CRANFIELD UNIVERSITY

P. P. IKETUBOSIN

Studies on Non-linear Dynamic Process Monitoring

SCHOOL OF ENGINEERING

PHD THESIS

CRANFIELD UNIVERSITY

SCHOOL OF ENGINEERING
DEPARTMENT OF OFFSHORE, PROCESS AND ENERGY
ENGINEERING

PHD THESIS

Academic Year 2010 - 2011

PABARA-EBIERE PATRICIA IKETUBOSIN

Studies on Non-linear Dynamic Process Monitoring

Supervisor: DR Y. CAO

July 2011

**This thesis is submitted in fulfilment of the requirements for the
degree of Doctor of Philosophy**

Abstract

Owing to the numerous benefits of process monitoring, the subject has attracted a lot of attention in the last two decades. Process monitoring is an art of identifying abnormal deviations in a process from the normal operating condition using various techniques. Generally, the development of these monitoring techniques is geared towards applying these techniques to industrial processes. In addition, most industrial processes are dynamic and non-linear in nature. Therefore, in the development of the monitoring algorithms, the dynamic as well as the non-linear properties of the plant should be taken into consideration.

Process monitoring techniques like the Principal Component Analysis (PCA) and Partial Least Squares (PLS) regression analysis were developed based on the assumption that the process data is normally distributed. Nevertheless, this assumption of normality is invalid for most industrial processes due to the non-linear nature of these plants. For such processes, the distribution of the process variables in general will be non-Gaussian, therefore making the widely applied PCA and PLS approaches inappropriate for the monitoring of plants. To address this limitation of the PCA and PLS for Dynamic processes, the Dynamic PCA (DPCA) and dynamic PLS (DPLS) approaches were developed.

The challenge of efficiently monitoring process plants with dynamic and non-linear characteristics is the motivation for this study. The overall aim of this study is to develop process monitoring strategies that are able to take the dynamic and non-linear properties of the plant into account. With these strategies, more efficient performance monitoring of the plant can be achieved.

To address the challenge of efficiently monitoring process plants with both dynamic and non-linear properties, in this work, existing multivariate monitoring techniques like Dynamic Principal Component Analysis (DPCA), Dynamic Partial Least Square Regression (DPLS) and the Canonical Variate Analysis (CVA) are extended us-

ing Kernel Density Estimations (KDE) resulting in the novel **DPCA with KDE**, **DPLS with KDE** and **CVA with KDE** techniques. In addition, another novel approach, **State Space Independent Component Analysis** (SSICA) is developed to improve performance monitoring. Furthermore, the extended and developed techniques in this work are evaluated using simulated data of the Tennessee Eastman Process (TEP) Plant and a waste water treatment plant. The TEP plant is a complex dynamic and non-linear process that was developed by the Eastman Chemical Company. The TEP plant is commonly employed in the process monitoring community. For this reason, the TEP plant is a good platform for comparison and is therefore, adopted for the evaluation of the various monitoring algorithms in this work. Furthermore, the simulated waste water treatment plant, which has been extensively used and is widely accepted as a benchmark to evaluate monitoring and control strategies was also adopted to evaluate the monitoring methods in this work.

The techniques developed in this work are compared with some existing techniques and are able to significantly improve the process monitoring performance over the existing techniques.

Acknowledgments

I would like to thank my supervisor, Dr Yi Cao immensely for his constant guidance and directions through this entire PhD project. The exposure I have gotten from him during the my research period is far more than I ever imagined I would gather. His continuous guidance has given birth to three refereed conference and journal publications of very high standard.

It is also a pleasure to thank the many people who have contributed in various ways to the completion of this PhD dream. Dr Meihong Wang has constructively criticised my work from the start of this PhD to this point, making very useful suggestions to me especially during my reviews. My gratitude also goes to Mrs Samantha Skears for all the support I got from her during my time in the Process Systems Engineering Group.

My colleagues past and present in the group have been really wonderful, encouraging creativity and innovation. The broad range of knowledge and skills within the group have positively influenced the research path in this PhD project.

Very special thanks to my husband George Iketubosin for all the support and understanding he has shown during this PhD programme. Without his support none of this PhD dream would be true. I would also like to appreciate my baby Eliana Iketubosin for being the best baby a studying mother can ever ask for.

My gratitude also goes to my parents, Mr and Mrs Nathan B. Odiowei for all the love, support and sacrifices they have made through this PhD programme and indeed, my entire life time. Also, I would like to thank my uncles Prof Steve Odi-Owei and General Stanley Diriyai and their families for the great support I have gotten from them all. My gratitude also goes to my cousin Inebimo for her constant phone calls and support.

Contents

References   **149**

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

The importance of improving process monitoring strategies is emphasised by the ever increasing demand for better efficiency in the chemical, pharmaceutical, manufacturing, food and waste water treatment industries to mention a few. Process monitoring is a means of identifying variations in a process from the normal operating process. In addition, process monitoring is an important asset-management technology to maintain high performance of automation systems in operating processes. Process monitoring consists of detection and diagnosis. Detection involves identifying abnormal deviations in a process from normal operating conditions while diagnosis involves investigating the reason for the occurrence of the deviations in the process. The major goal of process monitoring is to detect abnormal deviations early and investigate the reason for the occurrence of such abnormal deviations.

Generally, process monitoring techniques involve two steps; the off-line training and the on-line monitoring. The off-line training consist of developing models that reflect the normal operating process and then from the estimated model, deriving a

control limit which is able to determine the status of the monitored process. The on-line monitoring involves estimating latent variables from the new data using the model parameters determined in the off-line training step and then using the derived control limit to determine whether the new process is 'in-control' or 'out-of-control'. Processes are said to be in control when the abnormal deviations mentioned above are absent and said to be 'out-of-control' when abnormal deviations are present. The success of a process monitoring technique depends greatly on;

- How accurately the developed model reflects the normal operating process

- The accuracy and appropriateness of the derived control limit.

A process monitoring model is a mathematical representation of a process. These monitoring models can be statistical or mechanical models. The statistical models are those models that are developed based on historical process data whereas the mechanical models are established by physical and chemical reactions and depend on detailed process information.

Most industrial processes have a large number of process variables that could be auto-correlated and cross-correlated. For such processes, a mechanical model will be time consuming and expensive to build. Statistical models on the other hand, do not require detailed properties of the process and are not expensive to build. Owing to these benefits of the statistical models, statistical models have been adopted in this study. Furthermore, the control limit mentioned above can be defined as a reference mark that determines the 'in-control' or 'out-of-control' status of a monitored process.

1.2 Motivation for Study

Process monitoring approaches based on statistical models are referred to as Multivariate Statistical Process Monitoring (MSPM) techniques. These MSPM techniques are a collection of useful tools for the early detection and diagnosis of abnormal conditions in a process.

Amongst MSPM techniques, the PCA and PLS have been reported the most. Also, the PCA and PLS techniques are static models that are inadequate for dynamic processes. To address this limitation, dynamic extensions of the PCA and PLS approaches known as dynamic PCA (DPCA) and dynamic PLS (DPLS) were developed [1, 2]. Nevertheless, the DPCA and DPLS are not the best approaches for dynamic systems as they may not be able to capture some important dynamic behaviours [3, 4]. Also, the latent variables extracted from the DPCA and DPLS approaches are not guaranteed to yield accurate and minimal dynamic representations [3].

The Canonical Variate Analysis (CVA) on the other hand is reported to be an efficient solution for dynamic processes [3, 4, 5, 6, 7]. The CVA is a linear dynamic monitoring approach to estimate the minimum number of state variables, which is reported to be more efficient than the DPCA and DPLS approaches [4].

Traditionally, the DPCA, DPLS and CVA approaches discussed above are generally associated with the T^2 and Q metrics [1, 2, 3, 8, 9, 10, 11, 12]. Both metrics are estimated based on the Gaussian assumption. However, most industrial plants are non-linear, following a non-Gaussian distribution. For such non-linear processes, the Gaussian assumption required by the T^2 and Q metrics is invalid. As a result, the traditional DPCA, DPLS and CVA approaches are inappropriate for non-linear processes and may not be able to correctly identify the underline faults.

In addition, the states obtained from the CVA are only de-correlated and not statis-

tically independent. The problem can be addressed by employing the Independent Component Analysis (ICA). The ICA technique recovers a few statistically independent source signals known as the independent components (ICs) from collected process measurements by assuming that these ICs are non-Gaussian. The ICA is well suited to non-linear plants following a non-Gaussian distribution.

Also, most published ICA approaches have employed the static PCA in the pre-processing stage for whitening. However, the PCA is not appropriate for dynamic systems, consequently making the PCA associated ICA inappropriate for dynamic systems. To address this limitation of the PCA associated ICA, the dynamic ICA (DICA) was proposed in which the dynamic PCA is employed for the whitening stage. Nevertheless, the performance of the DICA is still unsatisfactory.

The development of process monitoring techniques is geared towards applying these techniques to industrial processes in order to improve process condition monitoring. It is well known that most industrial processes to which these monitoring techniques are applied are both dynamic and non-linear. This non-linearity makes the process variables driven by noise and disturbances to follow a non-Gaussian distribution. In practice to achieve efficient condition monitoring of industrial plants, the dynamic and non-linear properties of the plants should be taken into consideration.

So far, monitoring techniques have been developed to address either the dynamic or non-linear properties associated with most industrial plants. Furthermore, an attempt has been made to simultaneously address the dynamic and non-linear issues associated with industrial plants, however unsatisfactorily. This emphasises the desperate demand for appropriate and efficient monitoring of industrial plants with both dynamic and non-linear characteristics, which is the motivation for this work.

1.3   Aim and Objectives of Study

The aim of this work is to improve process condition monitoring by developing monitoring strategies that are able to simultaneously address the dynamic and non-linear issues commonly associated with most industrial processes. Furthermore, the objectives of this research include:

- Literature review on process monitoring approaches

- Undertake a study on Dynamic Principal Component Analysis (DPCA).

- Undertake a study on Dynamic Partial Least Squares (DPLS).

- Undertake a study on Canonical Variate Analysis (CVA).

- Carried out a study on Independent Component Analysis (ICA).

- Undertake a study on Dynamic Independent Component Analysis (DICA).

- Carried out extensive study on Kernel Density Estimations (KDE).

- Carried out a study on the Tenneessee Eastman Process (TEP) Plant.

- Carried out a study of a Waste Water Treatment Plant.

1.4   Work Done

The efficiencies of the traditional DPCA, DPLS and CVA approaches for dynamic processes have been reported [1, 2, 3]. However, the association of these approaches with control limits estimated based on the Gaussian assumption such as the Hotelling's T^2 and the Q metrics makes these approaches insufficient for non-linear processes. One solution to the problem of monitoring non-Gaussian processes due to the non-linearity of such processes is to directly estimate the probability density function

(PDF) of the T^2 and Q metrics through the kernel density estimations (KDE) [13]. The KDE is a well established approach to estimate the PDF particularly for univariate random processes [14]. Hence, the KDE is suitable for the T^2 and Q metrics which are univariate although the underlying processes are multivariate.

In this work, to address the limitations of the DPCA, DPLS and CVA for non-linear processes, all three approaches are extended by deriving more appropriate control limits using kernel density estimations (KDE). The resulting development is the novel DPCA with KDE, DPLS with KDE and CVA with KDE approaches. Furthermore, all the proposed KDE approaches are applied to the Tenneessee Eastman Process (TEP) Plant. Their monitoring performances are compared with those of their non-KDE counterparts before comparing the novel approaches, one with another. The improvement of the proposed approaches over their non-KDE counterparts is also illustrated in this thesis. Amongst the KDE approaches developed in this work, the CVA with KDE is able to significantly improve the monitoring performance over the DPCA with KDE and the DPLS with KDE approaches.

Although the CVA with KDE technique was superior to the other approaches, the state variables obtained from the CVA are only de-correlated but not statistically independent, hence are not efficient enough for non-linear process monitoring. On the contrary, the ICs extracted from the ICA approach are not only decorrelated but are also statistically independent. Thus, the ICA [15, 16] is a possible solution to the limitation of the CVA and is therefore considered in this work.

Conventionally, most of the reported studies on the ICA have utilised the PCA for the whitening and dimension reduction in the pre-processing stage before employing the ICA to extract the statistically independent components. In the pre-processing stage the measurement variables are de-correlated by the whitening procedure and then the process dimension reduced to minimize mathematical complexities. However, the use of the PCA in the pre-processing stage of the ICA makes such ICA

approaches inappropriate for dynamic process monitoring due to the limitations of the PCA for dynamic processes mentioned above. This means that although the ICA approach is well suited for non-linear processes, it is unsuitable for dynamic processes.

To address this limitation of the ICA for dynamic processes, a dynamic ICA (DICA) method was developed where the DPCA was employed in the pre-processing stage for whitening before applying the ICA for the extraction of the statistical independent components from the principal components. The so called DICA approach is an attempt to simultaneously address the dynamic and non-linear issues associated with most real time processes. Nevertheless, the DICA like the DPCA is not the best approach to capture dynamic behaviour from process measurements [3]. As a result, the statistical advantage of the ICA is not fully exploited by the DICA, making the monitoring of dynamic processes using the DICA technique still unsatisfactory.

To address the limitation of the ICA and DICA techniques mentioned above, an efficient dynamic ICA-based process monitoring technique is developed in this work. In the proposed ICA-based approach, the CVA which is reported to be the best approach to capture dynamic behaviour is employed for the pre-processing stage of the ICA to construct a state space before applying the ICA algorithm to extract the independent components from the constructed state space, resulting in the new State Space Independent Component Analysis (SSICA). The control limits of the proposed SSICA approach are determined using the KDE. The proposed SSICA approach is able to improve condition monitoring over the CVA and DICA approaches as illustrated in this work.

1.5 Case Studies

The algorithms developed and considered in this work are evaluated by applying them to simulated data from the Tennessee Eastman Process (TEP) Plant and a Waste Water Treatment Plant. Both case studies are briefly discussed below.

1.5.1 Tennessee Eastman Process Plant

The Tennessee Eastman Process (TEP) Plant is one of the case studies employed in this work. It was created by the Eastman Chemical Company to provide a realistic industrial process to evaluate control and monitoring strategies. The TEP is based on a simulation of an actual industrial process that has been modified for proprietary reasons. The choice of the TEP in this work is motivated by the fact that the TEP plant has both dynamic and non-linear properties and is therefore a good representation of most industrial processes. In addition, the TEP plant is commonly employed in the process monitoring community. This makes it possible to compare the algorithms developed in this work with already existing algorithms from other published works. A detailed description of the TEP plant is also presented in this thesis.

1.5.2 Activated Sludge Model No. 1

The Activated Sludge Model No.1 (ASM1) is a simulated waste water treatment plant (WWTP) focusing on the biological treatment of a waste water treatment process [17, 18]. The WWTP is non-linear in nature and subject to dynamic changes. The ASM1 plant is chosen in this work for the evaluation of the developed algorithms because of its dynamic and non-linear properties . Also, the ASM1 plant is accepted as a standard in the monitoring community due to its extensive use [15, 17, 18, 19].

1.6   Contributions of Study

From this work, some monitoring techniques have been developed that are able to simultaneously take both the dynamic and non-linear properties of plants into account. Four novel monitoring approaches have been developed while some existing monitoring techniques are also considered. In addition, these techniques have been applied to two simulated case studies which are the Tenneessee Eastman Process Plant (TEP) and the Activated Sludge Model No. 1 (ASM1). The contributions from this study are listed below;

- Developed a DPCA based approach that is adapted to non-linear plants using the KDE resulting in the novel DPCA with KDE approach.

- Application of the novel DPCA with KDE approach to the TEP Plant.

- Developed a DPLS based approach that is adapted to non-linear plants using the KDE resulting in the novel DPLS with KDE approach.

- Application of the novel DPLS with KDE approach to the TEP Plant.

- Developed a CVA based approach that is adapted to non-linear plants using the KDE resulting in the novel CVA with KDE approach.

- Application of the novel CVA with KDE approach to the TEP plant

- Developed a novel State Space Independent Component Analysis (SSICA) for an efficient monitoring of real-time plants with both dynamic and non-linear properties.

- Application of the novel SSICA approach to the TEP plant

- Application of the novel SSICA approach to a Waste Water Treatment Plant.

- Application of the existing DICA approach to a Waste Water Treatment Plant for the purpose of comparison.

- Application of the novel DPCA with KDE approach to a Waste Water Treatment Plant.

- Application of the novel DPLS with KDE approach to a Waste Water Treatment Plant.

- Application of the novel CVA with KDE approach to a Waste Water Treatment Plant.

1.7 Publications

Journal Publications

- **P. Odiowei and Y. Cao** (2009). Nonlinear Dynamic Process Monitoring using Canonical Variate Analysis and Kernel Density Estimations. IEEE Transactions on Industrial Informations 6 (1), Pages 36 - 45.

- **P. P. Odiowei and Y. Cao** (2010). State Space Independent Component Analysis for Nonlinear Dynamic Process Monitoring. Chemometrics Intelligent Laboratory Systems, Volume 103, Issue 1, Pages 59 - 65.

Conference Publications

- **P. P. Odiowei and Y. Cao** (2008). Kernel Density enhanced PCA for Process Monitoring of a Waste Water Treatment Plant (WWTP). 1st Postgraduate Research Conference, 29th - 30th September, 2008, Dundee, United Kingdom.

10

- **P. P. Odiowei and Y. Cao** (2009). Nonlinear Dynamic Process Monitoring using Canonical Variate Analysis and Kernel Density Estimations. 10th International Symposium on Process Systems Engineering - PSE09.

1.8 Thesis Outline

This thesis consists of six chapters and the content of each chapter is summarised below.

Chapter 1: Introduction

This chapter presents a brief discussion on the background of this dissertation, outlining the motivation for the study as well as the aim and objectives of this research work.

Chapter 2: Literature Review

The fundamentals of process monitoring is discussed in this chapter. Also, a review of literature relevant to the work is provided, highlighting the current status of process monitoring.

Chapter 3: Extended Dynamic Approaches using Kernel Density Estimations

In this chapter, the description of the novel KDE approaches developed in this work is presented. Furthermore, the evaluation of these KDE approaches using the Tennessee Eastman Process Plant is illustrated. Also, the developed KDE approaches are compared with their non-KDE counter-parts to demonstrate their ability to

improve condition monitoring over their non-KDE counter-parts.

Chapter 4: State Space Independent Component Analysis

A novel State Space Independent Component Analysis (SSICA) is developed with the principles behind its development explained in this chapter. The proposed SSICA is applied to the TEP plant to evaluate the technique. It is also compared with the CVA and DICA methods, to demonstrate its ability to improve condition monitoring over both approaches.

Chapter 5: A Case Study

A Waste Water Treatment Plant is adopted as a case study in this chapter. The KDE approaches developed in this work are first applied to the Waste Water treatment Plant and then their monitoring performances compared with those of their non-KDE counter-parts. Furthermore, the proposed SSICA is also applied to the Waste Water Treatment Plant and its monitoring performance compared with those of the CVA and the DICA approaches. the monitoring performance of the SSICA is also compared with those of the KDE approaches.

Chapter 6: Conclusions

This final chapter presents the conclusions and outlines recommendations the future research.

Chapter 2

Literature Survey

This chapter presents an overview of process monitoring with an up to date review of literature relevant to this work. The history and the current status of process monitoring are discussed along with some existing process monitoring approaches.

2.1 Process Monitoring

In the process industry, there is an increasing demand to satisfy stringent safety and environmental regulations. Process operations that were considered acceptable at one time may no longer be adequate. To satisfy these stringent regulations, industrial processes are generally operated under closed-loop control. Generally, process controllers are designed to maintain satisfactory operations by compensating for the effect of disturbances and changes occurring in the process. However, there are changes in the process which these controllers cannot handle adequately. These changes can be detected early using process monitoring techniques.

Process monitoring is a means of detecting abnormal condition in a process. The

major goal of process monitoring is to detect abnormal condition early and diagnose the reason for the occurrence of such abnormal condition. These abnormal conditions are generally referred to as faults, which can occur due to changes in the process parameters, equipment failure, process noise and disturbances. Moreover, the presence of such faults in a process can affect the process operations and consequently, the process outputs. In extreme cases, these faults could result in fatal accidents involving the loss of lives. The recent Fukushima nuclear plant disaster due to a radiation leak is an example of such accidents that can occur when process operations go wrong. It is therefore, important to detect these faults early and diagnose the cause(s) of the faults in order to maintain safety during process operations. The importance of process monitoring is further emphasised by its applications in the chemical, manufacturing, pharmaceutical and food industries to mention a few.

Process monitoring was pioneered by Walter A. Shewart in the early 1920s and consists of fault detection, fault identification, fault diagnosis and process recovery [20]. Fault detection is the act of investigating the presence or occurence of a fault while fault identification is the identification of the variables that are most relevant to diagnosing the fault. Furthermore, fault diagnosis is the investigation of the reason for the occurence of the fault whereas process recovery also known as intervention is the elimination of the fault.

Process monitoring measures can be characterised as being analytical, knowledge-based or data-driven.

2.1.1   Analytical Methods

The analytical measures employ mathematical models often constructed from first principles [21, 22, 23, 24]. Based on the process input and output variables from the normal operating process, these analytical methods use detailed mathematical

models to generate features like the residuals ($\mathbf{r}$), parameter estimates ($\hat{\mathbf{p}}$) and state estimates ($\hat{\mathbf{x}}$). Faults are then detected by comparing the features estimated from the process observations with the features derived from the normal operating process [25]. The analytical methods that use residuals are commonly known as the analytical redundancy methods [21, 26]. These residuals are the difference between the plant observations and the mathematical models [21, 27]. The residuals will be zero in the absence of faults, disturbances, noise and/or modelling errors and will be non-zero if faults, disturbances, noise and/or modelling errors are present [20]. Three ways of generating the residuals on which the analytical redundancy methods are based are parameter estimation [28, 29], observer-based design [30, 31] and parity relations [21, 22].

- Parameter Estimation

  In this approach, the residuals are the difference between the nominal model parameters and the estimated model parameters [23, 24]. The model parameters can be estimated using standard parameter estimation techniques [32, 33]. This method is only appropriate if the process faults are associated with changes in the model parameters and appropriate mathematical models are available [21]. A threshold can be constructed based on the nominal model parameters to determine the presence of a fault while faults are detected based on the deviations in the nominal model parameters and the estimated model parameters. There is an indication of a fault if the changes in the estimated model parameters is greater than the estimated threshold.

- Observers

  The observer-based method is another way to generate residuals on which analytical methods are based. The observer is a device that estimates internal states using the system inputs and outputs and the model of a system. In this approach, the output of the system is estimated from the states or

a subset of the states [21, 34]. Thereafter, the difference between the measured plant output and the estimated plant output known as the estimation error [34] is employed as the vector of residuals [21, 30, 34]. Thresholds on the residuals of the output variables can then be determined to detect abnormal deviations [35]. However, the downside of this approach is that a detailed mathematical model is required [21].

- Parity Relations

  The parity relation is another popular method to generate residuals in which the residual is generated solely from the observations [21]. This method checks the consistency of the measurements with the mathematical equations of the system. Also, a linear dynamic transformation can be carried out on the parity relations so that the transformed residuals are appropriately employed for detceting faults in dynamic systems [36].

The analytical approaches have the advantage of incorporating physical understanding of the process into the process monitoring scheme. Nevertheless, the analytical measures are applied to systems with a relatively small number of process variables [21, 22]. This is because detailed models are required for the analytical approaches to be efficient. These detailed models are expensive to obtain for processes with a large number of variables. Hence, the use of analytical methods for most industrial plants can result in complex mathematical computations.

2.1.2   Knowledge-based Methods

An alternative method for process monitoring is the knowledge-based methods. These knowledge-based approaches use qualitative methods and are particularly suitable for systems without detailed mathematical models. Also, they are based on causal analysis [37, 38, 39], expert systems [40, 41, 42] and pattern recognition [43].

Causal analysis employs the concept of causal modelling of fault-symptom relationships whereas the expert systems are methods which imitate the reasoning behind human problem solving. Pattern recognition methods which include artificial neural network (ANN) and self organising maps, use the relationship between the data patterns and fault classes without explicitly modelling the internal states or structure. The ANN was developed in an attempt to mimic the computational structures of the human brain. It is a non-linear mapping between the input and output, consisting of interconnected neurons arranged in layers. One way to apply ANN for fault diagnosis is to assign the input neurons to process variables and the output neurons to fault indicators.

2.1.3   Data Driven Methods

Data driven methods of process monitoring are derived directly from the process measurements. These methods involve the use of statistical methods known as Statistical Process Monitoring (SPM). Originally SPM univariate methods were applied to monitor key product variables [44]. However, for most chemical and industrial processes, the process variables are not independent because correlations exist between the process variables. In such a situation, the univariate methods are inadequate since they do not consider the relations between the process variables. To address this limitation, the Multivariate Statistical Process Monitoring (MSPM) approaches also known as projection methods were introduced [45].

The efficiency of the data driven measures is highly dependent on the quantity and quality of the process data. This is not a problem for modern computer automated industrial processes where sophisticated data collection systems collect process information and make these process data readily available.

2.2 Multivariate Statistical Process Monitoring

The MSPM methods are statistically based approaches that were initially developed for applications in chemometrics [46] and then extended to process monitoring [47, 48, 49]. The basic idea of the MSPM approach is that a high dimensional space spanned by the process variables is projected onto an orthogonal space spanned by variables known as latent variables. These latent variables are linear combinations of the original variables. Generally, for most MSPM methods the first few latent variables are sufficient to describe the variation of the process and are employed to construct the monitoring model. Hence, the space spanned by these latent variables is called the model space. The remaining latent variables not included in the model space are assumed to be associated with noise and excluded to the residual space. The MSPM methods have the advantage of being able to handle large numbers of correlated process variables. Furthermore, the MSPM methods employ multivariate charts that use the Hotelling's T^2 statistic which is determined from the values of estimated latent variables in the model space. The T^2 statistic is a measure of the variation in the model space [50]. Several multivariate control charts based on the T^2 metric have been proposed [51, 52] and reviewed [53]. Another commonly used metric for MSPM methods is the Q metric also known as the Squared Prediction Error (SPE). The SPE is the measure of variation in the residual space and is also commonly employed for process monitoring [21, 54, 55].

The MSPM techniques are classified into the linear and non-linear methods. A relationship between two variables is said to be linear if a plot of their values on a co-ordinate system produces a straight line. The MSPM methods can also be classified as static and dynamic methods. The static methods assume the condition to be monitored is in steady state while the dynamic methods do not assume steady state and therefore account for the changes that occur in processes. These static and dynamic methods are discussed in the following sections.

2.2.1 Static Methods

Generally, static monitoring approaches assume that the observations are time independent, meaning that the past observation does not affect the present observation. These methods are suitable for steady state operations and include the Principal Component Analysis (PCA), Principal Component Regression (PCR) and Partial Least Squares (PLS) to mention a few.

2.2.1.1 Principal Component Analysis

Amongst the static MSPM techniques, the PCA is most widely reported [44, 56]. The PCA was introduced by Pearson [57] in 1901 and developed by Hotelling in 1933 [58]. Pearson described the PCA as an approach to find the closest fit of lines and planes to points in a space [57]. It was originally developed for chemical measurements in chemometrics to describe the relationship between variables [59, 60] before it was extended to process monitoring applications [44, 48, 61, 62, 63, 64]. The PCA decomposes the variation in a set of correlated process measurements to a set of de-correlated latent variables known as principal components (PCs). The PCA is scale dependent. Hence, it is important to scale or standardise the data to avoid measurements with large magnitudes overshadowing important measurements with small magnitudes. A graphical illustration of the PCA decomposition is presented in Figure 2.1 The PCA involves analysing the eigen structure of the covariance matrix of the data from the normal operating process for the estimation of the PCs. Moreover, the PCA is suitable for the analysis of steady-state data with uncorrelated measurements [1, 4, 65].

Consider a data matrix $\mathbf{X}$ with m observations and n variables. The PCA transformation can be presented as

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathbf{E} \tag{2.1}$$

19
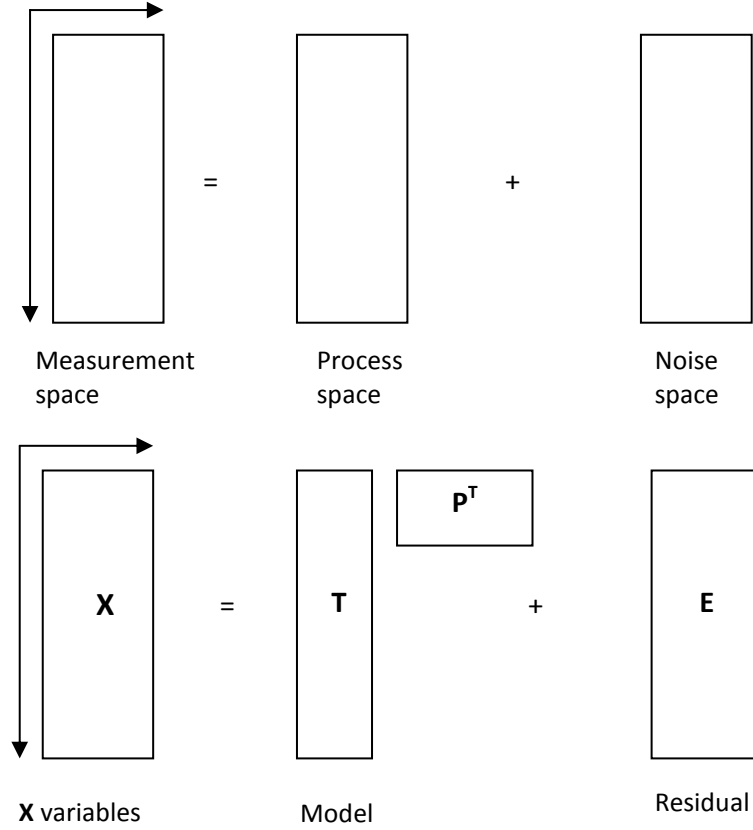
Figure 2.1: Decomposition of $\mathbf{X}$

where $\mathbf{T}$ is the score matrix, $\mathbf{P}$ is the loading matrix and $\mathbf{E}$ is the residual matrix [44, 65]. The loading vectors, i.e., the columns of matrix $\mathbf{P}$ are actually the eigen vectors of the covariance of the data matrix $\mathbf{X}$. Note, $\mathbf{X}$ has zero mean. The covariance matrix and its eigen value decomposition are expressed below;

$$\mathbf{S} = \frac{1}{m-1}\mathbf{X}^T\mathbf{X} = \mathbf{U}\mathbf{V}\mathbf{U}^T \qquad (2.2)$$

where m is the number of observations, $\mathbf{U}$ is the eigen-vector matrix and $\mathbf{V}$ is the eigen-value matrix. The principal components are the projections of the original variables along the directions determined by the eigen-vectors. In order to optimally capture the variations in the data, only the first a eigen-vectors $(p_1, p_2, ..., p_a)$ corresponding to the first a largest eigen-values of the covariance of $\mathbf{X}$, where $(a < n)$ are retained in the PCA models. The remaining $(n - a)$ eigen-vectors are excluded

to the residual space. From the eigen-vectors in the model and residual spaces, principal components can be derived for the model and residual spaces respectively. The retained a PCs define the subspace with the greatest variability, which could be described as the model space. The rest of the PCs not in the model space are considered to be associated with noise and are excluded to the residual space. By excluding the PCs that do not contribute significantly to the overall variation to the residual space, the dimensionality of the process data is correspondingly reduced. The PCA reconstructs the data matrix from the de-correlated PCs as

$$\widehat{\mathbf{X}} = \widehat{\mathbf{T}}\widehat{\mathbf{P}}^T = \mathbf{X}\widehat{\mathbf{P}}\widehat{\mathbf{P}}^T \tag{2.3}$$

where $\widehat{\mathbf{T}}$ and $\widehat{\mathbf{P}}$ are the score and loading matrices employed to estimate the retained a PCs in the model space.

The variations in the model space can be determined using the Hotelling's T^2 statistic which is defined as

$$T^2 = \sum_{i=1}^{a} \mathbf{t}_i{}^T \sigma_i{}^{-2} \mathbf{t}_i \tag{2.4}$$

where $\sigma_i{}^2$ is the estimated variance of the ith principal component, $\mathbf{t}_i$. However, the T^2 will only detect variations in the model space but is unable to detect the variations in the residual space. Hence, monitoring processes using the T^2 metric alone may not be sufficient. The variations in the residual space can be detected using the Squared Prediction Error (SPE) also known as the Q metric which is defined as;

$$Q = \sum_{i=1}^{n} (\mathbf{x}_{new,i} - \widehat{\mathbf{x}}_{new,i})^2 \tag{2.5}$$

where $\mathbf{x}_{new,i}$ is the ith measurement in the new data to be monitored and $\widehat{\mathbf{x}}_{new,i}$ is the estimated value of $\mathbf{x}_{new,i}$.

Furthermore, from the estimated T^2 and Q values, the control limits for the T^2 met-

ric [66] and that for the Q metric [67] can be determined. For each observation, the process is considered normal if the T^2 value is less than or equal to its upper control limit and the Q value is less than or equal to its upper control limit. On the other hand, when either the T^2 value or the Q value is greater than their corresponding control limits, it is considered as an indication of a fault.

MacGregor and Kourti [44] established a PCA model from normal operation data and judged the behaviour of online processes against the PCA model in order to detect deviations from the normal operating process. In their study, they applied the PCA to a mineral processing plant, a continuous polymerisation plant and an industrial batch polymerisation reactor, illustrating the efficiency of the PCA for fault detection and diagnosis.

For most industrial processes, a large number of process variables are involved and more than three PCs are often required to capture most of the variance. For such a case, generally, there is a difficulty to represent the operating envelope. To address this problem, Wang et al. [62] presented a PCA approach based on visualization using parallel co-ordinates, transforming the Euclidean space into parallel coordinates. They proposed a method of visualising multiple PCs with the aim of displaying the operative envelopes when three or more PCs are involved. Their approach was applied to the Manresa Waste Water Treatment Plant (WWTP) to demonstrate its efficiency. Also, Wang and Cui [68] developed a strategy for fault detection and diagnosis based on the PCA to efficiently detect and diagnose sensor faults in centrifugal chillers.

Furthermore, multi-way PCA (MPCA) has been proposed for monitoring batch and semi-batch processes [69, 70]. The MPCA is an extension of the PCA to handle data in three-dimensional arrays. The three dimensions arise from batch trajectories that consist of batch runs (**I**), process variables (**J**) and time (**K**). The data from the batch processes are organised into an array **X** of dimension (**I** x **J** x **K**). The basic idea

of the MPCA is to unfold the three-dimensional batch data into a two-dimensional data before applying PCA to the two-dimensional data. The unfolding of the three dimensional data is illustrated in Figure 2.2. MPCA decomposes the array $\mathbf{X}$ into
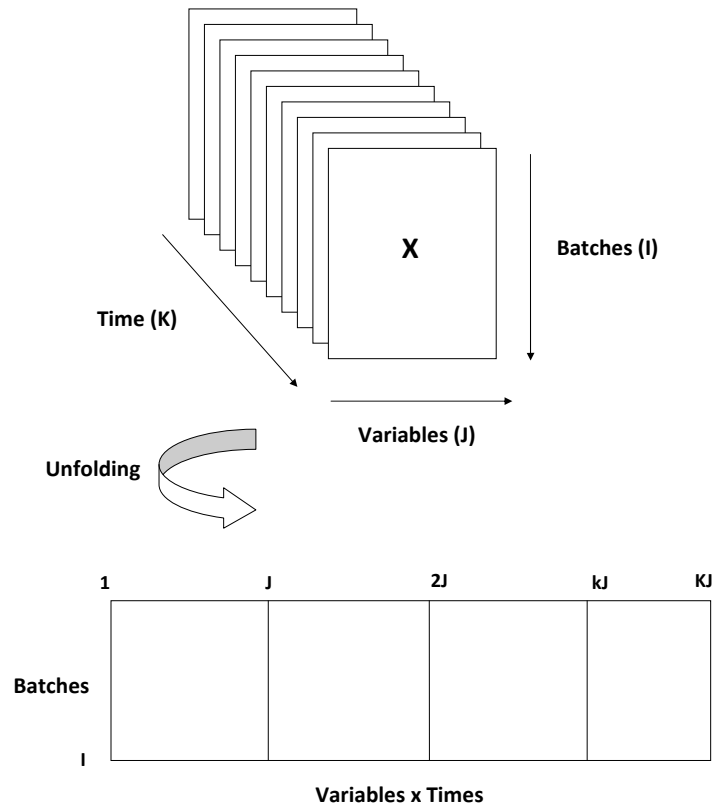
Figure 2.2: Decomposition of a three-dimensional batch data matrix $\mathbf{X}$

the score and loading matrices in a similar way to the static PCA. Kosanovich et al. [71] applied MPCA to an industrial batch process.

More recently, rather than using the Hotelling's T^2 and Q metrics, the use of a unified monitoring statistic as opposed the T^2 metric and the Q metric has been developed. The T^2 metric and the Q metric have been combined algorithmically [72, 73]. Also, a probabilistic model was proposed to provide a single likelihood-based statistic [74, 75, 76, 77]. A key advantage of these unified approaches is increased

sensitivity.

Although the PCA is flexible with an expanding role for process monitoring, standard PCA has major short comings. In the development of the PCA model, the following assumptions are made;

- The observations are time independent

- The observations follow a Gaussian distribution

Unfortunately, the assumption of time-independence made by the PCA may not be valid for processes subject to dynamic disturbances. Consequently, the employment of the PCA to monitor such dynamic processes will not be able to fully describe the entire process variation [9]. For this reason, the PCA is not suitable for dynamic processes [1, 4, 78, 79]. Furthermore, the assumption of normality may also be invalid for most chemical processes where strong non-linearity makes variables driven by noise and disturbances strongly non-Gaussian [4, 5].

