Newcastle University

Data Integration for the Monitoring of Batch Processes in the Pharmaceutical Industry

By

Chris Wai Leung Wong

王 偉 樑

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

School of Chemical Engineering and Advanced Materials
Newcastle University
United Kingdom

May 2007

Abstract

Advances in sensor technology has resulted in large amounts of data being available electronically. However, to utilise the potential of the data, there is a need to transform the data into knowledge to realise an enhanced understanding of the process. This thesis investigates a number of multivariate statistical projection techniques for the monitoring of batch fermentation and pharmaceutical processes. In the first part of the thesis, the traditional performance monitoring tools based on the approaches of Nomikos and MacGregor (1994) and Wold *et al.* (1998) are introduced. Additionally, the application of data scaling as a data pre-treatment step for batch processes is examined and it is observed that it has a significant impact on monitoring performance. Based on the advantages and limitations of these techniques, an alternative methodology is proposed and applied to a simulated penicillin fermentation process. The approach is compared with existing techniques using two metrics, false alarm rate and out-of-control average run length.

A further manufacturing challenge facing the pharmaceutical industry is to understand the differences in the performance of a product which is manufactured at two or more sites. A retrospective multi-site monitoring model is developed utilising a pooled sample variance-covariance methodology of the two sites. The results of this approach are compared with a number of techniques that have been previously reported in the literature for the integration of data from two or more sources.

The latter part of the thesis focuses on data integration using multi-block analysis. Several blocks of data can be analysed simultaneously to allow the inter- and intra- block relationships to be extracted. The methodology of multi-block Principal Component Analysis (MBPCA) is initially reviewed. To enhance the sensitivity of the algorithm, wavelet analysis is incorporated within the MBPCA framework. The fundamental advantage of wavelet analysis is its ability to process a signal at different scales so that both the global features and the localised details of a signal can be studied simultaneously. Both existing and the modified approach are applied to data generated from an experiment conducted in a batch mini-plant and that was monitored by both physical sensors and on-line UV-Visible spectrometer. The performance of the integrated approaches is benchmarked against the individual process and spectral monitoring models as well as examining their fault detection ability on two additional batches with pre-designed process deviations.

Acknowledgement

I would like to express my sincere thanks to my supervisor Professor Elaine Martin for her unlimited trust and faith in my ability through the important journey of my life. Her valuable support and encouragement are one of the critical motivations to drive me to the completion of this thesis. I would also like to thank Professor Julian Morris for his guidance in my early part of the PhD and Angela Bott for all her help and administrative support.

I am also grateful to Mr Richard Escott and Dr Christian Airiau from GlaxoSmithKline (GSK) Tonbridge site who provided great insight into the pharmaceutical industry. Great thanks for the organisation of placements at various GSK sites and for the data collected with its process understanding. The valuable support from Richard is more than what a typical industrial supervisor would provide.

I would like to acknowledge the financial support from the Engineering and Physical Sciences Research Council (EPSRC) for the industrial CASE award, the Overseas Research Students Award Scheme (ORSAS) and the Centre for Process Analytics and Control Technology (CPACT).

I would also like to thank all the friends and people I met during my stay in Newcastle in particular to Sophia Triadaphillou, Pol Moreno, Tien Kheng Khoo, Gwen Chen, Susan Lau and Ivan Man. It is the friendship and joy they bought that motivated every moment regardless of my mood level. My PhD life would never be possible without their distractions and fun. My great CPACT and departmental colleagues of Suresh, Ahmed, Daniel and Maria, Mahesh, Pieter and Dave, Pieter, Katarina, Marco, Ming and Vince gave me different levels of challenges. Great thanks also to Irene, Agnes, Alfred, Lawrence, Yvonne, the committee and members of Chinese Students Society, Staff from the Enterprise Centre, peers from the UK GRAD Programme and CMI Enterprisers. Especially to Joanne, whose love and companionship was appreciated, and is missed.

The greatest thanks is reserved for my family, my parents and my brother Charles, who always provide the love and support no matter when and what. They have made me progress to the next level of the journey.

Contents

Abbreviations and Acronyms

API	Active Pharmaceutical Ingredient
ARL	Out-of-control Average Run Length
CPC	Common Principal Component
CPCA	Consensus Principal Component Analysis
CWT	Continuous Wavelet Transform
DCS	Distributed Control Systems
DoE	Design of Experiments
DS	Design Space
DWT	Discrete Wavelet Transform
EWMA	Exponentially Weighted Moving Average
FDA	U.S. Food and Drug Administration
FIR	Finite Impulse Response
HPCA	Hierarchal Principal Component Analysis
HPLC	High Performance Liquid Chromatography
IR	Infrared
LDPE	Low-density Polyethylene
MBPCA	Multi-block Principal Component Analysis
MBPLS	Multi-block Partial Least Squares
MIR	Mid-Infrared
MPCA	Multiway Principal Component Analysis
MPLS	Multiway Partial Least Squares
MRA	Multiresolution Analysis
MSC	Multiplicative Signal Correction
MSPC	Multivariate Statistical Process Control
MSPCA	Multi-scale Principal Component Analysis
NAS	Net Analyte Signal
NIPALS	Non-linear Iterative Partial Least Squares
NIR	Near-Infrared
OSC	Orthogonal Signal Correction
PARAFAC	Parallel Factor Analysis
PAT	Process Analytical Technology
PCA	Principal Component Analysis
PCPC	Partial Common Principal Component
PLS	Partial Least Squares or Projection to Latent Structures

PRESS	Predicted Error Sum of Squares
QbD	Quality by Design
R&D	Research and Development
RMSEP	Root Mean Square Error of Prediction
SCADA	Supervisory Control and Data Acquisition Systems
SNV	Standard Normal Variate
SPC	Statistical Process Control
SPE	Squared Prediction Error
SPLS	Serial Partial Least Squares
SVD	Singular Value Decomposition
TLD	Tri-linear Decomposition
UCL	Upper Control Limit
UV-Visible	Ultra-violet Visible

Nomenclature

English Characters

A	Absorbance
a	Scaling parameter in wavelet analysis
a_{LB}	Molar absorptivity
a_{msc}	Multiplicative correction factor
B	Number of data blocks
b	Translation parameter in wavelet analysis
b_{LB}	Path length in cm
b_{msc}	Additive correction factor
$\mathbf{C}$	Correlation matrix
$\mathbf{C}^P$	Pooled correlation matrix
c_{ij}	Covariance between variables $\mathbf{X}_i$ and $\mathbf{X}_j$
c_{LB}	Molar concentration
cl	Control limit
c^{SPE}	Contributions for SPE
c^{T^2}	Contribution for Hotelling's T^2
c^t	Contributions for score
d	Inner regression coefficient
$\mathbf{E}$	Residual matrix
$\mathbf{E}_B$	Block residual matrix in multi-block analysis
$\mathbf{F}$	Residual matrix for $\mathbf{Y}$
$F(\tau, \omega)$	Time-frequency domain function
$F(\omega)$	Frequency domain
$f(t)$	Time domain function
G	Number of groups
H	Number of wavelet coefficients
H_0	Null hypothesis
I	Number of samples or batches
J	Number of variables
j_{block}	Number of variables in a block
K	Number of time points
L	Number of quality variables
L_p	Leverage for loadings

L_t	Leverage for scores
M	number of vanishing moments
$\mathbf{P}_B$	Block loading matrix in multi-block analysis
$\mathbf{P}^P$	principal component loadings of the multiple group model
$\mathbf{P}_T$	Super loading matrix in multi-block analysis
$\mathbf{p}_r$	Loading vector of principal component r
p	Criterion for the percentage of variation explained
$\mathbf{q}_r$	Output loading vector of latent variable r
R	Number of principal components selected in the model
r	r^{th} principal component
$\mathbf{S}$	Estimated variance-covariance matrix
$\mathbf{S}^P$	Pooled sample variance-covariance matrix
s	Standard deviation
$\mathbf{T}$	Super block matrix in multi-block analysis
$\mathbf{T}_B$	Block score matrix in multi-block analysis
$\mathbf{T}_T$	Super score matrix in multi-block analysis
$\mathbf{t}_r$	Score vector of principal component r
$t_{I-1,\alpha/2}$	Value of the Student's t-distribution with $I-1$ degrees of freedom
$\mathbf{U}$	Same column vector as does the scores matrix $\mathbf{T}$ but normalised to unit length
$\mathbf{u}_r$	Output score vector of latent variable r
v_b	Block weighting
$\mathbf{W}$	Diagonal matrix containing the square roots of the eigenvalues
$\mathbf{W}_T$	Super weight matrix in multi-block analysis
$\mathbf{w}_r$	Weight vector of latent variable r
$\underline{\mathbf{X}}$	Three-dimensional matrix with size I x J x K
$\mathbf{X}$	Two-dimensional matrix with size I x J
$\mathbf{X}_B$	Sub-block of two-dimensional matrix in multi-block analysis
$\mathbf{X}^*$	Mean-centred data matrix
$\mathbf{X}^+$	Auto-scaled data matrix
$\bar{\mathbf{x}}$	Vector of mean value
$\mathbf{Y}$	Two-dimensional matrix with size I x L
z_α	Standard normal deviate

Greek Characters

α	Significance level
λ	Eigenvalue
σ	Standard deviation
Λ	Diagonal matrix of a covariance matrix
$\Psi(t)$	Mother wavelet
$\Psi(\omega)$	Fourier transform
Σ	Variance-covariance matrix
β	Regression coefficient

Publications

Wong, C. W. L., R. E. A. Escott, A. J. Morris and E. B. Martin (2005), The integration of process and spectroscopic data for enhanced knowledge extraction in batch processes, *European Symposium on Computer-Aided Process Engineering-15*, Barcelona, 1141-1146.

Wong, C. W. L., A. J. Morris and E. B. Martin (2004), Evaluation of multivariate statistical approaches for batch performance monitoring, *Royal Society of Chemistry Emerging Chemometrician*, Glasgow.

Wong, C. W. L., E. B. Martin, A. J. Morris and R. E. A. Escott (2004), Realising the power of process spectroscopic integration in batch performance monitoring, *Advances in Process Analytics and Control Technology*, Bath.

Wong, C. W. L., R. E. A. Escott, A. J. Morris and E. B. Martin (2003), Multi-site performance monitoring in batch pharmaceutical production, *IFAC ADCHEM International Symposium on Advanced Control of Chemical Processes*, Hong Kong, 766-771.

Wong, C. W. L., R. E. A. Escott, A. J. Morris and E. B. Martin (2003), Multi-site performance monitoring in batch pharmaceutical production, *Advances in Process Analytics and Control Technology*, York.

Wong, C. W. L., E. B. Martin, A. J. Morris and R. E. A. Escott (2001), Mathematics improve therapeutic drug manufacture, *Britain's Younger Engineers*, London.

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation and Objectives

In today's competitive process manufacturing industry, innovation is one of the major strategies for pioneers to remain ahead of their competitors. Challenges facing the major players include the need to bring new products to the market place in the shortest time possible, the achievement of right-first-time production and the manufacture of consistently high quality product at minimal cost. A key facilitator to achieving these goals is the translation of data into information and knowledge. More specifically through the analysis of data, an enhanced level of understanding is attained about the process and product, leading to the control of the process in a flexible manner and ultimately improved process capability.

In the pharmaceutical industry, a major cost benefit, which can be realised by reducing the time from drug development to full-scale production, is the effective extension to the lifetime of a product patent. In the transfer of new products at the research and development (R&D) stage into full-scale production, the extraction and use of information from data generated at different process stages can potentially realise significant reductions in the time taken to launch new products. The converse of this concept, i.e. the transfer of knowledge back to R&D through the extraction of information from the production data will, in the longer term, help drive the more rapid introduction of a new product. Traditionally, the standard approaches have been to rely on science and engineering expertise and phenomenological modelling. Mechanistic based approaches may alone not be viable due to the complexity of the process thus there is a need to complement existing tools and skills with additional advanced technologies. Multivariate statistical projection techniques are one possible set of technologies and they have been successfully applied in support of product discovery and product manufacturing (Martin *et al.*, 1999; Martin and Morris, 2002).

The concept of data integration between new product development and full-scale manufacturing in the pharmaceutical industry is that of a two-way flow of information; an upstream flow of new product recipes, including process models for use in full-scale plant production (scale-up) and a downstream information flow in terms of product quality and production data from manufacturing (scale-down). The upstream flow of information will help realise the faster introduction of new products and formulations into the main production facilities through the establishment of a formalised link between the product development laboratory, the pilot plant and the main production facilities. A reduction in innovation time will accordingly be achieved since fewer trials may be necessary to achieve a stable formulation on the production plant. The

downstream flow will provide the product development team with a well-structured overview of production and related issues associated with plant production and their impact on product consistency and quality. Ultimately through the use of this information, the knowledge acquired will be incorporated into new formulations and products.

However, such a concept is compounded by a number of research challenges including minimal amounts of data, especially from the production plant from which the model is to be developed, relative to the laboratory scale data and the fact that processes exhibit neither linear nor steady-state behaviour. The information from the different development and production stages will differ not only in terms of measurements recorded, but also with respect to sampling frequency and the different underlying physical and chemical properties of the system. The method of multi-block analysis (Westerhuis *et al.*, 1998) underpins the research work reported in this thesis. For this methodology, process performance and process understanding can be realised by analysing different groups of variables that are split into conceptually meaningful blocks and then appropriate models are developed for each block that are subsequently integrated. This route potentially enables the comprehensive integration of the whole of the product development life cycle.

One of the key competitive pressures within the pharmaceutical industry is the need to compete on a global scale and hence a continuous supply of new products to market is necessary as a consequence of the move to globalisation and rationalisation. The transfer of product manufacture to different manufacturing sites around the world is becoming the norm. Furthermore, increasing manufacturing capacity is necessary to handle the increasing number of new products in the pipeline. Also to lower the manufacturing cost, existing products are being manufactured outside of Europe and the USA. Such a manufacturing strategy is typical of that adopted by the major players whose manufacturing sites are now spread across the world.

In multi-site manufacturing and the need to transfer between sites, the issue is how to utilise, most effectively, the data from the existing site gathered during development and production for more effective product transfer. Multi-site production also encapsulates the situation where product quality between sites differs. Through the application of multivariate statistical projection techniques, a between-site data comparison can be undertaken to isolate the causes of the differences and hence product quality will be enhanced through increased process understanding.

A multivariate statistical projection based technique that allows the monitoring of groups of products was proposed by Lane (1999). He proposed a multi-group technique which can simultaneously handle a range of products manufactured at different grades by a single

representation. In this thesis, investigations are undertaken to study further the applicability of the multi-group technique to multi-site monitoring. A pharmaceutical process which is manufactured at two different sites is used to investigate such a monitoring scheme.

The key philosophy underpinning these research areas is that of Statistical Process Control (SPC). SPC is concerned with the monitoring of a process over time to be able to detect the onset of changes in process behaviour, process disturbances and special events at an early stage in their development. A process is said to be in a state of "statistical control" if certain process or product variables remain close to their desired values and the only source of variation is "common cause" variation. An excellent review on SPC is given by Montgomery (1996).

With the rapid development in sensor technology, the traditional SPC approach has been found to be inappropriate for modern day processes since much of the data gathered is not utilised. SPC systems effectively detect or provide early warning of unusual events that are related to individual process or quality measurements. However, most industrial processes are multivariate in nature and univariate SPC provides no information about the interactions between variables. Therefore Multivariate Statistical Process Control (MSPC) systems are considered to be more appropriate since all the variables of interest are analysed simultaneously and information on the behaviour of each variable relative to the others can be extracted. MSPC, in particular process performance monitoring and fault detection, is the main focus of this thesis. The two key methodologies forming the basis of MSPC have been the multivariate statistical projection techniques of Principal Component Analysis (PCA) and Partial Least Squares (PLS).

Through the application of PCA and PLS to a data set, the covariance/correlation structure between variables can be identified and information on the process or quality parameters, can be decomposed to a few uncorrelated latent variables. Consequently, multivariate statistical monitoring methods can be applied in the latent variable space to implement a process performance monitoring and fault diagnosis scheme. PCA and PLS based process performance monitoring and modelling methodologies have been developed and applied to both continuous and batch systems (Nomikos and MacGregor, 1995a; Wise *et al.*, 1999). However, the classic PCA and PLS methods are not necessarily suitable to analyse the data sets generated from batch processes. The challenge of monitoring batch processes is the addition of a third dimension, that of batch, thus the data becomes three-way. Since the standard PCA and PLS algorithms are applicable to two-way matrices, the techniques have been extended to handle three-way array problems, i.e. multiway PCA and multiway PLS.

In the thesis, different multi-way techniques are investigated. Multiway PCA by Nomikos and MacGregor (1995b) are compared and evaluated against the batch observation level monitoring of Wold *et al.* (1998) for the monitoring of the performance of batch processes. From the evaluation, both the advantages and limitations are identified and a modified monitoring approach is proposed that forms an alternative approach for MSPC. All three approaches are first evaluated using a penicillin simulation application (Birol *et al.*, 2002) before being applied to data from an industrial pharmaceutical process.

An appropriate batch process analysis and monitoring scheme is crucial to the monitoring of the performance of batch processes. In the last decade, there has been a significant amount of research activity in the area. The success of the methodologies was because they are both viable and adaptable to the batch operating industries as well as achieving the objectives for batch analysis as defined by MacGregor in the 1990s and summarised by Kourti (2003):

1. *Develop methods for analysing the operation of batch processes by employing the large number of process measurements collected routinely by process computers during the progression of individual batches.*
2. *Develop effective procedures for the on-line monitoring of the progression of batch processes.*
3. *Create simple tools for easy technology transfer to industry; monitoring charts can be easily understood and responded to by the process operators.*
4. *Develop diagnostic tools capable of rapidly identifying the nature of any fault once it has been detected.*

The concept of monitoring batch trajectories has been achieved by projecting them onto reduced spaces using multiway PCA and multiway PLS. A range of industrial applications have been reported demonstrating their applicability and effectiveness (Neogi and Schalgs, 1998; Kourti, 2005). Other modelling methods have been investigated such as adaptive PCA (Rannar *et al.*, 1998), batch dynamic PCA (Chen and Liu, 2002) and the tri-linear approaches of PARAFAC (Bro, 1997) and Tucker (Smilde, 1992) to address a variety of issues related to batch performance monitoring. The methodologies were found to be complementary although the sensitivity of an approach is specific to the application (Westerhuis *et al.*, 1999; Chiang *et al.*, 2006).

Following the introduction of the Process Analytical Technology (PAT) initiative by the U.S. Food and Drug Administration (FDA, 2004), the importance of batch performance monitoring has further been enhanced since MSPC is recognised as a tool to help build quality into the product. Wold (2004) and Wold *et al.* (2006) proposed to look at PAT on four levels:

1. *On-line or at-line instrumentation for the immediate analysis of the end-product at each process stage can be applied in the pharmaceutical production, e.g. granulation, drying, mixing, tableting and coating. Multivariate analytical data are translated back to, for example, concentrations, pHs and temperatures before interpretation.*
2. *Acceptable profiles and signatures of data corresponding to acceptable product are defined without necessarily returning to concentrations of known compounds or other known properties. One can alternatively use classification to determine whether the sample is within specification.*
3. *On-line spectroscopic and other multidimensional sensor arrays are used to monitor the manufacturing process in real time. The approach and tools are basically the same as for level 2 but applied "continuously" over all time points and phases of the manufacturing processes, to ensure that these processes and trajectories are within specification at all times.*
4. *Monitoring of the whole process can be achieved by putting data from all the steps together with the raw material properties for a total view of the process. Steps include synthesis of active ingredient, characterisation of excipients and other raw materials, granulation, mixing until tableting, coating and packaging.*

The four levels do not necessarily have to be executed in the order presented, but can be used as guidance. These four levels reinforce the objectives proposed by MacGregor (Kourti, 2003) but the focus has now switched to considering spectroscopic sensors in addition to the traditional process measurements. Further details of monitoring batch processes using spectroscopy can be found in Chapter 5.

1.2 Contributions of the Thesis

The primary area of contribution of the thesis is in the field of batch process performance monitoring and fault detection. More specifically the key contributions include:

- The development of a new batch process monitoring approach. The proposed methodology draws together the advantages of two existing approaches suggested by Nomikos and MacGregor (1995b) and Wold *et al.* (1998). The existing approaches have their own individual advantages including the removal of the batch mean trajectories and hence the major non-linear effect is removed and the handling of unequal batch lengths.

However there is no existing approach that combines the advantages of the two schemes thereby delivering an enhanced monitoring scheme. The proposed approach addresses these core issues and hence provides an alternative methodology to those currently reported in the literature.

– An evaluation of the on-line monitoring performance of a number of approaches using a simulation of a penicillin fermentation process is undertaken. Two metrics, false alarm rate and out-of-control average run length are considered to evaluate monitoring performance and the fault detection capability of the different batch process monitoring methods. The penicillin simulation had been developed based on a realistic dynamic model therefore the process data generated contains the features of a batch process such as non-linearity, dynamics and multi-stage behaviour and hence is suitable for the comparative study.

– One of the challenges facing the pharmaceutical manufacturing industry is to understand the performance of a product when it is manufactured at two or more sites and where independent monitoring systems have been developed. A number of approaches are considered and investigated for the development of a multi-site monitoring scheme to enable the real source of differences between sites to be identified. In particular a multi-group model based on the pooled sample variance-covariance matrix for two sites is explored.

– The development of integrated approaches for the combination of process and spectral data based on multi-block and wavelet techniques is undertaken. Independent process monitoring schemes have been reported in many applications with the monitoring of processes using spectroscopy receiving increasing attention. The need to integrate different forms of data is emerging as a major research challenge as is the requirement to identify the additional benefits of such an approach. Data collected from an experiment conducted in a batch mini-plant was monitored by both on-line physical sensors and UV-Visible spectrometer and is used to investigate the different integration algorithms proposed in this thesis. It is observed that the application of integrating process and spectral data not only enables more straightforward interpretation, but also increases the scientific understanding of the process.

1.3 Layout of the Thesis

The following is a summary of the main components of the chapters in the thesis.

Chapter 1 has provided an introduction to the motivation and the objectives for undertaking research into enhanced process understanding through the application of batch process performance monitoring techniques and data integration. The contributions of the thesis were also identified and briefly reviewed.

Chapter 2 presents an overview of the underpinning methodologies of MSPC – PCA and PLS. A description of the algorithms is given, along with a discussion of the associated metrics regarding the development of a model representation. Particular attention is given to PCA since it is the primary underpinning technique of the thesis. The extensions of PCA – multiway PCA and multi-group PCA are then introduced for the development of batch process monitoring schemes. Both off-line analysis and on-line through batch monitoring are described.

In Chapter 3, the different aspects relating to the development of batch process performance monitoring are studied and evaluated. The issues of pre-treatment of three-way data are discussed since it is a core component in the development of a monitoring scheme. The two key existing monitoring approaches are then explained. Due to some of their limitations, a novel monitoring scheme is proposed and its performance in comparison with existing approaches is evaluated through a study on a penicillin simulation package – Pensim. The performance is evaluated through two MSPC monitoring indices.

Chapter 4 investigates the application of the methodologies described in Chapter 2 and Chapter 3 to a pharmaceutical process. The applicability of the multi-group algorithm to address the multi-site batch monitoring of a pharmaceutical process which is manufactured at two different sites is investigated. A number of approaches are proposed to address this research challenge.

Chapter 5 introduces the concept of two advanced methodologies that can facilitate the integration of different forms of data – multi-block PCA and wavelet analysis. Both techniques are reviewed with specific focus on their applications to the area of process monitoring. Another area of increasing importance is that of process spectroscopy. The characteristics and measurement techniques of different forms of spectroscopy are described and their potential for on-line analysis is highlighted. Spectral pre-treatment methods are then reviewed. Finally, a literature review on existing data integration methods is presented.

The aim of the work presented in Chapter 6 is to apply the techniques of multi-block PCA, and in conjunction with wavelet analysis, to investigate their applicability for the combination of engineering process and spectral data in a pharmaceutical process. An experimental design was carried out to generate the data from a number of batch runs in a mini-plant experimental set up. The nominal models were built from the batches conducted under normal operating conditions and the performance of the individual process and spectral data models were then compared against the two integrated models – multi-block PCA model and the conjunction of wavelet analysis and multi-block PCA. Two batches containing changes in the chemistry as well as physical disturbances were investigated to assess the performance of the different approaches in the presence of process deviations.

Chapter 7 summaries the key results and defines areas for future work.

Chapter 2

Methodologies for Batch Process Performance Monitoring

2.1 Introduction

In today's manufacturing and processing environment, batch and semi-batch processes play an important role in the production of high value added products. Industries where batch processes are widely utilised include the pharmaceutical, speciality chemical, food, polymer, semi-conductor and bio-chemical. Typically, the manufacture of a batch process involves the charging of a specific recipe of materials to a vessel, processing it under controlled conditions and then discharging the final product. On completion of a batch, a number of measurements are recorded in the quality control laboratory. The batch is considered successful if all the quality parameters lie within pre-defined specification limits. This procedure is termed off-line analysis since a sample of product is taken at the end of production for testing and the quality of the product is unknown until the final laboratory result is reported.

On the other hand, the monitoring of batch performance in real time contributes to ensuring that the batch is progressing according to the desired trajectory thereby leading to a high quality product and where a deviation is detected, it can be corrected in an appropriate manner to avoid the loss of the batch. In other words, quality is measured and controlled in real time to eliminate the lag of uncertainty.

The key factors to business success are product quality and consistency hence enhanced manufacturing performance and process understanding are critical to achieving consistently high-quality, right-first-time production. Multivariate Statistical Process Control (MSPC) (Kresta *et al.*, 1991) and more recently multivariate Six Sigma (Yang, 2004) are some of the tools that have been applied to achieve these objectives. MSPC and the associated techniques have started to be recognised by industry as valuable tools for enabling a deeper understanding of the process through the extraction of information from data and hence contribute to companies' business drivers.

In this chapter, an overview of the MSPC methodologies is presented. Initially the multivariate statistical projection methods of Principal Component Analysis (PCA) and Partial Least Squares (PLS), which form the basis of MSPC, are introduced along with the associated process performance representations and metrics. The methodologies are then extended to the multi-way and multi-group techniques which are applicable for the monitoring of batch processes.

2.2 Review of Batch Applications and Developments in Multivariate Statistical Process Control (MSPC)

Since the introduction of MSPC to the processing industries over the past two decades, the number of applications of MSPC for the monitoring of batch processes has increased significantly. A brief summary of some of the range of applications is presented in this section. As a consequence of the initial research into MSPC by Nomikos and MacGregor (1994), the philosophy of MSPC was extended to batch processes, more specifically Multiway Partial Least Squares (MPLS) (Nomikos and MacGregor, 1995a). In both papers, applications based on a simulation of a semi-batch reactor for the production of styrene-butadiene latex were reported. Other studies included the application of Multiway Principal Component Analysis (MPCA) to an industrial batch polymerisation reactor where an on-line monitoring scheme was developed (Nomikos and MacGregor, 1995b). A further seminal paper in this area has included a tutorial on the multivariate statistical methods of MPCA based on its application to three industrial case studies (Kourti and MacGregor, 1995): a historical data review of the catalytic cracking of crude oil, monitoring of a continuous polymerisation process and a batch polymerisation process.

A further reported implementation of MSPC was reported by Gallagher *et al.* (1996), who applied MPCA for the monitoring of a nuclear waste storage tank. An application to a two-stage batch polymerisation process for improved process understanding was presented by Kosanovich *et al.* (1996) whilst Neogi and Schlags (1998) applied MPCA and MPLS to an emulsion batch polymerisation process. In 1996, Martin and Morris (1996) proposed non-parametric confidence bounds for process performance monitoring charts and then applied the concept to the monitoring of a batch methyl methacrylate polymerisation reactor. Albert and Kinley (2001) applied MSPC techniques to a fermentation process whilst more recently, Garcia-Munoz *et al.* (2003) reported another successful application of MSPC for the monitoring of an industrial batch drying process.

Other reported developments have included the work of Tates *et al.* (1999) who presented an application relating to the monitoring of a poly-vinyl chloride batch process where batch contribution plots were used for fault detection and identification. Martin and Morris (2002) described a pharmaceutical application that focussed on the production of an active pharmaceutical ingredient where there were a limited number of batches available and different sets of operating conditions. In 2002, a framework for the monitoring of multi-stage, multi-phase processes was developed and applied to a pharmaceutical granule production (Undey and Cinar, 2002).

A new on-line monitoring strategy using an updated MPCA model for a simulated fed-batch fermentation process was proposed by Lee *et al.* (2003) and in 2004 Flores-Cerrillo and MacGregor (2004) developed a new method that takes account of information from previous batches and uses this information to optimise the current batch process.

2.3 Principal Component Analysis (PCA)

2.3.1 Background and Objectives

PCA was first reported by Pearson in 1901 (Pearson, 1901) and was described as "*finding lines and planes of closest fit to systems of points in space*". The present form of PCA was modified by Hotelling in 1933 (Hotelling, 1933) who proposed using the technique for analysing the covariance/correlation structure between a number of random variables. However its application was not wide-spread until the advent of more powerful computers. Since then, PCA has been applied in many diverse fields including chemistry, engineering, geology, psychology and sociology.

Application of the methodology to a multivariate dataset, **X**, results in the generation of new variables that are a linear combination of the original variables and are constrained to be mutually orthogonal (Jolliffe, 1986). Through the exclusion of those principal components associated with noise, a reduction in dimensionality of the problem is achieved and the remaining principal components characterise the main sources of variation. This multivariate projection technique is explored further in this chapter. Initially the methodology is presented in terms of the standard approach that is directly applicable to continuous processes prior to discussing its extension to batch processes.

2.3.2 Geometric Interpretation

The geometric concept of PCA is shown in Figure 2-1. The three axes represent three original variables and "o" denotes the original observations. The bold line defines the first principal component which is a line of best fit (in the least squares sense) to the three dimensional observations in the sample space and which captures the main source of variation in the data matrix **X**. The second bold line defines the second principal component which is orthogonal to the first principal component and describes the next greatest amount of variation in the data. The loading vectors define the location of the plane from the origin to the original variables. Each loading gives an indication of the importance of a specific variable in defining the orientation of the principal component. The location of each original observation on the plane is given by its

score vector. The scores are the distance from the origin of the plane along each principal component to the projection of the observation onto the principal component.

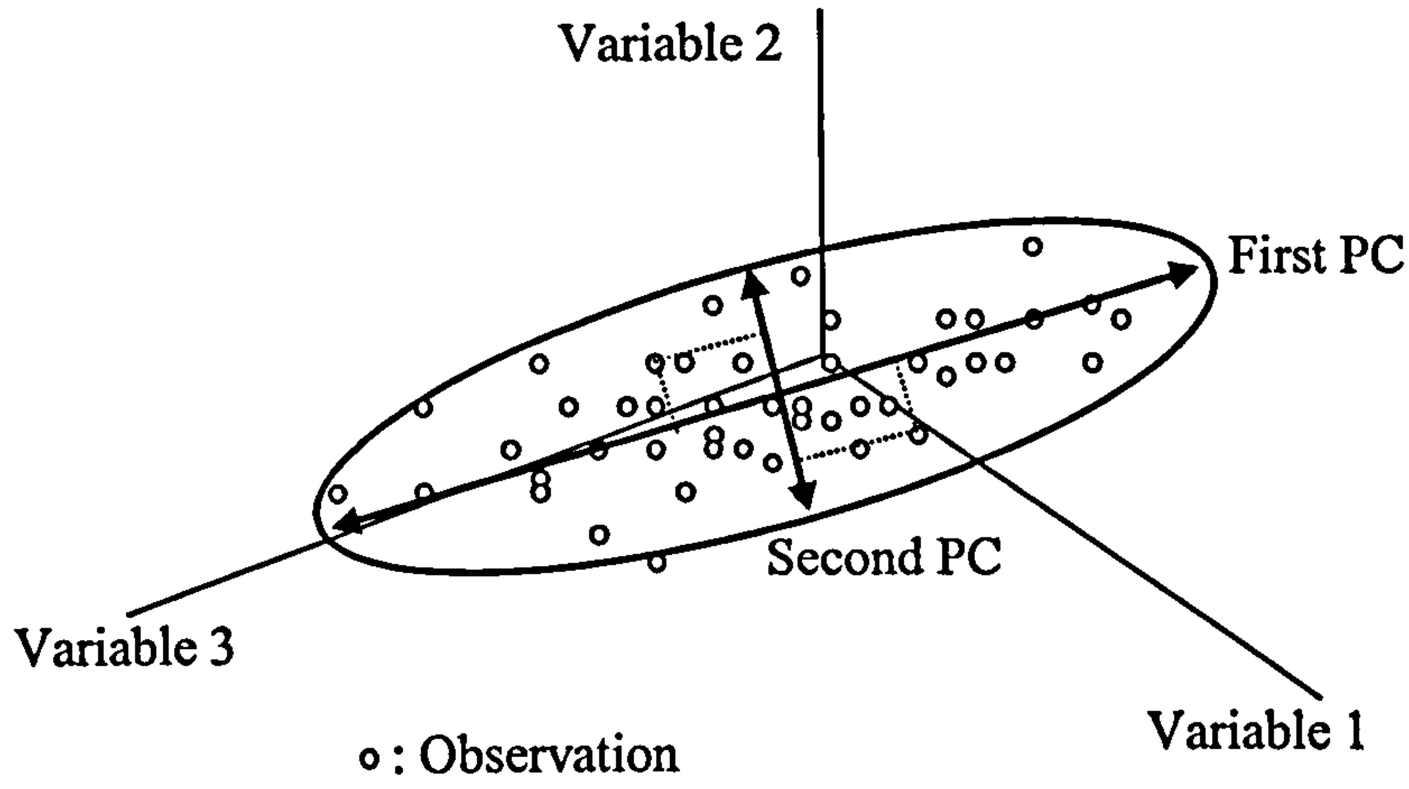


Figure 2-1 Graphical representation of principal component analysis

2.3.3 Mathematical Definition

PCA is a technique that linearly maps multi-dimensional data onto lower dimensional space with minimal loss of information. It can be considered as a variance maximisation technique (Jolliffe, 1986). A number of different methods can be applied to derive the principal components including Non-linear Iterative Partial Least Squares (NIPALS) (Geladi and Kowalski, 1986; Wold *et al.*, 1987) and Singular Value Decomposition (SVD) (Jolliffe, 1986). The NIPALS algorithm calculates the principal components sequentially whilst SVD derives all the components simultaneously.

Consider a data matrix **X** of order I x J where I is the number of samples and J is the number of variables. The covariance matrix of **X** is defined as:

$$\mathrm{cov}(\mathbf{X}) = \frac{\mathbf{X}^{\mathrm{T}}\mathbf{X}}{I-1} = \begin{pmatrix} c_{11} & \cdots & c_{1J} \\ \vdots & \ddots & \vdots \\ c_{1J} & \cdots & c_{JJ} \end{pmatrix} \qquad 2\text{-}1$$

where c_{jj^*} are the covariance between variables $\mathbf{x}_j$ and $\mathbf{x}_{j^*}$ ($j \neq j^*$) and the diagonal element c_{jj} is the variance of variable $\mathbf{x}_j$ ($j = 1, 2 \ldots J$). The variance explained by the individual principal components are the eigenvalues of the covariance matrix. Each column of the original matrix is normalised by either mean centring or auto-scaling. If auto-scaling is applied, Equation 2-1 gives the correlation matrix of **X**. Details of data scaling are given in Section 3.2.

PCA decomposes the data matrix **X** as a sum of the outer product of the vectors $\mathbf{t}_r$ and $\mathbf{p}_r$. The resulting decomposition can be subdivided into two parts, the internal structure (statistical model) and the extraneous structure (model residual):

$$\mathbf{X} = \mathbf{t}_1\mathbf{p}_1^T + \mathbf{t}_2\mathbf{p}_2^T + \ldots + \mathbf{t}_R\mathbf{p}_R^T + \mathbf{E} \qquad 2\text{-}2$$

$$\mathbf{X} = \sum_{r=1}^{R} \mathbf{t}_r\mathbf{p}_r^T + \mathbf{E} \qquad 2\text{-}3$$

where $\mathbf{t}_r$ is a vector of scores corresponding factors or components r, $\mathbf{p}_r$ is a loading vector, **E** is the residual matrix and R is the number of principal components included in the model which is less than or equal to the smaller dimension of **X**, i.e. $R = \min\{I, J\}$. The scores vectors, $\mathbf{t}_r$, contain information on how the samples (observations) are related to each other, whilst the loading vectors, $\mathbf{p}_r$, describe how the variables are inter-related. The loading vectors are the eigenvectors of the covariance matrix:

$$\mathrm{cov}(\mathbf{X})\mathbf{p}_r = \lambda_r\mathbf{p}_r \qquad 2\text{-}4$$

where λ_r is the eigenvalue associated with the eigenvector $\mathbf{p}_r$. The principal component scores, $\mathbf{t}_r$ ($r = 1, 2 \ldots R$), form an orthogonal set ($\mathbf{t}_r^T\mathbf{t}_{r^*} = 0$ for $r \neq r^*$), while the loading vectors, $\mathbf{p}_r$, are orthonormal ($\mathbf{p}_r^T\mathbf{p}_{r^*} = 0$ for $r \neq r^*$, $\mathbf{p}_r^T\mathbf{p}_{r^*} = 1$ for $r = r^*$). The scores vector can be described by:

$$\mathbf{t}_r = \mathbf{X} \cdot \mathbf{p}_r . \qquad 2\text{-}5$$

Thus the scores vector $\mathbf{t}_r$ is a linear combination of the original variables in the data matrix **X**, where the coefficients are defined by the loading vector $\mathbf{p}_r$. The principal components are arranged in descending order according to the associated eigenvalue λ_r ($r = 1, 2 \ldots R$). λ_r is a measure of the amount of variance explained by each principal component. In this context, the variance should be regarded as information inherent in the process. With the principal components arranged in descending order of λ_r, the first principal component captures the largest amount of information and each subsequent principal component captures the next greatest amount of variance.

The NIPALS algorithm decomposes the data in a manner such that the principal components are calculated sequentially. The algorithm is as follows with an arrow schematic given in Figure 2-2.

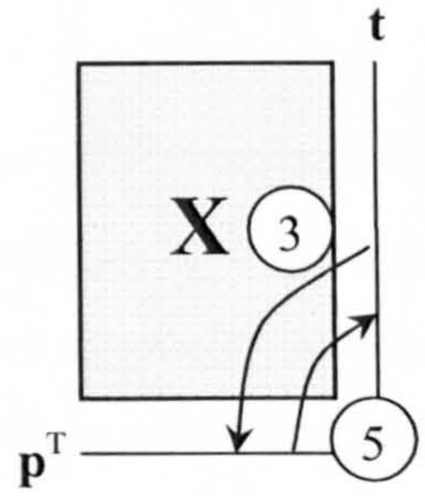


Figure 2-2 Arrow schematic of NIPALS PCA method

1. Centre and scale the data as appropriate. Set $r = 1$.
2. For each dimension r (principal component), select a vector $\mathbf{t}_r$ from $\mathbf{X}$.
3. Calculate the loading vector $\mathbf{p}_r$: $\mathbf{p}_r = \dfrac{\mathbf{X}^T \cdot \mathbf{t}_r}{\mathbf{t}_r^T \cdot \mathbf{t}_r}$. 2-6
4. Normalise the loading $\mathbf{p}_r$ to unit length: $\mathbf{p}_r = \dfrac{\mathbf{p}_r}{\|\mathbf{p}_r\|}$. 2-7
5. Calculate the score vector $\mathbf{t}_r$: $\mathbf{t}_r = \dfrac{\mathbf{X} \cdot \mathbf{p}_r}{\mathbf{p}_r^T \cdot \mathbf{p}_r}$. 2-8
6. Check for convergence. If $\mathbf{t}_r$ has not converged, go to Step (3). Convergence can be determined by summing the squared difference between corresponding elements of $\mathbf{t}_{rNew}$ and $\mathbf{t}_r$ relative to a pre-defined tolerance.
7. Perform deflation step: $\mathbf{X}_{New} = \mathbf{X} - \mathbf{t}_r \cdot \mathbf{p}_r^T$. 2-9
8. Replace $\mathbf{X}$ by $\mathbf{X}_{New}$ and calculate the next dimension, i.e. $r = r + 1$, go to step (2).
9. Stop when maximum number of principal components have been calculated or when desired number of principal components have been calculated.

By calculating the principal components simultaneously, SVD is a more efficient algorithm than NIPALS where the factors are calculated sequentially. SVD is based on the decomposition of a symmetric matrix, for example, the correlation matrix:

$$\mathrm{cov}(\mathbf{X}) = \mathbf{U} \cdot \mathbf{W} \cdot \mathbf{P}^T \qquad 2\text{-}10$$

where **U** is the scores matrix but normalised to unit length; **W** is a diagonal matrix containing the square root of the eigenvalues and $\mathbf{P}^T$ is the loading matrix.

2.3.4 Selection of the Number of Principal Components

An important step in the application of PCA is to determine the number of principal components that are required to adequately capture the major sources of variation in the data set. A highly correlated set of variables usually requires only a few principal components to be included in the model with those principal components excluded typically describing the noise in the process. Including more components in the model than those that explain the main sources of variation does not necessarily result in a better representation of the process. One potential impact is that the sensitivity of the model is affected. A number of techniques have been proposed to select the number of principal components including metrics that consider the cumulative percentage of variance explained and cross-validation based criteria.

Recall that each individual principal component is defined as a linear combination of the variables that explain the variation remaining after fitting the previous components. Therefore the percentage of total variance explained is an important criterion. The total variance can be calculated as follows:

$$\nu = \|\mathbf{X}\|^2 . \qquad 2\text{-}11$$

By fitting a model, matrix **X** can be expressed as:

$$\mathbf{X} = \hat{\mathbf{X}} + \mathbf{E} \qquad 2\text{-}12$$

where $\hat{\mathbf{X}}$ describes the systematic part and **E** denotes the residual. Thus the variance explained by the model is given by:

$$\hat{\nu} = \|\hat{\mathbf{X}}\|^2 \qquad 2\text{-}13$$

and the criterion for the percentage of variation explained by the model is given by:

$$p : 0.7 \le \frac{\hat{\nu}}{\nu} \le 0.9 . \qquad 2\text{-}14$$

The p value which was suggested by Jolliffe (1986) should typically lie between 0.7 and 0.9. However this rule is not always appropriate for some data sets where the level of noise is greater than 30% or less than 10%. Table 2-1 provides an example of a variance table for a data set which has five variables. If the p value rule is followed, two principal components should be included in the model, explaining 90.7% of the total variance.

Component	Eigenvalue (λ)	% variance explained	Cumulative % variance explained
1	3.352	67.05	67.05
2	1.182	23.65	90.70
3	0.285	5.70	96.40
4	0.135	2.70	99.10
5	0.045	0.90	100.00

Table 2-1 Eigenvalues and explained variances for a data set comprising five variables

Another popular technique is cross-validation (Wold, 1978). In the simplest case, every row (sample) in the data matrix is removed from the data set once and a model is computed with the remaining samples. The removed data is then predicted using the calculated PCA model and the Predicted Error Sum of Squares (PRESS) for the omitted sample is calculated. The optimal number of principal components is defined as that which gives the minimum residual error.

More specifically, the first step is to divide the data into a number of groups (G). For example if the data set contains 100 observations, then the data can be split into 4 groups of 25 observations. The next step is to exclude one group from the data and build a PCA model from the remaining groups, i.e. group 4 (*G4*) is excluded from the data and PCA is applied to the remaining groups *G1*, *G2* and *G3*. The first principal component is then used to calculate the principal component scores for the excluded group:

$$\mathbf{t}_{new_{G4}} = \mathbf{x}_{G4} \cdot \mathbf{p}_{(G1.G3)} \qquad 2\text{-}15$$

where $\mathbf{t}_{new_{G4}}$ is the principal component scores for group *G4* and $\mathbf{p}_{(G1:G3)}$ is the loading vector for the first principal component calculated from groups *G1*, *G2* and *G3*. The principal component scores are then used to estimate the values in group *G4*.

$$\hat{\mathbf{x}}_4 = \mathbf{t}_{new_{G4}} \cdot \mathbf{p}^{\mathrm{T}}_{(G1:G3)} \qquad 2\text{-}16$$

where $\hat{\mathbf{x}}_4$ is a vector of the estimates of the samples in group *G4* and $\mathbf{p}^T_{(G1:G3)}$ is the transpose of the loadings for principal component one. From this estimate, the PRESS is calculated:

$$\mathbf{E}_4 = \sum_{i=1}^{25} (\mathbf{x}_{4i} - \hat{\mathbf{x}}_{4i})^2 . \qquad 2\text{-}17$$

This procedure is repeated until every individual group has been left out once and the PRESS for each excluded group are then summed to give the total PRESS.

$$\text{Total PRESS} = \mathbf{E}_1 + \mathbf{E}_2 + \mathbf{E}_3 + \mathbf{E}_4 . \qquad 2\text{-}18$$

This procedure is repeated for an increasing number of principal components and the corresponding total PRESS is calculated. Figure 2-3 illustrates a typical plot for the total PRESS. In this case as the number of dimensions approach the number of significant principal components, the total PRESS decreases whilst by including additional principal component, that reflect noise in the data, a decrease in the goodness-of-fit of the sample prediction results. In this case five principal components would be recommended for selection.

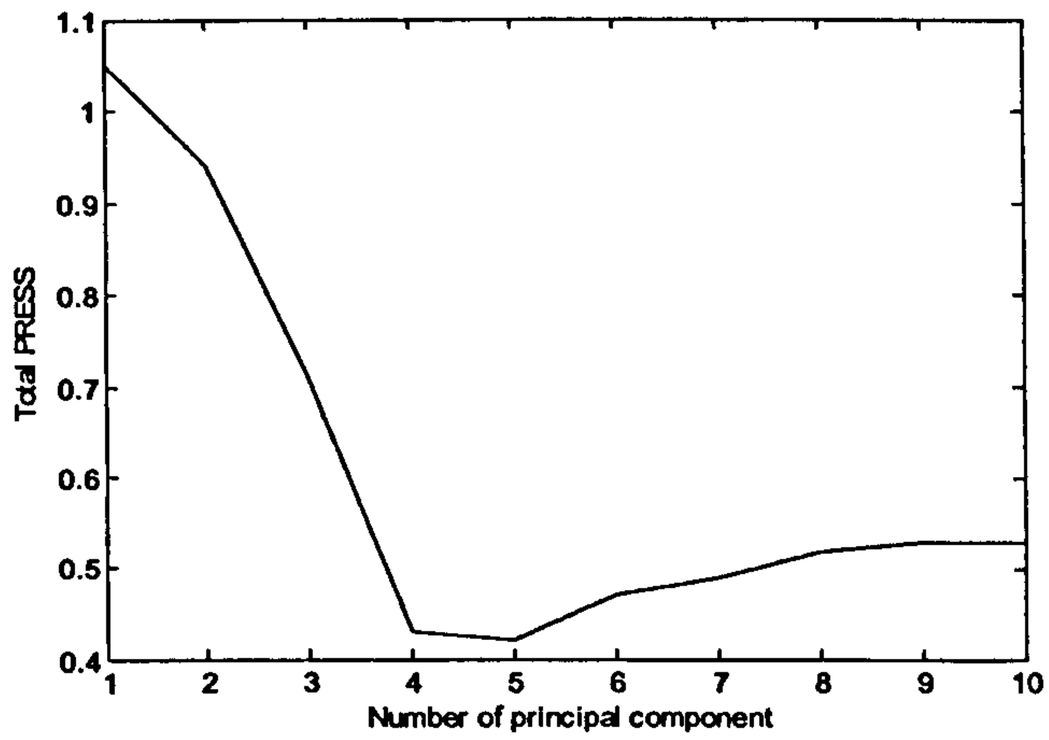


Figure 2-3 An example of the application of cross-validation

2.4 Process Performance Representations

2.4.1 Principal Components Scores and Loadings

As mentioned in Section 2.3.3 the eigenvectors of the covariance matrix are the loadings of the principal components and the principal component scores can be derived from Equation 2-5. Plots

of the principal components scores and loadings can illustrate the underlying relationship between the observations and the variables respectively.

The principal component scores plot can be represented as either a univariate or a bivariate representation. The control limits for the scores plots are derived from the statistical properties of the nominal data used to construct the process model. The control limits encapsulate the common cause variation present in the process and it is therefore assumed that the scores are normally distributed and independent. The control limits are given by:

$$\pm t_{I-1,\alpha/2} \cdot s_{est} \cdot \sqrt{1+\frac{1}{I}} \qquad 2\text{-}19$$

where I is the number of samples (batches), α is the significance level, s_{est} is the estimated standard deviation of the scores for a single principal component for all samples and $t_{I-1,\alpha/2}$ is the value of the Student's t-distribution with $I-1$ degrees of freedom. Typically α takes the values of 0.05 and 0.01 denoting the warning and action limits respectively.

Figure 2-4(a) provides an example of a univariate scores plot and Figure 2-4(b) shows a bivariate scores plot. Most of the samples lie within the confidence intervals with a few falling outside the 95% and 99% confidence limits as expected since when the number of samples is approaching infinity, 5% and 1% of samples are expected to lie outside the limits by chance respectively.

An example of a univariate and bivariate principal component loadings plots is shown in Figure 2-5. Analogous to the scores plot, similarities between variables can be visualised in a loading plot. For those variables that are positively correlated, they will fall in the same quadrant whilst those that are negatively correlated lie in diametrically opposite quadrants. However this behaviour must be reflected for all retained principal components. For example, variables 10 and 1 appear to be negatively correlated with variable 12. Likewise variable 7 is negatively correlated with variables 2 and 13. However this behaviour must be evident for the other retained principal components.

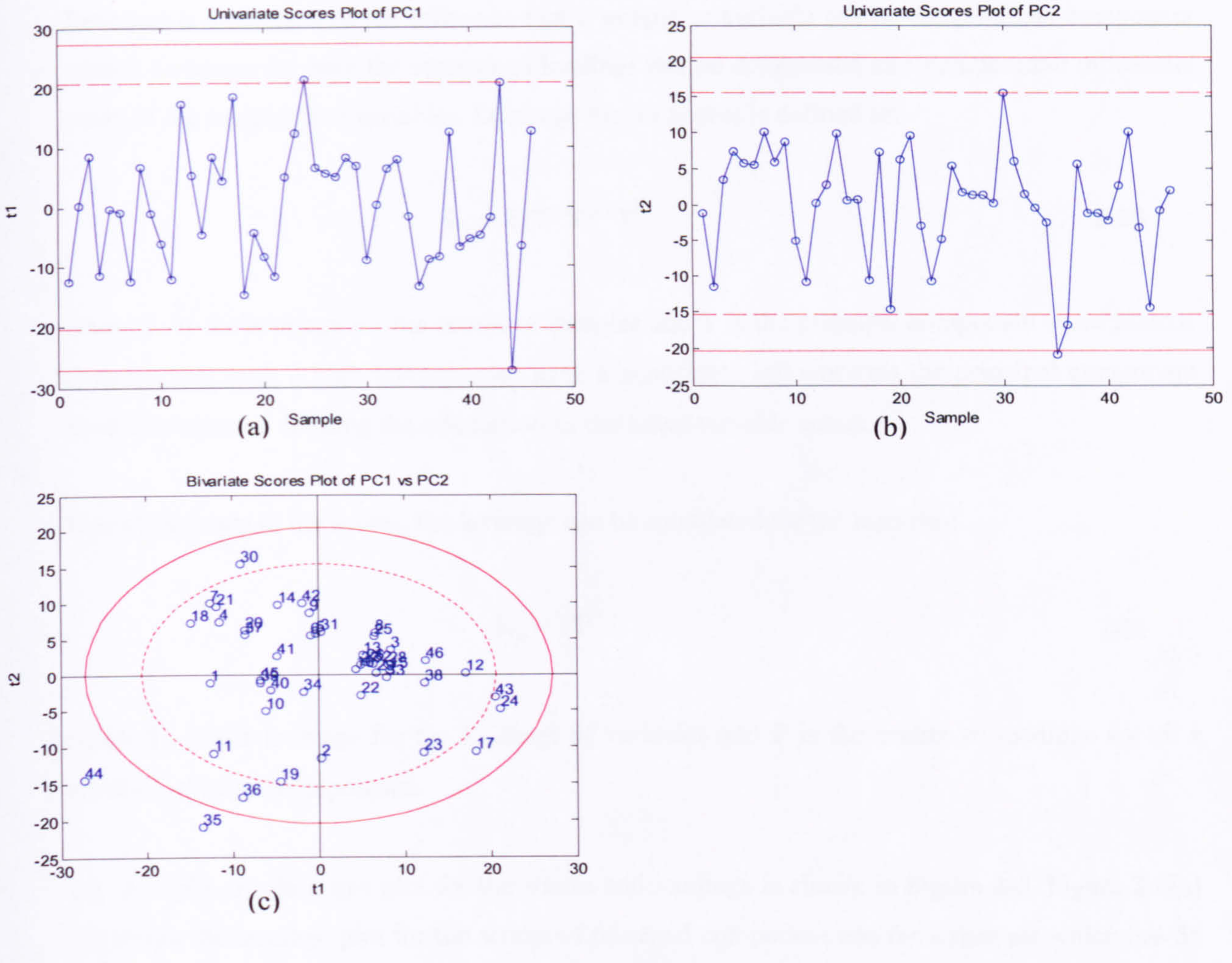


Figure 2-4 An example of (a) univariate scores plot of principal component one; (b) univariate scores plot of principal component two; (c) bivariate scores plot; key: "o" – principal component score; "---" – 95% warning limit; "—" – 99% action limit

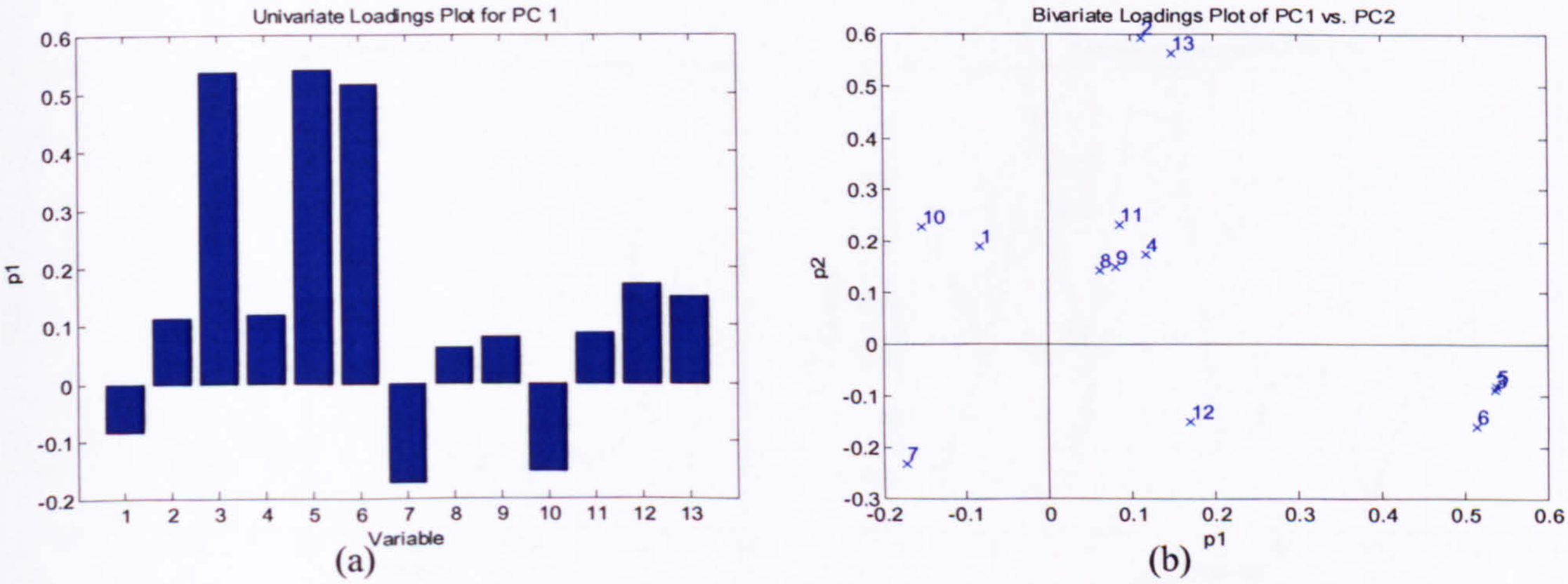


Figure 2-5 An example of (a) univariate loadings plot; (b) bivariate loadings plot

2.4.2 Leverage

Leverage is a measure of the influence that a sample or variable has on the principal component model. Leverage for both the scores and loadings can be determined as a comparative influential study of the samples and variables. Leverage for the scores is defined as:

$$\mathbf{L}_t = \mathbf{T}(\mathbf{T}^{\mathrm{T}}\mathbf{T})^{-1}\mathbf{T}^{\mathrm{T}} \qquad 2\text{-}20$$

where $\mathbf{L}_t$ is the leverage for the scores of samples and $\mathbf{T}$ is the principal component score matrix. Observations with a high leverage can have a significant influence on the principal component model in terms of defining the orientation of the latent variable space.

Complementary to the scores, the leverage can be calculated for the loadings:

$$\mathbf{L}_p = \mathbf{P}\mathbf{P}^{\mathrm{T}} \qquad 2\text{-}21$$

where $\mathbf{L}_p$ is the leverage for the loadings of variables and $\mathbf{P}$ is the matrix of loadings for all r retained principal components.

An example of a leverage plot for the scores and loadings is shown in Figure 2-6. Figure 2-6(a) illustrates the leverage plot for the scores of principal component one for a data set which has 56 samples. Sample 6 is shown to have higher leverage than the other samples. Figure 2-6(b) shows the leverage plot for the loadings of principal component one for the same data set which has 13 variables. Variables 3, 5 and 6 are shown to have high leverage which may be the rational for the high leverage of sample 6.

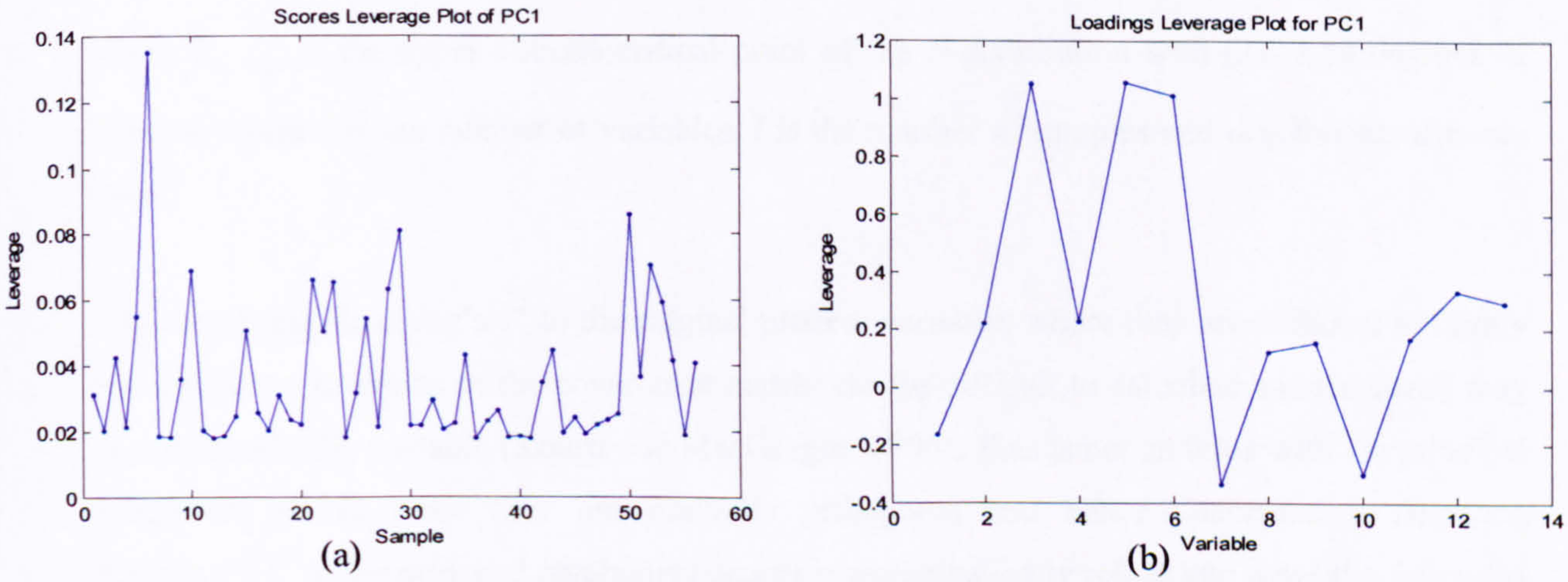


Figure 2-6 An example of a leverage plot for (a) scores; (b) loadings

2.4.3 Hotelling's T^2

A statistical metric that captures the behaviour of the retained principal components is that of Hotelling's T^2 (Hotelling, 1947). Hotelling's T^2 is directly related to the Mahalanobis distance (Mardia *et al.*, 1979) which is a distance metric that takes into account the variance-covariance matrix of the data. If the in-control variance-covariance matrix $\mathbf{\Sigma}$ is known, Hotelling's T^2 is defined as:

$$T^2 = (\mathbf{x} - \bar{\mathbf{x}})^T \Sigma^{-1} (\mathbf{x} - \bar{\mathbf{x}}) \qquad 2\text{-}22$$

where $\mathbf{x}$ is a vector of data and $\bar{\mathbf{x}}$ is the vector of mean values. Hotelling's T^2 follows a central χ^2 distribution with J degrees of freedom where J is the number of variables. A multivariate χ^2 chart can then be constructed by plotting the Mahalanobis distance versus time or sample number with an upper control limit (UCL) given by $\chi^2_{J,\alpha}$ where α is an appropriate significance level, i.e. $\alpha =$ 0.05 or 0.01 for the warning and action limits, respectively. However if $\mathbf{\Sigma}$ is unknown as is typically the case, it is more appropriate to estimate the following form:

$$T^2 = (\mathbf{x} - \bar{\mathbf{x}})^T \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}}) \qquad 2\text{-}23$$

where $\mathbf{S}$ is the estimated in-control variance-covariance matrix calculated from historical samples. An upper control limit T^2_{UCL} is then obtained based on the F-distribution (Tracy *et al.*, 1992):

$$T^2_{UCL} \sim \frac{J(I-1)}{I-J} F_{J,I-1,\alpha} \qquad 2\text{-}24$$

where $F_{J,I-1,\alpha}$ is the upper $100\alpha\%$ critical point of the F-distribution with $(J, I-1, \alpha)$ degrees of freedom where J is the number of variables, I is the number of samples and α is the significance level.

When applying Hotelling's T^2 to the original process variables where they are collinear or highly correlated, the inversion of the covariance matrix can be difficult to calculate and the result may be mathematically unstable (Kourti and MacGregor, 1996). This is not an issue with the principal component scores since they are mutually orthogonal and hence uncorrelated. Applying Hotelling's T^2 to the principal component scores is mathematically robust and takes the following form:

$$T^2 = \mathbf{t} \cdot \mathbf{\Lambda}^{-1} \cdot \mathbf{t}^{\mathrm{T}} = \sum_{r=1}^{R} \frac{\mathbf{t}_r^2}{\lambda_r} \qquad 2\text{-}25$$

where $\mathbf{\Lambda}$ is a diagonal matrix containing the r largest eigenvalues of the covariance matrix, $\mathbf{t}_r$ is a vector of the principal component scores for dimension r and λ_r is the eigenvalue of the covariance matrix for dimension r. An example of a Hotelling's T^2 monitoring chart is shown in Figure 2-7. Two samples lie outside the action limit hence consideration would need to be given to the implementation of corrective action. The methodology to tackle this issue is presented in Section 2.4.5.

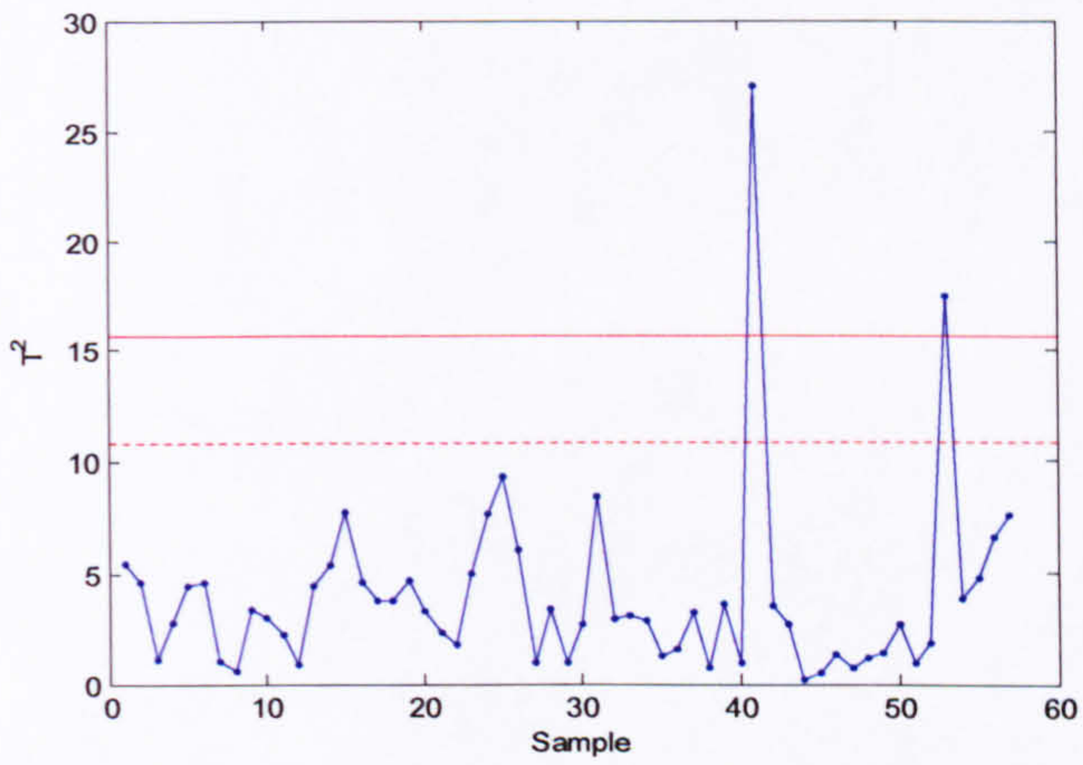


Figure 2-7 Example of a Hotelling's T^2 monitoring chart

2.4.4 Squared Prediction Error

Hotelling's T^2 can be used to monitor common cause variation within the space of the retained principal components. A second statistic that monitors the residual space is the Squared Prediction Error (SPE) or Q-statistic. It is the squared difference between the observed and the predicted values from the nominal representation and is defined as:

$$SPE_i = \sum_{j=1}^{J} (\mathbf{x}_{ij} - \hat{\mathbf{x}}_{ij})^2 \qquad 2\text{-}26$$

where J is the number of variables, $\mathbf{x}_{ij}$ is the original value and, $\hat{\mathbf{x}}_{ij}$ is the estimate of observation i for variable j and is given by:

$$\mathbf{t}_{ir} = \sum_{j=1}^{J} \mathbf{x}_{ij} \cdot \mathbf{p}_{jr} \qquad r = 1, 2 \ldots R \qquad 2\text{-}27$$

$$\hat{\mathbf{x}}_{ij} = \sum_{r=1}^{R} \mathbf{t}_{ir} \cdot \mathbf{p}_{rj} \qquad 2\text{-}28$$

where $\mathbf{t}_{ir}$ is the vector of scores for principal component r for observation i.

The Q-statistic can also be defined as the squared perpendicular distance of a multivariate observation from the reduced principal components space. Based on the assumption that the original data set is normally distributed, the control limits for the SPE can be derived as follows (Jackson and Mudholkar, 1979):

$$Q_\alpha = \theta_1 \left(\frac{z_\alpha \sqrt{2 \cdot \theta_2 \cdot h_0^2}}{\theta_1} + \frac{\theta_2 \cdot h_0 \cdot (h_0 - 1)}{\theta_1^2} + 1 \right)^{\frac{1}{h_0}} \qquad 2\text{-}29$$

where

$$\theta_i = \sum_{k=R+1}^{J} \lambda_k^i \qquad i = 1, 2, 3 \qquad 2\text{-}30$$

and

$$h_0 = 1 - \frac{2 \cdot \theta_1 \cdot \theta_3}{3 \cdot \theta_2^2} \qquad 2\text{-}31$$

z_α is the standard normal deviate corresponding to the upper ($1 - \alpha$) percentile, λ_k is the eigenvalue of the residuals, R is the number of principal components retained in the model and J is the number of variables. Box (1954) previously showed that:

$$Q_\alpha = f \cdot \chi^2_{h,\alpha} \qquad 2\text{-}32$$

where

$$f = \frac{\theta_2}{\theta_1} \qquad \text{and} \qquad h = \frac{\theta_1^2}{\theta_2}.$$

More recently, Nomikos and MacGregor (1995b) demonstrated that the two approximations, Equations 2-29 and 2-32, are equivalent. The Q value is the Euclidean distance of the position of a sample from the hyper-plane formed by the PCA representation. An example of a SPE monitoring chart is shown in Figure 2-8. Sample 53 lies outside the 99% action limit. A high SPE value is attained when the covariance structure in the data matrix does not follow the estimated

variance-covariance matrix and hence non-conforming behaviour infers a change in the nominal covariance structure.

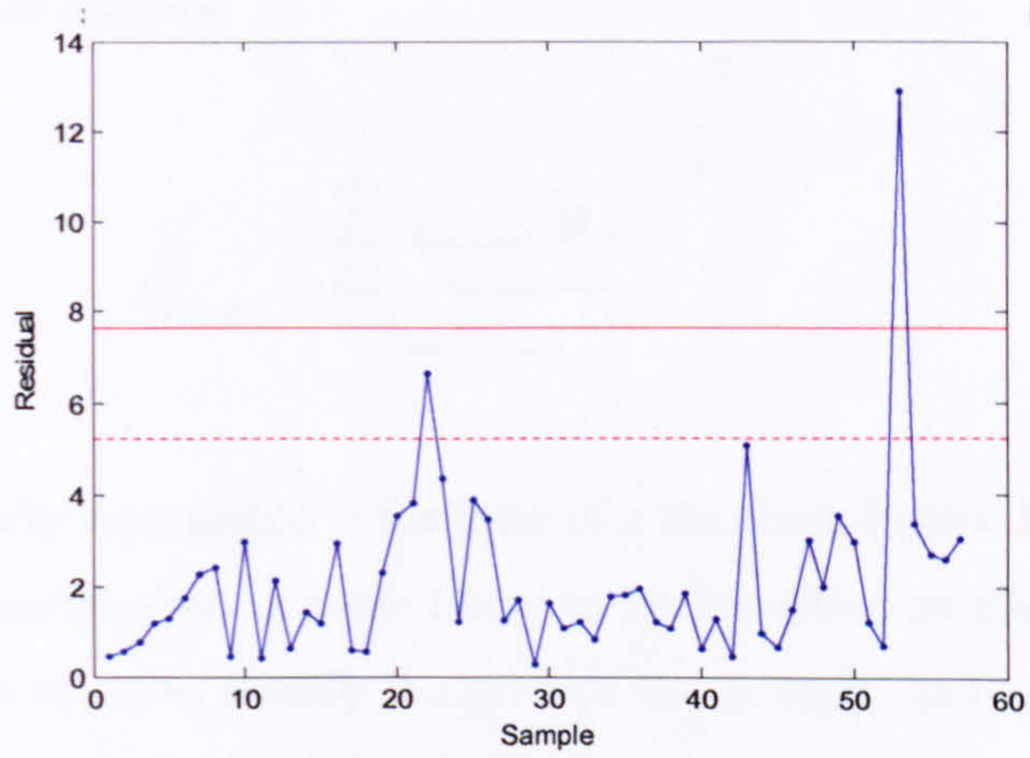


Figure 2-8 Example of a SPE monitoring chart

2.4.5 Contribution Plot

Once the multivariate statistical tools and metrics recognise a process to be out of statistical control, further analysis has to be performed to identify those variables indicative of the change. The early identification of such variables is crucial for on-line monitoring schemes, to ensure consistency of production and the safety of the plant. Therefore, the development of fault process identification tools is essential. One of the most commonly implemented methods is the contribution plot (Kourti and MacGregor, 1996; Miller *et al.*, 1998; Westerhuis *et al.*, 2000a). The method is based on calculating the contribution of the individual variables to the principal component score, the Hotelling's T^2 or the SPE for the non-conforming sample or the period of the deviation.

The principal component scores can be considered as a weighed sum of the process variables:

$$\mathbf{t}_{ir} = \sum_{j=1}^{J} \mathbf{x}_{ij} \cdot \mathbf{p}_{jr} = \mathbf{x}_{11}\mathbf{p}_{11} + \mathbf{x}_{12}\mathbf{p}_{21} + \ldots + \mathbf{x}_{1N}\mathbf{p}_{N1} \qquad 2\text{-}33$$

where $\mathbf{p}_{jr}$ is the loading vector for principal component r for variable j and $\mathbf{x}_{ij}$ is the vector for sample i for variable j. The J terms reflect the contribution of each of the j variable to the score $\mathbf{t}_{ir}$. Therefore for sample i, the contribution to the score for principal component r for variable j is given by:

$$c_{ir}^{t} = \mathbf{x}_{ij} \cdot \mathbf{p}_{jr} . \qquad 2\text{-}34$$

If interest is in a group of observations from i_{start} to i_{end}, then the contribution is calculated as the mean over the number of samples:

$$c_{i_{start:end},r}^{t} = \frac{\sum_{j=1}^{J} \mathbf{x}_{i_{start:end},j} \cdot \mathbf{p}_{jr}}{i_{end} - i_{start}} . \qquad 2\text{-}35$$

These values are typically represented in the form of a bar chart, Figure 2-9(a). It was conjectured that a large variable contribution is more likely to be identified as the main source of a fault therefore action can be taken to modify the process accordingly. In Figure 2-9(a), variable 5 is observed to have a large score contribution for the first principal component.

In the work of Kourti and MacGregor (1996) the contribution plot was proven to be an effective identification tool for fault detection but confidence limits were not considered to be an issue since the magnitude and direction of the contributions were considered to assist in the identification of the non-conforming variables. However Conlin *et al.* (2000) proposed a number of approaches to obtaining confidence limits for the individual variable contributions. The underlying premise was that the data followed a normal distribution and hence the following limits were applicable:

$$c^{t} \pm z_{\alpha/2} \cdot \hat{\sigma}_{c} \qquad 2\text{-}36$$

where $z_{\alpha/2}$ is the corresponding standard normal deviate and $\hat{\sigma}_{c}$ is the estimated standard deviation. The most direct method for estimating $\hat{\sigma}_{c}$ for a specific variable is to obtain the contributions for each sample for a specific variable for a particular principal component and then calculate the standard deviation of the contributions. The method works well with the only limitation being the assumption of normality. Additionally an absolute contribution is used for easier graphical interpretation. Figure 2-9(b) illustrates the same contributions as for Figure 2-9(a) with the addition of 95% and 99% confidence limits in its absolute contribution form. Variable 5 was interpreted previously to have a high contribution from Figure 2-9(a), however it falls within the confidence bounds. In this case all variables are within the statistical confidence bounds.

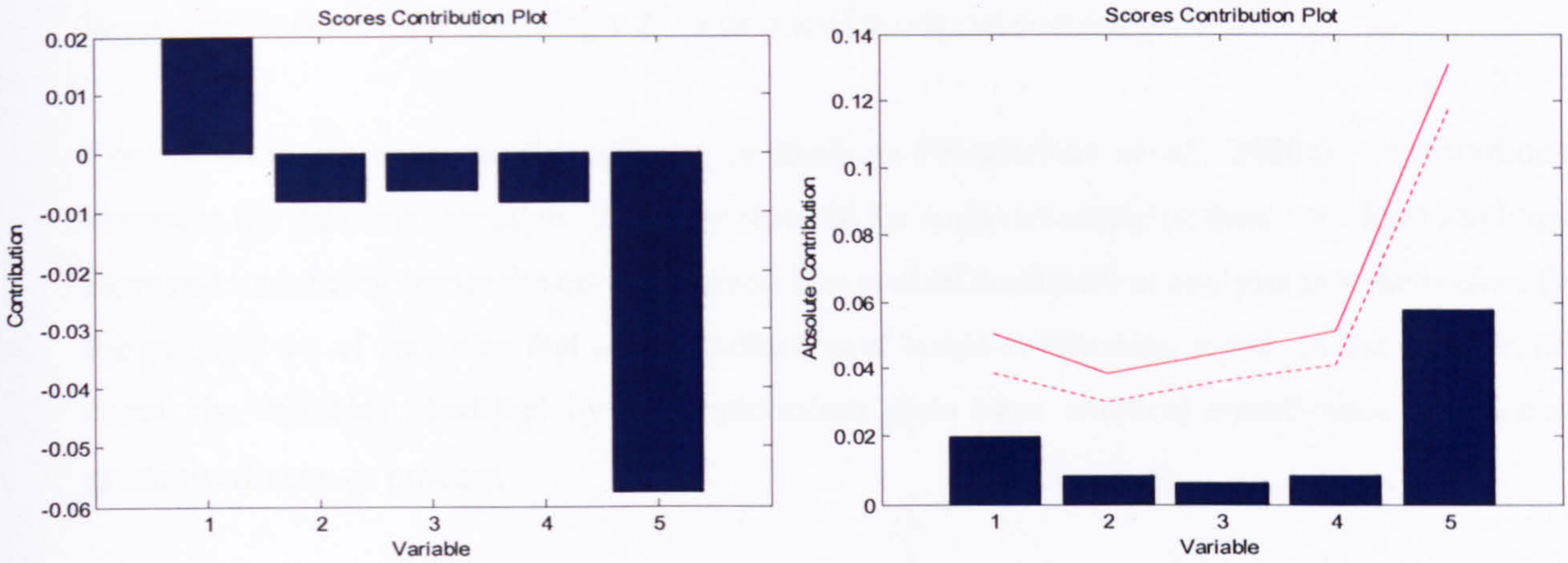


Figure 2-9 Example of a scores contribution plot: (a) no confidence limits; (b) with confidence limits in absolute contribution

In a similar manner, when a process deviation is detected from the Squared Prediction Error, the contribution plot for the SPE should be applied to identify the non-conforming variables:

$$c_{ij}^{SPE} = (\mathbf{x}_{ij} - \hat{\mathbf{x}}_{ij})^2 \qquad 2\text{-}37$$

where $\mathbf{x}_{ij}$ and $\hat{\mathbf{x}}_{ij}$ are the observed and predicted values for j variables, respectively.

The confidence limits for the SPE contributions are calculated in the same way as for the SPE (Section 2.4.4). For the Jackson and Mudholkar approach (1979), the variance-covariance matrix of the residuals from all the variables is replaced by the covariance of the residuals from variable j.

The Hotelling's T^2 contribution to each process variable can be calculated as proposed by Nomikos (1996):

$$c_{ij}^{T^2} = \sum_{r=1}^{R} \mathbf{\Lambda}_{rr}^{-1} \cdot \mathbf{t}_{ir}^{\mathrm{T}} \cdot \mathbf{x}_{ij} \cdot \mathbf{p}_{jr} \qquad 2\text{-}38$$

where $c_{ij}^{T^2}$ is the contribution summed over the R retained principal components, $\mathbf{\Lambda}_{rr}^{-1}$ is a diagonal matrix, $\mathbf{t}_{ir}$ is the score vector for sample i, $\mathbf{x}_{ij}$ is the vector for sample i and $\mathbf{p}_{jr}$ is the loading vector for variable j. The upper control limit (UCL) for the contributions of each variable is calculated as the mean of the contributions plus three times the standard deviation of the

contributions for each variable for all sample. The lower control limit is not considered since only large contributions infer Hotelling's T^2 to be out of statistical control.

Contributions are conceptually different to loadings (Westerhuis *et al.*, 2000a). Contributions represent the particular variables that were unusual for a given sample or samples whilst loadings represent variability across the entire data set. The goal of contribution analysis is to help identify the possible set of variables that are an indication of a non-conforming event. In the majority of cases, the variables identified by the contribution plots have practical significance which can guide the diagnosis process.

2.5 Partial Least Squares (PLS)

Partial Least Squares or Projection to Latent Structures (PLS), is a multivariate statistical regression technique. It is based on the analysis of two data blocks, the input matrix or independent block **X** and the output **Y**, or dependent variables. As for other regression techniques such as Multiple Linear Regression and Principal Component Regression, the coefficients β are derived using a least squares approach. The fundamental model for all regression techniques is:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{E}. \qquad 2\text{-}39$$

More specifically PLS models the relationship between **X** and **Y** by constructing a new set of variables referred to as latent variables where each latent variable is a linear combination of **X**. Consequently an approach similar to PCA is adopted for PLS whereby the dimensionality of both the process and quality space is reduced to a few pairs of latent variables. Thus given a data set of I samples with J process variables, $\mathbf{x}_j$, and L response variables, $\mathbf{y}_l$, i.e. $\mathbf{X}_{IxJ}$ and $\mathbf{Y}_{IxL}$ respectively, the PLS algorithm identifies the linear relationship between **X** and **Y** by performing ordinary least squares regression between each pair of input and output scores vectors, $\mathbf{t}_r$ and $\mathbf{u}_r$, respectively where R is the number of retained latent variables. Equation 2-39 is modified accordingly:

$$\mathbf{u}_r = \mathbf{t}_r d_r + \mathbf{e}_r \qquad r = 1, 2 \ldots R \qquad 2\text{-}40$$

where d_r is a coefficient and is obtained from the application of linear regression between the r-th vectors, $\mathbf{t}_r$ and $\mathbf{u}_r$:

$$d_r = \frac{\mathbf{u}_r^{\mathrm{T}} \cdot \mathbf{t}_r}{\mathbf{u}_r^{\mathrm{T}} \cdot \mathbf{u}_r}. \qquad 2\text{-}41$$

The decomposition is thus the product of each pair of input scores vectors, $\mathbf{t}_r$, and predicted output scores vectors, $\hat{\mathbf{u}}_r$, and a set of corresponding input and output loadings vectors $\mathbf{p}_r$ and $\mathbf{q}_r$. This provides an approximation model for the **X** variables and a prediction model for the **Y** variables:

$$\mathbf{X} = \sum_{r=1}^{R} \mathbf{t}_r \cdot \mathbf{p}_r^{\mathrm{T}} + \mathbf{E} \qquad 2\text{-}42$$

$$\mathbf{Y} = \sum_{r=1}^{R} \hat{\mathbf{u}}_r \cdot \mathbf{q}_r^{\mathrm{T}} + \mathbf{F} \qquad 2\text{-}43$$

where $\hat{\mathbf{u}}_r$ is the prediction of the scores $\mathbf{u}_r$ in an ordinary least squares sense and is given by:

$$\hat{\mathbf{u}}_r = \mathbf{t}_r \cdot d_r \qquad 2\text{-}44$$

E and **F** are the residual matrices for the **X** and **Y** matrices, respectively. A comprehensive review of PLS is given by Geladi (1986).

The most commonly used algorithm to perform PLS is Non-linear Iterative Partial Least Squares (NIPALS) (Wold, 1966). This algorithm sequentially extracts each pair of corresponding latent variables as a liner combination of the input and output variables prior to fitting the linear regression model. The prediction of the output scores is then evaluated. A summary of the algorithm is as follows and an arrow schematic is shown in Figure 2-10.

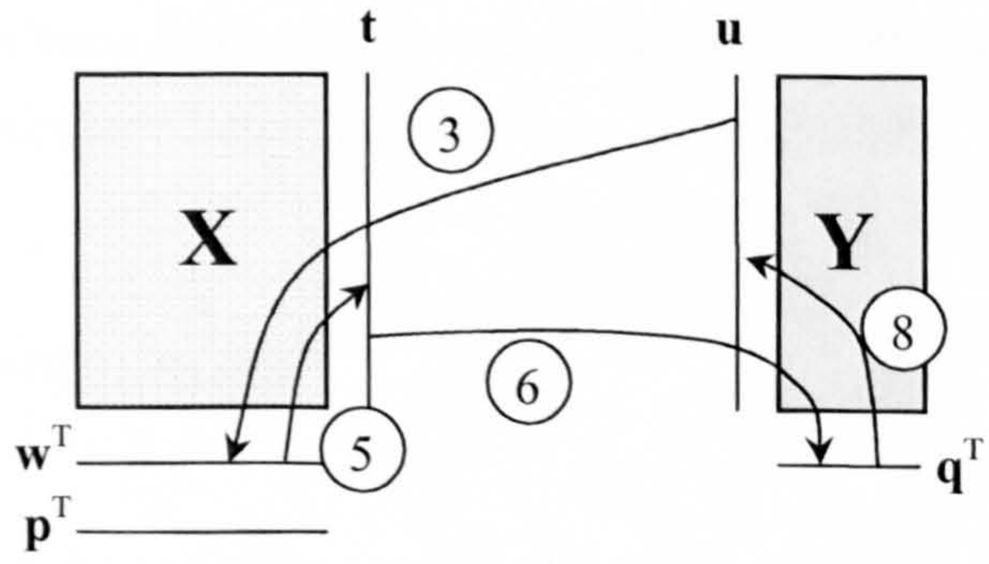


Figure 2-10 Arrow schematic of the NIPALS PLS method

1. Centre and scale **X** and **Y** matrices as appropriate. Set $r = 1$.
2. For each dimension r (latent variable), define the output scores **u** as a column of **Y**.
3. Regress **X** on $\mathbf{u}_r$: $\mathbf{w}_r^T = \frac{\mathbf{u}_r^T \cdot \mathbf{X}}{\mathbf{u}_r^T \cdot \mathbf{u}_r}$. 2-45
4. Normalise $\mathbf{w}_r$ to unit length: $\mathbf{w}_r = \frac{\mathbf{w}_r}{\|\mathbf{w}_r\|}$. 2-46
5. Calculate the input scores $\mathbf{t}_r$: $\mathbf{t}_r = \frac{\mathbf{X} \cdot \mathbf{w}_r}{\mathbf{w}_r^T \cdot \mathbf{w}_r}$. 2-47
6. Regress the columns of **Y** on $\mathbf{t}_r$: $\mathbf{q}_r^T = \frac{\mathbf{t}_r^T \cdot \mathbf{Y}}{\mathbf{t}_r^T \cdot \mathbf{t}_r}$. 2-48
7. Normalise $\mathbf{q}_r$ to unit length: $\mathbf{q}_r = \frac{\mathbf{q}_r}{\|\mathbf{q}_r\|}$. 2-49
8. Calculate the new output scores $\mathbf{u}_r$: $\mathbf{u}_r = \frac{\mathbf{Y} \cdot \mathbf{q}_r}{\mathbf{q}_r^T \cdot \mathbf{q}_r}$. 2-50
9. Check if the output scores $\mathbf{u}_r$ have converged. If Yes go to step (10) else go to step (3).
10. Calculate the **X** loadings: $\mathbf{p}_r^T = \frac{\mathbf{t}_r^T \cdot \mathbf{X}}{\mathbf{t}_r^T \cdot \mathbf{t}_r}$. 2-51
11. Regress $\mathbf{u}_r$ on $\mathbf{t}_r$: $d_r = \frac{\mathbf{u}_r^T \cdot \mathbf{t}_r}{\mathbf{u}_r^T \cdot \mathbf{u}_r}$. 2-52
12. Deflation step: $\mathbf{X}_{New} = \mathbf{X} - \mathbf{t}_r \cdot \mathbf{p}_r^T$. 2-53
13. Deflation step: $\mathbf{Y}_{New} = \mathbf{Y} - d_r \cdot \mathbf{t}_r \cdot \mathbf{q}_r^T$. 2-54
14. Replace **X** by $\mathbf{X}_{New}$ and **Y** by $\mathbf{Y}_{New}$ and calculate the next dimension, i.e. $r = r + 1$. Go to step (2).
15. Stop when maximum number of latent variables have been calculated or when desired number of latent variables have been calculated.

2.6 Multi-way Techniques

Consider data from a batch process that comprises I batches, where J variables are measured over K time intervals throughout the duration of the batch. The data can be arranged in a three-way data matrix $\underline{\mathbf{X}}$ (I x J x K) as shown in Figure 2-11. To analyse this three-dimensional matrix, there are two alternative strategies, a bi-linear or a tri-linear approach.

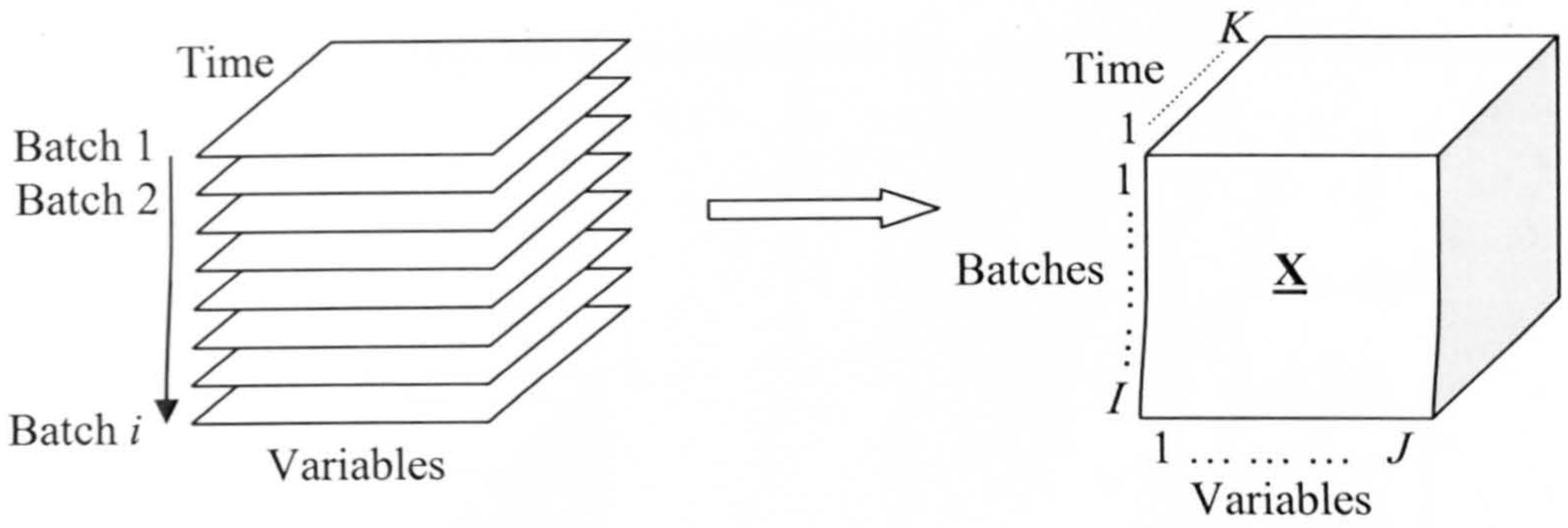


Figure 2-11 A three-way array of a batch process

For the bi-linear approach, the data is unfolded into a two-dimensional array and the bi-linear projection technique of PCA or PLS is applied. Wold *et al.* (1987) initially proposed multiway PCA (MPCA) for the decomposition of multi-way arrays. Nomikos and MacGregor (1994 and 1995b) then extended the application of MPCA to batch performance monitoring and subsequently to multiway PLS (MPLS) (Nomikos and MacGregor, 1995a). An alternative approach to that of Nomikos and MacGregor (1995b) was proposed by Wold *et al.* (1998). Details of the two different approaches are discussed in Sections 2.6.2 and 2.6.3.

The tri-linear approach retains the multi-linear structure of the data and uses multi-way techniques to decompose the multi-way array. A number of multi-way statistical techniques have been proposed including Tri-linear Decomposition (TLD) (Sanchez and Kowalski, 1990), Parallel Factor Analysis (PARAFAC) (Smilde and Doornbos, 1991; Bro, 1997), Tucker models (Geladi, 1989; Smilde, 1992) and multi-way covariate regression (Smilde and Kiers, 1999) for the decomposition of multi-way arrays. Each of the decomposition methods places a different set of constraints on the resulting matrices and vectors and in some cases the dimensionality of the original data form is retained. However the focus of the thesis is on the bi-linear approach because it is more pragmatic to the industrial situation and is discussed further in the next section.

2.6.1 Data Unfolding

To adopt the bi-linear projection approach, the batch data is first unfolded into a two-dimensional array. There are three approaches to unfolding a three-way matrix and for each approach, there are two ways to arrange the two-way matrices. Figure 2-12 to Figure 2-14 illustrate the six possible ways of unfolding a three-way matrix resulting in the following two-dimensional matrices: **A** (I x KJ), **B** (KI x J), **C** (K x IJ), **D** (IK x J), **E** (I x JK) and **F** (JI x K). The resulting matrices **A** and **E**, **B** and **D** are equivalent in that they are the same matrix structures but the columns and rows have been re-arranged. Matrix **C** is the transpose of **F** thus if PCA was applied

without centring and scaling, there would be a switch between the scores and loadings of the two matrices. Each of the three categories of unfolded matrices correspond to the analysis of different types of variability.

