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:

Design of a Multivariate Alarm Delay-Timers System for a Cement Rotary Kiln

Presented by:

RAYANE Belkacem
RABIA Amina

Supervisor:

Dr. KOUADRI Abdelmalek

Mrs. KACED Radhia (Co-Supervisor)

Registration Number:……..…../2019

We would like to dedicate this thesis to our loving parents …

Acknowledgements

Above all and for eternity, we are grateful to ***Almighty ALLAH*** for His mercy and compasion.

We would like to acknowledge our academic supervisor ***Dr. KOUADRI Abdelmalek*** for his supervision throughout the evaluation of this project, his patience, motivation, support and guidance with a sense of immense gratitude.

His valuable support was very helpful especially with the special circumstances that happened during our research period .

We would also like to thank ***Mrs BOURBIA Radia*** for sharing her valuable knowledge with us, and providing us with tips and assistance.

Also, we are sincerely grateful to all the staff of the Institute of Electrical and Electronics Engineering of the University of M'Hamed Bougara-Boumerdes, Teachers and workers.

Abstract

Alarm systems play an important role in industrial processes in the prevention of major accidents or disasters that lead to casualties and economical losses. In practice, however, nuisance alarms interfere with operators' judgment. A good thresholding system should be able to distinguish between normal and abnormal situations of sensors measurements. Accomplishing this became a very challenging task since various techniques already exist. In this thesis, a novel technique is used to enhance the performance of alarm systems. This technique is mainly based on modeling the process using the principal component analysis (PCA). Once the model is validated; Instead of the conventional control limits, a combined index is used with delay-timers to enhance the alarm system's performance. Simulated and industrial examples are discussed to prove the effectiveness of the proposed method in designing industrial alarm systems. This method proved its efficiency to fulfill high alarm performance.

Keywords—Industrial alarm systems, PCA, Multivariate system, Markov process, Delay-timers, False alarm rate, Missed alarm rate, Expected detection delay.

Table of contents

List of tables

List of figures

Nomenclature

Roman Symbols

φ	The combined index
Q	Q statistics
Q_α	Upper control limits for Q statistics
T^2	Hotelling's T^2 statistics
T_α	Upper control limits for the Hotelling's T^2
xp(t)	The process signal

Acronyms / Abbreviations

ALAP	Alarm Prioritization System
BEDAM	Bayesian Estimation Based on Dynamic Alarm Management
BS	Broken Stick method
CA	Correspondence Analysis
CPV	Cumulative Percent of Variance
CV	Cross Validation
DCS	Distributed Control System
EDD	Expected Detection Delay
EEMUA	Engineering Equipment and Materials Users Association
FAR	False Alarm Rate
FDI	Fault Detection and Isolation
GFC	Criterion of Goodness-of-Fit

LV	Latent Variable
MAP	Minimum Average Partial
MAR	Missed Alarm Rate
MSE	Mean-Square Error
MSPM	Multivariate Statistical Process Monitoring
MVSPC	Multivariate Statistical Process Control
PA	Parallel Analysis
PCA	Principal Component Analysis
PCs	Principal Component Subspace
PLS	Partial Least Square
REDD	Required Expected Detection Delay
RFAR	Required False Alarm Rate
RMAR	Required Missed Alarm Rate
ROC	Receiver Operating Characteristics Curve
RS	Residual Subspace
SPC	Statistical Processing Control
SPE	Squared Prediction Error
SVD	Singular Value Decomposition

Chapter 1

Introduction

1.1 Motivation

Nowadays, alarm management became a crucial task in industrial processes. One of the most important responsibilities of alarming systems is to enunciate and notify operators about abnormal situations, hence, it is inseparably attached to process safety.

As a definition, Alarm Management consists of methods, tools, standards (such as ISA-18.2 and EEMUA-191 [34, 15]), and activities that improve system performance by improving the effectiveness of alarm systems. Alarm systems are closely related to fault detection and Identification. First, a fault is detected then the operator is informed about the occurrence of fault by raising an alarm. According to the Abnormal Situation Management(ASM) Consortium statistics, the US petrochemical industry losses are about 20 billions $ due to abnormal situations [6], with a major accident every three years in average. A large number of safety incidents have showed the importance of alarm systems in the safety of plants, such as BP Texas refinery incident in 2005 due to failed management of instruments and alarms[24]. Hence, the design of alarm systems is becoming one of the glowing research fields in process control and automation [37].

The major problem in alarm design is the lack of specific techniques to meet standard requirements such as ISA-18.2(International Society of Automation, 2009) and EEMUA-191(Engineering Equipment and Materials Users' Association 2013)[15, 61].

Industrial processes suffer from faults due to their components malfunctioning or failures, and depending on the degree of the failures, these faults can even lead to serious incidents. Process alarm systems are used to inform the operator about possible process deviations from the normal state so that the operator can take appropriate corrective actions. Alarm activation and the corresponding response from the operator together serve as one of the critical layers of protection during the occurrence of a process fault.

The most common way in detecting an alarm state is to compare the value of a process variable $x(t)$ to constant trip point [41].

$$x(t) = \begin{cases} 1 & \text{if } xp(t) \geq x_{tp} (\text{or } xp(t) \leq x_{tp}) \\ 0 & \text{if } xp(t) \leq x_{tp} (\text{or } xp(t) \geq x_{tp}) \end{cases} \tag{1.1}$$

The changes of alarm variables from 0 to 1 and from 1 to 0 are respectively referred to as the alarm occurrence and alarm clearance as given by eq. (1.1).

Historically, alarms were expensive to add and difficult to implement. Every alarm from the sensor to the control room had to be hardwired in the past. Space in the control room was also very limited due to the installation of numerous input/output devices. These technical and physical limitations led to a limited number of alarms that were carefully designed and examined. The alarms were very reliable and could be fully trusted by operators [38].

Alarms are prioritized depending on the impact and severity of the abnormality on the operation. In the ideal case, for each abnormal situation one and only one alarm should be triggered. This is hardly the case in practice. With the rapid advances in control technology over the last few decades, the number of sensors deployed for a given plant has increased dramatically. In addition, DCS systems have made it even easier to configure alarms than previous hard wired systems. As a result, alarms are designed for almost every purposes of a plant without proper analysis and rationalization, leading to alarm overflow.

The alarm flood is a very crucial problem in process industries which is defined as a situation in where several hundreds of alarms appear on the screen within minutes of the upset condition and the appearance of these alarms has an adverse cascading effect on the complete Plant operation. We consider a situation to be an alarm flood when more than 10 alarms per 10 minutes arrive at the operator's panel [28]. According to a study conducted by Matrikon and published by (Rothenberg, 2009) as shown in table 1.1, the number of alarms appearing in various industries was significantly high as compared to the benchmarks given in the Engineering Equipment and Materials Users Association (EEMUA) benchmark. These studies have also highlighted and emphasized the need for proper alarm management [58].

Table 1.1 Cross industry alarm activation study (Adopted from (Rothenberg, 2009))

	EEMUA	Oil and Gas	Petrochem	Power	Others
Average alarm per day	144	1200	1500	2000	900
Average standing alarm	9	50	100	65	35
Peak alarm per10 mn	10	220	180	350	180
Average alarm / 10mn interval	1	6	9	8	5
Distribution%(high- low -Med)	80/15/5	20/40/35	20/40/35	20/40/35	20/40/35

Most of the alarms during alarm flooding are nuisance. Chattering alarms, fleeting alarms and stale alarms are all common forms of nuisance alarms [43]. Different types of alarms are defined as shown in table 1.2.

Table 1.2 Definition of different type of alarms.

Alarm type	Definition
Nuisance alarm	An alarm that annunciates excessively, unnecessarily, or does not return to normal after the correct response is taken.
Chattering alarm	An alarm that repeatedly transitions between the alarm state and the normal state in a short period of time.
Fleeting alarm	Fleeting alarms are similar short-duration alarms that do not immediately repeat.
Stale alarm	An alarm that remains in the alarm state for an extended period of time.
False alarm	A false alarm is an alarm that is raised when the variable is behaving normally.
Missed alarm	a missed alarm occurs when the process is behaving abnormally but no alarm is raised.

To improve the overall alarm system, we either redesign the whole system depending on the available process performances or by applying various alarm configurations (on the existing alarm system). Alarm systems are closely related to fault detection and identification/isolation, first a fault is detected then operator is informed about the occurrence of fault by raising an alarm.

Based on the method used to configure thresholds, the design of an alarm system can be classified into two categories: univariate and multivariate. In the univariate design, the alarm threshold and processing techniques are configured for an individual process variable. Whereas for multivariate design; they are designed for a latent variable which is usually a linear combination of several process variables [51]. However, incorrect setting of the alarm threshold may result in false, missed and redundant alarms and may delay alarm's activation.

To assess the performance of alarm systems, in a processing industry three indices are evaluated: the false alarm rate (FAR) , missed alarm rate (MAR) they both quantify the

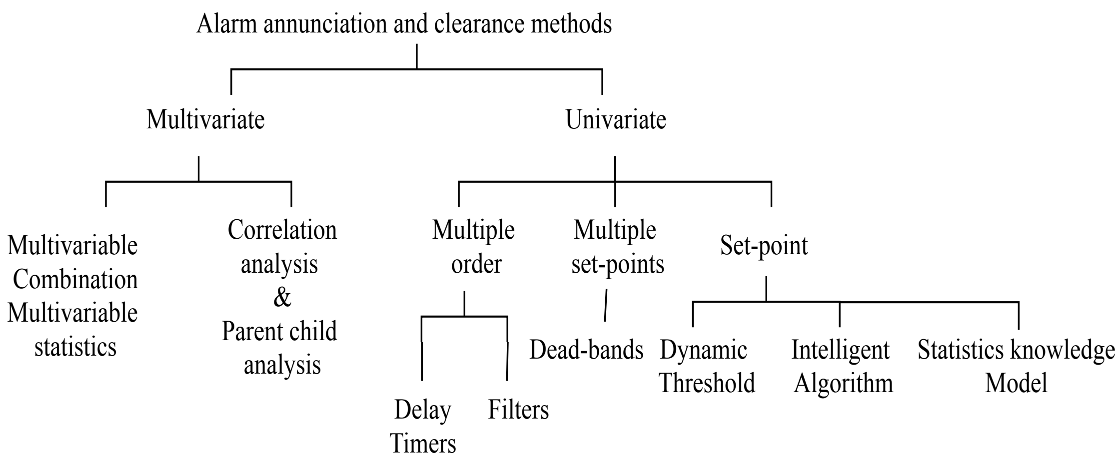


Fig. 1.1 Annunciation and clearance methods of alarm system design.

accuracy of the alarm; and the expected detection delay (EDD), which is the average time required to activate an alarm and quantifies the latency of the alarm system.

The knowledge of FAR, MAR and EDD is therefore important to include preventive measures in the system design. To reduce the number of unnecessary redundant alarms, a conventional approach is to review top ten worst alarms. In this method, from a list generated, alarms are reviewed one by one starting from the most frequent one and the root cause is identified [16]. Many of these generated alarms can simply be removed by properly adjusting alarm thresholds [39]. Alarm thresholds are normally set considering certain confidence range, e.g. $\mu \pm \sigma$ (where μ is mean and σ is standard deviation of the process variable) [37].

1.2 State of the Art

Alarm management has recently attracted academia as a potential research sector because of industrial demand. With the advancement of communication and computer technologies, the complexity of the process industries has increased many times. In order to maximize their installation reliability, efficiency, product quality and profitability, industry is now attempting to monitor thousands of process variables. The negative side is that hundreds or thousands of configured alarms are present and operators are lost in an information sea.

A lot of research has been done in this domain such as alarm system design, optimization and routing performance assessment [49]. Furthermore, there exist many problems in the designing path of alarm systems due to modeling noise and uncertainties.

According to many industrial studies, modern alarm systems suffer from alarm flooding; the reception of a large number of alarms that cannot be managed by the operator. These alarms are named nuisance alarms, came in many forms: chattering, repeating, staling alarms. The detection of chattering and repeated alarms has been investigated by [66] where an online method has been presented to detect and reduce them by regulating alarm threshold and employing delay-timers. As an extended work, by exploiting alarm duration and intervals, delay-timer has been designed to minimize nuisance alarms [67].

Alarming methods can be categorized into two categories; univariate and multivariate category. Despite multivariate method, univariate methods are massively applicable in industry, they are simple to design and implement since they use a single process variable. Threshold-based method, filters, dead-band and delay-timer are some of techniques employed in univariate alarm systems. Reference [62] has deployed rank order filters for alarm system design where FAR and MAR have been computed. Reference [4] presented time-dead band method where design procedures have been developed. Three types of delay-timer were introduced, conventional delay-timer [70], Genezalized delay-timer [2] as well as multi-setpoint delay-timer [61] for alarm system configuration. Based on an objective function (J), optimal alarm filter has been designed where J is minimal [13, 8]. However univariate alarming techniques are restricted in complicated alarming conditions.

Various techniques for managing these numerous alarms have been used in the literature. We review current and existing methods trends in the design and improvement of alarm systems in this sections.

1.2.1 Filters, Delay-timers and Dead-bands

Techniques like filters, dead-bands and delay timers are widely used in the process industry to reduce false and disturbing alarms. Rank filters are the most interesting among all non-linear filters because of two inherent characteristics: edge preservation and efficient noise attenuation with impulsive noise robustness [71]. Delay times are commonly used among all these methods for their simplicity and efficiency.

1.2.2 Alarm Threshold Design

In the development of the alarm system, univariate alarm systems is the most used method for design. Generally, one process variable is monitored and continually compared to a threshold. Incorrect setting of the threshold may, however, lead to false, missed or redundant alarms and may result in delayed alarm activation [35]. In [35, 36, 70] an overall design framework for alarm is introduced. The alarm system's performance is measured based on three metrics: FAR, MAR and the Expected Detection Delay (EDD). A threshold design process is presented in [35] for the various techniques of signal processing (filter, timer and dead-band).

1.2.3 Alarm Threshold Optimization

In current process industries, optimizing the alarm threshold is an important part of rationalizing the alarm system. In [21], the algorithm for change detection is proposed to obtain the adaptive alarm thresholds; it kept the false alarm rate constant, but failed to consider the influence of missed alarms. Nevertheless, existing alarm system design techniques focus mostly on one stable state of operation and are unable to deal effectively with alarm floods during transitions.

Bayesian estimation based on dynamic alarm management (BEDAM) method is applied to determine the dynamic alarm thresholds during transitions [42]. A new method for multivariate alarm thresholds is proposed in [26]. It sets only one limit for each alarm variable.

1.2.4 Alarm Priority Ranking

The alarm priority indicates the alarm's importance. The priority assignment method is applied to the rationalized alarm. Effective prioritization typically results in higher priorities being chosen less frequently than lower priorities [34]. Effective use of alarm priorities can enhance the operator's ability to manage alarms and provide a proper response.

In [12], an integrated model consisting of probability (P), impact (I) of potential hazards, and process safety time (t) is presented to prioritize the alarms raised. The proposed approach incorporates the process safety time rather than the response time of the operator. The Alarm Prioritization System (ALAP) is developed to calculate the severity of each alarm, categorize the alarms and rank the alarms based on their priority. The system will get and analyze trend data from DCS. To carry out fuzzy reasoning and achieve priority classification, the rules in the knowledge base will be combined. In addition, the capabilities of self-learning and self-tuning can be discussed so that in the near future the prototype system can perform automatic diagnosis of alarm patterns.

Multivariate methods, on the other hand, overcome limitations of the univariate methods; they are convenient in alarm management for situations where process variables relationships are very complex. Alarm threshold is an important part of alarm system design where FAR and MAR are directly affected. To optimize multivariate alarm thresholds, a new method based on FAR, MAR and correlation analysis is suggested in [26].

1.3 Report Contributions

The main objective of this thesis is to ***'design and assess a multivariate alarm system using the combined index φ'*** for a specific requirement on the FAR, MAR and EDD.

In spite of different studies and research works conducted so far, there are few works on using multivariate statistical approach with delay timers to improve the performance of the fault indicators, hence, the performance of alarm systems.

This study introduces the use of n-sample on-off delay-timer with a fixed threshold that has the capabilities of alarm nuisance reduction and performance improvement. In addition to that, the DIRECT based optimization technique is used to get optimal parameters for designing alarm systems [17]. Finally , an industrial case study is done to confirm the effectiveness of the proposed technique.

1.4 Report Organization

The organization of this report is as follows.

Chapter II discuss the necessary background used in the next chapters such as Multivariate Statistical Process Control(MVSPC), the Principal Component Analysis (PCA) and its use in fault detection and alarm generation. Also about Markov process that was used as a tool to compute the performance parameters.

In Chapter III, the design procedure of alarm system is illustrated with a simulated example.

Chapter IV shows a real process alarm system design and a discussion about the obtained results with a simulation and a comparison study. This thesis terminates with a conclusion and future works.

Chapter 2

Theoretical Background

2.1 Introduction

In this chapter, the background material necessary to understand the content of this thesis is discussed. In the following sections, an introduction to the performance specifications of alarm systems and configuration is given. To reduce alarm overflow, a delay timer is used which cause a time detection delay, this last is calculated using Markov processes. Therefore, both delay timer and Markov process are explained.

The proposed method for the design of alarm system is based on multivariate design method. So, a general framework for multivariate alarm design is introduced and the PCA method is explained step by step. Since the performance of the alarm system is evaluated in terms of three metrics, namely false alarm rate (FAR),missed alarm rate (MAR) and expected detection delay (EDD). Receiver Operating Characteristics (ROC) curve is introduced to visualize the trade-off on FAR and MAR with respect to the alarm threshold.

2.2 Multivariate Monitoring Process

In general, alarm systems can be designed according to the type of the system, univariate system design and multivariate system design. The most widely used statistical processing control (SPC) techniques involve univariate methods, mainly the observation and analysis of one variable at a time. In the univariate design, the alarm threshold and the processing technique (dead-zone, time-delay, filter, etc.) are individually designed for each process variable.

With the availability of DCS in almost any industrial plant, it's possible to measure hundreds, if not thousands of variables at a time, instead of just one. This huge amount of data is used for control and monitoring of the plant. As a result, univariate SPC methods and techniques provide little information about their mutual interactions. The variables measured are not all independent. Some of these variables are correlated due to sensor redundancy in the plant. Because of this correlation, when an anomaly occurs in the system, many process variables are affected. Monitoring all these variables will cause many nuisance alarms.

Multivariate Statistical Process Control (MVSPC) techniques simultaneously account for all variables of interest and can extract information about the behavior of each variable or characteristic relative to other variables by reducing dimensions and extracting features as it is illustrated by fig. 2.1. Multivariate methods are based on the idea that the behavior of a process can usually be expressed by a few independent variables, called latent variables. Latent (or virtual) variables are obtained as a linear combination of raw process measures. Several methods are available in the literature to obtain latent variables, including popular Principal Component Analysis (PCA) and Partial Least Squares (PLS) methods [20, 31, 56, 64, 33, 75].

In multivariate monitoring techniques, latent variables are monitored instead of individual process variables, resulting in fewer alarms for the same anomaly. Multivariate methods can manage the number of alarms and alarm delay, commonly alarm floods [60].

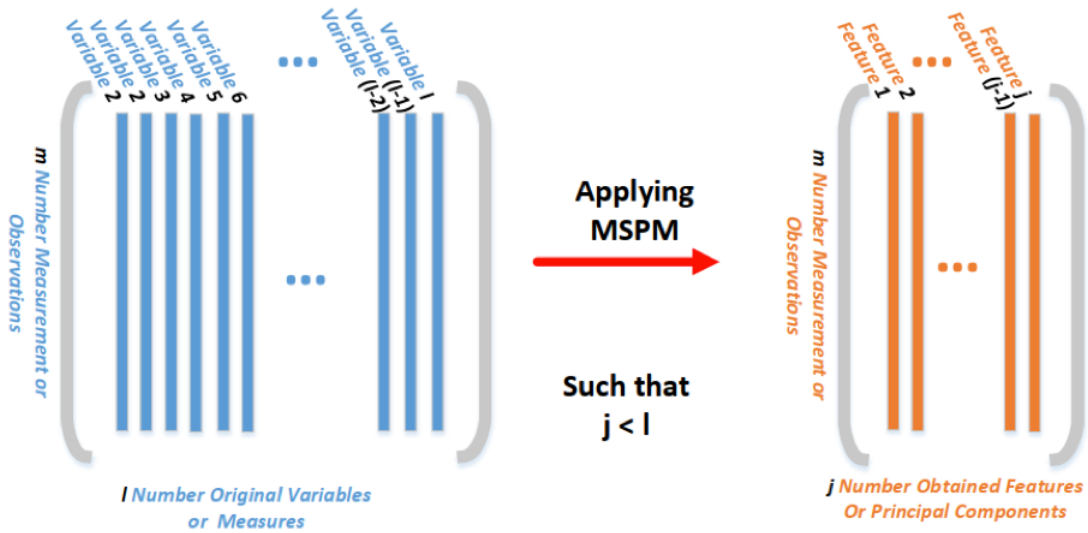


Fig. 2.1 Dimension Reduction and Feature Extraction by MSPM Methods.

2.3 Principal Component Analysis

Principal Component Analysis (PCA) is a standard statistical tool utilized for the monitoring of high-dimensional processes [1, 7, 27, 50, 5]. It is broadly used in nearly all areas of research where the manipulation of a large number of attributes is required.PCA is also used widely in statistical processes in general, usually used in MSPC [20] and in fault detection and isolation (FDI) [64]. It is a useful non-parametric method for obtaining relevant information from a complex data set.