2.2.1.2 Principal Component Regression

The PCR is another static linear MSPM method which is a simple extension of the PCA [18, 80]. The PCR is a regression analysis that uses the PCA for the esimation of the score matrix before employing the estimated score matrix to determine the regression coefficient. The first step of the PCR is the PCA decomposition shown in Equation (2.1) while the second step of the PCR is shown as

$$\mathbf{Y} = \mathbf{TB} \tag{2.6}$$

where $\mathbf{Y}$ is the output matrix and $\mathbf{B}$ is the regression matrix. Using the regression matrix ($\mathbf{B}$) and the score matrix ($\mathbf{T}$) from the PCA decomposition, $\mathbf{Y}$ can be predicted. Also, the accuracy of the predictions depend on the number of latent

variables retained in the model space which is determined by the number of PCs that give the best prediction of $\mathbf{Y}$. Process monitoring using PCR follows the same procedure as for PCA. Also, the decomposition of the data matrix ($\mathbf{X}$) in the PCR is done to maximise the captured variability. However, this is generally not optimal for prediction purposes. The PLS on the other hand is a better tool for such predictions and is discussed in the following section.

2.2.1.3   Partial Least Squares Regression

The PLS was developed by Herman Wold in 1966 [81] for econometrics although it later gained its popularity in chemometrics [82] and then in process monitoring [47, 83, 84]. The PLS is a robust multivariate regression algorithm based on the PCA approach of decomposing data matrices into latent variables. However, whilst the PCA decomposes a solitary data block into PCs, the PLS involves the decomposition of the independent variable ($\mathbf{X}$) and the dependent variable ($\mathbf{Y}$). The objective of the PLS is to model $\mathbf{X}$ in a way that the information in $\mathbf{Y}$ can be accurately predicted. The PLS achieves this by maximising the correlation between $\mathbf{X}$ and $\mathbf{Y}$ so that a linear relationship can be developed between two sets of abstract variables. The PLS decomposition of the $\mathbf{X}$ and $\mathbf{Y}$ matrices is shown below

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathbf{E} \tag{2.7}$$

$$\mathbf{Y} = \mathbf{U}\mathbf{Q}^T + \mathbf{F} \tag{2.8}$$

where $\mathbf{T}$ and $\mathbf{P}$ are the score and loading matrices for the independent variable ($\mathbf{X}$) while $\mathbf{U}$ and $\mathbf{Q}$ are the score and loading matrices for the dependent variable ($\mathbf{Y}$). The matrices $\mathbf{E}$ and $\mathbf{F}$ are the residuals for the independent and dependent variables respectively. Like the PCA, the PLS requires the data to be scaled to avoid super imposition of certain variables on other variables with relatively smaller

25

values. Apart from decomposing the **X** and **Y** blocks, the PLS also consists of connecting the latent variables that are extracted from the decomposition of the **X** and **Y** blocks. This is the regression step and is shown in (2.9).

$$\mathbf{U} = \mathbf{TB} \qquad\qquad (2.9)$$

where **B** is the regression matrix which describes the inner relation between **T** and **U**. In summary, the PLS can be described as a combination of an outer and inner model. The outer model is concered with the decomposition of **X** and **Y** blocks, while the inner model connects the latent variables that are extracted from the decomposition of **X** and **Y**. A common application of the PLS is to select the matrix (**Y**) to consist of the output variables while the matrix (**X**) consists of the other process variables [21, 44, 49, 85]. As in the PCR method, the choice of the number of latent variables to retain for process monitoring is important because this affects the accuracy of the predictions. Generally, cross validation is the tool to determine the appropriate number of latent variables to retain in the model space. The PLS can be carried out using the Non-linear Iterative Partial Least Squares (NIPALS) algorithm.

NIPALS is an algorithm for developing latent variables for the PCA or PLS. The development of the NIPALS algorithm was initiated by H. Wold [81] and later extended by S. Wold [86]. NIPALS starts with some guessed starting vector so that the estimated latent variables depend on the guessed starting vector. There are two types of NIPALS methods to model the predicted block; the PLS1 and PLS2. Both are very similar except that in PLS1 each predicted variable is modelled separately and the PLS model is built from sequencially calculated dimensions while in PLS2 all the predicted variables are modelled simultaneously [21]. Generally, the results from PLS1 and PLS2 are different particularly when there are several output variables [87]. The PLS2 is recommended for multivariate processes given the presence

of autocorrelations and cross correlations between the process measurements. In practice, the NIPALS algorithm gives more accurate results than the singular value decomposition (SVD) of the covariance matrix [87]. However, the NIPALS algorithm takes a longer time to calculate than the SVD covariance. The PLS has the following advantages;

- Modelling multiple predictors($\mathbf{X}$) and responses($\mathbf{Y}$) [21].

- Handling multicollinearity in the predictors($\mathbf{X}$) [88].

- It is robust when noise is present in the data [89].

- Accurate and robust where high levels of correlations exist [88, 90].

Process monitoring using the PLS follows the same procedure as for the PCA and PCR approaches. The PLS has been applied to process monitoring [47, 83, 84] and reported as an efficient process monitoring tool [84]. The Hotelling's T^2 and Q metrics have been recommended for monitoring the PLS score and residual spaces respectively [49]. Also, the unified monitoring approaches could be employed with the PLS.

Hoskuldsson [90] reformulated the PLS as an eigenvalue/eigenvector problem, different from the formulation illustrated above [81]. Generally, the first latent variable is the linear combination of $\mathbf{X}$ that will maximise the covariance between $\mathbf{X}$ and $\mathbf{Y}$. Also, the first PLS loading vector based on the reformulated PLS by Hoskuldsson [90] is the first eigen vector of the sample covariance matrix $\mathbf{X}^T\mathbf{Y}\mathbf{Y}^T\mathbf{X}$. A likelihood ratio test to determine the number of latent variables to retain in the model space was then established. He also showed the structure of the PLS decompositions of $\mathbf{X}$ and $\mathbf{Y}$ and the statistical aspects when it is used for model building.

Geladi and Kowalski [89] gave practical examples of the application of the PLS algorithm with simulated data, illustrating interesting properties of the PLS algo-

rithm by testing models and predictions to give a better understanding of the PLS technique. They chose the PLS1 method, using the residuals after each dimension for the estimation of the latent variables. The aim of their study was to present a worked example and then investigate the influence of random noise, non-linearities and interfering extra components on the process.

Lennox [91] proposed a novel application of the PLS for fault detection and diagnosis using the Tennessee Eastman Process (TEP) Plant. He demonstrated the use of the inner structure of the PLS model to provide information regarding the process and employed the provided information for fault detection and diagnosis. Also, his PLS algorithm was associated with three statistics; T^2, SPE_x and SPE_y. SPE_x is the square of the residual for the independent variables, $\mathbf{x}$ while SPE_y is the square of the residual for the dependent variables, $\mathbf{y}$. Besides, his PLS algorithm was reported to identify 12 of the 20 TEP faults suggested by Downs and Vogel [92].

Generally, the development of monitoring methods is geared towards applying these methods to industrial processes. The methods discussed above are extended in various ways to more efficiently monitor these industrial processes. Some extensions of the static methods are discussed below.

2.2.2   Dynamic Methods

The static methods in their development assume that the observations are time independent, which means that each observation is not dependent on the previous observation. However, for most chemical processes, this assumption of time independence is invalid due to the dynamic nature of these processes. Consequently, applying a static method to model dynamic processes will only reveal a linear static approximation but does not effectively characterise process dynamics. To address this limitation, lagged variables are introduced to model the dynamic relations [1, 2].

The dynamic extensions of the linear methods are essentially the same as the static linear methods except that in the extended dynamic approaches, the data matrix is composed of time shifted vectors [1, 2, 78]. Moreover, these dynamic extensions are reported to be more efficient than the static approaches for dynamic process monitoring [1, 2]. The dynamic extensions of the PCA and PLS approaches are discussed in the following sections.

2.2.2.1 Dynamic Principal Component Analysis

To address the limitation of the PCA due to the invalidity of the assumption of time independence already discussed in section 2.2.1.1, the Dynamic Principal Component Analysis (DPCA) was proposed [1]. The DPCA is an extension of the PCA for dynamic processes. In this extended PCA approach, the auto-correlations as well as the cross correlations between the process variables are taken into account by augmenting each observation vector with the previous l observations and stacking the data matrix as shown below

$$\mathbf{X}_A = \begin{bmatrix} \mathbf{x}_t{}^T & \mathbf{x}_{t-1}{}^T & \cdots & \mathbf{x}_{t-l}{}^T \\ \mathbf{x}_{t-1}{}^T & \mathbf{x}_{t-2}{}^T & \cdots & \mathbf{x}_{t-l-1}{}^T \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_{t+l-m}{}^T & \mathbf{x}_{t+l-m-1}{}^T & \cdots & \mathbf{x}_{t-m}{}^T \end{bmatrix} \tag{2.10}$$

where $\mathbf{X}_A$ is the augmented data matrix and $\mathbf{x}_t$ is an n-dimensional observation vector at time instance t. The DPCA approach consists of applying the PCA on the data matrix defined in Equation (2.10) Moreover, the DPCA is reported to be more efficient than static PCA for monitoring dynamic processes [1, 78, 79].

Ku et al. [1] proposed the DPCA, where they applied PCA on lagged variables using the well known time lag shift that is applied in system identification [4]. They augmented each observation vector with the previous observations, constructing a

data matrix in the form of the Hankel matrix illustrated in Equation (2.10) before applying the static PCA to the constructed data matrix for the estimation of the PCs. Their goal was to develop and utilize an easy to use DPCA model for the construction of multivariate charts to monitor the model and residual spaces of a dynamic process more efficiently. The PCs from the DPCA are employed in a similar way as those from the PCA for process monitoring in the model and residual spaces. Ku et al. [1] also examined the use of the DPCA for fault detection and fault isolation. Their DPCA model was reported to outperform the static PCA in detecting disturbances quicker.

Also, Lee et al. [19] proposed a method of dynamic sensor fault detection and identification using a DPCA based variable reconstruction for a Waste Water Treatment Plant (WWTP) [93]. Their technique consists of building the DPCA model [1] and then deriving the SVI (Sensor Validity Index) [64] of each measured variable using the construction method. The derived SVI was able to identify the abnormal sensors and their DPCA based approach was reported to be more efficient than the static PCA based approach.

In addition, Mina and Verde [78] developed a DPCA model in a similar way to Ku et al. [1] except that in their work, they also identified a set of nominal input-output relations for the purpose of diagnosis. They estimated the actual means of the input variables using exponentially weighted moving average (EWMA) and then estimated the means of the output variables using the input means. The aim of their study was to develop a DPCA model identifying the nominal input-output relationships. Their algorithm was applied to an interconnected three-tank system based on the Hotelling's T^2 metric. Their proposed algorithm was reported to be able to efficiently distinguish between normal changes in signals and deviations due to the occurence of faults.

Tsung [79] proposed an integrated approach using the DPCA [1] and the minimax

distance classifier which is an engineering model to simultaneously monitor and diagnose an automatic controlled process. Different from Ku et al. [1] who determined the number of PCs to retain and the order of the system iteratively, Tsung [79] determined the number of the PCs to retain and the order of the system from the analytical model. Furthermore, the control limits associated with his DPCA model was based on the Hotelling's T^2 and Q metrics. For the purpose of diagnosis, the author identified and isolated the root cause using the minimax distance classifier. Moreover, this approach was applied to an automatic machining process to demonstrate its efficiency and applicability.

Different from the DPCA approaches mentioned above, Srinivasan et al. [94] suggests building a dynamic model for the PCs obtained from the static PCA technique. Unfortunately, building a dynamic model for the PCs obtained from the static PCA will not be sufficient because the residual space will still remain static. Also, the employment of such a static residual space for the monitoring of dynamic process could lead to inaccurate conclusions irrespective of the reported success associated with process condition monitoring based on the residual space [1, 21, 67].

2.2.2.2 Dynamic Partial Least Squares

The PLS like the PCA is a static model that assumes that the observations are time independent. This assumption may not be valid for most industrial processes due to their dynamic nature, making the static PLS inappropriate and insufficient for monitoring dynamic processes.

To address the limitation of the static PLS for dynamic processes, Komulainen et al. [2] proposed the Dynamic PLS (DPLS) approach based on the Hotelling's T^2 metric. In the DPLS approach, the auto and cross correlations between the process variables are taken into account by incorporating time lags of the time series before

applying the PLS. The time lags are incorporated by forming an augmented data matrix as in Equation (2.10). The objective of their study was to develop an online monitoring system for an industrial dearomatization process. A fault in the analysers is a common disturbance in the dearomatization processes. In their work, data was collected over a period of time to reflect the most frequently occurring faults. Also, some computed variables were created, that captured the characteristics of the dearomatization process. These computed variables were constructed from lagged process measurements of the dearomatization process and then combined with the process variables on the basis of the correlations between the 2 set of variables. Thereafter, the PLS was applied to the combined variables and the different models tested before selecting the best mode. The performance of the DPLS with and without the computed variables was compared and it was reported that the DPLS with the computed variables performed better than the DPLS without the computed variables. In addition to being more efficient than the PLS, their approach correctly classified changes in process parameters as normal states and gave an alarm for an abnormal process state during the disturbance.

Lennox [91] extended the PLS to dynamic systems using the finite impulse response (FIR) and auto-regressive with exogenous (ARX) models [1, 2] for fault detection, diagnosis and isolation. His dynamic PLS aproach was applied to the TEP plant. Similarly, Lee et al. [95] extended the PLS using the FIR and ARX [1, 2] inputs to model a full-scale wastewater treatment plant in Korea. Their model was reported to give an impressive prediction performance.

Dynamic models are reported to be more suited to continuous processes [2]. Nevertheless, Fletcher et al. [96] adopted dynamic PLS (DPLS) for multi-way batch modelling while Chen and Liu [97] developed a dynamic model of the PLS known as the Batch DPLS (BDPLS) for on-line batch monitoring. The efficiency of both approaches was also reported [96, 97].

Although the DPCA and DPLS approaches discussed above are reported to improve the condition monitoring over the static PCA and PLS approaches [2, 91, 95], the latent variables extracted from the DPCA and DPLS approaches are not necessarily the minimal dynamic representations [3, 4, 9]. On the other hand, state space techniques like the Canonical Variate Analysis (CVA) have been reported to be the best solution for dynamic processes [3, 4, 5, 6, 7, 98]. The CVA is discussed in the following section.

2.2.2.3   Canonical Variate Analysis

Canonical Variate Analysis (CVA) is a linear dynamic dimension reduction technique to estimate the minimum number of state variables for dynamic process monitoring. CVA was first introduced by Hotelling [9, 99], adopted for use in dynamic systems for a limited class of processes by Akaike [9, 100] and adapted to general linear systems by Larimore [10, 100]. A basic concept introduced by Larimore in the CVA is the past and future vectors of the process. For an observation at time k of inputs and outputs, the past consists of past inputs and outputs occurring prior to time k while the future consists of future inputs and outputs occurring after time k. The past vector is defined in the same manner as the past matrix for autoregressive with exogeneous variables modelling [4] while the future includes process values at time k as well as the future values occurring after time k. Assume $\mathbf{y}_k \in \Re^{mq}$, for $k = (1, 2, \cdots, n)$ are process data collected under normal operating condition, the past $(\mathbf{y_{p,k}})$ and future $(\mathbf{y_{f,k}})$ vectors are defined as

$$
\mathbf{y}_{p,k} = \begin{bmatrix} \mathbf{y}_{k-1} \\ \mathbf{y}_{k-2} \\ \vdots \\ \mathbf{y}_{k-q} \end{bmatrix} \in \Re^{mq} \tag{2.11}
$$

33

$$\mathbf{y}_{f,k} = \begin{bmatrix} \mathbf{y}_k \\ \mathbf{y}_{k+1} \\ \vdots \\ \mathbf{y}_{k+q-1} \end{bmatrix} \in \Re^{mq} \qquad (2.12)$$

where mq is the length of the past and future vectors. The objective of CVA for process monitoring is to derive a low-dimensional representation of the process data that most accurately highlights the differences that exist between the normal operating data and the process data to be monitored. The CVA achieves this by determining the past that has the most information for the prediction of the future and then models a process by successively approximating the canonical variates. These canonical variates are linear combinations of the past and future vectors. Like most MSPM techniques, the first few canonical variates are adequate to describe the process bahaviour and are retained in the model space, while the remaining canonical variates are excluded to the residual space. The canonical variates in the model space are known as the state variables and can be employed for process monitoring [3, 7, 101]. The canonical variates in the residual space are also useful for process monitoring [12]. Moreover, the superiorioty of the CVA over the DPCA and DPLS techniques have been reported and demonstrated [4, 7, 101]. A detailed description of the CVA is presented in chapter 3.

Larimore [99] adopted the CVA to address the problem of modelling with a restricted order state space. Using the CVA, he optimally chose a number of linear combinations of the past for the prediction of the future. He further proposed the use of the Akaike Information Criterion (AIC) for the selection of the length of the past vector for CVA modelling and suggested the development of an ARX model for various model orders. The ARX model that minimised AIC defined the optimal length of the past vector. In addition, his CVA computation was based on SVD.

Also Negiz and Cinar [6] presented a study on CVA. Different from the study of Larimore [99], they determined the length of the past window to be the maximum significant lag after which the autocorrelation of the autoregressive model (AR) residuals become statistically insignificant. In addition to the CVA, they employed the PLS to develop another state space model and then compared the efficiencies of monitoring using these two state space models. From their findings, the CVA offerred a more robust tool for system identification in state space than the PLS approach.

Norvalis et al. [9] developed a tool for monitoring and diagnosis that combined canonical variate state space (CVSS) models with knowledge based systems (KBS). Faults were detected using the CVSS models while diagnosis was based on the KBS. Their technique was based on the Hotelling's T^2 charts for fault detection and was applied to a polymerisation reactor system. Also, contribution plots were employed to determine the process variables that contributed the most to the out-of-control T^2 values [44]. The contribution of the process variables from past to the current state variables was taken into account due to the dynamic nature of the plant considered in their work. Moreover, the efficiency of their technique was illustrated.

Juan and Fei [12] adopted the statistical CVA method based on Hotelling's T^2 charts and applied it to a chemical seperation process plant for fault detection. Their technique focussed on canonical correlations using the past process outputs only. In their CVA approach, the future data was predicted while the past data was reconstructed. The results from their study illustrated a good performance of the CVA model, showing that their technique was efficient in the presence of strong autocorrelations and cross correlations. It was also demonstrated that the precision of the CVA model improved with an increase in the length of the data employed for the CVA analysis.

Chiang et al. [21] different from the studies of Norvalis et al. [9] and Juan and

Fei [12], employed CVA to include the input and output variables for the estimation of the state variable. Their CVA model was based on SVD and associated with the Hotelling's T^2 and Q metrics. Also, their CVA approach was applied to the Tenneessee Eastman Process (TEP) plant and the results from their approach compared with those of the PCA and DPCA techniques also considered in their work. Based on their work, the CVA clearly outperformed the PCA and DPCA techniques. The superiority of their CVA approach over the PCA could be attributed to the fact that the CVA takes the auto-correlations and cross correlations into account whereas the static PCA does not. Also, the superiority of their CVA approach over the DPCA could be atttributed to the fact that the CVA is a more appropriate way to capture the dynamic behavior of plants than the DPCA [4, 10].

Simoglou et al. [10] presented a study in which they employed CVA and PLS to determine and compare the states. In their CVA approach, they employed the past vector consisting of inputs and outputs to determine the canonical variates [4, 21]. Different from all the works on CVA mentioned above, in the work of Simoglou et al. [10], each input and output included in the past vector has a different number of past values (lag). They demonstrated that building models where the number of past inputs and outputs were optimised for each model enhanced process monitoring performance.

The CVA is similar to the PLS in concept in that the CVA estimates linear combinations of the past values of the systems inputs and/or outputs that are correlated the most with linear combinations of the future values of the output of the process [10]. However, a comparison of the CVA based approaches and the PLS based approaches showed that the CVA based approaches are able to detect more fault and more rapidly than the PLS based approaches [4, 10].

2.2.3 Adaptive Methods

Most real-time industrial processes display a non-stationary behaviour. Process operating conditions can change due to a shift in operating point, short term disturbances or variations in parameters. The static approaches could interprete such changes as faults because these static methods rely on the assumption that the processes are stationary. Also, most industrial processes are time varying and static methods are not adequate for such time-varying processes. One way to address the time-varying and non-stationary problem is to use an adaptive model that does not require switching or tuning during the process changes [102, 103, 104, 105]. The initial model of the process can be updated to accommodate the time varying behaviour while detecing abnormal variations in the process. Wold [102] developed an exponentially weighted moving average (EWMA) approach for the PCA and the PLS. In the EWMA models, the more recent observations receive larger weighting than earlier ones. To avoid unwarranted adaptation of the PCA/PLS model, he proposed an approach in which the older PCA/PLS models are employed to determine the updated PCA/PLS models. Nevertheless, for the time-varying processes, the older process models may not accurately represent current process variable relationship so that utilising the older models to update the PCA/PLS may not be very accurate [104].

In addition, the recursive PCA (RPCA) [102, 104, 106] and recursive PLS (RPLS) [107, 108] techniques are adaptive extensions of the PCA and PLS to solve the time-varying and non-stationary problems using recursive means. The RPCA updates a PCA model as new observations become available while the RPLS method updates the PLS model when new observations are available. The recursive model can be regarded as a linearisation of the system at the current operational point [18]. Qin [103] integrated a moving window approach into the Recursive PLS. In his approach a PLS model was identified on the basis of a data set that is within a selected

window. However, the concern with his approach is that results from monitoring could be different for different window sizes [108].

2.2.4   Methods for Non-linear Systems

Most monitoring methods like the PCA are generally based on the assumption that the processes are linear. However, such methods are not appropriate for handling industrial problems which exhibit non-linear behaviour [109, 110, 111]. It is therefore necessary to extend the PCA to such non-linear processes in order to appropriately monitor such non-linear systems. One way to address this problem is to employ non-linear pre-treatment of the data if the relationship between the variables is known to be non-linear [18]. This involves the use of the square or logarithmic value of the variables [112]. Also the non-linear variants of the PCA have been developed and employed for non-linear process monitoring [113, 114, 115]. Kramer [113] proposed a non-linear PCA approach based on auto associative neural networks. However, the network proposed by Kramer is difficult to train because it has five layers [116]. Besides, it is also difficult to determine the number of nodes in each layer. Dong and McAvoy [111] proposed another non-linear PCA approach based on principal curves and neural networks while Jia et al. [117] proposed a non-linear PCA method based on an input-training neural network. Generally, these non-linear methods describe the relation between original variables and the scores using non-linear functions identified by a neural network as shown below

$$\mathbf{X} = \mathbf{F}(\mathbf{T}) + \mathbf{E} \tag{2.13}$$

where $\mathbf{T}$ is the matrix of non-linear PCs, $\mathbf{F}$ is the non-linear function equivalent to the loadings in linear PCA and $\mathbf{E}$ is the residual matrix. In recent years, a new non-linear PCA approach known as the Kernel PCA (KPCA) has been developed

to tackle the non-linear problem [115, 118, 119].

2.2.4.1   Kernel Principal Component Analysis

The Kernal PCA consists of mapping measurements from their original space into a higher dimensional feature space via non-linear mapping before computing the PCs in the feature space. The Kernel PCA was suggested by Scholkopf [115, 120]. The major advantage of the KPCA over other non-linear PCA approaches mentioned above is that it does not involve non-linear optimisation [115]. It requires only linear algebra, making it as simple as the standard PCA. Also, due to its ability to use different kernels, the KPCA can handle a wide range of non-linearities. Assume a training data set $\mathbf{X} \in \Re^{m \times n}$, each of the observations $\mathbf{x}_i$, $i = 1, 2, \cdots, m$ is an n-dimensional vector and can be mapped into an h dimensional feature space using a mapping function $\boldsymbol{\phi}_i = \boldsymbol{\Phi}(\mathbf{x}_i)$. The training data in the feature space can then be represented as

$\boldsymbol{\chi} = [\boldsymbol{\phi}_1 \ \boldsymbol{\phi}_2 \ \cdots \boldsymbol{\phi}_m]^T$. Note that $\boldsymbol{\chi}$ has zero mean. The sample covariance $(\mathbf{C})$ of the data set in the feature space can be estimated as;

$$(m - 1)\mathbf{C} = \boldsymbol{\chi}^T \boldsymbol{\chi} = \sum_{i=1}^{m} \boldsymbol{\phi}_i \boldsymbol{\phi}_i^T \qquad (2.14)$$

The KPCA in the feature space is equivalent to solving the following eigen-vector equation

$$\boldsymbol{\chi}^T \boldsymbol{\chi} \boldsymbol{\nu} = \sum_{i=1}^{m} \boldsymbol{\phi}_i \boldsymbol{\phi}_i^T \boldsymbol{\nu} = \lambda \boldsymbol{\nu} \qquad (2.15)$$

Note that $\boldsymbol{\phi}_i$ is not explicitly defined. The kernel trick which premultiplies Equation (2.15) by $\boldsymbol{\chi}$ is then applied below

$$\boldsymbol{\chi} \boldsymbol{\chi}^T \boldsymbol{\chi} \boldsymbol{\nu} = \lambda \boldsymbol{\chi} \boldsymbol{\nu} \qquad (2.16)$$

Defining

$$\mathbf{K} = \boldsymbol{\chi}\boldsymbol{\chi}^T = \begin{bmatrix} \boldsymbol{\phi_1}^T\boldsymbol{\phi_1} & \cdots & \boldsymbol{\phi_1}^T\boldsymbol{\phi_m} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\phi_m}^T\boldsymbol{\phi_1} & \cdots & \boldsymbol{\phi_m}^T\boldsymbol{\phi_m} \end{bmatrix} = \begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \cdots & k(\mathbf{x}_1, \mathbf{x}_m) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_m, \mathbf{x}_1) & \cdots & k(\mathbf{x}_m, \mathbf{x}_m) \end{bmatrix} \quad (2.17)$$

denoting

$$\boldsymbol{\alpha} = \boldsymbol{\chi}\boldsymbol{\nu} \quad (2.18)$$

we have

$$\mathbf{K}\boldsymbol{\alpha} = \lambda\boldsymbol{\alpha} \quad (2.19)$$

where $\mathbf{K}$ is the kernel matrix while $\boldsymbol{\alpha}$ and λ are an eigen-vector and eigen-value of $\mathbf{K}$ respectively. To solve for $\boldsymbol{\nu}$ in Equation (2.18), it is multiplied by $\boldsymbol{\chi}^T$

$$\boldsymbol{\chi}^T\boldsymbol{\alpha} = \boldsymbol{\chi}^T\boldsymbol{\chi}\boldsymbol{\nu} = \lambda\boldsymbol{\nu} \quad (2.20)$$

this shows that $\boldsymbol{\nu}$ is given by

$$\boldsymbol{\nu} = \lambda^{-1}\boldsymbol{\chi}^T\boldsymbol{\alpha} \quad (2.21)$$

In summary, to estimate the KPCA model, the eigen decomposition is first performed as in Equation (2.19) to obtain λ_i and $\boldsymbol{\alpha}_i$ before deriving $\boldsymbol{\nu}_i$ as in Equation (2.21). From (2.21), $\boldsymbol{\nu}_i^T\boldsymbol{\nu}_i = \lambda^{-1}$, however, using a normalised $\boldsymbol{\nu}$ ($\widetilde{\boldsymbol{\nu}}$),

$$\widetilde{\boldsymbol{\nu}}^T\widetilde{\boldsymbol{\nu}} = 1 \quad \text{and} \quad \widetilde{\boldsymbol{\nu}} = \sqrt{\lambda}\boldsymbol{\nu}$$

The matrix of the l leading eigen-vectors are the KPCA principal loadings in the feature space denoted as $\mathbf{V} = [\widetilde{\boldsymbol{\nu}_1} \ \widetilde{\boldsymbol{\nu}_2} \ \cdots \widetilde{\boldsymbol{\nu}_l}]$. From Equation (2.21), $\mathbf{V}$ is related

to the loading in the measurement space as

$$\mathbf{V} = \left[\lambda_1^{1/2} \boldsymbol{\nu}_1 \ \cdots \ \lambda_l^{1/2} \boldsymbol{\nu}_l \right]$$
$$= [\lambda_1^{1/2} \boldsymbol{\chi}^T \boldsymbol{\alpha}_1 \lambda_1^{-1} \ \cdots \ \lambda_l^{1/2} \boldsymbol{\chi}^T \boldsymbol{\alpha}_l \lambda_l^{-1}]$$
$$= [\lambda_1^{-1/2} \boldsymbol{\chi}^T \boldsymbol{\alpha}_1 \ \cdots \ \lambda_l^{-1/2} \boldsymbol{\chi}^T \boldsymbol{\alpha}_l]$$
$$= \boldsymbol{\Lambda}^{-1/2} \boldsymbol{\chi}^T \mathbf{P} \tag{2.22}$$

where $\mathbf{P} = [\boldsymbol{\alpha}_1 \ \cdots \ \boldsymbol{\alpha}_l]$ and $\boldsymbol{\Lambda} = \text{diag}[\lambda_1 \ \cdots \ \lambda_l]$ are the l eigen-vectors and eigen-values of $\mathbf{K}$, corresponding to the l largest eigen-values. For a given measurement $\mathbf{x}_k$ and its mapped vector $\boldsymbol{\phi}_k = \boldsymbol{\Phi}(\mathbf{x}_k)$, the principal components are estimated as

$$\mathbf{t}_k = \mathbf{V}^T \boldsymbol{\phi}_k \tag{2.23}$$

Equation (2.23) can be expressed as

$$\mathbf{t}_k = \boldsymbol{\Lambda}^{-1/2} \mathbf{P}^T \boldsymbol{\chi} \boldsymbol{\phi}_k$$
$$= \boldsymbol{\Lambda}^{-1/2} \mathbf{P}^T \mathbf{k}(\mathbf{x}_k) \tag{2.24}$$

and

$$\mathbf{k}(\mathbf{x}_k) = [\boldsymbol{\phi}_1 \ \boldsymbol{\phi}_2 \ \cdots \ \boldsymbol{\phi}_m]^T \boldsymbol{\phi}_k$$
$$= [\boldsymbol{\phi}_1^T \boldsymbol{\phi}_k \ \boldsymbol{\phi}_2^T \boldsymbol{\phi}_k \ \cdots \ \boldsymbol{\phi}_m^T \boldsymbol{\phi}_k]^T$$
$$= [k(\mathbf{x}_1, \mathbf{x}_k) \ k(\mathbf{x}_2, \mathbf{x}_k) \ \cdots \ k(\mathbf{x}_m, \mathbf{x}_k)]^T \tag{2.25}$$

Having estimated the PCs from KPCA as in (2.24), The T^2 can be estimated as shown in Equation (2.26).

$$T^2{}_{kpca} = (\mathbf{t}_k - \bar{\mathbf{t}})^T \mathbf{S}^{-1} (\mathbf{t}_k - \bar{\mathbf{t}}) \tag{2.26}$$

where $\mathbf{t}_k$ are the scores of the new batch at time k, $\bar{\mathbf{t}}$ is the mean of $\mathbf{t}$ and $\mathbf{S}$ is the covariance of the score, $\mathbf{t}$ in the feature space. In addition, the Q metric can be estimated as in (2.27) to (2.29).

$$\mathbf{e}_k = \tilde{\mathbf{V}}^T \boldsymbol{\phi}_k \tag{2.27}$$

where $\tilde{\mathbf{V}} = [\tilde{\boldsymbol{\nu}}_{l+1} \ \cdots \ \tilde{\boldsymbol{\nu}}_m]$

$$Q_{kpca} = \mathbf{e}_k^T \mathbf{e}_k = \boldsymbol{\phi}_k^T \tilde{\mathbf{V}} \tilde{\mathbf{V}}^T \boldsymbol{\phi}_k \tag{2.28}$$

Since, the dimension of the feature space is not known, it is not possible to know the number of residual components there. Hence, the loading matrix $\tilde{\mathbf{V}}$ cannot be explicitly calculated. However, the product $\tilde{\mathbf{V}} \tilde{\mathbf{V}}^T$ can be calculated as the projection orthogonal to the principal component space, which is

$$\mathbf{Q}_{kpca} = \boldsymbol{\phi}_k^T (\mathbf{I} - \mathbf{V} \mathbf{V}^T) \boldsymbol{\phi}_k = \boldsymbol{\phi}_k^T \boldsymbol{\phi}_k - \boldsymbol{\phi}_k^T \mathbf{V} \mathbf{V}^T \boldsymbol{\phi}_k \tag{2.29}$$

Another way to address the problem of monitoring non-linear processes is to employ the Independent Component Analysis (ICA). The ICA is well suited for non-Gaussian processes [15, 121] and is discussed in the following section.

2.2.4.2 Independent Component Analysis

Independent Component Analysis (ICA) is a statistical approach for revealing hidden factors that underlie sets of process measurements. These process measurements are generally dependent and may be combinations of latent variables that are not directly measured. The ICA was originally proposed to solve blind source separation problems which consists of recovering independent source signals after they have

been linearly mixed with an unknown mixing matrix. These independent source signals are also known as independent components (ICs). The basic challenge of the ICA is to estimate the ICs ($\mathbf{S}$) and the mixing matrix ($\mathbf{A}$) from the process measurements without any knowledge of $\mathbf{S}$ or $\mathbf{A}$. The ICs are assumed to be non-Gaussian and mutually independent. In this instant, a set of variables are said to be statistically independent from each other when the value of one variable cannot be predicted given the value of another variable. The ICA technique can be described as an optimising process of maximising the non-Gaussianity [121].

ICA Problem Definition

Assume we have m measured variables $\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_m}$ that are given as linear combinations of $n(\leq m)$ unknown ICs $\mathbf{s_1}, \mathbf{s_2}, ..., \mathbf{s_n}$. The relationship between the two vectors is given by

$$\mathbf{x} = \mathbf{As}$$
$$\mathbf{x} = [x_1, x_2, ..., x_m]^T$$
$$\mathbf{s} = [s_1, s_2, ..., s_n]^T \quad (2.30)$$

where $\mathbf{A} \in \Re^{m \times n}$ is the full rank mixing matrix. For a set of process data consisting of N observations, the preceeding relationship can be rewritten as

$$\mathbf{X} = \mathbf{AS}$$
$$\mathbf{X} \in \Re^{m \times N}$$
$$\mathbf{S} \in \Re^{n \times N} \quad (2.31)$$

The objective of the ICA is to calculate a seperating matrix $\mathbf{W} \in \Re^{n \times m}$ so that the components of the reconstructed data matrix $\hat{\mathbf{S}}$ is given as

$$\hat{\mathbf{S}} = \mathbf{W}\mathbf{X} \tag{2.32}$$

where $\hat{\mathbf{S}}$ is the estimation of $\mathbf{S}$. When $\mathbf{W}$ is the inverse of $\mathbf{A}$, $\hat{\mathbf{S}}$ is the best estimation if $n = m$ [122].

The ICA consists of a pre-processing step known as the whitening stage before applying the ICA algorithm for the extraction of the ICs. In the whitening step, the correlated measurements are linearly transformed into uncorrelated latent variables, shown below.

$$\mathbf{Z} = \mathbf{V}\mathbf{X} \tag{2.33}$$

where $\mathbf{Z}$ is the whitened matrix and $\mathbf{V}$ is the linear transformation matrix. The PCA is commonly employed for the whitening step of the ICA approach [15, 121, 123, 124]. Only some of the extracted ICs are retained in the model space for process monitoring. The selection of the ICs for the model space is an important part of the ICA. Selecting too many ICs will cause a magnification of noise and poor process monitoring performance. Selecting a few dominant ICs has the following advantages [15];

- Robust Performance

- Reduced mathematical complexity

Nevertheless, the ordering of the ICs is difficult and there is no standard criterion to achieve this. The ICs can be ordered based on their non-Gaussianity [125]. Back and Weigend [122] determined the component order based on the L_α norm (maximal signal amplitude) of each IC. They multiplied the corresponding row of $\mathbf{A}$ with the ICs to obtain the weighted ICs and then determined the dominant ICs to be those

ICs with the largest maximal signal amplitudes. Lee et al. [15] adopted the Euclidean norm in their study to determine the component order. Nonetheless, the Euclidean norm has its limitations for the extraction and ordering of the dominant ICs as demonstrated in the work of Lee et al. [126]. In practice, the data dimension could be reduced by selecting a few rows of the demixing matrix based on the assumption that the rows with the largest sum of squares have the greatest effect on the ICs [15, 126, 127].

The ICA is sometimes considered to be an extension of the PCA [128]. However, the objectives of the ICA are clearly different from those of the PCA. The PCA reduces the dimension of the data by projecting the correlated process variables on a lower set of uncorrelated (second order statistics) PCs while retaining most of the original variance. The ICA on the other hand decomposes process measurements into statistically (high-order statistics) independent components of lower dimensions. The PCA is only able to decorrelate variables, but not to make them independent. This is because the PCA imposes independence up to the second order statistics information (mean and variance) while constraining the direction vectors to be orthogonal. The ICA on the other hand has no orthogonality constraint and involves higher-order statistics [15, 124]. Therefore, the ICA is able to extract more statistically useful information than the PCA [5, 15, 16] and therefore, gives better monitoring results than the PCA based monitoring techniques. Several ICA approaches have been developed which include ICA by maximization of non-Gaussianity, ICA by maximum likelihood estimation, ICA by minimization of mutual information and the fast fixed point algorithm for ICA (FASTICA) [129, 130].

Lee et al. [15] proposed the use of the ICA for continuous process monitoring. The basic idea of their approach was to extract essential ICs that drive the process and then combine them with process monitoring techniques. They employed the Euclidean norm to sort out the rows of the demixing matrix in order to show only

those ICs that cause dominant changes in the process. The Euclidean norm was their choice for its simplicity and its efficiency in ICA monitoring. Three monitoring metrics; $I_d{}^2$ for the dominant ICs, the $I_e{}^2$ for the excluded ICs and the Q metric for the residual space were determined and then KDE employed to derive the control limit for all three statistics. Also, contribution plots were used for diagnosis of the faults. To illustrate the efficiency of their approach, they applied their method to a simple multivariate process example and a simulated waste water treatment plant. Moreover, they demonstrated that the ICA solution extracts the original source signal to a greater degree than the PCA.

Albazzaz and Wang [121] adopted the fast fixed-points (FASTICA) developed by Hyvarinen and Oja [129]. Different from the study of Lee et al. [15], they employed all the extracted ICs for process monitoring because they suggested that no single IC was more important than another. Hence, in their study, the number of ICs was determined by the number of PCs. In addition, they applied a box-cox transformation to change the non-Gaussian co-ordinates of the ICs to a Gaussian distribution in order to justify the use of control limits estimated based on the Gaussian assumption. Their approach was applied to the Manresa Waste Water Treatment Plant in spain to demonstrate the efficiency.

