Model-based Performance Monitoring of Batch Processes

By

Lindsay Anne McPherson

A Thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Department of Chemical Engineering and Advanced Materials

The University of Newcastle Upon Tyne

2007

ABSTRACT

The use of batch processes is widespread across the manufacturing industries, dominating sectors such as pharmaceuticals, speciality chemicals and biochemicals. The main goal in batch production is to manufacture consistent, high quality batches with minimum rework or spoilage and also to achieve the optimum energy and feedstock usage. A common approach to monitoring a batch process to achieve this goal is to use a recipe-driven approach coupled with off-line laboratory analysis of the product. However, the large amount of data generated during batch manufacture mean that it is possible to monitor batch processes using a statistical model. Traditional multivariate statistical techniques such as principal component analysis and partial least squares were originally developed for use on continuous processes, which means they are less able to cope with the non-linear and dynamic behaviours inherent within a batch process without being adapted. Several approaches to dealing with batch behaviour in a multivariate framework have been proposed including multi-way principal component analysis.

A more advanced approach designed to handle the typical characteristics of batch data is that of model-based principal component. It comprises of a mechanistic model combined with a multivariate statistical technique. More specifically, the technique uses a mechanistic model of the process to generate a set of residuals from the measured process variables. The theory being that the non-linear behaviour and the serial correlation in the process will be captured by the model, leaving a set of unstructured residuals to which principal component analysis (PCA) can be applied. This approach is benchmarked against the more standard approaches including multiway principal components analysis, batch observation level analysis.

One limitation identified of the model-based approach is that if the mechanistic model of the process is of reduced complexity then the monitoring and fault detection abilities of the technique will be compromised. To address this issue, the model-based PCA technique has been extended to incorporate an additional error model which captures the differences between the mechanistic model and the process. This approach has been termed super model-based PCA (SMBPCA). A number of different error models are considered including partial least squares (linear, non-linear and dynamic), autoregressive with exogenous (ARX) variables model and dynamic canonical correlation analysis. Through the use of an exothermic batch reactor simulation, the SMBPCA approach has been investigated with respect to fault detection and capturing

the non-linear and dynamic behaviour in the batch process. The robustness of the technique for application in an industrial situation is also discussed.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Professor Elaine B. Martin, Professor of Industrial Statistics, Department of Chemical Engineering and Advanced Materials for her encouragement, support, guidance and patience throughout my PhD. I would also like to thank Professor A. Julian Morris, Department of Chemical Engineering and Advanced Materials for his encouragement in completing my thesis.

I would also like to thank my colleagues in CPACT, Nicky and Ewan, my family and Zaid.

TABLE OF CONTENTS

TABLE OF FIGURES

1 CHAPTER ONE: OVERVIEW

1.1 Introduction

The use of batch processes is widespread across the manufacturing industries, dominating sectors such as pharmaceuticals, speciality chemicals and biochemicals. The main goal in batch production is to manufacture consistent, high quality batches with minimum rework or spoilage and also to achieve the optimum energy and feedstock usage.

In the past, the most commonly utilised method for operating a batch process to achieve these goals was the recipe driven approach, coupled with off-line laboratory-based analysis to determine whether the product is within specification. The batch recipe defines the sequence of process operations required to make the product, along with the corresponding quantities of raw materials and the desired operating conditions. This information can be determined through knowledge of the process, experimental design or simply by recording the critical parameters on previous successful batches and using these values as a template to replicate the quality in subsequent batches. To determine whether the batch has been successful, samples of the product are taken for analysis in the laboratory and the results of this analysis determine whether the batch satisfies the criteria for release.

There are a number of drawbacks associated with this approach. For example using off-line laboratory analysis to determine whether a batch has met the specifications for release can lead to problems. Firstly, the nature of off-line testing means that the batch is often complete by the time the results are received. Consequently non-conformance is only detected when it is too late to take any corrective action. This can result in a large amount of either re-work or the disposal of valuable product. Also, waiting for laboratory results materialises in a time lag which may slow down the release of the product.

In addition to this, the large amounts of data generated during the batch run are generally ignored as is the potentially valuable information about critical process parameters that could be gained from its analysis. Also, the recipe-driven technique is not consistent in terms of reproducing batch quality, since the recipe is often 'plugged in' and then left to run to completion with little or no intervention from the operators to optimise the operating conditions.

Retaining the market share, or expanding the customer base, in an increasingly aggressive global market, and ensuring that the progressively more rigorous regulations on health, safety and the environment are met, necessitates the use of more efficient and effective batch monitoring systems that utilize the large volumes of process data. Valuable information about how different process variables react and interact and the impact of these interactions on the final product can be ascertained if the correct variables are measured, monitored and analysed. Using this information can have a number of advantages including improving plant safety, reducing environmental and health risks, reducing operating costs through the early diagnosis of process malfunctions, the easier location of the source of the problem, a reduction in off-specification product, the avoidance of re-working batches, improved quality control and in general better process understanding.

There are currently a range of techniques available for the monitoring and/or control of batch processes. In this thesis the focus is on the use of multivariate statistical techniques for the monitoring of batch processes. A large number of variables can be, and are, measured on a batch process including temperatures, pressures, concentration, pH, viscosity, mixing speed, and flowrates, along with secondary measurements for variables such as concentration calculated using spectroscopic techniques such as near infrared, mid infrared and raman. Typically more than one of these variables impacts on the behaviour of a process which is why univariate techniques that focus on the analysis of one variable at a time can be inappropriate for batch monitoring, especially since it is often the interactions between these variables that result in the onset of abnormal conditions or deviations in the process. Therefore a multivariate approach that considers the behaviour of more than one variable is essential for efficient and effective batch monitoring.

1.2 Thesis Outline and Contributions

In Chapter 2 of the thesis, the basic multivariate statistical techniques of principal component analysis and partial least squares are introduced. The different multivariate approaches available for the monitoring of batch processes are discussed in Chapter 3, along with the issues of non-linearity and serial correlation and batches that are of variable duration. Traditional multivariate statistical techniques are based on the assumption that the data being analysed is linear and that the measurements are independent, i.e. not correlated in time. Therefore to apply these methods to batch data

they must be modified accordingly. In chapter 3 the standard techniques of multi-way Principal Component Analysis (PCA), multi-way Partial Least Squares (PLS) and Batch Observation Level PCA (BOL) are introduced. Multi-way PCA and multi-way PLS have been adapted from the standard techniques to deal with batch data, and the claim that they capture the non-linear behaviour is discussed. Additionally non-linear PCA and non-linear PLS are discussed as an alternative approach. Methodologies proposed to take the serial correlation into account include moving window PCA and adaptive PCA, along with the dynamic versions of PCA and PLS and state space models. These methods are also described briefly.

Chapter 4 also discusses batch monitoring techniques, but this time focussing on techniques that incorporate specific information about the process being modelled into the empirical model, such as the physicochemical relationships occurring in the process. One such technique discussed is Model-based Principal Component Analysis (MBPCA) which forms the basis of subsequent chapters. It is a method for combining a mechanistic model with a multivariate statistical technique and is designed to handle the typical characteristics of batch data. More specifically, the technique uses a mechanistic model of the process to generate a set of residuals from the measured process variables. The theory being that the non-linear behaviour and the serial correlation in the process will be described by the model, leaving a set of unstructured residuals to which principal component analysis can be applied. Other techniques discussed that utilise specific process information include variable transformations, data reconciliation and model identification.

Chapter 5 gives an in-depth explanation of the mass and energy balances and control algorithms (Luyben (1973)) that were used to develop a simulation of an exothermic batch reactor on which the different techniques for the monitoring of batch processes were tested.

Building on this, Chapter 6 describes a case study that utilizes the simulation of the exothermic batch reactor, the objective being to consider a number of the batch monitoring techniques introduced in Chapters 3 and 4 and to compare their performance, with respect to how well each technique deals with the batch characteristics of non-linear behaviour and serial correlation and how quickly an abnormality is detected and how easily the source of the abnormality can be identified from the information available. The initial study was based on the research of Lewin *et al* (1998), who applied Model-based PCA to the simulation of the batch reactor. By replicating parts of this research

and undertaking a similar study using multiway PCA and batch observation level PCA, a basis for the comparison of the batch monitoring abilities of each technique was formed.

In carrying out the case study in Chapter 6, the advantages and disadvantages of model-based PCA technique are highlighted. Although the concept of model-based PCA has the advantage of taking into account the non-linear and dynamic behaviour inherent in batch data, being able to cope with batches of different duration, and the ability to combine real process knowledge in the form of a mechanistic model with a multivariate model, the performance of the technique is dependent on the accuracy of the mechanistic model. With a reduced complexity model, that is not an exact match of the process, there will still be structure remaining in the residuals, thereby significantly reducing the fault detection ability of the technique.

This limitation led to the development of super model-based principal component analysis, which is introduced in Chapter 7. This technique builds on the original idea of MBPCA, but makes it more robust, by incorporating an additional error/residual model into the algorithm. This additional model removes any remaining structure present as a consequence of the reduced complexity of the model, providing a set of white noise residuals from which process abnormalities can be detected using linear techniques such as principal component analysis. Chapter 7 describes how the super model-based PCA algorithm is applied to the same simulation of the exothermic batch reactor used in the previous chapter to determine if it provides an improvement in terms of fault detection. The case study investigates different types of error model, and again the residuals are assessed to identify how well the technique deals with the characteristics of batch data and to determine which type of error model is the most appropriate.

Incorporating the additional residuals model into the model-based PCA algorithm goes some way towards dealing with the limitations of model-based PCA highlighted in Chapter 6. The objective of Chapter 8 is to look at the main issues that could potentially impact on the performance of super model-based PCA, such as model uncertainty and fault diagnosis, and examines different ways of resolving these potential problems to improve the performance and the robustness of the technique. Issues considered in this chapter include parametric uncertainty in the model, sensitivity to noise and plant-model mismatch. Consideration of these factors theoretically will give a more robust batch monitoring technique.

In summary, the objectives and contributions of the thesis are as follows:

- to examine the multivariate techniques currently used to monitor batch processes and to highlight their advantages and limitations through their application to data from a simulation of an exothermic batch reactor
- to examine the Model-based Principal Component Analysis technique, analyse its fault detection ability in comparison with other batch monitoring techniques, and to highlight limitations of the method
- to further develop the Model-based Principal Component Analysis technique to handle inaccuracies in the mechanistic model (super model-based principal component analysis)
- to investigate the different aspects that affect the Super Model-based Principal Component Analysis technique in an industrial situation

The fulfilment of these objectives has resulted in the following major contribution:

- the development of an advanced model-based technique for the monitoring of batch processes which addresses the issues commonly found in batch process data, such as non-linear and serially correlated behaviour, and additionally that can be made more robust to enable its application in an industrial situation.

2 CHAPTER TWO: BASIC MULTIVARIATE STATISTICAL TECHNIQUES

2.1 Introduction

In Chapter 3, different approaches specific to the monitoring of batch processes are discussed. Most of the techniques introduced are based on a variation of the multivariate statistical techniques of principal component analysis, Jollioffe *et al* (1986) and Wold *et al* (1987), and partial least squares (PLS), Geladi *et al* (1986) and Martin *et al* (1996). Therefore prior to discussing their application in a batch framework, this chapter summarises the theory behind these techniques.

Process data contains valuable information about how different variables interact and the impact that these interactions have on the final product. Successful implementation of monitoring systems will improve safety, reduce environmental and health risks, reduce costs through the early diagnosis of process malfunctions and the location of the problem source, thereby resulting in the reduction in off-specification product, and the avoidance of the need to re-work product. Hence this will give improved quality control and, overall, by adopting this approach a better understanding of the process through an enhanced understanding of the complex relationships between the variables will be attained through a detailed analysis of the process data.

A range of different types of variable can be monitored during a manufacturing process including temperatures, pressures, viscosity, flow rates and, more recently, secondary measurements of variables such as concentration through the use of spectroscopic instrumentation, such as with near infra-red (NIR) and mid infra-red (MIR) analysers. Typically more than one of these variables has an influencing effect on the behaviour of a process which is why univariate statistical analysis techniques, that focus on only one variable, tend to be inappropriate for the monitoring of manufacturing processes. The large number of variables monitored on chemical processes would require multiple control charts to be developed to adequately monitor the process and to keep track of the crucial variables. Furthermore, it is often the interaction between these variables that causes deviations in the process, therefore it is important that a multivariate approach to the analysis is adopted which considers the inter-relationships between the variables, otherwise misleading information may materialise which can result in inappropriate control actions being taken by an operator. This is demonstrated in Figure 1. Patton *et al* (2000), which shows the univariate control charts for two variables. The abnormal

observation does not appear to lie outside the control limits when monitored using the univariate control charts, however when the two variables are plotted on a multivariate control chart, the observation can clearly be to be outside of the normal operating limits. The multivariate control chart is discussed in more detail in section 2.2.

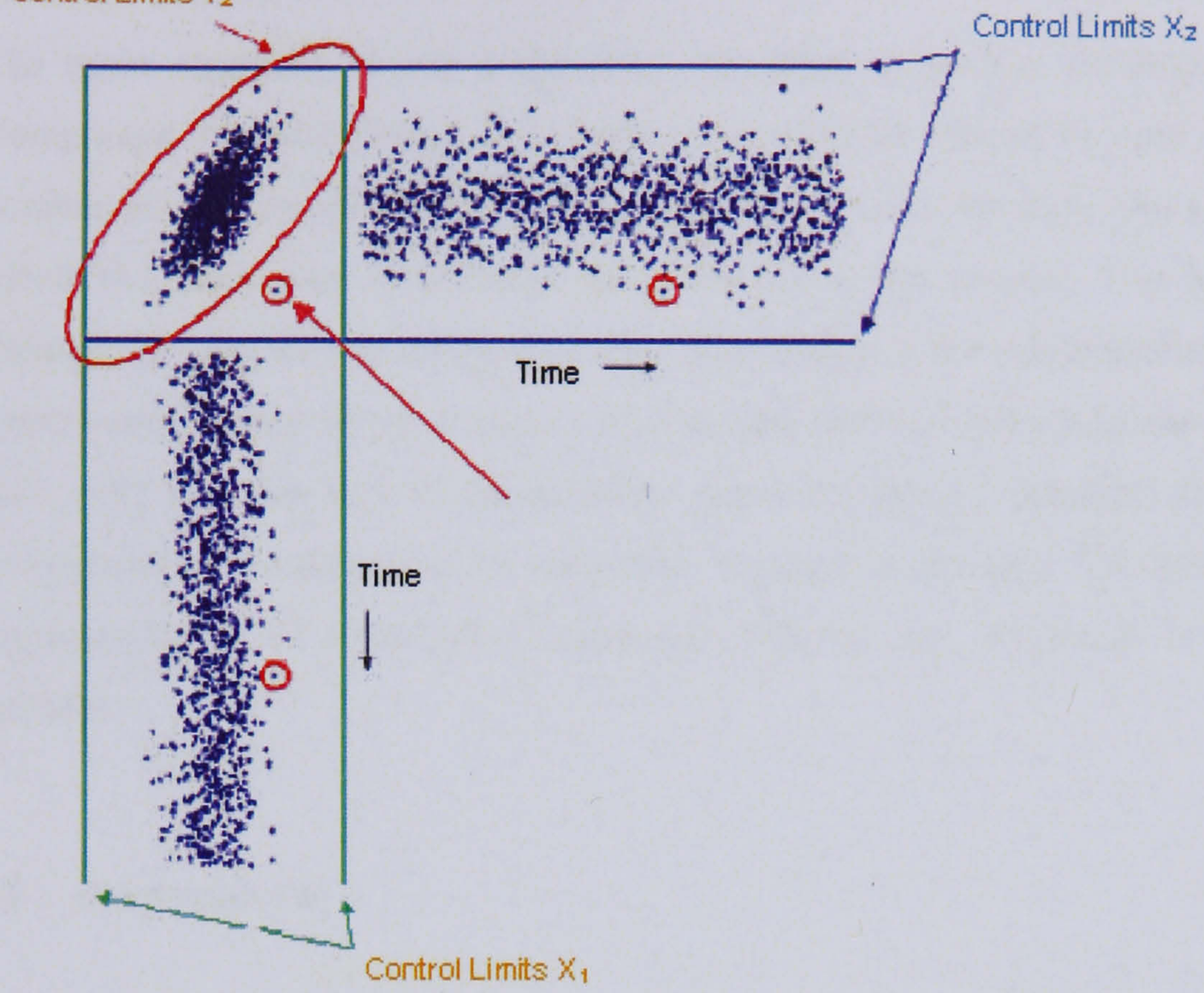


Figure 1 : Univariate control charts

As discussed previously, since process variables are usually highly correlated, their analysis necessitates a multivariate approach which can extract from the data the underlying trends that are determining the behaviour of the process. This is the goal of the basic multivariate statistical projection based techniques, such as principal component analysis (PCA), Jollioffe *et al* (1986) and Wold *et al* (1987), and partial least squares (PLS), Geladi *et al* (1986) and Martin *et al* (1996). The fundamental approach is based on the projection of the dataset onto a new subspace that is defined in terms of the principal components or latent variables that are linear combinations of the original variables and which describe the underlying variation in the process.

With respect to process monitoring, PCA and PLS were originally designed to deal with continuous processes and the steady state and linear relationships that exist between the variables in these processes, Wetherill *et al* (1991). The following sections briefly

describe how these fundamental techniques work. In Chapter 3, the focus on how they have been adapted to address the additional challenges that are associated with batch processes, i.e. non-linear behaviour and the presence of serial correlation.

2.2 Principal Component Analysis

2.2.1 Introduction

The main objective of the multivariate statistical projection technique of Principal Component Analysis (PCA) is to analyse the data of interest in such a way that the number of process variables being monitored can be compressed into a few dominant trends that determine or influence the behaviour of the process. This is accomplished through the application of an algorithm that results in the computation of a series of linear combinations of the original variables that summarise the inherent variation in the data with minimal loss of information, these are termed principal components and correspond to the directions of maximum variance in the data. The basic steps in the implementation of Principal Component Analysis are discussed in the following sections.

2.2.2 Pre-treatment

The first stage in the application of PCA is to identify a data matrix **X**, of order m rows (observations) and n columns (variables) from the data generated during the process. The observations are typically the time points of a continuous or batch process, and the variables are the sensor readings for a number of different measurements recorded on the process such as temperatures, pressures and flows.

The individual variables are then pre-processed to convert the matrix into a form more suitable for analysis. Pre-processing can comprise several different stages, such as outlier removal, transformation of the data set, filtering and scaling. An outlier, an observation that lies an abnormal distance from the rest of the distribution due to human error, measurement error, etc, may need to be eliminated from the data as outliers can distort projection-based analyses and lead to faulty conclusions being drawn about the status of the process. Therefore it is common practice to highlight and replace them with, for example, an interpolated value based on the other measurements. However before removing the outliers it is important to determine the cause and the frequency of the

abnormal value as it may be due to a recurring sensor fault etc and therefore some action requires to be taken.

Another pre-processing technique commonly applied is the transformation of the data, the goal being to normalise and/or linearise the data set. This step should only be considered where there is knowledge about important non-linear relationships between variables. The one issue with transforming process variables is that it can increase the complexity of the interpretation of the results. In a situation where the signal to noise ratio in the data is low, a filter may be applied to the data. The objective is to filter out the noise and convert the data into a smooth trend that is unaffected by noise. There are a raft of filtering algorithms, however the most commonly used are the first order uni-directional filter, first order bi-directional filters and median filters.

For process variables where there are a significant number of measurements missing due to instrument error or a similar issue, then a technique for in-filling the missing data must be applied, such as hold last value, averaging or interpolation. This is discussed in more detail in section 3.3. When considering the missing measurements, it is important to consider whether there is a pattern to the missing data, e.g. the data is always omitted at a certain time point or the instrumentation has reached saturation, or whether it has occurred randomly. If the data is missing randomly then a technique for in-filling can be applied, however if there is an identified pattern to the missing data then consideration should be given to the removal of the variable from the analysis in order to identify and correct the problem with the sensor.

Once the cleaning up of the data has been carried out, the data then needs to be scaled before the implementation of PCA. Scaling is required to be applied to the variables since when analysing multiple variables, it is likely they will have differing orders of magnitude. For example, a temperature may be of the order of 100°C, with the flow rate being 10 m^3/hr and the molecular weight being of the order of 10^3. Since the basis of PCA is to define the main source of variation, those variables with a large variance, and typically of large magnitude, will dominate the initial principal components. One method is to group and model similar variables but this is not ideal as it over-complicates the analysis by having to build several models. Hence it is important that the variables are scaled to even out their weightings. This will ensure that one variable does not inadvertently dominate the analysis, thereby masking the effect of other more critical variables in terms of determining the underlying behaviour of the process. Scaling the data can also normalise the distribution of the sample by adjusting the measurements

according to some transformation function in order to make them comparable with a specific point of reference. There are several methods for scaling the data, one popular technique is that of auto-scaling. In this case the data is mean-centred, that is the average value of each column/variable is calculated and subtracted from the data. Next the data is scaled to unit variance, that is the standard deviation (σ) of each column (variable) is computed and from this, the scaling weight ($1/\sigma$) is calculated, and then each column of data is multiplied by its scaling weight to give unit variance. Mean-centering the data improves the interpretability of the model. If the data is not mean-centred, then the first principal component will basically model the mean of the process variables, with the second principal component describing the direction of maximum variance in the data. The first principal component should be the most informative, modelling the maximum variance without any restrictions. Mean-centering the variables ensures that this happens, Nomikos *et al* (1994). Other forms of scaling include division by the standard deviation, mean centering and basic scaling of the data between 0 and 1 (subtract minimum for each column and divide by the range).

2.2.3 Application of Principal Component Analysis

After pre-treatment and scaling of the data, PCA is applied to the matrix **X**. This involves the eigenvector decomposition of the covariance matrix of **X**:

$$\text{cov}(\mathbf{X}) = \frac{\mathbf{X}^T\mathbf{X}}{m-1} \qquad (2.1)$$

The data matrix **X** is decomposed as the sum of the outer product of vectors $\mathbf{t}_i$ and $\mathbf{p}_i$:

$$\mathbf{X} = \mathbf{t}_1\mathbf{p}^T_1 + \mathbf{t}_2\mathbf{p}^T_2 + \ldots + \mathbf{t}_r\mathbf{p}^T_r + \ldots + \mathbf{t}_R\mathbf{p}^T_R \qquad (2.2)$$

where the eigenvector or loading vector, $\mathbf{p}_{i,}$ contains information on how the variables relate to each other and the score vector, $\mathbf{t}_i$, represents the projection of each observation onto the principal component and **R** defines the maximum number of principal components, and **r** the retained number of components, equal to the min{m,n}. The vector $\mathbf{p}_i$ allows the user to identify those variables that contribute to defining the main source of variability. For each $\mathbf{p}_i$:

$$\text{cov}(\mathbf{X})\mathbf{p}_i = \lambda_i\mathbf{p}_i \qquad (2.3)$$

where λ_i is the eigenvalue for eigenvector $\mathbf{p}_i$, and is a measure of the variance described by each $\mathbf{t}_i$, $\mathbf{p}_i$ pair. Eigen analysis is used to sort the eigenvectors and eigenvalues in terms of their contribution to explaining the underlying variance. The eigenvector corresponding to the largest eigenvalue has the same direction as the first principal component - it is a linear combination of the variables and describes the main direction of variability in the matrix **X**. i.e. the score vector $\mathbf{t}_i$ is a linear combination of the original n variables, with the coefficient defined by $\mathbf{p}_i$. The second principal component is orthogonal to the first principal component, and is a linear combination of the variables and explains the next greatest amount of variability in the data, and corresponds to the second largest eigenvalue. Generally the data can be explained with fewer latent variables than original variables without significant loss of information.

In general the PCA model is truncated after r components, with the remaining variance factors consolidated into a residual matrix (**E**):

$$\mathbf{X} = \mathbf{t}_1\mathbf{p}^T_1 + \mathbf{t}_2\mathbf{p}^T_2 + \ldots + \mathbf{t}_r\mathbf{p}^T_r + \mathbf{E} \qquad (2.4)$$

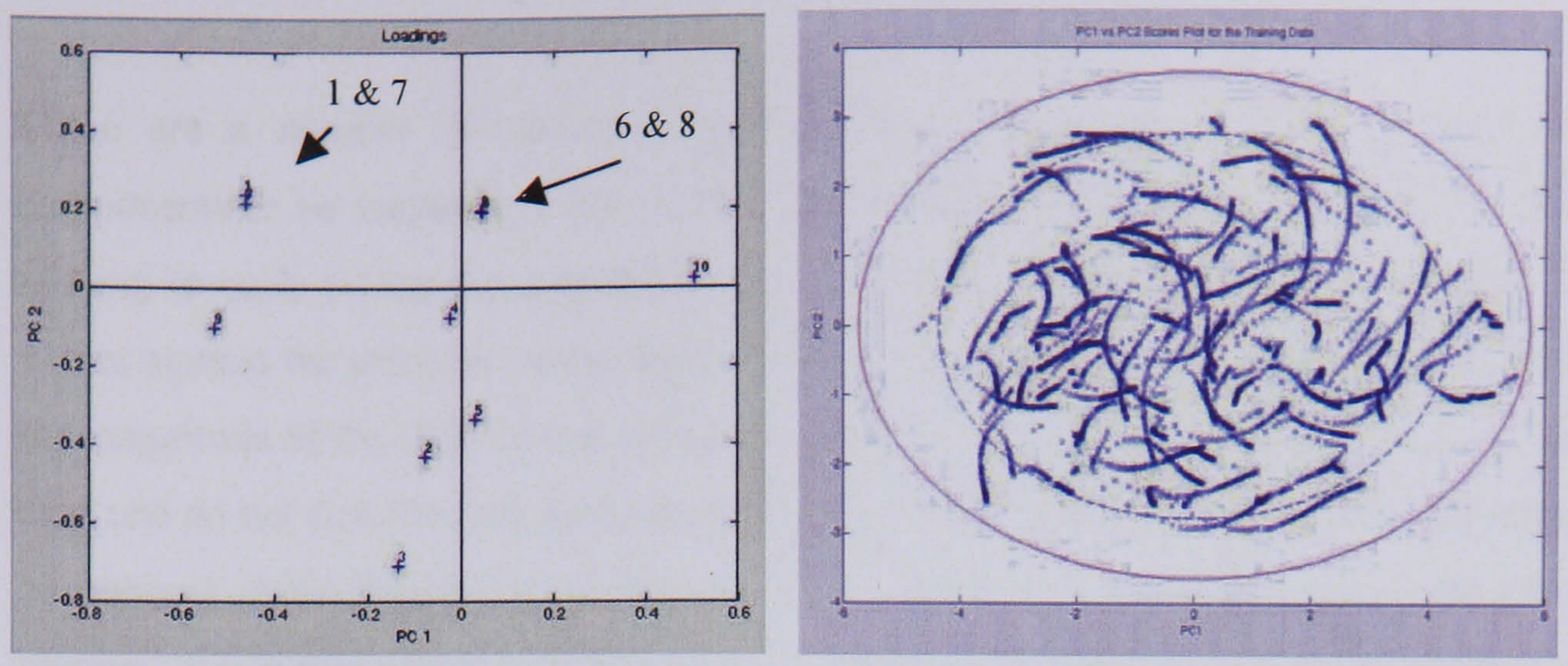
In equation 2.4, r must be less than or equal to the smallest dimension of **X**:

$$r \leq \min\{m,n\} \qquad (2.5)$$

The selection of r is discussed in section 2.2.5.

2.2.4 Loadings and Scores Plots

A bivariate plot of $\mathbf{p}_i$ vs $\mathbf{p}_{i1}$ (where $i \neq i_1$), a loadings plot, can be used to identify the inherent relationships between the variables, how they are correlated and how influential they are with respect to the representation. Variables which appear in the same quadrant of the plot are positively correlated whilst variables positioned in diametrically opposite quadrants are negatively correlated. Additionally, the distance between the variable and the origin of the plot indicates the strength of impact that it has on the model – the closer to the origin, the smaller the effect of the variable on the principal component. This can be seen graphically for 10 variables for a simulation of a continuous distillation process, Love (2007), in Figure 2.



(a) Bivariate loadings plot (b) Bivariate scores plot

Figure 2 : Bivariate loadings and scores plots

In Figure 2 (a), it can be observed that variable 4 is close to the origin and therefore has the least impact on both principal component one and two. Variables 6 and 8 are positively correlated, as they are closely situated within the same quadrant, as are variables 1 and 7. It can also be observed that variables 6 and 8 are negatively correlated with variables 2 and 3, which are situated in the opposite quadrant.

As stated, the principal component scores are the co-ordinates of the original observations as projected onto the new principal component plane, and the corresponding scores plot is shown in Figure 2 (b). The scores summarise the information contained in the original variables and allow the structure of the dataset to be visualised. It can be seen here that all the data is lying in one cluster, i.e. one mode of operation, however different 'paths' within the data can be seen within the cluster of data due to different stages within the process. The scores and loadings plots need to be studied simultaneously to understand process behaviour.

2.2.5 Selecting Number of Principal Components to be Retained

Once all R principal components have been calculated, the number of principal components retained for the definition of the process monitoring representation is required to be determined, r. As specified, the first principal component represents the direction of maximum variation in the process, and the second principal component represents the direction of next greatest variation etc, with the lower order components typically representing noise in the process as opposed to important underlying influential behaviour. Therefore it is important to carefully select the number of components included in the model to ensure that the information content is optimised.

There are a number of different methods for selecting the number of principal components to be included in the model. One technique is to consider the eigenvalues relating to each principal component, either by studying the values or by plotting the values against the principal component number. A typical rule of thumb when studying the magnitude of the eigenvalues is that principal components with an eigenvalue of less than one do not describe any systematic variation in the system and therefore should not be retained in the model. By plotting the eigenvalues against the principal component number, any sudden drops can be identified. This gives an indication of the number of components that should be retained (i.e. the number of components before the sudden drop). This can be seen in Figure 3 (Wise *et al* (2006)), where three principal components would be considered sufficient to represent the variation in the data set.

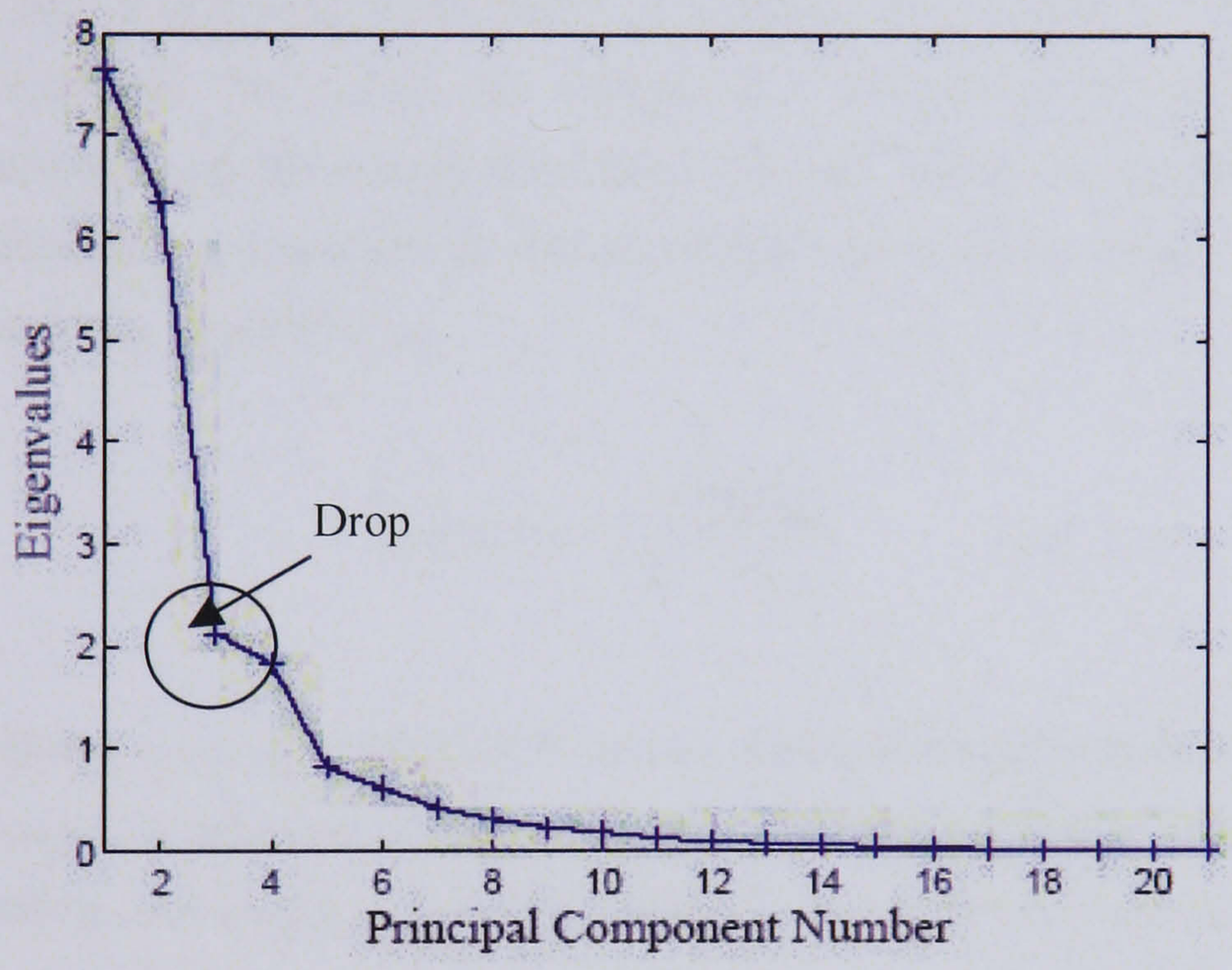


Figure 3 : Plot of eigenvalue versus principal component number

Cross-validation is another technique used to determine the optimum number of components to be retained in the model and can be used in combination with the aforementioned methodology. The basis of the technique involves the partitioning of the data matrix into several subsets. The number of subsets can range from two subsets up to the order of m, the number of samples in the data and is typically determined by the size of the data set and the associated computational burden. Principal component models are built on the data set with one subset omitted each time. The omitted subset is then projected onto the PC model and the values inferred. The predicted error sum of squares (PRESS) is computed for a different number of principal components retained in the

PCA process representation i.e. i,…min(m,n). This procedure is then repeated with each subset being omitted The total (PRESS) is then computed as the sum of all the individual PRESS's:

$$\text{PRESS}(r) = \frac{\sum_{n=1}^{N}\sum_{m=1}^{M} e_{nm,r}^{2}}{N \cdot M} \qquad r=1,\ldots.R \qquad (2.6)$$

where $e_{nm,r}$ is the mismatch between the n^{th} observation of the m^{th} variable given by the model based on the use of R principal components, is given by

$$e_{nm,r} = x_{nm} - \hat{x}_{nm,r} \qquad r <= 1,\ldots\min(m,n) \qquad (2.7)$$

N and M are the total number of observations and number of variables respectively in the original data matrix. The number of components to be included in the model is determined by the minimum value of the total PRESS in equation 2.6. An alternative metric used in combination with the PRESS value is the root mean squared error of cross validation (RMSECV):

$$\text{RMSECV}_r = \sqrt{\frac{PRESS_r}{n}} \qquad r=1,\ldots.R \qquad (2.8)$$

Plotting a curve of RMSECV versus principal component number will indicate the number of principal components that gives the minimum error, as for the PRESS statistic. It is useful to show the RMSECV plot in combination with the PRESS statistic to determine the number of components to be retained. An example plot of the RMSECV is given in Figure 4, Wise *et al* (2006).

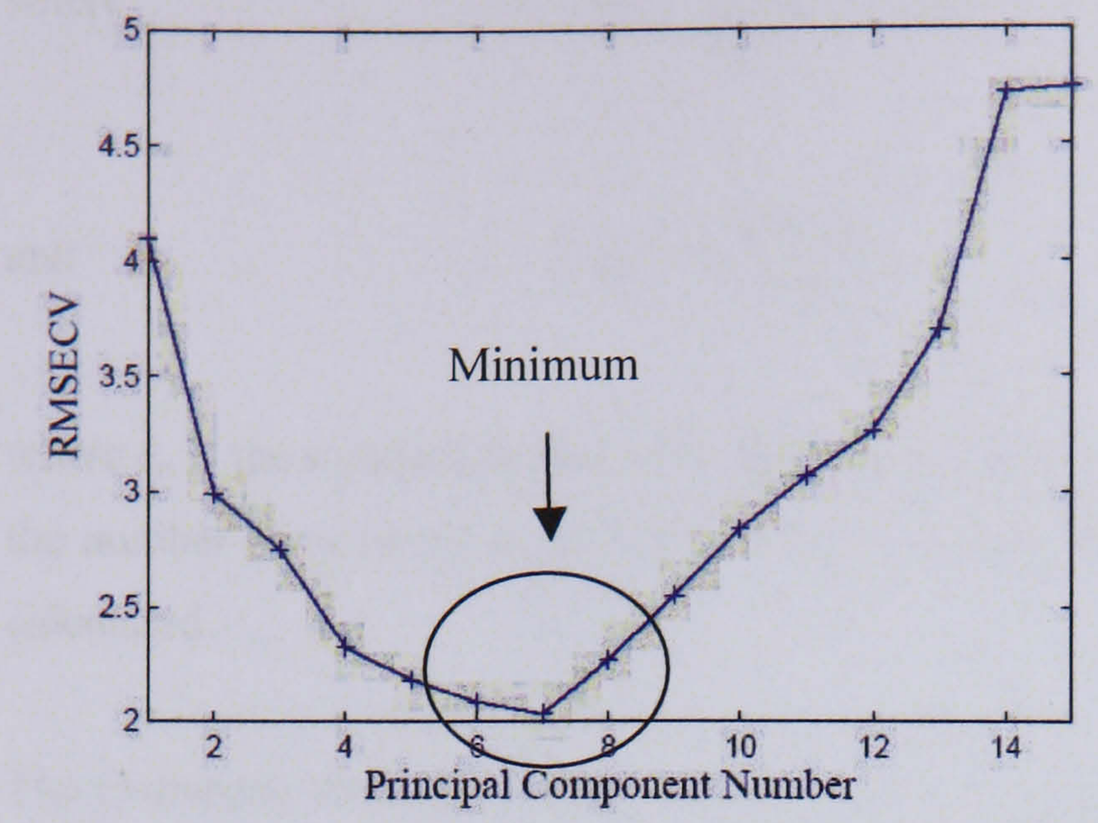


Figure 4 : Example root mean squared error of cross validation curve

2.2.6 Monitoring Statistics

Once the number of principal components required to adequately represent the data has been determined, the principal component model can be used to monitor subsequent process data. For this purpose, metrics such as the Q statistic and Hotelling's T^2 are employed. The Q statistic (also called the squared prediction error (SPE)) measures the goodness of fit of the process representation by describing the variation in the data not included in the monitoring representation:

$$Q_i = \mathbf{e}_i\mathbf{e}_i^T = \mathbf{x}_i(\mathbf{I}-\mathbf{P}_r\mathbf{P}_r^T)\mathbf{x}_i^T \qquad (2.9)$$

where $\mathbf{e}_i$ is the i^{th} row of the matrix $\mathbf{E}$, $\mathbf{P}_r$ is the matrix of r loading vectors and $\mathbf{I}_{nxn}$ is an identity matrix of appropriate size (n x n). The Q statistic is the difference between the sample and its projection onto the model, i.e. it indicates how well the sample is described by the process representation. It explains the random variation in the data that is not described by the model, for example the normal variation due to random noise. Confidence limits can be calculated for this statistic based on its random variation. The limit defines the distance from the plane that is considered normal based on the data included to develop the principal component model. When the Q statistic exceeds a predetermined limit, it indicates that there has been a significant change in the random:

$$Q_\alpha = \Theta_1\left[\frac{c_\alpha\sqrt{2\Theta_2 h_0^2}}{\Theta_1} + 1 + \frac{\Theta_2 h_0(h_0 - 1)}{\Theta_1^2}\right]^{1/h_0} \qquad (2.10)$$

where $$\Theta_i = \sum_{j=r+1}^{R} \lambda_j i \quad \text{for i = 1,2,3, ...} \qquad (2.11)$$

and $$h_0 = 1 - \frac{2\Theta_1\Theta_3}{3\Theta_2^2} \qquad (2.12)$$

where c_α is the standard normal deviate corresponding to the upper $(1 - \alpha)$ percentile, r is the number of principal components retained and R is the total number of components calculated.

The Q-statistic describes the variation outside of the model, whereas the Hotelling's T^2 statistic describes the variation of each sample within the model, i.e. it monitors variation within the score space:

$$T_i^2 = \mathbf{t}_i\lambda^{-1}\mathbf{t}_i^T = \mathbf{x}_i\mathbf{P}_r\lambda^{-1}\mathbf{P}_r^T\mathbf{x}_i^T \qquad (2.13)$$

where $\mathbf{t}_i$ is the i^{th} row of $\mathbf{T}_k$, the m by r scores matrix, $\mathbf{P}_r$ is the loading matrix and λ^{-1} is the diagonal matrix containing the inverse of the eigenvalues associated with the principal components retained in the model. Again confidence limits can be calculated:

$$T_{r,m,\alpha}^2 = \frac{r(m-1)}{m-r} F_{r,m-r,\alpha} \qquad (2.14)$$

where F represents the F distribution, r is the number of principal components retained, m is the number of samples involved in the construction of the model and α is the significance level of the confidence limits, e.g. for 95% confidence limits, α=0.05. The T^2 statistic describes the systematic variation in the process, as opposed to the random variation described by the Q statistic, therefore a deviation from the confidence limits in this case indicates a significant change in the systematic variation in the process.

2.2.7 Contribution Plots

It is important to be able to diagnose, or at least determine the variables responsible for the cause of any deviations from the T^2 confidence limits. Contribution plots are used to achieve this by calculating the contribution of each variable to the individual scores of the PCA model for those scores lying outwith the confidence limits:

$$\text{Contribution}_{ij} = \frac{t_i}{\sigma_i^2} p_{ij}(x_j - \mu_j) \qquad (2.15)$$

$$\text{Total Contribution} = \sum_{i=1}^{r} \text{Contribution}_{ij} \qquad (2.16)$$

where t_i is the out of control score, $p_{i,j}$ is the loading for the ith row and jth column of the matrix, σ is the standard deviation of row i, μ is the mean of column j and x_j is the out of control observation. The variables with the highest contribution values indicate those that could be the cause of the process abnormality.

Contribution plots can also be calculated for the SPE statistic. In this case, the contribution of each variable to a significant change in the SPE is the individual prediction error or residual:

$$\text{prediction error} = x_i - \hat{x}_i \qquad (2.17)$$

where the prediction $\hat{x}$ of the process variable is given by:

$$\hat{x}_i = x_i(\mathbf{I} - \mathbf{P}\mathbf{P}^{\mathrm{T}}) \qquad (2.18)$$

2.2.8 Example of PCA Monitoring Tools using a Continuous Distillation Column

This section illustrates the concepts and metrics described in the previous sections using a data from a simulation of a continuous distillation column distilling a binary mixture, Love (2007). There are 10 process variables in the analysis, including temperatures, feed flowrate, steam flowrate, reflux flowrate, bottoms take off rate and heads take off rate. The simulation length was 500 minutes, with variable measurements recorded every 0.1 minutes.

An example bivariate scores plot for a principal component model of a continuous distillation column simulation using three principal components is shown in Figure 5. Note the 95% and 99% confidence interval ellipses shown as dotted and solid lines. If data falls outside these limits during training then they may be outliers, however it is expected that 250 of the 5000 data points could lie outside the 95% limits, and 50 points

outside the 99% limits during normal operation. In Figure 5 the data remains within the limits, and the different trajectories of data within the ellipse represent the different step tests that were simulated during the process. Also shown are plots of the squared prediction error (SPE or Q statistic) and the T^2 statistic (Figure 6). These plots show that large peaks do not dominate the data and the statistical limits are not systematically breached, showing the data has been generated under normal operating conditions and is suitable for use for the development of a nominal model.

To apply the process representation to new data, $\mathbf{X}_{new}$, equation 2.19 is used, with the data having been first scaled and mean centred using the values determined for the normal (training) data:

$$\mathbf{T}_{new} = \mathbf{X}_{new}\mathbf{P} \qquad (2.19)$$

where $\mathbf{T}_{new}$ and $\mathbf{P}$ are the scores and loadings matrices respectively. The new scores and residuals are then calculated and plotted. If the new data breaches the confidence limits then there is likely to be a process problem. The contribution plots can indicate the variables responsible for the process problem.

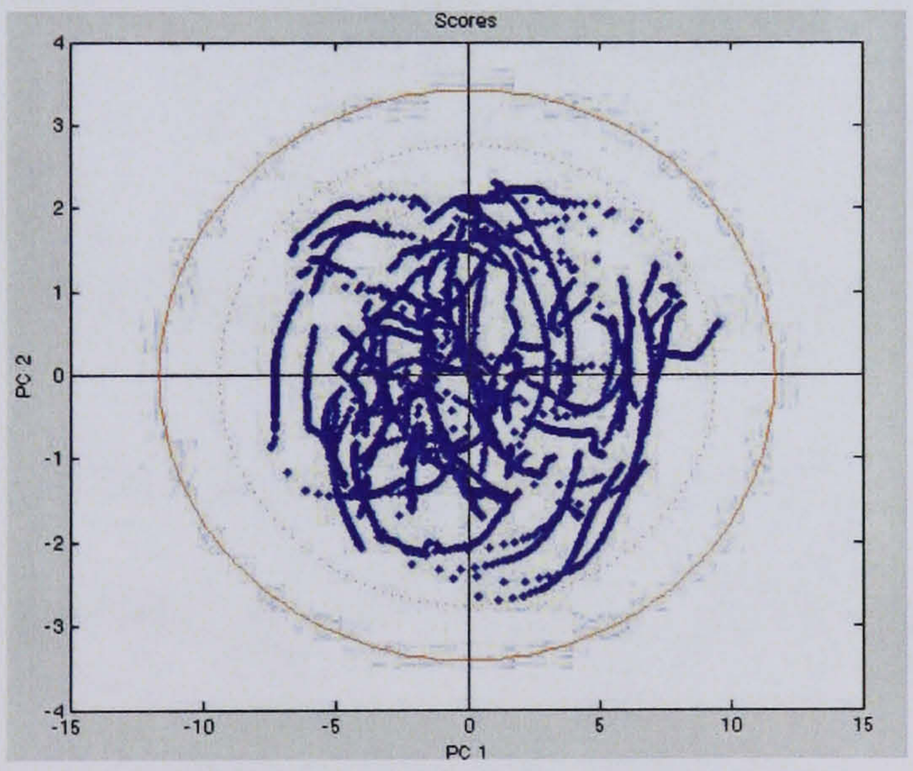
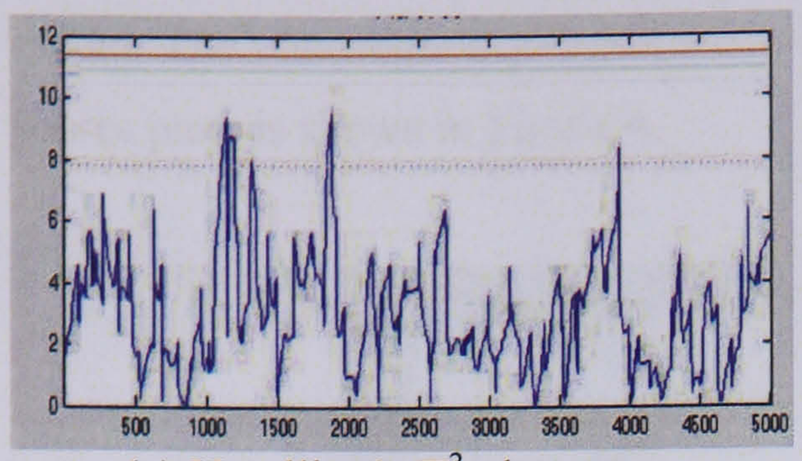
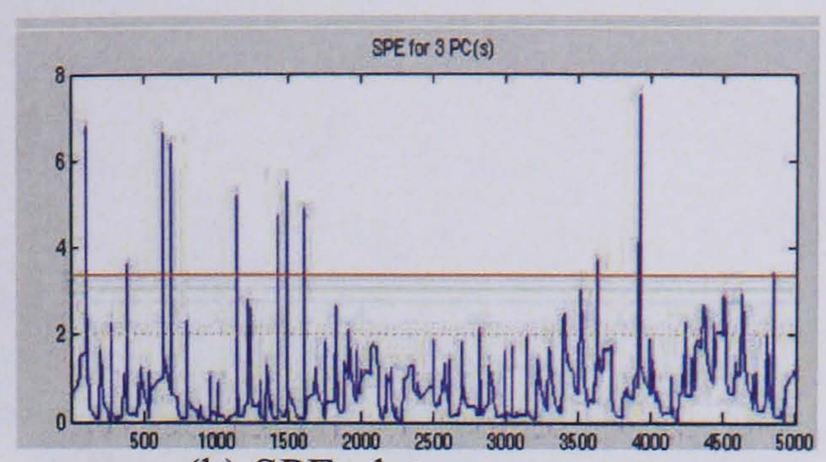


Figure 5: Bivariate scores plot of distillation data



(a) Hotelling's T^2 plot



(b) SPE plot

Figure 6: Hotelling's T^2 and SPE plot for nominal data

An example process fault, a sensor drift (introduced at data point 5500) is shown in Figure 7 and Figure 8. The experimental (i.e. the new data) is shown in green.

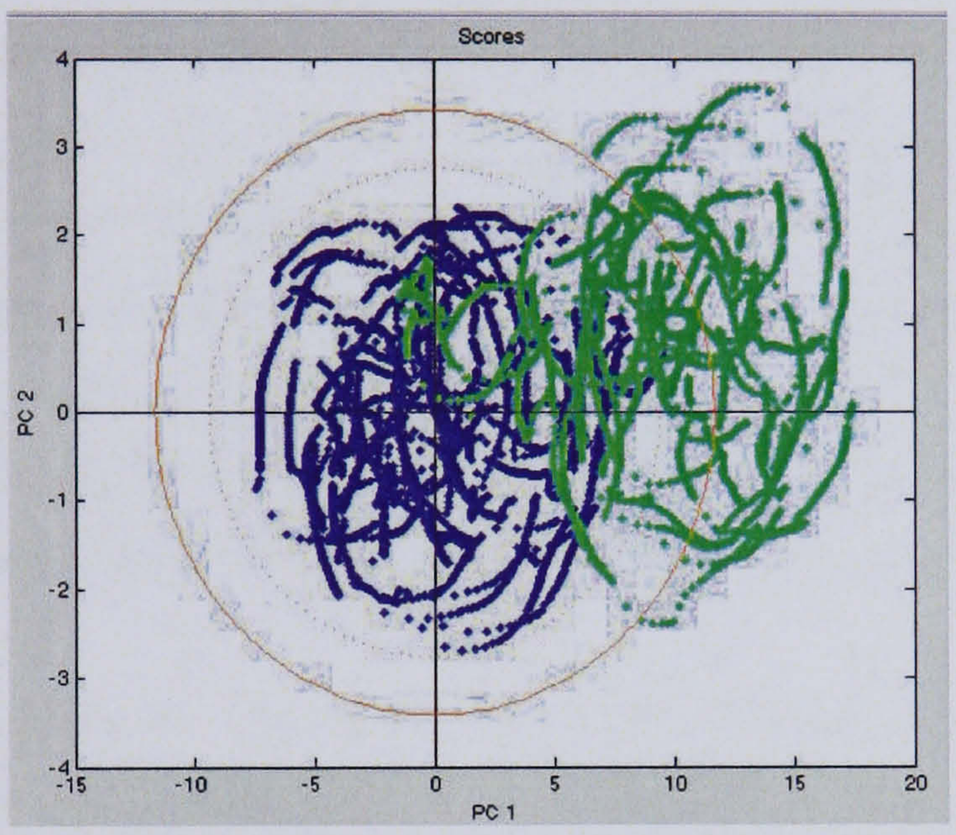
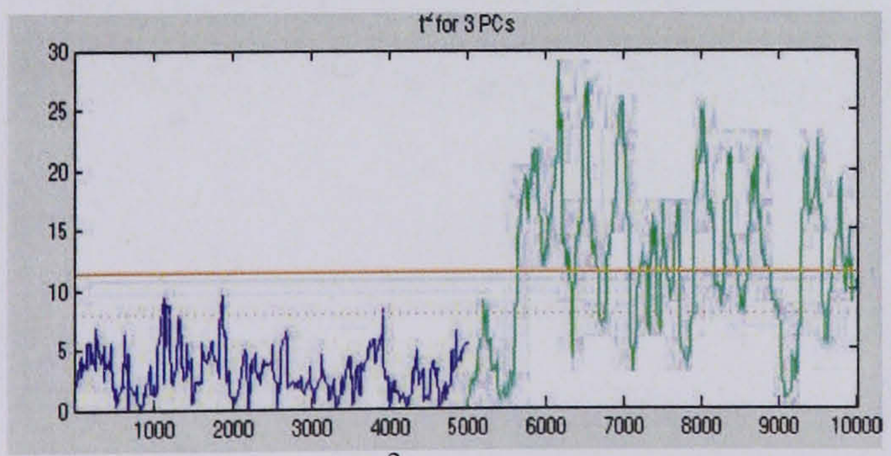
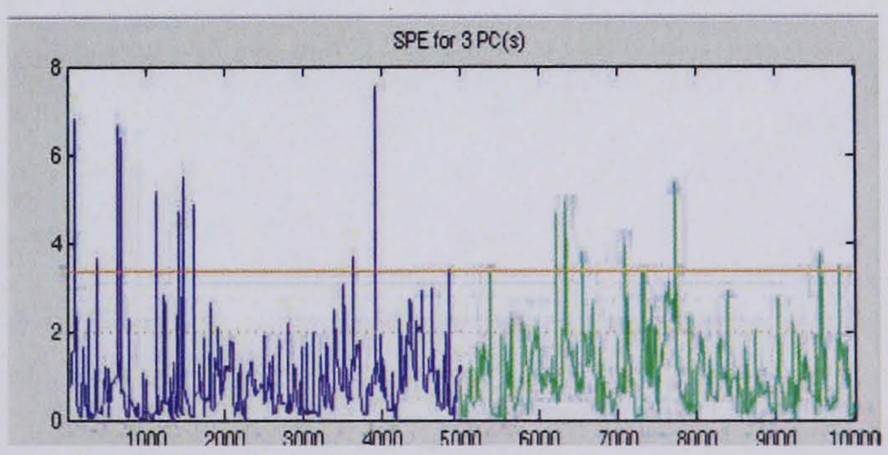


Figure 7 : Bivariate scores plot of new data (green)



(a) Hotelling's T^2 plot of new data



(b) SPE plot of new data

Figure 8: Hotelling's T^2 and SPE plots of new data (green)

It is clear from the scores plot and Hotelling's T^2 plot that a problem has occurred, the limit breach occurs 12 minutes after the fault has been introduced. The SPE plot shows no real evidence of a problem occurring, indicating that the fault is in the plane of the

model. The contribution plot for a typical point outside of the confidence limits on the scores plots is shown in Figure 9.

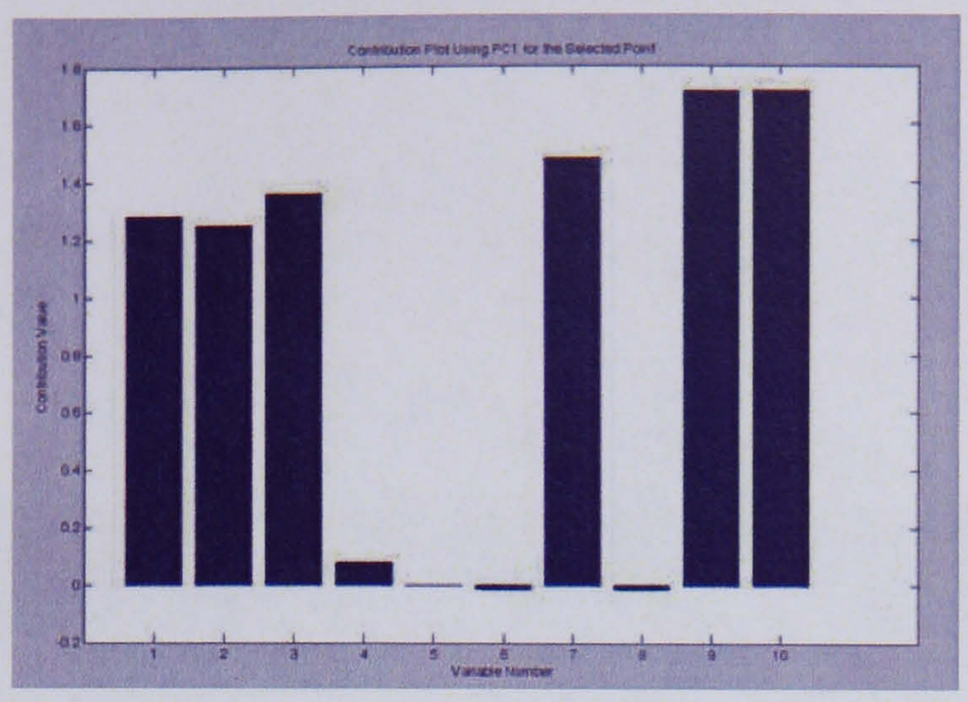


Figure 9: Contribution plot for sensor fault data point

The contribution plot shows that variables 1,2,3,7,9 and 10 show the greatest contribution. These are the top tray composition controller output, middle tray composition, base composition, reflux flow rate, bottoms take off flow rate and heads take off flow rate respectively. The fault introduced was in the composition measurement sensor, which would be expected to have a direct impact on variable 1, the top tray composition controller. However, as the fault has impacted on this variables it has also affected the flow of reflux and therefore also the top and bottom product flows.

The plots examined so far are for the off-line analysis of data and the retrospective analysis of process performance. However another valuable application of PCA is for on-line monitoring. In this case, the new data is analysed directly by the model (i.e. auto scaling and the application of the PCA process representation). An example of an on-line plot is shown in Figure 10, Wise *et al* (2006). Note the lighter shaded points which are old values which gradually fade to give a clearer impression of process movement.

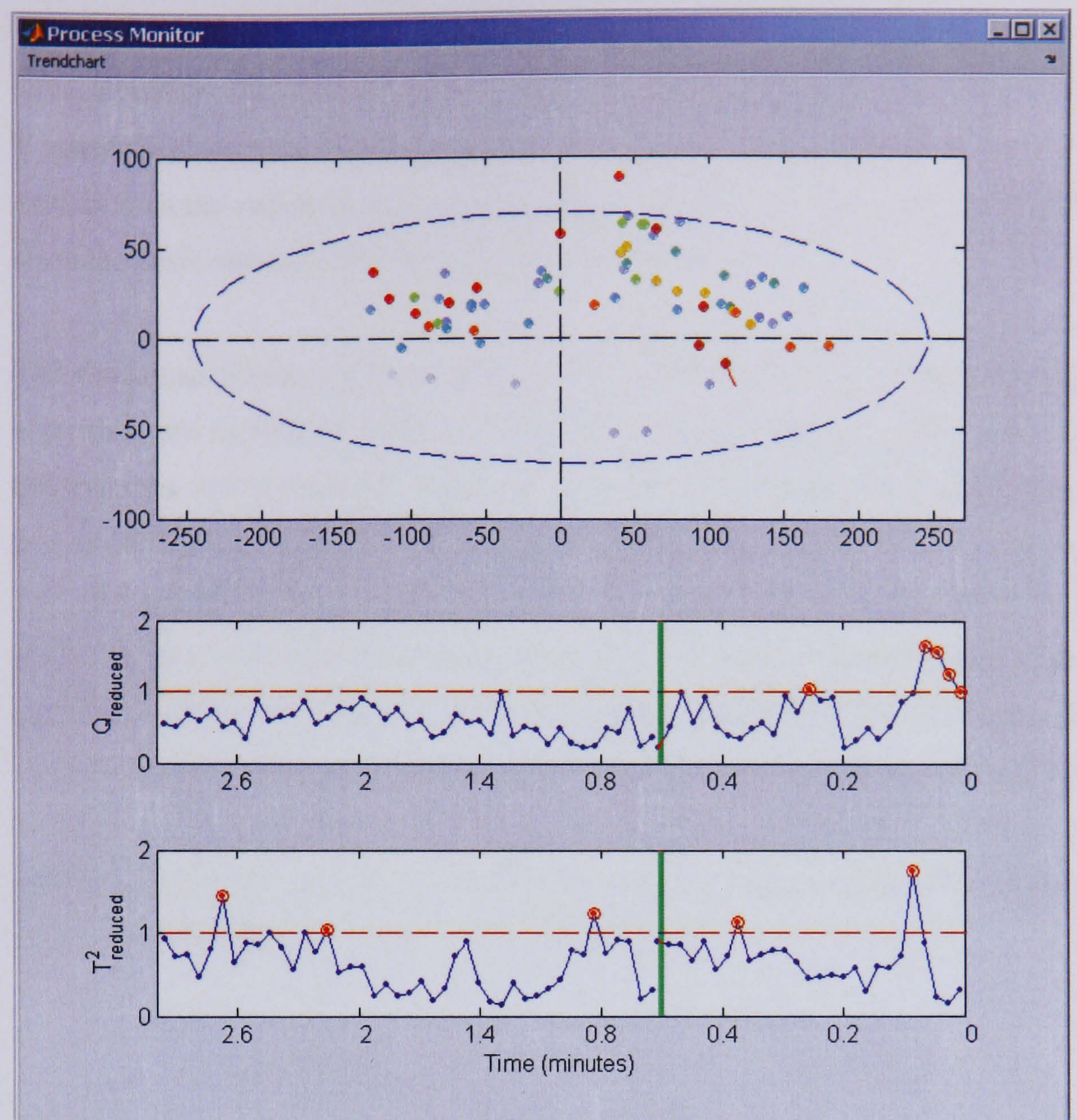


Figure 10 : On-line principal component analysis monitoring

2.3 Partial Least Squares

Partial Least Squares or Projection to Latent Structures (PLS), Wold *et al* (1984), can be used to study the relationship between the process data (**X**) and the product quality data (**Y**). PLS is a regression technique and is an extension of PCA. The key difference between PCA and PLS is that PCA is a maximum variance projection technique whilst PLS defines the maximum covariance between **X** and **Y**.

There have been a number of different applications of PLS, Geladi *et al* (1986), Lorber *et al* (1987), Martens *et al* (1989) and Höskuldsson (1996), including the development of multivariate calibration models for spectral data and the determination of the relationships between process variables at a certain point in time and a corresponding quality value such as purity or reactant conversion.

The basic operation of the PLS algorithm is that it simultaneously reduces the dimensionality of the two data sets **X** and **Y** to determine the latent vectors for the **X** and **Y** spaces that are most highly correlated. The latent vectors are computed such that they explain both the variation in **X** as well as the variation in **X** that is most predictive of **Y**, since the basic objective of PLS is to predict **Y** from **X**.

The fundamental steps in PLS, using the NIPALS (non-iterative partial least squares) algorithm, are as follows, de Jong *et al* (1993). Consider a matrix **X**, comprising of m observations and n variables and a matrix **Y** of quality variables (it is possible that **Y** may contain only one quality variable). The NIPALS algorithm determines the scores, **T**, and loadings matrix **P** and an additional set of vector weights, **W**, for the **X** data block. With multiple **Y** variables, the methodology also calculates scores, **U** and loadings, **Q**, for the quality **Y** data block. Inner coefficients, **b** (which relate the **X** and **Y** block scores), are also calculated. With NIPALS all the scores, weights, loadings and inner coefficients are calculated sequentially. First one column of **Y** is selected, $\mathbf{y}_j$, as a starting estimate for $\mathbf{u}_1$. This is normally the column of greatest variation. Commencing with the **X** data block:

$$\mathbf{w}_1 = \frac{\mathbf{X}^T\mathbf{u}_1}{\|\mathbf{X}^T\mathbf{u}_1\|} \tag{2.20}$$

where $\mathbf{w}_1$ is the weight vector, $\mathbf{u}_1$ is the quality scores vector, note that the 2-norm, $\| \|$, is defined as $\|\mathbf{a}\| = \sqrt{\mathbf{a}^T\mathbf{a}}$.

$$\mathbf{t}_1 = \mathbf{X}\mathbf{w}_1 \tag{2.21}$$

where $\mathbf{t}_1$ represents the data scores vector. In the **Y** data block:

$$\mathbf{q}_1 = \frac{\mathbf{Y}^T\mathbf{t}_1}{\|\mathbf{Y}^T\mathbf{t}_1\|} \tag{2.22}$$

$$\mathbf{u}_1 = \mathbf{Y}\mathbf{q}_1 \tag{2.23}$$

where $\mathbf{q}_1$ is the quality loadings vector. Convergence is checked by monitoring t_1 in equation 2.21 and comparing it with the value in the previous iteration. When the values converge, the algorithm proceeds to determine the **X** data block loadings with rescaling of the scores and weights:

$$\mathbf{p}_1 = \frac{\mathbf{X}^T\mathbf{t}_1}{\|\mathbf{t}_1{}^T\mathbf{t}_1\|} \tag{2.24}$$

$$\mathbf{p}_{1new} = \frac{\mathbf{p}_{1old}}{\|\mathbf{p}_{1old}\|} \tag{2.25}$$

$$\mathbf{t}_{1new} = \mathbf{t}_{1old}\ \|\mathbf{p}_{1old}\| \tag{2.26}$$

$$\mathbf{w}_{1new} = \mathbf{w}_{1old}\ \|\mathbf{p}_{1old}\| \tag{2.27}$$

The regression coefficient is then calculated for the inner relationship:

$$b_1 = \frac{\mathbf{u}_1{}^T\mathbf{t}_1}{\mathbf{t}_1{}^T\mathbf{t}_1} \tag{2.28}$$

Once the scores and loadings have been calculated for the first factor (or latent variable), the **X** and **Y** matrix residuals are determined as follows:

$$\mathbf{E}_1 = \mathbf{X} - \mathbf{t}_1\mathbf{p}_1{}^T \tag{2.29}$$

$$\mathbf{F}_1 = \mathbf{Y} - b_1\mathbf{t}_1\mathbf{q}_1{}^T \tag{2.30}$$

These steps are then repeated for the remaining latent variables but with **X** and **Y** replaced by the residuals, $\mathbf{E}_1$ and $\mathbf{F}_1$ with the subscripts incremented by one. PLS forms the matrix pseudo inverse:

$$\mathbf{X}^+ = \mathbf{W}(\mathbf{P}^T\mathbf{W})^{-1}(\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T \tag{2.31}$$

The scores for the new data, $\mathbf{X}_{new}$, are calculated as follows:

$$\mathbf{T}_{new} = \mathbf{X}_{new}\mathbf{W}(\mathbf{P}^T\mathbf{W})^{-1} \tag{2.32}$$

Finally predictions of the **Y** values for the new data are determined using:

$$\mathbf{Xb} = \mathbf{Y} \tag{2.33}$$

The predictions can then be compared to the actual values for process monitoring purposes. For example, the **Y** values could represent the end of batch quality of a process which are normally measured in the laboratory after manufacture is complete. By predicting these values through PLS, the quality of the batch can be monitored as it progresses, rather than waiting till the batch is complete, and appropriate action can be taken if the predicted quality is not as desired.

Note that, in the same manner as for PCA, the most useful information is contained in the first few latent variables. As with PCA, cross-validation is a technique that is commonly applied to determine the appropriate number of latent variables to retain in the PLS model to represent the process through the use of the PRESS statistic. An alternative technique is the information criterion. The most well known is Akaike Information Criterion (AIC, 1976):

$$AIC = 2r-2\ln L \qquad (2.34)$$

where r is the number of latent variables retained and L is the likelihood function. AIC takes into account the statistical goodness of fit of the model as well as the number of parameters that are required to be estimated to achieve the degree of fit by imposing a penalty as the number of latent variables increases. This is an important feature since the inclusion of too many latent variables will cause the magnification of noise and lead to poor process monitoring.

Selection of a suitable model allows the performance of subsequent batches to be monitored. Metrics such as Hotelling's T^2 and the Q statistic (or SPE) which were described in section 2.2 are also used for monitoring data in PLS control charts.

2.4 Conclusion

Principal component analysis and partial least squares have been shown to be valuable tools for the monitoring of continuous data. In Chapter 3, the adaptation of these techniques for application to a batch process is discussed.

3 CHAPTER THREE: REVIEW OF STANDARD BATCH MONITORING TECHNIQUES

3.1 Introduction

Many industries have seen a significant increase in batch production over the past two decades, particularly in the areas of pharmaceuticals, polymers, speciality chemicals and biochemicals. This growth has led to a need for process monitoring tools to be developed that specifically address the unique features of a batch process, including their finite duration, short product life cycle and dynamic and non-linear behaviour, Alexander *et al* (2002). This chapter reviews a number of techniques that are currently available for the analysis of data from batch processes, thereby enabling the subsequent development of a monitoring representation. In Chapter 6, the most commonly applied of these techniques are compared in terms of their ability to identify the onset of changes in operational behaviour and in Chapter 7, the techniques are developed further into a new monitoring methodology

Before the introduction of multivariate statistical analysis for the monitoring of batch processes, the most commonly recognised method for ensuring consistent, high quality batch production was the recipe-driven approach, Love (2007). This technique works by recording the recipe and the key variables necessary to produce a successful batch run and then this information is used as a template for future batch runs to reproduce the same quality. However there are a number disadvantages associated with this method, including the fact that large amounts of data generated during the batch run are not utilised and hence the information inherent within the data is not considered. Furthermore, utilising the recipe-driven approach does not guarantee the replication of the same product quality between batches. The recipe information, including the operating parameter set points, the material quantities and the sequence of operation is set at the beginning of a batch run. This often means that even if there are differences in the input conditions or materials, the same batch recipe is used each time, resulting in products of varying quality. Additionally, there is typically little or no interaction from the operator during the batch run unless an alarm condition occurs, therefore more subtle interactions between variables which may lead to an out of specification product are not identified. Furthermore, since the quality parameters are often not tested until the batch is complete, it is too late to take any corrective action if some of the parameters start to

deviate from the prescribed trajectory. Consequently the entire batch may require to be either re-worked or disposed of, which is a costly way of manufacturing a product.

One alternative to the recipe driven approach is to develop a mechanistic model of the process and use this to monitor batch performance. The advantage of this technique is that is uses real physical knowledge of the process to monitor its behaviour, making it easier for the user to comprehend than a black box model. However a significant amount of effort and computational time is required to be expended to ensure a comprehensive understanding is attained, hence enabling the modelling of the complex relationships that occur during a batch process. In practice some of the kinetic phenomena may never be fully understood, leading to reduced complexity mechanistic models, or at worst an inaccurate representation of the process, hence impacting on the subsequent effectiveness of the scheme.

To continue manufacturing in an increasingly aggressive global market, and ensuring retention of market share, whilst realising that the progressively more rigorous regulations are met (whether they be associated with operational safety and environmental protection for the polymer industries or strict drug regulations in the pharmaceutical industries), it is essential that an efficient and effective batch monitoring system that uses the significant amount of process data available is implemented online and is operated in real time.

3.2 Batch Data Matrix - Unfolding

In process performance monitoring, the techniques of PCA and PLS that were briefly introduced in Chapter 2 were originally applied to data from continuous processes. This thesis focuses on the monitoring of batch processes and to apply these multivariate techniques to data from batch processes, further consideration is required to be given to the data matrix, Nomikos *et al* (1995).

The main difference between data from a batch process compared with a continuous system is that the data structure is three-dimensional as opposed to two-dimensional, the added dimension being that of batch identifier. In a continuous process it is the relationships between the variables that are of consideration, whereas in a batch process not only is it important as to how the variables are correlated but also how they are

correlated over time. This results in a 3-dimensional batch data array $\underline{\mathbf{X}}(I \times J \times K)$, where I represents the number of batches, K is the number of time points and J is the number of variables. Figure 11 shows the batch data matrix. The data is decomposed into a two dimensional array and PCA is applied. There are six different possible methods for unfolding the data array, however only three are mathematically different, and only two of these are meaningful for the monitoring of batch processes. Figure 11 illustrates these two methods. Unfolding method one (time-wise unfolding) analyses the variability among batches in $\underline{\mathbf{X}}$ by summarising the information in the data with respect to the variables and their time variation, an (I x JK) matrix is formed, it is also possible to unfold method 1 in the format shown in Figure 12, this will not impact on the outcome of the MPCA technique. Method two summarises the data with respect to the variables to form a matrix of order (IK x J), often termed batch-wise unfolding.

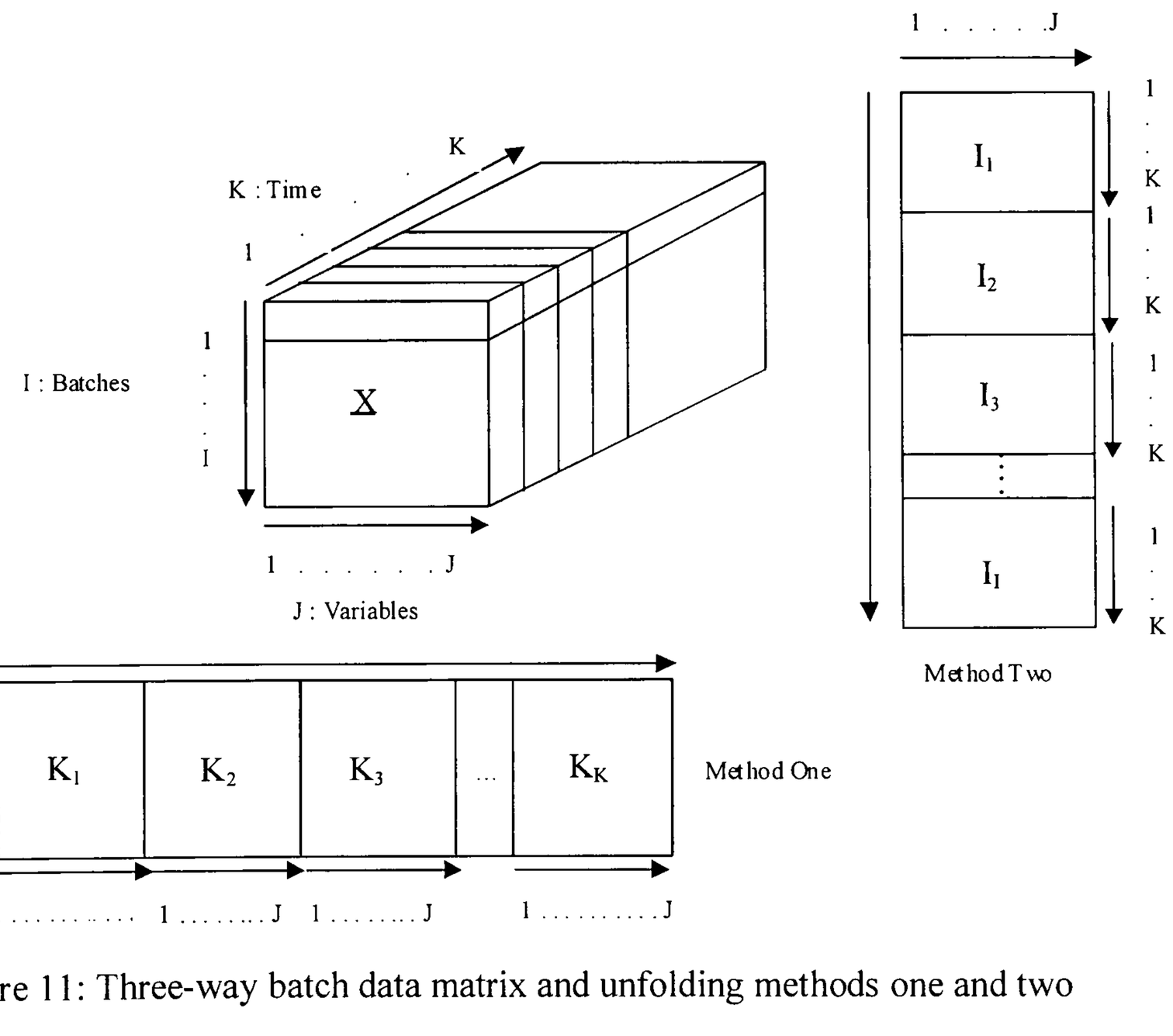


Figure 11: Three-way batch data matrix and unfolding methods one and two

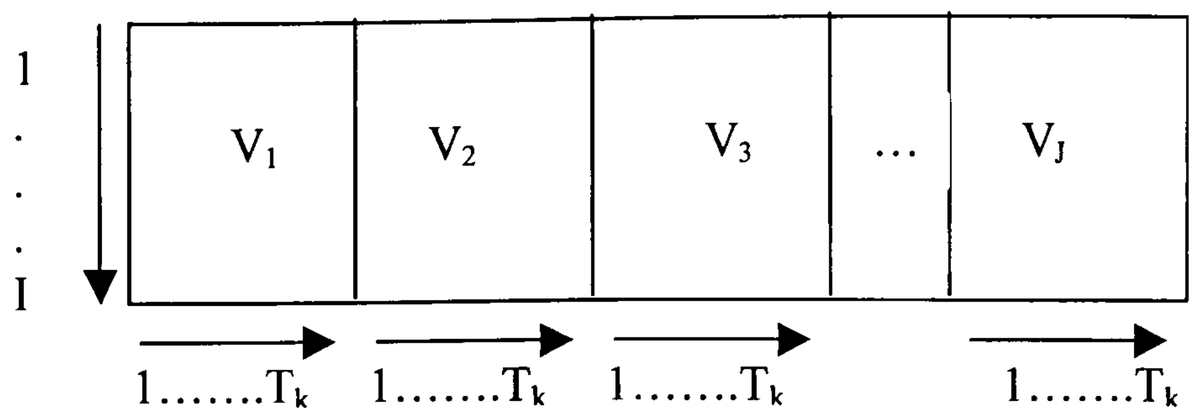


Figure 12: Alternative method one unfolding approach

Figure 13 shows the third method of unfolding the array. This method does not distinguish batch specific information from the time and variable information and is thus unsuitable for batch monitoring.

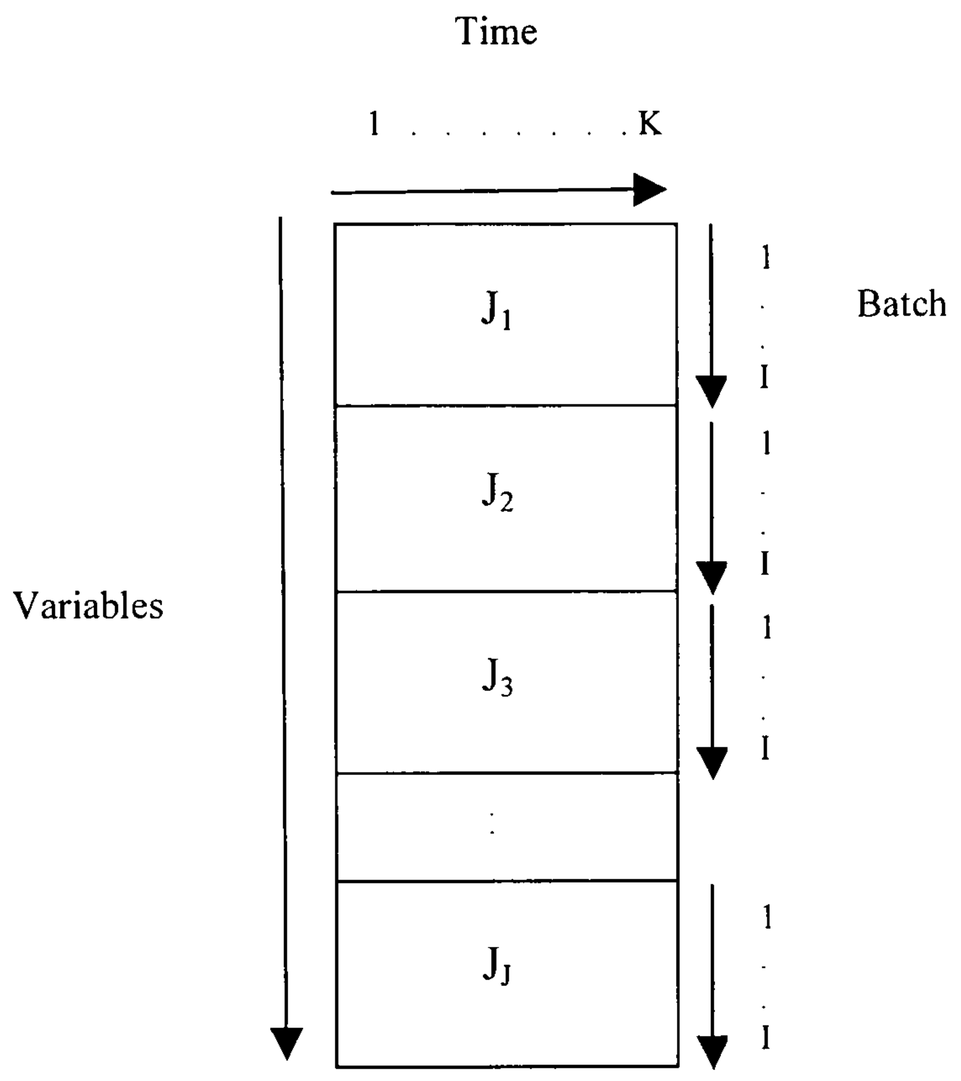


Figure 13: Unfolding method three

The subsequent sections describe how unfolding methods one and two work. In practice, the selection of the unfolding method depends on the final objective of the data analysis.

3.3 Multi-way Principal Component Analysis Using Unfolding Method One – End of Batch Technique

3.3.1 Overview

One of the first modifications of the standard multivariate analysis technique of PCA to enable it to handle batch data was introduced by Nomikos (1996), in the form of multi-way principal component analysis (MPCA). MPCA is an extension of the PCA technique (section 2.2) and the objective is to decompose the batch data array $\underline{\mathbf{X}}$ into a number of principal components that describe the normal variation of the process variables about a desirable time trajectory. The first step is to unfold the batch data array, $\underline{\mathbf{X}}$, into a two way array $\mathbf{X}$ (I ×JK), where each consecutive block defines a specific time period and contains the values for the J process variables at time point K_i for all I batches, i.e. Method 1 in Figure 11.

Unfolding the data matrix in this manner allows a PCA model to be built, as described in section 2.2. This approach allows the monitoring of the deviations of the batches from normal operating conditions. Prior to the application of the PCA algorithm, the unfolded data matrix is scaled, i.e. the mean of each column is subtracted from each element of the column, and each column is then divided by its standard deviation. Pre-treating the data in this way theoretically removes the main non-linear components from the data, allowing techniques such as linear PCA to be applied. Additionally by standardising the data, the differences between the measurement units for each variable are taken into account with each variable having an equivalent weighting. The centred and scaled data is then used to build a process representation and this forms the framework on which future batches are monitored. The model is given as follows, Nomikos *et al* (1994):

$$\underline{\mathbf{X}} = \sum_{l=1}^{r} \mathbf{t}_l \otimes \mathbf{P}_l + \underline{\mathbf{E}} \qquad (3.1)$$

where $\otimes$ is the outer product operator, r is the number of principal components retained, $\mathbf{t}_l$ is the score vector related to the batches, $\mathbf{P}_l$ is the loading matrix related to the variables and their time variation and $\underline{\mathbf{E}}$ is the residual matrix. The Q and Hotelling's T^2 statistics can then be calculated using the unfolded solution as described in section 2.2. MPCA explains the variation of the measured variables about their average trajectories.

Subtracting the average trajectory from each variable (accomplished by mean-centering the columns of the unfolded matrix **X**) removes the major non-linear behaviour of the process. The i^{th} element of the **t**-score vector corresponds to the i^{th} batch (or sample) and summarises the overall variation in this batch with respect to the other batches in the database over the entire history of the batch. The **P** loading matrix summarises the time variation of the measured variables about their average trajectories. The elements of **P** are the weights, which when applied to each variable at each time interval within a batch, give the **t** scores for that batch. This technique is also known as the end of batch (EOB) method or the Nomikos and MacGregor method (N&M).

3.3.2 MPCA Using Unfolding Method 1 – Example using Exothermic Batch Reactor

To demonstrate the monitoring metrics for the MPCA end of batch method, 50 batches from the exothermic batch simulation, Luyben (1976) described in detail in Chapter 4 are considered. A bivariate scores plot and the associated squared prediction error (SPE, or Q statistic) and Hotelling's T^2 plots for a two principal component model are shown in Figure 14 and Figure 15 respectively. It can be seen that the majority of batches are spread around the origin of the bivariate scores plot, with one batch lying outside the 95% control limits, this is statistically acceptable. The SPE plot and Hotelling's T^2 plot for the batches show that two batches exceed the 95% control limit, which is again acceptable for a data set of size fifty batches.

Figure 16 shows the univariate scores plot for all fifty batches for principal component one with the corresponding scores contribution plot for batch 38; one of the batches in the nominal data set. The contribution plot shows the behaviour of each variable over the duration of the batch, comparisons can be made with batches that remain in statistical control with those that are outside of the limits to assess which variables are the cause of the abnormal process behaviour. A univariate loadings plot showing the variables over time is shown in Figure 17 with an SPE contribution plot for batch 38, again showing how the variables behave over time. The highest contributions for variables 3 and 4, which are a jacket temperature and cooling valve in the reactor, occur around time point 100, whereas variable 1, the reactor temperature, gives the largest contribution to the variation in the batch at around time point 50, where the temperature is increasing to the point of reaction before the cooling system starts to operate.

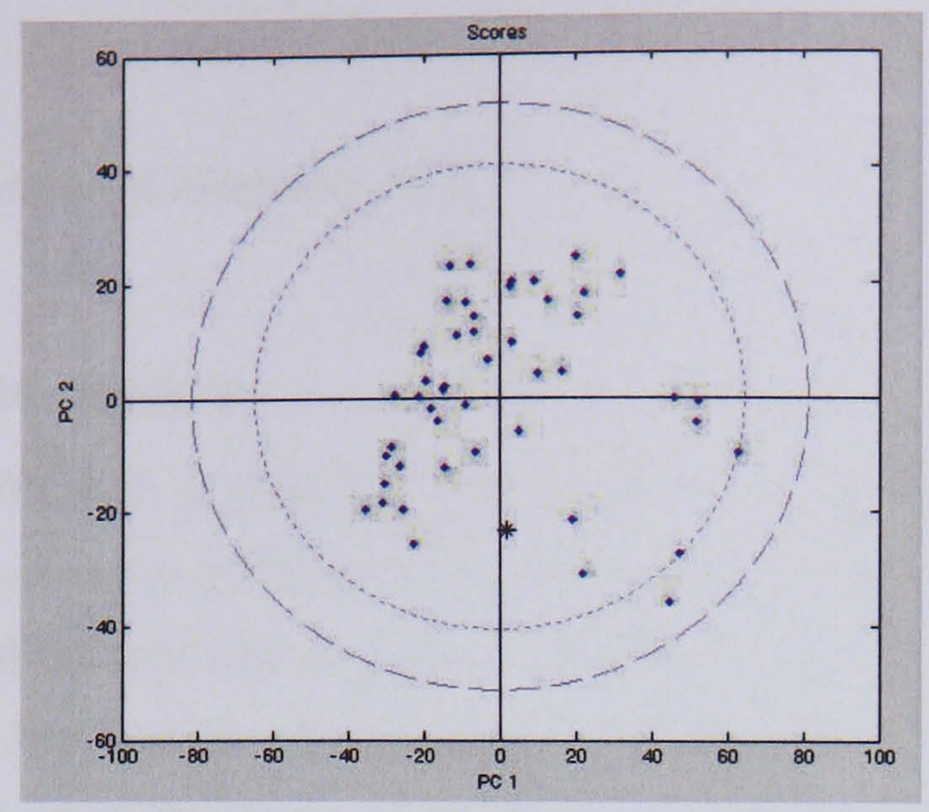
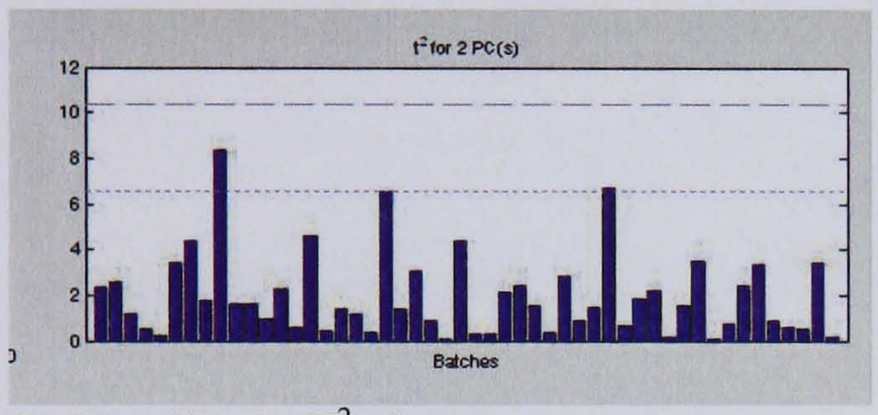
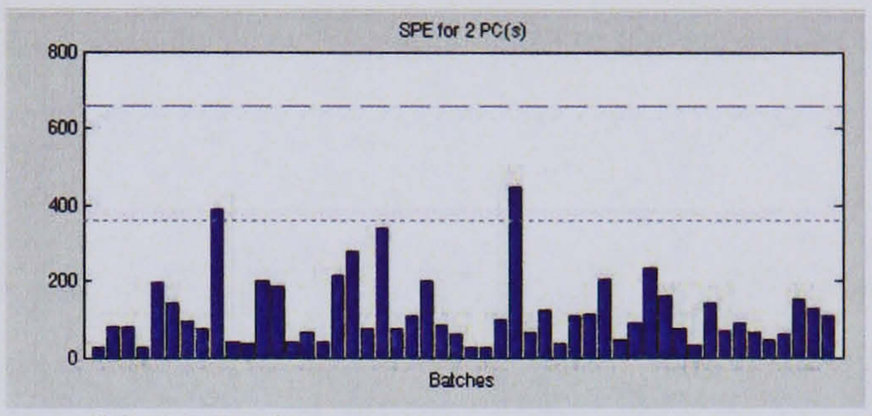


Figure 14 : Bivariate scores plot: two principal component model

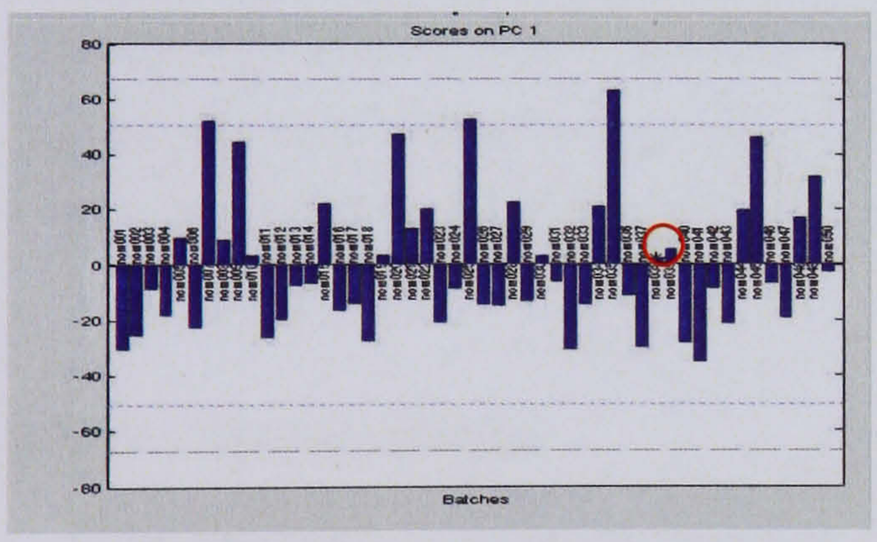


(a) Hotelling's T^2 plot



(b) SPE plot

Figure 15: Hotelling's T^2 plot and SPE plot for 50 batches



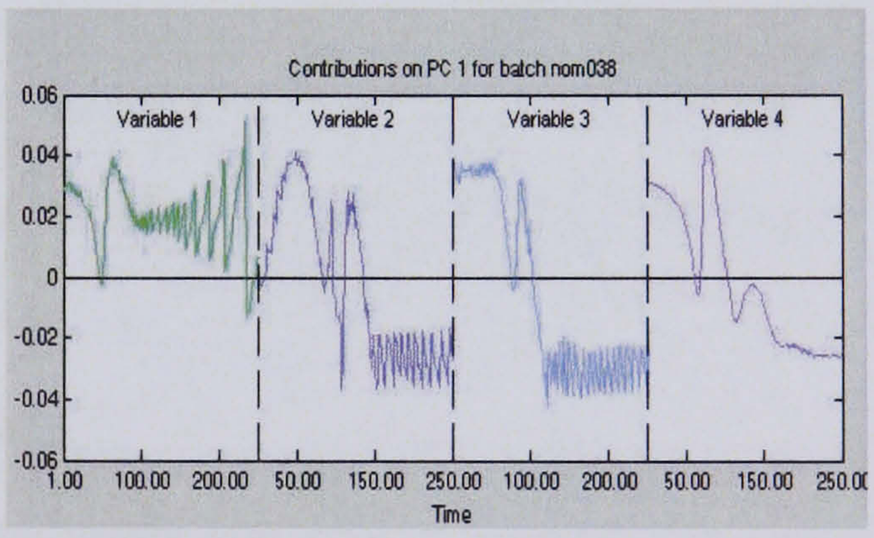
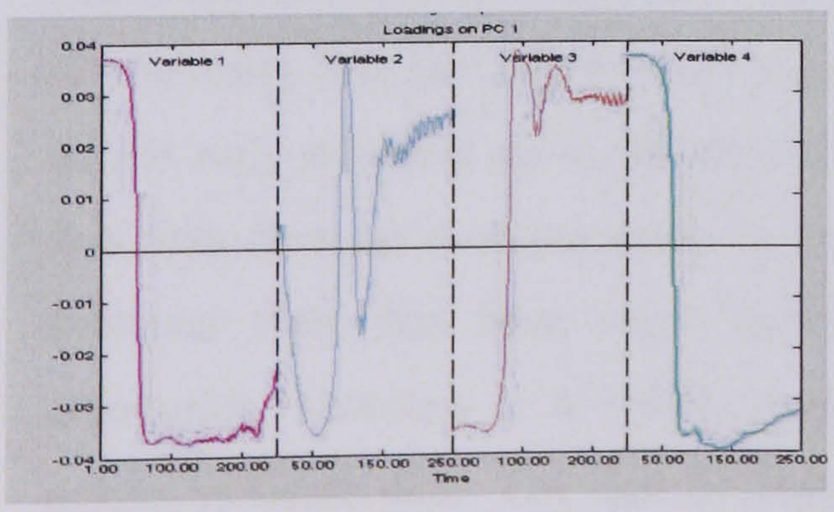


Figure 16 : A univariate scores plot for PC1 with a scores contribution plot for batch 38



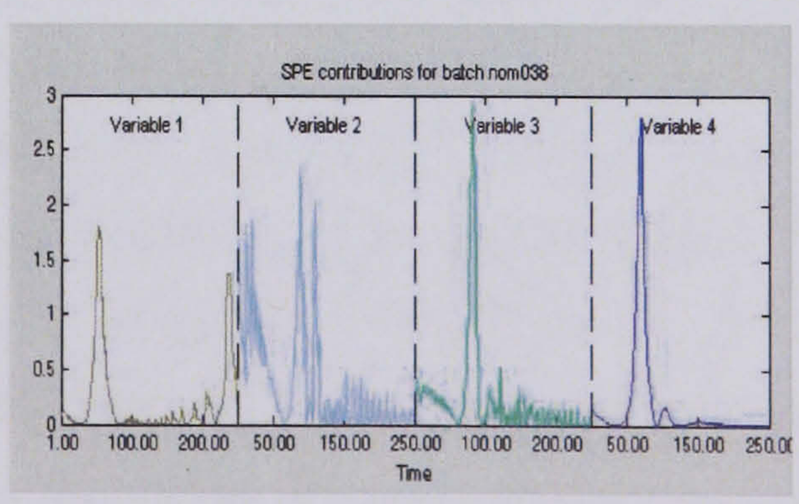


Figure 17 : A univariate loadings plot with an SPE contribution plot for batch 38

3.3.3 Limitations of Multiway PCA

3.3.3.1 Off-line Analysis

Although MPCA is a good tool for batch monitoring and analysis there are some limitations associated with the technique, the most important being that the technique does not deal directly with the serial correlation in the data. MPCA does not take into account the fact that the current process measurements are affected by the previous history of the batch because it is essentially treating the unfolded data array as a static dataset. As batch processes display dynamic behaviour, and the calculation of the control limits do not take this into account, abnormalities in the data may be masked or wrongly interpreted. This has been discussed by a number of authors, including Chen *et al* (2002).

Linked with this is the fact that batch data typically exhibits non-linear behaviour. Although the scaling method claims to remove the non-linear component from the data set by removing the average batch trajectory, this may not be the case, and therefore the application of a linear technique may result in variation in the data that is not explained. In practice, the non-linear aspects of the process are often described by the higher order components, and in a technique such as MPCA it is usually only the first few principal components that are retained to describe the process variation with the higher order ones being discarded through cross-validation as it is assumed they describe process noise.

3.3.3.2 On-line Analysis

While the issues of serial correlation and non-linear behaviour also apply to the on-line use of MPCA, there are specific issues associated with applying the technique on-line. One such issue with standard MPCA is that the full time history of the batch is required for the analysis to be carried. More specifically, when monitoring a process in real time, data is only available up to the current time point, and hence the direct application of new data from an evolving batch to the normal process representation is not possible. However there has been much interest in using the technique for on-line batch monitoring, Nomikos *et al* (1994) proposed three approaches for the in-filling of the unknown future data. The first of these is the current deviation approach where it is assumed that the batch trajectory will continue to deviate from the desired operating

conditions at the current level and the remaining time points are in-filled with this value. For the second method, the zero deviation approach, it is assumed that the batch trajectory will not deviate from the desired level of operation for the remaining duration of the batch and hence the data matrix is in-filled with zeros (i.e. the mean value of zero). In the final method, the missing data approach, the future batch data is considered to be missing and hence the scores and residuals are calculated using only the loadings up to the current time point. Therefore the loadings continue to evolve over the duration of the batch. This necessitates a new model being computed at every time point, or for blocks of time points, but this is possible with the advent of increased computational power.

These three techniques have their own identifiable limitations. For example the current deviation method assumes that the out of control variable(s) will remain in this state, no matter at which point in the batch they were detected. This may not necessarily be the case since the batch could recover itself before there was any detrimental effect on the product quality, or indeed new deviations could develop which mask the original fault. The zero deviations technique assumes that the out of control variables will return to their normal operating state prior to the batch being completed. Again this is not always true. Finally, the missing data approach gives poor performance monitoring at the beginning of the batch when only a small fraction of the loadings are used in the calculation. This can be crucial since in batch processes it is quite common for the majority of the reaction to occur during the early stages of the batch and therefore it is essential that there is adequate process monitoring at this stage. Therefore, when using MPCA, the most appropriate method of in-filling needs to be selected based on knowledge and experience of the process. It is possible that a combination of approaches may be best suited for some processes. However through the application of these techniques, it becomes possible to monitor the MPCA scores with respect to time.

The other limitation of MPCA is that the technique requires the batches to be of equal duration. In an industrial situation, this is unlikely to happen as the duration can vary significantly between batches. Much work has been carried out in this area and there are three proposed solutions. The first is to cut the batches to the same length (where length corresponds to the time of the last sample of the batch) i.e. batches longer than the new length are truncated whilst batches shorter than the new length are in-filled with missing values (for example using zero or first order interpolation, local averages or first order regression).

The second method involves the use of a technique such as dynamic time warping, DTW, Gollmer *et al* (1996). DTW is a method that was originally used for speech recognition. The concept is based on the fact that a computer has to be able to recognize a word, regardless of the speed used to pronounce it, hence DTW is used to re-order the digitized sound and fit it to a known pattern. Batch processes may progress at different speeds while still following the same process signature such as reactant conversion etc, therefore DTW can be used to re-order the batch data matrices to following a known pattern, such as the reactant conversion. The final method is to use a maturity or indicator variable. In this case the monitoring basis is switched from time to the maturity or indicator variable. The choice of the maturity index is process specific and it is important that the selected variable is monotonic, continuous and spans the full range of the other process variables included in the analysis.

Overall, MPCA provides a good tool for post-batch analysis and batch classification, however the fact that the process dynamics are not accounted for in the analysis is an important issue for consideration, as this is an important aspect of a batch process.

3.4 Multi-way Partial Least Squares (MPLS)

A natural extension to MPCA is multi-way Partial Least Squares, Nomikos *et al* (1995). A major issue in the monitoring of batch processes is the lack of available on-line quality measurements and since in practice the majority of batch processes operate under open loop with respect to quality, there is often no method of determining if the batch will produce a poor quality product until after it is complete and the analysis is carried out in the quality assurance laboratory. This can be extremely time-consuming. Multi-way partial least squares (MPLS) uses both the quality data and the process variable measurements to produce a model of the process that captures the covariance structure between the two data sets and hence provide inferential values of the final product quality parameters.

MPCA compresses the data to describe the main source of variation in the 3D array, $\underline{\mathbf{X}}$, the process data, while MPLS compresses the data in such a way that the resulting latent variables identify the variance of $\underline{\mathbf{X}}$ that is most predictive of $\mathbf{Y}$, the quality data, as described in section 2.3. Consequently MPLS can be used as an on-line inference tool for the final product quality, therefore providing a real time measurement of whether the batch is within specification. However, in some cases there may only be a limited

number of **Y** values, for instance laboratory analysis from the end of a batch, which could mean the model is not as representative of the full range of **X** data as it only uses data from the end of the batch as opposed to throughout its history.

Basically multi-way PLS can be thought of as performing a PLS analysis on a large two dimensional matrix. To build the model, the data array, $\underline{\mathbf{X}}$, is first unfolded in the same manner as for multi-way PCA, to give a matrix **X** of order I x JK (as shown in Figure 11). Likewise for **Y** which is a matrix of order I x M, (final quality variables m=1,2,…,M). **X** and **Y** are mean centered and scaled to unit variance, which means the MPLS model explains the variation of each variable around its average trajectory at each time point. PLS is then performed on the unfolded data set to decompose the **X** and **Y** matrices into a summation of r scores, **t** (I x 1), and loadings, **p** (JK x 1), **q** (M x 1) vectors, plus residual matrices, **E** (I x JK) and **F** (I x M):

$$\mathbf{X} = \sum_{l=1}^{r} \mathbf{t}_r \mathbf{p}_r^T + \mathbf{E} \qquad (3.2)$$

$$\mathbf{Y} = \sum_{l=1}^{r} \mathbf{t}_r \mathbf{q}_r^T + \mathbf{F} \qquad (3.3)$$

Collating the r, **t**, **p** and **q** vectors to give **T** (I x r), **P** (JK x r) and **Q** (M x r) matrices gives:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathbf{E} \qquad (3.4)$$

$$\mathbf{Y} = \mathbf{T}\mathbf{Q}^T + \mathbf{F} \qquad (3.5)$$

where **T** is given by:

$$\mathbf{T} = \mathbf{X}\mathbf{W}(\mathbf{P}^T\mathbf{W})^{-1} \qquad (3.6)$$

T represents the matrix of latent variable scores, with each row defining a single batch. The scores describe the overall variability of each batch with respect to the other batches. **P** and **W** summarise the variation over time of the variables about their average trajectories and their elements are the weights for each variable at each individual time

point, and **Q** relates the variation in the process measurements to the final product quality.

The advantage of MPLS over MPCA is that through the use of historical process data and quality measurements that represent in-specification product, it is possible to build a model that will infer the final product quality from early in the batch. However the same set of limitations hold for this technique as for MPCA with respect to the application of the technique on-line, in that there is a need to estimate the future observations for an evolving batch. The methods as described for MPCA (3.3.3.2) are again applicable. The other limitation is that it does not effectively capture the dynamic behaviour in the data.

A further difference between MPCA and MPLS, which can be seen either as an advantage or a disadvantage, depending on the monitoring objective, is that since MPCA works by only looking at the variance in the process measurements in matrix $\underline{\mathbf{X}}$, it will detect any irregularities in the process whether or not they impact on the final product quality. Conversely, MPLS will only identify process deviations that have been identified as being influential through the model on product quality, and thus any variable that exhibits abnormal behaviour that has not been included in the model will not be identified. This is a disadvantage if the operator is interested in all potential sources of abnormal behaviour and the potential consequences it may have on the equipment and process rather than just the behaviour that affects final product quality.

3.5 Multi-block Multi-way Techniques

Multiway PLS incorporates information about the product quality along with the process variables. Where additional batch information is available, such as initial batch conditions like raw material qualities, initial ingredient charges, or process measurements from a preceding process, a multi-way multi-block PLS technique, Kourti *et al* (1995), can be implemented.

The matrix and array structures for the technique are shown in Figure 18, matrix **Z** contains the additional data on the initial conditions, array $\underline{\mathbf{X}}$ contains the on-line measurements and matrix **Y**, the quality measurements. Array $\underline{\mathbf{X}}$ is unfolded as discussed previously in section 3.2 and then matrix **Z** and the unfolded $\underline{\mathbf{X}}$ are treated as two blocks and scaled and weighted appropriately. Multi-way PLS can then be applied as illustrated in section 3.4. The weighting of the blocks can be decided either by utilising process

understanding, or alternatively should no information be available, equal weightings can be assigned. It is recommended that several models should be derived with different weight ratios and the best model is that for which the cross-validated percentage explained in **Y** is equal to, or higher, than that explained by a PLS model using each block separately. Consequently this extension to multi-way PLS allows additional process information to be incorporated into the model, providing a powerful monitoring and tracking tool for new batches. However, the same limitations as for MPLS apply.

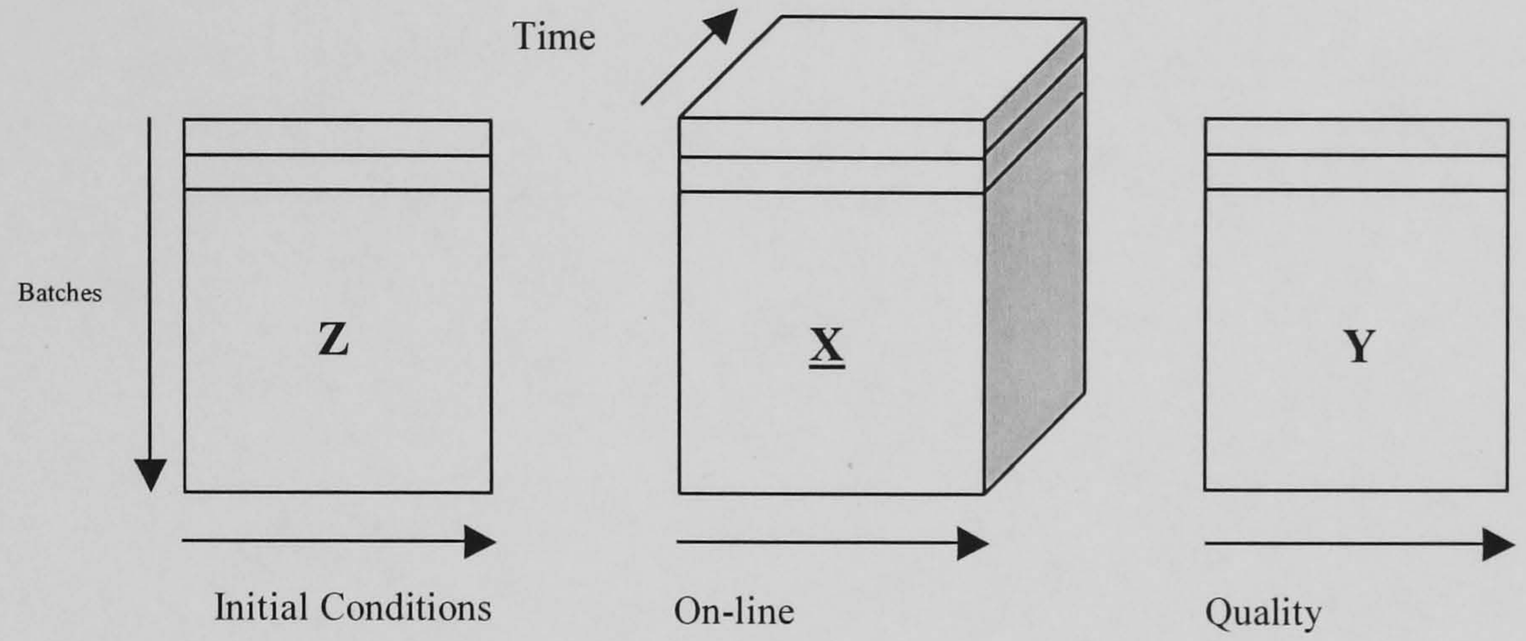


Figure 18 : Nature of the Batch Databases

3.6 Tri-Linear Batch Monitoring Techniques

Multi-way PCA and multi-way PLS are bi-linear techniques that have been widely applied for the monitoring of batch processes. However an alternative approach for dealing with 'box' style data has been proposed based on tri-linear methodologies including parallel factor analysis (PARAFAC) and Tucker3 models, Louwerse *et al* (2000). These methods do not require data unfolding, an advantage for data preparation, and instead retain the 3-dimensional structure of the data array. A two-way matrix representation of the PARAFAC model of $\underline{\mathbf{X}}$ with r components can be written as follows:

$$\underline{\mathbf{X}} = \mathbf{A}(\mathbf{C} \circ \mathbf{B})^{T} + \underline{\mathbf{E}} \qquad (3.7)$$

where **A** (I x R) represents the scores matrix and **C** (K x R) and **B** (J x R) are the loadings, 'o' is the Khatri-Rao product of **C** and **B** (partitioned into columns) and $\underline{\mathbf{E}}$ (I x J x K) denotes the residuals. Without altering the model, the vectors of **B** and **C** can be

normalised to unit length if the matching vectors of **A** are modified, i.e. inversely scaled. Interestingly there is only one solution to the PARAFAC model, this compares to the non-uniqueness of a bilinear model, i.e. there is no rotational ambiguity. Figure 19 provides a visualisation of the structure of the resulting analysis.

