MULTISCALE PROCESS MONITORING WITH SINGULAR SPECTRUM ANALYSIS

By

Syamala Krishnannair

Thesis presented in partial fulfilment of the requirements for the degree

of

Master of Science in Engineering

(Extractive Metallurgy)

In the Department of Process Engineering

at the University of Stellenbosch

Supervised by

Prof C Aldrich

December 2010

DECLARATION

I, the undersigned, hereby declare that the work contained in this thesis is my own original work and that I have not previously in its entirety or in part submitted it at any university for a degree.

Signature: ……..

Date: ……..

SUMMARY

Multivariate statistical process control (MSPC) approaches are now widely used for performance monitoring, fault detection and diagnosis in chemical processes. Conventional MSPC approaches are based on latent variable projection methods such as principal component analysis and partial least squares. These methods are suitable for handling linearly correlated data sets, with minimal autocorrelation in the variables. Industrial plant data invariably violate these conditions, and several extensions to conventional MSPC methodologies have been proposed to account for these limitations.

In practical situations process data usually contain contributions at multiple scales because of different events occurring at different localizations in time and frequency. To account for such multiscale nature, monitoring techniques that decompose observed data at different scales are necessary. Hence the use of standard MSPC methodologies may lead to unreliable results due to false alarms and significant loss of information.

In this thesis a multiscale methodology based on the use of singular spectrum analysis is proposed. Singular spectrum analysis (SSA) is a linear method that extracts information from the short and noisy time series by decomposing the data into deterministic and stochastic components without prior knowledge of the dynamics affecting the time series. These components can be classified as independent additive time series of slowly varying trend, periodic series and aperiodic noise. SSA does this decomposition by projecting the original time series onto a data-adaptive vector basis obtained from the series itself based on principal component analysis (PCA).

The proposed method in this study treats each process variable as time series and the autocorrelation between the variables are explicitly accounted for. The data-adaptive nature of SSA makes the proposed method more flexible than other spectral techniques using fixed basis functions. Application of the proposed technique is demonstrated using simulated, industrial data and the Tennessee

Eastman Challenge process. Also, a comparative analysis is given using the simulated and Tennessee Eastman process. It is found that in most cases the proposed method is superior in detecting process changes and faults of different magnitude accurately compared to classical statistical process control (SPC) based on latent variable methods as well as the wavelet-based multiscale SPC.

OPSOMMING

Meerveranderlike statistiese prosesbeheerbenaderings (MSPB) word tans wydverspreid benut vir werkverrigtingkontrolering, foutopsporing en –diagnose in chemiese prosesse. Gebruiklike MSPB word op latente veranderlike projeksiemetodes soos hoofkomponentontleding en parsiële kleinste-kwadrate gebaseer. Hierdie metodes is geskik om lineêr gekorreleerde datastelle, met minimale outokorrelasie, te hanteer. Nywerheidsaanlegdata oortree altyd hierdie voorwaardes, en verskeie MSPB is voorgestel om verantwoording te doen vir hierdie beperkings.

Prosesdata afkomstig van praktiese toestande bevat gewoonlik bydraes by veelvuldige skale, as gevolg van verskillende gebeurtenisse wat by verskillende lokaliserings in tyd en frekwensie voorkom. Kontroleringsmetodes wat waargenome data ontbind by verskillende skale is nodig om verantwoording te doen vir sodanige multiskaalgedrag. Derhalwe kan die gebruik van standaard-MSPB weens vals alarms en beduidende verlies van inligting tot onbetroubare resultate lei.

In hierdie tesis word ’n multiskaalmetodologie gebaseer op die gebruik van singuliere spektrumontleding (SSO) voorgestel. SSO is ’n lineêre metode wat inligting uit die kort en ruiserige tydreeks ontrek deur die data in deterministiese en stochastiese komponente te ontbind, sonder enige voorkennis van die dinamika wat die tydreeks affekteer. Hierdie komponente kan as onafhanklike, additiewe tydreekse geklassifiseer word: stadigveranderende tendense, periodiese reekse en aperiodiese geruis. SSO vermag hierdie ontbinding deur die oorspronklike tydreeks na ’n data-aanpassende vektorbasis te projekteer, waar hierdie vektorbasis verkry is vanaf die tydreeks self, gebaseer op hoofkomponentontleding.

Die voorgestelde metode in hierdie studie hanteer elke prosesveranderlike as ’n tydreeks, en die outokorrelasie tussen veranderlikes word eksplisiet in berekening gebring. Aangesien die SSO metode aanpas tot data, is die

voorgestelde metode meer buigsaam as ander spektraalmetodes wat gebruik maak van vaste basisfunksies. Toepassing van die voorgestelde tegniek word getoon met gesimuleerde prosesdata en die Tennessee Eastman-proses. ’n Vergelykende ontleding word ook gedoen met die gesimuleerde prosesdata en die Tennessee Eastman-proses. In die meeste gevalle is dit gevind dat die voorgestelde metode beter vaar om prosesveranderings en –foute met verskillende groottes op te spoor, in vergeleke met klassieke statistiese prosesbeheer (SP) gebaseer op latente veranderlikes, asook golfie-gebaseerde multiskaal SP.

ACKNOWLEDGEMENTS

I hereby express my sincere gratitude towards my promoter, Prof.Chris Aldrich for his unlimited support, guidance, patience and motivation for the successful completion of this study. I am also sincerely thankful to Dr.Gorden Jemwa for his continuous help in the fulfillment of this thesis and for his encouragement throughout the study.

I would like to thank the administrative staff in the Department of Process Engineering for their prompt help in all the administrative work related to the completion of this study. I am very grateful to Anglo Platinum for giving me the opportunity to work with their team in Process Optimization Department and also for using the results in my study. I would also like to thank NRF and University of Zululand's Research Committee for their financial support for this study. I am also very much grateful to my colleagues and friends in the University of Zululand for their support.

Finally, I would like to acknowledge my husband's and my kid's encouragement, patience and moral support which inspired me tremendously.

NOMENCLATURE

SPC	Statistical Process Control
CUSUM	Cumulative sum
EWMA	Exponentially Weighted Moving Average
PCA	Principal Component Analysis
PLS	Partial Least Squares
MSPCA	Multiscale Principal Component Analysis
SSA	Singular Spectrum Analysis
MSSSA	Multiscale Singular Spectrum Analysis
SVD	Singular Value Decomposition
SPM	Statistical Process monitoring
MSPC	Multivariate Statistical Process Control
ARL	Average Run Length
EOFs	Empirical Orthogonal Functions
RCs	Reconstructed Components
TCM	Tool Condition Monitoring
cPCA	Conventional Principal Component Analysis
$\square^{n\square m}$	Set of n-by –m real matrix
Λ	Diagonal matrix of Eigenvaues
λ_i	i^{th} Eigenvalue
$\|\cdot\|$	Euclidean Norm
E	Residual Matrix

e_i	i[th] row of the residual matrix
X	Matrix of process data
T	Matrix of score vectors
P	Matrix of loading vectors for X
ψ	Wavelet function
$\Re^M$	M -dimensional Euclidean Space
$\mathbf{X}$	Trajectory Matrix
$\mathbf{C}_\mathbf{X}$	Covariance Matrix
I	Identity Matrix
$N\ \mu,\sigma$	Gaussian distribution with mean μ and Standard deviation σ
$x\ t$	The observed value of the time series at time t
a_k	k[th] eigenvector
a_m, d_m	Discrete wavelet transforms parameters
ξ_t	Reconstructed component of the time series

TABLE OF CONTENTS

Chapter 1 Introduction

The last few decades have seen an increased emphasis on process monitoring and control in chemical and metallurgical industries as a result of, among other, a challenging economic environment, environmental and safety considerations, and dwindling natural resources. The detection and diagnosis of disturbances and faults that may negatively affect process behaviour and/or product quality has become critical in achieving operational excellence, which traditionally was narrowly defined in terms of profitability, cash flow and revenue. Hence, there has been increased focus on the development and application of advanced process control systems for monitoring, control and diagnosis of process operations. The development of these advanced control technologies is a great challenge, particularly for large scale systems such as those encountered in the chemical and metallurgical industries. Monitoring of these highly complex and large integrated systems, where information can be overloaded on thousands of variables sampled at high frequency rates, is inherently a difficult task (Bailey, 1984).

Statistical studies on industrial accidents have shown that more than 60% of accidents are a direct result of human errors (Venkatasubramanian, 2005). In the absence of succinct and reliable process condition indicators, the risk of incorrect process diagnosis is increased which may result in decisions that only worsen abnormal situations. A few recent major industrial accidents illustrate the risks associated with poor abnormal situation management: Union Carbide's accident in Bhopal, India in December 1984; Occidental Petrolium's Piper Alpha accident in July 1988; the explosion at Kuwait Petrochemical's Mina Al-Ahmedhi refinery in June of 2000 with an estimated loss of $400 million; explosion at the offshore oil platform of Petrobas, Brazil which caused it to sink into the sea in March 2001 at a loss of $5 billion; and the explosion at the AZF chemical plant which killed dozens of people in September 2001 (Lees, 1996; Venkatasubramanian, 2005).

Minor accidents occurring on a daily basis in industries also result in significant losses to businesses and society due to increased occupational injuries, illnesses and compensations thereof, which run into billions of dollars every year (Bureau of Labor Statistics, 1998; National Safety Council, 1999). For example, it was once estimated in the 1990s that U.S. petrochemical industries were incurring losses of US$20 billion every year (Nimmo, 1995). Similar instances of heavy losses include Nucor Corporation Inc. losing $100 million in a pollution control lawsuit and British economy losing $27 billion annually (Laser, 2000). Clearly, improved and reliable timeous detection of abnormal events or faults can help process and manufacturing industries meet their economic goals and social obligations as this ensures high quality production, reduction of product rejection rate and meeting stricter safety and environmental regulations.

Early detection of faults in physical systems requires a proper understanding of process behavior coupled with the judicious use of process monitoring and control techniques. Such an understanding is typically expressed as a mathematical process model which captures dynamic and stochastic aspects related to the evolution of the process. As part of process control requirements, such a model ideally contains information that enables the detection of faulty or abnormal conditions. Given a representative model of a process, potential faults can be detected by monitoring deviations of the actual process behavior from that predicted by the model. In the ideal case, such process models are based on fundamental principles governing process evolution. Unfortunately, these first-principles models are usually inadequate due to limited fundamental knowledge or difficult to obtain in many cases. An alternative is to derive control models on the basis of observed process data (Kano and Nakagawa, 2008), whose availability and volume has increased exponentially in modern times as a direct result of improvements in plant automation and instrumentation as well as data storage capacity. Exploiting process data can potentially yield critical plant status information with high frequency. Additionally, the use of process data provides for simple diagnosis of the source of an abnormal event.

Among data-driven process control technologies proposed in the last few years, statistical process monitoring (SPM) techniques have been widely accepted in industry because of their effectiveness and simplicity. SPM is based on the use of statistical methods to detect the existence and time of occurrence of changes that cause deviations in process performance (Negiz and Cinar, 1997). The basic framework for statistical process control was originally developed for industrial engineering applications such as tool making, where observed data are invariably stochastic. An extension of this framework to dynamic systems where the data are multivariate and highly correlated was first proposed by Kresta et al. (1991). Other modifications and extensions of this framework have since been proposed and the collective of these multivariate statistical techniques is commonly referred to as **multivariate statistical process control** (MSPC) (Kruger et al., 2007; Zhang and Dudzic, 2006). MSPC techniques exploit the high degree of redundancy in multivariate data to generate a reduced set of statistically uncorrelated variables that are subsequently used in deriving monitoring tools.

Although MSPC techniques and its several extensions are able to decorrelate process variables, these techniques are often not well-suited for dynamic process systems as encountered in metallurgical plants. Process data in these systems are invariably autocorrelated. However, classical MSPC techniques are unreliable when measurements are autocorrelated (Ku et al., 1995). Another practical limitation of MSPC methods is that the static models used rely on the assumption that the process operates at a steady state condition. As a result these techniques do not capture information about events that occur with different localization in time, space and frequency that is multiscale characteristics in data (Aradhye et al., 2003). Exploiting the *scale* properties in data allows for signal decomposition and therefore different representations of data. For example a detailed view of the signal is obtained by representing the signal in low scale (high frequency) components, whereas a non-detailed view is obtained by representing the signal in high scale (low frequency) components (Polikar, 1996). Advantages of such signal decompositions include detecting

oscillatory behaviour, noise separation and trend analysis which are useful in practical problems such as data rectification and gross error detection. In this thesis a multiscale process monitoring approach using singular spectrum analysis is proposed to address these limitations of classical MSPC techniques.

1.1 Process Monitoring and Diagnosis

The diagnosis of process operations broadly involves four hierarchical tasks, namely fault detection, fault identification, fault diagnosis and process recovery (Chiang et al., 2001). In **fault detection** the goal is to determine when a process or plant being monitored is out-of-control. Early detection of a fault condition is important in avoiding below quality product batches or system breakdown, and this can be achieved through proper design of an effective fault detection method. Once a fault condition has been positively detected, the next step is isolating the variables responsible for the out-of-control situation, a task referred to as **fault identification** (or fault isolation). Subsequent troubleshooting efforts are then mainly focused on relevant subsystems to **diagnose** or determine the source of the out-of-control status. Characteristics of the fault, including type of fault, location, magnitude as well as time of occurrence are determined. Finally, the system is corrected by elimination of the fault or its cause via a p**rocess recovery** phase to complete the process monitoring procedure. These tasks and their relationships are as outlined in Figure 1.

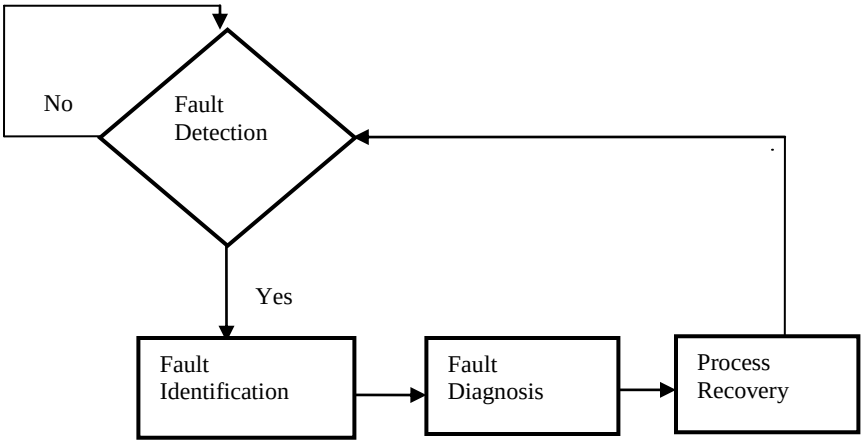

Figure 1: Process monitoring loop schematic.

As alluded to previously, successful implementation of the above procedure depends on the development of an appropriate process model that adequately describes the dynamic process. Such a model can be designed based on three different approaches, namely data-driven, analytical, and knowledge-based techniques. Analytical approach generally involves detailed mathematical models developed from first-principles to generate features such as residuals, parameter estimation and state estimation. Fault detection and diagnosis is performed by comparing these values with those associated with normal operating conditions either directly or after some transformation. The parameter estimation method and the observer-based methods are the two common methods used in the analytical approach for monitoring the process. Knowledge-based approaches are based on qualitative models which involve uncertain, conflicting, and non-quantifiable information. Expert systems, fuzzy logic, machine learning and pattern recognition are the most common knowledge based approaches used for monitoring, control and diagnosis in the process industries (Chiang et al., 2001; Uraikul et al., 2007).

For large scale systems it is often difficult to use analytical approaches because of the lack of accurate models. It is equally challenging to apply knowledge-based approach to such large scale systems because construction of a model requires large amount of effort and skill that a typical operator may not have. When large volumes of process data are available as in a modern state-of-the-art plant, data-based technologies provide an alternative approach to process monitoring that partially circumvents difficulties associated with analytical or knowledge-based methods. This is particularly appealing route as modern industrial processes are characterized by high instrumentation and process automation and, thus it is not uncommon to have large amounts of data collected every few seconds on such plants. In principle, data-based approaches exploit structure or regularities in data to derive mathematical or statistical models that describe expected process behaviour under normal operating conditions. The

derived models can then be used for monitoring, control and process optimization tasks.

Data-driven process monitoring statistics based on multivariate methods and their applications in fault detection in industrial processes are briefly introduced in the next section.

1.2 Multivariate Statistical Process Monitoring

Reliability of a data-driven method depends on the nature of process variations such as common cause and special cause variations (MacGregor and Kourti, 1995; Ogunnaike and Ray, 1994). Common cause variations arise from random noise, and hence inherent in process data, while all other variations that are not due to common cause variations are special cause variations. Common cause variations or disturbances are difficult to eliminate using standard process control strategies. Since the variations in the process data are unpredictable, Statistical Process Control (SPC) plays a major role in process monitoring schemes. Traditional process control charts such as Shewart charts, cumulative sum (CUSUM) charts and exponentially weighted moving average (EWMA) charts have proved very effective for univariate stochastic processes. Unfortunately, they are inadequate multivariate (dynamic) processes which exhibit high correlations among measured variables. Moreover, it is difficult to detect the important events that occur in these processes with univariate charts because of the low signal to noise ratio typically associated with each variable. For these reasons multivariate statistical methods have been proposed to handle dynamic process. As pointed out earlier, multivariate statistical summarizes relevant information in a low-dimensional space. This also has an effect of reducing noise levels through averaging (MacGregor and Kourti, 1995).

The need for multivariate statistical process control over univariate control can be motivated by considering the fault detection problem illustrated in Figure 2.

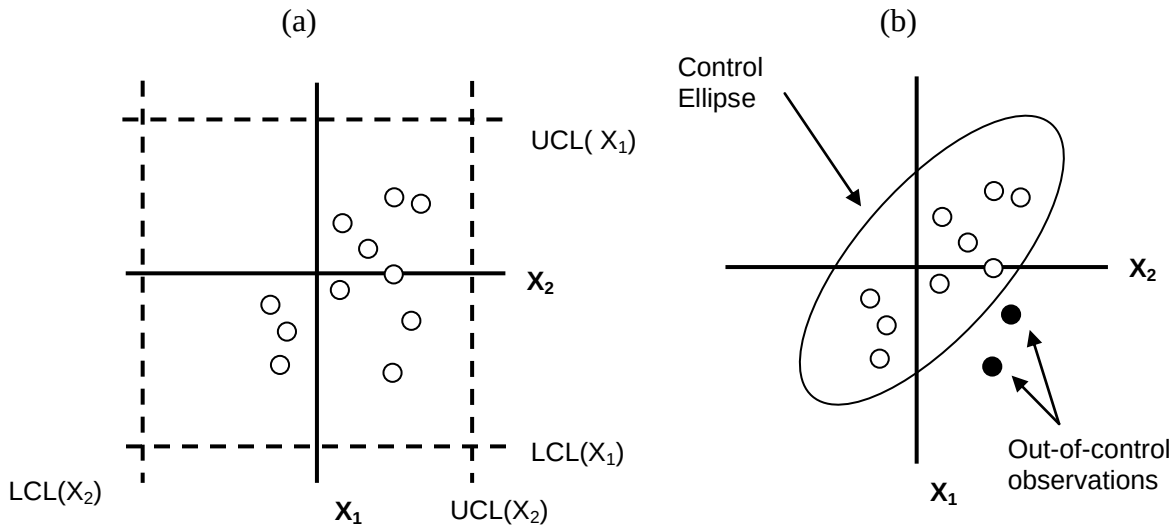

Figure 2: (a) Scatter plot of the multivariate data and (b) superimposition of control ellipse over the scatter plot of multivariate data.

In Figure 2(a) is a scatter plot summarizing the behavior of a two-variable system. Also shown in the same plot are the control limits when each variable is considered separately (as in traditional statistical quality control). These limits define an in-control region which, in this example, is shown delineated by a rectangle (see Mastrangelo et al., 1996; Tracy et al., 1992). Such a superimposition of univariate control charts defined for each variable does not exploit correlation structure between the variables. When correlation information is taken into account, the in-control region is defined by an elliptical region as shown in Figure 2(b). As can be seen from this plot, the two points in the lower right corner (solid circles), although falling within the control rectangle of Figure 2(a), are outside the control ellipse and, hence indicating a fault condition that would be missed with traditional statistical control strategies. The control ellipse exploits the correlation between the variables which results in the tilted shape of the in-control region.

In multivariate situations, the probability that a process is completely under normal operating control region is less than that in the univariate case (Montgomery, 1996). Similarly the probability that a multivariate process is completely out-of-control is less than that of a univariate case. Using multivariate

control charts the desired confidence level can be maintained by taking advantage of the cross correlation information between variables. Hence, the process can be analyzed for its stability without the added complication of maintaining many control charts at the same time.

Classical multivariate statistical process control methods, for example latent variable methods such as principal component analysis (PCA) and partial least squares (PLS), have been used in process monitoring problems. These are based on transforming a set of highly correlated variables to a set of uncorrelated variables (Kresta et al., 1991; MacGregor and Kourti, 1995). The use of PCA assumes data are approximately normally distributed and time independent (Jollife, 1986). As noted above, chemical processes are dynamic in nature, and exhibit highly auto-correlated process variables. Moreover, correlations between variables tend to be nonlinear. These characteristics can lead to an excess of false alarms or a significant loss of information when using linear PCA for process monitoring.

To address these limitations, several modifications to basic PCA have been proposed. Nonlinear principal component analysis (NLPCA) is used to capture nonlinear relationships among variables. Compared to linear PCA, NLPCA can explain more variance in smaller dimensions (Dong and McAvoy, 1996; Kramer, 1991; Tan and Mavrovouniotis, 1995). Similarly, dynamic PCA has been proposed to eliminate the effect of autocorrelation in process data by augmenting the data matrix with time-lagged variables (Ku et al., 1995; Luo et al., 1999; Lin et al., 2000). Adaptive PCA updates the model parameters continuously by exponential smoothing so as to get the model adjusted to suit new operating conditions (Wold, 1994). Multiway and multiblock PCA are suitable for batch process operations (Nomikosi and MacGregor, 1994; MacGregor et al., 1994; Wold et al., 1996). Moreover, multiblock PCA allows for efficient computation of very large datasets.

Conventional multivariate process monitoring methods detect fault conditions at a single scale since they represent the data in terms of basis functions at a fixed

resolution or scale in time and frequency. Data containing contributions with the same localization everywhere in the time frequency domain can be efficiently represented by these single scale methods. In practical situations process data usually contain contributions at multiple scales because of different events occurring at different localizations in time and frequency. Hence, a measured process signal reflects an aggregate of these different events, including underlying process dynamics, as depicted in the example in Figure 3(a). Here, the measured signal is constituted of different possible disturbances that can occur on a system. These events are associated with time-frequency localizations as shown in Figure 3(b). For example, a sudden change in the data such as sensor noise extends over a wide range in the frequency domain but a narrow range in the time domain. In contrast, a slow change such as equipment degradation extends over a wide range in the time domain and a narrow range in the frequency domain (Bakshi, 1999). To account for such multiscale nature, monitoring techniques that decompose observed data at different scales prior to analysis are necessary. To this end, multiscale approaches designed to handle and take advantage of the information contained at multiple scales have been developed for addressing process monitoring tasks.

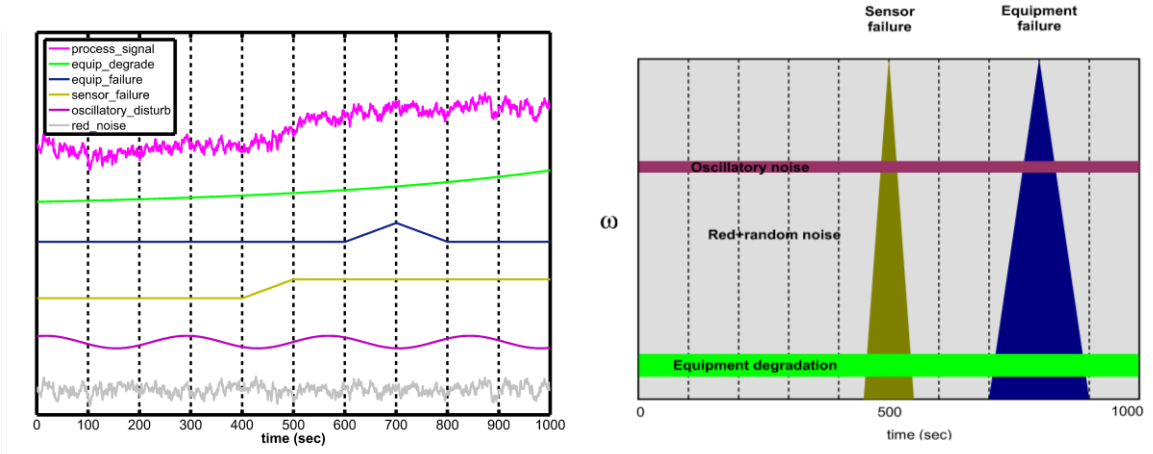

Figure 3(a) Illustration of a typical process signal and (b) its time-frequency representation (Bakshi, 1999)

An early development of a multiscale framework for statistical process monitoring can be attributed to Bakshi (1998) who proposed use of wavelets to decompose

data into several views or scales prior to the application of PCA. This has a two-fold effect, namely decorrelation across variables and elimination or reduction of autocorrelation individual variables. Wavelets are appropriate in this regard due to their time-frequency localization property. Several combinations of PCA with wavelets have been developed to monitor the process because of the ability of wavelets to compress multiscale features of the signal and approximately remove serial or auto correlations in time signals (Bakshi, 1998; Misra et al., 2002; Maulud et al., 2006; Rosen and Lennox, 2001). Multiscale Principal Component Analysis (MSPCA) approach adapts to the nature of the signal features and this approach has been extended to a nonlinear MSPCA by using neural networks to extract the latent nonlinear structure from the PCA transformed data (Fourie and Devaal, 2000; Shao et al, 1999; Zhinqiang and Qunxiong, 2005).