Although the ICA has been reported to be efficient for monitoring non-linear processes [15, 121], the ICA has its draw-backs.

- The ICA does not determine how many ICs to be extracted in order to establish an optimal ICA model. The ICs are generally extracted to the dimension of the given data or the dimension of the latent variables from the whitening stage [121] and this can incure a high computational load.

- The extracted ICs are not ordered by their importance and there is no standard criterion to order the ICs.

- Finally, the random initialization of the demixing matrix in the whitened space leads to different results in the ICA algorithm.

To address these three draw-backs of the ICA, Lee et al. [126] proposed a novel multivariate statistical monitoring technique based on a modified ICA. The basic idea of their aproach was to use the modified ICA for the extraction of the dominant ICs from the normal operating process data and then combine them with statistical process monitoring techniques. The objective of their modified ICA was to find a demixing matrix such that the elements of the extracted independent component vectors become as independent of each other as possible and are ordered by their variances. The PCA was first employed for the estimation of the few dominant PCs before updating the PCs using the FastICA algorithm. The ICs were then assumed to have the same variance as the PCs from which they were updated and each IC was ordered according to the variance of the ICs. In addition, the initialization in the modified ICA is based on the assumption that extracted PCs are good initial estimates of the ICs. This gives a consistent solution unlike the random initialization based traditional ICA algorithm. Different from the challenge of finding a demixing matrix in the traditional ICA, the challenge in the modified ICA is to find a matrix which has fewer parameters to estimate as a result of its orthogonality. For their modified ICA approach, the Hotelling's T^2 and Q metrics were employed for fault detection before contribution plots were constructed for diagnosis. Their approach was applied to a simulated Waste Water Treatment Plant, the TEP plant as well as a semiconductor etch process to demonstrate its superiority. The modified ICA was reported to be able to extract a few dominant ICs, determine the order of the ICs and give a consistent solution by avoiding random initialization. Moreover, the proposed approach detects various faults more efficiently than the PCA.

Kano et al. [131] proposed an ICA based process monitoring technique and deviced monitoring charts for each IC with the control limits estimated based on the average

run length (ARL). In their ICA based approach, data from the normal operating process was acquired and then normalised to zero mean and unit variance before determining a seperating matrix which was employed to calculate the ICs as in Equation (2.30). To demonstrate the feasibility of their approach, it was applied to a simple four variable system and a continuous-stirred-tank-reactor (CSTR) process. Also, their ICA approach was compared with the PCA and reported to detect faults earlier than the PCA approach.

Traditionally, the ICA is first based on the PCA to decorrelate the process measurements before applying the ICA algorithm for the extraction of the ICs [15, 16, 44, 62, 121, 132, 133, 134]. This allows the ICs to be interpreted by the simple geometry of the PCA. However, the PCA because of its static nature is not appropriate for dynamic processes. Consequently, the connection of the ICA with the static PCA makes such ICA approaches equally inappropriate for the monitoring of dynamic processes [5, 123, 127].

2.2.4.3 Dynamic Independent Component Analysis

To address the limitation of the traditional ICA for dynamic processes, the Dynamic ICA (DICA) was proposed [123]. Lee et al. [123] extended the ICA method to improve its monitoring performance for dynamic processes. Their DICA approach was first based on the DPCA for the pre-processing stage before applying the ICA for the extraction of the ICs. To demonstrate the efficiency of their DICA approach, it was applied to a simple multivariate dynamic process as well as the TEP plant. More importantly, their DICA approach was reported to be more efficient than the traditional ICA approach.

Similar to the work of Lee et al [123], Stefatos and Hamza [127] developed a DICA model [123] that they also applied to the TEP plant. In addition, they proposed

a contribution plots that took the spatial correlation between the process measurements and the serial correlation between observations into account. In their proposed approach for diagnosis, they assumed that all observations found as outliers were due to the same fault. Notwithstanding, when an outlying observation was a fault on its own, diagnosis was done with respect to that outlying observation. Above all, their approach was reported to be more efficient than the PCA, DPCA and traditional ICA schemes for fault detection and diagnosis. Also, their DICA algorithm was reportedly able to accurately detect and isolate the root causes for all the TEP faults.

Although the DICA is reported to be more efficient than the ICA for monitoring dynamic processes [123, 127], still the DICA like the DPCA is not the best approach to capture dynamic behaviour from process measurements [5]. As a result, the statistical advantage of the ICA is not fully exploited by the DICA and the performance of the DICA for dynamic monitoring is still not satisfactory.

Apart from the ICA based approaches, all the other monitoring approaches discussed above are generally associated with control limits derived based on the assumption that the estimated latent variables have a Gaussian distribution [1, 2, 9]. However, most industrial processes are non-linear, following a non-Gaussian distribution. For such processes, the Gaussian assumption is invalid and control limits estimated based on the Gaussian assumption may not be able to correctly identify the underlying faults. As a result, control limits of the Hotelling's T^2 and Q metrics both estimated based on the Gaussian assumption are restrictive and inappropriate for such non-linear industrial processes. One way to address the problem of monitoring such non-linear processes is by directly estimating the underlying probability density function (PDF) of the T^2 and Q metrics through the KDE to derive the correct control limit [4, 5, 13, 135].

2.2.4.4   Kernel Density Estimations

Kernel Density Estimations (KDE) is an efficient tool for the estimation of the probability density function of a process data. The KDE is easy to visualise conceptually and has applications in econometrics, chemometrics and process monitoring. Generally, density estimation methods are classified into parametric and non-parametric methods. The parametric methods are those that estimate the PDF with known underlying distributions whereas the non-parametric methods are those that estimate the PDF from the process data without a known underlying distributions. An example of such non parametric methods is the KDE. The KDE basically involves placing a kernel function with a probability mass equal to the inverse of the number of observations at each sampling point and then adding all the kernel functions together to form a kernel density estimate as shown below.

$$\hat{p}(x) = \frac{1}{Mh}\sum_{k=1}^{M}K\left(\frac{x-x_k}{h}\right) \tag{2.34}$$

where x_k, $k = 1, 2, \cdots, M$ are samples of x, h is the bandwidth and $K(\cdot)$ is the kernel function. Examples of kernel functions include the Gaussian kernels, Biweight kernel, Triangular kernel and Epanechnikov kernel. In kernel density estimations, the bandwidth selection is important because selecting a band-width too small will result in the density estimator being too rough, a phenomenon known as under-smoothed while selecting a band-width too big will result in the density estimator being too flat. An efficient use of the KDE technique requires an optimal selection of the band-width of the kernel. Several techniques have been proposed for data-driven band-width selection [136]. However, there is no single universally accepted approach to determine the band-width. The choice of the band-width is influenced by the purpose for which the density estimate is to be used.

One choice for band-width selection is to plot several nominal operating regions and

then select the estimate which is most similar to prior knowledge of the density. Another approach for band-width selection is the use of likelihood cross validation [13]. Also, the adoption of a rough estimation of the optimal band-width subject to mimimizing the mean integrated square error is reportedly efficient [14, 137]. The optimal bandwidth h_{opt} is derived as

$$h_{opt} = 1.06\sigma N^{-1/5} \qquad (2.35)$$

where σ is the standard deviation and N is the number of observations [137].

Martin and Morris [13] proposed a novel approach for constructing control limits based on the density of the process data with 99% confidence intervals. They referred to the control limit derived based on their approach as the M^2 metric. A likelihood based confidence region was constructed using the non-parametric bootstrap. Their approach was associated with the PCA and PLS methods. They combined techniques of standard bootstrap and KDE to overcome the limitations of the T^2 and Q metrics mentioned above. The band-width in their approach was selected using the least squares cross validation. In addition to their proposed approach, they derived control limits based on the Hotelling's T^2 metric for comparison. Both methodologies were applied to a continuous polyethylene reactor and a polymerisation reactor to demonstrate the efficiencies of both methodologies. In their work, some of the faults were undetected using the T^2 control limits, demonstrating the inadequacy of the Hotelling's T^2 metric for processes where the process measurements are dependent. Moreover, their M^2 metric was reported to be more efficient than the T^2 metric.

Chen et al. [135] adopted several KDE methods associated with PCA to monitor a gas melter process. The emphasis of their work was to demonstrate the efficiencies of three different density estimators which were verified based on the misclassification rates at given confidence intervals. The three density estimators were MISE

(mean integrated square error), ADMISE (adaptive mean integrated square error) and BAMISE (biased asymptotic mean integrated square error). They demonstrated that the KDEs could obtain nonparametric empirical density function for more efficient process monitoring. The control limits from the KDE detected the faults about 50 minutes earlier than the control limits from the parametric methods. Also, among the three density estimators employed in their work, the BAMISE was reported to be a more practical and efficient density estimator than the MISE and ADMISE.

Xiong et al. [128] extended the use of the PCA and ICA techniques using the KDE, to improve condition monitoring performance. To extend the applicable range of the PCA for non-linear processes, the PCA approach in their work was associated with control limits estimated based on the KDE, an approach which they referred to as PCA with KDE. Also, their proposed ICA approach was referred to as the ICA with KDE. Both approaches as well as the PCA and ICA without KDE techniques were applied to an industrial spheripol craft polypropylene catalyser reactor. Their PCA with KDE and ICA with KDE techniques were reported to improve monitoring precision over the PCA and ICA techniques without the KDE. However, the ICA without KDE approach considered in their work for comparison was associated with control limits estimated based the Gaussian assumption which is invalid due to the non-Gaussianity of the extracted ICs. Thus, the ICA without KDE considered in their work did not have appropriate control limits. This is expected to affect the efficieny of their ICA without KDE approach [4]. A suggestion is that an ICA approach with more appropriate control limits be employed in their work to form an unbiased platform for comparison with their proposed ICA with KDE approach.

Also, Odiowei and Cao [138] enhanced a PCA based process monitoring using the KDE. Their approach was applied to the Manresa Waste Water treatment plant and reported to be more efficient than the Gaussian assumption based Hotelling's T^2 statistic.

Jia et al. [117] developed a non-linear principal component analysis methodology based upon the input-training neural network as an appropriate methodology to handle the limitations of using static PCA for non-linear systems. In their study, the KDE was employed to define the action and warning limits while a differential contribution plot was derived to identify the potential sources of the process faults in the non-linear situations. The efficiency of their technique was demonstrated using an industrial fluidised bed reactor.

Table 2.1: Brief description of case studies

Case study	References	Description
Semi Conductor Etch Process	Lee et al. [126].	A modified ICA approach was applied to a semi conductor etch process. The process is dynamic in nature. In their study, only the machine state variables were considered for monitoring. The process data is available.
Dearomatisation Process	Komulainen et al. [2]	The DPLS approach was applied to a real plant data from the dearomatization unit of the Fortum Naantali Refinery in Finland. The process is dynamic, non-linear and continuous. Measurements were selected for monitoring based on auto-correlations and process knowledge. The process data is not available.
Chemical Separation Process	Juan and Fei [12]	The CVA was applied to the data from a dynamic real chemical separation process plant. The data is unavailable.
Industrial Sphericol Craft Polypropylene Catalyser	Xiong et al. [128]	The PCA with KDE and ICA with KDE approaches were applied to data from an industrial spheripol craft polypropylene catalyser reactor. The process is dynamic. The data is not available.
Industrial Fluidised Bed Reactor	Jia et al. [117]	Non-linear PCA was applied to data from an industrial fluidised bed reactor. The reactor is known to exhibit non-linear behaviour and is a real plant without available data.
An Interconnected three tank System	Mina and Verde [78]	A DPCA-based approach was applied to data from a simulated interconnected three tank system. The data from the three tank system is unavailable.
Tenneessee Eastman Process Plant	Lee et al. [123]	DICA was applied to the TEP plant. The TEP is a simulated plant with dynamic and non-linear properties. The TEP data is available and can be downloaded from http://brahms.scs.uiuc.edu
A Waste Water Treatment Benchmark (ASM1 model)	Lee et al. [15]	Applied the ICA to a simulated data obtained from a waste water treatment plant. It is a dynamic and non-linear Process with model available at the website of the COST working group (http://www.ensic.u-nancy.fr/COSTWWTP).
Polymerisation Reactor	Norvalis et al, [9]	Their proposed approach was employed to monitor data from a simulated polymerization plant. It is a dynamic and semi-batch process. The data from this plant is not available.
A simple Multivariate Process	Lee et al. [15]	The ICA was applied to a simple multivariate process with equations available in the paper of Lee et al. [15].
A Domestic Waste Water Treatment Plant in Korea	Lee et al. [95]	Non-linear DPLS was applied to the data from a real waste water treatment plant in Korea. The process has both dynamic and non-linear properties.

2.3 Summary of Case Studies

Generally, monitoring methods are applied to case studies to illustrate their efficiency. These case studies also create platforms to compare the efficiencies of various monitoring approaches. Hence, the choice of a suitable case study is very important. Unfortunately, finding such suitable case studies can be very challenging. To address this challenge, in this section, a summary of all the case studies reviewed in this chapter is presented in Table 2.1.

2.4 Conclusion

In this chapter a literature review report has been presented to explain the current status of process monitoring. The original univariate monitoring methods have been extended to multivariate processes. Furthermore, the multivariate monitoring methods were also extended in different ways to account for various characteristics of industrial processes. It is well known that chemical and industrial processes possess both dynamic and non-linear properties. The dynamic property of chemical processes was accounted for by extending some multivariate methods to their dynamic counterparts [1, 2, 78]. In addition, the non-linear properties were addressed by using various techniques [111, 113, 114, 115, 117]. These dynamic extensions as well as the non-linear extensions have been reported to be able to improve monitoring performance.

Nevertheless, to further improve the condition monitoring, both dynamic and non-linear properties should be considered together. Hence, monitoring techniques that can simultaneously account for the dynamic and non-linear properties are expected to further improve the condition monitoring over those methods that account for either of the properties alone.

Chapter 3

Extended Dynamic Approaches using Kernel Density Estimations

The development of three novel KDE-based algorihtms; DPCA with KDE, DPLS with KDE and the CVA with KDE approaches is explained in this chapter. Also, these algorithms are evaluated using the TEP plant and their monitoring performances compared with their non-KDE counterparts as well as one with another.

3.1   Introduction

It is well known that most chemical and industrial processes exhibit dynamic and non-linear properties. Static monitoring approaches like the PCA and the PLS discussed in the previous chapter have been extended to DPCA [1] and DPLS [2] respectively, in order to address the dynamic issues related to most chemical and industrial processes. Moreover, the DPCA and DPLS approaches are reported to be more efficient than the static PCA and PLS approaches for monitoring dynamic processes. In addition to the DPCA and DPLS methods, the CVA [99] was also

56

developed for dynamic process monitoring and reported for its efficiency [7, 101].

Although the DPCA, DPLS and CVA approaches are reported to be more efficient than the static approaches for dynamic process monitoring, for simplicity, these approaches are associated with control limits estimated based on the assumption that the variation of data is Gaussian. Nevertheless, most chemical processes are non-linear. For such processes, although the distribution of stochastic sources might be Gaussian, such as measurement noises and normally distributed disturbance, the variation of the process variables will be non-Gaussian. This means that for such processes, the assumption of Gaussianity is invalid. Even more, the control limits estimated based on the Gaussian assumption are unable to correctly identify the underlying faults.

The problem of monitoring non-linear processes with non-Gaussian variations can be addressed by directly estimating the underlying probability density function (PDF) of the control chart using the kernel density estimations (KDE). Probability density functions (PDFs) are useful engineering applications to describe distributions of random processes [75, 135]. The KDE is a well established approach to estimate the PDF without imposing a parametric model. The KDE is particularly suitable for metrics with univariate representations such as the Hotelling's T^2 and Q metrics [4].

In this chapter, three existing dynamic approaches (DPCA, DPLS and CVA) are extended using the KDE to adapt them to non-linear systems. As a result of the extensions in this chapter, three novel approaches; DPCA with KDE, DPLS with KDE and CVA with KDE are developed. The objective of the work is to develop monitoring approaches that are able to simultaneously address the dynamic as well as the non-linear issues related to most chemical and industrial processes. Furthermore, these developed monitoring algorithms are evaluated using the TEP plant to illustrate their efficiency in the monitoring of dynamic non-linear systems. In the following sections, the DPCA, DPLS and CVA algorithms are described. Thereafter,

the upper control limit and the KDE algorithm are also explained.

3.2 Dynamic Principal Component Analysis

Generally, for dynamic systems, the current values $(\mathbf{y}_k)$ of the process measurements are dependent on the past values $(\mathbf{y}_{k-i,\, i=1,2,\cdots,m})$. Hence, it is important to identify the relations between $\mathbf{y}_k$ and $\mathbf{y}_{k-i}$. Ku and coworkers [1] were able to achieve this with their proposed DPCA approach. The DPCA technique is basically the same as the static PCA, except that the data matrix consists of time shifted vectors in order to account for the auto-correlations and cross-correlations. This ability of the DPCA to account for these correlations makes the DPCA technique a more appropriate dynamic monitoring technique than the static PCA.

Assume a non-linear dynamic plant under consideration represented as follows;

$$\mathbf{y}_k = \mathbf{f}(\mathbf{y}_{k-1}, \mathbf{y}_{k-2}, \cdots, \mathbf{y}_{k-d}) + \mathbf{v}_k \tag{3.1}$$

where $\mathbf{y}_k$ is the measurement vector, $\mathbf{f}(.)$ is an unknown non-linear function, d is the number of lags and $\mathbf{v}_k$ is the measurement noise vector. Equation (3.1) can be linearised as

$$\mathbf{y}_k = \sum_{i=1}^{d} \mathbf{A}_i \mathbf{y}_{k-i} + \boldsymbol{\epsilon}_k \tag{3.2}$$

where $\mathbf{A}_i$ is an unknown matrix and $\boldsymbol{\epsilon}_k$ is a modelling error partially due to the underlying nonlinearity of the plant which has not been included in the linear model, as well as associated with the measurement noise vector $\mathbf{v}_k$. Due to the unknown nonlinearity, both $\mathbf{y}_k$ and $\boldsymbol{\epsilon}_k$ generally will be non-Gaussian although $\mathbf{v}_k$ might be normally distributed.

The matrix $\mathbf{Y}$ is defined as

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_k^T \\ \mathbf{y}_{k+1}^T \\ \vdots \\ \mathbf{y}_m^T \end{bmatrix} \tag{3.3}$$

Due to the linearization in (3.2), $\mathbf{Y}$ can be decomposed as shown in Equation (3.4)

$$\mathbf{Y} = \mathbf{T}\mathbf{P}^T + \mathbf{E} \tag{3.4}$$

where $\mathbf{T}$ is the score matrix consisting of the principal components (PCs), $\mathbf{P}$ is the loading matrix and $\mathbf{E}$ is the residual matrix. Due to the nonlinearity, both the PCs in $\mathbf{T}$ and the modelling errors in $\mathbf{E}$ are non-Gaussian.

Assume the data matrix $\mathbf{Y}$ taken from a dynamic system working under normal conditions. Each column in $\mathbf{Y}$ represents an auto-correlated time series. To account for these correlations, time lags are applied to each of the time series to form an augmented data matrix with time shifted vectors ($\mathbf{Y}_A$). The augmented data matrix $\mathbf{Y}_A$ is constructed as shown below;

$$\mathbf{Y}_A = \begin{bmatrix} \mathbf{y}_t^T & \mathbf{y}_{t-1}^T & \cdots & \mathbf{y}_{t-l}^T \\ \mathbf{y}_{t-1}^T & \mathbf{y}_{t-2}^T & \cdots & \mathbf{y}_{t-l-1}^T \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{y}_{t+l-m}^T & \mathbf{y}_{t+l-m-1}^T & \cdots & \mathbf{y}_{t-m}^T \end{bmatrix} \tag{3.5}$$

where $\mathbf{y}_t$ is the n-dimensional observation vector at time instance t. The DPCA approach consists of applying the PCA on the augmented data matrix defined in Equation (3.5). Moreover, the DPCA is reported to be more efficient than static PCA for monitoring dynamic processes [1, 78, 79]. The augmented data matrix $\mathbf{Y}_A$ is normalised to zero mean and unit variance before carrying out eigen-value

decomposition to avoid variables with large values imposing superficial variability. This normalisation is performed as shown in Equation (3.6)

$$\mathbf{Y}^*_{Aj} = (\mathbf{Y}_{Aj} - \mu_{yj})/\mathbf{S}_{yj} \tag{3.6}$$

where $\mathbf{Y}^*_{Aj}$ is the normalised vector of $\mathbf{Y}_{Aj}$ which is the jth column of the augmented data matrix $\mathbf{Y}_A$, μ_{yj} is the mean of $\mathbf{Y}_{Aj}$ and $\mathbf{S}_{yj}$ is the standard deviation of $\mathbf{Y}_{Aj}$. The covariance matrix ($\mathbf{C}$) is estimated and then the eigen-value decomposition of the covariance matrix obtained as shown.

$$\mathbf{C} = \mathbf{Y}_A^{*T}\mathbf{Y}_A^*/(m - l) = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T \tag{3.7}$$

where $m+1$ is the number of observations, l is the number of lags, $\mathbf{V}$ is the matrix of eigen-vectors and $\mathbf{\Lambda}$ is a diagonal matrix of eigen-values with a decreasing magnitude. The DPCA model can be expressed in terms of the loading matrix consisting of eigen-vectors corresponding to the a largest eigen-values.

$$\mathbf{Y}^*_A = \mathbf{T}\mathbf{P}^T + \mathbf{E} \tag{3.8}$$

where $\mathbf{T}$ and $\mathbf{E}$ are the principal component and residual matrices respectively. The PCs estimated from the eigen-vectors corresponding to the a largest eigen-values are retained in the model space while those estimated from the remaining $n - a$ eigen-vectors are excluded to the residual space. This way, the DPCA is able to capture the variations in the data while reducing the dimension and minimising the effect of random noise. A Scree plot based on maximum variance [21] can be employed to determine the number of principal components (PCs) to retain in the DPCA model space.

From the retained a PCs in the model space, the Hotelling's T^2 metric can be

derived to determine whether or not the monitored process is in control. Given a new observation vector defined as

$$\mathbf{y}_{k,new} = \begin{bmatrix} \mathbf{y}_k \\ \mathbf{y}_{k-1} \\ \vdots \\ \mathbf{y}_{k-l} \end{bmatrix} \in \Re^{ml}, \quad \tilde{\mathbf{y}}_{k,new} = (\mathbf{y}_{k,new} - \bar{\mathbf{y}})\mathbf{S}_y^{-1} \tag{3.9}$$

where

$$\bar{\mathbf{y}} = [\mu_{y1} \ \mu_{y2} \ \cdots \ \mu_{y(l+1)n}]^T \ , \quad \mathbf{S}_y = \text{diag}[\mathbf{S}_{y1} \ \mathbf{S}_{y2} \ \cdots \ \mathbf{S}_{y(l+1)n}]^T$$

and n is the number of measurements. The DPCA score can be estimated as

$$\mathbf{t}_{new} = \mathbf{P}\tilde{\mathbf{y}}_{k,new} \tag{3.10}$$

Then, the T^2 calculated as in Equation (3.11).

$$T^2{}_{dpca} = \mathbf{t}_{new}^T \mathbf{S}^{-1} \mathbf{t}_{new} \tag{3.11}$$

where $\mathbf{S} = \frac{1}{m-1}\mathbf{T}^T\mathbf{T}$.

In addition, the PCs estimated from the eigen-vectors corresponding to the $n - a$ singular values can be monitored by using the Q statistic developed by Jackson and Mudholkar [67] and defined in Equations (3.12) and (3.13) respectively.

$$\mathbf{e} = (\mathbf{I} - \mathbf{P}\mathbf{P}^T)\tilde{\mathbf{y}}_{k,new} \tag{3.12}$$

where $\mathbf{e}$ is the residual. The Q metric is estimated as in (3.13)

$$Q_{dpca} = \mathbf{e}^T\mathbf{e} \tag{3.13}$$

The T^2 and Q metrics estimated in (3.11) and (3.13) can be employed for process monitoring.

3.3 Dynamic Partial Least Squares

In the PCA and DPCA based approaches, the PCs in the model spaces are expected to retain most of the useful predictive data. However, these PCs may not contain the discriminatory power required to diagnose faults. Fortunately, this is not the case with the PLS and the PLS based approaches. Hence, in this work, a PLS based approach, the DPLS is also considered.

Consider a non-linear dynamic plant under consideration represented as follows:

$$\mathbf{y}_k = \mathbf{f}(\mathbf{y}_{k-1}, \cdots, \mathbf{y}_{k-d}, \ \mathbf{x}_k, \ \mathbf{x}_{k-1}, \cdots, \mathbf{x}_{k-d}) + \mathbf{v}_k \qquad (3.14)$$

$\mathbf{f}(.)$ is an unknown non-linear function while $\mathbf{v}_k$ is a measurement noise vector. Equation (3.14) can be linearised as

$$\mathbf{y}_k = \sum_{i=1}^{d} \mathbf{A}_i \mathbf{y}_{k-i} + \sum_{i=0}^{d} \mathbf{B}_i \mathbf{x}_{k-i} + \boldsymbol{\eta}_k \qquad (3.15)$$

where $\mathbf{A}_i$ and $\mathbf{B}_i$ are unknown matrices, while $\boldsymbol{\eta}_k$ is collective modelling error partially due to the underlying nonlinearity of the plant which has not been included in the linear model, as well as associated with the measurement noise vector $\mathbf{v}_k$. Due to the unknown nonlinearity, the collective modelling error, $\boldsymbol{\eta}_k$, as well as the measurements, $\mathbf{x}_k$ and $\mathbf{y}_k$ will be non-Gaussian although $\mathbf{v}_k$ might be normally distributed. From Equation (3.15) above, let

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_k^T \\ \mathbf{x}_{k+1}^T \\ \vdots \\ \mathbf{x}_m^T \end{bmatrix}, \mathbf{Y} = \begin{bmatrix} \mathbf{y}_k^T \\ \mathbf{y}_{k+1}^T \\ \vdots \\ \mathbf{y}_m^T \end{bmatrix}$$

Due to the linearization in (3.15), the following linear decomposition can be assumed.

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathbf{E}$$

$$\mathbf{Y} = \mathbf{U}\mathbf{Q}^T + \mathbf{F} \tag{3.16}$$

where $\mathbf{T}$ and $\mathbf{U}$ are score matrices, $\mathbf{P}$ and $\mathbf{Q}$ are loading matrices while $\mathbf{E}$ and $\mathbf{F}$ are the residuals of $\mathbf{X}$ and $\mathbf{Y}$ respectively. The latent variables in $\mathbf{T}$ and $\mathbf{U}$ as well as the modelling errors in $\mathbf{E}$ and $\mathbf{F}$ are non-Gaussian. Note, in this work, the standard DPLS approach is extended to non-linear models.

Taking the historical data matrices $\mathbf{X}$ and $\mathbf{Y}$ for the independent and dependent variables, the data matrices with time shifted vectors are constructed as shown in (3.17) and (3.18).

$$\mathbf{X}_A = \begin{bmatrix} \mathbf{x}_t^T & \mathbf{x}_{t-1}^T & \cdots & \mathbf{x}_{t-l}^T \\ \mathbf{x}_{t-1}^T & \mathbf{x}_{t-2}^T & \cdots & \mathbf{x}_{t-l-1}^T \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_{t+l-m}^T & \mathbf{x}_{t+l-m-1}^T & \cdots & \mathbf{x}_{t-m}^T \end{bmatrix} \tag{3.17}$$

$$\mathbf{Y}_A = \begin{bmatrix} \mathbf{y}_t{}^T & \mathbf{y}_{t-1}{}^T & \cdots & \mathbf{y}_{t-l}{}^T \\ \mathbf{y}_{t-1}{}^T & \mathbf{y}_{t-2}{}^T & \cdots & \mathbf{y}_{t-l-1}{}^T \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{y}_{t+l-m}{}^T & \mathbf{y}_{t+l-m-1}{}^T & \cdots & \mathbf{y}_{t-m}{}^T \end{bmatrix} \tag{3.18}$$

where $\mathbf{X}_A$ and $\mathbf{Y}_A$ are the augmented data matrices for the independent and dependent variables respectively, while $\mathbf{x}_t$ and $\mathbf{y}_t$ are the independent and dependent observation vectors at time instance t.

The augmented data matrices $\mathbf{X}_A$ and $\mathbf{Y}_A$ are normalised to zero mean and unit variance to avoid variables with large values imposing superficial variability. The normalisation of the augmented data matrix for the independent variables is shown in (3.19).

$$\mathbf{X}_{Aj}^* = (\mathbf{X}_{Aj} - \mu_{xj})/\mathbf{S}_{xj} \tag{3.19}$$

where $\mathbf{X}_{Aj}^*$ is the normalised vector of $\mathbf{X}_{Aj}$, which is the jth column of $\mathbf{X}_A$, μ_{xj} is the mean of $\mathbf{X}_{Aj}$ and $\mathbf{S}_{xj}$ is the standard deviation of $\mathbf{X}_{Aj}$. Also, the normalisation of the augmented data matrix for the dependent variable is illustrated in (3.20).

$$\mathbf{Y}_{Aj}^* = (\mathbf{Y}_{Aj} - \mu_{yj})/\mathbf{S}_{yj} \tag{3.20}$$

where $\mathbf{Y}_{Aj}^*$ is the normalised vector of $\mathbf{Y}_{Aj}$, which is the jth column of $\mathbf{Y}_A$, μ_{yj} is the mean of $\mathbf{Y}_{Aj}$ and $\mathbf{S}_{yj}$ is the standard deviation of $\mathbf{Y}_{Aj}$.

The DPLS approach consists of applying the PLS on the augmented data matrices constructed above. The PLS is described in the following section.

3.3.1   Partial Least Squares Regression

Partial Least Squares (PLS) Regression is a predictive process monitoring technique that involves two sets of variables; the predictor matrix ($\mathbf{X}$) and the response matrix ($\mathbf{Y}$). In the PLS, the predictor matrix ($\mathbf{X}$) is decomposed into a score matrix ($\mathbf{T}$) and a loading matrix ($\mathbf{P}$) that are correlated with the response matrix ($\mathbf{Y}$). Generally, the score matrix ($\mathbf{T}$) extracted from the decomposition of the predictor matrix ($\mathbf{X}$) is employed for the estimation of the score matrix ($\mathbf{U}$) of the response matrix ($\mathbf{Y}$) while the score matrix ($\mathbf{U}$) of the response matrix ($\mathbf{Y}$) is further employed to construct predictions for the responses. The score and loading matrices from the decomposition of $\mathbf{X}$ and $\mathbf{Y}$ can be estimated using the NIPALS algorithm or the singular value decomposition (SVD). However, in this study, the NIPALS algorithm is adopted. The NIPALS algorithm is explained in the following section.

3.3.2   Nonlinear Iterative Partial Least Squares Algorithm

There are two NIPALS methods to model the predicted block; PLS1 and PLS2 methods. The PLS1 is similar in operation to the PLS2 except that for the PLS1 approach each predicted variable is modelled separately whereas all the predicted variables are modelled simultaneously in the PLS2 approach [21]. Although the PLS2 requires a longer computation time [21, 139], the PLS2 is adopted in this study because of the simplicity of having to work with a single model. The predictor and response matrices are decomposed as shown in Equations (3.21) and (3.22).

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T \tag{3.21}$$

$$\mathbf{Y} = \mathbf{U}\mathbf{Q}^T \tag{3.22}$$

In the PLS approach, the decomposition of $\mathbf{X}$ and $\mathbf{Y}$ is followed by a regression

step where the PLS regresses the estimated $\mathbf{Y}$ score ($\mathbf{U}$) to the $\mathbf{X}$ score ($\mathbf{T}$) using a regression matrix ($\mathbf{B}$). The regression step is shown in (3.23).

$$\mathbf{U} = \mathbf{TB} \tag{3.23}$$

The decomposition and regression steps result in the reconstruction of a new set of predictor and response variables, $\mathbf{X}_{new}$ and $\mathbf{Y}_{new}$ respectively.

For the reconstruction of $\mathbf{X}_{new}$ and $\mathbf{Y}_{new}$, a reduced set of estimated score variables based on the retained latent variables is utilised. Moreover, the choice of latent variables is very important to have a good prediction. This is because a retained latent variable number too high will cause a magnification of noise and result in poor process monitoring performance while a retained latent variable number too low will result in the loss of some important process information. Cross validation is an efficient tool to determine the number of latent variables to retain in the model space [21].

PLS projects the row vectors of $\mathbf{X}$ and $\mathbf{Y}$ on to a reduced dimensional subspace, spanned by the weight vector ($\mathbf{w}_n$) which is estimated as shown in Equation (3.24) with $\mathbf{w}_n$ scaled to unit length.

$$\mathbf{w}_n = \mathbf{X}^T \mathbf{u}_n / ||\mathbf{X}^T \mathbf{u}_n|| \tag{3.24}$$

From the weight vector ($\mathbf{w}_n$) estimated in Equation (3.24), the score vector ($\mathbf{t}_n$) for $\mathbf{X}$ is estimated as shown in Equation (3.25).

$$\mathbf{t}_n = \mathbf{X}\mathbf{w}_n \tag{3.25}$$

The loading and score vectors for the response $\mathbf{Y}$ are also estimated as shown in

Equations (3.26) and (3.27).

$$\mathbf{q}_n = \mathbf{Y}^T\mathbf{t}_n/||\mathbf{Y}^T\mathbf{t}_n|| \tag{3.26}$$

$$\mathbf{u}_n = \mathbf{Y}\mathbf{q}_n \tag{3.27}$$

The score vectors are projections of the rows of the data blocks, $\mathbf{X}$ and $\mathbf{Y}$ onto a dimensional subspace [140, 141]. The weight and score vectors are determined iteratively till there is convergence and then the regression coefficient (b_n) for the training data is estimated as shown in Equation (3.28) below.

$$b_n = \mathbf{t}_n{}^T\mathbf{u}_n/\mathbf{t}_n{}^T\mathbf{t}_n \tag{3.28}$$

Equation (3.28) is the performance of the linear regression between $\mathbf{t}_n$ and b_n to produce the inner relationship. From the regression coefficient estimated in Equation (3.28), a new value of $\mathbf{u}_n$ ($\mathbf{u}_{nnew}$) is estimated as shown in Equation (3.29)

$$\mathbf{u}_{nnew} = \mathbf{t}_n b_n \tag{3.29}$$

$$\boldsymbol{\epsilon}_n = \mathbf{u}_n - \mathbf{u}_{nnew} \tag{3.30}$$

$$\boldsymbol{\epsilon}_n = \mathbf{u}_n - \mathbf{t}_n b_n \tag{3.31}$$

Equation (3.30) shows the estimation of the prediction error, $\boldsymbol{\epsilon}_n$ can be represented differently as in Equation (3.31) by substituting Equation (3.29) in Equation (3.30). The loading matrix ($\mathbf{P}$) in Equation (3.21) is the predictor loading matrix with the columns ($\mathbf{p}$) as the loading vectors. These predictor loading vectors are estimated as

$$\mathbf{p}_n{}^T = \mathbf{t}_n{}^T\mathbf{X}||\mathbf{t}_n{}^T\mathbf{X}|| \tag{3.32}$$

Furthermore, the response loading vectors are estimated as

$$\mathbf{q}_n{}^T = \mathbf{u}_{nnew}{}^T \mathbf{Y} || \mathbf{u}_{nnew}{}^T \mathbf{Y} || \tag{3.33}$$

The loading vectors are the projections of the columns of $\mathbf{X}$ and $\mathbf{Y}$ onto a reduced dimensional subspace [140]. The residuals are estimated by subtracting the variation of the nth pair of score vectors from the $\mathbf{X}$ and $\mathbf{Y}$ matrices. This process is known as deflation. The deflation process for $\mathbf{X}$ and $\mathbf{Y}$ matrices is shown in Equation (3.34) and Equation (3.35) respectively.

$$\mathbf{E}_n = \mathbf{X} - \mathbf{t}_n \mathbf{p}_n^T \tag{3.34}$$

$$\mathbf{F}_n = \mathbf{Y} - \mathbf{u}_{nnew} \mathbf{q}_n{}^T \tag{3.35}$$

where $\mathbf{E}_n$ and $\mathbf{F}_n$ are the residual matrices estimated by subtracting the matrices $\mathbf{t}_n \mathbf{p}_n{}^T$ and $\mathbf{u}_{nnew} \mathbf{q}_n^T$ from $\mathbf{X}$ and $\mathbf{Y}$ respectively. The matrices $\mathbf{t}_n \mathbf{p}_n{}^T$ and $\mathbf{u}_{nnew} \mathbf{q}_n^T$ are component matrices that describe the underlying structure between the predictor and response variables [141]. Using the component matrices in Equation (3.34) and Equation (3.35), the original measured predictor and response variables can be reconstructed as shown in Equation (3.36) and Equation (3.37). Hence, Equation (3.34) and Equation (3.35) now become Equation (3.36) and Equation (3.37) respectively as shown below.

$$\mathbf{X} = \mathbf{E}_n + \mathbf{t}_n \mathbf{p}_n{}^T \tag{3.36}$$

$$\mathbf{Y} = \mathbf{F}_n + \mathbf{u}_{nnew} \mathbf{q}_n{}^T \tag{3.37}$$

The vectors, $\mathbf{t}_n$, $\mathbf{u}_{nnew}$, $\mathbf{p}_n$ and $\mathbf{q}_n$ are stored as columns of the matrices $\mathbf{T}$, $\mathbf{U}_{new}$, $\mathbf{P}$, and $\mathbf{Q}$ respectively.

In summary, the PLS can be described as a combination of two models, the outer model concerned with the decomposition of $\mathbf{X}$ and $\mathbf{Y}$ demonstrated in Equations (3.36)

and (3.37) and the inner model which links the score and loading variables produced by the decomposition of $\mathbf{X}$ and $\mathbf{Y}$ in Equations (3.32) and (3.33). In the present study, the NIPALS algorithm described in Equation (3.24) to Equation (3.35) was employed to estimate the score and loading vectors. In this work, the cross validation was employed to determine the number of latent variables to retain in the model space.

3.3.3 Cross Validation

A major challenge of the PLS is to extract the optimum number of latent variables required to accurately model the responses. The cross validation however, is reported to be an efficient routine to achieve this goal [21]. Generally, cross validation involves splitting the training data into the construction and validation sets. Different types of cross validation methods include the Leave-one-out cross validation (LOO-CV), the v-fold cross validation and the Leave-multiple-out cross validation (LMO-CV). For the purpose of this thesis, the LMO-CV [21, 89] is adopted because the LOO-CV has a greater tendency of causing over-fitting [141, 142].