PCA depends upon the eigen-decomposition of positive semi-definite matrices and upon the singular value decomposition (SVD) of rectangular matrices. PCA is an orthogonal transformation used to reduce the dimensionality of a data-set, which consists of a large number of interdependent attributes, while retaining as much variation as possible in the original data-set. This process is performed by projecting the original attribute set into two orthogonal sub spaces called Principal Components subspace (PCs) and the residual subspace (RS) [46]. The principal components are not correlated and are ordered so that the former retain most of the variations present in all the original attributes. So mainly (PCs) will capture the normal variation of data where RS ideally captures only noise . Thus, a reduced dimension PC model can be used to detect and diagnose abnormalities in the original system in a robust way.

The quality of the PCA model can be evaluated using cross-validation techniques such as the bootstrap, cross validation and the jackknife. PCA can be generalized as correspondence analysis (CA) in order to handle qualitative variables and as multiple factor analysis (MFA) in order to handle heterogeneous sets of variables. Mathematically, PCA depends upon the eigen-decomposition of positive semi-definite matrices and upon the singular value decomposition (SVD) of rectangular matrices[69]. Furthermore, among all linear dimension reduction techniques, PCA has the best mean-square error [68, 18].

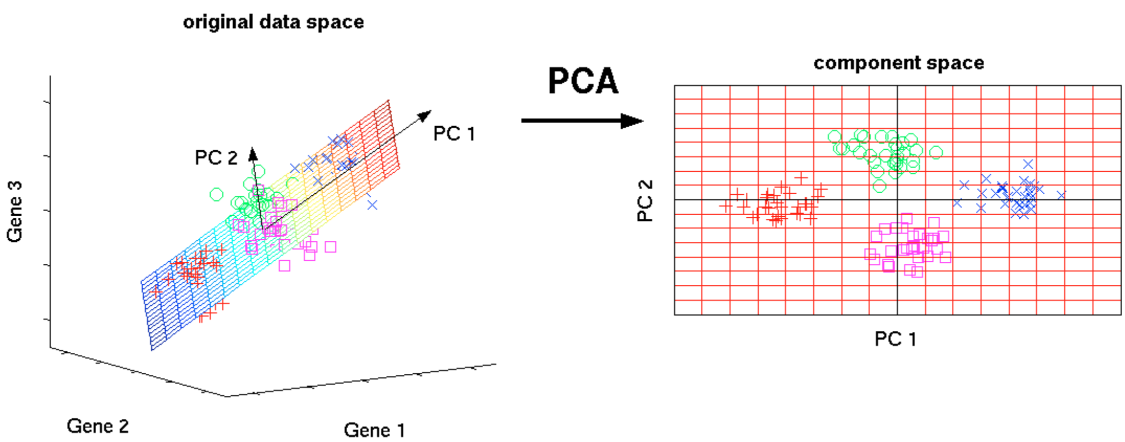


Fig. 2.2 Orthonormality and Dimension Reduction from 3D to 2D Using PCA [53]

2.3.1 Working Principle of PCA

Since a change of basis is applied, PCA is clearly assuming a linear relationship between the original data [65]. The main goals of Principal component analysis are:

1. Reducing the data dimensionality.

2. Transforming the data to a set of uncorrelated variables.

Let $X \in \mathbb{R}^{n \times m}$ be the original data matrix where m is the number of variables and n is the number of observations:

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{pmatrix} \tag{2.1}$$

Since different physical variables and sensors have different magnitudes and scales. Performing PCA on un-normalized variables will result in extremely heavy loads for high variance variables. Thus, in order to eliminate the effect of different scales and units, of the different sensors we normalize the data matrix by removing the mean of the data from each point and dividing it by its standard deviation using the scale parameter vectors $\bar{X}$ and ϵ as the mean and variance vectors, respectively. This will give a data matrix centered at the origin with zero mean and a unit variance. The PCA problem can be seen as the problem of finding an appropriate transformation $P \in \mathbb{R}^{m \times m}$ applied to X results in a new data set $T \in \mathbb{R}^{n \times m}$ i.e

$$T = XP \tag{2.2}$$

PCA is used to reduce the dimensions of the dataset while maintaining the maximum variation of the original dataset. The covariance matrix defines both the spread (variance) and the orientation (covariance) of the dataset. Therefore, the covariance/correlation matrix S of the dataset is constructed as follow:

$$\Sigma \approx \frac{1}{n-1} X^T X \tag{2.3}$$

Where our data set is expressed by the matrix $X \in \mathbb{R}^{n \times m}$. The covariance matrix as represented in eq. (2.3) is a biased estimator.

Using Singular Value Decomposition (SVD) on Σ yields to:

$$\Sigma = P \Lambda P^T \tag{2.4}$$

We can notice that the covariance (correlation) matrix is *symmetric* and P is *orthonormal*. Where:

- $PP^T = P^T P = I$
- Λ: is a diagonal matrix whose non-zero elements are the corresponding eigenvalues λ_i of the covariance matrix Σ.
- P: is the loading matrix ($m \times m$) consisting of the eigenvectors corresponding to the eigenvalues of Σ.

The eigenvalues in the diagonal matrix Λ represent the variance and the corresponding eigenvectors represent the direction. Hence arranging the eigenvalues in a decreasing order gives an ordered orthogonal basis from the greatest variance to the smallest one.

Let $T \in \mathbb{R}^{n\times m}$ as given by eq. (2.2) be the score matrix. Its columns are called score vectors ti, each of them associated with the corresponding principal component PCi is the score matrix consisting of the component vectors.Equation (2.5) gives the relation between the original data matrix X and the transformation matrix P.

$$X = TP^T \tag{2.5}$$

If only j number of PCs is kept from an original data set of m variables the previous matrices are expressed as follow:

$$P = [\hat{P}\tilde{P}]; \hat{P} \in \mathbb{R}^{m\times j} and \tilde{P} \in \mathbb{R}^{m\times(m-j)} \tag{2.6}$$

$$T = [\hat{T}\tilde{T}]; \hat{T} \in \mathbb{R}^{n\times j} and \tilde{T} \in \mathbb{R}^{n\times(m-j)} \tag{2.7}$$

$$\Lambda = \left(\begin{array}{c|c} \hat{\Lambda} & 0 \\ \hline 0 & \tilde{\Lambda} \end{array}\right); \hat{\Lambda} \in \mathbb{R}^{j\times j} and \tilde{\Lambda} \in \mathbb{R}^{(m-j)\times(m-j)} \tag{2.8}$$

We then express the data matrix X as the combination of the modeled variations and the non-modeled variation of X by the projection on the principle space and the residual space as follow:

$$X = \hat{X} + E \tag{2.9}$$

$$\hat{X} = \hat{T}\hat{P}^T \tag{2.10}$$

$$E = \tilde{T}\tilde{P}^T \tag{2.11}$$

In summary, to get the PCA model we follow the next steps:

1. Normalize the original data matrix X.
2. Calculate the covariance matrix Σ using eq. (2.3).
3. Perform SVD on Σ as given in eq. (2.4).

   where Λ is a diagonal matrix whose elements are the eigenvalues(λ_i) of Σ sorted in descending order, columns of matrix P are the eigenvectors of Σ.
4. Determine the appropriate number ‘j’ of principal components PCs using any method exists in the literature.

5. Compute the loading matrix $\hat{P}$ whose elements are the first 'j' columns of P.

6. compute the projection matrix C and $\tilde{C}$ using $\hat{P}$ and equation 2.12.

$$C = \hat{P}\hat{P}^T; \tilde{C} = \tilde{P}\tilde{P}^T = (I - C) \tag{2.12}$$

The original data matrix which has m dimension space is now substituted by the 'j' PCs and $(m - j)$ RS and then all the linear correlation between the measured variables is now removed.
When a new measurement vector $x \in \mathbb{R}^{1\times m}$ observed, it can be decomposed into two parts, using PCA model

$$x = \hat{x} + e \tag{2.13}$$

where $\hat{x}$ and e are the projection of the test vector x in the two orthogonal principal and residual subspaces, respectively. Hence, the PCA model is ready now to be used for fault detection and alarm generation.

2.3.2 Choosing the number of principal component (PCs)

Choosing the number of principle components representing the original dataset is one of the most crucial parts and plays a major role in the goodness of the PCA model. However, an exact method to estimate the number of retained components that is guaranteed to pick out the best number of PCs still unavailable; even though, many methods exists in the literature, depending on the application. In this section, some of the widely used methods are presented.

Guttman-Kaiser criteria

Kaiser-Guttman is one of the simplest methods for which we consider the number of principal components.It was found in the early 1954, and still being the most used method for the estimation of the number of components in PCA and factor analysis. It consists of picking eigenvalues equal to or greater than one (the average eigenvalue), to have a practical significance. This means that we only take components whose eigenvalues are greater than 1 [25]. A modification to the Kaiser criterion was suggested in [44] in order to deal with the significance problem, a reasonable choice is to select a cut-off less than 1, the value proposed in [45] was 0.7 which will be referred to as J7 criteria. The difference between the number of picked components using Kaiser's and Jolliffe's methods can be dramatic.

Cumulative Percent of Variance (CPV)

Another commonly used method for adaptive models proposed in [76] is the percentage cumulative variation CPV, in which the number of PCs we retain (j) contributes to a specific cumulative percentage of the total change in original data, calculated as follows:

$$CPV = \frac{\sum_{i=1}^{j} \lambda_i}{\sum_{i=1}^{m} \lambda_i} \times 100 \tag{2.14}$$

Choosing high CPV will result in a very noisy-sensitive model. On the other hand, low CPV gives a model that is less sensitive to anomalies.Usually, a percentage between $80-85\%$ is used in some context while some references encourages the use of percentages greater than 85%, but the optimal selection that guarantees the best representation for the data is a data-dependent and still out of reach. The optimal value of the CPV that best represents the data may vary widely according to the amount of noise present in the measurement and the redundancy between the different variables. Due to the absence of a clear rule, CPV turns out to be subjective and the fact that the cumulative percentage increases with the increase of the number of components makes the method uncertain.

Cross Validation

Cross validation (CV) is one of the technique used to test the effectiveness of a machine learning models, it is also a re-sampling procedure used to evaluate a model if we have a limited data. CV works based on the principle of dividing the available set of data into two parts, training and testing sets. The training set is used to estimate a model while the testing set is used to validate that model [57]. Basically, a PCA model have to be constructed for $i=1,2,\cdots,m$ and the one that best satisfies the selected GFC is selected. Reference [7] discussed Cross validation in details.

Parallel Analysis

Parallel analysis is a sample-based adaptation for the population-based k1 rule, which implies Monte Carlo simulation process [29]. The methodology is based on comparing the eigenvalues of the Covariance/Correlation matrix of the original data set with those of a randomly sampled data set from normally distributed population. A component is considered to be significant once the associated eigenvalue is larger than the one generated from the random sample. Based on the fact that a method based on random sampling may exhibit a large variability, it is usually required to repeat the generation of the pseudo-eigenvalues for k-times. Having a set of ordered pseudo-eigenvalues for each experiment, one could increase the estimation by taking either the average eigenvalue for each component or by assessing the empirical distribution of the eigenvalues associated with a given component to determine an upper confidence limit for each pseudo-eigenvalue with a pre-selected percentile [23].

The algorithm of PA as stated in [29] is summarized as follow:

1. Calculate the eigenvalues of the covariance/correlation matrix: $\lambda = \{\lambda_1, \lambda_2, \cdots, \lambda_m\}$

2. Generate k-sets of random data with the same size as X ($Z_i \in \mathbb{R}^{n \times m}, i = 1, 2, \cdots, k$)

3. Compute the eigenvalues of Z_i and store them in $\Xi \in \mathbb{R}^{k \times m}$ (Row-wise).Wehre ξ_{ij} is the j^{th} eigenvalue of the i^{th} generated data-set.

$$\Xi = \begin{bmatrix} \xi_{11} & \xi_{12} & \cdots & \xi_{1m} \\ \xi_{21} & \xi_{22} & \cdots & \xi_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \xi_{k1} & \xi_{k2} & \cdots & \xi_{km} \end{bmatrix} \tag{2.15}$$

   Do either:

   - Compute the average eigenvalue $\hat{\Xi} = [\bar{\xi_1}, \bar{\xi_2}, \cdots, \bar{\xi_m}]$
   - Compute the α-percentile of each column $\hat{\Xi} = [\xi_{1\alpha}, \xi_{2\alpha}, \cdots, \xi_{m\alpha}]$ where α usually selected between 0.95 and 0.99.

4. Retain only the components for which $\hat{\Xi}_i > \lambda_i$.

The study carried in [77] concludes that PA is the best of the studied methods, it was found to be able to retain the exact number of components in about 92% of the cases with 66% of the estimation errors are over estimations. Furthermore, PA was found to be insensitive to the distributional form of the data [14, 19].

There exist many other methods to determine the required number of principal components such as Scree-plot, Bootstrapp, Minimum Average Partial (MAP)and Broken Stick method (BS). These methods are explained in details in [7, 19, 14].

2.3.3 PCA in Multivariate Systems

PCA is often used in multivariate alarm systems as a modeling tool either PCA-Based Fault Detection or PCA-Based Fault identification.

Hotelling's T^2

Hotelling's T^2 statistic measures variations in PCS [30], it is a fault detection index and it was proposed by Hotelling (1947)[31], it is given by:

$$T^2 = \sum_{i=1}^{n} \frac{T_{ij}}{\lambda_j} = x^T P \Lambda^{-1} P^T x \tag{2.16}$$

Its control limit noted by T_α decides weather the process is healthy or faulty, and it is given by eq. (2.17).

$$T_\alpha = \frac{(n^2-1)a}{n(n-a)} F_\alpha(a, n-a) \tag{2.17}$$

where $F_\alpha(a, n-a)$ is the critical value of the Fisher–Snedecor distribution with a and $(n-a)$ degrees of freedom. i.e a is the number of retained principal component.

Q Statistics

The Squared Prediction Error (SPE) index which is known also as Q statistics measures the projection of the sample vector on the residual subspace (RS) [40]. It is a measure of the difference or residual between the sample and its projection in the model.

$$Q = ||e||^2 = ||(I-C)x||^2 \tag{2.18}$$

Jackson and Mudholkar [1979]([20, 40]) developed an expression for the control limit of SPE noted by Q_α. The process is under healthy state if inequality given in eq. (2.19) is satisfied.

$$SPE \leq Q_\alpha \tag{2.19}$$

And:

$$Q_\alpha = \theta_1 \left[c_\alpha \frac{h_0\sqrt{2\theta_2}}{\theta_1} + 1 + \frac{\theta_2 \theta_0 (\theta_0 - 1)}{\theta_1^2} \right]^{\frac{1}{\theta_0}} \tag{2.20}$$

where: $\theta_i = \sum_{j=a+1}^{m} \lambda_j^i$, for $i = 1, 2, 3$; and $h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2}$

and c_α represents the $1-\alpha$ percentile of the standard normal distribution.

In practice, a single index is better used than both T2 and Q for process monitoring. Yue and Qin [73] have proposed a practical alternative, namely a combined index.

The Combined Index φ

Reference [73] propose an index which is a combination between SPE and T^2 , while reference [56] suggest a combined index without given its control limit.

$$\varphi = \frac{SPE(x)}{Q_\alpha} + \frac{T^2(x)}{T_\alpha} = x^T \Phi x \sim g\chi^2 \tag{2.21}$$

We use the approximate distribution to calculate the confidence limits of the combined index. The distribution of φ can be approximated using $g\chi^2(h)$.

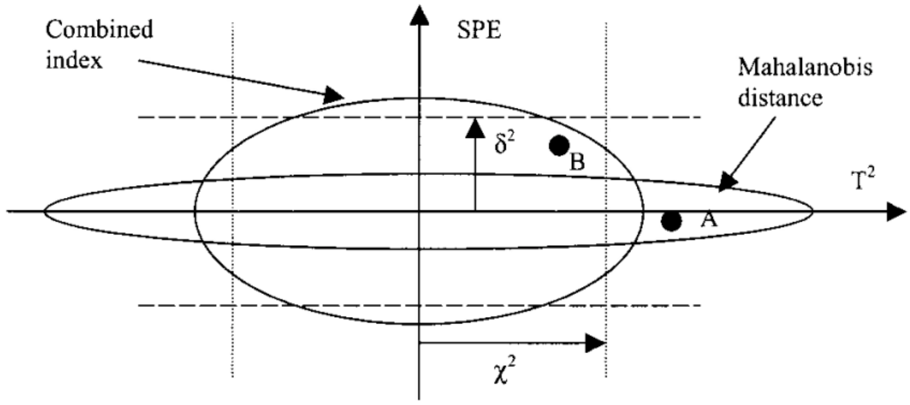


Fig. 2.3 Comparison between confidence regions for four statistical indices.

In this study, we will use this index to perform alarm design and performance assessment on the cement plant process.

2.4 Alarm Performance Assessment

The decision to trigger an alarm always depends on a single variable that can be a real process variable or a latent variable (multivariate techniques). The alarm is always triggered by this variable, exceeding a certain limit [38, 35]. However, incorrect settings for alarm limits or thresholds result in a poorly performing alarm system. To evaluate the performance of the alarm system, we must calculate the accuracy of the design False Alarm Rate (FAR) and Missed Alarm Rate (MAR). Both provide a measure of the accuracy of alarm systems.

Considering the process variable in fig. 2.4. Assume that the first and second parts of the data constitute both normal and abnormal conditions, respectively. The ideal monitoring system should not trigger any alarm for the first part and the second part should be continuously alarmed. Therefore, an alarm triggered in the first part is seen as a false alarm. Likewise, in the second part if the alarm has been removed, it is considered to be an alarm that is missed. Good design with higher precision and accuracy shows low, false and missing alarm rates.

According to the minimum probability of error and the Bayesian decision theory [32], the false alarm probability is the area under the normal data distribution curve (p1 in fig. 2.4), and given by eq. (2.22). The missed alarm rate is the area under the abnormal data distribution curve (q1 in fig. 2.4), and given byeq. (2.23) [70, 32, 54].

$$FAR = \int_{x_{tp}}^{+\infty} q(x)dx \tag{2.22}$$

$$MAR = \int_{-\infty}^{x_{tp}} p(x)dx \tag{2.23}$$

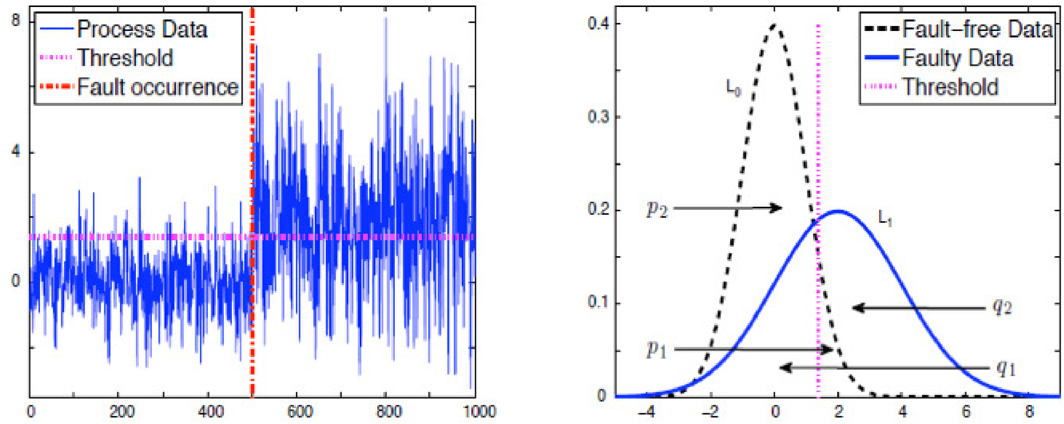


Fig. 2.4 Process data and corresponding PDFs(p1,p2,q1,q2 denotes probabilities and L0,L1 denotes likelihoods for fault-free and faulty casses).

where q(x) is the pdf of process variable (latent variable) in the normal state (L_0) and p(x) is the pdf of process variable (latent variable) in the abnormal state (L_1).

For the alarm variable of fig. 2.4, if a lower threshold is selected, there will be a lot of false alarms. However, only a few alarms will be missed. On the other hand, if a high threshold is selected, the number of false alarms will be significantly reduced, but many alarms will be missed. Therefore, a tight alarm limit will cause many false alarms, but few missed alarms. And a loose alarm limit will only cause a few false alarms, but a lot of missed alarms. This is a very important alarm system design trade-off: the false alarm rate compared to the missed alarm rate.

The accuracy of the alarm system can be demonstrated by a receiver operating characteristic curve (ROC) [74]. The ROC curve is the plot of the probability of missed alarm versus the one of false alarms as the trip point changes from $-\infty$ to $+\infty$. Since the design of alarm system seeks to minimize the false alarm and missed alarm rates, the ideal point in the ROC curve is the origin. However, in most cases this is not possible because of the overlap between normal and abnormal data distributions [59]. The optimum curve point is then the closest point to the origin as shown in fig. 2.5. The ROC curve and the optimum point also depend on types of alarm design method used and weight given on FAR and MAR.

Other than the false and missed alarm rates, the time required to trigger an alarm is another important performance indicator in alarm systems. Once an anomaly has occurred, alarms may not be raised instantly due to various system delays brought about by various reasons including network delays, poor implementation, hardware problems, sensor failure and data loss [3]. We call this time difference between the actual instance of the fault and the one when an alarm is raised by detection delay.

