

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI[®]

University of Alberta

STATISTICAL PROCESS MONITORING AND MODELING USING PCA AND PLS

by

Haitao Zhang



A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of **Master of Science**.

in

Process Control

Department of Chemical and Materials Engineering

**Edmonton, Alberta
SPRING 2000**



National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file *Votre référence*

Our file *Notre référence*

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-60199-4

Canada

University of Alberta

Library Release Form

Name of Author: Haitao Zhang

Title of Thesis: Statistical Process Monitoring and Modeling using PCA and PLS

Degree: Master of Science

Year this Degree Granted: 2000

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as hereinbefore provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

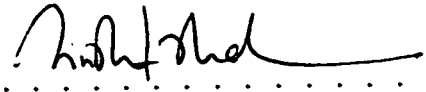
Haitao Zhang
.....
Haitao Zhang
CME 536
University of Alberta
Edmonton, AB
Canada, T6G 2G6

Date: *Jan 28, 2000*
.....


University of Alberta

Faculty of Graduate Studies and Research

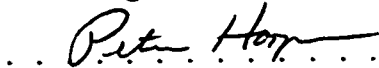
The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **Statistical Process Monitoring and Modelling using PCA and PLS** submitted by Haitao Zhang in partial fulfillment of the requirements for the degree of **Master of Science in Process Control**.



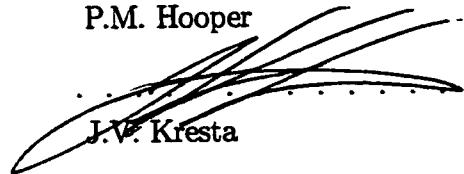
S.L. Shah (Supervisor)



B. Huang



P.M. Hooper



J.V. Kresta

Date: 27 January 2000

Abstract

Advancements in instrumentation and computer techniques allow modern industrial plants to collect and store a large number of measurements in a database. These databases contain extremely useful information about processes and are usually analogous to a “gold mine”. However, without efficient tools, little information can be extracted from these database. Recently, some multivariate statistical tools, such as Principal Component Analysis (PCA) and Partial Least Squares (PLS) have become popular data mining tools.

Unlike the traditional univariate statistical process control tools, multivariate statistical analysis methods take the correlation between variables into account and study many variables simultaneously. The original variables are usually projected into subspace through some dimension-reduced methods. The latent variables in the reduced dimensional space are then utilized to extract useful information about a process. The main objective of this thesis is to investigate these multivariate analysis tools (PCA and PLS) and their practical applications. The method of using PCA for statistical process monitoring is illustrated through a simulated flow rate system. Recursive PCA approach is implemented in a pilot scale plant. An empirical dynamic PLS model is built based on a industrial data set. A root cause diagnosis strategy is presented by taking advantage of PLS weighting vector. The multiscale monitoring strategy which extends the suitability of PCA for SPM based on autocorrelated measurements is demonstrated through a simulated example.

To My Parents

Acknowledgements

I would like to thank my supervisor Dr. Sirish Shah for his excellent and enthusiastic guidance and great support during the entire course of this research. His broad knowledge and deep understanding in the research area led me into the field of process control.

I would like to thank Professors Shah, Huang, Forbes, Chen, Marquez, Wiens for being excellent teachers. It is from them that I learned not only the fundamentals of process control and statistics, but also the methods and tools to carry out my research. Thanks to Jame Kresta for providing industrial data and great help during the course of applying statistical theory into practice. Thanks to Arun for teaching me wavelets, which expanded my research area greatly. I am also indebted to Bob, Jack and Walter for their help with computer and lab equipment.

I am really fortunate to pursue my degree in the CPC group at the U of A. I had a wonderful time in the past two years in this wonderful group. Arun, Lanny, Dongguang, Camron, Dan, Yale, Jie, Xin, Weihua, Huilan, Bushan, Ramish, Sachin, Shoukat, all of you have been such wonderful friends. Many original ideas were generated from the group seminar and free discussion with you. It is all of you that make the CPC group such a good place to stay.

Thank my wife Man Liu for her support, patience, understanding and encouragement during the course of pursuing my degree. This research cannot be done without her love.

Thank the Faculty of Graduate Study and Research and the Natural Science and Engineering Research Council of Canada for their financial support.

Contents

1	Introduction	1
1.1	Multivariate Statistical Process Control	1
1.2	Organization of the Thesis	2
2	Basic Theory of Multivariate Statistical Process Control	4
2.1	Introduction	4
2.2	Univariate Statistical Control Charts	5
2.3	Principal Components Analysis	8
2.3.1	PCA Algorithm	8
2.3.2	Optimal Dimension of a PCA Model	11
2.3.3	PCA for Multivariate Statistical Process Control (MSPC) . .	12
2.4	Partial Least Square	14
2.4.1	PLS Algorithm	14
2.4.2	Nonlinear and Dynamic PLS Model	18
2.5	Conclusions	19
3	PCA in Statistical Process Monitoring– A Simple Example	20
3.1	Introduction	20
3.2	Building of PCA Model	21
3.3	Sensor Fault Detection	23
3.3.1	Abrupt Sensor Fault Detection	23
3.3.2	Ramp-type Sensor Failure Detection	25
3.4	Limitations of PCA Model for SPM	27
3.5	Assembly of the measurement matrix	32
3.6	Minor Components	34

3.7 Conclusion	38
4 Recursive PLS and PCA	39
4.1 Introduction	39
4.2 Recursive PLS	39
4.3 Recursive PCA	43
4.4 Applying RPCA to a CSTH System	47
4.5 Analysis and Conclusions	50
5 Industrial Data Analysis	55
5.1 Introduction	55
5.2 Problem Description	55
5.3 Data Pre-processing	57
5.4 Variable Selection	59
5.5 PLS modeling	64
5.6 Dynamic PLS modeling	67
5.7 Root Cause Diagnosis	71
5.8 Concluding Remarks	79
6 MSPCA-Combination of Wavelets with PCA	80
6.1 Introduction	80
6.2 Introduction to Wavelets	81
6.3 Statistical Assumption of SPE and T^2	84
6.4 Combining Wavelets with PCA-MSPCA	85
6.5 MSPCA for Dynamic Process Monitoring-A illustrated Example . . .	88
6.6 Conclusion & Discussion	91
7 Conclusions	93
7.1 Contributions of thesis	93
7.2 Future Work	95
Bibliography	97
A Color Figures in Thesis	103

List of Figures

2.1	Correlation between the steam demand and the boiler fuel flow	7
2.2	A schematic diagram of PLS regression procedure.	17
3.1	Configuration of flow rate system	21
3.2	Snapshot of flow rate measurements	21
3.3	Eigenvalues of covariance matrix	23
3.4	Two PCs can capture 99% of information in the system.	23
3.5	The <i>SPE</i> plot for the normal operation data.	24
3.6	Hotelling T^2 chart for the normal operation data.	24
3.7	Measurements of flow rates with sensor failure between 300–400. . . .	26
3.8	Hotelling T^2 monitoring chart for the system with abrupt sensor failure.	26
3.9	<i>SPE</i> chart for the system with abrupt sensor failure	26
3.10	<i>SPE</i> contribution plot at sample instant 398 (Abrupt sensor fault). .	27
3.11	The measurements for flow rates with incipient sensor fault.	28
3.12	The <i>SPE</i> plot for the flow rate measurements with incipient sensor failure.	28
3.13	The <i>SPE</i> contribution plot at sample 293 (incipient sensor fault). . .	28
3.14	Flow rates measurements with the sensor fault in F_1 between 100–300.	29
3.15	The <i>SPE</i> monitoring chart detect the abnormal situation between 100– 300	29
3.16	<i>SPE</i> contribution plot at sample instant 249 (Ramp sensor fault) . . .	30
3.17	Three temperature measurements are added to the flow rate system. .	30
3.18	<i>SPE</i> plot for the new PCA model with temperature measurements. .	31
3.19	<i>SPE</i> contribution plot at sample time 295.	32
3.20	The <i>SPE</i> chart (top) is for PCA_1 ; <i>SPE</i> chart (bottom) is for PCA_2 .	34

3.21	Third principal component in <i>PCA1</i> (top); Third principal component in <i>PCA2</i> (bottom)	35
3.22	Thrid PC in <i>PCA2</i> is truely in residual space, while third PC in <i>PCA1</i> still have some autocorrelation.	35
3.23	The significant value in ACF of the third PC matches with the time delay value.	36
3.24	First two PCs explaine the variation of the data. The last PC contains the residuals of energy balance equation.	37
3.25	Third loading gives the correlation between variables	37
4.1	The schematic diagram of the CSTH system	51
4.2	Hotelling T^2 and SPE both detect the temperature and level disturbance.	51
4.3	The Hotelling T^2 and SPE without recusively updating of the model.	52
4.4	Hotelling T^2 and SPE of recursive PCA model, forgetting factor $\mu = 0.995$, model updated every one sample.	52
4.5	Hotelling T^2 and SPE of recursive PCA model, forgetting factor $\mu = 0.985$, model updated after every ten samples.	53
4.6	Recursive PCA can detect both flow rate and temperature disturbance when $\mu = 0.998$	53
4.7	Recursive PCA can detect large flow rate disturbance, but fail to detect small temperature disturbance when $\mu = 0.99$	54
5.1	Trend plots of the total emission and net emission.	56
5.2	The raw measurements of v14 and v36.	59
5.3	Filtered signals of v14 and v36	60
5.4	Comparison of spectrum of orginal and filtered signal (v20)	60
5.5	Time-series of variable 39	63
5.6	Correlation coefficients between total emission and other process variables	63
5.7	First loading coefficients in PLS model vs. correlation coefficients	63
5.8	The cluster analysis display the correlated variables	64
5.9	The cross validation suggest three latent variables should be kept in the model.	65

5.10	The Model output vs. the actual measurements using static PLS model.	66
5.11	A plot of the cumulative PRESS. This suggests that 6 latent variables should be kept in the PLS model.	68
5.12	The fitting error is small	68
5.13	Comparison of model output and actual measurements.	69
5.14	The model prediction vs. the actual measurements when the validation data is scaled by its own mean and variance.	70
5.15	The comparison of model outputs and actual measurements with EWMA updating of mean and variance	72
5.16	Trend plots of V39 (upper), v52(middle), and net emission(lower) . .	73
5.17	First loadings based on different section of data.	74
5.18	First weighting vector in recursive PLS model using all the available data.	75
5.19	First weighting vector in recursive PLS model with EWMA updating method.($\lambda = 0.9$)	75
5.20	First weighting vector in recursive PLS model with fix data length, window length 300, updated every 20 sample.	76
5.21	First weighting vector in recursive PLS model with EWMA updating. ($\lambda = 0.7$)	76
6.1	$f(n)$ is the signal at finest scale. By passing a series of high and low pass wavelets filters, it is decomposed into signals at different scales.	83
6.2	The ACF of z indicates that the data is time-dependent.	85
6.3	ACF of PC-1 (dynamic PCA model) indicating highly autocorrelated scores.	86
6.4	Original data of Z (a); Wavelets coefficients from first H.F. band to the Fourth H.F. band (b)-(e); Scaling function coefficients at the coarsest scale (lowest frequency band) (f)	87
6.5	ACF of reconstructed data from the second level H.F. band.	87
6.6	ACF of wavelets coefficients at first H.F. band or finer scale (a), second H.F band (b), fourth H.F band or coarser scale (c), and fifth H.F. band or coarsest scale(d).	88

6.7	SPE chart with dynamic PCA	89
6.8	T^2 Chart with dynamic PCA	89
6.9	SPE chart with MSPCA at finer level.	90
6.10	T^2 chart with MSPCA at coarsest level.	90
6.11	SPE chart with MSPCA at coarsest level.	90

List of Tables

3.1	Comparison of variance captured by two PCA models	33
4.1	Variance captured by latent variables in dynamic PCA model	48
5.1	Variance captured by latent variables in standard PLS model	65
5.2	Variance captured by latent variables in dynamic PLS model	69
5.3	Correlation coefficients between net emission and v51 based on different section of data	72

Chapter 1

Introduction

1.1 Multivariate Statistical Process Control

Statistical Process Control (SPC) is a philosophy based on the use of a variety of tools for monitoring the performance of a process. SPC provides the basis for achieving continuing improvements in product quality and productivity. An essential part of SPC is to employ a set of statistical tools to establish control charts for a process. Should some abnormal events occur in the process, the control charts should be able to detect the occurrence of special events and diagnose the possible causes. SPC has roots that date back to the 1920s in the work of Dr. Walter A. Shewhart. Since then Shewhart (Shewhart, 1931), CUSUM (Page, 1954; Woodward and Goldsmith, 1964) and EWMA (Roberts, 1959; Hunter, 1986; Lucas and Saccucci, 1990) monitoring charts have been widely accepted and implemented in most industries. In the past, traditional statistical charts were mainly focused on monitoring key product quality variables in plants. These monitoring charts examine one variable at a time as if each variable is independent of another. Unfortunately, measurements from a process are rarely independent; more often than not they are correlated to each other and are affected by some inherent relationship in the process. The univariate charts, which ignore the inter-relationships between variables, make the monitoring and diagnosis of special events very difficult.

In recent years, some multivariate statistical tools, such as Principal Component Analysis (PCA), Partial Least Square (PLS), have become more and more important in SPC. In contrast to the classical univariate SPC tools, these tools take the correlations between variables into account and monitor a set of correlated variables

simultaneously. Moreover, by projecting the original measurements into a latent subspace, latent variables are monitored in a reduced dimensional space. A PCA or PLS model is built on good historical data of normal or “nominal” process operation. This model can then be used to monitor or predict the future behavior of the process based on the assumption that operation in the future should follow the same pattern or template of this normal historical data. Any significant deviation from that assumption should be detected as a “special event”, such as sensor failure, disturbance, or possible process drift.

PCA is concerned with the study of one data block (X). X usually contains all process variables including quality variables. The idea of PCA was first introduced by Pearson (1901), and developed by Hotelling (1933). Their ideas are further reviewed by many other researchers (Jolliffe, 1986; Wold *et al.*, 1987; Jackson, 1991). In recent years, PCA has been used as a multivariate SPC tool in areas of application such as process monitoring (Kresta *et al.*, 1991; Wise and Gallagher, 1996; Chen and McAvoy, 1998; Rannar *et al.*, 1998), gross error detection (Tong and Crowe, 1995), sensor fault identification (Dunia *et al.*, 1996), Validation of plant operating modes (Zullo, 1996), etc.

Partial least squares (PLS) is also known as projection to latent structures. The pioneering work in PLS was done in the late sixties by Wold (1966) in the field of econometrics. The objective was to find relations between blocks of data by relating their latent variables. A detailed description of the PLS algorithm is given by Geladi and Kowalski (1986) and Hoskuldsson (1988). In PLS, two block of variables, X (contains the process variable) and Y (contains the quality variables), are usually included in the analysis. In this case, one would like to find an optimum latent variable transformation such that the transformed variables not only explain the variations in X , but also have most capability in predicting Y . This data-driven empirical statistical model approach is extremely useful under the situation where either a first principal model is difficult to obtain or the measured variables are highly correlated (colinear) to each other. The PLS methods have been extensively researched and applied in the chemometrics field. More recently, interest in PLS is spreading to the chemical engineering area. (MacGregor *et al.*, 1994; Nomikos and MacGregor, 1995; Kourti *et al.*, 1995; Lakshminarayanan, 1997).

1.2 Organization of the Thesis

The organization of the paper is as follows.

In Chapter 2, some univariate statistical control methods, such as Shewhart, CUSUM and EWMA, are reviewed and their limitations discussed. The basic algorithm of PCA and PLS are introduced in detail to provide necessary background for further discussion.

When PCA is used for statistical monitoring, the multivariate statistical charts, such as Hotelling T^2 and Q statistic (also know as Square Prediction Error or SPE) can be employed to monitor the process operating performance. The methodology of using PCA for statistical process monitoring (SPM) is illustrated in detail through a simulated example in chapter 3. Some related issues such as the importance of taking the time delay into account when applying PCA and the advantage of minor PCA are also emphasized in this chapter.

Conventional PCA and PLS models are most suitable for dealing with a steady state process. However, a practical process is usually a dynamic and time-varying one. Directly applying PCA and PLS method to monitor or model such a process often results in false alarms and model-plant mismatch. To adapt to the process drift and change of operating point, recursive updating of PCA and PLS models on-line is essential. The algorithms for recursive PCA and PLS are introduced in chapter 4. A pilot scale plant is then used to demonstrate the idea of recursive PCA (RPCA). Some issues of applying RPCA in practice are also discussed through this example.

Chapter 5 presents the study of applying PLS to deal with a industrial data set. In this industrial data analysis, some practical issues relating to PLS modeling, such as data pretreatment, variable selection and dynamic PLS, are discussed. A novel method which uses the weighting vectors in the recursive PLS model to do root cause diagnosis is discussed in this chapter.

In Chapter 6, the limitations of using T^2 and Q statistics in monitoring autocorrelated measurements are pointed out. Unfortunately, this limitation has not received enough attention in past years and as a result the control charts are misused by practitioners. Multiscale PCA (combining wavelets decomposition with PCA) is proposed in this thesis to overcome this limitation. The multiscale PCA approach not

only meets the underlying assumption of the control charts, but also demonstrates sensitivity in detecting small disturbances. The thesis ends with concluding remarks and suggestions for future research directions.

Chapter 2

Basic Theory of Multivariate Statistical Process Control

2.1 Introduction

The development of computer and instrumentation technology has enabled modern process industries to obtain huge volumes of data. It is often the case that even in a single operating unit, hundreds of process variables may be recorded and monitored. However, not all process variables are independent of each other, i.e. they are usually colinear or correlated in some way. In this case, much information in the collected data is redundant. How to effectively extract the useful information and eliminate the redundancy in the data is a challenging problem. Without effective data analysis tools, little information can be extracted from the data, which results in the situation that the process industry is generally “data-rich but information-poor”. In recent years, Principal Component Analysis (PCA) and Partial Least Square (PLS) have become popular multivariate analysis tools which allow us to extract useful information from process data and analyze the data in greatly reduced dimensional space.

This chapter is organized as follows: in section 2.2, different types of univariate control charts are introduced and limitations of these charts are pointed out. The basic theory of PCA and related issues, such as the selection of the optimal subspace dimension and PCA for Statistical Process Monitoring (SPM) are described in section 2.3. The algorithm of PLS is outlined in section 2.4, together with the discussion on dealing with the nonlinear and dynamic processes. The chapter ends with concluding remarks in section 2.5.

2.2 Univariate Statistical Control Charts

Univariate statistical methods, such as Shewhart Chart, CUSUM and EWMA have been widely used to monitor industrial processes for many years. These methods are briefly introduced here.

- Shewhart Chart

In a Shewhart chart, a sequence of samples (denoted as x_i) are plotted against time. Upper and lower control limits for the samples are established around the process mean (α) based on the “three sigma” rule, i.e. $\alpha \pm 3\sigma$, where σ is the standard deviation of the observations. Whenever the most recent measured point or a consecutive sequence of points are outside the control limits, an abnormal condition is said to be encountered and attention is focused to diagnose the source of the problem. In practice, α can be estimated from the sample average and σ from the sample standard deviation.

- CUSUM

In a CUSUM chart, rather than plotting a sample value itself, one can plot the cumulative sum of observations, $\sum_{i=1}^T x_i$. Or, instead one can plot the cumulative sum of the deviations, $\sum_{i=1}^T d_i$, where $d_i = x_i - c$, and c is a constant value, usually being the target value of the variable. Thus the CUSUM chart is the quantity

$$S_T = \sum_{i=1}^T d_i = \sum_{i=1}^T (x_i - c) \quad (2.1)$$

plotted against time T .

If the mean of samples equals the target value, then S_T will simply appear as a series of random numbers around zero. However, if the mean deviates from the target value by a small amount δ , then each time S_T will be added by this small number δ . Therefore, S_T will increase or decrease against time depending on the sign of δ . Should S_T proceed beyond a pre-determined control limit, then alarm can be announced and appropriate corrective action can be taken.

- EWMA

The exponentially weighted moving average (EWMA) is a statistic which gives less weight to old data, and more weight to new data. The expression for the EWMA is

$$Z_i = \lambda x_i + (1 - \lambda)Z_{i-1}, \quad 0 < \lambda \leq 1 \quad (2.2)$$

The starting value Z_0 is usually set to be the target value. Z_i is the output of EWMA and x_i is the observation from the process. Equation 2.2 can also be written as

$$Z_i = \lambda \sum_{j=1}^i (1 - \lambda)^{i-j} x_j + (1 - \lambda)^i Z_0 \quad (2.3)$$

$$= \sum_{j=1}^i w_j x_j + (1 - \lambda)^i Z_0 \quad (2.4)$$

where $w_j = \lambda (1 - \lambda)^{i-j}$ is the weight for x_j which falls off exponentially for past observations. The forgetting factor λ determines how fast EWMA forgets the data history. At its extremity, $\lambda = 1$, in which case the EWMA is equal to the most recent observation and gives the same result as Shewhart chart. As λ approaches zero, EWMA approximates the CUSUM criteria which gives the equal weights to the history observations.

The standard deviation of the EWMA (Hunter, 1986) can be shown to be

$$\sigma_{EWMA} = [\lambda / (2 - \lambda)]^{0.5} \sigma \quad (2.5)$$

and the three-sigma control limits for the EWMA can be established as follows

$$\tau \pm 3\sigma_{EWMA} \quad (2.6)$$

where τ is the target value. The process is considered out of control whenever Z_i falls outside the range of the control limits.

Traditional Shewhart, CUSUM and EWMA charts serve as strictly univariate monitoring tools only, i.e., only one variable can be plotted and monitored on one chart. With computers widely hooked up to industrial processes, it is very common

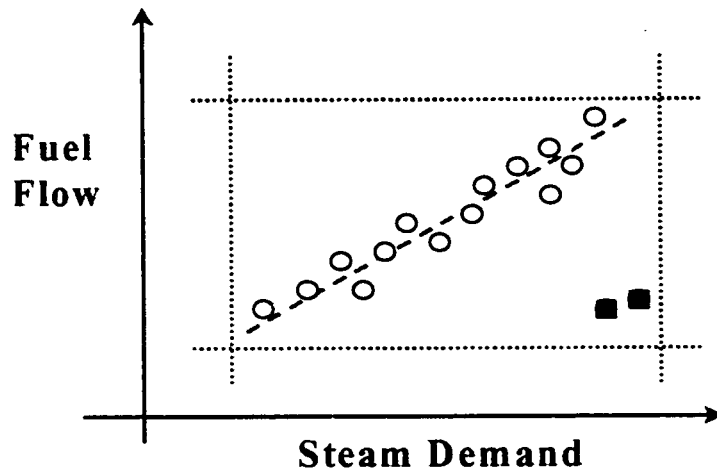


Figure 2.1: Correlation between the steam demand and the boiler fuel flow

that even in a single operating unit, hundreds of variables need to be displayed and monitored. Thus it is difficult to use univariate charts to monitor each process variable individually. Even if one could use such charts, the univariate monitoring charts ignore the correlation between variables and treat each variable independently. However, variables are usually correlated to each other, and their performance should be evaluated relative to other variables. Thus univariate statistics cannot make a correct interpretation and diagnosis of a process. The limitation of using univariate methods can be illustrated by the boiler example shown in Figure 2.1. Increasing the steam demand of boiler should require the corresponding increase of boiler fuel flow. Should the fuel flow measurement fail in such a way that it gives an incorrect reading when the steam flow increases, then the univariate control limits may not be able to detect such a fault. This is clearly seen in Figure 2.1. Although most of the points (○) are around the correlation line (---), two points (●) break up the normal correlation and yet are within the univariate control limits. In this case, the univariate chart fails to detect such a fault.

2.3 Principal Components Analysis

2.3.1 PCA Algorithm

The goal of principal components analysis is to explain the variance/covariance structure through an orthogonal set of linear combination of original variables in the reduced dimensional space. Although n principal components are required to capture the total system variability, due to the dependency and colinearity, usually much of the variation can be accounted for by a only small number of principal components.

Consider a properly scaled data matrix or measurement matrix $X_{m \times n}$, where m represents the number of samples, n represents the number of process variables. In the following discussion, it is assumed that the scaled data is zero mean centered with unit variance.

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ & & \cdots & \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}_{m \times n} \quad (2.7)$$

Let Σ be the covariance matrix associated with these process variables. In PCA the first principal component (latent variable) t_1 is a linear combination of original data that accounts for the maximum variance in the data. Denoting the coefficients of the first linear combination by p_1 (loading vector), then

$$t_1 = Xp_1 \quad (2.8)$$

The procedure of finding p_1 can be expressed as the following optimization problem:

$$\begin{aligned} \max_{p_1} t_1^T t_1 &= p_1^T X^T X p_1 \\ \text{s.t. } p_1^T p_1 &= 1 \end{aligned} \quad (2.9)$$

Note that if there is no constraint on p_1 , then an infinitely large p_1 will maximize $t_1^T t_1$. To avoid this, a unit length constraint of p_1 is usually required. To find the solution of the above optimization problem, introduce a Lagrangian multiplier λ . The constrained problem can then be reformulated as the following unconstrained

problem: find the maximum value of θ , where

$$\max_{p_1} \theta = p_1^T X^T X p_1 - \lambda(p_1^T p_1 - 1) \quad (2.10)$$

The partial derivative of θ w.r.t. p_1 set equal to zero gives

$$\frac{\partial \theta}{\partial p_1} = 2X^T X p_1 - 2\lambda p_1 = 0 \quad (2.11)$$

$$X^T X p_1 = \lambda p_1 \quad (2.12)$$

Now, the problem of finding p_1 has been changed to finding the eigenvector of $X^T X$ with λ as the corresponding eigenvalue. Since the variability explained by the first PC is proportional to λ , p_1 should be the eigenvector associated with the largest eigenvalue of matrix $X^T X$. p_1 is also referred to as the loading vector.

After finding the first PC, the information unexplained by the first PC in the data matrix is

$$E_1 = X - t_1 p_1^T \quad (2.13)$$

E_1 is called residual matrix. The next step is to find a second linear combination to explain the maximum variability in E_1 . Applying the same procedure to E_1 (Equation 2.9–2.12) and adding one more constraint $t_1^T t_2 = 0$, we can find the second loading vector, p_2 . It turns out that p_2 is the eigenvector associated with the second largest eigenvalue of $X^T X$. This procedure can be continued until n principal components are obtained. Usually, due to redundancy and noise in the data, ' A ' principal components ($A \ll n$) can capture much of the variability in X , and X can be expressed in the following way:

$$X = t_1 p_1^T + t_2 p_2^T + \cdots + t_A p_A^T + E = \hat{X} + E \quad (2.14)$$

$$\hat{X} = T P^T = \sum_{i=1}^A t_i p_i^T \quad (2.15)$$

$$E = T_e P_e^T = \sum_{i=A+1}^n t_i p_i^T \quad (2.16)$$

where A is the number of principal components retained in the PCA model. P is the loading matrix with each loading vector p_i as its column. It describes the relationships between variables. T is the principal component score matrix, which describes the observed variables in the transformed basis space spanned by the orthogonal vectors, p_i . Ideally A is chosen such that there is no significant process information left in E where E represents the random error. PCA models are formed by retaining only the loading vectors that describe the systematic variations in the data. Adding extra PCs to the model would only fit random error, and increase the prediction error.

Note that PCA requires proper scaling of raw data in order to obtain reasonable results. If improperly scaled, some unimportant variables may dominate the variance in X matrix due to the units of measurement. In general, the variables should be scaled based on their relative importance. Although it is common to scale the data to have unit variance and zero mean (auto-scaling), caution should be used to scale the variables that are almost constant, since in this case small variations may be highly amplified.

PCA is very closely related to Singular Value Decomposition (SVD) (Jackson, 1991). Through SVD, a data matrix X can be decomposed as:

$$X = U\Sigma V^T \quad (2.17)$$

Here, columns of V are equal to the PCA loading vectors and Σ is a diagonal matrix containing the singular values, which are equal to the square roots of the eigenvalue of the covariance matrix of X .

If one loading vector is plotted vs. another (for example p_1 vs. p_2), a loading plot is obtained which shows the relationship between the original variables and the principal components. The variables that are highly correlated to each other tend to cluster together. One can thus infer about the relative importance and influence of the original variable through this plot.

A scores plot is formed if one score vector is plotted vs. another (for example t_1 vs. t_2). Score plots reveal the relationship between the observations or samples and the principal components. By observing the changes in the score plots, one can monitor drifts and other abnormal events in the process.

2.3.2 Optimal Dimension of a PCA Model

In PCA, it is very important to select the optimal number of PC's to be retained in the model. There are many approaches for selecting the dimension A , such as cross-validation (Wold, 1978), cumulative percent variance (Malinowski, 1991), Xu and Kailath's approach (Xu and Kailath, 1994), Akaike information criterion (Akaike, 1974; Wax and Kailath, 1985), minimum description length criterion (Rissanen, 1978), and variance of reconstruction error (Qin, 1998). Some of these approaches are described below:

1. Cumulative Percent Variance

Cumulative Percent Variance measures the percent variance captured by the first k PC's, which can be expressed:

$$CPV(k) = \frac{\sum_{j=1}^k \lambda_j}{\sum_{j=1}^n \lambda_j} 100\% \quad (2.18)$$

k principal component is chosen if the $CPV(k)$ can explain a predetermined variance, say 95%.

