0.0021	0.0019	0.0017	126	0.0151	0.0145	0.0134	165	0.0197	0.0188	0.0174
10	-0.1398	-0.1351	-0.1251	49	-0.0338	-0.0331	-0.0307	88	0.0026	0.0025	0.0022	127	0.0153	0.0147	0.0135	166	0.0196	0.0189	0.0175
11	-0.1354	-0.1309	-0.1214	50	-0.0326	-0.0313	-0.0294	89	0.0031	0.0030	0.0026	128	0.0155	0.0148	0.0137	167	0.0197	0.0190	0.0175
12	-0.1311	-0.1268	-0.1177	51	-0.0309	-0.0302	-0.0280	90	0.0037	0.0034	0.0031	129	0.0156	0.0151	0.0138	168	0.0197	0.0191	0.0176
13	-0.1270	-0.1228	-0.1140	52	-0.0297	-0.0285	-0.0268	91	0.0041	0.0040	0.0035	130	0.0159	0.0150	0.0140	169	0.0198	0.0191	0.0176
14	-0.1230	-0.1190	-0.1105	53	-0.0282	-0.0273	-0.0255	92	0.0047	0.0043	0.0040	131	0.0158	0.0155	0.0141	170	0.0199	0.0192	0.0177
15	-0.1191	-0.1152	-0.1070	54	-0.0269	-0.0260	-0.0243	93	0.0050	0.0049	0.0043	132	0.0162	0.0154	0.0143	171	0.0199	0.0192	0.0177
16	-0.1152	-0.1116	-0.1036	55	-0.0256	-0.0248	-0.0231	94	0.0057	0.0051	0.0048	133	0.0162	0.0157	0.0144	172	0.0200	0.0192	0.0178
17	-0.1116	-0.1080	-0.1004	56	-0.0243	-0.0236	-0.0220	95	0.0058	0.0059	0.0051	134	0.0164	0.0159	0.0146	173	0.0201	0.0193	0.0178
18	-0.1080	-0.1045	-0.0972	57	-0.0231	-0.0223	-0.0209	96	0.0064	0.0061	0.0056	135	0.0166	0.0160	0.0147	174	0.0201	0.0194	0.0179
19	-0.1045	-0.1011	-0.0940	58	-0.0217	-0.0213	-0.0198	97	0.0068	0.0066	0.0059	136	0.0167	0.0161	0.0148	175	0.0202	0.0193	0.0179
20	-0.1011	-0.0979	-0.0910	59	-0.0207	-0.0199	-0.0187	98	0.0072	0.0069	0.0063	137	0.0169	0.0162	0.0150	176	0.0201	0.0195	0.0179
21	-0.0978	-0.0947	-0.0880	60	-0.0195	-0.0190	-0.0177	99	0.0076	0.0073	0.0067	138	0.0170	0.0163	0.0151	177	0.0202	0.0196	0.0180
22	-0.0946	-0.0916	-0.0851	61	-0.0184	-0.0178	-0.0167	100	0.0080	0.0077	0.0070	139	0.0172	0.0165	0.0152	178	0.0203	0.0196	0.0180
23	-0.0915	-0.0885	-0.0823	62	-0.0173	-0.0168	-0.0157	101	0.0084	0.0080	0.0073	140	0.0172	0.0167	0.0153	179	0.0203	0.0196	0.0181
24	-0.0884	-0.0856	-0.0796	63	-0.0162	-0.0158	-0.0148	102	0.0087	0.0084	0.0077	141	0.0174	0.0168	0.0154	180	0.0204	0.0196	0.0181
25	-0.0855	-0.0827	-0.0769	64	-0.0153	-0.0148	-0.0139	103	0.0092	0.0085	0.0080	142	0.0175	0.0169	0.0155	181	0.0205	0.0196	0.0182
26	-0.0826	-0.0799	-0.0743	65	-0.0141	-0.0139	-0.0129	104	0.0093	0.0093	0.0083	143	0.0176	0.0170	0.0156	182	0.0205	0.0197	0.0182
27	-0.0798	-0.0772	-0.0718	66	-0.0133	-0.0128	-0.0121	105	0.0099	0.0093	0.0086	144	0.0178	0.0171	0.0158	183	0.0206	0.0197	0.0183
28	-0.0771	-0.0746	-0.0694	67	-0.0122	-0.0121	-0.0112	106	0.0100	0.0098	0.0089	145	0.0179	0.0172	0.0159	184	0.0206	0.0198	0.0183
29	-0.0744	-0.0720	-0.0670	68	-0.0115	-0.0109	-0.0104	107	0.0104	0.0100	0.0092	146	0.0180	0.0172	0.0160	185	0.0206	0.0198	0.0183
30	-0.0718	-0.0695	-0.0647	69	-0.0105	-0.0103	-0.0096	108	0.0107	0.0103	0.00								

N°	X1	X2	X3	N°	X1	X2	X3	N°	X1	X2	X3	N°	X1	X2	X3
196	0.0209	0.0202	0.0186	235	0.0217	0.0208	0.0192	274	0.0217	0.0212	0.0194	313	0.0219	0.0212	0.0195
197	0.0210	0.0202	0.0187	236	0.0216	0.0209	0.0192	275	0.0219	0.0210	0.0194	314	0.0219	0.0211	0.0195
198	0.0210	0.0202	0.0187	237	0.0218	0.0207	0.0193	276	0.0218	0.0211	0.0194	315	0.0219	0.0211	0.0195
199	0.0211	0.0202	0.0187	238	0.0215	0.0210	0.0192	277	0.0218	0.0211	0.0194	316	0.0220	0.0211	0.0195
200	0.0210	0.0203	0.0187	239	0.0218	0.0207	0.0193	278	0.0219	0.0210	0.0195	317	0.0219	0.0212	0.0195
201	0.0211	0.0203	0.0188	240	0.0215	0.0211	0.0192	279	0.0218	0.0211	0.0194	318	0.0221	0.0210	0.0195
202	0.0211	0.0203	0.0188	241	0.0217	0.0209	0.0193	280	0.0220	0.0210	0.0195	319	0.0218	0.0213	0.0195
203	0.0211	0.0204	0.0188	242	0.0217	0.0209	0.0193	281	0.0218	0.0212	0.0194	320	0.0220	0.0211	0.0195
204	0.0211	0.0205	0.0188	243	0.0217	0.0209	0.0193	282	0.0220	0.0209	0.0195	321	0.0219	0.0212	0.0195
205	0.0212	0.0204	0.0188	244	0.0217	0.0209	0.0193	283	0.0217	0.0213	0.0194	322	0.0219	0.0212	0.0195
206	0.0212	0.0204	0.0188	245	0.0217	0.0210	0.0193	284	0.0219	0.0210	0.0195	323	0.0219	0.0211	0.0195
207	0.0212	0.0205	0.0189	246	0.0218	0.0208	0.0193	285	0.0218	0.0211	0.0195	324	0.0219	0.0212	0.0195
208	0.0213	0.0204	0.0189	247	0.0216	0.0211	0.0193	286	0.0219	0.0211	0.0195	325	0.0219	0.0211	0.0195
209	0.0212	0.0205	0.0189	248	0.0217	0.0209	0.0193	287	0.0219	0.0211	0.0195	326	0.0220	0.0211	0.0195
210	0.0214	0.0204	0.0189	249	0.0217	0.0210	0.0193	288	0.0219	0.0211	0.0195	327	0.0220	0.0211	0.0195
211	0.0212	0.0206	0.0189	250	0.0217	0.0210	0.0193	289	0.0219	0.0210	0.0195	328	0.0219	0.0212	0.0195
212	0.0214	0.0205	0.0190	251	0.0218	0.0209	0.0194	290	0.0218	0.0212	0.0195	329	0.0219	0.0212	0.0195
213	0.0212	0.0207	0.0189	252	0.0217	0.0210	0.0193	291	0.0220	0.0209	0.0195	330	0.0220	0.0211	0.0195
214	0.0213	0.0206	0.0190	253	0.0219	0.0208	0.0194	292	0.0217	0.0213	0.0194	331	0.0219	0.0212	0.0195
215	0.0213	0.0206	0.0190	254	0.0217	0.0211	0.0193	293	0.0220	0.0210	0.0195	332	0.0220	0.0211	0.0195
216	0.0214	0.0206	0.0190	255	0.0219	0.0208	0.0194	294	0.0218	0.0212	0.0195	333	0.0219	0.0212	0.0195
217	0.0214	0.0206	0.0190	256	0.0216	0.0212	0.0193	295	0.0219	0.0212	0.0195	334	0.0220	0.0211	0.0195
218	0.0214	0.0207	0.0190	257	0.0218	0.0209	0.0194	296	0.0219	0.0211	0.0195	335	0.0219	0.0213	0.0195
219	0.0215	0.0205	0.0191	258	0.0217	0.0211	0.0194	297	0.0219	0.0211	0.0195	336	0.0221	0.0210	0.0195
220	0.0213	0.0208	0.0191	259	0.0218	0.0210	0.0194	298	0.0219	0.0211	0.0195	337	0.0218	0.0213	0.0195
221	0.0215	0.0207	0.0191	260	0.0218	0.0210	0.0194	299	0.0219	0.0211	0.0195	338	0.0219	0.0212	0.0195
222	0.0214	0.0207	0.0191	261	0.0218	0.0210	0.0194	300	0.0220	0.0210	0.0195	339	0.0219	0.0212	0.0195
223	0.0215	0.0207	0.0191	262	0.0219	0.0209	0.0194	301	0.0218	0.0212	0.0195	340	0.0219	0.0212	0.0195
224	0.0215	0.0207	0.0191	263	0.0217	0.0211	0.0194	302	0.0219	0.0211	0.0195	341	0.0220	0.0212	0.0195
225	0.0215	0.0208	0.0191	264	0.0220	0.0208	0.0194	303	0.0219	0.0211	0.0195	342	0.0220	0.0211	0.0195
226	0.0216	0.0206	0.0192	265	0.0217	0.0212	0.0194	304	0.0219	0.0212	0.0195	343	0.0220	0.0211	0.0195
227	0.0215	0.0208	0.0191	266	0.0219	0.0209	0.0194	305	0.0219	0.0211	0.0195	344	0.0220	0.0211	0.0195
228	0.0216	0.0206	0.0192	267	0.0217	0.0212	0.0194	306	0.0219	0.0212	0.0195	345	0.0220	0.0211	0.0195
229	0.0215	0.0209	0.0192	268	0.0218	0.0211	0.0194	307	0.0220	0.0210	0.0195	346	0.0219	0.0213	0.0195
230	0.0216	0.0208	0.0192	269	0.0218	0.0210	0.0194	308	0.0218	0.0213	0.0195	347	0.0220	0.0212	0.0195
231	0.0215	0.0209	0.0192	270	0.0218	0.0211	0.0194	309	0.0221	0.0209	0.0195	348	0.0219	0.0212	0.0195
232	0.0216	0.0208	0.0192	271	0.0219	0.0210	0.0194	310	0.0218	0.0213	0.0195	349	0.0219	0.0212	0.0195
233	0.0216	0.0208	0.0192	272	0.0218	0.0211	0.0194	311	0.0219	0.0211	0.0195	350	0.0220	0.0212	0.0195
234	0.0216	0.0209	0.0192	273	0.0219	0.0209	0.0195	312	0.0219	0.0212	0.0195	351	0.0219	0.0212	0.0195

TABLE 4.3 – Données centrées

Le nuage est maintenant centré autour de l'origine, mais il reste l'effet de taille due au différence des unités (niveau, débit), cela veut dire que les variables sont hétérogènes, pour remédier à cette situation un prétraitement est tout d'abord effectué nous allons réduire les variables par (2.18) en les dévissant par leurs écart-types afin d'accorder aux variables la même importance, on a donc le tableau (4.4) :