Detection Delay = Fault occurrence - Alarm rising

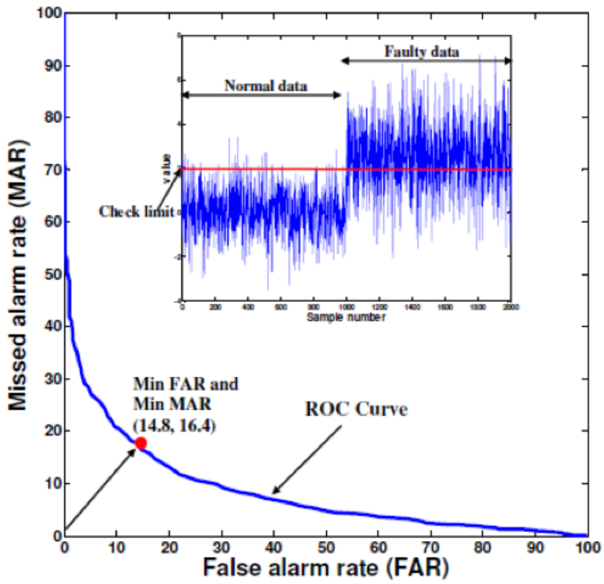


Fig. 2.5 Example of a simple ROC curve showing the selection of the best threshold.

Techniques such as delay timers, dead-bands and filters are widely utilized in alarm systems and enhance the effectiveness of the control method, but also increase detection time [2]. However, delays in detection may continue to happen even if the above techniques are not used, as it is directly related to the position of alarm threshold.

For safety reasons it is important to estimate the average time required to trigger an alarm known as the expected detection delay (EDD), which provides a measure of the latency of the alarm systems.

2.5 Processing the Alarm Variable

Regardless of the model's accuracy or design quality, the alarm variable carries some noise. Due to this noise, comparing the alarm variable to a threshold may cause false and chattering alarms. To improve alarm quality and reduce the number of false warnings, the alarm variable should be processed. Processing reduces variable noise and simplifies the design of the threshold. PCA itself considered as noise decoupling tool that proves its efficiency in noise rejection. Simple processing techniques can reduce false alarms if used correctly, such as filtering, dead-band and alarm delays [38, 3]. In this thesis, only delay timer would be discussed but to understand the working principle of the delay timer first, the Markov chain process has to be explained.

2.6 Markov Process

A *Markov process* is an independent process where outcome at any time instant depends only on the outcome that precedes it and none before that [55]. In this thesis Markov processes are used to estimate the FAR, MAR and EDD for delay-timers. One can describe the Markov chain process as follow:

Let $\{X_t, t = 0, 1, 2, \cdots\}$ be a finite set of random variables indexed by time t, and therefore $\{X_t, t = 0, 1, 2, \cdots\}$ is referred to as a discrete stochastic time process. This random variables set takes values from $\{i = 0, 1, 2, \cdots\}$ non-negative integer set, unless otherwise specified. For example, if $X_t = i$ then it is said that the process is in state i at time t [72].

A stochastic process $\{X_t\}$ is said to have the Markovian property if:

$$P\{X_{t+1} = j \mid X_t = i, X_{t-1} = k_{t-1}, \cdots, X_1 = k_1, X_0 = k_0\} = P\{X_{t+1} = j \mid X_t = i\} \tag{2.24}$$

for $(t = 0, 1, 2, \cdots)$, and all states $(i, j, k_0, k_1, \cdots, k_{t-1})$.

The conditional probabilities $P = \{X_{t+1} = j \mid X_t = i\}$ are called one-step transition probabilities for each i and j, if:

$$P\{X_{t+1} = j \mid X_t = i\} = P\{X_1 = j \mid X_0 = i\} \tag{2.25}$$

then the Markov chain is said to be stationary and $P = \{X_{t+1} = j \mid X_t = i\}$ is usually denoted by P_{ij}. Since the one-step transition probabilities do not depend on the step numbers, this implies that the transition probabilities do not change in time. The n-step transition probabilities, usually denoted by $P(n)_{ij}$, are defined as follows: For each i, j, and n (where $n = 1, 2, \cdots$):

$$P^(n)_{ij} = P\{X_{t+n} = j \mid X_t = i\} = P\{X_n = j \mid X_0 = i\} \tag{2.26}$$

for all $t = 0, 1, 2, \cdots$.

One method for calculating these n-step transition probabilities is the Chapman-Kolmogorov [1] equations:

$$P^{(n)}_{ij} = P^{(r)}_{ik} P^{(n-r)}_{kj} \tag{2.27}$$

for all i, j, n,and $0 < r < n$.

Applying the Chapman-Kolmogorov equations, it is easy to show that :

$$P^{(n)}_{ij} = P^{n}_{ij} = P_{ij} P^{n-1}_{ij} = P^{n-1}_{ij} P_{ij} \tag{2.28}$$

[1] An identity relating the joint probability distributions of different sets of coordinates on a stochastic process.

Where P is the one-step transition matrix, thus the n-step transition matrix can be obtained by calculating the nth power of the one-step transition matrix [63].

A Markov chain is ergodic if there is a positive probability to pass from any state to any other state in one step. A Markov chain with more than one state and just one out-going transition per state is either not irreducible or not aperiodic, hence cannot be ergodic [22]. For a Markov chain, a simple test for ergodicity is using eigenvalues of its transition matrix. The number 1 is always an eigenvalue. If all other eigenvalues are positive and less than 1, then the Markov chain is ergodic. This follows from the spectral decomposition of a non-symmetric matrix.

To summarize everything if S Is a set of independent states; the process begins in one of these states and moves from one state to another successively. If the chain is currently in state S_i, then at the next step it moves to state S_j with a probability denoted by P_{ij} (transition probability) which does not depends on which state the chain was in before the current state. With probability P_{ii}, the process can also remain in its current state i. The starting state is specified by the initial probability distribution, defined on S. This is usually done by specifying a certain state as the initial state [72, 47, 55].

If we consider a Markov Process with a limited number of states, we can define the probability matrix (Transition probability matrix) as follow:

$$P = \begin{pmatrix} P_{11} & P_{12} & \cdots & P_{1j} & \cdots \\ P_{21} & P_{22} & \cdots & P_{2j} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ P_{i1} & P_{i2} & \cdots & P_{ij} & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix} \tag{2.29}$$

Where P_{ij} are transition probabilities [47].

A probability vector π is called invariant for the Markov process if $\pi = \pi P$. Alternatively, π is a left eigenvector of P with eigenvalue 1 [52].

For the left eigenvector π of the matrix P to exist ($\pi = \pi P$) the Markov process must satisfy the following two conditions as explained in [70, 3, 2, 54]:

1. The summation of all entries in each row of P is 1 ($\sum_j P_{ij} = 1$).

2. All the entries of P are non-negative $P_{ij} > 0$.

To satisfy these conditions and the definition of Markov process, we make the following assumptions on the process data:

- Process data is independent and identically distributed (I.I.D.), i.e., at each sampling instant, the process data (random variable) has the same probability distribution as the other instants and all are mutually independent.

- Probability density functions of the fault free and faulty data are known. These distributions can be estimated from historical data.

In alarm systems, a process might be in two states from the alarm perspective: alarm state (A) and no alarm state (NA). In this thesis, Markov processes are used to model these states, estimate the detection delay for delay-timers and calculate the probabilities of false and missed alarms.

2.7 Conclusion

In this chapter, PCA as a multivariate statistical approach is presented. The mathematical background of PCA models was explained as well as the selection of the model parameters, such as the number of principal components and the loading matrix. Furthermore, the use of PCA-based fault detection and identification using the Hotelling's T^2 and Q statistics was presented with an additional index called the Combined ϕ which proves its effectiveness as a single index. In addition to that, alarm performance assessment was discussed with the aid of Markov process.

Chapter 3

Alarm System Design Using Delay-Timers

3.1 Introduction

A *delay-timer* is a simple but efficient way to reduce false and missed alarm rates and make the system less sensitive to noise and external process problems. Generally, an alert is decided on the basis of one sample; that is, if one sample exceeds the alarm limits, it should be raised and the alarm should be removed if one sample returns to an acceptable operating range. In practice, operators only wait a while before taking the appropriate action to ensure that the alarm is persistent [38]. This can be used to design alarms to improve the quality of the surveillance system. In such case, only one sample exceeding alarm limits will not raise the alarm.

In alarm systems, two types of delays are applied: On and off delay timers. In general, on delays timers are useful to prevent fleet alarms, and off delays timers can lock the alarm when repetitive alarms are raised. The main design parameter of a delay-timer is its length n. Although this technique may significantly reduce false alarms; detection-delay is the unwanted consequence in alarm activation [2, 74].

3.2 Working Principle of Delay-Timers

A standard n-sample on-delay is an alarm configuration triggered by the system status switching from non-alarm NA to alarm state A when the variable continually exceeds the alarm threshold of at least n samples consecutively in normal conditions. Similarly, once the alarm is triggered, it is only clear if m consecutive samples fall below the alarm threshold and the system switches from alarm A to non-alarm NA. This is referred to as an m-sample off-delay timer configuration. For on-delay, if the system returns to normal operating status during the intermediate states, alarm is not activated and vice versa for the

Table 3.1 EEMUA Recommendations for alarm delay

Signal Type	Recommended Maximum Delay
Flow	15 sec
Level	60 sec
Pressure	15 sec
Temperature	60 sec

off-delay case. In terms of safety, on-delay timers are more sophisticated than off-delay timers since are associated with the alarm activation [2, 38, 35, 70].

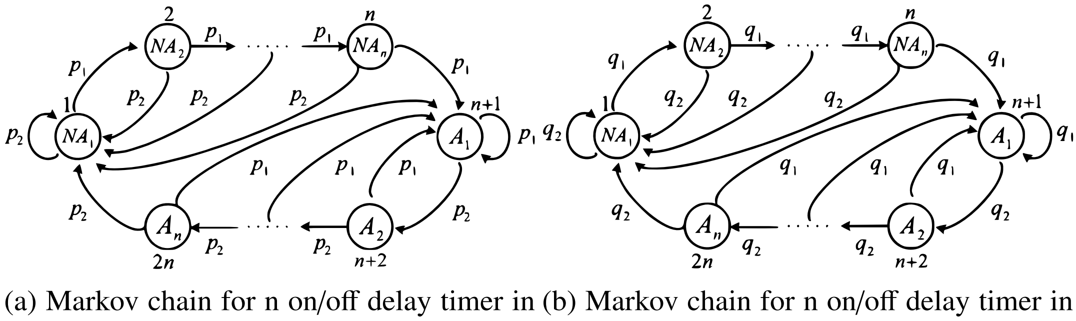
As the system should work in the normal range, returning to normal state (alert clearing) should be easier than drifting out of it (alarm raising). Therefore, m is normally chosen to be less than n. We assume, however, $m = n$ in our work for simplicity. Delays in the various loops as shown in table 3.1 are recommended by industry standards (EEMUA, 2007) [38].

Increasing the delay timer sample number will reduce the FAR and MAR rates, while increasing the time delay [74]. Thus the threshold limit and the number of delay number n form the two main alarm design parameters.

3.3 Markov chain & Delay-Timers

A Markov chain diagram can model the transition between the alarm and non-alarm states of a process variable with a delay-timer.

Assuming that in a fault-free operating region, the probability that one sample exceeds the alarm limit is p_1 and that one sample is within the alarm limit is p_2. As with the faulty area of operation, q_1 is the probability that one sample is within the alarm limit, and q_2 is the probability that one sample is above the alarm limit. Thus p_2 is $1-p_1$ and q_2 is $1-q_1$ and it is possible to construct a Markov process for both normal and abnormal operating region as shown in fig. 3.1.



(a) Markov chain for n on/off delay timer in normal condition

(b) Markov chain for n on/off delay timer in abnormal condition

Fig. 3.1 Markov diagrams for normal and abnormal condition

Provided that the process variable is initially within the normal operating range, without an alarm (NA state in the Markov process). If the next sample remains within the range (with a probability of p_2), no alarm will be triggered and NA status will remain. If a sample exceeds the alarm limit (with probability p_1), the system will switch to an intermediate state, NA_1, but no alert will be activated. In NA_1 state, if the following sample does not exceed the threshold, the system will return to NA. The anomalous sample is viewed as an outlier for this situation and has no impact. On the other hand, if the following sample is out of range (with a probability of p_1), the system will enter a second intermediate state NA_2 and this procedure will be delayed.
In each of NA_i's intermediate states, if a sample returns to normal, the Markov process returns to the NA state. If the process is in the state of NA_{n-1}, the last $n-1$ samples are all outside the acceptable range. If another sample exceeds the limit, the alarm is triggered. Therefore, the process will go to alarm state A with a probability p_1. A similar reasoning is used to clear the alarm and return the process to the non-alarm NA's state. Note that the alarm is always triggered in intermediate states A_i and will not be cleared until the process reaches NA state [35, 37, 38].

3.4 Performance Indices for Conventional On/Off Delay

In this section, we will propose expressions for the false alarm rate (FAR), missed alarm rate (MAR) and the detection delay (EDD) using the Markov process and the *conventional* on/off *delay-timer*.
It is very important to conduct a careful safety analysis in designing the delay-timer parameters (time and alarm limit).

3.4.1 Probability of False Alarm

If the system is supposed to work in the normal region, any triggered alarm will be regarded as a false alarm. According to the Markov's proposed process the probability of false alarm is then the sum of the probabilities of alarm states (A_i).

$$P(\text{false alarm}) = P_{FA} = P(A_1) + P(A_2) + \cdots + P(A_n) \tag{3.1}$$

Here, $P(A_i)$ indicates the probability of the steady state A_i and can be calculated using the Markov chain theory. For the Markov diagram shown in fig. 3.1a, the probability of false alarm in the case of n-sample alarm on/off delay is:

$$FAR = \frac{p_1^n \sum_{i=0}^{n-1} p_2^i}{p_1^n \sum_{i=0}^{n-1} p_2^i + p_2^n \sum_{i=0}^{n-1} p_1^i}. \tag{3.2}$$

Proof:

Let T_{ik} be the number of steps taken for the transmission from state i to another state k, and $P_{ik}^{(l)}$ as the probability of $T_{ik} = l$ thus:

$$P_{ik}^{(l)} = P(T_{ik} = l)$$

For $l = 1$, the matrix P of one step transition probability is:

$$P = \begin{bmatrix} p_2 & p_1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ p_2 & 0 & p_1 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ p_2 & 0 & \cdots & 0 & p_1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & p_1 & p_2 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & p_1 & 0 & p_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & p_1 & 0 & 0 & \cdots & p_2 \\ p_2 & 0 & 0 & \cdots & p_1 & 0 & 0 & \cdots & 0 \end{bmatrix} \tag{3.3}$$

Here, the element locating at i^{th} row and the j^{th} column of the matrix is the one step transition probability of the state from i to j. Based on the theory of the stationary distribution.

For an irreducible ergodic Markov chain, the limiting probability $\pi_k = \lim_{l\to\infty} P_{ik}^{(l)} > 0$ exists ,being independent of the initial state, and satisfy the equality (3.4)

$$\sum_k \pi_k = 1 \tag{3.4}$$

Because the Markov chain has only a finite number of states, these limiting probabilities satisfy Equation (3.5)

$$\Pi = \Pi P \tag{3.5}$$

Where:

$$\Pi = [\pi_1, \pi_2, \cdots, \pi_{2n}] \tag{3.6}$$

Rewriting Equation (3.5) as:

$$\begin{cases} P_2(\pi_1+\pi_2+\cdots+\pi_n)+p_2\pi_{2n}=\pi_1 \\ p_1\pi_1=\pi_2 \\ p_1\pi_2=\pi_3 \\ \vdots \\ p_1\pi_{n-1}=\pi_n \\ P_1(\pi_n+\pi_{n+1}+\cdots+\pi_{2n})=\pi_{n+1} \\ p_2\pi_{n+1}=\pi_{n+2} \\ \vdots \\ p_2\pi_{2n-1}=\pi_{2n} \end{cases} \tag{3.7}$$

From eq. (3.7) we have:

$$\begin{cases} \pi_2=p_1\pi_1 \\ \vdots \\ \pi_n=p_1^{n-1}\pi_1 \end{cases} \tag{3.8}$$

and

$$\begin{cases} \pi_{n+2}=p_2\pi_{n+1} \\ \vdots \\ \pi_{2n}=p_2^{n-1}\pi_{n+1} \end{cases} \tag{3.9}$$

Using eq. (3.8),eq. (3.9) and eq. (3.4), the first equality in eq. (3.7) can be, respectively, written in the following form:

$$\begin{cases} \pi_1(1+p_1+\cdots+p_1^{n-1})+\pi_{n+1}(1+P_2+\cdots+P_2^{n-1})=1 \\ P_2\pi_1(1+p_1+\cdots+p_1^{n-1})+p_2^n\pi_{n+1}=\pi_1 \end{cases}$$

So, we can easily obtain π_{n+1} as:

$$\pi_{n+1}=\frac{p_1^n}{p_1^n(1+p_2+\cdots+p_2^{n-1})+p_2^n(1+p_1+\cdots+p_1^{n-1})} \tag{3.10}$$

From eq. (3.1) we get:

$$\begin{aligned} FAR &= P(A_1)+P(A_2)+\cdots+P(A_n) \\ &= \pi_{n+1}+\cdots+\pi_{2n} \\ &= \pi_{n+1}(1+p_2+\cdots+p_2^{n-1}) \\ &= \frac{p_1^n(1+p_2+\cdots+p_2^{n-1})}{p_1^n(1+p_2+\cdots+p_2^{n-1})+p_2^n(1+p_1+\cdots+p_1^{n-1})} \end{aligned}$$

which proves Equation (3.2).

3.4.2 Probability of Missed Alarm

The missed alarm rate is the sum of all probabilities of no alarm states in the Markov chain under the faulty region (fig. 3.1b) of operation when delay-timers are applied.$P_{MA} = P(NA_1)+P(NA_2)+\cdots+P(NA_n)$ The same procedure can be followed for MAR calculation. With the difference that probability density function for abnormal data is used. If q_1 is the probability of one sample of abnormal data falling within the normal range and $q_2 = 1-q_1$. For calculation of MAR, the probability distribution of faulty data and the transition probability matrix, Q should be used, where Q can be obtained replacing probabilities p_1, p_2 by q_2, q_1, respectively, in equation 3.3, the missed alarm probability is given by :

$$MAR = \frac{q_1^n \sum_{i=0}^{n-1} q_2^i}{q_1^n \sum_{i=0}^{n-1} q_2^i + q_2^n \sum_{i=0}^{n-1} q_1^i}. \tag{3.11}$$

where the matrix $Q \in \mathbb{R}^{2n\times 2n}$ is:

$$Q = \begin{bmatrix} q_1 & q_2 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ q_1 & 0 & q_2 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ q_1 & 0 & \cdots & 0 & q_2 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & q_2 & q_1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & q_2 & 0 & q_1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & q_2 & 0 & 0 & \cdots & q_1 \\ q_1 & 0 & 0 & \cdots & q_2 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

Equation (3.11) can be proved by taking a procedure similar to that in the proof of FAR.

3.4.3 The Expected Detection Delay

Besides the false alarm and missed alarm rates, detection delay is another important performance index in alarm systems. Detection delay is the difference of samples between the actual instance of fault occurrence and the instance of alarm raised. If a process variable moves from fault-free region of operation into faulty region of operation at time T_o and alarm is raised at time T_r, then detection delay (DD) is given by:

$$DD = T_r - T_o \tag{3.12}$$

Assume the system is in the normal state of operation. Once a fault occurs, the time required to move from no alarm (NA_i) states to alarm (A_j) states is the detection delay. Using only the first transitions from NA_i to A in fig. 3.1a and at the moment of fault occurrence, the Markov model of the system switches from the fault-free model (represented by P) to the faulty model (represented by Q). To estimate the EDD, concept of *hitting time* is used. If the Markov model is divided into two sub-spaces, D for all alarm states and ε for all no alarm states, the hitting time is the time required to move to D from ε assuming the system initially started in sub-space ε [52]. The expected delay is given by [2]:

$$EDD = \frac{p_2^{m-1}\left(p_1^n q_1 \sum_{i=0}^{n-1} q_2^i + P_2\left(\sum_{j=0}^{n-1} p_1^j \sum_{k=0}^{n-j-1} q_2^k - q_2^n \sum_{i=0}^{n-1} p_1^i\right)\right)}{q_2^n\left(p_2^m \sum_{i=0}^{n-1} p_1^i + p_1^n \sum_{i=0}^{m-1} p_2^i\right)} \tag{3.13}$$

Equation (3.13) can be modified for different values of m and n to get special cases. In this study we take m=n for simplicity. The solution procedure can refer to reference [3].

3.5 Alarm System Performance

In this section, we discuss the advantages and disadvantages of the conventional delay-timer in terms of three criteria, namely, *accuracy* (ROC curve of false and missed alarm rates), *latency* (detection delay with respect to change in parameter design (settings)) and *sensitivity* (change in performance with any change in design settings).

3.5.1 *Accuracy and latency*

The receiver operating characteristic (ROC) curve is the plot of two alternatives when the decision variable is changed [53]. In this study, the ROC curve is plot of FAR vs MAR when the design parameters change. For our case, this curve will determine the accuracy of the alarm system to detect any anomalies in the system with less possible miss or false detection.

The latency of Alarm system is determined by the detection delay. So, the best parameters are those lead to less detection delay.
Later on in this chapter, a simulated example will illustrate the design procedure and shows all the system parameter for a specified performance level.