Efficient extraction of deterministic and stochastic features at various scales using wavelets depends on a number of factors, the most significant being the choice of basis function or mother wavelet for the optimal orthogonal expansion of the signal for the application in hand. The mother wavelet should have some desirable properties such as good time frequency localizations and general admissibility properties including various degrees of smoothness (number of continuous derivatives) and large number of vanishing moments (ensures maximum number of zeros of the polynomial at the highest discrete frequency) (Daubechies,1992; Ganesan et al., 2004; Meyer,1992).

A large number of wavelet bases that meet these requirements, such as completeness, time-frequency localization, and orthogonality (or limited redundancy for non-orthogonal basis functions) have been proposed in literature. Given the huge library of wavelets that exists, choosing an appropriate wavelet basis function for a specific purpose remains a difficult task for practitioners. In addition, while the optimal multiscale decomposition of the signal can be obtained by an automatic time-varying adjustment of the mother wavelet's shape, it does not allow for adjusting the nature of the analyzing function that is adaptable to the signal. Hence, other alternatives such as *Singular Spectrum Analysis* that are not susceptible to these limitations have been proposed.

Singular spectrum analysis (SSA) is a non-parametric data analysis method which requires no assumptions to be made about the data and can be applied to small samples. SSA is based on the singular value decomposition (SVD) of a trajectory or lagged covariance matrix obtained from a time series (Golyandina et al., 2001). SSA is data adaptive and only uses information obtained from the spectral decomposition of the data, thereby overcoming most of the limitations in using the short and noisy time series. SSA can decompose a time series into deterministic and stochastic components and, hence, can be used where data compression and signal-to-noise ratio enhancement are required (Jemwa and Aldrich, 2006).

Applications of SSA were initially limited to the field of climatology (Broomhead and King, 1986; Vautard and Ghil, 1989; Vautard et al., 1992) but have since been applied to various other fields such as biosciences (Mineva and Popivanov, 1996), geology (Rozynski et.al., 2001), economics (Ormerod and Campbell, 1997), solar physics (Kepenne, 1995), and recently process systems engineering (Barkhuizen, 2003; Barkhuizen and Aldrich, 2003 ; Botha, 2006; Jemwa and Aldrich, 2006).

1.3 Problem Statement

Although multivariate statistical process monitoring have proven very effective diagnostic tools, there are still a few challenges that are yet to be adequately addressed, particularly for dynamic, nonlinear systems as encountered in mineral processing and metallurgical systems. Classical MSPM methods based on linear latent variable projection techniques are suitable for handling linearly correlated data. Industrial chemical and metallurgical processes are generally dynamic and multiscale in nature and the use of standard MSPM tools may lead to unreliable results due to false alarms and significant loss of information. Progress has been achieved with the introduction of multiscale monitoring methods using wavelets. However, wavelets require *a priori* specification of the basis function. In order to provide for an optimal multiscale decomposition of data, it is desirable to adjust

the shape of the analyzing wavelet to the signal, instead of searching through the extensive “libraries” of the mother wavelets (analyzing functions). Also, it is desirable to modify the shape of the analyzing wavelet in time and scale especially if the data set is not stationary. The inherent constraint of a unique mother wavelet does not allow for this flexibility. Hence, improved data adaptive process monitoring techniques are required that do not suffer from limitations of existing state-of-the-art wavelet-based methods.

In this study a multiscale process monitoring technique based on SSA and PCA is proposed. This approach explicitly accounts for autocorrelation in process variables. Singular spectrum analysis can be regarded as equivalent to the use of a data adaptive wavelet transform (Yiou et al., 2000). However, unlike wavelet analysis which use fixed basis functions, SSA uses data-adaptive basis functions and, therefore, can be expected to provide more flexibility than other spectral techniques. In addition, SSA provides a qualitative decomposition of a signal into deterministic and stochastic parts that can be useful in other application such as data rectification and gross error detection.

1.4 Objectives of the Study

The main objective of this study is the development of a multiscale process monitoring method using singular spectrum analysis (MS-SSA) for the early and reliable detection of the anomalies or undesirable deviations in process systems. An analysis of the properties of the framework is presented. Subsequently, the proposed technique is demonstrated using simulated and industrial data. Also, a comparative analysis is given using the Tennessee Eastman Challenge problem as a benchmark.

As part of the study a review of the related literature on the recent multivariate statistical process monitoring methods and its applications are also done in detail.

1.5 Thesis Outline

The thesis is organized as follows: In Chapter 2 the advantage of multiscale multivariate methods in process monitoring over the single scale multivariate monitoring methods are reviewed in detail. Recent applications of multiscale monitoring methods and their limitations are also reviewed. Chapter 3 discusses Singular Spectrum Analysis and its recent applications in various fields such as geophysics, climatology and life sciences are discussed. An alternative multiscale monitoring strategy based on singular spectrum analysis (SSA) is proposed. A general strategy of a multiscale process monitoring algorithm is presented with the development of multiscale process monitoring method with SSA. Its features are investigated using a simple one-dimensional system.

In chapter 4 MS-SSA methodology is demonstrated and assessed by means of four case studies: two simulated systems as well as data from an industrial plant as well as the Tennessee Eastman Challenge. Chapter 5 concludes the thesis and opportunities for future research in this regard are listed.

Chapter 2 Multivariate Statistical Process Control: A Literature Review

Performance monitoring and early detection of abnormal events is critical in achieving set product quality objectives as well as general continuous process improvement. Examples of such abnormal events include among other, drifts and shifts in the mean or the variance of one or more process variables. To this end, a range of statistical process monitoring techniques has been proposed as a means for achieving stated plant objectives. These included classical charting techniques such as Shewhart, cumulative sum (CUSUM), and exponentially weighted moving average (EWMA) control charts used in monitoring the performance of processes to detect changes in process performance. However, these charts are not suitable for multivariate processes where observed variables tend to be significantly correlated. To effectively handle these cases, multivariate extensions of these univariate methods have been developed. These are based on the projection of measured variables onto latent structures. More specifically, methods based on the use of principal component analysis (PCA), partial least squares (PLS) and related variants have gained a lot of attention over the last couple of decades in the monitoring of multivariable processes (Ku et al., 1995; Kresta et al., 1991; MacGregor et al., 1994). These groups of fault detection and diagnosis tools are generally referred to as multivariate statistical process control (MSPC) methods.

Despite their wide acceptance and success, MSPC methods are often not adequate for processing multivariate data with multiscale or autocorrelated measurements (Aradhye et al., 2003). As discussed before, a typical process signal is an aggregate of events at different localizations in time, space, and frequency from a variety of sources (see Figures 3(a) and (b) in Chapter 1). Conventional multivariate methods and their extensions are single scale (that is, same time-frequency localizations at all locations) in nature and relate variables at the scale of the sampling interval (Bakshi, 1998). Because of fixed time and

frequency resolution, single scale monitoring methods may not be effective in detecting shifts related to such data (Bakshi, 1999). More generally, single scale methods are very sensitive to sudden oscillations but they are not efficient in extracting hidden patterns and frequency-related information. The use of spectral analysis methods such as Fourier transforms, power spectral density and coherence functional analysis can overcome some of these limitations of conventional multivariate methods. Bakshi (1999) showed that wavelet analysis provides a convenient basis to develop a multiscale process monitoring framework because of the time frequency localization and multiresolution properties of the wavelet transform.

In this chapter conventional univariate SPC methods and PCA-based monitoring and diagnostics are reviewed. Multiscale process monitoring strategy using wavelets and PCA, which is aimed at overcoming limitations associated with the classical MSPC methods in process monitoring, is then discussed. Recent applications of multiscale monitoring methods and corresponding limitations are also reviewed.

2.1 Classical Statistical Process Control

Classic univariate control charts analyze data at a fixed scale or resolution, which makes them detect changes at that single scale. More formally, the linear transformation of data in these charts has been done at fixed frequencies and extract features in the domain of time as illustrated in Figure 4 (Hunter, 1986). Shewhart charts represent data at the sampling interval or at the finest scale which is effective for detecting large mean shifts. Shewhart charts use only information about the process contained in the last observed point and ignore any information given by the entire sequence of points. This limitation of Shewhart charts can be overcome by the use of CUSUM, moving average (MA) and EWMA charts. On the one hand CUSUM charts represent data at the scale of all measurements or at the coarsest scale and directly incorporate all of the information in the sequence of sample values by plotting the cumulative sums of

the deviations of the sample values from a target value. MA and EWMA charts, which fall in-between these two extremes viz. Shewhart and CUSUM, are very effective in detecting small mean shifts. The MA chart monitors the process location over time based on the average of the current subgroups and one or more prior subgroups and hence it gives equal importance to past data within its moving window. On the other hand, in EWMA the average of the samples is computed in a way that gives less and less weight to data as they are further removed in time from the current measurement.

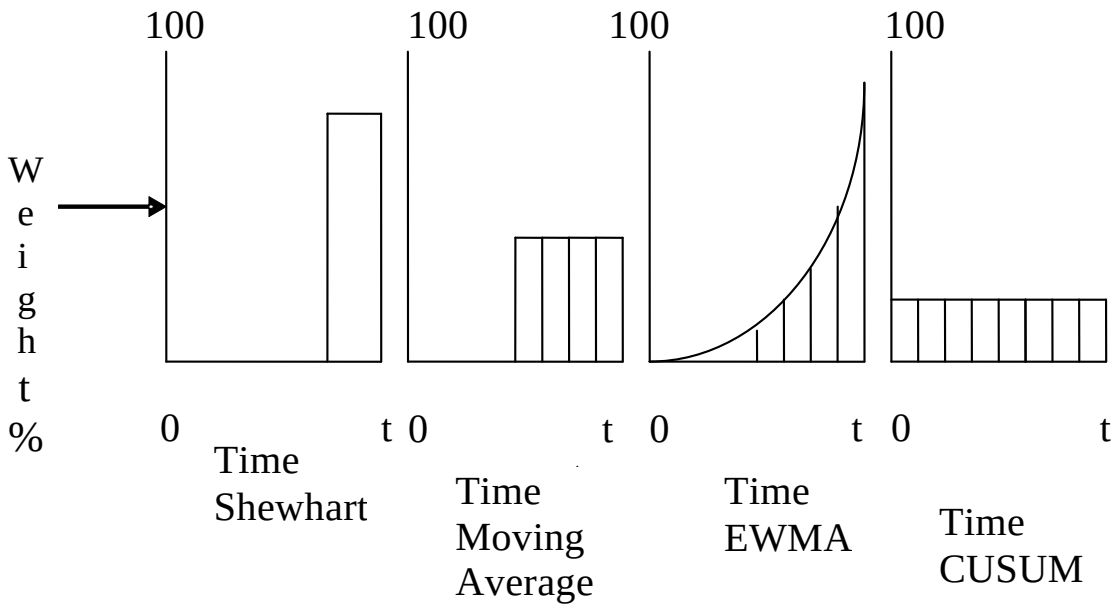

Figure 4 The traditional multivariate control charts. (Ganesan et al., 2004)

The classical SPC approaches from the perspective of stochastic industrial processes, unfortunately, do not perform well for applications that exhibit high correlations in observed process variables. These methods treat the variables independently and also extract the magnitude of deviation in each variable independently of all others, ignoring the correlation structure. Hence, process deviations or abnormal events in the process may not be detected (see Figure 2).

Using multivariate methods, such as PCA and PLS, process variables are treated simultaneously and correlation information is exploited to derive improved monitoring and fault diagnosis (MacGregor and Kourti, 1995). These techniques have found wide application in the process control community and are discussed next.

2.2 Principal Component Analysis (PCA)

PCA is a linear multivariate statistical method generally used for data compression and information extraction by projecting a high-dimensional dataset onto a space with significantly fewer dimensions. Specifically, PCA transforms a set of highly correlated variables into new set of uncorrelated variables, called principal components (PC). Principal components are orthogonal to each other and are also a linear combination of the original variables. In most cases the first few principal components that explain most of the variation in the data are retained. In order to uniformly handle variables with different amplitude and frequency measurements are normally mean centered and scaled prior to performing PCA (Rosen and Lennox, 2001).

Given a data matrix $X \in \mathbb{R}^{n \times m}$ with n observations and m process variables, PCA decomposes the data matrix X into sum of the product of the PC scores t_i and PC loadings p_i, that is

$$X = TP' = \sum_{i=1}^{m} t_i p_i' . \qquad (1)$$

The principal component loadings are orthonormal to each other and denote the direction of the hyperplane that captures the maximum possible residual variance (variance that are not captured by the model) in the measured variables[1] (Bakshi, 1998). The scores and the loadings are obtained by singular value decomposition of the data matrix or, alternatively, eigenvalue decomposition of

[1] Note that the term *measured variable* in this thesis refers to the observed variables that are not controlled by the system

the covariance matrix of X Singular value decomposition of the data matrix is given by

$$X = U\Lambda^{\frac{1}{2}}V \tag{2}$$

where Λ is the diagonal matrix containing the eigenvalues of the covariance matrix of X, and the eigenvalues (λ_i) are the variances of the principal components. The loadings and scores are obtained via $P' = V$ and $T = U\Lambda^{\frac{1}{2}}$ respectively (Bakshi, 1998). The principal components are ordered according to the variance explained by the transformed features, with the leading principal component, that is $t_1 = Xp_1$, being a linear combination of the columns of X that has maximum variance subject to $\|p_1\| = 1$ where p_1 is the eigenvector of the covariance matrix of X

$$\text{Cov } X = n^{-1}X'X \tag{3}$$

The second principal component is orthogonal to the first principal component and explains the maximum residual variance (after t_1) subject to $\|p_2\| = 1$, and so forth for all *m* components.

For data sets with large number of variables one often finds that multiple measurements of the same variable, or constraining relationships between different variables result in ill-conditioning or collinearity problems due to the redundancy in the data set, resulting in several of the eigenvalues to be equal or close to zero. This redundancy can be removed from the data by representing it with smaller number of principal components whose eigenvalues are greater than a very small positive number (Ku et al., 1995). Hence by selecting k non-zero eigenvalues the data matrix can be approximated as

$$X = T_k P_k' + E = \sum_{i=1}^{k} t_i p_i' + E \tag{4}$$

$$\hat{X} = T_k P_k' = \sum_{i=1}^{k} t_i p_i'. \tag{5}$$

Here $k \leq \min\ m,n\ ,\hat{X}$ represents the reconstructed data and E is the residual matrix $E = X - \hat{X}$. Several techniques are available to assist in selecting the appropriate *k*, for example, percent variance, parallel analysis, scree plots and cross-validation (Jackson, 1991).

For applications considered later, the percent variance criteria have been chosen for selecting appropriate number of PCs. This method determines k by calculating the smallest number of loading vectors needed to explain specified minimum percentage of the total variance $\theta_{\text{threshold}}$, typically 90% or 95%:

$$k > \min\left\{ d \Big/ \frac{\sum_{i=1}^{d} \lambda_i}{\sum_{i=1}^{m} \lambda_i} > \theta_{\text{threshold}} \right\} \qquad (6)$$

The discarded eigenvalues are assumed to correspond to PCs explaining high-frequency variations in data, probably due to the influence of noise. The subspace spanned by $\hat{X}$ is referred to as the score space and that spanned by E the residual space. A geometric representation of PCA is illustrated for a three-dimensional system in Figure 5 where the data are well explained by two principal components.

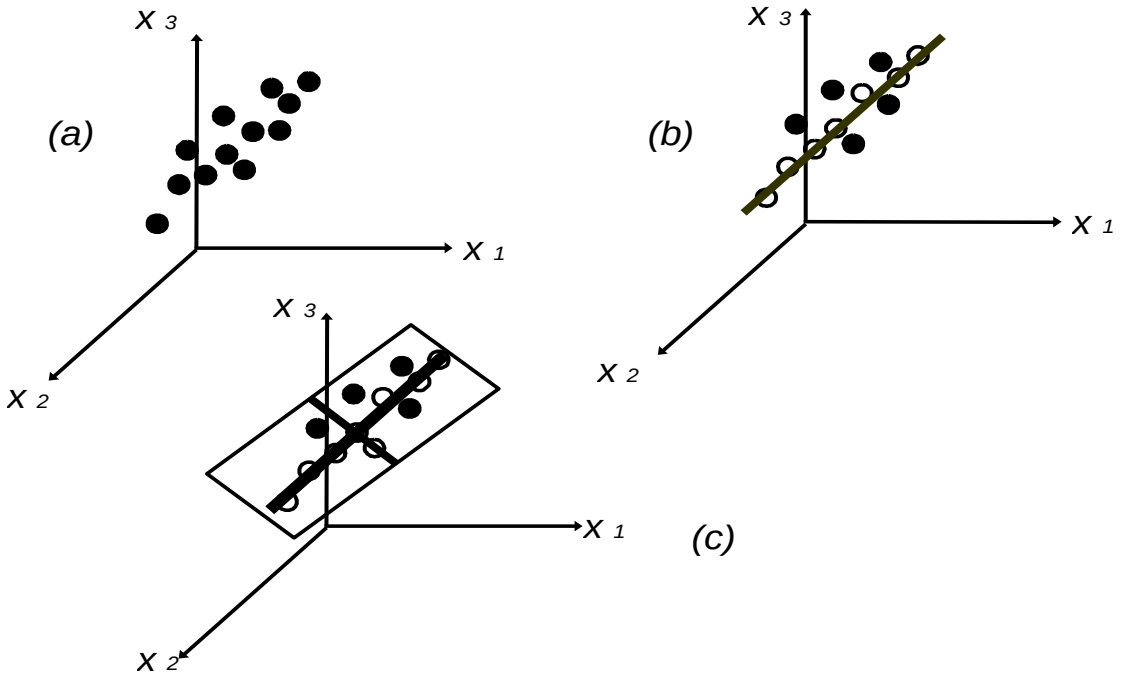

Figure 5 Geometric representations of the steps in principal component analysis for a 3-dimensional system showing (a) the data points in the observation space, (b) the first principal component, (c) the plane defined by the first two principal components. This figure indicates that the derivation of principal components is based on the successive projection of lines through three dimensional space.

2.2.1 Process Monitoring Using PCA

For monitoring a process with PCA, two-dimensional score plots (t_1 vs t_2), Hotelling's T^2 statistic and squared prediction errors (SPE) or Q statistics are typically used. The two-dimensional score plots are used when most of the variation is well explained by the first two PCs. In situations where more than two PCs are retained the use of two-dimensional score plots is cumbersome, even though abnormal variations in the process can be detected by the scores that move out of the confidence limit in the two dimensional score plot (Kresta et al., 1991). Hotelling's T^2 statistic explains the variation within the score space by using all the retained PCs and hence, explain most of the variation in the data. T^2 statistic can provide a better performance in monitoring when the number of retained PCs is greater than two. T^2 is the sum of normalized squared scores given by

$$T_i^2 = \sum_{j=1}^{k} \frac{t_{ij}^2}{\lambda_j} = x_i P_k \lambda^{-1} P_k{}^T x_i^T \quad (7)$$

where T_i^2 is the T^2 value for the i^{th} row of measurements k is the number of scores selected, t_{ij} is the score corresponding to the i^{th} row and j^{th} loading , x_i is the i^{th} observation in X and P_k is the matrix of k loading vectors retained in the PCA model. Confidence limits for T^2 can be calculated by means of the F-distribution as follows

$$T_{k,n,\alpha}^2 = \frac{n-1\ k}{n-k} F_{k,n-k,\alpha} \quad (8)$$

where $F_{k,n-k,\alpha}$ is the upper $100\cdot\alpha$ % critical point of the F-distribution with k and $n-k$ degrees of freedom.

If a new event occurs, which is not captured in the PCA model, the T^2 chart based on the first k PCs may not be sufficient for detecting the fault. Such events change the nature and the dimensions of the relationship between the process variables and are detected using both the T^2 statistic and the Q statistic (Kresta et al., 1991).

The Q statistic or squared prediction error (SPE) measures variability that breaks the normal process correlation. Mathematically, Q is obtained as the sum of the squared errors in the residual space or the sum of variations in the residual space, which is defined as

$$Q_i = \sum_{j=1}^{k} \left(x_{ij} - \hat{x}_{ij} \right)^2 = e_i e_i^T = x_i (I - P_k P_k{}^T) x_i{}^T \quad (9)$$

where $\hat{x}_{ij}$ is the predicted value of x_{ij}, e_i is the i^{th} row of the residual matrix E and I is the identity matrix of appropriate size. The Q statistic is thus a measure of the amount of variation in each sample not captured by the retained PCA model. The upper confidence limit for Q can be calculated based on all the eigenvalues λ_i of the covariance matrix of X i.e.

$$Q_\alpha = \Lambda_1 \left[1 + \frac{c_\alpha \ 2\Lambda_2 \theta^2 \ ^{\frac{1}{2}}}{\Lambda_1} + \frac{\Lambda_2 \theta \ \theta - 1}{\Lambda_1^2} \right]^{\frac{1}{\theta}} \quad (10)$$

where $\Lambda_i = \sum_{j=k+1}^{m} \lambda_j^i$ for $i = 1,2,3,$ $\theta = 1 - \frac{2\Lambda_1 \Lambda_3}{3\Lambda_2^2}$,and c_α is the standard normal deviate corresponding to the upper $1-\alpha$ percentile.

The values of these two statistics are also calculated for the new data set, that is the scores of the new data are calculated by projecting these data onto the k principal component loadings calculated from equation 5;

$$t_{new,i} = X'_{new} p_i \quad (11)$$

The residuals are calculated as follows:

$$e_{new} = X_{new} - \hat{X}_{new} \quad (12)$$

where $\hat{X}_{new} = P_k t_{k,new}$

If, at a specific point(time), T^2 or Q for the new data set is outside the calculated control limits, the process is judged to be out of control at that point.

2.2.2 Fault diagnosis using contribution plots

When a fault has been detected using the T^2 and Q statistics, it is important to identify the cause of the out-of-control status. This can be achieved using contribution plots. In a PCA model two types of contribution plots are needed to identify the fault since two types of multivariate control charts are used, i.e., by Q-chart for residuals and Hotelling's T^2 chart for systematic variations within the model structure (Teppola et al., 1998). PCA contributions plots are defined as the contribution of each process variable to the individual score of the T^2 or Q statistic. Note that the role of variable contribution plots in fault identification is to show which of the variables are related to the fault rather than to reveal the

actual size of it. The variables with high contribution to the contribution plots are simply the signature of such faults (Kourti, 2005).

Contribution Plots: Hotelling's T^2 Statistic

For the T^2 statistic value of an observation, the variable contributions to an out-of-limits value are obtained as a bar plot of the mean of the absolute value of $T\sqrt{\Lambda^{-1}}P'$ which shows how each variable is involved in the calculation of T^2 value at that point. T is the matrix containing the score values of all the variables at that scale and P is the corresponding loading matrix. The matrix Λ is a diagonal matrix of the eigenvalues. The inverse of this matrix normalizes the score values of different PCs. In order to decide whether the individual variable contribution to the T^2 value is significant or not, one can either compute control limits for the contribution plots or one can compare the size of the variable's contribution under faulty conditions with the size of the same variable's contribution under normal operating conditions. In other words variables with the largest contribution to the T^2 value often indicate the source of the fault. The control limit for individual variable contribution will be the length of T^2 interval, that is the square root of the T^2-limit (Jackson, 1991; Johnson and Wichern, 1992; Teppola et al., 1998).

Contribution Plots: Q Statistic

When an out-of-control situation is detected using the Q chart, bar graphs of the ratio of residual variance of each variable in the testing and training set show the variations of each process variable in the residual space. This is computed by generating the residual matrix E_{new} and E_{old} of the testing and training data set by the following equation:

$$E_{new} = X_{new}\left(I - PP^T\right) \tag{13}$$

where X_{new} is the new data matrix (testing data) and P is the loading matrix containing the retained PCs in the PCA model.

Similarly,

$$E_{old} = X_{old} \quad I - PP^T \tag{14}$$

where X_{old} is the old data matrix (training set).Then finding the ratio of residual variance, that is $\mathrm{var}(E_{new})/\mathrm{var}(E_{old})$, can assist in identifying the variables responsible for the variations in the residual space. Variables with large variation in the residual space will show a large value of the residual variance and will be also be out of the control limits of the Q chart.

2.2.3 MSPC Extensions

PCA is based on the assumption that process operates at a steady state operating condition and each of the variables is uncorrelated in time. In practice, chemical processes exhibit dynamic behavior and, therefore, in addition to being cross correlated, variables exhibit some degree of autocorrelation arising from, for example, throughput changes, controller feedback and the presence of unmeasured disturbance. Moreover, the high sampling frequency relative to the dominant process time constant and process inertia may lead to incorrect decisions due to false alarms when using PCA. To address these and other drawbacks, several extensions of PCA have been developed to account for non-Gaussianity, autocorrelation and nonlinearity in observed data. These are briefly introduced below.

Dynamic PCA incorporates both static and dynamic process characteristics (Kresta et al., 1991; Ku et al., 1995). Nonlinear PCA was proposed by incorporating the principal curves concept into an artificial neural network model (Dong and McAvoy, 1996). Kramer (1991) proposed auto-associative neural networks for extracting nonlinear principal components in high-dimensional data. A multivariate monitoring method based on multiblock PCA algorithm was proposed to monitor not only the entire process, but also each unit of the process. With multiblock PCA the data matrix is divided into multiple blocks of variables and captures the relationship between the sub-blocks by applying PCA

to each block as well as to all the blocks taken together (MacGregor et al., 1994; Wold et al., 1996). Multiway PCA was proposed to monitor time-varying batch processes. Information is extracted from the trajectories of all the process variables by projecting them onto principal components. Recursive or adaptive PCA was developed to circumvent difficulties associated with monitoring the time varying nature of certain processes (such as waste water treatment operations. Changes in process conditions are monitored by updating the mean, variance and the covariance structure of the monitored variable recursively or by using an exponential memory function (Dayal and MacGregor, 1997; Rosen and Lennox, 2001). Moving PCA was proposed to detect process deviations by monitoring changes in the direction of the PCs (Kano et al., 2000).