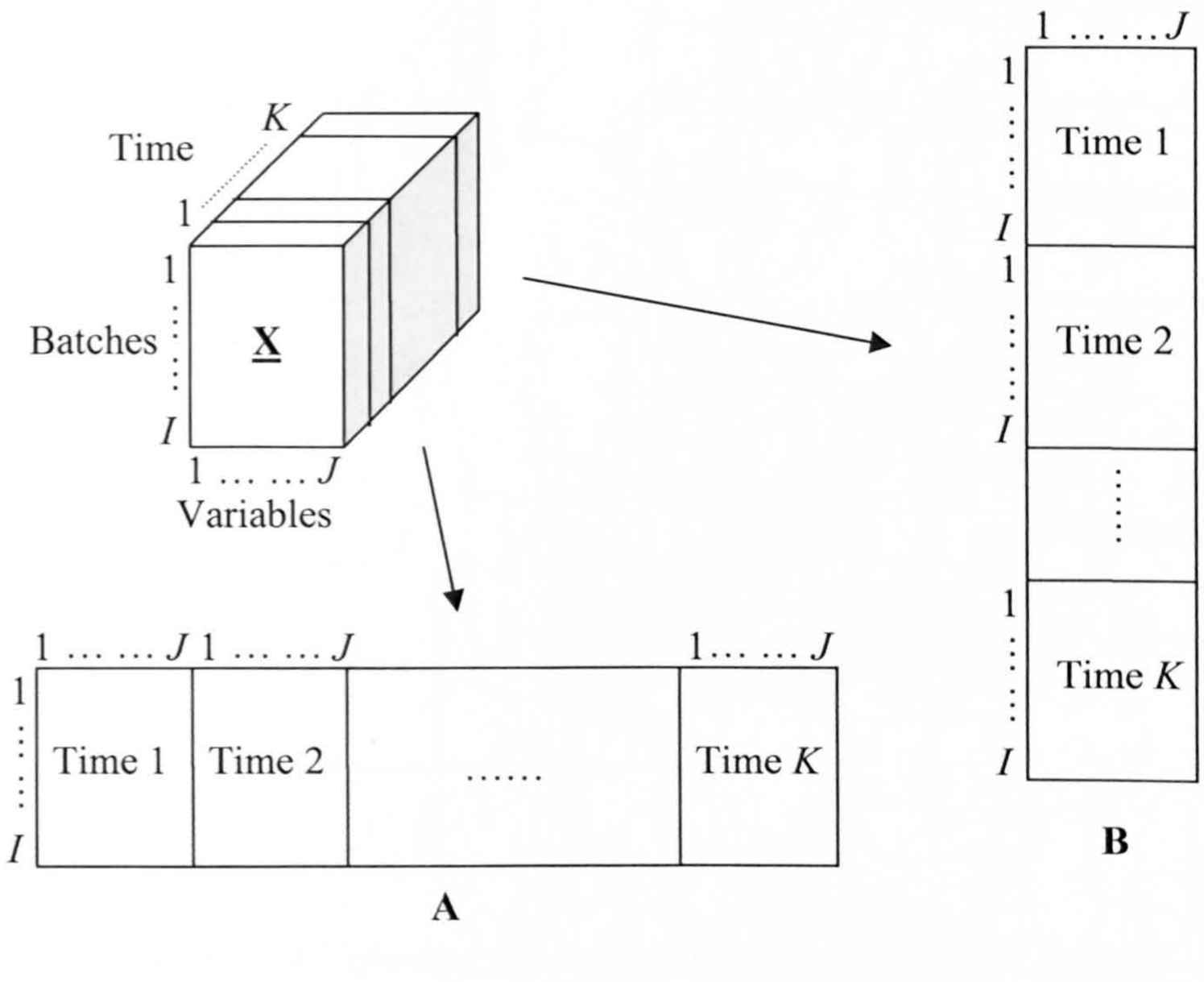


Figure 2-12 Unfolding approach one – Time mode A and B

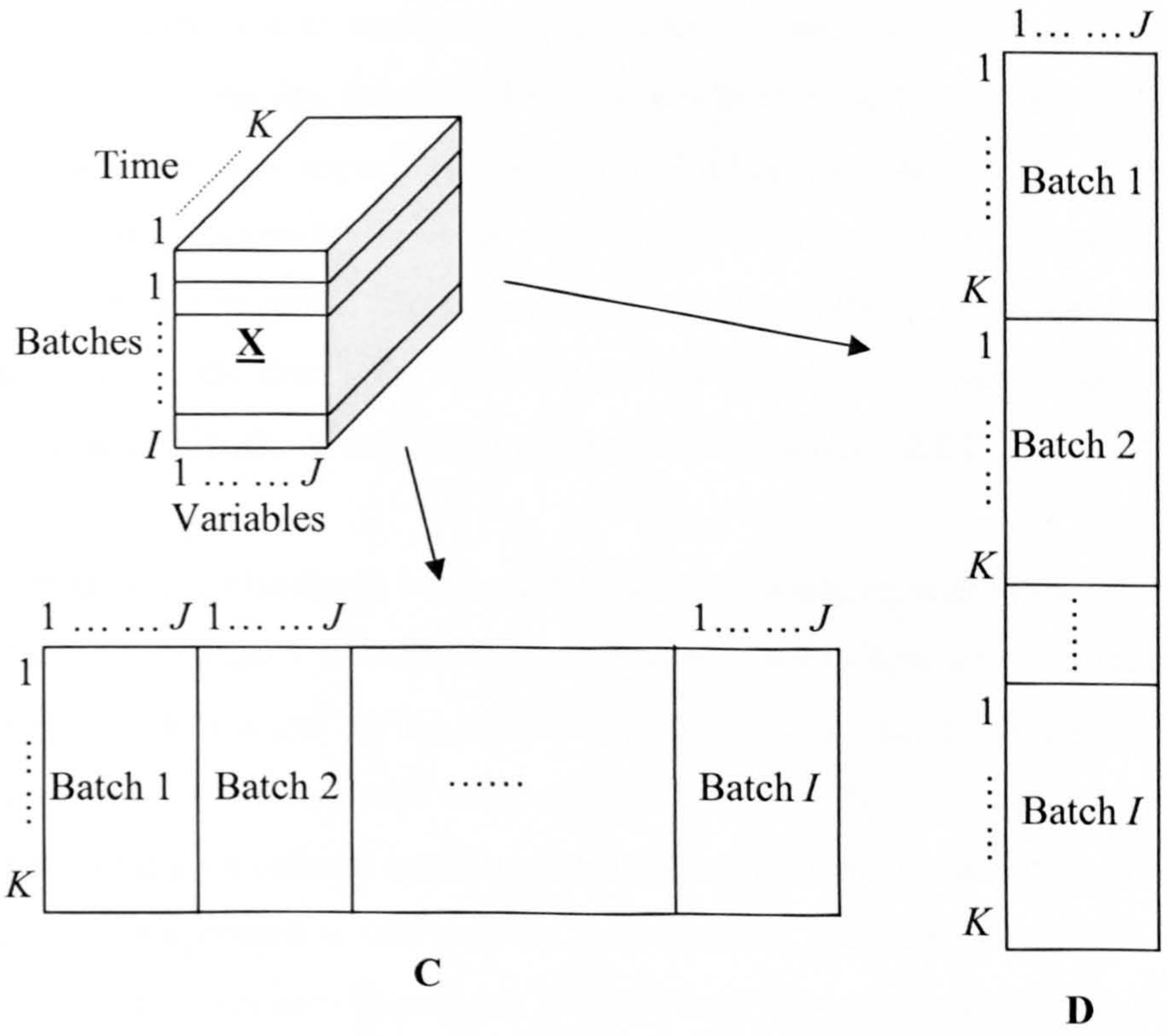


Figure 2-13 Unfolding approach two – Batch mode C and D

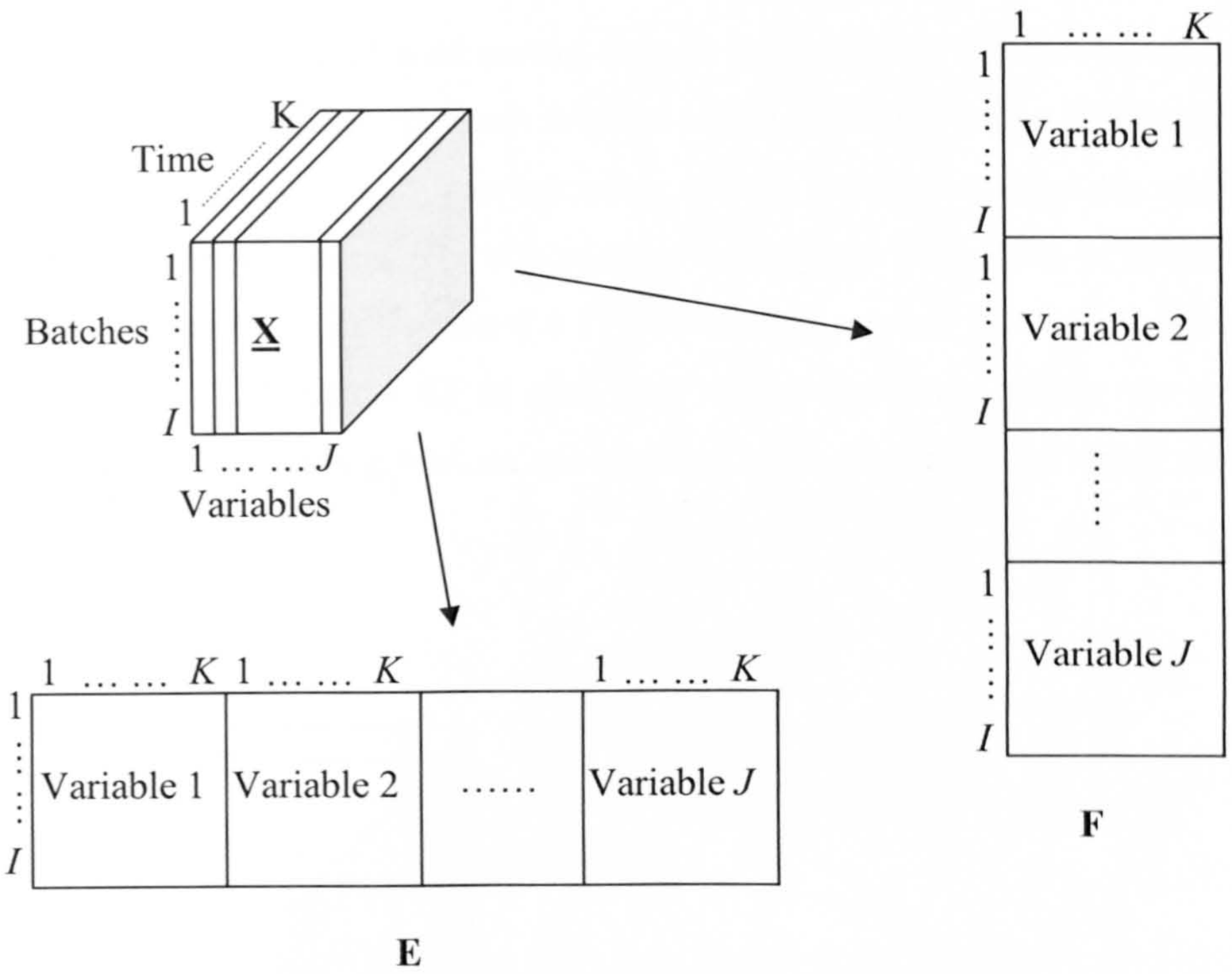


Figure 2-14 Unfolding approach three – Variable mode E and F

When the three-way matrix is unfolded to give matrix **A**, i.e. each column vector contains the measurements for variables at each time point over all batches, and each row comprises the measurements for all variables from one batch. It has been postulated that in batch performance monitoring, this is the most meaningful way of unfolding three-way matrix (Westerhuis *et al.*, 1999). This unfolding approach allows the study of the common-cause batch-to-batch variation at each time point for one variable. More specifically, it allows the comparison of the performance of each batch at a specific time point against a group of batches that were collected under normal operating conditions. Details of this approach are given in Section 2.6.2.

A second approach to multivariate batch performance monitoring was introduced by Wold *et al.* (1998). The three-way matrix is unfolded according to the structure given in matrix **D** with each row containing measurements of the variables at a particular time point in a batch and each column comprising measurements of each variable across all batches and time points. It is then centred by subtracting its column mean and scaling by its standard deviation. This is a different centring and scaling approach to that used by Nomikos and MacGregor and in this case the score trajectories are used to monitor the process. Details of this approach are given in Section 2.6.3.

2.6.2 The Nomikos and MacGregor Approach

The relationship between Multiway PCA and PCA is that MPCA is equivalent to performing PCA on a large two-dimensional data matrix formed by unfolding the three-way matrix. The Nomikos and MacGregor (N&M) approach is based on the unfolding of the three-way batch data matrix whereby vertical slices (I x J), corresponding to each point in time, are placed side by side to form a two-dimensional matrix (I x KJ) as shown in Figure 2-15. This is equivalent to the unfolded matrix **A** described in Section 2.6.1. The loading vectors can be re-arranged into a different format (unfolded matrix **E**) so that they summarise more clearly the evolution of individual variables. The loading vectors are identical to those calculated using matrix **A** but differ in the final ordering.

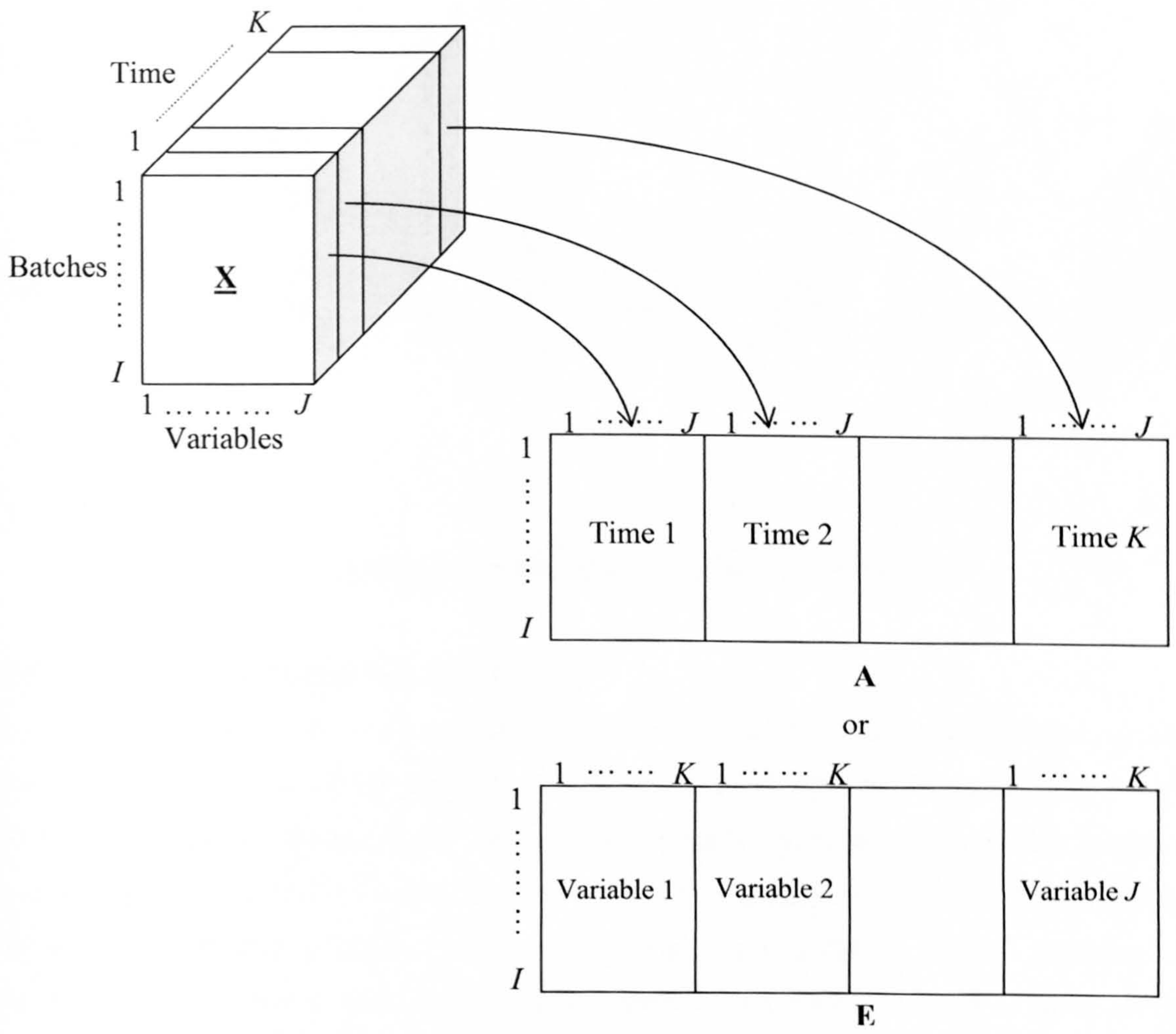


Figure 2-15 Procedure for unfolding a three-way array

MPCA has the same goals and benefits as PCA and has been shown to be statistically and algorithmically consistent with PCA (Nomikos and MacGregor, 1994). The objective of MPCA is in accordance with the principles of PCA thus the three-dimensional array, $\underline{\mathbf{X}}$, is decomposed as

the summation of the product of the score vectors ($\mathbf{t}_r$) and the loading matrices ($\mathbf{P}_r$) plus a residual $\mathbf{E}$ that is minimised in a least squares sense:

$$\underline{\mathbf{X}} = \sum_{r=1}^{R} \mathbf{t}_r \otimes \mathbf{P}_r + \underline{\mathbf{E}} \qquad 2\text{-}55$$

where R is the number of principal components retained. Each element of the score vector represents the overall variability of a single batch with respect to the other batches. The individual loading vectors summarise the time evolution of the variables about their mean trajectories and describe the covariance structure of the data. The MPCA decomposition of a three-way matrix can also be described graphically (Figure 2-16).

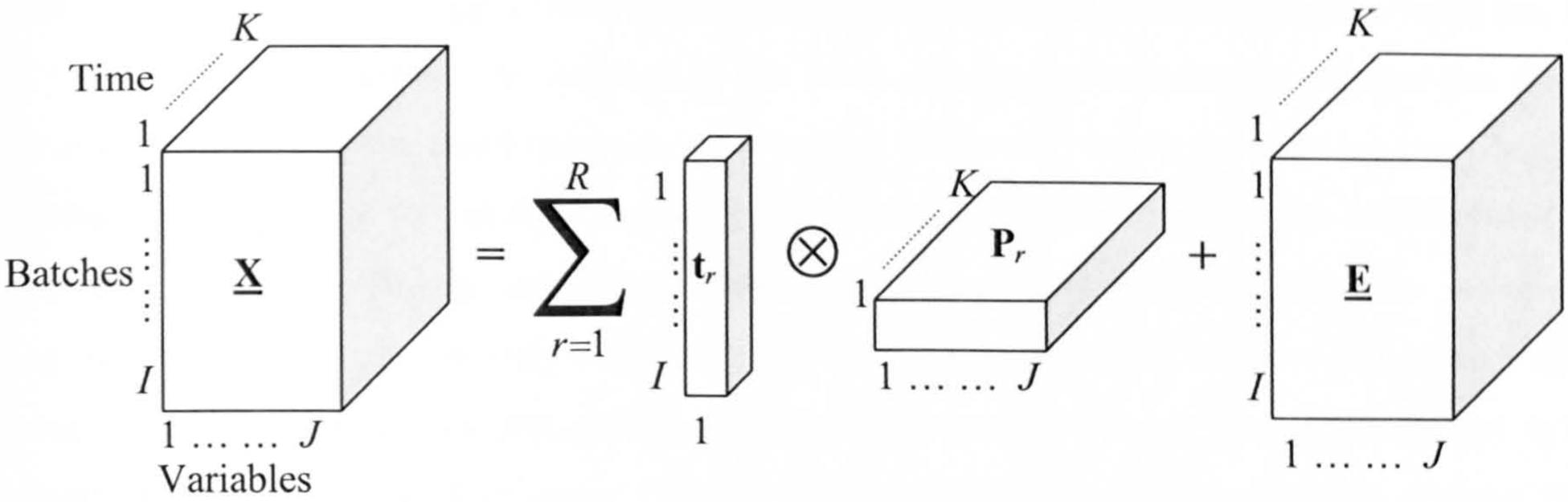


Figure 2-16 Decomposition of a three-way array by MPCA

2.6.2.1 Analysis of Historical Batch Data

The analysis of historical data gathered from a batch process gives an indication of how a process behaves. This is the first step prior to building a process representation. The usual practice is to include "good" quality batches for the model development and exclude unacceptable quality or abnormal batches. The definition of a "good" batch is typically defined as a batch that has satisfied all the necessary quality specifications hence should exhibit normal variation and behaviour under controlled operating conditions. Building a model with a set of good batches defines the normal operating region and the normal batch-to-batch variation, i.e. it is assumed that any future "good" batch run will be within the defined limits and have a similar correlation structure. A summary of the steps is given below:

Stage A: Historical data analysis and nominal model building

1. Collect data from historical batches reflective of good operation.

2. Check for missing data and apply data pre-processing techniques as necessary, see Section 3.2.1.
3. If necessary, apply batch length alignment, see Section 3.2.3.
4. Unfold the three-way matrix $\underline{\mathbf{X}}$ (I x J x K) to $\mathbf{X}$ (I x KJ).
5. Centre and scale the data to remove the batch mean trajectory, see Section 3.2.2.
6. Apply PCA to build nominal model.
7. Determine the limits for the control charts to be used as the basis of the nominal representations, i.e. univariate and bivariate scores, Hotelling's T^2 and SPE.
8. Check if the model is valid and contains nominal batch-to-batch variation. If abnormal variation is detected, model should be re-built to reflect only common cause variation.

2.6.2.2 On-line Batch Monitoring

The off-line analysis was based on the availability of the full batch history and the main interest was in monitoring end of batch behaviour. It is more effective if the monitoring of a batch can be in real-time, i.e. throughout the duration of the batch. An on-line monitoring scheme has many advantages including the rapid detection of abnormal behaviour thereby preventing poor quality production, the ability to fine-tune the control parameters at each stage for the achievement of consistent production and the capability to diagnosis faults on-line. However, one issue is that for any new batch, the data are only available from the beginning of the batch to the current time point. Hence in terms of the data matrix, the values beyond the current time point are yet to be recorded. Nomikos and MacGregor (1995b) proposed a number of methods to address this problem. Based on the strategy that the batch data matrix is estimated with appropriate values such that the trajectory is completed according to a suggested criterion, the principal component score and SPE values at the current time point can be calculated utilising both the known and inferred values. The calculation of the principal component scores $\mathbf{t}_k$ and the SPE value for each time point (SPE_k) of a new batch is as follows:

1. For a vector of new observations $\mathbf{x}_{new}$ at time point k, the vector is standardised according to its mean and standard deviation corresponding to the k^{th} time point from the nominal data.
2. The values of vector $\mathbf{x}_{new}$ from time point $k+1$ to the end of batch are estimated using one of the proposed methods (See below for the zero deviation, current deviation and missing data approaches).
3. The scores $\mathbf{t}_k$ and SPE_k values are calculated:

$$\mathbf{t}_k = \mathbf{x}_{new} \cdot \mathbf{P}_{nominal}$$

$$\mathbf{e}_k = \mathbf{x}_{new} - \mathbf{t}_k \cdot \mathbf{P}_{nominal}^{\mathrm{T}}$$

$$SPE_k = \sum_{j=1}^{J} \mathbf{e}_{jk}^2$$

where $\mathbf{e}_{jk}$ is the prediction error of variable j at time point k.

4. Return to step (1) for next time point, update $\mathbf{x}_{new}$.

The loading matrix $\mathbf{P}_{nominal}$ obtained from the nominal reference data is used throughout the algorithm as it contains the nominal structural information for the entire history of the batches. The observation vector for the new batch consists of three parts: the past measurements, the current measurement and the future unknown measurements (Figure 2-17). Three methods were proposed for estimating the future unknown measurements of $\mathbf{x}_k$ (Nomikos and Macgregor, 1995b).

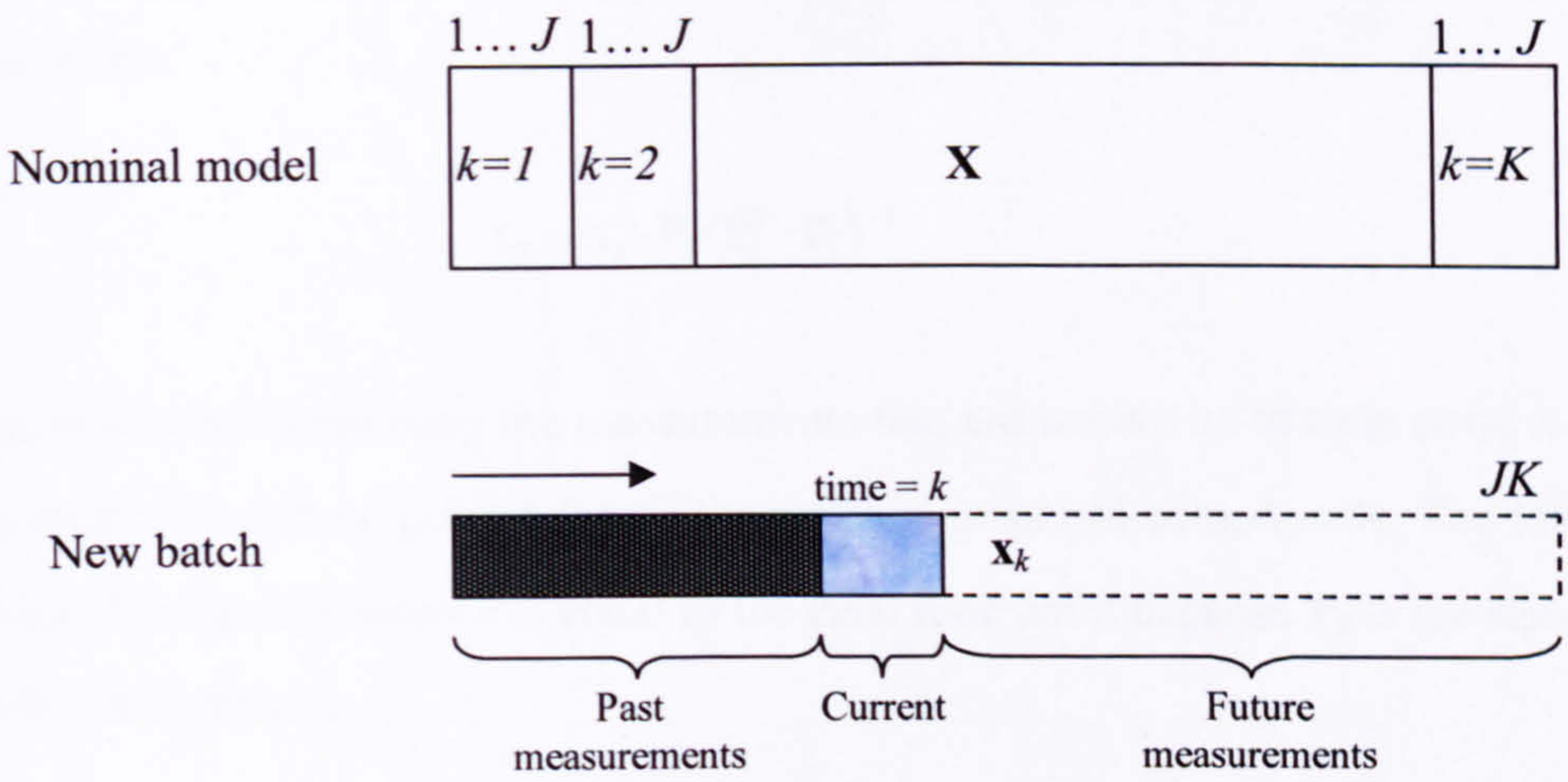


Figure 2-17 On-line batch monitoring scheme

Zero Deviations Approach

The first approach to in-filling the unknown observations of $\mathbf{x}_k$ is to assume that the future trajectory follows the desired level of operation as defined from the nominal operating conditions. The assumption is that the batch will operate normally for the remaining duration of the batch, i.e. follow the mean trajectory. Thus the unknown measurements of $\mathbf{x}_k$ are in-filled with zeros. This can be described as an optimistic scenario as there is no indication that a deviation will occur and there is also a limitation with respect to the detection capability of this approach during batch start-up.

Current Deviations Approach

The second approach for in-filling is to assume that the future measurements will continue to deviate at the same level as present at time point k. Thus the unknown part of $\mathbf{x}_k$ is in-filled with the offset values at time point k. This can be described as a pessimistic situation because it is assumed that a sensor measurement which is out of the desired operating region will remain so for the rest of the batch.

Missing Data Approach

The missing data approach uses the ability of PCA to handle missing data to overcome the need for in-filling. The unknown future observations are regarded as missing values thus only that part of the loadings up to time point k is used to compute the scores and residuals. The loadings can be described as "growing" in time until the end of batch. When the current time point is available, it is added to the past measurements in $\mathbf{x}_k$. The principal component scores $\mathbf{t}_k$ at time point k are then calculated:

$$\mathbf{t}_k = \mathbf{x}_k \cdot \mathbf{P}_k (\mathbf{P}_k^T \cdot \mathbf{P}_k)^{-1} \qquad 2\text{-}56$$

where $\mathbf{x}_k$ is a vector containing the measurements that are known up to time point k and $\mathbf{P}_k$ is the loading matrix up to time point k for all the retained principal components. The term $(\mathbf{P}_k^T \cdot \mathbf{P}_k)^{-1}$ will be equal to identity when k is equal to the final time point because $\mathbf{P}_k$ is the same as $\mathbf{P}$ which is an orthogonal matrix.

The selection of the most appropriate approach is dependent on the application and a combination of approaches can also be considered. Once the scores and SPE values have been calculated using one of the approaches for time point k, the values are plotted on the monitoring charts. The statistical control limits for the on-line principal component scores plot are derived in the same way as for standard PCA (Equation 2-19). Hotelling's T^2 for a new batch can be calculated as described in Equation 2-23:

$$T_k^2 = (\mathbf{t}_{new,k} - \bar{\mathbf{t}}_k)^T \mathbf{S}_k^{-1} (\mathbf{t}_{new,k} - \bar{\mathbf{t}}_k) \qquad 2\text{-}57$$

where $\mathbf{t}_{new,k}$ are the scores of the new batch at time point k, $\bar{\mathbf{t}}_k$ contains the means of the columns of the score matrix $\mathbf{T}_k$ for the nominal batches at time point k and $\mathbf{S}_k$ is the variance-covariance matrix of $\mathbf{T}_k$ and is an estimate of the variation across the nominal batches. Since the scores are a linear combination of a number of variables, they are independent and normally distributed in the

batch direction according to the central limit theorem. Therefore the control limits for Hotelling's T^2 follow a F-distribution:

$$T_{UCL}^2 \sim \frac{R(I^2-1)}{I(I-R)} F_{R,I-R,\alpha} \qquad 2\text{-}58$$

where I is the number of nominal batches, R is the number of principal component retained in the model, $F_{R,I-R,\alpha}$ is the critical value of the F-distribution with R and I–R degrees of freedom for a significance level α.

The SPE statistic for a new batch at time point k is given by:

$$SPE_k = \sum_{j=1}^{J} \mathbf{e}_{jk}^2 \qquad 2\text{-}59$$

where $\mathbf{e}_{jk}$ is the prediction error of variable j at time point k. The control limits of SPE can be derived as (Box, 1954):

$$Q_{UCL} = f \cdot \chi_{h,\alpha}^2 \cdot \qquad 2\text{-}60$$

The approximate values of f and h are determined by equating the mean ($\mu = fh$) and variance ($\nu = 2f^2h$) of the $f \cdot \chi_h^2$ distribution to the sample mean ($\bar{x}$) and variance (σ^2) of the SPE of the reference batch data at each time point. This is one way to estimate the values of f and h provided that the number of SPE values for the model reference data is sufficient:

$$f = \frac{\sigma^2}{2\bar{x}} \qquad 2\text{-}61$$

and

$$h = \frac{2\bar{x}^2}{\sigma^2}. \qquad 2\text{-}62$$

Hence the control limit for the SPE for time point k is given by:

$$SPE_{UCL} = \frac{\sigma^2}{2\bar{x}} \cdot \chi^2_{\frac{2\bar{x}^2}{\sigma^2},\alpha} \qquad 2\text{-}63$$

where $\chi^2_{\frac{2\bar{x}^2}{\sigma^2},\alpha}$ is the critical value of the chi-squared distribution with $\frac{2\bar{x}^2}{\sigma^2}$ degrees of freedom for significance level α. The use of a χ^2-distribution implicitly assumes normality of the errors which may not always be the case in practice. However, as the parameters of the χ^2-distribution used in calculating the SPE limits are acquired directly from the moments of the sampling distribution of the reference batch data, this approximating distribution is found to work well even in cases where the errors are not normal (Nomikos and MacGregor, 1995b). A summary of the steps for the monitoring of a new batch is outlined below.

Stage B: Monitor the performance of new batches

1. For a new batch at time point *k*, arrange the vector in the format $\mathbf{x}_{New}(1 \times KJ)$.
2. Apply the same scaling factors to the vector as calculated in Step 5, Stage A.
3. Project the new vector onto the loading matrix to calculate the new scores. For on-line monitoring, select the appropriate estimation method.
4. Calculate the metrics of Hotelling T^2 and SPE.
5. Project the new metrics onto the control charts and determine whether the process is within statistical process control (i.e. confidence limits).
6. If an abnormal situation is detected, identify the variables contributing to the fault through the application of contribution analysis.

2.6.2.3 Multiway Partial Least Squares (MPLS)

The technique of Multiway Partial Least Squares (MPLS) is similar to that of MPCA but as for PLS, the product quality data **Y** is introduced. The process block $\underline{\mathbf{X}}$, comprises *I* batches where *J* process variable measurements are recorded over *K* time intervals. The Y-block consists of the final quality measurements for each batch and forms a two-way matrix (*I* x *L*, batch x quality variable). The unfolding of the $\underline{\mathbf{X}}$ matrix is performed as for MPCA and as the Y-block is two-dimension, there is no requirement to unfold the matrix. After unfolding, both the **X** and **Y** matrices are mean centred and standardised to unit variance prior to applying the standard PLS algorithm.

2.6.3 The Wold, Kettaneh, Friden and Holmberg Approach

Another variant of batch performance monitoring was introduced by Wold *et al.* (1998). This approach is based on unfolding the three-way matrix in the batch direction to give a two-way matrix comprising *I* x *K* rows and *J* columns, assuming equal batch lengths. This is equivalent to the unfolded matrix **D** described in Section 2.6.1. The model for the unfolded matrix $\mathbf{X}_{IK \times J}$ is given by:

$$\mathbf{X}_{IKxJ} = \sum_{r=1}^{R} \mathbf{t}_r \mathbf{p}_r^{\mathrm{T}} + \mathbf{E} \qquad 2\text{-}64$$

where $\mathbf{t}_r$ is a score vector of size (1 x IK), $\mathbf{p}_r$ is a loading vector of length (J x 1), $\mathbf{E}$ is the residual matrix and R is the number of principal components included in the model.

Each row comprises the data for all variables for an individual batch at time point k in contrast to the unfolded matrix of the N&M approach where each row comprises the data for all variables and time points for a batch. The data are then centred and scaled to unit variance across the whole column as described in detail in Section 3.2.2. The scores matrix will now describe the trajectory of the evolving batches. This is a powerful approach for the monitoring of a batch as it evolves but the non-linear time-varying trajectories in the data are evident in the first few principal components.

It has been argued by Westerhuis *et al.* (1999) that by applying PCA to this form of matrix, additional principal components are required to be retained to describe the same systematic variation in the data as for the N&M approach. Furthermore, for the purpose of batch performance monitoring, it focuses on the monitoring of the variation described by the mean trajectories of the variables over all batches and all time points as opposed to the batch-to-batch variation which may be of greater interest.

The advantages of this approach are that it does not require any in-filling mechanism for future measurements and there is no need to perform batch length equalisation. However, mean centring and scaling of the unfolding matrix does not remove the mean trajectories from the data since it only captures the covariance among the variables which is not of major interest in monitoring. This also results in the non-linear and time-varying trajectories not being removed from the data matrix and these are reflected in its resulting scores. Furthermore, the loadings reflect an overall perspective of the variables ignoring the time factor, i.e. it is assumed the covariance structure does not change over time.

In the original paper of Wold *et al.* (1998), there is also an approach based on a PLS model in which the **y** vector contains either the batch evolution time or the batch maturity index. This results in potentially more components being retained in the model to explain sufficient variation in the **X** matrix. This is a consequence of the fact that the score matrix **T** calculated from **X** is strongly affected by the direction of variation inherent in **y**. Hence the first principal component will be dominated by these variables in the data matrix **X** that are correlated with **y**, i.e. those

variables that increase or decrease with time. The second principal component displays a quadratic effect with time for **X** variables and the third principal component, the cubic effect of the **X** variables and so on.

The $\mathbf{y}_{predicted}$ vector using batch evolution time can also allow an estimate of how far the new batch has progressed. This gives an early indication of when to terminate the batch. However, if a batch maturity variable can be identified as one of the **X** variables to be the **y** vector, the $\mathbf{y}_{predicted}$ vector contains a strong contribution from this variable but also from other **X** variables that are correlated with this maturity variable. Thus this is the preferred index since there is a strong correlation between **X** and **y**.

A summary of the batch process monitoring approach proposed by Wold *et al.* (1998) is outlined below.

Stage A: Historical data analysis and nominal model building

1. Collect data from historical batches reflective of good operation.
2. Check for missing data and apply data pre-processing techniques, see Section 3.2.1.
3. Unfold the three-way matrix $\underline{\mathbf{X}}$ (I x J x K) to $\mathbf{X}$ (IK x J). Batches may be of unequal duration but it is assumed for notational ease they are of equal length.
4. Construct a **y** vector with either batch evolution time or batch maturity index.
5. Centre and scale the data to remove the grand mean.
6. Either apply PCA to the **X** only matrix or apply PLS to **X** and **y** to build the nominal model.
7. Once the scores are obtained, they are re-arranged time-wise to determine the confidence limits ± 2 and 3 standard deviations at each time point, Figure 2-18.
8. Calculate the monitoring metrics using scores, $\mathbf{y}_{predicted}$, Hotelling's T^2 and SPE.
9. Check if the model is valid. If abnormal variation is detected, the model should be re-built to reflect only common cause variation.

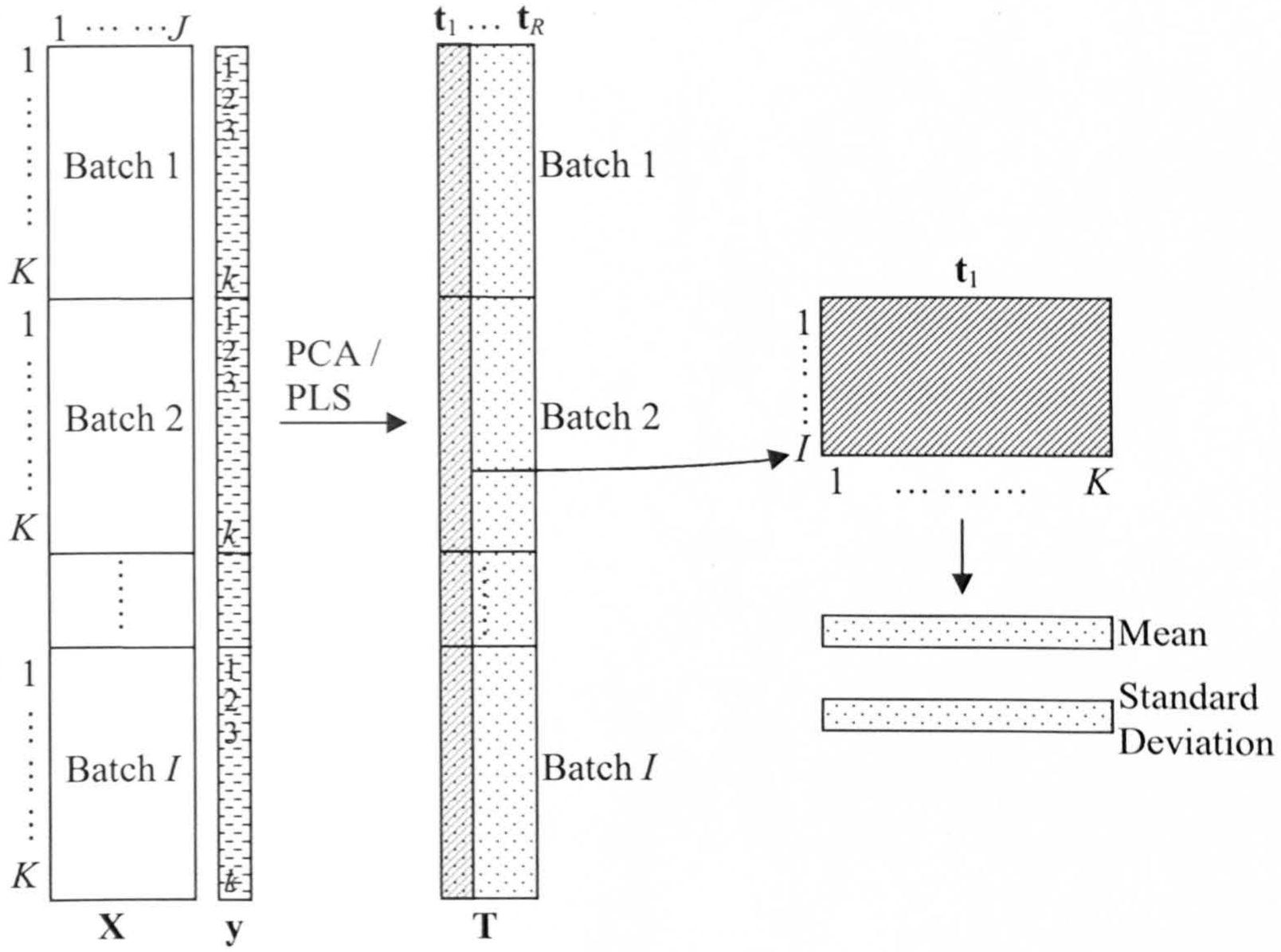


Figure 2-18 Illustration of Wold et al. approach and re-arrangement of scores matrix

Stage B: Investigate the performance of unseen batches

1. For a new batch at time point k, arrange the vector in the format $\mathbf{x}_{New}(K \times J)$, $k = 1, 2 \ldots K$.
2. Apply the same scaling factors to the vector from Step 5, Stage A.
3. Project the new vector onto the loading vector to calculate the new scores.
4. Calculate the monitoring metrics using the scores, $\mathbf{y}_{predicted}$, Hotelling's T^2 and SPE.
5. Project the new metrics onto the control charts and determine whether the process is within statistical process control (i.e. confidence limits).
6. If an abnormal situation is detected, identify the variables contributing to the fault through the application of contribution analysis.

2.7 Multi-group Techniques

2.7.1 Multi-group Principal Component Analysis

Principal Component Analysis (PCA) is typically viewed as a single group methodology where a separate model is required for the monitoring of individual products. This is evidenced from the literature where reported MSPC applications have focused on the monitoring of a single manufactured product (Kosanovich *et al.*, 1996; Neogi and Schlags, 1998). Limitations of this

approach are where interest is in multiple products and hence there is a need for sufficient data for each product type to develop separate model representations. Lane (1999) proposed a multi-group technique that can simultaneously handle a number of products by a single representation, multi-group PCA thereby addressing the data issue. This technique is an extension to PCA and is based on the assumption that a common eigenvector subspace exists for the variance-covariance matrices of the individual product situations. Through a pooled sample variance-covariance matrix, the principal component loadings of the multi-group model can be calculated and a process representation can be developed as described in Section 2.4.

2.7.2 Review of Other Multiple Group Monitoring Tools

The first application of PCA to multiple groups was proposed by Krzanowski (1979) who compared the educational performance of students from a number of colleges by developing a descriptive technique utilising a geometrical approach for comparing the subspace spanned by the first few principal components of the different groups. The methodology was based on the utilisation of the same set of variables but was split into a number of different groups and a comparison of the angles between the subspaces spanned by the principal components of the different groups was undertaken.

Subsequent to this approach, a formal statistical approach, Common Principal Component (CPC) analysis, was proposed by Flury (1984) to address the problem of the simultaneous analysis of group data by assuming that the eigenvectors of the individual variance-covariance matrices are statistically equivalent across all the groups, while the eigenvalues are allowed to vary. This resulted in the retention of lower order components in the model. However interest is often associated with the largest eigenvalues, i.e. first few components. Therefore this technique is potentially sub-optimal due to the inclusion of lower order components in the model.

A more appropriate representation, the Partial Common Principal Component Model (PCPC), was proposed by Flury (1987). This representation requires only the first few principal components to be common between the groups with the remaining principal components only being specific to a particular group. However, the underlying assumption of both approaches is that the same set of variables are utilised. This assumption is not always practical in some industrial processes and applications since different variables may be recorded for different products.

Thorpe (1983) proposed using the eigenvectors of the pooled sample variance-covariance matrix of the individual groups to approximate the principal component loadings. Consider the situation

where there are G groups, the pooled sample variance-covariance matrix $\mathbf{S}^P$ is defined as a weighted sum of the g individual variance-covariance matrices $\mathbf{S}_1$, $\mathbf{S}_2$, ..., $\mathbf{S}_g$:

$$\mathbf{S}^P = \frac{(I_1 - 1)\mathbf{S}_1 + (I_2 - 1)\mathbf{S}_2 + \cdots + (I_G - 1)\mathbf{S}_G}{\sum_{i=1}^{g} I_i - G} \qquad g = 1, 2 \ldots G \qquad \text{2-65}$$

where I_g is the number of samples within group g. Although not theoretically proven, Krzanowski (1984) hypothesised that the eigenvectors of the weighted sum of the variance-covariance matrix is the same as for the g sample variance-covariance matrices. Consequently, the pooled sample variance-covariance matrix is a good estimate of the eigenvectors that are common to all the individual groups.

2.7.3 Development of the Pooled Correlation Model

The pooled variance-covariance matrix for multiple group Principal Component Analysis was reviewed in the previous section. In this section two illustrations of the methodology are presented, a simple two-group case comprising two sets of identical variables (Lane, 1999). The second case is where one group has five variables and group two comprises four variables of which three variables are common to both groups. The situation is then extended to the approach of batch process monitoring where two groups of data can be considered as two grades of the same product having identical and grade specified process variables.

The first step in developing the model is to standardise the matrices. Data scaling has a significant impact on the final analysis and is discussed in Section 3.2.1. In this chapter, the two matrices are standardised separately by auto-scaling:

$$\mathbf{X}_g^+ = \frac{\left(\mathbf{X}_g - \bar{\mathbf{X}}_g\right)}{s_g} \qquad g = 1, 2 \ldots G \qquad \text{2-66}$$

where $\mathbf{X}_g^+$ is the auto-scaled data matrix for group g and s_g is the standard deviation. The individual correlation matrices for each group are then calculated:

$$\mathbf{C}_g = \frac{\mathbf{X}_g^{+\mathrm{T}} \cdot \mathbf{X}_g^+}{I_g - 1} \qquad g = 1, 2 \ldots G \qquad \text{2-67}$$

where $\mathbf{C}_g$ is the correlation matrix for group g. If the different groups comprise identical variables, the pooled correlation matrix is a weighted average of the individual correlation matrices:

$$\mathbf{C}^P = \frac{(I_1 - 1)\mathbf{C}_1 + (I_2 - 1)\mathbf{C}_2 + \cdots + (I_G - 1)\mathbf{C}_G}{\sum_{i=1}^{g} I_i - G} \qquad 2\text{-}68$$

where $\mathbf{C}^P$ is the pooled correlation matrix. Once the pooled correlation matrix has been obtained, the normal procedure of calculating PCA is applied with the eigenvectors of the pooled correlation matrix defining the principal component loadings, $\mathbf{P}^P$, of the multiple group model. The principal component scores for the individual groups are then calculated:

$$\mathbf{T}_{g,R} = \mathbf{X}_g^+ \cdot \mathbf{P}_R^P \qquad 2\text{-}69$$

where $\mathbf{T}_{g,R}$ is the matrix of scores for group g, $\mathbf{X}_g^+$ is the auto-scaled data matrix of group g and $\mathbf{P}_R^P$ is the matrix of common principal component loadings with R principal components retained. In a similar manner, Hotelling's T^2 and SPE can also be obtained for the different groups. A flow diagram of the framework for the development of a multi-group PCA model is summarised in Figure 2-19.

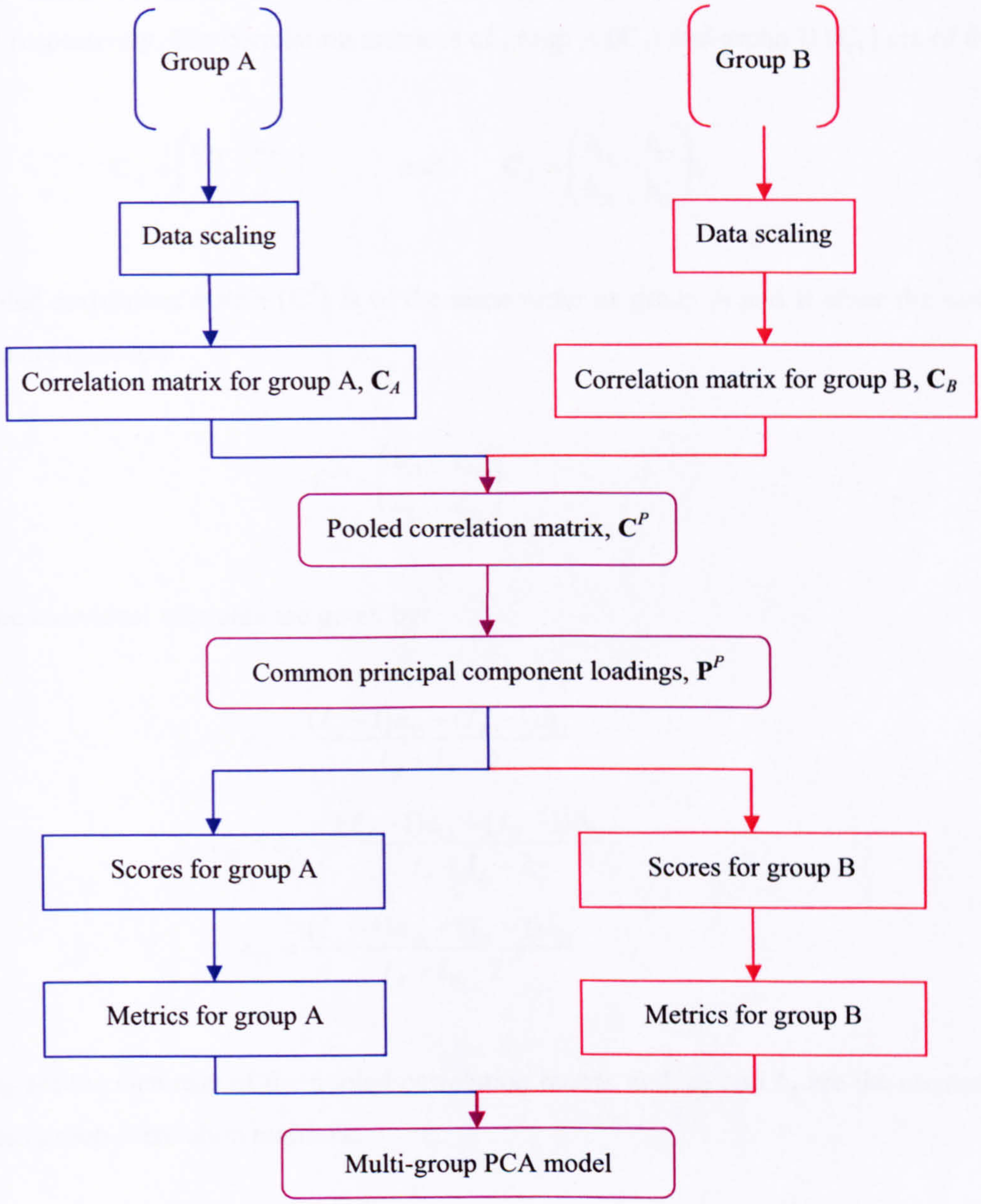


Figure 2-19 Summary of the multi-group PCA model development

Case One

Consider two matrices (group A and group B) which comprise two identical sets of variables. The first step in constructing the pooled correlation matrix is to standardise each of the variables for each group according to Equation 2-66. This results in the mean of each variable being removed and then normalised to unit variance. The correlation matrix for each group can then be calculated:

$$\mathbf{C}_A = \frac{\mathbf{A}^{+\mathrm{T}} \cdot \mathbf{A}^{+}}{I_A - 1} \qquad \text{and} \qquad \mathbf{C}_B = \frac{\mathbf{B}^{+\mathrm{T}} \cdot \mathbf{B}^{+}}{I_B - 1} \qquad 2\text{-}70$$

where $\mathbf{A}^+$ and $\mathbf{B}^+$ are the scaled data matrices and I_A and I_B are the number of samples in data set A and B respectively. The correlation matrices of group A ($\mathbf{C}_A$) and group B ($\mathbf{C}_B$) are of the form:

$$\mathbf{C}_A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \qquad \text{and} \qquad \mathbf{C}_B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}. \tag{2-71}$$

The pooled correlation matrix ($\mathbf{C}^P$) is of the same order as group A and B since the same set of variables are included:

$$\mathbf{C}^P = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} \tag{2-72}$$

where the individual elements are given by:

$$c_{11} = \frac{(I_A - 1)a_{11} + (I_B - 1)b_{11}}{I_A + I_B - 2} \tag{2-73}$$

$$c_{12} = c_{21} = \frac{(I_A - 1)a_{12} + (I_B - 1)b_{12}}{I_A + I_B - 2} \tag{2-74}$$

$$c_{22} = \frac{(I_A - 1)a_{22} + (I_B - 1)b_{22}}{I_A + I_B - 2} \tag{2-75}$$

where c_{ij} are the elements of the pooled correlation matrix and, a_{ij} and b_{ij} are the elements of the individual group correlation matrices.

Case Two

The second case is slightly more complex, i.e. a different number of variables form the basis of each group and the variables themselves differ between groups. Consider the case where matrix **A** comprises five variables. In contrast, matrix **B** includes four variables of which the first three (1 – 3) are the same variable as for matrix **A**. For illustration, the additional two variables in group A are denoted variables 4 and 5 and the fourth variable in group B is variable 6. The first step in constructing the pooled correlation matrix is again to standardise each of the variables in each group according to Equation 2-66. The correlation matrix for each group can then be calculated:

$$\mathbf{C}_A = \frac{\mathbf{A}^{+\mathrm{T}} \cdot \mathbf{A}^+}{I_A - 1} \qquad \text{and} \qquad \mathbf{C}_B = \frac{\mathbf{B}^{+\mathrm{T}} \cdot \mathbf{B}^+}{I_B - 1} \tag{2-76}$$

where $\mathbf{A}^+$ and $\mathbf{B}^+$ are the scaled data matrices. The correlation matrices for group A ($\mathbf{C}_A$) and group B ($\mathbf{C}_B$) are of the form:

$$\mathbf{C}_A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{pmatrix} \quad \text{and} \quad \mathbf{C}_B = \begin{pmatrix} b_{11} & b_{12} & b_{13} & b_{16} \\ b_{21} & b_{22} & b_{23} & b_{26} \\ b_{31} & b_{32} & b_{33} & b_{36} \\ b_{61} & b_{62} & b_{63} & b_{66} \end{pmatrix}.$$

With the pooled correlation matrix ($\mathbf{C}^P$) given by:

$$\mathbf{C}^P = \begin{pmatrix} c_{11} & c_{12} & c_{13} & c_{14} & c_{15} & c_{16} \\ c_{21} & c_{22} & c_{23} & c_{24} & c_{25} & c_{26} \\ c_{31} & c_{32} & c_{33} & c_{34} & c_{35} & c_{36} \\ c_{41} & c_{42} & c_{43} & c_{44} & c_{45} & c_{46} \\ c_{51} & c_{52} & c_{53} & c_{54} & c_{55} & c_{56} \\ c_{61} & c_{62} & c_{63} & c_{64} & c_{65} & c_{66} \end{pmatrix}.$$

The individual elements are calculated as follows:

$$c_{11} = \frac{(I_A - 1)a_{11} + (I_B - 1)b_{11}}{I_A + I_B - 2} \tag{2-77}$$

$$c_{12} = c_{21} = \frac{(I_A - 1)a_{12} + (I_B - 1)b_{12}}{I_A + I_B - 2} \tag{2-78}$$

$$c_{13} = c_{31} = \frac{(I_A - 1)a_{13} + (I_B - 1)b_{13}}{I_A + I_B - 2} \tag{2-79}$$

$$c_{14} = c_{41} = a_{14} \tag{2-80}$$

$$c_{15} = c_{51} = a_{15} \tag{2-81}$$

$$c_{16} = c_{61} = b_{16} \tag{2-82}$$

$$c_{22} = \frac{(I_A - 1)a_{22} + (I_B - 1)b_{22}}{I_A + I_B - 2} \tag{2-83}$$

$$c_{23} = c_{32} = \frac{(I_A - 1)a_{23} + (I_B - 1)b_{23}}{I_A + I_B - 2} \tag{2-84}$$

$$c_{24} = c_{42} = a_{24} \tag{2-85}$$

$$c_{25} = c_{52} = a_{25} \tag{2-86}$$

$$c_{26} = c_{62} = b_{26} \tag{2-87}$$

$$c_{33} = \frac{(I_A - 1)a_{33} + (I_B - 1)b_{33}}{I_A + I_B - 2} \quad 2\text{-}88$$

$$c_{34} = c_{43} = a_{34} \quad 2\text{-}89$$

$$c_{35} = c_{53} = a_{35} \quad 2\text{-}90$$

$$c_{36} = c_{63} = b_{36} \quad 2\text{-}91$$

$$c_{44} = a_{44} \quad 2\text{-}92$$

$$c_{45} = c_{54} = a_{45} \quad 2\text{-}93$$

$$c_{46} = c_{64} = 0 \quad 2\text{-}94$$

$$c_{55} = a_{55} \quad 2\text{-}95$$

$$c_{56} = c_{65} = 0 \quad 2\text{-}96$$

$$c_{66} = b_{66}. \quad 2\text{-}97$$

where c_{ij} are the elements of the pooled correlation matrix, a_{ij} and b_{ij} are the elements of the individual group correlation matrices and I_A and I_B are the numbers of observations in each group. From the pooled correlation matrix, it is observed that some of the elements do not require to be pooled between groups, i.e. some elements in the pooled correlation matrix are equal to the elements in the individual group correlation matrix or zero. This is a consequence of some variables not being common between the groups, e.g. variables 1 and 4. Additionally, $\mathbf{C}^P$ comprises zero entries. These represent the fact that variable 4 is associated with matrix **A** but variable 6 is not included in matrix **A**.

2.7.3.1 Extension to Batch Applications

To extend the situation to batch processes, an example can be considered where two different grades of a product are monitored. For example grade A comprises 20 production batches with 5 process variables and grade B includes 50 batches with 4 process variables. The construction of a batch multi-group nominal model can adopt the batch performance monitoring approaches described in Section 2.6 with an additional step of calculating the pooled sample correlation matrix from various groups before applying standard PCA. The calculation of the principal component scores of the batches from different groups is then performed on the common principal component loadings. A summary of the generic multi-group batch monitoring approach is described below:

1. Collect data from historical batches reflective of good operation for different groups of data.

2. Check for missing data and apply data pre-processing techniques as necessary, see Section 3.2.1.
3. If necessary, apply batch length alignment, see Section 3.2.3.
4. Unfold the three-way matrices $\underline{\mathbf{X}}$ to $\mathbf{X}$. The form of $\mathbf{X}$ depends on the batch monitoring approach adopted, see Section 2.6.
5. Centre and scale the individual group of data separately, see Section 3.2.2.
6. Calculate the individual elements of the correlation matrices for different groups.
7. Calculate the pooled correlation matrix, $\mathbf{C}^P$.
8. Apply PCA to build the nominal model and extract the common principal component loadings, $\mathbf{P}^P$.
9. Calculate the principal component scores for different groups based on $\mathbf{P}^P$.
10. Determine the limits for the control charts to be used as the basis of the nominal model, i.e. univariate and bivariate scores, Hotelling's T^2 and SPE.

2.8 Summary

In this chapter, an overview of Multivariate Statistical Process Control (MSPC) methodologies for batch process performance monitoring has been presented. MSPC plays an important role in the processing industries in terms of ensuring safe process operation and the manufacture of robust in-specification product. The underpinning techniques of MSPC are the bi-linear statistical projection methods of Principal Component Analysis (PCA) and Partial Least Squares (PLS). In PCA a large data set comprising variables that are correlated can be decomposed into a few independent new variables, principal components, to capture the key variability in the data set. PLS includes a second matrix which is typically the product quality information. In this case the dimensionality of the two matrices is reduced simultaneously and the best correlation structure is described by a series of latent variables. The process performance representations are described by defining the appropriate confidence limits for a number of metrics including the univariate and bivariate scores, Hotelling's T^2 and the SPE.

To handle three-way batch process data, PCA and PLS have been extended, Multiway PCA (MPCA) and Multiway PLS (MPLS). In these cases the three-way data is unfolded into two-dimensional arrays and PCA and PLS are then applied. The procedures for analysing historical data and the monitoring of on-line batch process are discussed. Another variant of batch process monitoring as based on the research of Wold *et al.* (1998) results in the three-way data being unfolded into a different format so that the resulting scores matrix describe the trajectories of a

evolving batch. This is a powerful way for the monitoring and diagnosis of a batch as it evolves but the non-linear time-varying trajectories in the data are still present.

Finally, a multi-group technique based on PCA was discussed. Such a technique enables the simultaneous monitoring of product grades or manufacturing sites. The method is based on the assumption that a common eigenvector subspace exists for the sample covariance matrices of the individual groups. The development of the pooled correlation model is discussed with two case examples. The extension of the concept to batch applications is also discussed. The next chapter will evaluate the performance of existing and proposed methodologies for process monitoring of batch processes.

Chapter 3

Performance Evaluation of Batch Process Monitoring Methodologies

3.1 Introduction

Prior to applying the multivariate projection techniques of PCA/PLS or their extensions, the data may require to be pre-treated. If care is not taken in the data pre-treatment stage, the overall outcome of the analysis can be misleading. One of the objectives of this chapter is to describe the various stages required to pre-treat the data and the different methods available. Based on the discussions in Chapter 2, together with what is described in this chapter for data pre-treatment, the fundamental principles of the proposed approach for the monitoring of batch processes are formulated. The introduction of this proposed approach is the focus of the chapter and its performance is evaluated against existing global and local approaches using a simulated data set. Two performance indices are utilised for the comparison of the different methodologies, i.e. false alarm rate and out-of-control average run length.

3.2 Pre-treatment of Batch Data

A significant amount of data is recorded in today's manufacturing environment. This forms the basis of the development of process monitoring schemes. Three key stages are considered in developing a monitoring scheme – historical data collection, data pre-treatment and the development of the final model using the appropriate multivariate statistical projection based techniques. Key to the implementation of a process performance monitoring scheme is the acquisition of representative data that captures normal behaviour. In industrial processes, data is normally acquired from commercial control software such as supervisory control and data acquisition systems (SCADA) and distributed control systems (DCS) and is stored in the data historian. Following the retrieval of the data, the first step is to perform a preliminary analysis to obtain an overview of the data using standard data visualisation tools.

3.2.1 Data Pre-processing

Data pre-processing is an area that has received limited attention in the literature, but in practice it can be the key to the success or failure of the process performance scheme. A number of different data types may be collected on a process, including process data, quality measurements and spectroscopic data. The process variables measure the physical phenomena of the process and are normally collected on-line with a high sampling rate. In contrast, measurements that characterise the quality of the product are typically recorded off-line in the quality control laboratory. More recently the increasing use of in-process spectroscopic analysers gives rise to another source of data. Spectroscopy provides real-time, high-quality chemically rich information enabling an

understanding of the chemical behaviour of a process to be attained. The techniques described in this section are primarily applicable to process and quality data. A more detailed discussion on the pre-processing techniques for spectroscopic data will be presented in Section 5.4.4.

The pre-processing of batch process data and quality outputs includes a number of core tools: the identification and handling of outliers, missing data and the filtering of noisy variables. Typically these operations are implemented for each variable and for each individual batch.

Missing Data

The multivariate statistical techniques of PCA and PLS require the data matrix to be complete however in practice it is common that certain records of data may be missing. In situations, where the missing values can be dealt with prior to performing the analysis, then this operation should be executed. One of two approaches can be adapted either discard the incomplete samples or if the output values are missing, in-fill the missing values using a method such as the mean, median, last recorded value or interpolate between the observations immediately before and after the missing value. In practice in batch processes, the usual approach is to in-fill since deletion will result in a distortion of the underlying relationship and also in the time course of the data between variables.

Outlier Detection and Handling

Any measurements of variables that appear to be inconsistent with the rest of data set are defined as outliers. These "outlying" observations can have a significant impact on the subsequent analysis. For example in PCA they can affect the direction of variability in the data and in modelling both reduce the accuracy and impact on the structure of the model. Outlier detection can be as simple as using time series or scatter plots for visual screening. Alternatively to identify multivariate outliers, PCA is typically applied and both Hotelling's T^2 and the SPE are considered. The usual approach to handling outliers following their detection is to treat them in a similar manner to missing data.

Noisy Data – Filtering

Industrial process data contains some degree of random variation or noise that may result low signal-to-noise ratio. These random effects can mask the major sources of variation in the process therefore the monitoring models must be robust to the levels of noise observed. Some of the more commonly applied filters include the unidirectional filter which is an Exponentially Weighted Moving Average (EWMA) of the previous samples; the bi-directional filter in which the filtering is performed in two directions and the median filter where a median value is taken over a pre-defined window size. The statistical techniques of PCA and PLS can also act as a filter to remove

the influence of small amounts of noise. For relatively large levels of noise, the application of advanced filtering techniques such as the wavelet transform (see Section 5.3) can be applied.

Data Transformation

Process data can be transformed mathematically by substituting the values of a variable with the values of a function of a variable. Some commonly used mathematical functions include the power, inverse, the logarithm and the exponential function. Applying a mathematical transformation to some process variables may produce a more appropriate statistical batch model than the raw data or the process non-linearity may be reduced within the system. A transformation should only be applied if it is beneficial to the statistical model as it can alter the correlation structure between the variables thus care must be taken.

Data pre-processing is an important stage in the analysis, but it is recommended that the modification of the raw data is minimised in order to retain the original process trends and information inherent in the data. The outcome of this step is a data set that comprises common cause variation. The next stage of development is to address the issues of data scaling.

3.2.2 Data Scaling

Performing data centring and scaling is an important consideration prior to the monitoring and modelling stage (Bro and Smilde, 2003). In this section, two-way matrices are the main focus since it is assumed the three-way data matrix is unfolded according to one of the approaches described in Section 2.6.1.

Centring

Centring involves the removal of constants across columns or rows in a data matrix. This action shifts the variable trajectories to a common baseline. The widely accepted method in multivariate analysis is mean-centring in which the average value of each variable is calculated and subtracted from the individual data values. It has also been referred to as column centring since variables are arranged in columns (Gurden *et al.*, 2001).

For the Nomikos & MacGregor approach (N&M), the unfolded data is arranged in format A as described in Section 2.6.1 (matrix **A**). The mean-centred data therefore allows the subsequent analysis to focus on batch-to-batch variation for the same time point of a variable (Figure 3-1). This enables the extraction of common cause variation between a number of nominal batches when constructing the nominal model representation. Mathematically it can be expressed as:

$$\mathbf{x}^*_{ijk} = \mathbf{x}_{ijk} - \frac{\sum_{i=1}^{I} \mathbf{x}_{ijk}}{I} \qquad i = 1, 2 \ldots I, j = 1, 2 \ldots J, k = 1, 2 \ldots K \tag{3-1}$$

where $\mathbf{x}^*_{ijk}$ is the centred vector of data and I is the number of batches.

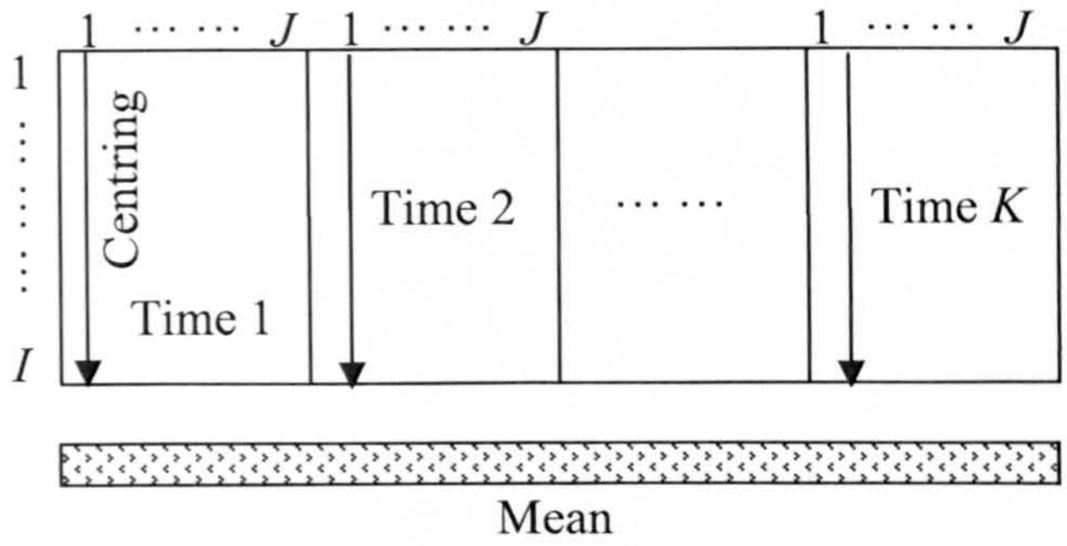


Figure 3-1 Illustration of mean-centring of N&M unfolding approach

For the Wold *et al*. approach, the unfolded data is arranged in the format of matrix **D** as described in Section 2.6.1. In this situation the mean batch trajectory is not removed by centring, instead the grand mean is taken across all batches and time points for a specific variable (Figure 3-2). Mathematically it can be expressed as:

$$\mathbf{x}^*_{ikj} = \mathbf{x}_{ikj} - \frac{\sum_{k=1}^{K}\sum_{i=1}^{I} \mathbf{x}_{ikj}}{IK} \qquad i = 1, 2 \ldots I, j = 1, 2 \ldots J, k = 1, 2 \ldots K \tag{3-2}$$

where $\mathbf{x}^*_{ikj}$ is the centred vector of data and J is the number of variables.

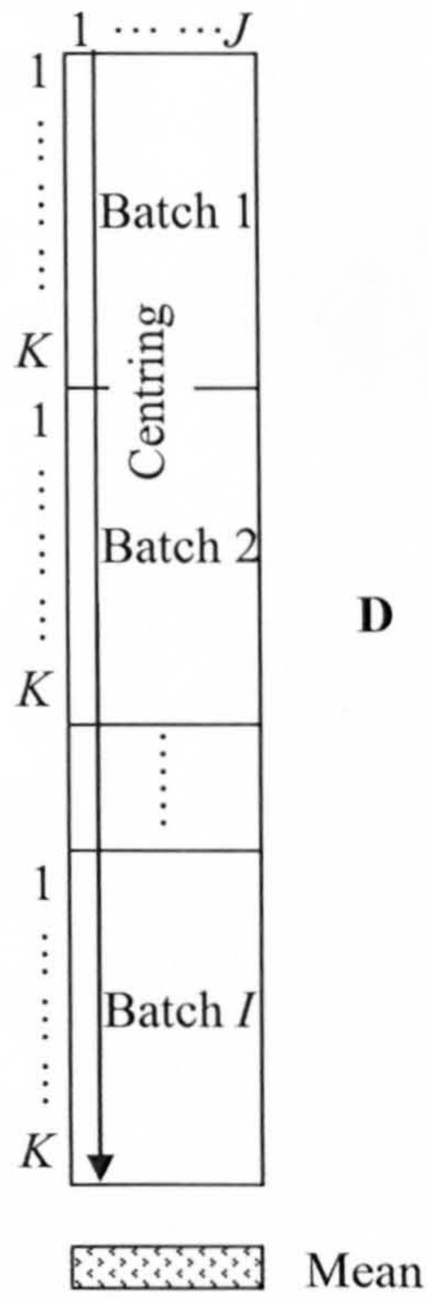
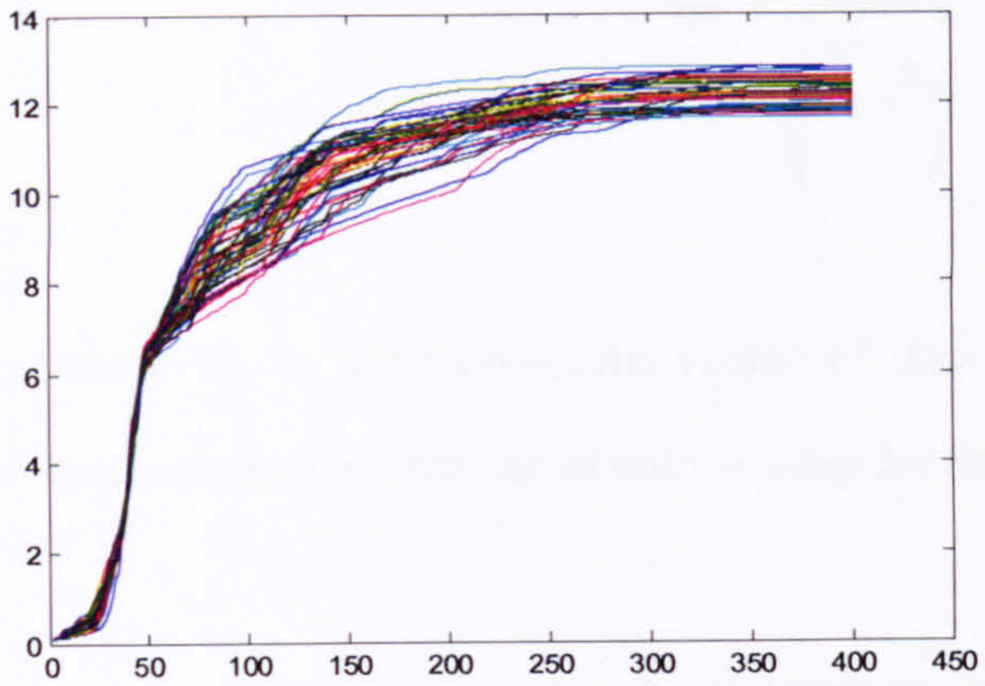
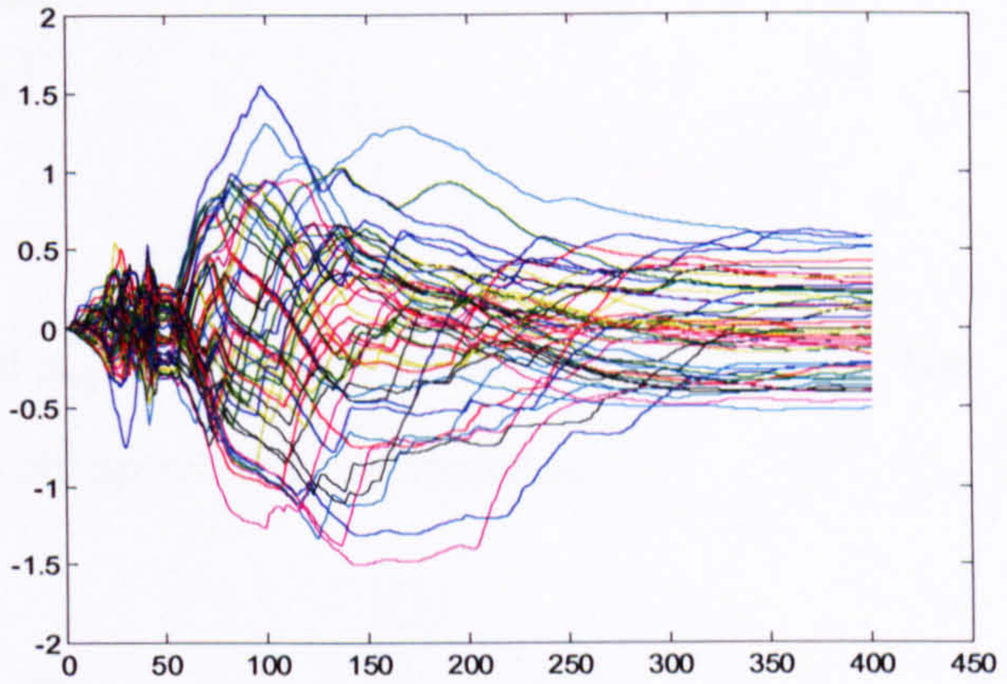


Figure 3-2 Illustration of mean-centring of Wold unfolding approach

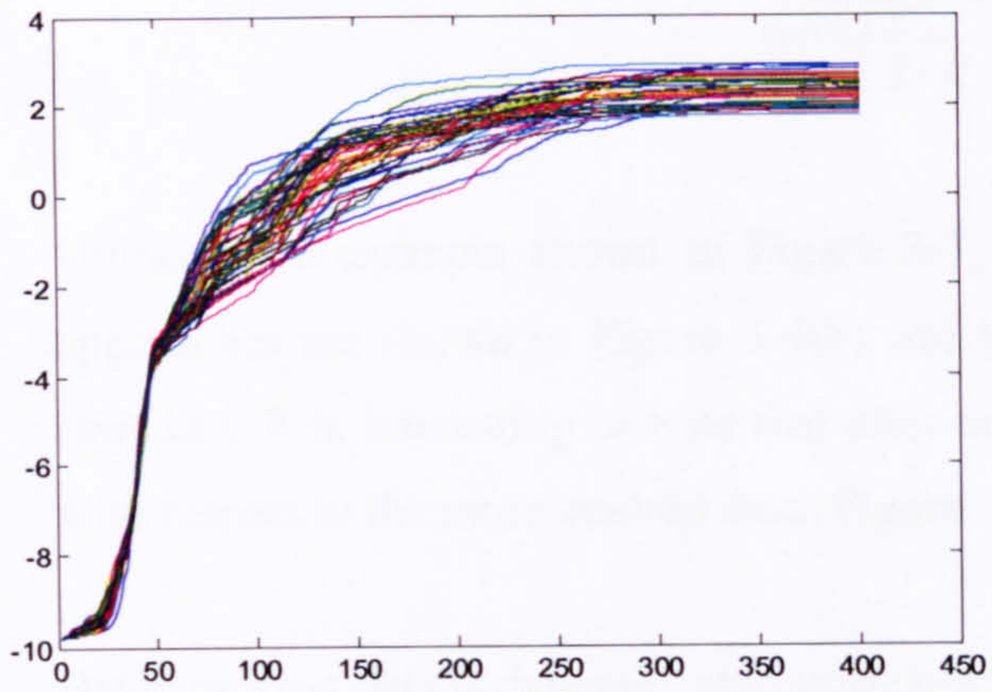
An example of a variable trajectory is shown in Figure 3-3. Figure 3-3(a) shows the raw data for 50 batches. Figure 3-3(b) and (c) illustrate the mean-centred data for the N&M and Wold unfolding approaches respectively. A significant difference is observed between the two different ways of unfolding and centring the matrices. The mean trajectories are removed in the case of the N&M approach. Consequently applying PCA to this data results in systematic variation in all the variable trajectories about their mean trajectories as shown in Figure 3-3(b). However, for the Wold methodology, the non-linear and time-varying trajectories are retained after applying centring, consequently the scores from the application of the multivariate projection technique describe the average trajectories. The trace in Figure 3-3(c) shows the same pattern as the raw data thus the time-varying trajectory still exists although the domain of the data is shifted to zero-mean.



(a)



(b)



(c)

Figure 3-3 Example of mean-centred data: (a) raw data; (b) N&M unfolding approach; (c) Wold unfolding approach

Scaling

The second step after centring is that of scaling. Since the scale of the variables differ, it is necessary to scale the data so that each variable exhibits similar level of variability. In particular in the application of the bi-linear techniques of PCA and PLS, a variable with a large variance will dominate the first few principal components or latent variables. If no prior information is known about the variables, each variable should be scaled to have equal variance thereby ensuring that each variable is given an equal chance of contributing to the final model. In the case where prior information is available, this insight should be reflected in the scaling.

The most widely applied procedure is that of auto-scaling. For the N&M approach, the mathematical expression of auto-scaling is expressed as:

$$\mathbf{x}_{ijk}^{+} = \frac{\mathbf{x}_{ijk}^{*}}{\sqrt{\frac{\sum_{i=1}^{I}\left(\mathbf{x}_{ijk} - \bar{\mathbf{x}}_{jk}\right)^2}{I-1}}} \qquad 3\text{-}3$$

where $\mathbf{x}_{ijk}^{+}$ is the auto-scaled vector of data and $\mathbf{x}_{ijk}^{*}$ is the mean-centred vector of data. The mathematical expression of auto-scaling for the Wold approach is expressed as:

$$\mathbf{x}_{ikj}^{+} = \frac{\mathbf{x}_{ikj}^{*}}{\sqrt{\frac{\sum_{k=1}^{K}\sum_{i=1}^{I}\left(\mathbf{x}_{ikj} - \bar{\mathbf{x}}_{j}\right)^2}{I \cdot K - 1}}}. \qquad 3\text{-}4$$

Utilising the example shown in Figure 3-3, the auto-scaled results for the N&M and Wold approaches are shown in Figure 3-4(b) and (c) respectively. Figure 3-4(a) shows the reference raw data. It is interesting to note that after auto-scaling, the scale of Figure 3-4(b) has changed with respect to the mean centred data, Figure 3-3(b).

Batch process data is dynamic, auto-correlated and cross-correlated. By performing normalisation, the major non-linear and dynamic components in the data are claimed to be removed (Nomikos and MacGregor, 1995b). Consequently if a process disturbance occurs, the correlation structure between the variables will change and hence the model will identify a change in normal operation.

Variable Weighting

When a large number of variables are being considered, blocking of the variables may be an option. Blocking can be in terms of a group of related variables, a specific operating region of a process, or a group of batches that were processed in the same reaction vessels. If blocking is applied, an additional scaling factor is typically introduced and domination from a set of high total variance variables can be avoided. Two types of weighting can be considered: hard and soft block scaling. Hard block scaling scales the variables in a block so that the sum of their variances is equal to unity and is expressed as:

$$v_b = \frac{1}{\sqrt{j_{block}}} \qquad 3\text{-}5$$

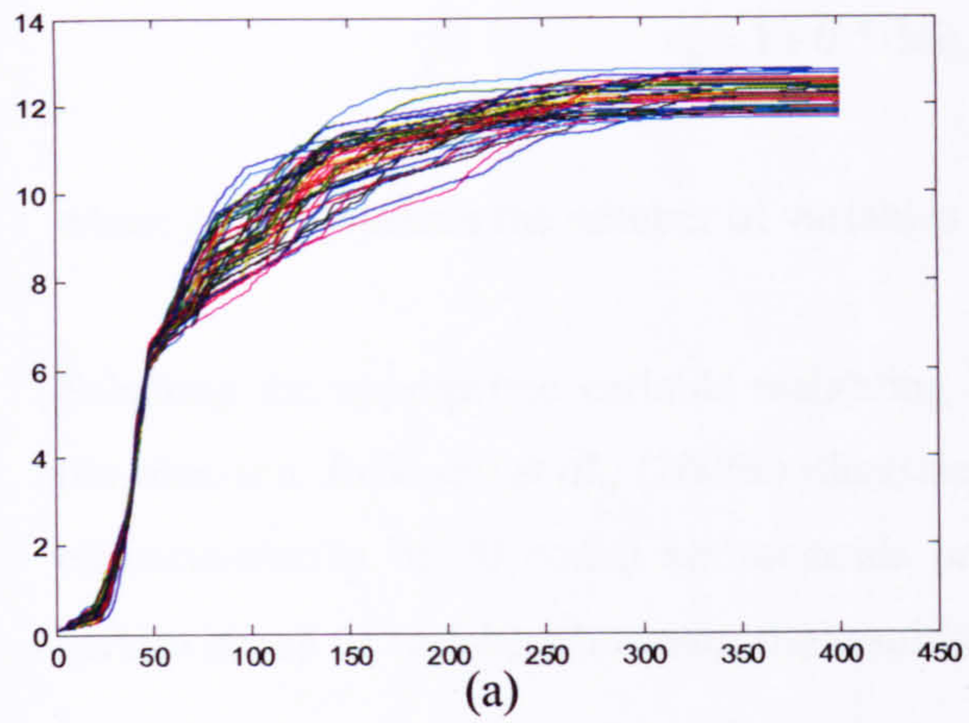


Figure 3-4 Example of mean-centred and auto-scaled data: (a) raw data; (b) N&M unfolding approach; (c) Wold unfolding approach

where j_{block} represents the number of variables in block b. In soft block scaling, each block of variables is scaled so that the sum of the variable's variances is equal to the square root of the number of variables in that particular block. The scaling weight is expressed as:

$$v_b = \frac{1}{\sqrt[4]{j_{block}}} \qquad 3\text{-}6$$

where j_{block} represents the number of variables in block b.

Mild weighting was applied in hierarchical models (Wold *et al.*, 1996) and is based on the number of variables contained within each block. For instance to focus the modelling on a particular block of variables, mild weighting can be applied to up-weight a particular block that has greater importance and conversely down-weight another block which is of less importance. Mild weighting is expressed as:

$$v_b = 1 + 0.5 \cdot \log_{10} \cdot j_{block} \qquad 3\text{-}7$$

where j_{block} represents the number of variables in block b.

Selecting the appropriate variable weighting is dependent on understanding of the problem and the situation. Eriksson *et al.*, (2006a) illustrated the impact of variable weighting to a multivariate characterisation of 20 coded amino acids problem. A subtle difference was found between a certain group of variables however the resulting scores plots were similar.

3.2.3 Alignment of Batch Length

In industrial situations, total batch duration and the duration of various stages within a batch are not normally of equal length for different batches even where a sophisticated control strategy is implemented. Figure 3-5 illustrates the unequal batch issue for the two unfolding methods. As observed from Figure 3-5 the impact of unequal batch length duration will affect the subsequent analysis for the N&M method whilst for the Wold approach the multivariate projection techniques of multi-way PCA and PLS can still be applied directly.

Since the application of N&M approach requires the data matrix of order $IxJK$ where K is the same for all batches, a number of methods have been proposed to handle the issue of unequal batch duration, including cutting to minimum length, change of axis to indicator variable and multivariate Dynamic Time Warping (DTW) (Rothwell, 1999). Alternatively if there is a large historical database available, it may be possible to select batches that are roughly of equal duration since the large population of data should have included sufficient nominal batch variation. With respect to the Wold approach, time points are the row vectors that arranged in a vertical direction hence length equalisation is not an issue and this is observed to be a significant advantage of this approach.

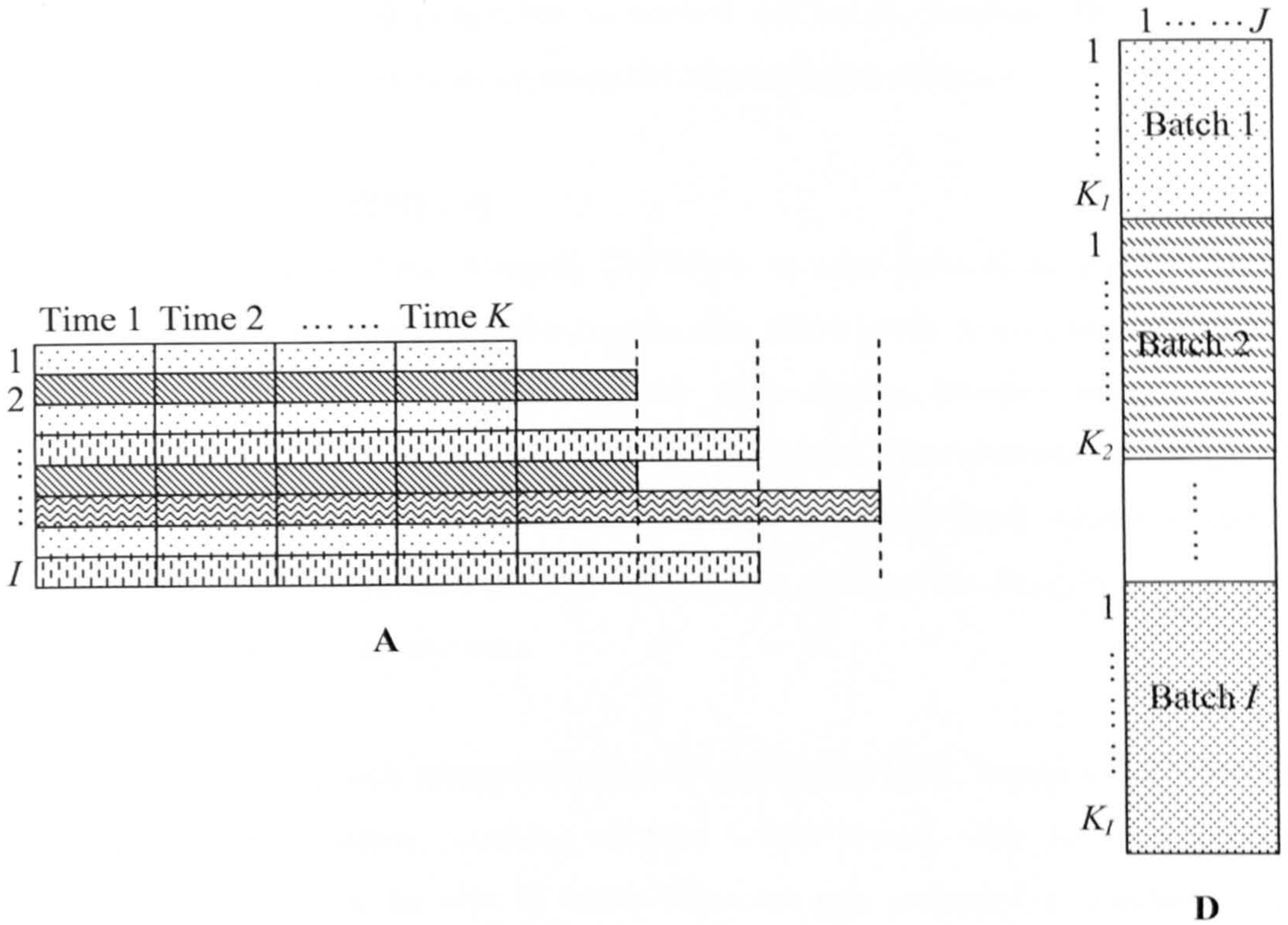


Figure 3-5 Illustration of unequal batch lengths for two unfolding methods

3.2.3.1 Cutting to Minimum Length

This method simply cuts the length of all batches to the selected minimum batch length, K_{min}. When adopting this method, care needs to be taken when defining K_{min} since it can impact on the quality of the process representation as well as on the range of process faults identified. This is because important process information may be removed. This method is applied when the batch durations in the reference database are not significantly different. It is process dependent and is a user defined method.

3.2.3.2 Change of Axis to Indicator Variable

Typically observations are recorded with respect to time, however for this approach, the observations are re-sampled based on another variable which has the properties of being smooth, continuous and monotonic. This surrogate variable, will span the range of all other variables within the dataset and the measurements of other variables are re-sampled at equal intervals based on this variable.

This method is applicable where a variable exhibits the afore mentioned properties (Martin *et al.*, 2002). Adopting this approach, results in no information being lost and time is included as an additional variable. In many industrial processes, time is treated as a control parameter but this is not necessary the best control strategy since the change in physical or chemical entities may be

more critical in determining batch completion. It is often that over-reaction or under-reaction occurs if time is a registered parameter to control the batch duration. The application of a surrogate variable may result in more representative monitoring schemes.

3.2.3.3 Dynamic Time Warping

The objective of Dynamic Time Warping (DTW) is to map features in a test pattern onto a reference pattern. This is analogous to the equalisation of the batch process trajectories by similar events in each batch run being aligned. The time-varying features within the data are synchronised by time normalisation. This process is known as "time warping" as the data patterns are translated, compressed and expanded in localised segments until similar features in the patterns between the reference data and the test data are matched resulting in the test data being of the same length as the reference data.

DTW has its origins in speech recognition (Sakoe and Chiba, 1978; Myers *et al.*, 1980) and is a flexible, deterministic, pattern matching scheme which works with pairs of patterns. The implementation of DTW in the area of batch processes was proposed by Gollmer and Posten (1996). The application focused on the detection of important process features during a fermentation process. A univariate DTW scheme was proposed to identify phases in batch cultivation and the detection of faults in a fed-batch cultivation. In 1998, multivariate DTW was proposed by Kassidas *et al.* (1998). The multivariate approach considers the variables simultaneously whilst the univariate approach aligns the trajectory within a variable. In the paper of Kassidas *et al.* (1998), a framework for off-line and on-line batch trajectory equalisation utilising industrial data from a polymerisation reactor was reported. Ramaker *et al.* (2003) went on to propose the application of dynamic time warping to spectral data for batch processes. The DTW procedure was modified by defining new weighting factors to address sections that contain no warping information.

3.2.4 Process Data Assessment

Prior to applying multivariate projection techniques, a preliminary screening of the data at a single batch level is carried out. At this stage, it is not necessary for the data to be formatted into a three-way matrix. There are occasions that even by looking at the individual variables and their trajectories for an individual batch, process problems can be identified immediately without applying more sophisticated methods.

A number of underlying assumptions concerning the data such as independence and normality require to be investigated prior to accepting a model. These assumptions are often violated in

practical applications. The effect of the violation of the normality assumption is not pursued in detail on this work.

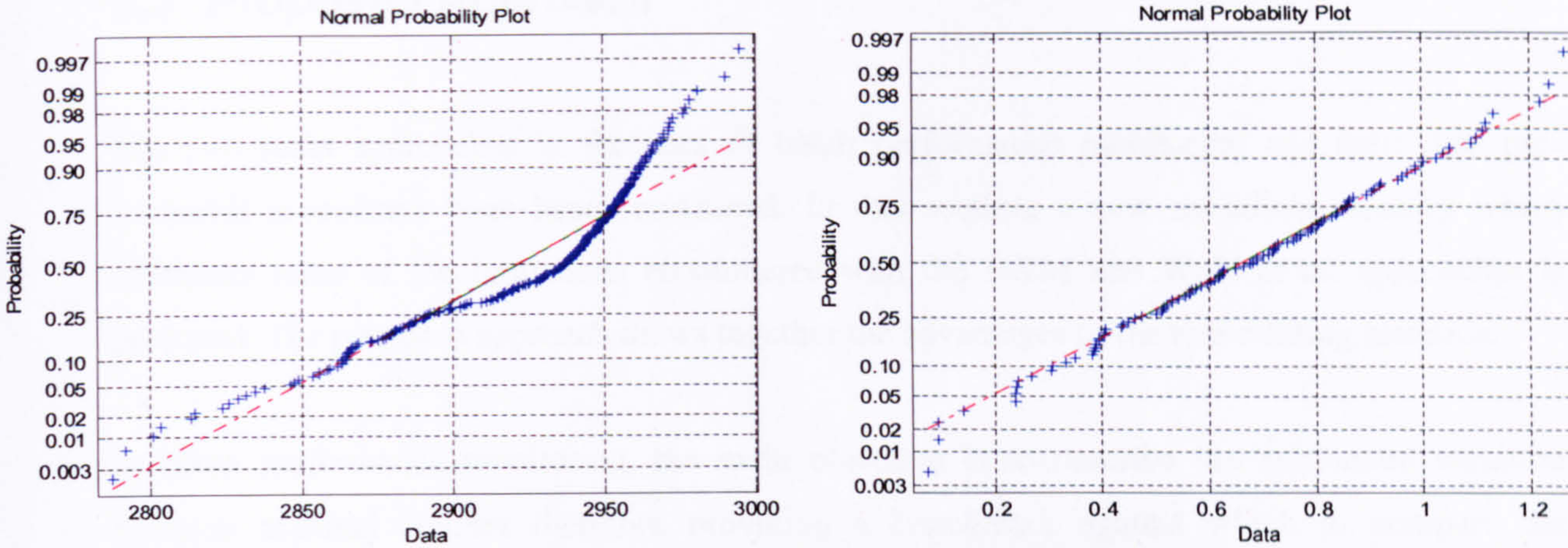
The normality for each variable is assessed by utilising a normal probability plot. An important requirement of MSPC is that the variables follow a normal distribution. This type of behaviour can be identified through the use of a normal probability plot. The normal probability plot is constructed by ordering the data from the smallest to largest, the order number (n) is then used to calculate the probability:

$$\text{probability} = \frac{2n-1}{2I} \qquad 3\text{-}8$$

I is the number of samples. Figure 3-6 shows the normal probability plots for the original data and the N&M unfolding and scaling approach. If the data lie approximately on a straight line then the data is normally distributed. In this case, auto-scaling has removed the non-normal component in the data resulting in normally distributed data.