A step by step procedure of the LMO-CV is presented below;

- Step 1: The cross validation procedure in this study involves dividing the training data into subsets with each subset employed as a validation data set while the remaining subsets form the construction data set.

- Step 2: The construction data is used to build a one latent variable model.

- Step 3: The validation set is used to validate the model formed in step 2.

- Step 4: With the predictions from the validation set, the PRESS (Prediction Residual Sum of Squares) and RSMEV (Root Mean Square Prediction Error for validation data set) are estimated.

- Step 5: The procedure is repeated excluding each subset block only once.

- Step 6: The total PRESS for one latent variable is estimated by summing the individual PRESS values for each subset block. The total RSMEV for one latent variable is also estimated by summing the individual RSMEV values.

- Step 7: Step 1 to step 6 are repeated for 2, 3... min (n, m) latent variables, obtaining a series of PRESS values, where n is the number of observations and m is the number of process variables.

- Step 8: The number of latent variables that give the minimum value of both PRESS and RSMEV is identified as the number of latent variables to retain for the PLS analysis reported in this chapter.

In the cross validation process, two matrices $\mathbf{X}_{tr}$ and $\mathbf{Y}_{tr}$ are involved for both the construction and validation data sets [139, 143]. $\mathbf{X}_{tr} \in \Re^{m \times n}$ whereas $\mathbf{Y}_{tr} \in \Re^{m \times p}$, where m is the number of rows, n is the number of variables in the predictor variables and p is the number of variables in the response variables. The validation data set is of size d while the construction data is of size $m - d$. Hence, $\mathbf{X}_{m-d}$ and $\mathbf{Y}_{m-d}$ for the construction data sets are applied to the validation data set. Equation (3.38) shows the estimation of the mean square prediction error for the construction data.

$$RSMEC = \sqrt{\Sigma(\mathbf{Y}_{m-d} - \mathbf{z}_{con})^2/m - d} \tag{3.38}$$

where $\mathbf{Y}_{m-d}$ is the response variable in the construction data set, the subscript $(m - d)$ is the number of rows of the construction data set and $\mathbf{z}_{con}$ is the predicted response of the construction data set. Using the validation data set, the mean square prediction error is estimated as in Equation (3.39) while the prediction residual sum of squares is estimated as shown in Equation (3.40).

$$RSMEC = \sqrt{\Sigma(\mathbf{Y}_{m-d} - \mathbf{z}_{val})^2/di} \qquad (3.39)$$

$$PRESS = \Sigma(\mathbf{Y}_{m-d} - \mathbf{z}_{val})^2 \qquad (3.40)$$

where $\mathbf{Y}_{m-d}$ is the response variable in the validation data set, the subscript $m-d$ represents the number of rows of the validation data sets and $\mathbf{z}_{val}$ is the predicted response of the validation data set. To estimate the appropriate number of latent variables min (m, n), each validation set is eliminated from the training data block just once to form the construction set from which the model is built. The PRESS is actually the statistic for lack of prediction accuracy [89]. The eliminated subset is then used as a validation set to validate the model. From the predictions based on the validation set, PRESS and RSMEV are estimated for each subset from 1,2,... min (m, n) latent variables. This process is repeated until each subset is removed only once to form the construction set. To estimate the total PRESS and RSMEV for one latent variable, the individual values of PRESS and RSMEV for each subset (validation data) at one latent are summed. The same thing is done for 2 to min (m, n) latent variables and the total PRESS and RSMEV for 2 to min (m, n) latent variables also summed for each latent variable. The PRESS and RSMEV were estimated for the normalised and non-normalised training data sets and the result of cross validation from the PRESS and the RSMEV were very similar.

From the DPLS approach, two sets of latent variables are estimated for the independent and dependent variables. Generally, it is the latent variables of the independent variables that is employed to represent the DPLS model space. Also, from cross validation analysis, a latent variables can be retained in the model space as illustrated in Equation (3.25). From the latent variables in the model space, the T^2 metric can be derived to determine whether or not the process is in control. For the purpose of

process monitoring, a new observation vector can be defined as

$$
\mathbf{y}_{k,new} = \begin{bmatrix} \mathbf{y}_k \\ \mathbf{y}_{k-1} \\ \vdots \\ \mathbf{y}_{k-l} \end{bmatrix} \in \Re^{ml}, \quad \tilde{\mathbf{y}}_{k,new} = (\mathbf{y}_{k,new} - \bar{\mathbf{y}})\mathbf{S}_y^{-1} \tag{3.41}
$$

where

$$
\bar{\mathbf{y}} = [\mu_{y1} \ \mu_{y2} \ \cdots \ \mu_{y(l+1)n}]^T \ , \quad \mathbf{S}_y = \mathrm{diag}[\mathbf{S}_{y1} \ \mathbf{S}_{y2} \ \cdots \ \mathbf{S}_{y(l+1)n}]^T
$$

and n is the number of measurements. The score for $\mathbf{y}_{k,new}$ can be estimated as

$$
\mathbf{t}_{new,k} = \mathbf{P}\tilde{\mathbf{y}}_{k,new} \tag{3.42}
$$

The T^2 metric is derived as shown in (3.43).

$$
T^2{}_{dpls} = \mathbf{t}_{new,k}^T \mathbf{S}^{-1} \mathbf{t}_{new,k} \tag{3.43}
$$

where $\mathbf{S} = \frac{1}{m-1}\mathbf{T}^T\mathbf{T}$. In addition, the residual matrix $\mathbf{E}$ which is estimated as shown in Equation (3.36) consists of residual vectors. These residual vectors can be employed to estimate the Q metric to monitor the noise space. The Q metric is estimated as shown in (3.44) and (3.45).

$$
\mathbf{e} = (\mathbf{I} - \mathbf{P}\mathbf{P}^T)\tilde{\mathbf{y}}_{k,new} \tag{3.44}
$$

where $\mathbf{y}_{new,k}$ is the new data to be monitored, while $\mathbf{t}$ and $\mathbf{p}$ are the score and loading vectors respectively.

$$
Q_{dpls} = \mathbf{e}^T\mathbf{e} \tag{3.45}
$$

The T^2 and Q metrics estimated in Equations (3.43) and (3.45) can be employed for process monitoring.

Although the DPCA and DPLS approaches described above are reported to be able to improve the monitoring performance over the static PCA and PLS for dynamic processes, the DPCA and DPLS methods are not the best approach to capture the dynamic behaviour [3]. On the other hand, state space techniques like the Canonical Variate Analysis (CVA) are reported to be the best solution for dynamic processes [3, 4, 5, 6, 7]. Therefore, in this work, the CVA is also considered.

3.4   Canonical Variate Analysis

More recently, monitoring techniques based on Canonical Variate Analysis (CVA) have been developed with control limits derived based on the Gaussian assumption [9, 12, 21]. The CVA is a linear dimension reduction technique to estimate the minimum number of state variables for dynamic process monitoring. Different from the published works on the CVA, in this study, the CVA is extended to dynamic non-linear systems by identifying state variables directly from the process measurements and deriving appropriate control limits using the KDE for the detection of abnormal conditions.

Assume the non-linear dynamic plant under consideration represented as follows;

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k) + \mathbf{w}_k$$

$$\mathbf{y}_{k+1} = \mathbf{g}(\mathbf{x}_k) + \mathbf{v}_k \tag{3.46}$$

where $\mathbf{x}_k \in \Re^n$ and $\mathbf{y}_k \in \Re^m$ are state and measurement vectors respectively, $\mathbf{f}(\cdot)$ and $\mathbf{g}(\cdot)$ are unknown non-linear functions, whereas $\mathbf{w}_k$ and $\mathbf{v}_k$ are plant disturbances and measurement noise vectors respectively. It is clear that such an unknown non-

linear dynamic system is generally difficult to deal with for monitoring. However, at a stable normal operating point, the non-linear plant can be approximated by a linear stochastic state space model as follows;

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \boldsymbol{\varepsilon}_k$$

$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \boldsymbol{\eta}_k \tag{3.47}$$

where $\mathbf{A}$ and $\mathbf{C}$ are unknown state and output matrices respectively. Due to the unknown nonlinearity, the collective modelling errors, $\boldsymbol{\varepsilon}_k$ and $\boldsymbol{\eta}_k$ generally will be non-Gaussian although $\mathbf{w}_k$ and $\mathbf{v}_k$ might be normally distributed processes.

Instead of dealing with the unknown non-linear system (3.46) directly, in this work, the approximated linear state space model given in (3.47) is considered through the standard CVA approach. Although the linear model (3.47) is easier to deal with than the non-linear system (3.46), the collective errors $\boldsymbol{\varepsilon}_k$ and $\boldsymbol{\eta}_k$ have to be treated as non-Gaussian processes. This leads to the direct PDF estimation of the associated T^2 and Q metrics through the KDE approach explained in section 3.6.

In the CVA approach, firstly, the measurement vector $\mathbf{y}_k$ is expanded by q past and future measurements to give the past and future observation vectors $\mathbf{y}_{p,k}$ and $\mathbf{y}_{f,k}$ respectively.

$$\mathbf{y}_{p,k} = \begin{bmatrix} \mathbf{y}_{k-1} \\ \mathbf{y}_{k-2} \\ \vdots \\ \mathbf{y}_{k-q} \end{bmatrix} \in \Re^{mq}, \quad \tilde{\mathbf{y}}_{p,k} = \mathbf{y}_{p,k} - \bar{\mathbf{y}}_{p,k} \tag{3.48}$$

$$\mathbf{y}_{f,k} = \begin{bmatrix} \mathbf{y}_{k} \\ \mathbf{y}_{k+1} \\ \vdots \\ \mathbf{y}_{k+q-1} \end{bmatrix} \in \Re^{mq}, \quad \tilde{\mathbf{y}}_{f,k} = \mathbf{y}_{f,k} - \bar{\mathbf{y}}_{f,k} \tag{3.49}$$

74

where $\bar{\mathbf{y}}_{p,k} = \frac{1}{M-1} \sum_{k=q+1}^{q+M} \mathbf{z}_k \mathbf{z}_k^T$ and $\bar{\mathbf{y}}_{f,k}$ are the sample means of $\mathbf{y}_{p,k}$ and $\mathbf{y}_{f,k}$ respectively, and the products of mq represents the lengths of the past and future observation vectors respectively. The length of the past and future observations can be determined by checking the autocorrelation of the square sum of the process variables such that the correlation can be neglected when the time distance is larger than the number of lags determined.

These past and future observations are stochastic processes. Their sample-based covariance and cross-covariance matrices can be estimated through the truncated Hankel matrices as follows;

$$\mathbf{\Sigma}_{pp} := \frac{1}{M-1} \sum_{k=q+1}^{q+M} \tilde{\mathbf{y}}_{p,k} \tilde{\mathbf{y}}_{p,k}^T = \frac{1}{M-1} \mathbf{Y}_p \mathbf{Y}_p^T \qquad (3.50)$$

$$\mathbf{\Sigma}_{ff} := \frac{1}{M-1} \sum_{k=q+1}^{q+M} \tilde{\mathbf{y}}_{f,k} \tilde{\mathbf{y}}_{f,k}^T = \frac{1}{M-1} \mathbf{Y}_f \mathbf{Y}_f^T \qquad (3.51)$$

$$\mathbf{\Sigma}_{fp} := \frac{1}{M-1} \sum_{k=q+1}^{q+M} \tilde{\mathbf{y}}_{f,k} \tilde{\mathbf{y}}_{p,k}^T = \frac{1}{M-1} \mathbf{Y}_f \mathbf{Y}_p^T \qquad (3.52)$$

where $\mathbf{Y_p}$ and $\mathbf{Y_f}$ are past and future truncated M-column Hankel matrices respectively, and defined as follows;

$$\mathbf{Y}_p = \begin{bmatrix} \tilde{\mathbf{y}}_{p,q+1} & \tilde{\mathbf{y}}_{p,q+2} & \cdots & \tilde{\mathbf{y}}_{p,q+M} \end{bmatrix} \in \Re^{mq \times M} \qquad (3.53)$$

$$\mathbf{Y}_f = \begin{bmatrix} \tilde{\mathbf{y}}_{f,q+1} & \tilde{\mathbf{y}}_{f,q+2} & \cdots & \tilde{\mathbf{y}}_{f,q+M} \end{bmatrix} \in \Re^{mq \times M} \qquad (3.54)$$

For a set of measurements with total N observations, the last element of $\mathbf{y}_{p,q+1}$ in (3.48) is $\mathbf{y}_1$, whereas the last element of $\mathbf{y}_{f,q+M}$ in (3.49) should be $\mathbf{y}_N$. Therefore, the maximum number of columns of these Hankel matrices is

$$M = N - 2q + 1 \qquad (3.55)$$

The CVA aims to find the best linear combinations, $\mathbf{a}^T(\tilde{\mathbf{y}}_{f,k})$ and $\mathbf{b}^T(\tilde{\mathbf{y}}_{p,k})$ of the future and past observations so that the correlation between these combinations is maximised. The correlation can be represented as follows:

$$\rho_{fp}(\mathbf{a}, \mathbf{b}) = \frac{\mathbf{a}^T \mathbf{\Sigma}_{fp} \mathbf{b}}{(\mathbf{a}^T \mathbf{\Sigma}_{ff} \mathbf{a})^{1/2} (\mathbf{b}^T \mathbf{\Sigma}_{pp} \mathbf{b})^{1/2}} \tag{3.56}$$

Let $\mathbf{u} = \mathbf{\Sigma}_{ff}^{-1/2} \mathbf{a}$ and $\mathbf{v} = \mathbf{\Sigma}_{pp}^{-1/2} \mathbf{b}$. The optimization problem can be casted as:

$$
\begin{aligned}
\max_{\mathbf{u}, \mathbf{v}} \quad & \mathbf{u}^T (\mathbf{\Sigma}_{ff}^{-1/2} \mathbf{\Sigma}_{fp} \mathbf{\Sigma}_{pp}^{-1/2}) \mathbf{v} \\
\text{s.t.} \quad & \mathbf{u}^T \mathbf{u} = 1 \\
& \mathbf{v}^T \mathbf{v} = 1
\end{aligned}
\tag{3.57}
$$

According to linear algebra theory, the solution, $\mathbf{u}$ and $\mathbf{v}$ are left and right singular vectors of the scaled Hankel matrix, $\mathbf{H} = \mathbf{\Sigma}_{ff}^{-1/2} \mathbf{\Sigma}_{fp} \mathbf{\Sigma}_{pp}^{-1/2}$ and the maximal correlation $\sigma = \max_{\mathbf{a}, \mathbf{b}} \rho_{fp}(\mathbf{a}, \mathbf{b})$ is the corresponding singular value of $\mathbf{H}$. If the rank of the scaled Hankel matrix, $\mathbf{H}$ is r, then there are r non-zero singular values, σ_i, $i = 1, 2, \cdots, r$ in the descending order and correspondingly r pairs of the left and right singular vectors, $\mathbf{u}_i$ and $\mathbf{v}_i$ for $i = 1, 2, \cdots, r$. Singular values and vectors can be collected in the following matrix form of the singular value decomposition (SVD) as proposed by Larimore [10, 99].

$$\mathbf{H} := \mathbf{\Sigma}_{ff}^{-1/2} \mathbf{\Sigma}_{fp} \mathbf{\Sigma}_{pp}^{-1/2} = \mathbf{U} \mathbf{D} \mathbf{V}^T \tag{3.58}$$

where

$$\mathbf{U} = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_r \end{bmatrix} \in \Re^{mq \times r} \tag{3.59}$$

$$\mathbf{V} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_r \end{bmatrix} \in \Re^{mq \times r} \tag{3.60}$$

$$\mathbf{D} = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{bmatrix} \in \Re^{r \times r} \tag{3.61}$$

Furthermore, the canonical variates can be directly estimated from the past observation vector $\tilde{\mathbf{y}}_{p,k}$ as illustrated in (3.62).

$$
\begin{aligned}
\mathbf{z}_k &= \begin{bmatrix} \mathbf{b}_1^T \\ \mathbf{b}_2^T \\ \vdots \\ \mathbf{b}_r^T \end{bmatrix} \tilde{\mathbf{y}}_{pk} = \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_r^T \end{bmatrix} \mathbf{\Sigma}_{pp}^{-1/2} \tilde{\mathbf{y}}_{pk} \\
&= \mathbf{V}^T \mathbf{\Sigma}_{pp}^{-1/2} \tilde{\mathbf{y}}_{pk} = \mathbf{J} \tilde{\mathbf{y}}_{pk}
\end{aligned}
\tag{3.62}
$$

where $\mathbf{J} = \mathbf{V}^T \mathbf{\Sigma}_{pp}^{-1/2} \in \Re^{r \times mq}$ is the transformation matrix, which transforms the mq-dimensional past measurements to the r-dimensional canonical variates. These canonical variates are normalised with a unit sample covariance.

$$
\begin{aligned}
&\frac{1}{M-1} \sum_{k=q+1}^{q+M} \mathbf{z}_k \mathbf{z}_k^T \\
={}& \mathbf{V}^T \mathbf{\Sigma}_{pp}^{-1/2} \left(\frac{1}{M-1} \sum_{k=q+1}^{q+M} \tilde{\mathbf{y}}_{p,k} \tilde{\mathbf{y}}_{p,k}^T \right) \mathbf{\Sigma}_{pp}^{-1/2} \mathbf{V} \\
={}& \mathbf{V}^T \mathbf{\Sigma}_{pp}^{-1/2} \mathbf{\Sigma}_{pp} \mathbf{\Sigma}_{pp}^{-1/2} \mathbf{V} = \mathbf{V}^T \mathbf{V} = \mathbf{I}
\end{aligned}
$$

From (3.62), the canonical variate space spanned by all the estimated canonical variates can be separated into the state space and the residual space based on the

77

order of the system. According to the magnitude of the singular values, the first a dominant singular values are determined and the corresponding a canonical variates retained as the state variables where $a < r$. In addition, the remaining $(r - a)$ canonical variates are said to be in the residual space. Equation (3.63) below shows the entire canonical variate space ($\mathbf{z}_k \in \Re^r$) spanned by the state variables ($\mathbf{x}_k \in \Re^a$) and the residual canonical variates ($\mathbf{d}_k \in \Re^{r-a}$)

$$\mathbf{z}_k = \begin{bmatrix} \mathbf{x}_k{}^T & \mathbf{d}_k{}^T \end{bmatrix}^T \tag{3.63}$$

The state variables ($\mathbf{x}_k$) are a subset of the canonical variates ($\mathbf{z}_k$) estimated in (3.62). Hence the state variable like the canonical variates is defined as a linear combination of the past observation vector $\tilde{\mathbf{y}}_{p,k}$, $\mathbf{x}_k = \mathbf{J}_x \tilde{\mathbf{y}}_{p,k}$, where $\mathbf{J}_x = \mathbf{V}_x{}^T \boldsymbol{\Sigma}_{pp}{}^{-1/2}$ with $\mathbf{V}_x$ consisting of the first a columns of $\mathbf{V}$ defined in (3.58). Like the canonical variates, the state variables also have the unit covariance. Once the states of the system are determined, the state and output matrices, $\mathbf{A}$ and $\mathbf{C}$ can then be estimated through linear least squares regression. However, the determination of the state and output matrices $\mathbf{A}$ and $\mathbf{C}$ are omitted from the rest of the thesis since these matrices are not used in this work. Moreover, the state variables and the canonical variates in the residual space can both be employed for process monitoring as will be illustrated in this work.

Traditionally, it was assumed that $\boldsymbol{\varepsilon}_k$ and $\boldsymbol{\eta}_k$ in Equations (3.47) are normally distributed, as well as the state, measurement and residual vectors, $\mathbf{x}_k$, $\mathbf{y}_k$ and $\mathbf{e}_k$ since a linear combination of multivariate Gaussian variables is also normally distributed. For N samples of data, the number of samples of the states available is M, given in (3.55). For the normally distributed a-dimensional state vector, $\mathbf{x}$ with M samples, $\mathbf{x}_k$, $k = 1, 2, \cdots, M$. The T^2 statistic defined in (3.64) can be used to test whether

78

the mean μ of $\mathbf{x}$ is at the desired target τ.

$$T_k{}^2 = (\mathbf{x}_k - \tau)^T \mathbf{S}^{-1}(\mathbf{x}_k - \tau) \tag{3.64}$$

where $\mathbf{S}$ is the estimated covariance of $\mathbf{x}$. If $\mu = \tau$, then $CT^2 \sim F(a, M - a)$, where $C = \frac{(M-1)(M+1)a}{M(M-a)}$. Equation (3.64) can be simplified as the state covariance matrix, $\mathbf{S} = \mathbf{I}$. Furthermore, since the past and future observations, $\tilde{\mathbf{y}}_{p,k}$ and $\tilde{\mathbf{y}}_{f,k}$ have zero means, the desired target for the state is $\tau = 0$. With these simplifications in place, the T^2 metric for the state space is represented in (3.65).

$$T_{kcva}^2 = \mathbf{x}_k{}^T \mathbf{x}_k \tag{3.65}$$

In addition, the Q metric is introduced for the CVA approach to test the significance level of the prediction error represented in the scaled past observation space. According to (3.62), the prediction error for the scaled past measurement and the corresponding Q meric are then defined in (3.66) and (3.67) respectively.

$$\mathbf{e}_k = (\mathbf{I} - \mathbf{V}_x \mathbf{V}_x^T)\mathbf{\Sigma}_{pp}^{-1/2}\tilde{\mathbf{y}}_{p,k} = \mathbf{F}\tilde{\mathbf{y}}_{p,k} \tag{3.66}$$

where $\mathbf{F}$ is the states excluded to the residual space.

$$Q_{kcva} = \mathbf{e}_k{}^T \mathbf{e}_k \tag{3.67}$$

Generally, statistical monitoring methods consist of first developing a statistical model and then determining control limits based on monitoring metrics to judge whether or not the processes are in control. The T^2 and Q metrics are the most commonly employed metrics for process monitoring. The T^2 metric represents the variations in the model space while the Q metric represents the variations in the residual space. In this work, the T^2 and Q metrics have been derived for each of the

DPCA, DPLS and CVA methods.

For the purpose of process monitoring, the T^2 metric requires a corresponding control limit, T^2_{UCL} to be derived and employed for the assessment of the process. In the following section, a T^2 control limit is derived to correspond to the T^2_{dpca}, T^2_{dpls} and T^2_{kcva} metrics derived for the DPCA, DPLS and CVA approaches. Similarly, the Q metric requires a corresponding control limit to judge the status of the process. In this work, a corresponding control limit for the Q metric is derived to correspond to the Q_{dpca}, Q_{dpls} and Q_{kcva} metrics derived for the DPCA, DPLS and CVA approaches.

3.5 Upper Control Limits

In this section, upper control limits (UCL) are derived for the T^2 and Q metrics under the normal distribution assumption. This is because the efficiency of a monitoring approach greatly depends on the appropriateness of the control limit as will be illustrated in later section of this work. This means that the condition monitoring performance of the monitoring method can be enhanced by the use of an appropriate control limit.

For the T^2 metric, if the latent variables are normally distributed, the corresponding UCL $T^2_{\mathrm{UCL}}(\alpha)$ for a significance level is derived as in (3.68).

$$T_{\mathrm{UCL}}{}^2(\alpha) = \frac{a(M-1)^2}{M(M-a)} F_{a,M-a}(\alpha) \tag{3.68}$$

where $F_{u,v}(\alpha)$ is the critical value of the F-distribution with u and v degrees of freedom for a significance level α. By comparing $T_k{}^2$ against $T_{\mathrm{UCL}}{}^2(\alpha)$ in real-time, an abnormal condition is then determined when $T_k{}^2 > T_{\mathrm{UCL}}(\alpha)$.

Therefore, the systems (3.2), (3.15) and (3.47) can be monitored by plotting their

corresponding $T_k{}^2$ against time, k, along with a UCL, $T^2{}_{\text{UCL}}(\alpha)$ corresponding to a significance level, α, that has the probability, $P(T_k{}^2 > T^2{}_{\text{UCL}}(\alpha)) = \alpha$.

Furthermore, given a level of significance, α, and based on the assumption of normality, the threshold, $Q_{\text{UCL}}(\alpha)$ of the Q-metric is estimated as shown below [67].

$$Q_{\text{UCL}}(\alpha) = \theta_1 \left(\frac{h_0 c_\alpha \sqrt{2\theta_2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1{}^2} \right)^{\frac{1}{h_0}} \tag{3.69}$$

where $\theta_i = \sum_{j=n+1}^{r} \lambda_j{}^i$, $h_0 = 1 - \frac{2\theta_1 \theta_3}{3\theta_2{}^2}$ and c_α is the normal deviate corresponding to $(1 - \alpha)$ percentile. In (3.69), λ_j is the eigenvalue of the covariance of the measured data. For the CVA error represented in (3.66), it should be the covariance of the scaled past observations, $\boldsymbol{\Sigma}_{pp}^{-1/2} \tilde{\mathbf{y}}_{p,k}$, $i.e.$

$$\lambda_j = \lambda \left(\sum_{k=q+1}^{q+M} (\boldsymbol{\Sigma}_{pp}^{-1/2} \tilde{\mathbf{y}}_{p,k})(\boldsymbol{\Sigma}_{pp}^{-1/2} \tilde{\mathbf{y}}_{p,k})^T \right) = 1$$
$$j = 1, 2, \cdots, r$$

Therefore, the calculation $Q_{\text{UCL}}(\alpha)$ can be simplified by letting $\theta_i = (r - a)$ and $h_0 = 1/3$ in (3.69). By comparing Q_k against $Q_{\text{UCL}}(\alpha)$ in real-time, an abnormal condition is determined when $Q_k > Q_{\text{UCL}}(\alpha)$.

Unfortunately, both control limits derived in (3.68) and (3.69) are based on the assumptions that the latent variables from the DPCA, DPLS and CVA approaches as well as their prediction errors are Gaussian. However, when the collective modelling errors of the systems (3.2), (3.15) and (3.47) are non-Gaussian processes, this assumption is not valid. Hence, $T^2{}_{\text{UCL}}(\alpha)$ and $Q_{\text{UCL}}(\alpha)$ derived in (3.68) and (3.69) can no longer be used as control limits for non-linear dynamic process monitoring.

One solution to this issue is to estimate the PDF directly for these T^2 and Q metrics through a non-parametric approach [13, 135]. Amongst various PDF estimating approaches, the kernel density estimation (KDE) approach [13, 135] is selected for

this work. The KDE algorithm is discussed in the following section.

3.6 Control Limit through Kernel Density Estimation

The KDE is a well established approach to estimate the PDF particularly for uni-variate random processes [14]. Therefore, it is particularly suitable for the T^2 and Q metrics which are univariate although the underlying processes are multivariate. Assume x is a random variable and its density function is denoted by $p(x)$. This means that

$$P(x < b) = \int_{-\infty}^{b} p(x)dx \tag{3.70}$$

Therefore, by knowing $p(x)$, an appropriate control limit can be determined for a specific confidence bound, α using (3.70). The estimation of the probability density function $\hat{p}(x)$ at point x through the kernel function, $K(\cdot)$ is defined as follows:

$$\hat{p}(x) = \frac{1}{Mh}\sum_{k=1}^{M}K\left(\frac{x - x_k}{h}\right) \tag{3.71}$$

where x_k, $k = 1, 2, \cdots, M$ are samples of x and h is the bandwidth. A rough estimation of the optimal bandwidth h_{opt} subject to minimising the approximation of the mean integrated square error can be derived in (3.72), where σ is the standard deviation [137].

$$h_{opt} = 1.06\sigma N^{-1/5} \tag{3.72}$$

By replacing x_k with the $T_k{}^2$ metrics obtained in Equations (3.11), (3.43) and (3.65) for the DPCA, DPLS and CVA methods, the above KDE approach is able to estimate the underlying PDFs of the T^2 metric. Also by replacing x_k with the Q_k metrics obtained in Equations (3.13), (3.45) and (3.67) for the DPCA, DPLS and CVA

methods, the above KDE approach is also able to estimate the underlying PDFs of the Q metric. The corresponding control limits, $T^2_{\text{UCL}}(\alpha)$ and $Q_{\text{UCL}}(\alpha)$ can then be obtained from the PDFs of the T^2 and Q metrics for a given confidence bound, α by solving the following equations respectively.

$$\int_{-\infty}^{T^2_{\text{UCL}}(\alpha)} p(T^2)dT^2 = \alpha$$
$$\int_{-\infty}^{Q_{\text{UCL}}(\alpha)} p(Q)dQ = \alpha \tag{3.73}$$

The T^2 and Q metrics are complementary. A fault may cause a significant deviation in the state space but not necessary results in a similar level of significance in the error space, vice versa. Therefore, in this work, a fault is then identified ($F_k = 1$) if either $T^2_k > T^2_{\text{UCL}}(\alpha)$ or $Q_k > Q_{\text{UCL}}(\alpha)$ conditions are satisfied, i.e.

$$F_k = (T^2_k > T^2_{\text{UCL}}(\alpha)) \oplus (Q_k > Q_{\text{UCL}}(\alpha)) \tag{3.74}$$

where $\oplus$ represents a logical OR operation. By using the fault detection condition (3.74), the monitoring performance becomes insensitive to the number of states, a since any ignored variances in the T^2 metric by reducing n will be recovered by Q metric.

By summarising the analysis presented in the previous sections, a new extension of CVA using KDEs for non-linear dynamic process monitoring is proposed to identify underlying faults subject to non-Gaussian processes. Similarly, by combining the DPCA with the monitoring metrics and the KDE algorithms, another extension of the DPCA, the novel DPCA with KDE is developed. In addition, the DPLS is extended using the KDE algorithm, resulting in the novel DPLS with KDE approach. The step by step procedure of these KDE approaches are presented in the following sections.

3.6.1   DPCA with KDE

In this section the DPCA is extended to non-linear systems using the KDE. The proposed DPCA with KDE approach consists of first estimating PCs using the standard DPCA before deriving appropriate control limits using the KDE.

DPCA with KDE Algorithm

1. Obtain the training data from the normal operating process

2. Construct the augmented data matrix as in Equation (3.5)

3. Normalise the augmented data matrix as in Equation (3.6)

4. Estimate the covariance matrix and perform singular value decomposition as in Equation (3.7)

5. Estimate the PCs as in Equation (3.8) and determine the order to retain in the model space, while excluding the rest to the residual space.

6. Estimate T^2 and Q metrics as in Equations (3.11) and (3.13)

7. Derive control limits using KDE as in Equation (3.73)

On-line Monitoring

1. Collect real-time monitoring data and apply normalisation or scaling as that used in the modelling

2. Estimate the PCs by projecting the scaled data to be monitored onto the loading matrix as in Equation (3.8)

3. Calculate the T^2 and Q metric as in Equations (3.11) and (3.13).

4. Use the determined KDE control limits to judge whether or not the process is in control

3.6.2 DPLS with KDE

The standard DPLS is extended to non-linear systems in this section. The proposed DPLS with KDE consists of the standard DPLS and KDE approaches. The procedure of the proposed DPLS with KDE is enumerated below.

DPLS with KDE Algorithm

1. Obtain the training data from the normal operating process and determine the independent and dependent variables ($\mathbf{X}$) and ($\mathbf{Y}$) respectively.

2. Construct the augmented data matrices for the ($\mathbf{X}$) and ($\mathbf{Y}$) as in Equations (3.17) and (3.18) respectively

3. Normalise the augmented data matrices as in Equations (3.19) and (3.20) respectively

4. Estimate the score and loading matrices of ($\mathbf{X}$) and ($\mathbf{Y}$) as in Equations (3.24) to (3.33)

5. Estimate the latent variables as in Equation (3.32)

6. Determine the number of latent variables to retain using cross validation

7. Estimate T^2 and Q metrics as in Equations (3.43) and (3.45)

8. Derive control limits using KDE as in Equation (3.73)

On-line Monitoring

1. Collect real-time monitoring data

2. Estimate the latent variables by projecting the monitoring data onto the weight vector as in Equation (3.25)

3. Determine the T^2 and Q metrics as in Equation (3.43) and (3.45) respectively

4. Use the determined KDE control limits to judge whether or not the process is in control

3.6.3 CVA with KDE

The CVA with KDE consists of first of employing the CVA for the extraction of the states and then deriving appropriate control limits using the KDE. The steps of the proposed CVA with KDE are listed below.

CVA with KDE Algorithm

1. Obtain the training data from the normal operating process

2. Construct the past and future observation vectors as in Equations (3.48) and (3.49).

3. Determine the maximum number of columns in the truncated hankel matrix as in Equation (3.55)

4. Determine the past and future truncated hankel matrices as in Equations (3.53) and (3.54)

5. Estimate the scaled hankel matrix and perform the singular value decomposition as in Equation (3.58)

6. Estimate the canonical variates as in Equation (3.62)

7. Separate the estimated canonical variates into the state and residual spaces based on the order of the system as in Equation (3.63)

8. Estimate T^2 and Q metrics as in Equations (3.64) and (3.67)

9. Derive control limits using KDE as in Equation (3.73)

On-line Monitoring

1. Collect real-time monitoring data

2. Determine the past and future observation vectors as in Equations (3.48) and (3.49)

3. Derive the canonical variates for the state and residual spaces

4. Determine T^2 and Q metrics from the esimated canonical variates based on the monitored data

5. Use the determined KDE control limits to judge whether or not the process is in control

All three KDE-based approaches developed in this work are applied to the Tenneessee Easman Process Plant, which is described in the following section.

3.7   Tennessee Eastman Process Plant

Ideally, monitoring techniques should be applied to real industrial plants. Unfortunately, such plants are not readily accessible. As a result, most research work in process monitoring is based on computer simulation of industrial processes. For this reason, the monitoring algorithms developed in this work are evaluated using the Tennessee Eastman Process (TEP) plant.

The TEP process is a large dimensional, non-linear process with unknown mathematical representation as the simulation is intentionally distributed as an undocumented FORTRAN program [92]. It was originally created by the Eastman chemical Company to provide a realistic industrial process for the evaluation of process monitoring approaches. The TEP plant [92] has 5 main units which are the reactor, condenser, separator, stripper and compressor [21, 92]. Streams of the plant consists of 8 components; A, B, C, D, E, F, G and H. Components A, B and C are

gaseous reactants which were fed to the reactor to form products G and H. The TEP data used for this work consists of two blocks; the training and test data blocks. Each block has 21 data sets corresponding to the normal operation (Fault 0) and 20 fault operations (Fault 1 - Fault 20). The sampling time for most of the process variables in the TEP plant is 3 minutes. A total of 52 measurements are collected for each data set of length, $N = 960$ representing 48-hour operation with a sampling rate of 3 minutes. However, 19 of the 52 measurements, 14 of them sampled at 6 minute interval and 5 of them sampled in every 15 minutes, have not been included in the DPCA and CVA studies due to the measurement time delay. Nevertheless, these variables were employed as the response variables for the work on DPLS. Different from the work reported by Chiang [21], 11 manipulated variables are treated the same as other measured variables because under feedback control, these variables are not independent any more. The simulation time of each operation run in the test data block is 48 hours and the various faults are introduced only after 8 hours. This means that for each of the faults, the process is in-control for the first 8 simulation hours before the process gets out of control at the introduction of the fault. All 20 faults have been studied in this work. Also in this work, the normal operating process data will be referred to as the training data. The test data block is based on a simulation of a real-time industrial process where the operating conditions, components and kinetics are modified for proprietary reasons. Furthermore, the analysis in this research work is based on the test data blocks alone. Also, a graphical presentation of the TEP plant is given below with the 5 main units high-lighted.

The gaseous reactant A is fed to the reactor through stream 1 while gaseous reactants A and C along with the inert B are fed to the reactor through stream 4. Also, gaseous reactants D and E are fed to the reactor through streams 2 and 3 respectively. The reactions that take place in the reactor result in the formation of the liquid products G and H, whereas F is the by-product. The TEP reactions are shown below;
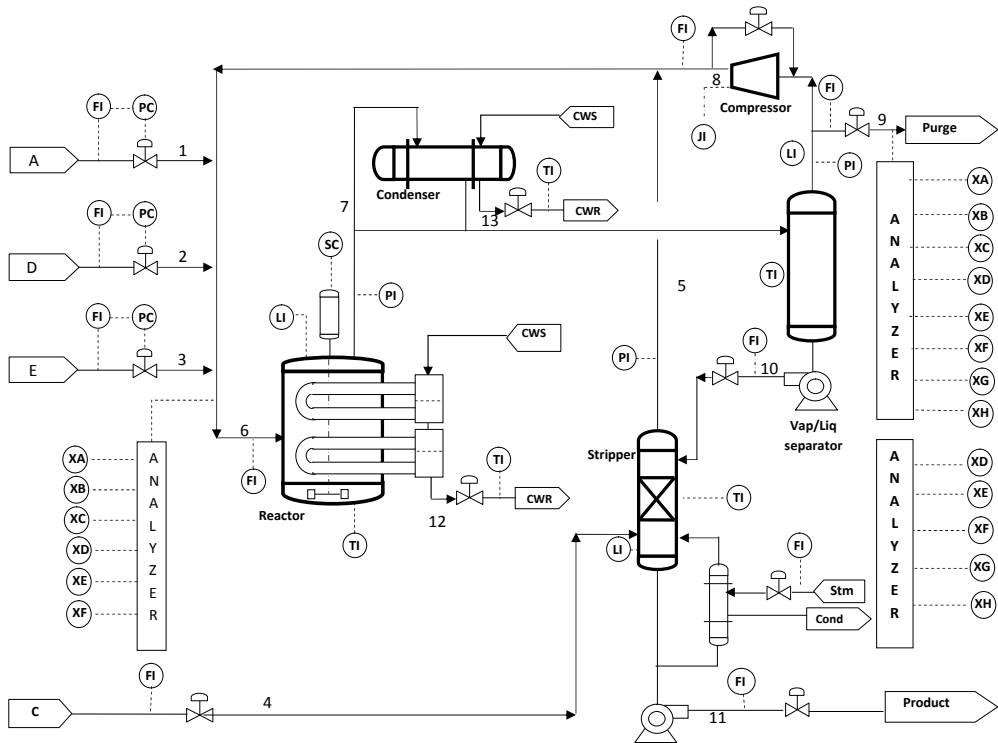
Figure 3.1: Graphical description of TEP plant

$$A(g) + C(g) + D(g) \rightarrow G(liq)$$

$$A(g) + C(g) + E(g) \rightarrow H(liq)$$

$$A(g) + E(g) \rightarrow F(liq)$$

$$3D(g) \rightarrow 2F(liq) \tag{3.75}$$
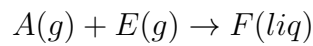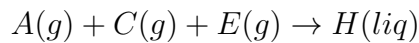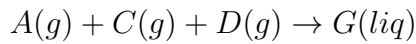
The product from the reactor is cooled through a condenser before it is fed to a vapour-liquid separator. The vapour from the separator is then recycled to the feed via the compressor. Also, some of the recycle stream is purged via stream 9 to avoid the inert and by-product from accummulating in the process. The condensed com-

ponents from the separator in stream 10 is pumped to the stripper. In the stripper, stream 4 strips the remaining reactants from stream 10, that are combined with the recycle stream through stream 5. Generally, for fault detection and identification problems [4, 5], the TEP plant is considered under a closed-loop control as described in [92].