These extensions to PCA have mainly been used for multivariate analysis of process data at uniform scale, with a few being extended to multiscale analysis, which is discussed later in this chapter.

2.2.4 Limitations of MSPC and Related Approaches

As highlighted earlier, process data are multiscale in nature due to contributions from events occurring with different localizations in the time frequency space. However, process monitoring using PCA and its extensions assume steady state operation and do not take into account the non-stationary process behavior, i.e. they operate on data collected at a fixed scale. Techniques such as Western Electric rules for identifying patterns (Western Electric, 1956) and combined Shewhart and CUSUM charts for detecting shifts of large and small magnitudes (Lucas, 1982) have been proposed as a solution to the single scale nature of SPC methods. However, these techniques represent data at all scales at the finest scale using the single scale approach of SPC methods and are computationally costly (Aradhye et al., 2003; Bakshi, 1998).

Another disadvantage of conventional MSPC based on PCA is that the obtained models are contaminated by an embedded error from noisy data whose magnitude is proportional to the number of retained PCs in the model (Malinowski, 1991). This limited ability of PCA method to remove the error by

eliminating some components deteriorates the quality of the model represented by the retained PCs leading to unreliable performance of PCA in many applications. Specifically, detection of small deviations may not be possible while the detection of large deviations is delayed due to the presence of errors that are leaked into the model by the retained PCs. The quality of gross-error[2] detection and estimation of the missing data by PCA is also affected by contaminated error in the PCA model (Bakshi, 1998). Therefore, methods that can separate the underlying error from the process are desirable for improved performance of MSPC methods based on PCA (Bakshi, 1998).

Finally, in single scale multivariate methods data along each PC is monitored by single scale charts. For example, the traditional multivariate control charts such as Hotelling T^2 chart are single scale and are suitable for extracting information only in the time domain because they represent data at fixed frequencies in the entire time domain (Ganesan et al., 2004). Methods for effectively handling multiscale characteristics associated with data are therefore desirable. In the next section a multiscale framework based on wavelets is discussed.

2.3 Multiscale Process Monitoring: Theory

It is useful to consider a physical analogy for a conceptual appreciation of the multiscale character of process data. Suppose one is given a road map, geographical map of a city or any image. The default representation is assumed to be at a “low scale” and the resolution is “fine”; in other words the neighboring roads, places or the adjacent pictures are close in distance and can be seen distinctly on the map at this fine resolution/low scale. These low scale representations make the map dense and will be difficult to identify the places clearly. For a detailed view on a particular area from the picture or the map, the scale must be increased (i.e., zooming) even though the resolution becomes coarse. By varying the scale at which the object is viewed, different levels of

[2] Gross errors are errors that occur when a measurement process is subject occasionally to large inaccuracies

details are possible at the different resolutions. At a coarse resolution these details correspond to larger structures providing a larger context of the image. So it is natural practice to have the first image details done at coarse resolution and then increase the resolution. Therefore, any multiscale representation aims at showing features on a scale ranging from the fine representations to the very coarse representations (Mallat, 1989; Tangirala, 2001).

The concept of representing a time signal at different resolutions is encapsulated in the preceding example, where the sampling interval determines the scale or the resolution of the signal. In fact, representing a continuous time signal with a sampled signal is analogous to the above example. Here the sampled signal is considered to be a single scale representation of the continuous time signal. Many chemical processes have different signals at different sampling intervals or rates, which makes the sampled signals to have different time resolutions or scales. The sampling interval of each signal determines its finest resolution, where each signal is an approximation of the underlying continuous signal. This way of multiscale representation can be considered as a special case of the general multiscale representation of any process signal in which the process signals are transformed into different scales (resolution), namely from finest scale to coarser scale and therefore exhibiting a multiscale behavior.

Multiscale representation of a signal is essential for monitoring and control of the process operations for a number of reasons including:

(a) Physical and chemical phenomena inherently occur at different spatial and time scales. Deterministic events usually occur at different locations and with different localization in time and frequency and the stochastic events such as measurement noise, disturbance and faults are scale and time dependent. In other words they occur in different time zones and frequency bands, as highlighted in Figure 3(a) and (b). Thus to identify a fault in a system it may be necessary to monitor observed signals in both the time and frequency domain due to this characteristic multiscale nature.

(b) Conventional monitoring and fault diagnosis methods are not suited for creating process models for several interrelated operational tasks such as closed loop feedback control, adaptive control, fault diagnosis and scheduling and planning of operating procedures, which are deployed at different time scales. Therefore, models that describe process behaviour at different time and frequency scales are essential to account for these plant-wide influences.

(c) Measurements of process behavior (variables) by sensors are done at different sampling rates including control actions at correspondingly different rates. For example, variables that change slowly with time do not need a fast sampling rate and variables that change quickly with time need a high sampling rate. A multiscale process model is essential for optimal fusion of measurement information at various time scales with control actions, since a model that represents the data at different time scales matches the sampling rate of various measurements and application rates of control actions (Stephanopoulos et al., 2008).

In response to above process realities, multiscale approaches have been developed to solve the problems in process monitoring and fault diagnosis. Multiscale principal component analysis (MSPCA) is a multiscale extension of the conventional PCA-based statistical process control methodology which has attracted a lot of attention in recent years. In MSPCA wavelet decomposition is first applied to decorrelate individual signals while in a subsequent step application of PCA removes cross-correlation between the variables at each scale as determined by the wavelet analysis. A brief overview of multiresolution analysis with wavelets that form the basis of the multiscale approach using PCA is reviewed in the next sections.

2.4 Multiresolution Analysis and Wavelets

Wavelets are a family of basis functions that provide a mapping from the time domain to the time-frequency domain. Wavelets can be used to decompose the

signal into different resolutions by projecting onto the corresponding wavelet basis functions using the so-called multiresolution analysis (MRA) introduced by Mallat, (1989). A wavelet set is constructed from a fundamental basis function or the mother wavelet by a process of translation and dilation. The wavelet set is defined as

$$\psi_{ab}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \tag{15}$$

where ψ is the mother wavelet function, a the dilation parameter and b the translation parameters. Orthogonality between the decomposed signals is achieved by defining a dyadic grid across which the dilation and translation parameters are selected. The location of the wavelet in the time domain is determined by the translation parameter, while the location in the frequency domain and scale of the time frequency localization are determined by the dilation parameter.

In the theory of multiresolution analysis any signal can be decomposed using a family of wavelet basis functions based on convolution with the corresponding filters (Mallat, 1989). Thus multiscale representation of a signal can be achieved by expressing data as a weighted sum of orthonormal basis functions that are localized both in time and frequency. An example of multiscale representation of a signal obtained by projecting it onto the corresponding basis function is shown in Figure 6. In this illustration Daubechies wavelet and the corresponding scaling functions are used to decompose the signal into multiple scales. The fine scale features (high frequency components) in Figure 6(c-f) are captured by the wavelet coefficients and the low frequency content of the original signal (Figure 6(b)) is captured by a set of basis functions called scaling function or father wavelet, which has the shape of a low-pass filter.

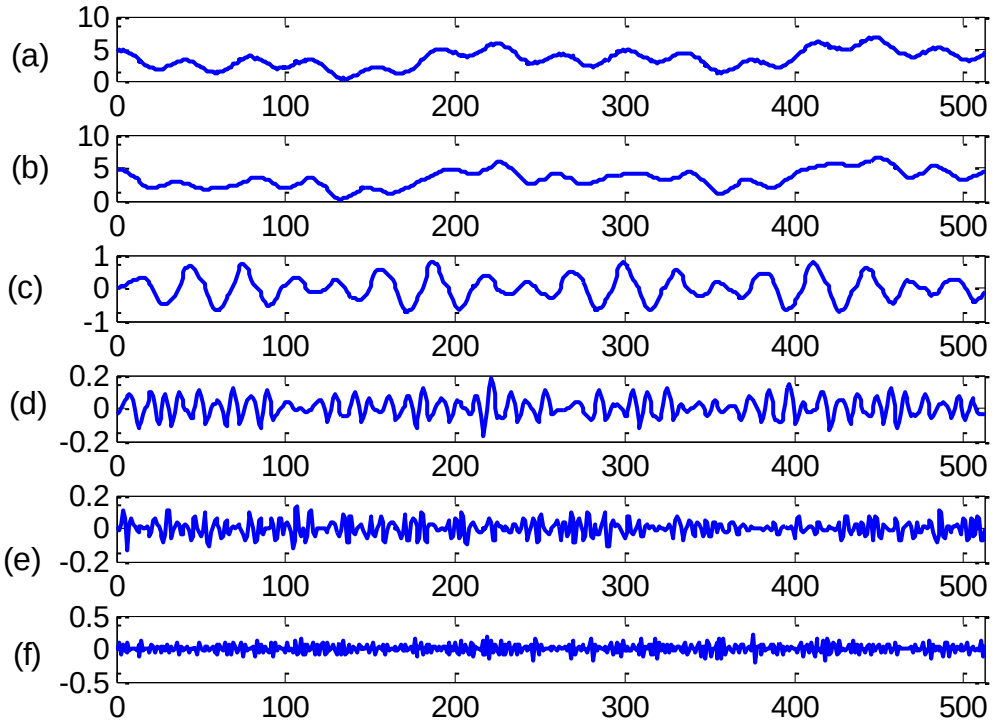

Figure 6: Multiscale representation of data by MRA. Decomposition of an (a) observed signal into multiple scale representations using Daubechies wavelet. Here the depth of decomposition level, L=4 is used. The significant or deterministic component (b) explaining the most variation in the signal is associated with the low frequency components, with (c)-(f) progressively explaining less variability. In particular, the last signal is associated with high frequency components.

More formally, convolution with the filter H represents projection on the scaling function, and convolution with the filter L represents projection on a wavelet. The corresponding coefficients thus obtained are referred to as the scaling function coefficients and wavelet coefficients. The coefficients at different scales can be computed as

$$a_m = Ha_{m-1} \tag{16a}$$

$$d_m = La_{m-1} \tag{16b}$$

where a_m is the vector of scaling function coefficients at the scale m and d_m is the vector of wavelet coefficients. Here a_m represents high scale low frequency components and d_m represents low scale high frequency components. In

wavelet decomposition the original data are considered to be the scaling function coefficients at the finest scale, i.e., $x = a_0$. In terms of the original data x equations 16a and 16b can also be represented as

$$a_m = H_m x \tag{17a}$$

$$d_m = L_m x. \tag{17b}$$

with H_m denoting application of the H filter m times while L_m is denotes application of the H filter $(m-1)$ times and the L filter once. Decomposition and reconstruction of the original data can be undertaken by reassembling the signal from its wavelet coefficients at all scales d_m, $m = 1,\ldots,L,$ and scaling function coefficients at the coarsest scale, a_L (Bakshi, 1998; Wang and Romagnoli, 2005)

Wavelet transforms have found applications in data compression (Mallat, 1989), filtering (Donoho et al., 1995), function estimation (Sadlar and Swami, 1999), feature extraction (Bakshi and Stephanopoulos., 1994) and multivariate statistical process monitoring and control where their ability to compress multiscale features and approximately decorrelate signals is exploited (Bakshi, 1998; Rosen and Lennox, 2001; Teppola and Minkkinen, 2000; Wang and Romagnoli, 2005). In signal estimation wavelets are used for denoising signals based on thresholding. Wavelet thresholding is also used in detecting outliers. In system identification several regression techniques that use wavelets have also been developed (Alsberg et al., 1997; Engel, 1994; Ogden and Parzen, 1996; Sjoberg et al., 1995).

In considering the use of wavelets in multiscale process monitoring methods, several issues must be addressed for effectiveness (Ganesan et al., 2004):

- Selection of a wavelet basis function
  Wavelet transforms require the choice of an analyzing function or the mother wavelet. Ideally a mother wavelet that provides a complete orthogonal expansion of the signal must be used. Also, the selected basis

function must have very good time frequency localization property in order to concentrate the spectral energy within an optimal band for the problem at hand.

- Depth of wavelet decomposition

  It is essential to optimize the depth of the decomposition of the time-domain data for maintaining the quality of the filtered signal. Excessive decomposition will compress the data potentially eliminating important features and over-smoothing of the signal. On the other hand, selection of improper depth of wavelet decomposition may lead to retention of a considerable portion of nose in the reconstructed signal.

- Length of the testing window

  For off-line and on-line monitoring the length of the testing window needs to be selected properly for optimal decomposition of the signal to reduce excessive smoothing and also increased sensitivity to smaller shifts at lower scales. For example, in offline analysis the (dyadic) length of the signal is considered a single window, while for online monitoring the dyadic window length is maintained by moving the window in time by including the most recent measurement.

- Border distortion or end-effects

  Non-causal nature of most of the wavelets causes a boundary problem in the last scaled signal of finite length. As a result the original signal cannot be represented accurately at the edges in on-line monitoring. Although techniques exist to minimize the boundary problem effect, they tend to be computationally costly due to the added constraints on the time periods and scales analyzed. Hence, large datasets are required for analyzing large scale or low frequency features with wavelets to minimize the effect of border distortion.

- On-line and off-line monitoring

  Wavelets are non-causal and, hence induce a time delay in the computation of coefficients at non-dyadic locations as well as at coarser

scales in online applications (Bakshi and Nounou, 2000). While this delay in the computation of wavelet coefficients can be eliminated using a moving window of dyadic length and taking the last coefficients at all scales, the orthonormality property (and, therefore, the decorrelation effect) of the coefficients is lost.

The choice of a proper basis function that is adapted to the different features of the monitoring signal is one of the most important tasks in wavelet based process monitoring methods. For example Figure 7 shows the multiscale decomposition of the same signal using Haar wavelet as shown in Figure 6. In Figure 7 the reconstructed signal at different scales are not perfectly reconstructed compared to the corresponding signals using Daubechies wavelet in Figure 6. This is because Daubechies wavelet is smoother than Haar wavelet and is capable of capturing smooth low-frequency features in the signal. This is clearly visible in comparing low-frequency components in Figure 6(b) and in Figure 7(b). Typical characteristic plots of the Haar and Daubechies basis functions are given in Figure 8 and show that the Haar wavelet is adapted to step functions while Daubechies wavelet is adapted to periodic functions similar to the signals used in Figures 6-7.

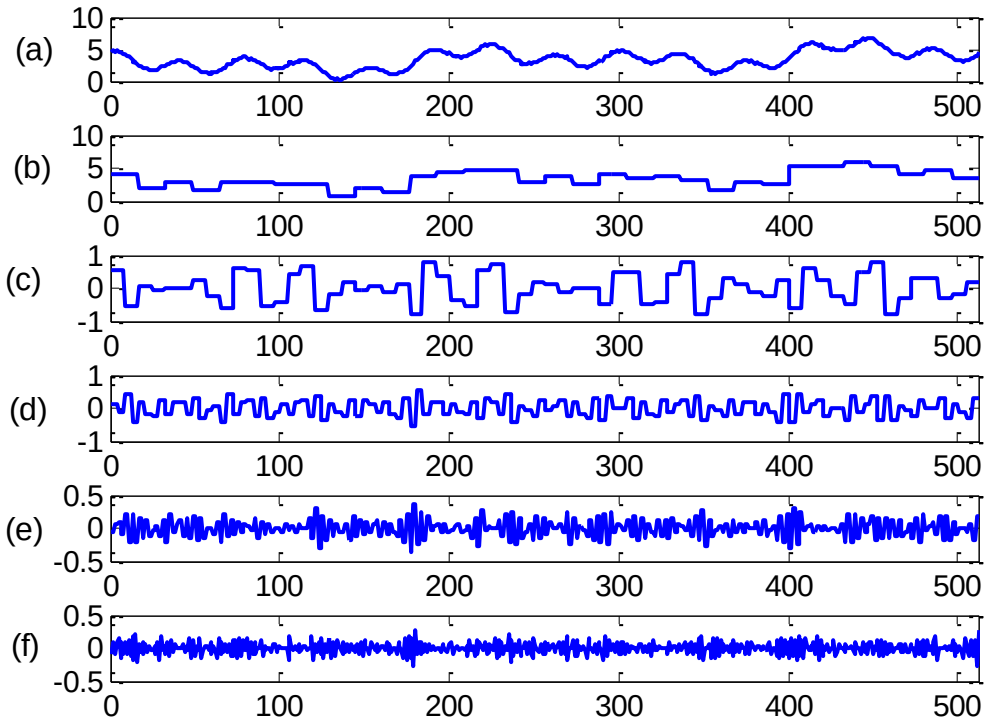

Figure 7 Decomposition of an (a) observed signal into multiple scale representations using Haar wavelet. Here the depth of decomposition level, L=4 is used. The significant or deterministic component (b) explaining the most variation in the signal is associated with the low frequency components, with (c)-(f) progressively explaining less variability. In particular, the last signal is associated with high frequency components

To satisfy the conflicting requirements of basis function – completeness, localization in both time and frequency, orthogonality or limited redundancy (non-orthogonal bases) – a large number of wavelet basis functions have been introduced. Unfortunately, searching through the extensive libraries of mother wavelets for a suitable basis function is not always practical. Automatic time-varying adjustment of the shape of the analyzing wavelets to the signal is preferable instead of searching for a suitable wavelet function. Moreover, for optimal multiscale decomposition of a signal it may also be desirable to modify the shape of the analyzing wavelet in time and scale, particularly for non-stationary data. The inherent constraint of a unique mother wavelet (analyzing function) does not allow for this flexibility. There is, therefore, a need for data adaptive wavelet transforms (Yiou et al, 2000).

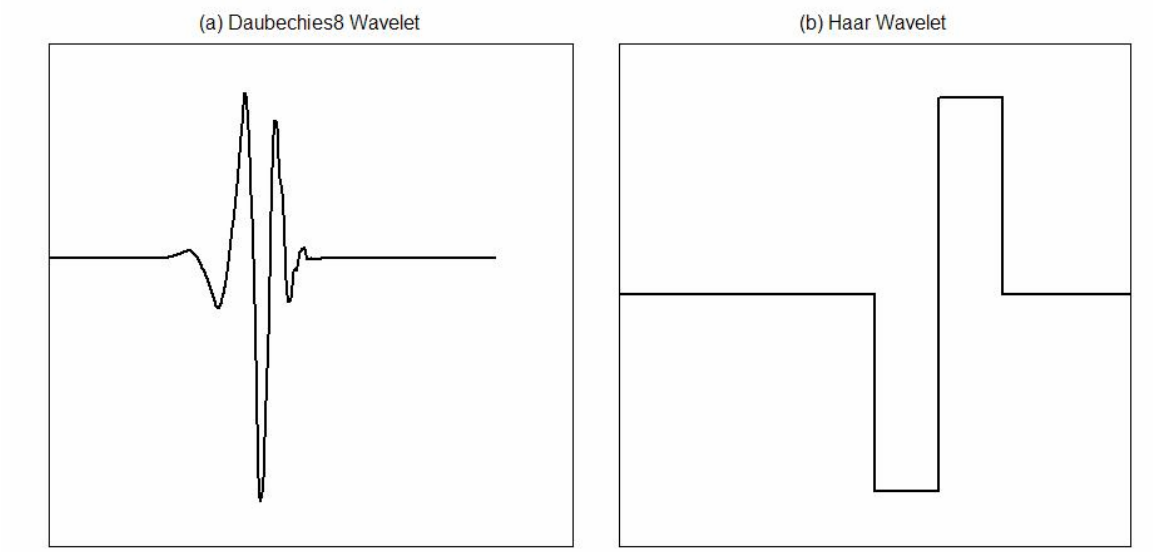

Figure 8 Typical characteristic plots of basis functions used in (a) Daubechies and (b) Haar wavelets.

2.5 Multiscale Process Monitoring using Wavelets

Wavelets have been used in multiscale analysis of observed data, particularly in recent years, due to appealing properties of wavelets discussed above. Grossman and Morlet, (1984) applied wavelets to change detection problems paving the way for the use of wavelet transform as a general tool for signal processing in various statistical applications such as change-point detection, edge detection and discontinuity analysis (Mallat and Zhong, 1992; Song and Jutamulia, 2000; Wang, 1995), as well as statistical process monitoring and control (Bakshi, 1998). Kosanovich and Piovoso, (1997) proposed to filter process data using a univariate finite median hybrid filter to obtain Haar wavelet transform coefficients and subsequently develop a PCA model using these coefficients. Bakshi, (1998) first established a unified multiscale statistical process monitoring for both univariate and multivariate data with the development of multiscale PCA. In this framework, relationships between variables are extracted using PCA while wavelet analysis captures autocorrelation structure. Multiscale PCA exploits properties of wavelet

transforms to extract deterministic features using a smaller number of wavelet coefficients, with stochastic components spreading their energy across all the coefficients, that is, deterministic and stochastic components are separated with different time frequency localization.

Multiscale PCA (MSPCA) methodology proceeds by decomposing each variable on a selected family of wavelets and computing independent PCA models for the coefficients at each scale. The control limits for the T^2 and Q control charts at each scale are computed using data collected from a process under normal operating conditions. As new data are collected wavelet coefficients are calculated at each scale based on the selected discretization procedure, and the scores at each scale are monitored using T^2 and Q control charts. If there are violations at the specified limit at any scale, those scales are considered significant and the signal is reconstructed back to the time domain using all the significant coefficients. The process status is determined by comparing the monitored statistics (e.g. T^2 and Q) reconstructed for the significant scales with the expected limits for the corresponding scales under normal operating conditions. Thus, the MSPCA method extracts the significant signal features and also adapts to the nature of the signal features. MSPCA improves on the standard PCA approach in statistical process monitoring because it integrates the tasks of feature extraction and process monitoring, eliminating the need for pre-filtering the data (Bakshi, 1998).

Aradhye et al., (2003) presented a theoretical analysis of the performance of multiscale SPC compared against conventional SPC methods based on their average run lengths. Although multiscale SPC did not perform better than the conventional SPC for detecting specific types of changes such as large and small shifts of a certain size, its average performance in detecting a range of changes in different types of measurements was better.

Kano et al., (2002) proposed integrating multiscale methods with monitoring methods for detecting changes over the distribution of process data as well the principal directions obtained from a PCA model. In order to monitor the process

and simultaneously perform early fault diagnosis Misra et al., (2002) proposed a MSPCA procedure which makes use of variable grouping and the analysis of contribution plots when an event is detected in the control charts at any scale. A wavelet-partial least square (Wavelet-PLS) model was proposed for both data analysis and process monitoring (Teppola and Minkkinen, 2000). A PLS model is constructed based on the filtered measurements obtained by filtering scales associated with low frequency components such as seasonal fluctuations and other long-term variations

Yoon and MacGregor, (2004) presented an approach based on multiscale representation of data in the original time domain that include the successive extraction of principal components for all variables represented at all scales, according to the decreasing magnitude of the associated eigenvalues of the covariance matrix at each scale. The loadings for the variables at the same scale contain non-zero entries due to orthogonal property of the wavelet coefficients and, therefore, each extracted principal component strictly provides the information corresponding to the specific scale. Although the results obtained with this approach are not very different with traditional MSPCA for the same number of PCs, this approach allows for a ranking of the relevant structures underlying data variability with respect to the contributions from the variables covariance at different scales (Reis and Saraiva, 2006).

Shao et al., (1999) and Fourie and de Vaal, (2000) proposed nonlinear process monitoring algorithms similar to MSPCA using neural networks. Zhiqiang and Qunxiong (2005) proposed a wavelet-based adaptive multiscale nonlinear PCA based on an improved input training neural network to monitor slow and feeble changes of the process signal that cannot be detected using conventional PCA. A multiscale orthogonal nonlinear strategy was developed for improved initial fault interpretation by using the optimal wavelet decomposition and the orthogonal NLPCA, in which the process monitoring can be performed by two scales only, namely, approximation and the highest level detail functions (Maulud et al., 2006). Choi.et.al, (2008) proposed a methodology, where the concept of

kernel PCA is combined with multiresolution analysis for nonlinear multiscale process monitoring.

In order to provide physical insight into the different scales under analysis as well reduce the number of available monitoring tasks, Rosen and Lennox, (2001) proposed combining components at different scales in the original time domain that is not in the wavelet transform domain using prior knowledge about a process. In this methodology there is no reconstruction stage as in the basic MSPCA method and a better adaptation in the mean to the non-stationary data is obtained by omitting the coarser scale components in the monitoring procedure.

A 2D wavelet transformation was applied to the compressed data from near-infrared (NIR) spectra collected over time and then a PCA model was estimated for this 2-D compressed matrix for monitoring the deviations of the incoming spectra from spectra that is collected during the normal operation (Trygg et al., 2001). Luo et al., (1999) proposed frequency band selection methodology to select frequency bands and subsequently data analysis is performed for sensor fault detection. Sun et al., (2003) used wavelet-domain hidden Markov models (HMM) to develop a framework for detecting abnormal situations. Lee et al., (2005) used adaptive multiway PCA (MPCA) to update the covariance structure at each scale to deal with changing process conditions in batch process monitoring.

Multiscale methods reviewed in this section focused on strategies and methodologies to detect process features where conventional multivariate process monitoring are not effective. Advantages of multiscale methods include denoising, ability to analyze signals localized in time and frequency, handling small and large shifts in the mean and/or variance within the same monitoring framework, feature extraction, automatically handling autocorrelation, different data distributions, errors in data, and nonlinearities (Ganesan et al., 2004).

In the next chapter an alternative multiscale monitoring strategy based on singular spectrum analysis (SSA) is proposed. The data adaptive nature of SSA

provides more flexibility compared to other spectral techniques for decomposing signals into multiple scales such as wavelets.

