

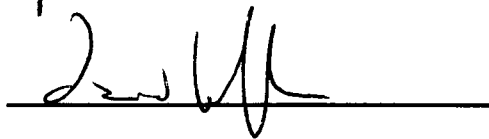
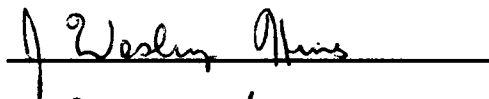
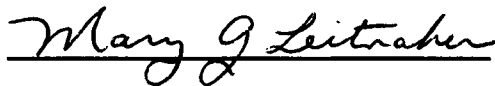
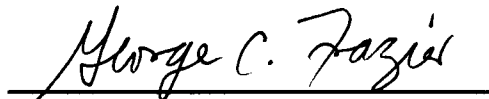
To the Graduate Council:

I am submitting herewith a dissertation written by Nitin Kaistha entitled "A Generic SPC Framework for the Characterization of Batch Profiles". I have examined the final copy of the dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Chemical Engineering.



Dr. C. F. Moore, Major Professor

We have read this dissertation
and recommend its acceptance:



Accepted for the Council:



Associate Vice Chancellor and
Dean of The Graduate School

**A GENERIC SPC FRAMEWORK FOR THE
CHARACTERIZATION OF BATCH PROFILES**

A Dissertation

Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Nitin Kaistha

December 1999

ACKNOWLEDGEMENTS

The help, support and guidance received from the various faculty members, staff, colleagues and friends in the completion of this work is gratefully acknowledged. It would be impossible to list all the people due to space and memory constraints. So please excuse any inadvertent omissions.

First of all, the moral support and guidance from Dr Charles Moore, formally author's advisor and informally akin to his father, is acknowledged. Suffices to say, that this work would not have been possible without him pushing in various subtle and not-so-subtle ways. The time spent here at the university shall always remain a cherished memory. The author would also like to express his sincere gratitude to the committee members – Dr Frazier, Dr Leitnaker, Dr Hines and Dr Keffer, for providing valuable suggestions and reviewing this work. The inputs of Dr Wang in the linear algebra course and through various informal discussions are also gratefully acknowledged.

All the staff members at the department of chemical engineering have been very helpful in all possible ways. The author would like to express his sincere appreciation for the warm help from the department secretaries, Betty and Sancy. The coffee at the department office has always been very special.

On the personal front, the help from the various friends is gratefully acknowledged. The author's family has always been a source of inspiration and support and this work is dedicated to them. Lastly, the author would like to acknowledge the constant help coming from his Master, Shri P Rajagopalachari.

ABSTRACT

This work develops a generic Statistical Process Control (SPC) framework for characterizing the systematic variability seen in a historical database of batch profiles in terms of meaningful scale parameters and studying the correlation with the final product quality. The complete framework, in contrast to existing methods, is geared towards giving meaningful results that can be easily connected to the actual process. The variability in the profiles is partitioned into two parts – consistent and inconsistent. The consistent variability is characterized using scale parameters. The consistent variability is further partitioned as along the time and measurement axes leading to time and magnitude scale parameters for the same, respectively. Time scaling refers to the alignment of events in a batch while magnitude scaling refers to the use of projection methods for extracting the systematic variability along the measurement axis. The framework thus integrates the time synchronization of the batch trajectories and the subsequent application of multivariate techniques for monitoring and quality correlation purposes.

Tools for time and magnitude scaling are developed. The use of Dynamic Time Warping (DTW) for time scaling is studied. A novel technique based on translation of feature vectors is also developed and is shown to be especially suited for time scaling batch profiles. Magnitude scaling is based on the application of projection methods such as Principal Component Analysis (PCA) for extracting directions of systematic measurement axis variability in the profiles. PCA is modified to an evolving factor type method, for extracting easily interpretable factors.

Multivariate SPC charts for process monitoring on the scale parameters and the residuals remaining after scaling are developed. The framework presented is an offline analysis and can be readily adapted for online monitoring purposes.

A polymethyl methacrylate batch polymerization simulation is used in order to demonstrate the application of the SPC framework. The importance of time scaling in explaining a significant amount of the overall variability in the profiles and also significant correlation with the final batch quality is established. Comparisons with multiway principal component analysis (MPCA), the existing framework, show the better sensitivity of the proposed technique to special cause disturbances. The ease in the engineering interpretation of the results as contrasted to MPCA is also established.

TABLE OF CONTENTS

CHAPTER		PAGE
1.	Introduction	1
2.	Literature Review	6
2.1	Batch Processes	6
2.2	Multivariate Regression and Projection Methods	12
2.2.1	Multivariate Regression	12
2.2.2	Projection Methods	14
2.2.2.1	Principal Component Analysis (PCA)	14
2.2.2.2	Partial Least Squares (PLS)	17
2.3	Time Alignment	19
2.4	Cluster Analysis	20
2.5	Polymethyl Methacrylate (PMMA) Polymerization Model ...	21
2.6	Summary	22
3.	Modeling Philosophy	23
3.1	Motivation	23
3.2	Partitioning of Variability	24
3.2.1	Consistent Variability	26
3.2.2	Inconsistent Variability	28
3.3	Mathematical Formulation	29

4.	Time Scaling Tools	35
4.1	Introduction	35
4.2	Time Scaling Tools	35
4.2.1	Linear Interpolation	36
4.2.2	Indicator Variable	37
4.2.3	Dynamic Time Warping	40
	4.2.3.1 Mathematical Formulation	40
	4.2.3.2 Example Application	47
4.2.4	Shapes and Features	50
4.2.5	Discussion	58
4.3	Extraction of Feature Vectors	61
4.3.1	Evolving Factor Analysis	62
4.3.2	Mathematical Formulation	62
4.3.3	Example Application	66
4.3.4	Discussion	75
5.	Magnitude Scaling Tools	78
5.1	Introduction	78
5.2	Principal Component Analysis (PCA)	78
5.3	Extraction of Meaningful Magnitude Scale Parameters	93
	5.3.1 Custom Scale Parameters Example	94
	5.3.2 Evolving Factors Analysis	98
5.4	Discussion	103

6.	SPC Charts	105
6.1	Introduction	105
6.2	Profile Projection Review	107
6.3	SPC Charts on Scale Parameters	109
6.4	SPC Charts on Residuals	113
7.	Batch Process Simulation	116
7.1	Introduction	116
7.2	Process Description	117
7.3	Model Equations	120
7.4	Simulated Historical Database	129
8.	Example Application	134
8.1	Introduction	134
8.2	Profile Analysis	134
8.2.1	Time Scaling	135
8.2.2	Magnitude Scaling	141
8.2.3	SPC Monitoring Charts	147
8.2.4	Comparison with MPCA	156
8.2.5	Manipulated Variables	165
8.3	Quality Correlations	167
8.4	Discussion	173
9.	Conclusions and Future Work	178
9.1	Summary	178

9.2	Conclusions	179
9.2	Future Work	182
LIST OF REFERENCES		186
APPENDICES		195
	APPENDIX A.....	196
	APPENDIX B.....	203
VITA		205

LIST OF FIGURES

FIGURE	PAGE
3.1 An example batch profile	25
3.2 Schematic of the modeling philosophy	33
4.1 Linear interpolation for time scaling	36
4.2 Extreme variation in batch duration	38
4.3 Time scaling using indicator variable.....	39
4.4 Schematic illustrating the slope constraint m	42
4.5 Global search constraint in DTW	44
4.6 DTW algorithm flow chart	46
4.7 Application of DTW for time alignment	48
4.8 Example of translation for extracting feature time	53
4.9 Feature templates used for time alignment	55
4.10 Time scaling using feature alignment.....	56
4.11 Indicator variable may not align all the consistent events	57
4.12 Spurious misalignment in DTW	60
4.13 Heater power profiles	67
4.14 Selected candidate features and maxminD	69
4.15 Extracted consistent features	71
4.16 A closer look at the initial part of the profiles	72
4.17 d vectors for a few profiles for the consistent features	73
4.18 Extracted times for the consistent features	73

4.19	Effect of shape factor on feature time extraction	74
5.1	Heater power profiles after time scaling	81
5.2	Scree plot for PCA of mean centered profiles	83
5.3	PCA on mean centered heater power profiles.	84
5.4	Standard deviation of heater power profile across time	85
5.5	Scree plot for PCA on autoscaled heater power profiles	86
5.6	PCA on autoscaled heater power profiles	87
5.7	Reactor temperature profiles	88
5.8	PCA on mean centered reactor temperature profiles	90
5.9	PCA on autoscaled reactor temperature profiles	91
5.10	Comparison between mean centered and autoscaled residuals	92
5.11	Use of factors obtained by division into zones	96
5.12	Comparison between projections using PCA and factors	97
5.13	Flowchart of the algorithm for extracting magnitude scale factors	99
5.14	Use of factors obtained by proposed EFA approach	102
6.1	Organization of batch data array	107
6.2	Pictorial schematic for magnitude scaling	110
6.3	PC model equations and statistics	112
6.4	Each batch as a slice in the 3D data array	115
7.1	PMMA batch process schematic	118
7.2	Free radical polymerization reaction mechanism	121
7.3	Kinetic rate equations for the reaction mechanism	124
7.4	Gel effect constitutive equations and parameters	125

7.5	Dynamic process model equations	126
7.6	Controller equations and parameters	128
7.7	Nominal profiles generated from the simulation	132
8.1	Effect of time scaling. Raw and time scaled profiles	138
8.2	Percentage variance explained by time scaling	140
8.3	Factors used for magnitude scaling for all the measured variables	142
8.4	Effect of magnitude scaling. Mean centered profiles and residuals remaining	144
8.5	Percentage variance explained by time and magnitude scaling	146
8.6	Univariate SPC charts on the time scale parameters with 3σ limits	148
8.7	Univariate SPC charts on the magnitude scale parameters with 3σ limits	150
8.8	Q statistic bar plot for all the batches with 99.7% confidence limit	153
8.9	Residuals for the abnormal batches	155
8.10	Scree plot for MPCA	158
8.11	Retained PC loadings for MPCA	158
8.12	Univariate SPC charts on scores for MPCA	159
8.13	Q statistic plot for MPCA	161
8.14	Residuals for abnormal batches	162
8.15	Magnitude scale parameters for batches 98 and 100	164
8.16	Quality predictions for MW_n using time scale parameters	171
8.17	Mean square error of prediction for all the quality variables	172
8.18	PC loadings for raw and time scaled profiles	175

LIST OF TABLES

TABLE	PAGE
7.1 Kinetic parameters used in the model.	125
7.2 Physical properties and other parameters for the process	127
7.3 Nominal and abnormal operation ranges for some process parameters	131
8.1 Time scaling zones and the method for each zone	137
8.2 Factors for each of the profiles	143
8.3 SPC chart violations for the time and magnitude scale parameters and the Q statistic	156
8.4 Correlation coefficient of time scale parameters with quality	169

Chapter 1

Introduction

Batch processes, as the name indicates, are used for the production of discrete quantities of product. Any process that treats raw material as a discrete entity for a finite amount of time to produce a finished product may be defined as a batch process. The simplest example is cooking a dish where a finite amount of raw material is processed in vessel for a finite duration of time to produce cooked food. In the process industry, virtually every unit operations can be run in the batch mode – crystallization, blending operations, distillation etc being the common examples. Batch processes are widely used for producing specialty chemicals, e.g., pharmaceuticals, polymers, biochemicals etc. These processes are ideally suited for manufacturing value added specialty chemicals where product volume is small, market demand is uncertain, the procedure is subject to change and operational flexibility is desired to produce different grades of the same product or a variety of products using the same equipment as dictated by market demand. With increasing emphasis on the production of such value added ‘niche’ area products, especially in the developed countries, there is substantial interest in the design, monitoring and control issues involved.

Batch processes are transient in nature with their state changing with time. The products are typically produced by following a specified recipe which usually consists of charging the batch vessel with a recipe of materials, processing under controlled conditions so that variables having significant effect on the product

characteristics follow a set trajectory and finally discharging of the product. On completion of the batch, a range of quality measurements is usually made at the quality control laboratory on a sample of the product, characterizing the batch as acceptable or poor quality. These quality variables are usually difficult to measure online. However data on several easily measurable variables such as temperatures, pressures and flows are taken online and considering the recipe based nature of batch processes, typically follow a trajectory or profile.

These profiles provide vital information characterizing the operation of the batch and should be used for monitoring the operation of a batch. Their dynamic non-linear characteristics however, pose a challenge and limit the use of monitoring methods as used in the continuous industry which operate around a fixed target or steady state. Statistical Process Control (SPC) charts, the primary monitoring tool, have thus been applied only to the final quality measurements at the end of each batch since the final desired product quality is a fixed target. The method characterizes only the batch to batch variability. Abnormal conditions can however develop during the course of a batch (within batch variability), which may lead to a poor quality product unless the problem is detected and remedied. The need for monitoring the profiles, which would reflect such abnormalities, as the batch develops is obvious. In spite of this, most industrial batch processes are run without any effective form of monitoring to characterize their operation and for the most part, rely simply on the precise sequencing and automation of the major steps in each run. Tools are therefore needed

for characterizing the variability in the batch profiles and studying the effect of such variability on the final batch quality.

The profiles are a 'fingerprint' of the batch operation and data on several successful and some unsuccessful batches is usually available in a historical database. This historical database can be used to empirically model the normal batch operation. The approach can also be extended to model the final batch quality from the profiles in case such quality data is available. The main difficulty lies in handling the vast amount of non-linear time trajectory data on highly correlated measurements with batches having variable duration.

This work addresses the above issues and develops a systematic generic SPC framework for the analysis and characterization of batch profiles using traditional multivariate approaches. The framework is applicable to any batch process with one or more online measurements and the trajectories for several past batches stored in a historical database. The emphasis is on quantifying the variability in the profiles in terms of simple meaningful scale parameters. These scale parameters quantify the consistent variability in the profiles.

Variability in batch profiles can occur in two ways, along the time scale and along the measurement scale. Variation in the duration of the batch is an example of the former while variation in the measurement itself is an example of the latter. SPC charts are proposed for the scale parameters and for the residuals remaining after extracting the consistent variability in the scale parameters. Techniques such as simple interpolation, time warping and extraction of shapes and features are used for

characterizing the time scale variability. Measurement scale variability is characterized using least square projection methods. The scale parameters which contain important information on the within batch variability are used to model the final batch quality by existing regression methods such as principal component regression (PCR) and partial least squares (PLS).

The emphasis of our approach is to partition the variability in a meaningful way so that abnormal scale parameters provide vital physical clues for diagnostic purposes and the effect on batch quality. The analysis of operational and quality control problems in past batches is thus facilitated. The framework is also easily adapted for monitoring the progress of new batches in the SPC sense.

Recently, much work has been done in the area of batch process monitoring which apply techniques from chemometrics. Historical batch data is usually 3 dimensional with batches, measurements and their time profiles forming the three dimensions. These methods unfold the 3-way data into a matrix and perform regular PCA or PLS on the huge unfolded matrix. Even though some applications have been reported which demonstrate the monitoring ability of the techniques, applications where the techniques are able to predict the batch quality well are scarce. Further, the 'black box' treatment to the data makes interpretation of the results cumbersome and therefore provides little insight for process understanding and subsequent improvement. Interpretation is further confounded by the standardization of the unfolded matrix so that all the variables and their time trajectories are on the same scale before analyzing the data. The issue of time scaling, a major source of

variability in batch profiles is also not considered. It is therefore felt that a framework oriented specifically towards meaningfully analyzing the variability in the batch profiles is needed. This work is an attempt in that direction.

The body of the dissertation consists of nine chapters. A detailed review on the relevant literature in the area of batch process monitoring is presented in chapter 2. In chapter 3, the philosophy for partitioning the variability in the empirical approach is discussed. Chapter 4 develops techniques for characterizing the time scale variability while Chapter 5 does the same for measurement scale variability. Chapter 6 is devoted to the development of control charts for the scale parameters and the residuals remaining. Chapter 7 describes a polymethyl methacrylate batch process simulation that is used for generating profiles in order to demonstrate the methodology. The demonstration forms the core of chapter 8. Results are compared with recently proposed multivariate techniques for monitoring batch processes. The dissertation concludes with a summary of the developed technique, conclusions drawn on the basis of experience with simulated and industrial data sets, unresolved issues and directions for further research in chapter 9.

Chapter 2

Literature Review

This chapter briefly reviews the existing literature on batch processes and the techniques that will be used in empirically modeling the batch profiles. The chapter is divided into five distinct sections. The first section reviews past work in the area of batch processes. Literature on relevant multivariate techniques such as Principal Component Analysis (PCA) and Partial Least Squares (PLS) is reviewed in the second section. The third section reviews techniques for time scaling. The fourth section briefly touches upon clustering analysis. The final section lists references for modeling the methyl methacrylate polymerization reactor, which is simulated to demonstrate the application of our developed methodology on a realistic process.

2.1 Batch Processes

Batch processes are very important to the chemical and manufacturing industries and are used for a variety of purposes. Shaw defines a batch process as consisting of a sequence of one or more steps (or phases) that must be performed in a defined order producing a finite quantity of finished product [1]. As pointed out by Fisher, batch operations are used when product quantities are small, reaction times are long, feedstock supplies and market demand are uncertain, manufacturing procedures are likely to change, and/or a large variety of small volume products are to be produced [2]. Thus batch processes are used in the production of specialty chemicals, e.g. pharmaceuticals, for separations, e.g. batch distillation and for the processing of

materials, e.g. the injection molding of plastics. Industrialized nations are now seeing a change in emphasis from continuous to batch processes reflecting a movement towards flexible specialty chemicals and biochemicals [2, 3].

Batch processes are different from the continuous and the discrete processes. Staples provides a comparison of the three types of processes from a control perspective [4]. Some key differences between a batch and a continuous process, are noted by Martin [5]. These processes are dynamic by nature and involve a sequence of phases carried out on a discrete recipe of material [2]. The control and monitoring issues involved are thus very different from those of continuous processes, which usually operate around a steady state. The unique nature of batch processes leads to a variety of automation, scheduling, monitoring and control issues. Industry has therefore moved towards a standard for batch processes [6]. The challenge lies in adjusting production schedules due to their dynamic nature and unpredictable market situation. References [7-9] are a few articles that study frameworks for scheduling batch processes. The various control issues involved in batch processes are elucidated in references [1-5, 10-13].

For the monitoring part, traditionally, the process industry has applied statistical process control (SPC) techniques to detect statistically significant deviations on simple control charts and the batch industry is no exception. Mamzic provides an overview of SPC techniques as applied in the process industry [14]. Oakland covers the subject of SPC in a basic but comprehensive manner, with emphasis on practical industrial applications [15]. Traditional SPC monitoring approaches focus on detecting

deviations on measurements of important variables from a target. Control charts initially proposed by Shewhart are used to monitor the variation in a measurement and detect any statistically significant deviations from the target [16]. Upper and lower control limits are computed with measurements lying outside these limits being deemed out of control and therefore requiring investigation to determine a cause.

These techniques are well suited for monitoring measurements in the continuous industry, which operate about a steady state or target. The dynamic or unsteady nature of batch processes severely limits their applicability. Traditional SPC has thus been used to monitor quality measurements on the final batch product in the batch. Vander Wiel et al report an application of SPC on quality measurements on a polymer resin produced in a commercial scale reactor [17]. On-line measurements on quality are usually not available so that these SPC methods serve only to monitor batch-to-batch variability in the quality variables. The possibility of detecting abnormal operation and taking some corrective action as the batch progresses in time is precluded. Hahn and Cockrum have studied a case where some on-line quality measurements are available [18].

Typically, on-line measurements are usually taken on several easily measured variables such as pressures, temperatures and flows. These measurements then provide an indirect means for monitoring the batch and studying the effect of deviations on the final product quality. Marsh and Tucker realized that such measurements do follow a certain dynamic pattern and proposed a simple Shewhart type SPC technique for monitoring the profiles [19]. The main criticism lies in their univariate orientation and

the lack of statistical reasoning. Pattern recognition has also been applied to batch profiles for qualitative monitoring purposes [20].

The Shewhart type control charts suffice for monitoring single quality or process variable measurements; however, chemical processes are typically multivariate in that several measurements on different process or quality variables are taken. Such multivariate measurements are handled by using traditional control charts based on the χ^2 and Hotelling's T^2 statistics [21-28]. Recently, Tracy et al have developed multivariate control charts for individual observations [24]. These multivariate methods, work well when the measurement space is not too large or ill-conditioned.

Chemical process or quality measurements are usually highly correlated to one another as the measurements are reflections of the same underlying physical phenomena. This causes the data to be ill-conditioned and therefore traditional multivariate techniques have limited application. Newer approaches based on Principal Component Analysis (PCA) and Partial Least Squares (PLS) are able to handle large ill-conditioned measurement spaces by projecting the data onto better conditioned subspaces and then applying traditional multivariate statistical methods on the subspace. PCA and PLS have their origin in chemometric research and are treated in a subsequent section. Wise et al gave a theoretical basis for using principal components for monitoring multivariate processes by drawing comparisons between state space and principal component models [29].

The SPC monitoring technique using projection methods is data driven and basically finds latent vectors that explain the significant systematic variation in the

data. The data is thus divided into two orthogonal subspaces, the projection onto the latent vector space and the residuals. The latent vectors have some nice orthogonality properties so that traditional T^2 statistic based SPC charts are used to monitor the variation in the latent vector subspace. The residuals are deemed as due to noise and common cause variability in the process. SPC charts based on the squared sum of the residuals are used to monitor the variation in this subspace. Both the charts together monitor the variability in the whole measurement space and are thus complementary to each other. Multivariate process monitoring using PCA and PLS is treated in detail in references [30-34].

There has been a significant amount of work done on monitoring continuous processes using these projection methods. Several industrial applications in the continuous process industry have been reported [36-39]. Some simulation studies have also been reported [34, 35].

The multivariate SPC charts based on PCA and PLS for the continuous process industries operating about a target have been extended to handle batch profile data [40, 41, 42]. Batch data is inherently 3-way in nature with the batches, the different measured variables and their time profiles forming the three dimensions. The approach basically unfolds the three dimensional data array into a 2 dimensional matrix and applies regular PCA or PLS to the huge unfolded data matrix with χ^2 or T^2 based statistics for control charting, as in the continuous industries. The approach is called Multiway PCA (MPCA) or Multiway PLS (MPLS). An overview of the multivariate SPC techniques can be found in references [43, 44].

The monitoring capability of the approach has been demonstrated on both simulated and industrial data sets. Nomikos demonstrates the application of MPCA on a simulated styrene butadiene latex reactor and also on an industrial batch polymerization process [40, 41]. Kourti et al report a couple of industrial applications [45]. Gallagher et al have applied MPCA to monitor a nuclear waste storage tank [46]. Debashis and Schlags study the application of MPCA and MPLS methodologies for monitoring emulsion batch polymerization processes [47]. The linear MPCA technique has been extended by Dong and McAvoy to non-linear principal components [48].

As such, the MPCA approach and its non-linear extension remain data compression techniques that have aided in the analysis of vast amount of historical batch data. Their main criticisms lie in the 'black box' approach with little emphasis on understanding and interpreting the principal components. The scaling used to standardize the different measurements across the different phases also is questionable. The issue of time scaling, a major source of variability is also not given due consideration. The SPC charts, though few in number, are therefore difficult to interpret and subsequently diagnose, thus limiting their applicability in the real world. Miller has proposed the use of contribution plots for assigning a 'numerical' cause for abnormalities detected on such control charts [49]. Nomikos applies contribution plots for diagnosing abnormal batches based on such contribution plots [50].

Addressing the issues listed in the previous paragraph forms the core of this research. Literature on regression techniques and projection methods such as PCA and

PLS, which form the basis of the empirical modeling approach for batch process monitoring and quality predictions, is now reviewed.

2.2 Multivariate Regression and Projection Methods

In this section, a brief overview of the various projection methods for characterizing the variability along the measurement scale in the batch profiles is given. Regression techniques used for predicting an output from a set of input variables are also reviewed as these techniques are used for making predictions on the final batch quality from the profile data.

2.2.1 Multivariate Regression

Multivariate regression is a subject in itself and is treated exhaustively elsewhere [51, 52, 70, 71]. Consider the prediction of an output y from a set of measurements x that are assumed to have an effect on the output. Both x and y are row vectors. Typically, as a first pass, a linear model is built to fit the available data, historical or from designed experiments, so that a 'best fit' model is obtained. Usually 'best fit' is understood in the least squares sense.

In matrix notation, consider the linear regression equation

$$Y = XB \quad \dots\dots\dots(2.1)$$

where

Y: n by m output matrix (n samples y of m output variables)

X: n by p input matrix (n samples x of p input variables)

B: p by m regression coefficient matrix

The regression coefficients **B** are found by inverting the **X** matrix. Typically, the **X** matrix is tall and skinny, since several samples are available, i.e., $n \gg p$. Thus, even though a theoretical inversion is not possible, the pseudo inverse of **X** is taken, which involves the projection of **Y** onto the column space of **X** and subsequent regression of **X** onto the projected **Y**. The solution thus obtained is

$$\mathbf{B} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \quad \dots\dots\dots(2.2)$$

$(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ is known as the pseudo inverse of **X** and minimizes the squared sum of error in prediction of **Y** [52]. Any new sample **x** can be used for prediction using the regression coefficients **B** as

$$\mathbf{y}_{\text{pred}} = \mathbf{x} \mathbf{B} \quad \dots\dots\dots(2.3)$$

The above method works fine as long as the **X** matrix is tall and skinny, ie there are more samples than the number of measurements so that $\mathbf{X}^T \mathbf{X}$ is not rank deficient and hence invertible. Even though a rank deficiency may not exist, an ill conditioned or almost singular $\mathbf{X}^T \mathbf{X}$ matrix upon inversion would give unstable estimates of the regression coefficients with high variance [55]. This high variance would render SPC control charts useless.

Process data, as noted before, can be highly ill conditioned. Further, in the case of batches, the profile matrices are fat and short ($p \gg n$). Thus techniques are needed that retain most of the significant information in the profiles in a few parameters. These few parameters can then be regressed with the output by traditional pseudo inverse type methods. Projection methods such as PCA and PLS are able to handle such ill-

conditioned problems without significant loss in information and are reviewed in the following paragraphs.

2.2.2 Projection methods

PCA and PLS are projection methods that have been extensively used in chemometric research for the calibration of spectra to predict the concentrations of the various species [54-76]. The analogy between spectra and time aligned batch profiles makes these projection methods an ideal choice for quantifying the variability in the batch profiles and predicting the batch quality outcomes. The techniques are briefly reviewed below.

2.2.2.1 Principal Component Analysis (PCA)

PCA is a linear exploratory data analysis technique that finds the directions of maximum variance in a data set [53-56]. These directions are known as the principal component loadings and the projections of the data samples onto the principal loadings are known as the sample scores. Specifically, consider a data matrix $\mathbf{X}$, which is n by m with the various samples along the rows. Each sample consists of measurements on m different variables. PCA decomposes the $\mathbf{X}$ matrix as a product of scores and loading matrices and residuals

$$\mathbf{X} = \mathbf{T}\mathbf{V}^T \dots\dots\dots(2.4)$$

where

$\mathbf{T}$: Sample scores matrix

V: Loading matrix

The loading matrix, **V**, is orthonormal and its columns span the row space of **X**. Each column of **V** represents a linear combination of the original variables and the sample scores are the projections onto these loadings. The loadings or principal components (PCs) are calculated so that the first PC is in the direction of maximum variance in the row space of **X**, the second PC is in the direction of maximum variance of the residual remaining and so forth. These loadings turn out to be orthonormal to each other since the covariance matrix $\mathbf{X}^T\mathbf{X}$ is symmetric, leading to orthogonal eigenvectors, which form the PCs [55]. The scores are therefore decorrelated. PCA thus represents a rotation of the original laboratory frame of reference to a newer one, comprising of the PCs as the basis vectors, so that in the rotated frame, the data appears uncorrelated.

PCA is based on the singular value decomposition (SVD) of a matrix **X** which is explained in [52]. Wang gives a geometric interpretation of SVD and its relation to PCA [56]. SVD decomposes a matrix as follows

$$\mathbf{X} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T \quad \text{.....(2.5)}$$

where

U: n by n orthogonal matrix. Columns of **U** span the column space of **X**

Σ : n by m diagonal matrix of singular values in descending order

V: m by m orthogonal matrix. Columns of **V** span the row space of **X**

From the PCA perspective, **V** is the loadings matrix while **U Σ** forms the scores matrix. The singular values measure the variance of the data along the corresponding

PC. Ill conditioned X matrices therefore result in very small singular values towards the end indicating that variation along the corresponding PCs is probably due to noise. Such PCs can then be ignored and only the first few significant ones are kept. Rejection of the last few 'insignificant' PCs removes the near singularity in the X matrix thus improving its conditioning. Assuming r out of m PCs are retained, the PCA decomposition thus becomes

$$X = T_r V_r^T + E \quad \text{.....(2.6)}$$

where

T_r : n by r matrix of retained scores

V_r : m by r matrix of retained PC loadings

E : n by m matrix of residuals remaining (Not explained by the PCs)

The method thus compresses the significant systematic variation in the data into a few scores. In the PC subspace, the projected data is well conditioned so that traditional multivariate regression techniques can be readily applied. The regression of the retained scores T_r with output Y is possible and is called Principal Component Regression (PCR). The rejection of insignificant PCs leads to robust matrix inversion. The regression equations in section 2.2.1 can be used with the retained scores matrix T_r replacing the X matrix. Dong extends the method to non-linear PCA [57].

The power of PCA lies in detecting and rejecting non-significant directions, thus improving the conditioning of the matrix. It is an unsupervised method that takes into consideration only the X data. It is therefore quite possible that information that is predictive of outcomes Y but is insignificant in terms of its variance may be lost in the

higher PCs that are rejected. This problem is partially taken care of by PLS which is the subject of the following subsection.

2.2.2.2 Partial Least Squares (PLS)

Partial Least Squares or Projection to Latent Structures is an supervised technique that takes into consideration the variation in the Y space along with that in the X space to extract latent vectors (similar to principal components) that are more predictive of Y . PLS is based on the NIPALS algorithm [58] and decomposes the X and Y matrices as follows

$$\mathbf{X} = \mathbf{T}_r \mathbf{V}_r^T + \mathbf{E} \quad \dots\dots\dots(2.7a)$$

$$\mathbf{Y} = \mathbf{U}_r \mathbf{W}_r^T + \mathbf{F} \quad \dots\dots\dots(2.7b)$$

where $\mathbf{T}_r$ and $\mathbf{U}_r$ are the retained scores matrices, $\mathbf{V}_r$ and $\mathbf{W}_r$ are the retained loadings and $\mathbf{E}$ and $\mathbf{F}$ are the residuals remaining for the X and Y matrices respectively. For understanding purposes, PLS rotates the X and Y spaces so that the scores in the two spaces line up which improves the predictions. The regression is done between the retained scores by means of a linear inner relation

$$\mathbf{U}_r = \mathbf{T}_r \mathbf{B} \quad \dots\dots\dots(2.7c)$$

In subsequent chapters, the subscript 'r', indicating the number of retained PCs or latent vectors on the loading and score matrices is dropped for ease of notation and it is understood that the first few significant PCs are retained.

While PCA calculates PCs so as to maximize the variance ($\mathbf{X}^T \mathbf{X}$) explained in X , PLS calculates the loadings by maximizing the covariance $\mathbf{X}^T \mathbf{Y}$ explained. Geladi

and Kowalski give a tutorial on PLS which should be a good starting point for understanding PLS [59]. Höskuldsson explains the geometric properties such as orthogonality of the scores etc of PLS and is probably the best reference for a geometric understanding [61]. Interpretations from various other viewpoints such as maximizing a covariance, rotation of the subspaces and so forth are also given. Other references are also available [60, 62, 63]. Modifications to the original NIPALS algorithm have been proposed for computationally more efficient latent vector extraction [64, 65, 66]. Haaland and Thomas relate PLS to other quantitative regression methods such as PCR [67]. Similarities between the basic framework of PCA and PLS are explained by Höskuldsson [68]. Thomas and Haaland contrast the various multivariate methods such as PCR, PLS and multiple linear stepwise regression (MLSR) [69]. The subject of multivariate analysis has received much attention in chemometric literature and the reader is referred to [70, 71] for a comprehensive treatment of the subject.

The method of PCA and PLS has been extended to handle 3 or higher dimensional data arrays and is called Multiway PCA (MPCA) or Multiway PLS (MPLS) [72, 74]. MPCA and MPLS form the basis of monitoring schemes for batch data, which as discussed earlier, are three dimensional. Wangen and Kowalski have extended PLS to multi-block PLS to handle physically distinct blocks of data [73]. Wold extends the PLS algorithm to a non-linear inner relation [75, 76].

The above projection techniques taken from chemometric research are used to compress the variability in the batch profiles along the measurement scale. These

techniques are meaningful only when the various profiles are aligned along the time axis. Relevant past work in the area of time alignment is listed in the next section.

2.3 Time Alignment

Batches have variability along the time axis in that different batches may not progress at the same rate due to physical reasons. Examples would include slower heat transfer due to reactor fouling, slower kinetics due to impurities in the raw materials etc. The rate of progress of the batch may vary from batch to batch so that different batches have different lengths. Inconsistencies in operator intervention also cause variability in the time development of batch profiles. Before any meaningful statistical analysis can be made, the timing differences in the batches need to be reconciled [40]. Darnell showed that the variability along the time scale needs to be treated before any attempt can be made at meaningfully explaining the variability along the measurement scale [78]. A simple profile scaling model was also developed.

Konstantinov and Yoshida recognized that various profiles do follow a certain dynamic pattern and applied qualitative reasoning based on the occurrence of distinct events in the profiles [77]. The methodology is applied to an industrial fermentation process. The approach is however very qualitative and lacks statistical reasoning.

Debashis and Schlags have used reaction extent as an indicator variable for aligning the profiles onto the same time scale [47]. Other methodologies rely on simple linear interpolation, padding the profiles with zeros and so on. These are however ad hoc and do not lead to alignment of the features in the profiles since batch processes

are highly non-linear and therefore may need local compression and expansion of the time axes of the profile for a good alignment. Dynamic Time Warping (DTW), a technique from speech recognition is capable of such local compression and expansion of the time axes [80-84].

DTW is a distance based optimization that uses the principle of optimality [79] to search a non-linear warping of the time axis of two profiles so that a time normalized distance is minimized. The basic technique and minor modifications can be found in references [80-83]. The theory of time warping is explained in a later chapter. For present purposes, suffices to say that it shifts some feature vectors in time, compresses some and/or expands others so that a minimum distance is achieved. It is therefore well suited to handle the alignment problem in batch profiles and a recent application to batch profiles has been reported [84].

2.4. Cluster Analysis

Batch profiles can have different shapes depending on the recipe being operated. Before a monitoring model can be built, these different shapes need to be classified. Further more, within a monitoring model, the various scale parameters may form separate clusters depending on the kind of fault occurring, say a heat transfer problem or excessive impurities in the raw materials, for example. Automatic clustering and classification techniques are therefore needed.

Clustering is a field in itself and the reader is referred to references [85, 86] for an exhaustive treatment of the subject. For our purposes, it is sufficient to note that

most clustering algorithms are distance based methods. The clustering can be done in two ways - by specifying the number of cluster centers to be identified or by giving a distance tolerance. In the former, the cluster centers are initialized and iteratively updated while in the latter, a new cluster center is formed whenever a sample exceeding the distance tolerance from the cluster centers is encountered.

The various techniques briefly reviewed in the last few sections are adapted for application to batch profiles. The application of the adapted techniques on batch profiles must be demonstrated on a realistic batch process. A simulated batch reactor for polymerizing methyl methacrylate to polymethyl methacrylate is used to generate profiles for the purpose. Literature on the methyl methacrylate polymerization for modeling the batch reactor is now listed.

2.5 PolyMethyl Methacrylate (PMMA) Polymerization Model

The techniques developed for SPC monitoring of batch processes are demonstrated using a simulation of a polymethyl methacrylate batch reactor. The polymerization reaction occurs in toluene solvent with AIBN as the initiator. The polymerization occurs through free radical polymerization with initiation, propagation, chain transfer and termination as the steps [87-90]. The reaction is characterized by a particularly strong gel effect, which causes autoacceleration in the overall rate of polymerization [91]. The kinetics of the free radical polymerization are highly non-linear and are modeled by using the quasi steady state hypothesis (QSSA) on the intermediate species [88]. Baillagou et al have developed a computer simulation model

that does not incorporate the QSSA approximation [89-90]. The gel effect has been modeled by using the constitutive equations of Chiu et al [91]. The molecular weight distribution and polydispersity of the produced polymer are modeled by using the method of moments [90]. Soroush and Kravaris use the model to implement non-linear control on an experimental batch polymerization reactor [92]. The authors have also elaborated on the framework for batch process in another publication design [93].

2.6 Summary

In this chapter, a brief outline of the relevant work in the area of batch process monitoring has been given. Literature on multivariate empirical modeling techniques such as PCA and PLS has also been listed. References for the batch process simulation model and time scaling techniques are also enumerated.

The next few chapters are devoted to the development of a framework for analyzing the variability in the batch profiles in a meaningful way that enhances process understanding. The modeling philosophy is described in the next chapter. The subsequent chapters deal with the tools that must be used in order to implement the philosophy.

Chapter 3

Modeling Philosophy

3.1 Motivation

Today's batch processes are well automated with several online measurements available on easily measured variables such as pressures, temperatures and flows. These profiles contain information on the operation of the batch and are typically available in a historical database. Variability exists in these profiles and the database on past successful operation can be used to model it in the 'normal' batch profiles for statistical process control (SPC), the assumption of course being that 'abnormal' operation would cause high variability in the profiles thus causing statistical control limit violations. Further, the online measurements can provide indirect clues about the final batch quality, typically measured offline after the completion of the batch. Finally, a model of past normal batch operation would lead to deeper process understanding.

An empirical model of the normal batch operation is therefore desirable for

- (a) SPC monitoring of new batches against the model of past successful batches.
- (b) Using the model for understanding the correlation with final batch quality.
- (c) Helping develop deeper process understanding and subsequent process improvement.

It may be pointed out that the historical database on batch profiles is huge. A data set on 100 batches with 10 measured variables with each batch lasting about 5 hours with a sampling rate of 1 per minute would consist of $100 \times 10 \times 5 \times 60 = 300,000$

points. Any model should therefore try to model the variability in the batch profiles in terms of simple numbers (scale parameters) that meaningfully capture the variability in the batch profiles. This leads to the idea of partitioning the variability in the batch profiles, which is dwelled upon in the next section.

3.2 Partitioning of Variability

First of all, variability in a batch profile must be defined. Figure 3.1 plots profiles for several batches on a particular measured variable. As can be seen from the plot, all the profiles are transient in nature with a similar underlying shape. An average of the 'normal' profiles characterizes the normal operation and is known as the reference profile. It captures the typical transients in the profile, the major nonlinear component in batch profiles. Given a reference profile, the difference between a raw profile and the reference profile is an indication of the variability in the raw profile and measures by how much the new profile is different from the normal operation reference profile. This variability about the reference is to be partitioned in a systematic manner.

Variability in batch profiles is broadly classified as occurring because of two sources, consistent and inconsistent. These two sources are best explained by an example. Consider a batch reactor on which the reactor level is measured. Sensor measurement noise would always be present in the profile and is an example of inconsistent variability.

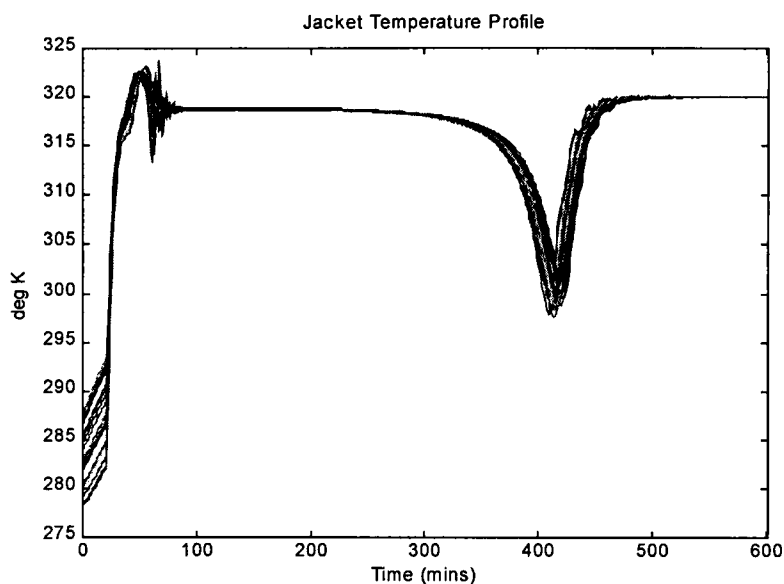


FIGURE 3.1: An example batch profile

In order to explain the consistent part, let us say that there is some variability in the initial charge to the reactor. Assuming constant reaction mixture density, the level profiles would be flat so that the difference between a profile and the reference profile would be a constant across time. Here the term reference profile refers to the mean of all the 'normal' operation level profiles and would again be flat with its value indicating the average amount charged to the reactor. The constant difference between a raw profile and the reference profile can be attributed to the variability in the initial charge, a distinct physical cause. Assuming that variability in the initial charge to the reactor always occurs, this shift would always be present in the level profiles i.e. the shift occurs consistently in all batches. The variability is thus due to a source that is consistently present in all the batches and is therefore called consistent variability.

Looked at in another way, consistent sources of variability cause the profiles to differ from the reference in a particular way. In the above example, all the profiles differ consistently from the reference by a shift. A scale parameter can be proposed in order to quantify such consistent sources of variability about the reference profile. In the example, the scale parameter would be the shift (above or below) from the reference. Subtracting this shift from the profiles would lead to projected profiles that are very close to the reference and the difference of the projected profile from the reference would be an indication of the remaining inconsistent variability.

The above paragraphs summarize the modeling philosophy in simple terms. The objective is to partition the 'normal' variability about the reference profile into two parts, consistent and inconsistent. The consistent variability in a raw profile is captured in scale parameters leading to a projected profile that is very close to the reference. The consistent part is thus captured in the scale parameters while the inconsistent part remains in the residuals remaining. The algebra for obtaining the projected profile by scaling is dealt with in a later section. The next section dwells on the partitioning of consistent variability in batch profiles.

3.2.1 Consistent Variability

The consistent variability in batch profiles can further be partitioned into two parts, along the time scale and along the magnitude scale. Time scale variability is again best explained by means of an example.

Consider the case where an operator has to open a valve in order to let the reactor product out. The product flow profile would show a sudden jump from zeros to a high value when the valve is opened. The time at which this is done may vary from batch to batch depending on the reaction rates. Thus if the reaction rate is slow, the time taken to achieve a particular conversion would be large while for fast rates, less time would be taken. Clearly, variability exists along the time axis and the time axis of a profile has to be adjusted so as to match up with that of the reference, before any comparison can be made between the reference profile and the new profile. This adjustment of the time axis is known as *time scaling*. Its need is further amplified when we consider that a batch typically consists of various phases with variability existing in the duration of each of the phases. The variability along the time scale can be extracted in terms of the times of occurrence of any consistently identifiable checkpoints that occur as a batch progresses in time. The duration of each of the phases would be simple examples of such time scale parameters quantifying the time scale variability.

Once the batches are time scaled, consistent variability along the measurement axis of the batch profiles can be accounted for in terms of scale parameters and this is referred to as *magnitude scaling*. The parameters so obtained are known as magnitude scale parameters. The profile shift parameter as explained in the previous section is an example of magnitude scaling.

The consistent variability in the batch profiles is thus quantified by the time scale and magnitude scale parameters. Note that the model building exercise leads to

a better understanding of the process and the scale parameters provide vital clues about the physical sources of variability in the profiles.

3.2.2 Inconsistent Variability

The discussion on the consistent variability in the last subsection indicates that variability in batch profiles does not always occur in the profiles and therefore cannot be captured in terms of scale parameters, forms the 'inconsistent' part. Special causes that do not occur consistently from batch to batch but are of interest for detection are an example of 'inconsistent' sources. Given that the consistent variability in a raw profile has mostly been extracted in terms of scale parameters to give the projected profile, the residuals remaining between the reference profile and the projected profile provide an indication of inconsistent variability. These residuals can be lumped in a squared sum of error (SSE) type statistic thus giving a parameter quantifying the inconsistent variability.

The variability in the batch profiles is thus quantified in terms of simple numbers, the scale parameters and the SSE type statistic. Control charts on the scale parameters and the SSE type statistic can be proposed with statistical limits being obtained from the 'normal' batches. These control charts can then be used for SPC monitoring. Also, the scale parameters, that contain information on most of the consistent variability in the profiles, can be correlated with the final batch quality and used to build a model for predicting the batch quality, if high correlation exists. The

whole exercise leads to a better understanding of the variability in the process, which may lead to subsequent process improvement. This, in short, forms the gist of our modeling philosophy. A brief mathematical formulation of the profile scaling follows in the next section.

3.3 Mathematical Formulation

This section briefly describes the scaling technique (profile projection method) used for characterizing the consistent variability in the profiles to obtain the scale parameters. As explained previously, the scale parameters capture the consistent variability in the profiles along the time scale and along the magnitude scale. These scale parameters are proposed so as to capture the consistent common cause variability in the batch profiles.

Darnell showed that variability along the time scale needs to be handled independent of the magnitude scale variability. Also all the profiles need to be put on a common time axis for studying the magnitude scale variability. For time alignment, any consistently occurring checkpoints as the batch develops in time can be used. These times would vary from batch to batch and serve as indicators of similar underlying dynamic phenomena, thus justifying their use for time alignment. Simple linear interpolation can be used between these checkpoints. The use of indicator variables such as reaction conversion is also possible. However, such indicator variables are often not available and one has to rely on the consistent checkpoints for time scaling.

Once the batches are put on a common time axis, consistent variability along the magnitude scale can be extracted by using physical models or from empirical techniques such as Principal Component Analysis (PCA). Darnell proposed a simple two parameter model based on considerations on the type of variability that is usually seen in batch profiles. The model is basically a simple linear regression with intercept between the time scaled profile ($\mathbf{p}$) and the reference profile ($\mathbf{r}$). The two magnitude scale parameters are the slope and the intercept. Recently, data driven techniques, such as PCA and PLS, have been used by many researchers, to extract the most significant directions of variations in the profiles from the normal batch data. The scale parameters are the projections on these significant directions and are also called scores.

Notwithstanding how these directions are obtained, from physical considerations or from empirical techniques, given the projection directions, the projection methodology remains simple matrix algebra. Let $\mathbf{p}_i$ denote the i^{th} time scaled batch profile and $\mathbf{r}$ the reference profile where $\mathbf{r}$ is obtained as the mean of normal batch profiles. Both $\mathbf{p}_i$ and $\mathbf{r}$ are row vectors. The consistent variability in $\mathbf{x}_i = \mathbf{p}_i - \mathbf{r}$, is to be scaled by projecting on the columns of $\mathbf{H}$, where the columns of $\mathbf{H}$ contain the directions of projection. Note that $\mathbf{H}$ may be obtained from PCA in which case $\mathbf{H} = \mathbf{V}$, the PC loadings matrix, or from other methods like evolving factors analysis. The directions for projection may also be proposed based on the operator's judgement about the typical variability seen in the profiles.

