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 **Control**

Option: **Control**

Title:

Reduced Kernel PCA Technique for Fault Detection in Complex Systems: A Fractal Dimension-based Approach

Presented by:

Mr.BOULMERDJ Anes

Mr.BOUNAAS Alaeddine.

Supervisor:

Dr.KOUADRI Abdelmalek.

Mr.KAIB Mohammed Taher Habib (Co-Supervisor).

Registration Number:………/2021

Abstract

Fault detection and diagnosis is an important problem in process engineering. It is the central component of abnormal event management (AEM) which has attracted a lot of attention recently.

This thesis discuses different classes of FDD approaches for process monitoring. In addition, it presents main results of fault detection and diagnosis in a cement manufacturing plant using three monitoring techniques. The techniques are based on multivariate statistical analysis and a threshold strategy. The process is statistically modeled using Principle Component Analysis (PCA), kernel PCA and new proposed reduced KPCA to cope with the computational problem introduced by KPCA. The proposed RKPCA method consists on reducing the number of observations in a data matrix using a proposed algorithm based on fractal dimension.

The Hotelling's T^2, Q in addition to the new proposed index called the combined statistic φ are used as fault indicators for testing PCA, KPCA and the suggested approach RKPCA carried out using the cement rotary kiln system. The three methods are compared to in terms of False Alarms Rate (FAR), Missed Alarms Rate (MDR), Detection Time Delay (DTD) and the cost function (J).

The obtained results demonstrate the effectiveness of the proposed technique in reducing the number of observations from 768 to 11, leading to an 11x11 kernel matrix instead of 768x768, hence, diminishing computational time and storage requirement. Moreover, it has effectively detected the different types of faults when using statistical indices.

Keywords: Fault Detection; Principal Component Analysis (PCA); Kernel Principal Component Analysis (KPCA); Reduced Kernel Principal Component Analysis (RKPCA); Fractal Dimension; the Combined Index.

Praise is to Allah by Whose grace good deeds are completed, Praise is to You as befits the Glory of Your Face and the greatness of Your Might.

This study is wholeheartedly dedicated to:

*My **dear Parents,** who gave me an endless support, a lot of inspirations and emotions, who kept pushing me to this point and still. May god bless them.*

*My beloved wonderful sisters **Marwa** and **Rihab** and the whole family.*

B.Alaeddine

*My **beloved Parents**, who have continually provided me with strength, motivation and unlimited support. **My brother and Sisters**. My whole Family.*

B.Anes

To all our friends with whom we shared the best memories. To every person that left a good mark during our journey.

AKNOWLEDGEMENT

We would like to show our heartfelt appreciation to our supervisor, Dr. ***Abdelmalek KOUADRI****, for providing us with essential advice during this project. His energy, passion, and motivation have greatly impacted us. Working and studying under his supervision was a wonderful pleasure and honor.*

*We would like to express our gratitude to Mr.****KAIB Mohammed Taher Habib*** *for his assistance. We are very grateful for all of his time and suggestions in making this experience successful and interesting.*

We would also want to thank all of the teachers of the Institute of Electrical and Electronics Engineering of the University of M'Hamed Bougara- Boumerdes for their assistance and support over the previous five years.

TABLE OF CONTENT

ABSTRACT

DEDICACE

AKNOWLEDGEMENT

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

NOMENCLATURE

GENERAL INTRODUCTION ………………………….…………………………**1**

CHAPTER I. FAULT DETECTION AND DIAGNOSIS

I.1. Introduction…………………………………………………………………**4**

I.2. Definitions of fault detection and diagnosis……………………………………...**4**

I.3. Desirable attributes of a fault detection and diagnosis system……………………...**6**

I.4. Transformation of measurements in a diagnostic system…………………………...**8**

I.5. Classification of fault detection and diagnosis methods……………………………**9**

I.5.1. Quantitative model based methods…………………………………………**10**

I.5.2. Qualitative Model-Based Methods…………………………………………**12**

I.5.3. Process History (Data-driven) Methods……………………………………**13**

I.6. A comparison of various approaches…………………………………………**16**

I.7. Statistical Process Monitoring (SPM)……………………………………….**16**

I.8. Conclusion…………………………………………………………………**17**

CHAPTER II. MULTIVARIATE STATISTICAL APPROACHES

II.1. Introduction………………………………………………………………**18**

II.2. Principal Component Analysis (PCA)………………………………………**18**

II.2.1. Statistical process modeling using PCA…………………………………..**18**

II.2.2. Model dimension selection………………………………………………..**20**

II.2.3. Fault detection Indices ……………………………………………………......…**20**

II.2.4. PCA main drawbacks……………………………………………………………..**24**

II.3. Kernel Principal Component Analysis (KPCA)…………………………………**24**

II.3.1. Statistical process modeling using KPCA………………………………………**25**

II.3.2. Fault detection using KPCA……………………………………………………**29**

II.3.3. KPCA main drawback…………………………………………………………..**31**

II.4. The proposed approach Reduced Kernel principal component (RKPCA)……….**32**

II.4.1. Fractal Dimension……………………………………………………………**32**

II.4.2.Correlationdimension…………………………………………..……….…**33**

II.4.3. Data Reduction using Correlation Dimension ……………………..………….**34**

II.5. Conclusion ………………………………………………………..……..**35**

CHAPTER III. APPLICATION, RESULTS & DISCUSSION

III.1. Introduction……………………………………………………………………**36**

III.2. Process description……………………………………………………………**36**

III.3. Computation of monitoring performance metrics………………………………..**38**

III.4. Application procedure…………………………………………………………**39**

III.4.1. Application of PCA, KPCA and the proposed approach RKPCA to cement rotary kiln process……………………………………………………………………...**39**

III.5. Results and discussion…………………………………………………………..**42**

III.5.1. PCA monitoring model………………………………………………………...**42**

III.5.2. KPCA monitoring model………………………………………………………**44**

III.5.3. The proposed RKPCA monitoring model……………………………………..**50**

III.5.4. Comparison between PCA , KPCA and the proposed approach RKPCA…….**60**

III.6. Conclusion ……………………………………………………………………**63**

GENERAL CONCLUSION……………………………………………………………**64**

REFERENCES

LIST OF TABLES

Table 1. Comparison of diagnosis methods…………………………………...……..**16**

Table 2. PCA-based fault detection algorithm……………………………...……..**23**

Table 3. KPCA-based fault detection algorithm……………………….….…...**31**

Table 4. Data reduction using Correlation Dimension ……….…………...………..**35**

Table 5. Process variables of the cement rotary kiln……………………..….…….…**37**

Table 6. Data sets used in the application……………………………..………….…..**39**

Table 7. Simulated sensor faults introduced at 500-1000 s……………..……..………**39**
..
Table 8. FAR contributed by T^2, Q, under NOC using PCA. …………..………..…...**42**

Table 9. Missed detection rate (MDR), False Alarm Rate (FAR), detection time delay (DTD) and the cost function J values for the eight faults of cement rotary kiln using PCA… ………………………………………….……………………….…**43**

Table 10. FAR contributed by T^2, Q, φ under NOC using KPCA…………...……….**44**

Table 11. Missed detection rate (MDR), False Alarm Rate (FAR), detection time delay (DTD) and the cost function J values for the eight faults of cement rotary kiln using KPCA……………………………………..………………………………………**46**

Table 12. RKPCA time complexity analysis………………………………… …..**54**

Table 13. FAR contributed by T^2, Q, φ under NOC using the proposed RKPCA…....**55**

Table 14. Missed detection rate (MDR), False Alarm Rate (FAR), detection time delay (DTD) and the cost function J values for the eight faults of cement rotary kiln using the proposed RKPCA..…………………………………………………………...**57**

Table 15. FAR (%) of faults monitoring results ……….……………………… **61**

Table 16. MDR (%) and DTD of faults monitoring results…………………………. **62**

Table 17. A comparative table between PCA, KPCA and the proposed RKPCA using the average J value contributed by T^2, Q and φ…………………………………… **62**

LIST OF FIGURES

Figure 1. Time dependency of faults………………………………………………….……… **4**

Figure 2. Basic fault models……………………………………………………….….…. **5**

Figure 3. Transformations in a diagnostic system……….. **8**

Figure 4. Classification of fault detection and diagnosis methods……………….….…. **10**

Figure 5. Control chart for SPC…………………………………………………..….. **17**

Figure 6. Data projection on two PCs ……………………………………….……... **21**

Figure 7. Non-linear data transformation to linear feature space $\mathcal{F}$ ……………………**25**

Figure 8. Ain El Kebira cement plant ……………………………………...…… **37**

Figure 9. T^2, Q and φ monitoring results of healthy process operation using KPCA ….**45**

Figure 10. T^2, Q and φ monitoring results of real involuntary process fault in the cement rotary kiln using KPCA……………………………………………………….…...….…..**47**

Figure 11. T^2, Q and φ monitoring results of sensor faults using KPCA……………..… **49**

Figure 12. Graphically representation of the original data (786 observations)……… **50**

Figure 13. Bar chart represents the Kullbeck-Liebler divergence and the difference in mean and variance between the original and reduced data…………………………. **53**

Figure 14. T^2, Q and φ monitoring results of healthy process operation using RKPCA... **56**

Figure 15. T^2, Q and φ monitoring results of real involuntary process fault in the cement rotary kiln using the proposed RKPCA…………………………………...……..…..**58**

Figure 16. T^2, Q and φ monitoring results of sensor faults using RKPCA…………..…… **60**

NOMENCLATURE

FDD: Fault Detection and Diagnosis

FDI: Fault Detection and Identification

CPV: Cumulative Percent Variance

DTD: Detection Time Delay

FAR: False Alarm Rate

KPCA: Kernel Principal Component Analysis

MDR: Missed Alarm Rate

MSPC: Multivariate Statistical Process Control

NN: Neural Network

NOC: Normal Operation Condition

PC's: Principal Components

PCA: Principal Component Analysis

PLS: Partial Least Square

RKPCA: Reduced Kernel Principal Component Analysis

SPC: Statistical Process Control

SPE: Squared Prediction Error

GENERAL INTRODUCTION

With the advent of industry, current industrial processes are transforming into smart ones. In particular, many modernized industrial processes are equipped with several well-elaborated sensors to gather process-related data for discovering faults existing or arising in processes as well as monitoring the process status. For this change of industrial environments with the full-automation of equipment and process, more cautious supervision that includes process control and suitable corrective actions is required to guarantee the process efficiency. Some changes in the process the controllers cannot handle adequately, these changes are called faults. The types of faults occurring in industrial systems include process parameter changes, disturbance parameter changes, actuator problems, and sensor problems. Here, comes the role of Fault Detection and Diagnosis.

Various approaches have been proposed, tackling this issue from different angles. These can be broadly divided into model based approaches and knowledge based approaches. Model based approaches generally utilize results from the field of control theory and are based on parameter estimation or state estimation. The approach is based on the fact that a fault will cause changes in certain physical parameters which in turn will lead to changes in some model parameters or states. It is then possible to detect and diagnose faults by monitoring the estimated model parameters or states. When using this approach, it is essential to have the knowledge about the relationships between faults and model parameters or states. Furthermore, quite accurate models are required. Knowledge based approaches generally utilize results from the field of Machine Learning. Several knowledge based fault diagnosis approaches have been proposed. These include the quantitative based approach and the qualitative simulation based approach. In the quantitative based approach, faults are usually diagnosed by causally tracing symptoms backward along their propagation paths. In the qualitative simulation based approach, the model of a process is used to predict the behavior of the process under the normal operating condition and various faulty conditions. Fault detection and diagnosis is then performed by comparing the predicted behavior with the actual observations. To develop knowledge based diagnosis systems, knowledge about process structure and qualitative models of process under various faulty conditions are required. The development of a knowledge based diagnosis system is generally effort demanding.

For industrial processes, the most easily obtainable knowledge is usually the process measurement data. During process operations, a large number of process variables are measured and these measurement data are routinely collected and stored by computers. Process monitoring based on statistical analysis of process data has been investigated by several researchers recently. Multivariate statistical process control (MSPC) is a collection of such techniques that are based on the analysis of measurements system. These approaches are generally based on principal component analysis (PCA) or partial least squares (PLS) techniques. PCA have been widely reported in literature for fault detection, isolation and reconstruction, the latest can reduce the dimensionality of process data by projecting them down to a low dimensional latent variable space. Process monitoring can then be performed in this latent variable space. Faults can be detected and classified by inspecting the plots of the Squared Prediction Errors Q, using the Hotteling's T^2 index or the combined index φ. This approach has been shown to be very effective in some situations, especially when the number of faults is not large.

Despite proven performance of PCA for linear system processing, it remains a method of linear projection and cannot reveals nonlinearities. To overcome this limitation, many extensions of PCA for handling nonlinear systems have been developed such as principal curves using neural networks and kernel principal component analysis (KPCA). KPCA has been developed by Schölkopf, it consists to first map input space into a reproducing kernel Hilbert space called feature space via nonlinear kernel function and then to compute principal components (PCs) in that feature space by applying conventional linear PCA. Compared with nonlinear extension of PCA, the advantages of KPCA are that it does not involve any nonlinear optimization problem, making it simple as standard PCA. Indeed, KPCA can handle a wide range of nonlinearities by the possibility to use different kernels. KPCA has been applied successfully for process monitoring in various fields such as chemical technology, face recognition and engineering in medicine. Despite proven performances of KPCA, kernel principal component analysis (KPCA) suffers from a high computational cost and requires the storage of the symmetric kernel matrix (computation time increases with the number of samples). Among several solutions that have been proposed, only few methods are available for dimensionality reduction with KPCA, we mention k means clustering and reduced kernel principal component analysis (RKPCA).

In this project, we propose a new reduced KPCA algorithm based on the fractal dimension. It consists on using the fractal dimension to select a reduced set of observations that "sufficiently" approaches the system behavior. Several techniques used to determine the fractal dimension of a given set; in this work we have used the Correlation dimension that is based on the distances between the variables to identify the fractal dimension. Main advantage of the proposed fractal dimension based RKPCA is that it entails maximum lower memory and computation time by retaining only principal observations. Similar to the PCA, the Hotelling's T^2 statistic, the Q statistic and the combined index φ are indices commonly used in KPCA-based process monitoring, hence, also used in the proposed RKPCA.

The proposed RKPCA method has been tested on a cement manufacturing plant and compared to KPCA and Euclidian distance based RKPCA in terms of False Alarm Rate (FAR), Missed Alarm Rate (MDR), Detection Time Delay (DTD), the Cost function (J) and the Execution Time (ET).

This report is organized as follow: Chapter I presents the theoretical definitions of fault detection and diagnosis field, methods and classification. Chapter II the mathematical background of PCA, KPCA and the proposed method RKPCA are explored and their application in complex processes. Chapter III provides the detailed description of the cement plant followed by the experimental setup. The obtained results are discussed and concluding remarks are drawn.

Chapter I.

Fault Detection and Diagnosis

I.1. Introduction

Fault detection and diagnosis has been an active area of research for the last few decades, which is an essential part of modern industries to ensure safety and product quality [1]. This has encouraged the development of many fault diagnosis methods and which can be classified into three categories: Quantitative Model based methods, Qualitative model based methods, and process history based methods [2].

In this chapter, we first address the definitions and terms used in the domain of process fault diagnosis. In the next part, we present a list of 10 desirable characteristics that one would like a diagnostic system to possess. Section 4 discusses the data transformations that occur during the diagnostic decision-making process. Section 5 contains a classification of fault diagnostics methods.

I.2. Definitions of fault detection and diagnosis

A fault is generally defined as a departure of an observed variable or calculated parameter from an accepted range [3]. More specifically, a fault is an unpermitted deviation of at least one characteristic property of a variable from an acceptable behavior. This means that a fault may lead to the malfunction or failure of the system [4]. Essentially, Fault Detection and Diagnostics, or FDD, is the process of uncovering errors in physical systems while attempting to identify the source of the problem.

The time dependency of faults can be distinguished, as shown in Fig. 1:

- Abrupt fault (stepwise): The cause of the fault and/or the effects remains continuing until corrected.
- Incipient fault (drift-like): The effect on process remains constant until corrected.
- Intermittent fault (with interrupts): The effect disappears and reappears in time.

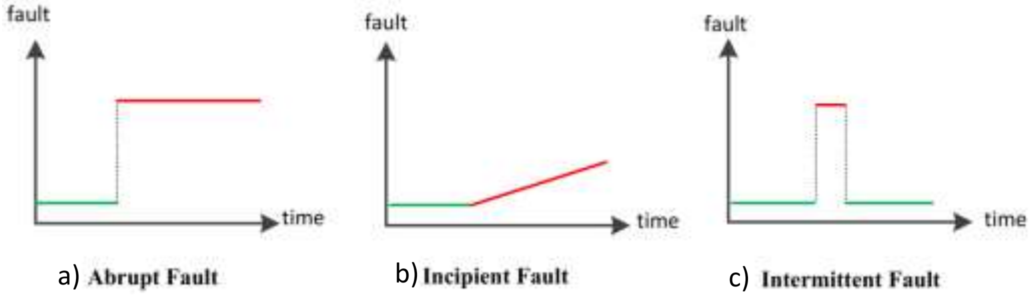


Figure 1. Time dependency of faults: a) Abrupt; b) Incipient; c) Intermittent [29].

Faults are also categorized as additive or multiplicative (Fig.2). The addition of fault characteristics influences a variable in additive faults. These types of faults might emerge as deviations from a normal or desirable value in a process measure. Multiple faults impact the variable as a product and frequently occur as changes in the process parameters [4].

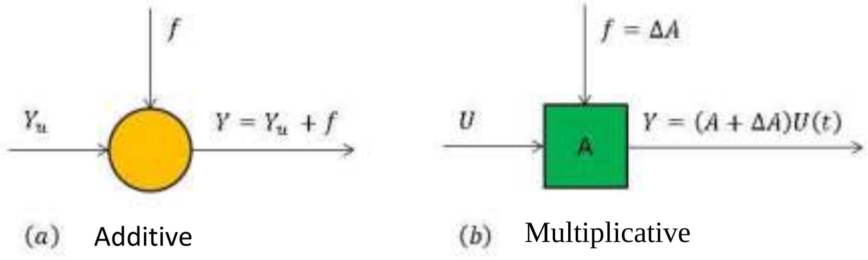


Figure 2. Basic fault models: (a) Additive and (b) Multiplicative faults [5].

Faults may include:

- **Malfunctioning sensors and actuators**

Errors with actuators and sensors are often caused by a fixed failure or a constant bias (positive or negative). Failure of one of the instruments might cause the plant status variables to vary beyond permissible limits. The goal of diagnostic is to discover any instrument failure that might severely affect the operation of the control system [5].

- **Gross parameter changes in a model**

Parameter failures occur when a disturbance from the environment enters the system via one or more independent variables. A variation in the concentration of the reactant in a reactor feed from its usual or steady state value is an example of such a malfunction. The concentration is an independent variable in this case [5].

- **Structural changes**

Changes in the structure come when the process itself changes. They occur when the process equipment is badly damaged. Structural defects affect the flow of information between different variables. The failure of a controller is an example of a structural failure. Other examples are a stuck valve, a fractured or leaked pipe etc. [5].

I.3. Desirable attributes of a fault detection and diagnosis system

In order to compare various diagnostic approaches or to assess whether a fault detection and diagnosis system is successful, it is useful to identify a set of desirable characteristics that a diagnostic system should possess [5].

I.3.1. Quick detection and diagnosis

The diagnostic system should respond quickly to malfunctions. A system that is designed to detect failure quickly will be sensitive to high frequency influences. This makes the system sensitive to noise and can lead to false alarm during normal operation [5].

I.3.2. Isolability

Isolability is the ability of the diagnostic system to distinguish between different failures. Most of the classifiers work with various forms of redundant information. There is only a limited freedom for classifier design. Due to this, a classifier with high degree of isolability would usually do a poor job in rejecting modelling uncertainties and vice versa [5].

I.3.3. Robustness

A diagnostic system satisfying robust feature means its performance should be insensitive to the effect of various noise and modeling uncertainties [5].