3.5.2 *Sensitivity*

During practical process operations, it is needy to adjust alarm design parameters for many reasons. Therefore, it is very important to investigate the degree to which changes in alarm design parameters affect performance indices. Here, the parameter that changes most frequently is the alarm limit (threshold) and the number of delay-timers n. We define sensitivity as the ratio of the infinitesimal change in the function (FAR, MAR or EDD) to the infinitesimal change in the alarm threshold x_{tp}. For a fixed delay-timer n the sensitivity can be expressed as:

$$S_F^{x_{tp}} = \lim_{\Delta x_{tp} \to 0} \frac{\text{Fractional change in the function, } F}{\text{Fractional change in the threshold, } x_{tp}} = \lim_{\Delta x_{tp} \to 0} \frac{\Delta F / F}{\Delta x_{tp} / x_{tp}} = \frac{x_{tp}}{F} \frac{\partial F}{\partial x_{tp}} \tag{3.14}$$

Here F can be any of the FAR, MAR, and EDD expressions derived earlier.

3.6 Proposed Method

In the following section, a case study is represented which shows how the proposed method of alarm system design is applied to a multivariate system.
For example, consider 9 measurement variables of the following system:

- ***Simulated Example***

$$\begin{cases} x_1 = \mu_1 + \epsilon \\ x_2 = 2\mu_1 + \epsilon \\ x_3 = x_2 + \epsilon \\ x_4 = 2x_3 + x_1 + \epsilon \\ x_5 = \mu_2 + \epsilon \\ x_6 = 3\mu_1 + x_4 + \epsilon \\ x_7 = x_1 + x_6 + 2x_3 \\ x_8 = 2x_6 + 3x_5 + \epsilon \\ x_9 = x_8 + x_5 + x_1 + \epsilon \end{cases} \tag{3.15}$$

Where: ϵ is a white noise
$\mu_1 \sim \mathcal{N}(0, 0.1)$; $\mu_2 \sim \mathcal{N}(3, 0.2)$

Next, the training data-set is stored in a data-matrix X. Then, we apply PCA algorithm. $X = [x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9]^T \in \mathbb{R}^{1500 \times 9}$. While designing alarm systems, some specifications should be respected to assure its good performance. PCA method has been used to reduce the dimension of the data, the number of retained principle components is determined using the Kaiser's criterion, two principal components are retained as expected since the independent used variable were 2. The loading matrix is calculated using eq. (2.3) and eq. (2.4). Using the procedure explained in section 2.3.1, the PCA model is built. The initial thresholds for both Hotelling T^2 and SPE are respectively calculated using eq. (2.17) and eq. (2.20):

$$T_\alpha = 9.2510 \tag{3.16}$$

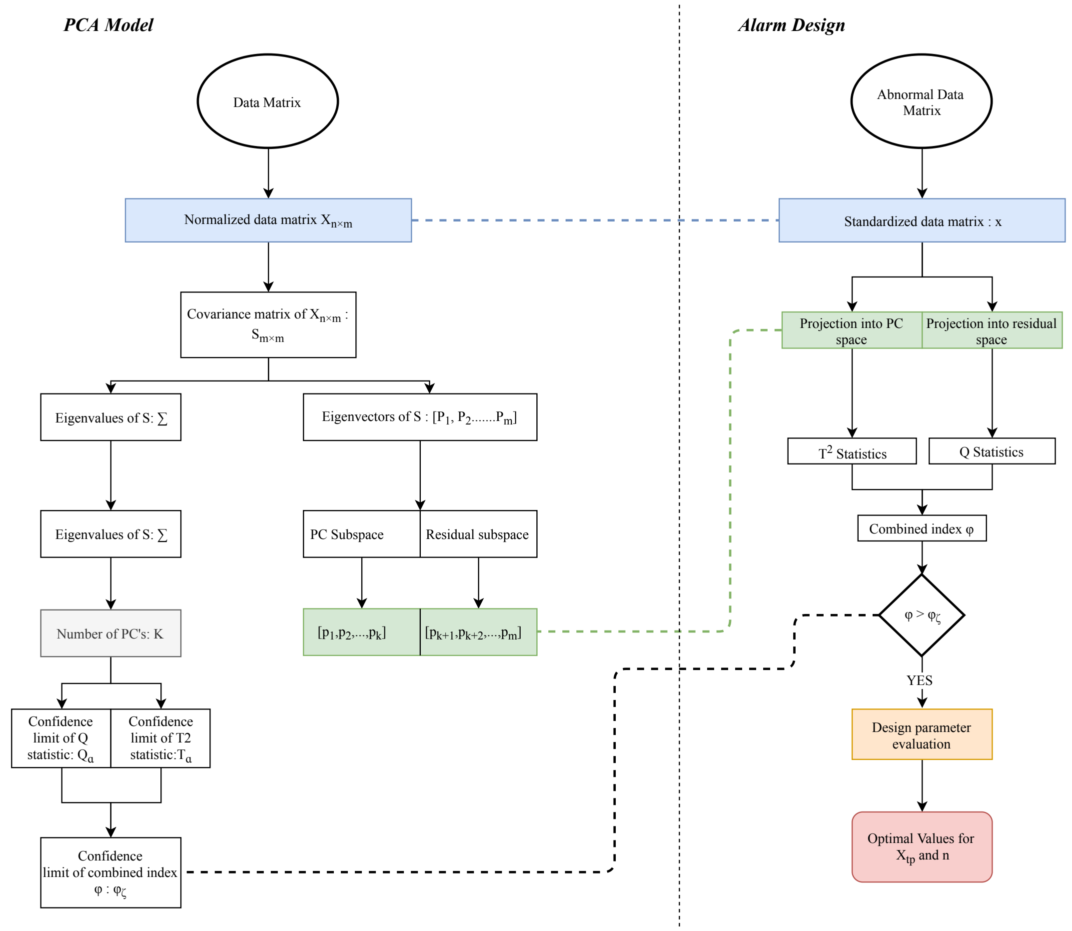


Fig. 3.2 Flow chart for the proposed method

$$Q_\alpha = 0.0643 \tag{3.17}$$

After validating the model with testing data-set, and using eq. (2.21) the combined index φ is used in the alarm system design. Figure 3.4 monitor the process under normal and abnormal situation (For the smallest possible abnormality), the fault occurrence is from the sample 1500 till the end of simulation.

With the aid of maximum likelihood estimation (MLE) method, we can estimate the parameters of the PDFs functions for both fault-free and faulty process data using the kernel function.

Given the PDFs of the normal and abnormal condition of the simulated process in fig. 3.5, the alarm system design will be based on the three performance indices FAR, MAR and EDD as explained before.

The main problem that one can face when applying linear data-driven techniques, such as PCA, to design alarm system or even for process monitoring using a fixed control limits is constantly the high number of false/missed alarms. These false/missed alarms may be due to the presence of random noise in the system, dynamics, nonlinear behaviours,

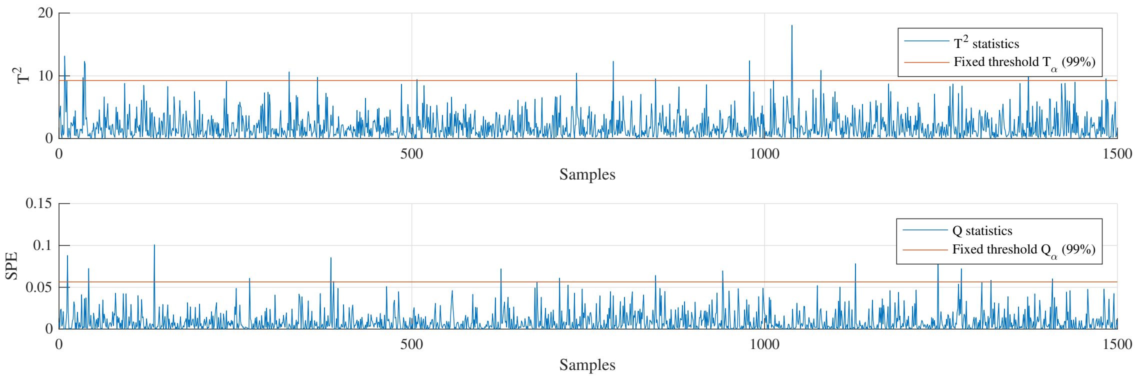


Fig. 3.3 Training data-set.

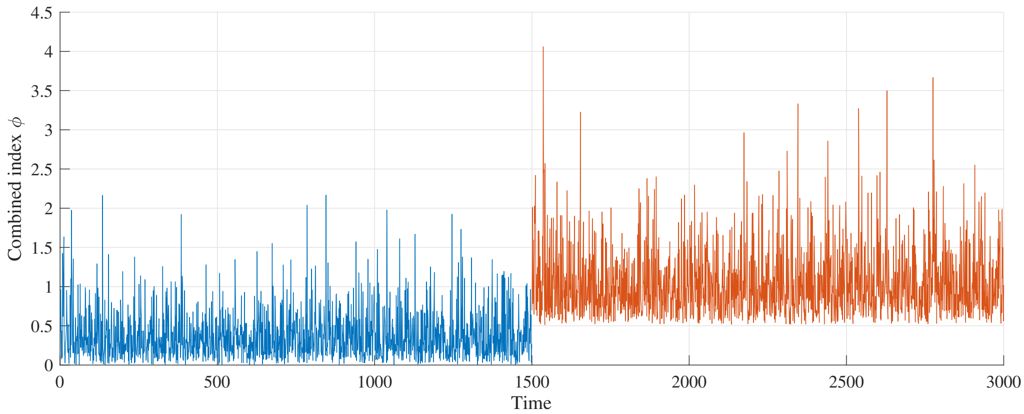


Fig. 3.4 The combined index $\varphi(t)$ of system with a variation from normal to abnormal condition at the time instant 1500.

changing in operation set-point and the mathematical expression by which the control limits are calculated. To address this problem, if a lower threshold is selected, there will be a lot of false alarms. However, only a few alarms will be missed. On the other hand, if a high threshold is selected, the number of false alarms will be significantly reduced, but many alarms will be missed. Therefore, a tight alarm limit will cause many false alarms, but few missed alarms. And a loose alarm limit will only cause a few false alarms, but a lot of missed alarms. This is a very important alarm system design trade-off that will be used as a design parameter.

The design problem is formulated on how to choose the right trip point x_{tp} and number of sample delay n that both satisfy certain requirements on the previous performances (FAR, MAR and EDD). Three cases will be investigated in this thesis

- Case I: Design x_{tp} for a fixed value of n.
- Case II: Design n for a fixed value of x_{tp}.
- Case III: Design both n and x_{tp}.

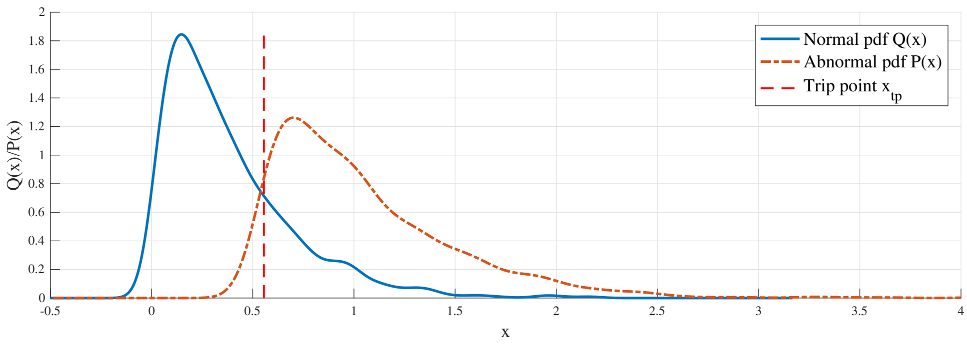


Fig. 3.5 The PDF's of the combined index $\varphi(t)$ in normal and abnormal condition.

3.6.1 Design of x_{tp} for a fixed value of n

For a fixed value of n, the design of x_{tp} is based on two compromises between FAR and MAR / EDD. We take system 3.15 as an example to show the design procedures. the latent variable generated in fig. 3.4 is used as a single signal that detect anomalies in the system. The goal is to design x_{tp} for the alarm on/off delay-timer. If we define for example $n = 3$ to meet the requirements $FAR \leq 5\%, MAR \leq 5\%$, and $EDD \leq 6h$ ($h = 1s$) ; the relation between FAR, MAR, EDD and x_{tp} is obtained using eq. (3.2), (3.11) and (3.13) , and is illustrated in the three plots of fig. 3.6.

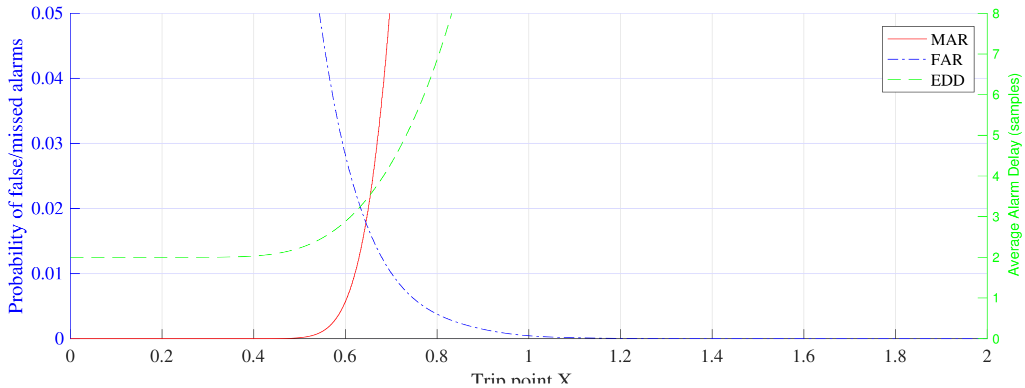


Fig. 3.6 The relation between FAR, MAR/EDD and the trip point x_{tp} for $n = 3$.

Two trade-offs between FAR and MAR/EDD are clearly noticed as x_{tp} varies. We notice that FAR is decreasing when x_{tp} grows larger, but MAR and EDD are increasing, and vice versa. Therefore, FAR, MAR, and EDD requirements will impose their own valid ranges based on these trade-offs. The intersection of these ranges will be the final choice to meet all these requirements; if the intersection is empty ($\varnothing$), there is no way to meet all of them by changing x_{tp} alone.

In this case, based on fig. 3.6 satisfying the requirement on FAR, MAR and EDD require $x_{tp} \geq 0.52$, $x_{tp} \leq 0.71$ and $x_{tp} \leq 0.59$ respectively. The intersection of these ranges gives a valid range for $x_{tp} \in [0.52, 0.59]$.

3.6.2 Design of n for a fixed value of x_{tp}

The number of on/off delay-timer is an important parameter since it affect FAR/MAR. For a fixed threshold x_{tp}, the design of n is based on the trade-off between FAR/MAR and EDD. The objective is to design n for system (3.15) such that $MAR \leq 3\%$, $FAR \leq 3\%$ and $EDD \leq 10h$ for $h = 1s$.

The relationship between FAR / MAR / EDD with the fixed value of x_{tp}, like in the previous example, are obtainable from eq. (3.2), (3.11) and (3.13). The three plots in fig. 3.7 show the relation between these parameters for a fixed threshold $x_{tp} = 0.971$.

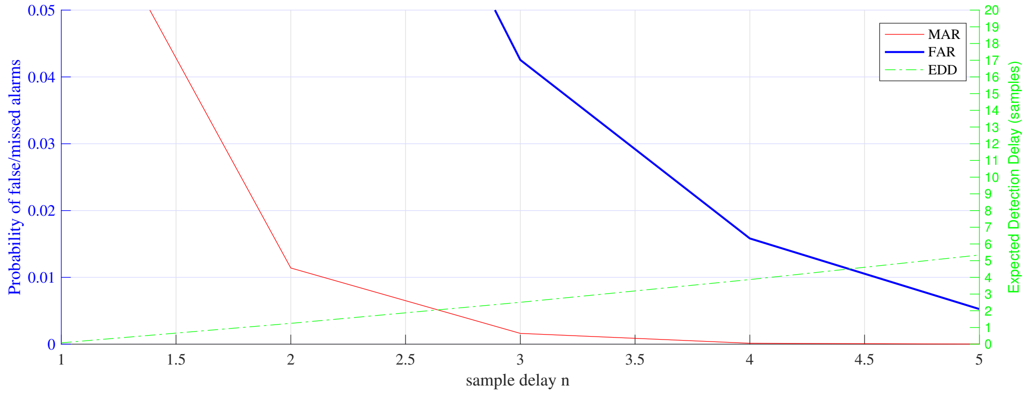


Fig. 3.7 The relation between FAR, MAR/EDD and the number of simple delay n for $x_{tp} = 0.97$.

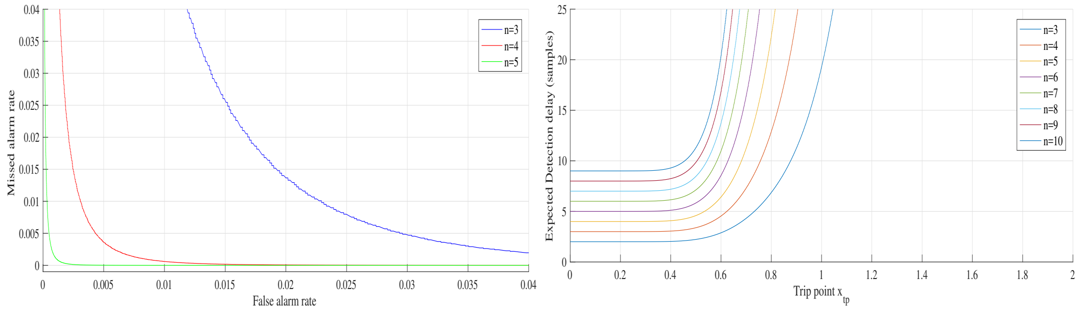
Here two trade-offs exist between the FAR /MAR and EDD. A larger value of n leads to decreases of FAR and MAR but to an increase of EDD. These trade-offs will limit the possible range of n. To meet the specified requirement, it is clear from fig. 3.7 that: from MAR $n \geq 2$, from FAR $n \geq 4$ and from EDD $n \leq 8$. So the valid range for n to meet all the requirements is the intersection of these ranges $n \in [4,8]$.

3.6.3 Design Both n and x_{tp}

The design process is more complicated if both x_{tp} and n are free. It is fundamental that the valid values of n are first determined and then design x_{tp} for each valid value of n. The following example shows the design procedure, where the same latent variable is used.

The design procedure takes the following steps: the plots of MAR versus FAR (ROC) can be obtained by varying xtp for different values of n as shown in fig. 3.8a. We choose from this curve the range of n which can comply with the FAR and MAR requirements. Second, using eq. (3.13) for different values of n, the relation between EDD and x_{tp} is shown in fig. 3.8b. Then, we find the valid range of n in order to meet the EDD condition. From the obtained range of n.

For a criterion of $MAR \leq 4\%$, $FAR \leq 4\%$ and $EDD \leq 8h$ for $h = 1s$. The range of n is found from both graphs as $n \in [3,9]$.



(a) The ROC curves for different number of sample delay n.

(b) The relation between EDD and x_{tp} for different number of sample delay n.

Fig. 3.8 Relation between FAR, MAR/EDD for different values of n

Now, we can look for the corresponding range of x_{tp} by exploring each value of n. Table 3.2 shows the valid ranges of x_{tp} for each value of $n \in [3,9]$.

Table 3.2 The valid ranges of x_{tp} for each $n \in [3,9]$

n	x_{tp}(FAR/MAR)	x_{tp}(EDD)	x_{tp}
3	[0.58,0.67]	[0.50,0.81]	[0.58,0.67]
4	[0.50,0.72]	[0.35,0.70]	[0.50,0.70]
5	[0.45,0.75]	[0.32,0.63]	[0.45,0.63]
6	[0.40,0.81]	[0.30,0.57]	[0.40,0.57]
7	[0.38,0.88]	[0.29,0.52]	[0.38,0.52]
8	[0.36,0.92]	[0.28,0.46]	[0.36,0.46]
9	[0.35,0.97]	[0.20 0.25]	$\varnothing$

However, only one pair of values for x_{tp} and n can be implemented in the alarm system. Thus, it is necessity to have an optimal value for both x_{tp} and n. Depending on the priority between *FAR*, *MAR* or *EDD* for each specific application. In reality, many optimization criteria can be found in the literature or can be formulated to obtain the optimal values of x_{tp} and n. One option is to minimize EDD under certain upper borders of the FAR and MAR. The optimum values of x_{tp} and n must be obtained, despite a change in the optimization criterion according to the same principal.

One solution is to choose a weighted-sum cost function as

$$J(x_{tp},n) = \omega_1 \frac{FAR}{RFAR} + \omega_2 \frac{MAR}{RMAR} + \omega_3 \frac{EDD}{REDD} \tag{3.18}$$

Here, RFAR, RMAR and REDD are the requirements of *FAR*, *MAR* and *EDD* respectively. ω_1, ω_2 and ω_3 are the weights of *FAR*, *MAR* and *EDD* respectively. Hence, the optimal values of x_{tp} and n are the ones minimizing the cost function in (3.18) ,i.e:

$$(x_{tp}, n) = \arg minJ(x_{tp}, n).$$

For this simulated example, and for simplicity let the importance of the three indices be equal to one. i.e ($w_i = 1; i = 1,2,3$). Because of the non-linearity of FAR, MAR and EDD, the DIRECT method which is s a sampling algorithm that will globally converge to the minimal value of the objective is applied with a step of 0.001 to search for the optimal values of x_{tp} and n within their valid ranges [17, 48]. Table 3.3 shows the minimum cost of $J(x_{tp}, n)$ for each value of n and its corresponding threshold. The optimum values that satisfy the design requirements are x_{tp} = 0.5549 and n = 5.