N°	X1	X2	X3	N°	X1	X2	X3	N°	X1	X2	X3	N°	X1	X2	X3
1	-4.5803	-4.3349	-3.9947	40	-1.1480	-1.1605	-1.1815	79	-0.0667	-0.0621	-0.0715	118	0.3124	0.3130	0.3153
2	-4.3152	-4.2478	-3.9868	41	-1.1061	-1.1121	-1.1364	80	-0.0481	-0.0500	-0.0556	119	0.3194	0.3154	0.3205
3	-4.1315	-4.1075	-3.9471	42	-1.0619	-1.0710	-1.0914	81	-0.0341	-0.0354	-0.0397	120	0.3217	0.3275	0.3232
4	-3.9757	-3.9696	-3.8835	43	-1.0201	-1.0250	-1.0490	82	-0.0202	-0.0209	-0.0238	121	0.3310	0.3227	0.3311
5	-3.8362	-3.8414	-3.8040	44	-0.9782	-0.9863	-1.0066	83	-0.0039	-0.0064	-0.0106	122	0.3287	0.3396	0.3338
6	-3.7060	-3.7155	-3.7113	45	-0.9387	-0.9452	-0.9669	84	0.0101	0.0081	0.0053	123	0.3380	0.3372	0.3391
7	-3.5827	-3.5994	-3.6159	46	-0.8992	-0.9065	-0.9271	85	0.0240	0.0178	0.0186	124	0.3426	0.3444	0.3444
8	-3.4664	-3.4857	-3.5153	47	-0.8620	-0.8702	-0.8874	86	0.0333	0.0371	0.0318	125	0.3473	0.3493	0.3497
9	-3.3571	-3.3768	-3.4146	48	-0.8271	-0.8315	-0.8503	87	0.0496	0.0468	0.0450	126	0.3519	0.3517	0.3550
10	-3.2502	-3.2679	-3.3139	49	-0.7852	-0.8000	-0.8132	88	0.0612	0.0613	0.0583	127	0.3566	0.3565	0.3576
11	-3.1478	-3.1663	-3.2159	50	-0.7573	-0.7565	-0.7788	89	0.0729	0.0734	0.0689	128	0.3612	0.3589	0.3629
12	-3.0479	-3.0671	-3.1179	51	-0.7178	-0.7298	-0.7417	90	0.0868	0.0831	0.0821	129	0.3635	0.3662	0.3656
13	-2.9525	-2.9703	-3.0199	52	-0.6899	-0.6887	-0.7099	91	0.0961	0.0976	0.0927	130	0.3705	0.3638	0.3709
14	-2.8595	-2.8784	-2.9272	53	-0.6550	-0.6597	-0.6755	92	0.1101	0.1049	0.1060	131	0.3682	0.3759	0.3735
15	-2.7688	-2.7864	-2.8345	54	-0.6248	-0.6282	-0.6437	93	0.1170	0.1194	0.1139	132	0.3775	0.3735	0.3788
16	-2.6781	-2.6993	-2.7444	55	-0.5945	-0.5992	-0.6119	94	0.1333	0.1243	0.1272	133	0.3775	0.3807	0.3815
17	-2.5944	-2.6122	-2.6596	56	-0.5643	-0.5702	-0.5828	95	0.1356	0.1436	0.1351	134	0.3821	0.3856	0.3868
18	-2.5107	-2.5276	-2.5749	57	-0.5364	-0.5387	-0.5536	96	0.1496	0.1484	0.1484	135	0.3868	0.3880	0.3894
19	-2.4293	-2.4453	-2.4901	58	-0.5038	-0.5145	-0.5245	97	0.1589	0.1605	0.1563	136	0.3891	0.3904	0.3921
20	-2.3502	-2.3679	-2.4106	59	-0.4806	-0.4806	-0.4954	98	0.1682	0.1678	0.1669	137	0.3938	0.3928	0.3974
21	-2.2735	-2.2904	-2.3311	60	-0.4527	-0.4589	-0.4689	99	0.1775	0.1775	0.1775	138	0.3961	0.3952	0.4000
22	-2.1991	-2.2154	-2.2543	61	-0.4271	-0.4298	-0.4424	100	0.1868	0.1872	0.1854	139	0.4007	0.4001	0.4027
23	-2.1270	-2.1404	-2.1801	62	-0.4015	-0.4056	-0.4159	101	0.1961	0.1944	0.1934	140	0.4007	0.4049	0.4053
24	-2.0549	-2.0703	-2.1086	63	-0.3759	-0.3814	-0.3920	102	0.2031	0.2041	0.2040	141	0.4054	0.4073	0.4080
25	-1.9875	-2.0001	-2.0371	64	-0.3500	-0.3572	-0.3682	103	0.2147	0.2065	0.2119	142	0.4077	0.4098	0.4106
26	-1.9200	-1.9323	-1.9682	65	-0.3271	-0.3355	-0.3417	104	0.2170	0.2259	0.2199	143	0.4100	0.4122	0.4133
27	-1.8549	-1.8670	-1.9020	66	-0.3085	-0.3088	-0.3205	105	0.2310	0.2259	0.2278	144	0.4147	0.4146	0.4186
28	-1.7921	-1.8041	-1.8384	67	-0.2829	-0.2919	-0.2967	106	0.2333	0.2380	0.2358	145	0.4170	0.4170	0.4212
29	-1.7293	-1.7412	-1.7748	68	-0.2667	-0.2629	-0.2755	107	0.2426	0.2428	0.2437	146	0.4193	0.4170	0.4239
30	-1.6689	-1.6807	-1.7139	69	-0.2434	-0.2484	-0.2543	108	0.2496	0.2501	0.2517	147	0.4193	0.4243	0.4239
31	-1.6084	-1.6251	-1.6530	70	-0.2248	-0.2242	-0.2331	109	0.2566	0.2573	0.2570	148	0.4263	0.4219	0.4292
32	-1.5549	-1.5622	-1.5947	71	-0.2039	-0.2048	-0.2146	110	0.2635	0.2646	0.2649	149	0.4240	0.4315	0.4292
33	-1.4968	-1.5089	-1.5391	72	-0.1829	-0.1879	-0.1934	111	0.2705	0.2718	0.2702	150	0.4286	0.4291	0.4318
34	-1.4456	-1.4533	-1.4834	73	-0.1667	-0.1685	-0.1748	112	0.2798	0.2743	0.2782	151	0.4310	0.4315	0.4345
35	-1.3922	-1.4025	-1.4305	74	-0.1457	-0.1516	-0.1563	113	0.2821	0.2864	0.2835	152	0.4310	0.4340	0.4371
36	-1.3410	-1.3517	-1.3775	75	-0.1318	-0.1298	-0.1377	114	0.2891	0.2888	0.2914	153	0.4356	0.4364	0.4398
37	-1.2922	-1.3009	-1.3272	76	-0.1085	-0.1201	-0.1192	115	0.2961	0.2960	0.2967	154	0.4379	0.4364	0.4424
38	-1.2410	-1.2549	-1.2768	77	-0.0992	-0.0935	-0.1033	116	0.3007	0.3033	0.3020	155	0.4403	0.4388	0.4424
39	-1.1968	-1.2041	-1.2291	78	-0.0760	-0.0838	-0.0874	117	0.3077	0.3057	0.3073	156	0.4403	0.4412	0.4451

N°	X1	X2	X3	N°	X1	X2	X3	N°	X1	X2	X3	N°	X1	X2	X3
157	0.4449	0.4412	0.4477	196	0.4868	0.4896	0.4927	235	0.5054	0.5041	0.5086	274	0.5054	0.5138	0.5139
158	0.4426	0.4485	0.4477	197	0.4891	0.4896	0.4954	236	0.5031	0.5065	0.5086	275	0.5100	0.5090	0.5139
159	0.4472	0.4460	0.4504	198	0.4891	0.4896	0.4954	237	0.5077	0.5017	0.5113	276	0.5077	0.5114	0.5139
160	0.4472	0.4509	0.4530	199	0.4914	0.4896	0.4954	238	0.5007	0.5090	0.5086	277	0.5077	0.5114	0.5139
161	0.4496	0.4533	0.4557	200	0.4891	0.4920	0.4954	239	0.5077	0.5017	0.5113	278	0.5100	0.5090	0.5166
162	0.4519	0.4533	0.4557	201	0.4914	0.4920	0.4980	240	0.5007	0.5114	0.5086	279	0.5077	0.5114	0.5139
163	0.4542	0.4533	0.4583	202	0.4914	0.4920	0.4980	241	0.5054	0.5065	0.5113	280	0.5124	0.5090	0.5166
164	0.4542	0.4557	0.4583	203	0.4914	0.4944	0.4980	242	0.5054	0.5065	0.5113	281	0.5077	0.5138	0.5139
165	0.4589	0.4557	0.4609	204	0.4914	0.4969	0.4980	243	0.5054	0.5065	0.5113	282	0.5124	0.5065	0.5166
166	0.4566	0.4581	0.4636	205	0.4938	0.4944	0.4980	244	0.5054	0.5065	0.5113	283	0.5054	0.5162	0.5139
167	0.4589	0.4606	0.4636	206	0.4938	0.4944	0.4980	245	0.5054	0.5090	0.5113	284	0.5100	0.5090	0.5166
168	0.4589	0.4630	0.4662	207	0.4938	0.4969	0.5007	246	0.5077	0.5041	0.5113	285	0.5077	0.5114	0.5166
169	0.4612	0.4630	0.4662	208	0.4961	0.4944	0.5007	247	0.5031	0.5114	0.5113	286	0.5100	0.5114	0.5166
170	0.4635	0.4654	0.4689	209	0.4938	0.4969	0.5007	248	0.5054	0.5065	0.5113	287	0.5100	0.5114	0.5166
171	0.4635	0.4654	0.4689	210	0.4984	0.4944	0.5007	249	0.5054	0.5090	0.5113	288	0.5100	0.5114	0.5166
172	0.4659	0.4654	0.4715	211	0.4938	0.4993	0.5007	250	0.5054	0.5090	0.5113	289	0.5100	0.5090	0.5166
173	0.4682	0.4678	0.4715	212	0.4984	0.4969	0.5033	251	0.5077	0.5065	0.5139	290	0.5077	0.5138	0.5166
174	0.4682	0.4702	0.4742	213	0.4938	0.5017	0.5007	252	0.5054	0.5090	0.5113	291	0.5124	0.5065	0.5166
175	0.4705	0.4678	0.4742	214	0.4961	0.4993	0.5033	253	0.5100	0.5041	0.5139	292	0.5054	0.5162	0.5139
176	0.4682	0.4727	0.4742	215	0.4961	0.4993	0.5033	254	0.5054	0.5114	0.5113	293	0.5124	0.5090	0.5166
177	0.4705	0.4751	0.4768	216	0.4984	0.4993	0.5033	255	0.5100	0.5041	0.5139	294	0.5077	0.5138	0.5166
178	0.4728	0.4751	0.4768	217	0.4984	0.4993	0.5033	256	0.5031	0.5138	0.5113	295	0.5100	0.5138	0.5166
179	0.4728	0.4751	0.4795	218	0.4984	0.5017	0.5033	257	0.5077	0.5065	0.5139	296	0.5100	0.5114	0.5166
180	0.4752	0.4751	0.4795	219	0.5007	0.4969	0.5060	258	0.5054	0.5114	0.5139	297	0.5100	0.5114	0.5166
181	0.4775	0.4751	0.4821	220	0.4961	0.5041	0.5060	259	0.5077	0.5090	0.5139	298	0.5100	0.5114	0.5166
182	0.4775	0.4775	0.4821	221	0.5007	0.5017	0.5060	260	0.5077	0.5090	0.5139	299	0.5100	0.5114	0.5166
183	0.4798	0.4775	0.4848	222	0.4984	0.5017	0.5060	261	0.5077	0.5090	0.5139	300	0.5124	0.5090	0.5166
184	0.4798	0.4799	0.4848	223	0.5007	0.5017	0.5060	262	0.5100	0.5065	0.5139	301	0.5077	0.5138	0.5166
185	0.4798	0.4799	0.4848	224	0.5007	0.5017	0.5060	263	0.5054	0.5114	0.5139	302	0.5100	0.5114	0.5166
186	0.4798	0.4823	0.4848	225	0.5007	0.5041	0.5060	264	0.5124	0.5041	0.5139	303	0.5100	0.5114	0.5166
187	0.4821	0.4823	0.4874	226	0.5031	0.4993	0.5086	265	0.5054	0.5138	0.5139	304	0.5100	0.5138	0.5166
188	0.4821	0.4848	0.4874	227	0.5007	0.5041	0.5060	266	0.5100	0.5065	0.5139	305	0.5100	0.5114	0.5166
189	0.4821	0.4848	0.4874	228	0.5031	0.4993	0.5086	267							

variables retenues.

Comme on peut le voir dans le tableau (4.5), la matrice permet d'observer les relations positives, cela signifie que les trois variables augmentent et diminuent ensemble.

Un lien entre les variables se traduit par une corrélation linéaire non nulle, indépendamment du sens de la relation la valeur absolue du coefficient permet de mesurer l'intensité de la relation entre les trois variables : plus la relation est proche de 1, plus la relation est forte, plus elle s'approche de 0 moins elle n'a de signification. En regardant les coefficients de notre matrice, on déduit l'existence d'une forte liaison entre les trois variables, donc un nombre de variables (facteurs) réduit suffit pour présenter l'ensemble de l'information.

	niveau1	niveau2	niveau3
niveau1	1,00000	0,99989	0,99913
niveau2	0,99989	1,00000	0,99960
niveau3	0,99913	0,99960	1,00000

TABLE 4.5 – Matrice de corrélation

4.4.4 Les vecteurs et les valeurs propres

La diagonalisation de la matrice de corrélation donne les vecteurs et les valeurs propres qui représentent respectivement les axes principaux d'inertie maximum et les variances expliquées par les facteurs indépendants les uns les autres, donnés par le tableau (4.6).

Du fait que les facteurs sont hiérarchisés et prennent des parts décroissantes de la variance, sachant que la variance est utilisée pour prendre en compte la dispersion d'une variable quantitative, les premiers axes concentrent généralement l'essentiel de l'information. Pour évaluer la qualité de représentation donnée par chaque axe, on mesure la dispersion du nuage sur l'axe concernée autour de barycentre donné par le pourcentage des variances (taux d'inertie) par (2.44), on peut lire dans la deuxième colonne du tableau (4.7) que le premier axe correspond à la plus grande valeur propre (2,999) concentre à lui seul presque la totalité de l'ensemble de l'information (99,97%) puisque l'inertie totale du nuage qui correspond à la trace de la matrice de corrélation égale à 3. Puisque le second et le troisième axe regroupent chacun respectivement (0,03%) et (0,001%), on remarque que même pas (1%) de l'inertie totale n'est expliquée par le deuxième et le troisième axe, alors l'information contenue sur ces deux axes, peut être considérée comme résiduelle .

Les vecteurs propres			Les valeurs propres		
0.5773	0.6161	0.5359	2.9991	0	0
0.5774	0.1560	-0.8014	0	0.0009	0
0.5773	-0.7721	0.2657	0	0	0.0000

TABLE 4.6 – Les vecteurs et les valeurs propres

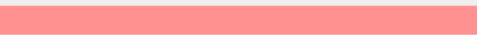
Axis	Eigen value	Difference	Proportion (%)	Histogram	Cumulative (%)
1	2,999078	2,998176	99,97 %		99,97 %
2	0,000902	0,000883	0,03 %		100,00 %
3	0,000019	-	0,00 %		100,00 %
Tot.	3,000000	-	-	-	-

TABLE 4.7 – Valeurs propres des axes et leurs contributions

Le vecteur propre correspond à la plus grande valeur représente la combinaison des signaux les plus significatifs dans l'ensemble des données et explique la majeure partie de l'information.

4.4.5 Cordonnées factorielles des individus

Les cordonnées des individus sur les nouveaux axes sont obtenus par une projection orthogonale des individus sur c'est axes.

Les cordonnées fournies par MATLAB conformément à l'équation (2.49) sont illustrés sur le tableau (4.8) et représentées par la figure (4.3).

