People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research

University M'Hamed BOUGARA - Boumerdes

Institute of Electrical and Electronic Engineering
Department of Power and Control

Final Year Project Report Presented in Partial Fulfilment of the Requirements for the Degree of

Master

In **Electrical and Electronic Engineering**

Option: **Control**

Title:

Interval Valued PCA-Based Approach For Fault Detection In Complex Systems.

Presented by :

-**LOUHAB Salah Eddine.**

-**LOUIFI Abdelhalim.**

Supervisor:

Dr. KOUADRI Abdelmalek.

Co-supervisor:

Mr. ROUANI Lahcene.

Registration Number:...../2019

Acknowledgements

At the end of this work, we want to express our deep gratitude and appreciation to our supervisor Dr. KOUADRI Abdelmalek, and to our co-supervisor Mr.Rouani Lahcene, for the valuable suggestions, support and guidance. We are sincerely grateful to all the staff of the Institute of Electrical and Electronics Engineering of the University of M'Hamed Bougara- Boumerdes, Teachers and workers, especially the Control Department members, for the great help and support during the last five years. A special thanks to all of our teachers, we have learned a lot from you so thank you very much.

Abstract

The aim of this study is to emphasis on the detection of process sensor faults based on Principal Component Analysis (PCA). In real life case, the uncertainties of the sensor data are influencing the system and causing some difficulties in the control decision making, which in turn evokes and increases the number of false alarms and imprecise decisions. In its standard form, PCA makes no distinction between data points and the associated measurement errors which vary depending on experimental conditions. As a result, a contemporary way of representing the influence of these uncertainties on sensors has been used, namely, a representation of data in the form of interval-valued.
Process modeling has been performed based on PCA for interval-valued data, where four of the most known methods have been tested. To limit the rate of false alarms, a threshold, with a certain confidence level, has been developed for both of the Hotelling's T^2, Q-statistics, and new ϕ statistics to detect the process's faults. To confirm the ability of the proposed approach, synthetic data has been implemented, simulated, and tested on the proposed sensor fault detection. Finally, cement rotary kiln data have been tested to validate the proposed approach in reducing false alarms and missed detection rates.

Contents

List of Figures

List of Tables

INTRODUCTION

With increasing demands for efficiency and product quality and progressing integration of automatic control systems in high-cost and safety-critical processes, fault detection and fault diagnosis have gained a very important role. However, these advances in such systems and processes have made it more difficult and challenging to detect the presence of abnormal operation before complete failures occur. Early fault detection and the quick and accurate diagnosis was acknowledged for its ability to prevent catastrophic damages, avoid defects, and improve the quality. The ancestor of fault detection and diagnosis (or, statistical process control (SPC)) dates back to the early 1920s works of Dr. Walter. A. Shewhart of the Bell Telephone Laboratories. With the advent of the computer and its increasing application in decentralized process automation systems, the 1970s was the beginning of computationally more involved and soft-based fault detection algorithms. In particular, Japan had the first successful application of SPC methods. As a result, American industry suffered extensively from the Japanese competition that has led, in turn, to renewed interest in SPC methods. Since then, many U.S. companies have begun extensive programs to develop and implement their fault detection methods in their manufacturing, engineering, and other business organizations.
Over the last two decades, Fault Detection and Diagnosis (FDD) algorithms have many applications to a wide range of industrial processes. They have been the subject of intensive research over the past two decades. Current FDD methods can be classified into model-based and data driven-based methods. The use of neural networks, artificial intelligence, and statistical methods is widely common as data-driven approaches for fault detection.
Due to the complexity of industrial processes, a large trend of application of data-driven methods can be observed from the amount of research carried in that particular field. Furthermore, within data-driven approaches, principal component analysis (PCA) and partial least squares (PLS) are the two standard methods in data-driven fault detection and diagnosis. The use of statistical methods to construct control charts usually leads to a considerable amount of false detections. This last problem arises as a natural consequence of the presence of background noise and statistical uncertainties. Many efforts have been spent in the field in order to increase the sensitivity so that small deviations can be detected, ameliorate the detection time, decrease the delay in deciding about the presence of the fault, and decreasing the rate of false alarms in the control chart.

Thereby, the performance can be improved significantly by taking into account additional information on all measurements to keep the systems in operation. In other words, more in-depth knowledge of the precision of the measurements is needed. Indeed, more complete information on measurements can be achieved by taking into account the uncertainty of

sensor measurement. A recent solution which considers such additional information can be addressed by describing a set of measurement uncertainty in terms of interval-valued data. Thus, the new interval-valued measurement encloses all the recorded sensor measurements, during a period of time between the minimum and maximum values, which correspond to the bounds of the interval. However, fault detection techniques, and more specifically those based on statistical methods like PCA, have been developed for the analysis of single-valued variables. However, they are unadapted for the interval-valued case.

First, it is worth mentioning that the PCA model is based on the eigendecomposition of the covariance data matrix. Therefore, considering the new type of interval-valued data, the determination of such a model is a quite hard task, given the absence of well-established mathematical methods in the matter (for the time being). Therefore, many researchers investigated the use of geometrical approaches, giving birth to many well approximated interval-valued PCA methods, such as vertices PCA (VPCA), centers PCA (CPCA), midpoints-radii PCA (MRPCA), and complete information PCA (CIPCA), and many others, which tend to work for narrow intervals only, and are not considered in this work.

Therefore, the aim of this work is to extend the PCA based approaches to deal with interval-valued data. In other words, we have to define new fault detection strategies for interval-valued data. For that, we investigate the use of the four most known interval-valued PCA methods CPCA, VPCA, MRPCA, and CIPCA, and extend various fault detection techniques based PCA to the interval-valued PCA case to detect the presence of faults in a complex industrial system. A simulation example and a cement rotary kiln data have been taken as an application to eliminate the existence of false alarms without deteriorating the detection time and the sensitivity to the different deviations.

Chapter 1

FAULT DETECTION AND DIAGNOSIS

1.1 Introduction

Fault detection and diagnosis is a very important problem for process engineering. It is the central component of abnormal event management (AEM) where it has attracted a lot of attention. AEM deals with time detection, diagnosis and correction of abnormal conditions of faults in a process. Early detection and diagnosis of process faults while operating can help avoid abnormal events and reduce productivity loss.
There are several ways to detect faults which could be summarized into two classes: model based and data driven. For the model based, a mathematical modal has to be extracted to describe the process, where data driven is based on the availability of the amount data for many industrial process.
In this chapter we will give an idea about faults and their classifications, to gather informations in order to be able to select and use the best fault detection and diagnosis approach.

1.2 Terminology

At the beginning it is necessary to get familiar with fault detection and diagnosis Nomenclature to simplify for us this field.

- *Fault:* A fault is an unpermitted deviation of at least one characteristic feature of the system from the acceptable, usual, standard condition. It can be also defined as a departure from an acceptable range of an observed variable or a calculated parameter associated with a process [69].

- *Failure:* A failure is defined as a permanent interruption of a system's ability to

perform a required function under specified operating conditions [25]. Types of failure can be classified into :

1. number of failures: single, multiple.
2. predictability: random failure (unpredictable, e.g. statistically independent from operation time or other failures), deterministic failure (predictable for certain conditions), systematic failure or causal failure (dependent on known conditions).

- *Malfunction:* It is defined as an intermittent irregularity in the fulfillment of a systemâ€™s desired function [25]. Both failures and malfunctions are events caused by faults. In other words, we can say that a fault appears as failure or a malfunction.
- *Disturbance:* It is an unknown and uncontrolled input acting on a system.
- *Error:* It is a deviation between a measured or computed value of an output variable and its true or theoretically correct one.
- *Residual:* A fault indicator, based on deviation between measurements and model equation based computations.
- *Symptom:* A change of an observable quantity from normal behavior [14].
- *Fault detection:* Finding if there is any fault in the system and also the time of that fault [34]. It determines whether the supervised process is working in a normal (or healthy) operating mode.
- *Fault isolation:* It is the determination of the kind, location and time of detection of fault follows fault isolation.
- *Fault identification:* It is the determination of the process variables that are responsible for the fault and thus the information can be used to diagnose the fault. It helps to focus on the subsystems where the fault occurred in a large plant [14].
- *Fault diagnosis:* Fault diagnosis amounts to determining the kind, size, location of the fault including fault isolation and identification is called as fault diagnosis.
- *Monitoring:* The continuous real task of determining the conditions of a physical system by recording information, recognizing and indication anomalies in the behavior is the process of monitoring.
- *Reliability:* Ability of a system to perform a required function under stated conditions, within a given scope, during a given period of time.

- *Safety:* Ability of a system not to cause danger to persons or equipment or the environment.
- *Availability:* Probability that a system or equipment will operate satisfactorily and effectively at any period of time [25].
- *Supervision:* It is the processes of monitoring and taking appropriate actions to maintain the operation in the case of fault [14].
- *Integrity:* The integrity of a system is the ability to detect faults in its own operation and to inform a human operator [25].

1.3 Fault Classification

There are many reasons for the appearance of faults. They stem for examples from: wrong design or assembling, missing maintenance, ageing, corrosion.... With regard to the operation phase they may be already present or they may appear suddenly with a small or large size or in steps. According to the time dependency, faults can be classified into :

1. *Stepwise faults (Abrupt):* they occur instantaneously often as a result of hardware problems (failures).
2. *Driftwise faults(Incipient):* they represent slow in time parametric changes often as a result of aging they are more difficult to detect.
3. *Intermittent faults:* they are faults that appear and disappear repeatedly for instance due to a partially damaged wiring.

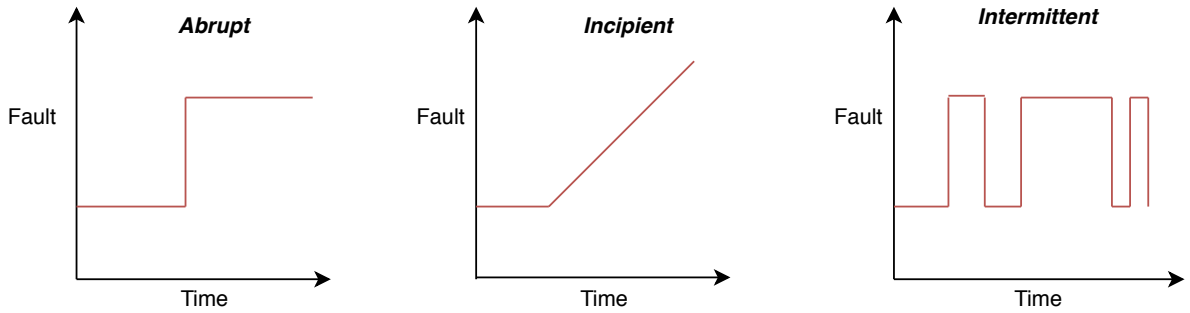


Figure 1.1: time characteristics of faults

The faults can be classified into :

1. *Additive faults*: it affects a variable by addition of the fault [2], and it can be described mathematically as :

$$Y(t) = Y_u(t) + f(t) \tag{1.1}$$

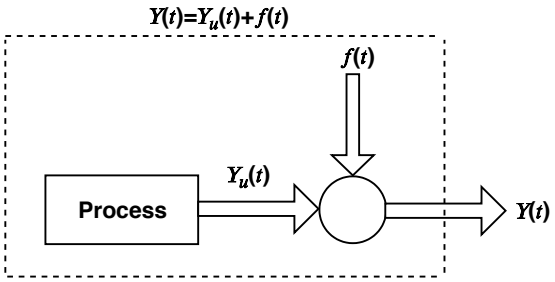


Figure 1.2: Additive fault model for an output signal

2. *Multiplicative faults*: they are changes in some plant parameters, they can change the plant outputs which depend on the magnitude of the known inputs. Those faults describe the damages and malfunction of the plant equipment. It is described mathematically as follow :

$$Y(t) = (\alpha + \Delta\alpha(t))U(t) \tag{1.2}$$

Or it can be written as :

$$Y(t) = Y_u(t) + f(t)U(t) \tag{1.3}$$

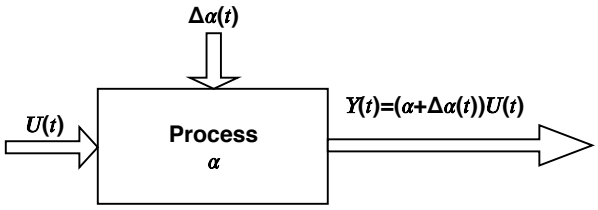


Figure 1.3: Multiplicative fault model for an output signal

Also faults that occur in process plant can be classified into three types :

1. *sensor faults*: they represent incorrect reading from the sensors this can be due to broken wires, lost contact with the surface etc. In which case the reading shown by the sensor is not related to the value of the measured physical parameter. This can for instance be a gain reduction, a biased measurement or increased noise [34] [2].

2. *Actuator faults*: they represent partial or complete loss of control action. Total actuator fault can occur for instance as a result of a breakage cut or burned wiring short cuts or the presence of outer body in the actuator. In case of partially failed actuator only part of the normal actuation is produced, it can be result of hydraulic pneumatic leakage reduced input voltage or increased resistance.

3. *Component faults*: faults that we are not able to consider them as sensor faults or actuator faults. They occur due to structural damages of the components. The dynamical behavior of the system can be changed because of these faults [25].

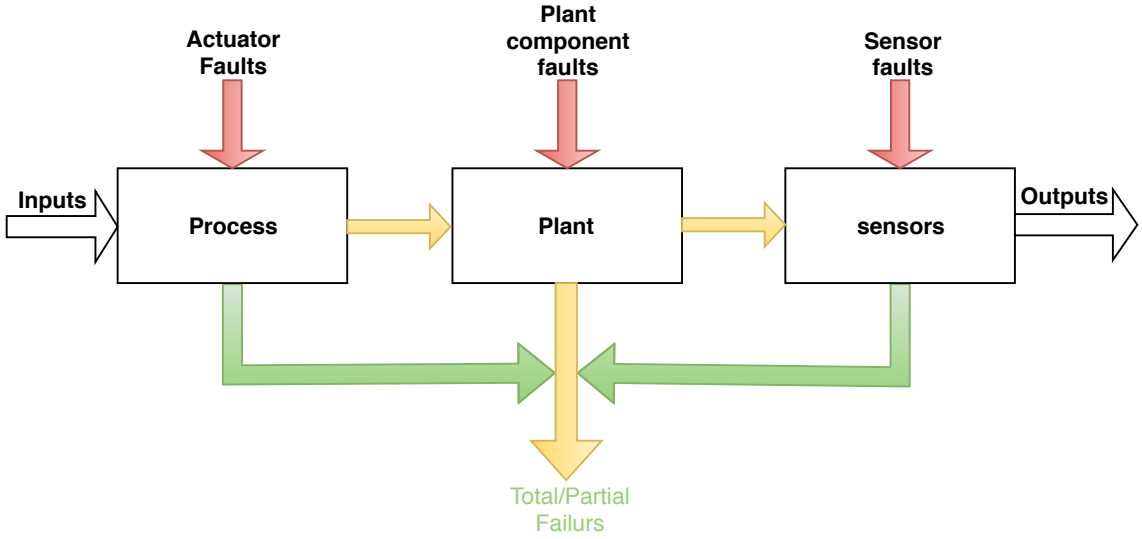


Figure 1.4: Falut Models

1.4 Desirable characteristics of a fault diagnostic system

It is useful to identify a set of desirable characteristics that a diagnostic system should possess as efficient methodology, it is used to compare the different diagnostic approaches, there are several characteristics that are considered in FDD which are :

- *Quick detection and diagnosis*: diagnosis system should respond quickly in case of fault occurance by detecting and diagnosing the fault, the FDD system should respond as fast as possible to detect a failure or a malfunction and must be quick in identifying the roots of the abnormal process state. A system that is designed to detect a fault quickly will be sensitive to high frequency influences this makes the system sensitive to noise and can lead to frequente false alarms during normal operating conditions [61].

- *Isolability*: ability of the diagnosis system to distinguish between different failures [69], but due to uncertainties it is difficult to achieve a good isolability.
- *Robustness*: one would like the diagnosis system to be robust to various noise and uncertainties, and also to resist the environmental conditions like dust or humidity.
- *Novelty identifiability*: one of the minimal requirements of a diagnosis system is to be able to deside on a given current process conditions whether the process is operating normally or abnormally (fault) and if it is abnormal whether the cause is known or unknown.
- *Adaptability*: process operating conditions can change due to environmental changes, also due to disturbances, and production quantities and qualitives, thus the fault diagnosis system should be adaptable to these changes.
- *Multiple fault identifiability*: it is difficult problem to identify multiple faults due to the interacting nature of most faults so the fault diagnosis system may be able to decouple these interactions [61].
- *Explanation facility*: an important factor to provide explanation on how the fault is originated and propagated to the current situation.
- *classification error estimate*: error measures would be useful to project confidence and reliability on the diagnostic decisions made by the system. This characteristic makes the diagnostic system more recommended.
- *Storage and computational requirements*: Quick and real time solution require algorithms and implementation which are computationally less complex, but might entail high storage requirements. Often, an FDD system with the ability to balance on these requirement is highly needed [69].
- *Modelling requirements*: for fast and easy deployment of real time diagnostic classifiers, the modelling effort should be as minimal as possible.

Before the final diagnostic decision is made, it is important also to know the different transformations that the process measurements go through.The most four essential spaces where the diagnosis process can be at any instant are :

1. *Measurement space*: the inputs of the FDD system $x = [x_1, x_2, x_3, ...x_m]$ where 'm' refers to number of variables (measurements).
2. *Feature space*: in this space variables are analyzed and combined to extract useful features about process bahavior, the space can be seen as space of points $y = [y_1, y_2, y_3, ..., y_i]$ where y_i is the i^{th} feature obtained as a function of the measurements by utilizing a priori problem knowledge.

3. *Decision space*: The mapping from the feature space to decision space is usually designated to meet some objective function (such as minimizing the misclassification). This transformation is achieved ed by either using a discriminant function or in some cases using simple threshold functions, The decision space is a space of points $d = [d_1, d_2, ..., d_k]$, where 'k' is the number of decision variables obtained by suitable transformations of the feature space.

4. *Class space*: The class space is the final interpretation of the diagnostic system delivered to the user. The transformations from decision space to class space is again performed using either threshold functions, template matching or symbolic reasoning as the case may be, the class space is a set of integers $c = [c_1, c_2, c_3..., c_M]$, where 'M' is the number of failure classes, indexing the failure classes indicating categorically to which failure class (or classes) including normal region a given measurement pattern belongs [69] [14].

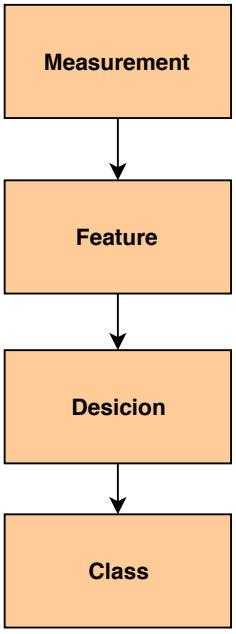


Figure 1.5: Transformations in a diagnostic system

1.5 Classification of fault detection techniques

FDD methods can be classified into model-based methods and data-driven methods. They use simulation models and measurement data. We can distinguish between them by the knowledge used to diagnose the cause of faults. The main components in a diagnosis classifier are the type of knowledge and the type of diagnostic search strategy. Diagnostic search strategy is usually a very strong function of the knowledge representation scheme which in turn is largely influenced by the kind of a priori knowledge available. Hence, the

type of a priori knowledge used is the most important distinguishing feature in diagnostic systems [69].

The basic priori knowledge that is needed for fault diagnosis is the set of failures and the relationship between the observations (symptoms) and the failures. A diagnostic system may have them explicitly, or it may be inferred from some source of domain knowledge. The priori domain knowledge may be developed from a fundamental understanding of the process using first-principles knowledge. Such knowledge is referred to as deep, causal or model-based knowledge. On the other hand, it may be gleaned from past experience with the process, this knowledge is referred to as shallow, compiled, evidential or process history-based knowledge. The model-based methods use priori-knowledge to identify the differences between model simulation results and actual operation measurement. They are divided into qualitative and quantitative modeling methods. The models are developed based on some physical knowledge related on the process which is necessary to be understood. In quantitative models this understanding is expressed in terms of mathematical functional relationships between the inputs and outputs of the system (like transfer function). The qualitative models use rule-based methods developed based on priori-knowledge. Qualitative models use the qualitative rule relationships to detect and diagnose faults instead of quantitative mathematical equations. The rules are derived from expert knowledge, process history data and quantitative models simulation data. Expert knowledge is normally summarized to a database in the form of if-then statements. Data-based models are divided into qualitative and quantitative data based. There are different ways in which data can be transformed and presented as a priori knowledge to a diagnostic system [[2] [44] [69] [38]]. The following chart summarize the different methods which are used in FDD:

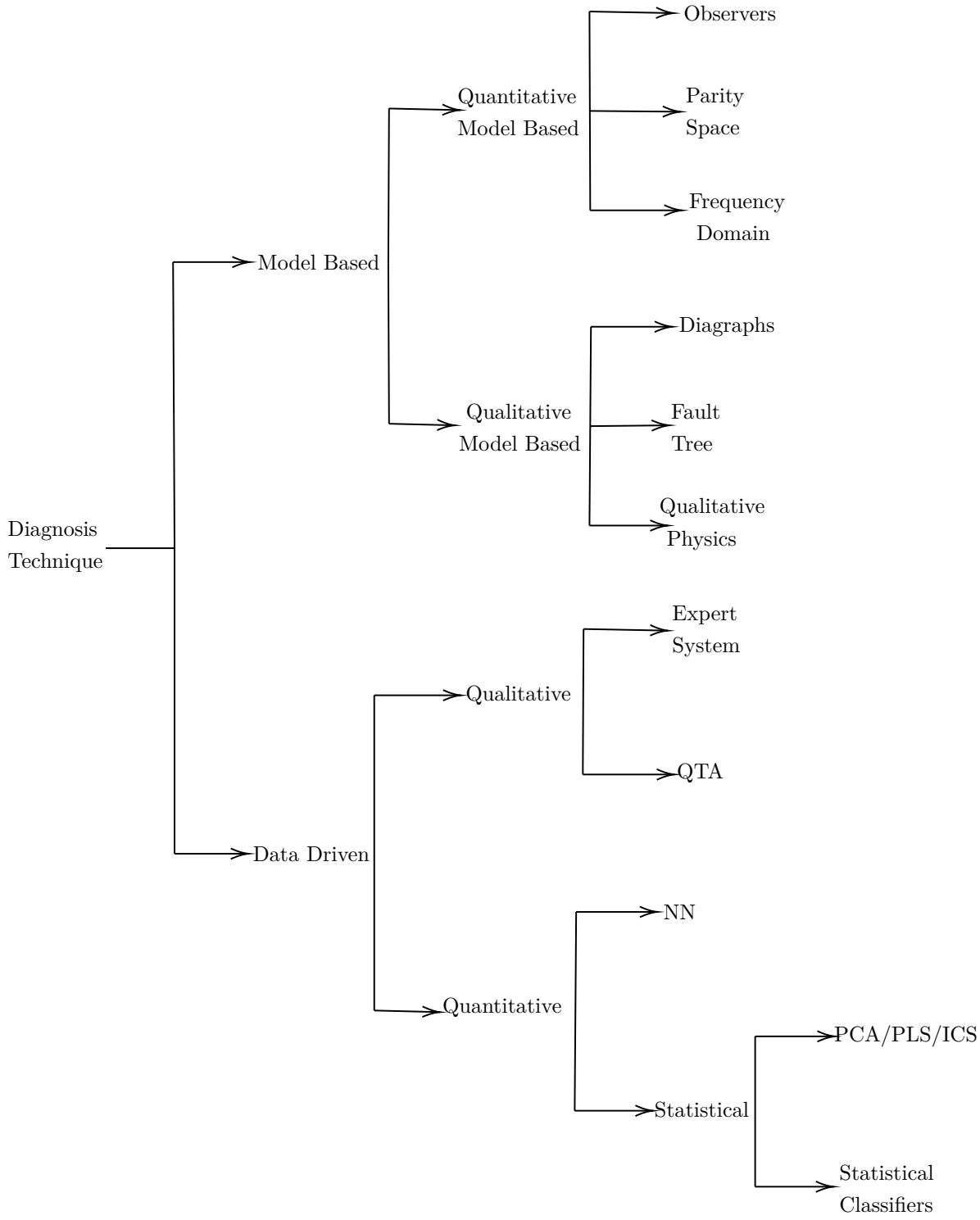


Figure 1.6: Classification of fault detection methods

1.5.1 Model based fault detection methods

Different approaches for fault detection using mathematical models have been developed in the last 20 years, model based methods are based on Analytical redundancy which is achieved from the functional dependence among the process variables and is provided

by a site of algebric equations or temporal relationships among the states and the inputs/outputs of the system, it's main objective is to check the actual system behavior against the system model for consistency, this approach assumes that the structure and the parameters of the model are precisely known [44].

Quantitative Model-based Methods

They are based on using general input-output and state space models to generate residuals and they can be classified into :

1. *Observers*: it is based on estimating the outputs of the system from the measurements [2] [44].
2. *Parity space*: this method compares the process behavior with a process model describing nominal, it's idea is to check the consistency of the mathematical equations of the system by using the actual measurements [2] [44].
3. *Frequency Domain*: Residuals are generated in the frequency domains via factorization of the transfer function of the monitored system [2] [35].

We can see that FDD using quantitative models are often used in applications because they provide the most accurate estimators of output when they are well formulated so faulty operation can be easily distinguished from normal operation without any excessive work, but it also have weakness points, which are :

- They require many inputs to describe the system.
- The effort required to develop a model is significant.
- They can be complex.

Qualitative Model-based Methods

They are based on various forms of qualitative knowledge used in fault diagnosis and they can be classified to :

1. *Qualitative physics approaches*: based on detailed knowledge of the physical relationships and characteristics of all components in a system, and based on them a set of detailed mathematical equations based on mass, momentum, and energy balances along with heat and mass transfer relations are developed and solved,qualitative physics knowledge follows two steps,first to derive qualitative equations from the differential equations termed as confluence equations,the second is the derivation of qualitative behavior from the Ordinary Differential Equations (ODEs) [2].