If $\mathbf{ms}_i$ denotes the scale parameter vector, then the projection directions in the columns of $\mathbf{H}$ are to be scaled by $\mathbf{ms}$ so that

$$\mathbf{x}_i = \mathbf{ms}_i \mathbf{H}^T + \mathbf{e}_i \quad \dots\dots\dots(3.1a)$$

or

$$\mathbf{e}_i = \mathbf{x}_i - \mathbf{ms}_i \mathbf{H}^T \quad \dots\dots\dots(3.1b)$$

so that the least square error $\mathbf{e}_i \mathbf{e}_i^T$ is minimum. The 'best' solution is obtained as

$$\mathbf{ms}_i = \mathbf{x}_i (\mathbf{H}^T)^+ \quad \dots\dots\dots(3.1c)$$

where $(\mathbf{H}^T)^+$ denotes the pseudoinverse of the $\mathbf{H}^T$. If $\mathbf{H}$ is full column rank, then $(\mathbf{H}^T)^+$ is obtained as $\mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1}$. In case of batch profiles, $\mathbf{H}$ would typically be column full rank. Rank deficiency would indicate the presence of a redundant direction for the scaling which implies meaningless scale parameter and should necessarily be avoided. Also, in case PCA is used to obtain the projection directions, the $\mathbf{H}$ matrix is the same as the loadings matrix $\mathbf{V}$ is orthogonal so that $\mathbf{H}^T \mathbf{H} = \mathbf{V}^T \mathbf{V} = \mathbf{I}$. Thus the projection equations for PCA become

$$\mathbf{ms}_i = \mathbf{t}_i = \mathbf{x}_i \mathbf{V} \quad \dots\dots\dots(3.2a)$$

and

$$\mathbf{e}_i = \mathbf{x}_i - \mathbf{t}_i \mathbf{V}^T = \mathbf{x}_i (\mathbf{I} - \mathbf{V} \mathbf{V}^T) \quad \dots\dots\dots(3.2b)$$

Note that if all the consistent common cause variability is extracted in the scale factors, the SSE $\mathbf{e}_i \mathbf{e}_i^T$ would be small in case the profile is normal. Special causes that have an effect on the profile (observable) would usually not be accounted for by the scaling and would give abnormally high $\mathbf{e}_i \mathbf{e}_i^T$. Extreme common cause variability in the form of abnormal scale parameters is also possible which can be

interpreted as along the time scale and/or the magnitude scale. Profiles may be termed as abnormal based on statistical control limit violations on SPC charts for the scale parameters and charts based on $\mathbf{e}_i \mathbf{e}_i^T$. The control limits can be obtained from the normal operation batches. The detailed formulation of the control charts and the derivation of the control limits is given in chapter 6.

Qualitatively, it is worth noting that since the consistent variability is extracted in the scale parameters, $\mathbf{e}_i \mathbf{e}_i^T$ (after scaling) would be quite small when compared with $\mathbf{x}_i \mathbf{x}_i^T$ (no scaling) for the normal operation batches. This gives tighter control limits on control charts based on $\mathbf{e}_i \mathbf{e}_i^T$ than if the charts were based on $\mathbf{x}_i \mathbf{x}_i^T$. This makes the monitoring charts sensitive to subtle special cause deviations not accounted for by the scale parameters, a very desirable feature.

In summary, the modeling philosophy attempts to explain the systematic variability in the batch profile in terms of simple time and magnitude scale parameters. These scale parameters and the SSE statistic on the residuals remaining after projection are used for SPC monitoring of new batches with control limits obtained from past 'normal' batch operation. Correlation of the scale parameters with the final batch quality may also be studied and high correlation can be used to build a regression model from the scale parameters giving online predictions of the batch quality. The complete exercise leads to a better understanding of the variability in the batch process. Figure 3.2 depicts the above philosophy.

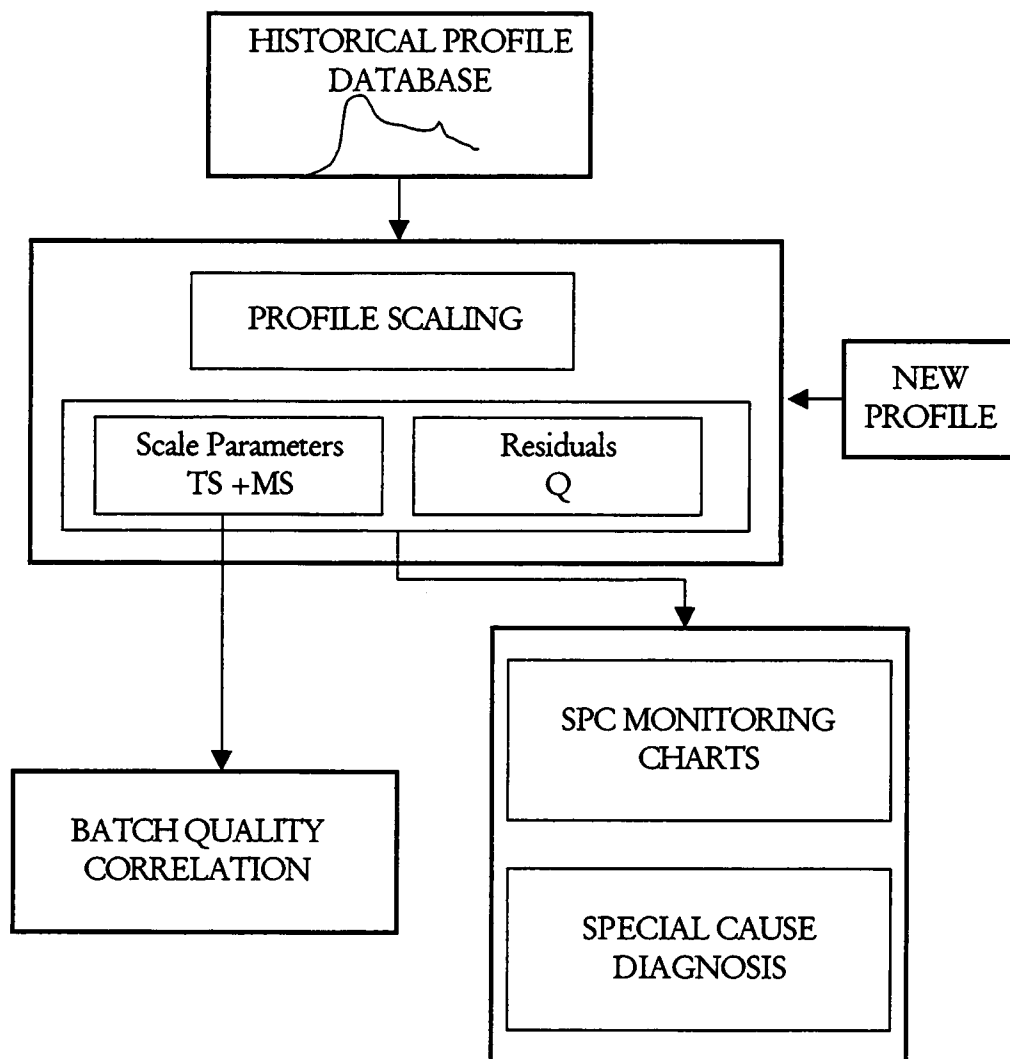


FIGURE 3.2: Schematic of the modeling philosophy

The next two chapters explain the tools to be used for scaling the profiles for consistent variability. Chapter 4 details the time scaling tools while chapter 5 explains the methodology for extracting the magnitude scale directions to be used for projection.

Chapter 4

Time Scaling Tools

4.1 Introduction

Variability along the time axis forms the major component of the overall variability in batch profiles and the time axes of all the batches must be synchronized before any reasonable analysis of the variability along the measurement axis can be done. This is known as time scaling. Reasons such as varying reaction rates due to variations in the raw materials, seasonal variations in the heat transfer characteristics of the equipment or inconsistent operator intervention can be cited for such time scale variability.

Various tools can be used for synchronization of the trajectories. These vary in complexity from simple linear interpolation of the time axis to a complex non-linear warp of the time axis using dynamic time warping. Extraction of the times of occurrence of various events within a batch is the intermediate case in terms of complexity. This chapter briefly describes these tools with example applications along with a discussion of issues involved in their application to real industrial batch processes. A simple methodology for the extraction of consistent events that can be used for time scaling is also presented.

4.2 Time Scaling Tools

In this section, tools for alignment of the time axis of a batch profile with that of the reference profile are briefly discussed. Time scaling is necessary since all the

batches should have the same number of points before any empirical techniques such as PCA can be applied for analyzing the profiles. Further, significant variation in the duration of batches exists. Examples where the typical batch duration may vary from say six to ten hours are not uncommon. The need for time scaling is therefore imperative. A brief description of the tools for time scaling follows.

4.2.1 Linear Interpolation

A simple technique for time alignment is to linearly interpolate all the profiles to be of the same length as the reference profile. This would suffice in case the variation in the duration of a batch is not much and is illustrated in figure 4.1. Typically, since batches usually consist of various phases, linear interpolation can be used within a phase so that all the profiles are of equal length for that particular phase.

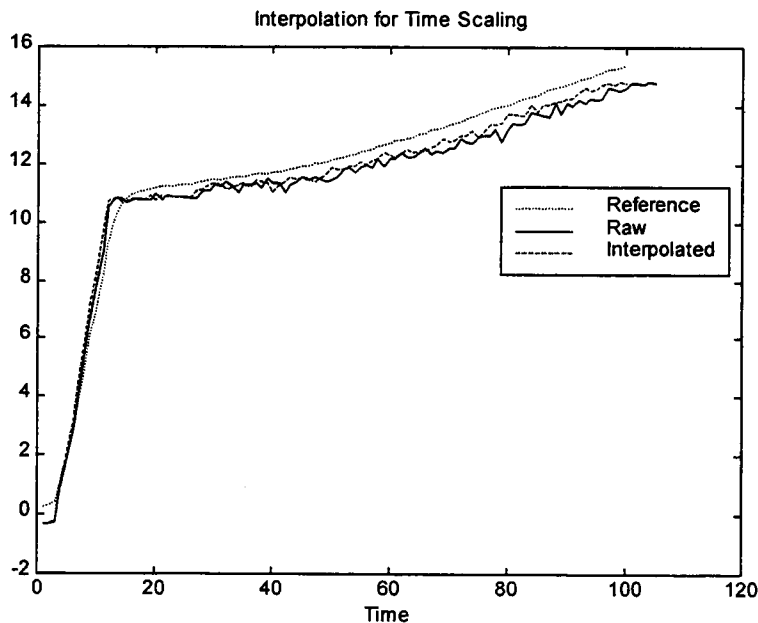


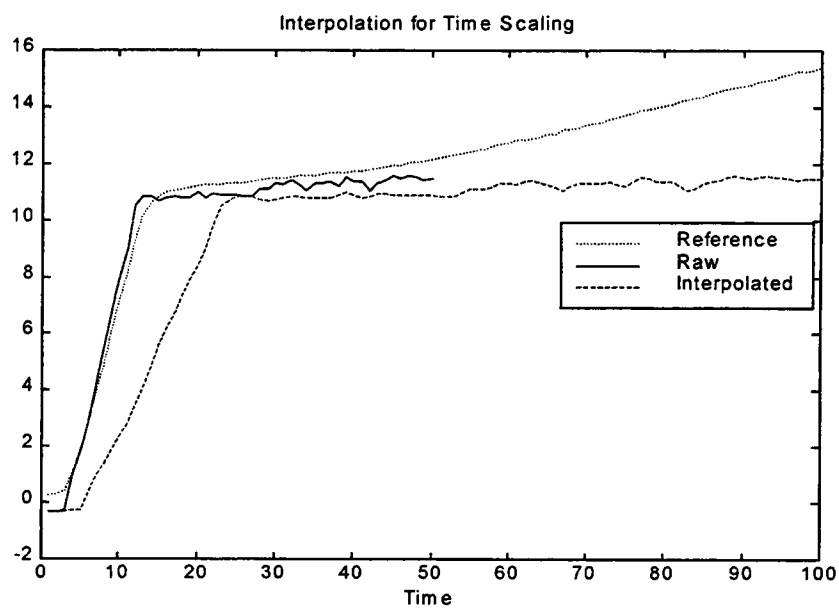
FIGURE 4.1: Linear interpolation for time scaling

Interpolation may not work for all the phases since in the same batch, some phases may be longer while others shorter than in the reference implying the need for simultaneous compression and expansion of the respective phases. It is thus suitable for time scaling phases in a batch.

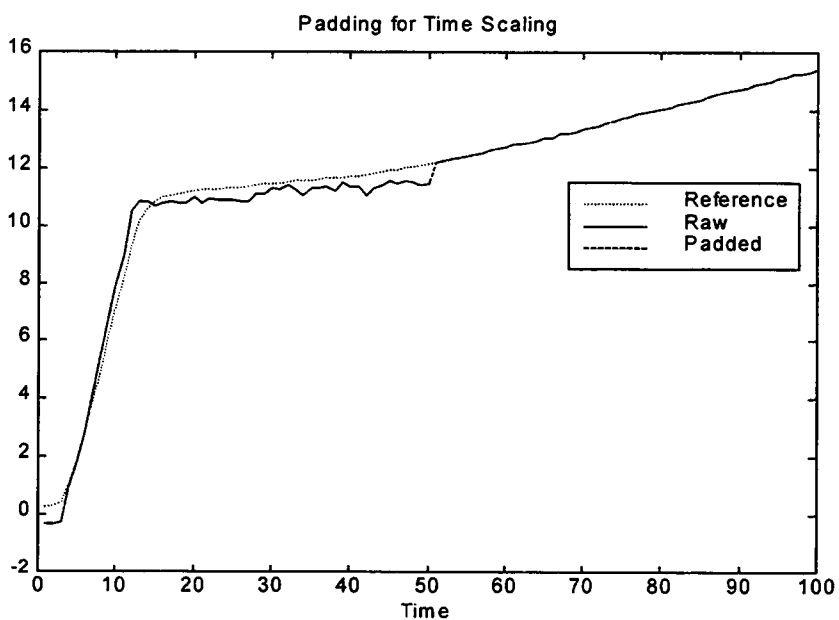
When significant variation exists in the duration of a phase, linear interpolation can introduce some spurious misalignment as can be seen in figure 4.2 (a). In such cases, points can be added or removed from the profile depending on its duration. The padding could be by the last available value or by the values in the reference profile. This is also shown in figure 4.2 (b) where the reference profile values are padded for unavailable data so that the profile and the reference are of equal length.

4.2.2 Indicator Variable

Non-linear time scaling is also possible in case an indicator variable is available [40, 47]. An indicator variable is one which monotonically increases (or decreases) with time. Its value should be consistent at the beginning and the end of the batch (or phase). An example would be the reaction conversion as the batch (or phase) progresses in time. The time of a profile can be mapped to that of the reference by finding the time in the reference closest to the indicator variable value in the present batch. The method is illustrated in figure 4.3.



(a)



(b)

FIGURE 4.2: Extreme variation in batch duration
 (a) Interpolation introduces spurious misalignment
 (a) Padding with reference value

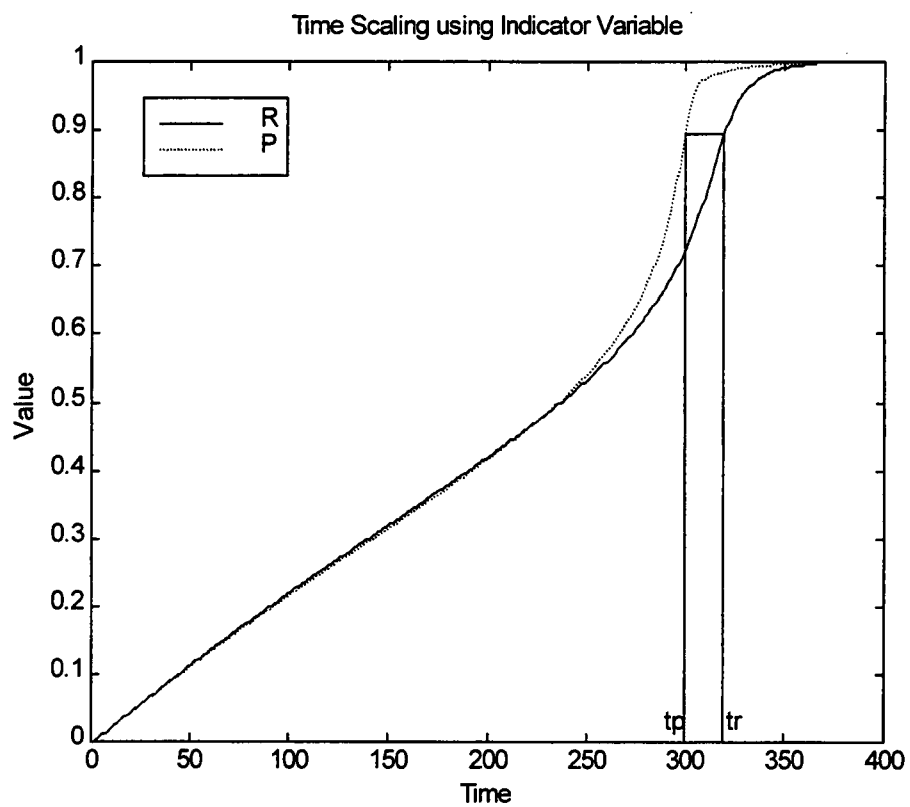


FIGURE 4.3: Time scaling using indicator variable

Typically, such indicator variables would be reactor levels where the product density is significantly different from the reactant density. For example in free radical polymerizations, when the density of the polymer formed is more (or less) than that of the monomer and the solvent, the formation of polymer leads to a decrease (or increase) in the reactor level thus giving a direct indication of the reaction conversion. Advances in analytic measurement technology such as near infrared, Raman and mass spectroscopy has made the online measurement of concentration possible in several systems of practical interest. Such measurements can also be used as indicator variables that monitor the reaction conversion. The approach is capable of local compression and expansion of the time axis of a profile. For cases where an indicator variable is not available, dynamic time warping, a speech recognition technique capable of nonlinear time scaling, may be used and is described in the next section.

4.2.3 Dynamic Time Warping

Dynamic time warping (DTW), a distance based technique from speech recognition, is a method capable of local compression and expansion of the time axis of the profile so that a time normalized distance from the reference is minimized [80-84]. DTW has recently been applied for the synchronization of batch trajectories [84]. A brief mathematical formulation follows.

4.2.3.1 Mathematical Formulation

Let $\mathbf{R}$ denote a reference trajectory and $\mathbf{P}$ denote a new trajectory where

$$\mathbf{R} = \mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_i, \dots, \mathbf{r}_I \quad \dots\dots\dots(4.1a)$$

$$\mathbf{P} = \mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_j, \dots, \mathbf{p}_J \quad \dots\dots\dots(4.1b)$$

Note that both $\mathbf{R}$ and $\mathbf{P}$ could be multivariate and that usually, $I \neq J$. DTW finds a mapping F of K points on the $I \times J$ time grid

$$F = \{c(1), c(2), \dots, c(k), \dots, c(K)\}, \max(I, J) \leq K \leq I + J \quad \dots(4.2a)$$

where $c(k)$ is an ordered pair

$$c(k) = [i(k), j(k)] \quad \dots\dots\dots(4.2b)$$

that maps time $i(k)$ on trajectory $\mathbf{R}$ to time $j(k)$ on $\mathbf{P}$. F can thus be viewed as a path on the time grid so that a normalized distance between $\mathbf{R}$ and $\mathbf{P}$ is minimized and is also known as the warp function.

For physically meaningful warping, certain constraints are imposed on the warp function. In order to ensure that events are compared in their natural time order, F should be continuous and monotonous. This is expressed as

$$i(k+1) \geq i(k)$$

$$j(k+1) \geq j(k)$$

Excessive warping is avoided by imposing upper and lower bounds on the slope of F . This is illustrated in figure 4.4 which indicates that the k^{th} path point can only be reached from a specified set of predecessors. Note from the figure that the predecessors chosen are such that two consecutive steps cannot be one along the horizontal and the other along the vertical. In other words, the slope is constrained to have a continuous derivative.

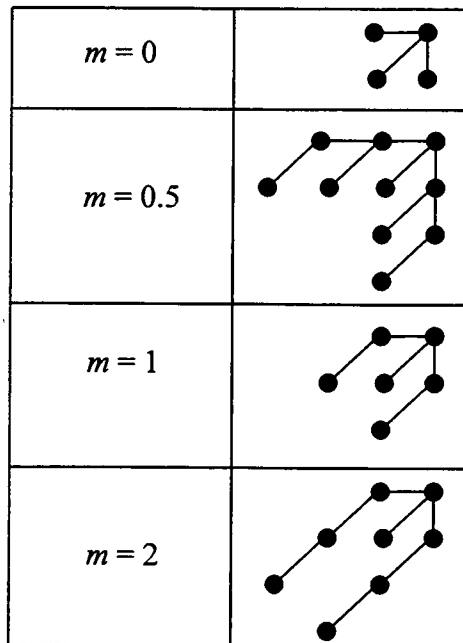


FIGURE 4.4: Schematic illustrating the slope constraint m

Another way of imposing the slope constraint is to allow a maximum number of consecutive horizontal or vertical transitions m on the path so that the slope is constrained between m and $1/m$. To avoid excessive warping, the search space is limited to within a band about the $i = j$ line on the $I \times J$ time grid. This is shown in figure 4.5.

Finally, constraints on the end point are imposed. If the end points are known with certainty,

$$c(1) = (1, 1)$$

$$c(K) = (I, J)$$

Relaxed end-point constraints which specify an allowable region for the first and last path points are also possible [82].

The various DTW algorithms can be classified as symmetric or asymmetric based on the weight given to **R** and **P**. Symmetric algorithms give equal weight to both **R** and **P** while asymmetric algorithms give more weight to **R**. Thus, symmetric algorithms give the same warp function **F** if the roles of **R** and **P** are interchanged while asymmetric algorithms may give a different warp function. Nevertheless, both can be formulated within a common framework using dynamic programming.

The time normalized distance between two trajectories is defined as [80]

$$D = \frac{\sum_{k=1}^K d(k).w(k)}{N(w)} \dots\dots\dots(4.3)$$

where $d(k)$ is a distance measure between $R(i(k))$ and $P(j(k))$ and is usually the Euclidean distance. $N(w)$ is a normalization factor that makes D independent of the

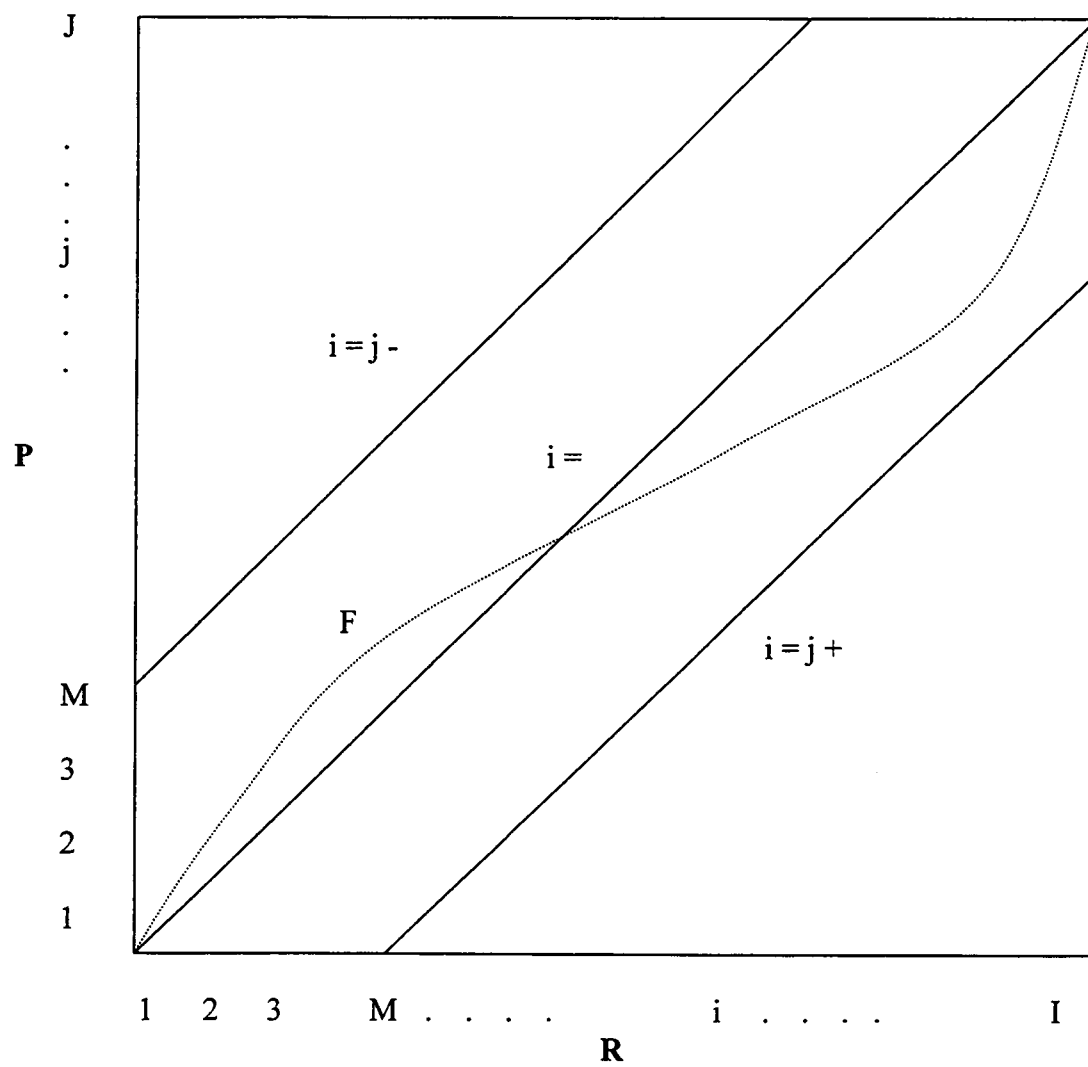


FIGURE 4.5: Global search constraint in DTW

number of points on the path and its value depends on the type of weighting function $w(k)$ used. The optimum is obtained as

$$D^*(I, J) = \min_F [D(I, J)] \quad \dots\dots\dots(4.4a)$$

and the optimum path is obtained as

$$F^* = \operatorname{argmin}_F [D(I, J)] \quad \dots\dots\dots(4.4b)$$

Since $N(w)$ is chosen to be independent of the path, the minimization problem is

$$D^*(I, J) = \frac{1}{N(w)} \min_F \sum_{k=1}^K d(k)w(k) \quad \dots\dots\dots(4.5)$$

where $N(w)$ has been taken out because of path independence. The optimization is now possible recursively by dynamic programming, which uses the principle of optimality [79]. At any point $[i, j]$ on the time grid, the accumulated distance is found using the following simple rule

$$D_A(i, j) = \min \begin{cases} D_A(i-1, j) + w_h d(i, j) \\ D_A(i-1, j-1) + w_d d(i, j) \\ D_A(i, j-1) + w_v d(i, j) \end{cases} \quad \dots\dots\dots(4.6)$$

In the above equation, the subscripted w 's indicate the weight used for horizontal, vertical and diagonal transitions on the time grid. These weights can be chosen so that the diagonal transition is favoured ($w_d < [w_h, w_v]$). The algorithm is symmetric in case $w_h = w_v$ and is unsymmetric otherwise. The optimum path is not known till the recursion is complete and the decision about the optimal predecessor must be stored at each of the grid points in the search space. The path is then backed out after the recursion is complete. A complete flowchart of the algorithm is given in figure 4.6.

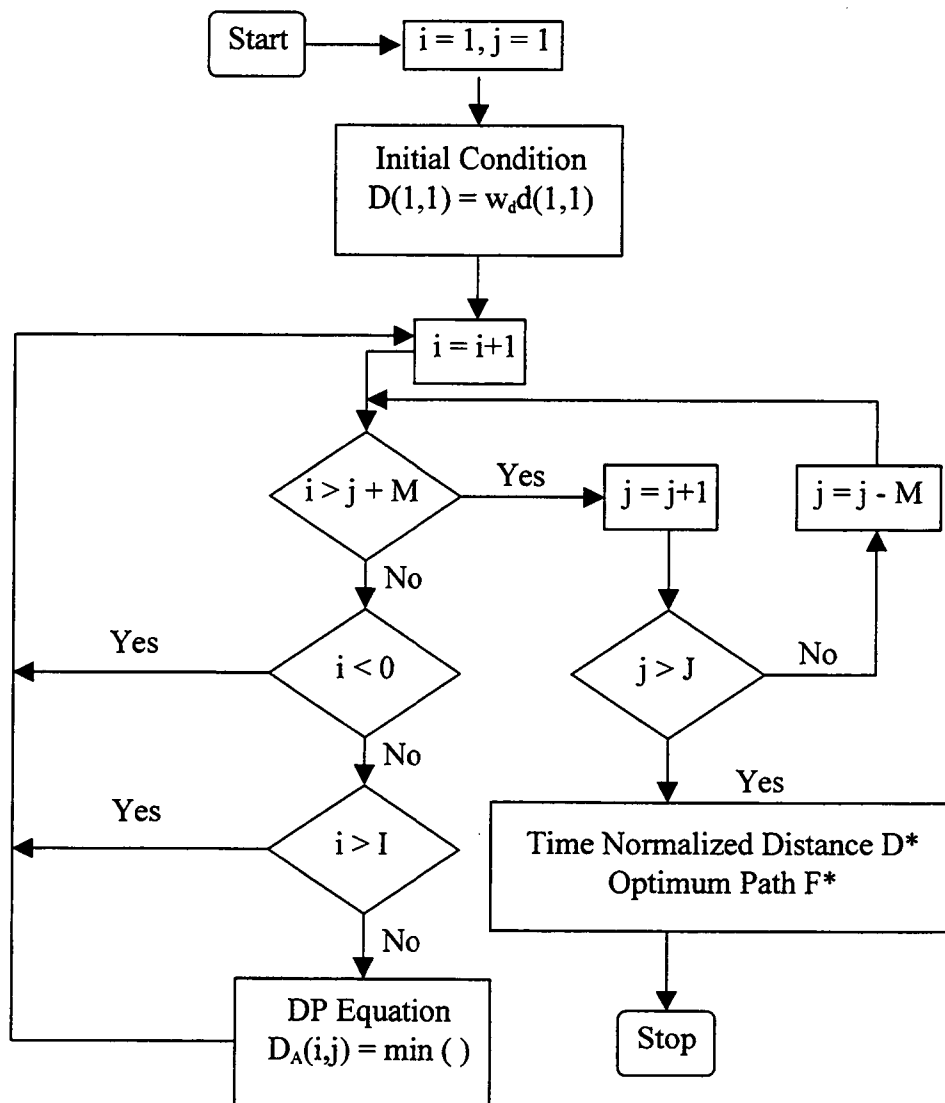


FIGURE 4.6: DTW algorithm flowchart

4.2.3.2 Example Application

The application of DTW is demonstrated in figure 4.7, which plots the raw and the warped profiles. The warping function is also plotted. The multivariate profiles are scaled between 0 and 1 so that measurement units do not affect the distance measure. As seen from the figure, the significant events in the batch - the initial ramp up in the blue and red profiles, the step down in the yellow profile along with the simultaneous step up in the purple and cyan profiles, the saturation in the yellow profile along with the step down in the cyan and purple profiles, and finally the step up in the green profile, are all aligned by DTW. The warp function is similar to the indicator variable and is obtained based on the trajectory followed by **P**. DTW is thus suitable for time alignment of distinct events in batch trajectories. However, physically unreasonable warping can also occur due to the extreme oscillations in the measurements. This aspect shall be further clarified in the discussion section later.

One way to avoid such unrealistic warping would be to use weight w that favor diagonal transitions ($w_d \gg [w_h, w_v]$). In the example presented, w_h and w_v are set to 4 while w_d is set to 1. Also, tighter limits on the slope of the warp function may be used. Another way to avoid such unreasonable warping is to apply a weighting factor to the various measurements so that the most consistent variable is given the highest weight as reported in a recent application [84]. The approach however is iterative, computationally intensive and in the example presented in the reference, gives almost complete weight to a variable that is similar to an indicator variable.

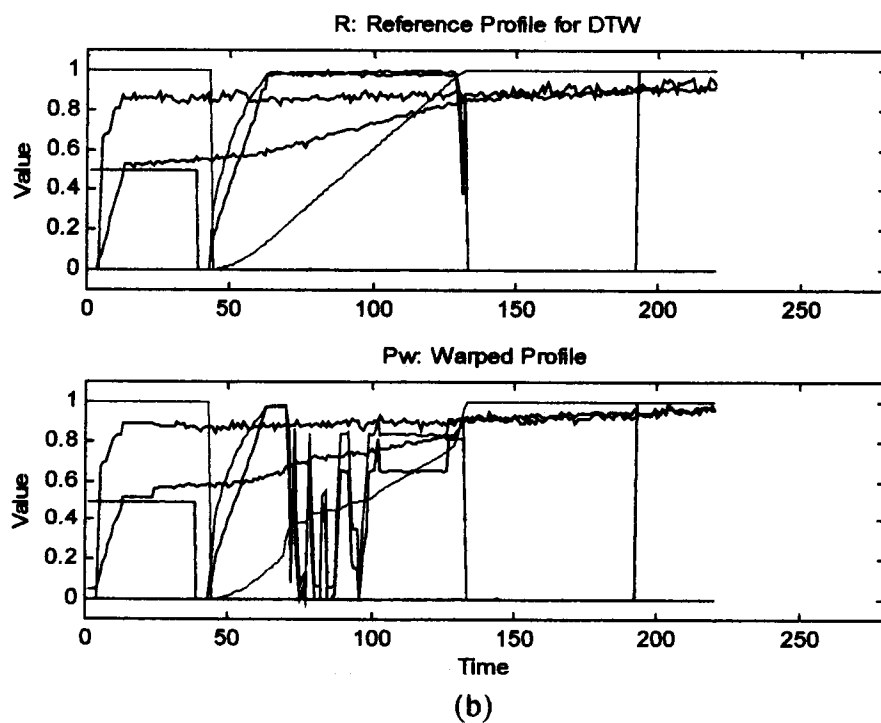
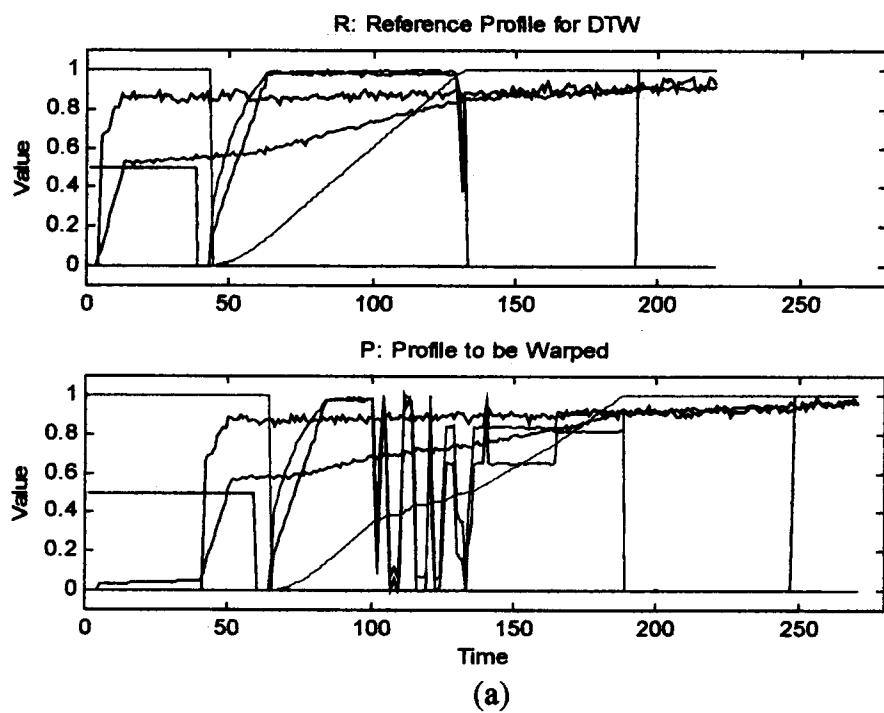


FIGURE 4.7: Application of DTW for time alignment
 (a) R and P before DTW
 (b) R and P after DTW

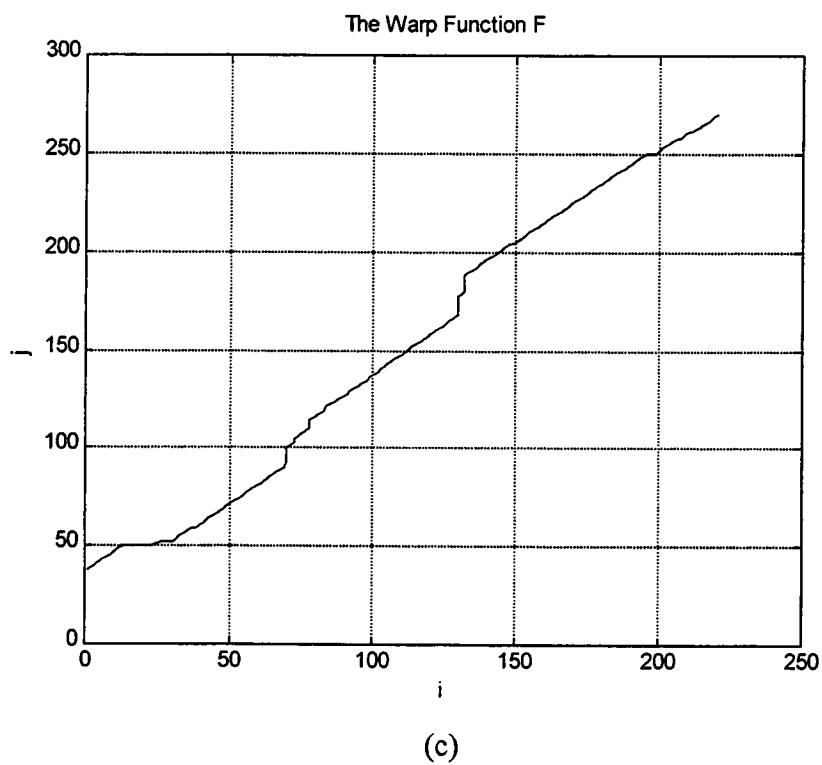


FIGURE 4.7 (continued): Application of DTW for time alignment
(c) The warp function F

Such indicator variables are usually not available in batch processes, which severely limits the application of DTW in a physically reasonable manner.

Typically, batch processes consist of various phases, which are characterized by distinct physical events such as a step change or sharp peaks and valleys etc. These can be easily identified and used for alignment. Methods like linear interpolation or padding can be used between these distinct events for time scaling. This would lead to the alignment of the various important shapes and features in the profiles in a physically reasonable manner. The idea is discussed in detail in the next section.

4.2.4 Shapes and Features

Batch profiles consist of several distinct shapes and features such as step changes, peaks, valleys, inflection points etc. These distinct features that occur consistently from batch to batch are indicative of similar underlying dynamic phenomena and can be used for time alignment. Konstantinov and Yoshida used a technique based on fuzzy logic for extracting such features and using an expert system for qualitative reasoning based on the time of occurrence of such events [77]. Derivatives on the profiles are used for matching a snapshot of the profile with a template shape. The method is applied to a batch fermentation process. However, since the first and second derivatives are used for pattern matching, the technique can only handle reasonably continuous and noise free profiles. Here a simpler distance based technique for extracting the times of occurrence of features in a profile is

developed. The technique, since it is distance based, is well suited for handling profiles containing distinct features such as step changes.

Consider a profile $\mathbf{p}$ consisting of a distinct feature vector $\mathbf{f}$ where both $\mathbf{p}$ and $\mathbf{f}$ are a row vectors. Feature extraction using a distance measure is based on the fact that when the feature $\mathbf{f}$ is translated across the profile $\mathbf{p}$, and a normalized SSE calculated for each of the times, the times at which a feature $\mathbf{f}$ occurs in $\mathbf{p}$ would show a very small SSE. Minima on the SSE curve thus indicate the times of occurrence of the feature $\mathbf{f}$ in the profile $\mathbf{p}$. The concept is similar to that of convolution of two vectors ($\mathbf{f}$ and $\mathbf{p}$) except that instead of the overlap between the vectors, the MSE is calculated.

More formally, if $\mathbf{p}$ and $\mathbf{f}$ consist of I and J points respectively, then the translation results in a vector $\mathbf{d}$ of length $K = I+J-1$ where the element $\mathbf{d}(k)$ is obtained as

$$\mathbf{d}(k) = [\mathbf{p}_k - \mathbf{f}_k] [\mathbf{p}_k - \mathbf{f}_k]^T / n_k \quad \dots\dots\dots(4.8)$$

where $\mathbf{p}_k$ and $\mathbf{f}_k$ denote the overlapping parts of the profile and the feature respectively and are obtained as

$$\mathbf{p}_k = \begin{array}{ll} \mathbf{p}(1:k) & \text{if } k < J \\ \mathbf{p}(k-J+1:k) & \text{if } k \geq J \text{ and } k \leq I \\ \mathbf{p}(k-J+1:I) & k > I \end{array} \quad \dots\dots\dots(4.9a)$$

$$\mathbf{f}_k = \begin{array}{ll} \mathbf{f}(1:k) & \text{if } k < J \\ \mathbf{f}(1:J) & \text{if } k \geq J \text{ and } k \leq I \\ \mathbf{f}(k-I+1:J) & k > I \end{array} \quad \dots\dots\dots(4.9b)$$

In equation 4.8, n_k is a normalization factor and equals the number of points used in the calculation of $\mathbf{d}(k)$ and is obtained as

$$n_k = \begin{cases} k & \text{if } k < J \\ J & \text{if } k \geq J \text{ and } k \leq I \\ J-(k-I) & \text{if } k > I \end{cases} \dots\dots\dots(4.9c)$$

n_k is thus less than J towards the beginning and the end of the translation, since the number of points of overlap between p and f is less than J . Note that in the above expressions, it is assumed that $J < I$, or the feature template vector f is shorter than the whole profile p .

The above technique gives a simple means of extracting the times of occurrence of a distinct feature in a profile. A profile may consist of more than one consistent feature. The same translation technique can then be applied to all such features. The feature times are extracted as the minima on the respective d vectors. Having obtained the feature times, techniques such as padding and linear interpolation may be used between the times so that the length of the profiles equals that of the reference.

As an example, consider the profiles used in the DTW example. The concept is demonstrated on a particular variable as plotted in figure 4.8 (a), which shows the reference r and the profile p before and after time scaling. Figure 4.8 (b) shows the profile p , the feature vector f and the d that is calculated by translation. Note the minimum in d which corresponds to the occurrence of the feature in profile p . Once the time of occurrence of the feature is obtained, time alignment with the reference r is accomplished by either adding or removing points from p or by interpolation as described earlier. In the present example, since p is longer in duration than r , additional points are removed. The time alignment is clearly evident from the figure.

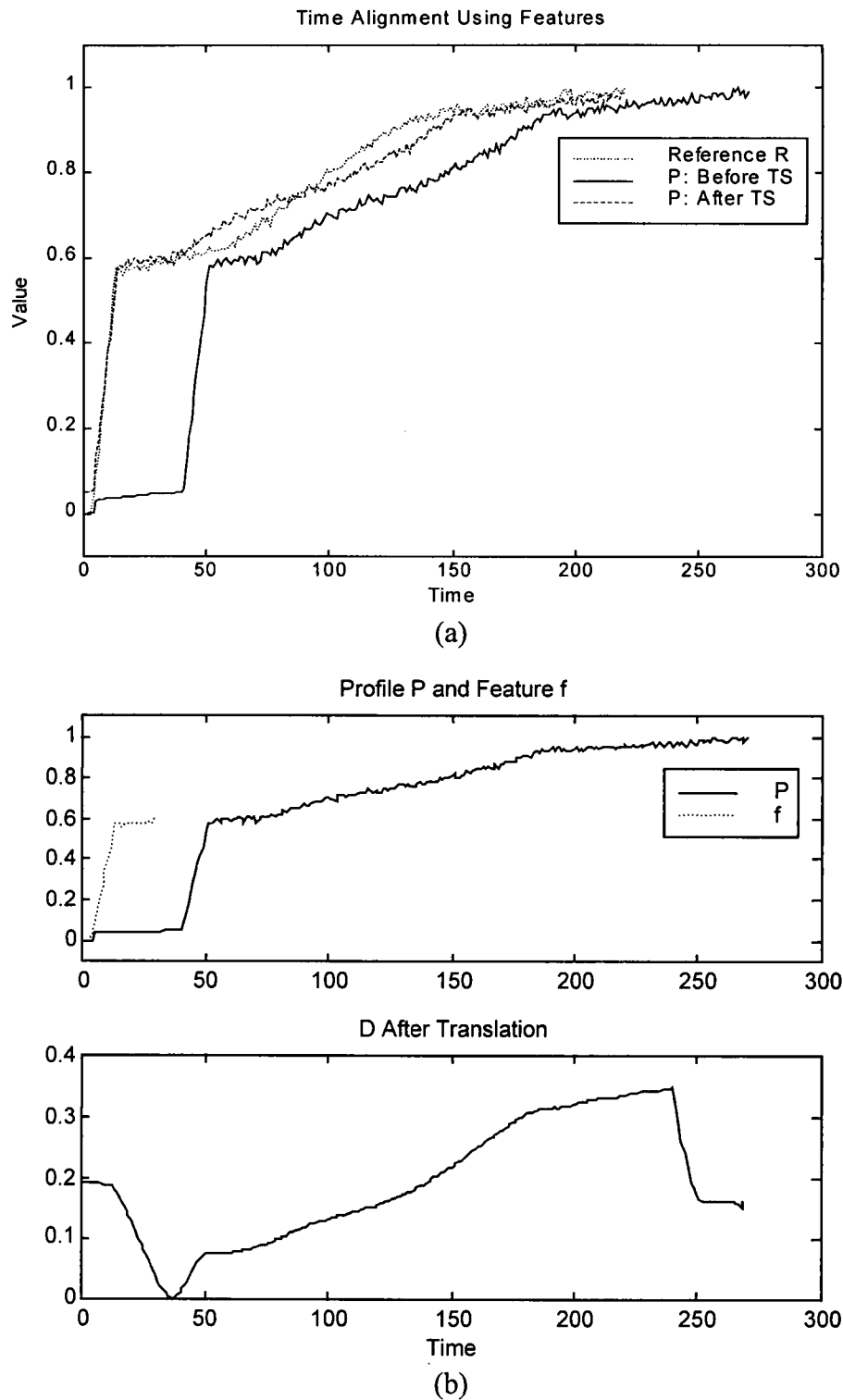


FIGURE 4.8: Example of translation for extracting feature time
 (a) Reference profile R and raw and time scaled profile P
 (b) Profile P, feature f and d resulting after feature translation

All the other profiles are treated in a similar manner and the features used for time alignment in the various profiles are plotted in figure 4.9. Extra points at the beginning and the end of the profiles are removed. Interpolation is done between features. The resulting time scaled and the raw profiles are plotted in figure 4.10. The method gives reasonable time alignment with minimal computational effort when compared with DTW. Also, the alignment seems physically reasonable.