I.3.4. Novelty identifiability

A fault detection and diagnosis system should be able to decide whether a process is in a normal or malfunction operation and, if an abnormal condition occurs, whether the causes are from known or novel unknown malfunction [5].

I.3.5. Classification error estimate

An important practical requirement for a diagnostic system is in building the user's confidence on its reliability. This could be greatly facilitated if the diagnostic system could provide a priori estimate on classification error that can occur. Such error measures would be useful to project confidence levels on the diagnostic decisions by the system giving the user a better feel for the reliability of the recommendations by the system [5].

I.3.6. Adaptability

It is also desirable to have extendable systems. This would allow processes to change due to changes in the external inputs, structural changes and also changes in the environmental conditions. Thus the diagnostic system should be adaptable to changes [5].

I.3.7. Modelling requirements

The amount of modelling required for the development of a diagnostic classifier is an important issue. For fast and easy deployment of real-time diagnostic classifiers, the modelling effort should be as minimal as possible [5].

I.3.8. Explanation facility

A diagnostic system should be able to explain where a fault originated and how it propagated in the system [5].

I.3.9. Multiple fault identifiability

This refers to the ability of a diagnostic system to identify and correctly classify multiple faults that may even coexist in a system. This is a rather difficult requirement mainly due to nonlinearities and coupling/ interactions that generally exist between the states and the potential fault sources of a dynamical system. Another reason is that some faults in an engineering system are extremely difficult to model because of their complexity [5].

I.3.10. Storage and computational requirements

This criterion is specifically required for the fast real-time implementation of diagnostic classifiers. Then, the systems should be reasonably balanced between high storage capacities and less computational complexity [5].

I.4. Transformation of measurements in a diagnostic system

To attempt a comparative study of various diagnostic methods it is helpful to view them from different perspectives. In this sense, it is important to identify the various transformations that process measurements go through before the final diagnostic decision is made. Two important components in the transformations are the a priori process knowledge and the search technique used. Hence, one can discuss diagnostic methods from these two perspectives. In general, one can view the diagnostic decision making process as a series of transformations or mappings on process measurements. Fig. 3 shows the various transformations that process data go through during diagnosis.

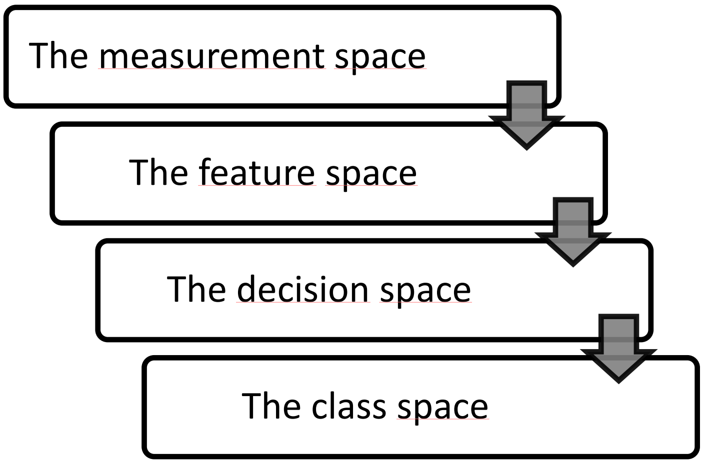


Figure 3. Transformations in a diagnostic system [5].

The measurement space is a space of measurements x_1, x_2,. . .,x_N with no priori problem knowledge relating to these measurements, these are the input to the diagnostic system. The feature space is a space of points y = (y_1,. . .,y_i) where y_i is the i^{th} feature obtained as a function of the measurements by utilizing a priori problem knowledge. Here, the measurements are analyzed and combined with the aid of a priori process knowledge to extract useful features about the process behavior to aid diagnosis. The mapping from the feature space to decision space is usually designate to meet some objective function (such as minimizing the misclassification). This transformation is achieved 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_K], where K is the number of decision variables, obtained by suitable transformations of the feature space. The class space is a set of integers c = [c_1. . . 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. The class space is thus the final interpretation of the diagnostic system delivered to the user. The transformations from decision space to class space are again performed using either threshold functions, template matching or symbolic reasoning as the case may be [5].

I.5. Classification of fault detection and diagnosis methods

There is a large number of literatures about systems of fault diagnostics varying from analytical techniques to artificial intelligence and statistics. The classification of these fault diagnosis methods very often is not consistent. This is primarily due to the fact that academics generally focus on a given branch [4].

The main difference between the approaches of fault diagnosis is about the diagnostic knowledge utilized. At the limits, diagnostics may be based on prior knowledge (e.g., models based entirely on first principles) or empirically driven (e.g., by black-box models). Both techniques utilize models and both employ data, but the diagnostic approach is fundamentally different[6]. Usually, the model-based approach is built out using a basic understanding of the process physics. It can be classified into quantitative model-based and Qualitative model-based. In quantitative models this understanding is expressed in terms of mathematical functional relationships between the inputs and outputs of the system. In contrast, in qualitative model equations, these relationships are expressed in terms of qualitative functions centered on different units in a process [5].

Purely process history approaches (i.e., methods based on black-box models) use no priori knowledge of the process but, instead, derive behavioral models only from measurement data from the process itself. In this latter case, the models may not have any direct physical significance [6]. There are different ways in which this data can be transformed and presented as a priori knowledge to a diagnostic system. This is known as the feature extraction process from the process history data, and is done to facilitate later diagnosis. This extraction process can mainly proceed as either quantitative or qualitative feature extraction (Fig. 4) [5].

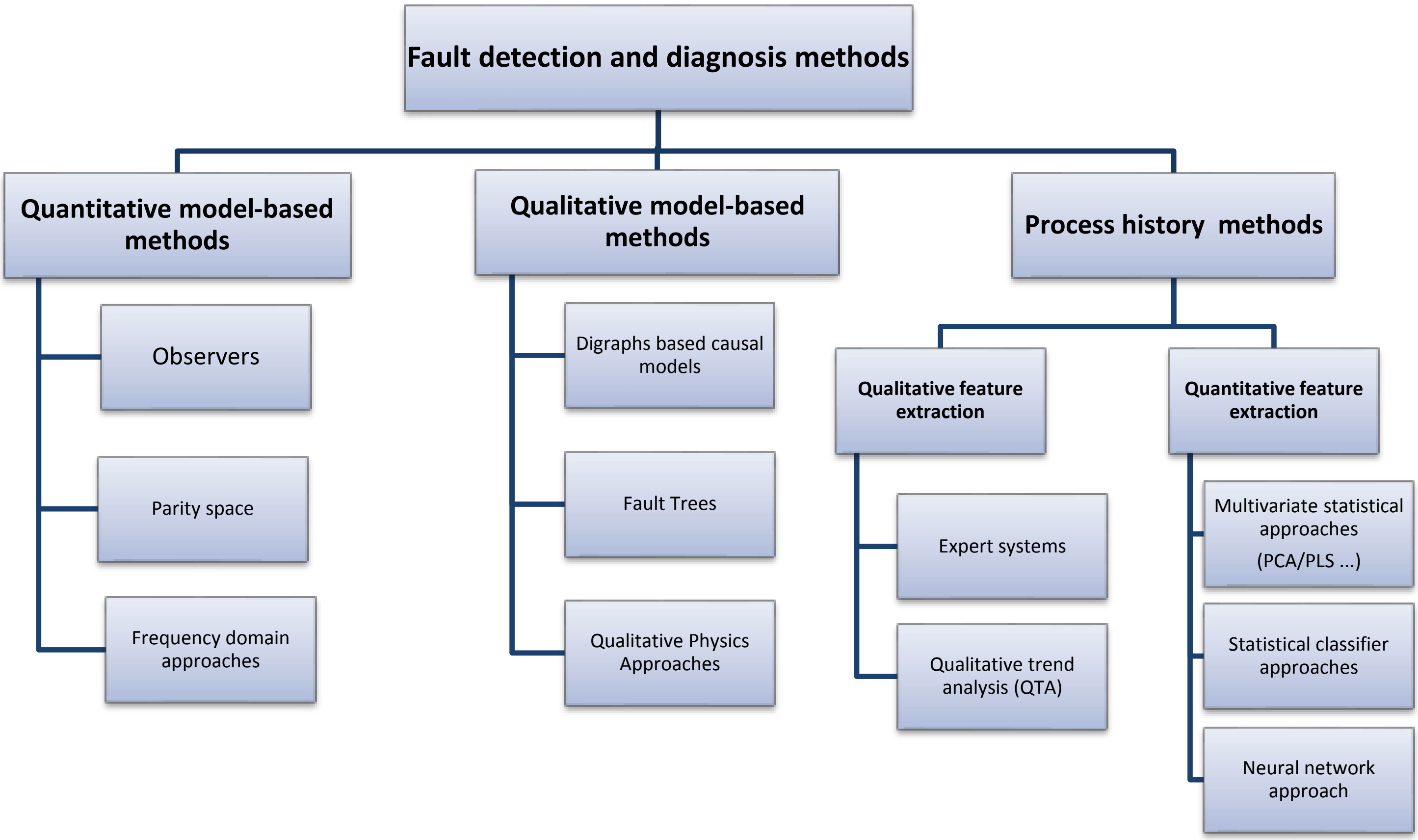


Figure 4. Classification of fault detection and diagnosis methods.

I.5.1. Quantitative model based methods

Model-based methods rely on analytical redundancy by using explicit mathematical models of the monitored process, plant, or system to detect and diagnose faults [30]. The essence of this concept is to check for consistency between the actual outputs of the monitored system and the outputs obtained from a (redundant. i.e. not physical) analytical mathematical model. Therefore, any inconsistency expressed as residuals, can be used for detection and isolation purposes. These residuals should be close to zero when no fault occurs but show 'significant' values when the underlying system changes [5].

The quantitative model-based approaches have been based on using general input-output and state space models to generate residuals. These approaches can be classified into observers, parity space and frequency domain approaches.

I.5.1.1. Observers

The main concern of observer-based FDI is the generation of a set of residuals which detect and uniquely identify different faults. The method develops a set of observers, each one of which is sensitive to a subset of faults while insensitive to the remaining faults and the unknown inputs [5,6].

I.5.1.2. Parity space

It is essential to check the parity (consistency) with sensor outputs (measurements) and know process inputs of plant models. The so-called residual or value of the parity equations is zero under perfect stable operation conditions [5].

I.5.1.3. Frequency domain approaches

Residuals are also generated in the frequency domains via factorization of the transfer function of the monitored system.

Strengths of Quantitative Models. Strengths of fault detection and diagnosis based on quantitative models include:

• Models are based on sound physical or engineering principles.
• They provide the most accurate estimators of output when they are well formulated.
• Detailed models based on first principles can model both normal and “faulty” operation; therefore, “faulty” operation can be easily distinguished from normal operation.
• The transients in a dynamic system can only be modeled with detailed physical models[30].

Weaknesses of Quantitative Models. The weaknesses of fault detection and diagnosis based on quantitative models include:
• They can be complex and computationally intensive.
• The effort required to develop a model is significant.
• These models generally require many inputs to describe the system, some for which values may not be readily available.
• Extensive user input creates opportunities for poor judgment or input errors that can have significant impacts on results [6].

I.5.2. Qualitative Model-Based Methods

Fault detection and diagnostics based on qualitative modeling techniques represent another broad category that is based on a priori knowledge of the system. Unlike quantitative modeling techniques in which knowledge of the system is expressed in terms of quantitative mathematical relationships, qualitative models use qualitative relationships or knowledge bases to draw conclusions regarding the state of a system and its components (e.g., whether operations are “faulty” or “normal”) [6].

Qualitative model-based methods can be classified into:

I.5.2.1. Digraphs based causal models

Causal graphs provide a good way to represent physical cause-effect relations between different process variables that are of interest for fault diagnosis. In the causal directed graph models, the nodes denote the variables, while the directed edges between the nodes represent the causal relations between these variables, through which faults can propagate. The Signed Directed Graph (SDG) method, the simplest causal directed graph method, uses pure qualitative information, which can give rise to ambiguous fault diagnosis [7].

I.5.2.2. Fault Trees

Fault tree analysis (FTA) describes all possible causes of a specified system state in terms of the state of the components within the system. This will be achieved with the use of coherent and non-coherent fault trees. A coherent fault tree is constructed from AND and OR logic, therefore only considers component failed states. The non-coherent method expands this allowing the use of NOT logic which implies that the existence of component failed states and working states are both taken into account [8].

I.5.2.3. Qualitative Physics Approaches

The detailed physical models are based on detailed knowledge of the physical relationships and characteristics of all components in a system. Using this detailed knowledge for mechanical systems, a set of detailed mathematical equations based on mass, momentum, and energy balances along with heat and mass transfer relations are developed and solved. Detailed models can simulate both normal and “faulty” operational states of the system (although modeling of faulty states is not required by all methods).qualitative physics approach is represented in mainly two approaches. The first approach is to derive qualitative equations from the differential equations termed as confluence equations [9]. Considerable work has been done in this area of qualitative modeling of systems and representation of causal knowledge [10].The other approach in qualitative physics is the derivation of qualitative behavior from the Ordinary Differential Equations (ODEs). These qualitative behaviors for different failures can be used as a knowledge source [11].

Strengths of Qualitative Models. Strengths of qualitative models are:

• They are well suited for data-rich environments and noncritical processes.

• These methods are simple to develop and apply.

• Their reasoning is transparent, and they provide the ability to reason even under uncertainty.

• They possess the ability to provide explanations for the suggested diagnoses because the method relies on cause-effect relationships.

• Some methods provide the ability to perform FDD without precise knowledge of the system and exact numerical values for inputs and parameters [6].

Weaknesses of Qualitative Models. Weaknesses of FDD based on qualitative models include:

• The methods are specific to a system or a process.

• Although these methods are easy to develop, it is difficult to ensure that all rules are always applicable and to find a complete set of rules, especially when the system is complex.

• As new rules are added to extend the existing rules or accommodate special circumstances, the simplicity is lost.

• These models, to a large extent, depend on the expertise and knowledge of the developer [6].

I.5.3. Process History (Data-driven) Methods

In process history based methods, only the availability of large amount of historical process data is needed. This data can be transformed and presented as a priori knowledge to a diagnostic system using different ways. And this is known as feature extraction.

This extraction process can be either qualitative or quantitative in nature. Two of the major methods that extract qualitative history information are the expert systems and trend modelling methods. Methods that extract quantitative information can be broadly classified as non-statistical or statistical methods [12].

I.5.3.1. Qualitative feature extraction

Two of the major methods that extract qualitative history information are expert systems and Qualitative Trend Analysis [12].

- **Expert systems**

An expert system is generally a very specialized system that solves problems in a narrow domain of expertise. The main components in an expert system development include: 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. The main advantages in the development of expert systems 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 [10, 12].

- **Qualitative trend analysis (QTA)**

Trend analysis and prediction are important components of process monitoring and supervisory control. Trend modeling can be used to explain the various important events that happen in a process, to diagnose malfunctions and to predict future states. From a procedural perspective, in order to obtain a signal trend not too susceptible to momentary variations due to noise, some kind of filtering needs to be employed [10].

I.5.3.2. Quantitative feature extraction

Methods that extract quantitative information can be broadly classified as Non-statistical or statistical methods. Neural networks are an important class of non-statistical classifiers. Principal component analysis (PCA)/partial least squares (PLS) and statistical pattern classifiers form a major component of the statistical feature extraction methods [10, 12].

- **Multivariate statistical approaches:**

Multivariate statistical techniques are powerful tools capable of compressing data and reducing its dimensionality so that essential information is retained and easier to analyze than the original huge data set; and they are able to handle noise and correlation to extract true information effectively. Multivariate statistical process control methods, such as Principal Component Analysis (PCA) and Partial Least Squares (PLS), have been used in process monitoring problems. These are based on transforming a set of highly correlated variables to a set of uncorrelated variables [13, 14]. Principal component analysis (PCA) probably is the most popular among these techniques [16, 15]. PCA is capable of compressing high-dimensional data with little loss of information by projecting the data onto a lower-dimensional subspace defined by a new set of derived variables (principal components (PCs)) [17].

- **Statistical classifier approaches**

Fault diagnosis is essentially a classification problem and hence can be cast in a classical statistical pattern recognition framework. Fault diagnosis can be considered as a problem of combining, over time, the instantaneous estimates of the classifier using knowledge about the statistical properties of the failure modes of the system [11, 18, 19].

- **Neural network approach**

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 and radial basis (ii) The learning strategy such as supervised and unsupervised learning [12].

Fault detection and diagnosis methods based on process history are well suited to problems for which theoretical models of behavior are poorly developed or inadequate to explain observed performance and where training data are plentiful or inexpensive to create or collect. This approach provides black-box models, which are easy to develop and do not require an understanding of the physics of the system being modeled with a generally manageable computational requirement.

Beside all the advantages listed earlier, the most significant drawbacks is that most of the models cannot be used to extrapolate beyond the range of the training data and a large amount of training data is needed, representing both normal and faulty operation. The models are specific to the system for which they are trained and rarely can be used on other systems. Process data-based methods are suitable where no other methods exist. Some are applicable for virtually any kind of pattern recognition problems [6, 11].

I.6. A comparison of various approaches

Table 1 gives a comparison of various methods in terms of the desirable characteristics of diagnostic systems. In the table only some representative methods in each of the three approaches (quantitative model-based, qualitative model-based, process history based) are chosen for comparison. A check mark would indicate that the particular method (column) satisfies the corresponding desirable property (row). A cross would indicate that the property is not satisfied and a question mark would indicate that the satisfiability of the property is case dependent.

Table 01. Comparison of diagnosis methods [12].

	Observer	Digraphs	Expert systems	QTA	PCA	Neural networks
Quick detection and diagnosis	√	**?**	√	√	√	√
Isolability	√	**x**	√	√	√	√
Robustness	√	√	**x**	√	√	√
Novelty identifiability	**?**	√	**x**	**?**	√	√
Classification error	**x**	**x**	**x**	**x**	**x**	**x**
Adaptability	**x**	√	√	**?**	**x**	**x**
Explanation facility	**x**	√	√	√	**x**	**x**
Modelling requirement	**?**	√	√	√	√	√
Storage and computation	√	**?**	√	√	√	√
Multiple fault identifiability	√	√	**x**	**x**	**x**	**x**

√: suitable; x: not suitable; ? : not assessed

I.7. Statistical Process Monitoring (SPM)

Statistical Process monitoring is an analytical decision making tool which allows to see when a process is working correctly and when it is not. Variation is present in any process, deciding when the variation is natural and when it needs correction is the key to quality control.

A control chart sometimes called a Shewhart chart, a statistical process control chart, or an SPC chart—is one of several graphical tools typically used in quality control analysis to understand how a process changes over time. Data are plotted in time order. A control chart always has a central line for the average, an upper line for the upper control limit, and a lower line for the lower control limit. These lines are determined from historical data. By comparing current data to these lines, you can draw conclusions about whether the process variation is consistent (in control or healthy state) or is unpredictable (out of control, affected by special causes of variation).

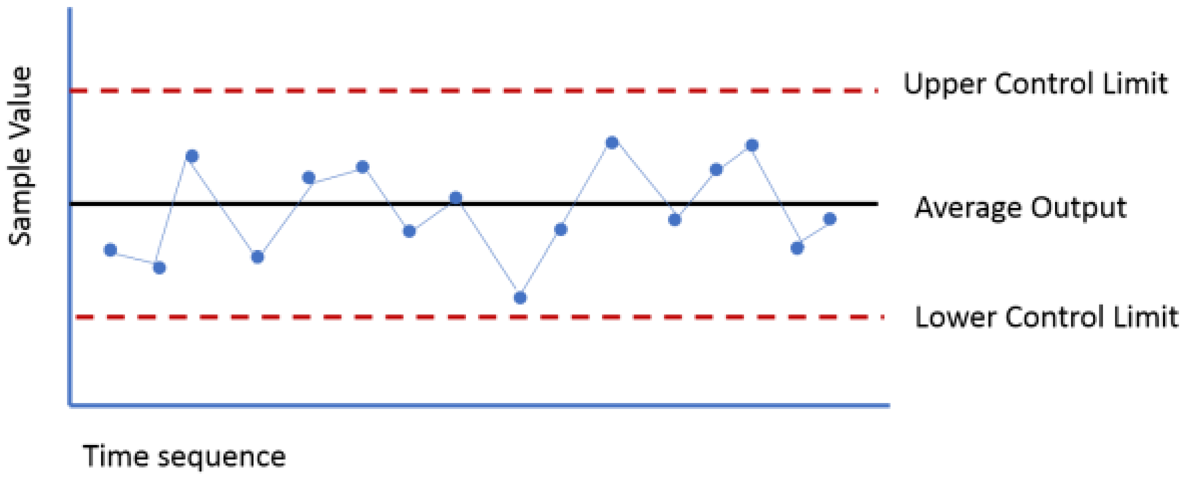


Figure 5. Control chart for SPC

.