Auto-correlation is an issue in process performance monitoring. It measures the correlation between observations at different time points. The majority of batch manufacturing processes exhibit dynamic behaviour and rarely remain at steady state. A steady state process implies that the process variables vary about a fixed time point. The time series structure of the data has an effect on the distribution of false alarms. Since most monitoring techniques are based on the assumption of steady state, the presence of dynamic or auto-correlated behaviour can affect the detection capability of the monitoring model. Figure 3-7 shows the impact of scaling and how the auto-correlation structure in the data is affected. Some degree of auto-correlation is observed in the original raw data for an individual variable as shown in Figure 3-7(a) however the auto-correlation effect is removed by applying auto-scaling to the N&M unfolded data (Figure 3-7(b).

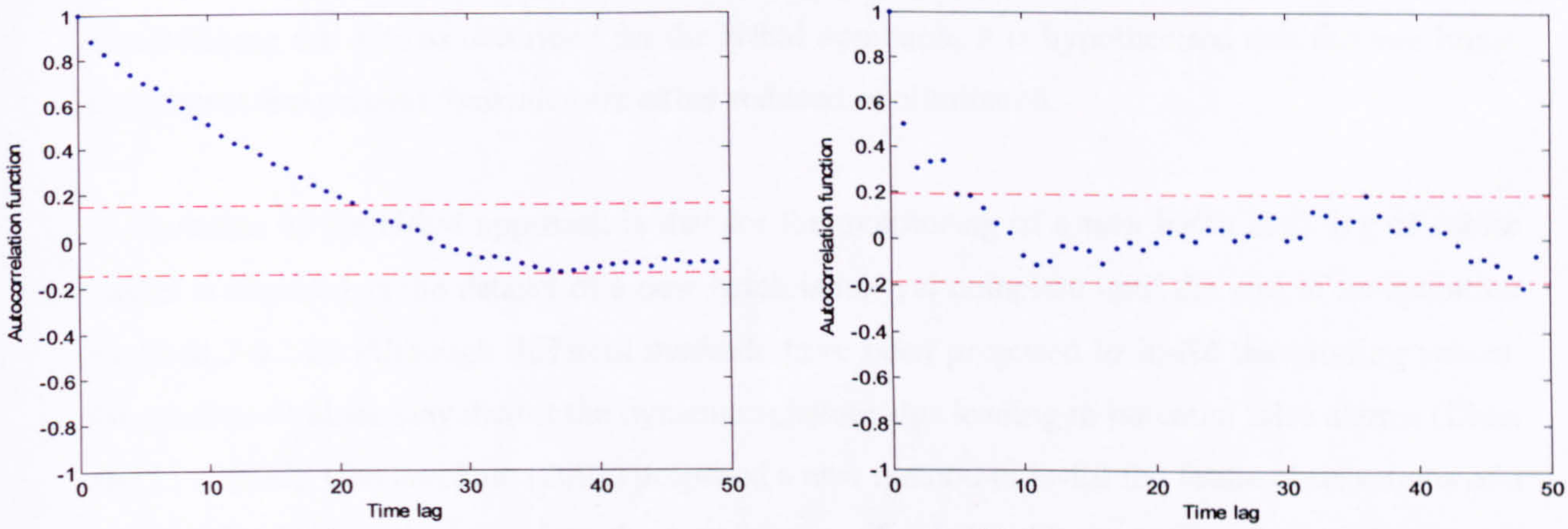
In contrast by applying the Wold unfolding and scaling approach, the non-normal component and auto-correlation structure in the data are not removed and remained in the data set (Figure 3-8).



(a) Original data

(b) N&M unfolded and scaled data

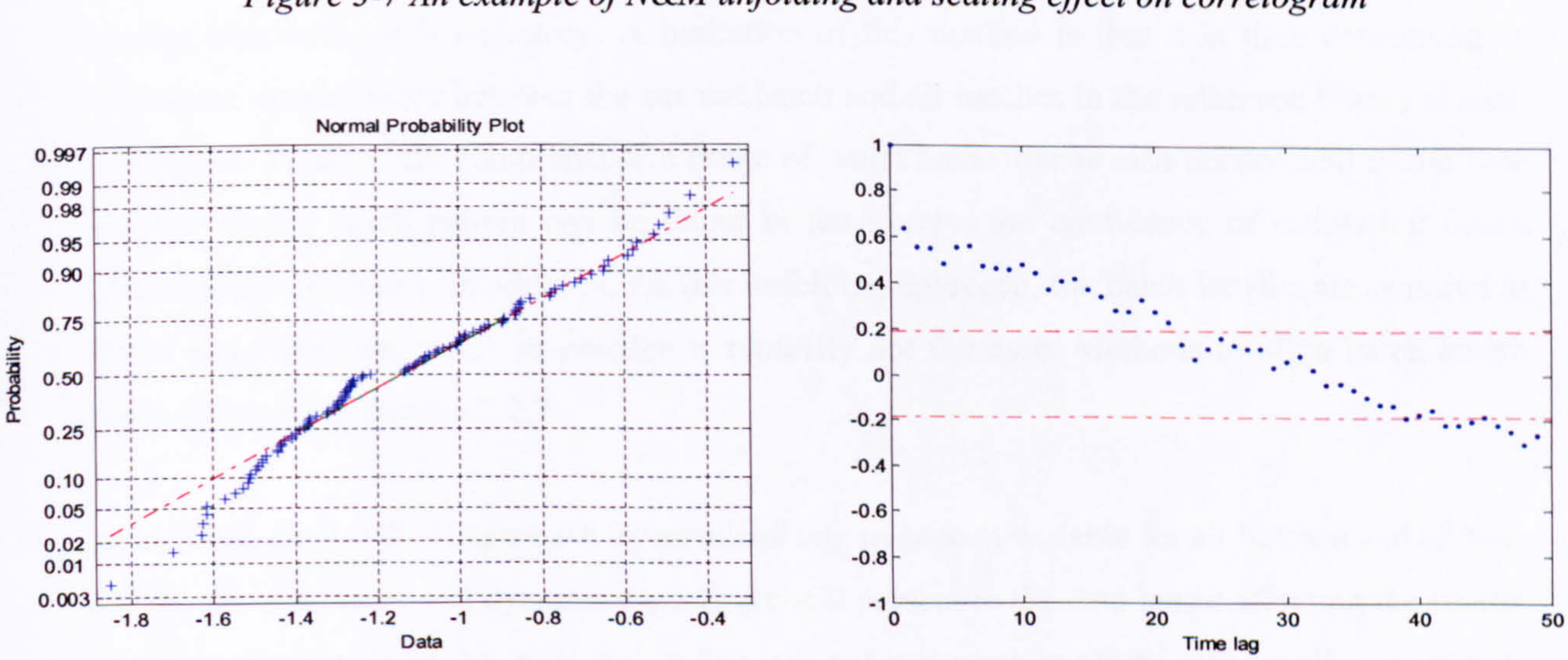
Figure 3-6 An example of N&M unfolding and scaling effect on normal probability plot



(a) Original data

(b) N&M unfolded and scaled data

Figure 3-7 An example of N&M unfolding and scaling effect on correlogram



(a) Normal probability plot

(b) Correlogram

Figure 3-8 Data assessment effect on Wold unfolding and scaling approach

3.3 Proposed Approach

The two main approaches in the area of batch performance monitoring and their data pre-treatment procedures have been introduced. In this section, a new modelling strategy which addresses some of the limitations encountered with the N&M and Wold *et al.* approaches is proposed. The proposed approach draws together the advantages of the two existing methods.

In batch performance monitoring, the main objective is to describe the systematic variation between nominal batches therefore providing a benchmark against which to compare the performance of new batches. For the development of a robust monitoring scheme, it is assumed that the data and hence the principal component scores, are independent and normally distributed. However, as most monitoring techniques are based on the assumption of steady state, the presence of dynamic and non-linear behaviour may affect process detection capability. By standardising the data as described for the N&M approach, it is hypothesised that the non-linear component and process dynamics are either reduced or eliminated.

A limitation of the N&M approach is that for the monitoring of a new batch in-filling of future values is required as the dataset of a new batch is not yet complete until the end of its operation (Section 2.6.2.2). Although different methods have been proposed to in-fill the missing values, the predicted values may distort the dynamic relationships leading to potential false alarms (Chen and Liu, 2002). Cho and Kim (2003) proposed a new method to in-fill the future observations of a new batch. This method requires the building up of a batch library where the nominal batch trajectories are stored and then the prediction of new observations is based on searching for a similar historical batch trajectory. A limitation of this method is that it is time consuming to determine the similarity between the current batch and all batches in the reference library at each time point. A relatively comprehensive range of batch behaviour is also needed and in the case that no similar batch pattern can be found in the library, the confidence of estimating future observations decreases. In addition, for this unfolding approach, the batch lengths are required to be of equal duration which in practice is typically not the case. Methods to align batch length were discussed in Section 3.2.3.

In contrast, for the Wold approach by standardising over each variable for all batches and all time points, the non-linear and dynamic aspects are still present in the data hence affecting the scores. However the benefit of this approach is that it is not necessary to align batch lengths or estimate future observations. The inclusion of the **Y** matrix, batch maturity, results in some unique features in the final representation. The first component typically explains the variation in the **X** matrix

that is most highly correlated to time. The second component represents the variation that changes quadratically over time and so on. Overall batch maturity drives the covariance structure in the **X** matrix. Employing this approach has provided a reference as to how a new batch progresses when compared against the nominal batches so that any deviation in the new batch can be detected through the principal component scores.

For the proposed approach, the first part is based on unfolding the three-way data matrix $\underline{\mathbf{X}}$ (I x J x K) into a two-dimensional matrix **X** (I x KJ) as per the N&M approach. The unfolded data is then centred and scaled to unit variance. This procedure enables a reduction in the non-linear and dynamic behaviour in the data thereby focusing the modelling on the nominal deviations about the mean trajectories. The unfolded matrix is then rearranged into the form **X** (IK x J) as shown in Figure 3-9, i.e. the Wold *et al.* approach. The bi-linear technique of PCA is then applied to this data matrix and the resulting scores vector describes the variation of the through batch mean trajectories. The loadings contain information about the overall variable information. The directions of variable and batches are now both preserved hence the potential to monitor batch-to-batch variations and through batch behaviour for each variable is now captured in a single monitoring model. The overall graphical representation of the proposed approach is presented in Figure 3-9.

The approach has a number of advantages for the on-line monitoring of batch processes. The first is that the mean trajectory of the batches is removed therefore the resulting data matrix is more suitable for the application of the multivariate statistical projection techniques due to the reduction of the non-linear and dynamic components. Furthermore, the proposed approach does not require the prediction of future missing values and the alignment of different batches of different durations. In general, the proposed approach should provide an alternative way of monitoring batch processes. Its performance is evaluated against other methods in Section 3.4. A summary of the steps is given below.

Stage A: Historical data analysis and nominal model building

1. Collect data from historical batches reflective of good operation.
2. Check for missing data and apply data pre-processing techniques as necessary, see Section 3.2.1.
3. Unfold the three-way matrix $\underline{\mathbf{X}}$ (I x J x K) to **X** (I x KJ).
4. Centre and scale the data to remove the batch mean trajectory, see Section 3.2.2.
5. Re-arrange the unfolded matrix **X** (I x KJ) to **X** (IK x J) where K can differ between batches. For simplicity, equal batch length is assumed.
6. Apply PCA to build the nominal model.

7. Re-arrange the scores time-wise to determine the 95% and 99% confidence limits, i.e. ± 2 and 3 standard deviations respectively at each time point (Figure 3-9).
8. Calculate Hotelling's T^2 and the SPE metrics and associated confidence limits.
9. Investigate if the model is valid and contains nominal batch-to-batch variation. If abnormal variation is detected, model should be re-built to reflect only common cause variation, i.e. remove abnormal nominal batches.

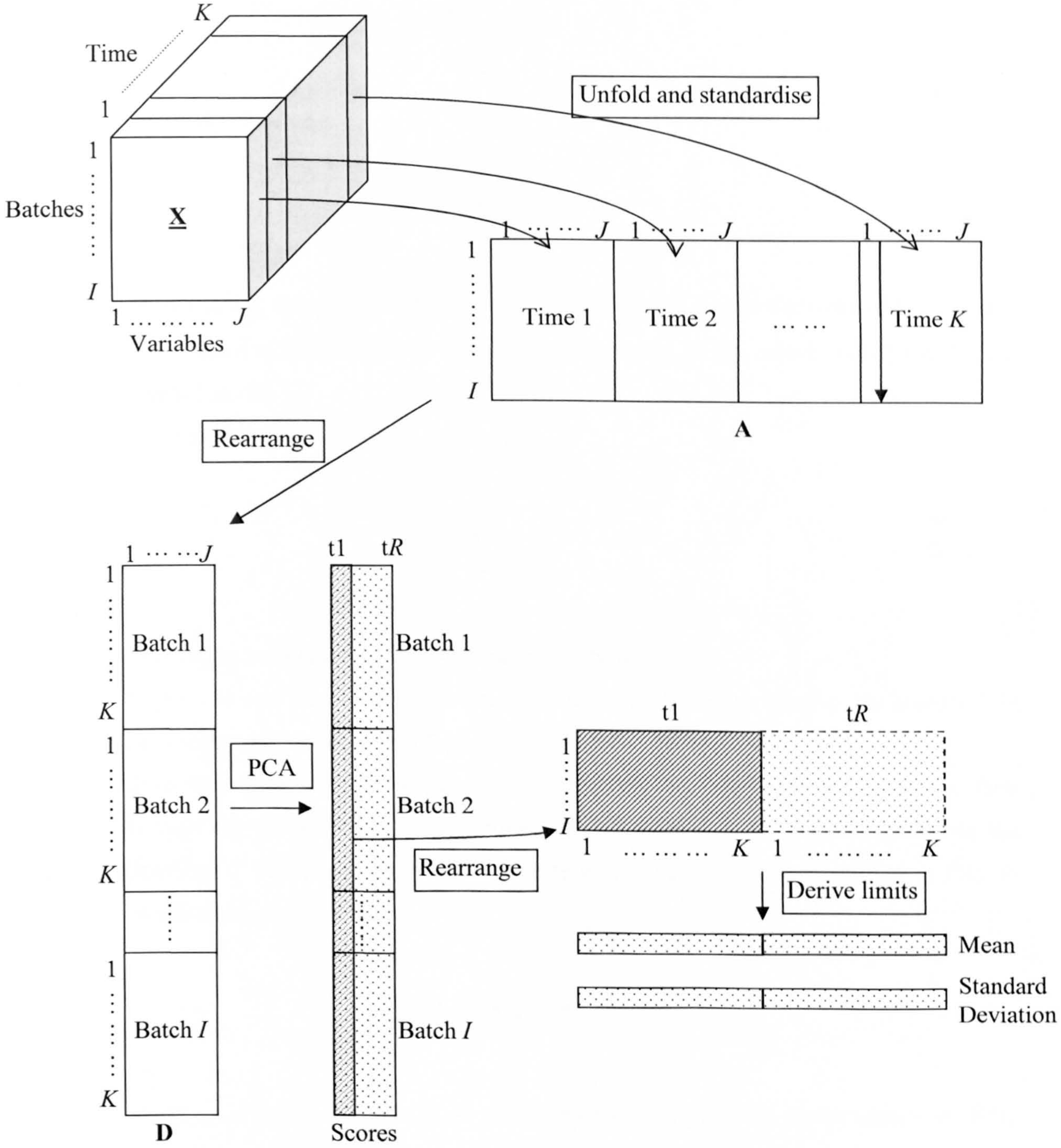


Figure 3-9 Illustration of the proposed approach for on-line batch monitoring

Stage B: Investigate the performance of unseen batches

1. For a new batch at time point k, arrange the vector in the format $\mathbf{x}_{New}(1 \times KJ)$, $k = 1, 2 \ldots K$.
2. Apply the same scaling factors to the vector from Step 4, Stage A.
3. Re-arrange the new batch to the format $\mathbf{X}_{New}(K \times J)$, $k = 1 \ldots K$.
4. Project the new vector onto the loading matrix to calculate the new scores:

$$\mathbf{t}_{New,kj} = \mathbf{X}_{New,kj} \cdot \mathbf{p} . \qquad 3\text{-}9$$

5. Calculate Hotelling T^2:

$$T_k^2 = (\mathbf{t}_{New,kj} - \bar{\mathbf{t}}_k)^T \mathbf{S}_k^{-1} (\mathbf{t}_{New,kj} - \bar{\mathbf{t}}_k) \qquad 3\text{-}10$$

where $\mathbf{t}_{New,kj}$ is the scores for the new batch at time point k, $\bar{\mathbf{t}}_k$ is the mean of the columns of the score matrix and $\mathbf{S}_k$ is the covariance matrix of the scores calculated for the nominal model.

6. Calculate the SPE:

$$SPE_k = \sum_{j=1}^{J} \mathbf{e}_{kj}^2 \qquad 3\text{-}11$$

where $\mathbf{e}_{kj}$ is the prediction error of variable j at time point k.

7. Project the new metrics onto the control charts and determine whether the process is in statistical process control (confidence limits).
8. If an abnormal situation is detected, identify the variables contributing to the fault through the application of contribution plots (Section 2.4.5). The contribution to the Hotelling T^2 for a new set of batches at time k, $\mathbf{x}_{New,kj}$, can be represented as follows (Westerhuis *et al.*, 2000a):

$$c_k^{T^2} = \mathbf{t}_{New,kj}^{\mathrm{T}} \cdot \mathbf{S}_k^{-1} \cdot (\mathbf{x}_{New,kj} \cdot \mathbf{p}_j^{\mathrm{T}})^{\mathrm{T}} \qquad 3\text{-}12$$

where $c_k^{T^2}$ is the contribution of the variables to T_k^2. The variable contribution to the SPE, c_k^{SPE}, can be calculated as:

$$c_k^{SPE} = e_{kj}^2 . \qquad 3\text{-}13$$

3.4 On-line Monitoring Performance Evaluation

3.4.1 Introduction

Following the introduction of the proposed batch process monitoring approach, its performance for on-line monitoring and fault detection capability is evaluated and compared against the afore mentioned existing approaches. The focus is on those approaches that develop models of the process data. Simulated batch process data is considered for the evaluation of the different techniques. In addition, two types of modelling are considered – a global model and a local model. Control charts of the principal component scores, Hotelling's T^2 and SPE statistics are compared and used to assess the performance of the model. Two performance indices, the false alarm rate and the out-of-control average run length (ARL), are considered. These are equivalent to type I and type II errors respectively.

In a statistical hypothesis, two kinds of errors may occur. If the null hypothesis (H_0) is rejected when it is true, then a type I error has occurred. If the null hypothesis is not rejected when it is false, a type II error has been made. The probabilities of these two types of errors are summarised as:

$$\alpha = P\{\text{type I error}\} = P\{\text{reject } H_0 \mid H_0 \text{ is true}\}$$
$$\beta = P\{\text{type II error}\} = P\{\text{fail to reject } H_0 \mid H_0 \text{ is false}\}$$

The properties, advantages and limitations of the different approaches are discussed and final recommendations are drawn that address the hypothesis of whether superior performance is evident in terms of the proposed approach over existing methods.

3.4.2 False Alarm Rate

A false alarm is where the control chart identifies a fault, when in practice a fault has not occurred. A high false alarm rate may indicate the fault detection technique is too sensitive. On the other hand, a low false alarm rate increases the rate of missed detection and hence the technique is not able to recognise a real fault. If the type I error is not satisfactory, the limits may be required to be adjust so that the required level of α is attained. A balance between type I and type II errors is required for assured process performance monitoring. The false alarm rate is calculated as the percentage of samples over the entire batch duration that fall outside the 95%

and 99% confidence limits based on an average over the reference data of nominal batches. As the batches are identified to reflect nominal operating conditions, it is assumed that the targets are within statistical control. Therefore a crossing of the control limit is considered to be a false alarm however theoretically it is expected that 5% and 1% of samples may lie outside the 95% and 99% confidence limits respectively. In the case of the local model approach where the process is separated into two phases, the false alarm rates for the different phases will be reported as the summation of the false alarms from the individual local models. In all cases the metrics of the principal component scores, Hotelling's T^2 and SPE are assessed.

3.4.3 Out-of-Control Average Run Length

A missed detection is the situation where a fault has not been diagnosed, although a fault has occurred. This is a similar measure to the ARL which is commonly employed for fault detection. The ARL is defined as the average number of observations between the fault occurring and the fault being detected for a number of batches. The ARL can only be calculated from theory or using numerical simulations but not from industrial data. To evaluate the out-of-control ARL, additional batches are generated. The batches are generated under the same conditions as the nominal batches except that a fault is introduced. By determining the number of observations between the occurrence of the fault and its detection for the principal component scores, Hotelling's T^2 and SPE, a measure of the out-of-control ARL is obtained.

3.4.4 Description of the Data and Modelling Approaches

Penicillin simulation data is used for the evaluation of the different approaches (Birol *et al.*, 2002). The production of secondary metabolites such as penicillin has been studied widely in both academia and industry (Atkinson and Mavituna, 1991). Such an antibiotic is generally produced using filamentous micro-organisms. The form of this target product is not usually developed during the cell growth stage. Hence, it is common practice to first grow the micro-organisms in a batch culture followed by a fed-batch operation to promote the synthesis of the antibiotic. The penicillin fermentation can be divided into four physiological phases (pre-culture, exponential cell growth, stationery and cell death) and two operational phases. The first operational mode is a batch process in which the first two physiological phases are developed. The last two physiological phases are performed as fed-batch operations. In the first operational phase, most of the cell is cultivated during the initial pre-culture stage and penicillin starts to be produced during the exponential growth phase. Then the process is switched to fed-batch operation where glucose is fed until the end of the fed-batch mode to maintain high penicillin productivity (Birol *et al.*, 2002). A typical time series structure of penicillin cultivation is shown in Figure 3-10.

A simulation of penicillin production, Pensim, which was developed by Birol *et al.* (2002) is used as the basis of the comparative study of the different monitoring methodologies. The development of this simulation was based on a realistic dynamic model of a fed-batch penicillin fermentation process therefore the process is considered to exhibit non-linear, dynamic and multi-stage features. The input and output structure of the process is described in Figure 3-11.

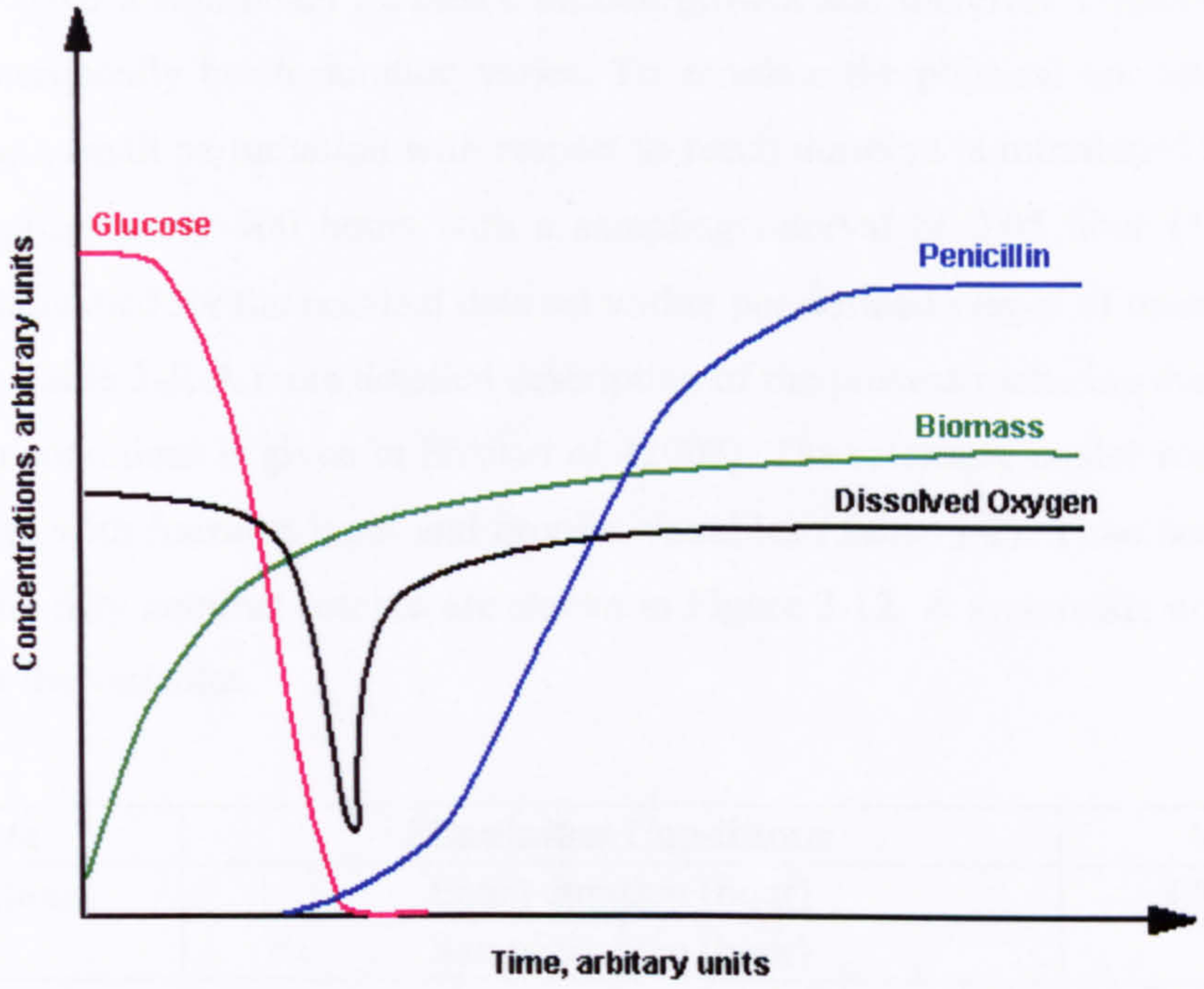


Figure 3-10 Time series of penicillin cultivation

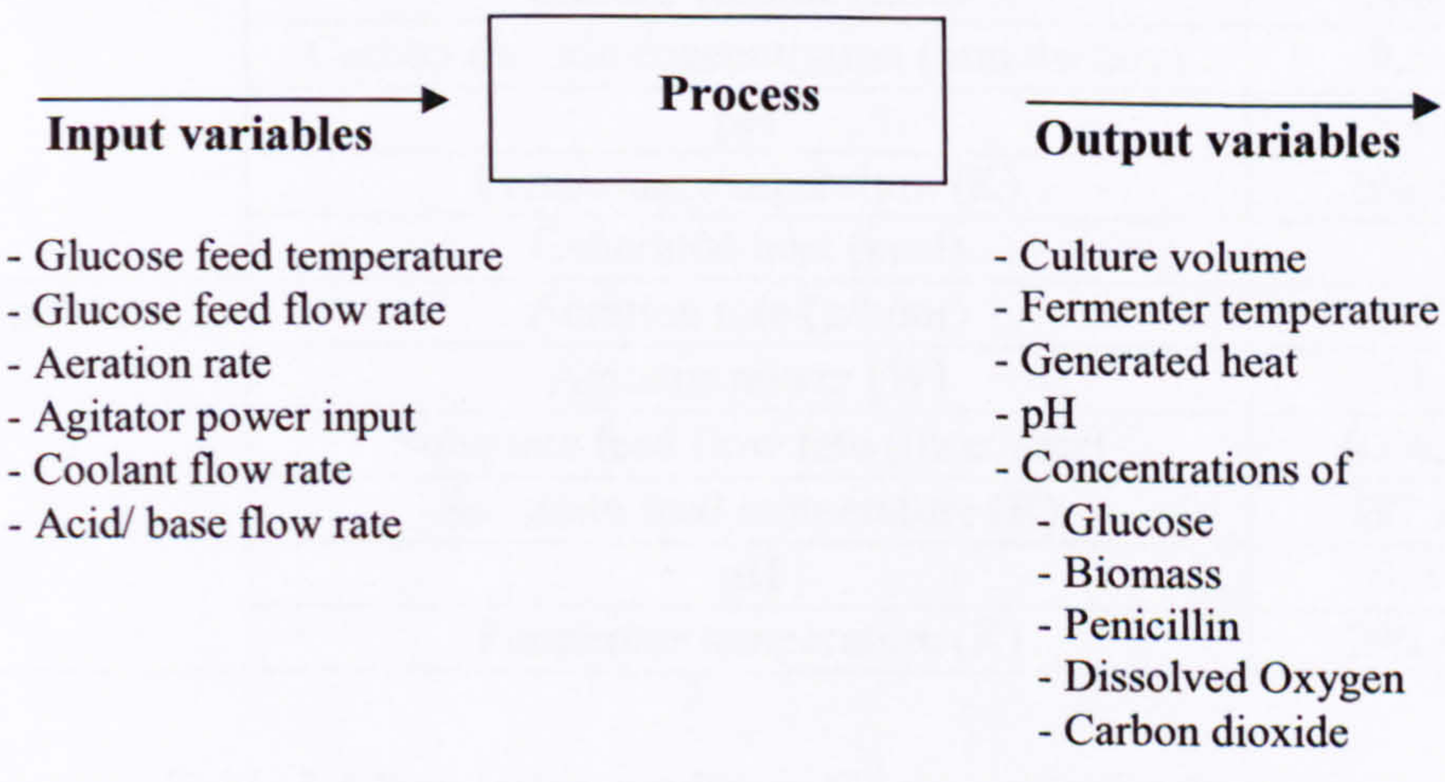


Figure 3-11 Input and output structure of the penicillin process

The simulated process is regulated by closed-loop proportional-integral-derivative (PID) controllers of pH and temperature since these variables play an important role with respect to the quality and quantity of the final product whilst glucose addition is performed under open-loop

control. The system switches itself from batch mode to fed-batch operation when the glucose (carbon source) reaches a threshold value of 0.3 g/l. This usually takes about 45 hours. At this stage, the process switches to fed-batch operation where glucose is fed constantly until the end of the process.

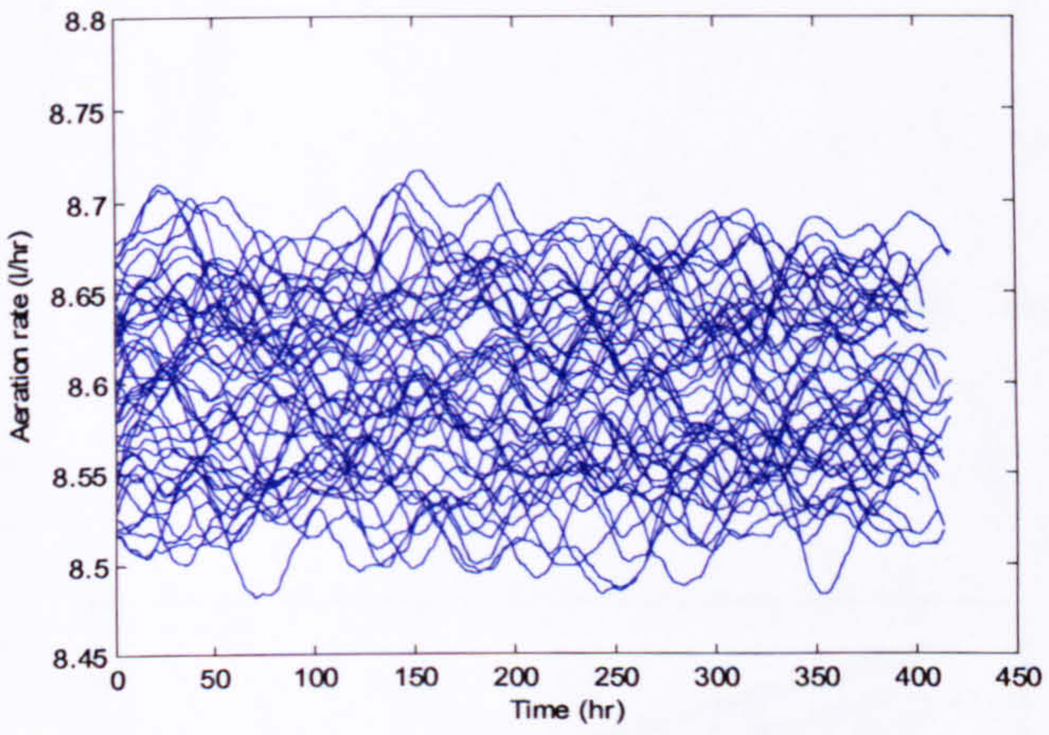
In a batch fermentation process, slight changes in operating conditions during critical periods of operation may have a significant influence on cell growth and therefore influence final quality and yield. Consequently batch duration varies. To simulate the physical uncertainty present in each batch run, a small perturbation with respect to batch duration is introduced but the average duration is approximately 400 hours with a sampling interval of 0.05 hour (3 minute). Fifty batches were simulated for the nominal data set within per-defined ranges of operation which are summarised in Table 3-1. A more detailed description of the process including the state equations and simulation conditions is given in Birol *et al.* (2002). The reference model was built from the nominal batches with fourteen input and process variables (Table 3-2). Time series plots of the variables for the fifty nominal batches are shown in Figure 3-12. A systematic nominal variation is observed for the variables.

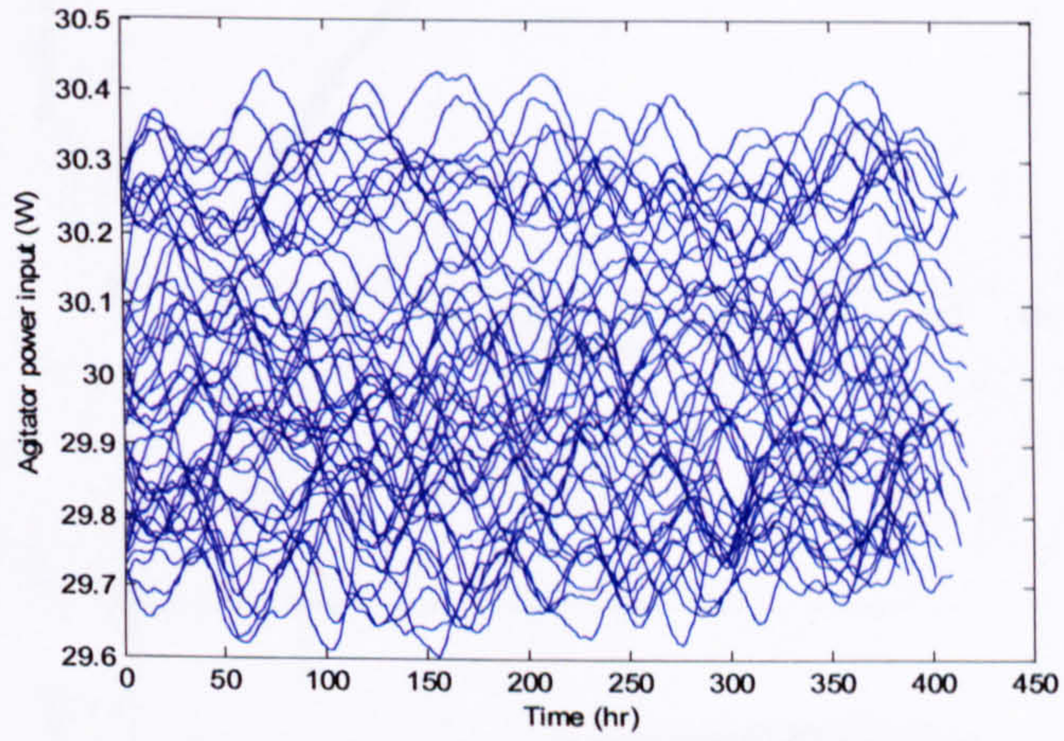
State	Simulation Conditions	Value
Duration	Batch duration (hour)	400 ± 5%
	Sampling time (hour)	0.05
Initial conditions	Substrate concentration (g/litre)	15 ± 5%
	Dissolved oxygen concentration (mmole/litre)	1.16
	Biomass concentration (g/litre)	0.1 ± 5%
	Penicillin concentration (g/litre)	0
	Culture volume (litre)	100 ± 5%
	Carbon dioxide concentration (mmole/litre)	0.5 ± 1%
	pH	5 ± 0.1%
	Fermentor temperature (K)	298 ± 0.1%
	Generated heat (kcal)	0
Set points	Aeration rate (g/hour)	8.6 ± 1%
	Agitator power (W)	30 ± 1%
	Substrate feed flow rate (litre/hour)	0.042 ± 1%
	Substrate feed temperature (K)	297 ± 0.1%
	pH	5 ± 0.1%
	Fermentor temperature (K)	298 ± 0.1%

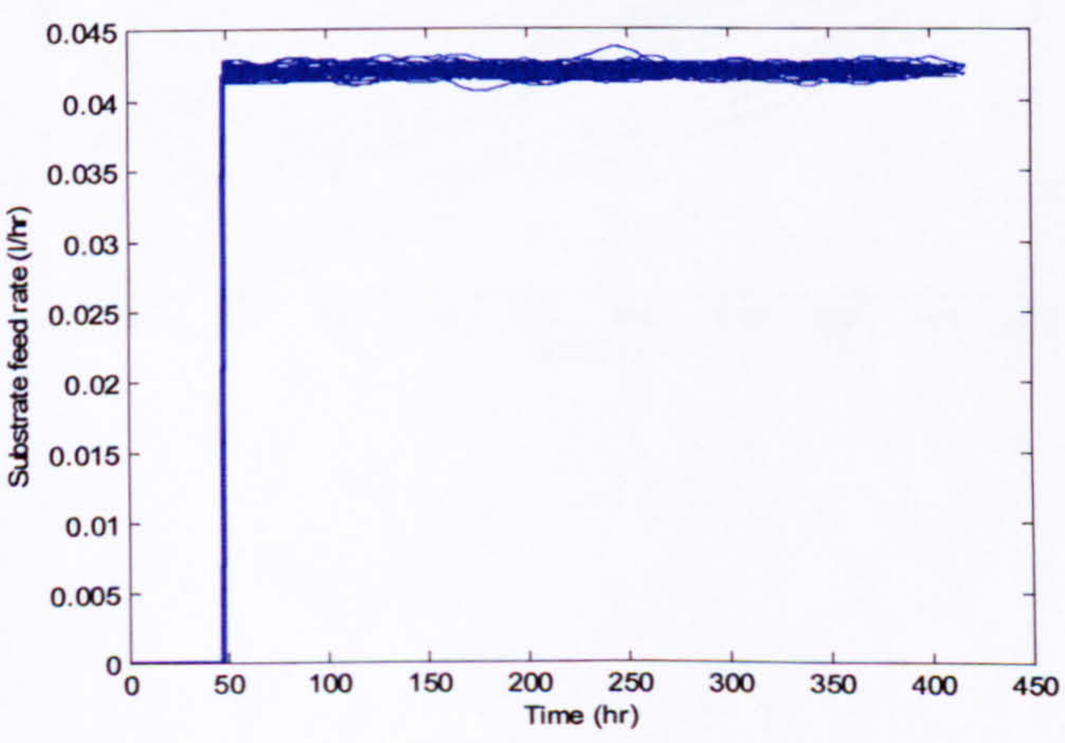
Table 3-1 Simulation conditions for the penicillin process

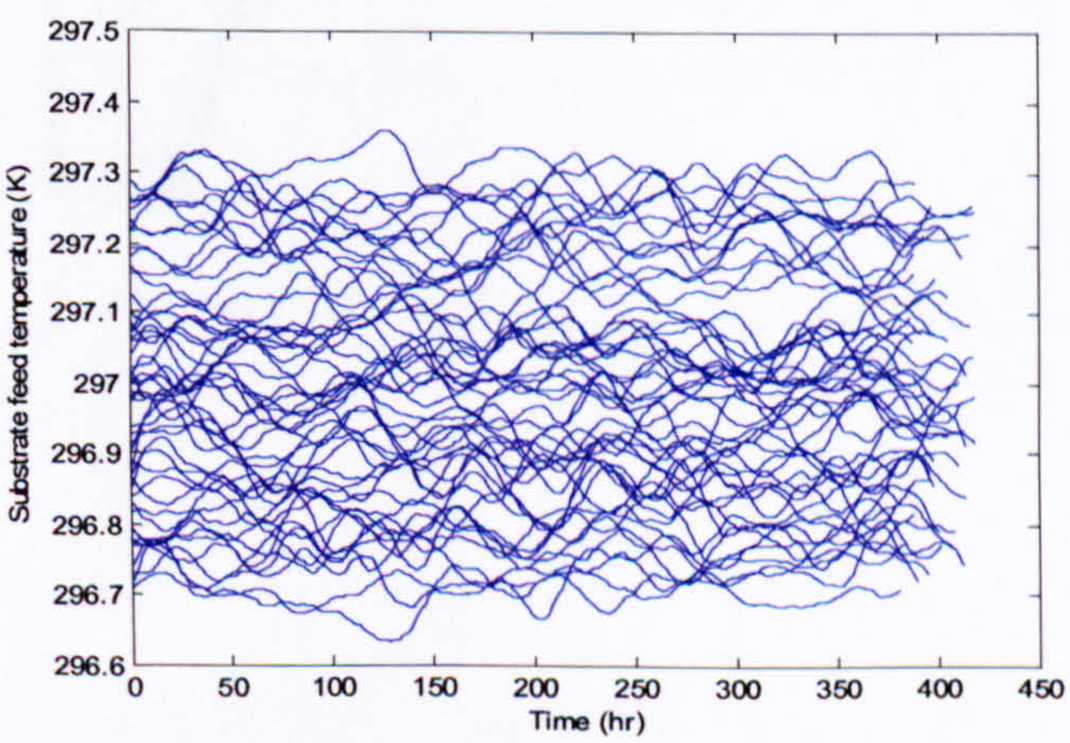
Variable Number	Definition	State
1	Aeration rate (litre/hour)	Input
2	Agitator power input (W)	Input
3	Substrate feed rate (litre/hour)	Input
4	Substrate feed temperature (K)	Input
5	Substrate concentration (g/litre)	Process
6	Dissolved oxygen concentration (mmole/litre)	Process
7	Biomass concentration (g/litre)	Process
8	Penicillin concentration (g/litre)	Process
9	Volume (litre)	Process
10	Carbon dioxide concentration (mmole/litre)	Process
11	pH	Process
12	Fermentor temperature (K)	Process
13	Generated heat (kcal)	Process
14	Cooling water flow rate (litre/hour)	Input

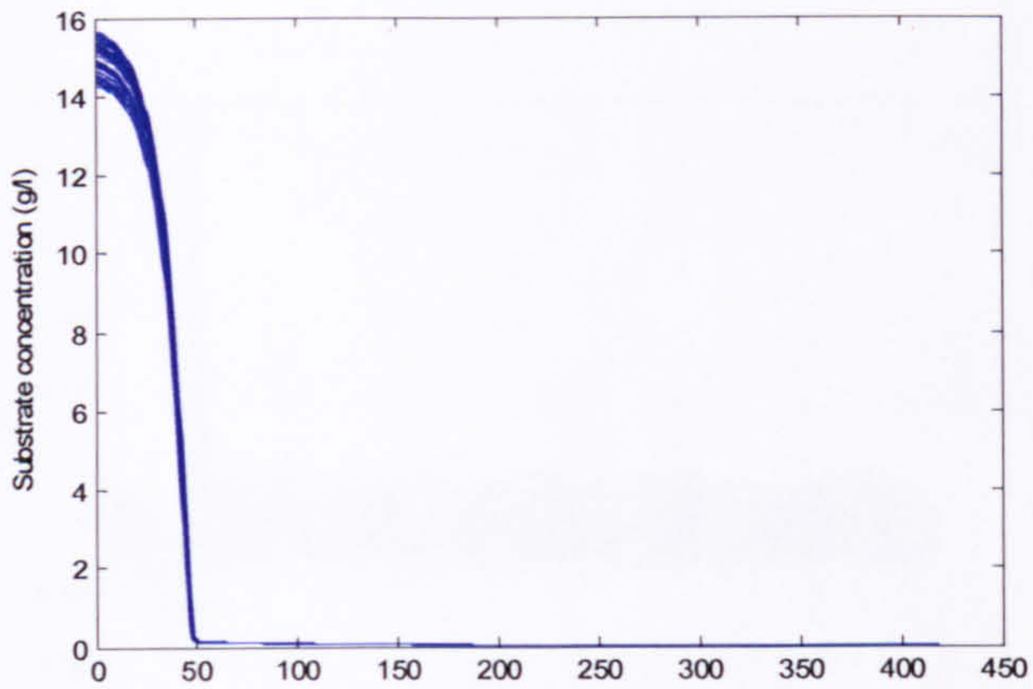
Table 3-2 Input and process variables

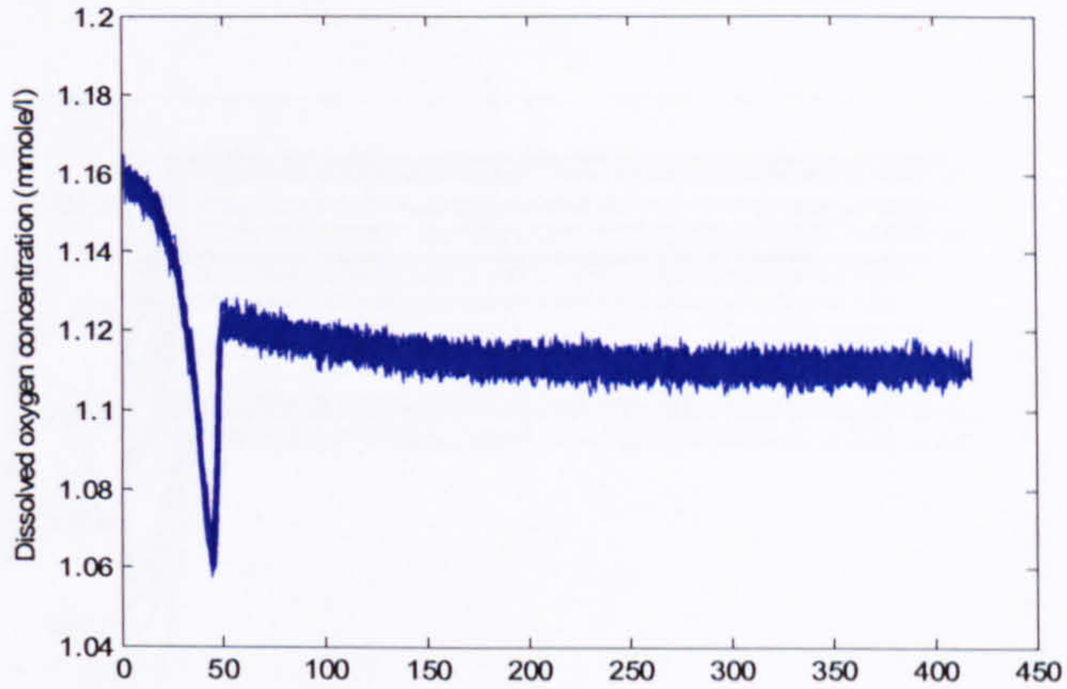


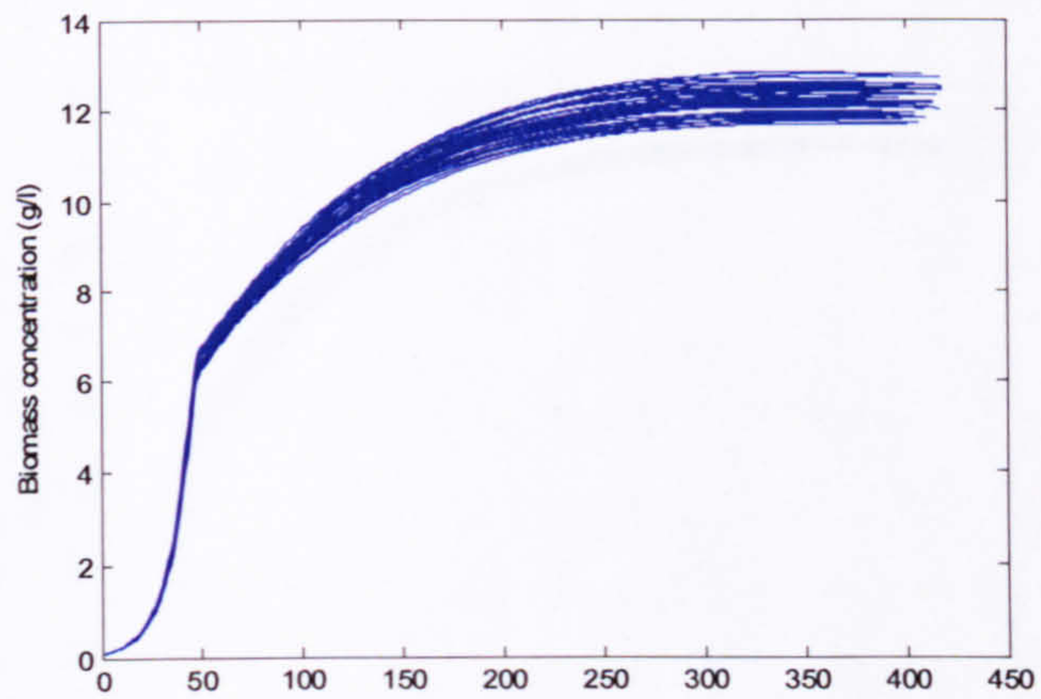


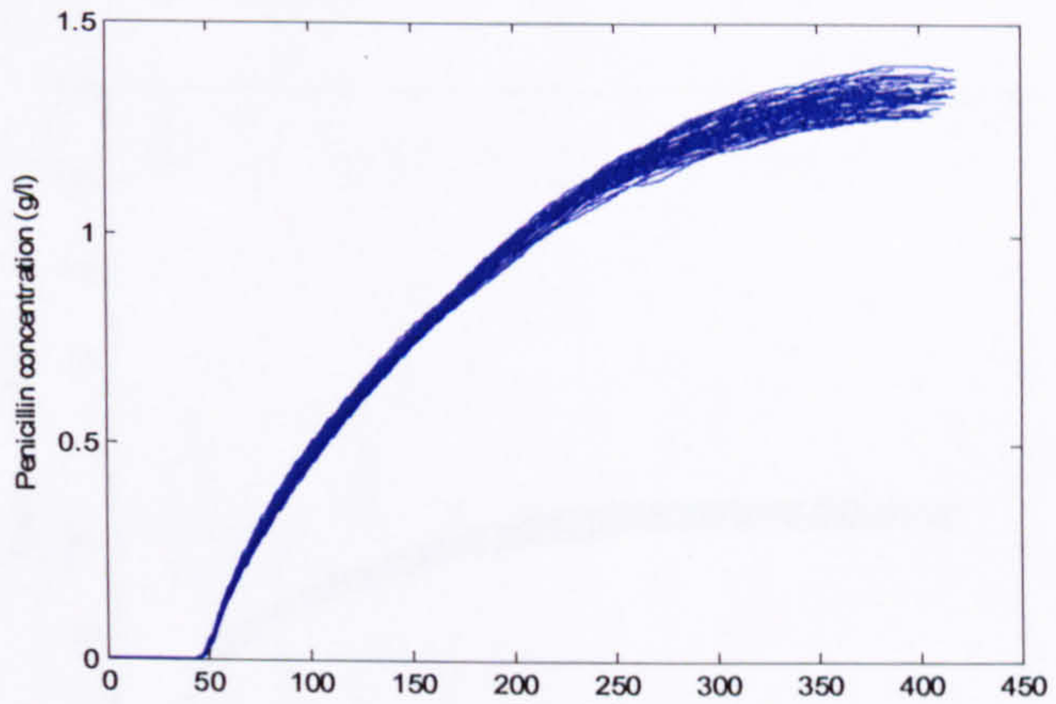


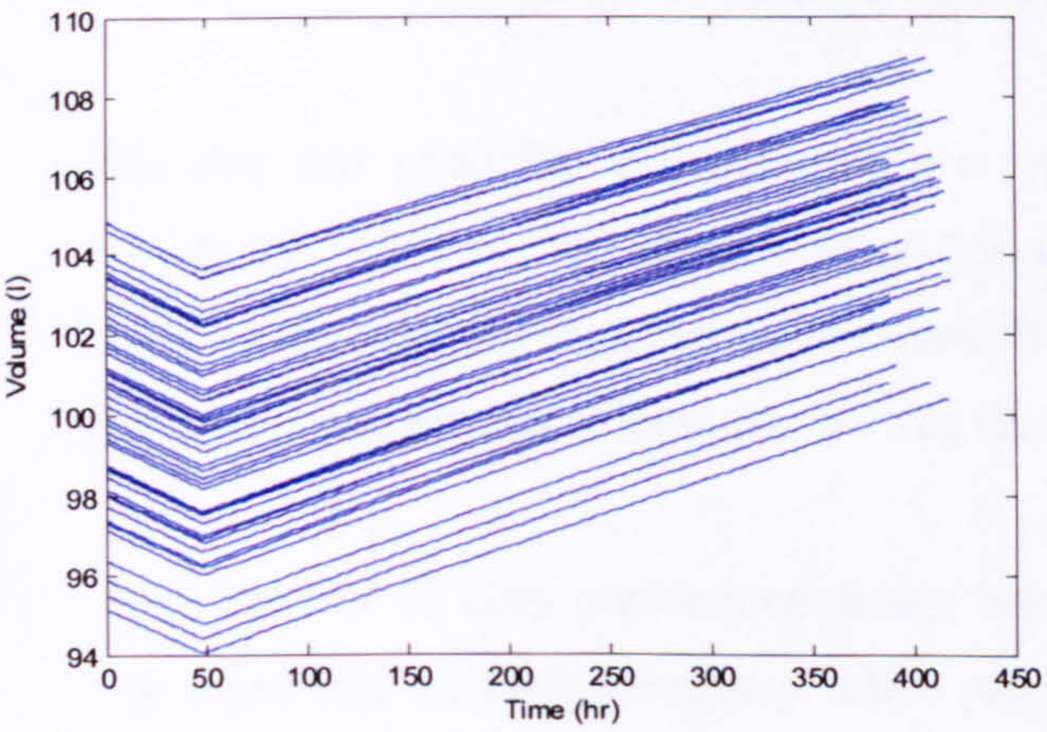


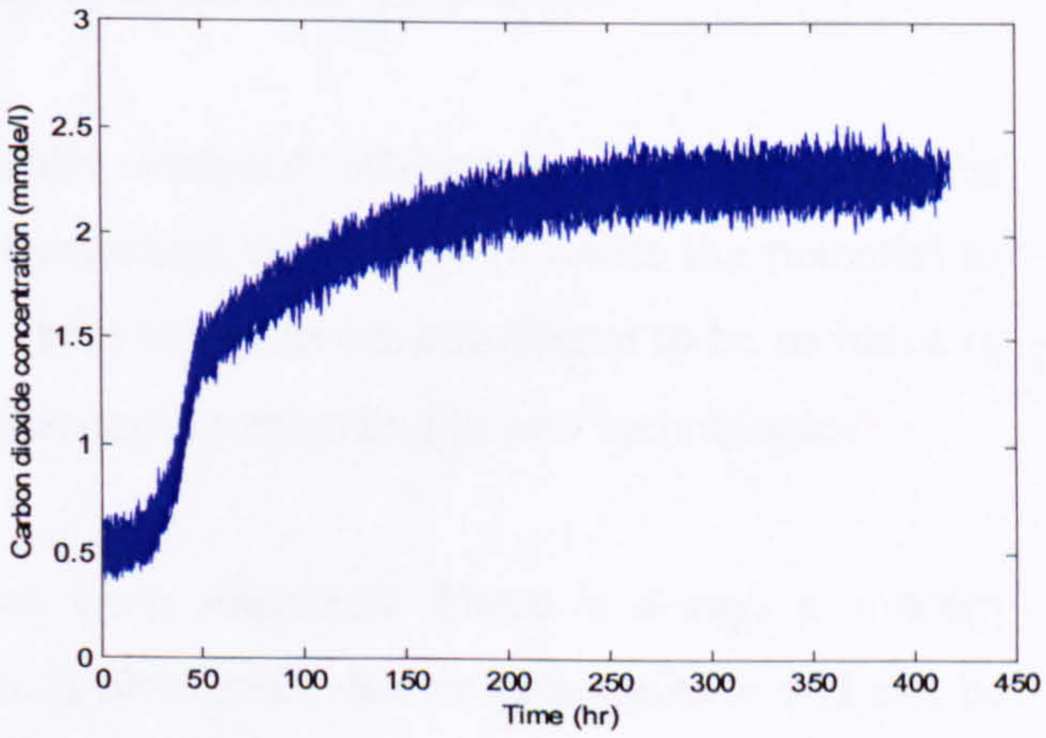

Substrate concentration (g/l)
Time (hr)


Dissolved oxygen concentration (mmole/l)
Time (hr)


Biomass concentration (g/l)
Time (hr)


Penicillin concentration (g/l)
Time (hr)


Volume (l)
Time (hr)


Carbon dioxide concentration (mmole/l)
Time (hr)

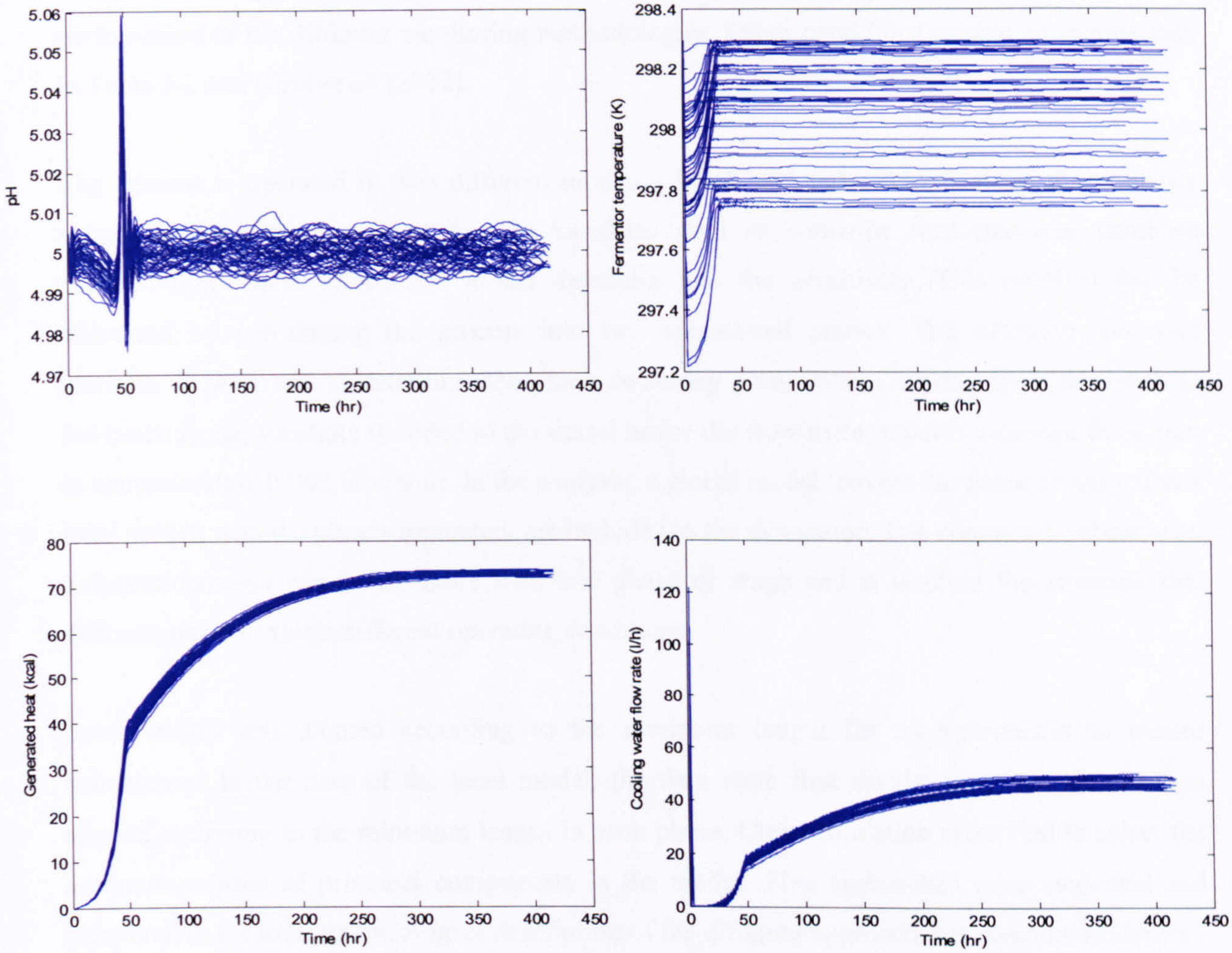


Figure 3-12 Time series plots of input and process variables

Biomass and penicillin concentration are typically analysed off-line in a quality assurance laboratory. However, the advance in on-line measurement technology provides the potential to measure the quality parameters in real-time thus these variables are considered to be included in the monitoring scheme to simulate the way that industry is responding to new technologies.

A further set of fifty pre-defined faulty batches were simulated. There is always a concern expressed that the fault introduced into a process is observable due to its magnitude and can be identified from the univariate analysis therefore the true potential of the multivariate techniques cannot be realised. One of the faults introduced in the original paper (Birol *et al.*, 2002) was that of a 95% step decrease in agitation power. This is a clear malfunction resulting from the control system and is an alarm that will be reported. This kind of fault is often not sufficient to test the true performance of monitoring approaches. A subtle abnormality should be the main interest for fault detection. Therefore the subtle fault introduced for this thesis was a gradual decrease of 0.1

% in the substrate feed rate (variable 3) after 80 hours of operation. These adjustments enabled the generation of a subtle process deviation thereby enabling the evaluation of the true performance of the different monitoring methodologies. Other conditions remain as summarised in Table 3-1 and Birol *et al.* (2002).

The process is operated in two different modes – batch and fed-batch. A sharp discontinuity occurs at the switching point for the variables such as substrate feed rate and substrate concentration which affects the model dynamics and the sensitivity. This problem can be addressed by partitioning the process into two operational phases. The substrate feed rate (variable 3) was used to determine the phase switching point. When moving from the batch to fed-batch mode, substrate is added to the vessel hence the flow meter records a change from zero to approximately 0.042 litre/hour. In the analysis, a global model, covers the entire process, and a local model, with the phases separated, are included in the evaluation. It is common to observe an industrial process which has more than one phase or stage and it is often the situation that different phases exhibit different operating conditions.

Batch length was aligned according to the minimum length for all approaches to ensure consistency. In the case of the local model, the data were first divided into two phases then aligned according to the minimum length in each phase. Cross validation is applied to select the optimum number of principal components in the model. Five approaches were evaluated and compared in the assessment. A brief description of the different approaches is summarised below.

1. Nomikos & MacGregor MPCA approach using zero deviation for in-filling future observations. See Section 2.6.2.2. Auto-scaling was applied to the unfolded matrix.
2. Nomikos & MacGregor MPCA approach using current deviation for in-filling future observations. See Section 2.6.2.2. Auto-scaling was applied to the unfolded matrix.
3. Nomikos & MacGregor MPCA approach using missing data method for in-filling future observations. See Section 2.6.2.2. Auto-scaling was applied to the unfolded matrix.
4. Wold *et al.* approach with **y** vector being the batch maturity index. See Section 2.6.3. Auto-scaling across batch and time trajectories was applied to the unfolded matrix.
5. Proposed PCA approach with auto-scaling being applied to the unfolded matrix. See Section 3.3.

3.4.5 Results and Discussion

The results for the false alarm rate for the different approaches with a 95% confidence limit are shown in Figure 3-13. Figure 3-13(a) illustrates the comparison of metrics between the different

representations using a global model and Figure 3-13(b) focuses on the comparison of the methodologies using a local model with two phases.

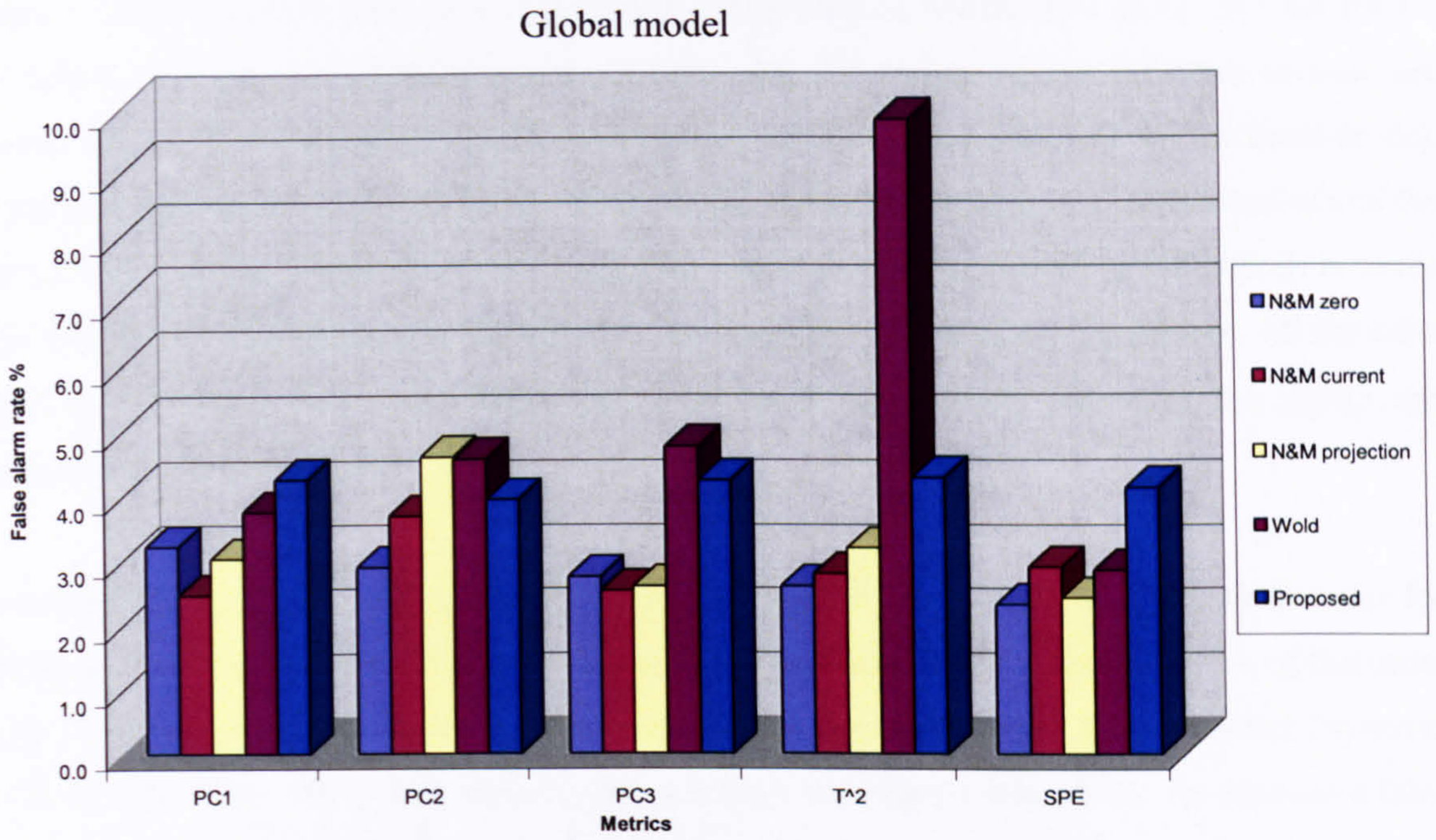


Figure 3-13 False alarm rate for metrics of different approaches for (a) global model; (b) local model

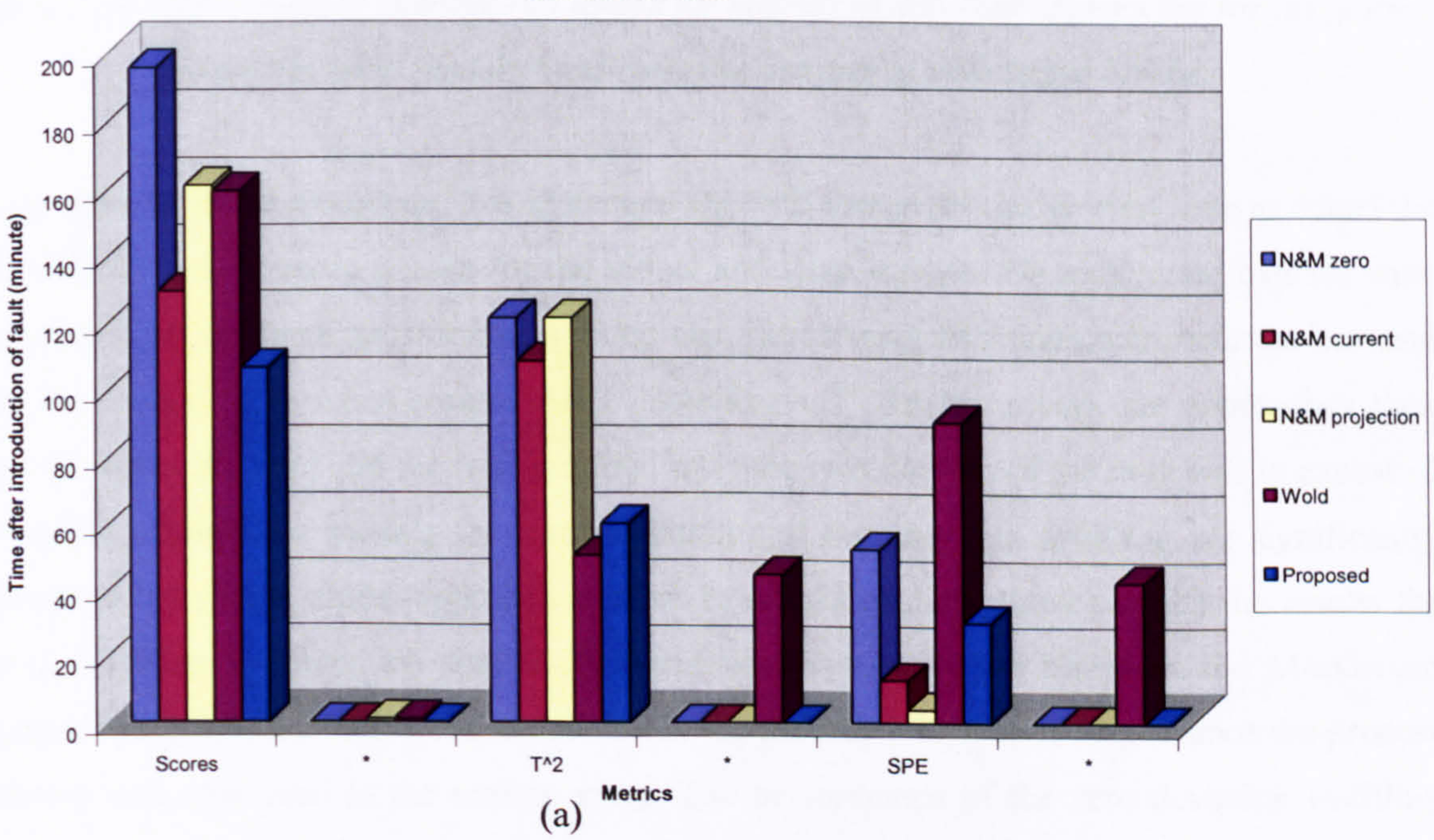
In general it is expected to have a false alarm rate of 5% for a 95% confidence limit to reflect the type I error. The results can be collated into three groups. The first group includes the approaches proposed by Nomikos and MacGregor (1 – 3). This group of approaches requires the estimation of future observations for real time monitoring. Nevertheless, overall this group has the lowest false alarm rate for the different representations and the ranges of the different metrics are between 2% to 3% for the global model with more variation being observed for the local model. For the first principal component for the local model, a 1% false alarm rate is obtained whilst the other metrics exhibit similar results to the global model except for Hotelling's T^2 which is in the range, 1% to 2%. The difference between the three methodologies is not significant. All the false alarm rates are well below the theoretical acceptance value hence indicating the group one approaches are too optimistic.

The second group of approach is based on the approach of Wold *et al.* The false alarm rates for principal components one to three for both the global and local model are similar, i.e. of the order of 4%. The false alarm rate for the SPE for both models is slightly lower than expected. However the Hotelling's T^2 for the global model exhibits a high false alarm rate, 9.5%. An excessive false alarm rate can reduce fault detection accuracy but it is interesting to observe that an improvement in the performance of a model is achieved through the local model. The false alarm rate of Hotelling's T^2 is between 4% to 5% which lies in the acceptable region. This has revealed that for data containing process dynamics and non-linearity, when the monitoring is performed across the whole process and the two distinct process stages are ignored monitoring performance is impaired. When the process is monitored through a local model that takes into account the different stages, the methodology is proven to provide acceptable results.

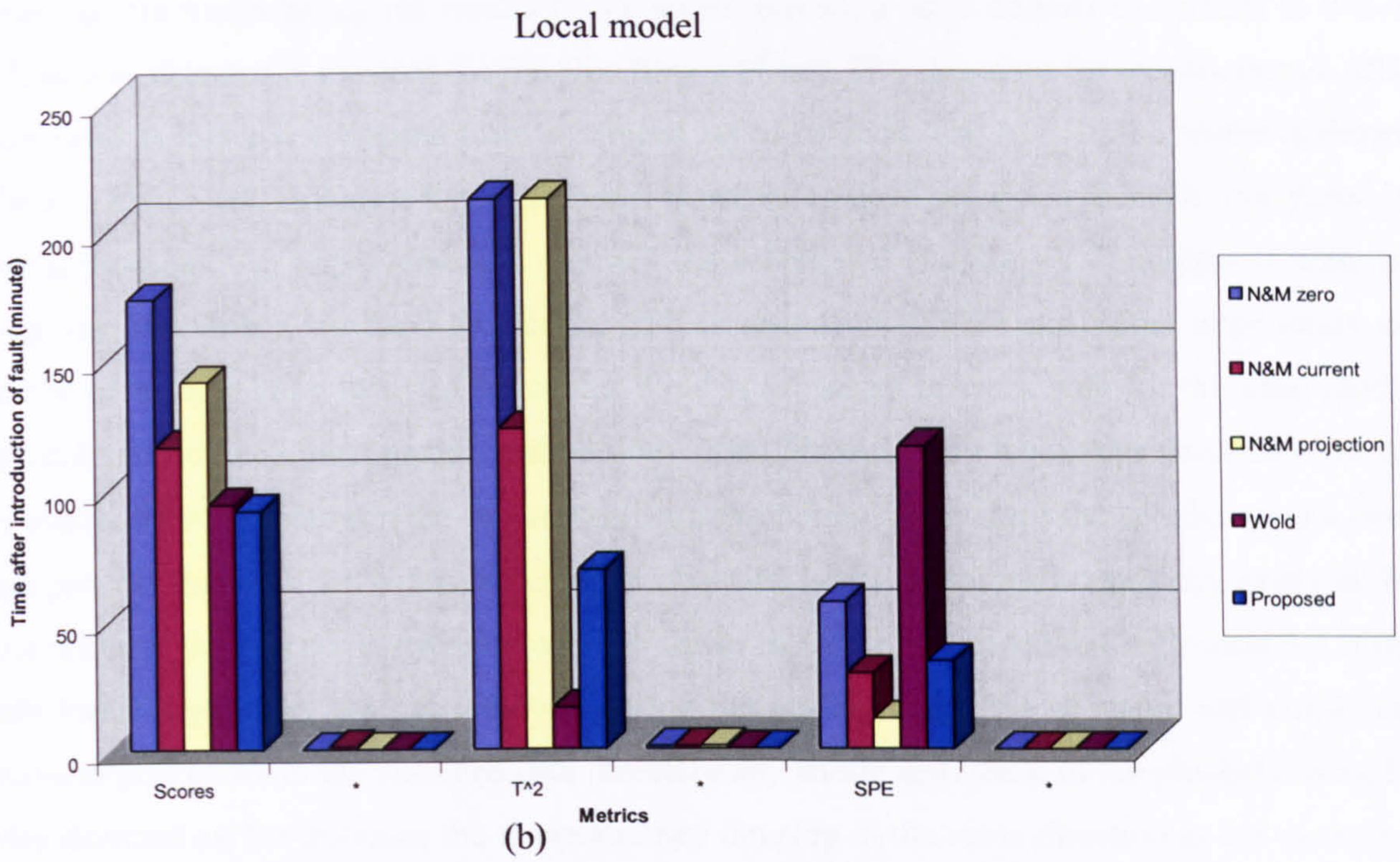
The third group of approaches is based on the proposed methodology. Again the false alarm rate for principal components one to three for both the global and local models is at reasonable level which is around 4% except for the third principal component of the local model that exhibits a 5% false alarm rate. The Hotelling's T^2 and SPE give similar results and overall the performance is fairly consistent across the metrics for both the global and local models. This methodology has given an overall good performance for the type I error for the simulated training data set.

The results of the false alarm rate have a direct impact on the result of the out-of-control average run length. The hypothesis is that if a low false alarm rate is observed, the model is less sensitive than expected therefore the ability to detect a fault is likely to be lower hence the out-of-control average run length can be longer.

To evaluate this hypothesis, the results for the out-of-control ARL for the global and local models with a 95% confidence limit are given in Figure 3-14. The results obtained were as follows. First, fifty faulty batches were generated with a downward drift of 0.1% from 80 hours of operation to the end of the process. Random noise was introduced as described in Table 3-1. These batches were then monitored using the principal component scores, Hotelling's T^2 and SPE confidence limits based on the nominal representation that were developed for the five approaches. For the principal component scores, the first three components are compared in the control chart and the principal component which gives the lowest out-of-control ARL was selected for determining the lowest ARL. From a practical prospective, it is the control chart that gives the lowest out-of-control ARL that is of main interest. In addition, the determination of the ARL should be in conjunction with the index of the number of batches where the fault was not detected. This is represented by an indicator * for each metric. The lower the index, the better and more robust is model performance.



(a)



(b)

*Figure 3-14 Out-of-control average run length for metrics of different approaches for (a) global model; (b) local model; * indicates the number of undetected batches*

For the approaches in group one, it is in general the SPE that gives the shortest time to detect the fault with the results being similar for the global and local models. No undetected batches were encountered. Of the three approaches in group one, the missing data approach performs best and gives the lowest SPE out-of-control ARL. Hotelling's T^2 for the group one approaches took longer to detect the fault and the time required after the introduction of the fault was in excess of 100 minutes. The local models for zero deviation and missing data in-filling are significantly longer than the global model approach and are in excess of 200 minutes whilst the results for current deviation in-filling are similar. This indicates that when the Nomikos and MacGregor approach is applied, it is not necessary to divide the process into several stages since the process dynamics was addressed in the scaling stage. The performance of the zero deviation in-filling method is limited at the beginning of each stage since not enough measurements are used to estimate the trajectory hence greater uncertainty is evident in terms of fault detection. This can impact on the sensitivity of detecting the deviation. The out-of-control ARL of the principal component scores perform the worst of all the metrics for the group one approaches. An excessively long detection time is observed although the principal component which exhibits the shortest detection time was selected. Nevertheless, the current deviation in-filling approach has the lowest ARL among the group which is in consistent with the results for Hotelling's T^2.

The Wold *et al.* approach revealed a significant difference between the global and local modelling. By interrogating the results of the global model, a large number of batches in which no fault was detected is reported for the Hotelling's T^2 and SPE therefore the out-of-control ARL is not valid as this was estimated from a reduced set of batches. The ARL of the scores is shown to be at a high level. However the local model is shown to have competitive results compared to other approaches. An improvement is seen for the scores and Hotelling's T^2 compared with the group one approaches but the ARL for the SPE is also at the higher end. More importantly no undetected batches were reported hence the Wold *et al.* seems to work well for the local model approach but not the global model methodology. This fits well with the results described before. The main difference between the two modelling approaches is that since the simulation is a two-stage process (batch to fed-batch switching at about 45 hour), a major discontinuity exists in the process hence the process is dynamic and non-linear in nature. As described in Section 3.3 if the batch mean trajectories are not removed during the scaling stage, the dynamic and non-linear behaviour still exists in the unfolded data therefore any subtle difference in the process cannot be easily detected as, for instance, the deviation may develop in the same direction as the dynamics of the process. The advantage of applying the local model is that the process dynamics and non-linearity were reduced by dividing the process into several phases.

Finally, the results for the proposed approach is investigated. By interrogating the global model, the out-of-control ARL is in general lower than for the other methods hence the detection of a fault is quicker. The result is more apparent for the scores and Hotelling's T^2 than the SPE in which the current deviation and missing data approaches of Nomikos and MacGregor performed better. However, the proposed approach is more consistent across the metrics and no undetected batches are reported. Similar performance is identified for the local model and the level of out-of-control ARL is close to the global model with a slight improvement for the scores. This implies that the proposed approach is robust for detecting a fault and shown to perform better than other approaches in terms of the detection of a subtle deviation. If a local approach is employed for the Wold *et al.* method, the result is comparable to the proposed approach.

3.5 Conclusions and Recommendations

The focus of this chapter was to perform an evaluation of five MSPC based approaches for the monitoring of the performance of a batch process, in particular to consider a new approach. The methodologies behind the different data pre-treatment methods and the philosophy of data unfolding were discussed and the impact illustrated through a simulation study. The issue of the

batch characteristics of non-linear behaviour and process dynamics were addressed. The most important distinction between the approaches is the way the correlation structure in the data is modelled and the scaling approach is applied. For the monitoring of a new batch, the in-filling of future observations is required for on-line monitoring using the Nomikos and MacGregor methodology and the batch lengths require to be equalised. Despite these two limitations, the loadings structure captures the systematic variation about the time trajectories for all variables for all batches. In contrast there is no underlying assumptions for the monitoring of a new batch using the Wold *et al.* approach and no batch equalisation is necessary. The concept of the proposed approach is to combine the advantages of the two existing methods and hence it is assumed that the covariance structure is fixed over the duration of a batch. One limitation is that the loadings describe the summation of both the batch and time trajectories for a variable.

Five approaches were demonstrated and compared using a simulated penicillin production. Since the process is multi-stage, both global and local models were developed and assessed. Subtle fault was considered to test the potential of the techniques. The performance of the different approaches was evaluated using two performance indices, the false alarm rate and the out-of-control average run length. Some recommendations on the application of the different methods are summarised below:

- For the purpose of batch performance monitoring, the Nomikos and MacGregor approach of auto-scaling the unfolded data is most appropriate as the between batch behaviour is captured and the non-linear and dynamic components in the data are reduced or eliminated. Therefore this procedure of data scaling is recommended.
- The on-line N&M approaches require in-filling of future observations although a number of methods have been proposed, at the start of a batch, the inferences may be inaccurate hence the monitoring model is not realistic. Hence monitoring a batch process on-line using the Wold *et al.* philosophy (batch stacking) gives better performance and requires both less computation and makes no assumption concerning values of subsequent observations.
- The issue of detecting a fault during the start-up of a process has to be addressed. Of the N&M approaches, the zero deviation and missing data approaches were shown to be less sensitive than the current deviation approach for both the global and local models. This problem does not occur for the Wold *et al.* approach as no estimation of subsequent observation is required to perform on-line prediction.
- In terms of fault detection capability, both the current deviation approach and the proposed approach exhibited good performance with the later approach being shown to be more robust and reliable. In addition, the proposed approach also demonstrated the advantages

observed from the Wold *et al.* approach. Overall, the evaluation showed that the proposed approach has merits for batch process monitoring and fault detection. The application of the proposed approach to an industrial data set is investigated in Chapter 6.

The development of multi-site monitoring approaches on industrial process data will be investigated in the next chapter.

Chapter 4

Application of Industrial Process Data Analysis

4.1 Introduction

In the pharmaceutical industry, the manufacture of drugs is as important as generating new products from research and development. Manufacturing challenges facing the industry include the need to reduce the time between product development and full-scale production, the achievement of right-first-time manufacture and the manufacture of consistently high quality product with minimal environmental impact. A contribution to addressing these challenges is to utilise the data collected from the process and to convert it into information and ultimately knowledge, thereby enabling an enhanced understanding of the process to be achieved. These drivers have resulted in process data analysis and performance monitoring becoming an integral part of process operation (Miletic *et al.*, 2004).

With the drive towards globalisation and the need to transfer product manufacture to different sites around the world, relying on single source production can be a potential threat with respect to the survival of a company. In practice when the product is manufactured at two or more sites, independent monitoring systems may have been developed. Adding a further level of complexity is that, to date for many industrial processes, performance monitoring systems are developed for individual process units, as opposed to the complete process. A major challenge is now to identify the sources of the differences in process operation and product variation, between the sites, taking into account multiple units.

In this Chapter, multivariate statistical projection techniques are applied to an industrial application where consideration is given to the manufacture of an Active Pharmaceutical Ingredient (API) at multiple manufacturing sites. Historical process data of the API is available for the two manufacturing sites with the goal being the investigation of the subtle differences in product quality. Traditional approaches to multi-site production have been to rely on an individual site's science and engineering expertise and phenomenological understanding. The existing mechanistic based approaches may alone not be viable due to a lack of understanding of the process and the significant time taken for model development. Therefore an empirical data based approach may enable more rapid and effective process understanding and model development with respect to scale-up, multi-site manufacture and product transfer. Within this chapter a number of approaches are investigated for the development of multi-site monitoring.

4.2 Process Description

The process forming the basis of this study is a single stage within a multi-stage liquid phase synthetic route for the production of an active pharmaceutical ingredient that is carried out at two manufacturing sites. The chemistry step involves an exothermic addition that is controlled by the reactant addition rate and the reactor temperature and is of approximately one hour duration. This is a critical stage since the quality of the final product is highly dependent on this step. Post addition, the mixture is stirred for one hour to ensure the reaction is complete and it is then cooled and quenched by the addition of an aqueous solution. Distillation is subsequently applied resulting in an oil which is then re-dissolved. The slurry is finally stirred, filtered and dried to obtain the final product for this stage of the manufacturing process.

Although different plant configurations and procedures are employed at the two sites, a number of similar process variables are monitored. The process data at both sites is acquired from the reactor probes linked to the data historians and was extracted for the subsequent analysis. The process variables include reactor temperature, reactor pressure, reactant addition, level, agitation rate and vapour temperature. Process data from 57 batches from Site A and 144 batches from Site B was available for the analysis. The batch information is summarised in Table 4-1.

Site A	Site B
57 batches	144 batches
5 process variables	4 process variables
1. Reactor temperature (°C) 2. Reactor pressure (bar) 3. Reactant addition (kg) 4. Level (litres) 5. Agitation rate (rpm)	1. Reactor temperature (°C) 2. Reactor pressure (bar) 3. Reactant addition (kg) 6. Vapour temperature (°C)

Table 4-1 Batch information for the two manufacturing sites

4.3 Data Pre-processing

The first step in the development of a nominal model is the acquisition of representative data of the process. The raw data collected from the data historian may contain data that is not relevant to the critical step of the process. Time series plots are used for the pre-screening of the raw data. Figure 4-1 illustrates the time series plots for all variables for all batches for site A with the same information being shown for site B in Figure 4-2.

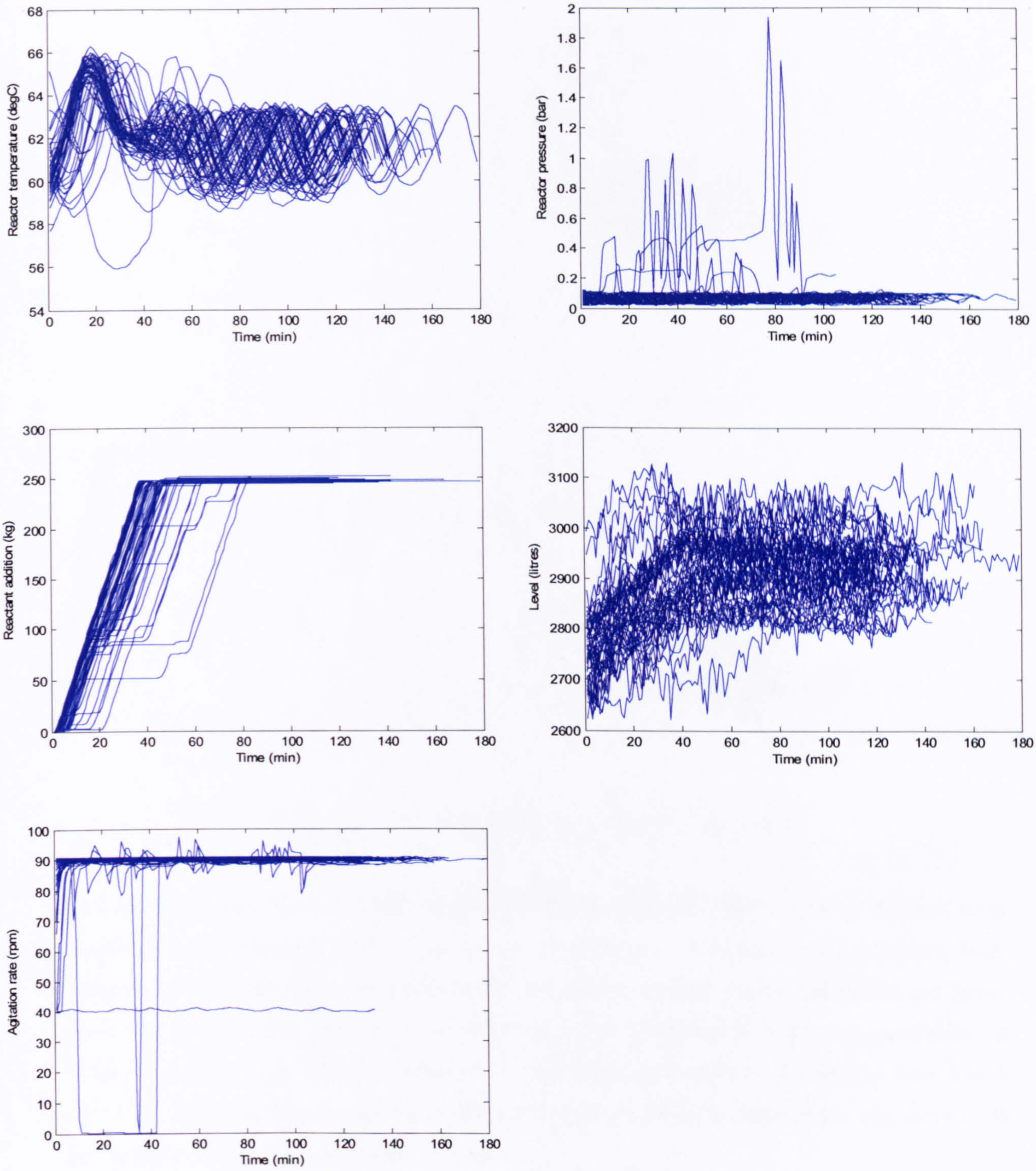


Figure 4-1 Time series plots of all variables for site A

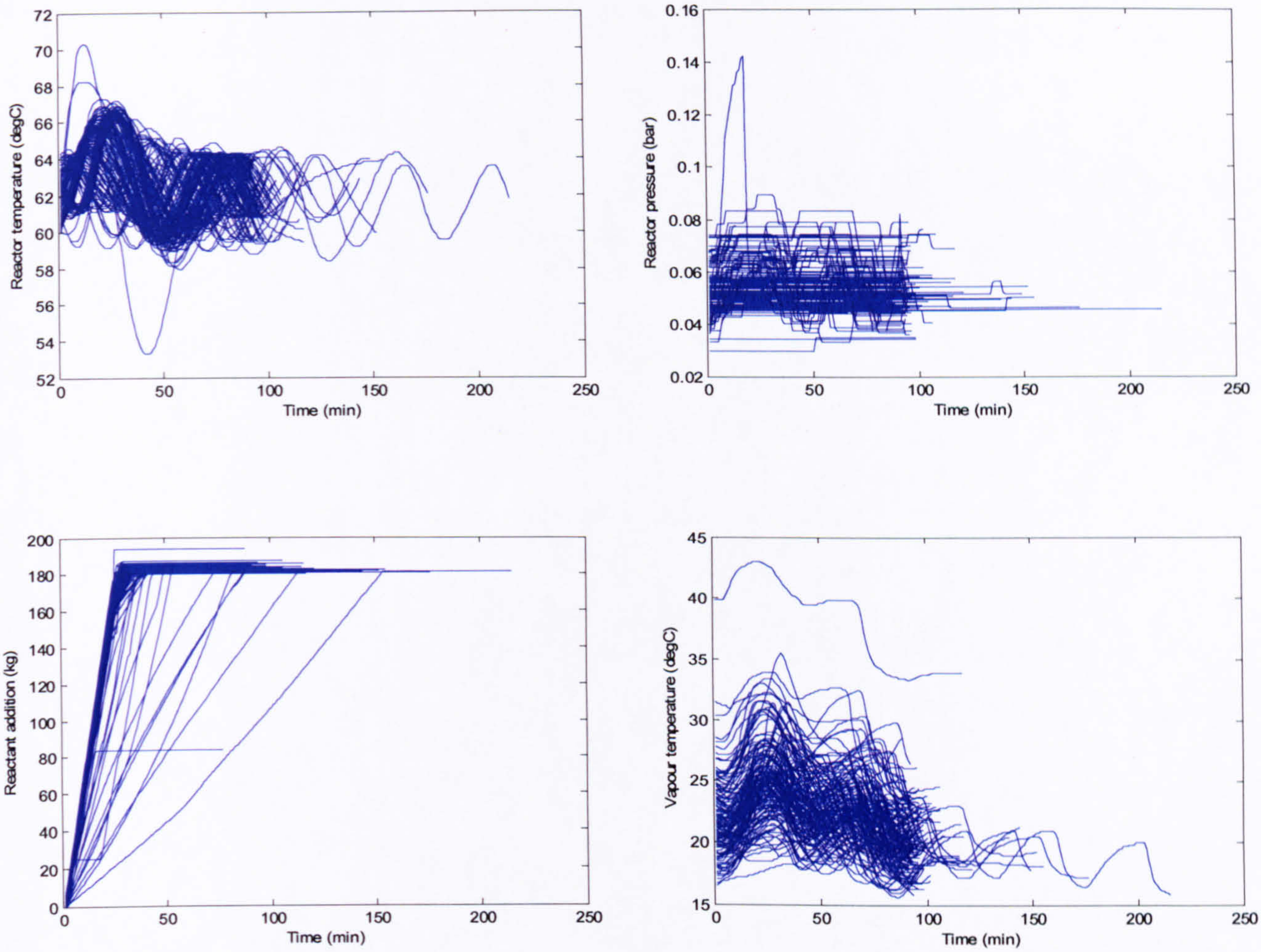


Figure 4-2 Time series plots of all variables for site B

From the time series plots, a number of outliers can be observed. These may be a result of some batches not-being operated at the typical process conditions or else due to sensor failures. Such univariate outliers require to be removed for the data to exhibit similar behaviour and hence enable the true process deviations within sites to be identified through the application of multivariate techniques. As a consequence of this preliminary analysis, 11 batches from site A and 19 batches from site B were removed resulting in a total of 46 batches for site A and 133 batches for site B for the subsequent analysis.