2. Eigenvalue One Criterion

Only those PC's whose variances (equals to the corresponding eigenvalues of $X^T X$) are greater than one are retained in the model.

3. Average Eigenvalue

This approach selects the eigenvalues which are greater than the mean of all eigenvalues and discards eigenvalues smaller than the mean.

4. Cross Validation

The whole data set is divided into several blocks. Each time, one block of data is left out, and a PCA analysis is performed on the remaining blocks of data and the prediction error sum of squares (PRESS) is calculated based on the data block which is left out. The procedure is repeated until each block of data has been left out once. Adding all the resulting PRESS together, gives a cumulative PRESS. The optimal order of PCA is the order that minimizes

the cumulative PRESS. Although this method is lengthy and has no sound statistical background, it is effective in practical use. More references on this method are given by Stone (1978) ; Eastment and Krzanowski (1982) ; Geladi and Kowalski (1986b).

2.3.3 PCA for Multivariate Statistical Process Control (MSPC)

Once a PCA model based on data representing historical normal operation is obtained, it can be utilized to monitor future deviation of the process from normality. Let $x_{new}(1 \times n)$ denote the multivariate observation from the process. Projecting it to the latent PCA space gives the scores $t_{new}(1 \times A)$:

$$t_{new} = x_{new}P \quad (2.19)$$

Prediction of the original variables from t_{new} is given by:

$$\hat{x}_{new} = t_{new}P^T = x_{new}PP^T \quad (2.20)$$

The model residual is:

$$e_{new} = x_{new} - \hat{x}_{new} \quad (2.21)$$

Two dimensional score plots (usually t_1 vs. t_2), Hotelling T^2 , and Q statistics (SPE) are usually used to monitor the process. If the first two principal components can explain a large part of the variance, the two dimensional scores plot provides an intuitive insight into process variations. Any abnormal shift in the process will cause the scores (t_{new}) to move out of the confidence limits in the 2-D score plots.

However, the T^2 chart is more appropriate if the variance explained by the first two PCs is relatively small. The T^2 statistic based on the first A PCs is defined as

$$T^2 = \sum_{i=1}^A \frac{t_i^2}{\lambda_i} \quad (2.22)$$

where λ_i is eigenvalue of the covariance matrix of X . The T^2 statistic given in Equation (2.22) can be considered as an ellipsoid in a A dimensional space.

Confidence limits for T^2 at confidence level $(1 - \alpha)$ relate to the F -distribution as follows:

$$T_{m,A}^2 = \frac{(m-1)A}{m-A} F_{A,m-A} \quad (2.23)$$

where $F_{A,m-A}$ is the upper $100\alpha\%$ critical point of the F distribution with A and $m - A$ degree of freedom (Tracy *et al.*, 1992).

Variations in the process could be associated with the breakdown of the correlation structure among the measured variables while still within the confidence regions of T^2 chart. For this reason, monitoring the process only via T^2 chart is not sufficient. To overcome this problem, the SPE chart is also used in conjunction with the T^2 chart.

Let $x_i(1 \times n)$ denote an observation whose prediction from PCA model is given by $\hat{x}_i = x_i P P^T$. The p dimensional error vector is given by $e_i = x_i - \hat{x}_i$. The SPE is then defined as

$$SPE = e_i e_i^T \quad (2.24)$$

SPE can be considered as a measure of the plant-model mismatch. The confidence limits for SPE are given by Jackson and Mudholkar (1979). This test suggests the existence of an abnormal condition when $SPE > Q_\alpha$, where Q_α is defined as:

$$Q_\alpha = \Theta_1 \left[1 + \frac{c_\alpha h_0 \sqrt{2\Theta_2}}{\Theta_1} + \frac{\Theta_2 h_0 (h_0 - 1)}{\Theta_1^2} \right]^{\frac{1}{h_0}} \quad (2.25)$$

where

$$\Theta_i = \sum_{j=p+1}^n \lambda_j^i; \text{ for } i = 1, 2, 3 \quad (2.26)$$

$$h_0 = 1 - \frac{2\Theta_1\Theta_3}{3\Theta_2^2} \quad (2.27)$$

c_α is the confidence limits for the $1 - \alpha$ percentile in a normal distribution.

Once an abnormal event is detected in a SPE chart, the SPE contribution plot can usually be utilized to isolate the fault. The fractional contribution of each process variables to the overall SPE at sample instant i can be computed as:

$$\alpha_j = \frac{SPE_j}{SPE} \quad (j = 1, 2, \dots, n) \quad (2.28)$$

where SPE_j denotes the square of the j^{th} element of the error vector e_i . If the contribution from some variables is significant, then these variables are most likely to be the cause of the abnormality (Miller *et al.*, 1993). Although the contribution plots cannot unequivocally diagnose the cause, they will provide much greater insight into possible cause and thereby greatly narrow the search (MacGregor and Kourti, 1995).

2.4 Partial Least Square

2.4.1 PLS Algorithm

Consider a dependent data matrix Y ($m \times j$) and an independent data matrix X ($m \times n$), where m is the number of samples or observations, j and n are the number of dependent and independent variables, respectively. Y and X may be related by

$$Y = X\theta + \text{error} \quad (2.29)$$

If the matrix $X^T X$ is non-singular, then the best estimate of θ in a least square sense is given by the well known least square regression:

$$\theta = (X^T X)^{-1} X^T Y \quad (2.30)$$

If the matrix $(X^T X)$ is singular, or nearly so, the direct inversion of $X^T X$ is impossible or error prone, and another representation of the inverse has to be found.

• PLS1

The objective of PLS is to maximize the covariance matrix between linear combinations of X and Y block. For the sake of simplicity, we first consider the case where only one dependent variable is in Y block, thus $Y = y$. This is also known as *PLS1* in the literature. Any vector t in the column space of X can be written as a linear combination of column of X , $t = Xw$ for some vector w . The target here is to find a w such that t is good at describing Y , i.e. such that it maximizes the correlation between X and Y . Description means a linear measure between Y and X . The choice of such a linear measure made by PLS is $x_i^T y$, where x_i is the i^{th} column of X . w is determined as $w = (w_1, w_2, \dots, w_n)$ with $w_i = (x_i^T y)$. In terms of X and Y , we can write w as $w = X^T Y$. In the numerical computations w is scaled to have unit length to ensure numerical stability.

$$w = X^T Y / \|X^T Y\| \quad (2.31)$$

When the first score vector is computed as $t_1 = Xw$, we want to determine second score vector that is orthogonal to the previous one. We do this by computing the residual matrix

$$E_1 = X - t_1 p_1^T \quad (2.32)$$

here p_1 is the rotated w and defined by

$$p_1 = \frac{X^T X w}{t_1^T t_1} = \frac{X^T t_1}{t_1^T t_1} \quad (2.33)$$

The second score vector is computed as $t_2 = E_1 w_{new}$. The residual matrix formed in Equation 2.32 ensures that different t vectors form an orthogonal basis due to the fact

$$t_1^T t_2 = t_1^T E_1 w_{new} = t_1^T (X - t_1 p_1^T) w_{new} = (t_1^T X - t_1^T X) w_{new} = 0 \quad (2.34)$$

This procedure is continued until the newly computed vector t_{new} does not contribute much in describing y .

• PLS2

We now discuss the case where two or more dependent variables are in the Y block which is also known as PLS2 problem. The solution to this problem is described under the optimization framework as follows:

$$\begin{aligned} \max_{w_1, q_1} \text{cov}(t_1, u_1) &= t_1^T u_1 = w_1^T X^T Y q_1 \\ \text{s.t. } w_1^T w_1 &= 1 \\ q_1^T q_1 &= 1 \end{aligned} \quad (2.35)$$

where $t_1 = X w_1$ and $u_1 = Y q_1$ are linear combination of X and Y respectively.

To solve the above problem, two Lagrangian multipliers λ_1 and λ_2 are introduced and an unconstrained optimization problem is formed as to find the maximum value of Θ , where Θ is

$$\Theta = w_1^T X^T Y q_1 - \lambda_1 (w_1^T w_1 - 1) - \lambda_2 (q_1^T q_1 - 1) \quad (2.36)$$

Let the first derivative of Θ with respect to w_1^T and q_1^T be set to zero, i.e.

$$\begin{aligned} \frac{\partial \Theta}{\partial w_1^T} &= X^T Y q_1 - 2\lambda_1 w_1 = 0 \\ \frac{\partial \Theta}{\partial q_1^T} &= Y^T X w_1 - 2\lambda_2 q_1 = 0 \end{aligned} \quad (2.37)$$

We then have

$$X^T Y q_1 = 2\lambda_1 w_1 \quad (2.38)$$

$$Y^T X w_1 = 2\lambda_2 q_1 \quad (2.39)$$

Solving the combined Equation 2.38 and 2.39, we obtain the following eigenvalue problems:

$$X^T Y Y^T X w_1 = 4\lambda_1 \lambda_2 w_1 \quad (2.40)$$

$$Y^T X X^T Y q_1 = 4\lambda_1 \lambda_2 q_1 \quad (2.41)$$

w_1 is expressed as an eigenvector associated with the largest eigenvalue of $X^T Y Y^T X$, and q_1 is an eigenvector associated with the largest eigenvalue of $Y^T X X^T Y$.

If $X^T Y = U S V^T$ through SVD decompositions, w_1 and q_1 can be proved (Hoskuldson, 1988, Manne, 1987) to be the first column of matrix U and V . Once t_1 and u_1 are obtained by $t_1 = X w_1$ and $u_1 = Y q_1$, a linear inner relation between u_1 and t_1 are found to be

$$\hat{u}_1 = b_1 t_1 \quad (2.42)$$

where $b_1 = u_1^T t_1 / t_1^T t_1$, is the linear regression coefficient. The X and Y blocks are indirectly related through the inner relation between u_1 and t_1 .

$\hat{u}_1 q_1^T$ can be interpreted as the part of the information in Y block that has been explained by the first PLS dimension. Similarly $t_1 p_1^T$ represents the information used up in X block. The X and Y block residuals can be further calculated as follows:

$$E_1 = X - t_1 p_1^T \quad (2.43)$$

$$F_1 = Y - b_1 t_1 q_1^T$$

where p_1 is defined similarly as in Equation 2.33 to ensure that the next score vector is orthogonal to the previous one. The above procedure is continued until the useful information in Y cannot be explained further by adding one more latent variable. Figure 2.2 (Lakshminarayanan, 1997) illustrates this iterative procedure.

In practice, how to choose a proper number A for the latent variables is a crucial step in PLS. If all the latent variables are used in modeling, the model may fit the noise and therefore reduce the predictive ability of the model. The cross validation method has been suggested by S. Wold (1978) to determine a proper number of latent variables.

In matrix form, the total information explained in Y block can be expressed as:

$$Y_{\text{exp}} = b_1 t_1 q_1^T + b_2 t_2 q_2^T + \cdots + b_A t_A q_A^T = T B Q^T \quad (2.44)$$

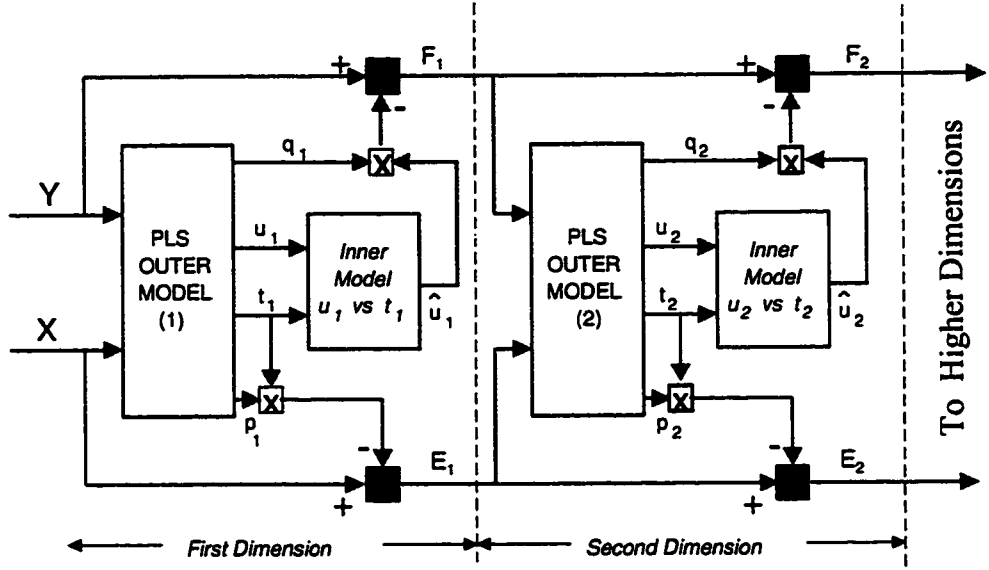


Figure 2.2: A schematic diagram of PLS regression procedure.

where B is a diagonal matrix with b_1, b_2, \dots, b_A as the diagonal elements, $T = [t_1 \ t_2 \ \dots \ t_A]$, $Q = [q_1 \ q_2 \ \dots \ q_A]$.

Note that both in *PLS1* and *PLS2*, the weighting vector w_i ($i = 2, 3, \dots, A$) are directly related to different residual matrix E_i through Equation 2.45

$$t_i = E_i w_i, \quad i = 2, 3, \dots, A \quad (2.45)$$

and not to the original data matrix X . This obscures the relation between the weighing vectors and original variables. Suppose R represents the relation between X and T :

$$T = XR \quad (2.46)$$

Since each latent vector t_i , $i = 1, 2, \dots, A$ lies in the column space of X , R can be computed by regressing T on X (de Jong, 1993):

$$R = X^\dagger T \quad (2.47)$$

where X^\dagger is the pseudo-inverse of X . R has been shown to have the following relation with W and P (Helland, 1988):

$$R = W(P^T W)^{-1} \quad (2.48)$$

Combining Equation 2.46 and 2.44, we have

$$Y = XRBQ^T + F \quad (2.49)$$

$$Y_{\text{exp}} = XC_{\text{pls}} \quad (2.50)$$

where F is the final residual matrix and $C_{\text{pls}} = RBQ^T$, is the PLS regression coefficient matrix.

In the classical PLS algorithm the sequential calculation of PLS dimension is done iteratively. In general, iterative algorithms often become inefficient when data structure is large. To overcome this difficulties, a fast and memory-saving PLS regression algorithm for data matrices with large number of samples was proposed by Lindgren *et al.* (1993). de Jong (1993) developed another novel algorithm for PLS, SIMPLS, which can calculate the PLS factors directly as the linear combination of the original variables.

2.4.2 Nonlinear and Dynamic PLS Model

Although PLS provides a robust calculation method for correlated data, its major restriction is that only linear information can be extracted from the data. Since many practical data is inherently nonlinear in nature, it is desirable to have a method to model nonlinear relations between variables. One possible approach is to include the nonlinear transformation of original variables in X block (such as exponential, square root and logarithm). However, this approach suffers from the requirement of priori knowledge of the process and expansion of data matrix.

Another approach is to move the nonlinearity into PLS inner model. Wold *et al.* (1989) use the quadratic model for inner relation between u_i and t_i . This method is limited to handling the quadratic inner relations and ineffective for other types of nonlinearity. Qin and McAvoy (1992a) proposed to incorporate neural network into the PLS modeling. In their method, a multi-input multi-output nonlinear modeling task is decomposed into linear outer relations and simple nonlinear relations which are performed by a single-input and single-output networks. The proposed neural net PLS (NNPLS) gives better results than the PLS modeling method and the direct neural network approach.

To deal with dynamics in a process, we can include the past values of the input and output variables in the X block. In this case, C_{pls} in Equation 2.49 can be interpreted as finite impulse response (FIR) coefficients (Ricker, 1988) or as a multivariate autoregressive moving average (ARMA) model (Qin and McAvoy, 1992b). The disadvantage of the method is that the dimension of X can be very large, especially with many input variables. Kasper and Ray (1993) developed a modified PLS modeling procedure which permits the dynamics to be expressed as part of the inner relation and obviates inclusion of lagged values of the input and output variables in the input data matrix. Moreover, the approach can optimize the dynamic transformation so that the resulting model has maximum predictive ability. Lakshminarayanan *et al.* (1997) suggested that the inner relation in PLS (t_i and u_i) can be captured by a dynamic ARX model instead of the static model to deal with a dynamic system.

2.5 Conclusions

Data when properly analyzed by statistical tools yield meaningful information. Univariate statistics are inappropriate in handling large amounts of multivariate data from processes and may not allow extraction of useful information from the data. To overcome this difficulty, multivariate statistical methods, such as PCA and PLS, have become important tools for data mining. These methods treat all the data simultaneously and explain the correlation between the data in a reduced dimensional space. The algorithms of PCA and PLS and some related issues are discussed in this chapter. In the following chapters, we will further demonstrate how these multivariate tools can be effectively used in statistical modeling and monitoring areas.

Chapter 3

PCA in Statistical Process Monitoring— A Simple Example

3.1 Introduction

The basic idea of using PCA for statistical process monitoring (SPM) is to build a PCA model based on the data when the system is operating at a nominal or normal state. This model can then be used to monitor the possible disturbance, sensor failure, and other abnormal events in the process that may lead the process to depart from its “normal” state. In SPM, usually two statistics are employed. One is the Hotelling T^2 statistic, which measures the variation within the PCA model. The other is the Q statistic, also known as the square prediction error (SPE), which indicates how well the measurements conform to the PCA model or how large is the deviation from the nominal model. An abnormal event is encountered if some points exceed the confidence limits in either the SPE or T^2 charts.

In this chapter, a simulation example is utilized to illustrate the application of PCA in SPM. The chapter is organized as follows: In section 2.2, the steps required in building a PCA model for a simple flow rate system is outlined. The use of this model to detect various sensor faults is illustrated in section 2.3. In section 2.4, we discuss the limitation of using PCA as a fault isolation tool. The possible ways to overcome this shortcoming are also proposed in this section. The importance of minor principal components are emphasized in section 2.5. The chapter ends with concluding remarks in section 2.6

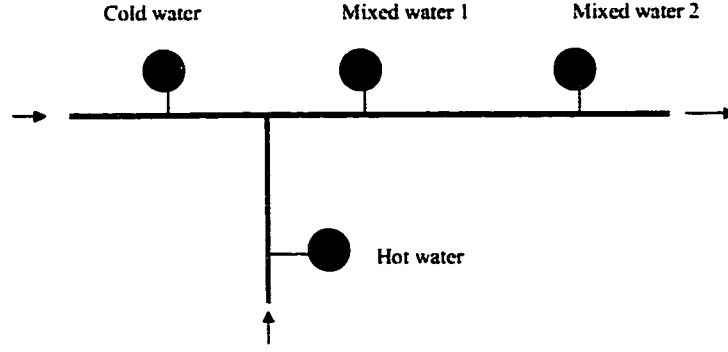


Figure 3.1: Configuration of flow rate system

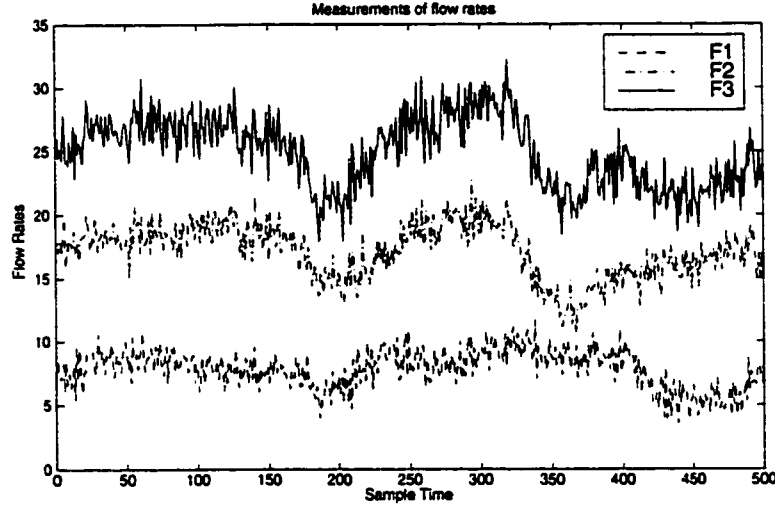


Figure 3.2: Snapshot of flow rate measurements

3.2 Building of PCA Model

Let us consider a flow rate system, as described in Figure 3.1. F_1 , F_2 are cold and hot water flow rates measurements respectively. F_3 is the mixed water flow measurement. For incompressible fluids with no storage capacity, it is expected that $F_3 = F_1 + F_2$. F_4 is a redundant measurement of F_3 .

F_1 , F_2 are two independent measurements. F_3 and F_4 are clearly dependent on the F_1 , F_2 . The snapshot of four flow measurements are displayed in Figure 3.1. Only three measurements are displayed here, since F_4 is only slightly different from F_3 due to the presence of measurement noise.

These four flow rate signals (0-500) are considered to be measured when the system

is operating at normal condition. A PCA model is built based on the data compiled from the measurement matrix arranged as follows:

$$X = [F_1(t) \quad F_2(t) \quad F_3(t) \quad F_4(t)] \quad (3.1)$$

Two linear relationships exist among flow rate measurements if there is no measurement noise:

$$F_1(t) + F_2(t) = F_3(t) \quad (3.2)$$

$$F_3(t) = F_4(t) \quad (3.3)$$

Strictly speaking, $F_3(t) \neq F_1(t) + F_2(t)$ and $F_3(t) \neq F_4(t)$ due to the measurements noise.

First, all the measurements are auto scaled to have unit variance and zero mean, i.e.

$$F_s = \frac{F - \bar{F}}{\sigma_F} \quad (3.4)$$

where the subscript s denotes the scaled measurements, \bar{F} is the sample average and σ_F is the sample standard deviation. The covariance matrix is then calculated based on the scaled data:

$$\text{cov}(X) = \frac{X_s^T X_s}{n - 1} \quad (3.5)$$

Four eigenvalues of covariance matrix are shown in Figure 3.3. As would be expected, only two are significant. This is so because only two of the vectors in X_s are independent. Columns 3 and 4 are linear combination of columns 1 and 2. The first PC explains 78.34% of the total variance; the second PC explains 21.47% of the total variance. Totally, two PCs can capture almost 99% of all the useful information in the system (Figure 3.4).

Thus, only two PCs need to be retained in the PCA model. The *SPE* and Hotelling T^2 charts for the normal operation data are shown in Figure 3.5 and 3.6.

Since the *SPE* plot shown here is based on normal operating data, one should expect that almost all the data will lie within 99% confidence interval. Similarly, the data points in the Hotelling T^2 chart are also within the 99% confidence limits.

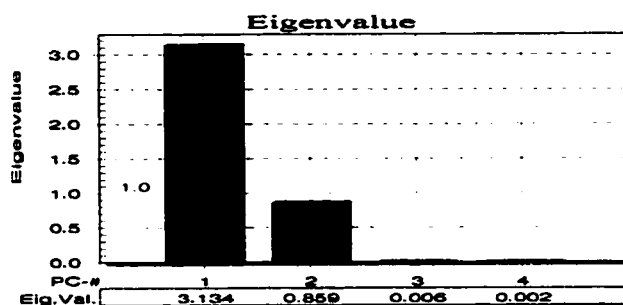


Figure 3.3: Eigenvalues of covariance matrix

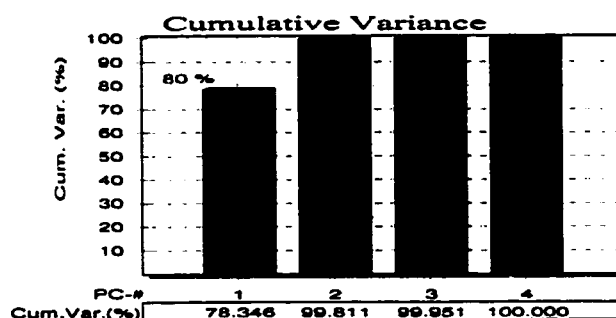


Figure 3.4: Two PCs can capture 99% of information in the system.

3.3 Sensor Fault Detection

3.3.1 Abrupt Sensor Fault Detection

In this section, an abrupt sensor failure is simulated by adding a small constant deviation (the magnitude of the deviation is equal to one in this example) to F_3 between sample times 300–400. This could represent a sudden sensor offset or mis-calibration. The measurements of three flow rates are displayed in Figure 3.7. Clearly, it is not easy to detect the sensor failure by observing the trend plot of the raw measurements.

Applying the PCA model, as generated from the nominal process, to monitor the system, the Hotelling T^2 and *SPE* charts are displayed in Figures 3.8–3.9. The

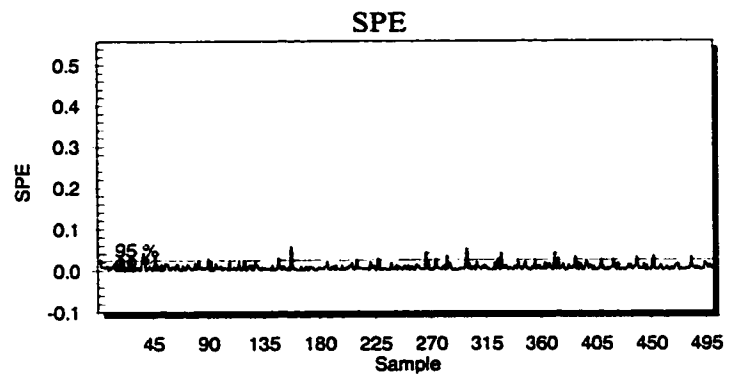


Figure 3.5: The SPE plot for the normal operation data.

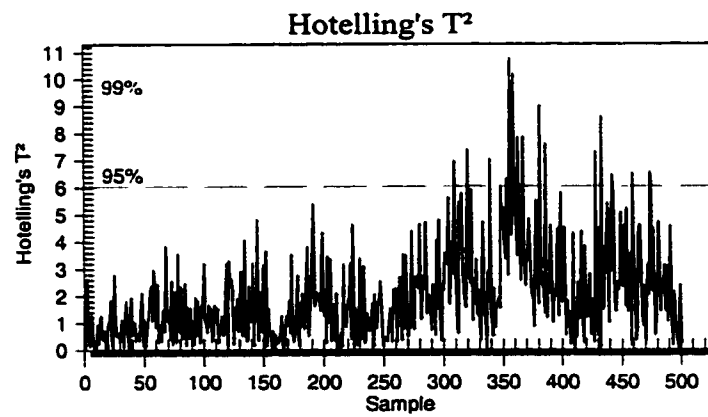


Figure 3.6: Hotelling T^2 chart for the normal operation data.

95% and 99% detection limits are also shown in these figures. Many measurement points are out of 95% limits in the T^2 chart. However, in the SPE chart, only the measurements between 300-400 are out of detection limits. How should we interpret these results?

The data set used to build the PCA model is different from the data set used for validation. F_1 and F_2 are two independent flow rates. They are always changing. Thus, the system is in a varying state. Whenever the system shifts from the normal operating states (defined by the normal measurements used for the PCA model), the measurements will flare up in T^2 chart. Several points being out of control limits in Figure 3.8 implies that the system has drifted beyond previous operation. However, we notice that the T^2 chart is not sensitive in detecting the sensor fault. (All points between the 300-400 sample rates are within the control limits). This is due to the small magnitude of the fault which is comparable to the dynamic transients in the process. In other words, the variability of the process and the noise “mask” the sensor fault.

Although the system may drift away from the normal operating point, if the relationships between the measurements continue to hold, only small prediction errors should be expected. The sample points before 300 and after 400 in *SPE* chart show this case (Figure 3.9). However, when the sensor fault occurs during 300-400 sample intervals, the correlations between variables break down. This results in a large *SPE* which can be seen clearly in Figure 3.9.

Once an abnormal situation is detected in the system, the *SPE* contribution plot usually can be utilized to isolate the fault. Figure 3.10 shows the *SPE* contribution plot at sample instant 398. The percent contribution related to the mixed water 1 (F_3) is 68.61%, which is quite large compared to the contributions from other measurements. This indicates that F_3 is responsible for the abnormality, i.e. this sensor is likely to be faulty.

3.3.2 Ramp-type Sensor Failure Detection

A ramp-type or a slow drift sensor failure is simulated by adding a ramp change to normal measurements of F_4 between 100-300 sample intervals (Figure 3.11). Again, it is not easy to detect the fault by observation. However, the *SPE* monitoring chart

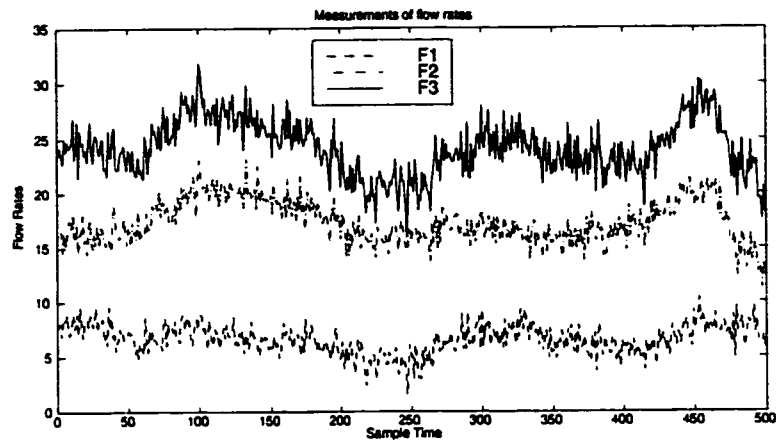


Figure 3.7: Measurements of flow rates with sensor failure between 300-400.