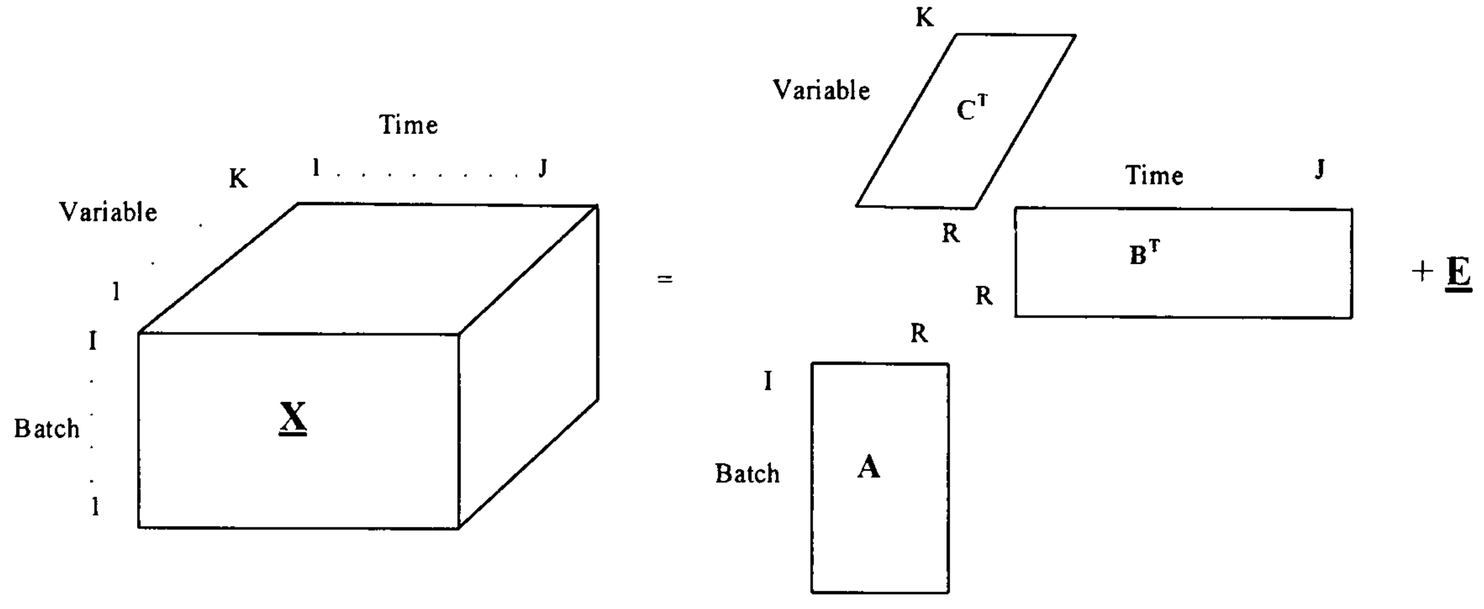


Figure 19 : Three dimensional PARAFAC analysis representation

Another tri-linear method commonly applied is the Tucker3 model. A two-way matrix representation of the Tucker3 model of $\underline{\mathbf{X}}$ (I x J x K) with Q, S and T components (for the first, second and third modes respectively) is given by:

$$\mathbf{X} = \mathbf{AH}\,(\mathbf{C} \otimes \mathbf{B})^{T} + \mathbf{E} \qquad (3.8)$$

Where **A** (I x Q) represents the scores, **B** (J x S) and **C** (K x T) are the loading matrices, $\otimes$ is the Kronecker product, **H** (Q x ST) is a two way representation of $\underline{\mathbf{H}}$ (Q x S x T), i.e. the weights for all possible component interactions, and **E** (I x J x K) is the residual array. Without changing the model, the matrices **A**, **B** and **C** can be made column-orthogonal such that the variance in the model is accounted for in $\underline{\mathbf{H}}$. Figure 20 illustrates the structure of the resulting analysis.

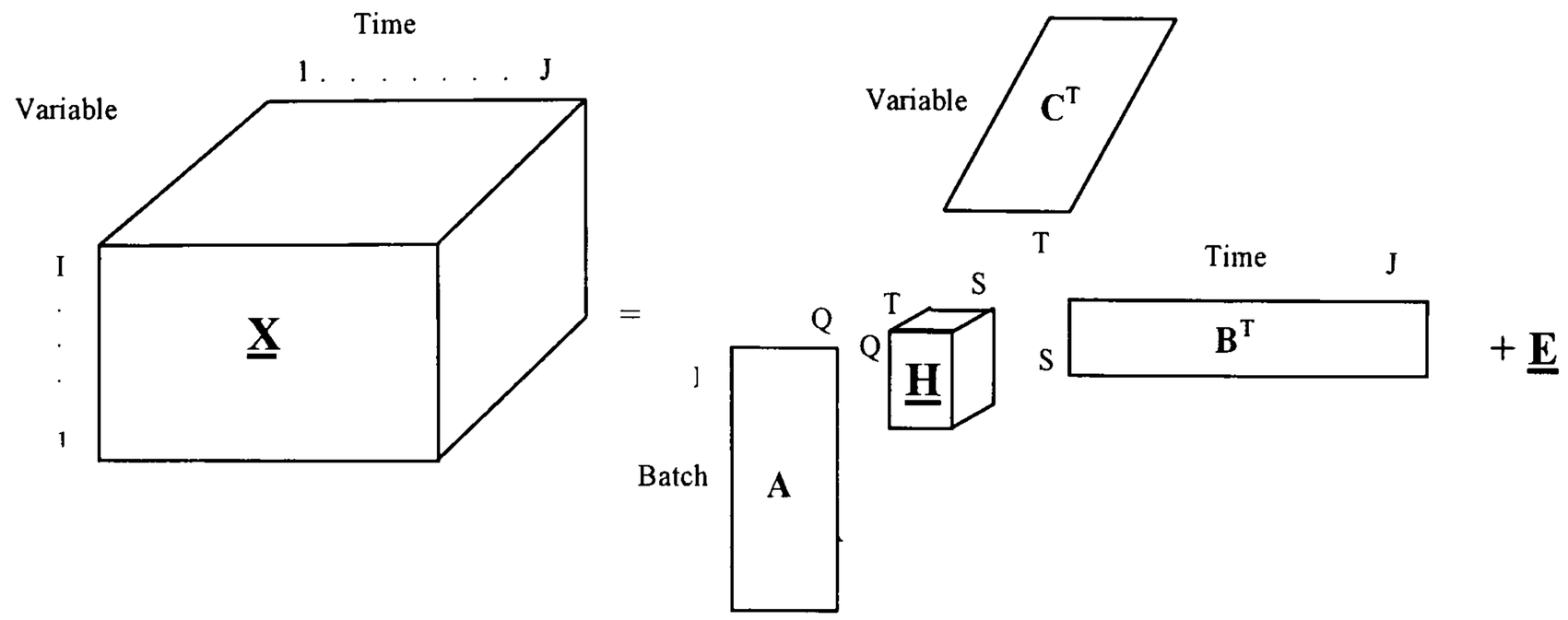


Figure 20 : Three dimensional Tucker3 analysis representation

In contrast to the PARAFAC methodology, the Tucker3 approach is not structurally unique, therefore Tucker3 models can be complex and give ambiguous results. Although these methods have a simpler data preparation step they may not always adequately represent the data and detect faults as well as MPCA, this is because MPCA estimates a parameter for every component, for each time point, of every variable and hence this large number of parameters can describe a significant amount of variability in the data. For example, the PARAFAC model only estimates a parameter for each variable over the time points and a parameter for each time point over the variables, therefore there is some averaging and hence less variation is explained compared to MPCA or the Tucker3 model (which does allow interactions between a particular component of a certain mode with all components of the other modes). The tri-linear approaches are not explored further in this thesis as the focus is on the bi-linear techniques.

3.7 Multi-way Principal Component Analysis Using Unfolding Method Two – Batch Observation Level Analysis

3.7.1 Overview

MPCA can be referred to as a batch level form of process monitoring. An alternative approach to MPCA for batch performance monitoring that was developed by Wold *et al* (1998) is that of batch observation level (BOL) analysis. This is an observation level based approach or more specifically a through batch modelling method. The advantage of the batch observation level technique is that it is designed to allow the behaviour of

the batches to be monitored as they evolve as opposed to waiting till the process is complete to assess batch performance. This is a consequence of the manner in which the batch data matrix $\underline{\mathbf{X}}$ is unfolded whereby the variable trajectory is not disturbed. Applying the observation level technique rather than a batch level technique enables the inference of batch maturity, i.e. how close to completion is the batch, as well as using the method to understand and monitor the evolution of the batch.

The first stage in the analysis is the unfolding of the three-way batch data matrix. In batch observation level analysis, the data is unfolded into a (IK x J) matrix, as shown in Figure 11, Method two. This method of unfolding is used to preserve the dynamic direction of the process variables, time, i.e. the history of each variable is retained. This is in contrast to the unfolding method (method one) used in N&M end of batch MPCA which preserves the direction of the batches, either all the measurements taken at the same time are collated, Figure 11, or all the measurements taken for each variable are kept together, Figure 12. So in method two, for batch observation analysis, the unfolded batches form a stack with each row corresponding to the variable values at a specific time point, for a specific batch, the method retains the direction of the variables, i.e. the trajectory of each variable over the batch. The unfolded matrix is then mean centred and scaled, i.e. each column is standardised. The means are in this case the mean of the entire variation for each variable for all time points and batches, i.e. the global mean. The impact of this is that rather than removing the average trajectory (as with MPCA), the average trajectory will be described by the scores values of the first principal component.

Utilising the unfolded batch data matrix, a PLS model is then built between the unfolded data matrix **X** (IK x J) and a dummy variable, **y**. The dummy variable is typically the local batch time but may be an alternative variable such as monomer conversion or yield, this is shown in Figure 21. The model is as follows:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{T} + \mathbf{E} \tag{3.9}$$

$$\mathbf{y} = \mathbf{T}_{R}\mathbf{c} + \mathbf{f} \tag{3.10}$$

where **T** (IK x r) represents the scores matrix for the r retained latent variables, **P** (J x r) is the loadings matrix, **c** is the vector attained by regressing **y** on the PLS scores, and **E** (IK x J) and **f** are the residuals. Each individual score vector corresponds to the evolution of each batch over time. The number of latent variables, r, is selected to ensure the **X**

data is adequately summarised according to equation 3.9 and the **y** is predicted according to equation 3.10.

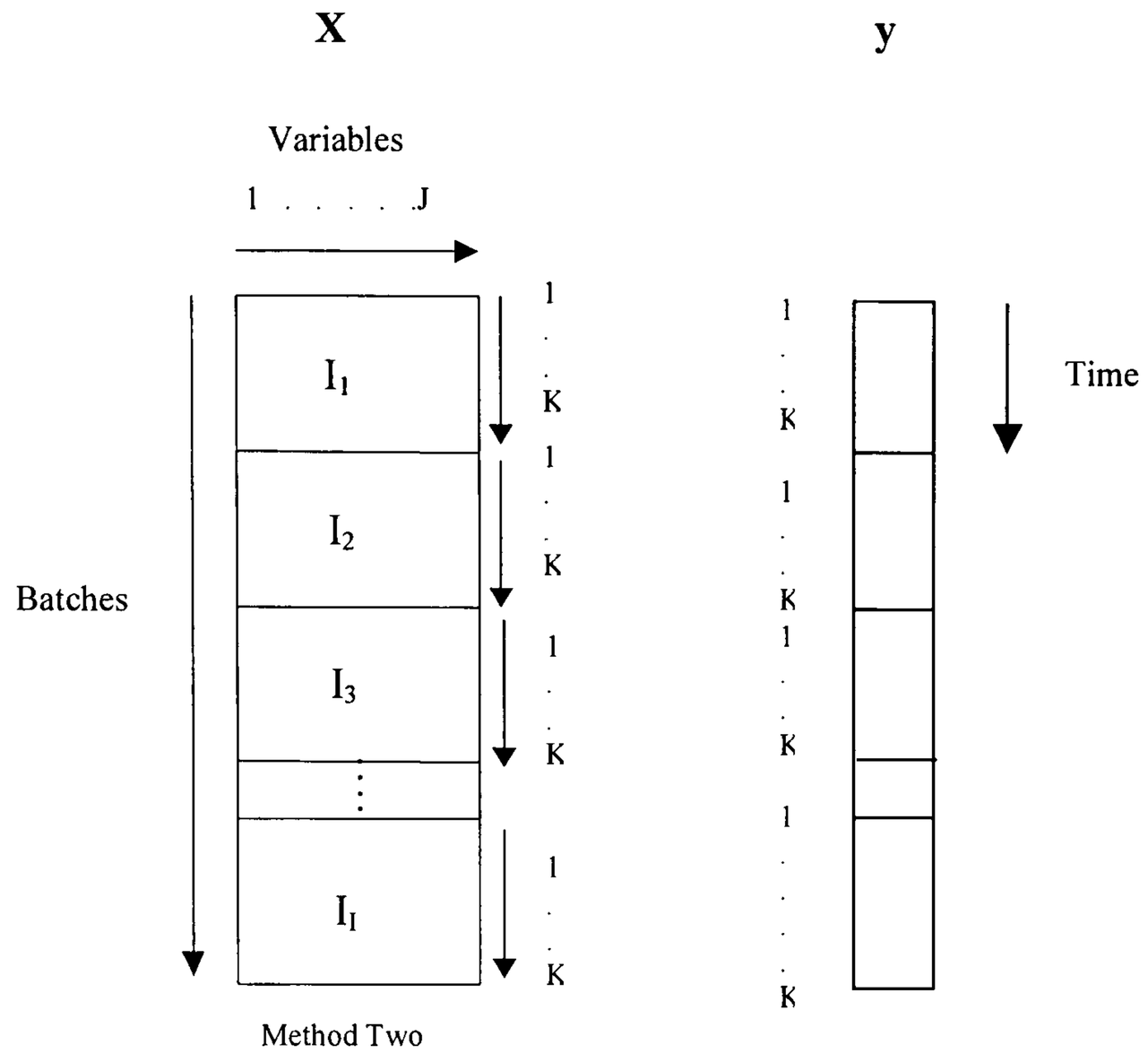


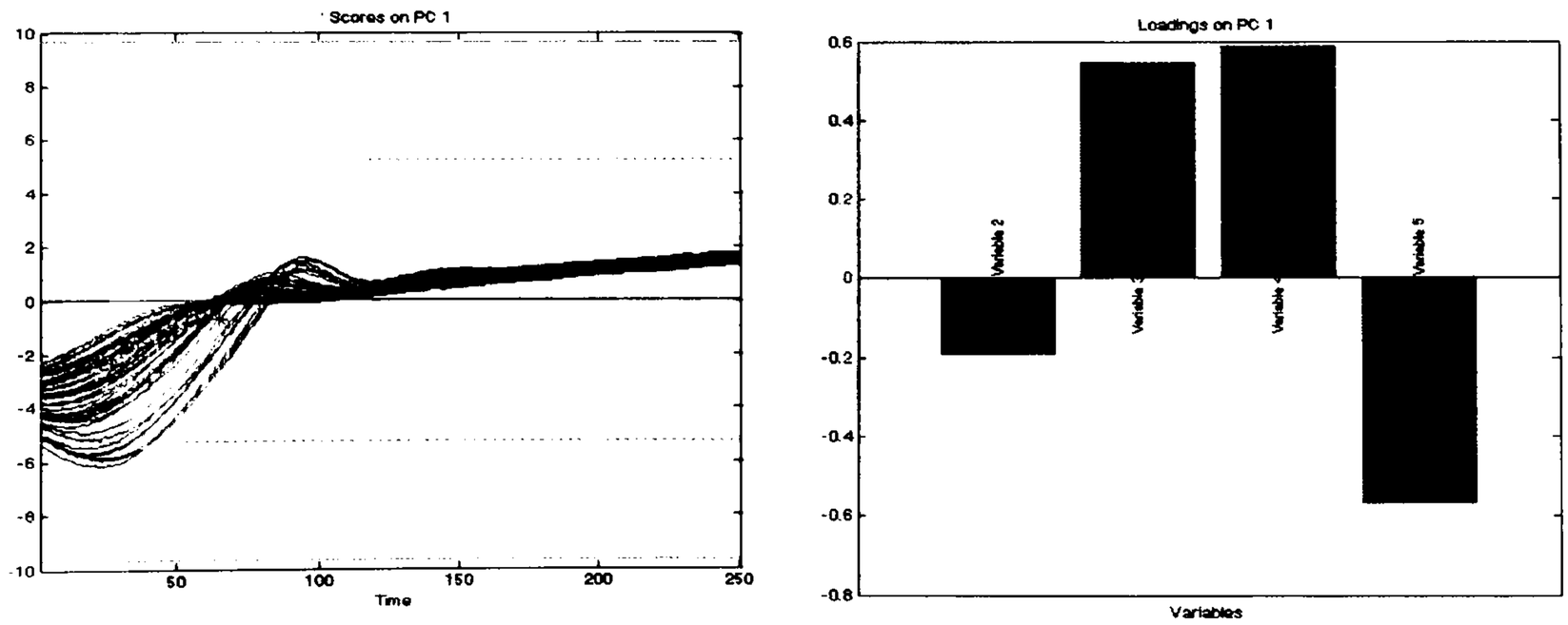
Figure 21 : The unfolded matrix and dummy variable

The local time, $\mathbf{y}_{pred}$, predicted by the PLS model can be used as a maturity index to indicate how far the batch has progressed. One advantage of the batch observation level technique is that it does not require all batches to be of the same duration, since the unfolding of the 3-dimensional batch matrix is performed over the variable direction. However, if there are batches that are greatly varying in duration it is possible to use the $\mathbf{y}_{pred}$ maturity variable to realign them in terms of interpolated data (e.g. 0%, 5%, 10%, etc of completion) so that each one is of equal length if it is so desired. Once the re-alignment has been performed the PLS calculation is carried out again with **y**, equation 3.10, this time being the value of the maturity index, often the 'aligned' $\mathbf{y}_{pred}$ from the first analysis, as opposed to the local batch time. In addition to the benefit of not requiring equal length batches, it is straightforward to monitor the batch as it is evolving rather than waiting for a complete batch, or having to make assumptions about future batch behaviour, as is required with end of batch (N&M) MPCA.

3.7.2 Batch Observation Level Monitoring Charts

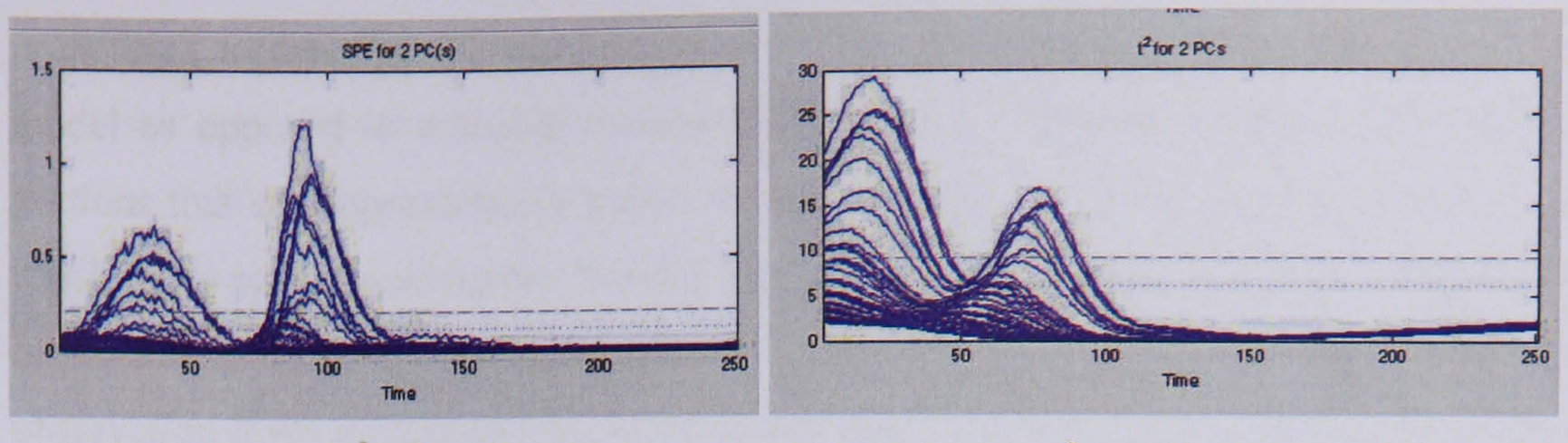
Once the PLS calculation has been performed, the next stage of the batch observation level technique is a statistical analysis of the scores to generate the control charts from which the process can be subsequently monitored. This step is achieved by rearranging the resulting matrix over the batch dimension, that is the scores for each batch form a separate row, $\mathbf{X}_T$ (I x rK). The 95 and 99% control limits are then calculated at each time point as opposed to globally. When each new sample point of an evolving batch is obtained the variables are standardised using the mean and standard deviation calculated for each variable in the nominal data set and the scores calculated. The point is then plotted, and, utilising the control limits from the normal data set, the status of the process is identified.

A univariate scores plot for PC1 of the 50 nominal batches from the exothermic batch simulation is shown in Figure 22 (a), and the corresponding loadings plot is shown in Figure 22 (b). It is easier to monitor the evolution of the batches over time using the univariate scores plot in comparison to the bivariate scores plots. It can be observed that there is greater variation in the batch behaviour during the first 100 time points. The loadings plot shows that variable 1, reactor temperature, has the least impact on principal component 1, and that variables 2 and 3, the wall and jacket temperatures respectively, are negatively correlated with variable 4, the cooling valve.



(a) Univariate scores plot for PC1 (b) Univariate loadings plot for PC1

Figure 22: Univariate scores and loadings plot for PC1

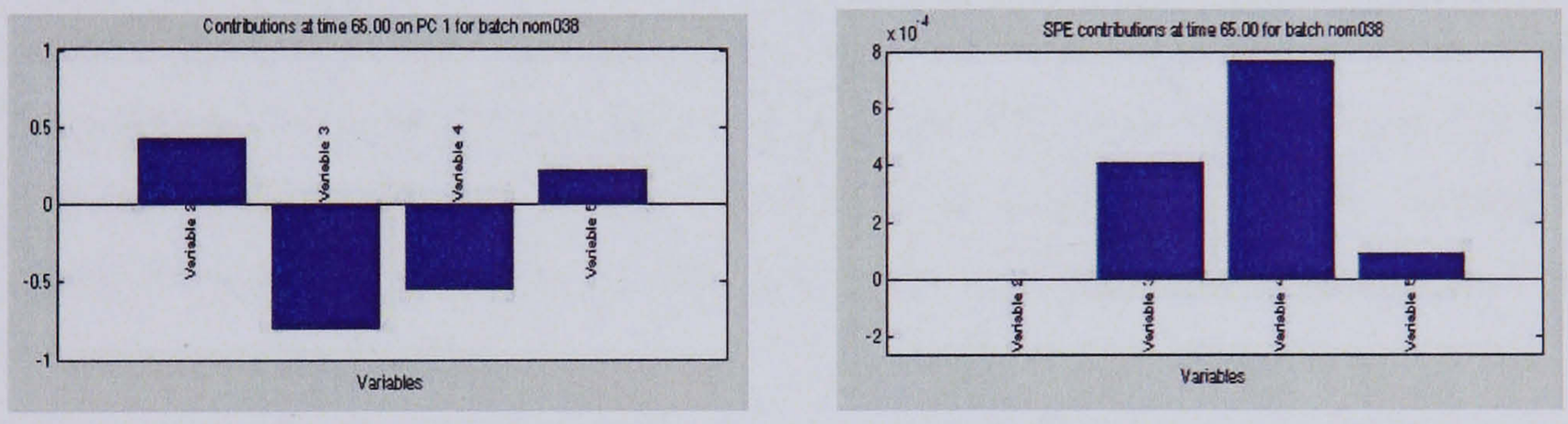


(a) SPE plot for 50 nominal batches (b) Hotelling's T^2 plot for 50 nominal batches

Figure 23: SPE and Hotelling's T^2 plots for 50 nominal batches

Figure 23 shows the SPE and Hotelling's T^2 plots for the fifty nominal batches. Both these plots show the evolution of the monitoring statistics over time, however as the control limits have been calculated globally, the first phase of the batch appears to be out of statistical control because greater variation occurs during this time point.

The associated scores contribution plot and SPE contribution plot for one of the nominal batches at time point 65 is shown in Figure 24. The scores contribution plot shows that variable 2, the reactor wall temperature, has the greatest impact on the principal component trajectory at time point 65, whereas from the SPE contribution plot, it is variable 3, the jacket temperature, that has the largest impact.



(a) Scores contribution plot (b) SPE contribution plot

Figure 24: Scores and SPE contribution plots for one nominal batch at time point 65

3.7.3 Batch Observation Level Analysis Conclusions

To summarise, the main advantage of the batch observation level technique for performance monitoring is that it is designed to monitor a batch as it evolves, as opposed to having to wait for data from a completed batch or having to make assumptions about the future behaviour of the process. However, the BOL approach does have some disadvantages, these are mainly related to its failure to incorporate the dynamics and non linear behaviour inherent within a batch process in the process representation. The BOL technique does not explicitly provide any information about the process dynamics, which is a key factor with batch processes and also an issue with MPCA. Additionally, it

is difficult to describe the non-linearity in a process using a single global linear PLS model as opposed to multiple models that reduce a non-linear problem into smaller sections that are approximately linear. In end of batch MPCA (unfolding method one), this issue is partially addressed through the removal of the major non-linear component of the data by applying scaling to each time point for all individual variables. The batch observation level technique uses a global approach to scaling and hence does not deal with the non-linear behaviour and therefore process behaviour may be masked by the non-linear behaviour inherent within the data. As a consequence, it has been proposed that batch observation level analysis could be carried out using the scaling approach proposed by Nomikos *et al* (1996) for MPCA prior to performing the PLS calculation. This was not investigated further in subsequent chapters of this thesis, however it should be considered for future work. An alternative option is to apply a local modelling technique to the batch observation level analysis, as proposed by Fletcher *et al* (2001). In this latter research, multiple local linear models were used to describe different sections of the process as opposed to the use of a single global model.

Another issue that has been raised with respect to the batch observation level analysis is that for this technique, the scores are originally calculated using a principal component analysis performed on the batch data matrix unfolded in the batch-wise direction. When the technique is then used for on-line monitoring, the scores are rearranged and unfolded over the batch dimension. This could be problematic and lead to a loss of information about the behaviour of the process since the scores were initially derived to explain the variation between variables.

To summarise, BOL is an alternative batch performance monitoring technique that allows the progress of evolving batches to be monitored. However the technique does not adequately address the issues of dynamics and non-linear behaviour inherent in batch data.

3.8 Adaptive Monitoring with Hierarchical, Moving Window and Adaptive PCA

3.8.1 Hierarchical PCA

One issue that has been identified with multi-way PCA is that is does not take into account the non-steady state nature of a batch process. Extensions to MPCA have been proposed that address some of the non-linear and dynamic features of a batch process,

these include hierarchical PCA, Ränner *et al* (1998) and van Sprang *et al* (2002), moving window PCA, Lennox *et al* (2001), and adaptive PCA, Li *et al* (2000). Although these techniques work in slightly different ways, they aim to implicitly include the time behaviour in the analysis of the data by computing several local PCA models to capture the correlation between variables at a number of different time stages in the process. The application of these approaches to batch processes is appropriate since batches tend not to conform to a steady-state and their behaviour changes as the reaction progresses.

Firstly, in the case of hierarchical PCA, for each frontal slice of the 3D array $\underline{\mathbf{X}}$, (I x J), a block score vector is computed, $\mathbf{b}_{Ki}$, Figure 25.

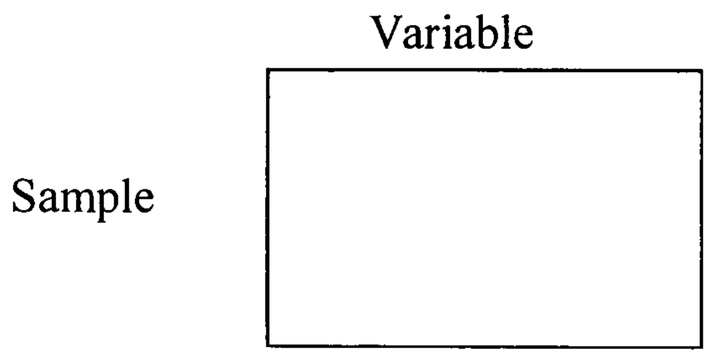


Figure 25: Frontal slice of **X** (IxJ)

This score vector explains the local variation at time point K_i. The block scores vector is placed in a consensus matrix $\mathbf{B}_{Ki}$, as is a 'super' scores vector $\mathbf{t}_{Ki-1}$, which describes the process variation up to time point (K-1). The scores, $\mathbf{b}_{Ki}$, are weighted using an adaptive parameter d. Basically, if d is large in magnitude, greater weight is given to the current process measurement in comparison to the previous measurements. If d is equal to zero, the block super scores are equivalent to the first super score. The larger the magnitude of d, the more variance in the data, is explained by the model. d typically varies between 0.3 and 1.0. With the newly updated $\mathbf{B}_{Ki}$, a new super scores vector, $\mathbf{t}_{Ki}$ is computed that represents the total variation in the process up to the current time point, K_i. This procedure is repeated throughout the duration of the batch. When using this technique for monitoring evolving batches, local models are built for each slice of the batch matrix $\mathbf{X}_{Ki}$ and the scores are continually rescaled to account for the variance that is captured by the different models. Figure 26 provides a summary of the hierarchical PCA algorithm.

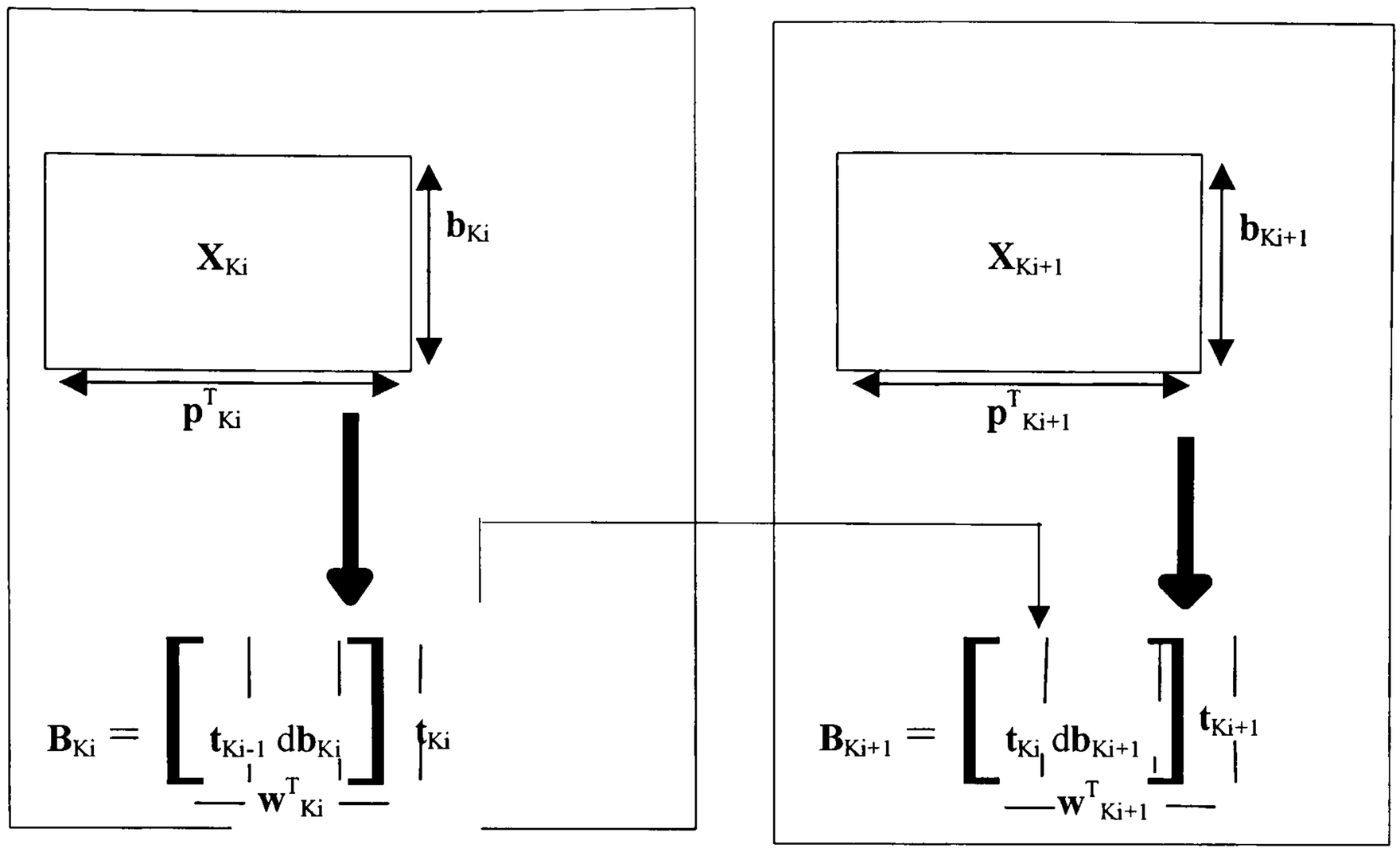


Figure 26 : Schematic of Hierarchical Principal Component Analysis

The advantage of using hierarchical PCA to monitor batch processes is that through using the adaptive parameter, d, the model can be adjusted to match different stages in the batch process. However there is no current method of tuning the parameter and thus d is selected based on experience. Also, as this technique uses local models, there is no requirement for in-filling as is required for the end of batch and batch observation multi-way techniques.

3.8.2 Moving Window PCA

A methodology with a similar objective to hierarchical PCA is that of moving window PCA. Again this technique aims to model the time variation in the data by computing a number of local PCA models throughout the duration of the batch. At any time during a batch, a PCA model is constructed over a moving window of time. The length of the moving window, K_d, is required to be selected and a matrix is then constructed containing the data from the K_d window length to the current time point, K_c, as shown in Figure 27.

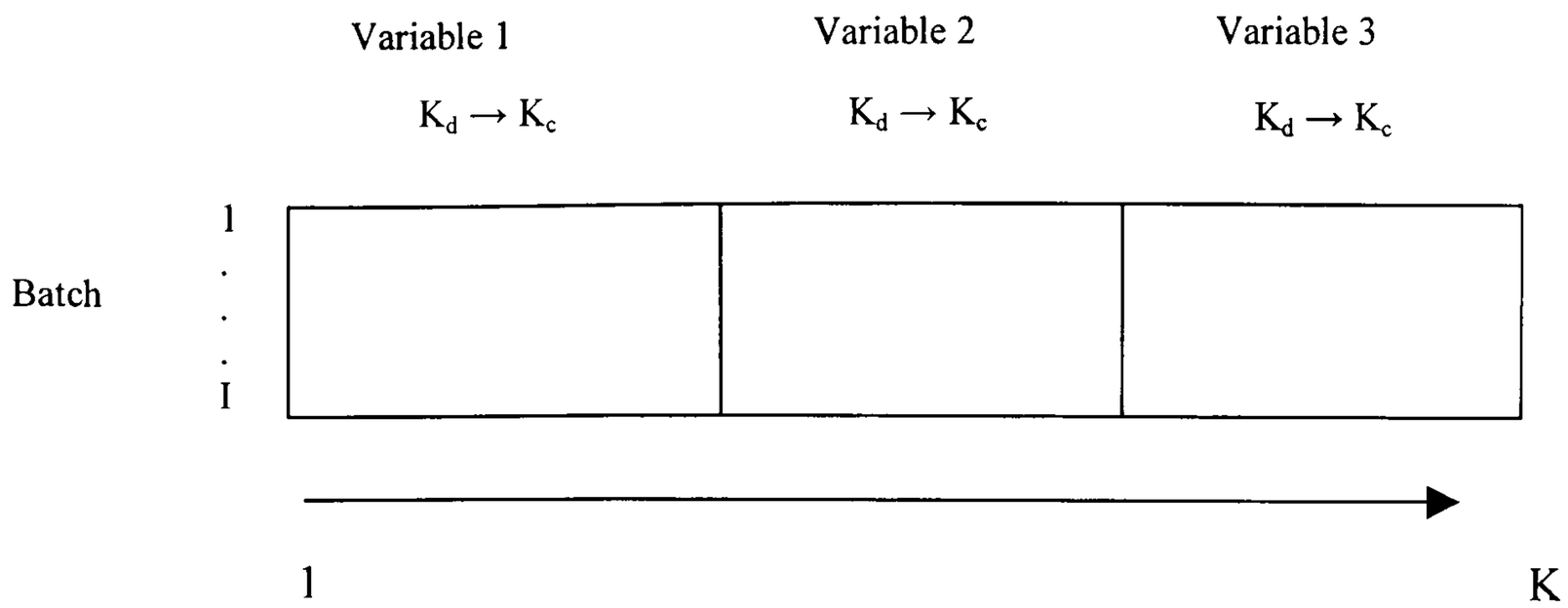


Figure 27 : Layout of Moving Window PCA

The moving window approach uses data unfolding method one and the matrix is then scaled to remove the mean trajectory. At each time point a PCA model is calculated utilising this matrix and then applied to the current batch. The main advantage of using the moving window technique is that the method handles non-linear behaviour since a series of local linear models are built as opposed to relying on a single global linear model. This is a much more efficient method for capturing the non-linear behaviour in the process. In addition, the technique incorporates lagged values of the process data in the model and consequently captures the dynamic behaviour inherent in a batch process. Additionally, moving window PCA has no requirement for equal length batches or for in-filling of the data since the monitoring statistics are only calculated using the current time window. Additionally, the technique does not need to use all the available batches for modelling, for example batches missing some data at a particular time can be omitted from the model during those time periods.

However, the main limitation of adaptive monitoring with hierarchical PCA, moving window PCA and related techniques is that they are computationally intensive. This feature can make them undesirable for the monitoring of faster industrial processes, where the emphasis is on rapid computation of the model to ensure problems are identified and corrective actions taken as quickly as possible. Again, increasing computing power will reduce the impact of this issue, however they do address the dynamic and non-linear behaviour.

3.9 Dynamic Projection Based Techniques

As has been stated previously, a defining feature of a batch process is its dynamic as well as non-linear behaviour. This is one aspect of batch behaviour that is not addressed by the standard multivariate monitoring techniques of multiway principal component analysis and partial least squares. Some enhancements to the methodologies for continuous processes have been developed. Ku *et al* (1995) incorporated auto-regressive structures into the fundamental techniques, more specifically the data are stacked with the current observation vector and the previous, δ, observations:

$$\mathbf{X}_{\delta} = \begin{bmatrix} (\mathbf{x}(K_i))^T & (\mathbf{x}(K_i - 1))^T & \dots & (\mathbf{x}(K_i - \delta))^T \\ (\mathbf{x}(K_i - 1))^T & (\mathbf{x}(K_i - 2))^T & \dots & (\mathbf{x}(K_i - \delta - 1))^T \\ . & . & . & . \\ . & . & . & . \\ (\mathbf{x}(K_i + \delta - K))^T & (\mathbf{x}(K_i + \delta - K - 1))^T & \dots & (\mathbf{x}_{Ii}(K_i - K))^T \end{bmatrix} \qquad (3.11)$$

where $\mathbf{x}(K_i) = [x_{1,Ki}\ x_{2,K}\ \dots\ x_{J,Ki}]^T$ is the J-dimensional observation vector at time point K_i.

Autoregressive batch dynamic MPCA (BDPCA) and MPLS (BDPLS), Chen *et al* (2001), were developed based on the assumption that the outputs of a batch process are conditioned on the process conditions and the variable trajectories of the specific batch. To achieve this, the array of process data $\underline{\mathbf{X}}$ used in the MPCA and MPLS calculations is unfolded to utilise time-delayed (sometimes called lagged in signal processing) values of previous process measurements. A BDPCA auto-regressive model structure for batch I_i at time point K_j can be represented as follows:

$$\mathbf{X}_{Ii}(K_j) = [\ (\mathbf{x}_{Ii}(K_j))^T\ (\mathbf{x}_{Ii}(K_j - 1))^T\ (\mathbf{x}_{Ii}(K_j - \delta))^T\ ] \qquad (3.12)$$

where δ is the window length of the dynamic process and:

$$\mathbf{X}_{Ii} = \begin{bmatrix} \mathbf{X}_{Ii}(\delta+1) \\ \mathbf{X}_{Ii}(\delta+2) \\ \cdot \\ \cdot \\ \cdot \\ \\ \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{X}_{Ii}(K) \end{bmatrix} = \begin{bmatrix} (\mathbf{x}_{Ii}(\delta+1))^T & (\mathbf{x}_{Ii}(\delta))^T & \ldots & (\mathbf{x}_{Ii}(1))^T \\ (\mathbf{x}_{Ii}(\delta+2))^T & (\mathbf{x}_{Ii}(\delta+1))^T & \ldots & (\mathbf{x}_{Ii}(2))^T \\ & \cdot & & \cdot \\ & \cdot & & \cdot \\ (\mathbf{x}_{Ii}(K))^T & (\mathbf{x}_{Ii}(K-1))^T & \ldots & (\mathbf{x}_{Ii}(K-\delta))^T \end{bmatrix} \quad (3.13)$$

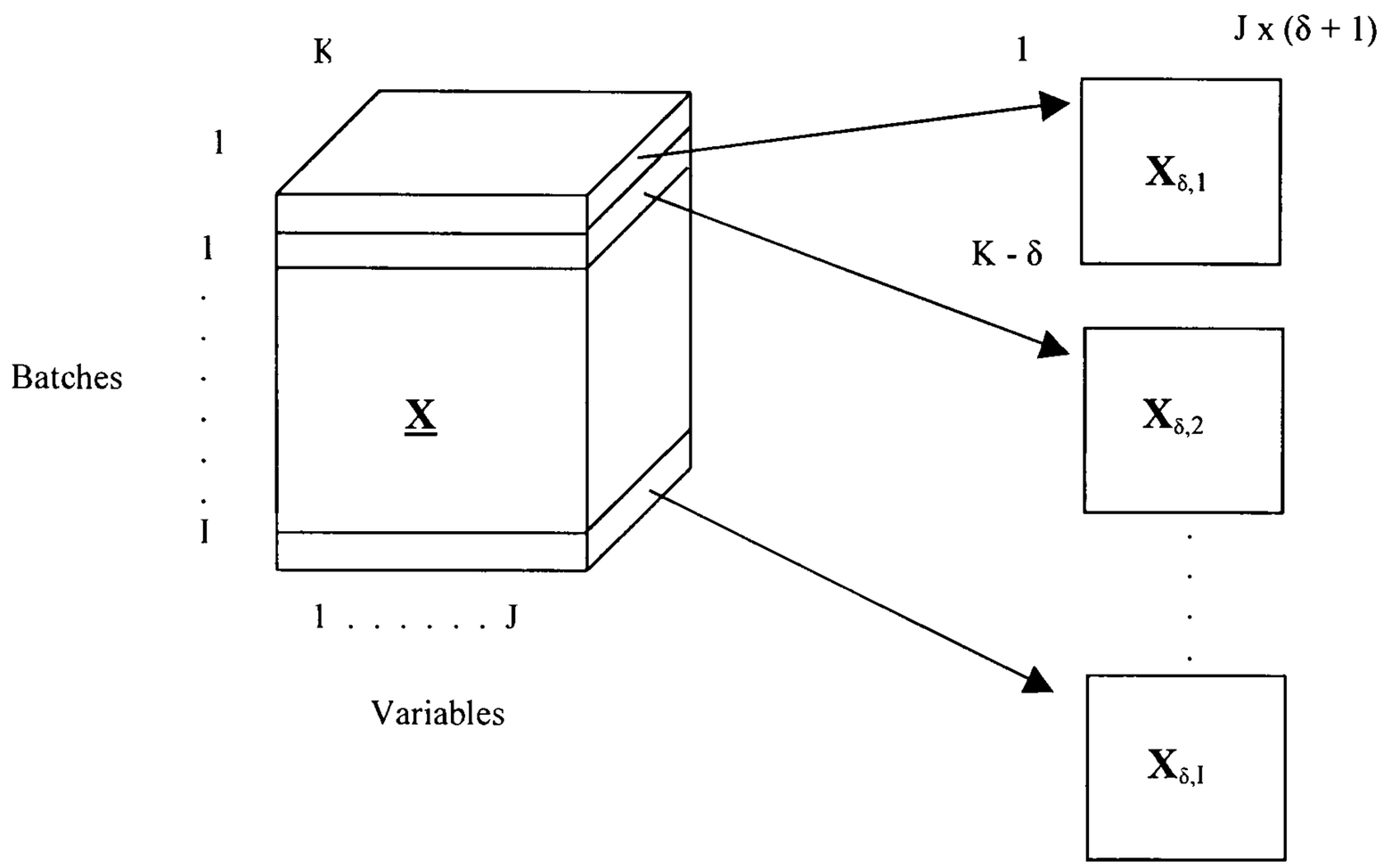


Figure 28 : Arrangement of a three-way array incorporating a time lagged window

Figure 28 shows the arrangement of the three-way array incorporating a time lagged window.

The matrix for each batch in the nominal data set can be built as follows. The lagged window, δ, can be set as defined by Ku *et al* (1995). A typical approach to determining the value of δ is to examine the autocorrelation or partial autocorrelation plots for the original variables thereby establishing the delay required for each variable. Depending on how similar the delay value is for each variable in the BDPCA model, it can then be structured using an average delay value calculated from all the variables, or alternatively each variable in the batch can be lagged separately. In the latter case, the number of samples for each variable can be reduced to the same minimum size after lagging has been performed (as only one or two observations would typically need to be removed

and this should not have any effect on the monitoring ability of the technique). The latter technique, i.e. variable delay, requires additional computational time but will give a better model fit if some of the variables differ significantly in terms of dynamic behaviour.

Once the lag matrix has been formulated for each batch, the individual covariance matrices, $\mathbf{S}_i$, are calculated and then pooled to form an average covariance matrix, $\mathbf{S}_{X\delta X\delta avg}$, Figure 29. This covariance matrix describes the average dynamic relationships between the variables. Utilising the average covariance matrix, a PCA model is calculated and control limits are derived for the monitoring of future batches.

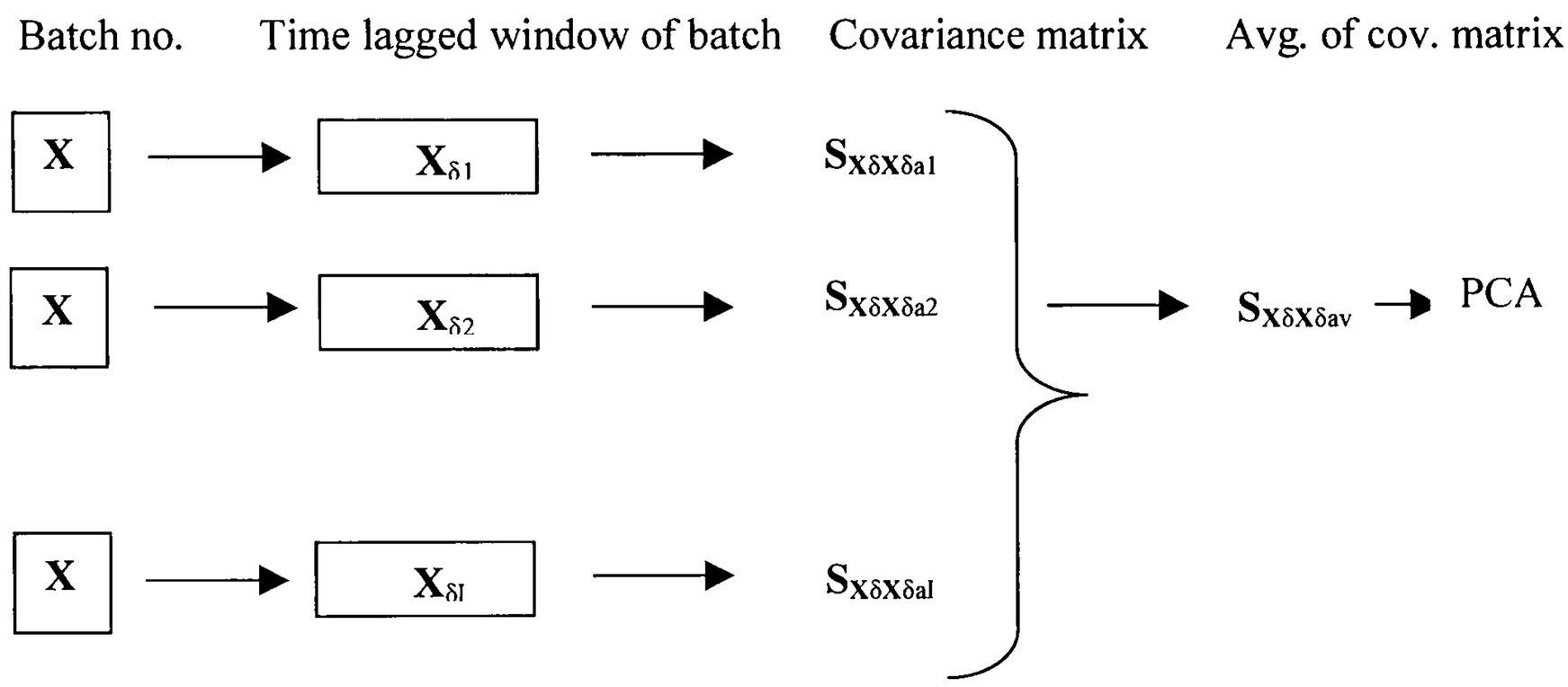


Figure 29 : Schematic of batch dynamic principal component analysis

This approach is also applicable for PLS, BDPLS. Consequently by modifying the multi-way PCA and PLS techniques to incorporate time-delayed process variables, the dynamic behaviour of the batches is factored into the model, providing a more appropriate monitoring tool when deployed on-line. In addition, the issue of the requirement to estimate the future behaviour of the batch is also removed as the BDPCA and BDPLS techniques are based on current and past data observations as opposed to future values.

In summary, the steps for modelling and then monitoring are as follows:

Modelling

- Collect normal historical batch data sets
- Select the number of time delays of all batches to extract the dynamic relationships and obtain the average time delay
- Construct time delayed data matrix
- Develop BDPCA model
- Compute the scores for all batches at each time point and set up control limits for each individual time point

Monitoring

- Record data from new batch run and establish time-delayed data window
- Project data onto retained principal components of BDPCA model
- Assess whether current data is within control limits
- Continue monitoring if answer is yes or investigate if no

However, BDPCA and BDPLS do not deal with the non-linear behaviour in the batch data since the lagged covariance matrix is calculated across the entire batch trajectory. This may mask the effect of an abnormality entering the process. Furthermore, it has also been claimed that by lagging the variables, the dynamics in the process are not adequately captured, Fletcher *et al* (2001). The reasoning behind this is that a batch process typically operates in several phases throughout its duration, and by calculating a single covariance matrix, the batch is treated as a single phase. This could be misleading since the interactions between variables often varies throughout the duration of a batch and a single covariance matrix only provides an average description of these variable interactions. Again this could mask any abnormal behaviour in the process. Also, since the BDPCA technique calculates a covariance matrix based on the time dimension, the method does not explain the batch to batch variation that occurs in the process, unlike end of batch MPCA (unfolding method one). Non-linear approaches to PCA are outlined in section 3.10. Research continues to progress in this area with a recent publication by Lu *et al* (2005) proposing the use of a 2D-DPCA model for on-line batch monitoring. In this method, the process data is augmented to include lagged measurements in both the time and batch directions. PCA is then applied to the 2-D augmented data matrix to capture the cross-correlations of the process variables and the autocorrelations in both the time and batch directions.

3.10 Non-linear Principal Component Analysis and Partial Least Squares

Of the batch monitoring techniques discussed so far, few have specifically addressed the issue of non-linearity in the data. There are typically two routes to dealing with the non-linear behaviour inherent in batch processes. The first approach is to transform the data using a non-linear function and then apply a linear regression technique to the transformed variables. This is essentially how model-based PCA (section 4.2) and multi-way PCA, for example, work. An alternative approach is to use a complex non-linear regression approach.

Although the multi-way PCA and PLS techniques (using unfolding method one and auto-scaling) go some way towards removing the non-linear behaviour from the data through the removal of the average trajectory from the individual variables, they are still basically linear techniques. This is a key issue associated with the application of these linear methods for batch modelling and then monitoring. This is related to the fact that these techniques may not be efficient in compressing non-linear data and therefore to capture the non-linear data lower order principal components require to be retained. However, as mentioned previously, minor principal components are often ignored in a multi-way analysis since it is assumed that they simply describe noise or negligible variance-covariance structures in the data whereas in fact they may contain important information on the non linear behaviour in the data. Consequently to ensure that the non-linear information is included in the analysis, a larger number of components would require to be retained, but this is difficult to asses if standard cross validation or other latent variable selection techniques are applied as it may require a priori knowledge that the process is non-linear. As described in section 3.8.2, the moving window and adaptive PCA methods attempt to tackle the issue of non-linear behaviour through the use of several local linear models to describe the batch behaviour over its duration. However, there is an alternative approach to these techniques that specifically addresses the non-linear nature of the data, that again is an extension of the PCA and PLS algorithms.

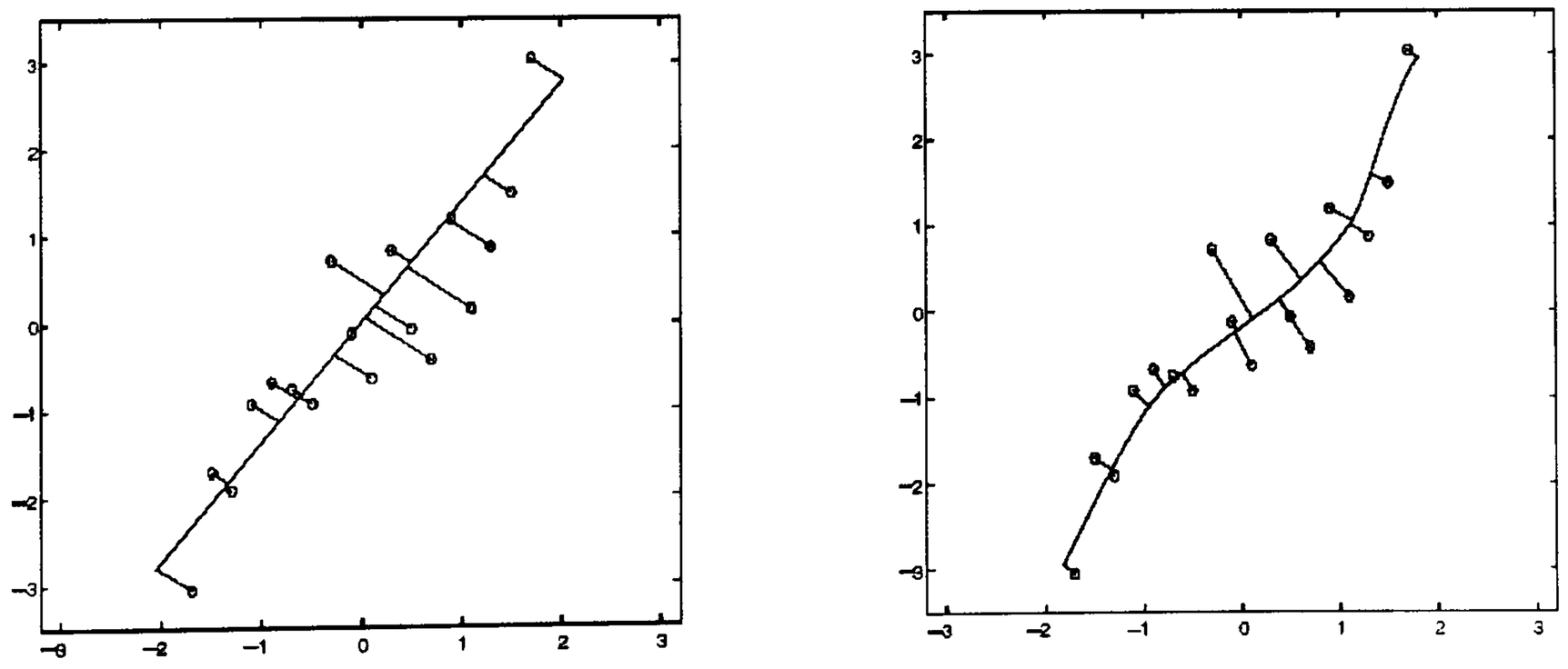
For the application of PCA to variables exhibiting non-linear behaviour, a non-linear mapping needs to be applied to map the original variables onto a reduced dimensional space. There are a number of alternative non-linear PCA approaches that have been applied to non-linear continuous processes, Jia *et al* (2000), as well as batch processes,

the most common being that of artificial neural networks. An alternate approach was that based on principal curves, Dong *et al* (1994).

For linear PCA, the sum of the orthogonal deviations between a straight line and the data is calculated and the first principal component is that line which minimises the error, whereas with non-linear PCA it is the deviations between the data and a smooth curve, termed a principal curve, that is used, Figure 30. These principal curves are generalisations of the principal components and are calculated in two stages, the projection stage, in which the data points are projected onto the curve, and secondly the smoothing stage where smoothing techniques are used to smooth the principal curve. This principal curve algorithm results in a table that contains the original process data, the non-linear scores and the reconstructed data from the non-linear principal components. The next step in the algorithm is to calculate the non linear loadings. This is achieved through the application of an artificial neural network, created using two, three layer neural networks, which identifies a non-linear function that approximates the curve:

$$\mathbf{X} = F(\mathbf{T}) + \mathbf{E} \qquad (3.14)$$

where **T** represents the principal component scores, F denotes the non-linear PCA loading function and **E** defines the residuals.



Standard Non-linear

Figure 30 : a) Linear Principal Component b) Non-linear Principal Component

Generally successful results have been achieved using this technique to monitor batch processes, however, since NLPCA is also performed on the unfolded batch matrix, as for end of batch MPCA, the issues with MPCA are the same as for NLPCA when the technique is applied on-line. Mainly this relates to the fact that the model requires the

entire batch trajectory for the monitoring of the process so an in-filling method has to be used to carry out on-line analysis. However, as the technique does take into account the process non-linear behaviour, better monitoring performance is generally achieved, as demonstrated by Jia *et al* (2000), where non-linear PCA was applied to an industrial fluidized bed reactor exhibiting highly non-linear behaviour, giving improved performance over the application of a linear PCA approach. Another problem with non-linear PCA is that it is more difficult to interpret than the linear version. With non-linear PCA, the movement of scores and therefore fault detection diagnosis can become more complex. It has been proposed that to combat this problem, a cumulative scores plot is used, where the accumulated scores increase the further away from the origin the score moves, Zhang *et al* (1996).

A non-linear PLS approach, has been developed whereby the non-linear features are incorporated into the framework of linear PLS to give a non-linear algorithm. This has been achieved for continuous processes using a polynomial expansion, Wold *et al* (1989) and Baffi *et al* (1999), linear smoothing, Frank *et al* (1990), spline function, Wold *et al* (1992), or neural network function, Qin *et al* (1992) and Baffi *et al* (1999), (2000).

Neural networks can be affected by the correlations between variables. More specifically, in situations where there is high correlation between the variables, as in batch processes, neural networks have the undesirable effect of increasing the variance of the noise in the predictions.

It has been demonstrated that a more stable modelling tool is achieved if a neural net is incorporated within the linear PLS framework with weight updating in the PLS input outer models, Baffi *et al* (1999), or with both inner and outer weight updating procedures, Baffi *et al* (2000). With sigmoidal feed forward neural networks, a set of networks is generated, one for each pair of latent variables, to identify the inner regression model. This is a generic approach to non-linear PLS modelling since with a neural network, no functional relationship is required to be assumed a priori. Again the technique requires the adaptation of the error-based updating procedure to recalculate the outer input weights, otherwise the technique is not suitable for data where the variable relationships are highly non-linear.

Another type of neural network that can be used for non-linear PLS is that of the radial basis function (RBF), Irwin *et al* (1995). This technique generates a set of RBF networks with Gaussian activation functions, again with one for each pair of latent variables, to

model the relationship between the inputs and outputs. The RBF technique is considered superior to the sigmoidal method because its training networks are simpler and faster. Overall the neural network PLS algorithms give improved performance over the quadratic non-linear PLS due to neural networks' universal approximators, Baffi *et al* (2000).

In all cases once a NLPLS model is built, new data can be treated and then passed through the algorithm to predict, for example, composition measurements on-line, for continuous processes. Control charts can be used to detect abnormal operation such that corrective action can be taken. Finally the application of NLPLS fault detection to batch processes is a new research area covered in this thesis, in Chapter 7, through the use of the technique in tandem with super model based techniques.

3.11 Batch Monitoring using State Space Models

An alternative approach to employing a non-linear algorithm to compensate for the non-linear behaviour in a batch process is to use a state space model, as suggested by Negiz *et al* (1997), Lee *et al* (2004), Simoglou *et al* (2002). The technique proposed by Simoglou *et al* (2002) is ideal for monitoring batch processes as it builds several local state space models over the duration of the batch. To construct these models, the system states are identified at each time interval and a least squares solution is used to calculate the state space model matrices:

$$\mathbf{t}_{t+1} = \mathbf{C}_t\mathbf{t}_t + \mathbf{w}_t \tag{3.15}$$

$$\mathbf{y}_t = \mathbf{H}_t\mathbf{t}_t + \mathbf{e}_t \tag{3.16}$$

where **t** represents the system states, **y** is the process measurements, **w** and **e** are the state residuals and the output residuals with covariance matrices **Q** and **R** respectively. **C** and **H** are the state space model matrices. To calculate the system states, partial least squares, Martin *et al* (1996), principal component regression, Lakshminarayanan *et al* (2000), or canonical variate analysis, Shaper *et al* (1994), can, for example, be used. These techniques use linear combinations of past measurements to best predict the future process measurements. When setting up the model, the batch data matrix is unfolded and scaled as for multi-way PCA and the future and past delays are then configured with the system states identified using either PLS, PCR or CVA. The time varying state space

model is then calculated using a least squares approach. A crucial step in the algorithm is the selection of the number of past and future delays. Since the advantage of the state space technique is that at each time interval a dynamic model is created that describes the relationship between the past and future measurements, it is essential that the model order is carefully selected to ensure that it adequately describes process behaviour. Additionally the number of delays for the future and the past will not necessarily be the same and therefore need to be identified separately. The suggested method is that of Akaike's Information Criterion (AIC, 1976).

There are a number of advantages to the use of time varying state space techniques. Firstly equal batch lengths are not a requirement, and also that there is no requirement to in-fill data or estimate the future behaviour of a batch, all of which are problems associated with the end of batch multi-way techniques. In addition to this, the main advantages with the method are in dealing with the non-linear and dynamic structures in the batch data. By building several local linear models as opposed to a single model, the non-linear behaviour in the process can be accounted for, and the dynamic state space model that is created at each time interval describes the dynamic relationships in the data and therefore the non-steady-state behaviour of the batches is captured. Another advantage of this technique is that it does not assume that the number of latent variables is constant throughout the duration of the batch since it is recalculated for each time interval. This is beneficial since the non-steady state nature of batch processes means that they operate in a number of different phases, each of which may require a different number of latent variables to adequately monitor the process.

As with many of the other techniques, the drawback of using a method such as this is the large amount of computational time required to generate a large number of local models over the batch duration. Having a large number of models to build and maintain can make the monitoring method less likely to be adopted in practice.

3.12 Conclusion

A number of methodologies have been proposed for the monitoring of batch processes and a selection of these have been summarised in this chapter, including multi-way techniques, dynamic methodologies, non-linear PCA and PLS and model-based analysis, the latter of which is the focus of this research.

Of the methods considered, the end of batch multi-way technique is one of the most commonly applied in multivariate batch performance monitoring. However this method does not fully tackle the non-steady state and non-linear behaviour of the batch process as it is essentially a linear technique. Furthermore it is more applicable as an end of batch analysis techniques as opposed to a through batch monitoring methodology such as batch observation level analysis (BOL). Likewise, BOL does not address the dynamic and non-linear behaviour inherent within a process. The adaptive PCA and Moving Window methods do address some of the challenges associated with batch monitoring as they use several local linear models to describe the non-linear behaviour that materialises over the whole process. By utilising several local models along with time-delayed variables, the dynamics in the data can also be considered to some extent. The other method for dealing with the non-linear behaviour in the data is to use a multivariate technique with a non-linear algorithm. This approach has been shown to produce some good results but does not completely address non linear behaviour and also interpretability of the resulting model is limited.

In Chapter 4 a number of batch monitoring techniques that take into account additional information about the process are discussed, and then in Chapter 5 the more common of the batch monitoring techniques discussed in Chapters 3 and 4 are applied to the monitoring of an exothermic batch reactor simulation. The goal is to assess the effectiveness of each technique in an attempt to determine which of the methods is the most appropriate as a batch monitoring tool.

4 CHAPTER FOUR: PROCESS SPECIFIC BATCH MONITORING TECHNIQUES

4.1 Introduction

Before the introduction of multivariate statistical techniques for batch performance monitoring, one of the most common ways to monitor the behaviour of a batch process, other than using univariate methods, was to develop a mechanistic/first principles model of the process in question and use this to determine if the process is behaving according to plan. However there are a number of problems associated with this technique, primarily associated with the amount of time required to generate a first principles model that accurately describes the physical and chemical phenomena occurring during a batch process. Also it is possible that there are some parts of the process that cannot be comprehensively described such as the cell growth kinetics. However, many practitioners find the idea of using a monitoring tool that is based on the actual chemical relationships occurring in the process, and therefore incorporating real process understanding, appealing, particularly in comparison to many of the techniques discussed in Chapter 3, which were empirical, data-based techniques, often called 'black-box' approaches. Whilst these techniques provide an excellent tool for batch performance monitoring, one limitation associated with this type of approach is that the interpretation of the underlying physical and chemical behaviour of the variable relationships/process abnormalities becomes difficult, and therefore reduces the potential to understand the underlying process behaviour.

This chapter looks at techniques which combine additional process specific information, such as mechanistic models, with an empirical model to produce a modelling/monitoring tool. This makes efficient use of all the available information without creating a modelling process that is too time-consuming and therefore expensive. The combination of additional information through mechanistic understanding of the model within an empirical batch monitoring tool can be useful in comparison to using a solely empirical approach in terms of improving control, performance monitoring and fault diagnosis as well as enhancing process understanding.

4.2 Model-based Principal Component Analysis

4.2.1 Overview

Model-based principal component analysis, Wachs *et al* (1998), is one of a number of ways in which physical and chemical information about a batch or continuous process can be incorporated into a performance monitoring tool. The approach combines a reduced complexity mechanistic model of the process with a multivariate statistical technique to give a reliable monitoring tool.

This type of mechanistic model-based approach is ideal for application to batch processes as it allows a linear PCA procedure to be applied to a non-linear, dynamic process. The idea behind model-based PCA is to develop a mechanistic model of the process and then perform a multivariate statistical analysis on the portion of the data that is not explained by the mechanistic model. Therefore in the situation where the model is perfect any data that is not predicted by the model will simply be white noise. Consequently when an abnormality enters the process it will not be described by the model and will therefore manifest itself in the white noise residuals and hence it will be detected by the standard PCA monitoring methods.

To apply MBPCA, a first principles model of the process is initially generated. This generally comprises a set of ordinary differential equations that describe the mass and energy balances, for example, involved in the reaction. A matrix of batch data is then collected from the process in question under normal operating conditions, $\mathbf{Y}$, and the model is computed at the process variable sampling rate to give a set of model predicted variable measurements, $\mathbf{Y}_T$. The model predicted measurements are subtracted from the actual process measurements to give a set of residuals ($\mathbf{E} = \mathbf{Y} - \mathbf{Y}_T$). Auto scaling is then performed on the residuals prior to the application of PCA. Since the mechanistic model should capture the dynamic and non-linear behaviour present in the process, by subtracting the model from the measured data, these characteristics should be removed from the data. The result is therefore a set of linear, steady-state residuals which are suitable for monitoring using a standard PCA approach. A schematic for the model-based PCA technique is shown in Figure 31.

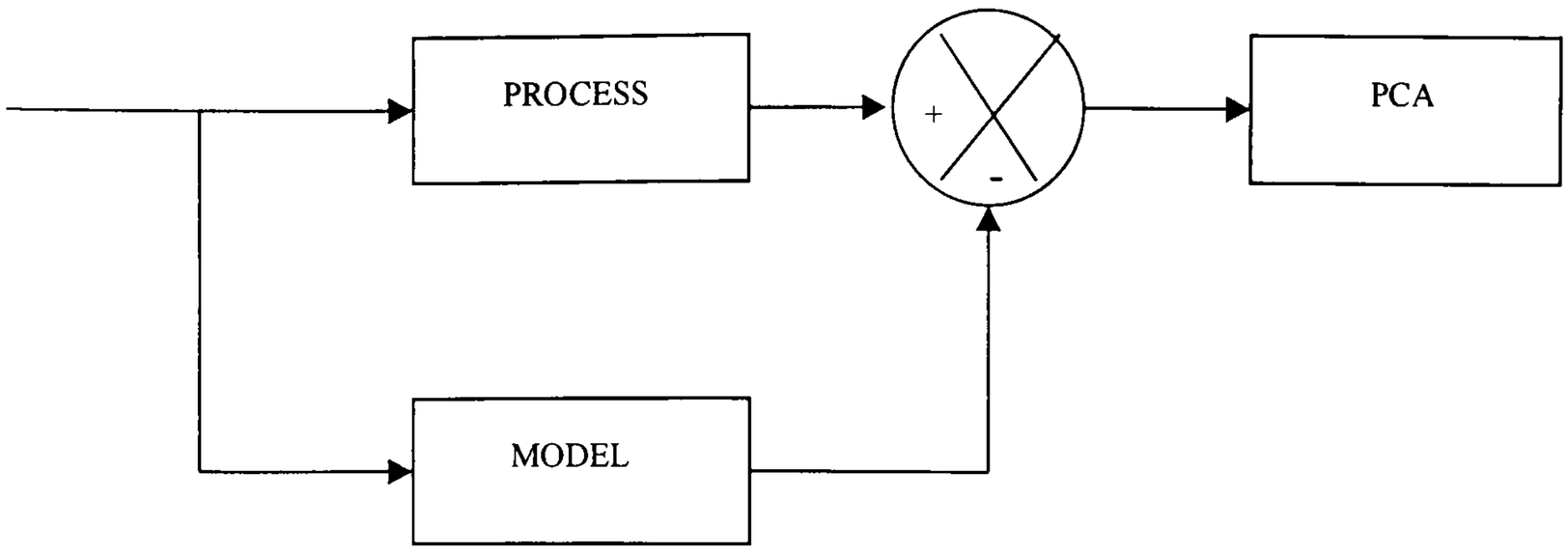


Figure 31 : Schematic of model-based PCA

4.2.2 Limitations of Model-based PCA

Model-based PCA has been applied successfully to a number of different fault detection situations, including the batch reactor simulation that formed the basis of the case study in Chapter 6, the Tennessee Eastman simulation, Wachs *et al* (1998), and an industrial ethylene compressor Rotem *et* al (2000). In all three situations, Wachs *et al* (1998) hypothesised that the main drawback to applying model-based PCA was that its performance was susceptible to model uncertainties. Therefore it was recommended that a fault detection factor was considered to predict the potential fault detection ability of a given model. This fault detection factor, F_{MB}, is a method for pre-diagnosing the fault detection capability of the model-based algorithm for a particular fault, and to determine its sensitivity to uncertainty. The calculation of the factor, F_{MB}, is based on the ratio between the mean shift in the scores from the normal operating conditions (NOC) to the control limit (either 95 or 99%) ($\Delta st_{d,j}$) and the range of variability of the scores within the NOC (Δst_j). The summed scores, st, are defined as:

$$st_{i,j} = t_{i,j} + t_{i-1,j} + \ldots + t_{i-s+1,j} \qquad (j = 1\ldots m,\ i = 1\ldots n) \qquad (4.1)$$

where m represents the variables, n the observations, s represents the number of recursively summed scores and $t_{i,j}$ are the scores. The mean shifts are defined for each principal component as:

$$(\Delta st_j) = 0.5(st_{j,max} - st_{j,min}) \quad (4.2)$$

$$(\Delta st_{d,j}) = st_{d,j} - st_{j,NOC} \quad (4.3)$$

The F_{MB} ratio in equation 4.4 is computed for all principal components retained, and the fault detection factor, F_{MB}, is calculated as the Euclidean norm of the scaled perturbations of the scores, where $st_{j,max}$ and $st_{j,min}$ are the maximum and minimum scores:

$$F_{MB} = \sqrt{\sum_{j=1}^{r} (\Delta st_{dj} / \Delta st_j)^2} \quad (4.4)$$

A value of unity or greater indicates that the model has good fault detection potential, whilst values close to zero indicate poor performance of the algorithm. The technique has proved successful for the determination of the model capabilities in the three simulation studies reported by Wachs *et al* (1998) and Rotem, *et al* (2000).

The key to using model-based PCA is that the process being monitored must be able to be described using a first principles model. However, for more complex reactions, it may be better to use a different monitoring tool as the first principles model may be too complicated to create easily. In some industrial cases, these phenomenological models can be easily derived and may already exist for control purposes or simulation analyses. If not, a large number of man hours may be required to generate a model that adequately describes the process behaviour and, as mentioned before, it can be a difficult task to create a model that perfectly describes the many chemical reactions occurring. With a perfect model, the detection of any process abnormalities in the residuals is straightforward, however if the model is of reduced complexity, the less likely it is to remove the non-linear behaviour and dynamics inherent in the batch data, meaning that the monitoring of the residuals using a standard PCA technique will not provide the level of performance required. Depending on the accuracy of the mechanistic model, MBPCA can be an excellent tool for performance monitoring since it attempts to deal with both the non-linear and dynamic structures in the data, as well as not requiring data on the complete batch duration to carry out the analysis on-line and it does not require batches of equal duration. The main limitation of the method is that the performance of the

model-based tool is dependent on plant-model mismatch. Improvements to the MBPCA technique are proposed in Chapter 7.

Another issue identified by Wachs *et al* (1998) with respect to model-based PCA is that the less accurate the model is, the greater the level of autocorrelation present in the residuals, making them less efficient with respect to the monitoring of the process. Thus there is a problem if a perfect model of the process is not available. Again this is addressed using the super model-based PCA approach discussed in Chapter 7 since it attempts to model the mismatch between the model and the plant.

In addition to applying the fault detection factor, Wachs *et al* (1998) carried out a study into the robustness of the model-based technique with respect to the modelling of the uncertainty by applying different levels of uncertainty to the parameters in the first principles model and then investigating the average run length with respect to fault detection. The results showed that some types of fault were affected more than others with respect to the uncertainty in the parameters, however the conclusions could not be generalised since they were significantly affected by the amount of noise present in the data. A similar study, whereby the effect of a poor model on the efficiency of the monitoring technique is studied, is carried out using the super model-based techniques along with a study into the effect of noisy data. The results of these studies are discussed in Chapter 8.

4.2.3 Data Reconciliation

Data reconciliation, Amand *et al* (2001), is another technique that utilises process specific information to improve the performance monitoring ability of an empirical modelling technique. The objective of this method is to remove any errors from the process measurements thereby giving true estimates of all the process state variables and the unmeasured process parameters. This technique has been applied by Amand *et al* (2001), prior to the determination of the projection matrix for principal component analysis for fault detection. To apply data reconciliation, a mechanistic model of the process is developed as a set of constraints, for example mass and energy balances. The measured data, which contains random and systematic errors, is then modified to conform to the mass and energy balances of the mechanistic model i.e. so it does not violate any of the constraints. A least squares method is then applied to minimise the

adjustments and the measurement covariance matrix is used as a weight matrix to correct the measurements.

A limitation of basic data reconciliation is that it is not particularly effective for dealing with process faults that have a significant impact on the process measurements. This is because a gross error will effectively change the behaviour of the process, i.e. the original mass and energy balances will no longer be valid. Therefore although the technique can detect failing instrumentation, it will not identify a major abnormality entering the process. This is because data reconciliation will modify process measurements that are already correct, to fit the model, not realising the problem lies in the mismatch between the model and the plant and hence the fault is not detected.

The application of data reconciliation ensures the data conforms to the mass and energy balance constraints of the process, consequently the linear combinations captured in the PCA model using the reconciled variables are more closely representative of the actual process. This approach is an improvement over standard empirical models such as PCA since it takes into account the true physical relationships/reactions occurring in the process. Through the inclusion of additional mechanistic information about a batch process, the resulting model describes more accurately the process behaviour and provides greater insight into process behaviour. For example a mass balance might be denoted by:

$$F_1 + F_2 = F_3 \tag{4.5}$$

where F_1 and F_2 are inlet stream mass flows and F_3 is the outlet mass flow. A statistical analysis of the noisy measurements may materialise in the model:

$$a.F_1 + b.F_2 = F_3 \tag{4.6}$$

where a and b are coefficients that are not equal to unity. Applying data reconciliation to pre-process the data before the PCA algorithm will ensure that the principal components represent the actual process behaviour, as opposed to the variations in the data introduced by noisy signals.

4.3 Model Identification

A number of model identification techniques utilise first principles knowledge of a process to improve their performance. For example in a technique, similar to model-based PCA, process measurements are first filtered using a mechanistic model to remove the effects of the known dynamics, Love (2007), resulting in a residual data set that can then be used to model empirically the process dynamics not captured by the mechanistic model. It is important to model the dynamics not accounted for by the process model in order to describe the behaviour of the whole process. By focusing the analysis on the dynamics not captured by the mechanistic model, increased insight into the process can be achieved, as well as a reduction in model uncertainties and sensitivity to noise. The technique is applied in two stages, first the mechanistic model is applied to the data. The resulting residuals can then be modelled using Canonical Variate Analysis (CVA) as the identification method, for example, Kemna *et al* (1995). They successfully applied the methodology to a stirred tank heating process and a recovery boiler.

4.4 Neural Network Hybrid Techniques

A raft of hybrid/grey box modelling and monitoring techniques have been proposed in the literature prior to model-based PCA being proposed. These methods have included techniques involving neural networks and fuzzy logic. However, as for all empirical based approaches, implementing a neural network in isolation may not reflect the physical reality of the process and may give predictions that violate the fundamental process constraints. Also this family of models is only valid over the training range of the data and therefore may not be applicable to the different phases in which a batch process may potentially operate. In addition, the application of a neural network to capture the non-linear and dynamic structures in a process requires the estimation of a large number of parameters. Consequently this necessitates a large number of samples and significant training time and, if not implemented appropriately, the model may be over fitted.

The limitations of mechanistic models has been discussed in previous chapters, with respect to the fact that a comprehensive physical model of a process is unlikely to be available and is expensive to develop, although the increasing need for operator training simulators may alleviate this problem. Therefore to address the limitation of these two

types of modelling methods, grey box or hybrid modelling techniques have been developed to make use of all the process information available, Psichogios *et al* (2004). Typically these methods combine a reduced complexity model of the process with a neural network to capture any remaining unknown dynamics, Lachman-Shalem *et al* (2001), to improve the process fault detection ability of the model.

Xiong *et al* (2002) developed a grey box technique for application in generic model control, where a process model was embedded in the controller. This technique has potential application for process fault detection. Xiong *et al* (2002) considered two case studies, a simulated exothermic batch reactor and a real time continuous stirred tank reactor, and used a first principles model to derive the basic model structure, allowing the variable interactions to be specified from physical knowledge. They then implemented a neural network to model the non-linear characteristics of the process of interest. It was proposed that the neural network could be integrated with the mechanistic model either serially or in parallel, Figure 32. A series approach is considered when some unknown parameters, **p**, are required to be estimated and **u** and **y** are the input and output variables respectively. If a comprehensive first principles model description of the process is not available then a parallel approach is required, for example in the case of a reactor system, where the heating and cooling system may be well understood but the kinetics may not have been modelled.

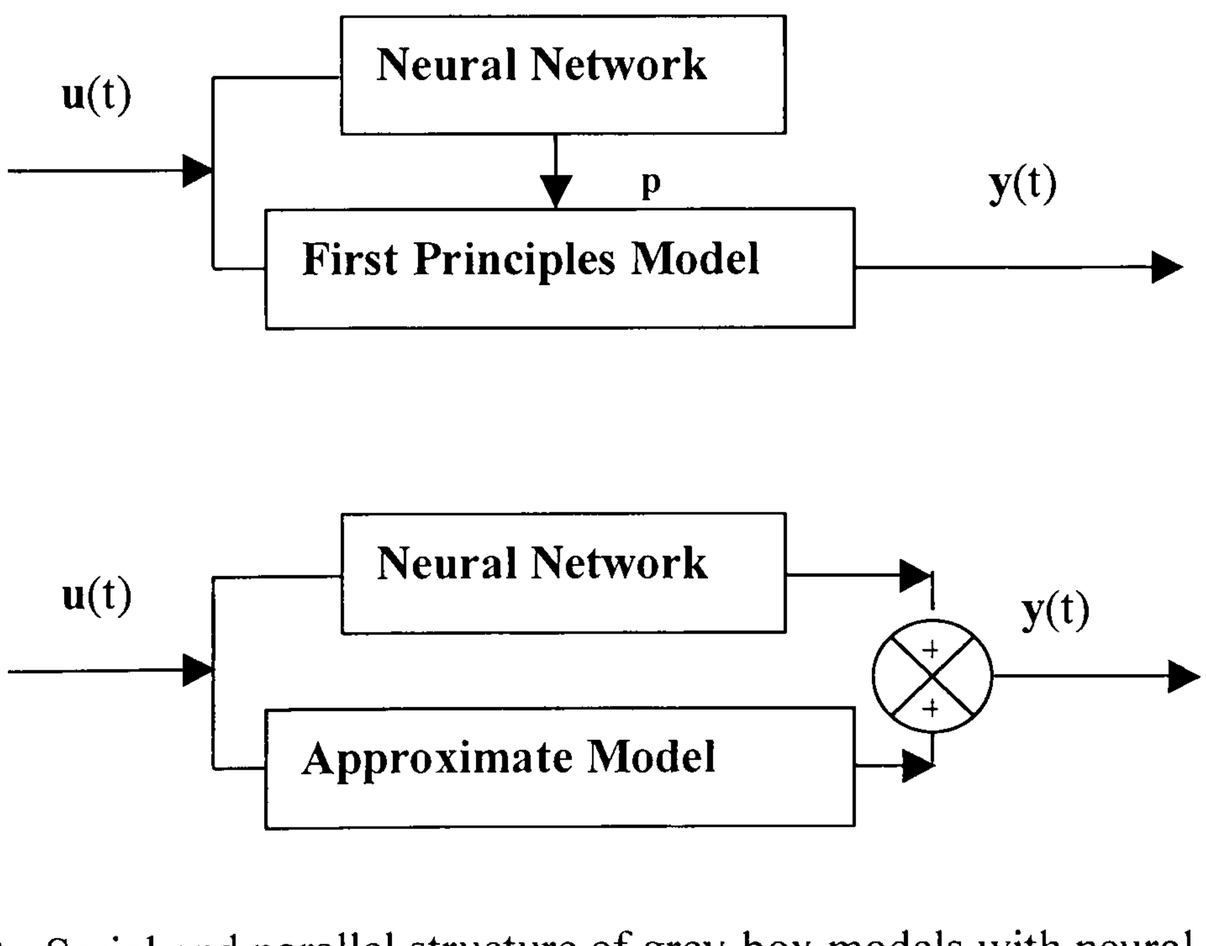


Figure 32 : Serial and parallel structure of grey-box models with neural networks

Xiong *et al* (2002) adopted a parallel approach and the technique was applied successfully in conjunction with a generic model control structure to the two simulations, i.e. the simulated exothermic batch reactor and the continuous stirred tank.

With the parallel approach, the output of the grey box model is the summation of the two separate model outputs. This arrangement allows both the model mismatch and the process disturbances to be captured. It was stated that the advantages of using this grey approach are that it generates a model that is more easily trained and updated than a 'pure' neural network and is also superior for prediction purposes. However one limitation is that hybrid neural networks require significant computational time.

Feyo de Azevedo *et al* (1997) also investigated the hybrid combination of a mechanistic model with a neural network by carrying out a study on the modelling of a biochemical process, the fed-batch production of baker's yeast. They compared the results of the hybrid technique with a mechanistic model and a neural network. Both the mechanistic model and the neural network gave unreliable predictions of the experimental data, with the hybrid model out-performing both approaches in terms of the prediction of batch experimental and training data. The main drawback associated with the technique is the computational time that was required for the training of the hybrid network was greater than for the individual techniques in isolation.

Molga *et al* (1999) employed a hybrid first principles neural network to model a liquid-liquid reacting system, again using first principles knowledge to produce mass and heat balance equations. They then applied a neural network approach to describe the unknown kinetic expressions in the system. This approach worked well as it required significantly less experimental effort and produced a flexible prediction tool (i.e. it required just five experimental data sets for training,). Model-based neural networks were also proposed by Fontaine *et al* (2001). They used first principles models in the learning phase of the neural network to produce a model of an isothermal reaction and showed that it outperformed the model produced using a pure neural network.

The studies of Zorzetto *et al* (2000) into grey box modelling involved introducing different levels of first principles knowledge into a neural network in an attempt to produce a model of a batch beer production process that captured the process mechanism. To fully model a bioprocess, there are typically three different types of equation required: mass and energy balances, rate equations and equations relating the rate parameters to the reaction temperature (Arrhenius-type equations). Since developing a rigorous model involving all three equation types would be time-consuming, they proposed the use of a hybrid technique. To assess if it was viable to model the bioprocess using a hybrid technique, they compared a model built using a pure black box neural network, a model built using the mass and energy balances in their differential

form with a neural network to describe the microbial growth rate equations and, finally, a model where the mass and energy balances and Monod type growth rate equations were used in their traditional forms and the neural network was only used to model the dependence between the rate equation parameters and the temperature, the third type of equation required. The sum of squared error between the model outputs and the experimental data was used as a performance index in this study. The results showed that the first type of hybrid model, using the mass and energy balances along with the neural network, provided the most accurate type of model. Including additional equations such as the Monod growth rates did not improve the performance of the hybrid model, however both types of hybrid model were a significant improvement over the black box approach.

4.5 Diagnostic Model Processor

A hybrid diagnostic technique, diagnostic model processor (DMP) was proposed in a study by Rengaswamy *et al* (2000). They compared the methodology to an ellipsoidal neural network technique specifically with respect to its fault diagnosis ability. The DMP method is a quantitative model-based diagnosis technique that is based on a series of first principles model equations. Associated with these equations is a set of underlying assumptions that serve as the a priori knowledge about the process. In contrast, the ellipsoidal neural network technique utilises historical data as its prior knowledge. In the diagnostic model processor, as with model-based PCA, the residuals generated from the model equations are the basis of the approach. During normal operating conditions the assumption is made that the residuals will exhibit white noise behaviour. For each model equation, there is a set of underlying assumptions about the faults. These represent the assumptions on which the associated model equation is based. If the assumptions for a particular model equation holds, the equation will be satisfied and the residuals will be approximately equal to zero. If the assumptions are not satisfied then the residuals will not be zero and the onset of a fault will have been detected. By examining the direction and the extent to which the model equation is violated, the most likely fault type can be estimated. An ultra high temperature sterilisation process was used to compare the performance of DMP with the data-based ellipsoidal neural net technique. Both methods performed well, however in some cases the DMP approach provided more than one suggestion for the cause of a fault. However it is possible that this can be seen as an advantage as the user can then apply their own knowledge of the process to determine the source of the problem from the given options.

An additional limitation of DMP is that it can be affected by the amount of noise in the data since its performance is critically dependent on the accurate tuning of a number of threshold parameters. This can be a difficult issue if there is a low signal to noise ratio present in the data which may be the case with a batch process. Furthermore, as a fault diagnosis method, DMP relies on the fault having already been modelled, therefore any new process abnormalities may either go undetected or be misclassified, which could have severe consequences. As mentioned before, this technique relies on an accurate first principles model of the process being available, and this is not always the case. With respect to the ellipsoidal neural networks, the methodology is limited to fault diagnosis over the data range in which the network was trained. Additionally, the ellipsoidal neural network does not give causal explanations as to the fault decision. However the technique is more robust with respect to its sensitivity to noise and the diagnosis of new faults. It was concluded from this study that an additional hybrid neural network technique should be developed which combined the advantages of both methods.

4.6 Fuzzy Logic

An alternative to using an artificial neural network was proposed by van Lith *et al* (2002). In their work they proposed the use of fuzzy logic as part of their hybrid modelling technique. In this case, a model of a fed batch penicillin fermentation, based on a framework of dynamic mass and energy balances, was enhanced with algebraic and fuzzy equations formulated in the state-space form to describe the un-modelled dynamics such as mass transformations and transfer rates. In this case, fuzzy models were developed for the non-linear and time dependent penicillin growth rate and the product formation rate where the relationships were unknown but the structural dependencies were (substrate and biomass concentrations). This resulted in a dynamic process model with a relatively high level of interpretability.

The advantage of hybrid modelling is that the knowledge and information available about a process can be combined to produce a realistic model, as opposed to producing either a first principles model or an empirical representation. This is especially true when using fuzzy logic since human experience of the process can be used to design the fuzzy relationships/structures and therefore capture operator experience about the dependencies of relevant phenomena occurring in the process. This type of hybrid fuzzy

first principles modelling is appropriate for use on a batch process, or any non-linear process covering a wide operating domain. This is because the methodology uses several fuzzy sub-models hence the problem is reduced to a number of linear problems, making it easier to model the dynamic behaviour of the process during the different stages, and also making it simpler to isolate and eliminate undesirable behaviour in one sub-model.

As with other hybrid techniques, the hybrid fuzzy method can be applied either serially or in parallel, (Figure 33). With the serial approach, the fuzzy logic sub-models are used to calculate the model variables (**p**) that are required by the physical part of the mechanistic model. In the parallel approach, where a major part of the system is represented by a fuzzy model, the two types of model are generated separately, and the outputs of the fuzzy logic block and the physical model are combined to produce the overall model. As previously, **u** and **y** are the input and output variables respectively. If the use of a first principles models is possible and favoured over black box models, the preferred approach is to use the serial semi-parametric approach since in this way the mechanistic model structure is left intact and only certain unknown phenomena are modelled using the fuzzy sub-models.

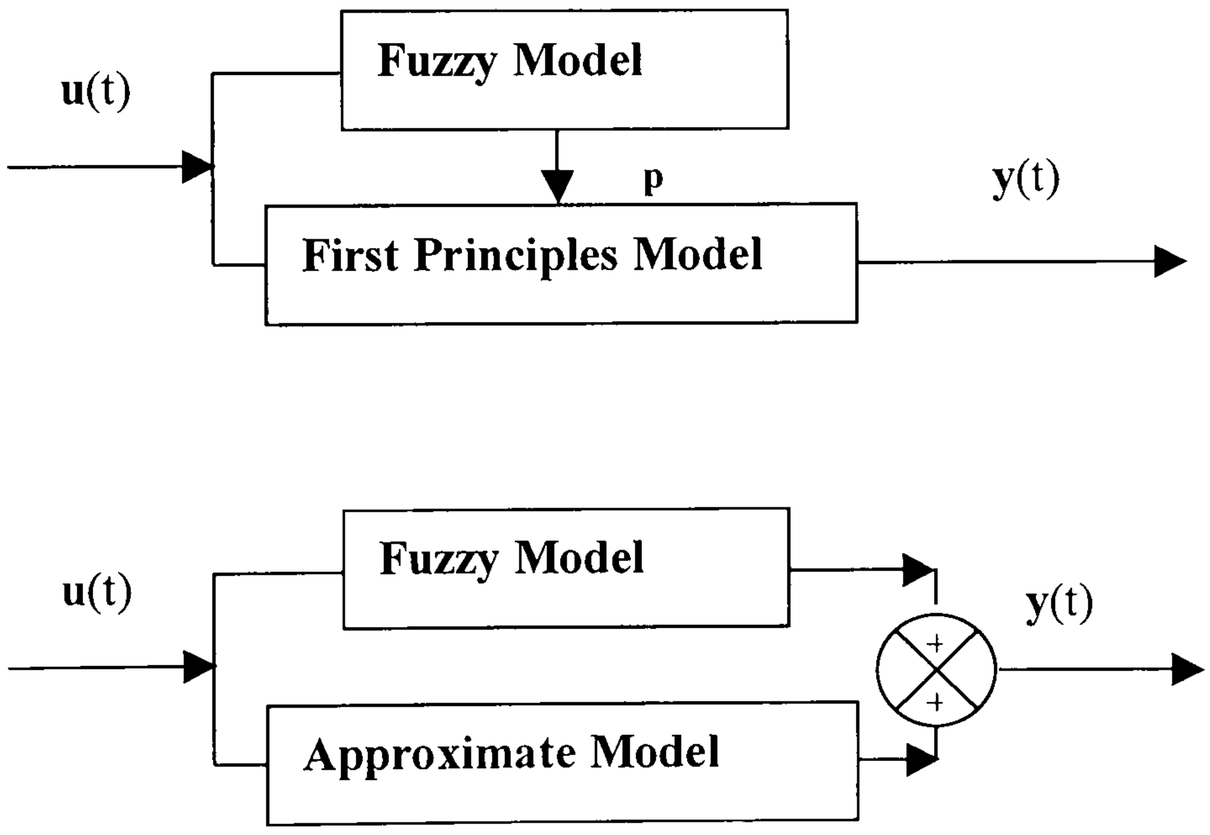


Figure 33 : Serial and parallel structure of grey-box models with fuzzy logic

To generate a fuzzy hybrid model, the model framework/structure is designed using first principles knowledge and process expertise. The key process variables and the mathematical equations for the non-linear parameters are thus described. In the next stage, the behaviour of the sub-processes described by the fuzzy parameters is

determined, typically using state estimation techniques. Van Lith *et al* (2002) used an Extended Kalman Filter. The estimates were then used in the sub-model identification stage, in this case fuzzy clustering, where the set of objects is divided into self-similar groups. This technique was implemented since no prior knowledge about the structure was available. The sub-models were then integrated into the hybrid models. This step is straight forward since the general structure of the system was a framework of accumulation balances along with an algebraic and two fuzzy equations. Finally model performance was improved through the optimisation of the fuzzy logic parameters. Again a model such as this can be used in tandem with techniques such as PCA, Yang *et al* (1999), which is something that should be considered for future work.

4.7 Conclusions

There are a raft of different methods for incorporating additional information about a process into a modelling or monitoring tool. Process-specific methods such as model-based PCA and model-based neural networks combine information about the physicochemical relationships and phenomena in the process with the process measurements to produce a more realistic model. All the techniques discussed in this chapter have shown that by including additional information about the process, a monitoring/prediction tool is provided that gives more accurate results and provides greater insight into the process than a standard black box approach. In Chapters 6 and 7 the model-based PCA approach is investigated further.

Chapter 6 takes the more common of these batch monitoring techniques and applies them to the monitoring of an exothermic batch reactor simulation. In doing this, the effectiveness of each technique can be assessed in an attempt to determine which of the methods is the most appropriate as a batch monitoring tool. Chapter 7 builds on the model-based approach to improve it further through use of additional modelling techniques.

Model-based PCA shows the potential to be an effective batch performance monitoring technique. It incorporates a mechanistic model of the process into the multivariate monitoring framework, and therefore attempts to address both the dynamic and non-linear behaviour occurring in the process. However the drawback of this method is that its effectiveness as a monitoring tool decreases when the mechanistic model is not a perfect match of the process, a problem that is likely to occur in an industrial situation.

In addition, mechanistic models can be expensive and difficult to create, hence the interest in hybrid techniques.

5 CHAPTER FIVE : EXOTHERMIC SIMULATION OF A BATCH REACTOR

5.1 Introduction

In subsequent chapters, a number of batch monitoring techniques are studied utilising data generated from a simulation of a batch process. For this purpose, Luyben's simulation of an exothermic batch reactor Luyben (1973) was used to create the plant data for the monitoring techniques under consideration, including multiway PCA, batch observation level analysis and model-based PCA. This chapter provides a detailed description of the simulation and how it has been developed.

5.2 Process Description

Figure 34 shows a diagram of the exothermic batch reactor under investigation. The operation of the batch reactor is as follows. The reactants are initially charged into the vessel at the start of the process and then, to bring the vessel contents up to the correct temperature to initiate the reaction, steam is fed into the jacket. Once the desired temperature has been reached, the reaction begins. However due to the exothermic nature of the process this is accompanied by the release of heat energy. Therefore to keep the reaction stable, it is necessary to remove the excess heat by feeding cooling water into the vessel jacket. Luyben designed the simulation so that the cooling water is fed into the jacket to force the reactor temperature to follow the correct temperature profile to produce the desired products. In Luyben's system, the required temperature profile is sent to the temperature controller as the set point signal. In the simulation that was used in this research, the temperature control has been simplified so that a constant set point is used thereby giving a less complex control system, however this still gave similar variable profiles (see Chapter 5) to those attained by Luyben.

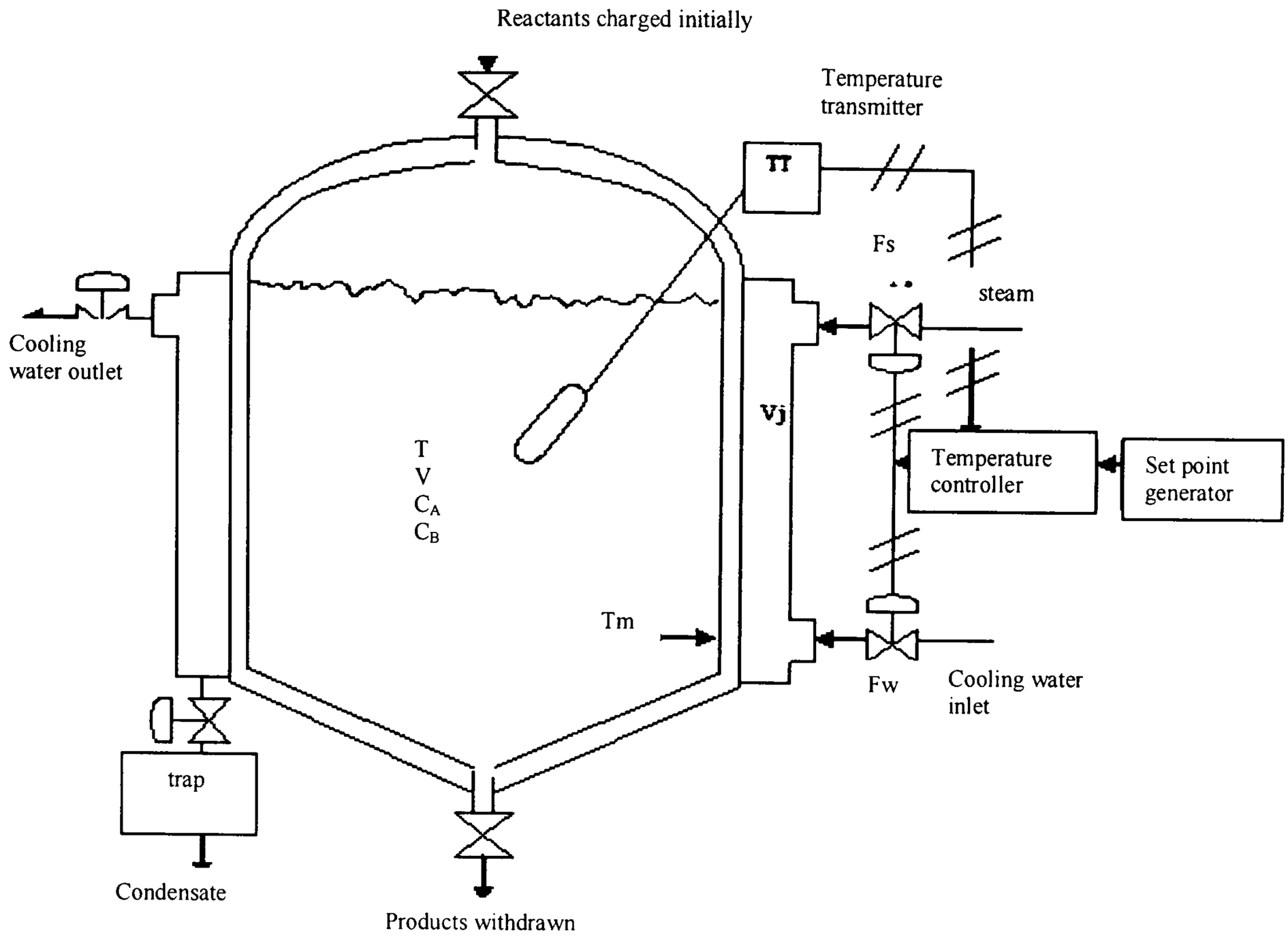


Figure 34: Schematic of the exothermic batch reactor

As the temperature in the reactor increases, the following first order consecutive reactions occur:

$$A \xrightarrow{\kappa 1} B \xrightarrow{\kappa 2} C \qquad (5.1)$$

In this reaction, the desired product is component B. Consequently the timing of the reaction has to be carefully controlled. If the reaction is terminated early, the percentage of component A that has been reacted will be too small, giving a low conversion and a low yield of product B. If the reaction continues for too long, component B will convert to by-product C, again giving a low yield of the desired product. Therefore there is an optimum batch duration that maximises the amount of product produced.

In addition to the optimum batch duration, there is also an optimum temperature profile at which to operate the reactor. The relationship between the rate at which the reaction proceeds and its temperature is represented by the Arrhenius equation:

$$\kappa = F * \exp^{(-Ea/RT)} \qquad (5.2)$$

where κ is a rate coefficient, F is a constant, Ea is the activation energy, R is the universal gas constant (8.314 $*10^{-3}$ kJ/molK), and T is the temperature (° Kelvin). At higher temperatures, the probability that two molecules will collide increases resulting in higher kinetic energy, which in turn affects the activation energy of the reaction. Consequently, the activation energy is the energy required to initiate the reaction. If the rate coefficients, κ_1 and κ_2, in this exothermic batch reaction have the same activation energies (i.e. the effect of temperature on each of the rates is equal) then the reaction needs to be operated at the maximum possible temperature to minimise the batch duration (and therefore reduce costs). Clearly this temperature would be limited by certain factors including the effect of the temperature on the reactants and the maximum operating temperature and pressure of the vessel and its associated components.

In the case where κ_1 is more temperature dependent than κ_2, the preferred option is again to run at the maximum possible temperature as this favours the production of product B. If κ_1 is less temperature dependent than κ_2, the control options are slightly more complicated. Ideally, the reaction has to be started off at a high temperature to initiate the reaction, the temperature then has to be reduced to prevent too much of product B reacting to form by-product C. Therefore there is an optimum temperature profile at which the batch reaction should be operated. In this case κ_1 is less temperature dependent than κ_2 and so the simulation was controlled using the optimum temperature profile.

In the following sections the equations used to develop the mathematic model of the process are described. The initial values of the parameters used to run the simulation are given in Section 5.5.

5.3 Mathematical Model

5.3.1 Overall Process

A number of assumptions were made to develop the mathematical model of the process. First of all it was assumed that the density of the reaction liquid, ρ, is constant, therefore the total continuity equation for the vessel contents is given by:

$$\frac{d(\rho V)}{dt} = 0 - 0 \tag{5.3}$$

Additionally it is assumed that there is no flow in or out of the reactor during the reaction and since ρ, the liquid density, is assumed to be constant, $dV/dt = 0$, that is the volume of liquid in the reactor is constant.

For each of the reactants, A and B:

$$V\frac{dC_A}{dt} = -V\kappa_1 C_A \tag{5.4}$$

$$V\frac{dC_B}{dt} = V\kappa_1 C_A - V\kappa_2 C_B \tag{5.5}$$

where V is the volume of the reactor contents, C_A and C_B are the concentrations of components A and B respectively, and κ_1 and κ_2 are the rate constants of the reactions.

The Arrhenius equation, described in section 5.2, for the specific reaction is as follows:

$$\kappa_1 = \alpha_1 e^{-E_1/RT} \tag{5.6}$$

$$\kappa_2 = \alpha_2 e^{-E_2/RT} \tag{5.7}$$

where α_i is a constant, and E_i is the activation energy. It is assumed that neither of these parameters varies with temperature.

The energy balances for the process, equation 5.9, and the metal wall, equation 5.10, have been derived using a lumped model for the reactor metal wall and the enthalpy equation 5.8:

$$h = C_p T \tag{5.8}$$

$$\rho V C_p \frac{dT}{dt} = -\lambda_1 V\kappa_1 C_A - \lambda_2 V\kappa_2 C_B - h_i A_i (T - T_M) \tag{5.9}$$

$$\rho_M V_M C_M \frac{dT_M}{dt} = h_o A_o (T_J - T_M) - h_i A_i (T_M - T) \tag{5.10}$$

where,

ρ	density	h_i	Inside film coefficient
V	Volume	h_o	Outside film coefficient
C_p, C_m	Heat capacity	A_i	Inside area
λ_1, λ_2	Exothermic heats of reaction	A_o	Outside area
κ_1, κ_2	Rate constants		

The subscripts 'M' and 'J' indicate the metal wall and the jacket respectively.

5.3.2 Heating Phase

To model the process during the heating phase of the reaction, the following set of equations are used.

The mass balance:

$$V_J \frac{d\rho_j}{dt} = F_s \rho_s - W_c \qquad (5.11)$$

where F_s is the flowrate of steam into the jacket and W_c is the rate of condensation of the steam. For the simulation, it is assumed that the liquid condensate is immediately drawn off through a steam trap, as shown in Figure 34.

Energy balance for the steam:

Change in energy = Energy in (steam) – Energy out (to process) – Energy (lost to condensate)

$$V_J \frac{d(U_J \rho_J)}{dt} = F_S \rho_S H_S - h_o A_o (T_J - T_M) - W_c h_c \qquad (5.12)$$

where U_j is the internal energy of the steam in the jacket, H_s is the enthalpy of the incoming steam, and h_c is the enthalpy of the liquid condensate. For this simulation, the internal energy changes are neglected since they are negligible with respect to the latent heat effects. Therefore the steady state energy balance can be simplified to give:

$$W_c = \frac{h_o A_o (T_J - T_M)}{H_s - h_c} \tag{5.13}$$

Using the perfect gas law and a vapour pressure equation, the jacket temperature and pressure can be calculated iteratively if ρ_j is known:

$$\rho_J = \frac{M}{RT_J} \exp\left(\frac{A_w}{T_J} + B_w\right) \tag{5.14}$$

where M is the molecular weight of the steam, and A_w and B_w are the vapour-pressure constants for water. Equation 5.14 is used to calculate T_j, and then P_j can be calculated using a vapour pressure equation.

The energy balance for the reactor jacket is considered unnecessary because of the small mass of metal involved.

5.3.3 Cooling Phase

Once the reaction temperature has been reached, cooling water is used in the jacket to remove the excess heat. Under the assumption that the jacket is perfectly mixed, the following energy balance is used to describe the heat transfer:

$$\rho_J V_J C_J \frac{dT_J}{dt} = F_w C_J \rho_J (T_{Jo} - T_J) + h_o A_o (T_M - T_J) \tag{5.15}$$

It should be noted that generally the outside film coefficient is significantly different for condensing steam and cooling water, however to keep the model simple only one value of h_o has been used.

5.4 Control of the Batch Reactor

As described earlier, the control of the exothermic batch reactor involves heating the vessel contents up to the required temperature and then a cooling medium is used to remove the excess heat of reaction so that it does not turn into a runaway reaction. This is more difficult to control than an endothermic reaction, where if the source of heat is removed, the reaction will stop. It can be difficult combining both heating and cooling

because the objective is to heat the contents up to the optimum temperature as quickly as possible without overshooting the set point. For this type of operation, split range control valves are often used. In split range control (Love (2007)) the output of the controller is split between two (or more) valves, in this case the cooling water valve and the steam valve.

During the heating phase of the simulation, the steam valve is kept open and the cooling water valve closed. Once the required temperature is reached, the temperature controller then closes the steam valve and opens the cooling water valve enough to control the reactor temperature at its set point. As the reaction progresses, the cooling water valve is opened further to maintain the correct temperature profile. It should be noted that during the heating phase, the cooling water outlet valve remains closed and the condensate valve remains open (as observed in Figure 34). During the cooling phase, the cooling water outlet valve should be opened whenever the cooling water is being fed to the jacket, and the condensate valve should be closed.