Thus, in the absence of an indicator variable, it may be a simpler method of time scaling. It must also be pointed out that even when an indicator variable is available, its use for time scaling may not align all the significant events even though such events do occur consistently from batch to batch. Such events occur independent of the rate at which a batch progresses and the methodology then provides a direct means of extracting their times of occurrence. This is demonstrated in figure 4.11, which plots the reaction conversion and the heater power profile for a polymethyl methacrylate batch polymerization reactor simulation. Details of the simulation can be found in chapter 7. For present purposes, it is sufficient to note that the reaction consists of a significant gel effect towards the end of the batch (at high conversion) when the reaction rate increases. Due to the high reaction rates, the heating needs to be turned off and coolant fed into the cooling jacket to keep the reactor temperature at its set point. The heater switching off, thus is a consistent event that occurs from batch to batch as can be seen in the figure. However, the use of reaction conversion for time scaling does not lead to the alignment of the heater switch off time as is seen in figure 4.11 (b). These times may be important in that they could have a significant

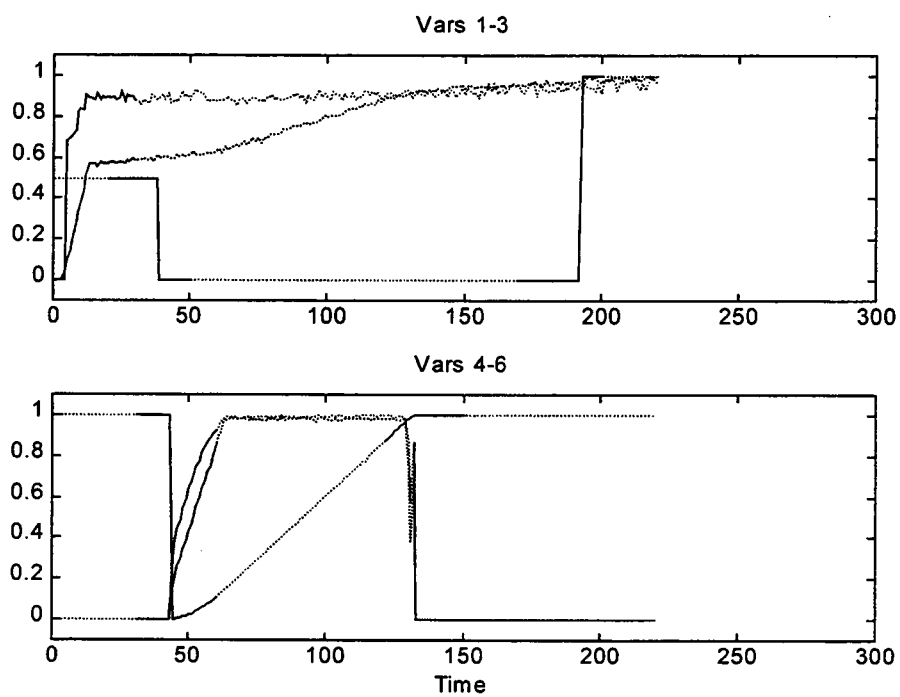
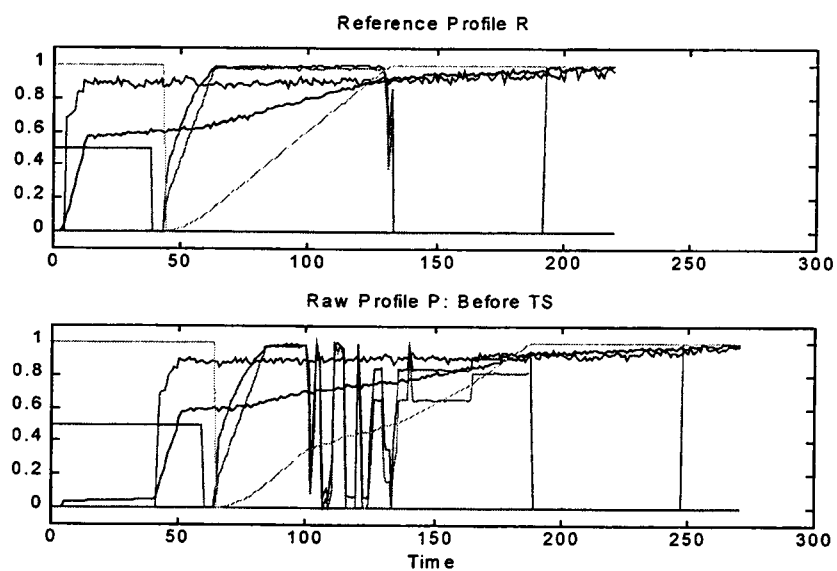
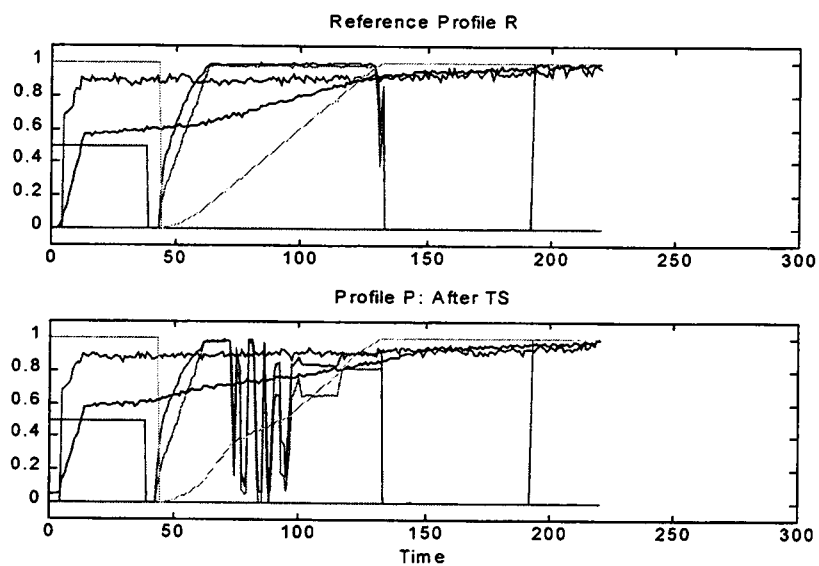


FIGURE 4.9: Feature templates used for time alignment.
Dotted lines: Complete reference profile
Solid lines: Features for time scaling



(a)

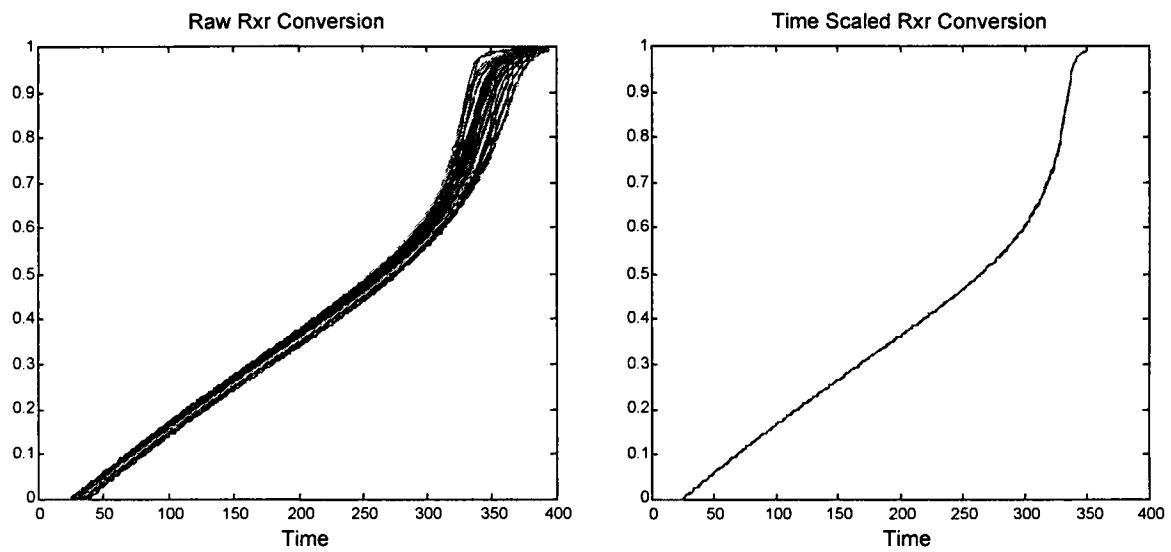


(b)

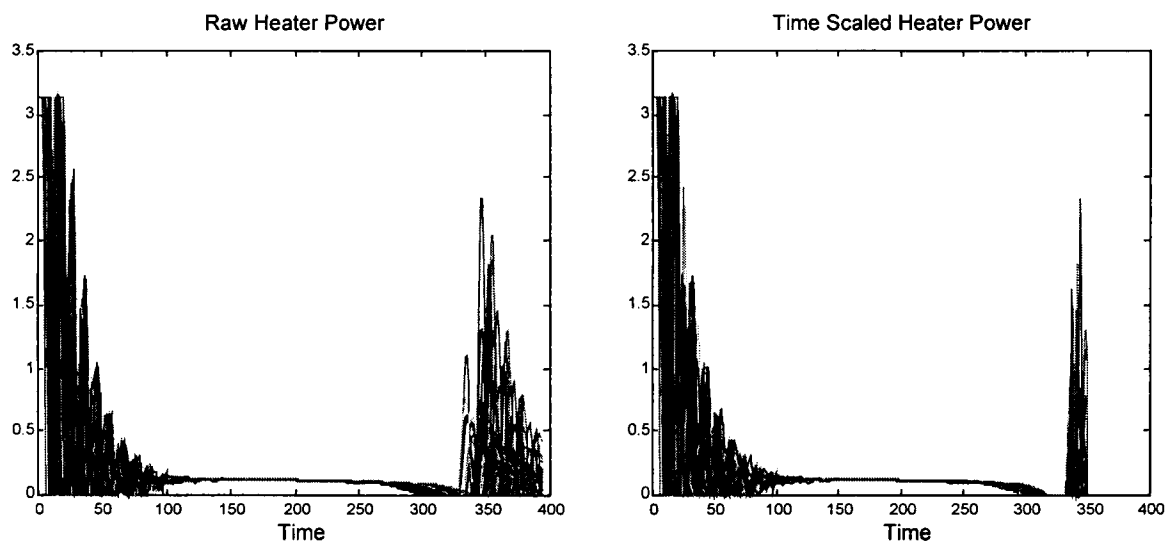
FIGURE 4.10: Time scaling using feature alignment.

(a) R and P before time scaling

(b) R and P after time scaling



(a)



(b)

Figure 4.11: Indicator variable may not align all the consistent events
 (a) Reaction Conversion
 (b) Heater Power

correlation with the final batch quality. The methodology presented above can be used to extract the time of occurrence of such a consistent event.

4.2.5 Discussion

In the last few subsections, various tools for time scaling have been described and some general comments are in order. An important issue is, no matter what tool is used, some information is lost when a batch is longer in duration than the reference. One way to avoid such information loss would be take the batch with the longest duration as the reference. In case a batch consists of several phases, a reference can be constructed such that each phase in the reference is the batch with the longest duration for that phase. Different phases in the reference may then be from different batches. This would minimize the information loss due to time scaling. For shorter duration reference profiles, the amount of information lost due to time scaling can be quantified in terms of a sum of squared error type statistic. Very high values on such a statistic would indicate excessive information loss.

Another issue is the decision on the method to use for time scaling. Unfortunately, there is no definite answer to question and judgement must be used. As a general guideline, an indicator variable should be used whenever available. Typically, if a batch consists of various phases, say reactor charging, preheating, reaction, cooling and product discharge, an indicator variable would at best be available for the reaction phase. The other phases are characterized by distinct events and feature extraction would be a good way of extracting their times of occurrence.

Once the feature times are extracted, interpolation may be used for time scaling between the features. However, in case the variation in the phase duration is high, padding or removal of data points may be needed.

As for dynamic time warping, it is a computationally intensive tool and can lead to spurious features due to excessive warping, as was shown in the example. Also, being a distance-based technique, it is sensitive to the type of scaling used to preprocess the profiles. Further, since the time scaled profiles are to be subsequently analyzed for magnitude scale variability, only noise free variables can be used to avoid physically unreasonable warps due to noise. Also, in case significant variation exists in the duration of a phase, spurious features may occur since such phases are characterized by sharp features that affect the distance measure severely so that the warping must align these features leading to misalignments elsewhere in the profile. This is illustrated in figure 4.12, which zooms into the initial part for the blue profile in figure 4.7. Note the spurious misalignment in the profile that occurs early in the batch as evident from the figure. For comparison, the time scaled profile using feature extraction is also plotted. The problem lies in limited control over the warp function because of the constraints imposed. Applicability of DTW to batch profiles is thus limited to cases where severe variability in the phase lengths does not occur. For such cases, feature extraction in combination with linear interpolation would suffice.

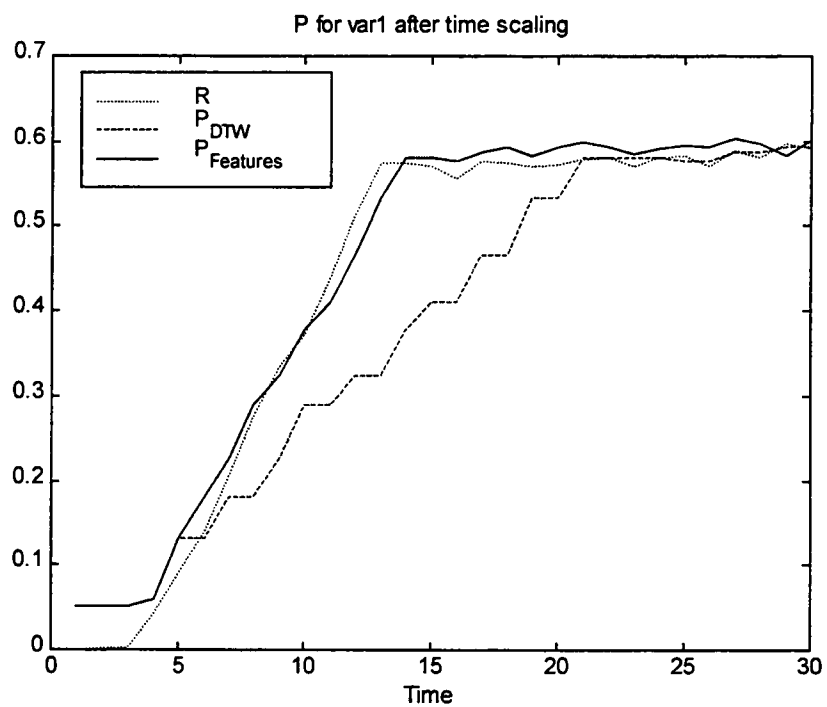


FIGURE 4.12: Spurious misalignment in DTW
 (a) R and P before and after DTW
 (b) Misalignment in initial period of variable 2.

4.3 Extraction of Feature Vectors

In the previous section, it is assumed that the feature vector $\mathbf{f}$ is known so that the translation method gives the time of occurrence of the feature vector in a profile. The extraction of suitable features $\mathbf{f}$ for time scaling forms the subject of this section. Feature vectors to be used should be extracted, primarily based on a physical understanding of the process. Batch processes usually consist of various distinct phases, eg: a batch reactor with reactor loading, preheating, catalyst addition for reaction, cooling and finally product discharge as its distinct phases. In most processes, the beginning (or end) of each of the phases is usually characterized by a distinct feature. In the reactor example, a flow valve opening to let the reactants into the vessel would indicate the beginning of the loading phase. Similarly the opening of the product valve would indicate the beginning of the product discharge phase. A large step increase in the reactor temperature set point would indicate the beginning of the preheating phase while a large decrease in the set point would imply the beginning of the cooling phase. Since each batch follows the same operating procedure, the distinct features would occur consistently from batch to batch indicating the possibility of using them for the extraction of time scale parameters. A physical understanding of the process along with a simple look at the profiles make such consistent features obvious.

As a further aid to examining the consistency of such features, translating the candidate features in a specified snapshot (time window) of each the batches would lead to a minima on the $\mathbf{d}$ vectors for the snapshot of each batch that contains the

feature. For features that occur consistently from batch to batch, such minima should be seen in the $\mathbf{d}$ vectors for all the batches. This property can be exploited to check for the consistent occurrence of a feature. It also suggests that the data itself can be analyzed for the occurrence of consistent features by taking all possible features in the profile serially and translating on a snapshot of the profiles and looking for the occurrence of minima. Consistent features can then be identified as the ones that lead to a valley shaped minima and give the least magnitude minima in the valley. Such a data driven methodology especially suited to detecting consistent sharp features in batch profiles is developed in the next subsection.

4.3.1 Evolving Factor Analysis

Consider a set of batch profiles from which consistently occurring features are to be extracted. The consistent features may be misaligned in time in the various profiles. Given a reference profile $\mathbf{r}$, the approach consists of taking a moving window of candidate features of a fixed length $J+1$ and translating the feature across a snapshot of the profiles to obtain $\mathbf{d}$ for all the profiles. The best candidates for consistent features are the ones that show the least MSE on the $\mathbf{d}$ vectors for all the profiles. A formal description follows.

4.3.2 Mathematical Formulation

Consider a set of N batch profiles

$$\mathbf{P} = \{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n, \dots, \mathbf{p}_N\}$$

of durations

$$L = \{l_1, l_2, \dots, l_n, \dots, l_N\}$$

The duration l_n of each profile p_n may be of different. A representative profile is chosen to serve as the reference r containing I points. Each candidate feature is constrained to be of length $J+1$ so that in all, there are at most $I-J$ members in the set of candidate feature vectors. Of these features, only the ones that have a large enough range (max-min) are meaningful for using as candidate features to be tested for consistent occurrence in the profiles. Suppose that we then have M such candidates where $M < I-J$. The candidate feature set is thus

$$F = \{f_1, f_2, \dots, f_m, \dots, f_M\}$$

with the beginning time index for the m^{th} feature given by i_m so that

$$f_m = r(i_m:i_m+J)$$

Each candidate feature f_m is translated across a snapshot S_m of the profiles P with

$$S_m = \{s_1, s_2, \dots, s_n, \dots, s_N\}$$

The members of S_m are the snapshots s_n of each profile p_n ($n = 1$ to N) centered around the feature beginning time index i_m so that

$$s_n = p_n(b_{ni}:e_{ni})$$

where

$$b_{ni} = \max(1, i_m-K); n = 1:N$$

$$e_{ni} = \min(i_m+K, l_n); n=1:N$$

The starting and end points of the snapshot s_n are denoted by b_{ni} and e_{ni} respectively. Translating the candidate feature f_m along the snapshot S_m , the d_{nm} vector is obtained for each batch using equation 4.8 in

$$\mathbf{D}_m = \{\mathbf{d}_{1m}, \mathbf{d}_{2m}, \dots, \mathbf{d}_{nm}, \dots, \mathbf{d}_{Nm}\}$$

Repeating this process of translation for all the candidate features ($i = 1$ to M) results in the array

$$\mathbf{D} = \{\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_m, \dots, \mathbf{D}_M\}$$

In order to ease implementation, all the $\mathbf{d}$ vectors can be pre or post padded to be of equal length. The padding may be done by -1 since negative sum of squares are not possible. Consider the m^{th} feature vector f_m , which is $J+1$ element long. Also consider a snapshot s_n that is full length, ie, it does not hit the beginning or the end of the profile. Then

$$b_{ni} = i_m - K$$

$$e_{ni} = i_m + K.$$

The d_n vector for such a snapshot would contain $2K+J+1$ elements. Also, the $J+K+1^{\text{th}}$ point in the d_n vector then corresponds to an overlap of f_m with $p_n(i_m:i_m+J)$. This correspondence can be maintained in all the snapshots that are not full length by padding with -1 . Thus in case the end of the profile is reached in the snapshot, -1 s are post padded while pre-padding is done if the beginning of the profile is reached. This allows for putting $\mathbf{D}$ in a 3 dimensional array form with easy visualization and book-keeping of the time indices. Assume that the 3D array contains 2D matrices with each matrix being composed of the $\mathbf{d}$ vectors for the N batches. There are thus M such

matrices for each of the candidate features. $\mathbf{D}$ is therefore a $[(2K+J+1) \text{ by } N \text{ by } M]$ array.

The huge $\mathbf{D}$ array must be characterized for consistently occurring valleys that indicate consistent features. Since in batch processes, it is typically the sharp features that characterize the beginning or end of a phase, a simple way is to extract the minima of the valleys in the $\mathbf{d}$ vectors around a reasonable vicinity of the $J+K+1$ point. Consistent features would then show small magnitude minima for all the profiles. The best features can then be extracted as the ones that show the least magnitude minima. Inconsistent features would not show valley like minima for all the batches.

Some times, the features in the batch profiles may be occurring consistently, in that the inherent shape is the same but some variability exists in the magnitude of features. A distinct step with a different magnitude in the steps would be an example. In such cases, even though the valley would be consistently observed from profile to profile, some residual error shall always remain owing to the magnitude scale variability. A simple way to extract this variability is to use a shape factor SF to project the overlapped portion of snapshot onto the feature in a least squares sense and then obtaining the MSE for $\mathbf{d}_n$. The shape factor can be constrained between a minimum and a maximum, to avoid meaningless scaling when the overlapped portion is not in the image of the feature.

Also since this is a distance based technique, valleys from consistently occurring but subtle features would generally be masked in the $\mathbf{D}$ array by the not-so-

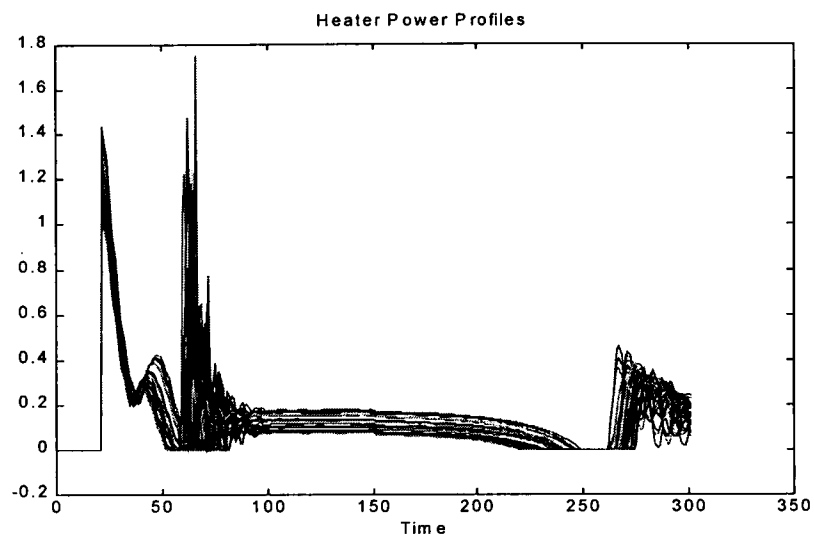
subtle features. The $\mathbf{d}_{nm}$ vectors must therefore be scaled by a suitable normalization so that the whole array can be observed on a common basis. Dividing $\mathbf{d}_{nm}$ by the range of the feature $\mathbf{f}_m$, can serve such a purpose and is used in the example study. Thus, for each candidate feature $\mathbf{f}_m$, $\mathbf{D}_m$ is normalized as

$$\mathbf{DN}_m = \mathbf{D}_m / (\max(\mathbf{f}_m) - \min(\mathbf{f}_m))$$

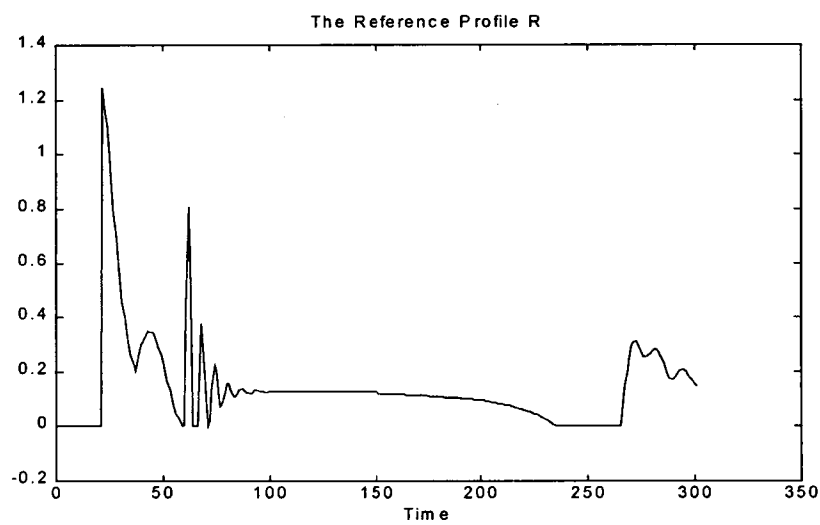
Note that in the above expression, $\mathbf{D}_m$ is a matrix the columns $\mathbf{d}_{nm}$, adequately pre and/or post padded with -1 s to make them of length $2K+J+1$. Using such a normalization, we obtain $\mathbf{DN}$ as the 3D array containing the normalized $\mathbf{d}_{nm}$. Features that show minima of a small magnitude for all the profiles can be considered as consistent. Of these, the ones that show the *least* magnitude minima are considered the 'best' features. This completes a brief formulation of the methodology. An example demonstration follows.

4.3.3 Example Application

As an example, consider a set of profiles as shown in figure 4.13(a). These profiles are for the heating rate of an electrical heater and have been generated by using the PMMA batch reactor simulation described in chapter 7. Three consistent features are obvious from the figure. The initial peak in the response at about time 20, the settling to a value of zero around time 220 and finally the sharp rise at about time 250. These features shall be referred to as features 1, 2 and 3 respectively.



(a)



(b)

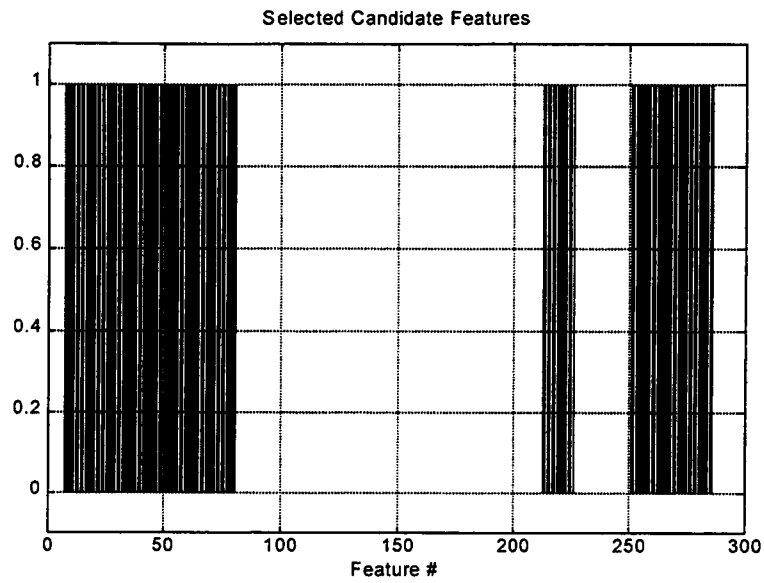
FIGURE 4.13: Heater power profiles

(a) Example profiles

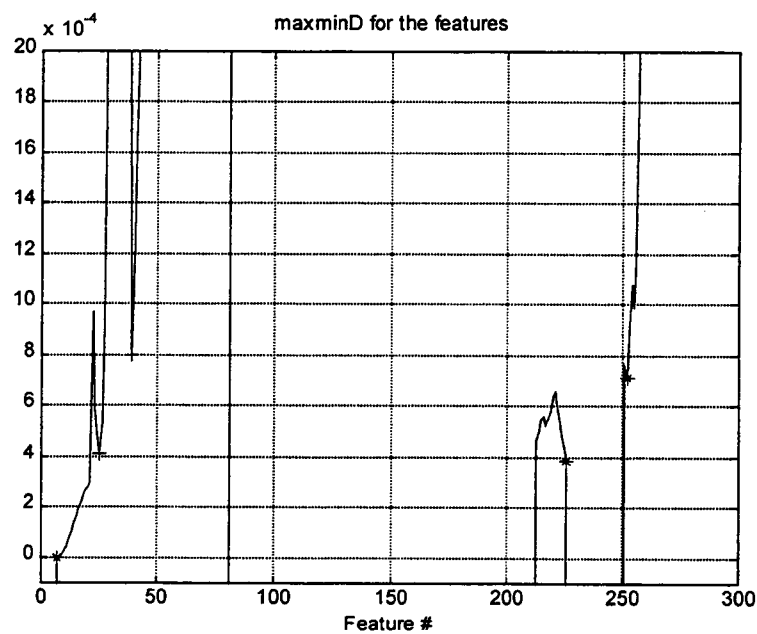
(b) Reference profile

Of the three features, feature 2 is subtle while features 1 and 3 are not so subtle. Also, some variability in the magnitudes of all the features is seen. A shape factor must therefore be used. It is constrained to lie between 0.75 and 1.5. The data set consists of 30 profiles. The first batch is used as the reference r and is plotted in figure 4.13(b). It contains 301 elements. Using a feature size of $J = 15$, in all 286 features are possible of which any feature whose range is below 0.2 is deemed unfit. The number of candidate features is thus reduced to 132. Note that physical considerations based on an understanding of the process can also be used to select the candidate features.

The $\mathbf{DN}$ array is now obtained for the 29 batch profiles using $K = 60$ for the snapshots. Its size is therefore 136 by 29 by 132. In order to extract the most consistent features, the minima along the columns within a window of 30 to 100 for the column indices in $\mathbf{DN}_m$ are obtained. The first and last few points in the normalized $\mathbf{d}_{nm}$ are not used to avoid edge effects due to incomplete overlap in the beginning and end of the translation. The maximum magnitude of the minima over all the batches for each of the candidate features is now obtained and for lack of a better word, is referred to as $\mathbf{maxminD}$. It is a vector of length 132 and is used to ascertain the consistency of a candidate feature. Figure 4.14(a) is a bar plot that shows the candidate features. A value of 1 indicates that the feature has a large enough range to be considered a candidate while 0 indicates a small range. Three time zones are evident in the plot. Figure 4.14(b) plots $\mathbf{maxminD}$ for the features. It is set to -1 for features that do not qualify as candidates.



(a)

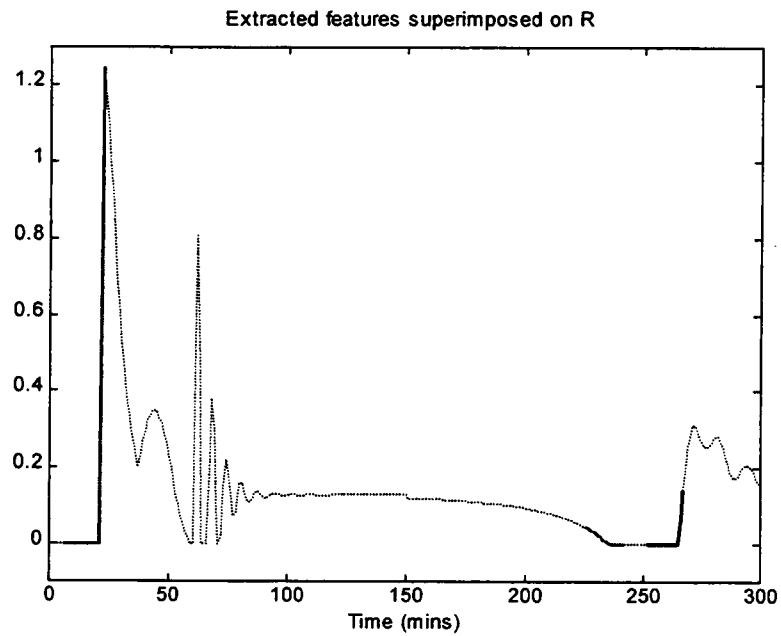


(b)

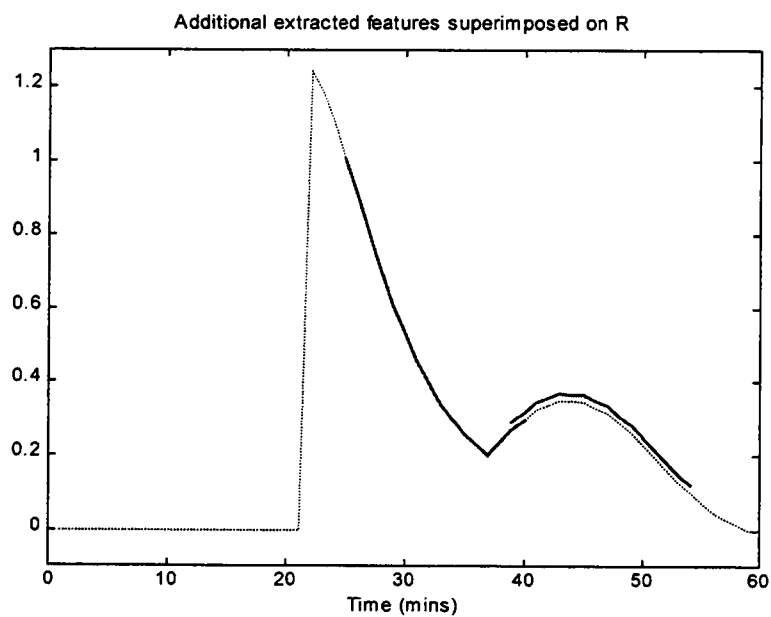
FIGURE 4.14: Selected candidate features and **maxminD**
 (a) Selected candidate features
 (b) **maxminD** for the features

Three separate zones of candidate features are seen from the figure and for present purposes, it is assumed that each zone has a single consistent feature. The minimum value for **maxminD** in a particular zone therefore qualifies as the most consistent feature in that zone. These are indicated as '*' in the plot. The candidate features so obtained for each of the zones are plotted in figure 4.15(a). Additionally, it may be pointed out that the features corresponding to points marked '+' are also of interest and are plotted in figure 4.15(b).

Figure 4.16 takes a closer look at the initial part of the profile. Another consistently occurring feature in the form of the heater power turning back up is seen. It is referred to as feature 4. The features in figure 4.15(b) correspond to this consistent event in the heater profiles. Since the two extracted features are very close to each other in time, only the first one (time 25 to 40) is used. The technique is thus able to identify the consistently occurring features in a profile. The times of occurrence of the four consistent features are now extracted by finding the minima on the respective translation $\mathbf{d}_{nm}$ for all the batches. Figure 4.17 plots the $\mathbf{d}_{nm}$ vectors for a few representative profiles for the 4 consistent features. Figure 4.18 shows example profiles along with the event times extracted by using the proposed translation with shape factors.



(a)



(b)

FIGURE 4.15: Extracted consistent features

(a) Best features in zones 1, 2 and 3

(b) Additional features in zone 1

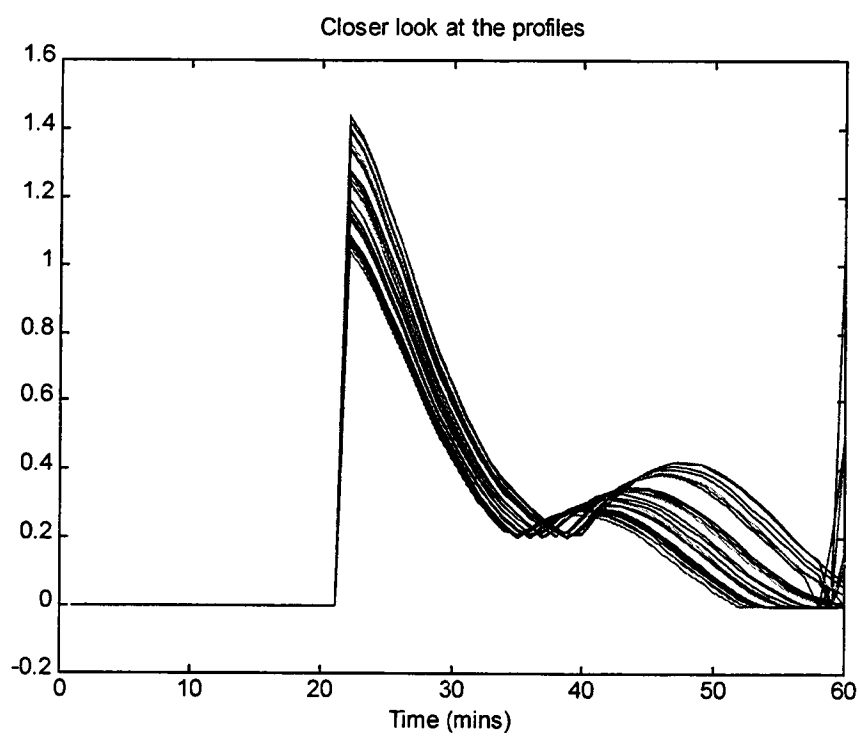


FIGURE 4.16: A closer look at the initial part of the profiles

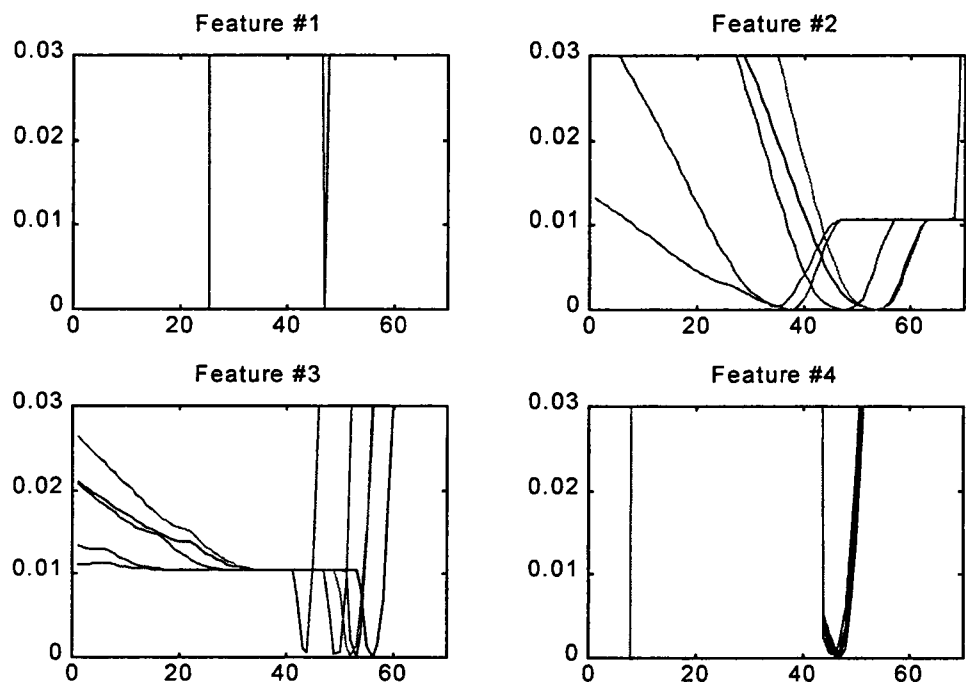


FIGURE 4.17: d vectors for a few profiles for the consistent features

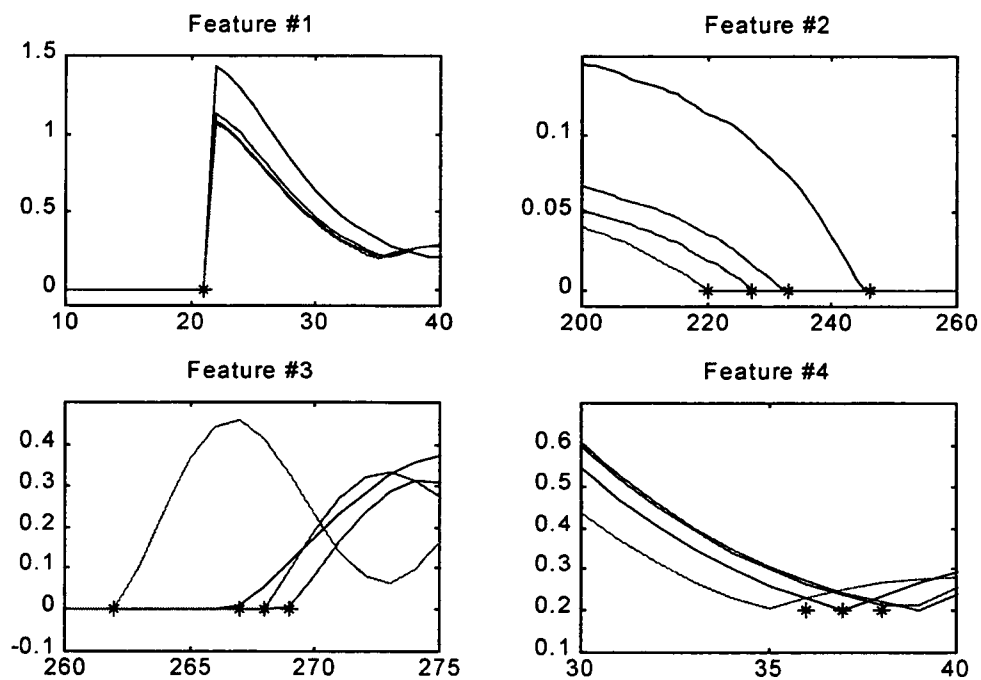


FIGURE 4.18: Extracted times for the consistent features

Note that the technique is able to identify the times for features 1, 2 and 3 accurately. These features are sharp enough so that the times extracted coincide with the discontinuity in the features. This is not the case with feature 4 as it is not so sharp. The shape factor is thus not able to adequately compensate for the magnitude scale variability. Better results may be expected in case a smaller feature size is used.

In order to emphasize the effect of using the shape factor, figure 4.19 compares the extracted times with and without using SF for feature 2, which is sharp but subtle. Note that the exact time at which the heater profile settled to a zero value is not as accurately determined when SF is not used. The reason is that when the rate of settling of the heater profile is not the same as that in the feature, the minima obtained contains in part the effect of the magnitude scale variability (different settling slopes) and in part the effect of time misalignment.

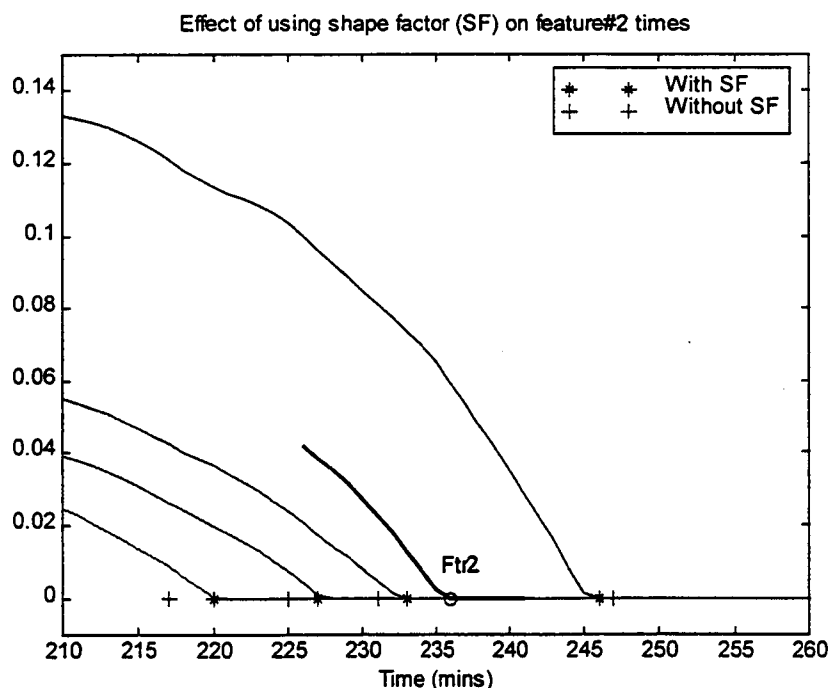


FIGURE 4.19: Effect of shape factor on feature time extraction

The use of the shape factor compensates for the magnitude scale variability and therefore the minima is primarily due to the effect of time misalignment alone, resulting in accurate event detection. The example application thus demonstrates the applicability of the simple translation technique for sharp event detection in batch profiles.

4.3.4 Discussion

The translation technique demonstrated above is based on a simple underlying concept, that on translating a sharp feature, a small SSE will result when the feature matches with a snapshot of the profile. Its simplicity makes it easy to implement and computationally less expensive than say DTW. The ability of the technique to extract consistent features that are sharp enough has clearly been demonstrated in the last section. However, when the features are not so sharp and contain significant magnitude scale variability (features 2 and 4), the accuracy of determining the time of an event is reduced.

One way to partially overcome the problem is to use a shape factor as was demonstrated to work well for feature 2. The shape factor however was not able to compensate for the magnitude scale variability in feature 4 well enough so that the event determination accuracy was not as good. The reason is that the feature itself is too long and a smaller size feature template would contain much less of the magnitude scale variability improving the accuracy to time extraction. This suggests that variable size templates should be used and the one that gives the best recognition

should be used. The idea, even though appealing at first sight, is practically impossible considering the tremendous computational load that it would entail to make it solely data driven. Specifically, the implementation would involve an additional step that involves varying the feature length so that now, a 4D DN array would be obtained with the 4th dimension being the variable feature length. Usually, a profile consists of only a few features that are obvious to naked eye. It is therefore felt that rather than computing for all the possible candidates, only physically meaningful features be used as the candidates. For such features, the best length feature can easily be obtained by varying the feature length between a maximum and a minimum limit. This would ease the computational burden a lot and also do away with any reasoning that needs to be applied in order to identify the 'best' features since that reasoning is implied in the initial candidates that are used.

Lastly, a few words on the scaling used before calculating the MSE in $\mathbf{d}$ are in order. In the example presented, a least squares projection was used to filter out simple magnitude scale variability effects in order to improve the accuracy of event time extraction. Another type of variability that occurs frequently in batch processes is of the 'dilation' kind. Thus even though an event occurs consistently, it takes place at a variable rate. As an example, consider a peak that may be dilated or constricted in the profiles. For accurate time extraction, a dilation factor must therefore be used. This leads to the idea of 'translation + dilation' which is similar to wavelets. In case variability of the dilation type occurs in a physically meaningful feature, the use of a dilation factor before calculating MSE in order to remove the effect of dilation on the

MSE would enhance the accuracy of the feature time extraction. Indeed, it is anticipated that the use of a dilation factor coupled with a smaller size template for feature 4 in the demonstration example of the last subsection would lead to much better accuracy in the extraction of the feature times.

This concludes a comprehensive treatment of the tools to be used for time scaling. The various issues have also been discussed. The next chapter focuses on the tools for magnitude scaling.

Chapter 5

Magnitude Scaling Tools

5.1 Introduction

This chapter briefly describes the tools that are used in order to extract consistent magnitude scale parameters. As discussed previously, after time scaling the batch profiles, variability along the measurement axis about the reference batch trajectory that consistently occurs in a particular way can be characterized using magnitude scale parameters. A simple example would be a flat profile that settles to a different value from batch to batch. In this case, the reference is a flat profile and the magnitude scale parameter is the shift above (or below) the reference that would bring the profile closest to the reference. PCA remains the primary tool for extracting such consistent sources of variability. The chapter is divided into two sections. The first section critically evaluates the application of PCA to batch profiles. In the second section, an evolving factor type approach based on PCA is developed that is ideally suited for extraction of variability that occurs in a particular way in batch profiles. The approach is applied to example batch profiles and its utility in extracting physically meaningful scale parameters is demonstrated.