I.8. Conclusion

The main purpose of this chapter is to provide certain concepts and terminology used in the subject of fault detection and diagnosis, as well as to explore several defect detection approaches from various perspectives. We additionally compared these techniques to a common set of desirable diagnostic system characteristics that we specified in section 3. In order to gain a basis in the field, a definition of SPC was researched. Due to the complexity of the FDD field, a detailed analysis of each approach is a time-consuming process. In the next chapters, we will concentrate on multivariate statistical techniques.

Chapter II.

Multivariate Statistical Approaches

II.1. Introduction

Due to consistent product quality demand and higher requirements in safety, the process monitoring performance has become a key factor in improving productivity and safety. Process systems are using large amount of data from many variables that are monitored and recorded continuously every day. Several multivariate statistical techniques for fault detection, analysis of process and diagnosis have been developed and used in practice [20].

This chapter is in three parts. In the first, PCA is defined, and what has become the standard derivation of PCs, in terms of eigenvectors of a covariance matrix, and it its use in Fault Detection is presented. However, the PCA identifies only linear structure in a given dataset, as it is nothing but a linear projection. To overcome this problem, many studies have been proposed to define nonlinear extensions of PCA. The kernel principal component analysis (KPCA) is among the most popular nonlinear statistical methods, the second part gives a review of the kernel PCA method for process monitoring, the process and the drawbacks. The last part discusses the proposed Reduced Kernel PCA, technique of reduction and its application in FDD.

II.2. Principal Component Analysis

Principal Component Analysis (PCA) is a multivariate statistical technique that analyzes data described by several inter-correlated quantitative dependent variables in order to extract the important information as new orthogonal variables called principal components. PCA depends upon the Eigen-decomposition of positive semidefinite matrices and upon the singular value decomposition (SVD) of rectangular matrices [21]

PCA is used for exploratory data analysis and examination of the relationships among a group of variables. Hence it can be used for dimensionality reduction. [22]

II.2.1. Statistical process modeling using PCA

The first step in the PCA algorithm is to construct a data or feature matrix $\boldsymbol{X}$, where each sample is represented as one column and the number of rows represents the dimension, note that $\boldsymbol{X} \epsilon\, \boldsymbol{R}^{\boldsymbol{N} \times \boldsymbol{m}}$ is a normalized data with zero mean and unit variance. Calculating the lower dimensional space with the PCA technique can be done using the covariance or SVD methods [23,24].

The covariance matrix is calculated as follows:

$$\boldsymbol{C} = \frac{1}{N-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X} \tag{2. 1}$$

The covariance matrix of $\boldsymbol{X}$ can be decomposed using SVD as:

$$\boldsymbol{C} = \boldsymbol{P\Lambda P}^{\boldsymbol{T}} \tag{2. 2}$$

Where the matrix $\boldsymbol{P} \boldsymbol{\epsilon} \boldsymbol{R}^{\boldsymbol{m} \times \boldsymbol{m}}$ is the principal component loading vectors. While, $\boldsymbol{\Lambda} \boldsymbol{\epsilon} \boldsymbol{R}^{\boldsymbol{m} \times \boldsymbol{m}}$ is a diagonal matrix of the eigenvalues of $\boldsymbol{Cov} \boldsymbol{\epsilon} \boldsymbol{R}^{\boldsymbol{m} \times \boldsymbol{m}}$ decreasingly ordered.

The data matrix is given as:

$$\boldsymbol{X} = \boldsymbol{TP}^{\boldsymbol{T}} \tag{2. 3}$$

where $\boldsymbol{T}$ is called the score matrix, and it is given by:

$$\boldsymbol{T} = \boldsymbol{XP} \tag{2. 4}$$

The aim of PCA method is to represents the data by fewer sufficient components. Thus, using $\boldsymbol{\ell} < \boldsymbol{m}$ of the components, by decomposing he matrix $\boldsymbol{P}$ can be decomposed into:

$$\boldsymbol{P} = [\ \widehat{\boldsymbol{P}} \quad \widetilde{\boldsymbol{P}}\ ] \tag{2. 5}$$

Where $\widehat{\boldsymbol{P}}$ contains only the ℓ first columns of $\boldsymbol{P}$. $\boldsymbol{\ell}$ Represents the number of retained principal component, the decomposition of the used matrices becomes:

$$\boldsymbol{P} = [\ \widehat{\boldsymbol{P}}_{\boldsymbol{m}\times\boldsymbol{l}} \quad \widetilde{\boldsymbol{P}}_{\boldsymbol{m}\times(\boldsymbol{m}-\boldsymbol{l})}\ ] \tag{2. 6}$$

$$\boldsymbol{T} = [\ \widehat{\boldsymbol{T}}_{\boldsymbol{n}\times\boldsymbol{l}} \quad \widetilde{\boldsymbol{T}}_{\boldsymbol{n}\times(\boldsymbol{m}-\boldsymbol{l})}\ ] \tag{2. 7}$$

$$\boldsymbol{\Lambda} = \begin{bmatrix} \widehat{\boldsymbol{\Lambda}}_{\boldsymbol{l}\times\boldsymbol{l}} & \boldsymbol{O}_{\boldsymbol{l}\times(\boldsymbol{m}-\boldsymbol{l})} \\ \boldsymbol{O}_{(\boldsymbol{m}-\boldsymbol{l})\times\boldsymbol{l}} & \widetilde{\boldsymbol{\Lambda}}_{(\boldsymbol{m}-\boldsymbol{l})\times(\boldsymbol{m}-\boldsymbol{l})} \end{bmatrix} \tag{2. 8}$$

The data matrix $\boldsymbol{X}$ can be decomposed as:

$$\boldsymbol{X} = \boldsymbol{X}\widehat{\boldsymbol{P}}\widehat{\boldsymbol{P}}^{\boldsymbol{T}} + \boldsymbol{X}\widetilde{\boldsymbol{P}}\widetilde{\boldsymbol{P}}^{\boldsymbol{T}} = \boldsymbol{XA} + \boldsymbol{X}(\boldsymbol{I}-\boldsymbol{A}) = \widehat{\boldsymbol{X}} + \boldsymbol{E} \tag{2. 9}$$

Where $\widehat{\boldsymbol{X}} = \boldsymbol{XA}$ and $\boldsymbol{E} = \boldsymbol{X}(\boldsymbol{I}-\boldsymbol{A})$ are the projection of the observed sample onto the principal component subspace and the residual subspace respectively [24]. Note that matrix $\boldsymbol{A}$ is idompotent.

II.2.2. Model dimension selection

The effectiveness of the PCA model depends on the number of principal components (PCs) are to be used for PCA. Selecting an appropriate number of PCs introduces a good performance of PCA in terms of processes monitoring. Several methods for determining the number of PCs have been proposed such as; the scree plot , the cumulative percent variance (CPV), the cross validation and the profile likelihood [20]. In this study here, the cumulative percent variance method is utilized to come up with the optimum number of retained principal components. The cumulative percent variance is computed as follows:

$$CPV = \frac{\sum_{i=1}^{i=l} \boldsymbol{\lambda}_i}{\sum_{i=1}^{i=m} \boldsymbol{\lambda}_i} \% \qquad \textbf{(2. 10)}$$

Where $\boldsymbol{l}$ is the number of the retained PC's having their sum of variances greater than a certain percentage of the total variance [24], the percentage chosen based on the CPV determines the quality of the constructed model.

II.2.3. Fault Detection Indices

For fault detection, the PCA model of the process is developed, based on normal operating process data, and then used to check new measurement data. The differences between the new measurement data and their projections to the built model, the residuals are then subjected to some sort of statistical test to determine if they are significant. Usually the SPE statistic, also called squared prediction error ($\boldsymbol{Q}$), and the Hotelling's ($\boldsymbol{T^2}$) statistic are used to represent the variability in the residual subspace and principal component subspace [26].

A third monitoring index $\boldsymbol{\varphi}$ is introduced, which is a combination of the $\boldsymbol{T}^2$ and $\boldsymbol{Q}$ indices, weighted by their control limits for monitoring the principal and residual space simultaneously [27].

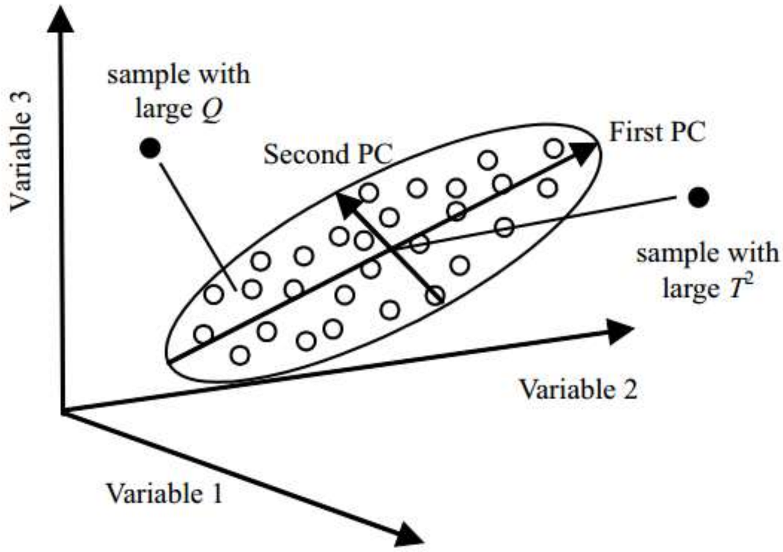


Figure 6. Data projection on two PCs [28]

II.2.3.1. Q-statistic or squared prediction error (SPE)

It is possible to detect new events by computing the squared prediction error SPE or $\boldsymbol{Q}$ of the residuals for a new observation. $\boldsymbol{Q}$ statistic [29], [30], is computed as the sum of squares of the residuals. Also, the $\boldsymbol{Q}$ statistic is a measure of the amount of variation not captured by the PCA model [20], it is defined as:

$$Q = \|E\|^{2} = \boldsymbol{X}(\boldsymbol{I}-\boldsymbol{A})(\boldsymbol{I}-\boldsymbol{A})^{\boldsymbol{T}}\boldsymbol{X}^{\boldsymbol{T}} \quad \textbf{(2. 11)}$$

The monitored system, meanwhile, is accepted to be in normal operation if:

$$Q \leq Q_{\alpha} \quad \textbf{(2. 12)}$$

The threshold of Q-statistic is calculated by the following relation [20]:

$$Q_{\alpha} = \theta_1 \left[\frac{c_{\alpha} h_0 \sqrt{2\theta_2}}{\theta_1} + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} + 1\right]^{1/h_0} \quad \textbf{(2. 13)}$$

Where $\theta_i = \sum_{j=l+1}^{i=m} \boldsymbol{\lambda}_{\boldsymbol{j}}^{\boldsymbol{i}}$ {i = 1,2,3}, $h_0 = 1 - \frac{2\theta_1\theta_3}{\theta_2^2}$ and $\boldsymbol{c}_{\boldsymbol{\alpha}}$ is the value of the normal distribution with $\boldsymbol{\alpha}$ is the level of significance at the instant of an unusual event, when there is a change in the covariance structure of the model, this change is going to be detected by a high value of $\boldsymbol{Q}$.

The threshold of $\boldsymbol{Q_\alpha}$ statistic can also be calculated by the following [31]:

$$Q_\alpha = g\chi^2_{h,\alpha} \tag{2. 14}$$

Where $g = \theta_2/\theta_1$ and $h = \theta_1^2/\theta_2$

II.2.3.2. Hotelling's T2 statistic

The Hotelling's $\boldsymbol{T^2}$ statistic measures the variability in the principle components subspace. It depends on the first eigenvalues that capture the most variations of data [32], it is determined by:

$$\boldsymbol{T^2 = X\widehat{P}\,\widehat{\Lambda}^{-1}\,\widehat{P}^{\,T}\,X^{\,T}} \tag{2. 15}$$

The threshold is calculated with conditions that the process is normal and the data follows a multivariate normal distribution according to the following relation:

$$T^2_\alpha = \frac{L(N^2-1)}{N(N-L)}F_{L,N-L,\alpha} \tag{2. 16}$$

Where $\boldsymbol{F_{L,N-L,\alpha}}$ is the Fisher normal distribution function with $\boldsymbol{L}$ and $\boldsymbol{N-L}$ degrees of freedom. $\boldsymbol{\alpha}$ is the level of significance (confidence interval). When the number of observations N is high, The $\boldsymbol{T^2}$ statistic threshold can be approximated with $\boldsymbol{\chi^2}$ distribution [32].

$$T^2_\alpha = g\chi^2_{h,\alpha} \tag{2. 17}$$

Where $\boldsymbol{g}$ and $\boldsymbol{h}$ can be set to 1.

II.2.3.3. φ Statistic

The combined index, $\boldsymbol{\varphi}$ statistic, has been first proposed by Yue and Qin .The index is a combination of both $\boldsymbol{T^2}$and $\boldsymbol{Q}$ statistics, it gives informations about the variability in the whole measurement space and used for monitoring the principal and residual space simultaneously[27]. For a new measurement vector $\boldsymbol{X}$, $\boldsymbol{\varphi}$ is defined:

$$\varphi = \frac{T^2}{T^2_\alpha} + \frac{Q}{Q_\alpha} = \boldsymbol{X^T wX} \tag{2. 18}$$

Where :

$$\boldsymbol{w} = \frac{\widehat{\boldsymbol{P}}\,\widehat{\boldsymbol{\Lambda}}^{-1}\,\widehat{\boldsymbol{P}}^{\boldsymbol{T}}}{T_\alpha^2} + \frac{\boldsymbol{I} - \widehat{\boldsymbol{P}}\,\widehat{\boldsymbol{P}}^{\boldsymbol{T}}}{Q_\alpha} \qquad \textbf{(2. 19)}$$

In a similar way, the $\boldsymbol{\varphi}$ statistic threshold can be determined by [32]:

$$\varphi_\alpha = g\chi^2_{h,\alpha} \qquad \textbf{(2. 20)}$$

$\boldsymbol{g}$ and $\boldsymbol{h}$ are given as follow [32,26] :

$$g = \frac{trace(Cov\,w)^2}{trace(Cov\,w)} \qquad \textbf{(2. 21)}$$

$$h = \frac{[trace(Cov\,w)]^2}{trace(Cov\,w)^2} \qquad \textbf{(2. 22)}$$

Table 02. PCA-based fault detection algorithm [24]

Algorithm 1: PCA-based fault detection algorithm
1. Offline monitoring - Obtain training fault free data set that represent the normal operations. - Scale the data to zero mean and unit variance. - Compute the covariance matrix $\boldsymbol{C}$ using **(2. 1)**. - Calculate the eigenvectors and eigenvalues of $\boldsymbol{C}$. - Determine how many principal components to be used. Many techniques can be used in this regard. In this work, the CPV criterion is used **(2.10).** - Calculate $\boldsymbol{Q_\alpha}$, $\boldsymbol{T^2}_{\boldsymbol{\alpha}}$ and $\boldsymbol{\varphi}_\alpha$ using **(2.13) (2.16) (2.20)** respectively. **2. Online monitoring** - Obtain testing data (possibly faulty data). - Scale the data using mean and variance of the training set. - Calculate $\boldsymbol{Q}$, $\boldsymbol{T^2}$ and $\boldsymbol{\varphi}$. - Check for faults: if $\boldsymbol{Q} \leq \boldsymbol{Q_\alpha}$, $\boldsymbol{T^2} \leq \boldsymbol{T^2}_{\boldsymbol{\alpha}}$ and $\boldsymbol{\varphi} \leq \boldsymbol{\varphi}_\alpha$, then declare a fault.

II.2.4. PCA main drawbacks

The selection of the optimal number of principal components (PCs) in fault detection using principal components analysis (PCA) is a critical and sensitive operation because overestimating the number of components will results in a contamination of the extracted information by adding noise dimensions with no useful information. This will lead to an important amount of false alarms. In the other hand, underestimating the number of components results in a loss of information and a misdetection of faults [33].

The PCA technique produces new uncorrelated variables called principal components (PC) with each component is linear combinations of original variables. However, the majority of processes, data have nonlinear relationships. In fact, PCA only defines a projection of linear data. Hence, it is incapable to analyze and represent the data with nonlinear characteristics. This limitation and nonlinearity problem have motivated various researchers to develop nonlinear extensions, such as Kernel PCA presented in the next section [34].

II.3. Kernel Principal Component Analysis (KPCA)

Kernel Principal Component Analysis (KPCA) is among the most popular dimensional reduction and analysis method. It extends the conventional linear PCA to deal with nonlinear modes[34] that can only be effectively performed on a set of observations that vary linearly. When the variations are nonlinear, the data can always be mapped into a higher-dimensional space in which they vary linearly. That is, according to Cover's theorem, the nonlinear data structure in the input space is more likely to be linearly separabale after high-dimensional nonlinear mapping [35].

The main objective of Kernel PCA is to model process data with non-linear structure. It consists to transform the nonlinear input data space into a linear data in a new high dimensional feature space denoted $\mathcal{F}$ and to perform PCA in that space, where the feature space $\mathcal{F}$ is nonlinearly transformed from the input space with a non-linear mapping function [34].

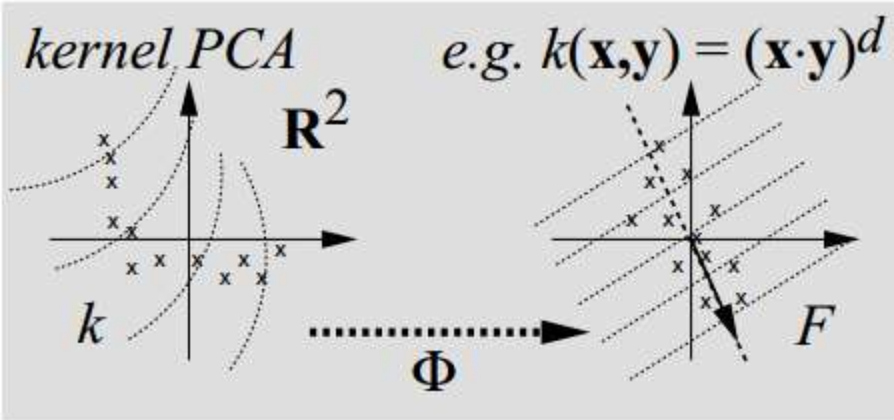


Figure 7. Non-linear data transformation to linear feature space $\mathcal{F}$ [48]

II.3.1. Statistical process modeling using KPCA

We assume a distribution consisting of N data points and m variabals $\boldsymbol{X_i} \boldsymbol{\epsilon} \boldsymbol{R^{N \times m}}$. Before performing linear PCA, these data points are mapped into a higher-dimensional feature space $\mathcal{F}$ through a non-linear mapping :

$$\phi: X_i \epsilon R^m \rightarrow \phi(X_i) = \phi_i \epsilon\ R^h \quad \textbf{(2. 23)}$$

Where $\boldsymbol{R^h}$ $(h >> m)$ be an arbitrarily large or possibly infinite dimension. The data is arranged in $\mathcal{F}$ as $\boldsymbol{\chi} = [\phi(x_1) \dots \phi(x_i) \dots \phi(x_N)]^{\boldsymbol{T}} \in \mathbb{R}^{\boldsymbol{N \times h}}$ define the data matrix in the feature space $\mathcal{F}$, then the covariance matrix $\boldsymbol{C_\phi}$ can be expressed as :

$$\boldsymbol{C_\phi} = \frac{1}{N} \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \boldsymbol{\phi}(\boldsymbol{X_i}) \boldsymbol{\phi}(\boldsymbol{X_i})^{\boldsymbol{T}} = \frac{1}{N} \chi \chi^{\boldsymbol{T}} \epsilon R^{h \times h} \quad \textbf{(2. 24)}$$

The principal components of the mapped data ϕ_i are computed by solving the eigenvalue decomposition of $\boldsymbol{C_\phi}$such that:

$$\boldsymbol{\lambda v} = \boldsymbol{C_\phi v} \quad \textbf{(2. 25)}$$

Where are $\boldsymbol{\lambda}$ eigenvalues and $\boldsymbol{v} \ \epsilon\ \mathcal{F}\{\boldsymbol{0}\}$ their corresponding eigenvectors. The importance of the eigenvectors is indicated by the magnitude of their corresponding eigenvalues.