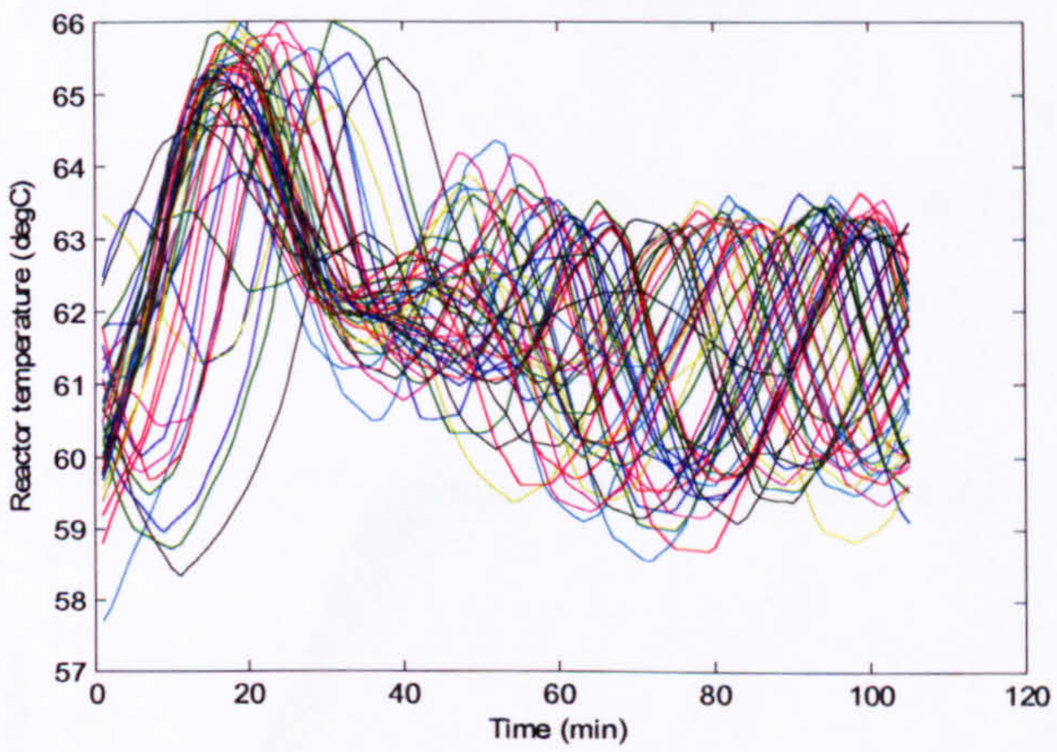
One major issue with industrial data, relating to batch processes, is that the batches are rarely of equal duration. Methodologies for the equalisation of batch lengths were discussed in Section 3.2.3 and in this application the technique of cutting to a pre-defined batch length is applied. The rational for this was that since the focus of the analysis is the exothermic addition, there is no loss of information when this form of batch equalisation is applied. Batch lengths of 105 minutes for site A and 85 minutes for site B were selected. The difference of equalised batch lengths between

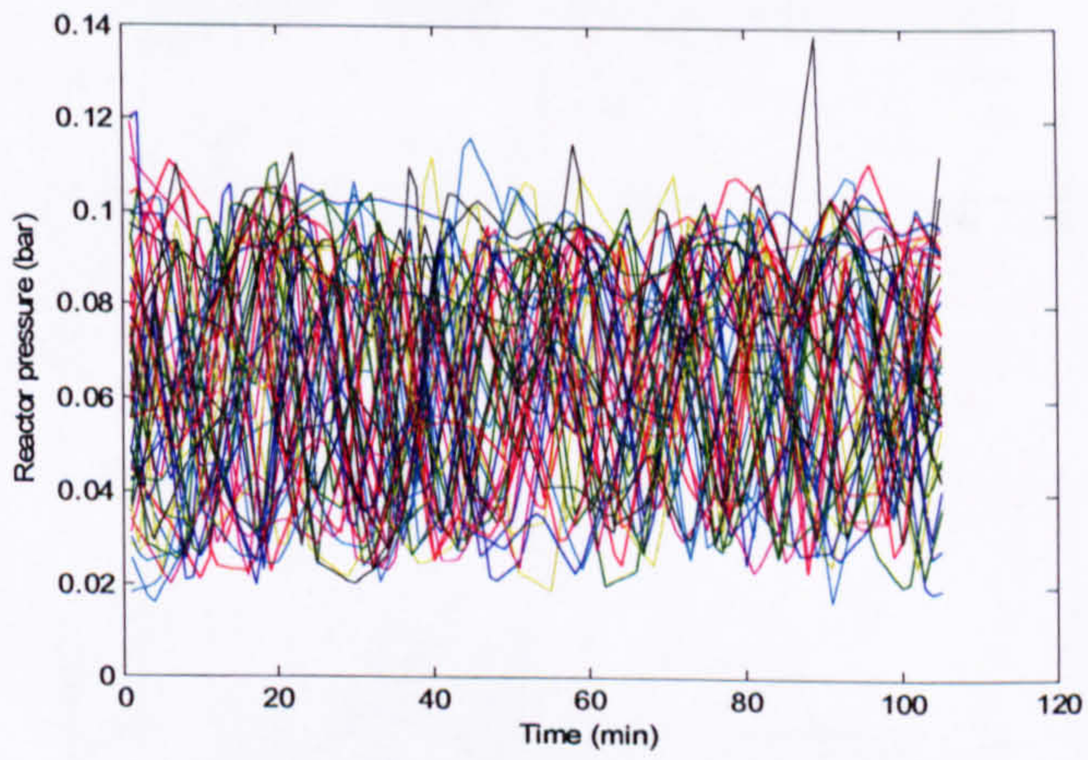
sites is due to the alignment to the minimum batch length for individual site. A consequence of the application of data pre-processing is that a set of representative data for each site is attained.

From the engineering prospective, it is important to approach a problem using engineering principles hence the process data may require a transformation from the raw data. An assumption is made for this application that the transfer of recipe between sites is specified for the feed on a per unit volume basis. This can be calculated by dividing the raw data of reactant addition by level. However, level measurement is not readily available for site B. It is therefore required to postulate level in order to calculate the unit per volume basis. The level time series profile for site B is calculated by

$$B:level\ profile = \frac{B:Reactant\ addition\ /\ time\ point}{A:Reactant\ addition\ /\ time\ point} \cdot A:mean\ level\ profile \qquad 4\text{-}1$$

The resulting data for sites A and B are plotted in Figure 4-3 and Figure 4-4 respectively. The postulated level for site B is plotted in Figure 4-5 as reference however it is not included in the subsequent analysis.





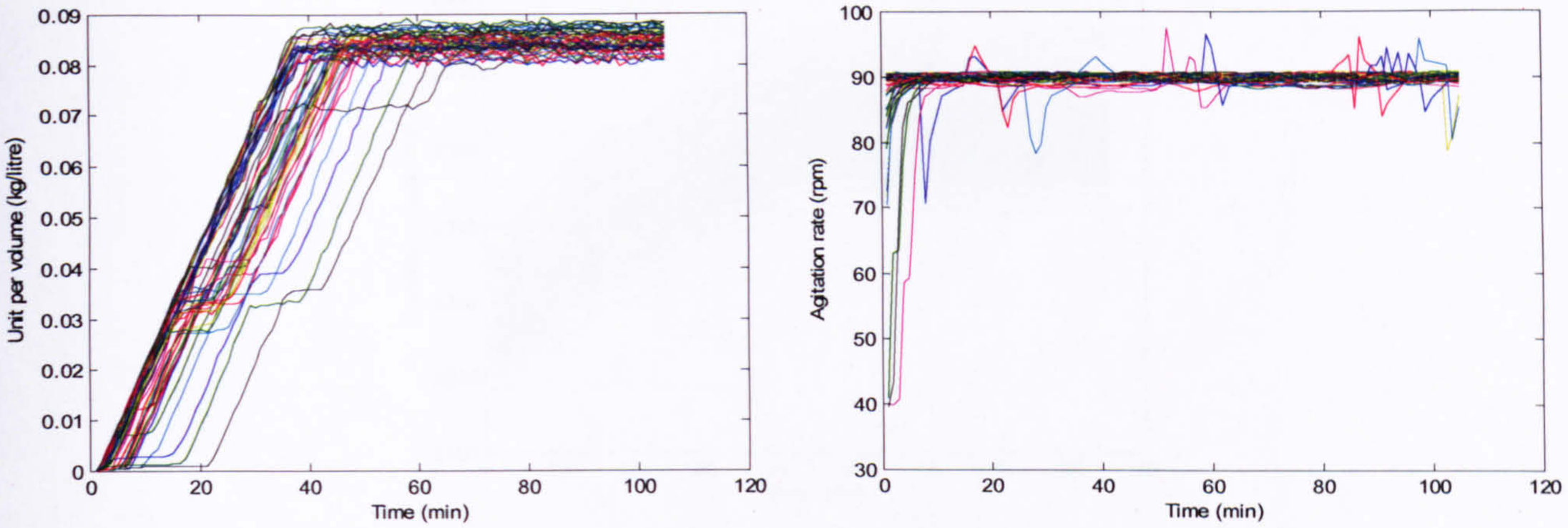


Figure 4-3 Time series plots of pre-processed batches for site A

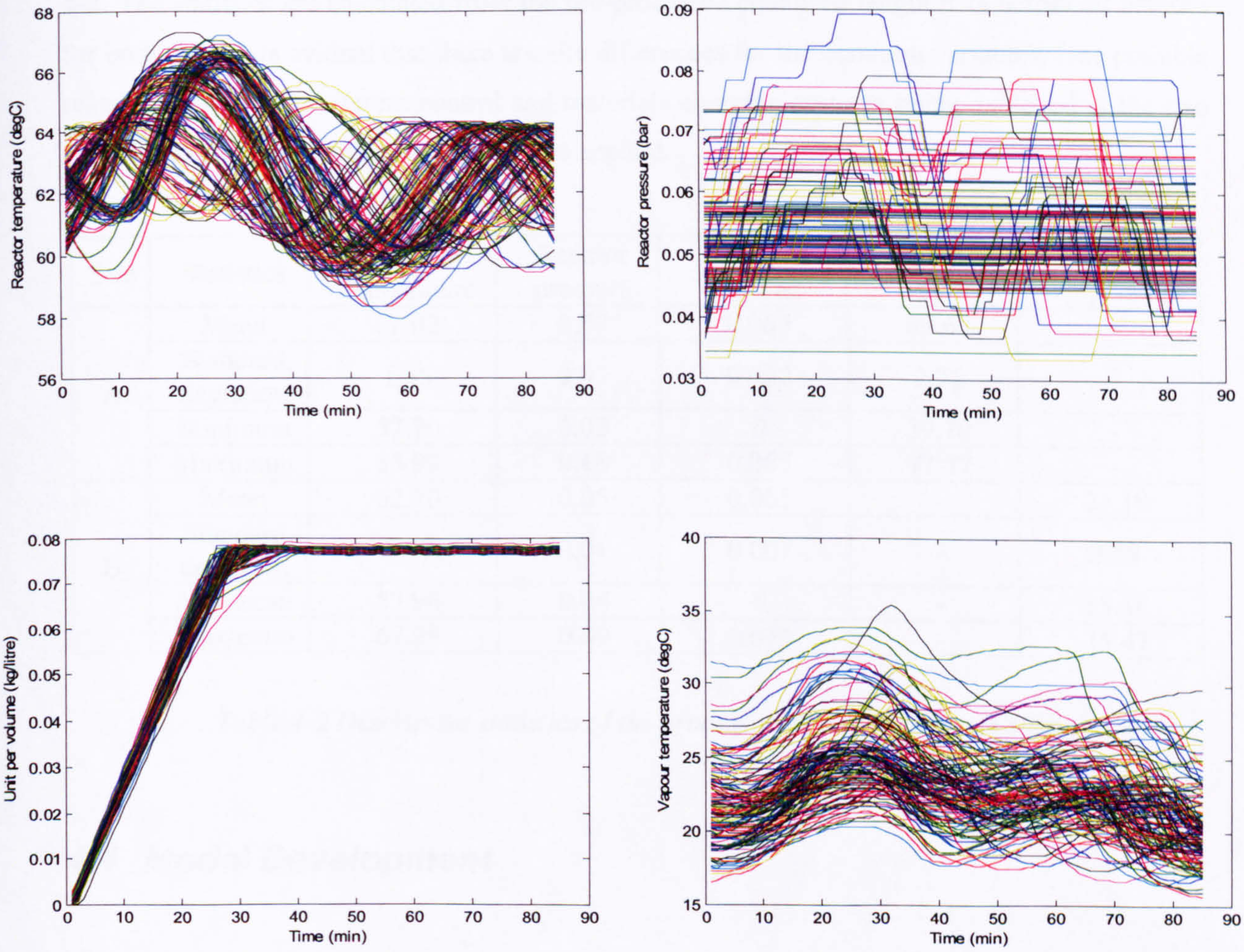


Figure 4-4 Time series plots of pre-processed batches for site B

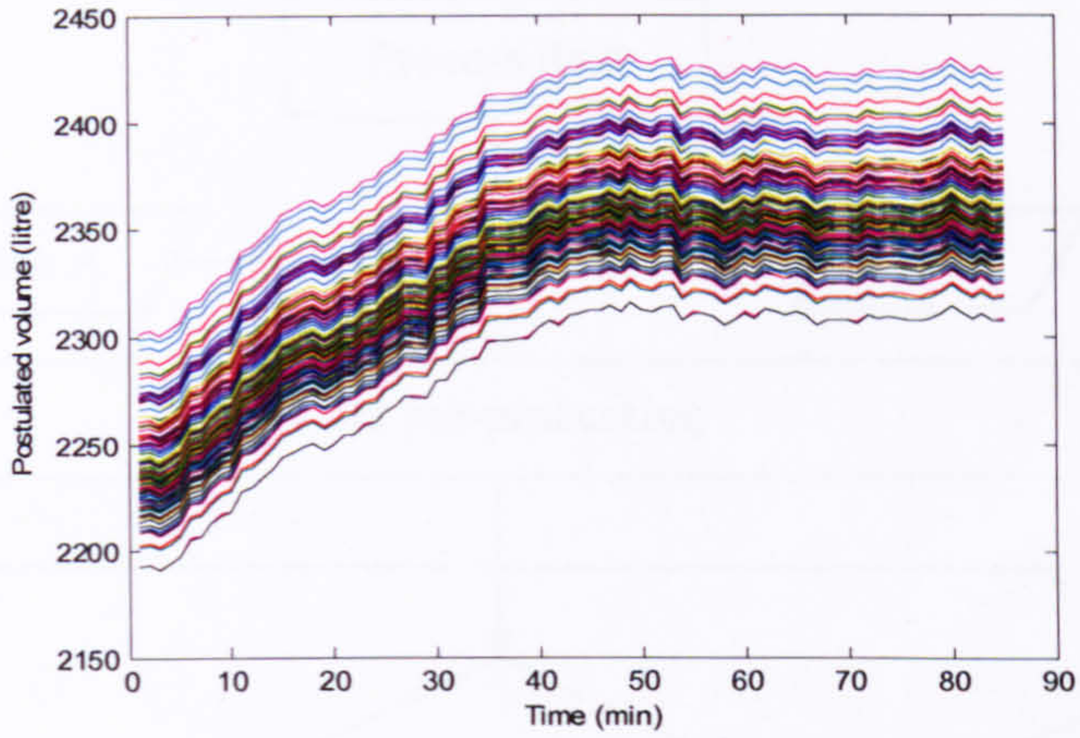


Figure 4-5 Time series plot of postulated level for site B

The descriptive statistics for the process variables for the two sites can also be informative, Table 4-2. The statistics are calculated from the pre-processed equalised length data across all batches for both sites. It is evident that there are site differences for the common variables. One possible reason is due to the different control and materials charging systems being deployed at the two sites therefore a different control strategy is applied.

Site	Statistics	Reactor temperature	Reactor pressure	Unit per volume	Agitation rate	Vapour temperature
A	Mean	62.02	0.07	0.065	89.67	-
	Standard deviation	1.55	0.03	0.027	2.21	-
	Minimum	57.70	0.02	0	39.70	-
	Maximum	65.99	0.48	0.085	97.47	-
B	Mean	62.70	0.05	0.065	-	23.19
	Standard deviation	1.87	0.01	0.007	-	3.19
	Minimum	57.96	0.04	0	-	15.58
	Maximum	67.28	0.09	0.079	-	35.47

Table 4-2 Descriptive statistics of the process variables for the two sites

4.4 Model Development

The focus of the analysis is the development of integrated multi-site model and the development of individual site models forms the benchmark of the different approaches being developed. Having pre-screened and equalised the duration of the batch data, the next step is to build a nominal model. The three-way batch data is unfolded and modelled using the Nomikos and

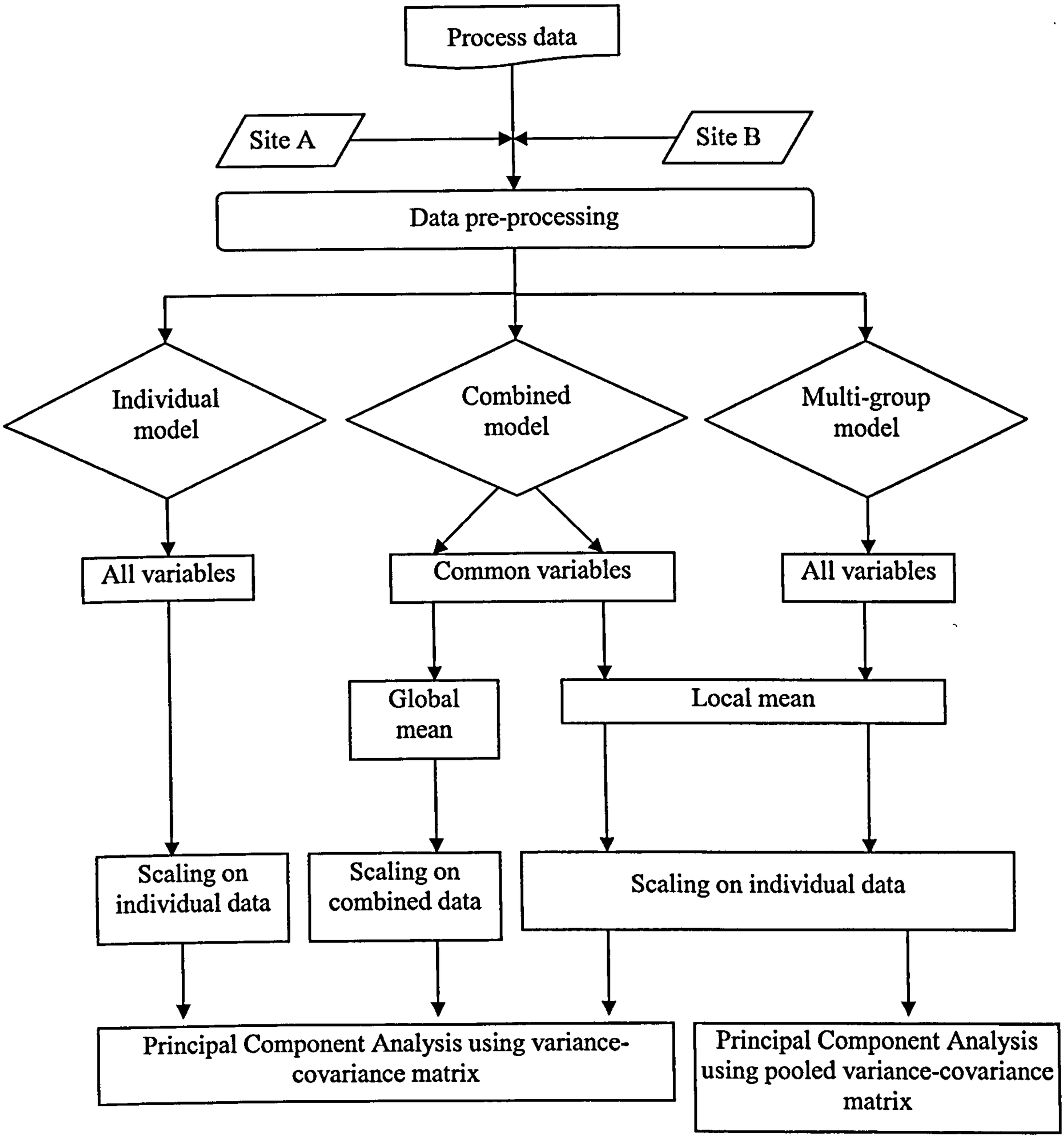


Figure 4-6 Overview of different monitoring approaches

MacGregor approach described in Section 2.6.2. This approach is suitable for the analysis of historical batch data with the time trajectories for each variable being described by the principal component loadings (Kourti, 2005). The overall batch information is summarised in the principal component scores. A number of different multiway PCA approaches that combine the two data sets are developed and compared against the two individual site models. The data matrices comprising the common variables from the two sites were adopting different scaling procedures applied. The first approach resulted in the removal of the global mean and standard deviation of

each variable (calculated from the data for the two sites) with the second approach requiring the local mean and standard deviation for each individual variable for each site to be removed. Finally a multi-group model based on the pooled sample variance-covariance matrix was developed using all the variables monitored at both sites. Figure 4-6 provides an overview of the different approaches.

4.4.1 Individual MPCA Model

46 batches of data from site A and 133 batches of data from site B are included in the nominal model development. The models are constructed for each individual site. Table 4-3 summarises the amount of variance captured by each of the principal components for both sites. Six principal components for site A and seven principal components for site B are selected by cross-validation (Wold, 1978) to be retained in the subsequent analysis. 58% of the underlying variability for site A is explained whilst 91% of the underlying variability for site B is explained. The difference in the number of principal components retained in the model and the variability explained is determined by the covariance structure of the data set. The application of cross-validation identifies the optimal number of principal components in the model that captures the key sources of variation. It should also be noted that the number of principal components in the model is greater than the number of variables for each site. This is because the total number of variables is the product of variables and time points. For example there are 4 variables and 105 time points for site A therefore the total number of "variables" is 420. Figure 4-7 illustrates the scree plots for the two sites.

Principal component	Site A (5 variables)		Site B (4 variables)	
	Individual % variance captured	Cumulative % variance captured	Individual % variance captured	Cumulative % variance captured
1	15.73	15.73	35.75	35.75
2	12.50	28.23	17.24	52.98
3	9.37	37.59	13.54	66.52
4	8.95	46.54	11.62	78.14
5	6.89	53.43	7.29	85.43
6	4.14	57.57	3.84	89.28
7	3.77	61.34	1.58	90.86
8	3.42	64.77	1.46	92.32

Table 4-3 Percentage of variance explained by the individual principal components

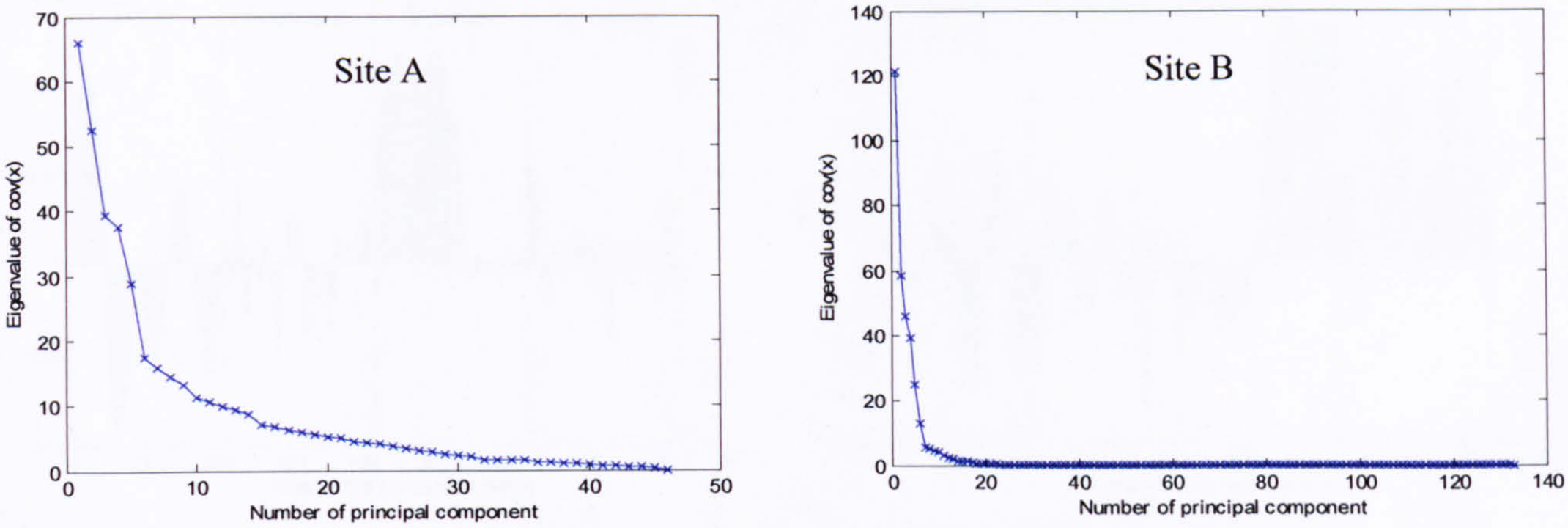


Figure 4-7 Scree plots for site A and site B

The univariate loadings plots of principal component one and two for site A and site B are shown in Figure 4-8 and Figure 4-9 respectively. From the loadings plots of multiway PCA, process behaviour over time for the different variables can be examined. The dotted lines are used to differentiate between different variables. It is interesting to observe from the loadings how the influence of variables change over batch duration. As observed from Figure 4-8 for site A for the first principal component, unit per volume (variable three) has a large influence throughout the duration of the batch exhibiting an increase at the start of the batch before falling off whilst for the second principal component, it exhibits a decrease then a sharp increase at about the half way of the batch duration. By interrogating the loadings plot for site B (Figure 4-9), reactor pressure (variable two) and vapour temperature (variable four) have a large positive influence across the batch duration whilst the second principal component reveals a large influence from unit per volume. The loadings can also be represented by bivariate scatter plot and they are shown in Figure 4-10 and Figure 4-11 for site A and site B respectively.

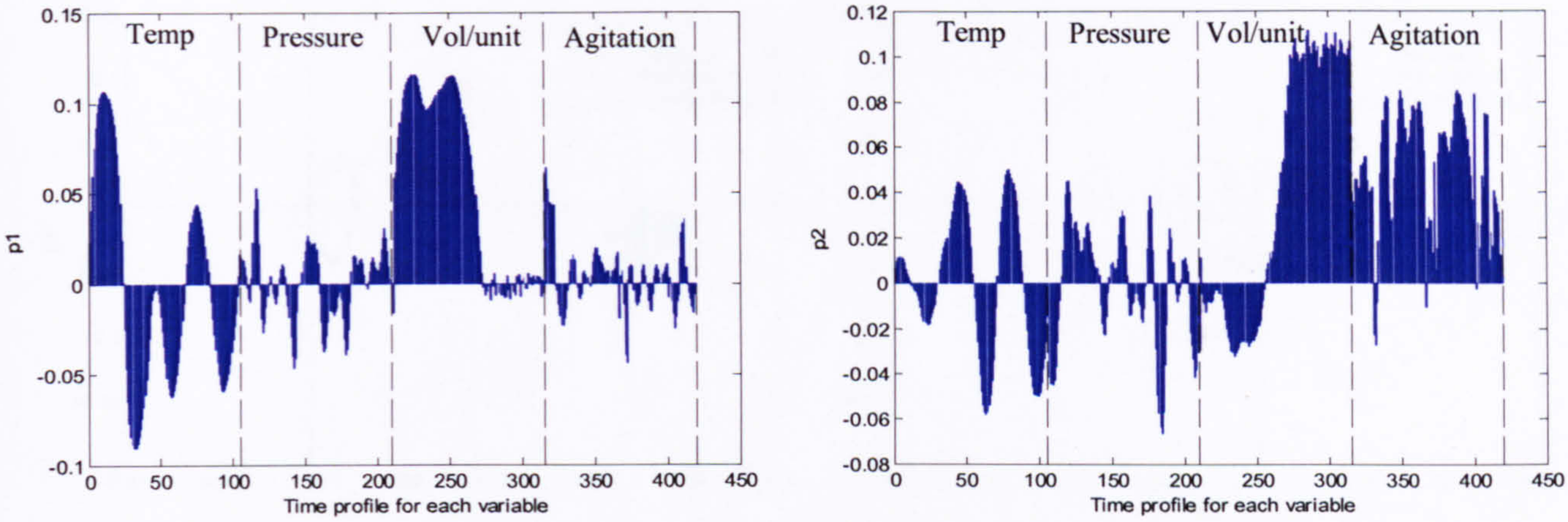


Figure 4-8 Univariate loadings plots of principal components one and two for site A

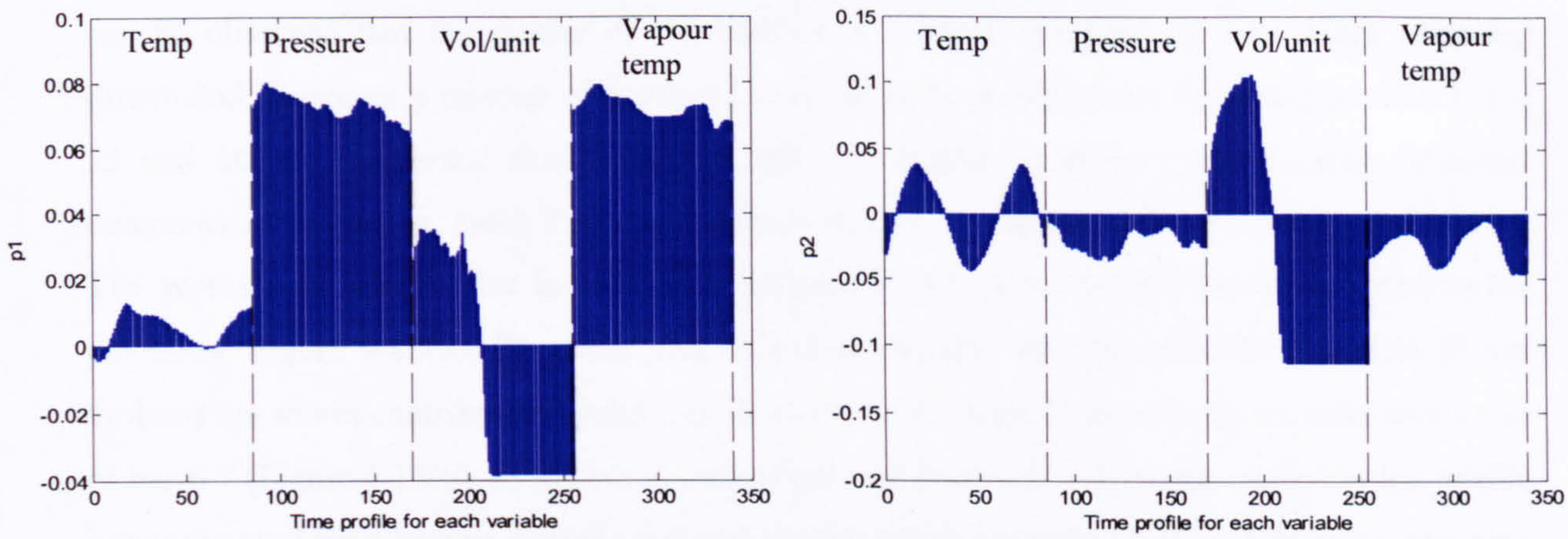


Figure 4-9 Univariate loadings plots of principal components one and two for site B

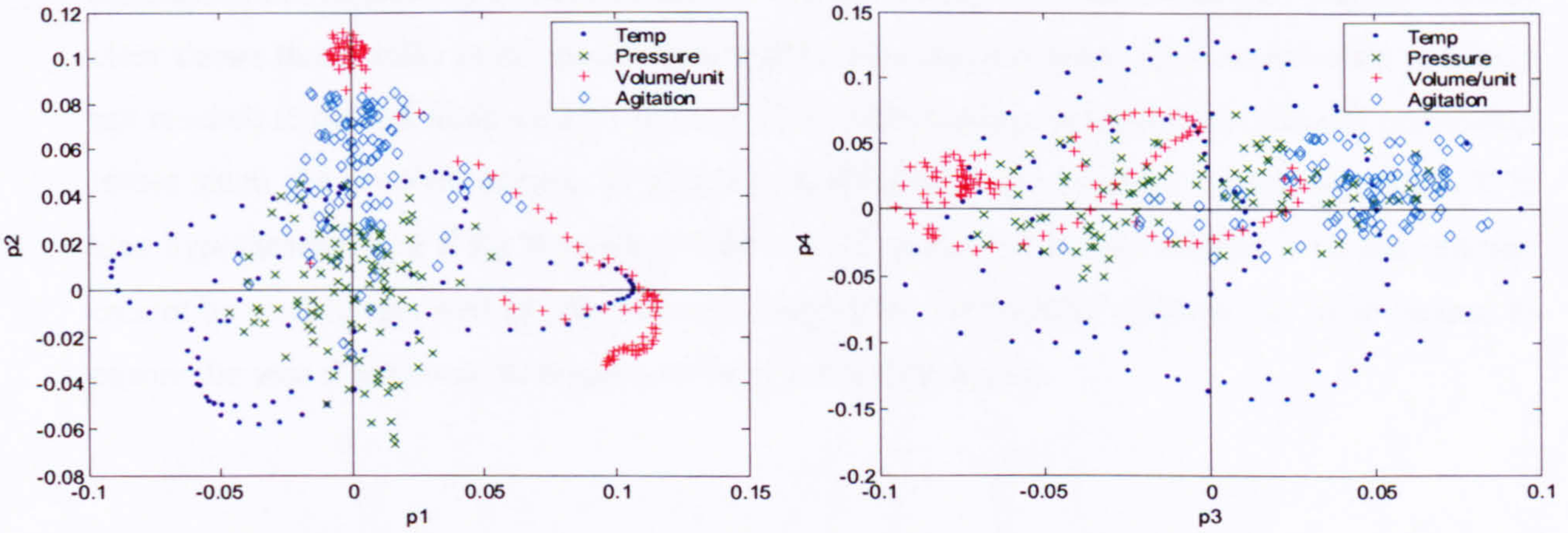


Figure 4-10 Bivariate loadings plots for first four principal components for site A

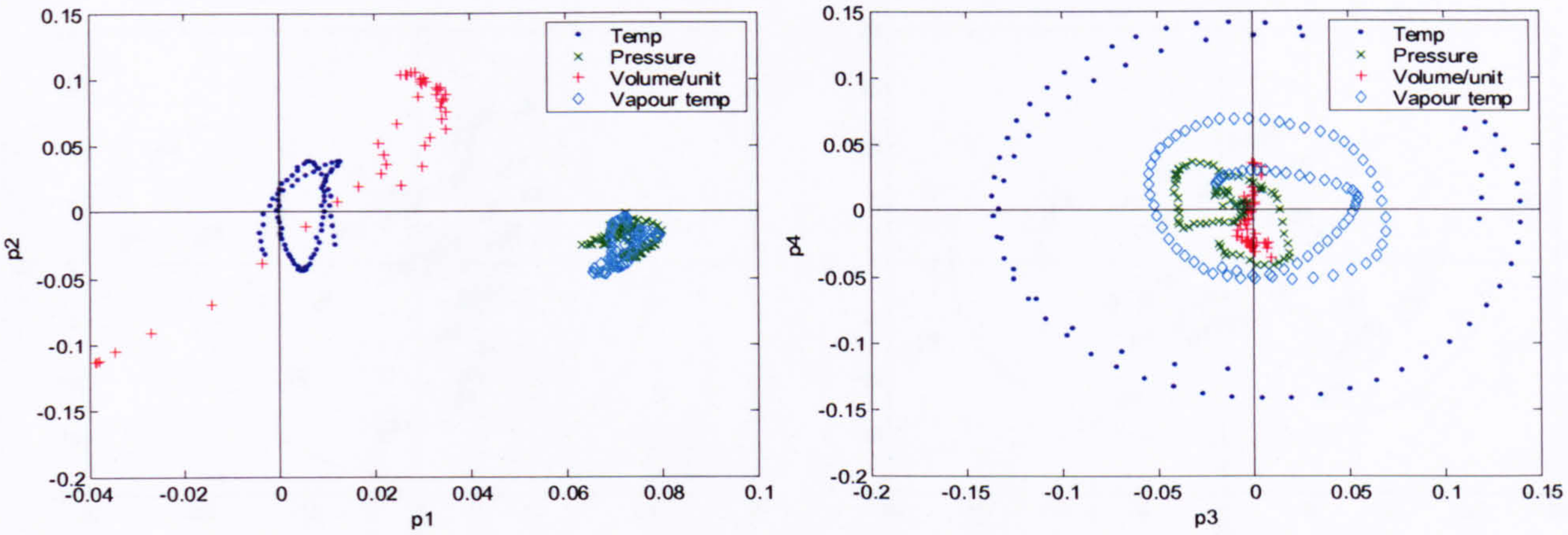


Figure 4-11 Bivariate loadings plots for first four principal components for site B

From the bivariate scores plots of the first six principal components for site A (Figure 4-12), it can be observed that the scatter of the batches is random inferring the scores are normally distributed. However a number of batches lie out with the action limits including batches 7, 19, 35 and 44. It is expected that a number will lie outside by chance. Investigating principal component five and six, batch 7 lies well outside the action limits and this is investigated further. The scores contribution plot is used to investigate which of variable(s) has contributed to the deviation, Figure 4-13(a). From the plot it is observed that the concentration (variable 3) has violated the scores contribution limits. This is further confirmed by examining the time series plot of batch 7 (Figure 4-13(b). From this it is observed that there was a different concentration profile hence the total reaction completion time was longer than for average batches. It is also interesting to investigate batch 35 since for principal component scores one to four, it is identified to lie out-of-statistical control limits. The scores contribution plot (Figure 4-14(a) identifies that the concentration (variable 3) has contributed to the deviation. The time series plot (Figure 4-14(b) clear shows that a delay in the concentration at process start-up hence the concentration profile is not reached to default limit until 63 minutes. The concentration is a good indicator of processing issues since if a problem occurs, an action to hold the reactant addition is normally taken. It is also hypothesised that if the reactant addition is too quick, the heat generated from the reaction cannot be readily removed by the jacket service hence the reactant addition has to be paused to ensure the vessel temperature remains in the specification region.

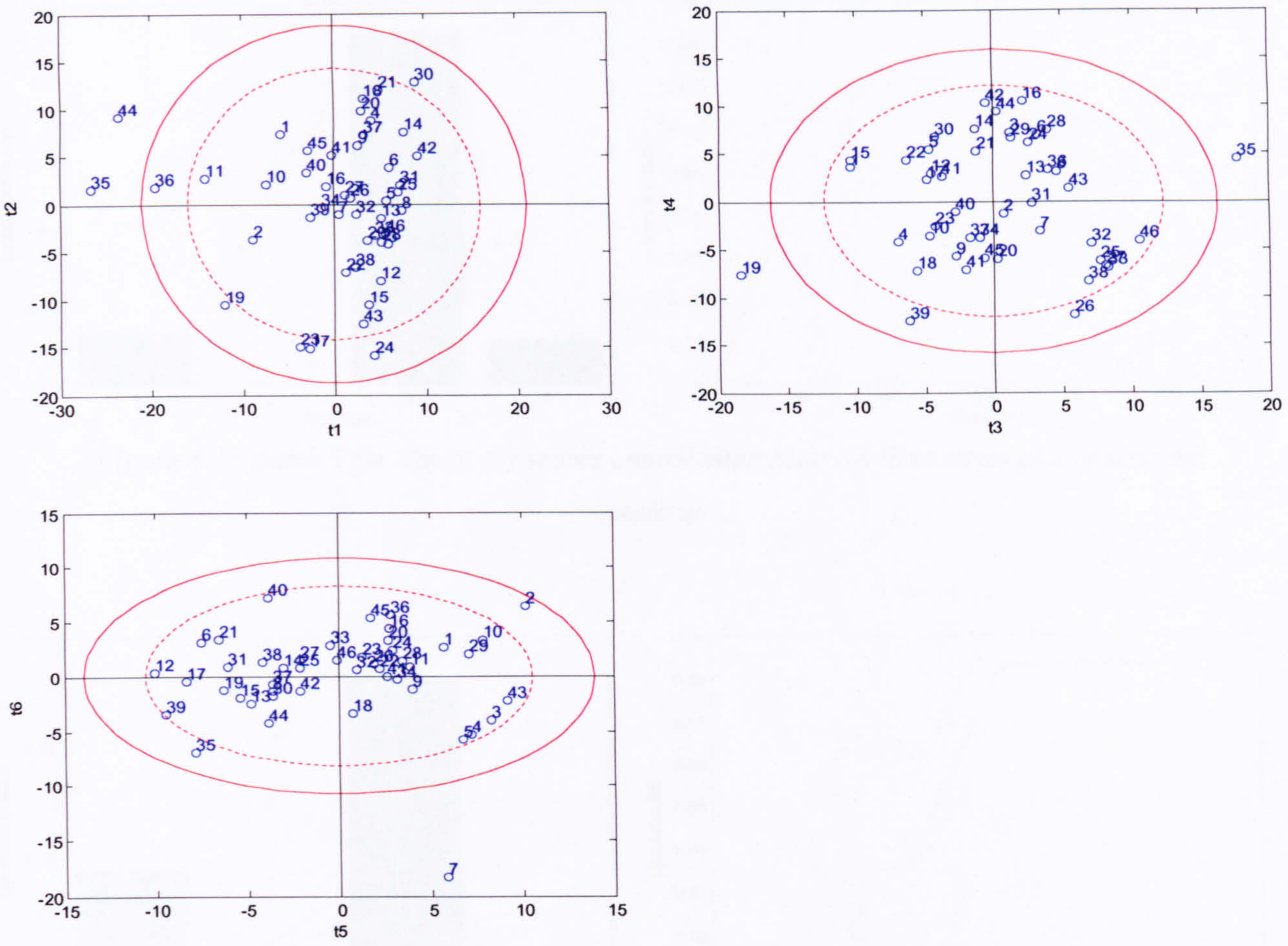


Figure 4-12 Bivariate scores plots for six principal components for site A

The scores plots of principal component one to four for site B are shown in Figure 4-15. The scores are also randomly distributed with a few batches lying out with the limits including batches 22, 30, 33, 49, 51 and 119. The batches were interrogated with scores contribution plots to identify these variables that contributed to the deviation and to identify if any remedial action require to be taken to prevent similar processing problems to be occurred.

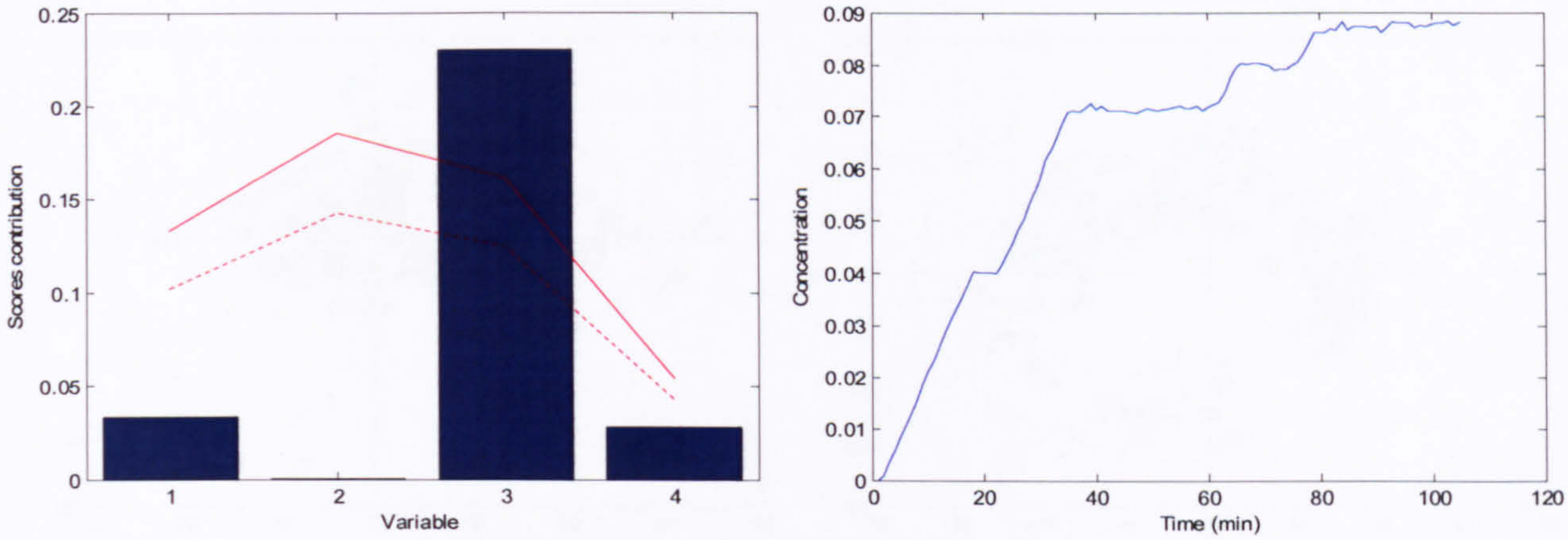


Figure 4-13 Batch 7 for site A: (a) scores contribution plot; (b) time series plot of reactant addition

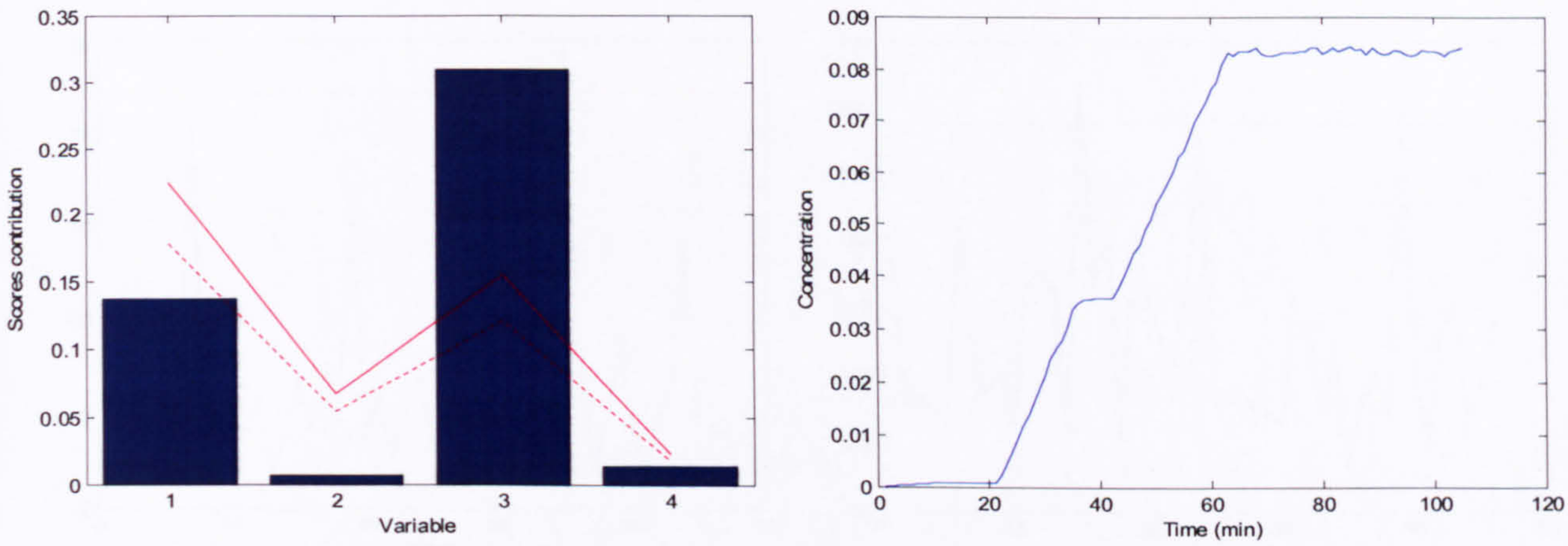


Figure 4-14 Batch 35 for site A: (a) scores contribution plot; (b) time series plot of reactant addition

The individual site models have been represented in terms of principal component scores plot. Alternative metrics that can be used to are Hotelling's T^2 and the Squared Prediction Error as shown in Figure 4-16 and Figure 4-17 respectively for site A and B. Figure 4-16 shows that batch 35 lies outside the Hotelling's T^2 limits for site A as was observed from principal components one and two. These charts form the benchmark against which the combined approaches developed in the following sections are compared.

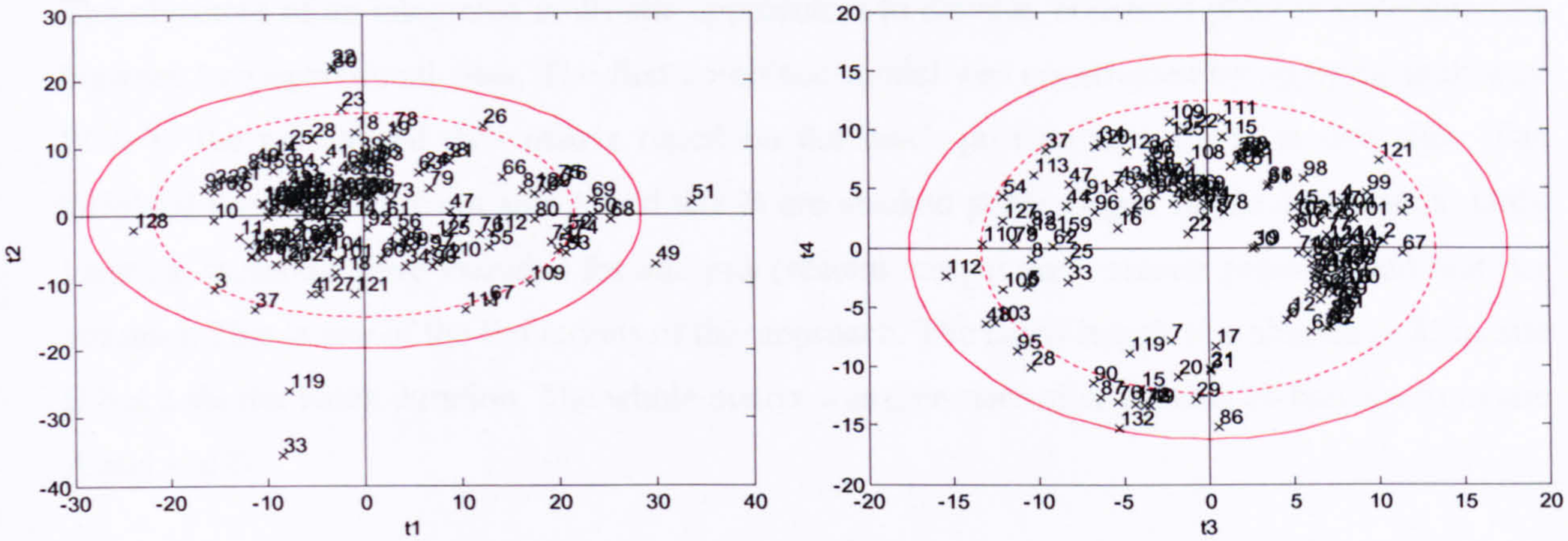


Figure 4-15 Bivariate scores plots of principal components one to four for site B

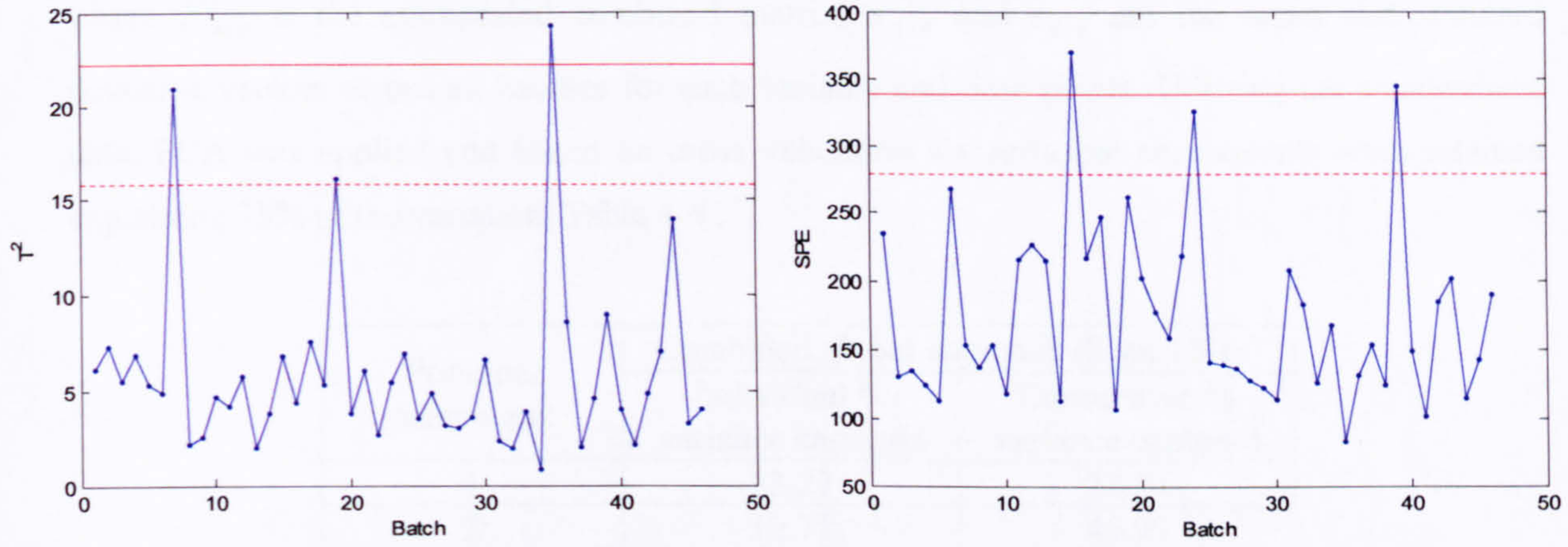


Figure 4-16 Hotelling's T^2 and SPE monitoring charts for site A

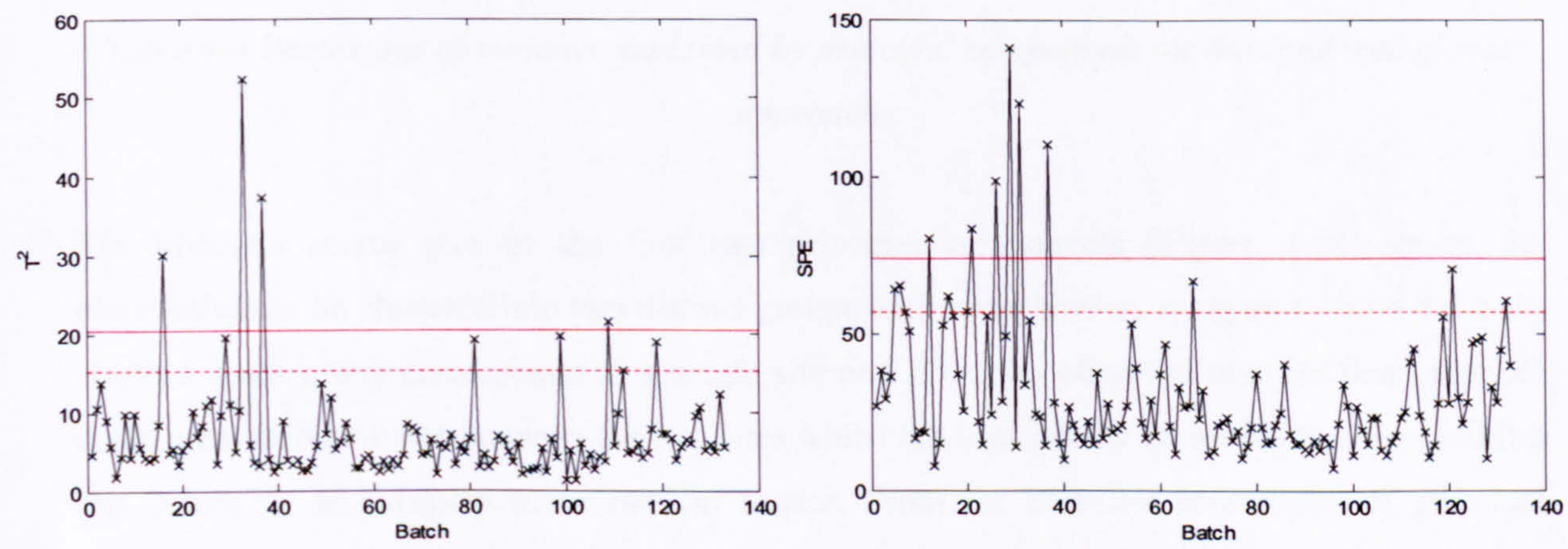


Figure 4-17 Hotelling's T^2 and SPE monitoring charts for site B

4.4.2 Combined MPCA Model with Global Based Approach

The objective of an integrated multi-site approach is to develop enhanced process understanding between two operational sites. The first combined model was constructed by applying multiway PCA to the normalised data matrix based on the batch process data from the two sites. The unfolded process data from site A and site B are stacked generating a single data matrix. Only identical variables were included for analysis (reactor temperature, reactor pressure and unit per volume). This is one of the limitations of this approach. The batch length was aligned to 85 as site B has a shorter batch duration. The whole matrix was then normalised across all batches from site A and site B:

$$\mathbf{X}^{+}_{A+B} = \frac{\mathbf{X}_{A+B} - \bar{\mathbf{x}}_{A+B}}{\mathbf{s}_{A+B}} \qquad 4\text{-}2$$

where $\mathbf{X}^{+}_{A+B}$ is the auto-scaled combined matrix, $\bar{\mathbf{x}}_{A+B}$ and $\mathbf{s}_{A+B}$ are the mean and standard deviation vectors across all batches for each variable and time points. Utilising the standardised data, PCA was applied and based on cross validation six principal components were retained explaining 75% of the variation, Table 4-4 .

Principal component	Combined global approach (3 variables)	
	Individual % variance captured	Cumulative % variance captured
1	33.27	33.27
2	12.71	45.97
3	10.25	56.22
4	8.13	64.35
5	7.49	71.84
6	3.08	74.92

Table 4-4 Percentage of variance explained by principal components for the combined global approach

The bivariate scores plot of the first two principal components (Figure 4-18) shows the observations to be clustered into two distinct groups with some batches lying away from the main clusters. Each group corresponds to a single site and it can be observed that the first principal component differentiates between the two sites whilst the lower order components do not exhibit this behaviour and display more random scatter. From the bivariate scores plot of principal component one and two, it can be seen that there are differences between the two sites and both "within" and "between" group variation is captured resulting in a process representation that does

not satisfy the fundamental assumption of normal distribution hence potentially impacting on the fault detection and diagnostic capabilities of the approach. It is also observed that some outlying batches from site A and site B differ to those identified in the individual site analysis, demonstrating the potential limitation of this approach as it provides conflicting information to the previous analysis. Batch 7 from site A was again revealed to be different from the nominal group and this was not detected by this approach. It is conjectured that this may due to the influence of certain variables being excluded from the analysis.

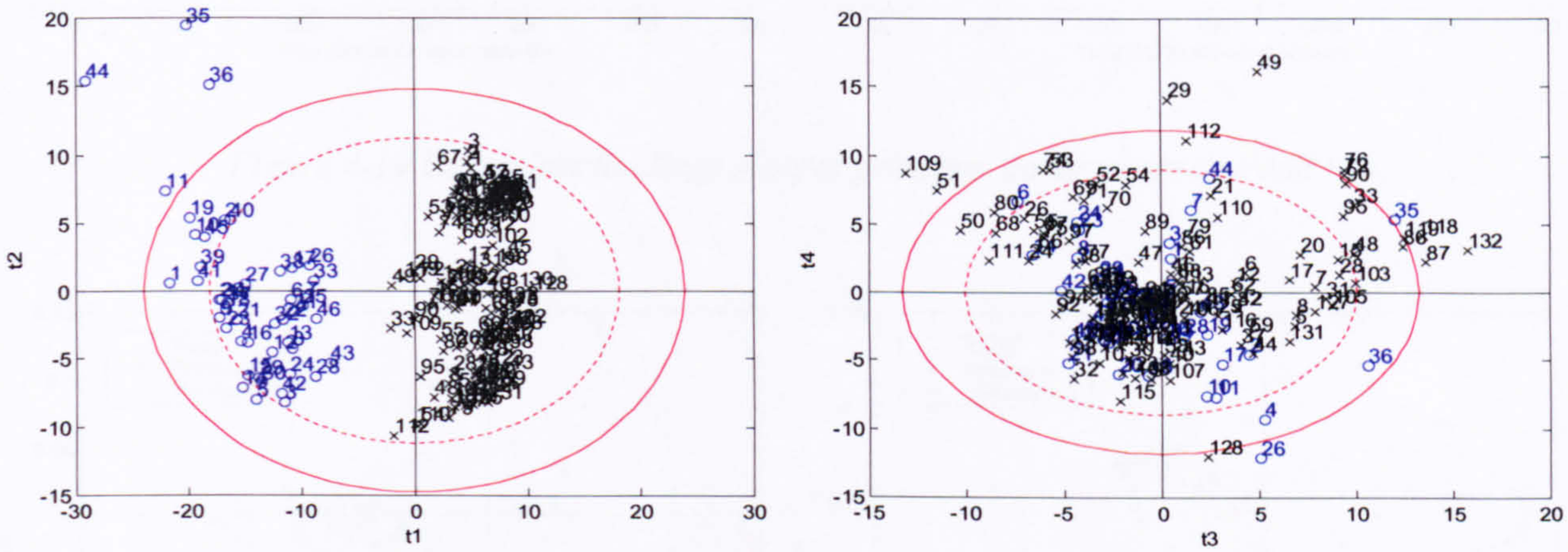


Figure 4-18 Bivariate scores plot of principal components one to four – Site A: "o"; Site B: "x"

Figure 4-19 and Figure 4-20 show the univariate and bivariate loadings plots for principal component one and two, and principal component one to four respectively. There is no strong trend for a particular variable. However, by interrogating the Hotelling's T^2 and SPE monitoring charts (Figure 4-21), the batches from site A are identified as exhibiting non-conforming behaviour from the SPE plot whilst batches from site B do not violet the limits. The two metrics present different types of information than the individual models thus this approach has limitation to be considered as a representative multi-site process analysis.

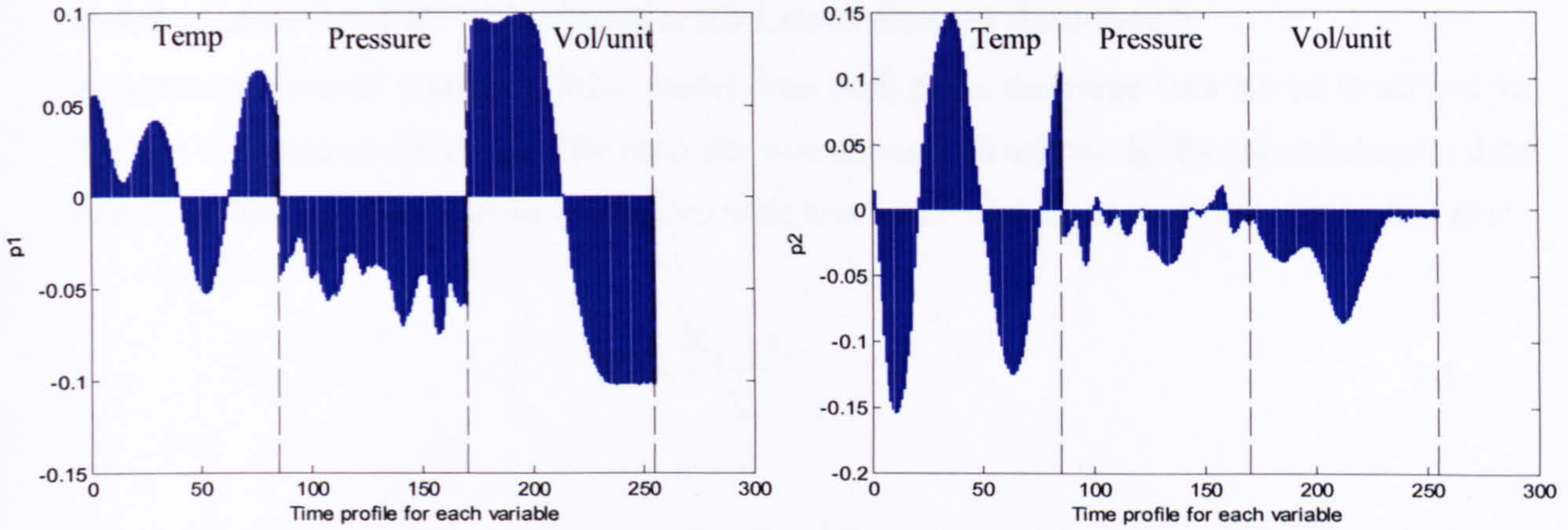


Figure 4-19 Univariate loadings plots of principal components one and two

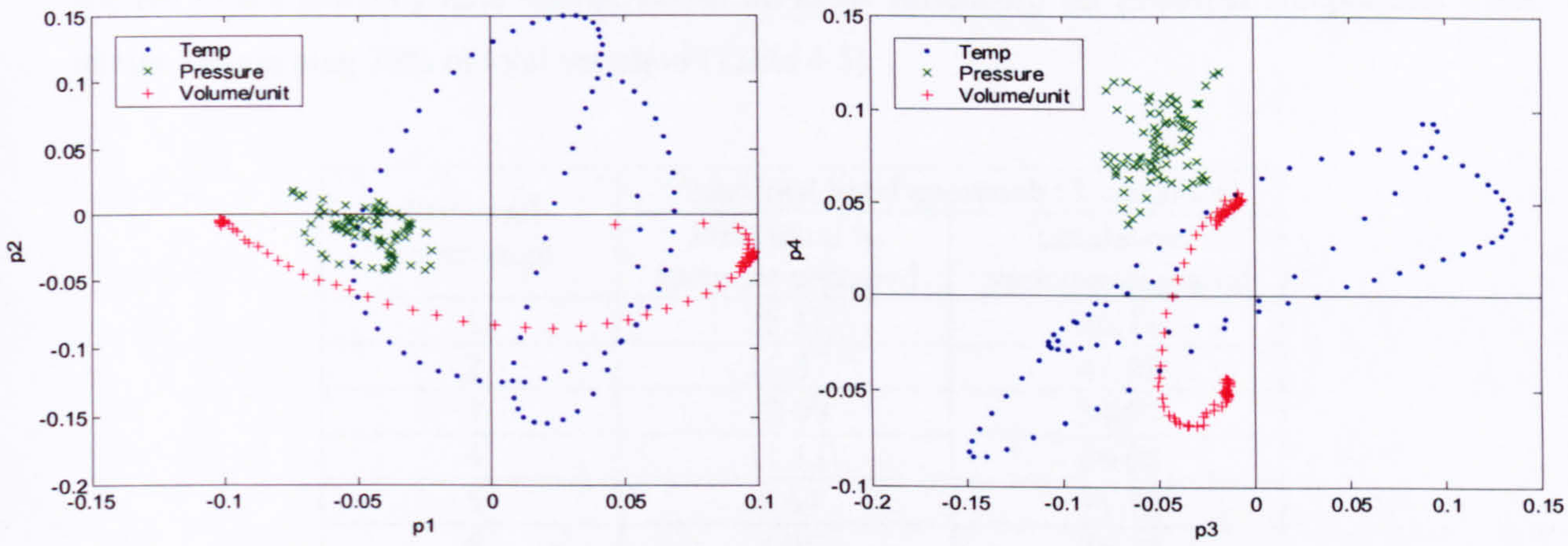


Figure 4-20 Bivariate loadings plots of principal components one to four

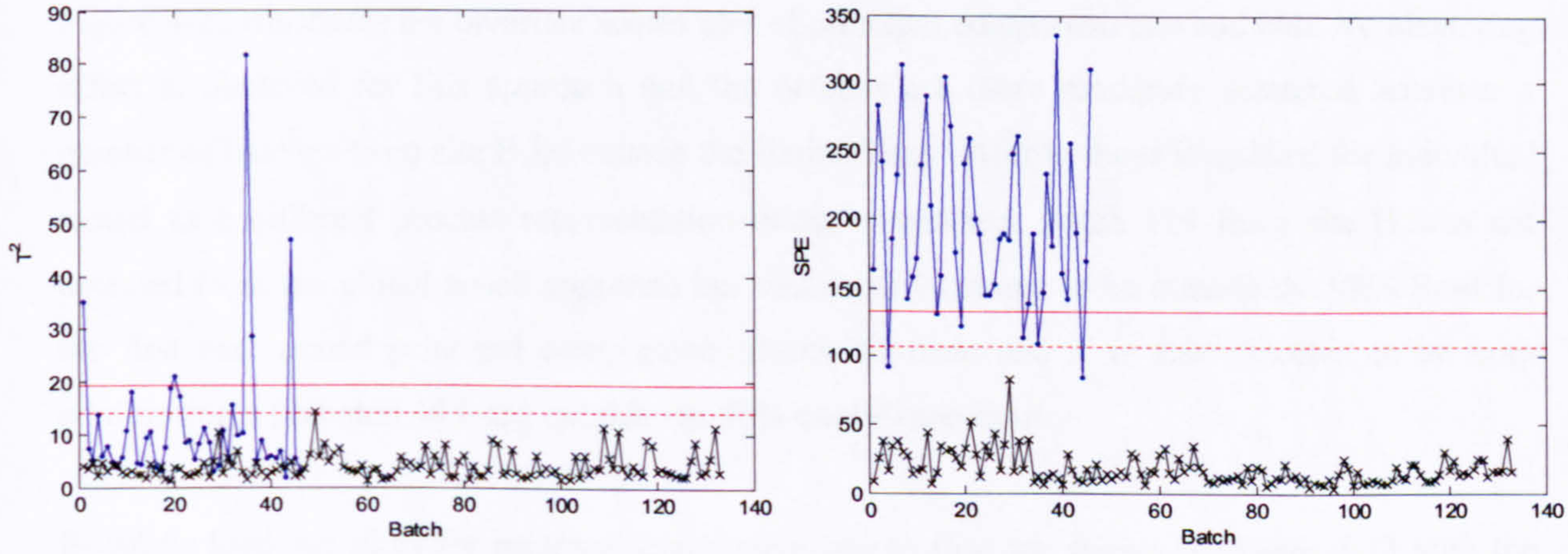


Figure 4-21 Hotelling's T^2 and SPE monitoring charts

4.4.3 Combined MPCA Model with Local Based Approach

A second combined multiway PCA model was built from the same data set as described in Section 4.4.2. However, the data for each site was normalised separately. By normalising the data matrix in this way, the variation of each variable was removed with respect to the individual site:

$$\mathbf{X}_g^+ = \frac{\mathbf{X}_g - \bar{\mathbf{x}}_g}{\mathbf{s}_g} \qquad 4\text{-}3$$

where $\mathbf{X}_g^+$ is the auto-scaled data matrix for site g, $\bar{\mathbf{x}}_g$ and $\mathbf{s}_g$ are the mean and standard deviation vectors for site g. The individual site data is normalised with respect to within site variation therefore separate means and standard deviations are calculated for each site. PCA was then applied to the resulting data matrix. Based on cross validation, six principal components were retained explaining 76% of total variation (Table 4-5).

Principal component	Combined local approach (3 variables)	
	Individual % variance captured	Cumulative % variance captured
1	25.11	25.11
2	16.87	41.98
3	12.99	54.97
4	11.11	66.08
5	5.64	71.72
6	4.05	75.78

Table 4-5 Percentage of variance explained by principal components for the combined local approach

Figure 4-22 illustrates the bivariate scores plot of principal component one and two. No clustering effect is observed for this approach and the batches are more randomly scattered however a number of batches from site B lie outside the limits. They differ to those identified for individual model as a different process representation being considered. Batch 119 from site B was not detected from the global based approach but this batch is shown to be outside the 99% limit for the first and second principal components. Batch 35 from site A is also detected to be non-conforming with batch 44 lying outside the 95% confidence limit.

Bivariate loadings plots for principal component one to four are shown in Figure 4-23 with the univariate plots given in Figure 4-24. Reactor pressure is observed to have a significant influence in terms of determining the direction of greatest variability for principal component one and the

weighting of unit per volume increases in the principal component two after process start-up. This has revealed that the reactor pressure is operated differently between the two sites and the initial start-up of unit per volume addition was similar until the scale differences became apparent. Consequently it can be concluded that the between site variation is captured in this approach. However, by interrogating the Hotelling's T^2 and SPE monitoring charts (Figure 4-25), the batches from site A exhibit non-conforming behaviour in the SPE plot whilst the batches from site B do not violet the limits.

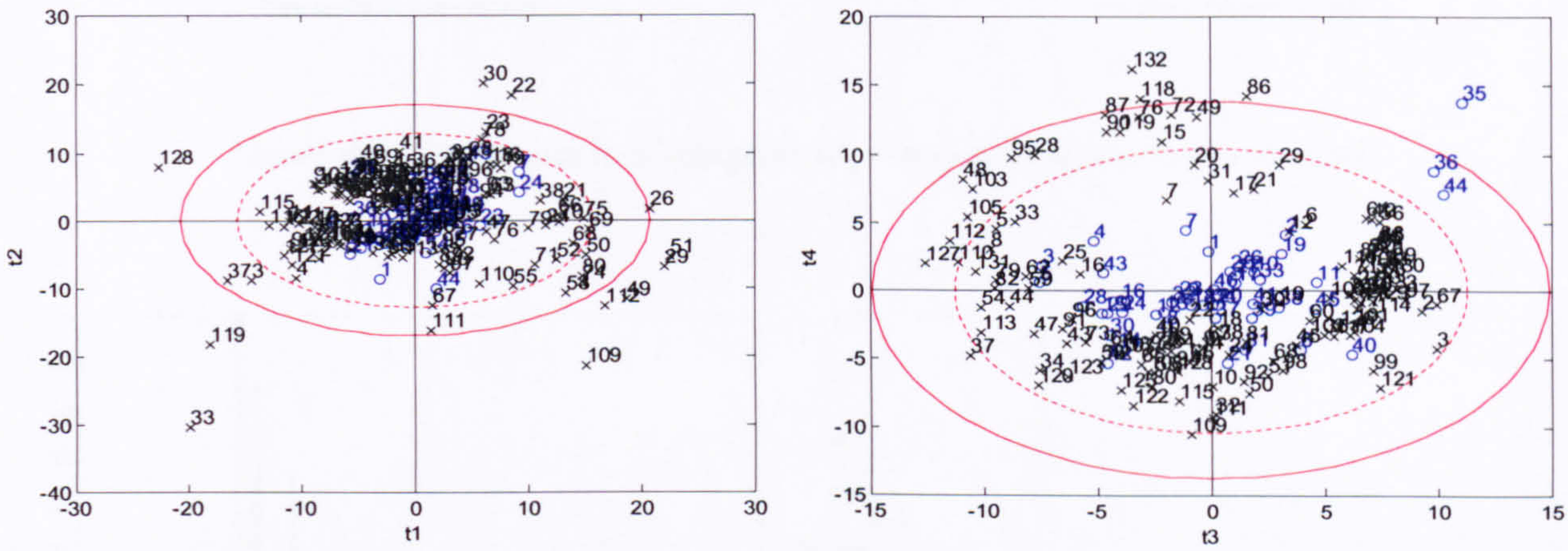


Figure 4-22 Bivariate scores plot of principal components one to four – Site A: "o"; Site B: "x"

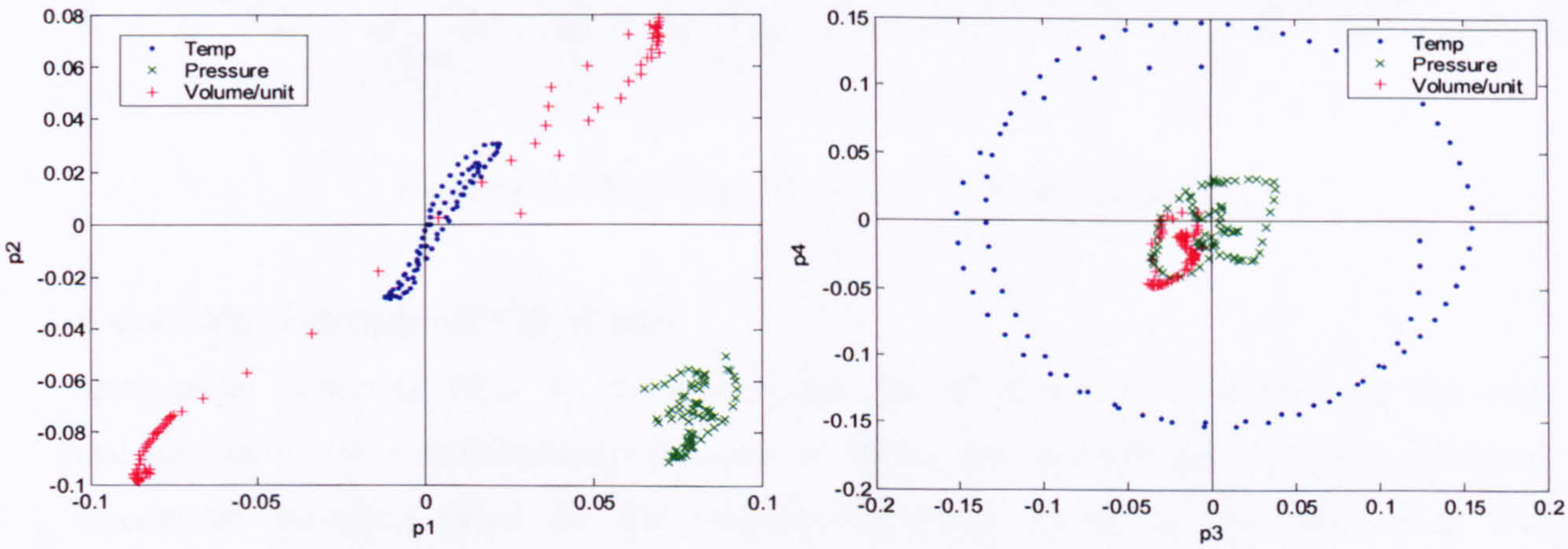


Figure 4-23 Bivariate loadings plots of principal components one to four

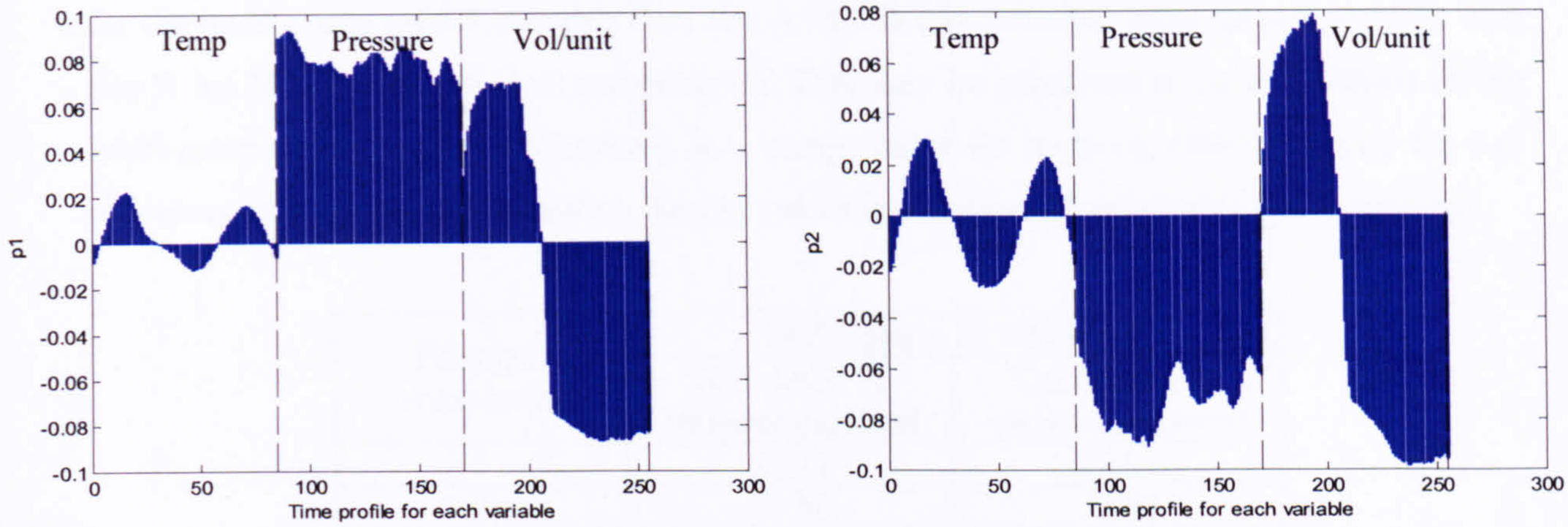


Figure 4-24 Univariate loadings plots of principal components one and two

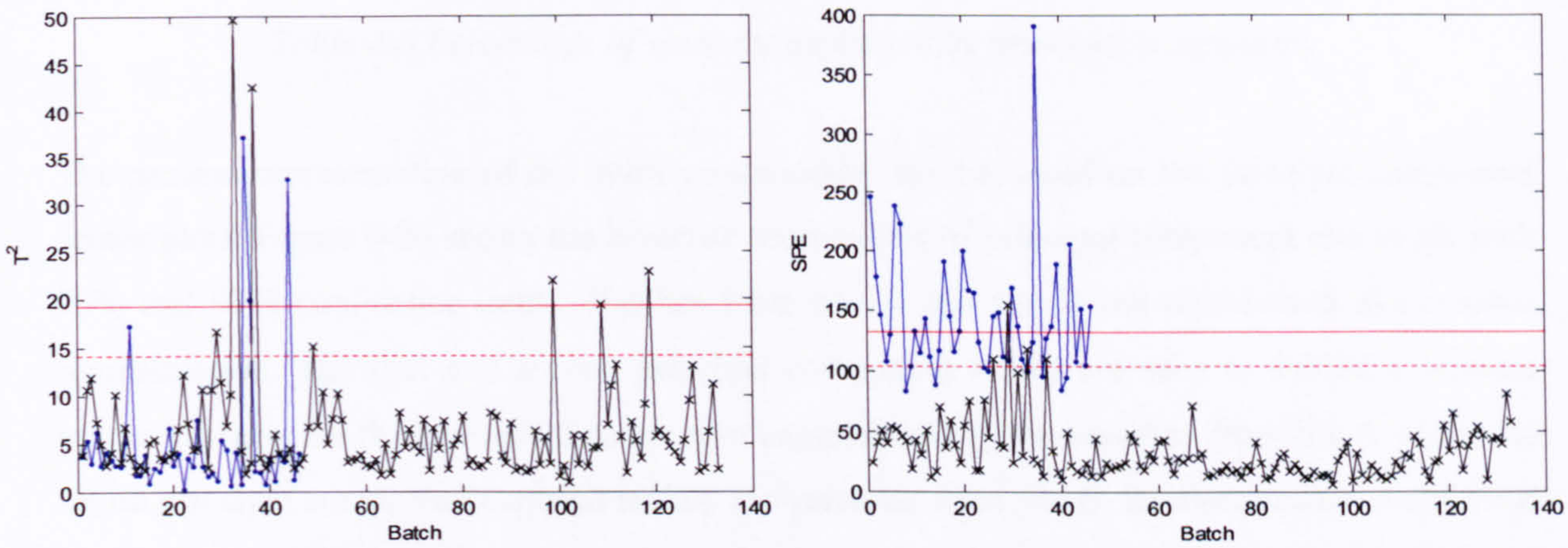


Figure 4-25 Hotelling's T^2 and SPE monitoring charts

4.4.4 Multi-group MPCA Model

Mutli-group multiway PCA is investigated for the simultaneous monitoring of the two manufacturing sites. Multi-group modelling is based on the assumption that a common eigenvector subspace exists for the variance-covariance matrix of individual sites. The development of a multi-group model is described in Section 2.7.3, in particular case two study describes the situation where there were five variables in one group and four variables in the second group and where three variables are the same. This can be adapted for the implementation of this approach.

To determine the optimal number of principal components to be retained in the model, cross-validation was applied. Table 4-6 shows that 67% of the variation is explained by seven principal

components. Compared with the individual site models, the total amount of variability explained for the multi-group model is higher than site A by 9% (six principal components) and less than site B by 24% (seven principal components). This may be perceived to be a limitation of the multi-group model approach. However, it is compensated for by being able to monitor the two processes into a single representation thereby reducing a number of monitoring charts required.

Principal component	Multi-group (6 variables)	
	Individual % variance captured	Cumulative % variance captured
1	26.11	26.11
2	10.50	36.62
3	8.09	44.70
4	7.14	51.85
5	6.26	58.11
6	5.03	63.13
7	3.75	66.88

Table 4-6 Percentage of variance explained by principal components

The process representation of the multi-group model can be based on the principal component scores plots. Figure 4-26 shows the bivariate scores plots of principal component one to six with 95% and 99% confidence limits. Batches from site A and site B are represented in the same representation. The first and second principal component scores are seen to exhibit a random scatter however the third principal component explains mainly the variation from site A whilst the fourth principal component explains mainly the variation from site B. Random scatter is observed again from the fifth principal component scores onward. The outlying batches in site A of 19 and 35 from the individual model are detected in the multi-group model and the outlying batches in site B of 33, 49, 51 and 119 from the individual model are also detected in the multi-group model. Batch 44 from site A and batches 22 and 30 from site B were detected to be outlying in the scores plot from individual models and they fall between the 95% and 99% confidence limits in the multi-group model. Some other batches that fell between the 95% and 99% confidence limits in the individual models are now outlying the 99% confidence limit. The overall representation in multi-group model mirrors the individual models however only a single representation is required for the monitoring of two manufacturing sites although there is an impact to the sensitivity of the detection. This may due to the weighting of batches between the two sites therefore the variation is better for one site than the other.