5.2 Principal Component Analysis (PCA)

The purpose of this section is to show the relevance of Principal Component Analysis (PCA) as applied to batch profiles and underscoring the need for modifications to handle batch profiles. Throughout the section, example profiles are

used to illustrate the various issues that are raised. PCA remains the backbone of exploratory data analysis techniques and has already been described in reasonable detail earlier and the reader is referred to chapter 2 for the mathematical details. For present purposes, it is sufficient to note that PCA of a matrix $\mathbf{X}$, results in principal component loadings $\mathbf{V}$ such that in the rotated space with columns of $\mathbf{V}$ as the basis vectors, the sample data points (rows of $\mathbf{X}$) appear decorrelated. The principal components (PCs) are obtained so that the variance explained by the successive PCs is in descending order. The first PC is thus in the direction of maximum variance in $\mathbf{X}$, the second PC in the direction of maximum variance of the residuals remaining after projecting $\mathbf{X}$ onto the first PC and so forth. In the simple 2 D case, where each sample is only 2 element long ($\mathbf{X}$ has 2 columns), if the $\mathbf{X}$ matrix is mean centered and the data appears as an ellipse about the origin, the first PC is in the direction of the major axis of the ellipse while the second PC is in the direction of the minor axis of the ellipse. Extension to higher dimensions can now be easily understood.

More formally, as in equation 2.6, the PCA decomposition of $\mathbf{X}$ is

$$\mathbf{X} = \mathbf{T}\mathbf{V}^T + \mathbf{E} \quad \text{.....(5.1)}$$

In case high correlation exists between the PCs, the first few PCs explain most of the significant variation in the $\mathbf{X}$ matrix so that the higher PCs can be rejected as noise or 'insignificant'. This leads to the residual matrix $\mathbf{E}$, which contains the 'insignificant' variation in the higher PCs. Note that if all the PCs are retained, $\mathbf{E} = \mathbf{0}$. PCA is thus an effective technique for data compression of highly ill-conditioned matrices with large redundancy. In such cases, only the first few significant PCs need only be

retained.

It must be noted that PCA is a scaling dependent technique and results that use different ways to scale the columns in $\mathbf{X}$ may be very different and must be interpreted in accordance with the initial matrix scaling used. Typically, in literature, two methods for scaling the $\mathbf{X}$ matrix are prevalent. The first method mean centers the columns of $\mathbf{X}$ so that the scaled matrix has 0 mean. Such a scaling is mostly used when the units on each of the columns in $\mathbf{X}$ are the same. A simple example would be calibration spectra from say a near infrared spectroscope. In other cases, typically when the units on the various measurements are different, the scaling is done so that the columns of the scaled matrix have zero mean and unit variance. This is known as autoscaling and gives a convenient way to put all the measurements (columns of $\mathbf{X}$) on a common unit variance basis.

In batch profiles, the rows of the $\mathbf{X}$ matrix are the profiles from the historical database. Each column represents a time point in the batch. High correlation is expected between the various columns of $\mathbf{X}$ because of autocorrelation in time. Correlation is also expected between the trajectories from batch to batch due to the recipe based operation of batch processes. PCA should thus serve as an effective tool in compressing most of the measurement axis variability in the profiles into a few numbers.

Any new profile $\mathbf{x}_i$ can be projected on the retained PCs to obtain the scores $\mathbf{t}$ and the residuals $\mathbf{e}_i$ as follows from equation 3.2

$$\mathbf{t}_i = \mathbf{x}_i \mathbf{V} \dots\dots\dots(5.2)$$

$$\mathbf{e}_i = \mathbf{x}_i - \mathbf{x}_i \mathbf{V} \mathbf{V}^T \quad \dots\dots\dots(5.3)$$

The scores on the retained PCs then form the scale parameters.

As an example application of PCA on batch profiles, consider again the heater power profiles generated from the PMMA batch reactor simulation. The simulation details can be found in chapter 7. The profiles are plotted in figure 5.1. The data set contains 30 batch profiles that have been time scaled. Zones containing a particular type of variability are obvious from the figure. The first zone contains the initial smooth oscillations from time 20 to about 60. This is followed by oscillations of large amplitude from time 60 to about 100 (zone2) after which the profiles settle to a value and subsequently reduce to 0 at about time 384 (zone 3). The heater is off from 385 to 415 (zone 4). Finally the heater again comes on at time 416 and after a few oscillations, settles to a constant value (zone 5). PCA is performed on the $\mathbf{X}$ matrix using the two scaling techniques – mean-centering and autoscaling. $\mathbf{X}_m$ refers to the mean centered matrix while $\mathbf{X}_s$ refers to the autoscaled matrix.

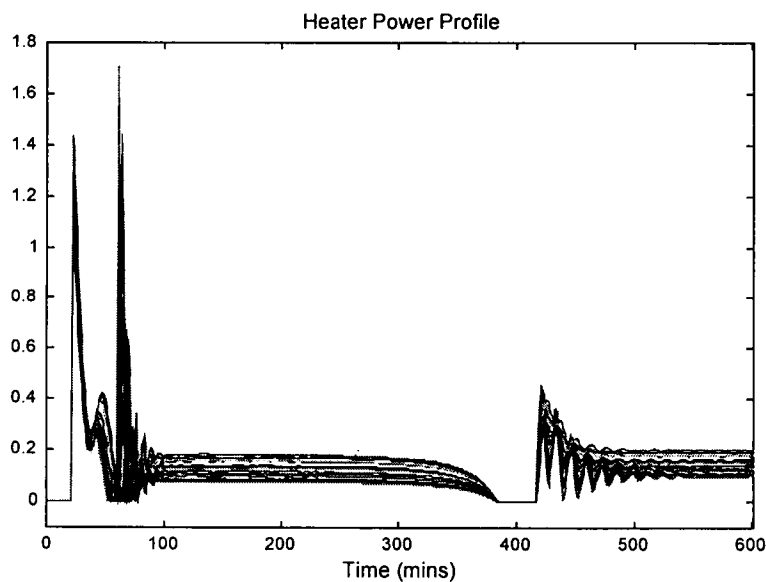


FIGURE 5.1: Heater power profiles after time scaling

The mean centered matrix X_m is studied first. Figure 5.2 contains the scree plot that plots the cumulative explained variance with respect to the number of PCs. The fact that only 2 PCs are needed to explain more than 80% of the variance in the profiles indicates the high degree of redundancy in the profiles. The two PC loadings are plotted in figure 5.3(a). Note that all the PCs have high loadings in zone 2 that contains oscillations of high amplitude. These oscillations are not systematic but since PCA is a variance based technique, high loadings are given to this zone as a significant amount of variance in the profiles exists in here.

Zones 1, 2 and 5 contain systematic variation in the form of smooth oscillations and profiles settling to a value. The PCs also have loadings in these zones so that after projection, the residuals remaining should be almost zeros. These residuals (E) are plotted in figure 5.3(b). The mean centered profiles (X_m) are also plotted in the figure. Indeed, all of the systematic variability is explained in the two PCs. However, the interpretation of the three retained PCs is not a trivial task since very high weight is given to zone 2 in all the PCs which somewhat obscures the systematic variability in zones 1, 3 and 5.

This disadvantage can be partially overcome by autoscaling, which gives high weight to time points that have low variance so that the division in the autoscaling is by a small standard deviation. In other words, the time zones that are the most consistent, ie, low variance are given greater weightage.

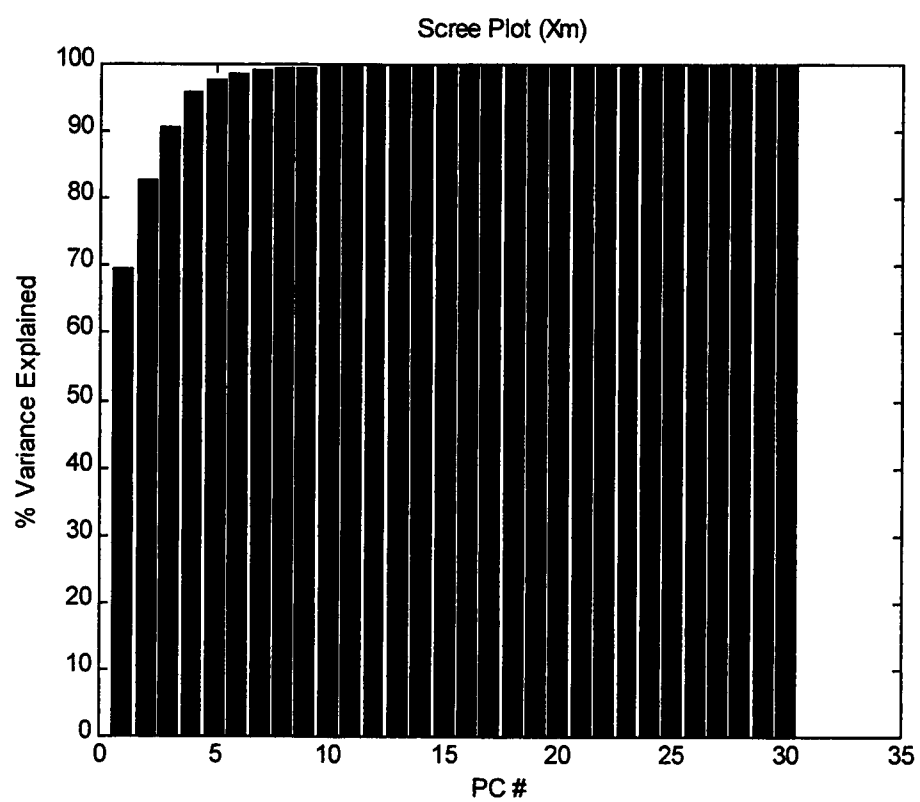
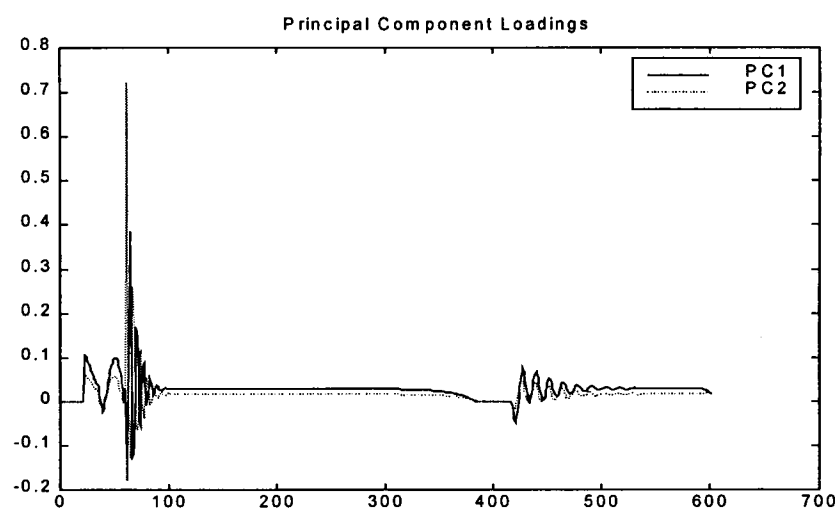
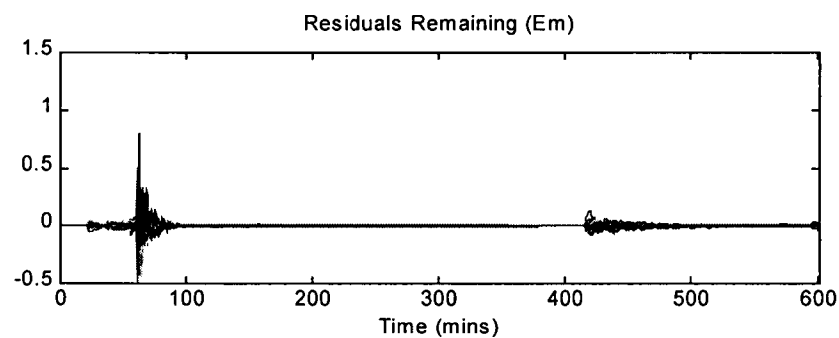
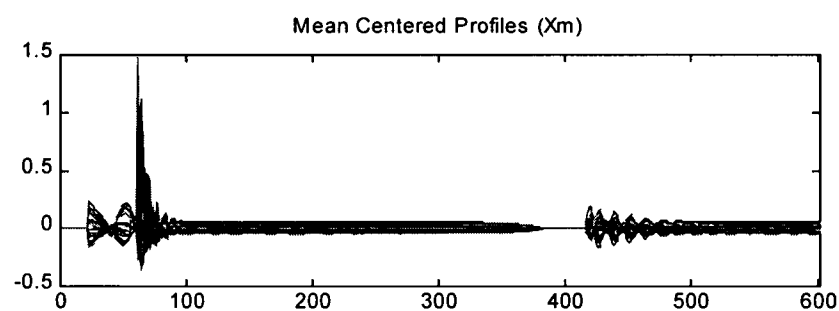


FIGURE 5.2: Scree plot for PCA of mean centered profiles



(a)



(a)

FIGURE 5.3: PCA on mean centered heater power profiles
 (a) Retained PC loadings
 (b) Residuals remaining after PCA

Figure 5.4 plots the standard deviation of the profiles. Zone 2 has a high standard deviation so that it is given the least weight while the other zones that contain the systematic variability are given higher weight in the matrix preprocessing. The scree plot for the autoscaled profiles (X_s) is plotted in figure 5.5. It is seen that a single PC explains more than 80% of the variability in the autoscaled profiles again indicating the high degree of redundancy in the data.

A single PC is retained and the PC loading is plotted in figure 5.6(a). Note that the PC loadings give high weight to the systematic variability zones. The residuals (E) after projection are plotted in figure 5.6(b). The autoscaled profiles (X_s) are also plotted. Note the correspondence between the PC loadings and the zones in which the residuals are very small.

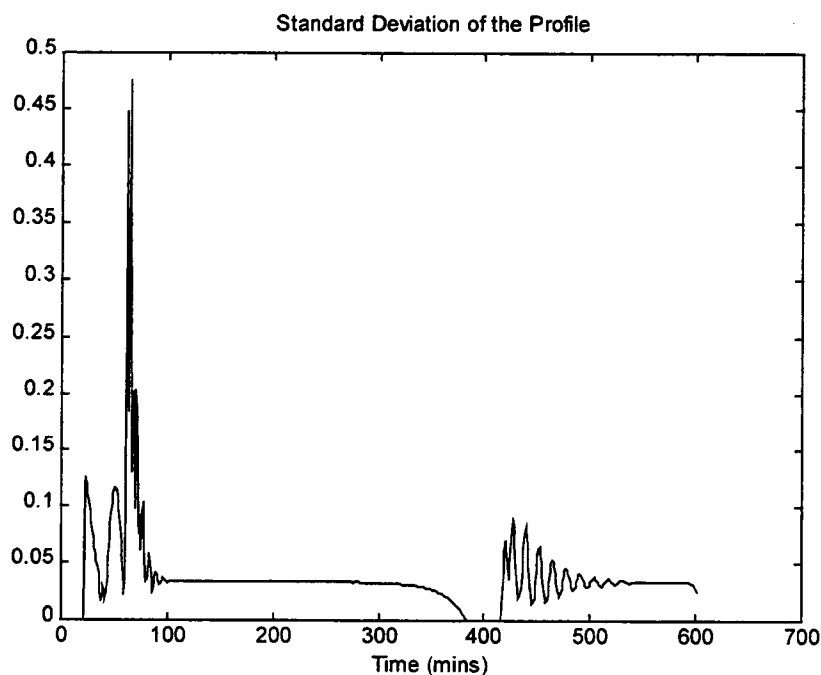


FIGURE 5.4: Standard Deviation of the heater power profile across time

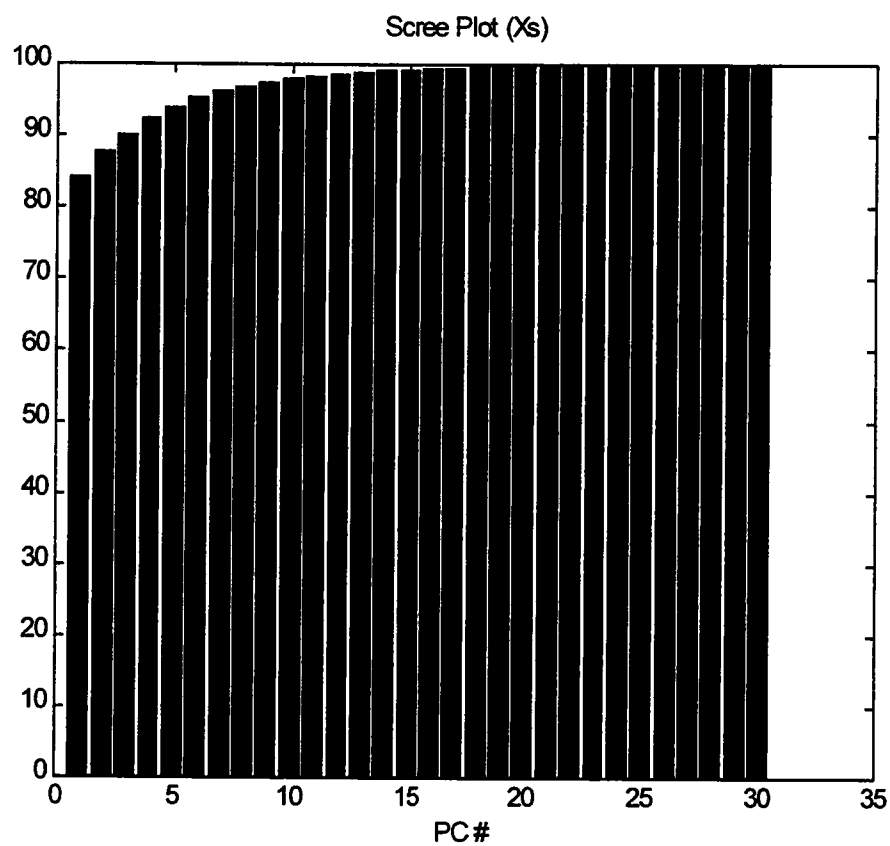
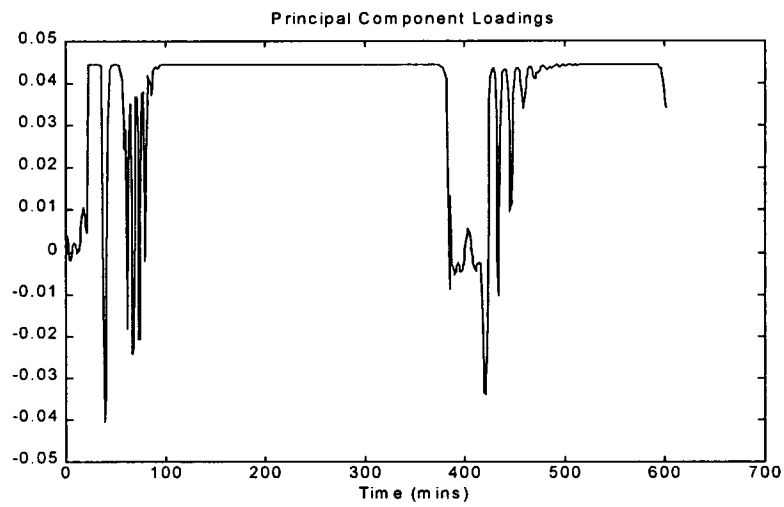
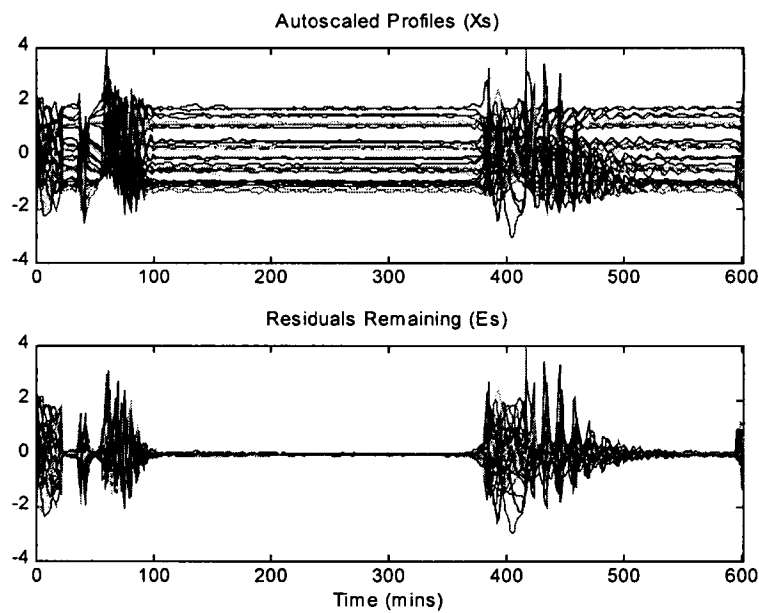


FIGURE 5.5: Scree plot for PCA on autoscaled heater power profiles



(a)



(b)

FIGURE 5.6: PCA on autoscaled heater power profiles
 (a) Retained PC loading
 (b) Residuals remaining after projection

The approach, even though appealing at first sight has its own disadvantages. Firstly, the interpretation of the PCs is obscured by the autoscaling, which involves division by the column standard deviation. The profiles themselves are thus hypothetical. Secondly, in case there are zones that contain systematic variability that is also high variance along with zones that are of low variance, the high variance systematic variability is given lower weight so that it may not be adequately explained in the first few retained PCs.

This is illustrated again by an example. Consider the profiles plotted in figure 5.7. These are profiles for the reactor temperature for the PMMA batch reactor simulation. The variability in the temperature profile exists initially from time 1 to about 40. This zone corresponds to the initial loading and heating of the reactor so that a single PC should suffice for the mean centered profiles in $\mathbf{X}_m$.

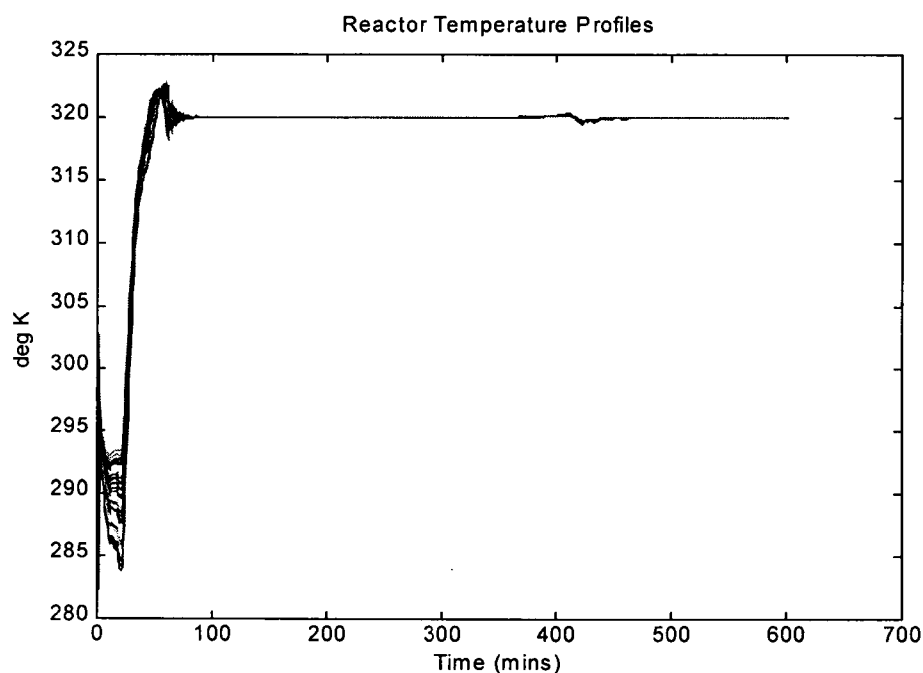
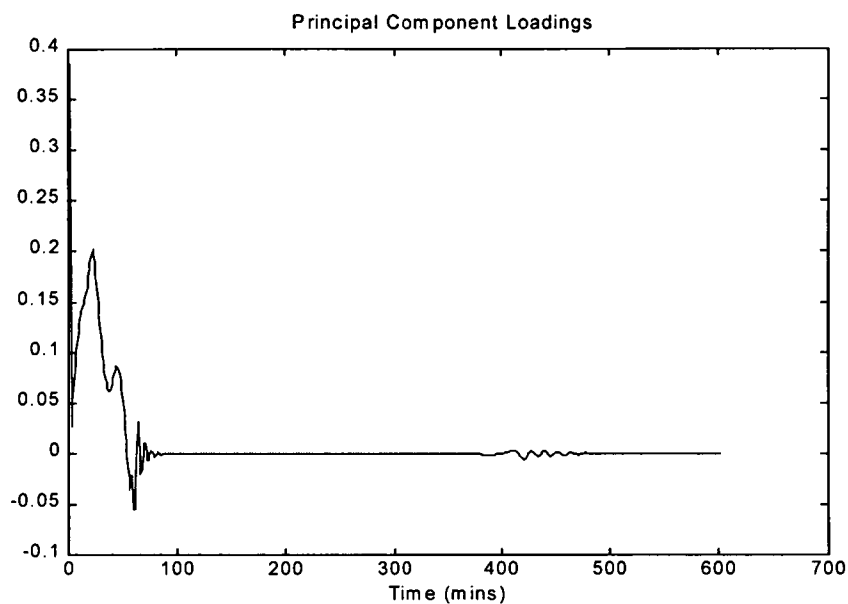


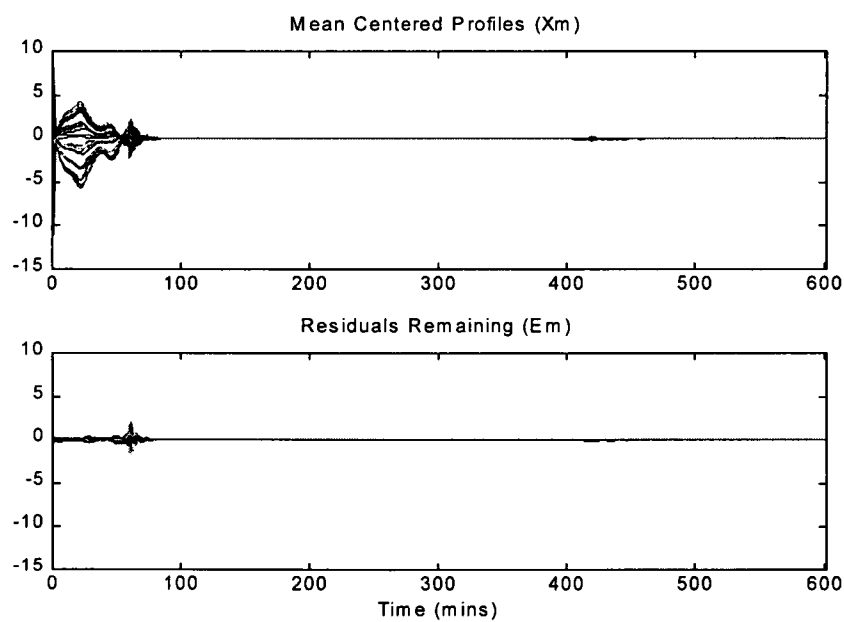
FIGURE 5.7: Reactor temperature profiles

As before PCA is applied to the mean centered profiles. A single PC explains more than 90% of the variability in the profiles. The PC loading is plotted in figure 5.8(a) while the projected profiles are plotted in 5.8(b). The mean centered profiles are also plotted in figure 5.8(b). Note that almost all of the variability in the initial start up phase is explained by the first PC.

The autoscaling case (X_s) is studied next. Due to the matrix preprocessing, the unsystematic part of the in the flat part of the profile is given very high weight so that autoscaling makes the data somewhat spherical (better conditioned) and the first four PCs explain <50% of the variability in the profiles. A single PC is retained and the PC loading is plotted in figure 5.9(a). The projected profiles are plotted in figure 5.9(b). The autoscaled profiles are also plotted in the figure. Note that autoscaling obscures the variability in the initial preheat phase. Thus even though the variation in the startup phase is systematic, a significant amount remains unexplained. Even with three PCs retained, very little increase in the amount of systematic variability in the preheat zone is observed (data not shown). This is because the autoscaling tends to give high weight to the flat zones with negligible variability so that the first PC is able to explain only a part of the variability in the preheat zone. A significant portion of the variability is thus washed down to higher PCs as a result of the autoscaling. This is not desirable.

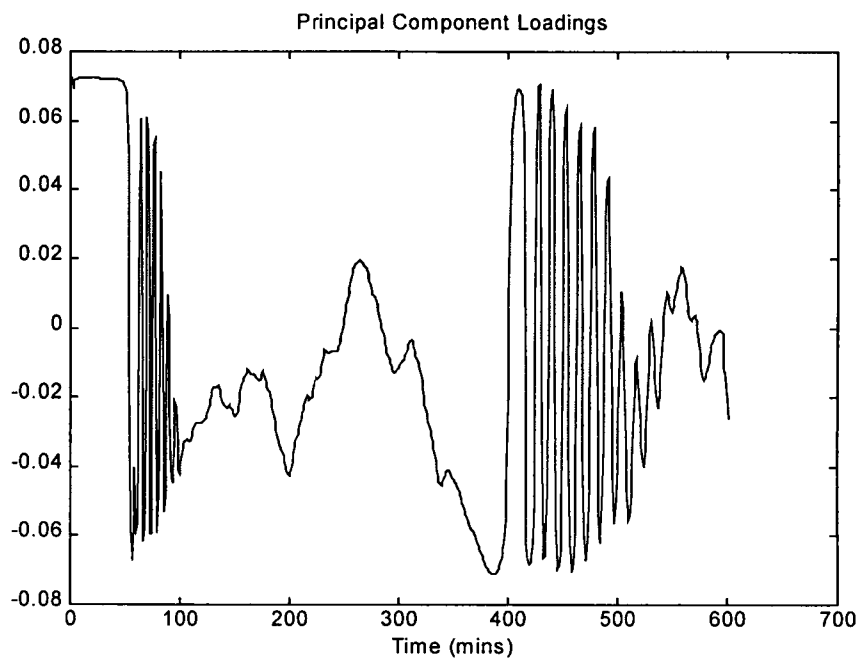


(a)

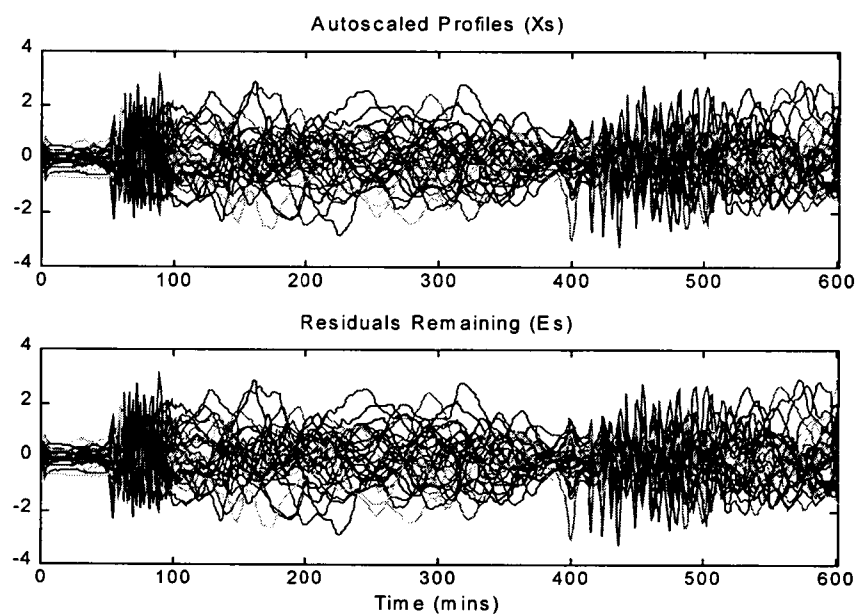


(b)

FIGURE 5.8: PCA on mean centered reactor temperature profiles
 (a) Retained PC loadings
 (b) Residuals remaining after projection



(a)



(b)

FIGURE 5.9: PCA on autoscaled reactor temperature profiles
 (a) Retained PC loadings
 (b) Residuals remaining after projection

Finally the residuals are multiplied by the profile standard deviation so that a comparison can be made with the residuals for the mean centering case. Figure 5.10 plots these residuals. Note that PCA using autoscaling is unable to explain the systematic variability in the startup period. Our experience has shown that industrial examples of such batch profiles are quite common.

From the above few examples, it is obvious that PCA by itself is not a very good way for analyzing the variability in the profiles. The basic problem lies in the fact that in batch profiles, the correlation structure of the profiles changes from phase to phase so that a single scaling technique (mean centering or autoscaling) for all the phases is unable to extract the systematic variability in all the phases.

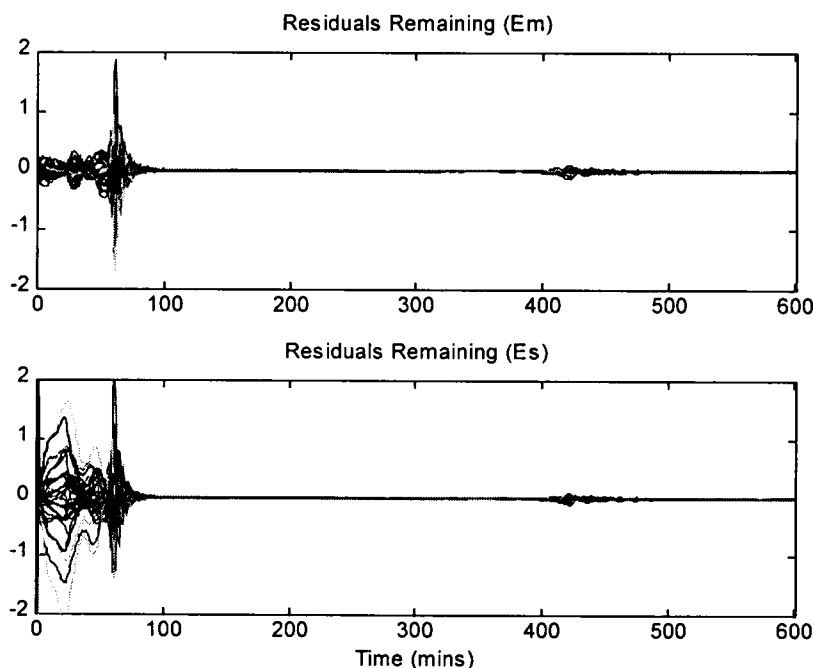


FIGURE 5.10: Comparison between mean centered and autoscaled profile residuals

In the next section, the division of profiles into zones for extracting the systematic variability in a particular zone to give easily interpretable factors is demonstrated. A simple methodology based on an evolving factor analysis type approach to extract such systematic variability in batch profiles is also developed.

5.3 Extraction of Meaningful Magnitude Scale Parameters

Consistent variability along the measurement axis in batch profiles typically occurs 'in a particular way'. A very simple example has been discussed earlier in chapter 3. In the example, variability exists in the level of a batch reactor due to initial charging variation. The level profiles in the various batch profiles are thus shifted above or below the reference by a constant amount. Variability thus occurs in a particular way, ie a shift, so that a parameter can be proposed to quantify the effect. In most batch profiles, variability occurs in a particular way so that 'custom' scale parameters can be proposed for the profiles.

Having said this, it is necessary to point out that a batch consists of various phases and the underlying physical phenomena occurring between the various phases are very different. We thus have as an example, a reactor loading phase, a preheating phase, a reaction phase, a product-curing phase and finally a product discharge phase. It is therefore reasonable to expect that since the underlying physical phenomena are different from phase to phase, each phase should have its own characteristic measurement axis variability. Extending the reactor level example further, it is expected that in the loading phase, the variability would mostly be in the form of

varying slopes (rates of addition) of the level profile. The shift parameter is therefore not well suited for the loading phase. Thus, variability is expected to occur ‘in a particular way’ only in zones or phases of the profile so that each such zone should have its own scale parameter. In fact, this is the problem that was faced in the heater profile PCA example earlier – it takes more than one PC to explain the variability in a particular zone (zone 3) since each PC also contains information from other zones. The PCs are thus difficult to interpret and make little physical sense even though maximum amount of variance is explained.

A more direct approach is to divide the profile into zones that have variability of a particular kind and use the PCs for that zone. This would lead to factors that make physical sense and are therefore easy to interpret. The word ‘factors’ has been deliberately used since PCs imply orthogonality of the scores. The fact that the whole profile has been divided into zones implies that the scores from zone to zone may be correlated and therefore non-orthogonal. This is the price that one has to pay in order to enhance the interpretability of the factors. The above points are illustrated below by means of an example which uses ‘custom’ scale parameters.

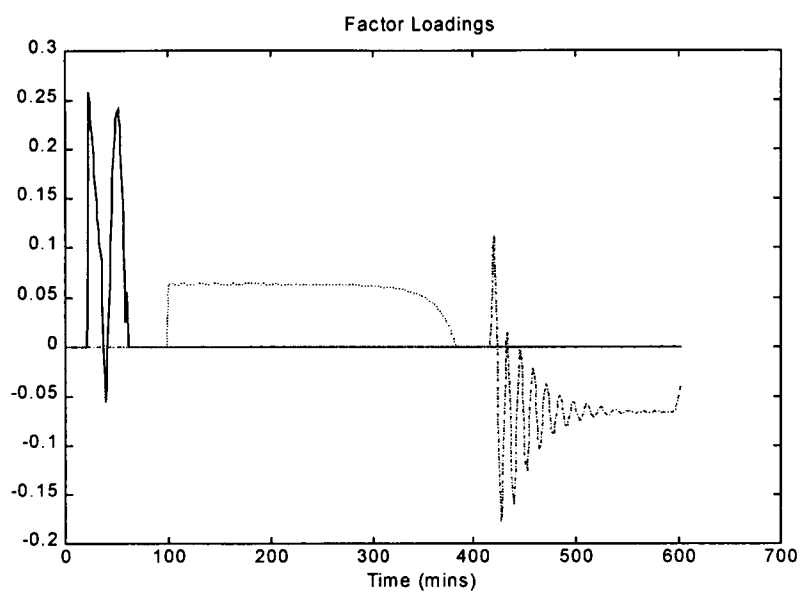
5.3.1 Custom Scale Parameters Example

The heater power profile example is again used as a demonstration. Five zones are obvious from the profile (see figure 5.1). Zone 1 extends from time 0 to 60 where the profile picks up from 0, decays and oscillates smoothly. Zone 2 contains the wild oscillations from time 61 to 110 and is therefore not subjected to PCA. Zone

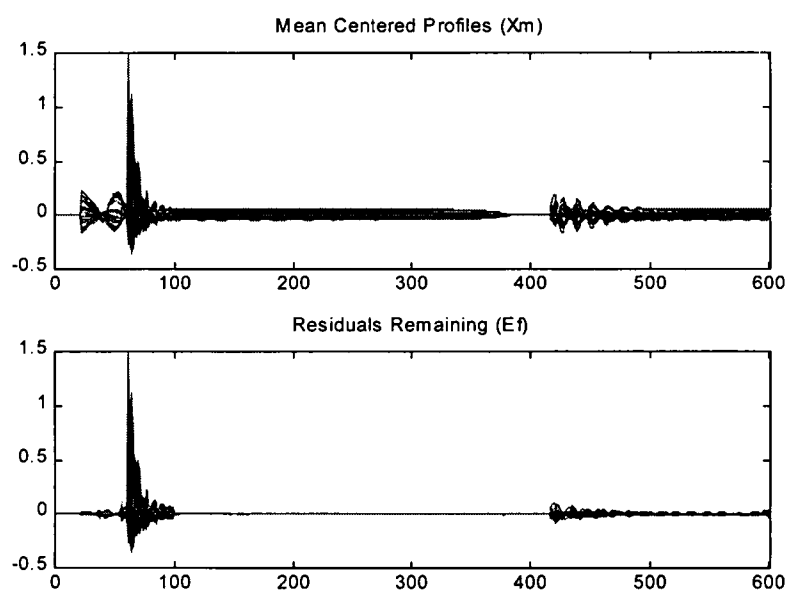
3 extends from time 111 to 384 where the profile settles to a value and subsequently goes to 0. Zone 4 extends from time 385 to 415 where the heater is off (value is 0) and is not subjected to PCA. Zone 5 extends from time 416 to the end of the batch. Zones 1, 3 and 5 are subjected to PCA and the factors and the projected profiles are plotted in figure 5.11.

Note that zeros are padded so that each factor extends for the whole duration of the batch. Also notice that a single factor is able to explain most of the variability in zone 3. In the corresponding PCA example in section 5.1, it took 2 PCs to explain similar variability in zone 3. For comparison purposes, figure 5.12 plots the E matrices for zone 3 using the present approach and also using the 2 PCs from PCA on X_m in section 5.1.

From the figure, it is seen that information from other zones causes the single factor systematic variability in zone 3 to be partitioned between the 2 PCs even though a single factor is enough for the zone. Note using the factor approach, almost all of the variability is explained in zone 3 as evident from the near zero residuals. The residuals remaining in the PCA case are somewhat higher. This implies that the factor approach would be much better suited for extremely subtle special cause detection in zone 3. Also, the proposed approach of dividing the batch into zones and then obtaining the factors significantly enhances the interpretability of the factors. This is very desirable since, in the global picture, the idea is to enhance process understanding by looking at systematic sources of variation.



(a)



(b)

FIGURE 5.11: Use of factors obtained by division into zones
 (a) Factor loadings
 (b) Residuals remaining after projection

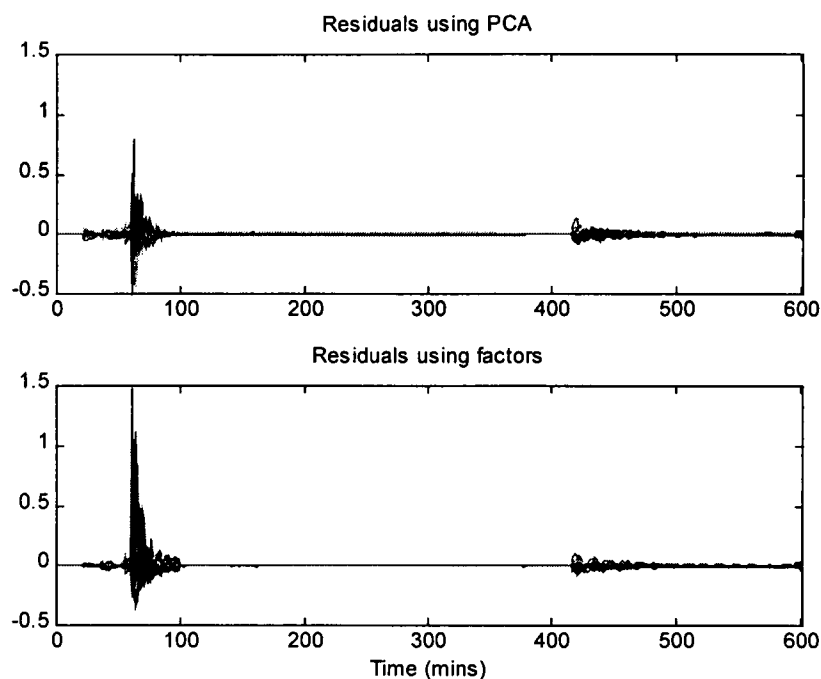


FIGURE 5.12: Comparison between projections using PCA and factors

PCA is the primary tool that has been used by various researchers for analyzing batch data [41-48]. The basic approach has been to perform PCA and then subsequently interpret the results that are obtained. The emphasis thus is on post analysis result interpretation. The method that has been described in this section is proactive in the sense that care is taken in proposing the scale parameters so that minimal effort is needed in interpreting the scale parameters that are well understood. Each approach has its own advantages and disadvantages.

In the next subsection, a simple algorithm is proposed to extract the systematic sources of variability from the profiles. The algorithm is based on an evolving factors type approach.

5.3.2 Evolving Factors Analysis

This subsection develops an algorithm that can be used to extract the zones containing systematic variability that occurs in a particular way in the profiles. An example is used to illustrate its application to batch profiles. The basic approach consists of taking a moving window and performing PCA on the windowed profiles. In case variability in the windowed portion is of a particular type, a single PC would explain more than a pre-specified tolerance (e.g. 90%) of the overall variance. In such a case, a factor is considered to have been found. The window is now extended to include more points and the first PC obtained for the extended window till the tolerance limit is violated. Upon violation of the tolerance, the factor is stored and search is begun for a new factor from the point of tolerance violation. A simple flowchart of the algorithm is given in figure 5.13.

The flowchart indicates that finding a factor and extending the window are the key to the algorithm and criteria for each need to be carefully formulated for general applicability. Criteria for finding a factor are first described. First and foremost, in order for a snapshot to qualify as a factor, the amount of variance explained by the first PC of the snapshot should be a significant fraction of the overall variability in the snapshot. In other words, the ratio of the MSE after and before projection should be much smaller than 1.