All solutions $\boldsymbol{v}$ with $\boldsymbol{\lambda \neq 0}$ lie in the spam of $[\phi(x_1) \dots \phi(x_i) \dots \phi(x_N)]$ this has two useful consequences[k]. First, there exist coefficients $\boldsymbol{\alpha_i}$ such that:

$$\boldsymbol{v} = \sum_{i=1}^{N} \alpha_i \phi(X_i) \tag{2. 26}$$

Second, we may instead consider the set of equations:

$$\lambda \langle \phi(X_k), v \rangle = \langle \phi(X_k), C_\phi v \rangle \qquad k = 1, \dots, N \tag{2. 27}$$

Combining equations **(2. 26)** and **(2. 27)**, we get:

$$\lambda \sum_{i=1}^{N} \alpha_i \langle \phi(X_k), \phi(X_i) \rangle = \frac{1}{N} \sum_{i=1}^{N} \alpha_i \left\langle \phi(X_i), \sum_{j=1}^{N} \phi(X_j) \right\rangle \times \langle \phi(X_{ij}), \phi(X_i) \rangle \qquad \boldsymbol{k = 1, \dots, N} \tag{2. 28}$$

The trick herein is that the PCA can be computed such that the vectors ϕ_i appear only within scalar products. Thus, the mapping need not be explicitly computed and only the dot products of two vectors in the feature space are needed. Instead, we only work with a kernel function $k(x, y)$, which replaces the scalar product. This is called the kernel trick (35,36).

Now defining a matrix $\boldsymbol{K} \in \boldsymbol{R}^{\boldsymbol{N \times N}}$ by:

$$\boldsymbol{K}_{ij} = k(x_i, x_j) = \langle \phi(X_i), \phi(X_j) \rangle = \boldsymbol{\chi\chi^T} \tag{2. 29}$$

$$\boldsymbol{K} = \begin{bmatrix} \phi_1^T\phi_1 & \dots & \phi_1^T\phi_N \\ . & & . \\ . & \dots & . \\ . & & . \\ \phi_N^T\phi_1 & \dots & \phi_N^T\phi_N \end{bmatrix} = \begin{bmatrix} k(x_1, x_1) & \dots & k(x_1, x_N) \\ . & & . \\ . & \dots & . \\ k(x_N, x_1) & \dots & k(x_N, x_N) \end{bmatrix} \tag{2. 30}$$

The left hand side of equation **(2.28)** can be expressed as:

$$\boldsymbol{\lambda} \sum_{i=1}^{N} \alpha_i \langle \phi(X_k), \phi(X_i) \rangle = \boldsymbol{\lambda} \sum_{i=1}^{N} \alpha_i \boldsymbol{K_{ki}} \qquad \textbf{(2. 31)}$$

And the right hand side of equation **(2.28)** can be given by:

$$\frac{1}{N} \sum_{i=1}^{N} \alpha_i \left\langle \phi(X_i), \sum_{j=1}^{N} \phi(X_j) \right\rangle \times \langle \phi(X_{ij}), \phi(X_i) \rangle = \frac{1}{N} \sum_{i=1}^{N} \alpha_i \sum_{j=1}^{N} K_{ij} K_{ji} \qquad \textbf{(2. 32)}$$

this reads :

$$\boldsymbol{\lambda} N \boldsymbol{K} \boldsymbol{\alpha} = \boldsymbol{K}^2 \boldsymbol{\alpha} \qquad \textbf{(2. 33)}$$

where $\boldsymbol{\alpha}$ denotes the column vector with entries $\boldsymbol{\alpha_1}, \boldsymbol{\alpha_2}, \dots\ \boldsymbol{\alpha_N}$.

To find solutions of equation **(2.33)**, we solve the eigenvalue problem:

$$N \boldsymbol{\lambda} \boldsymbol{\alpha} = \boldsymbol{K} \boldsymbol{\alpha} \qquad \textbf{(2. 34)}$$

Eigen-decomposition of the kernel matrix K is equivalent to performing PCA in $\boldsymbol{R^h}$ [i]. This yields eigenvectors $\boldsymbol{\alpha_1};\ \boldsymbol{\alpha_2}; \dots ; \boldsymbol{\alpha_N}$ with eigenvalues $\boldsymbol{\lambda_1} \geq \boldsymbol{\lambda_2} \geq \cdots ; \geq \boldsymbol{\lambda_N}$. The dimensionality of the problem can be reduced by retaining only the 1st $\boldsymbol{l}$ eigenvectors using the cumulative percent variance method given in equation **(2.10)**.

We normalize $\boldsymbol{\alpha_1};\ \boldsymbol{\alpha_2}; \dots ; \boldsymbol{\alpha_N}$ by requiring that the corresponding vectors in $\boldsymbol{\mathcal{F}}$ be normalized, that is:

$$\langle v_k, v_k \rangle = 1 \qquad k = 1, \dots , \ell \qquad \textbf{(2. 35)}$$

Using eq **(2. 35)** and **(2. 26)** we get the normalization condition:

$$\begin{aligned} 1 &= \sum_{i,j=1}^{N} \alpha_i^k \alpha_j^k \langle \Phi(\mathrm{x}_i) \cdot \Phi(\mathrm{x}_j) \rangle = \sum_{i,j=1}^{N} \alpha_i^k \alpha_j^k K_{ij} \\ &= \langle \alpha^k \cdot \boldsymbol{K} \alpha^k \rangle = \lambda_k \langle \alpha^k \cdot \alpha^k \rangle \end{aligned} \qquad \textbf{(2. 36)}$$

Thus,

$$\langle \alpha_k, \alpha_k \rangle = \frac{1}{\lambda_k} \qquad k = 1, \dots, \ell \tag{2. 37}$$

Which shows that :

$$v_i = \sum_{j=1}^{N} \frac{\alpha_j^i}{\sqrt{\lambda_i}} \phi_j = \frac{1}{\sqrt{\lambda_i}} \boldsymbol{\chi}^{\boldsymbol{T}} \boldsymbol{\alpha}^{\boldsymbol{i}} \tag{2. 38}$$

We denote $\widehat{\boldsymbol{P}}_{\boldsymbol{N}\times\boldsymbol{l}} = [v_1, v_2, \dots, v_l]$ the principal subspace, which is denoted:

$$\widehat{\boldsymbol{P}}_{\boldsymbol{N}\times\boldsymbol{l}} = \left[\boldsymbol{\chi}^{\boldsymbol{T}} \alpha_1 \lambda_1^{-\frac{1}{2}}, \dots, \boldsymbol{\chi}^{\boldsymbol{T}} \alpha_\ell \lambda_\ell^{-\frac{1}{2}}\right] = \boldsymbol{\chi}^{\boldsymbol{T}} \widehat{\boldsymbol{P}}^{\boldsymbol{T}} \widehat{\boldsymbol{\Lambda}}^{-\frac{1}{2}} \tag{2. 39}$$

By projecting $\boldsymbol{\phi}(\boldsymbol{X})$ onto principl eigenvectors in the feature space F we get [24,35]:

$$\hat{\boldsymbol{t}} = \widehat{\boldsymbol{P}}^{\boldsymbol{T}} \boldsymbol{\phi} = \boldsymbol{\Lambda}^{-\frac{1}{2}} \widehat{\boldsymbol{P}}^{\boldsymbol{T}} k(x) \in \mathbb{R}^{\boldsymbol{\ell}} \tag{2. 40}$$

Or :

$$\hat{\boldsymbol{t}}_{\boldsymbol{k}} = \langle v_k, \phi(X) \rangle = \sum_{i=1}^{N} \alpha_i^k \langle \phi(X_i), \phi(X) \rangle, k = 1, \dots, k \tag{2. 41}$$

II.3.1.1. Computing Dot Products in Feature Spaces

The kernel representations allow us to compute the value of the dot product in F without having to carry out the map $\boldsymbol{\phi}$. This method substitute a priori chosen kernel functions K for all occurrences of dot products[k], several kernel functions exists:

Laplacian kernel :

$$\mathrm{k}(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|}{\sigma}\right) \tag{2. 42}$$

Gaussian kernel [Radial Basis Function (RBF)]:

$$\mathrm{k}(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad \textbf{(2. 43)}$$

Where $\boldsymbol{\sigma}$ is specied a priori by the user.

Polynomial function:

$$\mathrm{k}(x_i, x_j) = \langle x_i, x_j \rangle^d \quad \textbf{(2. 44)}$$

where $\boldsymbol{d}$ is a positive integer.

Sigmoid function:

$$\mathrm{k}(x_i, x_j) = \tanh(\beta_0 \langle x_i, x_j \rangle + \beta_1) \quad \textbf{(2. 45)}$$

where $\boldsymbol{\beta_0}$ and $\boldsymbol{\beta_1}$ are fixed by the user to satisfy Mercer's theorem[34].

Kernel functions provides a low-dimensional KPCA subspace that represents the distributions of the mapping of the training vectors in the feature space. A specific choice of kernel function implicitly determines the mapping and the feature space $\boldsymbol{\mathcal{F}}$.[35]

Before applying KPCA, mean centering in the high dimensional space should be performed. This can be done by substituting the kernel matrix $\boldsymbol{K}$ with [35]:

$$\mathbf{K}_{\boldsymbol{c}} = \mathbf{K} - \mathbf{1}_{\boldsymbol{N}}\mathbf{K} - \mathbf{K}\mathbf{1}_{\boldsymbol{N}} + \mathbf{1}_{\boldsymbol{N}}\mathbf{K}\mathbf{1}_{\boldsymbol{N}} \quad \textbf{(2. 46)}$$

Where

$$\mathbf{1}_{\boldsymbol{N}} = \frac{1}{N}\begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix} \in \boldsymbol{R}^{N \times N} \quad \textbf{(2. 47)}$$

II.3.2. Fault detection using KPCA

The KPCA-based monitoring method is similar to that using PCA in that the three indices: $\boldsymbol{Q}$, The Hotelling's $\boldsymbol{T}^2$ and the combined and can be interpreted in the same way in the feature space [37].

II.3.2.1. Q-statistic

The squared prediction error ($\boldsymbol{Q}$) is usually used for fault detection using KPCA. However, the conventional KPCA does not provide any approach of data reconstruction in the feature space. Thus, the computation $\boldsymbol{Q}$ index is difficult in the KPCA method.

A proposed expression to calculate $\boldsymbol{Q}$ in the feature space H, which is shown as follows[38]:

$$Q = \mathrm{k}(x, x) - k_x^T \widehat{\boldsymbol{P}} \widehat{\boldsymbol{\Lambda}}^{-\mathbf{1}} \widehat{\boldsymbol{P}}^{\boldsymbol{T}} k_x \quad \textbf{(2. 48)}$$

The confidence limit for $\boldsymbol{Q}$ index can be calculated using the $\boldsymbol{\chi}^{\mathbf{2}}$-distribution and is given by:

$$Q_\alpha = g\chi^2_{h,\alpha} \quad \textbf{(2. 49)}$$

$\boldsymbol{g}$ is a weighting parameter included to account for the magnitude of $\boldsymbol{Q}$ and $\boldsymbol{h}$ accounts for the degrees of freedom. If $\boldsymbol{a}$ and $\boldsymbol{b}$ are the estimated mean and variance of the $\boldsymbol{Q}$ [35, 38] Where $\boldsymbol{g} = b/2a$ and $h = 2a^2/b$.

II.3.2.2. T²-statistic

The Hotelling's $\boldsymbol{T}^2$ index is calculated as:

$$\boldsymbol{T}^{\mathbf{2}} = \hat{\boldsymbol{t}}\, \widehat{\boldsymbol{\Lambda}}^{-\mathbf{1}}\, \hat{\boldsymbol{t}}^{\boldsymbol{T}} \quad \textbf{(2. 50)}$$

The $\boldsymbol{T}^{\mathbf{2}}$ is calculated using kernel functions as [38]:

$$\boldsymbol{T}^{\mathbf{2}} = k(x)^{\mathrm{T}} \widehat{\boldsymbol{P}} \boldsymbol{\Lambda}^{-\mathbf{1}} \widehat{\boldsymbol{P}}^{\mathbf{T}} k(x) \quad \textbf{(2. 51)}$$

The control limit for $\boldsymbol{T}^2$ is calculated using $\boldsymbol{\mathcal{F}}$ distribution as in conventional PCA given by equation **(2. 17)**.

III.3.2.3. φ Statistic

The combined index $\boldsymbol{\varphi}$ is proposed in [27],[25],That it aims to monitor the principal and residual space in the feature space simultaneously by combining the $\boldsymbol{Q}$ and $\boldsymbol{T}^2$ indices weighted by their thresholds, and it is defined as as in PCA eq **(2. 18)**.

The combined index $\boldsymbol{\varphi}$ has a control limit which is given in PCA or the same as $\boldsymbol{Q_\alpha}$ eq **(2. 49)**.

Table 03. KPCA-based fault detection algorithm [39].

Algorithm2: KPCA-based fault detection
1. Offline monitoring
- Acquire normal operating data and normalize it using the mean and standard deviation of each variable.
- Decide on the type of kernel function to use and determine the kernel parameter.
- Compute the kernel matrix $\boldsymbol{K}$ of the NOC using equation **(2.29)** and normalize it using equation **(2. 46)**.
- Solve the eigenvalue problem given in equation **(2.34)** and normalize the eigenvectors using equation **(2.38)**.
- Calculate the monitoring statistics ($\boldsymbol{\varphi}$, $\boldsymbol{T}^2$ and $\boldsymbol{Q}$) of the normal operating data.
- Determine the control limits of $\boldsymbol{\varphi}$, $\boldsymbol{T}^2$ and $\boldsymbol{Q}$ charts.
2. Online monitoring
- Obtain new data for each sample and scale it with the mean and variance of the model.
- compute the kernel vector of the normal new measured data.
- Mean center the test kernel vector using **(2. 46)**.
- Calculate the monitoring statistics ($\boldsymbol{\varphi}$, $\boldsymbol{T}^2$ and $\boldsymbol{Q}$) of the test data.
- Monitor whether $\boldsymbol{\varphi}$, $\boldsymbol{T}^2$ and $\boldsymbol{Q}$ exceeds its control limit calculated in the Offline monitoring.

II.3.3. KPCA main drawback

Kernel PCA is computationally expensive. Most time consuming is the extraction of the eigenvectors of $\boldsymbol{K}$, which is $\boldsymbol{O}(\boldsymbol{N}^{\mathbf{3}})$ if extracting all eigenvectors [40]. Kernel PCA is also memory exhaustive: the $\boldsymbol{N} \times \boldsymbol{N}$ matrix $\boldsymbol{K}$ need an $\boldsymbol{O}(\boldsymbol{N}^{\mathbf{2}})$ to be stored.

Furthermore, testing is expensive. For each new data point, the kernel function needs to be evaluated N-times. However, this number could be reduced using so-called ‘reduced set methods’ [36].

In order to surmount the problem of high computational cost and storage, we have investigated the use of a reduced version of KPCA called reduced kernel principal component (RKPCA)

II.4.Proposed approach Reduced Kernel principal component (RKPCA)

In the case of dynamic system, monitoring algorithms based on KPCA suffer from computation complexity as the amount of computer memory and the number of observations increases, which implies that it contains redundant information as well as noise due to measurement errors. To overcome this burden, possible solutions among others consist either to use kernel k means for clustering, to use an incremental approach for fast calculation of the kernel PCs , or to search a reduced data set that approaches sufficiently the system behavior with original model in the offline phase before applying it online [41].

In this section, we improve the use of the RKPCA method for monitoring by introducing the Fractal Dimension as the new reduction method for identification of the observations containing the principal information.

II.4.1. Fractal Dimension

Mandelbrot written in his book ‘The Fractal Geometry of Nature (1982)’ where he explained the fractal phenomena in nature: ‘A fractal is a shape made of parts similar to the whole in some way’.

Fractals are a relatively new mathematical concept for describing the geometry of irregularly shaped objects in terms of fractional numbers rather than integers. The key parameter for fractal analysis is the fractal dimension, which is a real noninteger number, differing from the more familiar Euclidean or topological dimension. The latter is an integer, with a value of one for a line, and two for a surface. The fractal dimension for a line of any shape varies between one and two, and for a surface of any shape, between two and three. This ratio provides a statistical index of complexity comparing how detail in a pattern changes with the scale at which it is measured. Fractal fragmentation theory provides a means by which the entire size distribution of material can be quantified [42, 44]

There exist several techniques for the determination of the Fractal Dimension, we mention : The Hausdorff dimension which is purely a description of the geometry of the fractal set, Box counting dimension that has a number of practical limitations, particularly at a high embedding dimension, and so a variety of other algorithms have also been developed. The most popular method to compute dimension is to use the correlation Dimension, which estimates dimension based on the statistics of pairwise distances [43].

In this study we are going to use the correlation algorithm based on the comparison done in [43]

II.4.2. Correlation dimension

As the most widely used quantitative parameter to describe attractors, correlation dimension is a measure of the complexity of the system related with its degrees of freedom [45]. It is obtained from the correlations between random points on the attractor. Consider the set $\{ Xi, i = 1 \ldots N \}$ of points on the attractor, obtained e.g. from a time series, i.e. $Xi = X(t + iz)$ with a fixed time increment $\boldsymbol{z}$ between successive measurements. Due to the exponential divergence of trajectories, most pairs (Xi, Xj) with $\boldsymbol{i} \neq \boldsymbol{j}$ will be dynamically uncorrelated pairs of essentially random points. The points lie however on the attractor. Therefore they will be spatially correlated. We measure this spatial correlation with the correlation integral $\boldsymbol{C(r)}$[46], defined according [43] as:

$$C(r) = \frac{1}{N(N-1)} \sum_{i \neq j} H\left(r - \|X_i - X_j\|\right) \tag{2. 52}$$

Where $\boldsymbol{H}$ is the Heaviside function. In words,

$$C(r) = \frac{\#\ of\ distances\ less\ than\ r}{\#\ of\ distances\ all\ together}$$

$\boldsymbol{C}\ (\boldsymbol{R})$ is a measure of the probability that two arbitrary points $\boldsymbol{X_i}, \boldsymbol{X_j}$ will be separated by a distance less than $\boldsymbol{r}$ [45].

If the number of data and the embedding dimension are adequately large, the correlation dimension is given by [45]:

$$D = \lim_{r \to 0} \frac{\log C(r)}{\log r} \tag{2. 53}$$

However, there are a variety of practical issues and potential pitfalls that come with making an estimate from finite data. Extracting dimension directly from the correlation integral according to Eq. **(2.53)** is extremely inefficient, since the convergence to $\boldsymbol{D}$ as $\boldsymbol{r} \to \boldsymbol{0}$ is logarithmically slow. Taking a slope solves this problem. [43]

II.4.3. Data Reduction using Correlation Dimension

The idea is to compute the Correlation dimension of the original data, then take off observations one by one and compute the corresponding new dimension, only $\boldsymbol{R}$ samples that changes the dimension when deleted are retained. The process is explained in the algorithm 3.

Table 4. Data reduction using Correlation Dimension.