Table 3.3 The optimal value of x_{tp} corresponding to the minimal cost function $J(x_{tp}, n)$ for each value of n.

n	x_{tp}	Min $J(x_{tp}, n)$
3	0.6081	1.1933
4	0.5981	0.8010
5	0.5549	0.7993
6	0.5173	0.8562
7	0.4849	0.9364
8	0.4649	1.0301

After the design process, a simulation could be done using Matlab Simulink (Stateflow) to implement Delay-timers with the obtained results from the Design. The output of this simulation is the Alarm signal. This will be done in the Industrial case study later on.

3.7 Conclusion

In this chapter, three performance specifications of alarm delay-timers, namely, the FAR, MAR and EDD, are calculated for conventional delay-timers. Some preliminary sensitivity analysis are also discussed. A simulated example is presented to show the steps of the proposed method and design procedure for designing multivariate alarm system. As a future extension, further study on the sensitivity advantage of conventional delay-timers is required by a systematic approach.

Chapter 4

Industrial Application

In this chapter, a real application is done to illustrate the performance assessment of the alarm system. The Data was taken from the rotary kiln process that construct the PCA model. More details about the process is availlable in Appendix B.

4.1 Process Description

Ain El Kbira Cement Company, or SCAEK, was founded in 1998 as a cement factory. It went public in mid-2016. The company produces ordinary and special cement mixtures. SCAKs headquarters and factory are located in the region of Setif, northern Algeria. Its annual production capacity is estimated at 3,200,000 tons of cement. Cement production is composed of five areas; beginning with the quarry where the limestone rocks are dug out and crushed, comes after the raw meal area where limestone, clay and iron ore mix are grinded to some fineness. Then the material is fed to the cook area where material undertakes many physical and chemical operations, such as drying, dehydration, decomposition of carbonates. Some reactions are solid others involves the liquid phase (sintering) and third are achieved while cooling. Almost the physical transformations happen in the preheat tower whereas the main chemical reactions happen in the rotary kiln under high temperature gradient. The product of this area is called clinker. Clinker will be mixed to other additives and undertakes finish grinding in the subsequent area to produce the cement. The fifth area is dedicated to packing and expedition. Ain El Kebira cement plant in the Algerian east, where the work is conducted, is a dry process consisting of four cyclone levels in the suspension preheater and short rotary kiln of 5.4 (m) shell diameter (without brick and coating) and 80 (m) length, with 3 incline. The kiln is spun up to 2.14 (rpm) using two 560 (kws) asynchronous motors and producing clinker of density varying from 1300 to 1450 (kg/m3) under normal conditions. Two natural gas burners are used, the main one in the discharge end and the secondary in the first level of preheater tower [9].

Table 4.1 Data sets used for this application

Data sets	Size	Sampling time(s)	Description
Training set	$X \in R^{768\times 44}$	20	Normal operation data, used to construct PCA model
Testing set	$X \in R^{11000\times 44}$	1	Normal operation data, used to validate the PCA model
Process fault	$X \in R^{2084\times 44}$	1	Normal/faulty operation data, Real process fault
Sensor faults (6 sets)	$X \in R^{1500\times 44}$	1	Sensor fault simulations

4.2 Design Procedure

While designing an alarm , some specifications should be respected to assure the good performance of the alarm system. As done in section 3.6, the design's method is as follow: PCA method has been used to reduce the dimension of the data, the number of retained principle components is determined using the Kaiser criterion due to its simplicity and effectiveness. In designing the alarm system, we have used the combined as an index to detect the deviation of the process from normal to abnormal condition. This was conducted after building the PCA model first and validated it using the training and testing data-sets respectively. From the **real fault**, the process data consists of healthy and faulty part; no information are available about the real occurrence time of the fault. We have taken a window of 300 samples length and move it to find that the range [251,550] samples is the appropriate choice to fit the PDF functions. since other choices yields either in no overlapping at all or gives the same PDF. We chose the intersection of the PDFs to be as minimum as possible to describe the real minimum deviation possible and hence yield to the best performance with higher deviation. The same procedure will be applied later to both T^2 and Q statistics signals generated from the same process for comparison reason.

The relevant signals that describe the kiln system are precede history data. These data were collected from the historian of the plant. The 44 variables was taken from an open loop configuration to avoid the interaction between sensors; the application has no need for pre-processing and filtering the data since PCA is a noise decoupling method.

Table 4.1 lists the different data-sets used in order to construct and validate the PCA model, then evaluate and compare the performance of the designed Alarm system.

Table 4.2 Simulated sensor faults introduced at (500 -1000 s)

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 = -4×10^{-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
SFault(8)	12	+5.5%, −4.5%	Additive fault: Intermittent fault, changing intervals and amplitudes

4.3 Design results using the combined index

The objective is to asses the parameters of the alarm system and redesign them if necessary to feet the following requirements: $MAR \leq 4\%$,$FAR \leq 4\%$ and $EDD \leq 5s$. Start with building the PCA model, validate it and construct the ϕ index in the normal and abnormal situation. The basic trip point of the system is the control limit $\phi_\alpha = 1.816$.

1. First, we build the PCA model for the system using the healthy data-set. The obtained model is used to construct the Hotelling'S T^2 and Q statistics as shown in fig. 4.1. The initial thresholds are the ones obtained from eq. (2.17) and eq. (2.20).

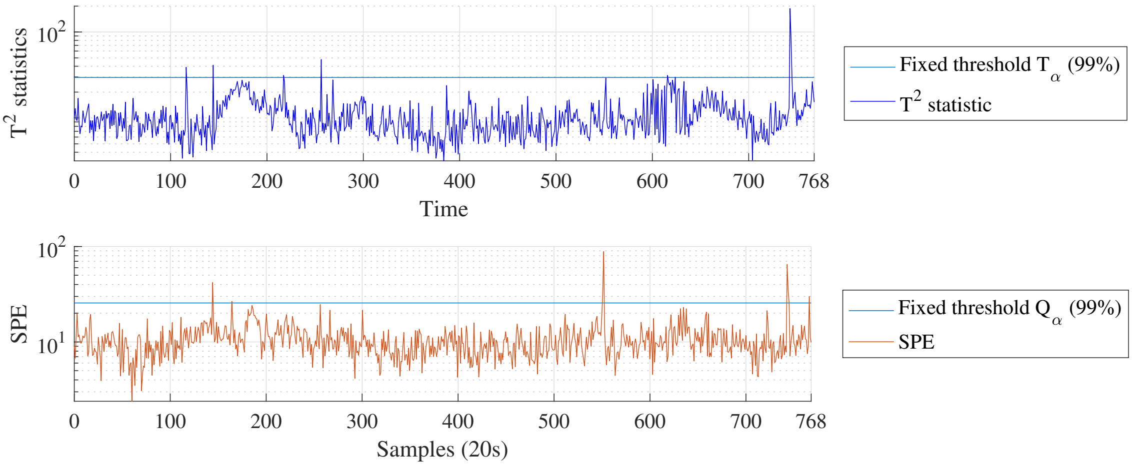


Fig. 4.1 Training data-set Q and T^2 for Rotary kiln system.

After that, and using eq. (2.21) we build the combined index $\varphi(t)$ as shown in fig. 4.2.

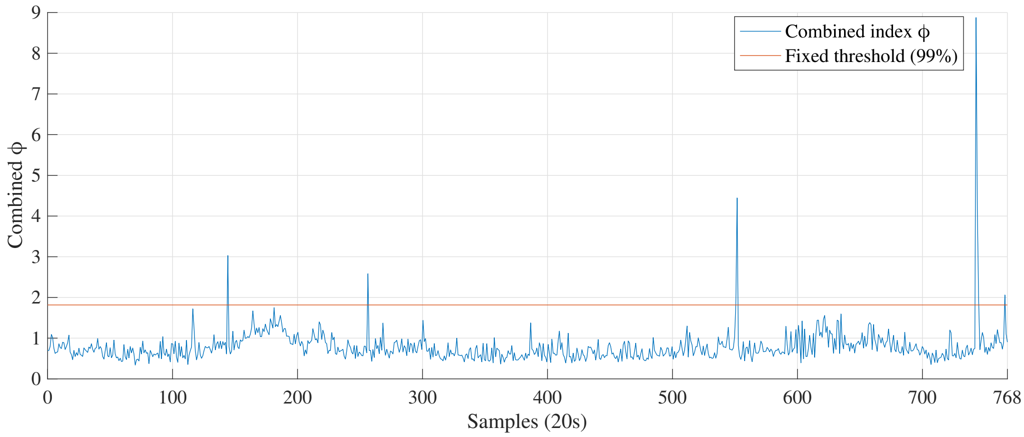


Fig. 4.2 The Combined index $\varphi(t)$ for the training data-set.

Finally, we validate the system using the testing data-set. As we notice from fig. 4.3, there exists many false alarms due to some uncertainties that could happened during the acquisition. That is why we will try to reduce those false alarms using Delay-timers.

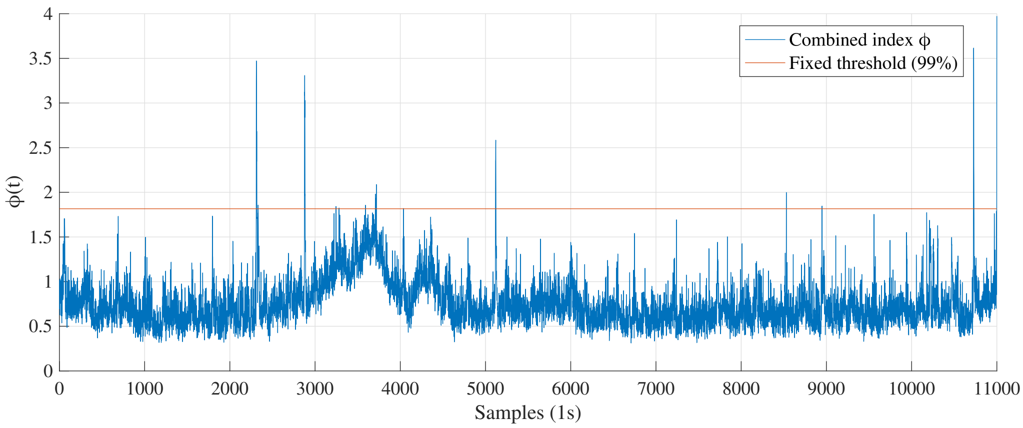


Fig. 4.3 The Combined index $\varphi(t)$ for the testing data-set.

2. The next step is to construct the combined ϕ index in the normal and abnormal state using a real fault occurs in the system. The nature and the real occurrence time of the fault is undetermined. The basic trip point of the system is the control limit $\phi_\alpha = 1.816$.

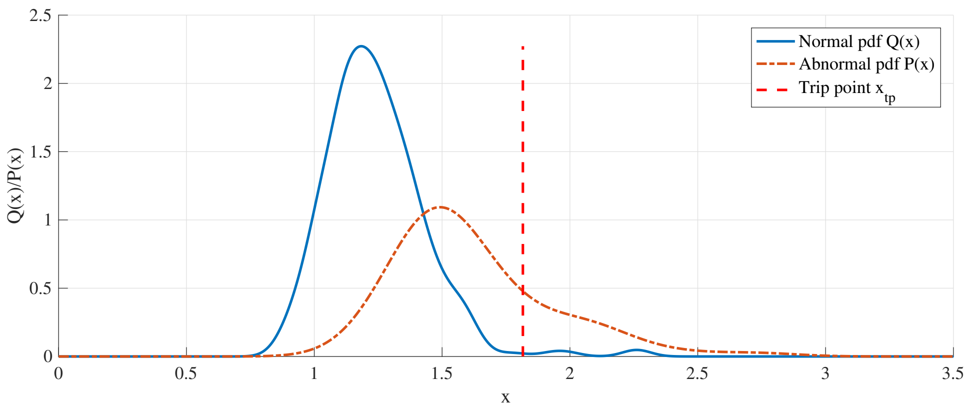


Fig. 4.4 Approximated PDFs for healthy and faulty data of the combined index ϕ using kernel function.

Using the estimated PDFs in Fig.4.4, the performance indices FAR, MAR and EDD are calculated from equations (3.2), (3.11) and (3.13) , in the absence of delay-timer ; $FAR = 1.4\%$, $MAR = 45\%$ and $EDD = 1.22s$. The design requirements are not satisfied.

3. Considering the design procedure discussed in section 3.6, the first step is trying to design x_{tp} in the base case $n = 1$, according to the relations between FAR/MAR, EDD and x_{tp} (fig. 4.5), based on the specifications $MAR \leq 4\%$,$FAR \leq 4\%$ a valid range for x_{tp} will be imposed. $x_{tp} \in [0.50, 1.19]$ for FAR and $x_{tp} \in [1.59, 2.50]$ no intersection means that it is impossible to satisfy all requirements with $n = 1$ for every possible value of x_{tp}.

Design for fixed threshold x_{tp}

In this case, since the choose of a fixed delay-timer $n = 1$ does not yields to the required results. On solution is either increase the delay number n or simply design n for a fixed

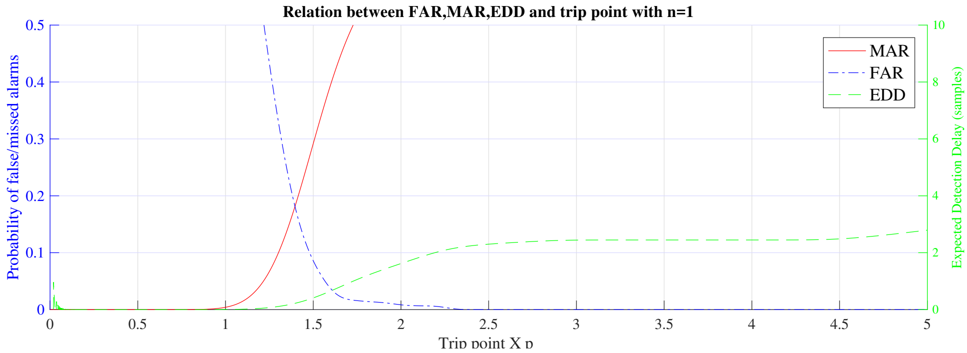


Fig. 4.5 The relation between FAR/MAR/EDD and the trip point x_{tp} for $n = 1$.

value of x_{tp}. Choosing $x_{tp} = 1.65$, the relation between FAR, MAR and EDD is represented in fig. 4.6.

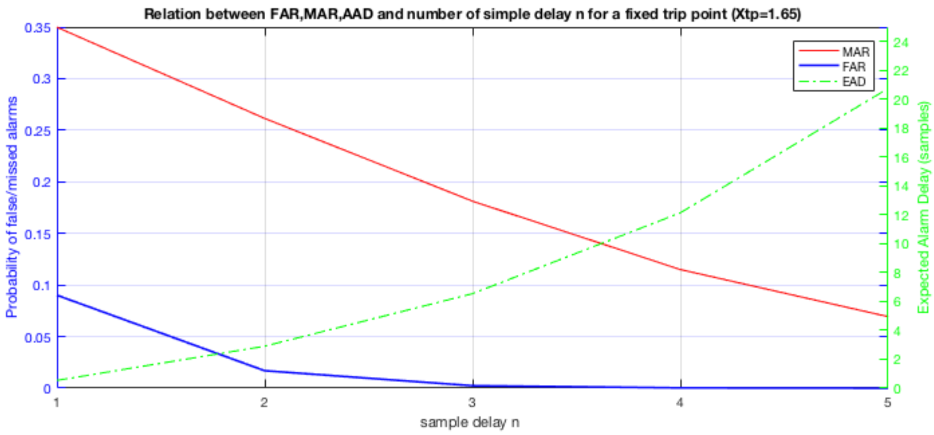


Fig. 4.6 The relation between FAR/MAR/EDD and the delay number n for $x_{tp} = 1.65$.

From the relation between FAR/MAR and EDD for a fixed threshold, we can see that the valid ranges for n that gives best results according to the previous requirement are $n \geq 2$, $n \geq 7$ and $n \leq 3$ for FAR, MAR and EDD respectively. Finding a valid range that satisfies all the reqiurments will be impossible with this value for x_{tp}.

Design for both x_{tp} and n

Another possible way is to design the number of on/off delay-timer n and the threshold x_{tp} simultaneously, following the procedure done in section 3.6 the valid ranges of n are decided by the three requirements in *FAR*, *MAR* and *EDD* from the relations between *FAR*/*MAR*/*EDD*. Hence, from fig. 4.7 it is clear that $n \leq 6$ and from the ROC curve in fig. 4.7a $n > 2$ i.e ($2 < n \leq 6$). The valid ranges for x_{tp} from both *FAR*/*MAR* and *EDD* are given in Table 4.3.

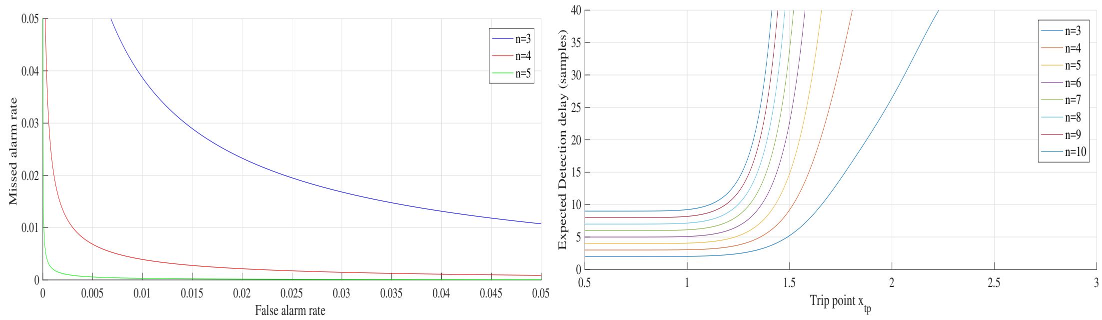
In practice, only one pair of x_{tp} and n will be implemented in the Alarm system. In order to find an optimal value for x_{tp} and n, the DIRECT method is applied to the performance measure given by eq. (3.18) with the same weights $w_1 = w_2 = w_3 = 1$ to bring

Table 4.3 The valid ranges of x_{tp} for each $n \in [3,6]$ and the minimum of the cost function

n	x_{tp}(FAR/MAR)	x_{tp}(EDD)	Range of x_{tp}	Min $J(x_{tp}, n)$	x_{tp}
3	[1.37,1.43]	[0.50,1.49]	[1.37,1.43]	1.3963	1.4353
4	[1.33,1.48]	[0.50,1.36]	[1.33,1.36]	1.3593	1.3628
5	[1.31,1.53]	[0.50,1.24]	$\varnothing$	/	/
6	[1.29,1.55]	[0.50,0.98]	$\varnothing$	/	/

equal significance to each performance index. The results are shown in Table 4.3. The application of DIRECT method was the same as in the simulated example (section 3.6). The optimum values are:

$x_{tp} = 1.3628$ for $n = 4$ yields to $FAR = 1.7\%$, $MAR = 0.24\%$ and $EDD = 5s$, which satisfy the specified requirements.



(a) The ROC curves for different number of sample delay n for the cement rotary kiln.

(b) The relation between EDD and x_{tp} for different sample delay n for the cement rotary kiln.

Fig. 4.7 Relation between FAR, MAR/EDD for different values of n

4.4 Design Results using Q and T^2 indices

In this section, we redesign the previous alarm system but using the Q and T^2 indices instead of the combined $\varphi(t)$, using the same requirements stated before, $MAR \leq 4\%$,$FAR \leq 4\%$ and $EDD \leq 5s$.

The figures in the next page represents the ROC and EDD curves for deferent values of n and X_{tp} for both Q and T^2.

The results of designing both n and x_{tp} for Q and T^2 are obtained from fig. 4.11a, fig. 4.11b, fig. 4.12a, and fig. 4.12b and summarized in Table 4.4. As we can see, there is no valid range to fulfill these requirement for both Q and T^2, the intersection set is empty ($\varnothing$). We could find a valid range for those two indices if the conditions of the design were less strict.

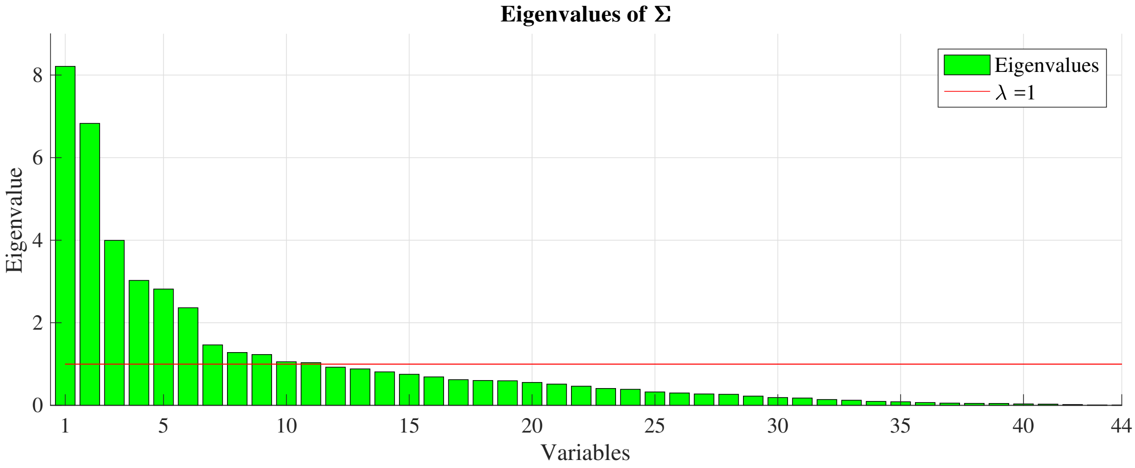


Fig. 4.8 The eigenvalues of the covariance matrix Σ,11 principle components are retained using Kaiser's rule.