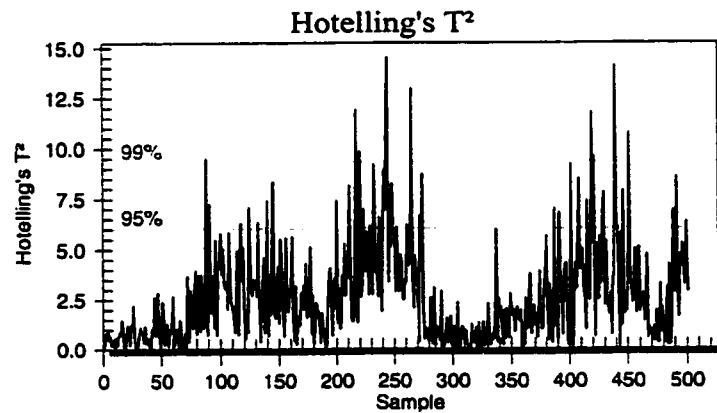


Figure 3.8: Hotelling T^2 monitoring chart for the system with abrupt sensor failure.

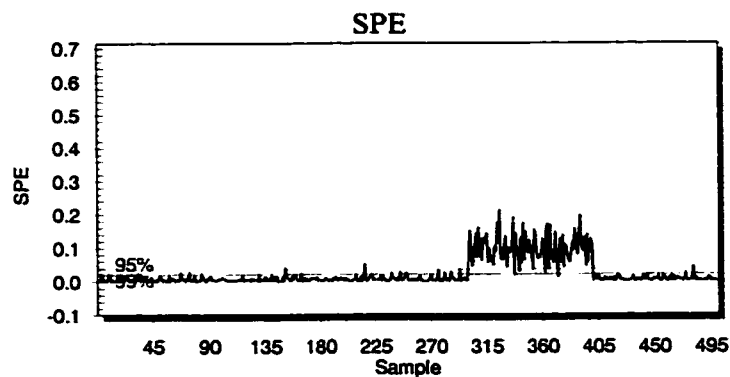


Figure 3.9: SPE chart for the system with abrupt sensor failure

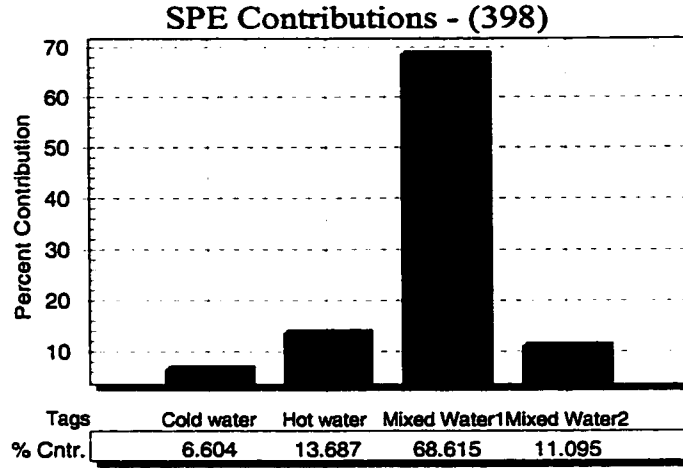


Figure 3.10: *SPE* contribution plot at sample instant 398 (Abrupt sensor fault).

(Figure 3.12) indicates that an incipient fault has occurred. To diagnose the cause for the fault, the contribution plot was checked picking up F_4 as the candidate (Figure 3.13).

3.4 Limitations of PCA Model for SPM

In this section, we will show the limitation of PCA model as a diagnostic tool through the simulated flow rate example. Possible remedies to overcome these shortcomings are also proposed and discussed.

Let us consider the following scenario for the flow rate system. Suppose that only three flow rates are measured (F_1, F_2, F_3). An abrupt sensor failure related to measurements of F_1 occurs between samples 100-300 (Figure 3.14). The *SPE* monitoring chart does indeed detect such a fault because the relationship between the variables breaks down (Figure 3.15). However, the *SPE* contribution plot at sample instant 249 (Figure 3.16) considers F_3 as the possible cause, since it indicates the largest contribution to *SPE*. In this case, the *SPE* contribution statistic is unable to identify the correct faulty sensor. Why?

Note that there are three variables in the system, but only one linear relationship between them ($F_1 + F_2 = F_3$). The diagnosis mechanism of PCA requires redundancies among the measured variables. Without enough redundancies, the PCA model

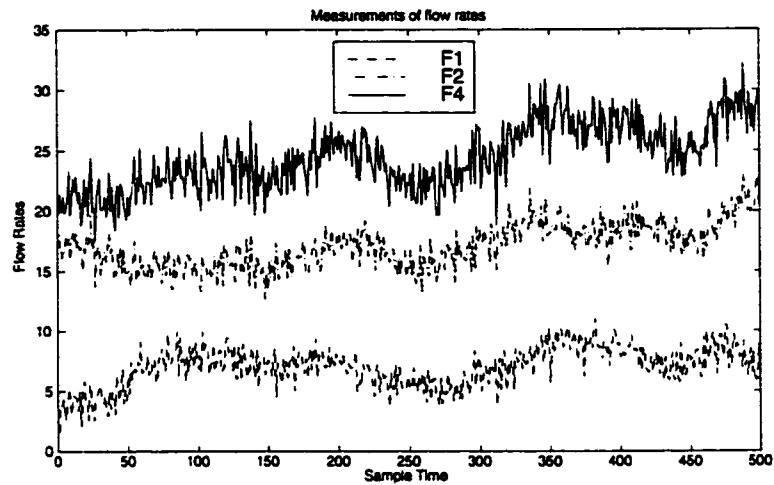


Figure 3.11: The measurements for flow rates with incipient sensor fault.

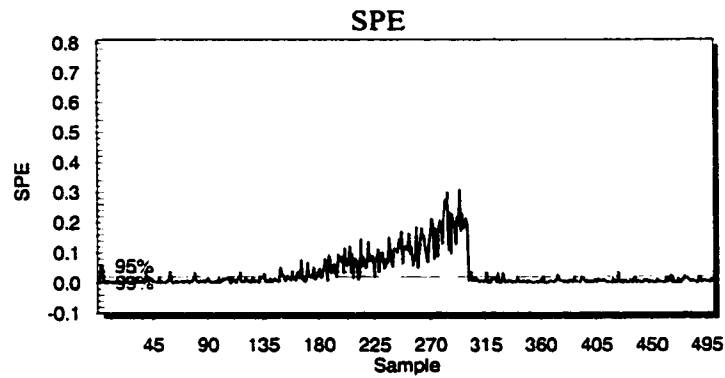


Figure 3.12: The *SPE* plot for the flow rate measurements with incipient sensor failure.

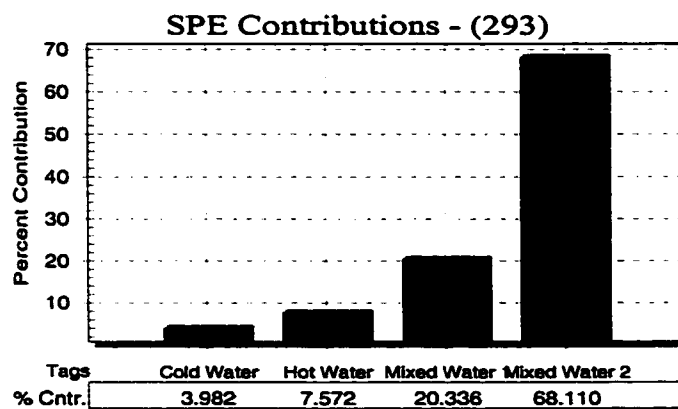


Figure 3.13: The *SPE* contribution plot at sample 293 (incipient sensor fault).

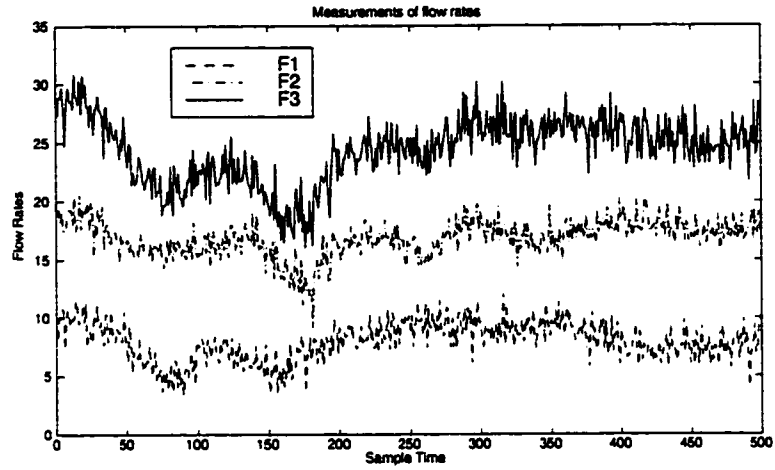


Figure 3.14: Flow rates measurements with the sensor fault in F_1 between 100-300.

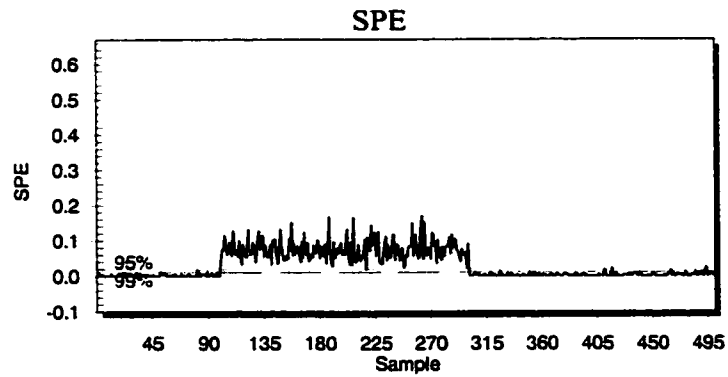


Figure 3.15: The *SPE* monitoring chart detect the abnormal situation between 100-300

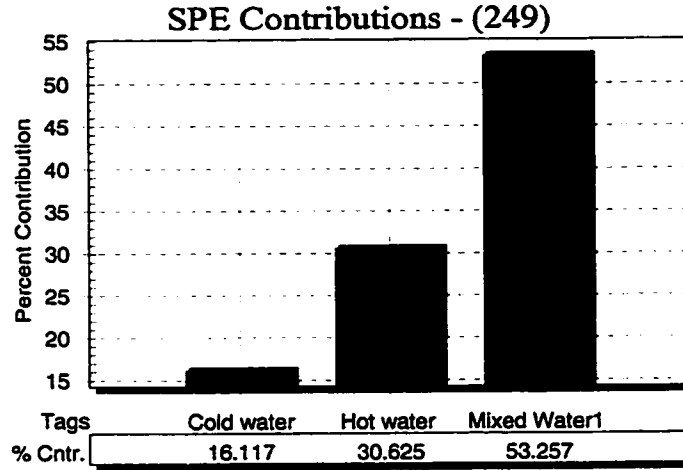


Figure 3.16: *SPE* contribution plot at sample instant 249 (Ramp sensor fault)

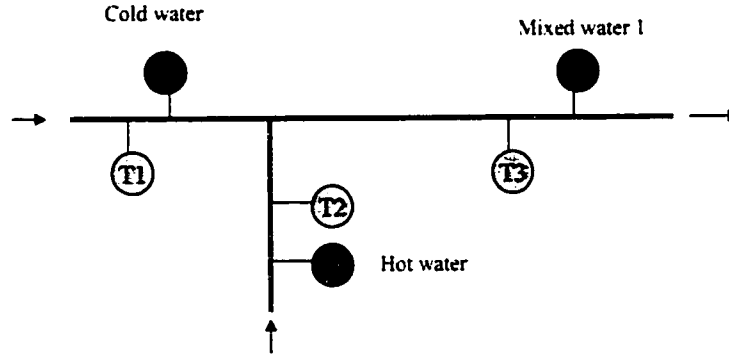


Figure 3.17: Three temperature measurements are added to the flow rate system.

gives the wrong diagnostic result.

To overcome this shortcoming, more constraints need to be added to the system. In this example, we can either add redundant measurements for F_1 or utilize the temperature measurements in the system. A more detailed schematic plot for the flow rate system, together with the temperature measurements is shown in Figure 3.17.

The new measurement matrix X is then arranged as follows:

$$X_{new} = [F_1 \quad F_2 \quad F_3 \quad F_1 T_1 \quad F_2 T_2 \quad F_3 T_3] \quad (3.6)$$

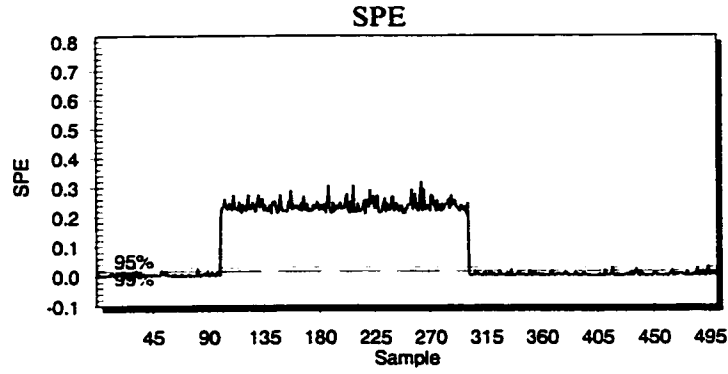


Figure 3.18: SPE plot for the new PCA model with temperature measurements.

Now two linear relationships exist in X_{new} :

$$F_1 + F_2 = F_3 \quad (\text{Continuity equation assuming constant density}) \quad (3.7)$$

$$F_1 T_1 + F_2 T_2 = F_3 T_3 \quad (\text{Energy balance equation assuming constant heat capacity})$$

A new PCA model based on the new measurement matrix can then be used to diagnose sensor faults more precisely. The same sensor fault as introduced in F_1 (cold water) between samples 100-300 is now simulated again. The *SPE* plot is shown in Figure 3.18. The *SPE* clearly flares up in the plot indicating the fault. Comparing to the *SPE* plot for the flow rate system (Figure 3.15), the *SPE* for the new model is more sensitive. The reason is that there are now two terms in X_{new} related to F_1 . Both terms contribute to the *SPE*.

The contribution plot for *SPE* (Figure 3.19) shows that the measurements related to F_1 and $F_1 T_1$ is the source of the fault. If we assume that the most likely cause of the fault is failure of only one sensor at a time, then the sensor for F_1 is the candidate. Since it is highly unlikely that both sensors (F_1 and T_1) could have failed at the same time, we are first led to investigate the case where F_1 is likely the source of the problem.

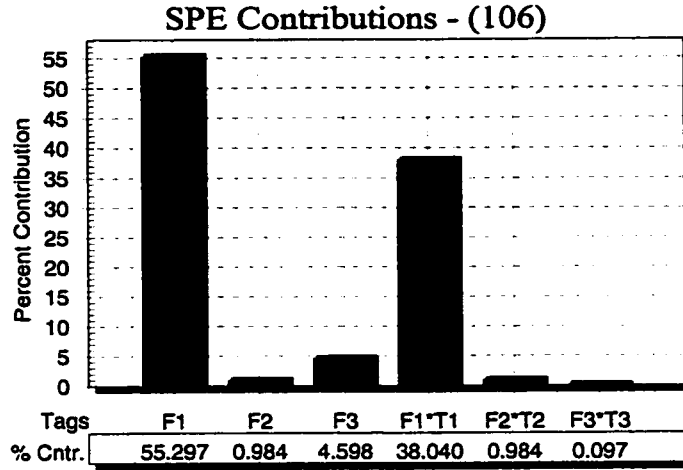


Figure 3.19: *SPE* contribution plot at sample time 295.

3.5 Assembly of the measurement matrix

In the previous discussion, we have ignored the time delay possibly caused by the transportation lag. Physically there is no delay in a system when incompressible fluid is being transported without any accumulation capacity. The data considered here is simulated data where a delay has been inserted for illustration purpose only and we assume this delay causes a lag between the measurements of F_3 and F_4 . What happens if there is a large time delay between F_3 and F_4 ?

In such case, caution should be used in arranging the measurement matrix so that the correct time delay can be taken into account. This point is illustrated by the following example.

Assume that the time delay between F_3 and F_4 is 4 sample intervals. Consider two measurement matrices:

$$\begin{aligned}
 X1 &= [F_1(t) \quad F_2(t) \quad F_3(t) \quad F_4(t)] \\
 X2 &= [F_1(t) \quad F_2(t) \quad F_3(t) \quad F_4(t+4)]
 \end{aligned} \tag{3.8}$$

Two ways of arranging the data matrix are considered here. $X1$ contains the four flow rate measurements, but the delay term in F_4 is ignored, while $X2$ includes a delay term in F_4 . The variance captured by the two PCA modes is listed in Table 3.1. Two

<i>PCA1</i> (based on <i>X1</i>)			
Principal Component #	Eigenvalue of $cov(x)$	% Variance Captured by This PC	% Variance Captured Total
1	2.77	69.29	69.29
2	0.863	21.58	90.87
3	0.362	9.05	99.92
4	0.003	0.08	100
<i>PCA2</i> (based on <i>X2</i>)			
1	3.14	78.38	78.38
2	0.87	21.43	99.82
3	0.00543	0.14	99.95
4	0.00194	0.05	100

Table 3.1: Comparison of variance captured by two PCA models

principal components in *PCA1* capture 90.87% of the total variance, whereas two PCs can capture 99.82% of the total variance in *PCA2*. With a properly arranged matrix, i.e. the raw measurements lagged correctly as in *X2*, the same number of PCs is able to capture even more information in the measurement matrix.

Two PCs are retained in both models since there are two linear relations between variables. The *SPE* charts for both models are displayed in Figure 3.20. We can see a fairly large number of points out of the 95% confidence interval in the *SPE* chart for *PCA1*, which implies that a significant amount of information remains unexplained in the residual space. In contrast, the *SPE* in *PCA2* is small. This means that almost all the information can be explained by the *PCA2* model. (only 1% of variance remains unexplained).

Further comparison of third principal component in each model (Figure 3.21) indicates that the variance of third PC in *PCA1* is much bigger than that in the *PCA2*. The autocorrelation function (ACF) of the third score (the residual score) in *PCA2* (Figure 3.22) is almost zero which means the third PC is true white noise, and therefore truly represents white-noise residual terms that cannot be captured by any other dynamics. Note that the ACF of the third PC in *PCA1* is significant at lag 4, which is exactly the same as the time delay value in the system. This match is not a coincidence. Further experiments show that the significant value in ACF of the third PC does have a relationship with the time delay. For example, if the time delay is

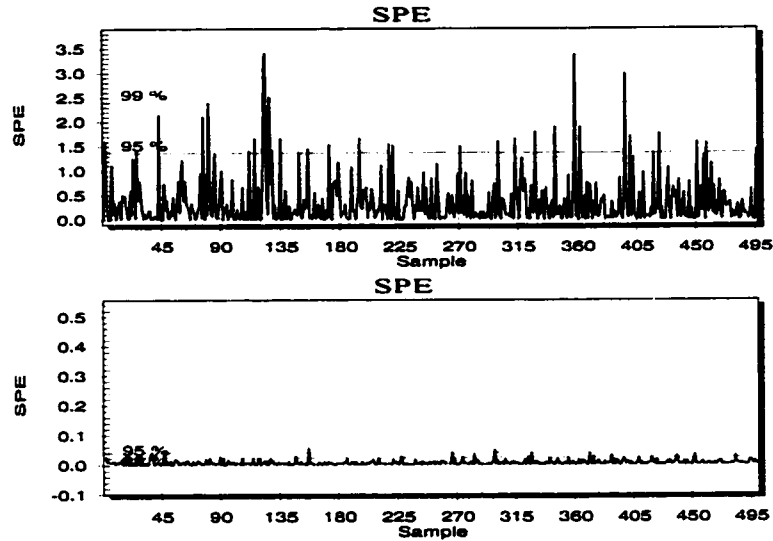


Figure 3.20: The SPE chart (top) is for $PCA1$; SPE chart (bottom) is for $PCA2$.

set at one sample interval, the ACF of the third PC (Figure 3.23) clearly flares up at lag 1. This relationship thus provides an alternative to find the time delay in the system.

In summary, the $X2$ matrix is correctly assembled and takes the process delay into account. The PCA model based on this matrix can explain the true relationship between variables ($F_4(t) = F_3(t - 4)$). On the other hand, the $X1$ matrix is not correctly assembled. As a result the PCA model cannot capture the true relationship. This model is therefore unreliable for process monitoring.

To build a correct PCA model, it is important to include lagged variables when time delays exist.

3.6 Minor Components

Minor components refer to the components that are not retained in a PCA model. As we know, the principal components explain most of variation in the data. Only unimportant or residual information is left in minor components. Minor components usually contain residuals for the constraints or redundancies that exist between variables. Few studies (Schoukens *et al.*, 1997) have taken the minor components into account when doing PCA analysis. The loading vectors related to the minor compo-

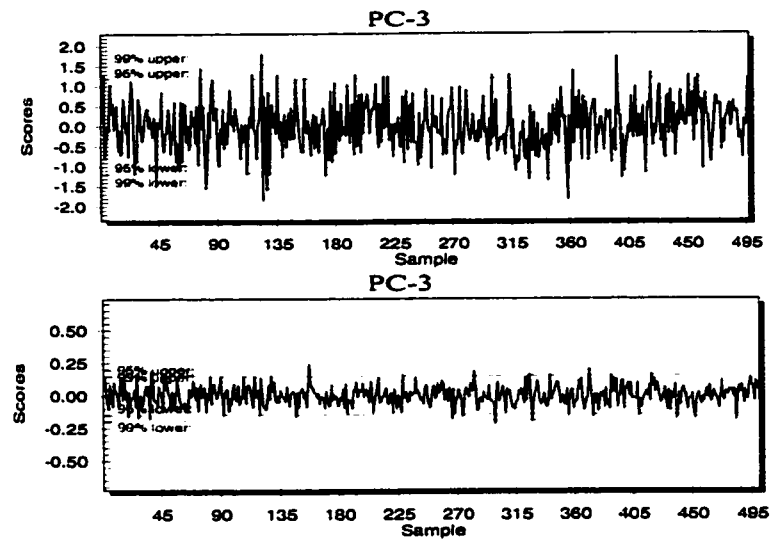


Figure 3.21: Third principal component in *PCA1* (top); Third principal component in *PCA2* (bottom)

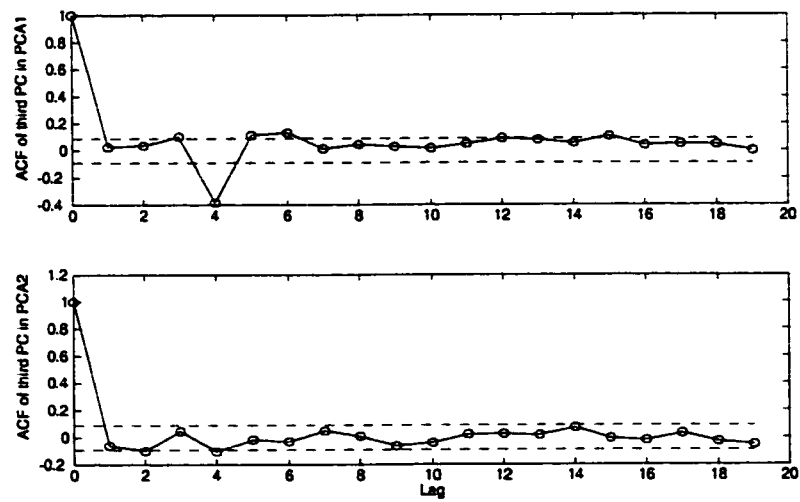


Figure 3.22: Thrid PC in *PCA2* is truly in residual space, while third PC in *PCA1* still have some autocorrelation.

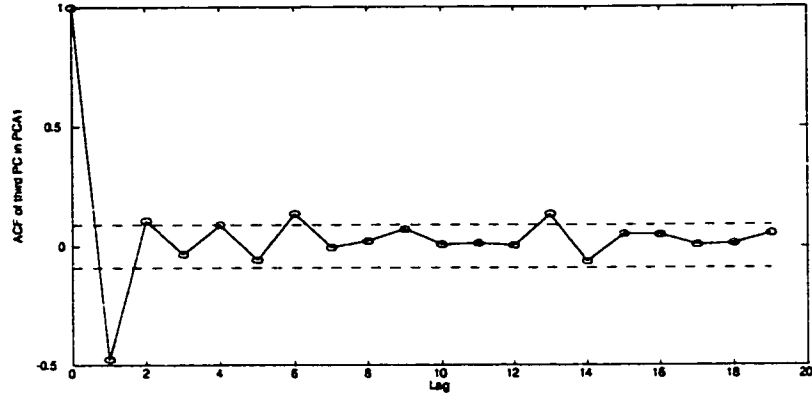


Figure 3.23: The significant value in ACF of the third PC matches with the time delay value.

nents actually describe the correlations between variables. To illustrate the point, let us consider the following example:

$$X = [F_1T_1 \quad F_2T_2 \quad F_3T_3] \quad (3.9)$$

A PCA model is built based on the measurement matrix X , which include three energy terms. Three principal components are plotted in Figure 3.24. First two PCs explain most of variation in the data. The last PC is the residual term for energy balance constraint between variables.

The last loading vector (Figure 3.25) provides the correlation between the energy terms, that is:

$$\begin{aligned} \text{Residual} &= 0.579F_1T_1 + 0.576F_2T_2 - 0.576F_3T_3 \approx 0 \\ \Rightarrow F_1T_1 + F_2T_2 &\approx F_3T_3 \end{aligned} \quad (3.10)$$

Equation 3.10 is the energy balance equation (a constraint that relates three energy terms). The last loading vector or the minor component reveals the relationship between them.

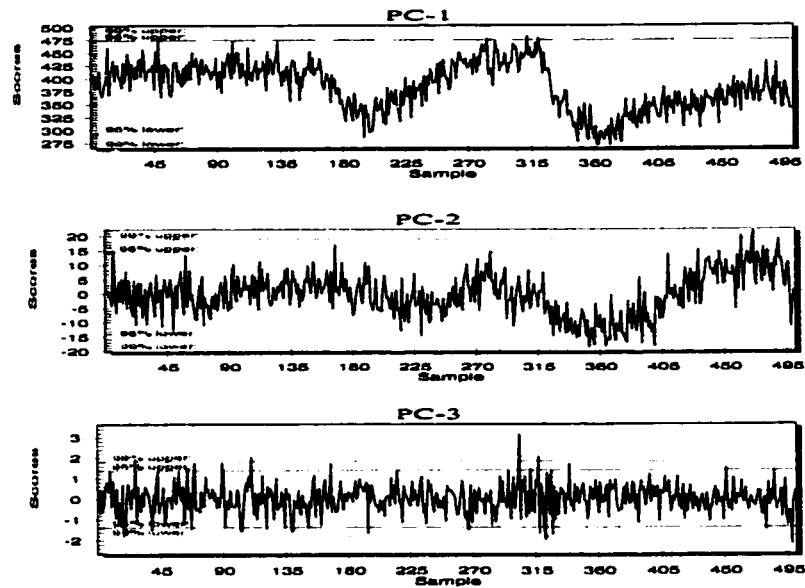


Figure 3.24: First two PCs explain the variation of the data. The last PC contains the residuals of energy balance equation.

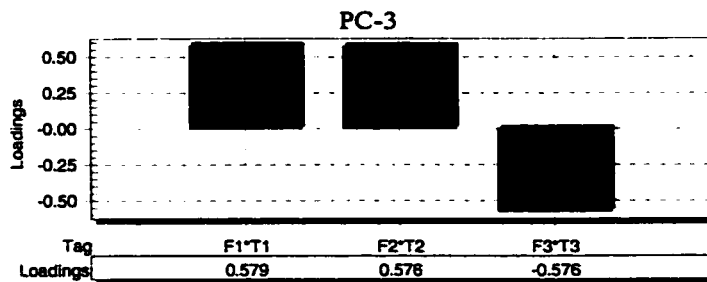


Figure 3.25: Third loading gives the correlation between variables

3.7 Conclusion

The basic methods of using PCA for SPM have been demonstrated via applying PCA to monitor a simple flow rate system. How to detect various sensor faults and fault isolation have been discussed in detail.

The limitation of using PCA model and possible ways to overcome these shortcomings are also illustrated. One should be cautious in assembling the measurement matrix to ensure a successful application of PCA, especially when dealing with the time delay. Finally, the minor components and associated loading vectors, which are important in revealing the relationships between variables are emphasized.

Chapter 4

Recursive PLS and PCA

4.1 Introduction

Multivariate statistical methods are rapidly being introduced in a variety of manufacturing processes, chemical process industry, steel rolling, pulp and paper, pharmaceuticals, food, beverage, and cosmetics (Wold and Sjostrom, 1998). Conventional PCA and PLS are most suitable for steady state processes where measurements are stationary signals and are only weakly autocorrelated. However, a common scenario is that chemical processes are usually dynamic, time-varying and nonstationary. Direct application of PLS and PCA to these chemical processes for modeling and monitoring usually results in large model-plant mismatch and many false alarms. Adaptive algorithms for these statistical tools are thus essential for monitoring such time-varying processes. In this chapter, recursive PLS and PCA algorithms are briefly described and an application to a pilot scale Continuous Stirred Tank Heater (CSTH) system is used to illustrate the necessity and effectiveness of these recursive algorithms.