Algorithm 3: Data reduction using Correlation Dimension
1. Calculating the fractal dimension : - Acquire normal operating data $\boldsymbol{x}$ and normalize the data using the mean and standard deviation of each variable, to get $\boldsymbol{X}$. - Calculate the Euclidian distance $\boldsymbol{r}$ and the correlation integral $\boldsymbol{C}(\boldsymbol{r})$ using **(2. 52)**. - Draw $\boldsymbol{logC}(\boldsymbol{r})$ versus $\boldsymbol{log\ r}$. - Extract the slope $\boldsymbol{s}$, where the fractal dimension $\boldsymbol{Df} = \boldsymbol{s}$. **2. Reducing data :** -For each $\boldsymbol{i} = \boldsymbol{1}, \dots, \boldsymbol{N}$ and size of $\boldsymbol{X_{red}} < \boldsymbol{D_f}$: - Delete $\boldsymbol{x}_i$. - Calculate the new fractal dimension $\boldsymbol{D}_i$, following the same steps of part 1. - If $\boldsymbol{D}_i = \boldsymbol{D}_f$, go to the next sample. - Else, restore $\boldsymbol{x}_i$.

After the reduction, KPCA is applied on the new reduced data.

II.5. Conclusion

During this chapter, the theoretical background of PCA models was established as well as the selection of the model parameters, such as the number of PCs and the loadings. Furthermore, KPCA approach and its theoretical background has been proposed to cope with the problem of linearity. However, Due to its high computational cost, the RKPCA technique was introduced with its theory to deal with this issue. The foundations of fault detection using this three methods was presented using the well-known Hotelling's $\boldsymbol{T}^2$, sum squared error $\boldsymbol{Q}$ and the combined index $\boldsymbol{\varphi}$.

Chapter III.

Application, Results & Discussion

III.1. Introduction

In this chapter, the introduced RKPCA method based on Fractal Dimension is tested on the data collected from a cement factory under Normal Operation Conditions and different fault types (abrupt, incipient and intermittent) that can occur and disturb the normal operation of the system. Then compare it with results of other techniques: PCA, different types of KPCA, RKPCA based on Euclidean distance. The comparison is done in terms of False Alarm Rate (FAR), Missed Alarm Rate (MDR), Detection Time Delay (DTD) and the Cost function (J).

III.2. Process description

The first step in the dry cement production process is to produce flour like raw material by milling limestone, clay and iron ore mix. This raw material is then fed to the kiln system at the upper end of the preheat tower which is composed of a series of suspending cyclones where heat exchange between feed material and hot gas stream exhausting the rotary kiln is made. Drying, dehydration and carbon expulsion are initiated wherein. Afterwards, in the rotary kiln which is a huge rotating furnace, several chemical reactions occur between calcium and silicon dioxide yielding a new chemical structure called clinker. In the kiln discharge, a cooler system is used to cool hot clinker to preserve its properties using forced air. As final step, clinker and natural gypsum are milled together to get what is commonly known as cement.

Ain El Kebira cement plant (first production line) shown in Figure 8 , where our case study is conducted, is located in the east of Algeria. It has a rotary kiln of 5:4m shell diameter and 80m length, with 3◦ incline. The kiln can be rotated up to 2:14 rpm as maximum speed using two 560KW s asynchronous motors and produces clinker with density from 1300 to 1450kg=m3 under the normal operation. The plant works with two natural gas burners, the main one placed in the discharge end and the secondary located in the first level of the pre-heater tower without any tertiary air conduct. The data used to monitor the process is recorded from 44 sensors. The sensors measure temperatures, pressures, speeds, and motors current. Table 5 illustrate the different variables used in the process monitoring. In general, two main datasets are collected from the plant which is used to develop and validate RKPCA model, and afterwards to test under different faulty situations the established monitoring scheme [24].

Figure 8. Ain El Kebira cement plant [24]

Table 5. Process variables of the cement rotary kiln [24]

Signal	Description	Unit
1, 3, 5, 7	Depression of gases in outlets of cyclones: one, two, three, and four respectively, tower I	mbar
2, 4, 6, 8	Temperature of gases in outlets of cyclones: one, two, three, and four respectively, tower I	°C
10	Depression of gas in inlet of cyclone four, tower I	mbar
17, 19, 21, 23	Depression of gases in outlets of cyclones: one, two, three, and four respectively, tower II	mbar
18, 20, 22, 24	Temperature of gases in outlets of cyclones: one, two, three, and four respectively, tower II	°C
12, 25	Temperature of the material entering the kiln from tower I, and tower II respectively	°C
9, 15	Power of the motor driving the exhauster fans of tower I, and tower II respectively	kW
11, 16	Speed of the exhauster fans of tower I, and tower II respectively	r.p.m
13	Depression of gas in the outlet of the smoke filter of tower I	mbar
14, 26	Temperature of gas in the outlet of the smoke filter: tower I, and tower II respectively	°C
27	The sum of the powers of the two motors spinning the kiln	kW
28	Temperature of excess air from the cooler	°C
31	Temperature of the secondary air	°C
29, 32, 33	pressure of air under the static grille, repression of: fan I, fan II, and fan III respectively	mbar
30, 34	Speed of the cooling fan I, and fan III respectively	r.p.m
35, 37, 39	pressure of air under the chamber I, II, and III of the dynamic grille, repression of fan IV, fan V, and fan VI respectively	mbar
36, 38, 40	Speed of cooling fan IV, fan V, and fan VI respectively	r.p.m
41	Speed of the dynamic grille	strokes/min
42	Command issue of the pressure regulator for the speeds of the draft fans of cooler filter	r.p.m
43	Flow of fuel (natural gas) to the main burner	m^3/h
44	Flow of fuel (natural gas) to the secondary burner (pre-calcination level)	m^3/h

III.3. Computation of monitoring performance metrics

Monitoring performance was based on five metrics: False Alarms rate (FAR), Missed Alarms rate (MDR), Detection Time Delay (DTD), the Cost function J.

III.3.1. False Alarms Rate

Calculated as the percentage of faulty samples under healthy state, and given by:

$$\boldsymbol{FAR} = \frac{\#\ of\ faults\ under\ healthy\ state}{\#\ of\ samples\ under\ healthy\ state} \times \mathbf{100}\% \qquad \textbf{(2. 53)}$$

III.3.2. Missed Detection Rate

Calculated as the percentage of healthy samples under faulty state, and given by:

$$\boldsymbol{MDR} = \frac{\#\ of\ healthy\ samples\ under\ faulty\ state}{\#\ of\ samples\ under\ healthy\ state} \times \mathbf{100}\% \qquad \textbf{(2. 54)}$$

III.3.3. Detection Time Delay

Defined as the time required for indicating the fault after its occurrence.

$$\boldsymbol{DTD} = time\ of\ detection - time\ of\ occurance \qquad \textbf{(2. 56)}$$

III.3.4. The Cost function

The cost function is a way to evaluate the results in terms of performance matrices. Using the three evaluation criteria introduced previously, a cost function is of the form :

$$\boldsymbol{J} = \boldsymbol{q_1}\frac{\boldsymbol{FAR}}{\boldsymbol{FAR_d}} + \boldsymbol{q_2}\frac{\boldsymbol{MDR}}{\boldsymbol{MDR_d}} + \boldsymbol{q_3}\frac{\boldsymbol{DTD}}{\boldsymbol{DTD_d}} \qquad \textbf{(2. 57)}$$

Where $\boldsymbol{FAR_d}$, $\boldsymbol{MDR_d}$ and $\boldsymbol{DTD_d}$ are the desired values of $\boldsymbol{FAR}, \boldsymbol{MDR}$ and $\boldsymbol{DTD}$ respectively. Whereas $\boldsymbol{q_1}, \boldsymbol{q_2}$ and $\boldsymbol{q_3}$ are their respective weights.

III.4. Application procedure

This work is based on the real-time data collected by process computers from the cement plant. Table 6 lists the different data sets used in order to construct and test the proposed RKPCA then evaluate and compare its performances to that of PCA and KPCA.

Table 6. Data sets used in the application [24].

Data sets	Size	Sampling interval (s)	description
Training set	$X_0 \in R^{768\times44}$	20	Normal operation data, used to construct FD scheme
Testing set	$X_0 \in R^{11000\times44}$	1	Normal operation data, used to test the FD scheme
Process fault	$X_0 \in R^{2084\times44}$	1	Normal/Faulty operation, process fault
Sensor faults (6 sets)	$X_0 \in R^{1500\times44}$	1	Sensor fault simulations

The dataset is divided into training, testing, faulty sub-datasets. The training dataset consists of 768 observations (one sample each 20s) which are collected under the healthy operating conditions of the plant during 4 hours and 15 min. It is used to extract a reduced training dataset via Correlation Dimension by which a RKPCA model is built. While the testing dataset contains 11000 samples (one sample each second) which is used to test and validate the developed RKPCA model. The faulty dataset includes 1500 samples for each subset, where various simulated sensors faults are carried out to evaluate the monitoring scheme performance. These simulated faults Sfault1 to Sfault7 lie in the interval {500 – 1000}, in the intermittent case, the fault is introduced in the intervals {500-580, 610-660, 700-740, 800-830, 870-900 and 975-1000} with amplitude 5.5%, 4.5%, 5%, 5.5%, 5% and 4.5% respectively[24]. The faults characteristics are briefly and adequately described and reported in Table 7. [24]

Table 7. Simulated sensor faults introduced at 500-1000 s [24].

Fault	Faulty variables	Fault magnitude	Description of the fault
SFault(1)	16	(0, 5%)	Additive random fault, with mean 0, and variance 0.05
SFault(2)	44	−2%	Abrupt additive fault, bias $b = -0.02$
SFault(3)	30	+2%	Additive fault: Linear drift from 0% to 2%; slope $K_s = 4 \times 10^{-5}$
SFault(4)	34	−2%	Additive fault: Linear drift from 0% to −2%; slope $= -4 \times 10^{-5}$
SFault(5)	12, 18, 43	[+, −, +]2%	Abrupt additive fault ±2% (multiple)
SFault(6)	4, 6, 8, 14, 24	[+,+,+,−,−]2%	Additive fault: Linear drift from 0% to ±2% (multiple); slope $= \pm 4 \times 10^{-5}$
SFault(7)	11	+4.5%–5.5%	Additive fault: Intermittent fault, changing intervals and amplitudes

III.4.1.Application of PCA, KPCA and the proposed approach RKPCA to cement rotary kiln process

The steps to follow to apply PCA, KPCA and the proposed approach RKPCA previously described, on the cement plant, to build and evaluate the models are explained.

III.4.2.1. PCA monitoring model

To build PCA monitoring model, the training data is normalized then used to build the PCA model by using the steps described in **Table2 (Algorithm 1)**. The CPV is set to 90% of the total variance in the training dataset, which results in retaining PC's. Then, the statistical control limits $\boldsymbol{Q_\alpha}, \boldsymbol{T_\alpha^2}$ and $\boldsymbol{\varphi_\alpha}$ at the confidence 99% are determined.

The performance of the PCA model is evaluated using the monitoring performance metrics $\boldsymbol{FAR}, \boldsymbol{MDR}, \boldsymbol{DTD}$ and the Cost Function **J**. In this study, $\boldsymbol{FAR}, \boldsymbol{MDR}$ and $\boldsymbol{DTD}$ are of the same importance $\boldsymbol{q_1 = 1}, \boldsymbol{q_2 = 1}$ and $\boldsymbol{q_3 = 1}$ and $\boldsymbol{FAR_d}$, $\boldsymbol{MDR_d}$ and $\boldsymbol{DTD_d}$ values are 1%, 1% and 1 respectively.so the cost function is of the form:

$$\boldsymbol{J = FAR + MDR + DTD} \qquad \textbf{(2. 58)}$$

III.4.2.2. KPCA monitoring model

The **Algorithm 2** is followed to obtain the KPCA model using the training data after normalization. The same CPV is set to 90% of the total variance in the training data, which results in retaining the PC's. Another, important parameter for kernel-based methods in model development for process monitoring is the choice of the kernel function and its width. The radial basis kernel $\mathbf{k}(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$ is utilized in this work and normalized using equation **(2.40)**. The value of the kernel parameter c depends on the process being monitored and has been set to $c = 2\sigma^2 = 13\ 000$ in this application.

Then, the statistical control limits $\boldsymbol{Q_\alpha}, \boldsymbol{T_\alpha^2}$ and $\boldsymbol{\varphi_\alpha}$ at a 99% confidence level are determined.

These parameters are used to evaluate $\boldsymbol{FAR}, \boldsymbol{MDR}$ and $\boldsymbol{DTD}$**,** and calculate the Cost function **(2.58).**

For KPCA modeling, an important criteria is taken in consideration: the execution time. Kernel PCA is known by its high computational and storage problem and our purpose is to decrease this time by the proposed RKPCA approach.

III.4.2.3. Proposed RKPCA monitoring model

The reduction is done following the Algorithm 3, $\boldsymbol{r}$ and $\boldsymbol{C}$ produced are vectors of $\frac{\boldsymbol{m}\times(\boldsymbol{m}-\mathbf{1})}{\mathbf{2}}$ elements. To determine the slope from $\boldsymbol{log}(\boldsymbol{C})$ vs $\boldsymbol{log}(\boldsymbol{r})$ curve, we take the second third of the points and draw a fitting line through them. At the end of the algorithm we end up with $\boldsymbol{R}$ number of samples equal to the fractal dimension.

After the normalization of the new $\boldsymbol{R}\times\boldsymbol{R}$ reduced training data matrix, a conventional KPCA model is applied, then, the statistical control limits $\boldsymbol{Q}_{\boldsymbol{\alpha}}, \boldsymbol{T}_{\boldsymbol{\alpha}}^{\mathbf{2}}$ and $\boldsymbol{\varphi}_{\boldsymbol{\alpha}}$ at a confidence level of 99% are determined. The new model is then evaluated using the monitoring performance metrics $\boldsymbol{FAR}, \boldsymbol{MDR}, \boldsymbol{DTD}$ and the Cost Function **J** of the form **(2. 58).**

III.5. Results and discussion

III.5.1. PCA monitoring model

The results of PCA monitoring model are taken from [49].

The CPV is set to 90%, this results in retaining 20 PC's($\ell = 20$). The thresholds $\boldsymbol{T_\alpha^2}$, $\boldsymbol{Q_\alpha}$ and $\boldsymbol{\varphi_\alpha}$ deduced from the training set are computed using 99% as the confidence interval, the limits are $\boldsymbol{T_\alpha^2 = 51.13}$, $\boldsymbol{Q_\alpha = 1.525 \times 10^{-4}}$ and $\boldsymbol{\varphi_\alpha = 1.7238}$

The performance of the fault detection model based PCA in terms of FAR set is summarized in table 8. From this table, it can be seen that globally the three monitoring indices showed good results of FAR. The FAR contributed by the PCA model in the training set are good. This clearly indicates the accuracy of the model. In the testing set, a low FAR is obtained for the rotary kiln monitoring using PCA. Besides, it is even negligible when using a confidence level of 99%.

Table 08. FAR contributed by $\boldsymbol{T}^2, \boldsymbol{Q}$, $\boldsymbol{\varphi}$ under NOC using PCA

Method	**NOC Data**	**index**	**FAR (%)**
PCA	Training data	T^2	1.04
		Q	1.04
		φ	1.04
	Testing data	T^2	0.08
		Q	0.65
		φ	0.57

Table 9. Missed detection rate **(MDR)**, False Alarm Rate **(FAR),** detection time delay **(DTD)** and the cost function **J** values for the eight faults of cement rotary kiln using PCA using 99% as the confidence interval.

Faults	Performances	T^2	Q	φ
Process fault	MDR FAR DTD	1.98 1.67 30.00	1.68 14.05 0.00	1.14 14.52 2.00
	J	**33.65**	**15.73**	**17.66**
Random and single	MDR FAR DTD	2.40 0.30 1.00	1.20 0.80 1.00	1.40 0.60 1.00
	J	**3.70**	**3.00**	**3.00**
Abrupt and single	MDR FAR DTD	0.00 0.90 0.00	7.60 1.70 0.00	0.00 1.30 0.00
	J	**0.90**	**9.30**	**1.30**
Drift and single 1	MDR FAR DTD	6.80 0.70 27.00	10.80 0.40 5.00	5.80 1.10 27.00
	J	**34.50**	**16.20**	**33.90**
Drift and single 2	MDR FAR DTD	21.20 0.10 90.00	2.20 0.10 1.90	2.40 0.10 10.50
	J	**111.30**	**4.20**	**13.00**
Abrupt and multiple	MDR FAR DTD	0.00 0.20 0.00	0.00 1.50 0.00	0.00 0.60 0.00
	J	**0.20**	**1.50**	**0.60**
Drift and multiple	MDR FAR DTD	22.60 1.00 111.40	15.60 1.30 71.40	15.80 1.50 54.90
	J	**135.00**	**88.30**	**72.20**
Intermittent and single	MDR FAR DTD	0.00 1.04 0.00	0.00 1.85 0.00	0.00 1.37 0.00
	J	**1.04**	**1.85**	**1.37**
J average		40.03	17.51	17.87
J total			25.13	

III.5.2. KPCA monitoring model

III.5.2.1. NOC results

The thresholds $\boldsymbol{Q_\alpha}, \boldsymbol{T^2_\alpha}$ and $\boldsymbol{\varphi_\alpha}$ with a 99% confidence level deduced from the training data are found to be:

$$\boldsymbol{T^2_\alpha = 39.0767 \qquad Q_\alpha = 3.534 \times 10^{-4} \qquad \varphi_\alpha = 2.1944}$$

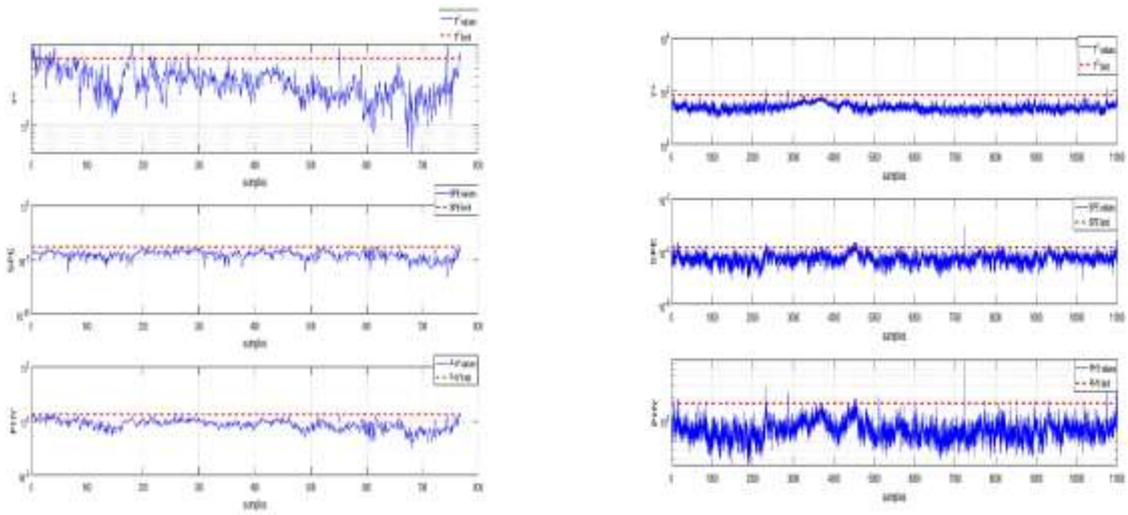
Performances of the fault detection model based KPCA in terms of FAR are tabulated in table10.

The FAR values provided by the three monitoring indices $\boldsymbol{T^2}, \boldsymbol{Q}$ and $\boldsymbol{\varphi}$ shows good results which confirm the accuracy of the KPCA model. In the validation phase, the model has provided the best performances in the three monitoring indices using 99% confidence interval

Table 10. FAR contributed by $\boldsymbol{T}^2, \boldsymbol{Q}, \boldsymbol{\varphi}$ under NOC using KPCA.