Chapter 3 Multiscale Singular Spectrum Analysis

Singular Spectrum Analysis (SSA) is a data-adaptive spectral technique for time series analysis. It derives its data-adaptive characteristic from the basis functions obtained from the data, unlike in wavelets where fixed basis functions are specified a priori. This property also makes it possible to extract reliable information from short and noisy time series. Essentially, multiscale representation of a signal using Singular Spectrum Analysis (SSA) decomposes the signal into deterministic and stochastic constituents. SSA uses an augmented (information) matrix called the lag-correlation matrix derived from the data to define basis functions referred to as empirical orthogonal functions (EOFs) computed using, typically, PCA (Yiou et al., 2000). In fact, SSA is a particular application of the Karhunen-Loéve expansion theorem that is also used in PCA analysis (Vautard et al., 1992). Similar to PCA, these EOFs are ranked in decreasing order with respect to the spectral or singular values of the information matrix. Moreover EOFs depend only on the data being analyzed, making SSA more flexible than other spectral techniques.

In this chapter SSA methodology and its recent applications in various fields such as geophysics, climatology and life sciences are reviewed. A methodology of multiscale process monitoring method using SSA for signal decomposition is proposed and its properties are demonstrated using a simulated example.

3.1 Singular Spectrum Analysis (SSA)

Singular spectral time series analysis involves singular value decomposition (SVD) of a trajectory or lagged covariance matrix obtained from the original time series, followed by reconstruction of the series using subsets of eigenfunctions and corresponding principal components. The time series is first embedded into an *M*-dimensional time series known as the *trajectory matrix*. Singular value decomposition is then applied to decompose the trajectory matrix into a sum of

elementary matrices. Subsequently, the elementary matrices that contribute to the norm of the original matrix are grouped, with each group giving an approximation of the original matrix. Finally, the smoothed approximation of the time series is recovered by diagonal averaging of the elementary matrices obtained from the decomposing the trajectory matrix. An outline of the basic SSA methodology is summarized in Figure 9 (Golyandina et al., 2001) and the procedural steps involved in SSA in detail below.

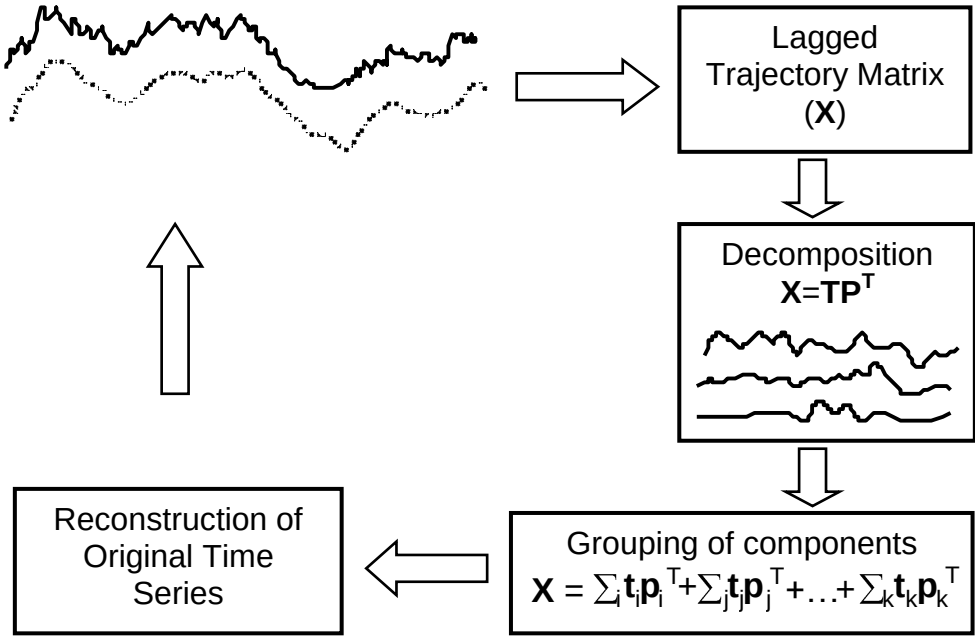

Figure 9 Decomposition and reconstruction of a time series by use of singular spectrum analysis.

Step 1: Embedding

Given a time series $x\ t\ , t = 1, 2, ..., N$, a trajectory matrix is constructed by sliding a window of length M along the time series to give lagged vectors $\mathbf{x}_i \in \Re^M$:

$$\mathbf{x}_i = \ x\ i\ , x\ i+1\ , ..., x\ i+M-1\ \ ^T, 1 \le i \le K, K = N - M + 1\ \ 1 \qquad (18)$$

The vectors x_i thus formed are collected in an augmented multidimensional time series known as the trajectory matrix $\mathbf{X} = \ \mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_\mathbf{k}$:

$$\mathbf{X} = \left[x_{ij} \right]_{i,j=1}^{K,M} = \begin{bmatrix} x\ 1 & x\ 2 & \dots & x\ M \\ x\ 2 & x\ 3 & \dots & x\ M+1 \\ \vdots & \vdots & \ddots & \vdots \\ x\ K & x\ K+1 & \dots & x\ N \end{bmatrix} \quad (19)$$

The trajectory matrix $\mathbf{X} \in \Re^{NxM}$ is a Hankel matrix, that is,

$$x_{ij} = x\ i + j - 1\ , 1 \le i \le K, 1 \le j \le M$$

The SSA decomposition results are conditioned on the window length *M*. An optimal window length is typically chosen such that the longest time scales of underlying system dynamics are captured and also a better separation of the signal from noise is achieved. A heuristic commonly used in nonlinear time series studies is taking the first zero crossing of the autocorrelation function of a time series as the window size *M*. This ensures linear independence between the two successive observations (Abarbanel, 1997). Another approach is to select *M* as the value corresponding to the first minimum of the mutual information (Shaw, 1984).

Step 2: Singular Value Decomposition

A $M \times M$ covariance matrix C_X of X is constructed from the trajectory matrix, i.e.

$$\mathbf{C}_\mathbf{X} = \frac{1}{K}\mathbf{X}^T\mathbf{X} \quad (20)$$

where $K = N - M + 1$ and spectral decomposition of the trajectory matrix is obtained as solution of the eigenvalue problems

$$\mathbf{C}_\mathbf{X}\mathbf{a}_k = \lambda_k \mathbf{a}_k, \quad k = 1, 2, ..., M \quad (21)$$

where a_k and λ_k are the k^{th} eigenvector and eigenvalue respectively. The square roots of the (non-negative) eigenvalues $\sqrt{\lambda_k}$ are called *singular values*, and the set of ordered singular values $\sqrt{\lambda_1} \ge \sqrt{\lambda_2} \ge ... \ge \sqrt{\lambda_M}$ is called the *singular spectrum*, hence the term *singular spectrum analysis*. The ordering implies that

the k^{th} eigenvalue explains at least as much of the variance in the data than the $k+1^{\text{th}}$ eigenvalue.

Broomhead and King (1986a) performed the SVD on covariance matrix obtained from the trajectory matrix $\mathbf{X}$ to obtain the eigenfactors of $\mathbf{C_X}$ Thus by setting $d = \max\ k \in 1,2...,M\ / \lambda_k > 0$ and $\mathbf{v}_k = \mathbf{X}'\mathbf{a}_k \big/ \sqrt{\lambda_k}$, for $k = 1,...,d$ SVD of the trajectory matrix $\mathbf{X}$ can be written as

$$\mathbf{X} = \sum_{k=1}^{d} \sqrt{\lambda_k}\mathbf{a}_k \mathbf{v}_k^T = \mathbf{X}_1 + \mathbf{X}_2 + ... + \mathbf{X}_d \qquad (22)$$

where $\mathbf{X}_k = \sqrt{\lambda_k}\mathbf{a}_k \mathbf{v}_k'$ are bi-orthogonal matrices with rank-one known as elementary matrices Thus rank of $\mathbf{X} = d$.

Similar to PCA, the spectral decomposition of the trajectory matrix $\mathbf{X} \in \Re^{KxM}$ can be written as a product of a score matrix $\mathbf{T} \in \Re^{KxM}$ and a transposed loading matrix $\mathbf{P} \in \Re^{MxM}$ .Thus, in more precise term the trajectory matrix can be expressed as the sum of the outer product of the individual pairs of vectors $\mathbf{t}_i$ and $\mathbf{p}_i$ by setting $\mathbf{P}_i = \sqrt{\lambda_i}\mathbf{v}_i$ and $\mathbf{T}_i = \mathbf{a}_i$:

$$\mathbf{X} = \mathbf{T}_i \mathbf{P}_i^T = \mathbf{t}_1\mathbf{p}_1^T + \mathbf{t}_2\mathbf{p}_2^T + ... + \mathbf{t}_d\mathbf{p}_d^T \qquad (23)$$

Since SSA is simply PCA performed on the trajectory matrix, mathematical and statistical properties of PCA extend to SSA. Specifically, the leadings PCs capture most of the information when variables are highly correlated in the observation space. Furthermore, representing the data by the first few PCs minimizes the approximation error. Also, the first few PCs have minimal entropy with respect to the inputs (assuming data are normally distributed). The number of principal components that should be retained is application specific. For example, a few PCs that explain 60% of variance can be considered adequate for visualization purposes. On the other hand for modeling purposes a PCs explaining a larger proportion of the variance in the data may be necessary (Aldrich and Barkhuizen, 2003; Jemwa and Aldrich, 2006).

Step 3: Grouping of Components

The objective of this step is to separate the additive components of the time series by which the signal is expressed as the sum of intrinsic dynamical components and external noisy components. Moreover through grouping the large part of the information for the signal is compressed by projecting the time series on the subspace of the principal components corresponding to largest singular values (Vitanov et al., 2008).

Each eigenvector from the preceding can be used to construct a time series of length $K = N - M + 1$ by projecting the vectors along each principal direction *k*:

$$\mathbf{t}_k\ t\ = \sum_{j=1}^{M} x\ t + j - 1\ \mathbf{a}_k\ \ j \tag{24}$$

for $t = 1, 2, ..., K$ to obtain principal component scores $\mathbf{t}_k$. These represent the new coordinates of the data in the rotated coordinate space. Hence, for $p < M$ leading components selected to represent the time series, the p-dimensional score vectors of the decomposed matrix $\mathbf{T}$ are given by

$$\mathbf{t}\ t\ = \left[\mathbf{t}_1\ t\ , \mathbf{t}_2\ t\ , ..., \mathbf{t}_p\ t\ \right]^T, \quad t = 1, 2, ..., K \tag{25}$$

There are several ways to group or select the number of leading principal components to retain for extraction of signal features such as trends and oscillations. A common method examines the shape of the eigenvectors. For example, a trend can be extracted by grouping the indices from the eigenvectors that vary slowly. More refined trends or oscillations can be extracted by grouping those indices from eigenvector pairs whose scatter plot resemble a circle or a polygon (Thomakos et al., 2002). Visual inspection of the energy or the percentage contribution of the i^{th} principal component in the analyzed time series can also be used to select components to retain. The percent contribution of an eigenvector (or fraction of the information content explained) is given by the ratio (Tzagkarakis.et.al, 2007)

$$R_i = \frac{\lambda_i}{\sum_{j=1}^{M} \lambda_j}. \qquad (26)$$

This ratio also represents the contribution of the elementary matrix $\mathbf{X}_i$ in the expansion of the trajectory matrix $\mathbf{X}$ (Alonso and Salgado, 2005).

Step 4: Reconstruction

Convolution of a set of principal components $\mathbf{T}$ with corresponding eigenvector or principal directions recovers phase information lost in the preceding decomposition:

$$\tilde{x}\ t+j-1\ = \sum_{k=1}^{p\le M} t_k\ \ t\ \mathbf{a}_k\ \ j\ \ , t=1,...,K \qquad (27)$$

for $t=1,2...,N$ and $j=1,2,...,M$

The elementary matrix described above is no longer a Hankel matrix. However, the reconstructed time series can be approximated by taking the average of the elements on the corresponding diagonals of the elementary matrices obtained in the grouping stage. Through diagonal averaging or Hankelization, the elementary matrix is transformed into a principal component of length *N* to create reconstructed components (RCs) of the original series. The diagonal averaging is performed according to

$$\tilde{x}\ i\ = \begin{cases} \frac{1}{i} & \sum_{j=1}^{i}\sum_{k=1}^{p} t_k\ \ i-j\ \ \mathbf{a}_k\ \ j \qquad 1\le i\le M-1 \\ \frac{1}{M} & \sum_{j=1}^{M}\sum_{k=1}^{p} t_k\ \ i-j\ \ \mathbf{a}_k\ \ j \qquad M\le i\le K \\ \frac{1}{N-i+1} & \sum_{j=1-N+M}^{M}\sum_{k=1}^{p} t_k\ \ i-j\ \ \mathbf{a}_k\ \ j \qquad K+1\le i\le N \end{cases} \qquad (28)$$

where $K=N-M+1$.

A linear decomposition of the original series into p reconstructed components (RCs) contains fitted values of the reconstruction and residual series by the application of diagonal averaging on the elementary matrices (Thomakos et al., 2002). The decomposition of the signal can thus be expressed as

$$\xi_t = z_t + r_t, \quad t = 1, 2, \ldots, N \tag{29}$$

where $z_t = \sum_{j=1}^{p} \xi_t^{\,j}$ represents the p reconstructed components of the original series containing fitted values of the reconstruction and $r_t = \xi_t - z_t$ represents the residual series capturing the error in the reconstruction. In most physical applications z_t is associated with deterministic components (trends) and r_t is associated with stochastic components (noise) in the data. The reconstruction step allows for the separation of underlying signal from the noise.

One of the properties of reconstruction step is that it preserves the phase of the time series, so that $x(t)$ and $\tilde{x}(t)$ can be superimposed on the same time scale, $1 \le t \le N$. This is considered an advantage of using RCs over PCs of length K as these do not contain direct phase information within the embedding dimension *M*. Another property of this step is that no information is lost through the reconstruction procedure since the sum of all individual RCs gives the original time series. If there exist any short oscillation spells in the signal, RCs allow their precise localization in time whereas PCs do not. Hence, a key advantage in using RCs instead of PCs is the recovery of epochs (Rozynski.et.al., 2001).

Figure 10 shows an example of multiscale representation of a signal based on SSA decomposition. The original univariate autocorrelated and noisy time series is shown plotted in Figure10 (a). The decomposition of the original data into five different scales is achieved by using SSA algorithm as described above. The univariate data is embedded using a window size of length five. The fine scale (high frequency) features are plotted in Figure10(c-f). The coarse scale (low frequency) features are plotted in 10b. The coarse scale signals are reconstructed by using those PCs with largest eigenvalues and the remaining components are reconstructed by using those PCs with eigenvalues decreasing in the ascending order. Application of SSA results in the decomposition of a signal into principal signals that capture distinct features in the original signal namely, the deterministic mean and random variations (Tzagkarakiz et al., 2007). The components which carry most of the information content of the original signal

are exactly the first few deterministic ones, which can be predicted more accurately because of their slow-varying nature.

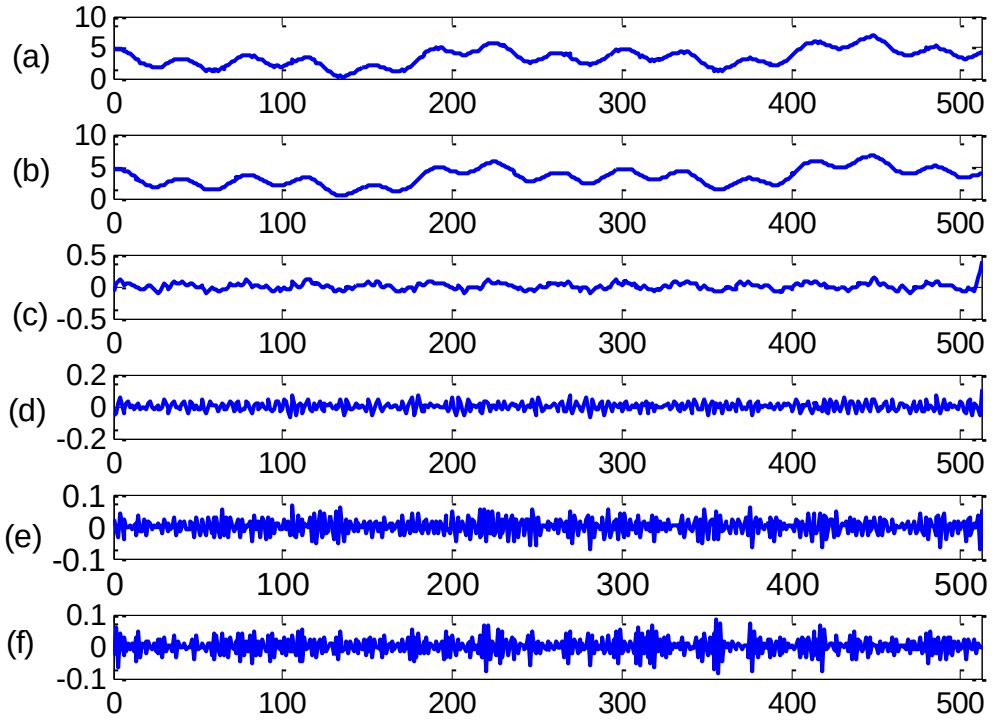

Figure 10 Decomposition of an (a) observed signal into multiple scale representations using singular spectrum analysis. Here a sliding window M=5 was used. The significant or deterministic component (b) explaining the most variation in the signal is associated with the leading eigenelement, with (c)-(f) progressively explaining less variability. In particular, the last signal is associated with high frequency components.

In Figure 10, it is also clear that the basis function that is used in SSA decomposition is adapted to the different features of the monitoring signal. This is because the reconstructed signals at different scales are perfectly reconstructed as shown by Figure10.

SSA has widely been applied in the climatology as well as other geophysical sciences. These applications of SSA include detection of climatic oscillations and regime changes in the amplitude of paleoclimatic time series (Vautard and Ghill, 1989); extracting global warming trends and oscillatory modes from the noisy components of the global surface air temperature time series (Vautard et al.,

1992); climatic oscillations recorded in the Guliya ice core, their relationships with the solar radiation variations and occurrence of the sub-orbital climatic oscillations in the record (Yang, et al., 2006). SSA has also been used in a comparative study on climatic oscillations for extracting relevant trends and oscillations in the time series (Barkhuizen, 2003; Ghil and Yiou, 1996).

A few applications of SSA in biosciences have also been reported. Using EEG (electroencephalogram) data, Mineva and Popivanov, (1996) investigated the dynamics of the single trial readiness potentials in human beings. SSA has also been used in ultrasonic analysis (Pereira and Macial, 2001), later extended to the characterization of the properties of trabecular bones (Pereira et al., 2004). Other related applications include ultrasonic detection and imaging of brachytherapy seeds (Mamou and Feleppa, 2007); study of temporal and spatial variations in shoreline positions to identify characteristic patterns in the shoreline response (Rozynski et al., 2001); analysis of time series with missing data (Schoellhamer, 2001); and signal-to-noise ratio enhancement (Carniel et al., 2006). SSA has also been used in econometrics (Thomakos et al., 2002); biomechanical analysis (Alonso et al., 2005); machine condition monitoring (Alonso and Salgado, 2005; Salgado and Alonso, 2006, 2007; Wang et al., 2001); computer networks behavioral analysis (Tzagkarakis et al., 2007; Wu and Gong, 2000); safety control and monitoring in nuclear power plants (Palomo et al., (2003); system identification in metallurgical reactors (Aldrich and Barkhuizen, 2003); and time series classification (Jemwa and Aldrich, 2006).

3.2 Basic Steps in MS-SSA

As discussed in Chapter 2, traditional process monitoring methods may not perform reliably on data exhibiting multiscale characteristics. Preprocessing of these data to filter high frequency components (noise) can potentially improve the performance of data-based fault diagnostic systems. To this end, integrated feature extraction and process monitoring approaches that eliminate the need for pre-filtering the data have been proposed (Bakshi, 1998). Despite the recent

advancements in wavelet based multiscale methods, the complexity of devising a data adaptive basis function for the optimal multiscale decomposition of process data still remains a challenge in process monitoring tasks. In this section, a SSA based multiscale process monitoring is proposed as an alternative approach to existing spectral methods.

The basic framework of multiscale process monitoring method using SSA is shown in Figure 11. The input data can either be univariate or multivariate with any of the following attributes: stationary or non-stationary; Gaussian or non-Gaussian; random or gross errors; independent or correlated; linear or nonlinear; or, deterministic or stochastic. Similar to other data analysis methods, pre-screening of the input data for outliers or missing data can be considered as an initial step in methodology. The variables are then decomposed into different resolutions (scales) using SSA. The time-domain signal at each scale is subsequently reconstructed preserving the phase of the original time series. After the reconstruction step statistical process control (SPC) techniques can be applied to monitor for significance on the reconstructed data at different scales. Once a fault is detected by the monitoring step fault identification step is conducted to identify cause of the fault at that scale. This is done by using contribution plots of the variables. More detailed discussion of the above basic steps will follow with the proposed multiscale process monitoring method with SSA (MS-SSA) in the next section.

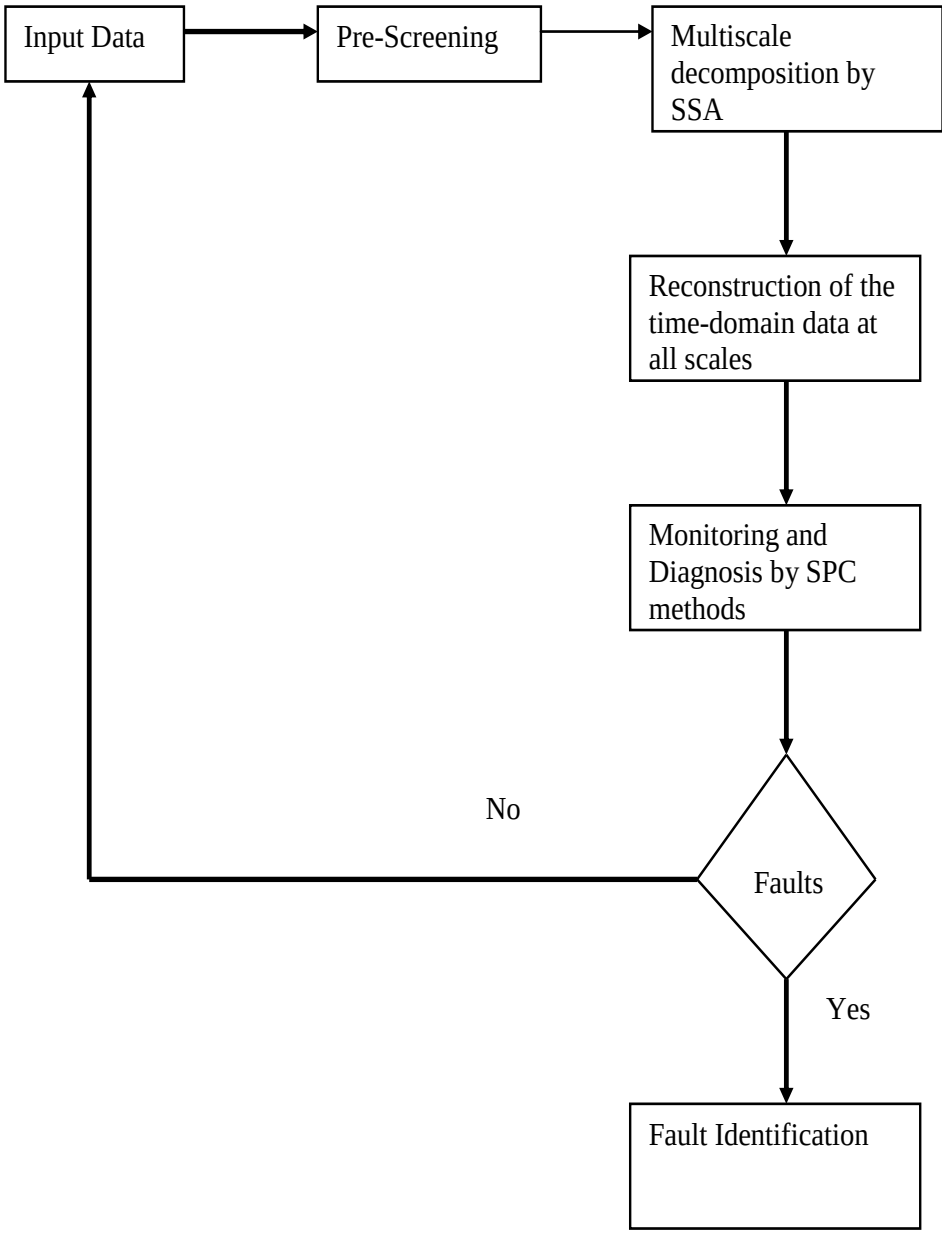

Figure 11: Basic steps in multiscale statistical process monitoring method.

3.2.1 MS-SSA -Methodology

The MS-SSA methodology for process monitoring consists of decomposing each variable by SSA into multiple scales after which a PCA model is developed using the reconstructed variables at each scale, thereby accounting for the correlation between the variables. Control limits for scores and residuals are computed as in

classical multivariate SPM but now for reconstructed multivariate data at each scale based on original data representing normal operation. For new data, a statistically significant change is detected if the scores or residuals of the reconstructed data violate the control limits at any scale.

Multiscale decomposition of the signal by SSA is illustrated in Figure 12 and summarized in the following steps:

1. Decomposition of each process variable into different resolution (multiple scale) is achieved by applying SSA to each variable using an appropriately selected embedding window size, for example, the maximum of the first decorrelation points of the variables under consideration.
2. Group the reconstructed trajectory matrices as a product of principal component scores and loading vectors according to the scale.
3. Reconstruct the scaled versions of the original data matrix by diagonal averaging to obtain the multiscale approximation of the original data matrix.
4. Apply PCA to the reconstructed data matrix, select appropriate number of PCs and determine control limits on the monitored indexes T^2 and Q at each scale.
5. Once a fault has been detected in the above step, identify the variables responsible for the out of control status detected by T^2 and Q. This is done by the implementation of contribution plots of the variable contributing to T^2 and Q statistics at the time when they violate the control limits at the corresponding scales.