In this simulation, all the instrumentation is pneumatic, therefore the controller output pressure P_c varies between 3 and 15 psig. The valves are tuned so that the steam valve is 100% open when the controller output pressure is 15 psig and is fully closed at 9psig. Conversely the water valve will be 100% open at 3psig and fully closed at 9psig, as shown in Figure 35.

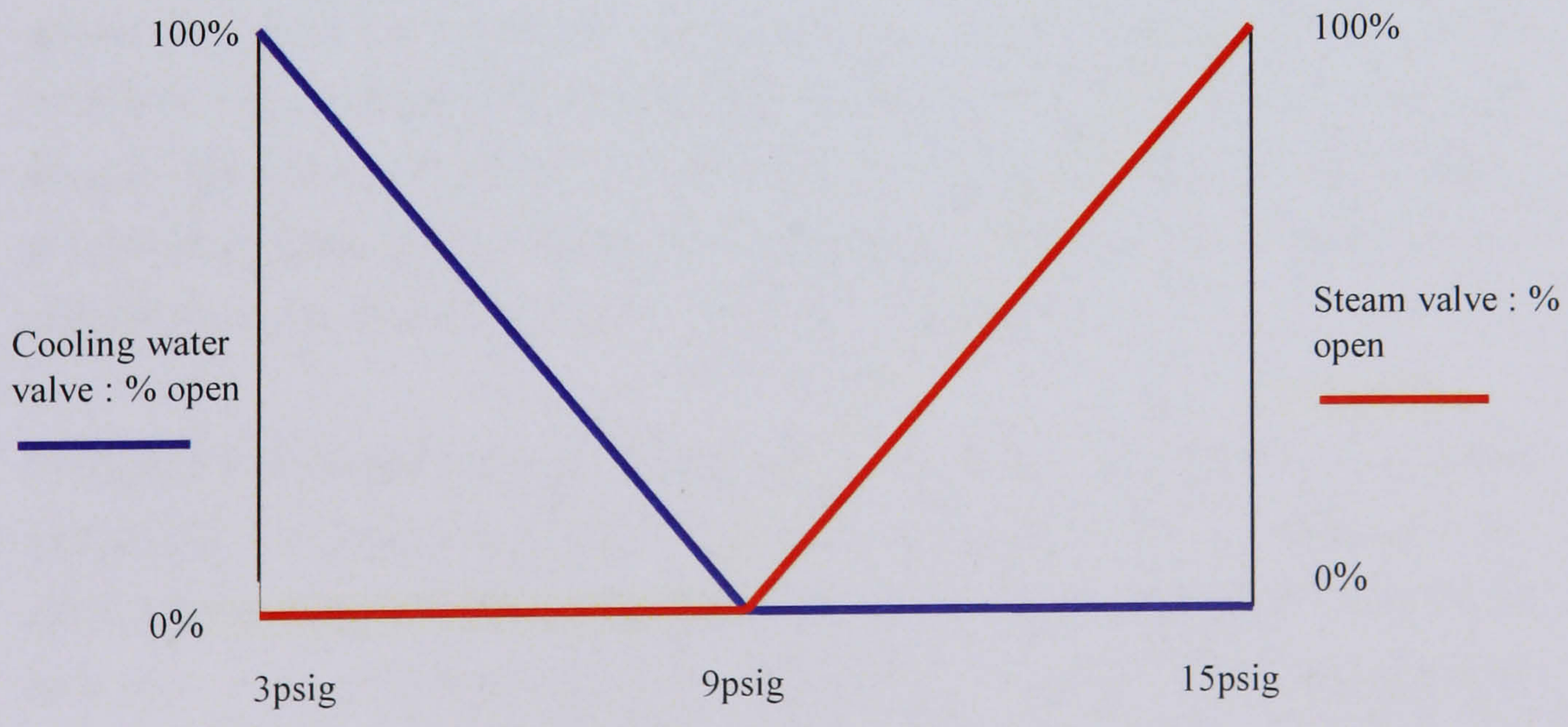


Figure 35: Split range control valve diagram of reactor heating and cooling system

An endothermic reaction is inherently safe because if the control system fails causing the heating/cooling system to fail, the reaction will stop as it needs heat energy to continue. However with an exothermic reaction if the control system fails, causing the heating /cooling system to fail, a runaway reaction will occur as the reaction will continue to release heat with no method of removal. The split range valves have been designed to operate in a safe manner such that if there is a failure in the control system, the reaction will not be allowed to go out of control. The cooling water valve is a fail open valve, meaning that if the instrument air pressure is lost or similar, the valve will fail to the open position, allowing the cooling water to continue to fill the jacket. The steam valve is a fail close valve, so in the event of a control system failure, the valve will fail closed – removing any heat source from the reactor vessel. This method of valve configuration will stop the reaction safely. In addition to this, many reactors also have inhibitor dumping systems where a chemical which stops the reaction is dumped into the reactor as soon as an emergency condition or severe control system failure is reached, although this is not a necessary part of this simulation.

The type of control algorithm used with the split range control valves is a proportional feedback controller. A feedback controller works by calculating the error between the measured signal (such as the reactor temperature) and the controller setpoint (the target temperature) and sending an output signal to the control valve (or valves) to try and reduce the error between the two. This is the simplest form of feedback controller as it does not contain the integral and derivative parts of a conventional feedback controller. The proportional action in the control algorithm acts on the error in the signal, the greater the error, the greater the change in output signal. The user can define the sensitivity of the controller by adjusting the controller gain (denoted by Kc) – increasing the gain makes the controller more sensitive to changes in error. Tuning of the algorithm is required to ensure that Kc is sufficiently large to act on the error with a rapid response without forcing the process to become oscillatory and unstable.

To keep the simulation simple, integral action and derivative action have not been included in this control algorithm. The purpose of integral action is to eliminate the offset from the setpoint, i.e. after the initial proportional 'kick' has died away there will be a residual offset between the setpoint and the measured signal. The integral action acts on this offset through the integration of the error, it slowly closes or opens the valve and forces the variable towards the setpoint – it has a more long term effect than the proportional action. The derivative action is used to stabilise and increase the speed of the controller response. It is dependent on the slope of the error and not on the

magnitude. This type of action should not be included in the algorithm if the process is noisy since it has the effect of amplifying the spikes in the data, forcing the controller output to swing wildly.

For simplification, only the proportional feedback controller was used in this simulation. In the simulation, the bias of the controller was set to 7psig (meaning that when there is zero error between the measured signal and the setpoint, the controller output pressure is 7psig):

$$P_c = 7 + K_c(P^{SET} - P_{TT}) \qquad (5.16)$$

In Luyben's original model, a pneumatic function generator was used to supply the set point, the slightly simplified version of the model that was used in this research uses a fixed set point, which still provided acceptable variable profiles.

The exothermic batch reactor simulation described in detail in this chapter is used in the subsequent chapters to generate data used in the assessment of a number of different batch monitoring techniques.

5.5 Initial Model Parameters

The initial values for the parameters used in the simulation are summarised in Table 1:

Model Parameter	**Definition**	**Initial Value**
κ_1	Rate constant reaction 1	$6.2949*10^{-10}$
κ_2	Rate constant reaction 2	$5.4263*10^{-5}$
α_1	Constant – reaction 1	729 min^{-1}
α_2	Constant – reaction 2	6567.6 min^{-1}
E_1	Activation energy – reaction 1	15000 Btu/lb mol
E_2	Activation energy – reaction 2	20000 Btu/lb mol
C_a	Concentration of reactant A	0.8 lb mol A/ft^3
C_b	Concentration of reactant B	0
R	Universal gas constant	8.314 kJ/mol K
T	Reactor temperature	80K
Tm	Wall temperature	80K
Tj	Jacket temperature	80K

Th	Temperature of heating medium	260K
Tc	Temperature of cooling medium	80K
λ_1	Exothermic heat of reaction – reaction 1	-40000 Btu/lb mol
λ_2	Exothermic heat of reaction – reaction 2	-50000 Btu/lb mol
ρ	Density	50 lb/ft^3
ρ_j	Density of water	62.3 lb/ft^3
Cp	Heat capacity	1 Btu/lb K
C_j	Heat capacity of water in jacket	1 Btu/lb K
C_m	Heat capacity of reactor vessel	0.12 Btu/lb F
V_j	Volume of jacket	18.83 ft^3
V_m	Volume of vessel	42.5 ft^3
h1	Heat transfer coefficient – reaction 1	0.071
h2	Heat transfer coefficient – reaction 2	0.177
Φ	Jacket space velocity	1.058 ft/s
Uh	Position of heating valve	0.65
Uc	Position of cooling valve	0
θ_1	Calculated value ($\rho_m * C_m * V_m$)	578.765
θ_2	Calculated value ($\rho_j * C_j * V_j$)	1173.109

Table 1: Initial parameter values for exothermic batch simulation

6 CHAPTER SIX: CASE STUDY OF STANDARD MONITORING TECHNIQUES

6.1 Introduction

In Chapters 3 and 4, a selection of the current techniques available for the monitoring of the performance of batch processes were reviewed. The objective of this chapter is to assess the performance monitoring abilities of a number of the more commonly applied batch monitoring tools, and compare their performance using the simulation of the exothermic batch reactor, Luyben (1973) described in Chapter 5. The comparison is based on how well each technique deals with the characteristics of a batch process, including non-linear and dynamic behaviour, how rapidly each technique detects the onset of an abnormality in the process and how easily the source of the abnormality can be identified.

6.2 Exothermic Batch Simulation

Luyben's two stage exothermic batch simulation, as described in detailed in Chapter 5, is used. As discussed, it consists of two first order reactions which take place consecutively in the reactor. The desired product from the reaction being component B:

$$A \xrightarrow{\kappa_1} B \xrightarrow{\kappa_2} C \qquad (6.1)$$

The mathematical model described in Chapter 5 is summarised in equations (6.2-6.6). The 4th order Runge Kutta method was used to numerically integrate the ordinary differential equations in the mathematical model.

$dCa/dt = \kappa 1 * Ca$	(6.2)
$dCb/dt = \kappa 1 * Ca - \kappa 2 * Cb$	(6.3)
$dT/dt = \left[\kappa 1 * ca * \lambda 1 + \kappa 2 * cb * \lambda 2/\rho * cp + h1 * (Tm - T)\right]$	(6.4)
$dTj/dt = \phi * \left[Uh * \left(Th - Tj\right) + Uc * \left(Tc - Tj\right)\right] - \left(h2 * (Tj - Tm)\right)/\theta 2$	(6.5)
$dTm/dt = \left[h2 * (Tj - Tm) - h1 * (Tm - T)\right]/\theta 1$	(6.6)

where,

Ca, Cb : reactant concentrations	$\kappa 1$, $\kappa 2$: rate constants
T : reactor contents temperature	$\lambda 1$, $\lambda 2$: heats of reaction
Tm : reactor wall temperature	ρ : density
Tj : reactor jacket temperature	*Cp* : heat capacity
h1, *h2* : heat transfer coefficients	*Uh, Uc* : heating and cooling media valve positions
ϕ : jacket space velocity	

The simulated batch duration was three hundred minutes, with measurements being recorded every minute for the reactor temperature, wall temperature, jacket temperature and cooling valve position. Of this batch time, the first fifty minutes were considered to be the safe start-up time, as determined by Lewin and Lavie (1990). During this period, the batch is being brought into control and settles into its normal operating window. Consequently the first fifty minutes of operation are excluded from the analysis.

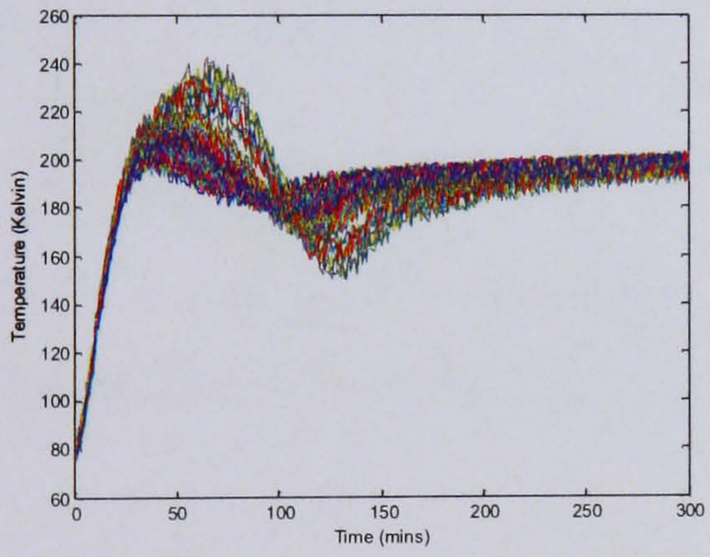
As the focus of this research was not how to deal with batches of differing lengths, the batches generated for this study were of equal length. This made the application of techniques such as MPCA more straight forward since they are not designed to easily accommodate differing batch lengths. However the issue of batch length is an important and widely studied subject in batch monitoring and a number of papers have been published on the subject (Section 3.3).

6.3 Data Set

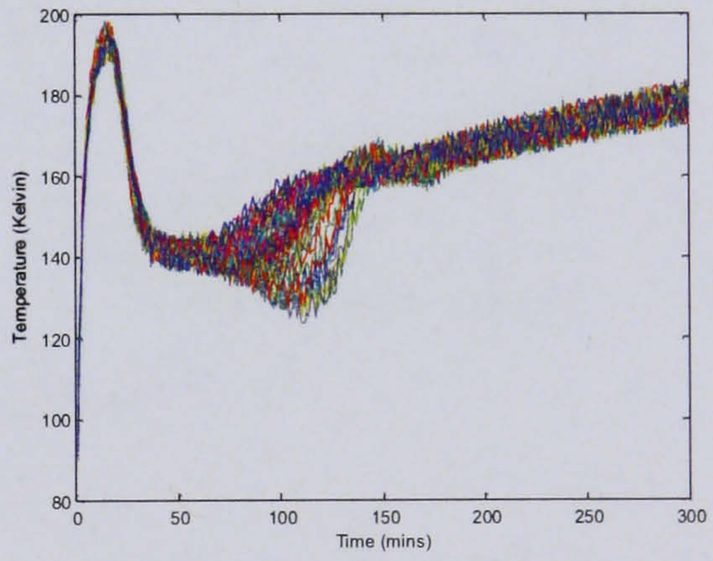
The case study, which focuses on the comparison of a number of different monitoring techniques, necessitated the generation of a data set that reflected normal operating conditions. Lewin *et al* (1998) undertook a study on the benefits of model-based PCA using the exothermic batch simulation, hence the same method for generating data was adopted in this study. In generating the data set, four process variables were recorded throughout the batch - three reactor temperatures (vessel content (T), reactor wall (T_m),

and jacket outlet(T_j)) and the position of the cooling water control valve (U_c). When Lewin carried out his studies into model-based PCA, he assumed that the reactant concentrations would not be available on-line, therefore for comparison, the same procedure was adopted in this study. Nowadays it is relatively common to measure the concentration of reactants on-line, for example using spectroscopic techniques such as Near-Infrared and Raman.

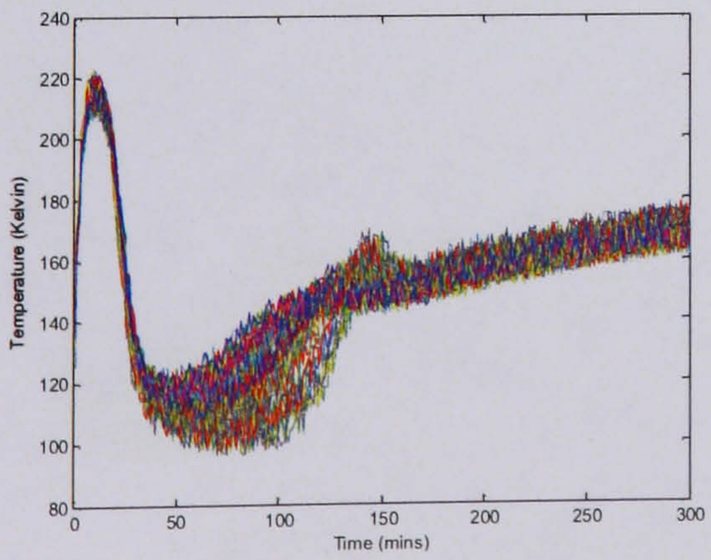
To test the batch monitoring techniques, a nominal data set comprising fifty batches was generated under normal operating conditions using the Runge Kutta method, with an additional twenty batches generated as a validation data set to investigate the performance of the model. To create these data sets, the temperature variables, heat transfer coefficients and activation energies were corrupted with random noise generated from a Gaussian distribution with a signal to noise ratio of 1.5. In reality it would only be the temperature measurements that would be affected by noise, however in the interest of increasing the variation between the batches, noise was also added to the heat transfer coefficients and activation energies. The plots in Figure 36 show the trajectories for each of the four variables for the fifty nominal batches. These plots show the full duration of the batch, i.e. three hundred minutes, however as stated earlier, for the purposes of the case study, the first 50 minutes were excluded.



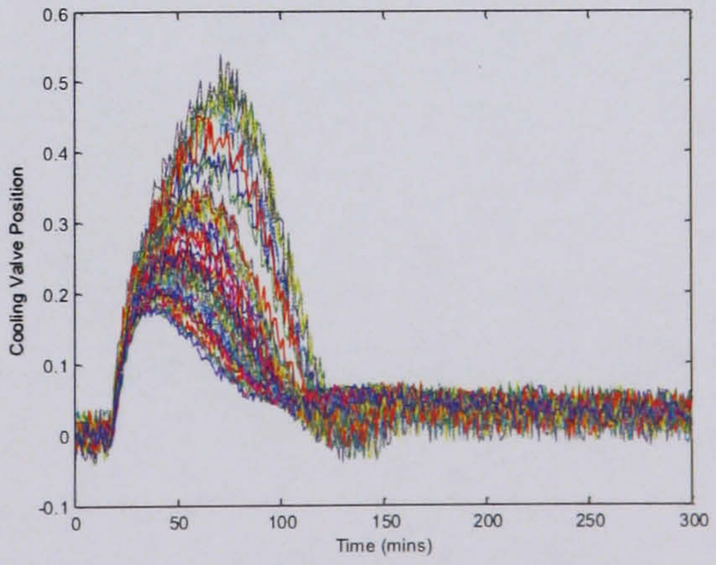
a) Variable 1 : Reactor Temperature



b) Variable 2 : Wall Temperature



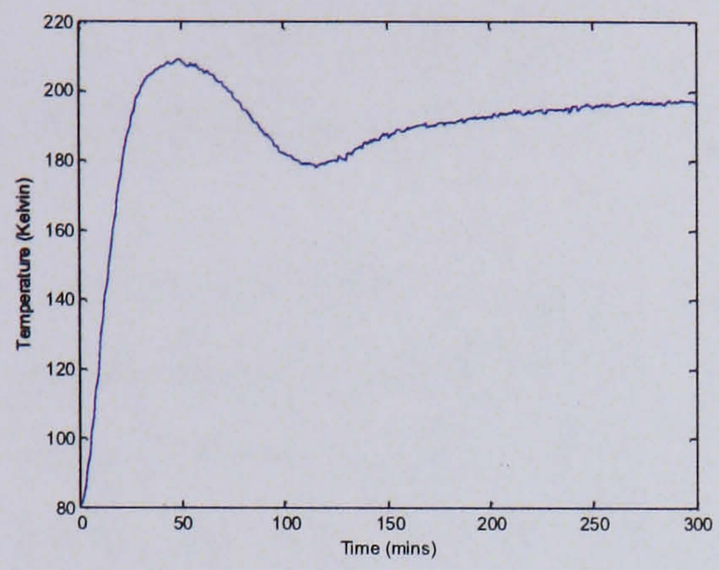
c) Variable 3 : Jacket Temperature



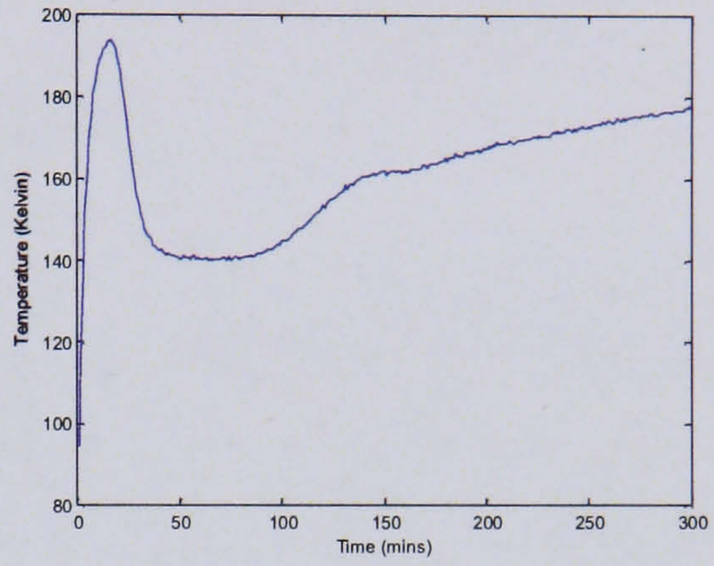
d) Variable 4 : Cooling Water

Figure 36: Variable trajectories of the nominal data set

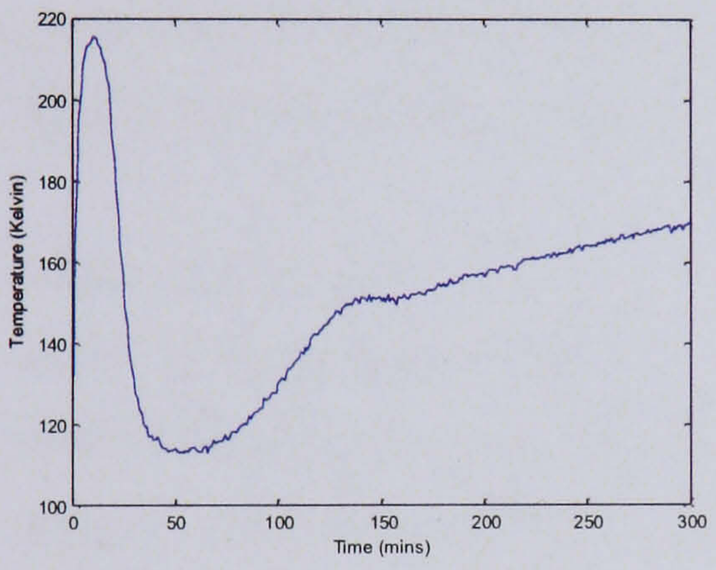
By studying the nominal data set, it is possible to obtain an indication of which variables and which time periods during the batch are responsible for the main sources of variation in the process. Furthermore, this can provide information on where there are opportunities to reduce process variation. Figure 37 (a)-(d) show the average trajectories for each of the variables in the simulation whilst Figure 38 (a)-(d) show the standard deviations around those average trajectories.



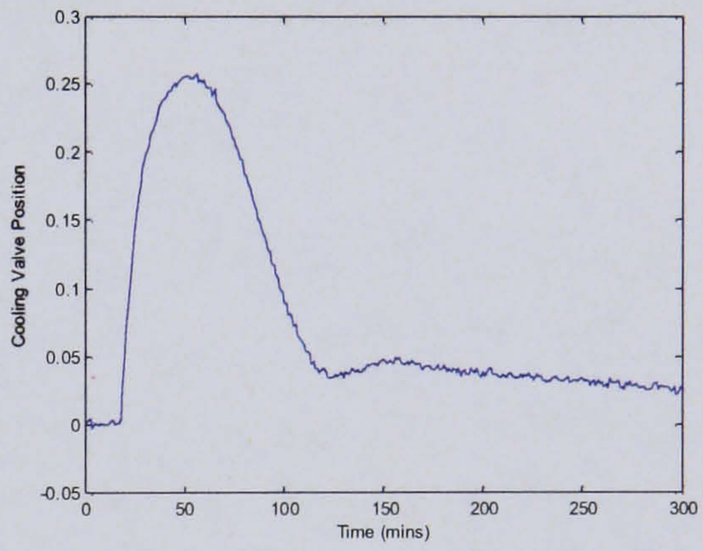
a) Reactor Temperature



b) Wall Temperature

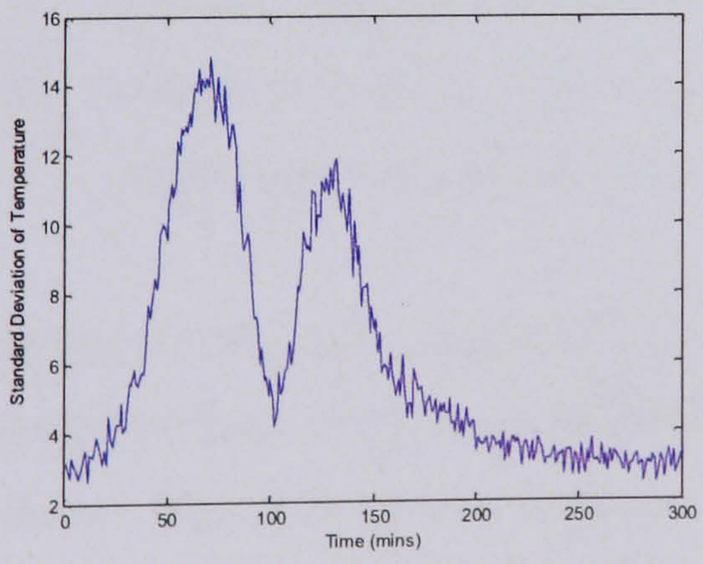


c) Jacket Temperature

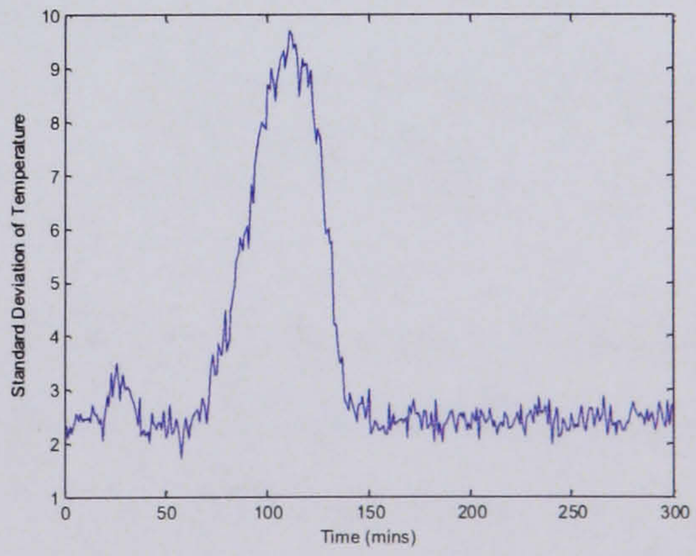


d) Valve Position

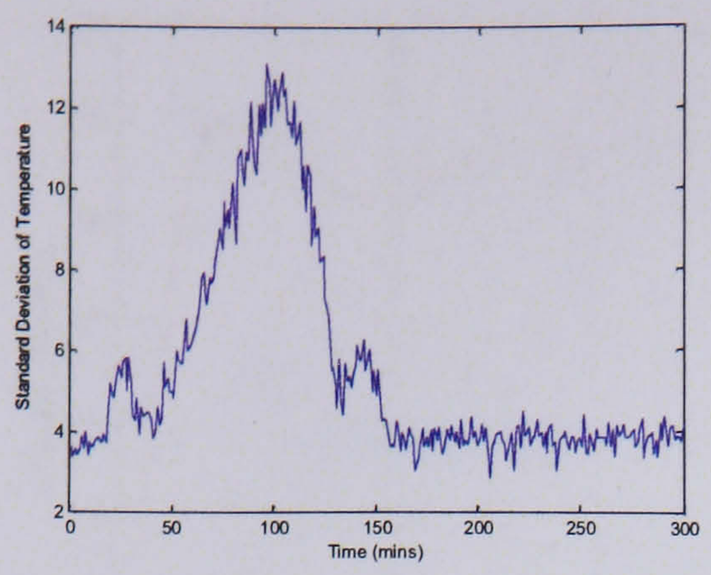
Figure 37: Average trajectories of the process variables



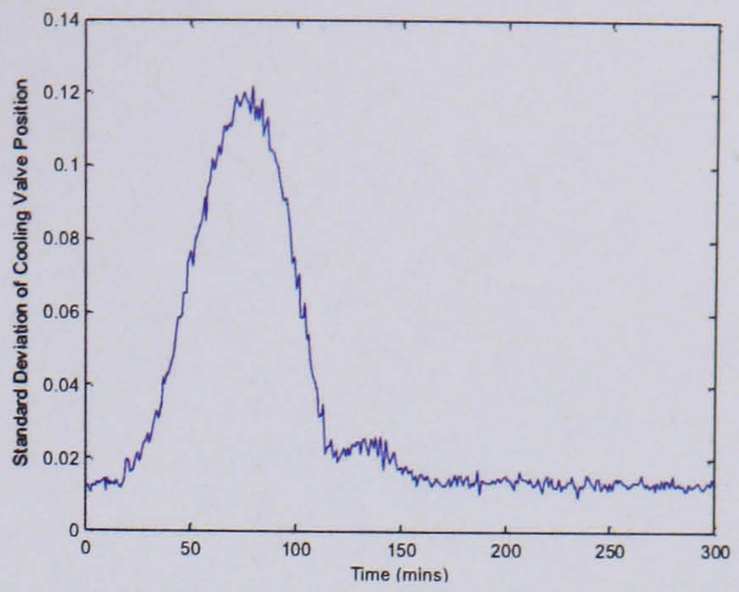
a) Reactor Temperature



b)Wall Temperature



c) Jacket Temperature



d) Valve Position

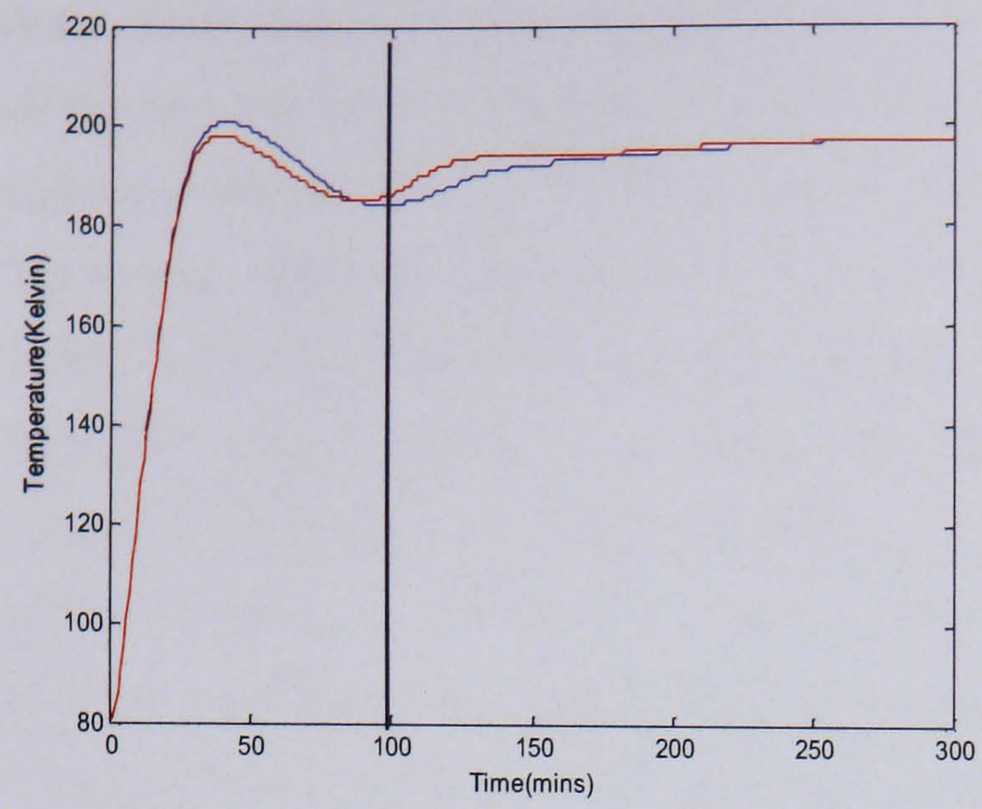
Figure 38: Standard deviations of the process variables

From the plots of the variable standard deviations, it is evident that the variation is greatest for all variables between 70 and 100 minutes. This behaviour is reflected in the plots of the average trajectories, which all go through a peak or trough during this period before starting to level out. This behaviour relates to the two stages of the reaction, that is between 70 and 100 minutes, the heat is applied to initiate the reaction, and the time period after 100 minutes where the reaction has begun to give out heat energy and hence cooling water is fed in to the reactor to remove the excess heat.

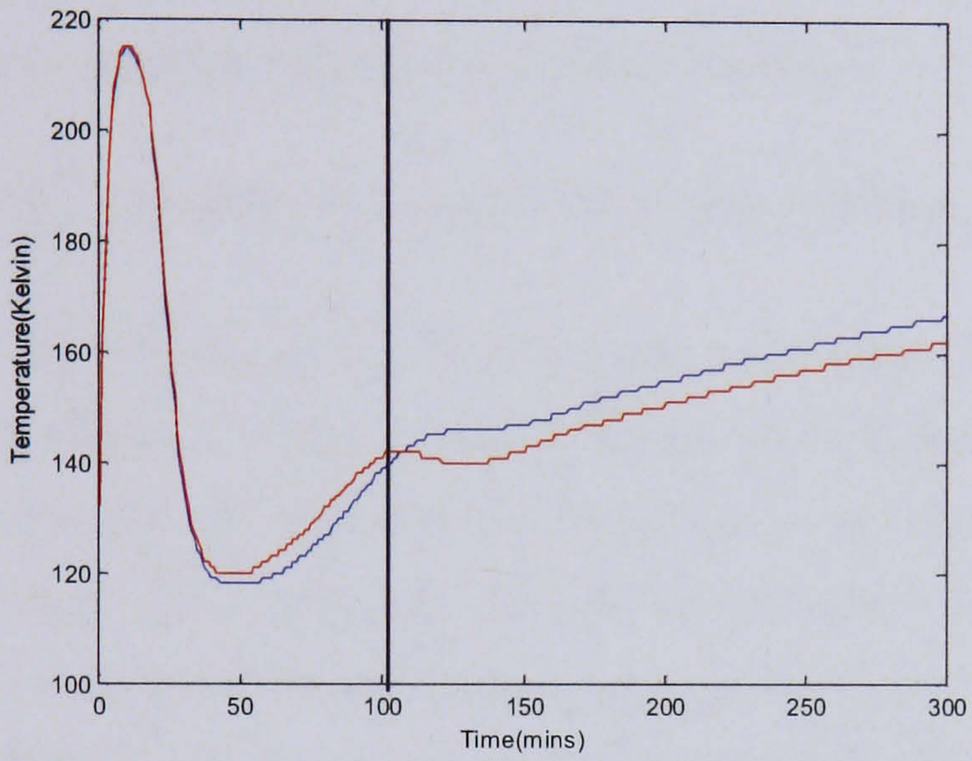
As the main objective of the case study is to examine the fault detection abilities of each of the monitoring techniques, in addition to the nominal and validation data sets, data sets containing different fault types were generated. To generate this data, three different types of process abnormality were considered, based on typical faults that could occur during a batch process:

i) a temperature sensor fault occurs after 100 minutes
ii) a decrease in the heat transfer coefficients is introduced after 150 minutes, which indicates that some form of fouling has occurred
iii) a fault with the cooling water valve which materialises as a change in the split range controller pressure output. The fault occurs after 150 minutes.

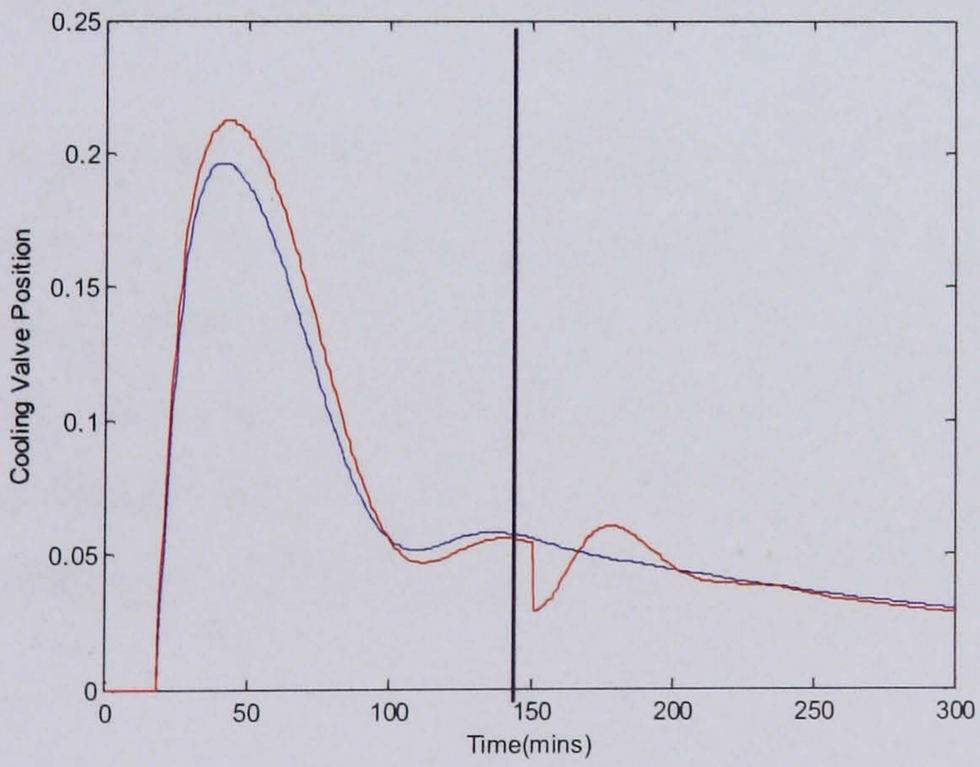
In Figure 39 (a)-(c), trajectories illustrating the difference between a batch operating under normal operating conditions and a batch containing one of the three faults are shown. The nominal trajectories are in blue while the fault trajectories are in red. The plots also show where the fault is introduced. It should be noted that due to the noise introduced to the data, the trajectories are not expected to be identical before the fault is introduced.



a) Fault type one: Reactor temperature (nominal: blue, fault: red)



b) Fault type two: Jacket temperature (nominal: blue, fault:red)



c) Fault type three: Cooling valve position (nominal: blue, fault: red)

Figure 39: Plots of nominal batch trajectories compared to abnormal batch trajectories for fault variables

A set of fifty batches for each fault type were generated. In the subsequent sections, each of the selected batch monitoring techniques (MPCA, BOL and MBPCA) are used to build a model of the process during normal operation from the fifty nominal batches. The twenty validation batches are then used to test the performance of the model, and finally the faulty batch data sets are then projected onto the control charts developed from these nominal models to see whether the process abnormalities are detected, and, if detected, how quickly the change in process operation is observed using each of the different techniques. As well as examining the plots to assess the fault detection abilities of these techniques, the ease of diagnosis of the fault is also considered, as are the dynamic and non-linear properties of the batch data set.

6.4 Nominal Models for Off-line Analysis

6.4.1 Multiway Principal Component Analysis

The advantages and disadvantages of multiway PCA were discussed in Chapter Two, particularly with reference to the fact that the technique performs more efficiently as an end-of-batch classification technique as opposed to a through-batch monitoring tool, since, although straightforward to implement, approximations need to be made to in-fill the data post the current time point to the end of the batch. Therefore the case study of MPCA focuses on the fault detection abilities of the technique as an end-of-batch method as opposed to a through-batch monitoring technique.

A multiway PCA model was built from the scaled nominal data set and cross-validation was carried to determine how many principal components should be retained. Table 2 shows the R^2, the fraction of the sum of squares explained by the current component, and Q^2 values, the fraction of the total variation of X that can be predicted by a component as estimated by cross-validation, and from this information four principal components were retained to represent the main sources of variation in the data.

No PCs ret	R^2x	Cum R^2x	Eigen-values	Q^2	Cum Q^2
1	0.646	0.646	32.3	0.634	0.634
2	0.258	0.904	12.9	0.706	0.892
3	0.0783	0.982	3.92	0.8	0.978
4	0.0123	0.994	0.614	0.678	0.993
5	0.00246	0.997	0.123	0.357	0.996
6	0.00211	0.999	0.105	0.653	0.998

Table 2: Results of cross validation analysis for multiway principal component analysis

The scores plots for the first four principal components are shown in Figure 40 a) and b), with a 95% control limit being calculated:

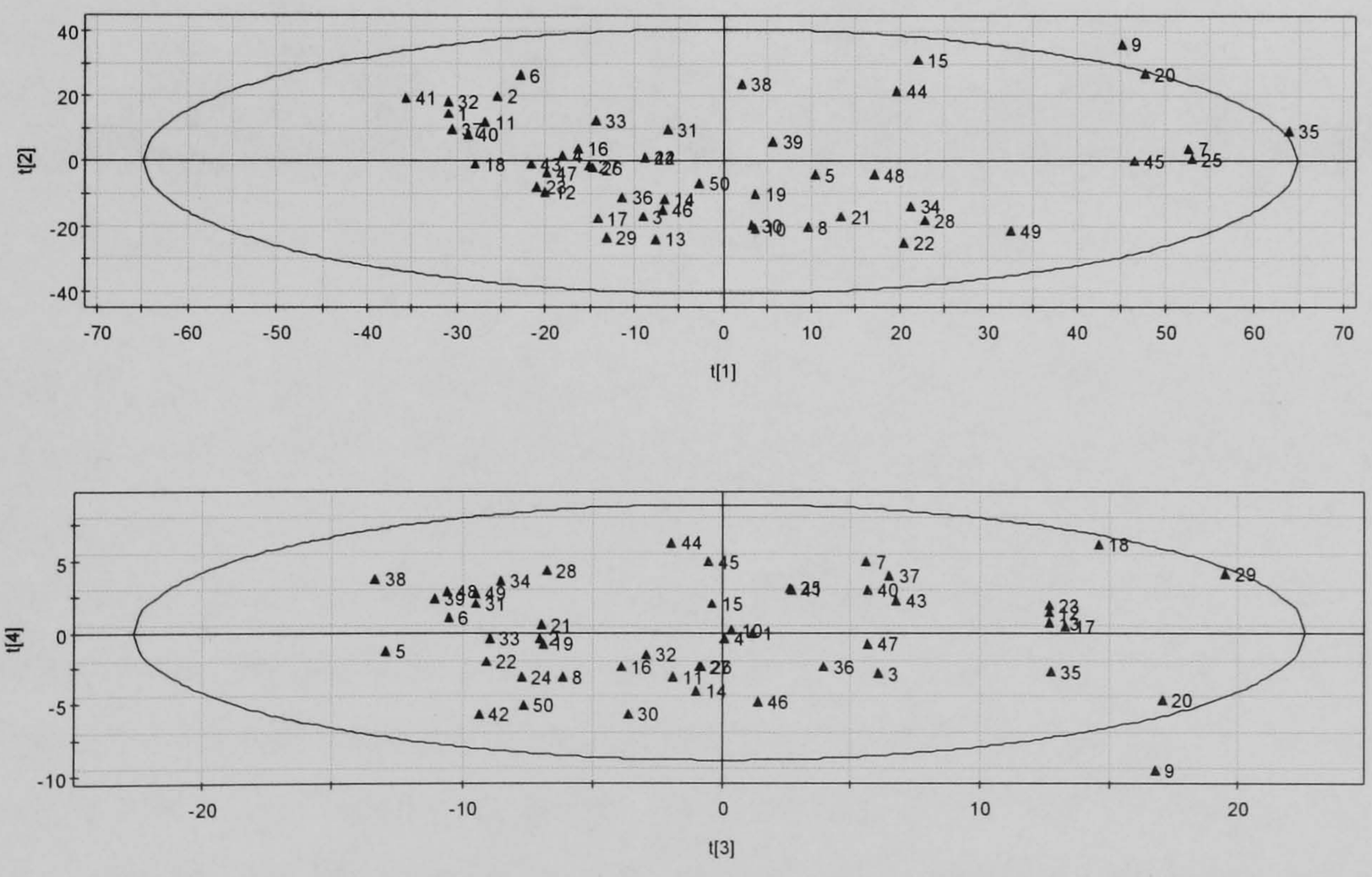


Figure 40: Bivariate scores plot for the nominal batch data set

The control charts in Figure 40 show that the majority of batches lie within the normal operating region, with only two batches, numbers 9 and 35, falling just outside of the confidence limits. For 95% control limits, as $n \rightarrow \infty$, 5% of the points in the data set will lie outside the limits, whilst for 99% control limits, 1% of the points will lie outside the

limits. In this case, the limits are set at 95%, and therefore the number of batches lying outside the limits is statistically acceptable.

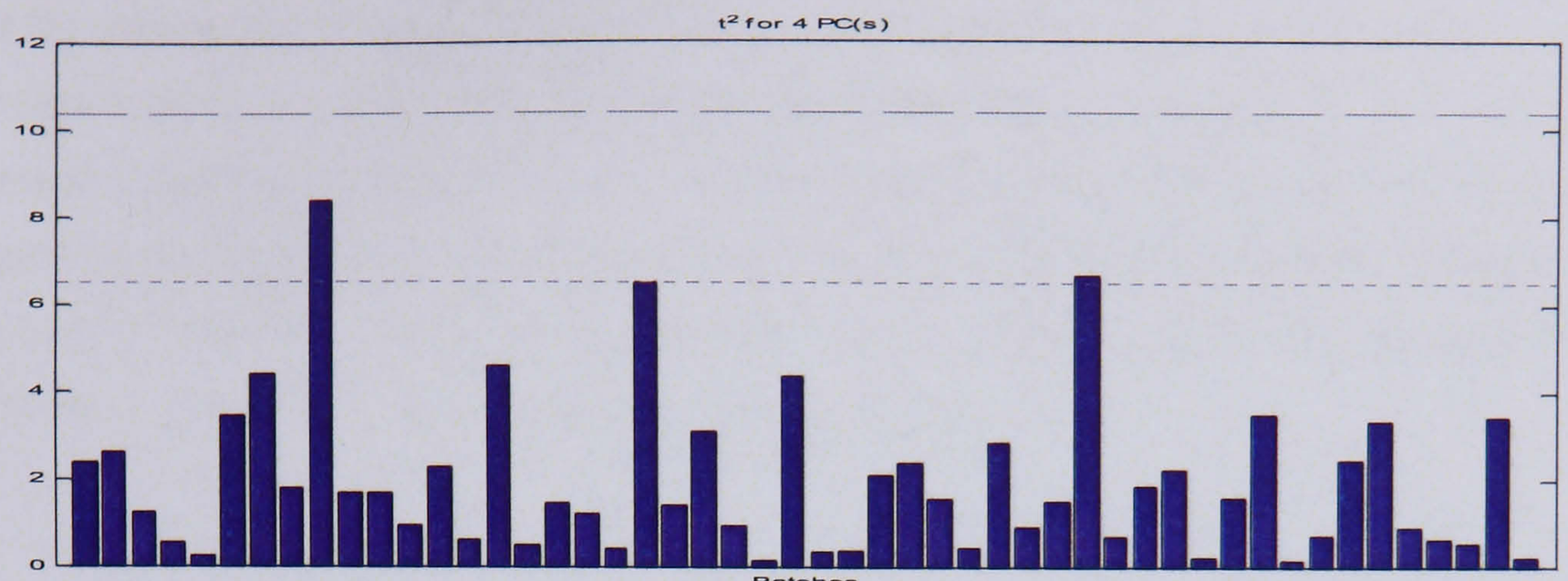


Figure 41: Hotelling's T^2 plot of nominal batch data

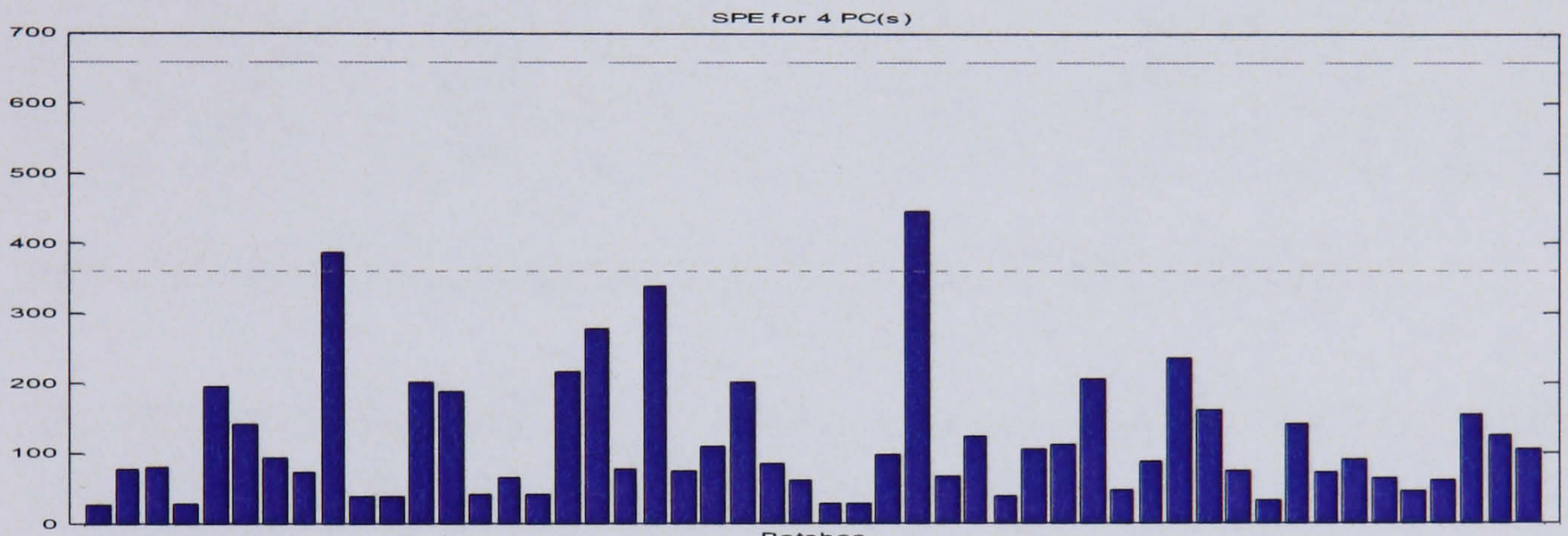


Figure 42: Square Prediction Error plot of nominal batch data

Figure 41 and Figure 42 show the Hotelling's T^2 and SPE metrics respectively. As expected, the Hotelling's T^2 plot demonstrates similar behaviour to the bivariate scores chart, batches 9 and 35 exceed the 95% limit, with batch 20 being close to the limit. The rest of the batches lie within the limits. The SPE plot also indicates that batch 9 lies outside the limits, however batch 29 is also considered to be out of statistical control with respect to the squared prediction error. The SPE statistic represents the variation in the data not accounted for by the model, consequently since the batch lies within the limits on the other plots, indicates that there could be some special cause variation within the batch, not captured by the model, as opposed to a change in the common cause variation that would be detected by the Hotelling's T^2 plot. However, an acceptable number of nominal batches lie within the limits to indicate that the model is valid. This will be investigated further in section 4.4, where the validation data set is projected onto the MPCA model.

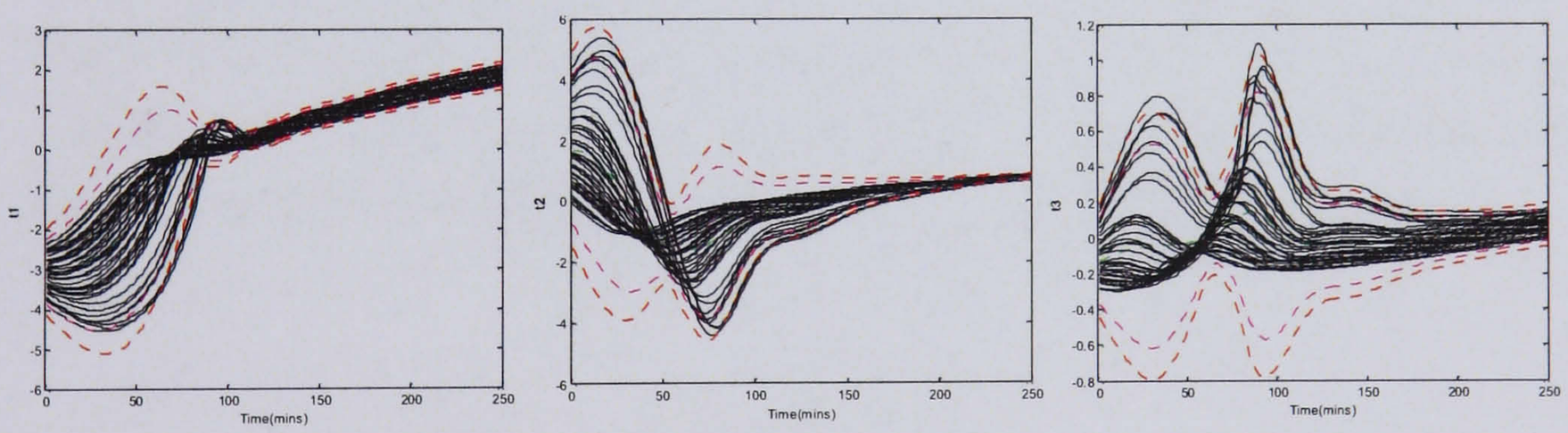
6.4.2 Batch Observation Level Analysis

Batch Observation Level (BOL) analysis is a through-batch technique which can be used for the monitoring of the performance of the evolving batches. The same nominal data set of fifty batches as used in the multiway principal component analysis study was used to generate a batch observation level model of the process under normal operating conditions. Again cross validation was used to determine how many latent variables should be retained. From these results it was decided to retain 3 latent variables to represent the variation in the process, although two would be sufficient.

No. ret PC	R^2x	Cum. R^2x	Eigen val.	R^2y	Cum. R^2y	Q^2	Cum. Q^2
1	0.665	0.665	2.66	0.811	0.811	0.811	0.811
2	0.327	0.992	1.31	0.057	0.868	0.301	0.868
3	0.006	0.998	0.032	0.00943	0.877	0.071	0.877

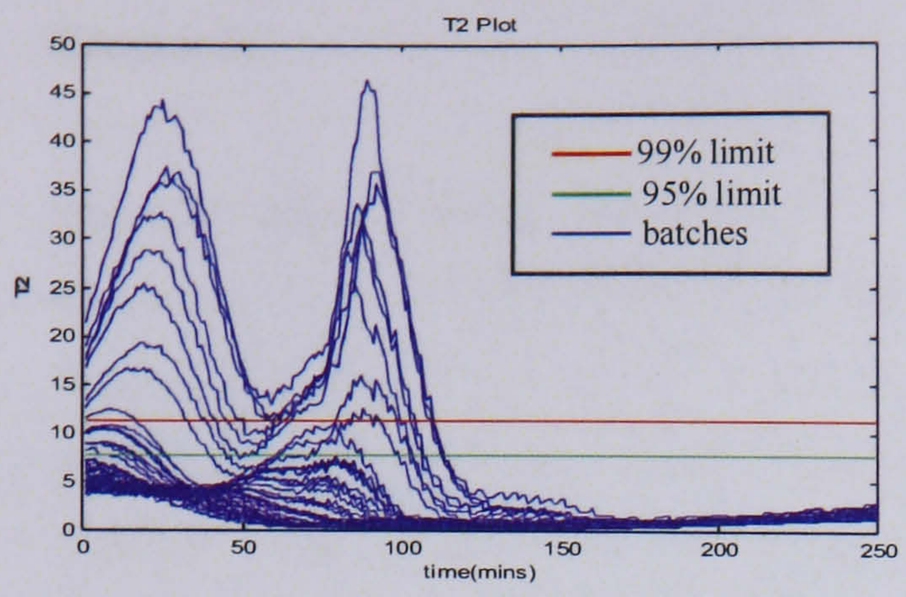
Table 3: Results of cross validation analysis for batch observation level analysis

The univariate scores plots for the nominal batch observation level model are shown in Figure 43, with the Hotelling's T^2 and SPE plots in Figure 44.

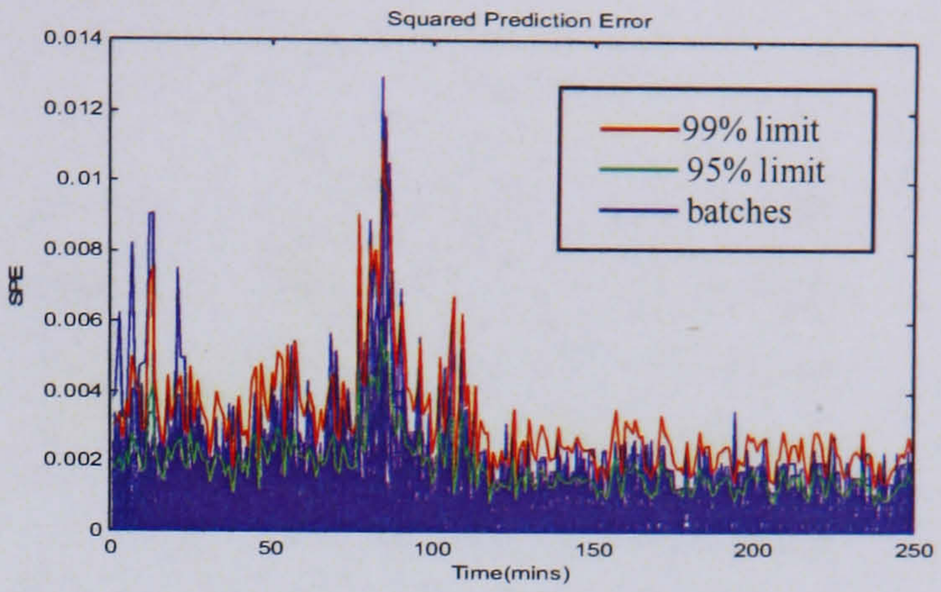


a) Latent variable 1 b) Latent Variable 2 c) Latent Variable 3

Figure 43: Scores plots of batch observation level model of nominal data set



a) Hotelling's T^2 plot for nominal data set



b) SPE plot of nominal data set

Figure 44: Hotelling's T^2 and SPE plots for nominal data set

The univariate scores control charts for the nominal data show that the control limits are largest in magnitude during the first 100 time points of the batch (this does not include the initial 50 time points already removed). After time point 100, batch behaviour becomes more consistent. In the control charts, only 2 batches deviate outside the limits, which is acceptable. The Hotelling's T^2 plot in Figure 44 also shows that there is a significant level of variation in the batches in the first 100 time points, with a number of batches moving outside the control limits during this time, before settling down in the second half of the batch. The batches have different phases during their operation, with the behaviour being significantly different between the start-up heating phase and the cooling phase. Therefore global monitoring, as is seen with the Hotelling's T^2 plot, is not ideal as it uses the same control limits for the duration of the batch. The SPE plot uses variable control limits, which takes into account the variation in batch behaviour, and again only 1 or 2 batches lie outside these limits during the 250 minutes.

6.5 Model-based Principal Component Analysis

The final technique considered in this chapter is that of model-based PCA. The basis of this approach is a first principles model of the process. The first principles model was generated following the same method as described in the research of Lewin. The five mass and energy balances defined for the exothermic simulation were used (eqns 6.2 – 6.6). Plant model mismatch was created by introducing small offsets to some of the kinetic parameters. The values of the rate constants (k_1 and k_2) and heats of reaction (λ_1 and λ_2) were increased by 7%, consequently the model and plant data were not an exact

match, as would be the case in an industrial situation (baseline parameters are defined in Chapter 5).

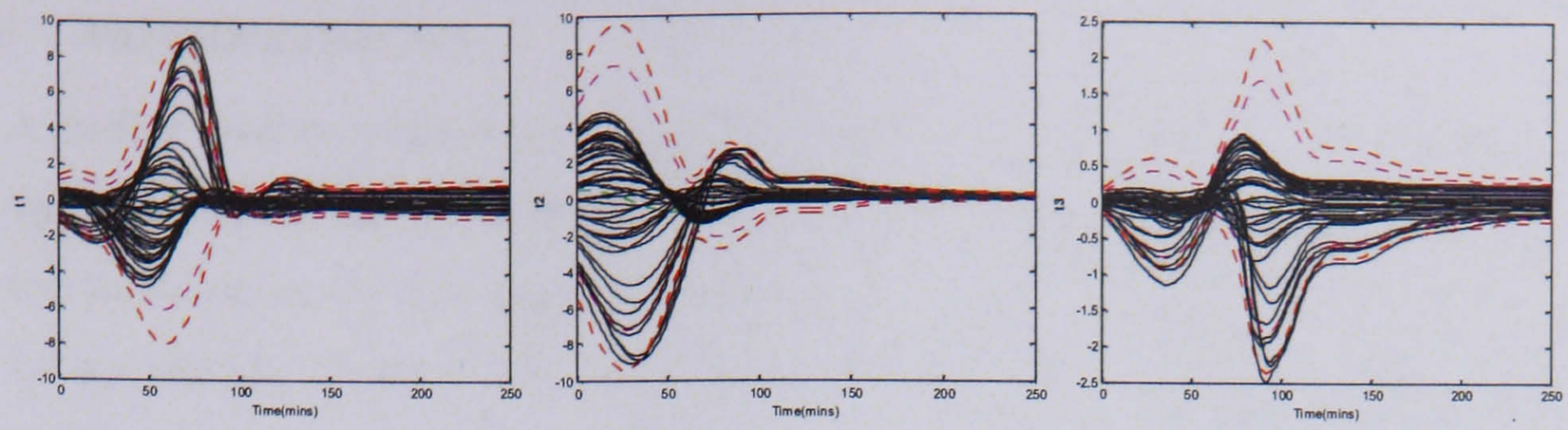
The advantage of using model-based PCA is that as well as being a through batch technique, it also addresses both the non-linear and dynamic aspects of a process through the use of a mechanistic model of the reaction which should capture these aspects of process behaviour. The theoretical rationale is that the mechanistic model of the process describes what is happening during the batch, including the dynamic and non-linear properties of the process, therefore when the model-predicted values of the variables are subtracted from the actual process variables, the dynamic and non-linear characteristics which are present in both should be removed, leaving only unstructured white noise which can be modelled. Any abnormalities entering the process will not have been captured by the mechanistic model and therefore should be present in the residuals and thus detectable.

The mechanistic model was solved at the same time intervals as the nominal batch data values were collected, and the model-predicted values for each variable trajectory were then subtracted from the actual values of the process variables to give the residuals that form the basis of the nominal data set. The residuals were then globally scaled, Wold *et al* (1998) before performing batch observation level analysis to give a nominal model. Table 4 gives the results for the cross validation analysis.

No ret PCs	R^2x	Cum. R^2x	Eigen-values	R^2y	Cum. R^2y	Q^2	Cum. Q^2
1	0.41	0.41	1.64	0.0136	0.0136	0.0136	0.0136
2	0.559	0.968	2.24	0.00111	0.0148	0.00113	0.0147
3	0.023	0.991	0.125	0.00526	0.02	0.00534	0.02

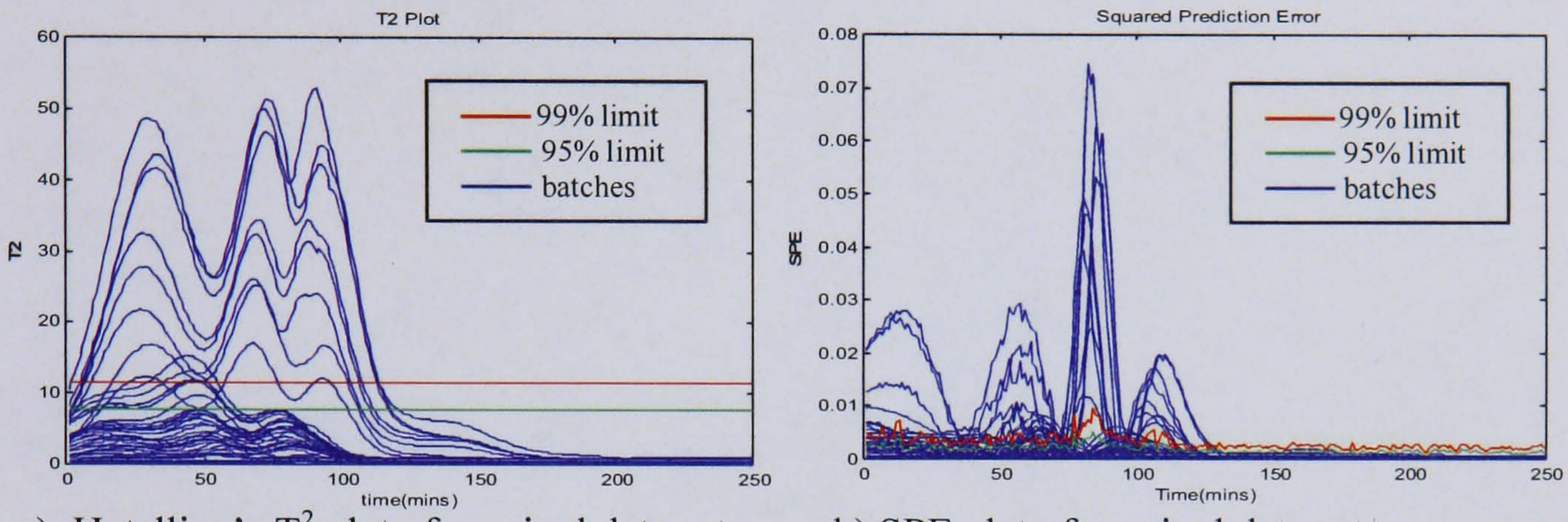
Table 4: Cross validation results for model-based PCA

The control charts for the nominal MBPCA model are shown in Figure 45, along with the Hotelling's T^2 and SPE plots in Figure 46.



a) Latent Variable 1 b) Latent Variable 2 c) Latent Variable3

Figure 45: MBPCA scores control charts for the nominal data set



a) Hotelling's T^2 plot of nominal data set b) SPE plot of nominal data set

Figure 46: Hotelling's T^2 and SPE plots of the nominal data set

From the MBPCA scores control charts, Figure 45, it can be observed that there is still strong structure in the data with respect to time, with a large amount of variation evident in the first 100 time points. It was expected that the application of the model-based PCA technique would remove this structure and it would not be seen so strongly in the model of the residuals. However, parameter errors were introduced to the data to ensure there was a plant model mismatch, therefore the model-based algorithm is not efficient in removing the structure inherent within the data. The Hotelling's T^2 and SPE plots also show the strong dynamic structure in the data at the start of the batch. Again it is noted that a significant number of batches exceed the control limits during the first 100 minutes of the batch, i.e. in excess of the one batch that would be expected for the 99% control limit and three batches for the 95% control limit in the nominal data set.

6.6 Validation Data Set

A further twenty batches were generated under normal operating conditions for the validation of the models built from the nominal data set. Using the validation data set, the performance of each model was examined with respect to the number of batches falling outside of the confidence limits. With the 95% and 99% confidence limits developed for the nominal models, it is expected that 5% of the batches would lie outside the 95% limits and 1% of the batches would be outside the 99% limits.

6.7 Multiway Principal Component Analysis

The projection of the twenty validation batches onto the multiway PCA scores plot is shown in Figure 47, with the Hotelling's T^2 and SPE plots in Figure 48 and Figure 49 respectively.

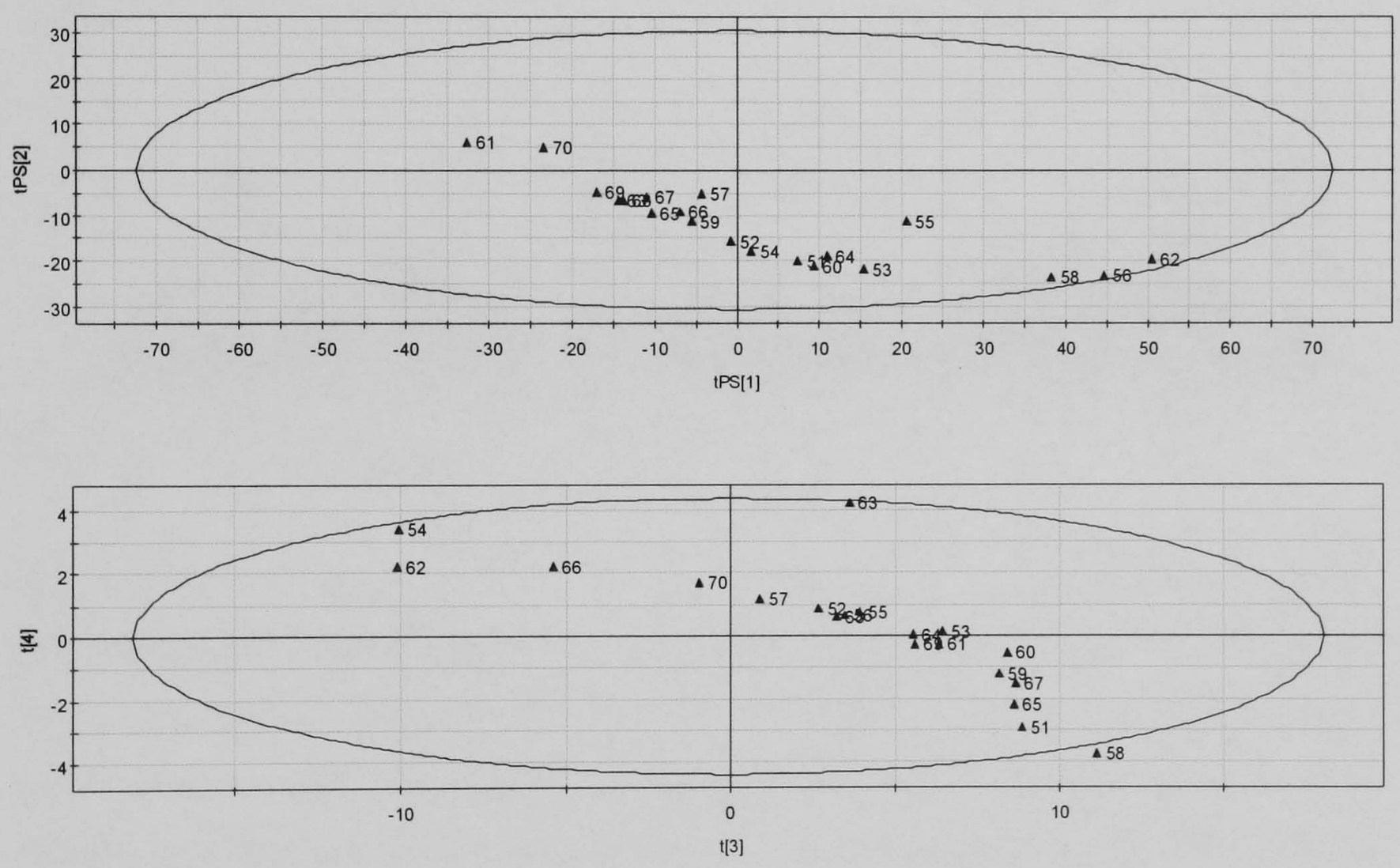


Figure 47: Validation batches projected onto MPCA nominal model

The bivariate MPCA scores plots in Figure 47 shows some structure in the batches as they all tend towards the bottom half of the chart, however the batch numbering indicates they are randomly placed and not following a particular trend.

The bivariate scores plots of the validation data show that 3 batches lie close to the limits. This is acceptable for the 95% confidence limit.

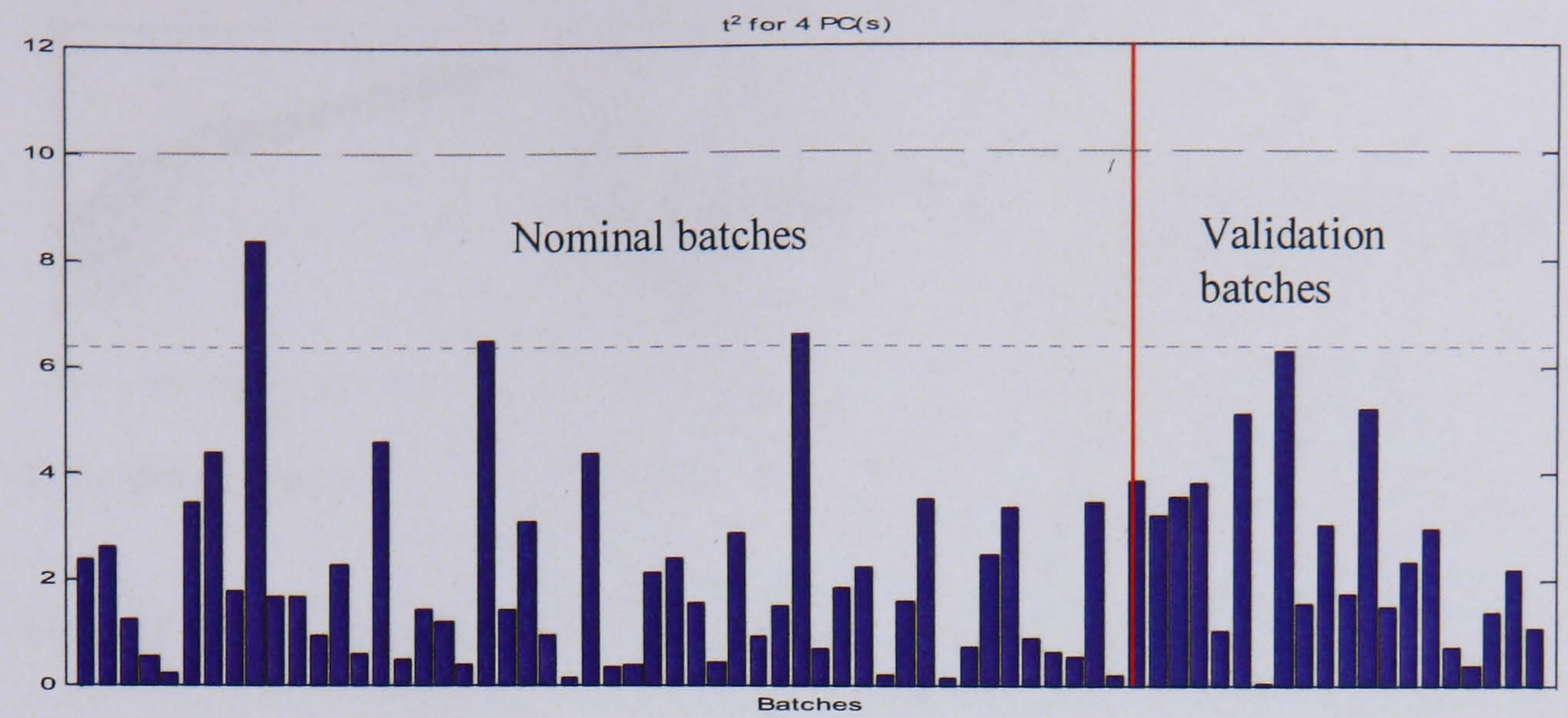


Figure 48: Hotelling's T^2 plot of the validation batches

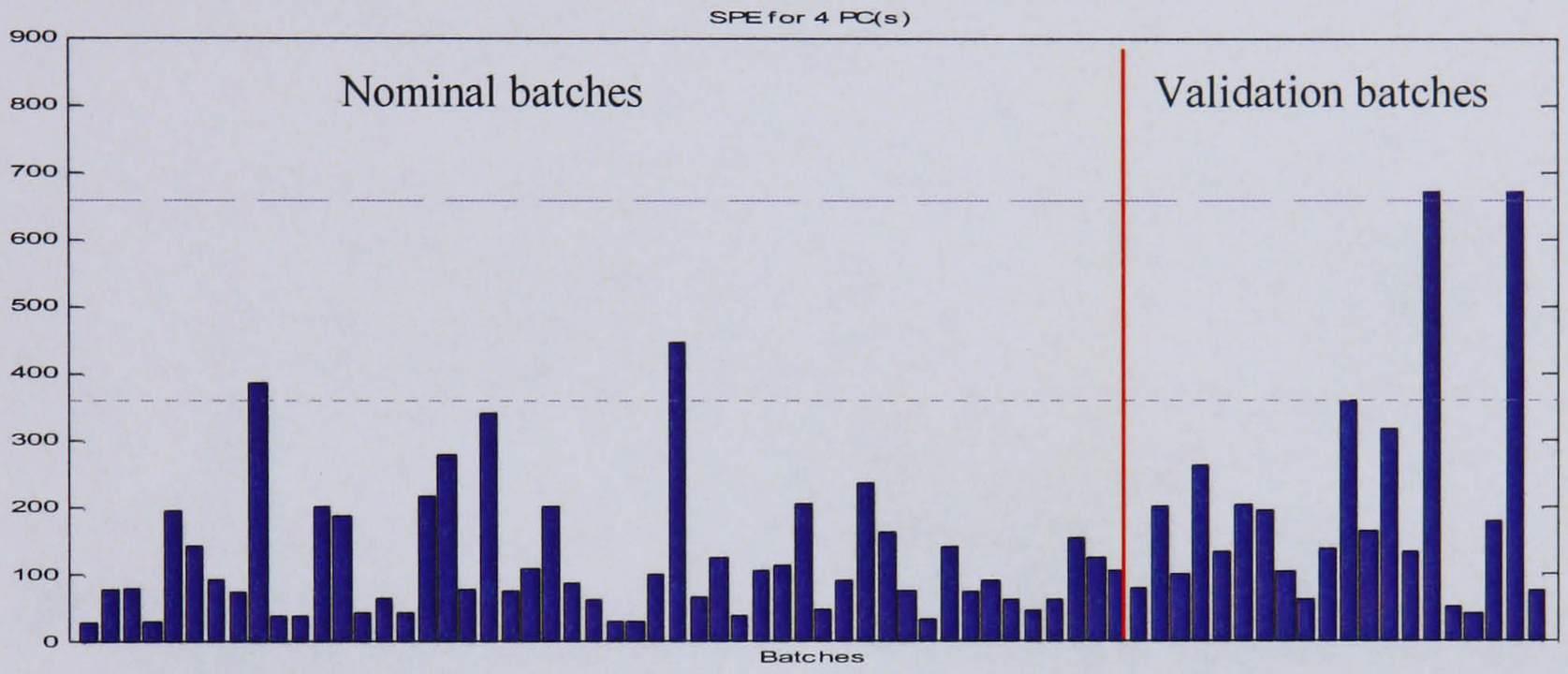
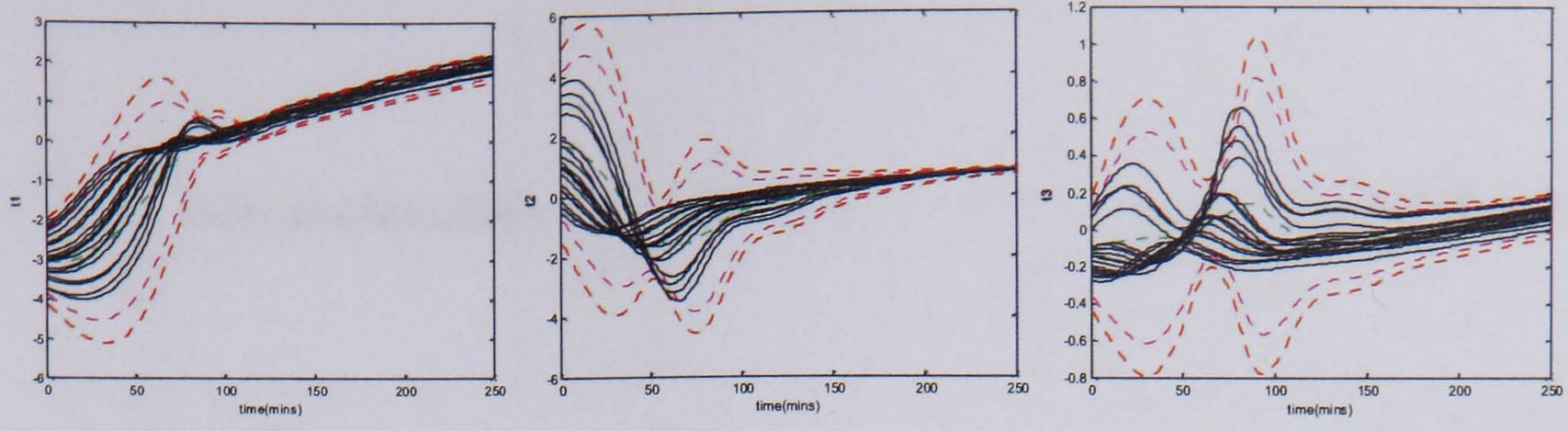


Figure 49: SPE plot of the validation batches

The Hotelling's T^2 plot shows that all validation batches lie within the control limits, indicating that the batches contain only normal common cause variation. The SPE plot in Figure 49 identifies 2 batches just exceeding the 99% limit. This usually indicates there may be some kind of change in the correlation structure of the variables. However these batches have been generated in the same way as the nominal data so it is likely the two batches have a higher prediction error due to random variation and in practice a batch can exceed the limits by chance.

6.7.1 Batch Observation Level Analysis

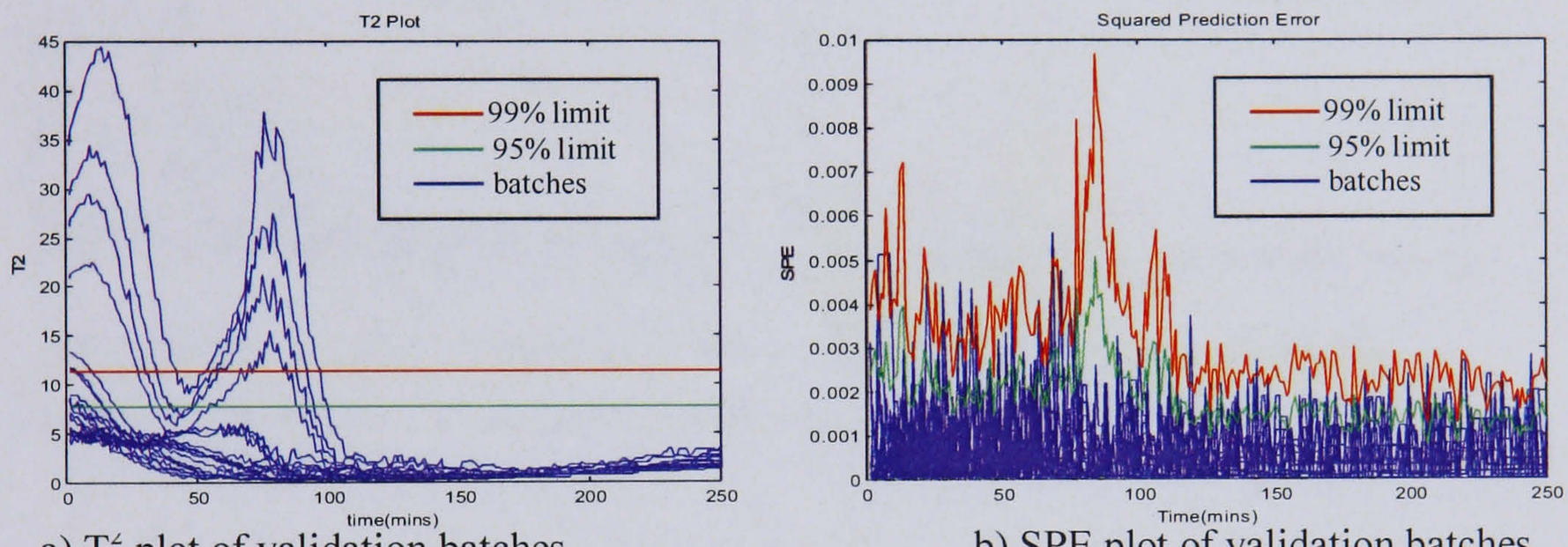
The projection of the twenty validation batches onto the batch observation level PCA scores plot is shown in Figure 50, and the Hotelling's T^2 and SPE plots are shown in Figure 51.



a) Latent Variable 1 b) Latent Variable 2 c) Latent Variable 3

Figure 50: Batch observation level scores plots of the validation batches

The projection of the validation batches onto the nominal control limits shows that one batch moves away from the limits, therefore the model is acceptable.



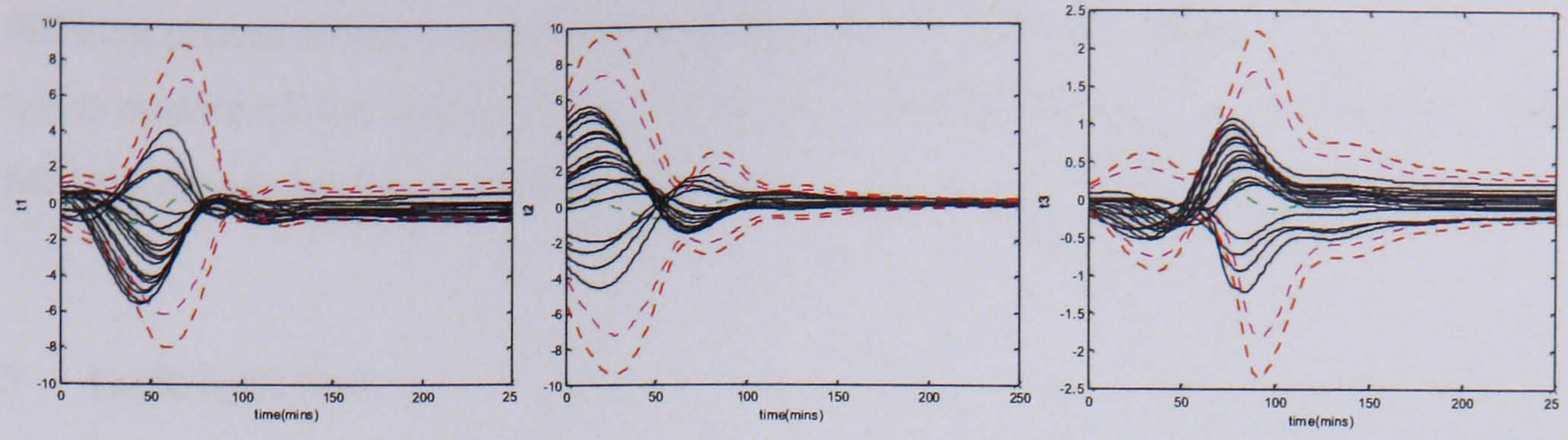
a) T^2 plot of validation batches b) SPE plot of validation batches

Figure 51: Hotelling's T^2 and SPE plots of validation batches

Four validation batches deviate from the Hotelling's T^2 control limits, as a consequence of the large amount of variation during the first part of the process, with one moving outside the SPE control limits. This indicates that the batch observation level model is acceptable, as the validation batches do not show any differences to common cause variation, other than at the start which was also seen with the nominal batches, or in the correlation structure between variables.

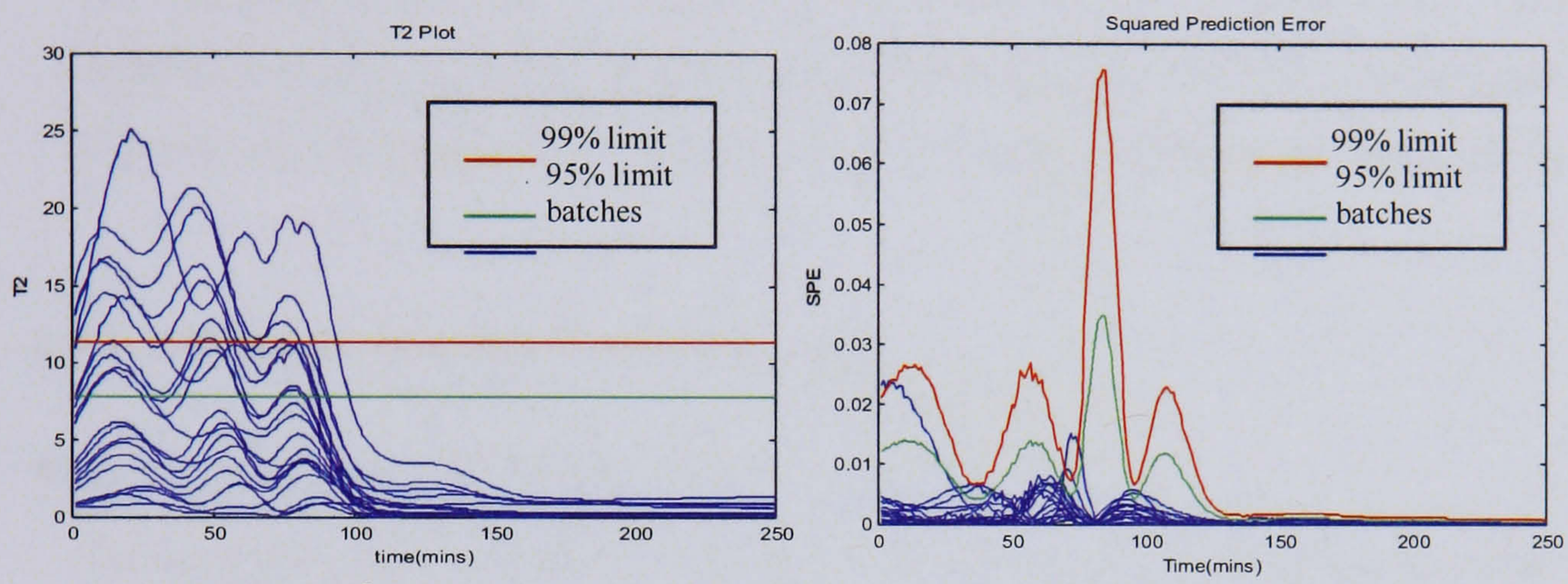
6.7.2 Model-based Principal Component Analysis

The projection of the twenty validation batches onto the model-based PCA scores plot is shown in Figure 52, with the Hotelling's T^2 and SPE plots in Figure 53.



a) Latent Variable 1 b) Latent Variable 2 c) Latent Variable 3

Figure 52: MBPCA - batch observation level plots for validation batches



a) Hotelling's T^2 plot for validation batches b) SPE plot for validation batches

Figure 53: Hotelling's T^2 and SPE plots for validation batches

Figure 52 shows that two validation batches move outside the scores control limits for the first latent variable at the start of the batch, and also on the third latent variable at around time point 75. Again the Hotelling's T^2 plot shows that several batches move outside the limits at the start of the process, as was observed with the nominal data set, whilst only one validation batch exceeds the SPE control limits, which is statistically acceptable.

6.8 Overall Conclusions

The three model types gave different results when tested using the validation batch data set. Multiway PCA shows all validation batches remaining inside the scores control limits, whereas with the batch observation level and model-based approaches there were one and two batches outside of the limits respectively. With the Hotelling's T^2 plots, all validation batches remained inside the control limits for MPCA, but for the BOL and MBPCA models, several batches lay outside the limits. This is because the control limits were calculated globally for the Hotelling's T^2 plots and therefore did not take into account the

different phases of batch operation. The SPE plots for BOL and MBPCA showed only one batch outside of the control limits, which is acceptable for the size of data set, whilst MPCA had two batches outside of the limits.

6.9 Fault Type One

Once the validity of each model type had been assessed, the fault detection abilities of the techniques were then compared. The first fault studied was a temperature sensor fault which occurred 100 minutes into the batch. In the analysis, since the first 50 minutes have been removed, the control charts will show the fault occurring after 50 minutes. In the following section, the 50 batches of fault type one are projected onto the three different nominal models developed using MPCA, BOL and MBPCA to investigate how effective each technique is with respect to detecting changes that occur in the process.

6.9.1 Multiway Principal Component Analysis

6.9.1.1 Fault Detection (Fault Type 1)

The bivariate scores control charts in Figure 54 show the fifty batches generated for fault type one projected onto the model, with Figure 55 and Figure 56 illustrating the Hotelling's T^2 and SPE charts.

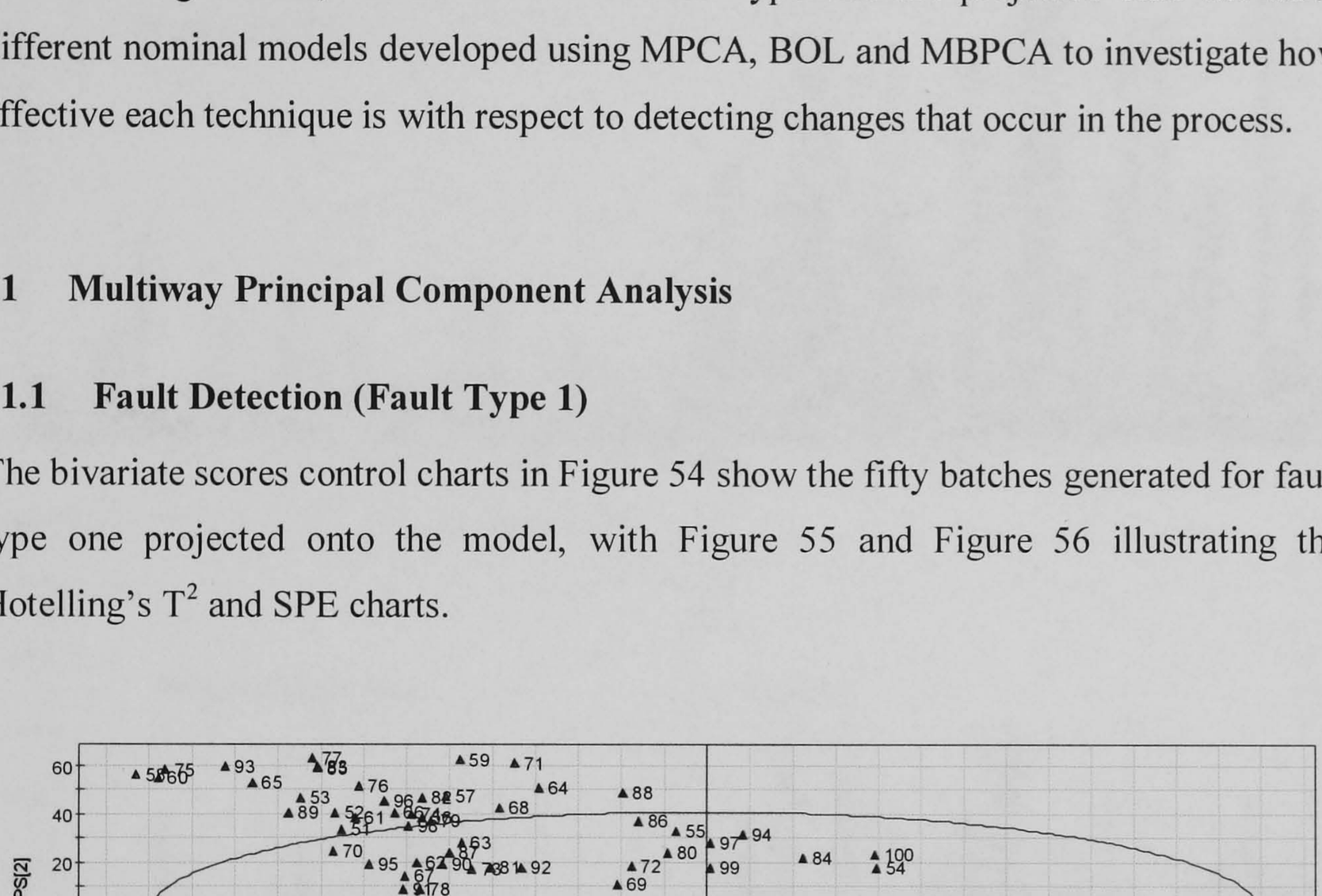


Figure 54: MPCA scores plots for fault type 1

Of the fifty batches of fault type one projected, twenty five batches are not detected as being abnormal by the MPCA control charts. In total 25 batches are observed to be out of control for the first two principal components. This means that approximately half the batches lie within the control limits of the first 2 principal components. The plot of principal components 3 and 4 shows approximately ten batches lying outside of the control limits, of these batches five were also lying outside of the limits on principal components 1 and 2. Therefore in total 31 batches were detected as being abnormal using the MPCA bivariate scores plots.

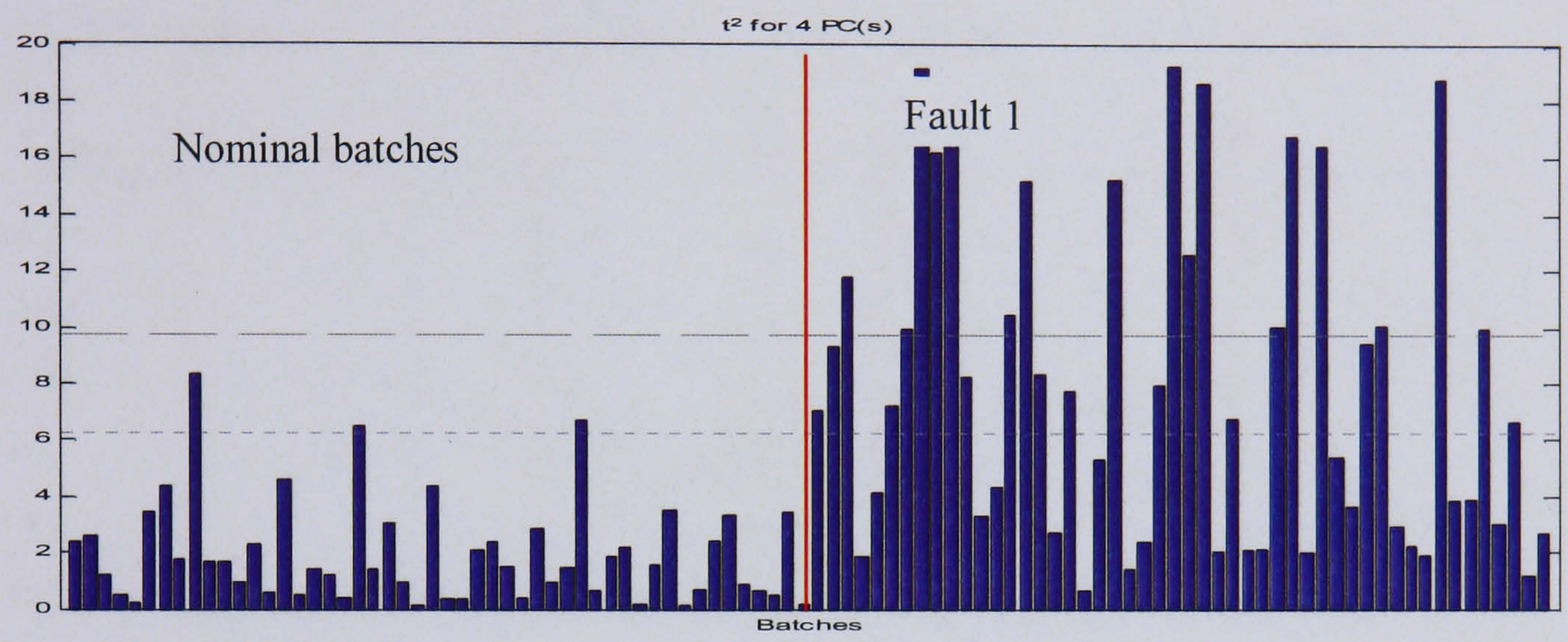


Figure 55: Hotelling's T^2 plot of fault type 1

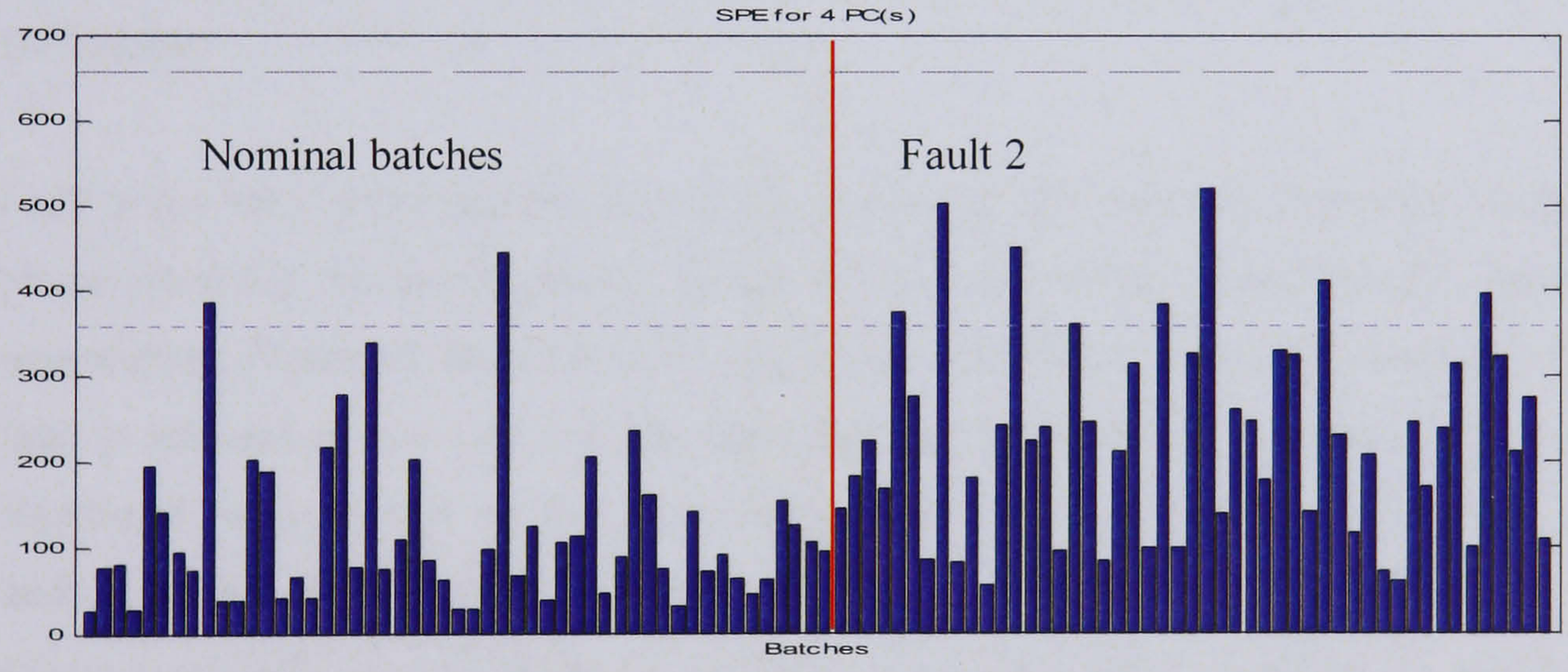


Figure 56: SPE plot of fault type 1

The Hotelling's T^2 plot, Figure 55, shows that 17 batches exceed the control limits (99%), with only 8 batches exceeding the SPE control limits (95%). For fault type 1, fault detection by MPCA is not successful. Although the technique deals with the non-linear behaviour in the data through the application of scaling, it does not capture the dynamic behaviour. This may be why so few faulty batches have been detected by the technique. Alternatively, the faults may be too subtle to be differentiated from the well-behaved batches, this is discussed further in subsequent sections.

6.9.1.2 Fault Diagnosis (Fault Type 1)

To determine the cause of the outlying batches, contribution plots for a number of the batches that lie outside of the confidence limits for fault type one were examined using contribution plots, as discussed in Chapter 2. Figure 54 showed batch 93 lying outside the confidence limits for principal component 2. The corresponding contribution plot is shown in Figure 57.

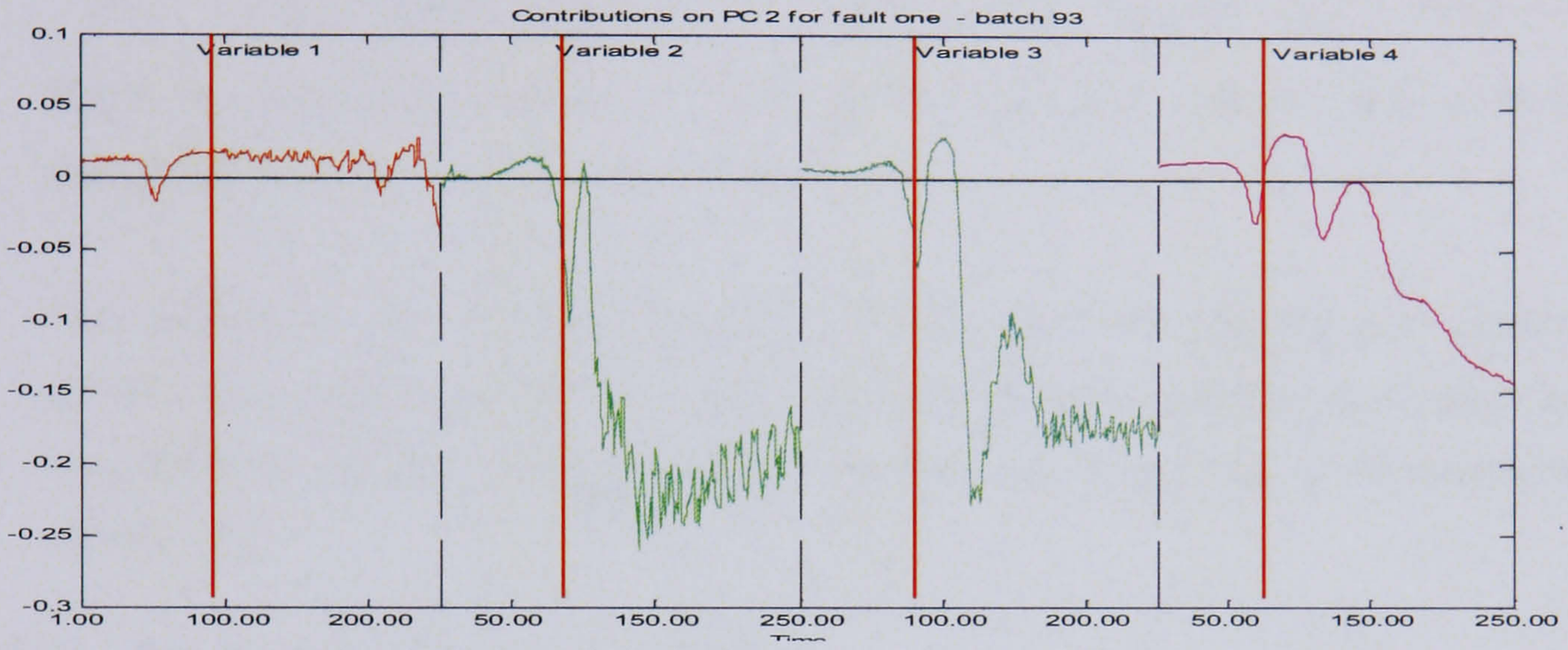


Figure 57: Contribution plot for PC2 for batch 93 (red lines indicate where the fault is introduced)

Fault 1 is a fault in the reactor temperature sensor, variable one, and therefore it would be expected that this would cause a change in the contributions in variable one, reactor temperature. Figure 57 shows a small dip in the contribution of variable one after the fault is introduced, however it is the other three variables whose contribution is more significant with respect to the batch lying outside of the control limits. This is understandable as the exothermic batch simulation uses the values of variable one to calculate variables two and three, the wall and jacket temperature, therefore the effect of a fault in variable one would impact on the other parameters.