Method	**NOC Data**	**index**	**FAR (%)**
KPCA	Training data	T^2	2.994792
		Q	0.651042
		φ	0.911458
	Testing data	T^2	2.809091
		Q	0.518182
		φ	0.336364

Figure 9 shows the monitoring results of KPCA model in normal operation (training and testing set) using the 3 indices. The horizontal line in red represents the index threshold with a confidence level of 99%



a) Training data set b) Testing data set

Figure 9. T^2, Q and φ monitoring results of healthy process operation using KPCA.

(a) testing data set , (b) training data set.

III.5.2.2. Involuntary real process fault results

Table 11 summarizes the performances of KPCA-based model in terms of MDR, FAR, DTD and Cost function **J**.

The table shows that the detection time of the fault **Drift and multiple** is delayed when the three monitoring indices are used with the confidence level 99%.

In terms of FAR and MDR, high amount of false alarms is shown by the three indices except for $\boldsymbol{\varphi_\alpha}$ where FAR equals to 2.67. Whereas, the amount of missed detected samples is approximately zero and also for the delay of the detection (seconds).

In order to give a global view of the performances in detecting this fault, the cost function is used. Minimum values of **J** are seen when the thresholds $\boldsymbol{T^2_\alpha}$ and $\boldsymbol{\varphi_\alpha}$are utilized.

Table 11. Missed detection rate **(MDR)**, False Alarm Rate **(FAR),** detection time delay **(DTD)** and the cost function **J** values for the eight faults of cement rotary kiln using KPCA using 99% as the confidence interval.

Faults	Performances	T^2	Q	φ
Process fault	MDR	0.06	0.75	0.25
	FAR	7.57	10.02	2.67
	DTD	1	4	1
	J	**8.63**	**14.77**	**3.92**
Random and single	MDR	5.58	1.19	1.79
	FAR	0.6	0.4	0
	DTD	0	0	0
	J	**6.18**	**1.59**	**1.79**
Abrupt and single	MDR	0	5.78	0
	FAR	1.2	1	0.9
	DTD	0	0	0
	J	**1.2**	**6.78**	**0.9**
Drift and single 1	MDR	4.99	10.57	6.98
	FAR	1.9	0.3	0.5
	DTD	19	6	28
	J	**25.89**	**16.87**	**35.48**
Drift and single 2	MDR	15.16	2.39	4.79
	FAR	0.3	0	0
	DTD	1	6	11
	J	**16.46**	**8.39**	**15.79**
Abrupt and multiple	MDR	0	0	0
	FAR	0.63	0.94	0.21
	DTD	1	0	0
	J	**1.63**	**0.94**	**0.21**
Drift and multiple	MDR	18.16	15.96	18.36
	FAR	1.8	0.8	0.9
	DTD	110	110	110
	J	**129.96**	**126.76**	**129.26**
Intermittent and single	MDR	0	0	0
	FAR	4.13	1.9	1.66
	DTD	0	0	0
	J	**4.13**	**1.9**	**1.66**
J average		24.26	22.25	23.62
J total			23.38	

Figure 10 shows the monitoring results of KPCA model applied on the real process fault using the three indices. The horizontal line in red represents the index threshold with a confidence level of 99%.

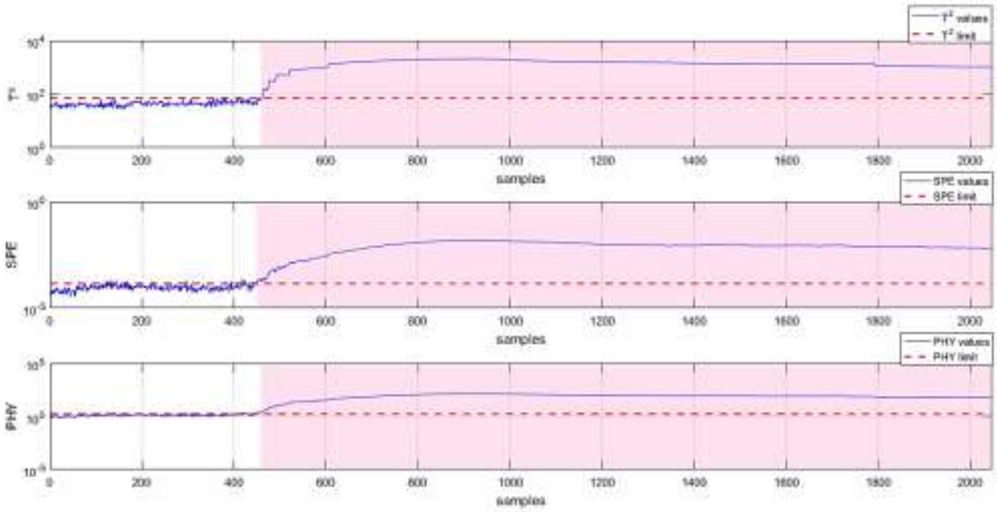


Figure 10. T^2, Q and φ monitoring results of real involuntary process fault in the cement rotary kiln using KPCA

III.5.2.3. Simulated sensor faults results

The performances of the KPCA based model in detecting all these faults are summarized in table 11.

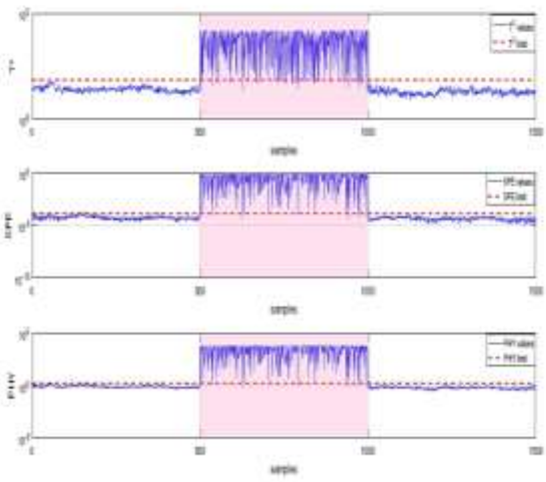
From the table, abrupt fault in single sensor (Sfault2) or multiple sensors (Sfault5) and intermittent fault in single sensor (Sfault7) are detected immediately and efficiently with no delay and zero amounts of missed detected samples. Furthermore, the quantity of false alarms is insignificant.

The KPCA based model has also promptly detected the random fault (Sfault1) and the drift fault in single sensor (Sfault4) with the three monitoring indices. Whereas the detection is delayed in the drift fault in single sensor (Sfault3) to 20s and to 1min30s when using to detect the drift fault in single sensor (Sfault6).

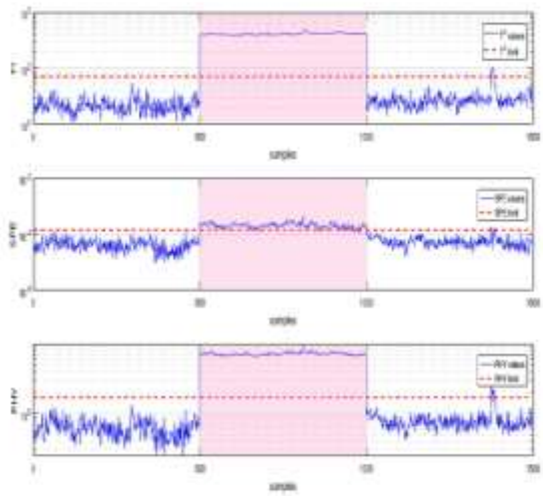
In terms of FAR, negligible values in detecting random fault in single sensor (Sfault1) and drift fault in single sensor (Sfault3, Sfault4) are noticed. Similarly, the amounts of missed alarms are acceptable except the detection of: Sfault3, Sfault4 and Sfault6 using the monitoring indices.

Globally, the performance of KPCA model in terms of average **J** has shown that random fault (Sfault1), abrupt fault in single sensor (Sfault2) or multiple sensors (Sfault5) and intermittent fault in single sensor (Sfault7) are efficiently detected particularly. Whereas, drift fault in single sensor (Sfault4) are effectively detected using the statistic $\boldsymbol{Q}$ and as well as (Sfault3) using the $\boldsymbol{Q}$. Drift fault in multiple sensors (Sfault6) has shown high values of **J** in the three monitoring indices.

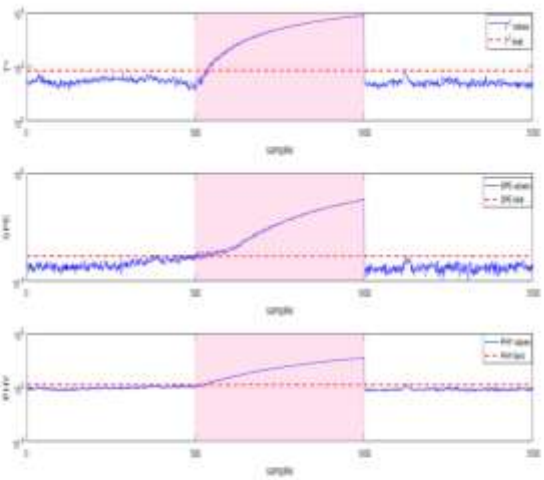
Figure 11 show the 3 indices monitoring results using KPCA. The detection of single as well as multiple sensor faults of abrupt, random, intermittent, and drift types can be noticed from the graphs



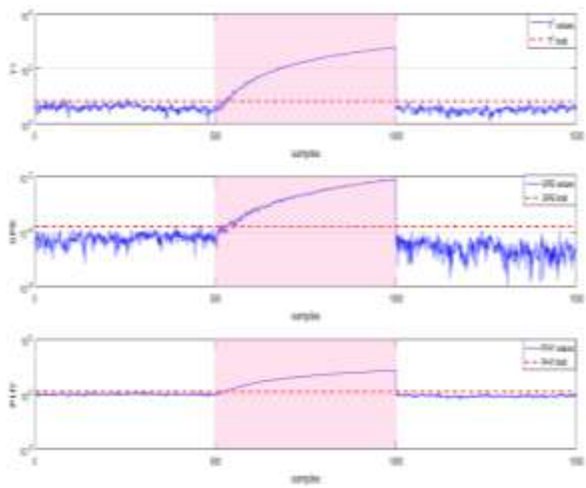
a) Random fault in single sensor (Sfault1)



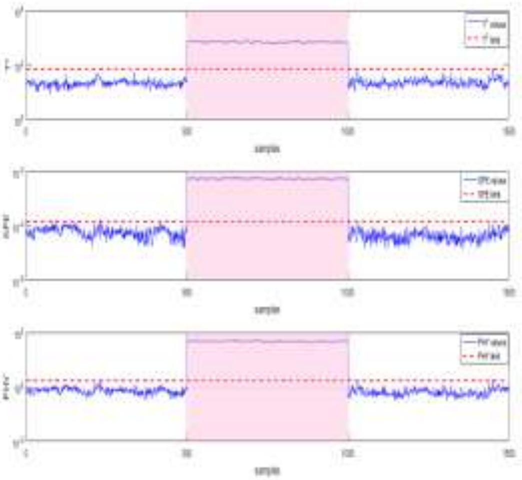
b) Abrupt fault in single sensor (Sfault2)



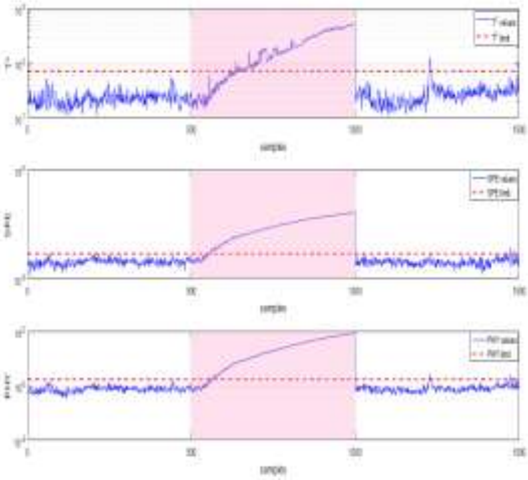
c) Drift fault in single sensor (Sfault3)



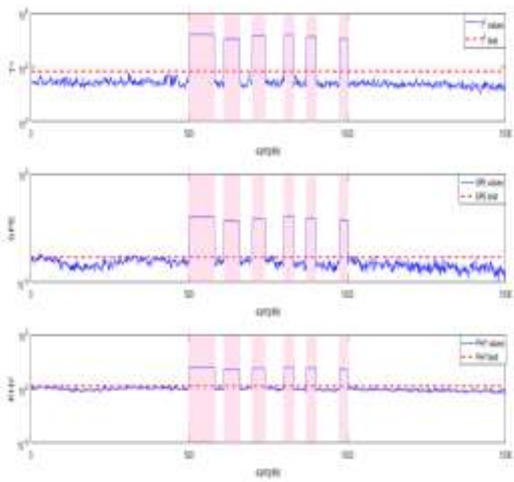
d) Drift fault in single sensor (Sfault4)



e) Abrupt fault in multiple sensors (Sfault5)



f) Drift fault in multiple sensors (Sfault6)



g) Intermittent fault in single sensor (Sfault7)

Figure 11. T^2 , Q , φ monitoring results of sensor faults using KPCA. **(a)** Sfault1; **(b)** Sfault2; **(c)** Sfault3; **(d)** Sfault4; **(e)** Sfault5; **(f)** Sfault6; **(g)** Sfault7.

III.5.3. Proposed RKPCA monitoring model

III.5.3.1. Graphical representation of the original dataset (768 observations) using logarithmic distance

$\boldsymbol{log}(\boldsymbol{C})$ vs $\boldsymbol{log}(\boldsymbol{r})$ is plotted after calculating the correlation integral $\boldsymbol{C}$ from **(2.52)** and the Euclidian distance $\boldsymbol{r}$ between the different pairs of variables.

The following graph is obtained:

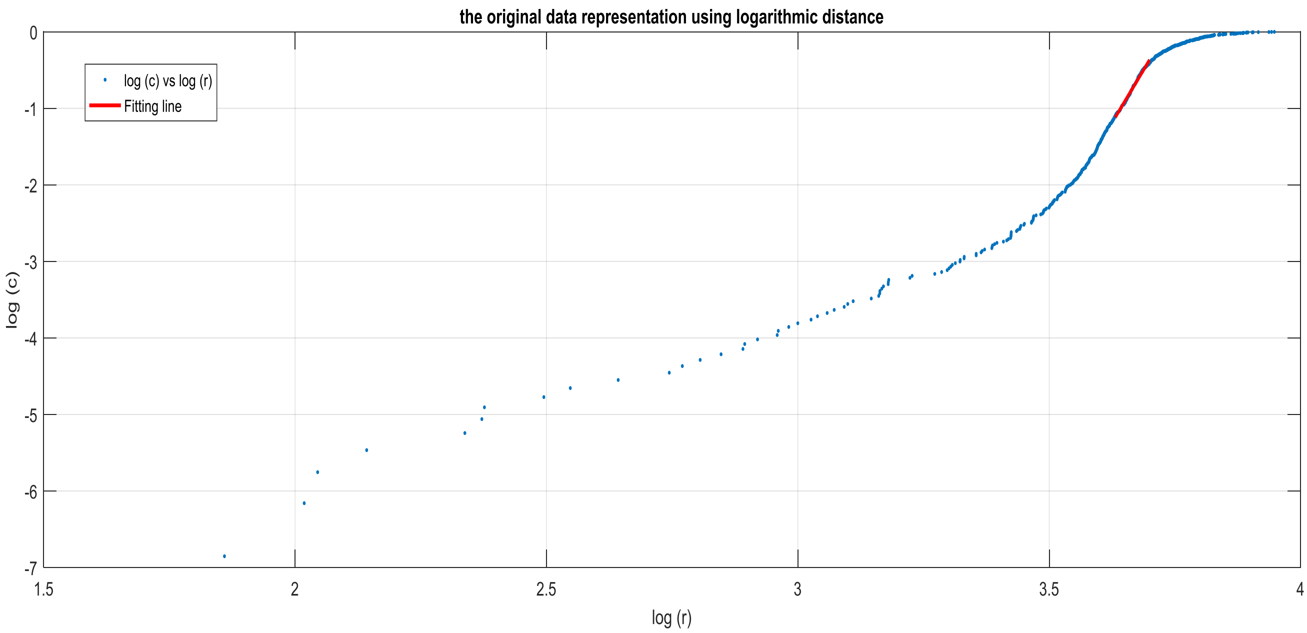


Figure 12. Graphically representation of the original data (786 observations)

III.5.3.2.Reduced data and characteristics

After plotting the data, we take the second third of the dataset (the second third where the distance $\boldsymbol{r}$ tends to 0 as it is mentioned in eq.**(2.53)**) which represents approximately a linear line and its slope is the fractal dimension of the dataset.

The obtained fractal dimension through integral correlation technique is 11. So, the original data can be reduced and represented only with 11 independent observations.

A regression analysis is used to analyses the fitting line. Regression Analysis is a method of statistically analyzing data. The purpose is to estimate the relationships between a dependent variable (often called the 'outcome variable') and one or more independent variables (often called 'predictors', 'covariates', or 'features') and to establish a mathematical model to observe specific variables how they predict. More specifically, regression analysis can help people understand the amount of change in the dependent variable when only one independent variable changes.

A reasonable form of a relationship between $\boldsymbol{Y}$ and $\boldsymbol{X}$ is the linear relationship:

$$\boldsymbol{Y = b0 + b1x}$$

Where, of course, *b0* is the **intercept** and *b1* is the **slope**.

The regression model has given the following results:

```
b =

   10.9984
  -41.0558

>> bint

bint =

   10.9003   11.0966
  -41.4157  -40.6959

>> stats

stats =

   1.0e+04 *

    0.0001    4.8649         0    0.0000
```

The regression prediction equation established is:

$$\boldsymbol{Y = -41.0550 + 10.9984 \times X}$$

Where:

b0= 10.9984 represents the slope of the line.

b1= -41.0557 represents the intersection with the Y axis.

bint values are the intervals in which b values range with a 95% confidence intervals, this implies:

$$10.9003 \leq b1 \leq 11.0965$$

$$-41.4156 \leq b0 \leq -40.6959$$

stats gives the statistics of the regression model, the first value represents the decision coefficient or the coefficient of determination R^2 .

R^2=0.9935,

R-Squared (R^2 or the coefficient of determination) is a statistical measure in a regression model that determines the proportion of variance in the dependent variable that can be explained by the independent variable. In other words, R^2 shows how well the data fit the regression model (the goodness of fit).

The coefficient of determination suggests that the model fit to the data explains 99.35% of the variability observed in the response, meaning that a significant linear regression relationship exists between the response Y and the predictor variables in X.

Based in the results of the regression analysis, the second third of dataset is approximately linear with a slope of 11, which represents the fractal dimension.

III.5.3.3. Homogeneity test

In order to check the similarity between the original and the new reduced data sets, we expose it to a homogeneity test. We compare the sets in terms of mean difference, variance defense and Kulbeck-liebler Divergence.

Kulbeck-liebler Divergence

Kullback-Leibler divergence is an information-based measure of disparity among probability distributions. Given distributions P and Q defined over X, with Q absolutely continuous with respect to P.

Kullback-Leibler divergence is given by [47] :

$$D(P \parallel Q) = \int P(x) \log \frac{P(x)}{Q(x)} \mathrm{dx} \qquad \textbf{(2. 59)}$$

The value of Kullbeck-Liebler Divergence ranges between 0 and ∞, where 0 means P and Q are identical and ∞ means that they are completely different.

In Figure 13, the mean and variance difference between the normalized original data and the selected reduced set is represented in bar graph, in addition to kullbeck-Liebler values corresponding to each of the 44 variables.