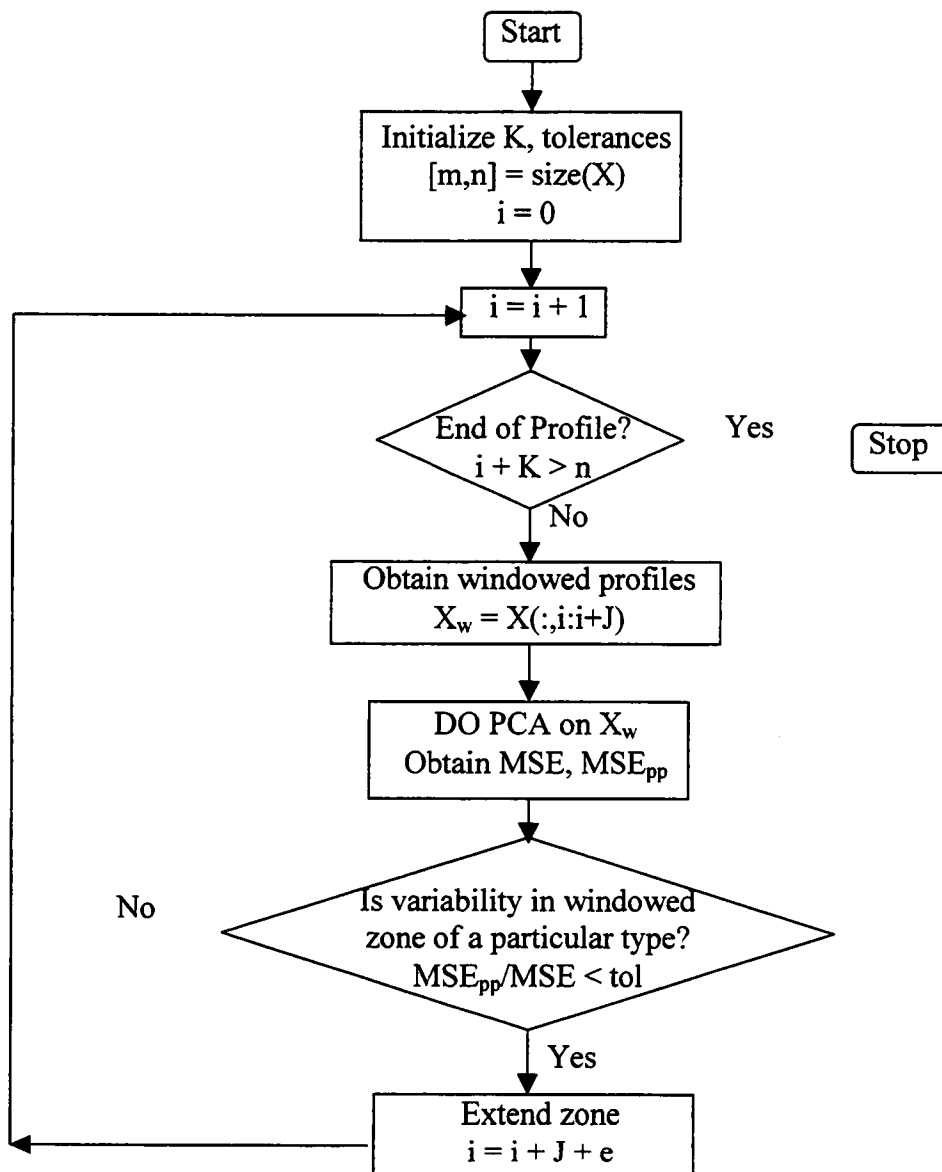


FIGURE 5.13:Flowchart of the algorithm for extracting magnitude scale factors

More formally, let $\mathbf{X}_w$ denote the windowed profiles where the snapshot extends from time s to $s+K$. The MSE before projection is

$$\text{MSE}_x = \text{trace}(\mathbf{X}_w \mathbf{X}_w^T) / (K+1)I \quad \dots\dots\dots(5.4)$$

I denotes the number of profiles. It is assumed that the profiles are column mean centered. A PCA on $\mathbf{X}_w$ with projection using the only the first PC leads to the residual error remaining $\mathbf{E}_w$ as

$$\mathbf{E}_w = \mathbf{X}_w - \mathbf{X}_w \mathbf{v}_w \mathbf{v}_w^T \quad \dots\dots\dots(5.5)$$

where $\mathbf{v}_w$ denotes the first PC of the snapshot. The MSE after projection is obtained as

$$\text{MSE}_e = \text{trace}(\mathbf{E}_w \mathbf{E}_w^T) / (K+1)I \quad \dots\dots\dots(5.6)$$

The criteria for finding a factor then boils down to the ratio $\text{MSE}_e / \text{MSE}_x$ being much much less than 1. A reasonable tolerance like 0.2 may be used. In some cases, an optional criteria may additionally be used for avoiding snapshots that contain high variance features that last for a very short duration. A statistic such as the number of points in the snapshot that contain more than a prespecified threshold (say 10%) of the MSE in the whole snapshot can be used. In case the number of points is less than a prespecified tolerance, the snapshot need not be considered.

Once a snapshot qualifies as a factor, it must be extended. The criteria for extension are simple. The ratio $\text{MSE}_e / \text{MSE}_x$ of the extended snapshot should still meet the tolerance criteria described earlier. In all, the algorithm has two parameters – the minimum length of the window and the ratio tolerance. An optional parameter for avoiding sharp features may also be used.

Let $\mathbf{H}$ denote the matrix containing the factors in the columns with adequate

zero padding. Since PCA is used to obtain the factors in a zone, each column of the matrix is unit length as zero padding does not alter the magnitude. Also, since the zones are assumed to be *non-overlapping*, the columns are orthogonal so that $\mathbf{H}$ is an orthogonal matrix. The projection equations therefore are the same as for PCA and the scale parameters are obtained as obtained from equation 3.2 as

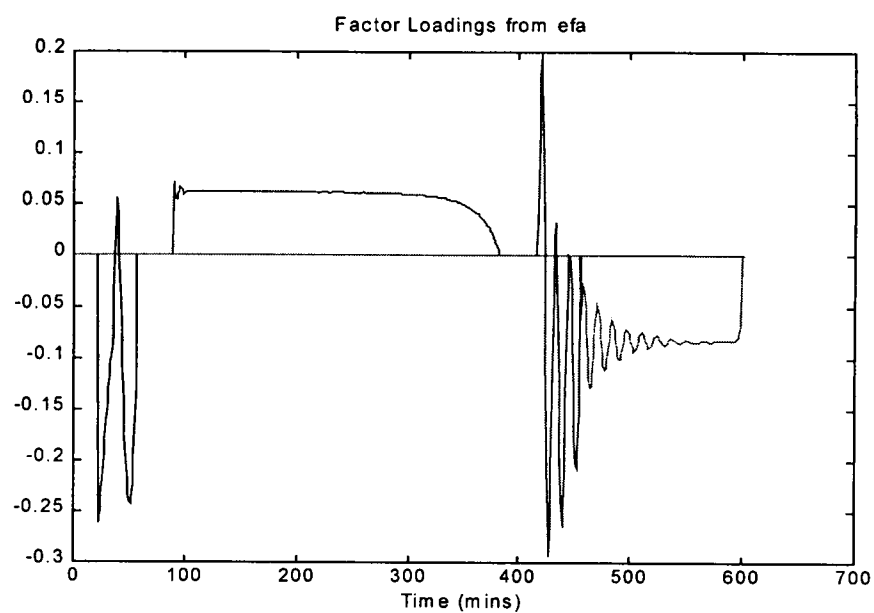
$$\mathbf{MS} = \mathbf{X} \mathbf{H} \quad \text{.....(5.6a)}$$

while the residuals after projection are obtained as

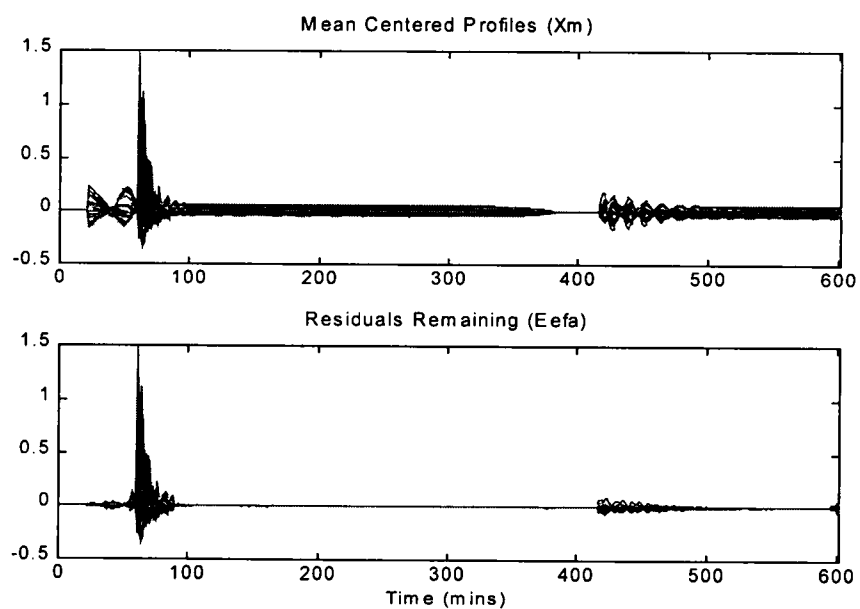
$$\mathbf{E} = \mathbf{X} (\mathbf{I} - \mathbf{H}\mathbf{H}^T) \quad \text{.....(5.6b)}$$

The above equations are the matrix versions of equations 3.2. The scale parameters in $\mathbf{MS}$ may not however be decorrelated as opposed to the decorrelated scores in PCA. In fact high correlation is expected due to autocorrelation between the various zones.

As an example, the algorithm is applied to the heater power profile. The minimum window size used is 50 and the ratio tolerance is taken as 0.2. To avoid sharp features, the number of points that contribute to the snapshot MSE should be greater than 10. The factors thus obtained are plotted in figure 5.14(a). Note that the algorithm successfully identifies the principal factors in the various zones. The mean centered profiles ($\mathbf{X}$) and residuals ($\mathbf{E}$) are plotted in figure 5.14(b). The residuals compare very well with the residuals using PCA as was described in section 5.1 (see figure 5.3). Also, the factors are very easy to interpret, a very desirable feature. The approach thus retains the power of projection methods in special cause detection with the additional benefit of meaningful factors.



(a)



(b)

FIGURE 5.14: Use of factors obtained by proposed EFA approach
 (a) Factor loadings
 (b) Residuals remaining after projection

The proposed algorithm in its present form is able to extract the sources of variability of a particular type as the factors. Depending on the profiles, these factors may need to be further fine tuned for profile projection. As an example, it may be the case that two factors are obtained for zones that are adjacent to each other. In such a case, it may be desirable to combine the two zones into one and obtain as factors the first two PCs. Exercising such an option would however be case specific and no general guidelines can be proposed as such. Suffices to say here that a physical understanding of the process would help a lot resolving such cases. Even so, the obtained factors do provide valuable insight about the predominant type of variability in a zone that is useful from the process understanding standpoint.

5.4 Discussion

In this chapter, the issue of characterizing the variability along the measurement axis has been addressed. From the various illustrations, it is clear that PCA as such is unable to cope with the change in the correlation structure from phase to phase so that the profiles must be divided systematically into time zones that contain a particular type of variability. The division into time zones should be done based on a physical understanding of the process. Usually, a look at the profiles makes such a division obvious so that 'custom' scale parameters can be proposed for the time zones. As an automated technique, an evolving factor type approach was developed to extract the consistent sources of measurement axis variability. The example showed the applicability of the algorithm to batch profiles.

Factors so extracted are very easy to interpret resulting in meaningful scale parameters. The real benefit of the technique lies in these well understood scale parameters. Such an understanding is critical for process improvement. The technique also retains the power of using projection methods for subtle special cause deviations. The only compromise is the resulting scale parameters with correlation between them. This fact can be utilized in detecting abnormal batch operation in case the correlation structure is violated. The batch would then show up as an outlier in the scale parameter subspace. Multivariate control charts can be proposed for the scale parameter space. Further, since the projection methods extract most of the consistent sources of variability in the scale parameters, profiles with special causes would show high residuals (E) after projection. Therefore, formal control charts on the residuals are also needed. The formulation of these control charts forms the subject matter of the next chapter.

Chapter 6

SPC Charts

6.1 Introduction

Statistical process control (SPC) charts are used extensively in the process industry for monitoring purposes. The SPC charts are usually on measurements or parameters that are considered important for the health of a process. These charts use control limits for discriminating between normal and aberrant samples. For process monitoring, the control limits are usually obtained from nominal data from past operation. Normal samples lie within the control limits while aberrant samples show excessive deviation from the nominal operation causing control limit violation. A statistically significant deviation is said to have occurred in case such a violation occurs indicating a significant deviation from the nominal operation.

In batch processes, the consistent variability in the nominal profiles is characterized by the scale parameters. Specifically, the use of time and magnitude scale parameters has been proposed for quantifying the consistent variability in this work. These scale parameters are important indicators of the amount of variability of a particular type in a batch and are therefore important indicators of the state of a batch. SPC charts on the scale parameters are therefore needed. As an illustration, consider again the reactor level example where the profiles are shifted above or below the reference by a fixed amount due to variability in the initial reactor charging. The shift scale parameter for the profile is denoted by S . For nominal operation, S is expected to lie within a bound say $\pm S_{\max}$, about S . In the case where too much (or too

little) material is added to the reactor, the shift parameter would violate the maximum bound for nominal operation. The simple example highlights the application of control charts on the scale parameters for abnormal operation detection.

The variability that is not accounted for by the scale parameters, remains in the residuals after projection. Normal profiles are therefore expected to show small residuals. On the other hand, profiles that contain special causes not accounted for by the scale parameters, would typically show large residuals. SPC charts are therefore also needed on the residuals remaining after projection.

The two types of SPC charts, on the scale parameters and the residuals, complement each other. Violations on the scale parameter charts indicate excessive variability of the scale parameter type while SPC violations on the residuals indicate special causes unaccounted for by the scale parameters.

Typically, batch processes contain measurements on several variables that are correlated to each other. The profile data is therefore inherently 3 way in nature with the various batches, the measurements and time forming the 3 dimensions. In chapters 4 and 5, the focus was on single profile data ie a 2D matrix. In order to avoid confusion, the projection methodology is restated with a consistent nomenclature in section 6.1. It may be considered as a review of the methodology with an extension to handling the 3D data due to the multiple on-line measurements in a batch. The section is also necessary since the control charts would use the various quantities proposed. In section 6.2, control charts on the scale parameters are formulated while section 6.3 proposes control charts for the residuals remaining after projection.

6.2 Profile Projection Review

Consider a set of I 'normal' operation batches from the historical database as the reference set. Each batch contains profiles on K measurements such as temperatures, pressure and flows etc. The first step in the scaling methodology is to time scale the batches using consistently occurring features for time alignment. The time scale parameters so obtained are denoted by TS , which is a matrix with each row containing the scale parameters for a batch. The batches after time scaling are the same duration with J points. The complete data set thus has three dimensions – batches (I), time (J) and measurements (K). Subscripts i , j and k are used to refer to these dimensions respectively. The 3 way array containing the batch profiles is denoted by P_{3d} and is shown in figure 6.1. The k^{th} slice of the 3 way array is the time scaled profiles for the k^{th} measurement variable with each profile being along the I rows.

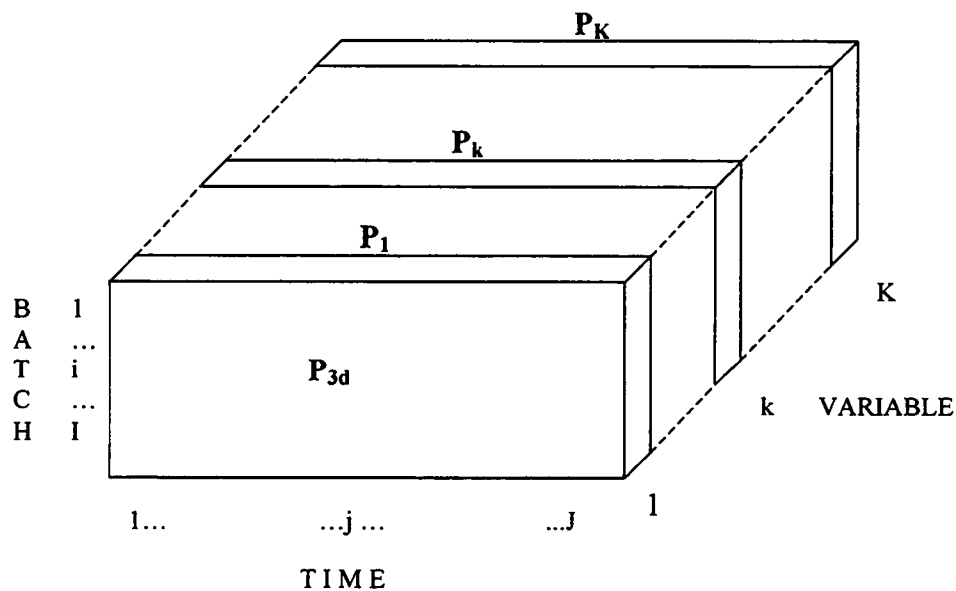


FIGURE 6.1: Organization of batch data array

The first step in magnitude scaling consists of mean centering the columns of the K slices. Let $\mathbf{X}_{3d}$ denote the array resulting after subtracting the mean (also called reference) from the columns of the $\mathbf{P}_{3d}$ array. The magnitude scaling methodology consists of obtaining factors for consistent variability seen in each of the K measurement profiles and projecting onto the factors to obtain the scale parameters. The projection equations in 5.4 and 5.5 in chapter 5 must be applied to the factors for each of the K measurements. Let $\mathbf{H}_k$ ($k = 1$ to K) denote the factors for the k^{th} measurement. As indicated in chapter 5, $\mathbf{H}_k$ is orthogonal so that the projection of the various profiles that are also called the magnitude scale parameters for the k^{th} profile is

$$\mathbf{MS}_k = \mathbf{X}_k \mathbf{H}_k \quad \text{.....(6.1)}$$

The residuals remaining after projection are obtained as

$$\mathbf{E}_k = \mathbf{X}_k - \mathbf{MS}_k \mathbf{H}_k^T \quad \text{.....(6.2a)}$$

Or substituting for $\mathbf{MS}_k$

$$\mathbf{E}_k = \mathbf{X}_k (\mathbf{I} - \mathbf{H}_k \mathbf{H}_k^T) \quad \text{.....(6.2b)}$$

In the above equations, $\mathbf{X}_k$ denotes the k^{th} slice along the third dimension in the $\mathbf{X}_{3d}$ array and is a matrix that contains the mean centered batch profiles for the k^{th} measurement. The equations given above are directly obtained from the projection equations 3.2 in chapter 3. The above process is repeated for all the measurements, $k = 1$ to K , to obtain the magnitude scale parameters and the residuals. The residuals can again be put in a 3 way array form similar to $\mathbf{P}_{3d}$ and is called $\mathbf{E}_{3d}$. The time scale

and magnitude scale parameters for all the profiles can be combined into a scale parameter matrix **SP** as

$$\mathbf{SP} = [\mathbf{TS} \quad \mathbf{MS}_1 \quad \mathbf{MS}_2 \quad \dots \quad \mathbf{MS}_k \quad \dots \quad \mathbf{MS}_K] \quad \dots\dots(6.3)$$

The projection methodology is pictorially described in figure 6.2. The methodology leads to the scale parameters **SP** and the residuals E_{3d} . The next two sections are devoted to the development of SPC charts on the scale parameters and the residuals.

6.3 SPC Charts on Scale Parameters

The scale parameters for the reference set of normal operating batches are obtained in **SP**. The simplest SPC charts are univariate control charts with approximate $\pm 3\sigma$ as the 99% confidence limits. All the scale parameters can be put on a common basis using autoscaling so that the upper and lower control limits are ± 3 since after autoscaling, the standard deviation of each column is 1. More formally, if $\mathbf{SP}_a$ denotes the autoscaled parameter matrix, the $1-\alpha$ limits on the r th scale parameter would be

$$\mathbf{SP}_{s,r,\alpha} = \pm t_{I-1,\alpha/2} \quad \dots\dots\dots(6.4)$$

Where I is the number of samples (rows) and $t_{I-1,\alpha/2}$ corresponds to the probability point on the single sided t-distribution with $I-1$ degrees of freedom and area $\alpha/2$. The limits can also be proposed based on engineering judgement. Such charts should suffice for gross error detection and are used in the present work for demonstration purposes.

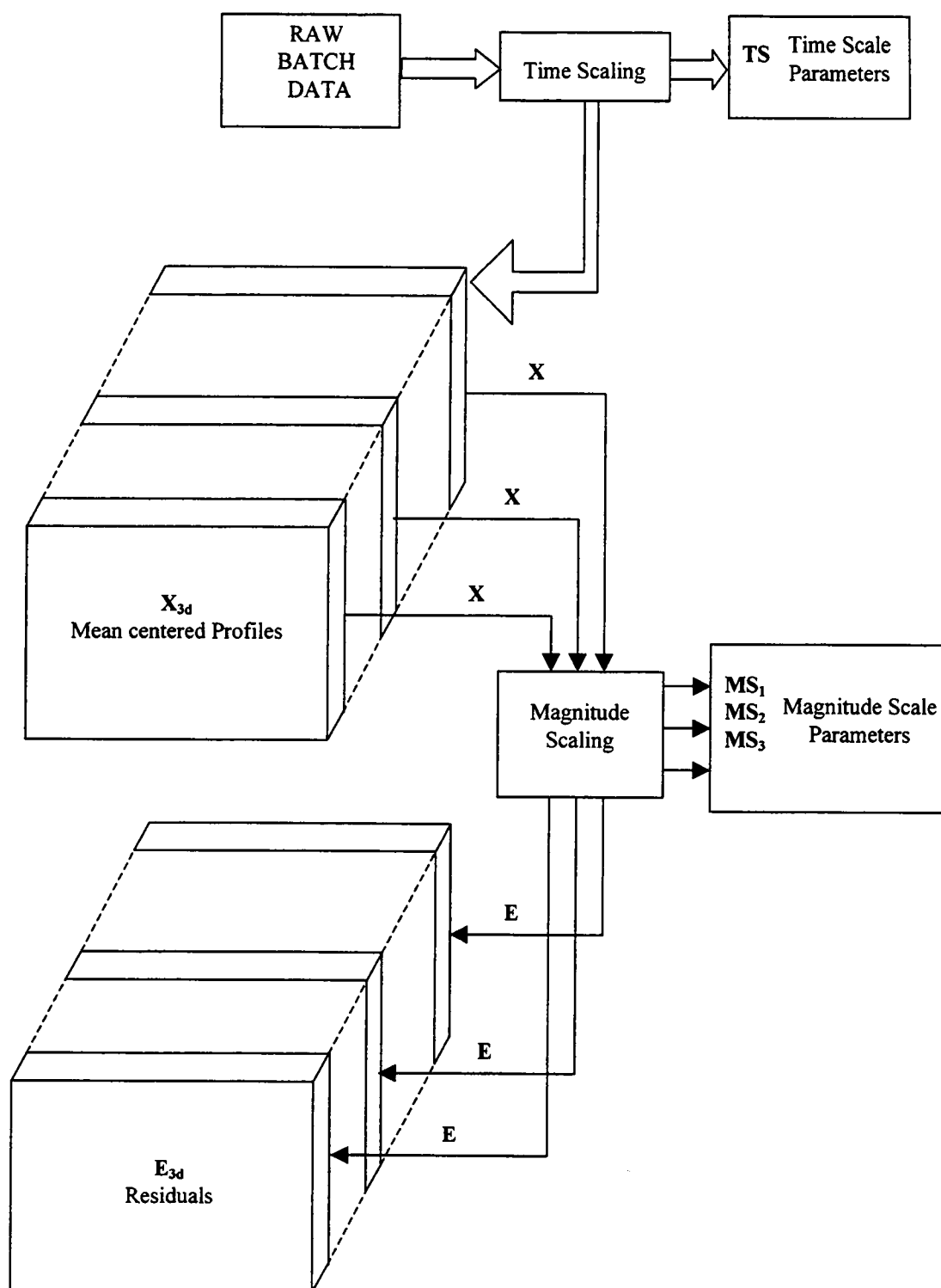


FIGURE 6.2: Pictorial schematic for magnitude scaling

For subtler error detection, multivariate control charts based on the Hotelling's T^2 statistic can be used [21]. The statistic is a quadratic form with inverse of the covariance matrix as the defining matrix. Collinearity between the columns of $\mathbf{SP}$, is possible causing matrix inversion problems. PCA can be used in order to avoid such problems in calculating the relevant statistics. PCA has already been described earlier in chapter 2. If r PCs are retained in the PCA model, then the multivariate statistics used to characterize the samples are - the Hotelling's T^2 and the Q statistic. Figure 6.3 tabulates the equations for the PCA model and the multivariate statistics and the confidence limits on the statistics. The limits on the Q statistic were proposed first by Jackson [25].

In the equations, the matrix λ_r for PCA is diagonal and denotes the covariance of the scores. The diagonality results due to the decorrelated scores in PCA. From the theory of PCA, the diagonal elements of λ_r can also be obtained as the eigenvalues of the matrix $\text{cov}(\mathbf{X})$. Note that both the T^2 and the Q statistic are quadratic forms. The T^2 statistic is a weighted distance of the projection of the samples on the PC hyperplane from the origin and measures the variability in the PC subspace. The Q statistic on the other hand, measures the perpendicular distance of the samples from the PC hyperplane.

In the case of SPC charts for the scale parameters, the autoscaled matrix $\mathbf{SP}_s$, serves as the $\mathbf{X}$ matrix. When collinearity exists between the scale parameters, the Q statistic is of primary importance since abnormal operation where the correlation structure breaks down would result in out of control Q statistic.

PCA Model

PC Equation (r retained PCs)

$$\mathbf{X} = \mathbf{T}_r \mathbf{V}_r + \mathbf{E} = \mathbf{U}_r \boldsymbol{\Sigma}_r \mathbf{V}_r^T + \mathbf{E}$$

Sample Scores

$$\mathbf{t}_i = \mathbf{x}_i \mathbf{V}_r$$

Sample Residuals

$$\mathbf{e}_i = \mathbf{x}_i - \mathbf{t}_i \mathbf{V}_r^T = \mathbf{x}_i (\mathbf{I} - \mathbf{V}_r \mathbf{V}_r^T)$$

Multivariate Sample Statistics

Hotelling's T^2

$$T_i^2 = \mathbf{t}_i \text{cov}(\mathbf{t})^{-1} \mathbf{t}_i^T = \mathbf{t}_i \boldsymbol{\lambda}_r^{-1} \mathbf{t}_i^T$$

$$\text{where } \boldsymbol{\lambda}_r = \frac{\boldsymbol{\Sigma}_r^2}{I - 1}$$

Q statistic

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

Confidence Limits

Hotelling's T^2

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

Q Statistic

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

$$\text{where } \Theta_i = \text{trace} \left(\frac{\mathbf{E} \mathbf{E}^T}{I - 1} \right)^i \text{ for } i = 1, 2, 3$$

Variable Descriptions

$\mathbf{X}$	Data Matrix (I by J)
$\mathbf{V}_r$	Retained PC loadings (J by r)
$\mathbf{T}_r$	Scores on retained PCs (I by r)
$\mathbf{E}$	Residuals remaining (I by J)
$\mathbf{x}_i$	i^{th} sample row vector (1 by J)
$\mathbf{t}_i$	i^{th} sample scores (1 by r)
F	Critical value of F distribution
c_α	standard normal deviate

FIGURE 6.3: PCA model equations and statistics

Note that in case high correlation does not exist between the various scale parameters, matrix inversion problems are non-existent so that the T^2 statistic alone can be used to monitor the complete scale parameter space since all the PCs are retained. Alternately, the univariate control charts on the scale parameters would suffice for monitoring purposes. The T^2 chart is however convenient since it is a single chart with information on all the scale parameters. The real utility of the multivariate charts for subtle error detection exists only when high correlation exists between the scale parameters.

A word of caution on the use of confidence limits on the scale parameters is perhaps in order. The scale parameters contain the deterministic information in the profiles so that the assumption of normality of the retained scores is questionable. The confidence limits, univariate or multivariate T^2 , are therefore only rough guidelines and may be altered based on process knowledge. The next section proposes multivariate charts on the residuals remaining (E_{3d}) after projection.

6.4 SPC Charts on Residuals

The projection methodology extracts the consistent sources of variability in the scale parameters to result in the residual array E_{3d} . SSE type statistics can be proposed on the residuals. The task is simplified since most of the systematic variability is extracted in the scale parameters so that the residuals remaining are stochastic in the sense that there is no structure remaining in the residuals. These residuals for the K different measurements per batch can therefore be combined

together to give a single SSE type statistic rather than K different statistics for each measured variable, thus simplifying the picture.

In order to combine the residuals for the various measurements, some type of scaling must be used since the units may be different on the various measurements. Autoscaling is ideal in this case since the residuals are assumed non-systematic. Thus, every time point for each variable is put on a unit variance basis. Let E_{3d} denote the autoscaled array where each column of the array is obtained from E_{3d} by mean centering and dividing by the column standard deviation. Referring to figure 6.4, the autoscaled residuals of the i^{th} batch for all the measurements are therefore along the i^{th} horizontal slice of the array. The combined residual statistic for the i^{th} batch may be obtained as

$$Q_i = \sum_{j=1}^J \sum_{k=1}^K E_s(i, j, k) \quad \text{.....(6.6)}$$

Approximate confidence limits on the statistic can be obtained using Jackson's expression as indicated in section 6.2. This simple statistic can be conveniently used for detecting special causes not accounted for by the scale parameters.

A few comments are in order. Since the systematic variability is already extracted in the scale parameters, the residuals in the zones containing the systematic variability are expected to be very small. The autoscaling would then punish very small deviations heavily in these zones making subtle special causes easily detectable.

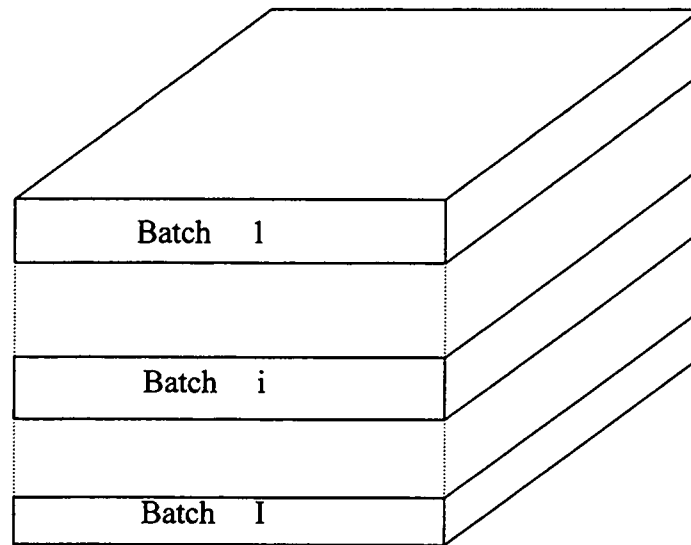


FIGURE 6.4: Each batch as a slice in the 3D data array

Note that since the autoscaling is done *after* the extraction of the systematic variability, the SPC charts are more sensitive to special causes than traditional methods that use autoscaling *before* applying PCA. The inadequacy of autoscaling in extracting meaningful scale parameters has already been demonstrated in chapter 5. The proposed SPC methodology thus combines the extraction of meaningful scale parameters along with the ability to detect extreme subtle causes. These points shall be illustrated in chapter 8 with a demonstration on a realistic batch process simulation. The description of the simulation forms the subject matter of the next chapter.

Chapter 7

Batch Process Simulation

7.1 Introduction

This chapter describes a batch polymerization reactor simulation. The simulation is used to generate realistic batch profiles to demonstrate the application of the methodology for profile characterization outlined in earlier chapters. The system modeled is a batch reactor polymerizing methyl methacrylate monomer in toluene solvent to produce polymethyl methacrylate (PMMA). Azobisisobutyronitrile (AIBN) is used as the initiator. The system is chosen due to the highly non-linear modeling equations, a typical characteristic of industrial batch processes. Also, research work in the past have proposed modeling equations for the polymer quality which gives a convenient means for obtaining quantitative quality parameters. Such quantitative measures can be used by regression techniques such as multiple linear regression (MLR), Principal component Regression (PCR) and partial least squares (PLS) for predicting the quality from the scale parameters that characterize the batch operation. Study of the correlation of the scale parameters with the polymer quality also forms an important part of the overall profile characterization framework and the quality data generated from the simulation is again used for demonstration purposes. The batch reactor system and the modeling equations are briefly described in the following sections.

7.2 Process Description

The batch process is based on the experimental set up used by Souroush and Kravaris for the application of nonlinear control techniques and a schematic is drawn in figure 7.1 [92]. The set up consists of a reactor with a cooling system to remove the heat released during the polymerization. The heat removal occurs by heat transfer between the coolant circulating in the jacket around the reactor. Water is used as the cooling fluid and is recirculated in the cooling circuit using a centrifugal pump.

The control system consists of a PID controller to manipulate the heat input to the cooling system to track the reactor temperature to the set point. A split range at the controller output is used to manipulate the heat input to the circulating coolant. Thus, in case the controller output is positive ie heat must be input, the electric heater is switched on while if the output is negative ie heat must be removed, the coolant flow valve is opened to let in coolant.

The operating procedure consists of 3 distinct phases – reactor charging, preheating to the reaction set point and finally AIBN addition indicating the beginning of the reaction. The reactor is charged with the monomer and the solvent using the two inlets provided for the purpose. The flow profiles for the two components are pre-specified. The charging step begins at time $t_{mi} = 1$ min after the simulation is started when the monomer flow inlet is opened. The monomer inlet is closed at $t_{mf} = 16$ mins. The solvent inlet is opened at $t_{si} = 10$ min and closed at $t_{sf} = 20$ min. The charging phase therefore lasts till 20 mins after the beginning of the simulation.

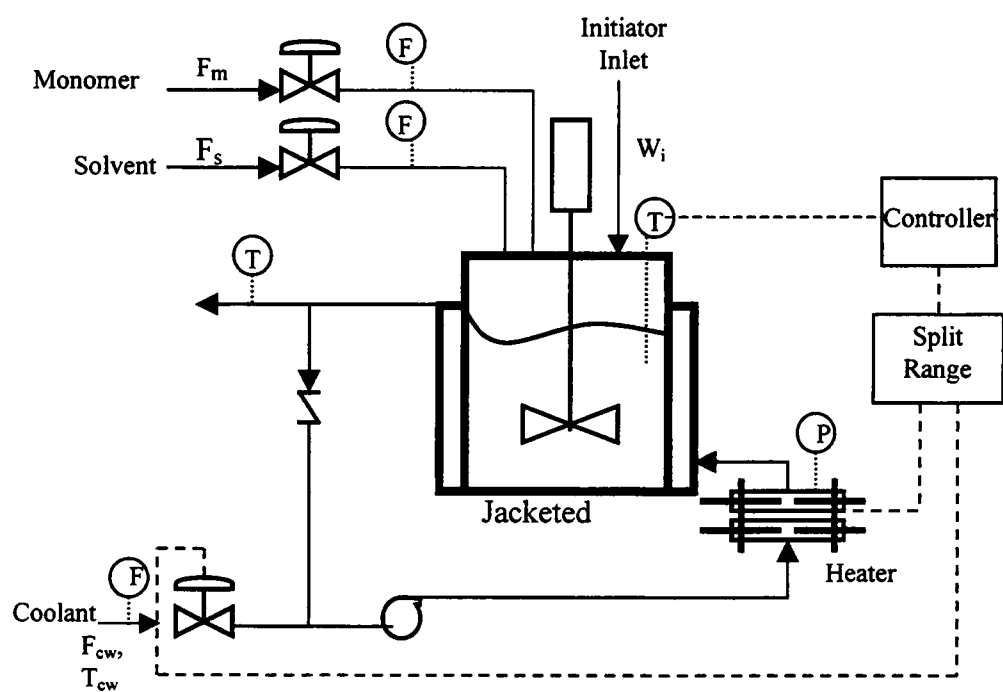


FIGURE 7.1: PMMA batch process schematic

The way the simulation is setup, no variability exists in the times at which the valves are opened or closed. Variability can however exist in the flows of the monomer and solvent, denoted by F_m and F_s , respectively.

The preheat phase begins after the charging is complete at $t_h = 30$ mins when the controller is enabled. Upon controller enabling, the reactor temperature set point is a constant at 320 K, the reaction temperature. In order to avoid reset windup, the integral mode is enabled only after the temperature rises to within 5K of the set point. It thus takes about half an hour for the reactor temperature to rise to the set point. No variability exists in the time at which the control system is enabled.

Finally, the initiator AIBN is added to the reactor indicating the beginning of the reaction at about 1 hour (60 mins) after the beginning of the simulation. Nominal variability exists in the time t_{ii} at which the initiator is added. The time thus varies randomly between 55 to 65 mins from batch to batch. Variability is also possible in the amount of initiator charged (W_i). The variability in t_{ii} simulates an operator mediated task. The previous phases of charging and preheating, where no variability exists in the times, simulate automated tasks. The batch is considered complete when the conversion reaches around 0.995 (about 6 hours after AIBN addition for nominal operation).

The measurements include the reactor temperature (T), the jacket temperature (T_j), the coolant flow (F_{cw}), the electric heater power (P), the monomer (F_m) and solvent (F_s) flows and the weight (W_i) of initiator added. The polymer quality is characterized using the number average molecular weight (MW_n), the weight average

molecular weight (MW_w) and the polydispersity index (PDI). MW_n measures the average chain length while PDI is a measure of the spread in the molecular weight distribution of the polymer. For many industrial applications, these quality parameters are considered sufficient for characterizing the polymer product.

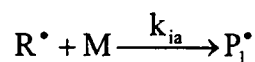
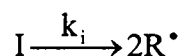
This completes a brief overview of the PMMA batch reactor process description. The next section describes in brief, the mathematical equations that are used to model the batch polymerization process.

7.3 Model Equations

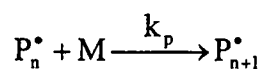
The process is modeled using dynamic material and energy balance equations on the reactor and the cooling system. In the modeling equations, the physical properties such as density and specific heat are considered constant. The values used are calculated at the reactor set point temperature. Thus since the reactor temperature is controlled, the assumption is good once the set point is reached or in other words, once the reaction begins. The model is therefore reasonable in the reaction phase where most of the non-linearity occurs due to the kinetics of the reaction.

The kinetics of methyl methacrylate polymerization reaction in toluene solvent is modeled using the free radical polymerization mechanism [87-92]. The reaction mechanism is given in figure 7.2 [89, 90]. The steps in the polymerization include initiation, propagation, chain transfer and chain termination.

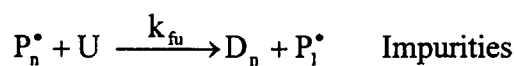
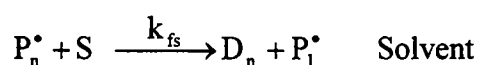
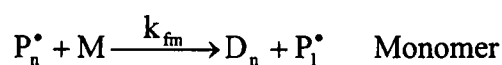
Initiation



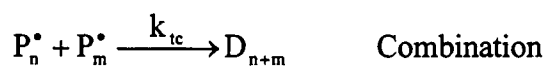
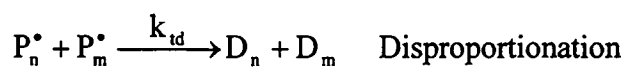
Propagation



Chain Transfer



Termination



where

I	Initiator
R	Primary radical
M	Monomer molecule
S	Solvent molecule
U	Impurities in the monomer
$P_n^\bullet$	Live polymer chain of length n
D_n	Dead polymer chain of length n

FIGURE 7.2: Free radical polymerization reaction mechanism

In the initiation step, the initiator forms a free radical, which in turn causes the monomer to form an alkyl free radical. Polymerization occurs primarily through the propagation step where monomer molecules are added to existing alkyl free radicals leading to higher chain length radicals. These radicals are called live chains since they can propagate by addition of more and more monomers. Dead chains or dead polymer molecules that cannot propagate, are produced by chain transfer and chain termination. It is evident from the mechanism that the competition between the propagation and dead chain producing steps, determines the quality of the polymer. Indeed, in practical free radical polymerizations, impurities in small concentrations that significantly enhance the dead chain production by chain transfer are known to have a drastic impact on the polymer quality [94]. A hypothetical impurity capable of such chain transfer is therefore included in the reaction mechanism.

The most rigorous kinetic model for the reaction should be based on all the reacting species including P_n and D_n for $n = 1$ to ∞ . This is however impractical and the method of moments using only the first few moments on the live or dead chain distribution is used as a good approximation [90]. The live and dead chain distributions are characterized by their 0th, 1st and 2nd moments. The moments are defined as

$$\lambda_i = \sum_n n^i [P_n] \text{ for } i = 0, 1 \text{ and } 2 \quad \dots\dots\dots(7.1a)$$

for the live chains and as

$$\mu_i = \sum_n n^i [D_n] \text{ for } i = 0, 1 \text{ and } 2 \quad \dots\dots\dots(7.1b)$$

for the dead chains. These moments that partially describe the chain distribution can be used to obtain the polymer quality parameters as

$$MW_n = \frac{\lambda_1 + \mu_1}{\lambda_0 + \mu_0} \dots\dots\dots(7.2a)$$

and

$$MW_w = \frac{\lambda_2 + \mu_2}{\lambda_1 + \mu_1} \dots\dots\dots(7.2b)$$

The polydispersity index (PDI) is obtained as

$$PDI = \frac{MW_w}{MW_n} \dots\dots\dots(7.2c)$$

The reaction rate equations on the various species and the moments are tabulated in figure 7.3. The kinetic parameter values are tabulated in table 7.1 [89].

The PMMA polymerization reaction is characterized by a strong gel-effect when the reaction rates increase at high monomer conversions. The gel effect is a highly nonlinear process and is modeled by using the constitutive equations of Chiu et al that uses the free volume concept as in figure 7.4 [91].

The gel-effect occurs primarily since the polymer density is significantly higher than the monomer density leading to a volume contraction. So at high conversions, the reaction rates significantly increase owing to low reaction mixture volume causing an increase in the reactant concentrations. Such an automatic increase in the reaction rates is therefore also known as autoacceleration. Using the kinetic model, the dynamic equations for modeling the batch reactor system are obtained as in figure 7.5.

$$r_M = (k_p + k_{fm})[M]\mu_0$$

$$r_I = -k_i[I]$$

$$r_{\mu_0} = 2fk_i[I] - k_t\mu_0^2$$

$$r_{\mu_1} = 2fk_i[I] + k_p[M]\mu_0 + (k_{fm}[M] + k_{fs}[S])(\mu_0 - \mu_1) - k_t\mu_0\mu_1$$

$$r_{\mu_2} = 2fk_i[I] + k_p[M](2\mu_1 + \mu_0) + (k_{fm}[M] + k_{fs}[S])(\mu_0 - \mu_2) - k_t\mu_0\mu_2$$

$$r_{\xi_0} = (k_{fm}[M] + k_{fs}[S])\mu_0 + (k_{td} + 0.5k_{tc})\mu_0^2$$

$$r_{\xi_1} = (k_{fm}[M] + k_{fs}[S])\mu_1 + k_t\mu_0\mu_1$$

$$r_{\xi_2} = (k_{fm}[M] + k_{fs}[S])\mu_2 + k_t\mu_0\mu_2 + k_{tc}\mu_1^2$$

where the rate constants are as follows

k_i = initiator decomposition

k_p = propagation

k_{fm} = chain transfer to monomer

k_{fs} = chain transfer to solvent

k_{fu} = chain transfer to impurities in the monomer

k_{td} = termination by disproportionation

k_{tc} = termination by chain combination

$k_t = k_{td} + k_{tc}$ = overall termination

FIGURE 7.3: Kinetic rate equations for the reaction mechanism

TABLE 7.1: Kinetic parameters used in the model

Rate Constant (k)	Arrhenius Frequency (A)	Activation Energy (E)	Equation
k_i	$1.0533 \cdot 10^{15} \text{ s}^{-1}$	$1.2877 \cdot 10^4 \text{ kJ} \cdot \text{kmol}^{-1}$	$k = A \exp^{(-E/RT)}$
k_{p0}	$4.9167 \cdot 10^5 \text{ m}^3 \cdot \text{kmol}^{-1} \cdot \text{s}^{-1}$	$1.8283 \cdot 10^4 \text{ kJ} \cdot \text{kmol}^{-1}$	do
k_{fm}	$4.6610 \cdot 10^9 \text{ m}^3 \cdot \text{kmol}^{-1} \cdot \text{s}^{-1}$	$7.4479 \cdot 10^4 \text{ kJ} \cdot \text{kmol}^{-1}$	do
k_{fs}	$4.9658 \cdot 10^8 \text{ m}^3 \cdot \text{kmol}^{-1} \cdot \text{s}^{-1}$	$6.6150 \cdot 10^4 \text{ kJ} \cdot \text{kmol}^{-1}$	do
k_{fu}	-	-	$k_{fu} = 1 \cdot 10^{-4} k_p$
k_{t0}	$9.8000 \cdot 10^9 \text{ m}^3 \cdot \text{kmol}^{-1} \cdot \text{s}^{-1}$	$2.9442 \cdot 10^3 \text{ kJ} \cdot \text{kmol}^{-1}$	$k = A \exp^{(-E/RT)}$
k_{td}/k_{tc}	$2.5278 \cdot 10^3 \text{ m}^3 \cdot \text{kmol}^{-1} \cdot \text{s}^{-1}$	$1.7178 \cdot 10^4 \text{ kJ} \cdot \text{kmol}^{-1}$	do
$k_{\theta p}$	$3.0233 \cdot 10^{13} \text{ s}^{-1}$	$1.1700 \cdot 10^5 \text{ kJ} \cdot \text{kmol}^{-1}$	do
$k_{\theta t}$	$1.4540 \cdot 10^{20} \text{ s}^{-1}$	$1.4584 \cdot 10^5 \text{ kJ} \cdot \text{kmol}^{-1}$	$k = [I]_0 A \exp^{(-E/RT)}$

Initiator Efficiency $f = 0.58$

$$k_t = \frac{k_{t0}}{1 + \frac{\xi_0 k_{t0}}{D k_{\theta t}}}$$

$$k_p = \frac{k_{p0}}{1 + \frac{\xi_0 k_{p0}}{D k_{\theta p}}}$$

with

$$D = \exp\left(\frac{2.303(1 - \phi_p)}{A(T) + B(1 - \phi_p)}\right)$$

$$A = 0.168 - 8.21 \times 10^{-6} (T - T_{gp})^2$$

$$B = 0.03$$

where

T_{gp} is the glass transition temperature of PMMA

ϕ_p is the volume fraction of the polymer

FIGURE 7.4: Gel effect constitutive equation and parameters

Reactor Material Balance

$$\frac{d[M]V}{dt} = r_M V$$

$$\frac{d\mu_i V}{dt} = r_{\mu_i} V \text{ for } i = 0, 1 \text{ and } 2$$

$$\frac{d\xi_i V}{dt} = r_{\xi_i} V \text{ for } i = 0, 1 \text{ and } 2$$

Reactor Energy Balance

$$\rho C_p \frac{dT V}{dt} = r_H V - U_x A_x (T - T_j)$$

Jacket Energy Balance

$$m_w C_{pw} \frac{dT_j}{dt} = U_x A_x (T - T_j) + U_\infty A_\infty (T_j - T_\infty) + [P_{htr} - F_{cw} \rho_w C_{pw} (T_j - T_{cw})]$$

with

$$\text{instantaneous jacket heat transfer coefficient} = U_x = U_0 [\alpha + (1 - \alpha) \exp(-\beta x_m^\gamma)]$$

$$\text{instantaneous jacket heat transfer area} = A_x = A_0 (1 + \epsilon x_m)$$

$$\text{volume expansion factor} = \epsilon = \varphi_{m0} \left(\frac{\rho_m}{\rho_p} - 1 \right)$$

$$\text{initial monomer volume fraction} = \varphi_{m0} = \frac{N_{m0} \rho_m}{\sum_{i=m,s,i} N_{i0} \rho_i}$$

$$\text{monomer conversion} = x_m = 1 - \frac{N_m}{N_{m0}}$$

$$\text{rate of heat generation} = r_H = (-H_p) k_p \mu_0$$

$$\text{heat actually input by the heater to the jacket } P_{htr} = \eta_{htr} P$$

where P is the controller output

FIGURE 7.5: Dynamic process model equations

In addition to the material balance equations, the reactor and jacket energy balance equations are also given in the figure. Note that the quasi steady state approximation (QSSA) is not used in the simulation equations since the approximation breaks down in the gel effect due to accumulation of free radicals.

The physical properties and the equipment parameters for use in the equations are tabulated in table 7.2 [92]. The PID controller tuning constants and the split range implementation equations are tabulated in figure 7.6. Note that gain scheduling is used so that the controller is tuned to be aggressive in the reaction phase as indicated by the high gain. This is necessitated to remove the high reaction heats in the gel effect phase to avoid significant temperature overshoots.