2. *Fault tree approach*: they are used to analyse the system's safety and reliability, it is a top-down analysis technique which describes the relationship between basic events,intermediate conditions, and top events, it is build using AND/OR gates.

3. *Diagraphs(Causal model approaches)*: a Signed Directed Graph (SDG) graphically represents a process system, a digraph is a graph with directed arcs between the nodes and SDG is a graph in which the directed arcs have a positive or negative sign attached to them. The directed arcs lead from the cause nodes to the effect nodes, it can be obtained from differential algebraic equations for the process [2].

Although these methods are easy to develop, it is difficult to ensure that all rules are always applicable, Qualitative methods provide shortcuts and may offer the best way to meet analytical needs where the other approaches needs time or cost prohibitive [35].

1.5.2 Data Driven fault detection methods

There are different ways to represent data as a priori knowledge to a diagnostic system,which is known as feature extraction (qualitative or quantitative).

Qualitative Data Driven Methods

there are two basic approaches for qualitative feature extraction which are :

1. *Expert system approaches*: an expert system is the system that solves problems in a narrow domain of expertise. The essential components in this system development are: knowledge acquisition, choice of knowledge representation, the coding of knowledge in a knowledge base, the development of inference procedures for diagnostic reasoning and the development of input/output interfaces. It's advantages in the development for diagnostic problem-solving are: ease of development, transparent reasoning, the ability to reason under uncertainty and the ability to provide explanations for the solutions provided [2] [69].

2. *Qualitative trend analysis (QTA) approaches*: they are important for monitoring the process and supervising the control, they do malfunction diagnosis and predict future states, in order to obtain a signal trend not too susceptible to momentary variations due to noise, some kind of filtering needs to be done [69].

Quantitative Data-Driven Methods

They solve the diagnostic problem as a pattern recognition problem, and it is based on classification of data points to pre-determined classes. Statistical methods use knowledge of a priori class distributions to perform classification. An example is a Bayes classifier

which uses the density functions of the respective classes. Approaches that use quantitative methods are classified to statistical like (PCA,PLS), and non-statistical like Neural Networks (NN) [2] [69].

1. *Multivariate Statistical approaches*: they are capable of compressing data and reduce it's dimensionality,they have been extensively reported in the literature.

    - *Principle component Analysis(PCA)*: the main assumptions in this method are that the data follows a Gaussian distribution and that all the samples are independent of one another, pca is based on orthogonal decomposition of the covariance matrix of the process variables along directions that explain the maximum variation data [2], more detailed informations about pca is explained in second chapter [62].
    - *Partial Least Squares (PLS)*: it is used to reduce dimension of both process variables input X and output Y.
    - *Independent Component Analysis (ICA)*: it is a computational technique for revealing hidden factors that underlie sets of random variables,measurements,or signals [19]. Independent component analysis attempts to decompose a multivariate signal into independent non-Gaussian signals. As an example, sound is usually a signal that is composed of the numerical addition, at each time t, of signals from several sources, The question then is whether it is possible to separate these contributing sources from the observed total signal. When the statistical independence assumption is correct, blind ICA separation of a mixed signal gives very good results [40], It is also used for signals that are not supposed to be generated by mixing for analysis purposes.

2. *Statistical classifier approaches*: fault diagnosis is essentially a classification problem and hence can be cast in a classical statistical pattern recognition framework [69].

3. *Neural network approaches*: considerable interest has been shown in the literature in the application of neural networks for the problem of fault diagnosis. Neural networks have been proposed for classification and function approximation problems. In general, neural networks that have been used for fault diagnosis can be classified along two dimensions: (i) the architecture of the network such as sigmoidal, radial basis and so on, (ii) the learning strategy such as supervised and unsupervised learning. Different network architectures have been used for the problem of fault diagnosis [2] [69] [19].

In this work, multivariate statistical approaches are used to represent data as a priori knowledge more informations are expressed in chapter 3 and 4.

1.6 Process Monitoring

In any system no matter how the process is good a background noise will exist, and the causes of this noise which are chance causes and assignable causes which defines the acceptable variability and the non-acceptable variability, the presence of any kind of assignable causes is exhibiting some kind of abnormal operation that have to be detected and removed effectively in minimum time.
Statistical process monitoring (SPM) is a graphical representation for the process quality characteristics that has been measured or computed from an acquired data and using certain classifier and used to determine whether the process is under statistical control (healthy state) or not [46].
SPM has found wide applications in different industrial processes, including chemicals, polymers, microelectronics manufacturing and pharmaceutical processes. The tasks involved in SPM typically include: (i) fault detection; (ii) fault identification and diagnosis; (iii) fault estimation, which assesses the fault magnitude; (iv) fault reconstruction, which estimates the fault-free values to keep control and monitoring on-going even if some faults have occurred, the Hotelling's T^2 statistic and the Q-statistic which is also known as the squared prediction error (SPE) are used for the detection of an out-of-control situation.
SPM is an endless loop that have to be always acting on any system, the following figure summarized the Process monitoring loop [58] :

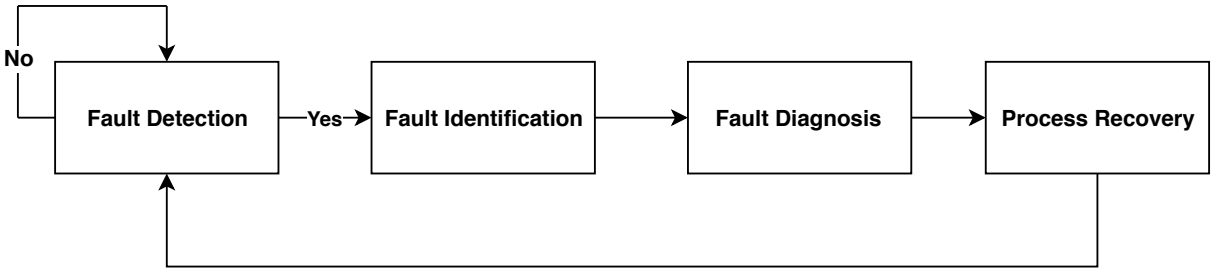


Figure 1.7: process monitoring criteria

1.7 Conclusion

In this chapter, the basic terminology of fault detection and diagnosis have been presented. Also, the different approaches for fault detection and diagnosis were explained briefly. Statistical process monitoring was investigated in order to achieve a background in the field. The next chapter will build upon the aforementioned prerequisites to introduce the basic adopted tool in this manuscript: the principal component analysis.
Further study is build according to data-driven methods (PCA) for fault detection based on it's relative simplicity acceptable efficiency.

Chapter 2

PRINCIPAL COMPONENT ANALYSIS AND IT'S APPLICATION FOR FAULT DETECTION

2.1 Introduction

Data-driven methods require the availability of sufficient data from the process. For large-scale processes, such as chemical plants, the development of model-based fault-detection or the gathering of expert knowledge on the process both require a considerable amount of efforts. For that data-driven analysis, methods offer a sufficient way, especially based on methods of multivariate statistical analysis and mostly using PCA which have received considerable amount of attention.

2.2 Linear PCA Principle

Principal Component Analysis (PCA) procedure for orthogonal linear transformation of the input data to a new coordinate system, among the common data analysis methods, principal components analysis is widely used to discover and to visualize the main structure of a multidimensional data set that are usually correlated to each other. Precisely, it aims at reducing the complexity of the multivariate data via a projection into a subspace which preserves maximum variance of the original space in minimum number of dimensions. PCA can be defined as a linear orthogonal projection of the original correlated data into a new set of uncorrelated data that explain the trend of the process.
The main objective of PCA is to transform a set of dependent variables to a set of uncorrelated variables, called Principal Components (PC's), which are order so that the first

few components contains most of the information in the original data set.
The next figure represents an illustration for PCA with two dimensional data X1 and X2. The first principal component is a line through the widest part of the ellipse and the second component is the line that goes through the less wide part of the ellipse. The best-fit line (or axis in this case) corresponds to the first principal component which goes through the wide part of points, in PCA the more correlated the original data, the better this line will explain the actual values of the observed measurements, this line will best explain all the observations with minimum residual error. In other words, the line goes in the direction of maximum variance of the projections.

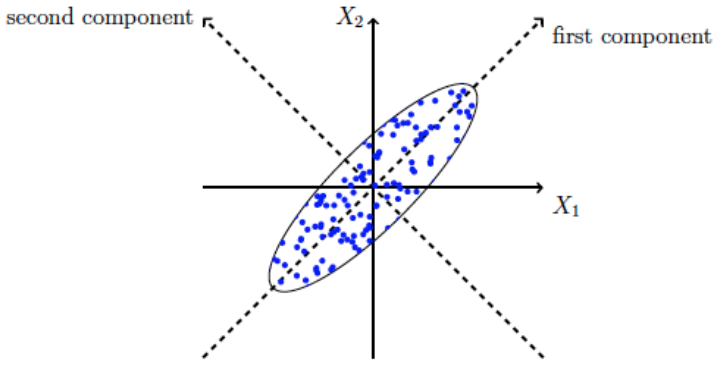


Figure 2.1: Illustration of PCA in 2-D data

2.3 Mathematical principle of PCA

Let $X = [X_1, X_2, X_3, ...X_m]$ be a set of m random variables of n observations, and let Σ be the variance-covariance matrix X has a $rank = min(n, m)$, the rank is brought down using an approximation of X by a matrix of lower rank say ℓ where $\ell < rank(X)$.
We want to define m linear observations of X that represent the information in X more parsimoniously.

$$\Sigma = \frac{1}{n-1} X^T X \tag{2.1}$$

The covariance matrix Σ is of dimension m x m, real, and symmetric. The diagonal elements are the variances of the individual random variables, while the off-diagonal elements are their covariances.
Let $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq ....\lambda_m$ be the eigenvalues of the covariance matrix Σ, and $p_1, p_2, p_3, ...p_m$ the corresponding eigenvectors. The sample covariance matrix Σ of data can then be represented by it's eigen decomposition as :

$$\Sigma = P^T \Lambda P \tag{2.2}$$

Where P are known as the principle component leading vectors with $p^T p = I$, and λ regroups eigenvalues of the covariance matrix sorted in decreasing order, they define the variance explained by each corresponding eigen vector.
Eigen-decomposition is used for PCA, because it is a technique to reduce dimension by:

- Taking linear combinations of the original variables.
- Each linear combination explains the most variance in the data it can.
- Each linear combination is uncorrelated with the others

Given that $P = [p_1, p_2, p_3, ..., p_m]$ is the eigen vector matrix of data matrix X, and $\Lambda = diag(\lambda_1, \lambda_2, ...\lambda_m)$ contains the associated eigenvalues positioned diagonally, PCA can be viewed as a linear mapping of the form [32]:

$$T = XP \tag{2.3}$$

Where $T \in \mathbb{R}^{n \times m}$ is the principal component matrix, the eigenvectors P are the coefficients for the linear transformation, the projection can be reversed back with:

$$X = TP^T \tag{2.4}$$

The principal components (PC's) are sorted by descending order of the eigenvalues. If the first ℓ PC's have most of the variation of X, then the first ℓ eigenvectors $P_\ell = [p_1, ..., p_\ell]$ as "factors" underlying the set $X_1, ..., X_m$, so the mapping is performed from $\mathbb{R}^m$ to a lower dimensional space $\mathbb{R}^\ell$.
By choosing ℓ eigenvectors of P corresponding to ℓ largest eigen values we transform the space of the measured variables into the reduced dimension space $\widehat{T}$ which allows to obtain the estimate reduced rank matrix $\widehat{X}$ and matrix $P_{m-\ell} \in \mathbb{R}^{m \times (m-\ell)}$ generated by $(m - \ell)$ columns of P allows the projection in the residual sub-space $\widetilde{T}$, so the estimation error or residuals E can be calculated, the relations can be expressed as follows :

$$\Sigma = \begin{bmatrix} P_\ell^T & P_{m-\ell}^T \end{bmatrix} \begin{bmatrix} \Lambda_\ell & 0 \\ 0 & \Lambda_{m-\ell} \end{bmatrix} \begin{bmatrix} P_\ell \\ P_{m-l} \end{bmatrix} \tag{2.5}$$

$$\widehat{T} = XP_\ell \tag{2.6}$$

$$\widehat{X} = XP_\ell P_\ell^T = XC_\ell \tag{2.7}$$

$$\widetilde{T} = XP_{m-\ell} \tag{2.8}$$

$$E = X - \widehat{X} = X(I - P_\ell P_\ell^T) = X(I - C_\ell) \tag{2.9}$$

PCA is considered as one of the most famous methods in the area of latent variable modelling. This statistical model have shown to be a very powerful tool in dealing with complex process data, especially chemical processes. Various uses of such model include process optimization and control, predictive modelling, and process monitoring.

2.4 Determining number of useful component

While conducting PCA on a set of variables, the number of extracted components is equal to the number of variables being analyzed, necessitating to decide how many of these components are truly meaningful and worthy of being kept. The expectation is that the first few components will account for meaningful information (high variance), and that the last components will tend to account for less meaningful information (low variance). Therefore, the next step of the analysis is to determine how many useful components should be retained for the PCA model. However, the most common problem in the use of PCA is specifically the determination of this number of components to retain Several comparisons have been lead in literature that focus on the effectiveness of various criteria. All the methods used to select the number of principal components gives fairly good results, and have each advantages and drawbacks, it is then advisable to carefully choose the criterion that is most suitable for the desired application. In what follows, some of the most common criteria that may be used in choosing the number of components are described.

2.4.1 Kaiser rule

It is one of the most used criteria to solve the number of components problem, this approach simply retains the components corresponding to an eigenvalue greater than 1 ($\lambda_i > 1$) [33], Thus components with eigenvalues less than 1.0 are viewed as trivial, this rule was criticized for the significance problem and many suggestions and modifications was done to deal with this problem, but The general advice is to use this method as a starting point.

2.4.2 Cumulative Percentage of Variance

PCA is all about variance of data high variance means useful information and low variance means noise, PC's are ordered according to decreasing variance, so we have to retain ones that their cumulated variance (CV) approximates the total variance of the original variables, The amount of variance preserved by l-Dimensions out of a total of m

available measurement types can be calculated based on the eigenvalues of the covariance/correlation matrix as :

$$CPV(\ell) = \frac{\sum_{i=1}^{\ell} \lambda_i}{\sum_{j=1}^{m} \lambda_j} \tag{2.10}$$

Hence, we have to choose ℓ so that the cumulative percentage of variance $CPV(\ell)$ is high. Generally values in the range 70-80% are considered satisfactory,but for more accurate models higher CPV ratio is often required (over 90%). Furthermore, while we would like to conserve as much as possible of the variance, we want to retain as few principal components as possible and to keep the noise out of interference. Therefore, the decision becomes a matter of trading-off the amount of conserved variance and the number of retained components [67].

2.4.3 Cross-Validation Criterion

This method is based on fitting the model to only part of the available data (training set), and measure the goodness of the models with different numbers of extracted components fit the other part of the data (validation set). Various choices exist but we prsesnt only the predicted residual sum of squares ($PRESS$) [74], which is given by :

$$PRESS(\ell) = \frac{1}{Nm} \sum_{k=1}^{N} \sum_{i=1}^{m} (\widehat{x}_i^{\ell}(k) - x_i(k))^2 \tag{2.11}$$

N is the size of the used validation set, we pick the first value of ℓ that minimizes $PRESS(\ell)$ and consider it as the number of principle components.

2.4.4 Parallel Analysis (PA)

This method is based on comparing the eigenvalues of the Covariance/Correlation matrix of the original data set with those of a randomly sampled data set from normally distributed population [21]. When the eigenvalues from the random data are larger then the eigenvalues from the pca or factor analysis you known that the components or factors are mostly random noise, once the associated eigenvalue is larger than the one generated from the random sample the associated component is considered to be significant.
PA is summarized in the following steps [23] :

1. calculate the eigen values of the covariance matrix: $\lambda = [\lambda_1, \lambda_2, ..., \lambda_m]$
2. generate k-sets of random data with same size as X ($Y_i \in \mathbb{R}^{n \times m}, i = 1, 2, ..., k$).
3. compute eigen values of Y_i and store them in $\Psi \in \mathbb{R}^{k \times m}$ with δ_{ij} is the j^{th} pseudo-

eigenvalue for the i^{th} experiment.

$$\Psi = \begin{bmatrix} \delta_{11} & \delta_{12} & \cdots & \delta_{1m} \\ \delta_{21} & \delta_{22} & \cdots & \delta_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{k1} & \delta_{k2} & \cdots & \delta_{km} \end{bmatrix} \tag{2.12}$$

after calculating the eigenvalues we compute the average eigenvalue as :
$\widehat{\Psi} = [\overline{\delta_1}, \overline{\delta_2}, ..., \overline{\delta_m}]$

4. Retain the components for which: $\widehat{\Psi} > \lambda_i$

2.4.5 Bootstrap method

It is similar to cumulative percentage method. Basically, this method is based on randomly drawing n samples from the data set with replacement in order to form the bootstrap sample.bootstrap is then used to train the model. But due to the random sampling, the bootstrap method exhibits a large variability.

2.4.6 Broken Stick method (BS)

This method is based on considering the variance shared by the principal axes to be embedded in a stick of a unit length, then if we broke it randomly to m segments, the k^{th} longest segment would be :

$$E(\pounds_k) = \frac{1}{m} \sum_{i=k}^{m} \frac{1}{i} \tag{2.13}$$

If we consider that the PCA model divides the variance randomly on all the m-dimensions, the fraction of the variance explained by each dimension would be the same as the relative lengths obtained by breaking the sticks randomly into m pieces [39], after putting the eigenvalues and the BS predictions in decreasing order we compare the individual eigenvalues with individual BS predictions. BS model is easy to calculate and it showed a good performance with acceptable accuracy. Some studies showed that broken stick model has a tendency to underestimate the number of retained components, and a major problem with this model is its independence from the sample size [8]. Nevertheless, BS model is easy to calculate and it showed a good performance with acceptable accuracy, at least when the studied data is highly correlated [51].

2.4.7 Minimum Average Partial (MAP)

MAP is a method based on the matrix of partial correlation [68], it involves a complete PCA followed by the examination of the matrices of partial correlations [49]. The average of squared partial correlations is used as a Goodness-of-fit measure, when the minimum average squared partial correlation is reached, the residual matrix resembles an identity matrix and no further components are extracted,the statistic is :

$$f_0 = \sum_{i=1,i\neq j}^{m} \sum_{j=1,j\neq i}^{m} \frac{\rho_{ij}^2}{m(m-1)} \tag{2.14}$$

If $f_0 > f_1$ then no components would be extracted.
The more significant components are partialled out, the systematic variance decreases, then the unsystematic variance causes the average statistic of partial correlations (ASPC) to increase again [49]. MAP shows less variability and not affected by the sample size, but it neglect non-trivial components when they have small loading.

2.5 Fault indicators

The procedure begins by establishing a principal component plan or model under normal operations, then an index is calculated to evaluate the process performance Among the commonly used indices which are the Squared Prediction Error (SPE) that calculates the perpendicular distance between a new observation and the established estimate using principal components, and the Hotelling's T^2 statistic that represents the variability in the principal component subspace. Based on the calculated index, a proper threshold can be established for fault detection, where exceeding values with respect to that threshold are considered as faulty entries.
One of the most powerful tools in process quality control is the statistical control chart first developed in the 1920's by Walter Shewhart. Multivariate control charts however were introduced in 1947 by Harold Hotelling [24]. They allow to aggregate information concerning a few process variables on one control chart using some statistic. This statistic is a measure of distance between the values of these variables while taking into account the structure of correlations between variables in the form of covariance matrix. In the case of PCA, it can serve as model established from collected data on a process, where new observation measurements in the testing set can be projected into the lower dimensional space. This new data can then be decomposed into a principal component part and a residual part. Each of these projections can be monitored separately or jointly using control charts, where several Shewhart type statistics can be used for this purpose.

2.5.1 Hotelling T^2 statistic

This statistic provides an indication of unusual variability within the principal subspace. It is calculated as the scaled squared 2-norm of first principal components vector $\hat{t}(k)$. Or the value of one sample is equal to the sum of squares of the adjusted (unit variance) projections on each of the principal components, the formula of the Hotelling's T^2 is given by :

$$T^2(k) = \hat{t}(k)^T \Lambda_\ell^{-1} \hat{t}(k) = \sum_{i=1}^{\ell} \frac{t_i^2(k)}{\lambda_i} \tag{2.15}$$

Λ_ℓ is the diagonal matrix of the ℓ largest eigenvalues $\lambda_i = [\lambda_1, ..., \lambda_\ell]$ of the covariance matrix Σ.
The process is considered normal for a given significance level α if $T^2 < \tau_\alpha^2$. The control threshold τ_α^2 can be obtained by different approaches, but it is frequent to use the following expression [65] :

$$\tau_\alpha^2 = \frac{\ell(N-1)^2}{N(N-\ell)} F_{\ell,(N-\ell),\alpha} \tag{2.16}$$

Where $F_{\ell,(N-\ell),\alpha}$ is the $(1-\alpha)\times 100$ percentile of the fisher distribution with ℓ and (N-l) degrees of freedom.

2.5.2 Squared Prediction Error (SPE)

Also known as the Q-statistic, it measures the total sum of variations in the residual space meaning the observations corresponding to $(m-\ell)$ smallest eigenvalues. It is defined as the residual distance and is computed as the squared 2-norm of this residuals, or, as its sum of squares. For an observation vector $x(k)$, the SPE is given by :

$$SPE(k) = \|e(k)\| = \sum_{i=1}^{m} (e_i(k))^2 \tag{2.17}$$

Given that:

$$e(k) = x(k) - \hat{x}(k) = (1 - C_\ell)x(k) \tag{2.18}$$

A threshold can be applied to define the the normal variations of noise for the SPE statistic, and a violation of this threshold would indicate that the noise has significantly changed, i.e. indicating the presence of abnormal situation or fault in the data. The control threshold for the SPE statistic δ_α^2 with a significance level α has been approximated

as [30] :

$$\delta_\alpha^2 = \theta_1 \left[\frac{h_0 c_\alpha \sqrt{2\theta_2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]^{1/h_0} \tag{2.19}$$

Where $\theta_i = \sum_{j=l+1}^{m} \lambda_i^j$, and λ_j is the j-th eigenvalue of the covariance matrix Σ. $i = 1, 2, 3$. $h_0 = \frac{2\theta_1\theta_3}{3\theta_2^2}$, and c_α is $(1-\alpha) \times 100$ percentile for a standard normal distribution.
The next scheme represent the idea of how fault detection works, first of all it converts the data collected from the process into a few meaningful measures, and thereby assist the operators in determining the status of the process, and if necessary in diagnosing the faults, then we build the model the establishment of an accurate PCA model is one of the key issues in the design of a PCA fault detection scheme. Considering that the process will be monitored by one/several statistics (SPE in this case), the indicators need to be computed along with their corresponding control thresholds. The next phase is the exploitation of the constructed model, i.e. fault detection phase. Now if we consider new data available from the process, fault detection is defined as the practice of determining whether these new observations from the process are in control or not. This is done by calculating the new monitoring charts, and controlling them using the computed thresholds during the previous phase. A faulty condition is declared if the computed statistic exceeds its detection threshold.

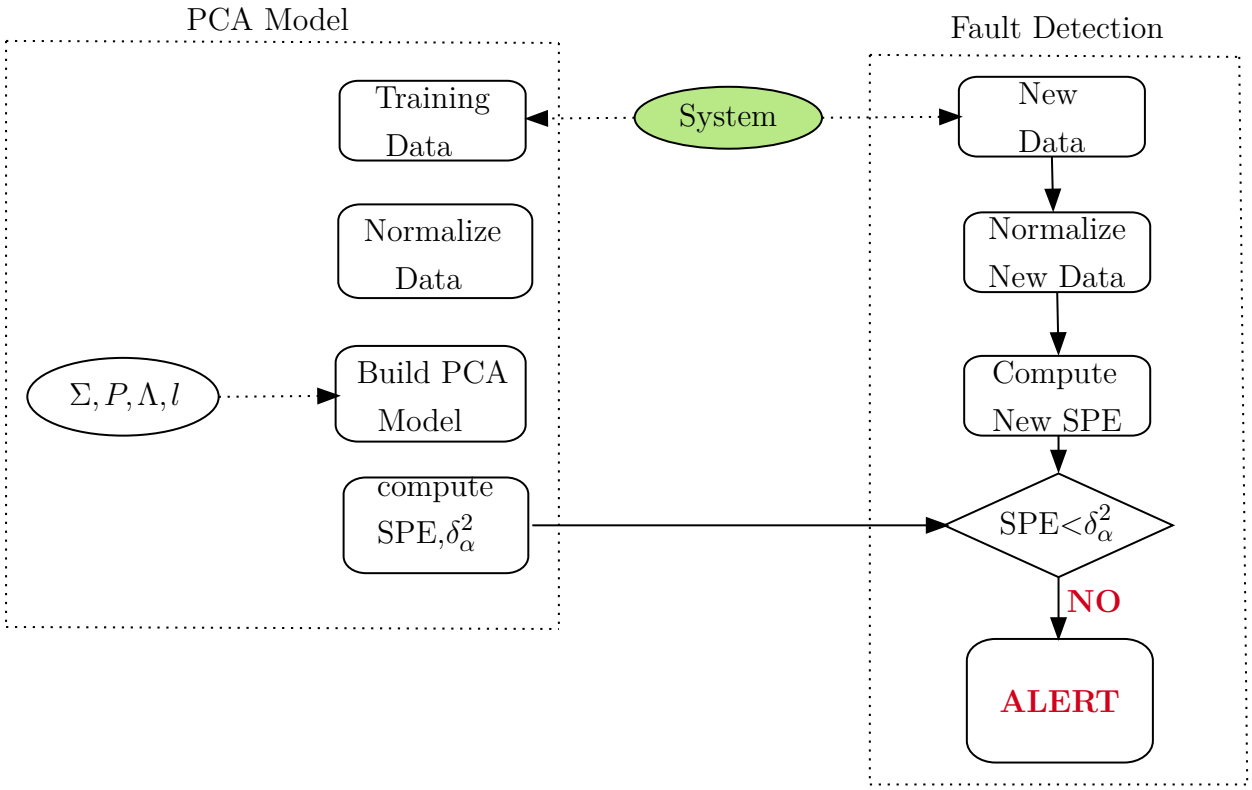


Figure 2.2: Fault Detection Scheme Based on PCA Model

In the fault detection phase, first the SPE statistic for the new observation is computed, then the SPE is checked against the corresponding threshold δ_α^2 If the SPE statistic exceeds the threshold $SPE > \delta_\alpha^2$ then a fault is detected.On the contrary, if the SPE is in-control $SPE < \delta_\alpha^2$ then the process is considered in control.