3.7.1 Tennessee Eastman Process Variables

The TEP variables mentioned above are described in Table 3.1. However, only 11 of the 12 manipulated variables (MVs) and the 41 measured variables are described in Table 3.1 because the agitator speed which is supposed to be the 12th manipulated variable is not actually manipulated.

3.7.2 Tennessee Eastman Process Faults

The TEP plant has 21 scenarios corresponding to Faults $0 - 20$, with Fault 0 being the data simulated at normal operating condition (no fault) and Faults 1 - 20 corresponding to data sets from the simulated fault processes, each with a specified fault as listed in Table 3.2.

3.8 Monitoring Performance

The monitoring performance in this study is assessed based on the percentage reliability which is defined as the percentage of the samples outside the control limits [144] within the last 40 hour faulty operation. Hence a monitoring technique is said to be better than another technique if the percentage reliability of this technique is numerically higher than the percentage reliability of another. Also, the monitoring

Table 3.1: Tennesseee Eastman Process variables

ID	Description	ID	Description
x_1	A Feed (Stream 1)	x_{27}	Component E (Stream 6)
x_2	D Feed (Stream 2)	x_{28}	Component F (Stream 6)
x_3	E Feed (Stream 3)	x_{29}	Component A (Stream 9)
x_4	Total Feed (Stream 4)	x_{30}	Component B (Stream 9)
x_5	Recycle Flow (Stream 8)	x_{31}	Component C (Stream 9)
x_6	Reactor Feed Rate (Stream 6)	x_{32}	Component D (Stream 9)
x_7	Reactor Pressure	x_{33}	Component E (Stream 9)
x_8	Reactor Level	x_{34}	Component F (Stream 9)
x_9	Reactor Temperature	x_{35}	Component G (Stream 9)
x_{10}	Purge Rate (Stream 9)	x_{36}	Component H (Stream 9)
x_{11}	Product Separator Temperature	x_{37}	Component D (Stream 11)
x_{12}	Product Separator Level	x_{38}	Component E (Stream 11)
x_{13}	Product Separator Pressure	x_{39}	Component F (Stream 11)
x_{14}	Product Separator Underflow (Stream 10)	x_{40}	Component G (Stream 11)
x_{15}	Stripper Level	x_{41}	Component H (Stream 11)
x_{16}	Stripper Pressure	x_{42}	MV to D Feed Flow (Stream 2)
x_{17}	Stripper Underflow (Stream 11)	x_{43}	MV to E Feed Flow (Stream 3)
x_{18}	Stripper Temperature	x_{44}	MV to A Feed Flow (Stream 1)
x_{19}	Stripper Steam Flow	x_{45}	MV to Total Feed Flow (Stream 4)
x_{20}	Compressor Work	x_{46}	Compressor Recycle Valve
x_{21}	Reactor Cooling Water Outlet Temperature	x_{47}	Purge Valve (Stream 9)
x_{22}	Separator Cooling Water Outlet Temperature	x_{48}	MV to Separator Pot Liquid Flow (Stream 10)
x_{23}	Component A (Stream 6)	x_{49}	MV to Stripper Liquid Product Flow (Stream 11)
x_{24}	Component B (Stream 6)	x_{50}	Stripper Steam Valve
x_{25}	Component C (Stream 6)	x_{51}	MV to Reactor Cooling Water Flow
x_{26}	Component D (Stream 6)	x_{52}	MV to Condenser Cooling Water Flow

Table 3.2: Brief description of TEP plant faults

Fault	Description	Type
1	A/C Feed Ratio, B Composition Constant (Stream 4)	Step
2	An increase in B while A/C Feed ratio is constant (stream 4)	Step
3	D Feed Temperature (Stream 2)	Step
4	Reactor Cooling Water Inlet Temperature	Step
5	Condenser Cooling Water Inlet Temperature	Step
6	A loss in Feed A (stream 1)	Step
7	C Header Pressure Loss - Reader Availability (Stream 4)	Step
8	A,B,C Feed Composition (Stream 4)	Random variation
9	D Feed Temperature (Stream 2)	Random variation
10	C Feed Temperature (Stream 4)	Random variation
11	Reactor Cooling Water Inlet Temperature	Random variation
12	Condenser Cooling Water Inlet Temperature	Random variation
13	Reaction Kinetics	Slow drift
14	Reaction Cooling Water Valve	Sticking
15	Condenser Cooling Water Valve	Sticking
16	Unknown	Unknown
17	Unknown	Unknown
18	Unknown	Unknown
19	Unknown	Unknown
20	Unknown	Unknown

performance is assessed by the detection delay which is the time period it takes to detect a fault after the introduction of the fault. The false alarm rate was also investigated although it has not been reported because all the KDE associated approaches discussed in this chapter had zero false alarm rates. The 99% confidence interval is adopted in this study.

The variability of the training data is characterised by the extracted canonical variate state space model. Firstly, the number of time lags for past and future observations is determined from the autocorrelation function of the summed squares of all measurements as shown in Figure 3.2 against $\pm 5\%$ confidence bounds. The autocorrelation function indicates that the maximum number of significant lags in this study is 16. Hence both p and f are set to 16. The length of the past and future observations (mq) is 528 according to (3.48) and (3.49). The number of columns of the truncated Hankel matrices according to (3.55) is $M = 929$. The singular value decomposition is then performed on the scaled Hankel matrix as in (3.58).
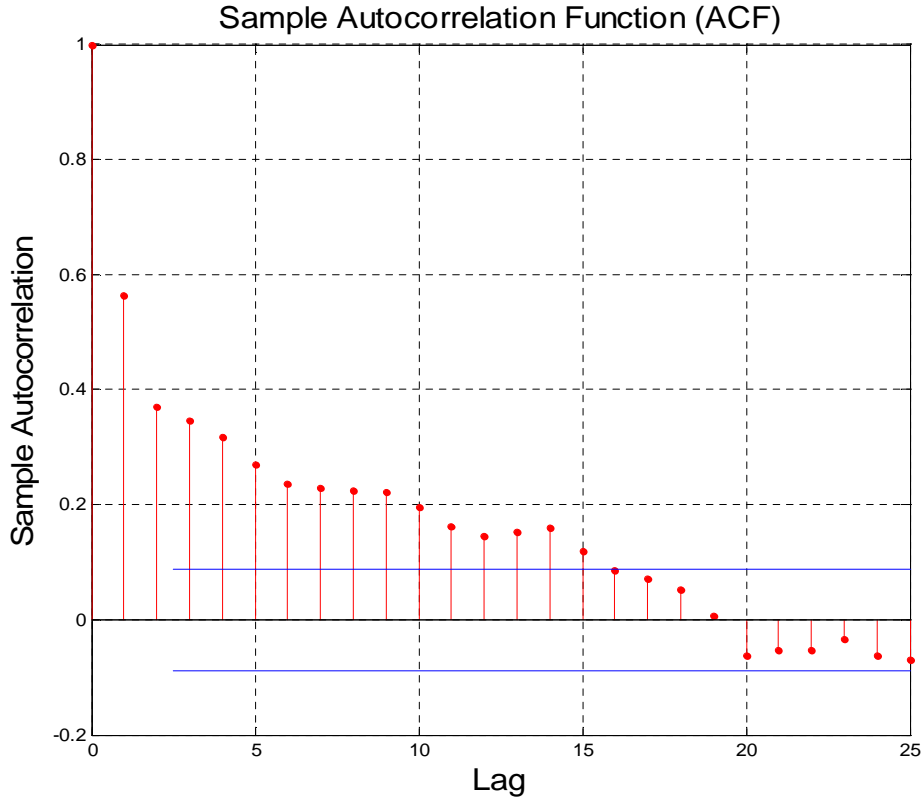
Figure 3.2: Autocorrelation function of the summed squares of all measurements

Several ways have been suggested to determine the order (a) of the system for CVA based approaches amongst which the dominant singular values [3, 21] and the Akaike Information Criterion (AIC) are most widely adopted. The former method was adopted in this study to determine the order of the system. The singular values from the scaled Hankel were normalised to have the values ranging between 0 and 1 and then the order determined based on the dominant normalised singular values. For the TEP case study, it is noticed that the singular values of the scaled Hankel matrix **H** in (3.58) decrease slowly. If a is determined from these singular values, it will be unrealistically large as indicated in Figure 3.3, which shows the normalised sum of squares of residual singular values against the number of states. As mentioned already, the value of a is not sensitive to monitoring performance for this work due to the fault detection condition (3.74) adopted. Hence, a more realistic number of singular values, $a = 26$ represented by circles in Figure 3.3 is
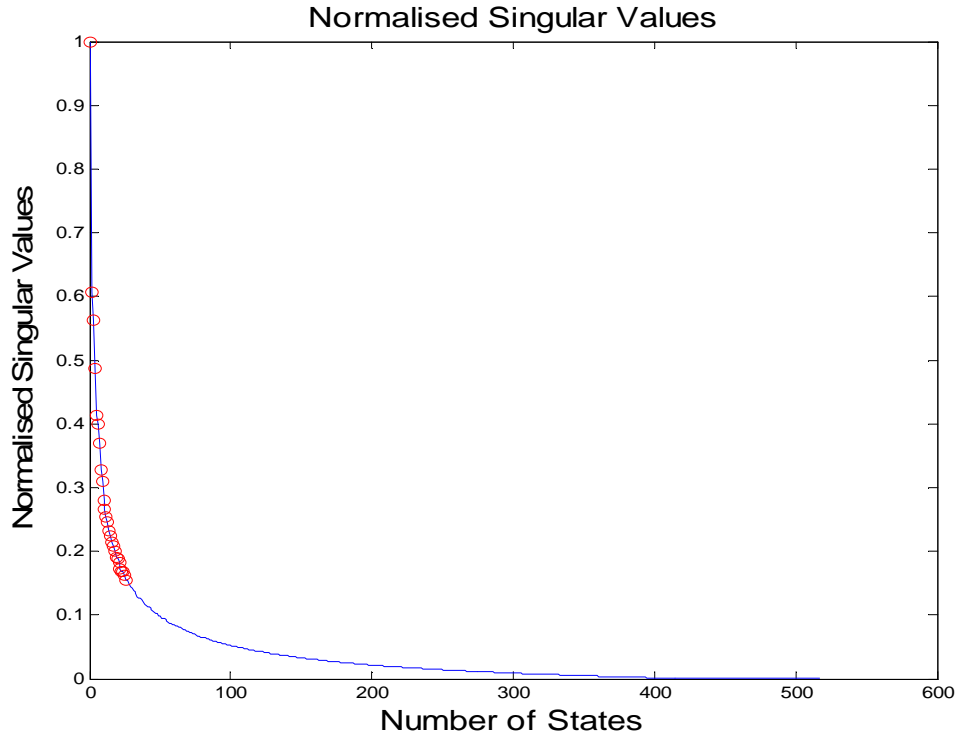
Figure 3.3: Normalised singular values from the scaled hankel matrix

employed to represent the model space. Also, to make a fair comparison of the proposed KDE-based approaches developed in this chapter one with another as well as with their non-KDE counterparts, the number of lag and the order to determine the dimension of the latent variables are the same for all the approaches developed and considered in this chapter. In addition, the monitoring criterion is also the same for the developed KDE-based approaches as well as their non-KDE counterparts.

3.9   Comparison of Monitoring Approaches

In this section, the monitoring performances of the approaches considered in this chapter are presented. The reliabilities of all KDE and non-KDE approaches considered in this chapter are presented in Table 3.3 while their detection delays are presented in Table 3.4.

Table 3.3: Reliability (%) comparison

Fault IDs	CVA with KDE	CVA	DPCA with KDE	DPCA	DPLS with KDE	DPLS
1	99.75	99.75	99.38	99.25	99.25	99.25
2	99.5	98.5	98	97.88	98.13	98
3	73.03	37.2	0	0	0.2497	0
4	99.88	99.88	99.88	99.88	99.88	99.88
5	99.88	99.88	29.09	27.84	28.21	26.47
6	99.88	99.88	99.88	99.88	99.88	99.88
7	99.88	99.88	99.88	99.88	99.88	99.88
8	99.88	98.75	97.25	97.13	97	97
9	92.26	75.28	0.2497	0	0.2497	0
10	96.63	96.25	39.08	28.21	36.83	29.46
11	99.38	99.38	99.88	98.63	97.88	97.75
12	99.5	99.5	98.13	98.13	98	98
13	96.13	96.13	95.01	95.01	94.76	94.76
14	99.88	99.75	99.75	99.75	99.75	99.75
15	99.5	99.5	0.1248	0	0.1248	0
16	99.13	99.13	35.83	26.22	26.97	21.6
17	98.13	98.13	97.75	97.75	97.75	97.75
18	99.25	99.25	98.63	98.5	98.63	98.5
19	99.88	99.88	90.51	87.02	84.64	79.28
20	97.63	97.25	79.15	76.9	73.91	71.41

Furthermore, the monitoring performances of each KDE approach is compared with its non-KDE counterpart and then all the KDE approaches are also compared.

3.9.1 Monitoring Performance of DPCA with KDE

In this section, the proposed DPCA with KDE and the traditional DPCA are both applied to the simulated TEP data. To make a fair comparison of both approaches, the order to determine the principal components are the same. In this work, 26 PCs were retained in the model space, while the maximum number of significant lags was 16.

The results presented in Table 3.3 above show that for 12 of the TEP faults (1, 2, 5,

Table 3.4: Detection delay (minute) comparison

Fault IDs	CVA with KDE	CVA	DPCA with KDE	DPCA	DPLS with KDE	DPLS
1	9	9	18	21	21	21
2	15	15	51	54	48	51
3	15	39	-	-	1125	-
4	6	6	6	6	6	6
5	6	6	12	12	12	12
6	6	6	6	6	6	6
7	6	6	6	6	6	6
8	30	33	69	72	75	75
9	33	45	2115	-	1125	-
10	84	93	210	210	219	219
11	18	18	24	24	24	24
12	15	15	48	48	51	51
13	96	96	123	123	129	129
14	6	9	9	9	9	9
15	15	15	1140	-	1125	-
16	24	24	111	111	216	219
17	48	48	57	57	57	57
18	21	21	36	39	36	42
19	6	6	36	39	36	42
20	60	69	120	123	123	123

8, 9, 10, 11, 15, 16, 18, 19 and 20), the proposed DPCA with KDE is able to detect more of the fault than the traditional DPCA while for the remaining 8 faults (3, 4, 6, 7, 12, 13, 14 and 17) the detections by both approaches are equal. Furthermore, for 6 of the TEP faults (1, 2, 8, 18, 19 and 20) the DPCA with KDE is able to slightly improve the monitoring performance over the traditional DPCA while for 2 faults (9 and 15), the improvement of the DPCA with KDE over the traditional DPCA in terms of detection delay is significant. Also, for the remaining 12 faults (3, 4, 5, 6, 7, 10, 11, 12, 13, 14, 16 and 17), both approaches achieved equal detection delays.

The efficiency of the proposed DPCA with KDE over the traditional DPCA is due to the fact that the DPCA with KDE takes the non-linearity of the plant into account

which the traditional DPCA does not do. Also, the superiority of the DPCA with KDE over the traditional DPCA is emphasised for faults that are non-linear in nature. This means that for linear faults, the proposed DPCA with KDE approach may not demonstrate advantages over the traditional DPCA. This explains why the performance of the DPCA with KDE is the same as the traditional DPCA, for some of the TEP faults. Furthermore, the DPCA with KDE approach and the traditional DPCA both have zero false alarm rates for all the TEP faults employed in this study.

3.9.2   Monitoring Performance of DPLS with KDE

In this section, the proposed DPLS with KDE and the traditional DPLS i.e. with control limits derived based on the Gaussian assumption are both applied to the TEP data. To make a fair comparison for both approaches, the same measurements were selected as the predictor and response variables based on their sampling rates. 33 of the 52 TEP measurements with a sampling rate of 3 minutes were selected as the predictor variables while the remaining 19 measurements, 14 of them sampled at 6 minutes interval and 5 of them sampled at 15 minutes interval were employed as the response variables. Also, 26 latent variables were selected for the model spaces of both approaches.

For 11 of the faults (2, 3, 5, 9, 10, 11, 15, 16, 18, 19 and 20), the DPLS with KDE approach has a better reliability than the traditional DPLS while for the remaining 9 faults (1, 4, 6, 7, 8, 12, 13, 14 and 17), the reliabilities of both approaches considered in this section are the same. Furthermore, for 4 of the faults (2, 16, 18 and 19), the DPLS with KDE is able to slightly improve the detection delay while the improvement in detcetion delay is significant for 3 faults (3, 9 and 15). Also, the DPLS with KDE and the traditional DPLS have equal detection delays for the remaining 9 faults (1, 4, 6, 7, 8, 12, 13, 14 and 17). The superiority of the DPLS with KDE over the traditional DPLS is emphasised for faults that are non-linear in

nature. This explains why the performance of the DPLS with KDE is the same as the traditional DPLS, for some of the TEP faults. Both approaches have zero false alarm rates for all twenty TEP faults employed in the current work.

3.9.3 Monitoring Performance of CVA with KDE

In this section, the CVA with KDE and traditional CVA are both applied to the TEP data. The results presented in Table 3.3 above show that for 7 of the faults (2, 3, 8, 9, 10, 14 and 20), the proposed CVA with KDE has higher reliabilities than the traditional CVA, while for the remaining 13 faults (1, 4, 5, 6, 7, 11, 12, 13, 15, 16, 17, 18 and 19) the reliabilities of both approaches considered in this section are equal. Moreover, the CVA with KDE has lower detection delays for 6 of the faults (3, 8, 9, 10, 14 and 20), while for the remaining 14 faults (1, 2, 4, 5, 6, 7, 11, 12, 13, 15, 16, 17, 18 and 19), both approaches considered in this section have equal detection delays. In addition, both approaches considered in this section have zero false alarm rates for all twenty faults employed in this work. The proposed CVA with KDE is able to improve the monitoring performance over the traditional CVA. Moreover, the superiority of the proposed CVA with KDE over the traditional CVA is particularly emphasised for those faults (3, 9 and 15) that are generally difficult to detect by most monitoring techniques.

3.9.4 Comparison of KDE Approaches

In this section, the monitoring performances of the KDE approaches developed in this chapter are compared. To make a fair comparison amongst the KDE approaches, the number of lags and the order to determine the dimension of the latent variables are the same for all the KDE approaches. The CVA with KDE is able to improve the reliability for 17 of the faults (1, 2, 3, 5, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18,

19 and 20) over the DPCA with KDE while for the remaining 3 faults (4, 6 and 7) the reliability of the DPCA with KDE is the same as that of the CVA with KDE. In addition, in terms of detection delays, the CVA with KDE is also able to improve the monitoring performance over the DPCA with KDE. For 17 of the faults (1, 2, 3, 5, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 and 20), the CVA with KDE is able to detect the faults earlier than the DPCA with KDE while for the remaining 3 faults (4, 6 and 7), the detection delays of the CVA with KDE and the DPCA with KDE are the same.

The CVA with KDE is also able to improve monitoring performance over the DPLS with KDE in terms of reliabilities. For 17 of the faults (1, 2, 3, 5, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 and 20), the CVA with KDE has a better reliability than the DPCA with KDE while for the remaining 3 faults (4, 6 and 7), the reliability of the CVA with KDE is the same as that of the DPCA with KDE. In addition, the CVA with KDE is able to detect the faults earlier than the DPLS with KDE for 17 faults (1, 2, 3, 5, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 and 20), while for the remaining 3 faults, the detection delays of the CVA with KDE is the same as that of the DPLS with KDE.

The results show that the DPCA with KDE has a better reliability over the DPLS with KDE for 9 faults (5, 8, 10, 11, 12, 13, 16, 19 and 20) while for 3 of the faults (1, 2 and 3), the DPLS with KDE has a better reliability than the DPCA with KDE. For the remaining 8 faults the reliabilities of the DPCA with KDE and the DPLS with KDE are the same. For 7 of the faults (1, 8, 10, 12, 13, 16 and 20), the DPCA with KDE is able to detect the faults earlier than the DPLS with KDE, while for 4 of the faults (2, 3, 9 and 15), the DPLS with KDE is able to detect the faults earlier than the DPCA with KDE. The detection delays of the DPCA with KDE and the DPLS with KDE are the same for 9 faults (4, 5, 6, 7, 11, 14, 17, 18 and 19).

For most of the faults considered in this work, the CVA with KDE outperformed the

DPCA with KDE and DPLS with KDE approaches. The superiority of the CVA with KDE is particularly emphasised for faults such as 3, 9 and 15, that are difficult to monitor by most monitoring algorithms.

To further illustrate the superiority of the proposed CVA with KDE, the T^2 and Q monitoring charts of all the proposed approaches for Fault 9 are presented in Figure 3.4. In Figure 3.4, sub-figures in the left column and the right column are for the T^2 and Q charts respectively; whilst the first, second and third rows are for CVA, DPCA and DPLS approaches respectively. Upper control limits obtained based on the Gaussian assumption are represented as dash-dot lines, whilst the UCLs determined by the KDE approach are shown in dashed lines.
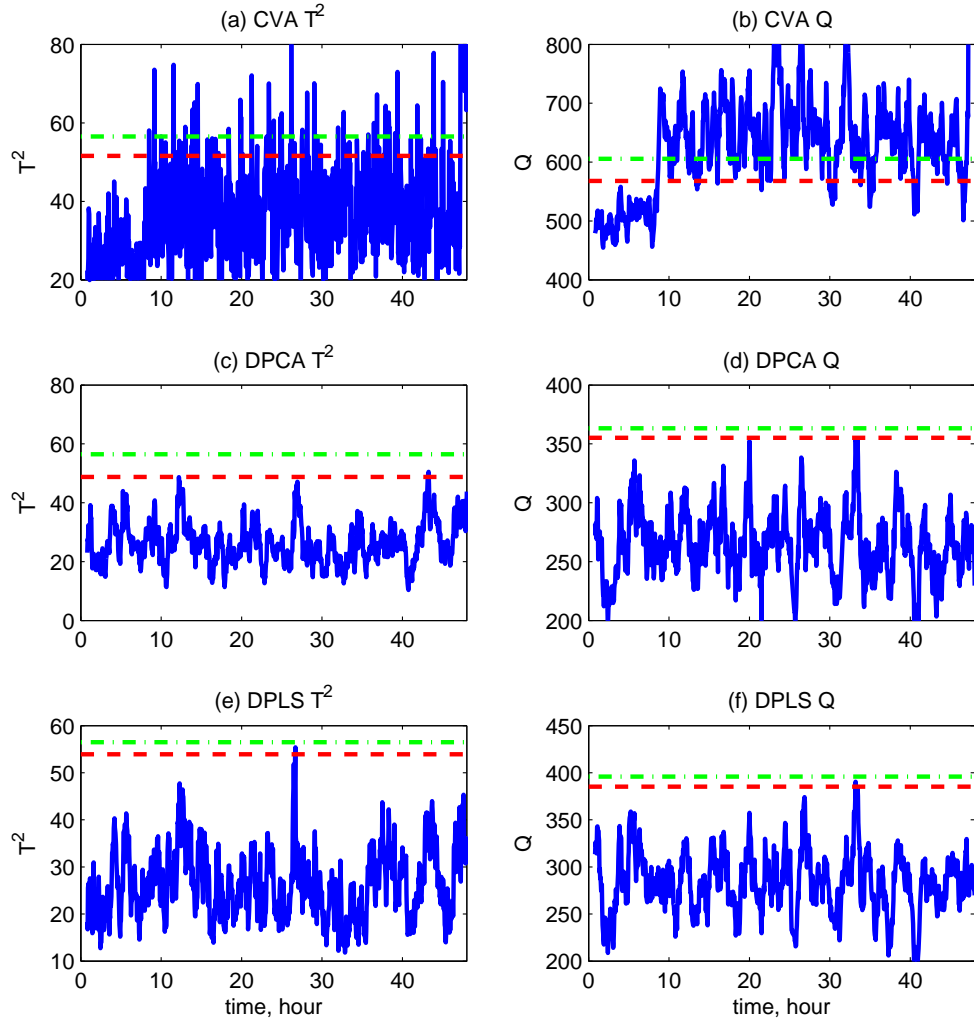
Figure 3.4: Monitoring charts for Fault 9

Figure 3.4 clearly indicates that only the CVA model is able to reveal the difference in dynamic behaviour between the normal operation and the operation with Fault 9. This is because the CVA approach is a better way than the DPCA and DPLS approaches to capture dynamic behaviours [3, 4, 5]. Both T^2 and Q metrics produced by the DPCA and the DPLS approaches have no identifiable difference between the normal and faulty operations. Furthermore, the CVA with KDE approach gives tighter UCLs for both metrics resulting in a higher percentage of reliability and

earlier fault detection than the traditional CVA approach.

To further justify the employment of the control limits derived using the KDE in this work, the distribution of some of the derived latent variables are illustrated in Figure 3.5.
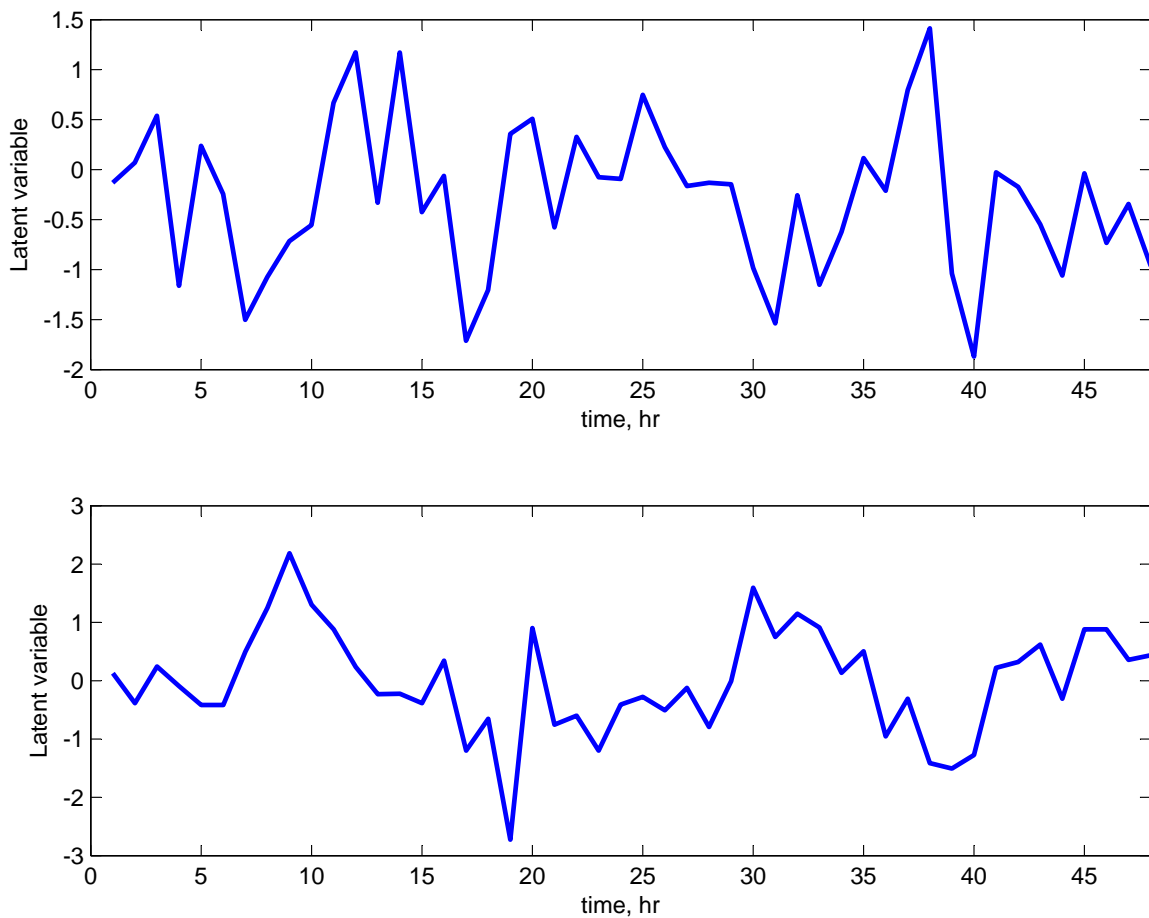
Figure 3.5: Latent Variable Distribution

The KDE is an appropriate solution for non-Gaussian distributions such as that of the latent variables illustrated in Figure 3.5.

3.10 Chapter Summmary

In this chapter, three dynamic approaches are extended to non-linear systems using the KDE resulting in the development of the novel DPCA with KDE, DPLS with KDE and CVA with KDE approaches. A detailed description of all three approaches as well as the existing approaches on which they are based is presented. To demonstrate the efficiency of the novel approaches developed in this chapter, these approaches are applied to the TEP plant. A description of the TEP plant is also presented in this chapter. The monitoring performance of the developed techniques are first compared with their non-KDE counterparts and then compared one with another.

Generally, the KDE approaches developed in this chapter are able to improve the monitoring performance over their non-KDE counterparts for most of the TEP faults employed. Nonetheless, the monitoring performance of the CVA with KDE is better than the monitoring performance of the DPCA with KDE and the DPLS with KDE as well as their non-KDE counterparts. Although the monitoring performance of the CVA with KDE is not significantly better than that of the CVA for some of the faults, the monitoring performance of the CVA with KDE is significantly better than that of the CVA without KDE for faults that are particularly difficult to detect by most monitoring algorithms.

The employment of the KDE with the existing DPCA, DPLS and CVA approaches, allows these approaches to address the non-linear issues associated with most industrial plants in addition to their ability to adddress dynamic issues.

Chapter 4

State Space Independent Component Analysis

The current chapter presents a detailed discussion of a novel state space independent component analysis (SSICA) technique developed in this work. The proposed SSICA approach is evaluated by applying it to the TEP plant. Furthermore, the monitoring performance of the proposed SSICA is compared with those of the existing dynamic ICA (DICA) and CVA approaches.

4.1 Introduction

The CVA technique discussed in the previous chapter is an efficient dynamic process monitoring tool as discussed already [4, 7, 101]. However, the CVA is generally a linear approach not appropriate for non-linear systems. For this reason in this work, the CVA is extended to non-linear systems using the KDE as explained in the previous chapter. Also, it was demonstrated in the previous chapter that the CVA with KDE approach developed in this work is able to improve the monitoring

performance over the existing CVA with control limits derived based on the Gaussian assumption as well as the DPCA and DPLS based approaches. However, the states obtained from the CVA are only de-correlated and not statistically independent.

A solution to the problem of monitoring non-linear systems is the Independent Component Analysis (ICA) [15, 16, 132]. The ICA technique recovers a few statistically independent source signals from collected process measurements by assuming that these independent signals are non-Gaussian. A set of variables are said to be statistically independent from each other when the value of one variable cannot be predicted by giving the value of another variable. The measured variables in process systems are in general not independent but are combinations of several independent sources that are not directly measured. Furthermore, the ICA can be employed for the identification of these independent sources. To identify the unmeasured ICs from measurements, the ICA algorithm involves a pre-processing stage known as the whitening stage to eliminate the cross correlation between the process variables before extracting the independent components [15, 16, 19, 121, 123, 132, 133, 134].

Conventionally, most published ICA studies have utilised the PCA for whitening and dimension reduction in the pre-processing stage [15, 16, 121]. This allows the ICs to be interpreted by the simple geometry of the PCA [19]. However, the connection of the ICA with PCA makes such ICA approaches not appropriate for dynamic process monitoring due to the static nature of the PCA.

To address this limitation of the ICA for dynamic processes, Lee et al. [123] extended the ICA methods and proposed the dynamic ICA (DICA) approach to improve the monitoring performance. In the so called DICA approach, a dynamic extension of the PCA (DPCA) is applied to an augmented data set to account for the auto-correlations in the pre-processing stage before extracting the ICs from the decomposed PCs. However, the DICA, like the DPCA, is not the best approach to capture the dynamic behaviour from process measurements [3]. As a result, the statistical

advantage of the ICA is not fully exploited by the DICA and the performance of the DICA in dynamic process monitoring is still not satisfactory as to be illustrated in this work.

Nevertheless, the state-space models like the Canonical Variate Analysis (CVA) on the other hand are reported to be efficient tools for dynamic process monitoring [3, 4, 6, 10, 12, 21, 99]. The CVA is a dimension reduction technique that is based on state variables and is well suited for auto-correlated and cross-correlated process measurements. This makes the CVA based approaches a better choice than the PCA based approaches for dynamic process monitoring.

To derive an efficient tool for non-linear dynamic process monitoring, in this study, the CVA, rather than the PCA, is proposed as the pre-processing tool to associate with the ICA resulting in a novel State Space Independent Component Analysis (SSICA) approach. In the proposed SSICA approach, the CVA is adopted as a dimension reduction tool to construct a state space and perform the dynamic whitening in the pre-processing stage before applying the ICA to the constructed state space in order to identify the statistically independent components. The SSICA approach is developed for non-linear dynamic process monitoring and applied to the Tennessee Eastman Process Plant as a case study.

Moreover, it is demonstrated that generally, the proposed SSICA is able to improve the process monitoring performance over the existing DICA technique, which is reported to be an improvement of the traditional ICA [15, 127]. Also, the overall performance of the proposed SSICA is better than the CVA, which was reported to be an efficient dynamic monitoring tool [3, 4, 6, 10, 12, 21, 99]. The performance improvements of the SSICA over the DICA and the CVA include an increase in detection reliability and a decrease in both detection delay and false alarms.

4.2 State Space Independent Component Analysis Algorithm

The development of process monitoring techniques is geared towards applying these techniques to industrial processes in order to improve process performance monitoring. It is well known that most industrial processes to which the monitoring techniques are applied are dynamic and non-linear. The SSICA approach is developed to deal with such processes. Firstly, from a general non-linear dynamic system, a linearized state space model can be constructed from normal operation data through the CVA. The non-Gaussian collective modelling errors are then extracted as statistically independent components through the SSICA. The obtained state space independent components together with the residuals are used for process monitoring by comparing the upper control limits estimated through the KDE approach. These algorithms are described in details as follows.

4.2.1 Canonical Variate Analysis

Consider a non-linear dynamic plant represented as:

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k) + \mathbf{w}_k$$

$$\mathbf{y}_k = \mathbf{g}(\mathbf{x}_k) + \mathbf{v}_k \tag{4.1}$$

where $\mathbf{x}_k \in \Re^n$ and $\mathbf{y}_k \in \Re^m$ are state and measurement vectors respectively, $\mathbf{f}(\cdot)$ and $\mathbf{g}(\cdot)$ are unknown non-linear functions, whereas $\mathbf{w}_k$ and $\mathbf{v}_k$ are plant disturbances and measurement noise vectors respectively. Clearly, it is difficult to monitor such unknown non-linear dynamic systems directly. Fortunately, under normal operation conditions, the plant in Equation (4.1) can be approximated by a linear stochastic

state space model as:

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \boldsymbol{\varepsilon}_k$$

$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \boldsymbol{\eta}_k \tag{4.2}$$

where $\mathbf{A}$ and $\mathbf{C}$ are unknown state and output matrices respectively, whereas $\boldsymbol{\varepsilon}_k$ and $\boldsymbol{\eta}_k$ are collective modelling errors partially due to the underlying nonlinearity of the plant, which has not been included in the linear model, and partially associated with process disturbance and measurement noise, $\mathbf{w}_k$ and $\mathbf{v}_k$ respectively. Note, as a result of the nonlinearity of the physical plant represented in Equation (4.1), the collective modelling errors, $\boldsymbol{\varepsilon}_k$ and $\boldsymbol{\eta}_k$ in Equation (4.2) generally will be non-Gaussian although $\mathbf{w}_k$ and $\mathbf{v}_k$ might be normally distributed.

To monitor the linear dynamic process represented in (4.2) without knowing matrices $\mathbf{A}$ and $\mathbf{C}$, the CVA is employed to extract the state variables $\mathbf{x}_k$ from process measurements, $\mathbf{y}_k$. The CVA is based on the so called *subspace identification*, where the process measurements are stacked to form the past and future spaces through the past, $\mathbf{y}_{p,k}$ and future, $\mathbf{y}_{f,k}$ observations defined as follows.

$$\mathbf{y}_{p,k} = \begin{bmatrix} \mathbf{y}_{k-1} \\ \mathbf{y}_{k-2} \\ \vdots \\ \mathbf{y}_{k-q} \end{bmatrix} \in \Re^{mq}, \quad \tilde{\mathbf{y}}_{p,k} = \mathbf{y}_{p,k} - \bar{\mathbf{y}}_{p,k} \tag{4.3}$$

$$\mathbf{y}_{f,k} = \begin{bmatrix} \mathbf{y}_{k} \\ \mathbf{y}_{k+1} \\ \vdots \\ \mathbf{y}_{k+q-1} \end{bmatrix} \in \Re^{mq}, \quad \tilde{\mathbf{y}}_{f,k} = \mathbf{y}_{f,k} - \bar{\mathbf{y}}_{f,k} \tag{4.4}$$

where the first subscripts of $\mathbf{y}_{p,k}$ and $\mathbf{y}_{f,k}$ indicate either the past (p) or the future (f) observations, respectively, whilst the second subscripts stand for the reference sampling point, where the past and future observations are defined. The sample means of the past and future observations are represented as $\bar{\mathbf{y}}_{p,k}$ and $\bar{\mathbf{y}}_{f,k}$ respectively, whilst $\tilde{\mathbf{y}}_{p,k}$ and $\tilde{\mathbf{y}}_{f,k}$ are pre-processed past and future observations with zero means.