These are discussed in detail next.

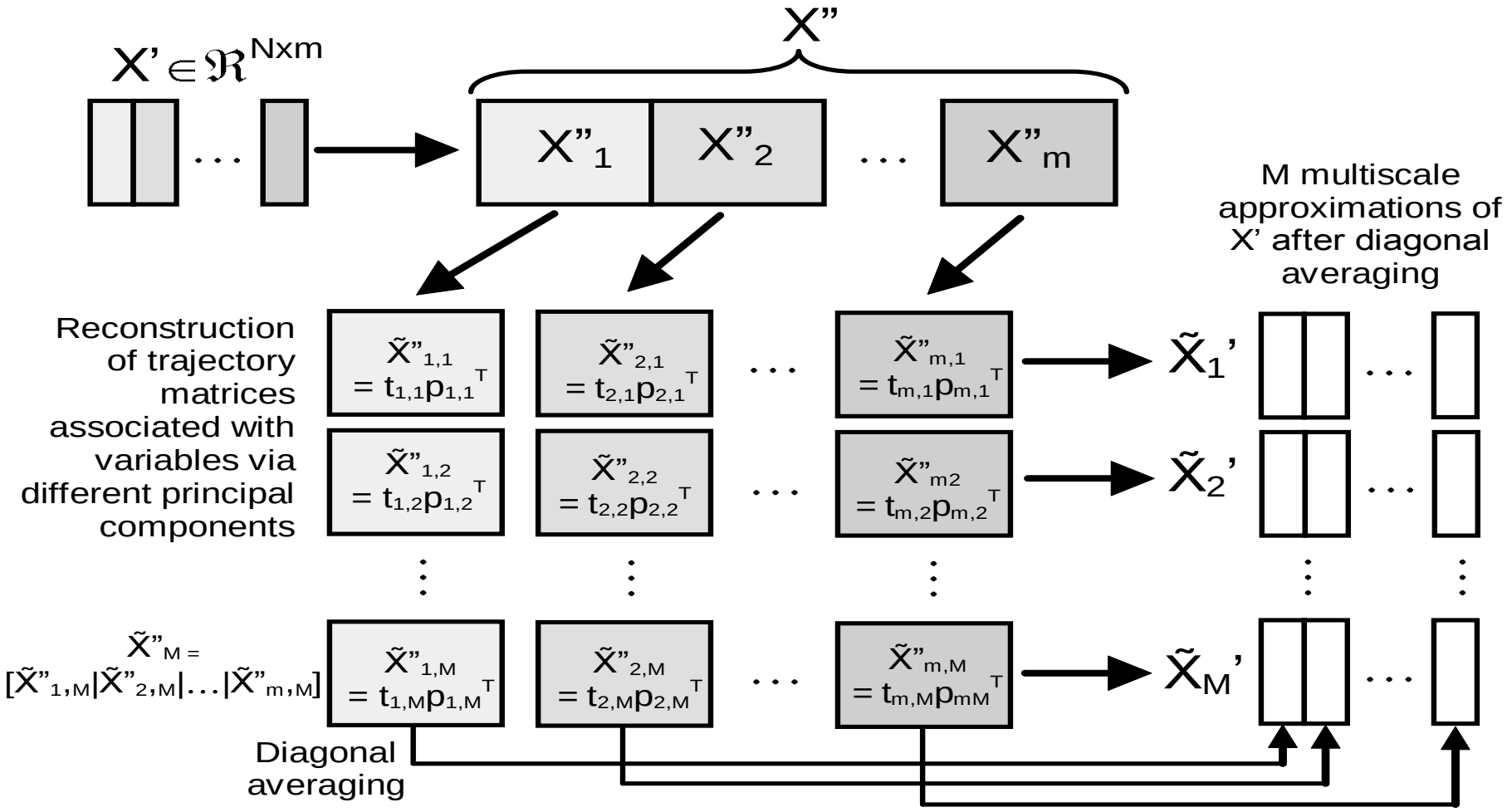

Figure 12 A schematic summary of multivariate data decomposition using SSA-based multiscale resolution analysis.

Step1: Multiscale decomposition of data with SSA

Decomposition of the process variables at different resolution levels can be accomplished by applying SSA to each variable separately using a common window of size *M*. More formally, given N observations on m variables $X^{'} \in \Re^{N \times m}$, trajectory matrix for each variable $\mathbf{X}_k^{''}$ is computed by augmenting each variable $X_l^{'}\ t\ : t = 1,...N\ , 1 \le l \le m$ with M lagged copies of itself using (19). The augmented trajectory matrix thus formed can be represented as:

$$\mathbf{X}^{''} = \mathbf{X}_1^{''}, \mathbf{X}_2^{''}, ..., \mathbf{X}_m^{''} \tag{30}$$

Process signals in general are contaminated by noise of finite length. Hence, choice of the embedding dimension or the window size M is important to ensure that the original system and its reconstruction are diffeormorphically equivalent (Wang et al., 2001). If the window size is too small, the reconstructed signal becomes compressed along the main diagonal or identity line of the embedding space, which results in little information gain. This is called redundance between successive delay coordinates (Casdagli et al., 1991; Wang et al., 2001). Smaller window size may also cause the several neighboring peaks in the spectrum of the data set to coalesce at the coarse resolution (Vautard et al., 1992). For a too large window size the successive delay coordinates become causally unrelated and the resulting reconstruction will not be representative of the true dynamics, a phenomenon referred to as irrelevance (Casdagli et al., 1991; Wang et al., 2001). Moreover, a large window size may mix noise and certain components of time series with a complex structure (Alonso et al., 2005). The structure of the trajectory matrix depends on the choice of M and, therefore, ideally a selected value for M should separate deterministic and stochastic components in the data. However, information regarding signal to noise ratio in observed data is usually not available or difficult to estimate for most systems. In practice, choice of the embedding dimension for SSA decomposition is mainly based on (pseudo-)heuristics, namely first minimum of mutual information criteria and the first zero

of the autocorrelation function (Abarbanel, 1997). The two approaches both aim to choose an optimal window length that captures the global behavior of the system. A primary difficulty in estimating the window size using mutual information is the need to first estimate a probability distribution on the systems states. Incorrect or ill-considered binning may result in poor choice of the embedding window.

In estimating the window size as the first zero crossing of the autocorrelation function for the data one need to ensure linear independence between the two "state" variables, $x\ t$ and $x\ t-M$ (Bray and Wikswo, 2002). This method for the choice of window size is simple. It also gives much more reliable results than the mutual information criterion, since autocorrelation functions are useful for determining residuals. In addition to that it is also useful for detecting periodic components in data and for identifying the dominant power law noise type (white, flicker, random walk, flicker walk and random run) for the particular data type. In the case studies considered later, selection of window length will be based on the first zero of the autocorrelation.

The window length *M* for embedding the data in the delay coordinate space can be selected such that points of different lagged vectors in each variable $X_i'(l), X_i'(k), (l \neq k), 1 \leq i \leq m$ are linearly independent. In this study, the window length M is the first maximal decorrelation point of variables in the process system, or the highest value of M where the sample autocorrelation function of each variable $c_i\ M$ passes through zero for the first time. In mathematical terms, M is the $first.\max\ M / c_i\ M\ = 0$, where

$$c_i(M) = \frac{\sum_{j=1}^{N} (X_i'(j+M) - \bar{X}_i')(X_i'(j) - \bar{X}_i')}{\sum_{j=1}^{N} (X_i'(j) - \bar{X}_i')^2} \qquad (31)$$

and $\bar{X}_i' = \frac{1}{N}\sum_{j=1}^{N} X_i'(j)$ is the arithmetic mean of the corresponding variable.

In some systems this choice of the embedding dimension can give poor results and, as an alternative, the embedding dimension can instead be based on the first decorrelation point or $1/e$ (Schuster, 1988; Tsonis, 1992). Other studies have shown that embedding the data in three or four dimensions gives more reliable results than the above criteria (Michael, 2005) Irrespective of the used criterion, the choice of embedding dimension must ensure that the lag is: (a) large enough for the coordinates to carry as much new information as possible, and (b) small enough for the various coordinates not to be far apart. In this study the choice of embedding dimension based on the first maximum point where the sample autocorrelation is zero gave much more reliable results than choosing some other points as a decorrelation point for the autocorrelation function of the variables in the data.

The trajectory or lagged matrix $\mathbf{X}^{"}$ in (30) can be expressed as the product of a score matrix $\mathbf{T} \in \Re^{KxmM}$ and a transposed loading matrix $\mathbf{P} \in \Re^{mMxmM}$. Thus each variable $X_j^{'}$, $j = 1,2,...,m$ is decomposed by expressing its corresponding trajectory matrix in terms of an ordered series of score and loading vector products using (23).

Where each product $t_{j,i}\, p_{j,i}^{T} = \tilde{\mathbf{X}}_{j,i}^{"}$, for $j = 1,...,m$ and $i = 1,...,M$

Hence $\mathbf{X}_j^{'} = \mathbf{X}_j^{"} = \tilde{\mathbf{X}}_{j,1}^{"} + \tilde{\mathbf{X}}_{j,2}^{"} + ... + \tilde{\mathbf{X}}_{j,M}^{"}, 1 \le j \le m$. (32)

Since eigenvalues are the variance of the signal in the corresponding PC direction, the smallest eigenvalues can be interpreted as noise, while the largest eigenvalue represents the signal components (Mamou and Feleppa, 2007). SSA can decompose the signal without making any assumption about the frequency content of each PC. Therefore the PCs with which the trend was obtained share some frequency bands with the PCs that represent the detrended signal (Golyandina.et.al, 2001). This is contrast to other classical spectral analysis that decompose the signal into components with disjoint frequency spectra (Salgado and Alonso, 2006).

There is no specific criteria for selecting important scales for fault detection and diagnosis, and selection of important scales for monitoring is largely based on heuristic criteria. For example, in the first few scales the percentage of the total variance of the time series concentrated in the selected number of PCs is larger than some predetermined value. These PCs are used to reconstruct the signal in the corresponding scales. Hence the significant amount of information about the data can be extracted by monitoring those scales in most applications (Vitanov.et.al, 2008).

Step 2: Reconstruction of the time domain data at all scales.

To recover the time domain signal corresponding to the transformed elementary matrices of each variable at each scale, the averages along the diagonals of each elementary matrix are computed. The result is different components that represent the original time series in each scale.

Scaled versions of the original data matrix $\mathbf{X}^{'}$ are reconstructed by

$$\tilde{\mathbf{X}}_i^{'} = \left[r_1 \;\; t_{1,i} p_{1,i}^T \quad r_2 \;\; t_{2,i} p_{2,i}^T \quad . \;\; . \;\; . \quad r_m \;\; t_{m,i} p_{m,i}^T \;\; \right], \text{ for } 1 \le i \le M \qquad (33)$$

where $r_k \;\; . , 1 \le k \le m$ are the diagonal averaging functions obtained by convolving the corresponding PCs with eigenvectors (EOFs) in the k^{th} variable. The term $r_k (t_{k,i} p_{k,i}^T)$ represents the k^{th} reconstructed component (RC) in (33) obtained from variable k for the i^{th} resolution (scale). The k^{th} reconstructed component at time i for the variable l can be obtained by using (28), which results in M representations of the data, as in the above Figure 12.

For process monitoring purposes, the M PCs obtained from decomposition of $\mathbf{X}'$ are used as basis vectors to project the test data on the subspace of the PCs corresponding to largest singular values into different levels. In most cases, the significant amount of information about the test data can be compressed in the low-dimensional projection subspace. The PCs corresponding to the large singular values have large amplitude oscillations with low frequencies, where those corresponding to small singular values are small amplitude, high frequency

oscillations. Therefore, information for large-amplitude low periodic components is compressed in the leading PCs while information for small amplitude high periodic components is compressed in the remaining PCs. The projection on the subspace defined by a few leading PCs (with largest singular values) performs a filtering (Vitanov et al., 2008).

Step 3: Monitoring the reconstructed signal using SPC methods

After the decomposition and reconstruction stages the M approximations of the original signal or the selected approximations can be monitored separately using multivariate statistical process control methods, that is, each reconstructed $\tilde{\mathbf{X}}_j'$ is decomposed using PCA as described in section 2.2. Similarly, the appropriate PCs retained at each scale are selected using, for example, percent variance criterion (see equation (6), section 2.2). After the selection of appropriate number of PCs in each scale, control limits on the monitored performance indices such as Hotelling's T^2 and Q statistics are evaluated (Kresta et al., 1991). The process is said to be out of control at a specific scale if the values of T^2 or Q for the reconstructed new data points violate the control limits at that scale.

The significance level (α) is the likelihood of incorrectly assigning an event as abnormal because of the false positive error (Type 1 error). The use of multiple tests increases false positive error (Yoon and MacGregor, 2004). This is because the possibility of finding at least one test statistically significant increases in the case of multiple tests due to chance fluctuations. Thus, for a set of independent tests, the significance level of each test must be adjusted such that the overall significance for all tests taken together equals the nominal value. One solution to the above problem is the use of Bonferroni bounds to adjust the significance values α at each level (Alt and Smith, 1988; Bakshi, 1998; Yoon and MacGregor, 2004). For a M level decomposition, the Bonferroni adjustment is given by

$$\alpha_{adj} = 1 - \left(1 - \alpha_{nominal}\right)^{\frac{1}{M}}. \qquad (34)$$

These adjustments ensure that the risk of incorrectly finding a fault continues to be in the desired significance level, provided each of the tests is statistically independent. This type of adjustment is useful when one uses the monitoring plots only for fault detection. Adjusting the significance level with Bonferroni rule is not appropriate in the case that the monitoring plots are used only for the purpose of fault diagnosis because it is assumed that a fault is already detected (Yoon and McGregor, 2004).

Step 4: Fault identification using contribution plots

As discussed in Chapter 2, contribution plots are used to narrow the possible sources of the fault at those scales where the T^2 and Q plots violate control limits. Otherwise the multiscale monitoring loop in Figure 11 can be repeated for the new set of data. The variables with largest contribution to the respective statistics are considered major contributions to the fault in the case studies of this thesis. It is to be noted that contribution plots may not explicitly identify the cause of an abnormal event, but they determine the entries in the fault conditions that are not consistent with the normal operating conditions due to the "smearing" effect (Qin, 2003). The variables with largest contribution are considered the major contributors to the different fault conditions. In some case studies the size of a variable's contribution under faulty conditions is compared to the size of the same variable's contribution under normal operating conditions as some variables have naturally larger variations (Ralston et al., 2004).

3.2.2 MS-SSA Methodology: An Illustration

To demonstrate the proposed MS-SSA methodology and its potential in process monitoring applications, univariate data generated from a Gaussian distribution with zero mean and unit standard deviation as used in Aradhye et al., (2003) is considered. Abnormal operation is introduced via a shift change of magnitude 5 at time t=30, which persists until t=60 before the process returns to normal operating condition. An embedding window size M=3 was chosen for decomposing the signal using SSA. The control limits at each scale were fixed to

be equal to within three times the standard deviation of the reconstructed signal of the data under normal operating conditions at each scale.

The monitoring results using MS-SSA are shown in Figure 13 for four different times after the mean shift has been introduced. It can be seen that the absolute values of the control limits are scale dependent. The shift at *t*=30 is detected in the first scale in Figure13 (b) at instance of occurrence, which indicates MS-SSA sensitivity to the sudden change in the process mean. SSA-based multiscale method also detects the shift at time t=34, which also shows the nature of MS-SSA (data-adaptive nature) after the shift has persisted for some time. The shift is detected at the first scale in Figure 13(b). This is because the PCs in that scale has greater variance compared to PCs in other scales and hence carry most of the variations in the data. This shows that the first few scales can give some reliable information on the occurrence of the fault in MS-SSA. The plots at t=61 and t=64 again show the behavior of MS-SSA when the process has returned to normal operations. At t=61 the first two scales continue to violate the detection limit, but the return to normal operation is picked up at t=64 in the first scale of Figure 13(b). These plots show that MS-SSA is effective in detecting mean shift in the first two scales.

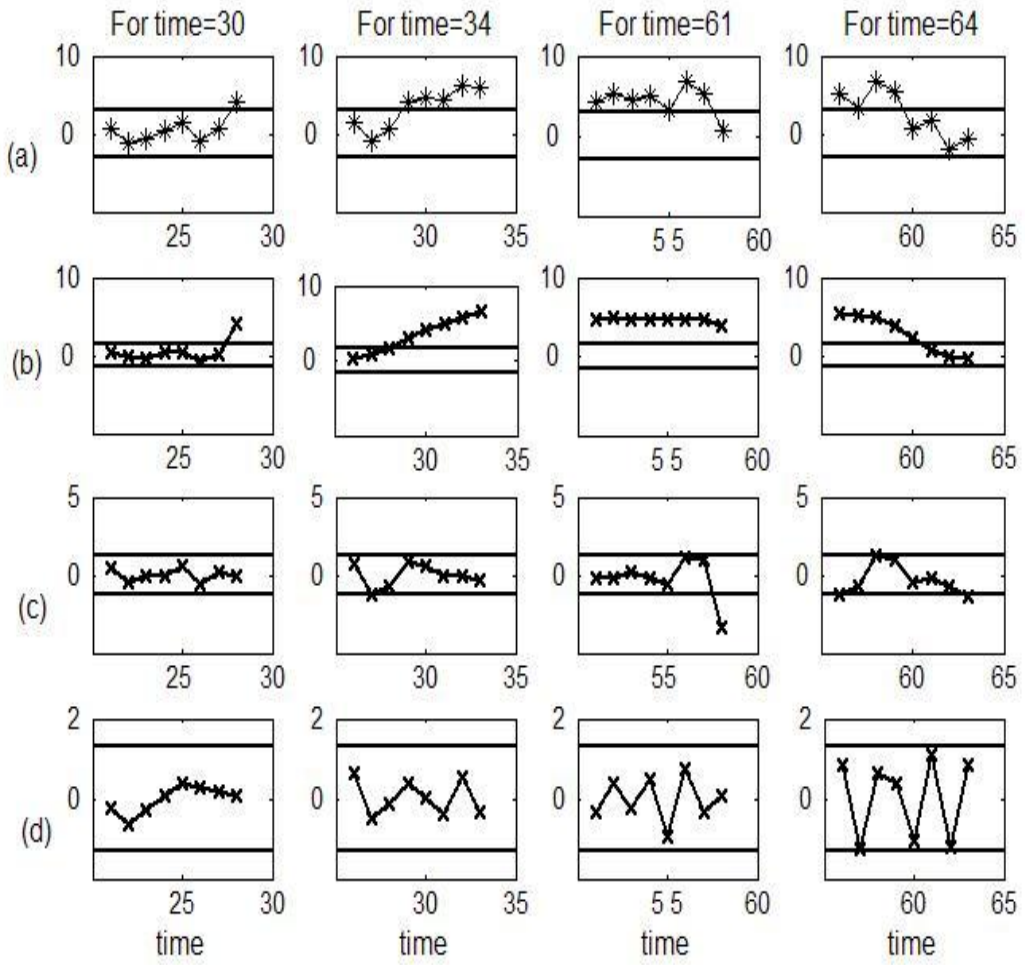

Figure 13.Illustration of MS-SSA methodology. (a) Original signal used for monitoring measurement at selected time, (b), (c), (d) represent reconstructed signals at scale 1, 2, and 3 respectively.

For comparative purposes, a multiscale analysis using MSPCA was performed as above using Haar wavelet with a decomposition level of 3 and the results are shown in Figure 14. The finest scale detects the shift when it occurs for the first time at t=30 as shown by in Figure14 (d). Figure14 (e) shows the reconstructed signal with the coefficients outside the limits and the limit at that particular time, *t*=30. The reconstructed signal and the values of the limit at t=30 in Figure14e form the corresponding points in Figure14d. The Figures for t=34 show how MSPCA detect the shift after the shift has persisted for some time. The shift at t=34 is detected by the wavelet and the last scaled coefficients as shown in Figure14 b and a. The detection limits at different time as shown in Figure14e is

computed based on variance of normal data at the selected scales. In MSPCA only the last scaled signals detects the shift when it persists for some time, while in MS-SSA the signals at the first scale detects the shift as indicated in Figure13b at t=34 when it persist for some time. The plots at t=61 and 64 illustrate the behaviour of MSPCA when the process has return to the normal operation. The last scale signal at t=61 in Figure14a continues to violate the limit as the process return to normal. But the signal at the finest scale in Figur14d and the reconstructed signal in Figure14e at t=61 show that the process has return to normal. Finally the last scaled signal at t=34 in Figure14a stops violating the limit and the plots show that the process has gone back to the normal operation (Aradhye et al., 2003).

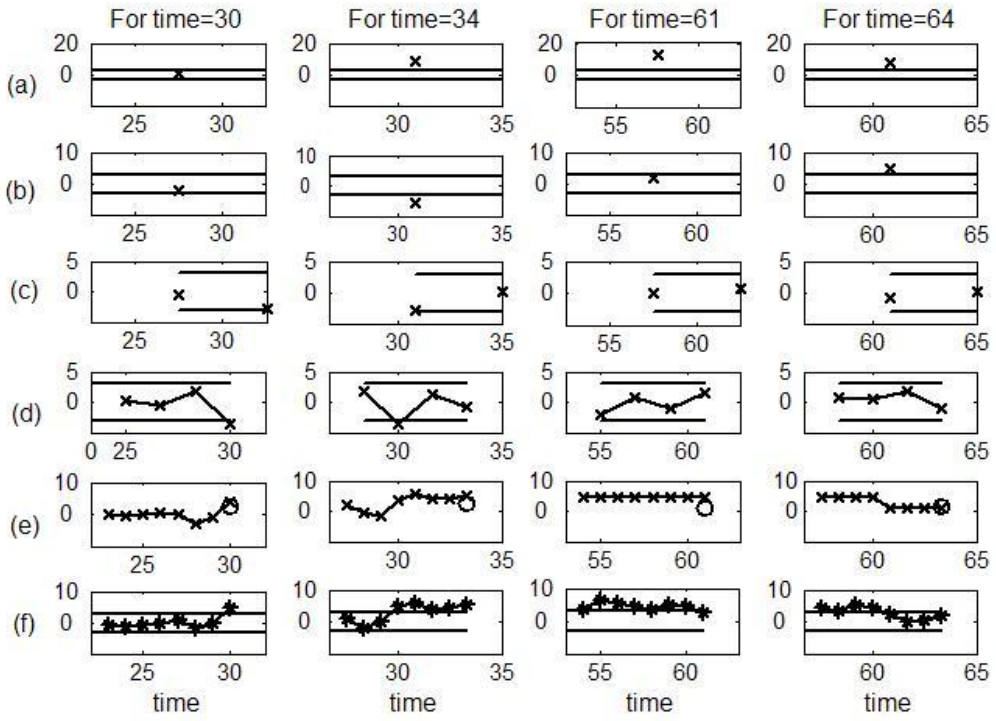

Figure 14 Illustration of MSSPC methodology.(a), represents wavelet coefficients at m=4,(b) represents wavelet coefficients at m=3 ,(c),represents wavelet coefficients at m=2 (d) represents wavelet coefficients at m=1, (e) represents reconstructed signal and detection limit(O) corresponding to selected time and (f) represent original signal used for monitoring measurement at selected time.

Chapter 4 Evaluation of MS-SSA Process Monitoring Methodology

In this chapter the MS-SSA methodology as proposed in the previous chapter is evaluated by means of four case studies: (i) a simulated linear autocorrelated multivariate system, (ii) a simulated 2x2 system process previously studied by Kano et al., 2002 (iii) the Tennessee Eastman Challenge problem used as a benchmark in many plant-wide process control studies (Downs and Vogel, 1993), and, (iv) industrial data from a milling circuit of a South African precious metals mining concern. The performance of the MS-SSA approach is compared against classical multivariate statistical process control based on PCA as well as wavelet-based multiscale process monitoring. The effect of choice of the embedding dimension in the use of MS-SSA is also investigated using data from the first case study.

4.1 Case Study I: A Simulated Multivariate Linearly Autocorrelated Process

In the first case study data generated from a simple linear autocorrelated multivariate process system with three process variables are considered. Fault conditions are introduced as described below. Also, different datasets of the system are generated for different autocorrelations and noise levels to study the effect of these parameters on the technique.

Denoting by $N(\mu,\sigma)$ a Gaussian noise model with a mean of μ and standard deviation σ, let $x(t)$ be a normally distributed variable sample at time t with zero mean and unit standard deviation, i.e. $x(t) \sim N(0,1)$. Three "observed" process variables for the simulated system, $x_1(t)$, $x_2(t)$ and $x_3(t)$, are generated according to the following equations:

$$x_1(t) = 0.9x(t) + \varepsilon_1(t), \quad \varepsilon_1(t) \sim N(0, 0.01), \tag{35a}$$

$$x_2(t) = 0.5x(t) + \varepsilon_2(t), \quad \varepsilon_2(t) \sim N(0, 0.01) \tag{35b}$$

$$x_3(t) = x_2(t) + \varepsilon_3(t), \quad \varepsilon_3(t) \sim N(0, 0.02). \tag{35c}$$

From the data generating equations, it is clear that variables x_2 and x_3 are highly correlated while variable x_1 has the most significant autocorrelation. The strong autocorrelation in equation 35a implies a strong time-dependence in the signal. The multivariate n × 3 data matrix $\mathbf{X}_0$ describing normal operating behaviour is given by,

$$\mathbf{X}_0(t) = \left[x_1(t) \quad x_2(t) \quad x_3(t) \right] \tag{36}$$

Three abnormal or fault conditions are induced by changing (i) the autocorrelation of x_1 to ± 0.5 and (ii) the standard deviation of the noise in x_3 to 0.1. In addition, control test data in which no fault was present was also generated, that is Case 0. The settings of the above abnormal conditions are summarized in Table1. A total of 100 data sets were generated in each of the three test cases.