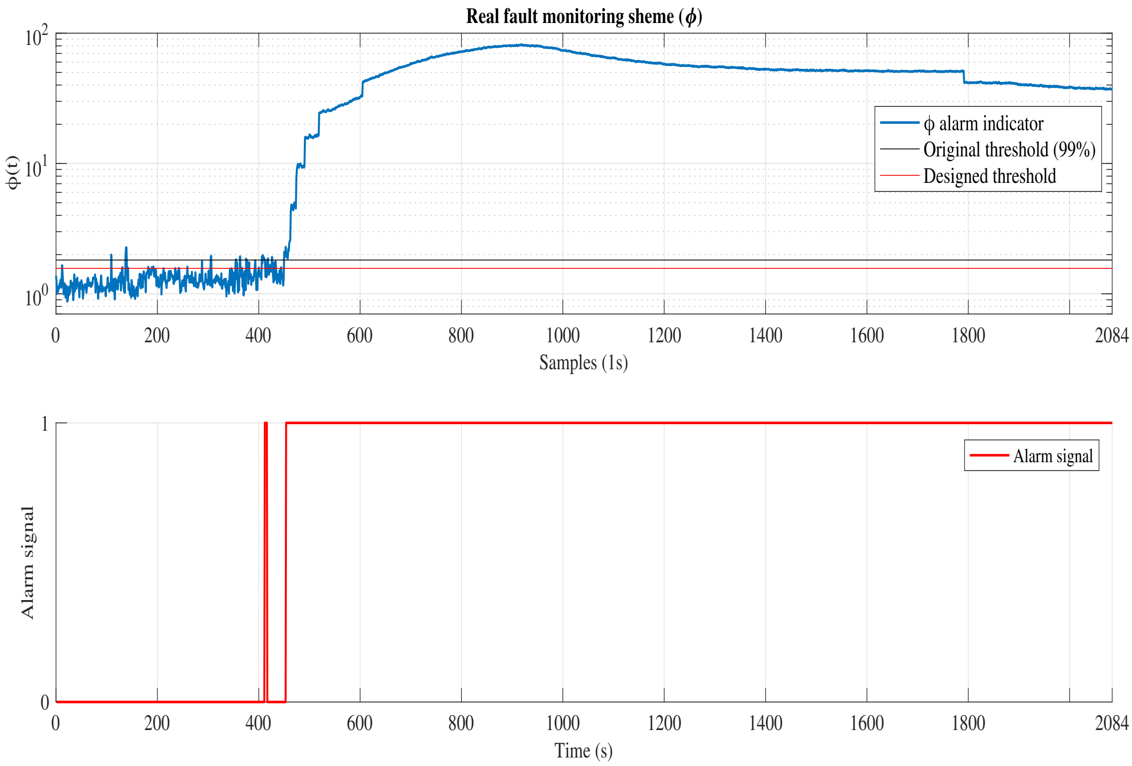
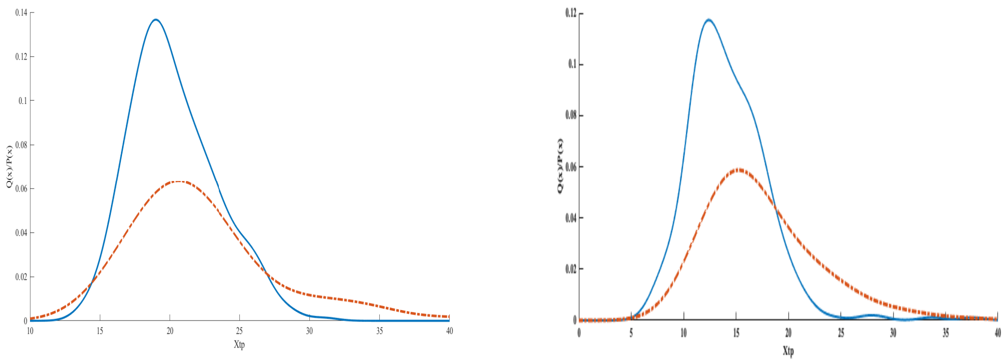


Fig. 4.9 The combined index $\varphi(t)$ monitoring results of a real process fault in the cement rotary kiln.

Table 4.4 Valid ranges of x_{tp} for each value of n (For Q and T^2).

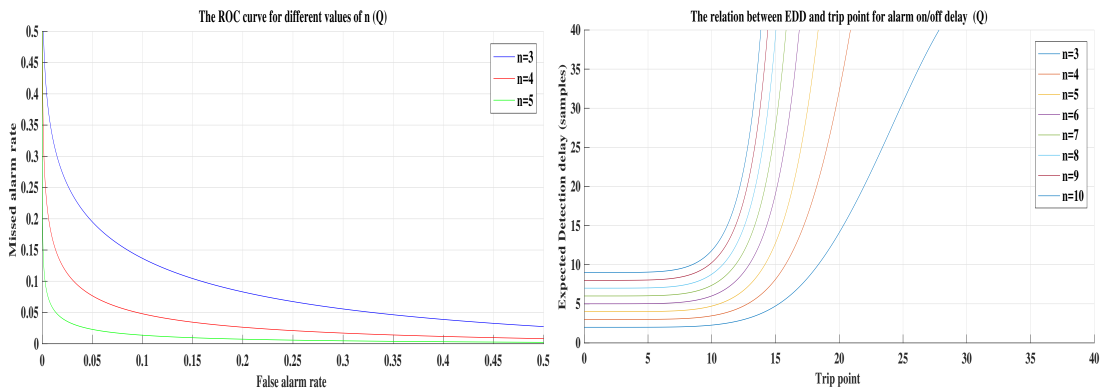
n	x_{tp} From FAR/MAR		x_{tp} From EDD		Valid range for x_{tp}	
	Q	T^2	Q	T^2	Q	T^2
3	/	/	[0, 15.31]	[0, 20.26]	∅	∅
4	/	/	[0, 12.25]	[0, 17.93]	∅	∅
5	[15.64,16.03]	/	[0,11.27]	[0, 15.71]	∅	∅
6	[15.31,16.48]	[21.18,21.47]	[0,7.85]	[0, 11.33]	∅	∅



(a) Approximated PDFs of T^2 for healthy and faulty data.

(b) Approximated PDFs of Q statistics for healthy and faulty data.

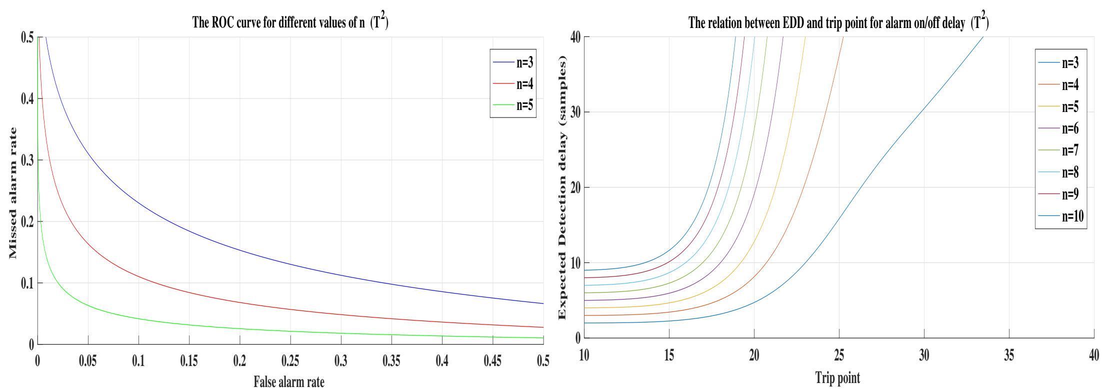
Fig. 4.10 Approximated PDFs (Q statistics and Hotelling's T^2).



(a) The ROC curves for different number of sample delay n (Q).

(b) The relation between EDD and x_{tp} for different sample delay n (Q).

Fig. 4.11 Relation between FAR, MAR/EDD for different values of n (Q statistics)



(a) The ROC curves for different number of sample delay n (T^2).

(b) The relation between EDD and x_{tp} for different sample delay n (T^2).

Fig. 4.12 Relation between FAR, MAR/EDD for different values of n (T^2 statistics)

4.5 Simulation

In this section, the Stateflow toolbox of the MATLAB Simulink has been used To implement the *Delay-Timer*. Using the obtained results from section 4.3, Figure 4.13 illustrates the implemented 4-sample delay-timer. In this figure phi_test1 is the generated signal, and phi_a is the threshold (the designed trip point x_{tp}). NA1, NA2, NA3 and NA4 are “No Alarm” states and A1, A2, A3 and A4 are “Alarm” states.

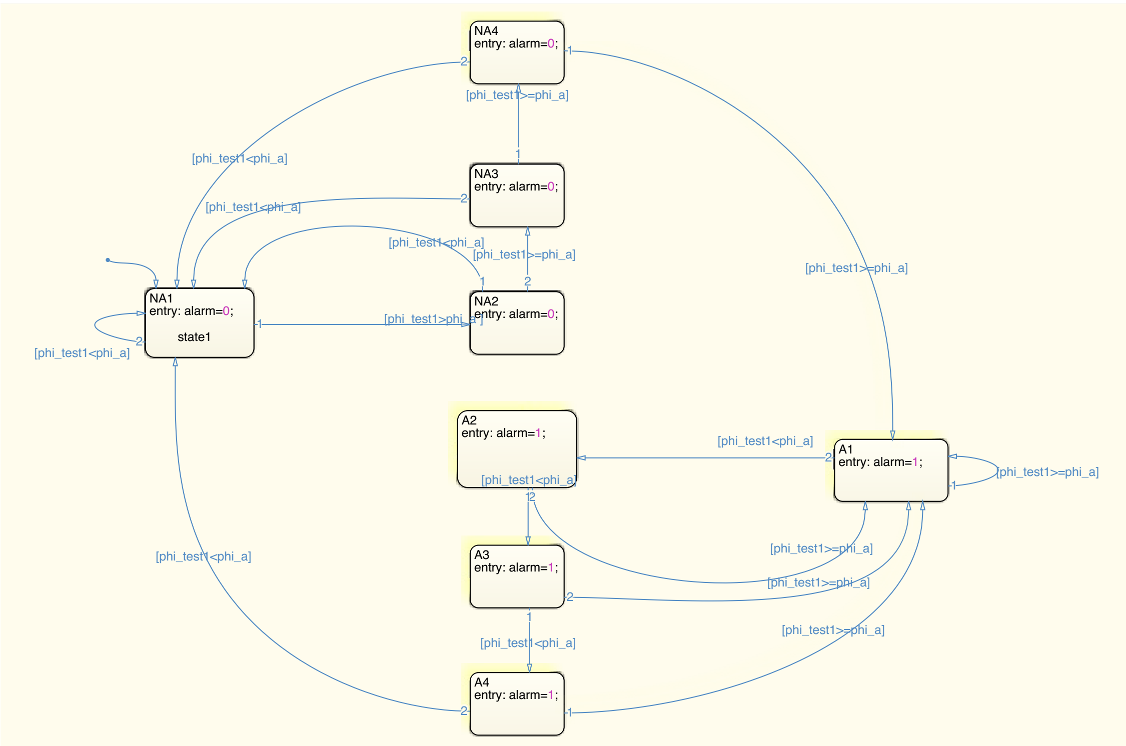


Fig. 4.13 (4-Sample) Delay-Timer implemented by Stateflow toolbox in MATLAB Simulink

The simulation was done on the real fault and the set of simulated faults stated in table 4.2 and plotted in. fig. 4.15. The results shown in fig. 4.16 demonstrates the alarm annunciation of simulated faults with the designed threshold achieving by mentioned implemented delay timer in State flow (fig. 4.13).

For the real faults shown in fig. 4.9, since the exact time of occurrence was unknown, a perturbation is detected from sample 430. Since more than four samples stay above the threshold an alarm has raised. An expected delay of maximum 5s still exists. The alarm signal will still activated starting from instant (453s). The use of 4-delay simple with the right threshold allow to enhance the performance of alarm system. If we compare this application with the one used in [9] where A. Bakdi et al. have used EWMA-based adaptive threshold monitoring scheme for the same system (Rotary kiln) and with the same data-sets, we found that these performances are good enough since the use of delay-timer with a fixed threshold gives almost same results. However, a false alarm has rised (430s) but still, because the design requirements set before allow a maximum of 4% FAR.

For other faults, this technique gives best results in term of FAR/MAR. As can be seen in fig. 4.15 and fig. 4.16, when the designed threshold is considered, with the corresponding simple delay, the requirements are satisfied. We can notice in most faults, a delay of maximum 4 simple is always present, specially in alarm clearance. It is evident that this type of design has better performance related to the both FAR/MAR and EDD and it causes fewer nuisance alarms.

4.6 Sensitivity

As seen in Chapter3, the sensitivity is very important in the design of any system. Therefore, it is necessary to investigate the degree to which changes in alarm design parameters affect performance indices. In this case, these parameters are mainly the threshold for a fixed alarm delay n. Equation (3.14) define the sensitivity as the ratio of the infinitesimal change in the function (FAR, MAR or EDD) to the infinitesimal change in the alarm limit (x_{tp}) for a fixed delay-timer n.

$$S_F^{x_{tp}} = \lim_{\Delta x_{tp} \to 0} \frac{\Delta F / F}{\Delta x_{tp} / x_{tp}} = \frac{x_{tp}}{F} \frac{\partial F}{\partial x_{tp}}$$

Here, F can be any of the FAR, MAR, and EDD quantities derived in earlier sections. As a specific example, eq. (3.2) for the false alarm rate of $n = 4$ on/off delay is considered as it is found by the design results. Replacing $p_2 = 1 - p_1$ and denoting $p_1 = p$, it can be shown that the sensitivity for n=4 is:

$$S_{FAR}^{x_{tp}} = x_{tp} \frac{4\left(p^7 - 3p^6 + 3p^5 - p^4 - 2p^3 + 3p^2 - 3p + 1\right)}{p\,(p-2)\,(p^2 - 2p + 2)\,(2p^4 - 4p^3 + 6p^2 - 4p - 1)} P(x_{tp}) \tag{4.1}$$

$P(x_{tp})$ represents the value of the likelihood function in fig. 4.4.

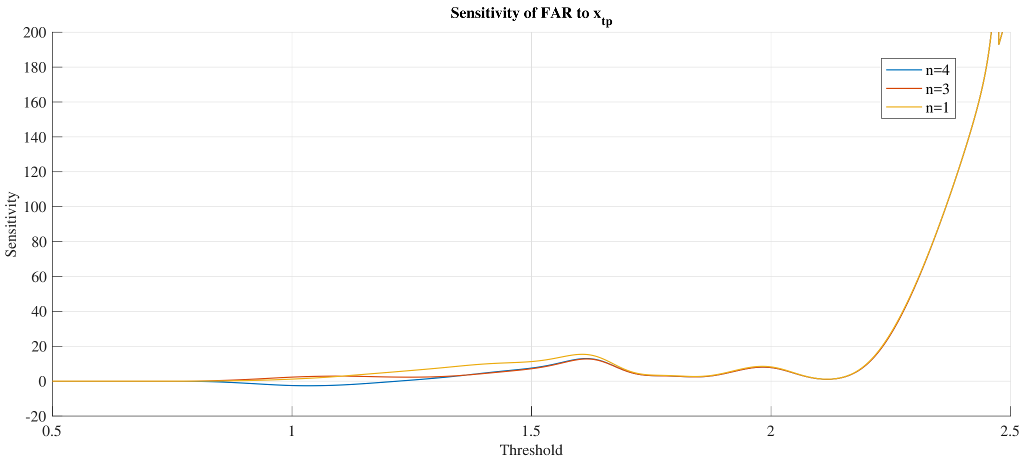
Consider the sensitivity curves in fig. 4.14 which are obtained for the set of data of the rotary kiln system designed before, with the threshold changing from 0.5 to 2.5 . It can be seen that, a lesser value of n is more sensitive to the threshold change comparative to a lower one. This indicates the conventional delay-timer is more sensitive to changes in threshold in general when n gets larger. Also, we can notice small alternating for the sensitivity, but in the region of interest the sensitivity gets larger when threshold increases, and for a fixed threshold for example $x_{tp} = 1.3$, the sensitivity is larger for large value of n.

Sensitivity analysis is very important as small changes in the distributions of normal and abnormal data (latent variable), which may be regarded equivalently as small changes in the threshold, can cause significant deviations in FAR and MAR; such sensitivity is not desirable in practice.

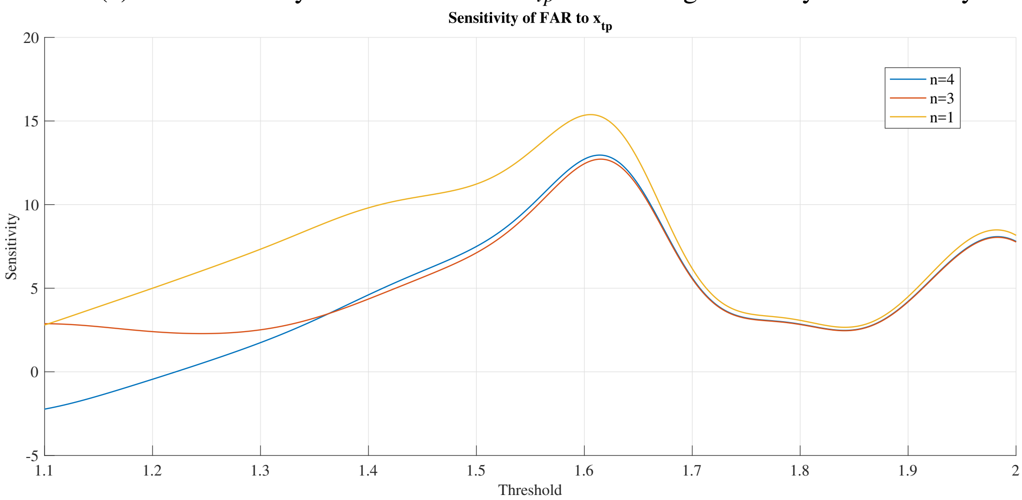
4.7 Results

Some results are highlights in the following notes:

- The use of the combined index φ guarantee good results for a strict design ($MAR \leq 4\%$,$FAR \leq 4\%$ and $EDD \leq 5s$), compared to Q and T^2.
- The use of Delay-Timers made the system more accurate since it reduce (FAR/MAR), and less sensitive to change in system parameters.
- The simulation performed using Stateflow toolbox in MATLAB Simulink, confirms the design results. Where the FAR/MAR does not exceeds its maximum limit.
- A maximum detection delay of 4s was present in most detected faults.
- The sensitivity of FAR to the threshold gets larger when x_{tp} increases.
- For a fixed threshold, in general the performance indices are less sensitive to alarm design parameter (x_{tp}) when n gets larger.



(a) The sensitivity curves of FAR to x_{tp} in the designed rotary kiln alarm system.



(b) Zoom view when threshold is in the interest region ($1.1 \leq x_{tp}) \leq 2$).