4.2 Recursive PLS

Partial Least Square regression is effectively used in process modeling and monitoring to deal with a large number of variables with collinearity. In most of applications, the PLS regression is a batch-wise modeling approach, which means that the data are collected and stored in computer, and PLS regression is carried out on the historical batch of data. Although it overcomes the collinearity, the batch PLS has a limitation in that it is difficult to incorporate the newly available data in the model. Although

one can simply merge new data with the old one and rebuild the model, it is computationally inefficient because the old data are modeled repeatedly. Furthermore, the algorithm will ultimately run out of the computer memory since the data length keeps on increasing.

For adaptive process modeling and monitoring, it is essential to find an efficient approach to update PLS model recursively and efficiently. The basic idea of recursive PLS (RPLS), which was first proposed by Helland *et al.* (1991), was to use loading matrices to express old data. Although the algorithm is computationally efficient, it may suffer from losing information in the original data matrix and result in numerical errors. In addition, the algorithm is not an exponentially updated one. Qin (1998) modified this method and further extended it to block-wise RPLS with a moving window and a forgetting or discounting factor for steady and dynamic process modeling.

The key idea of RPLS is to update the covariance matrix efficiently. Consider a PLS regression coefficient matrix which can be expressed in a general framework:

$$C_{PLS} = (X^T X)^\dagger X^T Y \quad (4.1)$$

where $(\cdot)^\dagger$ denotes the generalized inverse defined by the PLS algorithm and X, Y are the input and output matrices. An explicit expression for the PLS coefficient matrix can be found in (Hoskuldson, 1988). Using singular value decomposition (SVD), X and Y can be decomposed as

$$\begin{aligned} X &= TP^T + E_A \\ Y &= TBQ^T + F_A \end{aligned} \quad (4.2)$$

where E_A and F_A are residual matrices for X and Y (refer to Chapter 2).

When a new pair of data matrix $(X_1 \ Y_1)$ is available, the augmented data matrices are

$$X_{new} = \begin{bmatrix} X \\ X_1 \end{bmatrix} \text{ and } Y_{new} = \begin{bmatrix} Y \\ Y_1 \end{bmatrix} \quad (4.3)$$

The new PLS model can be calculated as

$$C_{PLS}^{new} = \left(\begin{bmatrix} X \\ X_1 \end{bmatrix}^T \begin{bmatrix} X \\ X_1 \end{bmatrix} \right)^\dagger \begin{bmatrix} X \\ X_1 \end{bmatrix}^T \begin{bmatrix} Y \\ Y_1 \end{bmatrix} \quad (4.4)$$

In Equation 4.4, matrices $X^T X$ and $X^T Y$ can be denoted as

$$\begin{aligned} X^T X &\simeq PT^T TP^T = PP^T \\ X^T Y &= PT^T TBQ^T + PT^T F_A = PBQ^T \end{aligned} \quad (4.5)$$

where T is an orthonormal matrix and it is orthogonal to F_A (Qin, 1998). Therefore, the new PLS regression coefficients become

$$C_{PLS}^{new} = \left(\begin{bmatrix} P^T \\ X_1 \end{bmatrix}^T \begin{bmatrix} P^T \\ X_1 \end{bmatrix} \right)^\dagger \begin{bmatrix} P^T \\ X_1 \end{bmatrix}^T \begin{bmatrix} BQ^T \\ Y_1 \end{bmatrix} \quad (4.6)$$

Comparing Equation 4.4 with 4.6, we can see that performing PLS regression on the data pair

$$\begin{bmatrix} P^T \\ X_1 \end{bmatrix}, \begin{bmatrix} BQ^T \\ Y_1 \end{bmatrix}$$

results in the same regression model as performing PLS on data pair

$$\begin{bmatrix} X \\ X_1 \end{bmatrix}, \begin{bmatrix} Y \\ Y_1 \end{bmatrix}$$

In other words, instead of using the augmented input and output matrices, we can use the old model and the new data to obtain a new PLS model. To extend this result, we can further build a PLS model based on newly available data pair (X_1, Y_1) . Denote the matrices associated with (X_1, Y_1) through PLS regression by $(T_1, W_1, P_1, B_1, Q_1)$. It can be easily proved that performing PLS on data pair:

$$\begin{bmatrix} P^T \\ P_1 \end{bmatrix}, \begin{bmatrix} BQ^T \\ B_1Q_1^T \end{bmatrix} \quad (4.7)$$

results in the same regression model as performing PLS on data pair:

$$\begin{bmatrix} X \\ X_1 \end{bmatrix}, \begin{bmatrix} Y \\ Y_1 \end{bmatrix} \quad (4.8)$$

Similarly, it can be concluded that applying PLS regression on multi-block data pair $(X_1, Y_1), (X_2, Y_2) \cdots (X_s, Y_s)$ is equivalent to applying PLS regression on

$$\begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_s \end{bmatrix}, \begin{bmatrix} B_1Q_1^T \\ B_2Q_2^T \\ \vdots \\ B_sQ_s^T \end{bmatrix} \quad (4.9)$$

Since the number of rows in P (number of variables) is usually much less than that in X (number of samples), updating a PLS model based on sub-models other than the whole data set requires less computation. Note that Equation 4.5 holds under assumption that E_A is near zero. In practice it is essential to check whether $\|E_A\| \leq \epsilon$. Otherwise, Equation 4.5 does not hold and the RPLS algorithm will result in numerical errors. Qin (1998) pointed out the essentiality of this step in recursive PLS.

The cross validation method is widely used to determine the number of latent variables in PLS regression (Wold, 1978). Usually the subspace dimension reported by cross validation is less than 'A' which is required to satisfy recursive PLS. This means that when the model is used for prediction, we use fewer latent variables. However, more latent variables need to be retained to facilitate the model updates in the future.

Newly available data usually contain more important information about current process than the old ones. To adapt to the process change, it is desirable to give more weight to the new data while forgetting old data. This can be done by either using a moving window or a forgetting factor. For a moving window approach, the model based on old data block is omitted from Equation 4.9 when the model based on new data block is available. Then PLS regression can be performed on this updated data matrices. The forgetting factor approach uses a forgetting factor λ ($0 < \lambda < 1$) to discount the old data exponentially.

Suppose that at step s , we have a PLS model with associated matrices P_s , B_s , Q_s . When the next data block is available, the new model with a forgetting factor can be obtained by performing PLS using

$$\begin{bmatrix} P_{s+1}^T \\ \lambda P_s^T \end{bmatrix}$$

as the input matrix and

$$\begin{bmatrix} B_{s+1}Q_{s+1}^T \\ \lambda B_sQ_s^T \end{bmatrix}$$

as the output matrix. Where P_{s+1} , B_{s+1} , Q_{s+1} are the matrices associated with the new data block. In other words, the new PLS model can be obtained by deriving a sub-model based on the new data block and then combining it with the old model with a forgetting factor.

Dayal and MacGregor (1997a) addressed RPLS by combining the kernel PLS algorithm (Dayal and MacGregor, 1997a) with the recursive updating of the covariance matrix as follows:

$$\begin{aligned}(X^T X)_t &= \lambda_t (X^T X)_{t-1} + x_t^T x_t \\ (X^T Y)_t &= \lambda_t (X^T Y)_{t-1} + x_t^T y_t\end{aligned}\tag{4.10}$$

where, x_t and y_t are the new input and output row vector. $(X^T X)_t$ and $(X^T Y)_t$ are the updated covariance matrices at time t . When calculating the current covariance matrix, the previous data are discounted exponentially with a forgetting factor λ ($0 < \lambda < 1$). This strategy of using the kernel algorithm to calculate the updated covariance matrices in Equation 4.10 has been shown to be much faster than the conventional PLS (Dayal and MacGregor, 1997b).

In ordinary PLS, the data are usually scaled to have zero mean and unit variance before performing PLS regression. In a time-varying process, the mean and standard deviation of variables may change over time. Using the old mean and standard deviation to scale the newly available data may result in a bias term. In this case, one can use the updated mean and standard deviation to scale the data to avoid the bias. This step will be discussed in the next chapter. Another approach to handle the mean-centering problem is to introduce an intercept term d , as in the general linear regression model. The input and coefficient matrices are simply modified to include this term as shown below:

$$Y = XC + I \cdot d = [X \ I] \begin{bmatrix} C \\ d \end{bmatrix}\tag{4.11}$$

where I is a unity column vector with all its elements equal to 1 and d is a constant scalar. Such a bias term can help reduce the model prediction error.

4.3 Recursive PCA

Traditional PCA model works well with a stationary process. However, chemical processes usually demonstrate time-varying behaviour, such as catalyst deactivation, heat exchanger scaling, equipment aging, sensor and process drifting, maintenance and cleaning. A lot of false alarms are usually resulted if a time-invariant PCA model

is used to monitor such aforementioned processes. To improve the robustness of process monitoring, an adaptive process monitoring strategy is essential. Gallagher *et al.* (1997) pointed out the need to consider the drifting mean and variance in a non-stationary system when he used the PCA model to monitor a semiconductor etching process. Wold (1994) extended the standard multivariate models with exponentially weighted moving average (EWMA) models based on multivariate scores from PCA and PLS. A complete recursive PCA scheme, which was recently proposed by Qin *et al.* (1999), considered the following issues:

1. Recursive update of the correlation matrix when building the PCA model, including the update of the mean and variance.
2. Efficient algorithm for the computation of PCA.
3. Recursive determination of the number of principal components.
4. Recursive determination of the confidence limits for SPE and T^2 to facilitate adaptive monitoring.

Herein, we give a brief review of the algorithms for updating the correlation matrix and we use the same notation as in Qin *et al.* (1999). $X^0 \in \mathbb{R}^{n \times m}$ denotes the raw data matrix with n samples and m variables. The raw matrix is so scaled that data in each column have zero mean and unit variance.

The conventional PCA involves the singular value decomposition of the correlation matrix or covariance matrix. The correlation matrix is estimated using the scaled data:

$$R \approx \frac{1}{n-1} X^T X \quad (4.12)$$

Let $X_1^0 \in \mathbb{R}^{n_1 \times m}$ be the raw data block used to build initial PCA model. The mean of the variable can be expressed as the vector:

$$b_1 = \frac{1}{n_1} (X_1^0)^T I_{n_1 \times 1} \quad (4.13)$$

where $I_{n_1 \times 1} = [1, 1, \dots, 1]^T \in \mathbb{R}^{n_1}$. The scaled data vector is then given by:

$$X_1 = (X_1^0 - I_{n_1 \times 1} b_1^T) \Sigma_1^{-1} \quad (4.14)$$

where $\Sigma_1 = \text{diag}(\sigma_{1,1}, \dots, \sigma_{1,m})$ and $\sigma_{1,i}$ is the standard deviation of the i^{th} variable ($i = 1, \dots, m$).

The correlation matrix based on X_1 is

$$R_1 = \frac{1}{n_1 - 1} X_1^T X_1 \quad (4.15)$$

Assume that b_k , X_k and R_k have been calculated based on k^{th} block of data and we want to calculate b_{k+1} , X_{k+1} and R_{k+1} when the next block of data $X_{n_{k+1}}^0 \in \mathbb{R}^{n_{k+1} \times m}$ is available. Denoting

$$X_{k+1}^0 = \begin{bmatrix} X_k^0 \\ X_{n_{k+1}}^0 \end{bmatrix} \quad (4.16)$$

for all the $k + 1$ blocks of data. The mean b_{k+1} is related to b_k in the following way:

$$\left(\sum_{i=1}^{k+1} n_i \right) b_{k+1} = \left(\sum_{i=1}^k n_i \right) b_k + \left(X_{n_{k+1}}^0 \right)^T 1_{n_{k+1}} \quad (4.17)$$

Denoting $N_k = \sum_{i=1}^k n_i$, equation 4.17 can be rewritten as:

$$b_{k+1} = \frac{N_k}{N_{k+1}} b_k + \frac{1}{N_{k+1}} \left(X_{n_{k+1}}^0 \right)^T 1_{n_{k+1}} \quad (4.18)$$

The recursive calculation of X_{k+1} is given by

$$\begin{aligned} X_{k+1} &= [X_{k+1}^0 - 1_{k+1} b_{k+1}^T] \Sigma_{k+1}^{-1} \\ &= \left[\begin{bmatrix} X_k^0 \\ X_{n_{k+1}}^0 \end{bmatrix} - 1_{k+1} b_{k+1}^T \right] \Sigma_{k+1}^{-1} = \begin{bmatrix} X_k^0 - 1_k \Delta b_{k+1}^T - 1_k \Delta b_k^T \\ X_{n_{k+1}}^0 - 1_{n_{k+1}} b_{k+1}^T \end{bmatrix} \Sigma_{k+1}^{-1} \\ &= \begin{bmatrix} X_k \Sigma_k \Sigma_{k+1}^{-1} - 1_k \Delta b_{k+1}^T \Sigma_{k+1}^{-1} \\ X_{n_{k+1}} \end{bmatrix} \end{aligned} \quad (4.19)$$

where

$$\begin{aligned} X_k &= (X_k^0 - 1_k b_k^T) \Sigma_k^{-1} \\ X_{n_{k+1}} &= (X_{n_{k+1}}^0 - 1_{n_{k+1}} b_{k+1}^T) \Sigma_{k+1}^{-1} \\ \Sigma_j &= \text{diag}(\sigma_{j,1}, \dots, \sigma_{j,m}), \quad j = k, k+1 \\ \Delta b_{k+1} &= b_{k+1} - b_k \end{aligned} \quad (4.20)$$

Since all the $k + 1$ blocks of mean-centered data are:

$$X_{k+1}^0 - 1_{k+1} b_{k+1}^T = \begin{bmatrix} X_k^0 - 1_k \Delta b_{k+1}^T - 1_k \Delta b_k^T \\ X_{n_{k+1}}^0 - 1_{n_{k+1}} b_{k+1}^T \end{bmatrix} \quad (4.21)$$

the standard deviation for the i^{th} variable is

$$\sigma_{k+1,i}^2 = \frac{\left\| \begin{array}{c} X_k^0(:,i) - 1_k \Delta b_{k+1}(i) - 1_k \Delta b_k(i) \\ X_{n_{k+1}}^0(:,i) - 1_{n_{k+1}} b_{k+1}(i) \end{array} \right\|^2}{N_{k+1} - 1} \quad (4.22)$$

where $X_k^0(:,i)$ is the i^{th} column of the associated matrix; $b_{k+1}(i)$ and $\Delta b_{k+1}(i)$ are the i^{th} elements of the associated vector. Equation 4.22 can be rewritten as

$$(N_{k+1} - 1) \sigma_{k+1,i}^2 = \|X_k^0(:,i) - 1_k b_k(i)\|^2 + N_k \Delta b_{k+1}^2(i) + \|X_{n_{k+1}}^0(:,i) - 1_{n_{k+1}} b_{k+1}(i)\|^2 \quad (4.23)$$

Introducing a new notation,

$$\sigma_{k,i}^2 (N_k - 1) = \|X_k^0(:,i) - 1_k b_k(i)\|^2 \quad (4.24)$$

then Equation 4.23 can be rewritten as:

$$(N_{k+1} - 1) \sigma_{k+1,i}^2 = (N_k - 1) \sigma_{k,i}^2 + N_k \Delta b_{k+1}^2(i) + \|X_{n_{k+1}}^0(:,i) - 1_{n_{k+1}} b_{k+1}(i)\|^2 \quad (4.25)$$

The correlation matrix can be similarly derived as:

$$\begin{aligned} R_{k+1} &= \frac{1}{N_{k+1} - 1} X_{k+1}^T X_{k+1} \\ &= \frac{N_k - 1}{N_{k+1} - 1} \Sigma_{k+1}^{-1} \Sigma_k R_k \Sigma_k \Sigma_{k+1}^{-1} + \frac{N_k}{N_{k+1} - 1} \Sigma_{k+1}^{-1} \Delta b_{k+1} \Delta b_{k+1}^T \Sigma_{k+1}^{-1} \\ &\quad + \frac{1}{N_{k+1} - 1} X_{n_{k+1}}^T X_{n_{k+1}} \end{aligned} \quad (4.26)$$

Since the old data do not represent the current process, it is usually desirable to ignore the old data exponentially. The update of mean, variance and covariance with a forgetting factor are thus computed as follows:

$$b_{k+1} = \mu b_k + (1 - \mu) \frac{1}{n_{k+1}} \left(X_{n_{k+1}}^0 \right)^T 1_{n_{k+1}} \quad (4.27)$$

$$\sigma_{k+1,i}^2 = \mu (\sigma_{k,i}^2 + \Delta b_{k+1}^2(i)) + (1 - \mu) \frac{1}{n_{k+1}} \|X_{n_{k+1}}^0(:,i) - 1_{n_{k+1}} b_{k+1}(i)\|^2 \quad (4.28)$$

$$R_{k+1} = \mu \Sigma_{k+1}^{-1} (\Sigma_k R_k \Sigma_k + \Delta b_{k+1} \Delta b_{k+1}^T) \Sigma_{k+1}^{-1} + (1 - \mu) \frac{1}{n_{k+1}} X_{n_{k+1}}^T X_{n_{k+1}} \quad (4.29)$$

where $0 < \mu \leq \frac{N_k}{N_{k+1}}$ is a forgetting factor. A small value of the forgetting factor tends to discount the old data quickly. If $\mu = \frac{N_k}{N_{k+1}}$, it is equivalent to no forgetting.

The number of principal components may change over time. Thus it is necessary to determine the number of PCs recursively. There are many methods for determining the number of PCs, as discussed in Chapter 2. However, not all approaches are suitable for recursive PCA. For example, the cross-validation is not suitable because the old data may contain little information about the current process.

When a PCA model is used as a process monitoring tool, the Hotelling T^2 and SPE are the two main monitoring charts used. The confidence limits for these two statistics also change over time since the confidence limits for T^2 rely on the number of PC retained in the model, while the limits for SPE depend on the eigenvalues of the correlation matrix. Details about these issues can be found in Qin *et al.* (1999).

The procedure of using RPCA for process monitoring is summarized as follows:

1. Build a PCA model based on an initial data block ($k = 1$) collected from normal operating condition. Calculate the confidence limits for T^2 and SPE.
2. When a new sample is available, calculate T^2 and SPE based on the old model. If either of the two statistics exceeds the confidence limit obtained in the previous step, then an abnormal event is detected. At this point model updating is stopped. Otherwise, collect the data until a block size (n_{k+1}) is reached.
3. Update the covariance matrix using Equation 4.26 or 4.29. Rebuild the PCA model based on the updated covariance matrix. Calculate new confidence limits for T^2 and SPE. Set $k = k + 1$, and go back to step 2.

In practice, step 2 is important because the data that contains the sensor fault and other abnormal conditions should not be included in updating the model. However, sometimes it is not easy to discriminate the normal process drift from a small disturbance and sensor fault. This point will be further addressed in the next section.

4.4 Applying RPCA to a CSTH System

In this section, a continuous stirred tank heater (CSTH) system is used to illustrate adaptive process monitoring through RPCA. This system is available in the Computer

LV#	Eigenvalues	% variance captured this PC	% total variance captured
1	4.21	52.58	52.58
2	1.16	20.12	72.70
3	0.85	10.73	83.43
4	0.70	8.83	92.26
5	0.42	5.25	97.51
6	0.103	1.28	98.80
7	0.0082	1.02	99.82
8	0.0014	0.18	100

Table 4.1: Variance captured by latent variables in dynamic PCA model

Process Control laboratory at the University of Alberta. A schematic diagram of the CSTD is shown in Figure 4.1. High temperature steam passes through a heater coil. The cold water is heated by the steam coil and exits through a long copper tube. Four thermocouples located at different sections of the long exit pipe provide temperature signals at different locations. The flow rates of steam and cold water are available as manipulative variables to control the temperature and level of the water in the tank.

Flow rates of steam and cold water, temperature and level are the four monitored measurements. Taking the dynamics of the process into account, the lag one measurements of four variables are also included in the data matrix:

$$X = [F_{steam}(t) \ F_{cold}(t) \ T(t) \ L(t) \ F_{steam}(t-1) \ F_{cold}(t-1) \ T(t-1) \ L(t-1)] \quad (4.30)$$

Performing principal component analysis to the data matrix in Equation 4.30 is also known as the dynamic PCA, which has been suggested by Ku *et al.* (1995). The strategy of including lagged measurements in X allows one to capture the dynamic relationships between variables in this system. Several inter-relationships exist between these measurements, for example, the increase of F_{cold} not only raises the level, but also acts as a disturbance to the temperature and effects the steam flow rates. Cross validation suggests that four principal components, which can capture 92% variation of measurements, should be retained in the model (Table 4.1).

Figure 4.2 shows the Hotelling T^2 and SPE charts for a fixed PCA model when a temperature disturbance (between samples 50-150) and a cold water flow rate disturbance (between samples 450-470) are introduced. Both statistics clearly show up and exceed the 99% confidence limits indicating the occurrence of some abnormal condition.

In Figure 4.3, a small variation of the level setpoint is introduced. To follow the change of the setpoint, the process moves to another state which is a normal operation in the system. However, both Hotelling T^2 and SPE for a fixed PCA model exceed their confidence limits, resulting in a false alarm. Therefore, it is inferred that a time-invariant PCA cannot be used to monitor a time-varying system effectively in practice.

Figure 4.4 and 4.5 show the T^2 and SPE calculated from a recursive PCA model. In this experiment, the set point change begins at sample instant 500 as we update the PCA model recursively. Figure 4.4 shows the results when the model is updated every one sample with forgetting factor $\mu = 0.995$. Figure 4.5 is the case where the model is updated after every 10 samples with a forgetting factor $\mu = 0.985$. Both statistics are well within the confidence limits and significantly reducing the probability of false alarms. Note that the confidence limits for SPE vary with time after sample 500 because of the variation of eigenvalues of the covariance matrix. However, in this experiment, the number of principal components are fixed at 4, resulting in a constant confidence limit for T^2 as shown in Figure 4.4 and 4.5.

When applying RPCA for adaptive process monitoring, the value of the forgetting factor μ determines how efficiently the PCA model tracks the process drift. A small forgetting factor discounts the old data heavily and allows the model to follow the process change quickly, but may sacrifice the sensitivity of the model to detect small disturbances. Figure 4.6 shows the case where a recursive PCA can detect both a small temperature disturbance (between 150-250) and a rather large cold water disturbance (between 550-570) when $\mu = 0.998$ is used. However, when a smaller value of forgetting factor $\mu = 0.99$ is used, both T^2 and SPE fail to detect the small temperature disturbance (Figure 4.7). This means that RPCA model by itself cannot discriminate the normal process drift from a slow drift caused by a small disturbance. On the other hand, an abrupt and a large cold water disturbance usually causes the Hotelling T^2 or SPE to violate the confidence limits significantly. RPCA can quickly detect such a violation and will stop updating the model until the process returns to its normal state.

4.5 Analysis and Conclusions

In this chapter, recursive PLS and PCA algorithms have been described and an application example is used to illustrate the effectiveness of the algorithm. To make those statistical tools more robust and successful in practice, it is essential to develop the recursive algorithms for PCA and PLS to adapt to the process drift. Although progress has been made in developing and implementing these recursive algorithms, there are still a lot of practical issues that need to be dealt with. Here, we point out some directions for future work. First, when the recursive PCA is applied on-line, the measurements are usually very noisy. If the data is directly included in the recursive PCA model, the true correlation between variables may suffer from such noisy signals. Thus an effective on-line filter needs to be found and used together with recursive PCA. Second, the measurements may include some sensor faults. If so, these measurements should not be used to update the PCA model since it will destroy the correlation between variables. A mechanism is thus necessary to separate the faulty sensor signal from the normal signal before updating the model (Qin and Li, 1999). In addition the determination of an optimal forgetting factor to trade off the sensitivity and robustness of recursive PCA for a given process needs further investigation. Finally, recursive PCA shown herein is a static model updating technique. It may not work for fully dynamic processes. Thus dynamic RPCA which is suitable in monitoring a fully dynamic process should be investigated.

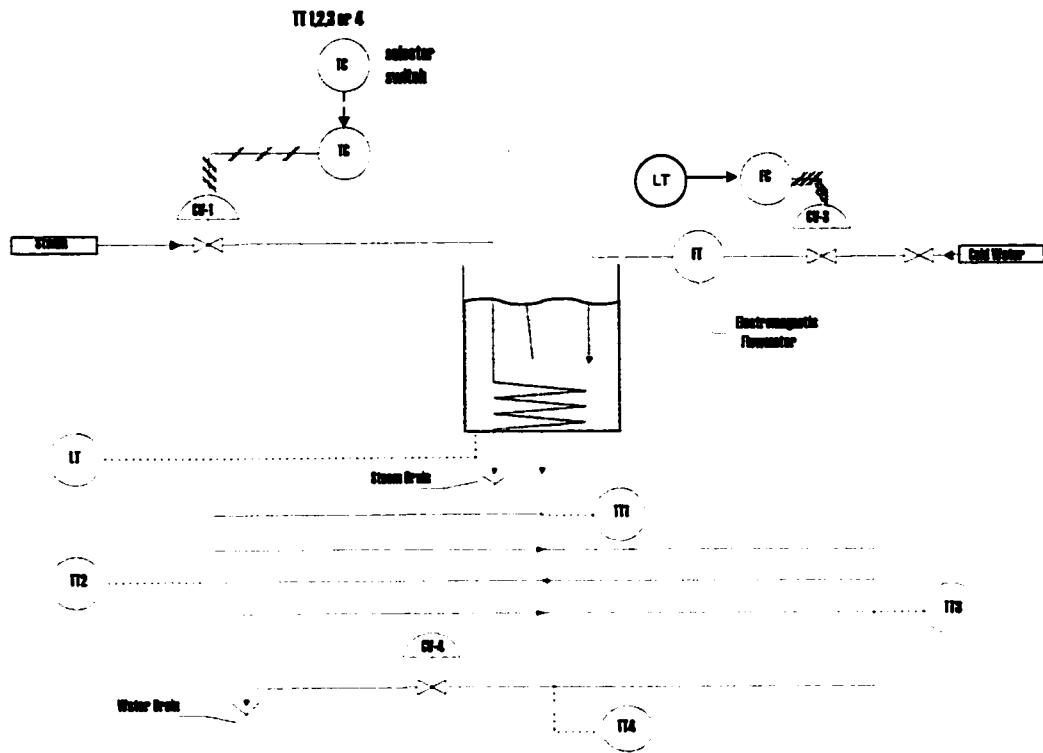


Figure 4.1: The schematic diagram of the CSTD system

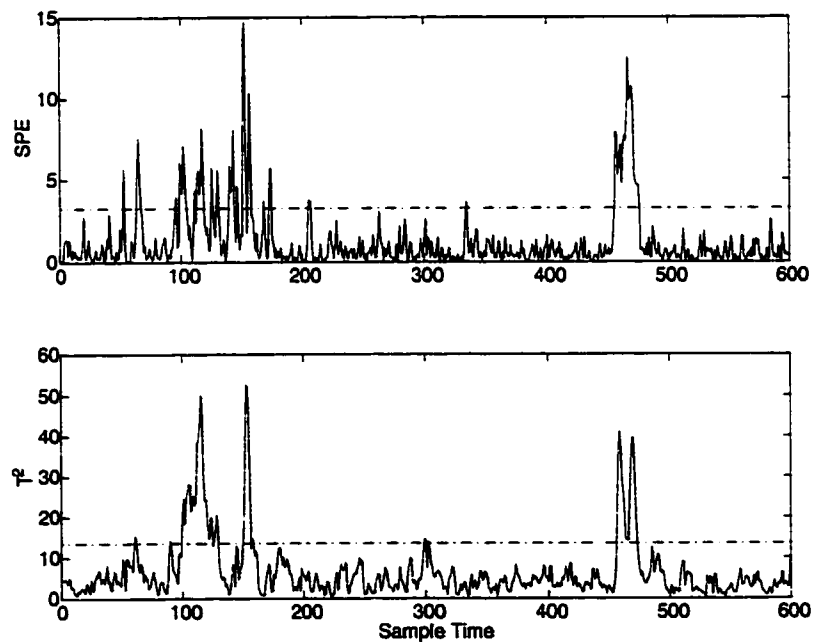


Figure 4.2: Hotelling T^2 and SPE both detect the temperature and level disturbance.