Table 1 Settings of fault conditions.

Case	**Fault Type**	**Fault Size**
0	Normal Condition	N/A
1	Change of autocorrelation of x_1	$0.9 \rightarrow \pm 0.5$
2	Change of the standard deviation of the noise in x_3	$0.02 \rightarrow 0.1$

Typical plots of the variables for both normal and fault conditions are shown in Figures 15-17 for the three test cases

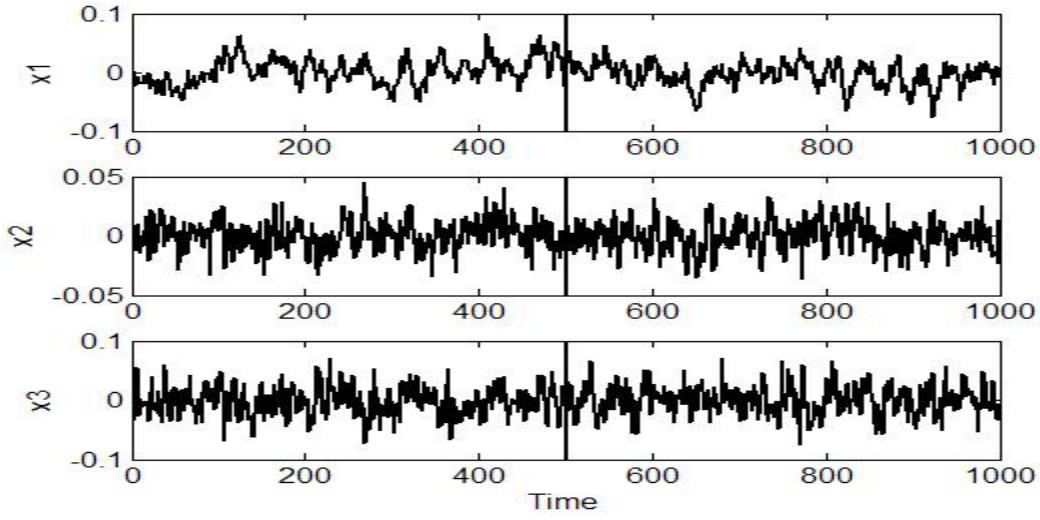

Figure 15 Time series plots of observed data during normal and fault conditions (case 0 in Table 1). The vertical line at time = 500 indicates change point from normal to the faulty conditions,

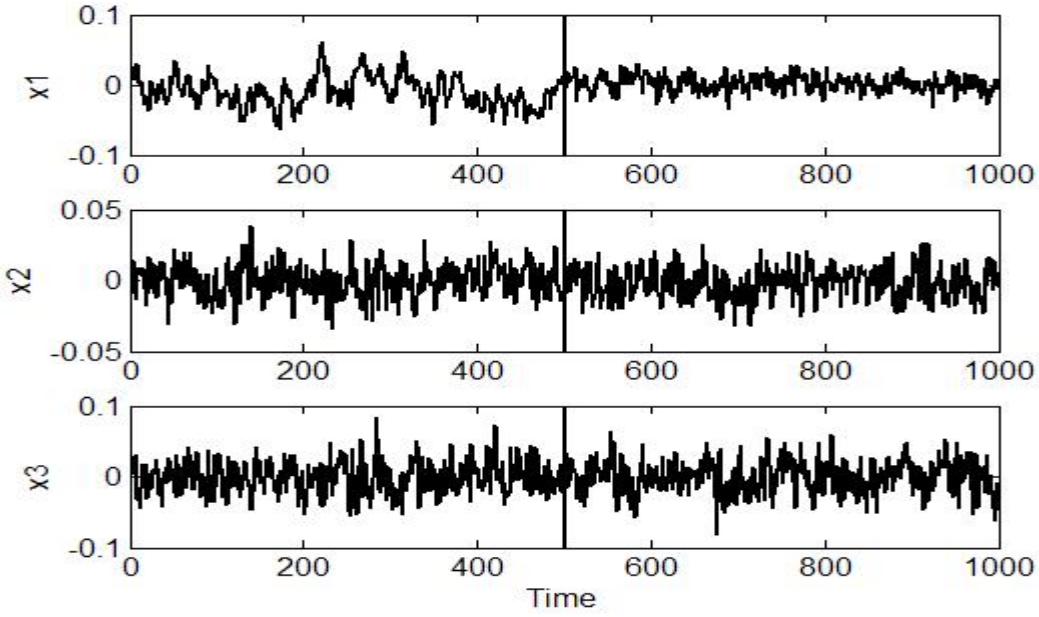

Figure 16 Time series plots of observed data during normal and fault conditions (case 1 in Table 1). The vertical line at time = 500 indicates change point from normal to abnormal conditions.

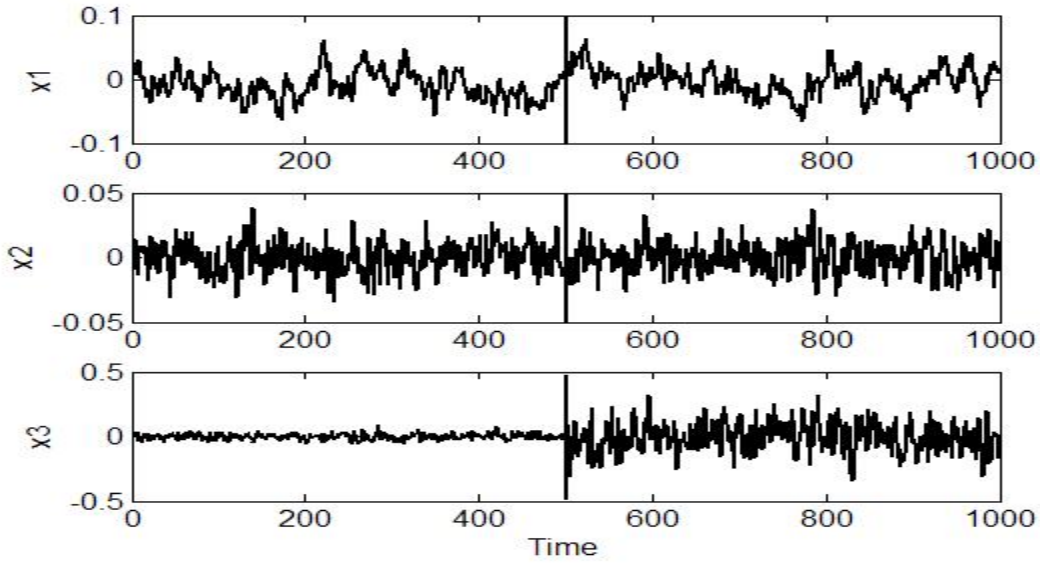

Figure 17 Time series plots of observed data during normal and fault conditions (case 2 in Table 1). The vertical line at time = 500 indicates change point from normal to abnormal conditions.

MS-SSA and conventional PCA were evaluated over a set of 100 realizations for each of the three test cases. The average reliability index computed across the 100 simulations was used for performance evaluation (Kano et al., 2002). The reliability index is defined as the fraction of test samples violating the control limit.

4.1.1 Results: cPCA

The spectral information from the PCA model for the normal operating condition is shown in Figure 18, from which two PCs explaining about 71% of the total variance where retained. The control limits for both T^2 and Q statistics were set at 95% confidence level and are shown in Figures 19-22. The results are summarized in Table 2.

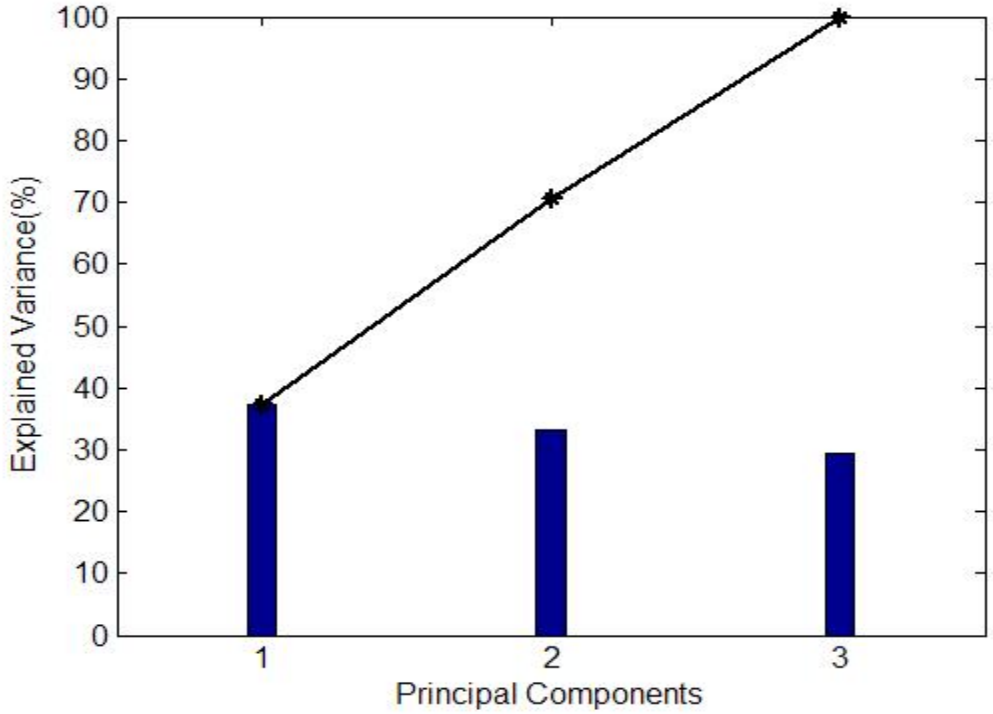

Figure 18 Bar plot shows the percentage of variance explained by PCs in normal operating conditions in PCA. A line plot shows the cumulative variance explained by respective PCs.

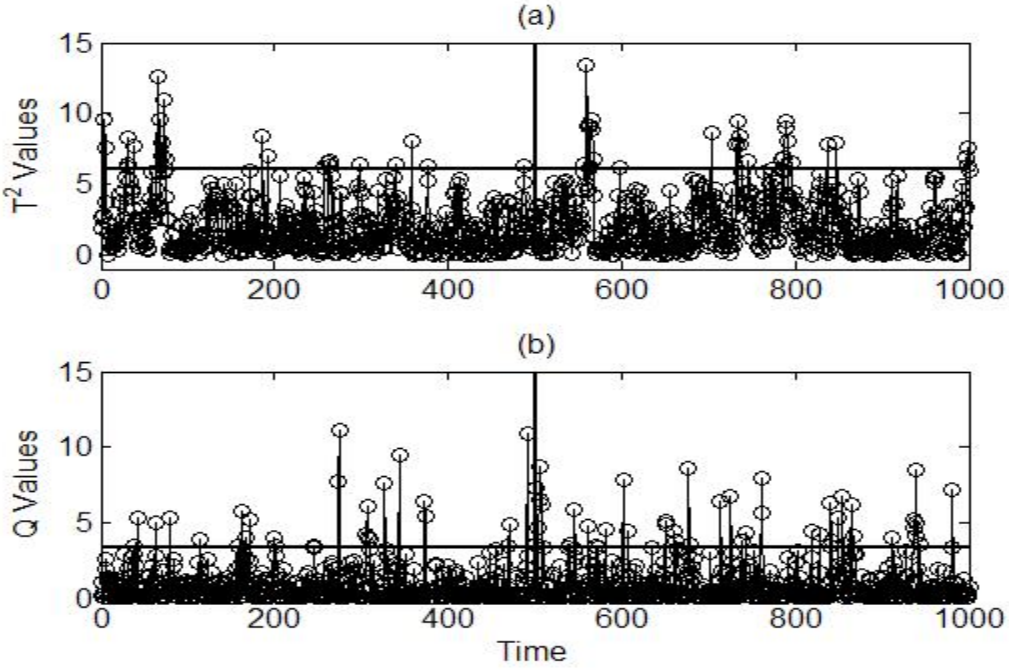

Figure 19 Conventional PCA: (a) Hotelling's T^2 and (b) Q statistics for Case 0 with 95% confidence limit during normal and abnormal conditions (shown separated by a vertical line at time=500).

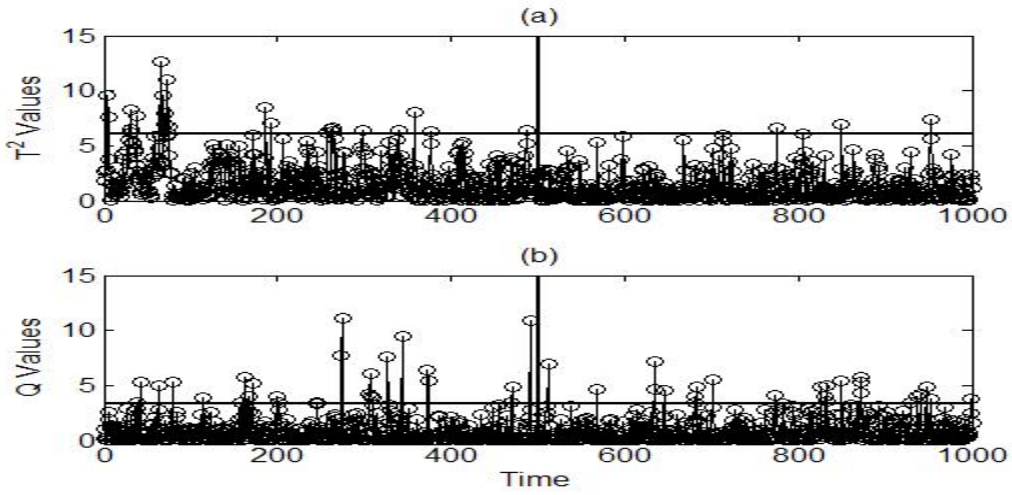

Figure 20 Conventional PCA: (a) Hotelling's T^2 and (b) Q statistics for Case 1 with 95% confidence limit during normal and abnormal operating conditions shown separated by a vertical line at time=500. The autocorrelation of x_1 in normal condition is 0.9 and that in abnormal condition is 0.5.

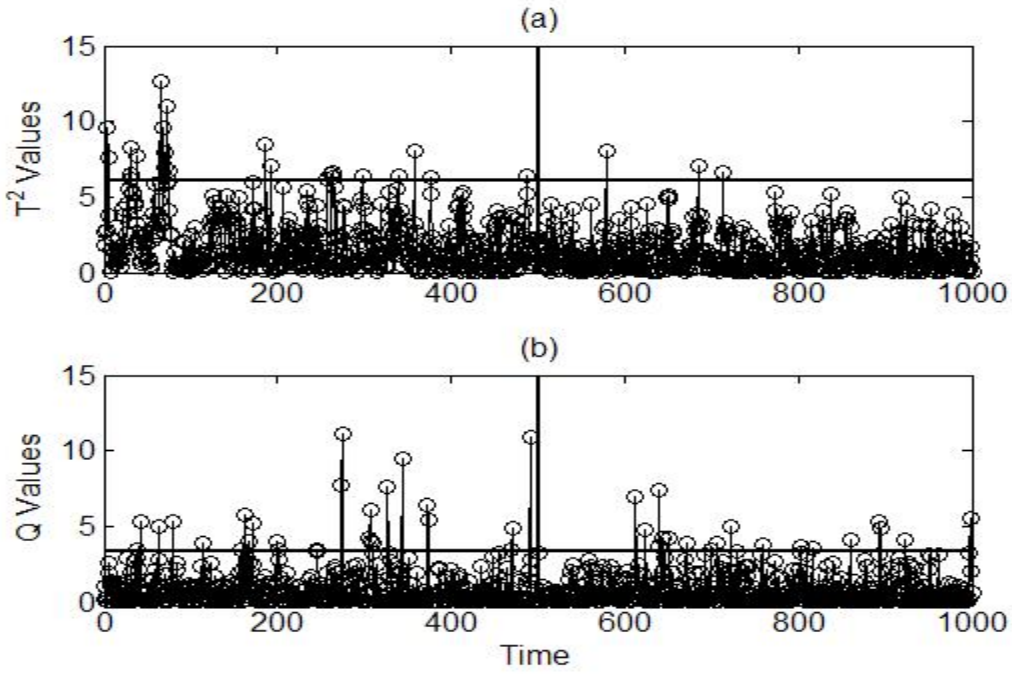

Figure 21 Conventional PCA: (a) Hotelling's T^2 and (b) Q statistics for Case1 with 95% confidence limit during normal and abnormal operating conditions shown separated by a vertical line at time=500. The autocorrelation of x_1 in normal condition is 0.9 and that in abnormal condition is -0.5

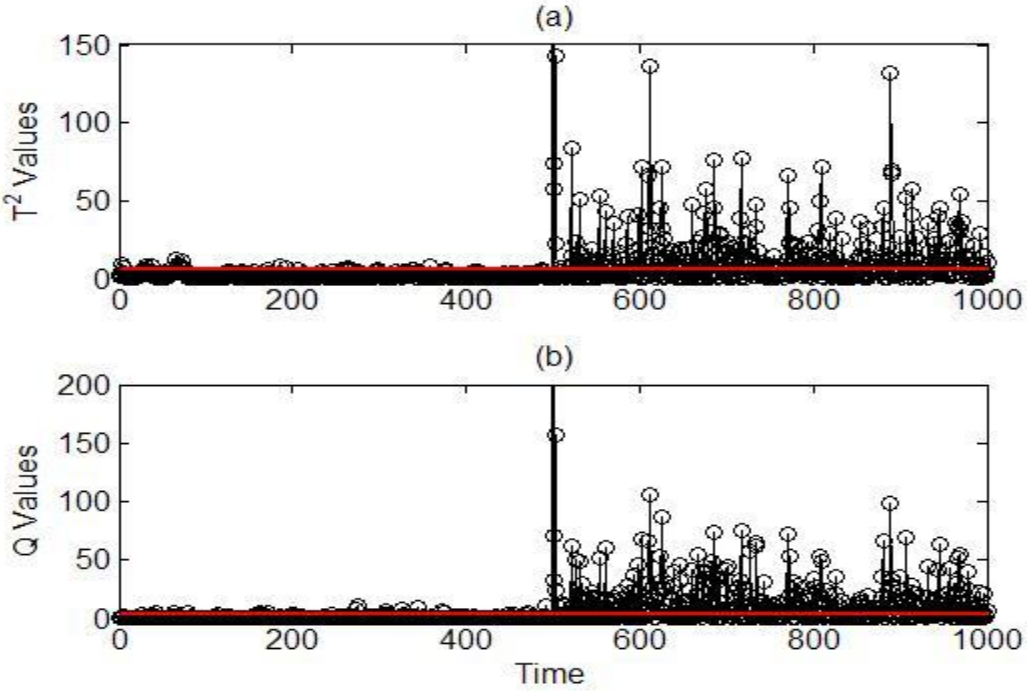

Figure 22 Results for cPCA: (a) Hotelling's T^2 and (b) Q statistics for Case 2 with 95% confidence limit superimposed .during normal and abnormal operating conditions shown separated by a vertical line at time=500. The standard deviations of the noise level in x_3 under normal and abnormal conditions are 0.02 and 0.1 respectively.

The reliability measure in Case 0 is close to 5%, which indicates that the control limits are successfully determined in conventional PCA approach. As discussed in Chapter 2, cPCA is best suited for analyzing steady state data and can give misleading results when data are correlated, as is apparent in Figure 20 for example. In this case only a few samples violated 95% confidence limit in both the T^2 and Q statistics, although changes were induced in the variables. Therefore, both T^2 and Q charts of PCA fail to detect the change in autocorrelation of variable one.

In the second case (case 2), the % reliability increased as the noise level in the data increased. Moreover, data points are highly correlated in time in this case and, hence, the application of static PCA to such systems can give misleading information. Since the data points are autocorrelated the selected PCs for monitoring statistics in conventional PCA also exhibit significant autocorrelation.

This leads to a violation of control limits in T^2 chart of PCA. The control limits calculated for T^2 chart based on the time independency assumptions are not valid in this case study. In general, the use of conventional MSPC methods in data exhibiting significant autocorrelation (and, therefore, in violation of the time independence assumption) can be unreliable, resulting in false alarms although the process is within expected normal operating regime.

4.1.2 Results: MS-SSA

The first step in MS-SSA is to create a reference model by applying linear SSA to the normal data $\mathbf{X}$. The parameter *M*, the window length for embedding the data in the delay coordinate space, is selected based on the first point of maximal decorrelation of variables in normal operating condition (NOC) using sample autocorrelation functions of the variables as discussed in Chapter 3. The autocorrelation functions of the three variables under NOC are shown in Figure 23 as functions of time delay. The window length is selected according to

$$\begin{aligned} M &= \max\ M_{x_1}, M_{x_2}, M_{x_3} \\ &= \max(5,5,19) \\ &= 19 \end{aligned}$$

where M_x is the first point of zero autocorrelation for variable *x*. Note that the decay rate of the autocorrelation functions is dependent on the noise levels as well as the strength of the autocorrelation.

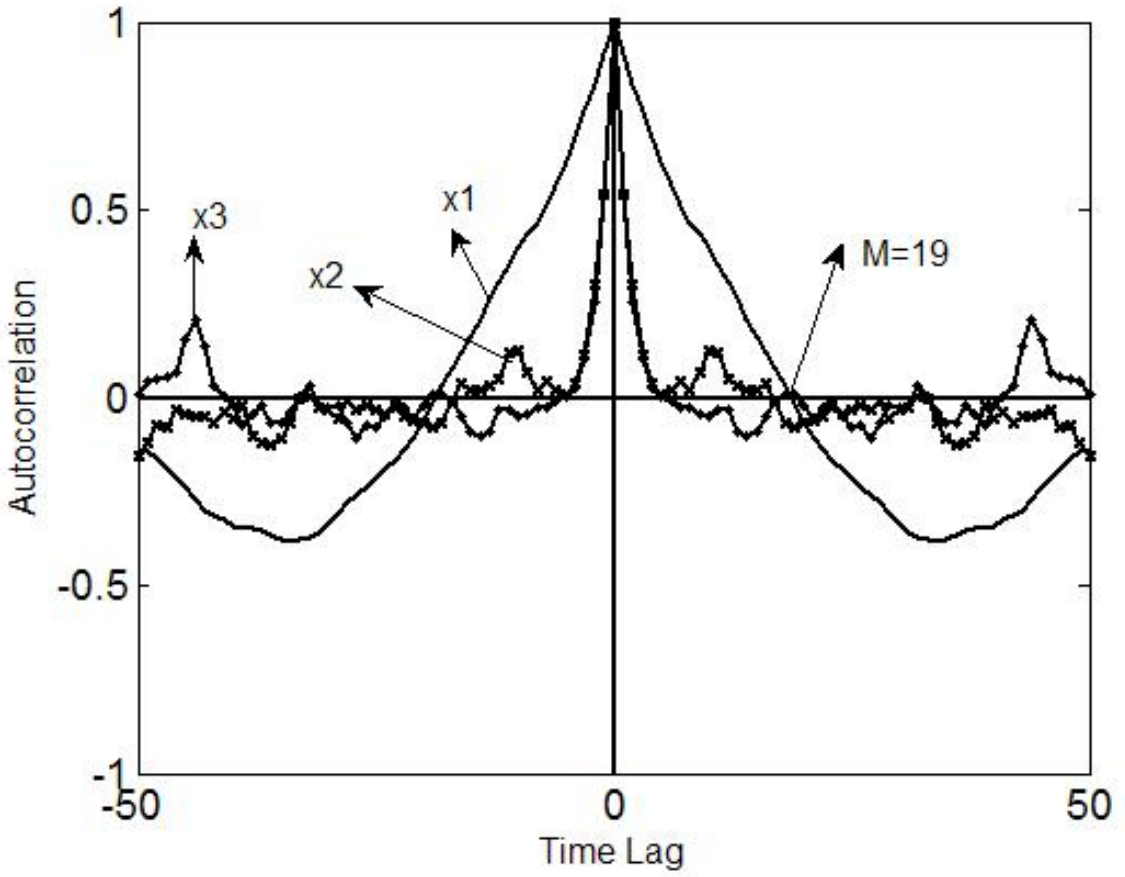

Figure 23 Sample autocorrelation functions of the variables The first maximal decorrelation point is 19.

On the basis of the chosen window size, a 19-level multiscale representation of each variable in $\mathbf{X}$ is obtained and is shown in Figure 24(a), (b) and(c). Using reconstructed data at each level, PCA models were built by retaining the first two PCs, each such model explaining at most 70% of total variance in the reconstructed data, Figure 25.