TABLE 7.2: Physical properties and other parameters for the process

$MW_m = 100.12 \text{ kg}\cdot\text{kmol}^{-1}$	$MW_s = 92.14 \text{ kg}\cdot\text{kmol}^{-1}$	$MW_i = 164.21 \text{ kg}\cdot\text{kmol}^{-1}$
$\rho_m = 915.1 \text{ kg}\cdot\text{m}^{-3}$	$\rho_s = 842.0 \text{ kg}\cdot\text{m}^{-3}$	$\rho_i = 915.0 \text{ kg}\cdot\text{m}^{-3}$
$\alpha = 0.5$	$B = 7.0$	$\gamma = 3.0$
$U_{\infty}A_{\infty} = 5.57\cdot 10^{-3} \text{ kJ}\cdot\text{s}^{-1}\cdot\text{K}^{-1}$	$P_{\max} = 3.13 \text{ kJ}\cdot\text{s}^{-1}$	$F_{\text{cwmax}} = 3.0\cdot 10^{-4} \text{ m}^3\cdot\text{s}^{-1}$
$m_w = 4.0 \text{ kg}$	$C_{pw} = 4.2 \text{ kJ}\cdot\text{kg}^{-1}\cdot\text{K}^{-1}$	$\rho_w = 1000 \text{ kg}\cdot\text{m}^{-3}$
$C_p = 2.2 \text{ kJ}\cdot\text{kg}^{-1}\cdot\text{K}^{-1}$	$H_p = -5.78\cdot 10^4 \text{ kJ}\cdot\text{kmol}^{-1}$	$d = 8\cdot 10^{-2} \text{ m (reactor dia)}$

PID Controller

$$\text{Controller output } u = K_c \left[e + \frac{1}{\tau_i} \int e dt + \tau_D \frac{de}{dt} \right]$$

Controller Tuning

Before Reaction Begins

$$K_c = 0.04 \text{ kJ} \cdot \text{K}^{-1} \quad \tau_i = 200 \text{ s} \quad \tau_D = 0 \text{ s}$$

After Reaction Begins

$$K_c = 1 \text{ kJ} \cdot \text{K}^{-1} \quad \tau_i = 2000 \text{ s} \quad \tau_D = 0 \text{ s}$$

Split Range Coordination

$$\text{Heater Power } P = \begin{cases} u & \text{if } 0 \leq u \leq P_{\max} \\ P_{\max} & \text{if } u \geq P_{\max} \\ 0 & \text{if } u < 0 \end{cases}$$

$$\text{Coolant Flow } F_{\text{cw}} = \begin{cases} \frac{-u}{C_{\text{pw}} \rho_w (T_j - T_{\text{cw}})} & \text{if } -F_{\text{cwmax}} C_{\text{pw}} (T_j - T_{\text{cw}}) \leq u < 0 \\ F_{\text{cwmax}} & \text{if } u < -F_{\text{cwmax}} C_{\text{pw}} (T_j - T_{\text{cw}}) \\ 0 & \text{if } u \geq 0 \end{cases}$$

FIGURE 7.6: Controller equations and parameters

Finally, autocorrelated noise is added to the various measurements and also some of the model parameters to simulate stochastic disturbances in the process. Specifically, autocorrelated noise is added to the jacket heat transfer coefficient U and the heater power efficiency η_{hr} .

This completes a brief description of the simulation. SIMULINK is used to implement the batch reactor system. The next section briefly describes the nominal operation set generated from the simulation and the various abnormal operation scenarios.

7.4 Simulated Historical Database

A total of 100 batches are simulated with 96 batches forming the nominal reference set and the remaining 4 simulating special cause disturbance scenarios. Variability in the process is introduced by varying the following parameters

1. Monomer and Solvent Flow Rates (F_m and F_s)
2. Initiator addition time (t_{ii})
3. Weight of initiator charged (W_i)
4. Mole fraction of impurities in the monomer

These parameters are specified at the beginning of each simulation and do not change through the course of the simulation. Negligible variability is kept in the amounts charged to the reactor, as is expected in recipe based industrial batch operation. These measurements are therefore not subject to analysis.

Variability is also introduced as the simulation proceeds in time by pre-specifying the following profiles

1. The jacket heat transfer coefficient
2. The heater efficiency
3. The coolant temperature
4. The oxygen concentration in the reaction mixture

The above profiles allow for abnormalities that develop as the batch progresses in time. For example the heat transfer coefficient may go down significantly during a batch due to a loss in the agitator power. Similarly, an increase in the concentration of oxygen in the reaction mixture would slow down the reaction rate as oxygen is a known reaction inhibitor. The decrease in the reaction rate is simulated as the initiator efficiency f reducing linearly with oxygen concentration.

Several disturbance scenarios are possible by abnormally varying any of the above parameters. Of all the possibilities, four particular ones of interest used in this research are

1. The oxygen concentration ramps up during a batch. This is possible in case nitrogen bubbling fails so that the dissolved oxygen content increases and would slow down the reaction.
2. The coolant temperature increases midway through the batch.
3. The heater efficiency ramps down during a batch. In an industrial setting where saturated steam is used as the heating fluid, such a decrease in efficiency is seen during boiler pressure surges.

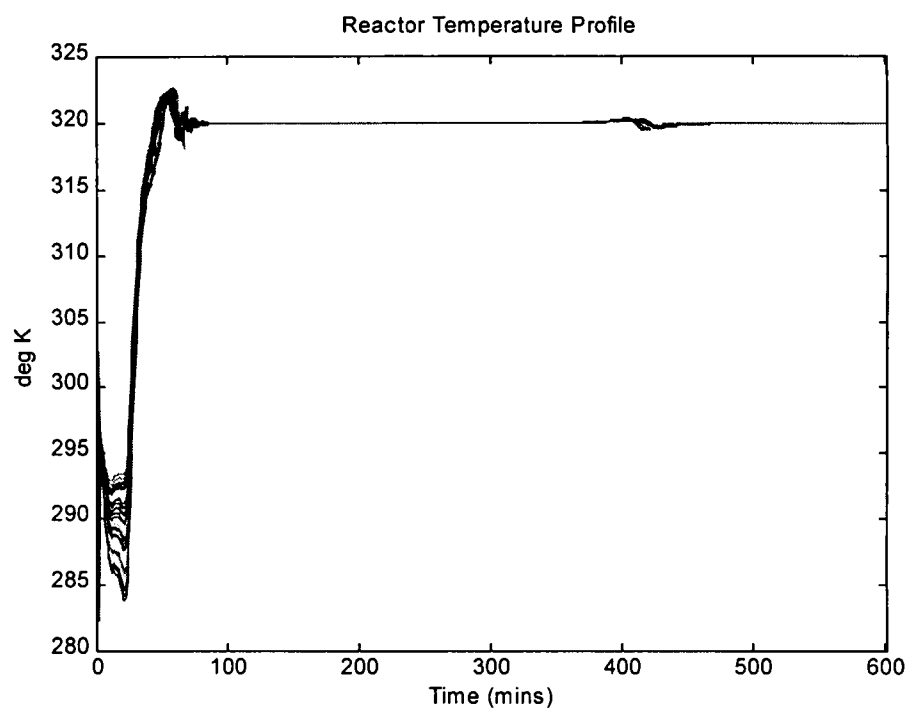
4. The agitator power goes down as the batch progresses in time causing a decrease in the jacket heat transfer coefficient.

One batch simulating each of the above cases is simulated so that four special cause batches are also a part of the complete historical database. Table 7.3 tabulates the nominal ranges of the various parameters used for generating the 96 batches in the reference set. Ranges for the parameters in case of a disturbance scenario are also given in the table. The profiles generated for the nominal set are plotted in figure 7.7.

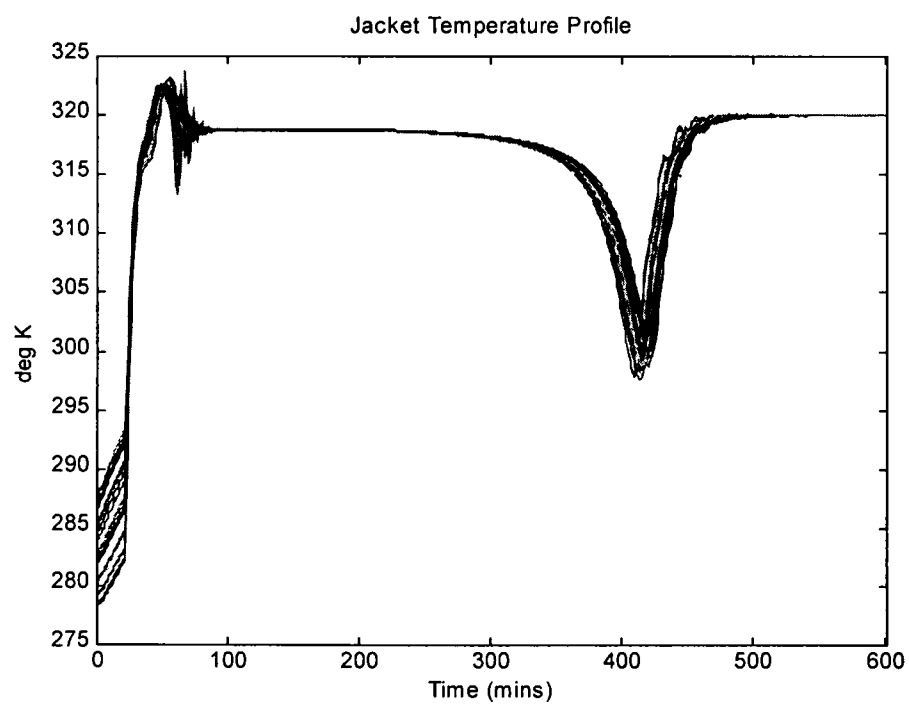
This completes a description of the historical database that is generated using the simulation. The next chapter applies the scaling methodology proposed previously to the simulated data set consisting of nominal and abnormal batches. It is hoped that it demonstrates the coming together of all the concepts illustrated in bits and pieces in the past few chapters.

TABLE 7.3: Nominal and abnormal operation ranges for some process parameters

Parameter	Type	Nominal Range	Abnormal Operation
$F_m (10^5 \text{ kmol} \cdot \text{s}^{-1})$	S	1.0021 - 1.0023	NA
$F_s (10^5 \text{ kmol} \cdot \text{s}^{-1})$	S	0.6853 - 0.6855	NA
$T_{ii} (\text{min})$	S	55 - 65	NA
$W_i (\text{grams})$	S	32.00 - 32.02	NA
Monomer purity (%)	S	96 - 100	
$T_{cw} (\text{K})$	P	275 - 285	300
$U_0 (\text{J} \cdot \text{m}^2 \cdot \text{s}^{-1} \cdot \text{K}^{-1})$	P	195 - 205	150
η_{htr}	P	0.98 - 1.02	0.9
$\text{O}_2 \text{ content (ppm)}$	P	0	10

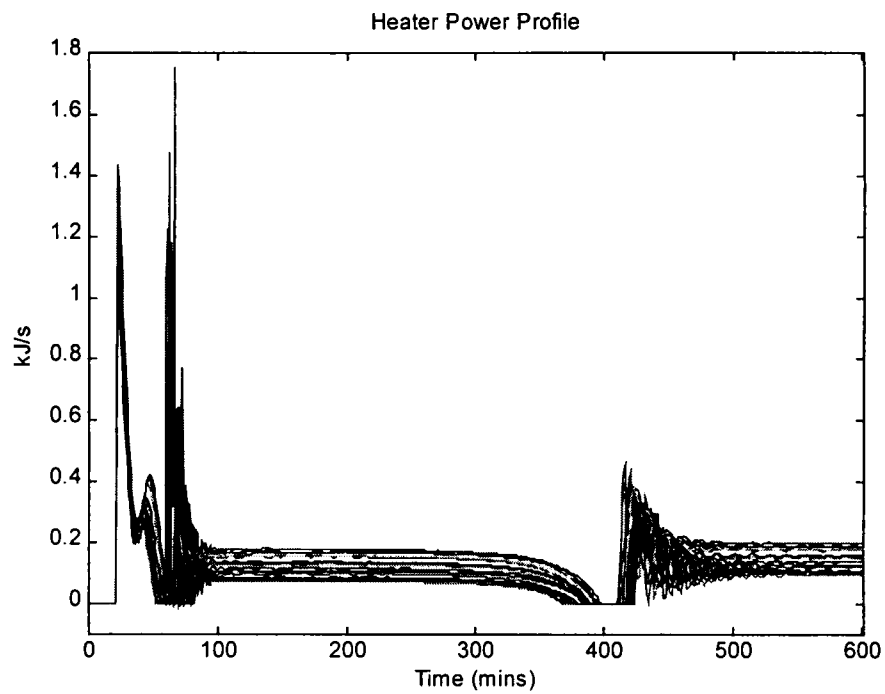


(a)

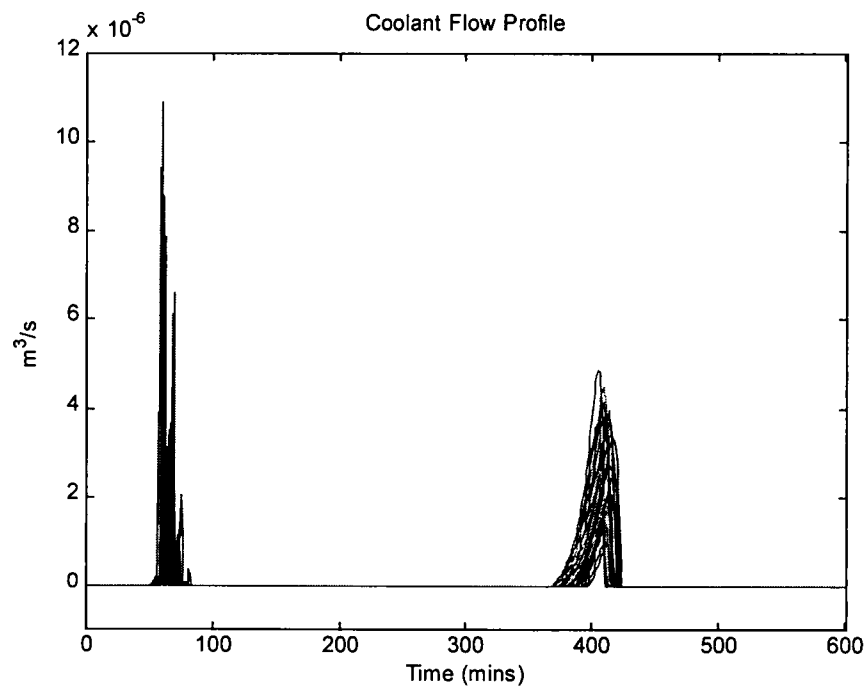


(b)

FIGURE 7.7: Nominal profiles generated from the simulation
(a) Reactor temperature
(b) Jacket temperature



(c)



(d)

FIGURE 7.7 (continued): Nominal profiles generated from the simulation

(c) Heater Power

(d) Coolant Flow

Chapter 8

Example Application

8.1 Introduction

In this chapter, the framework for SPC on batch profiles is applied to the data generated from the PMMA simulation described in the last chapter. The chapter represents the coming together of all the concepts described earlier into a coherent whole. The first section is devoted to the empirical model development exercise. Specifically, the time scale and magnitude scale parameters for the profiles are obtained along with a demonstration of batch monitoring using SPC charts. Results are also compared with the existing MPCA technique. Finally, a few words on the importance of manipulated variables for monitoring purposes are given. In the second section, correlation of the scale parameters with the final batch quality is studied. Throughout the chapter, the various issues are discussed at the appropriate place.

8.2 Profile Analysis

The complete data set consists of 100 batches with 96 nominal operation batches and 4 special cause batches. From the simulation description in the last chapter, profiles of the four measured variables – reactor temperature (T), cooling jacket temperature (T_j), heater power (P) and coolant flow (F_{cw}) are studied. These profiles have already been plotted in figure 7.8. The nominal operation data set is used to formulate an empirical model using time and magnitude scaling and obtain

the control limits for the control charts for SPC monitoring in the next few subsections.

8.2.1 Time Scaling

The time axes of the various batches can be put on a common scale by the alignment of various events/features that occur consistently from batch to batch. A simple technique for the extraction of such consistent events has been described in chapter 5. From the process description in the last chapter, it is known that no time scale variability exists in the loading and preheat phase of the process as the operation is automated so that the valves are opened and closed at fixed times after the beginning of the simulation. Similarly, the time at which the control system turns on for preheating is also fixed. Variability however exists in the time at which the catalyst is added so that the time of addition of catalyst can be used as a checkpoint. Assume that the DCS records the time so that it is available even though the exact time is not very obvious from the profiles.

Once the catalyst is added, reaction begins and time scale variability occurs since the reaction rates vary from batch to batch. However, since a measure of the reaction conversion is not available (a typical industrial scenario), the indicator variable approach is ruled out so that consistent events must be used for time scaling.

PMMA polymerization is characterized by a strong gel effect at high polymer conversion and results in high reaction rates. This is signified in the profiles by the jacket heater turning off and coolant flow turning on to remove the high heats of

reaction in the gel effect. Since the gel effect is a characteristic of the process, this heater shutting off event occurs consistently in all the profiles. The heater shut off time can thus be used as a consistent checkpoint for time scaling.

Once the gel effect sets in, the monomer is consumed at a faster rate so that at very high conversion when almost all of the monomer has polymerized, the reaction rate reduces. The heater power now turns on in order to keep the reactor at the reaction temperature set point. This is again a characteristic of the process so that the heater turn on event occurs consistently in all the profiles. It can therefore be used for time scaling.

The two checkpoints for time scaling are thus the heater shut off and subsequent heater turn on. Heater shut off indicates the beginning of autoacceleration while heater turn on indicates almost complete monomer conversion. In chapter 5, these two events were obtained using an automated algorithm. In this chapter therefore, the physical understanding aspect is emphasized.

The time of occurrence of these events can be easily obtained using the feature translation technique described in chapter 5. Of the several possible features that can be used for the purpose, the one that gives the least SSE upon translation gives the best recognition accuracy and is therefore used. The reader is referred to chapter 5 for the details. It must be pointed that by focussing only on the candidate features in the vicinity of the heater shut off and turn on events, the computational burden is considerably reduced since only a few relevant candidate features that are physically meaningful, are used from all the possible features.

As for events in the other profiles, suffices to say that the alignment of the features in the heater power profile results in the alignment of features in other profiles too. As an example, the heater turning on causes the jacket temperature to rise again so that aligning the heater turn on should result in the jacket temperature valleys to be aligned. The heater profile events are used for time alignment since they are sharp and therefore can be obtained accurately using the translation technique of chapter 5.

Three events are therefore used for time scaling – 1. Initiator addition, 2. Heater shut off and 3. Heater turning on. There are thus four (3+1) distinct phases from the starting to the end of a batch. Once the times of occurrence of the time scale events is known, linear interpolation or padding must be used so that each phase in the various batches is of the same length. The various phases and the method used for time scaling each of the phases are tabulated in table 8.1. The raw and time scaled profiles for each of the measurements are plotted in figure 8.1.

TABLE 8.1: Time scaling zones and the method for each zone

<i>TIME ZONE</i>	<i>TIME SCALING TECHNIQUE</i>
1 – 60 min	Pad Mean value
61 – 390 min	Linear Interpolation
391 – 426 min	Linear interpolation
427 – 601 min	Pad Last Available Value

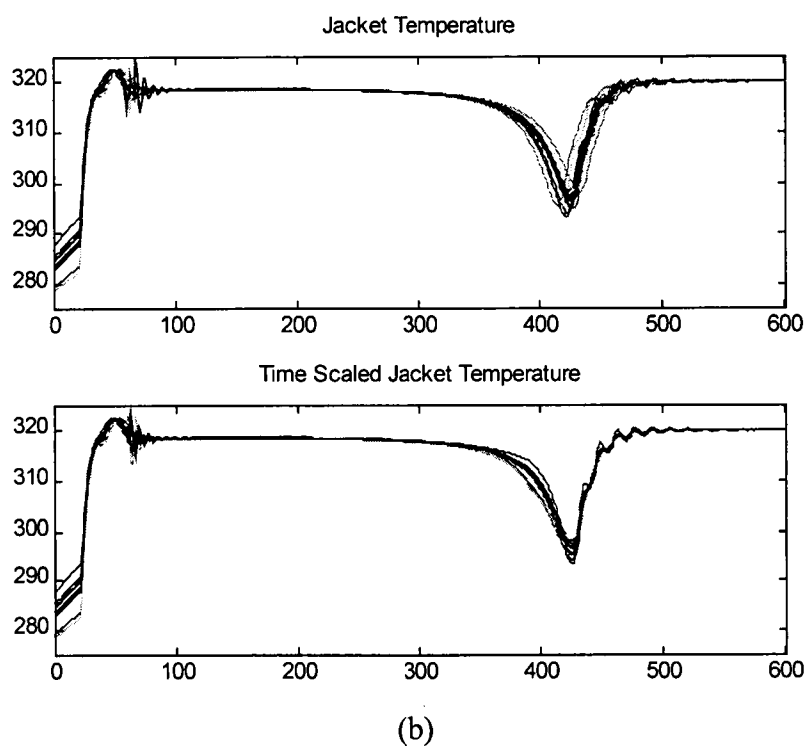
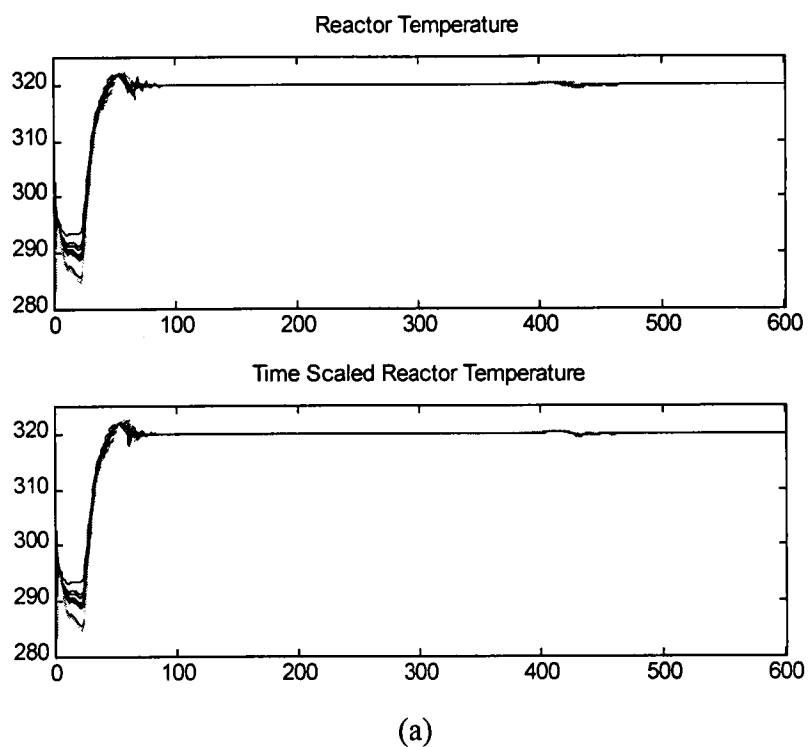


FIGURE 8.1: Effect of time scaling. Raw and time scaled profiles
(a) Reactor temperature
(b) Jacket temperature

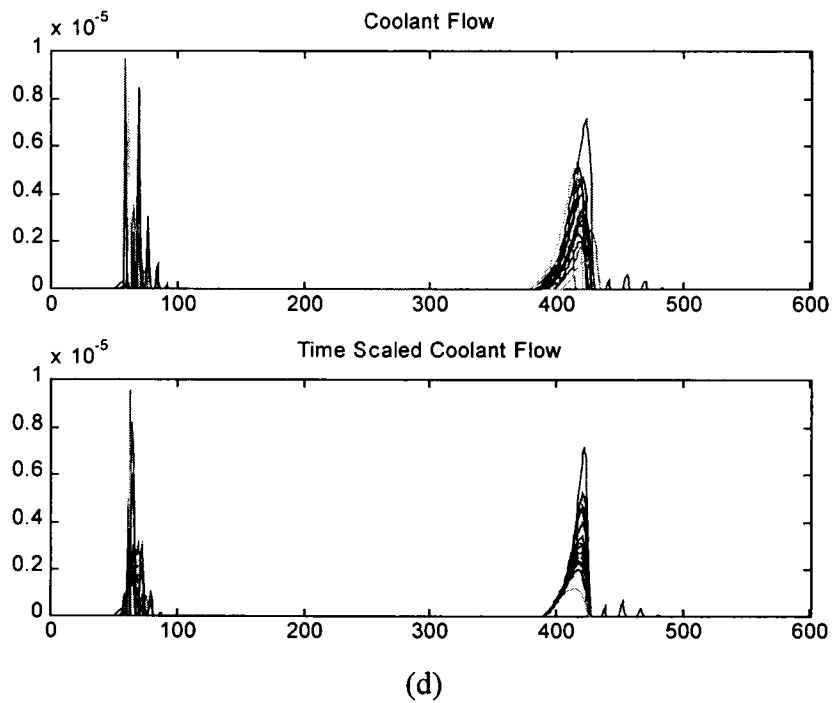
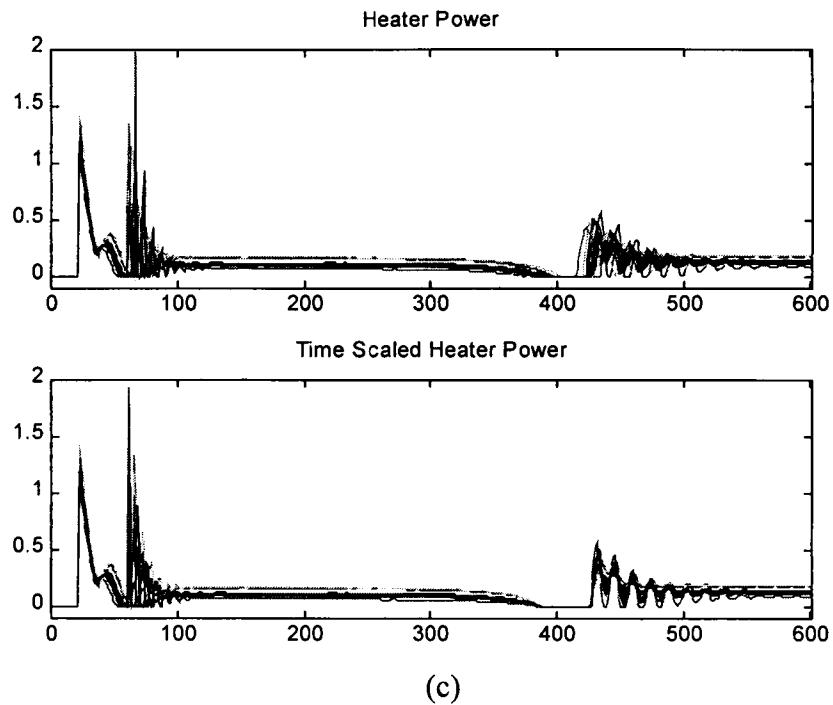


FIGURE 8.1 (continued): Effect of time scaling. Raw and time scaled profiles
 (c) Heater power
 (d) Coolant flow

From the figure, it is seen that time scaling explains a significant amount of the total variability in the profiles. As a quantitative measure, the MSE of the profiles before and after time scaling is plotted in figure 8.2 for each of the measured variables. In the figure, the MSE is obtained after mean centering the profiles. Thus if $\mathbf{X}_k$ denotes the mean centered profile matrix, raw or time scaled, where each row is a profile, the MSE is obtained as

$$\text{MSE} = \text{trace}(\text{diag}(\mathbf{X}_k \mathbf{X}_k^T)) / (I.J) \quad \dots\dots\dots(8.1)$$

where I is the number of rows and J is the number of time points in the batches. In order to view the results on the same plot, the variance explained by time scaling is expressed as a percentage of the total variance. The amount of data that is lost due to time scaling is assumed negligible in comparison to the overall MSE of the profiles.

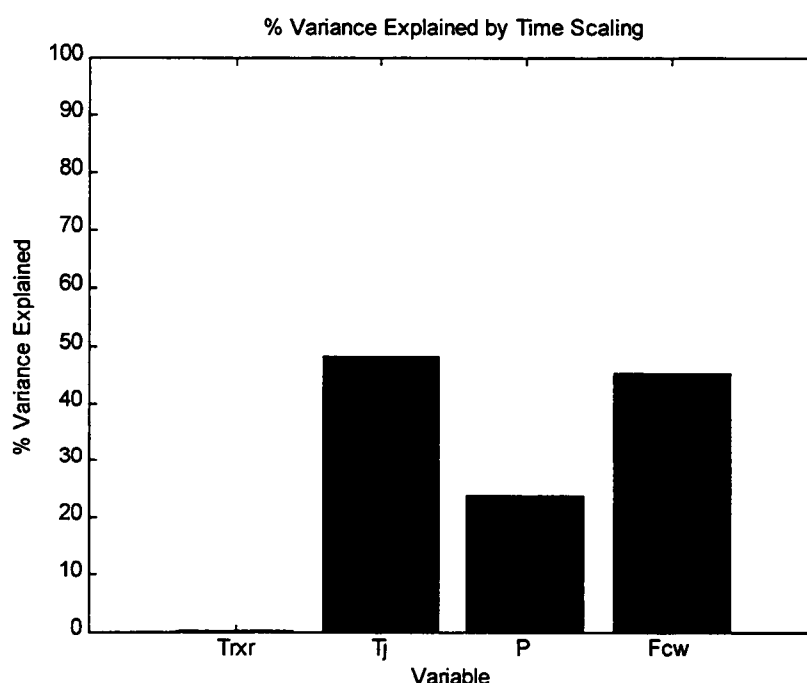


FIGURE 8.2: Percentage variance explained by time scaling

The bar plot shows that more than 40% of the variance in the jacket temperature and coolant flow profiles is accounted for by time scaling. Also about 25% of the overall variability in the heater profile is explained by time scaling. It is therefore clear that time scaling accounts for a significant amount of variability in the batch profiles underscoring its importance in the analysis of any batch process historical database.

8.2.2 Magnitude Scaling

Once the profiles are time scaled, the consistent variability along the measurement axis is modeled by magnitude scaling. In chapter 6, a simple evolving factor analysis (EFA) type algorithm was developed in order to extract the factors for magnitude scaling. In this chapter, for the sake of completeness, the use of 'custom' scale parameters is demonstrated. The factors so obtained are very similar to those obtained by the EFA approach referred to earlier. This again demonstrates the fact that a simple look at the profiles coupled with a physical understanding of the process can provide valuable insight about the type of variability in the profiles, thus simplifying the extraction of the scale factors.

The basic approach for extracting the magnitude scale factors remains the same as was described in chapter 5 – that of finding time zones (or phases) that show a particular type of variability occurring consistently across all profiles, e.g. a shift. In case the variability is of a particular type, a single PC would explain most of the variability in the zone. In this section, rather than using the EFA approach to extract

such zones, the profiles are directly divided into zones that contain variability of a particular type and the factors are directly obtained using the singular value decomposition (SVD) to extract the factors. The division into zones is based on the user's judgement. In most cases, such a division is obvious from the data itself so that the subjectivity is not a major disadvantage. The significant reduction in the computational burden when compared with the EFA approach makes the use of such 'custom' factors very attractive in most cases. The factors obtained for each of the measured variables are plotted in figure 8.3. Table 8.2 tabulates the number of factors for each variable along with the time zones used. A physical description of the factor is also provided. Note that each of the factors is very easy to interpret.

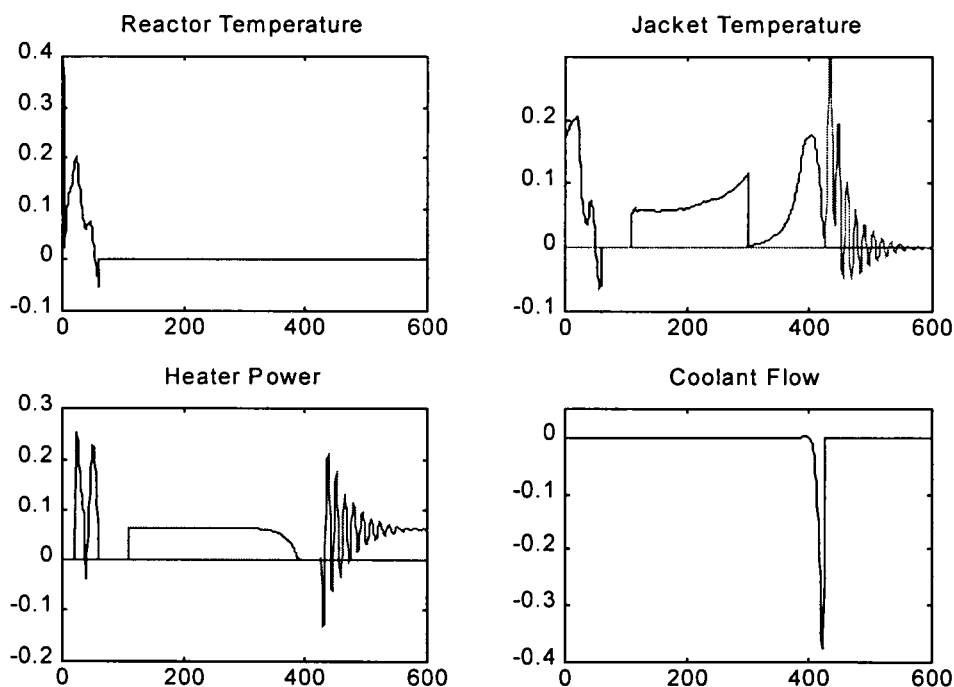


FIGURE 8.3: Factors used for magnitude scaling for all the measured variables

TABLE 8.2: Factors for each of the profiles

<i>Variable</i>	<i># of Factors</i>	<i>Factor Time Zones</i>	<i>Factor Description</i>
Reactor Temperature	1	0 – 60min	Startup Dynamics
Jacket Temperature	4	0 – 60min 110 – 300min 310 – 426min 427 – 600min	Startup Dynamics Slope Valley Oscillations
Heater Power	3	1 – 60min 110 – 400min 426 – 600min	Startup Dynamics Level & Slope Oscillations & Level
Coolant Flow	1	390 – 430min	Flow Overshoot

The projection methodology discussed in chapter 6 is now applied to extract the scale parameters. The mean centered profiles and the residuals remaining after projection for a few representative batches is plotted in figure 8.4. The reduction in the variability in the time zones corresponds directly to the factors in figure 8.3. As with time scaling, a significant amount of variability is explained by magnitude scaling. Figure 8.5 plots the percentage of variance explained by both time and magnitude scaling. Let MSE_o , MSE_{ts} and MSE_{tsms} denote the mean square errors for the raw, time scaled and both time and magnitude scaled profiles, respectively as given by equation 7.1. Then the percentage variance explained by time scaling is obtained as

$$\%_{ts} = 100.(MSE_o - MSE_{ts})/MSE_o \quad \dots\dots\dots(8.2a)$$

while the percentage variance explained by magnitude scaling is obtained as

$$\%_{ms} = 100.(MSE_{ts} - MSE_{tsms})/MSE_o \quad \dots\dots\dots(8.2b)$$

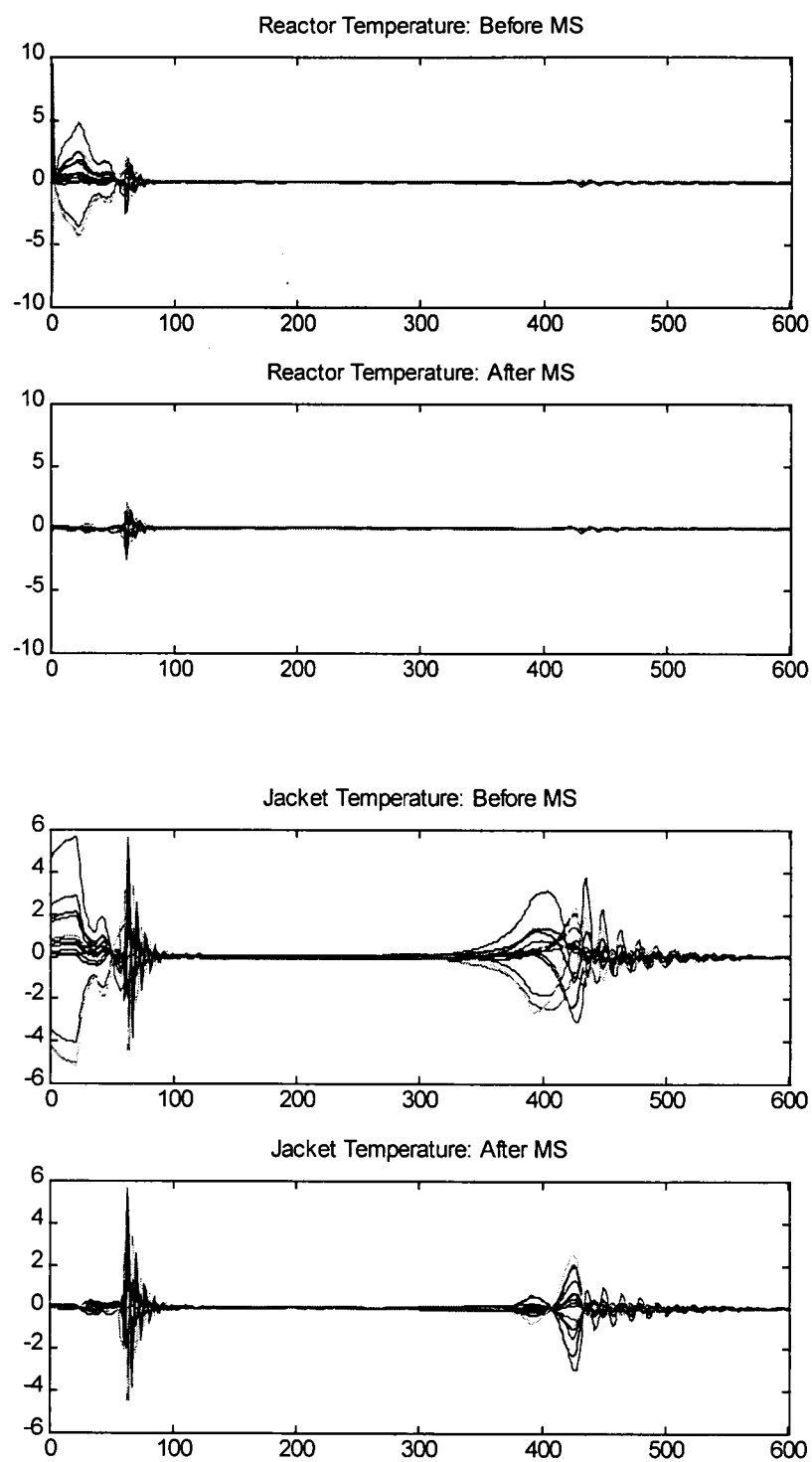


FIGURE 8.4: Effect of magnitude scaling. Mean centered profiles and residuals remaining.

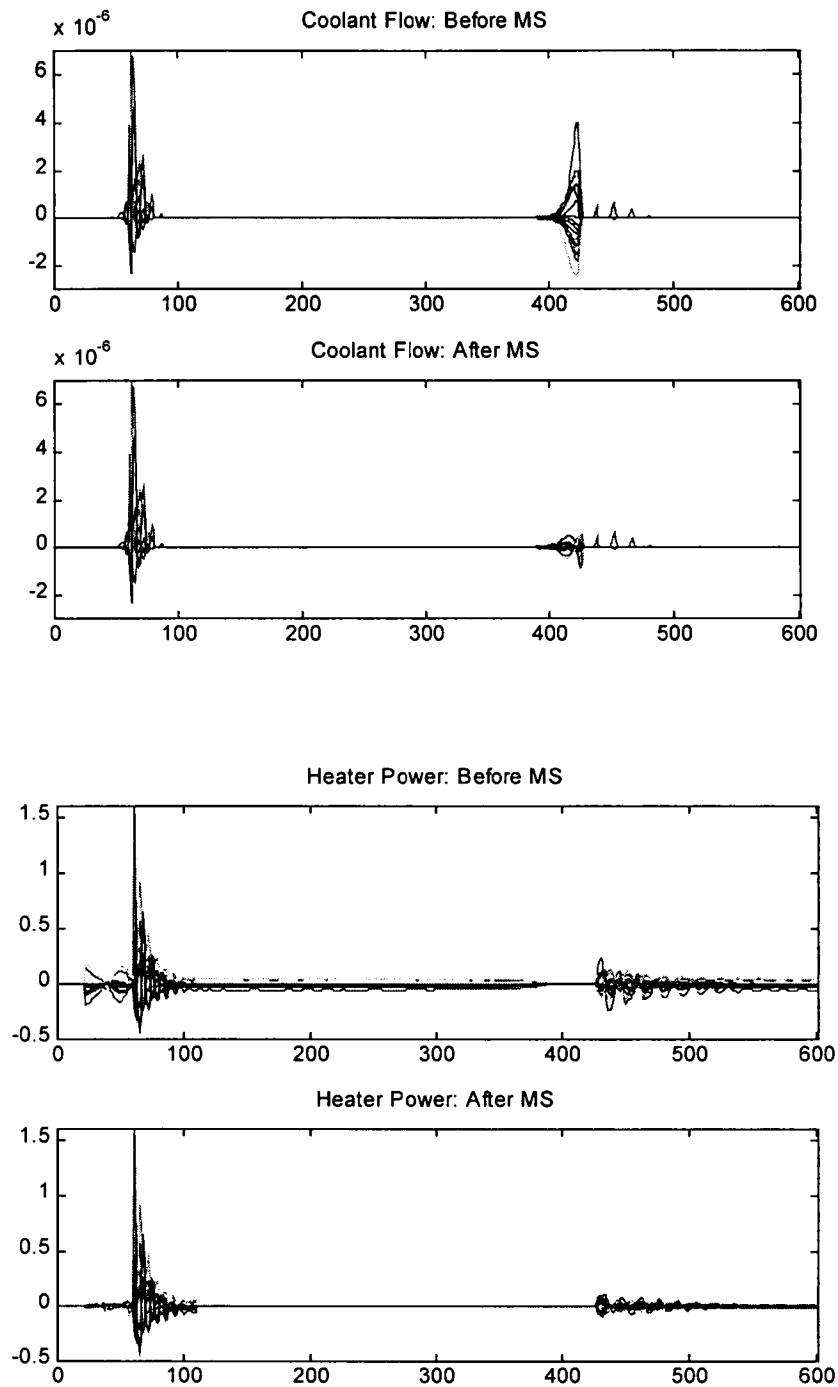


FIGURE 8.4 (continued): Effect of magnitude scaling. Mean centered profiles and residuals remaining.

From the figure, it is seen that magnitude scaling explains almost 90% of the variability in the reactor temperature profile. The approximate figures for the jacket temperature, heater power and coolant flow profiles are 50, 30 and 20 percent respectively. Thus magnitude scaling accounts for a significant amount of the total variability in the profiles. The variance explained by time scaling and magnitude scaling for each of the measured variables is plotted in figure 8.5. For the reactor temperature and jacket temperature profiles, the total variance explained is more than 90%. For the other profiles, the unexplained variance consists mostly of the oscillations. In the next section, the use of SPC charts on the scale parameters and the residuals remaining for monitoring purposes is studied.

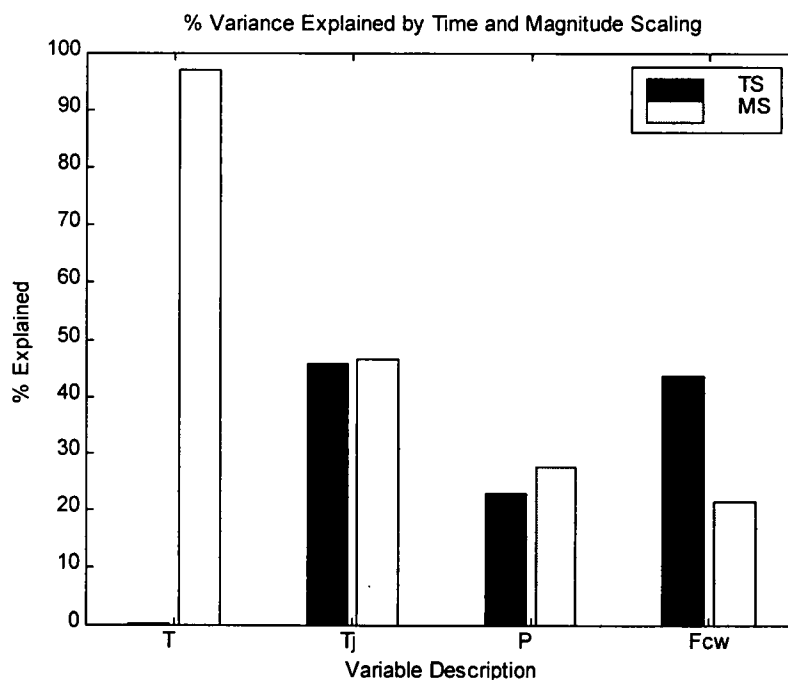


FIGURE 8.5: Percentage variance explained by time and magnitude scaling

8.2.3 SPC Monitoring Charts

SPC charts on the scale parameters and the residuals remaining after scaling were proposed in chapter 6. These charts can be used for monitoring the batch profiles for any abnormal operation. Four special cause disturbance scenarios are used to study the monitoring capabilities – 1. The dissolved oxygen concentration increases (batch 97), 2. The cooling water temperature increases (batch 98), 3. The heater efficiency decreases (batch 99) and 4. The jacket heat transfer coefficient decreases (batch 100). These scenarios were described in chapter 7. Each disturbance is simulated to occur at time 170min after the start of the simulation as a ramp that saturates to a final value.

First of all SPC charts for the time scale parameters are studied. Figure 8.6 plots the duration of each of the four phases for the nominal batches and the four special cause batches. The special cause batches are plotted as a different color. Note that all the time scale parameters appear to be within the 3σ limits except for batch 97 that has excessively long second and third phases. This is understandable since in the batch, the dissolved oxygen content in the reaction mixture increases midway so that the overall reaction rate slows down resulting in the longer phase lengths. The abnormal time scale parameters point to this slowing down and provide vital clues for assigning a physical cause for the abnormality. In the present case, the excessively long phase duration points to stoichiometric problems.