In Figure 54, batch 95 lies outside the control limits for principal components 3 and 4, the corresponding contribution plot for this batch is shown in Figure 58:

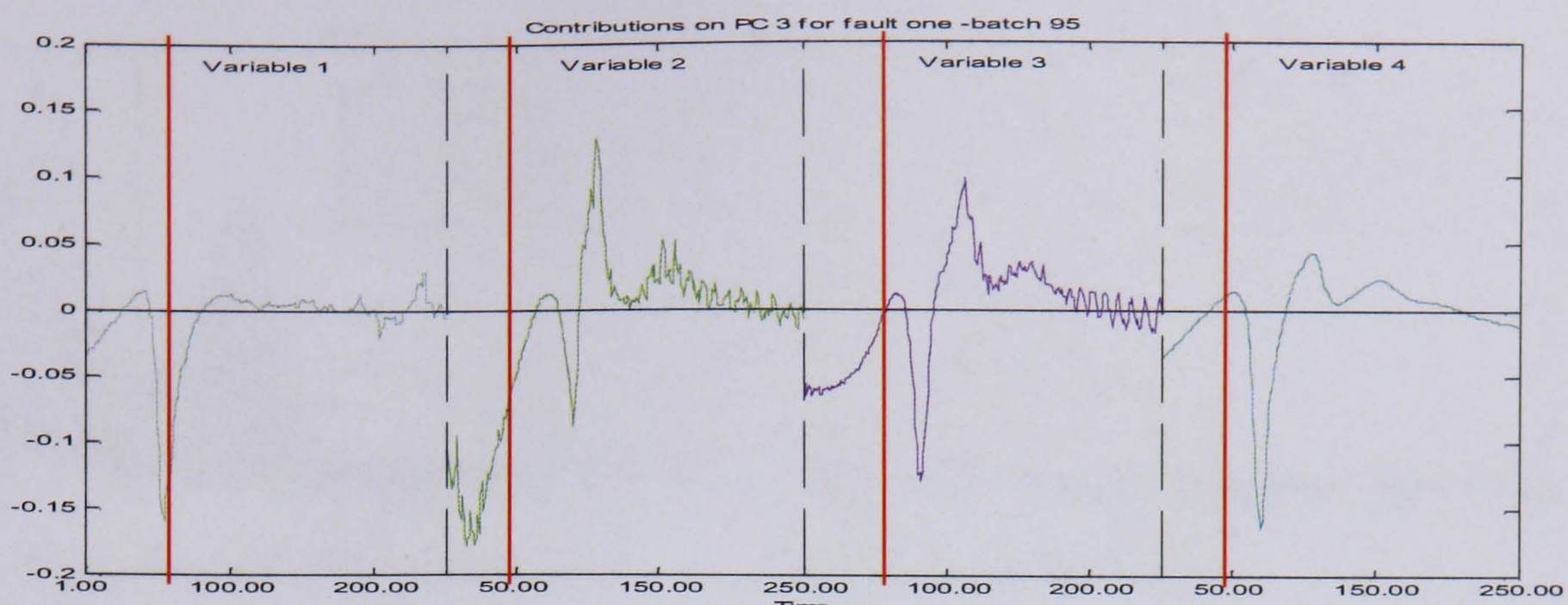


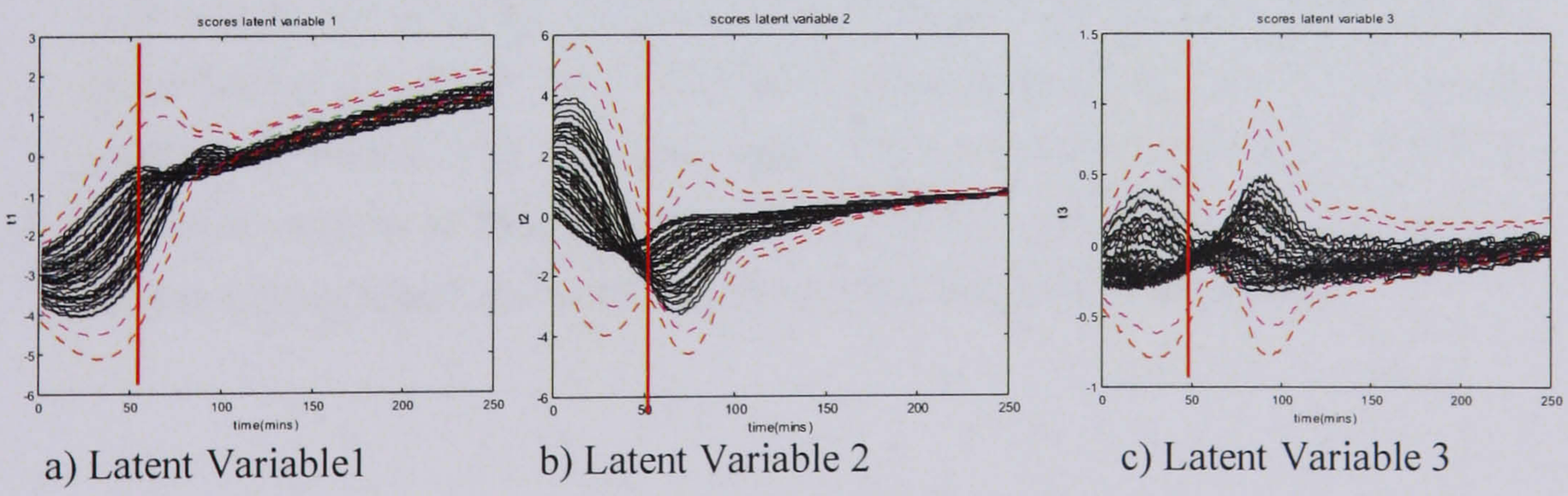
Figure 58: Contribution plot for PC3 for batch 95 (red lines indicate where fault is introduced)

The contribution plot in Figure 58 shows a large dip in the variable one (reactor temperature) contributions after the reactor temperature fault is introduced. Changes in the other three variables are observed after this time point, as they react to the change in variable one.

6.9.2 Batch Observation Level Analysis

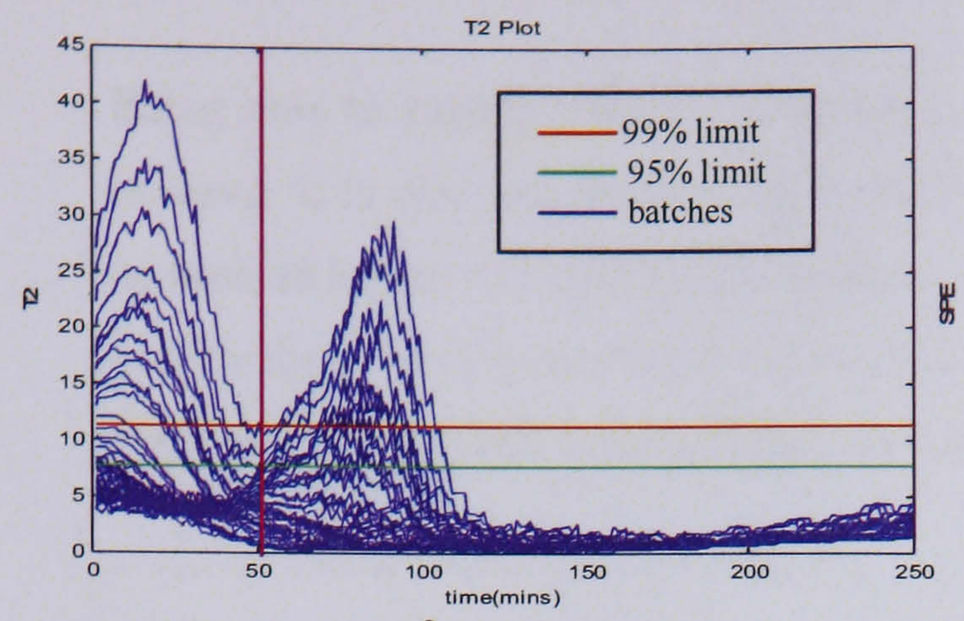
6.9.2.1 Fault Detection (Fault Type 1)

Using the nominal model, the fault type 1 data set was projected onto the batch observation level control chart (the point of fault introduction is indicated by the red lines) The resulting plots are shown in Figure 59, followed by the Hotelling's T^2 and SPE plots (Figure 60).

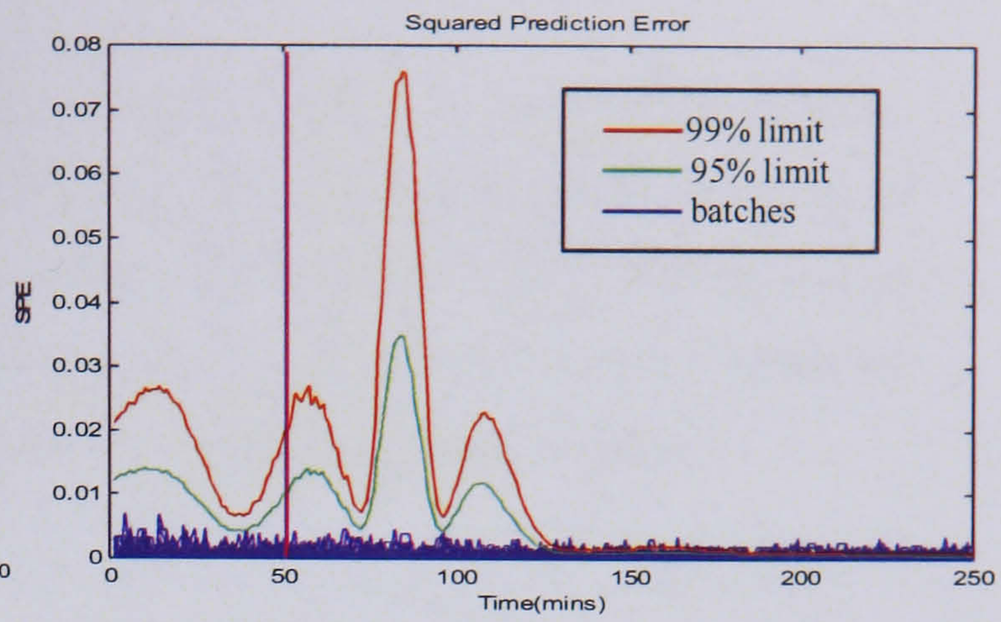


a) Latent Variable1 b) Latent Variable 2 c) Latent Variable 3

Figure 59: Batch observation level control charts for fault type 1



a) Hotelling's T^2 plot for fault type 1



b) SPE plot for fault type 1

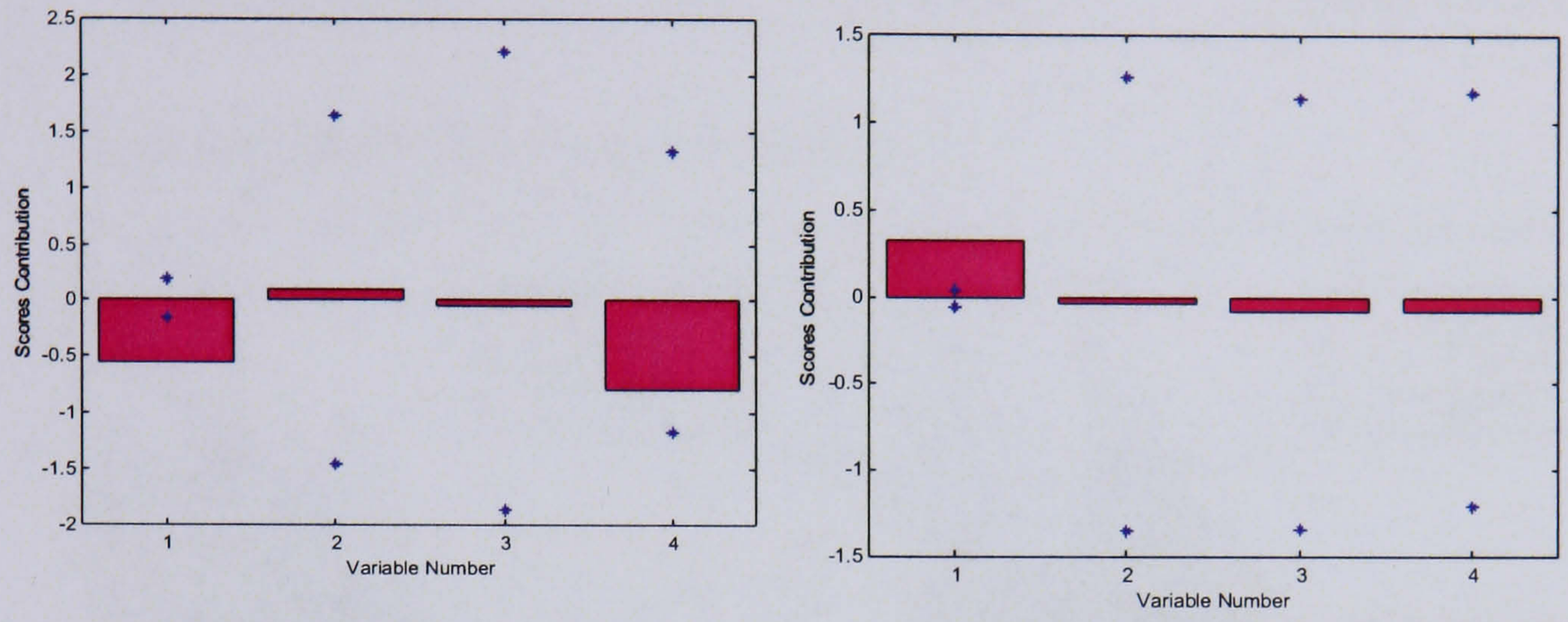
Figure 60: Hotelling's T^2 and SPE plots for fault type 1

The fault type one batches start to deviate from the control limits after time point 90 in latent variable 1, approximately 40 time points after the fault is first introduced. The average out of control run length (ARL) has been calculated and is reported at the end of this section. The ARL is calculated as the time between the introduction and the detection of the fault and is calculated for each of the 50 batches and then averaged. Latent variable 2 does not detect any batches as being abnormal, whilst for latent variable 3 some of the batches breach the control limits towards the end of the batch. Approximately 60% of the batches were detected as being abnormal using the BOL technique.

The Hotelling's T^2 plot of the fault one data set was also investigated. These plots are generally used to detect if the variation in the experimental data (the fault data in this case) is greater than normal common cause variation. From Figure 60, it is not possible to determine which batches are outside of the control limits due to the fault since many of the batches are outside the limits prior to the introduction of the fault. This was also noted in the nominal and validation model and is partially a consequence of the large amount of variation in the data at this time. The SPE plot in Figure 60 does not show any batches moving outside the control limits until approximately time point 125.

6.9.2.2 Fault Diagnosis (Fault Type 1)

Being able to rapidly detect a batch going out of control is one part of the problem, however it is also important to be able to isolate the cause of the batch going out of control. In Figure 61 contribution plots are shown for one of the batches that has moved outside the control limits in the bivariate scores plots, at the point just after it leaves the limits (100 mins), and then later on in the batch (140 mins) for latent variable 1.



(a) Time point 100 (b) Time point 140

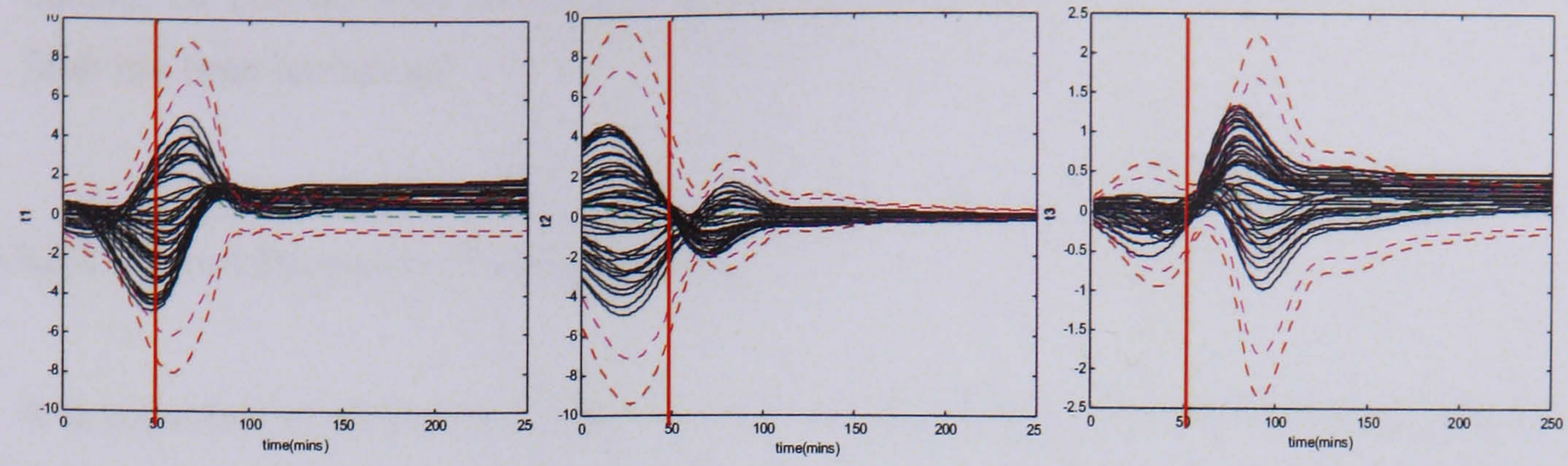
Figure 61: Contribution plot for out of control batch for fault type 1

The blue stars indicate the confidence limits for each variable, as calculated from the nominal data set. Figure 61(a) shows that variable 1, the reactor temperature, has the greatest contribution to the batch exceeding the control limits. Figure 61(b) shows that variable 1, the reactor temperature, is still making a significant contribution to the batch being outside the limits, but the contributions from the other variables have decreased. These charts indicate that the reactor temperature is the cause of the batches lying outside of the control limits, which is the variable that was modified in the simulation study.

6.9.3 Model-based Principal Component Analysis

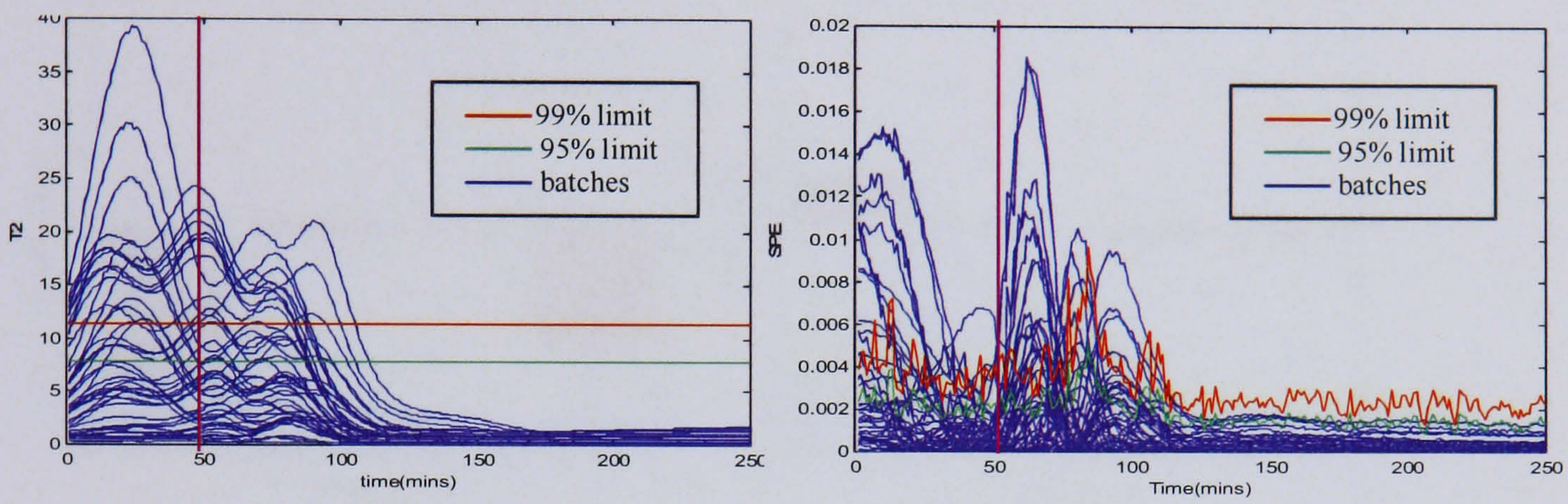
6.9.3.1 Fault Detection (Fault Type One)

The model-based principal component analysis algorithm was applied to the data set of batch data containing fault type one. To do this, the values from the solved mechanistic model were subtracted from each abnormal batch to create a set of residuals, it is these residuals that were projected onto the MBPCA control limits calculated from the residuals of the nominal data set. The resulting control charts are in Figure 62, and the associated Hotelling's T^2 and SPE plots in Figure 63.



a) Latent Variable 1 b) Latent Variable 2 c) Latent Variable 3

Figure 62: MBPCA control charts for fault type 1



a) Hotelling's T^2 plot of fault type one b) SPE plot of fault type one

Figure 63: Hotelling's T^2 and SPE plot of fault type one

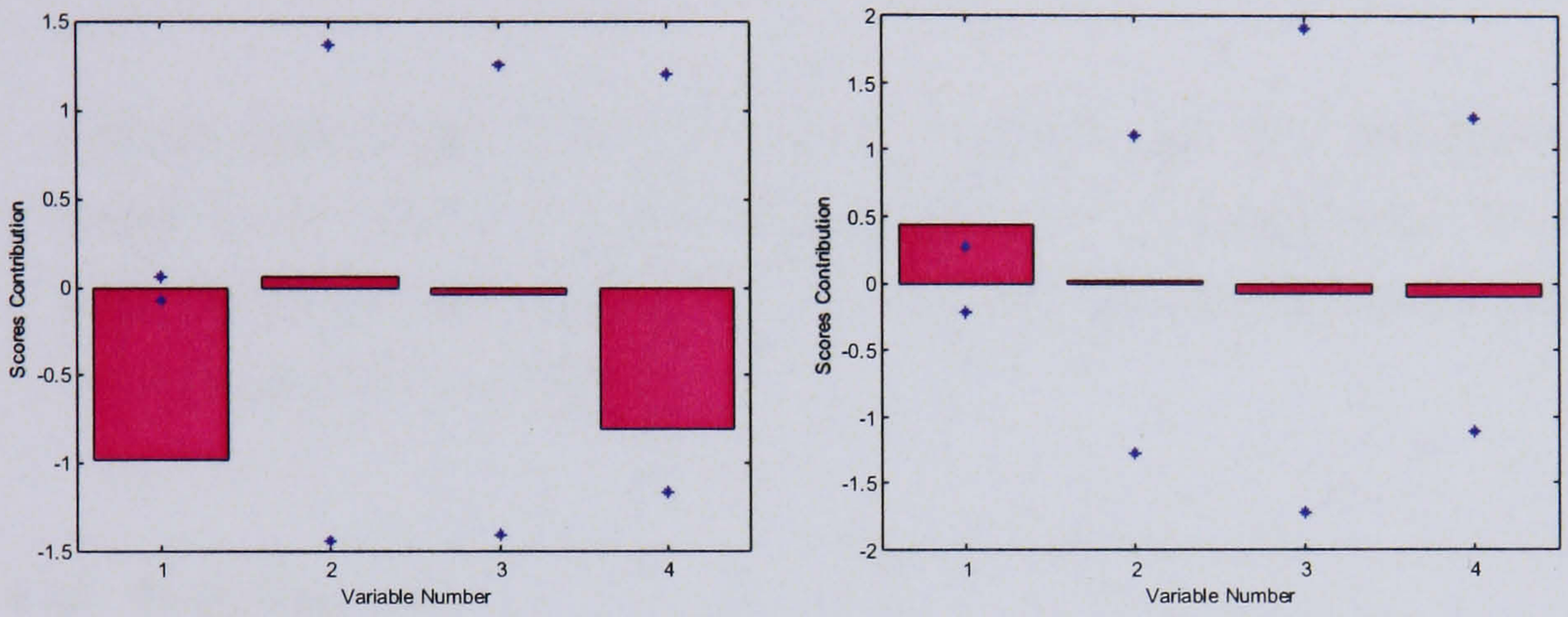
Although the MBPCA control charts in Figure 62 show the scores calculated from the residuals of the batches, it can be observed that a large amount of structure remains in the data. This can be attributed to the fact that the mechanistic model used to create the residuals is not a perfect match of the process, therefore structure remains. However, the out of control average run length for fault type one for the MBPCA model is approximately 64 minutes for latent variable 2. This is compared with the other batch analysis techniques at the end of the chapter.

The Hotelling's T^2 plot for fault type one, Figure 63, does not show a change in the behaviour after the fault is introduced, since a number of batches are already outside of the control limits and no additional batches exceed the limits after the fault has been introduced. The SPE chart in Figure 63 also has some batches already outside of the control limits prior to the introduction of the fault, this is partially due to the noise in the data and also due to the non-linear behaviour as the covariance structure differs between the different phases of the batch After the fault is introduced at time point 50, several

batches do deviate from the control limits, approximately 10-20 time points after the fault has been introduced.

6.9.3.2 Fault Diagnosis (Fault Type One)

It is important to determine whether the cause of fault type one can be determined from the model-based technique. Contribution plots were calculated for a typical out of control batch for fault type one just after the trajectory moves outside the control limits at time point 100, and then 50 minutes later at time point 150, Figure 64:



a) Time point 100

b) Time point 150

Figure 64: Contribution plot for out of control batch – fault type 1

Figure 64(a) shows that variable 1, the reactor temperature, has the most significant effect in terms of the batch leaving the control limits. Figure 64(b) shows the contribution plot 50 minutes later (100 time points after the fault was first introduced). This plot shows that the reactor temperature is responsible for the batch deviation from the normal operating limits. Therefore it is possible to tell from the MBPCA contribution charts which variable is responsible for fault type 1.

6.9.4 Summary of Fault Type One

The fault detection and fault diagnosis results for fault type one are summarised for each of the batch monitoring techniques in Table 5, and the results are discussed in detail at the end of the chapter. The Hotelling's T^2 statistic has not been included in this analysis because it is uninformative due to the number of batches outside of the control limits prior to the fault being introduced.

Technique	% Batches Detected (control chart)	% Batches Detected (SPE)	ARL (control chart)	ARL (SPE)
MPCA	54	16	N/A	N/A
BOL	57	20	83	63
MBPCA	86	32	64	127

Table 5: Summary of fault detection results

It can be observed from Table 5 that MBPCA gives the best fault detection results with respect to the number of batches detected. The ARL for fault type 1 is shorter for MBPCA when examining the scores control charts, but the BOL ARL is shorter when examining the SPE control charts.

6.10 Fault Type Two

The second fault type considered in the investigation of the batch monitoring techniques was the decrease in the heat transfer coefficients, which indicates that some form of fouling has occurred. Again, 50 batches of this fault were generated using the exothermic batch simulation, and the data set was then studied using each of the monitoring techniques.

6.10.1 Multiway Principal Component Analysis (Fault Type Two)

6.10.1.1 Fault Detection (Fault Type Two)

The 50 abnormal batches were projected onto the MPCA control limits generated from the nominal data set to investigate the effect of the changes in the process on the monitoring technique. Figure 65 shows the bivariate scores control charts, followed by the Hotelling's T^2 and SPE plots in Figure 66 and Figure 67 respectively.

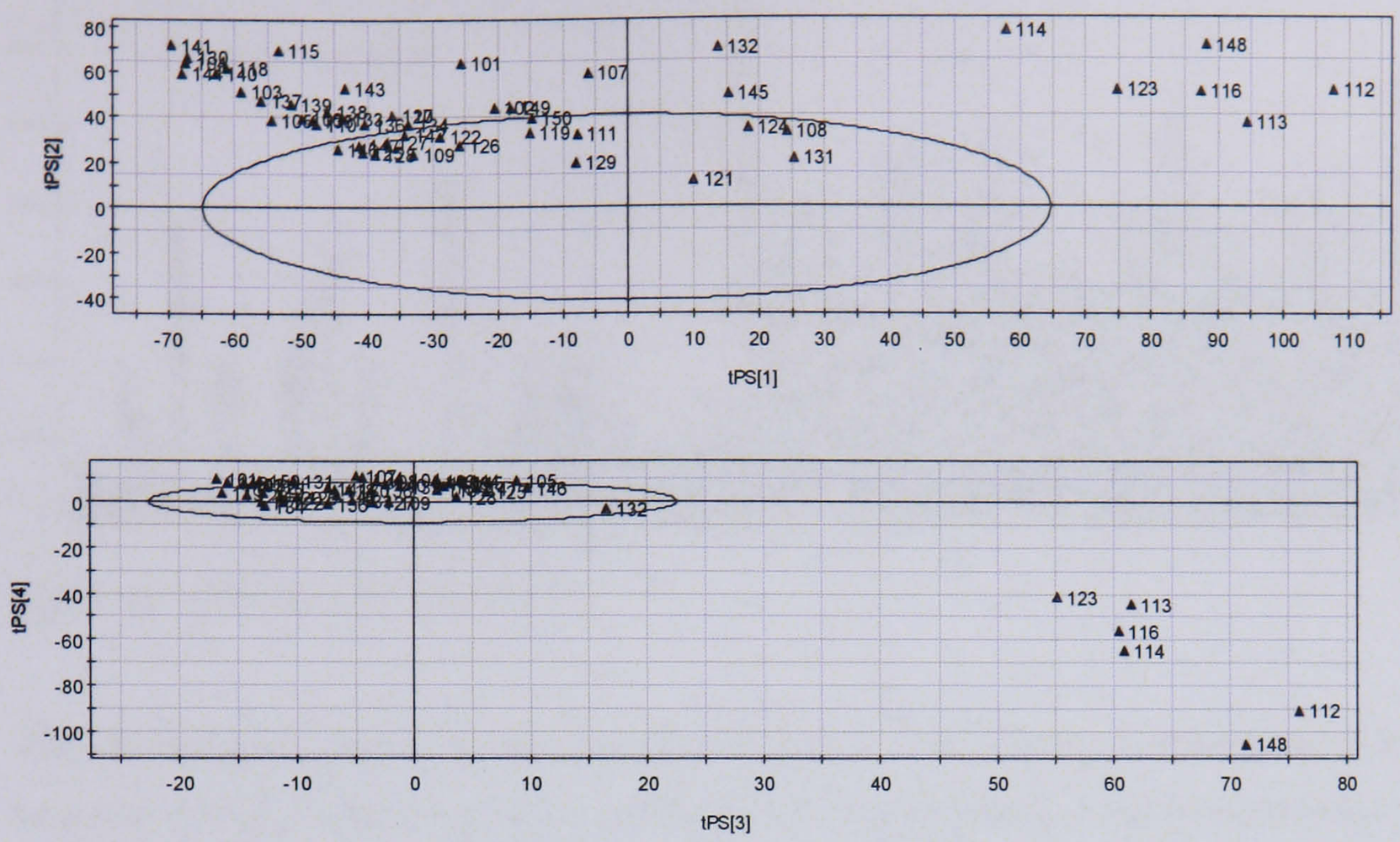


Figure 65: MPCA control charts for fault type 2

The MPCA technique is more successful at detecting the fault type 2 batches than the fault type 1 batches, with approximately 75% of the abnormal batches lying outside of the control limits, the majority being detected from principal component 2.

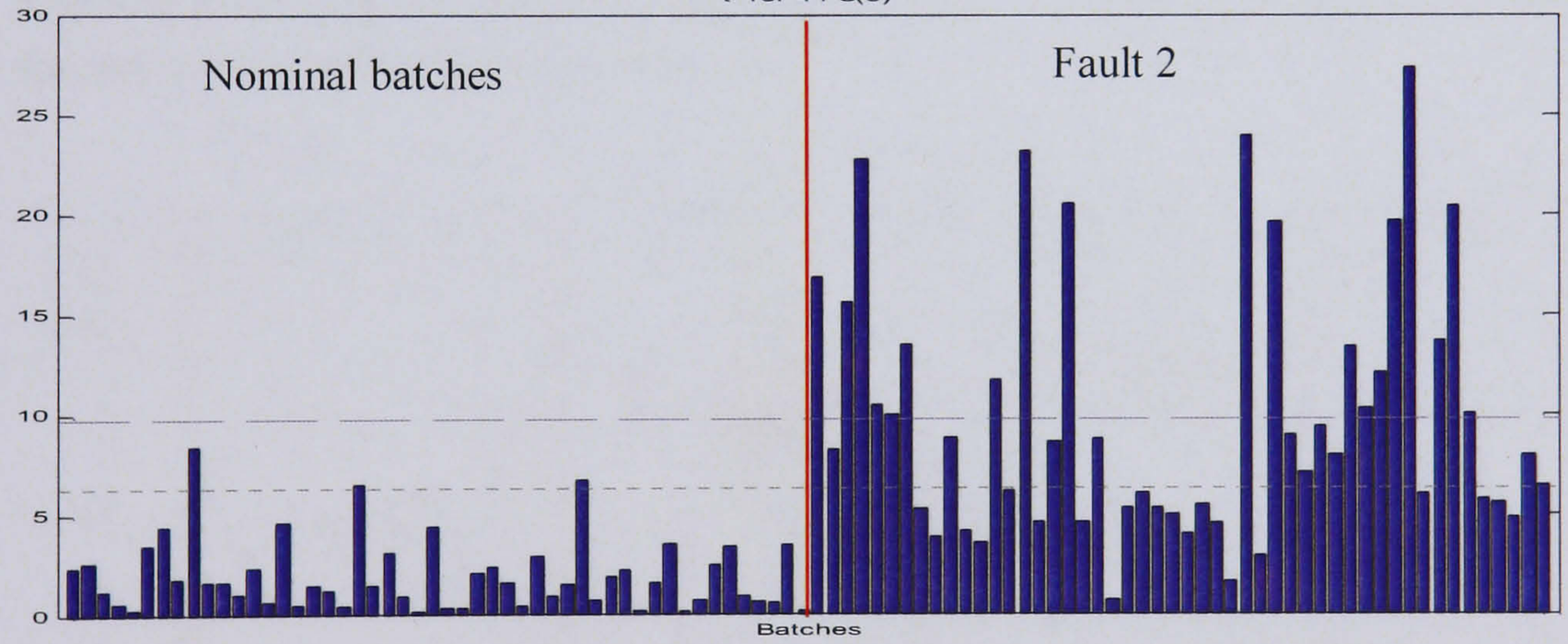


Figure 66: Hotelling's T^2 plot for fault type 2

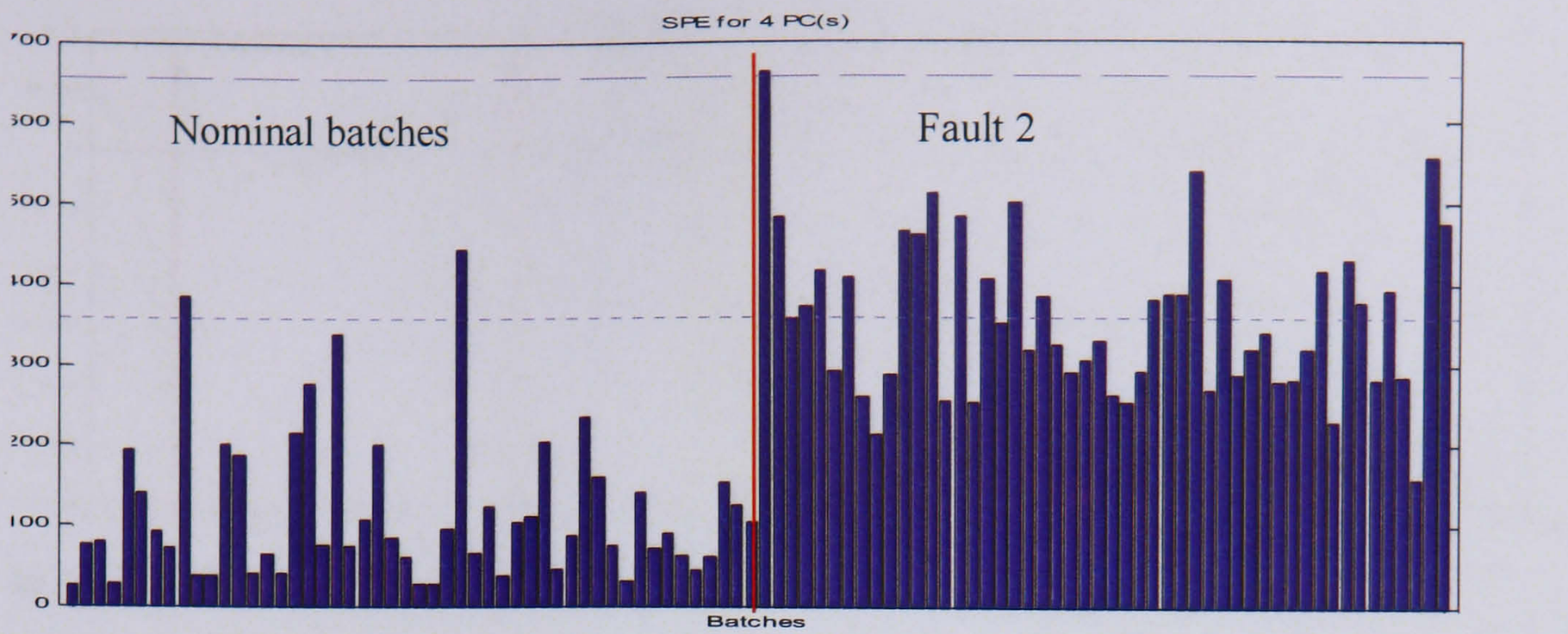
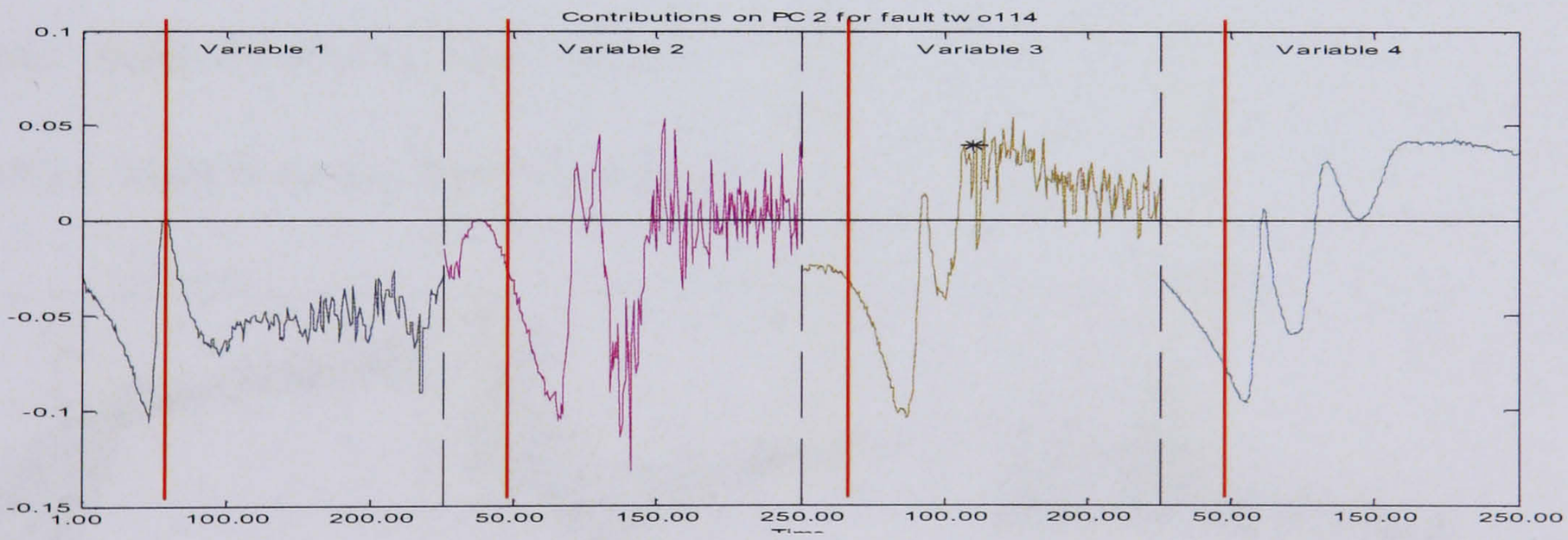


Figure 67: SPE plot for fault type 2

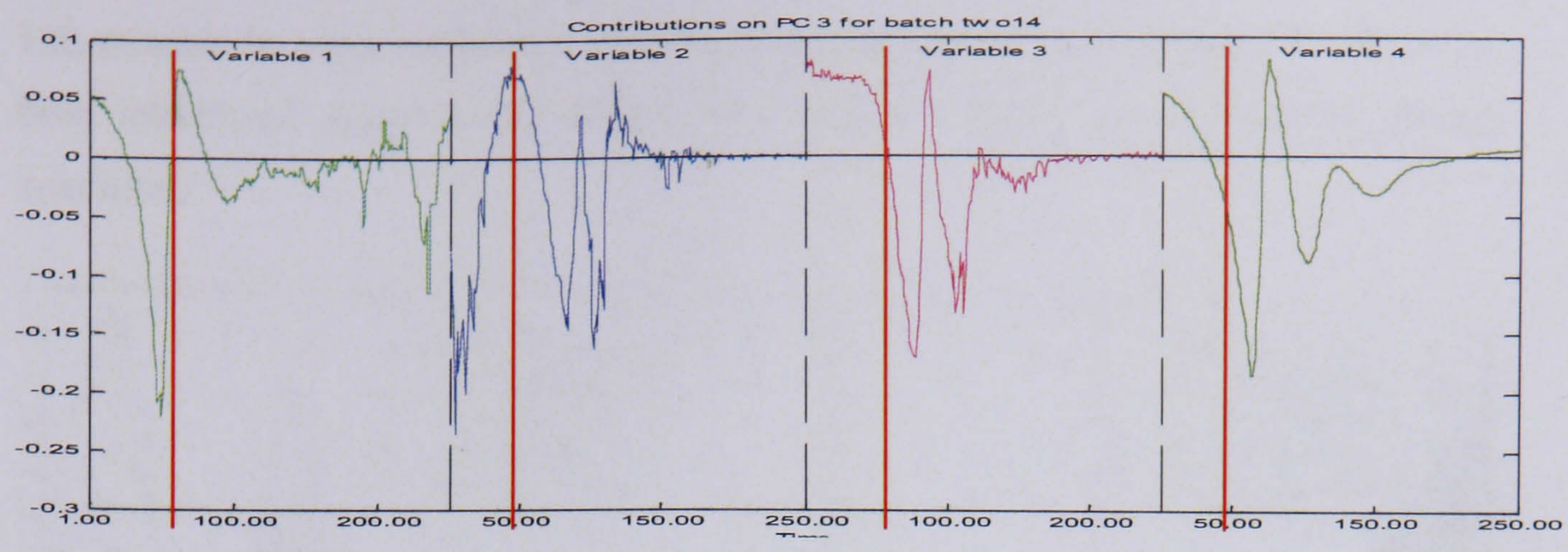
The Hotelling's T^2 and SPE plots shown in Figure 66 and Figure 67 respectively are not as successful as the bivariate scores control charts at detecting the out of control batches, however the Hotelling's T^2 plot does detect 19 batches as abnormal and the SPE, 23 batches, in total 31 batches lie outside the limits.

6.10.1.2 Fault Diagnosis (Fault Type Two)

To determine the cause of the outlying batches, a contribution plot was generated for outlying batch, number 114, for principal components 2 and 3, as it exceeded the limits for both these components (Figure 68).



a) Contribution plot for batch 114, PC2



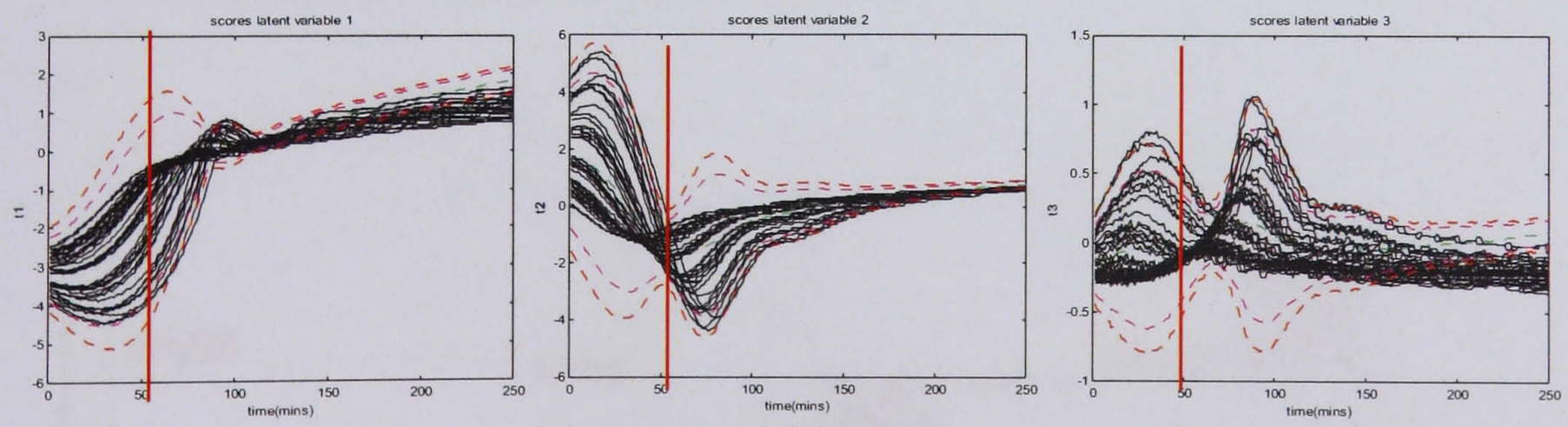
b) Contribution plot for batch 114, PC3

Figure 68 : Contribution plot for batch 114

Fault type two was a decrease in the heat transfer coefficients and was introduced at time point 100. It would be expected that this fault had an effect on the temperature variables (variables 1-3). The contribution plot for principal component 2 show variables 2 and 3, the wall temperature and the jacket temperature, having the most significant effect on the behaviour of the batches, and variable 4, the cooling valve position, also exhibits a change in behaviour after the fault has been introduced, which is expected as the cooling valve reacts to changes in the reactor temperatures.

6.10.2 Batch Observation Level Analysis

6.10.2.1 Fault Detection (Fault Type Two)

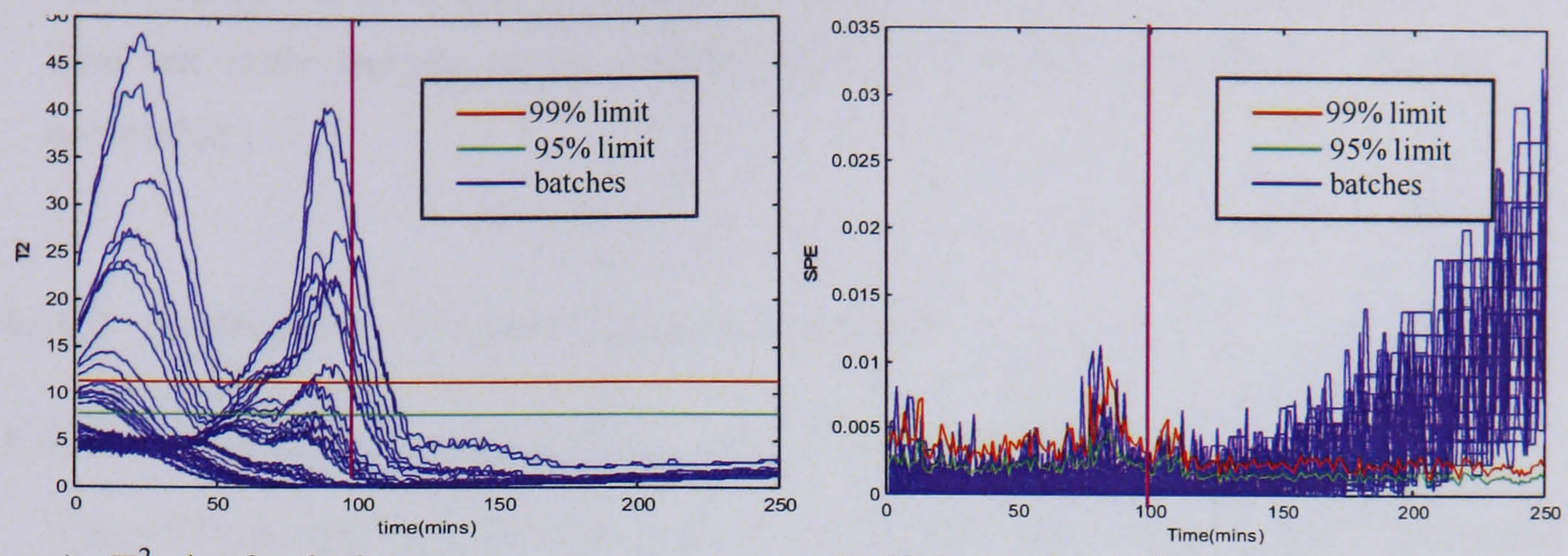


a) Latent Variable 1 b) Latent Variable 2 c) Latent Variable 3

Figure 69: Batch observation level control charts for fault type 2

Figure 69 shows the control charts for the second fault type projected onto the BOL nominal control limits. The second fault type, the decrease in heat transfer coefficient indicates that the effect of fouling can be first detected in some of the batches at around

120 minutes in latent variable 1. This is approximately 20 minutes after the fault has been introduced. Again latent variable two does not detect any changes in process operation.



a) T^2 plot for fault type two

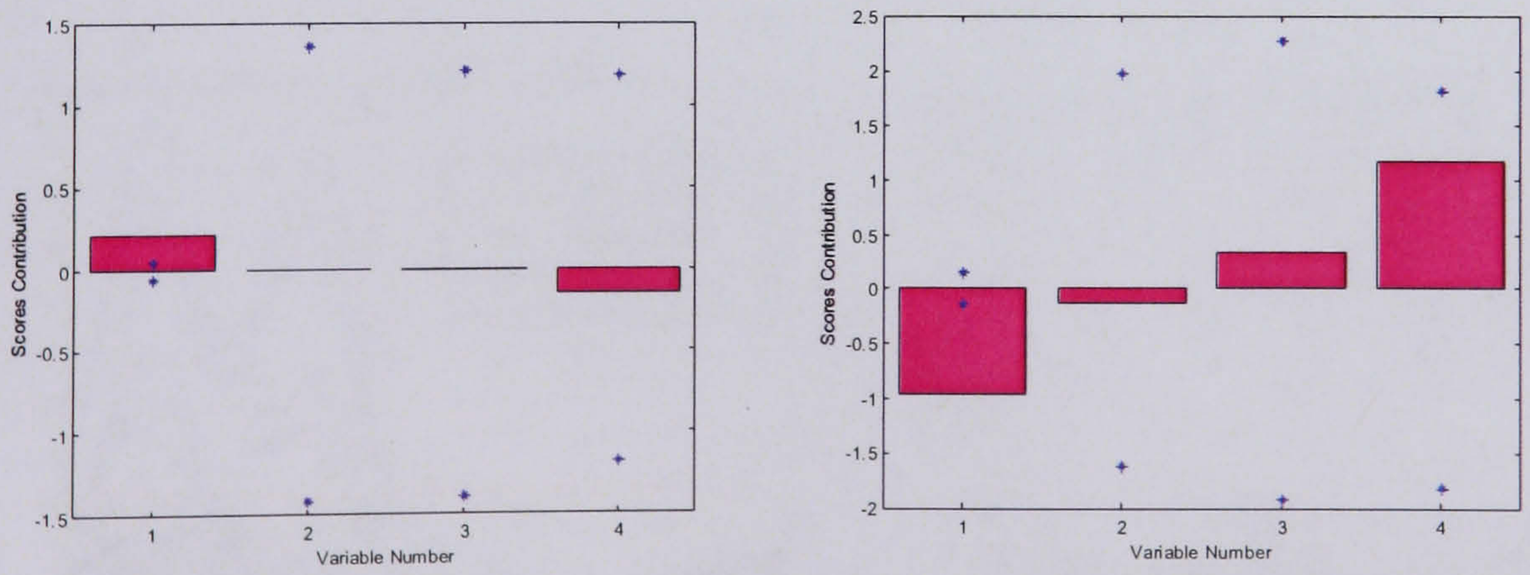
b) SPE plot for fault type two

Figure 70: Hotelling's T^2 and SPE plot for fault type two

The Hotelling's T^2 and SPE plots are shown in Figure 70. The Hotelling's T^2 plot does not detect a change in batch behaviour after the fault has been introduced. This is in contrast to the SPE plot which shows all 50 batches move outside of the control limits after the fault has occurred at time point 100.

6.10.2.2 Fault Diagnosis (Fault Type Two)

The batch observation level technique was fairly successful in terms of the detection of the second fault type, however it is also important to isolate the cause of the fault. In Figure 71 the contribution plots for a batch that moved outside of the control limits are shown, first at the time point after the batch has left the normal operating limits, and then 50 minutes further on into the batch.



a) Time point 160

b) Time point 210

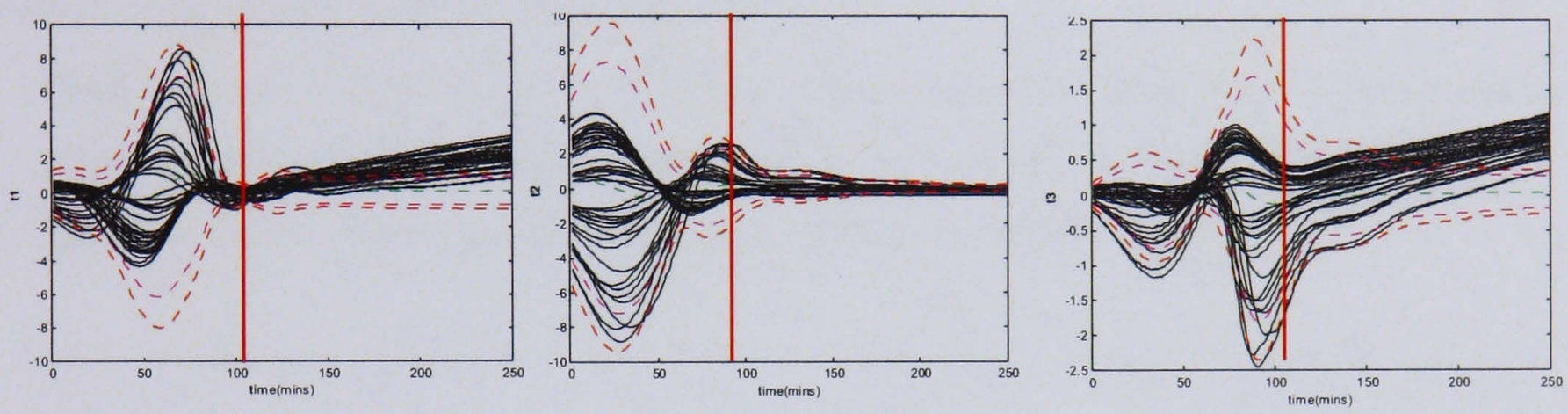
Figure 71: Contribution plot for out of control batch for fault type 2

The contribution plot for fault type two shows variable 1, reactor temperature, as having the most effect on the batch moving outside of the control limits, as the batch progresses the effect of the valve position also becomes more significant, as the split range control valve attempts to deal with the effect of the change in heat transfer coefficient, but it does not move outside of the contribution limits, which indicates its behaviour is acceptable.

6.10.3 Model-based Principal Component Analysis

6.10.3.1 Fault Detection (Fault Type Two)

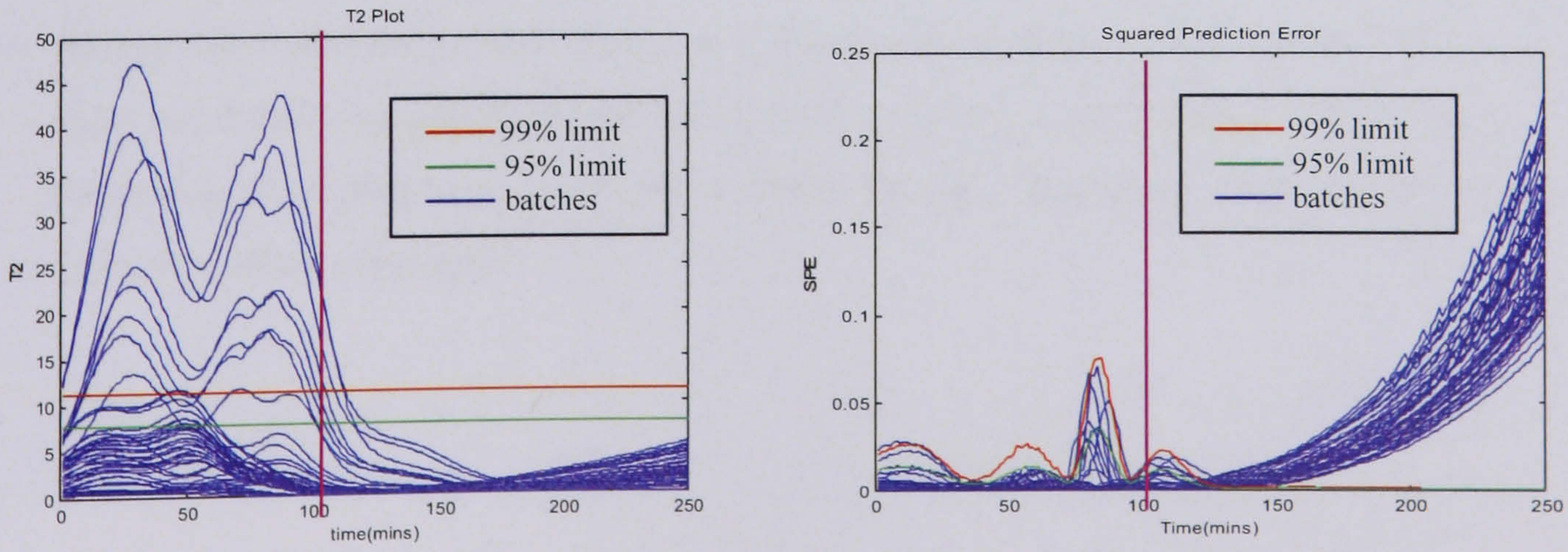
The MBPCA algorithm was then applied to the set of fault two batch data. The control charts used to assess the fault detection ability of the technique are shown in Figure 72.



a) Latent Variable 1 b) Latent Variable 2 c) Latent Variable 3

Figure 72: MBPCA control charts for fault type two

The chart for latent variable 1 shows that the majority of batch trajectories clearly move outside of the control limits at approximately time point 150. Similar behaviour can be seen in latent variable 3. The Hotelling's T^2 and SPE plots for the model-based data are shown in Figure 73.



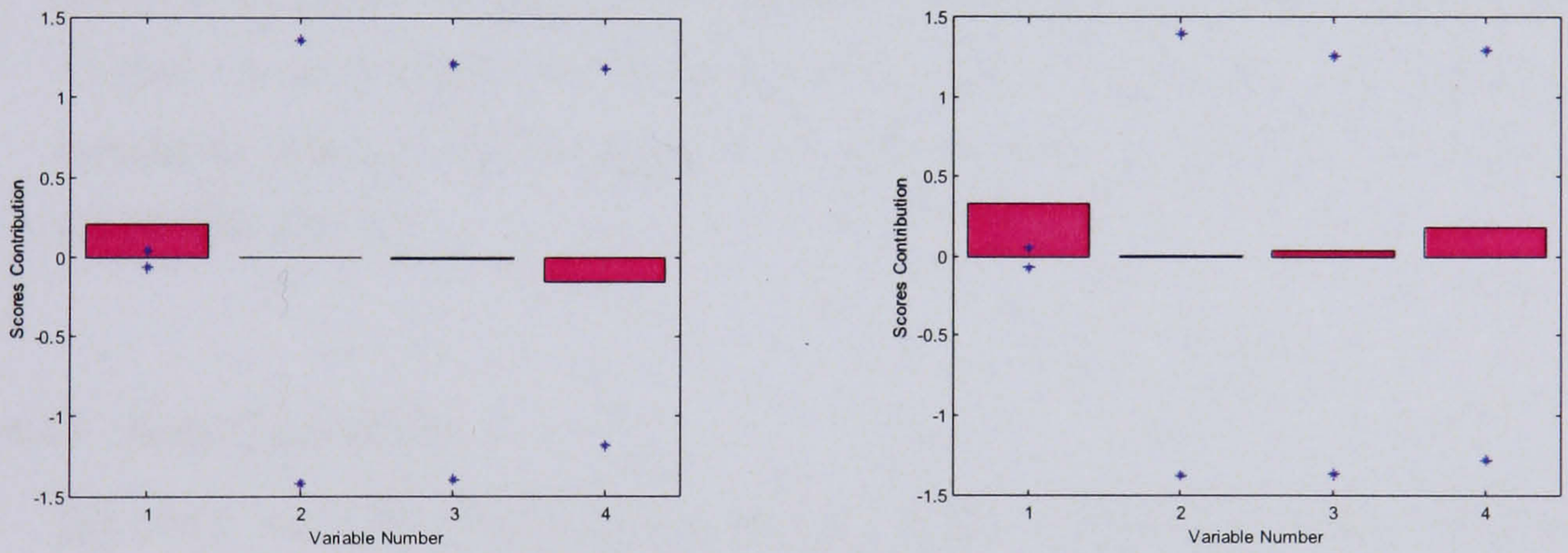
a) Hotelling's T^2 plot of fault type two b) SPE plot of fault type two

Figure 73: Hotelling's T^2 and SPE plot of fault type two

In Figure 73, the Hotelling's T^2 plot, ignoring the batches that are initially outside of the control limits, it can be seen that there is a change in batch behaviour approximately 50 minutes after the fault has been introduced as the batches start to move towards the control limits, although they do not exceed them. In the SPE plot however, the batches clearly move outside of the limits between 25 and 50 minutes after the fault has been introduced. The fact that the fault is more easily detected on the SPE chart as opposed to the Hotelling's T^2 chart indicates that the fault has caused a change in the correlation structure between the variables. The fault in question is a decrease in the heat transfer coefficient, which would have an effect on the way in which the variables relate to each other.

6.10.3.2 Fault Diagnosis (Fault Type Two)

Fault type two was detectable by the model-based technique. However it is important to determine if the fault can be diagnosed from the data. In Figure 74, the contribution plots for a batch that moved outside of the control limits are shown.



a) Time point 150 b) Time point 250

Figure 74: Contribution plots for out of control batch – fault type two

As has been discussed, fault type 2 is a change in the heat transfer coefficient and is expected to have an effect on the temperature variables. Figure 74(a) shows the reactor temperature (variable one) is the only variable having a significant effect, and in Figure 74(b) this effect continues.

6.10.4 Summary of Fault Type Two

The fault detection and fault diagnosis results for fault type two are summarised for each of the batch monitoring techniques in Table 6:

Technique	% Batches Detect. (control chart)	% Batches Detect. (T^2)	% Batches Detect. (SPE)	ARL control chart	ARL T^2	ARL SPE
MPCA	54	19	23	N/A	N/A	N/A
BOL	86	0	100	42	-	79
MBPCA	100	0	100	45	-	42

Table 6: Summary of batch monitoring methods

MBPCA detects all abnormal batches using the scores control charts and SPE control charts, however the MBPCA ARL for the control chart is slightly longer for MBPCA than it is for BOL. The BOL technique also demonstrates good results. The Hotelling's T^2 chart is not a useful tool in detecting this type of fault, as it is a change in the correlation structure of the variables as opposed to a change in the common cause variation in the data.

6.11 Fault Type Three

The third fault type investigated in the case study was the fault in the cooling water valve, which occurs after time point 150 (time point 100 on the control charts as the first fifty time points were removed before analysis). 50 batches containing this fault were generated, and the data set was then investigated using the three batch monitoring techniques.

6.11.1 Multiway Principal Component Analysis

6.11.1.1 Fault Detection (Fault Type Three)

The 50 batches of fault type three were projected onto the MPCA bivariate scores control limits. The resulting plots are shown in Figure 75.

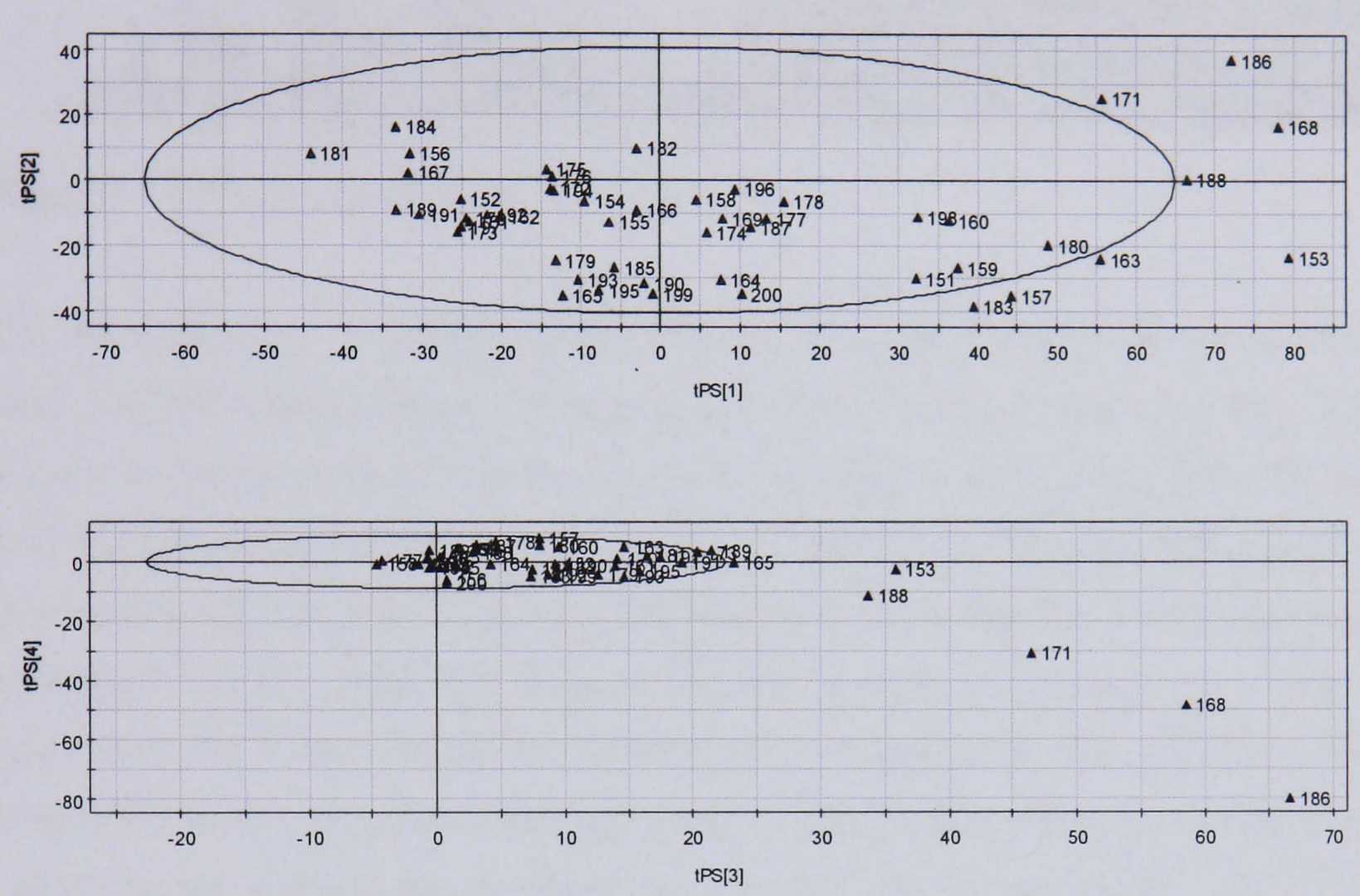


Figure 75: MPCA bivariate scores control charts for fault type 3

The control charts in Figure 75 show that the MPCA technique is not very successful at detecting the problem with the cooling valve, with only 8 batches lying outside of the control limits. The Hotelling's T^2 and SPE plots for the third fault type are shown in Figure 76 and Figure 77 respectively.

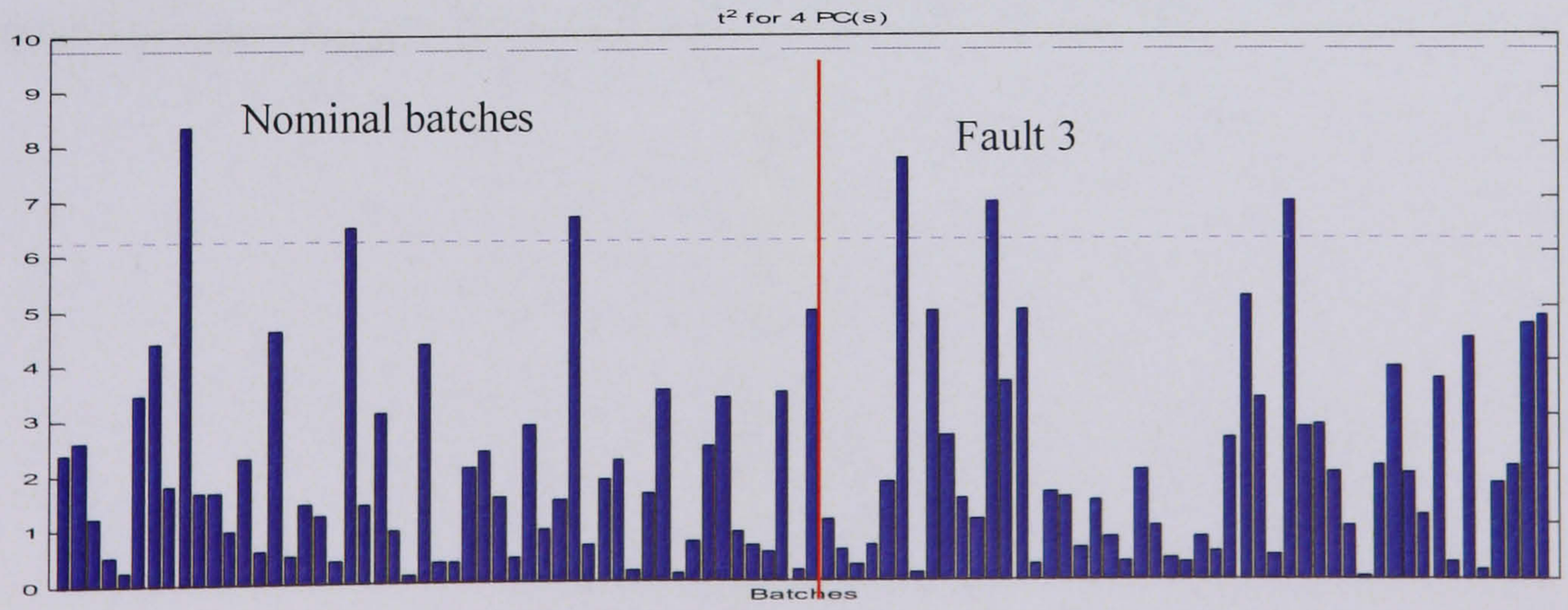


Figure 76: Hotelling's T^2 plot for fault type three

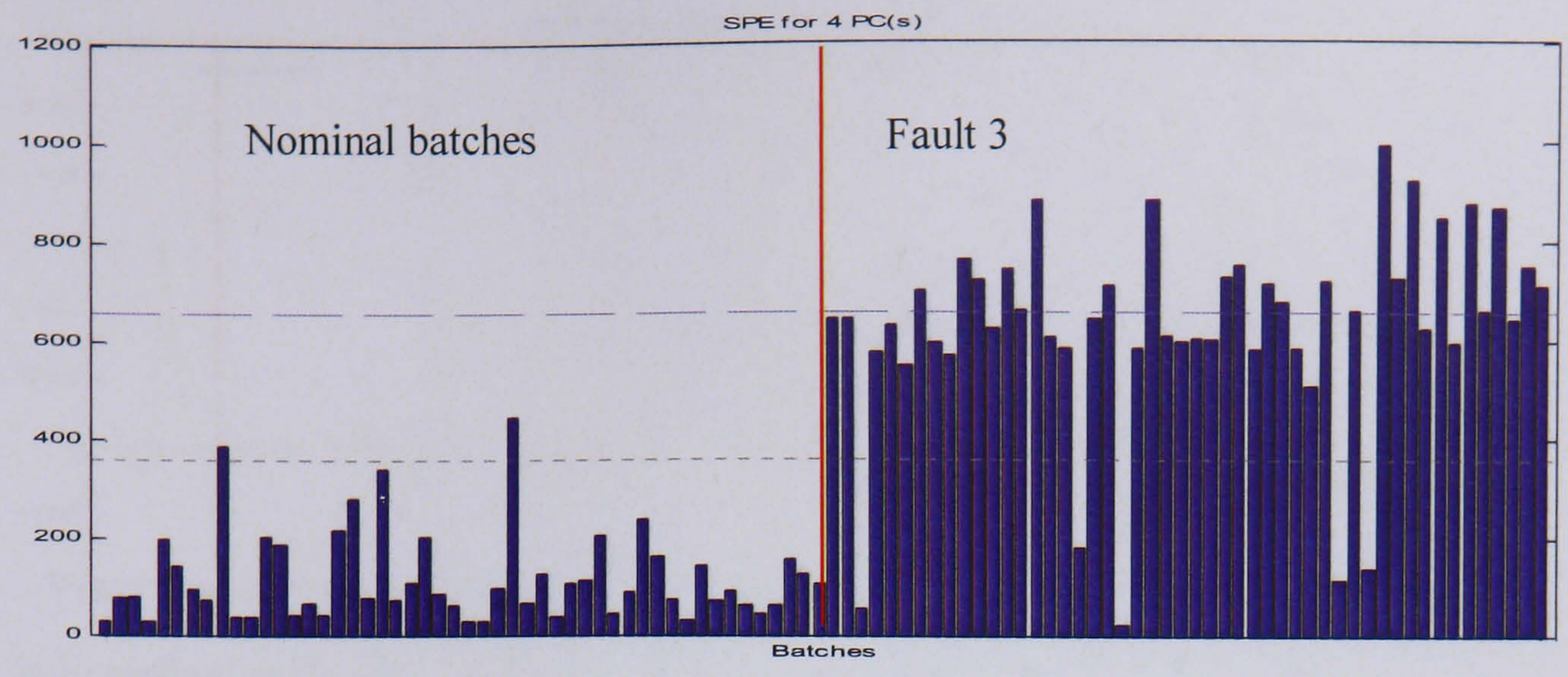
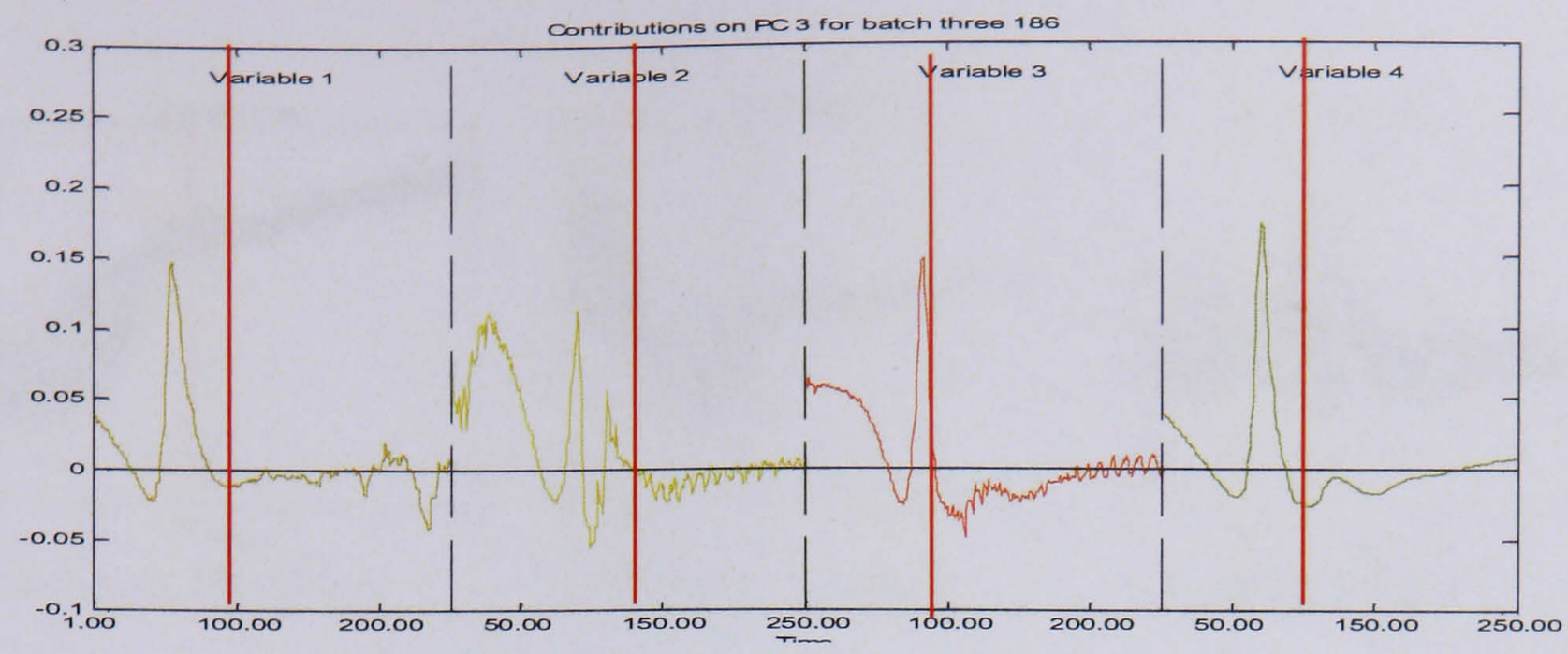


Figure 77: SPE plot for fault type three

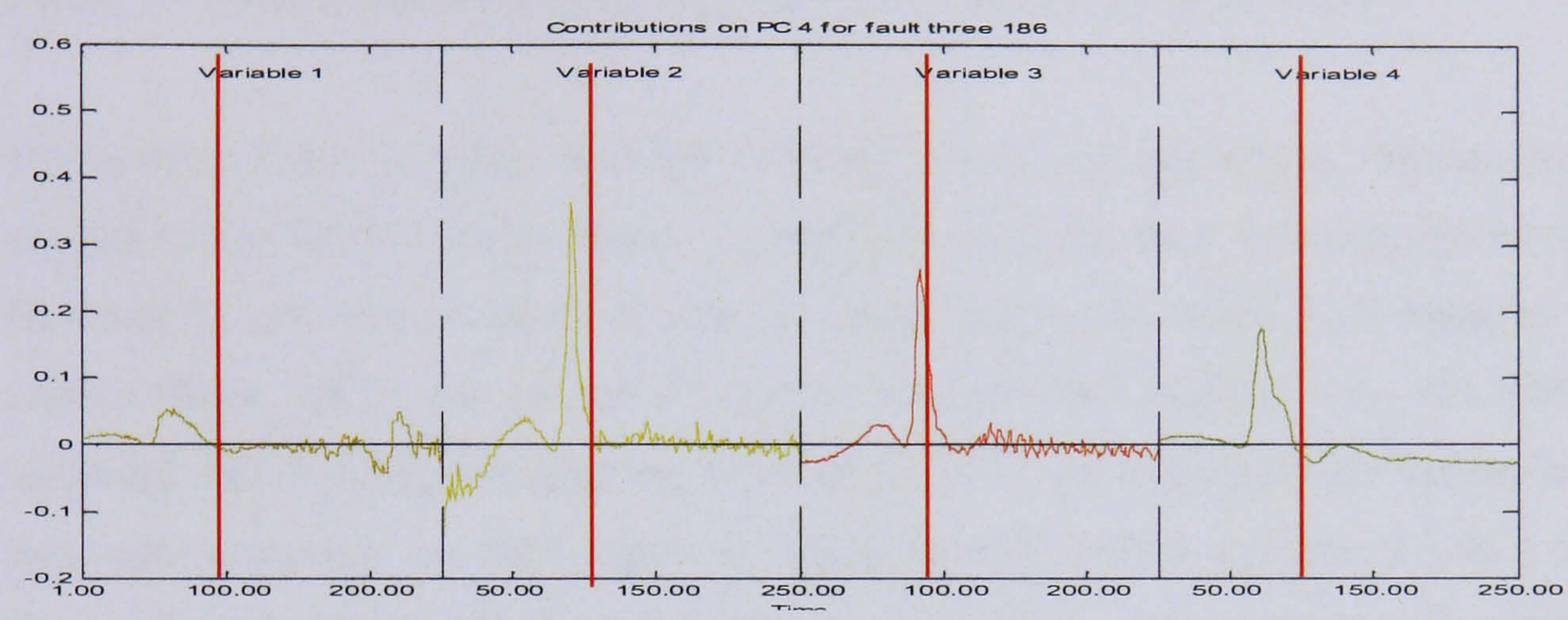
Using the Hotelling's T^2 statistic, only 3 batches are detected as being out of statistical control. The SPE chart in Figure 77 shows that the fault three batches display very different behaviour to the nominal data, as they all lie fairly close to the control limits. 20 batches exceed the 99% control limits, with 45 of the 50 batches exceeding the 95% control limit. Again the SPE chart is more successful at detecting the abnormalities in the batch than the Hotelling's T^2 statistic. This may be due to the type of fault, in this case it is a fault to the cooling valve that is affecting the batch behaviour. The impact that this fault has is that the cooling valve is slow to react to the required temperature increase or decrease. This will have an effect on the relationships between the variables that could change the correlation structure between them.

6.11.1.2 Fault Diagnosis (Fault Type Three)

The third fault type is a fault in the cooling water, therefore it would be expected that one of the main ways in which this fault would manifest itself would be in variable 4, the position of the cooling valve. Figure 78 shows the contribution plots for batch 186, which exceeds the control limits for principal components 3 and 4.



a) Contribution plot for batch 168 – PC3



a) Contribution plot for batch 168 – PC4

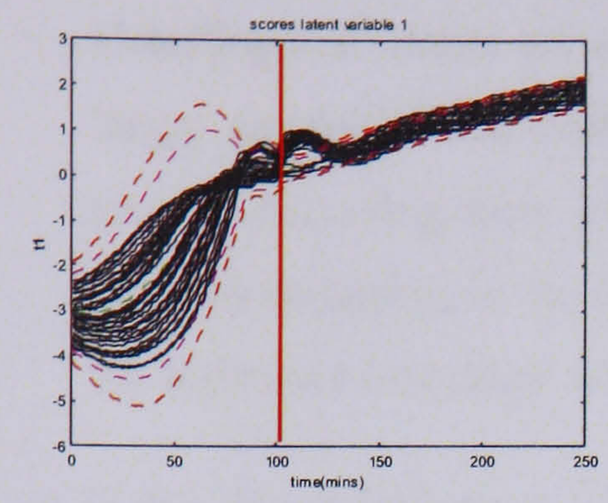
Figure 78: Contribution plots for fault type 3

From Figure 78 (a) it is difficult to ascertain the cause of the batch moving outside of the control limits. The large contributions observed in variable four occur prior to the fault being introduced in to the process. However, there are some changes in variable 4 after the fault occurs, likewise in variables 2 and 3, the wall temperature and jacket temperature respectively. It is expected that these variables would be affected by a change in the cooling and heating behaviour of the reactor. Figure 78 (b) again does not easily identify the cause of the batch being projected outside of the normal operating limits.

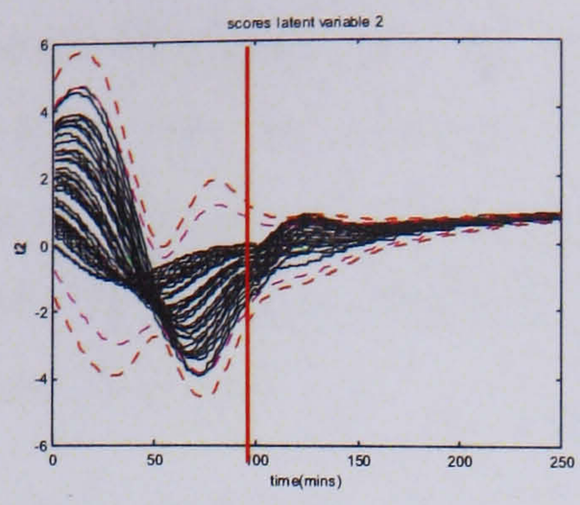
6.11.2 Batch Observation Level Analysis

6.11.2.1 Fault Detection (Fault Type Three)

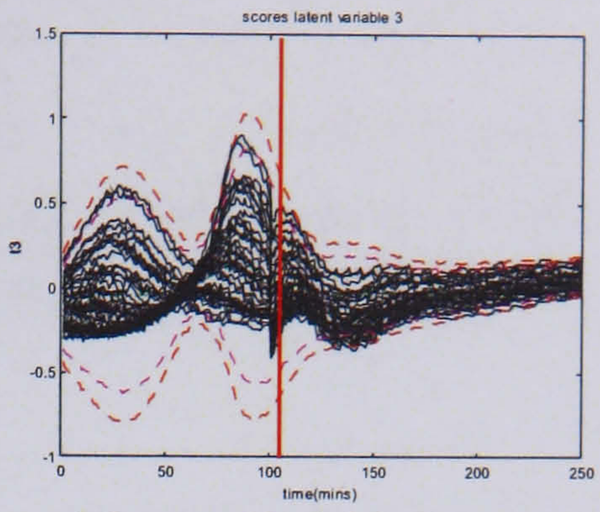
The 50 batches generated containing fault type three were also projected onto the batch observation level model. The resulting control charts are shown in Figure 79.



a) Latent Variable 1



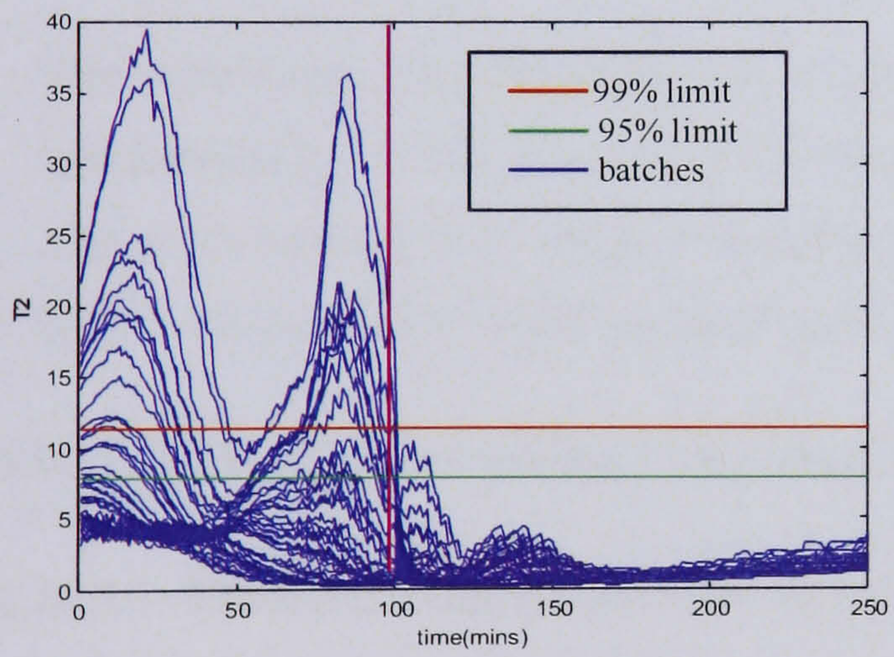
b) Latent Variable 2



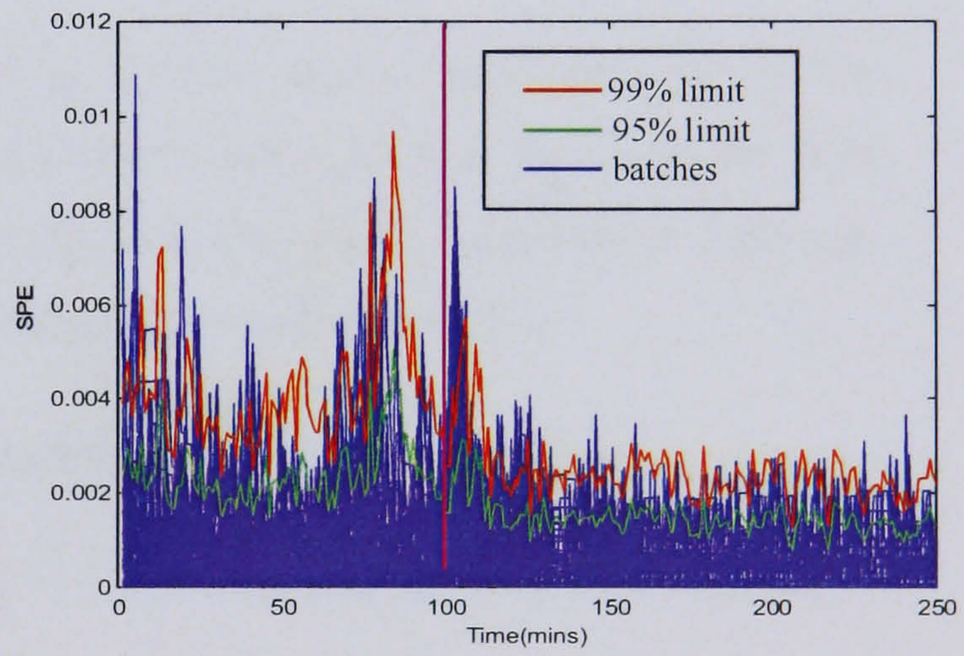
c) Latent Variable 3

Figure 79: Batch observation level scores control charts for fault type three.

The control chart for latent variable 1 clearly shows the majority of batches move outside of the control limits almost immediately after the fault has been introduced. However it can also be observed that the batch trajectories move back between the control limits before the end of the batch. This possibly explains why the MPCA technique had difficulty in detecting the problem as it was examining the batches after they were complete. i.e. back within statistical control. Latent variables 2 and 3 only show a limited number of batches deviating from the limits. The Hotelling's T^2 and SPE fault detection charts are shown in Figure 80.



a) Hotelling's T^2 plot of fault type 3



b) SPE plot of fault type 3

Figure 80: Hotelling's T^2 and SPE plots of fault type 3

The Hotelling's T^2 plot does not show the batches moving outside of the control limits, again, in contrast, a significant number of the batches do exit the SPE control limits directly after the fault has been introduced, but again they return to normal operating conditions before the batch is complete. The reason for the SPE chart being more successful at detecting non-conforming batches than the Hotelling's T^2 plot could be due

to the type of fault, as discussed with MPCA. However another explanation is that as the Hotelling's T^2 limits are calculated globally over the batch, and there is a significantly larger amount of variation in the first 100 minutes of the batch, any changes to the process occurring after 100 minutes may not be detected if they are smaller than the changes occurring in the initial batch phase. This does not occur with the SPE chart as the limits are calculated at each time point.

6.11.2.2 Fault Diagnosis (Fault Type Three)

A contribution plot for a batch of fault type three that moved outside of the control limits is shown in Figure 81, at time point 106, directly after the fault is introduced.

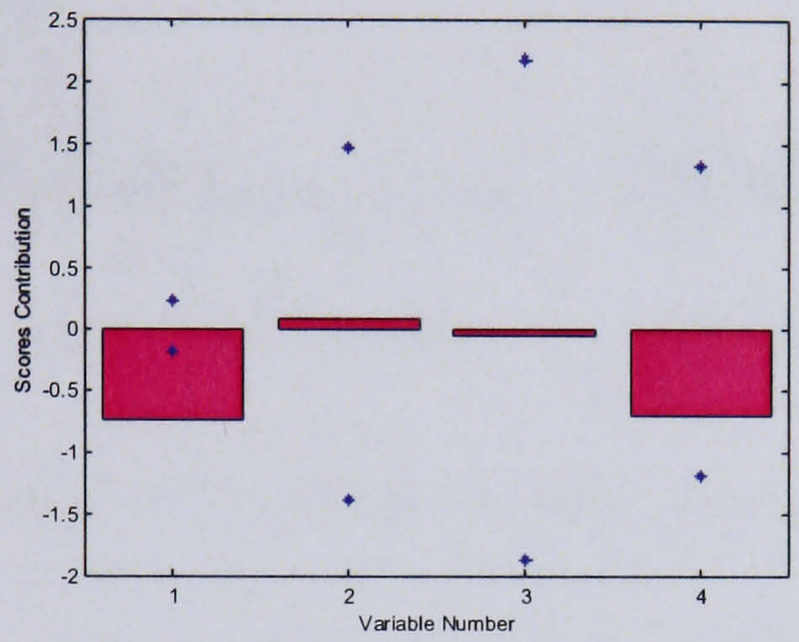
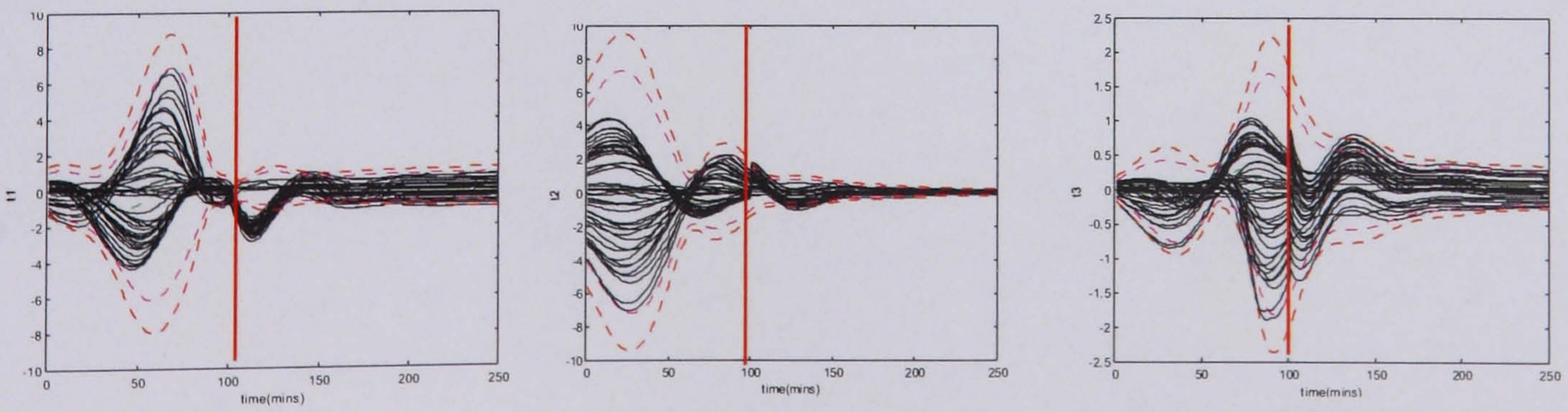


Figure 81: Contribution plot for abnormal batch for fault type three

The contribution plot shows that the reactor temperature has a significant effect on the batch trajectory moving outside of the control limits, this is to be expected as the fault affects the heating and cooling, which directly affects the reactor temperature. Although it does not exceed the limits, the cooling valve position is also important.

6.11.3 Model-based Principal Component Analysis

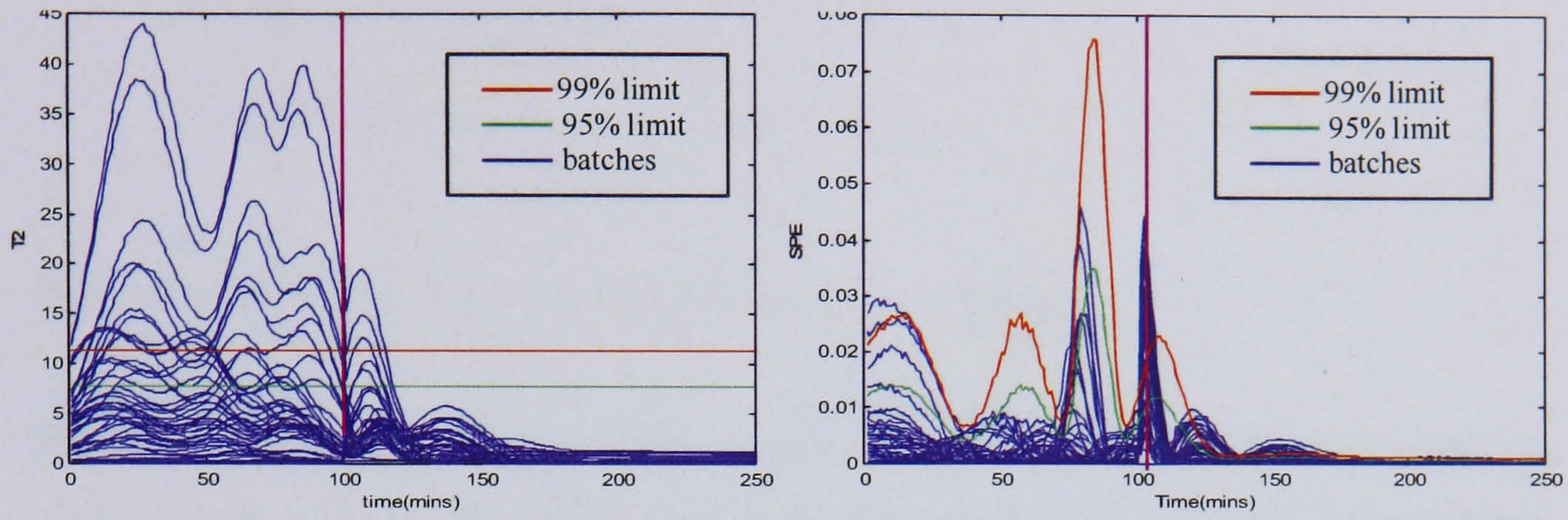
6.11.3.1 Fault Detection (Fault Type Three)



a) Latent Variable 1 b) Latent Variable 2 c) Latent Variable 3

Figure 82: MBPCA scores control charts for fault type three

The MBPCA control charts show a definite change in process behaviour after the valve fault has been introduced at time point 100. The majority of batches deviate from the control limits within a few minutes of the fault occurring in latent variable 1, before returning to the normal operating conditions approximately 50 minutes later. In latent variables 2 and 3, a change in the process trends can be seen, although the batches do not move outside of the limits.



a) Hotelling's T^2 plot for fault type three b) SPE plot for fault type three

Figure 83: Hotelling's T^2 and SPE plot for fault type three

With the Hotelling's T^2 and SPE plots shown in Figure 83, again it is only in the SPE chart that the change in process behaviour can be seen, with the majority of batches exceeding the control limits within a few minutes of the fault being introduced.

6.11.3.2 Fault Diagnosis (Fault Type Three)

To assess how successful the model-based technique is in terms of the diagnosis of the faults, contribution plots for a batch as it moves outside the control limits were generated, at time point 111.

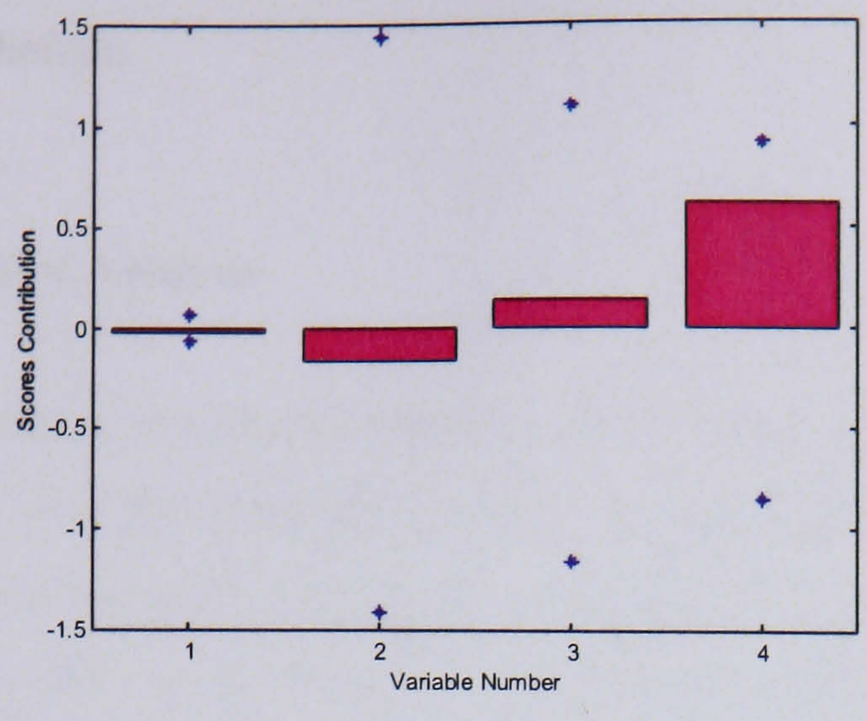


Figure 84: Contribution plot for fault type 3 at time point 111.

Fault type 3 is caused by a fault in the cooling water valve, variable 4, the valve position, shows the most significant contribution to the batch as it moves out of the scores limits, however, none of the variables exceeds their contribution limits at this time point.

6.11.4 Summary of Fault Type Three

The fault detection and fault diagnosis results for fault type three are summarised for each of the batch monitoring techniques in Table 7.

Technique	% Batches Detect. (control chart)	% Batches Detected (T^2)	% Batches Detected (SPE)	ARL control chart	ARL T^2	ARL SPE
MPCA	16	6	40	N/A	N/A	N/A
BOL	90	0	64	18	-	6
MBPCA	90	0	74	22	-	5

Table 7: Summary of batch fault detection metrics

The BOL and MBPCA techniques both show a high detection rate from the scores chart, particularly compared to the MPCA technique. The ARL for the MBPCA technique is slightly longer than for BOL. MPCA demonstrates poor results in detecting fault type three, possibly because the batches return to statistical control before the end of the

batch, making the abnormality difficult to detect when applying MPCA as an end of batch technique.

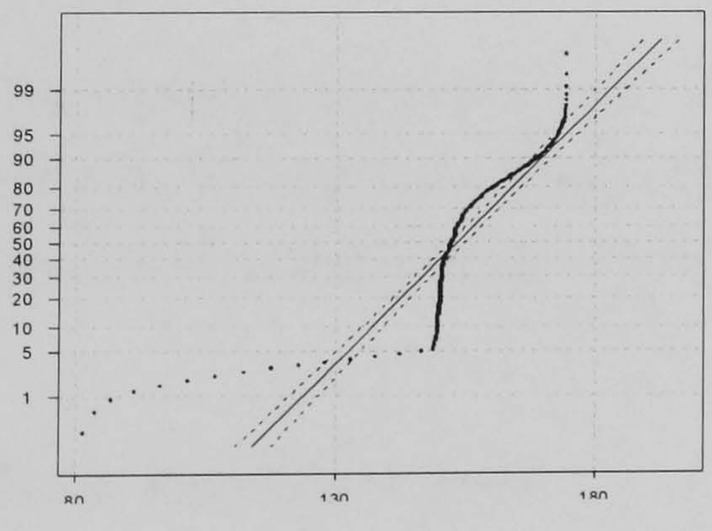
6.12 Residual Analysis

Batch processes are characterised by non-linear behaviour and serial correlation between variables. It is the presence of these characteristics that can make the application of multivariate statistical techniques challenging to apply in terms of the monitoring of batch processes. In this section, the residuals from the different batch techniques that were investigated in the case study were examined to assess their effectiveness in dealing with the non-linear and serially correlated behaviour of the exothermic batch process.

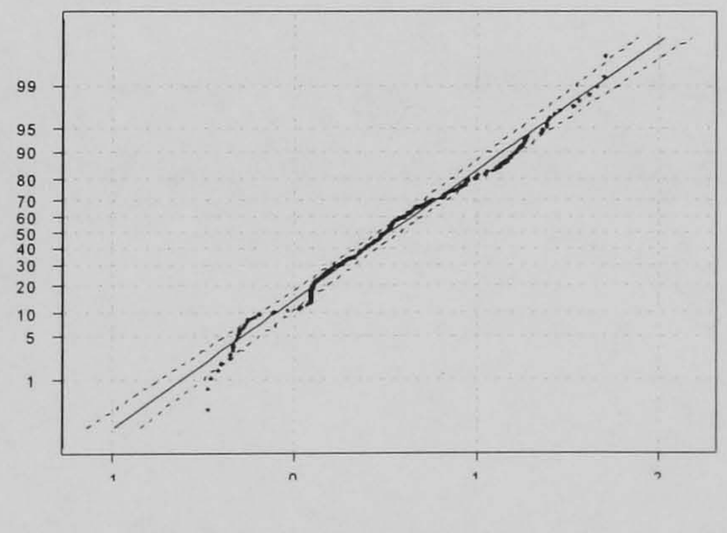
6.12.1 Multiway Principal Component Analysis

6.12.1.1 Non-linear Behaviour

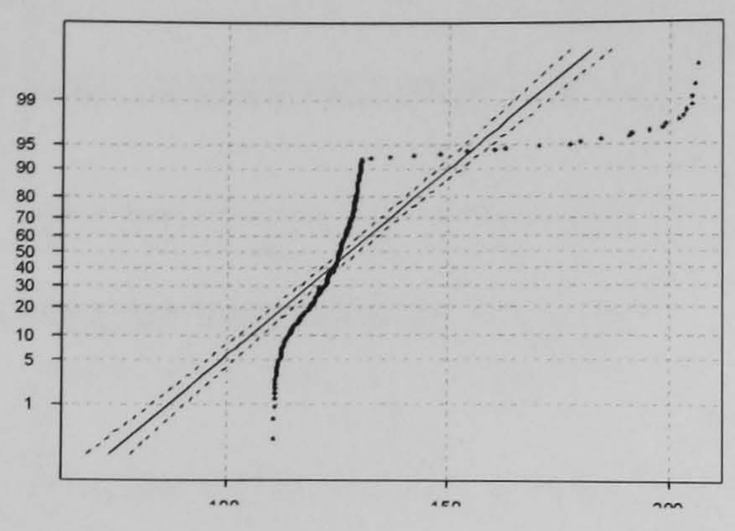
The multiway PCA technique claims to remove the main non-linear component from the data through the application of the scaling technique, which removes the average trajectory of the batches, and PCA is then applied to the deviations about the mean trajectory. To determine the validity of this statement, plots of the individual variables after scaling were examined, using a normal probability plot. In this case, only the 250 time points included in the multivariate analyses were considered. The probability plots for the raw data and the data after scaling are shown in Figure 85.



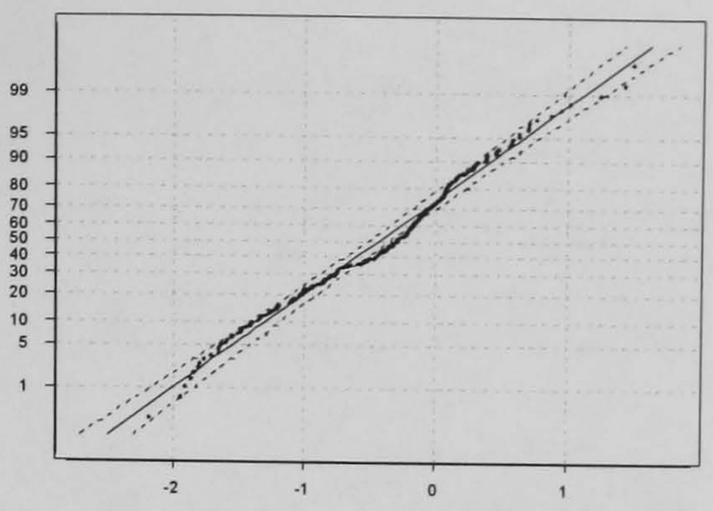
a) Reactor temperature – raw data



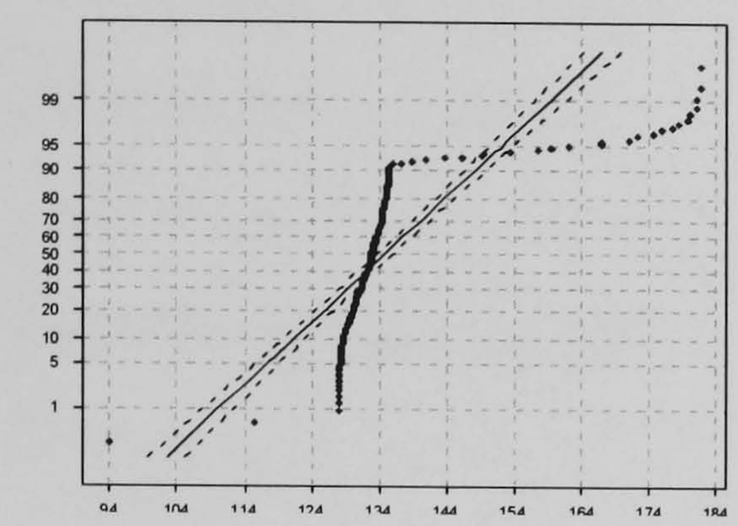
b) Reactor temperature – scaled data



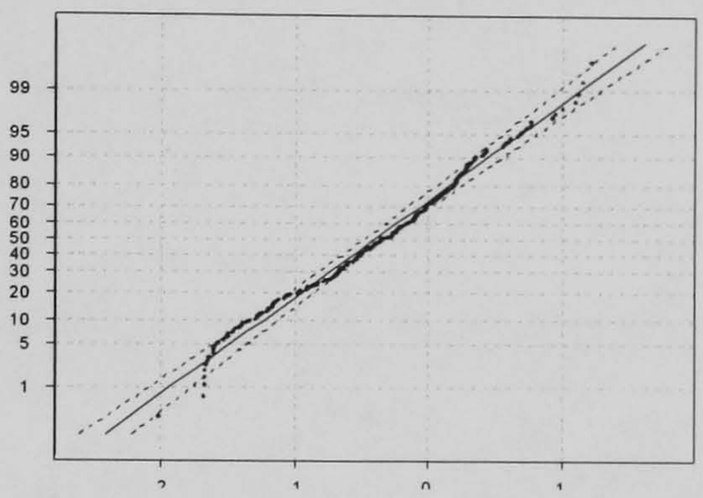
c) Wall temperature – raw data



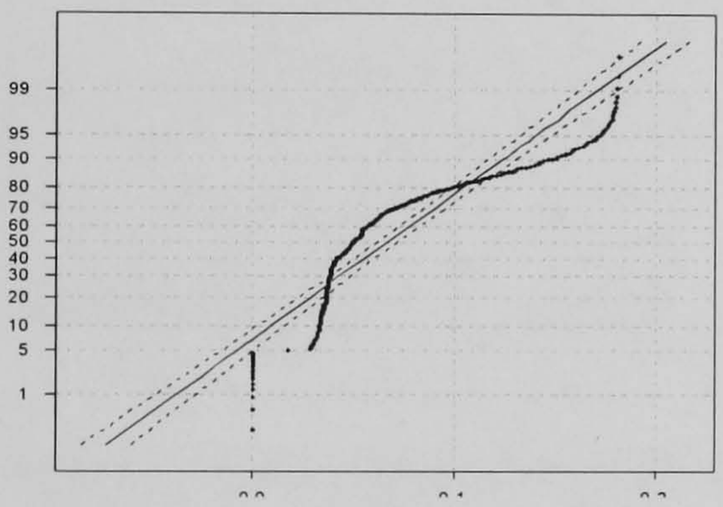
d) Wall temperature – scaled data



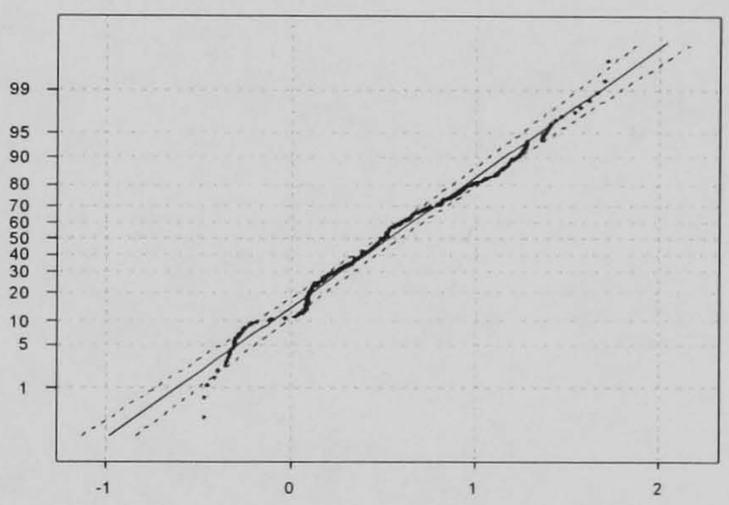
e) Jacket temperature – raw data



f) Jacket temperature – scaled data



g) Cooling valve position – raw data



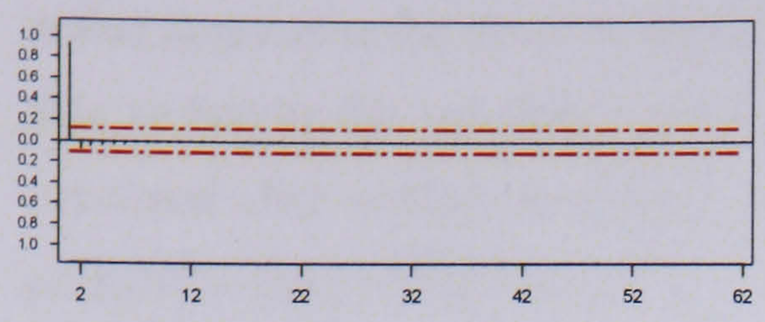
h) Cooling valve position – scaled data

Figure 85: Probability plots of raw and scaled variables

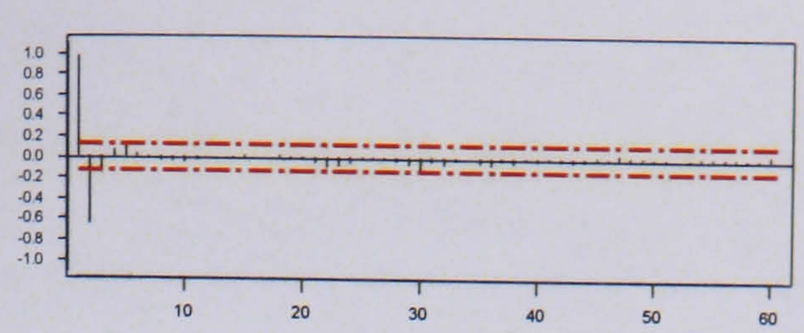
The probability plots for the raw variables show a fairly high degree of non-linearity in the data, but after the application of the data scaling algorithm the resulting data can be observed to exhibit approximate normality.