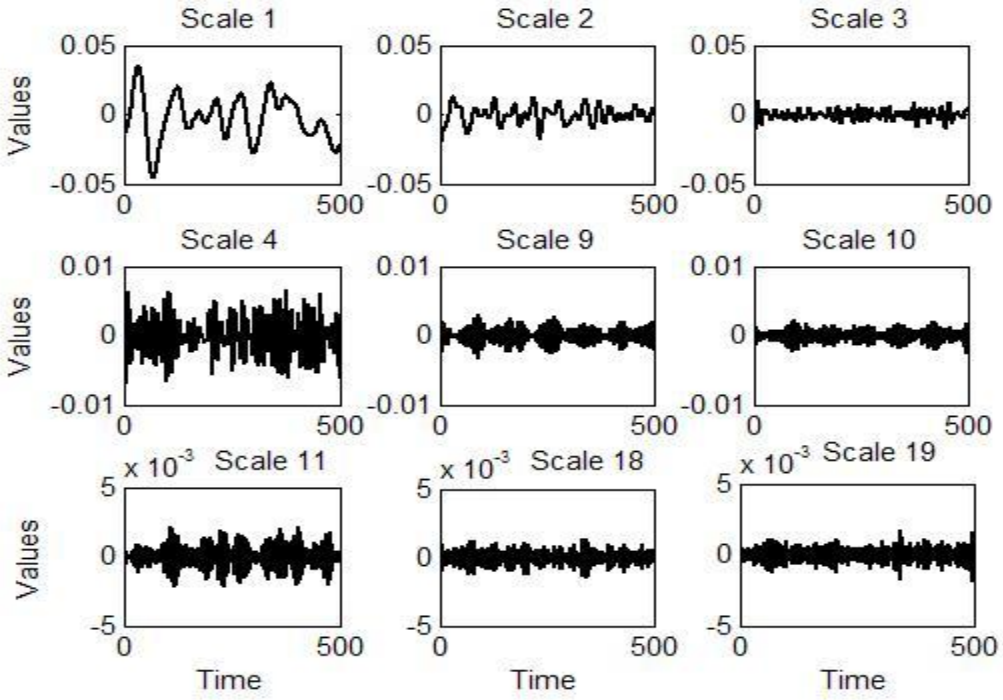

Figure 24(a) Multiscale representation of x_1 in normal operating condition using MS-SSA

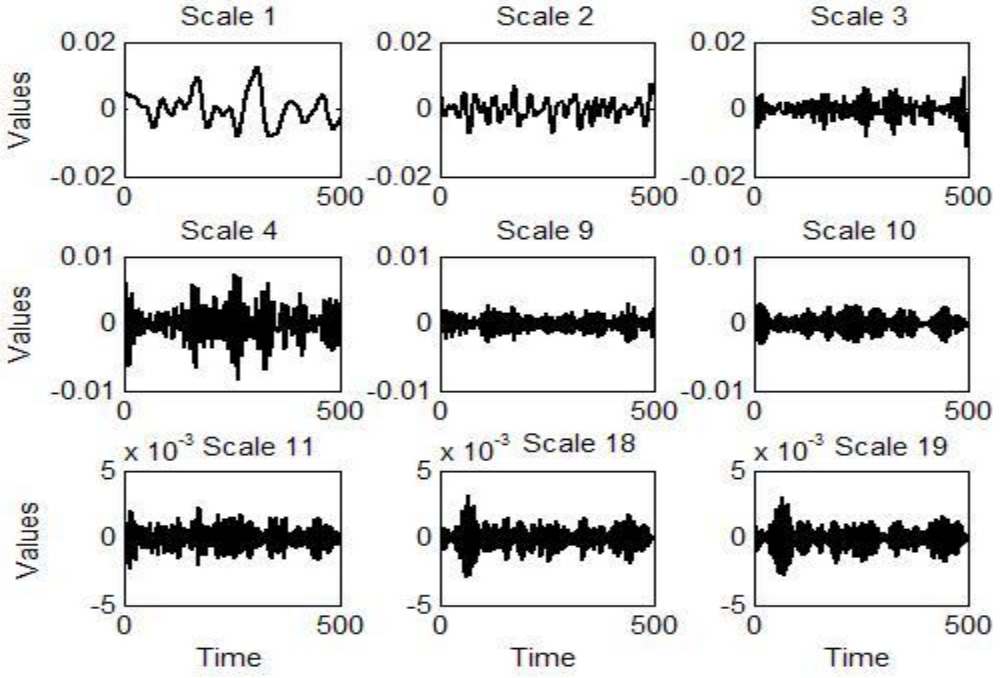

Figure 24(b) Multiscale representation of x2 in normal operating condition using MS-SSA

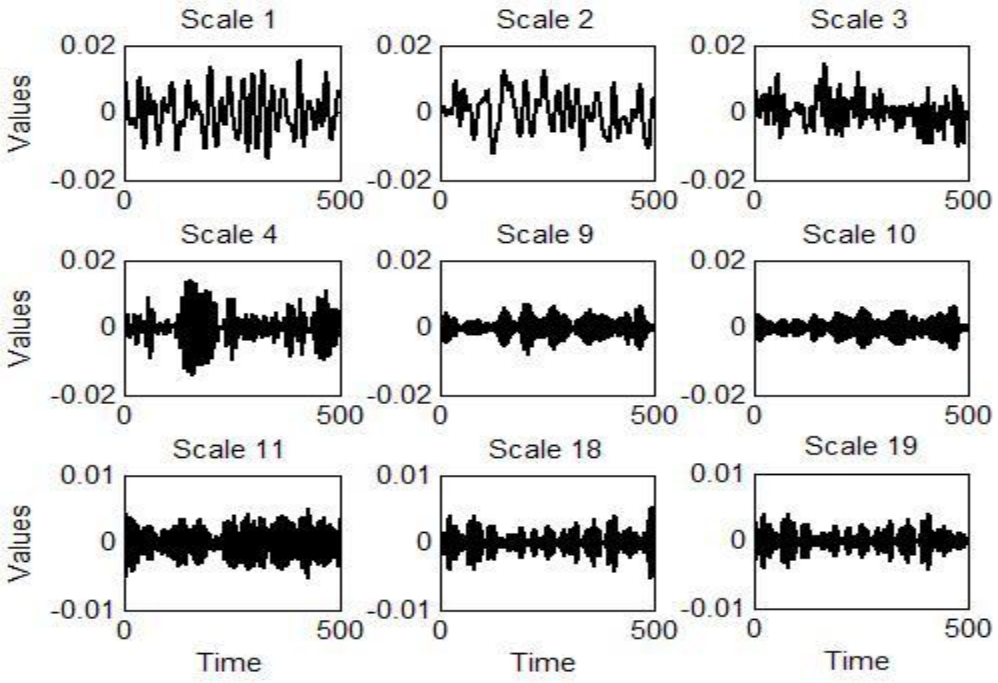

Figure 24(c) Multiscale representation of x3 in normal operating condition using MS-SSA

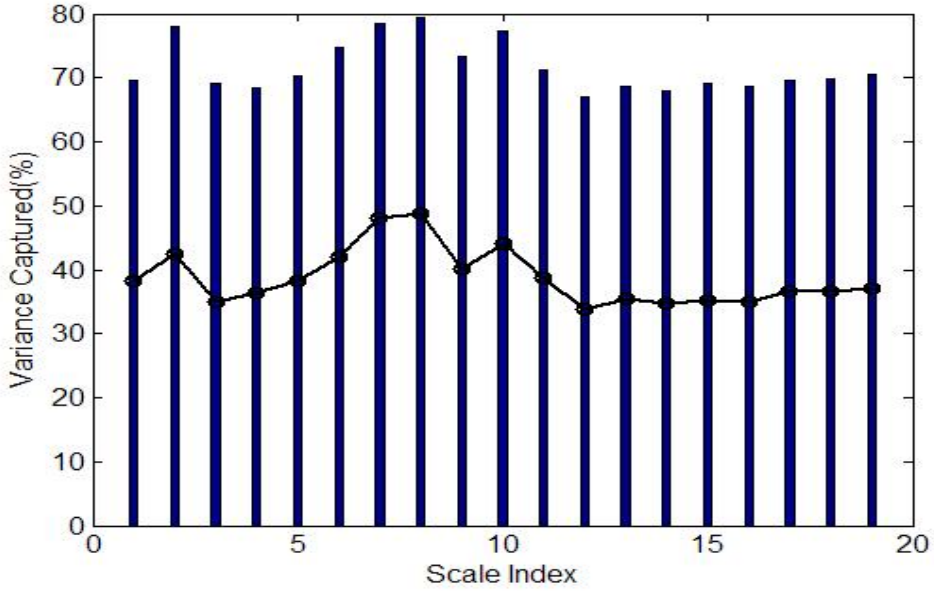

Figure 25 Barplot shows the percentage of variance captured by first two PCs in all scales and the line plot shows the percentage of variance captured by first PCs in all scales.

The Bonferroni adjusted significance level α_{adj} at each scale is evaluated as 0.0026 $=1-(1-\alpha)^{1/19}$, where the target overall significance level α is 0.05. Because the adjusted significance is less than the target, the upper control limits of T^2 and Q statistics in MS-SSA at each scale are thus higher than in cPCA case.

For the test data in each case, the multiple scales are obtained by projecting the data onto the respective loadings obtained from the normal data with SSA and the monitored T^2 and Q statistics are computed as described in the MS-SSA procedure in Chapter 3. For the purpose of comparsion, the wavelet-based MSPCA approach was also evaluated using the Haar wavelet as the basis function and a decomposition level of 4. In each of the multiscale methods, the control limits were set using the Bonferroni adjustment. The performances of MS-SSA and conventional PCA are evaluated on the basis of % reliability as shown in Figures 26 (a-d) and 27 (a-d) respectively, and also summarized in Table 2 for comparison. These figures display the spread of variance in the data as well as the presence of outliers in each fault case compared to that in a normal operating condition. The width of boxes in each case provides the visual estimation of the expected range of data compared to the normal operating condition. The numerical values in Table 2 also confirm the median values of the reliability percentage in each case as indicated by the line in the box plots.

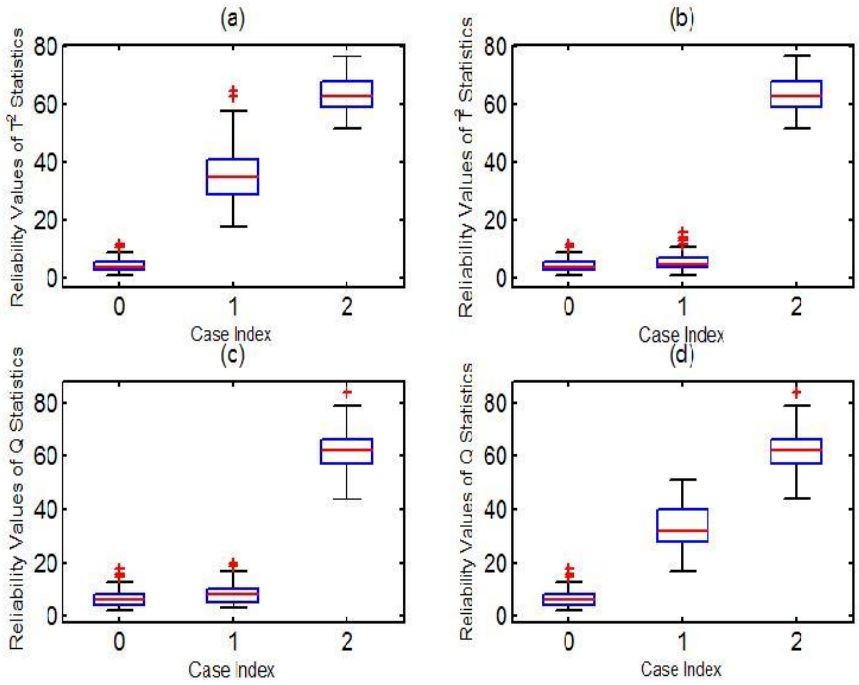

Figure 26 (a-d) Box plots showing the reliability of T^2 *and* Q *statistics in the performance of MS-SSA in all cases. In Figure 26(a) and (d) the autocorrelation of x1 in case1 is -0.5 and in Figure 26(b) and (c) the autocorrelation of x1 in case1 is 0.5.*

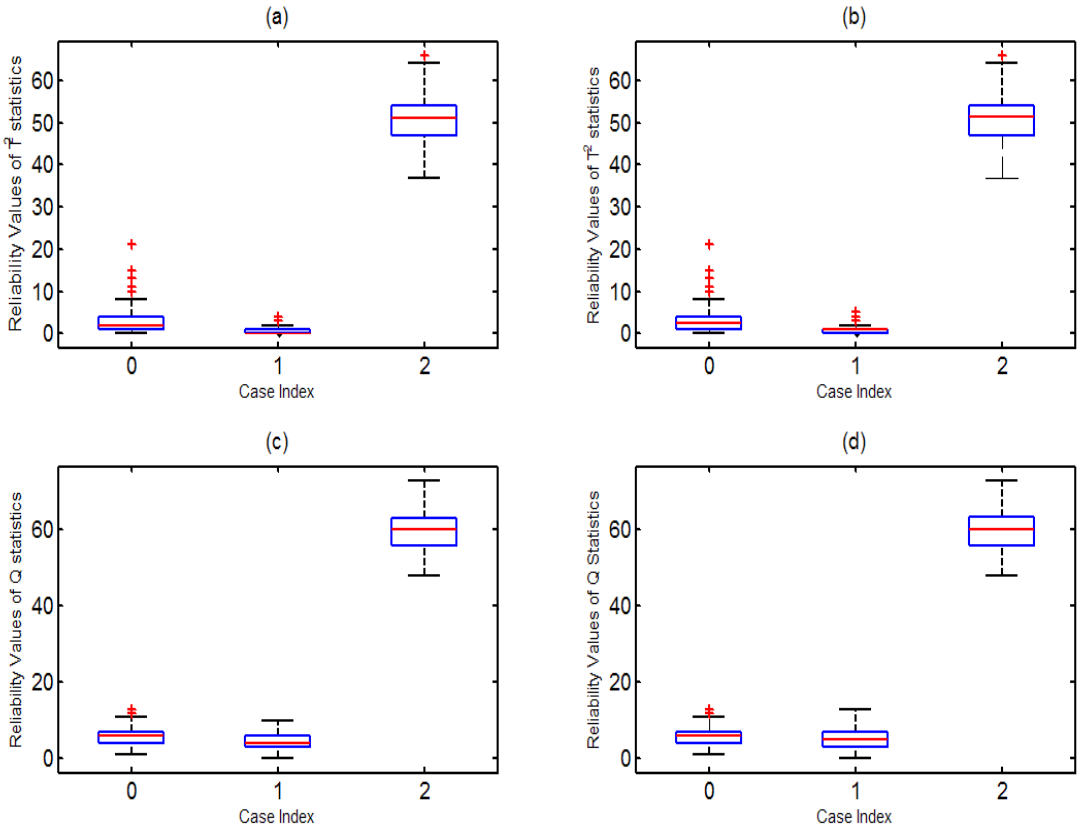

Figure 27 (a-d) Box plots showing the average reliability of T^2 statistics in the performance of cPCA in all cases. The autocorrelation of x1 in case1 is 0.5.

Table 2. Reliability (%) of MS-SSA, PCA and MSPCA

Method	**Monitoring Statistic**	**Case 0**	**[3]Case 1**	**Case 2**
MS-SSA	T^2	4.5	5.9/35.9	63.5
	Q	6.5	8.6/33.7	60.9
PCA	T^2	3.5	0.7/0.9	51.3
	Q	5.4	5.1/5.1	61.4
MSPCA	T^2	2.6	1.2/2.3	51.9
	Q	3.4	0.12/2.4	51.8

3 Reliability (%) of monitoring statistics in Case 1 when the autocorrelation of x1 is changed to 0.5 and -0.5.respectively from 0.9.

The control limits as evaluated in Case 0 for both MS-SSA and MSPCA compare well with the expected false alarm rate (of 5%), since the reliability measures are close to 5%. Hence the multiscale models of MS-SSA and MSPCA developed in the normal operating conditions can be expected to detect the corresponding faults in Case 1 and Case 2. In these latter cases, MS-SSA performs better than cPCA as well as the comparable MS-PCA approach, especially for detecting changes in autocorrelation of x_1 in Case 1.The generally better performance attained using MS-SSA compared to MSPCA can be attributed to use of data adaptive basis functions instead of a fixed mother wavelet.

The contribution plots in Figures 28(a-d) and 29(a-d) show the variable contributions to values of the T^2 and Q statistics for the different fault cases, as explained in the MS-SSA methodology. Corresponding contribution plots for classical PCA-based MSPC are shown in Figure 30(a-d). In Case 1, Figure 28(b) shows that the size of the contribution of x_1 has chnaged in both charts compared to respective variables in normal condition in most of the cases. In the last few scales contribution of x_1 has significantly increased in both charts. In Case 2, Figure 28(d) shows that x_3 has the highest contribution to the T^2 and Q values. This is possibly because of the fault that is introduced in Case 2 causing x_3 to deviate more from other variables. In case 2, data was corrupted with high frequency noise that is expected to appear in the high frequecy scales of Q statistics in MS-SSA model as shown in the Figure 29(d). It has to be noted that these contribution plots may not identify the exact fault when there is a strong correlation among the process variables(Qin, 2003). In that case drawing conclusions from these dagnostic method may give misleading information on the fault condition.

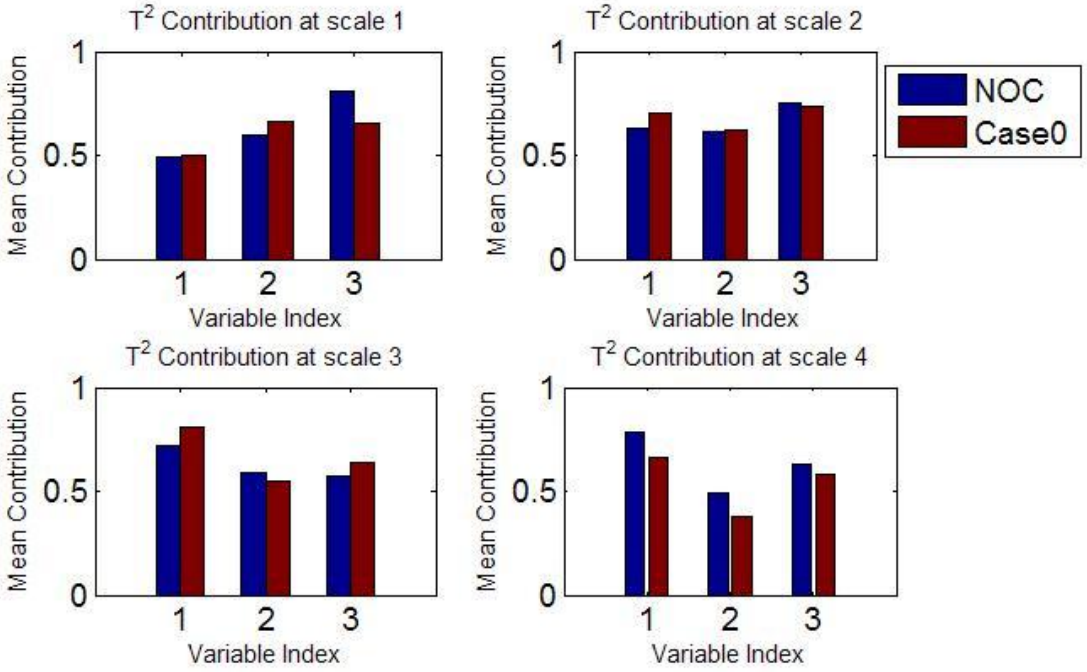

Figure 28 (a) Variable contributions to T2 statistics for case 0 indicating that all the variables are operating under normal condition.

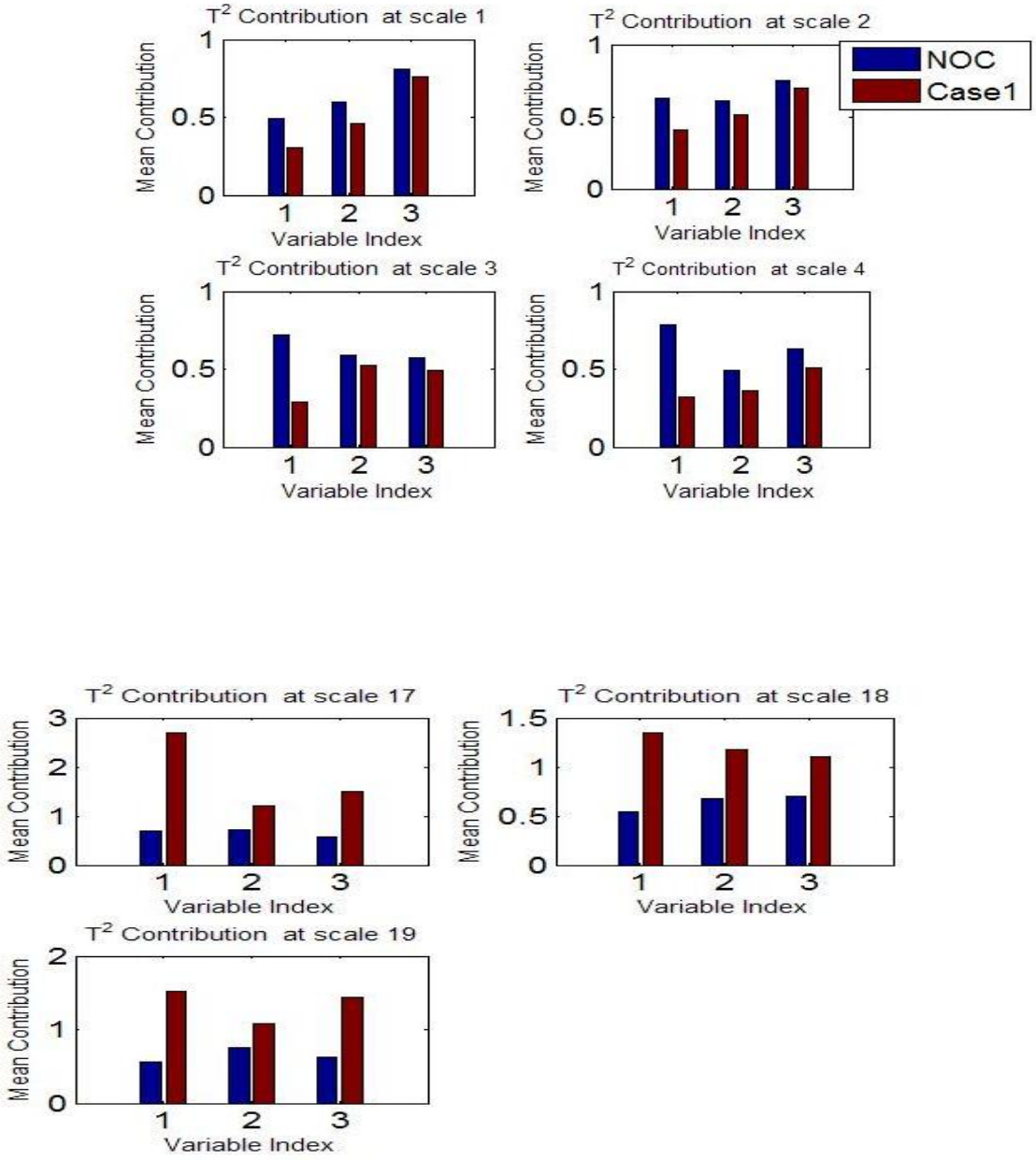

Figure 28 (b) Variable contributions to T2 statistics in case 1 when autocorrelation of variable x1 is changed to 0.5.

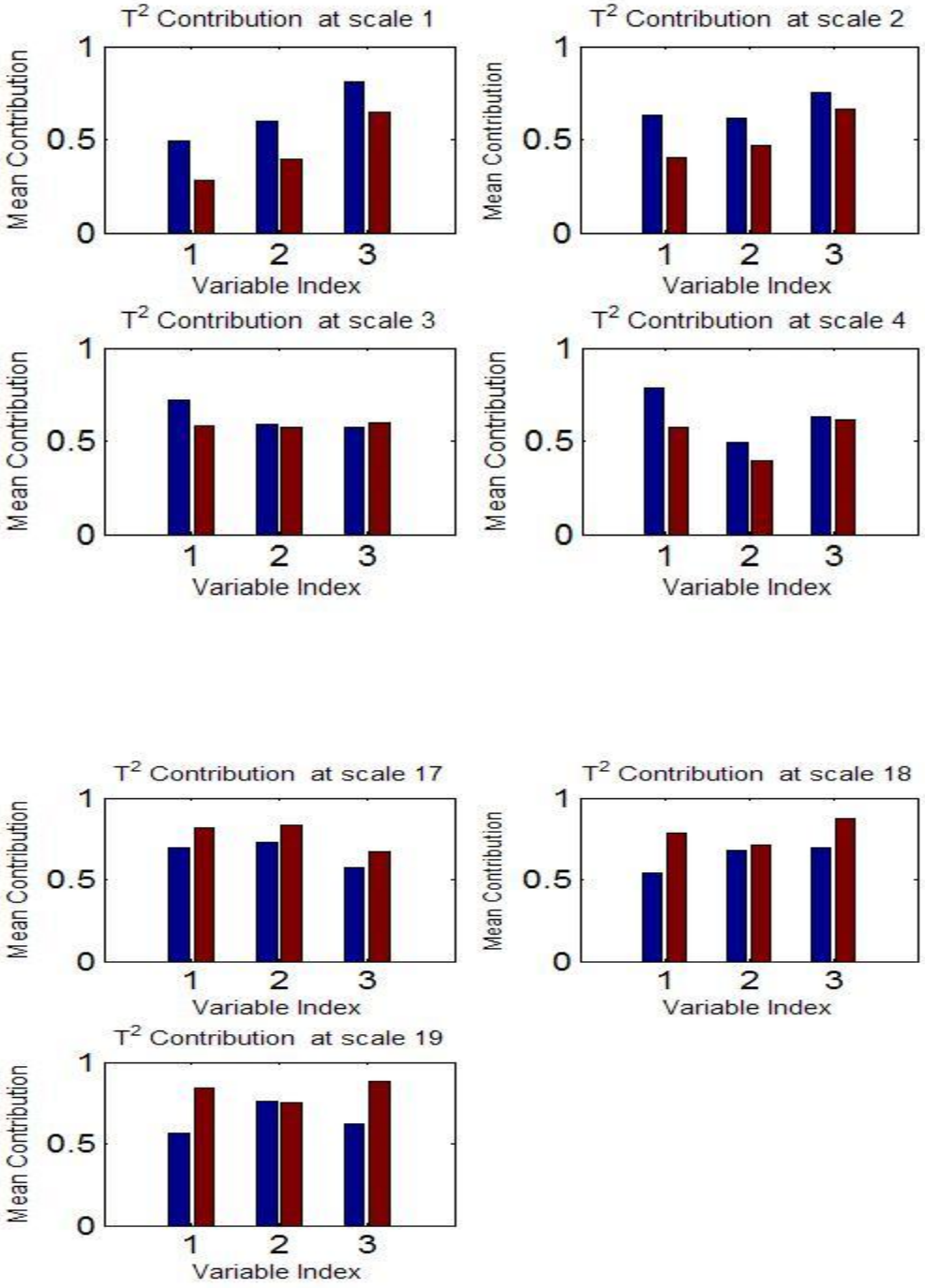

Figure 28 (c) Variable contributions to T2 statistics for case1 when autocorrelation of variable x1 is changed to -0.5.