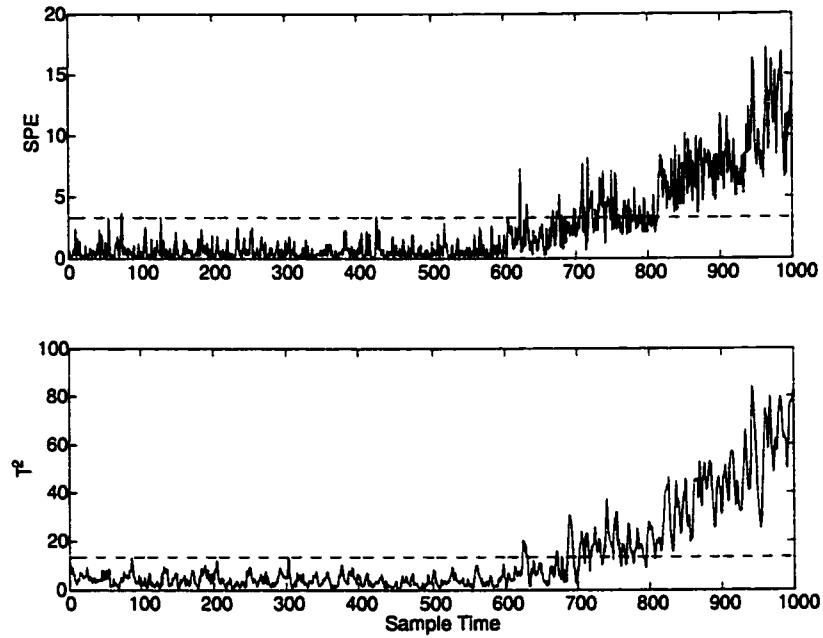


Figure 4.3: The Hotelling T^2 and SPE without recursively updating of the model.

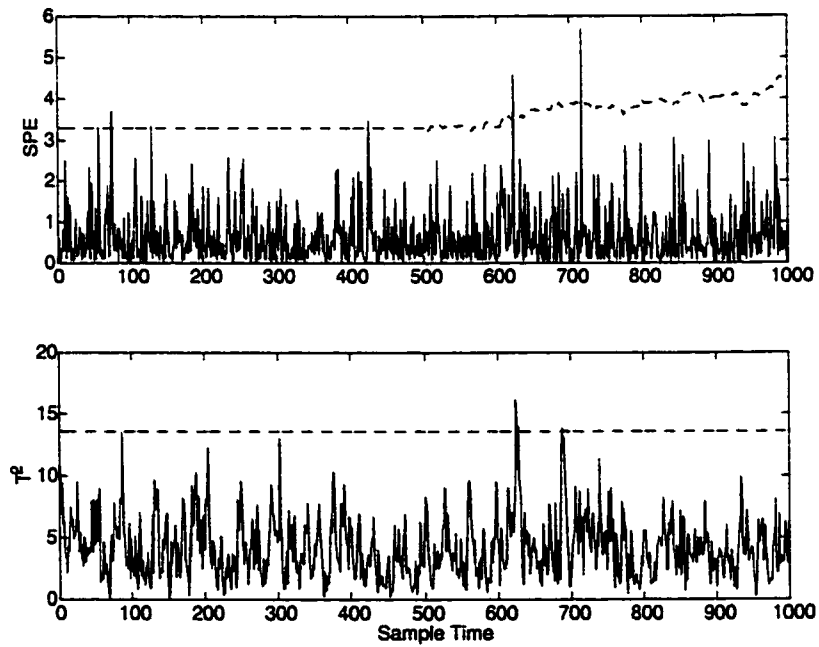


Figure 4.4: Hotelling T^2 and SPE of recursive PCA model, forgetting factor $\mu = 0.995$, model updated every one sample.

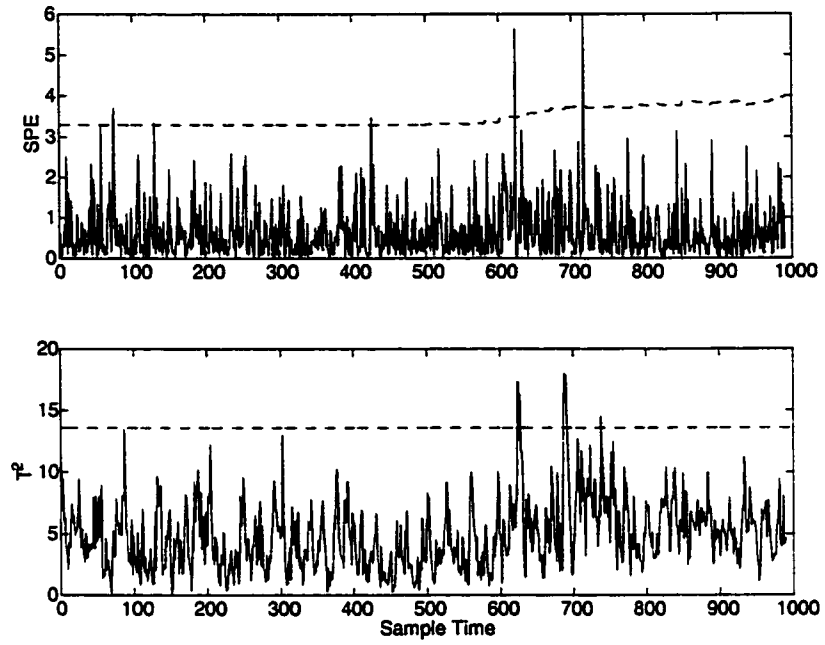


Figure 4.5: Hotelling T^2 and SPE of recursive PCA model, forgetting factor $\mu = 0.985$, model updated after every ten samples.

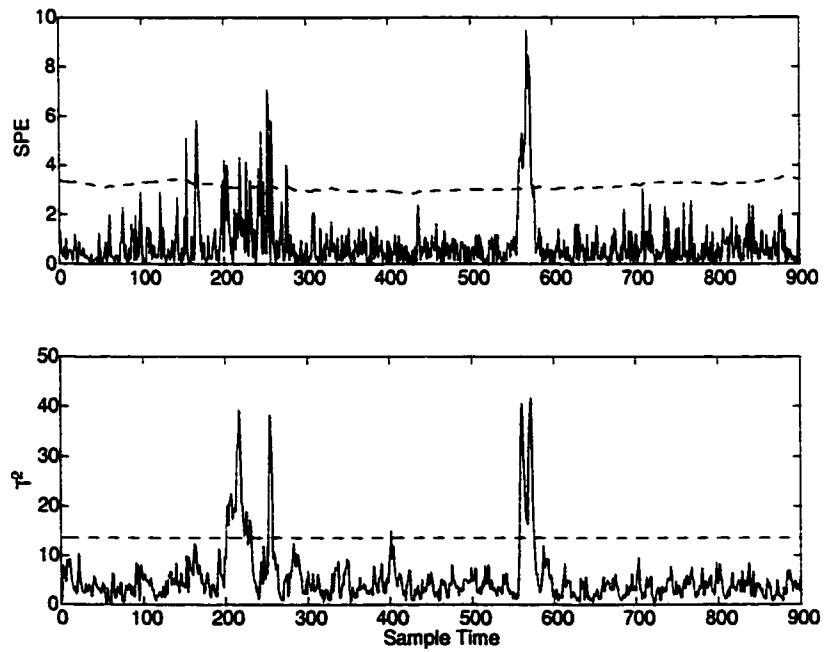


Figure 4.6: Recursive PCA can detect both flow rate and temperature disturbance when $\mu = 0.998$.

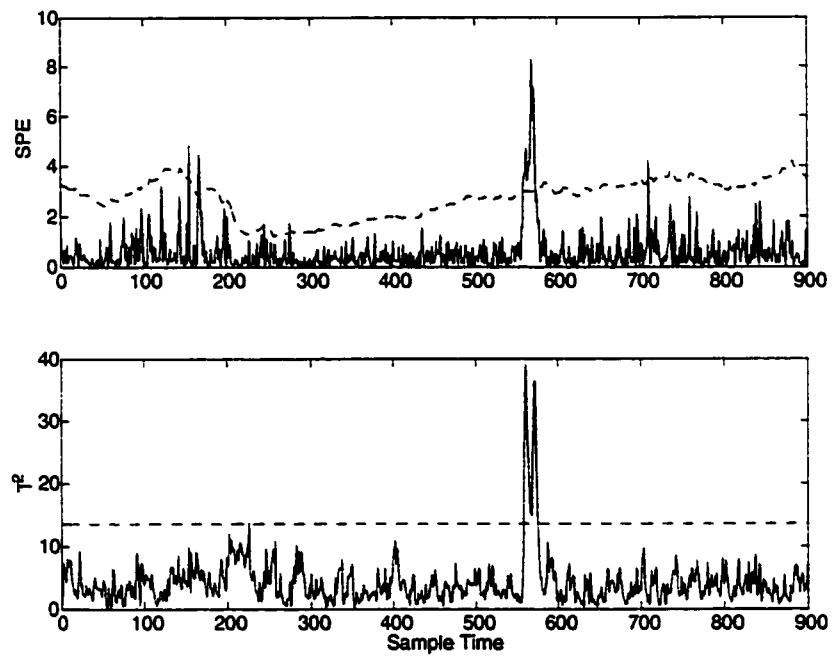


Figure 4.7: Recursive PCA can detect large flow rate disturbance, but fail to detect small temperature disturbance when $\mu = 0.99$.

Chapter 5

Industrial Data Analysis

5.1 Introduction

PLS has been proved as a good alternative to the classical multiple linear regression and principal component regression methods because of its robustness in dealing with colinear measurements. Many successful industrial applications of PLS have been reported recently (Wise, 1991; Morud, 1996; Blom, 1996; Mujunen *et al.*, 1996; Lakshminarayanan, 1997).

With the present-day's focus on environmental monitoring, the ability to predict and report certain emissions, such as CO , SO_2 , NO_x , etc. is becoming an important part of daily routine work in chemical plants. In this chapter, we show that a PLS based empirical model can be used to predict emission levels in the situation when strong colinearity exists among variables. Some issues in PLS modeling, such as data pretreatment, variable selection, dynamic PLS modeling, are also discussed in detail. Finally we show how the PLS weighting vector can help in diagnosing the root causes that influence the emission.

5.2 Problem Description

In the Syncrude industrial example that is considered here, there are 4 units and a total of 74 measured variables. (In respect of proprietary information, a detailed description of the process is omitted. Variable numbers, instead of the true tag names, are used to represent the measurements). Two variables are related to the emission of SO_2 : the total emission generated from four units and the net emission

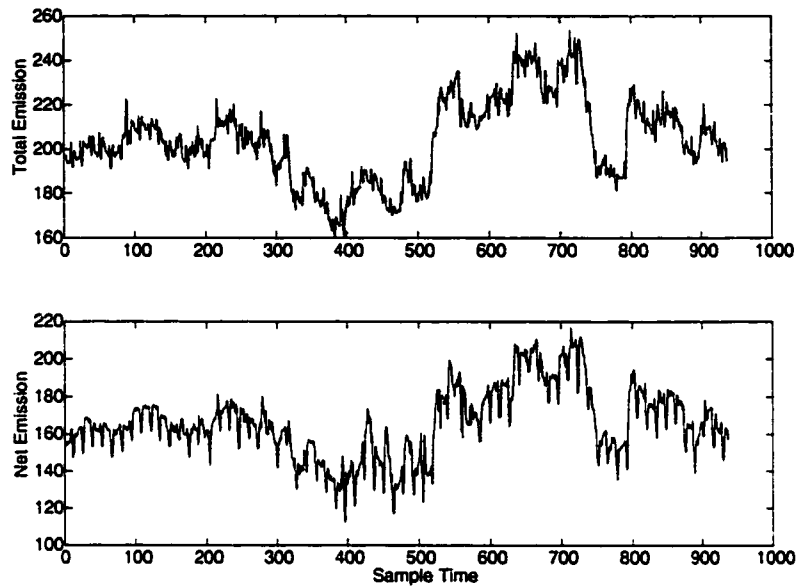


Figure 5.1: Trend plots of the total emission and net emission.

(a calculated value eliminating the emission from one unit) are two variables that need to be monitored. Figure 5.1 shows the time-series of total emission and net emission. There is clear correlation between the trend plots of the two emission variables. Although the values between sample instants 0-300 is higher than that between sample instants 300-500, they are both considered to be at normal level. However, for some unknown reasons, the emission went up after sample instant 500 and lasted for a period of time and then went back to the normal level around sample instant 750.

The objective of this data analysis was twofold:

- First, to investigate what are the possible reasons that can explain the high emission level whenever it occurs? In other words, which variables are the likely candidates that effect the emission?
- Second, is it possible to build an inferential model to predict the emission level? If so, a good prediction model can help in taking preventative action to prevent high emission levels.

5.3 Data Pre-processing

Process data is usually contaminated by the random and gross errors due to measurement noise, sensor failure, and human error. The statistical analysis tools (PCA, PLS, etc.) are data driven methods, that rely on good quality data. Since bad data can severely effect the analysis results, data pre-processing is essential before doing further analysis.

Data pre-processing usually refers to transforming data (taking logarithm, difference, etc.), filtering, truncation, replacing or deleting bad and missing data. Proper choice of the data preprocessing methods depends on the nature of the data. Here, we describe some filtering algorithms that are commonly used in dealing with the chemical process data (Seborg *et al.*, 1989).

The samples of measured variables are denoted as $\cdots x_{t-1}, x_t, x_{t+1} \cdots$ and the corresponding filtered variables are denoted as $\cdots y_{t-1}, y_t, y_{t+1} \cdots$, where t refers to the current sampling instant.

- Moving average filter

A moving average filter averages a specified number of past data points, by giving equal weight to each data points. The algorithm is expressed as

$$y_t = \frac{1}{J} \sum_{i=t-J+1}^t x_i \quad (5.1)$$

J is the number of past data points that are being averaged.

- Exponential filter

The output of an exponential filter can be expressed as:

$$y_t = \alpha x_t + (1 - \alpha)y_{t-1} \quad (5.2)$$

Equation 5.2 indicates that the filtered measurement is a weighted sum of the current measurement x_t and the filtered value at the previous sampling instant y_{t-1} . Limiting cases for α are

$\alpha = 1$: No filtering (the filter output is the raw measurement x_t).

$\alpha \rightarrow 0$: The measurement is ignored.

- An exponential filter is more effective than the moving average filter, since it gives more weights to the most recent data. Both exponential and moving average filter are examples of low pass filters.
- Noise-Spike Filter

A noise spike is said to occur if a noise measurement changes suddenly by a large amount and then returns to the normal value at next sampling instant.

Noise-spike filters are used to set a maximum allowable change of current measurement from previous one. The noise-spike filter can be expressed as:

$$y_t = \begin{cases} x_t & \text{if } |x_t - y_{t-1}| \leq \Delta x \\ y_{t-1} - \Delta x & \text{if } y_{t-1} - x_t > \Delta x \\ y_{t-1} + \Delta x & \text{if } y_{t-1} - x_t < -\Delta x \end{cases} \quad (5.3)$$

If a large change in the measurement occurs, the filter replaces the measurement by the previous filter output plus (or minus) the maximum allowable change.

The trend plot of two process variables (v14 and v36) are displayed in Figure 5.2. Many spikes are observed in these trend plots. These spikes may result from failure of the sensors or due to some other reasons. If included in the data analysis, these spikes will certainly influence the variance/covariance information. In this case study, a median filter is used to filter out spikes.

If the window length of a median filter is $2M + 1$. The output of the median filter at time instant t is

$$X_t = \text{median}(X_{t-M}, \dots, X_{t-1}, X_t, X_{t+1}, \dots, X_{t+M}) \quad (5.4)$$

The median operator in Equation 5.4 reorders $X_{t-M}, \dots, X_{t-1}, X_t, X_{t+1}, \dots, X_{t+M}$ in a descending or ascending way, and then takes the middle (or the median) value of the reordered queue.

If a data segment consisting of a spike has n data points, a length M ($M > 2n$) median filter can eliminate such a spike. Thus, proper selection of M requires the knowledge of the maximum duration of gross errors. When such knowledge is available, the length of median filter can be determined so that a complete elimination of gross errors is achieved.

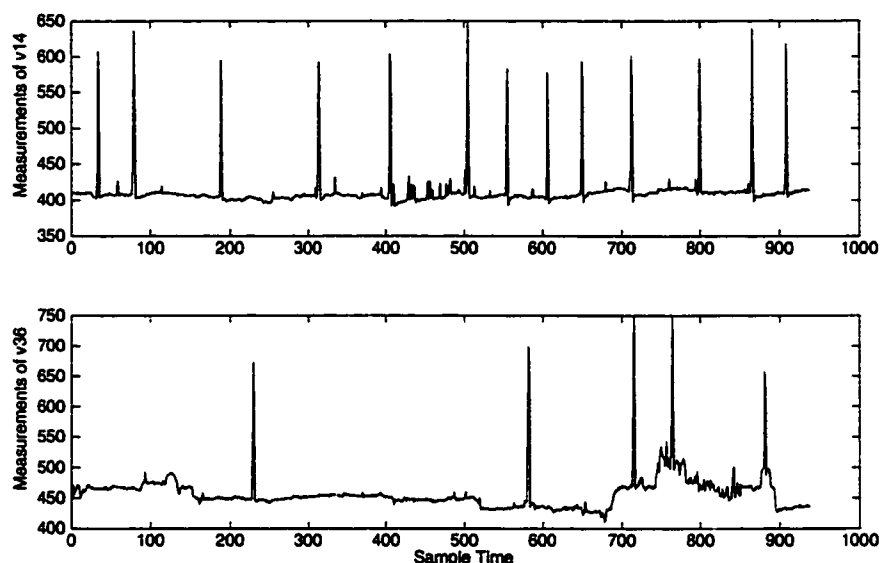


Figure 5.2: The raw measurements of v14 and v36.

A length 7 median filter is used to eliminate the spikes in all measurements, since maximum data window of big spikes in the data set are 7. The filtered v14 and v36 are shown in Figure 5.3. All the big spikes have been eliminated after filtering. Note that the median filter algorithm shown in Equation 5.4 is non-causal, because it requires the future data values $X_t, X_{t+1}, \dots, X_{t+M}$ as the input. Thus this filter is not suitable for on-line use.

A median filter is a low pass filter. With increase of the filter data window length, more information in high frequency part will be lost, but the low frequency information is well retained. This point is illustrated by the spectrum plots (Figure 5.4) of the original and filtered signal v14. Since the information in chemical process data is mainly localized in the low frequency part, and the noise is mainly localized in the high frequency part, the median filter is effective in getting rid of noise and gross errors in measurements.

5.4 Variable Selection

Variable selection is an important subject in statistical modeling method due to the fact that not all the recorded variables are relevant to the dependent variables (variables in Y block). In fact if some irrelevant variables are deleted from the predictor

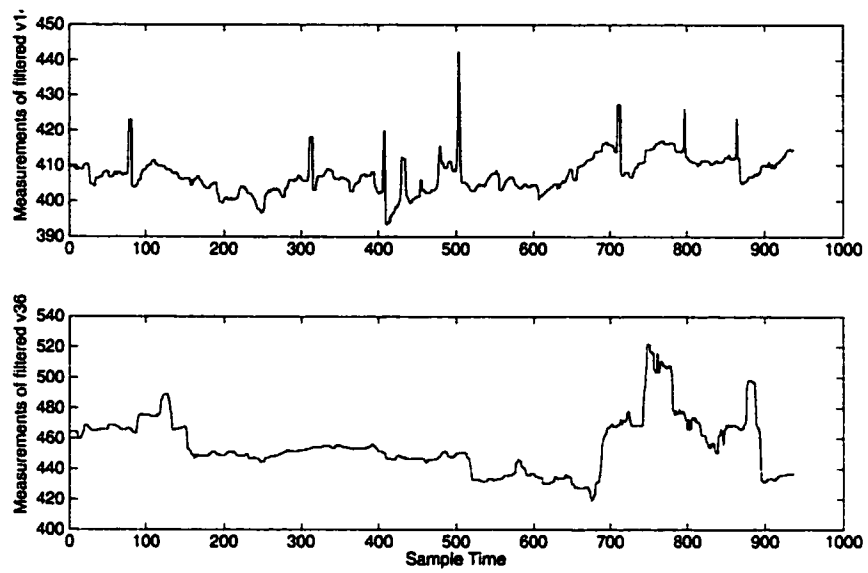


Figure 5.3: Filtered signals of v14 and v36

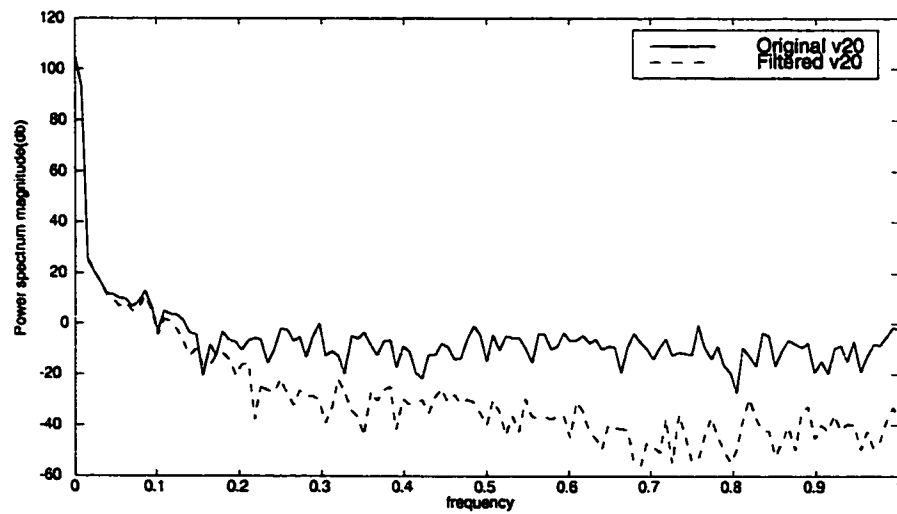


Figure 5.4: Comparison of spectrum of orginal and filtered signal (v20)

matrix (X block), then the model prediction capability can be improved. If irrelevant variables are retained in the model, these variables can potentially deteriorate the quality of the model. Therefore, it is necessary to eliminate totally irrelevant variables and only keep the relevant variables in the model. Generally, variable selection techniques can be classified into several categories (Forina *et al.*, 1999):

- Subset selection.

A number of regression models are built using different subsets of the predictors, the performance of each model is evaluated and the best one is selected based on some criteria, for example, minimum square prediction error. Genetic algorithms (GA) are well known to have the advantage of searching over all possible subsets but within a reasonable time. Moreover, GA offer a number of possible optimal or near-optimal subsets (Leardi *et al.*, 1992).

- Dimension-wise selection.

Martens and Naes (1989) suggest replacing the small PLS weights in each latent variable with zero, so that the corresponding predictors are cancelled from the latent variable. Other similar PLS weights related methods can be found in interactive variable selection (Lindgren *et al.*, 1994), automatic variable selection (Forina *et al.*, 1986), and the intermediate least square method (Frank, 1987).

- Model-wise elimination.

The model is developed with all the predictors:

$$y = b_1x_1 + b_2x_2 + \dots + b_nx_n \quad (5.5)$$

If the associated regression coefficients in the regression model is small. The corresponding predictor can be eliminated. Another criterion that can help decide the importance of the predictor is ISE (iterative stepwise elimination) (Boggia *et al.*, 1997), defined as

$$\gamma_i = \frac{|b_i|s_i}{\sum_{i=1}^n |b_i|s_i} \quad (5.6)$$

where s_i is the standard deviation of the predictor i . The predictor with the minimum importance is eliminated in each elimination cycle, and the model is

built again with the remaining predictors. The final model is the one that has the maximum predictive ability. For other model-wise elimination methods, readers can refer to uninformative variable elimination PLS (Centner *et al.*, 1996), and the iterative predictor weighting PLS method (Forina *et al.*, 1999).

Although there are 72 process variables measured in this process, more careful scrutiny of these variables indicates that some variables have little relationship with change in emission observed. One such variable (v39) is shown in Figure 5.5 as an example. A further correlation analysis tell us that the correlation coefficient between v39 and total emission is -0.08, which means that relationship between these two variables can be ignored. The correlation coefficients between the total emission and other variables are shown in Figure 5.6. As we expect, the total emission and net emission have strong correlation (correlation coefficient = 0.97). Meanwhile, many other variables (v4, v15, v16, v17, etc.) show very weak correlation with the emission. These variables, if included in the inferential model, may influence the prediction accuracy since their variation appears not to influence the emission in the real process, but may cause variation of the emission level in the model. An absolute value of 0.3 is set as the threshold in this correlation analysis. Any variable, if its correlation coefficient fall in the range (-0.3 0.3), will not be included in the model. By using this as a variable selection criterion, only 33 variables are retained in the model for further analysis.

In this study, the correlation analysis is used to select the important variables. Many other methods, as discussed above, can be used to do variable selection. For example, we can build a preliminary PLS model. The variables which have little contribution to the inferential model, should have relatively small first weighting coefficients. Thus they can be deleted from the model. Figure 5.7 gives the comparison between the correlation coefficients and the first PLS weighting vector. The variation of first weighting vector in PLS is similar to that of correlation coefficients and these two methods are expected to give the similar results.

It is difficult to build a process model based on first principles and the process knowledge for the emission, because the mechanism influencing the emission is not known. In this case, building an empirical model is a realistic choice. Furthermore, the

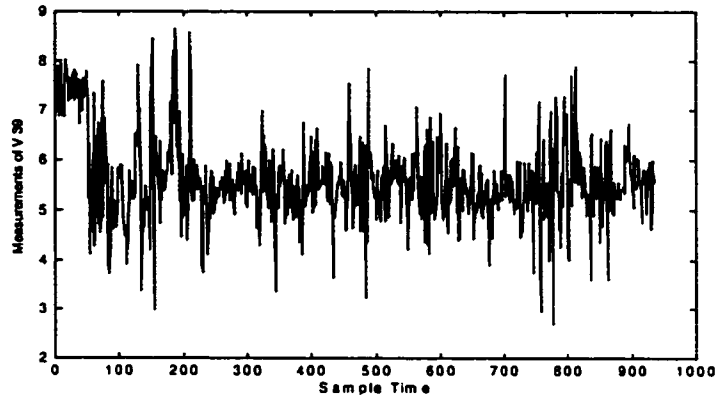


Figure 5.5: Time-series of variable 39

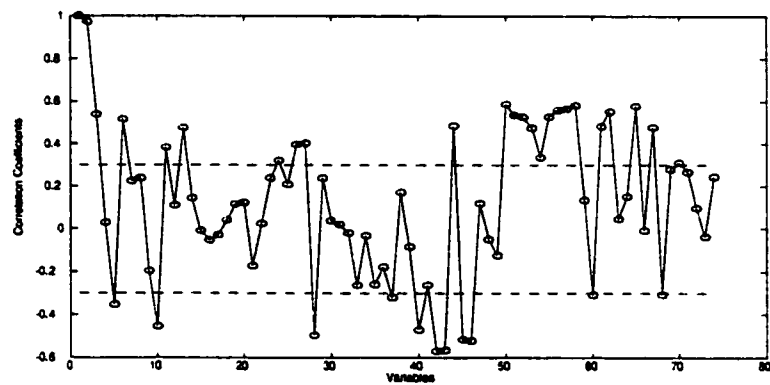


Figure 5.6: Correlation coefficients between total emission and other process variables

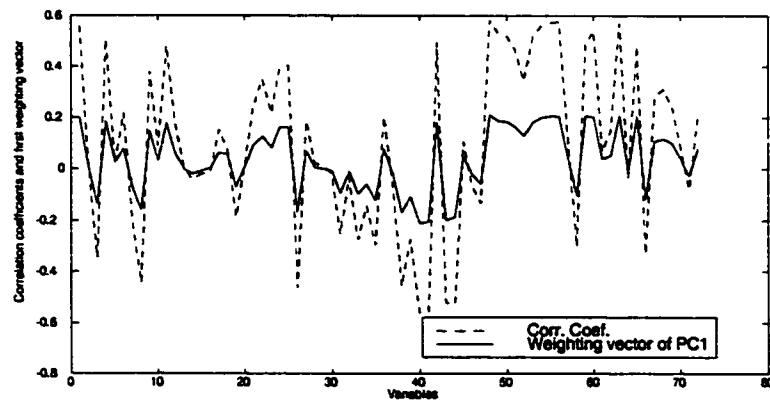


Figure 5.7: First loading coefficients in PLS model vs. correlation coefficients

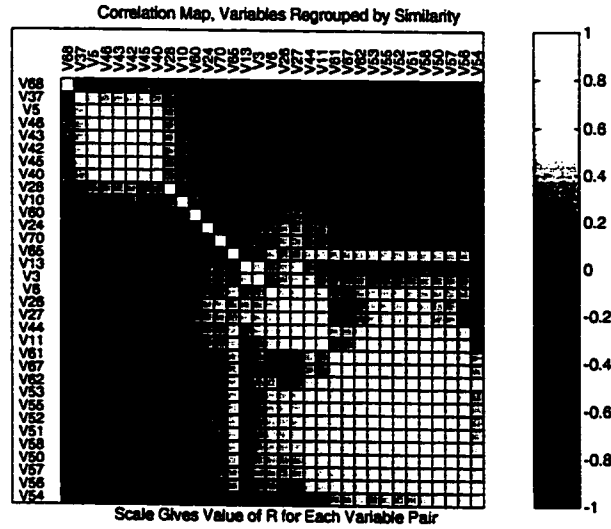


Figure 5.8: The cluster analysis display the correlated variables

ordinary least squares method is not numerical stable because of the highly correlated variables. Figure 5.8 shows the correlation map with variables regrouped by similarity. Highly correlated variables form several clusters in the figure which implies that the PLS method is a better choice in dealing with the collinearity in the data.