Fig. 4.14 Sensitivity curves for different values of n

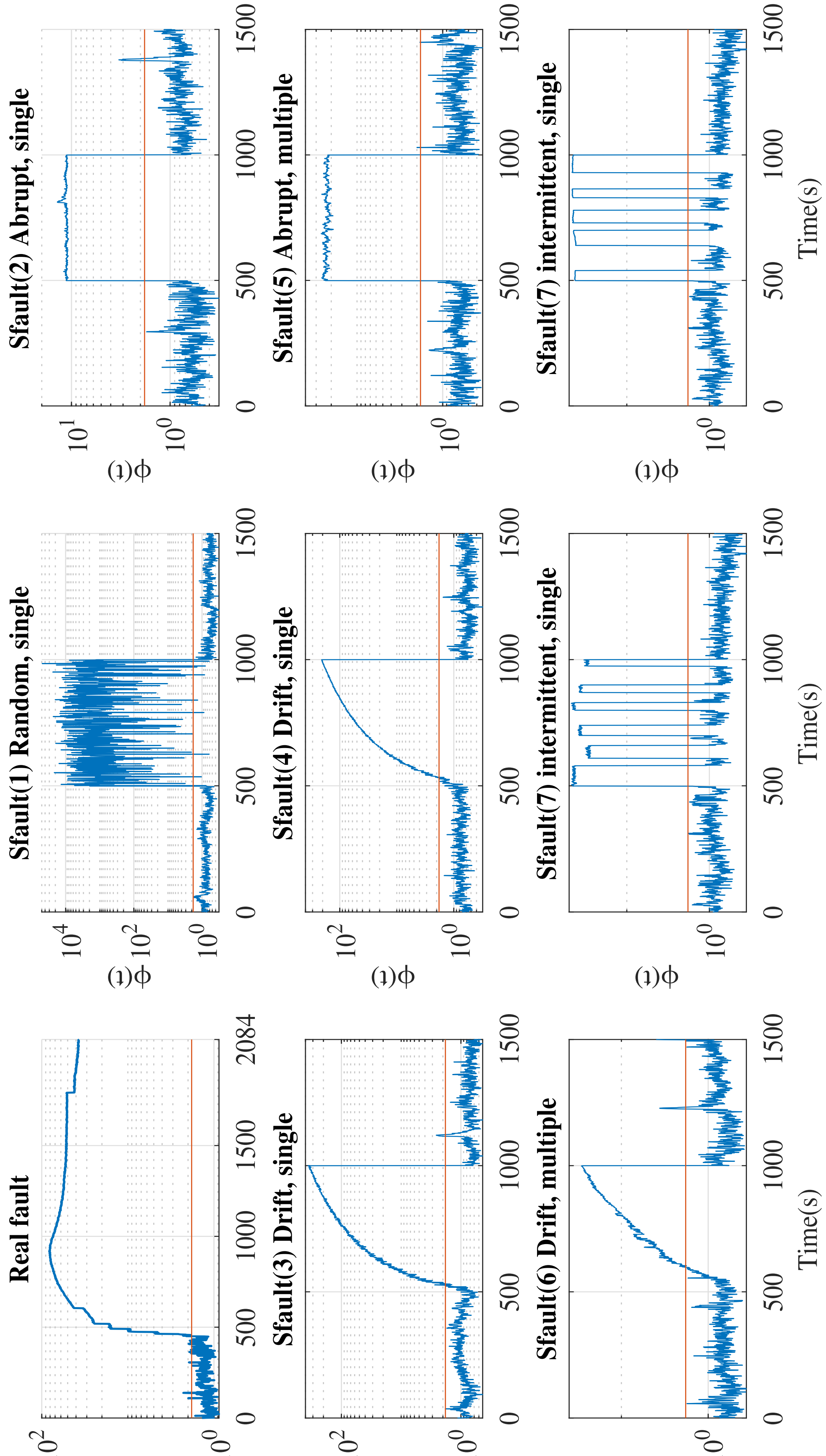


Fig. 4.15 Monitoring different faults using the combined index ϕ and a fixed threshold

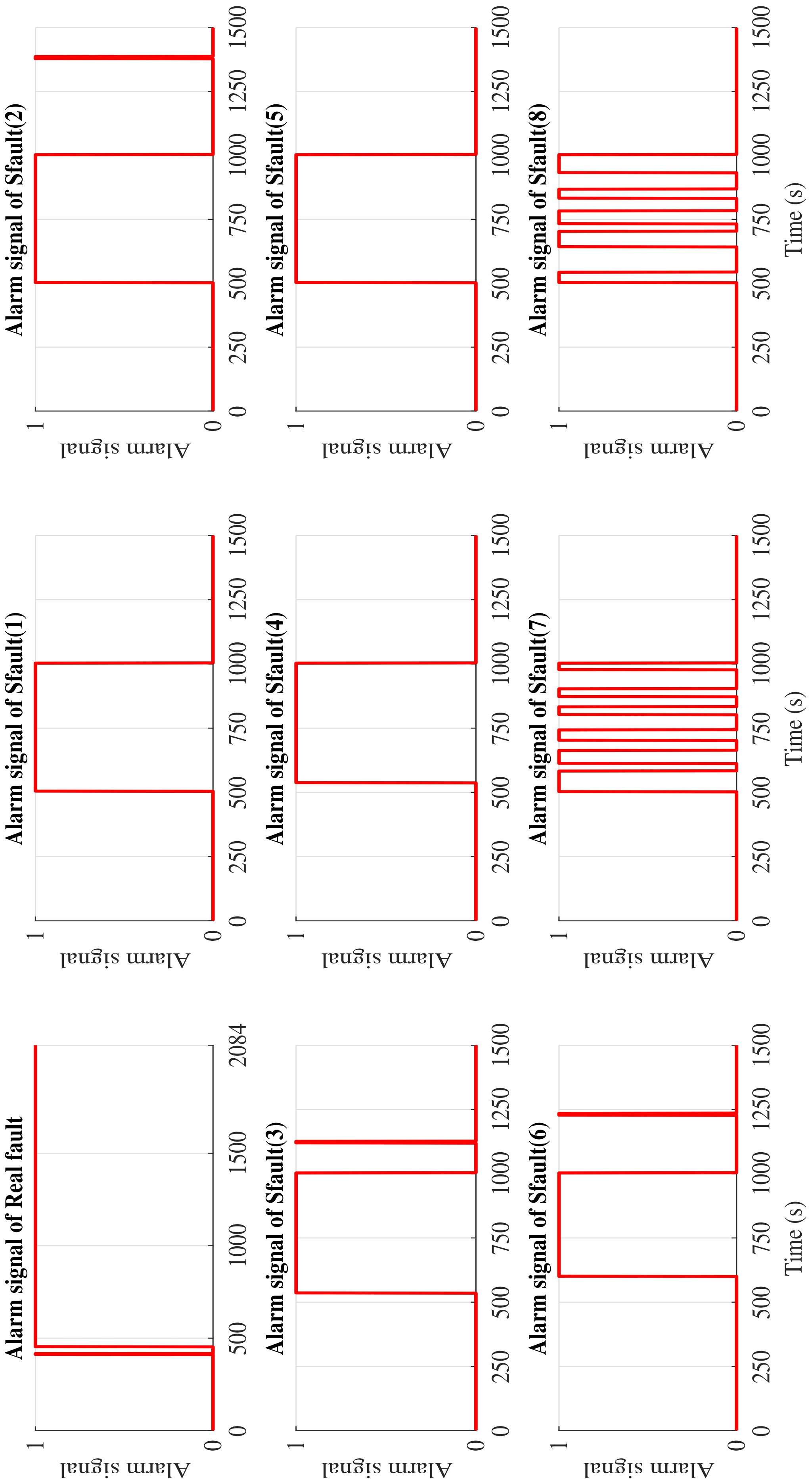


Fig. 4.16 Alarm signal of different faults using the Design results.

4.8 Conclusion

In each alarm system or condition monitoring system, decision-making plays an important and vital role. Thresholding is an inseparable part of decision making, and if thresholding had chosen wisely by designers, the system would have the best performance. It is evident that an excellent thresholding system must separate normal and abnormal data from sensor measurements completely. In this chapter, some methods such as simple thresholding with delay-timers for multivariate system was proposed. It also, use the combined index as a single signal and use it in the design of Alarm system. The numerical results of performance assessment show that the proposed method give good results comparable to other monitoring sheme methods. Also, the alarm annunciations showed that the proposed methodology has a less chattering outcome. The sensitivity of the system was also discussed in the last section. Finally, the efficiency of the mentioned methodology in the industrial area has been proven by an industrial case study.

Conclusion

In this dissertation, we researched efficiency evaluation and systematic design for multivariate alarm systems, based on three performance assessment indices, namely FAR, MAR, and EDD. This design is based on the on/off delay-timer using the Markov theory.

This thesis discussed quantitative relationship between the commonly used univariate alarm design methods and associated performance measures then adapt them with multivariate alarm design. In the literature model-based fault detection is covered extensively. Although the signal-based method is most widely used and easy to implement compared to model-based methods, it did not receive much attention in academia. This thesis uses the combination between process History-based and signal-based to develop and design a good alarm system. ISA 18.2 and EEMUA 191 standards are recently published as a guideline for alarm management and design; but industries are still far from what is recommended by these standards. In order to act in accordance with ISA 18.2 and EEMUA 191, there is no alternate other than following a systematic alarm system design and rationalization method. Delay-timers and deadbands are two very common alarm design techniques used in the industry. These methods can be implemented easily in most modern DCS systems and are effective in reducing false and nuisance alarms. However, there is a common misconception that 8-second or any other configuration of delay-timer delays the alarm activation by eight seconds or so, which is hardly the case in reality. In Chapter 3, the detection delay is calculated for fixed limit checking with delay-timers. The alarm systems with delay-timers are modeled using Markov processes; and historical process data is used to estimate the invariant probability vector. Also an optimal alarm design method is discussed that allows compromise between the false alarm rate, the missed alarm rate and the detection delay.

The combined index φ was implemented and used to model the alarm scheme as a single signal, describing anomalies within the multivariate system.

A simulated example showed the design procedure using PCA and the combined index, the FAR, MAR, and EDD computations have been validated on Chapter 3, three cases have been examined on the basis of the three performance indices for systemic design of the trip point and / or the number of sample delay n.

For a discussion of advantages and disadvantages of the proposed method with the conventional delay-timers, their performance is compared in terms of three criteria, namely, accuracy (ROC curve with false and missed alarm rates), latency (detection delay with respect to change in settings) and sensitivity (change in performance with any change in design settings).

An industrial case study was provided to illustrate the design procedure and to prove the efficiency of this method with real process data. A comparison between the use of the combined index and the statistical indices (Q and T2) confirms its effectiveness.

A prospective research is to use the adaptive limit in the design of the alarm system instead of the fixed one. This can improve efficiency and deliver the highest results in terms of alarm annunciation and alarm clearance.

Appendix A

Basic of Statistical measures & Linear Algebra

This section is dedicated to clarify the algebraic terms and theorems used within the thesis. However, a little prior exposure to linear algebra is still required from the reader.

Statistics Basics

Mean

The expected value (also called the mean) of a random variable is the sum of the product of each possible value and its corresponding probability. In mathematical terms, if the random variable is X, the possible values are x_1, x_2, . . . x_n, and the corresponding probabilities are P(x_1),P(x_2),...,P(x_n) then

$$E(X) = \sum_{i=1}^{n}(x_i P(x_i)) \tag{A.1}$$

In a data set that is identically distributed, the mean is simply:

$$\mu = E(X) = \frac{\sum_{i=1}^{n} x_i}{n} \tag{A.2}$$

Variance

The variance σ^2 of a random variable, which takes on discrete values x and has mean μ, is given by the equation

$$\sigma^2 = \sum(x-\mu)^2 P(x) = \sum(x^2 P(x)) - \mu^2 \tag{A.3}$$

In a data-set, variance is also a measure of the spread of the data in a data set with mean $\bar{X}$

$$\sigma^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1} \tag{A.4}$$

Variance is claimed to be the original statistical measure of spread of data.

Standard Deviation: The standard deviation σ is simply the square root of the variance.

Covariance

The Covariance ids a measure of how much each of the dimensions varies from the mean with respect to each other. Alternatively, it is a measure to see if there is a relationship between the 2 dimensions (or more).

We can define the covariance between one dimension and itself as the Variance. i.e $var(X) = Cov(X,X)$

So the covariance between two dimensions is defined as:

$$Cov(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{n-1} \tag{A.5}$$

Covariance calculations are used to find relationships between dimensions in high dimensional data sets (usually greater than 3) where visualization is difficult. A Covariance matrix C is used to represent these relations. For 3-D dimensions (variables) C could be defined as:

$$C = \begin{bmatrix} cov(X,X) & cov(X,Y) & cov(X,Z) \\ cov(Y,X) & cov(Y,Y) & cov(Y,Z) \\ cov(Z,X) & cov(Z,Y) & cov(Z,Z) \end{bmatrix} \tag{A.6}$$

PROPERTIES:

- Diagonal: variances of the variables.
- $cov(X,Y) = cov(Y,X)$, hence this matrix is symmetrical about the diagonal (upper triangular).
- m-dimensional data will result in $m \times m$ covariance matrix.

Linear Algebra

Matrices

A matrix is a rectangular array of numbers written between rectangular brackets (or large parenthesis).

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

An important attribute of a matrix is its size or dimensions, i.e., the numbers of rows and columns. The matrix above has 3 rows and 4 columns, so its size is 3×3 (Square matrix). A matrix of size $m \times n$ is called an $m \times n$ matrix [11].

- **Matrix representation of a collection of vectors**: Matrices are very often used as a compact way to give a set of indexed vectors of the same size. For example, if $x_1, ..., x_N$ are n-vectors that give the n feature values for each of N objects, we can collect them all into one $n \times N$ matrix

$$X = [x_1 \ \ x_2 \ \ \cdots \ \ x_N],$$

- **Special Matrices**
  - **Zero matrix**: A zero matrix is a matrix with all elements equal to zero, referred sometimes as $O_{m \times n}$.
  - **Identity matrix**: An identity matrix is another common matrix. It is always square. Its diagonal elements are ones. It is donated by I_n
  - **Diagonal matrices**: A square $n \times n$ matrix A is diagonal if $A_{ij} = 0$ for $i \neq j$.
  - **Triangular matrices**: A square $n \times n$ matrix A is *upper triangular* if $A_{ij} = 0$ for $i > j$, and it is *lower triangular* if $A_{ij} = 0$ for $i < j$.
  - **Symmetric matrices**: A square matrix A is symmetric if $A = A^T$, i.e., $A_{ij} = A_{ji}$ for all i, j.
  - **Hermitian matrices**: A matrix $A \in \mathbb{C}^{n \times n}$ said to be self-adjoint or Hermitian if $A^{\mathcal{H}} = A$ here: $A^{\mathcal{H}} = (A^*)^{\mathcal{H}}$ with (*) denotes the complex-conjugate and (T) denotes the matrix transpose. The set of Hermitian matrices of order n is denoted by $\mathcal{H}_n$. A real matrix $A \in \mathbb{R}^{n \times n}$ is Hermitian if and only if $A^T = A$.

- **Matrix Transpose**: If A is an $m \times n$ matrix, its transpose, denoted A^T (or sometimes A' or A^*), is the $n \times m$ matrix given by $(A^T)_{ij} = A_{ji}$.
  - $(A + B)^T = A^T + B^T$.
  - $(AB)^T = B^T A^T$.

- **Matrix multiplication**: $A = [a_{ij}]_{m \times p}$; $B = [b_{ij}]_{p \times n}$; $AB = C = [c_{ij}]_{m \times n}$,where $c_{ij} = row_i(A).col_j(B)$.

- **Outer vector product**: let $a = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$; $b^T = [b_a \ \cdots \ b_m]$; $ab = c$, where c is $n \times m$ matrix.

- **Matrix-vector multiplication**: If A is $m \times n$ and x is an n-vector, then the matrix-vector product $y = Ax$ is the m-vector y with elements

$$y_i = \sum_{k=1}^{n} A_{ik} x_k \quad ; i = 1, \cdots, m. \tag{A.7}$$

- **Determinant and Trace**: A determinant is a function of a square matrix that reduces it to a single number. The determinant of a matrix A is denoted $|A|$ or det(A). The determinant of a square matrix $A = [a_{ij}]$ which has a size of $n \times n$ is:

$$det(A) = \sum_{j=1}^{n} a_{ij} A_{ij} \quad ; i = 1, \cdots, n. \tag{A.8}$$

$$A_{ij} = (-1)^{i+j} det(M_{ij})$$

The Trace of a matrix $A = [a_{ij}]$ is

$$tr(A) = \sum_{j=1}^{n} a_{jj} \tag{A.9}$$

- **Matrix Inversion**: For square matrix A, the inverse exist if A is non-singular. i.e (det(A)$\neq$ 0) A^{-1} is an inverse iff:

$$AA^{-1} = A^{-1}A = I \tag{A.10}$$

For non-square matrices, the existence of a pseudo-inverse required $A^T A$ non-singular.

$$A^{\#} = [A^T A]^{-1} A^T A^{\#} A = I \tag{A.11}$$

Inner Product in Vector spaces

Let $\mathcal{V}$ be a vector space over the field $\mathcal{F}$, where $\mathcal{F}$ is either $\mathbb{R}$ or $\mathbb{C}$. An inner product on $\mathcal{V}$ is a function $\langle .,. \rangle : \mathcal{V} \times \mathcal{V} \to \mathcal{F}$ such that for all $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \mathcal{V}$ and $\alpha, \beta \in \mathcal{F}$, we have the following:

1. $\langle \boldsymbol{u}, \boldsymbol{u} \rangle \geq 0$ and $\langle \boldsymbol{u}, \boldsymbol{u} \rangle = 0$ if and only if $\boldsymbol{u} = 0$.

2. $\langle \alpha \boldsymbol{u} + \beta \boldsymbol{v}, \boldsymbol{w} \rangle = \alpha \langle \boldsymbol{u}, \boldsymbol{w} \rangle + \beta \langle \boldsymbol{v}, \boldsymbol{w} \rangle$

3. $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \langle \boldsymbol{v}, \boldsymbol{u} \rangle^*$

We define a real (or complex) inner product space as a vector space $\boldsymbol{v}$ over $\mathbb{R}$ (or $\mathbb{C}$), along with an inner product defined on it and under standard inner product (dot product, scalar product), two fundamental spaces are defined as:

1. The n-dimensional Euclidean Space: is the inner product space $\mathbb{R}^n$ under the standard inner product, defined by:

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \boldsymbol{u}^T . \boldsymbol{v} = \sum_{i=1}^{n} \boldsymbol{u}_i \boldsymbol{v}_i \tag{A.12}$$

2. The n-dimensional Unitary Space: is the inner product space $\mathbb{C}^n$ under the standard inner product (dot product), defined by:

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \boldsymbol{v}^{\mathcal{H}} . \boldsymbol{u} = \sum_{i=1}^{n} \boldsymbol{u}_i \boldsymbol{v}_i^* \tag{A.13}$$

The Euclidean norm of a vector $\boldsymbol{u}$ is defined as the inner product of $\boldsymbol{u}$ with itself and it is defined as:

$$||\boldsymbol{u}||_2 = |\boldsymbol{v}|^2 = \langle \boldsymbol{u}, \boldsymbol{u} \rangle = \boldsymbol{u}^T . \boldsymbol{u} = \sum_{i=1}^{n} \boldsymbol{u}_i^2 \tag{A.14}$$

An inner product is a measure of collinearity:

- $\boldsymbol{u}$ and $\boldsymbol{v}$ are orthogonal iff: $\boldsymbol{u}^T \boldsymbol{v} = 0$
- $\boldsymbol{u}$ and $\boldsymbol{v}$ are collinear iff: $\boldsymbol{u}^T \boldsymbol{v} = ||\boldsymbol{u}|| ||\boldsymbol{v}||$

Linear independence

A set of n-dimensional vectors $x_i \in \mathbb{R}^n$, are said to be linearly independent if none of them can be written as a linear combination of the others. In other words:

$$\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_k x_k = 0 \quad \text{iff} \quad \alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$$

Span

A span of a set of vectors x_1, x_2, ... , x_k is the set of vectors that can be written as a linear combination of x_1, x_2, ... , x_k.

$$\text{span}(x_1, x_2, ..., x_k) = \{c_1 x_1 + c_2 x_2 + ... + c_k x_k | c_1, c_2, ..., c_k \in \mathbb{R}\}$$

Basis

A collection of n linearly independent n-vectors (i.e., a collection of linearly independent vectors of the maximum possible size) is called a *basis*.

A basis for $\mathbb{R}^n$ is a set of vectors which:

1. Span $\mathbb{R}^n$, i.e. any vector in this n-dimensional space can be written as linear combination of these basis vectors.

2. Are *linearly independent*.

Orthogonal/Orthonormal Basis

- An *orthonormal basis* of an a vector space V with an inner product, is a set of basis vectors whose elements are mutually orthogonal and of magnitude 1 (unit vectors).

- Elements in an *orthogonal basis* do not have to be unit vectors, but must be mutually perpendicular. It is easy to change the vectors in an orthogonal basis by scalar multiples to get an orthonormal basis, and indeed this is a typical way that an orthonormal basis is constructed (or Gram-Schmidt Orthonormalization Process).

- Two vectors are orthogonal if they are perpendicular, i.e., they form a right angle, i.e. if their inner product is zero.

$$\boldsymbol{u}^T \boldsymbol{v} = \sum_{i=0}^{n} \boldsymbol{u}_i \boldsymbol{v}_i = 0$$

The following statements are equivalent to saying that a real matrix $A \in \mathbb{R}^{n \times n}$ is orthonormal.

1. A has orthonormal columns.

2. A has orthonormal rows.

3. $A^{-1} = A^T$

4. $||A.x||_2 = ||x||_2$ for every $x \in \mathbb{R}^n$.

Eigenvalues and Eigenvectors

An *eigenvector* is a nonzero vector that satisfies the equation:

$$Ax = \lambda x \tag{A.15}$$

where A is a square matrix, λ is a scalar, and x is the eigenvector. λ is called an eigen- value. Eigenvalues and eigenvectors are also known as, respectively, characteristic roots and characteristic vectors, or latent roots and latent vectors [10]. The term Eigen-pair is used to address the pair (λ, x). An eigenvalue can be either simple or repeated. The spectrum of A, σ(A), is the multiset of all eigenvalues of A, with eigenvalue λ appearing $m(\lambda)$ times (algebraic multiplicity) in σ(A). The geometric multiplicity, $q(\lambda)$, of an eigenvalue λ is the number of linearly independent eigenvectors associated with the eigenvalue λ [7].

1. An eigenvalue λ is simple if $m(\lambda) = 1$.

2. An eigenvalue λ is semi-simple if $m(\lambda) = q(\lambda)$.

All eigenvectors of a symmetric matrix are perpendicular to each other, no matter how many dimensions we have.

One of the simplest forms that a matrix A can be transformed into, using equivalence or similarity transformation, a diagonal form. Based on the eigenvalues of the matrix A, the following results stands:

- Let $\lambda_1, \lambda_2, \cdots, \lambda_r$ be distinct eigenvalue of A, with $r \leq n$. If $A \in \mathbb{C}^{n\times n}$ then A is *diagonalizable* if and only if $m(\lambda_i) = q(\lambda_i)$ for $i = 1, \cdots, r$. If $A \in \mathbb{R}^{n\times n}$ then A is *diagonalizable* by non-singular matrix $M \in \mathbb{R}^{n\times n}$ if and only if all the eigenvalues of A are real and $m(\lambda_i) = q(\lambda_i)$ for $i = 1, \cdots, r$.

- A is diagonalizable if and only if A has n-linearly independent eigenvectors.

- If A has n distinct eigenvalues, then A is diagonalizable, otherwise we can transform it into *Jordan form*.

Singular Value Decomposition (SVD)

A *singular value decomposition* of a matrix $A \in \mathbb{C}^{n\times m}$ with rank$(A) = r$ and $m \leq n$ is a factorization of the form:

$$A = U.\Sigma.V^{\mathcal{H}} \tag{A.16}$$

Where $\Sigma = diag(s_1, s_2, \cdots, s_p) \in \mathbb{R}^{n\times m}$, $p = min(m,n)$ and $s_1 \geq s_2 \geq s_3 \geq \cdots \geq s_p \geq 0$. Both $U \in \mathbb{C}^{n\times n}$ and $V \in \mathbb{C}^{m\times m}$ are unitary.