1	-7.4536	-.4134	-.0418	40	-2.0149	.0240	.0010	79	-.1156	.0045	-.0050	118	.5431	-.0022	.0003	157	.7701	-.0028	.0038
2	-7.2457	-.2426	-.0326	41	-1.9368	-.0226	-.0034	80	-.0887	-.0055	-.0005	119	.5515	-.0016	.0035	158	.7729	-.0031	-.0033
3	-7.0356	-.1383	.0291	42	-1.8616	.0214	-.0007	81	-.0631	.0041	-.0004	120	.5614	-.0003	-.0042	159	.7758	-.0026	.0019
4	-6.8293	-.0699	.0190	43	-1.7864	.0217	-.0039	82	-.0375	.0027	-.0004	121	.5686	-.0015	.0068	160	.7801	-.0039	-.0013
5	-6.6289	-.0253	.0121	44	-1.7154	.0208	-.0012	83	-.0120	.0048	.0002	122	.5785	-.0023	-.0073	161	.7843	-.0042	-.0013
6	-6.4275	.0030	.0057	45	-1.6459	.0208	-.0024	84	.0136	.0034	.0003	123	.5856	-.0010	.0010	162	.7857	-.0027	-.0001
7	-6.2342	.0234	.0040	46	-1.5778	.0206	-.0017	85	.0349	.0032	.0035	124	.5955	-.0011	-.0009	163	.7886	-.0033	.0019
8	-6.0433	.0350	.0019	47	-1.5124	.0185	-.0003	86	.0590	.0018	-.0035	125	.6040	-.0016	-.0009	164	.7900	-.0030	-.0001
9	-5.8593	.0416	.0000	48	-1.4485	.0173	-.0028	87	.0817	.0031	.0010	126	.6112	-.0024	.0010	165	.7942	-.0021	.0031
10	-5.6765	.0468	-.0032	49	-1.3848	.0194	.0043	88	.1044	.0023	-.0009	127	.6182	-.0009	.0004	166	.7958	-.0052	.0007
11	-5.5022	.0500	-.0038	50	-1.3236	.0168	-.0065	89	.1242	.0031	-.0015	128	.6253	-.0017	.0023	167	.7985	-.0034	.0000
12	-5.3306	.0514	-.0036	51	-1.2640	.0167	.0032	90	.1455	.0030	.0017	129	.6324	-.0012	-.0015	168	.8014	-.0051	-.0013
13	-5.1631	.0495	-.0041	52	-1.2058	.0157	-.0064	91	.1654	.0028	-.0021	130	.6381	-.0014	.0055	169	.8028	-.0037	.0000
14	-5.0028	.0486	-.0033	53	-1.1490	.0152	-.0018	92	.1853	.0023	.0031	131	.6452	-.0030	-.0047	170	.8070	-.0039	.0000
15	-4.8438	.0482	-.0037	54	-1.0951	.0141	-.0024	93	.2023	.0028	-.0027	132	.6523	-.0017	.0036	171	.8070	-.0039	.0000
16	-4.6891	.0481	-.0010	55	-1.0425	.0127	-.0010	94	.2221	.0033	.0057	133	.6580	-.0026	-.0015	172	.8099	-.0045	.0019
17	-4.5416	.0478	-.0034	56	-.9914	.0134	-.0003	95	.2392	.0016	-.0065	134	.6665	-.0031	-.0015	173	.8126	-.0027	.0012
18	-4.3954	.0472	-.0039	57	-.9404	.0130	-.0028	96	.2577	.0008	.0006	135	.6721	-.0019	-.0002	174	.8156	-.0044	.0000
19	-4.2520	.0447	-.0037	58	-.8908	.0143	.0030	97	.2747	.0022	-.0020	136	.6764	-.0021	-.0002	175	.8155	-.0033	.0032
20	-4.1158	.0441	-.0023	59	-.8410	.0114	-.0040	98	.2904	.0009	.0000	137	.6836	-.0030	.0018	176	.8170	-.0040	-.0019
21	-3.9809	.0421	-.0021	60	-.7970	.0116	.0006	99	.3074	.0000	.0000	138	.6878	-.0032	.0018	177	.8212	-.0042	-.0019
22	-3.8502	.0403	-.0019	61	-.7502	.0114	-.0019	100	.3230	.0011	-.0006	139	.6948	-.0016	.0011	178	.8226	-.0028	-.0007
23	-3.7225	.0392	-.0037	62	-.7061	.0105	-.0006	101	.3371	.0018	.0007	140	.6992	-.0029	-.0021	179	.8241	-.0048	.0000
24	-3.5991	.0393	-.0023	63	-.6636	.0116	.0001	102	.3529	-.0006	-.0005	141	.7048	-.0017	-.0008	180	.8255	-.0034	.0013
25	-3.4783	.0365	-.0034	64	-.6238	.0099	-.0018	103	.3656	.0008	.0059	142	.7090	-.0020	-.0008	181	.8255	-.0040	.0032
26	-3.3605	.0355	-.0032	65	-.5798	.0100	.0028	104	.3827	-.0008	-.0063	143	.7133	-.0022	-.0008	182	.8297	-.0036	.0013
27	-3.2470	.0346	-.0031	66	-.5415	.0092	-.0030	105	.3953	.0016	.0033	144	.7204	-.0030	.0012	183	.8326	-.0043	.0032
28	-3.1377	.0340	-.0030	67	-.5032	.0092	.0035	106	.4082	-.0012	-.0030	145	.7247	-.0033	.0012	184	.8340	-.0039	.0013
29	-3.0284	.0334	-.0028	68	-.4648	.0074	-.0054	107	.4210	-.0009	.0002	146	.7276	-.0039	.0031	185	.8340	-.0039	.0013
30	-2.9234	.0331	-.0027	69	-.4307	.0077	.0010	108	.4338	-.0016	.0002	147	.7318	-.0028	-.0027	186	.8354	-.0035	-.0006
31	-2.8212	.0320	.0013	70	-.3938	.0065	-.0027	109	.4451	-.0002	-.0005	148	.7375	-.0029	.0044	187	.8383	-.0041	.0013
32	-2.7204	.0297	-.0050	71	-.3598	.0081	-.0021	110	.4579	-.0009	-.0004	149	.7417	-.0029	-.0046	188	.8397	-.0037	-.0006
33	-2.6239	.0309	-.0017	72	-.3257	.0073	.0012	111	.4691	.0004	-.0011	150	.7445	-.0024	.0005	189	.8397	-.0037	-.0006
34	-2.5302	.0281	-.0041	73	-.2944	.0060	-.0007	112	.4805	.0004	.0041	151	.7488	-.0026	.0005	190	.8425	-.0043	.0013
35	-2.4394	.0281	-.0021	74	-.2619	.0072	.0019	113	.4919	-.0004	-.0030	152	.7517	-.0043	-.0007	191	.8439	-.0040	-.0006
36	-2.3499	.0266	-.0014	75	-.2305	.0049	-.0032	114	.5019	-.0018	.0009	153	.7573	-.0031	.0006	192	.8439	-.0029	.0026
37	-2.2633	.0258	-.0025	76	-.2008	.0064	.0064	115	.5132	-.0005	.0003	154	.7602	-.0037	.0025	193	.8453	-.0036	-.0026
38	-2.1782	.0256	.0014	77	-.1709	.0040	-.0057	116	.5231	-.0006	-.0017	155	.7629	-.0019	.0018	194	.8468	-.0046	.0013
39	-2.0958	.0239	-.0030	78	-.1427	.0076	.0032	117	.5316	.0000	.0015	156	.7659	-.0036	.0006	195	.8482	-.0042	-.0006

N°	t1	t2	t3	N°	t1	t2	t3	N°	t1	t2	t3	N°	t1	t2	t3
196	.8482	-.0042	-.0006	235	.8765	-.0028	.0020	274	.8851	-.0053	-.0044	313	.8894	-.0045	-.0012
197	.8511	-.0048	.0013	236	.8765	-.0038	-.0012	275	.8850	-.0032	.0020	314	.8880	-.0049	.0007
198	.8511	-.0048	.0013	237	.8780	-.0037	.0058	276	.8851	-.0043	-.0012	315	.8880	-.0049	.0007
199	.8524	-.0034	.0026	238	.8766	-.0049	-.0044	277	.8851	-.0043	-.0012	316	.8893	-.0035	.0020
200	.8525	-.0044	-.0006	239	.8780	-.0037	.0058	278	.8866	-.0053	.0027	317	.8894	-.0045	-.0012
201	.8553	-.0051	.0014	240	.8780	-.0045	-.0064	279	.8851	-.0043	-.0012	318	.8892	-.0024	.0052
202	.8553	-.0051	.0014	241	.8794	-.0044	.0007	280	.8879	-.0038	.0039	319	.8894	-.0056	-.0044
203	.8567	-.0047	-.0006	242	.8794	-.0044	.0007	281	.8865	-.0039	-.0031	320	.8893	-.0035	.0020
204	.8581	-.0043	-.0025	243	.8794	-.0044	.0007	282	.8865	-.0042	.0059	321	.8894	-.0045	-.0012
205	.8581	-.0032	.0007	244	.8794	-.0044	.0007	283	.8865	-.0050	-.0063	322	.8894	-.0045	-.0012
206	.8581	-.0032	.0007	245	.8808	-.0040	-.0012	284	.8866	-.0053	.0027	323	.8880	-.0049	.0007
207	.8610	-.0049	-.0006	246	.8794	-.0034	.0039	285	.8866	-.0063	-.0005	324	.8894	-.0045	-.0012
208	.8609	-.0039	.0026	247	.8809	-.0051	-.0044	286	.8880	-.0049	.0007	325	.8880	-.0049	.0007
209	.8610	-.0049	-.0006	248	.8794	-.0044	.0007	287	.8880	-.0049	.0007	326	.8893	-.0035	.0020
210	.8623	-.0024	.0039	249	.8808	-.0040	-.0012	288	.8880	-.0049	.0007	327	.8893	-.0035	.0020
211	.8624	-.0045	-.0025	250	.8808	-.0040	-.0012	289	.8866	-.0053	.0027	328	.8894	-.0045	-.0012
212	.8652	-.0041	.0026	251	.8823	-.0050	.0027	290	.8880	-.0059	-.0024	329	.8894	-.0045	-.0012
213	.8638	-.0042	-.0044	252	.8808	-.0040	-.0012	291	.8865	-.0042	.0059	330	.8893	-.0035	.0020
214	.8653	-.0051	-.0006	253	.8822	-.0040	.0059	292	.8865	-.0050	-.0063	331	.8894	-.0045	-.0012
215	.8653	-.0051	-.0006	254	.8822	-.0037	-.0052	293	.8879	-.0038	.0039	332	.8893	-.0035	.0020
216	.8666	-.0037	.0007	255	.8822	-.0040	.0059	294	.8880	-.0059	-.0024	333	.8894	-.0045	-.0012
217	.8666	-.0037	.0007	256	.8823	-.0047	-.0063	295	.8894	-.0045	-.0012	334	.8893	-.0035	.0020
218	.8680	-.0033	-.0013	257	.8823	-.0050	.0027	296	.8880	-.0049	.0007	335	.8908	-.0041	-.0031
219	.8681	-.0047	.0046	258	.8837	-.0057	-.0025	297	.8880	-.0049	.0007	336	.8908	-.0045	.0059
220	.8696	-.0064	-.0037	259	.8837	-.0047	.0007	298	.8880	-.0049	.0007	337	.8894	-.0056	-.0044
221	.8709	-.0040	.0007	260	.8837	-.0047	.0007	299	.8880	-.0049	.0007	338	.8894	-.0045	-.0012
222	.8695	-.0054	-.0005	261	.8837	-.0047	.0007	300	.8879	-.0038	.0039	339	.8894	-.0045	-.0012
223	.8709	-.0040	.0007	262	.8836	-.0036	.0039	301	.8880	-.0059	-.0024	340	.8894	-.0045	-.0012
224	.8709	-.0040	.0007	263	.8837	-.0057	-.0025	302	.8880	-.0049	.0007	341	.8907	-.0031	.0000
225	.8713	-.0036	-.0012	264	.8836	-.0025	.0071	303	.8880	-.0049	.0007	342	.8893	-.0035	.0020
226	.8724	-.0049	.0046	265	.8851	-.0053	-.0044	304	.8894	-.0045	-.0012	343	.8893	-.0035	.0020
227	.8723	-.0036	-.0012	266	.8836	-.0036	.0039	305	.8880	-.0049	.0007	344	.8893	-.0035	.0020
228	.8														

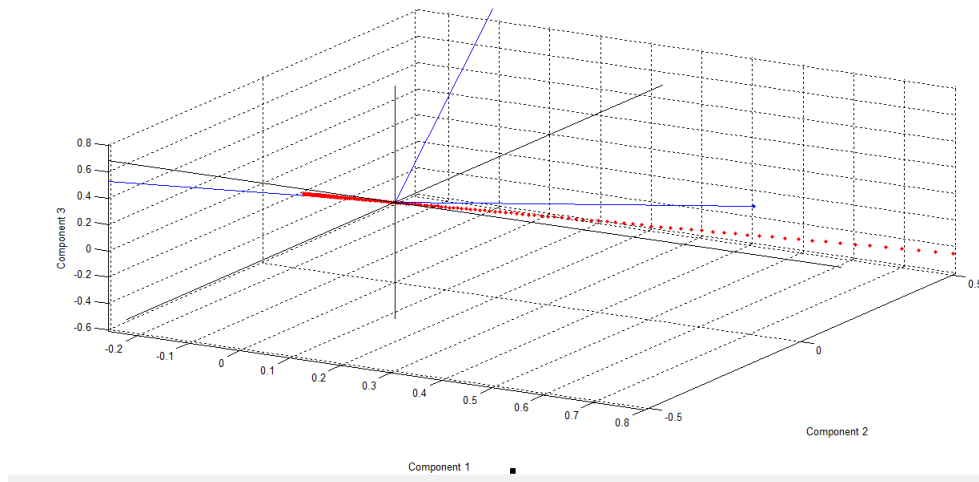


FIGURE 4.3 – Les composantes principales

4.4.6 Choix de nombre d'axe

Après avoir mise en évidence l'existence d'un lien linéaire entre les variables, l'étape suivante consiste à déterminer le nombre de composantes à retenir pour l'identification du modèle ACP, en utilisons le critère PCV (Pourcentage Cumulé de la Variance) (3.8) ainsi le critère de Kaiser (3.9), on remarque sur le tableau (4.7) et la figure (4.4) obtenue par le logiciel SPSS, que les deux critères aboutissent au même résultat qui consiste à ne retenir qu'un seul composante qui correspond au premier axe puisque lui seul présente presque la totalité de l'information, à noter que, selon le critère PCV, si la variance expliquée par le premier axe est faible, on aurait éventuellement pu prendre le deuxième et le troisième axe si ces derniers présentent une variance de tel sorte que les trois axes dépassent les 90%.

En regardant la corrélation des variables avec les composantes principales pour déterminer la signification concrète de la composante retenue, donnée par la figure (4.5), on voit sur le cercle de corrélation entre les deux premières composantes principales et les trois variables que la première composante traduit la majorité d'information, puisque presque toutes les variables sont fortement corrélées avec cette dernière et loin de l'origine donc on peut dire que la composante (t1) explique au mieux la variabilité des données originales, alors on déduit que la validité de ce choix est vérifiée.

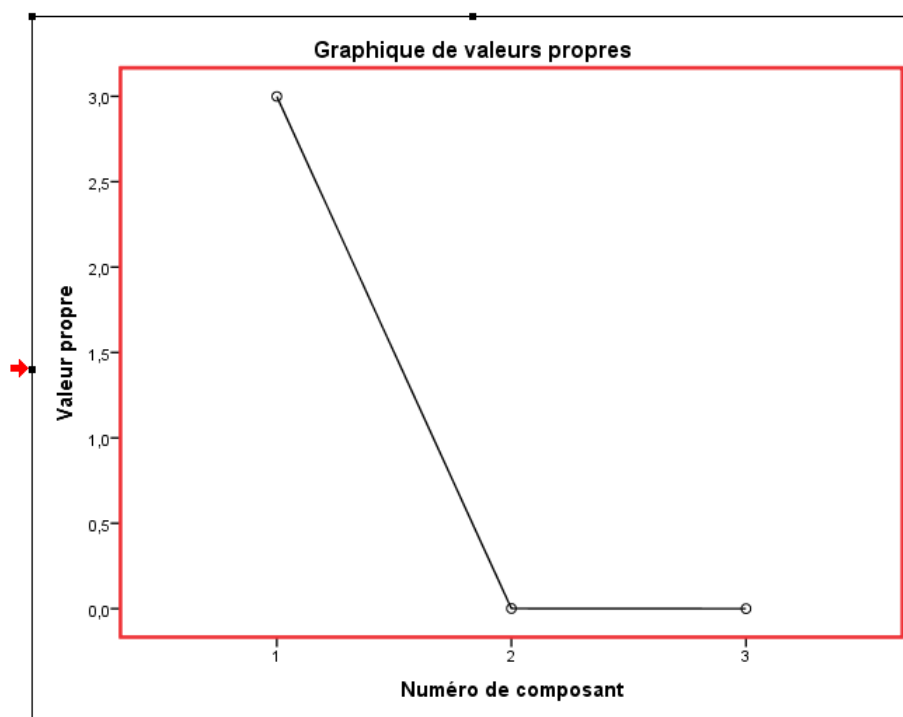


FIGURE 4.4 – Graphique de valeurs propres

4.5 Interprétation des axes

L'analyse est réalisée à l'aide des individus et variables contribuant le plus à l'axe. En effet si une variables a une forte contribution positive à l'axe, les individus ayant une forte contribution positive à l'axe sont caractérisés par une valeur élevée de la variable.