5.5 PLS modeling

The whole data set (936 samples) is divided into two parts. One set (including sample instants 0-850) is used for modeling. The other data set (including sample instants 851-936) is used for validation. In this study, only one variable (total emission) is included in Y block (known as PLS1 modeling). The X block contains 33 variables after variable selection. To ensure that the results are invariant to the choice of measurement units, all the variables are auto-scaled (zero mean, unit variance). In a static PLS model, the data matrices are arranged as follows:

$$\begin{aligned} X &= [X(t-1)] \\ Y &= [Y(t)] \end{aligned} \quad (5.7)$$

$X(t-1)$ contains 33 process variables at lag one measurement. Y contains the measurements of total emission at current time. The objective of the static PLS model

	X block		Y block	
LV#	This LV	Total	This LV	Total
1	54.99%	54.99%	48.91%	48.91%
2	7.05%	62.04%	22.93%	71.83%
3	3.00%	65.03%	4.20%	76.03%

Table 5.1: Variance captured by latent variables in standard PLS model

is to predict the emission value at next sampling time using 33 process variables at current time.

Cross validation is used to determine the number of latent variables in the PLS model. The cumulative PRESS (prediction error sum of squares) plot (Figure 5.9) indicates that 3 latent variables is a good choice. Table 5.1 lists the variance explained by each latent variable. With three latent variables, 65.03% percent of information in X block is utilized to explain 76.03% percent of information in Y block.

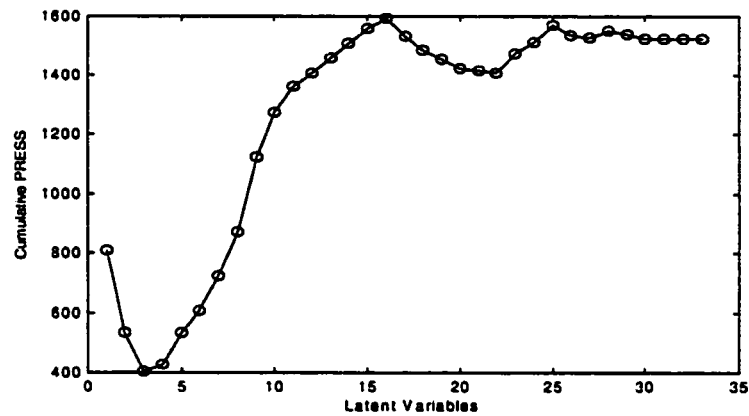


Figure 5.9: The cross validation suggest three latent variables should be kept in the model.

The model outputs, together with the real measurements, are displayed in Figure 5.10. We can see that although the model output can follow the trend of actual measurements, the fitting error is relatively large. The prediction error (after sample instant 850) is also large. The possible reasons for this big model-plant mismatch may be due to:

- The underlying assumption that the process is at a steady state does not hold all the time. The process appears to have gone through a dynamic transition

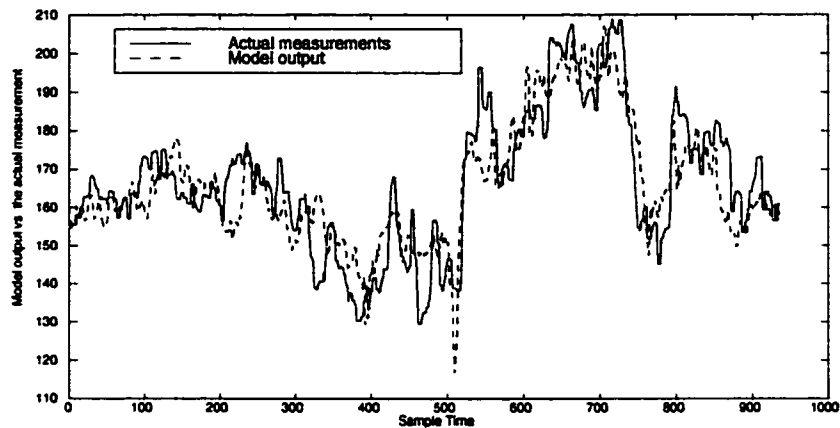


Figure 5.10: The Model output vs. the actual measurements using static PLS model.

around sample time 500. Thus, the static PLS model, which assumes that the process is at steady state, may not work well in modeling the dynamics in the process.

- The standard PLS assumes a linear algebraic model structure. The real process is nonlinear. If the process operates at a single steady state, it can be linearized by a PLS model around the operating point. However, it is difficult for PLS to account for the nonlinearity at different steady states.
- We may not measure all the variables that affect SO_2 emissions. In other words, some of variables that influence the emission level may not be included in the model due to unobservability.
- Some of the variation in the process may be random, thus unpredictable.

The possible remedies to improve the model are:

- Use PLS to extract the relationship between variables from normal process data. Including more data collected from different normal states can help PLS “learn” more about the process.
- Build a dynamic PLS model by including the lagged measurements of emission as the model input.

- Build a nonlinear PLS model. This can be done in two ways. One method is to incorporate some nonlinear elements (nonlinear transformation of original measurements) in the X matrix, which require the apriori knowledge about the process. The other method is to use some nonlinear modeling techniques to build nonlinear inner relationship for the latent variables. Neural network has been reported as one of these tools (Qin and McAvoy, 1992a).

In this analysis, we have attempted to build a dynamic PLS model and a nonlinear PLS model and found that a nonlinear model that ignores the dynamics in the process did not perform well in contrast to a dynamic linear model. Thus, we will only focus on the discussion of the dynamic linear model.

5.6 Dynamic PLS modeling

There are a number of approaches to deal with dynamics in a process. One method is to include a relatively large number of lagged values of the input variables in the input data matrix. This approach is used by Ricker (1988) and this results in a PLS Finite Impulse Response (FIR) models. This model is parsimonious in the parameters due to the dimensionality reduction capabilities of PLS. Another possibility is to include the lagged values of both the inputs and outputs in the input data matrix. This approach is used by Qin and McAvoy (1992) and this results in a parsimonious multivariate ARMA model. Both FIR and ARMA methods require a substantial increase in the dimension of X matrix. With the FIR approach, the column dimension of X is proportional to the specified length of the impulse response model. With the ARMA approach, the column dimension of X is proportional to the sum of the orders of the autoregressive and moving average parts of the model. These dimension can reasonably range from 5 to 50 times the original data dimensions for the FIR model and from 2 to 6 times for the ARMA model. Thus, there can be a significant increase in the column dimension of X . To overcome this difficulty, Kaspar and Ray (1993) proposed another method which permits the dynamics to be expressed as part of the inner relation and obviates inclusion of lagged values of the input and output variables in the input data matrix. The resulting dynamic PLS model can be easily used to design a control system by employing precompensators and postcompensators

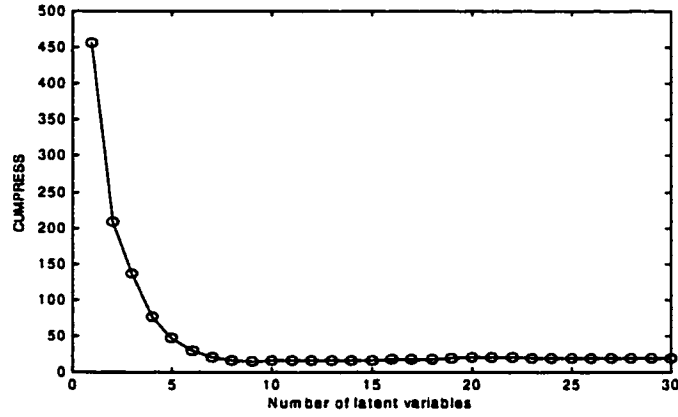


Figure 5.11: A plot of the cumulative PRESS. This suggests that 6 latent variables should be kept in the PLS model.

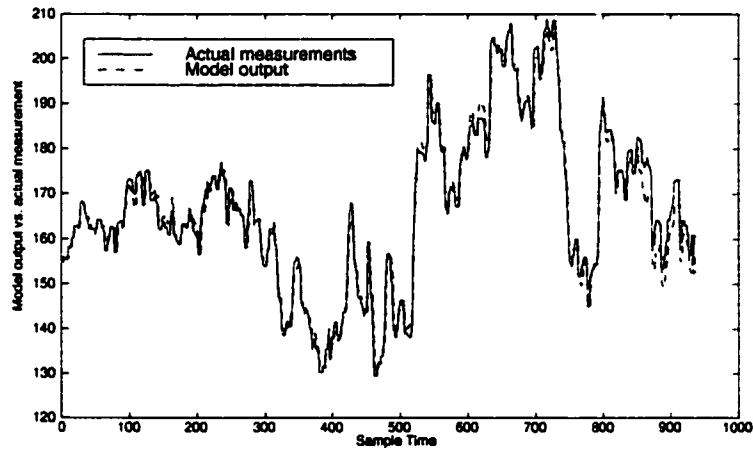


Figure 5.12: The fitting error is small

constructed from the input and output loading matrices.

In the emission data analysis, a first order ARX model is utilized to capture the dynamics in the process. The data matrix for dynamic PLS modeling is arranged as follows:

$$X = [X(t-1) Y(t-1)] \quad (5.8)$$

$$Y = [Y(t)] \quad (5.9)$$

$X(t-1)$ is same as that in the static PLS model. However, lag one measurement of total emission is also included in X matrix.

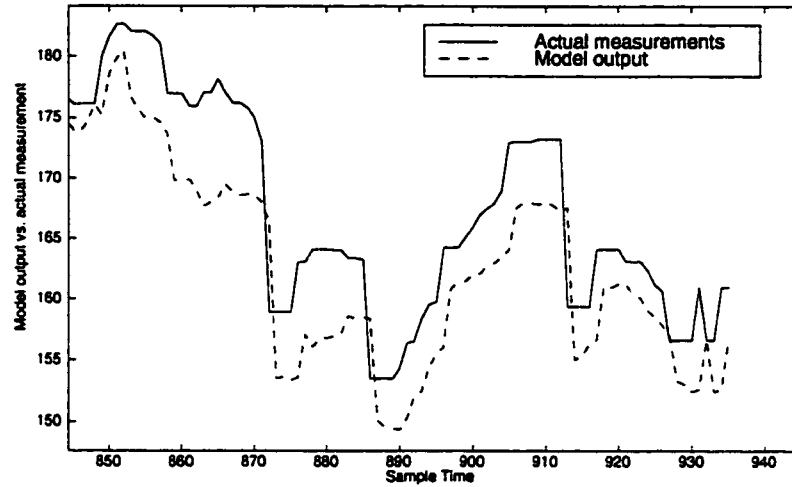


Figure 5.13: Comparison of model output and actual measurements.

	<i>X</i>		<i>Y</i>	
LV#	This LV	Total	This LV	Total
1	47.05	47.05	53.34	53.34
2	10.34	57.39	26.07	79.41
3	6.12	63.51	8.22	87.63
4	5.18	68.69	4.81	92.45
5	4.03	72.73	2.83	95.27
6	2.41	75.14	2.04	97.32

Table 5.2: Variance captured by latent variables in dynamic PLS model

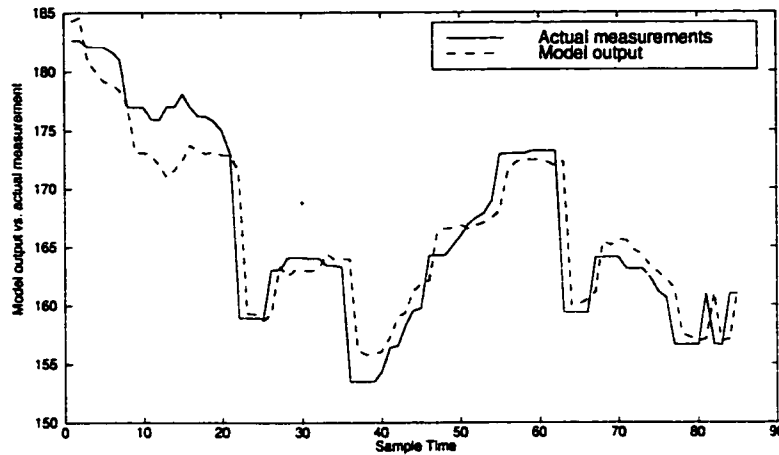


Figure 5.14: The model prediction vs. the actual measurements when the validation data is scaled by its own mean and variance.

The cumulative PRESS plot (Figure 5.11) indicates that 6 latent variables should be retained in the dynamic PLS model. The variance captured by the latent variables is shown in Table 5.2. Six latent variables can explain 75.14% of X and 97.32% of Y . The comparison between the model output and the actual measurement is shown in Figure 5.12. Very small fitting error (samples between 0-850) is observed in the plot. The zoomed figure related to validation data (after sample instants 850) is shown in Figure 5.13. The prediction error is large. However, the prediction and actual measurements trend very well.

Recall that the data used for building a PLS model was autoscaled. When the model is used for prediction, the same mean and standard deviation are employed to scale the recently available data. Scaling the data in this way assumes that the signal is stationary. In other words, the mean and variance of the data are not expected to change over time. However, the emission measurements are nonstationary signals and thus it is incorrect to scale the new measurements using the old mean and variance. If the validation data is scaled using the mean and variance estimated based on samples 850-936, we observed a significant improvement of the model prediction (Figure 5.14). Although this treatment of scaling is non-causal, and cannot be used on-line, it highlights the points that the new data should not be scaled using the old mean and variance in a nonstationary process.

To deal with the drifting mean and covariance in the process, the new process data should be scaled by the updated mean and covariance when the new data is available. For this purpose, we use an exponentially weighted moving average (EWMA) method to update mean and variance, as shown in Equation 5.10:

$$\begin{aligned} M_{t+1} &= \alpha M'_t + (1 - \alpha)M_t \\ V_{t+1} &= \beta V'_t + (1 - \beta)V_t \end{aligned} \quad (5.10)$$

where M_{t+1} is the mean used for scaling at current time. M_t is the previous mean used for scaling. M'_t is the estimated mean based on new process data. (length of new process data is m , which is a user specified parameter). α is a weighting factor to the new data ($0 \leq \alpha \leq 1$). The notation for the variance can be explained in a similar way. When the method is used on-line, the user needs to choose α , β and m . A large α and β tend to give the current measurements more weights and forget the old measurements quickly. The choice of m determines interval length needed to update the mean and variance. For example, if $m=5$, the mean and variance will be updated every 5 sample intervals. Let α and β equal to 0.025 and $m = 2$ in this analysis. Figure 5.15 shows better results in comparison with Figure 5.13 when the updated mean and variance are used.

PLS modeling combined with EWMA gives better performance when dealing with the nonstationary process. However, at least two issues need to be addressed with this method. One issue is how to choose a proper m . In other words, how to decide when we need to update the mean and variance. Some rules or algorithms that are able to detect the drifts of mean and variance need to be employed. The other issue is proper choice of weighting factors α and β . Strategies for identifying optimal weighting parameters need to be developed.

5.7 Root Cause Diagnosis

As we have seen in Figure 5.1, emission is at normal level before sample instant 500 and exceeds a certain criterion after sample instant 500. This case happens occasionally in the process. Whenever this occurs, engineers try to find the causes that result in undesirably high level emission. In the past, this was done by viewing

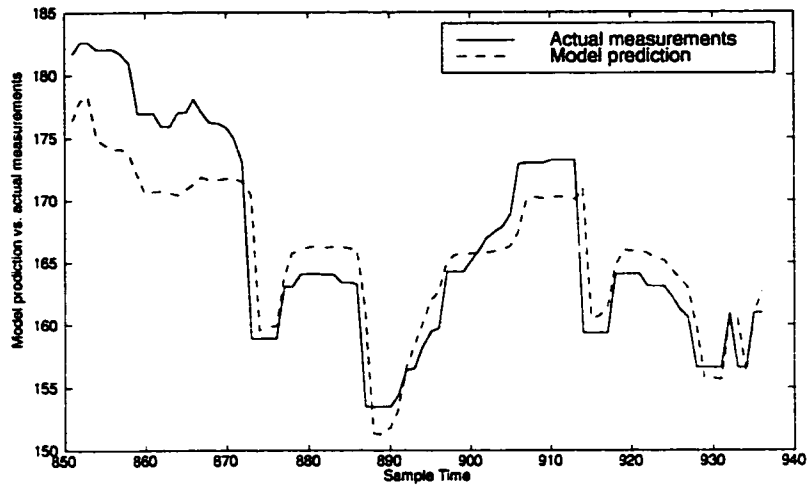


Figure 5.15: The comparison of model outputs and actual measurements with EWMA updating of mean and variance

Data Range	Corr. Coef.
100–300	-0.349
400–600	0.726
600–800	0.0017

Table 5.3: Correlation coefficients between net emission and v51 based on different section of data

the trend plot of individual process variables and the possible causes were identified based on engineer's experience. This univariate method is not only time consuming (over 100 process variables in the process need to be examined), but also subject to errors due to collinearity and noise masking the real root cause.

Figure 5.16 shows the trend plots of v39, v51, and the net emission, respectively. The measurements of V39 is almost at stable state, and unlikely to be the variable that causes the variation of emission around sample instant 500. V51 is almost constant before sample instant 500; however, it has a significant change around sample instant 500. Since the emission also undergoes a large change at sample instant 500, v51 is likely to be one of the variables pointing to the root cause. Table 5.3 shows the correlation coefficients between net emission and v51 at different periods of data. The correlation coefficients based on sample instants 400-600 is the largest, which indicates that v51 is highly correlated to the net emission during this period.

As shown in Table 5.3, the correlation coefficients between two variables change

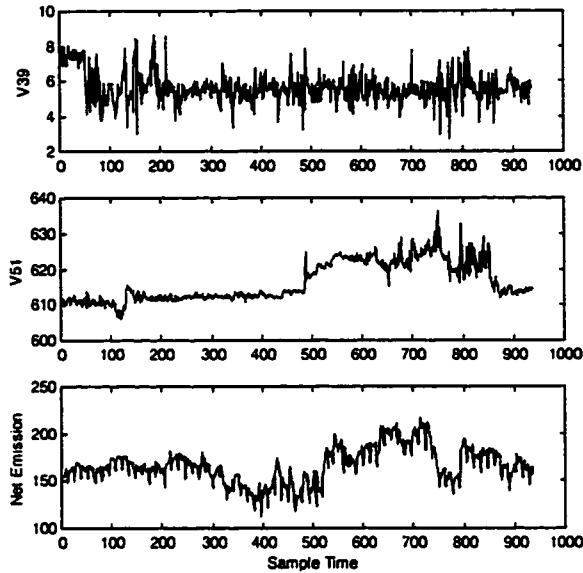


Figure 5.16: Trend plots of V39 (upper), v52(middle), and net emission(lower)

in magnitude and sign over time. This implies that the correlation matrix R is not a constant matrix. If the process is at an ideal steady state (all measurements are almost constant and only affected by measurement noise), the values of off-diagonal elements in R should be very small. In other words, variables should show very weak correlations between each other at steady state. The data collected at steady period would thus reveal little relationships among variables.

The PLS model based on different sections of data may reveal different relationship between the emission and the other process variables. This is evident since the process undergoes different states or changes during the time when the data was collected and the correlation also changes during this period. The data used for PLS modeling thus cannot reflect all the normal variation and conditions in the process. Thus a PLS model built off-line may have a large model-plant mismatch if it cannot be updated on-line. Should the plant change, the model should also be updated to reflect the current state of the process. This is the idea of adaptive modeling that has been investigated by many researchers (Ljung and Soderstrom, 1983; Goodwin and Mayne, 1987; Eykhoff, 1974).

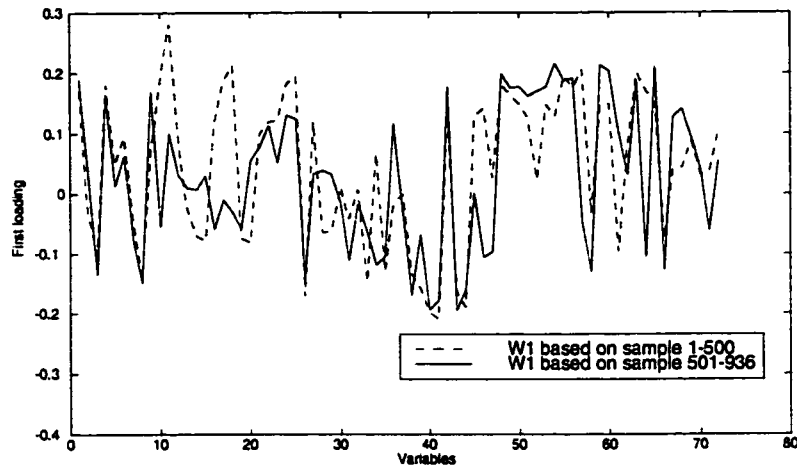


Figure 5.17: First loadings based on different section of data.

Weighting vectors in a PLS model are important in the sense that they reveal the relative importance of process variables to response variables (total emission and net emission in this analysis). The first weighting vector (associated with first principal component) in the PLS model provides more information than other weighting vectors, since the first PC usually explains the most variation in X block and at the same time contribute most to the prediction of Y compared to other PCs. Figure 5.17 compares the first weighting vectors based on the data between sample instants 1-500 and 500-936. Some weighting values remain unchanged (for example, weightings related to variables 1-10). However, other weightings show a large variation, such as variables 18, 45, etc. This phenomenon implies that continuous monitoring of the weightings for process variables may reveal the causes for the variation of emission when there is a significant change.

Color coding or chromatic graphing is utilized here to facilitate an easy interpretation of the results. In the following color figures, variables are reordered based on the similarity between each other (refer to Figure 5.8). (Note: all the color figures will be displayed in the text in black and white to facilitate the printing and the corresponding color figures are attached at the end of this thesis).

There are several ways to update the model recursively (refer to the discussion in Chapter 3). Here we compare the results from three methods:

1. When the new measurements are available (say 20 samples), the initial model

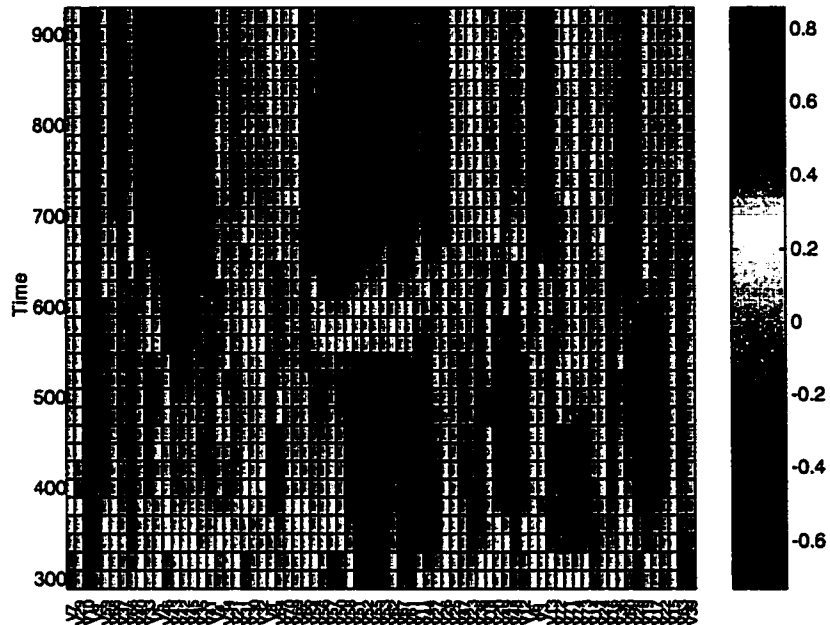


Figure 5.18: First weighting vector in recursive PLS model using all the available data.

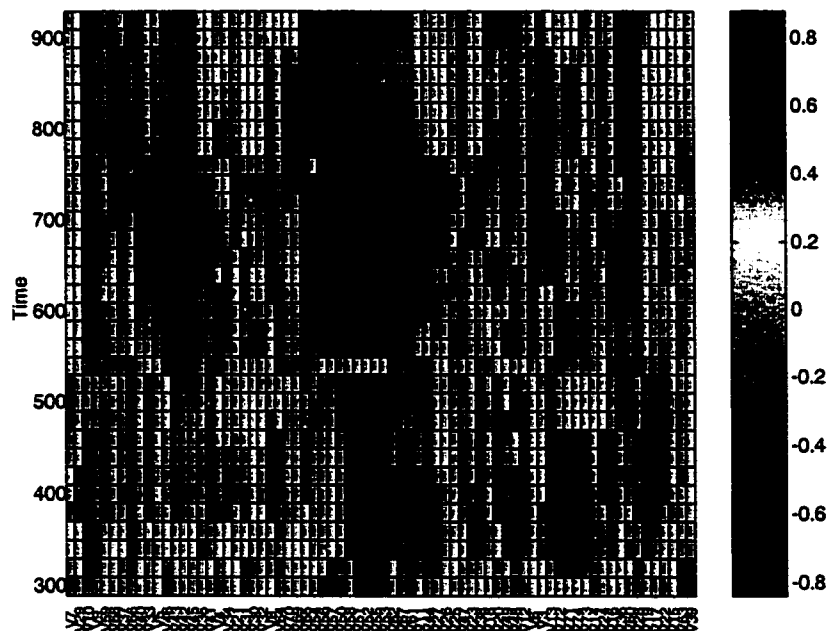


Figure 5.19: First weighting vector in recursive PLS model with EWMA updating method. ($\lambda = 0.9$)

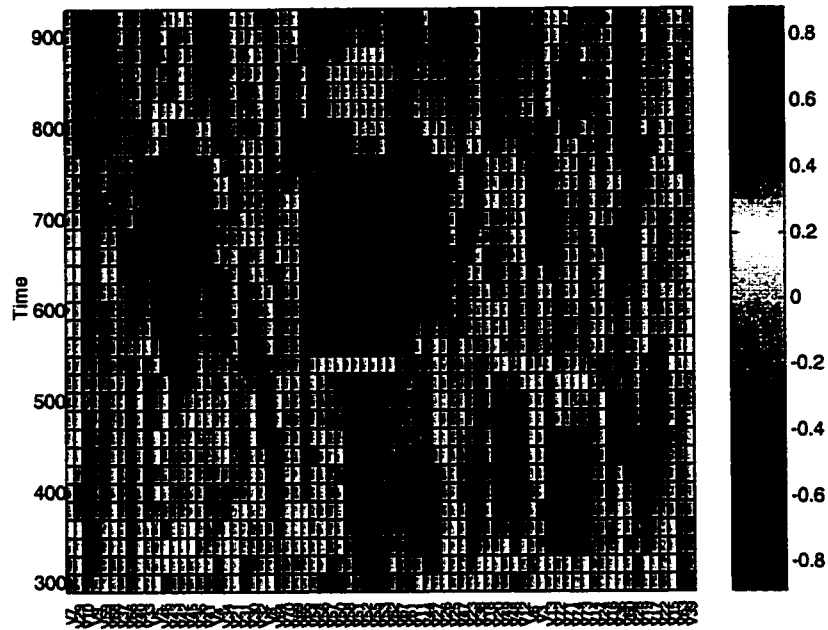


Figure 5.20: First weighting vector in recursive PLS model with fix data length, window length 300, updated every 20 sample.

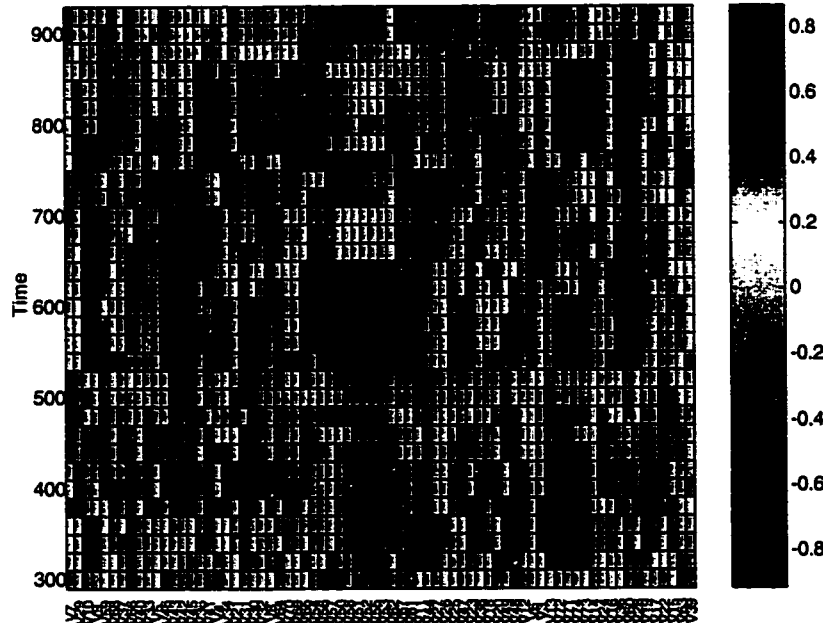


Figure 5.21: First weighting vector in recursive PLS model with EWMA updating. ($\lambda = 0.7$)

is updated including the old data together with the new one. No discounting of the old data is considered in this case. (Figure 5.18)

2. Exponentially weighted moving average updating of the model. The old data is discounted in building the model. The new data carry more weights than the old one. (Figure 5.19 and Figure 5.21)
3. A fixed length moving window is set to include the same amount of data in building the model. When some fresh data is available, it is included in the model. At the same time, some old data is deleted. (Figure 5.20)

In all three methods, the initial models are built using samples 0-300. All models are updated whenever 20 new observations are available. 72 process variables are included in X matrix. Total emission variable is included in Y matrix. First weighting vector in a static PLS model (lag one measurement of net emission is not included in Y matrix) is calculated recursively.