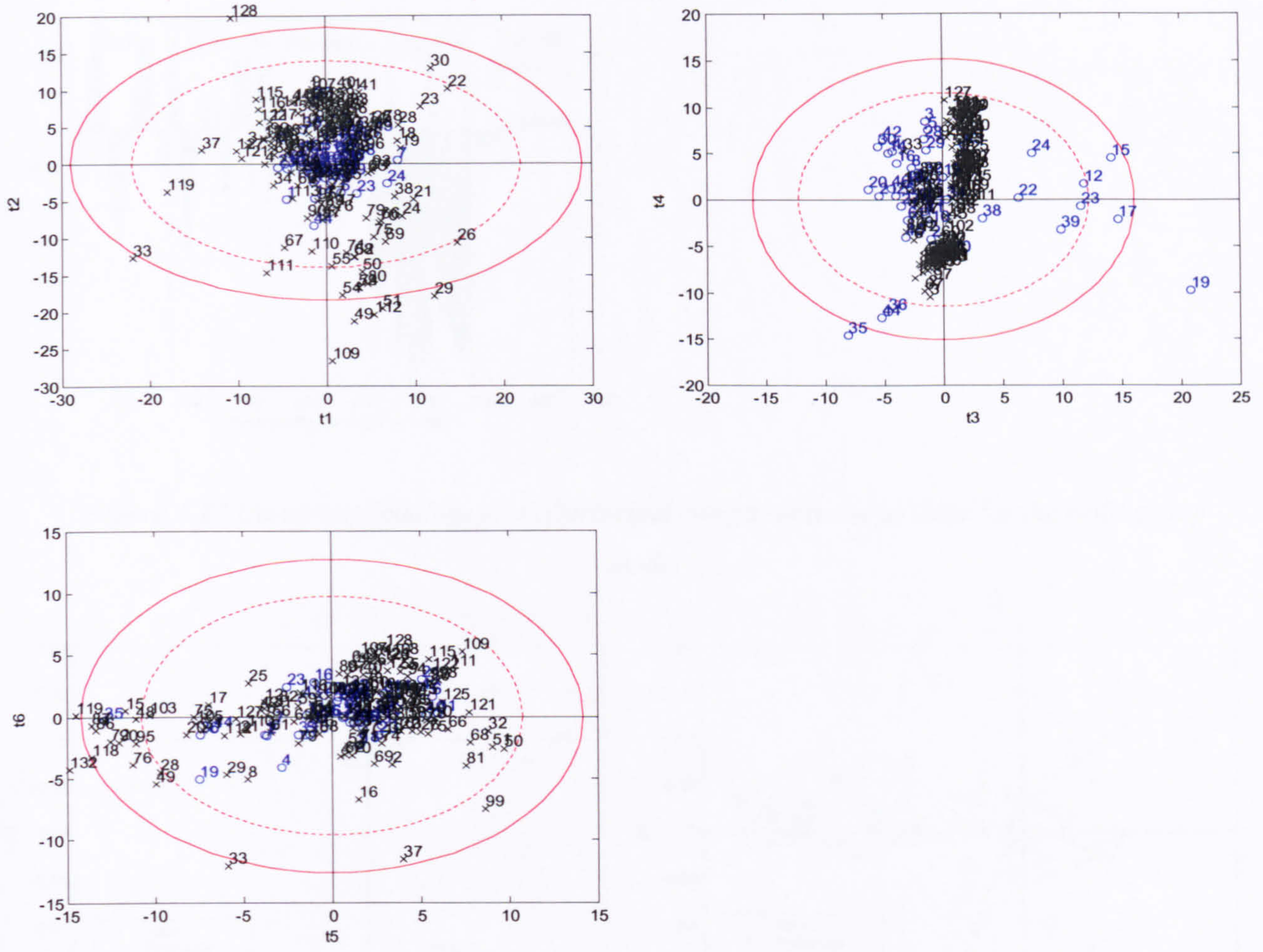
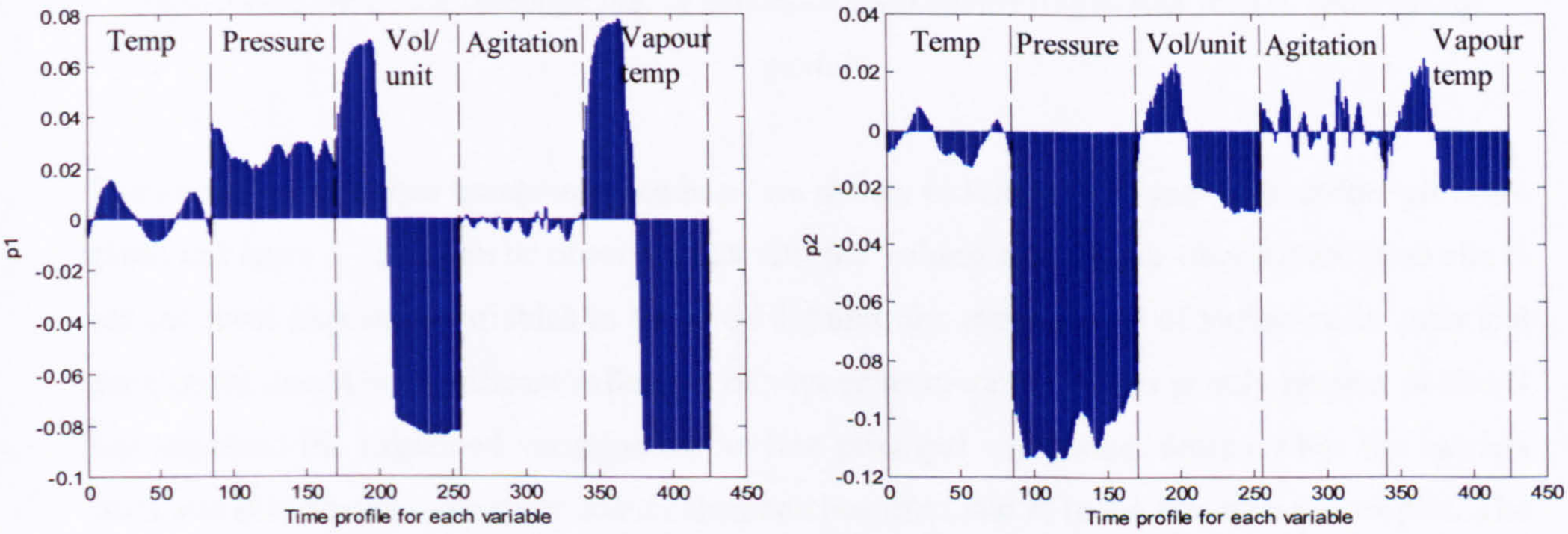


Figure 4-26 Bivariate scores plot of principal components one to six for the multi-group model



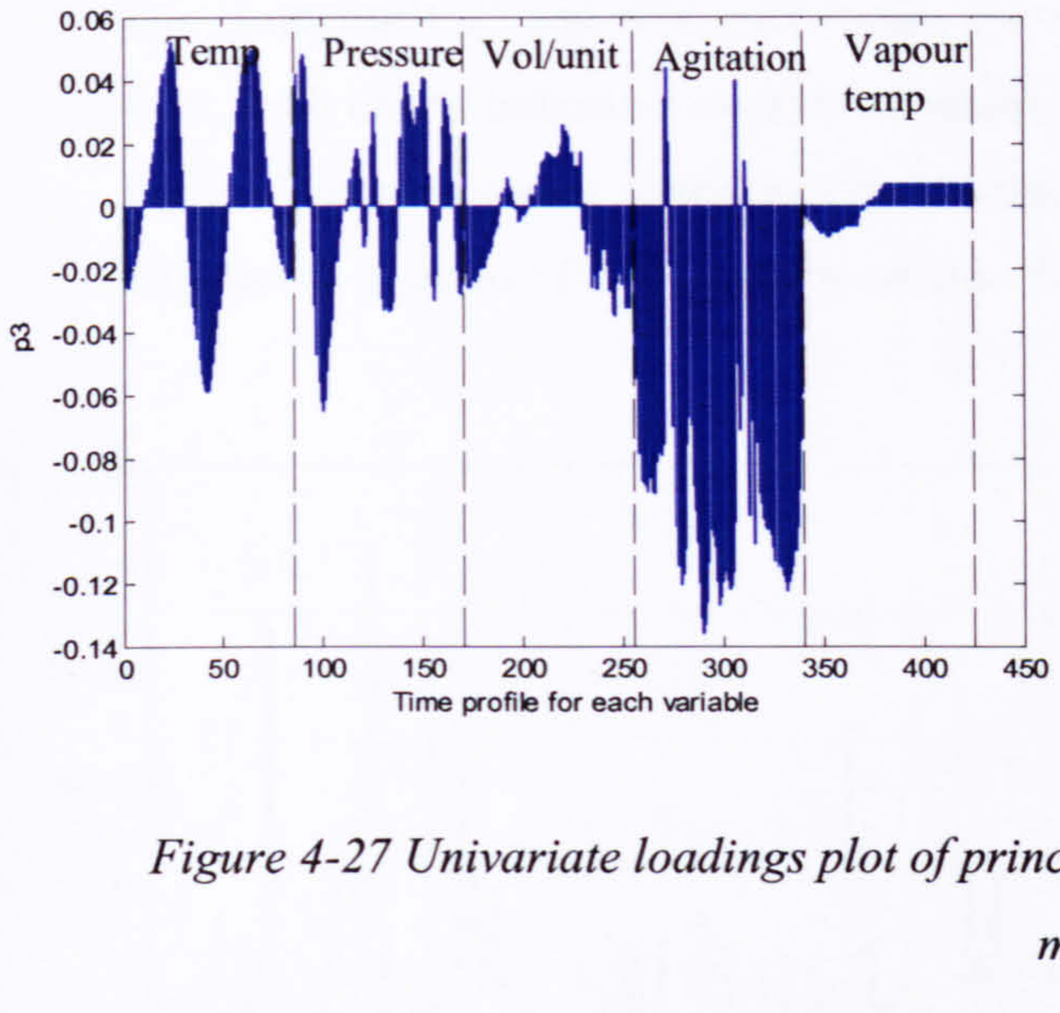


Figure 4-27 Univariate loadings plot of principal components one to three for the multi-group model

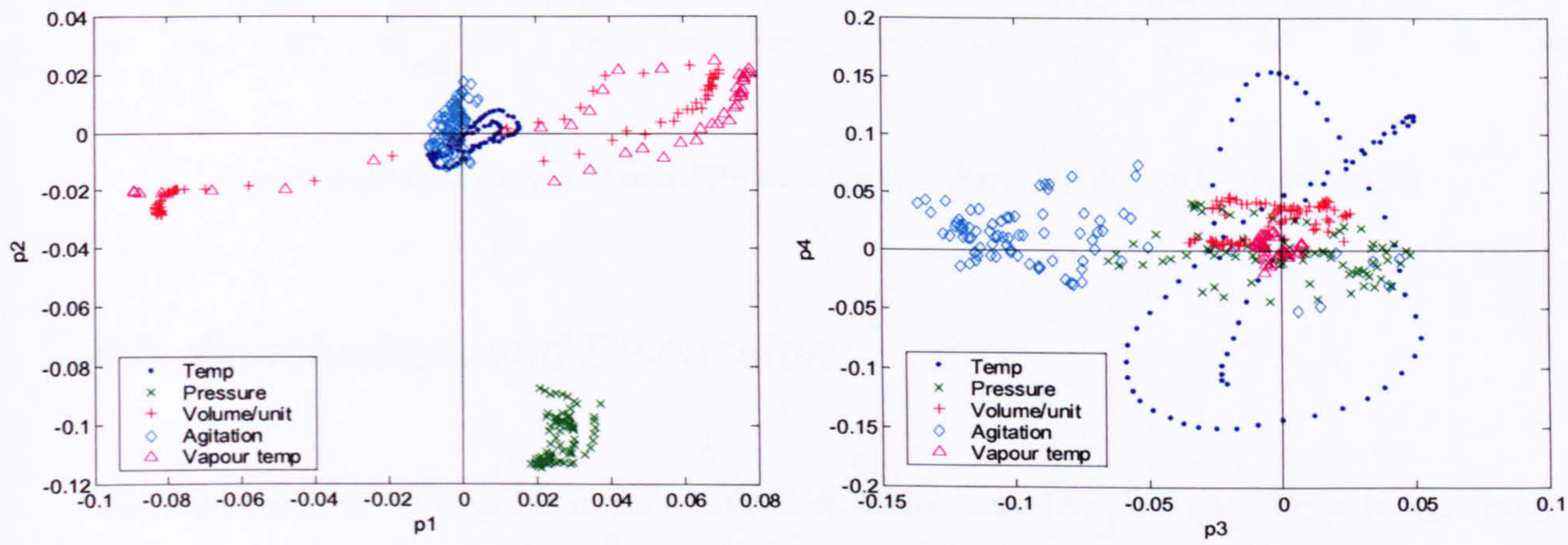


Figure 4-28 Bivariate loadings plot of principal components one to four for the multi-group model

The univariate principal component loadings are shown in Figure 4-27 and the bivariate plots are given in Figure 4-28. It can be observed that unit per volume and vapour temperature from site B are the most important variables in terms of defining the main source of variation for principal component one. The significant influence of vapour temperature which is only present in site B has impacted the explained variation of the first principal component scores. Thus the batches from site B is shown to be more scatter than batches from site A in the bivariate scores plot. The same principle applies to the second principal component loadings where the major influence of pressure from site A has driven the non-normal scatter of principal component two scores. The loadings of principal component three display an influence from the common variables between the two sites whilst the forth principal component exhibits the impact from the agitation rate from site A hence the non-normal scatter for site B in the bivariate scores plot.

The Hotelling's T^2 and SPE monitoring charts for the multi-group model are shown in Figure 4-29. Both charts indicate a number of batches from both sites deviate from the normal operating region. The monitoring charts exhibit similar patterns to the individual models although some differences in terms of these batches outside the 99% confidence limits.

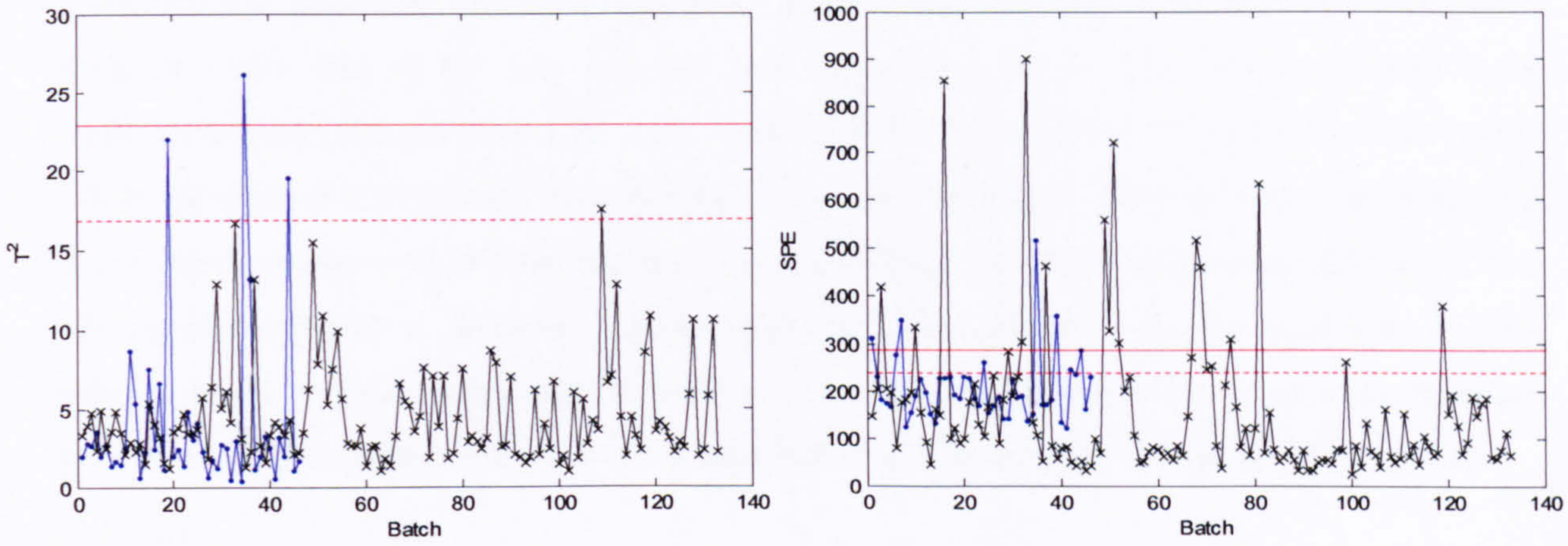


Figure 4-29 Hotelling's T^2 and SPE monitoring charts for the multi-group model

4.5 Conclusions and Discussion

The capabilities of multi-group models have been demonstrated by their application to data from a multi-site drug intermediate batch process. Pre-screening of the data was initially performed to remove any non-common cause variation. Batch length equalisation was then achieved through the cutting of the batches to a pre-defined length to ensure that the analysis focused on the main area of interest. Multi-way principal component analysis was then applied to the pre-processed data. The first approach analysed the data from each site individually. Two combined models where the data was scaled differently were then studied. The multi-group models developed were hypothesised to eliminate between cluster variation but also enable the monitoring of the two sites using a single model. This development theoretically provides a powerful monitoring tool for enhanced process understanding and hence potentially will minimise the differences between process operation across different manufacturing plants in the future. In addition, based on the multi-group approach, it is possible to utilise the methodology to assist in the transfer of a process to a new site or to scale-up a process from pilot plant to full scale manufacturing plant. The concept is based on the assumption that the chemistry and kinetics of the reaction remain unchanged even though the process operates at a different scale or on different equipment.

The advantage of being able to develop a single model for two, or more, sites is that it enables an enhanced understanding of the subtle differences in performance between the two manufacturing processes. In addition it can help facilitate the transfer of a process to a new site by providing a baseline monitoring model with the model being updated as new batches are manufactured. The scores plot detects those batches that move outside the statistical control region for the two sites utilising one representation. Thus, the application has demonstrated that the multi-group model has acceptable detection and diagnostic properties although the overall sensitivity may be reduced compared with that of the corresponding individual site models. Only three variables were common between the two sites with three variables differing hence the full potential of the multi-group approach is not realised. This also highlights the complexity of the industrial situation, that is a common product can be manufactured using a different set of procedures and equipment. The multi-group approach is, however, compensated for by being able to monitor a number of sites using a single process model and the number of control charts needed to monitor the different sites in a highly regulated and responsive manufacturing environment is significantly reduced.

The main source of difference between the sites was that of scale. This aspect has previously limited the comparison between sites. The investigation of within site variation is as critical as between site variation in order to understand the sources of process variability in relation to the product quality. The removal of the global mean and standard deviation enables the between site variation to be captured in the analysis, however within site variation is affected. It is apparent that a scale difference between sites can dominate the analysis therefore the true underlying variability cannot be realised. In contrast, the removal of the local mean and standard deviation enables the within site variation to be captured in the analysis thus the final representation is a mixed site model. The approach was shown to identify a number of batches out with the statistical control limits that were not detected in the individual models. The source of variation captured is different and is again limited by the inclusion of common variables only. However, the local approach is applied to all variables. This ensures that the within site variation is captured. The between site variation is then extracted by pooling the two matrices resulting in a final representation that not only captures the within site variation but also the differences between sites.

Future work should include the development of a multi-group PLS approach to capture the product quality to understand how the process variation has contributed to the difference in quality between sites.

The next chapter will introduce the advanced methodologies for data integration of batch processes and an overview of handling spectral data.

Chapter 5

Advanced Methodologies for Data Integration of Batch Processes

5.1 Introduction

Traditionally batch process performance monitoring has been implemented using physical process parameters (Neogi and Schlags, 1998; Martin and Morris, 2002) and has played an important role in a number of processing industries including pharmaceutical, chemical, polymer and paper. However, there is always a need to improve process understanding and subsequent monitoring, optimisation and control schemes. With the advent of the Process Analytical Technology (PAT) approach, the use of different forms of on-line spectroscopy to attain a detailed understanding of the process has become of increasing importance. In most studies, the spectral data has been analysed independent of the process data (Gurden *et al.*, 2002; van Sprang *et al.*, 2003). However it is hypothesised that by integrating the process (physical state) and spectroscopic (chemical state) measurements for multivariate statistical data analysis and monitoring, an improved process understanding and fault diagnosis can be achieved. A number of approaches have been proposed including the integration of wavelet analysis with Principal Component Analysis (PCA) and Partial Least Squares (PLS) (Bakshi, 1998; Misra *et al.*, 2002; Lu *et al.*, 2003; Hui *et al.*, 2003). The aim of this Chapter is to introduce the two methodologies that are utilised to form the basis of two data integration algorithms – multi-block PCA and the wavelet transform. Finally this Chapter provides an overview of a number of spectroscopic techniques and discusses the handling of spectral data. Applications of process spectroscopy for batch process monitoring are reviewed along with a review of existing methodologies for data integration.

5.2 Multi-block Principal Component Analysis and Its Application

5.2.1 Objectives

With the latest advances in information technology and on-line instrumentation, there is a need for advanced data analysis methods that can handle different data sources including process and spectral data or different forms of spectroscopic data, e.g. Raman and Near-Infrared. The multi-block family of techniques is one set of algorithms that allows the simultaneous analysis of several data sets, thereby enabling the extraction of the underlying relationships existing both between and within the data blocks.

The major advantage of multivariate statistical projection methods is their ability to handle highly correlated variables by reducing the dimensionality of the problem to a few latent variables that capture the main sources of variability. Multi-block based projection techniques have further expanded the range of problems that can be tackled. The goal is to divide a set of inter- or intra-related variables into meaningful blocks thereby realising easier interpretation of the data and hence providing an enhanced understanding of the process. For example for multi-block principal component analysis, the problem is reduced to defining two levels: the sub-level (block level) and the super level. The sub-level represents the behaviour of the individual block whilst the super level reflects the integration of the information from the blocks.

Figure 5-1 illustrates four block constructs. The Class 1 structure is where each individual block comprises both different samples (batches) and variables. Each set of samples or variables represents a structure. In this case, a relationship may exist between different blocks of data, however, there is no direct evidence of a relationship between the data blocks. The multi-site data considered in Chapter 4 falls into this class where there were a different numbers of batches and a different set of variables for the two sites, although some common variables existed between the sites.

For the second class, the samples are common to each block but the variables may differ between the different blocks. This is equivalent to decomposing a complex data set into a series of sub-groups where the inter-relationship between blocks is of interest. This is the class of data that is the focus of this chapter. An example is where different types of data such as physical, chemical and spectral data are collected for a specific system over the same time period with the aim of monitoring the process through the utilisation of the different forms.

Class 3 is where the blocks comprise the same set of variables but the samples differ between blocks. An example is where the same form of spectroscopic data is generated for different chemical systems. A multiplexer can be used for the monitoring of different physical unit operations or same type of unit operations but various manufacturing lines (McLennan and Kowalski, 1995).

In the fourth class, both samples and variables are common. This can be considered to be a common three-way format. This class is not considered in the thesis. However an example of the data can be obtained by running a high through-put experiment where there are many samples analysed at regular intervals with the same number of sensors.

A specific problem will have a preferred solution type for example a multi-group approach can analyse class 1 structure whilst multi-block analysis is suitable for analysis class 2 structure. The latter method is further investigated in the next section.

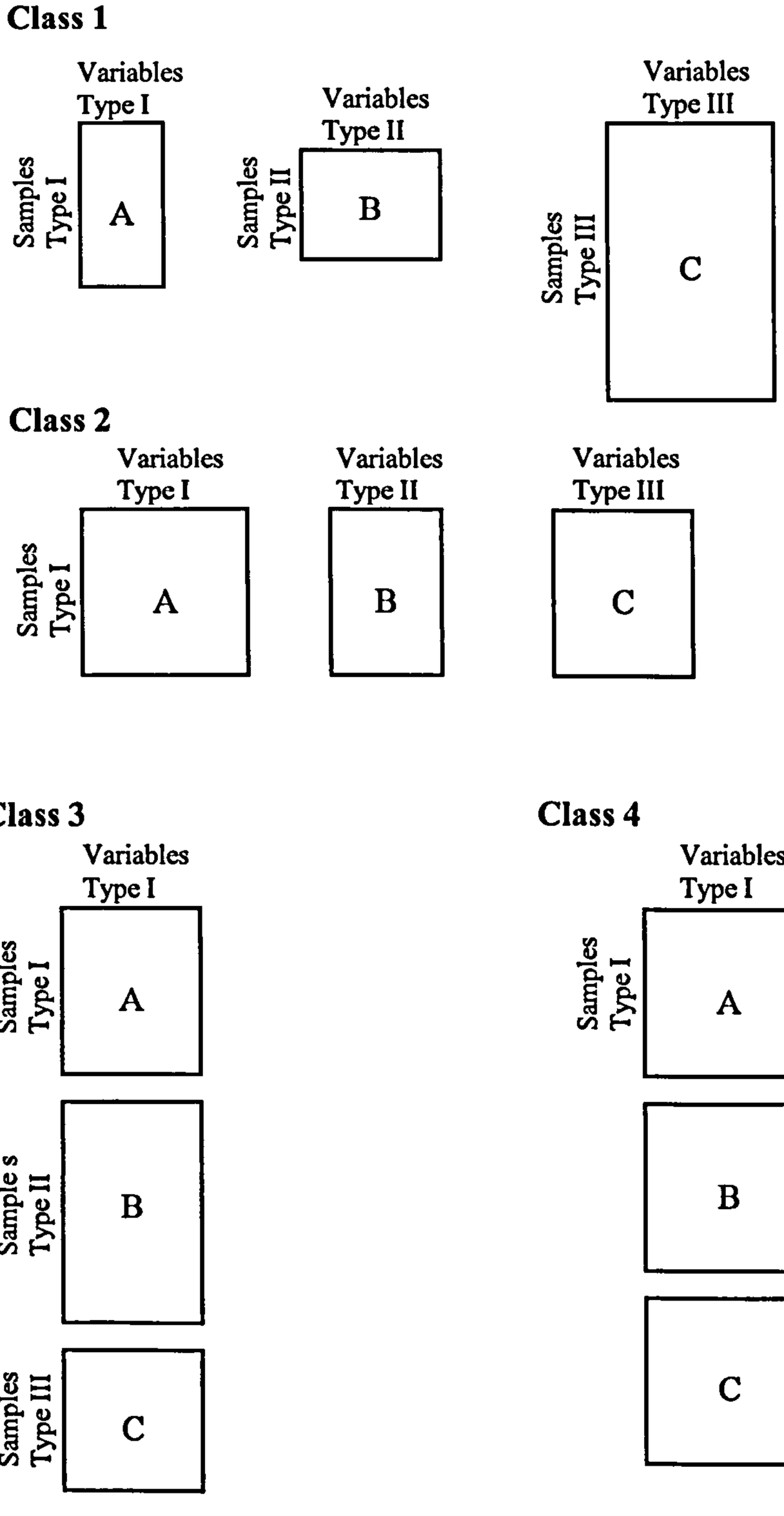


Figure 5-1 Different types of multi-block data structure

5.2.2 Background

Multi-block data analysis has its origins in path analysis and path modelling in the fields of sociology and econometrics. Wold *et al.* (1987a) proposed the concept of consensus PCA and

hence laid the foundation for multi-block analysis in process monitoring. Since this time a number of other algorithms have been proposed including consensus PCA, hierarchical PCA and multiblock PCA. Westerhuis *et al.* (1998) provided a comprehensive analysis of the different multi-block methods and showed that a number of the properties of the multi-block methods are aligned with those of standard PCA and PLS. More recently Qin *et al.* (2001) identified further overlap in terms of the properties of PCA and PLS.

5.2.3 Consensus Principal Component Analysis (CPCA)

The concept of Consensus Principal Component Analysis (CPCA) was introduced by Wold *et al.* (1987a) as a method to compare blocks of variables measured on the same samples. The algorithm is based on Non-linear Iterative Partial Least Squares (NIPALS) and an arrow schematic of the CPCA algorithm is given in Figure 5-2.

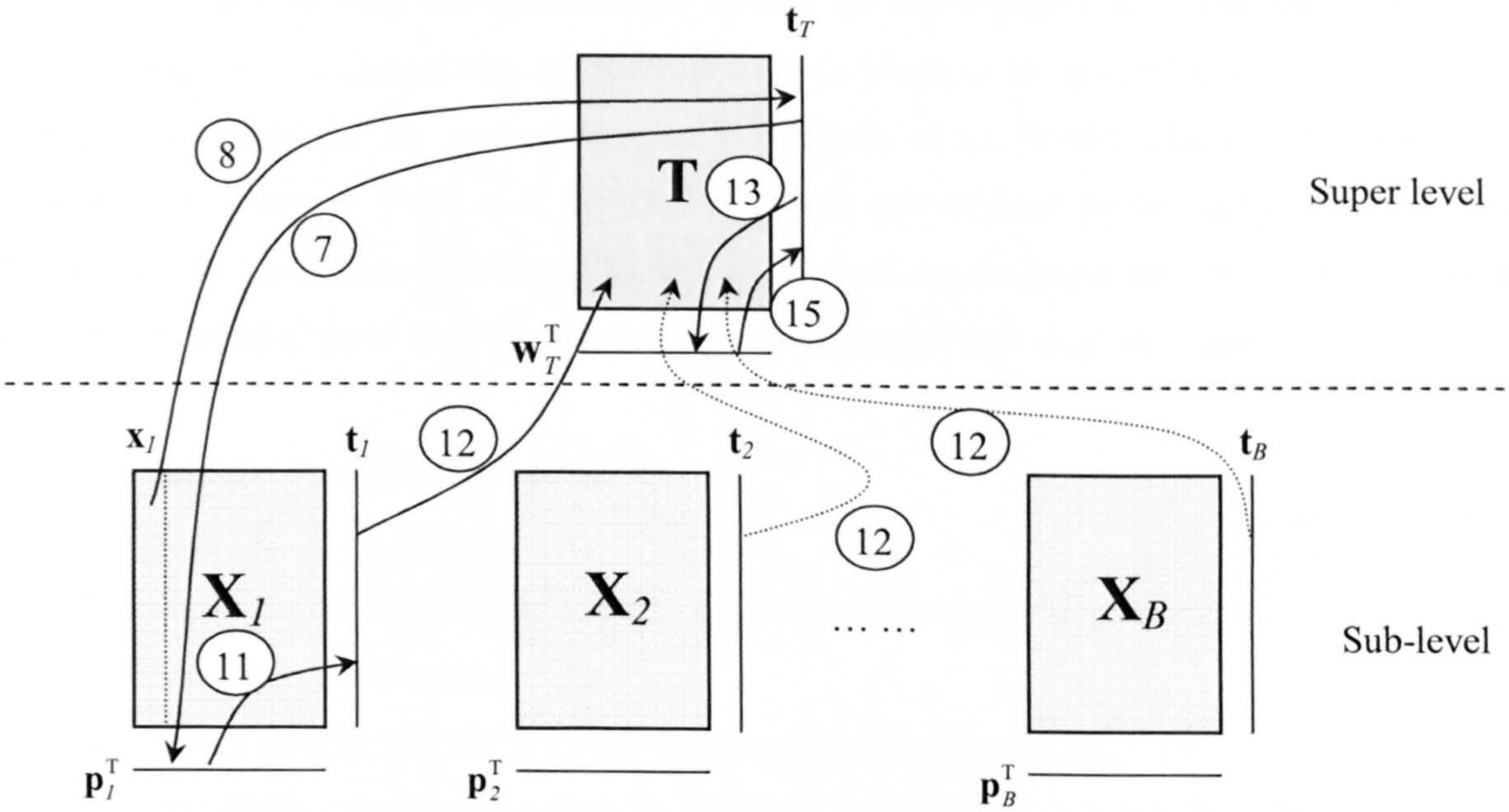


Figure 5-2 Arrow schematic of NIPALS consensus PCA

Class 2 data is considered, that is samples (batches) are common between blocks. In this case, the number of batches is the same between blocks however the number of variables may differ between blocks. The batch dataset is split into a number of blocks $\underline{\mathbf{X}}_1$, $\underline{\mathbf{X}}_2$ … $\underline{\mathbf{X}}_B$. There are no set rules how to define a block but it is important to understand the process and the objectives of the analysis before creating the blocks. The data are checked for missing data and appropriate data pre-processing techniques are first applied as described in Section 3.2.1. If necessary, the batch length should be equalised as discussed in Section 3.2.3. The proposed monitoring approach is adopted into the CPCA algorithms. The three-way matrices $\underline{\mathbf{X}}_b$ (I x J x K) where $b = 1, 2 \ldots B$ are

then unfolded to $\mathbf{X}_b$ (I x JK) and normalised according to Section 3.2.2. Appropriate weighting should be applied to the blocks to achieve equal variance. The unfolded matrices $\mathbf{X}_b$ (I x JK) are then re-arranged to $\mathbf{X}_b$ (IK x J) and the consensus PCA algorithm is applied.

For each principal component r, a starting super score, $\mathbf{t}_{Tr}$ is selected as the first column of one of the blocks and this vector is regressed on all blocks to give the block variable loadings $\mathbf{p}_{br}$ (b = 1, 2 ... B). The block scores $\mathbf{t}_{br}$ are calculated and combined into a super block $\mathbf{T}_r$. The super scores are then regressed on the super block to give the super weights (loadings of the super block) $\mathbf{w}_{Tr}$ of the block scores with the super weight being normalised to unit length. A new super score is calculated. The procedure is repeated until the super score converges. After convergence, all the blocks are deflated using the super score and the second principal component is then determined by repeating the process on the residual matrix.

The relationship between the block scores $\mathbf{t}_{br}$ and the super scores $\mathbf{t}_{Tr}$ is defined by the super weights $\mathbf{w}_{Tr}$ which indicates the contribution of each block to the overall scores. The algorithm presented is based on the methodology of Westerhuis *et al.* (1998) who modified the initial algorithm proposed by Wold *et al.* (1987a) to address convergence problems by normalising the block variable loadings to unit length as well as normalising the super weight. A summary of the CPCA procedure aligned with the proposed monitoring approach is given below:

1. Split the batch dataset into blocks.
   $\underline{\mathbf{X}} \rightarrow [\underline{\mathbf{X}}_1\ \underline{\mathbf{X}}_2 \ldots \underline{\mathbf{X}}_B]$
2. Check for missing data and apply data pre-processing techniques as necessary, see Section 3.2.1.
3. If necessary, apply batch length alignment, see Section 3.2.3.
4. Unfold the three-way matrices $\underline{\mathbf{X}}_b$ (I x J x K) to $\mathbf{X}_b$ (I x KJ), b = 1, 2 ... B.
5. Centre and scale the data appropriately, see Section 3.2.2.
6. Apply appropriate weighting to achieve equal variance between blocks, see Section 3.2.2.
7. Re-arrange the unfolded matrices $\mathbf{X}_b$ (I x KJ) to $\mathbf{X}_b$ (IK x J), b = 1, 2 ... B. Set r = 1.
8. For each dimension (principal component), let first column of $\mathbf{X}_b$ be the starting vector for the super score $\mathbf{t}_{Tr}$.
9. The starting super score $\mathbf{t}_{Tr}$ is regressed on all blocks $\mathbf{X}_b$ to give the block variable loadings $\mathbf{p}_{br}^{\mathrm{T}}$.

5-1

$$\mathbf{p}_{br} = \frac{\mathbf{X}_b^{\mathrm{T}} \cdot \mathbf{t}_{Tr}}{\mathbf{t}_{Tr}^{\mathrm{T}} \cdot \mathbf{t}_{Tr}} \qquad b = 1, 2 \ldots B \,.$$

10. Normalise $\mathbf{p}_{br}$ to unit length.

    $$\| \mathbf{p}_{br} \| = 1 \,.$$

11. Block scores $\mathbf{t}_{br}$ are then calculated.

    $$\mathbf{t}_{br} = \mathbf{X}_b \cdot \mathbf{p}_{br} \,. \qquad \text{5-2}$$

12. The block scores $\mathbf{t}_{br}$ are combined to give the super block $\mathbf{T}_r$.

    $$\mathbf{T}_r \leftarrow [\mathbf{t}_{1r}\, \mathbf{t}_{2r} \ldots \mathbf{t}_{Br}] \,.$$

13. Standard PCA is performed on the super block $\mathbf{T}_r$.

    $$\mathbf{w}_{Tr} = \frac{\mathbf{T}_r^{\mathrm{T}} \cdot \mathbf{t}_{Tr}}{\mathbf{t}_{Tr}^{\mathrm{T}} \cdot \mathbf{t}_{Tr}} \,. \qquad \text{5-3}$$

14. Normalise $\mathbf{w}_{Tr}$ to unit length.

    $$\| \mathbf{w}_{Tr} \| = 1 \,.$$

15. Calculate the super score vector $\mathbf{t}_{Tr}$.

    $$\mathbf{t}_{Tr} = \mathbf{T}_r \cdot \mathbf{w}_{Tr} \,. \qquad \text{5-4}$$

16. Check for convergence. If $\mathbf{t}_{Tr}$ has not converged, go to Step (9).

17. Calculate the block loadings $\mathbf{p}_{br}^{\mathrm{T}}$.

    $$\mathbf{p}_{br} = \frac{\mathbf{X}_b^{\mathrm{T}} \cdot \mathbf{t}_{Tr}}{\mathbf{t}_{Tr}^{\mathrm{T}} \cdot \mathbf{t}_{Tr}} \,. \qquad \text{5-5}$$

18. Perform deflation step.

    $$\mathbf{X}_{bNew} = \mathbf{X}_b - \mathbf{t}_{Tr} \cdot \mathbf{p}_{br}^{\mathrm{T}} \qquad b = 1, 2 \ldots B \,. \qquad \text{5-6}$$

19. Replace $\mathbf{X}_b$ by $\mathbf{X}_{bNew}$ and calculate the next dimension, i.e. $r = r + 1$, go to step (8).

20. Stop where maximum number of components have been calculated or when desired number of principal components have been calculated.

5.2.4 Hierarchical Principal Component Analysis (HPCA)

Hierarchical Principal Component Analysis (HPCA) was introduced by Wold *et al.* (1996) as a modified version of the multi-block PCA methods. An arrow schematic of the algorithm based on NIPALS is given in Figure 5-3.

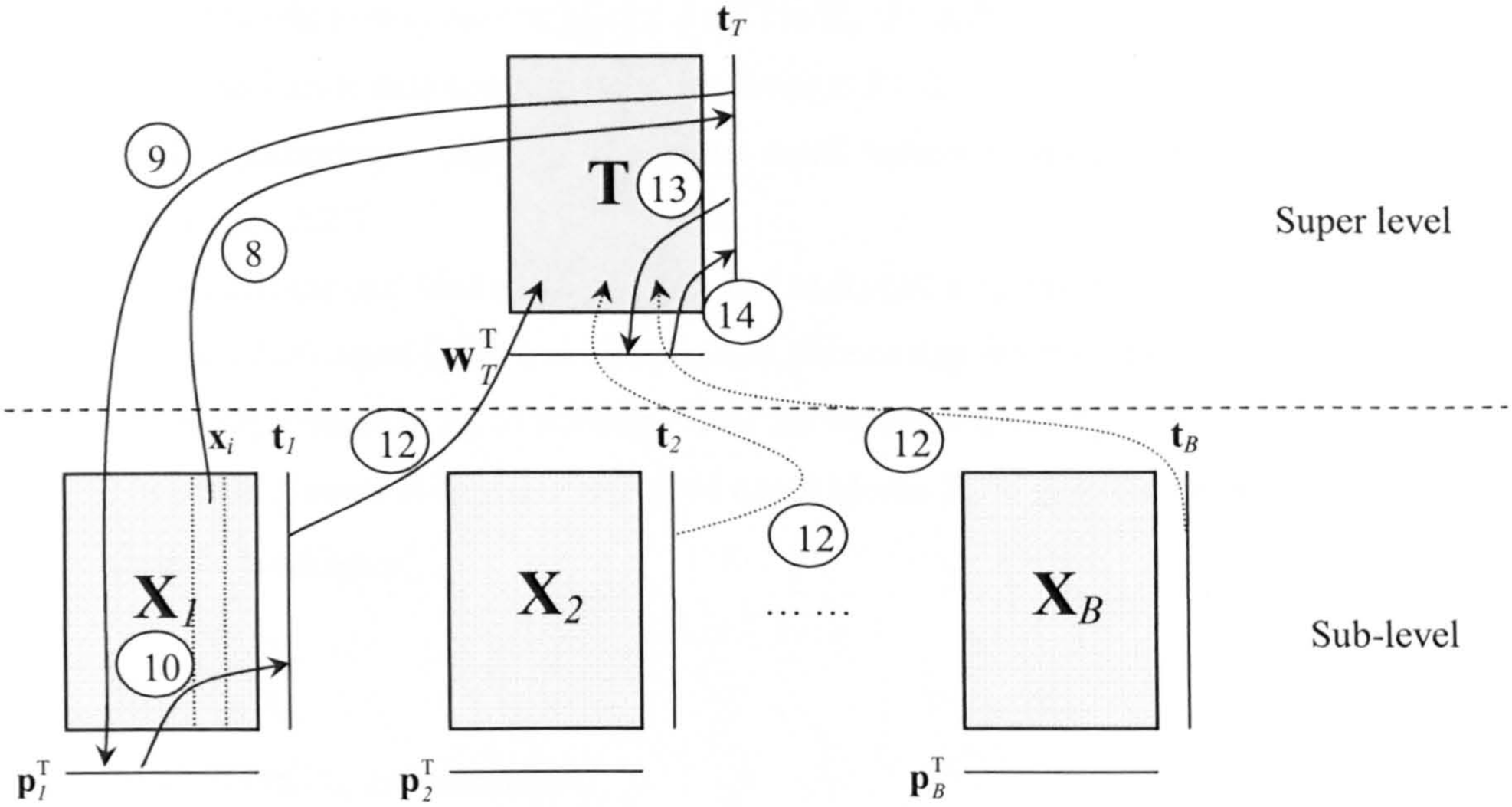


Figure 5-3 Arrow schematic of NIPALS hierarchical PCA

For each principal component *r*, a starting super score $\mathbf{t}_{Tr}$ is selected as that eigenvector which has the largest eigenvalue following the decomposition of $\mathbf{X}_b^T\mathbf{X}_b$. The starting super score is regressed on all blocks of $\mathbf{X}_b$ to give the block variable loadings $\mathbf{p}_{br}^T$ ($b = 1, 2 \ldots B$). The block scores $\mathbf{t}_{br}$ ($b = 1, 2 \ldots B$) can then be calculated from the block loadings. The block scores are then normalised to unit length prior to being combined into a super block matrix $\mathbf{T}_r$. The super score is then regressed on the super block to give the super weight $\mathbf{w}_{Tr}$ of each block score to the super score. The super score is then normalised to unit length and a new super score is calculated. This procedure is repeated until the super score converges to a pre-defined precision. After convergence, all the blocks are deflated using the super score and the second principal component is determined by repeating the process on the residual matrix. The algorithm was modified by Westerhuis *et al.* (1998) since the original algorithm converged to different solutions depending on the starting vector. The block score is also normalised to unit length in addition to the super score to address the issue. A summary of the HPCA procedure for handling batch processes with the proposed monitoring approach is as follows:

1. Split the batch dataset into blocks.
   $\underline{\mathbf{X}} \rightarrow [\underline{\mathbf{X}}_1\ \underline{\mathbf{X}}_2 \ldots \underline{\mathbf{X}}_B]$
2. Check for missing data and apply data pre-processing techniques as necessary, see Section 3.2.1.
3. If necessary, apply batch length alignment, see Section 3.2.3.

4. Unfold the three-way matrix $\underline{\mathbf{X}}_b$ (I x J x K) to $\mathbf{X}_b$ (I x KJ).
5. Centre and scale data appropriately, see Section 3.2.2.
6. Apply appropriate weighting to achieve equal variance between blocks, see Section 3.2.2.
7. Re-arrange the unfolded matrix $\mathbf{X}_b$ (I x KJ) to $\mathbf{X}_b$ (IK x J). Set $r = 1$.
8. For each dimension (principal component), choose eigenvector of the largest eigenvalue of $\mathbf{X}_b$ as starting vector for the super score $\mathbf{t}_{Tr}$.
9. The starting super score $\mathbf{t}_{Tr}$ is regressed on all blocks $\mathbf{X}_b$ to give the block variable loadings $\mathbf{p}_{br}^{\mathrm{T}}$.

$$\mathbf{p}_{br} = \frac{\mathbf{X}_b^{\mathrm{T}} \cdot \mathbf{t}_{Tr}}{\mathbf{t}_{Tr}^{\mathrm{T}} \cdot \mathbf{t}_{Tr}} \qquad b = 1, 2 \ldots B \,. \tag{5-7}$$

10. Block scores $\mathbf{t}_{br}$ are calculated.

$$\mathbf{t}_{br} = \mathbf{X}_b \cdot \mathbf{p}_{br} \,. \tag{5-8}$$

11. Normalise $\mathbf{t}_{br}$ to unit length.

$$\| \mathbf{t}_{br} \| = 1 \,.$$

12. The block scores $\mathbf{t}_{br}$ are combined to give the super block $\mathbf{T}_r$.

$$\mathbf{T}_r \leftarrow [\mathbf{t}_{1r}\ \mathbf{t}_{2r} \ldots \mathbf{t}_{Br}] \,.$$

13. Standard PCA is performed on the super block $\mathbf{T}_r$.

$$\mathbf{w}_{Tr} = \frac{\mathbf{T}_r^{\mathrm{T}} \cdot \mathbf{t}_{Tr}}{\mathbf{t}_{Tr}^{\mathrm{T}} \cdot \mathbf{t}_{Tr}} \,. \tag{5-9}$$

14. Calculate the super score vector $\mathbf{t}_{Tr}$.

$$\mathbf{t}_{Tr} = \mathbf{T}_r \cdot \mathbf{w}_{Tr} \,. \tag{5-10}$$

15. Normalise $\mathbf{t}_{Tr}$ to unit length.

$$\| \mathbf{t}_{Tr} \| = 1 \,.$$

16. Check for convergence. If $\mathbf{t}_{Tr}$ has not converged, go to Step (9).
17. Perform deflation step.

$$\mathbf{X}_{bNew} = \mathbf{X}_b - \mathbf{t}_{Tr} \cdot \mathbf{p}_{br}^{\mathrm{T}} \,. \tag{5-11}$$

18. Replace $\mathbf{X}_b$ by $\mathbf{X}_{bNew}$ and calculate the next dimension, i.e. $r = r + 1$, go to step (8).
19. Stop where maximum number of components have been calculated or when desired number of principal components have been calculated.

The difference between CPCA and HPCA is that the super score in HPCA is normalised to unit length whilst the normalisation is performed for the super weight in CPCA. Additional research has shown that if the starting eigenvector of the matrix is selected to be the largest eigenvalue, this forces the algorithm to a specific solution hence the objective function for HPCA is not as

clear as for CPCA where the objective is to maximise the variance in **X** (Westerhuis *et al.*, 1998). The introduction of the extra normalisation step helps stabilise the direction of the final solutions.

Another feature of HPCA is that when the block scores are orthogonal, the super score of that principal component is the mean of all the block scores. Normalisation of the block scores transfers the explanation of different variables in each block to the loadings hence it is more difficult to reveal between blocks variation.

5.2.5 Multiblock Principal Component Analysis (MBPCA)

Chen and McAvoy (1997, 1998) proposed an alternative multi-block PCA algorithm, Multiblock Principal Component Analysis (MBPCA). The MBPCA algorithm is based on the concept of PCA but the block variable loadings are used to form the super block instead of the scores. The algorithm is described schematically through the arrow schematic shown in Figure 5-4.

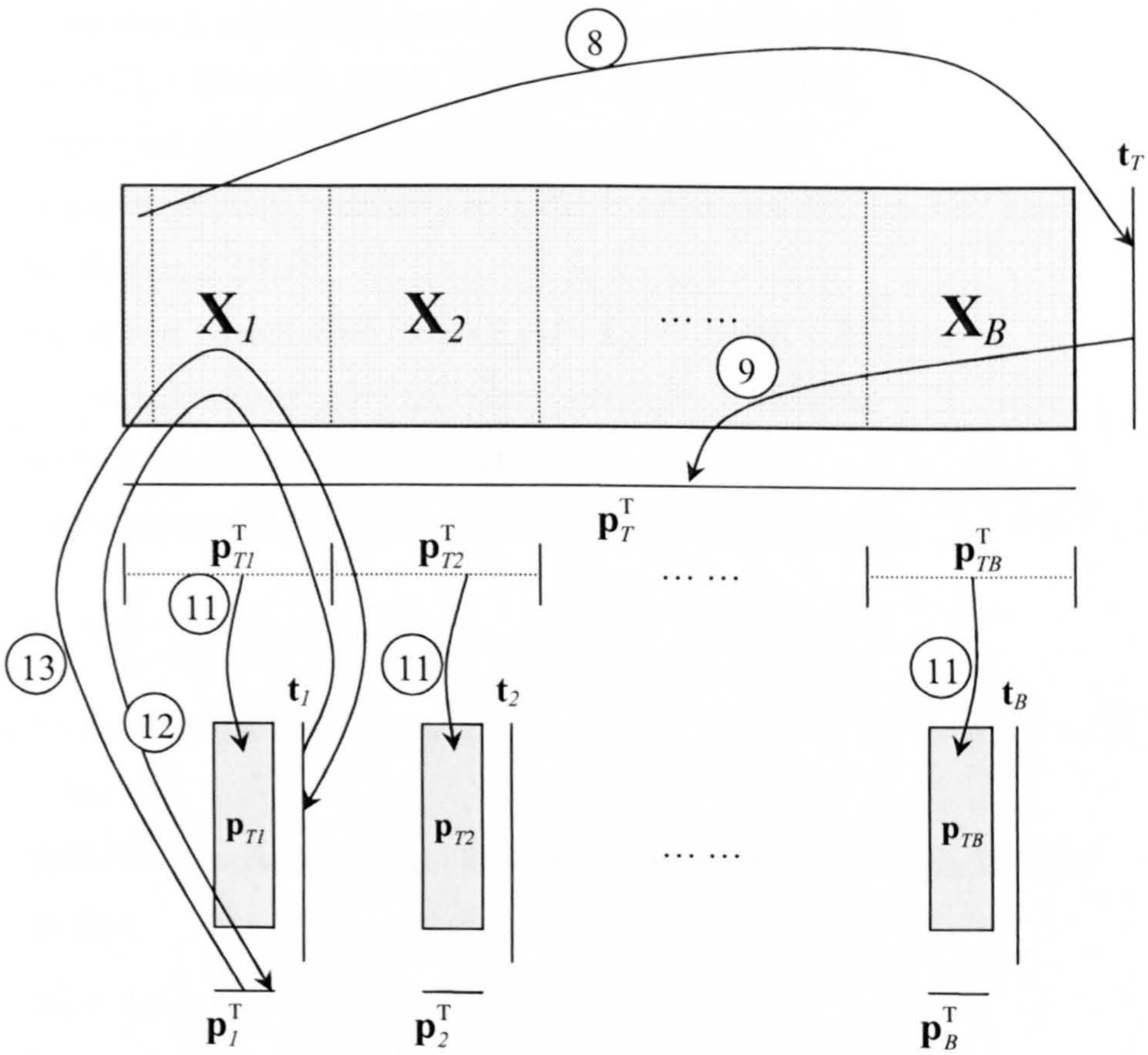


Figure 5-4 Arrow Schematic of NIPALS multiblock PCA

For each principal component, a starting super score, vector $\mathbf{t}_{Tr}$, is selected as the first column of the **X** matrix. The starting score is regressed on **X** to obtain the super loading $\mathbf{p}_T^{\mathrm{T}}$. The super loading is then divided into different blocks in the same way as for **X**. In each block, the super loading is regressed on the block variables to give the block scores $\mathbf{T}_{br}$ ($b = 1, 2 \ldots B$). The block loading $\mathbf{P}_{br}$ ($b = 1, 2 \ldots B$) is calculated following the convergence of the block scores. After convergence, all the blocks are deflated and the resulting residual matrices are combined to form the new **X** matrix. The second super loading is determined by repeating the overall process on the residual matrix. A summary of the MBPCA procedure aligned with the proposed monitoring approach is as follows:

1. Split the batch dataset into blocks.
   $\underline{\mathbf{X}} \rightarrow [\underline{\mathbf{X}}_1\ \underline{\mathbf{X}}_2 \ldots \underline{\mathbf{X}}_B]$
2. Check for missing data and apply data pre-processing techniques as necessary, see Section 3.2.1.
3. If necessary, apply batch length alignment, see Section 3.2.3.
4. Unfold the three-way matrix $\underline{\mathbf{X}}_b$ (I x J x K) to $\mathbf{X}_b$ (I x KJ).
5. Centre and scale data appropriately, see Section 3.2.2.
6. Apply appropriate weighting to achieve equal variance between blocks, see Section 3.2.2.
7. Re-arrange the unfolded matrix $\mathbf{X}_b$ (I x KJ) to $\mathbf{X}_b$ (IK x J). Set $r = 1$.
8. For each dimension (principal component), let first column of **X** be the starting vector for the super score $\mathbf{t}_{Tr}$.
9. Perform standard PCA on **X** to obtain the super loadings $\mathbf{p}_{Tr}^{\mathrm{T}}$.

$$\mathbf{p}_{Tr} = \frac{\mathbf{X}^{\mathrm{T}} \cdot \mathbf{t}_{Tr}}{\mathbf{t}_{Tr}^{\mathrm{T}} \cdot \mathbf{t}_{Tr}}. \qquad 5\text{-}12$$

10. Normalise $\mathbf{p}_{Tr}$ to unit length.
    $\| \mathbf{p}_{Tr} \| = 1$.
11. Split the super loadings $\mathbf{p}_{Tr}^{\mathrm{T}}$ into different blocks in the same way as **X** is divided.
    $\mathbf{p}_{Tr}^{\mathrm{T}} \rightarrow [\mathbf{p}_{T1r}^{\mathrm{T}}\ \mathbf{p}_{T2r}^{\mathrm{T}} \ \cdots\ \mathbf{p}_{TBr}^{\mathrm{T}}]$.
12. Block scores $\mathbf{t}_{br}$ are calculated.

$$\mathbf{t}_{br} = \frac{\mathbf{X}_b \cdot \mathbf{p}_{Tbr}}{\mathbf{p}_{Tbr}^{\mathrm{T}} \cdot \mathbf{p}_{Tbr}}. \qquad 5\text{-}13$$

13. Calculate the block loadings $\mathbf{p}_{br}^{\mathrm{T}}$. 5-14

$$\mathbf{p}_{br} = \frac{\mathbf{X}_b^{\mathrm{T}} \cdot \mathbf{t}_{br}}{\mathbf{t}_{br}^{\mathrm{T}} \cdot \mathbf{t}_{br}} .$$

14. Check for convergence. If $\mathbf{t}_{br}$ has not converged, go to Step (9).
15. Perform deflation step.

$$\mathbf{X}_{bNew} = \mathbf{X}_b - \mathbf{t}_{br} \cdot \mathbf{p}_{br}^{\mathrm{T}} \quad . \qquad 5\text{-}15$$

16. The residual matrices $\mathbf{X}_{bNew}$ are combined to form the matrix $\mathbf{X}_{New}$.

$$\mathbf{X}_{New} = [\mathbf{X}_{1New}\, \mathbf{X}_{2New} \ldots \mathbf{X}_{BNew}] \, .$$

17. Replace $\mathbf{X}_b$ by $\mathbf{X}_{bNew}$ and calculate the next dimension, i.e. $r = r + 1$, go to step (8).
18. Stop where maximum number of components have been calculated or when desired number of principal components have been calculated.

The first part of the MBPCA algorithm is similar to the CPCA algorithm in that standard PCA can be applied to calculate the block scores and loadings. The major difference is at the deflation stage where the block scores are used to deflate the residual matrices to obtain orthogonal block scores for MBPCA but super scores are used for deflation in CPCA to obtain orthogonal super scores.

5.2.6 Applications of Multi-block Techniques to Process Monitoring

The first multi-block application, reported by Wold *et al.* (1987b), applied CPCA for the analysis of sensory quality data for a number of wines. The quality characteristics of the wines as identified by a group of tasters such as body, bitterness and colour for each taster were placed in individual blocks and were represented by the block scores. The consensus of the tasters was captured by the super scores and the super weights explain the relative importance of each taster.

Wangen and Kowalski (1988) proposed an algorithm for dealing with complex chemical systems. The objective was to deal with applications where the blocks are in parallel but which may at the same time be connected in series. A schematic of the process demonstrated in the paper is presented in Figure 5-5.

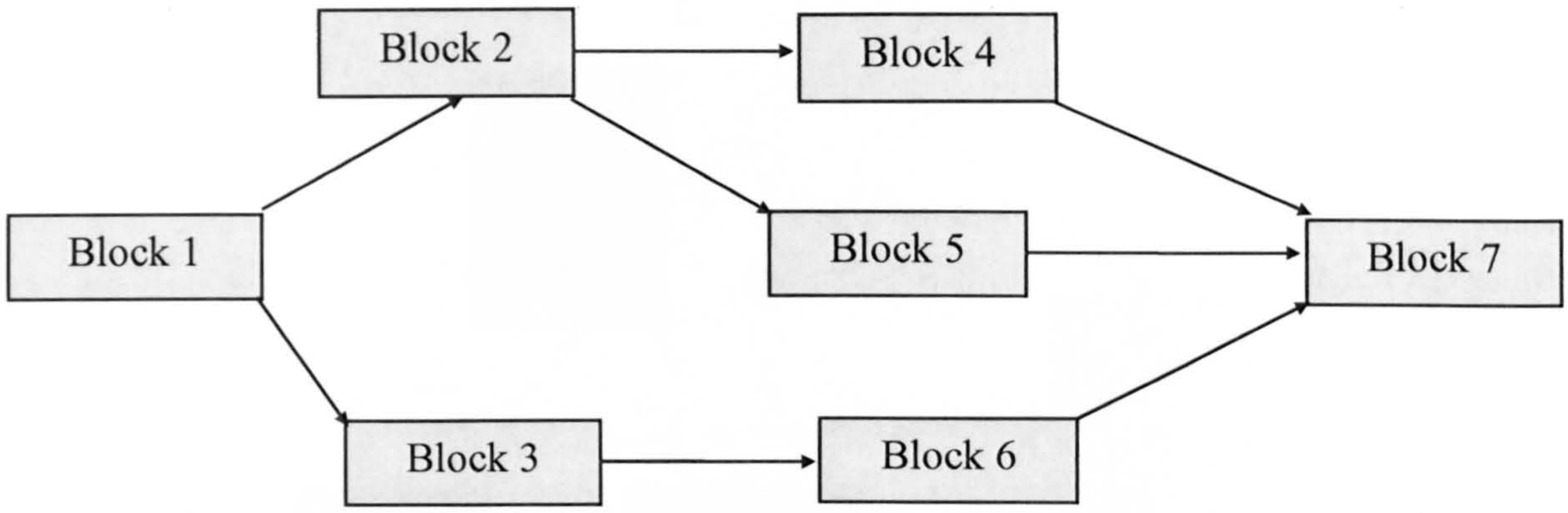


Figure 5-5 An illustration of the complex chemical system (Wangen and Kowalski, 1988)

The algorithm assumes that the process moves from left to right, with block 1 being used for prediction whilst block 7 is the quantity to be predicted. The interior blocks, blocks 2 to 6, are both to be predicted and also be used for prediction. If a block predicts more than one block (e.g. block 2), the scores vector of the predicted blocks (blocks 4 and 5) are combined to form a new block. Two-block PLS regression is then applied between the predictor block and the new combined block. Similarly the scores vectors for these blocks predicting a specific block (e.g. block 4, 5 and 6) are combined into a new block and PLS regression is applied between the new combined block and the predicted block 7. The algorithm can be viewed as an "alternative direction" approach where it cycles from right-to-left (backward phase) to calculate the **t** scores then cycles from left-to-right (forward phase) to calculate the **u** scores. The algorithm was demonstrated through an application using simulated data but no details of the simulation were presented.

MacGregor *et al.* (1994) reported the first on-line industrial application of multi-block PLS (MBPLS). The study considered low-density polyethylene produced at high pressure in a two-zone continuous tubular reactor. The process variables including temperatures and flow rates were divided into two blocks, one for each subsection of the reactor. Each process block was then regressed against a single block of quality variables, **Y**, using PLS. The resultant latent variable scores, **T**, were then used to construct the individual monitoring charts for each of the separate blocks. An overall monitoring chart was also constructed by combining the two sets of latent variable scores into a single consensus scores matrix. PLS was then applied to the consensus scores matrix which was used to construct the overall monitoring chart. A schematic of the two-block algorithm is presented in Figure 5-6.

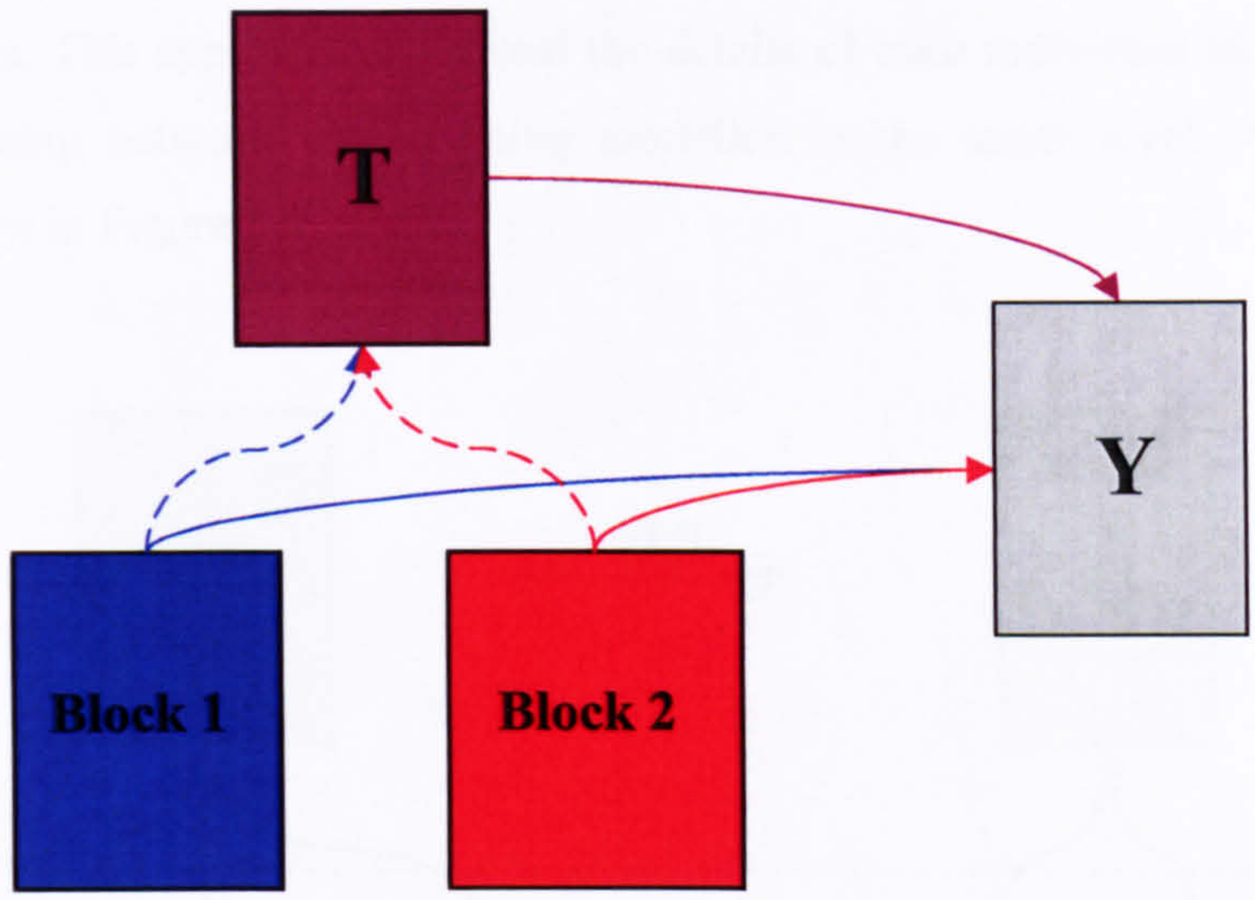


Figure 5-6 An illustration of MBPLS as described by MacGregor et al. (1994)

Although there are no specific rules for determining the block structure of a process, the multi-block technique is applied in situations where the process can be naturally blocked into sections. MacGregor *et al*. (1994) suggested that blocks should correspond to distinct units of the process and that there should be minimal correlation between the individual process units. However, in many industrial situations, the blocks may be joined in series and the upstream blocks may influence the behaviour of downstream blocks. At the same time, the downstream blocks may be better related to the final product quality therefore the blocking of variables is application specific.

Another MBPLS application was reported by Westerhuis and Coenegracht (1997). They applied the methodology to a two-stage wet granulation and tableting pharmaceutical process. The objective of the study was to examine the influence that the two groups of variables had on the final quality of the tablets. The first block comprised process variables from both stages and the composition of the powder mixture whilst the physical properties of the granulates were placed in a second block. The advantage of MBPLS was demonstrated for this application since improved interpretability resulted through the separation of events that related to individual block. The capability to focus specifically an individual blocks provided additional information over ordinary PLS.

Wold *et al*. (1996) demonstrated the application of hierarchical MBPLS to a residue catalytic cracker. Hierarchical MBPLS differs from the MBPLS algorithms in that the scores are calculated for individual quality blocks ($\mathbf{Y}_1$, $\mathbf{Y}_2$) as well as the individual process blocks ($\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_3$). The

individual scores are then combined into two super score blocks, one for the process scores (**T**), and one for the quality scores (**U**). A super level PLS is then performed on the combined process and quality blocks. This approach can reveal the details of each individual block at the sub level with the relationship between blocks being modelled at the super level. A schematic of the algorithm is shown in Figure 5-7.

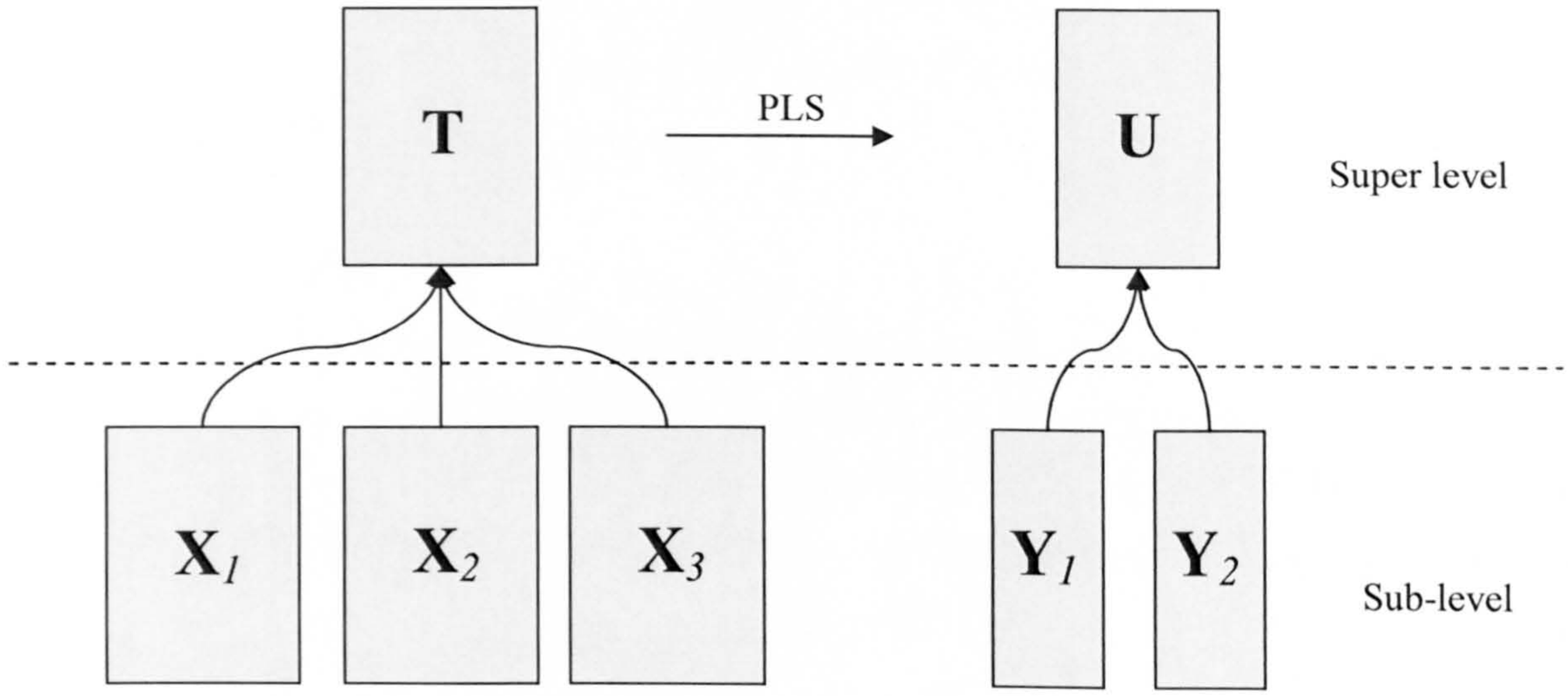


Figure 5-7 Schematic scheme of hierarchical MBPLS

Janné *et al.* (2001) applied hierarchical MBPLS to spectral data as a pre-treatment method. Spectral data typically comprises several hundreds of highly correlated variables that can be affected by light scattering. Data pre-treatment methods such as multiplicative scatter correction (see Section 5.4.4) can be applied to reduce the scattering effect. However in this paper hierarchical PCA / PLS was applied as an alternative data pre-treatment method where the spectral data corresponding to different substituents, e.g. carbonyl, amines and phenyl etc. were divided into several blocks. The authors concluded that the hierarchical pre-treatment method gave fewer components in the model and the ability to focus on specific spectral regions to enable more detailed analysis. The main advantage of multi-block techniques is the resulting ease of interpretation of the analysis.

More recently, an application of multi-block PCA was reported on a sequencing batch wastewater treatment process (Lee and Vanrolleghem, 2002). The process data considered was non-linear, time-varying and subject to significant disturbances from process operation. It is also a natural multi-phase batch process therefore multiblock PCA was applicable. Each phase of the process defines a block and consequently a local model approach can capture specific features within a

phase. The authors demonstrated the feasibility and effectiveness of multi-block analysis to detect local faults and the simplicity of data interpretation.

Another application using multi-block analysis was in the pharmaceutical industry (Lopes *et al.*, 2002). The application was the modelling of the production of an active pharmaceutical ingredient (API) in which inoculum and fermentation are two important stages known to affect the final API concentration. It was also found that inoculum growth is highly related to the fermentation productivity consequently a multi-block PLS model was built on the process stages to infer final API concentration.

5.2.7 Discussion

From the preceding discussion of applications reported in the literature, it was recognised that multi-block analysis generally enhances interpretability and understanding of the process compared with conventional unblocked model. More specifically by structuring the complex data into blocks and analysing it at different levels, a flexible framework is created whereby it is possible to focus on the major trends between blocks at the super level or to explore the details in a block at the sub level.

Wold *et al.* (1996) also demonstrated that multi-block models provide better predictive ability than models developed on unblocked data assuming that the data can be divided into conceptually meaningful blocks. Furthermore Eriksson *et al.* (2006b) considered the case where the data was divided into 4 blocks. The issue was that block $\mathbf{X}_3$ had a high proportion of missing samples (Figure 5-8) and if this data is treated as a whole, the samples would have had to be removed from the data set reducing significantly its overall size. However, by blocking its data, the samples missing in $\mathbf{X}_3$ can be retained in the other blocks although the issue of missing data in block $\mathbf{X}_3$ still exists at the super level.

Multi-block analysis can also be considered as an alternative to block scaling. Auto-scaling is most commonly used but a limitation of this approach is that it does not consider whether the variables are naturally grouped. With multi-block models, each block of data at the base level should have approximately the same influence on the higher level assuming an appropriate weighting is applied to individual blocks.

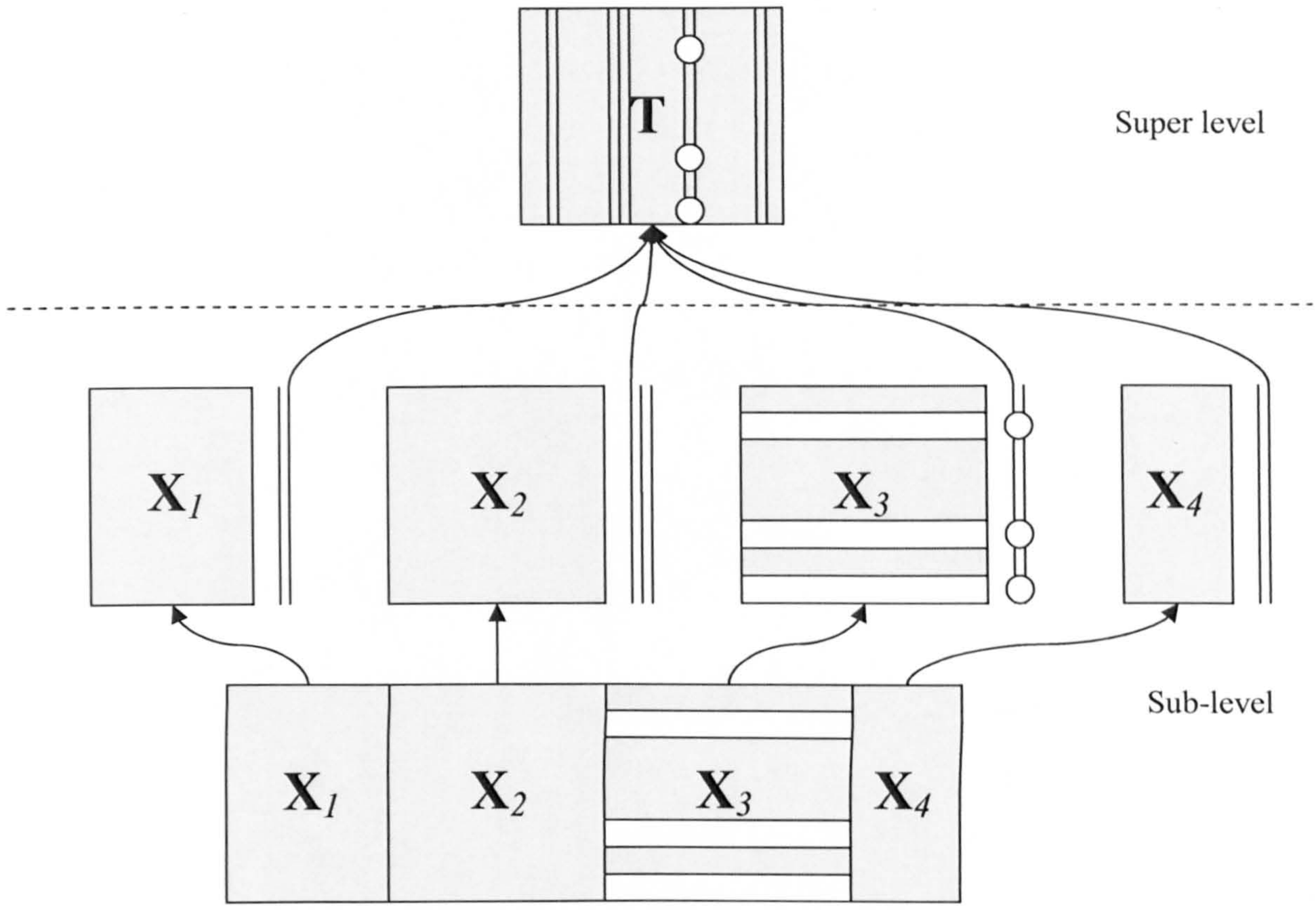


Figure 5-8 Illustration of a large matrix with missing data in block $\mathbf{X}_3$ (Eriksson et al., 2006b)

5.3 The Wavelet Transform and Its Applications

5.3.1 Introduction

Wavelets have attracted significant interest across the mathematics, signal processing and engineering communities. As pointed out by Mallat (1998), *"Wavelets are based not on a 'bright new idea', but on concepts that already existed under various forms in many different fields. The formalisation and emergence of this 'wavelet theory' is the result of a multi-disciplinary effort that brought together mathematicians, physicists and engineers, who recognised that they were independently developing similar ideas. For signal processing, this connection has created a flow of ideas that goes well beyond the construction of new bases or transforms"*. In the area of process monitoring, wavelet technologies have provided new tools and algorithms for the development of process representations. The focus of this section is to provide an introduction to the wavelet transform and describe its application in batch process monitoring.

A brief description of the Fourier transform and short-time Fourier transform is first given prior to defining the wavelet transform. A number of wavelet transforms and their reconstructions are introduced before discussing their properties. Following this specific examples of wavelet families are discussed. The theory of multiresolution analysis is then described since it forms the key theoretical basis of the wavelet transform. Finally, applications of the wavelet transform for batch process monitoring are reviewed.

The book "The World According to Wavelets" by Hubbard (1995) provides a good introduction to wavelets and requires little knowledge of mathematics. Daubechies (1992) describes the basic theory of wavelets whilst more technical books on wavelets include (Meyer, 1992; Meyer 1993). Additionally, there are some excellent tutorials on the wavelet transform including (Bentley and McDonnell, 1994; Graps, 1995; Alsberg *et al.*, 1997).

5.3.2 The Fourier Transform and Short-Time Fourier Transform

The behaviour of many signals can be studied either in the time domain or the frequency domain by applying appropriate mathematical techniques. For example the Fourier transform can decompose a signal into constituent sinusoids of different frequencies, i.e. it reduces the signal from a time-base to a frequency-base. Fourier transforms are appropriate for the analysis of stationary signals that do not develop with time. In an effort to address this limitation, the short-time Fourier transform can be utilised as it enables the signal to be analysed in the time-frequency domain.

The Fourier transform and its inverse establish a one-to-one relationship between the time domain function, $f(t)$, and the frequency domain, $F(\omega)$. The Fourier transform is defined as:

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{j\omega t} dt \,. \qquad 5\text{-}16$$

Using the identity

$$e^{j\omega t} = \cos k\theta + j \sin k\theta \qquad 5\text{-}17$$

the Fourier transform can be viewed as the decomposition of a function $f(t)$ into a sum of frequency components, the coefficients of which are given by the inner product of $f(t)$ and $e^{j\omega t}$. This transformation uses sines and cosines as basis functions for the mapping between the time domain and the frequency domain. The time independence of these basis functions result in a signal description purely in the frequency domain. The resulting spectrum $F(\omega)$ therefore

describes the overall energy of any frequency contained in the function *f(t)*. However, the spectrum does not give any information about the time location of the different frequency components.

For the analysis of non-stationary signals, a mathematical function is required to transform a signal into the time-frequency domain. A short-time Fourier transform can address such an analysis and is defined as:

$$F(\tau,\omega)=\int_{-\infty}^{\infty} f(t)g(t-\tau)e^{-j\omega t}dt\,. \qquad 5\text{-}18$$

The short-time Fourier transform can be viewed as the Fourier transform of the signal $f(t)g(t-\tau)$, which is the signal *f(t)*, windowed by the function *g(t)* about the time. Thus the short-time Fourier transform performs a linear mapping from a time domain function, *f(t)*, onto a time-frequency domain function, $F(\tau, \omega)$. However once the size of the time window is selected, the window is fixed for all frequencies and thus this approach lacks flexibility.

5.3.3 The Wavelet Transform

The fundamental advantage of wavelet analysis is its ability to process a signal at different scales so that both the global features and the localised details of a signal can be studied simultaneously. Wavelet analysis is capable of revealing aspects of the data that other signal analysis techniques may not identify including trends, breakdown points and self-similarity.

5.3.3.1 Definition of the Wavelet Transform

The wavelet transform can be used to represent the original signal, or function, as a linear combination of the wavelet functions. Thus the data can be represented using the corresponding wavelet coefficients. The wavelet transform can also be regarded as being the decomposition of a signal into a set of basis functions, i.e. wavelets. These wavelet functions are obtained from a single prototype wavelet, called the "mother wavelet", through dilation and translation operations.

The "mother wavelet" $\Psi(t)$ is a function whose Fourier transform $\Psi(\omega)$ satisfies the "admissibility condition" defined by:

$$\int_{-\infty}^{\infty}\frac{|\Psi(\omega)|^2}{\omega}d\omega=C_\Psi<\infty \qquad 5\text{-}19$$

where C_Ψ is a finite value. This condition implies that the signal has finite energy, i.e. does not contain any constant component:

$$\int_{-\infty}^{\infty} \Psi(t)dt = 0 . \qquad 5\text{-}20$$

This means that a wavelet must therefore be an oscillatory function with zero mean. For wavelets to be useful basis functions for analysis, the mother wavelet must posses certain properties:

- Smoothness: $\Psi(t)$ should be a smooth function with continuous derivatives.
- Good time localisation: $\Psi(t)$ together with its derivatives should decay rapidly.
- Good frequency localisation: $\Psi(\omega)$ should decay sufficiently fast as the frequency $\omega \to \infty$ and $\Psi(\omega)$ should be sufficiently flat near $\omega = 0$. The flatness at $\omega = 0$ is related to the number of vanishing moments of $\Psi(t)$. The l^{th} moment of a wavelet is defined as (Motard and Joseph, 1994):

$$\int_{-\infty}^{\infty} t^l \Psi(t)dt \qquad 5\text{-}21$$

and a wavelet is said to have M vanishing moments if

$$\int_{-\infty}^{\infty} t^l \Psi(t)dt = 0 \qquad \text{for } l = 0, 1 \ldots M \qquad 5\text{-}22$$

which is equivalent to

$$\left.\frac{d^l \Psi(\omega)}{d\omega}\right|_{\omega=0} = 0 \qquad \text{for } l = 0, 1 \ldots M \qquad 5\text{-}23$$

Wavelets with a large number of vanishing moments result in the frequency response of the wavelets being flatter when the frequency ω is small. These properties of wavelets along with the admissibility condition, infer that:

- Wavelets are band-pass filters since the frequency response decays sufficiently fast for a large frequency, ω, and is flat at $\omega = 0$

– $\Psi(t)$ is the impulse response of this filter which again decays sufficiently fast as t increases and is a zero-mean oscillatory function.

By performing scaling and translation operations on the mother wavelet $\Psi(t)$, a family of wavelet functions are created. They all have the same shape as the mother wavelet but differ in terms of size and position. They are denoted by:

$$\Psi_{a,b}(t)=\frac{1}{\sqrt{|a|}}\Psi\left(\frac{t-b}{a}\right) \qquad 5\text{-}24$$

where a is the scaling parameter and b is the translation parameter. The factor $\frac{1}{\sqrt{|a|}}$ is used to ensure that each wavelet function has the same energy as the mother wavelet.

The scaling operation can be seen as carrying out "stretching" or "compressing" operations on the mother wavelet. The resultant wavelets can be used to capture different frequency information with respect to the function being analysed. The "compressed" wavelet is used to fit the high frequency components, whilst the stretched wavelet is used to fit the low frequency components. The translation operation, involves "shifting" the mother wavelet along the time axis. This is used to capture the time information of the function being analysed.

5.3.3.2 The Continuous and Discrete Wavelet Transform

The selection of appropriate wavelet transforms is dependent on the type of input signal, dilation and translation parameters. The following section describes two popular transforms – the Continuous Wavelet Transform and the Discrete Wavelet Transform.

The Continuous Wavelet Transform

For the Continuous Wavelet Transform (CWT), the scaling parameter a and translation parameter b change continuously over time t. The CWT is defined as:

$$\mathrm{CWT}(a,b)=\left\langle f,\Psi_{a,b}\right\rangle=|a|^{-1/2}\int_{-\infty}^{\infty}f(t)\Psi^{*}\left(\frac{t-b}{a}\right)dt \qquad a,b\in R,a\neq 0 \qquad 5\text{-}25$$

where the asterisk denotes the complex conjugate and the notation $\langle,\rangle$ denotes the standard scalar product:

$$\langle f,g\rangle = \int_{-\infty}^{\infty} f(t)g^*(t)dt\,. \qquad 5\text{-}26$$

The CWT is a linear transform in the time-frequency domain which is similar to the short-time Fourier transform. Following the application of CWT, the CWT coefficients which are a function of time-shift and frequency (scale) parameters are determined. In a graphical representation, the *x*-axis usually represents the time-shift whilst the *y*-axis represents the scale. A log scale is typically used for the *y*-axis to enable a large range of scales to be accommodated. The lower values of scale correspond to the higher frequencies, whilst the higher values of scale correspond to the lower frequencies. The colour of the *x*-*y* block represents the magnitude of the CWT coefficients for different scales and time, according to the definition of the colour bar. Hence a time-scale coloration plot is obtained representing the time evolution from left to right and decreasing frequency from bottom to top. An example is shown in Figure 5-9 where two different frequencies of signal are plotted in Figure 5-9(a) and Figure 5-9(b) shows the coloration plot for the CWT coefficients.

The plot clearly shows that two different frequency elements exist at different scales. One can be seen at the higher scale and corresponds to the sinusoids at the lower frequency; the other is at the lower scale and corresponds to the sinusoids at the higher frequency.

The example reveals the flexibility and efficiency of the wavelet transform in the time-frequency domain. By providing short time windows in the high frequency region and long time windows in the low frequency region, both local behaviour and global features of the signal can be extracted.

The Discrete Wavelet Transform

For the discrete signal, the translation parameter *b* is proportional to the scaling parameter *a*:

$$a = a_0^m \text{ and } b = nb_0a_0^m\,. \qquad 5\text{-}27$$

The family of wavelets $\{\Psi_{m,n}(t)\}$ can then be defined as:

$$\Psi_{m,n}(t) = a_0^{-m/2}\Psi(a_0^{-m}t - nb_0) \text{ and } m,n \in Z\,. \qquad 5\text{-}28$$

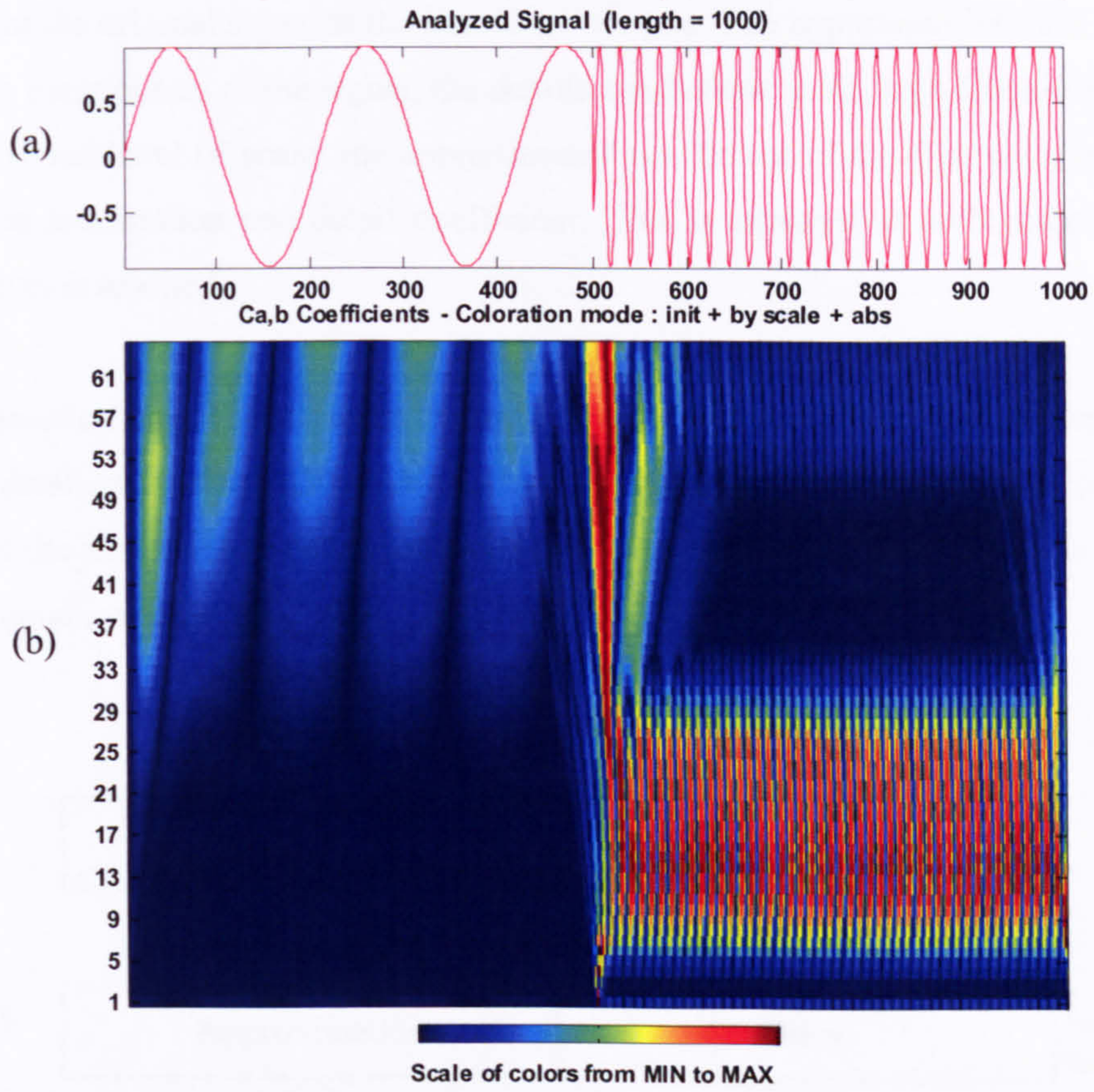


Figure 5-9 Example signal of sine curve and its continuous wavelet transform coefficient plot

The Discrete Wavelet Transform (DWT) is then given as follows:

$$(\mathrm{DWT})_{mn} = \langle f, \Psi_{mn} \rangle = a_0^{-m/2} \int_{-\infty}^{\infty} f(t)\Psi(a_0^{-m}t - nb_0)dt \,. \qquad 5\text{-}29$$

As a_0 tends to 1 and b_0 tends to 0, the discrete form approaches the continuous form. The discrete wavelet transform can be classified according to the type of wavelets used for the transformation. When orthonormal wavelets are used, the family of wavelets are linearly independent and form an orthonormal basis (Daubechies, 1988; Mallat, 1989a). In this case, the discrete wavelet transform becomes the orthonormal wavelet transform which is the form of wavelet considered in the thesis.

5.3.3.3 Multiresolution Analysis

Mallat (1989b) and Meyer's theory of Multiresolution Analysis (MRA) provides a systematic approach to constructing the orthonormal wavelet basis. The basic idea behind MRA is shown in Figure 5-10. It has two operating modes: firstly the decomposition of a signal then its

reconstruction. In the signal decomposition stage, an approximation and a detail coefficient are computed from the original signal at the first level of scale. The approximations are the high scale, low frequency components of the signal; the details are the low scale, high frequency components. Then at the second level of scale, the approximated coefficient of the first scale is decomposed into another approximation and detail coefficient. This is repeated at each scale until the pre-determined level is reached.

In the reconstruction stage the steps are reversed. For the last scale, the approximate coefficient is added to the detail coefficient. This results in a finer signal approximation of the next scale. The coefficients at the next level of detail coefficient are summarised and this process is repeated until the original signal is recovered.

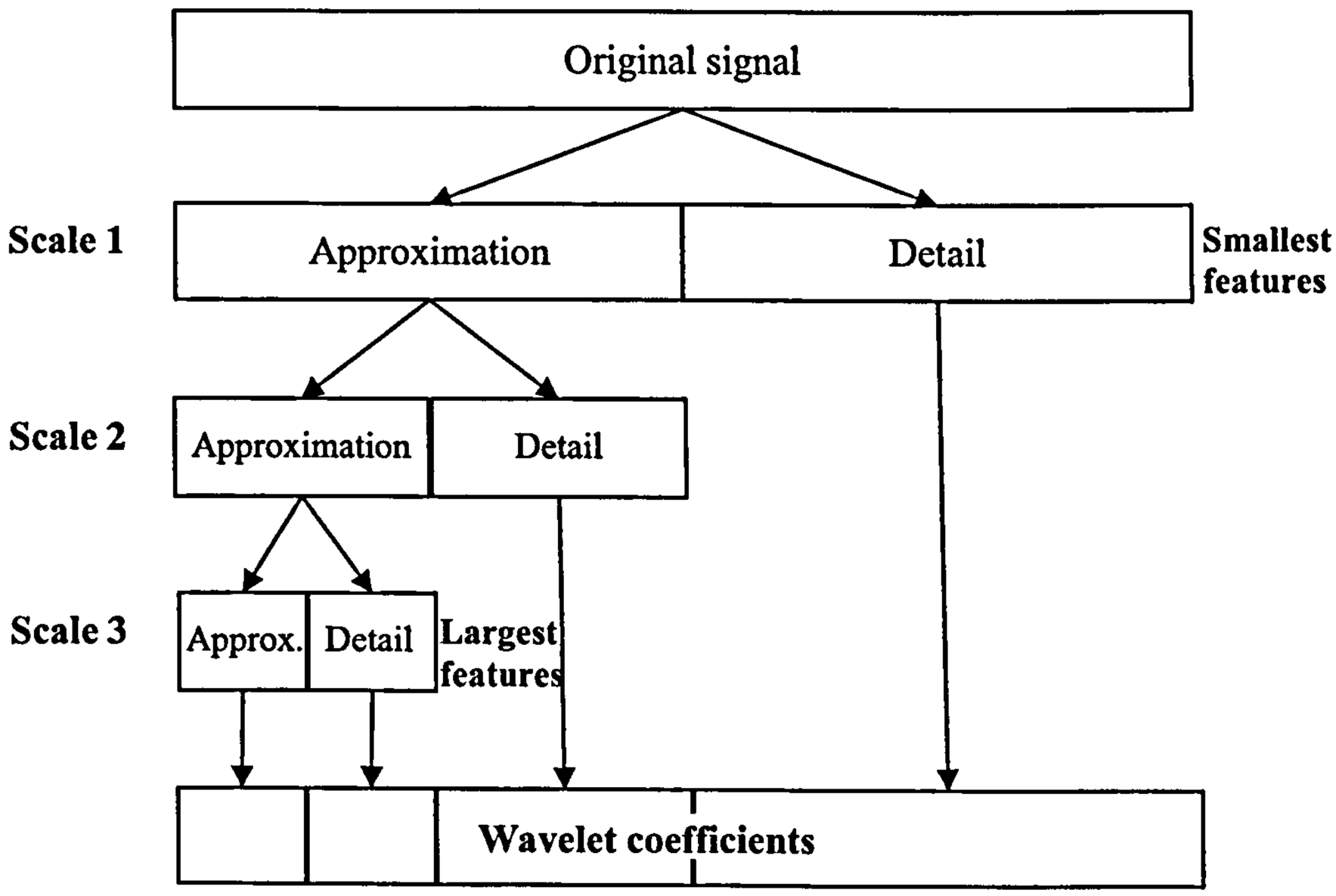


Figure 5-10 Process map of multiresolution analysis

One of the conditions for the application of MRA is a signal of dyadic length, i.e. 2^n where n is an integer. If the signal is not dyadic, padding of the original signal to the nearest dyadic length can be applied. One method is to pad the signal with zeros. However, this can cause unnecessary edge effects. According to Trygg and Wold (1998), if padding is done by linear padding, minimal edge effects are introduced. Linear padding is where a vector of linearly computed values are applied, starting with the last value of the original vector and ending with the first value. Teppola and Minkkinen (2000) considered two cases in terms of padding. The first case is related to the first

value of the signal and the same value is utilised whilst the second group is associated with the last value of the signal and this time the same last value of the signal is utilised.

5.3.3.4 Properties and Examples of Wavelets

Some important properties associated with wavelets include: smoothness, the number of vanishing moments, symmetry, orthogonality and time-frequency localisation (Motard and Joseph, 1994).

Smoothness

The smoothness of a wavelet is determined by the number of derivatives which exist. When approximating a function, it is recommended that the smoothness of the wavelet and the function should match reasonably well thus a more compact representation is obtained.

Vanishing Moments

The number of vanishing moments, M, is defined by Equation 5.23 and is weakly linked to the number of oscillations (Hubbard, 1995). The more vanishing moments a wavelet has, the greater the oscillation. Wavelets with a larger number of vanishing moments tend to enhance the sparseness of the resulting wavelet coefficient matrix. This is because the number of vanishing moments determines what the wavelet does not detect. For example a wavelet with one vanishing moment does not detect linear functions in a signal, as the wavelet coefficients of the linear function are zero. A wavelet with two vanishing moments does not detect quadratic functions either. As a result, the resultant wavelet coefficients would be sparse, and hence compression would be easier.

Symmetry

Symmetric wavelets are symmetric in shape and therefore do not distort the phase information of the transformed signal. Compactly supported wavelets are non-zero over a finite interval, either in the time and frequency domain.

Orthogonality

The discrete wavelet transform can be classified into orthogonal and non-orthogonal wavelets. Only orthogonal wavelets are considered in the thesis as they are linearly independent. The advantage of orthogonal wavelets is that no redundancy remains in the wavelet representation and exact reconstruction of the original signal can be achieved. Orthogonal wavelets include the Meyer wavelet, the Haar wavelet, the Battle-Lemarie wavelets and a class of orthonormal wavelets with compact support constructed by Daubechies. Of these wavelet families, the Haar wavelet is least useful in practice since it has poor frequency localisation. The Daubechies

wavelets are commonly used because they have good time-frequency localisation and compact support.

In practice, it is not possible to construct wavelets with an arbitrary combination of all the above properties. For instance, a wavelet with an infinite number of derivatives and compact support in both the time and frequency domain cannot be constructed. Similarly, it is not possible to construct a smooth, compactly supported, symmetric and orthogonal real-valued wavelet. The choice of wavelet is usually made by selecting the appropriate properties for the required application.

A number of wavelets with different property combinations are available in the literature. Four examples of orthogonal wavelet families are shown in Figure 5-11.

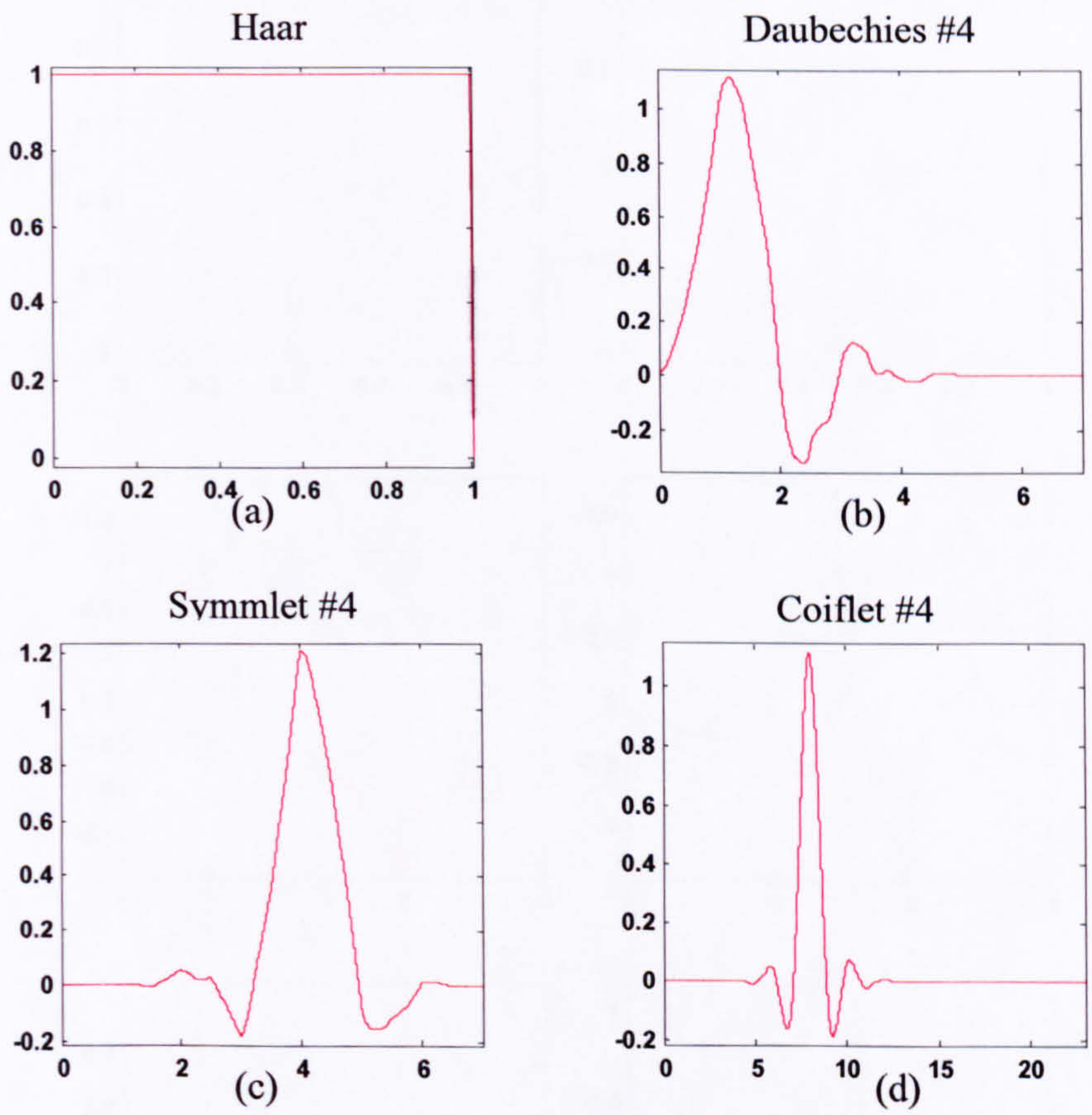


Figure 5-11 Scaling functions of some example orthogonal wavelets

The scaling function of the Haar wavelet is illustrated in Figure 5-11(a). The Haar wavelet is the simplest orthogonal wavelet which features compact support, symmetry and orthogonality at the expense of frequency localisation. Figure 5-11(b) shows the scaling function for one of the Daubechies families which has the properties of compact support and orthonormal wavelets with

4 vanishing moments. Figure 5-11(c) illustrates the scaling function for the symmlet wavelet with 4 vanishing moments. The properties of this family are similar to Daubechies but less asymmetric. Finally, Figure 5-11(d) shows the scaling function of the coiflet wavelet with 4 vanishing moments.