6.12.1.2 Serial Correlation

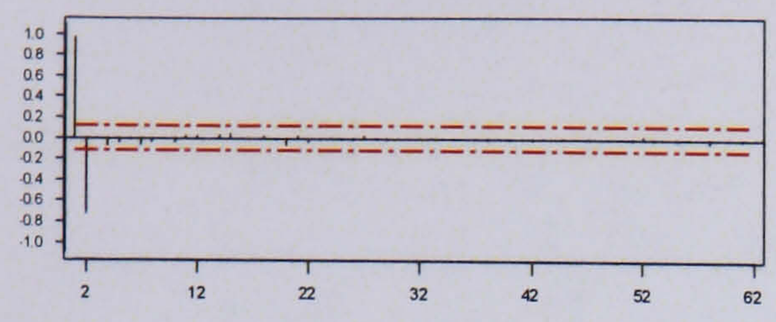
The second issue of concern with batch data is that of serial correlation. To investigate this issue partial autocorrelation plots were investigated for the raw and scaled data respectively, as shown in Figure 86.



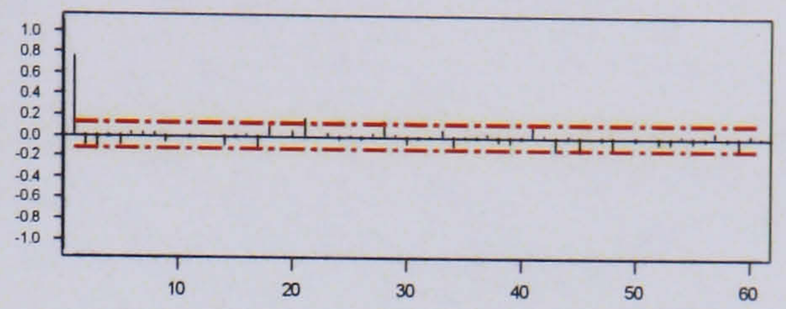
a)Reactor temperature – raw data



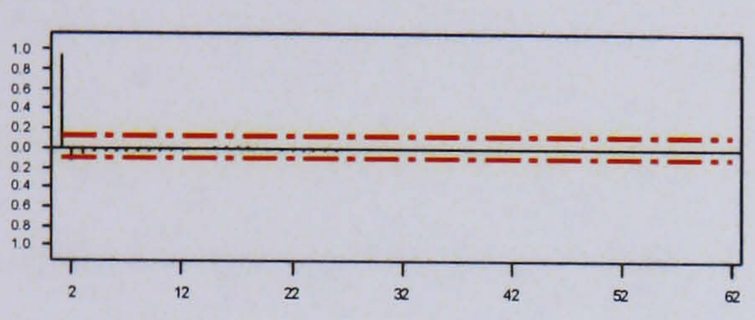
b) Reactor temperature – scaled data



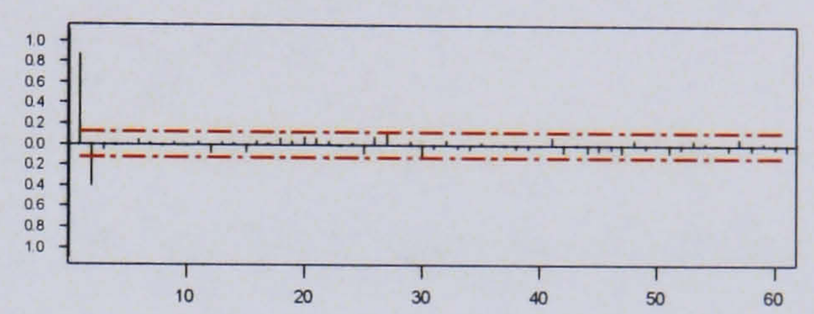
c) Wall temperature – raw data



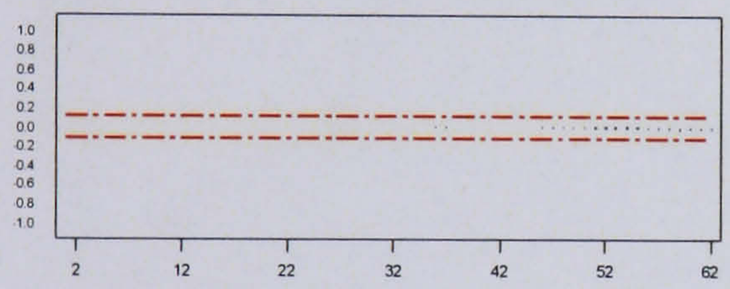
d) Wall temperature – scaled data



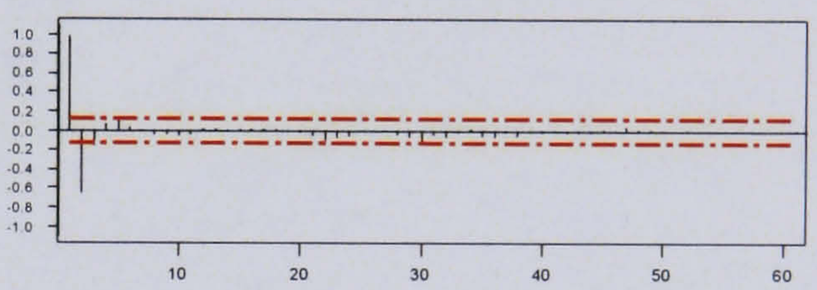
e) Jacket temperature – raw data



f) Jacket temperature – scaled data



g) Cooling valve position – raw data



h) Cooling valve position – scaled data

Figure 86: Partial autocorrelation plots of raw and scaled process variables

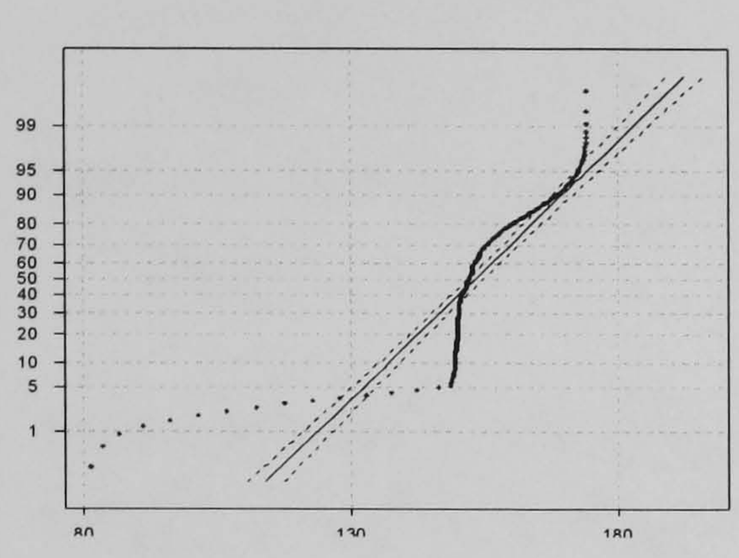
As can be seen, the scaling technique has not been successful in removing the dynamic structure in the batch data, with the dynamic structure decreasing for some of the variables but increasing for others. In general there is no overall improvement, but this is to be expected since the scaling technique was designed with the aim of tackling the issue of non-linearity rather than addressing the dynamics in the data.

6.12.1.3 Batch Observation Level Analysis

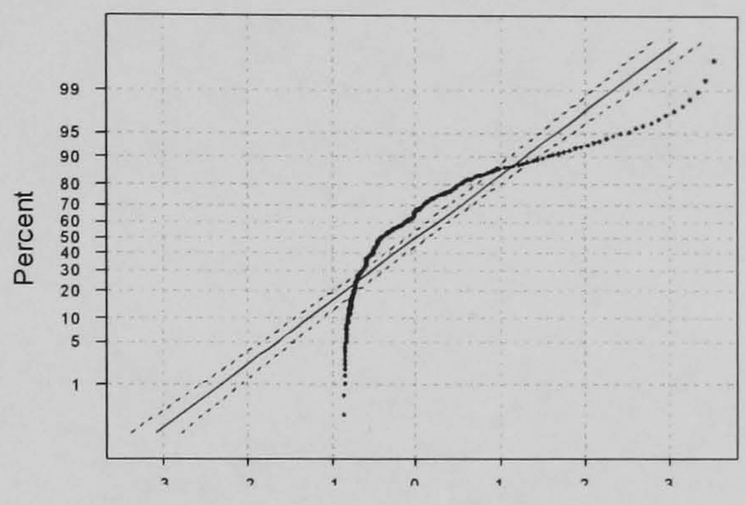
6.12.1.4 Non-linear Behaviour

The scaling method used for Multiway PCA is said to remove the non-linear behaviour in the data, and in the studies carried out on MPCA in this chapter it has been observed to be successful in doing so. The batch observation level technique however scales the data in a different way, instead of removing the average trace of the data, the data is

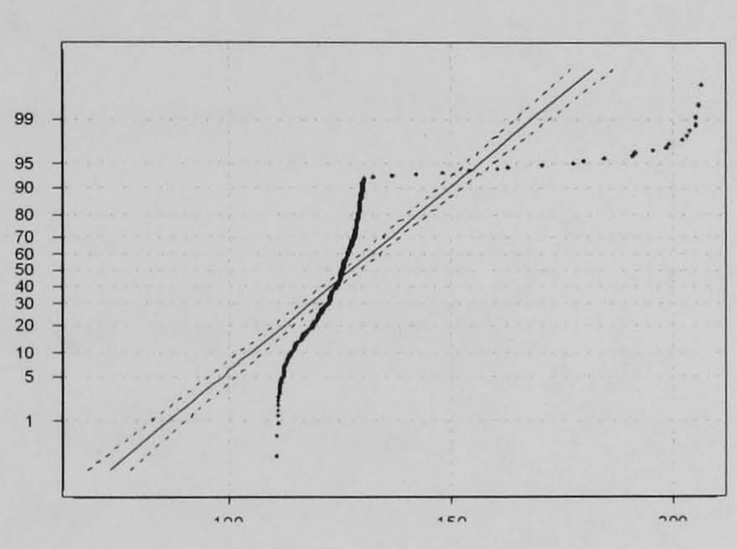
scaled to preserve the dynamic direction of the batches. To determine if this technique is able to handle the non-linear and dynamic structures in the data, the variables were examined after scaling. As before, the non-linear aspect was examined using normal probability plots of the data, shown in Figure 87.



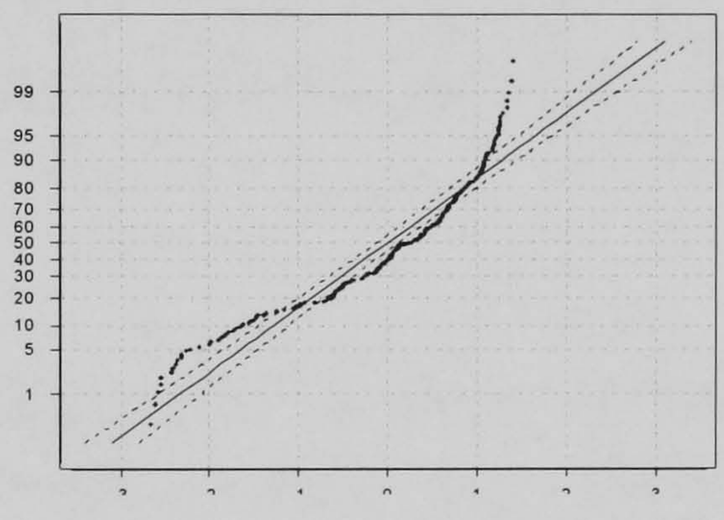
a) Reactor temperature – raw data



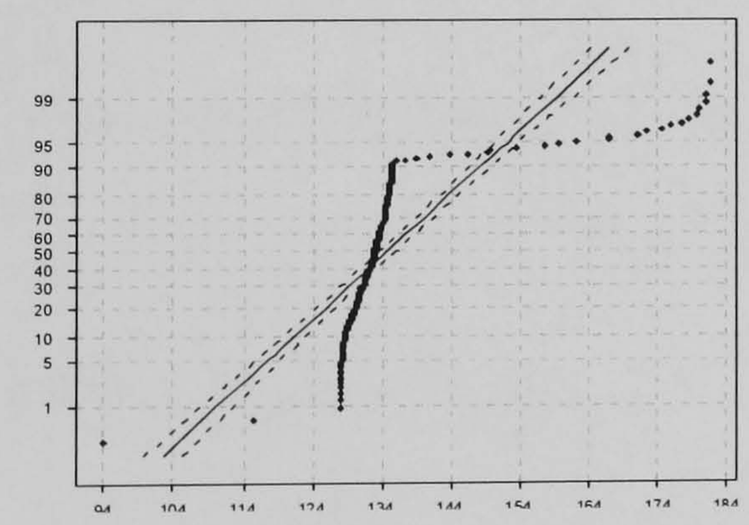
b) Reactor temperature – scaled data



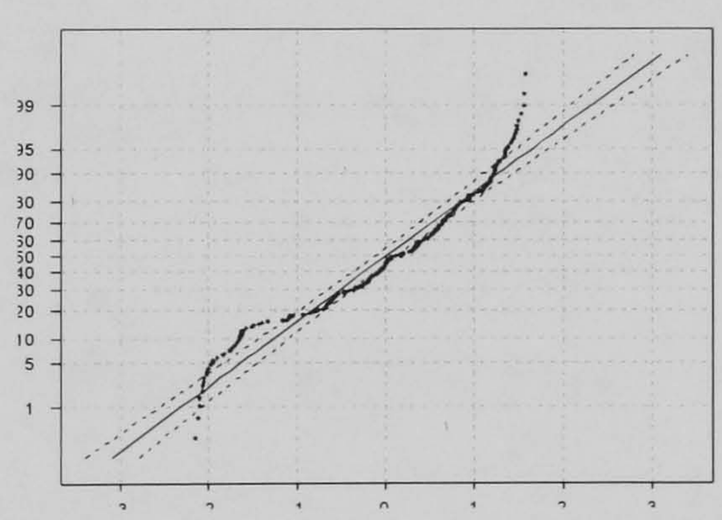
c) Wall temperature – raw data



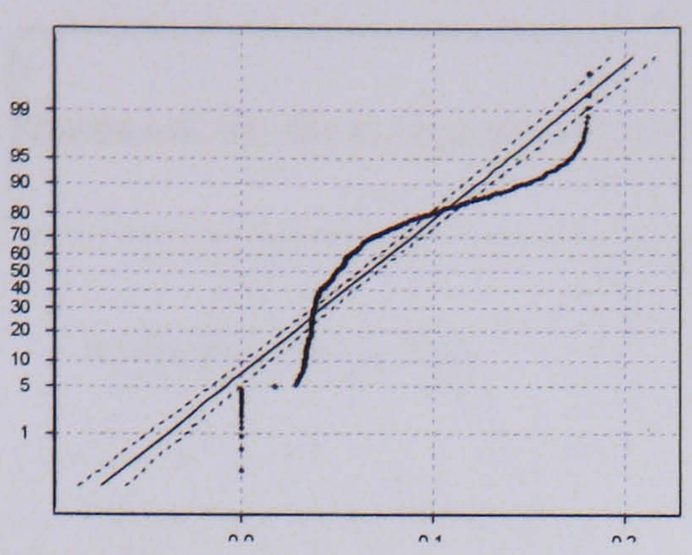
d) Wall temperature –scaled data



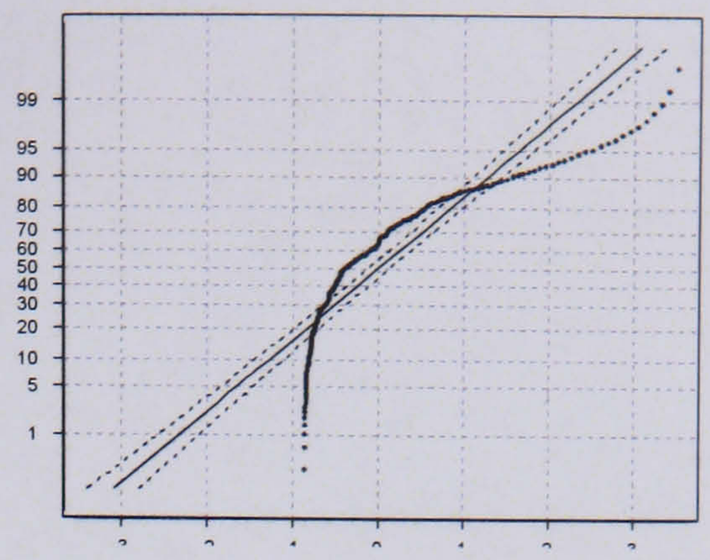
e) Jacket temperature – raw data



f) Jacket temperature - scaled data



g) Valve position – raw data



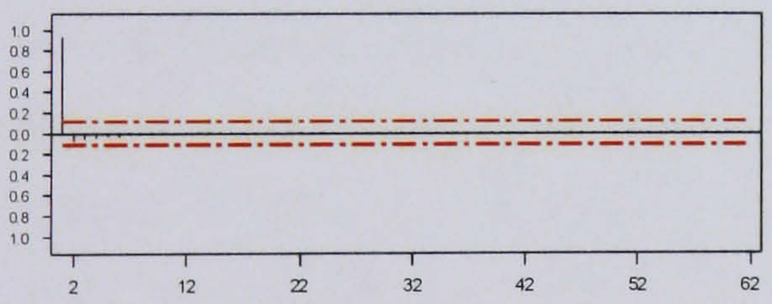
h) Valve position – scaled data

Figure 87: Probability plots of raw and scaled data for batch observation level analysis

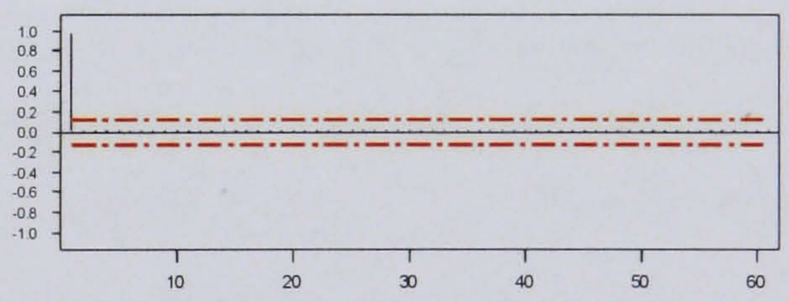
The probability plots of the data show variables one and four still exhibit significant non-linear behaviour compared to the original data, however there is a slight improvement. Variables 2 and 3 demonstrate a more significant reduction in non-linearity although they are still skewed. As expected, the BOL technique does not demonstrate the significant improvements that the MPCA scaling technique achieved.

6.12.1.5 Serial Correlation

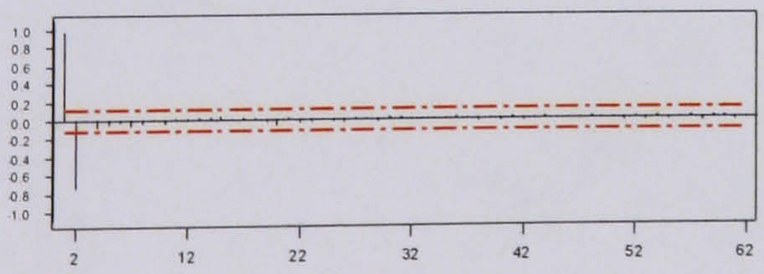
The MPCA technique was not effective in dealing with the serial correlation in the data. Figure 88 shows the PACF results for the batch observation level scaled data to see if this approach provides better results.



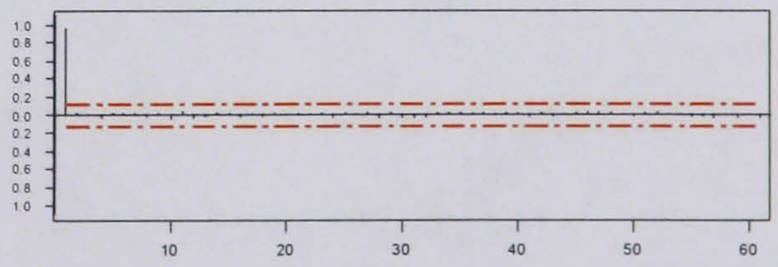
a) Reactor temperature – raw



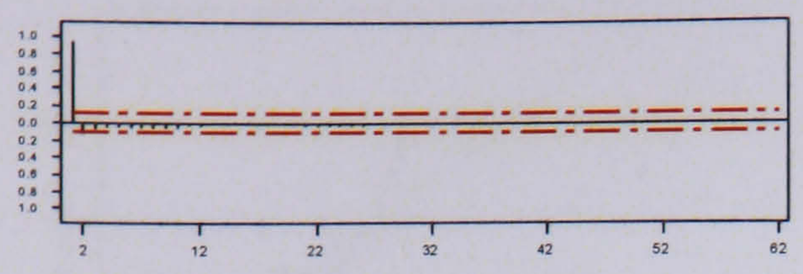
b) Reactor temperature – scaled



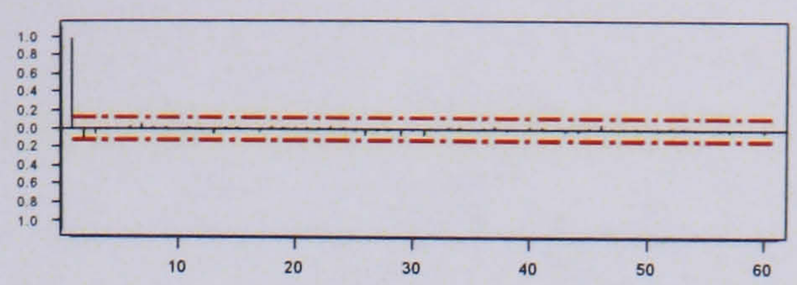
c) Wall temperature – raw



d) Wall temperature – scaled

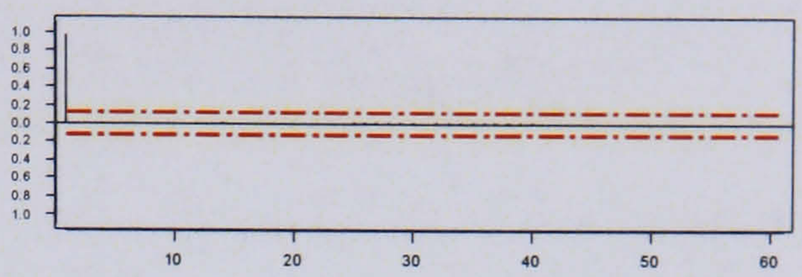


e) Jacket temperature – raw



f) Jacket temperature – scaled

g) Valve position – raw



h) Valve position – scaled

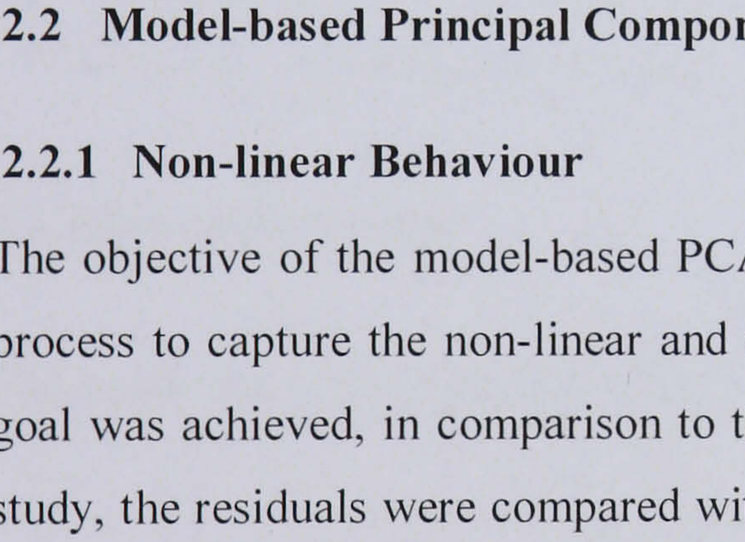
Figure 88: Partial autocorrelation plots for batch observation level data

The variables examined still display a high degree of serial correlation in the data, although compared with the partial autocorrelation plots from the original data, this type of scaling has produced a marked improvement in the structure, with all the variables now having a process order of 1 (as opposed to orders of 1,2,1,3 respectively in the original data).

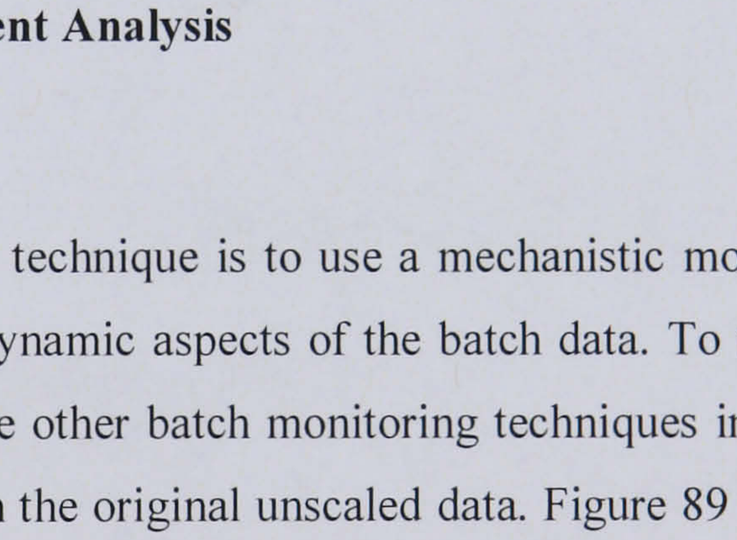
6.12.2 Model-based Principal Component Analysis

6.12.2.1 Non-linear Behaviour

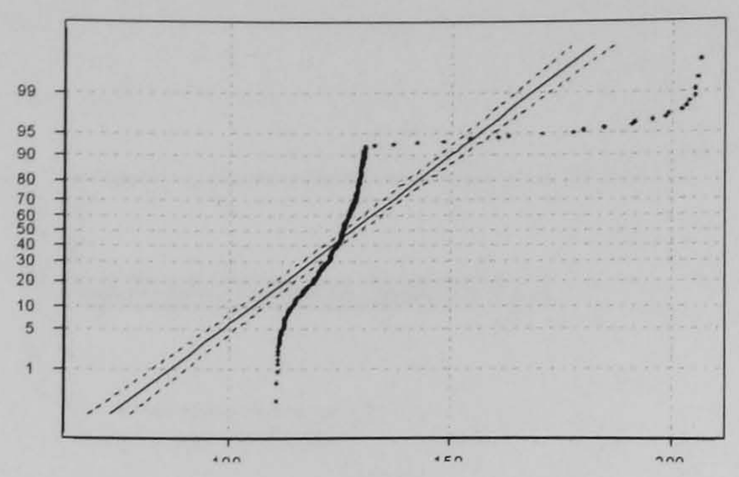
The objective of the model-based PCA technique is to use a mechanistic model of the process to capture the non-linear and dynamic aspects of the batch data. To test if this goal was achieved, in comparison to the other batch monitoring techniques in this case study, the residuals were compared with the original unscaled data. Figure 89 shows the normal probability plots.



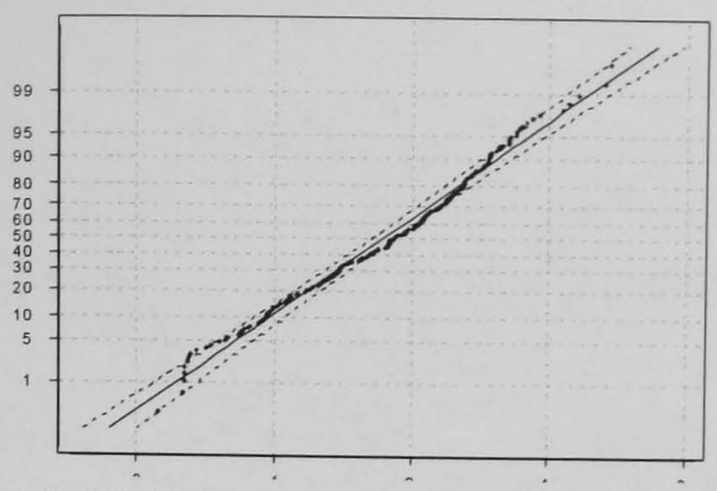
a) Reactor temperature – raw



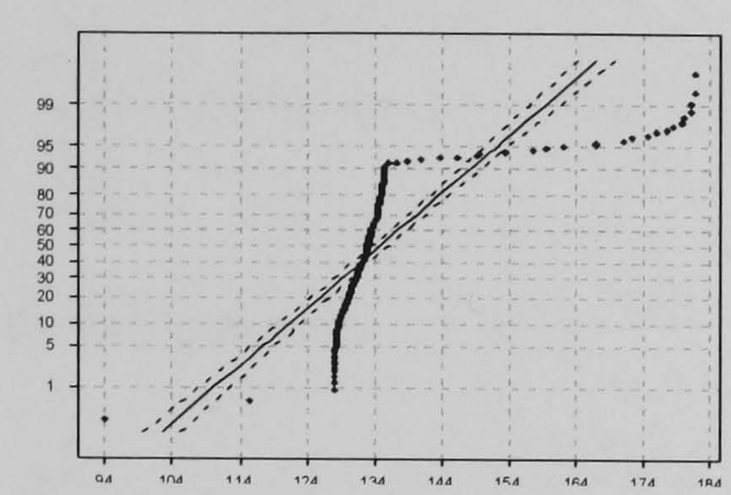
b) Reactor temperature – residual



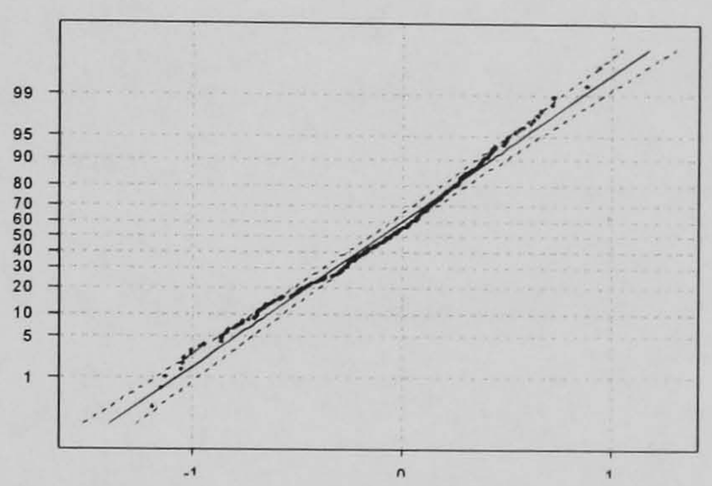
c) Wall temperature – raw



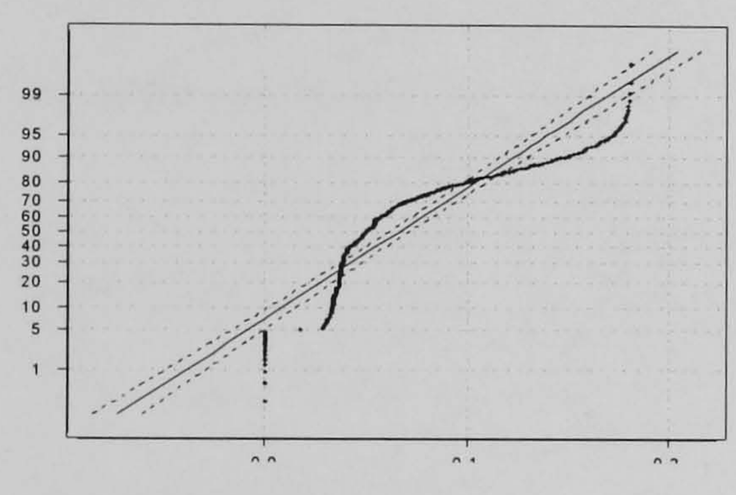
d) Wall temperature – residual



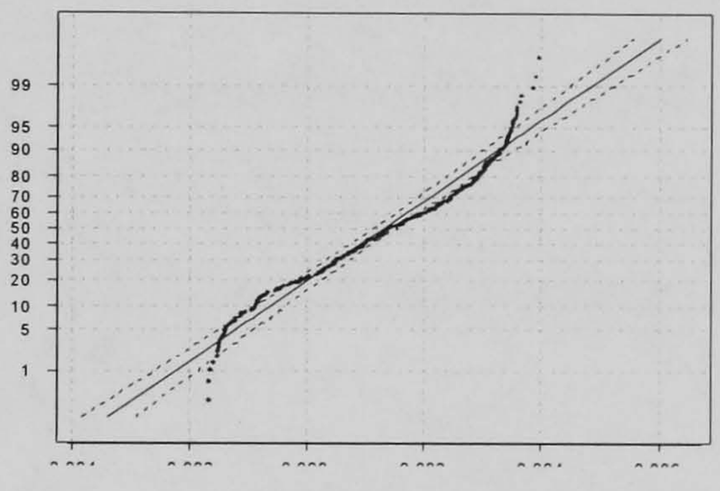
e) Jacket temperature –raw



f) Wall temperature –residual



g) Valve position – raw



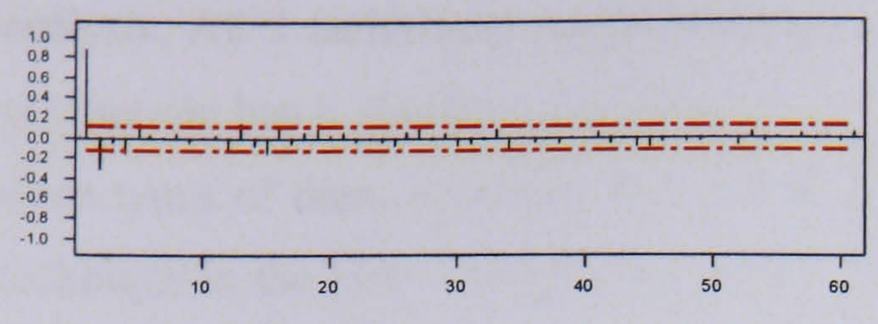
h) Valve position – residual

Figure 89: Normal probability plots for model-based PCA

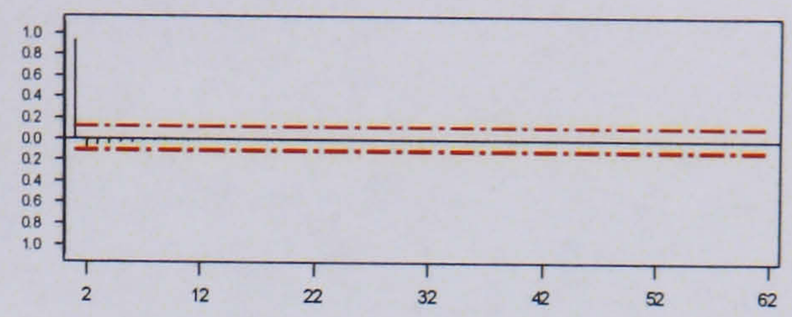
The probability plots show that the model-based technique has reduced the degree of non-linearity in the data, with all 4 variables showing an improvement, particularly variables 2 and 3 which are approximately straight lines, although the results are not as good as for MPCA.

6.12.2.2 Serial Correlation

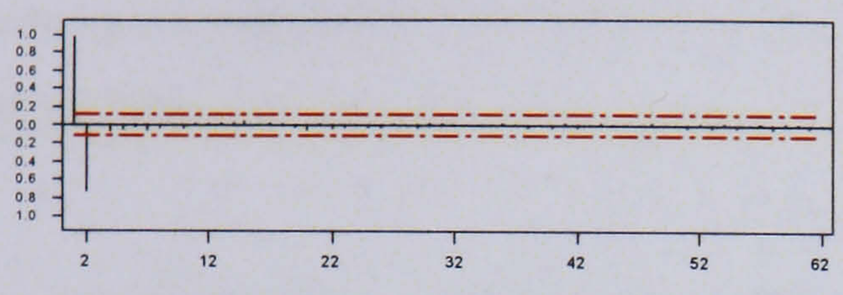
Partial autocorrelation plots were used to investigate the dynamic structures remaining in the model-based data, Figure 90.



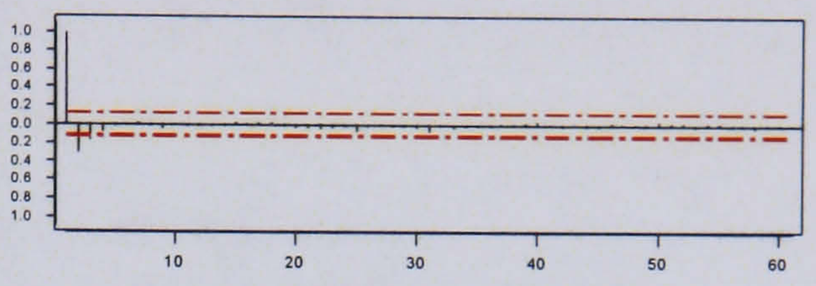
a) Reactor temperature – raw



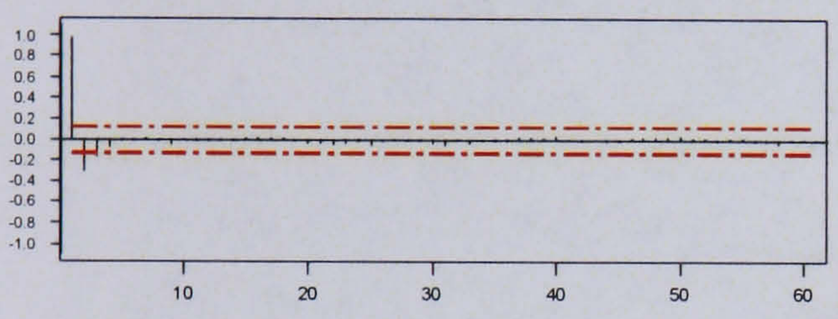
b) Reactor temperature – residual



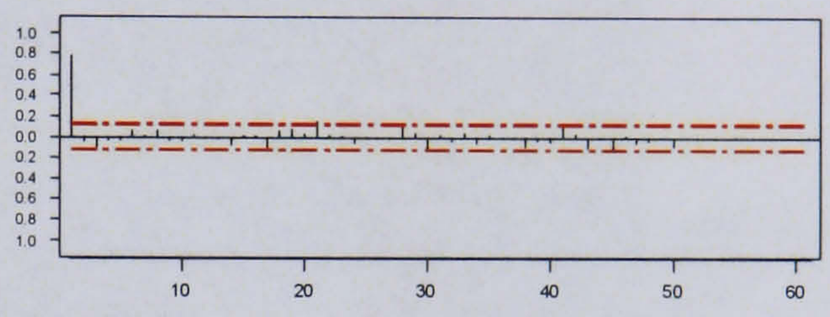
c) Wall temperature – raw



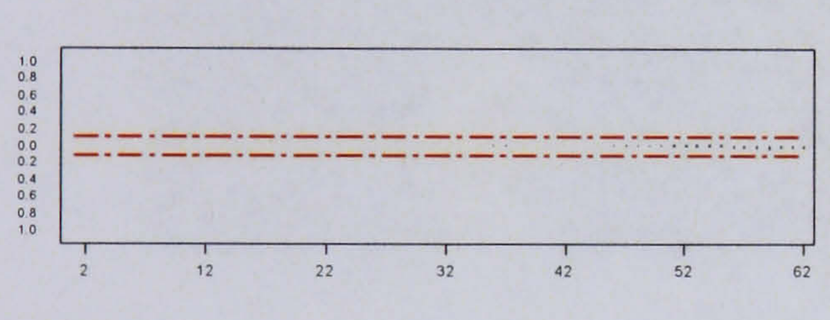
d) Wall temperature – residual



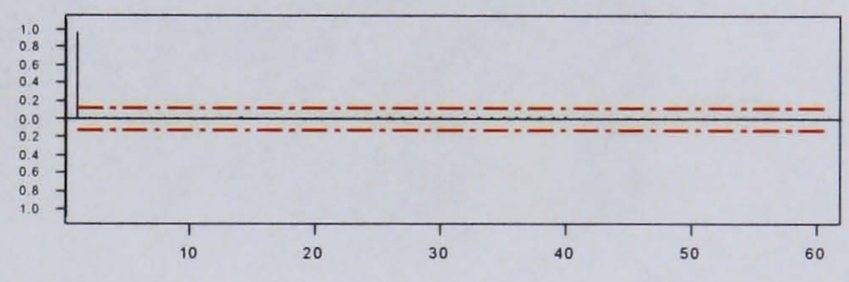
e) Jacket temperature – raw



f) Jacket temperature – residual



g) Valve position – raw



h) Valve position – raw

Figure 90: Partial autocorrelation plots for model-based PCA

Figure 90 shows the results for the partial autocorrelation function for the residuals. For some variables the serial correlation in the data is reduced, whilst it has increased for others. The reason that MBPCA has not been as successful at removing the serial correlation from the data as would be hoped is that the model and the plant are not a perfect match and therefore the model will not remove all the structure in the data.

6.13 Conclusions

In this chapter, three standard batch monitoring techniques were examined with respect to their ability to deal with the non-linear and dynamic characteristics inherent within batch processes, and how these effects impacted on their fault detection capability. Additionally, how easy it was to diagnose the potential source of the faults using each

method was investigated. The techniques compared were multiway principal component analysis, batch observation level analysis and model-based principal component analysis. All 3 techniques were investigated using the same nominal data set from an exothermic batch simulation, a validation data set of 20 normal operation batches, and three types of fault, of which 50 batches were generated for each, to determine which technique is the most suitable for the monitoring of batch processes. In Figure 91, the number of validation batches observed outside of the control limits for each monitoring technique is presented. This was calculated for all the principal components included in the models.

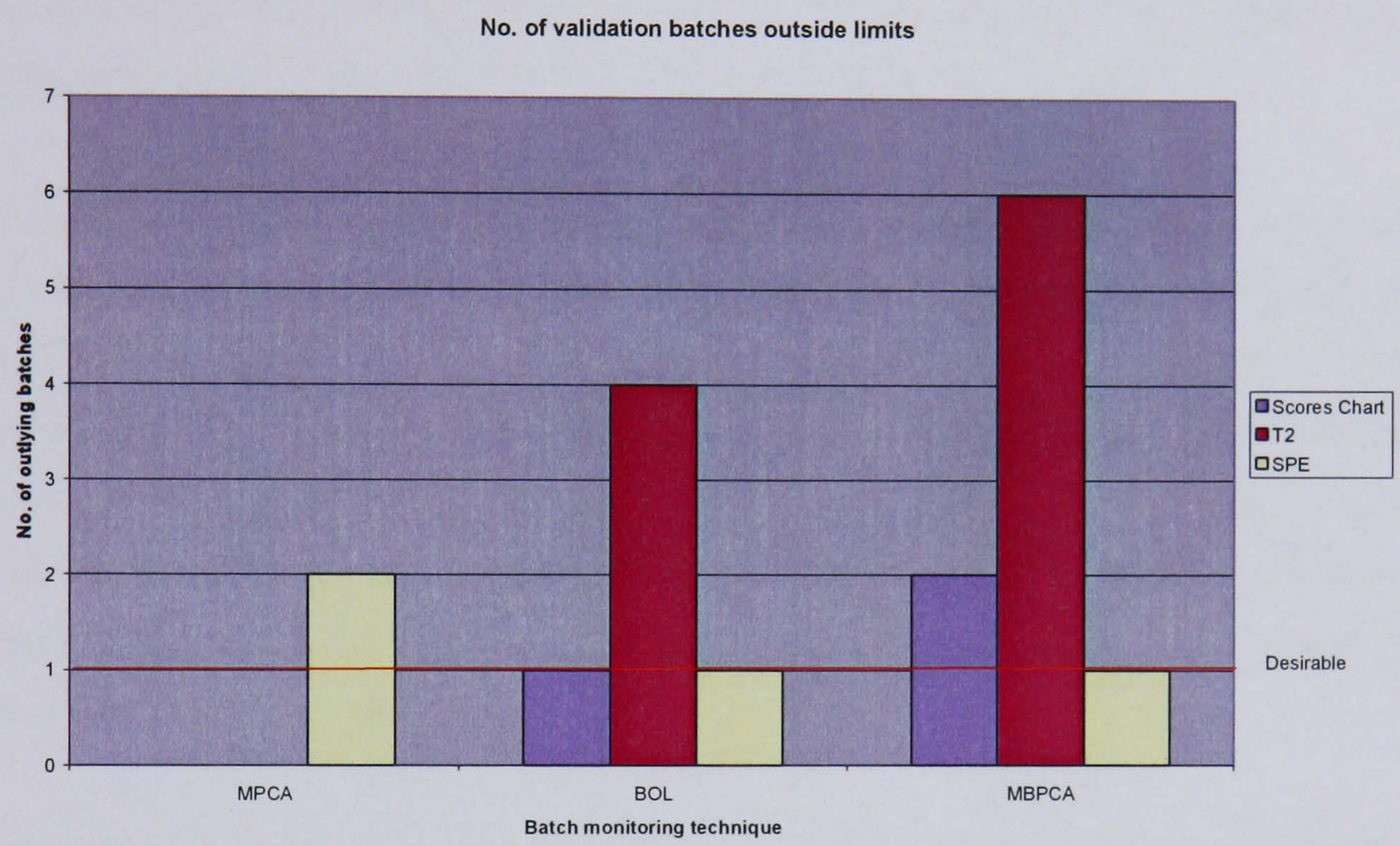


Figure 91: Number of validation batches falling outside control limits

The MPCA control chart and Hotelling's T^2 chart did not identify any of the 20 validation batches outside of the control limits, whilst 2 validation batches were outside for the SPE control limits. With both the BOL and MBPCA techniques, a larger than acceptable number of batches fell outside the control limits on the Hotelling's T^2 plot. There is a large amount of variation in the data during the first 100 minutes, as this is a different phase of operation of the batch, after this time the batch behaviour is fairly steady and the amount of variation is greatly reduced. The control limits for the Hotelling's T^2 plot are calculated over the entire length of the batch. As the batch comprises of two phases, one in which there is a large amount of variation and one in which the behaviour is fairly steady state, the limits represent an average of both phases. This means that during the first 100 minutes the limits are too low, therefore several batches lie outside the limits, and after 100 minutes the limits are too high, making the detection of abnormalities difficult.

Differences were evident in terms of how well each of the techniques dealt with the non-linear structure and serial correlation in the batch data. Multiway PCA demonstrated good results in terms of the removal of the non-linear structure in the data, as a consequence of the scaling methodology that was applied (as discussed in Chapter 3). However this scaling technique does not have a positive effect on the serial correlation in the data and there is still evidence of the data being serially correlated.

The batch observation level technique gave acceptable fault detection performance but did not successfully deal with the non-linear behaviour in the data. The technique gave better results than MPCA with respect to addressing the serial correlation.

The model-based PCA technique did not fulfil its potential as a batch fault detection tool that had the capability to deal with both the serial correlation and the non-linear behaviour in the data. This is because there was a mismatch between the plant and the model, due to the specific batch behaviour being not fully captured by the mechanistic model. However, MBPCA did give good results in terms of dealing with the non-linear behaviour in spite of the presence of plant model mismatch, so although the mechanistic model was not a perfect match of the plant it did manage to capture to some extent the non-linear behaviour in the process.

Two aspects of the fault detection abilities of the techniques were looked at. Firstly, how many of the abnormal batches did the technique detect for each fault type, and secondly, how rapidly did the technique detect an out of control batch. The second aspect only applied to the BOL and MBPCA techniques as MPCA was applied as an end of batch technique. It would be possible to use MPCA as a through batch technique using in-filling of the data, as discussed in Chapter Two, therefore allowing the average time for detection of a fault to be calculated, however this was not included in this work. Figure 92 – Figure 94 summarise the results for the percentage of batches detected for each of the monitoring techniques using the control charts, Hotelling's T^2 charts and SPE charts respectively. The results were calculated based on all the principal components included in the models.

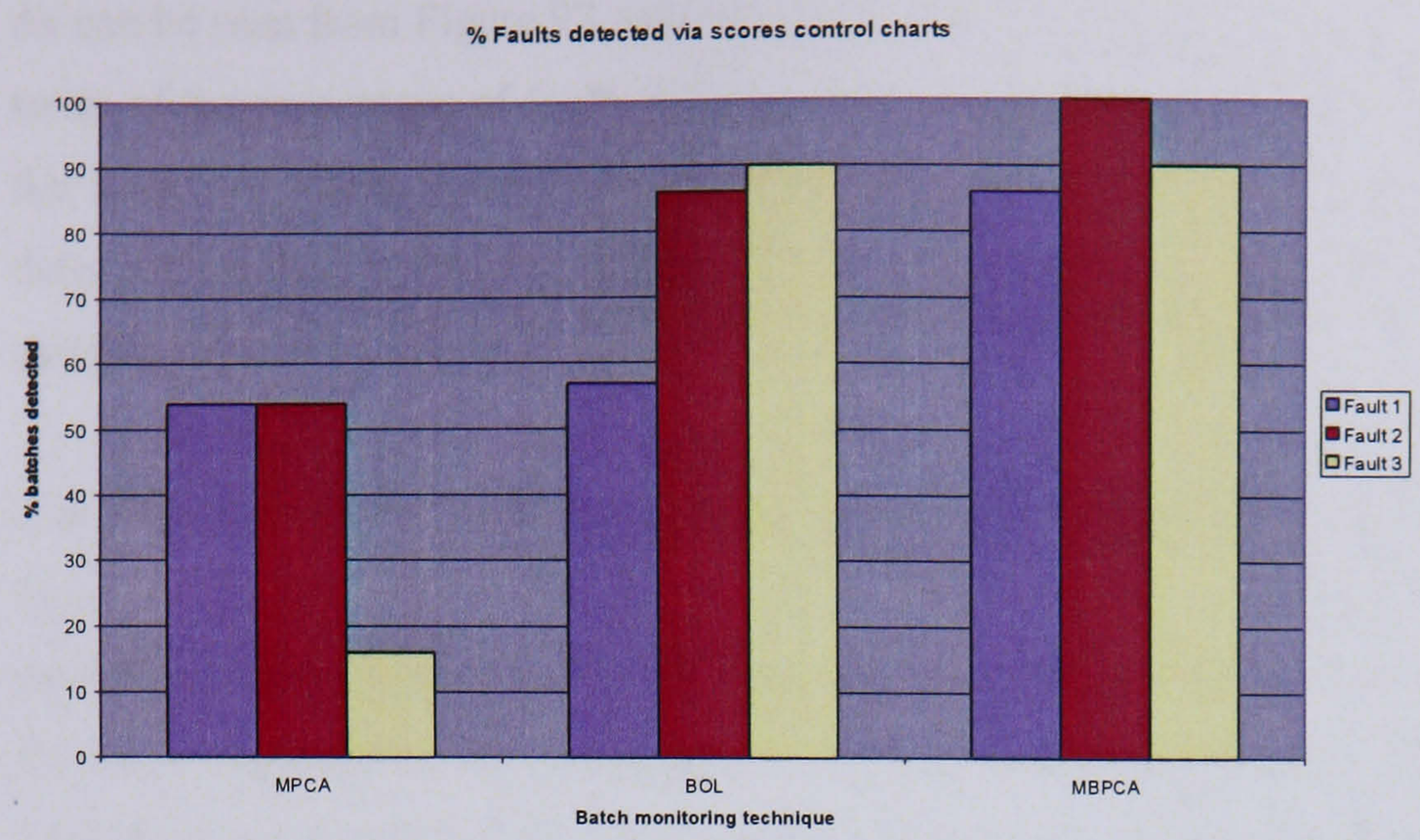


Figure 92: Percentage of faults detected via scores control charts

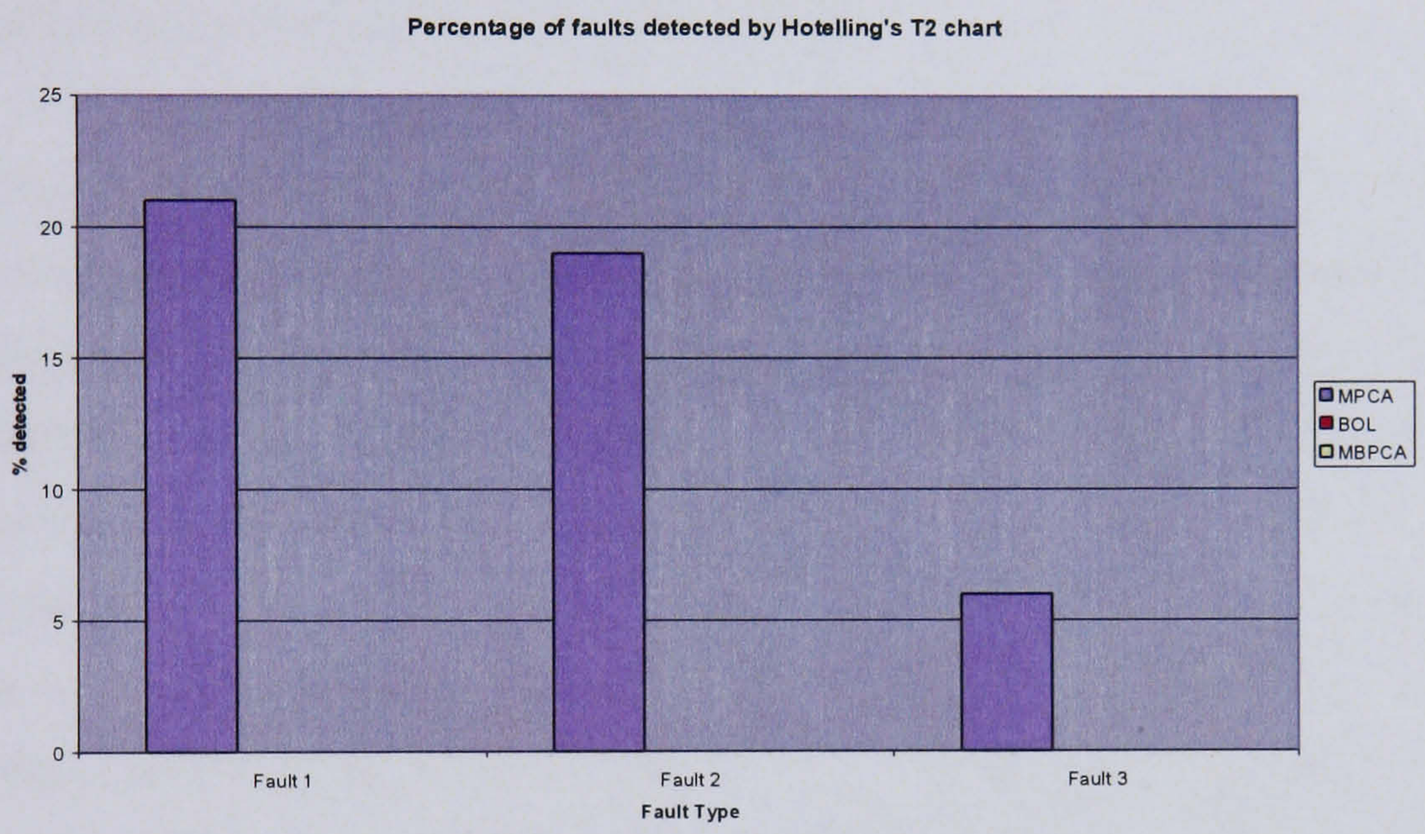


Figure 93: Percentage of faults detected by Hotelling's T^2 charts

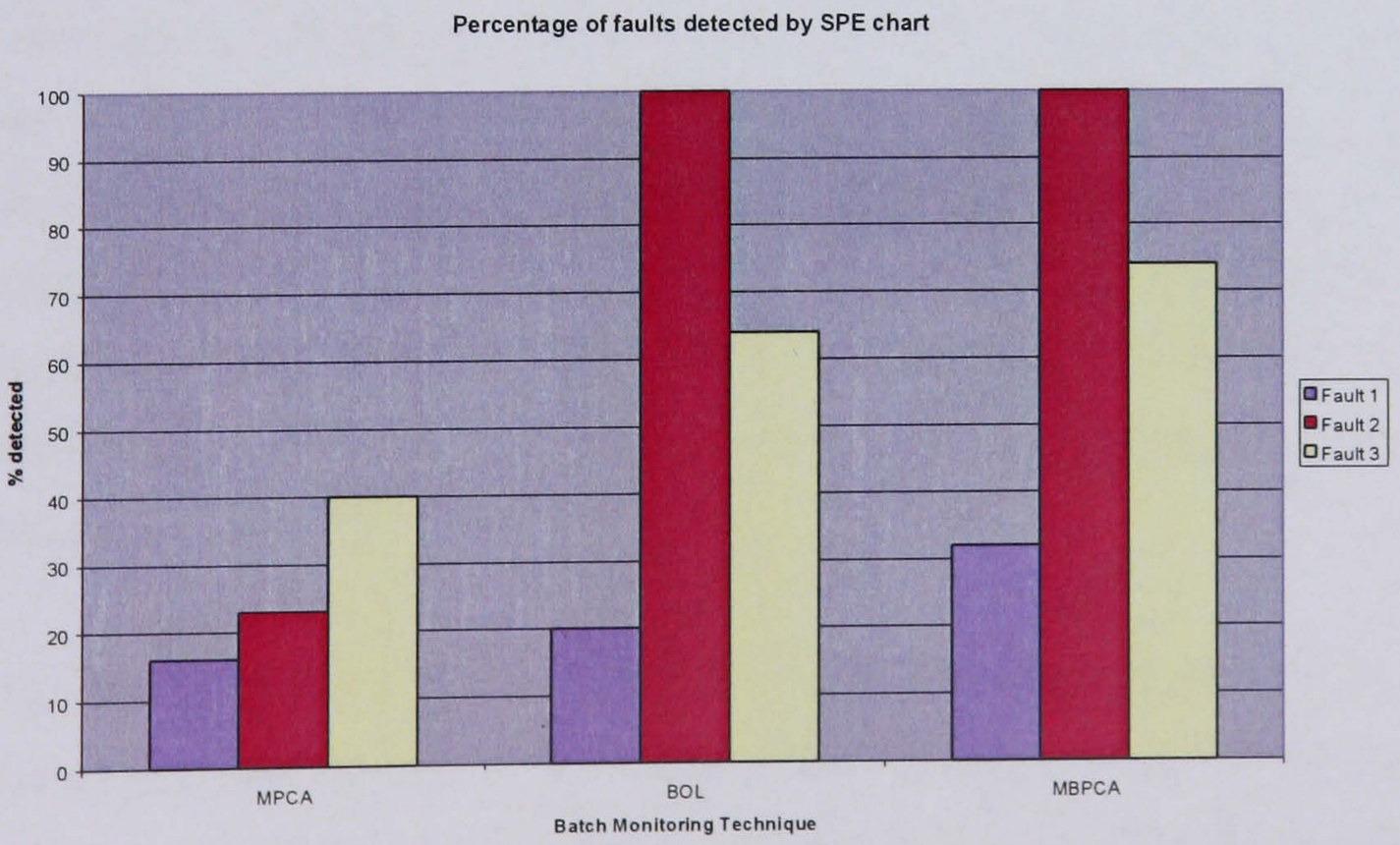


Figure 94: Percentage of faults detected by SPE charts

As can be seen from Figure 92 and Figure 94, model-based PCA is the most effective in terms of the percentage of faulty batches that were detected as being abnormal, via both the MBPCA scores control charts and the SPE control charts. However, the fault detection results for the batch observation level technique are similar to those achieved using the model-based algorithm.

It is difficult to make a judgement from the Hotelling's T^2 metrics included in Figure 93 because neither the BOL technique or the MBPCA technique managed to detect faults using this metric. This is because the variation occurring during the first 100 minutes of the batch is greater than the variation caused by the faults that are introduced. The control limits are calculated over the full length of the nominal batch data set, and so variation which is not as significant as that during the first 100 minutes may not be picked up by the Hotelling's T^2 control charts.

The fault detection results for the multiway PCA technique are not very good, particularly for fault type three, the valve fault, for the scores control charts. This is because once the fault was introduced, it deviated from the normal operating limits initially, before returning to normal operating conditions within a fairly short time period. In this study MPCA was applied as an end of batch technique, therefore it is possible that the batch trajectories were not out of control for a long enough period to have an effect on the end of batch scores. Examination of the SPE chart in Figure 94 shows that MPCA managed to detect 40% of the fault type 3 batches, compared to just 21% with the scores control chart. The squared prediction error looks at the variation in the data not predicted by the model and a change in this value indicates a change in the relationships between the variables, which would be expected from a fault in the cooling valve. This could be why the SPE plot is able to detect more abnormal process batches than the MPCA scores control charts. Another explanation for the poorer performance achieved overall by the MPCA technique is that the scaling method applied to the data has successfully removed the non-linear structure in the data, however the issue of serial correlation is still present. This will have impacted the fault detection ability of the technique.

The second aspect of the fault detection abilities of the techniques that was looked at was the average out of control run length i.e. the number of observations between the fault occurring in the process and the fault being detected by the control chart. For the batch to be considered to be going out of control, the batch trajectory had to remain outside of the control limits for at least three consecutive observations. The ARL could only be

assessed for the BOL and MBPCA techniques as MPCA is primarily an end of batch technique, although via techniques such as in-filling of data (see section 3.3) it can be applied as an on-line technique, however in this study the point at which the batch deviates from the control limits cannot easily be identified.

Figure 95 and Figure 96 summarise the ARLs calculated for the two through batch monitoring methods, via the scores control charts and the SPE charts. The Hotelling's T^2 statistics were not included as they were inconclusive for this data set. The run length for each batch was calculated as the shortest time for the fault to be detected on any of the principal components in the model.

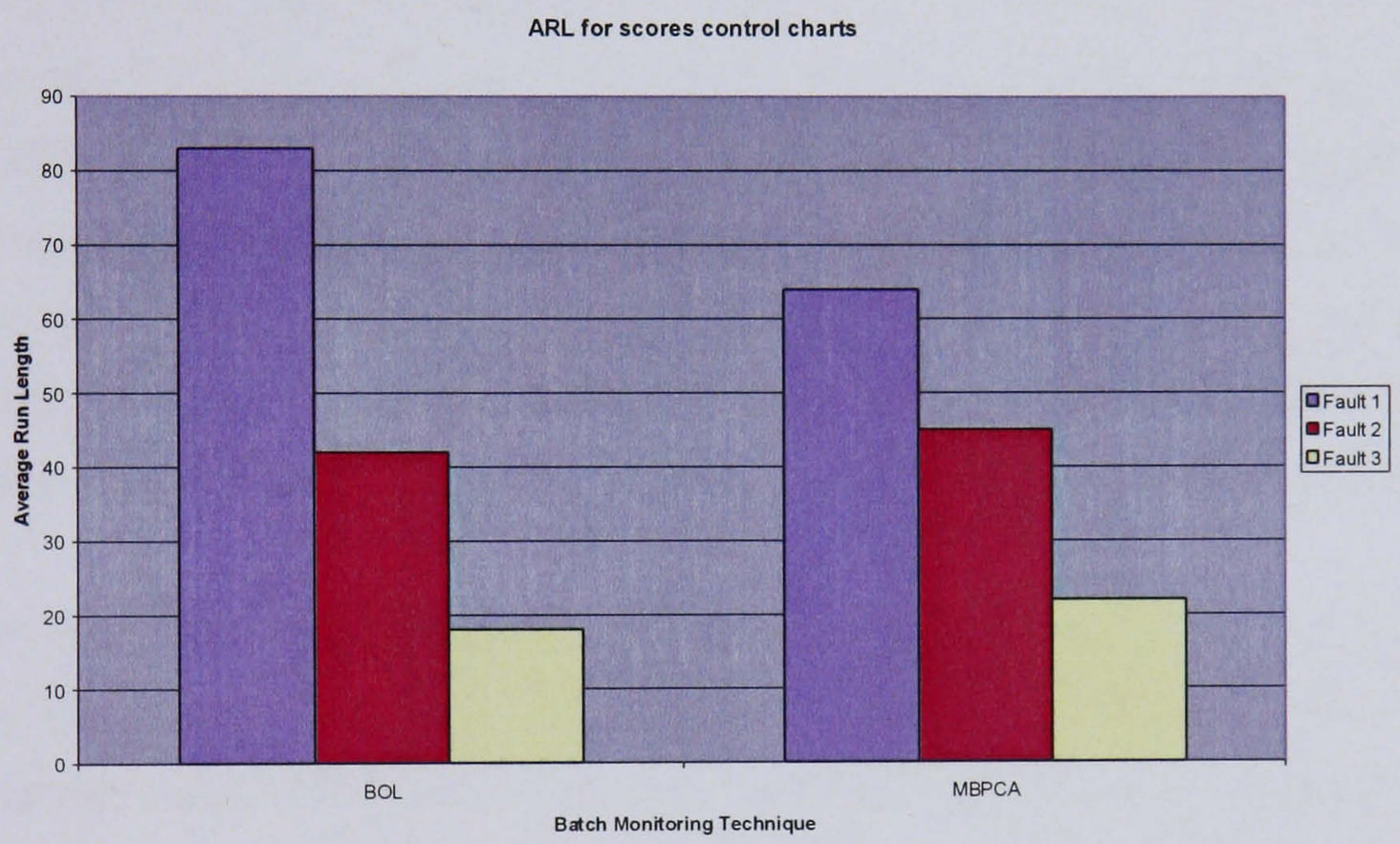


Figure 95: Average run lengths for scores control charts

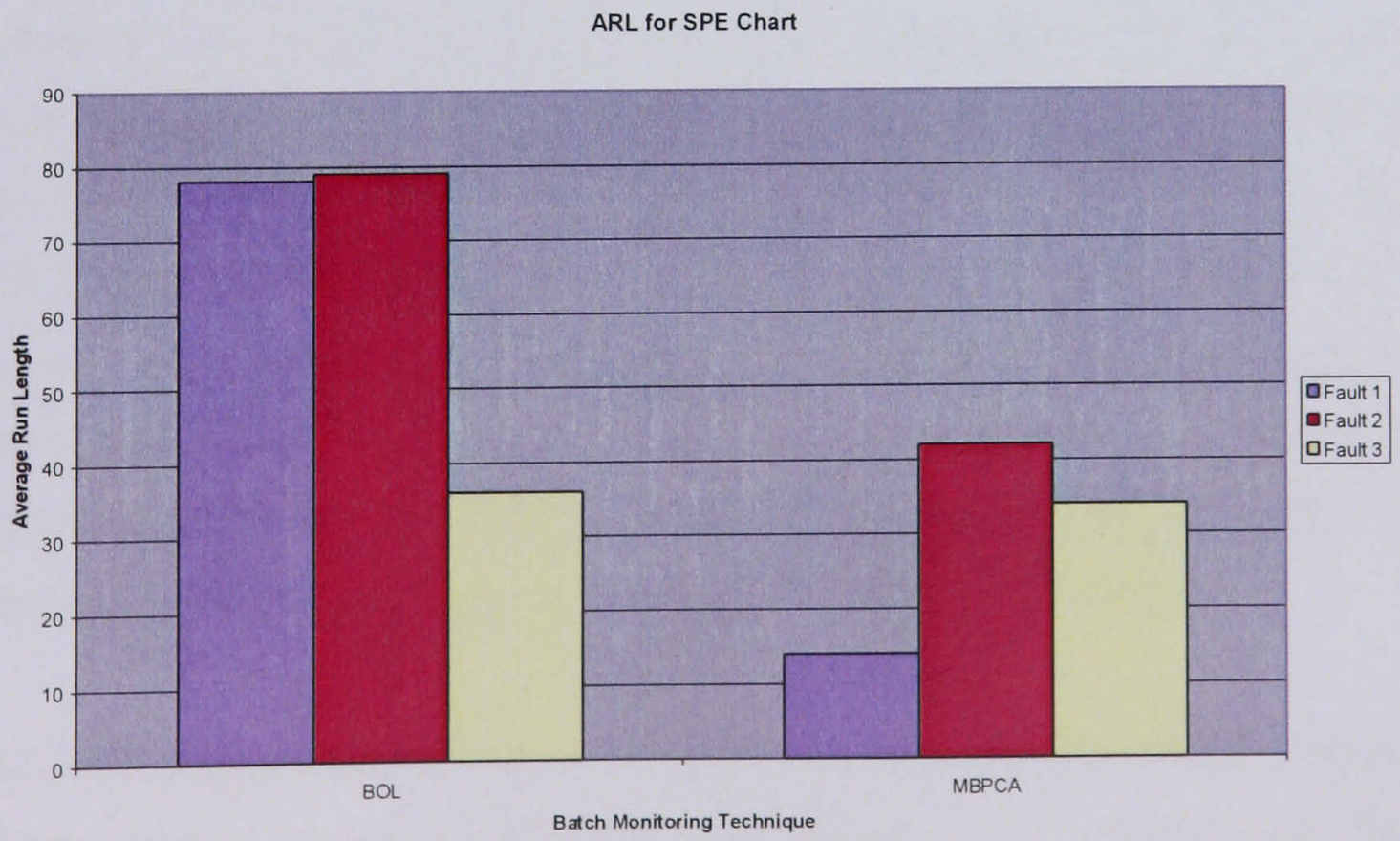


Figure 96: Average run lengths for SPE charts

It can be observed from the chart in Figure 95 that the average run lengths for the batch observation level technique and the model-based PCA technique are fairly similar. The BOL method produces a slightly shorter run length for faults 2 and 3, the heat transfer coefficient decrease and the cooling valve fault, whereas the MBPCA method has a shorter run length for the reactor temperature sensor fault. However, the difference between the two monitoring techniques is not significant. It would have been expected that the MBPCA technique would give shorter ARL results than the BOL technique, however the presence of plant model mismatch in the data means that the fault detection ability is impaired.

In terms of the average run lengths calculated from the SPE charts summarised in Figure 96, the BOL ARLs are significantly longer than the MBPCA ARLs for the first two faults. This could be because although there is a plant model mismatch present in the MBPCA data, it has still managed to remove a large amount of the structure in the data, leaving a set of residuals which are easier to model than the non-linear and dynamic process modelled in the BOL technique. The difference between the fault data and the residuals may be more obvious than the difference between the faults and the process data, therefore the SPE statistic would detect the MBPCA faults more rapidly.

The possibility of fault diagnosis for each technique was studied using contribution plots of the scores. As mentioned earlier in the chapter, the simulated fault impacted on variables other than those directly responsible for the abnormality because the values of all the process variables are calculated using the values of the variable with the simulated fault, as would occur with a complex fault which affects many different variables. This makes fault diagnosis more challenging. From the contribution plots from the MPCA scores it was possible to determine the cause, or related causes, of the first two faults, however it was not as clear for the third fault type, the valve fault. It would be expected that the first two faults would be easier to diagnose as they would have a strong effect on the temperature variables. The fault diagnosis performance for the BOL and MBPCA techniques was roughly the same, it was possible to determine the variables responsible for causing the faults in all cases, and the fourth variable, the valve position, did show a contribution for the valve fault in both cases.

From the analysis carried out, it is difficult to draw a conclusion as to whether the non-linear behaviour or the dynamic behaviour has more of an impact on the ability to detect faults on the exothermic batch process. Model-based PCA and BOL demonstrate fairly

similar fault detection abilities, with neither technique outperforming the other with respect to capturing the non-linear and dynamic behaviour in the process. MPCA is not as successful in the detection of abnormal batches, and although it deals with the non-linear behaviour in the data it does not address the dynamic structure. However, this does not necessarily indicate that the dynamic behaviour has more of an influence than the non-linear behaviour, the decrease in performance could be due to the fact that the technique was applied as an end of batch method as opposed to a through batch technique.

Of all the techniques examined in this case study, model-based PCA demonstrated the greatest potential as a batch monitoring technique. Even with a degraded model, the technique still gave fault detection results comparable to the other methods. Also, of the techniques studied MBPCA attempts to deal with both the non-linear and non-steady state behaviour of the batch process, features which the other methods did not jointly address, although the potential of the method was not fully realised in this study due to the error between the model and the plant. This indicates that the MBPCA technique has excellent potential as a batch performance tool if the issue of plant-model mismatch can be addressed. The next chapter deals with the extension of model-based PCA to make it more robust to plant-model mismatch.

7 CHAPTER SEVEN – SUPER MODEL-BASED MONITORING TECHNIQUES

7.1 Introduction

The evaluation of model-based multi-way PCA (Wachs and Lewin (1998)) on a simulation of a batch reactor, described in the previous chapter, showed how the technique used a mechanistic model of the process, to which small parameter errors were introduced, to try and remove the non-linear and dynamic structure in the data to make it more appropriate for use with a linear multivariate technique. From this analysis it could be seen that when a plant-model mismatch is present in the model-based technique there is only some reduction in the serial correlation and non-linearity in the data and hence the performance monitoring and fault detection of the monitoring scheme is not significantly enhanced in comparison to standard batch monitoring techniques. Significant improvements only materialise when a perfect model is fitted, as shown by Lewin (1998).

In an industrial environment, building a perfect model of a chemical process is not always feasible. The product life cycle of batch processes is often short and therefore it is not always possible to dedicate the time, resources and detailed studies necessary to obtain a thorough understanding of process behaviour required to build a comprehensive mathematical model for each product. Consequently this approach is not a realistic monitoring option. Furthermore, in some cases there will be processes where it will not be possible to model certain physical or biological phenomena.

However, the concept of model-based PCA still appears to be an appropriate approach for dealing with the characteristics of a batch process. Thus to overcome the limitations, a modified model-based approach, super model-based multi-way PCA, McPherson et al (2001) is proposed. This technique incorporates an additional stage whereby the structure remaining in the residuals, as a result of fitting a reduced complexity mechanistic model, is removed through the fitting of a model to the residuals. This results in a second set of residuals that are independent, identically and normally distributed. A number of techniques were investigated for the additional modelling stage including Partial Least Squares, Martin *et al* (1996), non-linear PLS, Qin *et al* (1992). AutoRegressive with eXogeneous input (ARX) model, Chen *et al* (2001), dynamic PLS,

Baffi *et al* (2000), dynamic non-linear PLS, Shi *et al* (2000) and dynamic Canonical Correlation Analysis (CCA), Schaper *et al* (1994).

The standard model-based technique and its performance monitoring results have previously been reported in Chapter 6. In this chapter, the super-model based techniques are evaluated and the results compared with the standard monitoring techniques. Section 7.2 of this chapter describes the super model-based algorithm. A case study utilising the exothermic batch reactor simulation that formed the basis of the analysis in Chapter 6 is then used to demonstrate the effectiveness of the technique. It is observed that the super model-based technique delivers improved performance monitoring in terms of a reduction in false alarm rates and enhanced fault detection ability in comparison with the standard MSPC batch monitoring techniques. The residuals are also assessed to see how well the technique addresses the issue of non-linear and dynamic behaviour.

7.2 Super Model-based Principal Component Analysis (SMBPCA)

7.2.1 Overview

As discussed in section 7.1, unless a perfect mechanistic model is available for application within model-based PCA, serial correlation and non-normality will be present in the residuals that are used to monitor the process. The presence of these characteristics will mask subtle changes in process behaviour, reducing the effectiveness of the performance monitoring scheme to that of a standard batch monitoring technique. As a perfect mechanistic model is typically not realisable in an industrial environment, a modified approach that allows for a reduced complexity model is proposed. Super Model-based Multi-way Principal Component Analysis (SMBMPCA) incorporates an additional residual modelling stage prior to the application of multi-way PCA to remove any remaining structure in the residuals. The resulting set of residuals is then assumed to be unstructured, i.e. they exhibit independent, identically and normal behaviour and can therefore be modelled using a linear statistical projection technique, such as multi-way PCA, Wold *et al* (1998). A schematic is shown in Figure 97 and the algorithm is summarised in section 7.2.2.

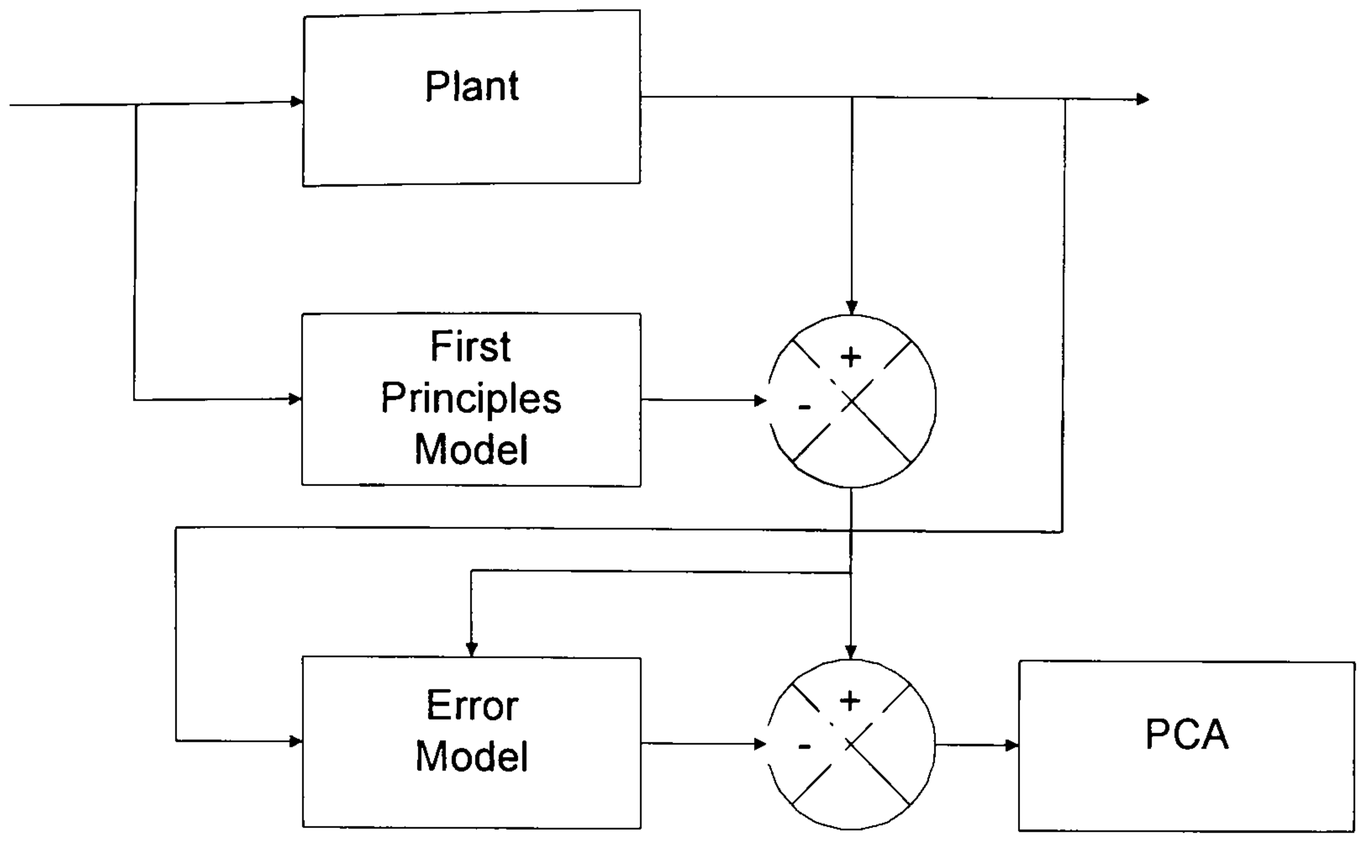


Figure 97: Schematic of super model-based Principal Component Analysis (SMBPCA)

7.2.2 Super Model-based PCA Algorithm

1. Data from the plant is collected under normal operating conditions.
2. The values inferred from the reduced complexity mechanistic model are calculated and subtracted from the plant measured values giving a matrix of structured residuals.
3. The structured residuals are standardised to zero mean and unit variance.
4. An error model is then used to infer the values of the structured residuals from the original variables.
5. The inferred values of the residuals are subtracted from the original structured residuals, resulting in a set of unstructured residuals.
6. The unstructured residuals are standardised to zero mean and unit standard deviation.
7. Batch observation level multi-way PCA is then applied to the standardised unstructured residuals and the confidence limits for the nominal representation are calculated for the monitoring metrics, scores, Hotelling's T^2 and squared prediction error.
8. The monitoring of new batches is then based on the confidence limits calculated from the analysis performed on the nominal data set.

7.2.3 Error Model Types

The super model-based multi-way PCA algorithm was investigated using a number of different error models, which can be summarised as local models, non-linear models and

dynamic models and were introduced in Chapter 3. The first type of error model used in conjunction with the SMBMPCA technique was a linear partial least squares model (PLS). Further investigation of the SMBMPCA technique was carried out by implementing it using a non-linear PLS model with a quadratic function. A number of different types of dynamic model were also tested, including an AutoRegressive with eXogenous (ARX) variables time series model, dynamic PLS, non-linear dynamic PLS and dynamic Canonical Correlation Analysis.

In the following sections, nominal models are built using the SMBMPCA technique with the different types of error model. These models are examined to determine how well the technique deals with the non-linear and dynamic structure in the data, and then the fault detection ability of each technique is assessed.

7.3 Nominal Models

7.3.1 Super model-based PCA with PLS error model

7.3.1.1 Nominal Model

To apply the super model-based PCA technique using a batch PLS error model, inferred values from the solved reduced complexity mechanistic model (reactor temperature, wall temperature, jacket temperature and cooling valve position) were subtracted from the same four process variables in the 50 nominal batches. This creates a set of structured residuals for each batch which are scaled to zero mean and unit variance. A PLS model is then built where **X**, the predictor variables are the original variables, and **Y**, is the set of structured residuals. The purpose of including the error model in the SMBMPCA model is to create a set of unstructured residuals. To do this, **Y**, the inferred values of the residuals generated from the PLS model, are subtracted from the original set of structured residuals produced using the mechanistic model and process variables. This, in theory, gives a set of unstructured residuals to which a linear statistical projection technique can be applied, in this case batch observation level PCA (detailed in Chapter 3).

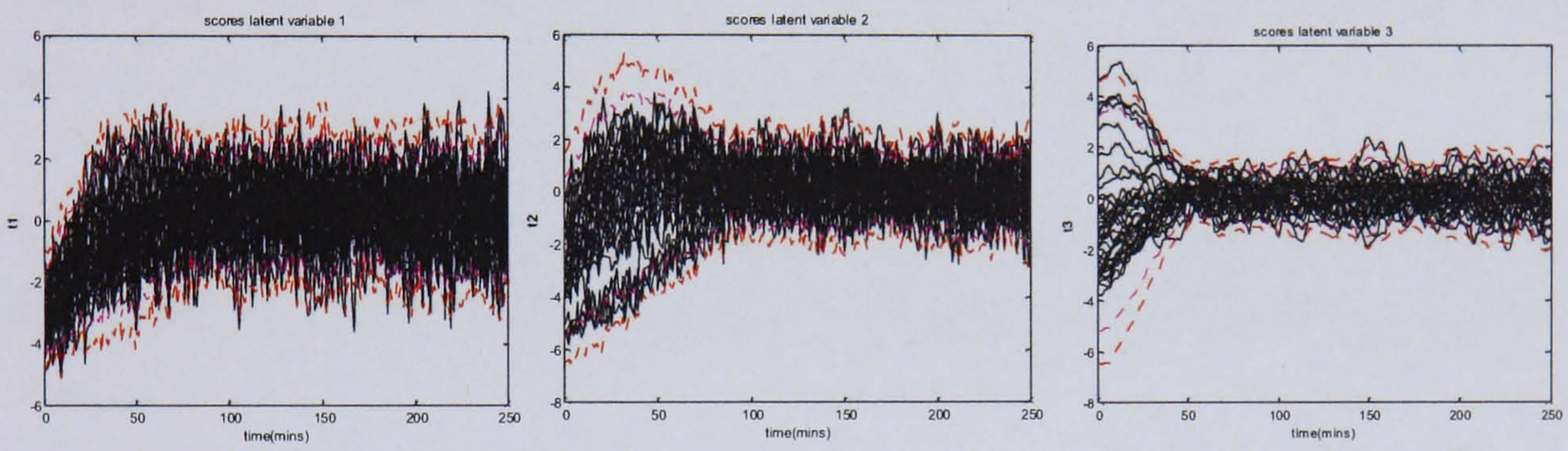
Batch observation level PCA is applied to the set of unstructured residuals created from the nominal data set, with time as the maturity variable, as with the case study in Chapter

6. Cross-validation was carried out to determine how many principal components should be retained, Table 8.

No. PCs	R^2x	Cum R^2x	Eigen-values	R^2y	Cum R^2y	Q^2	Cum Q^2
1	0.407	0.407	1.63	0.113	0.113	0.113	0.113
2	0.387	0.794	1.55	0.0242	0.138	0.0272	0.138
3	0.196	0.99	0.785	0.0128	0.151	0.0148	0.15

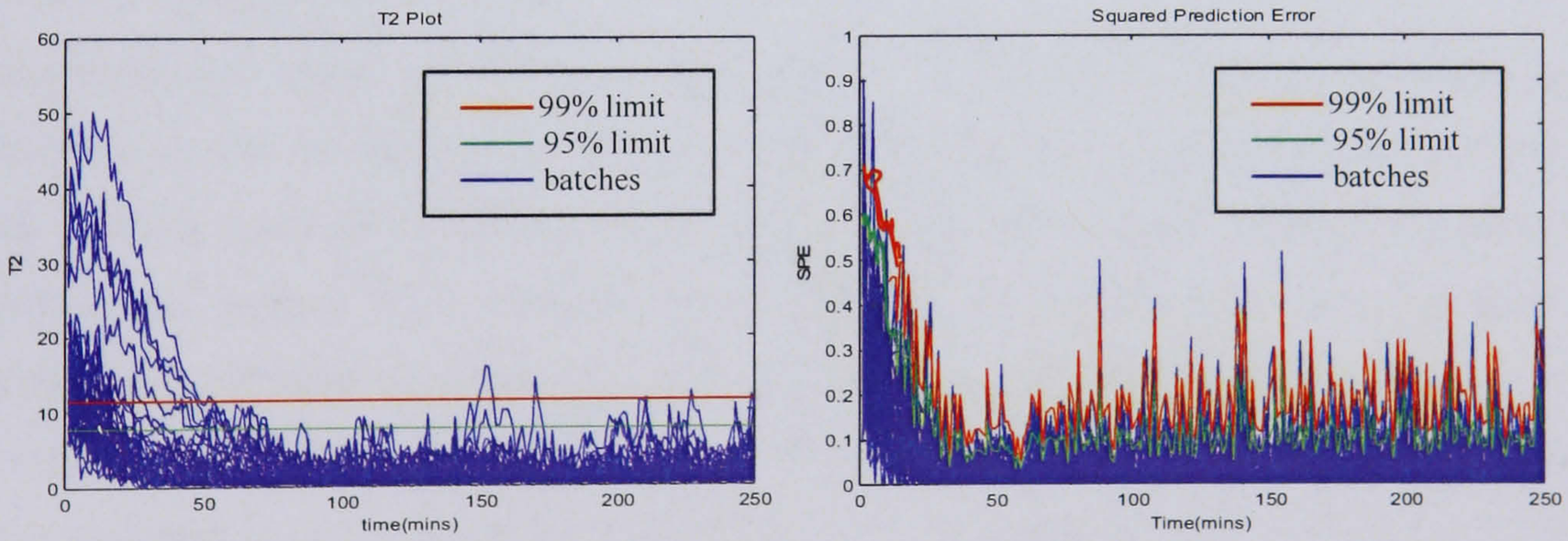
Table 8: Results of cross validation analysis for SMBMPCA using PLS error model

Three latent variables are retained in the model. The nominal plots of the SMBPCA with PLS model are shown in Figure 98 and Figure 99.



a) Latent variable 1 b) Latent variable 2 c) Latent variable 3

Figure 98: Control charts for nominal SMBMPCA with PLS model



a) Hotelling's T^2 plot of nominal data b) SPE plot of nominal data

Figure 99: Hotelling's T^2 and SPE plots of nominal data

There are two noticeable differences between the control charts produced when applying the model-based technique and those produced when applying the super model-based technique. Firstly, the structure in the data appears to have been reduced, plots of the nominal model-based control charts (Figure 45, Chapter 6) showed the magnitude of the

scores varying between -8 and 10. In Figure 98, the control limits have been smoothed, and only vary between -4 and 3, and are fairly constant after the first 50 time points. The second noticeable change is that the signal to noise ratio has increased, which could indicate that the objective of the super model-based technique, to remove the non-linear and dynamic behaviour inherent in the process leaving residuals containing white noise which can be analysed using a linear multivariate statistical technique, has been successful. The other possible explanation of this increase is noise is that the model has been over-fitted.

The Hotelling's T^2 plot of the nominal data set shows that during the first 50 time points a large number of the batch trajectories lie outside of the control limits, makes detection of process abnormalities difficult as it is expected that several of the fault batches will be outside of the control limits before the fault is introduced. The variable limits on Figure 99(b) mean that this is not a problem for the SPE control charts, although the batch trajectories do occasionally move inside and outside of the limits.

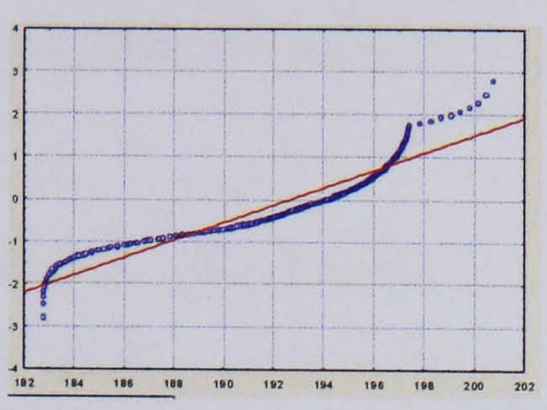
The effectiveness of the technique in removing the non-linear and dynamic structure is investigated in the next section, where the residuals involved in building the model are examined.

7.3.1.2 Non-linear and Dynamic Behaviour

7.3.1.2.1 Non-linear Structure

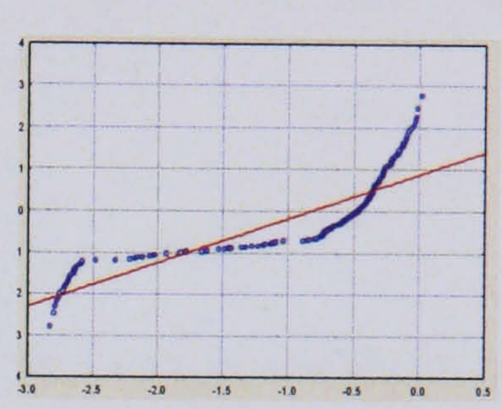
Super model-based PCA addresses the non-linear and dynamic behaviour inherent in batch processes through the use of error models to create a set of unstructured residuals on which a linear statistical technique, such as batch observation level analysis can be performed. Section 7.2.2 discussed how the initial set of residuals was created by subtracting the inferred mechanistic model values from the process variables. This set of residuals still contains structure if plant model mismatch is present. Therefore to ensure the structure is removed, an error model is built, inferring the values of the residuals from the original process variables. Subtracting the inferred residuals from the first set of residuals gives a set of unstructured residuals on which batch observation level analysis can be performed. It is this set of residuals which will be examined in this study, to determine if the technique has been successful in removing the structure before the BOL technique was applied.

An assessment of the non-linear structure present in the SMBMPCA with PLS model residuals is shown in Figure 100. Each row of the three plots shows the probability plot of the original variable, the corresponding residual from the application of the model-based PCA technique to that variable, and the residual from the application of the super model-based to the variable. Residual 1 indicates the residual of the reactor temperature, residual 2 the residual of the wall temperature etc.

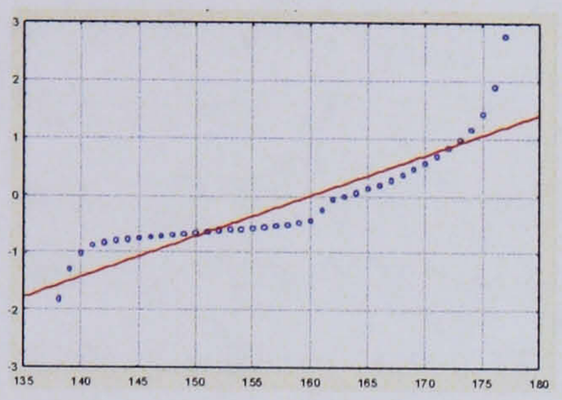


a) Reactor temperature – raw

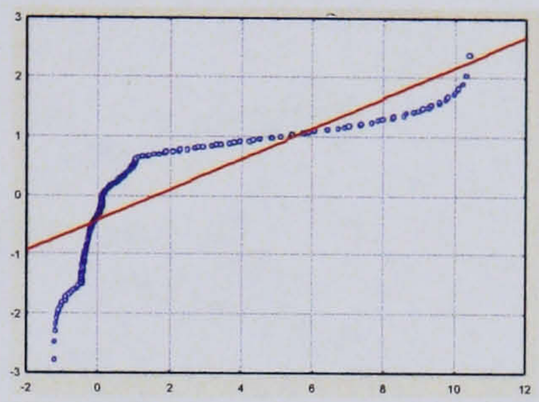
b) Model-based residual 1



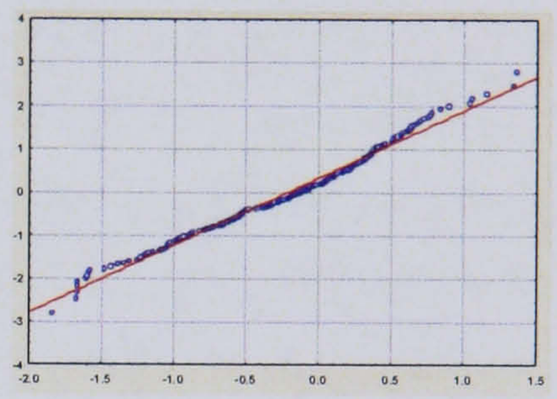
c)Residual 1



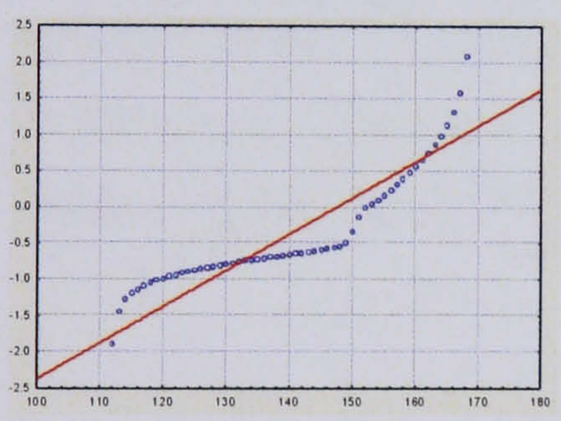
d) Wall temperature –raw



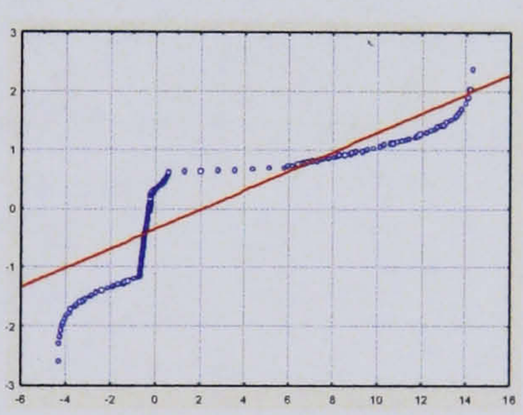
e) Model-based residual 2



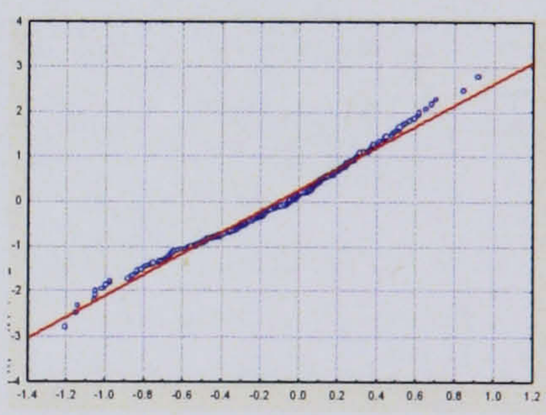
f) SMB residual 2



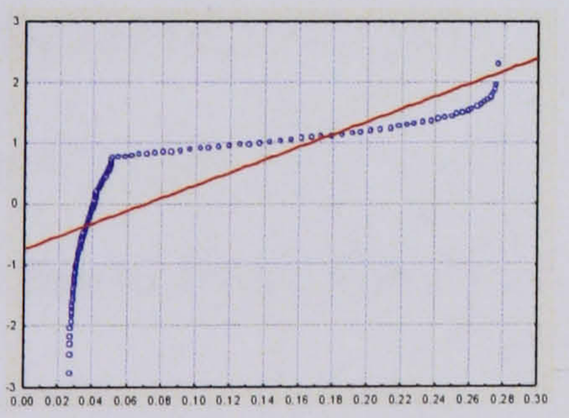
g) Jacket temperature – raw



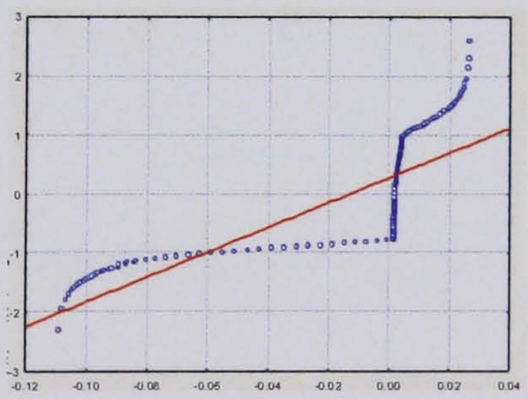
h) Model-based residual 3



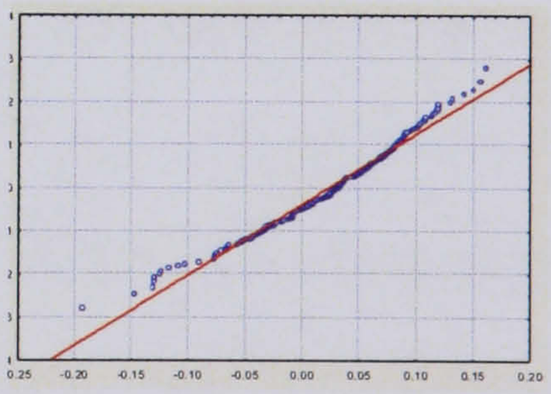
i) SMB residual 3



j) Cooling valve position – raw



k) Model-based residual 4



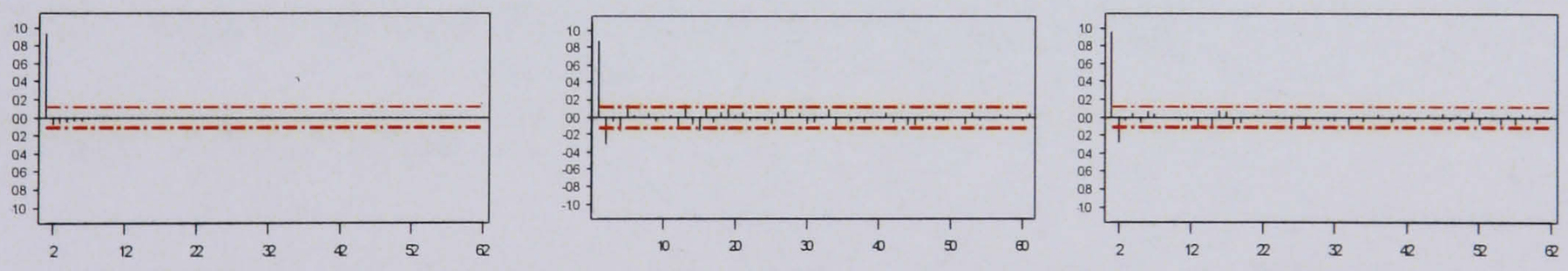
l) SMB residual 4

Figure 100: Normal probability plots of original data, model-based residuals and SMBMPCA residuals

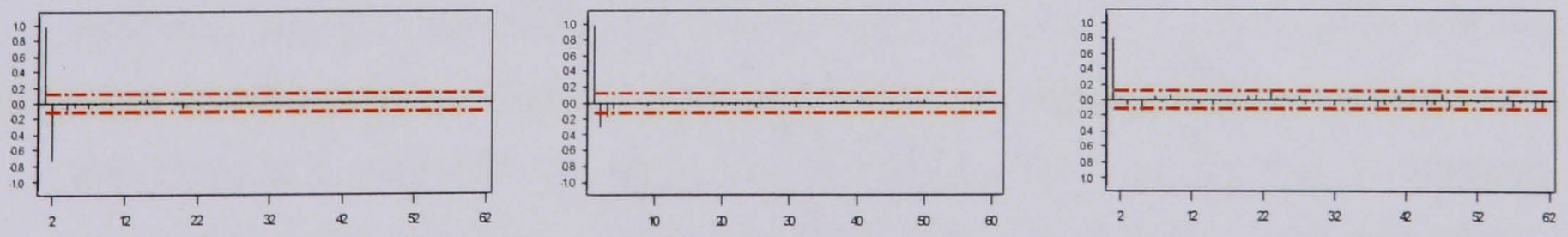
As observed from Figure 100, application of the super model-based technique using PLS has significantly reduced the level of non-linearity from the residuals, in comparison with the original data and the model-based residuals. Residuals 2, 3 and 4, originally the wall temperature, jacket temperature and valve position, now demonstrate fairly linear behaviour. However, residual 1, the reactor temperature, still demonstrates non-linear behaviour. This can be attributed to the fact that the partial least squares technique, Martin *et al* (1996), is not designed to cope with non-linear structure in the data and has therefore not modelled the residual behaviour as effectively as required.

7.3.1.2.2 Dynamic Structure

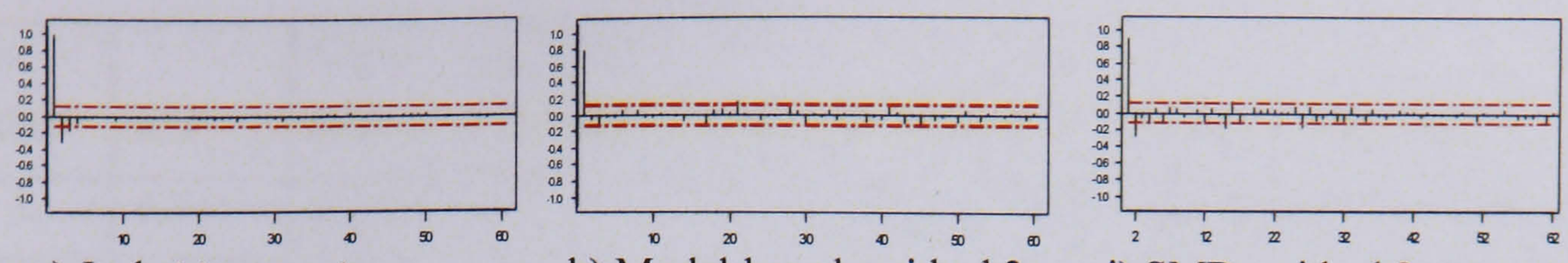
The other aspect of the residuals investigated was the dynamics, or structure, remaining in the data after the application of the super model-based technique. As with the probability plots, the residuals looked at were those generated after the application of the PLS model. One way of assessing the amount of structure remaining in the data is to look at partial autocorrelation plots of the residuals, as it has already been determined that the process in question is autoregressive. The partial autocorrelation plots define the process order of each variable or residual, which basically measures the strength of the correlation amongst consecutive observations. If the process order is reduced to zero then the technique has been successful in removing the dynamic structure from the residuals. Figure 101 shows the partial autocorrelation plots for the original data, the model-based residuals, and the super model-based residuals. This allows any improvements to be quantified.



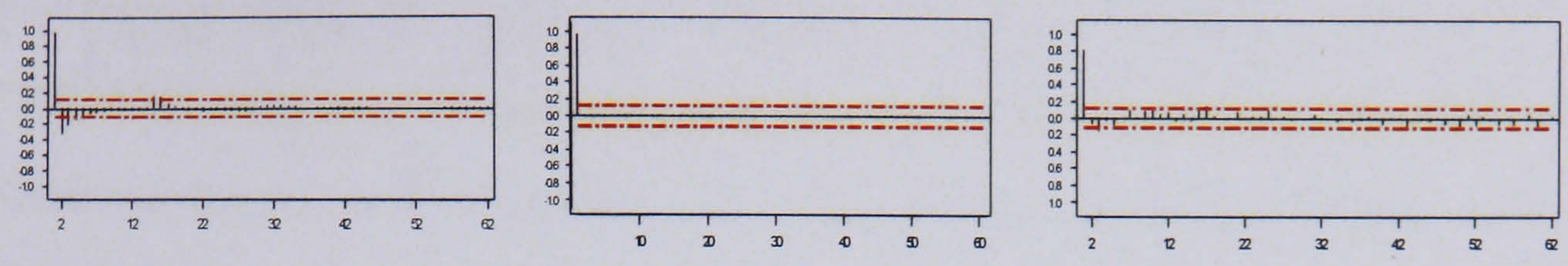
a) Reactor temperature – raw b) Model-based residual 1 c) SMB residual 1 (after PLS)



d) Wall temperature – raw e) Model-based residual 2 f) SMB residual 2



g) Jacket temperature h) Model-based residual 3 i) SMB residual 3



j) Valve position – raw k) Model-based residual 4 l) SMB residual 4

Figure 101: Partial autocorrelation plots for original data, model-based and SMBMPCA residuals

The application of SMBMPCA to the data has had reasonable results with respect to the reduction of the dynamic structure in the data. Variables 2, 3 and 4 (wall temperature, jacket temperature and valve position respectively) show a reduced degree of partial autocorrelation in comparison to the original process variables. Examining the data to determine whether the technique demonstrates improvements in comparison to the model-based method, again, 3 out of the 4 variables show an improvement. However, despite the improvement, all variables still display some structure.

7.3.2 Super model-based PCA with non-linear PLS error model

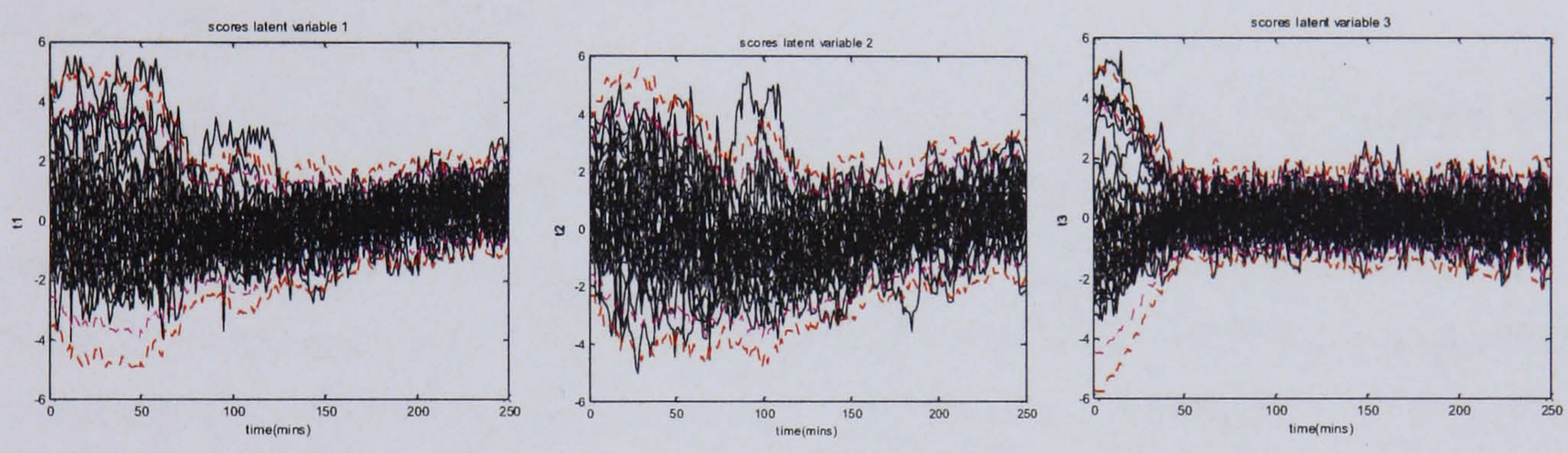
7.3.2.1 Nominal Model

SMBMPCA with a PLS error model achieved improvements in terms of removing the non-linear behaviour in the batch data, however the non-linearity was not completely removed. Therefore the second type of error model considered in conjunction with the super model-based technique was non-linear partial least squares, Baffi *et al* (1999). The non-linear PLS technique was applied to the residuals generated from the mechanistic model, the resulting residuals were then modelled using batch observation level analysis. The results of the cross validation analysis on the nominal data are summarized in Table 9.