To facilitate the interpretation, the variables are regrouped based on similarity. Cluster analysis indicates the similarity between variables and shows how the variables should be grouped together. The X axis gives the variable index. The Y axis gives the current time. Each pixel in the figures represent the colour coded weighting values for a variable. For example, the first row from bottom represents the weighting values for the corresponding variables in PLS model based on samples 0-300. When next 20 sample are available, the models can be rebuilt by means of any of the three methods discussed above. The updated weightings are then displayed in a new block of colors in a horizontal line above the old one. The procedure is repeated until the end of the data length.

Figure 5.18 shows the change of weighting vector of PLS model if the model is updated using all the available data (no discounting to the new data). Weighting values for some variables are fairly stable, e.g. v7, v39, v15. Meanwhile, the weighting values are small, which implies that these variables have little relationships with emission. On the other hand, weighting values for some variables, for example v51, v52, v53, v55 etc. have undergone significant changes, both in sign (from negative to positive) and magnitude.

We know that emission level rise to abnormal level after sample time 500. Thus the variables whose weighting values indicate a significant change after sample instant 500 are most likely to be the cause that influences the emission. v11 v12 v61 v67 v62 v53 v55 v52 v51 v58 v50 v56 v57 have been picked out as such candidates. The limitation of using the first method is that all the available data is given equal weights in building the model, which makes the weightings insensitive to the current states of the process. In the second method, exponential weighted moving average (EWMA) updating of PLS model is utilized. The new data is given more weightings than the old. The weights on the old data decay exponentially by using a forgetting factor λ ($0 < \lambda \leq 1$). A smaller λ will forget the old data faster, which also make the weighting vector more sensitive to the new data. Figure 5.19 and 5.21 show the results with $\lambda = 0.9$ and 0.7 , respectively. More variation of the color can be observed in Figure 5.21 due to the smaller λ . This allows us to find a cause and effect relationship between emission and other process variables during a short period of time. In practice, if emission level is relatively constant and only violates the monitoring limit occasionally, a big λ should be chosen to avoid drastic and frequent change of the color.

Figure 5.20 shows the variation of first weighting vector when a fixed length moving window is used. The results are similar to EWMA method since both methods discount the old data in some way. A small window length is equivalent to using a small λ in the EWMA method.

The diagnostic results obtained from above three methods are similar. These results have proved to be very helpful in finding the cause that affects the emission. Usually, when the process is at steady state, the emission level is relative constant. The color coding graph cannot provide much information in this case. However, if the emission rises to an abnormal level, the system will undergo a transition i.e. it is in a different state. In this case, the weighting values for some variables will vary significantly (color changes correspondingly in the graph). The possible reasons that effect the emission thus can be found among these variables. Although the method cannot provide a unequivocal diagnostic results, it can quickly narrow the search range and give clues to the possible causes. For example, by simply observing the color variation after the sample instant 500, we can quickly discard many irrelevant

variables and focus only on 13 variables.

5.8 Concluding Remarks

A dynamic PLS model is more appropriate if the measurements from a process display not only strong crosscorrelation but also strong autocorrelation. A dynamic PLS model can be achieved by including the lagged variables in the predictor matrix or expressing dynamics in the inner model. In a time-varying process, it is essential to update the mean and variance over time. The newly available data should be scaled using the updated mean and variance to avoid any possible bias in model prediction. The weighting vectors in a PLS model reveal the relative importance between prediction variables (variables in X block) and response variables (variables in Y block). Using color blocks to display the variation of the weighting vector, we show how this method can be effectively applied in diagnosing the root causes for abnormal emission.

Chapter 6

MSPCA-Combination of Wavelets with PCA

6.1 Introduction

Statistical Process Monitoring (SPM) via PCA involves the use of Hotelling T^2 and Q (also known as SPE) charts. Time-independency and normal distribution of the measurements and residuals of PCA model are required for obtaining the statistical limits for the T^2 and Q charts, respectively. Conventional PCA is thus ideally suited for monitoring steady state processes based on the assumption that the measurements are time independent (uncorrelated) and normally distributed. Typically, most of the processes are in dynamic state, with various events occurring such as abrupt process changes, slow drifts, bad measurements due to sensor failures, human errors, etc. Data from these processes are not only cross-correlated, but also auto-correlated. Applying conventional PCA directly to dynamic systems usually results in false alarms, making it insensitive to detect and discriminate different kinds of events.

Every event is associated with a certain frequency band according to its power spectrum. Wavelets are emerging tools for decomposing a signal into various frequency bands providing simultaneous time-frequency domain analysis. Wavelet decomposition of a signal results in approximately decorrelated wavelet coefficients of the stochastic part of the signal and a few set of large coefficients containing the trend of the process. Thus, wavelets provide a good separation between the deterministic and stochastic parts of a measured signal (Donoho *et al.*, 1995). Typically, measurements from chemical processes are autocorrelated and thus unsuitable for SPM using

PCA. Harris and Ross (1991) have suggested changes in the confidence limits when monitoring univariate data. Their results do not directly apply to the multivariate case. However, in this work, it is shown that through wavelets decomposition, the resulting wavelets coefficients are almost decorrelated. PCA models can then be built based on these wavelet coefficients over each frequency (scale) band. This strategy helps in overcoming the above mentioned shortcomings of conventional PCA for SPM. Consequently, MultiScale PCA (MSPCA) retains the statistical basis for the monitoring charts. Moreover, since each event occurs over a certain frequency band, MSPCA possess greater sensitivity in fault detection and process changes. These ideas are illustrated by a simulated example.

6.2 Introduction to Wavelets

The ideas of wavelets originated from the works in engineering (subband coding), physics (coherent states, renormalization group), and pure mathematics (Calderon-Zygmund operators). The name wavelets were coined in early eighties by some French researchers (Morlet *et al.*, 1982; Grossmann and Morlet, 1984). Since then, wavelets have attracted a great deal of attention from scientists and engineers in various disciplines. In this section, we give a brief introduction to wavelets and Mallat's multiresolution analysis, which is a necessary background for understanding Multiscale PCA. More details about wavelets theory and their applications can be found in many references (Daubechies, 1992; Chui, 1992; Koornwinder, 1993; Chan, 1995; Mallat, 1998).

The wavelet transform is a tool that transforms data, functions, or operators into different frequency components, and then studies each component with a resolution matched to its scale (Morlet *et al.*, 1982). Fourier transforms provide a perfect frequency resolution of a signal but completely lose the information in the time-domain, i.e., the Fourier transform cannot provide the time-frequency localization at the same time. This means that although we are able to get the frequencies present in a signal, we do not know when they are present.

The Wavelet transforms, in contrast, resolve signals both in time and frequency

domain. A wavelet of ψ is a function of zero average:

$$\int_{-\infty}^{+\infty} \psi(t) dt = 0 \quad (6.1)$$

Wavelets are obtained through dilation and translation of $\psi(t)$

$$\psi_{\tau,s}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-\tau}{s}\right) \quad (6.2)$$

Function $\psi(t)$ sometimes is called “mother wavelet” because it represents the prototype from which wavelets functions $\psi(s, \tau)$ are generated. τ is the translation parameter which determines the location of the wavelet in the time domain, and s is the dilation parameter which determines the location in the frequency domain.

Let $f(t)$ and $g(t)$ be two functions in $L^2(R)$. The functional transformation or inner product of $f(t)$ and $g(t)$ is defined as

$$\langle f(t), g(t) \rangle = \int_{-\infty}^{+\infty} f(t) \cdot g^*(t) dt \quad (6.3)$$

According to definition of inner product, the continuous wavelet transform of a signal can be defined as the inner product of the signal with the wavelets basis functions

$$wf(\tau, s) = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{s}} \psi^*\left(\frac{t-\tau}{s}\right) dt \quad (6.4)$$

A family of discrete wavelets is represented as the dyadic dilations and translations of a *mother* wavelet $\psi(k)$:

$$\psi_{m,n}(k) = 2^{-m/2} \psi(2^{-m}k - n) \quad (6.5)$$

where m and n are the dilation and translation parameters respectively. $\psi_{m,n}$ can constitute a set of orthonormal basis functions for $L^2(R)$ with the proper choice of ψ (Daubechies, 1992). Any function f in $L^2(R)$ can then be represented, up to arbitrarily small precision, by projections onto the set of basis functions $\psi_{m,n}$ (Daubechies, 1992).

The Discrete Wavelet Transform (DWT) analyzes the signal at different scales (or over different frequency bands) by decomposing the signal at each scale into a *coarse* approximation (low frequency information) and *detail* information (high frequency information). DWT employs two sets of functions, the *scaling* functions

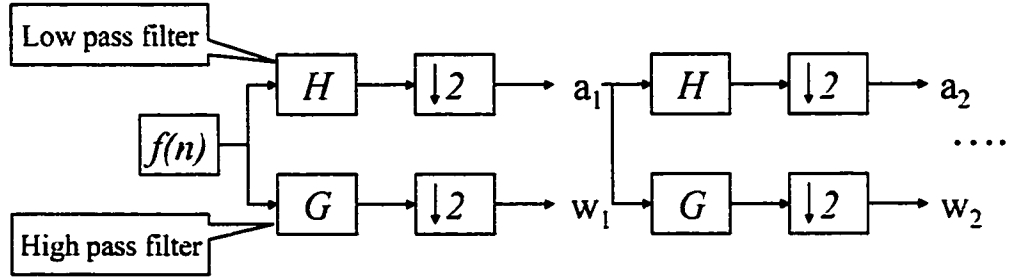


Figure 6.1: $f(n)$ is the signal at finest scale. By passing a series of high and low pass wavelets filters, it is decomposed into signals at different scales.

and *wavelet* functions, which are associated with low pass filter H and high pass filter G , respectively. Mallat's decomposition algorithms (Mallat, 1989) provide a simple means of transforming data from one level of resolution, to the next coarser level of resolution. Let a_{m-1} denote the scaling function coefficients at scale $m - 1$, the wavelets and scaling function coefficients at level m is

$$w_m = Ga_{m-1}, \quad a_m = Ha_{m-1} \quad (6.6)$$

Figure 6.1 shows this procedure. Note that each time after the decomposition of a_m , the resulting a_{m+1} is down sampled by 2. This is due to the fact that decomposition of a signal into the coarser scale will double the frequency resolution and result in redundancy of the information. In other words, only half of the number of samples is needed to characterizes the signal at current scale. The raw data from a process $f(n)$ can be considered at the finest level $f(n) = a_0$. Through multi-resolution decomposition, the raw data can be decomposed into several sets of wavelet coefficients at different scales and a set of scaling function coefficients at the coarsest scale. Coefficients at different scale contain the information over various frequency bands in the original signal.

The inverse Mallat transform is a transformation from the coarser level $m - 1$, back to the finer level m . For example, a_m can be reconstructed based on a_{m+1} and w_{m+1} , and the original signal may be completely reconstructed using its wavelets coefficients at all scales and scaling function coefficients at the coarsest scale.

Two important applications: data compression and data filter can be achieved through Mallat's transformation and inverse transformation. Since the total length of

all scaling function and wavelets coefficients together is equal to the length of original signal due to the down sampling, there is no extra burden to store those coefficients. Moreover, some wavelets coefficients may lie in the frequency band in which we are not interested, and thus can be discarded. This means that the useful information in a signal can be represented and stored by fewer wavelets and scaling function coefficients. On the other hand, when we use this length-reduced wavelets coefficients to reconstruct the original signal, the noise or irrelevant information localized in those uninterested frequency bands is automatically filtered out.

6.3 Statistical Assumption of SPE and T^2

Recall that the confidence limits calculated in Equation 2.23 are based on the assumptions that the measurements are time independent and normally distributed in the multivariate sense (Johnson and Wichern, 1992; Tracy *et al.*, 1992). The confidence limits in Equations 2.25-2.27 were derived assuming that errors were random with zero mean and Gaussian distribution (Jackson and Mudholkar, 1979). Although, the underlying distribution of the residuals and the scores can vary substantially from the Gaussian assumption without affecting the results due to the central limit theorem (Wise *et al.*, 1990). The highly autocorrelated data would certainly affect the confidence limits for the T^2 chart. In practice, measurements from dynamic chemical processes do not satisfy the assumptions on the measurements resulting in the loss of statistical basis for the T^2 charts. However, not enough attention has been paid to these underlying assumptions (Luo *et al.*, 1998). On the other hand, improper choice of number of PCs retained in the model may lead to the autocorrelated residuals, which will influence the confidence limits for SPE chart. Consequently, conventional PCA is not suitable for monitoring dynamic processes due to the presence of nonstationarities and time dependencies.

We illustrate this point by means of an example where the autocorrelated measurements: u , z and p are given by

$$u(k) = 0.7u(k-1) + w(k-1) \quad (6.7)$$

$$z(k) = 0.8z(k-1) + 0.3u(k-1) \quad (6.8)$$

$$p(k) = 0.5p(k-1) + 0.2u(k-1) \quad (6.9)$$

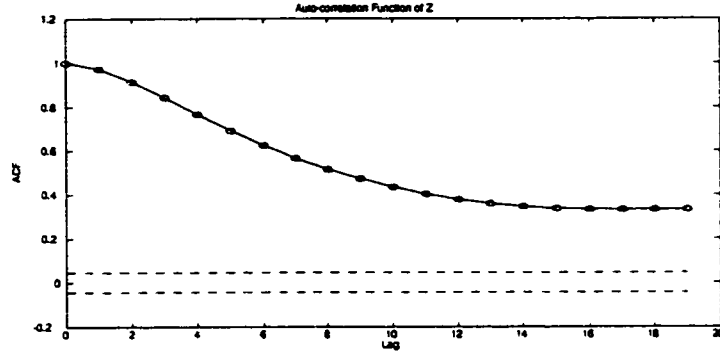


Figure 6.2: The ACF of z indicates that the data is time-dependent.

where $w(t)$ is the white noise with unit variance. The data matrix, X for a steady-state PCA analysis is arranged as follows:

$$X = [u(k) \ z(k) \ p(k)] \quad (6.10)$$

The auto correlation function(ACF) of z is shown in Figure 6.2. The strong autocorrelated feature indicates the strong time-dependence of the data. Consequently, the first PC also exhibits significant autocorrelation as shown in Figure 6.3. In this case, statistical control limits obtained for the T^2 chart based on the time independency assumption are not valid. In order to account for the dynamics/autocorrelation, we consider the lagged data matrix for the dynamic PCA model given in (Ku *et al.*, 1995):

$$X = [u(k) \ z(k) \ p(k) \ u(k-1) \ z(k-1) \ p(k-1)] \quad (6.11)$$

However, the dynamic PCA model using the lagged data matrix as shown in Equation 6.11 suffers from certain drawbacks. Firstly, the order of the process and the associated time delay are usually unknown. Time delay estimation of chemical processes is a non-trivial problem. It implies here that the lagged data matrix might not capture the true dynamics of the process. Secondly, the incorporation of lagged variables does not aid in decorrelating the measured variables. Thus, the assumption of time-dependency still remains violated, and hence invalidating the statistical basis obtained with the dynamic PCA model.

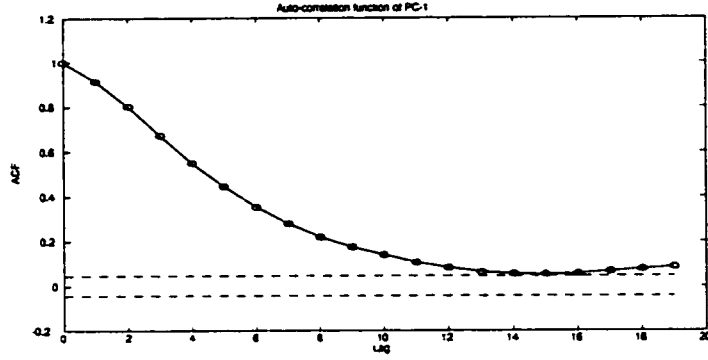


Figure 6.3: ACF of PC-1 (dynamic PCA model) indicating highly auto-correlated scores.

6.4 Combining Wavelets with PCA-MSPCA

Measured variables from a dynamic process may contain contributions from several events, such as process dynamics, sensor noise and fault, parameter drifts, and operator-induced actions. Every event has its own frequency and time features. For instance, components of measurement noise are mainly localized in the high frequency band, while basic process dynamics are mainly localized in the low frequency band. Since the data from practical processes are multiscale in nature, analyzing a process signal over various frequency ranges or scales may provide a multiscale, hierarchical description of the signal (Bakshi and Stephanopoulos, 1994). Consequently, PCA models built on wavelet transformed data at various scales or frequency bands would have some advantages. Each PCA model at a certain scale can be expected to have a greater ability to detect events whose spectrum is most significant in this scale. Here, PCA models are based on the wavelet coefficients, rather than the reconstructed data over certain frequency bands. The benefit of monitoring the wavelet coefficients is that these coefficients contain no significant correlation, while the reconstructed data still retains some time-dependent features. These issues are illustrated using the data generated in the earlier example. A 4-level wavelet transformation of the data is carried out yielding the representation of the signals over 5 frequency bands or scales (the lowest frequency band and 4 sets of higher frequency bands). Figure 6.4 shows the wavelets and scaling function coefficients at different scales. ACF of reconstructed data over the second level high frequency band is shown in Figure 6.5. Figure 6.6

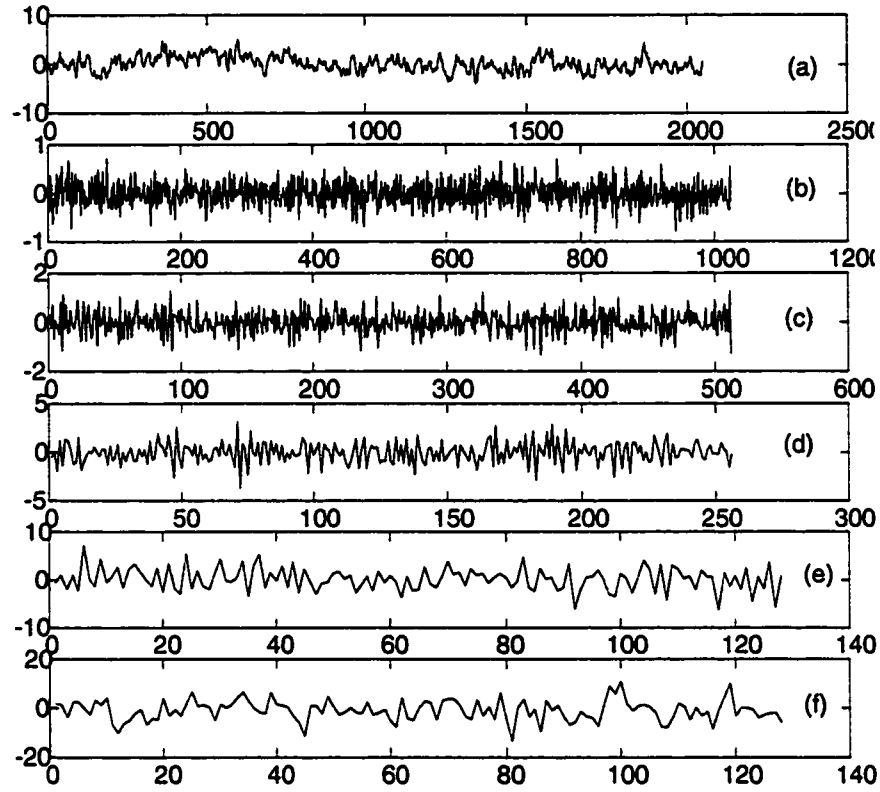


Figure 6.4: Original data of Z (a); Wavelets coefficients from first H.F. band to the Fourth H.F. band (b)-(e); Scaling function coefficients at the coarsest scale (lowest frequency band) (f)

shows the ACF of wavelets coefficients at different frequency band. It is easy to see that the wavelet coefficients contain no significant correlation while the reconstructed signal still contains significant correlation.

The idea of multiscale PCA is to build models for these wavelet coefficients individually at different scales. The PCA models at each scale are combined to monitor the process and are expected to be more sensitive to abnormal events than conventional PCA. Note that a dyadic length of data is required for wavelet decomposition. When the MSPCA is used online to monitor a process, it is necessary to incorporate the fresh data timely and also keep the dyadic length of data. Bakshi (1998) suggested a sliding window method to apply multiscale PCA on-line.

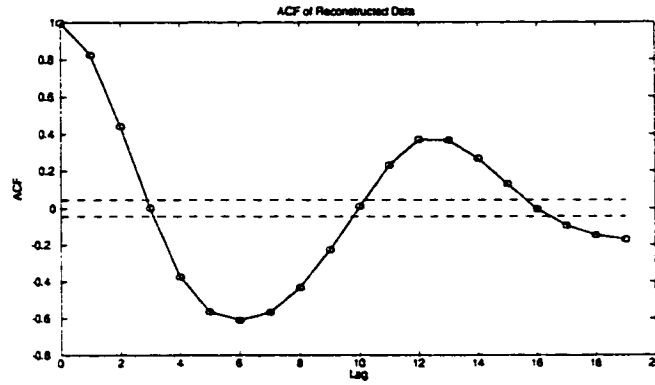


Figure 6.5: ACF of reconstructed data from the second level H.F. band.

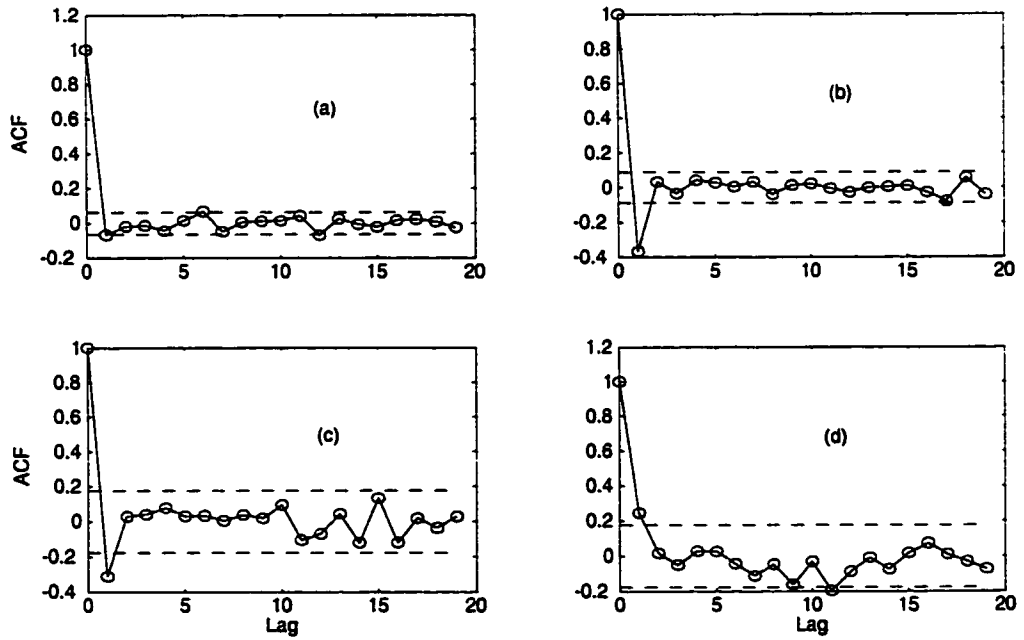


Figure 6.6: ACF of wavelets coefficients at first H.F. band or finer scale (a), second H.F. band (b), fourth H.F. band or coarser scale (c), and fifth H.F. band or coarsest scale(d).

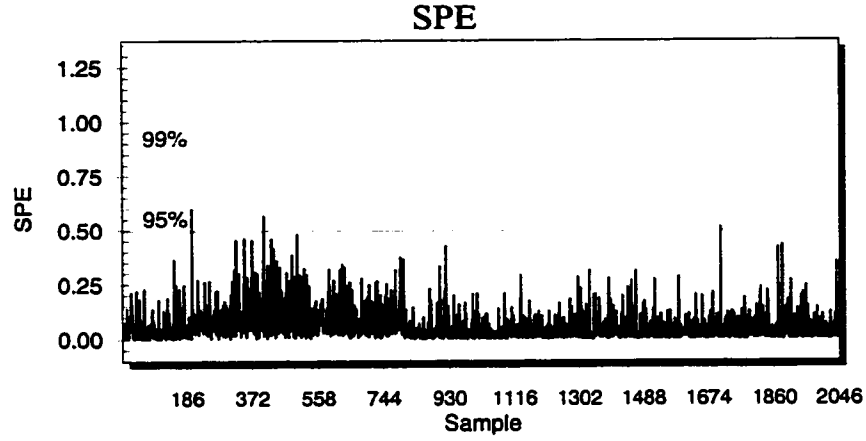


Figure 6.7: SPE chart with dynamic PCA

6.5 MSPCA for Dynamic Process Monitoring-A illustrated Example

In this section, we illustrate the effectiveness of MSPCA using the aforementioned example in Equations 6.7-6.9. A sensor failure is simulated by introducing a sudden mean shift of magnitude 1.5 in z between sample times 200-800. For comparison purposes, we show the results obtained using dynamic PCA. Results obtained using steady-state PCA are not shown here since it is well-known that steady-state PCA cannot perform well for dynamic processes.

The SPE and T^2 charts obtained using dynamic PCA are shown in Figures 6.7 and 6.8, where we have employed the same lagged data matrix as in Equation 6.11. The SPE chart is unable to detect the sensor fault. The T^2 chart can detect the sensor fault, but it cannot report a fault unequivocally since the scores and residuals go back and forth across the confidence limits.

Besides, the confidence limits in the T^2 chart are based on time-dependent measurements. Hence, the statistical basis is no longer valid.

Since sudden changes in processes contain mainly the high frequency component, the beginning and end of abrupt events are expected to reflect in the PCA model based on the wavelet coefficients at finer scale (first level of decomposition). The SPE chart shown in Figure 6.9 clearly indicates this point. The T^2 chart with MSPCA at finer scale is not shown here because of its insensitivity to such a change. A note of

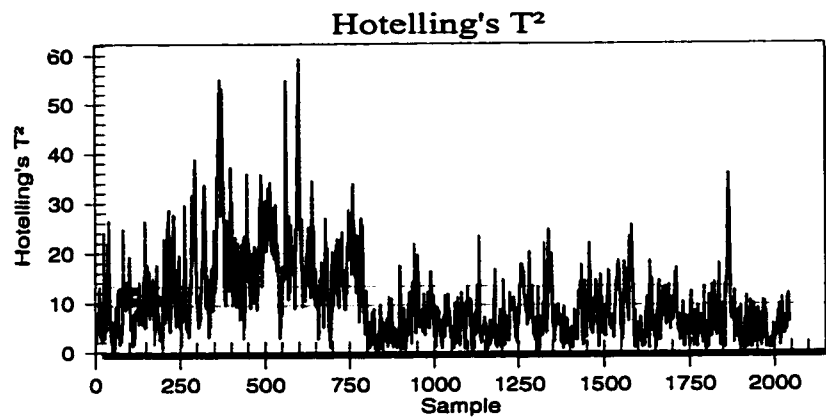


Figure 6.8: T^2 Chart with dynamic PCA

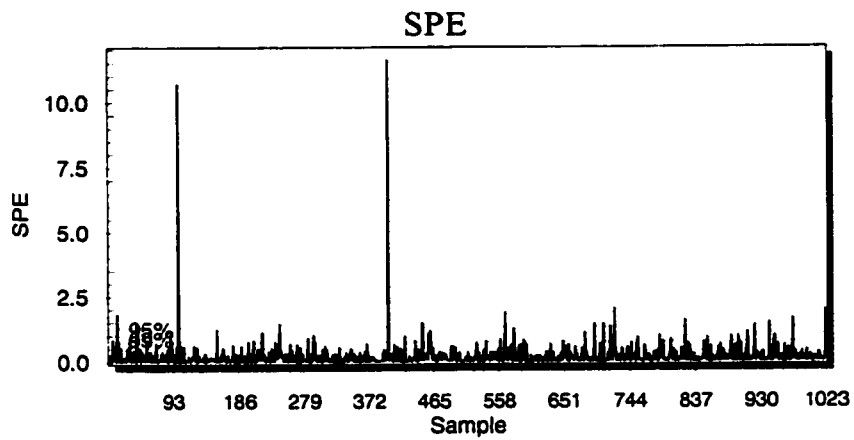


Figure 6.9: SPE chart with MSPCA at finer level.

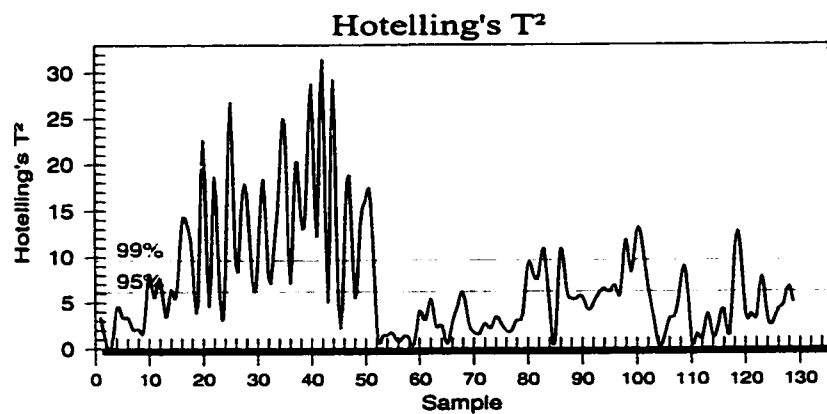


Figure 6.10: T^2 chart with MSPCA at coarsest level.