4.5.1 Étude des variables

Les variables associées aux axes factoriels représentent les combinaisons linéaires des variables initiales, les coefficients de ces relations permettent de calculer les coordonnées des individus dans le repère factoriel, de juger de leur proximité dans les différents plans et permettent aussi de calculer les corrélations entre les variables originales et les facteurs. Ses corrélations sont présentées dans le tableau suivant :

Attribute	Axis_1		Axis_2		Axis_3	
	Corr.	% (Tot. %)	Corr.	% (Tot. %)	Corr.	% (Tot. %)
niveau1	0,99983	100 % (100 %)	-0,01851	0 % (100 %)	-0,00236	0 % (100 %)
niveau2	0,99998	100 % (100 %)	-0,00469	0 % (100 %)	0,00353	0 % (100 %)
niveau3	0,99973	100 % (100 %)	0,02320	0 % (100 %)	-0,00117	0 % (100 %)
Var. Expl.	2,99908	100 % (100 %)	0,00090	0 % (100 %)	0,00002	0 % (100 %)

TABLE 4.9 – Corrélations variables-facteurs

Le facteur 1 est corrélé positivement au niveau 1, niveau 2, niveau 3. On peut aussi interpréter à partir des contributions des variables pour chaque axe retenu en regardant quels sont les variables qui participent le plus à la formation de l'axe (ce sont celles qui ont une grande coordonnée en valeur absolue). En pratique, on retient pour l'interprétation des variables dont la contribution est supérieure à la contribution moyenne. Le tableau suivant obtenu par Tanagra présente les coordonnées factorielles des variables. Les variables qui ont plus contribué à la formation de l'axe 1 sont : niveau 1, niveau 2, niveau 3.

Attribute	Mean	Std-dev	Axis_1	Axis_2	Axis_3
niveau1	0,1969670	0,0429417	0,5773385	-0,6160511	-0,5358744
niveau2	0,1791647	0,0412714	0,5774291	-0,1559625	0,8014059
niveau3	0,1507994	0,0376957	0,5772832	0,7721119	-0,2656826

TABLE 4.10 – Les coordonnées factorielles des variables

Ces résultats peuvent être visualiser graphiquement par la figure (4.5) obtenu par SPSS.

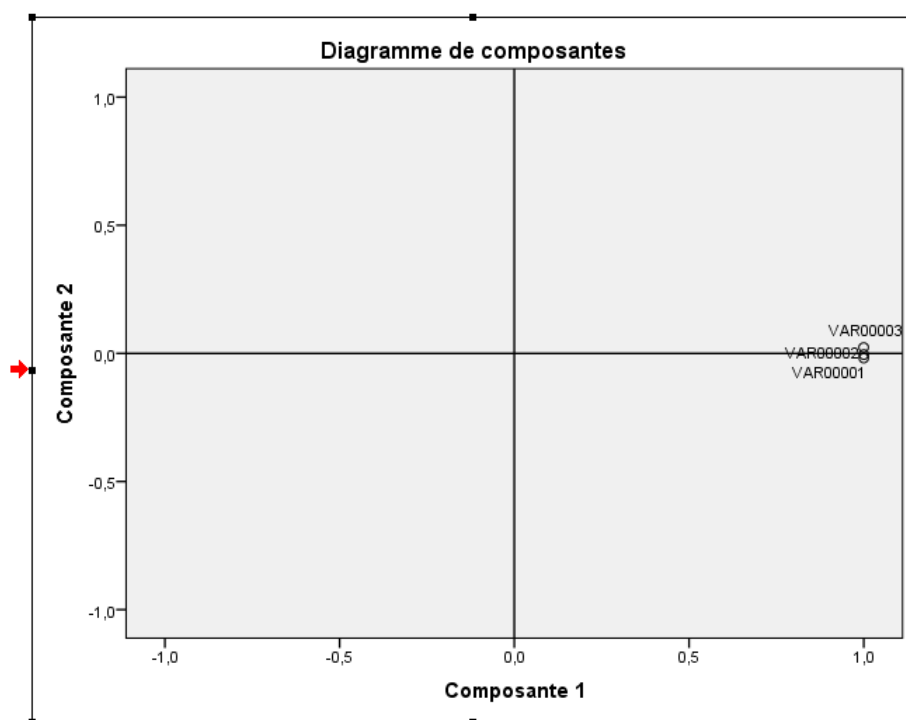


FIGURE 4.5 – Projection des variables sur le plan factoriel (1-2)

Les 3 variables sont très proches au premier axe factoriel (corrélation significative), donc elles sont bien représentées sur cet axe et les projection des individus est fidèle à la réalité.

4.5.2 Étude des individus

Puisque l'ACP est réalisée sur un tableau comportant 351 individu, dans ce cas on ne pourra pas interpréter les positions relatives de tous les individus car le nuage est tellement dense que l'on n'y verra pas grand-chose. Toutefois, si un individu est atypique il va sortir du nuage alors on pourra l'identifier pour éventuellement le supprimer et effectuer un nouveau passage sans cet individu. la projection des individus sur le plan factoriel axe1-axe 2 permet d'extraire la figure (4.6).

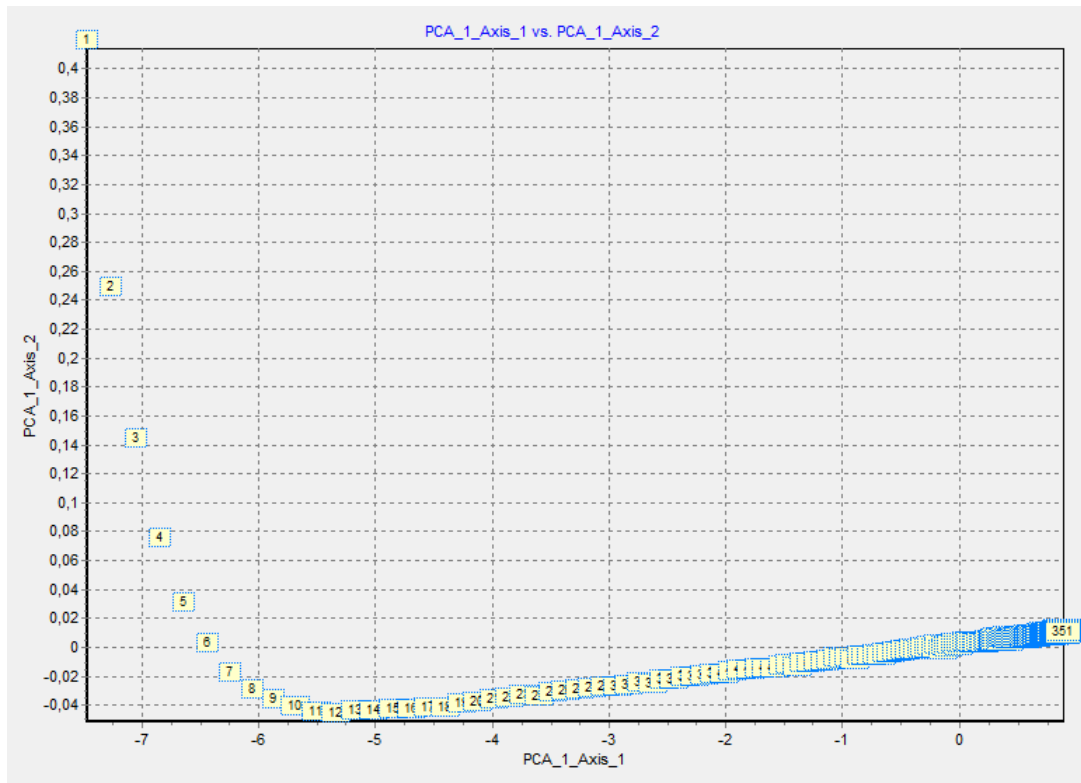


FIGURE 4.6 – Projection des individus sur les axes factoriels 1-2

Nous remarquons, dans la figure ci-dessus puisque le premier axe a un poids (valeur propre) égale à 2,99 et que le deuxième axe est pratiquement nulle, donc les données sont pratiquement alignées et les 3 variables qui décrivent le système sont nettement corrélées. Cela dit, les niveaux dans les 3 bacs varient dans le même sens simultanément, ce qui traduit la dynamique du système. Pour étudier la ressemblance entre individus, on prend par exemple le cas de la première observation et la dernière, et le cas des 5 dernières observations. On remarque dans le cas 1 que les deux points sont opposés, c'est à dire que les 3 niveaux de l'observation 1 sont opposés aux niveaux de l'observation 351, cela est vrai puisque l'observation 1 correspond aux mesures des niveaux à l'instant $t=0$ (vides) et l'observation 351 correspond aux mesures des niveaux en régime permanent.

de même pour le cas 2, on trouve que les 5 observations sont toutes presque dans la même position, cela correspond au régime permanent c'est à dire les même niveaux dans chaque bac pour les 5 observations.

Donc, on constate que la qualité de représentation des individus sur l'axe 1 est assez bonne, cette qualité est approuvée par le calcul du \cos^2 entre les différentes variables et l'axe 1 représenté par la figure (4.7) suivante :

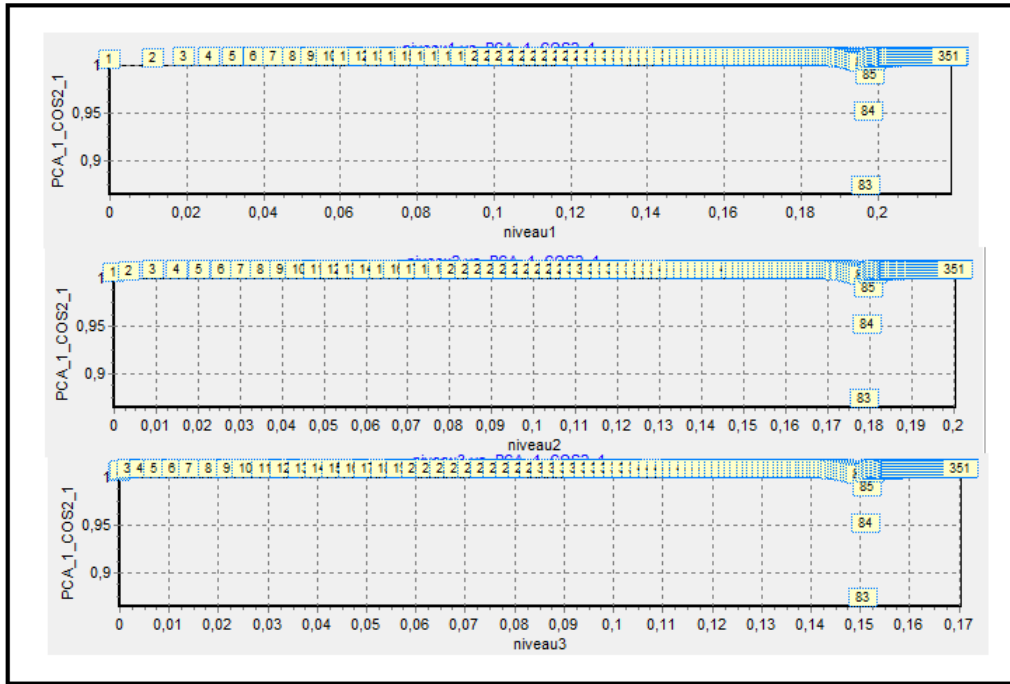


FIGURE 4.7 – mesures des qualités de représentation des variables sur les axes factoriels

Alors les approximations sont assez bien respecté. Quant au 5 premier points seront exclus du nuage et considéré comme étant des résiduels puisque elles contribues à la formation des axes 2, 3, supposé éliminé.

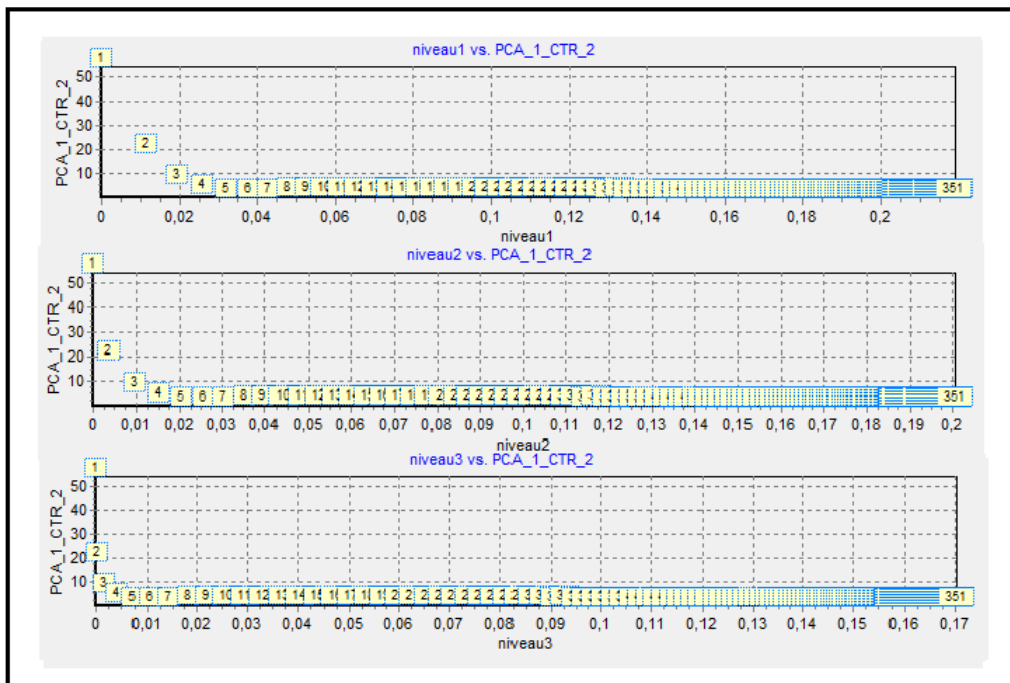


FIGURE 4.8 – Contribution des variables à l'axe 2

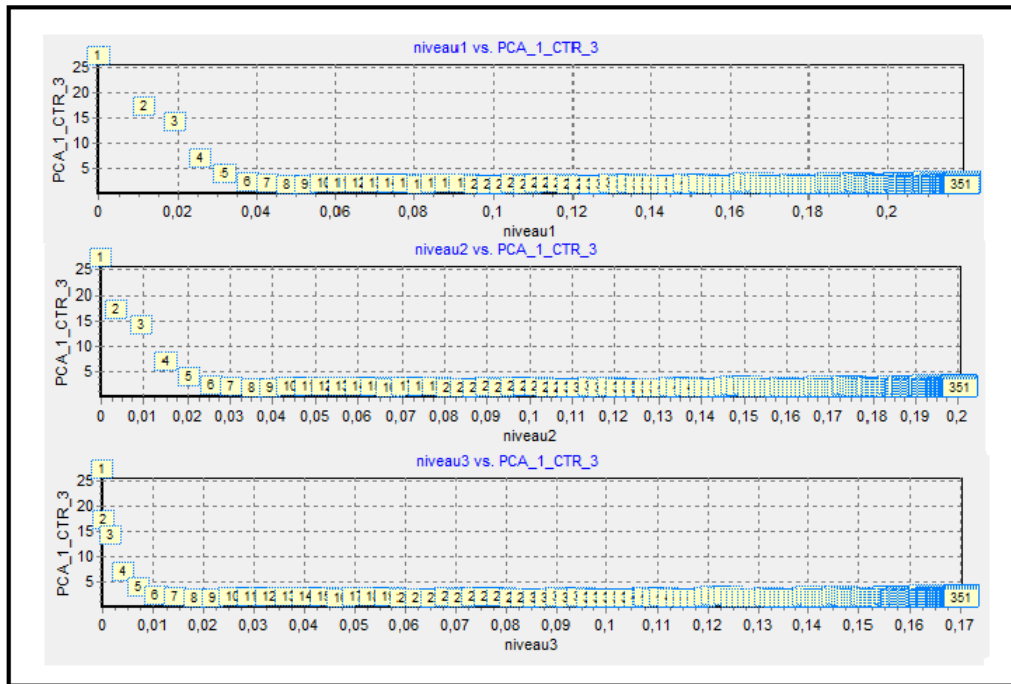


FIGURE 4.9 – Contribution des variables à l’axe 3

4.6 Modélisation par l’ACP

Après avoir défini les paramètres et la structure, le modèle ACP est obtenu à partir du premier vecteur propre par l’équation (3.6), qui permet de définir à la fois la projection permettant de réduire les données et la projection inverse pour l’estimation des données originelles. Les données réduites sont représentées par le premier vecteur du tableau (4.8) et l’approximation de la matrice des données par l’ACP se résume simplement à estimer les variables initiales par l’équation (3.5).