The past and future truncated Hankel matrices $\mathbf{Y}_p$ and $\mathbf{Y}_f$ are then defined in Equation (4.5) and Equation (4.6) respectively.

$$\mathbf{Y}_p = \begin{bmatrix} \tilde{\mathbf{y}}_{p,(q+1)} & \tilde{\mathbf{y}}_{p,(q+2)} & \cdots & \tilde{\mathbf{y}}_{p,(q+M)} \end{bmatrix} \in \Re^{mq \times M} \tag{4.5}$$

$$\mathbf{Y}_f = \begin{bmatrix} \tilde{\mathbf{y}}_{f,(q+1)} & \tilde{\mathbf{y}}_{f,(q+2)} & \cdots & \tilde{\mathbf{y}}_{f,(q+M)} \end{bmatrix} \in \Re^{mq \times M} \tag{4.6}$$

From the Hankel matrices defined above, the covariance of the past, future and cross-covariance matrices are estimated as follows:

$$\mathbf{\Sigma}_{pp} = E(\tilde{\mathbf{y}}_{pk}\tilde{\mathbf{y}}_{pk}^T) = \mathbf{Y}_p\mathbf{Y}_p^T(M-1)^{-1} \tag{4.7}$$

$$\mathbf{\Sigma}_{ff} = E(\tilde{\mathbf{y}}_{fk}\tilde{\mathbf{y}}_{fk}^T) = \mathbf{Y}_f\mathbf{Y}_f^T(M-1)^{-1} \tag{4.8}$$

$$\mathbf{\Sigma}_{fp} = E(\tilde{\mathbf{y}}_{fk}\tilde{\mathbf{y}}_{pk}^T) = \mathbf{Y}_f\mathbf{Y}_p^T(M-1)^{-1} \tag{4.9}$$

The goal of the CVA is to find the best linear combinations, $\mathbf{a}^T\tilde{\mathbf{y}}_{fk}$ and $\mathbf{b}^T\tilde{\mathbf{y}}_{pk}$ of the future and past observations so that the correlation between these combinations is maximised. The correlation can be represented as:

$$\rho_{fp}(\mathbf{a}, \mathbf{b}) = \frac{\mathbf{a}^T\mathbf{\Sigma}_{fp}\mathbf{b}}{(\mathbf{a}^T\mathbf{\Sigma}_{ff}\mathbf{a})^{1/2}(\mathbf{b}^T\mathbf{\Sigma}_{pp}\mathbf{b})^{1/2}} \tag{4.10}$$

Let $\mathbf{u} = \boldsymbol{\Sigma}_{ff}^{-1/2}\mathbf{a}$ and $\mathbf{v} = \boldsymbol{\Sigma}_{pp}^{-1/2}\mathbf{b}$. The optimization problem can be casted as:

$$\max_{\mathbf{u},\mathbf{v}} \quad \mathbf{u}^T(\boldsymbol{\Sigma}_{ff}^{-1/2}\boldsymbol{\Sigma}_{fp}\boldsymbol{\Sigma}_{pp}^{-1/2})\mathbf{v}$$
$$s.t. \quad \mathbf{u}^T\mathbf{u} = 1 \tag{4.11}$$
$$\mathbf{v}^T\mathbf{v} = 1$$

The solution of this problem can be obtained through the SVD on the scaled Hankel matrix, $\mathbf{H}$ as indicated in Equation (4.12).

$$\mathbf{H} = \boldsymbol{\Sigma}_{ff}^{-1/2}\boldsymbol{\Sigma}_{fp}\boldsymbol{\Sigma}_{pp}^{-1/2} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T \tag{4.12}$$

where $\mathbf{U} = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_{mq} \end{bmatrix} \in \Re^{mq \times mq}$, $\mathbf{V} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_{mq} \end{bmatrix} \in \Re^{mq \times mq}$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \ddots & & \vdots \\ 0 & 0 & \cdots & \sigma_{mq} \end{bmatrix} \in \Re^{mq \times mq}$$

From Equation (4.12) above, the canonical variate, $\mathbf{z}_k \in \Re^{mq}$ based on the past measurements can be derived as in Equation (4.13).

$$\mathbf{z}_k = \begin{bmatrix} b_1^T \\ b_2^T \\ \vdots \\ b_{mq}^T \end{bmatrix} \tilde{\mathbf{y}}_{p,k} = \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_{mq}^T \end{bmatrix} \boldsymbol{\Sigma}_{pp}^{-1/2}\tilde{\mathbf{y}}_{p,k} = \mathbf{V}^T\boldsymbol{\Sigma}_{pp}^{-1/2}\tilde{\mathbf{y}}_{p,k} = \mathbf{J}\tilde{\mathbf{y}}_{p,k} \tag{4.13}$$

where $\mathbf{J} = \mathbf{V}^T\boldsymbol{\Sigma}_{pp}^{-1/2} \in \Re^{mq \times mq}$ is the transformation matrix, which transforms the mq-dimensional past measurements to the mq-dimensional canonical variate space. The canonical variate estimated in (4.13) can be separated into the state and

residual spaces based on the order of the system, a. According to the magnitude of the singular values, the first a dominant singular values are determined and the corresponding a elements of the canonical variate retained as the state variables where $a < mq$. In addition, the remaining $(mq - a)$ elements of the canonical variate are said to be in the residual space. Equation (4.14) below shows the entire canonical variate space $(\mathbf{z}_k \in \Re^{mq})$ is spanned by the state variables $(\mathbf{x}_k \in \Re^a)$ and the residuals $(\mathbf{d}_k \in \Re^{mq-a})$, both of which are subsets of the canonical variate, $\mathbf{z}_k$.

$$\mathbf{z}_k = \begin{bmatrix} \mathbf{x}_k^T & \mathbf{d}_k^T \end{bmatrix}^T \tag{4.14}$$

The previous chapter showed that the state variables, $\mathbf{x}_k$ obtained through CVA provides a tool better than directly using the past or future observations to monitor the dynamic systems in (4.1). However, as shown in (4.2), the states are combinations of statistically independent non-Gaussian sources. To make process monitoring more efficient, identifying these sources from the states is desired. The corresponding algorithm is to be developed in the next section.

4.2.2 State Space Independent Component Analysis

According to Equation (4.2), $\mathbf{x}_k$ can be expressed as a linear combination of the initial state, $\mathbf{x}_0$ and the collective modelling errors, $\boldsymbol{\varepsilon}_j$, for $j = 0, 1, \ldots, k - 1$.

$$\mathbf{x}_k = \mathbf{A}^k \mathbf{x}_0 + \sum_{j=0}^{k-1} \mathbf{A}^j \boldsymbol{\varepsilon}_{k-1-j} \tag{4.15}$$

Equation (4.15) indicates that if $\mathbf{x}_0$ and $\boldsymbol{\varepsilon}_j$, $j = 0, \ldots, k - 1$ are mixtures of $m(\leq n)$ unknown independent components, $\mathbf{s}_j \in \Re^m$, for $j = 0, \cdots, k - 1$, then the states, $\mathbf{x}_k$, for $k = 1, \ldots M$ are also linear combinations of these unknown independent

components. More specifically, the relationship can be expressed as follows.

$$\mathbf{X} = \mathbf{B}_x \mathbf{S}_x \qquad (4.16)$$

where $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_M \end{bmatrix} \in \Re^{a \times M}$ is the state matrix, $\mathbf{B}_x = \begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_m \end{bmatrix} \in \Re^{a \times m}$ is an unknown mixing matrix, and $\mathbf{S}_x = \begin{bmatrix} \mathbf{s}_{x,0} & \cdots & \mathbf{s}_{x,M-1} \end{bmatrix} \in \Re^{m \times M}$ is unknown independent component matrix. The SSICA aims to estimate both mixing matrix, $\mathbf{B}_x$ and independent component matrix, $\mathbf{S}_x$, from the state matrix, $\mathbf{X}$ obtained through the CVA described above.

The problem can be solved through an existing ICA algorithm, such as the FastICA [145] to find a de-mixing matrix, $\mathbf{W}$ such that the rows of the estimated independent component matrix,

$$\hat{\mathbf{S}}_x = \mathbf{W}_x \mathbf{X} \qquad (4.17)$$

are as independent of each other as possible. Based on the "non-Gaussian represents independence" principle [145], the de-mixing matrix as well as the independent component matrix are obtained through iterative optimizations to maximize certain non-Gaussian criteria.

The ICA can be applied to the residual space spanned by $\mathbf{d}_k$. The independent component matrix in the residual space is obtained by applying the ICA algorithm to the residual matrix, $\mathbf{D}$ as follows.

$$\hat{\mathbf{S}}_d = \mathbf{W}_d \mathbf{D} \qquad (4.18)$$

where $\mathbf{D} = \begin{bmatrix} \mathbf{d}_1 & \cdots & \mathbf{d}_M \end{bmatrix} \in \Re^{(mq-a) \times M}$.

The ICA based process monitoring is frequently associated with the Mahalanobis

distance I^2, also known as the D-statistic [15, 123, 134]. The I^2 metric is the sum of the squared independent components extracted from the ICA algorithm.

$$I^2_{x,k} = \hat{\mathbf{s}}^T_{x,k}\hat{\mathbf{s}}_{x,k} \tag{4.19}$$

$$I^2_{d,k} = \hat{\mathbf{s}}^T_{d,k}\hat{\mathbf{s}}_{d,k} \tag{4.20}$$

where $\hat{\mathbf{s}}_{x,k}$ and $\hat{\mathbf{s}}_{d,k}$ are the k-th columns of $\hat{\mathbf{S}}_x$ and $\hat{\mathbf{S}}_d$, respectively. The M $I^2_{x,k}$ and $I^2_{d,k}$ values for $k = 1, \ldots, M$ are then used to derived the upper control limits, $I^2_{x,\mathrm{UCL}}(\alpha)$ and $I^2_{d,\mathrm{UCL}}(\alpha)$ using the bounded KDE algorithm described in the next section. It is worthy to note that the KDE employed in the SSICA approach in this chapter is bounded and different from the KDE of the CVA with KDE approach in the previous chapter.

For online monitoring, the ICs of the state and residual spaces is calculated from the new measurements, $\tilde{y}^{\mathrm{new}}_{p,k}$ using the transformation matrix, $\mathbf{J} = \begin{bmatrix} \mathbf{J}^T_x & \mathbf{J}^T_d \end{bmatrix}^T$ and the de-mixing matrices, $\mathbf{W}_x$ and $\mathbf{W}_d$ respectively.

$$\hat{\mathbf{s}}^{\mathrm{new}}_{x,k} = \mathbf{W}_x\mathbf{J}_x\tilde{\mathbf{y}}^{\mathrm{new}}_{p,k} \tag{4.21}$$

$$\hat{\mathbf{s}}^{\mathrm{new}}_{d,k} = \mathbf{W}_d\mathbf{J}_d\tilde{\mathbf{y}}^{\mathrm{new}}_{p,k} \tag{4.22}$$

The corresponding I^2 metrics for the new measurements are then obtained as follows.

$$I^{2,\mathrm{new}}_{x,k} = (\hat{\mathbf{s}}^{\mathrm{new}}_{x,k})^T\hat{\mathbf{s}}^{\mathrm{new}}_{x,k} \tag{4.23}$$

$$I^{2,\mathrm{new}}_{d,k} = (\hat{\mathbf{s}}^{\mathrm{new}}_{d,k})^T\hat{\mathbf{s}}^{\mathrm{new}}_{d,k} \tag{4.24}$$

A fault condition is then detected by the proposed SSICA if either I^2 metric is larger than the corresponding UCL.

4.3 Control Limit through Bounded KDE

The ICs are not Gaussian. Therefore, the UCL for the I^2 metric cannot be derived analytically. The kernel density estimation (KDE) is a well established approach to estimate the PDF of random processes [13, 14, 16]. Hence, it is a natural selection using the KDE to determine the UCL [4]. Considering both I^2 metrics are positive, a KDE algorithm with lower bound support is adopted in this work to estimate the UCL.

Let $y > 0$ be the random variable under consideration. Firstly, the bounded y is converted into unbounded x by defining $x = \ln(y)$. Then, the density function $p(x)$ can be estimated by the normal KDE algorithm. Finally, the density function of y is $p(\ln(y))/y$ as derived in (4.25).

$$P(y < b) = P(x < \ln(b)) = \int_{-\infty}^{\ln(b)} p(x)dx = \int_{0}^{b} p(\ln(y))\frac{1}{y}dy \qquad (4.25)$$

Therefore, by knowing $p(x)$, an appropriate control limit can be determined for a specific confidence bound, α using Equation (4.25). The estimation of the probability density function $\hat{p}(x)$ at point x through the kernel function, $K(\cdot)$ is defined as follows

$$\hat{p}(x) = \frac{1}{Mh} \sum_{k=1}^{M} K\left(\frac{x - x_k}{h}\right). \qquad (4.26)$$

where $x_k, \; k = 1, 2, \cdots, M$ are samples of x and h is the bandwidth. To avoid selecting a band-width too small or too big, an optimal band-width h_{opt} described in the previous chapter is adopted in a similar manner as the CVA with KDE approach

To use both I_x^2 and I_d^2 metrics together, the joint distribution of these two metrics has to be considered. In general, the joint probability of two random variables, x

and y is defined as follows.

$$P(x < a, y < b) = \int_{-\infty}^{a} \int_{-\infty}^{b} p(x, y) dx dy \qquad (4.27)$$

However, for the SSICA and the DICA, I_x^2 and I_d^2 are independent. Hence,

$$P(x < a, y < b) = P(x < a)P(y < b) \qquad (4.28)$$

Equation (4.28) can also be approximately applied to T^2 and Q metrics for the CVA because $\mathbf{x}$ and $\mathbf{d}$ in (4.14) are uncorrelated. This means the joint PDF estimation can be simplified by two univariate PDF estimations.

By replacing x_k in Equation (4.26) with $I_{x,k}^2$ and $I_{d,k}^2$ obtained in (4.19) and (4.20) respectively, the above KDE approach is able to estimate the underlying PDFs of the I_x^2 and I_d^2 metrics. The corresponding control limits, $I_{x,\text{UCL}}^2(\alpha)$ and $I_{d,\text{UCL}}^2(\alpha)$ can then be obtained from the PDFs of the I_x^2 and I_d^2 metrics for a given confidence level, α by solving the following equations respectively.

$$\int_0^{I_{x,\text{UCL}}^2(\alpha)} \frac{p(\ln(I_x^2))}{I_x^2} dI_x^2 \int_0^{I_{d,\text{UCL}}^2(\alpha)} \frac{p(\ln(I_d^2))}{I_d^2} dI_d^2 = \alpha \qquad (4.29)$$

$$\int_0^{I_{x,\text{UCL}}^2(\alpha)} \frac{p(\ln(I_x^2))}{I_x^2} dI_x^2 = \sqrt{\alpha} \qquad (4.30)$$

$$\int_0^{I_{d,\text{UCL}}^2(\alpha)} \frac{p(\ln(I_d^2))}{I_d^2} dI_d^2 = \sqrt{\alpha} \qquad (4.31)$$

In this work, a fault is then identified ($F_k = 1$) if either $I_{x,k}^{2,\text{new}} > I_{x\text{UCL}}^2(\alpha)$ or $I_{x,d}^{2,\text{new}} > I_{d\text{UCL}}^2(\alpha)$ conditions are satisfied, i.e.

$$F_k = (I_{x,k}^{2,\text{new}} > I_{x,\text{UCL}}^2(\alpha)) \oplus (I_{d,k}^{2,\text{new}} > I_{d,\text{UCL}}^2(\alpha)) \qquad (4.32)$$

where $\oplus$ represents a logical "OR" operation.

A graphical summary of the SSICA algorithm is presented in Figure 4.1.

(a)   (b)

Historical Data: Collect normal operating process data.

Pre-processing stage: Use CVA for whitening in the pre-processing stage, estimating the transformation matrix (**J**) and then constructing the state space.

Determination of model space: Using the correlations as explained in chapter 4, determine the model and residual spaces.

FASTICA: Using FastICA, find a demixing matrix (**W**) such that $\hat{S}_X = W_X X$ is as independent as possible.

Estimation of Residual ICs: Estimate the ICs for the residual space as shown below.
$$\hat{S}_D = W_D D$$

Estimate the I^2 metric: This is estimated for the model and residual spaces
$$I^2_{x,k} = \hat{s}^T_{x,k}\hat{s}_{x,k} \; , \; I^2_{d,k} = \hat{s}^T_{d,k}\hat{s}_{d,k}$$

Control Limit Estimation: For given α , solve
$$\int_0^{I^2_{x,UCL}(\alpha)} \frac{p(\ln(I^2_x))}{I^2_x} dI^2_x = \sqrt{\alpha}$$
$$\int_0^{I^2_{d,UCL}(\alpha)} \frac{p(\ln(I^2_d))}{I^2_d} dI^2_d = \sqrt{\alpha}$$

Collect Real-time Monitoring Data: This is the data that should be monitored to determine the presence of a fault.

Estimate the ICs: For online monitoring, the ICs of the state and residual spaces is calculated using the new data ($\tilde{y}^{new}_{p,k}$), the transformation matrix , $J = \begin{bmatrix} J^T_x & J^T_d \end{bmatrix}$ and the demixing matrices , W_x and W_d.
$$\hat{s}^{new}_{x,k} = W_x J_x \tilde{y}^{new}_{p,k}$$
$$\hat{s}^{new}_{d,k} = W_d J_d \tilde{y}^{new}_{p,k}$$

Estimate the I^2 metric for new measurements for the model and residual spaces:
$$I^{2,new}_{x,k} = (\hat{s}^{new}_{x,k})^T \hat{s}^{new}_{x,k}$$
$$I^{2,new}_{d,k} = (\hat{s}^{new}_{d,k})^T \hat{s}^{new}_{d,k}$$

Monitoring: A fault is detected if

either $I^{2,new}_{x,k} > I^2_{x,UCL}(\alpha)$ or
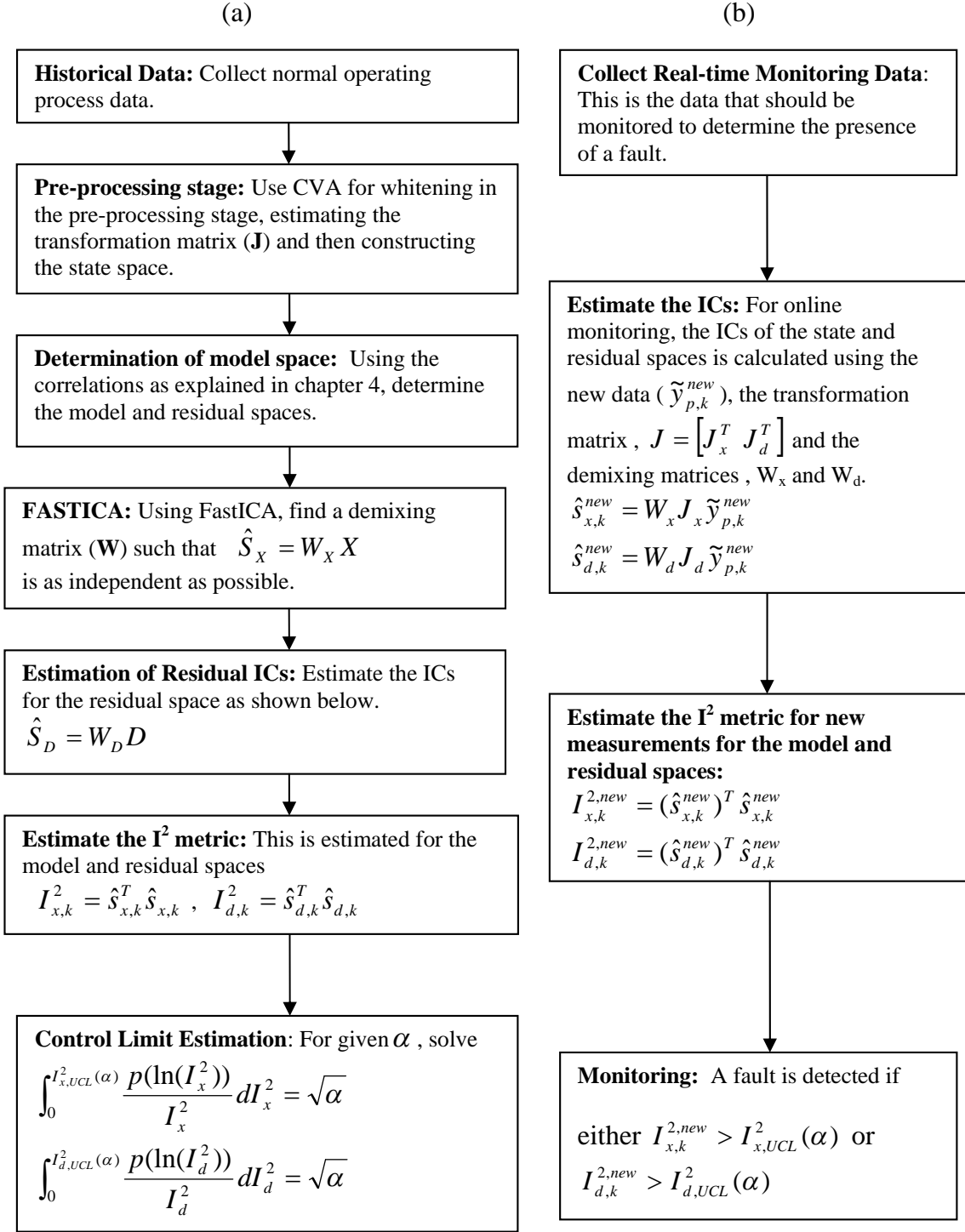$I^{2,new}_{d,k} > I^2_{d,UCL}(\alpha)$

Figure 4.1: Flow chart of SSICA algorithm

4.4   Monitoring Performance of SSICA

The proposed SSICA technique developed in this chapter is applied to data from the TEP plant to illustrate its efficiency. Details of the TEP plant are described in the previous chapter. The SSICA is assessed using the percentage reliabilities, detecion delays and false alarm rates. The monitoring performance of the proposed SSICA is compared with that of the DICA as well as the CVA to demonstrate the improvement by performing ICA on the state space obtained by the CVA.

For the pre-processing CVA described above, the number of state variables to retain in the dominant space is normally determined by the dominant singular values from the scaled Hankel matrix $\mathbf{H}$ in Equation (4.31). However, applying the CVA to the TEP case study showed that using the dominant singular values left an unrealistically large number of state variables in the dominant space. Hence a more realistic number of state variables 28 were retained in the dominant space so that 28 state variables were retained from the pre-processing stage of CVA. To make a fair comparison with the proposed SSICA, in the DICA, equal number of latent variables were retained in the dominant spaces while the rest of the latent variables spanned the excluded spaces. The percentage reliability and the detection delay of all twenty TEP faults for the proposed SSICA technique is compared with that of the CVA and DICA techniques and presented in Table 4.1. The corresponding false alarm rate of all faults is 0.6849% for the DICA and 0% for both the CVA and the SSICA techniques.

The superiority of the SSICA over the CVA and DICA techniques is demonstrated in Table 4.1. For 10 of the 20 faults (2, 3, 8, 9, 10, 13, 15, 16, 17 and 20), the SSICA is able to improve the monitoring performance over the existing DICA technique in both reliability and detection delay, while for the rest 10 faults (1, 4, 5, 6, 7, 11, 12, 14, 18 and 19), the SSICA maintains the same performance as the DICA

117

Table 4.1: Performance comparison

Fault	Reliability (%)			Detection Delay (minute)		
	SSICA	CVA	DICA	SSICA	CVA	DICA
1	99.75	99.75	99.75	9	9	9
2	99.63	99.63	99.50	12	12	15
3	73.03	70.04	19.48	15	21	21
4	99.88	99.88	99.88	6	6	6
5	99.88	99.88	99.88	6	6	6
6	99.88	99.88	99.88	6	6	6
7	99.88	99.88	99.88	6	6	6
8	99.00	98.88	98.75	18	30	33
9	91.64	90.01	46.82	18	39	48
10	96.75	96.38	96.13	18	90	96
11	99.38	99.38	99.38	18	18	18
12	99.50	99.50	99.50	15	15	15
13	96.25	96.13	96.13	18	96	96
14	99.88	99.88	99.88	6	6	6
15	99.63	99.63	99.50	12	12	15
16	99.38	99.38	99.25	18	18	21
17	98.38	98.25	98.13	18	45	48
18	99.25	99.25	99.25	21	21	21
19	99.88	99.88	99.88	6	6	6
20	97.63	97.50	97.13	18	63	72

approach. In addition, the SSICA is able to achieve reduced false alarm rates than the DICA for all the TEP faults. In terms of percentage reliability, the improvement of the SSICA over the DICA is significant ($> 0.5\%$) for 4 of the faults (3, 9, 10 and 20). Particularly, for faults 3 and 9, the improvement is extremely significant, over 40%. Meanwhile, the SSICA is able to reduce the detection delay significantly (> 10 minutes) for 6 faults (8, 9, 10, 13, 17 and 20). More specifically, applying the SSICA to faults 10 and 13 resulted in over one hour reduction in the detection delay.

In comparison between the SSICA and CVA, the performance of the SSICA is better than that of the CVA for 7 faults (3, 8, 9 10, 13, 17 and 20) in both the reliability and the detection delay, whereas the performance criteria of the remaining 13 faults (1, 2, 4, 5, 6, 7, 11, 12, 14, 15, 16, 18 and 19) are the same for both methods. In the reliability, the improvement on 2 faults (3 and 9) are significant (over 1%). Also,

significant improvements in the detection delay (> 10 minutes) are observed for 6 faults (8, 9, 10, 13, 17 and 20), for two of which (10 and 13), the improvements are over one hour.

To appreciate the capability of the SSICA, fault detection by these three methods along with fault propagation is further analysed for Faults 3 and 9. Both Faults 3 and 9 relate to the temperature of D feed (stream 2), one for step change (Fault 3) and another for random variations (Fault 9). These faults directly result in small deviations in the reactor cooling water outlet temperature, which can be easily corrected by the closed-loop control system by manipulating the cooling water flow. Therefore, both faults are generally difficult to be detected by most monitoring approaches [5].

Figure 4.2 shows a comparison of the fault detection along with the propagation of Fault 3 for the SSICA (a), CVA (b) and DICA (c) techniques, using F_k derived from Equation (4.32), whilst Figure 4.3 shows the fault detection along with the propagated Fault 9 process for these three techniques.

It is for such faults as Faults 3 and 9 that the superiority of the proposed SSICA technique over the CVA and particularly the DICA techniques is most outstanding as illustrated in Figures 4.2 and 4.3. The performance of the SSICA is better than that of the CVA and DICA techniques for both Faults 3 and 9. Particularly, it is clear that for both faults the SSICA is able to show a significant improvement of fault detection over the DICA technique within a few hours of the early stage of fault propagation. The improvement in the early fault propagation stage is important since it will give more time for operators to deal with the detected fault.

Although the DICA, also referred to as the ICA with delays is reported to be a more efficient dynamic monitoring tool than the traditional ICA [15], the proposed SSICA technique is able to significantly improve the monitoring performance over
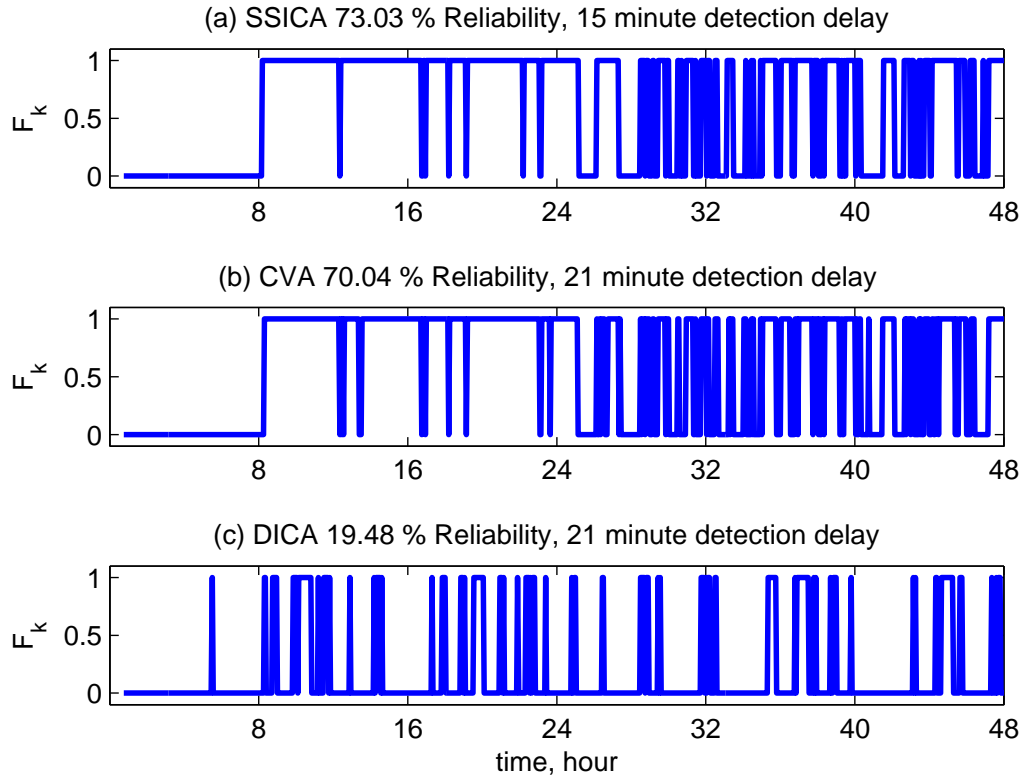
Figure 4.2: Comparison of fault detection along with the propagation of Fault 3

the DICA technique for most of the TEP faults considered in this work. This is because the pre-processing stage of the SSICA is based on the CVA, which is a more appropriate and more efficient dynamic monitoring tool than the DPCA on which the DICA technique is first based on. Furthermore, the efficiency of the SSICA over the CVA is owed to the fact that the SSICA is more suitable than the CVA to deal with non-Gaussian process measurement, separating the original sources to a greater degree than the CVA technique. Moreover, note that the results illustrated above demonstrate that there were no faults for which either the CVA or DICA techniques outperformed the proposed SSICA technique.

It is worth to note that the CVA approach adopted for comparison with the proposed SSICA is able to cope with certain level of non-linearities due to its association with the KDE for the determination of the UCL [4]. Moreover, the superiority of the
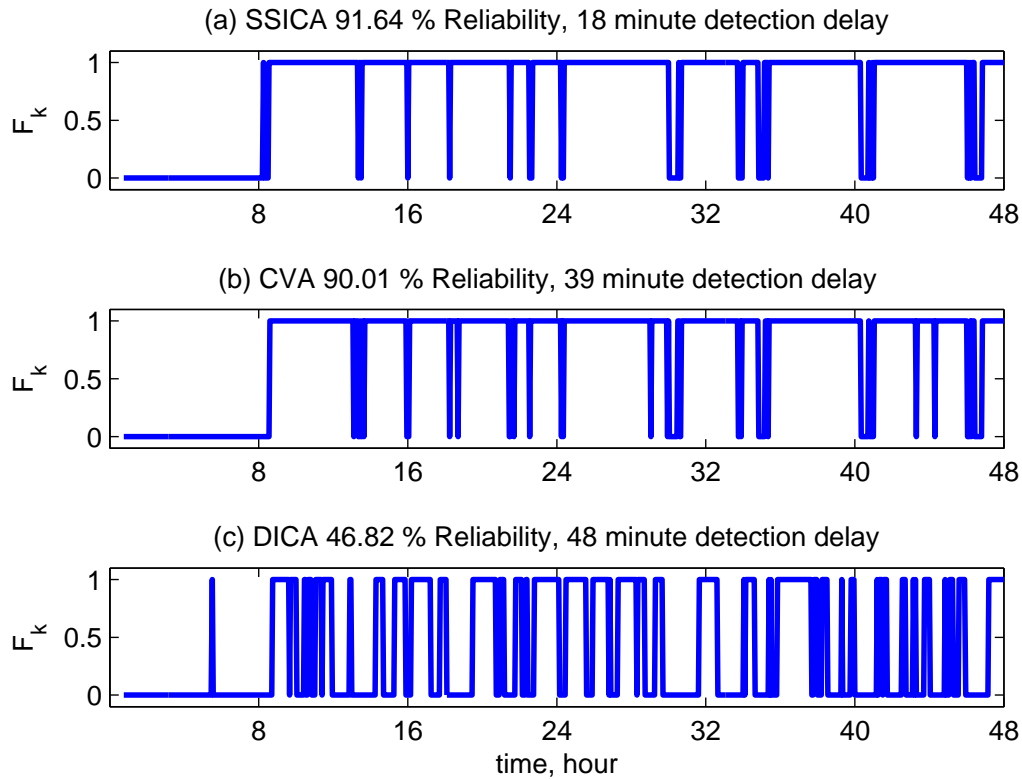
120

Figure 4.3: Comparison of fault detection along with the propagation of Fault 9

CVA over the DICA indicates that the dynamic issue has more impact on the fault detection performance than the non-linearity for the TEP process. This might be due to the feedback control, which widely propagates the transient response caused by a fault, as well as restricts the variations caused by a fault to relatively small level. This restriction on variation causes some faults to be difficult to detect without taking into account the correlations in time. Meanwhile, the effect of non-linearity on fault responses is also restricted so that the CVA with KDE approach is able to detect most faults adequately. This may also be the reason why the performance of the SSICA and the CVA is very close for most of the TEP faults.

4.5 Chapter Summmary

In this chapter, the existing ICA approach which is well suited to non-linear processes is extended to dynamic processes using the CVA resulting in the novel SSICA approach. The proposed SSICA consists of employing the CVA for the construction of the state space in the pre-processing stage before using the ICA for the extraction of the ICs from the constructed state space and then employing the KDE to derive appropriate control limits. To demonstrate the efficiency of the proposed SSICA, it is applied to the TEP plant. Moreover, a comparison of the monitoring performance of the SSICA with that of the CVA and DICA methods demonstrates the superiority of the proposed SSICA over both methods. This is because in the developement of the SSICA technique, the dynamic and non-linear properties commonly associated with most chemical processes are both taken into account. The DICA also takes the dynamic and non-linear properties into account, although unsatisfactorily. The CVA method on the other hand, only takes the dynamic properties into account.

Chapter 5

Case Study

This chapter describes a case study based on the ASM1 model of a Waste Water Treatment Plant (WWTP). Evaluation of all the monitoring algorithms developed in this work using the ASM1 model is illustrated. The KDE associated approaches are first evaluated and then compared with their non-KDE counterparts before a comparison of the KDE approaches is presented. Furthermore, the SSICA is evaluated using the ASM1 model and then a comparison of the results from the SSICA approach with those of the DPCA with KDE, DPLS with KDE, CVA with KDE and the DICA is presented.

5.1   Waste Water Treatment Plant

In the last few decades, waste water treatment has become an industry of high complexity due to the increasing requirements on the efficiency of the effluent water quality. Waste water treatment plants are large non-linear plants subject to dynamic changes. The major goal of waste water treatment plants is to remove pollutants from waste waters and recover the water quality. This recovery is based on physical, chemical and biological treatment processes. However, the biological treatment part

is considered for this case study.

The biological treatment consists of micro organisms like bacteria, unicellular and multi-cellular life forms that grow by feeding on the organic pollutants in the waste water. Also, the biological treatment aims to have only a certain amount of micro organisms in the process. This is achieved in an activated sludge (AS) reactor by separating the sludge from the water phase in a sedimentation unit and returning it to the biological reactor as recycled sludge. The excess sludge from the process is then removed and treated in a sludge treatment process which both stabilises and de-waters the sludge. This stabilised sludge can then be used as a fertiliser. However, sludge treatment is not discussed any further in this thesis because it is beyond the scope of this work.

The idea to develop a simulation benchmark for wastewater treatment plants was given birth to in the mid 1990s. The benchmark was developed in parallel by the European Cooperation in Science and Technology (COST) Actions 682/624 and the first IAWQ (International Association on Water Quality) Task Group on Respiratory-based Control of the Activated Sludge process [17]. Their goal was to develop a description of the organic carbon and nitrogen removal involving the nitrification and denitrification processes. In addition a mathematical representation simulating the behaviour of the bio-reactor was supposed to be created. The task group published their final result as the IAWQ Activated Sludge Model No. 1 (ASM1) in 1987 [17]. Moreover, this benchmark has been widely applied and is also employed in this work. The ASM1 model is described in the following section.

5.2 Activated Sludge Model No. 1

The ASM1 is a reference model for bioreactor in waste water treatment plants [17]. It is designed to treat an average flow of $20,000\text{m}^3$/day of waste water with an

average chemical oxygen demand (COD) concentration of 300mg/l. The COD is a measure of the amount of organic compounds in water.

5.2.1 Plant Configuration

The ASM1 plant consists of five reactors and a ten layer secondary settling tank which serves as a sedimentation unit. The first two reactors are anoxic while the last three are aerobic. Also, the reactors in the ASM1 plant are modelled as completely mixed models. A layout of the ASM1 plant is shown in Figure 5.1.
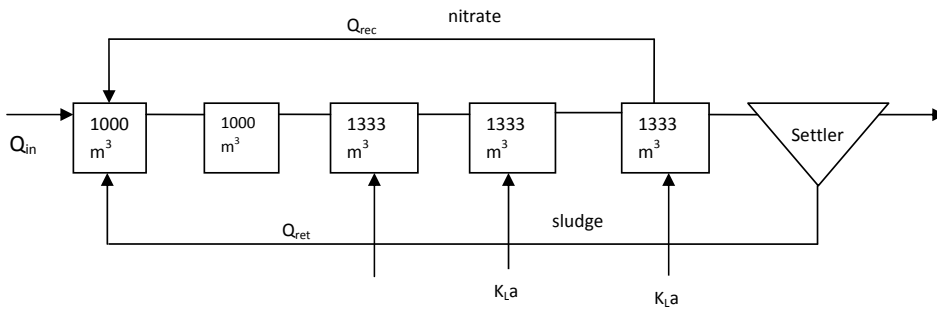

Figure 5.1: Layout of ASM1 plant

In addition, the physical dimensions of the plant are presented in Table 5.1.