2.5.3 The new indicator ϕ (Phi)

We represent a fault indicator denoted as ϕ which is based on T^2 and SPE indicators and their common thresholds (control limits) which are calculated for a certain control level (not less than 95%), ϕ is used to avoid the ambiguity in decision for the case where only one indicator detects the fault like we will see in the simulation example, specially in small faults, thus instead of reading two indicator results only one is analysed, ϕ matrix is computed as :

$$\phi(k) = \frac{T^2(k)}{\tau_\alpha^2} + \frac{SPE(k)}{\delta_\alpha^2} \tag{2.20}$$

Where τ_α^2 and δ_α^2 represent the control limits (Thresholds) for T^2 and SPE indicators respectively, calculated for a 95% control level.

2.6 Simulation Example

Let us consider the following example constituted of 6 variables, which are defined at different time samples k,we will use this example to apply fault detection strategy based on PCA model

$$\begin{cases} x_1(k) = 0.5v_1(k) - 1.3sin(k/N)cos(k/3) \\ x_2(k) = 0.9v_2(k) - 1.2cos(k/4)^3 e^{(-k/2N)} \\ x_3(k) = x_1(k) + x_2(k) \\ x_4(k) = 2x_1(k) + x_3(k) \\ x_5(k) = x_2(k) + x_3(k) \\ x_6(k) = 2x_1(k) + x_2(k) \end{cases} \tag{2.21}$$

Where $v_1(k)$ and $v_2(k)$ are realizations randomly generated from a normal distribution with zero mean and unit variance,ie.$\sim N(0,1)$,The raw data variables are presented in :

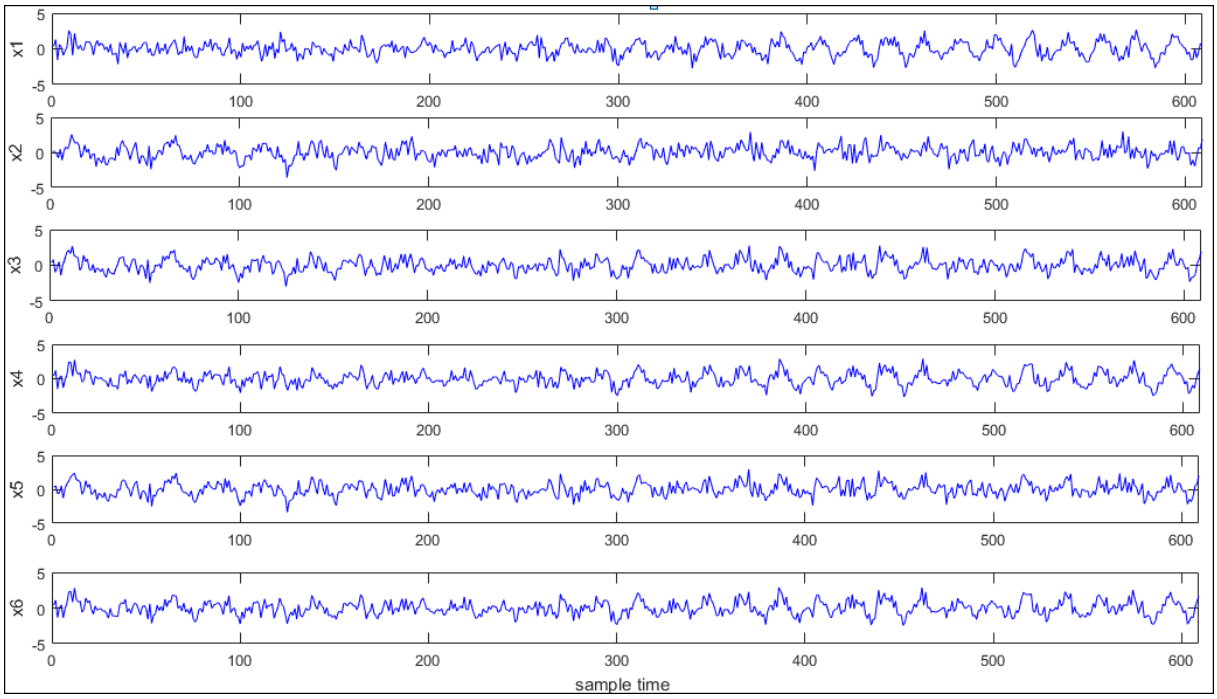


Figure 2.3: Simulation of Different variables of raw data

The first step is to center and standardize the data to remove any diiferences between the different variables of the data,after that we compute the covariance matrix as said in (2.1), then we perform an eigen-decomposition of Σ and by that we can obtain the eigenvalues and their corresponding eigenvectors. Matrices Λ and P are sorted in decreaing order and they are given by :

$$\Lambda = \begin{pmatrix} 4.7234 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.2766 & 0 & 0 & 0 & 0 \\ 0 & 0 & 5.2318\mathrm{e}^{-16} & 0 & 0 & 0 \\ 0 & 0 & 0 & 2.9326\mathrm{e}^{-16} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1.4133\mathrm{e}^{-16} & 0 \\ 0 & 0 & 0 & 0 & 0 & 3.7232\mathrm{e}^{-17} \end{pmatrix} \tag{2.22}$$

$$P = \begin{pmatrix} -0.3134 & 0.6480 & -0.2497 & 0.2316 & 0.4221 & 0.3124 \\ -0.3507 & -0.5729 & -0.3505 & -0.2331 & -0.4113 & 0.5679 \\ -0.4552 & -0.1296 & 0.7051 & -0.3609 & 0.4551 & -0.6144 \\ -0.4319 & 0.3053 & 0.3033 & -0.2299 & -0.3799 & -0.3830 \\ -0.4242 & -0.3431 & -0.0940 & 0.4290 & -0.2847 & 0.5679 \\ -0.4530 & 0.1556 & -0.4656 & -0.1202 & -0.1202 & 0.0653 \end{pmatrix} \tag{2.23}$$

By using one of the methods explained before of how to determine the number of principal components, Kaiser rule is applied (the Eigenvalue Greater Than One Rule) because it is the easiest one in this case, it can be seen that only 2 eigenvalues are greater than 1, so

the number of principal components is $\ell = 2$, the off-line thresholds are computed for the used statistics, T^2, Q-statistic (SPE), and ϕ.

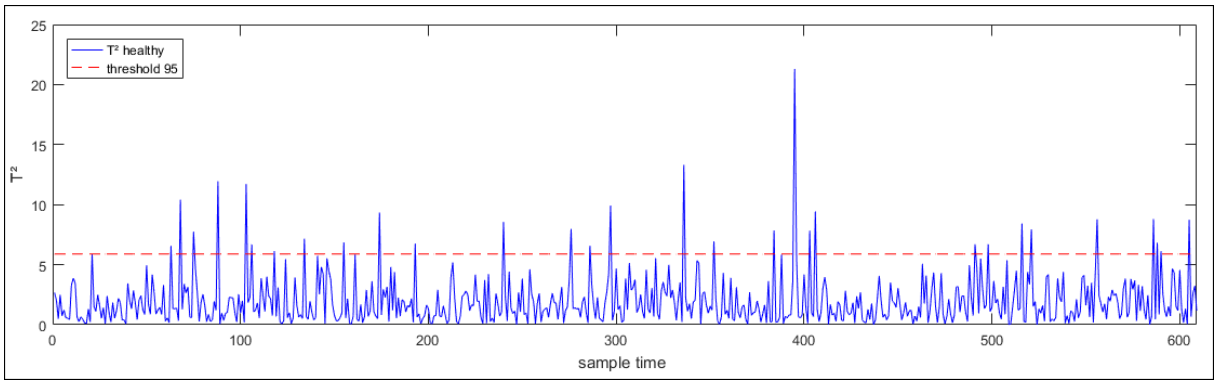


Figure 2.4: T^2 in healthy case

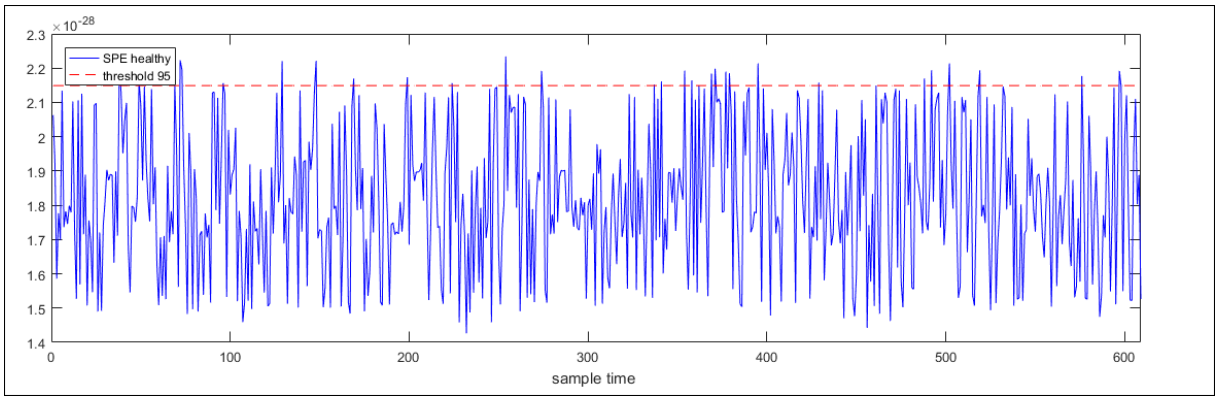


Figure 2.5: SPE in healthy case

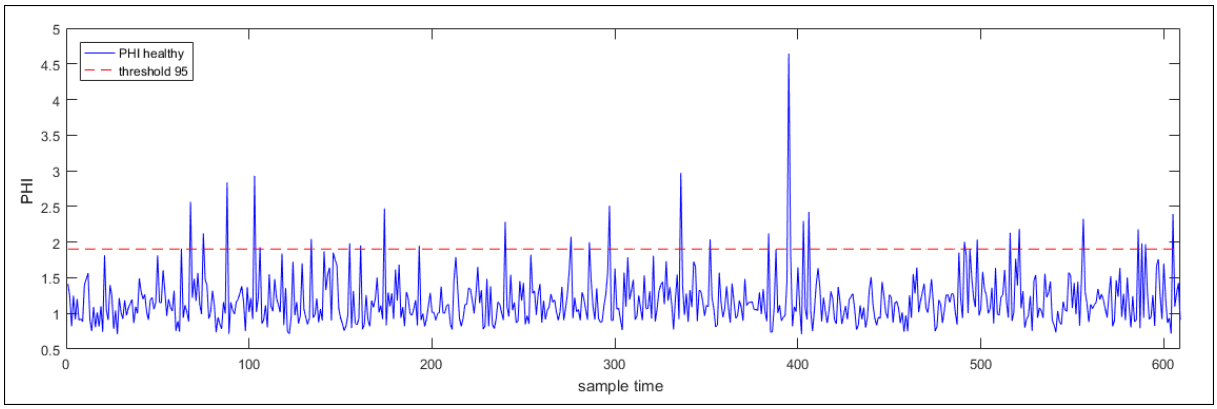


Figure 2.6: ϕ in healthy data

Now a new set of 609 samples data is generated for the example in (2.21), a fault is added in variable 4 from sample time 100 to 300 as a bias corresponding to 5% of the variable variation, the obtained results are as follows.

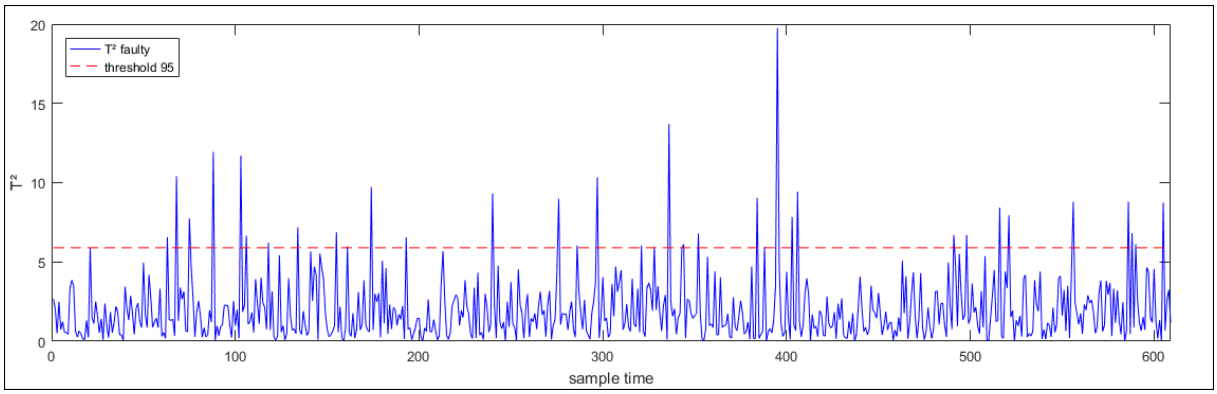


Figure 2.7: T^2 indicator after adding the fault

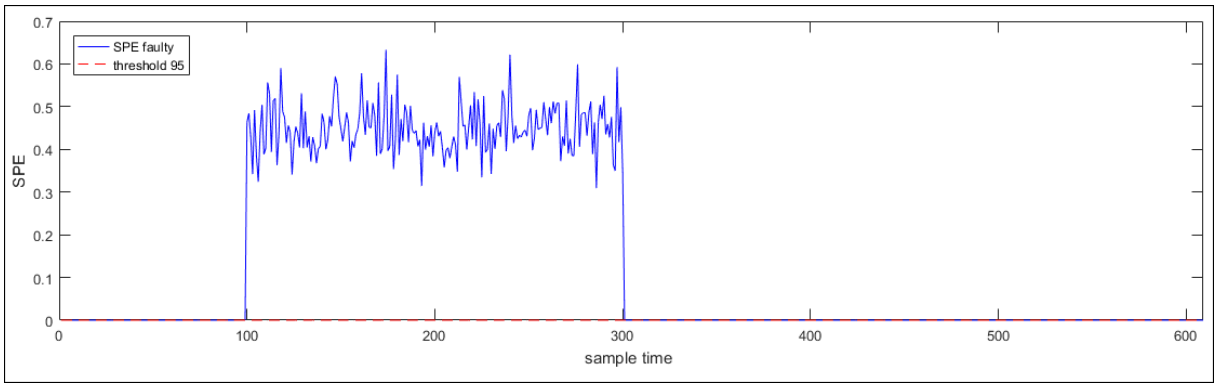


Figure 2.8: SPE indicator after adding the fault

The SPE indicator detected the injected fault correctly with the presence of a small amount of false alarms, and the T^2 statistic did not detect the fault in a good way, because the Hotelling's T^2 measures variations in the principal space, thus can only detect faults that are much higher in amplitude than the ones detected by the SPE statistic.
Now the ϕ fault indicator is represented in the faulty case using the same fault as T^2, and SPE.

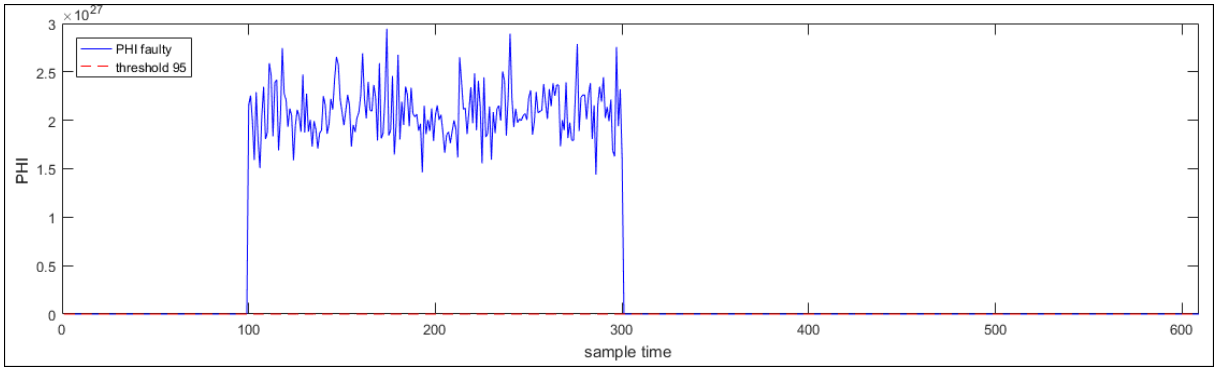


Figure 2.9: ϕ indicator after adding the fault

It can be seen That the fault is obvious, and highly exceeds the control limit, with a negligible rate of false alarms.

The performance of the statistical methods is evaluated based on the common indices : (i) fault detection time delay (time required for indicating the fault after its occurrence), (ii) False Alarms Rate (FAR), and (iii) Missed Detection Rate (MDR). The objective of a reliable monitoring scheme is to achieve shortest delays for detection and clearance of faults, moreover, lowest FAR and MDR. These last two metrics are defined respectively as follows:

$$FAR = 100\frac{N_{N,F}}{N_N}\% \tag{2.24}$$

$$MDR = 100\frac{N_{F,N}}{N_F}\% \tag{2.25}$$

Where $N_{N,F}$ is the number of normal samples detected as faults, and N_N is the number of normal samples. Whereas, $N_{F,N}$ is the number of faulty samples identified as normal, and N_F is the number of faulty samples.
The following table shows the results of FAR, and MDR after adding the fault to the data.

Table 2.1: FAR contributed by T^2, SPE, and ϕ

		FAR %		
	CL	T^2	Q-statistic	ϕ(Phi)
Abrupt fault	95%	8.91	14.19	8.87
	98%	3.97	5.90	3.95
	99%	2.87	3.97	2.65
Intermittent fault	95%	8.91	11.58	8.90
	98%	3.96	4.98	3.97
	99%	2.33	2.66	2.30
Incipient fault	95%	21.95	15.93	20.30
	98%	12.67	4.98	12.50
	99%	2.67	2.96	2.43

Table 2.2: MDR & DTD contributed by T^2, SPE, and ϕ

		MDR[%]			DTD[s]		
	CL	T^2	Q	ϕ	T^2	Q	ϕ
Abrupt fault	95%	95	1	0.6	1	0.8	0.5
	98%	98	0.4	0.3	0	0	0
	99%	99	0	0	0	0	0
Intermittent fault	95%	94.5	0.6	0.5	3	1	1
	98%	98	1.5	0.5	0	0	0
	99%	99	0.5	0.4	0	0	0
Incipient fault	95%	100	10	8	32	40	31
	98%	100	5	3.8	21	32	20
	99%	100	2	1.3	11	15	9

From the obtained results the T^2 statistic could not detect the injected fault with almost 100% miss detection rate and small fault alarm rate, good results for the new statistic indicator ϕ where we have small FAR with almost 0 [s] detection time delay.

2.7 Conclusion

In this chapter, we have seen how PCA reduces the data representation subspace and enables the determination of the redundancy relationships which are used after that to detect the faulty data. The detection of faults based on a PCA model is mainly performed using a certain statistic or indicator. Various statistics were presented, namely the T^2 and SPE indicators with their corresponding thresholds. Also, a new combined statistic was presented, ϕ, where it proves its effectiveness compared to other indicators.
PCA is a popular approach for sensor fault detection and isolation, however, it is limited in its classical off-line case to stationary processes, and cannot handle dynamic processes, or non-linear processes. Therefore, other PCA variants should be used depending on the process at hand, such as Recursive PCA, Moving window PCA, Non-linear PCA, or Kernel PCA.
The primary aim of this work is to have a more robust approach which regard the uncertainty of the process in which its data sets are always corrupted with uncertainties and noise. For that, the next chapter will address this issue by using an alternative form of PCA that takes into account these uncertainties and noise: the interval-valued PCA.

Chapter 3

PCA FOR INTERVAL-VALUED DATA

3.1 Introduction

The main idea of Multivariate Statistical Process Control (MSPC) approach is to extract the useful data information from the original data samples obtained in large quantities from a process under normal operating conditions, and construct some statistics for monitoring the state or condition of the process. So far, the most widely used MSPC method may be the principal component analysis approach. In the past decades, several improvements have been made to the traditional PCA in order to handle different natures of processes and data. However, none of them treats the particular case where process data are highly noisy due to sensor uncertainties, i.e. there is no appropriate robust approach which regards the uncertainties of sensor measurements, though it is mostly the case in practical circumstances.

In real life, data are approximate values given by the process sensors, and they are stained with uncertainties due to different causes including measurement noise, degradation from age, environmental exposure, and even maintenance interventions, etc. these factors affect the measured signals and can lead ultimately to a mass of misstatement and false alarms. A solution to avoid the influence of these uncertainties on sensors is to represent the data in the form of interval-valued data. Therefore, each measurement is no longer represented as a value, but as a set of values limited by the minimum and maximum recorded values [13]. Consequently, new analytical methods need to be developed to account for this special characteristic of data.
In the last two decades, various approaches were proposed. The first PCA adaptations for interval-valued data are known as the vertices PCA (VPCA) and the centers PCA (CPCA) [9] [10]. Both methods involve a transformation of the interval input matrix. The first

approach uses the vertices matrix to compute principal components, whereas the second uses centers matrix. Another PCA for interval-valued data is the midpoints-radii PCA (MRPCA) [50], which treats midpoints and interval ranges as two separate variables to enhance CPCA by incorporating radius. Also, an alternative approach using least squares for MRPCA has been proposed [16]. Furthermore, another analytical approach based on an interval-valued covariance matrix has been proposed [20]. This PCA for interval-valued data method with an enhanced covariance matrix calculation is called the complete information PCA (CIPCA) [70]. The following chapter tries to show that it is most likely possible to apply PCA for interval-valued data methods in monitoring uncertain systems by modeling sensors uncertainties in the form of intervals. Thus, ensuring maximum robustness of the diagnosis routine.

3.2 Interval valued-Data

In almost all cases, the true value cannot be measured and the collected data on a process are only approximations given by the sensors, and are thus imprecise. This is due to the uncertainties induced by measurement errors or determined by specific experimental conditions. In this section two different approaches to create the interval valued data are presented.

3.2.1 IVD using error of measurement

Let $X \in \mathbb{R}^{n \times m}$ be the data matrix which contains n samples of m process variables, where $x_j(k)$ is the *j-th* variable. To represent the measure by an uncertain or interval value rather than a single value. The uncertainty being unknown, we suppose that its variation is limited and can be represented by an interval of the form $[\underline{x}_j(k), \overline{x}_j(k)]$, where $\underline{x}_j(k)$ is the lower bound (LB) and $\overline{x}_j(k)]$ is the upper bound (UB) of the interval, precisely, they represent the minimum and maximum values registered for the j-th interval valued variable with respect to the k-th observation unit.
The new global interval matrix $[X]$ is formalized as a set of interval valued observation for the different description variables by the following :

$$\begin{pmatrix} [\underline{x}_1(1), \overline{x}_1(1)] & \cdots & [\underline{x}_m(1), \overline{x}_m(1)] \\ \vdots & \ddots & \vdots \\ [\underline{x}_1(n), \overline{x}_1(n)] & \cdots & [\underline{x}_m(n), \overline{x}_m(n)] \end{pmatrix} \tag{3.1}$$

Now let $\delta x_j(k)$ be the error in a measurement where $x_j(k)$ is the quantity of interest. $\delta x_j(k)$ is the difference between the true value $x_j^*(k)$ and the measurement result $x_j^c(k)$

and it is defind as :

$$\delta x_j(k) = x_j^c(k) - x_j^*(k) \tag{3.2}$$

Also the radius of the interval observation $x_j^r(k)$ defines an interval model while the center of the interval is given by the measurement $x_j^c(k)$. Thus, the standard intervation construction formula is given by :

$$[x_j(k)] = \left[x_j^c(k) - x_j^r(k), x_j^c(k) + x_j^r(k)\right] \tag{3.3}$$

Or it can represented in the form known as the midpoint-radius form by the couple (midpoint,radius), as :

$$[x_j(k)] = \left\{x_j^c(k), x_j^r(k)\right\} \tag{3.4}$$

Neither measurements nor estimates are 100% accurate, so in reality the actual value $x_j^*(k)$ of a quantity can differ from the result $x_j^c(k)$ obtained by the measurement once we perform a measurement result $x_j^c(k)$ we know that the actual (unknown) value $x_j^*(k)$ of the measured quantity belong to the interval $x_j^*(k) = [\underline{x}_j(k), \overline{x}_j(k)]$ where :

$$\underline{x}_j(k) = x_j^c(k) - \delta x_j(k) \tag{3.5}$$

And

$$\overline{x}_j(k) = x_j^c(k) + \delta x_j(k) \tag{3.6}$$

3.2.2 IVD using Moving Window approach

Another approach to deal with quantitative variables and to create the upper and lower bounds is using moving window method. Moving window (MW) is a criteria which cover the entire data matrix by increasing the sampled data and removing the previous sampled data at the same time. This method can avoid the impact of the data before the steady state on the data of the failure period, it can also make fault data characteristics more obvious when faults occur. For the data matrix $X_{n\times m}$ and by selecting the parameter f as the window length, then the k-th window matrix is as follows:

$$\begin{bmatrix} x_{k1} & x_{k2} & \cdots & x_{km} \\ x_{k+1,1} & x_{k+1,2} & \cdots & x_{k+1,m} \\ \cdots & \cdots & \cdots & \cdots \\ x_{k+f-1,1} & x_{k+f-1,2} & \cdots & x_{k+f-1,m} \end{bmatrix} \tag{3.7}$$

For the matrix $X_{n\times m}$ take the window length of f in total we need to calculate $n-f+1$ window matrix.

How to choose an optimal size of the moving window is another issue Generally, the window size should be large enough to cover sufficient sample data for modeling and monitoring. However, the large window size leads to a significant reduction in the computational efficiency. Furthermore, when the process changes rapidly, the window covers too much outdated sample data which leads to failure in tracking the process change, Although a smaller window size can enhance computational efficiency, data within the window may not properly represent the underlying relationships between the process variables. Besides, the too small window may result in a potential danger in that the generated model may adapt so quickly to process changes that abnormal behavior is undetected. Up to now, most model-updating approaches have empirically used a fixed window size. To the best of our knowledge, there is little concern for how to determine the optimal size of a moving window.