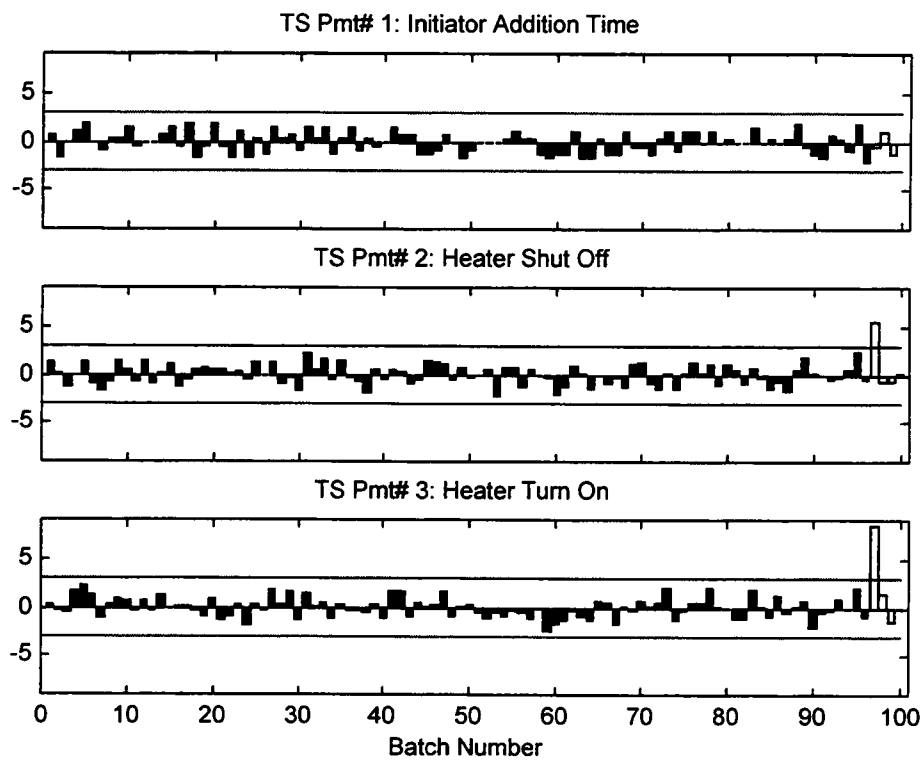


FIGURE 8.6: Univariate SPC charts on the time scale parameters with 3σ limits

Univariate SPC charts on the magnitude scale parameters are studied next. These are plotted in figure 8.7 for each of the scale parameters. SPC charts for the 2nd, 3rd and 4th factors on the jacket temperature show violations for batch 100 which simulates an abnormally low heat transfer coefficient that occurs midway through the batch. This is again easily interpreted as the jacket temperature decreases abnormally to remove the same amount of heat so that abnormally negative values are observed for the 2nd and 3rd scale parameters. The 4th scale parameter captures the excessive oscillations that occur due to the aggressive controller action. Thus, each of the SPC violations is easily interpretable. Batch 98 shows a violation on the 4th factor of the jacket temperature profile and is interpreted again as capturing the excessive oscillations occurring towards the end of the batch. These oscillations occur since the coolant temperature increases midway through the batch causing the aggressive controller to be oscillatory in the post gel-effect phase.

The SPC charts on the magnitude scale parameters for the heater profile show no violations. The SPC chart on the MS parameter for the coolant flow profile shows that batch 98 that simulates abnormally high cooling water temperature midway through the batch shows a control limit violation. This is again interpreted very easily since more coolant needs to be added in order to decrease the jacket temperature. This shows that the magnitude and time scale parameters that are proposed with a distinct interpretation provide vital physical clues for diagnosing an abnormal operation in a batch.

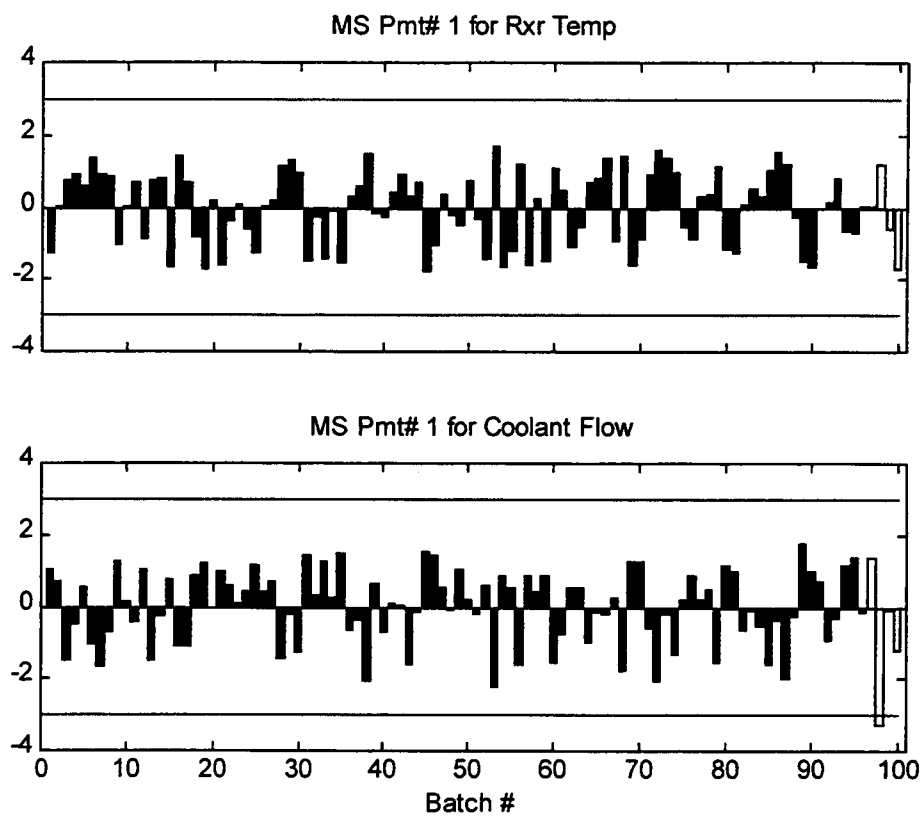


FIGURE 8.7: Univariate SPC charts on the magnitude scale parameters with 3σ limits.

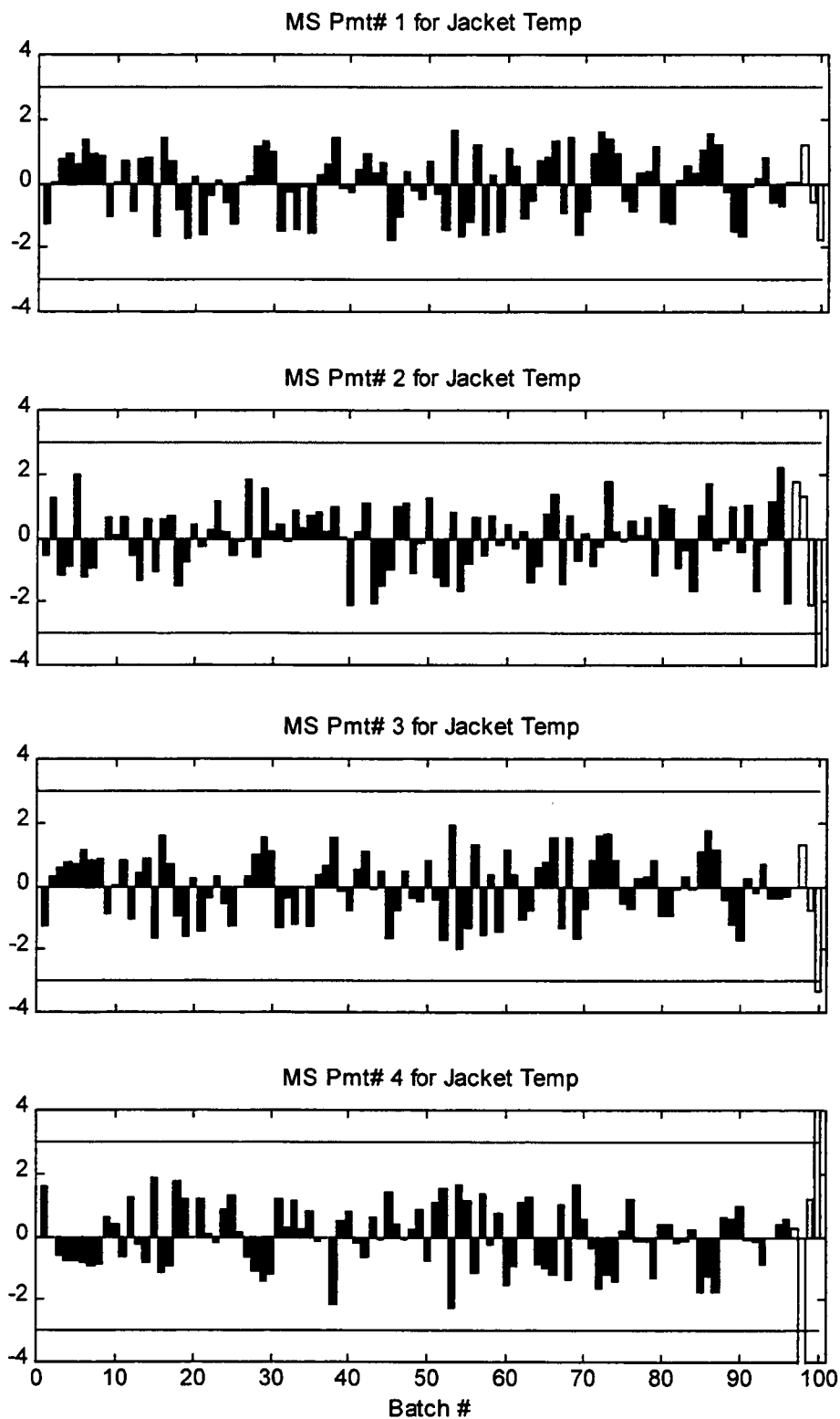


FIGURE 8.7 (continued): Univariate SPC charts on the magnitude scale parameters with 3σ limits.

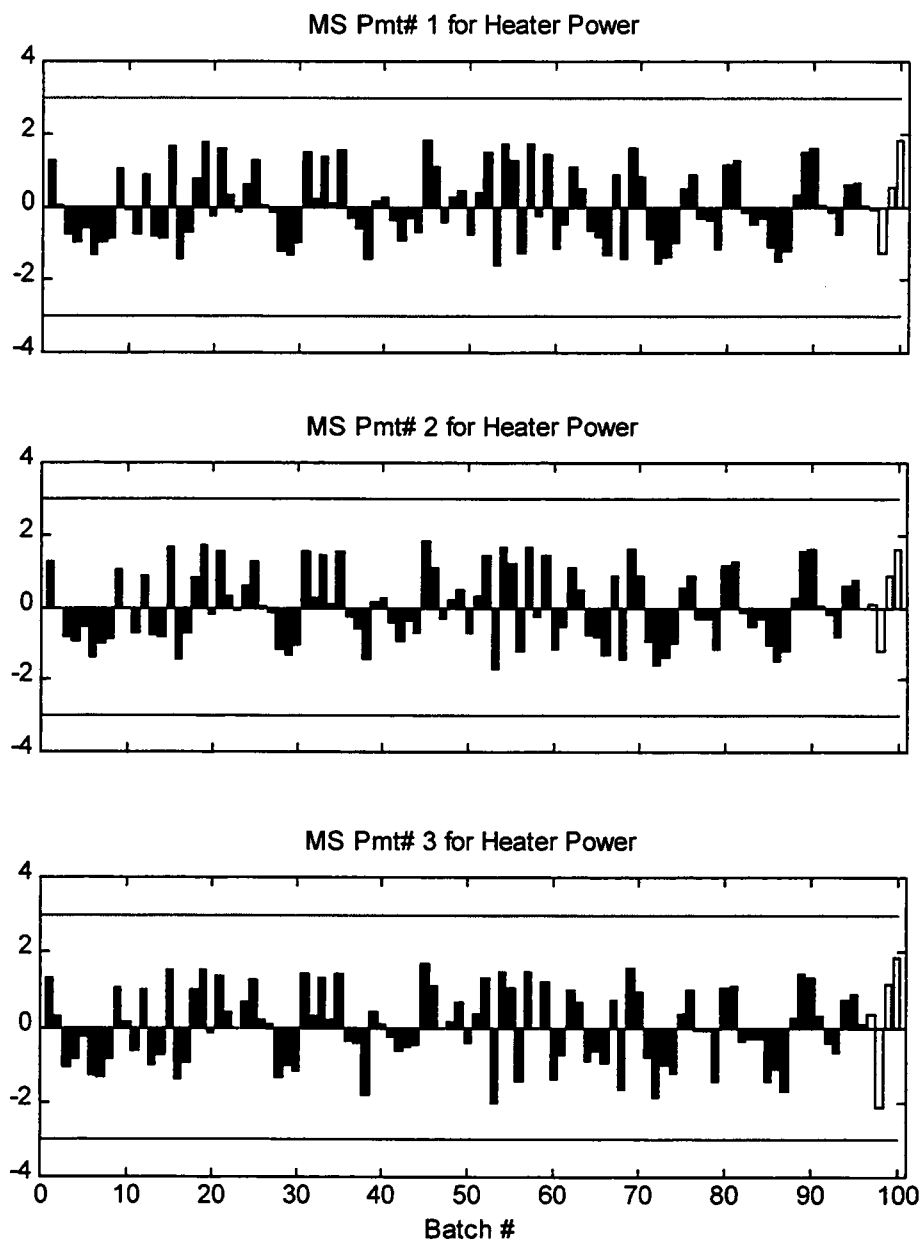


FIGURE 8.7 (continued): Univariate SPC charts on the magnitude scale parameters with 3σ limits

Even though an exact cause may not be assignable, distinct classes of problems such as stoichiometry problems or heat transfer problems can nevertheless be distinguished from the sets of scale parameters which show a violation on the SPC chart. Typically, in the industry, it is usually a single problem that causes operational or quality problems. The methodology then provides a direct means for boiling down the particular area of concern. Subsequent process improvement would then be based on a physical understanding of the process.

Finally, SPC charts on the Q statistic that indicates the lack of fit from the empirical model are plotted in figure 8.8. Note that all of the special cause batches show distinct violations of the 99.7% confidence limits.

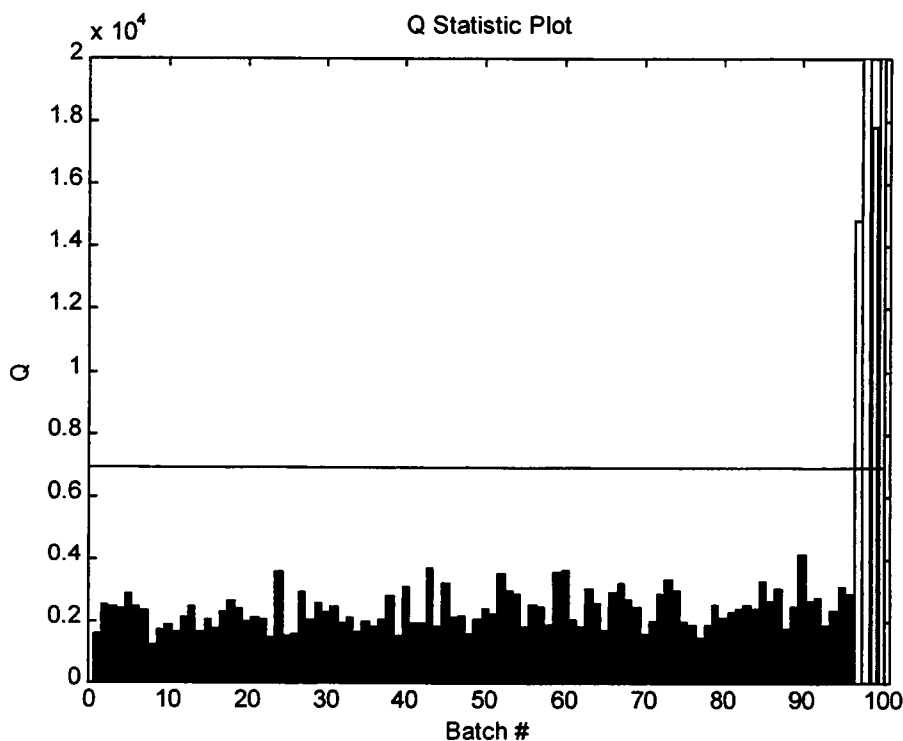


FIGURE 8.8: Q statistic bar plot for all the batches with 99.7% confidence limit

These violations must now be interrogated as to why the Q statistic is out of control. In order to do so, figure 8.9 plots the scaled residuals for the special cause batches. A few nominal operation residuals are also plotted for comparison purposes. Note that for all the special cause batches, the disturbance is introduced at 170min. From the figure, severe residuals are seen after 170min for batches 97, 99 and 100, which cause the Q statistic to be out of control. Note that batch 99, which simulates heater efficiency decrease also shows control limit violations. No violations were seen in the scale parameter charts for the batch. Batch 98 that simulates abnormally high coolant temperature also shows a control limit violation. However, the violation is seen occurring only in the post gel effect phase when coolant is pumped into the jacket. Earlier on, no coolant is pumped into the system and the coolant is simply recirculating so that the disturbance does not enter the system till the coolant flow starts.

The SPC charts show that the Q statistic complements the scale parameter charts in detecting abnormal operation. All the variability that is not detected in the scale parameters remains in the residuals and shows up in the Q statistic. They are thus complimentary to each other and together monitor the complete profile space. A summary of all the SPC charts for the abnormal batches is provided in table 8.3. Notice from the table that all the batches show an abnormally high Q statistic implying high residuals remaining after scaling. This completes the demonstration of the SPC monitoring methodology. The next section applies MPCA to the batch data set for comparison purposes.

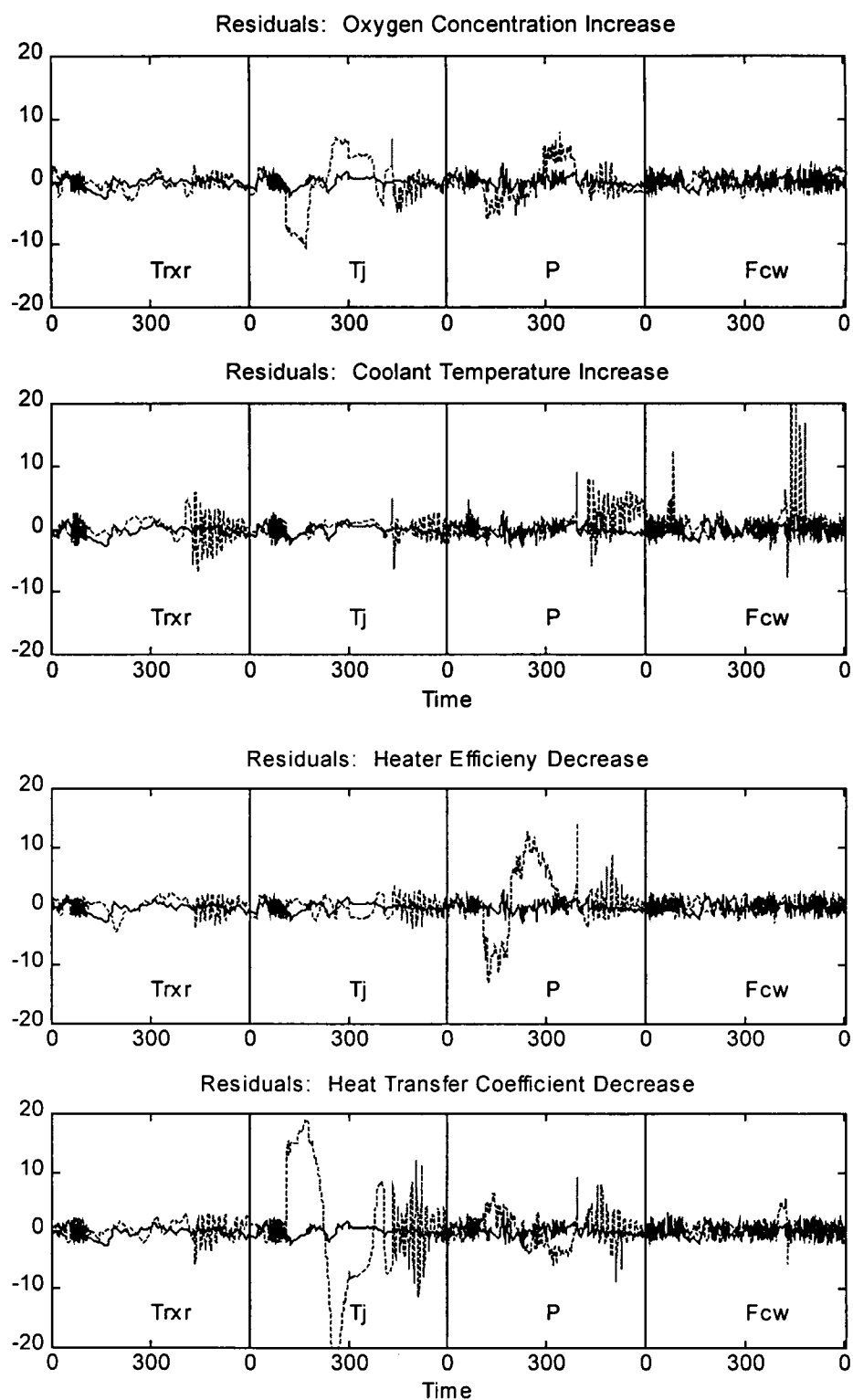


FIGURE 8.9: Residuals for the abnormal batches. Solid lines are for a normal batch. Dotted lines are for the abnormal batch

TABLE 8.3: SPC chart violations for the time and magnitude scale parameters and the Q statistic

Parameters		Batch 97	Batch 98	Batch99	Batch 100
TS Pmts	Initiator				
	Htr Off	+			
	Htr On	+			
MS Pmts	Trxr				
	Tj ₁				
	Tj ₂				+
	Tj ₃				+
	Tj ₄		+		+
	P ₁				
	P ₂				
	P ₃				
	Fcw		+		
Residuals (Q)		+	+	+	+

8.2.4 Comparison with MPCA

MPCA is applied to the batch data set as described in references [40, 41]. The 3D historical batch data array is unfolded into a 2D matrix and the nominal operation batches are autoscaled. Upon unfolding, the 2D matrix $\mathbf{X}$ consists of I batches as the rows. The total number of columns is $J.K$ where K is the number of measured variables while J is the total number of time points. For the present case $J = 601$ while $K = 4$. In order to reduce the computation load in calculating the principal components, alternate time points are taken to form the $\mathbf{X}$ matrix so that in all, the $\mathbf{X}$ matrix is 96 by 1204 (301x4). Note that the time scaled profiles are used in order to form the $\mathbf{X}$ matrix.

The scree plot for the PCA is plotted in figure 8.10. The first two 2 PCs are considered significant since they explain about 50% of the total variability and also since the 'knee' in the scree plot seems to occur at 2 PCs. The two PC loadings are plotted in figure 8.12. Demarcation lines between the loadings for the four measured variables are also plotted. Note the difficulty in correctly interpreting the meaning of the loadings.

The magnitude scale parameters consist of the projections on the PCs and are also known as scores. SPC charts on the two scores for all the batches are plotted in figure 8.12. Note from the figure that the scores on the first PC for all the normal batches are within the control limits. The scores on the second PC are also within limits for all the batches except for the low heat transfer coefficient case (last batch). However, an interpretation of the abnormally low score is not very straightforward from the PC loadings. Indeed, considerable effort is needed in correctly interpreting the abnormal scores. Researchers have proposed the contribution plot method to point out the variables that 'numerically' contribute to the out of control scores and Q statistic [50]. Notwithstanding such 'numerical' techniques, the difficulty in interpreting the scores in MPCA based techniques always exists thus limiting its use. On the other hand, the approach as has been proposed in this article, is geared towards giving factor loadings that are easy to interpret and connect to the actual process variability.

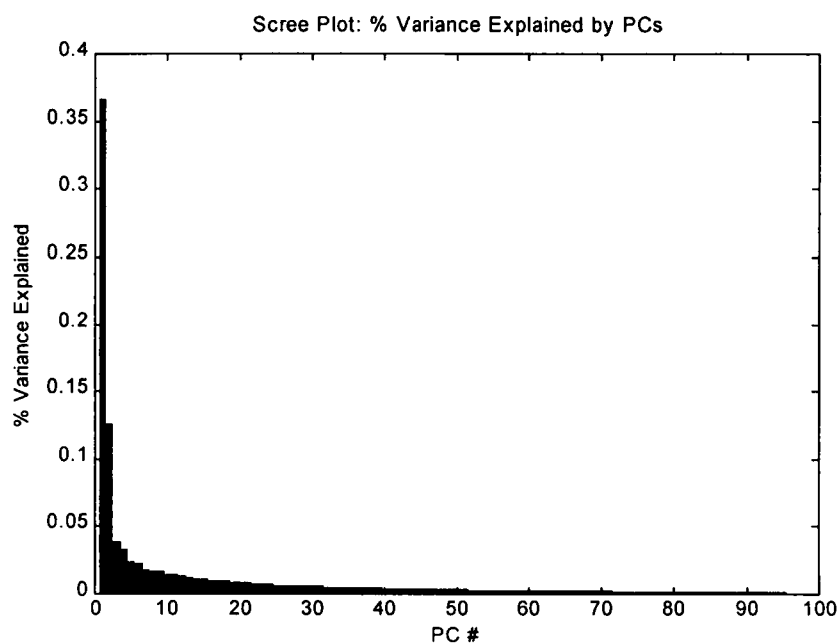


FIGURE 8.10: Scree plot for MPCA

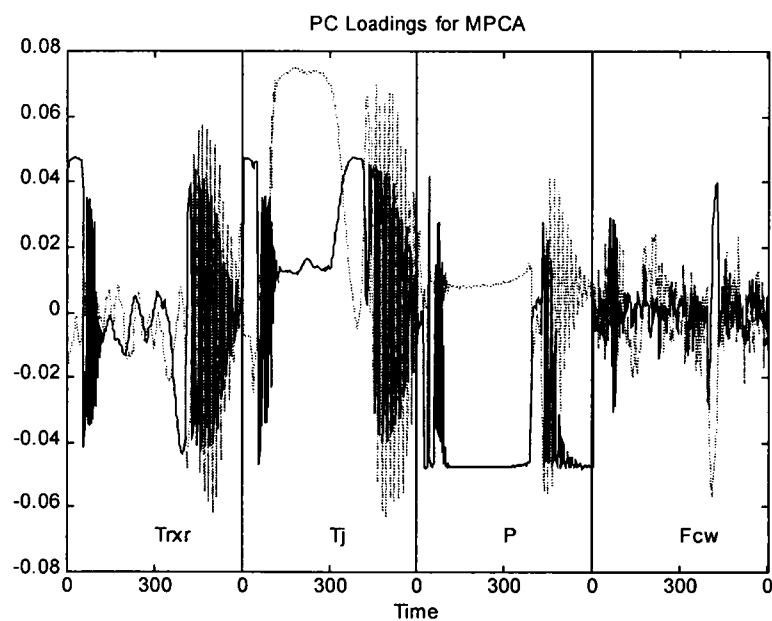


FIGURE 8.12: Retained PC loadings for MPCA
Solid: First PC
Dotted: Second PC

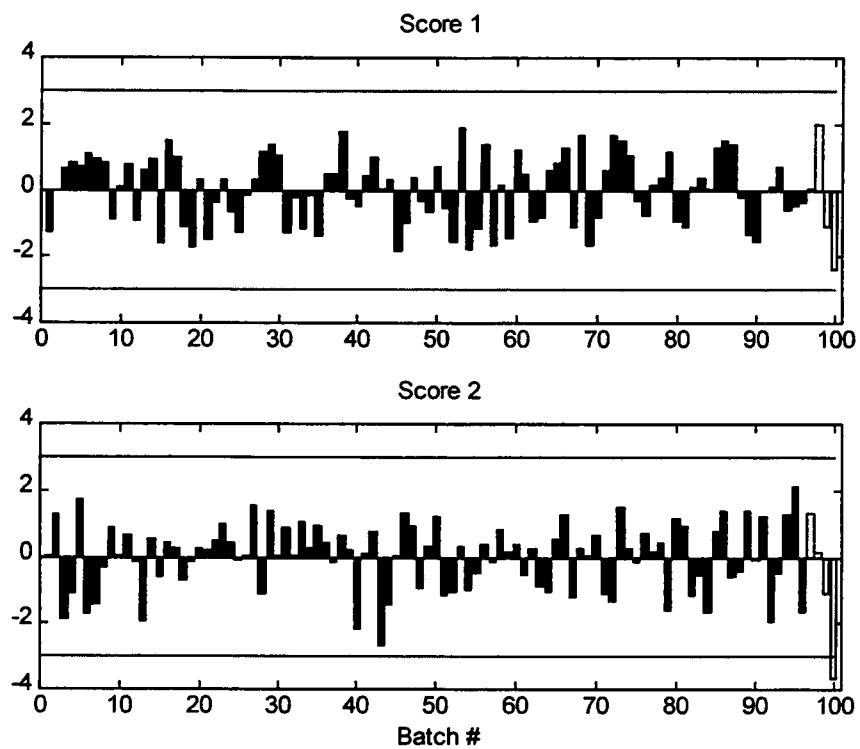


FIGURE 8.12: Univariate SPC charts on scores for MPCA

SPC charts on the Q statistic are plotted in figure 8.13. violations of the 99.7% confidence limits are observed for the Q statistic for all the abnormal batches as seen in figure 8.13. As in the previous subsection, figure 8.14 plots the residuals for all the special cause batches along with the residuals of a nominal batch for comparison. Note that for the same disturbance scenario, the control limit violations on the Q plot are not as severe as was seen in the earlier study especially for batches 97 and 99.

Batch 99 that simulates abnormally low heater efficiency, is worth further discussion. Note that most of the contribution to the high Q statistic comes due to the oscillations towards the end of the batch as seen from figure 8.14 (third subplot). A comparison with figure 8.9 that plots the residuals remaining for the scaling methodology proposed in this work, shows that a significant contribution to the Q statistic comes from time points much earlier in the batch, especially from the heater power profiles. In the simulation, the heater efficiency was reduced midway through the batch at about 170min so that significant residuals remain after scaling around time 170min in the heater power profiles which is not seen in the MPCA case. This shows the enhanced sensitivity of our proposed scaling methodology in detecting extremely subtle deviations from nominal operation.

The trick lies in autoscaling the residuals *after* the consistent variability is extracted in the scale parameters. MPCA, on the other hand, relies on autoscaling before extracting the systematic variability. Autoscaling after extraction of the consistent variability implies that the residuals for the scaled profiles would be smaller than the original unscaled profiles.

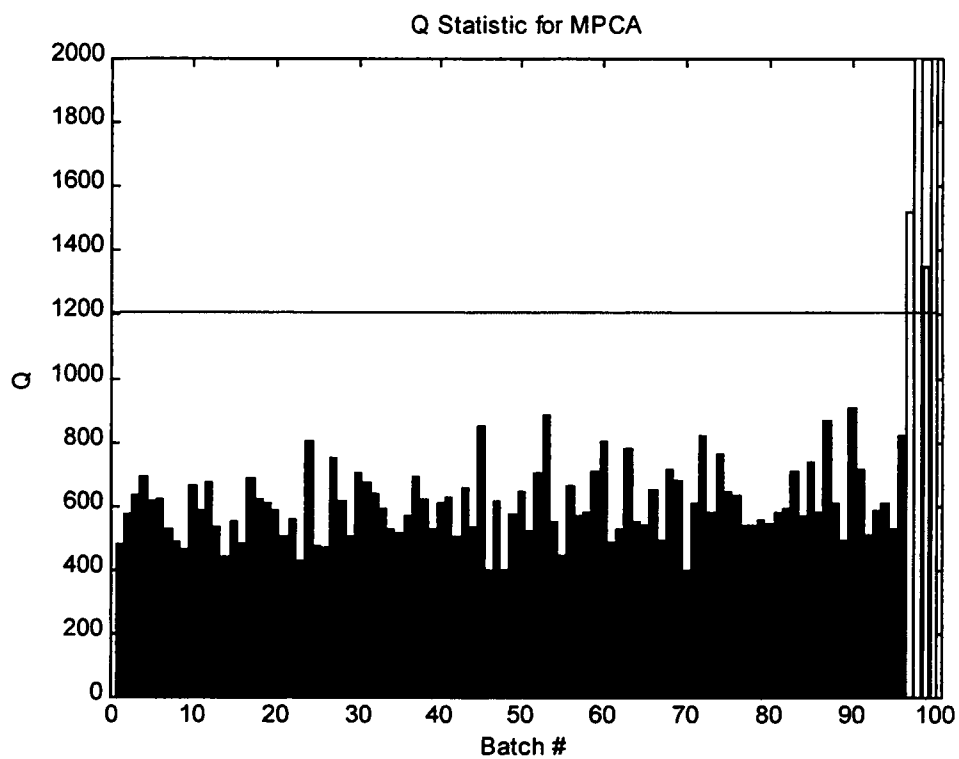


FIGURE 8.13: Q statistic plot for MPCA

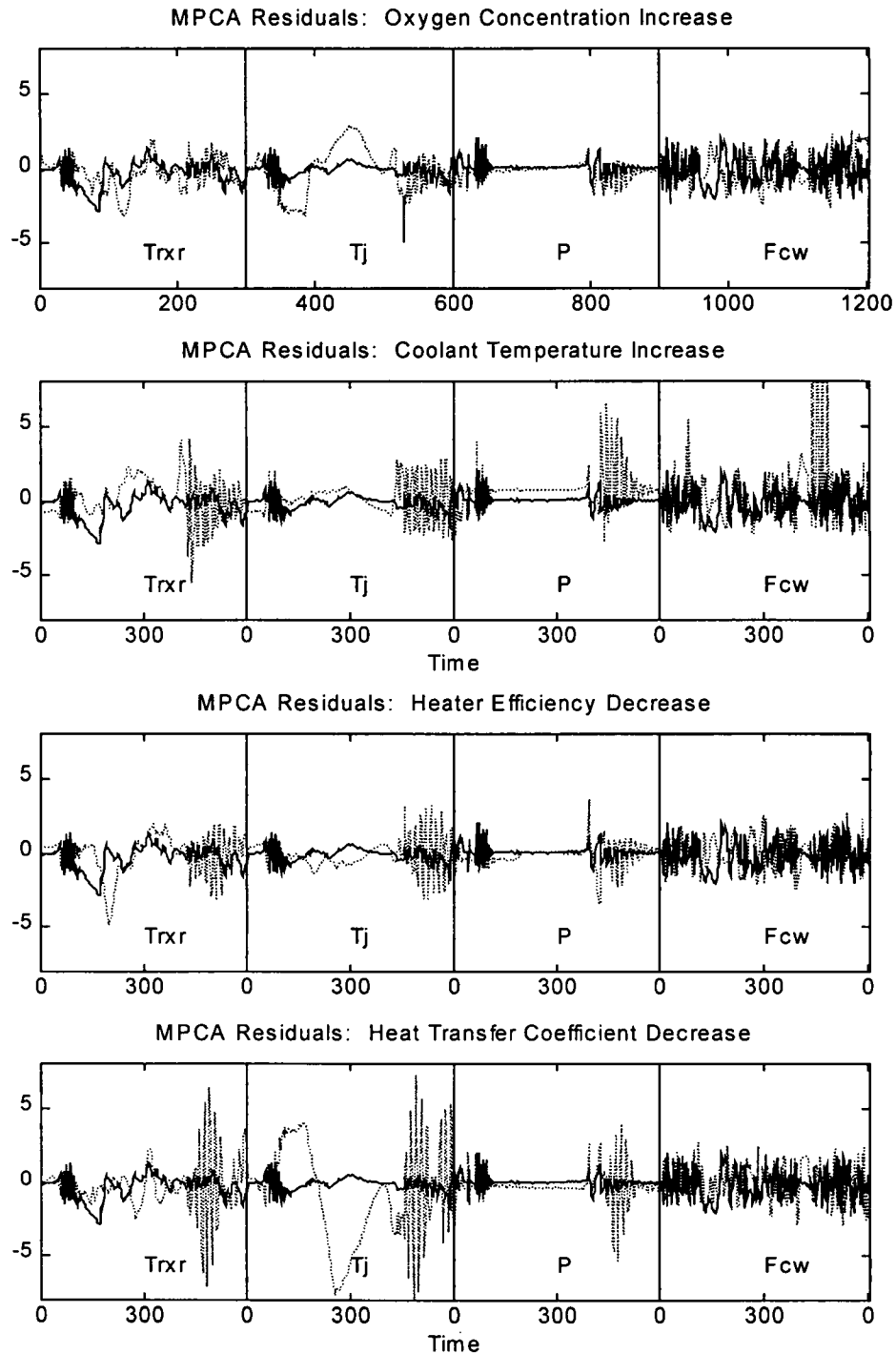


FIGURE 8.14: Residuals for abnormal batches

Thus, autoscaling would punish any high residuals much more severely than if the consistent variability was not extracted. This leads to higher sensitivity in the detection of special causes as seen in batch 99.

Scaling the profiles is thus of paramount importance for monitoring purposes. As a corroboration of the fact, an MPCA was performed on the unfolded X matrix by standardizing the four measured profiles with the overall mean standard deviation of the complete profile rather than standardizing each time point as in autoscaling. Results showed that none of the special cause batches were detected on the SPC charts.

Critics may feel that the above discussion on scaling is of minor significance for practical industrial monitoring applications. However, the real crunch comes in the well-understood scale parameters in our approach implying very easy interpretation of the results characterizing the batch. Thus diagnosing an abnormal batch for a physical reason is much easier since the abnormal scale parameters provide vital physical clues.

As an example, consider batch 98 and 100. Batch 98 has an abnormally high coolant temperature while batch 100 has an abnormally low heat transfer coefficient. The magnitude scale parameters for both the batches along with the 3σ limits are plotted in figure 8.15. Note that batch 98 shows a scale parameter violation for the coolant flow in addition to a scale parameter violation on the 4th magnitude scale parameter on the jacket temperature while batch 100 shows scale parameter violations for factors 2, 3 and 4, associated with the jacket temperature profile.

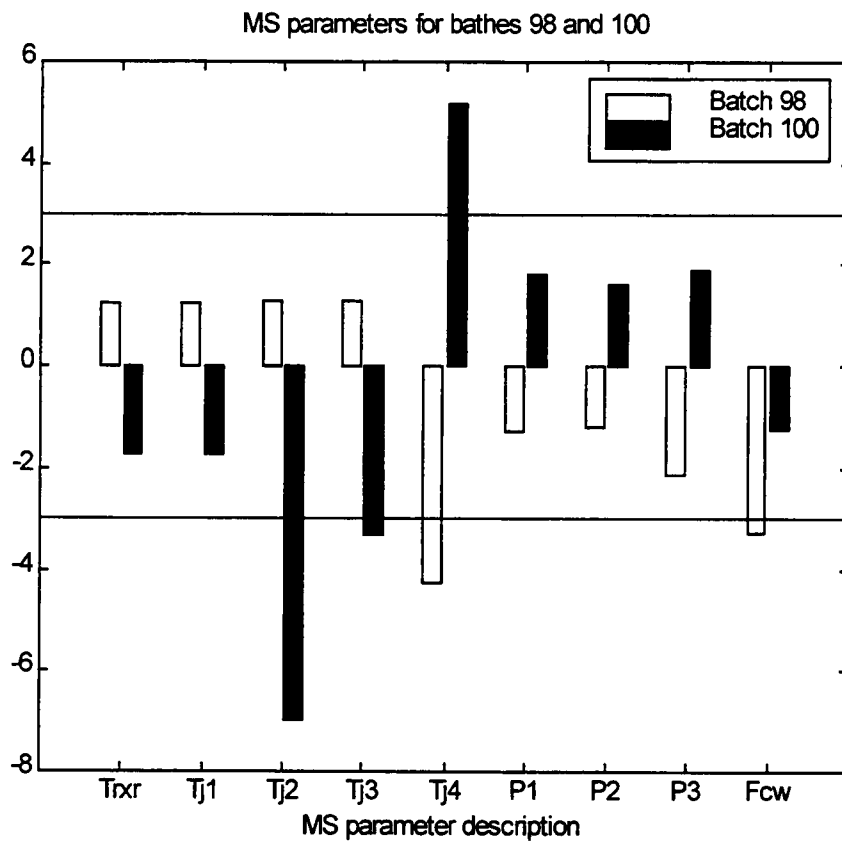


FIGURE 8.15: Magnitude scale parameters for batches 98 and 100
 Batch 98 Abnormal coolant temperature
 Batch 100 Abnormal jacket heat transfer coefficient

For batch 98, more than usual coolant flow is needed to remove the reaction heat as in normal operation indicating that the coolant may be hotter than usual. For batch 100, the SPC violations on the jacket temperature indicate that the jacket temperature is lower than usual so that either less than usual heat is being generated in the reaction or the heat transfer characteristics of the equipment may have changed. The former option is ruled out since the time scale parameters are within statistical control limits indicating usual reaction rate. The well understood scale parameters coupled with physical reasoning facilitates distinguishing and isolating the different special causes for the two batches. This is a very desirable feature and is a vital advantage since the approach enhances process understanding, the main goal of undertaking the analysis of the historical database. Note that a simple interpretation is not very easily possible from MPCA results that analyzes all the profiles together. Indeed, our approach retains the power of projection methods for the detection of subtle special causes along with giving simple and physically meaningful scale parameters.

8.2.5 Manipulated Variables

In the last few subsections, the application of the SPC framework for monitoring purposes has been demonstrated. Various disturbance scenarios were studied and a special mention must be made of manipulated variables. In any process, the manipulated variables are used for rejecting disturbances that enter the process in order to keep the controlled variable at its set point. Thus any disturbance (special

cause in SPC jargon) that enters the process, is compensated for by the control system by making adequate adjustments to the manipulated variables. From the transformation of variability standpoint, almost all of the variability in the form of disturbances is transformed to the manipulated variables by a good control system. Thus for the characterization of variability in a process, manipulated variables are very important since they contain most of the information about the disturbances or special causes. This fact is true of all controlled processes – batch or continuous.

In the present batch example, the reactor temperature is the controlled variable, while the heater power and coolant flow, are the manipulated variables. The jacket temperature is also a manipulated variable in the traditional cascade control sense. It is seen in all the disturbance scenarios that only the scale parameters on the manipulated variables show SPC violations. Also the main contribution to the Q statistic for the residuals comes from the residuals on these manipulated variables only. The contribution of the residuals for the controlled variable (reactor temperature) is minimal in all the cases. This is amply clear from figure 8.9 for the residuals of the abnormal batches. It can also be verified for the MPCA case from figure 8.14 for the abnormal batches. Thus, no matter what technique is used for magnitude scaling, manipulated variables are of paramount importance for characterizing the variability in the process.

Considering the above, for most processes with a reasonably good control system, the characterization of the variability in the manipulated variables would lead to a characterization of the variability in the process. It is with this forethought that it

was felt that MPCA may not be a good way of analyzing the variability in batch processes since all the variables are analyzed together. Further, autoscaling would unnecessarily complicate matters by giving more weight to the controlled variable since variability in the controlled variable is less while less weight is given to the manipulated variables. The proposed approach thus analyzes each of the variables separately. The resulting ease in interpreting the results due to the simplified approach has already been demonstrated earlier.

This completes the demonstration of the proposed SPC methodology with comparison with the existing MPCA monitoring technique. The next section demonstrates the study of the correlation of the scale parameters with the final batch quality.

8.3 Quality Correlations

The goal of a historical database analysis of any industrial process is to enhance process understanding and then use the enhanced understanding for improvements in the process. By improvement, it is usually meant tighter control on the product quality, increased energy efficiency, reduced cycle time and so on. Thus 'quality' is a very generic term and should be understood as encompassing not only the product quality parameters, but also the process efficiency parameters.

Having defined 'quality' in a generic sense, it is noted that in the batch profile SPC framework, the scale parameters characterize the various types of systematic variability in the process. It is therefore reasonable to expect some correlation of these

scale parameters with the quality parameters. Physically meaningful correlation can be utilized to build a regression model for predicting the quality from the scale parameters.

In the PMMA batch reactor simulation example, the correlation of the scale parameters with the product polymer quality – MW_n , MW_w and PDI is studied. The physically meaningful scale parameters for studying the correlation consist of parameters after the reaction phase begins (post initiator addition). As was pointed out in chapter 7, the polymer quality is determined by the competition between dead chain formation and live chain propagation. The time scale parameters after the reaction begins quantify the rate of progress of the batch and should therefore be studied for correlation with the final batch quality. Also, the reactor temperature effects the kinetics of the reaction so that any magnitude scale parameters for the reactor temperature can be used for studying the correlation. In the present case, there is no significant variability in the reactor temperature after the reaction begins so that only the time scale parameters need to be considered.

The time scale parameters of interest are proposed as follows. The first time scale parameter is the duration between the initiator addition and the heater shut off indicating the beginning of the gel effect. The second time scale parameter is taken as the duration between initiator addition and the heater turning on again indicating almost complete conversion of the monomer. The third time scale parameter is taken as the duration between the heater turn on and the end of the batch and indicates the time scale variability towards the end of the batch.

Before building a regression model, the complete data set is divided into 2 parts – 1. The training set for building the regression model and 2. The test set for validating the model on data not seen by the model. The training set consists of 60 randomly chosen batches from the nominal data set while the test set consists of the rest. The correlation coefficient of the three time scale parameters with the final batch quality parameters is tabulated in table 8.4. Note that very high correlation exists between the second time scale parameter and the polymer quality parameters. High correlation exists for the third time scale parameter while mild correlation exists for the first time scale parameter.

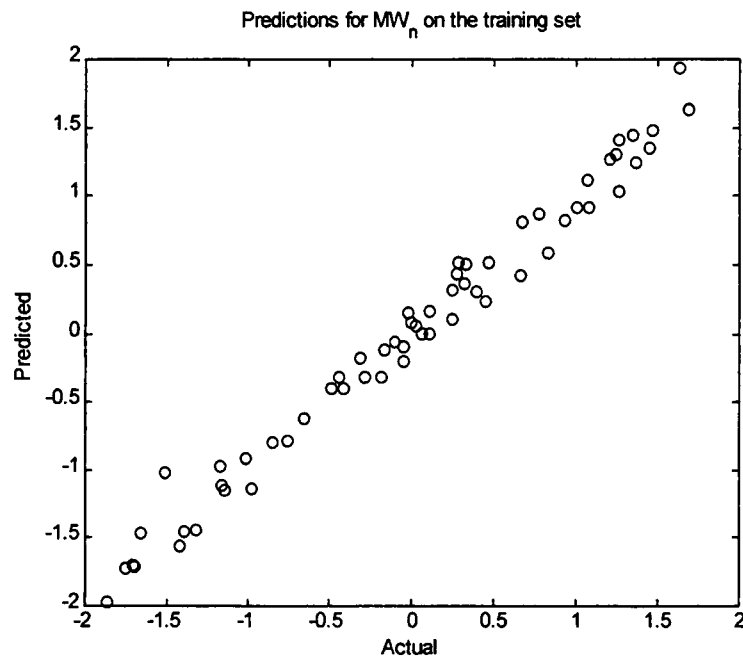
The correlation structure can be explained from a physical understanding of the process. The variability in the reactor loading is minimal so that almost all of the variability in the polymer quality is due to the impurities in the monomer. From past studies, it is known that the most nonlinear phase in the polymerization is the gel effect so that the monomer impurities effect the reaction kinetics most severely in the gel effect phase

TABLE 8.4: Correlation coefficient of time scale parameters with quality

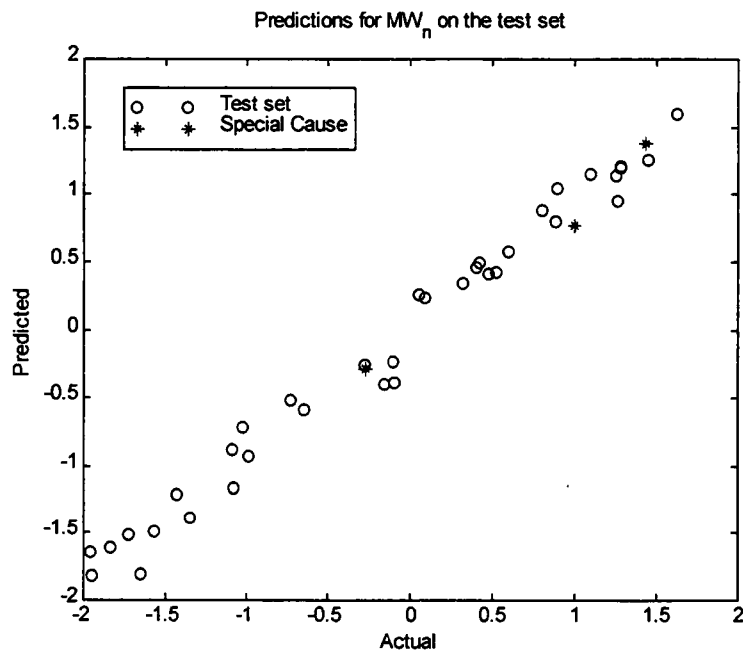
	TS 1	TS 2	TS 3
MW _n	0.4126	-0.9734	0.7467
MW _w	-0.3916	-0.9786	0.7574
PDI	-0.3805	-0.9804	0.7623

Thus high correlation is obtained with the second time scale parameter which signifies the end of the gel effect phase. The effect of the impurities is not as significant before the gel effect phase begins so that mild correlation is obtained for the first time scale parameter due to the small variability in the initial reactor charging. High correlation with the third scale parameter is due to the fact that the third time scale parameter itself is highly correlated with the second time scale parameter with a correlation coefficient ~ 0.7 . For regression purposes, all the time scale parameters represent reasonably independent information since the highest correlation coefficient of ~ 0.7 is not ill conditioned.