Table 5.1: Physical dimensions of ASM1 plant

Description	Type
Anoxic Reactors	$1000m^3$
Aerobic Reactors	$1333m^3$
Sedimentation Unit	$6000m^3$
Flow rate of internal recycle stream from tank 5 to tank 1	$55338m^3/day$
Flow rate of sludge recycle stream from bottom of settler to tank 1	$18446m^3/day$
Average influent flow rate	$18446m^3/day$

The ASM1 model is based on the activated sludge process (ASP), in which biological reactions occur in the presence of oxygen while settled sludge is partly recycled into the reactor to enhance the ability of the biological reactor. This work is based on the modified ASM1 model with the feedforward-feedback control strategy developed in the work of Wang [17].

5.2.2 Influent Characteristics

Three influent scenarios represented by three data files were developed by a working group on the benchmarking of WWTP, COST 624, which are dry weather, storm weather and rain weather [15]. Each of the files contains influent data for 14 days in 15-minute intervals, with 96 samples taken in each day. Therefore, for the 14 days represented in the influent data, 1345 samples also known as observations were taken. The influent scenarios were developed to mimic real waste water characteristics typical for a plant of the chosen size. In each reactor, the variables included in the ASM1 are listed in Table 5.2.

Table 5.2: ASM1 variables

Symbol	Variable
S_I	Inert organic matter
S_S	Readily biodegradable substrate
X_I	Particulate inert organic matter
X_S	Slowly biodegradable substrate
$X_{B,H}$	Active heterotrophic biomass
$X_{B,A}$	Active autotrophic biomass
X_P	Particulate product from biomass decay
S_O	Dissolved oxygen
S_{NO}	Nitrate and nitrite nitrogen
S_{NH}	Ammonia nitrogen
S_{ND}	Biodegradable organic nitrogen
X_{ND}	Particulate biodegradable organic nitrogen
S_{ALK}	Alkalinity

5.2.2.1 Dry Weather

The dry weather data displays a daily pattern corresponding to that of a dry weather for 14 days. Hence, there is no major upset or disturbance present in this data set. The influent flow, soluble and particulate loadings from the dry weather file is depicted in Figure 5.2.
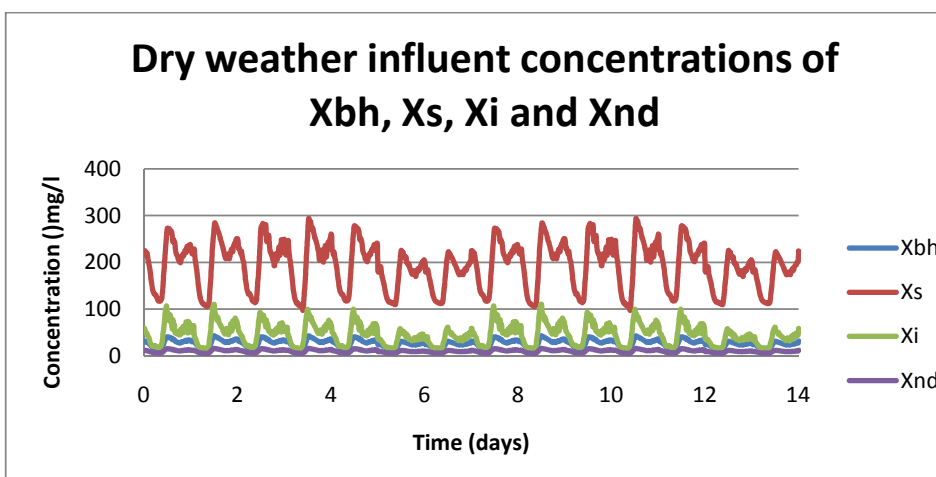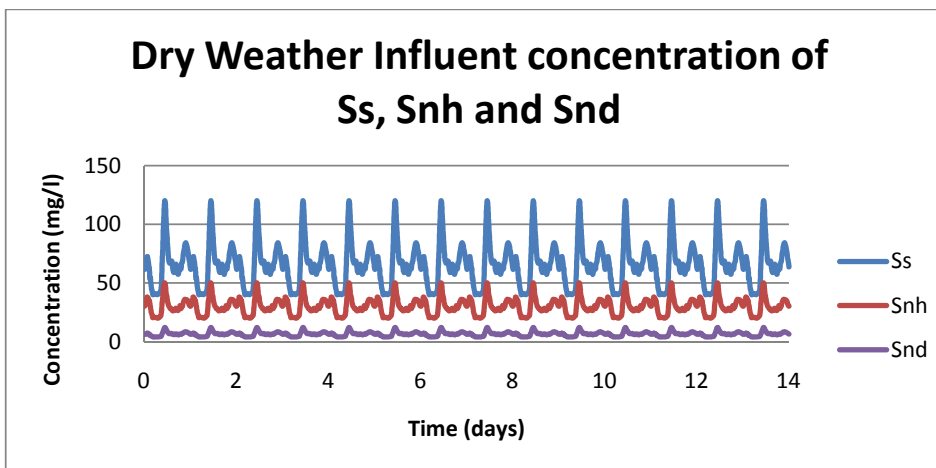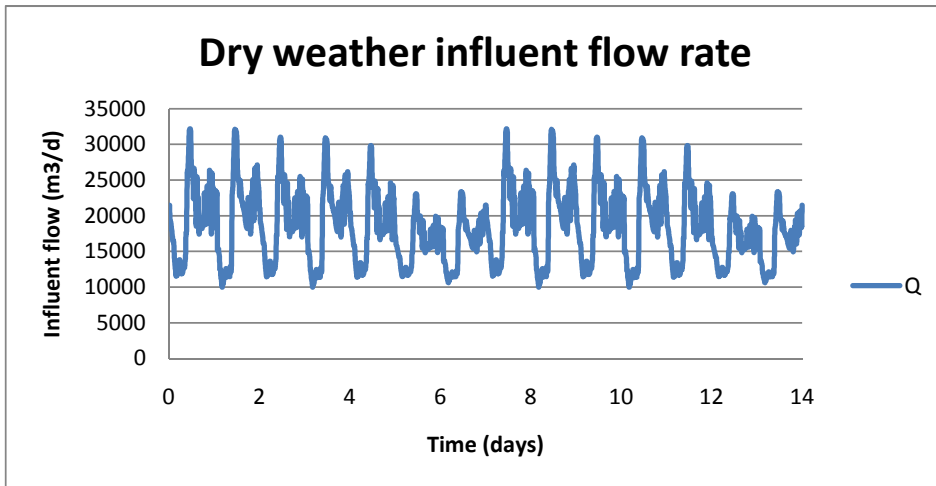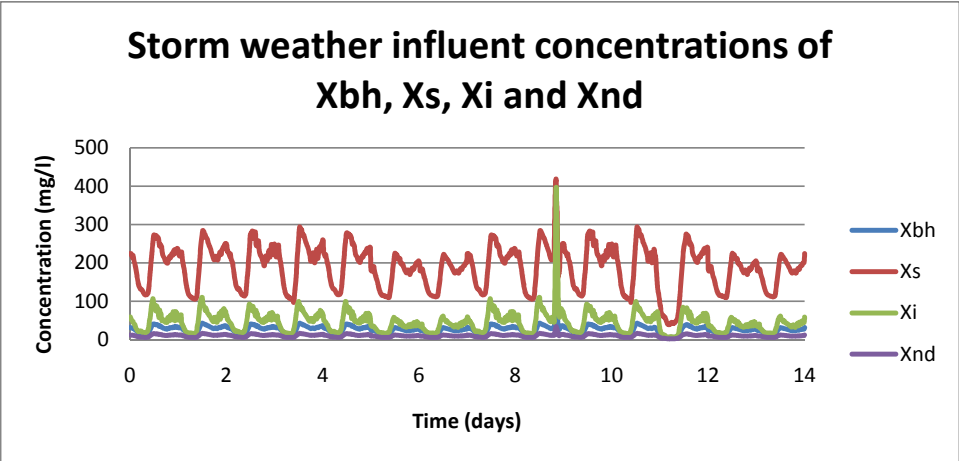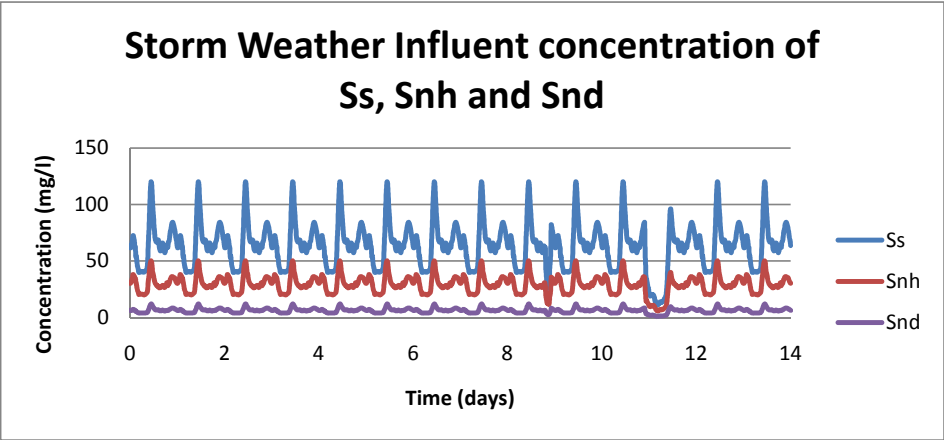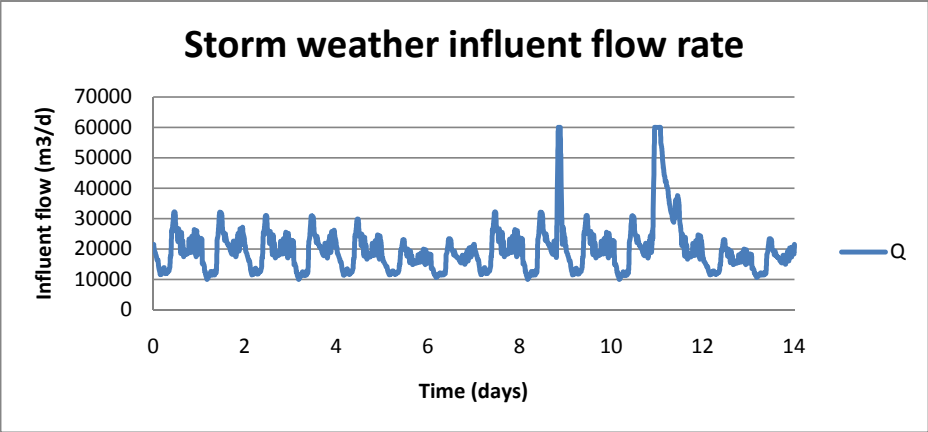
Figure 5.2: Plot of dry weather variables

Figure 5.3: Plot of storm weather variables

5.2.2.2 Storm Weather

The storm weather on the other hand has two major disturbances. The influent flow as well as the soluble and particulate loadings from the storm weather file are all shown in Figure 5.3. The first disturbance which occurs at the end of the 8th day is a short storm involving a rain event lasting for 3 hours. The influent flow reaches up to 3 times the average dry weather and the particulate pollutant loading increases up to 4 times the average dry weather loading as illustrated in Figure 5.3. As a result of the increased flow in the influent, there is a flush-out. This means that particulate matter present in the sewer system is flushed out due to the increase in the flow rate. The second disturbance in the storm data occurs at the start of the 11th day and involves a rain event lasting for 15 hours where the influent flow and the soluble pollutant loading is similar to that of the first disturbance. However, the particulate loading in the second disturbance is reduced to half the average dry weather loading because it is assumed that the sewers have been washed from the rain event in the first disturbance. As a result, the particulate matter is diluted resulting in low influent concentrations. During both disturbances in the storm data, the soluble compounds are diluted as illustrated towards the end of the 8th day and the 11th day as in the plot of the concentrations of Ss, Snh and Snd. In addition, the average influent flow rate during both disturbances increases from 20,000m3/day to 60,000m3/day3 as in the plot of the storm weather inluent flow rate in Figure 5.3.

5.2.2.3 Rain Weather

The rain weather data has a prolonged period of rain, where the rain starts on the 8th day and diminishes on the 10th day. The influent flow, soluble and particulate loadings from the rain weather file is depicted in Figure 5.4. The effect of the prolonged rain between the 8th and the 11th days is clearly shown in Figure 5.4. As a result of the prolonged rain, there is an increase in the influent flow rate while both

soluble and particulate matters are diluted as depicted in the plots of the soluble and particulate concentrations in Figure 5.4.
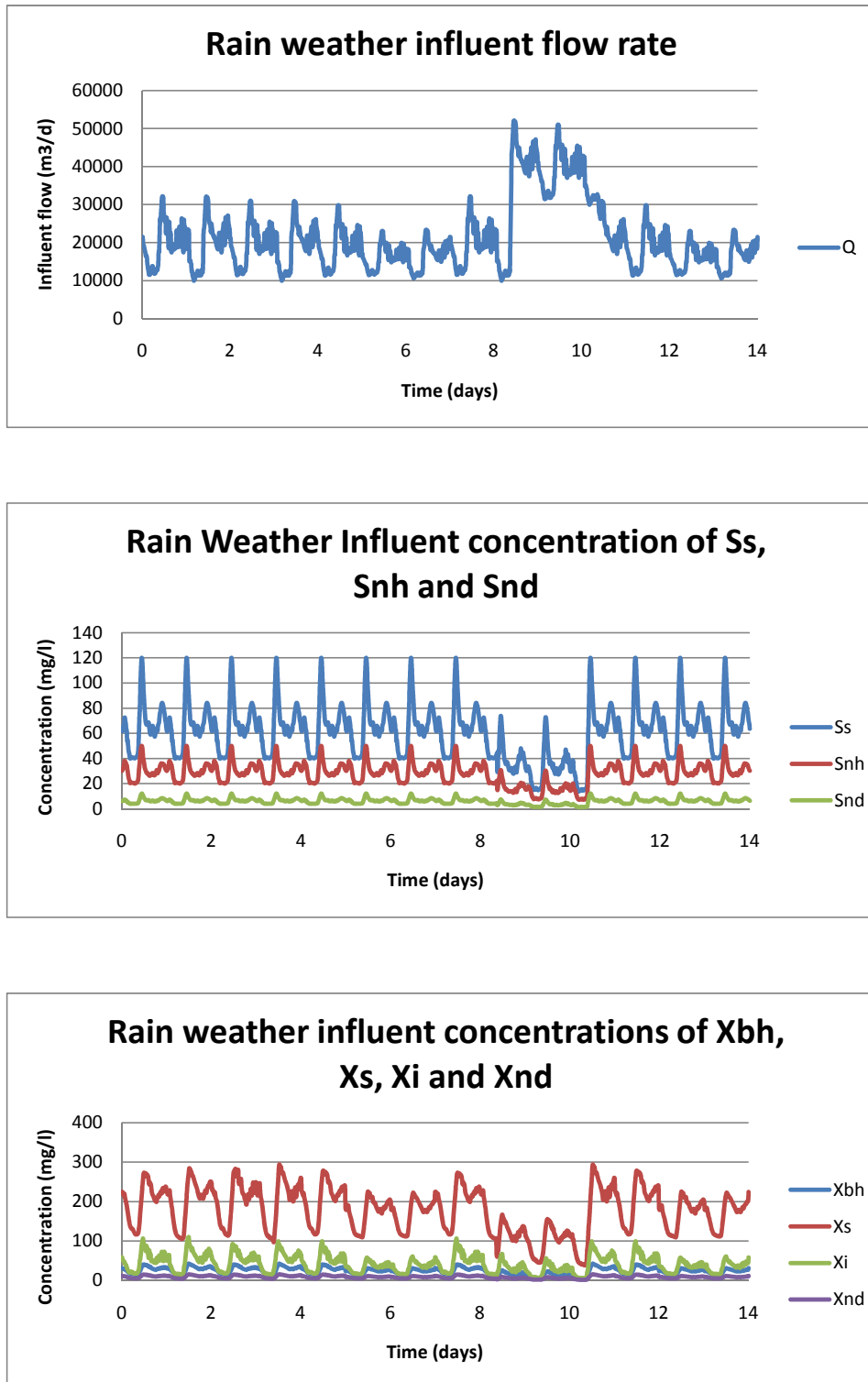

Figure 5.4: Plot of rain weather variables

5.2.3 Fault Scenarios

The training model in this work is based on a 14-day normal operating period of the dry weather. Generally, process monitoring involves the development of a model from the normal operating process and then employing the developed model to monitor on-line processes to determine the presence of a fault. For the purpose of this work, three fault scenarios (Faults A_1, A_2 and A_3) are simulated and evaluated. These fault scenarios involve different types of disturbances that could be encountered in a real waste water treatment plant.

The fault scenarios in this work were created by varying the values of some of the variables of the ASM1 model in its normal operating mode. Also, the created faults involve internal disturbances [15], which are variations caused by changes within the process that affect the process behaviour. For each of the fault scenarios created, the plant is operated in its normal mode for the first 7 days, before the various faults are introduced on the 8th day of the operation. These introduced faults are then allowed to remain in the process till the end of the process.

Among the many variables used in the benchmark, 6 were selected to build the monitoring system because they are important variables in the real waste water treatment plant systems [15]. The 6 measurements employed to build the monitoring system in this work are listed in Table 5.3;

Table 5.3: Monitoring variables for benchmark model

Symbol	Variable
Q_{in}	Influent flow rate
DO	Dissolved Oxygen in the fifth reactor
COD_e	Chemical oxygen demand
SNH_e	Influent ammonium concentration
BOD_e	Biological oxygen demand
K_La_5	oxygen transfer coefficient (reactor 5)

The nominal and fault data sets have 1345 observations and the 6 variables listed in Table 5.3 above. Generally, most changes in the biological process of the WWTP are

slow. Therefore, in this work, when faults are introduced, these faults are allowed to carry on till the end of the simulation process to allow the full impact of the fault to be felt in the process. The three fault scenarios created in this work are briefly described in the following sections.

5.2.3.1 Fault A_1

Fault A_1 involves a step decrease in the autotrophic growth rate. The autotrophic bacteria use inorganic carbon as their carbon source to bring about nitrification, which is the conversion of ammonia to nitrite and nitrate in the presence of sufficient air. The nitrate is then reduced to nitrogen gas in the absence of oxygen in the denitrification process and then released to the atmosphere. Generally, waste water treatment plants aim to remove nitrogen in order to improve the water quality. Hence, the nitrification rate is an important parameter in waste water treatment plants. As a result of the introduction of this fault, nitrification is inhibited and consequently, the nitrification rate which quantifies how quickly nitrification occurs is also affected. In the ASM1 model, the nominal value of the autotrophic growth rate is 0.5. In the Fault A_1 scenario, the autotrophic growth rate is changed on the 8th day to 0.2 and allowed to remain at that value for the rest of the simulation process.

5.2.3.2 Fault A_2

This fault involves a step decrease in the ammonification rate on the 8th day from $0.05\mathrm{m}^3/(\mathrm{gCODday})$ to $0.02\mathrm{m}^3/(\mathrm{gCODday})$ till the end of the process simulation. Ammonification is the conversion of nitrogen to ammonia while how quickly ammonification occurs is the ammonification rate. As a result of this fault, there is a decrease in the conversion of the nitrogen to ammonia. This results in more nitro-

gen in the water than required, which adversely affects the water quality. Faults that involve a change in the ammonification rate are generally well detected because the rate of conversion of nitrogen in waste water treatment plants has a significant effect on the growth of the micro-organisms and other aspects of the waste water treatment process.

5.2.3.3   Fault A_3

Fault A_3 involves a step change in the heterotrophic growth rate. The heterotrophic bacteria in the ASM1 use organic carbon as their carbon source to bring about denitrification, which is the reduction of the nitrate from the nitrification process to nitrogen gas. In the ASM1 model, the nominal value of the heterotrophic growth rate is 4. In the Fault A_3 scenario, the heterotrophic growth rate is doubled on the 8th day and allowed to remain at that value for the remaining duration of the simulation process. This fault is a change in the biological kinetic parameters which generally results in a small change in the output variables. Faults of this kind are generally more difficult to detect by majority of monitoring algorithms.

5.3   Monitoring Performance

The ASM1 model described above is further employed for the evaluation of all the monitoring methods developed in this work. The monitoring performance in this work is assessed based on the percentage reliability which is defined as the percentage of the samples outside the control limits [144] within the last 8 hour faulty operation. Hence a monitoring technique is said to be better than another technique if the percentage reliability of this technique is numerically higher than the percentage reliability of another. Two other parameters employed to assess the monitoring methods in this chapter are the detection delays and the false alarm

rates. The detection delay is the time period it takes to detect a fault after the introduction of the fault while the false alarm rates is the amount of non-fault operations that are mis-detected as fault operations by the monitoring algorithms.

The percentage reliabilities, detection delays and false alarm rates are the criteria for the assessment of the monitoring methods in this chapter. The 99% confidence interval is also adopted in this work. The evaluation of the developed monitoring algorithms in this chapter is similar to those in the previous chapters. To evaluate the monitoring methods in this chapter, the number of time lags for past and future observations is first determined from the autocorrelation function of the summed squares of all measurements as shown in Figure 5.5 against $\pm 5\%$ confidence bounds.
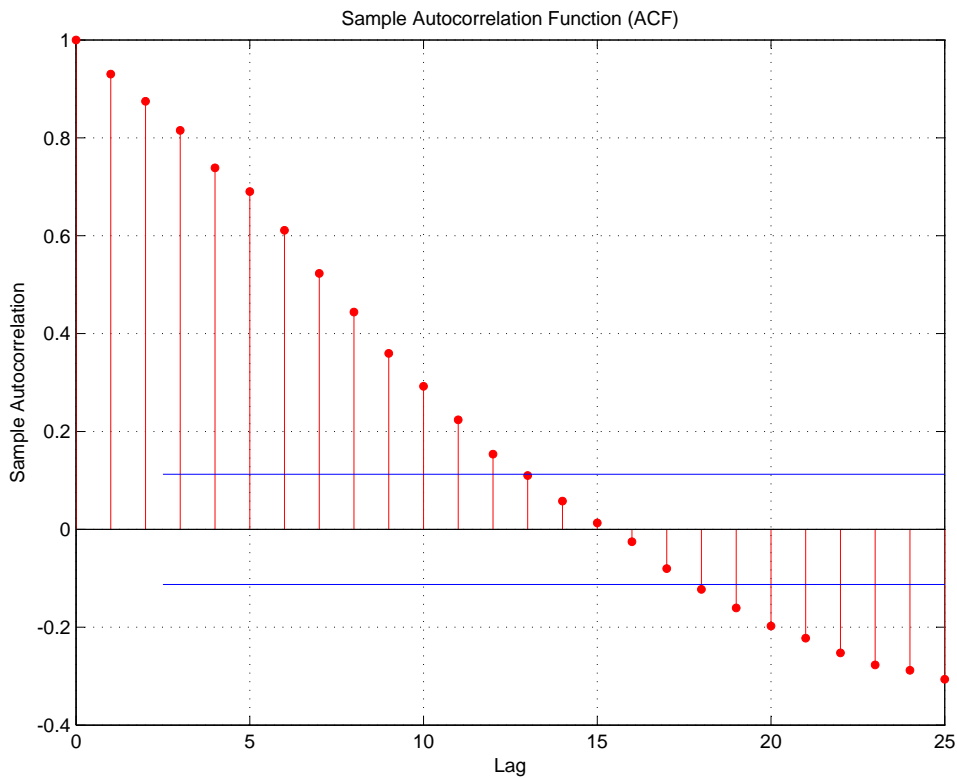

Figure 5.5: Autocorrelation function of the summed squares of all measurements

The autocorrelation function in Figure 5.5 above indicates that the maximum number of significant lags is 13. Hence the past vector (p) and the future vector (f)

are both set to 13. Also, the length of the past and future observations (mq) is 78 according to Equations (3.48) and (3.49). The number of columns of the truncated Hankel matrices according to (3.55) is $M = 1320$. To determine the order (a) of the system, the singular value decomposition is performed on the scaled Hankel matrix as in (3.58). The singular values obtained from the singular value decomposition of the scaled Hankel are normalised to have their values ranging between 0 and 1. Thereafter, the order of the system is determined based on the dominant normalised singular values. Normalised singular values ranging between the values of $0.1 - 1$ are considered dominant while normalised singular values with values less than 0.1 are considered not to be dominant. In the work reported in this chapter, there are 26 dominant singular values represented by the circles as illustrated in Figure 5.6.
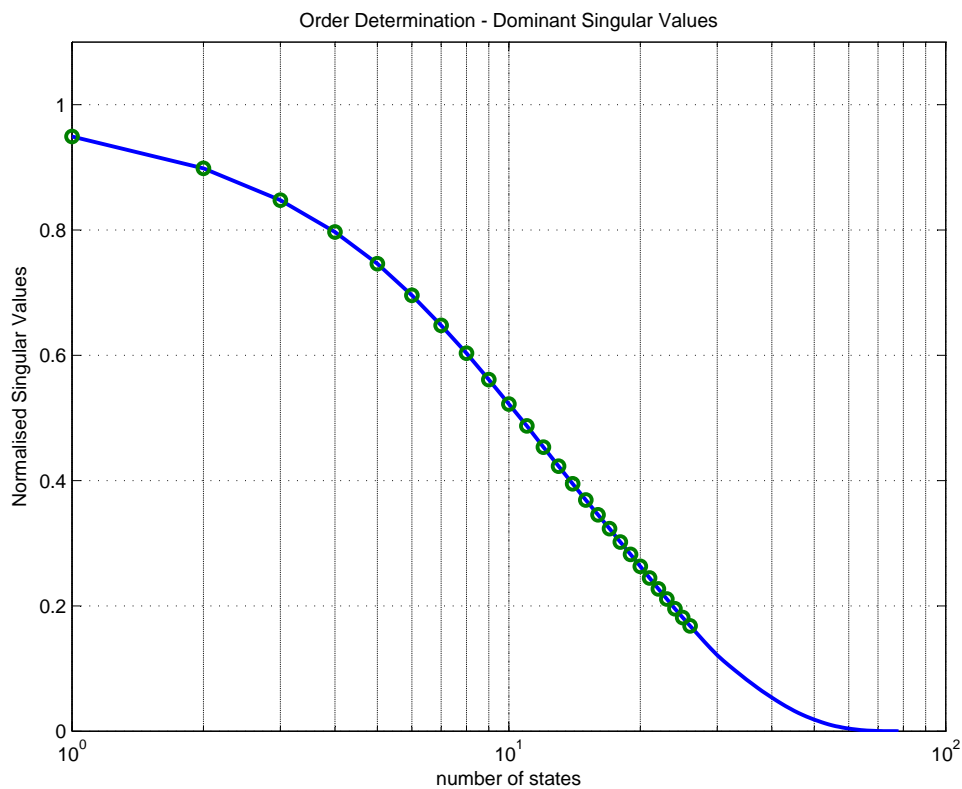

Figure 5.6: Normalised singular values from the scaled hankel matrix

Hence, in this work, $a = 26$. Sometimes, a fault may cause a significant deviation in the model space but not necessarily result in a similar level of significance in the

residual space, vice versa. Therefore, in the work in this chapter, a fault is identified ($F_k = 1$) if either $T^2{}_k > T^2{}_{\text{UCL}}(\alpha)$ or $Q_k > Q_{\text{UCL}}(\alpha)$ conditions are satisfied. This fault detection condition has the advantage of being more sensitive than employing either the T^2 or the Q metric alone. Also, by using this fault detection condition, the monitoring performance becomes insensitive to the number of states (a), since any ignored variances in the T^2 metric by reducing a will be recovered by Q metric.

In the following sections the monitoring performance of all the methods developed in this work are presented. Firstly, the KDE approaches are evaluated using the data produced by the ASM1 model and then compared with their non-KDE counterparts. The monitoring performance of the KDE approaches are also compared one with another. Furthermore, the SSICA is evaluated using the same data and its monitoring performance compared with those of the KDE approaches as well as the DICA approach. To make a fair comparison of the approaches considered in this chapter, equal number of latent variables are retained in the model spaces while the rest of the latent variables span the excluded spaces. In addition, equal number of lags for p and f are employed for all the approaches in this chapter.

5.4   Comparison of Monitoring Approaches

In this section, the reliabilities, detection delays and the false alarm rates of the monitoring approaches are presented in Table 5.4, Table 5.5 and Table 5.6 respectively.

5.4.1   Monitoring Performance of DPCA with KDE

In this section, the proposed DPCA with KDE is applied to the data from a waste water treatment plant. To illustrate the efficiency of the DPCA with KDE over

Table 5.4: Reliability comparison

Fault	Reliability (%)							
	CVA with KDE	CVA	DPCA with KDE	DPCA	DPLS with KDE	DPLS	SSICA	DICA
A_1	88.34	76.28	88.04	73.16	88.19	75.28	91.10	87.73
A_2	88.65	76.53	88.34	73.28	87.88	75.28	91.41	88.04
A_3	75.46	74.66	61.2	48.81	72.55	71.04	85.89	34.51

Table 5.5: Detection delay comparison

Fault	Detection Delay (min)							
	CVA with KDE	CVA	DPCA with KDE	DPCA	DPLS with KDE	DPLS	SSICA	DICA
A_1	285	2505	360	2601	465	1140	270	1215
A_2	285	2505	360	2610	465	1140	270	1235
A_3	285	2505	360	2601	465	1140	270	1230

the traditional DPCA, their monitoring performances are compared. The results in Table 5.4 show that, in terms of percentage reliability, the improvement of the DPCA with KDE over the traditional DPCA is significant ($> 14\%$) for the 3 faults employed. Meanwhile, the DPCA with KDE is also able to reduce the detection delay significantly (> 37 hours) for all 3 faults. This is very important in process monitoring because detecting the faults early is a major goal of process monitoring as this advantage gives operators more time to deal with the situation early in order to ensure safe operating processes. From the results illustrated in Table 5.6, the DPCA with KDE has a higher false alarm rates (0.06%)than the traditional DPCA for the 3 faults employed. This is because the DPCA with KDE has a tighter control limit with a greater chance of false alarm rates.

Table 5.6: False alarm rates comparison

Fault	False Alarm Rate (%)							
	CVA with KDE	CVA	DPCA with KDE	DPCA	DPLS with KDE	DPLS	SSICA	DICA
A_1	2.199	1.268	2.786	2.724	2.199	2.002	2.199	0
A_2	2.199	1.268	2.786	2.724	2.199	2.002	2.199	0
A_3	2.199	1.268	2.786	2.724	2.199	2.002	2.199	0

5.4.2 Monitoring Performance of DPLS with KDE

In this section, the monitoring performance of the DPLS with KDE is compared with the traditional DPLS. For the DPLS analysis in this section, the Chemical oxygen demand (COD_e) and the Biological oxygen demand (BOD_e) were selected as the response variables while the predictor variables consisted of the Influent flow rate (Q_{in}), Dissolved oxygen in the fifth tank (DO), Influent ammonium concentration (SNH_e) and the oxgen transfer coefficient in the fifth reactor (K_{La5}). As illustrated in Table 5.4, the improvement of the DPLS with KDE over the traditional DPLS is significant ($> 1.5\%$) for all the faults employed. In addition, the DPLS with KDE is able to detect the faults significantly earlier (> 11.25 hours) than the traditional DPLS as shown in Table 5.5. Nevertheless, Table 5.6 shows that the DPLS with KDE has a higher false alarm rates (0.2%)than the traditional DPLS for the 3 faults employed. This is because the DPLS with KDE has a tighter control limit with a greater chance of false alarm rates.

5.4.3 Monitoring Performance of CVA with KDE

In this section, the monitoring performance of the CVA with KDE is compared with that of the traditional CVA. The results illustrated in Table 5.4 show that for all the faults employed, the reliabilities of the CVA with KDE is higher ($> 0.8\%$) than the reliabilities of the CVA without KDE. Also, it is shown in Table 5.5 that for

all the faults, the proposed CVA with KDE is able to detect the faults earlier (37 hours) than the CVA without KDE. The superiority of the CVA with KDE over the CVA without can be attributed to the suitability and appropriateness of the control limit derived using the KDE for plants exhibiting nonlinearity. Nonetheless, the CVA with KDE is unable to improve the false alarm rates over the traditional CVA. This is because the control limit of the CVA with KDE is tighter than that of the traditional CVA, causing the CVA with KDE to have a higher false alarm rate than the traditional CVA.

Clearly, the KDE approaches are able to significantly improve the monitoring performance over the their non-KDE counterparts in terms of reliability and detection delays. This improvement is due to the ability of the KDE associated approaches to account for the non-Gaussianity of the plants that may be due to non-linearity, which the non-KDE approaches do not account for. However, the KDE approaches developed in this work are unable to improve the false alarm rates. This is because using the KDE, the UCL is reduced which results in a higher chance of false alarm rates.

It is illustrated in Table 5.4, Table 5.5 and Table 5.6 that generally, the CVA with KDE is superior to the DPCA with KDE and the DPLS with KDE approaches also developed in this work. From Table 5.4, the reliabilities of the CVA with KDE are higher than those of the DPCA with KDE for all faults. Also, the CVA with KDE has higher reliabilities than the DPLS with KDE for all faults. This is because the CVA is a better dynamic monitoring tool than the DPCA and DPLS approaches. The CVA is able to capture those dynamic behaviours that may not be captured by the DPCA and DPLS approaches. This way, the CVA with KDE does not only address the issues of non-Gaussianity which could be due to the non-linearity of the plant but also deals efficiently with the dynamic issues associated with industrial plants. For Faults A_1 and A_3, the DPLS with KDE has higher reliabilities than

the DPCA with KDE while for the remaining fault (A_2), the DPCA with KDE performed better than the DPLS with KDE in terms of reliabilities.

The illustrations in Table 5.5 show that the CVA with KDE is able to detect all 3 faults earlier (> 1 hours) than the DPCA with KDE. In addition, the CVA with KDE is also able to improve significantly (3 hours) the monitoring performance in terms of detection delay over the DPLS with KDE for all the faults. Although all the KDE approaches address the issue of nonlinearity of the plant in a similar way, the CVA based approach deals with dynamic issues better than the DPCA and DPLS based approaches. Hence, the ability of the CVA with KDE to detect the faults earlier than the DPCA with KDE and the DPLS with KDE is due to the fact that the CVA is a better dynamic monitoring tool than the DPCA and DPLS approaches. Meanwhile, the DPCA with KDE is able to detect the faults earlier (1.75 hours) than the DPLS with KDE for all 3 faults.

Table 5.6 shows that the false alarm rates of the CVA with KDE are lower than those of the DPCA with KDE for all 3 faults. However, the false alarm rates of the CVA with KDE and those of the DPLS with KDE are exactly the same for all the faults. Also, the DPLS with KDE is able to improve ($> 0.6\%$) the monitoring performance over the DPCA with KDE in terms of the false alarm rates for all of the faults.

In summary, the evaluation of the KDE associated approaches using the data simulated from the ASM1 model shows that the KDE approaches are able to improve the monitoring peformance over their non-KDE counterparts, based on the parameters employed in this work to assess the performance of the monitoring techniques.

5.4.4 Monitoring Performance of SSICA

The SSICA approach is evaluated using simulated data from the ASM1 model. The proposed SSICA is first based on the CVA to construct the state space before applying the ICA to extract the state space independent components. To demonstrate the improvement by performing ICA on the state space obtained by the CVA, the SSICA is compared with the CVA. The DICA [123, 127] is a dynamic extension of the ICA with similar objectives as the proposed SSICA approach. Therefore, in this work, the SSICA is also compared with the DICA to illustrate its efficiency. In addition, the SSICA is compared with the KDE approaches.

The reliabilities illustrated in Table (5.4) show that the SSICA is able to improve the monitoring performance in terms of the reliability over all the other methods for all the faults employed in this work. The improvement of the SSICA over the other methods is significant ($> 2.76\%$). The SSICA is able to significantly improve the reliabilities over the DICA approach. The efficiency of the SSICA over the DICA is particularly emphasised for the Fault A_3 process as illustrated. The superiority of the SSICA over the DICA is due to the fact that the SSICA is based first on the CVA which is a more appropriate dynamic monitoring tool than the DPCA [3, 4] on which the DICA is first based in the pre-processing stage. The SSICA also has higher reliabilities than the CVA with KDE, DPCA with KDE and the DPLS with KDE methods for all 3 faults. Particularly, the improvement of the SSICA over the KDE associated approaches is significant for Fault A_3. To illustrate this significant improvement of the SSICA over the other methods for Fault A_3, a comparison of the fault detection along with the propagation of Fault A_3 for all the techniques is presented in Figure 5.7.
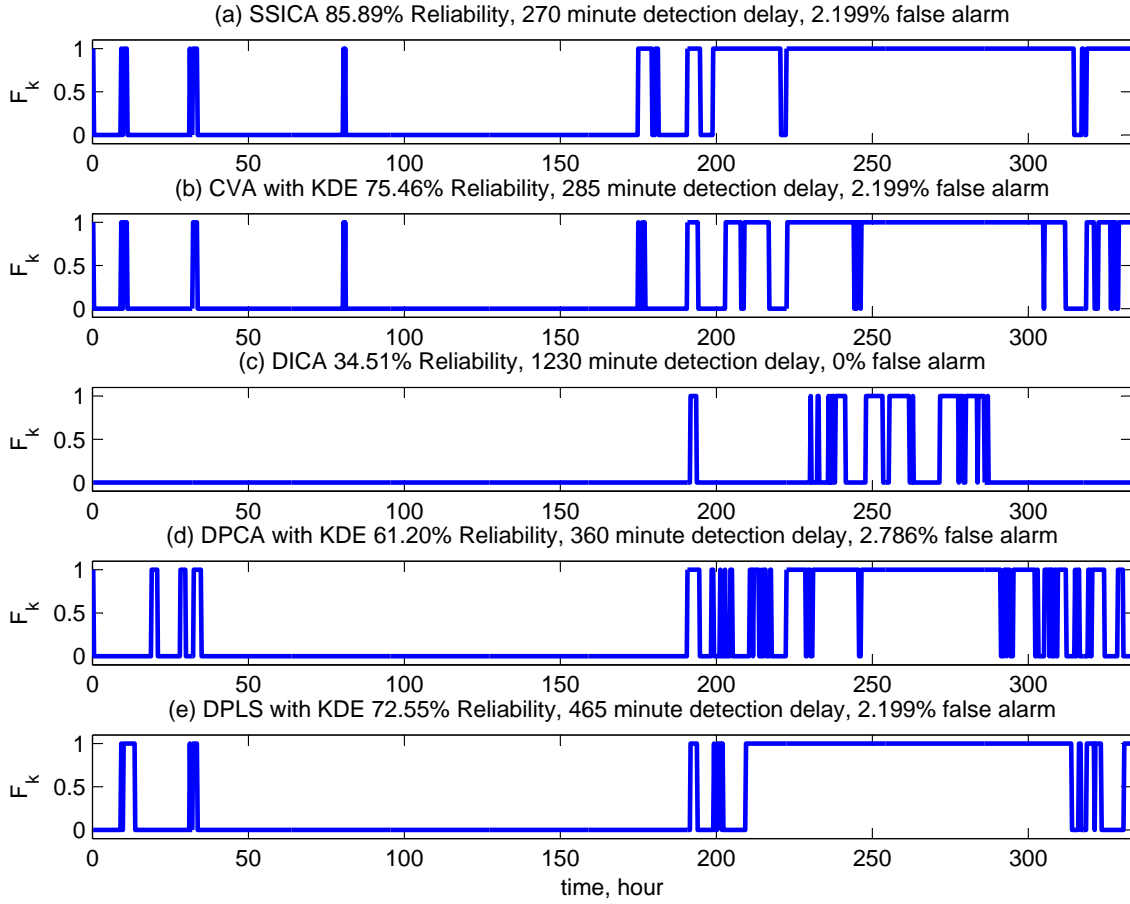
Figure 5.7: Comparison of fault detection along with the propagation of Fault A_3

Furthermore, the results illustrated in Table 5.5 show that the SSICA is able to improve the monitoring performance in terms of detection delays over the CVA with KDE, DICA and DPLS with KDE methods for all the faults. Also, the SSICA is able to detect the faults earlier than the DPCA with KDE for all faults.