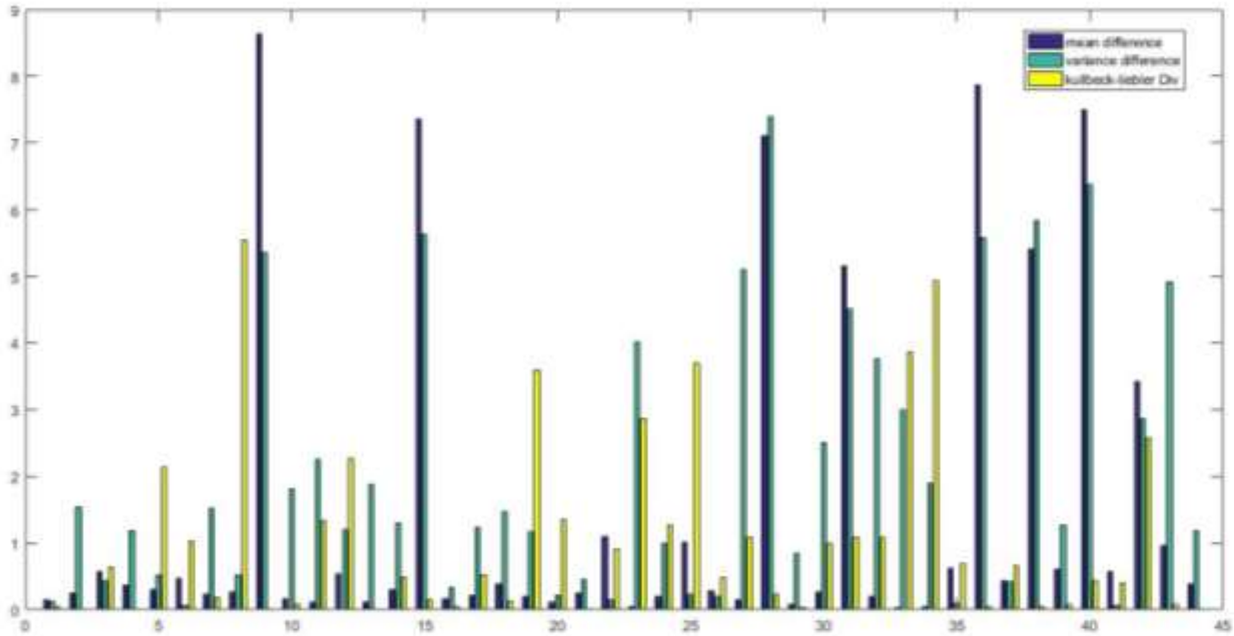


Figure 13. Bar chart represents the Kullbeck-Liebler divergence and the difference in mean and variance between the original and reduced data

We can see that the mean and variance differences are both high for variables 9, 15, 28,31,36,38 and 40; variables 32 and 43 have some variance difference whereas there mean difference is tiny. However, the corresponding KLD values are very low for the previous variables indicating high degree of similarity. Variables 8, 19, 25 and 34 show some probability distribution disparity, however the mean and variance difference is negligible. The remaining variables of the original and reduced data highly match in terms of mean, variance and probability distribution.

Taking into consideration the few observation selected to replace the original large set, we can say that the new reduced samples are satisfactory similar to the original.

III.5.3.4. Time and space complexity:

Time complexity is a function describing the amount of time an algorithm takes in terms of the amount of input to the algorithm. Whereas Space complexity describes the amount of memory (space) an algorithm takes in terms of the amount of input to the algorithm. These concepts are expressed in terms of 'Big O' notation. The time complexity is found to be $\mathcal{O}(l^3)$ as explained in table 12. The reduced KPCA uses a kernel model matrix of $l \times l$ matrix that needs to be stored, meaning that the space complexity is $\mathcal{O}(l^2)$

Table 12. RKPCA time complexity analysis.

Method		Cost
RKPCA	*Inti.: Training data* $X \in \mathbb{R}^{N\times m}$	
	Begin	
	$K_l = k(\boldsymbol{x}_i, \boldsymbol{x}_j)\big\|_{i,j=1\ldots l}$	l^2
	eigendecompose $K_l = \hat{P}_l \hat{\Lambda} \hat{P}_l^T$	l^3
	Compute T^2, Q *and* φ	l
	Compute T_α^2, Q_α *and* φ_α	l
	End	
Total	$l^3 + l^2 + 2l = \mathcal{O}(l^3)$	

Given that number of samples retained is $l = 11$, using the proposed RKPCA minimizes the time it takes to detect the faults with 99.99%, in addition to the reduction of storage from a 768×768 matrix to 11×11 kernel matrix with 99.97%.

III.5.3.5. NOC results

The monitoring results of the RKPCA model in normal operation (training and testing set) using the three indices $\boldsymbol{T}^2, \boldsymbol{Q}$ and $\boldsymbol{\varphi}$ (with confidence levels 99%) are obtained:

$$\boldsymbol{Q_\alpha = 1.525 \times 10^{-4}} \qquad \boldsymbol{T^2_\alpha = 51.13} \qquad \boldsymbol{\varphi_\alpha = \ 1.7238}$$

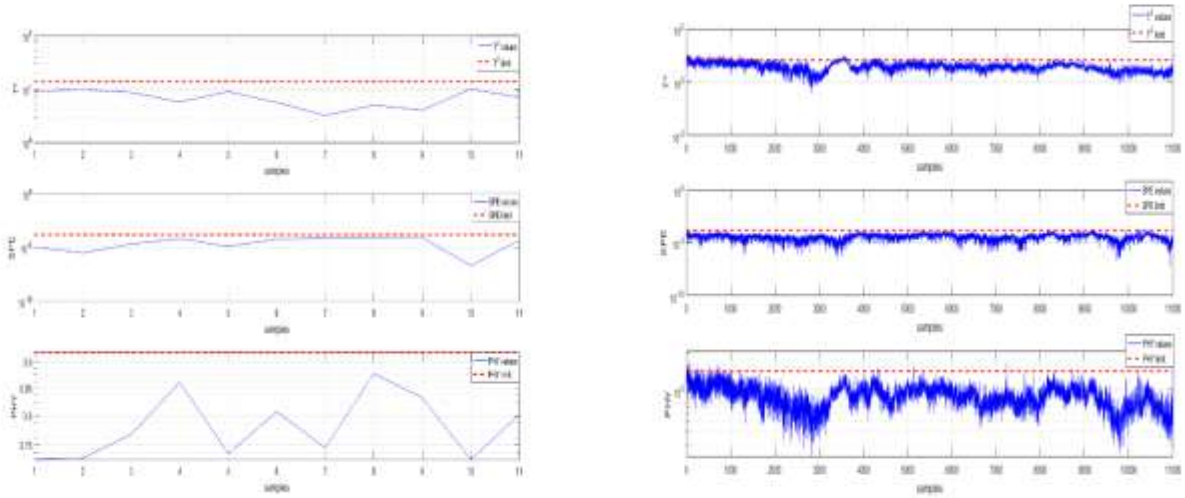
The model is assessed using the FAR values .

From table 13, The FAR values contributed by the RKPCA model in the training and testing sets are very good, where the FAR indicated by the T^2 is satisfactory which is around 4% .Whereas Q and φ shown the best performance with negligible FAR values. This clearly indicates the model's precision.

Table 13. FAR contributed by $\boldsymbol{T}^2, \boldsymbol{Q}$ and $\boldsymbol{\varphi}$ under NOC using RKPCA.

Method	NOC Data	Index	FAR (%)
RKPCA	Training data	T^2	4.817708
		Q	0.390625
		φ	0.260417
	Testing data	T^2	4.636364
		Q	0.127273
		φ	0.381818

Figure 14 shows the time evolution of the fault detection using the three indices based RKPCA for training and testing set. The horizontal line in red represents the threshold a) training data set , b) testing data set



a) Training data set b) Testing data set

Figure 14. T^2, Q and φ monitoring results of healthy process operation using RKPCA.

(a) testing data set , (b) training data set.

III.5.3.6. The involuntary real process fault results

The performance measures **MDR**, **FAR**, **DTD** and Cost function **J** recorded from using the proposed RKPCA-based model are summarized in table 14

From table14, $\boldsymbol{T^2}$ index has detected immediately the process fault (DTD = 0.00s). Whereas $\boldsymbol{Q}$ and $\boldsymbol{\varphi}$ indices shows a slight delay of 3 and 1 seconds. In terms of MDR, the proposed RKPCA has shown negligible amount of missed detected samples with all the indices. For the FAR, $\boldsymbol{Q}$ and $\boldsymbol{\varphi}$ indices provided acceptable results, however, we record considerable values when using $\boldsymbol{T^2}$ index.

Table 14. Missed detection rate (MDR), False Alarm Rate (FAR), detection time delay (DTD) and the cost function J values for the eight faults of cement rotary kiln using the proposed RKPCA using 99% as the confidence interval.

Faults	Performances	T^2	Q	φ
Process fault	MDR	0.31	0.43	0.06
	FAR	36.3	0.89	5.56
	DTD	0	3	1
	J	**36.61**	**4.32**	**6.63**
Random and single	MDR	3.19	24.35	4.39
	FAR	18.11	0	2.5
	DTD	0	0	0
	J	**21.31**	**24.35**	**6.89**
Abrupt and single	MDR	21.75	0	0
	FAR	1.9	0	0
	DTD	1	0	0
	J	**24.65**	**0**	**0**
Drift and single 1	MDR	19.76	5.78	7.98
	FAR	4.8	0.2	0.4
	DTD	78	27	28
	J	**102.56**	**32.98**	**36.38**
Drift and single 2	MDR	5.38	6.78	5.78
	FAR	0.8	0	0
	DTD	18	27	18
	J	**24.19**	**33.78**	**23.78**
Abrupt and multiple	MDR	0	0	0
	FAR	0.21	0.21	0.1
	DTD	18	0	0
	J	**18.21**	**0.21**	**0.1**
Drift and multiple	MDR	28.34	33.13	23.75
	FAR	0.2	0.1	0.2
	DTD	110	110	120
	J	**138.54**	**143.23**	**143.95**
Intermittent and single	MDR	0	3.83	0
	FAR	1.66	1.42	1.58
	DTD	0	0	0
	J	**1.66**	**5.26**	**1.58**
J average		45.97	30.5	27.41
J total			34.63	

To summarize, **J** has shown that the real process fault is detected effectively using $\boldsymbol{Q}$ and $\boldsymbol{\varphi}$. However, large values are provided by the $\boldsymbol{T}^2$ index.

Figure 15 shows the monitoring results of RKPCA model applied on the real process fault using the 3 indices. The horizontal line in red represents the index threshold with a confidence level of 99%.

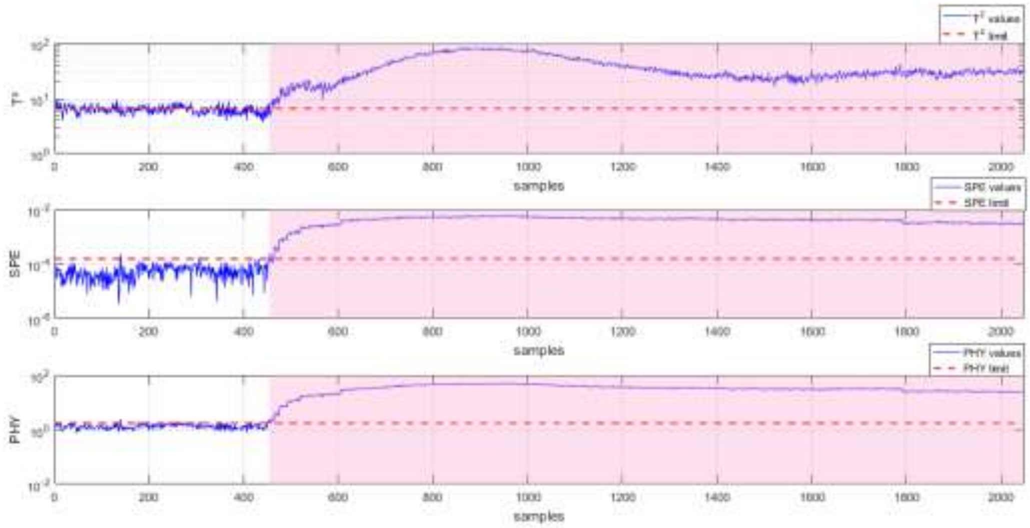
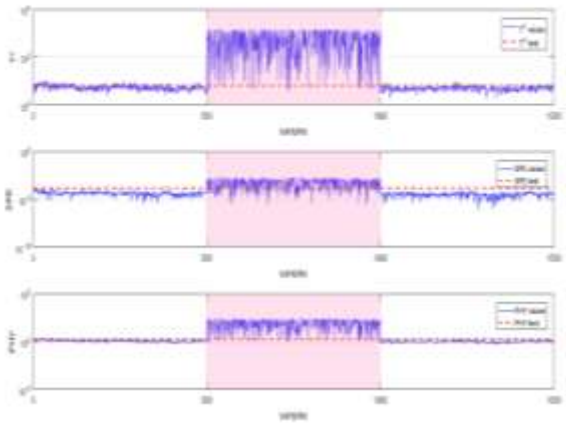


Figure 15. T^2, Q and φ monitoring results of real involuntary process fault in the cement rotary kiln using the proposed RKPCA

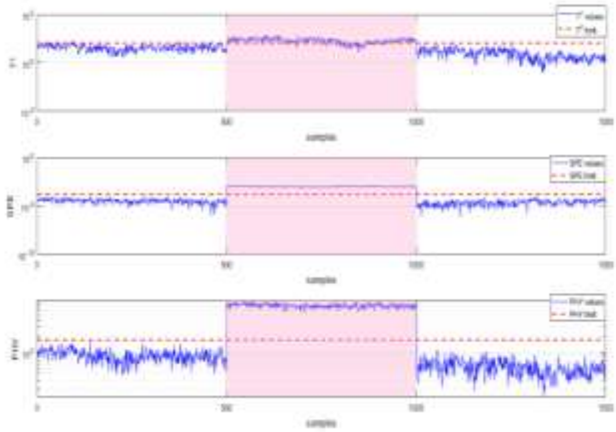
III.5.3.7. Simulated sensor faults results

From table 14, Abrupt fault in single sensor (Sfault2) or multiple sensors (Sfault5) have no missed detections nor time delay, except for the T^2 that missed a number of faults in the Sfault2 and had an 18 seconds delay in Sfault5, the FAR values are zero or negligible. The random fault (Sfault1) was detected with no delay using the three indices with very low FAR and MDR values besides the Q index that shows 24.35% MDR and T^2 that shows 18.11% FAR. For the Drift faults in single 1 & 2 (Sfault3, Sfault4) and in multiple variables and (Sfault6) we can notice a tiny false alarms rate, for the single faults (Sfault3, Sfault4), some detection delay occurs around 18 and 28 seconds which explains the MDR results that ranges from 5% to 7%. The Drift and multiple (Sfault6) marks the highest DTD of 1min 40 sec, that reflected on the MDR outcomes that raises to around 30%. The last type of faults tested is the Intermittent single sensor fault, the latest was detected with no time delay and negligible FAR and MDR values.

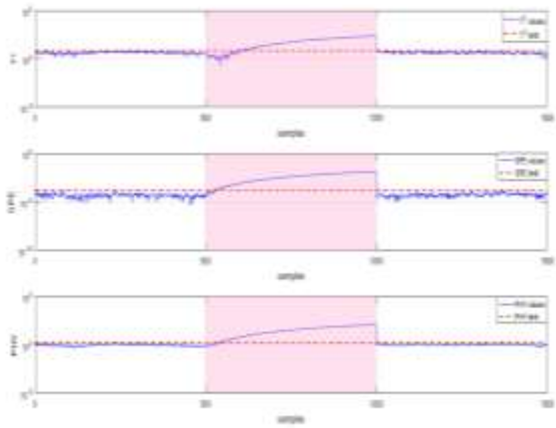
Figure 16 show the 3 indices monitoring results using KPCA. The detection of single as well as multiple sensor faults of abrupt, random, intermittent, and drift types can be noticed from the graphs.



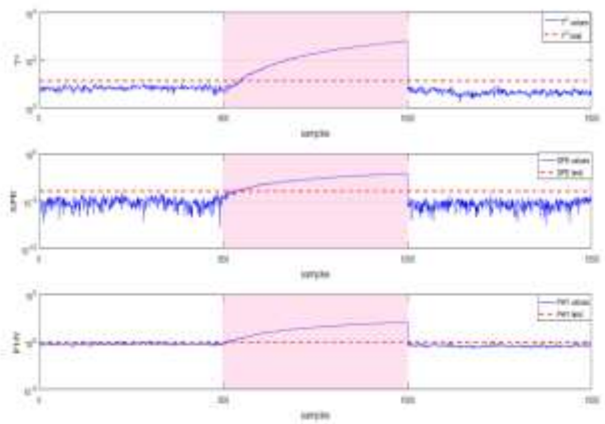
a) Random fault in single sensor (Sfault1)



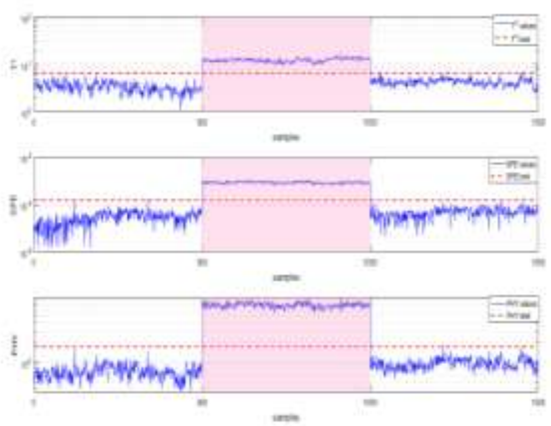
b) Abrupt fault in single sensor (Sfault2)



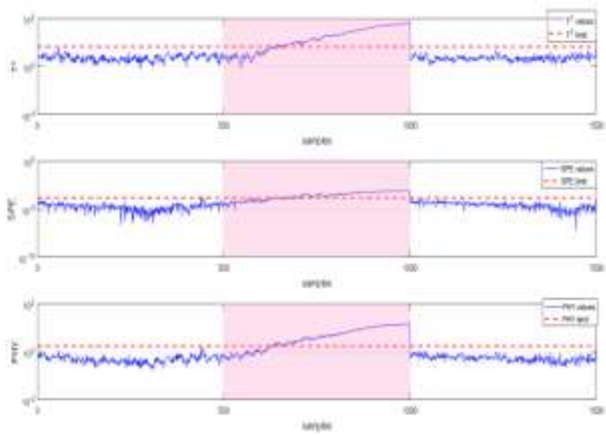
c) Drift fault in single sensor (Sfault3)



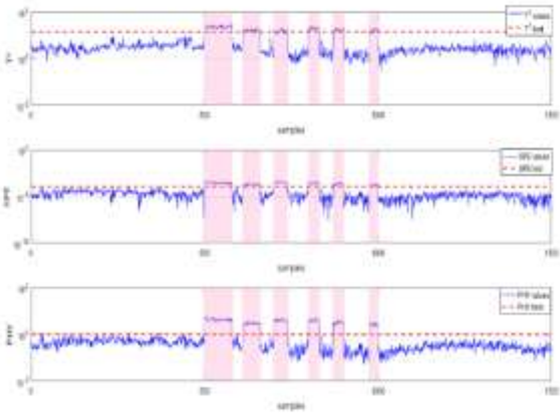
d) Drift fault in single sensor (Sfault4)



e) Abrupt fault in multiple sensors (Sfault5)



f) Drift fault in multiple sensors (Sfault6)



g) Intermittent fault in single sensor (Sfault7)

Figure 16. φ monitoring results of sensor faults using RKPCA. **(a)** Sfault1; **(b)** Sfault2; **(c)** Sfault3; **(d)** Sfault4; **(e)** Sfault5; **(f)** Sfault6; **(g)** Sfault7.

III.5.4. Comparison between PCA, KPCA, RKPCA based on Euclidian distance and the proposed approach RKPCA

The results achieved by the Euclidian distance RKPCA are taken from [49].

Based on the result of the correlation dimension, the proposed RKPCA approach has achieved the best minimization in the time and storage complexity compared to KPCA technique and the RKPCA based on the Euclidian distance, what we say is that the time gain is about 99.99% whereas the storage gain is 99.97%. This is due to the reduction in the number of samples from 768 to 11 samples, which means the number of observations has been reduced by approximately 153 times, and this is tremendous. This reduction in time is very important in nowadays processes.

Generally speaking, the KPCA shows better results when $\boldsymbol{T}^2$ and $\boldsymbol{\varphi}$ indices are used, with the minimum overall cost **J** with a value of **23.36** as shown in table 17. This implies that KPCA technique has dealt with the non-captured nonlinearities when PCA is used, and the loss of information resulted from the reduction in RKPCA.