Le tableau (4.11) représente ces estimations calculées par MATLAB.

N°	X1	X2	X3	N°	X1	X2	X3	N°	X1	X2	X3	N°	X1	X2	X3	N°	X1	X2	X3
1	-4.3028	-4.3035	-4.3028	40	-1.1632	-1.1634	-1.1632	79	-0.0668	-0.0668	-0.0668	118	.3135	.3136	.3135	157	0.4445	0.4446	0.4445
2	-4.1828	-4.1835	-4.1828	41	-1.1181	-1.1183	-1.1181	80	-0.0512	-0.0512	-0.0512	119	.3184	.3185	.3184	158	0.4462	0.4463	0.4462
3	-4.0616	-4.0623	-4.0616	42	-1.0746	-1.0748	-1.0746	81	-0.0364	-0.0364	-0.0364	120	.3241	.3242	.3241	159	0.4478	0.4479	0.4478
4	-3.9425	-3.9432	-3.9425	43	-1.0312	-1.0314	-1.0312	82	-0.0216	-0.0216	-0.0216	121	.3282	.3283	.3282	160	0.4503	0.4504	0.4503
5	-3.8268	-3.8274	-3.8268	44	-9903	-9904	-9903	83	-0.0070	-0.0070	-0.0070	122	.3340	.3341	.3340	161	0.4528	0.4529	0.4528
6	-3.7105	-3.7111	-3.7105	45	-9502	-9503	-9502	84	.0078	.0078	.0078	123	.3381	.3381	.3381	162	0.4536	0.4537	0.4536
7	-3.5989	-3.5995	-3.5989	46	-9108	-9110	-9108	85	.0201	.0201	.0201	124	.3438	.3438	.3438	163	0.4552	0.4553	0.4552
8	-3.4887	-3.4893	-3.4887	47	-8731	-8732	-8731	86	.0341	.0341	.0341	125	.3487	.3488	.3487	164	0.4560	0.4561	0.4560
9	-3.3824	-3.3830	-3.3824	48	-8362	-8363	-8362	87	.0471	.0471	.0471	126	.3528	.3529	.3528	165	0.4584	0.4585	0.4584
10	-3.2770	-3.2775	-3.2770	49	-7994	-7995	-7994	88	.0603	.0603	.0603	127	.3569	.3569	.3569	166	0.4594	0.4595	0.4594
11	-3.1763	-3.1768	-3.1763	50	-7641	-7642	-7641	89	.0717	.0717	.0717	128	.3610	.3610	.3610	167	0.4610	0.4611	0.4610
12	-3.0773	-3.0778	-3.0773	51	-7297	-7298	-7297	90	.0840	.0840	.0840	129	.3651	.3651	.3651	168	0.4626	0.4627	0.4626
13	-2.9806	-2.9811	-2.9806	52	-6961	-6962	-6961	91	.0955	.0955	.0955	130	.3684	.3684	.3684	169	0.4634	0.4635	0.4634
14	-2.8880	-2.8885	-2.8880	53	-6633	-6634	-6633	92	.1070	.1070	.1070	131	.3725	.3725	.3725	170	0.4659	0.4660	0.4659
15	-2.7962	-2.7967	-2.7962	54	-6322	-6323	-6322	93	.1168	.1168	.1168	132	.3766	.3766	.3766	171	0.4659	0.4660	0.4659
16	-2.7070	-2.7074	-2.7070	55	-6018	-6019	-6018	94	.1283	.1283	.1283	133	.3799	.3799	.3799	172	0.4675	0.4676	0.4675
17	-2.6218	-2.6222	-2.6218	56	-5724	-5725	-5724	95	.1381	.1381	.1381	134	.3848	.3849	.3848	173	0.4691	0.4692	0.4691
18	-2.5374	-2.5379	-2.5374	57	-5428	-5429	-5428	96	.1488	.1488	.1488	135	.3880	.3881	.3880	174	0.4708	0.4709	0.4708
19	-2.4546	-2.4550	-2.4546	58	-5142	-5143	-5142	97	.1585	.1586	.1585	136	.3905	.3906	.3905	175	0.4708	0.4709	0.4708
20	-2.3760	-2.3764	-2.3760	59	-4855	-4856	-4855	98	.1676	.1676	.1676	137	.3946	.3947	.3946	176	0.4716	0.4717	0.4716
21	-2.2961	-2.2965	-2.2961	60	-4601	-4602	-4601	99	.1775	.1775	.1775	138	.3971	.3971	.3971	177	0.4741	0.4742	0.4741
22	-2.2227	-2.2231	-2.2227	61	-4330	-4331	-4330	100	.1864	.1864	.1864	139	.4011	.4012	.4011	178	0.4748	0.4749	0.4748
23	-2.1489	-2.1493	-2.1489	62	-4076	-4077	-4076	101	.1946	.1946	.1946	140	.4036	.4037	.4036	179	0.4757	0.4758	0.4757
24	-2.0777	-2.0781	-2.0777	63	-3831	-3831	-3831	102	.2037	.2037	.2037	141	.4069	.4069	.4069	180	0.4765	0.4766	0.4765
25	-2.0080	-2.0083	-2.0080	64	-3601	-3602	-3601	103	.2110	.2110	.2110	142	.4093	.4094	.4093	181	0.4782	0.4783	0.4782
26	-1.9399	-1.9403	-1.9399	65	-3347	-3348	-3347	104	.2209	.2209	.2209	143	.4118	.4119	.4118	182	0.4790	0.4791	0.4790
27	-1.8744	-1.8747	-1.8744	66	-3126	-3126	-3126	105	.2282	.2282	.2282	144	.4159	.4160	.4159	183	0.4806	0.4807	0.4806
28	-1.8113	-1.8116	-1.8113	67	-2905	-2905	-2905	106	.2357	.2357	.2357	145	.4184	.4184	.4184	184	0.4814	0.4815	0.4814
29	-1.7482	-1.7485	-1.7482	68	-2683	-2684	-2683	107	.2430	.2430	.2430	146	.4200	.4201	.4200	185	0.4814	0.4815	0.4814
30	-1.6876	-1.6879	-1.6876	69	-2487	-2487	-2487	108	.2504	.2504	.2504	147	.4225	.4225	.4225	186	0.4822	0.4823	0.4822
31	-1.6286	-1.6289	-1.6286	70	-2273	-2274	-2273	109	.2569	.2570	.2569	148	.4258	.4258	.4258	187	0.4839	0.4840	0.4839
32	-1.5704	-1.5707	-1.5704	71	-2077	-2078	-2077	110	.2643	.2643	.2643	149	.4282	.4283	.4282	188	0.4847	0.4848	0.4847
33	-1.5148	-1.5150	-1.5148	72	-1880	-1881	-1880	111	.2708	.2708	.2708	150	.4298	.4299	.4298	189	0.4847	0.4848	0.4847
34	-1.4606	-1.4608	-1.4606	73	-1700	-1700	-1700	112	.2774	.2774	.2774	151	.4323	.4324	.4323	190	0.4864	0.4865	0.4864
35	-1.4082	-1.4085	-1.4082	74	-1512	-1512	-1512	113	.2840	.2840	.2840	152	.4340	.4341	.4340	191	0.4872	0.4873	0.4872
36	-1.3566	-1.3568	-1.3566	75	-1331	-1331	-1331	114	.2897	.2898	.2897	153	.4372	.4373	.4372	192	0.4872	0.4873	0.4872
37	-1.3066	-1.3068	-1.3066	76	-1159	-1159	-1159	115	.2962	.2963	.2962	154	.4388	.4389	.4388	193	0.4880	0.4881	0.4880
38	-1.2574	-1.2576	-1.2574	77	-0987	-0987	-0987	116	.3020	.3020	.3020	155	.4404	.4405	.4404	194	0.4888	0.4889	0.4888
39	-1.2099	-1.2101	-1.2099	78	-0824	-0824	-0824	117	.3069	.3069	.3069	156	.4421	.4422	.4421	195	0.4896	0.4897	0.4896

N°	X1	X2	X3	N°	X1	X2	X3	N°	X1	X2	X3	N°	X1	X2	X3
196	4996	4997	4996	235	5060	5061	5060	274	5110	5111	5110	313	5134	5135	5134
197	4913	4914	4913	236	5060	5061	5060	275	5109	5110	5109	314	5126	5127	5126
198	4913	4914	4913	237	5068	5069	5068	276	5109	5110	5109	315	5126	5127	5126
199	4921	4922	4921	238	5060	5061	5060	277	5109	5110	5109	316	5134	5135	5134
200	4921	4922	4921	239	5068	5069	5068	278	5118	5119	5118	317	5134	5135	5134
201	4937	4938	4937	240	5068	5069	5068	279	5109	5110	5109	318	5134	5135	5134
202	4937	4938	4937	241	5077	5078	5077	280	5126	5127	5126	319	5134	5135	5134
203	4945	4946	4945	242	5077	5078	5077	281	5117	5118	5117	320	5134	5135	5134
204	4954	4955	4954	243	5077	5078	5077	282	5118	5119	5118	321	5134	5135	5134
205	4953	4954	4953	244	5077	5078	5077	283	5118	5119	5118	322	5134	5135	5134
206	4953	4954	4953	245	5085	5086	5085	284	5118	5119	5118	323	5126	5127	5126
207	4971	4972	4971	246	5076	5077	5076	285	5118	5119	5118	324	5134	5135	5134
208	4970	4971	4970	247	5085	5086	5085	286	5126	5127	5126	325	5126	5127	5126
209	4971	4972	4971	248	5077	5078	5077	287	5126	5127	5126	326	5134	5135	5134
210	4978	4979	4978	249	5085	5086	5085	288	5126	5127	5126	327	5134	5135	5134
211	4979	4980	4979	250	5085	5086	5085	289	5118	5119	5118	328	5134	5135	5134
212	4995	4996	4995	251	5093	5094	5093	290	5126	5127	5126	329	5134	5135	5134
213	4987	4988	4987	252	5085	5086	5085	291	5118	5119	5118	330	5134	5135	5134
214	4995	4996	4995	253	5093	5094	5093	292	5118	5119	5118	331	5134	5135	5134
215	4995	4996	4995	254	5093	5094	5093	293	5126	5127	5126	332	5134	5135	5134
216	5003	5004	5003	255	5093	5094	5093	294	5126	5127	5126	333	5134	5135	5134
217	5003	5004	5003	256	5093	5094	5093	295	5134	5135	5134	334	5134	5135	5134
218	5011	5012	5011	257	5093	5094	5093	296	5126	5127	5126	335	5142	5143	5142
219	5011	5012	5011	258	5102	5103	5102	297	5126	5127	5126	336	5142	5143	5142
220	5020	5021	5020	259	5101	5102	5101	298	5126	5127	5126	337	5134	5135	5134
221	5027	5028	5027	260	5101	5102	5101	299	5126	5127	5126	338	5134	5135	5134
222	5020	5021	5020	261	5101	5102	5101	300	5126	5127	5126	339	5134	5135	5134
223	5027	5028	5027	262	5101	5102	5101	301	5126	5127	5126	340	5134	5135	5134
224	5027	5028	5027	263	5102	5103	5102	302	5126	5127	5126	341	5142	5143	5142
225	5035	5036	5035	264	5101	5102	5101	303	5126	5127	5126	342	5134	5135	5134
226	5036	5037	5036	265	5110	5111	5110	304	5134	5135	5134	343	5134	5135	5134
227	5035	5036	5												

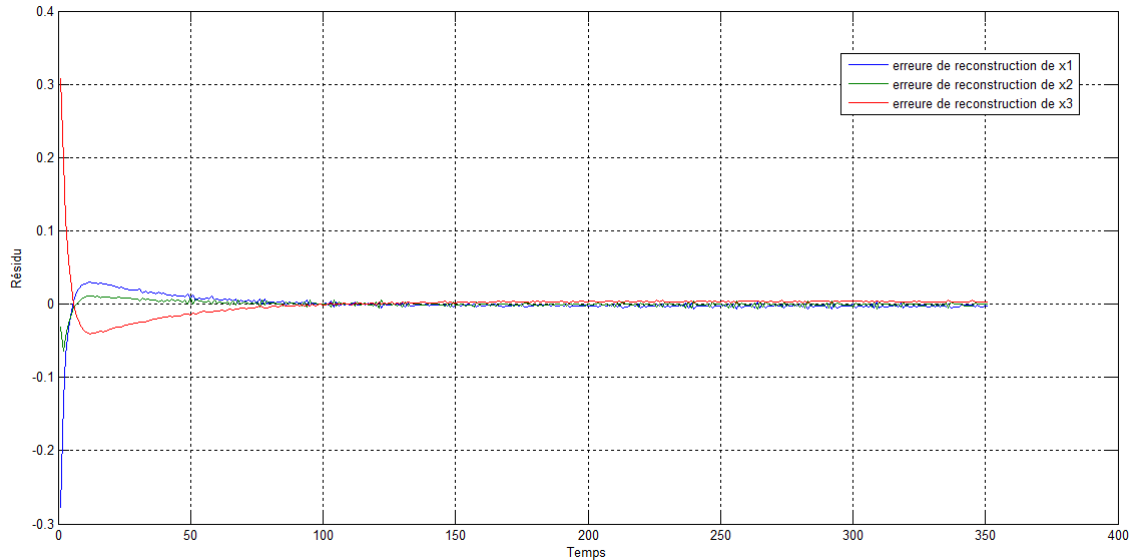


FIGURE 4.10 – L'évolution des erreurs d'estimations des mesures sans défauts

On remarque que l'écart est de faible valeur, ce qui implique que les estimations des mesures donné par le modèle ACP sont assez correcte en utilisant une seule composante, donc le modèle ACP est optimale.