No PCs	R^2x	Cum R^2x	Eigen-values	R^2y	Cum R^2y	Q^2	Cum Q^2
1	0.333	0.333	1.33	0.0209	0.0209	0.0208	0.0208
2	0.449	0.782	1.8	0.00226	0.0232	0.00225	0.023
3	0.184	0.976	0.737	0.0011	0.0243	0.00118	0.0242

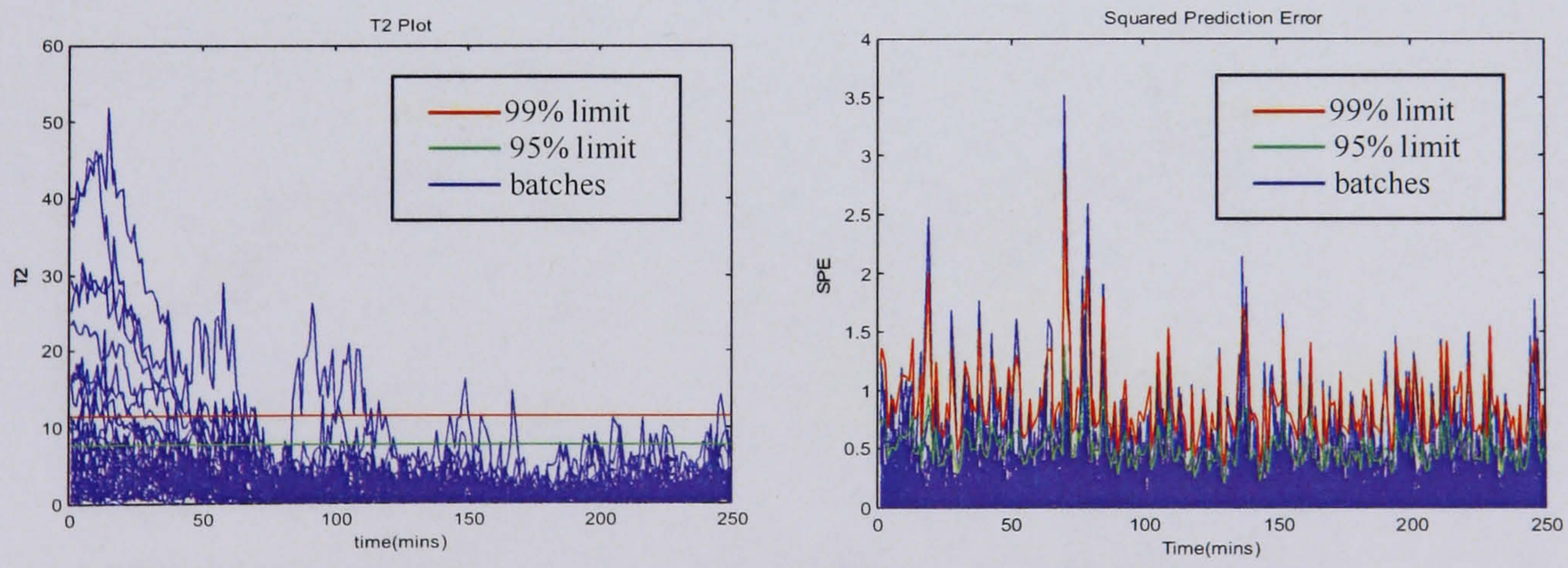
Table 9: Summary of cross-validation results for SMBMPCA with non-linear PLS error model

Three latent variables were retained to represent the nominal model, capturing 98% of the variance in the data. The batch observation level scores plots for these three latent variables are shown in Figure 102.



a) Latent variable 1 b) Latent variable 2 c) Latent variable 3

Figure 102: Nominal scores plots for SMBMPCA with non-linear error model



a) Hotelling's T^2 plot of nominal data b) SPE plot of nominal data

Figure 103: Hotelling's T^2 and SPE plot of nominal data

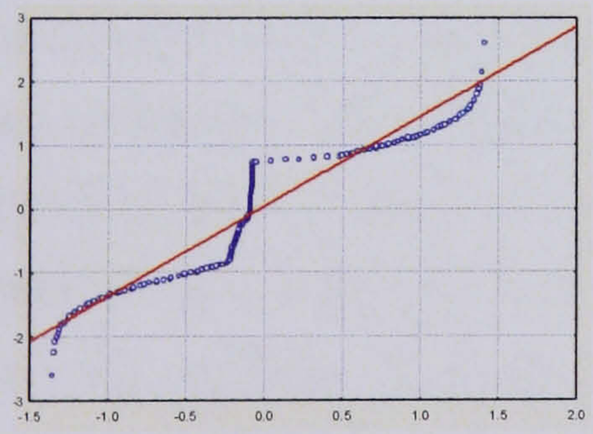
The plots of the nominal model demonstrate similar behaviour to those produced using the SMBMPCA method with a PLS model, the structure has been reduced and the noise

in the data has increased. Due to the noise in the data, two batches exceed the control limits. However this is acceptable for a confidence limit of 99%. The Hotelling's T^2 control chart of the nominal model again shows nine batches outside of the control limits during the first 50-100 minutes of operation, due to the large amount of variation that occurs during the first phase of operation of the process. The process attains steady state during the second phase of operation and only 1 batch lies outside of the limits. The SPE plot shows the majority of batches within the control limits, however the noise or over-fitting of the data causes two batches to exceed the control limits occasionally, but not for more than three consecutive time points and therefore they are not considered to be out of statistical control.

7.3.2.2 Non-linear and Dynamic Behaviour

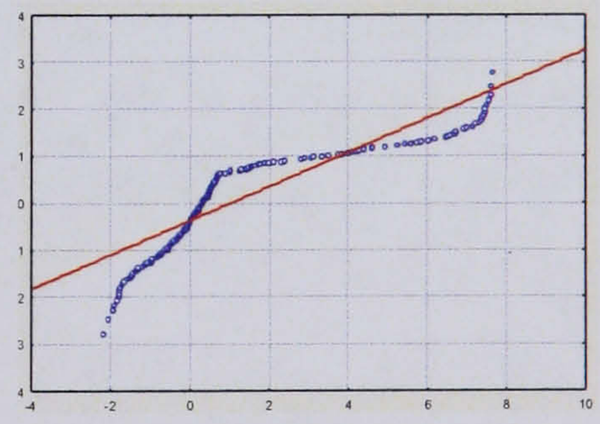
7.3.2.2.1 Non-linear Structure

The objective of the super model-based technique is to remove the non-linear and dynamic behaviour in the data so that a linear technique such as batch observation level analysis can be applied. To determine the effectiveness of the SMBMPCA with the non-linear PLS technique, the residuals generated after the application of the non-linear PLS technique were assessed i.e. before the BOL technique is performed on the data. The normal probability plots of the super model-based residuals are in Figure 104, with the plots for the model-based residuals for comparison. The plots of the raw data are shown in Figure 100.

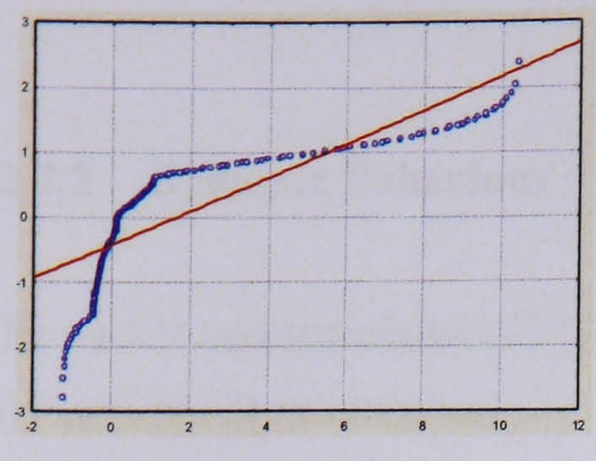


Reactor Temperature

a) Model-based residual 1

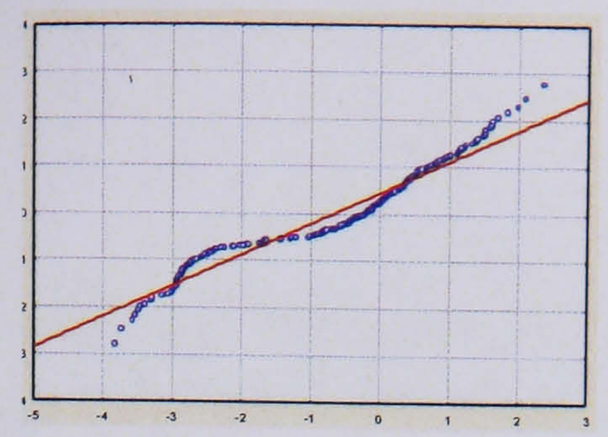


b) Super model-based residual 1

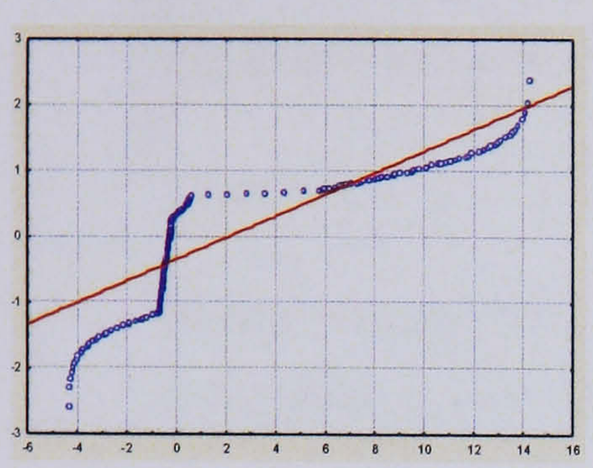


Wall Temperature

c) Model-based residual 2

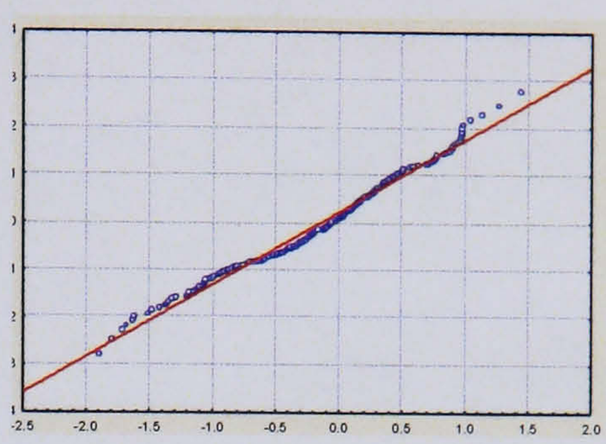


d) Super model-based residual 2

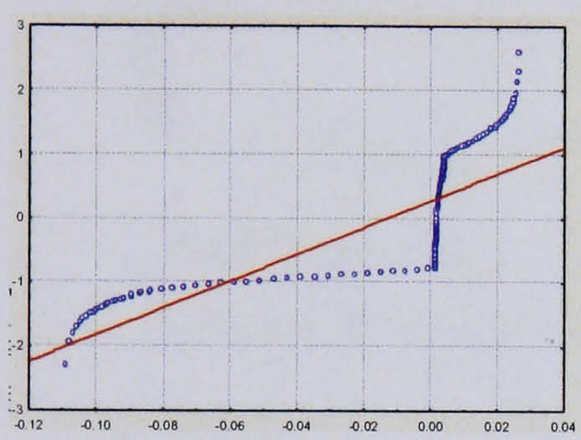


Jacket Temperature

e) Model-based residual 3

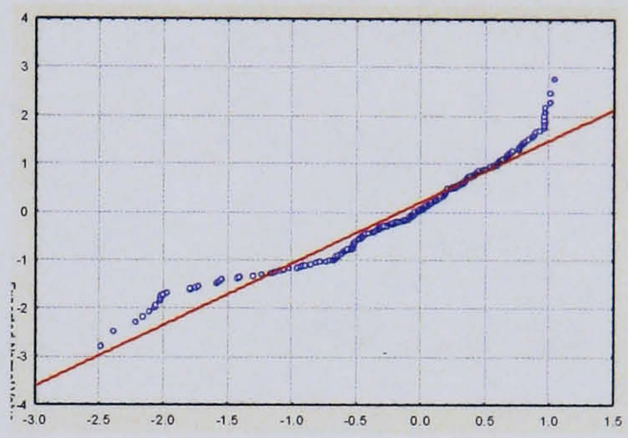


f) Super model-based residual 3



Cooling Valve Position

g) Model-based residual 4



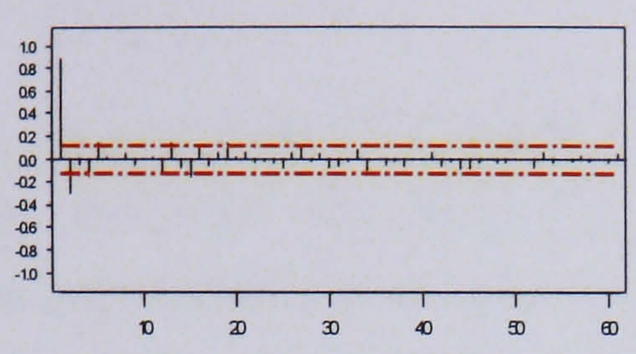
h) Super model-based residual 4

Figure 104: Probability plots for MBPCA and SMBPCA with non-linear PLS residuals

All 4 sets of residuals show an improvement in terms of a reduction in the non-linearity through the use of the SMBPCA approach in comparison to the model-based approach. However, when compared to the results achieved with the linear PLS error model, the reduction in non-linearity is not as good for variables 2,3 and 4. This should not be the case the non-linear PLS algorithm is designed to deal with this aspect of the data, whereas the linear PLS technique is not. This could be because the non-linear PLS algorithm chosen is not the most effective one for this application.

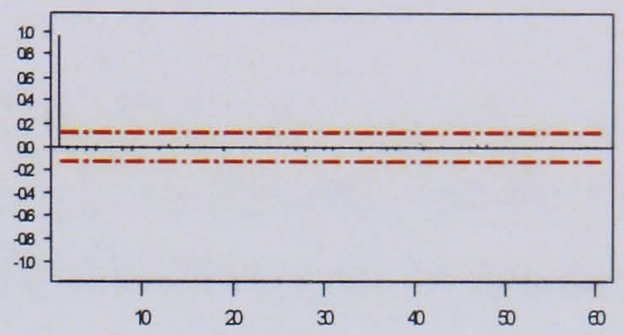
7.3.2.2.2 Dynamic Behaviour

The partial autocorrelation plots of the super model-based residuals are shown in Figure 105 with the plots of the standard model-based residuals for comparison.

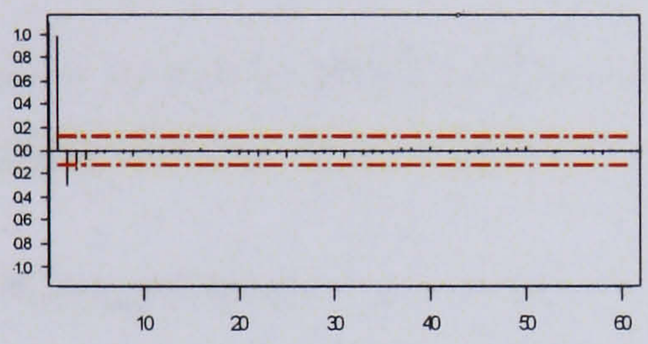


Reactor Temperature

a) Model-based residual 1

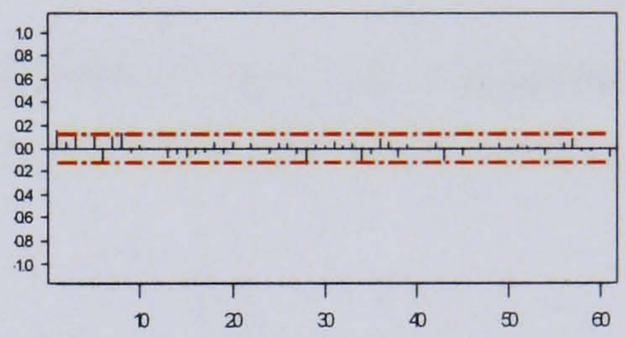


b) Super model-based residual 1

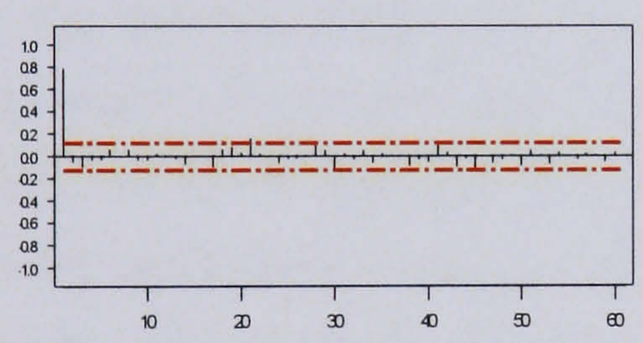


Wall Temperature

c) Model-based residual 2

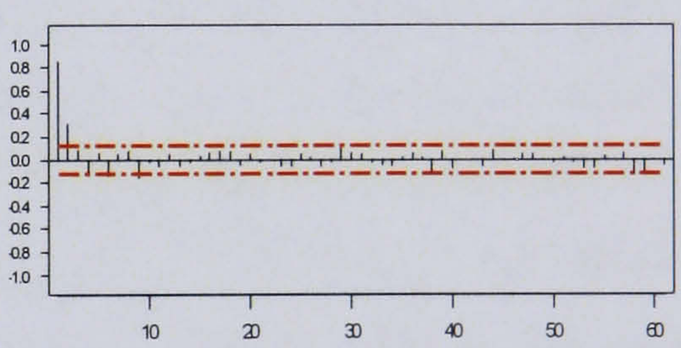


d) Super model-based residual 2

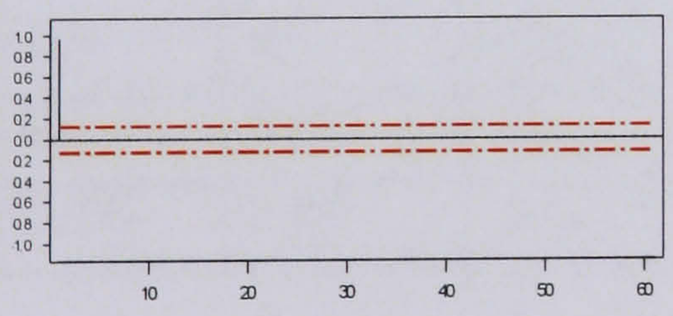


Jacket Temperature

d) Model-based residual 3

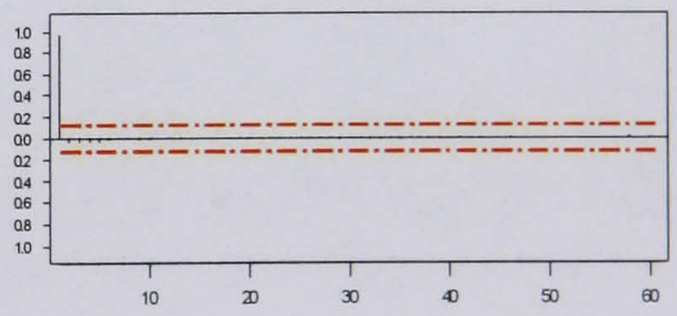


e) Super model-based residual 3



Cooling Valve Position

f) Model-based residual 4



g) Super model-based residual 4

Figure 105: Partial autocorrelation plots for model-based and SMB with resulting non-linear PLS residuals

In this case, the residuals show that the dynamic structure has been removed from variable 2, the wall temperature, whereas the improvement in the other three variables is negligible or non-existent, in fact the serial correlation is observed to increase in variable 3, the jacket temperature. The removal of the dynamic structure from variable 2, the wall temperature indicates that the dynamics have been well modelled by the non-linear PLS technique.

Overall, the non-linear PLS model is not as effective in removing the non-linear properties of the data as the linear PLS model and only partially removes the dynamic structure. Improvements in this area can be achieved through the use of a dynamic error model, as discussed in the following sections.

7.3.3 Super model-based PCA with autoregressive with eXogeneous (ARX) input model

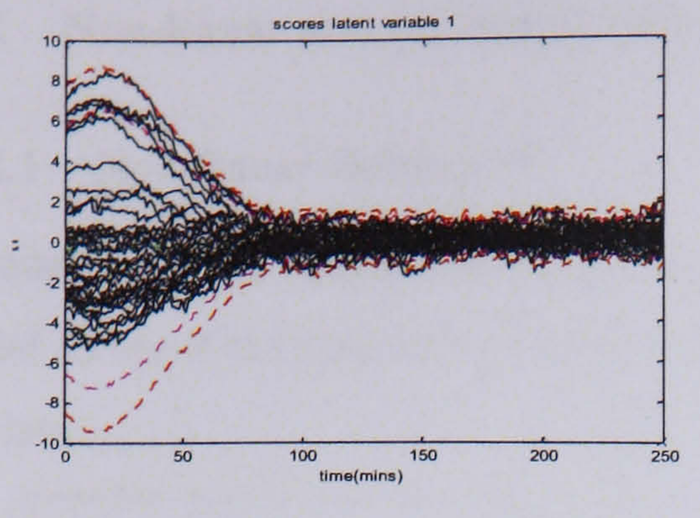
7.3.3.1 Nominal Model

Although the PLS techniques showed some promising results when dealing with the model-based residuals, they are not designed to deal with the dynamic behaviour present within the residuals. Better results could be expected from applying a time series technique which deals with the dynamic structure in the data. The super model-based technique was therefore applied using an ARX (Autoregressive with eXogeneous Input) time series model to deal with the dynamic structure of the residuals, as described in Chapter 6. Using the ARX model as an error model, the SMBMPCA technique was applied to the nominal data set of 50 batches.

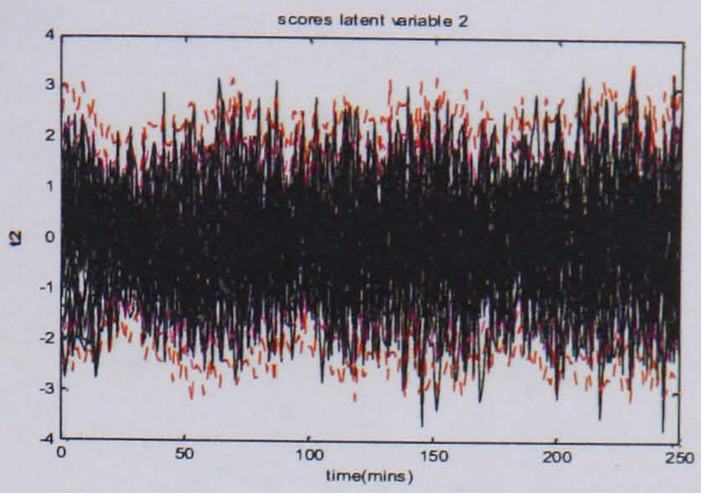
No PCs	R^2x	Cum R^2x	Eigen-values	R^2y	Cum R^2y	Q^2	Cum Q^2
1	0.439	0.439	1.75	0.0153	0.0153	0.0153	0.0153
2	0.533	0.971	2.13	0.00146	0.00168	0.00147	0.0168

Table 10: Summary of cross validation results for SMBMPCA with ARX error model

The nominal model was built from two latent variables. capturing 97% of the variance in the data:

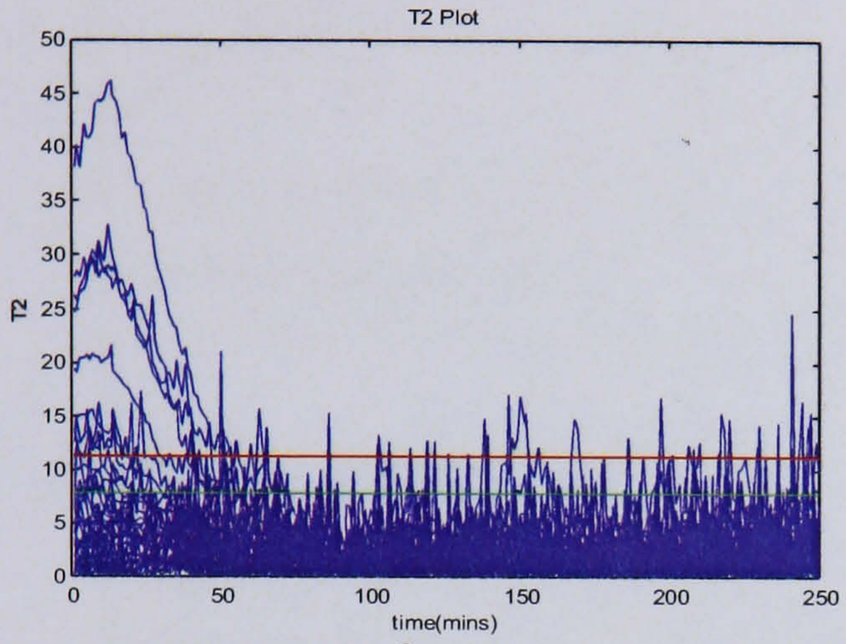


a) Latent variable 1

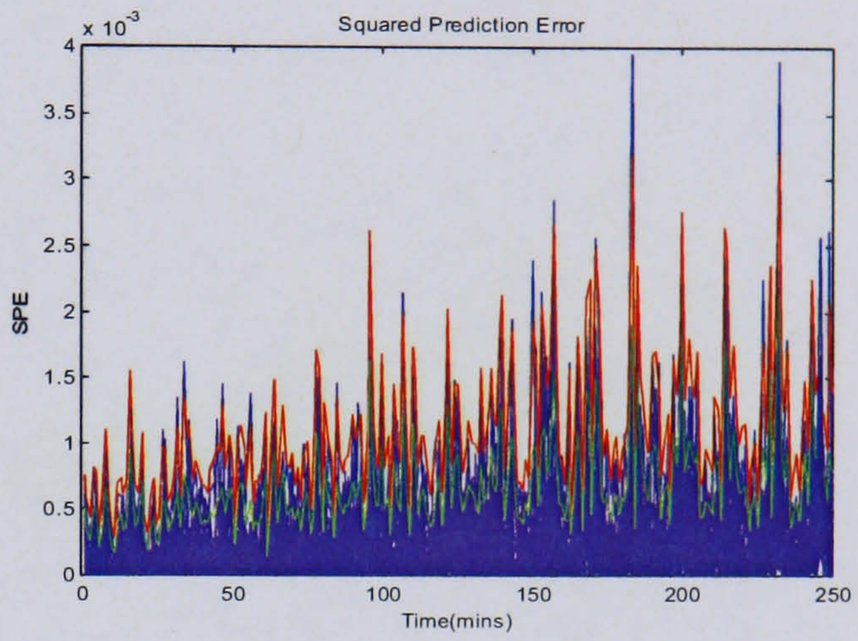


b) Latent variable 2

Figure 106: Nominal scores plots for SMBMPCA with ARX error model



a) Hotelling's T^2 plot for nominal data



b) SPE plot for nominal data

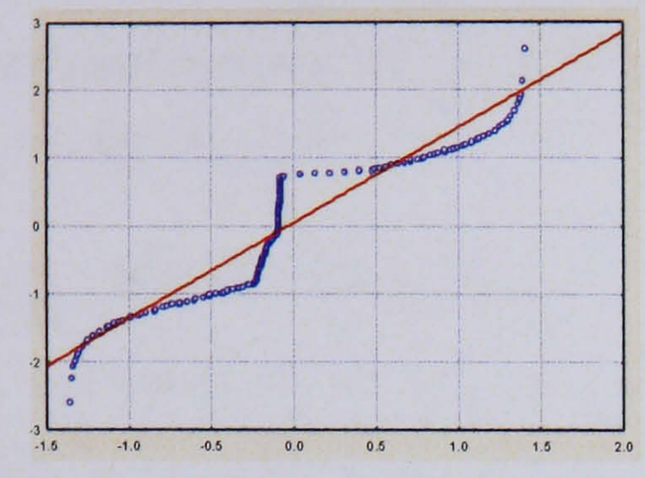
Figure 107: Hotelling's T^2 and SPE plots for nominal data

The control charts generated using SMBMPCA with an ARX error model display similar behaviour to that generated using the other types of error model. There appears to be less structure in the data, indicated by the approximately constant standard deviation witnessed in the control limits in latent variable 2. The control chart for latent variable 1 also demonstrates less structure, although there is still some variation in the first 100 time points, this is confirmed by the Hotelling's T^2 plot in Figure 107 (a), as there are still a number nominal batches out of control initially. However the number of out of control batches has decreased, which indicates that the ARX model has been more successful in removing structure from the data than the previous techniques. Again the SPE plot shows the batches remaining inside the control limits, with occasional spikes outside of the limits, but not for more than three consecutive time points.

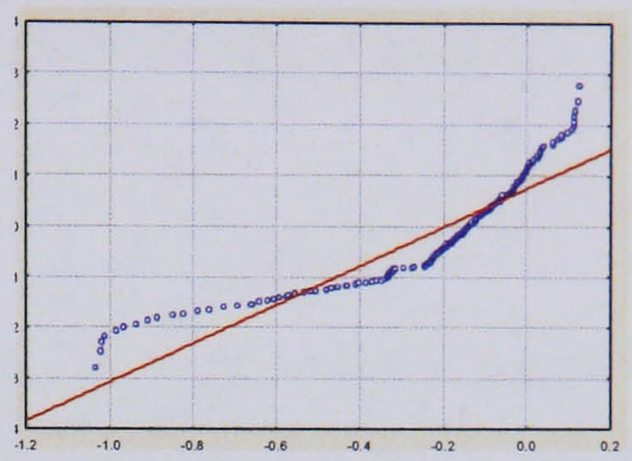
7.3.3.2 Non-linear and Dynamic Behaviour

7.3.3.2.1 Non-linear Behaviour

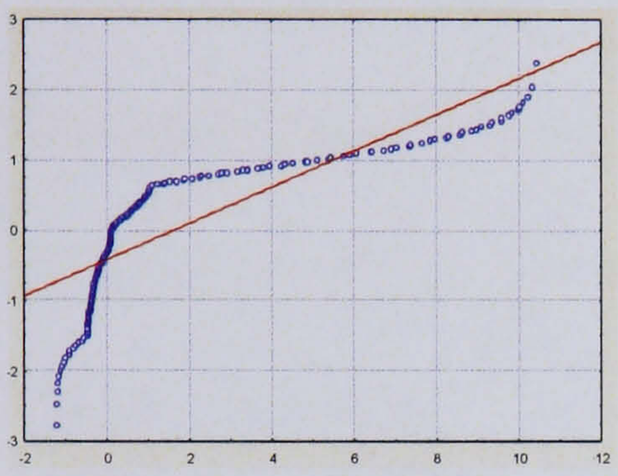
An assessment is made of the model residuals after the application of the ARX error model to see if the technique has been successful in reducing the non-linear structure in the data:



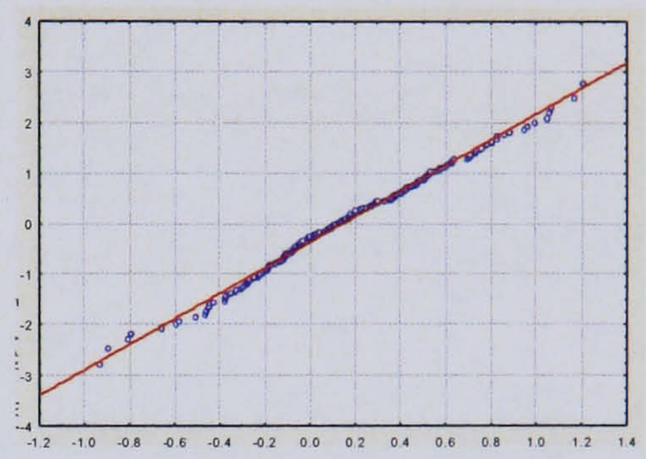
a) Model-based residual 1



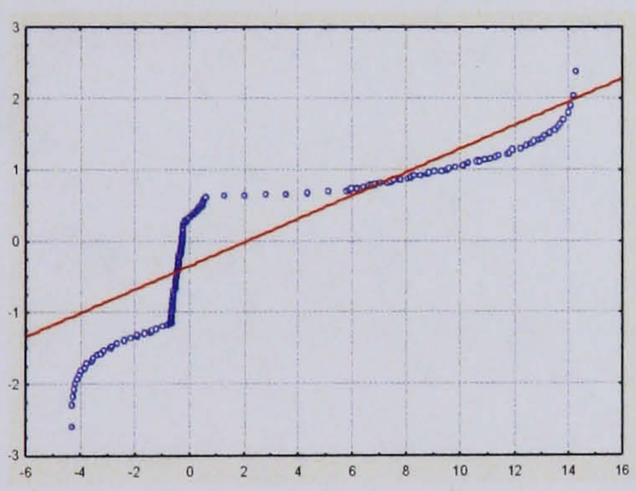
b) Super model-based residual 1



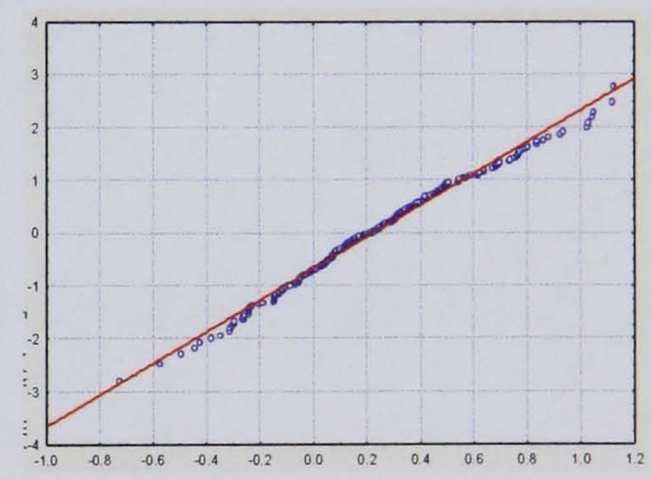
c) Model-based residual 2



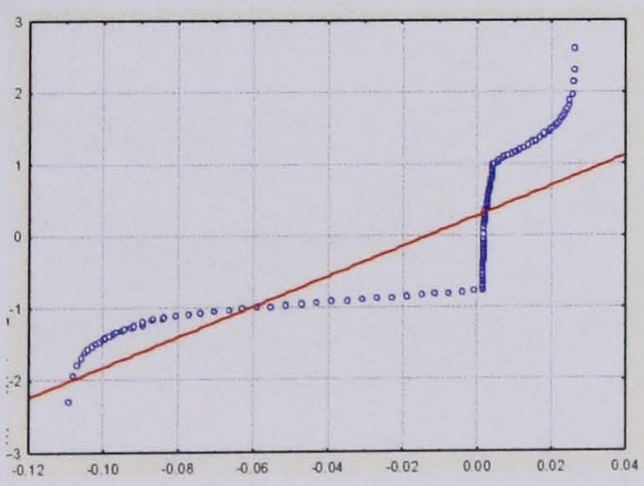
d) Super model-based residual 2



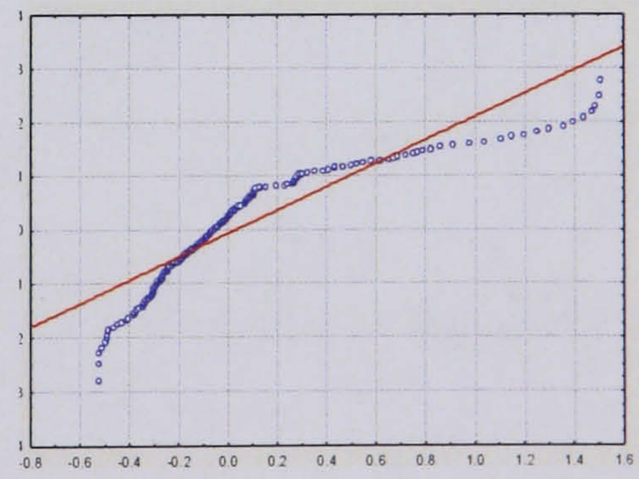
e) Model-based residual 3



f) Super model-based residual 3



g) Model-based residual 4



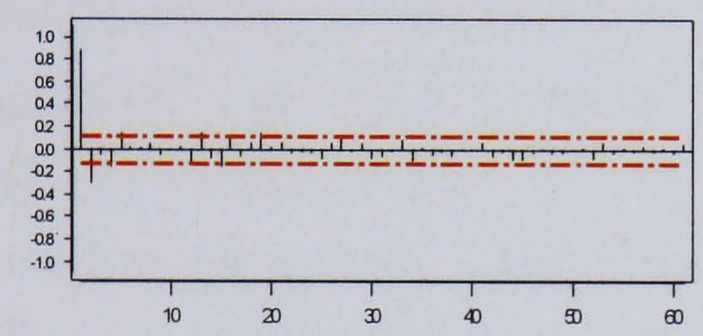
h) Super model-based residual

Figure 108: Normal probability plots for SMBMPCA with ARX error model

Although the ARX time series model was selected specifically to deal with the dynamic structure in the data, it does not address the non-linear behaviour. However from an analysis of the residuals it is evident that the non-linear behaviour in the standard residuals has been significantly reduced, especially in residual variables 2 and 3, the wall temperature and jacket temperature respectively. Residuals 1 and 4, the reactor temperature and cooling valve position, also demonstrate a reduction in the non-linearity in comparison to the model-based residuals, although not as successfully as for the other two residual variables.

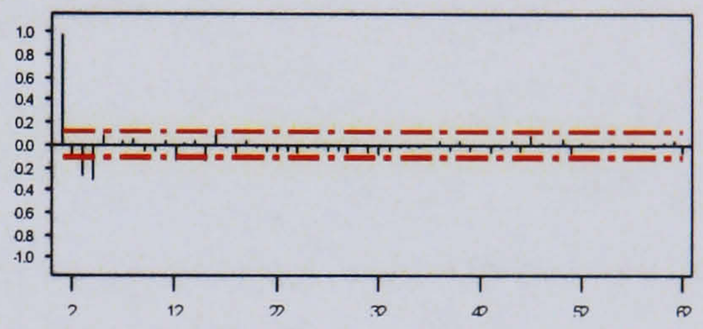
7.3.3.2.2 Dynamic Behaviour

To investigate the ability of the technique to capture the dynamic structure in the residuals, the partial autocorrelation plots were examined, Figure 109.

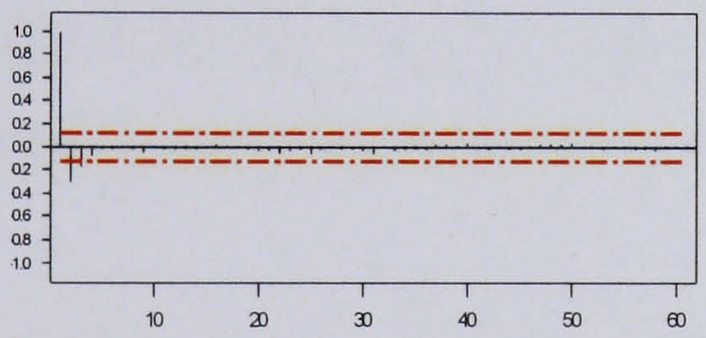


Reactor Temperature

a) Model-based residual 1

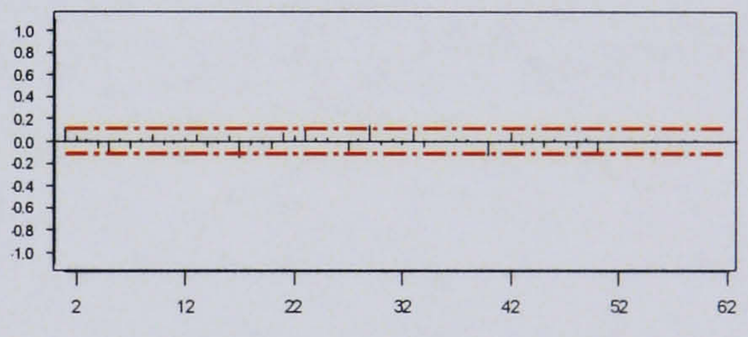


b) Super model-based residual 1

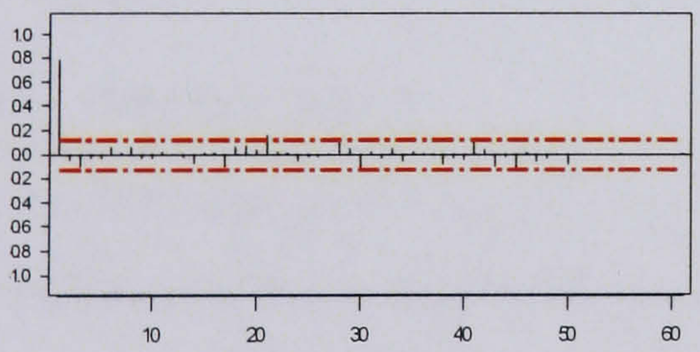


Wall Temperature

c) Model-based residual 2

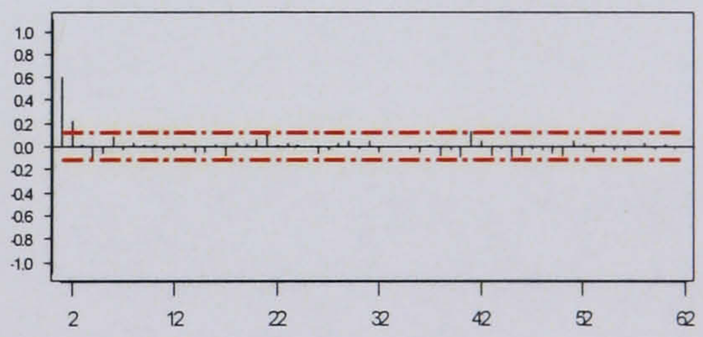


d) Super model-based residual 2

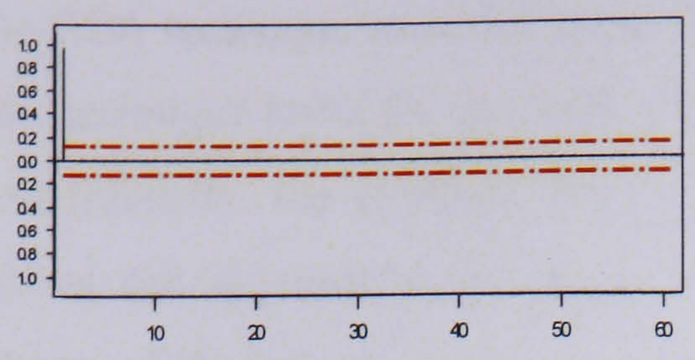


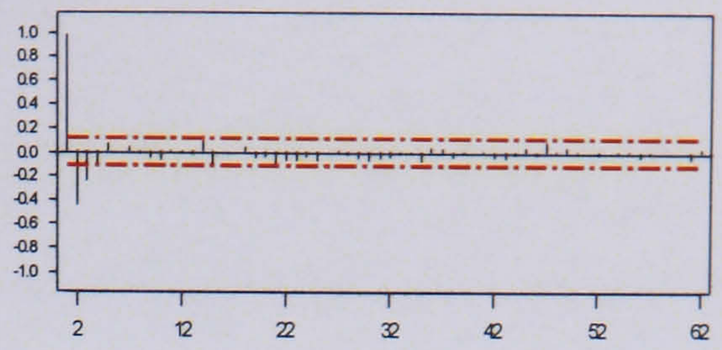
Jacket Temperature

e) Model-based residual 3



f) Super model-based residual 3





Cooling Valve Position

g) Model-based residual 4 h) Super model-based residual 4

Figure 109: Partial autocorrelation plots for MB residuals and SMBMPCA with ARX residuals

Comparison of the plots in Figure 109 with those showing the dynamic structure in the original data in Figure 100 show that the ARX time series model of the standard residuals has helped remove the structure in variables 2 and 3, the wall and jacket temperatures, but shows no improvement with respect to variables 1 and 4, the reactor temperature and valve position. The residuals were modelled using an ARX model and the fact that the structure has only been removed from two of the four residuals indicates that the lags built into the model for each residual may be inaccurate, consequently there is a need to lag each residual variable separately. In comparison to the residuals from the standard model-based technique there is only a significant reduction in the amount of structure in residual 2, the wall temperature, the structure in the other three variables is worse.

7.3.4 Super model-based PCA with dynamic canonical correlation analysis error model

7.3.4.1 Nominal Model

In the ARX approach, a lagged data matrix is created of the model inputs, the original variables, and the model outputs, the first set of model-based residuals, and least squares regression is used to build a prediction model. The disadvantage of this approach is that there are issues in terms of its implementation in the presence of correlated or collinear variables, for instance the covariance matrix is calculated over the entire trajectory of the batch, giving only an average description of the variation occurring between variables. The multi-phase nature of a batch process means that the correlations between the variables are likely to vary over the duration of the batch. By implementing the ARX data structure within the CCA framework, this limitation can be addressed. Application

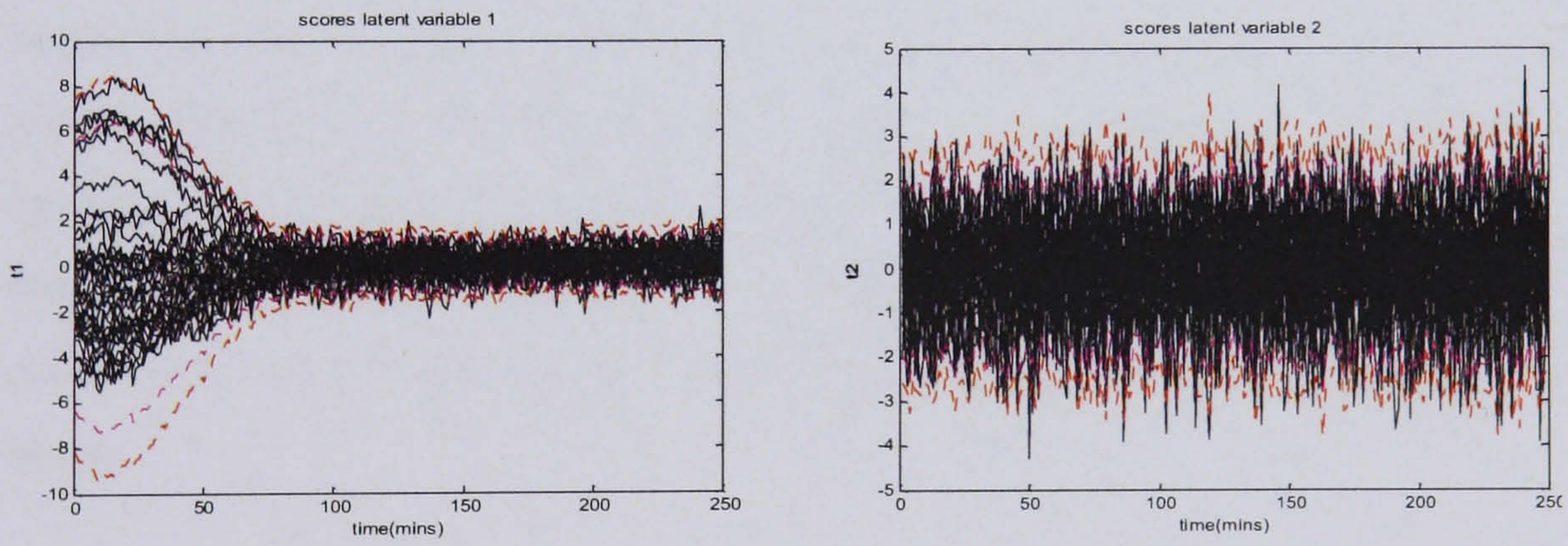
of the CCA technique involves the building a model similar to PLS but more balanced i.e. the technique looks for the correlation between X and Y, rather than regressing Y on X.. Additionally, the dynamic state space model that is created at each time interval describes the dynamic relationships in the data and therefore the non-steady-state behaviour of the batches is captured.

A nominal model was built using the data set of 50 batches, applying batch observation level analysis to the residuals generated after the application of the dynamic CCA technique. The results of the cross validation analysis are shown in Table 11.

No PCs	R^2x	Cum R^2x	Eigen-values	R^2y	Cum R^2y	Q^2	Cum Q^2
1	0.718	0.718	2.87	0.168	0.168	0.168	0.168
2	0.24	0.958	0.959	0.264	0.432	0.318	0.432

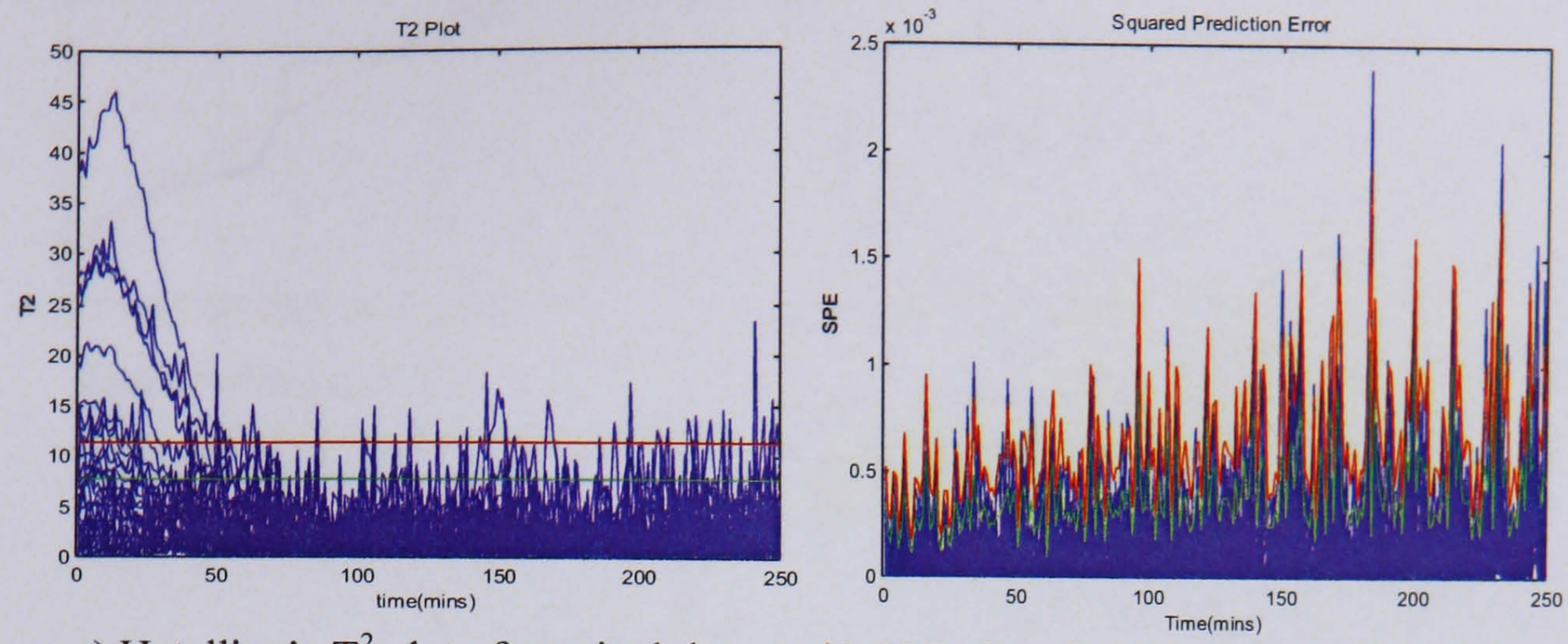
Table 11: Summary of cross validation results for SMBMPCA with Dynamic CCA error model

Two latent variables were retained, capturing 96% of the variation in the data. The nominal model control charts are shown in Figure 110:



a) Latent variable 1 b) Latent variable 2

Figure 110: Scores plots of nominal model for SMBMPCA with Dynamic CCA error model



a) Hotelling's T^2 plot of nominal data b) SPE plot of nominal data

Figure 111: Hotelling's T^2 and SPE plots of nominal data

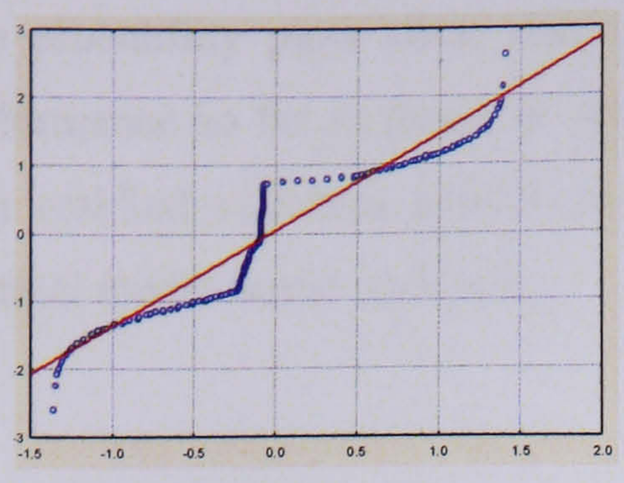
As with the other techniques investigated, the plots of the nominal model indicate that the application of the SMBMPCA with dynamic CCA has removed some of the structure in the data. The nominal data set of 50 batches has remained inside the control limits, apart from the occasional spike outside of the limits due to the noise in the data from overfitting, however they do not remain outside of the limits for three consecutive time points.

Figure 111 (a) and (b) show the Hotelling's T^2 and SPE control charts, nine of the batches lie outside of the control limits initially in the Hotelling's T^2 plot, however they return to the limits within the initial 50 minutes, as opposed to the 100 minutes taken to return the limits with the other techniques examined so far. This is because the error models implementing an ARX structure have to a greater extent captured the amount of structure in the data and therefore reduced the variation in the first 100 minutes of the batch.

7.3.4.2 Non-linear and Dynamic Behaviour

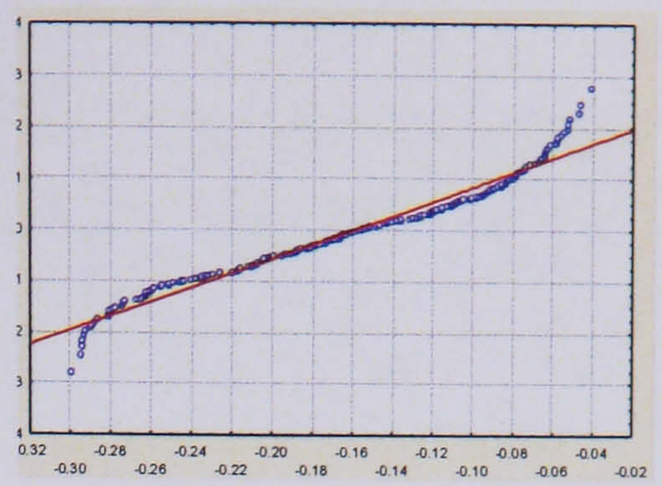
7.3.4.2.1 Non-linear Behaviour

The probability plots were examined to assess if the SMBMPCA technique effectively deals with the non-linear behaviour in the data.

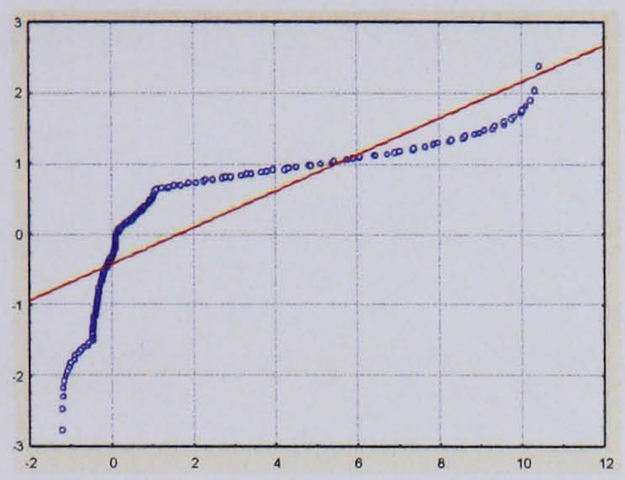


Reactor Temperature

a) Model-based residual 1

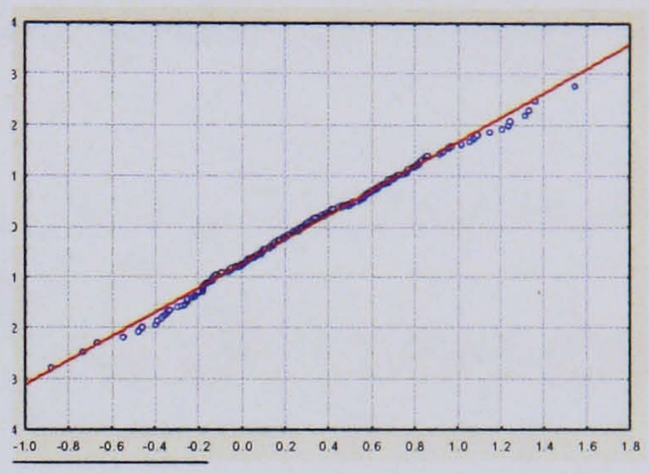


b) Super model-based residual 1

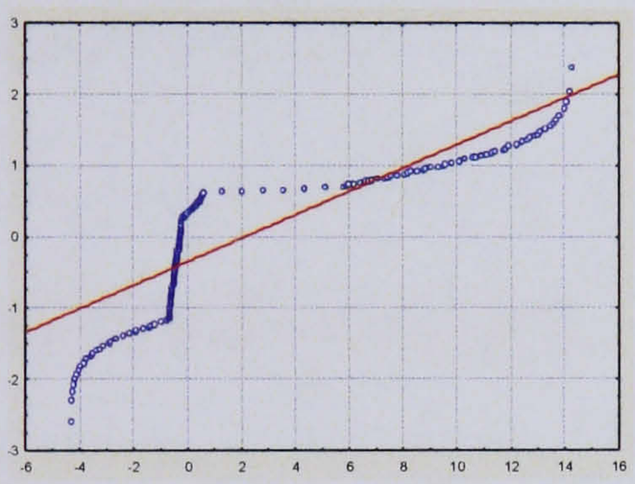


Wall Temperature

c) Model-based residual 2

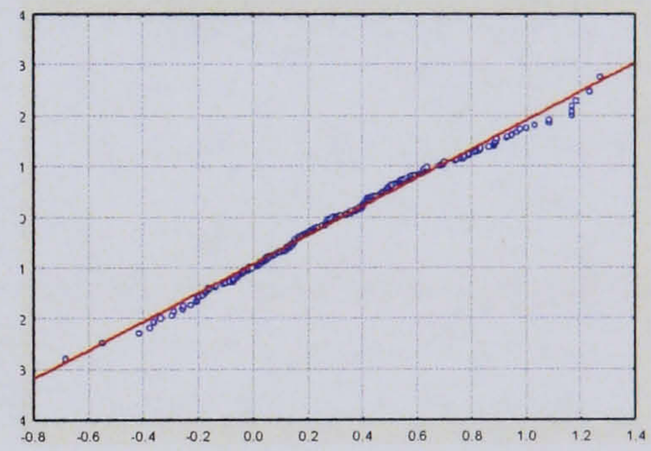


d) Super model-based residual 2

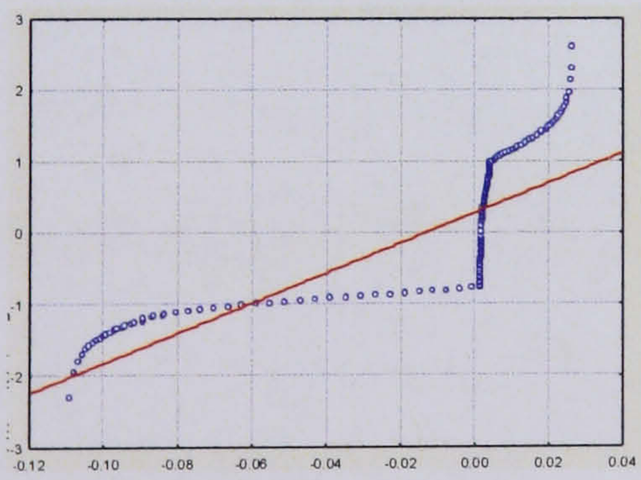


Jacket Temperature

e) Model-based residual 3

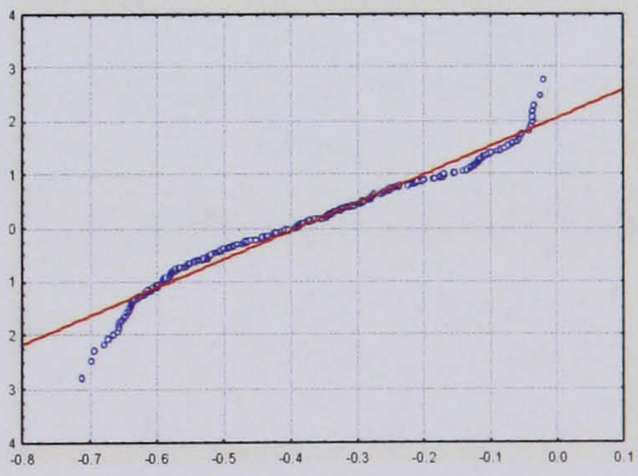


f) Super model-based residual 3



Cooling Valve Position

g) Model-based residual 4



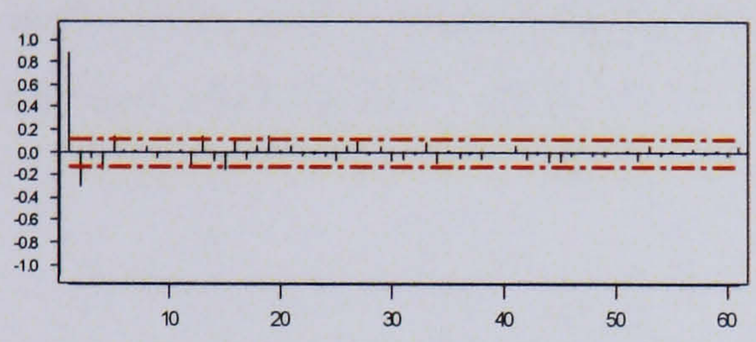
h) Super model-based residual 4

Figure 112: Normal probability plots for MB residuals and SMBMPCA with dynamic CCA residuals

The probability plots show that the dynamic CCA technique demonstrates the best performance so far in terms of dealing with the non-linearity in the residual data. All four residual variables exhibit approximately linear behaviour in comparison to the original model-based residuals.

7.3.4.2.2 Dynamic Behaviour

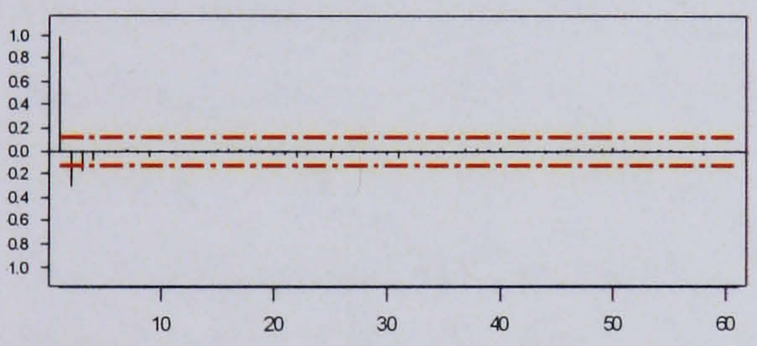
The efficiency of the dynamic CCA technique with respect to removal of the dynamic structure remaining in the residuals is examined through the partial autocorrelation plots:



Reactor Temperature

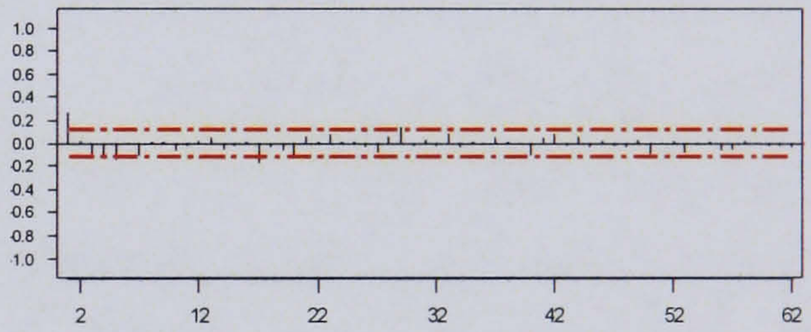
a) Model-based residual 1

b) Super model-based residual 1

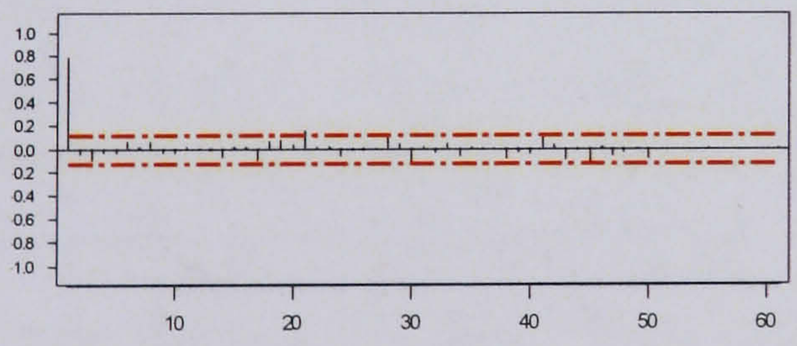


Wall Temperature

c) Model-based residual 2



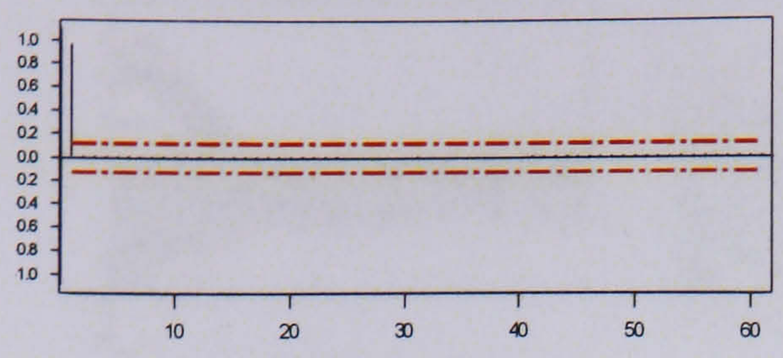
d) Super model-based residual 2



Jacket Temperature

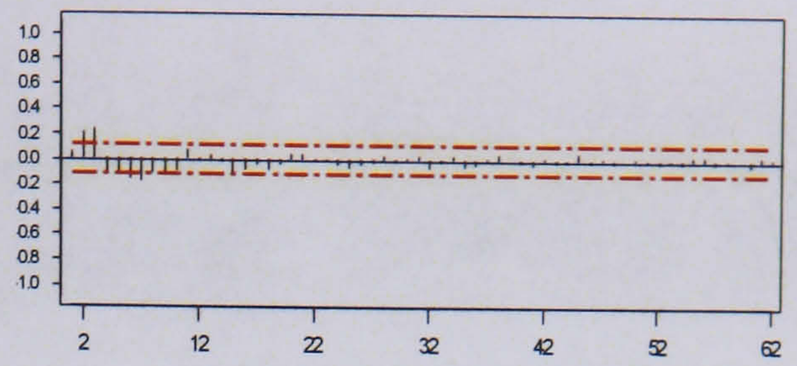
e) Model-based residual 3

f) Super model-based residual 3



Cooling Valve Position

g) Model-based residual 4



h) Super model-based residual 4

Figure 113: Partial autocorrelation plots for MB residuals and SMBMPCA with dynamic CCA residuals

The partial autocorrelation plots show how effective SMBMPCA with dynamic CCA is at removing the dynamic structure in the data. All 4 residuals display negligible dynamic structure after the super model-based technique has been performed.

7.3.5 Super model-based PCA with dynamic non-linear PLS

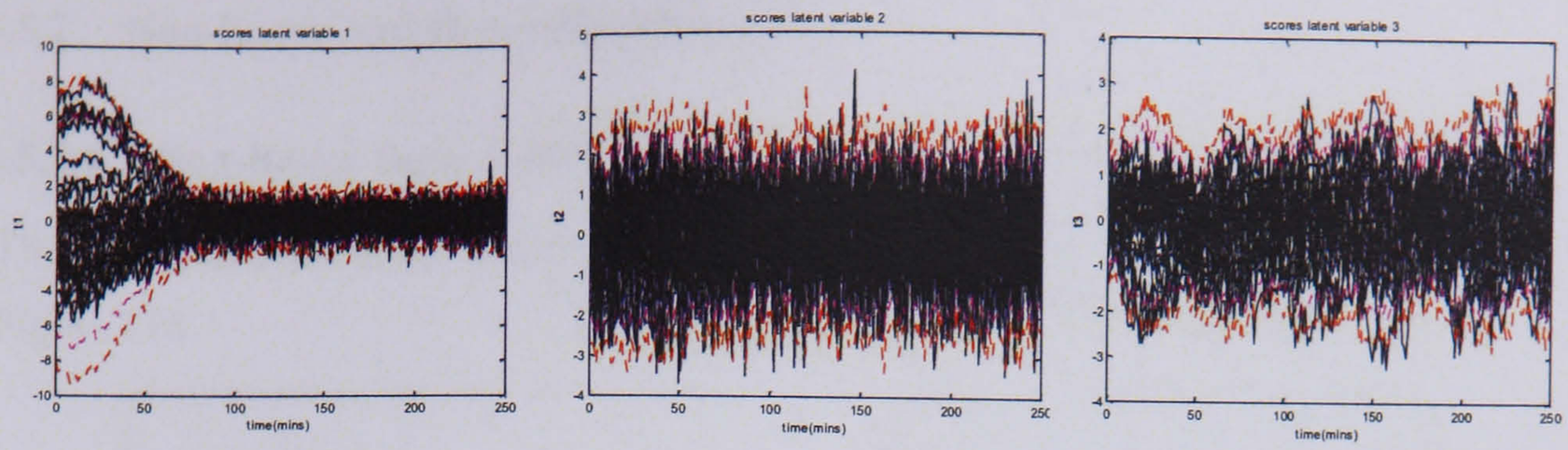
7.3.5.1 Nominal Model

The non-linear PLS technique applied in section 7.3.2 was effective in removing the non-linear behaviour from the residuals but was not as efficient in dealing with the serial correlation. Therefore the SMBPCA technique was investigated using the non-linear PLS technique embedded in the ARX framework to tackle the dynamic structure in the data. The residuals generated after application of the dynamic non-linear PLS error model were used to build a nominal batch observation level model.

No PCs	R^2x	Cum R^2x	Eigen-values	R^2y	Cum R^2y	Q^2	Cum Q^2
1	0.529	0.529	2.11	0.0139	0.0139	0.0138	0.0138
2	0.25	0.778	0.999	0.00175	0.00157	0.00159	0.0153
3	0.183	0.961	0.73	0.00198	0.00176	0.00197	0.0173

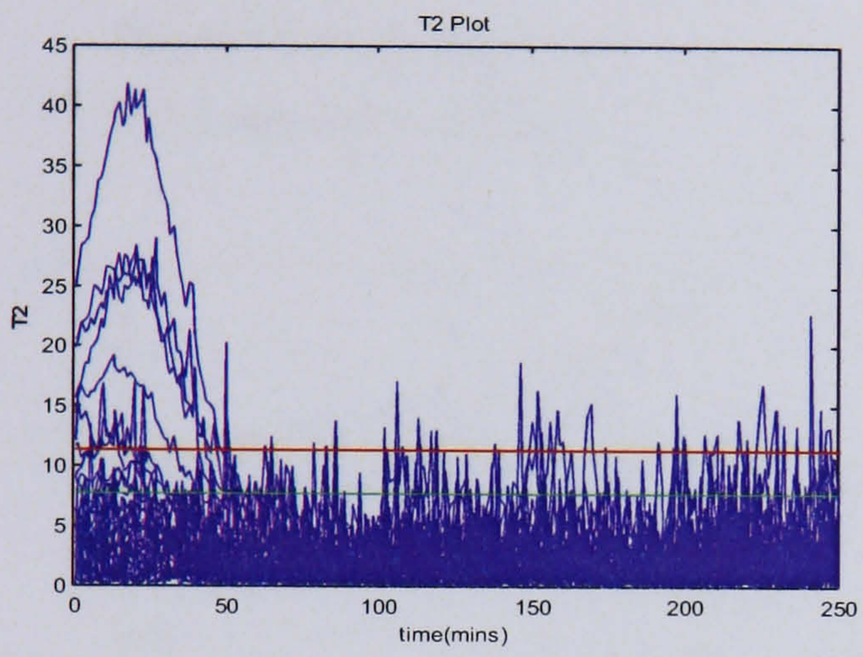
Table 12: Summary of cross validation results for SMBMPCA with Dynamic Non-linear PLS error model

Three latent variables were retained for the nominal model, capturing 96% of the variation in the data. The control charts for the nominal model are shown in Figure 114.

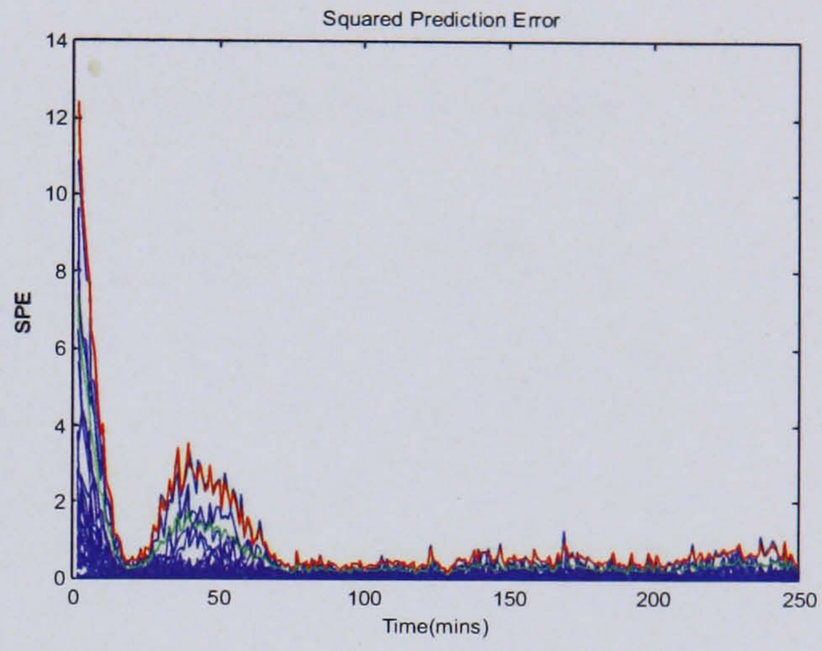


a) Latent variable 1 b) Latent variable 2 c) Latent variable 3

Figure 114: Scores plots for nominal model of SMBMPCA with dynamic non-linear PCA



a) Hotelling's T^2 plot of nominal data



b) SPE plot of nominal data

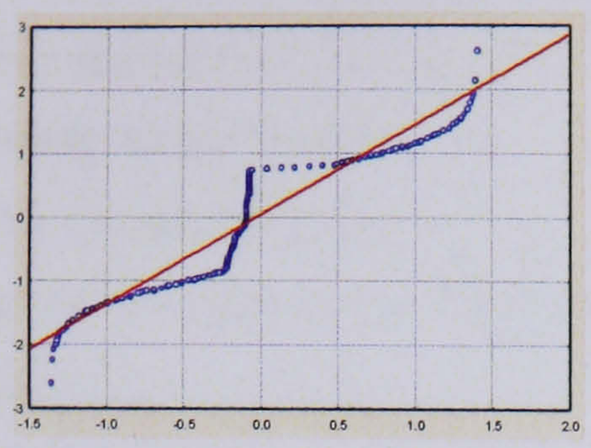
Figure 115: Hotelling's T^2 and SPE plots of nominal data

The scores control charts in Figure 114 follow a similar pattern to those generated by the previous SMBMPCA techniques. The dynamic structure in the data appears to have been significantly reduced in comparison to the model-based PCA control charts seen in Chapter 6, although the noise in the residuals appears to have increased, particularly in the latent variable 2 control chart. The Hotelling's T^2 control chart in Figure 115 (a) also displays noisy behaviour, with eight batches lying outside of the control limits initially, although the majority of the batches remain inside the control limits after the initial 100 time points, the first phase of operation. In Figure 115 (b), the SPE control chart of the nominal data shows all the batches remaining inside the control limits, again it is noted that the behaviour is more dynamic during the first 100 minutes of the batch.

7.3.5.2 Non-linear and Dynamic Behaviour

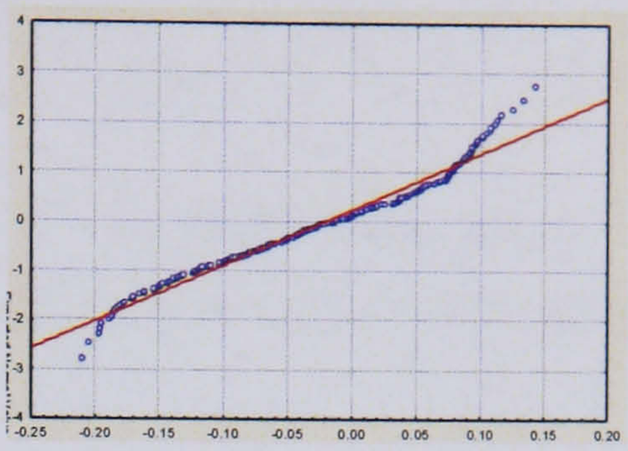
7.3.5.2.1 Non-linear Behaviour

The probability plots for the SMBPCA with dynamic non-linear PLS are illustrated in Figure 116.

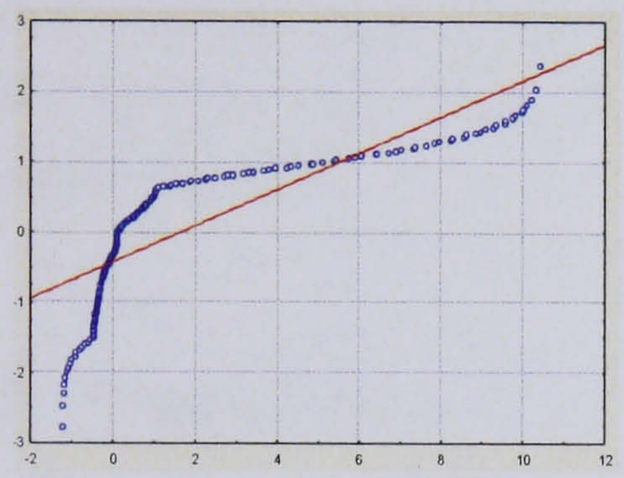


Reactor Temperature

a) Model-based residual 1

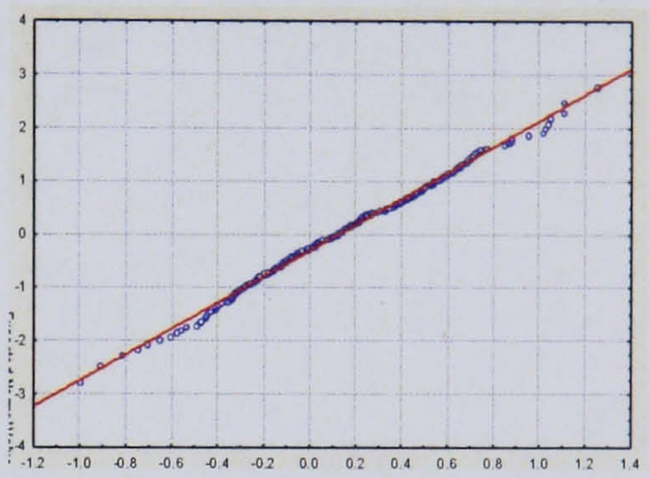


b) Super model-based residual 1

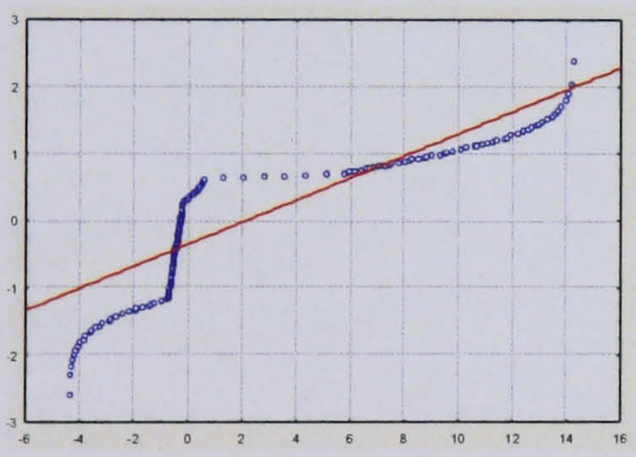


Wall Temperature

c) Model-based residual 2

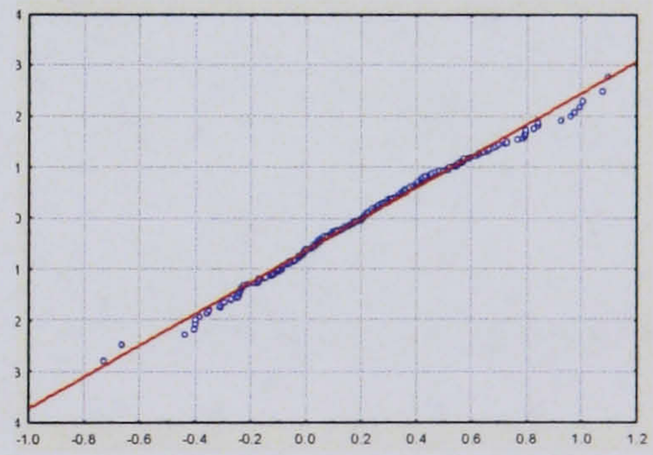


d) Super model-based residual 2

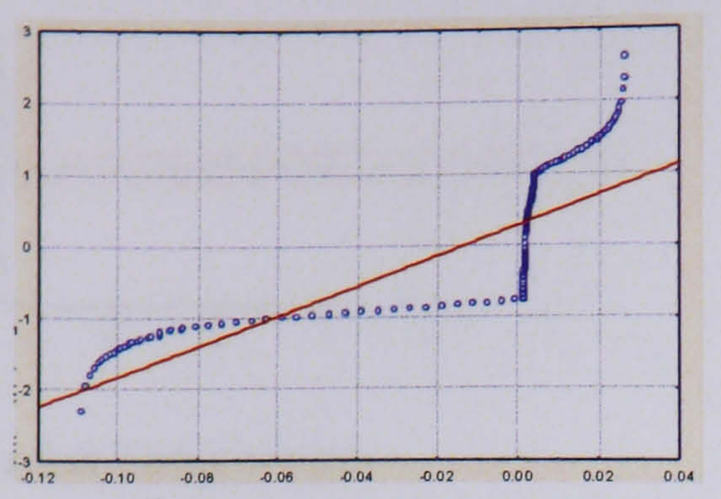


Jacket Temperature

e) Model-based residual 3

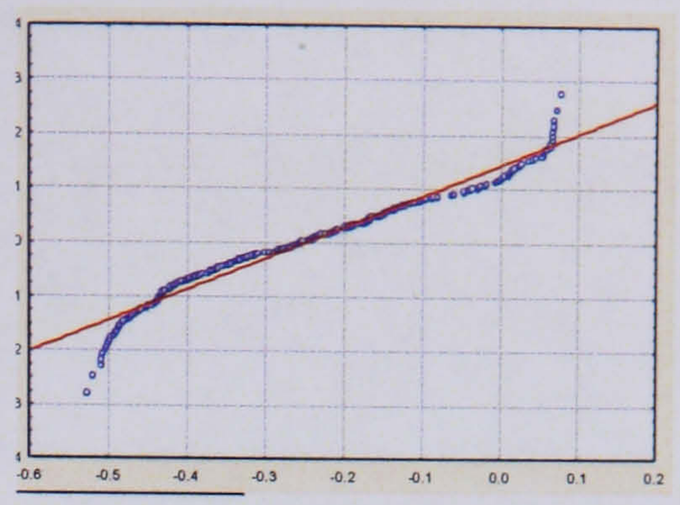


f) Super model-based residual 3



Cooling Valve Position

g) Model-based residual 4

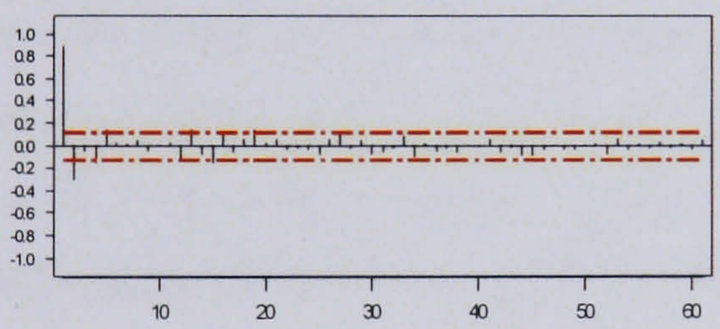


h) Super model-based residual 4

Figure 116: Probability plots for MB residuals and SMBMPCA with dynamic non-linear PLS residuals

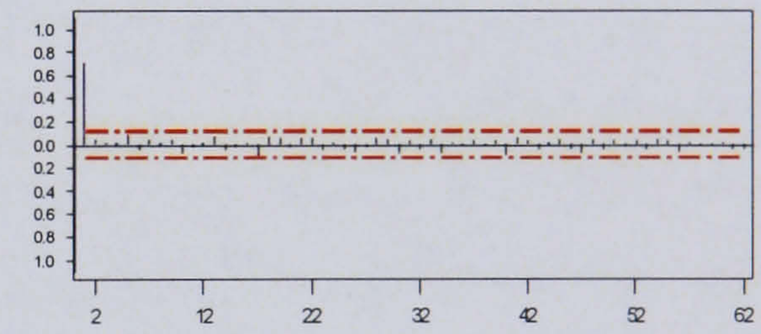
The probability plots of the super model-based residuals generated using dynamic non-linear PLS exhibit similar results to those achieved with the dynamic CCA model. Residual variables 2 and 3, the wall and jacket temperatures, are more linear than residual variables 1 and 4, the reactor temperature and valve position, however all 4 demonstrate linearity.

7.3.5.2.2 Dynamic Behaviour

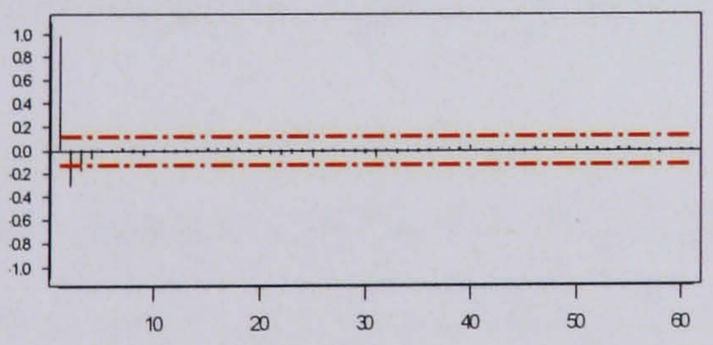


Reactor Temperature

a) Model-based residual 1

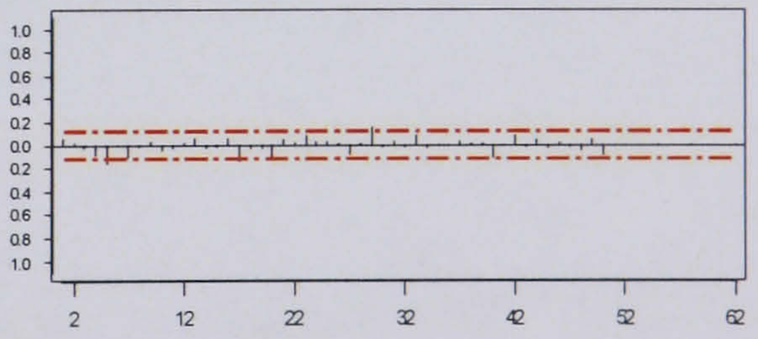


b) Super model-based residual 1

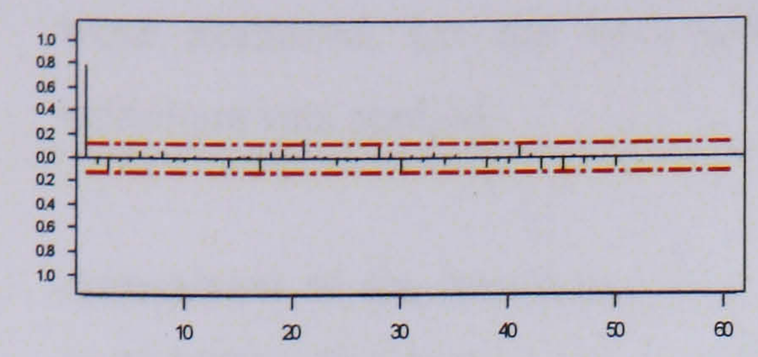


Wall Temperature

c) Model-based residual 2



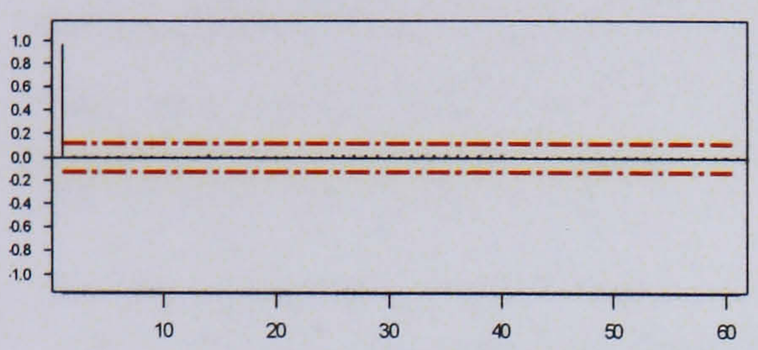
d) Super model-based residual 2



Jacket Temperature

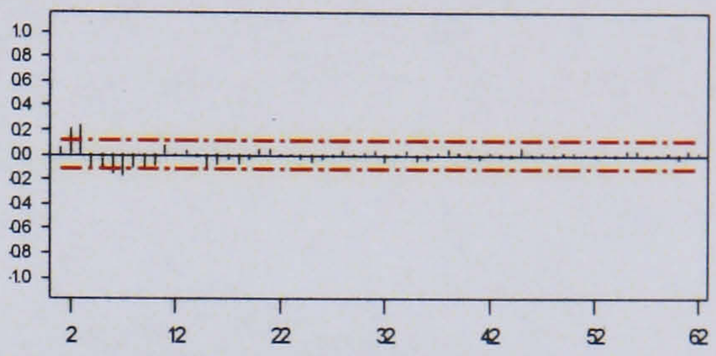
e) Model-based residual 3

f) Super model-based residual 3



Cooling Valve Position

g) Model-based residual 4



h) Super model-based residual 4

Figure 117: Partial autocorrelation plots for MB and SMBMPCA with dynamic non-linear PLS residuals

SMBMPCA with dynamic non-linear PLS completely removes the dynamic structure in residual variables 2 and 3, wall temperature and jacket temperature, and reduces the dynamic behaviour in residual variables 1 and 4, the reactor temperature and valve position, significantly in comparison to the original batch data, although the results are not as good as those attained using dynamic CCA. A better model fit may have been achieved using the non-linear PLS algorithm with a neural net model or radial basis function, as suggested by Baffi *et al* (2000). This would have further reduced the dynamic structure in the data. However there is also the possibility of over-fitting the model through the application of this technique, which can be reduced through the use of cross-validation and testing the model with validation data sets.

7.4 Summary of Residual Analysis Results

The super model-based PCA technique was developed to deal with two of the main characteristics present in batch data, non-linear behaviour and dynamic structure. The basis of the technique is to remove these properties from the batch data, allowing a linear multivariate statistical technique such as batch observation level PCA to be successfully applied to the resulting residuals. To assess if the SMBMPCA techniques investigated achieved these objectives, the residuals generated from the application of the error model

were examined, i.e. the unstructured residuals before the batch observation level technique was applied.

Assessment of the non-linear structure in the residuals was made initially via normal probability plots of the data. It is difficult to summarise the visual assessment of the probability plots generated for each SMBMPCA technique, therefore the values of the Shapiro-Wilks test statistic, Shapiro *et al* (1965), calculated at the same time, are summarised instead. The Shapiro-Wilks test calculates a W statistic that tests whether a data set is normally distributed. A smaller value of the W statistic indicates that there has been a departure from normality.

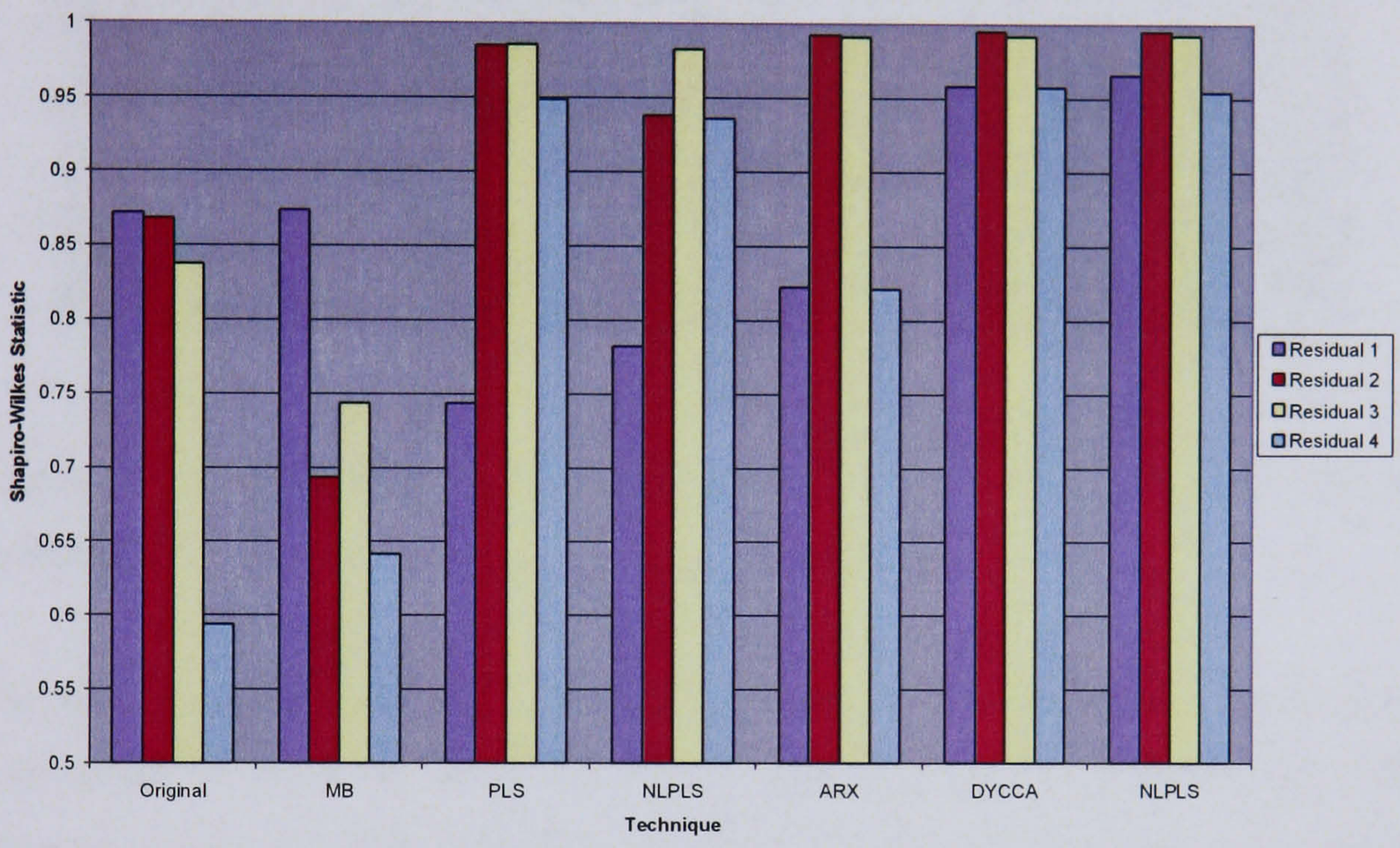


Figure 118: Summary of Shapiro-Wilks test for normality on model-based residuals

Analysing the results of the Shapiro-Wilks test in conjunction with the normal probability plots throughout the chapter, it can be seen that all five of the super model-based techniques demonstrate an improvement in the removal of the non-linear behaviour in comparison to the original and model-based techniques. In Figure 118 the dynamic CCA error model and the dynamic non-linear PLS model give the best results with respect to the removal of the non-linear structure, the probability plots generated previously for each model type support this theory.

An assessment was made of the dynamic structure in the residuals using partial autocorrelation plots to study the process order in the data, the objective being to reduce the process order, which is basically the strength of the correlation between consecutive observations. As with the non-linear behaviour, this objective was investigated on the

residuals generated from the error model i.e. before the batch observation level PCA technique was applied. A summary of the process orders for each technique is shown in Figure 119.

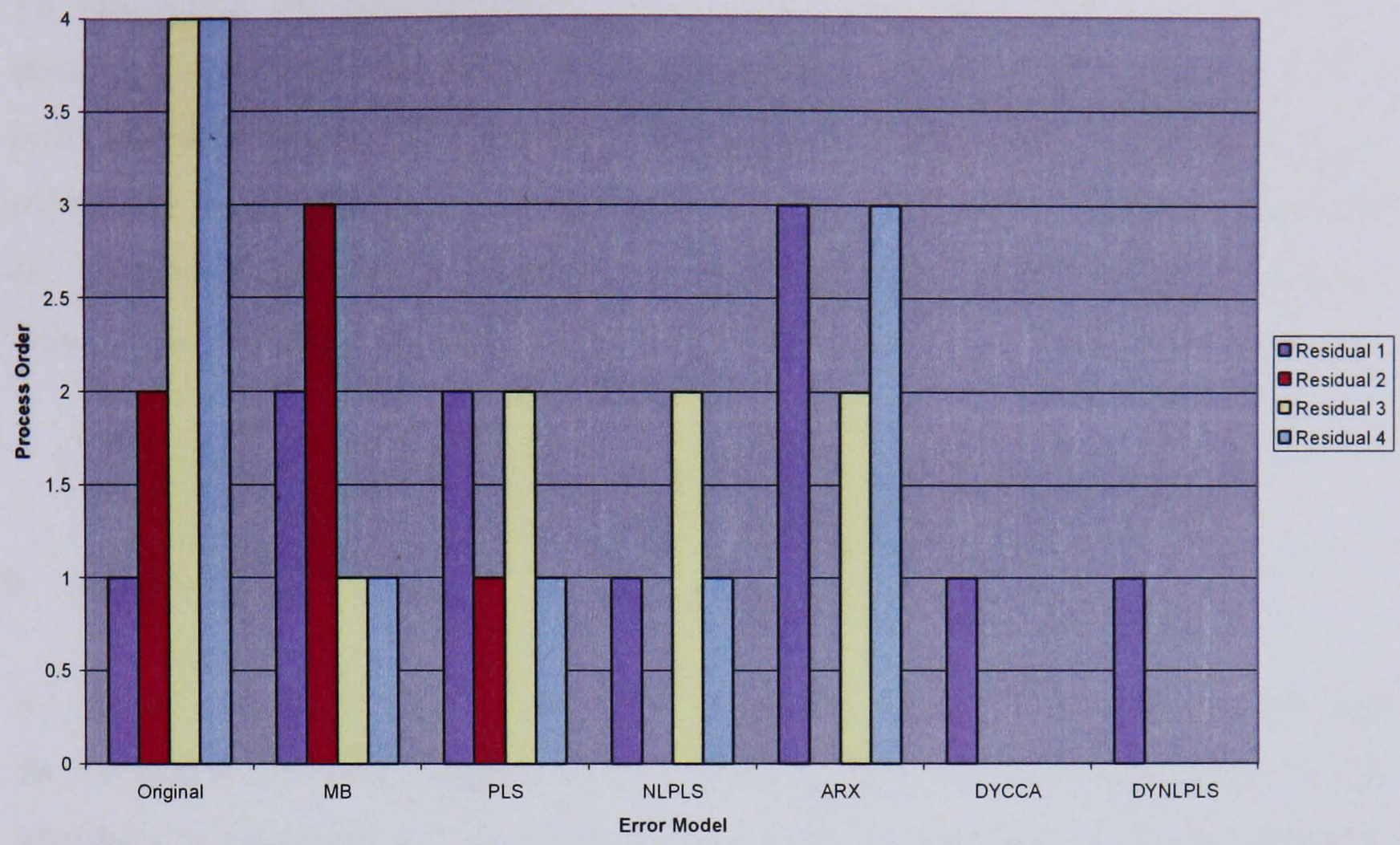


Figure 119: Summary of process order reduction for residuals on each SMBMPCA technique

The summary of the process orders for each residual variable shows that SMBMPCA with dynamic CCA and SMBMPCA with dynamic non-linear PLS are the most efficient in removing the dynamic structure from the data as the structure has been effectively removed from three of the four residual variables. The process order of residual variable 1, originally the reactor temperature, has remained the same. This is possibly due to poor modelling of the variable, although it is also interesting to note that this is the variable that was most significant in the previous contribution analyses. The effectiveness of the dynamic CCA and dynamic non-linear PLS techniques is not surprising as both model types incorporate an ARX framework into the error model to account for the serial correlation in the data. The model-based technique has increased the process order in some of the residual variables, this is due to the mismatch between the plant and the model. While the ARX model would be expected to perform well, an increase in the process order for two out of the four variables is observed. This is because, although the ARX model takes into account the dynamic nature of the data, it does not always work well in the presence of correlated and collinear data. Additionally, autocorrelations between observations are interdependent, therefore there is the possibility that removing

some of the dynamic structure from the process through the application of the ARX framework the other autocorrelations will be affected, compounding their effect as opposed to reducing it. This is potentially what has happened in this case.

To summarise, the best results in terms of removing the non-linear and dynamic structure in the data, are achieved through the use of the super model-based PCA technique with dynamic CCA error models or dynamic non-linear PLS models. In the subsequent section, the fault detection abilities of all the super model-based techniques are examined, to see if the removal of the structure from the residuals has a positive effect of the fault detection results.

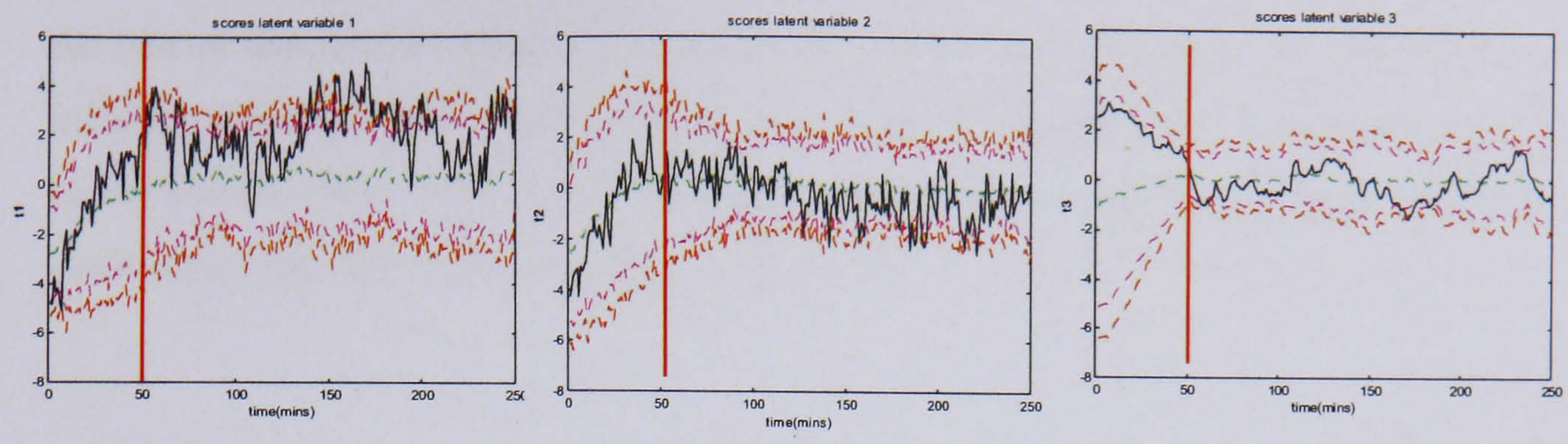
7.5 Fault Detection

As well as examining the residuals from the various SMBMPCA techniques, the fault detection ability of the method was also tested using the three sets of faulty batches introduced in Chapter 6. Following the format of the case study carried out in Chapter 6, the 50 batches for each fault type were projected onto the nominal model for each of the super model-based techniques investigated. The average run length (the average time between a fault occurring and the fault being detected by the control chart), was calculated over the 50 batches and averaged for each technique. The results are discussed in the following sections.

7.5.1 Super model-based PCA with PLS error model

The observation level scores plots for SMBPCA with PLS error model are shown in Figure 120. Due to the noisy residuals, only one batch for each fault type is shown on the scores control charts as otherwise it is difficult to identify what is happening. The same batch was used for each of the error models. However, all fifty batches are used in the calculation of the average run lengths.

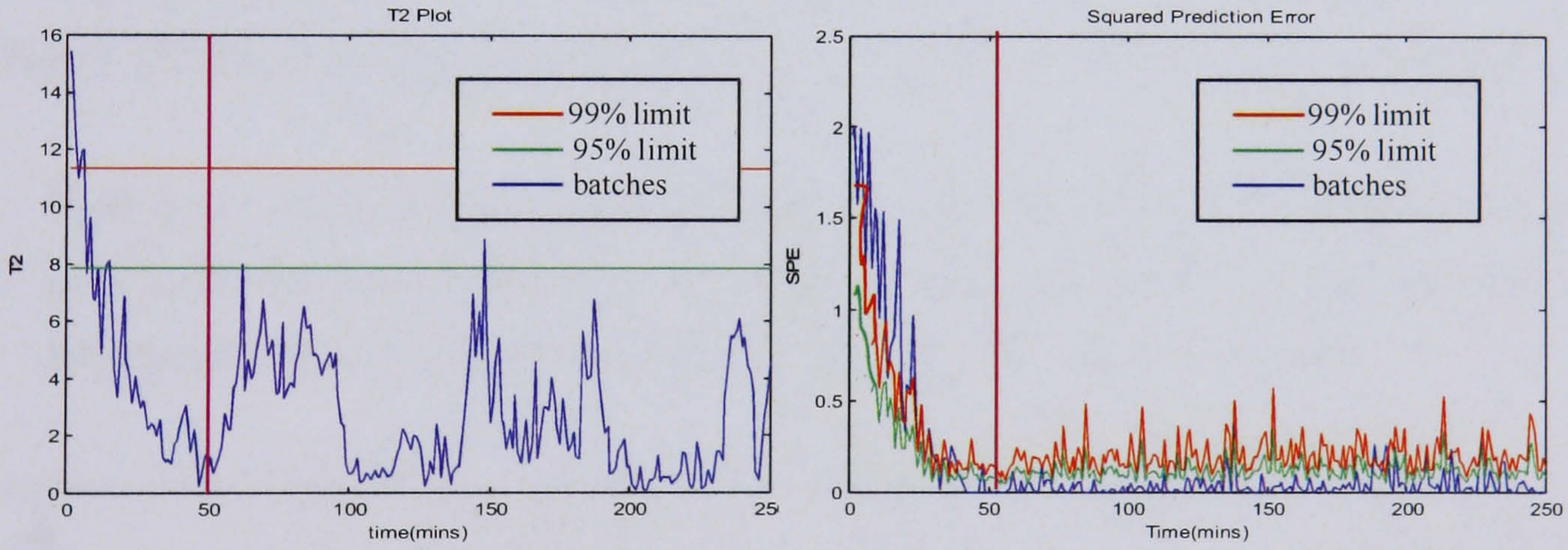
7.5.1.1 Fault Type One



a) Latent variable 1 b) Latent variable 2 c) Latent variable 3

Figure 120: Batch observation level control charts for fault type 1

The point of introduction of the fault is indicated by the red line on the control charts. Figure 120(a) shows there is a change in the behaviour of the batch almost as soon as the fault is introduced, the batch trajectory is initially moving upwards, after the fault is introduced, the batch starts to move downwards. However, the trajectory does not actually move outside the control limits until approximately 100 minutes after the fault has occurred. In the scores control charts for latent variables 2 and 3, a change is also observed in the batch behaviour after the fault has been introduced. Figure 121 shows the Hotelling's T^2 and SPE plots for the same batch.



a) Hotelling's T^2 plot for fault type 1 b) SPE plot for fault type 1

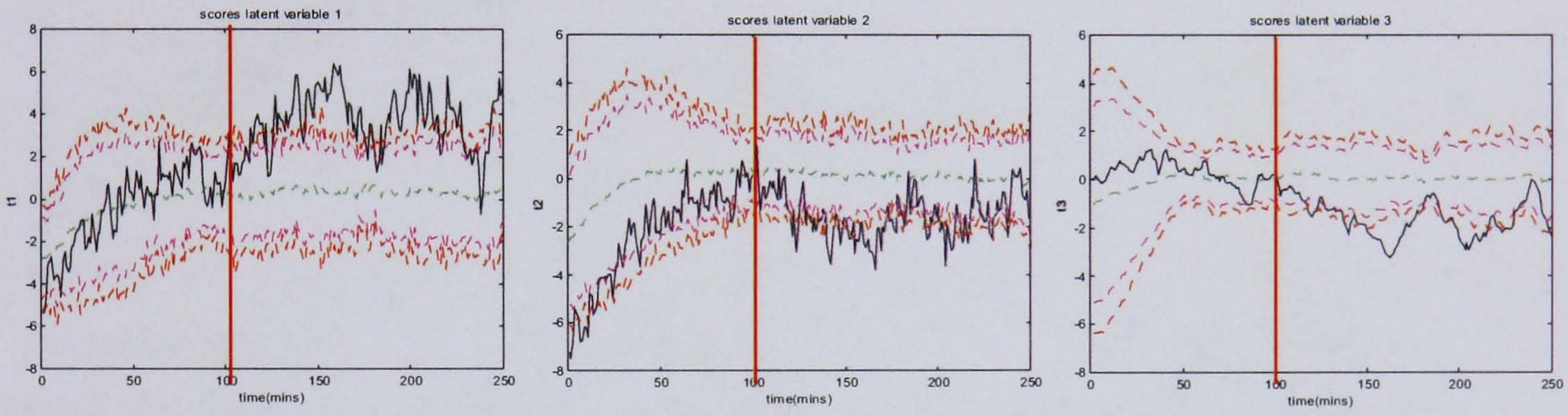
Figure 121: T^2 and SPE plots for fault type 1

Figure 121 (a), the Hotelling's T^2 control chart, shows the batch trajectory moving towards the control limits after fault type one, the temperature sensor fault, is introduced. The batch exceeds the 95% limit briefly at around time point 60 and then again at around time point 140. The SPE chart in Figure 121 (b) shows the batch being initially out of control, however, after the fault is introduced at time point 50 there are no further deviations from the SPE control limits. This indicates that there has not been a

significant change in the correlation structure between the variables, but instead indicates some kind of manipulation of the process parameters. This is consistent with the nature of the fault and the simulation used to generate it, because in the simulation the reactor temperature dictates the behaviour of the other variables. A temperature fault would not necessarily change the correlation structure between the variables but would cause them to move away from their normal operating trajectories, and so should be more easily detected by the Hotelling's T^2 metric.