Now to find our bounds (LB) and (UB) we have to extract the standard deviation and the mean for each variable of our data matrix, then calculate the control limit (threshold) for each variable for a certain control level (not less than 95%), and each time we slide the window for a given window length till we find $UB_1, UB_2, ..., UB_m$, same work is done for lowe limit $LB_1, LB_2, ..., LB_m$.

At the end we gather the lower bounds and upper bounds in separate matrices to form upper limit matrix and lower limit matrix $XU = [UB_1, UB_2, ..., UB_m]$, where XU denotes upper limit of data matrix X, same for the lower limit $XL = [LB_1, LB_2, ..., LB_m]$, the formula for lower and upper limits are as follows:

$$\begin{cases} xL_1(k) = \mu_1 - CL_1 * \sigma_1 \\ xL_2(k) = \mu_2 - CL_2 * \sigma_2 \\ \vdots \\ xL_m(k) = \mu_m - CL_m * \sigma_m \end{cases} \tag{3.8}$$

$$\begin{cases} xU_1(k) = \mu_1 + CL_1 * \sigma_1 \\ xU_2(k) = \mu_2 + CL_2 * \sigma_2 \\ \vdots \\ xU_m(k) = \mu_m + CL_m * \sigma_m \end{cases} \tag{3.9}$$

Where xU and xL denotes upper and lower states of the data matrix X, μ represents the mean which is calculated for each variable, CL is the control limit (threshold) evaluated at 95%, σ is the standard deviation computed for each variable of X.

3.2.3 Interval Arithmetic and Statistics

In this section we will review some basic interval arithmetic operations and statistics for interval computations which will be used later in PCA for the interval valued-data

Definition 3.2.3.1. *The midpoint of a generic interval [x], or center is given by:*
$x^c = \frac{\underline{x}+\overline{x}}{2}$

Definition 3.2.3.2. *The radius of a generic interval [x] is given by:*
$x^r = \frac{\overline{x}-\underline{x}}{2}$
And the range of an interval [x] is its width given by:
$w([x]) = \overline{x} - \underline{x}$

Definition 3.2.3.3. *An n-dimensional interval vector [x], is the ordered n-tuple of intervals given by:*
$[x] = [[x_1], [x_2], ..., [x_m]]$.
The $n \times m$ dimensional matrix X is the ordered m-tuple of interval-valued vectors$[x_i], i = 1, ..., m$ given by $X = \left[[\mathrm{x}_1]^T, [\mathrm{x}_2]^T, ..., [\mathrm{x}_m]^T\right]$

Definition 3.2.3.4. *the interval mean* $[m]$ *of of an interval-valued vector* [x] *is defined as:*
$[m] = \frac{1}{n}\sum_i [x_i]$

Definition 3.2.3.5. *The distance between two generic intervals [x] and [y] is given by:*
$d([x],[y]) = |x^c - y^c| + |x^r + y^r|$
Where $d([x],[y])$ satisfies the Euclidean distance properties.

3.2.4 Interval-valued Data Normalisation

Typically, some normalisation must be performed prior to processing data in order to objective or scale-invariant result. we will describe below some alternative normalisation methods for the case of interval-valued data.

3.2.4.1 normalisation using the dispersion of interval center and range

Let $\{[x_j(1)], [x_j(2)], ..., [x_j(k)]\}$ be a set of finite interval, so that $[x_j(k)] \subset \mathbb{R}\ \forall k \in \{1, ..., n\}$ and $[m_j]$ is their corresponding mean interval, then :
$\sum_{k=1}^n \left[(x_j^c(k) - m_j^c) + (x_j^r(k) - m_j^r)\right] = 0$ and $\sum_{k=1}^n d^2([x_j(k)], [m_j])$ is minimized. Where $x_j^c(k)$ and $x_j^r(k)$ are are the midpoints and the radius of $[x_j(k)]$ according to previous definitions, m_j^c and m_j^r are their respective means. We define the variance as the sum of squared distances with respect to the mean interval, as a consequence the variance σ_j^2 for interval valued data is defined as follows: $\sigma^2 = \frac{1}{n}\sum_{k=1}^n d^2([x_j(k)], [m_j])$ or it can also be

written according to the following formula :

$$\sigma^2 = \frac{1}{n}\sum_{k=1}^{n}\left(\left|x_j^c(k) - m_j^c\right| + \left|x_j^r(k) - m_j^r\right|\right)^2 \tag{3.10}$$

Where $[m_j] = \left[\frac{1}{n}\sum_{k=1}^{n}\underline{x}_j(k) \qquad \frac{1}{n}\sum_{k=1}^{n}\overline{x}_j(k)\right]$, $m_j^c = \frac{1}{n}\sum_{k=1}^{n}x_j^c(k)$ and $m_j^r = \frac{1}{n}\sum_{k=1}^{n}x_j^r(k)$
So the variance obtained will be :

$$\sigma^2 = \frac{1}{n}\left[\sum_{k=1}^{n}\left(x_j^c(k) - m_j^c\right)^2 + \sum_{k=1}^{n}\left(x_j^r(k) - m_j^r\right)^2 + 2\sum_{k=1}^{n}\left|x_j^c(k) - m_j^c\right|\left|x_j^r(k) - m_j^r\right|\right] \tag{3.11}$$

The distance between intervals can be generalized to the Euclidean distance in the space $\mathbb{R}^m$, So a normalized interval is :

$$\left[\frac{1}{\sigma}\left(x_j^c(k) - m_j^c - \left|x_j^r(k) - m_j^r\right|\right), \frac{1}{\sigma}\left(x_j^c(k) - m_j^c + \left|x_j^r(k) - m_j^r\right|\right)\right] \tag{3.12}$$

3.2.4.2 normalisation using the dispersion of the interval centers

This method normalises such that the resulting transformed midpoints have zero mean and unit variance for each variable.
The mean value and the dispersion of all interval midpoint are given by:

$$m_j = \frac{1}{n}\sum_{k=1}^{n}\frac{\left(\underline{x}_j(k) + \overline{x}_j(k)\right)}{2} \qquad and \qquad \sigma_j^2 = \frac{1}{n}\sum_{k=1}^{n}\left(\frac{\underline{x}_j(k) + \overline{x}_j(k)}{2} - m_j\right)^2 \tag{3.13}$$

With this notation the normalised interval is defined with boudaries:

$$\left[\frac{\underline{x}_j(k) - m_j}{\sigma_j}, \frac{\overline{x}_j(k) - m_j}{\sigma_j}\right] \tag{3.14}$$

3.2.4.3 normalisation using the dispersion of the interval boundaries

The interval boundaries normalisation method normalises the data such that for each variable $[X_j]$ the n intervals $[x_j(k)]$ such that the mean and the joint dispersion of the rescaled interval boundaries are 0 and 1, respectively. The joint dispersion of a variable $[X_j]$ is defined by:

$$\sigma_j^2 = \frac{1}{n}\sum_{k=1}^{n}\frac{\left(\underline{x}_j(k) - m_j\right)^2 + \left(\overline{x}_j(k) - m_j\right)^2}{2} \tag{3.15}$$

So for $k = 1, ..., n$ the intervals $[x_j(k)] = [\underline{x}_j(k), \overline{x}_j(k)]$ are transformed into:

$$\left[\frac{\underline{x}_j(k) - m_j}{\sigma_j}, \frac{\overline{x}_j(k) - m_j}{\sigma_j}\right] \tag{3.16}$$

3.2.4.4 normalisation using the global range

This normalisation methods transforms, for a given variable, the intervals $[x_j(k)] = [\underline{x}_j(k), \overline{x}_j(k)]$, $k = 1, ..., n$ such that the range of the n rescaled intervals is the unit interval [0, 1].
Let Min_j=min$\left\{\underline{x}_j(k), ..., \underline{x}_j(k)\right\}$ and Max_j=max$\{\overline{x}_j(k), ..., \overline{x}_j(k)\}$ be the extremal lower and upper boundary values. With this notation, the interval is transformed into normalised interval with boundaries:

$$\left[\frac{\underline{x}_j(k) - Min_j}{Max_j - Min_j} \qquad \frac{\overline{x}_j(k) - Min_j}{Max_j - Min_j}\right] \tag{3.17}$$

3.3 Interval-valued PCA

The key to perform PCA on a set of data is the calculation of correlation (or covariance) matrix and it's eigen-decomposition, real life applications makes the problem of computing eigenvalues of a matrix more complicated by imposing uncertainties and errors on collected measurements to solve this problem the measurements will be represented as intervals. Many works on the eigenvalue problem were developed, and several approximation algorithms were proposed. Unfortunately, the theoretical background for the eigenvalue problem of symmetric interval-valued matrices is not wide enough and there are few practical methods. Few existing approaches to interval-valued PCA use these methods, and are known as interval algebra based PCA's, but only work for very narrow intervals, especially with large sample size data, hence limiting their applicability. Alternatively, and given the absence of well established mathematics in the matter, several intervalvalued PCA methods perform a codification of the initial interval-valued data set based on a certain geometrical representation.
For fault detection of interval-valued PCA, an accurate PCA model have to be used. In this section, four most known interval PCA approaches: Vertices PCA, Centers PCA, Midpoints-Radii PCA and the Complete information PCA are presented, discussed , and investigated for use in process modelling. In order to avoid unwanted differences among the variables. Without loss of generality, we assume in the following that interval-valued observations have been normalized according to methods mentioned before in section (3.2.4).

3.3.1 Centers PCA

PCA for interval-valued data methods presented are mainly based on the analysis of coded/ transformed matrices from initial interval-valued data. The Centers PCA (CPCA) decomposes the correlation matrix of the centers coded matrix X^c, so the data matrix X

is coded in terms of centers by the following :

$$\begin{pmatrix} x_1^c(1) & \cdots & x_m^c(1) \\ \vdots & \ddots & \vdots \\ x_1^c(n) & \cdots & x_m^c(n) \end{pmatrix} \tag{3.18}$$

Where $x_j^c(k)$ is the midpoint or center of the interval. Then, the CPCA is performed by applying classical PCA on centers matrix X^c, and the covariance matrix Σ_c is calculated as:

$$\Sigma_c = \frac{1}{n-1} X^{cT} X^c \tag{3.19}$$

A straightforward way of calculating covariance Σ_c for the interval-valued matrix $[X]$, as:

$$\Sigma_c = \begin{pmatrix} \frac{1}{4n}\sum_{k=1}^{n}(\underline{x}_1(k)+\overline{x}_1(k))^2 & \frac{1}{4n}\sum_{k=1}^{n}(\underline{x}_1(k)+\overline{x}_1(k))(\underline{x}_2(k)+\overline{x}_2(k)) & \cdots & \frac{1}{4n}\sum_{k=1}^{n}(\underline{x}_1(k)+\overline{x}_1(k))(\underline{x}_m(k)+\overline{x}_m(k)) \\ \frac{1}{4n}\sum_{k=1}^{n}(\underline{x}_2(k)+\overline{x}_2(k))(\underline{x}_1(k)+\overline{x}_1(k)) & \frac{1}{4n}\sum_{k=1}^{n}(\underline{x}_2(k)+\overline{x}_2(k))^2 & \cdots & \frac{1}{4n}\sum_{k=1}^{n}(\underline{x}_2(k)+\overline{x}_2(k))(\underline{x}_m(k)+\overline{x}_m(k)) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{4n}\sum_{k=1}^{n}(\underline{x}_m(k)+\overline{x}_m(k))(\underline{x}_1(k)+\overline{x}_1(k)) & \frac{1}{4n}\sum_{k=1}^{n}(\underline{x}_m(k)+\overline{x}_m(k))(\underline{x}_2(k)+\overline{x}_2(k)) & \cdots & \frac{1}{4n}\sum_{k=1}^{n}(\underline{x}_m(k)+\overline{x}_m(k))^2 \end{pmatrix} \tag{3.20}$$

The principal components of classical PCA on centers matrix, $T_1^c, ..., T_m^c$, are obtained via mapping using $T^c = X^c P$ given that $P = p_1, ..., p_m$ are the eigenvectors from spectral decomposition of matrix Σ_c. The interval-valued principal components are then constructed as:

$$\begin{cases} \underline{t}_j(k) = \sum_{i=1,p_{ij}<0}^{m} \overline{x}_i(k)p_{ij} + \sum_{i=1,p_{ij}>0}^{m} \underline{x}_i(k)p_{ij} \\ \\ \overline{t}_j(k) = \sum_{i=1,p_{ij}<0}^{m} \underline{x}_i(k)p_{ij} + \sum_{i=1,p_{ij}>0}^{m} \overline{x}_i(k)p_{ij} \end{cases} \tag{3.21}$$

Where pij is the i^{th} element of the j-th column of eigenvector matrix P. Similarly, the interval-valued estimates based on CPCA model for the first ℓ components are obtained as:

$$\begin{cases} \underline{\hat{x}}_j(k) = \sum_{i=1,c_{ij}<0}^{m} \overline{x}_i(k)c_{ij} + \sum_{i=1,c_{ij}>0}^{m} \underline{x}_i(k)c_{ij} \\ \\ \overline{\hat{x}}_j(k) = \sum_{i=1,c_{ij}<0}^{m} \underline{x}_i(k)c_{ij} + \sum_{i=1,c_{ij}>0}^{m} \overline{x}_i(k)c_{ij} \end{cases} \tag{3.22}$$

Where c_{ij} is the i^{th} element of the j-th column of matrix C_ℓ.
CPCA performed on interval input data is based on a numerical centers codification of the data, a treatment with classical PCA analysis technique, and finally a transformation of classical results into interval description.

3.3.2 Midpoints-Radii PCA

The midpoints-radii PCA (MRPCA) introduce an approach PCA for interval-valued data in a different way. Problems with statistical analysis of interval data using standard interval arithmetic can be avoided by representing them using interval midpoints and ranges, i.e. $[X] = \{X^c, X^r\}$. The midpoints-radii PCA (MRPCA) on interval-valued data is a hybrid method that is resolved in terms of midranges (X^r) and midpoints (X^c), given by t definitions of radius and mid-point explaind before and their interconnection. MRPCA can be considered as an improvement over CPCA by including radius.
Considering the definition of interval distance, the variance for interval-valued data, can be expressed by :

$$\sigma^2 = \frac{1}{n}\left[\sum_{k=1}^{n}\left(x_j^c(k) - m_j^c\right)^2 + \sum_{k=1}^{n}\left(x_j^r(k) - m_j^r\right)^2 + 2\sum_{k=1}^{n}\left|x_j^c(k) - m_j^c\right|\left|x_j^r(k) - m_j^r\right|\right] \quad (3.23)$$

The covariance is defined as the sum of three components: the variance between midpoints, the variance between radii, and twice the measure of congruence between midpoints and radii. It follows that the global covariance matrix Σ is given by :

$$\Sigma = \frac{1}{N}\left(X^{cT}X^c\right) + \frac{1}{N}\left(X^{rT}X^r\right) + \frac{1}{N}\left(\left|X^{cT}X^r\right| + \left|X^{rT}X^c\right|\right) \quad (3.24)$$

According to MRPCA, two independent PCA's are singly exploited on the two matrices of midpoints and radius. The solutions are given by the following eigen-systems :

$$X^c\Sigma_c^{-1}P^c = \Lambda^c P^c \quad (3.25)$$

$$X^r\Sigma_r^{-1}P^r = \Lambda^r P^r \quad (3.26)$$

Where Λ^c, P^c and Λ^r, P^r are, respectively, the eigenvalues and eigenvectors of the two partial eigen-decomposition of midpoints and radii matrices, and given that :

$$\Sigma_c = \frac{1}{N}\left(X^{cT}X^c\right) and\, \Sigma_r = \frac{1}{N}\left(X^{rT}X^r\right) \quad (3.27)$$

The two independent PCA's on midpoints and radius do not however cover the whole variance in (3.34), to include variance of the interconnection between midpoints and radius given by the term $\left(\left|X^{cT}X^r\right| + \left|X^{rT}X^c\right|\right)$, at the same time, giving a graphical interpretation of the interval-valued units. The radius coordinates are thus rotated and then superimposed on the midpoints PC's as supplementary points, in order to get a logical graphical representation of the statistical units based on MRPCA model. Two choices for rotation of radius are defined below :

The first choice for computing the radius rotation matrix can be achieved by maximizing the Tucker congruence coefficient between midpoints and radii and it is given by :

$$f(A) = \sum_{j=1}^{m} \left(\frac{a_j^T X^{cT} X_j^r}{\left(a_j^T X^{cT} X^c a_j\right)^{1/2} \left(X_j^{rT} X_j^r\right)^{1/2}} \right) \tag{3.28}$$

Under the constraint $A^T A = 1$, and given that $A = [a_1, ..., a_m]$ is the rotation matrix of radii, and X_j^r denotes the radii of j-th interval-valued variable $[X_j]$.
A second choice for computing rotation matrix is expressed in terms of singular value decomposition.

$$X^{cT} X^r = P \Lambda^{cr} Q^T \tag{3.29}$$

Where Λ^{cr} is the eigenvalues matrix of covariance $X^{cT} X^r$, the matrix A is defined as :

$$A = Q P^T \tag{3.30}$$

Let us consider that we have performed two classical PCA's on X^c and X^r, then principal components of midpoints radii are given by :

$$\begin{cases} T^c = X^c P^c \\ \\ T^r = X^r P^r \end{cases} \tag{3.31}$$

It follows that the construction of the interval-valued component vector $[\mathbf{t}(k)]$ using rotated ranges via matrix A is :

$$\begin{cases} \underline{\mathbf{t}}(k) = \mathbf{t}^c(k) - A\mathbf{t}^r(k) \\ \\ \overline{\mathbf{t}}(k) = \mathbf{t}^c(k) + A\mathbf{t}^r(k) \end{cases} \tag{3.32}$$

Hence, for a number of components ℓ, the interval-valued estimates of vector $[\mathrm{x}(k)]$ are obtained similarly, by projecting rotated ranges estimates $\hat{x}^r(k)$ on the estimated midpoints $\hat{x}^c(k)$ as :

$$\begin{cases} \underline{\hat{\mathrm{x}}}(k) = \hat{x}^c(k) - A\hat{x}^r(k) \\ \\ \overline{\hat{\mathrm{x}}}(k) = \hat{x}^c(k) + A\hat{x}^r(k) \end{cases} \tag{3.33}$$

3.3.3 Vertices PCA

The Vertices Principal Component Analysis (V-PCA), offers the possibility to detect the underlying structure of the two-way interval valued data set stored in $X_{n\times m}$:

$$X = \begin{pmatrix} [\underline{x}_1(1), \overline{x}_1(1)] & \cdots & [\underline{x}_m(1), \overline{x}_m(1)] \\ \vdots & \ddots & \vdots \\ [\underline{x}_1(n), \overline{x}_1(n)] & \cdots & [\underline{x}_m(n), \overline{x}_m(n)] \end{pmatrix}$$

The i-th row of X pertains to the i-th observation unit, $i = 1, ..., n$. As each observation unit is characterized by m interval valued variables it can be represented as a hyperrectangle in $\Re^m$, and the number of vertices of each hyperrectangle is 2^m. Hence an an interval valued vector $[\mathrm{x}(k)] = [[\underline{x}_1(k), \overline{x}_1(k)], ..., [\underline{x}_m(k), \overline{x}_m(k)]]$ in an m dimensional space, can be represented by a numerical data matrix X_i of 2^m lines and m columns. Vertices are given by all the possible combinations between the bounds of the vector $[\mathrm{x}(k)]$ As an example, given a two dimensional interval vector $[\mathrm{x}(k)] = [[\underline{x}_1(k), \overline{x}_1(k)], [\underline{x}_2(k), \overline{x}_2(k)]]$ k)]], vertices matrix V (k) is constructed as follows:

$$X_i = \begin{pmatrix} \underline{x}_1(k) & \underline{x}_2(k) \\ \underline{x}_1(k) & \overline{x}_2(k) \\ \overline{x}_1(k) & \underline{x}_2(k) \\ \overline{x}_1(k) & \overline{x}_1(k) \end{pmatrix} \tag{3.34}$$

In other words, each bound of the interval-valued observations is a vertex of the corresponding hyper-cube, i.e. a point has 1 vertex, a segment has 2 vertices, a rectangle 4 vertices, and so on.

V-PCA does not directly summarize the interval valued data X, it is replaced by a single valued data matrix obtained as follows. Each interval valued row is transformed into the numerical matrix X_i as follows :

$$X_i = \begin{pmatrix} \underline{x}_1(k) & \underline{x}_2(k) & \cdots & \underline{x}_m(k) \\ \overline{x}_1(k) & \underline{x}_2(k) & \cdots & \underline{x}_m(k) \\ \vdots & \vdots & \ddots & \vdots \\ \overline{x}_1(k) & \overline{x}_2(k) & \cdots & \overline{x}_m(k) \end{pmatrix} \tag{3.35}$$

By stacking below each other all the matrices $X's$, $i = 1, ..., n$, we get the new numerical valued data matrix $X_{\texttt{V-PCA}}$ with $n2^m$ rows and m columns :

$$X_{\texttt{V-PCA}} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} \tag{3.36}$$

V-PCA is based on performing PCA on $X_{\texttt{V-PCA}}$, where it can be preprocessed as in the standard single valued case.
The application of PCA to the matrix $X_{\texttt{V-PCA}}$ will give $\hat{X}_{\texttt{V-PCA}} = T_{\texttt{V-PCA}} P^T$, and if P is chosen to be column-wise orthonormal, then $T_{\texttt{V-PCA}} = X_{\texttt{V-PCA}} P$. To facilitate the interpretation of the solution, for each observation unit, for each component, the segment containing all component scores for vertices associated with this observation unit. Specifically, with respect to the k-th component, $k = 1, ..., p$, if n_i denotes the set of all the vertices for the i-th observation unit, $i = 1, ..., n$. The dimension of matrix $X_{\texttt{V-PCA}}$ is huge which will make the PCA of a such matrix practically impossible. This computational problem can be overcome by considering a special property of PCA. Specifically, it is well-known that the columns of the component loadings matrix are the eigenvectors obtained from the eigen-decomposition of the cross-products matrix. Note that the eigenvectors are arranged in such a way that the first ones are associated with the highest eigenvalues. Dealing with the cross-products matrix $\Sigma_V = X^T_{\texttt{V-PCA}} X_{\texttt{V-PCA}}$, after simplification the covariance matrix Σ_V can be written as :

$$\Sigma_V = 2^{m-2} \begin{pmatrix} 2\sum_{k=1}^{n} \left(\underline{x}_1^2(k) + \overline{x}_1^2(k)\right) & \sum_{k=1}^{n} (\underline{x}_1(k) + \overline{x}_1(k))(\underline{x}_2(k) + \overline{x}_2(k)) & \cdots & \sum_{k=1}^{n} (\underline{x}_1(k) + \overline{x}_1(k))(\underline{x}_m(k) + \overline{x}_m(k)) \\ \sum_{k=1}^{n} (\underline{x}_2(k) + \overline{x}_2(k))(\underline{x}_1(k) + \overline{x}_1(k)) & 2\sum_{k=1}^{n} \left(\underline{x}_2^2(k) + \overline{x}_2^2(k)\right) & \cdots & \sum_{k=1}^{n} (\underline{x}_2(k) + \overline{x}_2(k))(\underline{x}_m(k) + \overline{x}_m(k)) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{k=1}^{n} (\underline{x}_m(k) + \overline{x}_m(k))(\underline{x}_1(k) + \overline{x}_1(k)) & \sum_{k=1}^{n} (\underline{x}_m(k) + \overline{x}_m(k))(\underline{x}_2(k) + \overline{x}_2(k)) & \cdots & 2\sum_{k=1}^{n} \left(\underline{x}_m^2(k) + \overline{x}_m^2(k)\right) \end{pmatrix} \tag{3.37}$$

The components can be extracted by performing the eigen-decomposition on Σ_V, indeed the obtained loading matrix P is columnwise orthonormal. However, this would require that we nevertheless use the huge matrix $X_{\texttt{V-PCA}}$, a short cut is used to define what may be called the positive, and negative component loadings matrices, respectively, P^+ and P^- with generic elements

$$p_{jk}^{+} = \begin{cases} p_{jk} & if p_{jk} \geq 0 \\ \\ 0 & otherwise \end{cases} \tag{3.38}$$

$$p_{jk}^{-} = \begin{cases} p_{jk} & if p_{jk} \leq 0 \\ \\ 0 & otherwise \end{cases} \tag{3.39}$$

Where p_{jk} gives the loading of variable j on component k. thus P^+ and P^- contains the non-negative (non-positive) elements of P whereas the negative (positive) elements of P are replaced by 0. In matrix notation, the bounds of the component scores matrix are given by :

$$\begin{cases} \underline{t}(k) = \overline{X}(k)P^- + \underline{X}(k)P^+ \\ \\ \overline{t}(k) = \overline{X}(k)P^+ + \underline{X}(k)P^- \end{cases} \tag{3.40}$$

Thus it can be seen that score matrices were computed without explicitly having to compute all the component scores for all the vertices. It follows that this computational approach to V-PCA finds the same component loadings and the same segments for the observation units, as the original computational approach to V-PCA. We only lose the component scores of all individual vertices, but not of the segments that enclose them. Estimates of interval measurements are also computed as said before:

$$\hat{X}_{\texttt{V-PCA}} = T_{\texttt{V-PCA}} P^T \tag{3.41}$$

The estimated interval-valued measurements for the ℓ principal components are then computed as:

$$\begin{cases} \underline{\hat{x}}_j(k) = \underline{t}(k)P^- \\ \\ \overline{\hat{x}}_j(k) = \overline{t}(k)P^+ \end{cases} \tag{3.42}$$

3.3.4 Complete Information PCA

A new interval PCA method called complete information based PCA (CIPCA) was developed in order to seize more information within interval observations. CIPCA accomplishes the derivation of interval-valued principal components and transforms PCA modeling into the computation of some inner products. Thus, leading to more accurate results and providing an efficient and effective way for conducting PCA on large-scaled interval data. First of all let's define the following:

- ***Inner product of interval-valued variables:***
  Any two interval-valued variables $[X_j]$ and $[X_{j'}]$ their inner product is given by :

$\Big\langle [X_j], [X_{j'}] \Big\rangle = \sum_{k=1}^{n} \Big\langle [x_j(k), x_{j'}(k)] \Big\rangle$ where :
$\Big\langle [x_j(k), x_{j'}(k)] \Big\rangle = \frac{1}{4} \Big(\underline{x}_j(k) + \overline{x}_j(k) \Big) \Big(\underline{x}_{j'}(k) + \overline{x}_{j'}(k) \Big)$

- ***Interval squared norm:***
  In the case of auto-correlation given by $\langle [X_j(k)], [X_j(k)] \rangle$ the inner product is equal to the squared norm extended to the interval case$\|[X_j]\|^2$, defined by :
  $\|[X_j]\|^2 = \sum_{k=1}^{n} \|[x_j(k)]\|^2$ where :
  $\|[x_j(k)]\|^2 = \frac{1}{3} \Big(\underline{x}_j^2(k) + \underline{x}_j(k) \overline{x}_j(k) + \overline{x}_j^2(k) \Big)$

Based on the above definitions of interval norm and inner product, and with Matrix $[X]$ being normalized, the covariance matrix Σ of $X_{n \times m}$ is given by :

$$\Sigma = \frac{1}{n} \begin{pmatrix} \langle [X_1], [X_1] \rangle & \langle [X_1], [X_2] \rangle & \cdots & \langle [X_1], [X_m] \rangle \\ \langle [X_2], [X_1] \rangle & \langle [X_2], [X_2] \rangle & \cdots & \langle [X_2], [X_m] \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle [X_m], [X_1] \rangle & \langle [X_m], [X_2] \rangle & \cdots & \langle [X_m], [X_m] \rangle \end{pmatrix} \tag{3.43}$$

The determination of the interval-valued principal components $[T]$ in CIPCA method is based on a linear combination algorithm for interval-valued variables.
Let us consider that we performed an eigen-decomposition of the covariance matrix Σ, where $\lambda_1, ..., \lambda_m$ and $p_1, ..., p_2$ are the resulting eigenvalues and eigenvectors respectively. The interval-valued principal components are thus computed as :

$$\begin{cases} \underline{t}_j(k) = \sum_{i=1}^{m} p_{ij} \left(\tau \underline{x}_i(k) + (1-\tau) \overline{x}_i(k) \right) \\ \\ \overline{t}_j(k) = \sum_{i=1}^{m} p_{ij} \left((1-\tau) \underline{x}_i(k) + \tau \overline{x}_i(k) \right) \end{cases} \tag{3.44}$$

With

$$\tau = \begin{cases} 0, p_{ij} \leq 0 \\ 1, p - ij \geq 0 \end{cases} \tag{3.45}$$

It follows that interval-valued estimates from CIPCA model are obtained as:

$$\begin{cases} \underline{\hat{x}}_j(k) = \sum_{q=1}^{m} C_{\ell qj} \Big(\tau \underline{x}_q(k) + (1-\tau) \overline{x}_q(k) \Big) \\ \\ \overline{\hat{x}}_j(k) = \sum_{q=1}^{m} C_{\ell qj} \Big((1-\tau) \underline{x}_q(k) + \tau \overline{x}_q(k) \Big) \end{cases} \tag{3.46}$$

With the same condition on τ, and given that $C_\ell = P_\ell P_\ell^T$, with ℓ is the number of chosen principal components.