4.7 Détection et localisation

Pour illustrer ce qui a été dit jusqu'à présent sur la détection et la localisation de défaut par l'analyse en composantes principales linéaire, nous allons introduire 3 défauts dans le système en créant 3 perturbations (fuites) sur les différents bacs, le premier défaut (F1) est simulé aux données et précisément à la variable $X_1(t)$ entre l'instant 50 et 150, le deuxième (F2) affecte $X_2(t)$ entre intervalle [75-175], quant à la troisième (F3) qui affecte $X_3(t)$ est prise entre [150-250], afin d'observer l'impacte de défauts sur les résidus. Chaque défaut est étudié indépendamment des autres.

L'évolution des variables en absence et avec présence de défauts représenté par les figures (4.11), (4.12), (4.13).

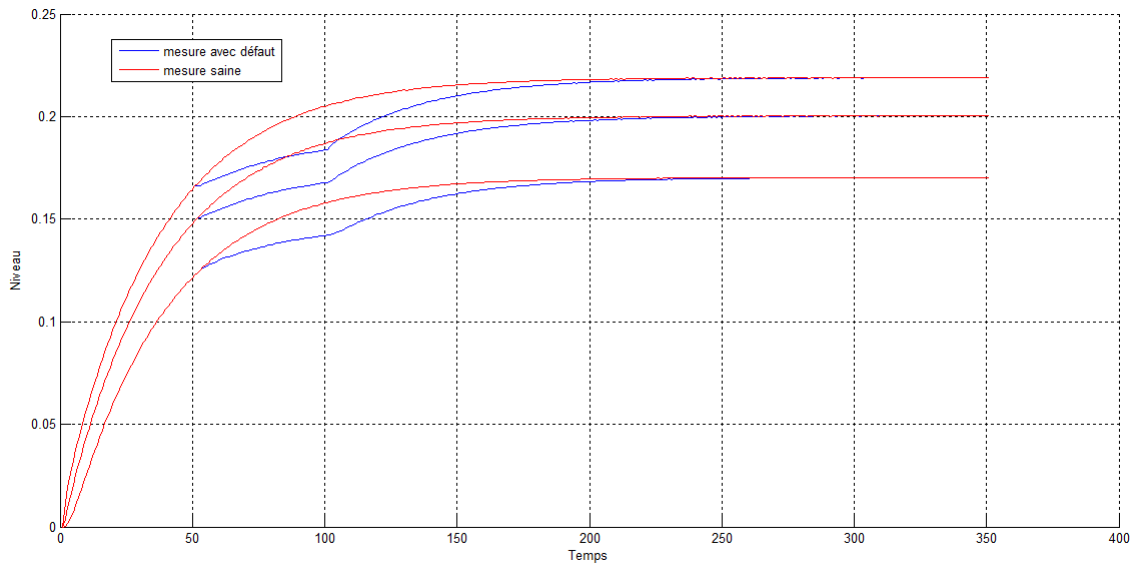


FIGURE 4.11 – L'évolution des mesures affectés par F1

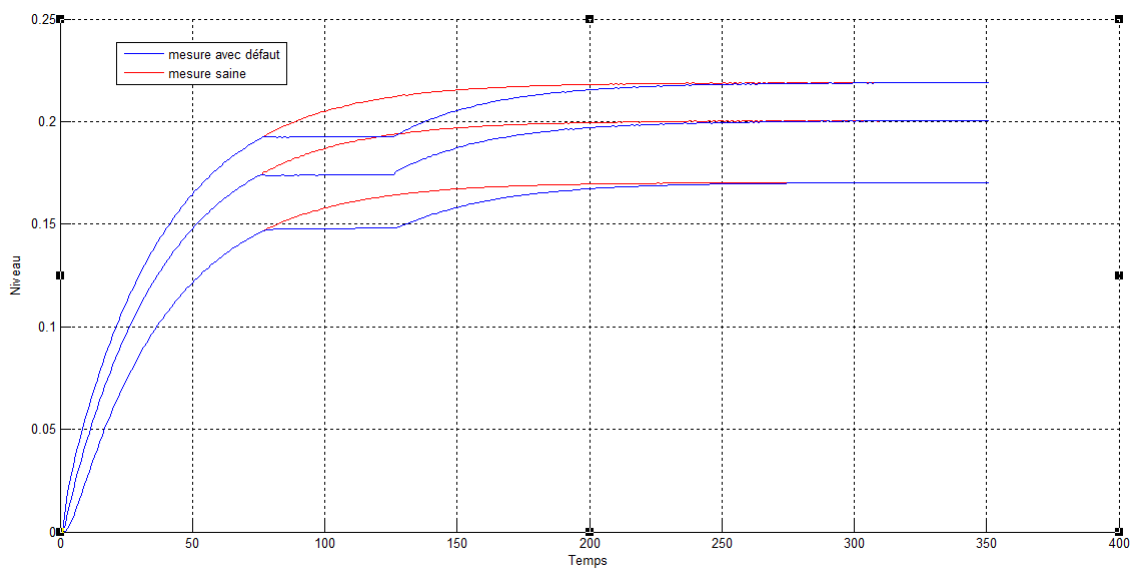


FIGURE 4.12 – L'évolution des mesures affectés par F2

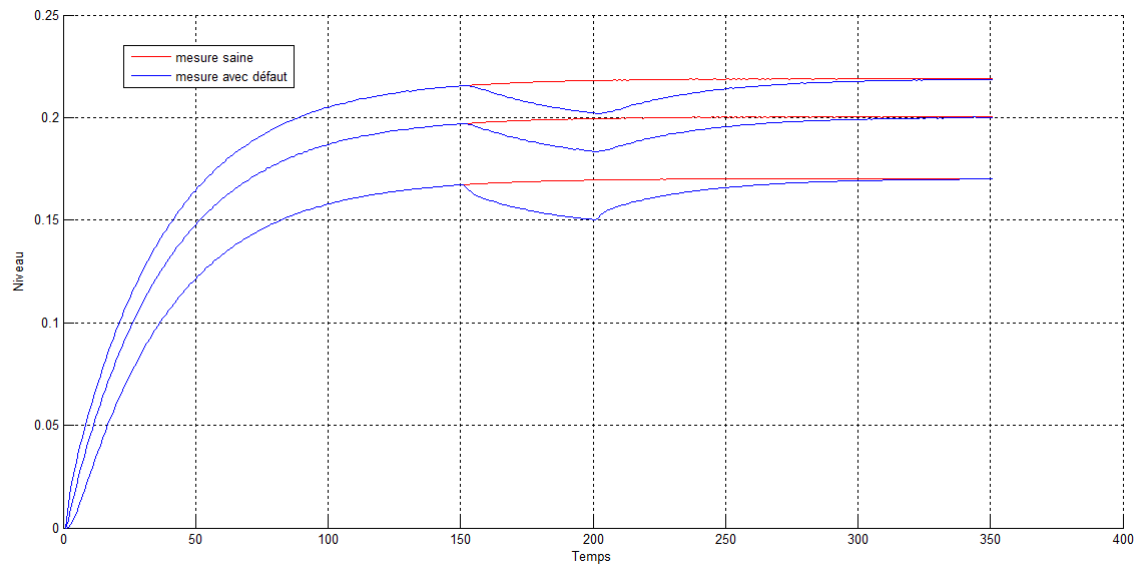


FIGURE 4.13 – L'évolution des mesures affectés par F3

L'effet de propagation des défauts sur les résidus est illustrer par les figures (4.12), (4.13), (4.14).

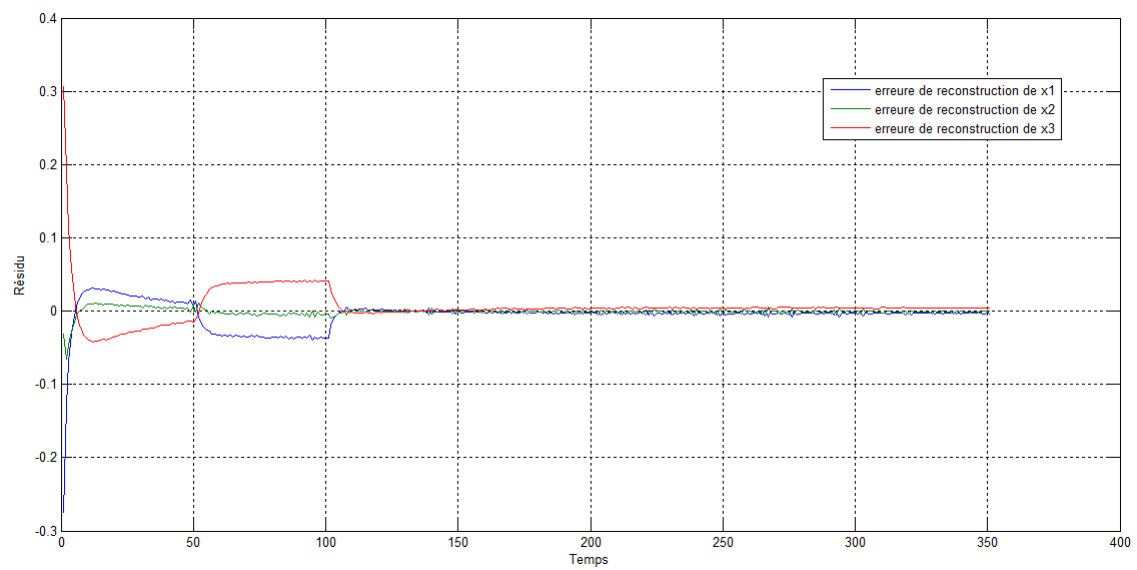


TABLE 4.12 – L'effet de F1 sur les résidus

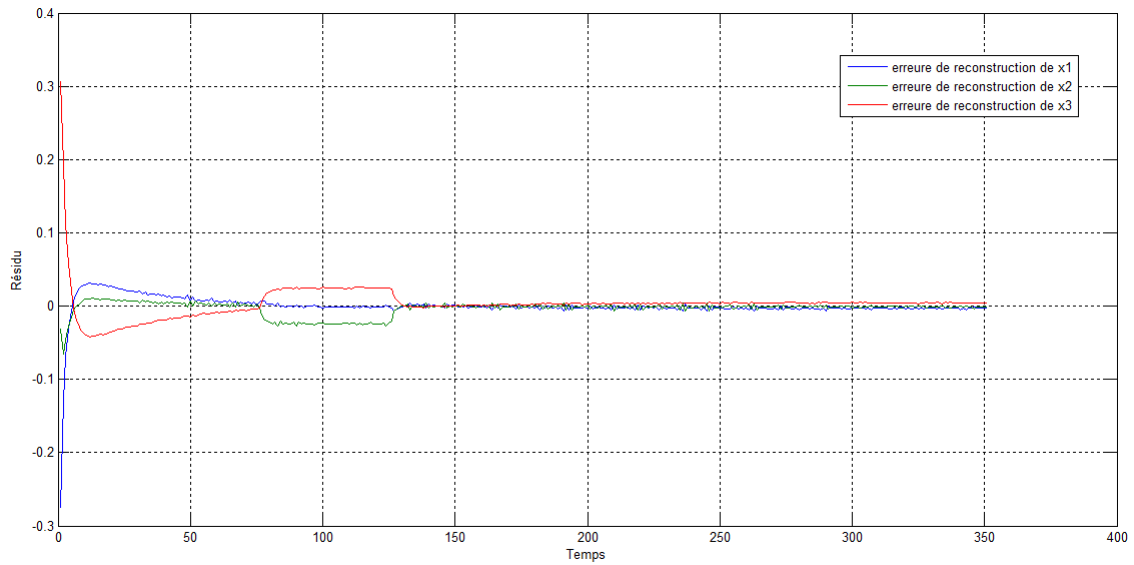


TABLE 4.13 – L’effet de F2 sur les résidus

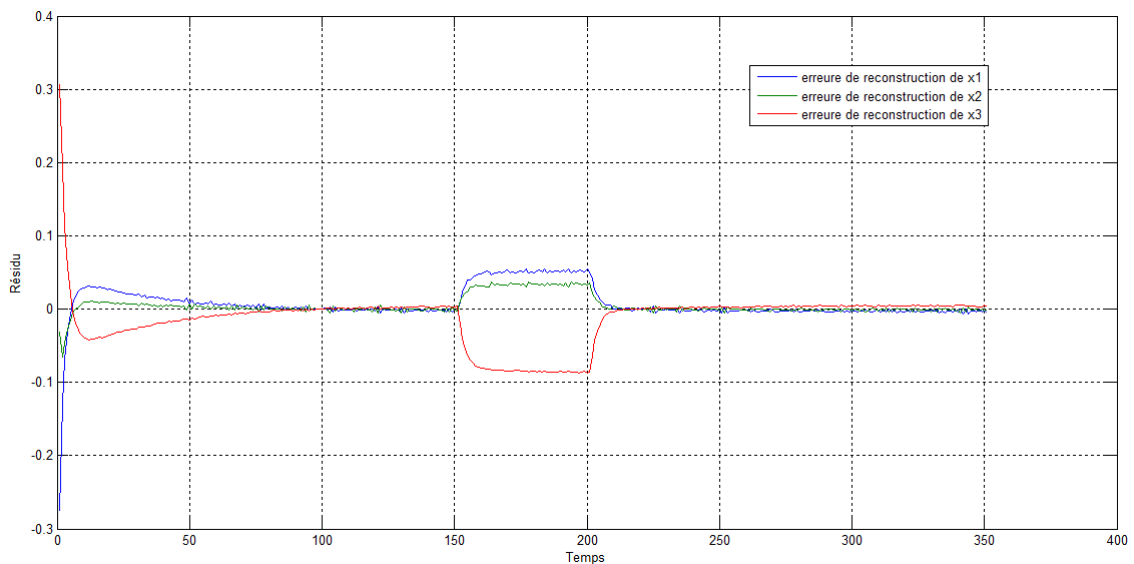


TABLE 4.14 – L’effet de F3 sur les résidus

On remarque qu’il y a une augmentation dans les quantité résiduel qui traduit un changement dans les corrélations des variables.

Pour détecter les défauts par l’ACP, nous calculons l’indice de détection de défauts SPE (3.22) pour différent cas sur les différents instantes de la plage de fonctionnement. La présence de défaut sera révéler par le dépassement du seuil calculer par les équations (3.24),(3.25),(3.26).

La figure (4.14) représente l’évolution de cet indice en absence de défaut, les figures (4.15), (4.16), (4.17) représente respectivement l’évolution de SPE en présence de défauts sur X1, X2, X3.