One of the most widely used families of wavelets are those derived by Daubechies (1988). These wavelets are orthonormal, compactly supported and have a maximum number of vanishing moments for the support. Selecting different numbers of vanishing moments results in scaling functions and wavelet functions with different degrees of smoothness. Examples of Daubechies' wavelets and their corresponding scaling functions with different number of vanishing moments are shown in Figure 5-12.

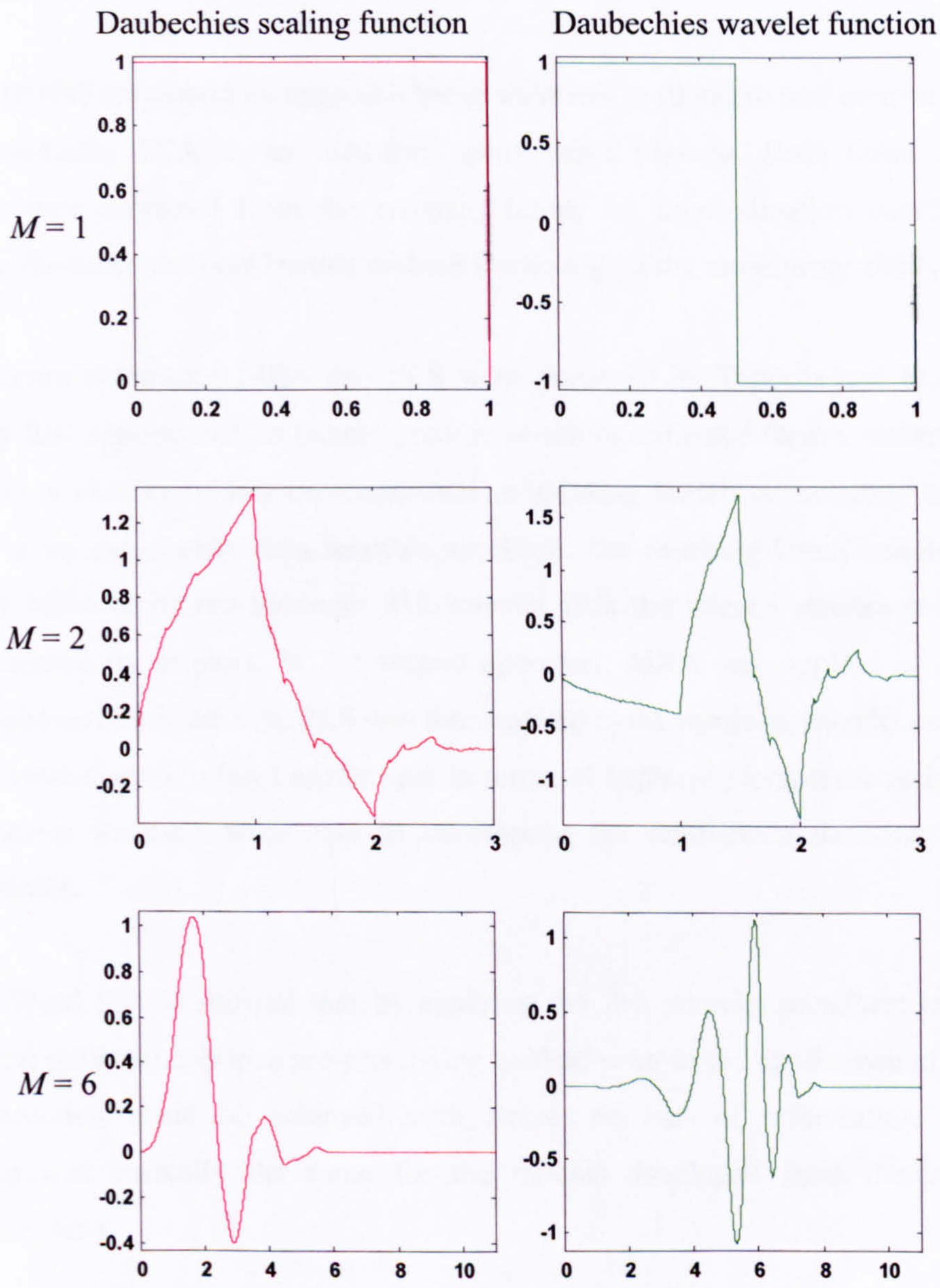


Figure 5-12 Daubechies' wavelets family with different vanishing moments

5.3.4 Applications of the Wavelet Transform to Process Monitoring

Wavelets can be used to assist in the development of process monitoring models. Kosanovich and Piovoso (1997) applied wavelets as a Finite Impulse Response median filter (FIR) to pre-filter the signals and then used wavelet functions for the decomposition of the filtered signal to the time-frequency domain to address the noise and systematic biases in the process data. This particular median filter was used since it offers reduced computational complexity over other median filters. The filtered values are then decomposed, using the Haar wavelet transform, into low and high frequency components to isolate process events in the wavelet domain rather than the time domain. The resulting Haar wavelet coefficients are then combined with PCA to monitor an industrial process which has 24 monitored variables. Improved sensitivity with respect to fault classification was achieved.

Shao *et al.* (1999) developed an approach using wavelets to de-noise and remove spikes prior to applying non-linear PCA to an industrial spray dryer process. Both linear and non-linear correlations were extracted from the analysis. Using the approximation coefficients reduced significantly the computational burden without impacting on the monitoring ability.

Two approaches combining MRA and PLS were proposed by Teppola and Minkkinen (2000, 2001). They first applied PLS to twenty process variables collected from a wastewater treatment plant to infer a number of key environmental monitoring variables including temperature and humidity. As an exploratory data analysis approach, the resulting latent variable scores were analysed by MRA using the Symmlet #10 wavelet with the latent variables at different scales being represented by bi-plots. In the second approach, MRA was applied as a pre-treatment method to the two blocks of data. PLS was then applied to the resulting wavelet coefficients. Both approaches were shown to have advantages in terms of capturing long-term variations and local deviations since wavelets were able to decompose the multi-scale data into high and low frequency scales.

Trygg and Wold (1998) showed that by applying the fast wavelet transform to the individual Near-Infrared (NIR) signals as a pre-processing method prior to the application of PLS modelling, good compression could be achieved with almost no loss of information. The predictive performance was basically the same for the models developed from the compressed and uncompressed data.

Trygg *et al.* (2001) have also applied the wavelet transform to monitor a wood chip manufacturing process using on-line NIR. The challenges of monitoring on-line in this situation included sampling, significant moisture variations, temperature variations and a moving conveyer belt. In particular, a temperature increase is known to produce spectral shifts in signals. Wavelets were shown to be able to detect such shifts when compared to PCA due to their ability to separate frequency bands. Again no loss of information was recorded when performing wavelet data compression.

An important application of wavelets was that of Multi-scale Principal Component Analysis (MSPCA) (Bakshi, 1998 and 1999). MSPCA is basically the application of PCA to the wavelet coefficients at each scale and then those loadings that satisfy a pre-determined threshold value are selected for subsequent analysis. This approach was shown to enable greater data compressibility and the decomposition of multi-scale features of a signal since nearly all signals were inherently multi-scale in nature. Several authors have extended the methodology to handle different situations (Rosen and Lennox, 2001; Misra *et al.*, 2002).

5.4 Batch Process Monitoring Using Spectral Data

5.4.1 Introduction

New on-line measurement technologies such as spectroscopy are particularly useful as a means of obtaining real time, high quality chemically rich information. When comparing spectroscopy to chromatography for the on-line monitoring of batch processes, it offers several advantages including speed, robustness and a non-destructive route, giving direct chemical insight into the spectroscopically active compounds present in a system. Several spectroscopic techniques are being developed for on-line process implementation, for example, Near-Infrared (NIR), Mid-Infrared (MIR), Ultra-violet Visible (UV-Visible) and Raman. Each technique has its own characteristics and process specific applications. These are discussed in more detail in the subsequent section in the context of how they can be used for real time batch process monitoring and integrated within the framework of MSPC. A detailed summary of on-line spectroscopic monitoring techniques are given in the review of Workman *et al.* (1999). The review provided information on chromatography techniques, infrared spectroscopy and imaging techniques, Raman and electronic spectroscopy (UV-Visible and fluorescence), mass spectroscopy, process chemometrics, flow and sequential injection analysis and ultrasonic analysis.

5.4.2 Considerations of On-line Spectral Data Process Analysis

On-line process analysis has to date provide qualitative and quantitative information about a chemical batch process in real time to help realise higher quality and more consistent product quality. It has typically focused on physical process parameters including temperature, pressure, flow rate, percentage of reagent added and agitator rate. Measurements from these sensors are normally measured and recorded automatically at regular time intervals by a process control system and are traditionally characterised by being cheaper and easier to implement than other sensors such as analytical devices and also require less rigorous calibration. This type of data describes the physical state of a chemical reaction or the physical interaction of liquids, solids and / or gas. Consequently, they play an important role in process design and development, process optimisation, process safety and control.

Complementary to physical measurements are those that provide information on the chemical properties. Thus for a comprehensive monitoring scheme, the two forms of information is important. Spectroscopy measures directly the specific chemical properties of the molecules thus the monitoring of concentration changes of reactants and the formation of products is viable. Additionally the onset of the presence of impurities and catalyst activities can also be tracked. Some of the features of spectral data are summarised below:

- Molecular spectroscopy deals with the interactions of molecular species with electromagnetic radiation, i.e. it is the measurement and interpretation of electromagnetic radiation absorbed, scattered or emitted by a chemical species (McLennan and Kowalski, 1995). The on-line spectra describe the chemical state of a process which relate to the molecular nature and bonding, for example, to measure the absolute concentrations of the target molecule in the process.
- It is common to employ one form of spectroscopy for a specific unit operation therefore the variables are usually measured in the same unit (e.g. absorbance). With the increasing importance of advanced measurement technologies, more than one form of spectroscopy can be implemented into a unit operation for multi-purpose monitoring. This may require more advanced spectral pre-processing techniques to cope with the large amount of data.
- Despite the complexity of batch operation, spectroscopic measurement consist of a linear contribution from each absorbing species according to the Lambert-Beer law which is a mathematical means of expressing how light is absorbed by matter:

$$A = a_{LB} \cdot b_{LB} \cdot c_{LB} \qquad 5\text{-}30$$

where A is the absorbance, a_{LB} is the molar absorptivity of the analyte for the specific wavelength, b_{LB} is the path length in cm and c_{LB} is the molar concentration. This linear characterisation of the data is highly suitable for analysis.

- Spectral data typically contains a high level of systematic variation hence the quality of data is usually high since the level of process noise is reduced. However, some physical conditions within a process can limit spectral acquisition and the quality of data. For example when operating the instrument at high environmental temperature.

Spectroscopy has proven its usefulness for the identification of chemical species in the process analytical laboratory. However when used for the on-line monitoring of batch processes, there are several potential issues that require to be considered compared with the laboratory based off-line analysis. Additionally, laboratory based instruments are rarely utilised in a plant environment as they require to withstand more harsh production conditions. Consequently, some of the main challenges of using spectroscopy for on-line batch process monitoring include:

- Selecting the appropriate spectroscopic technique is key to success. The form of spectroscopy has to be fit for purpose and off-line calibration and testing are usually required before considering on-line usage.
- The placement of the sensor / probe of the selected spectroscopic method has to be in the most appropriate part of the process to obtain a meaningful and informative result. The appropriateness of such a sampling location has to be designed carefully.
- In comparison to physical process sensors such as flow meters and temperature sensors, spectroscopic instrumentation is generally more sensitive to changes in physical process conditions such as temperature fluctuations, variations in input material, background noise and vibration. This can impact on the quality of the measured spectra.
- Spectral effects including light scattering and artefacts of spectrometer behaviour such as baseline and wavelength shifts can affect the quality of the measured spectra. Model robustness is important for the successful inference of the product concentration, for example, for monitoring applications. A range of mathematical pre-processing methods are available to remove undesired variation.
- When using a spectrometer in a highly regulated manufacturing environment, safety is crucial and different levels of certification are required.

5.4.3 Process Spectroscopy for On-line Analysis

Spectroscopy is the science of the interaction of electromagnetic radiation with matter. The output spectrum displays the intensity of radiation absorbed by a sample versus a quantity related to

photon energy, such as wavelength or frequency. Different spectroscopic techniques that can be implemented on-line include NIR, MIR, UV-Visible and Raman. These methods together with the spectral data pre-treatment methods are discussed in the next sections. Bakeev (2005) provides an overview of the different spectroscopic tools and the implementation strategies for on-line spectroscopy.

Mid-Infrared (MIR)

Mid-Infrared spectroscopy is one of the most versatile techniques available for the measurement of molecular species. The operational wavelength range of IR is from 12000 to 50 cm^{-1} (800 to 200 000 nm) with the MIR region lying in the range 4000 to 625 cm^{-1} (25 000 to 160 000 nm). This region covers the fundamental vibrations of most of the common chemical bonds and organic compounds. One of the biggest advantages of MIR is its broad applicability to a multitude of applications based on the information content of the MIR spectrum (Ruckebusch *et al.*, 2000; Van Sprang *et al.*, 2003). There are also sensitivity advantages compared with NIR in that it is as much as hundred times more sensitive.

Near-Infrared (NIR)

NIR can be considered to capture the chemical behaviour of the hydrogen atom in its various molecular manifestations. The frequency range of NIR is from approximately 4000 cm^{-1} up to 12500 cm^{-1} (800 – 2500 nm) and covers mainly the overtones and combinations of the lower energy fundamental molecular vibrations that include at least one hydrogen bond vibration. The weaker absorption compared with the fundamental vibrational bands (i.e. MIR) decreases in intensity range from between 10 – 100 times that of the original band. However, NIR is less sensitive to impurities in the sample. As a consequence of the reduced absorbency, the NIR beam is able to penetrate deeper into the test mixture, providing a more representative analysis. This increases the potential for the use of NIR in on-line applications.

Ultra-violet Visible (UV-Visible)

The UV-Visible spectrum is generated through absorption and fluorescence measurements in the UV and visible wavelength regions. The UV region normally studied is 200 to 400 nm and the visible region is 400 to 800 nm. Many organic compounds and some inorganic compounds will absorb in the UV-Visible region and normally broad spectral bands are produced which result in overlapping spectra of compounds in mixtures. The detection limits for analysis by UV-Visible spectrometry are lower than those of MIR and NIR spectrometry, giving more sensitive determinations.

Raman

Raman spectroscopy looks at the fundamental vibrational information contained in visible to NIR wavelength light scattered by a sample. This allows the use of very long runs of optical fibre cable, which allows the instrument to be located on the plant with relative ease and for samples to be taken from points that are physically located far from each other. Raman can detect the gain or loss energy due to Rayleigh scattering owing to interactions with energy levels involved with vibrational and rotational transitions. Vibrations in molecules are Raman active if there is a change in polarisation. This occurs during symmetrical vibrations, unlike IR where only unsymmetrical vibrations are active. IR requires a change of dipole moment. However, Raman spectroscopy is not a very sensitive technique. Compounds present below the 0.1% to 1% range are often not detected in typical samples.

5.4.4 Spectral Data Pre-treatment

As explained in Section 3.2, most data requires mathematical pre-treatment prior to any statistical analysis to remove artefacts in the data such as noise, outliers and non-linear behaviour. With respect to spectral data, the artefacts include base-line drift, multiplicative scatter effects and noise. Although such pre-treatments may result in improved model performance, it is important to understand the inherent assumptions of these pre-treatment methods. A number of researchers have evaluated the different pre-processing techniques and the consensus was that the pre-processing method of spectral data is important for information extraction, however, with regard to which is the most appropriate method for a particular spectra type, there is no consensus (Wold *et al.*, 1998; Eriksson *et al.*, 2000; Azzouz *et al.*, 2003 and Zeaiter *et al.*, 2005). A number of the key methods are summarised below.

Standard Normal Variate (SNV)

The SNV method (Barnes *et al.*, 1989) reduces the multiplicative effects caused by differences in sample path length. Each sample spectrum is corrected by the mean of all the wavelengths with the multiplicative adjustment being the standard deviation of the values over all wavelengths such that:

$$\mathbf{x}_{corr} = \frac{\mathbf{x}_i - \bar{\mathbf{x}}_i}{\sqrt{\frac{\sum_{i=1}^{J}(\mathbf{x}_i - \bar{\mathbf{x}}_i)^2}{J-1}}} \qquad i = 1, 2 \ldots J \qquad \text{5-31}$$

where $\mathbf{x}_i$ is the ith sample spectrum, $\bar{\mathbf{x}}_i$ is the mean over all J-wavelengths, for sample i and $\mathbf{x}_{corr}$ is the standard normal variate of the spectral data $\mathbf{x}_i$.

A limitation of this method is that the multiplicative effects are assumed to be uniform across the whole spectral range which is not always the case thus the underlying spectral information is distorted.

Multiplicative Signal Correction (MSC)

MSC reduces the shift in the spectra relative to each other by regressing them either against a chosen reference spectrum or against a calculated mean spectrum, which is then used as the reference for correction. The MSC method (Martens and Næs, 1989; Geladi *et al.*, 1985) has been effectively applied in many NIR diffuse reflectance applications where there are multiplicative variations between sample responses. It provides an estimate of the relationship of each sample scatter with respect to the scatter of a reference spectrum, thus the same level of scatter for all spectra is obtained. The MSC is modelled by the follow equation:

$$\mathbf{x} = a_{msc} \cdot \mathbf{x}_{ref} + b_{msc} + \mathbf{e} \qquad 5\text{-}32$$

where $\mathbf{x}$ is the sample spectrum, a_{msc} is the multiplicative correction factor, $\mathbf{x}_{ref}$ is a reference spectrum, b_{msc} is an additive correction factor and $\mathbf{e}$ is the residuals. For most applications, $\mathbf{x}_{ref}$ is the mean spectrum of the data set although this may not always be the case. The MSC correction factors a_{msc} and b_{msc} are estimated using least squares given the sample spectrum and reference spectrum. The corrected spectrum can then be calculated:

$$\mathbf{x}_{corr} = \frac{(\mathbf{x} - \hat{b}_{msc})}{\hat{a}_{msc}}. \qquad 5\text{-}33$$

This model assumes that any sample spectrum can be estimated as a multiple of the reference spectrum, i.e. the offset and multiplicative spectral effects are much larger than effects from changes in actual chemistry. As a result, this method can lead to poor modelling results since the chemical-based variations in the data are much greater than the correction factor variations.

Orthogonal Signal Correction (OSC)

The idea of OSC is to remove systematic information in the **X** matrix that is not correlated to the modelling of the **Y** responses to achieve better models (Wold *et al.*, 1998; Wold *et al.*, 2001). Spectra often contain systematic variation such as light scattering and differences in path length which is unrelated to the responses. OSC is a PLS-based method that removes the uncorrelated part of **X** from **Y**. The removed part is mathematically orthogonal to **Y**. The result of the

application of OSC is a model based on PLS components containing information about the correction of **X**. Then the regular PLS diagnostics such as scores and loadings are available to interpret exactly which kind of information was extracted from **X**.

Derivatives

Derivatives can be used to remove effects such as offset and background slope variations between samples given that the variables are expressed as a continuous physical property. This method is applicable to spectroscopic data as the variables are expressed as a continuous wavelength or wavenumber. A mathematical derivation of a function is used for such a correction and a discrete form of such a function, Savitsky-Golay (Gorry, 1990; Mark and Workman, 2003), can be used to calculate the derivatives. These filters are essentially local functions that are applied to each spectrum using a moving-window approach across the wavelength or wavenumber axis to evaluate the derivative. The use of such filters requires the specification of two parameters, those which define the width and resolution of the local functions.

In spectroscopy applications, a first derivative effectively removes baseline offsets in the spectral profiles. A second derivative results in the removal of both baseline offset differences between the spectra and differences in baseline slopes between spectra.

Derivatives can be applied directly to a single spectrum in contrast to other pre-treatment methods that require a set of data to be available.

Baseline Correction

This method is used to correct for a parallel baseline shift (Candolfi *et al.*, 1999) that may be due to variations in the sample presentation. Spectra are usually affected by the stability of the instrument, temperature and humidity. Baseline correction is achieved by subtracting a specific region from a reference spectrum that is relatively free of absorbance peaks through linear regression. In some cases, the mean absorbance of a number of spectra can be used for correction.

5.4.5 Applications of Batch Spectral Process Monitoring

An increasing number of batch process monitoring applications using on-line spectroscopy have been reported in the literature. Westerhuis *et al.* (2000b) presented an application of the on-line monitoring of a chemical batch reaction using UV-Visible spectroscopy. For the study, the MSPC philosophy was adopted with the behaviour of new batches being compared against the process signature based on good historical batches. It was also shown that by introducing some known deviations, a combination of control chart metrics such as the SPE and contribution plot were able

to detect such deviations successfully. This was one of the first reported applications and the authors concluded that the philosophy is promising however more applications needed to be investigated.

Batch process monitoring by spectroscopy using a grey modelling approach was proposed by Gurden *et al.* (2001). The authors defined a grey model as a model consisting of known sources of variation (usually described as "hard" or "white") and unknown sources of variation (usually described as "soft" or "black"). A first-order batch reaction was monitored by UV-Visible spectroscopy and thirty batches were generated under normal operating conditions. These batches defined the nominal model and a further nine non-conforming batches were used to test the monitoring and fault detection capability. The additional information incorporated into the model was claimed to improve model interpretability. Gurden *et al.* (2002) further explored such ideas and compared the ideas with the monitoring of batch processes using traditional engineering process variables. Although on-line analysers are more widely implemented in industry, there are still some industrial issues that affect model robustness. The approach of the grey model has the advantage of stabilising the model by improving interpretability with some known features of the process. The same data, as from the previous UV-Visible application, was used but new data with different types of process disturbances were investigated. It was concluded that such an approach provides good detection and diagnosis capability for a variety of process faults and they then proposed a number of future applications including the idea of combining spectroscopic and engineering process measurements.

Another application based on a different spectroscopy technique was reported by Van Sprang *et al.* (2003). MIR was used to monitor a batch polymerisation process and it was shown how a multivariate calibration model could be developed and used with the principles of MSPC. The calibration model was used to monitor the decrease in the monomer concentration whilst the monitoring model was used to detect deviations in experimental batch runs.

An on-line drying process for the manufacture of an active pharmaceutical ingredient by NIR spectroscopy was reported (Parris *et al.*, 2005). On-line NIR spectroscopy was utilised to detect the end-point of a drying process in contrast to the traditional off-line manual sampling approach of High Performance Liquid Chromatography (HPLC). Since most of the solvents to wet the API can be detected by NIR, this approach allowed a predictive model to be built to detect the end-point by correlating the on-line data with the off-line results.

5.5 Literature Review on Data Integration Approaches

The data used in batch process monitoring are generally divided into three categories: process data, quality data and spectroscopic data. The process data according to Gurden *et al.* (2002) is characterised by its heterogeneity as the process measurements are monitored by different sensors and are therefore recorded in different units. This type of data is very useful to characterise the physical form of the process and is a valuable and fundamental source of information. The quality data of product is a measure of its quality state and the specification limits are defined for each product characteristic. Spectral data according to Gurden *et al.* (2002) is characterised by its description of the process chemistry and relates to the molecular nature of the species. It is characterised by homogeneity since the wavelengths are measured in the same units.

In practice integrating different types of data can maximise the amount of knowledge extracted about a process. A number of researchers have considered the integration of different types of data and a few applications have been reported for batch process monitoring. The view from Gurden *et al.* (2002) is that spectroscopy is complemented by the process measurements and when the process and spectral data are combined, a full description of the process in terms of the chemical and physical states is realised. Two ideas were proposed for the conjunction of the data: data augmentation and multi-block analysis. In data augmentation, the spectral block of data is treated as an additional block of highly correlated variables and is added to the process data to form a large data set. The second approach is to treat the process and spectral data as separate entities but the data are combined using multi-block analysis as described in Section 5.2. Combining different spectroscopic techniques has also been considered as a consequence of the shift from traditional off-line analysis to multiple on-line measurements. Nevertheless, Gurden *et al.* (2002) only provided a conceptual framework of how the data can be integrated. They did not provide an application study.

Cimander and Mandenius (2002) discussed an application where multiple measurements including data from bioprocess sensors, the electronic nose, NIR, on-line HPLC and mass spectrometer were used for the monitoring of an experimental tryptophan production. A multi-analyser bioreactor system was considered and was connected to an expert system for process monitoring and control, Figure 5-13.

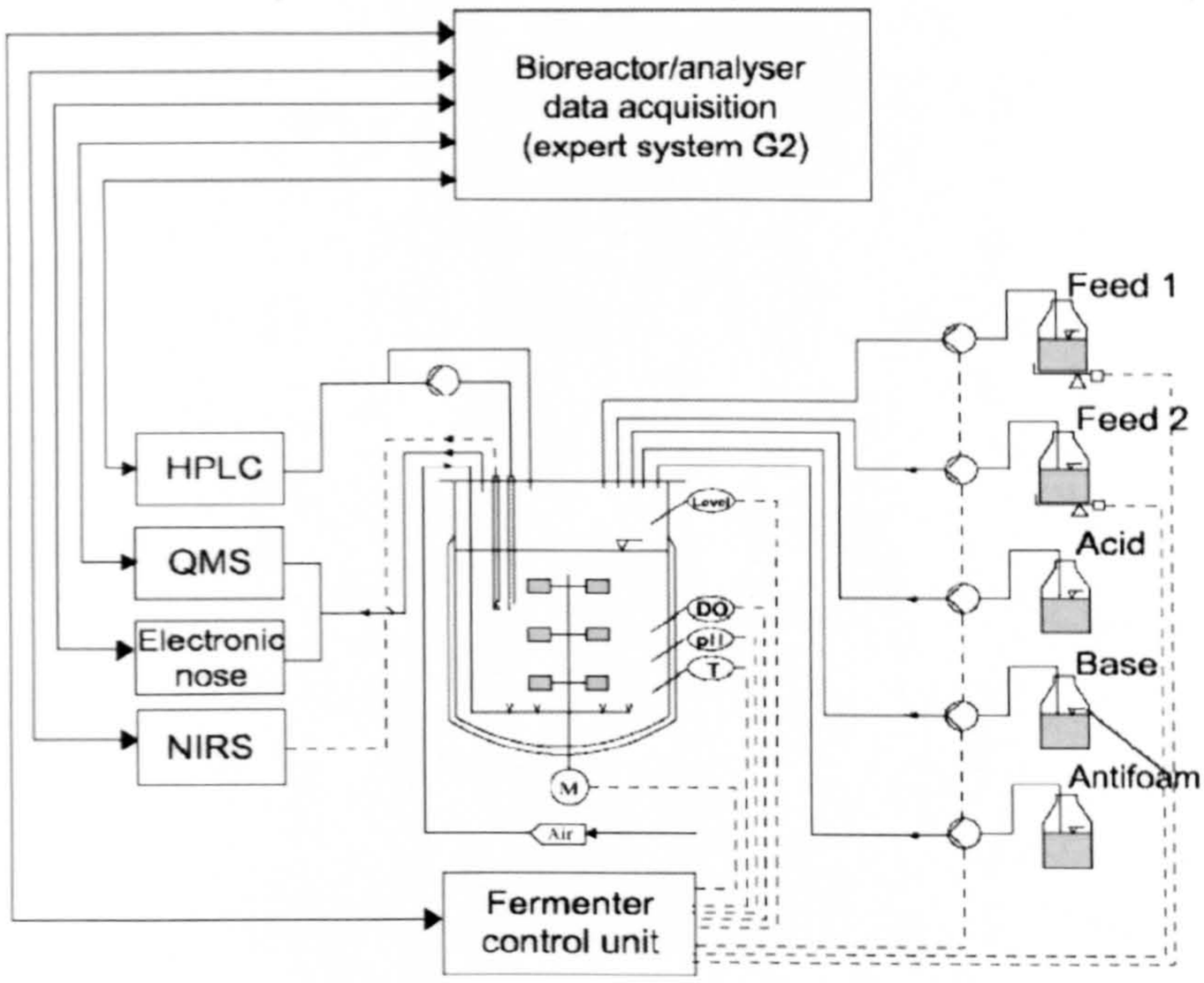


Figure 5-13 Multi-analyser bioreactor system (Cimander and Mandenius, 2002)

A set of representative variables were identified and used for the development of a MSPC representation using PLS. The authors concluded that a high level of correlation was observed between the process measurements and the final quality. However, no information on how the different types of data were combined and synchronised was provided nor whether the data was analysed through standard PLS or multiway PLS. It was also not clear from the paper, how the different spectral data were pre-processed and weighted between blocks.

An application of combining and comparing MIR and NIR measurements using multi-block PLS was described in Bras *et al.* (2005) for the modelling of soybean flour quality properties. Calibration models based on the individual data sets were built using PLS. The best PLS model was identified by applying the concept of the net analyte signal (NAS) where sensitivity, selectivity and limit of detection are considered. In multivariate calibration, NAS is used to define which wavelengths from a multivariate component spectrum are related to the quality of interest (Lorber *et al.*, 1997). The two spectral data sets were then combined using multi-block PLS. The data were first pre-processed by removing the non-informative region, then MSC was applied to remove spectral artefacts before mean-centring the resulting data. Two types of multi-block PLS models were considered: Multi-block PLS (MBPLS) and Serial PLS (SPLS). MBPLS (Westerhuis and Coenegracht, 1997) is an extension of PLS where the difference is in the grouping of the predictors, **X**, into blocks to improve the interpretability of the model. SPLS was

applied as an alternative multi-block algorithm (Berglund and Wold, 1999). The underlying principle is that the predictor blocks are treated in a serial mode rather than parallel and the response variables are used to provide the connection. Consequently, two models were built based on the two combinations, MIR + NIR and NIR + MIR. From the reported results, multi-block techniques were proven to be effective in identifying the effect from each spectral data set in relation to the flour properties. This is not possible using individual models.

Felicio *et al.* (2005) focused on comparing different PLS and multi-block PLS algorithms through their application to a set of MIR and NIR spectral data for the prediction of the flash point in gas oil, benzene and research octane number of gasoline. Single block PLS was applied to both the MIR and NIR data sets and MBPLS and SPLS were also considered. Model performance was compared using the metrics of Q^2 and Root Mean Square Error of Prediction (RMSEP). The only pre-processing applied was to divide the MIR and NIR spectra by their maximum absorbance value as the magnitude of absorbance in the NIR spectra was much higher than that for the MIR spectra. The results were compared and it was concluded that although SPLS gave good and robust results, the advantage over a single PLS model is small. The performance of MBPLS lay between that of PLS and SPLS.

5.6 Conclusions

This Chapter described two advanced methodologies that can facilitate the emerging need to integrate and combine different forms of data to attain increased process understanding of batch processes. The first technique was multi-block analysis. Different multi-block PCA methods were introduced and a number of advantages including enhanced interpretation of data structure, model stability and the capability to handle large amounts of data were identified. The wavelet transform was the second technique considered. It offers better features than both the Fourier transform and the short-time Fourier transform to handle a signal. The idea of multiresolution analysis is commonly applied to many applications.

A consequence of the technological advance in process spectroscopy is that previous traditional laboratory based spectroscopic techniques can be deployed in manufacturing processes. Differences exist between the use of spectroscopy off-line and on-line thus some considerations in terms of on-line analysis were highlighted. There are an increasing number of MSPC applications of on-line spectral data due to data availability and consequently the concept of data integration is now becoming of significant interest. This aspect was reviewed in the last section.

In the next chapter, an application of the integration of process and spectral data using the afore mentioned techniques is described.

Chapter 6

Application of Process and Spectral Data Integration to a Batch Mini-plant Experiment

6.1 Introduction

One of the objectives of the FDA PAT initiative is to achieve better control throughout the process by improving the scientific understanding of products and unit operations for Active Pharmaceutical Ingredient (API) and formulation manufacture. To investigate this hypothesis, a qualitative investigation into the combining of two data forms using multi-block and multiresolution analysis is undertaken. The results of the combined analysis are compared with those attained from separate analyses undertaken on the process and spectral data. To investigate the potential of the developed methodology, process and UV-Visible data from a batch mini-plant is considered.

By adopting an integrated approach, the relevant information can be presented in a single representation therefore reducing the complexity of the graphical presentation. The focus of this Chapter is to examine the potential of two data integration approaches that utilise multi-block and wavelet analysis. Both integrated approaches are performed on the basis of the proposed monitoring approach introduced in Chapter 3. The enhanced performance monitoring features are therefore captured. Design of experiment procedures are used to introduce controlled process variation into the test batches. In addition a variety of physical and chemical process manipulations are made to a number of batches to assess the fault detection abilities of the different approaches.

6.2 Design of Experiments

Design of Experiments (DoE) is a methodology for planning and executing experiments to extract the maximum amount of information from a limited set of experimental runs. A detailed description of experimental design and the underlying strategy can be found in Montgomery (2005) and Araujo and Brereton (1996a, 1996b, 1996c). DoE involves specifying a set of experiments that are likely to be the most informative with regard to a specific issue therefore the strategy is problem dependent. However, a common approach can be achieved by defining a standard reference experiment (centre point) and then various representative experiments are performed about this point in a systematic manner. Most experimentation in drug pharmaceutical development involves studying the impact of the change of several process variables (factors) to optimise processes and understand the relationship between factors and characteristics of the product quality (responses). Figure 6-1 illustrates the general model of inputs, factors and responses.

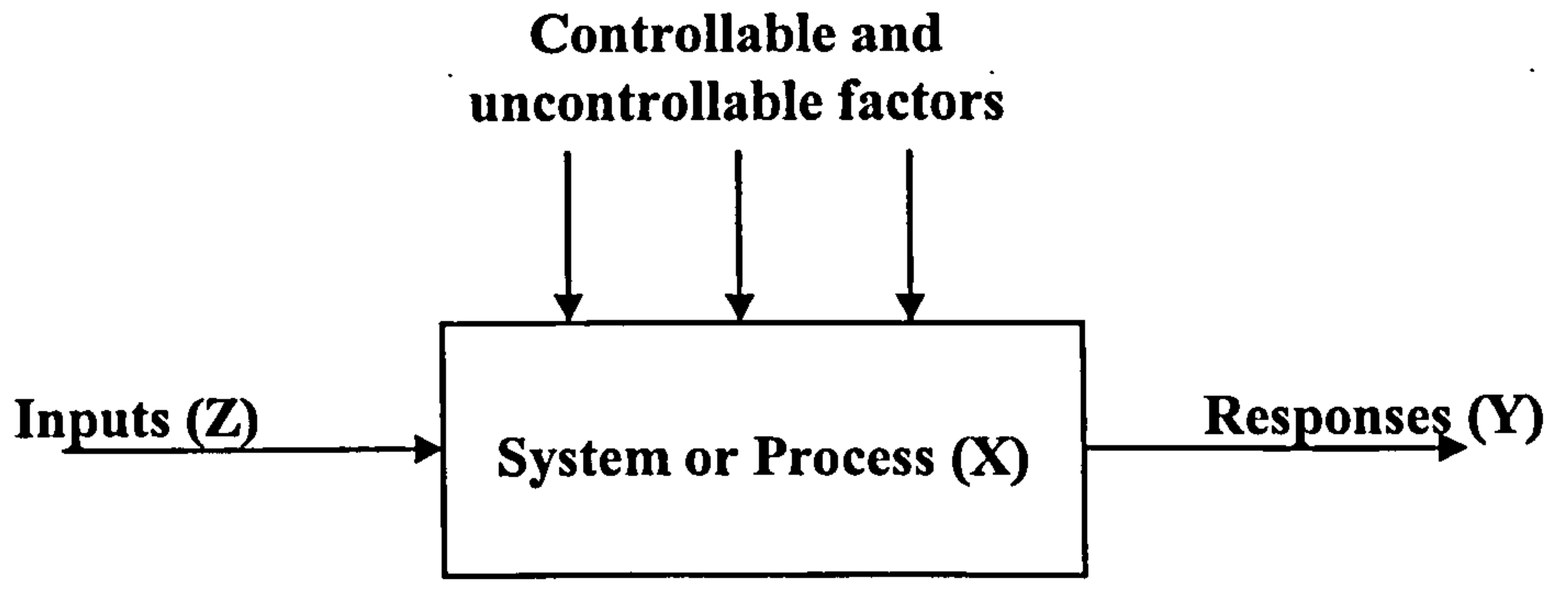


Figure 6-1 General model of inputs, factors and responses

It is important to identify the factors that have an influence on the responses to achieve optimal conditions. The factors can be either quantitative or qualitative. A quantitative variable is a continuous variable that can take any number between pre-defined ranges in the design. A qualitative factor is discrete such as on/off or species A or B. With DoE, the different factors are varied simultaneously over a set of experiments instead of the traditional approach of varying one factor at a time. This has improved experimental efficiency by reducing the number of experimental runs undertaken. A common experimental form is the factorial design where all possible combinations of inputs factors are considered. This is the approach adopted in the case study described in the next section.

6.3 Process Description

A simple reaction of nitrobenzene hydrogenation to aniline is conducted. The starting material is placed with the 5% Pd/C catalyst in a one-litre hastelloy vessel, fully baffled and operating with a pitched blade type impeller. The vessel is then pressurised with hydrogen and heated up to the desired temperature. The reaction is triggered by starting the agitator and bubbling hydrogen by hollow tubing into the solution. The process stops when hydrogen uptake is zero. Two critical factors were identified by the analytical chemists that would affect reaction performance, reaction temperature and the initial concentration of nitrobenzene. Therefore an experimental design was conducted to obtain controlled process variation. A factorial design with two factors and six centre points was performed to investigate the interaction of the critical factors on the formation of aniline. The structure of the experimental design batches is given in Table 6-1:

Batch	Reaction temperature (°C)	Initial concentration (M)	Duration (minute)
1	40	0.6	45.7
2	40	0.6	39.1
3	40	0.6	37.8
4	40	0.6	37.9
5	40	0.6	38.9
6	40	0.6	37.3
7	50 (H)	0.8 (H)	24.1
8	50 (H)	0.4 (L)	–
9	30 (L)	0.4 (L)	66.1
10	30 (L)	0.8 (H)	45.4
11	40	0.6	47.6
12	40	0.6	33.9

Table 6-1 Factorial design structure of aniline reaction: L – low setting, H – high setting

In addition, Figure 6-2 illustrates the graphical representation of this experimental design. Batches 1 – 6 are the centre points and were operated at 40°C and an initial concentration of 0.6M. The two levels of temperature were set to 30°C to 50°C and the initial concentration to 0.4 to 0.8M (batch 7 – 10). The results for batch 8 were lost during experimentation therefore no data was recorded. The goal of the study was to investigate whether an enhanced process monitoring and fault detection scheme could be achieved through an integrated approach, i.e. the conjunction of process and spectral data. Consequently two further batches (batch 11 and 12) were run with pre-designed process deviations which reflect both a change to the process as well as to the chemistry. The nominal reaction duration ranged from 37.3 to 45.7 minute with an average duration of 39.5 minute.

6.4 Data Pre-processing and Modelling

The graphical techniques associated with batch process data pre-screening (such as time series plots, trend plots and scatter plots) help in the understanding of the variable characteristics. By adopting this approach, spurious observations can be identified and treated. There are occasions where by examining the univariate time series of variables, process problems can be identified immediately without resorting to more complex methods. If univariate visualisation does not identify any outliers or spurious points, it is still necessary to perform multivariate visualisation to identify more subtle data problems.

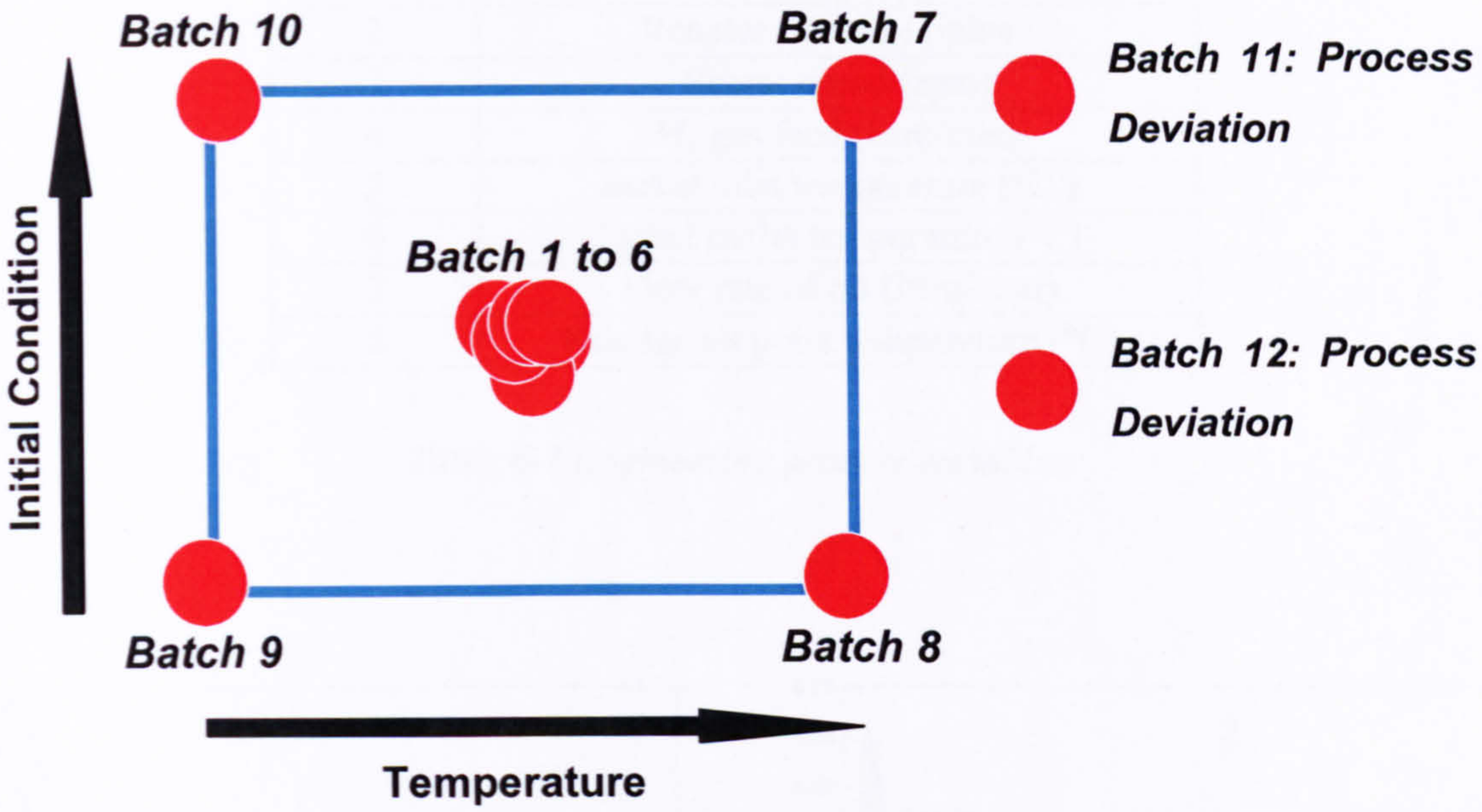


Figure 6-2 Graphical representation of the design of experiment batches

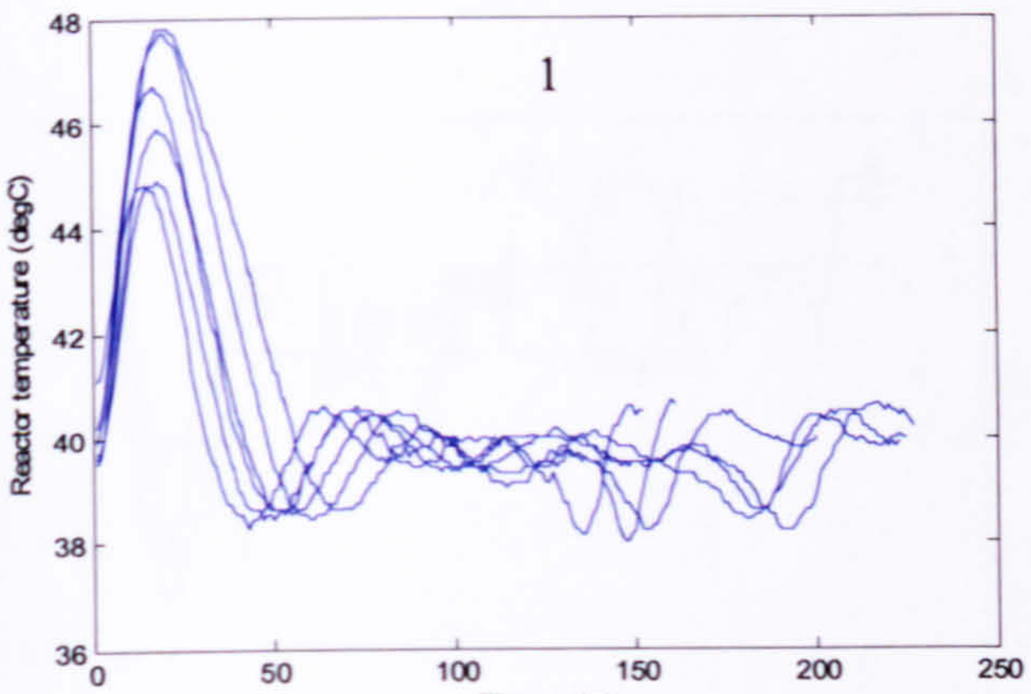
6.4.1 Engineering Process Measurements

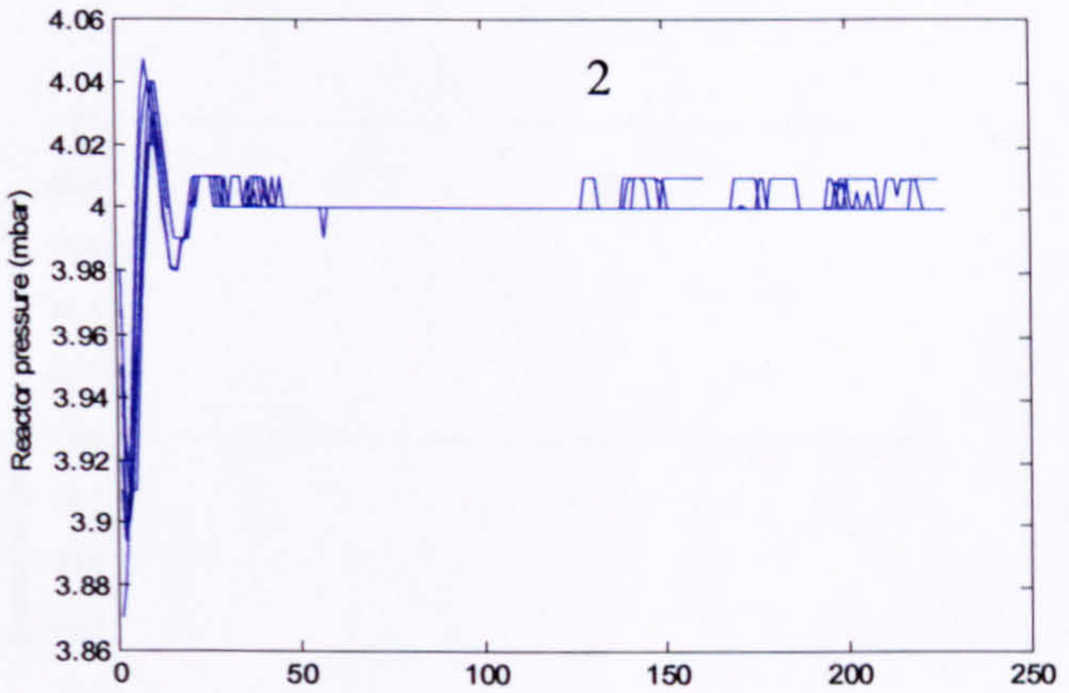
The engineering process measurements are listed in Table 6-2 and the time series plots of the nominal batches are shown in Figure 6-3. The sampling interval is every 10 seconds as to align to the sampling interval of spectral data. No spurious points are observed for the reference trajectories though there is an issue of varying duration of batches (Table 6-1). Since the focus of this section is on process monitoring and fault detection scheme, only the batches containing nominal and non-conforming behaviour are investigated, i.e. the centre point batches and the two batches incorporating process deviations. Batch 7, 9 and 10 are not considered in this study however results from those batches can be analysed as validation batches to confirm the validity of the DoE.

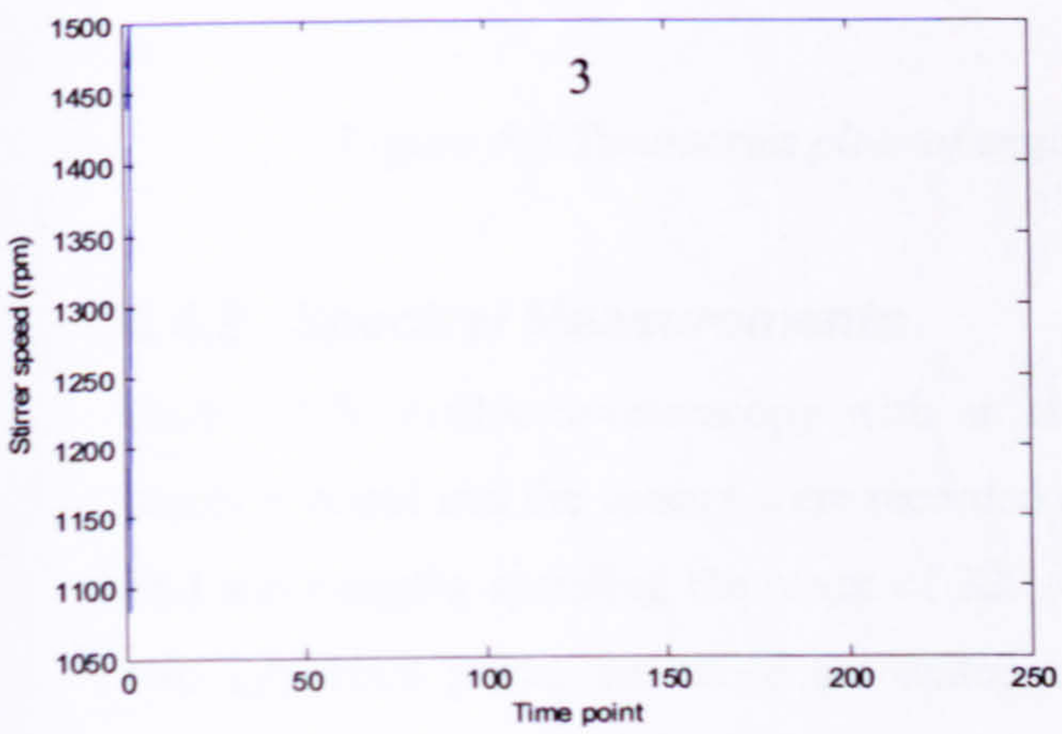
The pre-processing of process variables is an important step in the modelling of batch processes as described in Section 3.2. The engineering process variables were auto-scaled to unit variance and reactor set point temperature (variable 8) was excluded as it was not a critical processing parameter and was constant. However for a more complex process when set points are changed frequently, it can be considered as critical processing parameters for monitoring.

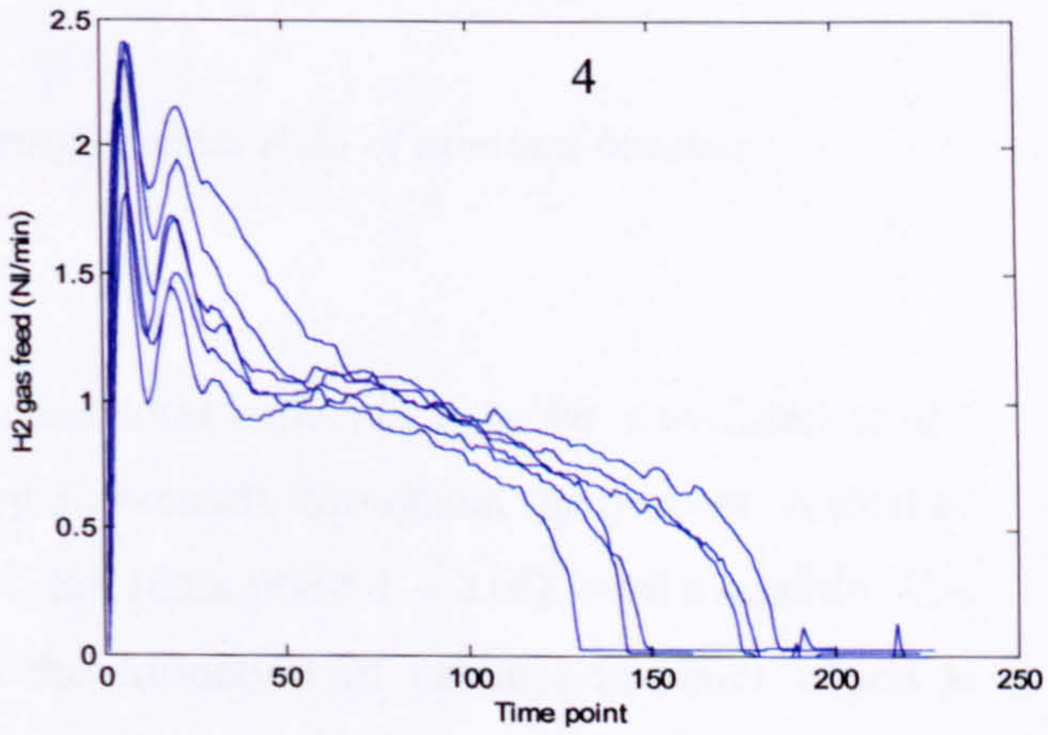
Variable	Description
1	Reactor temperature (°C)
2	Reactor pressure (mbar)
3	Stirrer speed (rpm)
4	H_2 gas feed (litre/min)
5	Jacket inlet temperature (°C)
6	Jacket outlet temperature (°C)
7	Flow rate of oil (litre/hour)
8	Reactor set point temperature (°C)

Table 6-2 Engineering process variables









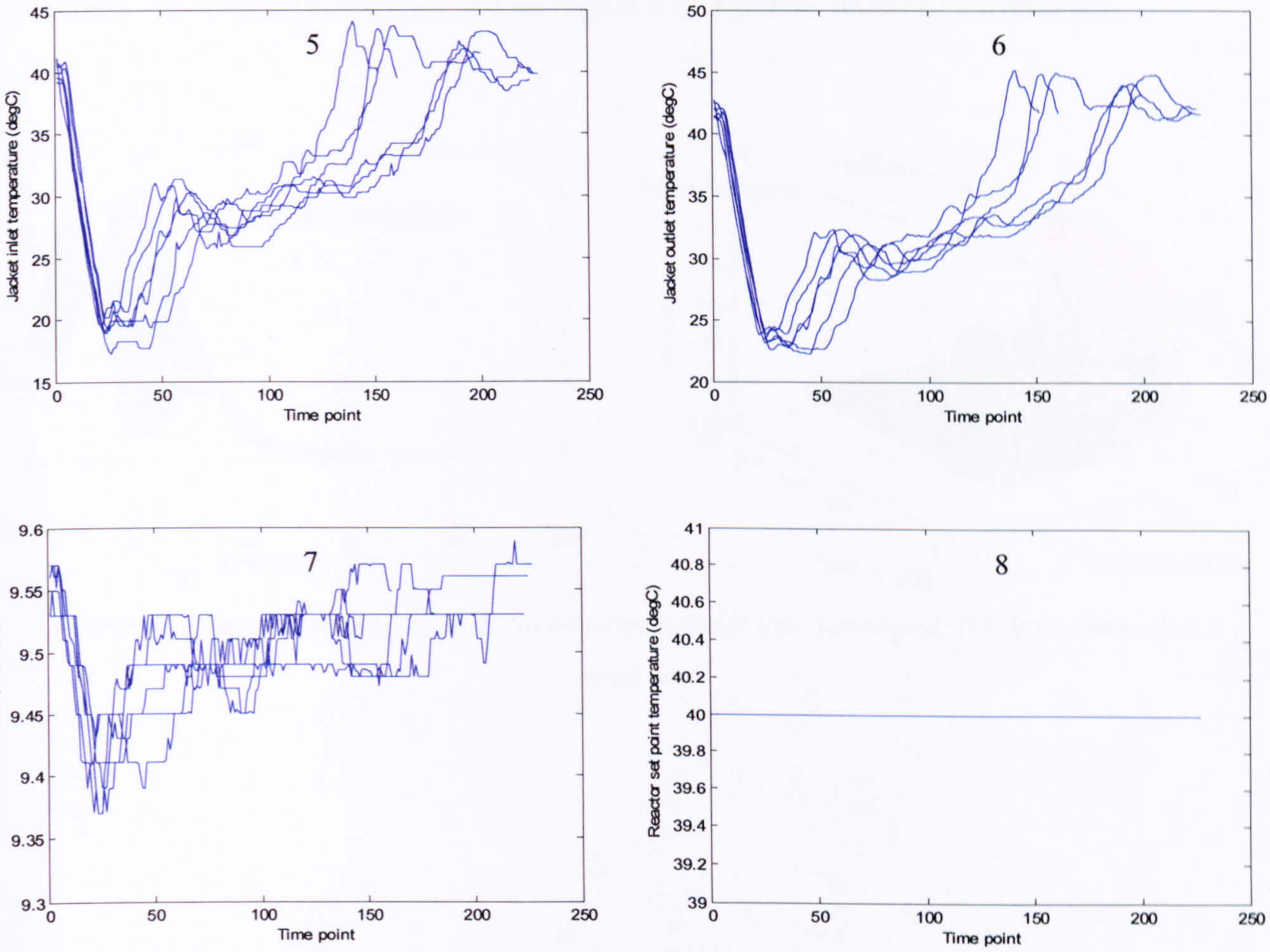


Figure 6-3 Time series plots of engineering process data of nominal batches

6.4.2 Spectral Measurements

On-line UV-Visible spectroscopy with an attenuated total reflectance probe was fitted to the reaction vessel and the spectra were recorded every 30 seconds throughout the process. A total of 180 wavelengths spanning the range of 220 – 400 nm (data point 1 – 216) were available. The two important peaks are those associated with the formation of product (aniline) which is absorbed at 236 nm (data point 19) and the depreciation of reactant (nitrobenzene) at 260 nm (data point 49). One of the challenges of data integration is to attain samples at the same time points for the disparate data sets. The process measurements may be recorded with a sampling interval of seconds but the time frame for the spectroscopic measurements is typically larger. In this study to realise the more rapid detection of a fault, a sample rate of ten seconds was selected and zero order linear interpolation of the spectral data was applied. A typical UV-Visible spectrum for batch 2 is represented as a two-dimensional time series and a three-dimensional mesh presentation is shown in Figure 6-4. The two important peaks are clearly shown in both

representations that the aniline peak increases from the start of the reaction whilst nitrobenzene peak decreases as it was consumed during the reaction. By examining the three-dimensional representation for all the nominal batches (Figure 6-5), a shift in the baseline is observed.

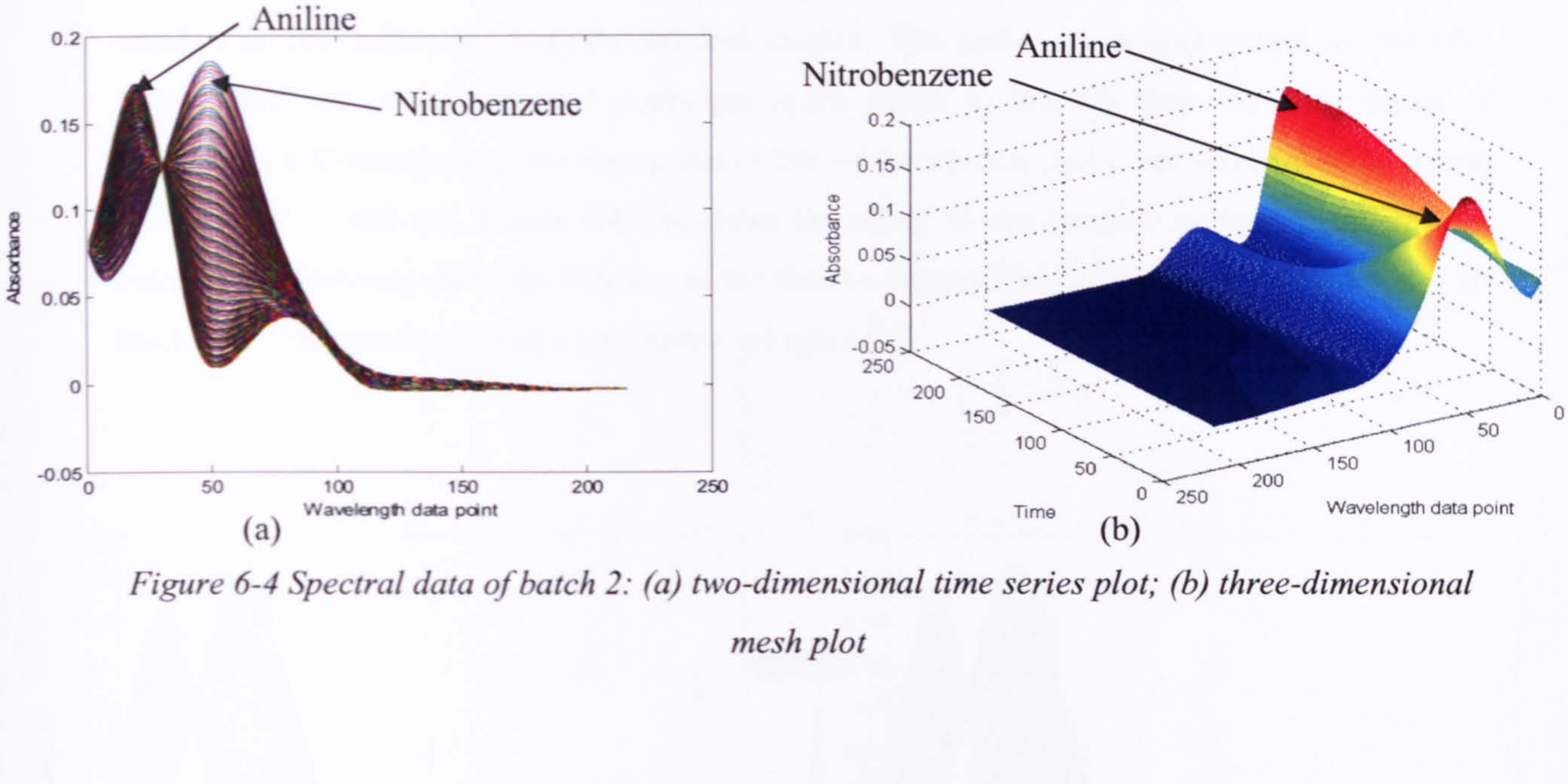


Figure 6-4 Spectral data of batch 2: (a) two-dimensional time series plot; (b) three-dimensional mesh plot

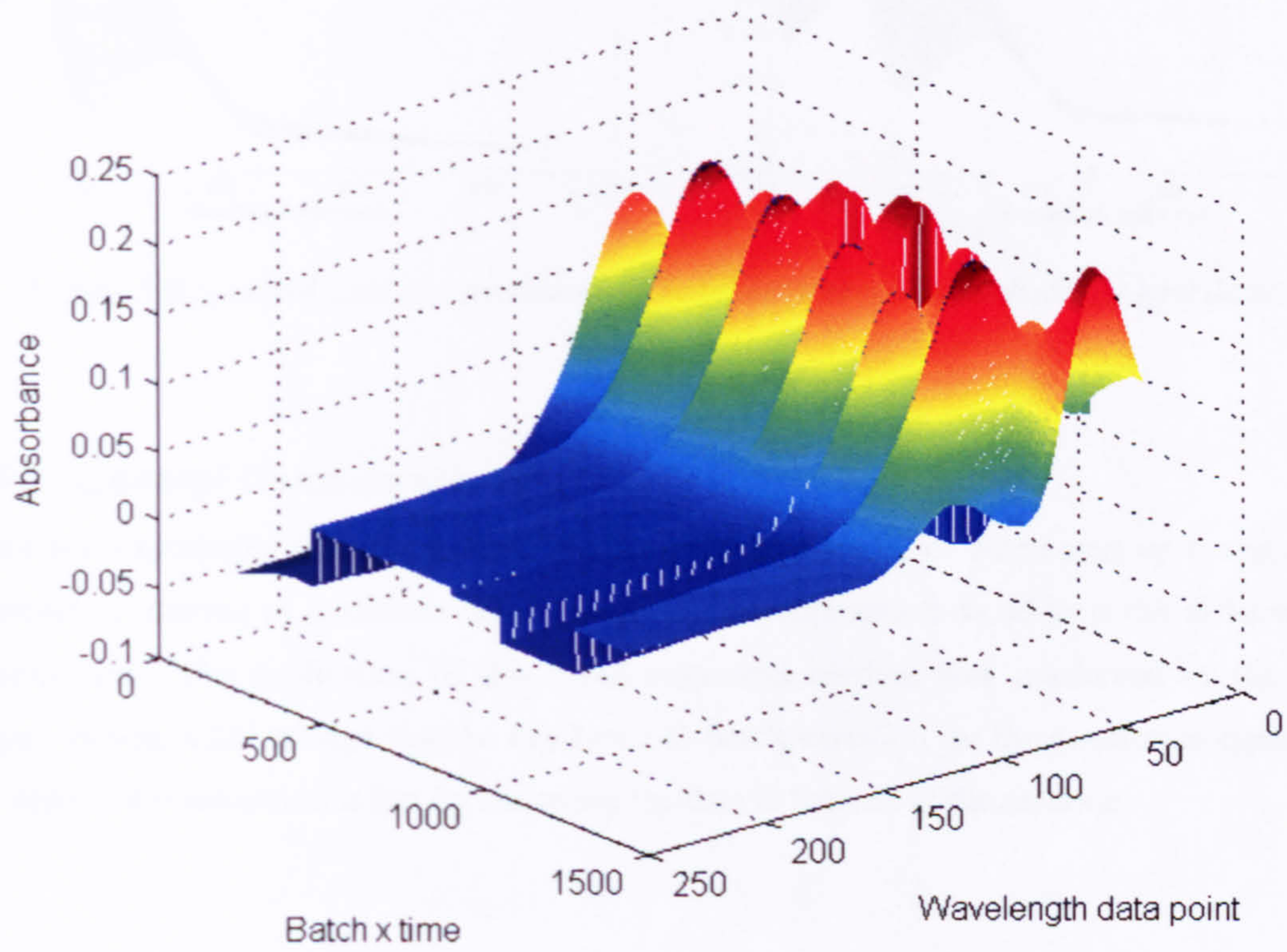


Figure 6-5 Three-dimensional mesh plot of spectral data of all nominal batches

It is important to standardise this variation before applying any form of scaling as the overall objective of the analysis is to model the batch-to-batch variation thus if a local difference exists, the global mean trajectories cannot be removed. A baseline correction technique is employed to remove this offset. A region of the spectrum is selected which is relatively free of absorbance peaks (no peak observed for the region), a line is then regressed through the region and this baseline is then subtracted from the original spectra. This technique is appropriate for the UV-Visible data set as the observed peaks are much easier to identify than for other forms of spectroscopy. Consequently, the region that is free of absorbance peaks was selected in the region between 380 – 400 nm. Figure 6-6 illustrates the effect of the baseline correction method for batch 5. It effectively shifts the baseline of the spectra to zero. This procedure is performed for all batches and the resulting spectra are shown in Figure 6-7.

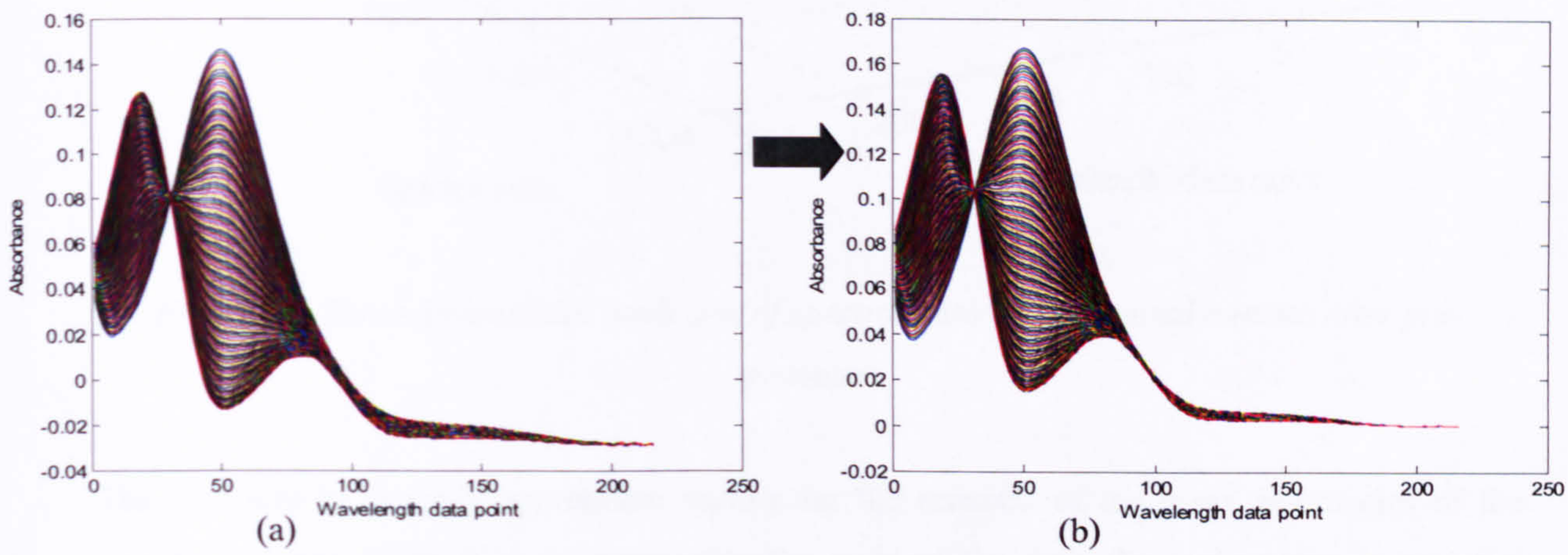


Figure 6-6 Spectral data pre-treatment by baseline correction method: (a) raw data; (b) corrected data

6.4.3 General Considerations

Since this experiment involves a rapid reaction, the monitoring of batch start up is critical. The approach of cutting to minimum batch length is thus appropriate to address the different batch lengths issue. The application of this batch alignment method was reinforced by the process engineers who acknowledge that the key batch-to-batch variation for the reaction is captured and no important information is lost by removing the data at the end of the reaction.

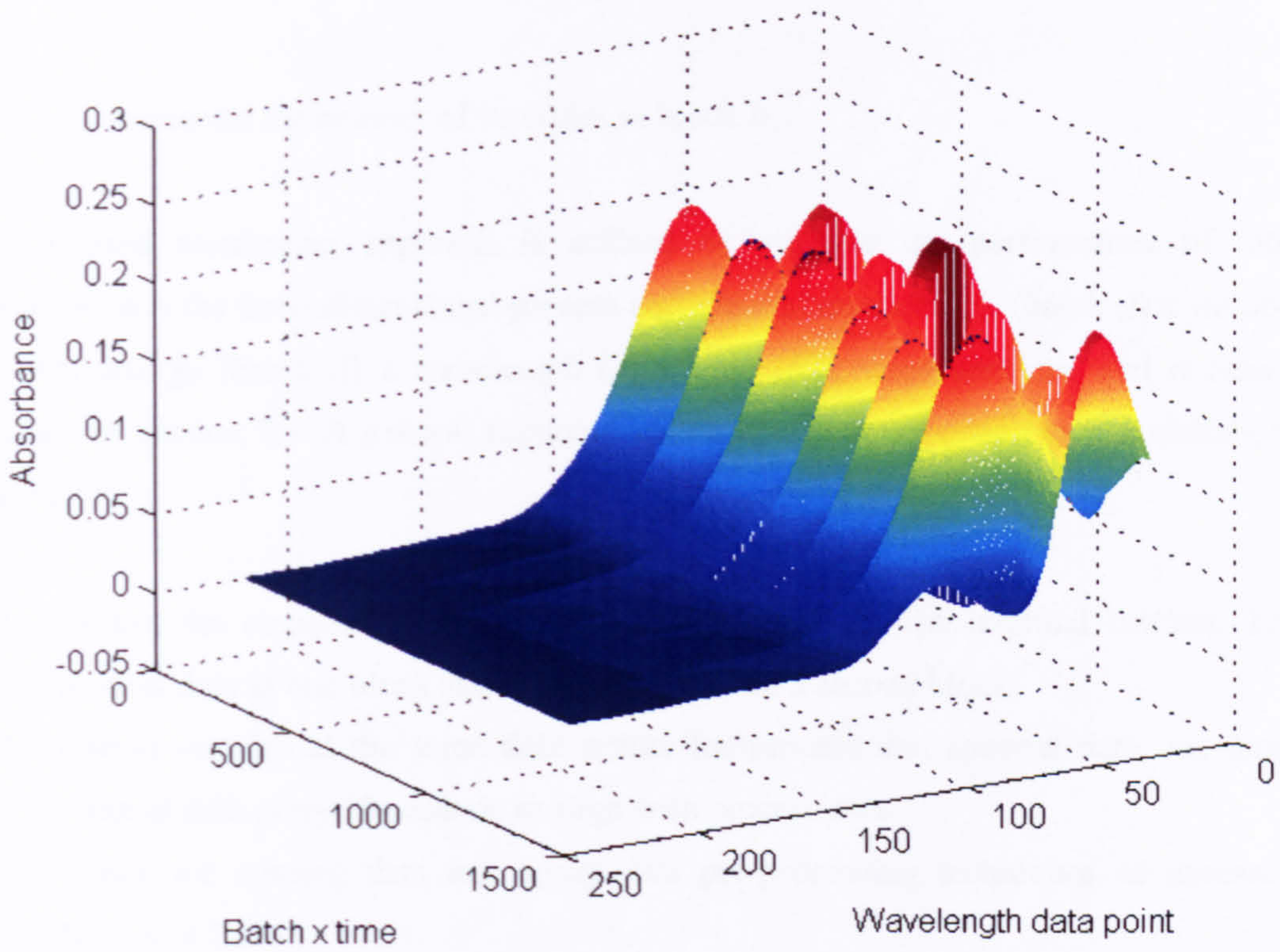


Figure 6-7 Three-dimensional mesh plot of spectral data of all nominal batches after pre-treatment

The next step is to apply appropriate scaling for the removal of the mean trajectories of the nominal batches. Since PCA is sensitive to the scale of the data, the scaling step is essential especially in the case of combining different forms of data for a single model analysis. The general practice of scaling process data is auto-scaling with mean-centring typically being applied to spectral data since the absorbance are of the same units. The process data was previously auto-scaled and to achieve consistency between the scaling domains, the spectral data are normalised by auto-scaling. This is generally feasible for all types of data.

Another factor to consider when different blocks of data are analysed using a multi-block approach is that of weighting. This is especially important when the number of variables between blocks differ significantly. The weighting factor is applied after scaling and before the application of its multi-block algorithms. Equation 3.5 is re-stated as:

$$v_b = \frac{1}{\sqrt{j_{block}}} \qquad 6\text{-}1$$

where j_{block} represents the number of variables in block b.

The proposed monitoring approach is utilised to evaluate the performance of integrated approaches thus the three-dimensional process and spectral matrices, $\underline{\mathbf{X}}_1$ (batch (I) x variable (J) x time (K)) and $\underline{\mathbf{X}}_2$ (batch (I) x wavelength (J) x time (K)), were unfolded and re-arranged as described in Section 3.3. A general summary of the data pre-processing and modelling steps is given below.

1. Collect the engineering process and spectral data for the nominal batches. Treat the process data as one block and the spectral data as a second block.
2. Obtain samples at the same time points for process and spectral data, i.e. interpolate spectral data every 10 seconds to align with process data.
3. Check for missing data and apply data pre-processing techniques as necessary, see Section 3.2.1.
4. Apply batch length alignment, i.e. cutting to minimum batch length, see Section 3.2.3.
5. Apply baseline correction method to adjust spectral baseline, see Section 5.4.4.
6. Unfold the three-way matrices $\underline{\mathbf{X}}$ (I x J x K) to $\mathbf{X}$ (I x JK).
7. Apply auto-scaling to the process and spectral data, see Section 3.2.2.
8. Apply weighting factor to achieve equal variance, see Section 3.2.2.
9. Re-arrange the unfolded matrices $\mathbf{X}$ (I x JK) to $\mathbf{X}$ (IK x J), see Section 3.3.

6.4.4 Deviated Batches

The two batches with pre-designed process deviations are utilised to evaluate monitoring and fault detection performance. The engineering process variables and spectral data for batches 11 and 12 are shown in Figure 6-8 and Figure 6-9 respectively. A series of disturbances are observed for both types of data. The process deviations were carefully designed that both physical and chemical disturbances were introduced to the process. It was hypothesised that the physical sensors and chemical spectra would be able to detect such disturbances. The details of the disturbances are summarised below:

Batch 11

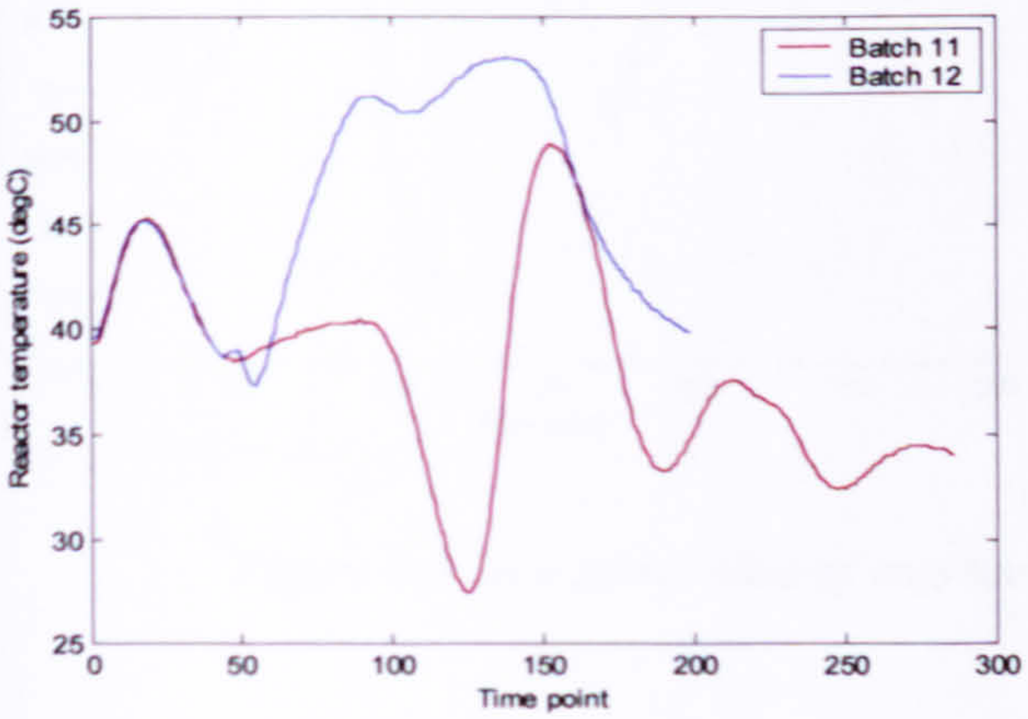
1. Discharge 10% less catalyst (simulate charging problem).
2. Simulate a series of temperature control problem.

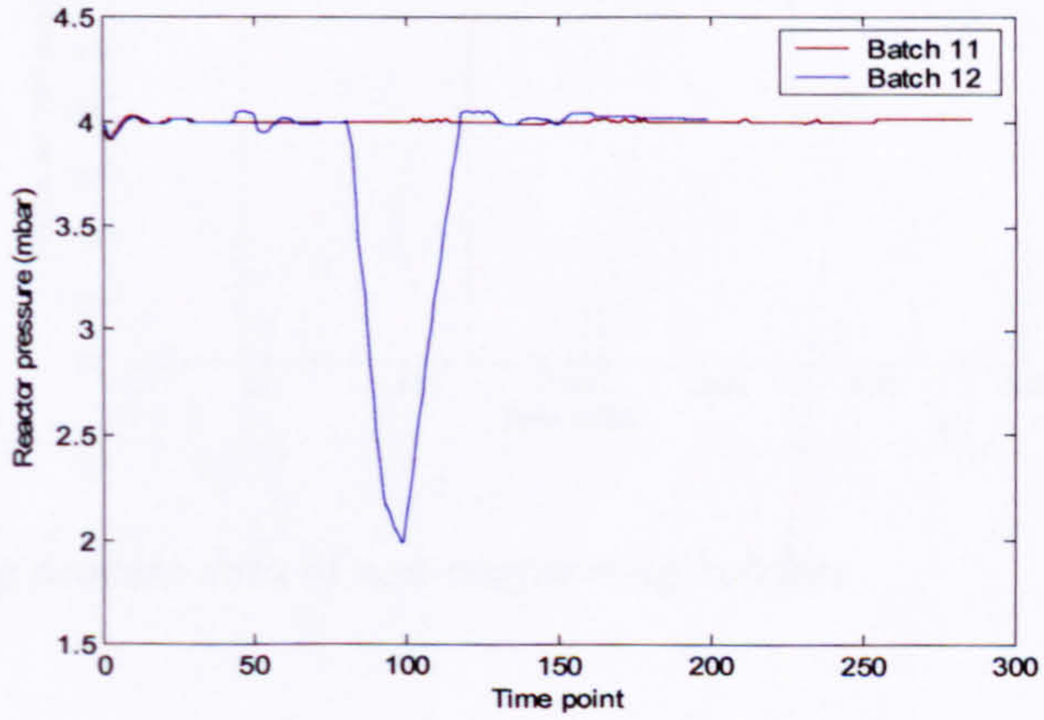
- Start at 7.1 min (data point 43): cool down the vessel from 40° to 30°C by jacket control.
- 14.9 – 19.5 min (90 – 117): Further cool down by internal vessel temperature probe.
- 19.6 min (118): Reheat the vessel back to 40°C.

Batch 12

1. Stirrer disturbance.
   - 7.1 – 8.1 min (43 – 49): Stop stirrer.
   - 8.2 min (50): Restart stirrer.
2. Simulate a series of pressure loss problem.
   - 13.3 – 15.3 min (80 – 92): Reduce pressure from 4 to 2 mbar.
   - 15.36 – 16.3 min (92 – 99): Remain the pressure at 2 mbar.
   - 16.3 – 21.4 min (99 – 129): Ramp pressure back to 4 mbar.
3. Stirrer disturbance.
   - 21.5 – 24.5 min (130 – 148): Reduce agitator speed from 1500 to 750 rpm (simulate less efficient gassing).
   - 24.6 min – end (149 – end): Increase agitator speed back to 1500 rpm.

Both sets of data were pre-processed according to the nominal batches and the same scaling factors were applied. The monitoring and fault detection capability of the monitoring schemes will be investigated in Section 6.6 and 6.7.





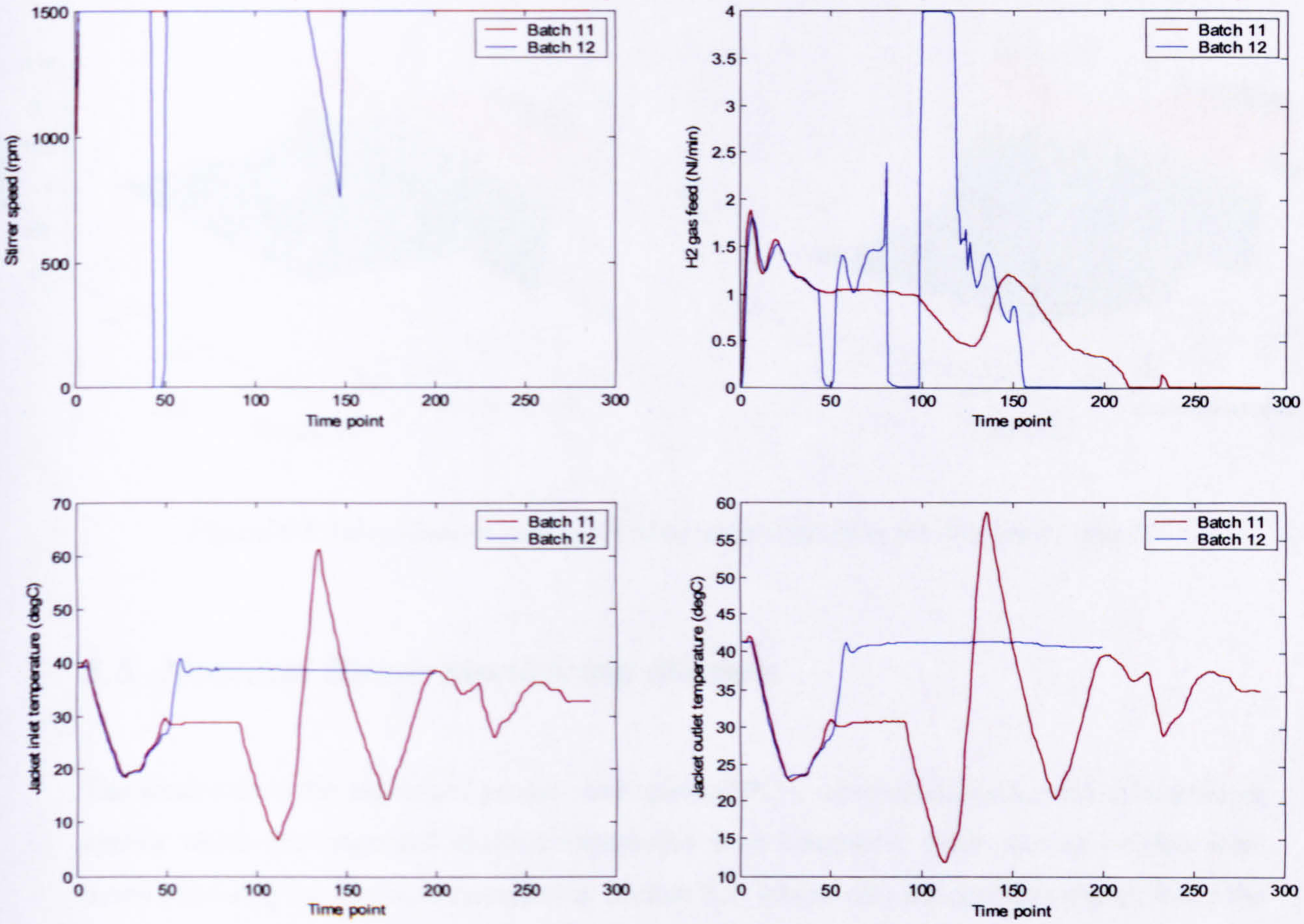


Figure 6-8 Time series plots of engineering process data of non-conforming batches

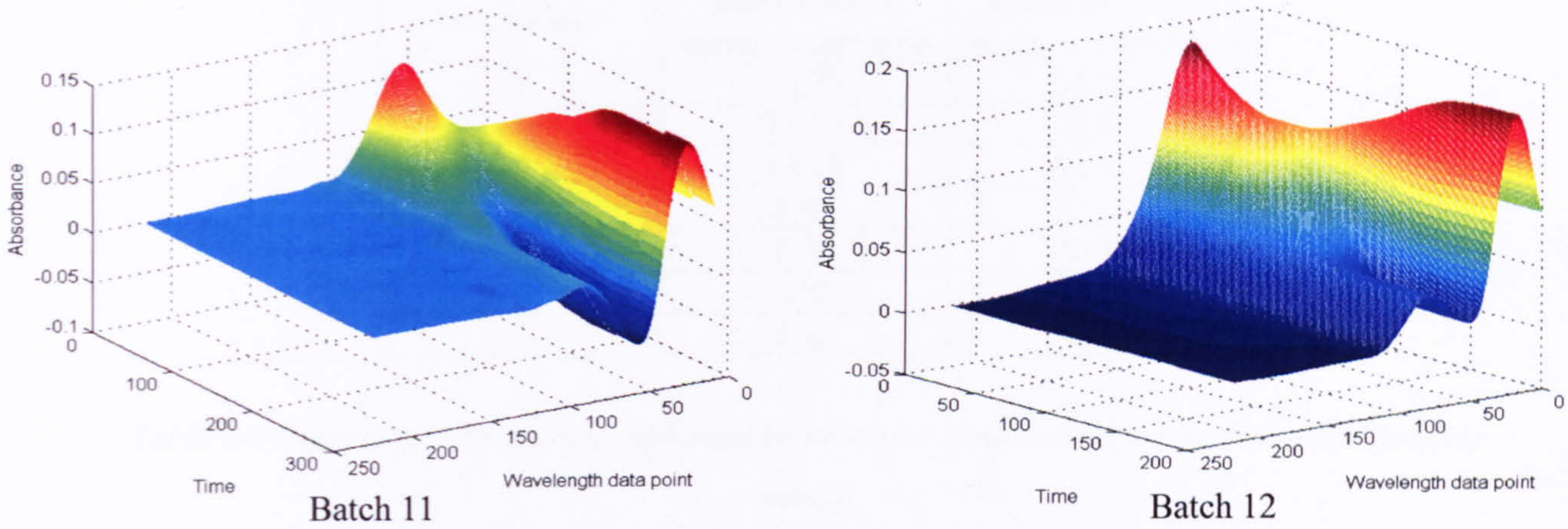


Figure 6-9 Three-dimensional mesh plots of spectral data for batches 11 and 12

6.5 Nominal Batch Monitoring Models

The results from the individual process and spectral PCA representations formed a benchmark against which the integrated analysis approaches were compared. Both nominal models were developed using the approach described in Section 3.3. Where only a single data source forms the basis of the analysis, the diagnostic metrics such as principal component scores provide a good indication of the state of the process. Hotelling's T^2 and Squared Prediction Error (SPE) can also provide additional indication of the process status. A further variant of the integrated approach that incorporated wavelet analysis was also developed and compared. Combining the proposed approach with multi-block PCA and / or with the wavelet transform is a novel strategy in the field of batch process monitoring.

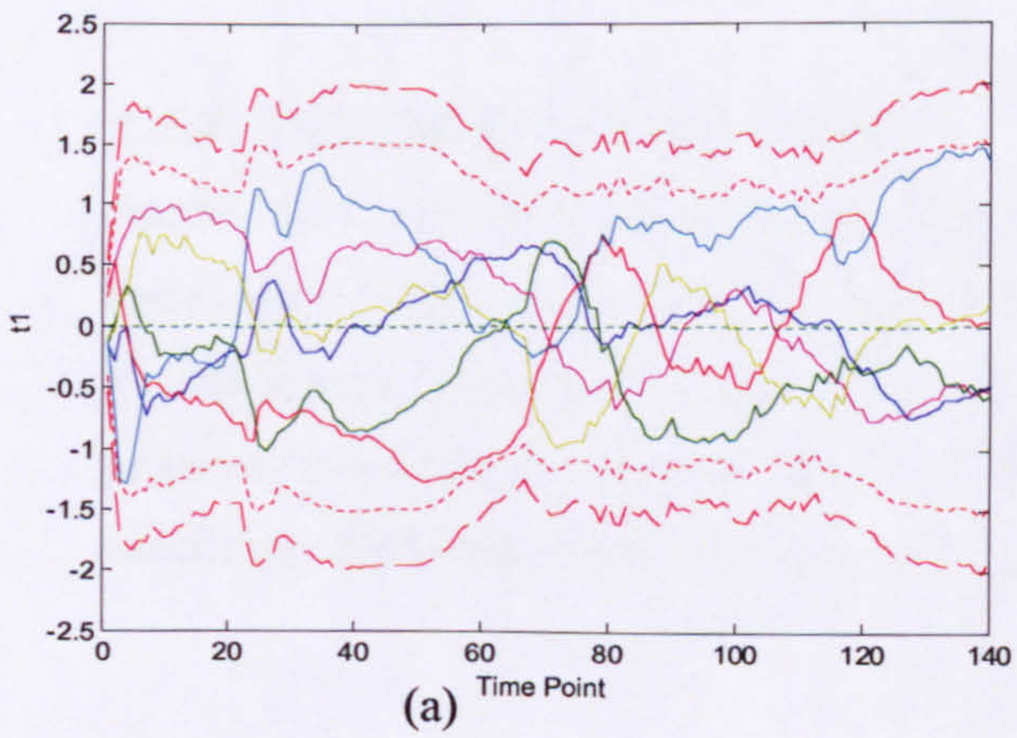
6.5.1 Individual Process Model

Three principal components were recommended by cross-validation to be retained for the individual process model, explaining 84% of the underlying variability. Table 6-3 shows the amount of variance captured by each of the principal component.