3.3.5 Determining the Number of PC's for interval-Valued PCA

The number of principal components has a significant impact on each step of the PCA based sensor for fault detection as well as it's performance. In section (1.3) various methods of determining the number of PC's were mentioned and detailed, which include: the predicted residual error sum of squares ($PRESS$) or cross validation criterion, Kaiser rule, The cumulative percentage of variance (CPV)..., these methods were developed only for single-valued PCA approach and do not cover the interval-valued PCA case. However, The interval-valued PCA methods detailed above are based on various approximations, and the resulting covariance matrix, eigenvalues and eigenvectors are single-valued. Thus, classical methods for determining the number of principal components could be employed for interval-valued PCA, that is, due to the fact that all the methods for selecting PC's depend partially or globally on the triplet Covariance, Eigenvalues, Eigenvectors.
In our computations we have used the cumulative percentage of variance (CPV) to determine the number of principle components, where we have to choose ℓ so that the cumulative percentage of variance $CPV(\ell)$ is high enough, the range used is 90% or more to get an accurate models.

3.4 Interval-valued PCA for Fault Detection

Fault detection based on PCA aims to detect deviations from typical process behaviour by evaluating whether a process is statistically in control or not, which is done by analysing different projections of the new process data into the principal or residual sub-spaces, which are obtained according to the PCA model. Application of PCA-based fault detection involves using several statistics for monitoring the condition of the process. The most used statistics are the Hotteling's T^2 test, and the squared prediction error SPE, this analysis is used for the case of single-valued data.
If a measurement is considered as the average of the collected measured values during a sampling period, it may cause an excessive rate of false alarms and missing detection When this type of data is used for a PCA based monitoring. In addition to that, to obtain a reliable PCA model representing the process, we need to use a good amount of samples, which are not always available.
However, describing the measurement as an interval-valued unit helps covering the whole averaging period, and thus yields more information about the process than the averaged measure. This quality of information gathered from the process positively impacts on the PCA model for intervalvalued data, which becomes consequently more robust to uncertainties and even to slight process variations.
In this section, we propose several fault detection strategies and fault detection indices based on PCA for interval-valued data methods, to achieve a maximum robustness toward

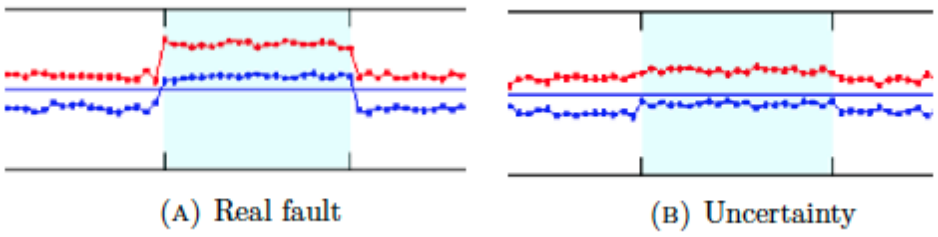
uncertainties.

3.4.1 Univariate Chart

A way to approach PCA based fault detection, is by separately analysing the residuals. First we define the interval-valued PCA estimation error, or residuals. Based upon the different methods presented above, then the interval-valued residuals $\underline{e}(k)$ and $\overline{e}(k)$, which represent the lower residual limit and the upper one respectively, they can be obtained via the difference between interval-valued estimates given in (3.22)(3.33)(3.42)(3.46) and original variables, or by projection on last $(m-\ell)$ components as:

$$\begin{cases} \underline{e}(k) = \underline{x}(k) - \overline{\hat{x}}(k) \\ \\ \overline{e}(k) = \overline{x}(k) - \underline{\hat{x}}(k) \end{cases} \tag{3.47}$$

In classical PCA, unusual events are projected onto residual space and can therefore be detected using several statistics. The same principle can be applied in the interval-valued PCA case. However, an interesting property of interval-valued residuals makes it possible to achieve proper fault detection in univariate case, i.e. without use of detection statistics, that will be discussed further.
The envelope created by residuals is a safe zone spanned by the interval of uncertainty in which every unusual event is not considered as a fault but as an uncertainty, as presented this Figure:



(A) Real fault (B) Uncertainty

Figure 3.1: Univariate Fault detection method using PCA for interval-valued data

In mathematical terms we can express the univariate interval-valued PCA detection method as follows
-Process is faulty, if :

$$\underline{e}_j(k)\overline{e}_j(k) > 0\,, \qquad j = 1,..,m \tag{3.48}$$

-Process is healthy, if:

$$\underline{e}_j(k)\overline{e}_j(k) < 0\,, \qquad j = 1,..,m \tag{3.49}$$

However, this method is restricted to processes with small number of variables, as it can be computationally cumbersome for large processes.

3.4.2 Multivariate Charts

Several statistics that measure variations in different projection sub-spaces are used in multivariate fault detection based on PCA model, T^2 and SPE (Q-statistic) are the most common fault indicators. In the case of interval valued-data, and based on an interval valued PCA model, an extension of these indicators is required in order to handle the new nature of data.

3.4.2.1 Interval-valued T^2

The decomposition of the principle explained for each interval-valued PCA model is given by :

$$[\hat{\mathbf{t}}(k)] = [[\underline{\mathbf{t}}_1(k), \overline{\mathbf{t}}_1(k)], ..., [\underline{\mathbf{t}}_\ell(k), \overline{\mathbf{t}}_\ell(k)]] \tag{3.50}$$

To calculate the Hotteling's T^2 we need the first ℓ eigenvalues, for the interval case, and for T^2 statistic eigenvalues are calculated as the corresponding empirical variances for the given bound, i.e. the variance of the interval principal components $\underline{\hat{\mathbf{t}}}(k)$ and $\overline{\hat{\mathbf{t}}}(k)$ for T^2 statistic.
For the first interval principal components, the interval $[T^2]$ statistic is calculated for the two bounds trough a combination of interval eigenvalues and interval principal components by :

$$\begin{cases} \underline{T}^2(k) = \underline{\hat{\mathbf{t}}}(k)^T \Lambda_\ell^{-1} \underline{\hat{\mathbf{t}}}(k) \\ \\ \overline{T}^2(k) = \overline{\hat{\mathbf{t}}}(k)^T \Lambda_\ell^{-1} \overline{\hat{\mathbf{t}}}(k) \end{cases} \tag{3.51}$$

3.4.2.2 Interval-valued SPE

For the interval valued case SPE statistic is calculated based on residuals as in the classical case , thus yielding an interval $[SPE]$ with an upper and a lower bound, corresponding

respectively to the upper and lower bounds of the calculated residuals as :

$$\begin{cases} \underline{SPE}(k) = \|\underline{\mathrm{e}}(k)\|^2 = \underline{\mathrm{e}}(k)^T \underline{\mathrm{e}}(k) \\ \\ \overline{SPE}(k) = \|\overline{\mathrm{e}}(k)\|^2 = \overline{\mathrm{e}}(k)^T \overline{\mathrm{e}}(k) \end{cases} \tag{3.52}$$

Given that $\underline{\mathrm{e}}(k) = [\underline{e}_1(k), ..., \underline{e}_m(k)]$, and $\overline{\mathrm{e}}(k) = [\overline{e}_1(k), ..., \overline{e}_m(k)]$, and
$[SPE](k) = \left[\underline{SPE}(k), \overline{SPE}(k)\right]$.

3.4.2.3 Thresholds for interval-valued statistics

An extension of the contro limit is done in order to compute $[T^2]$ and $[SPE]$ thresholds based on Box's quadratic form approximation, the limits for the corresponding index can be computed based on its estimate mean (a) and estimate variance (b) so that :

$$\delta_\alpha^2 = g^{index} \chi^2_{h^{index},\alpha} \tag{3.53}$$

Where $g = \frac{b}{2a}$ and $h = \frac{2a^2}{b}$, the control limits for the presented interval statistic are calculated for each bound of the statistic at hand. However, the obtained thresholds (for upper and lower bound) tend to be equal due to the symmetry of bounds, which leaves us with one control limit for the interval index.
In interval approaches PCA based fault detection FDI, a fault is detected if both bounds of the interval statistics exceed the detection threshold. However, the case where only one bound of the statistic exceeds the control threshold is not considered as a certain case of fault occurrence, so we may have some inaccuracy of the used control charts.

3.4.2.4 Combined Interval fault detection indices

Due to the interval nature of the statistics, specially when each bound yields a different decision from the other regarding the occurrence of faults. The idea of doing one indicator instead of upper and lower indices is used to avoid the ambiguity in the fault detection decision.
Since, the classical squared prediction error SPE, used as an index for fault detection based PCA approach, is a norm of residuals vector, the alternative form of this statistic norm for the interval-valued data may be achieved using the interval squared norm formula
Let us consider a new interval measurement vector given by:

$$[\mathrm{x}(k)] = [[\underline{x}_1(k), \overline{x}_1(k)]\, [\underline{x}_2(k), \overline{x}_2(k)] \ldots [\underline{x}_m(k), \overline{x}_m(k)]] \tag{3.54}$$

The proposed interval fault detection index is defined as the interval squared prediction error (ISPE) and is given by:

$$ISPE(k) = \|[\mathrm{e}(k)]\|^2 = \sum_{j=1}^{m} \|[e_j(k)]\|^2 \tag{3.55}$$

Where

$$\|[e_j(k)]\|^2 = \frac{1}{3}\left(\underline{e}_j^2(k) + \underline{e}_j(k)\overline{e}_j(k) + \overline{e}_j^2(k)\right) \tag{3.56}$$

Applying the same principle we calculate the interval Hotelling T^2, the statistic $[IT^2]$ is found as :

$$IT^2(k) = \left\| \frac{[\hat{\mathbf{t}}(k)]}{[\Lambda_\ell]^{1/2}} \right\|^2 \tag{3.57}$$

Where $[\hat{\mathbf{t}}(k)]$ is the first ℓ interval-valued components from data projection using interval PCA model, with $[\Lambda_\ell]$ being the corresponding eigenvalues which are calculated for the $[T^2]$ statistic.
Furthermore a new statistical control thresholds are computed based on Box's formula. Hence, the example of the confidence limit δ_α^2 for the $ISPE$ can be found from its approximate distribution as :

$$\delta_\alpha^2 = g\chi_{h,\alpha}^2 \tag{3.58}$$

Where g is the weighting parameter included to account for the magnitude of the $ISPE$, and h accounts for the degrees of freedom with a significance level of $1-\alpha$, typically selected to be 95% to 99%.
Accordingly, the control limits of the IT^2 is calculated similarly, based on Box's formula, for the corresponding degrees of freedom and significance level.

3.4.2.5 The interval new indicator $I\Phi$

This statistic also based on the Combined Interval fault detection indices IT^2 & $ISPE$ as follows :

$$I\Phi(k) = \frac{IT^2(k)}{\tau_\alpha^2} + \frac{ISPE(k)}{\delta_\alpha^2} \tag{3.59}$$

3.5 Simulation example

In order to demonstrate the use of the interval-valued PCA in fault detection, let us consider the previous simulation example based on 6 variables $j = 1, ..., 6$, let us also

recall that this data are based on two variables generated from two normal distributions, and four linear analytic redundancy relations between the remaining variables as follows:

$$\begin{cases} x_1(k) = 0.5v_1(k) - 1.3sin(k/N)cos(k/3) \\ x_2(k) = 0.9v_2(k) - 1.2cos(k/4)^3 e^{(-k/2N)} \\ x_3(k) = x_1(k) + x_2(k) \\ x_4(k) = 2x_1(k) + x_3(k) \\ x_5(k) = x_2(k) + x_3(k) \\ x_6(k) = 2x_1(k) + x_2(k) \end{cases}$$

To simulate the presence of measurement uncertainties, one of the two methods explained in (3.2) is used to generate the interval valued-data. To avoid the problem of the window length in the moving window method, we simply used the error measurement to present the data as an interval value.

For this example a variation δX was added to the generated data. This variation is a realization of centered variables corresponding to 10% of the variation range of each variable. Thus, we obtain the new interval data matrix $[X] = [X - \delta X\,,\, X + \delta X]$ The interval-valued variables of data is presented as :

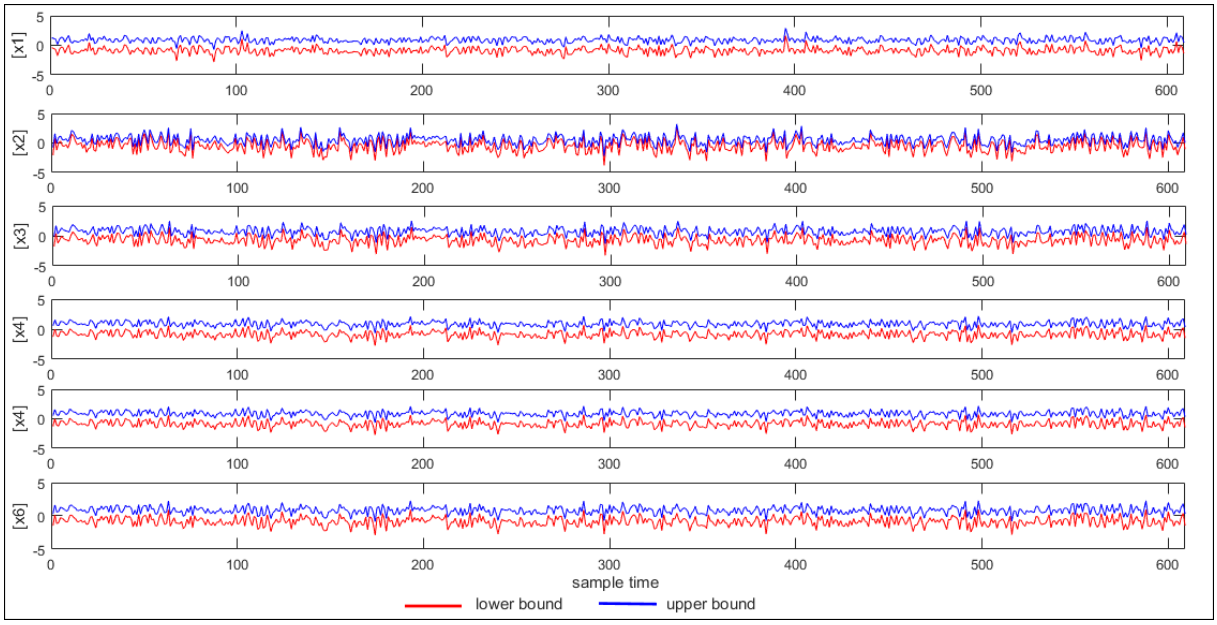


Figure 3.2: Generated Interval-valued variables

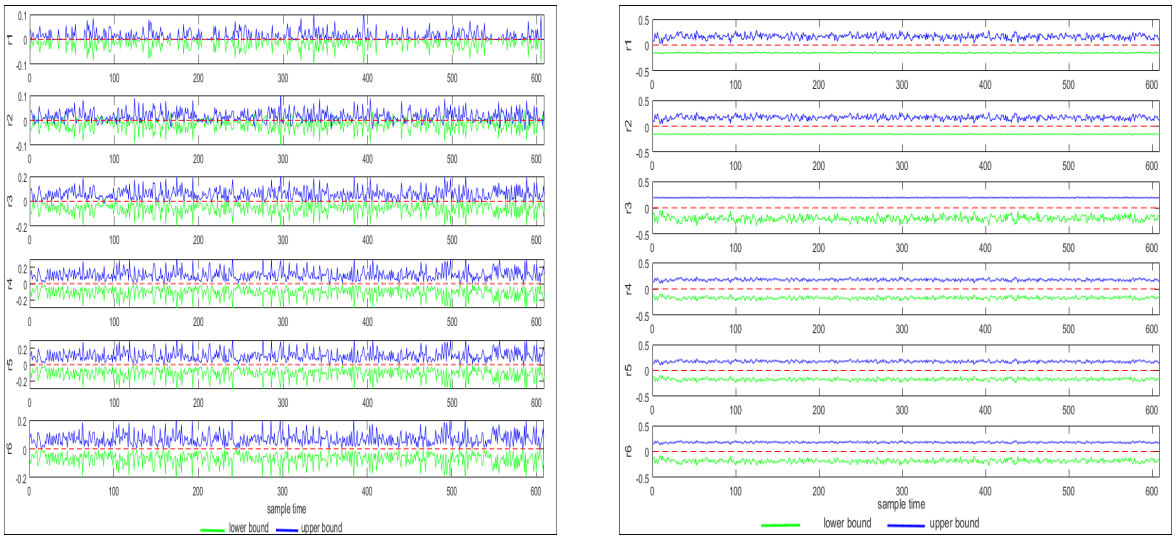
The first step is to center and standardize the data to zero mean and unit variance, according to section (3.2.4) there are four different methods to normalise the data, in this example each method is applied to see the differences, the shown figures results from standardizing the data using the global range normalisation method.

The eigen-decomposition of the covariance matrix of $[X]$ allows the determination of its

eigenvalues and eigenvectors, i.e. the interval-valued PCA model according to the various methods described previously. The number ℓ of components to be retained for the interval-valued PCA models are determined using the Cumulative Percentage of Variance (CPV), where ℓ is chosen as the value that makes $CPV = 90\%$, accordingly, $\ell = 2$ principal axes are kept in the interval PCA model.

Since we are using the the PCA for interval-valued model, let us first inspect the structure or nature of residuals. figure (3.3) represent the interval-valued residuals obtained through an interval-valued PCA model. We note that the upper and lower bounds of residuals which are, respectively, residuals of upper and lower estimations of interval data matrix, are forming an envelope around the zero line.

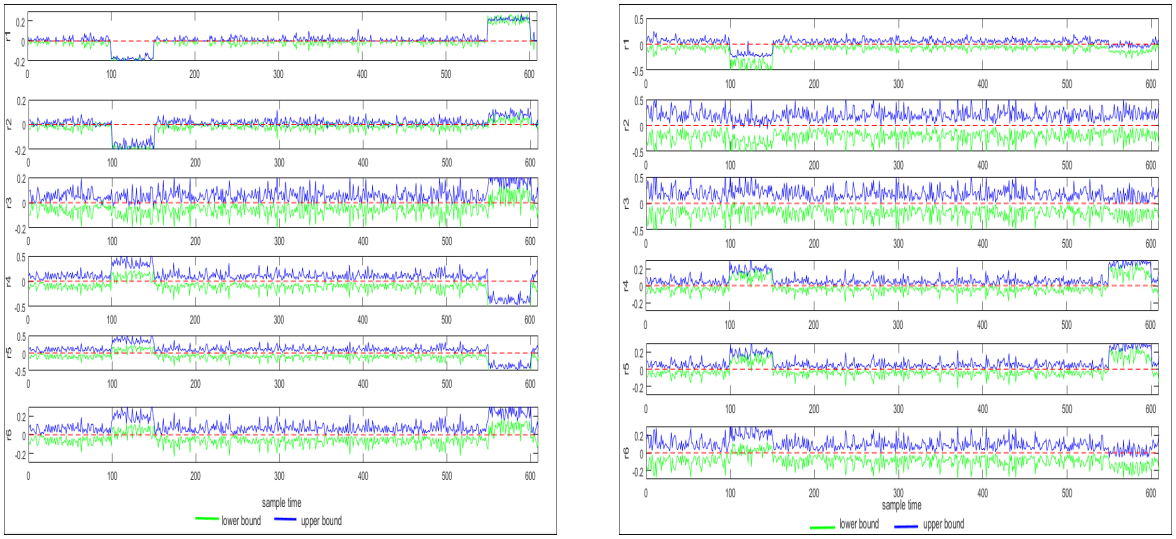
Remark: *In this section the results obtained is by applying only two methods of PCA for interval valued-data, which are the Centers PCA (CPCA), and Midpoints-Radii PCA (MRPCA) using the global range normalisation method, for fault detection instead of showing the upper and lower cases for the T^2, and SPE indicators which may cause an ambiguity in decision, the combined interval indicators are applied IT^2 and $ISPE$, also the new statistic $I\Phi$ is used to detect the fault.*



(a) (b)

Figure 3.3: Healthy interval-valued residuals using (a) CPCA and (b) MRPCA

The system of this example is simulated a second time on 609 samples, then an offset is added to the data to variables $x_4(k)$, and $x_5(k)$ from time sample 550 to 600, also an offset is added to $x_1(k)$ from time sample 100 to 150, simulated as a constant bias of amplitude equal to 5% of variation range of the changed variables. For the univariate fault detection case, an abnormal behavior cannot be considered as a fault unless one of the residuals bounds changes sign. We can clearly notice the behavior of the residuals in Figure (3.4), which represents the interval-valued residuals in faulty case as follows :



(a) (b)

Figure 3.4: Faulty interval-valued residuals for (a) CPCA and (b) MRPCA

From the figures a good results for the MRPCA, where in the healthy zone the residual envelope slightly changes in different directions (positive or negative) in sample area 550 to 600. In other words, both bounds of the residual shift in a direction, but do not however exceed the middle zero line (no sign change in residuals). In the fault occurrence zone, we note the sign change of the lower and upper bounds in residual $e_4(k)$ & $e_5(k)$ which confirms a faulty condition. Same result in the sample area 100 to 150. The other residuals are affected because they are linearly related to variable x_1. Whereas using the CPCA, it can be seen that in the faulty case, the other residuals are affected with a sign change, which make it difficult to know the faulty variables.

The multivariate charts of upper and lower may show an unclearness in decision of fault as shown in figure (3.6), where ϕ is used to detect the added fault of 5% from sample 100 to 300.