A simple linear regression model is therefore built to predict the product quality using all time scale parameters. The regression equations in chapter 2 are used. The predicted and actual values are plotted for the training set in figure 8.16(a). Note that both the X and Y matrices are autoscaled before regression so that the units on the plot are not the actual quality units. The model is validated by predictions on unseen data in the test set. The predicted and actual values for the test set are plotted in figure 8.16(b). Note from the plot that the points are almost about the 45 degree line for both the training and test set indicating that the model is good. Also, the predictions for the special cause batches 98 to 100 are plotted in the figure showing the good predictions. Batch 97 is not plotted since due to the increases amount of oxygen, the reaction kinetics is severely affected causing out of range product quality.



(a)



(b)

FIGURE 8.16: Quality prediction for MW_n using time scale parameters
 (a) Training set
 (b) Test set

To quantify the predictions, the root mean square error of prediction (RMSEP) for both the test and the training sets are plotted in figure 8.17. Since both the **X** and **Y** matrices in the regression are autoscaled, a $RMSEP \ll 1$ indicates a good model. Such small RMSEP is observed in figure 8.17. Also note that the RMSEP for both the test and the training set are comparable, thus validating the model.

This completes a demonstration of the correlation of scale parameters with the batch quality. In the example study, a very high correlation of one of the time scale parameters with the product quality enables building a good regression model. Such a model can be used for predicting the batch quality on-line and corrective action can be taken in case it is felt that the quality is not within specification. The corrective action must however be based on a physical understanding of the process.

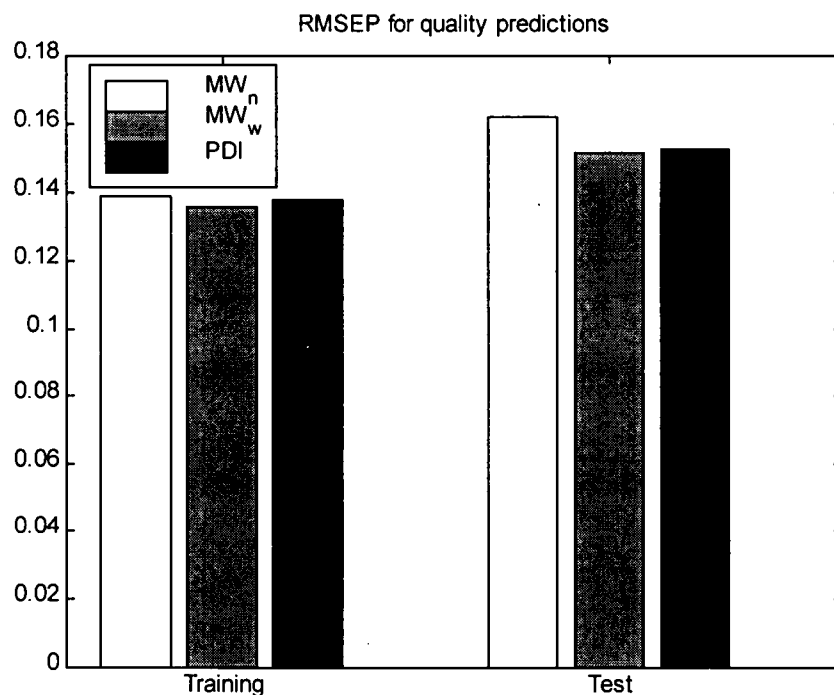


FIGURE 8.17: Mean square error of prediction for all the quality variables

In the present case, the second time scale parameter is not available till the gel effect phase is almost complete. Nevertheless, once the time scale event has occurred, a good prediction of the batch quality is possible. On-line corrective action may then be taken, for example by adding a suitable additive such as a reaction inhibitor, to have the desired effect on the product quality. Such corrective steps are however possible only through a physical understanding of the process. Indeed, the empirical modeling approach is only a correlational study and not a cause and effect study so that observed high correlation must be backed by physical reasoning about the cause. Also, the corrective action to be taken needs process understanding. This highlights the fact that statistical tools must always work in tandem with good engineering judgement for realizing the possibility of process improvement.

8.4 Discussion

This chapter demonstrates the application of SPC tools for the characterization of the variability in batch profiles in terms of simple scale parameters and the study of the correlation of the scale parameters with the batch quality. The technique as was demonstrated is an offline analysis. It can however be easily adapted for on-line monitoring by using on-line SPC charts on the scale parameters and the Q statistic as was proposed earlier in references [40, 41]. Briefly, SPC charts are proposed on the scale parameters and the Q statistic based on data available till the current time in a new batch. Other modifications such as filling up the future with zeros or an exponential decay to 0 and then applying the projection techniques on the

'anticipated' complete profiles can also be used. Control limits are obtained for each time point in the batch for the scale parameters and the Q statistic. The nominal operation data set is used for obtaining the limits.

The example study clearly brings out the importance of time scaling the profiles for a meaningful analysis. Time scaling explains a major portion of the variability in the profiles so that it is important for profile monitoring. High correlation of the time scale parameters with the final batch quality was also seen. It also gives a means for distinguishing one special cause from another, e.g. stoichiometry problems from other process problems. In the past work on the analysis of batch profiles, the aspect of time scaling was not given due attention. From our experience on industrial batch data and also the simulation example, it is felt that time scaling is probably the most crucial step, both for monitoring and quality correlation purposes.

To better appreciate the importance of time scaling, consider the coolant flow profiles from time 351 to 450 where the significant events in the gel effect are not aligned in time. A PCA on the mean centered profiles for both the time aligned and raw profiles is performed. It takes 3 PCs to explain about 93% of the overall variability in the raw profiles while a single PC for the time scaled profiles explains >95% of the variability. The PC loadings for both the cases are plotted in figure 8.18. Note that for the raw profiles, three PCs are needed in order to account for the time misalignment while for the time scaled profiles, only a single PC is needed since there is no time misalignment. Also, the interpretation of the PC is quite

straightforward when compared to the raw profiles. This illustrates the importance of time scaling in enhancing the interpretability of the results. Secondly, to emphasize the importance of time scaling for quality predictions, MPLS on the autoscaled raw profiles is performed using cross validation on the test set. The best RMSEP for the test set is 0.3 for any of the three quality variables which is higher than the worst RMSEP of 0.16 obtained for linear regression using only the time scale parameters in the previous section. Time scaling is thus important for both monitoring and quality correlation purposes.

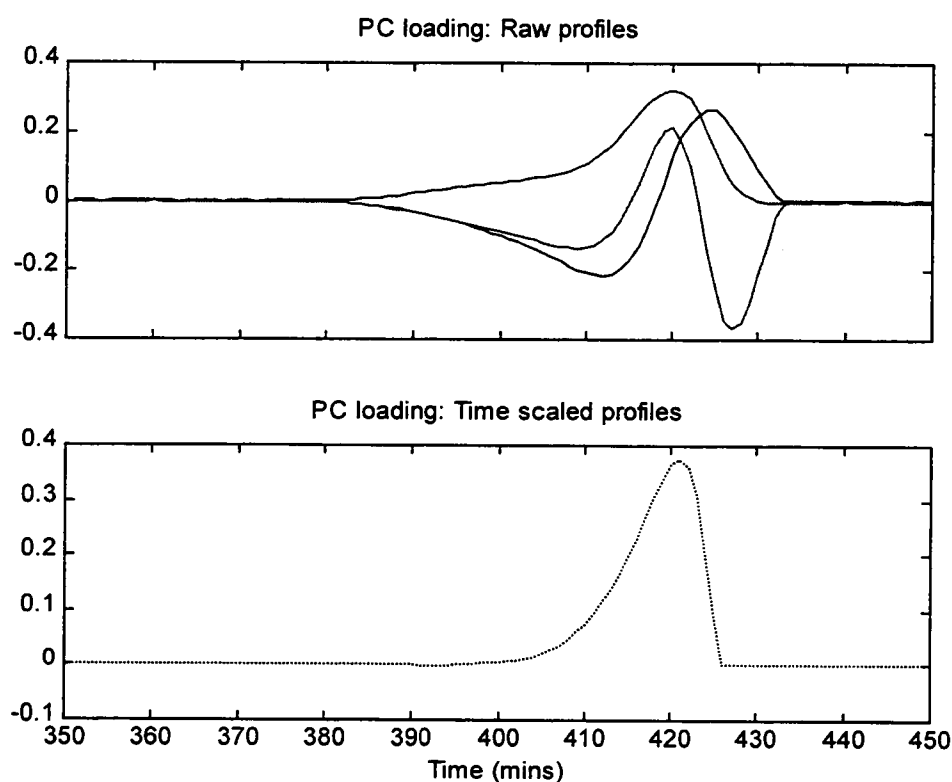


FIGURE 8.18: PC loadings for raw and time scaled profiles

In all the work, the emphasis has been on proposing physically meaningful scale parameters that facilitate process understanding. The various disturbance scenarios studied in the example study show that such meaningful scale parameters go a long way in giving vital clues that distinguish one fault from another.

As far as modeling the product quality is concerned, a high correlation of scale parameters with the product quality may not always be obtained. One principal reason for such an occurrence is the highly non-linear nature of the process rendering linear modeling techniques inadequate. Nonlinear techniques such as non-linear PLS (NLPLS) can be used in case such nonlinearities are detected. The other more important reason is the lack of adequate process measurements that are indirectly related to the process quality. Such cases call for better measurements that are reflective of the product quality. The technique can nevertheless be used for monitoring the process and characterizing the variability in the profiles. It may also be pointed out that only those special causes that are 'observable' or in other words are reflected in some or all the measurements can be detected by the monitoring technique. For example, suppose that in the PMMA simulation example, only the reactor temperature was measured, then all of the disturbance scenarios would have gone undetected. The reason is that all the SPC violations in the control charts occurred due to the manipulated variables. Manipulated variables are therefore important since they increase the 'observability' of various disturbance scenarios.

Lastly, it is again emphasized that the modeling technique is an empirical one with suitable modifications to handle the variability typically seen in batch profiles.

All empirical modeling techniques study the correlation structure of the data and in no way provide any cause and effect relationship. The cause and effect relationship comes only from a physical understanding of the process. Such an understanding must be used to assign reasons for violations on the SPC charts. Note that an SPC violation does not necessarily mean that the product would be of poor quality. It only means that a special cause not seen earlier in the empirical model has been encountered. In case it is felt that the quality would be poor, corrective action must again be taken based on engineering judgement. Thus a physical understanding is necessary for any process improvement/adjustments to be possible.

This completes a comprehensive demonstration of the proposed SPC monitoring technique along with a discussion of the various engineering issues. The next chapter summarizes the conclusions from this work and provides suggestions for further investigation.

Chapter 9

Conclusions and Future Work

This research work has focussed on the development of a generic SPC framework for characterizing the variability in batch profiles in a meaningful way. Useful tools for implementing the framework are also developed. Based on the various examples presented, several conclusions can be drawn. The preliminary work also points to some obvious and some not-so-obvious directions for future work. The summary of the work, conclusions and suggestions for future work are discussed in this chapter.

9.1 Summary

This work has developed an SPC framework for empirically modeling nominal operation batch profiles in a historical database to give easily interpretable scale parameters. The overall variability in the profiles is partitioned as along the time axis and the measurement axis with scaling being proposed to capture the consistent (characterizable) part in the time and magnitude scale parameters. The work studies existing tools (DTW and PCA) for implementing the profile scaling. Simple new techniques especially applicable to industrial batch profiles are also developed. The use of shapes and features is thus proposed for the extraction of consistent event times while PCA is modified to a moving window type approach to give easily interpretable factor loadings. SPC charts (univariate and multivariate) on the scale parameters and

the residuals remaining after scaling are proposed for monitoring batches against a baseline of past nominal operation. The application of the framework is demonstrated on profiles generated from a simulated PMMA reactor. The tentative conclusions or inferences that can be drawn based on this work are listed in the next section.

9.2 Conclusions

A systematic study of a historical database of profiles from past successful batch operation provides useful information for process improvement and monitoring. The historical database provides a convenient reference (or baseline) against which a new batch can be compared and monitored in the SPC sense. The partitioning of the variability along the time axis and the measurement axis is generic and leads to simple scale parameters (time and magnitude) for profile characterization that can be readily interpreted. Scale parameters that are highly correlated with the final batch quality can be used to build a regression model of the batch quality. The simulated example application clearly demonstrated all of the above aspects.

In general, the demonstration clearly establishes the importance of time scaling. Not only is it important for the meaningful application of multivariate techniques such as MPCA, the phase durations can be significantly correlated to the final batch quality parameters, and therefore useful for quality predictions. Time scaling is thus of paramount importance in the characterization of batch profiles. This aspect has hitherto been neglected in the recent publications with their sole emphasis

being on the application of multivariate techniques for only the measurement axis variability.

The application of the template based translation technique for feature time extraction for time scaling, showed that sharp features such as a valve shut off or a valve turn on are determined accurately. In cases where some variability occurs along the measurement axis, the use of a shape factor to scale the windowed profile before calculating the SSE statistic significantly increases the accuracy of the event time extraction as demonstrated in the example application.

The application of DTW for time synchronization, on an industrial example profile indicates that DTW is sensitive to noise and profiles that contain distinct smooth events but are noisy elsewhere can lead to spurious features. Its use should therefore be limited to smooth profiles containing little noise. The computational efficiency of the shapes and features technique makes it the preferred choice over DTW for time alignment. Further, DTW results in the warp function F , which in itself is a profile that requires characterization, another disadvantage limiting its application.

The example application on the PMMA simulation profiles shows that the time scale parameters also help in distinguishing between the various types of faults that can occur in a process – eg stoichiometry problems from other process problems.

As far as magnitude scaling is concerned, the division of the complete profile into time zones containing ‘consistent’ measurement axis variability results in easily interpretable factors that characterize most of the measurement axis variability. The

interpretability of the factors makes them very useful in enhancing process understanding and diagnosing abnormal operation. As demonstrated in the PMMA example, this is a distinct advantage over the MPCA technique that handles all the data as a single entity.

The application of the developed moving window PCA algorithm for the extraction of such zones shows much promise.

The PMMA example showed the monitoring capabilities of the technique for special cause detection using the proposed SPC charts. The scale parameters quantify the variability of a particular type and violations on a scale parameter SPC chart are easily interpretable and therefore diagnosed for possible physical reasons for the upset.

The residuals remaining after scaling are useful in detecting any special causes not explained by the scale parameters. These residuals can be autoscaled to give high sensitivity to subtle special causes. The PMMA simulation example showed that all the special cause batches are detected by the Q statistic, which combines the residuals into a single number.

Scaling the profile matrix seems to play a crucial role in special cause detection. The results on the PMMA simulation show that MPCA using standardization by the overall mean profile standard deviation does not detect any of the special causes in any of the SPC charts. On the other hand, MPCA using autoscaling and the methodology proposed in this work using autoscaling the residuals, detect all of the special cause batches on the Q statistic. The proposed

methodology of autoscaling the residuals after the consistent variability has been extracted in the scale parameters seems to give better sensitivity to special cause detection than MPCA using autoscaling.

The PMMA demonstration also shows that manipulated variables are the key to detecting special cause batches. This is intuitively understood by the fact that a good control system transforms almost all of the process variability to the manipulated variables in order to track the controlled variable. Thus, special attention must be paid when analyzing the manipulated variables.

In summary, the research work demonstrates the various benefits of undertaking a systematic study of historical batch databases for characterizing the variability in the batch operation for enhancing process understanding. This initial work is based on a sound modeling philosophy of partitioning the typical variability seen in batch profiles as along the time axis and the measurement axis. From the experience on various industrial data sets, it is felt that this basic underlying philosophy is applicable to all batch processes in a generic way. The tools to implement the philosophy can however be improved upon and some suggestions in that direction are the subject of the next section. Certain issues that merit further attention are also pointed out.

9.3 Future Work

This research work has primarily focussed on the offline analysis of batch profiles. The proposed framework can be readily adapted to be implemented online.

Future work should focus on formalizing the online analysis framework. The issue of time scaling a batch that is still developing in time needs to be addressed. In this direction, it is felt that a logic based scheme that takes into consideration the temporal order of the events to decide the present time of the batch can be implemented.

Time scaling may lead to information loss in case the profile is longer in duration than the reference. Future work should develop simple SSE type statistics in order to quantify the information loss.

Time scaling is based on the occurrence of consistent checkpoints. In this work, the use of feature templates in order to extract times of occurrence of consistent events has been demonstrated. Variability in the profiles causes a reduction in the recognition accuracy. The use of a shape factor to extract such variability enhances the recognition accuracy as was shown in chapter 5. Additionally, a dilation factor can also be used that dilates (or constricts) the feature so that 'best' overlap with the windowed profile occurs. Such dilation variability is usually seen in batch profiles due to the variability in the rate at which a batch progresses. Thus, the use of a dilation factor in addition to a shape factor would significantly enhance the recognition accuracy of the technique.

In recent years, the wavelet transform has been used for feature identification and is based on the 'translation + dilation' technique. The use of the wavelet transform for event detection merits further research.

As for magnitude scaling, there is scope for improving the preliminary algorithm that was developed for the automatic extraction of time zones that show

measurement axis variability of a particular kind. Any improvement on the present algorithm would be incumbent upon the criteria that are used in order to determine if a zone has been detected and also the criteria for extending the zone. Also, faster algorithms such as successive bisection of future time zones can be used to implement extension of a time zone.

Work can also be done in the direction of using clustering algorithms for automatically finding batches that are clustered together in the scale parameter space. Abnormal batches belonging to different clusters may be due to different physical reasons. This concept of clustering of scale parameters was demonstrated, albeit in a qualitative manner, in the PMMA simulation example. Clustering techniques may also be needed to sort out the historical batch database into batches that belong to say, different modes of operation or different product grades. Note that a separate SPC monitoring model needs to be built for each such cluster.

The proper analysis of manipulated variables forms the key to the success of the SPC monitoring technique. Sometimes, the manipulated variables are very noisy due to extremely aggressive controller tuning coupled with measurement noise. PCA is not very effective in filtering out such noise leading to meaningless control limits such as a 150% valve opening. Traditional filtering techniques may therefore be needed as a preprocessing step to get filtered profiles, which can readily be subjected to a meaningful analysis. Future work should focus on simulating such examples and studying the effect of such profile filtering on the results.

9.4 Industrial Data Acknowledgement

The primary motivation for this work has been the development of an SPC framework that can provide useful engineering information to the operators and process improvement groups in an industrial setting. The tools as proposed in this work have been developed after the analysis of industrial batch datasets provided by MCEC member companies. MCEC (Measurement Control and Engineering Center) is an Industry University Cooperative Research Center (IUCRC). The PMMA simulation has been used as a demonstration of the concept on an industrially relevant process since proprietary concerns limit the publishing of results on industrial datasets. The simulation has been implemented in SIMULINK. The code for implementing the profile scaling tools has been written in Matlab 5.1 and the various functions are listed in the appendix A. The code may be requested from MCEC, University of Tennessee, Knoxville TN 37996.

LIST OF REFERENCES

- [1] Shaw, W.T. (1982), *Computer Control of Batch Processes*, EMC Controls, Inc, Cockeysville, MD
- [2] Fisher, T.G. (1990), *Batch Control Systems: Design, Application, and Implementation*, Instrument Society of America
- [3] Love, J. (1987), "Batch Process Control", *Chemical Engineering*, 437, 34-35
- [4] Staples, R.A.C. (1987), "Cost Effective Batch Control Syetem Using Programmable Controllers", *ISA Transactions*, Paper #87, 1129
- [5] Martin, P.G. (1984), "Computer Control of Batch Processes", *Measurements and Control*, September
- [6] Craig, L.W. (1995), "ISA's SP88 Standard for Batch Process Control", *Chemical Processing*, 58(2), 5
- [7] Reklaitis, G.V. (1995), "Scheduling Approaches for the Batch Process Industries", *ISA Transactions*, 34, 349-358
- [8] Goodall, W.R. and R. Roy (1996), "Short Term Scheduling and Control in the Batch Process Industry Using Hybrid Knowledge Based Simulation", *International Journal of Production Research*, 34(1), 33-50
- [9] Nobuaki, I. and M. Masaaki (1997), "Generic Framework for an Online Scheduling and Control System in Batch Process Management", *Computers and Chemical Engineering*, 21(11), 1291-1310
- [10] Rippin, D.W.T. (1993), "Batch Process Systems Engineering: A Retrospective and Prospective Review", *Computers and Chemical Engineering*, 17 Supplement, 1-13
- [11] Georgiou, G., and J. Hale (1992), "Guide to Batch Process Control", *Process and Control Engineering*, 45, 42-43
- [12] Juba, M.R. and J.W. Hamer, "Progress and Challenges in Batch Process Control", *Chemical Process Control – CPCIII*, Proceedings of the Third International Conference, Asilomar, CA
- [13] Wilkie, J.D.F. (1990), "Batch Process Control"
- [14] Mamzic, C.L., *Statistical Process Control*, ISA: Practical Guides for Measurement and Control

- [15] Oakland, J.S. (1986), *Statistical Process Control: A Practical Guide*, John Wiley, New York, NY
- [16] Shewhart, W.A. (1931), *Economic Control Quality of Manufactured Product*, D Van Nostrand Co Inc, New York, NY
- [17] Vander Wiel, S.A., W.T. Tucker, F.W. Faltin and N. Doganaksoy (1992), "Algorithmic Statistical Process Control: Concepts and an Application", *Technometrics*, **34**, 286-297
- [18] Hahn, G.J. and M.B. Cockrum (1987), "Adapting Control Charts to Meet Practical Needs: A Chemical Processing Application", *Journal of Applied Statistics*, **14**, 35-52
- [19] Marsh, C.E. and T.W. Tucker (1991), "Application of SPC Techniques to Batch Units", *ISA Transactions*, **30**, 39-47
- [20] Kutsuwa, Y. and H. Matsuyama (1992), "Improved Method of Fault Diagnosis of a Batch Process Using a Pattern Recognition Technique", *International Chemical Engineering*, **32**, 102-114
- [21] Hotelling, H. (1947), "Multivariate Quality Control", *Techniques of Statistical Analysis*, Eisenhart, Hastay, and Wallis (Editors), McGraw Hill, New York, NY
- [22] Anderson, T.W. (1984), *An Introduction to Multivariate Statistical Analysis*, Wiley, New York, NY
- [23] Johnson, R.A. and D.W. Wichern (1988), *Applied Multivariate Statistical Analysis*, Prentice Hall, Englewood Cliffs, NJ
- [24] Tracy, N.D., J.C. Young and R.L. Mason (1992), "Multivariate Control Charts for Individual Observations", *Journal of Quality Technology*, **24**, 88-95
- [25] Jackson, J.E., and G.S. Mudholkar (1979), "Control Procedures for Residuals Associated with Principal Component Analysis", *Technometrics*, **21**, 341
- [26] Sparks, R.S. (1992), "Quality Control with Multivariate Data", *Australian Journal of Statistics*, **34**, 375-390
- [27] Box, G.E.P., W.G. Hunter and J.S. Hunter (1978), *Statistics for Experimenters*, John Wiley and Sons, New York, NY
- [28] Montgomery, D.C. (1996), *Introduction to Statistical Quality Control*, 3rd Ed. John Wiley, New York, NY

- [29] Wise, B.M., N.L. Ricker, D.F. Veltkamp and B.R. Kowalski (1990), "A Theoretical Basis for the Use of Principal Component Models for Monitoring Multivariate Processes", *Process Control and Quality*, **1**, 41-51
- [30] Veltkamp, D.J. (1993), "Multivariate Monitoring and Modeling of Chemical Processes Using Chemometrics", *Process Control and Quality*, **5**, 205-217
- [31] Wise, B.M. and N.B. Gallagher (1996), "The process Chemometrics Approach to Process Monitoring and Fault Detection", *Journal of Process Control*, **6**(6), 329-348
- [32] Wise, B.M. and N.B. Gallagher (1997), "Multi-way Analysis and Process Monitoring and Modeling", *AIChE Symposium Series*, **93**(316), 271-274
- [33] 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
- [34] Kourti, T. and J.F. MacGregor (1995), "Process Analysis, Monitoring and Diagnosis, Using Multivariate Projection Methods", *Chemometrics and Intelligent Laboratory Systems*, **28**, 3-21
- [35] Skagerberg, B., J.F. MacGregor and C. Kiparissides (1992), "Multivariate Data Analysis Applied to Low Density Polyethylene Reactors", *Chemometrics and Intelligent Laboratory Systems*, **14**, 341-356
- [36] Tano, K., P.O. Samskog, J.C. Garde and B. Skagerberg (1993), "Partial Least Squares Modeling of Process Data at LKAB: Predicting Chemical Assays in Iron Ore for Process Control", Presented at the *International Symposium on the Application of Computers and Operations Research in the Mineral Industries*, Canadian Institute of Mining, Metallurgy and Petroleum, Montreal, Canada
- [37] Dayal, B., J.F. MacGregor, P.A. Taylor, R. Kildaw and S. Marcikic (1994), "Application of Feedforward Neural Networks and Partial Least Squares Regression for modelling Kappa Number in a Continuous Kamyr Digester", *Pulp and Paper Canada*, **95**, 26-32
- [38] Houdouin, D., J.F. MacGregor, M. Hou and M. Franklin (1993), "Multivariate Statistical Analysis of Mineral Processing Plant Data", CIM Bulletin, *Mineral Processing*, **86**, 23-34
- [39] Wise, B.M., D.J. Veltkamp, B. Davis, N.L. Ricker and B.R. Kowalski (1988), "Principal Component Analysis for Monitoring the West Valley Liquid Fed Ceramic

Melter", in Post and Wades (Editors), *Waste Management '88 Proceedings*, Tucson, AZ

[40] Nomikos, P., and J.F. Macgregor (1994), "Monitoring of Batch Process using Multiway Principal Component Analysis", *AIChE Journal*, **40**, 1361-1375

[41] Nomikos, P. and J.F. MacGregor (1995), "Multivariate SPC Charts for Monitoring Batch Processes", *Technometrics*, **37**, 41-59

[42] Kourti, T., P. Nomikos and J.F. MacGregor (1995), "Analysis, Monitoring and Fault Diagnosis of Batch Processes Using Multi-block and Multiway PLS", *Journal of Process Control*, **5**, 277-284

[43] Martin E.B., and A.J. Morris (1996), "Overview of Multivariate Statistical Process Control in Continuous and Batch Process Performance Monitoring", *Transactions of the Institute of Measurement and Control*, **18**(1), 51-60

[44] Martin, E.B., A.J. Morris, M.C. Papazoglou and C. Kiparissides (1996), "Batch Process Monitoring for Consistent Production", *Computers and Chemical Engineering*, **20** Supplement A, S599-S604

[45] Kourti, T., J. Lee and J.F. MacGregor (1996), "Experiences with Industrial Applications on Projection Methods for Multivariate Statistical Process Control", *Computers and Chemical Engineering*, **20** Supplementary, S745-S750

[46] Gallagher, N.B., B.M. Wise and C.W. Stewart (1996), "Application of Multiway Principal Component Analysis to Nuclear Waste Storage Tank Monitoring", *Computers and Chemical Engineering*, **20** Supplement, S739-S744

[47] Debashis, N. and C.E. Schlags (1997), "Application of Multivariate Statistical Techniques for Monitoring Emulsion Batch Processes", *Proceedings of the American Control Conference: Piscataway, NJ*, **2**, 1177-1181

[48] Dong, D. and T.J. McAvoy (1996), "Batch Tracking via Nonlinear Principal Component Analysis", *AIChE Journal*, **42**(8), 2199

[49] Miller, P., R.E. Swanson and C.F. Heckler (1993), "Contribution Plots: The Missing Link in Multivariate Quality Control", Presented at Annual Fall Technical Conference of American Society for Quality Control (Milwaukee, WI) and the American Statistical Association (Alexandria, VA)

[50] Nomikos, P. (1996), "Fault Detection and Diagnosis in Batch Processes", Presented at the World Batch Forum

- [51] Rawlings, J.O. (1988), *Applied Regression Analysis: A Research Tool*, Wadsworth and Brooks, Pacific Grove, CA
- [52] Strang, G. (1988), *Linear Algebra and Its Applications*, Harcourt Brace Jovanovich Publishers, San Diego, CA
- [53] Wold, S., K. Esbensen and P. Geladi (1987), "Principal Component Analysis", *Chemometrics and Intelligent Laboratory Systems*, 2, 37
- [54] Jackson, J.E. (1991), *A User's Guide to Principal Components*, John Wiley, New York, NY
- [55] Jolliffe, I.T. (1986), *Principal Component Analysis*, Springer Verlag, New York, NY
- [56] Wang, T.W., S. Vedula and A. Khettry (1994), "Geometric Interpretation of SVD, rank, mean centering and scaling in applying multivariate statistical analysis methods", *Fifth International Conference on Process Control*, 267-270
- [57] Dong, D. and T.J. McAvoy (1996), "Nonlinear Principal Component Analysis- Based on Principal Curves and Neural Networks", *Computers in Chemical Engineering*, 20, 65-78
- [58] Wold, S., H. Martens and H. Wold (1983), *Lecture Notes in Mathematics*, Springer Verlag, Heidelberg, p283
- [59] Geladi, P. and B.R. Kowalski (1986), "Partial Least Squares Regression: A Tutorial", *Analytica Chimica Acta* 185, 182-195
- [60] Lorber, A., L.E. Wangen, and B.R. Kowalski (1987), "A Theoretical Foundation for the PLS Algorithm", *Journal of Chemometrics*, 1, 19-31
- [61] Höskuldsson, A. (1988), "PLS Regression Methods", *Journal of Chemometrics*, 2, 211-228
- [62] Phatak, A. and S. de Jong (1997), "The Geometry of Partial Least Squares", *Journal of Chemometrics*, 11, 311-338
- [63] Braak, C.J.F.T. and de Jong S. (1998), "The Objective Function of Partial Least Squares Regression", *Journal of Chemometrics*, 12, 41-54
- [64] de Jong, S. (1993), "SIMPLS: an Alternative Approach to Partial Least Squares Regression", *Chemometrics and Intelligent Laboratory Systems*, 18, 251-263

- [65] Wang, T.W., A. Khettry, M. Berry and J. Batra (1995), "SVDPLS: An Efficient Algorithm for Performing PLS", *First International Chemometrics Internet Conference, InCICNC'94*
- [66] Dayal, B. and J.F. MacGregor (1996), *Journal of Chemometrics*, **11**, 73-85
- [67] Haaland, D.M. and E.V. Thomas (1988), "Partial Least Squares Methods for Spectral Analysis. Relation to Other Quantitative Calibration Methods and the Extraction of Qualitative Information", *Analytical Chemistry*, **60**, 1193-1202
- [68] Höskuldsson, A. (1995), "A Combined Theory for PCA and PLS", *Journal of Chemometrics*, **9**, 91-123
- [69] Thomas, E.V. and D.M. Haaland (1990), "Comparison of Multivariate Calibration Methods for Quantitative Spectral Analysis", *Analytical Chemistry*, **62**, 1091-1099
- [70] Martens, H. and T. Naes (1989), *Multivariate Calibration*, John Wiley, New York, NY
- [71] Höskuldsson, A. (1992), *Prediction Methods in Science and Engineering*
- [72] Wold, S., P. Geladi, K. Esbensen and J. Ohman (1987), "Multi-Way Principal Components and PLS Analysis", *Journal of Chemometrics*, **1**, 41
- [73] Wangen, L.E. and B.R. Kowalski (1988), "A Multiblock Partial Least Squares Algorithm for Investigating Complex Chemical Systems", *Journal of Chemometrics*, **3**, 3-20
- [74] Geladi, P., (1989), "Analysis of Multi-Way (Multi-Mode) Data", *Chemometrics and Intelligent Laboratory Systems*, **11**, 11-30
- [75] Wold, S., N.K. Wold and B. Skagerberg (1989), "Nonlinear PLS Modeling I", *Chemometrics and Intelligent Laboratory Systems*, **7**, 53-65
- [76] Wold, S., N.K. Wold and B. Skagerberg (1989), "Nonlinear PLS Modeling II. Spline Inner Relation", *Chemometrics and Intelligent Laboratory Systems*, **14**, 71-84
- [77] Konstantinov, K.B. and T. Yoshida (1992), "Real-Time Qualitative Analysis of the Temporal Shapes of (Bio)process Variables", *AIChE Journal*, **38**, 1703-1715
- [78] Darnell, B. (1996), "Preprocessing Batch Profiles for Statistical Process Control", Master of Science Thesis, University of Tennessee, Knoxville

- [79] Bellman, R.E. and S.E. Dreyfus (1962), *Dynamic Programming*, Princeton University Press, Princeton, NJ
- [80] Itakura, F. (1975), "Minimum Prediction Residual Principle Applied to Speech Recognition", *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP23(1), 67
- [81] Myers, C., L.R. Rabiner and R. Rosenberg (1980), "Performance Tradeoffs in Dynamic Time Warping Algorithms for Isolated Word Recognition", *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP28(6), 623-635
- [82] Rabiner, L.R., A.E. Rosenberg and S.E. Levinson (1978), "Considerations in Dynamic Time Warping Algorithms for Discrete Word Recognition", *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP26(6), 575-582
- [83] Sakoe, H. and S. Chiba (1978), "Dynamic Programming Algorithm Optimization for Spoken Word recognition", *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP26(1), 43-49
- [84] Kassidas, A., J.F. MacGregor and P.A. Taylor (1998), "Synchronization of Batch Trajectories using Dynamic time Warping", *AIChE Journal*, 44 (4), 864-875
- [85] Everett, B (1988), *Cluster Analysis*, Wiley, New York, NY
- [86] James, M. (1985), *Classification Algorithms*, Wiley, New York, NY
- [87] Fogler, H.S. (1992), *Elements of Chemical reaction Engineering*, Prentice Hall, Englewood Cliffs, NJ, 351-359
- [88] Ray, W.H. (1972), "On the Mathematical Modeling of Polymerization Reactors", *Journal of Macromolecule Science Reviews and Macromolecule Chemistry*, C8(1), 1
- [89] Baillagou, P.E. and D.S. Soong (1985), "Major Factors Contributing to the Nonlinear Kinetics of Free Radical Polymerization", *Chemical Engineering Science*, 40(1), 75-86
- [90] Baillagou, P.E. and D.S. Soong (1985), "Molecular Weight Distribution of Products of Free Radical Nonisothermal polymerization with Gel Effect. Simulation for Polymerization of PolyMethyl Methacrylate", *Chemical Engineering Science*, 40(1), 87-104
- [91] Chiu, W.Y., G.M. Carratt and D.S. Soong (1983), "A Computer Model for the Gel Effect in Free Radical Polymerization", *Macromolecules*, 16(3), 348-357

- [92] Soroush, M. and C. Kravaris (1992), "Nonlinear Control of a Batch Polymerization Reactor: An Experimental Study", *AIChE Journal*, 38, 1429-1448
- [93] Soroush, M. and C. Kravaris (1993), "Optimal Design and Operation of Batch Reactors", *Industrial and Engineering Chemistry Research*, 32, 866-893

APPENDICES

APPENDIX A

MATRIX SCALING FUNCTIONS

AUTOSCALE: scales cols of X to 0 mean and unit variance.
Cols with 0 std dev are only mean centered

Call Syntax: `[Xs, mx, sx] = autoscale(X);`

Input: X: input matrix

Output: Xs: Scaled matrix

mx: Mean of cols of X (row vector)

sx: std dev of cols of X (row vector)

Calls: mncn

MNCN: Mean centers the columns of X

Call Syntax: `[Xm, mx] = mncn(X);`

Input: X: Input X matrix

Output: Xm: Mean Centered Matrix

mx: mean of cols of X (row vector)

OVERALLSTD: Standardizes X by its overall mean row std dev

Call Syntax: `[Xs, mx, sx] = overallstd(X);`

Input: X: Matrix to be standardized

Output: Xs: Scaled X matrix

mx: Mean of cols of X

sx: overall mean std dev

Calls: mncn

PRETREAT: Scales X matrix according to input option

Call Syntax: `[Xs, mx, sx] = pretreat(X, opt);`

Input: X: Matrix to be standardized

opt: option for scaling the X Matrix

0: No data scaling,

1: mncn only

2: autoscale

3: overall std scaling

Output: Xs: Scaled X matrix

mx: Mean of cols of X

sx: std dev

Calls: mncn, autoscale, overallstd

RESCALE: Rescales X by input mean and std dev

Call Syntax: `Xs = rescale(Xs, ms, sx);`

Input: X: Matrix to be rescaled

mx: mean to be subtracted (row vector)

sx: std dev to be divided (row vector) [optional]

Output: Xs: rescaled matrix

UNSCALE: Unscales X by input mean and std dev

Call Syntax: `Xu = Unscale(X, mx, sx);`

Input: X: Matrix to be unscaled
 mx: mean to be added (row vector)
 sx: std dev to be multiplied (row vector) [optional]
Output: Xu: unscaled matrix

DYNAMIC TIME WARPING FUNCTIONS

DTW: Dynamic time warping of P onto R

Call Syntax: `[g, pth] = dtw(R, P, nstp, M, W, E, sopt, dopt);`

Input: R: Reference
 P: New profile
 nstp: number of consecutive horizontal steps in the warp function
 M: Global constraint search window about the diagonal line
 E: Endpoint constraint (optional; 0 for fully known end points)
 sopt: symmetric option (optional; 0-Asymmetric, 1-symmetric)
 dopt: 1 - displays every 10 time points warped (optional [1])

Output: g: recursive distance
 pth: index of path followed on the time grid

Calls: `InWindow`, `Feasible`, `SlopeConstraint1`, `DP`

INWINDOW: Checks if time indices i,j for DTW are within the global search space

Call Syntax: `flag = InWindow(i,j,M)`

FEASIBLE: Finds the feasible predecessors of I,J on the time grid
subject to global constraint M and relaxed end pt E

Call Syntax: `[pth, stat] = Feasible(I, J, M, E)`

pth is a symbolic representation for the feasible predecessors

- 1: J-1 approach
- 2: J-1, I-1 approach
- 3: I-1 approach
- 4: starting point approach

stat is empty if no feasible predecessors

SLOPECONSTRAINT: Applies the slope constraint to feasible search points in
prev to give Previous which satisfy the local slope
constraints also

Call Syntax: `Previous = SlopeConstraint(i, j, prev, nstp)`

DP: Dynamic programming equation. Finds the predecessor that gives the least
g_new and also returns the path followed in pth

Call Syntax: [g_new,pth] = DP(R, P, I, J, path, W, sopt);

GETF: Extracts the optimum warp function F for DTW

Call Syntax: F = GetF(pth,g);

Calls: GetPrevIdx

FWARP: Apply warping F to P and R

Call Syntax: [Pw, Rw]=Fwarp(P,R,F,opt);

When opt = 0, only P is warped. Otherwise both P and R are warped to an imaginary axis

TIME SCALING FUNCTIONS

ICMSTSALL: Translates ftr template (with dilation) x over profiles Y

Call Syntax: [MSE, IC, MS] = icmstsall(Y, x, cnstr);

Input: Y: Profile matrix (time along rows)
x: ftr template (J by 1)
cnstr: Constraint on IC and MS (2 by 3)
first row-maximum (IC,MS,TS)
second row-minimum (IC,MS,TS)
use TS=1 for no dilation

Output: MSE: MSE after translation (3rd dimension for dilation)
IC: initial condition (intercept)
MS: magnitude scale (slope)

Calls: icmstrns

ICMSTRNS: Translates ftr template x over profiles Y using shape factors

Call Syntax: [MSE, sf]=icmstrns(Y, x, cnstr);

Input: Y: Profile column vector (K by 1)
x: ftr template (J by 1)
cnstr: Constraint on sf (2 by 2)[opt]
first row-minimum
second row-maximum

Output: MSE: MSE on translation
sf: shape factors (IC, MS; 2 by J+K-1)

FTRTIME: calculates the absolute ftr and event time

Call Syntax: [tf, te] = ftrtime(MSE, fe, K, h, clip);

Input: MSE: MSE of translation for a batch (m, n)
fe: Time of event in the feature (>0 <K+1)
K: ftr template length
h: time index of snapshot beginning that was translated

clip: initial and final index of MSE to be studied (clip edges)
 Output: tf: feature time (n,1)
 te: event time (n,1)

TIMESCALE: Time scales the input array according to given options

Call Syntax: Pts = timescale(P, R, t, tr, opt)

Input: P: Ix1 cell array of batch profiles (Each cell (batch) is Ki (time) by J)
 R: Reference profile matrix (K by J)
 t: Event time matrix excluding beginning and end time (I by M)
 tr: Corresponding event times in R
 opt: Options for time scaling (M+1 by J)
 0: Linear interpolation between events
 1.1: Pad/chop end with R
 1.2: Pad/chop beginning with R
 2.1: Pad/chop end with last available value
 2.2: Pad/chop beginning with last available value

Output: Pts: Time scaled batch data array (K by J by I)

Calls: interpolate and padchop

INTERPOLATE: Linearly interpolates cols of an input matrix to a length of npts

Call Syntax: P1 = interpolate(P, npts)

Input: P: Input matrix to be interpolated
 npts: Length to which P is to be interpolated (>1)

Output: P1: interpolated matrix with length npts

PADCHOP: Pads/chops cols of a matrix according to option

Call Syntax: Pts = padchop(P, l, opt, R);

Input: P: Input profile matrix (K by J)
 l: number of rows (time points) in Pts
 opt: option for padding/chopping (optional)
 'end': pad/chop at end
 'beg': pad/chop at beginning
 defaulted to 'end'
 R: optional. If input, values in R are used for padding.
 otherwise the last/initial value in P is used

Output: Pts: Padded/chopped profile (l by J)

MAGNITUDE SCALING FUNCTIONS

EFA: Moving window PCA for extraction of zones of systematic variability

Call Syntax: fctrs = efa(X, minl, tol);

Input: X: Profile matrix (batches along rows)

minl: minimum window length
 tol: three element vector of tolerances.
 tol(1) 1-the min variance that must be explained in a snapshot
 tol(2) min standard deviation in the snapshot
 tol(3) max standard deviation in the snapshot

Output: fctrs: Sytematic variability directions

Calls: extend

EXTEND: Tests if a snapshot contains systematic variability and if so extends the snapshot

Call Syntax: [load, endpt] = extend(X, h, minl, tol3);

Input: X: Profile matrix (batches along rows)

h: Snapshot start time index

minl: minimum length of the snapshot from the hinge

tol3: Three element tolerance criteria for extraction of zones

 tol3(1) 1-the min variance that must be explained in a snapshot

 tol3(2) min standard deviation in the snapshot

 tol3(3) max standard deviation in the snapshot

Output: load: snapshot factor loading (empty if factor not found)

endpt: end point index of snapshot. (-1 if factor not found)

Calls: loadcalc

LOADCALC: Uses svd to calculate the factor loadings

Call Syntax: [load, Xpp, mse, msepp] = loadcalc(X, npcs);

Input: X: Profile matrix (samples along rows)

npcs: Number of PCs to calculate

Output: load: PC loadings

Xpp: Residuals after projection on the PC

mse: mean square error of X

msepp: mean square error of Xpp

Calls: svds (built in matlab function)

PCA: Performs PCA on input data matrix using SVD.

Call Syntax: [T, P, S, vexp] = pca(X, n)

Input: X: Input matrix

n: Number of PCs to retain (optional)

opt: output options (optional)

Output: T: sample scores

P: sample loadings

S: eigenvalues of retained PCs

vexp: %age variance explained by each PC

SPC FUNCTIONS

SPE: Calculates SSE of the residuals in E (rows)

Call Syntax: `Q = spe(E);`

SPECL: Calculates 95% and 99% confidence limits on the Qstatistic (SSE)

Call Syntax: `Qcl = specl(E);`

HOTELLING: Calculates the Hotelling's statistic of T using the cov matrix (Sinv)

Call Syntax: `D = hotelling(T, Sinv);`

HOTELCL: Calculates the 95% and 99% confidence limits on the Hotelling's statistic

Call Syntax: `Dcl = hotelcl(T);`

MISCELLANEOUS FUNCTIONS

D3TOMAT: Unfolds the 3rd dimension of input 3d array

Call Syntax: `X = d3tomat(X3d);`

Input: `X3d:` 3d array to be unfolded (I,J,K)

Output: `X:` 2d unfolded matrix (I,J.K)

UNIQUE2: Extracts the unique elements in index. Output is sorted.

Call Syntax: `ndx=unique2(index)`

FINDNOT: Finds elements of x that do not equal any element in y

Call Syntax: `X = findnot(x, y)`

APPENDIX B

GLOSSARY

AIBN	azo iso butyro nitrile
DTW	dynamic time warping
EFA	evolving factor analysis
LV	latent variable
MLR	multiple linear regression
MPCA	multiway principal component analysis
MPLS	multiway partial least squares
MS	magnitude scale
MSE	mean square error
MSEP	mean square error of prediction
MW _n	number average molecular weight
MW _w	weight average molecular weight
NIPALS	non-linear iterative partial least squares
PC	principal component
PCA	principal component analysis
PCR	principal component regression
PDI	poly dispersity index
PID	proportional integral derivative
PLS	partial least squares
PMMA	polymethyl methacrylate
QSSA	quasi steady state approximation
RMSEP	root mean square error of prediction
SPC	statistical process control
SSE	sum of squared error
SVD	singular value decomposition
TS	time scale

VITA

Nitin Kaistha was born in Pune, India on November 29, 1974. He graduated from Central School #1, Kanpur in 1992. He entered the Indian Institute of Technology, Kanpur that same year and graduated with a Bachelor of Technology with a major in Chemical Engineering in 1996. He then entered the graduate program at the University of Tennessee, Knoxville that same year and received a Doctor of Philosophy degree with a Major in Chemical Engineering and Minor in Statistics in 1999. He is a member of Phi Kappa Phi and the Instrument Society of America.