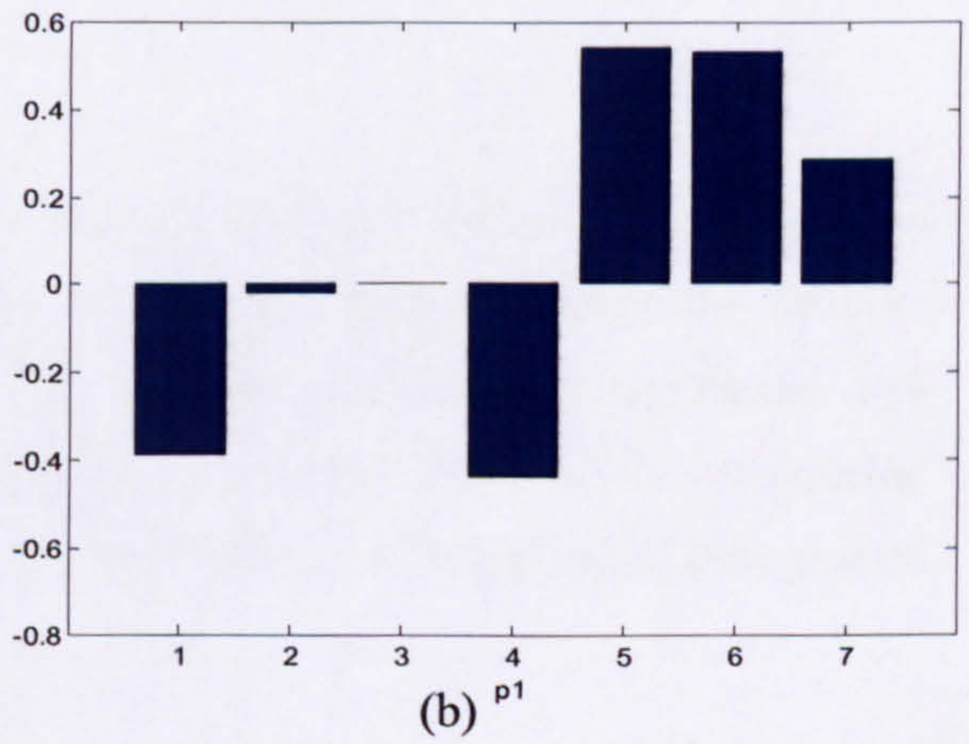
Principal component	Process model	
	Individual % variance captured	Cumulative % variance captured
1	54.73	54.73
2	17.15	71.88
3	12.13	84.01
4	8.43	92.43
5	6.36	98.79
6	0.92	99.70
7	0.30	100.00

Table 6-3 Percentage of variance explained by principal components for the individual process model

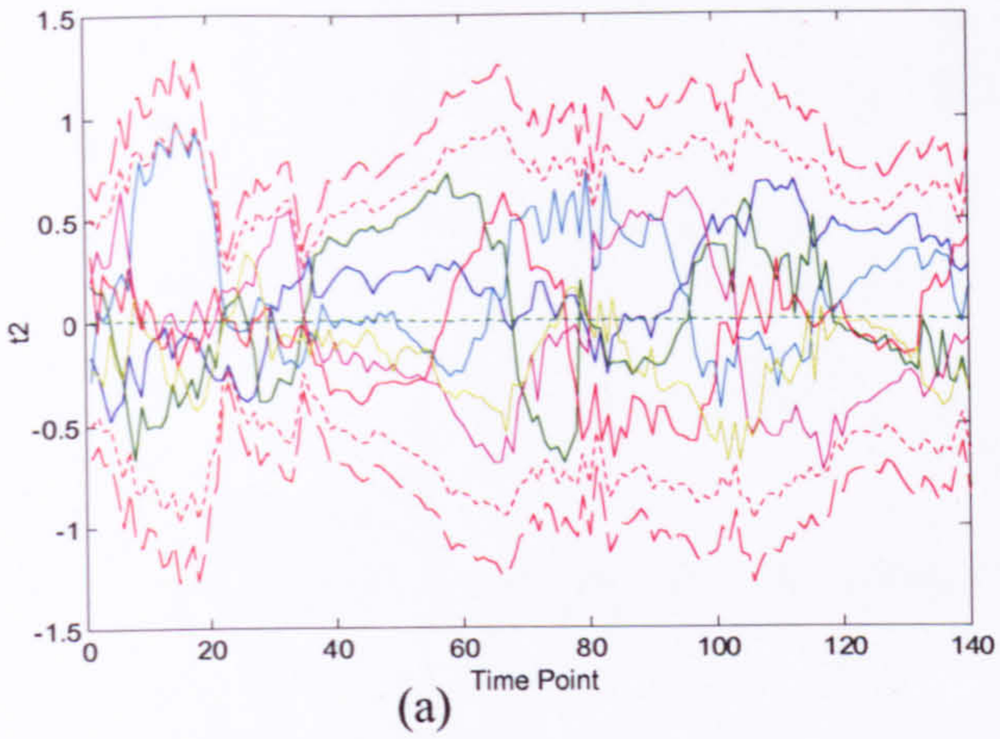
A nominal process representation based on the first three principal component scores and loadings for the engineering process variables are shown in Figure 6-10. The scores plot represents the normal batch variability throughout the process with the 95% and 99% statistical control limits defined as +/- 2 and 3 standard deviations respectively. By interrogating the scores plots in Figure 6-10, random variation is observed hence the model has captured nominal batch-to-batch variation. The loadings identify the key variables that determine the major sources of variation for each component. For principal component one, jacket inlet and outlet temperatures and flow rate of oil (variables 5 to 7) have a positive loading influence whilst reactor temperature and H_2 gas feed have the opposite influence. But all the variables in principal component two are shown to have same effect, i.e. they all have negative loadings. In particular flow rate of oil has a higher impact with respect to this principal component.



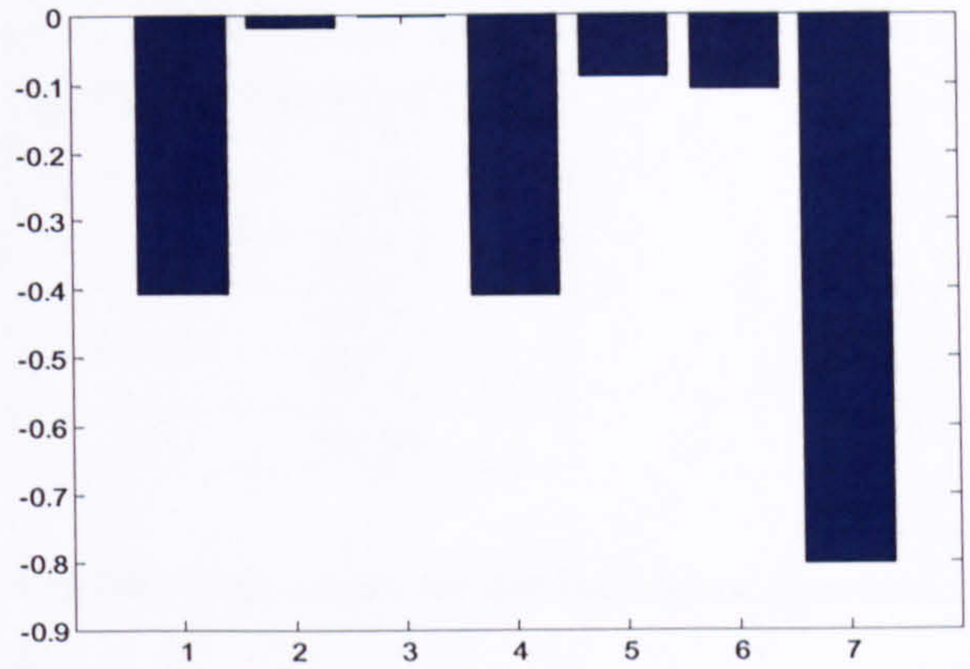
(a)



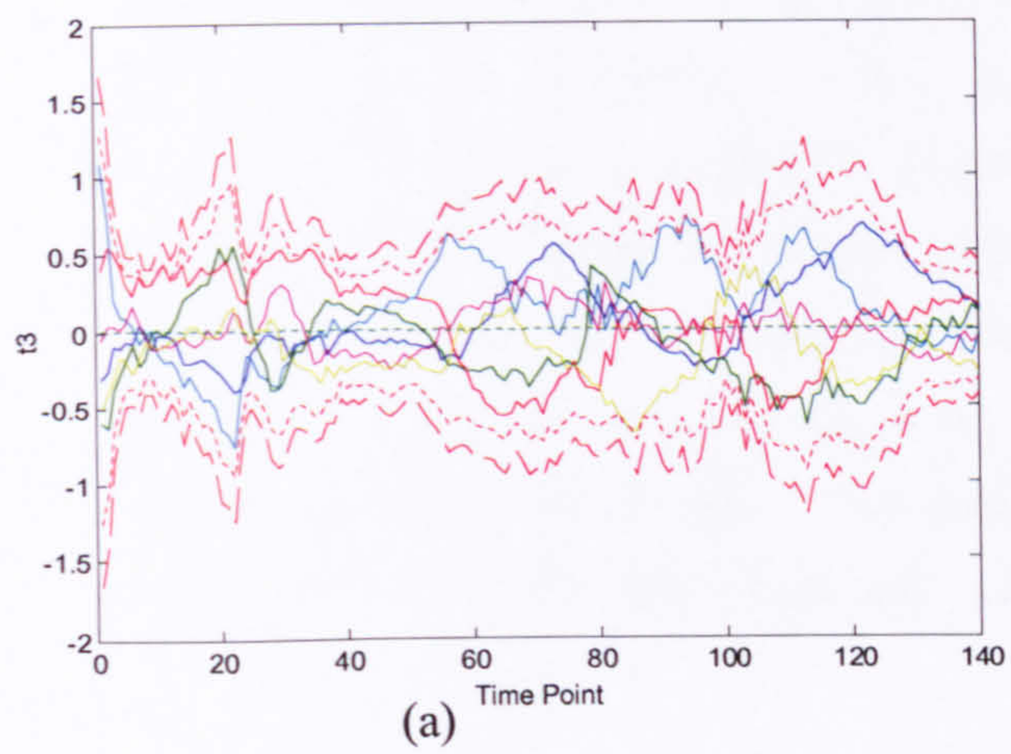
(b)



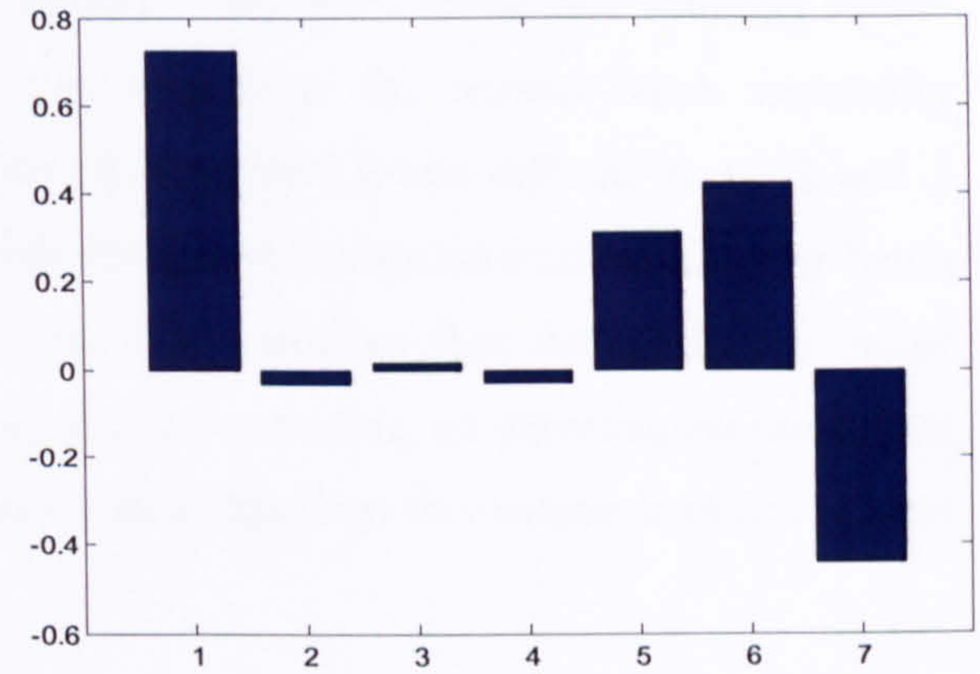
(a)



(b)



(a)



(b)

Figure 6-10 (a) Scores plots and (b) loadings plots for the individual nominal process model of principal components one to three

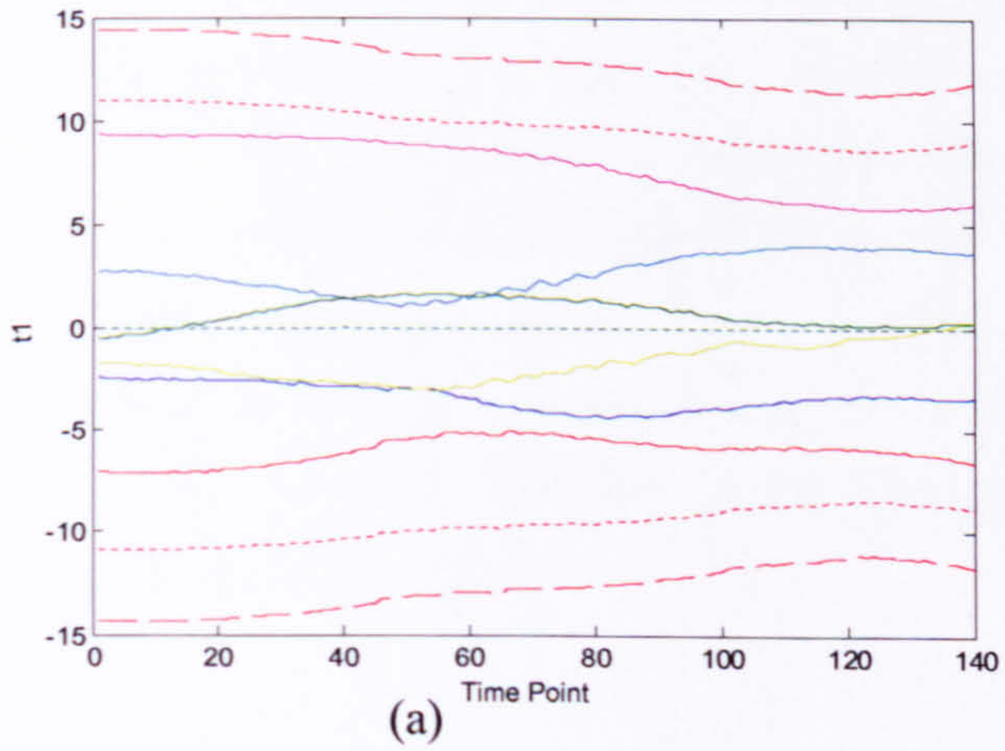
6.5.2 Individual Spectral Model

One of the factors for establishing a representative process model is to utilise important process knowledge. It was suggested by the analytical chemists that the region of interest for the UV-Visible spectra is 236 to 267 nm (data point 18 – 49). For the individual spectral model, two principal components are selected by cross-validation explaining 99% of the underlying variability. Table 6-4 shows the amount of variance captured by each of the principal components.

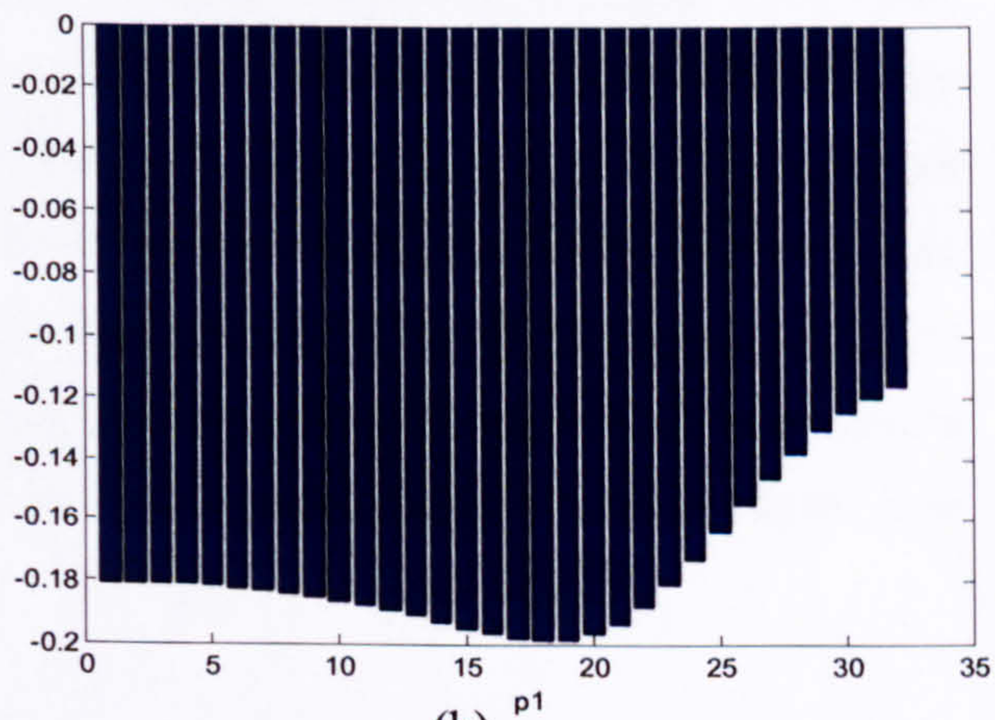
Principal component	Spectral model	
	Individual % variance captured	Cumulative % variance captured
1	78.53	78.53
2	20.85	99.38
3	1.74	99.23
4	0.53	99.77
5	0.15	99.91
6	0.05	99.96

Table 6-4 Percentage of variance explained by principal components for the individual spectral model

A nominal representation of the first and second principal component scores and loadings for the spectra are shown in Figure 6-11. The scores plot represents the normal batch variability throughout the process with the 95% and 99% statistical control limits defined as +/- 2 and 3 standard deviations respectively. Those scores plots are the reference against which new batch trajectories are compared. The loadings identify the key variables that determine the major sources of variation for each component. For principal component one, all wavelengths are shown to have negative loadings whilst a switching from negative loadings to positive loadings at data point 19 is shown in principal component two.



(a)



(b)

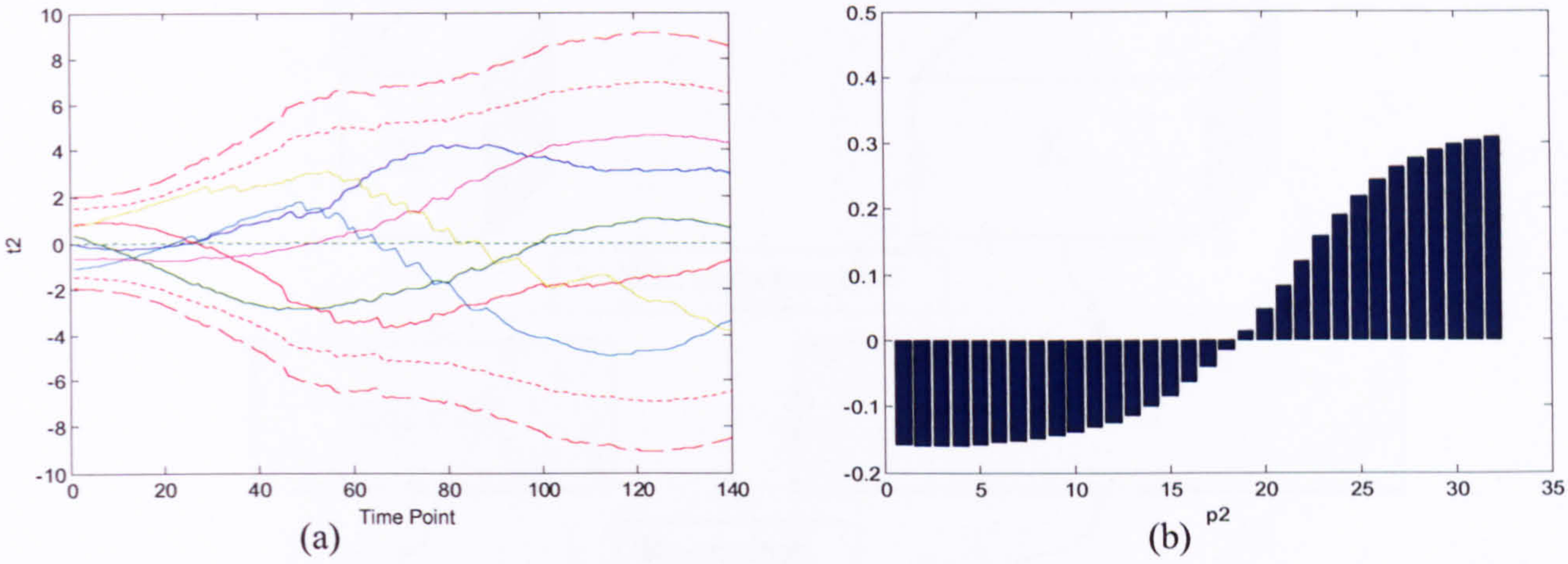


Figure 6-11 (a) Scores plots and (b) loadings plots for the individual nominal spectral model of principal components one and two

6.5.3 Integrated Multi-block PCA Model

For the integrated multi-block PCA approach, the process and spectral data are integrated using multi-block analysis, more specifically consensus PCA (CPCA) as discussed in Section 5.2.2. The generic problem of multi-block analysis is to identify underlying relationships between and within two possibly related data sets with the selection of CPCA being due to its enhanced performance compared with other multi-block algorithms (Smilde *et al.*, 2003). As discussed in Section 3.3, the proposed monitoring approach provides various advantages over other approaches such as the removal of the mean trajectories of the batches and no estimation of future observations required for on-line monitoring hence the CPCA algorithm is incorporated within the proposed monitoring approach for the first time. The success of the proposed monitoring approach will also be shown by its flexibility of incorporating additional algorithms for advanced batch performance monitoring. Figure 6-12 provides a schematic of the proposed integrated CPCA on-line monitoring scheme. The CPCA algorithm based on the NIPALS engine was described in Section 5.2.2.

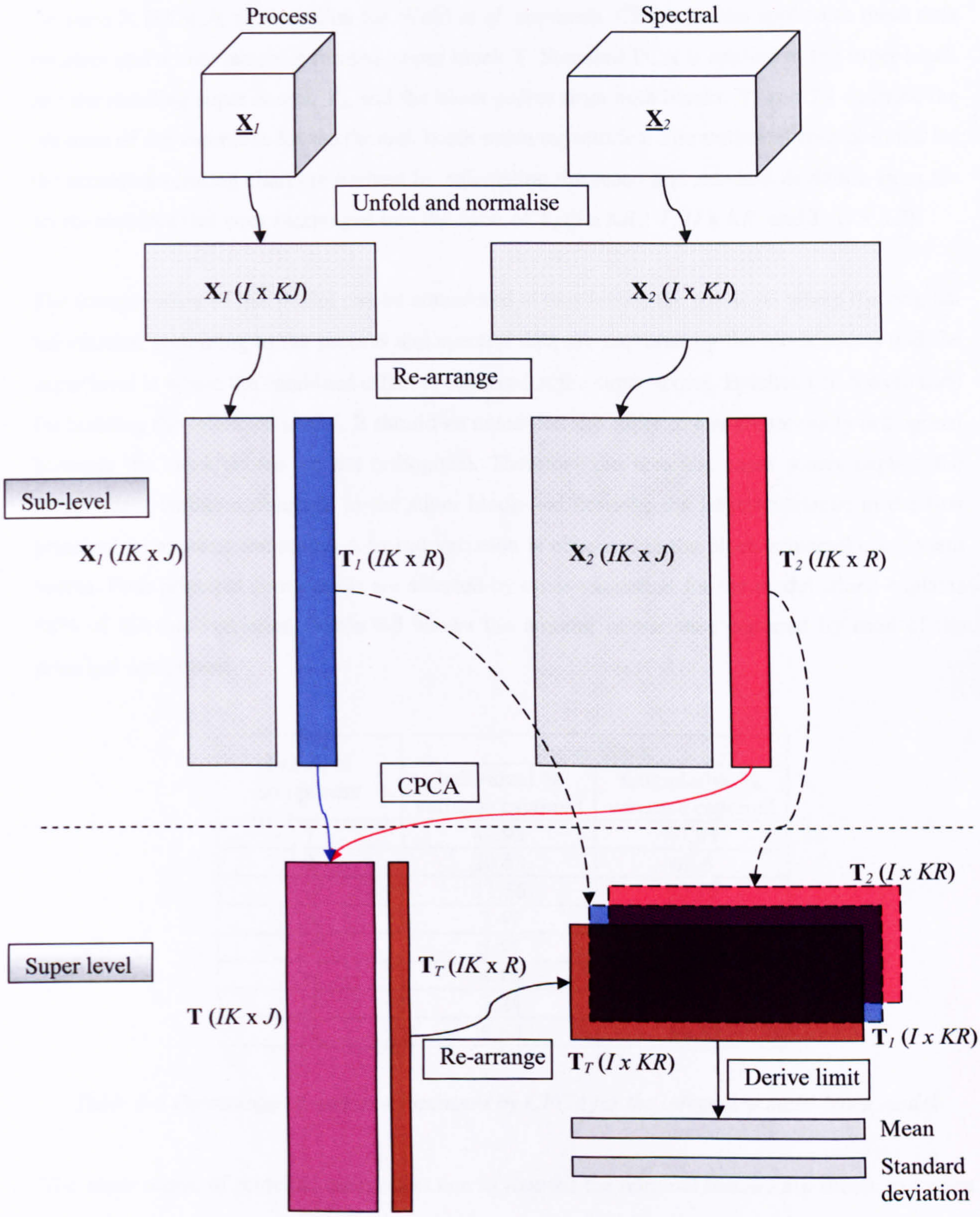


Figure 6-12 Proposed integrated CPCA on-line monitoring scheme

The three dimensional process and spectral matrices are placed in separate blocks, $\underline{\mathbf{X}}_1$ (batch (I) x variable (J) x time (K)) and $\underline{\mathbf{X}}_2$ (batch (I) x wavelength (J) x time (K)) respectively. The two blocks are then unfolded according to the Nomikos and MacGregor approach into the form $\mathbf{X}$ (I x

KJ) before they are centred and scaled to unit variance. The unfolded matrices are rearranged into the form **X** (*IK* x *J*), i.e. based on the Wold *et al.* approach. CPCA is then applied to these data matrices and a new matrix is formed, super block **T**. Standard PCA is applied to the super block and the resulting super scores, $\mathbf{T}_T$, and the block scores from both blocks, $\mathbf{T}_1$ and $\mathbf{T}_2$, describe the variance of the variations for the through batch mean trajectories. The statistical control limits for the scores monitoring chart are derived by calculating the mean and standard deviation from the scores matrices that were rearranged into the form of $\mathbf{T}_T$ (*I* x *KR*), $\mathbf{T}_1$ (*I* x *KR*) and $\mathbf{T}_2$ (*I* x *KR*).

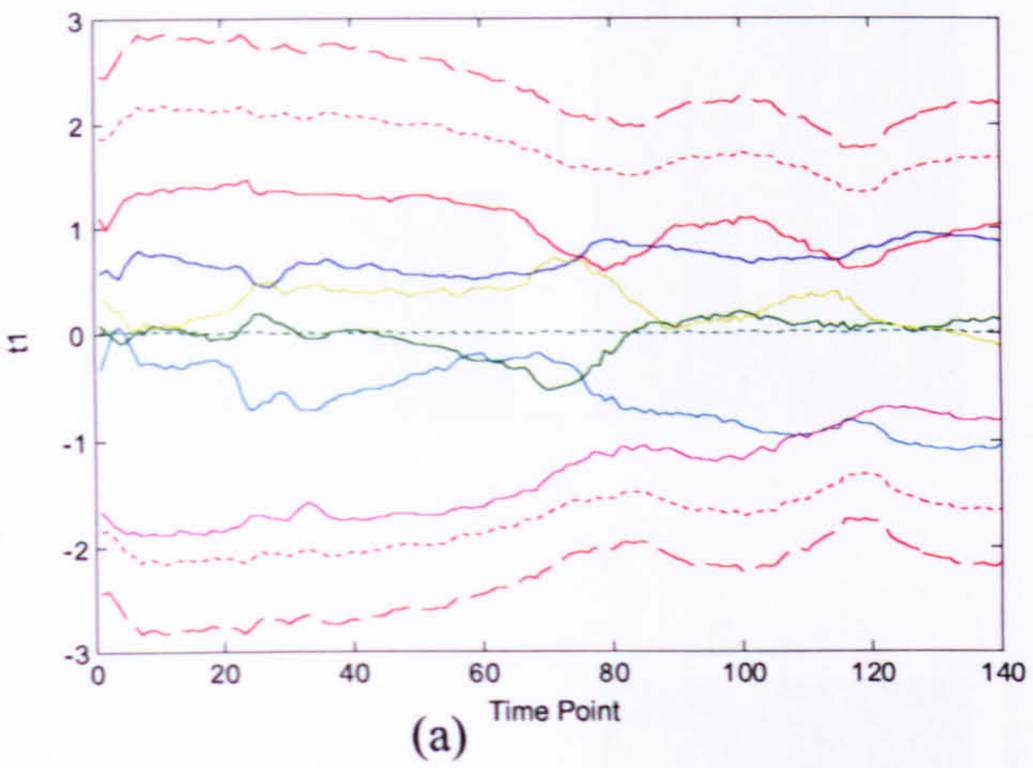
The interpretation of this model can be considered at two levels, the sub-level where the original information pertaining to the process and spectral data are captured by the block scores and the super level is where the combined effect is observed at the super scores. Batches 1 to 6 were used for building the reference model. It should be noted that the super scores are mutually orthogonal however the block scores are not orthogonal. Therefore the resulting super scores explain the independent variation observed in the super block and describe the largest variance in the first principal component and so on. A mixed variation is observed in the block principal component scores. Four principal components are selected by cross-validation for the model which explains 88% of the total variation. Table 6-5 shows the amount of variance captured by each of the principal component.

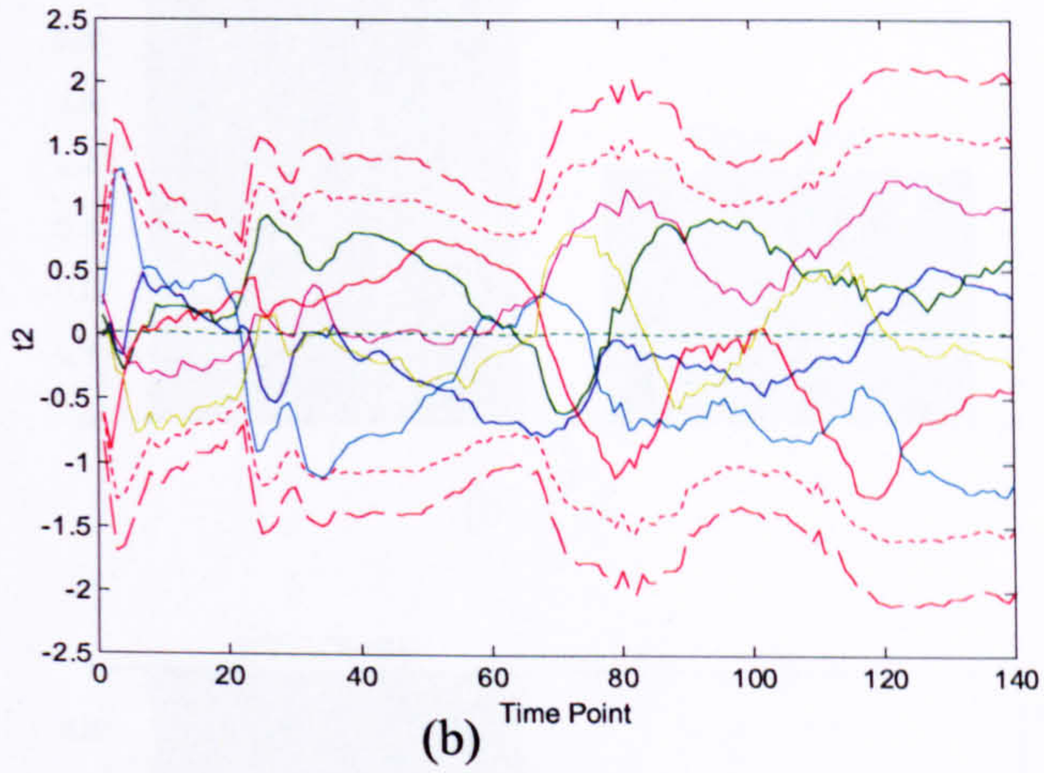
Principal component	Super block	
	Individual % variance captured	Cumulative % variance captured
1	47.93	47.93
2	20.68	68.6
3	11.56	80.16
4	7.47	87.63
5	5.57	92.65
6	4.16	96.81
7	2.24	99.05
8	0.64	99.69

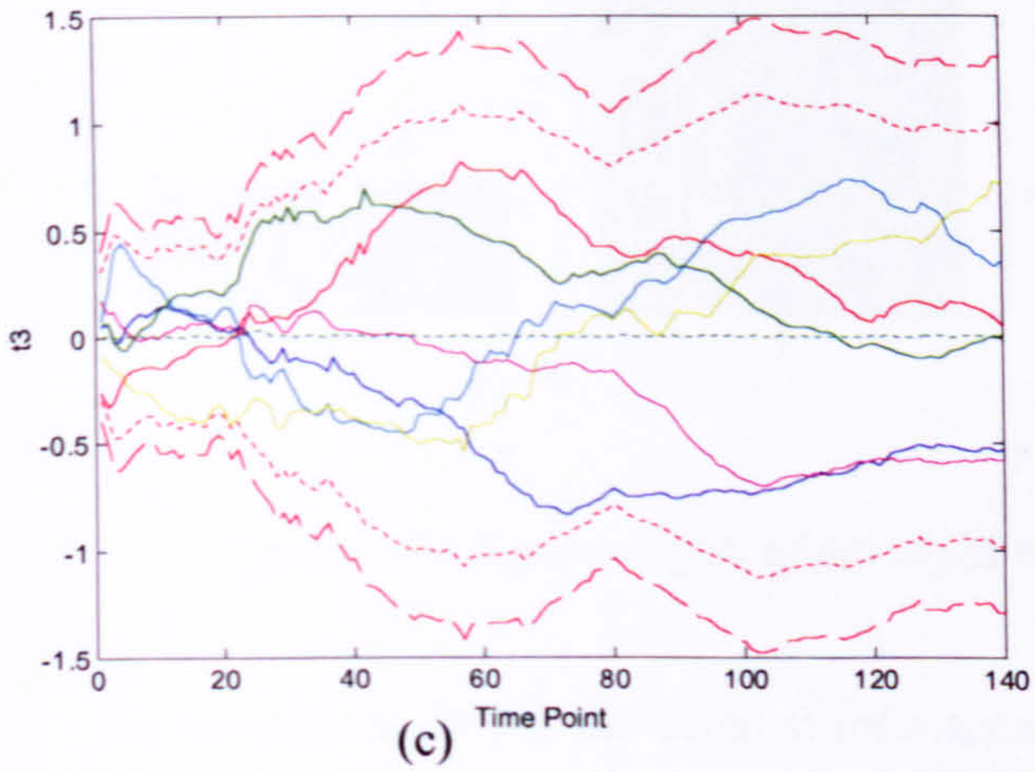
Table 6-5 Percentage of variance explained by CPCA for the integrated multi-block model

The super scores of principal component one to four for the nominal batches are shown in Figure 6-13. The scores are randomly distributed in the monitoring charts revealing the normal batch-to-batch variation is captured. The relevant super weights provide the block variability information of each principal component.

Figure 6-14 reveals that the spectral block (block 2) is more significant at principal component one and three whilst the process block (block 1) is more significant at principal component two and four.







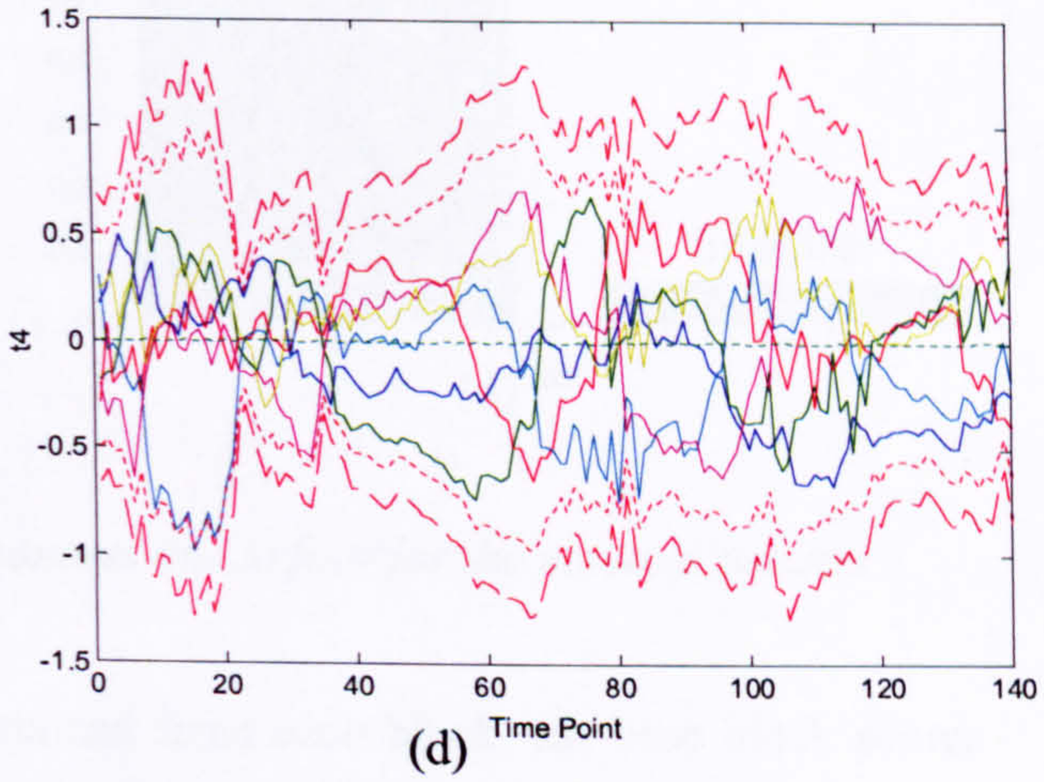
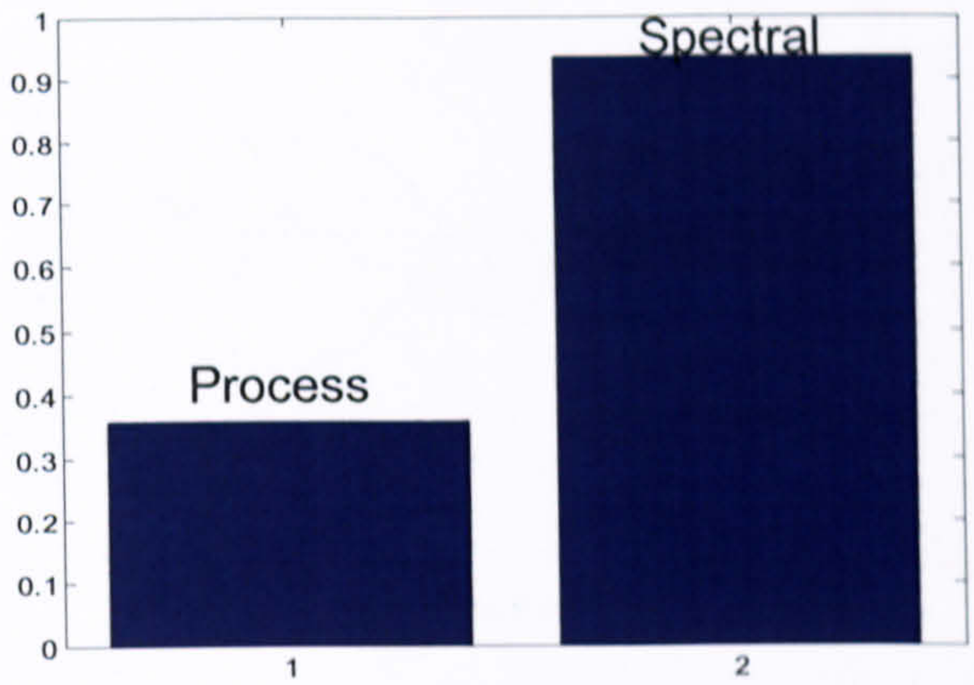
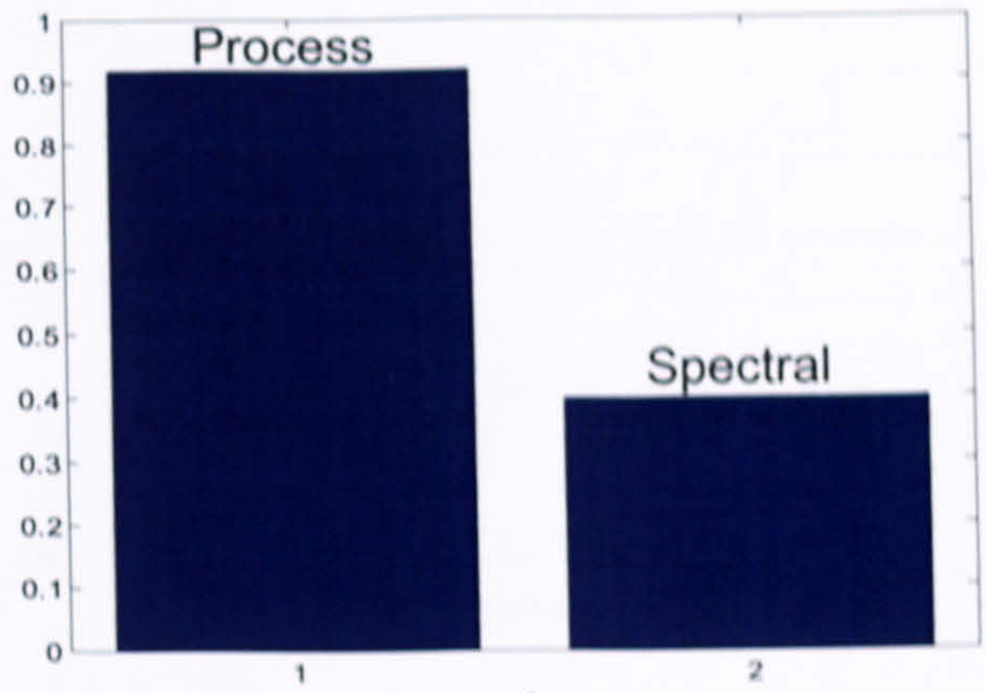


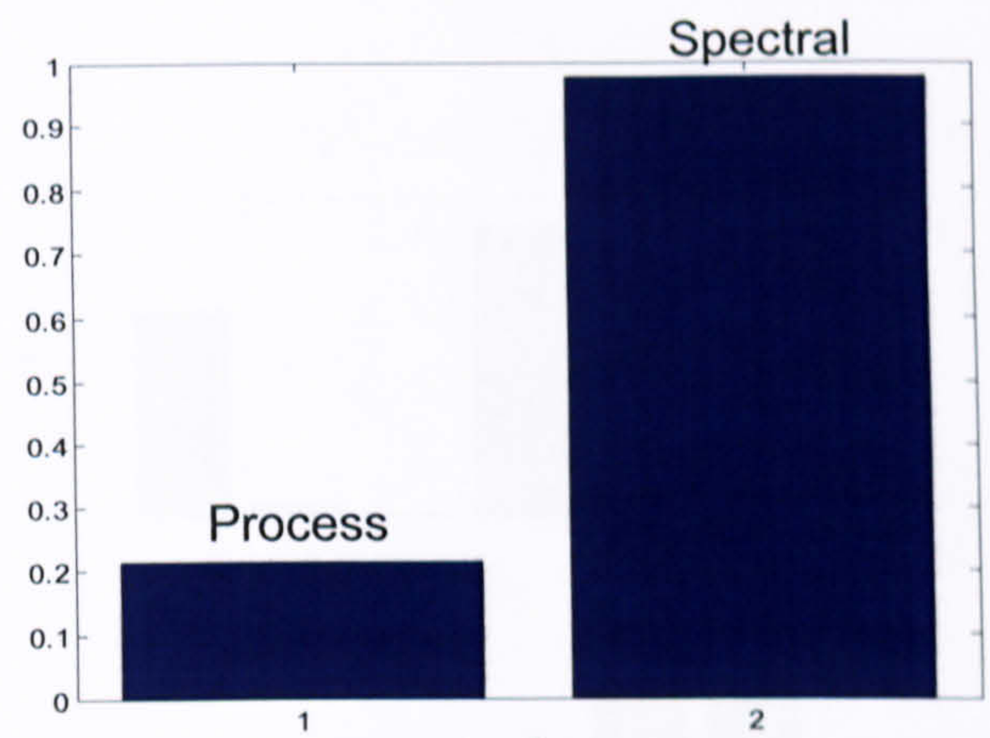
Figure 6-13 Super scores of principal components one to four for the nominal batches



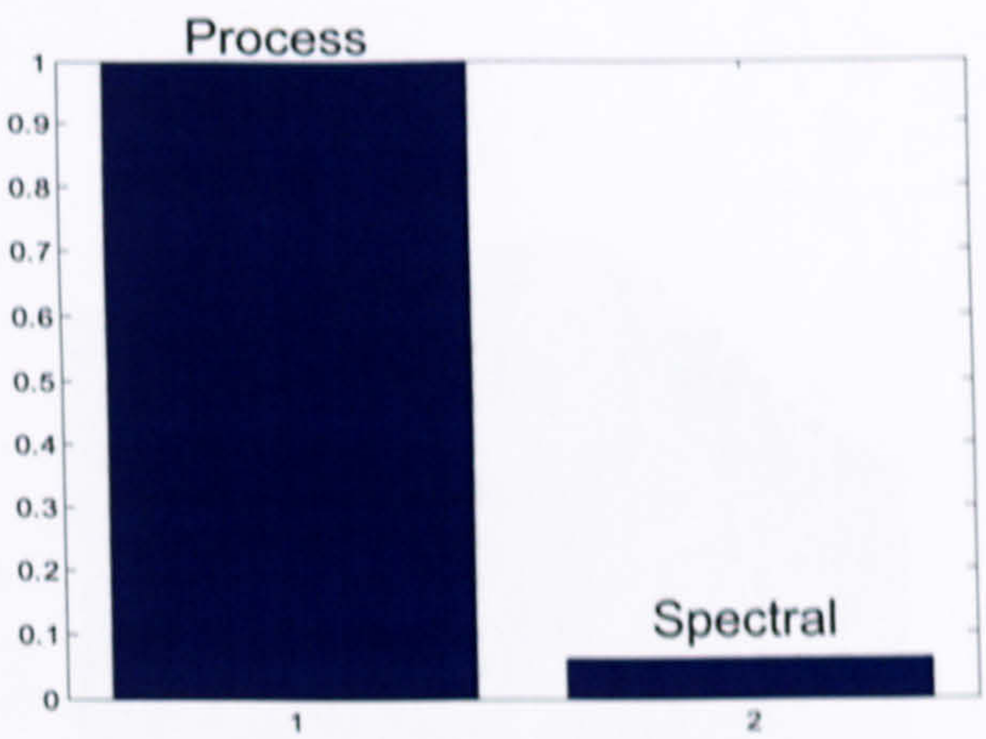
(a) p_1



(b) p_2



(c) p_3



(d) p_4

Figure 6-14 Super weights of principal components one to four for the nominal batches

To investigate further the detailed information attained from each block, the base block scores and base block loadings can be interrogated. Figure 6-15 illustrates the base block scores of principal component one in which Figure 6-15(a) shows the block scores for block 1 (process variables) and Figure 6-15(b) shows the block scores for block 2 (spectral block). By comparing the two monitoring charts to the individual models, a similar pattern is observed. This has further confirmed the research reported by Westerhuis *et al*. (1998) and Qin *et al*. (2001) that the results of CPCA can be calculated from the regular PCA method. The same situation applies also to the block loadings calculation. The block loadings of principal component one are shown in Figure 6-16. The block loadings are similar to those attained from the individual nominal representations.

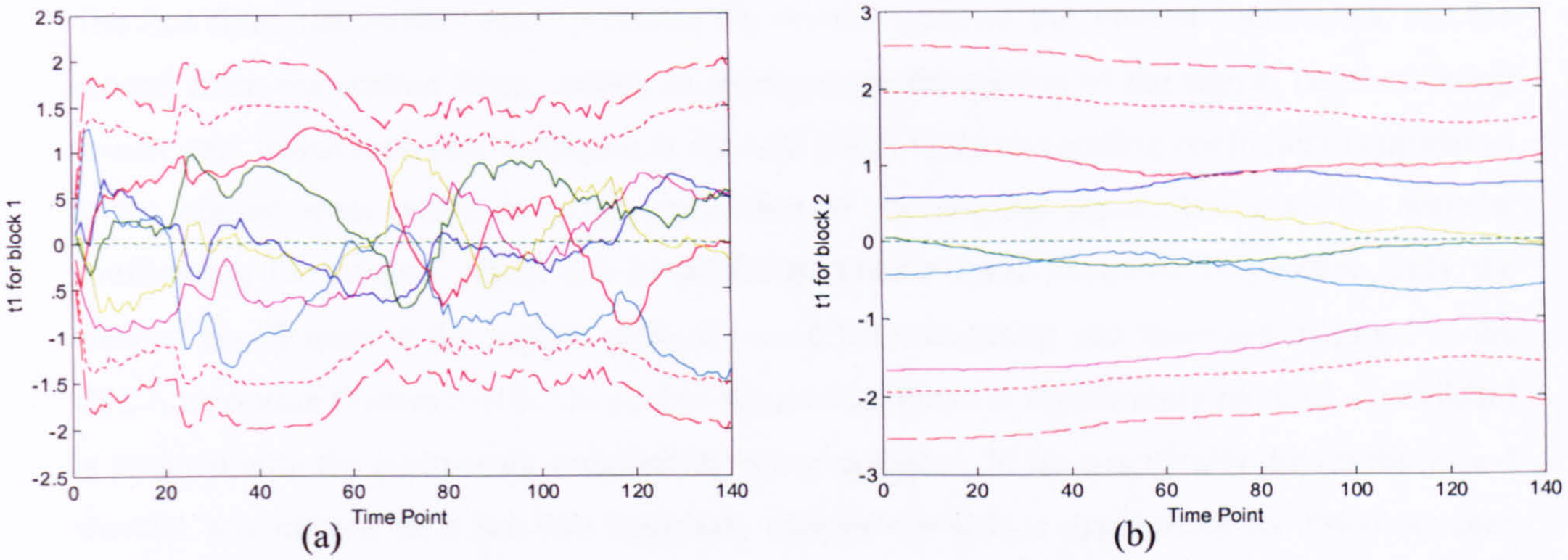


Figure 6-15 Block scores of principal component one of nominal batches

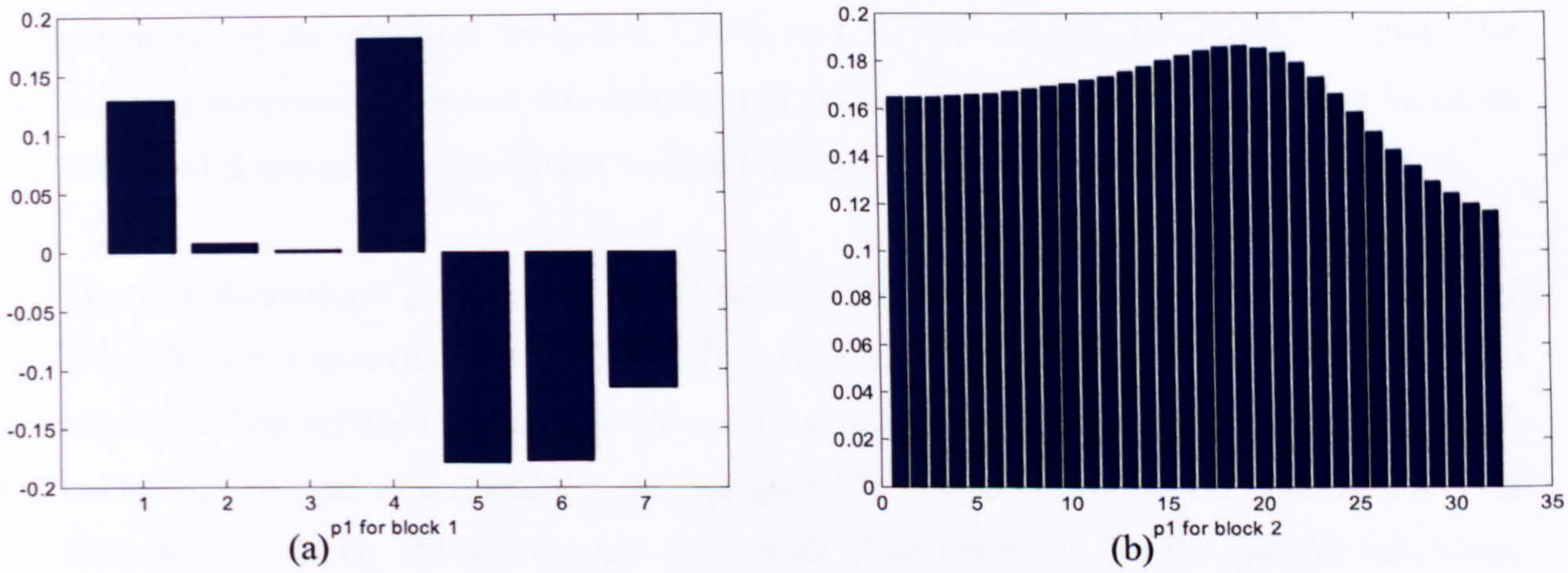


Figure 6-16 Block loadings of principal component one of nominal batches

6.5.4 Integrated Multi-block PCA Model Incorporating Wavelets

For the second approach, integration by multi-block analysis is performed as per the first approach but an additional step is included in terms of the pre-processing of the spectral data. The spectral data is decomposed using the discrete wavelet transform with the original signal being recursively decomposed at a resolution differing by a factor of two from the previous step. The signal thus requires to be of length 2^n, where n is an integer and hence the spectral data have to be padded to the nearest 2^n. In the application described, the UV-Visible spectrum comprises 216 wavelengths, thus it is padded to 2^8, i.e. 256. No pre-selection of range of interest is introduced.

A Finite Impulse Response (FIR) filter which comprises two filters is applied to the spectral data. The first filter, the wavelet filter, produces the detailed part, i.e. the wavelet coefficients, and the second filter, the scaling filter, creates an approximate description of the signal, i.e. the scaling coefficients which represent the signal at the next scale. Only one scaling coefficient is produced at the highest scale which is an approximation to the original signal. Using all the wavelet coefficients, the original signal can be perfectly reconstructed. However in practice, only the wavelet coefficients at the highest scale are used for monitoring and these are included in the CPCA algorithm (Figure 5-10). The size of the spectral block is significantly reduced as the detail is retained with the multi-scale components being extracted. More specifically the Daubechies-4 wavelet was chosen as it has two vanishing moments which is appropriate for low-frequency signals such as UV-Visible spectra. Five decomposition levels were selected hence the original spectral data was compressed to 6.4% its original size. The number of variables (wavelengths) for the spectral data is significantly reduced from the original number of wavelengths, i.e. 216 to 14 wavelet coefficients, resulting in the data being compressed 15-fold. Figure 6-17 provides a schematic of the proposed integrated CPCA and wavelet on-line monitoring scheme. The proposed monitoring approach was described in Section 3.3 and the CPCA algorithm based on the NIPALS engine was described in Section 5.2.2.

The three dimensional process and spectral matrices are treated into a separate block, $\underline{\mathbf{X}}_1$ (batch (I) x variable (J) x time (K)) and $\underline{\mathbf{X}}_2$ (batch (I) x wavelength (J) x time (K)) respectively. The two blocks are then unfolded according to Nomikos and MacGregor approach into the form $\mathbf{X}$ (I x KJ) before they are centred and scaled to unit variance. The unfolded matrices are rearranged into the form $\mathbf{X}$ (IK x J), i.e. utilising the part from Wold *et al.* approach. For the spectral data block, wavelet analysis is applied resulting a new matrix of wavelet coefficients. CPCA is then applied to the process block and the wavelet coefficients block. A new matrix of super block, $\mathbf{T}$, is formed. Standard PCA is applied to the super block and the resulting super scores, $\mathbf{T}_T$, and the block scores from both blocks, $\mathbf{T}_1$ and $\mathbf{T}_2$, describe the variance of the variations for the through batch mean trajectories. The statistical control limits for the scores monitoring chart are derived by calculating the mean and standard deviation from the scores matrices that were rearranged into the form of $\mathbf{T}_T$ (I x KR), $\mathbf{T}_1$ (I x KR) and $\mathbf{T}_2$ (I x KR).

To clarify further the data pre-processing procedures, the general summary is re-defined for this approach:

1. Collect engineering process and spectral data of nominal batches. Treat the process data as a block and the spectral data as a second block.

2. Attain samples at the same time points for the process and spectral data, i.e. interpolate the spectral data every 10 seconds to align with the process data.
3. Check for missing data and apply data pre-processing techniques as necessary, see Section 3.2.1.
4. Apply batch length alignment, i.e. cutting to minimum batch length, see Section 3.2.3.
5. Apply baseline correction method to adjust spectral baseline, see Section 5.4.4.
6. For spectral data, apply linear padding to the nearest dyadic length, i.e. 2^8.
7. For spectral data, perform Discrete Wavelet Transform with selected wavelet and decomposition levels, see Section 5.3.3.2.
8. Unfold the three-way matrices $\underline{\mathbf{X}}$ (I x J x K) to $\mathbf{X}$ (I x KJ).
9. Apply auto-scaling to process and spectral data, see Section 3.2.2.
10. Apply weighting factor to achieve equal variance, see Section 3.2.2.
11. Re-arrange the unfolded matrices $\mathbf{X}$ (I x KJ) to $\mathbf{X}$ (IK x J), see Section 3.3.

Four principal components are selected by cross-validation for the nominal model explaining 81% of the total variation. Table 6-6 shows the amount of variance captured by each of the principal component.

Principal component	Super block	
	Individual % variance captured	Cumulative % variance captured
1	36.66	36.66
2	26.08	62.73
3	11.25	73.98
4	7.46	81.44
5	6.41	87.87
6	4.38	92.25
7	3.07	95.32
8	2.35	97.67

Table 6-6 Percentage of variance explained by CPCA for the integrated multi-block model with wavelets

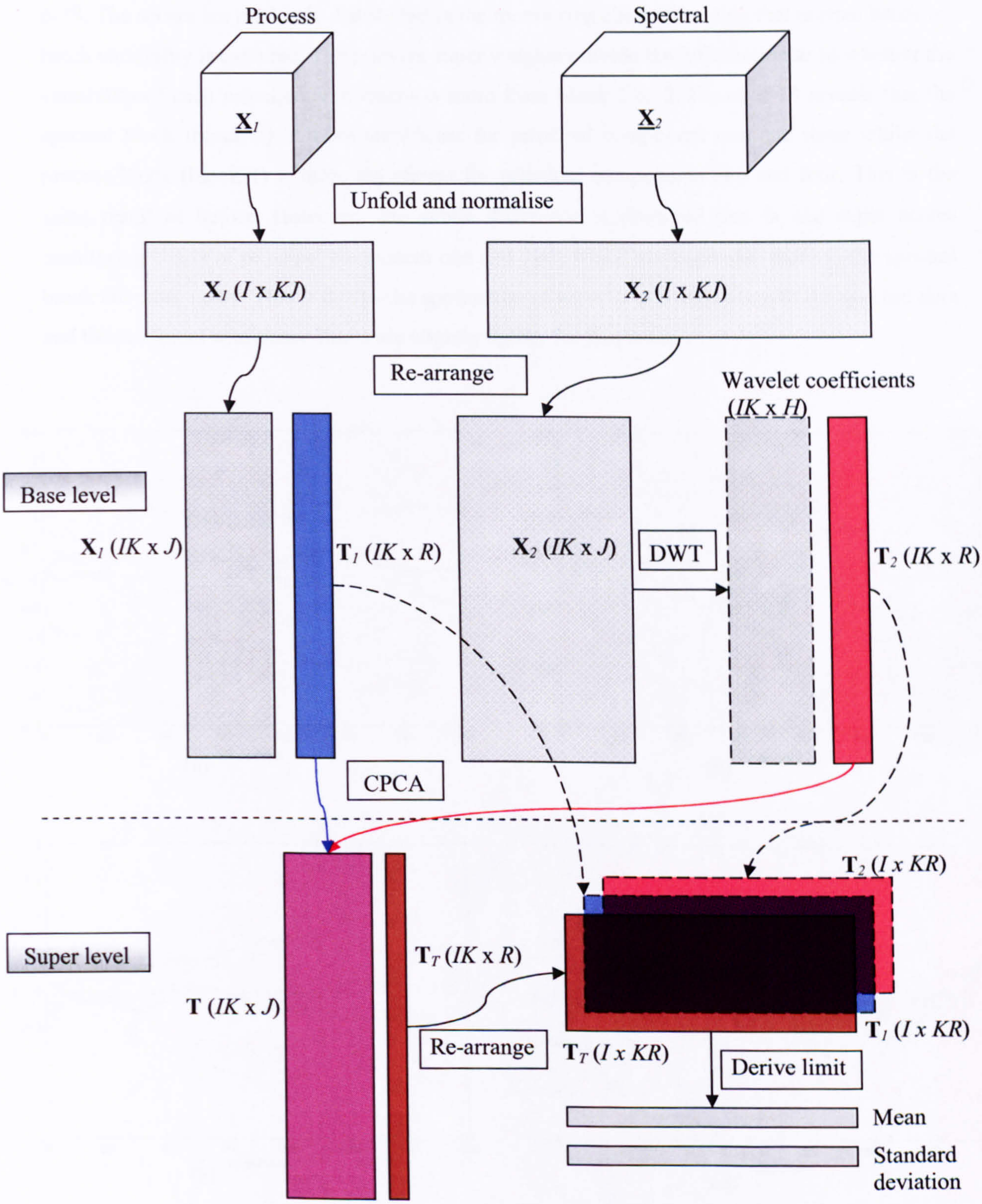


Figure 6-17 Proposed integrated CPCA and wavelets on-line monitoring scheme

The super scores of principal component one to four for the nominal batches are shown in Figure 6-18. The scores are randomly distributed in the monitoring charts revealing that normal batch-to-batch variability is captured. The relevant super weights provide the information as to whether the variability of each principal component is more from block 1 or 2. Figure 6-19 reveals that the spectral block (block 2) is more significant for principal component one and three whilst the process block (block 1) is more significant for principal components two and four. This is the same result as before. However, one subtle difference is observed that is, the super scores monitoring charts of principal component one and three which corresponded more to the spectral block are more noisy. This is due to the application of wavelet decomposition to the spectral data and the statistical confidence limits are slightly tighter for this model.

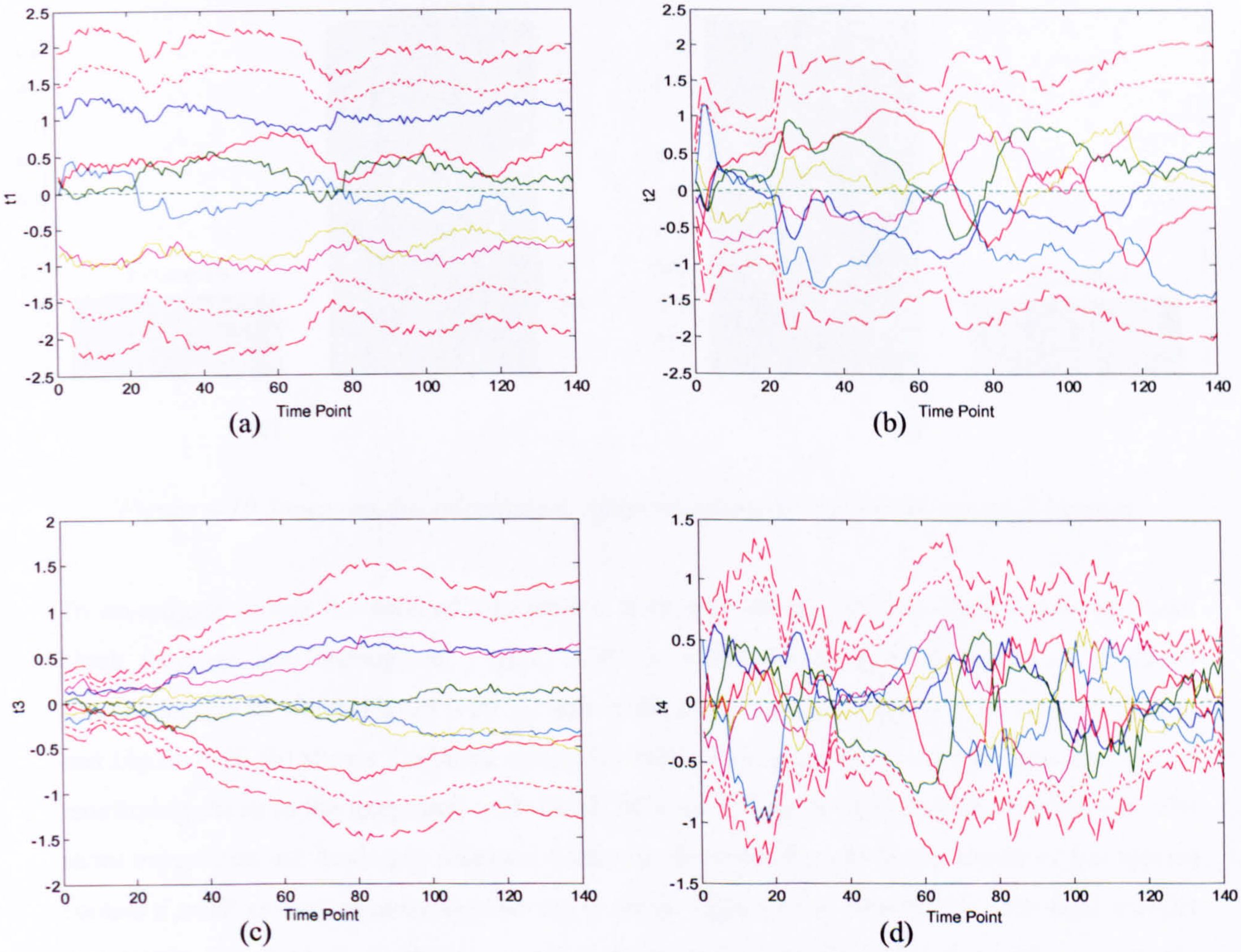


Figure 6-18 Super scores of principal components one to four for the nominal batches

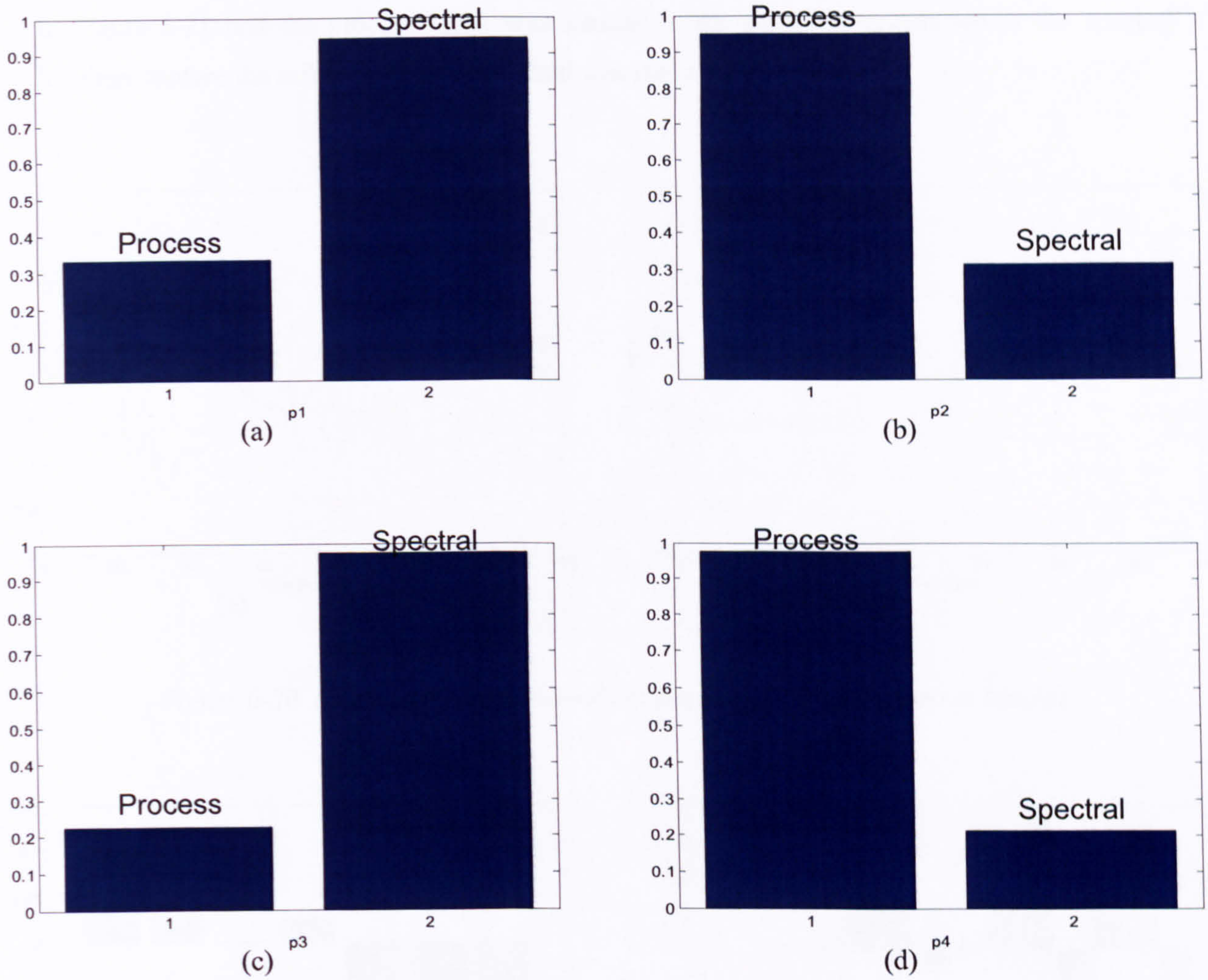


Figure 6-19 Super weights of principal components one to four for the nominal batches

To investigate further the detailed information from each block, the base block scores and base block loadings are interrogated. Figure 6-20 illustrates the base block scores of principal component one in which Figure 6-20 (a) shows the block scores for block 1 (process variables) and Figure 6-20 (b) shows the block scores for block 2 (spectral block). By comparing the two monitoring charts to the integrated multi-block PCA model, no major difference is observed. The score trajectories are randomly scattered in the model space but the block scores of the spectra contain a small amount of noise as observed from the super scores. Overall the additional wavelet step results in a final model that is almost identical to the integrated CPCA model however the number of variables required in the model is less than the integrated CPCA model. The number of wavenumbers in the spectral block was chosen to be reduced for the integrated CPCA model hence a direct comparison cannot be effective. Nevertheless, this has further confirmed that the experience and knowledge from experts is crucial to the success of establishing a monitoring scheme since the range of interest of for the wavelengths is captured in the analysis resulting the

noise being excluded from the model. The block loadings of principal component one are shown in Figure 6-21 and the process block was similar to the previous models whilst the spectral loadings capture the influence from individual wavelet coefficients.

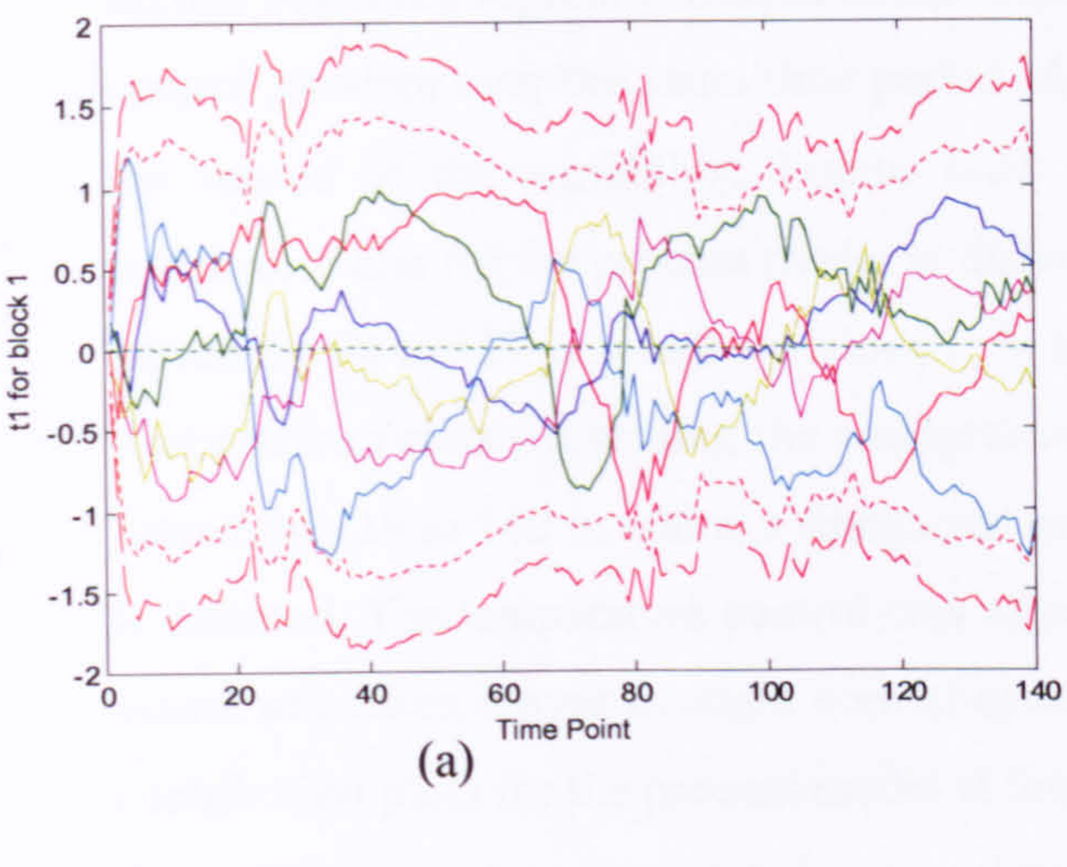


Figure 6-20 Block scores of principal component one for the nominal batches

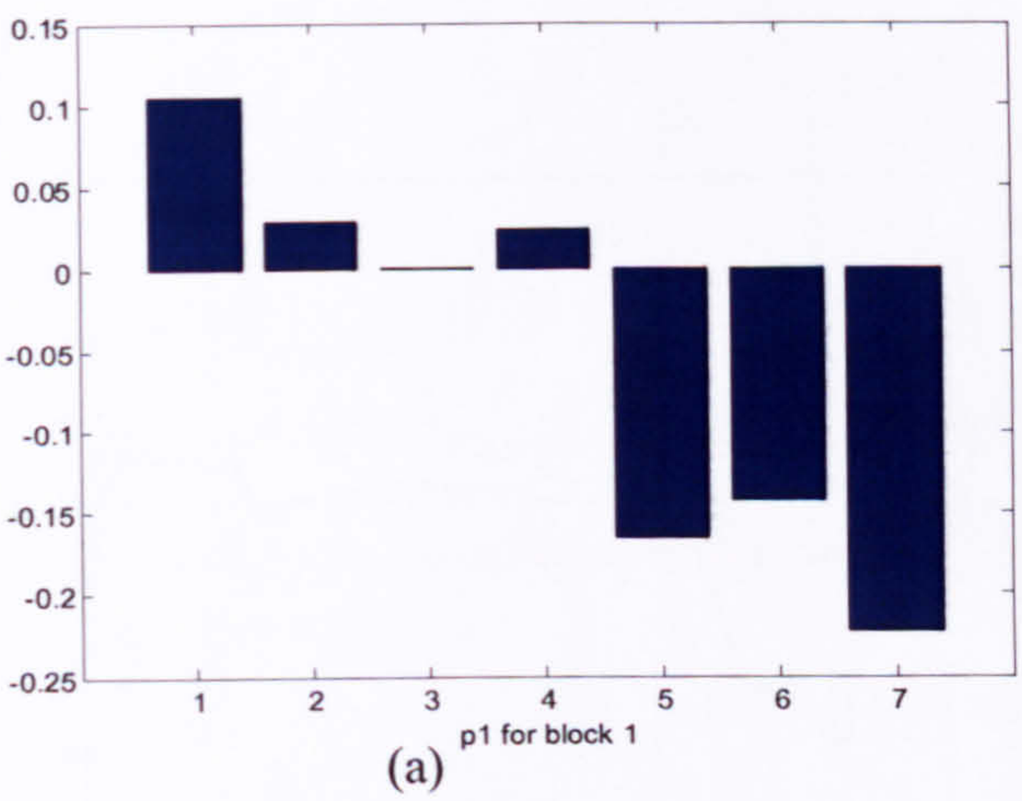
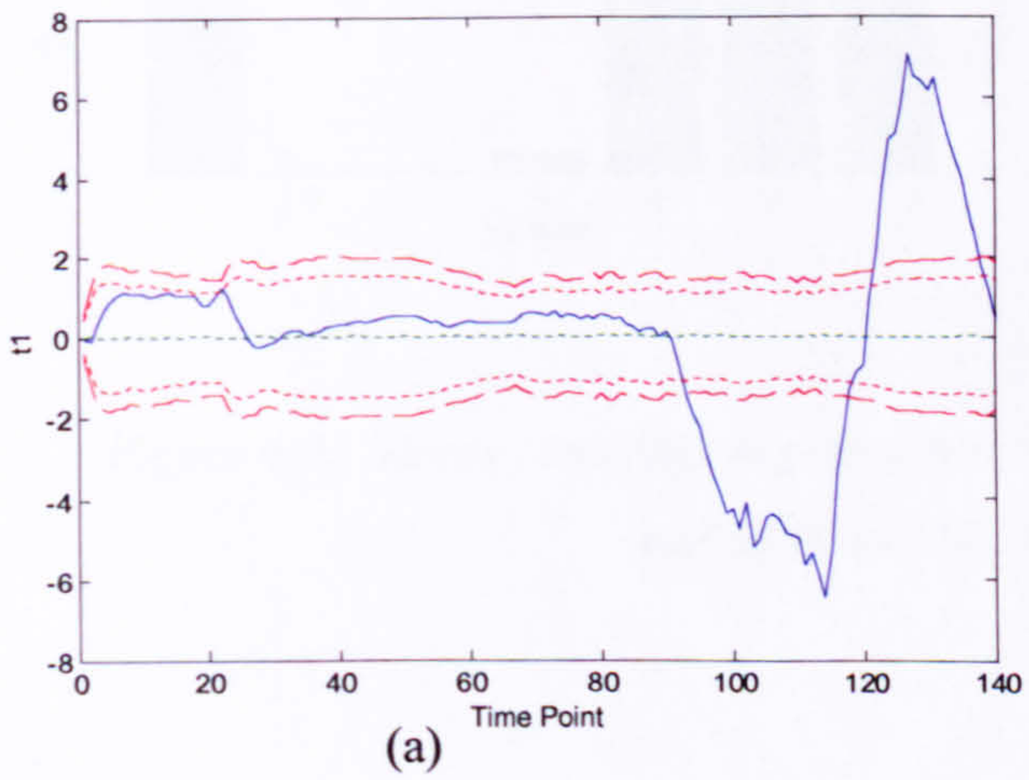


Figure 6-21 Block loadings of principal component one for the nominal batches

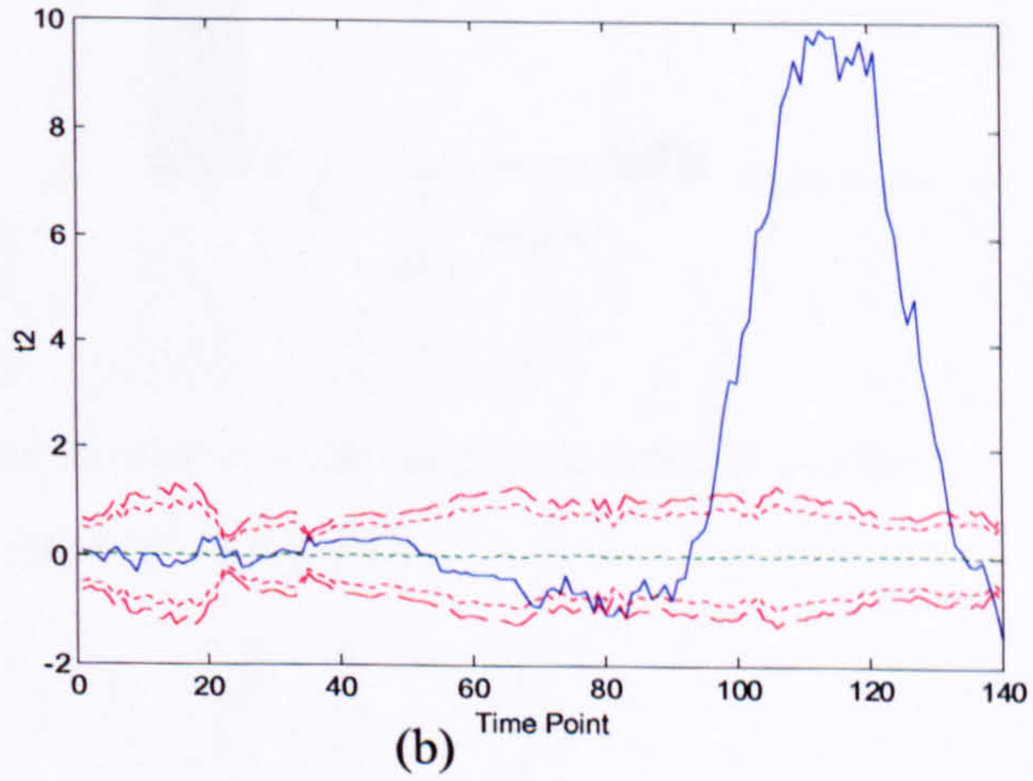
6.6 Analysis of Batch 11

Batch 11 is one of the abnormal batches that consists of catalyst discharge and a series of temperature control problems. This batch is projected onto the individual and integrated models to evaluate their performance with respect to monitoring and fault detection. The individual model monitoring charts again form the benchmark.

The process scores of principal component one to three for batch eleven are shown in Figure 6-22. Interrogating principal component one in Figure 6-22(a) reveals that the cooling down of the temperature is detected at time period 90 to 117 while when reheating the vessel back to 40°C, there was an overheating period detected from time period 118 before it dropped back to the normal operating region. Principal component two and three both detected a similar temperature control problem over the same time period. A scores contribution plot is then utilised to identify the source of the variability. Figure 6-23 shows the scores contribution plots of principal component one for the process model at different time periods. Figure 6-23(a) illustrates the time period for 90 to 117 in which variables 1, 5, 6 and 7 are out of statistical control and they are the temperature sensors revealing the temperature probe adjustment. Figure 6-23(b) shows the time period for 119 to 140 in which a significant contribution from the reactor temperature (variable 1) is detected. The temperature control was adjusted, back to normal operating region, however the vessel was taken longer to attain normal operating temperature. Figure 6-24 illustrates the scores contribution plots for the process model at time period 90 to 140 for principal component two and three. A temperature sensor problem was detected confirming the out-of-control signal was due to temperatures. The change in catalyst discharge is not captured from any of the process representations.



(a)



(b)

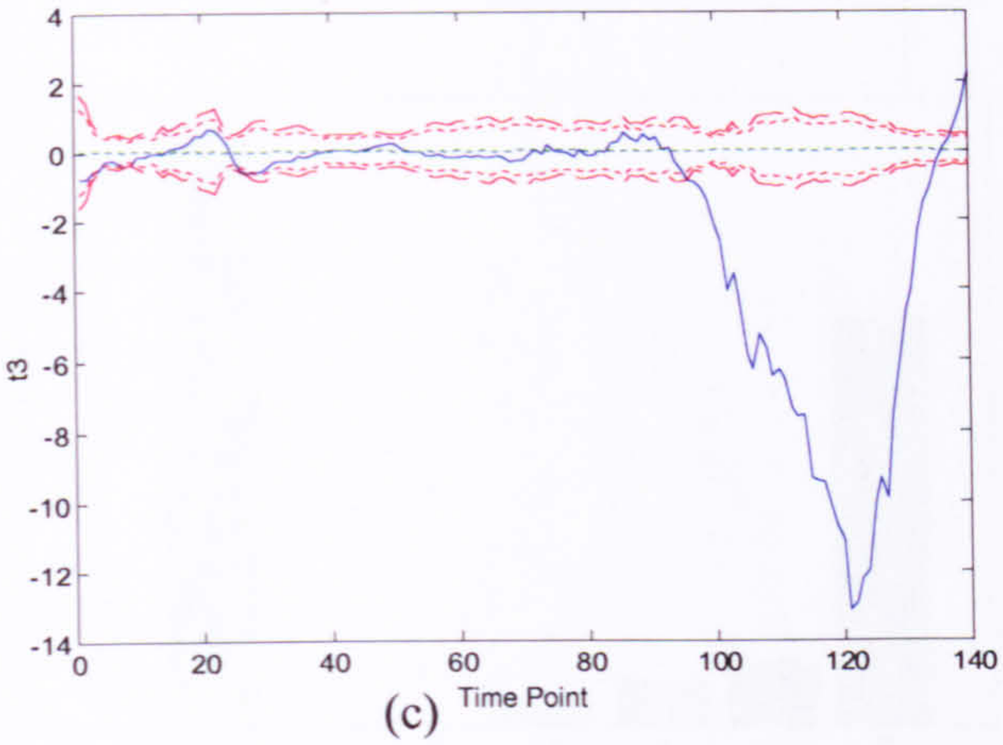


Figure 6-22 Scores of principal components one to three for the individual process model

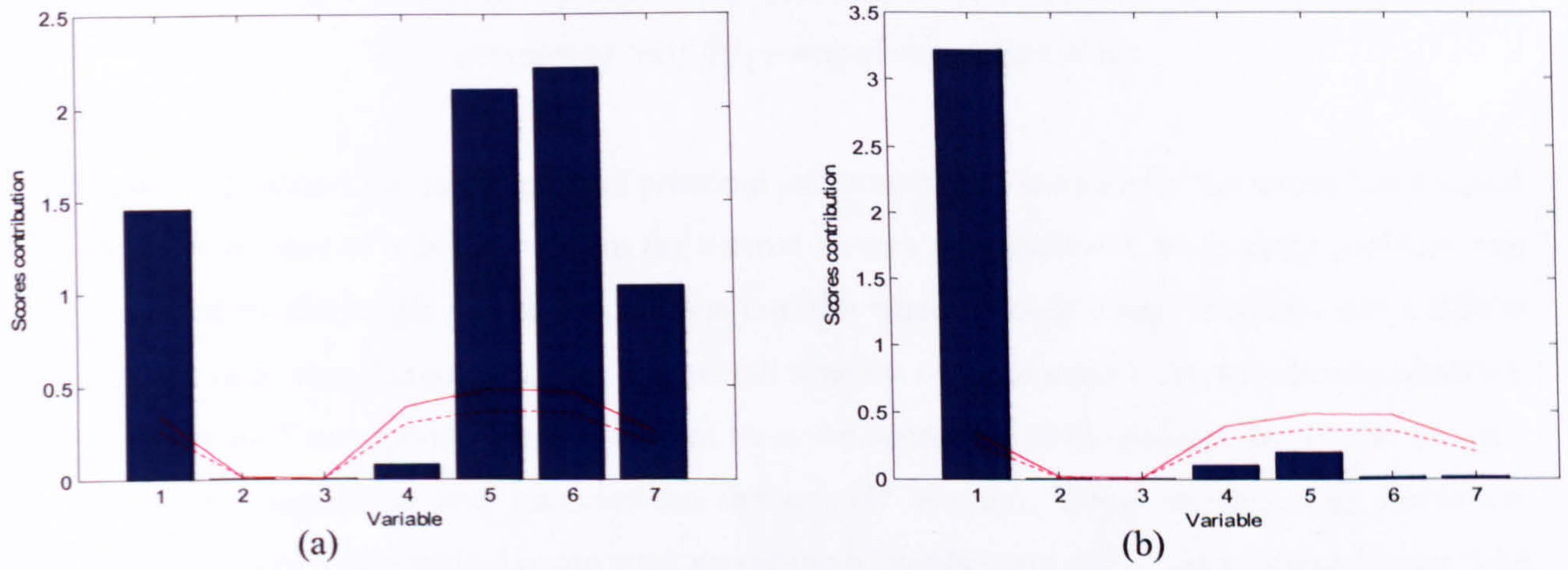


Figure 6-23 Scores contribution plot of principal component one for the process model: (a) time period 90 to 117; (b) time period 119 to 140

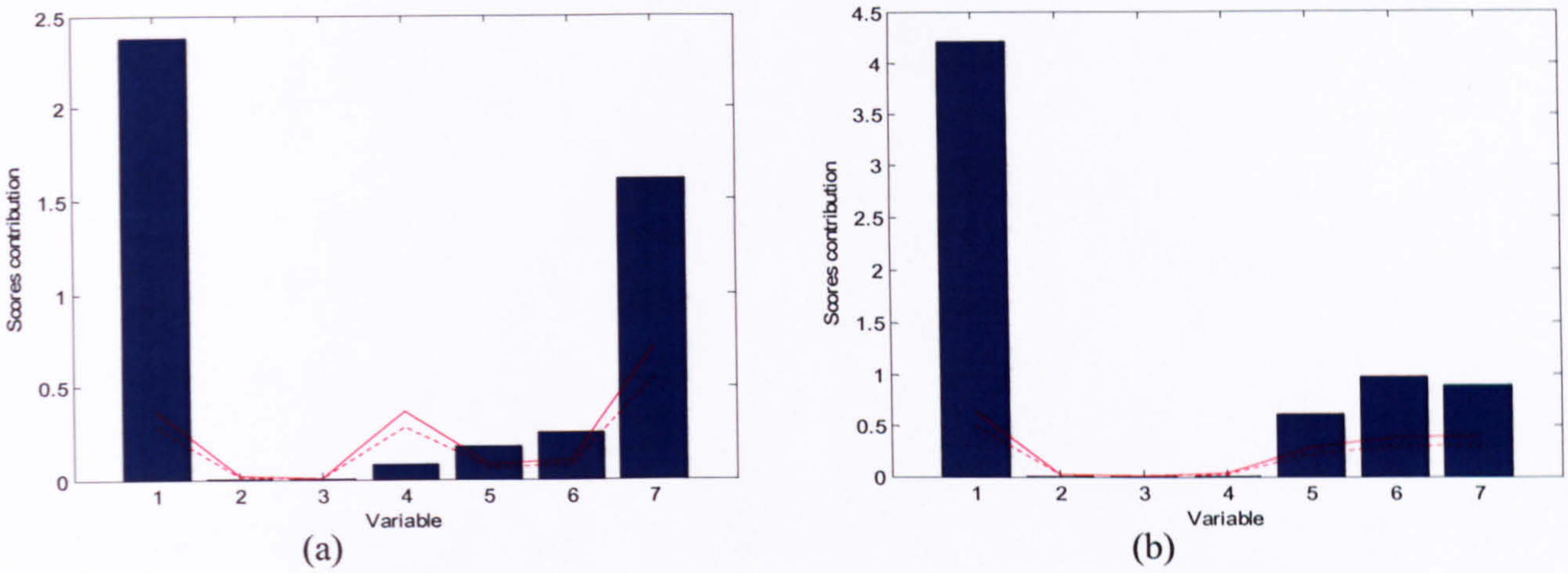


Figure 6-24 Scores contribution plot for the process model at time period 90 to 140: (a) principal component two; (b) principal component three

Figure 6-25 shows the scores plots of principal component one and two for the individual spectral model. Evidence of a deviation from the normal spectra was observed. A charging problem was simulated by discharging 10% less catalyst into the vessel therefore one would expect a slower reaction occurring thereby affecting the overall kinetics of the reaction. The trajectories observed in Figure 6-25 are out-of-statistical control from the beginning of the reaction hence the catalytic effect was significant and impacted on the overall reaction. When interrogating the scores contribution plot of principal component one at the beginning and end of the reaction, Figure 6-26, there is no significant difference throughout the reaction but a higher contribution is observed between wavelength data point 20 to 25. This has revealed that the rate of conversion from reactant to product changed as the data point 20 to 25 reflect the slope between the peaks of aniline and nitrobenzene. Scores contribution plot of principal component two represent the peaks of reactant and product and reflect the change in height (Figure 6-27). No physical disturbances appear to be detected from its spectral model thus one can argue that the individual models allow the source of the errors to be identified independently however the integrated approach will be observed to have the capability to summarise the overall effects in a single representation.