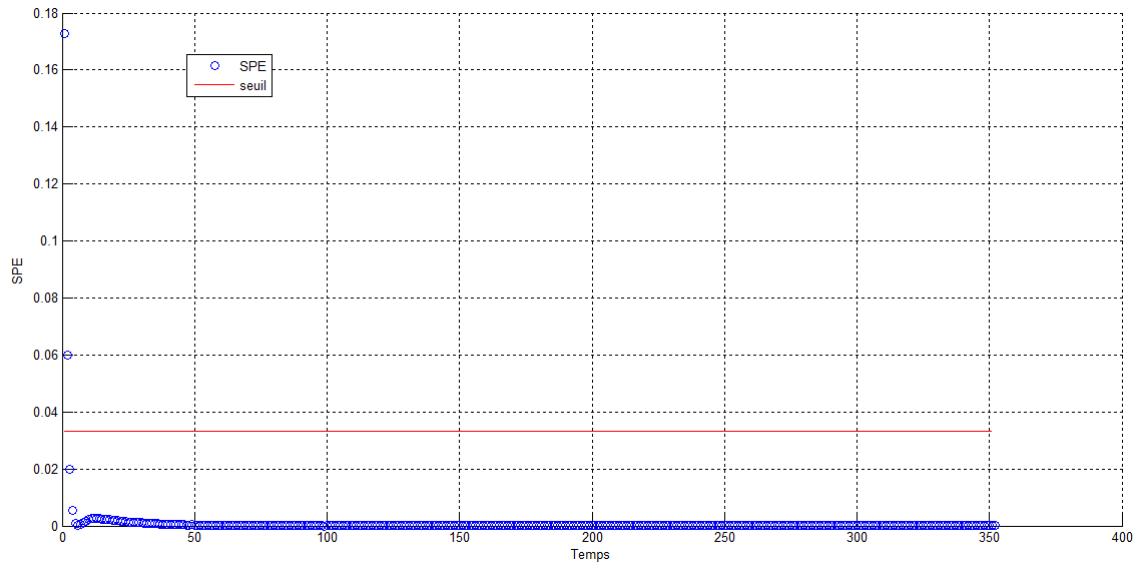


FIGURE 4.14 – L'évolution de SPE en absence de défaut

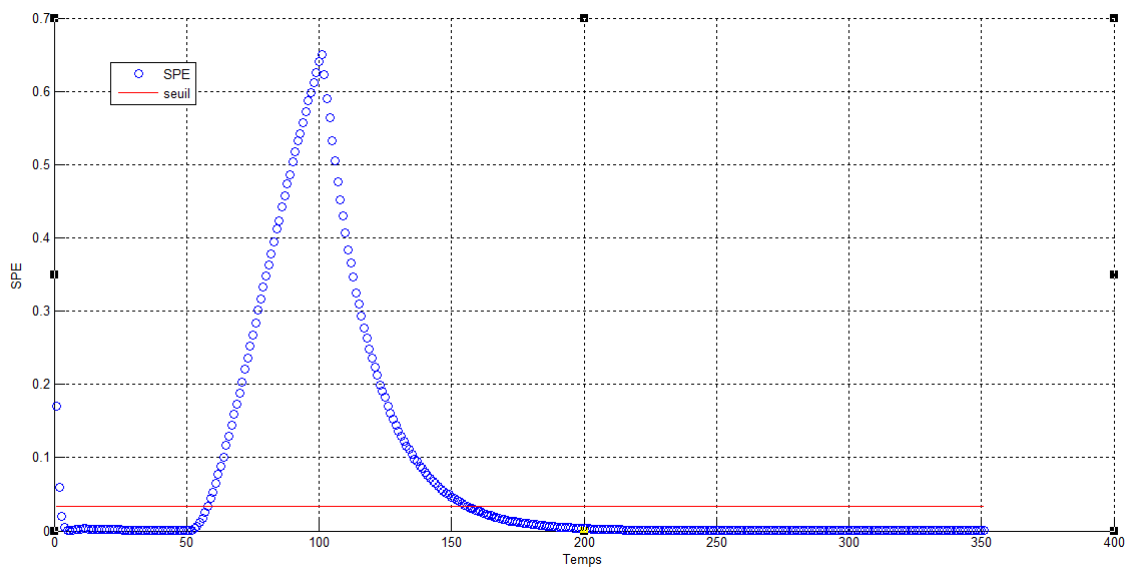


FIGURE 4.15 – L'évolution de SPE en présence de F1

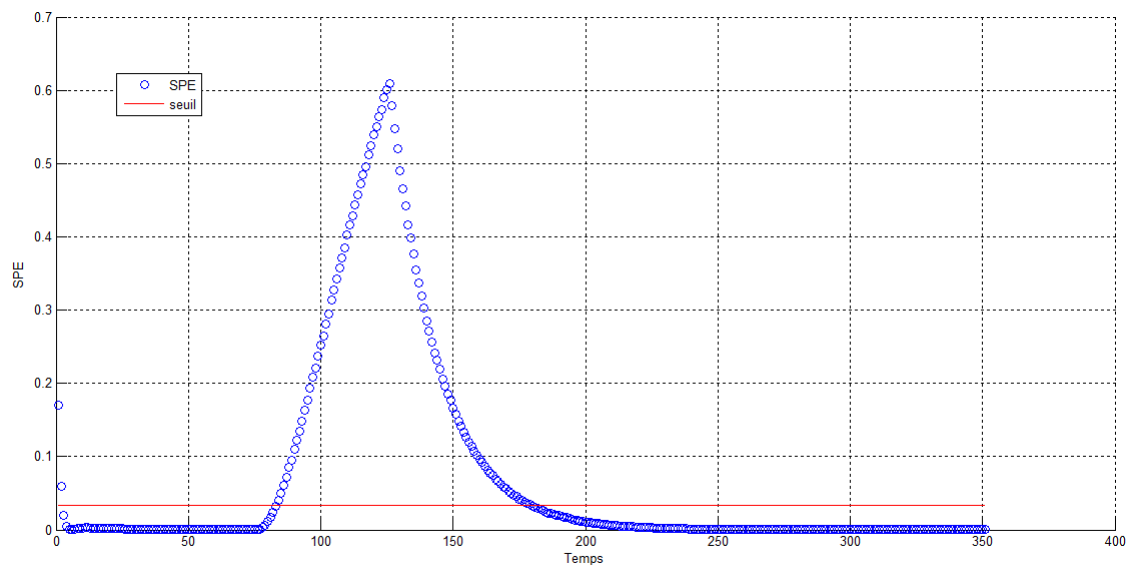


FIGURE 4.16 – L'évolution de SPE en présence de F2

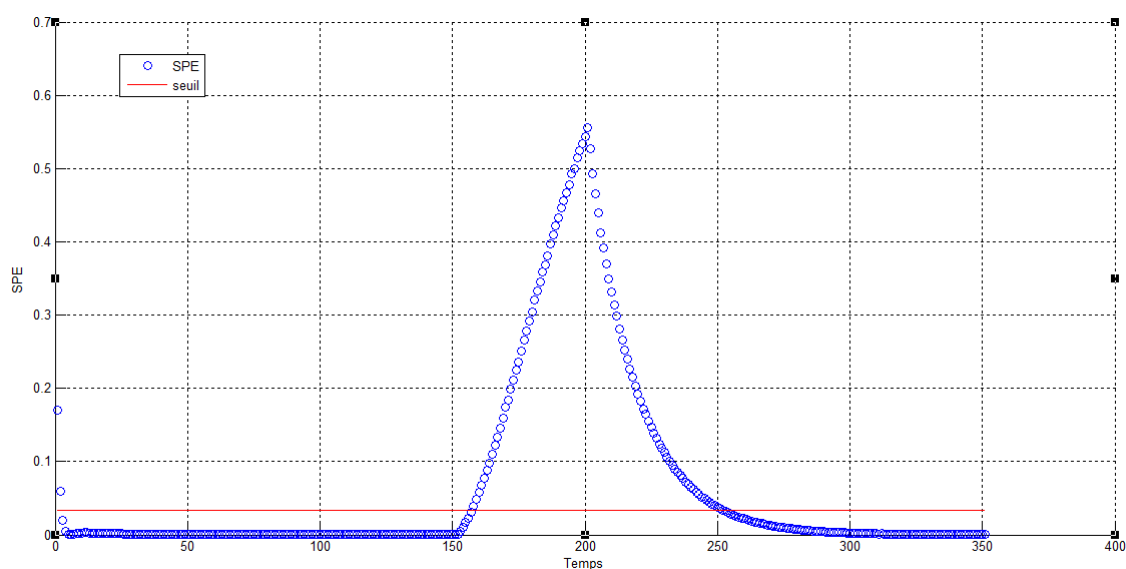


FIGURE 4.17 – L'évolution de SPE en présence de F3

On remarque des dépassements de seuil au moment d'introduction des défauts cela traduit la présence des défauts.

Une fois le défaut est détecté, il reste à localiser les variables en défauts. En se basant sur la méthode de localisation par le calcul de contribution à l'indice SPE.

La variable ayant la plus grande contribution est la variable en défaut. Les figures suivantes représentent les contributions pour les différents défauts.

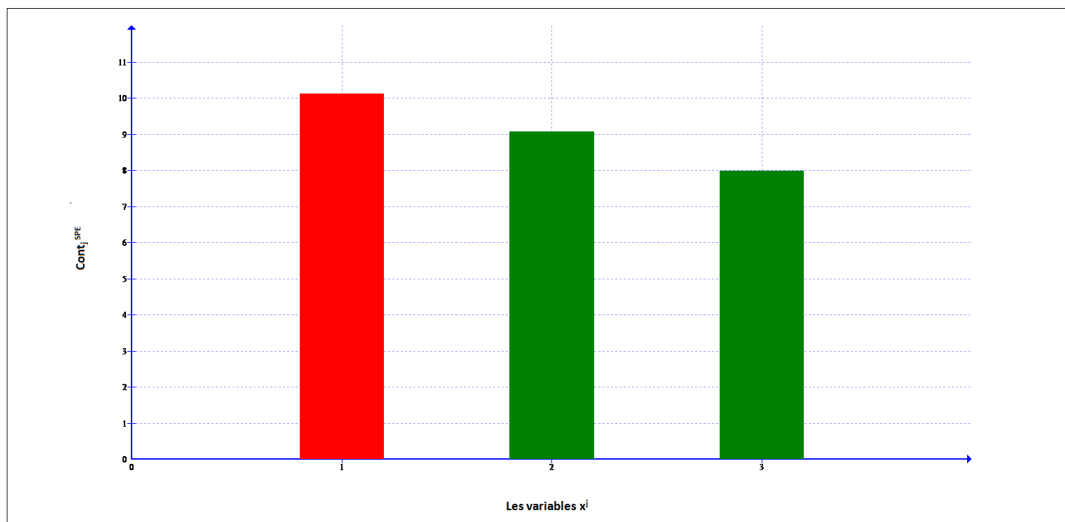


FIGURE 4.18 – Contributions à l'indice SPE en présence de F1

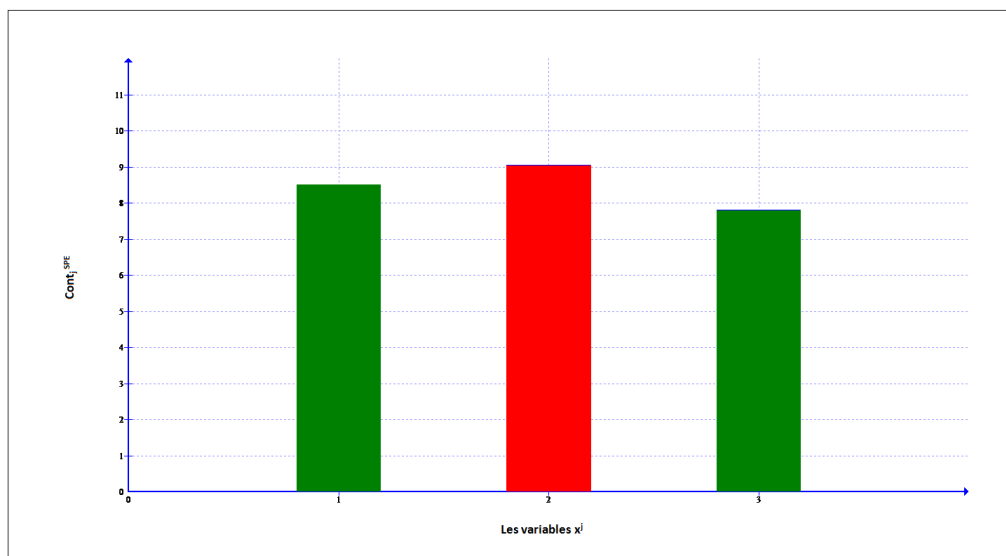


FIGURE 4.19 – Contributions à l'indice SPE en présence de F2

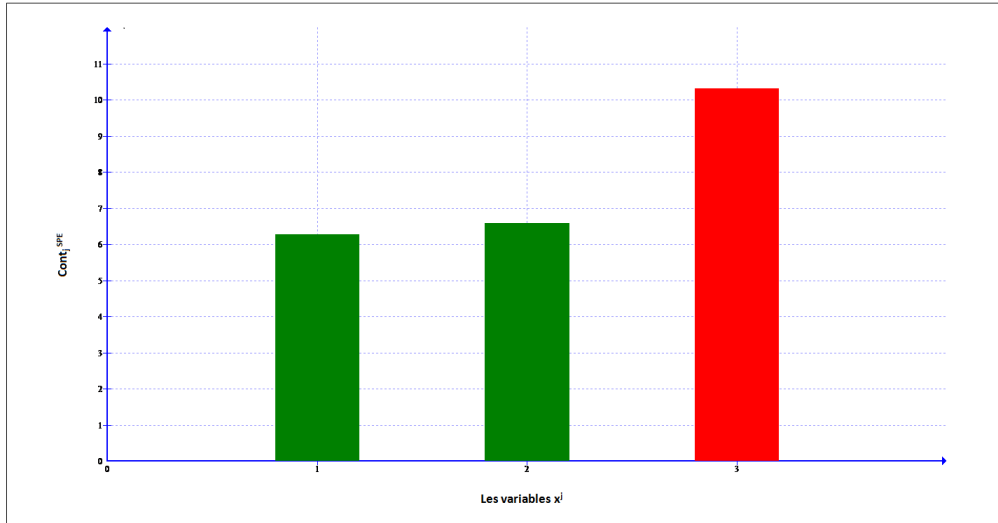


FIGURE 4.20 – Contributions à l'indice SPE en présence de F3

Les variables qui ont les contribution les plus grande correspond aux variables affecté pour chaque défaut, sont :

F1 : X1.

F2 : X2.

F3 : X3.

4.8 Conclusion

Dans ce chapitre, nous avons appliqué la méthode de l'ACP comme une technique de détection et d'identification sur un système hydraulique à trois cuves en cascade, dans laquelle nous avons simulé trois fuites d'amplitude constante. Dans notre étude nous avons arrivé à réduire le modèle ACP obtenu à un seul axe factoriel avec une perte d'informations minimal (0.031%) de l'information total qui nous a permet d'obtenir un modèle ACP de dimension (3*3) que nous avons utilisé par la suite pour surveiller et localiser les défauts présents dans le système dans le cas échéant. Afin de détecter les défauts, nous avons utilisé la statistique SPE pour fixé un seuil de détection qui indique la présence d'un défaut. Après la simulation des défauts, nous sommes arrivés à les détecter avec une fausse alarme sur une période minime pour chaque défaut, issue de la modélisation. Ensuite, nous avons localisé les défauts simulés en se basant sur le calcul de contribution SPE dans l'instant précis où le défaut est apparu. Donc L'ACP est une méthode très efficace pour le diagnostic des systèmes industriels.