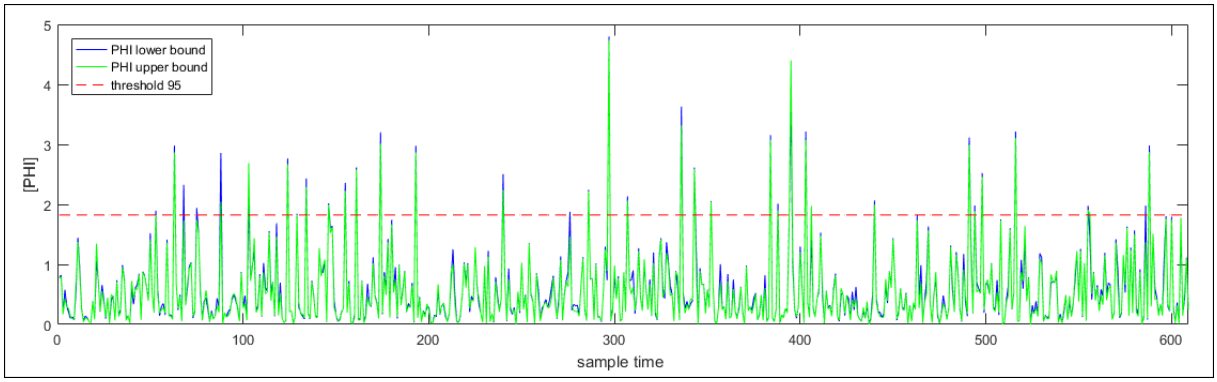


Figure 3.5: Time evolution of [ϕ] index in healthy case

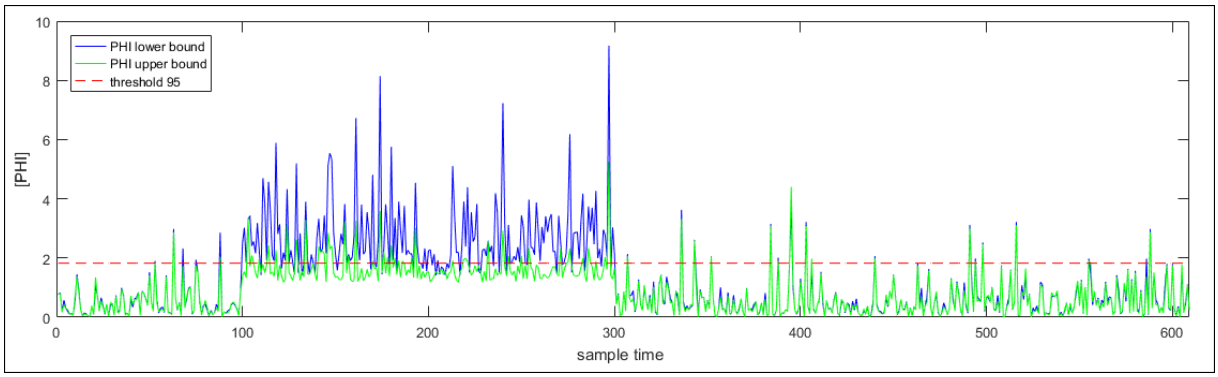
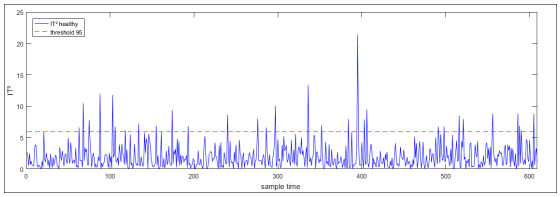


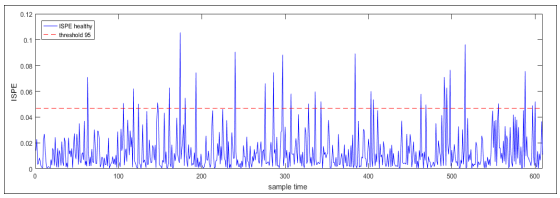
Figure 3.6: Time evolution of [ϕ] index in faulty case

It can be seen that only one bound exceeds the control threshold, this scenario presents an ambiguous decision. Thus, the detection rate of these interval statistics drastically decreases, once we consider this detection condition based on both bounds of the control statistic.

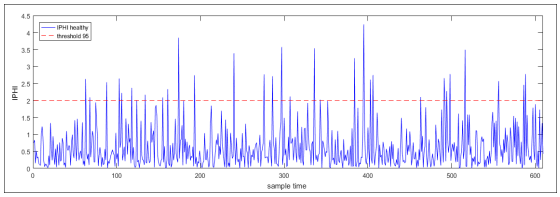
Using the multivariate charts (IT^2, $ISPE$, $I\Phi$) for fault detection instead of upper and lower cases gives a more clear results by adding the same simulated fault.



(a)

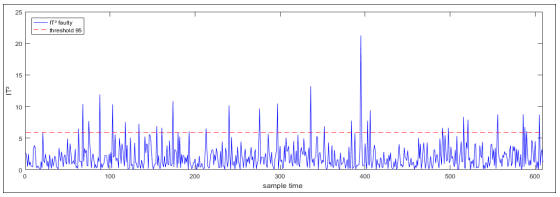


(b)

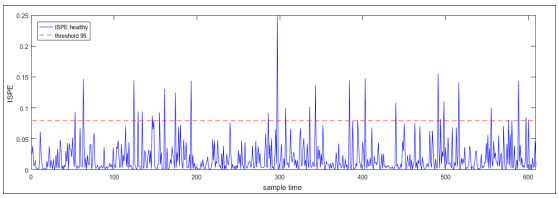


(c)

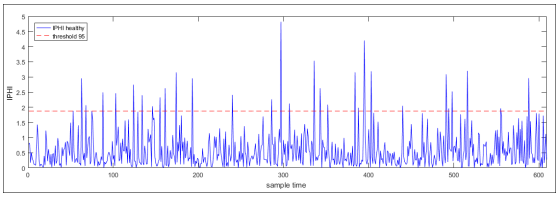
Figure 3.7: CPCA Results of healthy data using (a)IT^2, (b)$ISPE$, and (c)$I\Phi$ statistics.



(a)



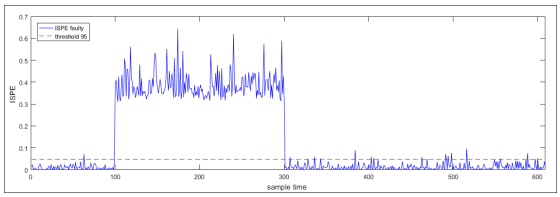
(b)



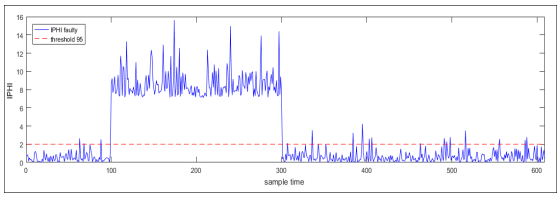
(c)

Figure 3.8: MRPCA Results of healthy data using (a)IT^2, (b)$ISPE$, and (c)$I\Phi$ statistics.

An offset of 5% is added to the data to variable $x_4(k)$ from time sample 100 to 400, because the Hotelling's IT^2 can not detect small faults, the results are shown below.

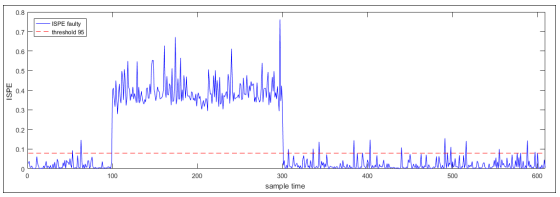


(a)

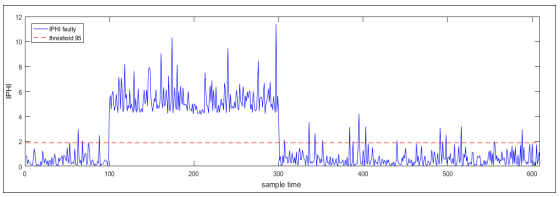


(b)

Figure 3.9: CPCA Results of faulty data using (a)$ISPE$, and (b)$I\Phi$ statistics.



(a)



(b)

Figure 3.10: MRPCA Results of faulty data using (a)$ISPE$, and (b)$I\Phi$ statistics.

Using both $ISPE$ & $I\Phi$ indicators show an improvement compared to SPE & ϕ indices in terms of fault detection and false alarms, i.e. enhancement of fault detection and reduction of false alarms. Where the fault clearly exceeds the control limit, with a small alarm rate registered. Both indices T^2 and IT^2 failed to detect the fault in this case. However, IT^2 outperforms the interval T^2 index in the case of high amplitude faults usually detected by the Hotelling's statistic.

Table 3.1: FAR % contributed by $ISPE$, and $I\Phi$

			FAR[%]	
Fault	Statistic	Confidence level [%]	CPCA	MRPCA
Abrupt	$ISPE$	95	3.6	2.9
		98	1.8	1.5
		99	0.9	0.2
	$I\Phi$	95	0.7	1.2
		98	0.5	0.5
		99	0.2	0.1
Drift	$ISPE$	95	4	3
		98	1	1.2
		99	0.3	0.7
	$I\Phi$	95	0.8	1
		98	0.1	0.3
		99	0	0
Intermittet	$ISPE$	95	16	14
		98	10	9
		99	4	4
	$I\Phi$	95	11	10
		98	5.9	6.6
		99	3	2.4

Table (3.1) shows an improvement in results of false alarm rate compared to classical PCA, where CPCA & MRPCA methods almost detect all the uncertainties except for the intermittent fault, the results also confirms the superiority of the new $I\Phi$ indicator over the other indices.

Table 3.2: MDR & DTD contributed by $ISPE$, and $I\Phi$ using CPCA & MRPCA

			MDR[%]		DTD[s]	
Fault	Statistic	Confidence level [%]	CPCA	MRPCA	CPCA	MRPCA
Abrupt	$ISPE$	95	0.2	0.5	00	00
		98	00	00	00	00
		99	00	00	00	00
	$I\Phi$	95	0.3	00	00	00
		98	0.1	00	00	00
		99	00	00	00	00
Drift	$ISPE$	95	4.65	3.82	30	24
		98	3.90	3.14	28	21
		99	2.16	2.69	20	18
	$I\Phi$	95	3.51	4.23	27	20
		98	3.22	3.11	19	16
		99	1.54	1.22	9	7
Intermittet	$ISPE$	95	0.71	1.13	2	1
		98	0.2	0.1	0.32	0.14
		99	00	00	00	00
	$I\Phi$	95	00	00	00	00
		98	00	00	00	00
		99	00	00	00	00

Table (3.2) shows a remarkable change in the miss detection rate and detection delay using the interval indices, and it can be seen that the $I\Phi$ indicator has detected the fault in time with a negligible delay, which proves again it's efficiency.

3.6 Conclusion

Using PCA for interval valued-data is a novel technique that emphasizes on uncertainties of measurement. The interval nature of the projections ensures the elimination of uncertainties during the fault detection procedure. Undoubtedly, taking benefits of any knowledge about uncertainties is one of the fundamental points of current research and development in multivariate statistical process monitoring. Hence, several interval-valued PCA models have been developed for more efficiency in analyzing the huge volume of data continuously emerging from various computerized industries.

In this chapter, The most known methods of the Interval valued-data were presented being VPCA, CPCA, MRPCA, and CIPCA approaches. Within this work, a new interval-valued statistics were presented to be used with interval-valued PCA along with their threshold calculation, which includes the combined indicators. The different indices are illustrated using CPCA, and MRPCA in a simulation example, and proven efficient in detecting faults in uncertain processes. However, it is necessary to view and compare the performances of such interval-valued models and statistics in real processes, which will be discussed in the next chapter.

Chapter 4

APPLICATION FOR CEMENT ROTARY KILN DATA

4.1 Determination of the number of retained components

The description of the cement rotary kiln data is explained, and detailed in appendix A. The first step in applying PCA and creating its model is to find the number of retained principal components. To do that, several methods were applied to the data set in order to pick the best and the most descriptive number of components to successfully separate the informative parts of the signal from the noise. The following table provides a summary for most of the applied methods.:

Table 4.1: Summary of the different methods used to select the number of retained principal components

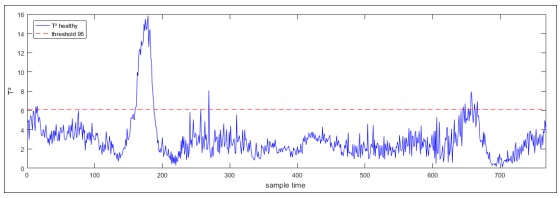
Method	*Number of retained PCs*	*Cumulative variance* [%]
Kaiser	11	75
CPV	8	60.35
	10	72.10
	13	90.94
	26	96.80
Bootstrap	14	72.97
MAP	4	43.67
Broken-Stick	7	57.71
Parallel analysis	11	67.11
PRESS	33	96.28

The results given by the different methods were widely spread, this phenomenon can be

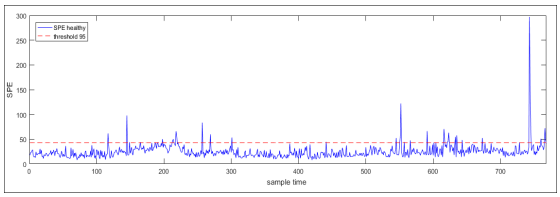
explained on the scope of the used data. Starting from Kaiser rule, the restriction of the retained components causes a great loss in the retained variance, which is supposed to be a measure for how much the dimension is informative. For BS, the method works on the basis of the random breaking under the assumption that the data to be tested against is random. In our data set, a high persistency was observed in the measurements of different sensors, which violates the assumption of randomness. MAP resulted in the most inferior amount of retained variance where more than half of the variance carried by the dimensions is considered as noise. In the other side, Cross Validation and CPV are apparently showing good behavior. For the CPV, all the chosen number of principal components (8 to 26) shows a small loss in the retained variance. The same thing applied to the cross-validation criteria where it results in a high percentage of variance.
Finally, in order to formalize the decision about the number of retained components, the number of components is varied from 8 to 33. In our work the method used to select the number of principal components is the Cumulative Percent of Variance (CPV) for not less than 90% of the variation, as a result, the retained number of components to be work with is chosen to be $\ell = 11$.

4.2 PCA based fault detection application for Single valued-data

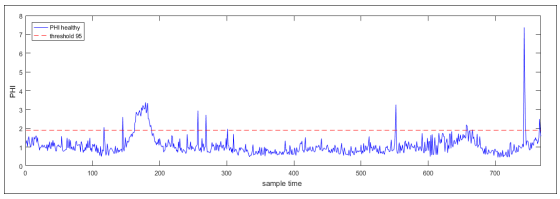
At the beginning, the training data and is constructed with a fixed threshold of 95% confidence level, using the T^2 & SPE statistics, and the ϕ statistic. Then these indices are used to monitor the real process fault.



(a)

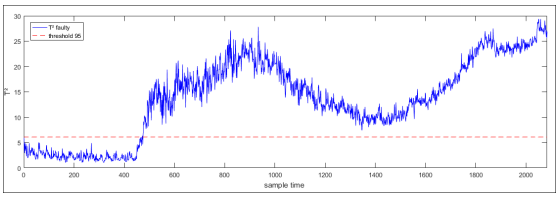


(b)

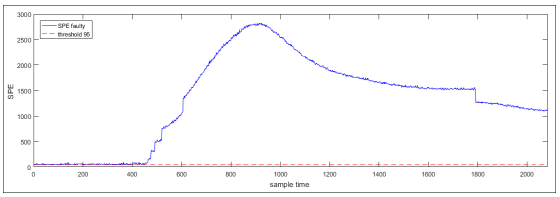


(c)

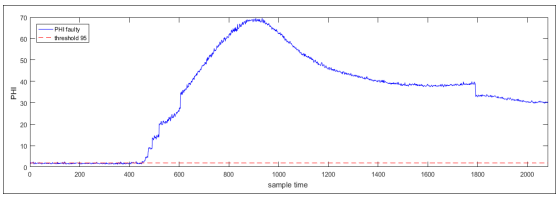
Figure 4.1: Results of normal process operation using (a)T^2, (b)SPE, and (c)ϕ statistics.



(a)



(b)



(c)

Figure 4.2: Results of real process fault using (a) T^2, (b) SPE, and (c) ϕ statistics.

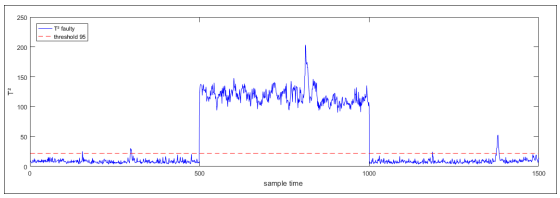
Table 4.2: FAR % contributed by T^2, SPE, ϕ using (95%, 98%, 99%) control limits.

Data set	Training set			Testing set			Faulty set (first 450 samples)		
Statistic	T^2	Q	ϕ	T^2	Q	ϕ	T^2	Q	ϕ
95_{CL}	5.44	4.19	4.75	4.33	6.12	4.98	18.59	13.41	12.23
98_{CL}	3.87	3.28	3.42	2.91	3.87	3.13	15.62	11.86	9.28
99_{CL}	2.86	2.34	2.53	1.52	2.49	1.76	10.63	8.97	6.52

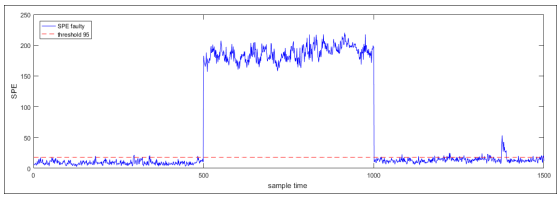
Table (4.2) lists FAR under the healthy operation of a cement plant. It has been evaluated for the training, testing and faulty data sets, and contributed by the SPE & T^2, and ϕ statistics using different fixed threshold percentages. FAR is the portion of faulty samples from the first 450 samples, where the process was ensured to be under healthy operation status.

It can be seen that a low FAR is obtained for the rotary kiln using the new indicator ϕ especially in the faulty data where it has a small value compared to the other statistics. Besides, the more control level used the smallest FAR registered. The indices clearly indicate few false alarms throughout 3000 samples. These false alarms are due to the transient state during a change in the process dynamics, where some set-points are changed by the operators.

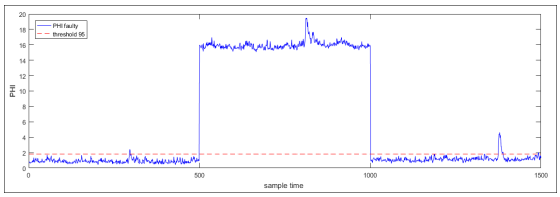
The simulated sensor faults that consist of abrupt (single), and drift (1) types are shown below using the same indices.



(a)

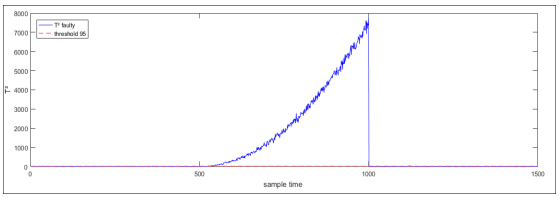


(b)

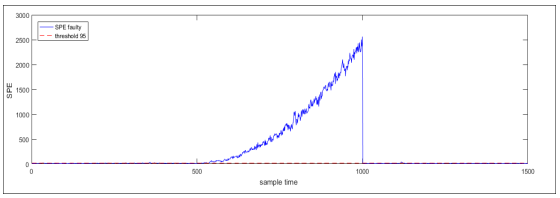


(c)

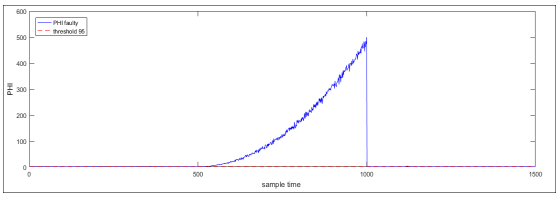
Figure 4.3: Results of abrupt (single) sensor fault using (a) T^2, (b) SPE, and (c) ϕ statistics.



(a)



(b)



(c)

Figure 4.4: Results of drift (single) sensor fault using (a) T^2, (b) SPE, and (c) ϕ statistics.

The Q & T^2, and ϕ statistics results of the detection of sensors faults are illustrated in table (4.3) showing the difference between the indices at a different control levels.

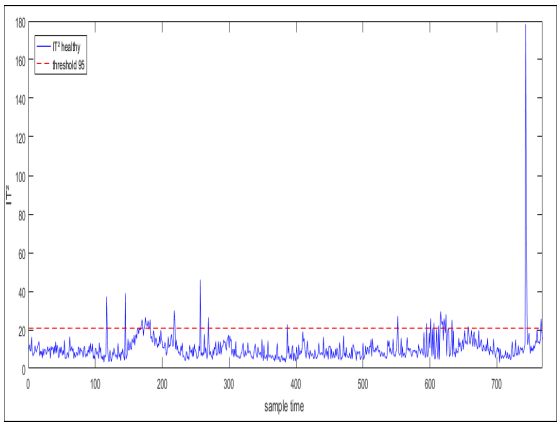
Table 4.3: Simulating sensor fault detection.

Fault	Statistic	Confidence level [%]	Abrupt	Drift 1	Drift 2	Random	Abrupt (mul)	Drift (mul)	Intermittent
MDR [%]	T^2	95	2.65	9.43	25.42	7.11	3.15	28.4	5.31
		98	1.23	8.56	23.33	6.23	2.98	28	5
		99	0.81	8.11	23.05	5.12	2.32	27	4.6
	Q	95	1.03	11.12	7.70	0.49	0.90	21	0.56
		98	0.78	10.08	7.32	0.32	0.00	20.12	0.00
		99	0.55	9.00	7.12	0.21	0.00	20.03	0.00
	ϕ	95	0.83	9.88	7.01	3.21	0.95	20	0.32
		98	0.49	9.31	6.50	2.98	0.00	19.13	0.00
		99	0.41	9.12	6.22	2.87	0.00	19.6	0.00
DTD [s]	T^2	95	4.91	46	130	0.00	10.24	145	1.3
		98	4.16	42	118	0.00	10.12	139	1
		99	3.22	40	110	0.00	10	131	1
	Q	95	3.44	46	50	0.00	0.67	102	0.00
		98	2.67	44	43	0.00	0.5	99	0.00
		99	2.56	43	36	0.00	0.00	98	0.00
	ϕ	95	3.78	42	42	0.00	0.71	110	0.00
		98	3.12	41	37	0.00	0.00	103	0.00
		99	2.34	41	34	0.00	0.00	100	0.00

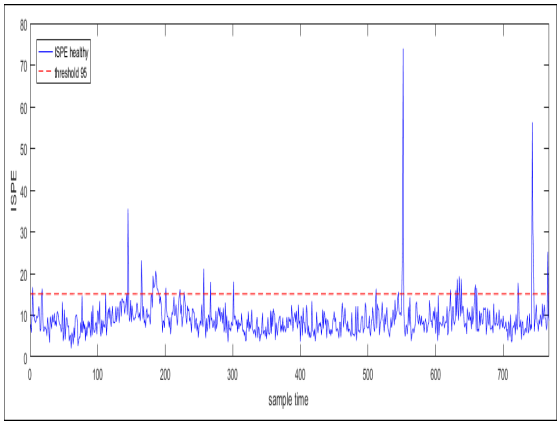
Figure (4.3) (4.4) show the obtained three monitoring indices through the occurrence of two different simulated faults. As it can clearly be seen from table (4.3), the random fault and abrupt faults in a single sensor or multiple sensors have been promptly detected regardless of their small amplitudes. However, the detection time of the drift faults in a single sensor, or in multiple sensors was slightly delayed, which means that the faults are not detected until they reach a certain amplitude. The intermittent fault is accurately detected via the three statistics, where the new ϕ statistic detected it instantaneously which proves its effectiveness.

4.3 PCA based fault detection application for Interval-valued data

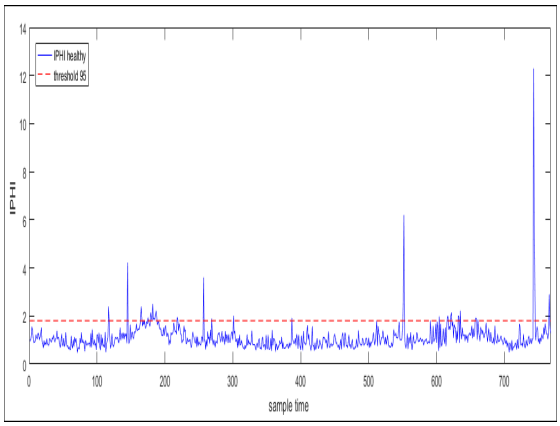
In the interval value data case, each measurement of the data sets used in this application is presented as interval-value instead of a single one. For that, a variation δX of 10%, as the variation range of each variable, has been added to obtain the interval data.
After recording the data measurements of the cement plant during the healthy state, a PCA model has been constructed that will be used to monitor the status of the cement process. On a similar way, a model of each method of interval-valued PCA (CPCA, MR-PCA, VPCA, CIPCA) has been built from the generated interval-valued data.
Remark: *In this application, the data has been normalized using the four different methods; the plotted graphs are the result of the best normalization method. In almost all cases, the global centers normalisation method was the best choice.*
In order to do a comparison between the four presented IVD-PCA methods, namely, VPCA, CPCA, MRPCA and CIPCA, the three statistical indicators (IT^2, $ISPE$, and $I\Phi$) have been computed. the results are presented on the following graphs for normal (healthy) process operation, and real process fault.



(a)

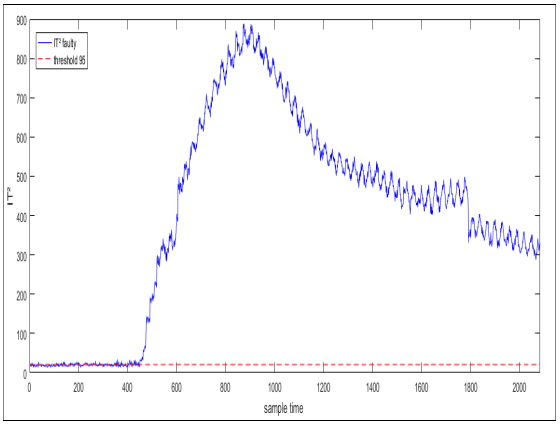


(b)

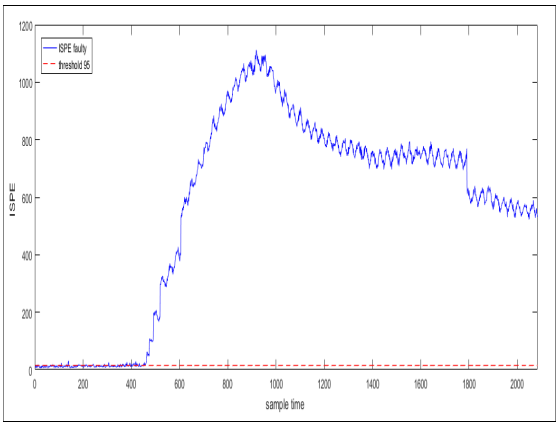


(c)

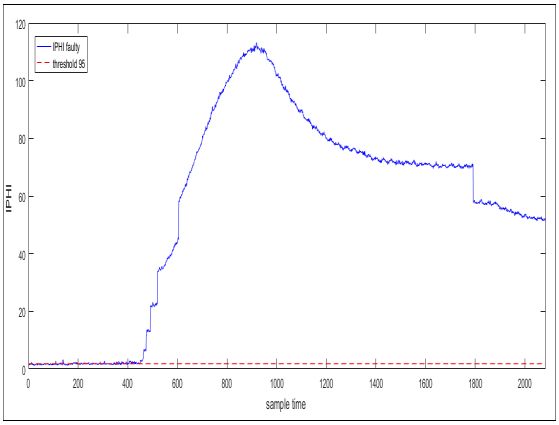
Figure 4.5: Results of the MRPCA interval-valued based method tested on the healthy data using (a) IT^2, (b) $ISPE$, and (c) $I\Phi$ statistics.