The performances of the proposed method in terms of FAR using the three indices have provided very good results, as well as the PCA and KPCA techniques, the values where better that the Euclidian distance based technique except for the $\boldsymbol{T}^2$ index where the values rise compared to the other methods. In addition, RKPCA technique has provided the best FAR values when $\boldsymbol{Q}$ index is used compared the other methods.

The values of MDR and DTD recorded where a little higher that the ones of the PCA method, whereas they were very close to the results of KPCA and the RKPCA based on Euclidian distance except for the multiple drift sensor fault (Sfault6), where the results are bigger due to the complexity of the fault which is considered as the most difficult type of faults due to the small development in the fault magnitude over time.

Based on table 15, KPCA has provided the best average **J** minimization. Yet, the proposed approach can perfectly detect all the faults with an average J of **33.38** which is better than the other RKPCA method. We can see that cost averages are **30.5** and **27.41** when $\boldsymbol{Q}$ and $\boldsymbol{\varphi}$ index are used respectively. Whereas, it has some difficulties when the detection is performed using $\boldsymbol{T}^2$ index.

The bold values highlight the best performance of the proposed method.

Table 15. FAR (%) of faults monitoring results

Method		Training	Testing	RPfault	SFault1	SFault2	SFault3	SFault4	SFault5	SFault6	SFault7
PCA	**T^2**	1.04	0.08	1.67	0.30	0.90	0.70	0.10	0.20	1.00	1.04
	Q	1.04	0.65	14.05	0.80	1.70	0.40	0.10	1.50	1.30	1.85
	φ	1.04	0.57	14.52	0.60	1.30	1.10	0.10	0.60	1.50	1.37
KPCA	**T^2**	2.99	2.80	7.57	0.60	1.20	1.90	0.30	0.63	1.80	4.13
	Q	0.65	0.51	10.02	0.40	1.00	0.30	0.00	0.94	0.80	1.90
	φ	0.91	0.33	2.67	0.00	0.90	0.50	0.00	0.21	0.90	1.66
RKPCA Eucludian distance based	**T^2**	0.69	0.00	0.00	0.20	0.20	0.10	0.10	0.20	0.10	0.96
	Q	0.69	3.75	23.57	12.10	39.80	3.60	7.80	27.00	19.80	18.63
	φ	0.69	0.88	10.24	1.80	14.10	0.70	0.70	4.10	2.70	3.69
The proposed RKPCA	**T^2**	4.81	4.63	36.3	18.11	1.90	4.80	0.8	0.21	0.20	1.66
	Q	**0.39**	**0.12**	0.89	**0.00**	**0.00**	**0.20**	**0.00**	**0.21**	**0.10**	**1.42**
	φ	**0.26**	0.38	5.56	2.50	**0.00**	**0.40**	**0.00**	**0.10**	**0.20**	1.58

Table 16. MDR (%) and DTD of faults monitoring results

Faults		RP Fault	SFault1	SFault2	SFault3	SFault4	SFault5	SFault6	SFault7
PCA	**T^2**	1.98 | 5	2.40 | 1	0.00 | 1	4.70 | 20	15.60 | 1	0.00 | 1	22.50 | 100	0.00 | 1
	Q	1.68 | 4	1.20 | 1	0.00 | 1	12.90 | 7	2.20 | 3	0.00 | 1	15.60 | 56	0.00 | 1
	φ	1.10 | 4	1.80 | 1	0.00 | 29	7.20 | 19	6.40 | 1	0.00 | 56	19.40 | 1	0.00 |1
KPCA	**T^2**	0.06 | 1	5.58 |0	0.00 | 0	4.99 | 19	15.16 | 1	0.00 | 1	18.16 | 110	0.00 | 0
	Q	0.75 | 4	1.19 | 0	5.78 | 0	10.57 | 6	2.39 | 6	0.00 | 0	15.96 | 110	0.00 | 0
	φ	0.25 | 1	1.79 | 0	0.00 | 0	6.98 | 28	4.79 | 11	0.00 | 0	18.36 | 110	0.00 | 0
RKPCA Eucludian distance based	**T^2**	3.61 | 56	43.00 | 0	1.20 | 0	9.20 | 42	21.0 | 103	0.00 | 0	42.4 | 203	0.00 | 0
	Q	0.18 | 17	1.00 | 1	0.00 | 0	3.20 | 5	1.00 | 1	0.00 | 0	8.40 | 21	0.00 | 0
	φ	0.72 | 1	1.80 | 1	0.00 | 0	4.40 | 9	1.60 | 2	0.00 | 0	11.00 | 55	0.00 | 0
Proposed RKPCA	**T^2**	0.31 **| 0**	3.19 | **0**	21.75 | 1	19.76 | 78	**5.38** | 18	**0.00 | 0**	28.34 | 110	**0.00 | 0**
	Q	0.43 **| 3**	24.35 | **0**	**0.00 | 0**	5.78 | 27	6.78 | 27	**0.00 | 0**	33.13 | 110	3.83 | 0
	φ	**0.06 | 1**	4.39 | **0**	**0.00 | 0**	7.98 | 28	5.78 | 18	**0.00 | 0**	23.75 | 120	**0.00 |** 0

Table 17. A comparative table between PCA, KPCA and the proposed RKPCA using the average J value contributed by T^2, Q and φ.

NOC Data	Index	Cost(J)	Average Cost J
PCA	**T^2**	40.03	25.13
	Q	17.51	
	φ	17.87	
KPCA	**T^2**	24.26	23.36
	Q	22.25	
	φ	23.62	
RKPCA (euclidien distance)	**T^2**	65.78	35.89
	Q	26.27	
	φ	15.64	
RKPCA (Fractal dimension)	**T^2**	45.97	33.38
	Q	30.50	
	φ	27.41	

III.6. conclusion

PCA, KPCA, and the proposed approach RKPCA were applied to a collection of healthy and faulty data collected from a cement factory in this section of the report. In terms of FAR, MDR, DTD, the cost function **J**, and execution time, the three approaches were analyzed. The proposed methodology's results have demonstrated its potential to reduce the computing time problem presented by KPCA. In addition to its efficiency to detect many types of faults (Abrupt, Drift, Intermittent).

GENERAL CONCLUSION

Fault Detection and Diagnosis approaches for effective process monitoring have gained considerable attention from both industrial as well as academic fields, and thus many useful process monitoring systems including FDD techniques have been exploited and implemented for several industrial processes. However, there are still lots of difficulties in the implementation of the FDD methods for real industrial processes due to the unique characteristics (e.g., multivariate, correlation, non-linearity...). This thesis presents present a classification of fault diagnosis methods and give very brief overview of each class. Fault diagnosis methods were classified into three main categories: qualitative model-based, quantitative model-based and history-based methods. Measurement data are usually the most easily available knowledge about a process, hence, history-based fault diagnosis methods are widely used where multivariate statistical process control (MSPC) are among the most efficient techniques that have seen growth in the last decade. Fault diagnosis systems based on such knowledge are generally easy to develop and maintain, well suited for highly non-linear systems and do not require understanding of the physics of the system being modelled. Through multivariate statistical data analysis, features associated with different faults can be discovered and used in fault diagnosis.

The well-known PCA technique is presented and used to detect the different types of faults with assumption of the linearity of the data. KPCA technique was then introduced to cope with the linear assumption of the PCA. It has provided good results in detecting the several faults; however, it has shown a big problem in terms of time consumption and space. The new RKPCA approach was proposed to deal with The high computational time and space requirement resulted from using KPCA method demands some further work on reducing the time and space. In this work we proposed the new reduction technique based on the fractal dimension. The idea behind the novel approach was to retain the samples that defines the data set and represents the essential information in the data matrix.

In this work, PCA, KPCA and the suggested approach RKPCA were used as multivariate statistical methods to monitor the cement rotary kiln system. The monitoring was based on the common used indices: Hotelling's $\boldsymbol{T}^2$ and $\boldsymbol{Q}$, in addition to a new proposed index called the combined index $\boldsymbol{\varphi}$. Then evaluated in terms of False Alarm Rate, Missed Alarm rate, Detection Time Delay, Time and Space complexity.

We can say that the results of the proposed RKPCA were not good as the ones of PCA and KPCA in terms of MDR and DTD; in the contrary the FAR values were pleasing. In the other hand, if we compare it to the RKPCA based on Euclidian distance, we can see that the results were better in terms of performance measures, and also in terms of size reduction were the outcome was incomparable.

This technique achieved the objectives it was set for compared to other techniques by far. The number of samples retained is only 1.43% of the original number of samples. Also, it saves 99.99% of the time and 99.97% of space compared the KPCA.

The reduced KPCA has shown promising results in Fault Detection and Diagnosis. With the fast development of industrial process and growth of its complexity and generated measurements data, the reduction using fractal dimension may be the alternative of other Fault Detection and Diagnosis methods. Further improvements to the performance of the proposed approach by improving the choice of the kernel function and its parameters, in addition to the specification of the number of principle components retained based on the process when applying the PCA in the feature space to obtain better performance.

REFERENCES

[1] Seongmin Heo and Jay H. Lee *"Fault detection and classification using artificial neural networks "* Department of Chemical and Biomolecular Engineering Korea Advanced Institute of Science and Technology 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea

[2] O.Taouali, I.Jaffel, H.Lahdhiri, M.Faouzi Harkat, H.Messaoud, "*New fault detection method based on reduced kernel principal component analysis (RKPCA)",* Int J Adv Manuf Technol ,2015.

[3] D.M.Himmelblau, *"Fault Detection and Diagnosis in Chemical and Petrochemical Processes",* Elvisar Press, Amsterdam, Himmelblau, 1978 .

[4] R.Iserman, *"Model-based Fault Detection and Diagnosis - Status and Applications",* Annual Reviews in Control, Vol.29, PP. 71–85, 2005.

[5] V.Venkatasubramanian, R. Rengaswamy, K. Yin Surya N. Kavuri, "*A review of process fault detection and diagnosis Part I: Quantitative model-based methods",* Computers and Chemical Engineering, vol. 27, pp.293-311, 2003.

[6] S. Katipamula, M.R. Brambley, "*Methods for fault detection, diagnostics and prognostics for building systems—a review, Part I"*, HVAC&R Research, Vol.12 (2), PP. 169–187, 2005.

[7] H.Cheng, M.Nikus, Sirkka-Liisa Jämsä-Jounela, *"Application of a causal digraph based fault diagnosis method with discrete state space model on a paper of machine simulator",* Helsinki University of Technology Laboratory of Process Control and Automation Kemistintie 1, FI-02150 HUT, Finland.

[8].E.E.HURDLE, L.M.BARTLETT and J.D.ANDREWS, *"System fault diagnostics using fault tree analysis. IN: Proceedings of the 16th Advances in Reliability Technology Symposium (ARTS)"*, Loughborough, UK, PP. 203-222, 2005.

[9] Q. YANG, *"Model-Based and Data Driven Fault Diagnosis Methods With Applications to Process Monitoring",* PhD Thesis, Electrical Engineering and Computer Sciences, Case Western Reserve University, Ohio, USA, 2004.

[10] Y.IWASAKI and H.A.SIMON, *"Causality in device behavior. Artificial Intelligence",* 29(1):3 – 32, 1986.

[11] A.Alkaya, "*Novel Data Driven-Based Fault Detection for Electromechanical and Process Control Systems*", Ph.D. thesis, Çukurova University, Institute of natural and applied sciences Department of electrical and electronics engineering, 2012

[12] V. Venkatasubramanian, R.Rengaswamy, S Kavuri, K.Yin, *"A review of process fault detection and diagnosis: Part III: Process history based methods".* Computers & Chemical Engineering, Vol.46, PP. 27-327, 2003

[13] J.V.KRESTA, J.F.MACGREGOR and T.E.MARLIN, *"Multivariate Statistical Monitoring of Process Operating Performance".* The Canadian, 1991.

[14] J.F.MACGREGOR, and T.KOURTI, *"Statistical Process Control of Multivariate Processes. Control Engineering Practice",* 3(3):403 – 414, 1995.

[15] S.Wold, K.Esbensen, & P.Geladi, *"Principal component analysis*», Chemometrics and Intelligent Laboratory Systems, *2*(1–3), PP. 37–52, (1987).

[16] I.Jolliffe, "*Principal component analysis* (2nd ed)", New York, NY: Springer-Verlag, (2002).

[17] B.M.Wise and N.B Gallagher, "*The process Chemometrics approach to process monitoring and fault detection"*, Journal of Process Control, VOL.*6*, PP. 329–348, (1996).

[18] H.OCAK, *"Fault Detection, Diagnosis and Prognosis of Rolling Element Bearings: Frequency Domain Methods and Hidden Markov Modeling".* PhD thesis, Case Western Reserve University, 2003.

[19] R.RENGASWAMY and V.VENKATASUBRAMANIAN, *"A fast training neural network and its updation for incipient fault detection and diagnosis".* Computers and Chem. Eng., 24(2-7):431–437, 2000.

[20] M. Mansouri, Marie-France Destain, H. Nounou, and M. Nounou. *'Enhanced Monitoring of Environmental Processes*' International Journal of Environmental Science and Development, Vol. 7, No. 7, July 2016.

[21] Herve Abdi and Lynne J. Williams *'Principal component analysis'*. Wiley Interdisciplinary Reviews Computational Statistics, 2(4), 433–459. Doi :10.1002wics.101

[22] G. T. Reddy *et al*., '*Analysis of Dimensionality Reduction Techniques on Big Data*,' in *IEEE Access*, vol. 8, pp. 54776-54788, 2020, doi: 10.1109/ACCESS.2020.2980942.

[23] Tharwat, A. *'Principal component analysis - a tutorial'*. International Journal of Applied Pattern Recognition, vol(3)n°(3), 197. Doi :10.1504ijapr.2016.079733

[24] F. Bencheikh, M.F. Harkat, A. Kouadri, A. Bensmail 'New reduced kernel PCA for fault detection and diagnosis in cement rotary kiln*' doi:* 10.1016/j.chemolab.2020.104091

[25] Carlos F. Alcala and S. Joe Qin *'Reconstruction-Based Contribution for Process Monitoring with Kernel Principal Component Analysis'* Ind. Eng. Chem. Res. **2010,** 49, 7849–7857.

[26] Qin, S. J. *'Statistical process monitoring: basics and beyond'*. Journal of Chemometrics, 17, 8(2003), . 480-502.

[27] H. Yue and S. Joe Qin, *"Reconstruction based fault identification using a combined index",* Ind. Eng. Chem. Res, VOL.40, PP.4403–4414, 2001.

[28] Slišković, Dražen; Grbić, Ratko; Hocenski, Željko *'Multivariate statistical process monitoring'* Tehnicki Vjesnik-Technical Gazette, 19 (2012), 1; 33-41 (međunarodna recenzija, članak, znanstveni).

[29].S.Mika, B.Schölkopf, A.Smola, K.R Müller, M.Scholz, G.R ̈tsch, *"Kernel PCA and de-noising in feature spaces",* Advances in Neural Information Processing Systems, VOL.11, PP.536-542, 1999.

[30].S.Romdhani, S.Gong, A. Psarrou, *"A multi-view nonlinear active shape model using kernel PCA",* Proceedings of BMVC, Nottingham,UK, PP. 483 492, 1999.

[31] P. Nomikos and J. F. MacGregor, *'Multivariate SPC charts for monitoring batch process'* Technometrics, vol. 37, no. 1, pp. 41–59, 1995.

[32] L. M. Elshenawy, *'Fault Detection and Isolation Indices for Large-Scale Systems'* 2019 14th International Conference on Computer Engineering and Systems (ICCES), 2019, pp. 269-274, Doi: 10.1109/ICCES48960.2019.9068161.

[33] D.M.Himes, R.H.Storer, C.Georgakis, "*Determination of the number of principal components for disturbance detection and isolation*." In: Proceedings of American control conference, IEEE Press, Piscataway, New Jersey, PP. 1279–1283, 1994

[34] Lahdhiri, H., Elaissi, I., Taouali, O. et al. *'Nonlinear process monitoring based on new reduced Rank-KPCA method'*. Stoch Environ Res Risk Assess 32, 1833–1848 (2018). Doi :10.1007/s00477-017-1467-z

[35] Jong-Min Lee, ChangKyoo Yoo, Sang Wook Choi, Peter A. Vanrolleghem, In-Beum Lee, *'Nonlinear process monitoring using kernel principal component analysis'* Chemical Engineering Science, Volume 59, Issue 1,2004,Pages 223-234,ISSN 0009-2509, Doi :10.1016/j.ces.2003.09.012.

[36] Heiko Hoffmann, "*Kernel PCA for novelty detection*" Volume 40, Issue 3, 2007,Pages 863-874, doi: 10.1016/j.patcog.2006.07.009.

[37] K.Kazor, R.W.Holloway, T.Y.Coth, A.S.Hering, *"Comparison of linear and nonlinear dimension reduction techniques for automated process monitoring of a decentralized Wastewater treatment facility"*, Stoch Environ Res Risk Assess, VOL.33, PP.1544-1544, 2016.

[38] Harkat, MF., Kouadri, A., Fezai, R. et al. *'Machine Learning-Based Reduced Kernel PCA Model for Nonlinear Chemical Process Monitoring'*. J Control Autom Electr Syst 31, 1196–1209 (2020). Doi :10.1007/s40313-020-00604-w

[39] R.T.Samuel, Y.Cao, *"Nonlinear process fault detection and identification using kernel PCA and kernel density estimation",* Systems science and control engineering, VOL.4, PP.165-174, 2016.

[40] D.Lopez-Paz, S.Sra,A. Smola,Z.Ghahramani, B.Schölkopf, *"Randomized nonlinear component analysis"*, In Proceedings of the 31st International Conference on Machine Learning, 2014.

[41] Jaffel, I., Taouali, O., Harkat, M.F. et al. *'Kernel principal component analysis with reduced complexity for nonlinear dynamic process monitoring'*. Int J Adv Manuf Technol 88, 3265–3279 (2017). Doi :10.1007/s00170-016-8987-4

[42] M.N Bagde, A.K Raina, A.K Chakraborty, J.L Jethwa, *'Rock mass characterization by fractal dimension'*. Engineering Geology, Volume 63, Issues 1–2,2002, Pages 141-155,ISSN 0013-7952, Doi:10.1016/S0013-7952(01)00078-3.

[43] James Theiler, *'Estimating fractal dimension'*. Journal of the Optical Society of America A: Optics, Image Science, and Vision, Volume 7, Issue 6, June 1990, pp.1055-1073 Doi:10.1364/JOSAA.7.001055 .

[44] Mandelbrot, Benoit B. '*The Fractal Geometry of Nature'*. 1983.

[45] Jafari Moghadam Fard, P., Moradi, M.H. & Saber, S. *'Chaos to randomness: distinguishing between healthy and non-healthy lung sound behaviour'*. Australas Phys Eng Sci Med 38, 47–54 (2015). Doi :10.1007/s13246-014-0316-3.

[46] Grassberger P., Procaccia I. (2004) *'Measuring the Strangeness of Strange Attractors'*. In: Hunt B.R., Li TY., Kennedy J.A., Nusse H.E. (eds) *'The Theory of Chaotic Attractors'*. Springer, New York, NY. Doi :10.1007/978-0-387-21830-4_12.

[47] Joyce J.M. (2011) *'Kullback-Leibler Divergence'*. In: Lovric M. (eds) International Encyclopedia of Statistical Science. Springer, Berlin, Heidelberg. Doi :10.1007/978-3-642-04898-2_327.

[48] B. Schölkopf, A. Smola and K. Müller, *'Nonlinear Component Analysis as a Kernel Eigenvalue Problem,'* in Neural Computation, vol. 10, no. 5, pp. 1299-1319, 1 July 1998, Doi: 10.1162/089976698300017467.

[49] S.Bennai, supervisor: Dr. Kouadri, '*Reduced Kernel PCA based Approach for Fault Detection in Complex Systems'*, Master project thesis, IGEE of Boumerdes university.