Conclusion générale

Dans le domaine du diagnostic, les méthodes basées sur le concept de redondance de l'information pour les processus où il est difficile voire même impossible de leur établir un modèle mathématique complet ont été développées afin d'identifier les états de fonctionnement d'un processus. Puisque les états de défaillance sont liés directement aux variables, alors l'analyse des mesures de variables qui permet d'identifier les causes correspondantes parfaitement au but du diagnostic qui se résume à constater l'apparition de défauts puis d'en trouver les causes. Dans ce contexte, les méthodes basées sur l'analyse en composantes principales (ACP) sont très intéressantes pour la mise en évidence des corrélations linéaires significatives entre les variables du processus pour autant formuler de façon explicite un modèle d'Entrées /Sorties du système.

En premier lieu, nous avons présenté le principe fondamental de l'analyse en composantes principales. Cette dernière est utilisée comme une technique de modélisation des relations entre les différentes variables du processus. Ainsi, les paramètres du modèle représentent les vecteurs propres de la matrice de corrélation ou de covariance des données prises lors d'un fonctionnement normal du processus. Cependant, pour la détermination de la structure du modèle, il faut déterminer le nombre des composantes à retenir (nombre de vecteurs propres). Pour cette raison, plusieurs critères de choix du nombre optimal des CPs ont été présentés.

Une fois le modèle est identifié, la procédure de détection et de localisation des défauts peut être effectuée par génération des indicateurs de défauts (résidus) en comparant le comportement observé du processus donné par les variables mesurées et le comportement sensé donné par le modèle ACP. La plupart des méthodes de détection à base de l'ACP utilisent généralement la statistique SPE (erreur quadratique) et la statistique T^2 de Hotelling pour la détection des défauts. Cependant, la SPE est un test qui cumule les erreurs de modélisation présentés sur chaque résidu et la statistique de T^2 n'est pas vraiment efficace vu que les conditions de cette statistique sont rarement vérifiées car elle est calculée à partir des premières composantes principales qui ne représentent en aucun cas des résidus et donc ne représente pas un indice de détection. Concernant le problème de localisation de la (ou les) variable en défauts, une approche est largement utilisée en analyse en composantes principales, consiste à calculer les contributions individuelles des variables à l'indice de détection (T^2 ou SPE). La variable ayant la plus forte contribution

est la variable incriminée.

Dans le dernier chapitre, nous avons présenté l'application de l'ACP linéaire à la modélisation, détection et à la localisation des défauts d'un système à 3 cuves, et on a obtenu des résultats fiables après l'implémentation du modèle ACP sur ordinateur en utilisant les logiciels (SPSS, Tanagra, MATLAB). Après comparaison des résultats obtenus par les 3 programmes utilisés, on constate que, avec un nombre réduit de variables on a pu modéliser le système avec une perte d'information minimale, on a réussi à détecter et à localiser les différentes variables incriminées qui affectent le fonctionnement du système avec une précision acceptable sans formuler implicitement le modèle d'Entrées/Sorties. Donc on conclut que l'ACP est une méthode prometteuse pour le diagnostic des systèmes industriels. Ainsi nous pouvons envisager, par rapport aux résultats obtenus dans ce travail, les perspectives suivantes :

- L'une des perspectives immédiates de notre travail est d'étendre cette nouvelle formulation pour les systèmes non linéaires. Dans ce travail, nous avons considéré que la structure des modèles des deux sous-systèmes est linéaire. Et par la suite, il serait intéressant de considérer des modèles avec une structure non linéaire.
- La prise en compte des erreurs de modélisation et des bruits de mesures : Dans notre étude, nous n'avons pas considéré les problèmes tels que les incertitudes et le bruit. Ces problèmes sont très importants dans la pratique et devront être étudiés de façon précise.
- Une autre perspective aussi importante est, l'étude du cas des défauts multiples.
- Enfin, dans cette thèse nous n'avons considéré que le problème de détection et la localisation de défauts provenant des trois cuves en cascade. Dans la continuité de ce travail, nous pouvons envisager à détecter les défauts multiples et à les localiser

Bibliographie

- [1] G. Zwingelstein. (1995). diagnostic des défaillances. Paris, Hermès.
- [2] S. Gentil. Supervision des procédés complexes. Hermes Science Publications. 2007.
- [3] B. Vincent, C. Duhamel, N. Tchernev, L. Ren. Détection de defaults d'un processus industriel basée sur l'analyse en composantes principales des données. 11eme congrès international de génie industriel – CIGI2015. Quebec , Canada , 25-28 Octobre 2015.
- [4] R.Isermann. (2011). Fault Diagnosis Applications : Model Based Condition Monitoring, Actuators, Drives, Machinery, Plants, Sensors, and Fault-tolerant Systems. Springer.
- [5] J. August. Applied reliability-centred maintenance, Penwell Publishing, 2000.
- [6] B.Mnassri, E. M. El Adel, B. Ananou, and M. Ouladsine. Détection et identification de defaults par analyse en composantes principales. In 3èmes journées doctorales / Journée nationales MACS, Anger, France, 2009.
- [7] M. Ayan. (2012). Observation et diagnostic de processus industriels a modèle non linéaire. Application aux machines électriques. Thèse Doctorat, Ecole Centrale de Nantes.
- [8] www.abb.com/transformers. Système de surveillance de diagnostic pour les applications industrielles, un système qui surveille et consigne les tensions transitoires. Transformer Service Affolternstrasse 44.P.O.Box 8131, 8050. Zurich, Suisse, 2011.
- [9]] L.Fortuna., S. Graziani, A. Rizzo, et M.G. Xibilia. (2006). Soft Sensors for Monitoring and Control of Industrial Processes (Advances in Industrial Control). Springer-Verlag New York, Inc.
- [10] C. Chakour. (2016). Diagnostic et surveillance des procédés industriel set de le environnement sur la base de l'analyse de données. Thèse Doctorat, Université Badji Mokhtar, Annaba.
- [11] H. Noura. Diagnostic et tolérance aux fautes. Ecole de formation Mécatronique : Recherche et Application, 2005.
- [12] J. Breuker, et W. Van De Velde. (1994). CommonKADS Library for expertise modelling :reusable problem solving components. IOS Press

- [13] G.Acosta, C.A. Gonzalez, et B. Pulido. (2001). Basic tasks for knowledge-based supervision in process control. *Engineering Applications of Artificial Intelligence*, 14(4), 441{455}
- [14] R. Isermann, and P. Ballé. Trends in the application of model-based fault detection and diagnosis of technical processes, *Control engineering practice*, Vol.5, No.5, pp. 709- 719, 1997.
- [15] O.Adrot. Diagnostic à base de modèles incertains utilisant l'analyse par intervalles : L'approche bornante » thèse de doctorat de l'institut national polytechnique de Lorraine, 2000.
- [16]] M.Alhaj Dibo. (Validation de données et diagnostic des systèmes incertains à l'aide de l'analyse par intervalle). thèse de doctorat de l'institut national polytechnique de Lorraine, 2005.
- [17]] J. Brunet. M. Labarrère, D. Jaume, A. Rault, M.Vergé. Détection et diagnostic de pannes : Approche par modélisation. Hermès, Paris, 1990.
- [18] B. Mnassri. Analyse de données multivariées et surveillance des processus industriels par analyse en composantes principales.Thèse doctorat. Université d'Aix-Marseille.12 octobre 2012. France.
- [19]] S. Methnani. Diagnostic, reconstruction et identification des défauts capteurs et actionneurs : application aux stations d'épurations des eaux usées. Autre. Université de Toulon. 2012, Français.
- [20] R. Isermann, R. Schwarz, and S. Stolz. Fault-tolerant drive-by-wire systems. *IEEE Control Systems Magazine*. 22(5). pp. 64-81. 2002.
- [21] Y. Tharrault, G. Maurot, J. Rogot, D. Fiorelli, S. Gillé. Identification de relations de redondance analytique pour le diagnostic de fonctionnement de capteurs d'une station d'équation. Poitiers. Novembre 2006.
- [22] Q. Yang, A. Aitouche, B. Ould Bouamama. Détection de defaults par redondance analytique de l'alimentation en air d'une pile à combustible de typePEM. Conférence Internationale Francophone de l'Automatique(CIFA), Nancy, France, Juin 2010. Redondance matérielle
- [23] V. Sircoulomb, G. Hobbs, H.Chafouk, J. Ragot. Analyse et synthèse de redondance de capteurs en vue d'améliorer l'estimation d'état d'un système, 4èmes Colloque Interdisciplinaire en Instrumentation, C2I 2007, Octobre, Nancy, France.
- [24] Professeur. D. Maquin Module surveillance des processus, Institut Polytechnique de Lorraine. Université Henri Poincare-Nancy1, Janvier, 2001.
- [25] I. Fagarasan, S. Ploix, S and S. Gentil. Causal fault detection and isolation based on a set-membership approach. *Automatica*. 40 : 2099-2110, 2004.

- [26] R.V. Beard. Failure accommodation in linear systems through self reorganization. Phdthesis, Massachusetts Institute of Technology. Dept. of Aeronautics and Astronautics, 1971.
- [27] T. Denoeux. Diagnostic par reconnaissance de formes : approches non probabilistes, IN B .Dubuisson,Ed, Diagnostic ,Intelligence Artificielle et reconnaissance de formes , *Treité IC2* ,Hermès, Paris ,2001.
- [28] Y. Grandvalet, and M. H. Masson. Chapitre : Réseaux de neurones pour le diagnostic, *Treité IC2, Diagnostic : Intelligence Artificielle et Reconnaissance des Formes*, ed. : B. Dubuisson, pp. 179-218, Hermès, 2001.
- [29] I. Jolliffe. *Principal Component Analysis*. Springer-Verlag, New York, 1986.
- [30] J. Colomer, and J. Meléndez. A family of FIR differentiators based on a polynomial least squares estimation, *Proc. Of the European Control Conference*, 2802-2807, 2001.
- [31] M. Basseville. Model-based stastical signal processing and decision theoretical approach to monotoring. In *Proc. IFAC symposium on Fault Detection, Supervision and Stafety of Technical Process*, pages :1-12, Washington, USA.2003.
- [32] L. Travé-Massuyès, P. Dague, F. Guérrin. *Le raisonnement qualitatif pour les sciences de l'ingénieur*. Hermès, 1997.
- [33] N. Kiupel, and P. M. Frank, A Fuzzy FDI decision making system for the support of the human operator, *IFAC Safeprocess'97*, August 26-28, pp. 731-736, Hull, UK, 1997.
- [34] J. Lunze, *Diagnosis of quantized systems*, 4th IFAC Symposium on Fault Detection. Supervision and Safety for Technical Processes (SAFEPROCESS-2000) Budapest, Hungary, vol. 1, pp. 28-39, 2000
- [35] J. Rabot, F. Kartz, D. Maquin. Espace de parité pour les systèmes linéaires incertains Synthèse, quelques résultats nouveaux et mise en œuvre. *Seizième Colloque GRETSI*.15-19 Septembre 1997,Gronoble.France.
- [36] S. Bachir, I. Bazine Ben Ameer, T. Poinot, K. Jellassi,J. C. Trigeassou. Estimation paramétrique pour le diagnostic des processus : Application à la bobine à noyau de fer. *Journal Européen des Systèmes Automatisés (JESA)*, Lavoisier, 2008, 42 (5), pp.509-539.
- [37] H. Abdelfettah . Estimation d'état et diagnostic de systèmes à commutation par filtrage multimodèle. *Automatique / Robotique*. thèse doctorat. Institut National Polytechnique de Lorraine - INPL, France, 2006.
- [38]] Roben Rodriguez Herrera, Danielle Salles,Le Gac .initiation à l'analyse factorielle des données.Ellipes Edition Marcketing S.A.,202,32,Rue Bargue 75740, Paris Cedex15 ,France.
- [39] M. Volle. *Analyse des données*, 49, rue Héricat ,75015,Paris, France.

- [40] B. Mnassri. Diagnostic de défauts par l'analyse en composantes principales. In 6^{èmes} Journées des doctorants du LSIS, Giens, Hyères, France, 2009.
- [41] B. Mnassri, E. M. El Adel and M. Ouladsine. Une généralisation sur les conditions suffisantes de détectabilité de défauts multidimensionnels par ACP. In Conférence Méditerranéenne sur l'Ingénierie sure des Systèmes Complexes, Agadir, Maroc, 2011.
- [42] M. Tamura, et S. Tsujita. (2007). A study on the number of principal components and sensitivity of fault detection using PCA. *Computers & Chemical Engineering*, 31(9), 1035-1046.
- [43] A. Farah. (2011). Modélisation d'un parc de machines pour la surveillance. Application aux composants en centrale nucléaire. Thèse Doctorat, Université de Lorraine.
- [44] Y. Zhang, C. M. Bingham and M. Gallimore. (2003). Fault detection and diagnosis based on extensions of PCA. *Advances in Military technology*, vol. 8, N° 2,
- [45] B. A. Anissa. Détection et localisation de défauts par Analyse en Composantes Principales certaine ou incertaine de type intervalle. Thèse Doctorat, Université de Monastir.
- [46] J. R. Fanjason. (2013). Diagnostic des systèmes à énergies renouvelables de type éolien. Thèse Doctorat, Université de Corse-Pascal Paoli.
- [47] M. Baligh. (2012). Analyse de données multivariées et surveillance des processus industriels par analyse en composantes principales. Thèse Doctorat, Université d'Aix-Marseille.
- [48] R. Qin S. J.Dunia, and T. F. Edgar. (1996). Identification of faulty sensors using principal component analysis. *AIChE Journal*, vol. 42, N° . 10, p. 2797-2812.
- [49] M. F. Harkat. (2003). Détection et Localisation de Défauts par Analyse en Composantes Principales. Thèse Doctorat, Université de Lorraine.
- [50] H. T. Eastment, et W.J. Krzanowski. (1982). Cross-Validatory Choice of the Number of Components from a Principal Component Analysis. *Technometrics*, 24(1), 73-77.
- [51] P. Besse, et L. Ferré. (1993). Sur L'usage de la validation croisée en analyse en composantes principales. *Revue de Statistique Appliqué*, 41(1), 71-76.
- [52] A. Raich, and A.Cinar. (1996). Process disturbance diagnosis by statistical distance and angle measures, In *Proceedings of IFAC Congress, San Francisco*, vol. N, p.283-288.
- [53] H. H. Yue and S.J. Qin. (2001). Reconstruction-based fault identification using a combined index. *Industrial & Engineering Chemistry Research*, vol. 40, p 4403-4414.
- [54] J. E. Jackson, G. S. Mudholkar. (1979). Control procedures for residuals associated with principal component analysis. *Technometrics*, vol. 21, N° 3.