(a)

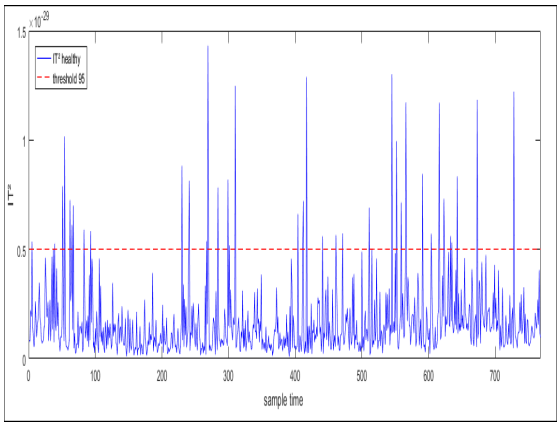


(b)

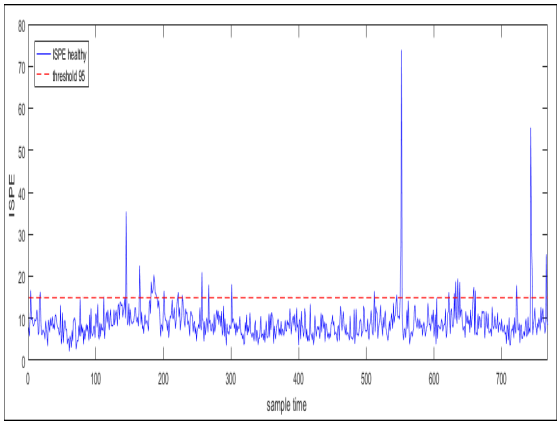


(c)

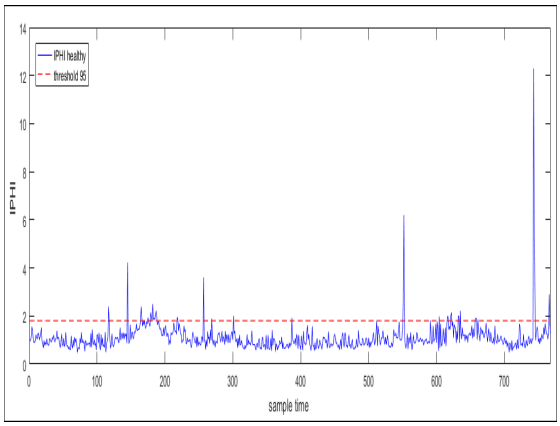
Figure 4.6: MRPCA interval-valued based method in detecting the real fault using (a)IT^2, (b)$ISPE$, and (c)$I\Phi$ statistics.



(a)

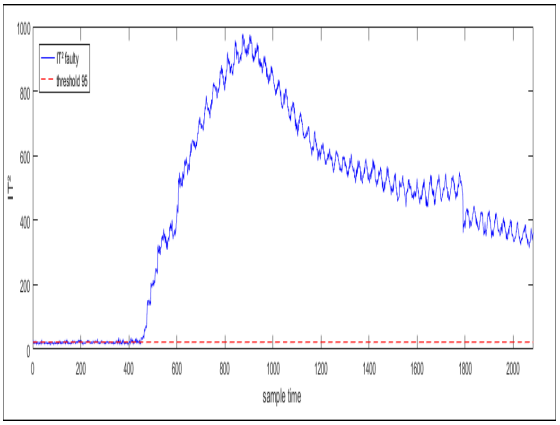


(b)

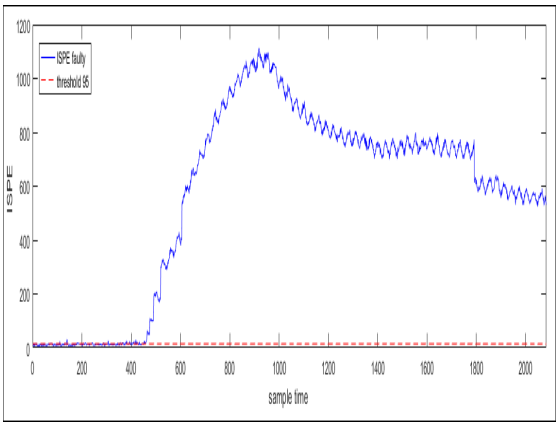


(c)

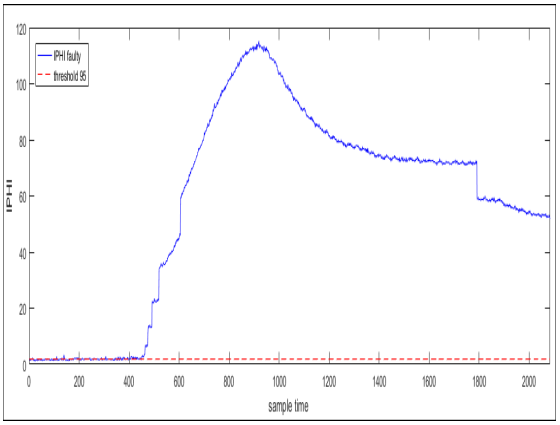
Figure 4.7: Results of the CPCA interval-valued based method tested on the healthy data using (a) IT^2, (b) $ISPE$, and (c) $I\Phi$ statistics.



(a)

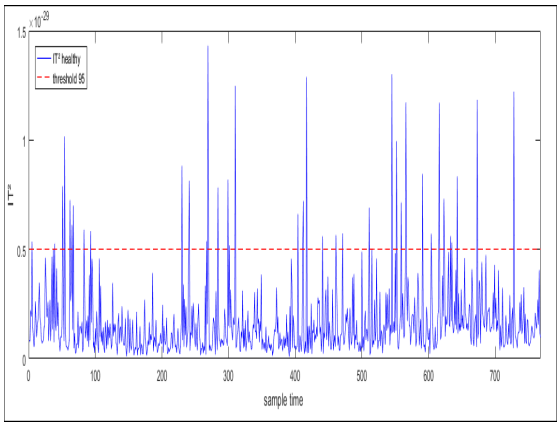


(b)

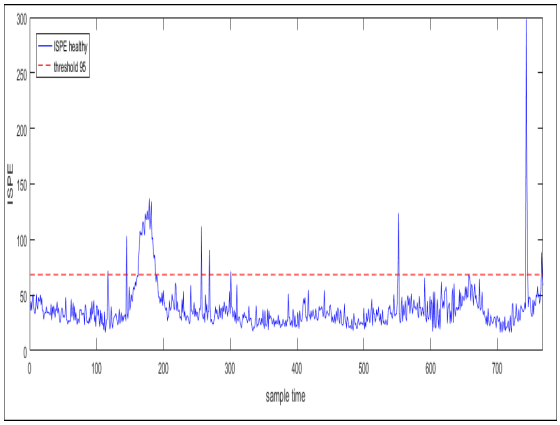


(c)

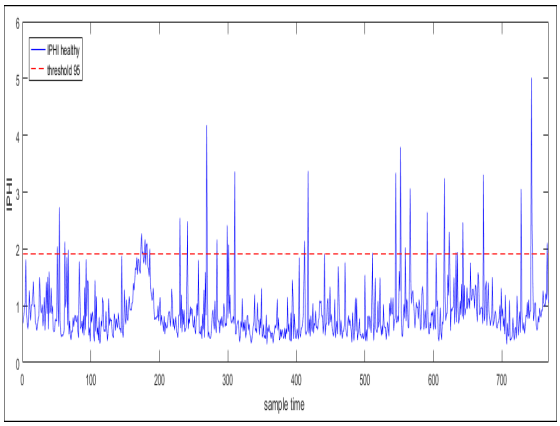
Figure 4.8: CPCA interval-valued based method in detecting the real fault using (a)IT^2, (b)$ISPE$, and (c)$I\Phi$ statistics.



(a)

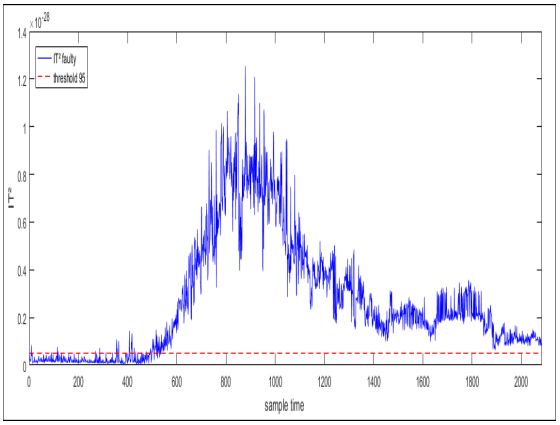


(b)

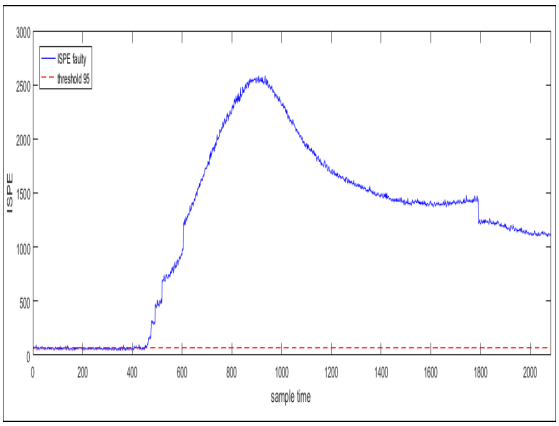


(c)

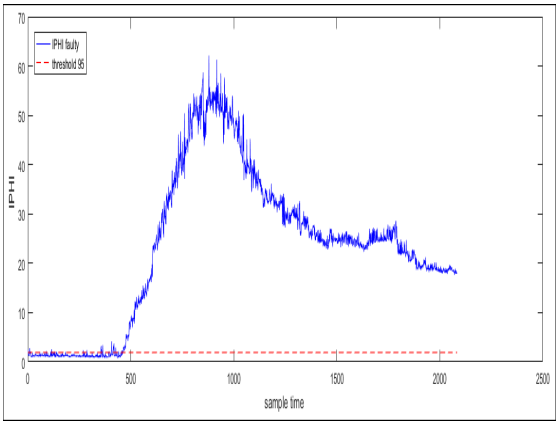
Figure 4.9: Results of the CIPCA interval-valued based method tested on the healthy data using (a) IT^2, (b) $ISPE$, and (c) $I\Phi$ statistics.



(a)

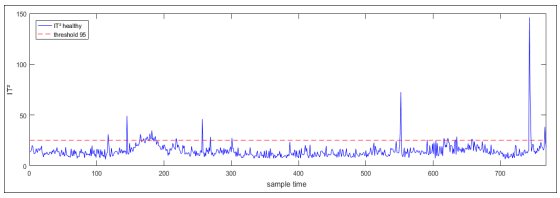


(b)

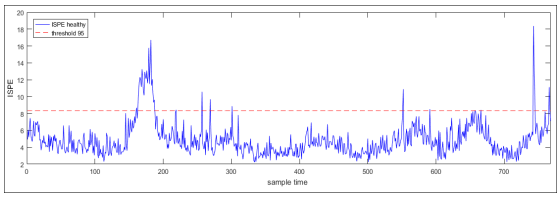


(c)

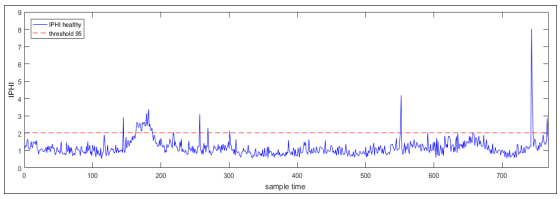
Figure 4.10: CIPCA interval-valued based method in detecting the real fault using (a)IT^2, (b)$ISPE$, and (c)$I\Phi$ statistics.



(a)

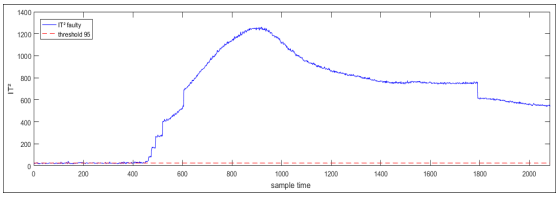


(b)

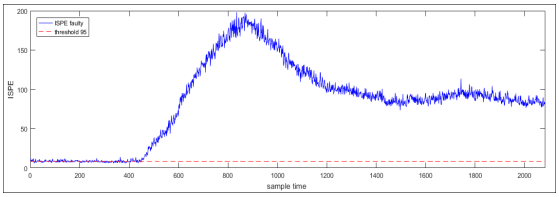


(c)

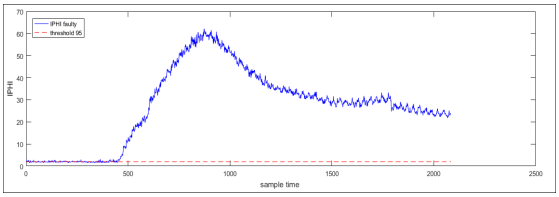
Figure 4.11: Results of the VPCA interval-valued based method tested on the healthy data using (a) IT^2, (b) $ISPE$, and (c) $I\Phi$ statistics.



(a)



(b)



(c)

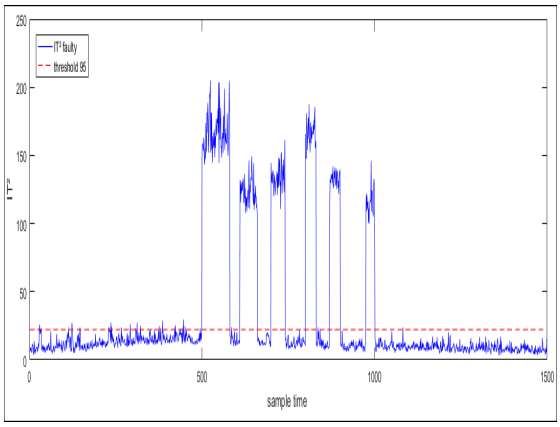
Figure 4.12: VPCA interval-valued based method in detecting the real fault using (a)IT^2, (b)$ISPE$, and (c)$I\Phi$ statistics.

The graphs show that the indices did correctly detect the real fault With a relative superiority of the $I\Phi$ indicator. The following table represent the FAR% under healthy operation of the cement plant for the testing, training, and the first 450 healthy samples of the faulty data, using a 95% confidence level.

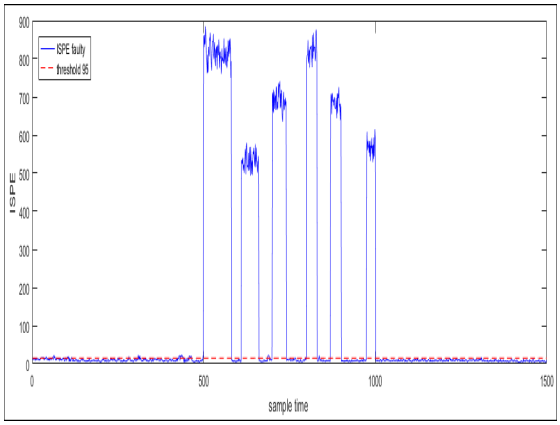
Table 4.4: FAR% contributed by IT^2, $ISPE$, $I\Phi$ using different IVD-PCA

	IT^2 /$ISPE$ /$I\Phi$		
PCA	Training set	Testing set	Faulty set
CPCA	2.5 / 1.8 / 1.1	2.2 / 3.4 / 2.5	3 / 0.6 / 0.4
MRPCA	1.9 / 1.3 / 0.4	1.4 / 3 / 1	2.3 / 0.3 / 0
VPCA	1.8 / 0.7 / 0.1	2 / 2.5 / 2.2	2.8 / 0.9 / 0.8
CIPCA	1 / 0 / 0	1.9 / 0.7 / 0.4	1.1 / 0.6 / 0

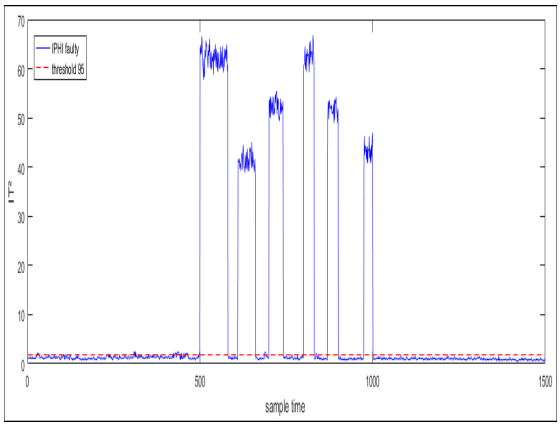
The obtained results prove the effectiveness of the IVD-PCA for uncertainties, where a big reduction has been remarked compared to classical PCA and using for 95% confidence level. Using CIPCA and MRPCA for example, a 0.2% of FAR has been recorded compared to a whopping 6.5% of the classical PCA method.



(a)

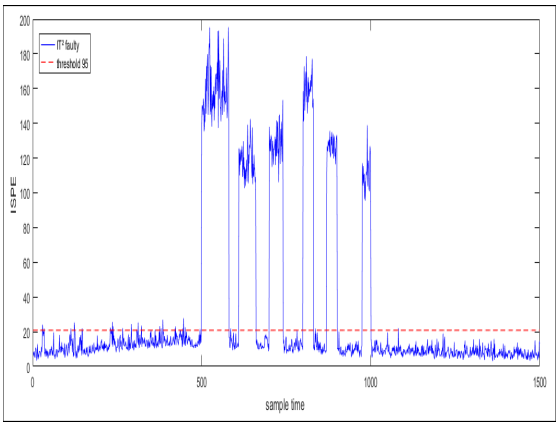


(b)

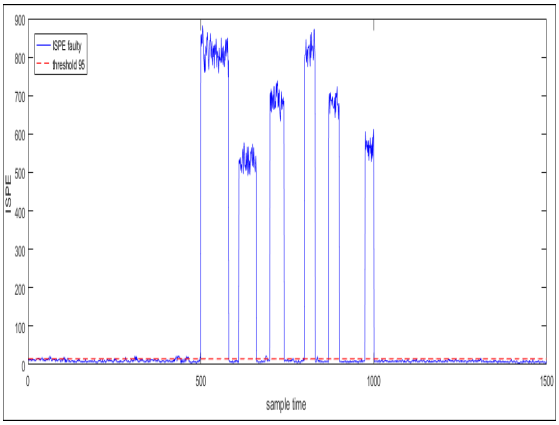


(c)

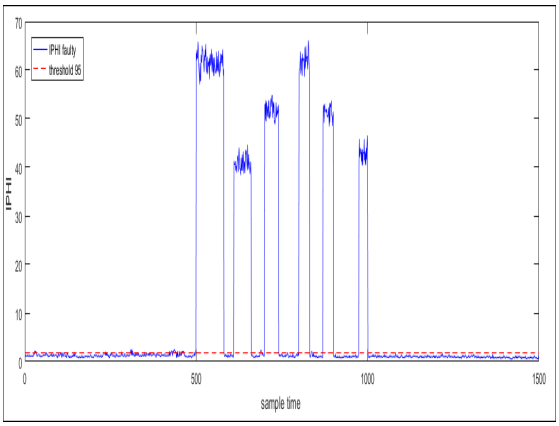
Figure 4.13: CPCA interval-valued based method in detecting intermittent fault using (a)IT^2, (b)$ISPE$, and (c)$I\Phi$ statistics.



(a)

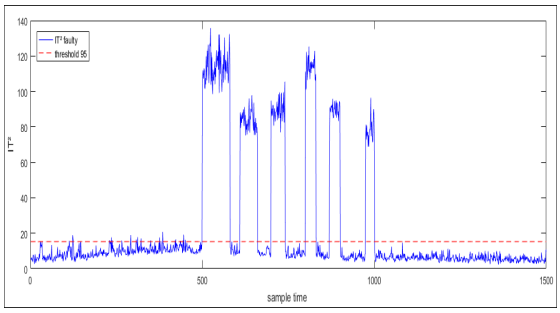


(b)

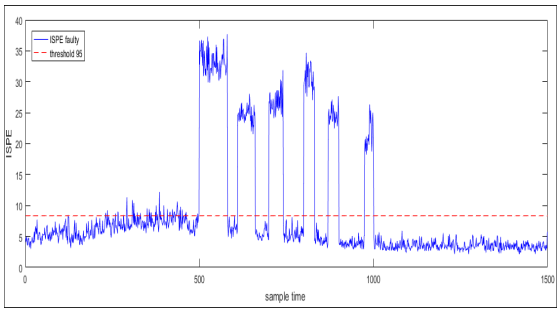


(c)

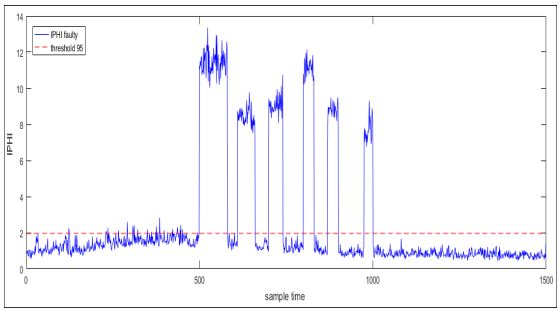
Figure 4.14: MRPCA interval-valued based method in detecting intermittent fault using (a)IT^2, (b)$ISPE$, and (c)$I\Phi$ statistics.



(a)

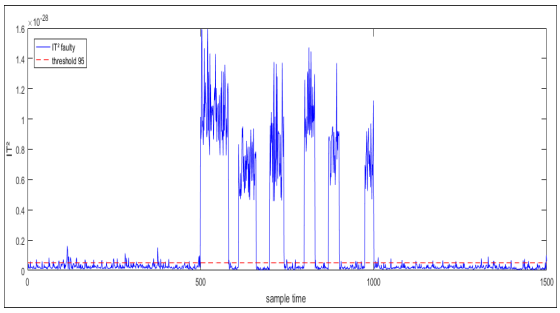


(b)

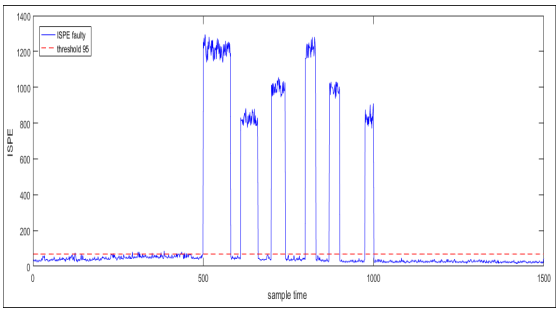


(c)

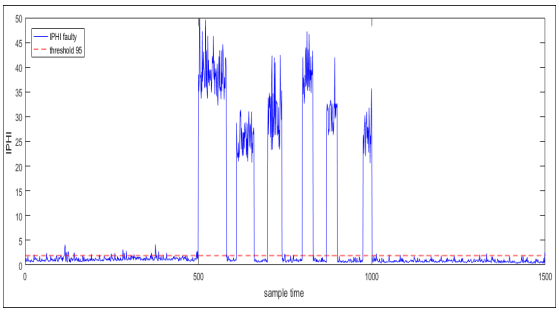
Figure 4.15: VPCA interval-valued based method in detecting intermittent fault using (a)IT^2, (b)$ISPE$, and (c)$I\Phi$ statistics.



(a)



(b)



(c)

Figure 4.16: CIPCA interval-valued based method in detecting intermittent fault using (a)IT^2, (b)$ISPE$, and (c)$I\Phi$ statistics.

The results of MDR[%], and DTD[s] of the simulated faults (abrupt, intermittent, drift, and random), using the different interval valued PCA are organized in the following table.

Table 4.5: MDR [%], and DTD[s] contributed by IT^2, $ISPE$, $I\Phi$ using different IVD-PCA

		IT^2 /$ISPE$/$I\Phi$						
	PCA	abrupt	drift 1	drift 2	random	abrupt (multiple)	drift (multiple)	intermittent
MDR [%]	CPCA	2/ 0.8/0.5	2.1/3.8 /1.6	5/ 3.2/1.4	2.1/0.6 /0	1.7/0.6/0.3	9/7.1/6.5	0.3/0/0
	MRPCA	1.2/ 0.4/0.1	2/2.3 /1.9	3/1.3/0	1.6/1/0	0.8/0/0	5/3.1 /8	0/0/0
	VPVA	1/ 0.9/0.5	4.11/3.3/2	11/7/8.4	3.9/2.3/2	2.3/2.1/1.6	20/14.2/11	1.4/0.4/0
	CIPCA	0.8/0.4 /0.2	11.12/1.1/0.4	7.70/0.6/0	0.49/0/0	0.90/0/0	3.3/2.7/2	0/0/0
DTD [s]	CPCA	2.6/1.7 /1.3	8/5/1	47/35 /11	0.4/0/0	1.8/0/0	12/10/8	0/0/0
	MRPCA	1.2/0.8/0.5	10/4/0	23/19/4	0/0/0	0.7/0 /0	10/5/4.2	0/0/0
	VPCA	3.1/2.2 /2	15/17/12	17/16/19	1.3/1/0.9	5/3.6/2	55/47/40	0.9/0.3/0.2
	CIPCA	1.9/0.6/0	46/0/0	15/2/1	0/0/0	0.47/0/0	6/4/3	0/0/0

The result of table (4.5) proves once again the efficiency of IVD-PCA over the classical PCA in detecting the simulated sensor faults, where more than 90% of MDR were eliminated using the CIPCA, which indicates it's superiority over the other interval valued-PCA.

4.4 Conclusion

In this section, the proposed strategy for fault detection using PCA for interval-valued data has been applied for process monitoring using a real process data, we have introduced different interval statistics based on interval-valued PCA model. A comparison between the four most known interval-valued PCA methods is performed, namely VPCA, CPCA, MRPCA and CIPCA. The application demonstrates the good performances obtained using PCA for interval-valued data for fault detection. Especially using the new indicator ϕ, where the smallest FAR, and MDR percentages were obtained using this statistic.

Conclusion and Perspectives

This thesis presented a new sensor fault detection strategy for systems subject to uncertainties, which is based on principal component analysis for interval-valued data. PCA is a faithful data exploratory tool that has long been used in several engineering fields and had known several improvements over time. Among the newest versions of PCA is PCA for interval-valued data. The purpose of this work was to investigate the applicability of this simple, yet a powerful method for the purpose of fault detection. Indeed, in the fault detection realm, one can believe that any additional information on the process can be beneficial, especially when this information is about uncertainties. We showed throughout this manuscript, that PCA for interval-valued data not only is applicable for fault detection and performs better than conventional PCA approach but also provides notable robustness towards uncertainties of measurements. However, most of the theory behind PCA based detection has been developed for single-valued data, so we had to forge most of the theory for an interval-valued PCA case. Starting from the four most know PCA models for interval-valued data, known as VPCA, CPCA, MRPCA and CIPCA, we had to define fault indicators of interval type. Overall, the main aspects of this dissertation can be summarized as follows.