From Table 5.6, the false alarm rate of the DICA is lower than the false alarm rates of the other methods considered in this chapter. This is because the DICA is not as sensitive as the SSICA or the KDE approaches developed in this work. Furthermore, the insensitivity of the DICA approach can be attributed to the the use of the DPCA for the pre-processing stage of the DICA. Meanwhile, the false alarm rates of the SSICA are the same as those of the CVA with KDE and DPLS with KDE for all of the faults. In addition, the false alarm rates of the SSICA are lower ($> 0.59\%$) than

those of the DPCA with KDE for all the faults.

5.5   Chapter Summmary

In this chapter, a detailed description of the ASM1 model is presented. Three fault scenarios are created based on the ASM1 model to evaluate the monitoring methods developed in this work. Also, a comparison of the monitoring performance of the methods developed in this work is presented. Generally, the CVA with KDE, DPCA with KDE and the DPLS with KDE approaches are able to improve the monitoring performance over their non-KDE counterparts based on the criteria employed in this work for the assessment of the monitoring methods. Among the KDE approaches, the CVA with KDE outperformed the DPCA with KDE and the DPLS with KDE approaches. In this chapter also, the SSICA is compared with the KDE associated approaches. The results show that the SSICA is able to further improve the monitoring performance over the KDE associated approaches. In addition, the SSICA is compared with the DICA. Moreover, the performance of the SSICA is superior to that of the DICA approach. Figures and Tables have been employed to illustrate these results in this chapter.

Chapter 6

Conclusions and Future Work

This chapter presents a summary of the conclusions drawn from this work and concludes with a discussion of recommendations for future research.

6.1   Summary of Thesis

The work performed in this thesis can be divided into three parts:

- A literature review on process monitoring and various process monitoring methods

- Development of novel monitoring approaches

- Evaluation of the developed monitoring approaches using simulated industrial plants

6.1.1 Outcomes of Work

In this work, various monitoring techniques were developed and evaluated. Dynamic methods like the DPCA, DPLS and the CVA are extended to non-linear systems using the kernel density estimations. The resulting techniques are the novel DPCA with KDE, DPLS with KDE and the CVA with KDE. The KDE approaches were first evaluated by applying them to the Tenneessee Eastman Process Plant. The percentage reliability, detection delays and the false alarm rates are the parameters employed to assess the efficiency of the monitoring methods. The KDE approaches were compared with their non-KDE counterparts to illustrate the improvement in the condition monitoring performance of the KDE approaches over their non-KDE counter-parts. Generally, the KDE approaches developed in this work were able to improve the condition monitoring over their non-KDE counterparts. However, there were some cases for which the performance of the KDE approaches were the same as that of their non-KDE counter-parts. This could be because the efficiency of the KDE approaches over their non-KDE counter-parts is due to the fact that the KDE approaches take the non-Gaussian distributions which may be caused by the nonlinearity of the plant into account which the non-KDE approaches do not do. Also, the superiority of the KDE approaches is emphasised for faults that are non-linear in nature. Hence, for linear faults, the KDE approaches may not demonstrate advantages over their non-KDE counter-parts. Moreover, it is important to note that there was no case for which a non-KDE approach performed better than a KDE approach in this work.

Furthermore, the KDE approaches developed in this work were also compared one with another. The DPCA with KDE outperformed the DPLS with KDE for some faults, while for the remaining faults, the performance of the DPLS with KDE was better than that of the DPCA with KDE. Amongst the KDE approaches, the CVA with KDE was superior to the DPCA with KDE and the DPLS with KDE ap-

proaches for some faults while for the remaining faults, the performance of the CVA with KDE is the same as those of the DPCA with KDE and the DPLS with KDE. The performance of all the KDE approaches are the same for those faults that are easily detected by most monitoring methods. However, for faults that are difficult to detect by most monitoring methods, the performance of the CVA with KDE is significantly better and particularly emphasised. Note that there was no case for which the CVA with KDE was outperformed by any of the other two KDE approaches. This is because the CVA is a better tool to capture dynamic behaviour than the DPCA and DPLS approaches. In addition, all the KDE approaches developed in this work had zero false alarm rates for all the TEP faults employed in this work.

In this work also, a novel State Space Independent Component Analysis (SSICA) was developed. The proposed SSICA extends the ICA to dynamic systems thereby addressing the limitations of the ICA for dynamic systems. In the proposed SSICA approach, the CVA is first employed for the construction of the state space in the pre-processing stage before applying the ICA algorithm to the constructed state space for the extraction of the non-Gaussian ICs and then employing the KDE for the the the derivation of appropriate and efficient control limits.

To evaluate the proposed SSICA approach, it was also applied to the TEP plant. Furthermore, to demonstrate the improvement of the proposed SSICA over the CVA, the monitoring performance of the SSICA is compared with that of the CVA. In addition, the performance of the proposed SSICA is compared with that of the DICA because both methods have similar objectives. For some of the faults, the performance of the SSICA is better than those of the DICA and CVA approaches although the significance of improvement over the CVA was not as high as that over the DICA. However, the performance of the SSICA was the same as those of the CVA and DICA approaches for the remaining faults. Moreover, it is important to note that all the TEP faults employed in this work, there is no fault for which the

CVA or DICA approaches performed better than the proposed SSICA in terms of reliability or detection delays. In addition, the SSICA was also able to reduce the false alarm rates over the CVA and DICA methods. The advantages of the SSICA over the DICA is attributed to the fact that the proposed SSICA is first based on the CVA which is a better dynamic tool than the DPCA on which the DICA approach is based. Furthermore, the efficiency of the proposed SSICA over the CVA is owed to the fact that the SSICA is more suited than the CVA to deal with non-Gaussian process measurements, separating the original sources to a greater degree than the CVA technique.

The KDE approaches were also applied to a simulated WWTP to demonstrate their efficiency. For the fault scenarios employed in this work, the KDE approaches were able to improve the the monitoring performance over their non-KDE counter-parts. Also, the CVA with KDE was able to further improve the monitoring performance over the DPCA with KDE and DPLS with KDE methods for all the WWTP faults employed in this work.

In summary, the KDE approaches were developed in this work to be able to simultaneously address the dynamic and non-linear issues commonly associated with most industrial plant. Moreover, these KDE approaches are able to improve the monitoring performance over the traditional non-KDE approaches. Also, the proposed SSICA is an efficient attempt to extend the ICA to dynamic systems, thereby accounting for the dynamic properties as well as the non-linearity of most industrial.

6.2 Future Work

There are a number of recommendations for further research to enhance the quality of the research work described in this thesis.

- Process monitoring consists of fault detection and diagnosis. So far, the work in this research has been on fault detection with no work done on diagnosis. Therefore, it is recommended in future to do diagnosis based on the novel approaches developed in this work for fault detection. Existing methods for diagnosis can be employed with the methods developed in this work.

- Like most CVA-based approaches, the CVA with KDE approach employed in this work is developed using the past vector. However, the future vector could also be employed for the development of the CVA model. Therefore, it is recommended to build a CVA model based on the future vector rather than the past vector as is done in the current work.

- In the current work, the SSICA is based on the FASTICA which performs the random initialisation of the demixing matrix in the whitened space. This random initialisation can result in different results in the ICA algorithm. Hence it is recommended in this work to employ the estimated states from the CVA for the determination of the demixing matrix.

- Non-linear models are another possible way to improve the monitoring of non-linear plants. Hence, a non-linear extension of the proposed SSICA is expected to improve the monitoring perfromance over the proposed SSICA.

Bibliography

[1] W. Ku, R. H. Storer, and C. Georgakis. Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 30:179–196, 1995.

[2] T. Komulainen, M. Sourander, and S. Jamsa-Jounela. An online application of dynamic PLS to a dearomatization process. *Computers and Chemical Engineering*, 28:2611–2619, 2004.

[3] A. Negiz and A. Cinar. Monitoring of multivariate dynamic processes and sensor auditing. *Journal of Process Control*, 8:375–380, 1998.

[4] P. P. Odiowei and Y. Cao. Nonlinear dynamic process monitoring using canonical variate analysis and kernel density estimations. *IEEE Transactions on Industrial Informatics*, 6(1):36–45, 2009.

[5] P. P. Odiowei and Y. Cao. State space independent component analysis for nonlinear dynamic process monitoring. *Chemometrics Intelligent Laboratory Systems*, 103:59–65, 2010.

[6] A. Negiz and A. Cinar. PLS, balanced and canonical variate realization techniques for identifying varma models in state space. *Chemometrics and Intelligent Laboratory Systems*, 38:209–221, 1997.

[7] A. Negiz and A. Cinar. Statistical monitoring of multivariable dynamic processes with state space models. *AIChE*, 43:2002–2020, 1997.

[8] T. J. Richard, K. Uwe, and E. C. Jonathan. Dynamic multivariate statistical process control using subspace identification. *Journal of Process Control*, 14:279–292, 2004.

[9] A. Norvalis, A. Negiz, and J. DeCicco. Monitoring by interfacing knowledge-based systems and multivariate statistical monitoring. *Journal of Process Control*, 10:341–350, 2000.

[10] A. Simouglou, E. B. Martin, and A. J. Morris. Statistical performance monitoring of dynamic multivariate processes using state space modelling. *Computers and Chemical Engineering*, 26:909–920, 2002.

[11] B. C. Juricek, D. E. Seborg, and W. E. Larimore. Fault detection using canonical variate analysis. *Ind. Eng. Chem. Res*, 43:458–474, 2004.

[12] L. Juan and L. Fei. Statistical modelling of dynamic multivariate process using canonical variate analysis. *Proceedings of IEEE International Conference of Information and Automation*, pages 218–221, 2006.

[13] E. B. Martin and A. J. Morris. Non-parametric confidence bounds for process performance monitoring charts. *Journal of Process Control*, 6(6):349–358, 1996.

[14] A. W Bowman and A. Azzalini. *Applied Smoothing Techniques for Data Analysis, The Kernel Approach with S-plus Illustrations*. Oxford, U.K.:Clarendon Press, 1997.

[15] J. Lee, C. Yoo, and I. Lee. Statistical process monitoring with independent component analysis. *Journal of Process Control*, 14:467–485, 2004.

[16] G. Chen, J. liang, and J. Qian. Chemical process monitoring and fault diagnosis based on independent component analysis. *5th World Congress on Intelligent Control and Automation*, page 1646, 2004.

[17] M. Wang. *Impact of Dissolved Oxygen Performance on the Operational Cost of Activated Sludge Process*. PhD thesis, Cranfield University, School of Engineering, 2007.

[18] C. Rosen. *A Chemometric Approach to Process Monitoring and Control with Applications to Wastewater Treatment Operation*. PhD thesis, Lund University, Department of Industrial Electrical Engineering and Automation, 2001.

[19] C. Lee, S. W. Choi, and I. B. Lee. Sensor fault diagnosis in a wastewater treatment process. *Water Science and Technology*, 53(1):251–257, 2006.

[20] R. Isermann. Model based fault detection and diagnosis methods. *Proceedings of American Control Conference, Piscataway, New Jersey, IEEE Press*, pages 1605–1609, 1995.

[21] L. H. Chiang, E. L. Russell, and R. D. Braatz. *Fault Detection and Diagnosis in Industrial Systems*. Springer, London, 2001.

[22] C. Xiong, Y. Zhao, and W. Liu. Fault detection method based on artificial immune system for complicated process. *International Conference on Intelligent Computing*, pages 625–630, 2006, Kunming, China, August 2006 Proceedings, Part II.

[23] R. Isermann. Process fault detection based on modelling and estimation methods: A survey. *Automatica*, 20:387–404, 1984.

[24] R. Isermann. Fault diagnosis of machines via parameter estimation and knowledge processing - tutorial paper. *Automatica*, 29:815–835, 1993.

[25] R. K. Mehra and J. Peschon. An innovations approach to fault detection and diagnosis in dynamic systems. *Automatica*, 7(4):637–640, 1971.

[26] M. Kitamura. Detection of sensor failures in nuclear plant using analytic redundancy. *Trans. Am. Nucl. Soc.*, 34:581–583, 1980.

[27] J. J. Gertler. *Fault Detection and Diagnosis in Engineering Systems*. Marcel Dekker, Inc., New York, 1998.

[28] A. Pouliezos, G. Stavrakakis, and C. Lefas. Fault detection using parameter estimation. *Quality and Reliability Engineering International*, 5:283–290, 1989.

[29] O. Moseler and R. Isermann. Model-based fault detection for a brushless DC motor using parameter estimation. *Proceedings of the 24th Annual Conference of the IEEE industrial Electronics Society, IECON*, 4:956–1960, 1998 Piscataway, New Jersey.

[30] P. M. Frank. Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy- a survey and some new results. *Automatica*, 26:459–474, 1990.

[31] R. Isermann and P. Ball. Trends in the application of model-based fault detection and diagnosis of technical processes. *Proceedings of the 13th IFAC World Congress*, N:1–12, 1998, Piscataway, New Jersey, IEEE Press.

[32] J. V. Beck and K. J. Arnold. *Parameter Estimation in Engineering and Science*. Wiley, New York, 1977.

[33] L. Ljung. *System Identification: Theory for the User*. Prentice-Hall, Englewood Cliffs, New Jersey, 1987.

[34] P. M. Frank and X. Ding. Survey of robust residual generation and evaluation methods in observer-based fault detection systems. *Journal of Process Control*, 7(6):403–424, 1997.

[35] A. S. Willsky. A survey on design methods for failure detection in dynamic systems. *Automatica*, 12:601–611, 1976.

[36] L. A. Mironovoski. Functional diagnosis of linear dynamic systems. *Automation and Remote Control*, 40:1198–1205, 1979.

[37] G. Lee, B. Lee, E. S. Yoon, and C. Han. Multiple-fault diagnosis under uncertain conditions by the quantification of qualitative relations. *Ind. eng. Chem. Res*, 38:988–998, 1999.

[38] K. J. Mo, G. Lee, D. S. Nam, Y. H. Yoon, and E. S. Yoon. Robust fault diagnosis based on clustered sympton trees. *Control Engineering Practice*, 5:199–208, 1979.

[39] K. J. Mo, Y. S. Oh, and E. S. Yoon. Development of operation-aided system for chemical processes. *Expert Systems with Applications*, 12:455–464, 1979.

[40] M. A. Kramer and F. E. Finch. *Fault Diagnosis of Chemical Processes*. Plenum Press, New York, 1989, In S. G. Tzafestas, editor, Knowledge-based System Diagnosis, Supervision and Control.

[41] R. S. H. Mah, K. D. Schnelle, and A. N. Patel. A plant-wide quality expert system for steel mills. *Computers and Chemical Engineering*, 15:445–450, 1991.

[42] R. Vaidhyanathan and V. Venkatasubramanian. Experience with an expert system for automated hazop analysis. *Computers and Chemical Engineering*, 20:1589–1594, 1996.

[43] R. J. Schalkoff. *Pattern Recognition: Statistical, Structural and Neural Approaches*. John Wiley and Sons, New York, 1992.

[44] J. F. MacGregor and T. Kourti. Statistical process control of multivariate processes. *Control Engineering Practice*, 3(3):403–414, 1995.

[45] J. E. Jackson. Quality control methods for several related variables. *Technometrics*, 1(4):359–377, 1959.

[46] P. Geladi. Notes on history and nature of partial least squares (PLS) modelling. *Journal of Chemometrics*, 2:231–246, 1988.

[47] J. Kresta, J. F. MacGregor, and T. E. Marlin. Multivariate statistical monitoring of process operating performance. *Canadian Journal of Chemical Engineers*, 69:35–47, 1991.

[48] T. Kourti and J. F. MacGregor. Multivariate spc methods for process and product monitoring. *Journal of Quality Technology*, 28:409–428, 1996.

[49] J. F. MacGregor. Statistical process control of multivariate processes. *Proceedings of IFAC Int. Symp. on Advanced Control of Chemical Processes*, pages 427–435, 1994, New York.

[50] B. Lennox, G. A. Montague, H. G. Hiden, G. Kornfeld, and P. R. Goulding. Process monitoring of an industrial fed-batch fermentation. *Biotechnology and Bioengineering*, 74(2):125–135, 2001.

[51] J. D. Healy. A note on multivariate cusum procedures. *Technometrics*, 29:409–412, 1987.

[52] R. B. Crosier. Multivariate generalizations of cumulative sum quality-control schemes. *Technometrics*, 30:291–303, 1988.

[53] C. A. Lowry, W. H. Woodall, C. W. Champ, and S. E. Rigdon. A multivariate exponentially weighted moving average control chart. *Technometrics*, 34:46–53, 1992.

[54] L. H. Chiang, E. L. Russell, and R. D. Braatz. Fault diagnosis in chemical processes using Fisher discriminant analysis, discriminant partial least squares, and principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 50:243–252, 1999.

[55] J. F. MacGregor, H. Yu, S. G. Munoz, and J. Flores-Cerrillo. Data-based latent variable methods for process analysis monitoring and control. *Computers and Chemical Engineering*, 29:1217–1223, 2005.

[56] S. J. Qin. Statistical process monitoring: Basics and beyond. *Journal of Chemometrics*, 17:480–502, 2003.

[57] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophhical Magazine*, 2(6):559–572, 1901.

[58] H. Hotelling. Analysis of complex statistical variables into principal components. *J. Educ. Psych.*, 24:417–441, 1933.

[59] P. D. Wentzell, D. Andrews, D. C. Hamilton, K. Faber, and B. R. Kowalski. Maximum likelihood principal component analysis. *Journal of Chemometrics*, 11:339–366, 1997.

[60] E. R. Malinowski. *Factor Analysis in Chemistry*. Wiley:New York, 1991.

[61] L. H Chiang, E. L. Russell, and R. D. Braatz. Fault diagnosis in chemical processes using fisher discriminant analysis, discriminant partial least squares and principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 50:243–252, 2000.

[62] X. Z. Wang, S. Medasani, F. Marhoon, and H. Albazzaz. Multidimensional visualization of principal compnent analysis. *Ind. Eng. Chem. Res*, 43:7036–7048, 2004.

[63] T. Kourti and J. F. MacGregor. Process analysis, monitoring and diagnosis using multivariate projection methods. *Chemometrics and Intelligent Laboratory Systems*, 28:3–21, 1995.

[64] R. Dunia, S. J. Qin, T. F. Edgar, and T. J. McAvoy. Identification of faulty sensors using principal component analysis. *American Institute of Chemical Engineers Journal*, 42(10):2797–2812, 1996.

[65] W. Lin, Y. Qian, and X. Li. Nonlinear dynamic principal component analysis for on-line process monitoring and diagnosis. *Computers and Chemical Engineering*, 24:423–429, 2000.

[66] H. Hotelling. Relation between two sets of variates. *Biometrika*, 28:321–377, 1936.

[67] J. E Jackson and G. S. Modholkar. Control proceedures for residual associated with principal component analysis. *Technometric*, 21:341–349, 1979.

[68] S. Wang and J. Cui. Sensor-fault detection, diagnosis and estimation for centrifugal chiller systems using principal-component analysis method. *Applied Energy*, 82:197–213, 2005.

[69] P. Nomikos and J. F. MacGregor. Monitoring batch processes using multiway principal component analysis. *American Institute of Chemical Engineering Journal*, 40(8):1361–1375, 1994.

[70] P. Nomikos and J. F. MacGregor. Multivariate spc charts for monitoring batch processes. *Technometrics*, 37:41–59, 1995.

[71] K. A. Kosanovich, M. J. Piovoso, K. S. Dahl, J. F. MacGregor, and P. Nomikos. Multi-way PCA applied to an industrial batch process. *In Proc. of American Control Conference*, pages 1294–1298, 1994, Balimore, Maryland, USA.

[72] H. H. Yue and S. J. Qin. Reconstruction-based fault identification using a combined index. *Ind. Eng. Chem. Res*, 40:4403–4414, 2001.

[73] Q. Chen, U. Kruger, M. Meronk, and A. Y. T. Leung. Synthesis of T^2 and Q statistics for process monitoring. *Control Engineering Practice*, 12:745–755, 2004.

[74] M. E. Tipping and C. M. Bishop. Probabilistic principal component analysis. *Journal of Royal Statistical Society B*, 61(3):611–622, 1999.

[75] T. Chen, J. Morris, and E. Martin. Probability density estimation via an infinite gaussian mixture model: Application to statistical process monitoring. *Appl. Statist*, 55:699–715, 2006.

[76] T. Chen and Y. Sun. Probabilistic contribution analysis for statistical process monitoring: A missing variable approach. *Control Engineering Practice*, 7(4):469–477, 2009.

[77] V. Kariwala, P. Odiowei, Y. Cao, and T Chen. A branch and bound method for isolation of faulty variables analysis. *Journal of Process Control*, 20(10):1198–1206, 2010.

[78] J. Mina and C. Verde. Fault detection for large scale systems using dynamic principal components analysis with adaptation. *International Journal of Computers, Communication and Control*, 2(2):185–194, 2007.

[79] F. Tsung. Statistical monitoring and diagnosis of automatic controlled processes using dynamic PCA. *International Journal of Production Research*, 38(3):625–637, 2000.

[80] I. T. Jolliffe. A note on the use of principal components in regression. *Journal of the Royal Statistical Society C (Applied Statistics)*, 31(3):300–303, 1982.

[81] H. Wold. *Estimation of Principal Components and related methods by iterative least squares*. New York: Academic Press, 1966, In P. R. Krishnaiah (Ed.), Mutivariate Analysis.

[82] S. Wold, M. Sjostrom, and L. Eriksson. PLS regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58:109–130, 2001.

[83] S. Wold, N. Kettaneh, and B. Skagerberg. Nonlinear PLS modelling. *Chemometric Intelligent Laboratory Systems*, 7(1-2):53–65, 1989.

[84] B. M. Wise and N. L. Ricker. Feedback strategies in multiple sensor systems. *AIChE Symposium Series on Process Sensing*, 85(267), 1989.

[85] R. Ergon. Informative PLS score-loading plots for process understanding and monitoring. *Journal of Process Control*, 14:889–897, 2004.

[86] S. Wold. Cross-validation estimation of the number of components in factor and principal component analysis. *Technometrics*, 20:397–405, 1978.

[87] G. Li, S. J. Qin, and D. Zhou. Geometric properties of partial least squares for process monitoring. *Automatica*, 46:204–210, 2010.

[88] U. Kruger, Y. Zhou, X. Wang, D. Rooney, and J. Thompson. Robust partial least squares regression: Part I, algorithm developments. *Chemometrics*, 22:1–13, 2007.

[89] P. Geladi and B. R. Kowalski. An example of 2-block predictive partial least-squares regression with simulated data. *Analytica Chimica ACTA*, 185:19–32, 1986.

[90] A. Hoskuldsson. PLS regression methods. *Journal of Chemometrics*, 2:211–228, 1998.

[91] B. Lennox. Integrating fault detection and isolation with model predictive control. *International Journal of Adaptive Control and Signal Processing*, 19:199–212, 2005.

[92] J. J Downs and E. Vogel. A plant-wide indsutrial process control problem. *Comput. Chem. Eng*, 17:245–255, 1993.

[93] M. Henze, C. P. L. Grady Jr., W. Gujer, G. Marais, and T. Matsuo. Activated sludge model no. 1. *Scientific and Technical Report No. 1*, 1987, IAWQ, London, UK.

[94] R. Srinivasan, C. Wang, W. K. Ho, and K. W. Lim. Dynamic principal components analysis based methodology for clustering process states in agile chemical plants. *Ind. eng. Chem. Res*, 43:2123–2139, 2004.

[95] D. S. Lee, M. W. Lee, S. H. Woo, Y. Kim, and J. M. Park. Nonlinear dynamic partial least squares modelling of a full-scale biological wastewater treatment plant. *Process Biochemistry*, 41:2050–2057, 2006.

[96] N. M. FLetcher, A. J. Morris, and E. B. Martin. Local linear and non-linear multi-way partial least squares batch modelling. *15th Triennial World Congress*, 2002, Barcelona, Spain.

[97] J. Chen and K. Liu. On-line batch process monitoring using dynamic pca and dynamic pls models. *Chemical Engineering Science*, 57:63–75, 2002.

[98] A. Simoglou, E. B. Martin, and A. J. Morris. A comparison of canonical variate analysis and partial least squares for the identification of dynamic processes. *Proceedings of the American Control Conference*, pages 832–837, 1999, San Diego, California.

[99] W. E. Larimore. System identification reduced order filtering and modelling via canonical correlation analysis. *Proceedings of the American Control Conference*, pages 445–449, 1983.

[100] C. D. Schaper, W. E. Larimore, D. E. Seborg, and D. A. Mellichamp. Identification of chemical processes using canonical variate analysis. *Computers Chemical Engineering*, 18(1), 1994.

[101] A. Simouglou, E. B. Martin, and A. J. Morris. Dynamic multivariate statistical process control using partial least squares and canonical variate analysis. *Computers and Chemical Engineering*, pages 277–280, 1999.

[102] S. Wold. Exponentially weighted moving principal component analysis and projection to latent structures. *Chemometrics Intelligent Laboratory Systems*, 23:149–161, 1994.

[103] S. J. Qin. Recursive pls algorithms for adaptive data modelling. *Computers and Chemical Engineering*, 22(4/5):503–514, 1998.

[104] L. M. Elshenawy, S. Yin, A. S. Naik, and S. X. Ding. Efficient recursive principal component analysis algorithms for process monitoring. *Ind. Eng. Chem. Res*, 49:252–259, 2010.

[105] N. B. Gallagher, B. M. Wise, S. W. Butler, D. D. White, and G. G. Barna. Development and benchmarking of multivariate statistical process control tools for a semiconductor etch process: Improving robustness through model updating. *Proceedings of the ADCHEM 97*, pages 78–83, 1997, Banff, Canada.

[106] W. Li, H. H. Yue, S. Valle-Cervantes, and S. J. Qin. Recursive pca for adaptive process monitoring. *Process Control*, 10:471–486, 2000.

[107] K. Helland, H. E. Berntsen, O. S. Borgen, and H. Martens. Recursive algorithm for partial least squares regression. *Chemometrics Intelligent Laboratory Systems*, 14:129–137, 1991.

[108] X. Wang, U. Kruger, and B. Lennox. Recursive partial least squares algorithms for monitoring complex industrial processes. *Control Engineering Practice*, 11:613–632, 2003.

[109] F. Jia, E. B. Martin, and A. J. Morris. Non-linear principal component analysis with application to process fault detection. *International Journal of Systems Science*, 31(11):1473–1487, 2000.

[110] L. Xu, E. Oja, and C. Y. Suen. Modified hebbian learning for curve and surface fitting. *Neural Networks*, 5:441–457, 1992.

[111] D. Dong and T. J. McAvoy. Nonlinear principal component analysis based on princial curves and neural network. *Computers and Chemical Engineers*, 20(1):65–78, 1996.

[112] C. F. Alcala and S. J. Qin. Reconstruction-based contribution for process monitoring with kernel principal component analysis. *Ind. Eng. Chem. Res*, 49(17):7849–7857, 2010.

[113] M. Kramer. Nonlinear principal component analysis using autoassociative neural network. *AIChE*, 37:233–243, 1991.

[114] D. Dong and T. J. McAvoy. Nonlinear principal component analysis based on princial curves and neural network. *Proceedings of American Control Conference*, pages 1284–1288, 1994.

[115] B. Scholkopf, A. Smola, and K. Muller. Nonlinear component analysis as a kernel eigen value problem. *Neural Comput.*, 10:1299–1319, 1998.

[116] B. Scholkopf, A. Smola, and K. Muller. Nonlinear process monitoring using kernel principal component analysis. *Chemical Engineering Science*, 59:223–234, 2004.

[117] F. Jia, E. B. Martin, and A. J. Morris. Non-linear principal component analysis with application to process fault detection. *International Journal of Systems Science*, 31(11):1473–1487, 2001.

[118] S. Mika, B. Scholkopf, A. J. Smola, K. R. Muller, M. Scholz, and G. Ratsch. Kernel pca and denoising in feature spaces. *Advances in Neural Information Processing Systems*, 11:536–542, 1999.

[119] S. Romdhani, S. Gong, and A. Psarrou. A multi-view nonlinear active shape modelling using kernel pca. *Proceedings of BMVC*, pages 483–492, 1999, Nottingham, UK.

161

[120] J. Cho, J. Lee, S. W. Choi, D. Lee, and I. Lee. Fault identification for process monitoring using kernel principal component analysis. *Chemical Engineering Science*, 60:279–288, 2005.

[121] H. Albazzaz and X. Z. Wang. Historical data analysis based on plots of independent and parallel co-ordinates and statistical control limits. *Journal of Process Control*, 16:103–114, 2006.

[122] A. D. Back and A. S Weigend. A first application of independent component analysis to extracting structure from stock returns. *Int. J. Neural Sys*, 8:473–484, 1997.

[123] J. Lee, C. Yoo, and I. Lee. Statistical monitoring of dynamic independent component analysis. *Chemical Engineering Sciences*, 59:2295–3006, 2004.

[124] T. Lee. *Independent Component Analysis: Theory and Applications*. Kluwer Academic Publishers, Boston, USA, 1998.

[125] A. Hyvarinen. Survey on independent component analysis. *Neural Comput. Surveys*, 2:94–128, 1999.

[126] J. Lee, S. J. Qin, and I. Lee. Fault detection and diagnosis based on modified independent component analysis. *AIChEJournal*, 52(10):3501–3514, 2006.

[127] G. Stefatos and A. B. Hamza. Dynamic independent component analysis approach for fault detection and diagnosis. *Expert Systems with Applications*, 37:8606–8617, 2010.

[128] L. Xiong, J. Liang, and J. Qian. Multivariate statistical process monitoring of an industrial polypropylene catalyzer reactor with component analysis and kernel density estimation. *Chinese Journal of Chemical Engineering*, 15(4):524–532, 2007.

[129] A. Hyvarinen and E. Oja. Independent component analysis: Algorithm and applications. *Neural Networks*, 13:411–430, 2000.

[130] A. Hyvarinen and E. Oja. A fast fixed point algorithm for independent component analysis. *Neural Computation*, 9:1483–1492, 1997.

[131] M. Kano, S. Tanaka, S. Hasebe, I. Hashimoto, and H. Ohno. Monitoring independent components for fault detection. *AIChE Journal*, 49:969–976, 2003.

[132] H. Albazzaz and X. Z. Wang. Statistical process control charts for batch operations based on independent component analysis. *Industrial and Engineering Chemistry Research*, 43(21):6731–6741, 2004.

[133] X. Liu, L. Xie, U. Kruger, T. Littler, and S. Wang. Statistical-based monitoring of multivariate non-gaussian systems analysis. *AIChE Journal*, 54(9):2379–2391, 2008.

[134] L. Hongguang and G. Hui. The application of independent component analysis in process monitoring. *Proceedings of the 1st International Conference on Innovative Computing, Information and Control*, page 97, 2006.

[135] Q. Chen, P. Goulding, D. Sandoz, and R. Wyne. Application of kernel density estimates to condition monitoring for process industries. *American Control Conference, Philadelphia, PA, USA*, 1998.

[136] M. C. Jones, J. S. Marron, and S. J. Sheather. A brief survey of band-width selection for density estimation. *J. Amer. Stat. Assoc.*, 91(433):401–407, 1996.

[137] S. Xiaoping and A. Sonali. Kernel density estimation for an anomaly based intrusion detection system. *Proceedings of the 2006 World Congress in Computer, Computer Science and Engineering and Applied Computing 2006*, page 161, 2006.

163

[138] P. P. Odiowei and Y. Cao. Kernel density enhanced PCA for process monitoring of a waste water treatment plant (WWTP). 1st *Postgraduate Research Conference, 29th - 30th September, Dundee, United Kingdom*, pages 123–129, 2008.

[139] N. Johan, L. Fredrik, H. Bjorn, K. Walter, J. M. V. Henk, and L. M. H. Joop. Classification of environmentally occurring chemicals using structural fragments and pls discriminant analysis. *Environ. Sci. Technol.*, 31(8):2313–2318, 1997.

[140] D. J. H. Wilson and G. W. Irwin. Pls modelling and fault detection on the tennessee eastman benchmark. *International Journal of Systems Science*, 31(11):1449–1457, 2000.

[141] Q. Chen and U. Kruger. Analysis of extended partial least squares for monitoring large-scale processes. *IEEE Transactions on Control Systems Technology*, 13(5):807–813, 2005.

[142] K. Baumannn, H. Albert, and M. V. Korff. A systematic evaluation of the benefits and hazards of variable selection in latent variable regression. Part 1. search algorithm, theory and simulations. *Journal of Chemometrics*, 75:339–350, 2002.

[143] P. Nomikos and J. K. MacGregor. Multi-way partial least squares in monitoring batch processes. *Chemometrics and Intelligent Laboratory Systems*, 30:97–108, 1995.

[144] M. Kano, K. Nagoa, S. Hasebe, I. Hashimoto, H. Ohno, R. Strauss, and B. R. Bakshi. Comparison of multivariate statistical process monitoring methods with applications to eastman challenge problem. *Comput. Chem. Eng*, 26:161–174, 2002.

[145] A. Hyvarinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Nueral*, 3:628–634, 1999.

[146] J. Lennox and C. Rosen. Adaptive multiscale principal component analysis for online monitoring of wastewater treatment. *Water Science and Technology*, 45(4-5):227–235, 2002.

[147] W. E. Larimore. Statistical optimality and canonical variate analysis system identification. *Signal Process*, 52:131–144, 1996.

[148] D. Hodge, L. Simon, and M. N. Karim. Data driven approaches to modelling and analysis of bioprocesses: Some industrial examples. *Proceedings of the American Control Conference*, pages 2062–2076, 2003, Denver, Colorado.

[149] W. Ku, R. H. Storer, and C. Georgakis. Uses of principal component analysis for disturbance detection and isolation. *Proceedings of CIMPROM*, 1994.

[150] T. R. Willemain and G. C. Runger. Designing control charts using an empirical reference distribution. *Journal of Quality Technology*, pages 31–38, 1996.

[151] N. Lu, Y. Yao, F. Gao, and F. Wang. Two-dimensional dynamic PCA for batch process monitoring. *AIChE Journal*, 51(12):3300–3304, 2005.

[152] B. Li, J. Morris, and E. B. Martin. Model selection for partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, 64:79–89, 2002.

[153] H. Akaike. Markovian representation of stochastic processes by canonical variables. *SIAMJ. Contr.*, 13:162–173, 1975.

[154] M. Scholz, F. Kaplan, C. L. Guy, J Kopka, and J. Selbig. Non-linear PCA: a missing data approach. *Bioinformatics*, 21(20):3887–3895, 2005.

[155] J. E. Jackson. Multivariate quality control. *Commun. Stat.-Theory Meth.*, 14:2657–2688, 1985.

[156] F. B. Alt. Multivariate control charts. *Encycclopedia of Statistical Science*, 6:110–122, 1985.

[157] C. Kiparissides J. F. MacGregor, C. Jaeckle and M. kotoudi. Process monitoring and diagnosis by multiblock pls methods. *AIChE Journal*, 40(5):826–838, 1994.

[158] T. Kourti, P. Nomikos, and J. F. MacGregor. Analysis, monitoring and fault diagnosis of batch processes using multiblock and multiway PLS. *Journal of Process Control*, 5(4):277–284, 1995.

[159] F. Jia, E. B. Martin, and A. J. Morris. Nonlinear principal component analysis for process fault detection. *Computers and Chemical Engineering*, 22(267):851–854, 1998.

[160] M. J. Hammer. *Water and Wastewater Technology*. Prentice Hall, Inc., Englewood Cliffs, New Jersey, USA, 1986.

[161] B. Wise and N. B. Gallagher. The process chemometric approach to process monitoring and fault detection. *Journal of Process Control*, 6:329–348, 1996.

[162] E. L. Russell, L. H. Chiang, and R. D. Braatz. Fault detection in industrial processes using canonical variate analysis and dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 51:81–93, 2000.

[163] K. Pearson. Contributions to the mathematical theory of evolution. ii. skew variation in homogeneous material. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and engineering Sciences*, 186:343–414, 1895.

[164] W. L. Brogan. *Model Control Theory*. Prentice-Hall, New Jersey, 1991.

[165] C. T. Chen. *Linear System Theory and Design*. Harcourt Brace College Publishers, Orlando, Florida, 1984.

[166] J. Zhang, E. B. Martin, and A. J. Morris. Process monitoring using non-linear statistical techniques. *Chemical Engineering Journal*, 67(3):181–189, 1997.

[167] F. Teymour. Dynamics of semibatch polymerization reactors: I. theoretical analysis. *AIChE Journal*, 43(1):145–156, 1997.

[168] J. F. Andrews. Dynamics and control of activated sludge. *Journal of Industrial Engineering and Management*, 2(3):464–498, 1998.