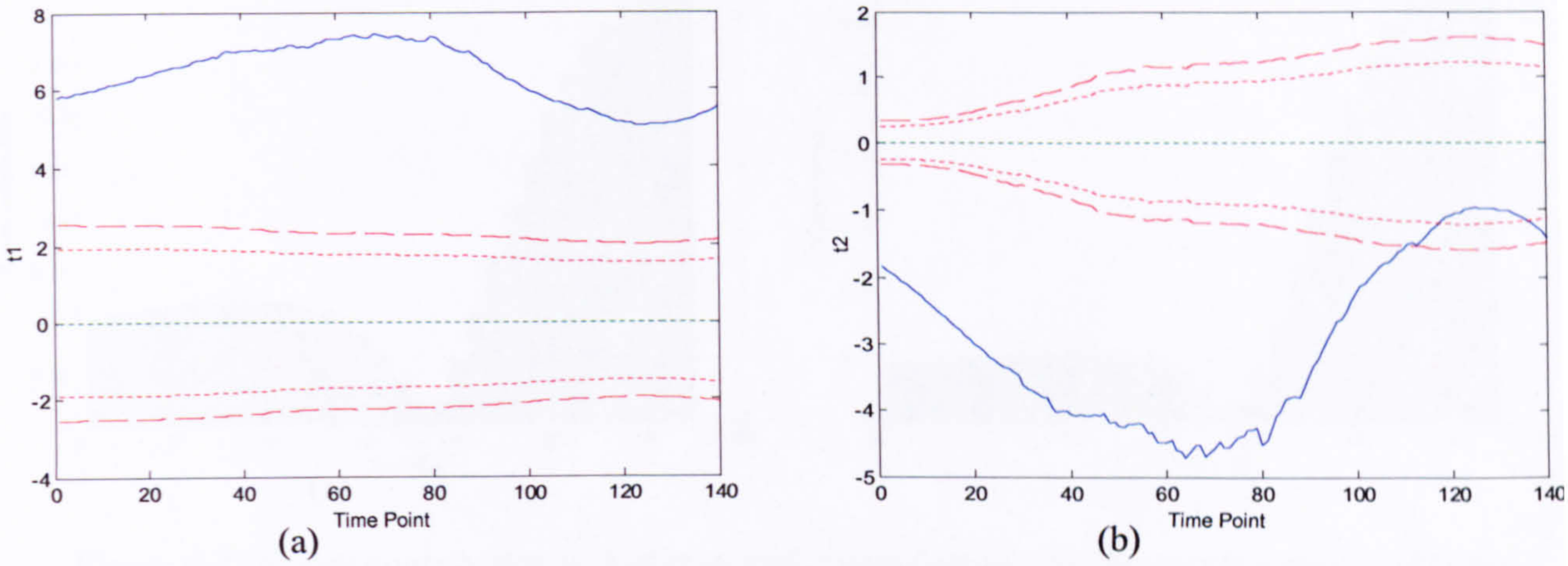


Figure 6-25 Scores of principal components one and two for the individual spectral model

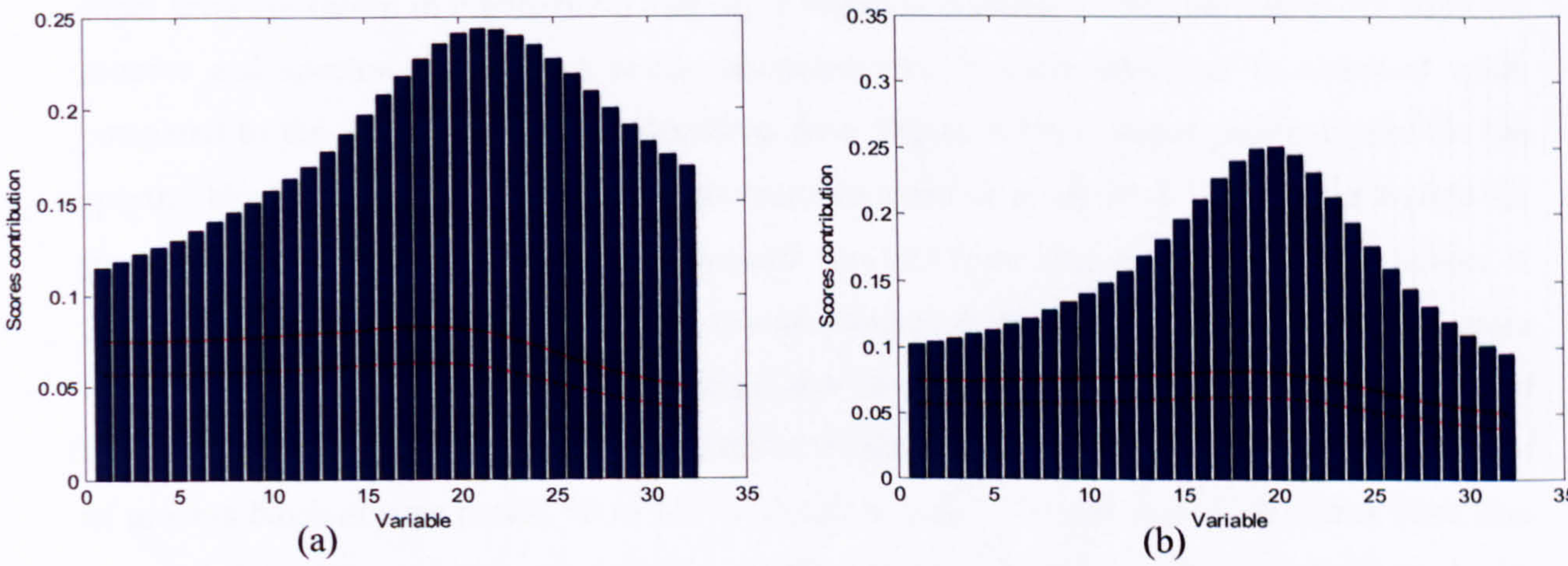


Figure 6-26 Scores contribution plot of principal component one for the spectral model: (a) time period 1 to 10; (b) time period 120 to 140

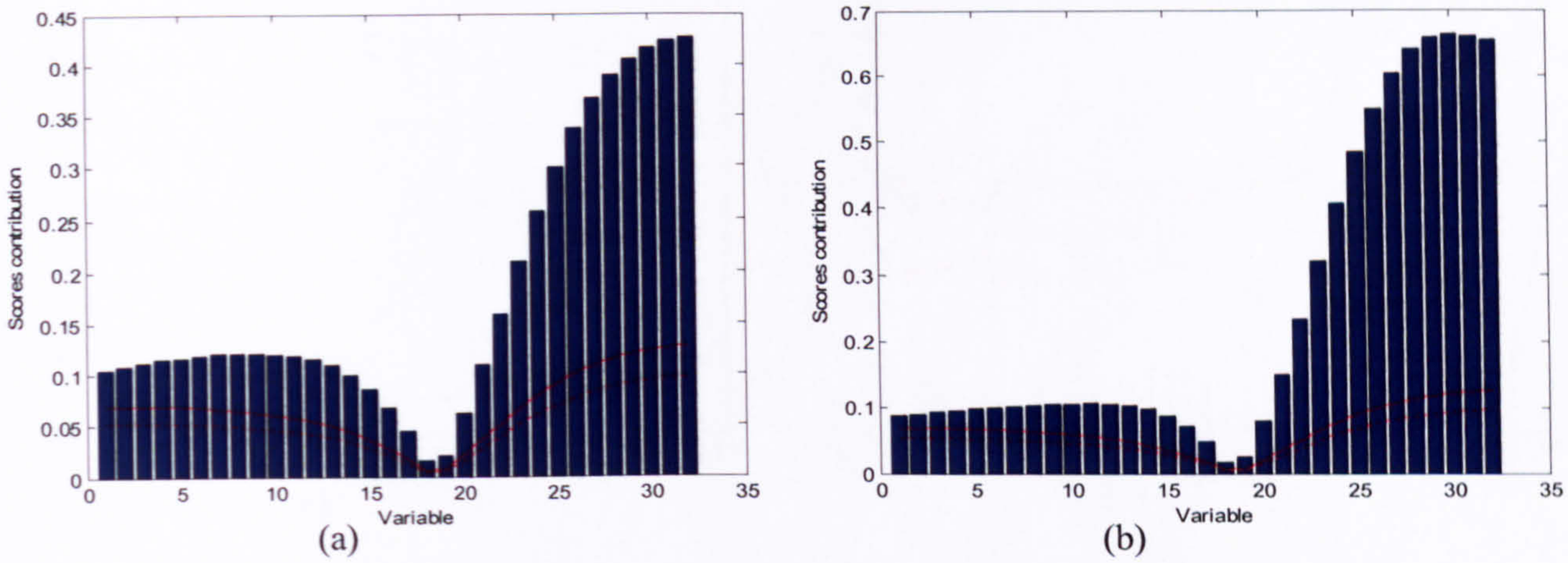


Figure 6-27 Scores contribution plot of principal component two for the spectral model: (a) time period 1 to 10; (b) time period 60 to 80

For the integrated multi-block PCA model, the scores of principal component one to four at the super level are shown in Figure 6-28. The super scores summarise clearly the deviations from the process and spectral blocks in a single representation. Similar behaviour is observed when compared to the individual models. Recalling from Figure 6-14, a higher super weight for the spectral block is observed to principal component one and three whilst a higher super weight for process block is attained to principal component two and four. This explains the super scores of principal component two and four having similar patterns as the process blocks. The temperature control problem can be identified from the process block scores of principal component one and two in Figure 6-29(a) and Figure 6-29(c) with verification using the block scores contribution plot of process block at time period 90 to 117 in Figure 6-30(a). The spectral block scores have also revealed the catalyst charging problem with the trajectory being out of the confidence limits exhibiting atypical behaviour of nitrobenzene hydrogenation as observed from the high contribution around the nitrobenzene peak in the spectral block scores contribution plot shown in Figure 6-30(b).

A process deviation can be observed at the start of the process and hence appropriate corrective action could have been taken to diagnose the possible source of the fault and if necessary terminate the process. Without the integration of the spectral information, process behaviour would not be questioned until a physical disturbance had been observed around time point 90. This has shown the importance of integrating the process and spectral information for on-line process monitoring.

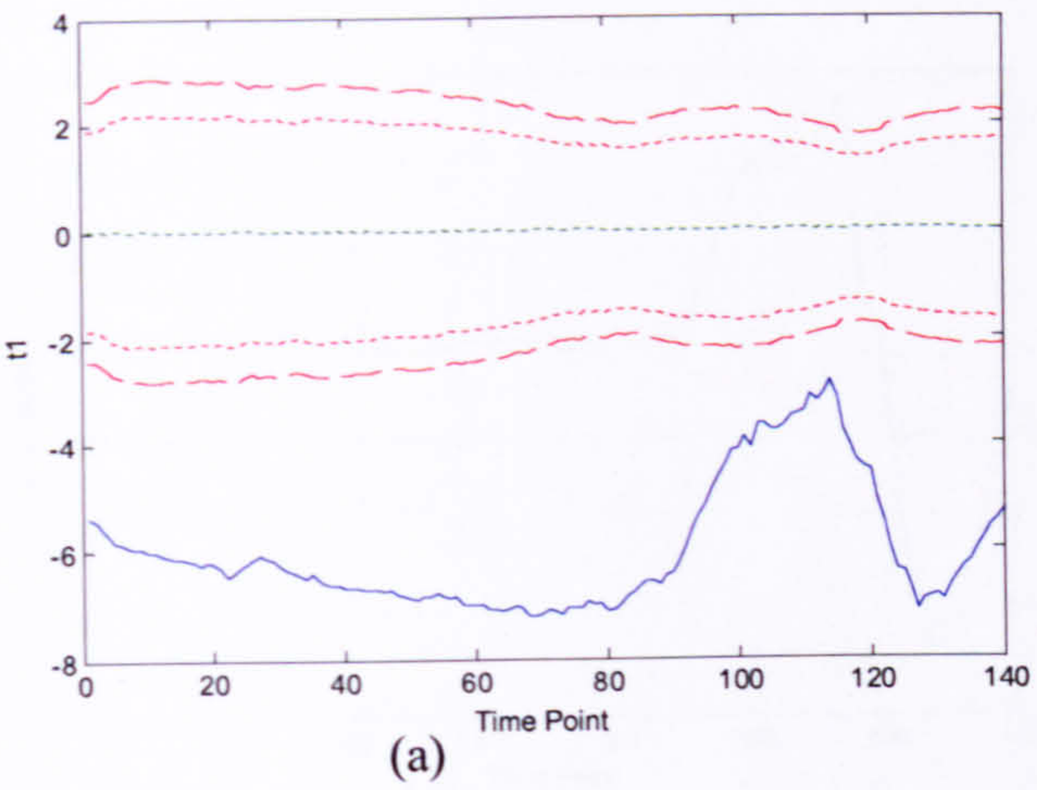
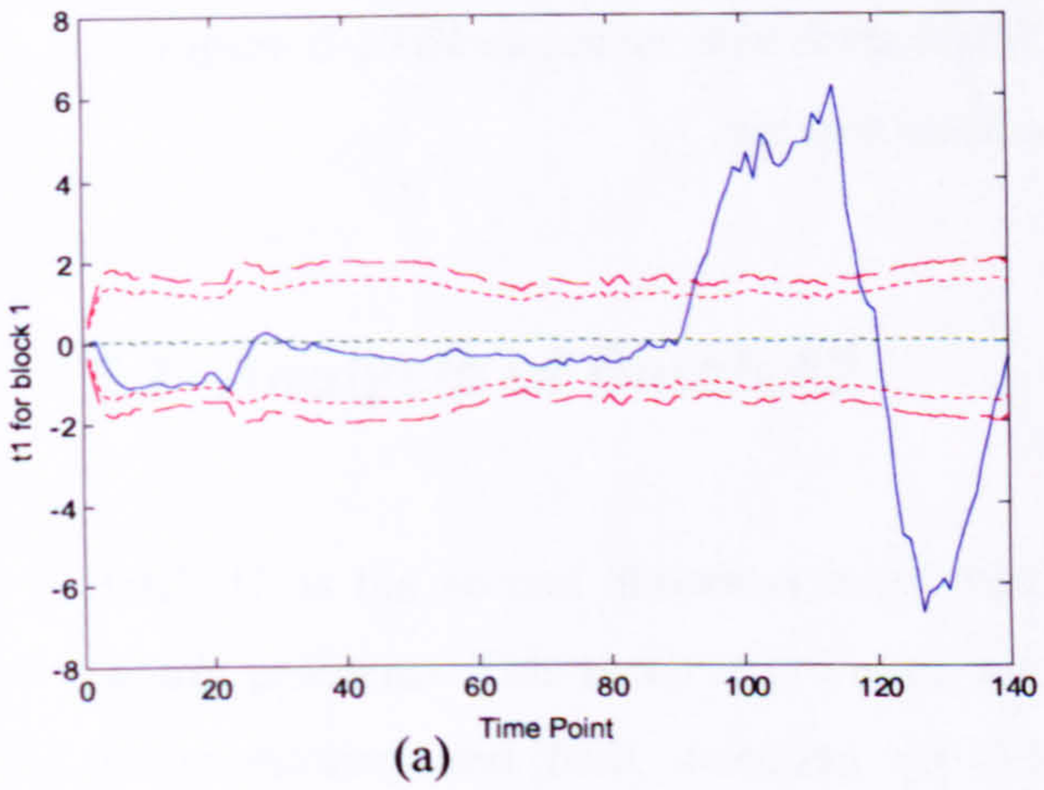


Figure 6-28 Super scores of principal components one to four for the integrated multi-block model



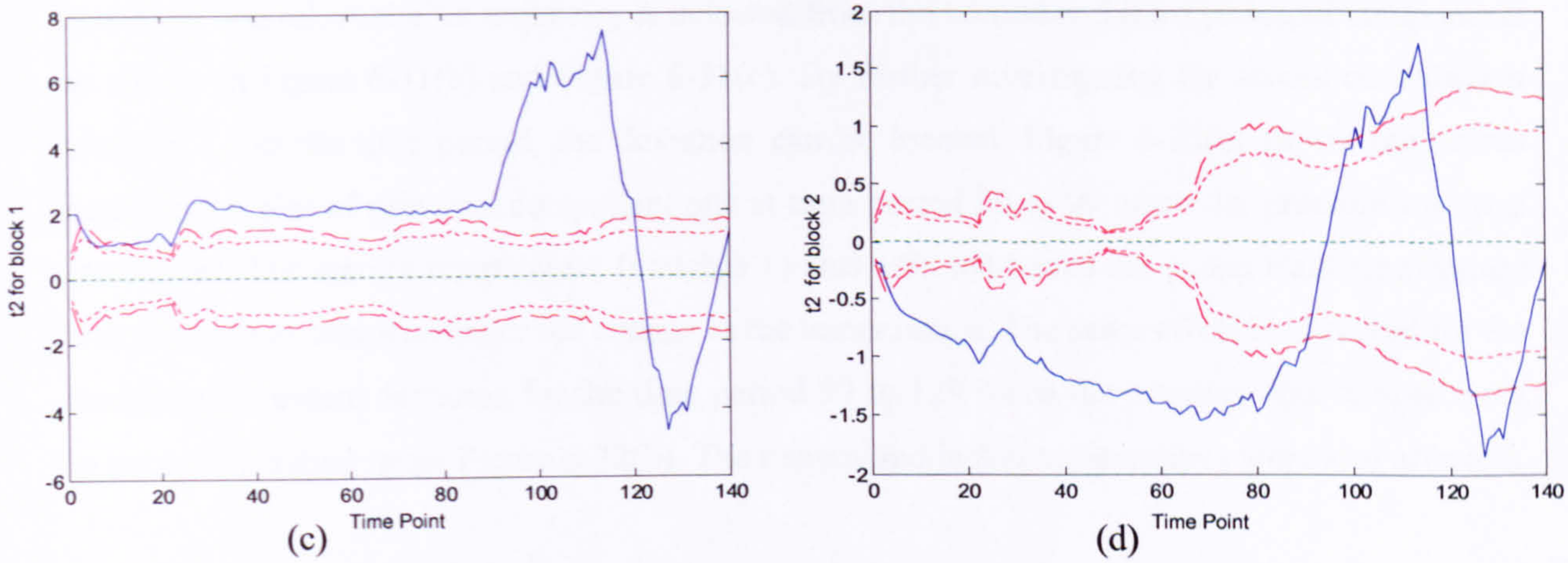


Figure 6-29 Block scores of principal components one and two for the integrated multi-block model

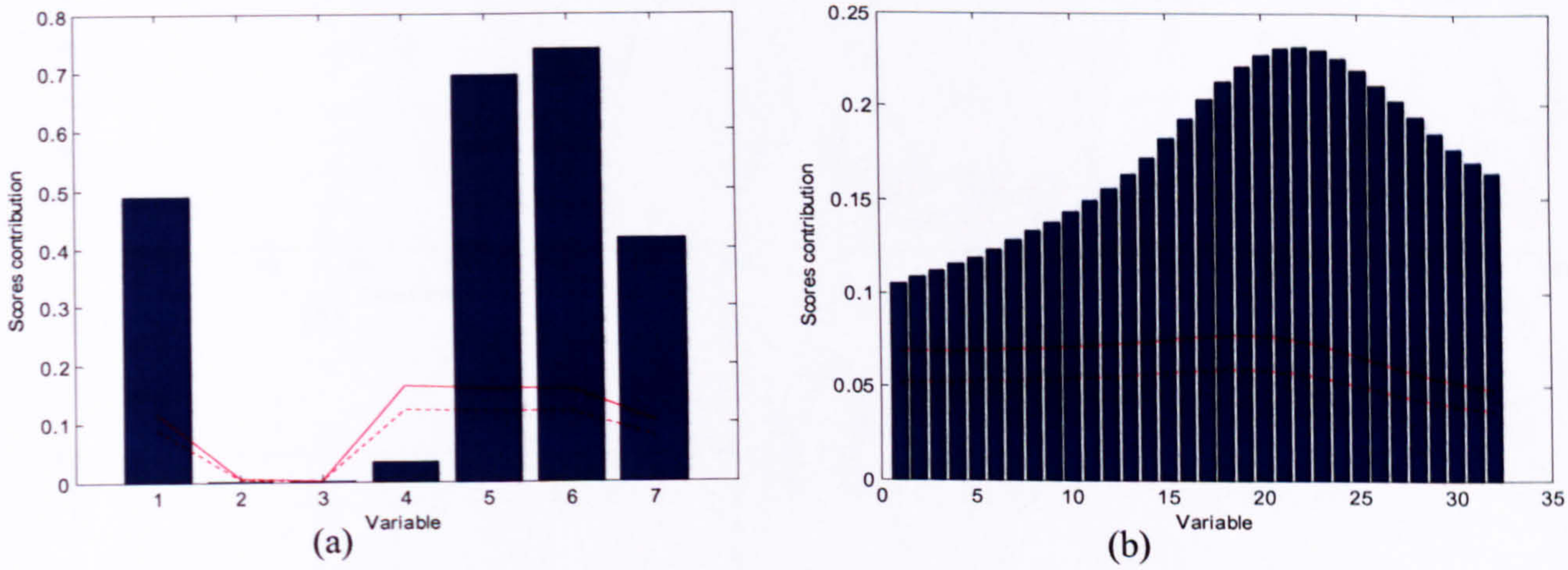


Figure 6-30 Block scores contribution plot: (a) process block at time period 90 to 117; (b) spectral block at time period 1 to 10

6.7 Analysis of Batch 12

Batch 12 is the second abnormal batch that consists of a series of pressure loss and agitator control problems. This batch is projected onto the individual and integrated models to evaluate the monitoring and fault detection capability of the representation. The individual model monitoring charts again form the basis of the benchmark analysis.

The principal component scores plots of the individual process model are shown in Figure 6-31. By interrogating the scores of principal component one, the trajectory is observed to move out of statistical control. A similar trajectory is detected from the second and third principal components as shown in Figure 6-31(b) and Figure 6-31(c). By further investigating the scores contribution plot for a specific time period, the deviation can be located. Figure 6-32(a) shows the scores contribution plot of principal component one at time period 80 to 99 when the pressure loss was introduced. The reactor temperature (variable 1) was affected hence the jacket fluid temperature was adjusted to compensate for the change in the temperature. The same effect is observed for the continuous pressure deviation for the time period 99 to 129 when the pressure was ramped back to normal operating range Figure 6-32(b). The reactor and jacket temperatures were also affected.

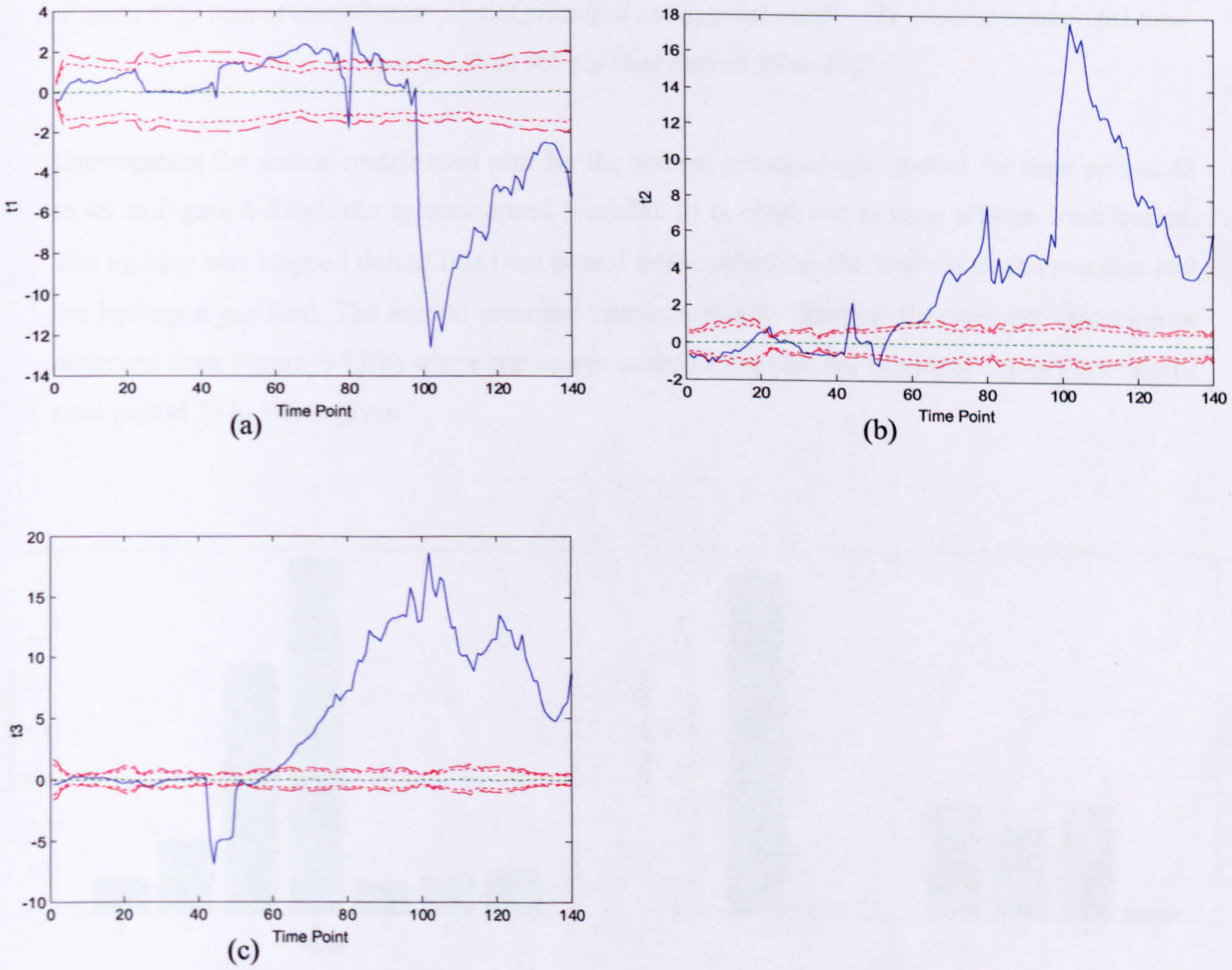


Figure 6-31 Scores of principal components one to three for the individual process model

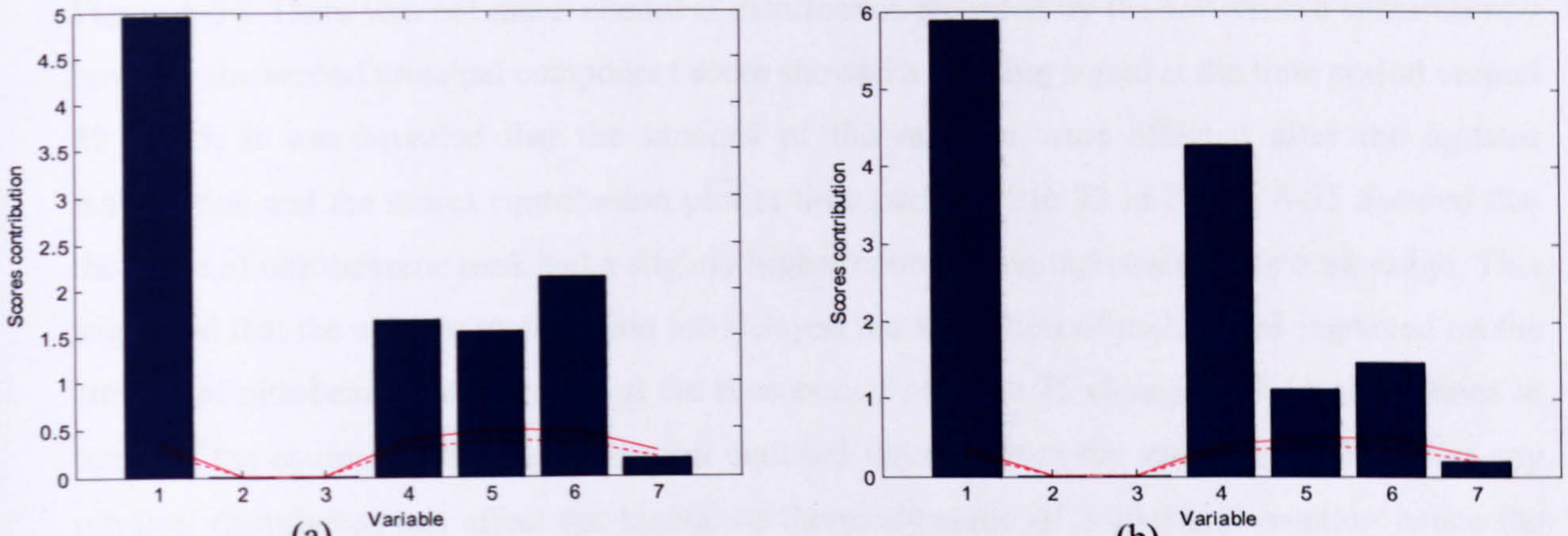


Figure 6-32 Scores contribution plot of principal component one for the process model: (a) time period 80 to 99; (b) time period 99 to 129

Interrogating the scores contribution plot for the second principal component for time period 43 to 49 in Figure 6-33(a), the agitator speed (variable 3) is observed to have a large contribution. The agitator was stopped during this time period hence affecting the kinetics of the reaction and the hydrogen gas feed. The second principal component also detected the pressure deviation as observed from Figure 6-33(b) where the scores contribution plot for principal component two of time period 70 to 140 is given.

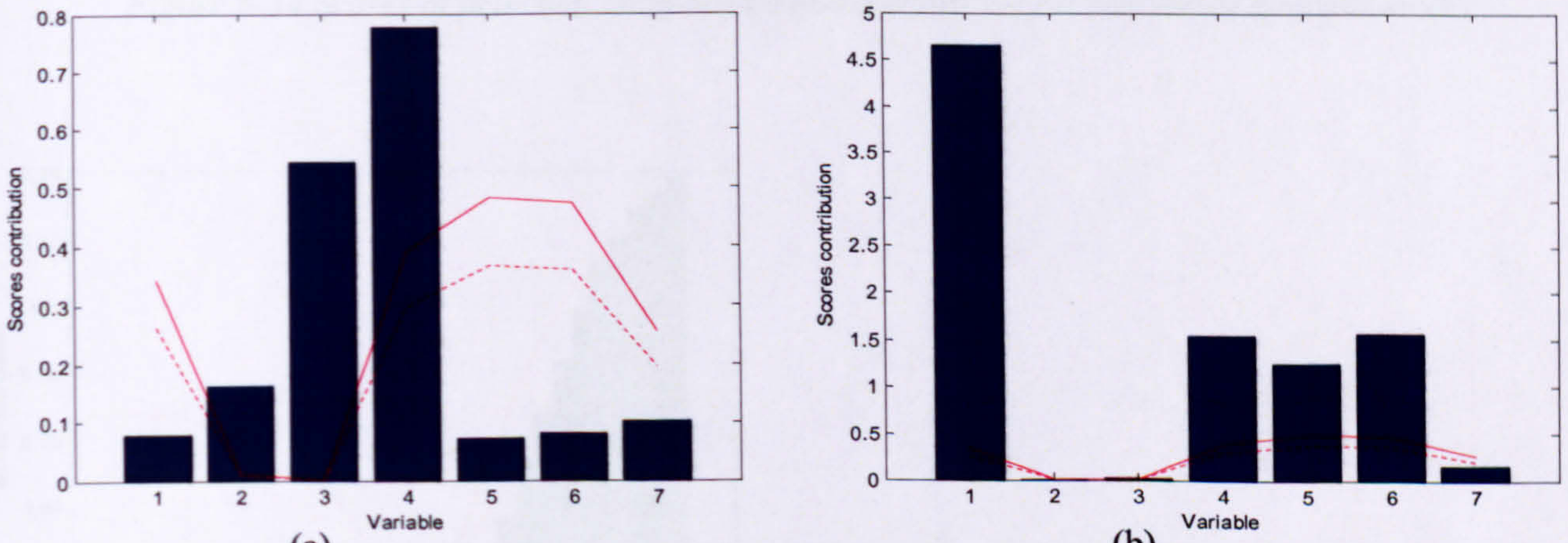


Figure 6-33 Scores contribution plot of principal component two for the process model: (a) time period 43 to 49; (b) time period 70 to 140

The principal component scores plots of the individual spectral model for batch 12 are shown in Figure 6-34. There was not much chemical disturbance recorded by the UV-visible spectroscopy however the second principal component score showed a warning signal at the time period around 50 to 75. It was revealed that the kinetics of the reaction were affected after the agitator malfunction and the scores contribution plot at time period 50 to 75 in Figure 6-35 showed that the range of nitrobenzene peak had a slightly higher contribution than the aniline peak range. This suggested that the agitator malfunction has delayed the formation of aniline and impacted on the amount of nitrobenzene and aniline at the time period of 50 to 75 changed. The malfunctions in terms of the equipment or sensors are not detected directly from the spectral data however any physical disturbance can affect the kinetic or thermodynamic of a chemical reaction hence the inherent effect is revealed.

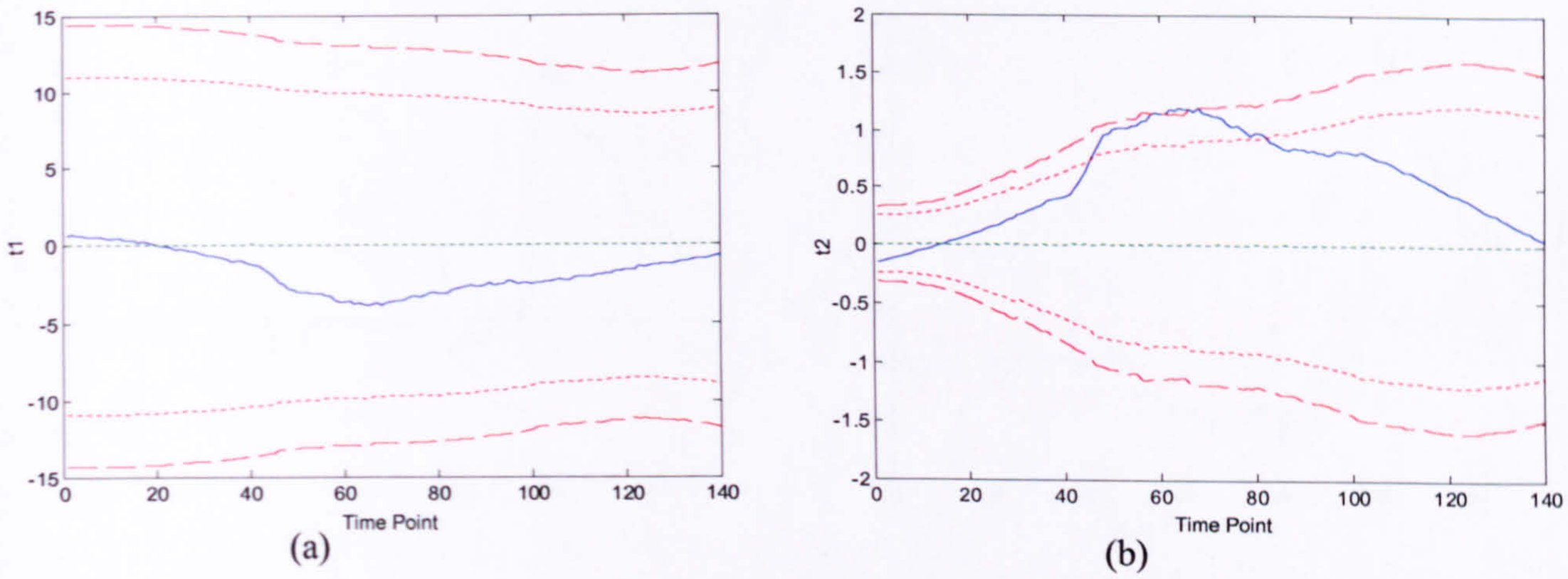


Figure 6-34 Scores of principal components one and two for the individual spectral model

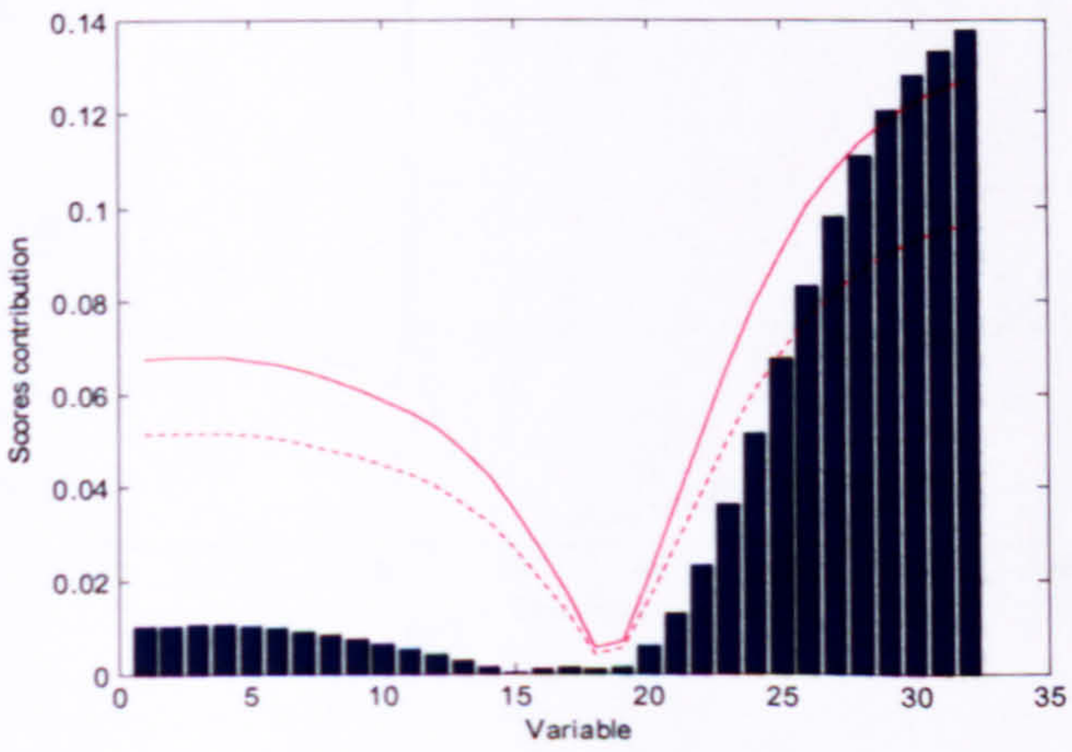


Figure 6-35 Scores contribution plot of principal component two for the spectral model at time period 50 to 75

For the integrated multi-block PCA model, the scores of principal component one to four at the super level are shown in Figure 6-36. The super scores summarise clearly the deviations from the process and spectral blocks in one single representation. The agitator disturbance introduced at time period 43 to 49 is detected by principal component four as an out-of-control signal whilst principal component one has revealed a shift in signal direction but it is still within statistical control. However, the consequence of agitator failure was clearly observed from its second to fourth principal components. The series of pressure loss problems between time period 80 to 129 are detected at different levels from different principal components. The first pressure drop starting at time period 80 was detected as significant for the second principal component. Nevertheless, the second pressure deviation starting at time point 99 is detected in most principal components.

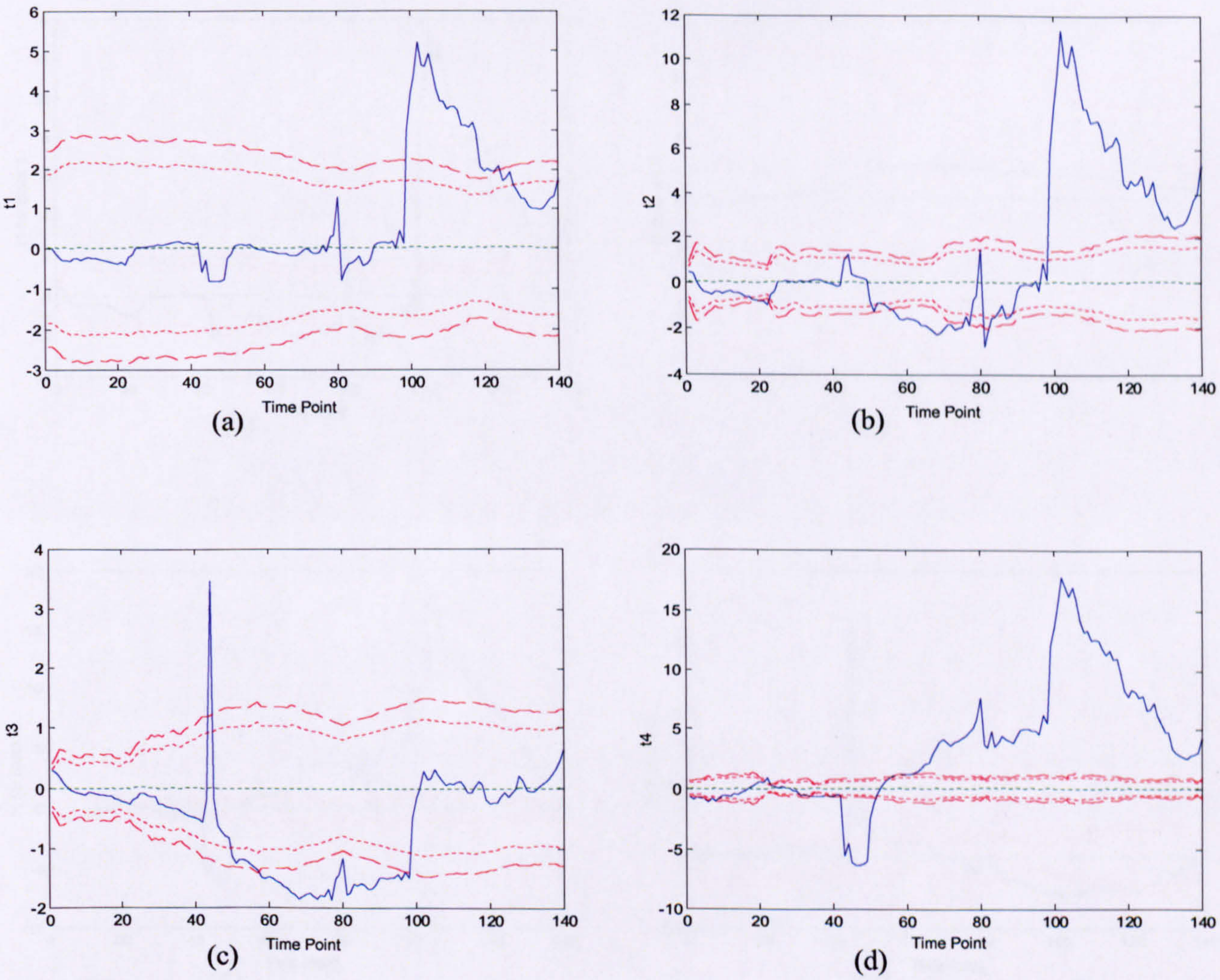


Figure 6-36 Super scores of principal components one to four for the integrated multi-block model

Since the super scores of principal component one and four have detected the agitator deviation and because the period of disturbance was short, the individual process model did not detect this deviation. Therefore the base blocks from the integrated multi-block PCA model of principal component one and four are investigated in terms of the fault detection capability as shown in Figure 6-37. Blocks 1 and 2 represent the process and spectral blocks respectively. The process block of principal component one has revealed a clear out-of-control signal at time period 43-49 and it is confirmed from the scores contribution plot shown in Figure 6-38(a) that agitator speed (variable 3) was identified as the main source of deviation. The spectral block of principal component one has revealed no atypical behaviour. The process scores contribution plot of principal component four shown in Figure 6-38(b) has also confirmed the atypical agitator behaviour as signalled in its block scores plot (Figure 6-37(c)).

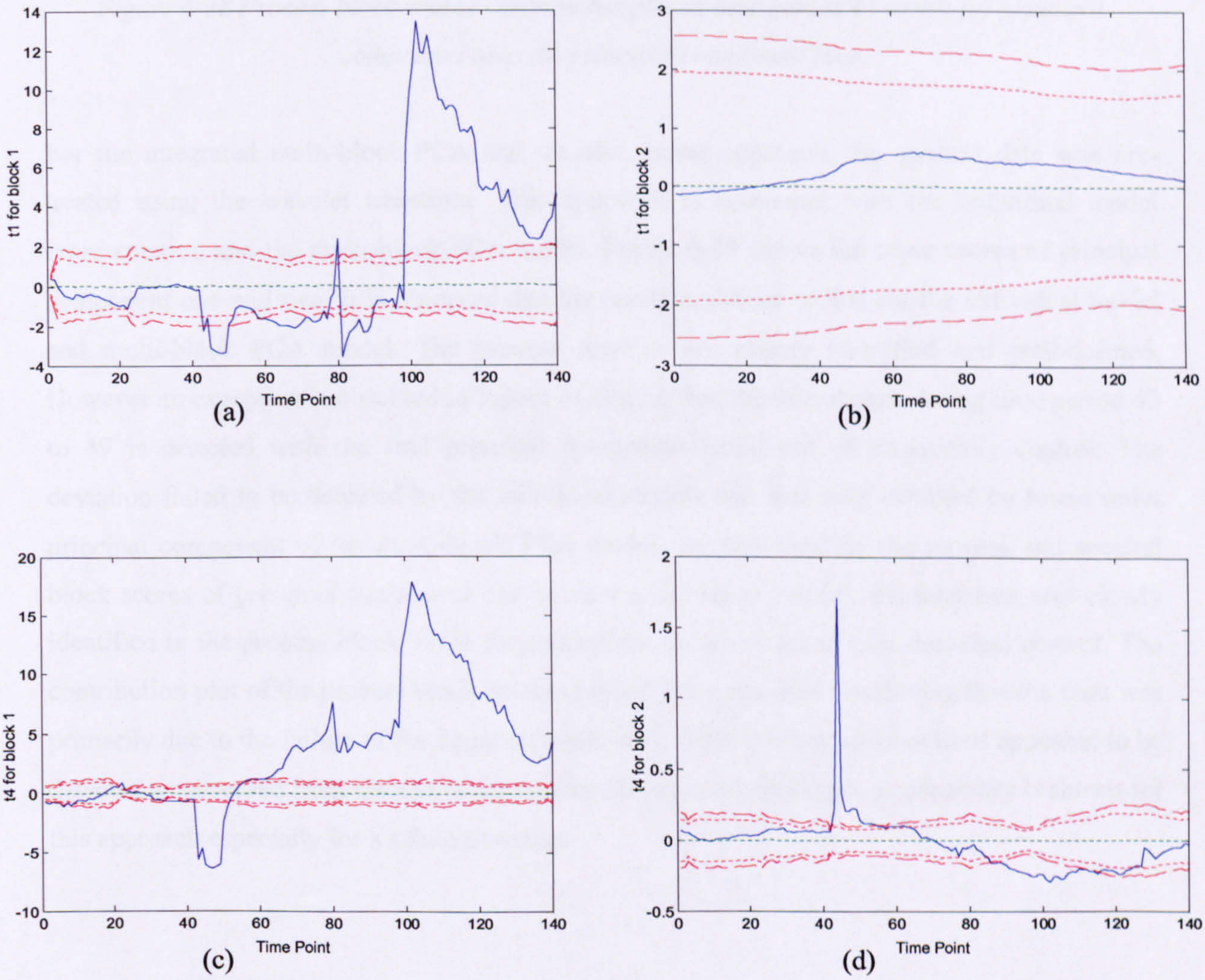


Figure 6-37 Block scores of principal components one and two for the integrated multi-block model

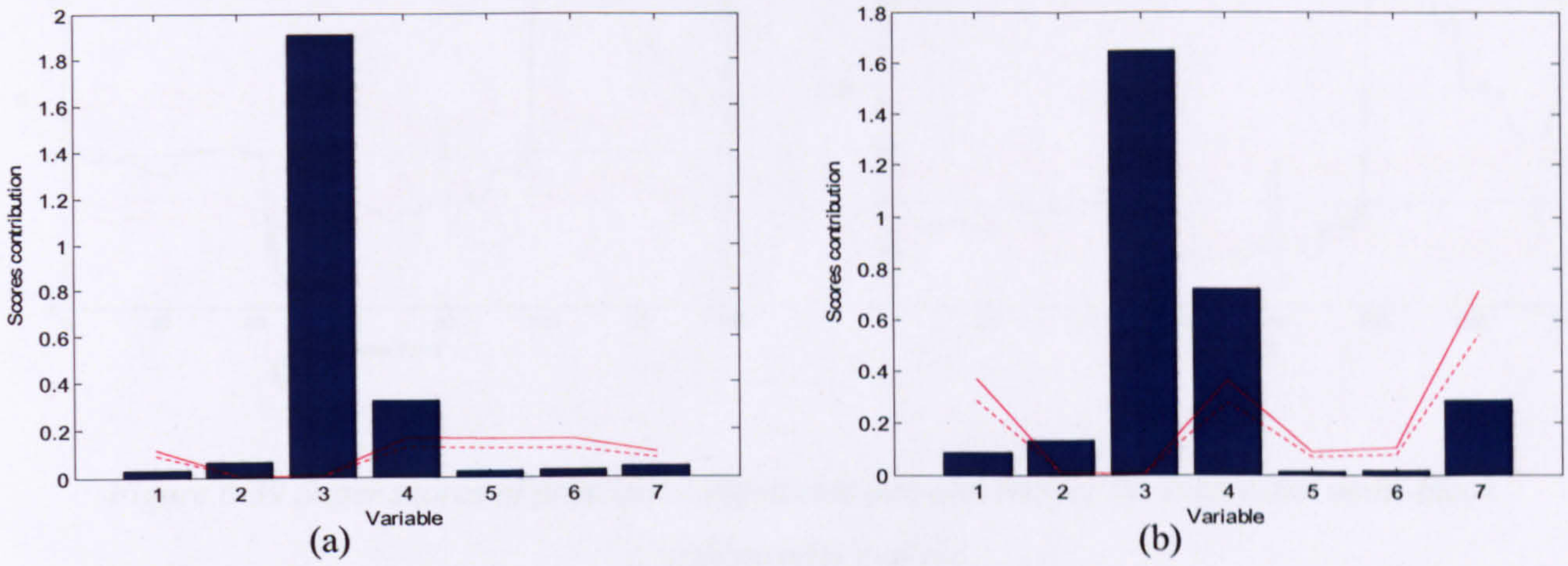


Figure 6-38 Process block scores contribution plot at time period 43 to 49: (a) principal component one; (b) principal component four

For the integrated multi-block PCA and wavelet model approach, the spectral data was pre-treated using the wavelet transform. This approach is compared with the individual model representation and the multi-block PCA model. Figure 6-39 shows the super scores of principal component one and two. It is observed that the result is similar to that for the individual model and multi-block PCA model. The process features are clearly identified and well-defined. However an exception is observed in Figure 6-39(a) in that the disturbance during time period 43 to 49 is detected with the first principal component being out of statistically control. The deviation failed to be detected by the individual models and was only detected by lower order principal component of the multi-block PCA model. By interrogating the process and spectral block scores of principal component one as shown in Figure 6-40(a), the deviation was clearly identified in the process block while the spectral block scores are within statistical control. The contribution plot of the process block scores (Figure 6-41) confirms the finding that the fault was primarily due to the failure of the agitator (variable 3). Other process malfunctions appeared to be detected as observed from the scores trajectories. Improved fault detection capability is shown for this approach especially for a subtle deviation.

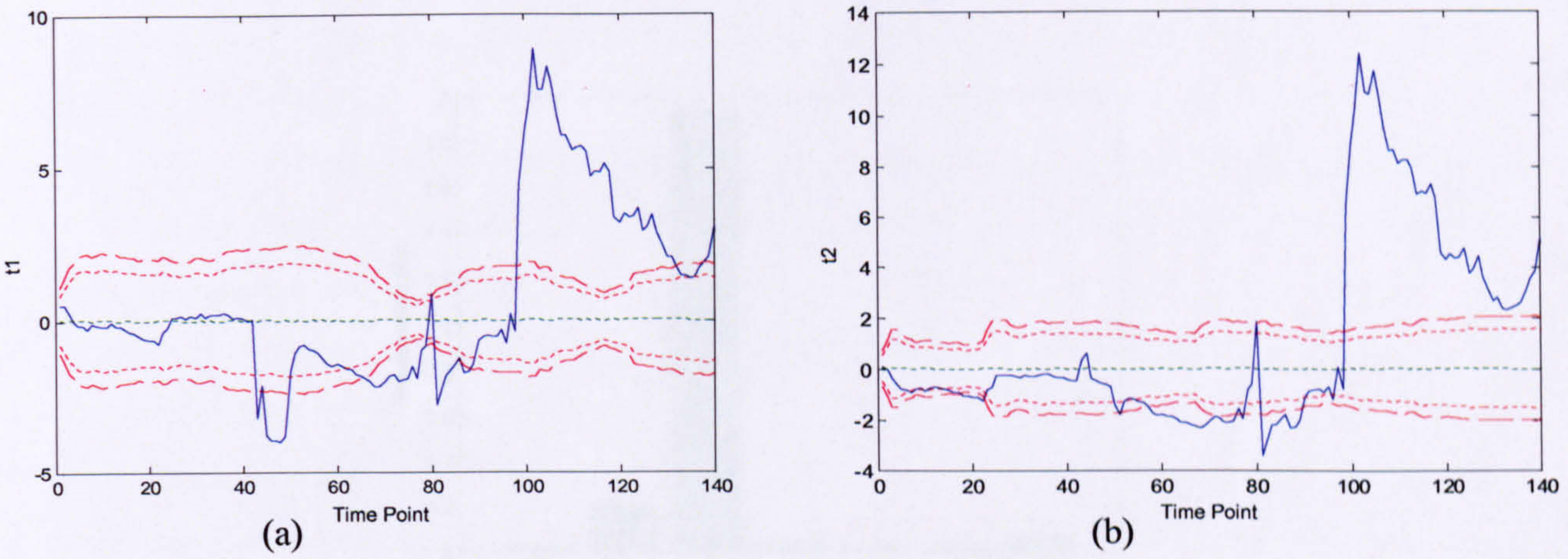


Figure 6-39 Super scores of principal components one and two for the integrated multi-block with wavelet model

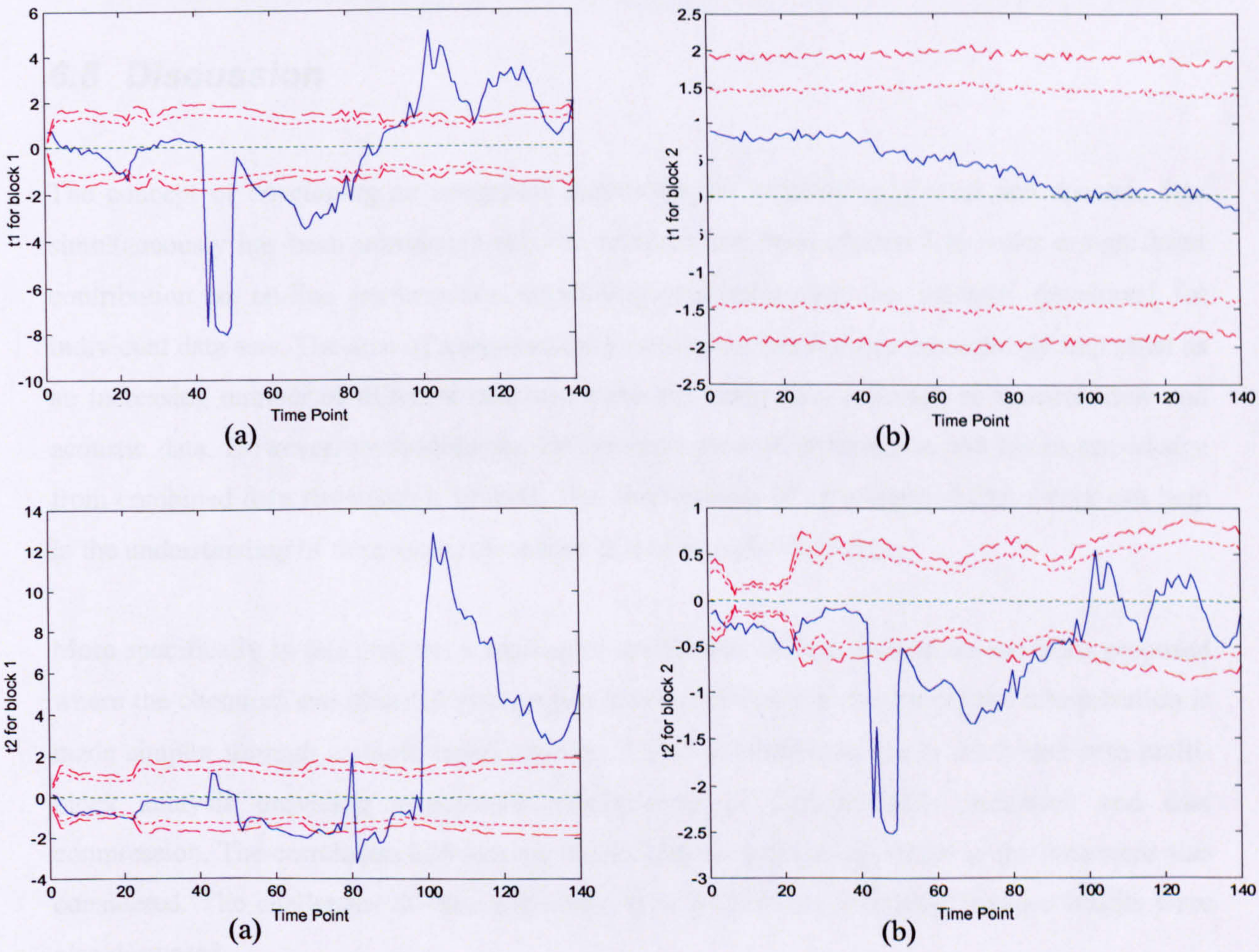


Figure 6-40 Base block scores of principal components one and two for the integrated multi-block with wavelet model

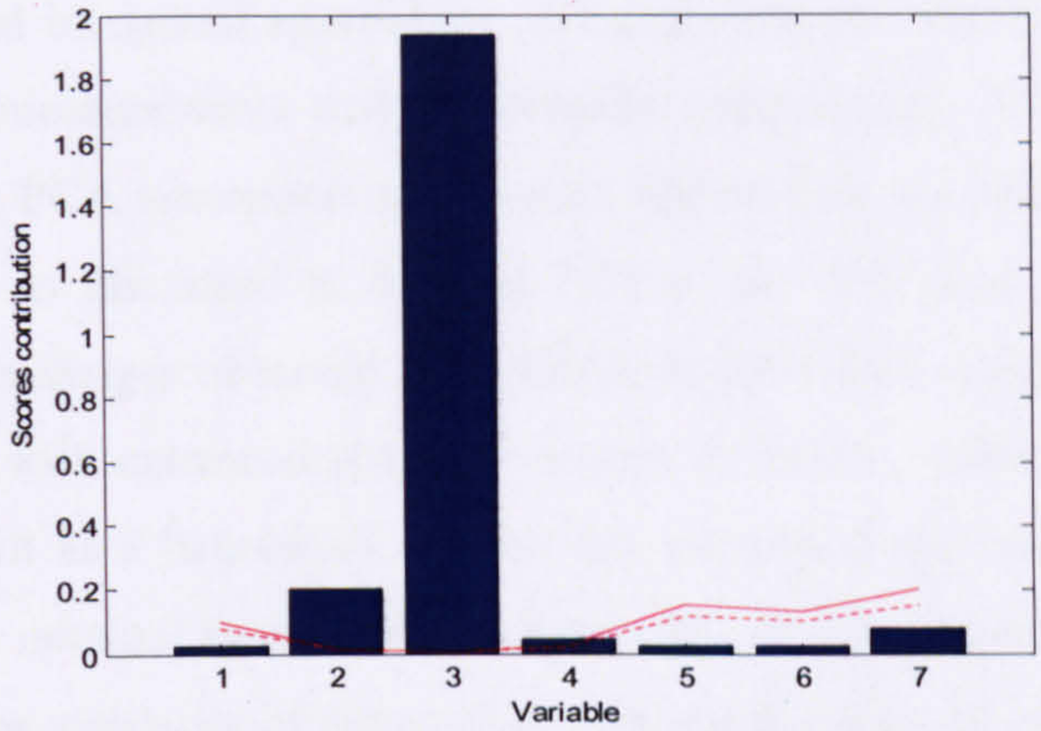


Figure 6-41 Process block scores contribution plot of principal component one at time period 43 to 49

6.8 Discussion

The concept of developing an integrated framework for monitoring process and spectral data simultaneously has been introduced and the research has been observed to make a significant contribution to on-line performance monitoring, outperforming the schemes developed for individual data sets. The area of integrated data monitoring has become increasingly important as an increasing number of different data structures are being recorded such as spectroscopic and acoustic data. However, methodologies for the extraction of information and hence knowledge from combined data structures is limited. The development of an integrated framework can help in the understanding of the process more than that of an individual entity.

More specifically in this chapter, a successful application of data integration has been proposed where the chemical and physical information is incorporated into the model but interpretation is made simpler through a single representation. Wavelet transformations is combined with multi-block analysis providing a powerful combination of dimensionality reduction and data compression. The correlation between the blocks and the multi-scale nature of the data were also considered. The challenges of time alignment, data scaling and weighing between blocks were also discussed.

A set of experimental design runs was performed on the reaction of nitrobenzene hydrogenation to aniline. Six centre point batches with same level of operating conditions were conducted hence a nominal model was constructed. Two further batches with pre-defined physical and chemical

deviations were run as the test sets to evaluate the performance monitoring and fault detection ability of the proposed integrated approaches. All experimental runs were monitored by on-line engineering process measurements and UV-Visible spectroscopy. The integrated multi-block PCA and multi-block PCA incorporating wavelets approaches are combined with the proposed monitoring approach as discussed in Section 3.3 for the first time. The results shown have confirmed that the advantages observed from different individual technique are combined into a single representation with enhanced ability to detect deviation. Individual process and spectral PCA model were built as a benchmark against the integrated approaches. From the integrated PCA approaches, the nominal models were comparable to those developed from the individual entity. This reflects the similarity of approaches forming the nominal model spaces. The analysis of batch 11 has revealed the importance of applying the integrated approach over the individual models. The deviation of catalyst charging cannot be observed from the engineering process measurements because the amount of catalyst affects intrinsically to the kinetics of reaction and it is only detectable by spectroscopic technique such as UV-Visible. The integrated approach has allowed the different sources of deviation to be identified in a single representation. The analysis of batch 12 using integrated multi-block PCA with wavelets has revealed a superior performance than the afore mentioned approach. A process deviation that was previously detected to be within the statistical control has now been revealed to be out of statistical control and was confirmed with the application of appropriate scores contribution plots. The ability of wavelets to separate multi-scale signal has significantly improved the detection of the shift of correlation structure for the new batch data against the nominal model.

For future work, the methodology will be extended to data collected from the monitoring of an industrial fermentation process where NIR and MIR spectral data are available together with the traditional process measurements. The data set will drive the integrated technique to its maximum performance as the knowledge contained in the combined NIR and MIR data are complex to be extracted (Triadaphillou, 2005).

Chapter 7

Conclusions and Future Work

7.1 Summary of Thesis

Utilisation of multivariate statistical projection techniques for the integration of various types and forms of data to realise the performance of monitoring of batch processes has been the focus in the thesis. Data overload is currently a real issue and existing techniques may not be appropriate for the task being considered such as the monitoring of the process at two sites and the simultaneous monitoring of process and spectral data. Reviewing the features of existing techniques, whilst at the same time developing novel and improved techniques is an important milestone to achieving enhanced process understanding, process modelling, process control and hence process capability. Principal Component Analysis (PCA) and Partial Least Squares (PLS) are the fundamental building blocks and have been applied to a number of different applications and industries. Chapter 2 provided an overview of the methodologies of multivariate statistical process control for batch processes including multi-way analysis and multi-group techniques. Multiway PCA and multiway PLS are extensions of PCA and PLS that handle three-way batch process data. Traditional batch monitoring techniques are reviewed and the data pre-treatment steps are examined. Multi-group PCA enables the simultaneous monitoring of different forms of data from different sites. Its extension to batch applications was discussed.

In Chapter 3, an evaluation of the performance of the existing and proposed monitoring methods was undertaken on a penicillin simulation study. The proposed approach incorporated the advantages of the existing methods in a single monitoring scheme in particular the correlation structure in the data and the use of a scaling approach that removes the non-linear and dynamic components. A further advantage is that the alignment of batch lengths is no longer required however the loadings are still fixed over the duration of a batch due to the summation of both the batch and time trajectories for a variable. This is a limitation of the proposed approach. The method was evaluated using two performance indices, the false alarm rate and the out-of-control average run length.

Chapter 4 highlighted an important application that is the monitoring of two manufacturing sites producing the same pharmaceutical drug intermediate. Different approaches were investigated for the detection of subtle differences in process performance between the two sites with respect to the product quality. Individual site models were developed as benchmarks prior to the construction of two models where the data was combined and scaled differently. The full potential of the combined models could not be realised since only three common variables existed between the two sites however the results have provided an indication that the main source of difference was the use of a different set of procedures and equipment. Nevertheless, the multi-

group PCA approach was shown to compensate for these limitations by assuming that a common eigenvector subspace existed for the variance-covariance matrix of the individual sites. The approach was shown to identify a number of batches out with the statistical control limits that were detected both in the individual and multi-group models.

Two advanced methodologies that can facilitate the emerging need for data integration were discussed in Chapter 5. The first technique was that of multi-block analysis based on PCA. Different variants of multi-block PCA were introduced and the relevant applications reported in the literature were reviewed. Multi-block analysis can provide enhanced process understanding at different levels by dividing a set of inter- or intra- related variables into meaningful blocks. The second technique utilises wavelet analysis to first decompose a signal into different scales so that both the global features and the localised details can be studied simultaneously. The application of wavelet analysis in the process industries was reviewed.

Process Analytical Technology (PAT) has resulted in the wide spread introduction of on-line process spectroscopy. Different forms of on-line spectroscopy were discussed including Near-Infrared, Mid-Infrared, UV-Visible and Raman. The pre-processing techniques of standard normal variate, multiplicative signal correction, orthogonal signal correction, derivatives and baseline correction were discussed. Finally a literature review on data integration revealed that there has only been a limited applications relating to the integration of different types of data.

The potential of data integration was demonstrated in Chapter 6 utilising process and UV-Visible spectral data. Both the technique of multi-block analysis and its conjunction with wavelet analysis were discussed. The individual process and spectral models were developed as a benchmark for performance comparison. An integrated multi-block PCA model was built comprising the process and spectral data as separate blocks and then a batch that contained a known process deviation and a change in the process chemistry was projected onto the model to test its monitoring and fault detection ability. The model was not only able to detect the deviations that resulted from the individual deviation, it also simplified the interpretation of the results by summarising the overall effects in a single representation. The multi-block PCA model that incorporated wavelets showed enhanced fault detection ability. A deviation that was not shown to be out-of-statistical control from the multi-block PCA model was clearly detected by the approach where wavelet analysis was first applied. The advantage of wavelets is that the signal is decomposed into different scales. The proposed monitoring approach introduced in Chapter 3 was applied throughout the entire analysis and it was clearly shown to be effective and flexible for a number of applications and resulted in improved results in batch performance monitoring compared to existing approaches.

In the pharmaceutical industry, the philosophy behind PAT is to incorporate the concept of Quality by Design (QbD) into every product and process. Quality should not be tested at the final quality control stage but at every stage of the pharmaceutical development and manufacturing processes. The techniques discussed in this thesis contribute to the establishment of a Design Space (DS) that represents the path towards good quality product. For example, the multi-group technique can be applied in the early stage of pharmaceutical development to establish the knowledge space utilising the data collected at the laboratory-scale and hence inform how pilot plant batches should be run since it was shown that it can handle scale differences. The multi-block technique can be widely applied to address many manufacturing challenges including the need to understand the impact of input materials and excipients on the final product quality. Data collected at different unit operations can be divided into natural blocks hence the tracking of variability can be broken down by unit operation.

7.2 Future Work

This thesis has illustrated a number of successful applications to address the emerging challenges in the pharmaceutical manufacturing industry. Novel methodologies were developed in conjunction with existing techniques for data integration. Since this is the first attempt to address the challenge, opportunities for further improvement and exploitation need to be carried out to enhance their performance and accuracy in an industrial environment. The theoretical work and practical implications are summarised below.

For the proposed monitoring algorithms discussed in Chapter 3, the following future work is recommended:

1. The limitation of deriving scores using a fixed loading for each variable needs to be modified to take account of the time-varying and dynamic nature of the batch data. For example, an adaptive approach of the algorithm could be investigated.
2. One of the advantages of the proposed approach is the handling of unequal batch length without requiring the estimation of future observations. However, the establishment of statistical control limits for the period from the minimum batch length, K_{min}, to the longest batch length in the nominal data set requires further study since an appropriate statistical distribution is required due to the reduced number of batches. The challenge is also to define what is the minimum number of batches for the statistical bounds to be valid.

3. The flexibility of the algorithms is that other types of scaling rather than auto-scaling can be applied to the unfolded data matrix. A batch maturity index can also be considered as the **Y** vector so that the algorithms can be extended to PLS in a similar fashion to the Wold *et al.* approach. Different variants of the proposed approach could be developed and compared in a comprehensive study for the evaluation of an optimised monitoring algorithm.

For the integrated multi-block PCA models discussed in Chapter 6, the recommendations for future work are as follows:

1. The integrated framework was only investigated for the proposed monitoring approaches. However the structure of the multi-block approaches allow for the incorporation of other algorithms. Therefore a comprehensive study could be conducted with alternative algorithms utilising simulated data and performance could be systematically assessed.
2. The methodology can be applied to other types of data and with more than two blocks and two levels. An example would be to combine the engineering process, MIR and NIR data for the monitoring of a fermentation process. A challenge is to assess whether the methodology can extract the information from different types of data that cannot be realised through the analysis of the different data forms in isolation (Triadaphillou, 2005).
3. If final quality data is available, the afore mentioned techniques can be extended to PLS for building a predictive model. This would constitute a major activity for the future.
4. Different types of wavelets and levels of decomposition can be further evaluated. Due to a large number of combinations being possible, experimental design can be applied for such an evaluation for the selection of the optimal wavelets and its decomposition level. This approach should further improve the understanding of the sensitivity on the modelling framework.

In the research into the on-line monitoring performance evaluation, other reported techniques can be included in the study such as the adaptive PCA (Rannar *et al.*, 1998), dynamic PCA (Chen and Liu, 2002) and multi-scale PCA (Bakshi, 1999) since the simulation data was dynamic, non-linear and multi-stage.

The multi-site application has further confirmed that the multi-group technique has significant practical implications in the area where data analysis was previously not viable due to difference in scale, equipment, measurement sensor and geographical location. Future work should include the development of a multi-group multiway PLS approach to capture the product quality to understand how the process variation has contributed to the difference in quality between sites. A

comparative study could be undertaken on data process scale-up, that is from laboratory to pilot plant and from pilot plant to full-scale production. Any deviations of quality identified can then be traced to the different modes of process operation.

The success of data integration should be measured by the level of improved process understanding and the extraction of underlying correlated behaviour that cannot be observed from individual analysis. In the case where external information such as knowledge of the product and batch genealogy are available, this should be incorporated into the empirical models for maximum knowledge mining (Ramaker *et al.*, 2002). Hybrid or semi-mechanistic models can be developed by fitting a first principal model to the data then the remaining systematic variation can be explained by the empirical approach. The major advantage of this approach is to realise the true scientific and engineering understanding of the process by knowing the mass and energy balances of the process with any further systematic variation being explained by the empirical model. This fits into the ultimate goal of PAT, that is to realise patient safety by understanding and controlling the quality of product into every stage of pharmaceutical development.

Reference

Albert, S. and R. D. Kinley (2001), Multivariate statistical monitoring of batch processes: an industrial case study of fermentation supervision, *TRENDS in Biotechnology*, 19, 53-62.

Alsberg, B. K., A. M. Woodward and D. B. Kell (1997), An introduction to wavelet transforms for chemometricians: A time-frequency approach, *Chemometrics and intelligent laboratory systems*, 37, 215-239.

Araujo, P. W. and R. G. Brereton (1996a), Experimental design I. Screening, *TrAC Trends in Analytical Chemistry*, 15, 26-31.

Araujo, P. W. and R. G. Brereton (1996b), Experimental design II. Optimization, *TrAC Trends in Analytical Chemistry*, 15, 63-70.

Araujo, P. W. and R. G. Brereton (1996c), Experimental design III. Quantification, *TrAC Trends in Analytical Chemistry*, 15, 156-163.

Atkinson, B. and F. Mavituna (1991), *Biochemical engineering and biotechnology handbook*, Stockton Press, New York.

Azzouz, T., A. Puigdomenech, M. Aragay and R. Tauler (2003), Comparison between different data pre-treatment methods in the analysis of forage samples using near-infrared diffuse reflectance spectroscopy and partial least-squares multivariate calibration method, *Analytica Chimica Acta*, 484, 121-134.

Bakeev, K. (2005), Process Analytical Technology: Spectroscopic tools and implementation strategies for the chemical and pharmaceutical industries, Blackwell publishing.

Bakshi, B. R. (1998), Multi-scale PCA with application to multivariate statistical process monitoring, *AIChE Journal*, 44, 1596-1610.

Bakshi, B. R. (1999), Multi-scale analysis and modeling using wavelets, *Journal of Chemometrics*, 13, 415-434.

Barnes, R. J., M. S. Dhanoa and S. J. Lister (1989), Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra, *Applied Spectroscopy*, 43, 772-777.

Bentley, P. M. and J. T. E. McDonnell (1994), Wavelet transforms - an introduction, *Electronics & Communication Engineering Journal*, 6, 175-186.

Berglund, A. and S. Wold (1999), A serial extension of multiblock PLS, *Journal of Chemometrics*, 13, 461-471.

Birol, G., C. Undey and A. Cinar (2002), A modular simulation package for fed-batch fermentation: penicillin production, *Computers and Chemical Engineering*, 26, 1553-1565.

Box, G. E. P. (1954), Some theorems on quadratic forms applied in the study of analysis of variance problems: effect of inequality of variance in one-way classification, *The Annals of Mathematical Statistics*, 25, 290-302.

Bras, L. P., S. A. Bernardino, J. A. Lopes and J. C. Menezes (2005), Multiblock PLS as an approach to compare and combine NIR and MIR spectra in calibrations of soybean flour, *Chemometrics and Intelligent Laboratory Systems*, 75, 91-99.

Bro, R. (1997), PARAFAC, Tutorial and applications, *Chemometrics and Intelligent Laboratory Systems*, 38, 149-171.

Bro, R. and A. K. Smilde (2003), Centring and scaling in component analysis, *Journal of Chemometrics*, 17, 16-33.

Candolfi, A., R. De Maesschalck, D. L. Massart, P. A. Hailey and A. C. E. Harrington (1999), Identification of pharmaceutical excipients using NIR spectroscopy and SIMCA, *Journal of Pharmaceutical and Biomedical Analysis*, 19, 923-935.

Chen, G. and T. J. McAvoy (1997), Multi-block predictive monitoring of continuous processes, *IFAC ADCHEM International Symposium on Advanced Control of Chemical Processes*, Banff, Canada, 73-77.

Chen, G. and T. J. McAvoy (1998), Predictive on-line monitoring of continuous processes, *Journal of Process Control*, 8, 409-420.

Chen, J. and K. C. Liu (2002), On-line batch process monitoring using dynamic PCA and dynamic PLS models, *Chemical Engineering Science*, 57, 63-75.

Chiang, L. H., R. Leardi, R. J. Pell and M. B. Seasholtz (2006), Industrial experiences with multivariate statistical analysis of batch process data, *Chemometrics and Intelligent Laboratory Systems*, 81, 109-119.

Cho, H. W. and K. J. Kim (2003), A method for predicting future observations in the monitoring of a batch process, *Journal of Quality Technology*, 35, 59-69.

Cimander, C. and C. F. Mandenius (2002), Online monitoring of a bioprocess based on a multi-analyser system and multivariate statistical process modelling, *Journal of Chemical Technology and Biotechnology*, 77, 1157-1168.

Conlin, A. K., E. B. Martin and A. J. Morris (2000), Confidence limits for contribution plots, *Journal of Chemometrics*, 14, 725-736.

Daubechies, I. (1988), Orthonormal bases of compactly supported wavelets, *Communications on Pure and Applied Mathematics*, XLI, 909-996.

Daubechies, I. (1992), *Ten lectures on wavelets*, SIAM, Philadelphia.

Ding, H., J. H. Liu and Z. R. Shen (2003), Drift reduction of gas sensor by wavelet and principal component analysis, *Sensors and Actuators B: Chemical*, 96, 354-363.

Eriksson, L., E. Johansson, N. Kettaneh-Wold, J. Trygg, H. Wikstrom and S. Wold (2006a), *Multi- and Megavariate Data Analysis Part I: Basic principles and applications*, Umetrics Academy.

Eriksson, L., E. Johansson, N. Kettaneh-Wold, J. Trygg, H. Wikstrom and S. Wold (2006b), *Multi- and megavariate data analysis. Part II: Advanced applications and method extensions*, Umetrics Academy.

Eriksson, L., J. Trygg, E. Johansson, R. Bro and S. Wold (2000), Orthogonal signal correction, wavelet analysis, and multivariate calibration of complicated process fluorescence data,, *Analytica Chimica Acta*, 420, 181-195.

Felicio, C. C., L. P. Bras, J. A. Lopes, L. Cabrita and J. C. Menezes (2005), Comparison of PLS algorithms in gasoline and monitoring with MIR and NIR, *Chemometrics and Intelligent Laboratory Systems*, 78, 74-80.

Flores-Cerrillo, J. and J. MacGregor (2004), Multivariate monitoring of batch processes using batch-to-batch information, *AIChE Journal*, 50, 1219-1228.

Flury, B. N. (1984), Common principal components in K-groups, *Journal of the American Statistical Association*, 79, 892-898.

Flury, B. N. (1987), Two generalisations of the common principal component model, *Biometrika*, 74, 59-69.

Gallagher, N. B., B. M. Wise and C. W. Stewart (1996), Application of multi-way principal components analysis to nuclear waste storage tank monitoring, *Computers and Chemical Engineering*, 20, S739-S744.

Garcia-Munoz, S., T. Kourti and J. F. MacGregor (2003), Troubleshooting of an industrial batch process using multivariate methods, *Industrial & Engineering Chemistry Research*, 42, 3592-3601.

Geladi, P. (1989), Analysis of Multi-way (multi-mode) data, *Chemometrics and Intelligent Laboratory Systems*, 7, 11-30.

Geladi, P. and B. R. Kowalski (1986), Partial Least Squares regression: a tutorial, *Analytica Chimica Acta*, 185, 1-17.

Geladi, P., D. MacDougall and H. Martens (1985), Linearisation and scatter correction for near-infrared reflectance spectra of meat, *Applied Spetroscopy*, 39, 491-500.

Gollmer, K. and C. Posten (1996), Supervision of bioprocesses using a dynamic time warping algorithm, *Control Engineering Practice*, 4, 1287-1295.

Gorry, P. A. (1990), General least-squares smoothing and differentiation by the convolution (Savitzky-Golay) method, *Analytical Chemistry*, 62, 570-573.

Graps, A. (1995), An introduction to wavelets, *IEEE Computational Science & Engineering*, 2, 50-61.

Gurden, S. P., J. A. Westerhuis and A. K. Smilde (2002), Monitoring of batch processes using spectroscopy, *AIChE Journal*, 48, 2283-2297.

Gurden, S. P., J. A. Westerhuis, R. Bro and A. K. Smilde (2001), A comparison of multiway regression and scaling methods, *Chemometrics and Intelligent Laboratory Systems*, 59, 121-136.

Gurden, S. P., J. Westerhuis, S. Bijlsma and A. K. Smilde (2001), Modelling of spectroscopic batch process data using grey models to incorporate external information, *Journal of Chemometrics*, 15, 101-121.

Hotelling, H. (1933), Analysis of a complex of statistical variables into principal components, *Educational Psychology*, 24, 417-441.

Hotelling, H. (1947), *Techniques of statistical analysis*, McGraw Hill, New York.

Hubbard, B. B. (1995), *The world according to wavelets*, Wellesley, Massachusetts.

Jackson, J. E. and G. S. Mudholkar (1979), Control procedures for residuals associated with Principal Component Analysis, *Technometrics*, 21, 341-349.

Janné, K., J. Pettersen, N. O. Lindberg and *T. Lundstedt (2001),* Hierarchical principal component analysis (PCA) and projection to latent structure (PLS) technique on spectroscopic data as a data pretreatment for calibration, *Journal of Chemometrics*, 15, 203-213.

Jolliffe, I. T. (1986), *Principal component analysis*, Springer-Verlag, New York.

Kassidas, A., J. MacGregor and P. A. Taylor (1998), Synchronisation of batch trajectories using dynamic time warping, *AIChE Journal*, 44, 864-875.

Kosanovich, K. A. and M. J. Piovoso (1997), PCA of Wavelet Transformed Process Data for Monitoring, *Intelligent Data Analysis*, 1, 85-99.

Kosanovich, K. A., K. S. Dahl and M. J. Piovoso (1996), Improved process understanding using multiway Principal Component Analysis, *Industrial & Engineering Chemistry Research*, 35, 138-146.

Kourti, T. (2003), Multivariate dynamic data modelling for analysis and statistical process control of batch processes, start-ups and grade transitions, *Journal of Chemometrics*, 17, 93-109.

Kourti, T. (2005), Application of latent variable methods to process control and multivariate statistical process control in industry, *International Journal of Adaptive Control and Signal Processing*, 19, 213-246.

Kourti, T. and J. F. MacGregor (1995), Process analysis, monitoring and diagonsis, using multivariate projection methods, *Chemometrics and Intelligent Laboratory Systems*, 28, 3-21.

Kourti, T. and J. F. MacGregor (1996), Multivariate SPC methods for process and product monitoring, *Journal of Quality Technology*, 28, 409-428.

Kresta, J., J. F. MacGregor and T. E. Marlin (1991), Multivariate statistical monitoring of process operating performance, *Canadian Journal of Chemical Engineering*, 69, 35-47.

Krzanowski, W. J. (1979), Between-group comparison of principal components, *Journal of the American Statistical Association*, 74, 703-707.

Krzanowski, W. J. (1984), Principal component analysis in the presence of group structure, *Applied Statistics*, 33, 164-168.

Lane, S. (1999), The extension of multivariate statistical process performance monitoring techniques to multiple group applications, PhD, *Department of Chemical and Process Engineering*, University of Newcastle-upon-Tyne, Newcastle-upon-Tyne.

Lee, D. S. and P. A. Vanrolleghem (2003), Monitoring of a sequencing batch reactor using adaptive multiblock principal component analysis, *Biotechnology and Bioengineering*, 82, 489-497.

Lee, J. M., C. Yoo and I. B. Lee (2003), On-line batch process monitoring using a consecutively updated multiway principal component analysis model, *Computers and Chemical Engineering*, 27, 1903-1912.

Lopes, J. A., J. C. Menezes, J. A. Westerhuis and A. K. Smilde (2002), Multiblock PLS analysis of an industrial pharmaceutical process, *Biotechnology and Bioengineering*, 80, 419-427.

Lorber, A., K. Faber and B. R. Kowalsky (1997), Analytical figures of merit for tensorial calibration, *Journal of Chemometrics*, 11, 419-461.

Lu, N. Y., F. L. Wang and F. R. Gao (2003), Combination method of principal component and wavelet analysis for multivariate process monitoring and fault diagnosis, *Industrial & Engineering Chemistry Research*, 42, 4198-4207.

MacGregor, J. F., C. Jaeckle, C. Kiparissides and M. Koutoudi (1994), Process monitoring and diagnosis by multiblock PLS methods, *AIChE Journal*, 40, 826-838.

Mallat, S. G. (1989a), Multiresolution approximations and wavelet orthonormal bases, *Transactions of the American Mathematical Society*, 315, 69-87.

Mallat, S. G. (1989b), A theory for multiresolution signal decomposition: the wavelet representation, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 11, 674-693.

Mallat, S. G. (1998), *A wavelet tour of signal processing*, Academic Press, San Diego.

Mardia, K. V., J. T. Kent and J. M. Bibby (1979), *Multivariate analysis*, Academic Press, London.

Mark, H. and J. Workman (2003), Derivatives in spectroscopy - Part III - Computing the derivative, *Spectroscopy*, 18, 106-111.

Martens, H. and T. Næs (1989), *Multivariate calibration*, John Wiley & Sons, New York.

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

Martin, E. B. and A. J. Morris (2002), Enhanced bio-manufacturing through advanced multivariate statistical technologies, *Journal of Biotechnology*, 99, 223-235.

Martin, E. B., A. Bettoni and A. J. Morris (2002), Monitoring of a batch continuous process using mass re-sampling, *Journal of Quality Technology*, 34, 171-186.

Martin, E. B., A. J. Morris and C. Kiparissides (1999), Manufacturing performance enhancement through multivariate statistical process control, *Annual Reviews in Control*, 23, 35-44.

McLennan, F. and B. R. Kowalski (1995), *Process analytical chemistry*, Kluwer Academic Publishers.

Meyer, Y. (1992), *Wavelets and operators*, University press, Cambridge.

Meyer, Y. (1993), *Wavelets: algorithms and applications*, SIAM, Philadelphia.

Miletic, I., S. Quinn, M. Dudzic, V. Vaculik and M. Champagne (2004), An industrial perspective on implementing on-line applications of multivariate statistics, *Journal of Process Control*, 14, 821-836.

Miller, P., R. E. Swanson and C. E. Heckler (1998), Contribution plots: A missing link in multivariate quality control, *International Journal of Applied Mathematics Computer Science*, 8, 775-792.

Misra, M., H. Yue, J. Qin and C. Ling (2002), Multivariate process monitoring and fault diagnosis by multi-scale PCA, *Computers and Chemical Engineering*, 26, 1281-1293.

Montgomery, D. C. (1996), *Introduction to statistical quality control*, John Wiley, New York.

Montgomery, D. C. (2005), *Design and analysis of experiments*, John Wiley and Sons.

Motard, R. L. and B. Joseph (1994), *Wavelet applications in chemical engineering*, Kluwer Academic Publishers, Boston.

Myers, C., L. R. Rabiner and A. E. Rosenberg (1980), Performance tradeoffs in dynamic time warping algorithms for isolated word recognitions, *IEEE Trans. on Acoustics, Speech and Signal Process*, 6, 623-635.

Neogi, D. and C. Schlags (1998), Multivariate statistical analysis of an emulsion batch process, *Industrial & Engineering Chemistry Research*, 37, 3971-3979.

Nomikos, P. (1996), Detection and diagnosis of abnormal batch operations based on multi-way principal component analysis, World Batch Forum, Toronto, May 1996, *ISA Transactions*, 35, 259-266.

Nomikos, P. and J. F. MacGregor (1995a), Multi-way partial least squares in monitoring batch processes, *Chemometrics and Intelligent Laboratory Systems*, 30, 97-108.

Nomikos, P. and J. F. MacGregor (1995b), Multivariate SPC charts for monitoring batch processes, *Technometrics*, 37, 41-59.

Nomikos, P. and J. F. MacGregor (1994), Monitoring batch processes using multiway principal component analysis, *AIChE Journal*, 40, 1361-1373.

Parris, J., C. Airiau, R. Escott, J. Rydzak and R. Crocombe (2005), Monitoring API drying operations with NIR, *Spectroscopy*, 20, 34-42.

Pearson, K. (1901), On lines and planes of closest fit to systems of points in space, *Philosophical Magazine*, 2, 559-572.

Qin, S. J., S. Valle and M. J. Piovoso (2001), On unifying multiblock analysis with application to decentralized process monitoring, *Journal of Chemometrics*, 15, 715-742.

Ramaker, H. J., E. N. M. van Sprang, J. A. Westerhuis and A. K. Smilde (2003), Dynamic time warping of spectroscopic BATCH data, *Analytica Chimica Acta*, 498, 133-153.

Ramaker, H. J., E. N. M. van Sprang, S. P. Gurden, J. A. Westerhuis and A. K. Smilde (2002), Improved monitoring of batch processes by incorporating external information, *Journal of Process Control*, 12, 569-576.

Rannar, S., J. F. MacGregor and S. Wold (1998), Adaptive batch monitoring using hierarchical PCA, *Chemometrics and Intelligent Laboratory Systems*, 41, 73-81.

Rosen, C. and J. A. Lennox (2001), Multivariate and multi-scale monitoring of wastewater treatment operation, *Water Research*, 35, 3402-3410.

Rothwell, S. G. (1999), Multivariate statistical process control of batch processes, PhD, *Department of Chemical and Process Engineering*, University of Newcastle-upon-Tyne, Newcastle-upon-Tyne.

Ruckebusch, C., B. Sombret, R. Froidevaux and J.-P. Huvenne (2000), On-line mid-infrared spectroscopic data and chemometrics for the monitoring of an enzymatic hydrolysis, *Applied Spetroscopy*, 55, 1610-1617.

Sakoe, H. and S. Chiba (1978), Dynamic programming algorithm optimisation for spoken word recognition, *IEEE Trans. on Acoustics, Speech and Signal Process*, 26, 43-49.

Sanchez, E. and B. R. Kowalski (1990), Tensorial resolution: a direct trilinear decomposition, *Journal of Chemometrics*, 4, 29-45.

Shao, R., F. Jia, E. B. Martin and A. J. Morris (1999), Wavelets and non-linear principal components analysis for process monitoring, *Control Engineering Practice*, 7, 865-879.

Smilde, A. K. (1992), Three-way analysis. Problems and prospects, *Chemometrics and Intelligent Laboratory Systems*, 15, 143-157.

Smilde, A. K. and D. A. Doornbos (1991), Three way methods for the calibration of chromatographic systems: comparing PARAFAC and three-way PLS, *Journal of Chemometrics*, 5, 345-360.

Smilde, A. K. and H. A. L. Kiers (1999), Multiway covariates regression models, *Journal of Chemometrics*, 13, 31-48.

Smilde, A. K., J. A. Westerhuis and S. De Jong (2003), A framework for sequential multiblock component methods, *Journal of Chemometrics*, 17, 323-337.

Tates, A. A., D. J. Louwerse, A. Smilde, G. L. M. Koot and H. Berndt (1999), Monitoring a PVC batch process with multivariate statistical process control charts, *Industrial & Engineering Chemistry Research*, 38, 4769-4776.

Teppola, P. and P. Minkkinen (2000), Wavelet-PLS regression models for both exploratory data analysis and process monitoring, *Journal of Chemometrics*, 14, 383-399.

Teppola, P. and P. Minkkinen (2001), Wavelets for scrutinizing multivariate exploratory models-interpreting models through multiresolution analysis, *Journal of Chemometrics*, 15, 1-18.

Thorpe, R. S. (1983), A review of numerical methods for recognising and analysing radical differentiation, *Numerical Taxonomy*, Berlin Heidelberg, Springer-Verlag, 404-419.

Tracy, N. D., J. C. Young and R. L. Mason (1992), Multivariate control charts for individual observations, *Journal of Quality Technology*, 24, 88-95.

Triadaphillou, S. (2005), Spectroscopic and process data fusion: enhanced monitoring of an industrial fermentation, PhD, *School of Chemical Engineering and Advanced Materials*, Newcastle University, Newcastle-upon-Tyne.

Trygg, J. and S. Wold (1998), PLS regression on wavelet compressed NIR spectra, *Chemometrics and intelligent laboratory systems*, 42, 209-220.

Trygg, J., N. Kettaneh-Wold and L. Wallbäcks (2001), 2D wavelet analysis and compression of on-line industrial process data, *Journal of Chemometrics*, 15, 299-319.

U.S. Department of Health and Human Services Foods and Drug Administration (2004), *Guideline for Industry, PAT – A framework for innovative pharmaceutical developing, manufacturing and quality assurance,* US, http://www.fda.gov/cder/guidance/6419fnl.pdf.

Undey , C. and A. Cinar (2002), Statistical monitoring of multistage, *multiphase batch processes, IEEE Control Systems Magazine*, 22, 40-52.

Van Sprang, E. N. M., H. J. Ramaker, H. F. M. Boelens, J. A. Westerhuis, D. Whiteman, D. Baines and I. Weaver (2003), Batch process monitoring using on-line MIR spectroscopy, *The Analyst*, 128, 98-102.

Wangen, L. and B. R. Kowalski (1988), A multiblock partial least squares algorithm for investigating complex chemical systems, *Journal of Chemometrics*, 3, 3-20.

Westerhuis, J. A. and P. M. J. Coenegracht (1997), Multivariate modelling of the pharmaceutical two-step process of wet granulation and tableting with multiblock partial least squares, *Journal of Chemometrics*, 11, 379-392.

Westerhuis, J. A., S. P. Gurden and A. K. Smilde (2000a), Generalized contribution plots in multivariate statistical process monitoring, *Chemometrics and Intelligent Laboratory Systems*, 51, 95-114.

Westerhuis, J. A., S. P. Gurden and A. K. Smilde (2000b), Spectroscopic monitoring of batch reactions for on-line fault detection and diagnosis, *Analytical Chemistry*, 72, 5322-5330.

Westerhuis, J. A., T. Kourti and J. F. MacGregor (1998), Analysis of multiblock and hierarchical PCA and PLS models, *Journal of Chemometrics*, 12, 301-321.

Westerhuis, J. A., T. Kourti and J. F. MacGregor (1999), Comparing alternative approaches for multivariate statistical analysis of batch process data, *Journal of Chemometrics*, 13, 397-413.

Wise, B. M., N. B. Gallagher, S. W. Butler, D. D. White and G. G. Barna (1999), A comparison of principal component analysis, multiway principal component analysis, trilinear decomposition and parallel factor analysis for fault detection in a semiconductor etch process, *Journal of Chemometrics*, 13, 379-396.

Wold, H. (1966), Non-linear estimation by iterative least squares procedures, *Research Papers in Statistics*, 411-444.

Wold, S. (1978), Cross-validatory estimation of the number of components in factor and principal components models, *Technometrics*, 20, 397-405.

Wold, S. (2004), The four levels of PAT (Process Analytical Technology) and associated risks and benefits in pharmaceutical production, *9th International Conference of Chemometrics in Analytical Chemistry*, Lisbon.

Wold, S., H. Antti, L. Lindgren and O. Jerker (1998), Orthogonal signal correction of near-infrared spectra, *Chemometrics and Intelligent Laboratory Systems*, 44, 175-185.

Wold, S., J. Cheney, N. Kettaneh and C. McCready (2006), The chemometric analysis of point and dynamic data in pharmaceutical and biotech production (PAT) -- some objectives and approaches, *Chemometrics and Intelligent Laboratory Systems*, 84, 159-163.

Wold, S., J. Trygg, A. Berglund and H. Antti (2001), Some recent developments in PLS modeling, *Chemometrics and Intelligent Laboratory Systems*, 58, 131-150.

Wold, S., N. Kettaneh and K. Tjessem (1996), Hierarchical multiblock PLS and PC models for easier model interpretation and as an alternative to variable selection, *Journal of Chemometrics*, 10, 463-482.

Wold, S., N. Kettaneh, H. Friden and A. Holmberg (1998), Modelling and diagnostics of batch processes and analogous kinetic experiments, *Chemometrics and Intelligent Laboratory Systems*, 44, 331-340.

Wold, S., P. Geladi, K. Esbensen and J. Ohman (1987), Multi-way principal components and PLS-analysis, *Journal of Chemometrics*, 1, 41-56.

Wold, S., S. Hellberg, T. Lundstedt, M. Sjostorm and H. Wold (1987a), PLS modelling with latent variables in two or more dimensions, *Frankfurt PLS meeting*, Frankfurt.

Wold, S., S. Hellberg, T. Lundstedt, M. Sjostorm and H. Wold (1987b), PLS model building: theory and application, *Frankfurt am Main*, Frankfurt.

Workman, J., D. J. Veltkamp, S. Doherty, B. B. Anderson, K. E. Creasy, M. Koch, J. F. Tatera, A. L. Robinson, L. Bond, L. W. Burgess, G. N. Bokerman, A. H. Ullman, G. P. Darsey, F. Mozayeni, J. A. Bamberger and M. S. Greenwood (1999), Process analytical chemistry, *Analytical Chemistry*, 71, 121R-180R.

Yang, K. (2004), Multivariate statistical methods and six-sigma, *International Journal of Six Sigma and Competitive Advantage*, 1, 76-96.

Zeaiter, M., J. M. Roger and V. Bellon-Maurel (2005), Robustness of models developed by multivariate calibration. Part II: The influence of pre-processing methods, *Trends in Analytical Chemistry*, 24, 437-444.