- The key performance of PCA for interval-valued data for sensor fault detection is its ability to neglect uncertainties of measurement by considering them as normal process variation. In other words, due to the interval nature of data, any information inside the interval is considered as a normal variation of the process, while any data outside this interval is considered as a fault.

- Two strategies are presented. The first strategy is a univariate fault detection, which has been performed on generated residuals from the model. The second strategy is a multivariate strategy which is based on fault detection statistics, with the new fault indicator ϕ, where we have seen it's efficiency.

- The proposed strategy is tested on a simulation example. Then a real application is presented in order to compare different interval-valued models and statistics which show substantial improvements in precision for detecting faults while perfectly handling uncertainties by considering them as process variations.

All in all, this research axis is proving itself very promising where several open issues deserve further study for interval-valued variables.

Appendix A

DESCRIPTION OF THE CEMENT ROTARY KILN

The cements plant divides into two parts (a) and (b) following material flow direction, from the top of the pre-heater tower downwards to the cooler. Rotary kiln is considered to be the hart of any cement plant. Its role is to heat materials to a point to which chemical reactions can take place. The experimental application of the proposed method was conducted on the rotary kiln of Ain El Kebira cement plant (first production line) in eastern Algeria. It is about 80 m length and 5.4 m diameter, with 3 degrees incline. The kiln is being rotated at a maximum speed of 2.14 rpm by two 560 kW asynchronous motors. The cement substance production can be summarized in the following phases. The first phase is to feed raw materials at the upper end of two parallel towers that include a series of four cyclones mounted on four floors. This phase mainly allows material drying and dehydration using the hot counter stream gas and a secondary fuel burner mounted in the tower bottom.

(a) pre-heater tower at the right along with rotary kiln laying in horizontal to the left.

(b) cooler system with its heat exchanger and filter to the left and kiln end appears to the right.

Figure A.1: Ain El Kebira cement plant

The second phase is that the pre-heated mixture enters the kiln and moves gradually downstream to the kiln's lower end due to its rotation. In this phase, the material is further heated using the main burner and hot secondary gas coming back from the cooler. In high temperatures (more than $1450C^{\circ}$) a liquid phase appears and complex chemical reactions are sped up before entering vitrification phase where materials become again solid but in a new structural form called clinker. The clinker is then milled with some additives to produce cement. The third phase is to cool the kiln product down to a temperature less than $100C^{\circ}$ by means of the cooler that consists of many fans, static and mobile grates.

Due to the complex dynamic, multivariable nature, nonlinear reaction kinetics, long time delays and variable raw material feed characteristics, the rotary kiln process is inherently difficult to model. It was declared that *"to the authors knowledge, there is no mathematical model that adequately describes the process behavior"*. Moreover, the product quality of industrial rotary kilns is usually measured after the clinker has cooled down which adversely limits the online supervision.

The following figure gives simplified schematic of the installation.

The applications in this work were made on a data collected on the cement plant of Ain El-Kbira, Setif, Algeria. from January 23^{th}, 2014 at 23:39:33PM to January 24^{th}, 2014 at 04:30:00AM with acquisition rate of one second. The current control performed in Process is manual centralized mode, where any action is based upon the human experience. The used data set consists of a continuous measurement for 4 hours, 50 minutes and 28 seconds (177428 seconds). The first 15300 seconds are collected when the system is surely in healthy state. The later 2084 seconds are collected in the presence of a fault in the system. The total data set holds information from 52 sensors spread along the process to monitor the different quantities, i.e., temperature, pressure, flow,...etc. Finally, amongst the 52 sensors, the sensors '331-01-07 / PE (or TE)' and '341-01-07 / PE(or TE)' are not used as feedback sensors for any control loop.

This chapter is dedicated to the investigation of PCA based approaches for fault detection in the case of single valued-data, and for the interval valued-data. The following figure gives simplified schematic of the installation :

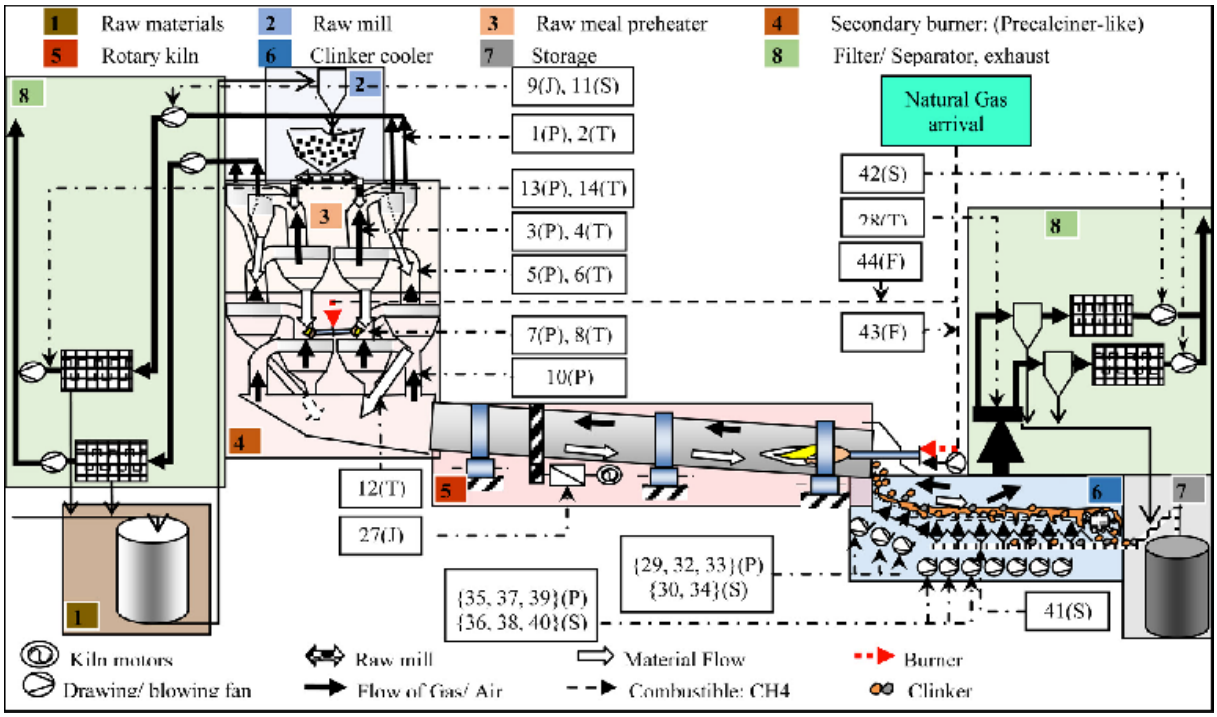


Figure A.2: Overview of the manufacturing process in the cement plant, depicting main unit operations (1-8) including the rotary kiln.

To construct and test the proposed fault detection scheme then evaluate and compare its performance, table (4.2) lists the different data sets.

Table A.1: Data sets used for the application.

Data sets	Size	Sampling intervals [s]	Description
Training set	$X_0 \in \mathrm{R}^{768\times44}$	20	Normal operation data,construct FD scheme
Testing set	$X_0 \in \mathrm{R}^{11000\times44}$	1	Normal operation data,to test FD scheme
Process fault	$X_0 \in \mathrm{R}^{2084\times44}$	1	Normal/Faulty operation, process fault
Sensor faults (7 sets)	$X_0 \in \mathrm{R}^{1500\times44}$	1	Sensor fault simulations

- *Training data:* the training data are collected during healthy operation, with a sampling rate of 1 sample each 20 s This set is used to construct the proposed fault detection scheme: building PCA model, computing the thresholds.

- *Testing data:* it is collected from the plant during healthy operation with a sampling interval of 1 s. This data set is used to test the accuracy and noise rejection of the proposed fault detection scheme compared to the fixed threshold scheme in terms of the false alarms.

- *Faulty process data:* this set consists of two regions; first region corresponds to healthy operation, whilst the second region corresponds to the process faulty operation. The fault occurs after 7 min and evolves slowly in the process. We use this data to check the ability of the proposed detection scheme to detect the fault.

- *Sensor faults:* this set contains 7 simulated sensor faults occurring in the rotary kiln process. It includes abrupt, random, intermittent, and slow drift additive faults that might be single as well as multiple faults. The original data is taken during healthy operation of the rotary kiln, and then faults are introduced from 500th sample to 1000th sample, the different faults are :

Table A.2: Simulated sensor faults, introduced at 500-1000[s].

Fault	Faulty variables	Fault magnitude	Description of the fault
Fault(1)	16	0.5%	Additive random fault, with mean 0, and variance 0.05
Fault(2)	44	-2%	Abrupt additive fault, bias $b = -0.02$
Fault(3)	30	+2%	Additive fault: Linear drift from 0% to +2%; slope $K_s = 4 \times 10^{-5}$
Fault(4)	34	-2%	Additive fault: Linear drift from 0% to -2%; slope $K_s = -4 \times 10^{-5}$
Fault(5)	12,18,43	[+,-,+]2%	Abrupt additive fault ±2% (multiple)
Fault(6)	4,6,8,14.24	[+,+,+,-,-]2%	Additive fault: Linear drift from 0% to ±2% (multiple); slope $K_s = \pm 4 \times 10^{-5}$
Fault(7)	11	+4.5%,-5.5%	Additive fault: Intermittent fault, changing intervals and amplitudes

1. fault1 is a random fault of mean 0 and variance 0.05 added to variable 16, which represents the speed of the induced draft fan installed in top of tower II for drawing gas from preheater cyclones to filters. This fault simulates a loss of accuracy in the speed sensor. It can also represent a vibration in the fan resulting in a small oscillation in the speed without changing its average value.

2. fault2 is an abrupt bias of -2% in feed of natural gas to the secondary burner which can happen due to many reasons: (i) a partial blockage in the natural gas tubes, (ii) leakage of natural gas, or (iii) just a bias error in sensor readings. Therefore, it has a bad impact on performance and safety of the process.

3. fault3 is a positive drift in the speed of the first blowing fan located in the cooling unit, where the measured speed increased linearly from 0% to 2% in 500[s].

4. fault4 is a negative drift in the speed of the third blowing fan similar to Sfault3. This fault may change the speed of cooling the clinker and result in a bad quality product.

5. fault5 represent multiple abrupt faults of magnitude 2%, occurring simultaneously at different sensors in positive and negative directions, respectively.

6. fault6 also represent a multiple abrupt faults of magnitude 2%, occurring simultaneously at different sensors in positive and negative directions, respectively.

7. fault7 is a positive intermittent fault in the speed of the exhauster fan of tower I, this actuator/sensor fault has a changing interval (non-periodic) with changing amplitude from 4.5% to 5.5%. These changes in speed can be caused by the actuator due to malfunctions in the drive's semiconductors or vibrations on the fan's shaft,

they can also be caused by the speed's sensor (tachometer). This fault is introduced in the intervals: 500-580, 610-660, 700-740, 800-830, 870-900, and 975-1000 with amplitudes 5.5%, 4.5%, 5%, 5.5%, 5%, and 4.5% respectively.

Bibliography

[1] Alfonso M Albano et al. "Singular-value decomposition and the Grassberger-Procaccia algorithm." In: *Physical review A* 38.6 (1988), p. 3017.

[2] Alkan ALKAYA. "Novel data driven-based fault detection for electromechanical and process control systems." PhD thesis. Adana, Turkey: Cukurova University, 2012: 38-43.

[3] Anissa Benaicha et al. "Fault detection and isolation with interval principal component analysis." In: *Proceedings Engineering & Technology*. Vol. 1. 2013, pp. 162–167.

[4] Patrice Bertrand and Françoise Goupil. "Descriptive statistics for symbolic data." In: *Analysis of symbolic data*. Springer, 2000, pp. 106–124.

[5] George EP Box et al. "Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification." In: *The annals of mathematical statistics* 25.2 (1954), pp. 290–302.

[6] James R Bunch and Christopher P Nielsen. "Updating the singular value decomposition." In: *Numerische Mathematik* 31.2 (1978), pp. 111–129.

[7] Emmanuel J Candès et al. "Robust principal component analysis?" In: *Journal of the ACM (JACM)* 58.3 (2011), p. 11.

[8] Richard Cangelosi and Alain Goriely. "Component retention in principal component analysis with application to cDNA microarray data." In: *Biology direct* 2.1 (2007), p. 2.

[9] Pierre Cazes et al. "Extension de l'analyse en composantes principales à des données de type intervalle." In: *Revue de Statistique appliquée* 45.3 (1997), pp. 5–24.

[10] Ahlame Chouakria. "Extension des méthodes d'analyse factorielle à des données de type intervalle." PhD thesis. Paris 9, 1998.

[11] JJ Daudin, C Duby, and P Trecourt. "Stability of principal component analysis studied by the bootstrap method." In: *Statistics: A journal of theoretical and applied statistics* 19.2 (1988), pp. 241–258.

[12] Bart De Ketelaere, Mia Hubert, and Eric Schmitt. "Overview of PCA-based statistical process-monitoring methods for time-dependent, high-dimensional data." In: *Journal of Quality Technology* 47.4 (2015), pp. 318–335.

[13] Edwin Diday. "Introduction à l'approche symbolique en analyse des données." In: *RAIRO-Operations Research-Recherche Opérationnelle* 23.2 (1989), pp. 193–236.

[14] Steven Ding. "Model-based fault diagnosis in dynamic systems using identification techniques, Silvio Simani, Cesare Fantuzzi and Ron J. Patton, Springer: London, 2003, 282pp. ISBN 1-85233-685-4." In: *International Journal of Robust and Nonlinear Control* 15.11 (2005), pp. 509–512.

[15] Xinghao Ding, Lihan He, and Lawrence Carin. "Bayesian robust principal component analysis." In: *IEEE Transactions on Image Processing* 20.12 (2011), pp. 3419–3430.

[16] Pierpaolo D'Urso and Paolo Giordani. "A least squares approach to principal component analysis for interval valued data." In: *Chemometrics and Intelligent Laboratory Systems* 70.2 (2004), pp. 179–192.

[17] Paul M Frank and Xianchun Ding. "Survey of robust residual generation and evaluation methods in observer-based fault detection systems." In: *Journal of process control* 7.6 (1997), pp. 403–424.

[18] Karl Ruben Gabriel. "The biplot graphic display of matrices with application to principal component analysis." In: *Biometrika* 58.3 (1971), pp. 453–467.

[19] Zhiqiang Ge and Zhihuan Song. "Process monitoring based on independent component analysis- principal component analysis (ICA- PCA) and similarity factors." In: *Industrial & Engineering Chemistry Research* 46.7 (2007), pp. 2054–2063.

[20] Federica Gioia and Carlo N Lauro. "Principal component analysis on interval data." In: *Computational Statistics* 21.2 (2006), pp. 343–363.

[21] Louis W Glorfeld. "An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain." In: *Educational and psychological measurement* 55.3 (1995), pp. 377–393.

[22] David A Harville and Daniel R Jeske. "Mean squared error of estimation or prediction under a general linear model." In: *Journal of the American Statistical Association* 87.419 (1992), pp. 724–731.

[23] John L Horn. "A rationale and test for the number of factors in factor analysis." In: *Psychometrika* 30.2 (1965), pp. 179–185.

[24] H Hotteling. *Multivariate quality control illustrated by the air testing of sample bombsites*. 1947.

[25] Rolf Isermann. *Fault-diagnosis systems: an introduction from fault detection to fault tolerance*. Springer Science & Business Media, 2006.

[26] Rolf Isermann. "Model-based fault-detection and diagnosis–status and applications." In: *Annual Reviews in control* 29.1 (2005), pp. 71–85.

[27] Rolf Isermann. "Process fault detection based on modeling and estimation methodsâ€"A survey." In: *automatica* 20.4 (1984), pp. 387–404.

[28] Rolf Isermann. "Supervision, fault-detection and fault-diagnosis methodsâ€"an introduction." In: *Control engineering practice* 5.5 (1997), pp. 639–652.

[29] Rolf Isermann and Peter Balle. "Trends in the application of model-based fault detection and diagnosis of technical processes." In: *Control engineering practice* 5.5 (1997), pp. 709–719.

[30] J Edward Jackson and Govind S Mudholkar. "Control procedures for residuals associated with principal component analysis." In: *Technometrics* 21.3 (1979), pp. 341–349.

[31] Mika Järvensivu, K Saari, and S-L Jämsä-Jounela. "Intelligent control system of an industrial lime kiln process." In: *Control Engineering Practice* 9.6 (2001), pp. 589–606.

[32] Ian T Jolliffe. "Mathematical and statistical properties of population principal components." In: *Principal component analysis* (2002), pp. 10–28.

[33] Henry F Kaiser. "The application of electronic computers to factor analysis." In: *Educational and psychological measurement* 20.1 (1960), pp. 141–151.

[34] Vamshi Krishna Kandula. "Fault detection in process control plants using principal component analysis." In: (2011).

[35] Srinivas Katipamula and Michael R Brambley. "Methods for fault detection, diagnostics, and prognostics for building systemsâ€"a review, part I." In: *Hvac&R Research* 11.1 (2005), pp. 3–25.

[36] HR Keller and DL Massart. "Peak purity control in liquid chromatography with photodiode-array detection by a fixed size moving window evolving factor analysis." In: *Analytica chimica acta* 246.2 (1991), pp. 379–390.

[37] Abdelmalek Kouadri et al. "An adaptive threshold estimation scheme for abrupt changes detection algorithm in a cement rotary kiln." In: *Journal of Computational and Applied Mathematics* 259 (2014), pp. 835–842.

[38] PREM KRISHNAN. "Applications of Multiv Ariate Analysis Techniques for Fault Detection, Diagnosis and Isolation." PhD thesis. 2011.

[39] P Legendre and L Legendre. "Numerical ecology, 2nd English edition Elsevier Science BV." In: *Amsterdam. 852p* (1998).

[40] Wei Lou, Guoying Shi, and Jun Zhang. “Research and application of ICA technique in fault diagnosis for equipments.” In: *2009 IEEE International Conference on Intelligent Computing and Intelligent Systems*. Vol. 4. IEEE. 2009, pp. 310–313.

[41] Rongfu Luo, Manish Misra, and David M Himmelblau. “Sensor fault detection via multiscale analysis and dynamic PCA.” In: *Industrial & Engineering Chemistry Research* 38.4 (1999), pp. 1489–1495.

[42] Robert L Mason, Nola D Tracy, and John C Young. “A practical approach for interpreting multivariate T 2 control chart signals.” In: *Journal of Quality Technology* 29.4 (1997), pp. 396–406.

[43] Robert L Mason, Nola D Tracy, and John C Young. “Decomposition of T 2 for multivariate control chart interpretation.” In: *Journal of quality technology* 27.2 (1995), pp. 99–108.

[44] Dubravko Miljković. “Fault detection methods: A literature survey.” In: *2011 Proceedings of the 34th international convention MIPRO*. IEEE. 2011, pp. 750–755.

[45] Paige Miller, Ronald E Swanson, and Charles E Heckler. “Contribution plots: A missing link in multivariate quality control.” In: *Applied mathematics and computer science* 8.4 (1998), pp. 775–792.

[46] Douglas C Montgomery and George C Runger. *Applied statistics and probability for engineers*. John Wiley & Sons, 2010.

[47] Bruce Moore. “Principal component analysis in linear systems: Controllability, observability, and model reduction.” In: *IEEE transactions on automatic control* 26.1 (1981), pp. 17–32.

[48] Paul Nomikos and John F MacGregor. “Multivariate SPC charts for monitoring batch processes.” In: *Technometrics* 37.1 (1995), pp. 41–59.

[49] Brian P Oâ€™connor. “SPSS and SAS programs for determining the number of components using parallel analysis and Velicerâ€™s MAP test.” In: *Behavior research methods, instruments, & computers* 32.3 (2000), pp. 396–402.

[50] Francesco Palumbo and Carlo N Lauro. “A PCA for interval-valued data based on midpoints and radii.” In: *New developments in Psychometrics*. Springer, 2003, pp. 641–648.

[51] Pedro R Peres-Neto, Donald A Jackson, and Keith M Somers. “How many principal components? Stopping rules for determining the number of non-trivial axes revisited.” In: *Computational Statistics & Data Analysis* 49.4 (2005), pp. 974–997.

[52] S Joe Qin. “Survey on data-driven industrial process monitoring and diagnosis.” In: *Annual reviews in control* 36.2 (2012), pp. 220–234.

[53] S Joe Qin, Hongyu Yue, and Ricardo Dunia. "Self-validating inferential sensors with application to air emission monitoring." In: *Industrial & engineering chemistry research* 36.5 (1997), pp. 1675–1685.

[54] Mark Rafferty et al. "Real-time multiple event detection and classification using moving window PCA." In: *IEEE Transactions on Smart Grid* 7.5 (2016), pp. 2537–2548.

[55] C Radhakrishna Rao. "The use and interpretation of principal component analysis in applied research." In: *Sankhyā: The Indian Journal of Statistics, Series A* (1964), pp. 329–358.

[56] Jiri Rohn and Assem Deif. "On the range of eigenvalues of an interval matrix." In: *Computing* 47.3 (1992), pp. 373–377.

[57] Evan L Russell, Leo H Chiang, and Richard D Braatz. *Data-driven methods for fault detection and diagnosis in chemical processes*. Springer Science & Business Media, 2012.

[58] Kristen Severson, Paphonwit Chaiwatanodom, and Richard D Braatz. "Perspectives on process monitoring of industrial systems." In: *Annual Reviews in Control* 42 (2016), pp. 190–200.

[59] Rumana Sharmin, Sirish L Shah, and Uttandaraman Sundararaj. "A PCA based fault detection scheme for an industrial high pressure polyethylene reactor." In: *Macromolecular Reaction Engineering* 2.1 (2008), pp. 12–30.

[60] Sigurd Skogestad. "Dynamics and control of distillation columns: A tutorial introduction." In: *Chemical Engineering Research and Design* 75.6 (1997), pp. 539–562.

[61] Ehsan Sobhani-Tehrani and Khashayar Khorasani. *Fault diagnosis of nonlinear systems using a hybrid approach*. Vol. 383. Springer Science & Business Media, 2009.

[62] Masayuki Tamura and Shinsuke Tsujita. "A study on the selection of model dimensions and sensitivity of PCA-based fault detection." In: *IFAC Proceedings Volumes* 37.9 (2004), pp. 529–534.

[63] Roberto Teti et al. "Advanced monitoring of machining operations." In: *CIRP annals* 59.2 (2010), pp. 717–739.

[64] Doan X Tien, K-W Lim, and Liu Jun. "Comparative study of PCA approaches in process monitoring and fault detection." In: *30th Annual Conference of IEEE Industrial Electronics Society, 2004. IECON 2004*. Vol. 3. IEEE. 2004, pp. 2594–2599.

[65] Nola D Tracy, John C Young, and Robert L Mason. "Multivariate control charts for individual observations." In: *Journal of quality technology* 24.2 (1992), pp. 88–95.

[66] Aditya Tulsyan and Paul I Barton. “Interval enclosures for reachable sets of chemical kinetic flow systems. Part 1: Sparse transformation.” In: *Chemical Engineering Science* 166 (2017), pp. 334–344.

[67] Sergio Valle, Weihua Li, and S Joe Qin. “Selection of the number of principal components: the variance of the reconstruction error criterion with a comparison to other methods.” In: *Industrial & Engineering Chemistry Research* 38.11 (1999), pp. 4389–4401.

[68] Wayne F Velicer. “Determining the number of components from the matrix of partial correlations.” In: *Psychometrika* 41.3 (1976), pp. 321–327.

[69] Venkat Venkatasubramanian et al. “A review of process fault detection and diagnosis: Part I: Quantitative model-based methods.” In: *Computers & chemical engineering* 27.3 (2003), pp. 293–311.

[70] Huiwen Wang, Rong Guan, and Junjie Wu. “CIPCA: Complete-Information-based Principal Component Analysis for interval-valued data.” In: *Neurocomputing* 86 (2012), pp. 158–169.

[71] Xun Wang, Uwe Kruger, and George W Irwin. “Process monitoring approach using fast moving window PCA.” In: *Industrial & Engineering Chemistry Research* 44.15 (2005), pp. 5691–5702.

[72] Mati Wax and Thomas Kailath. “Detection of signals by information theoretic criteria.” In: *IEEE Transactions on acoustics, speech, and signal processing* 33.2 (1985), pp. 387–392.

[73] John H Wilkinson and Christian Reinsch. *Handbook for Automatic Computation: Volume II: Linear Algebra*. Vol. 186. Springer Science & Business Media, 2012.

[74] Svante Wold. “Cross-validatory estimation of the number of components in factor and principal components models.” In: *Technometrics* 20.4 (1978), pp. 397–405.

[75] Svante Wold. “Exponentially weighted moving principal components analysis and projections to latent structures.” In: *Chemometrics and intelligent laboratory systems* 23.1 (1994), pp. 149–161.

[76] Svante Wold, Kim Esbensen, and Paul Geladi. “Principal component analysis.” In: *Chemometrics and intelligent laboratory systems* 2.1-3 (1987), pp. 37–52.

[77] John Wright et al. “Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization.” In: *Advances in neural information processing systems*. 2009, pp. 2080–2088.

[78] Shen Yin et al. “A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process.” In: *Journal of process control* 22.9 (2012), pp. 1567–1581.

[79] Q-T Zhang et al. “Statistical analysis of the performance of information theoretic criteria in the detection of the number of signals in array processing.” In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 37.10 (1989), pp. 1557–1567.

[80] Xiaodong Zhang, Marios M Polycarpou, and Thomas Parisini. “A robust detection and isolation scheme for abrupt and incipient faults in nonlinear systems.” In: *IEEE transactions on automatic control* 47.4 (2002), pp. 576–593.

[81] William R Zwick and Wayne F Velicer. “Comparison of five rules for determining the number of components to retain.” In: *Psychological bulletin* 99.3 (1986), p. 432.