7.5.1.2 Fault Type Two

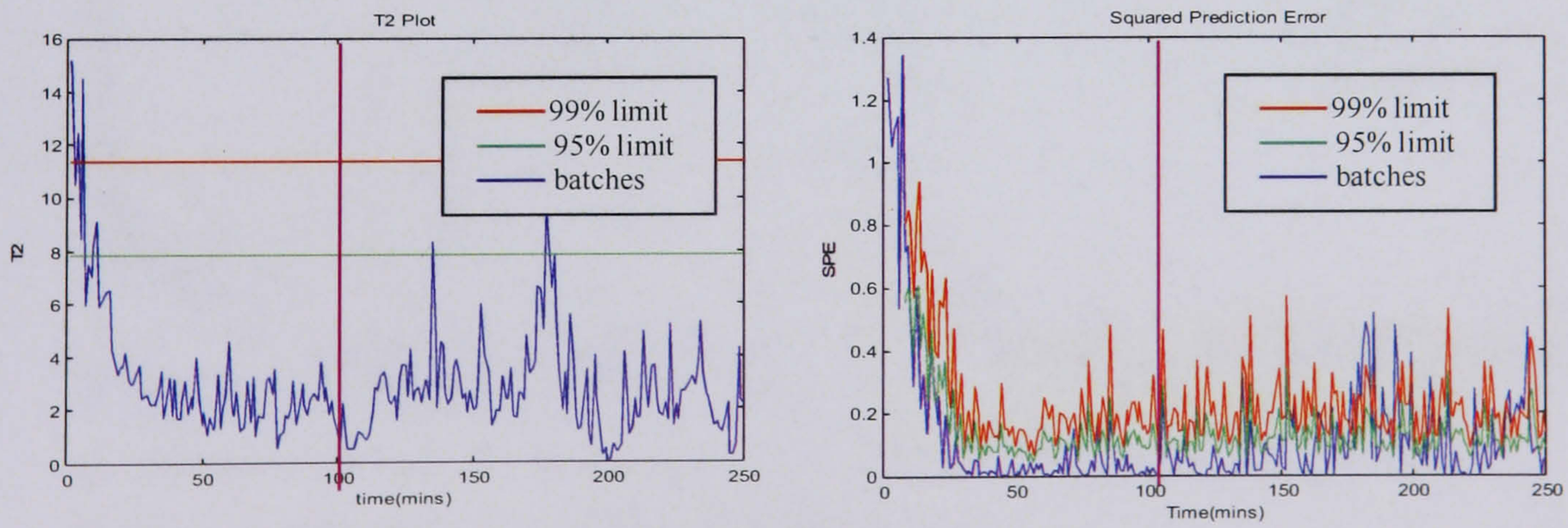
The scores control charts for the second fault type, the decrease in heat transfer coefficient are shown in Figure 122.



a) Latent variable 1 b) Latent variable 2 c) Latent variable 3

Figure 122: Batch observation level control charts for SMBMPCA with PLS – fault type 2

Fault type 2 is detectable by the SMBMPCA with PLS. The batch trajectory moves outside of the control limits within approximately 30 minutes of the fault being introduced. Figure 123 shows the Hotelling's T^2 and SPE plots for this fault.



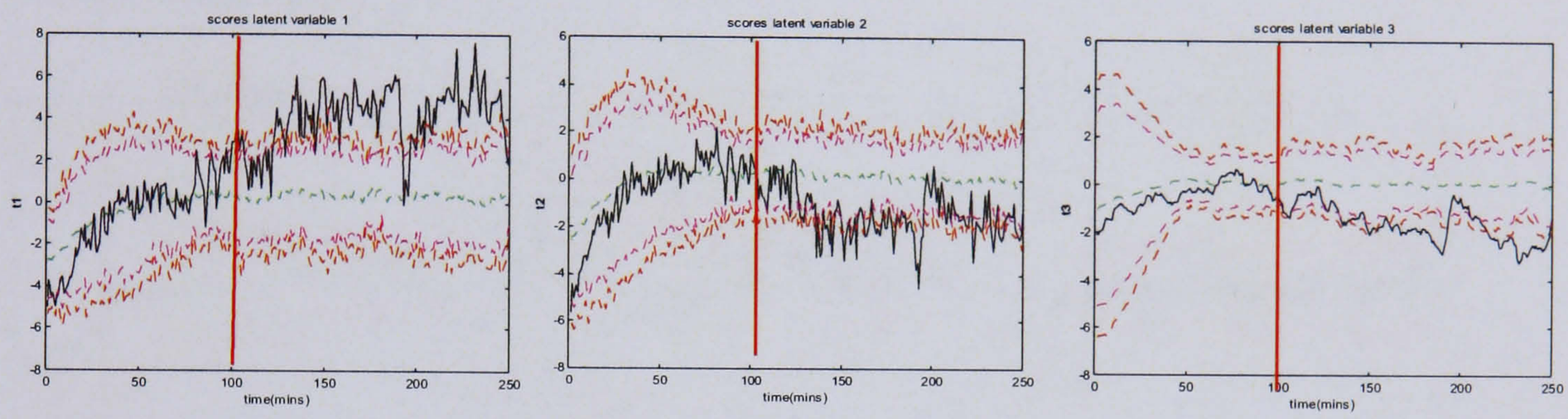
a) Hotelling's T^2 plot of fault type 2 b) SPE plot of fault type 2

Figure 123: Hotelling's T^2 and SPE plots of fault type 2

The plot of the Hotelling's T^2 statistic shows the batch trajectory breeching the 95% limit on 2 occasions, but not the 99% limit, and not for long enough for the excursion to be considered a fault. The SPE plots shows the batch moving outside of the control limits approximately 100 time points after the fault has occurred, although the noise present in the residuals makes this difficult to identify. The noise could be due to overfitting of the model.

7.5.1.3 Fault Type Three

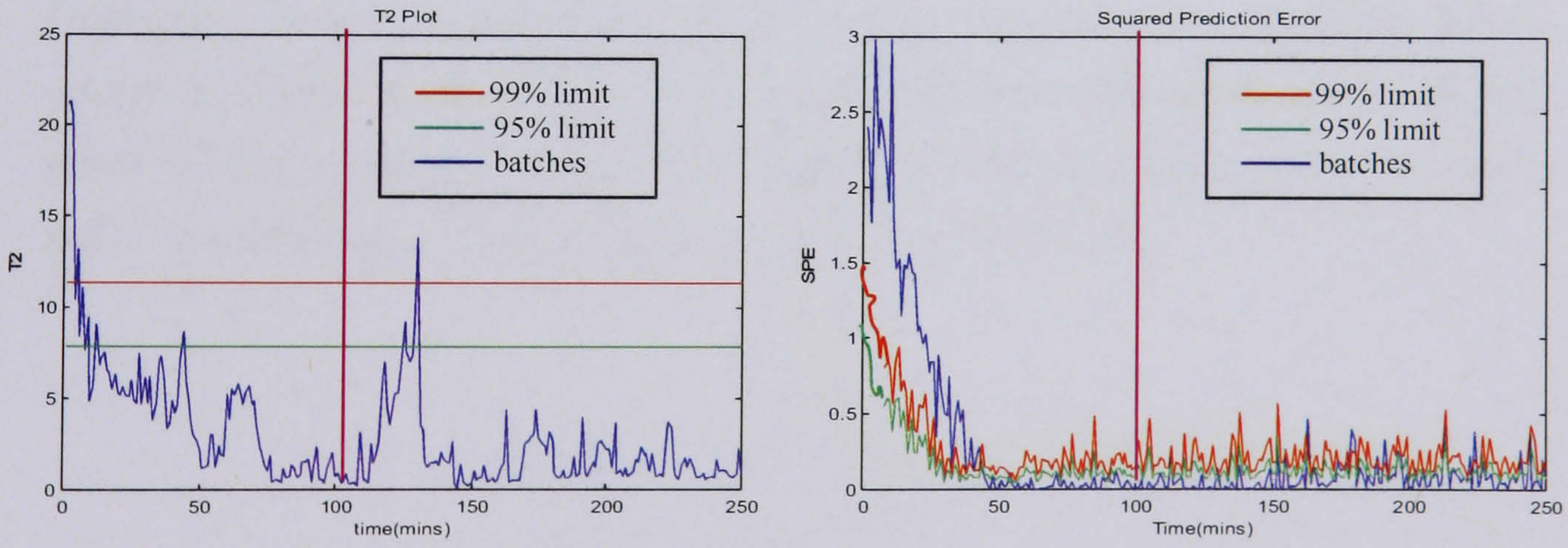
The control charts for the third fault type, the cooling valve fault, are shown in Figure 124:



a) Latent variable 1 b) Latent variable 2 c) Latent variable 3

Figure 124: Batch observation level control charts for SMBMPCA with PLS – fault type 3

The third fault type can be clearly seen moving outside the control limits within 40-50 minutes of the fault occurring for all three latent variables. The Hotelling's T^2 and SPE plots for the batch are shown in Figure 125.



a) Hotelling's T^2 plot for fault type 3 b) SPE plot for fault type 3

Figure 125: Hotelling's T^2 and SPE plots for fault type 3

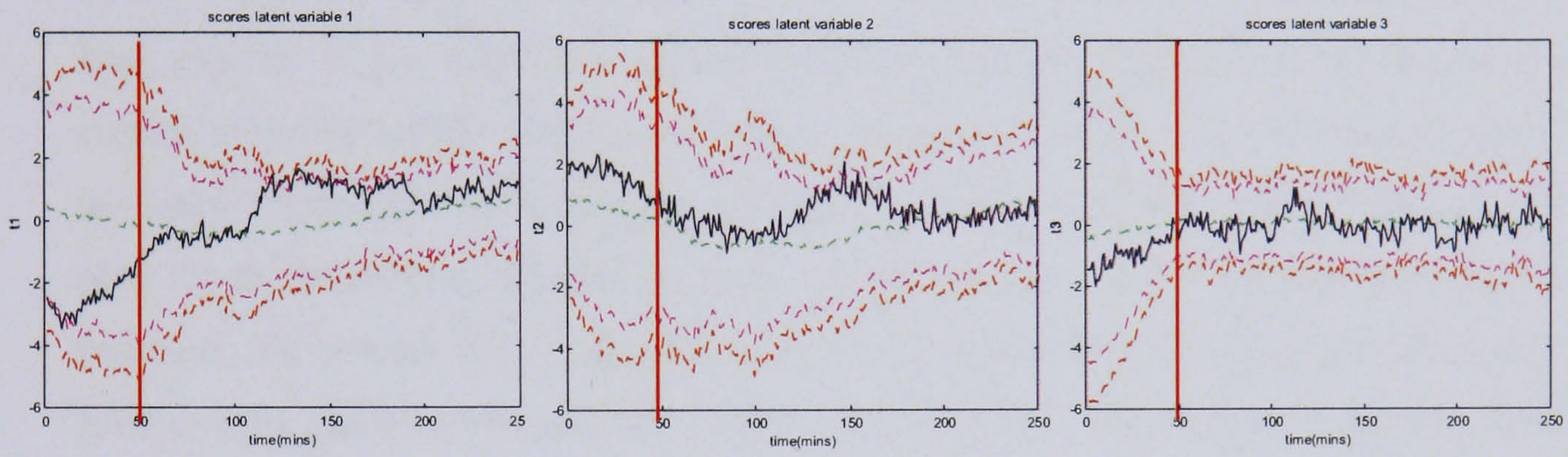
Figure 125 (a) shows the faulty trajectory deviating from the Hotelling's T^2 limits at around 125 minutes before re-entering them again shortly afterwards. This behaviour

was observed when fault three was investigated in Chapter 2 with the standard batch monitoring techniques. In the SPE plot, the batch briefly deviates from the control limits at around time point 160, but does not remain outside of the limits for long enough to be classified as a fault.

7.5.2 Super model-based PCA with non-linear PLS error model

The second type of error model used in conjunction with the super model-based technique was non-linear partial least squares. The control charts for each of the fault types are shown in the following sections.

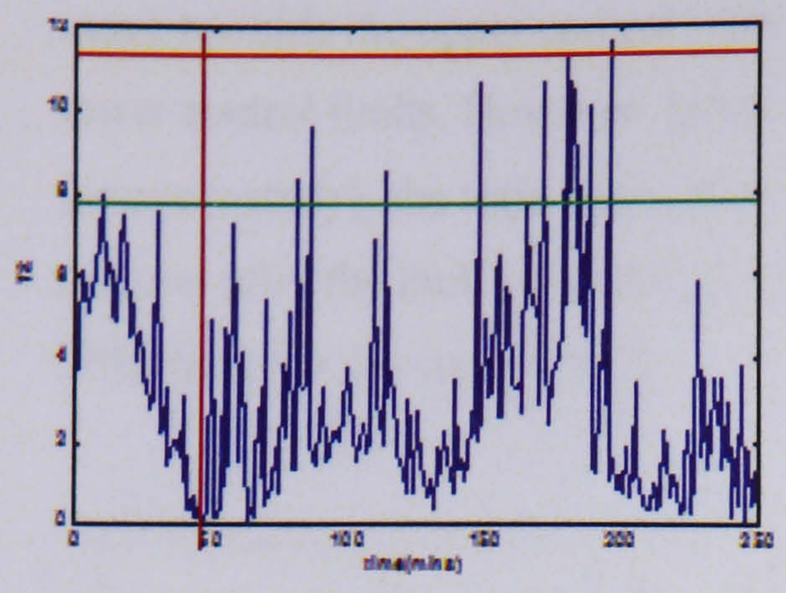
7.5.2.1 Fault Type 1



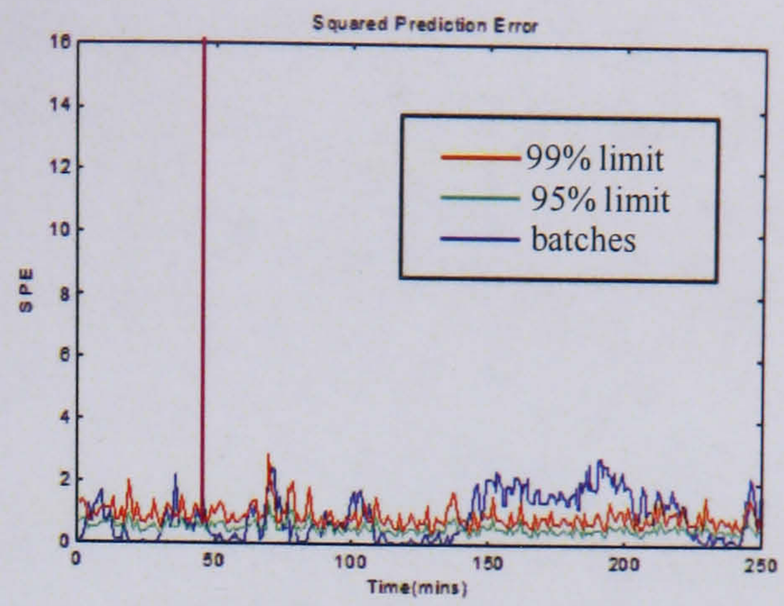
a) Latent variable 1 b) Latent variable 2 c) Latent variable 3

Figure 126: Batch observation level control charts for SMBMPCA with nonlinear PLS – Fault type 1

Fault type 1 is introduced at time point 50. At this point, there does not appear to be a noticeable change in the behaviour of the batch, although after approximately 60 time points the batch trajectory starts to move towards the control limits in latent variables 1 and 2. The Hotelling's T^2 and SPE plots are shown in Figure 127.



a) Hotelling's T^2 plot for fault type 1



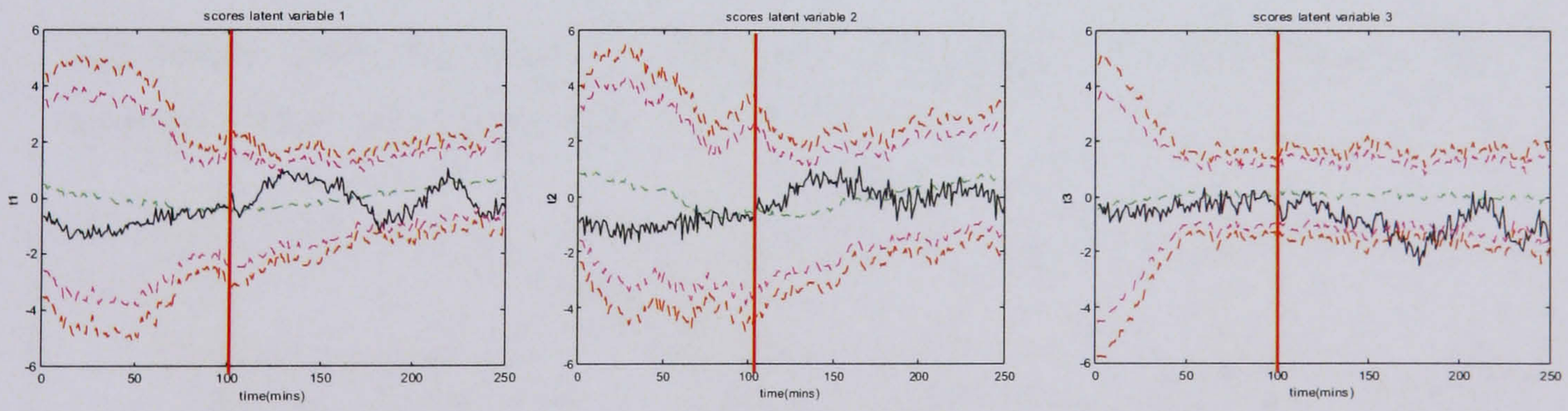
b) SPE for fault type 1

Figure 127: Hotelling's T^2 and SPE plots for fault type 1

The faulty batch trajectory is fairly noisy for the Hotelling's T^2 plot and therefore it is difficult to determine exactly when the batch moves outside of the control limits. However, on the basis that the batch is not considered to be out of control until it has been outside of the limits for 3 consecutive time points, the fault is not detected as cutting the 95% control limits until approximately time point 175, 125 minutes after it has been introduced. The batch never exceeds the 99% limits. In Figure 127 (b), the SPE plot, the deviation from the control limits does not occur until 100 minutes after the fault has been introduced. This is similar to what was witnessed with the batch observation level control charts, as the trajectory moves closer to the control limits at approximately time point 150.

7.5.2.2 Fault Type 2

The control charts for SMBMPCA with non-linear PLS are shown in Figure 128.

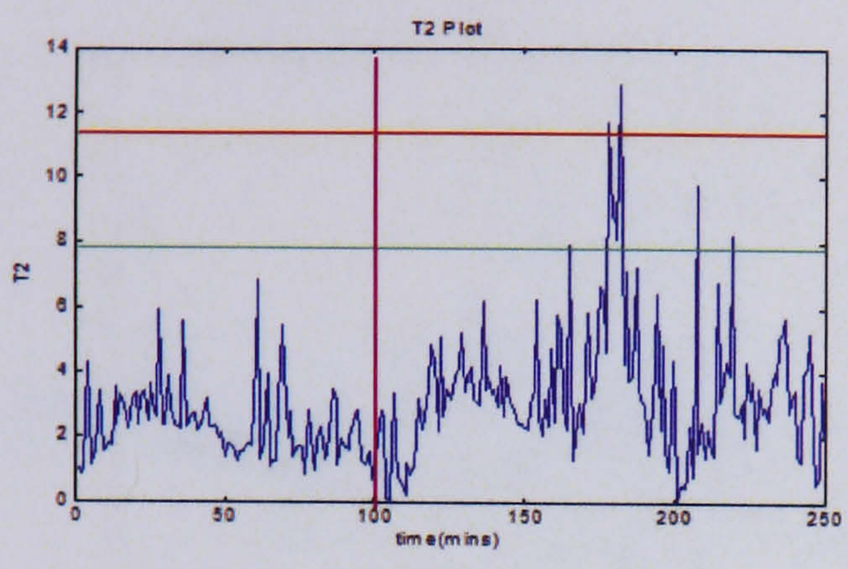


a) Latent variable 1 b) Latent variable 2 c) Latent variable 3

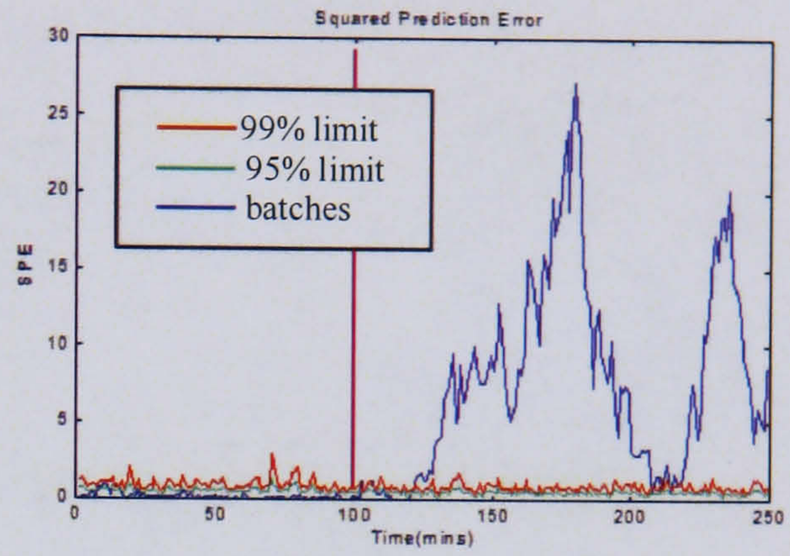
Figure 128: Batch observation level control charts for SMBMPCA with non-linear PLS – Fault type 2

Within approximately 10 minutes of the fault being introduced, there is a noticeable influence on the batch trajectory. For latent variables 1 and 2, it starts to increase and

move towards the upper control limits, whereas in latent variable 3 it moves towards the lower control limits. However, for the batch selected for these control charts (same batch for every study), the trajectory only deviates from the limits in latent variable 3, about 75 minutes after the fault has been introduced. Figure 129 illustrates the Hotelling's T^2 and SPE charts for the second fault.



a) Hotelling's T^2 plot for fault type 2



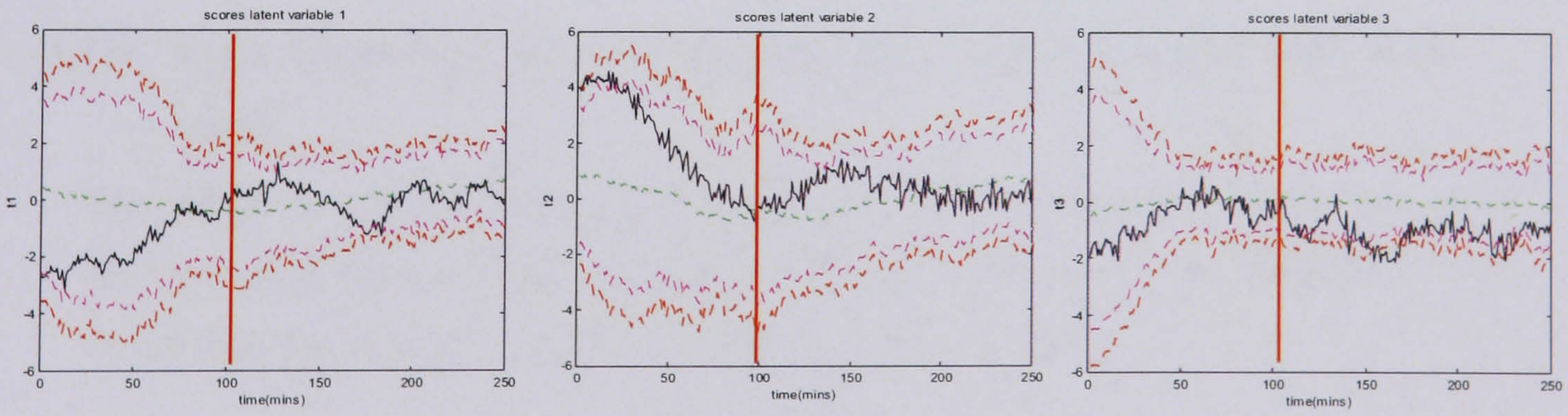
b) SPE plot for fault type 2

Figure 129: Hotelling's T^2 and SPE plots for fault type 2

The plot of the Hotelling's T^2 statisitic shows the batch moving towards the limits after the heat transfer coefficient fault has been introduced, before moving outside of the 99% limits approximately 75 minutes later. In the SPE plot the batch trajectory can be seen to deviate from the limits approximately 25 minutes after the fault occurs. The deviation from the limits is large, indicating a significant change in the correlation structure between the variables.

7.5.2.3 Fault Type 3

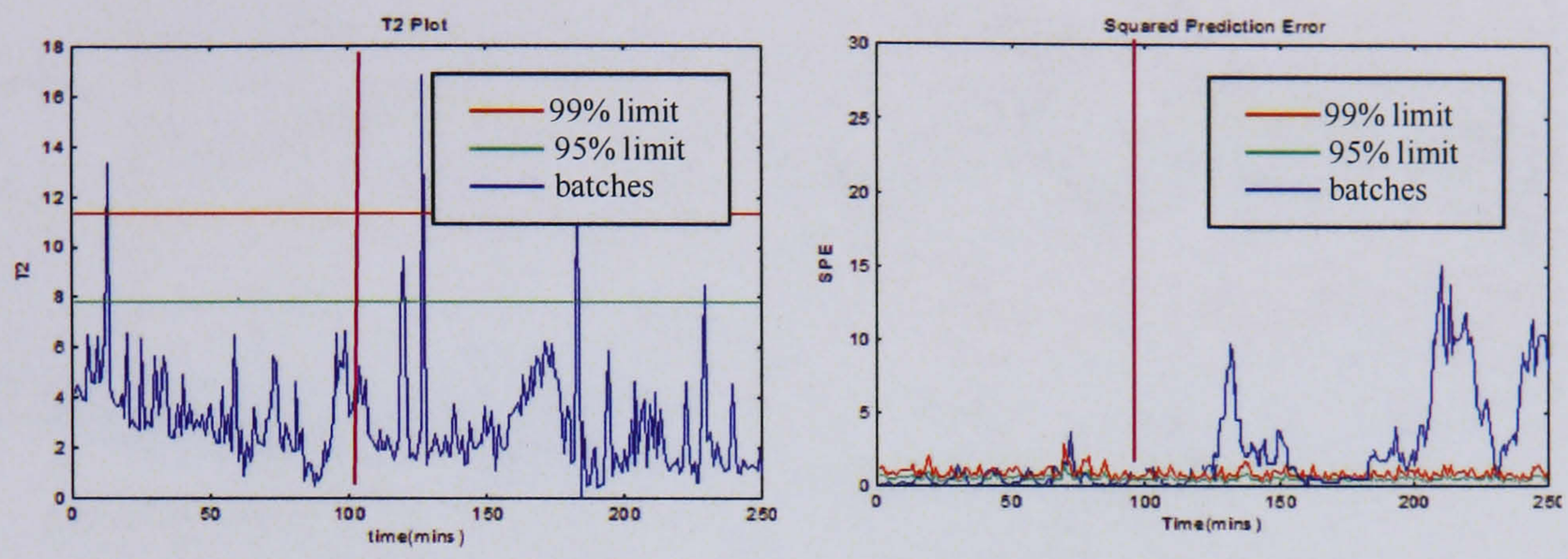
The control charts for Fault type 3, the fault in the cooling valve occurring after 100 minutes, are shown in Figure 130.



a) Latent variable 1 b) Latent variable 2 c) Latent variable 3

Figure 130: Batch observation level scores control charts for SMBMPCA with non-linear PLS – fault type 3

The batch observation level scores control charts show the batch trajectory moving towards the control limits after the fault has been introduced, although the trajectory only moves outside of the limits approximately 50 minutes later. The fault is more evident from the SPE and Hotelling's T^2 plots for fault type 3 in Figure 131.



a) Hotelling's T^2 plot for fault type 3 b) SPE plot for fault type 3

Figure 131: Hotelling's T^2 and SPE plots for fault type 3

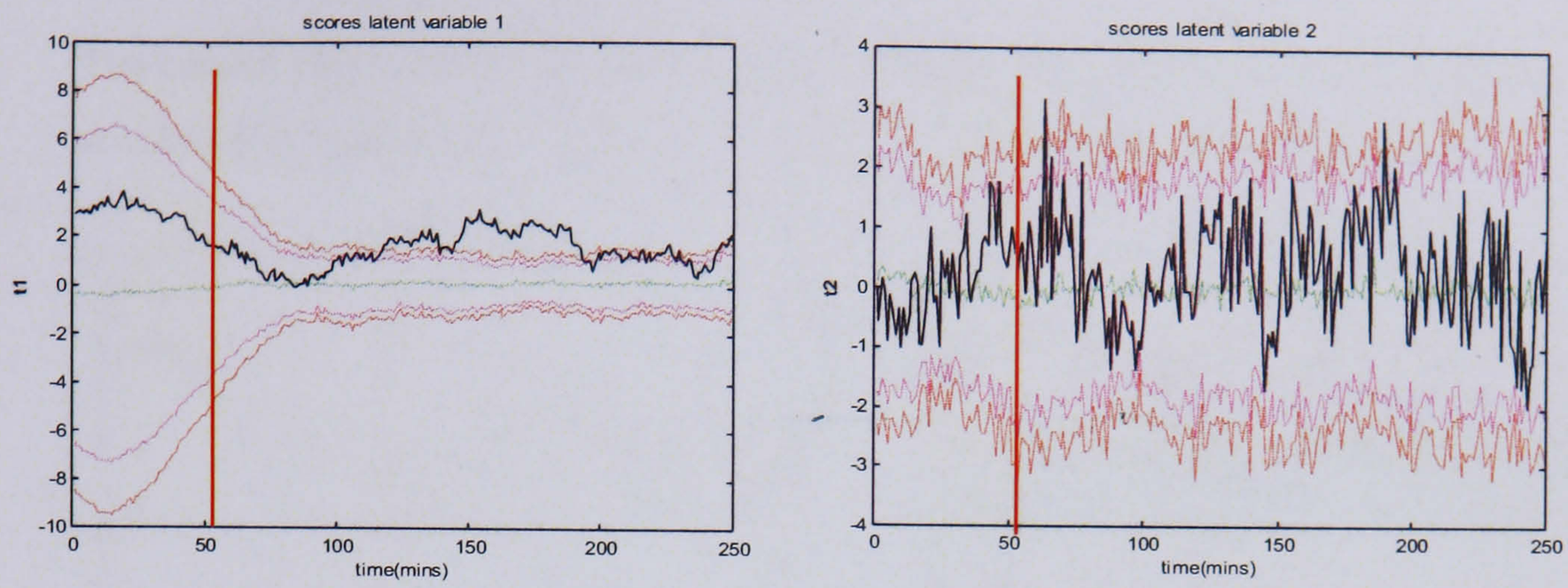
Both the Hotelling's T^2 and SPE control charts show the cooling valve fault causes the trajectories to move outside the control limits within 25 minutes of the fault being introduced. The batch then appears to move back into control, although it later deviates again with respect to the SPE chart. In the previous case study in Chapter 6 the cooling valve fault behaved in a similar way, deviating from the limits soon after it occurred, and then returning to normal operating conditions. This is because the fault only occurs when the cooling valve is moving, so during the second phase of operation where the batch exhibits steady state behaviour and the valve behaviour is fairly steady, the fault no longer has an impact.

7.5.3 Super model-based PCA with autoregressive with eXogeneous (ARX) input model

The SMBPCA with ARX error model technique demonstrated improvements in terms of the removal of the non-linear and dynamic structure in the batch data. The effect of this on the fault detection abilities of the technique is now examined.

7.5.3.1 Fault Type 1

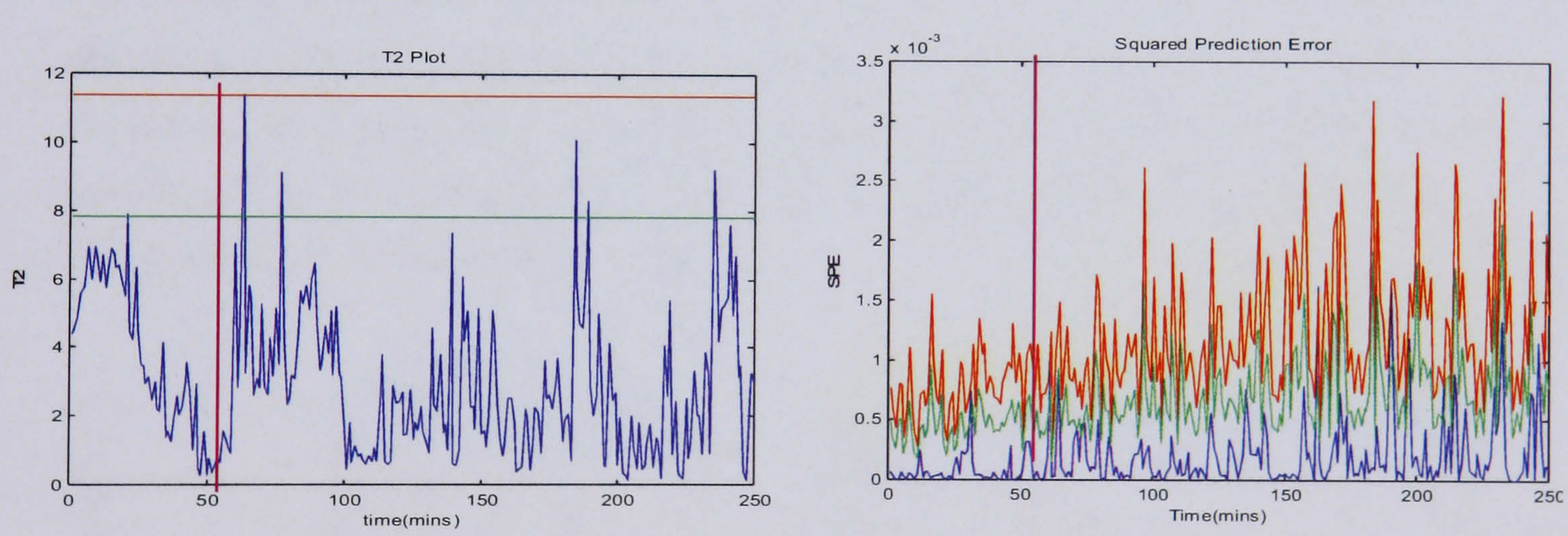
Figure 132 shows the batch observation level scores plots for fault type 1, the temperature sensor fault, for the two latent variables retained in the model.



a) Latent variable 1 b) Latent variable 2

Figure 132: Batch observation level control charts for SMBMPCA with ARX – Fault type 1

The batch trajectory moves outside of the control limits for latent variable 1, 50 minutes after the fault has been introduced, and although the trajectories move towards the control limits on latent variable 2, they do not deviate from them for a significant amount of time. The SPE and Hotelling's T^2 plots are shown in Figure 133:



a) Hotelling's T^2 plot for fault type 1 b) SPE plot for fault type 1

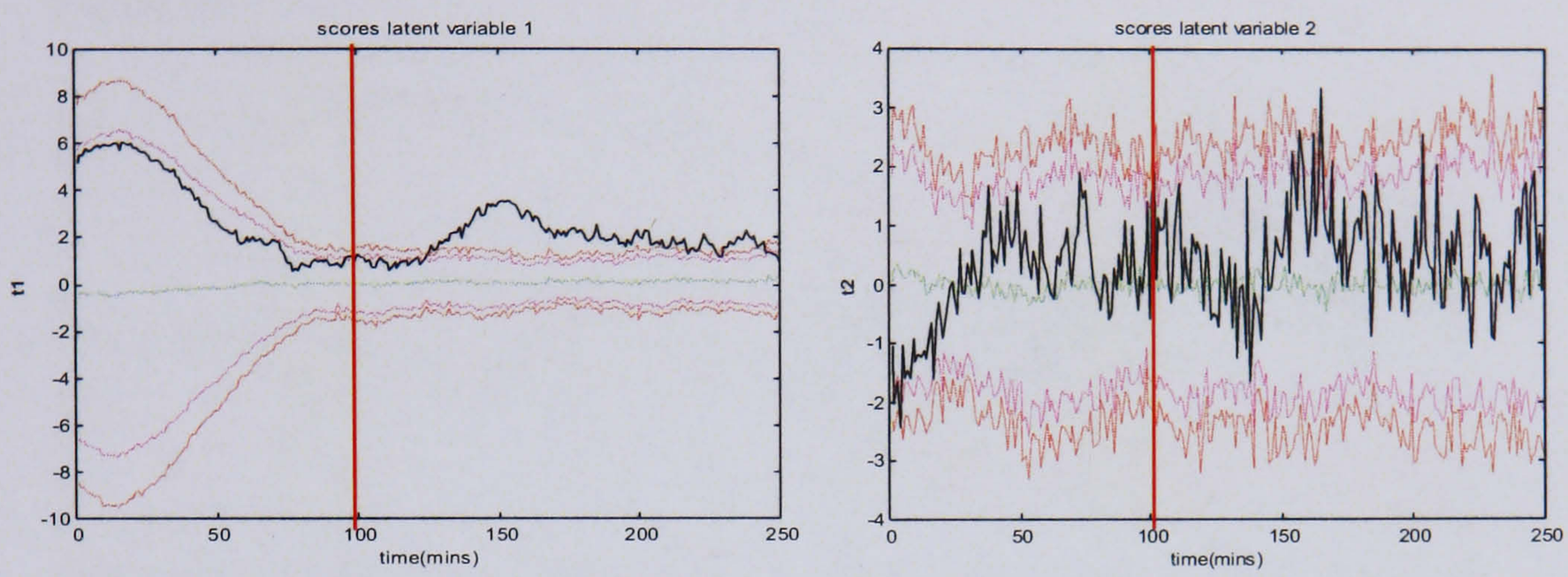
Figure 133: Hotelling's T^2 and SPE plots for fault type 1

The Hotelling's T^2 plot shows that the batch exceeds the 95% control limits soon after the fault has occurred, although it does not cross the 99% control limits. No fault is detected by the SPE control chart, which would indicate that there has not been a major change between the variables but instead a change in process conditions for example. This is an expected result because the reactor temperature affects all other variables in

the simulation, and so the fault would not change one variable in isolation, meaning there would not be a significant change in the correlations between the variables.

7.5.3.2 Fault Type 2

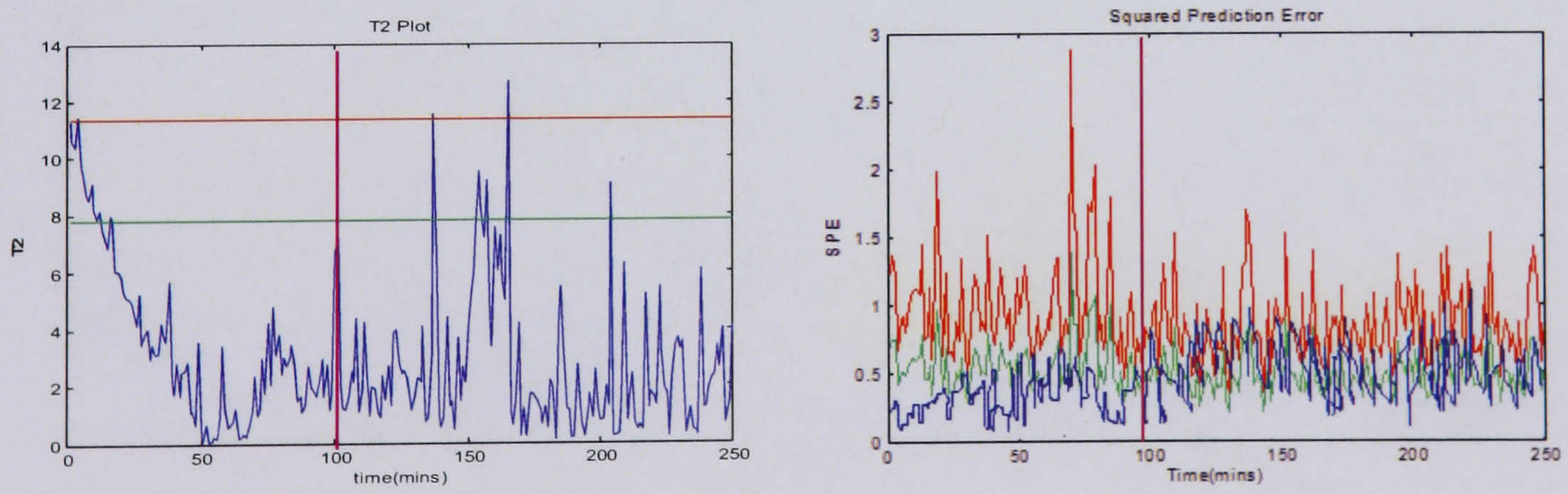
The control charts for the second fault type, the decrease in the heat transfer coefficient, are shown in Figure 134.



a) Latent variable 1 b) Latent variable 2

Figure 134: Batch observation level scores control charts for SMBPCA with ARX – Fault type 2

The non-conforming batch can be seen clearly moving outside of the control charts approximately 20 minutes after the heat transfer coefficient fault has been introduced. The process abnormality is not as evident from latent variable 2 due to the presence of noise. However, the trajectories can be seen moving towards the limits. The Hotelling's T^2 and SPE plots are shown in Figure 135.



a) Hotelling's T^2 chart of fault type 2 b) SPE chart of fault type 2

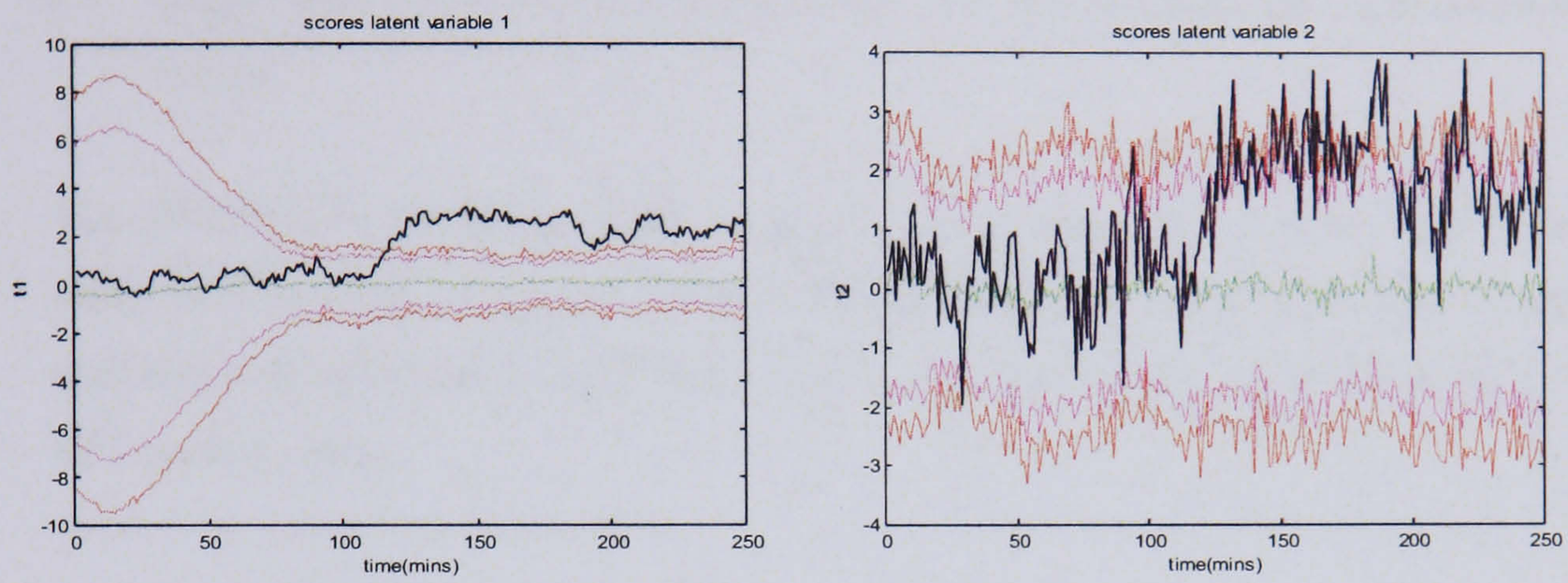
Figure 135: Hotelling's T^2 and SPE charts of fault type 2

The Hotelling's T^2 plot does not detect the batch as being abnormal until approximately 50 minutes after the fault has been introduced, and even then only the 95% limit is

exceeded. The SPE chart shows the batch trajectory moving towards the control limits after the fault has occurred, however it continues to fluctuate and does not fully move outside the 99% limit.

7.5.3.3 Fault Type 3

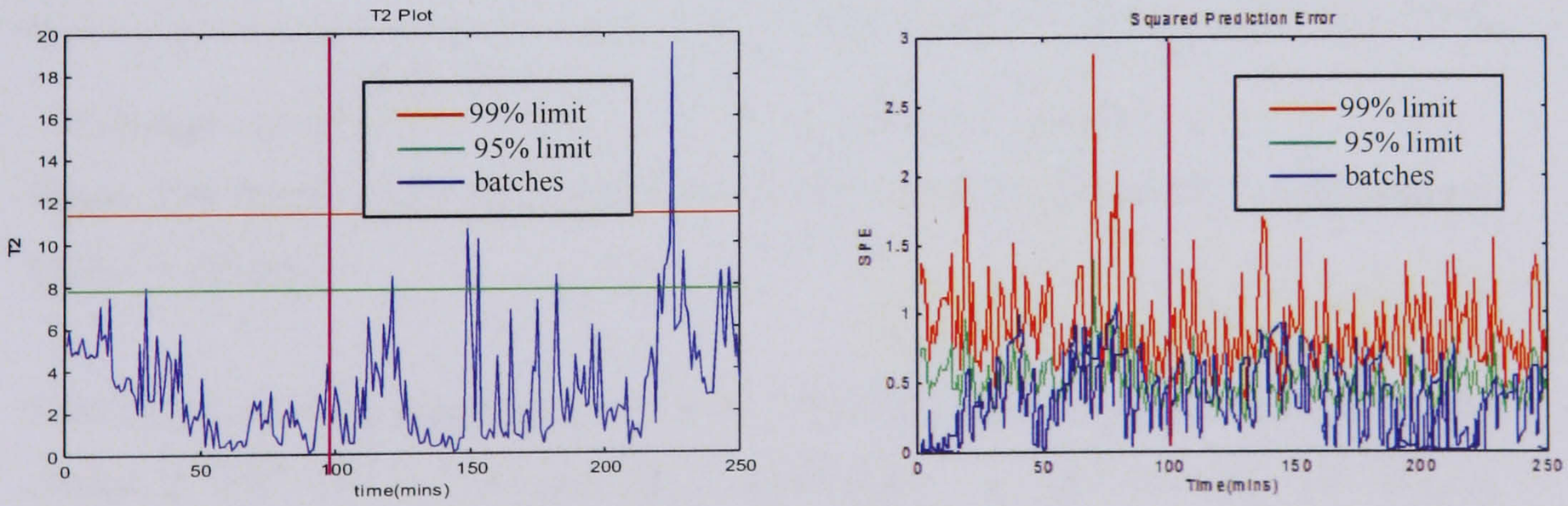
The control charts for fault type three, the cooling valve fault, are shown in Figure 136.



a) Latent variable 1 b) Latent variable 2

Figure 136: Batch observation level scores control charts for SMBMPCA with ARX – Fault type 3

After the cooling valve fault is introduced at time point 100, the batch trajectory moves outside the control limits for both latent variables. This happens within 15 to 20 minutes on latent variable 1, and within 50 minutes on latent variable 2. The Hotelling's T^2 and SPE plots are shown in Figure 137.



a) Hotelling's T^2 plot for fault type 3 b) SPE plot for fault type 3

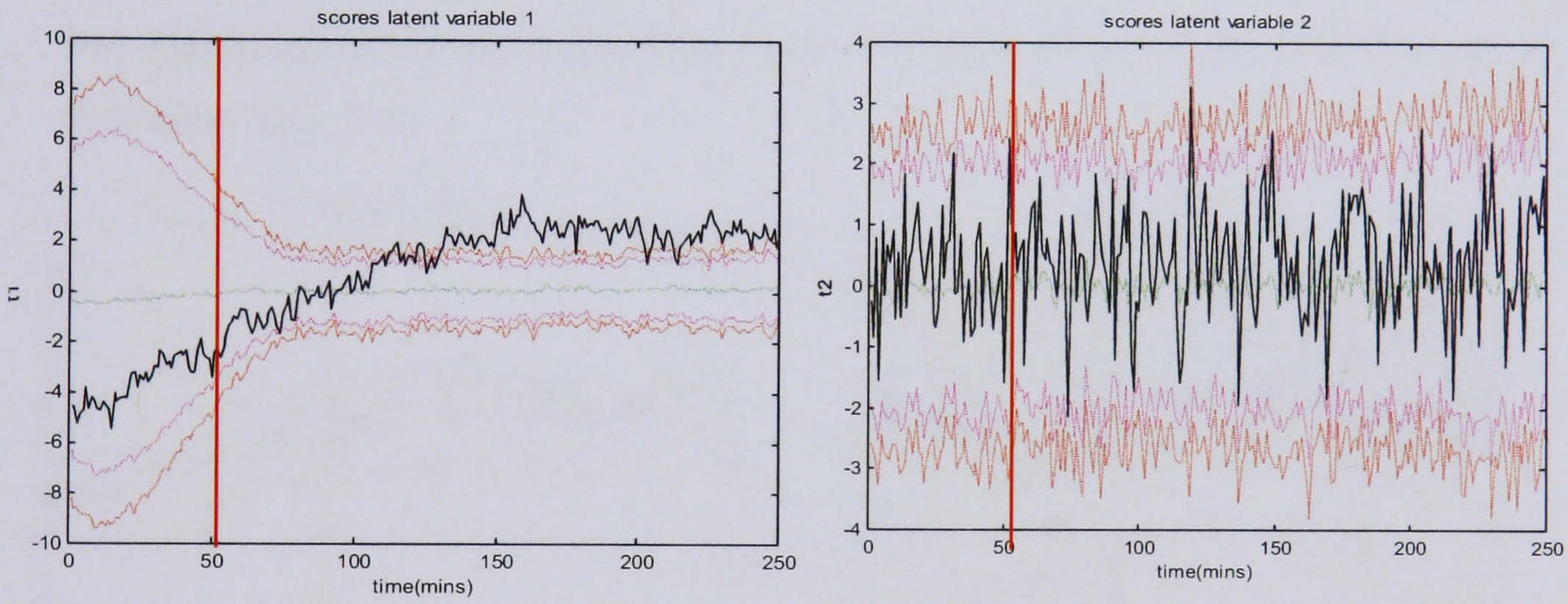
Figure 137: Hotelling's T^2 and SPE plot for fault type 3

The Hotelling's T^2 plot shows a small peak in the batch after the fault has been introduced which just exceeds the 95% limit, but not for more than three consecutive time points. The batch then exceeds the control limits towards the end of the batch, i.e. 100 minutes after the fault has occurred. In the SPE plot, the noise due to overfitting makes it difficult to tell if the fault has been detected.

7.5.4 Super model-based PCA with dynamic canonical correlation analysis error model

The SMBMPCA technique using a dynamic CCA error model gave good results in terms of the removal of the non-linear and dynamic structure in the data. The technique's fault detection abilities are assessed using the batch observation level, Hotelling's T^2 and SPE control charts.

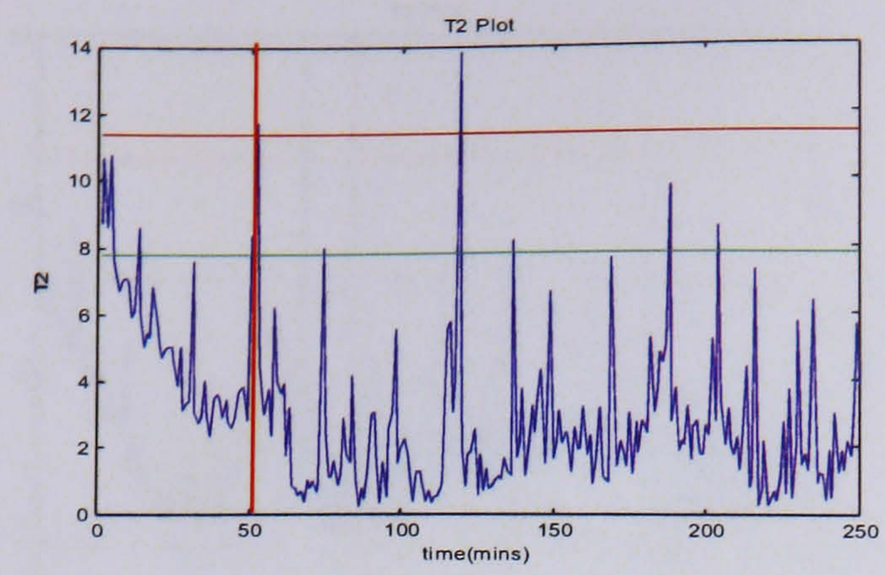
7.5.4.1 Fault Type 1



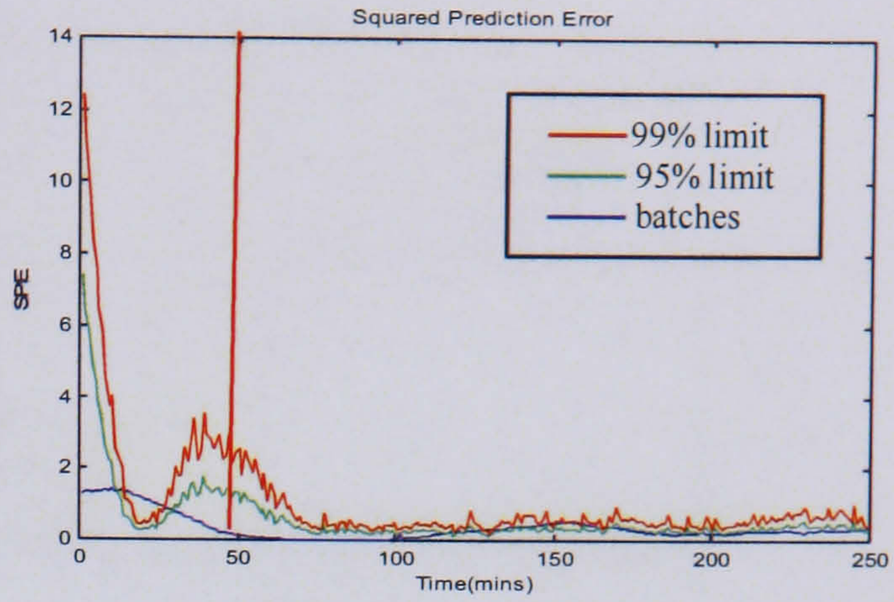
a) Latent variable 1 b) Latent variable 2

Figure 138: Batch observation level scores control charts for SMBMPCA with dynamic CCA – Fault type 1

Fault type 1, the temperature sensor fault, is introduced at time point 50. At this point the change in behaviour in latent variable 1 is not obvious as the trend continues moving towards the control limits before exceeding the approximately 75 minutes after the fault has occurred. The control chart for latent variable 2 does not show any difference in the batch behaviour. The Hotelling's T^2 and SPE control charts are shown in Figure 139:



a) Hotelling's T^2 plot for fault type 1



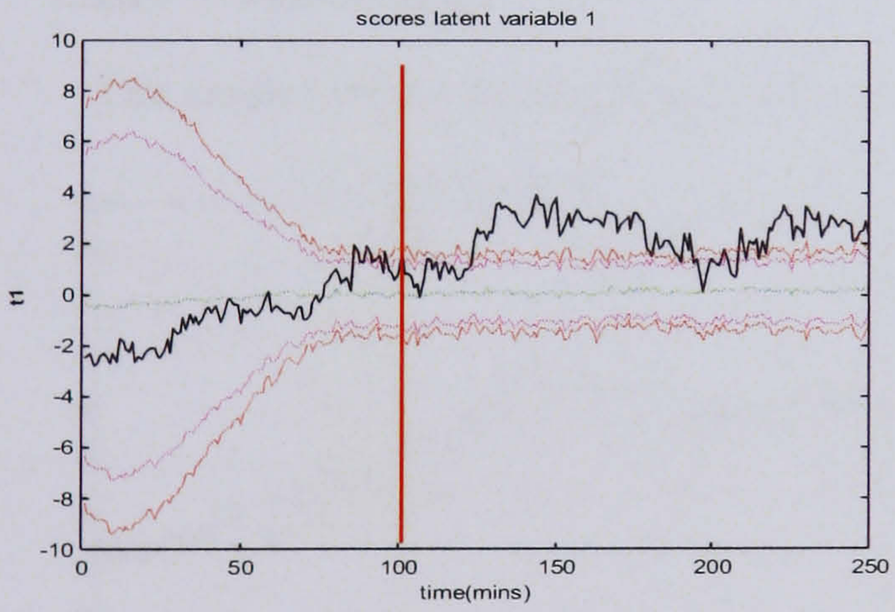
b) SPE plot for fault type 1

Figure 139: Hotelling's T^2 and SPE plots for fault type 1

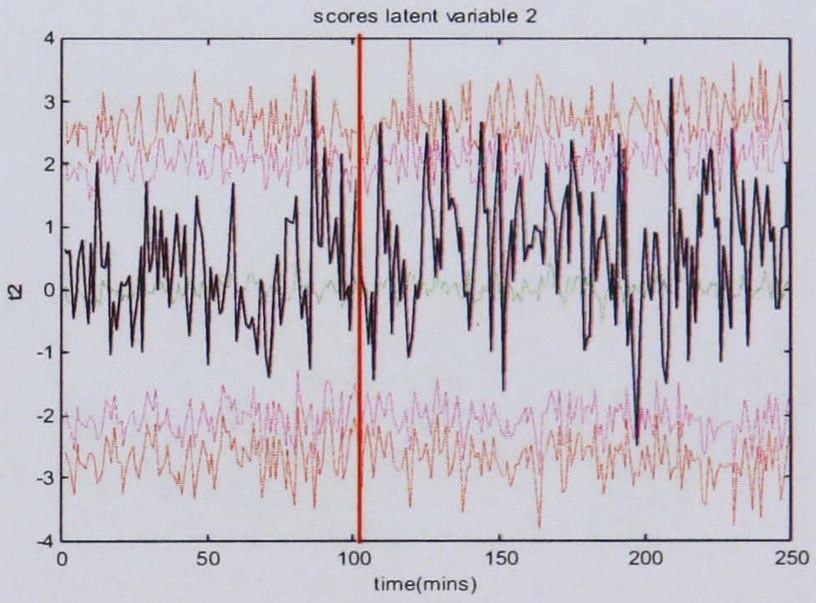
There is a spike outside the Hotelling's T^2 control limits as the fault is introduced, but the trajectory does not move outside the control limits for more than three consecutive time points. The SPE plot does not pick up any process abnormalities.

7.5.4.2 Fault Type 2

The control charts for the second fault type, the change in heat transfer coefficient are shown in Figure 140.



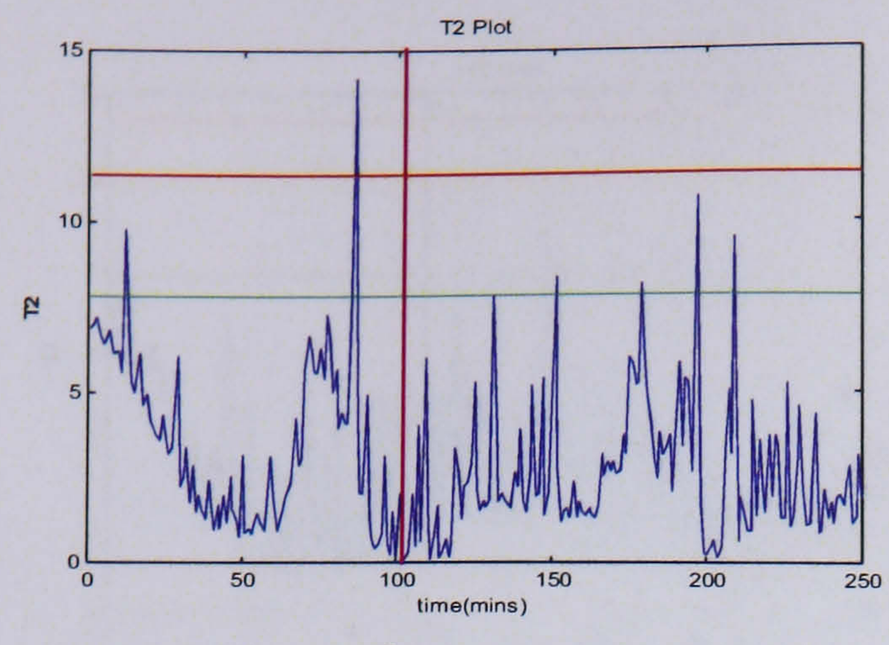
a) Latent variable 1



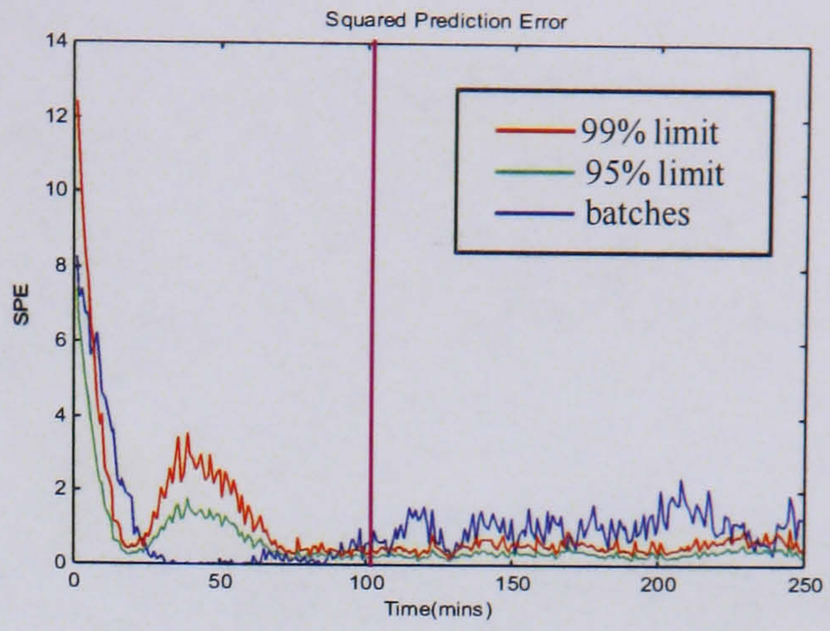
b) Latent variable 2

Figure 140: Batch observation level scores control charts for SMBMPCA with dynamic CCA – fault type 2

The faulty batch is detected leaving the control limits on the chart for latent variable 1 approximately 25 minutes after the fault is introduced. Latent variable 2 is too noisy to detect any problems with the batch.



a) Hotelling's T^2 plot for fault type 2



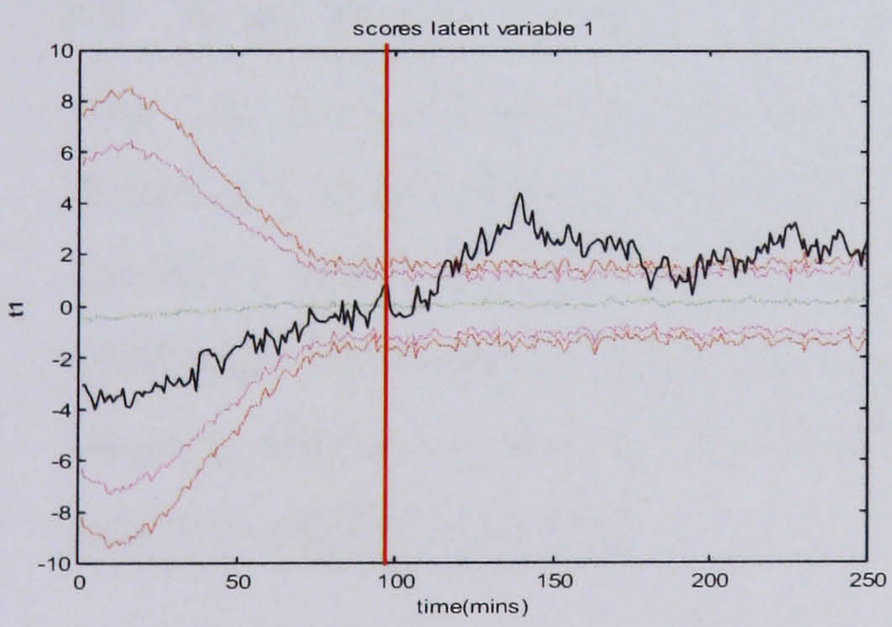
b) SPE plot for fault type 2

Figure 141: Hotelling's T^2 and SPE plots for fault type 2

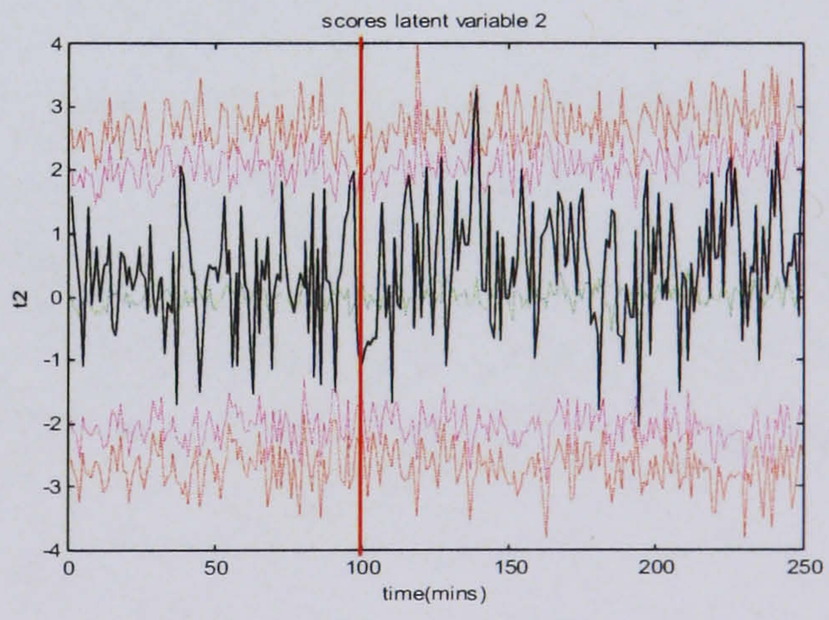
The Hotelling's T^2 plot in Figure 141 shows the batch trajectory moving outside of the control limits before the fault has occurred, it is assumed this is due to the noisy residuals. The SPE plot illustrates the trajectory deviating from the control limits in a few minutes of the fault being introduced and not returning to control. A fault that is more apparent on the SPE chart than the Hotelling's T^2 plot indicates there is a significant change in the correlation structure between the variables, which could be expected from a decrease in the heat transfer coefficient.

7.5.4.3 Fault Type 3

The control charts for the third fault type, the cooling valve fault, are in Figure 142.



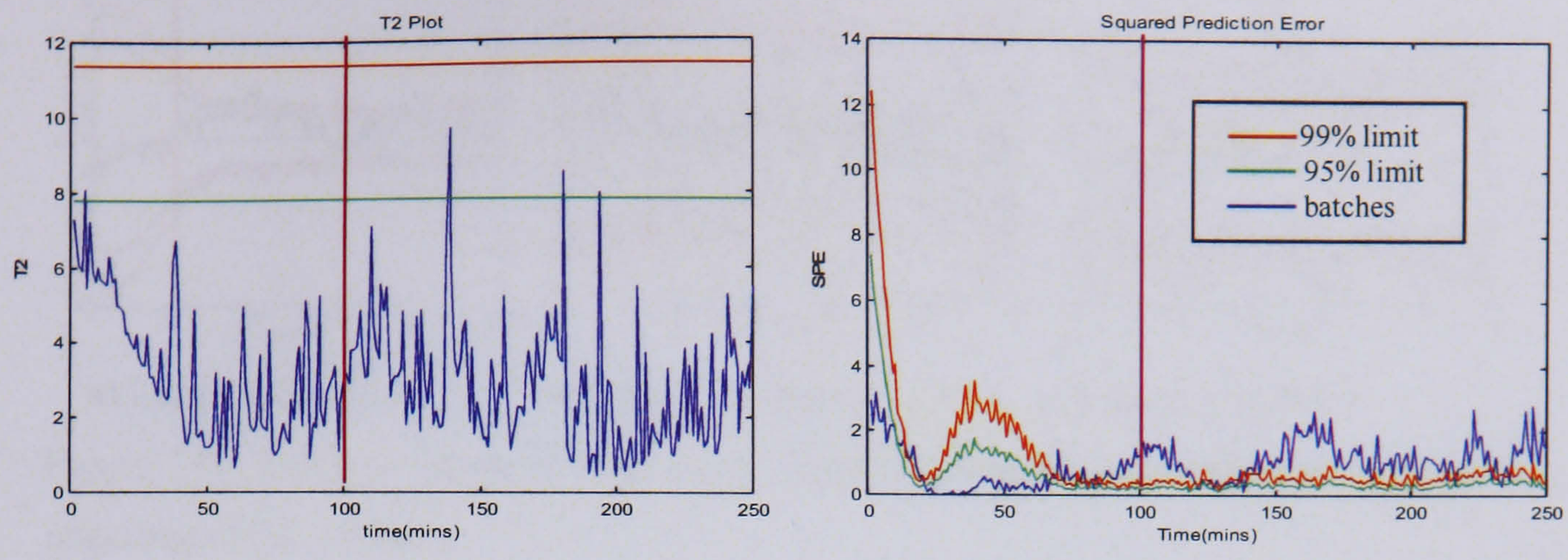
a) Latent variable 1



b) Latent variable 2

Figure 142: Batch observation level scores control charts for SMBMPCA with dynamic CCA – Fault type 3

The third fault is detected from latent variable 1 within the first 50 minutes of the fault being introduced. As witnessed with the cooling valve fault before the trajectory returns to the control limits and then exits again.



a) Hotelling's T^2 plot for fault type 3 b) SPE plot for fault type 3

Figure 143: Hotelling's T^2 and SPE plots for fault type 3

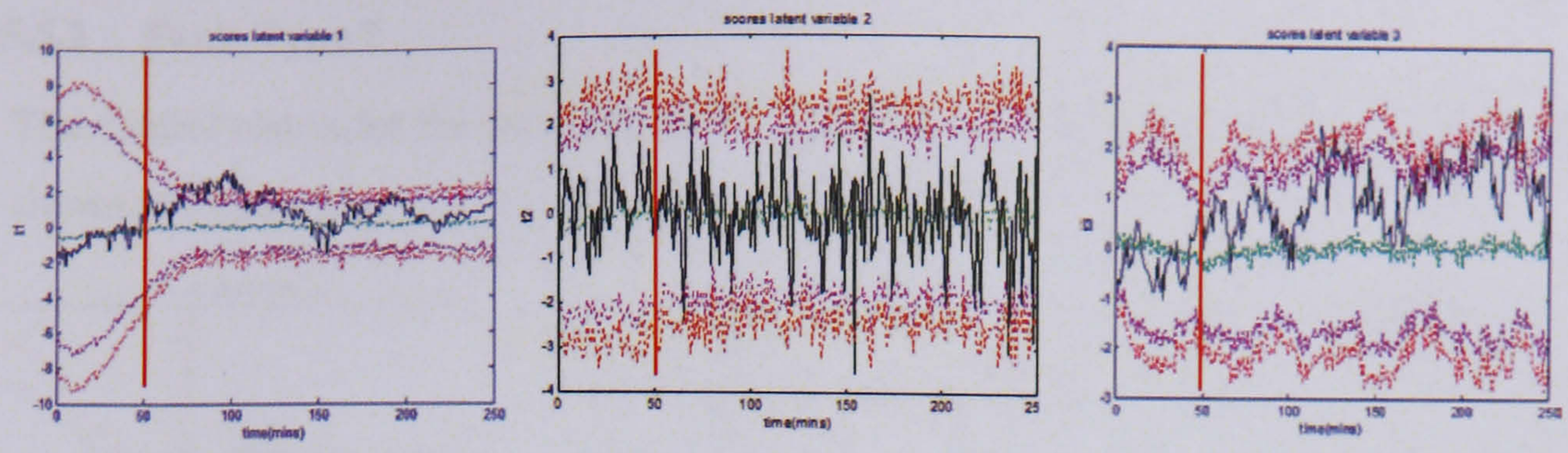
From the Hotelling's T^2 plot in Figure 143, the detection of the fault in the batch is difficult due to the noise in the data. The batch trajectory appears to leave the 95% control limits approximately 40 minutes after the fault has been introduced, however it never exceeds the 99% limits. The SPE plot shows the batch as already being outside of the control limits after the fault has been introduced, due to noise, however, it remains outside of the limits for the duration of the batch.

7.5.5 Super model-based PCA with dynamic non-linear PLS

The final error model used with the super model-based technique was dynamic non-linear PLS. The dynamic non-linear PLS method was applied to the structured residuals, creating a set of unstructured residuals on which the batch observation level analysis was performed. The technique dealt reasonably well with the removal of the non-linear and dynamic structure in the data. In the following sections the impact of this on its fault detection abilities is assessed.

7.5.5.1 Fault Type 1

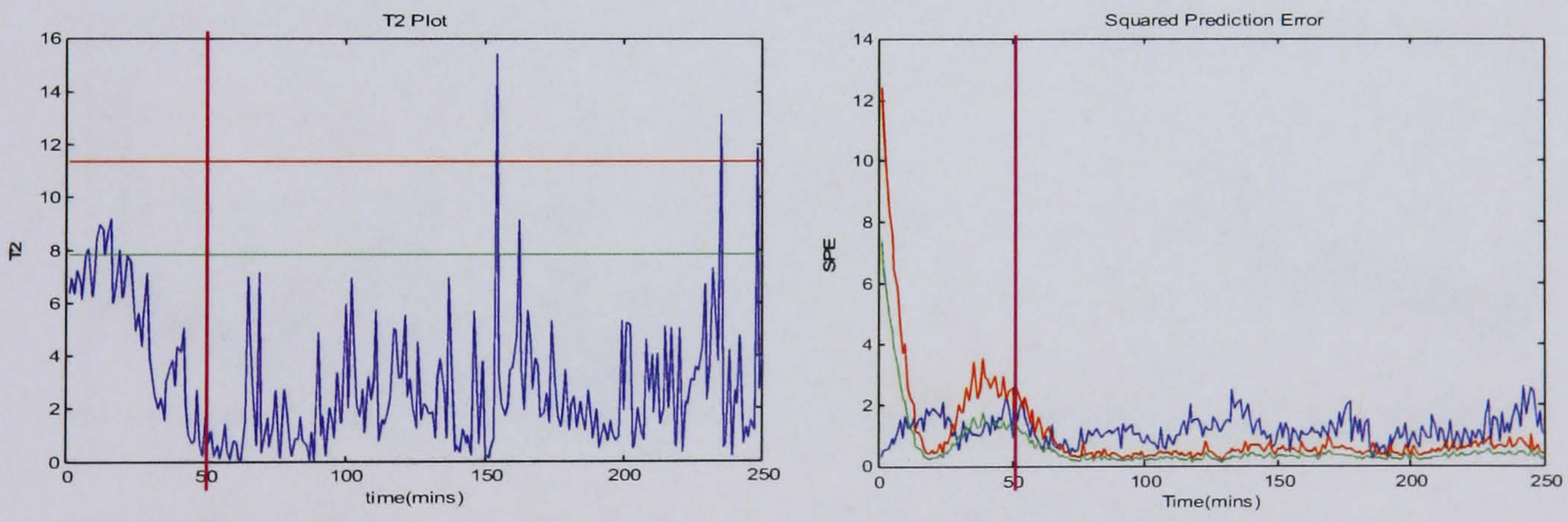
The batch observation level scores control charts incorporating the first fault type, the temperature sensor fault, are shown in Figure 144.



a) Latent variable 1 b) Latent variable 2 c) Latent variable 3

Figure 144: Batch observation level scores control charts for SMBMPCA with dynamic non-linear PLS – Fault 1

Fault type 1, the temperature sensor fault, is detected within 20 minutes from the batch observation level control charts, more rapidly than with any of the other methods investigated. It is difficult to detect the process abnormality in the Hotelling's T^2 control charts in Figure 145. This is partly because the fault is introduced at time point 50 and, during the first 100 time points the most variation is experienced in the batch, therefore detecting process deviations becomes more difficult. In the SPE chart in Figure 145, the batch does deviate from the control limits within 25 minutes of the fault occurring, however it also exits the control limits before the fault is introduced due to the noisy data most likely to be caused by overfitting of the model.

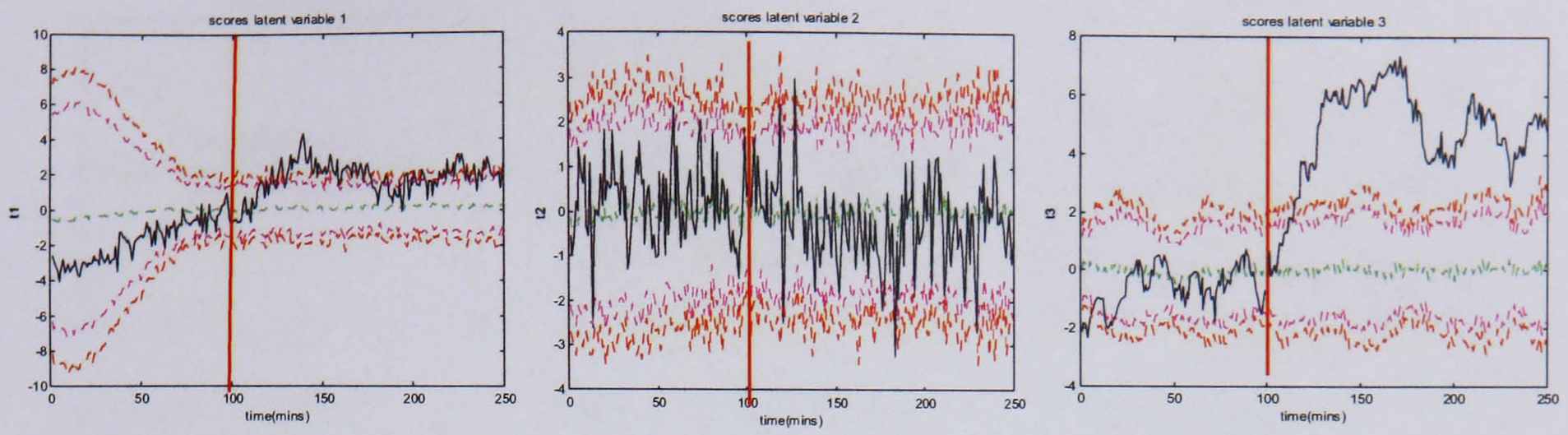


a) Hotelling's T^2 plot for fault type 1 b) SPE plot for fault type 1

Figure 145: Hotelling's T^2 and SPE plots for fault type 1

7.5.5.2 Fault Type 2

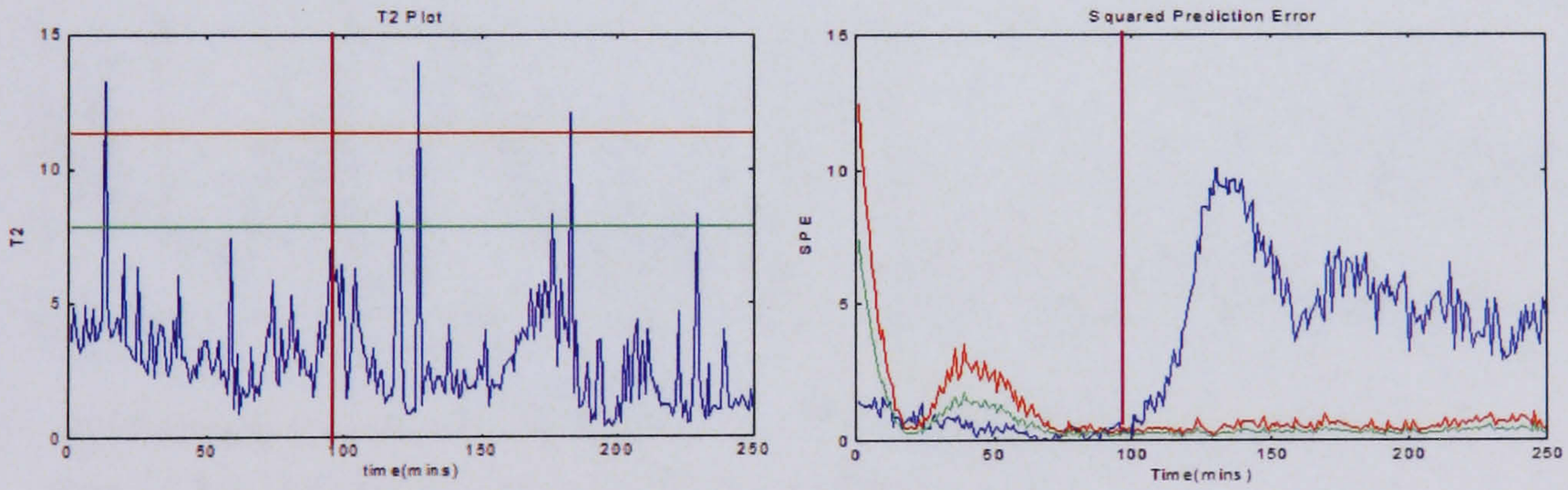
The control charts for the second fault type, the decrease in heat transfer coefficient, are shown in Figure 146.



a) Latent variable 1 b) Latent variable 2 c) Latent variable 3

Figure 146: Batch observation level scores control charts for SMBMPCA with dynamic non-linear PLS – Fault 2

The change in behaviour of the fault trajectory is noticeable as soon as the fault is introduced. For latent variable 1, the batch trajectory moves outside of the upper control limits approximately 20 minutes after the fault is first introduced. Similar behaviour is observed from latent variable 3, where the trajectory starts moving towards the control limits as soon as the fault occurs, exceeding the limits in under 20 minutes. The Hotelling's T^2 and SPE plots for the second fault type are shown in Figure 147.



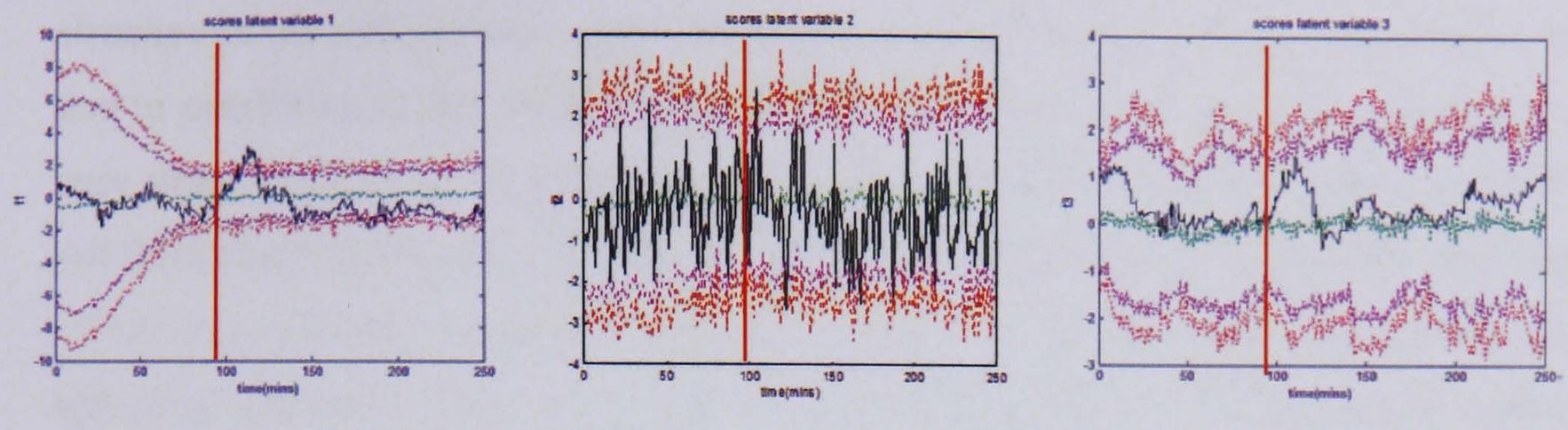
a) Hotelling's T^2 plot for fault type 2 b) SPE plot for fault type 2

Figure 147: Hotelling's T^2 and SPE plots for fault type 2

The Hotelling's T^2 plot shows the batch trajectory leaving the control limits within approximately 25 minutes of the fault being introduced, however it returns to operation within the limits shortly afterwards. However, the SPE plot shows the batch moving outside of the control limits as soon as the fault occurs and continues to increase in value and remains outside of the limits for the duration of the batch.

7.5.5.3 Fault Type 3

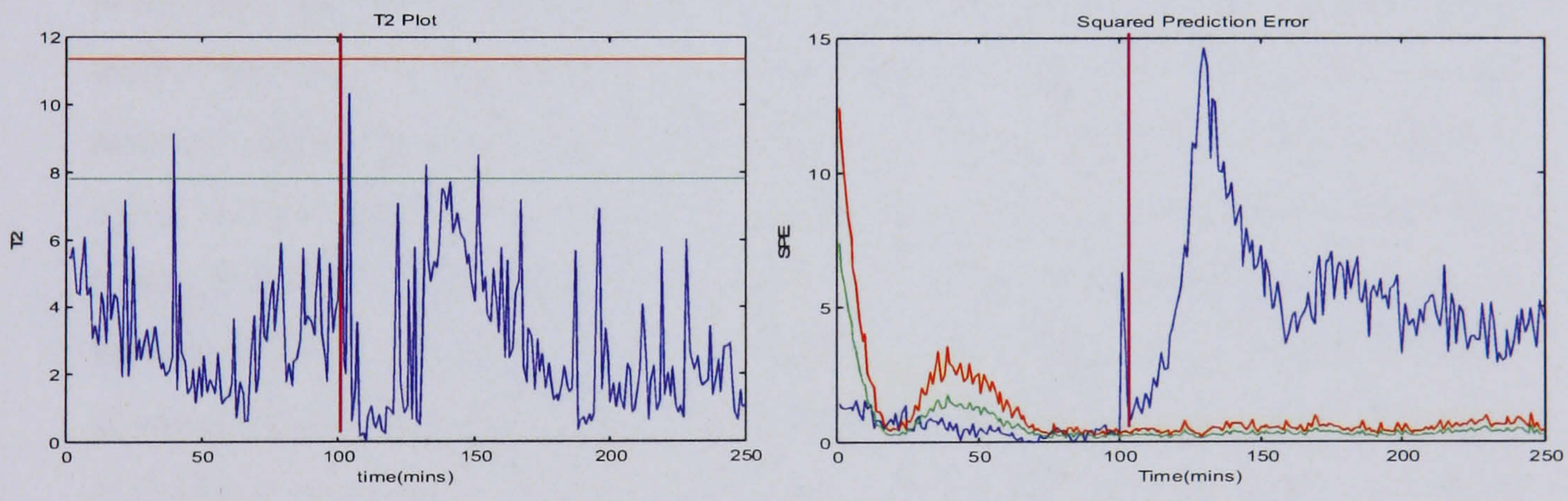
The batch observation level control charts for the third fault type, the cooling valve fault, are shown in Figure 148.



a) Latent variable 1 b) Latent variable 2 c) Latent variable 3

Figure 148: Batch observation level scores control charts for SMBMPCA with dynamic non-linear PLS – Fault 3

As the fault in the cooling valve is introduced, the batch trajectory for latent variable 1 moves outside of the control limits fairly quickly, within about 15 time points, before returning to normal operating conditions for the duration of the batch, as was seen with previous projections of this fault type. For latent variable 3 the batch is seen to respond quickly to the introduction of the fault, however it does not move outside of the limits.



a) Hotelling's T^2 plot for fault type 3 b) SPE plot for fault type 3

Figure 149: Hotelling's T^2 and SPE plot for fault type 3

Both the Hotelling's T^2 and SPE plots show the trajectory moving outside of the control limits almost immediately as the fault occurs, in the case of the Hotelling's T^2 plots, the trajectory then moves back inside the limits, whereas the SPE statistic for the faulty batch increases and remains outside of the limits for the duration of the batch.

7.6 Summary of Fault Detection Results

One effect of the application of the super model-based approach is that although the structure in the data is reduced, the residuals generated are fairly noisy, suspected to be due to overfitting of the model. Three different types of control chart were used in the case study, batch observation level scores control charts, Hotelling's T^2 charts and SPE. All three charts displayed noisy data, however with the Hotelling's T^2 and SPE plots this effect is particularly apparent, making it difficult to determine whether a batch is operating outside of its control limits. To address this problem a filter could be applied to the residual data to smooth the trajectories, as discussed further in Chapter 9. A filter was not used in this case study in order to give a true representation of the effects of the super model-based approach on the batch data and residuals.

For each of the techniques considered in this case study, the out of control average run length (ARL) was calculated. This is defined as the number of observations between a fault occurring in the process and the fault being detected by the control chart, calculated as an average over the full set of batches (fifty for each fault). An ARL value was calculated for each of the three different control charts for each technique. However due to the noise present in the residuals used to generate the control charts, the ARL values for the Hotelling's T^2 and SPE charts were difficult to calculate. Although there was also noise present in the batch observation level scores control charts, it was still possible to detect process abnormalities in the batch for some of the latent variables. Therefore, although the ARL values for the Hotelling's T^2 and SPE charts are included there is more confidence in the batch observation level scores results. In the case of the batch observation results, all three latent variables were examined, and the shortest run length of the three was used in the ARL calculation. As previously, the batch is only considered to be out of control if its trajectory remains outside of the control limits for at least three consecutive time points.

The average run lengths for each of the super model-based techniques are summarised in Figure 150, alongside the ARLs for the batch observation level analysis on the original data and the model-based PCA technique, as generated in the case study in Chapter 6. The summaries of the ARLs for the Hotelling's T^2 and SPE control charts are shown in Figure 151 and Figure 152 respectively.

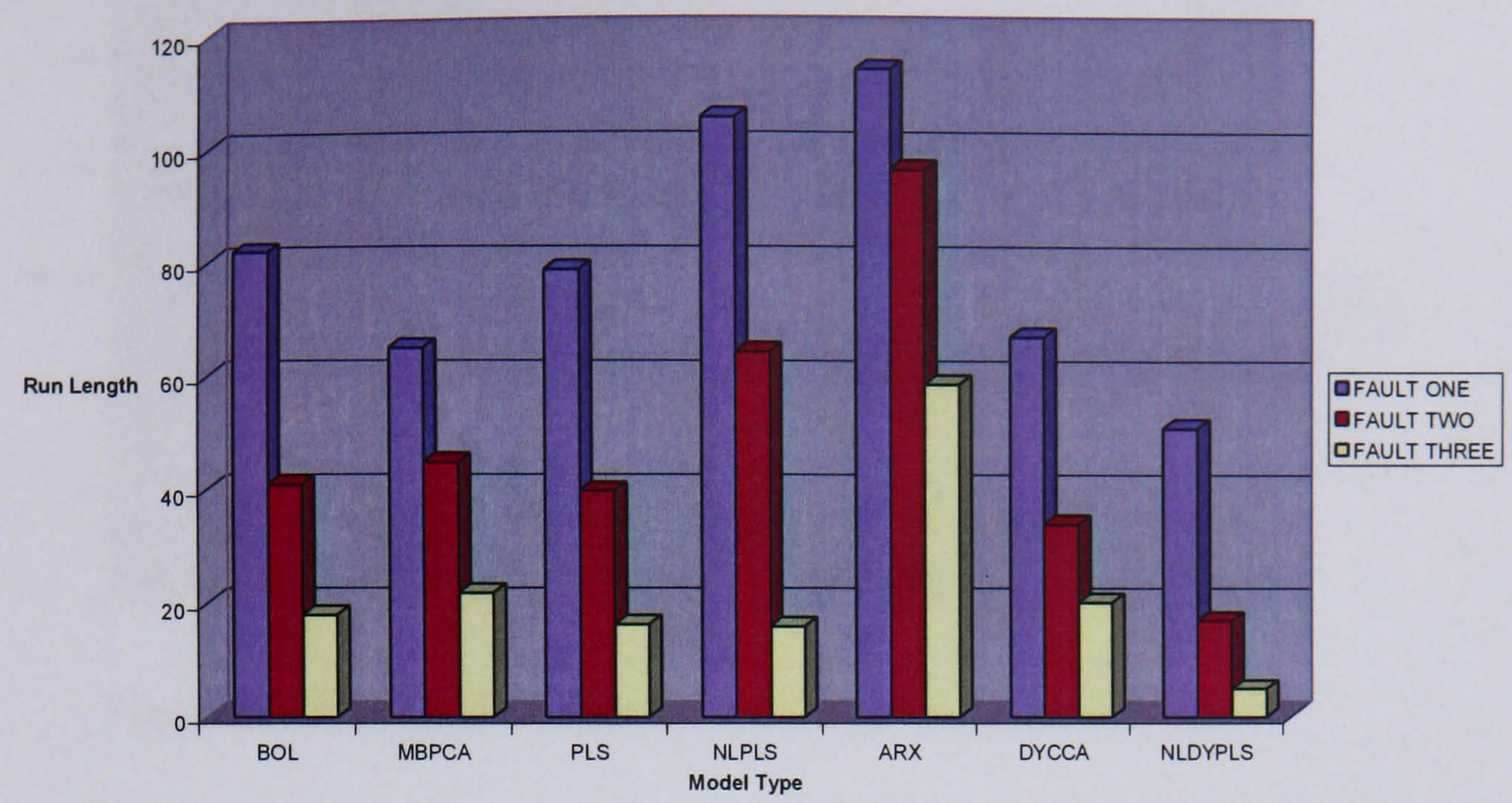


Figure 150: Average run length for batch observation level scores chart

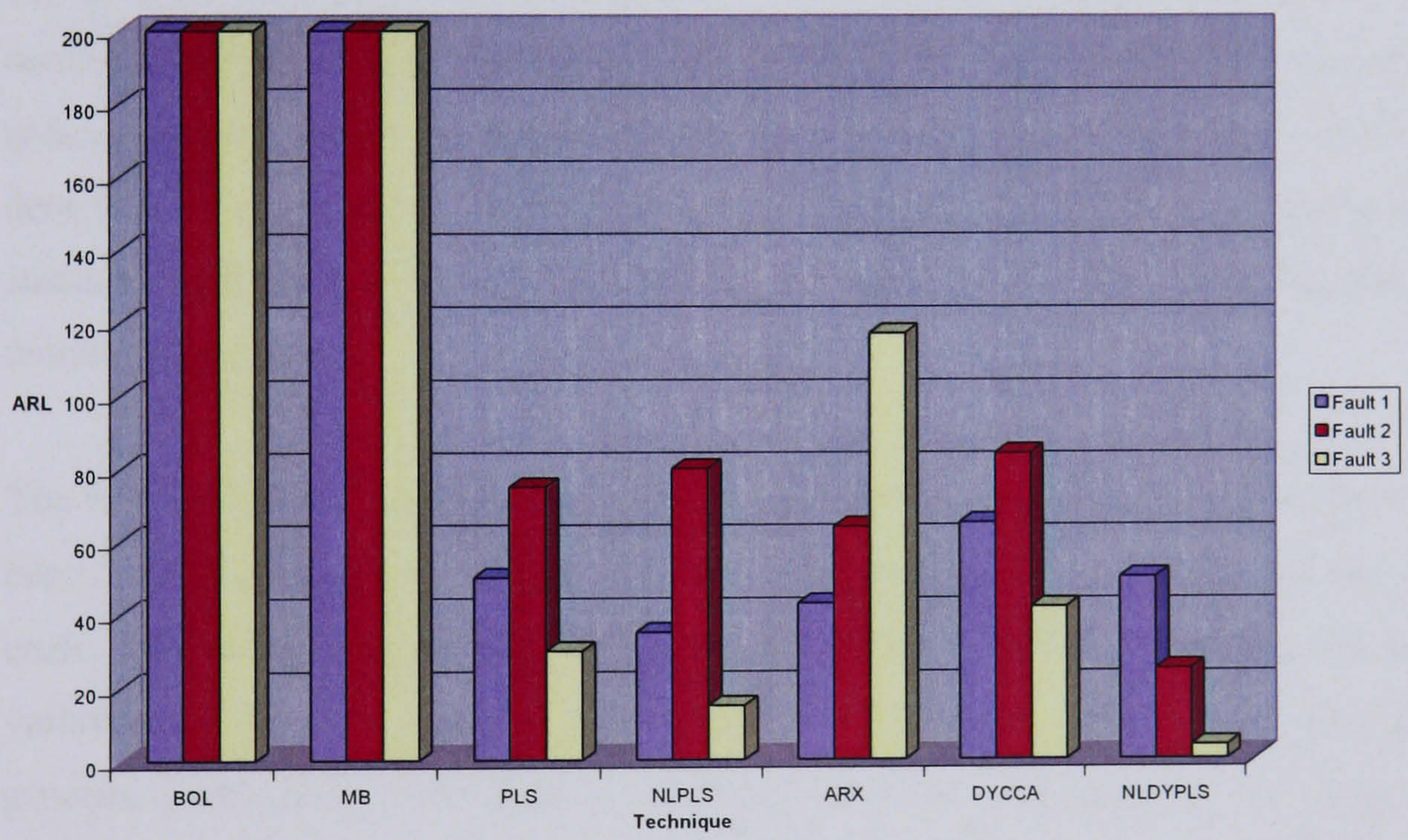


Figure 151: Average run length for Hotelling's T^2 chart

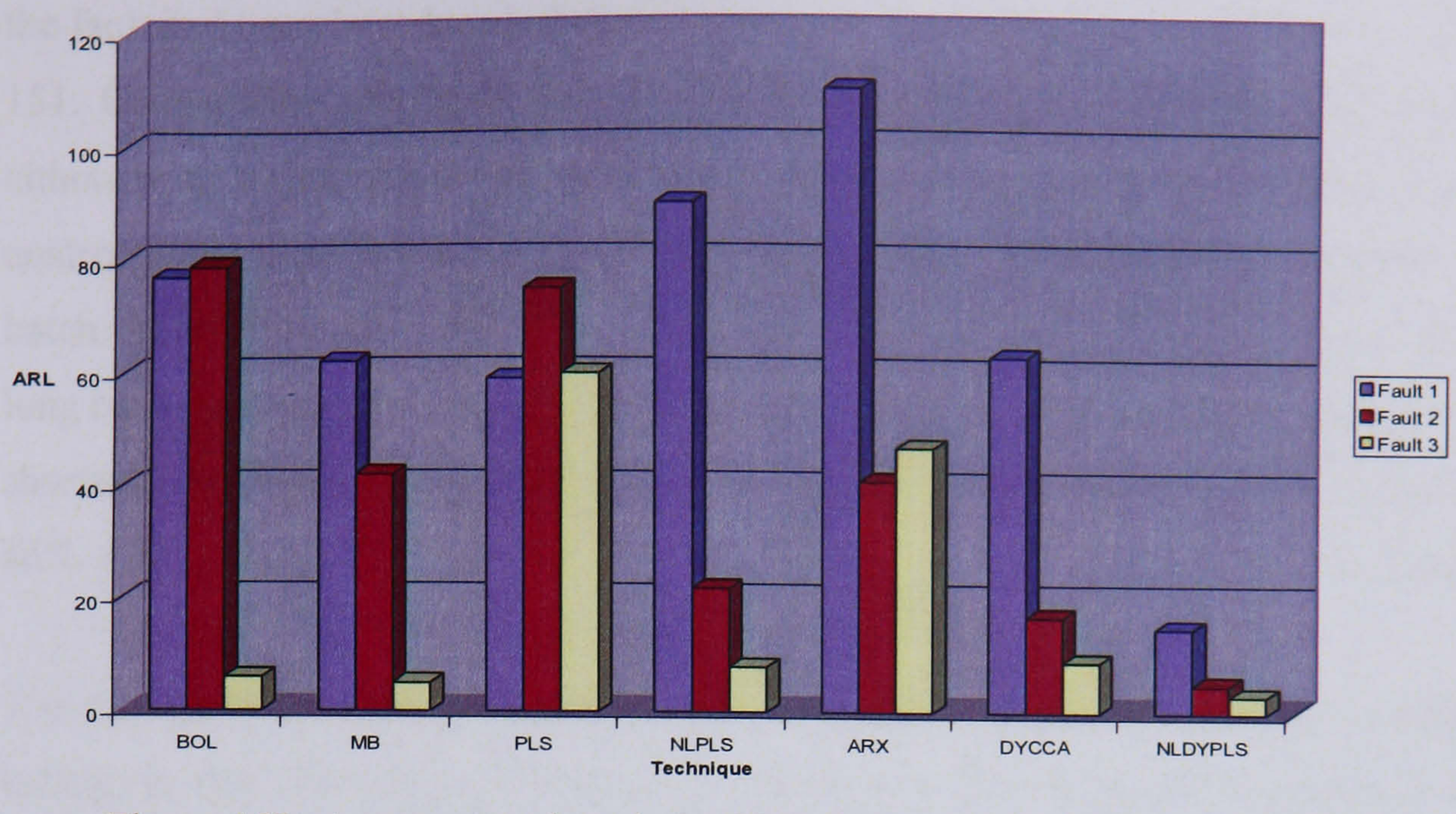


Figure 152: Average run length for SPE chart

The results of the super model-based residuals analysis shows that the dynamic CCA and non-linear dynamic PLS error models were the most effective in terms of removing the dynamic and non-linear structure from the residuals. Figure 150 shows similar results in terms of fault detection, with DYCCA and NLDYPLS demonstrating the greatest improvements in terms of a reduction in average run length. Overall, the detection times for each of the fault types have generally decreased in comparison to the standard batch observation level monitoring technique and to the standard model-based monitoring approach.

The first fault type, the reactor temperature sensor, was introduced 50 minutes into the batch, had the longest run length in terms of the batch observation level scores control chart. This is because the first 100 minutes modelled contain the greatest amount of variation in the data, making it more difficult to detect abnormalities entering the process, particularly if the fault is relatively small in comparison to the range of the process parameter and the noise in the process. However, a reduction in detection time of 30 minutes on average was observed using the super model-based approach with DYCCA and NLDYPLS. The run lengths of the NLPLS and ARX super model-based techniques increased for the first fault types. This is due to the fact that the modelling techniques were developed to deal with either the non-linear behaviour or the dynamic structure in the data and not both, whereas the DYCCA and NLDYPLS techniques take into account both characteristics. Figure 151 summarises the Hotelling's T^2 average run lengths. It is difficult to make any comparisons between the standard BOL approach and the MBPCA approach with the SMBMPCA approach in this case because the Hotelling's T^2 control charts generally showed the batches as being out of control before

the fault had been introduced, therefore faults are shown as being undetectable in Figure 151. Comparisons can be made between the different super model-based approaches, although as mentioned previously, due to the noise present in the residuals and the control limits there is not a high level of confidence in the results. The opposite of the batch observation ARL results are obtained, with the NLDYPLS error model giving a long run length using the Hotelling's T^2 control charts, and the NLPLS model giving the shortest. The average SPE run lengths in Figure 152 display a more similar story to the BOL ARL chart, with the NLDYPLS error model again giving the shortest run length.

The second fault type investigated was a decrease in the heat transfer coefficient, indicating that fouling has occurred in the reactor. This fault was introduced to the batches after 100 minutes and the standard batch observation level model took an average of 40 minutes to be detected. The average run lengths generated by the BOL control charts showed that the NLDYPLS and DYCCA techniques reduced this detection time by more than half. Again the ARL increases for the ARX and NLPLS models, as they do not fully address the structure contained in the residuals. The Hotelling's T^2 ARL plots also show the NLDYPLS technique having the shortest fault detection time, with the other model-based techniques all displaying similar results. The NLDYPLS error model also gives the best result in terms of the SPE run lengths, however again it must be mentioned that the confidence level is lower in the Hotelling's T^2 and SPE results.

The third fault introduced to the case study was the cooling valve fault, again introduced at time point 100. In this case a failure occurred in the split range control valve pressure output. This fault was more challenging to detect because it was small in magnitude and short in terms of the time in which it was present in the process. The only super model-based approach which significantly reduced the fault detection time was the NLDYPLS method with the ARL being more than halved. The other super model-based techniques either gave similar results to that of the standard BOL technique or worse results. This is because the fault was fairly small in magnitude and therefore difficult to detect amongst the noisy residuals. Again, similar results were demonstrated with the Hotelling's T^2 and SPE average run lengths.

7.7 False Alarm Rate

As well as considering the fault detection abilities of the super model-based techniques, the false alarm rate is also assessed. To be effective as a performance monitoring tool, the super model-based techniques must also have a low false alarm rate, an overly sensitive technique could cause too many spurious alarms, the consequence of this being that important alarms concerning the condition of the process could be ignored by the operator. To test the false alarm rate of these techniques, the super model-based approaches were applied to the set of 20 validation batches generated in Chapter 3. The residuals produced from the application of these techniques were then projected onto the corresponding control limits. The batches were generated under the same conditions as the nominal data set and are therefore expected to remain within the control limits for the duration of the batch. On this basis it is assumed that any deviation from the control limits constitutes a false alarm, and as with previous studies the trajectory must remain outside of the control limits for three consecutive time points to be considered out of control. All three latent variables were examined for each batch and a false alarm was registered if the trajectory was out of the confidence limits on any of the latent variable plots. Due to the noise problems experienced with the Hotelling's T^2 and SPE control charts, only the batch observation level control charts were assessed. Figure 153 shows the number of batches generating a false alarm out of the 20 batches in each validation data set.