- The diagonal entries of Σ are called the *singular values* of A.

- The columns of U are called left singular vectors of A.

- The columns of V are called right singular vectors of A.

Some remarks and properties of SVD are listed below:

1. The first r singular values are nontrivial while the last $p - r$ are zeros.

2. Every $A \in \mathbb{C}^{n\times m}$ has singular value decomposition. If $A \in \mathbb{R}^{n\times m}$ then U and V may be real.

3. The singular values of a matrix are unique [10].

4. If $A = U.\Sigma.V^{\mathcal{H}}$ is the singular value decomposition of A, then the following relations hold.
$$A.v_i = s_i.u_i, \quad A^{\mathcal{H}}.u_i = s_i.u_i, \quad u_i^{\mathcal{H}}.A.v_i = s_i, \quad i = 1, 2, \cdots, p$$

5. The nonzero singular values of A are the square roots of the nonzero eigenvalues of $A^{\mathcal{H}}.A$ or $A.A^{\mathcal{H}}$. The columns of V are eigenvectors of $A^{\mathcal{H}}.A$ and The columns of the matrix U are eigenvectors of $A.A^{\mathcal{H}}$ [10].

6. If $A \in \mathbb{C}^{n\times n}$ is Hermitian with eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$, then the singular values of A are $|\lambda_1|, |\lambda_2|, ..., |\lambda_n|$.

The insight underlying the use of SVD for these tasks is that it takes the original data, usually consisting of some variant of a word×document matrix, and breaks it down into linearly independent components. These components are in some sense an abstraction away from the noisy correlations found in the original data to sets of values that best approximate the underlying structure of the dataset along each dimension independently. Because the majority of those components are very small, they can be ignored, resulting in an approximation of the data that contains substantially fewer dimensions than the original. SVD has the added benefit that in the process of dimensionality reduction, the representation of items that share substructure become more similar to each other, and items that were dissimilar to begin with may become more dissimilar as well. In practical terms, this means that documents about a particular topic become more similar even if the exact same words don't appear in all of them [10].

Appendix B

Description of the Cement Rotary Kiln

Cement kilns, in general, are the heart of the cement production process; their capacity usually determines the overall capacity of the cement plant. Cement kiln is mainly used for calcining the cement clinker and it can be divided into dry-producing cement kiln and wet-producing cement kiln. In the wet process, raw meal is supplied at ambient temperature in the form of a slurry with about 40% of water. In modern works, the blended raw material enters the kiln via the pre-heater tower. Here, hot gases from the kiln, and probably the cooled clinker at the far end of the kiln, are used to heat the raw meal. As a result, the raw meal is already hot before it enters the kiln. The dry process is much more thermally efficient than the wet process.

As the most energy consuming, and the most critical and complex part in the production process, improvements of cement kilns has been the central concern of the cement manufacturing technology, thereby limiting the energy consumption, the green-house gases and improving the efficiency. To increase the kiln efficiency, a system of suspension preheater cyclones plus pre calciner are usually used. In such complex system, failures and malfunctions have a high probability of occurrence. Therefore, the need of an automatic system for detecting the presence of faults and determining their roots is mandatory.

Ain El Kebira cement plant in the Algerian east, where the work is conducted, is a dry process consisting of four cyclone levels in the suspension preheater and short rotary kiln

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

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

Fig. B.1 The cement plant of Ain El Kebira

Table B.1 Process variables of the cement rotary kiln.

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 Depression of gas in the outlet of the smoke filter of tower I	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	m3 /h
44	Flow of fuel (natural gas) to the secondary burner (pre-calcination level)	m3 /h

of 5.4 (m) shell diameter (without brick and coating) and 80 (m) length, with 3 incline. The kiln is spun up to 2.14(rpm) using two 560 (kws) asynchronous motors and producing clinker of density varying from 1300 to 1450 (kg/m3) under normal conditions. Two natural gas burners are used, the main one in the discharge end and the secondary in the first level of preheater tower. Figure B.2 gives simplified schematic of the cement rotary kiln with the used signals. The description of these signals is given in details in table B.1.

Due to the complex dynamic, multivariable nature, nonlinear reaction kinetics, long time delays and variable raw material feed characteristics, the rotary kiln process is inherently difficult to model. It was declared in [50] that “to the authors knowledge, there is no mathematical model that adequately describes the process behavior”.

This application was made with a data collected from the cement plant of Ain El-Kbira, Setif, Algeria. from January 23rd, 2014 at 23:39:33PM to January 24th, 2014 at 04:30:00AM with acquisition rate of one second. The current control performed in Process is manual centralized mode, where any action is based upon the human experience. The used data sets are described in details in Table 4.1. They consist of a continuous measurement for 4 hours, 50 minutes and 28seconds (177428 seconds). The first 15300 seconds are collected when the system is surely in healthy state. Another 2084 seconds are collected in the presence of a real fault in the system [7].

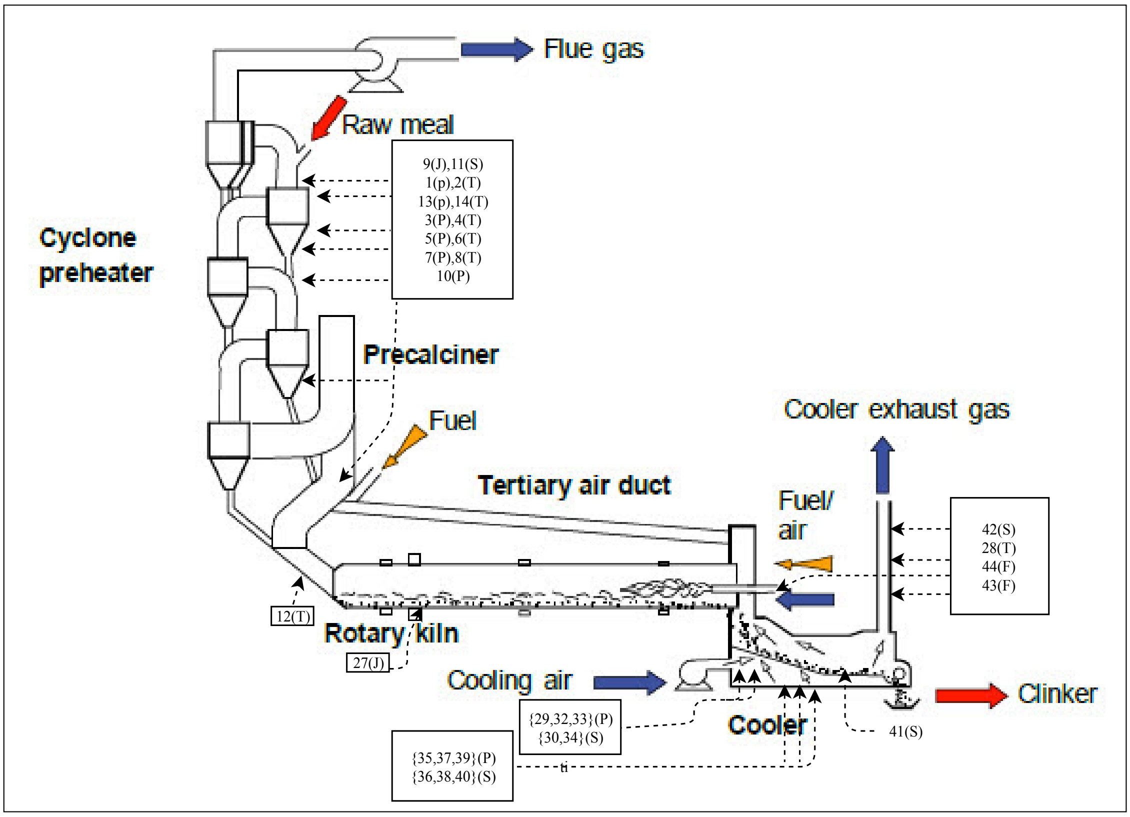


Fig. B.2 An overview about the cement plant rotary kiln process and the used signals, including:temperature(T), pressure(P), speed(S) and feed(F).

References

[1] Abdi, H. and Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4):433–459.

[2] Adnan, N. A., Cheng, Y., Izadi, I., and Chen, T. (2013). Study of generalized delay-timers in alarm configuration. *Journal of Process Control*, 23(3):382 – 395.

[3] Adnan, N. A., Izadi, I., and Chen, T. (2011). On expected detection delays for alarm systems with deadbands and delay-timers. *Journal of Process Control*, 21(9):1318–1331.

[4] Afzal, M. S., Chen, T., Bandehkhoda, A., and Izadi, I. (2018). Analysis and design of time-deadbands for univariate alarm systems. *Control Engineering Practice*, 71(96–107).

[5] Alkaya, A. and Eker, I. (2011). A new threshold algorithm based pca method for fault detection in transient state processes. In *2011 7th International Conference on Electrical and Electronics Engineering (ELECO)*, pages II–144. IEEE.

[6] Alrowaie, F., Kwok, K. E., and Gopaluni, B. (2014). Fault isolation based on general observer scheme in stochastic non-linear state-space models using particle filters. In *ADCONP Hiroshima*.

[7] ARIBI Yacine, GUERMI Hamza, S. b. K. (2016). Pca-based approach for fault detection in cement rotary kiln. Master's thesis, Institute of Electrical and Electronic Engineering, University M'Hamed BOUGARA – Boumerdes.

[8] Bahar-Gogani, M., Aslansefat, K., and Shoorehdeli, M. A. (2017). A novel extended adaptive thresholding for industrial alarm systems. In *2017 Iranian Conference on Electrical Engineering (ICEE)*, pages 759–765. IEEE.

[9] Bakdi, A., Kouadri, A., and Bensmail, A. (2017). Fault detection and diagnosis in a cement rotary kiln using pca with ewma-based adaptive threshold monitoring scheme. *Control Engineering Practice*, 66:64–75.

[10] Baker, K. (2005). Singular value decomposition tutorial. *The Ohio State University*, 24.

[11] Boyd, S. and Vandenberghe, L. (2018). *Introduction to Applied Linear Algebra: Vectors, Matrices, and Least Squares*. Cambridge university press.

[12] Chang, Y., Khan, F., and Ahmed, S. (2011). A risk-based approach to design warning system for processing facilities. *Process Safety and Environmental Protection*, 89(5):310 – 316.

[13] Cheng, Y., Izadi, I., and Chen, T. (2011). On optimal alarm filter design. In *4th International Symposium on Advanced Control of Industrial Processes*, pages 139–14.

[14] Dinno, A. (2009). Exploring the sensitivity of horn's parallel analysis to the distributional form of random data. *Multivariate behavioral research*, 44(3):362–388.

[15] Equipment, E. and Association, M. U. (2013). *Alarm systems: a guide to design, management and procurement*. Engineering Equipment and Materials Users' Association, London, England, second edition edition.

[16] F. Higuchi, I. Yamamoto, T. T. M. N. and Nishitani, H. (2009). Use of event correlation analysis to reduce number of alarms. In *10th International Symposium on Process Systems Engineering*, pages 1521–1526, Salvador.

[17] Finkel, D. E. (2003). *DIRECT optimization algorithm user guide*. Center for Research in Scientic Computation, North Carolina State University Raleigh, NC 27695-8205.

[18] Fodor, I. K. (2002). A survey of dimension reduction techniques. Technical report, Lawrence Livermore National Lab., CA (US).

[19] Franklin, S. B., Gibson, D. J., Robertson, P. A., Pohlmann, J. T., and Fralish, J. S. (1995). Parallel analysis: a method for determining significant principal components. *Journal of Vegetation Science*, 6(1):99–106.

[20] FuXiao, Wang, S., Xu, X., and Ge, G. (2009). An isolation enhanced PCA method with expert-based multivariate decoupling for sensor FDD in air-conditioning systems. *Applied Thermal Engineering*, 29:712–722.

[21] G Verdier, N Hilgert, J. P. V. (2008). Adaptive threshold computation for CUSUM-type procedures in change detection and isolation problems. *Computational Statistics & Data Analysis*, 52(4161–4174).

[22] Gagniuc, P. A. (2017). *Markov chains: from theory to implementation and experimentation*. John Wiley & Sons.

[23] Glorfeld, L. W. (1995). An improvement on horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and psychological measurement*, 55(3):377–393.

[24] Goel, P., Datta, A., and Mannan, M. S. (2017). Industrial alarm systems: challenges and opportunities. *Journal of Loss Prevention in the Process Industries*, 50:23–36.

[25] Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika*, 19(2):149–161.

[26] Han, L., Gao, H., Xu, Y., and Zhu, Q. (2016). Combining FAP, MAP and correlation analysis for multivariate alarm thresholds optimization in industrial process. *Journal of Loss Prevention in the Process Industries*, 40:471–478.

[27] HARKANE Mohamed Amine, S. b. K. (2018). On fault detection in a grid-connected photovoltaic system under two different operating modes. Master's thesis, Institute of Electrical and Electronic Engineering, University M'Hamed BOUGARA – Boumerdes.

[28] Haven, M. (1994). *Health and Safety Executive, The explosion and* $_r esattheTexacorefinery.HSEBooks,UK.$

[29] Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2):179–185.

[30] Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6):417.

[31] Hotelling, H. (1947). *Multivariate quality control-illustrated by air testing of sample bombsights*, chapter Techniques of Statistical Analysis, pages 11–184. Mcgraw-Hill, New York.

[32] Hu, J., Cai, S., and Zhang, L. (2017). Alarm management technology and its progress in process industries. In *2017 2nd International Conference on Control, Automation and Artificial Intelligence (CAAI 2017)*. Atlantis Press.

[33] Huo, M., Luo, H., Yin, S., and Kaynak, O. (2018). A Data-Driven Fault Detection Approach for Periodic Rectangular Wave Disturbance. In *IEEE Industrial Electronics Society*.

[34] ISA (2009). *Management of alarm systems for the process industries*. ISA 18.02, North Carolina.

[35] Izadi, I., Shah, S., Shook, D., Kondaveeti, S. R., and Chen, T. (2009a). Optimal alarm design. In *Fault Detection, Supervision and Safety of Technical Processes*, pages 651–656.

[36] Izadi, I., Shah, S. L., and Chen, T. (2010). Effective resource utilization for alarm management. In *49th IEEE Conference on Decision and Control (CDC)*, pages 6803–6808. IEEE.

[37] Izadi, I., Shah, S. L., Kondaveeti, S., and Chen, T., editors (2009b). *A framework for optimal design of alarm systems*, volume 14, Barcelona, Spain. 7th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes.

[38] Izadi, I., Shah, S. L., Shook, D. S., Kondaveeti, S. R., and Chen, T. (2009c). A framework for optimal design of alarm systems. *IFAC Proceedings Volumes*, 42(8):651–656.

[39] J. Ahnlund, T. B. and Spaanenburg, L. (2003). Rule-based reduction of alarm signals in industrial control. *Intelligent and Fuzzy Systems, IOS press*, 14:73–84.

[40] Jackson, J. E. and Mudholkar, G. S. (1979). Control procedures for residuals associated with principal component analysis. *Technometrics*, 21(3):341–349.

[41] Jiandong Wang, Fan Yang, T. C. and Shah, S. L. (2015). AN OVERVIEW OF INDUSTRIAL ALARM SYSTEMS. In *IEEE Transactions on Automation Science and Engineering*.

[42] Jianfeng Zhu, Yidan Shu, J. Z. F. Y. (2014). A dynamic alarm management strategy for chemical process transitions. *Journal of Loss Prevention in the Process industries*, 30:207–218.

[43] Jinqiu Hu, S. C. and Zhang, L. (2017). Alarm Management Technology and Its Progress in Process Industries. In *Advances in Intelligent Systems Research*, volume 134. 2nd International Conference on Control, Automation, and Artificial Intelligence (CAAI 2017).

[44] Jolliffe, I. (2011). *Principal component analysis*. Springer.

[45] Jolliffe, I. T. (1972). Discarding variables in a principal component analysis. i: Artificial data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 21(2):160–173.

[46] Jolliffe, I. T. and Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065):20150202.

[47] Jónás, T., Kalló, N., and Tóth, Z. E. (2014). Application of markov chains for modeling and managing industrial electronic repair processes. *Periodica Polytechnica Social and Management Sciences*, 22(2):87–98.

[48] Jones, D. R., Perttunen, C. D., and Stuckman, B. E. (1993). Lipschitzian optimization without the lipschitz constant. *Journal of optimization Theory and Applications*, 79(1):157–181.

[49] Kondaveeti, S. R., Izadi, I., L.Shah, S., Black, T., and Chen, T. (2012). Graphical tools for routine assessment of industrial alarm systems. *Computers & Chemical Engineering*, 46:39–47.

[50] Kouadri, A., Bensmail, A., Kheldoun, A., and Refoufi, L. (2014). An adaptive threshold estimation scheme for abrupt changes detection algorithm in a cement rotary kiln. *Journal of Computational and Applied Mathematics*, 259:835 – 842. Recent Advances in Applied and Computational Mathematics: ICACM-IAM-METU.

[51] Kourti, T. (2002). Process analysis and abnormal situation detection: From theory to practice. *IEEE Control Systems Magazine*, 22:10–25.

[52] Lawler, G. F. (2018). *Introduction to stochastic processes*. Chapman and Hall/CRC, second edition edition.

[53] Levy, B. C. (2008). *Principles of signal detection and parameter estimation*. Springer Science & Business Media.

[54] Lin, J., Sun, D., Bai, X., and Chen, L. (2018). A generalized alarm delay-timer's performance indices computing method. *Systems Science & Control Engineering*, 6(3):297–304.

[55] Papoulis, A. and Pillai, S. U. (2002). *Probability, random variables, and stochastic processes*. Tata McGraw-Hill Education.

[56] Raich, A. and Çinar, A. (1996). Statistical process monitoring and disturbance diagnosis in multivariate continuous processes. *AIChE JOURNAL*, 42:995–1009.

[57] Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., et al. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40(8):913–929.

[58] Rothenberg, D. H. (2009). *Alarm management for process control: a best-practice guide for design, implementation, and use of industrial alarm systems*. Momentum Press.

[59] Sandeep, K. (2009). Application of multivariate statistics for efficient alarm generation. *IFAC Proceedings Volumes*, 42(8):657–662.

[60] Schleburg, M., Christiansen, L., F.Thornhill, N., and Fay, A. (2013). A combined analysis of plant connectivity and alarm logs to reduce the number of alerts in an automation system. *Journal of Process Control*, 23:839–851.

[61] Su, J., Guo, C., Zang, H., Yang, F., Huang, D., Gao, X., and Zhao, Y. (2018). A multi-setpoint delay-timer alarming strategy for industrial alarm monitoring. *Journal of Loss Prevention in the Process Industries*, 54(1–9).

[62] Tan, W., Sun, Y., Azad, I. I., and Chen, T. (2017). Design of univariate alarm systems via rank order filters. *Control Engineering Practice*, 59:55 – 63.

[63] Van Kampen, N. G. (1992). *Stochastic processes in physics and chemistry*, volume 1. Elsevier.

[64] Venkatasubramanian, V., Rengaswamy, R., N.Kavuri, S., and KewenYin (2003). A review of process fault detection and diagnosis: Part III: Process history based methods. *Computers & Chemical Engineering*, 27:327–346.

[65] Wall, M. E., Rechtsteiner, A., and Rocha, L. M. (2003). Singular value decomposition and principal component analysis. In *A practical approach to microarray data analysis*, pages 91–109. Springer.

[66] Wang, J. and Chen, T. (2013). An online method for detection and reduction of chattering alarms due to oscillation. *Computers & Chemical Engineering*, 54:140–150.

[67] Wang, J. and Chen, T. (2014). An online method to remove chattering and repeating alarms based on alarm durations and intervals. *Computer and chemical engineering*, 67:43—-52.

[68] Wang, J. and Qin, S. J. (2002). A new subspace identification approach based on principal component analysis. *Journal of process control*, 12(8):841–855.

[69] Wise, B. M., Ricker, N., Veltkamp, D., and Kowalski, B. R. (1990). A theoretical basis for the use of principal component models for monitoring multivariate processes. *Process control and quality*, 1:41–51.

[70] Xu, J., Wang, J., Izadi, I., and Chen, T. (2012). Performance assessment and design for univariate alarm systems based on FAR, MAR, and AAD. In *IEEE Transactions on Automation Science and Engineering*, volume 9, pages 296–307.

[71] Yin, L., Yang, R., Gabbouj, M., and Neuvo, Y. (1996). Weighted median filters: a tutorial. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, 43(3):157–192.

[72] Yu, K.-Y. C. and Bricker, D. L. (1993). Analysis of a markov chain model of a multi-stage manufacturing system with inspection, rejection, and rework. *IIE transactions*, 25(1):109–112.

[73] Yue, H. H. and Qin, S. J. (2001). Reconstruction-based fault identification using a combined index. *Industrial & engineering chemistry research*, 40(20):4403–4414.

[74] Zang, H., Yang, F., and Huang, D. (2015). Design and Analysis of Improved Alarm Delay-Timers. *IFAC-PapersOnLine*, 48:669–674.

[75] Zhu, J., Ge, Z., and Song, Z. (2017). Distributed Parallel PCA for Modeling and Monitoring of Large-Scale Plant-Wide Processes With Big Data. *IEEE Transactions on Industrial Informatics*, 13:1877–1885.

[76] Zumoffen, D. and Basualdo, M. (2008). From large chemical plant data to fault diagnosis integrated to decentralized fault-tolerant control: pulp mill process application. *Industrial & Engineering Chemistry Research*, 47(4):1201–1220.

[77] Zwick, W. R. and Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological bulletin*, 99(3):432.