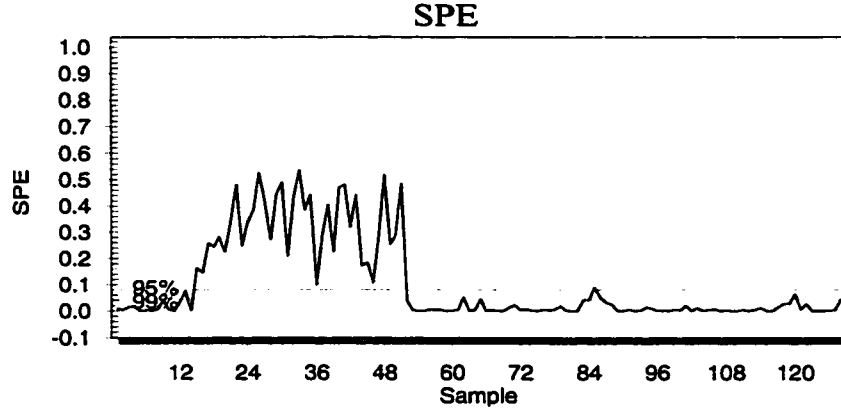


Figure 6.11: SPE chart with MSPCA at coarsest level.

caution is required while interpreting the time axis in the monitoring charts obtained using PCA on wavelet coefficients. At finer level, the time resolution is halved and therefore should be doubled when going back to the original time domain. While the wavelet coefficients at finer level (the highest frequency band) are primarily sensitive to sudden changes, the ones at the lower frequency bands detect the persistence in the fault. For this purpose, we make use of the T^2 and SPE charts using the coefficients at the coarser levels. Here, we only show the charts using the coarsest level coefficients. Figures 6.10 and 6.11 for MSPCA at coarsest level clearly show the persisting fault. The SPE chart is more sensitive than T^2 chart in this case. Since the original signals have been decomposed to four levels (5 frequency bands) in this example, we need to multiply the time index in Figure 6.10 and 6.11 by 2^4 to get the interval of fault persistence in the original time domain. The ACF of wavelet coefficients at the coarsest and finer levels are shown in Figure 6.6. The wavelet coefficients at these levels are almost decorrelated. Thus, the confidence limits for the monitoring charts can be correctly calculated in this case, providing a sound statistical criteria to detect an abnormal event.

6.6 Conclusion & Discussion

The underlying assumptions of applying PCA have not received enough attention in the past years. Since most industrial processes are dynamic systems, the measure-

ments are usually time dependent. Results of applying PCA to monitoring these systems will be severely affected without taking into account the time-dependency of the measurements. Wavelet transformation of measurements results in signals with no significant correlation. Monitoring charts with PCA models based on the transformed data comply with the underlying statistical assumptions. Furthermore, we have shown by example that MSPCA is more sensitive to small disturbances than dynamic PCA .

Typically, the wavelet coefficients at the coarsest level would primarily contain the deterministic part of a signal while the higher frequency bands contain mainly the stochastic part. In this example, it is a matter of coincidence that the wavelet coefficients at the coarsest level are approximately decorrelated as well. In practice, for reasons mentioned above, the confidence limits for the charts in Figures 6.10 and 6.11 may not be valid. In such a case, we could monitor deterministic or low-frequency changes with heuristics based on process knowledge.

In this work, we have discussed the issues involved in conventional PCA and a way to overcome some of the drawbacks therein. Extensions of off-line MSPCA to on-line MSPCA is possible with some enhancements in the methodology (Bakshi, 1998).

Chapter 7

Conclusions

7.1 Contributions of thesis

In summary, the contributions of the thesis are:

- **A simple and intuitive flow rate experiment has been employed to demonstrate the methodology of using PCA to do process monitoring. A physical set up for this experiment exists in the Chemical Engineering Department at the U of A. For the purpose of this illustration, simulation runs were conducted.**

This simulated example involves flow rate and temperature measurements of water flow rates. The correlation between these variables are simple and easy to be understood. Principal Component Analysis was used to detect different types of sensor faults happened in this experiment. The minor component, which was usually ignored by practitioners, was shown to be very useful to reflect the relationships between process variables. When a time delay exists between variables, it is essential to shift or lag the measurements correspondingly to allow the PCA model to capture the true relationship between variables.

- **A recursive PCA strategy was implemented on a pilot scale Continuous Stirred Tank Heater system.**

Most conventional methods work with fixed PCA or PLS model. When such a fixed model is used in practice, it usually results in a large model-plant mismatch when PLS is used for modeling, and false alarms when PCA is used for Statistical

Process Monitoring. This is because it is impossible to include all the normal data when building a model. On the other hand, the process can drift due to various reasons, such as the aging of equipments and catalysts or the process can be changed to meet different specification of the product. A fixed model cannot adapt the process change and therefore results in the aforementioned problems. In this thesis, the experiment results based on the CSTH system is used to illustrate this point. A recursive PCA algorithm is then implemented on this CSTH equipment for process monitoring. The results show that the on-line updated RPCA can follow the process change very well. The issue of choosing a proper forgetting factor to trade off the sensitivity and robustness of a PCA model is illustrated through this example.

- **A well performed PLS model has been built based on a real industrial data set. Using the same data set, continuous monitoring of the variation of PLS weightings has been proposed and verified as an effective method to find the root causes for the abnormal operations.**

A PLS model has been built to capture the relationships between the emission and other process variables in an industrial process. Although conventional PLS method is believed to be most suitable to model a steady state system, it has been shown in this thesis that a dynamic PLS model is more appropriate when the data display dynamic characteristics. Some issues in PLS modeling, such as data pretreatment, variable selection and choice of number of latent variables, are discussed in detail. Through color coding for PLS weightings, one can observe the color variation of the color blocks to find the possible root causes for abnormal events. In fact, the PLS weighting vectors reveal the relative importance of each variable in X block relating the variables in Y block. The color coding of the PLS weightings provides an intuitive and easy way to monitor the changes of relationships between variables.

- **Multiscale PCA (MSPCA) is proposed as an effective approach to overcome the limitation of using Hotelling T^2 and SPE in monitoring time-varying processes.**

The measurements from a time-varying process present strong autocorrelation. The underlying assumptions (time independent and random distribution) to obtain the control limits of T^2 and SPE are no longer valid in this case. The original signal can be decomposed into wavelets coefficients at different frequency bands through wavelets decomposition. In this thesis, we have shown that the resulting wavelets coefficients are almost decorrelated. Thus a PCA model based on these wavelets coefficients not only meet the underlying assumption of using statistical charts but also increase the sensitivity in detecting small disturbances. All these points have been well illustrated through a suitable simulation example.

7.2 Future Work

Here, we list some directions for future research.

- The optimal determination of the number of latent variables in PCA and PLS needs further investigation. Although there are a number of methods that can be used to do this, we still do not have a general guideline or rules to do so. Cross validation has been widely used to choose the number of latent variables. But it does not work well when the measurements are time-varying and autocorrelated. Thus more work needs to be done on this subject.
- When the recursive PCA is used to monitor a process, the RPCA may update the model to fit the small sensor drift other than the normal process drift. A strategy needs to be developed to differentiate the process drift from small sensor drift or an incipient fault. Recent work in this area has been reported by (Stork and Kowalski, 1999; Qin and Li, 1999).
- How to select a proper forgetting factor μ for a RPCA model needs further investigation. Forgetting factor in RPCA can trade off the robustness and sensitivity of RPCA model. A good choice of forgetting factor for a given process can allow RPCA to follow the normal process change and at the same time keep the sensitivity in detecting small disturbance.

- Conventional PLS method is a linear method and most suitable for dealing with the steady state process. Although there are already some studies in nonlinear and dynamic PLS. More work needs to be done in nonlinear and dynamic PLS to incorporate severe nonlinearities and dynamic in real industrial data.
- Both PCA and PLS are data-driven based method. Thus the acquisition of good quality data is the key in applying these statistical methods successfully. Real industrial data are usually contaminated with noise, disturbance, sensor failure and also contain a lot of missing data. These data cannot be used in building the model directly. A good data pretreatment strategy needs to be developed. Moreover, when the models are used on-line, a robust on-line filter or mechanism that can pretreat the real time on-line data is essential for a good application of these methods.
- Wavelets are an excellent tool that can provide the analysis of a signal both in time and frequency domain. Each type of abnormal event is associated with a inherent frequency pattern. Combining wavelets with PCA can decompose a signal into various frequency bands and make the detection of abnormal events more sensitive. Multiscale modeling has been reported recently by Bakshi (1999). It is worthwhile to carry out more work in multiscale data analysis and combine it with other statistical tools.

Bibliography

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Auto. Cont.* **19**, 716–723.
- Bakshi, B.R. (1998). Multiscale pca with application to multivariate statistical process control. *AIChE Journal* **44**(7), 1596–1610.
- Bakshi, B.R. (1999). Multiscale analysis and modeling using wavelets. *Journal of Chemometrics* **13**, 415–434.
- Bakshi, B.R. and G. Stephanopoulos (1994). Representation of process trends-iii. multiscale extraction of trends from process data. *Computers Chem. Engng.* **18**(4), 267–302.
- Blom, H.A. (1996). Indirect measurement of key water quality parameters in sewage treatment plants. *Journal of Chemometrics* **10**(5-6), 697–706.
- Boggia, R., M. Forina, P. Fossa and L. Mosti (1997). Chemometric study and validation strategies in the structure activity relationships of new cardiotoxic agents. *Quantitative Structure-Activity Relationships* **16**(3), 201–213.
- Centner, V., D.L. Massart, O.E. deNoord, S. deJong, B.M. Vandeginste and C. Sterna (1996). Elimination of uninformative variables for multivariate calibration. *ANALYTICAL CHEMISTRY* **68**(21), 3851–3858.
- Chan, Y.T. (1995). *Wavelet Basics*. Kluwer Academic Publishers.
- Chen, G. and T.J. McAvoy (1998). Predictive on-line monitoring of continuous process. *Journal of Process Control* **8**(5-6), 409–420.
- Chui, C.K. (1992). *Wavelets: A Tutorial in Theory and Application*. Academic Press, CA.
- Daubechies, I. (1992). *Ten Lecture on Wavelets*. SIAM, Philadelphia, PA.
- Dayal, B.S. and J.F. MacGregor (1997a). Improved pls algorithms. *Journal of chemometrics* **11**, 73–85.
- Dayal, B.S. and J.F. MacGregor (1997b). Recursive exponentially weighted pls and its application to adaptive control and prediction. *Journal of Process Control* **7**, 169–179.
- de Jong, S. (1993). Simpls: an alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems* **18**, 251–261.

- Donoho, D.L., L.M. Johnstone, G. Kerkycharian and D. Picard (1995). Wavelet shrinkage: Asymptopia?. *J.R. Stat. Soc. B* **57**, 301.
- Dunia, R., S.J. Qin, T. F. Edgar and T. J. McAvoy (1996). Identification of faulty sensors using principal component analysis. *AIChE* **42**, 2797–2812.
- Eastment, H. and W. Krzanowski (1982). Crossvalidatory choice of the number of components from a principal component analysis. *Technometrics* **24**, 73–77.
- Eykhoff, P. (1974). *System Identification*. Wiley Sons, New York.
- Forina, M., C. Casolino and C. Pizarro Millan (1999). Iterative predictor weighting (ipw) pls: a technique for the elimination of useless predictors in regression problems. *Journal of Chemometrics* **13**, 165–184.
- Forina, M., C. Drava and C. De La Pezuela (1986). Automatic selection of predictors in pls by means of statistical test on the correlation coefficient of the marginal least-squares regressions. *VI CAC (Chemometrics in Analytical Chemistry Conf.)*.
- Frank, I. (1987). Intermediate least squares regression method. *Chemometrics and Intelligent Laboratory Systems* **1**, 233–242.
- Gallagher, N.B., B.M. Wise, S.W. Butler, D.D. White and G.G. Barna (1997). Development and benchmarking of multivariate statistical process control tools for a semiconductor etch process: improving robustness through model updating. In: *Proceedings of ADCHEM 97*. Banff, Canada. pp. 78–83.
- Geladi, P. and B. R. Kowalski (1986a). Partial least-squares regression: A tutorial. *Analytica Chimica Acta* **185**, 1–17.
- Geladi, P. and B. R. Kowalski (1986b). An example of 2-block predictive partial least squares regression with simulated data. *Analytica Chimica Acta* **185**, 19–32.
- Goodwin, G.C. and D.Q. Mayne (1987). A parameter estimation perspective of continuous time model reference adaptive control. *Automatica* **23**, 57–70.
- Grossmann, A. and J. Morlet (1984). Decomposition of hardy functions into square integrable wavelets of constant shape.. *SIAM J. Math. Anal.* **15**, 723–736.
- Harris, T.J. and W.H. Ross (1991). Statistical process control procedures for correlated observations. *The Canadian Journal of Chemical Engineering* **64**, 48–57.
- Helland, I.S. (1988). On the structure of partial least squares regression. *Chemometrics and Intelligent Laboratory Systems* **17**, 581–607.
- Helland, K., H.E. Berntsen, O.S. Borgen and H. Martens (1991). Recursive algorithm for partial least squares regression. **14**, 129–137.
- Hoskuldson, A. (1988). Pls regression method. *Journal of Chemometrics* **2**, 211–228.
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.*
- Hunter, J.S. (1986). The exponentially weighted moving average. *Journal of quality Technology* **18**(4), 203–210.

- Jackson, J. Edward (1991). *A User's Guide to Principal Components*. Wiley-Interscience, New York.
- Jackson, J.E. and G.S. Mudholkar (1979). Control procedures for residuals associated with principal component analysis. *Technometrics* 21(3), 341-349.
- Johnson, R.A. and D.W. Wichern (1992). *Applied Multivariate Statistical Analysis*. Prentice-Hall, N.J.
- Jolliffe, I.T. (1986). *Principal Components Analysis*. Springer-Verlag, New York.
- Kaspar, M.H. and W.H. Ray (1993). Dynamic pls modelling for process control. *Chemical Engineering Science* 48(20), 3447-3461.
- Koornwinder, T.H. (1993). *Wavelets: An Elementary Treatment of Theory and Applications*. World Scientific Publishing, Singapore.
- Kourti, T., P. Nomikos and J.F. MacGregor (1995). Analysis, monitoring and fault diagnosis of batch processes using multi-block, multi-way pls. *Journal of Process Control* 5, 277-284.
- Kresta, J.V., J.F. MacGregor and T.E. Marlin (1991). Multivariate statistical monitoring of process operating performance. *Canadian Journal of Chemical Engineering* 69, 35-47.
- Ku, Wenfu, Robert H. Storer and Christos Georgakis (1995). Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems* 30, 179-196.
- Lakshminarayanan, S. (1997). Process Characterization and Control using Multivariate Statistical Techniques. PhD thesis. University of Alberta. Canada.
- Leardi, R., R. Boggia and M. Terrile (1992). Genetic algorithms as a strategy for feature selection. *Journal of Chemometrics* 6(5), 267-282.
- Lindgren, F., P. Geladi and S. Wold (1993). The kernel algorithms for pls. *Journal of Chemometrics* 7, 45-49.
- Lindgren, F., P. Geladi, S. Rannar and S. Wold (1994). Interactive variable selection (ivs) for pls. part 1: Theory and algorithms. *Journal of Chemometrics* 8, 349.
- Ljung, L. and T. Soderstrom (1983). *Theory and Practice of Recursive Identification*. MIT Press, Cambridge, Massachusetts.
- Lucas, J.M. and M.S. Saccucci (1990). Exponentially weighted moving average control schemes: Properties and enhancements. *Technometrics* 32(1), 1-12.
- Luo, R.M., M. Misra and D.M. Himmelblau (1998). Sensor fault detection via multiscale analysis and dynamic pca. In: *AIChE Annual Meeting*. Miami Beach, FL.
- MacGregor, J. F. and T. Kourti (1995). Statistical process control of multivariate processes. *Control Engineering Practice* 3, 403-414.
- MacGregor, J. F., C. Jaeckle, C. Kiparissides and M. Koutoudi (1994). Process monitoring and diagnosis by multiblock pls methods. *AIChE* 40(5), 826-838.

- Malinowski, E.R. (1991). *Factor Analysis in Chemistry*. Wiley-Interscience, New York.
- Mallat, S.G. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell. PMAI-11* p. 674.
- Mallat, S.G. (1998). *A wavelet tour of signal processing*. Academic Press, CA.
- Manne, R. (1987). Analysis of two partial-least-squares algorithms for multivariate calibration. *Chemometrics and Intelligent Laboratory Systems* (2), 283–290.
- Martens, H. and T. Naes (1989). *Multivariate Calibration*. Wiley, Chichester.
- Miller, P., R.E. Swanson and C.F. Heckler (1993). Contribution plots: The missing link in multivariate quality control. *Presented at 37th Annual Fall Conference. ASQC, Rochester, NY*.
- Morlet, J., G. Arens, I. Fourgeau and D. Giard (1982). Wave propagation and sampling theory. *Geophysics*. 47, 203–236.
- Morud, T.E. (1996). Multivariate statistical process control; example from the chemical process industry. *Journal of Chemometrics* 10(5-6), 669–675.
- Mujunen, S., P. Minkkinen, B. Holmbom and A. Oikari (1996). Pca and pls methods applied to ecotoxicological data: Ecobalance project. *Journal of Chemometrics* 10(5-6), 411–424.
- Nomikos, P. and J.F. MacGregor (1995). Multi-way partial least squares in monitoring batch process. *Chemometrics and Intelligent Laboratory Systems* 30, 97–108.
- Page, E.S. (1954). Continuous inspection schemes. *Biometrika* 41, 100–115.
- Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Phil. Mag.* 2, 559.
- Qin, S.J. (1998). Recursive pls algorithm for adaptive data modeling. *Computers Chem. Engng.* 22(4/5), 503–514.
- Qin, S.J. and T. J. McAvoy (1992a). Nonlinear pls modelling using neural networks. *Computers Chem. Engng.* 16(4), 379–391.
- Qin, S.J. and T.J. McAvoy (1992b). A data-based process modelling approach and its application. In: *Proceedings of the IFAC Conference on Dynamics and Control of Chemical Reactors*.
- Qin, S.J. and W. Li (1999). Detection, identification and reconstruction of faulty sensors with maximized sensitivity. Technical report. University of Texas at Austin, Dept. of Chemical Engineering.
- Qin, S.J., W. Li and H. Henry Yue (1999). Recursive pca for adaptive process monitoring. In: *Proc. of 14th World Congress of IFAC*. Beijing, China. pp. 85–90.
- Rannar, S., J.F. MacGregor and S. Wold (1998). Adaptive batch monitoring using hierarchical pca. *Chemometrics and Intelligent Laboratory Systems* 41, 73–81.

- Ricker, N.L. (1988). The use of biased least-squares estimators for parameters in discrete time pulse response models. *Industrial and Engineering Chemistry Research* 27, 343–350.
- Rissanen, J. (1978). Modeling by shortest data description. *Automatica* 14, 465–471.
- Roberts, S.W. (1959). Control chart tests based on geometric moving averages. *Technometrics* 1, 239–250.
- Schoukens, J., R. Pintelon, G. Vandersteen and P. Guillaume (1997). Frequency-domain system identification using non-parametric noise models estimated from a small number of data sets. *Automatica* 33, 1073–1086.
- Seborg, D.E., T.F. Edgar and D.A. Mellichamp (1989). *Process Dynamics and Control*. John Wiley and Sons, New York.
- Shewhart, W.A. (1931). *Economic control of quality of manufactured product*. Van Nostrand, Princeton, N.J.
- Stone, M. (1978). Cross-validation: a review. *Math. Operations-forsch. Statist., Ser. Statist.*
- Stork, C.L. and B.R. Kowalski (1999). Distinguishing between process upsets and sensor malfunctions using sensor redundancy. *Chemometrics And Intelligent Laboratory Systems* 46, 117–131.
- Tong, H. and C. M. Crowe (1995). Detection of gross errors in data reconciliation by principal component analysis. *AIChE* 41, 1712–1722.
- Tracy, N.D., J.C. Young and R.L. Mason (1992). Multivariate control charts for individual observations. *Journal of Quality Technology* 24, 88–95.
- Wax, M. and T. Kailath (1985). Detection of signals by information theoretic criteria. *IEEE Trans. Acoust. Speech, Signal Processing ASSP-33*, 387–392.
- Wise, B.M. (1991). Adaptive multivariate analysis for monitoring and modeling dynamic systems. PhD thesis. University of Washington. Seattle.
- 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.
- 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.
- Wold, H. (1966). *Multivariate Analysis*. Academic, New York.
- Wold, S. (1978). Cross-validatory estimation of the number of components in factor and principal components models. *Technometrics* 20, 397–405.
- Wold, S. (1994). Exponentially weighted moving principal components analysis and projection to latent structures. *Chemometrics and Intelligent Laboratory Systems* 23, 149–161.
- Wold, S. and M. Sjostrom (1998). Chemometrics, present and future success. *Chemometrics and Intelligent Laboratory Systems* 44, 3–14.

- Wold, S., K. Esbensen and P. Geladi (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems* 2, 37–52.
- Wold, S., N. Kettaneh-Wold and B. Skagerberg (1989). Nonlinear pls modelling. *Chemometrics and Intelligent Laboratory Systems* 7, 53–65.
- Woodward, R.H. and P.L. Goldsmith (1964). *Cumulative Sum Techniques*. Published for Imperial Chemical Industries, Oliver Boyd, London, England.
- Xu, G. and T. Kailath (1994). Fast estimation of principal eigenspace using lanczos algorithm. *SIAM J. Matrix Anal. Appl.* 15, 974–994.
- Zullo, L. (1996). Validation and verification of continuous plants operating modes using multivariate statistical methods. *Computers Chem. Engng* 20, S683–S688.

Appendix A

Color Figures in Thesis

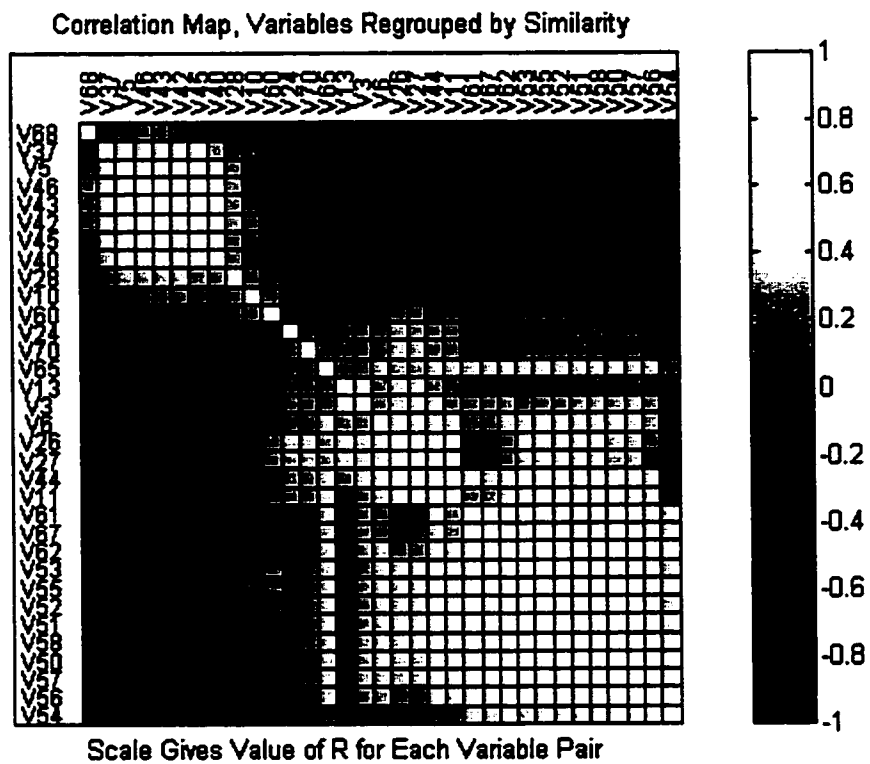


Figure 5.8 The cluster analysis display the correlated variables.

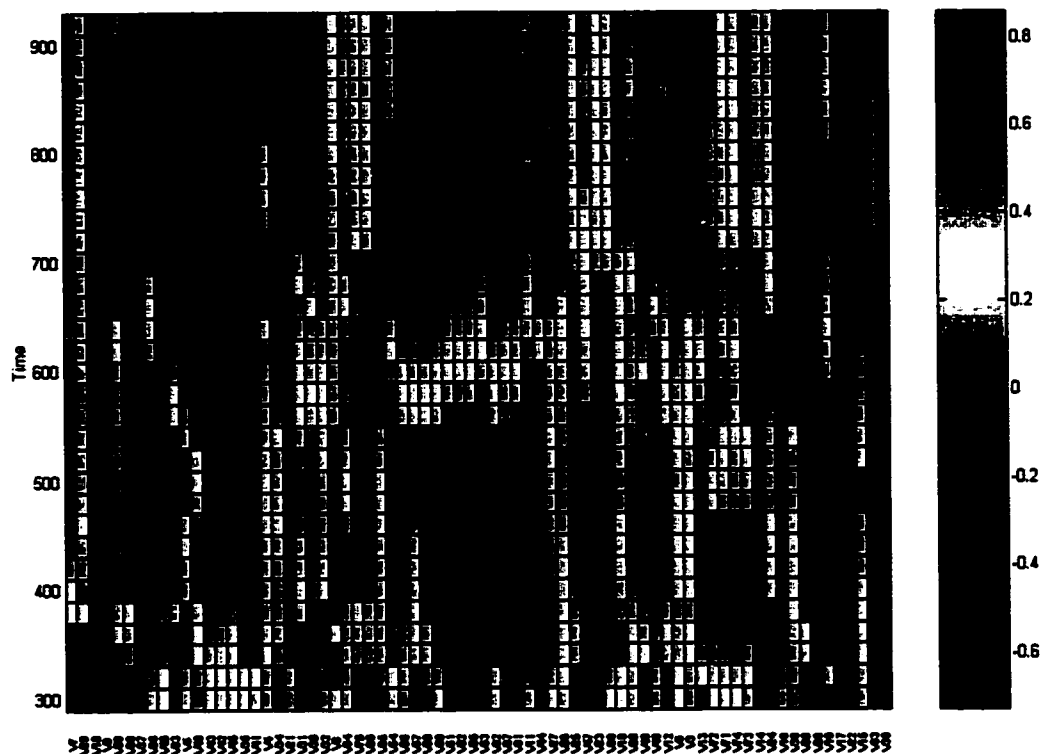


Figure 5.18 First weighting vector in recursive PLS model using all the available data.

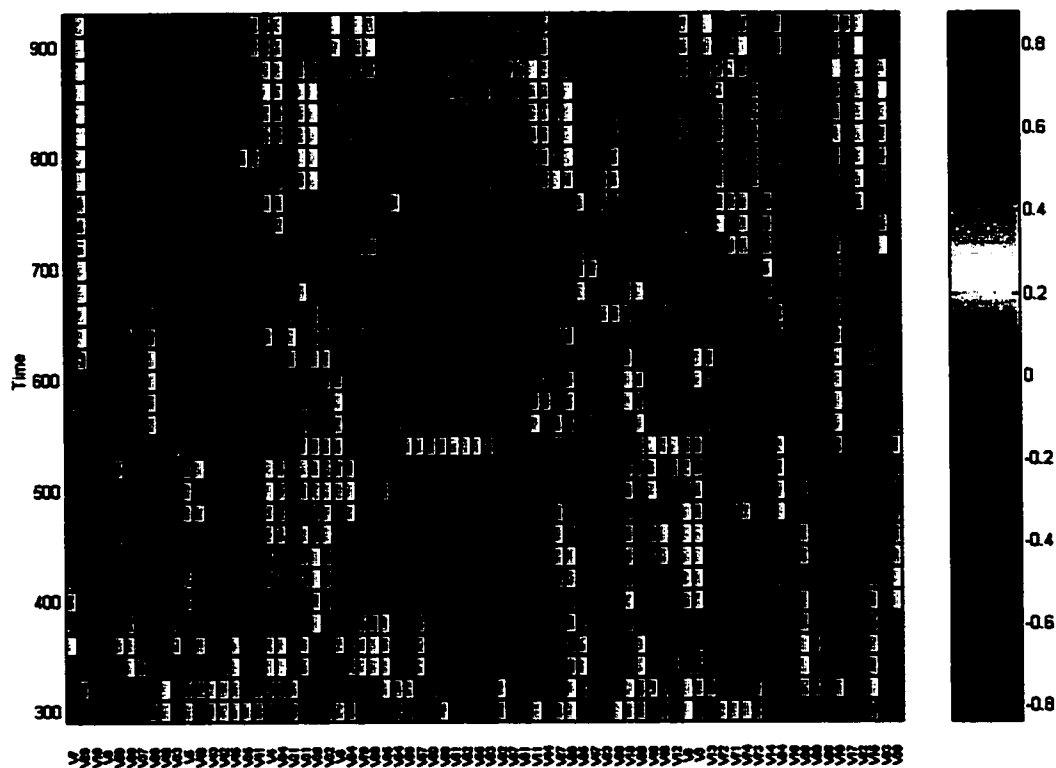


Figure 5.19 First weighting vector in recursive PLS model with EWMA updating method. ($\lambda=0.9$)