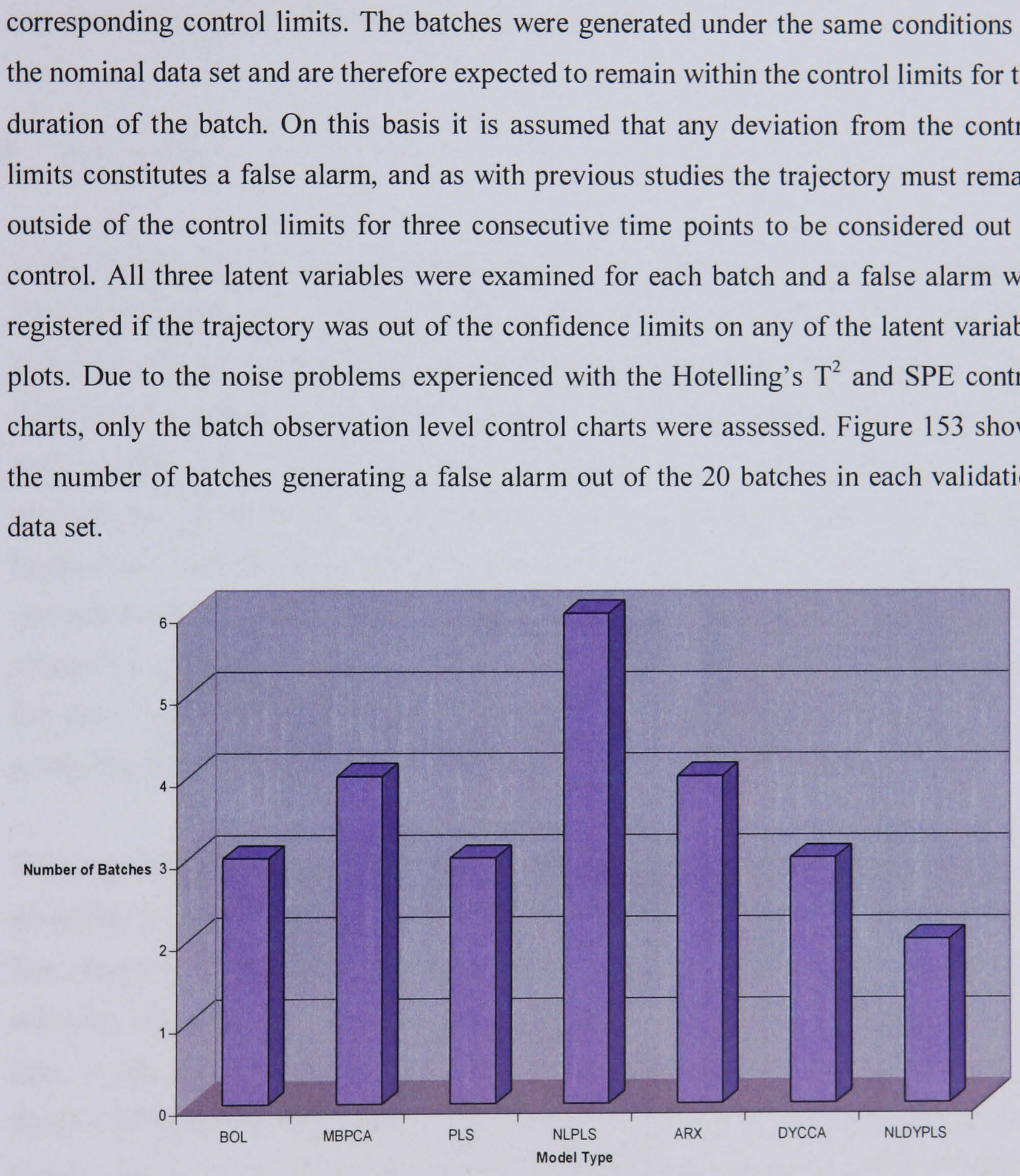


Figure 153: False alarm rates for each monitoring technique

Figure 153 shows that there is still a fairly high false alarm rate for each of the techniques studied. 99% of the batches should be in control in the data set, meaning only

1 batch should be out of statistical control. However this is not the case for any of the monitoring methods. This is likely to be due to the high level of noise in the data, meaning that lots of peaks occur in the control charts which stay out of control for more than 3 consecutive observations.

Of the monitoring techniques tested, the super model-based techniques did not perform significantly better than the normal and standard model-based techniques, and SMBPCA using non-linear PLS performed far worse. The best results are given by the technique using dynamic non-linear PLS, with a false alarm percentage rate of 5%, compared to non-linear PLS, which gave a false alarm rate of 24%.

7.8 Conclusions

In this chapter five different variations of the super model-based monitoring technique were investigated and compared to the standard batch observation level monitoring technique and the standard model-based PCA technique examined in Chapter 4. Five different error models were used with the super model-based approach to determine the most suitable approach for removing the dynamic and non-linear structure in the residuals, and therefore allowing a linear multivariate statistical approach to be applied to the data which should in turn give improved fault detection results over applying a standard batch monitoring technique, such as batch observation level analysis, to the original set of structured batch data. The objective of this case study was to determine if the super model-based monitoring techniques do indeed provide a better tool for monitoring batch processes than the standard batch monitoring technique.

The assessment of the batch monitoring techniques involved three factors, analysis of the model-based residuals, the fault detection ability and the false alarm rate of each method. The objective of the residual analysis was to determine if the super model-based approach was successful at removing the non-linear and dynamic properties from the data. In the model types studied, normal probability plots of the residuals and the Shapiro-Wilks W test were used to show that the non-linearity in the data was removed to some degree by the super model-based techniques. In general, the probability plots of the super model-based residuals all demonstrated improvements in linearity in comparison to the original data from the standard batch monitoring technique. As part of the residuals analysis, assessment was made as to whether the techniques were effective in removing the dynamic structure in the residuals. This was determined through the use

of partial autocorrelation plots of the residuals. Some removal of dynamic structure was seen in all of the super model-based approaches, however the dynamic CCA and non-linear PLS models gave the greatest improvements, with structure only remaining in one residual variable.

The objective of the super model-based techniques was to remove the non-linear structure and dynamics from batch data to provide a more effective monitoring tool. Therefore the next stage of the study was to see if the removal of these factors had an impact in terms of performance monitoring ability. To assess the fault detection ability of the techniques three different fault types were considered, a temperature sensor fault, fouling fault and valve problem, and tested on each of the model-based methods. The results generated in Chapter 6, where the same faults were investigated using the standard batch observation level monitoring technique and the standard model-based PCA approach, were also included for comparison.

Batch observation level scores control charts, Hotelling's T^2 control charts and SPE charts were used to calculate how long it took each method to detect any process abnormalities, as an average of the 50 batches included in each set of fault data, for each type of control chart. This gave the out of control average run length (ARL) which is summarised in Figure 150, Figure 151 and Figure 152. As discussed previously, the application of the super model-based techniques generates a set of noisy residuals, possibly due to overfitting of the model because once the super model-based technique had been applied, only a limited number of latent variables were needed to capture the majority of the variation in the process. This means that the Hotelling's T^2 and SPE metrics were difficult to interpret and a high degree of confidence is not given to the metrics calculated from these charts. However, detection of faults via the batch observation level scores charts was easier, and the results showed SMBPCA with non-linear dynamic PLS and with dynamic CCA gave the most improvement in terms of fault detection times in comparison with the standard batch monitoring technique. Some of the super model-based approaches did not show a reduction in fault detection times in comparison to the standard batch observation level technique, this is because the error models used, for example non-linear PLS and ARX time series models, only dealt with one of the key characteristics of batch processes i.e. non-linear or dynamic structure, and not both.

The final aspect of performance investigated in the case study was the false alarm rate generated by each technique, assessing the number of 'good' batches that the monitoring

techniques defined as being out of control at some point during the process. An overly sensitive monitoring method will cause too many false alarms which could lead to an operator ignoring important warnings when an actual fault occurs. Consequently, a technique which is not sufficiently sensitive will only identify process abnormalities once they have become more significant, and possibly harder to rectify. The false alarm rates were tested using a set of 20 validation batches, generated under normal operating conditions. The techniques tested did show that SMBPCA with Dynamic Non-linear PLS halved the false alarm rate of the standard batch monitoring technique and the simple model-based method, and SMBPCA with Dynamic CCA also showed significant improvements. However, the false alarm rates for all the model-based approaches tested were too high. This is due to the large amount of noise present in the residuals causing the trajectories to move in and out of the control limits regularly. This is something which is discussed further when looking at the robustness of the technique in Chapter 8 and in the future work section in Chapter 9 .

In conclusion, the testing carried out so far on the super model-based batch monitoring techniques shows that SMBPCA gives noticeable reductions in the dynamic and non-linear structure found in a batch process. As a result of this, these techniques are able to give improved fault detection and false alarm rates compared with the standard batch monitoring techniques. Of all the model types tested on this exothermic batch reaction, the best results were demonstrated by the SMBPCA with Dynamic CCA and SMBPCA with Dynamic Non-linear PLS techniques. The main drawback associated with the model-based techniques is the increased noise generated by application of the approach, suspected to be due to overfitting of the models, this is something which needs to be investigated further.

8 CHAPTER EIGHT: ROBUSTNESS OF SUPER MODEL-BASED PCA

8.1 Introduction

As has been mentioned previously, in Chapter 4, there are a number of issues associated with the development of a perfect mechanistic model of an industrial process. These include the fact that the accurate modelling of the many physical and chemical relationships occurring during a batch reaction can be a time-consuming and effort intensive task, and there is the possibility that a perfect model may not materialise since some physical, chemical or biological phenomena may not be fully understood or reproducible. Coupled with the large amount of time and computational effort required, model-based analysis maybe an infeasible choice for some processes, such as those that have a short product life cycle, for example due to the necessity of developing a new and complex model every time a different product is manufactured.

In this chapter some of the issues associated with applying the model-based and super model-based techniques in an industrial environment are examined. The focus of the investigation is the robustness of the technique, and to highlight areas where the technique fails and hence further research is required.

One of the main ideas behind super model-based PCA is that the use of an additional error model to model the residuals, enables the plant-model mismatch to be addressed without reducing the fault detection ability of the technique. However, in the case study undertaken in Chapter 7, the plant-model mismatch was relatively small. Hence it is necessary to expand the study to assess the impact of using a mechanistic model with uncertain parameters. This addresses the situation where the modeller cannot define them exactly, or where they vary, or where there are modelling errors present, for example an incorrect value has been assigned or a process condition has changed without the model being updated. A consequence of this is that if a signal exceeds the control limits it could be misinterpreted, that is it may be identified as a process fault but in practice it has been caused by a difference between the model and the plant.

Additionally, the study carried out, into the performance of the super model-based technique has focussed on a simulated environment with controlled levels of noise. The

levels of noise in an industrial environment are likely to be higher, therefore it is important to assess the potential effect this will have on the model-based technique, particularly as the application of the model-based technique has been observed to increase the levels of noise observed in the residuals.

8.2 Plant Model Mismatch

8.2.1 Introduction

Plant model mismatch is an issue that impacts on all model-based techniques. When an abnormality is detected by the SMBPCA control charts, or another MSPC control charts, the out of control signal could be caused either by a fault in the process or it could be due to plant-model mismatch. There are different reasons as to why mismatch between the model and plant can occur. It may be because the model is of reduced complexity or has been incorrectly modelled, or it may be because maintenance, or an upgrade, has been carried out on the plant without the modeller being informed. For example, a change to a valve, or pipework being replaced, may affect the characteristics of the process, for example resulting in different heat transfer coefficients, but this change is not reflected in the mechanistic model.

In process performance monitoring, it is important to determine whether an abnormality detected by the monitoring software is a fault on the plant or whether it is a mismatch between the plant and the model. This is particularly important with super model-based PCA, where the mechanistic model is critical to the success of the technique. Modelling errors can have a similar effect on the residuals, with the misdiagnoses of the mismatch as a fault resulting in inappropriate action being taken by the plant operators, leading to the process going further out of control. This first part of this chapter investigates methods of assessing whether an out of control signal is due to plant-model mismatch or due to an actual process fault. A further study is also carried out using the mechanistic model of the process where a plant feature has been changed. In this case the model mismatch was created by changing the volume of the reactor vessel. Data simulated from this modified model is then used with the original first principles model to investigate how the performance monitoring representation is affected and whether it is possible to identify the change as a mismatch as opposed to a process fault.

Identifying an abnormal signal as being a consequence of plant-model mismatch, as opposed to an actual fault, is a difficult task. Examination of the Squared Prediction Error (SPE) and Hotelling's T^2 statistic along with the batch observation level control charts may provide greater insight into the nature of the out of control signal. The squared prediction error (also known as the Q statistic and the DmodX statistic, distance to model) is the sum of the deviations between the predicted model values and the actual values. When the process is operating within its limits, the SPE statistic represents the variation in the data that is not captured by the model. Therefore when an unusual event, such as a fault or plant-model mismatch, occurs that introduces a change in the relationship between the variables, or the process mean, this will be indicated by a high value of the SPE.

Hotelling's T^2 statistic is effectively the distance from the data point to the centre of the model and therefore only detects variations in the data that are greater than the common cause variation already captured by the model. An out of control signal detected by the batch observation level control chart could be accompanied by both the SPE and Hotelling's T^2 statistics being outside their control limits, or by one of them exceeding the limits. If the SPE statistic moves outside the control limit then this indicates there has been a significant change in the correlation structure between the process variables. However, if only the Hotelling's T^2 statistic lies outwith the control limits, then this is an indication that the out of control signal has been caused by some kind of manipulation of the process parameters or that a change has occurred throughout the process. Hotelling's T^2 statistic will not identify a new type of special event that has not been included in the data used to build the nominal model, unlike the SPE statistic. Therefore it is possible that if the out of control signal detected from the BOL control chart is due to a plant model mismatch as opposed to an actual fault then it will be observed more clearly on the SPE chart as opposed to the Hotelling's T^2 control chart.

There are many different types of change detection algorithm, Basseville *et al* (1988), that could be applied to the problem. Typically they are based on the analysis of the residuals and/or the design of a decision-based system that monitors and responds to changes in the residuals. One approach to detecting model process mismatch is to use some type of knowledge based method such as an expert system, Leung *et al* (2000), where a database of known faults is generated. A pattern recognition technique, Yoon *et al* (2001), can then be used to classify the signature of the fault based on either a priori knowledge of the process or on statistical information extracted from the fault data, through the application of a neural network or a clustering based method. This

information is then used in conjunction with the database to determine if the out of control signal has a fault signature comparable to one in the database. If the fault signature does not match any of the existing signatures, the abnormality is either due to plant-model mismatch or is a completely new fault, not previously seen. The difficulty with this technique is that it requires comprehensive historical knowledge of all faults that could occur on the plant, or an accurate simulation of the process, and as discussed before this is not always possible.

There are a number of ways in which to build a fault signature database for use with a pattern recognition technique. Yoon *et al* (2001) proposed a steady state fault signature methodology which used an angle measure between the known fault signature and the new measurement vector signature to give a relative measure of the correlation between the two. An alternative possibility is to develop the SPE contribution plot for all the known faults, Yoon *et al* (2001). A pattern recognition technique is then used to monitor an evolving time series of the SPE contributions as it moved out of control and a rolling average is calculated. The average contributions from the new 'out of control' batch are then examined to determine if they were associated with a particular fault type from the historical database.

An alternative method that could be used to examine the fault data for plant model mismatch is that of variable reconstruction, Dunia *et al* (1996), Qin *et al* (2000), Harkat *et al* (2006). This is a technique that has been applied for the identification of faulty sensors. The method involves sequentially reconstructing the process variables from the PCA process representation, effectively estimating a process variable from the other process variables using the PCA model, although it can also be applied with other modelling techniques such as partial least squares (PLS), Wise *et al* (1991) and neural networks, Kramer (1991). By comparing the reconstructed variables with the original process measurements, a residual analysis is carried out, to identify the source of the difference. When this technique has been applied in sensor fault detection, a sensitivity index has been defined whereby if the residuals exceed the limit, the variable is assumed to be faulty, Dunia *et al* (1996).

8.2.2 Example of Plant Model Mismatch

The effect of plant-model mismatch on the super model-based technique is examined by modifying the process simulation to mimic a physical change on the plant. The

mechanistic model used in the super model-based technique was not changed, hence there was a 'real' mismatch between the plant and the model. The assumption is that the change has been carried out on the plant without the model being updated. There are a number of changes that could occur which would cause this mismatch, for example, the split range of the cooling valve could be altered, this would have a significant effect as the mechanistic model includes a control algorithm for the valve based on the split or alternatively the tuning of the control valve has been optimised and the gain changed. An alternative option is that the cooling water supply increases during the summer months, again making the model inaccurate as the heat balances would be using the incorrect cooling water temperature. Similarly there could be a change in the supply pressure from the steam main which would generate a difference between the model and the plant data. Another change that would have an effect on the plant and model would be the replacement of the reactor vessel with one of slightly different dimensions or fabricated from different materials. For this study, the volume of the reactor was changed to create the plant model mismatch.

The mechanistic model equations from Chapter 3 are shown in Equations (8.1)-(8.5).

$dC_a / dt = \kappa_1 * C_a$	(8.1)
$dC_b / dt = \kappa_1 * C_a - \kappa_2 * C_b$	(8.2)
$dT/ dt = [k_1 * C_a * \lambda_1 + \kappa_2 * C_b * \lambda_2/\rho * C_p + h_1 * (T_m - T)$	(8.3)
$dT_j/ dt = \varphi * [U_h * (T_h - T_j) + U_c * (T_c - T_j)] - (h_2 * (T_j - T_m))/ \theta_2$	(8.4)
$dT_m/ dt = [h_2 * (T_j - T_m) - h_1 * (T_m - T)]/ \theta_1$	(8.5)

The volume is represented in the model by Equations (8.6)-(8.9).

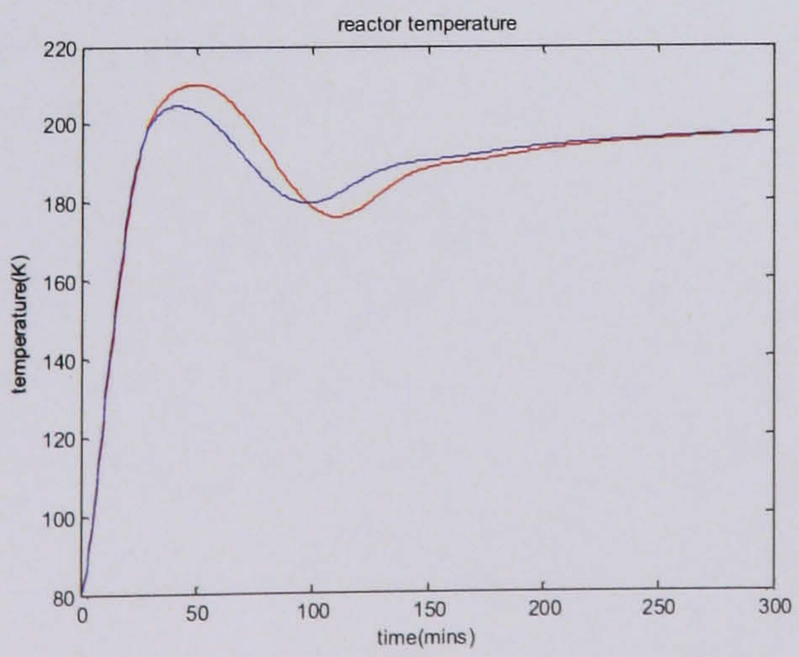
$\theta_1 = \frac{V_m \rho_m c_m}{V \rho C_p}$	(8.6)
$\theta_2 = \frac{V_i \rho_i c_i}{V \rho C_p}$	(8.7)
$h_1 = \frac{h_i A_i}{V \rho C_p}$	(8.8)
$h_2 = \frac{h_{ow} A_o}{V \rho C_p}$	(8.9)

The nominal values are shown in Table 13. To simulate the change in the volume of the reactor, the nominal values were increased by 5%.

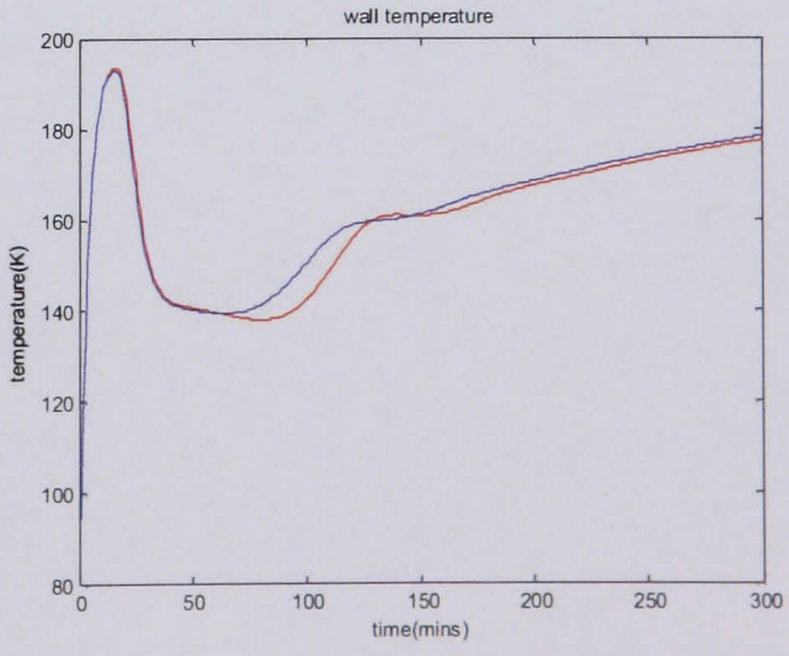
Parameter	Nominal Value	New Value
θ_1	0.272	0.2856
θ_2	0.552	0.57960
h1	0.071	0.07455
h2	0.177	0.18585

Table 13: Parameter changes to simulate plant model mismatch

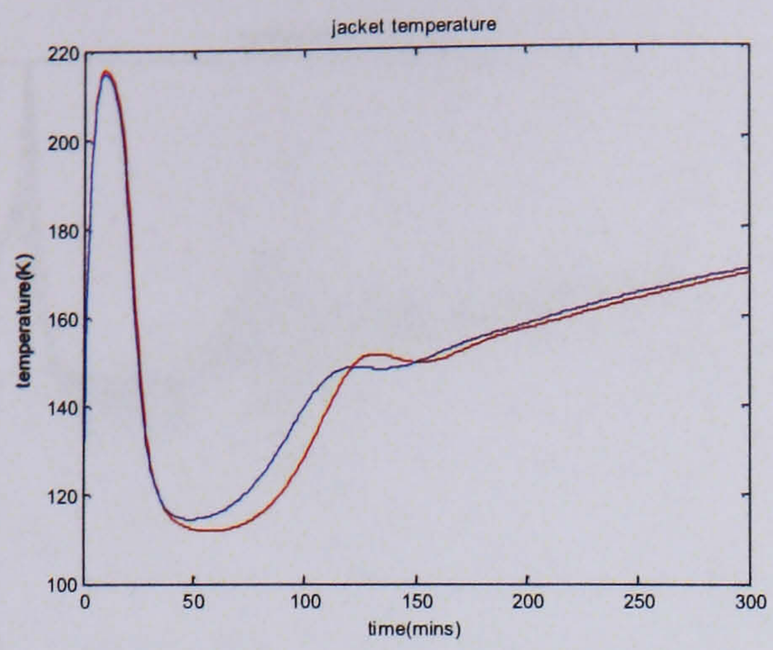
The impact of this change is demonstrated in the time series trajectories of the four variables in Figure 154. The original data is marked in red with the data containing the parameter changes given in blue.



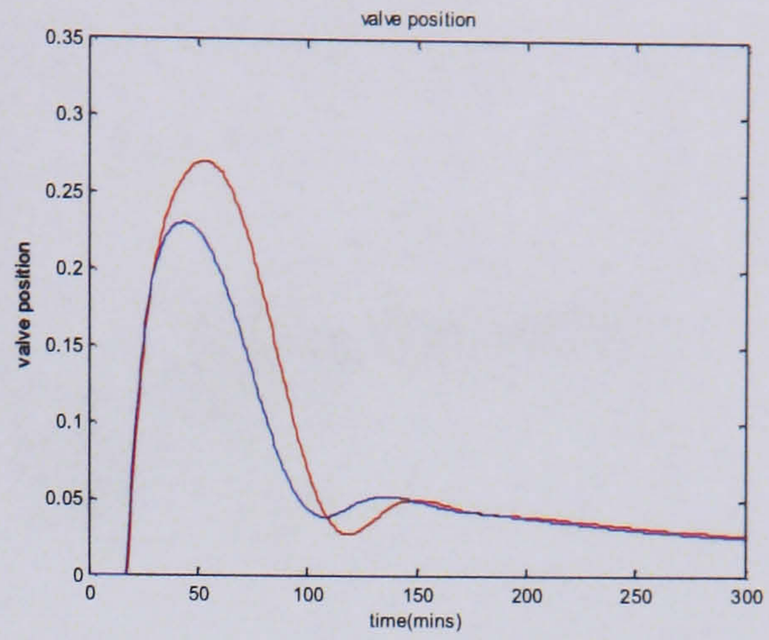
a) Reactor temperature



b) Wall temperature



c) Jacket temperature

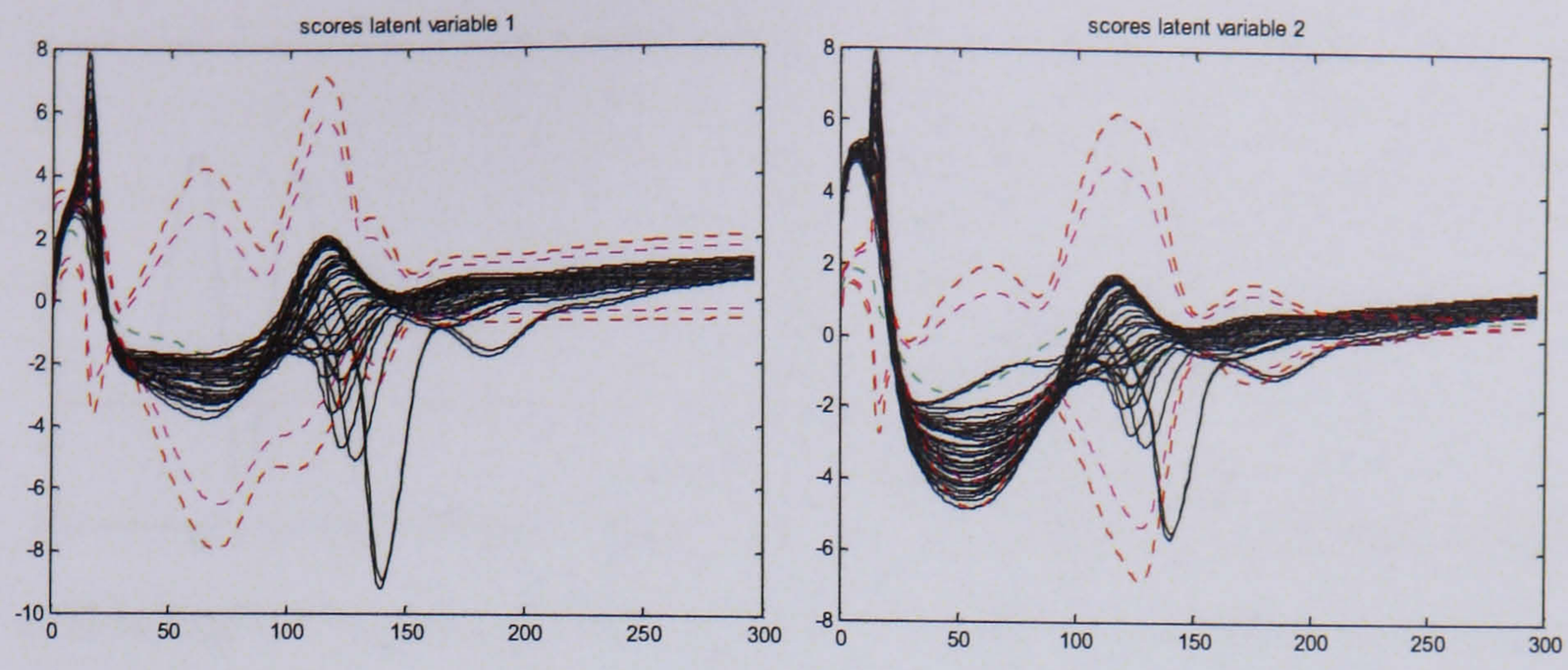


d) Valve position

Figure 154: Variable plots incorporating the parameter changes

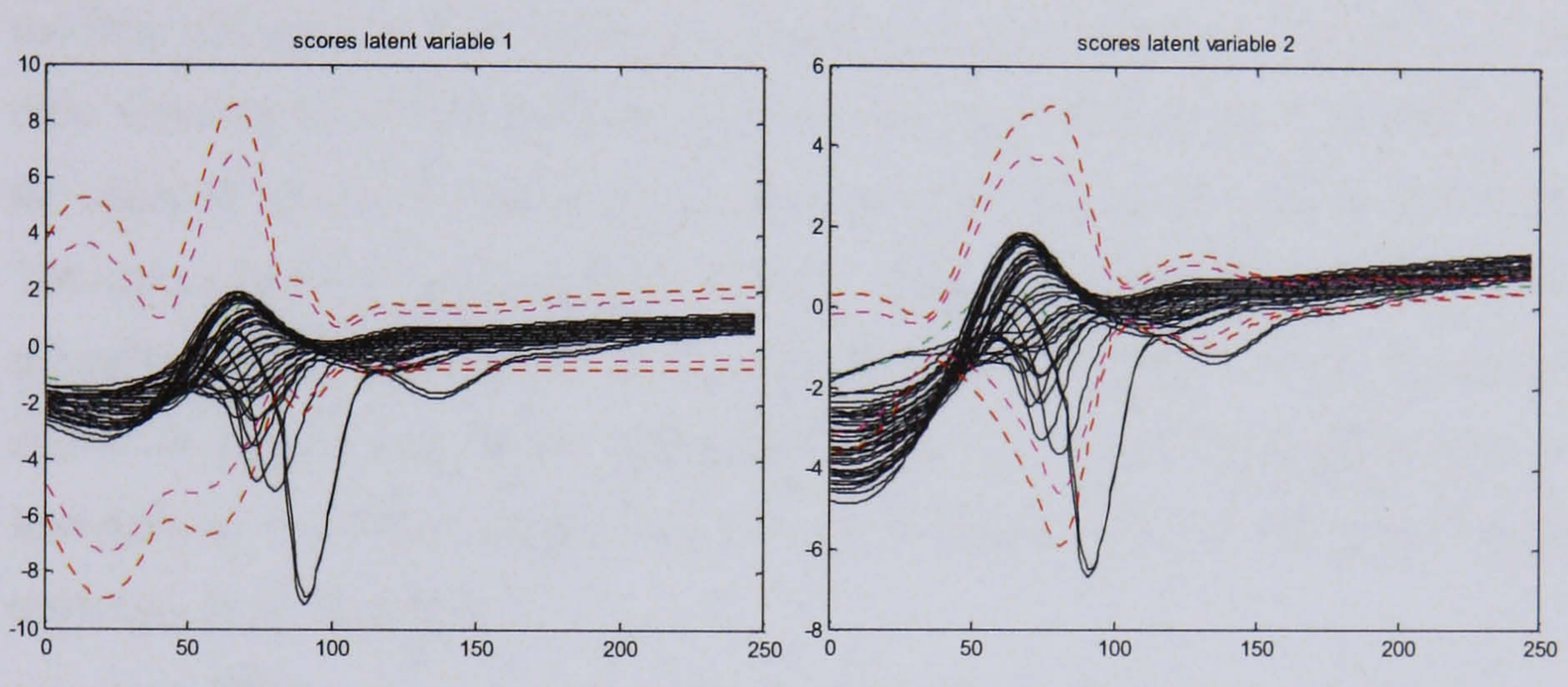
A new data set comprising fifty batches was generated incorporating the reactor volume change. The mechanistic model remained unchanged. The SMBPCA technique was then applied to the new set of data. The contribution analysis approach was used to investigate the new data set and the results were compared with the previously tested faults to see if there were any significant differences in the fault signatures. The SPE and Hotelling's T^2 statistics were also examined along with the batch observation level scores control charts.

The first 50 data points were removed prior to modelling since this is the time during which the batch is brought into control. However, plots of the variables including these points are included. Figure 155 shows the control charts for the super model-based PCA methodology with dynamic canonical correlation analysis applied, to model the structured residuals, for the set of plant model mismatch data, prior to the removal of the first 50 minutes, i.e. the batch start up is included. Figure 156 shows the control charts after the reaction start up section of the batch has been removed. The control limits are the same in both cases.



a) Latent Variable 1 b) Latent Variable 2

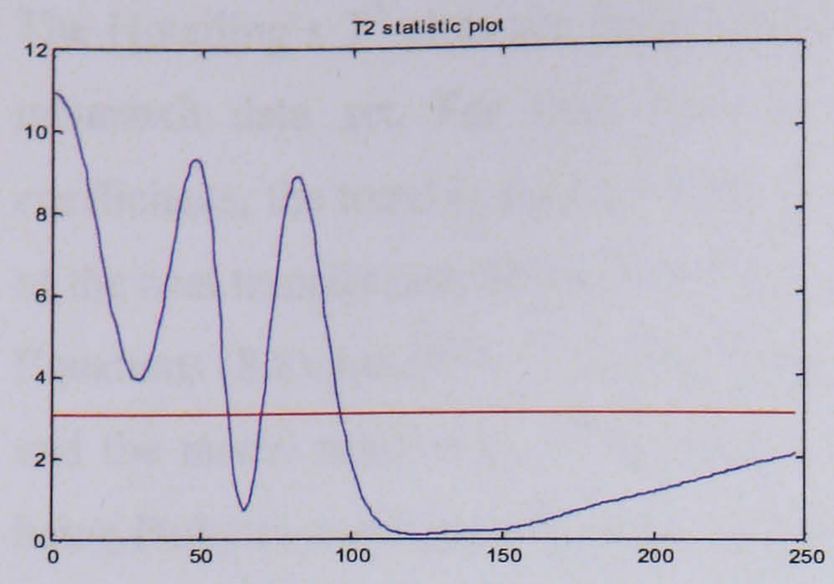
Figure 155 : Process model mismatch – Before batch start-up removed



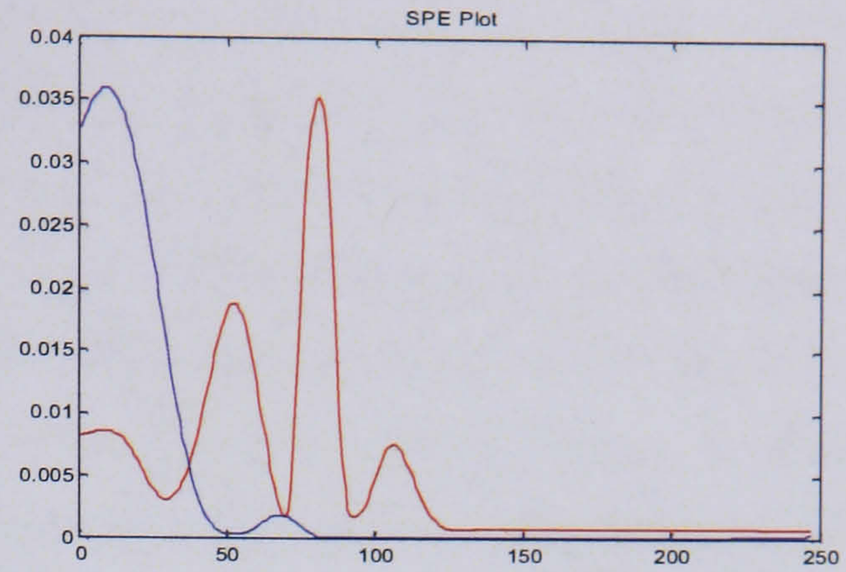
a) Latent Variable 1 b) Latent Variable 2

Figure 156 : Process model mismatch – After batch start-up removed

The control charts in Figure 155 clearly show the batches exceed the control limits as soon as the batch starts, after this the majority of the batches return to within the control limits for a short period of time before deviating outside the limits for the remainder of the batch. The change in the reactor vessel volume appears to have greatest influence at the start of the batch and not all the way through as would have been observed if a constant mismatch had materialised. Two of the batches can be observed to move more distinctly outside of the control limits in comparison to the other batches, this is due to the random noise added to the batches. The control charts in Figure 156 show what would be seen by the operator. The effect of the mismatch is not as apparent. The SPE and Hotelling's T^2 plots are shown in Figure 157 (limit shown in lighter red colour, trajectory shown in darker blue colour), only one batch from the data set is shown.



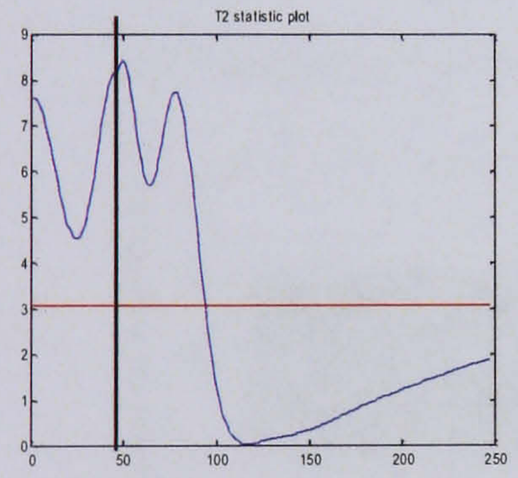
a) Hotelling's T^2 for model mismatch batch



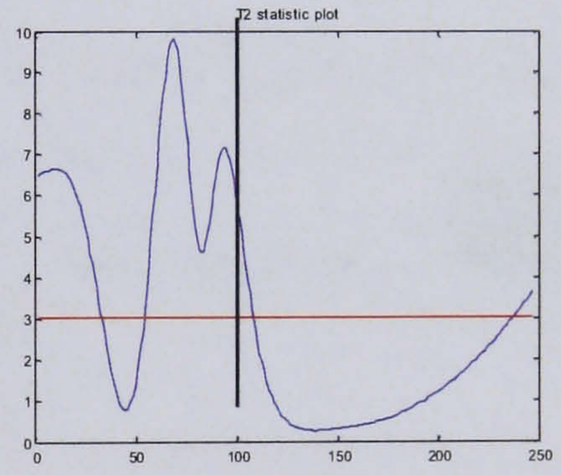
b) SPE for model mismatch batch

Figure 157 : Hotelling's T^2 and SPE plots for plant model mismatch data set

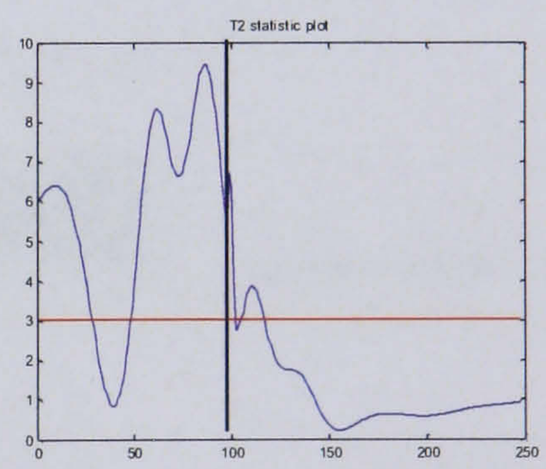
Both the Hotelling's T^2 and SPE statistics show the batch as being out of control during the first 100 minutes with the trajectories returning to inside the control limits after this time. Consequently since both statistics exceed the control limits, it is difficult to assign the cause of the out of control signal as being due to a fault or to plant model mismatch. The same plots were generated for the three faults considered in Chapter 6 to investigate the performance of SMBPCA. The results are shown in Figure 158 and Figure 159 (limit shown in lighter red colour, trajectory shown in darker blue colour, point of fault introduction in black). Again only 1 batch is shown for each fault type to allow the trajectory to be clearly seen.



a) Fault Type 1

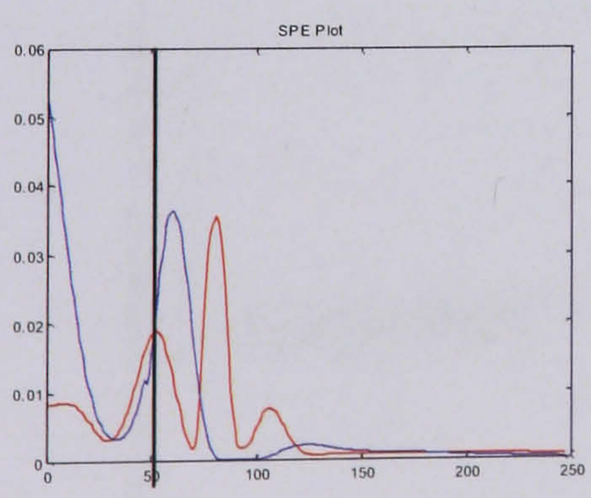


b) Fault Type 2

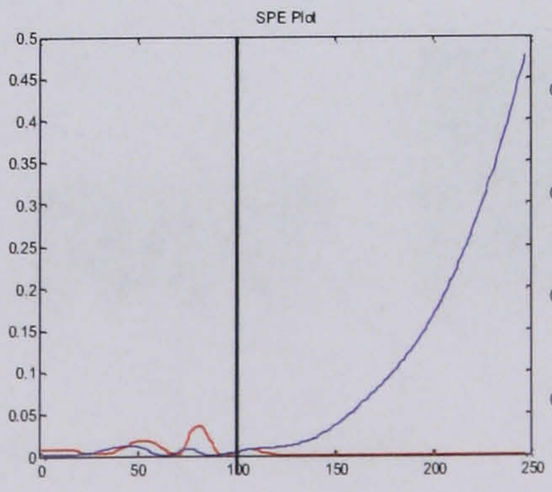


c) Fault Type 3

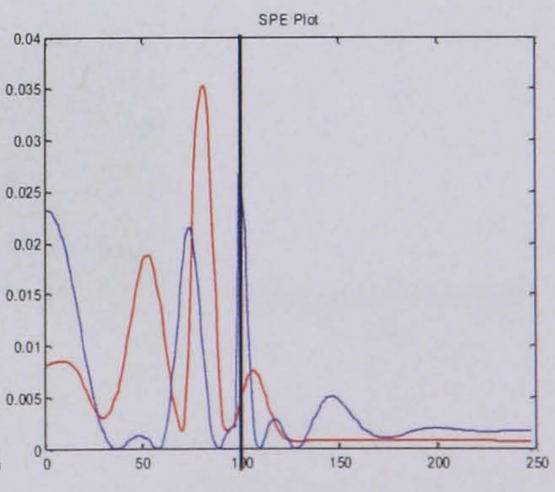
Figure 158 : Hotelling's T^2 for the 3 different fault types



a) Fault Type 1



b) Fault Type 2

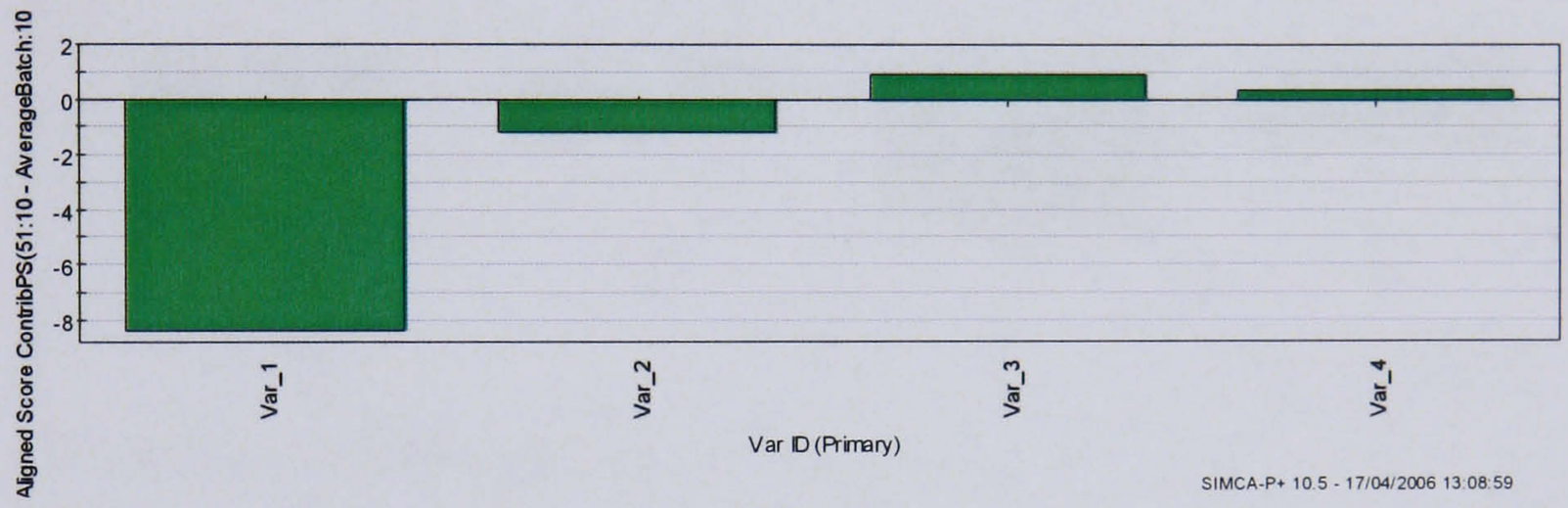


c) Fault Type 3

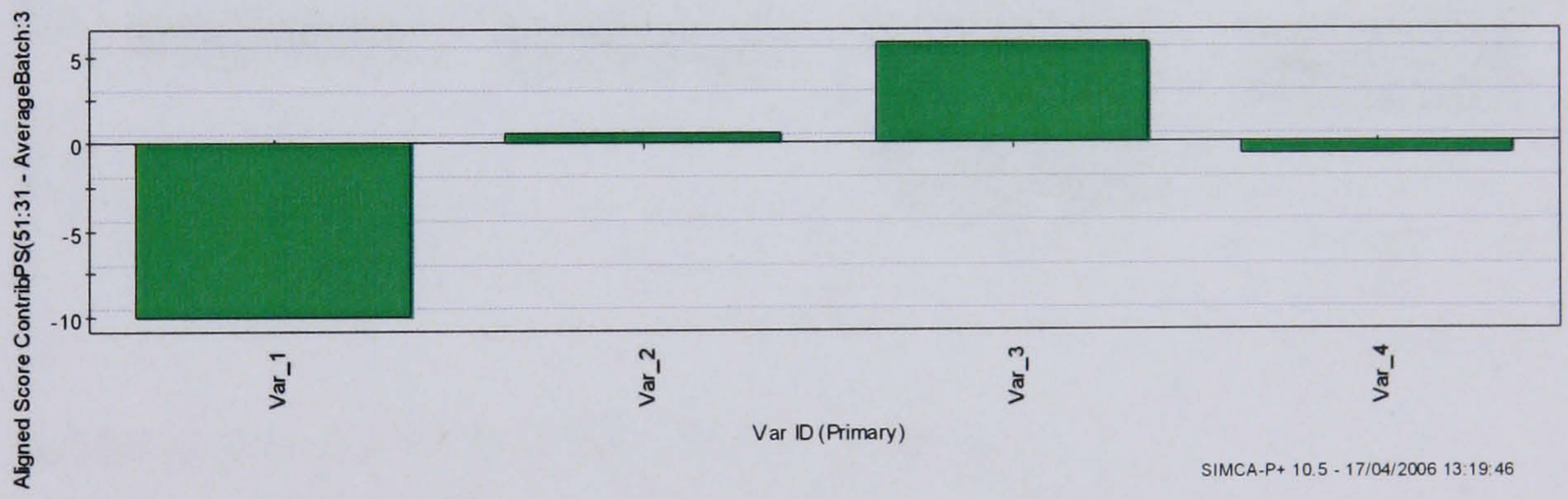
Figure 159 : SPE for the 3 different fault types used

The Hotelling's T^2 plots are fairly similar for all three fault types and the plant model mismatch data set. For fault type 2, which was a decrease in the heat transfer coefficients, the trend is similar to the plant model mismatch batch since the calculation of the heat transfer coefficients are affected by the volume of the reactor, as described in Equations (8.8) and (8.9). The SPE plots are more distinctive for each of the fault types and the model mismatch. Using these it may be possible to classify batches as either being faulty or containing plant-model mismatch if enough historical data on the process had been gathered.

The contribution plots, Figure 160 and Figure 161, were investigated to determine whether a fault or plant-model mismatch had occurred. Fault type 2, the change in the heat transfer coefficients was considered as it exhibits behaviour that is most close to the plant model mismatch data set. This is reflected in the Hotelling's T^2 plots being similar.



a) Contribution plots for batch containing mismatch at time point 10

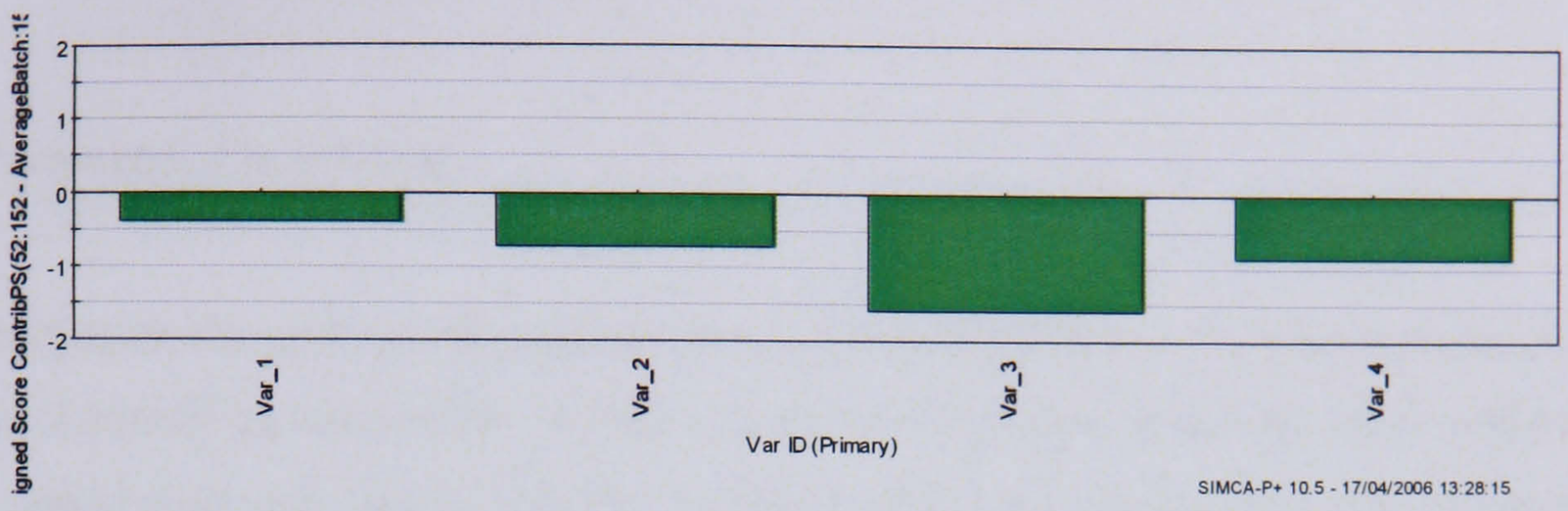


b) Contribution plots for batch containing mismatch at time point 31

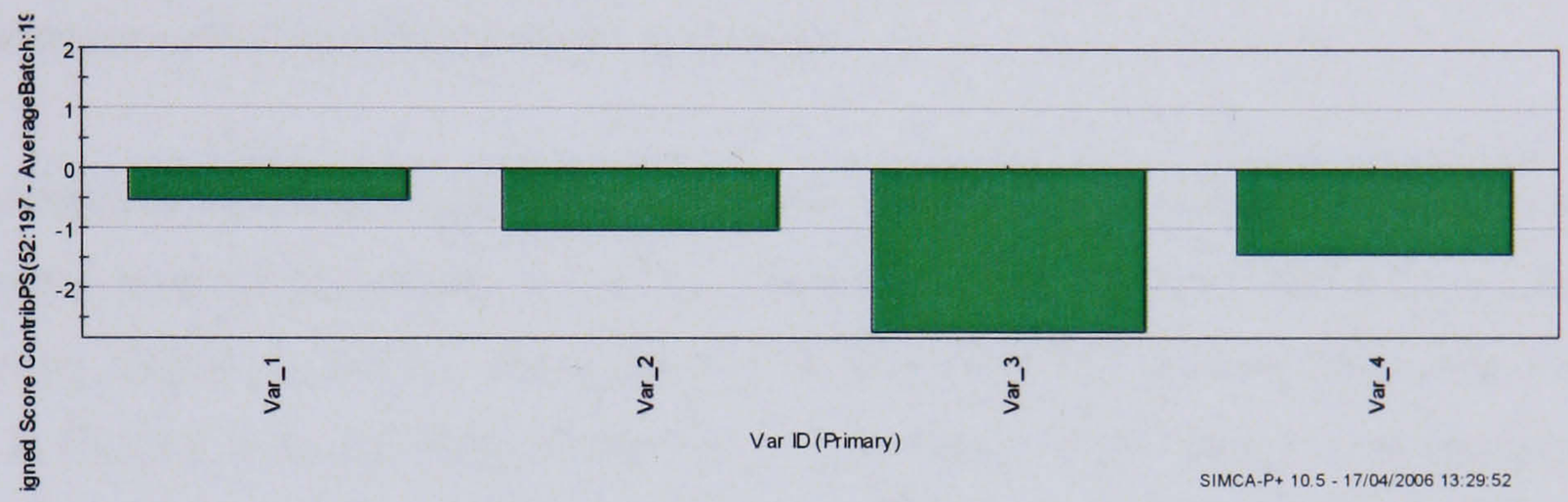
c) Contribution plots for batch containing mismatch at time point 50

Figure 160 : Contribution plots for batch as batch trajectory exceeds limits

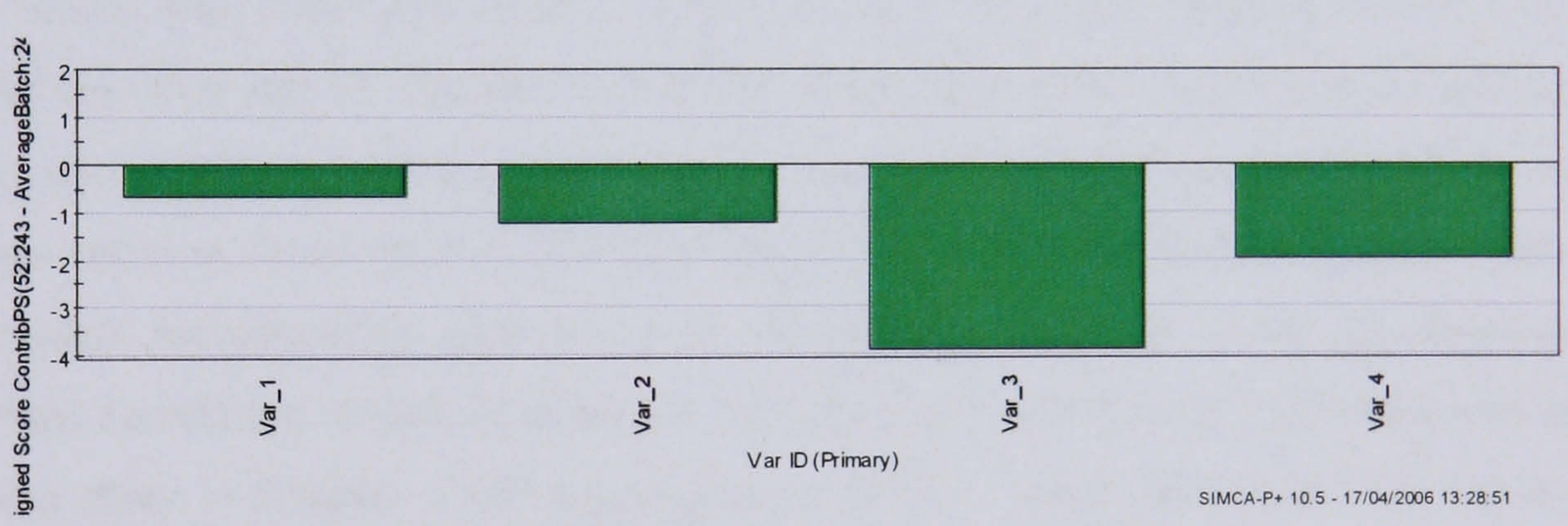
The contribution plots in Figure 160 show that variable 1, the reactor temperature, is responsible for the batch initially exceeding the control limits, and as the batch evolves it is variables 2 and 3, the jacket and metal temperatures that have an increasing effect.



a) Contribution plot for fault type 2 at time point 152



b) Contribution plot for fault type 2 at time point 197



c) Contribution plot for fault type 2 at time point 243

Figure 161 : Contribution plots for fault type 2 as fault evolves

The contribution plots for fault type 2, Figure 161, where there is a decrease in the heat transfer coefficients are different to those shown for the plant model mismatch, even though they both contain changes to the heat transfer coefficients. As fault type 2 evolves it is variable 3, jacket temperature, that consistently has the greatest effect on the batch exceeding the control limits, variables 1 and 2, reactor and wall temperature, which were most important in the batch containing the plant model mismatch, are less important. Although this is a simplistic approach to examining differences between the fault and the plant-model mismatch it has been shown to be effective for this case

because different patterns are apparent in the contribution plots for the fault and for the mismatch. It would be possible to build up a database of contribution plot fault signatures which could then be used as a reference for comparison when a batch exceeds its control limits. This is an area that requires further research.

8.3 Parametric Uncertainty

A variation on the problem of plant-model mismatch, discussed in the previous section, is that of parametric uncertainty, Rotem *et al* (2000). For some processes, there will be some uncertainty associated with the modelled parameters, particularly for complex processes. Therefore it is important to investigate how this uncertainty impacts on the performance of the super model-based technique.

To investigate this issue, a number of mechanistic models were generated, each one with increasing levels of uncertainty in specific parameters. The parameters selected were the frequency factors, α_1 and α_2, which are used in the Arrheinus equations, equations 5.6 and in Chapter 5, to calculate the reaction rate constants. These factors were changed from their nominal values of 729 and 6567.7 by 0.1%, 0,5%, 1%, 2%, 5% and 10%. Each model was then implemented in conjunction with super model-based PCA to analyse the same sets of simulated batch data, considered in previous chapters, and the impact on the fault detection ability of the method was investigated. Figure 162 shows the time taken to detect the first abnormal sample for each of the three fault types using the models incorporating each level of uncertainty. SMBPCA with the dynamic canonical correlation structured residuals analysis was selected as it performed well in the case study in Chapter 4 and computational effort is significantly less than for the dynamic non-linear PLS model, which gave the overall best results.

Faults 1 (sensor fault) and 2 (heat transfer coefficient decrease), Figure 162, both behave in a similar way, that is as the level of uncertainty in the parameters increases, the time taken to detect the first abnormal sample also increases before it reaches a plateau. A sample is detected to be out of statistical control when three consecutive samples have moved outside the control limit. For fault type 3, the control valve fault, the time taken to detect the problem increases with increasing uncertainty with the fault not being detected once the noise level is 10%. As the uncertainty in the data increases the control limits increase in magnitude, making the detection of abnormalities more difficult. From Figure 162 it can be observed that the ability to diagnose a cooling valve fault, fault 3, in

the exothermic batch reactor is more dependent on the uncertainty in the model than the model's ability to detect a fault with a temperature sensor, fault 1, or reactor fouling, fault 2. As the uncertainly level in the model increases beyond 6%, the cooling valve fault can no longer be detected, whereas the time taken to detect the first abnormal sample for both the reactor temperature sensor fault and the reactor fouling fault reach a plateau and do not increase further as the uncertainty increases. However, all three are significantly affected by the presence of parametric uncertainty.

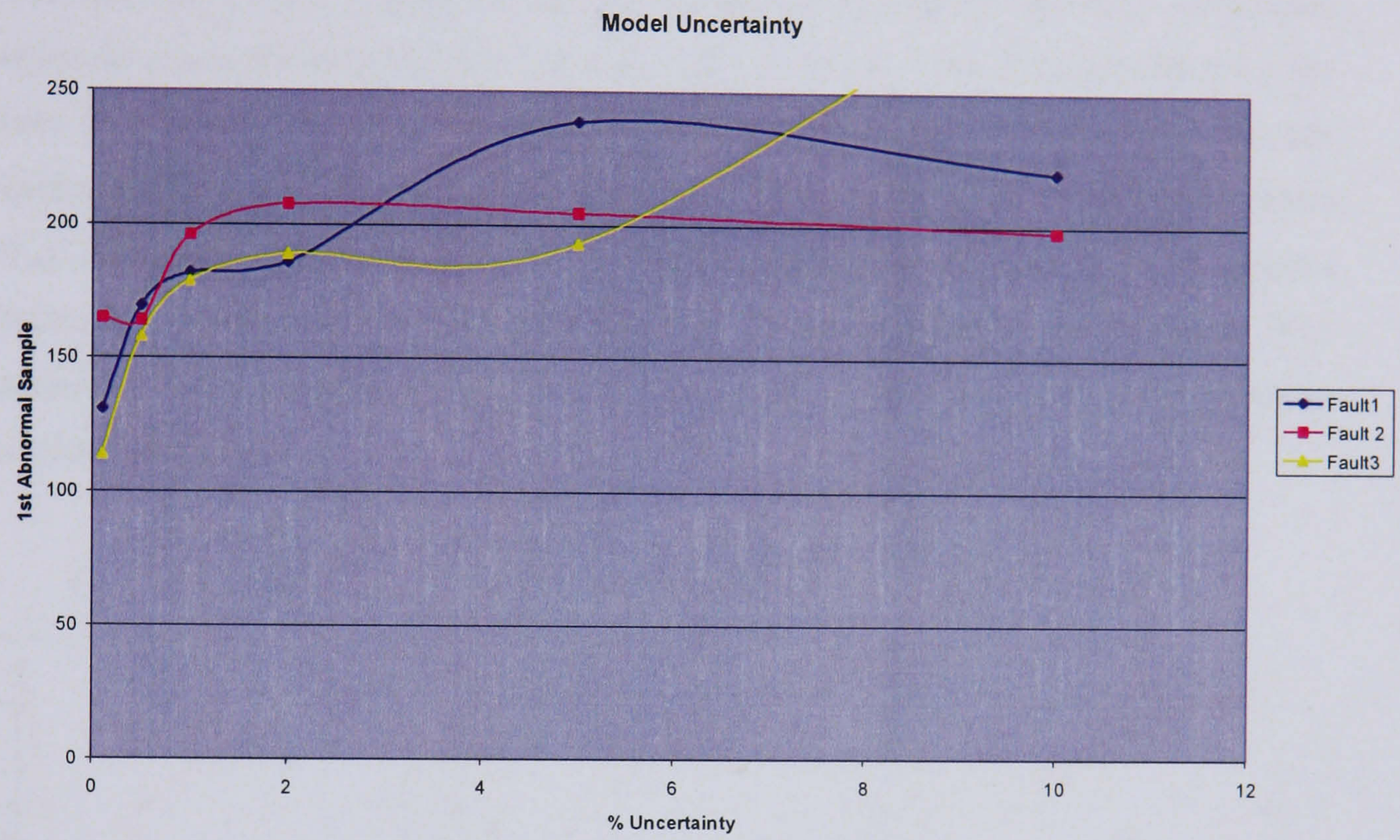


Figure 162 : Time for first abnormal sample versus level of uncertainty for the 3 fault types

8.4 Sensitivity to Noise

The main idea behind model-based PCA is the removal of the non-linear and dynamic behaviour inherent in the data using a mechanistic model, thereby generating a set of residuals that reflect white noise. Abnormalities in the process will create a different residual pattern to white noise and this change should be detectable using PCA. However, if the white noise in the process data has a high noise to signal ratio, particularly compared to a small abnormality in the process then it is unlikely that the technique would be able to detect the change in the process conditions. This is especially the case with industrial data as it often contains a high level of noise, therefore it is important to evaluate how robust the super model-based technique is, with respect to the

measurement noise, and also to determine a threshold signal to noise ratio where the technique fails to identify process faults.

To investigate the sensitivity of the technique to noise, which is defined as the ratio of the standard deviation of the simulated noise signal over the standard deviation of the original measured variable, a case study similar to the one carried out for model uncertainty in Section 8.3 was undertaken. The noise levels in the simulation of the exothermic batch reactor were modified, i.e. noise to signal ratios of 1%, 5%, 10% ,25%, 50% and 100% were considered, and 50 batches were generated for each noise level which served as the nominal data set and a further 50 data sets were generated for the three fault types. The parameters selected to indicate the measurement noise were the reactor temperature (T), the jacket temperature (T_J) and the reactor wall temperature (T_m). Figure 163 shows the plots of the reactor temperature variable (first 50 samples removed) for each of the 6 increasing noise levels. The super model-based PCA technique using dynamic CCA was then applied to the data using the standard mechanistic model.

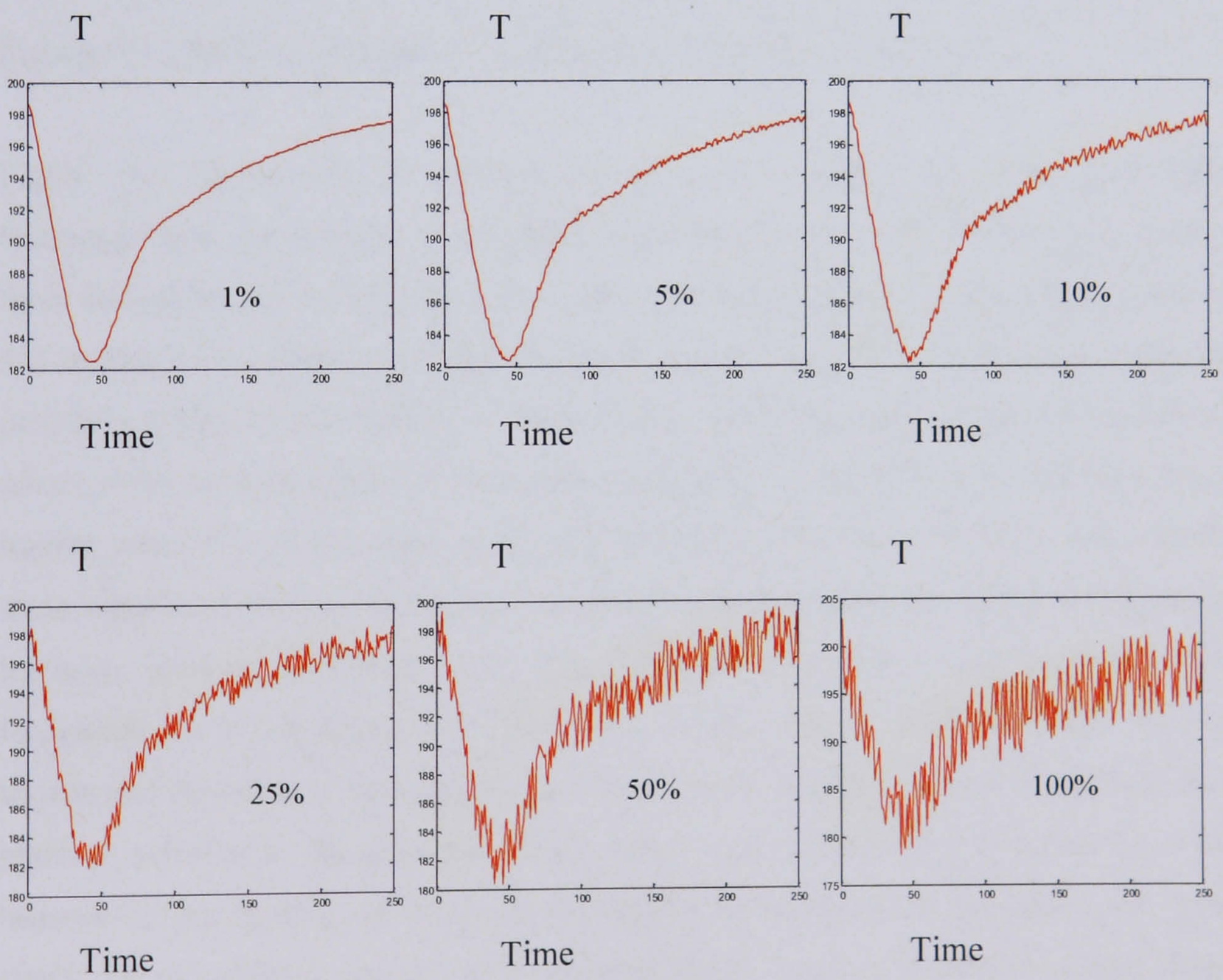


Figure 163 : Plots of increasing noise levels for reactor temperature variable

To assess the performance of the SMBPCA technique, the first abnormal sample for each fault type detected (as long as the batch remained outside the limits for three consecutive time points) by the technique was recorded for each of the different noise levels, Figure 164.

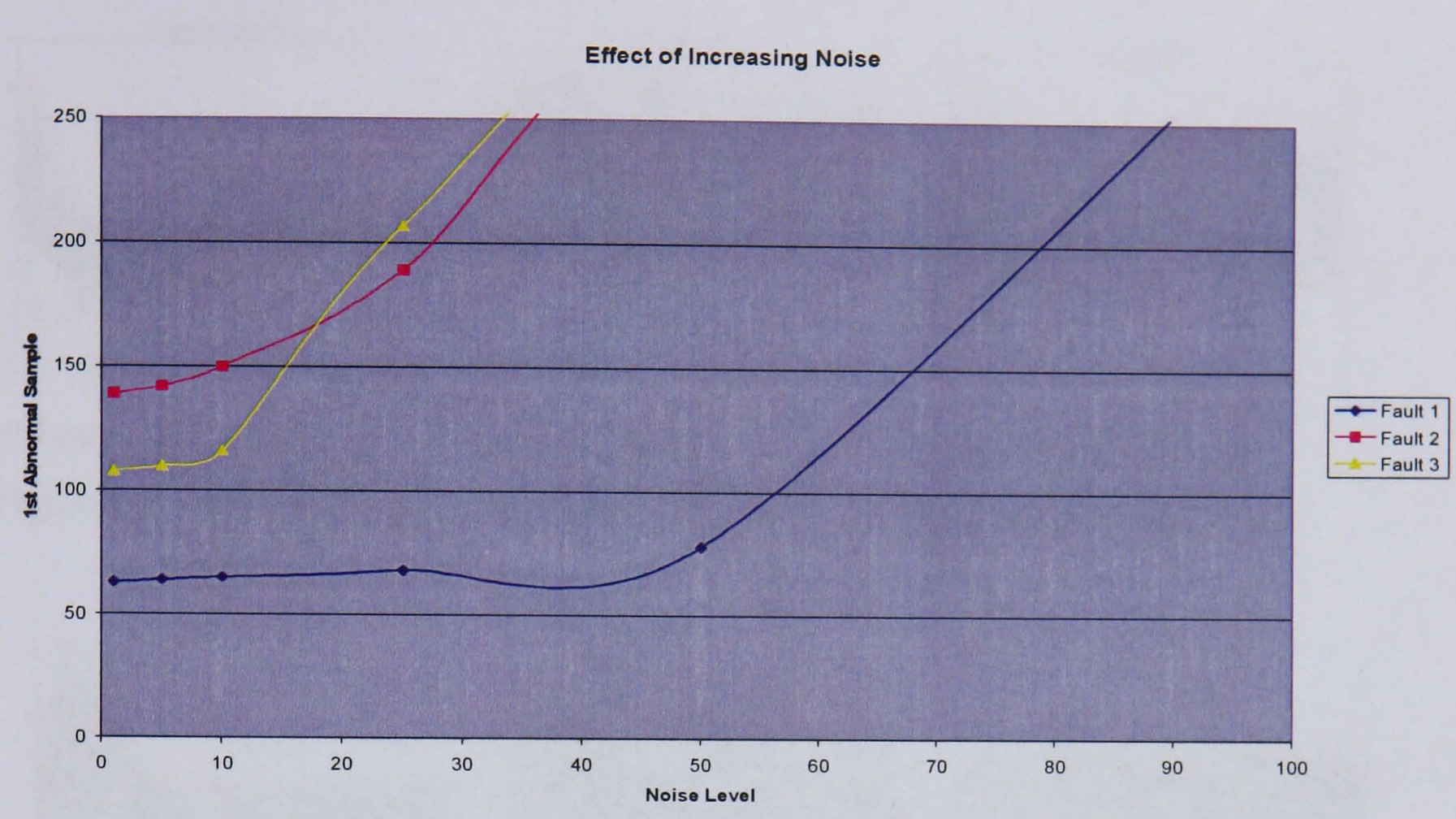
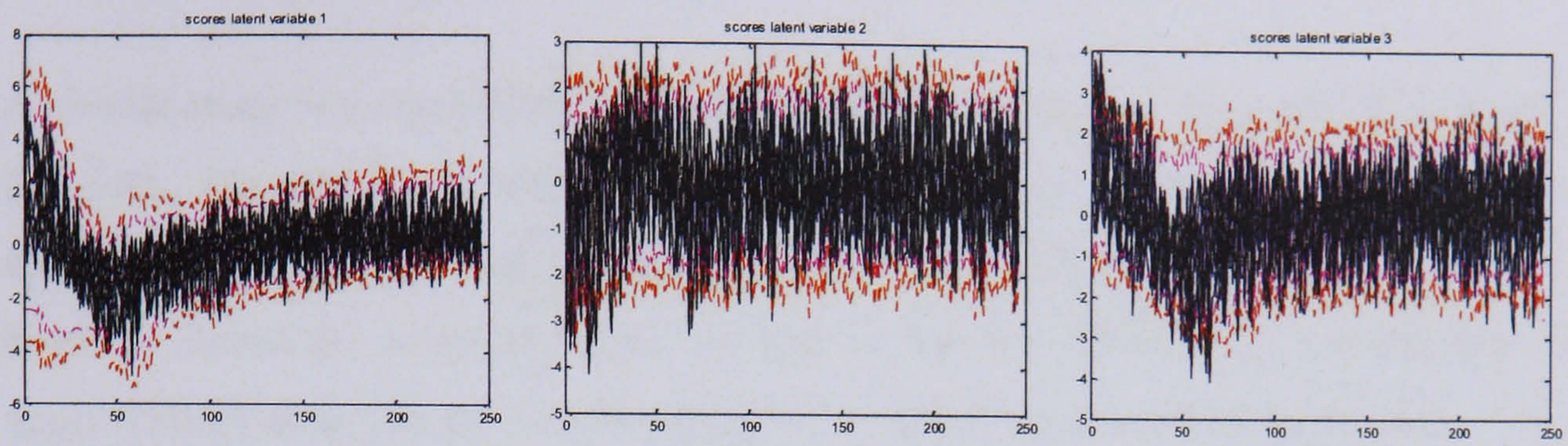


Figure 164 : Effect of increasing noise level on SMBPCA fault detection

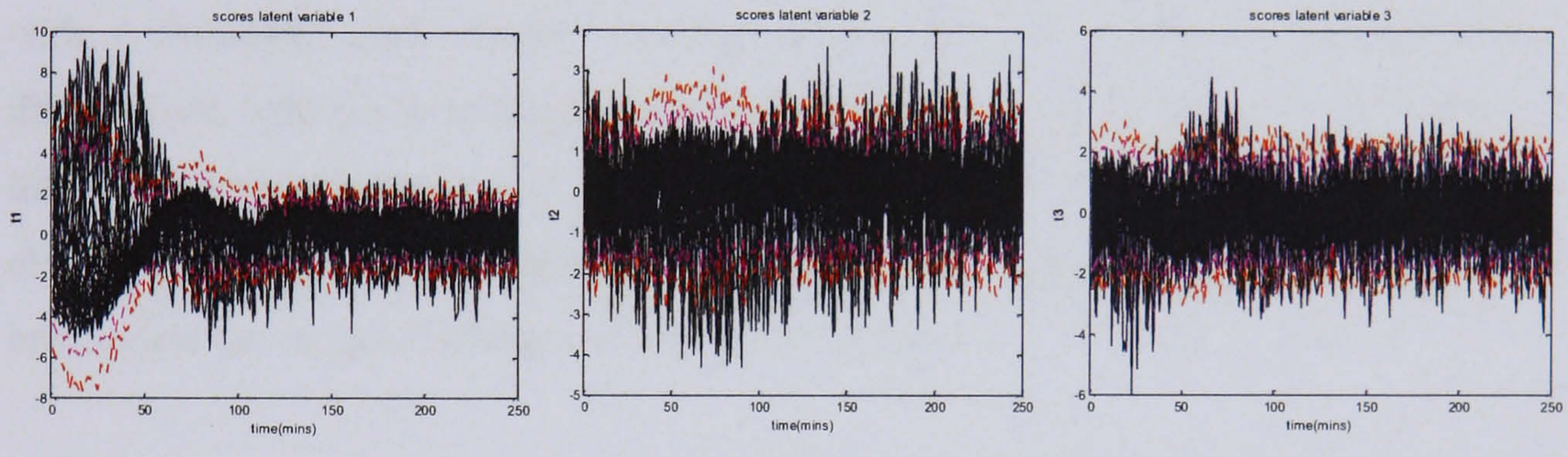
Figure 164 shows that the increase in noise level affects the super model-based technique after the noise to signal ratio increases beyond 10%. Prior to this point, all three fault types, the temperature sensor fault, the heat transfer coefficients decrease and the cooling valve fault all exhibit similar behaviour with no real decrease in the fault detection ability of the SMBPCA methodology. With fault type 1, the sensor fault, the effect of the noise to signal ratio increase is negligible at the 25% level and only starts to impact when it is of the order of 50 and 100%. However faults 2 and 3 demonstrate a more significant increase in the amount of time taken to detect an abnormal sample once the noise level has increased above 10%. This is because the control limits increase in magnitude as a consequence of the noise in the data and also because the batch trajectories fluctuate in and out of the control limits whether there is a problem on the plant or not due to the presence of the noise. This means there will also be a large increase in the number of false alarms. Figure 165 and Figure 166 show the control charts for the 100% noise to signal ratio for faults 2 and 3 respectively. One way of dealing with the noise would be to apply a filter to the data before presenting it to the model to remove the effect of noise.

It is difficult to explain why the detection of fault type 1 is much more weakly dependent on the noise level in the system, although it may be related to the fact that the sensor fault is on the reactor temperature sensor, and this is one of the variables to which the noise has been added, and also because the fault is larger in magnitude than the others.



a) Latent Variable 1 b) Latent Variable 2 c) Latent Variable 3

Figure 165 : Control charts for data with 100% noise to signal ratio – fault type 2



a) Latent Variable 1 b) Latent Variable 2 c) Latent Variable 3

Figure 166 : Control charts for data with 100% noise to signal ratio – fault type 3

8.5 Conclusions

The objective of this chapter was to investigate the robustness of the super model-based technique with respect to certain issues that could arise when the technique is applied in an industrial setting. The impact of issues such as model parameter uncertainty and sensitivity to noise were examined, along with investigating different possibilities for dealing with the plant model mismatch.

Applying SMBPCA to data from the simulation of the exothermic batch reactor indicated that certain fault types were more susceptible to parametric uncertainty in the model. Although the faults due to the sensor failure and reactor fouling were to some extent affected by model parameter uncertainty, it was noted that past a certain point of uncertainty, the sensitivity of the resultant model to detect the onset of faults was not

affected. However, with the control valve fault, as the uncertainty in the model parameter increased, it became impossible for the model to detect the process abnormality. One solution is to undertake a sensitivity analysis as it will help identify those model variables which are most critical in the model in terms of affecting its robustness.

A similar study was carried out to investigate the impact of increasing the noise levels in the data. The results of this study showed that up to a noise to signal level of approximately 10%, the fault detection abilities of SMBPCA were not greatly affected, however above this value, the ARL's began to increase significantly. Consequently to apply SMBPCA to data from noisy processes, filtering is required to be applied.

Plant model mismatch was also discussed, although a definitive solution to the problem was not proposed. A limited study was carried out using the exothermic batch reactor with a simulated plant model mismatch and it was shown that it was possible to differentiate between a fault and mismatch in the one case investigated. However, the difference between the model and the process will have an effect on the monitoring abilities of the approach and every effort should be made to accurately model the process and update the model as changes occur on the process.

9 CHAPTER NINE: CONCLUSIONS AND FURTHER WORK

9.1 Conclusions

The main focus of this research was to develop a technique for monitoring batch processes that took into account the non-linear and dynamic aspects of batch data.

The objectives of the thesis were as follows:

- to examine the multivariate techniques currently used to monitor batch processes and to highlight their advantages and limitations through application to data generated from a simulation of an exothermic batch simulation
- to examine the model-based principal component analysis technique, analyse its fault detection abilities in comparison with other batch monitoring techniques, and to highlight the limitations of the method
- to further develop the model-based principal component analysis technique to handle inaccuracies in the mechanistic model (super model-based principal component analysis)
- to investigate the different aspects that affect the super model-based principal component analysis technique in an industrial situation

The issue with batch data as opposed to continuous data is that the variables are correlated in time as well as with each other. One approach to addressing this is to unfold the data to form a two-dimensional data matrix to which the batch monitoring techniques can be applied. For example, for multiway PCA, the unfolding method results in a matrix to which PCA is applied. Originally MPCA proved valuable as an end-of-batch technique for process monitoring and classification. By unfolding the data and appropriate scaling, the mean trajectories of the batches are removed, and hence the technique goes some way towards removing the non-linear components typical of batch data, however it does not address the serial correlation in the data as it essentially treats the unfolded data as a static matrix. A further issue with MPCA is that is requires all batches to be of equal length, however this can be overcome either by cutting all batches to the same, or a selected, length, using an alternative index to time, such as percentage of reactant conversion, or by applying a technique such as dynamic time warping. It is also possible to apply MPCA as a through batch technique, although to do this a method

for in-filling the future unknown data must be implemented as MPCA requires batches that cover the duration of the process for analysis. Similar issues affect the application of multiway PLS, which incorporates quality data from the batch, and for the multi-block multiway PLS methodology which introduces the initial batch conditions as well as the process variables and end of batch quality results.

Adaptive techniques such as hierarchical PCA and moving window PCA do attempt to overcome the issue of serial correlation neglected by the multiway techniques by incorporating the time behaviour of batches into their analysis. This is achieved by computing several local PCA models as opposed to one global model, therefore capturing the different phases of the batch behaviour as it progresses. However, hierarchical PCA can be difficult to tune, and there is the possibility with moving window PCA that it could adapt to faults in the process rather than isolate them.

The batch observation level technique has the advantage of being a through-batch monitoring methodology. It uses a different method of unfolding which preserves the dynamic direction of the batches, allowing the batches to be monitored as they evolve as opposed to the analysis of complete batches. This also means that batches of differing durations can be included in the analysis. The limitations of the technique are that it does not take steps to address the dynamic and non-linear structure in the data. The Nomikos and MacGregor method of unfolding and scaling for MPCA means that the average trajectory of each variable at each time point is removed, which addresses the non-linear behaviour in the data. With the Wold *et al* (1998) unfolding and scaling method used in batch observation level PCA, only the global behaviour is removed, therefore the issue of non-linear structure is not addressed.

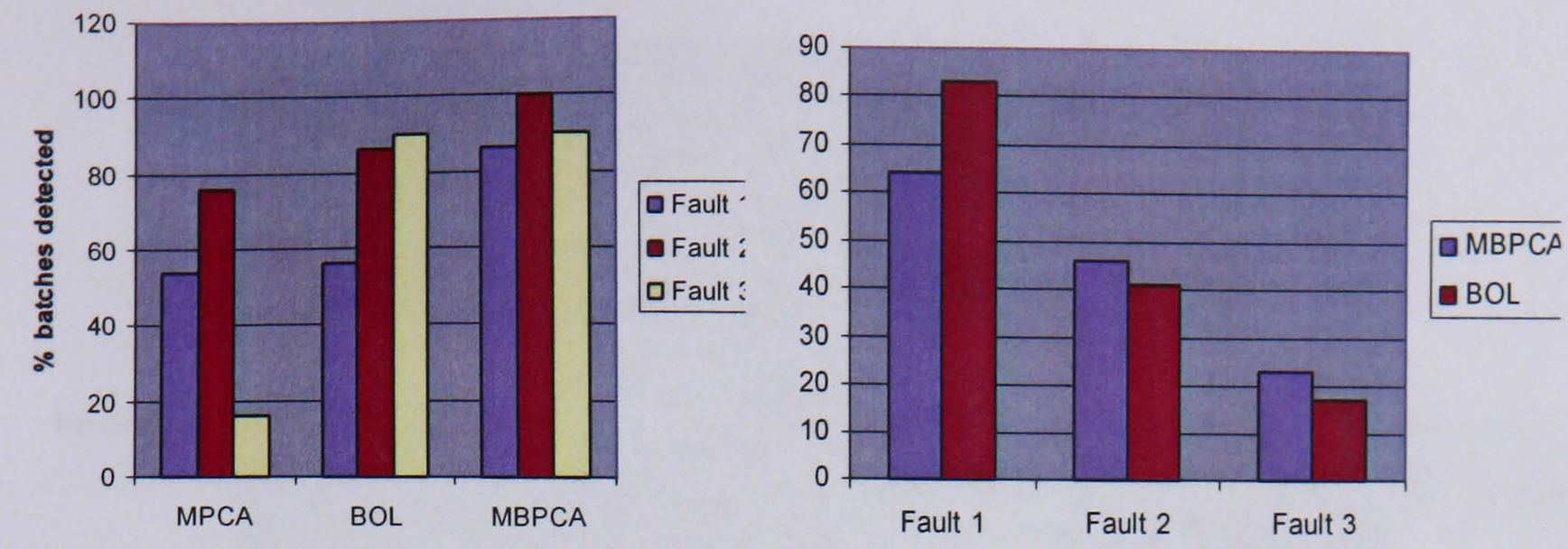
Dynamic MPCA and MPLS attempt to address the serial correlation in the batch data by incorporating an ARX model structure. However, these techniques do not deal with the non-linear behaviour inherent in the batch data as the lagged covariance matrix is calculated across the entire batch trajectory. The non-linear PCA and PLS techniques use a variety of different techniques to address the non-linearity in the data, for example neural networks or quadratic functions. The disadvantage of this type of technique is the interpretability of these black box models. They are also susceptible to overfitting.

Utilisation of state space models to monitor batch processes can be advantageous since they can be used for batches of unequal length, and more importantly they deal with the non-linear structure in the data by building several linear models over the duration of the

batch and also the dynamic state space model that is generated at every time point deals with the serial correlation in the data. However, defining the inputs and outputs for state space models can be challenging, and their interpretation can be difficult.

The model-based PCA technique attempts to deal with the non-linear and dynamic structure of batch data through the use of a mechanistic model which describes the physical and chemical relationships between the process variables. Applying this technique theoretically removes the non-linear and dynamic structure from the data, allowing a linear monitoring technique to be applied. It also has the advantage of including real physical knowledge about the process behaviour. However, the technique depends on an accurate model of the process being available which is not always possible. A reduced complexity model does not allow process abnormalities to be detected since the plant model mismatch masks changes in the plant. This drawback led to the development of the super model-based PCA technique, which incorporates an additional error model to address any inaccuracies between the model and the plant.

Before applying the super model-based PCA technique, investigations were carried out into the different batch monitoring techniques by comparing the performance of some of the key techniques using a simulation of an exothermic batch reactor. The fault detection and diagnostic abilities of multiway PCA, batch observation level PCA and model-based PCA were compared, as was their effectiveness in dealing with the non-linear and dynamic structures of the data. The results of this case study showed that the MBPCA technique detected the highest number of faulty batches, Figure 167. However the case study also showed that MBPCA only detected one of the faults more quickly than the BOL technique, Figure 167. The poor fault detection ability can be explained by the presence of a plant model mismatch between the process variables and those generated by the mechanistic model. The mismatch was introduced to mimic what could happen in an industrial situation, where a perfect model of the process was not available, and to test the robustness of the technique. However, although the technique demonstrated reduced fault detection ability, it still performed well in comparison to the multiway and batch observation level techniques.



a) % of faulty batches detected

b) out of control ARL

Figure 167 : Results of batch monitoring techniques case study

The drawback of the model-based PCA technique was identified as being related to the plant model-mismatches. Therefore the super model-based PCA technique was developed to enable a mechanistic model of the technique to be used even when the model is not a perfect match to the process. As mentioned previously, using a reduced complexity model of the process can result in its non-linear and dynamic structure being retained, thereby masking the impact of any abnormalities on the process. The SMBPCA technique addressed this by including an additional error model in the algorithm to remove any structure and non-linearity remaining in the data.

A number of different error models were investigated to determine the most appropriate. The results showed that SMBPCA with dynamic canonical correlation analysis and SMBPCA with dynamic non-linear PLS gave the best results with respect to fault detection and dealing with the typical batch characteristics. A comparison of the super-model based techniques with the batch monitoring techniques studied in Chapter 6 was undertaken and it was observed that SMBPCA gave the best performance as a monitoring technique of those tested, Figure 168.

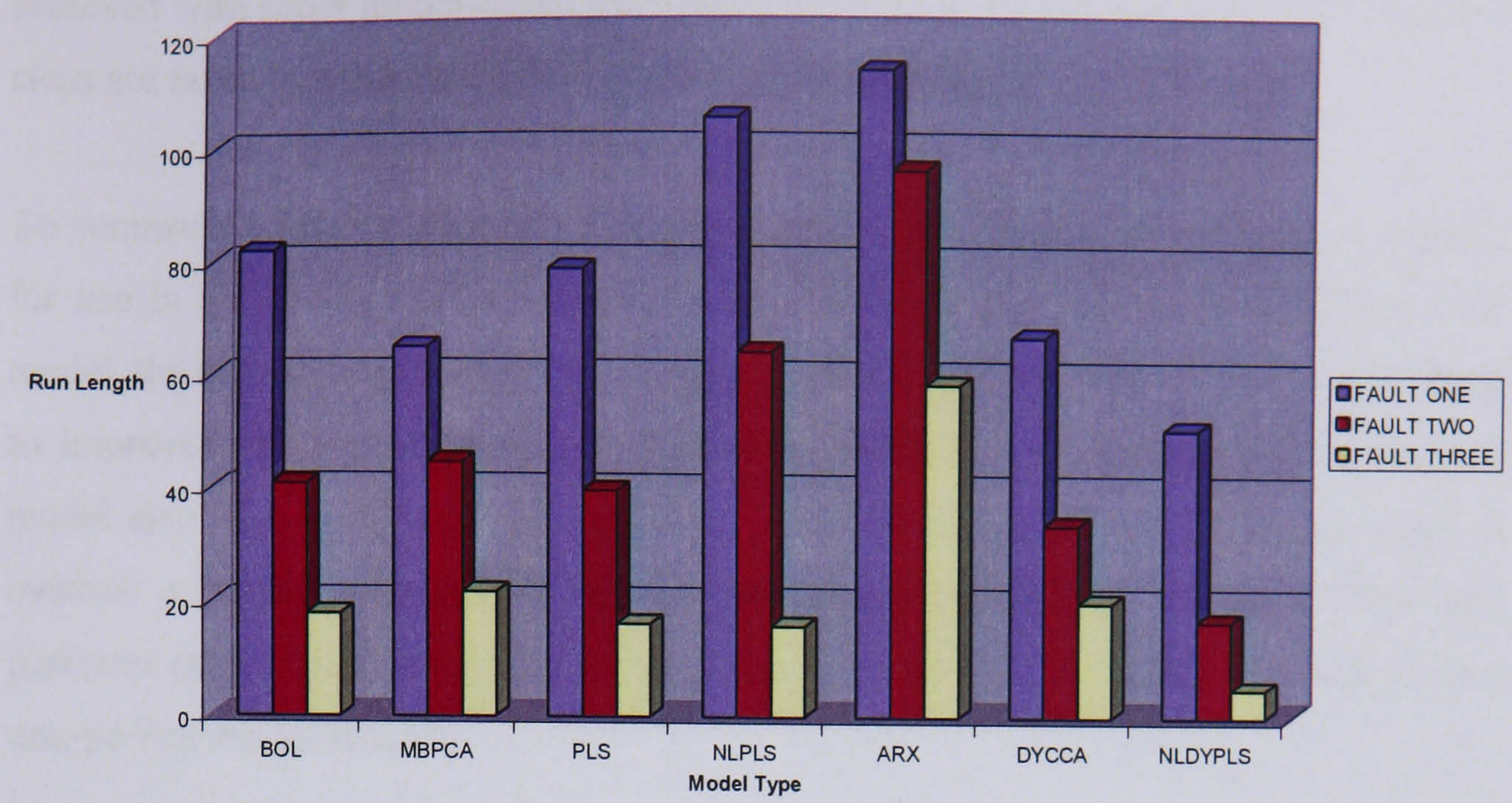


Figure 168: Out of control ARL for all batch monitoring techniques

The average run lengths for the SMBPCA technique using non-linear dynamic PLS were greatly reduced in comparison to those achieved using the batch observation level techniques, by more than 50% for two of the three faults tested. SMBPCA, using dynamic CCA, also demonstrated significant reductions in the ARL for two of the three fault types.

Although the case study on the super model-based monitoring techniques showed that that the SMBPCA technique out-performed the other batch monitoring techniques, it was also important to investigate the issues that would affect the technique if it was applied in an industrial situation. Therefore the robustness of the SMBPCA technique was examined with respect to model uncertainty, plant model mismatch and sensitivity to noise. With plant model mismatch, it was observed that it was possible for plant-model mismatch to be misinterpreted as faults. However it was also shown that there was potential for out of control signals to be classified either as a fault or as a mismatch if the correct pattern recognition or expert system was employed. The study into model uncertainty showed that the performance of the super model-based technique was not significantly affected by the increase in uncertainty in selected parameters until a certain limit was reached, after which the fault detection ability of the method decreased dramatically. A similar result was achieved for the study into the technique's ability to deal with instrumental noise in the process. Up to a certain level, the noise did not have a significant effect, but once past this level it became impossible to determine what was a fault and what was normal process activity. As would be expected, the best results were

achieved with super model-based PCA when a model as accurate as possible is used and steps are taken to reduce the effect of noise on the process.

To summarise, super model-based principal component analysis is a technique suitable for use in the monitoring of batch processes. Through the use of an additional error model, the non-linear and dynamic structures in the data are greatly reduced, which leads to improved process monitoring and fault detection. The addition of the extra error model also allows reduced complexity mechanistic models to be successfully used to monitor a batch process. Case studies carried out have shown that SMBPCA out-performs other batch monitoring techniques with respect to fault detection, even when a non-perfect model is used.

9.2 Future Work

The initial research and development into super model-based PCA has been undertaken in this thesis. However there are a number of areas that require further investigation to make the technique suitable for industrial application. The case study carried out in Chapter 7 showed that the application of the super model-based technique introduced additional noise into the residuals, making fault detection more complex due to varying control limits and batch trajectories fluctuating inside and outside of the control limits when the process was operating under normal conditions. It is likely that the cause of the noisy signal is due to over-fitting of the model, therefore research needs to be undertaken to determine if it is possible to reduce the noise through improved modelling of the process. A simple step that could be taken to improve the situation would be to apply a filter to the residuals to smooth them before the application of PCA, however consideration needs to be given to the desirability of including an additional step to the algorithm.

Another area which requires investigation is that of fault diagnosis. The super model-based technique demonstrated superior fault detection ability in comparison to other batch monitoring techniques investigated. However, the ability of the technique in terms of fault diagnosis was not investigated because it was deemed that contribution plots generated from the residuals of a set of residuals of a set of variables would not lead to a high level of interpretability from the operator. Therefore significant research needs to be carried out into possible methods for fault diagnosis using the SMBMPCA technique.

Initial studies were performed into the robustness of the SMBMPCA technique with respect to plant-model mismatch. There is scope for further work to be carried out with respect to the issue of discriminating between a mismatch and an actual fault, including the testing of pattern recognition techniques on the residuals. The issue of robustness can be further investigated by looking at the impact of model degradation, and how simplistic a mechanistic model can be before the technique offers no benefits over an empirical approach.

It would also be interesting to further develop the super model-based technique through the application of other types of error model, particularly hybrid models, such as a fuzzy logic model, as discussed in Chapter 4.

All the case studies carried out in this research used the same simulation of an exothermic batch reactor, therefore a key part of any future work should involve the application of the super model-based technique to an industrial data set and mechanistic model obtained from a plant. This would be possible to implement on a plant where a simulation of the process has already been designed for the purposes of operator training or for process development, as the equations used to build the simulator could be used for the mechanistic model.

10 References

Akaike, H., '*Canonical Correlation Analysis of Time Series and the Use of an Information Criterion*', System Identification: Advances and Case Studies', Academic Press, New York, 1976

Alexander, R. Clements, J. A., Guest, R., Leiper, K., Moffat, A. C.,Pearce, G., Pugh K., Wicks, S., '*New Technologies Forum 5: Multivariate Mathematical Approaches*', Medicines Control Agency, Royal Phamaceutical Society, Final Report, 2002

Amand, Th., Heyen, G., Kalitvenzett, B., '*Plant Monitoring and Fault Detection: Synergy Between Data Reconciliation and Principal Component Analysis*', Comp. & Chem. Eng., 25, 501-507, 2001

Baffi, G., Martin, E. B., Morris, A. J., '*Non-linear Projection to Latent Structures Revisited: the Quadratic PLS Algorithm*', Comp. and Chem. Eng., 23, 395-411, 1999

Baffi, G., Martin, E. B., Morris, A.J., '*Dynamic Non-linear Projection to Latent Structures Modelling*', Chem. & Int. Lab. Syst., 52, (1), 5-22, 2000

Baffi, G., Martin, E. B., Morris, A.J., '*Non-linear Projection to Latent Structures Revisited: the Neural Network PLS Algorithm*', Chemometrics Int Lab System., 52, (1), 5-22, 2000

Basseville, M., '*Detecting Changes in Signals and Systems – A Survey'*, Automatica, 24, (3), 309-326, 1998

Chen, J., Liu K.C., '*On-line Batch Process Monitoring using Dynamic PCA and Dynamic PLS Models*', Chem. Eng. Sci, 57, (1), 63-75, 2002

de Jong, S., '*SIMPLS: An Alternative Approach to Partial Least Squares Regression*', Chemom. Intell. Lab. Syst., 18, 251-263, 1993

Dong, D., McAvoy, T.J., '*Nonlinear Principal Component Analysis-based on Principal Curves and Neural Networks*', American Control Conference, 2, 1284-1288, 1994

Dunia, R., Qin, S. J., Edgar, T. F., McAvoy, T. J., '*Identification of Faulty Sensors Using Principal Component Analysis*', AIChE Journal, 42, (10), 2797-2812, 1996

Feyo de Azevedo, S., Dahm, B., Oliveria, F. R., '*Hybrid Modelling of Biochemical Processes: A Comparison with the Conventional Approach*', Computers & Chem. Eng., 21, S751-S756, 1997

Fletcher, N.M, Martin, E.B., Morris, A.J., Quinn, H., Hinge, B., '*The Monitoring of an Industrial Fed-Batch Fermentation Process*', 8th International Conference on Computer Applications in Biotechnology, 149-154, 2001

Fontaine, J. L., Germain, A., '*Model-based Neural Networks*', Computers & Chem. Eng., 25, 1045-1054, 2001

Frank, I. E., '*A Non-linear PLS Model*', Chemometrics Int. Lab. System, 8, 109-119, 1990

Geladi, P., Kowalski, B. R., ,*Partial Least Squares Regression: A Tutorial*', Anal. Chim. Acta., 185, 1-17, 1986

Gollmer, K., Posten, C., '*Supervision of Bioprocesses using a Dynamic Time Warping Algorithm*', Control Engineering Practice, 4, (9),1287-1295, 1996

H Martens, T Naes , '*Multivariate Calibration'* Chichester, England: John Wiley & Sons, 1989

Harkat, M. F., Mourot, G., Ragot, J., '*An Improved PCA Scheme for Sensor FDI: Application to an Air Quality Monitoring Network*', Journal of Process Control, 16, (6), 625-634, 2006

Höskuldsson, A., '*Prediction Methods in Science and Technology'*, Thor Publishing, Denmark, 1996.

Irwin, G. W., Warwick, K,. Hunt, K. J., '*Neural Network Applications in Control*'. The Institution of Electrical Engineers, London, ISBN: 0852968523, 1995

Jia, F., Martin, E.B., Morris, J., '*Non-linear Principal Components Analysis with Application to Process Fault Detection'*, Int Journal of Syst. Sci., 31(11), 1473-1487, 2000

Jollioffe, I., '*Principal Component Analysis'*, Springer-Verlag, New York, ISBN: 3540962697. 1986

Kemna, A.H., Mellichamp, D.A., '*Identification of Combined Physical and Empirical Models using Nonlinear a Priori Knowledge*', Control Eng. Practice, 3,3, 375-382, 1995

Kourti, T., Nomikos, P., MacGregor, J.F., '*Analysis, Monitoring and Fault diagnosis of Batch Processes using Multiblock and Multi-way PLS*', Journal of Process Control, 5. 277-284, 1995

Kramer, M.A., '*Non-linear Principal Component Analysis Using Autoassociative Neural Networks'*, AIChE Journal, 37, 233 (1991)

Ku, W., Storer, R. H., Georgakis, C.,'*Disturbance Detection and Isolation by Dynamic Principal Component Analysis*', Chemometrics Intel. Lab. Sys., 30, 179-196, 1995

Lachman-Shalem, S., Haimovitch, N., Shauly, E. N., Lewin, D. R., '*MBPCA Application for Fault Detection in NMOS Fabrication*', IEEE Trans on Semiconductor Manufacturing, March 2001

Lakshminarayanan, S., Fujii H., Grosman B., Dassau E., Lewin D.R., '*New Product Design Via Analysis of Historical Databases*', Computers & Chem. Eng., 24, (2), 671-676, 2000

Lee, J. H., Dorsey, A. W., '*Monitoring of Batch Processes Through State Space Models*' AIChE Journal, 50, 1198-1210, 2004

Lennox, B., Montague, G.A., Hiden, H., Kornfeld, G., '*Moving Window MSPC and its Application to Batch Processes*', IFAC, 2001

Leung, D., Romagnoli, J., '*Dynamic Probabilistic Model-based Expert System for Fault Diagnosis*', Computers & Chem Eng, 24, (11), 2473-2492, 2000

Lewin, D.R., Lavie, R., '*Designing and Implementing Trajectories in an Exothermic Batch Chemical Reactor'*, Ind. Eng. Chem. Res., 1990, 29, 89-96

Li, W., Yue, H. H., Valle-Cervantes, S., Qin, S. J., '*Recursive PCA for Adaptive Process Monitoring*', Journal of Process Control, 10, (5), 471-486, 2000

Lorber, A., L. E. Wangen and B. R. Kowalski, '*A Theoretical Foundation for the PLS Algorithm*,' Journal of Chemometrics, 1(19), 1987.

Louwerse, D.J., Smilde, A.K., *'Multivariate Statistical Process Control of Batch Processes based on Three-Way Models*', Chem. Eng. Sci., 55, 1225-1235, 2000

Love, J., '*Process Automation Handbook: A Guide to Theory and Practice*', Springer-Verlag, New York, ISBN-10: 1-84628-281-0, 2007

Lu, N., Yao, Y., Gao, F. '*Two-Dimensional Dynamic PCA for Batch Process Monitoring*', AIChE Journal, 51, (12), 3300-3304, 2005

Luyben, W.L., '*Process Modelling, Simulation and Control for Chemical Engineers*', McGraw-Hill, New York, 160-167, 1973

Martin, E.B., Morris, A.J., Zhang, J, '*Process Performance Monitoring using Multivariate Statistical Process Control*', Systems Eng. For Automation, 143, (2), 132-144, 1996

Martin, E.B., Morris, A.J., Zhang, J, '*Process Performance Monitoring using Multivariate Statistical Process Control*', Systems Eng. For Automation, 143, (2), 132-144, 1996

McPherson, L.A., Martin, E.B., Morris, A.J., '*Super Model-based Process Performance Monitoring'*, Advances in Process Control 6, September 2001, York.

Molga, E., Cherbański, R., '*Hybrid First-principle-neural-network Approach to Modelling of the Liquid-liquid Reacting System*', Chem. Eng. Sci., 54, 2467-2473, 1999

Negiz, A., Cinar, A., '*Statistical Monitoring of Multivariate Dynamic Processes with State Space Models*', AIChE Journal, 43, (8), 2002-2020, 1997

Nomikos, P., '*Detection and Diagnosis of Abnormal Batch Operation Based on Multi-way Principal Component Analysis*', World Batch Forum, Toronto, May 1996, ISA Transactions, 35, 259-266, 1996

Nomikos, P., MacGregor J.F., '*Multi-way Partial Least Squares in Monitoring Batch Processes*', Chemometrics Intel. Lab. Syst., 30, 97-108, 1995

Nomikos, P., MacGregor, J. F., '*Monitoring of Batch Processes Using Multiway Principal Component Analysis*', AIChE Journal, 40, (8), 1361-1375, 1994

Nomikos, P., MacGregor, J. F.,'*Multivariate SPC Charts for Monitoring Batch Processes*'. Technometrics, 37, (1), 41-59, 1995

Patton, R. J., Frank, P. M., Clark, R. N., '*Advances in Fault Diagnosis for Dynamic Systems*', Springer -Verlag, New York, ISBN: 3540199683, 2000

Psichogios, D.C., Ungar, L.H., '*A Hybrid Neural Network First Principles Approach to Process Modelling'*, AIChE Journal, 38, (10), 1499-1511, 2004

Qin, S. J., Dunia, R., '*Determining the Number of Principal Components for Best Reconstruction*', Journal of Process Control, 10, (2-3), 245-250, 2000

Qin, S. J., McAvoy, T. J., '*Non-linear PLS Modelling using Neural Networks*', Computers & Chem. Eng., 16, 379-391, 1992

Ranner, S., MacGregor, J. F., Wold, S., '*Adaptive Batch Monitoring using Hierarchical PCA*', Chemometrics Intel. Lab. Systems, 41, 73-81, 1998

Rengaswamy, R., Mylaraswamy, D., Årzeń, K.-E., Venkatasubramanian, V., '*A Comparison of Model-based and Neural Network-based Diagnostic Methods*', Eng. Appl. Of A. I., 14, (6), 805-818, 2001

Rotem Y., Lewin D. R., '*Assessing the Impact of Parametric Uncertainty on the Performance of Model-based PCA*', ADCHEM 2000, International Symposium on Advanced Control of Chemical Processes, 2000

Rotem, Y., Wachs, A., Lewin, D. R., '*Ethylene Compressor Monitoring Using Model-based PCA*', AIChE Journal, 46, (9), 1825-1836, 2000

Shaper, C.D., Larimore, W.E., Seborg, D. E., Mellichamp D. A., '*Identification of Chemical Processes Using Canonical Variate Analysis*', Computers & Chem. Eng., 18, (1), 55-69, 1994

Shapiro, S.S., Wilk, M.B., '*An Analysis of Variance Test for Normality (Complete Samples)*', Biometrika, 52, 2, 591-611 (1965)

Shi, R. & MacGregor, J.F.,' *Modelling of Dynamic Systems using Latent Variable and Subspace Methods'*. Journal of Chemometrics 14 (2000), 423-439

Simoglou, A., Martin, E.B., Morris, A.J., '*Statistical Performance Monitoring of Dynamic Multivariate Processes using State Space Modelling*', Computers & Chem Eng, 26, (6), 909-920, 2002

van Lith, P. F., Betlem, B. H. L., Roffel, B., '*A Structured Modelling Approach for Dynamic Hybrid Fuzzy-first Principles Models*', Journal of Process Control, 12, (5), 605-615, 2002

van Sprang E.N.M., Ramaker H.-J., Westerhuis J.A., Gurden S.P., Smilde A.K, '*Critical Evaluation of Approaches for On-line Batch Process Monitoring*', Chem. Eng. Sci., 57, (18), 3979-3991, 2002

Wachs, A., Lewin, D.R., '*Process Monitoring Using Model-based PCA'*, Proc. IFAC Symp of Dyn. And Control of Process Systems, Corfu, 1998.

Wetherill, G. B., Brown, D. W., '*Statistical Process Control: Theory and Practice*', Chapman and Hall, London, ISBN: 0412357003, 1991

Wise B. M., Shaver J. M., Gallagher N. B., Windig W., Bro R., Koch R. S., '*PLS_Toolbox 4.0 Manual*', Eigenvector Research Incorporated, Winatchee, 2006

Wise, B.M., Ricker, N.L.,*'Recent Advances in Multivariate Statistical Process Control: Improving Robustness and Sensitivity'*, Proc. IFAC. ADCHEM Symp., International Federation of Automatic Control, 125, 1991

Wold, S. et al, *'Modelling and Diagnostics of Batch Processes and Analogous Kinetic Experiments*', Chem & Int. Lab. Systs., 44, 331-340, 1998

Wold, S., '*Non-linear Partial Least Squares Modelling. II Spline Inner Function*'. Chemometrics Intel. Lab. System, 14, 71-84, 1992

Wold, S., Esbensen, K., Geladi, P.,*'Principal Component Analysis*', Chemometrics Intel. Lab. Syst., 2, 37-52, 1987

Wold, S., Kettaneh-Wold, N., Skagerberg, B. ,*'Non-linear PLS Modelling*', Chemometrics Intel. Lab. System, 7, 53-65, 1989

Wold, S., Ruhe, A., Wold, H., Dunn, W. .J, '*Collinearity Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses*' Siam J. Sci. Stat. Comput., 5, (3), 735-743, 1984

Xiong, Q., Jutan, A., '*Grey-box Modelling and Control of Chemical Processes*', Chem. Eng. Sci, 57, 1027-1039, 2002

Yang, T.N., Wang, S.D., *'Robust Algorithms for Principal Component Analysis'*, Pattern Recognition Letters 20, 927-933, 1999

Yoon, S., MacGregor, J. F., '*Fault Diagnosis with Multivariate Statistical Models Part I: Using Steady State Fault Signatures*', Journal of Proc. Cont., 11, (4), 387-400, 2001

Zhang, J., Martin, E. B., Morris, J., '*Non-linear performance monitoring'*, Control '96, UKACC International Conference on Control, 2, (427), 924- 929, 2-5 Sept. 1996

Zorzetto, L. F. M., Filho, M., Wolf-Maciel, M. R., '*Process Modelling Development Through Artificial Neural Networks and Hybrid Models*', Computers & Chem. Eng, 24, (2-7), 1355-1360, 2000