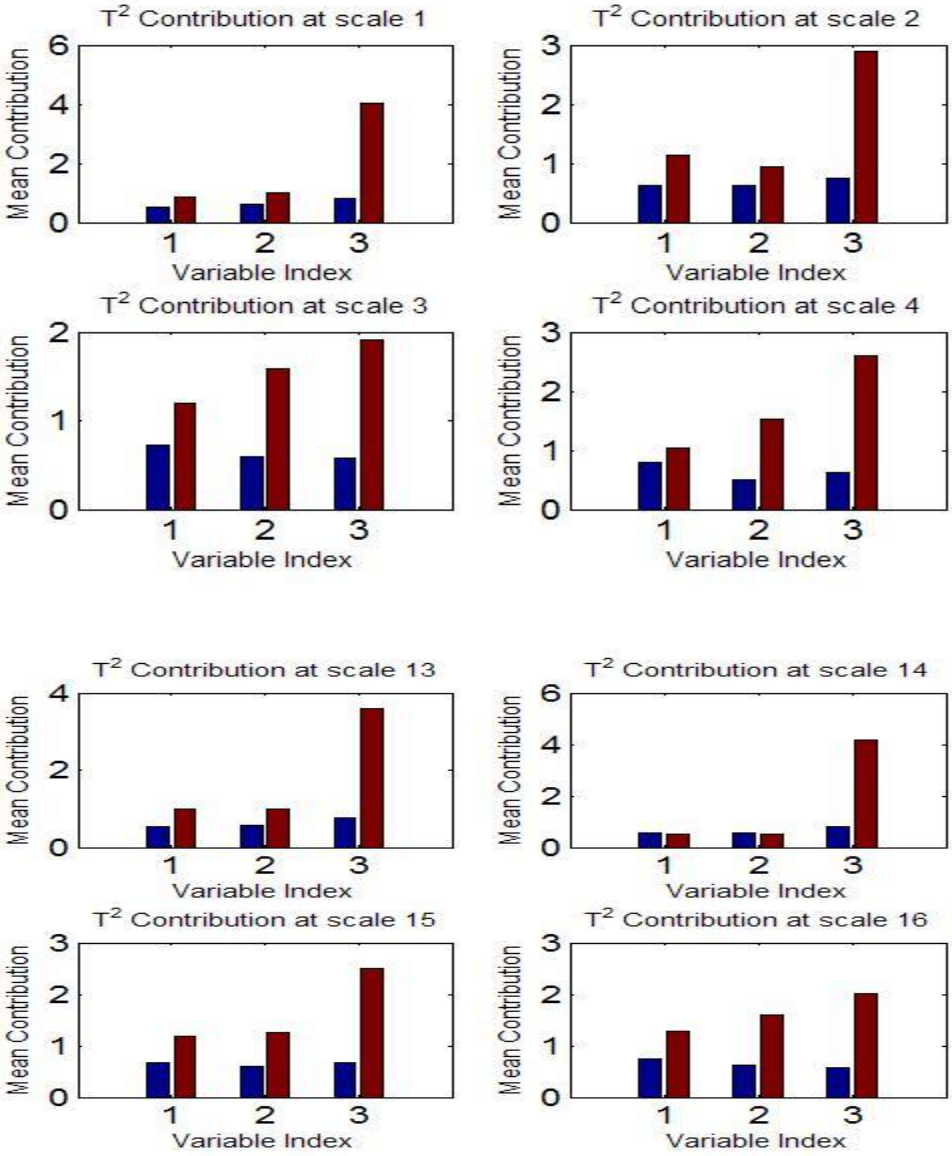

Figure 28 (d) Variable contributions to T^2 statistics for case 2.

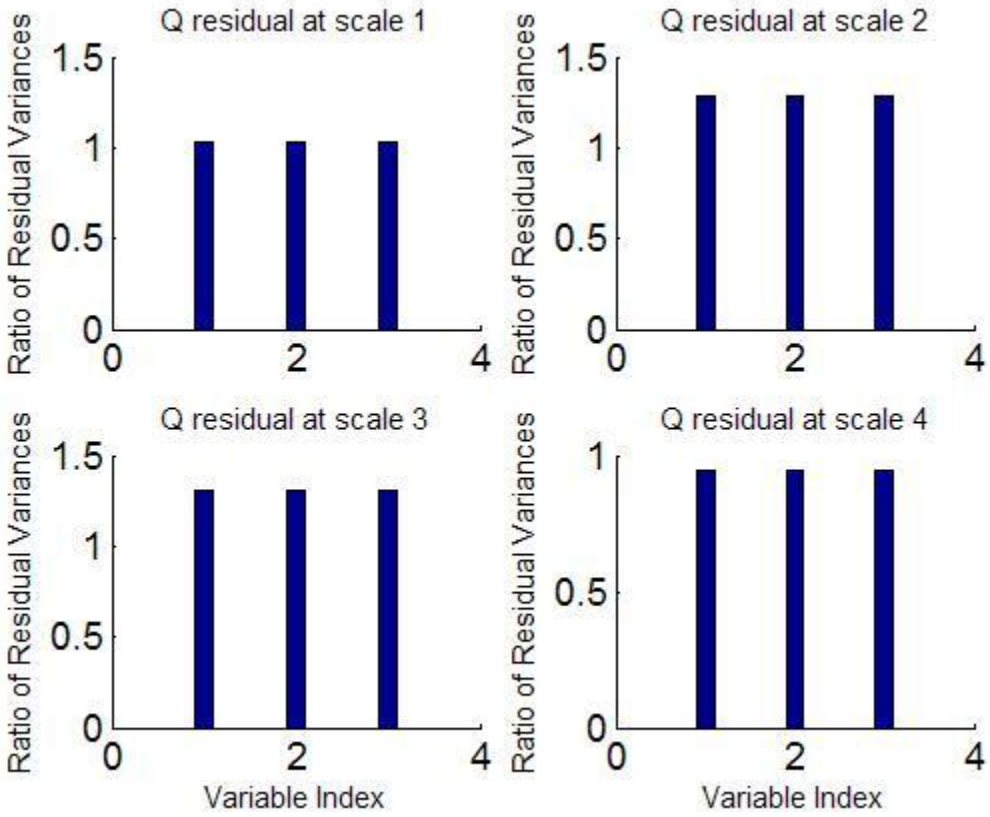

Figure 29 (a) Variable contributions to Q statistics for case 0.

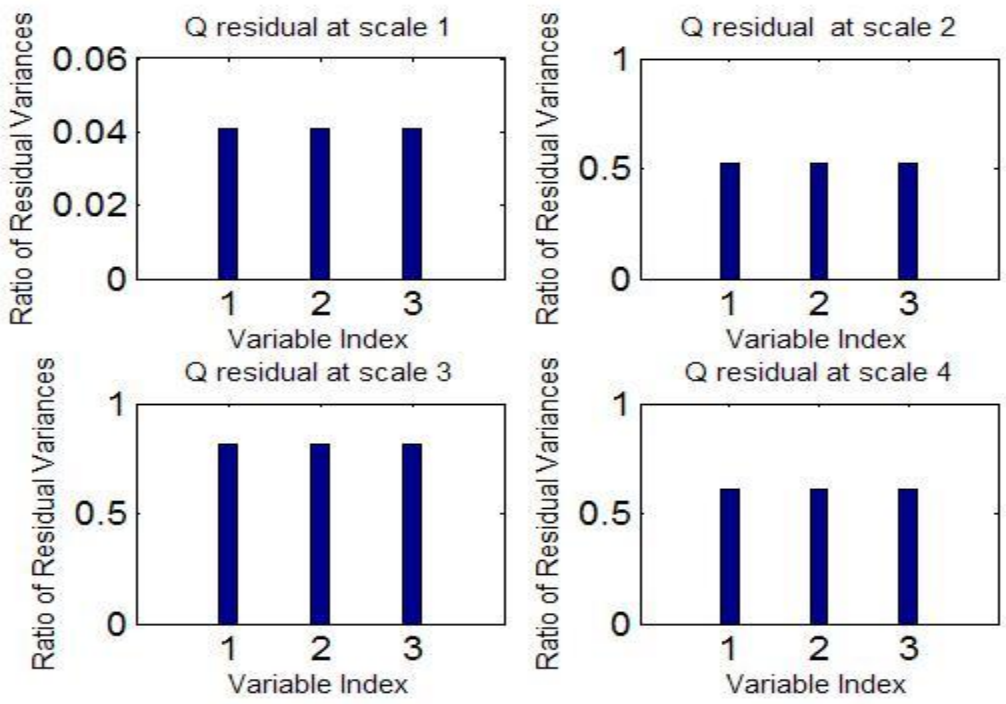

Figure 29 (b) Variable contributions to Q statistics for case 1 in which the autocorrelation of x_1 is changed to 0.5 from 0.9.

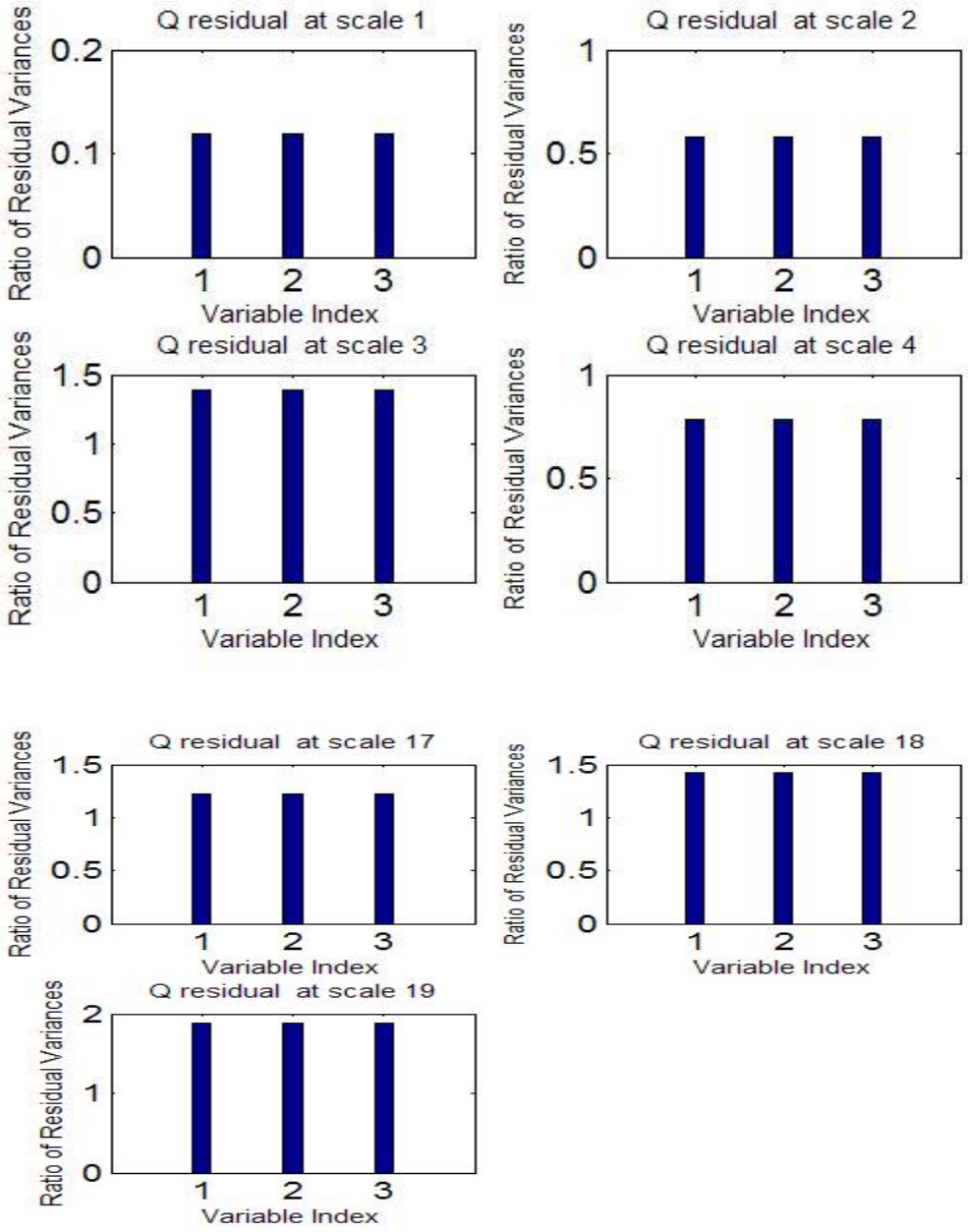

Figure 29 (c) Variable contributions to Q statistics in case1 when autocorrelation of x1 is changed to -0.5.

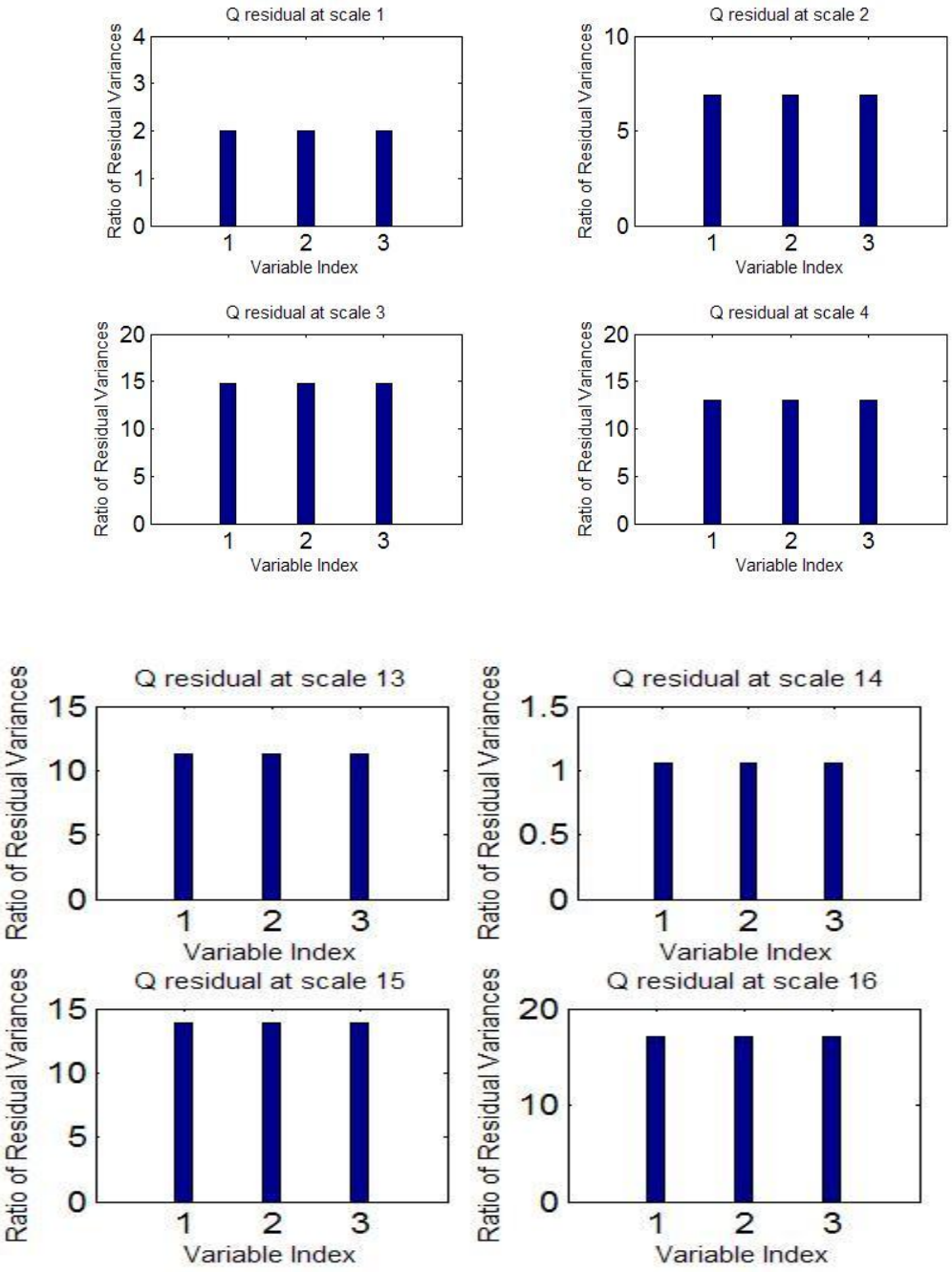

Figure 29 (d) Variable contributions to Q statistics for case 2.

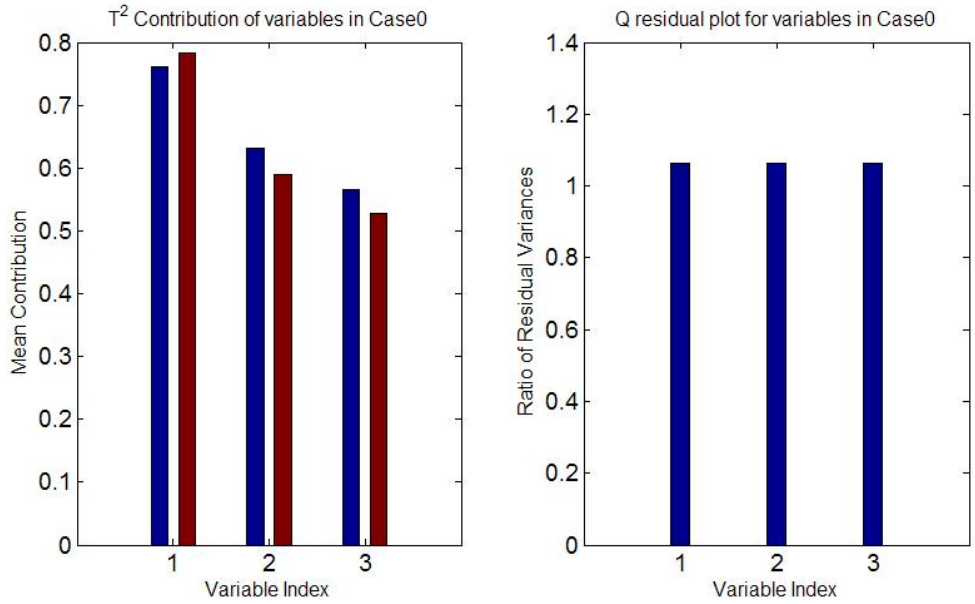

Figure 30 (a) Variable contributions to T^2 and Q statistics for case 0 using cPCA, with both plots showing all variables to be operating under normal conditions.

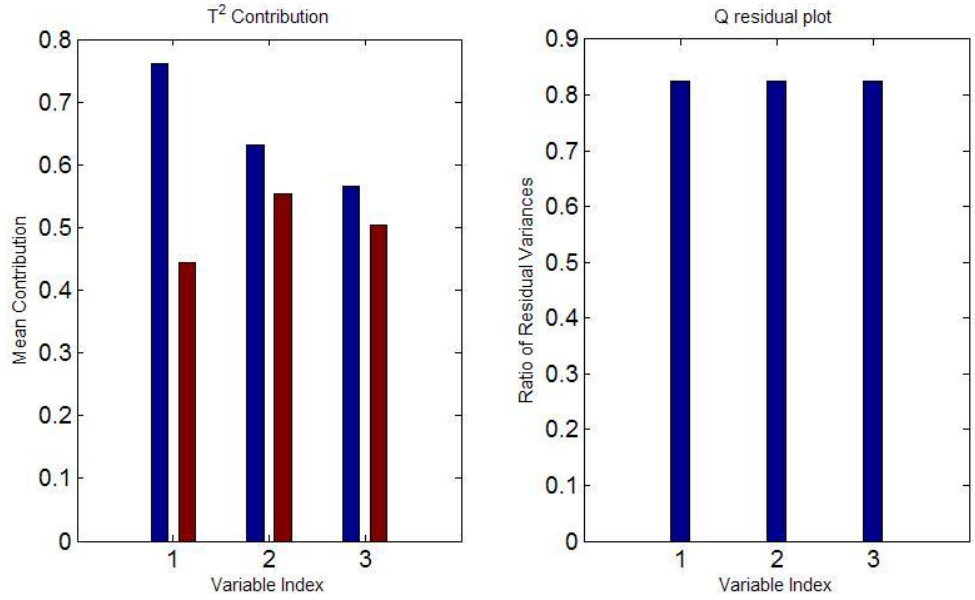

Figure 30 (b) Variable contributions to T^2 and Q statistics for case 1 using cPCA when auto correlation of x_1 is changed to -0.5 from 0.9.

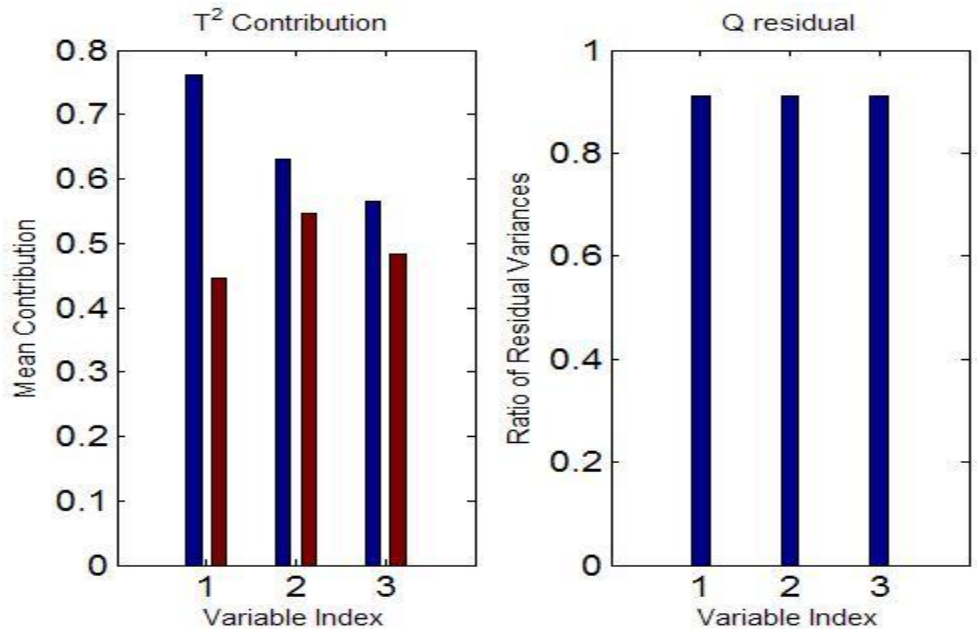

Figure 30 (c) Variable contributions to T^2 and Q statistics for case 1 using cPCA when auto correlation of x_1 is changed to 0.5.

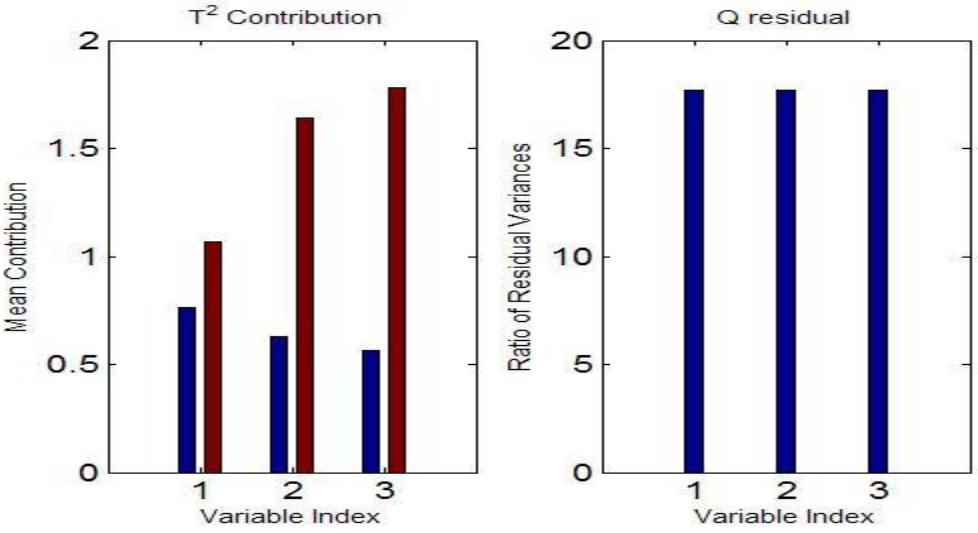

Figure 30 (d) Variable contributions to T2 and Q statistics for case 2 using cPCA.

From the above it can be seen that MS-SSA performed better than cPCA in detecting the changes in the autocorrelation of variable $x_{1,}$ that is Case 1. However, MS-SSA performed worse in detecting the change in positive autocorrelation (Case 1, where autocorrelation of $x_1$is chaged to 0.5 from 0.9).

This can be explained as follows: in positive autocorrelation, adjacent values do not differ as much as values that are further apart. In contrast, for negative autocorrelation adjacent values differ more than values that are farther apart (Stockwell,2007). Thus the change in autocorrelation from 0.9 to 0.5 makes adjacent values of the x_1 very close to the values of the respective variable in the reference model. Hence the distance between the data points in normal and abnormal situations does differ significantly compared to the case of negative autocorrelation in x_1. Nevertheless performance of MS-SSA in Case 1 compared to PCA also suggests advantages arising from the decorrelation by SSA: using MS-SSA allows for explicit time-dependency information in the data and, therefore, reliable fault detection capacity compared to classical MSPC approaches. This also underscores the increased sensitivity of multiscale methods to faults because of the separation of process behaviour into distinct scales in the search for changes in normal variability patterns (Reis and Saraiva, 2006).

A major problem when using MS-SSA is the choice of the window length *M*. There is no canonical rule for selecting the value of embedding dimension in SSA decomposition. However, the aim is to select a proper window length that produces separable and independent PCs. In general, the larger the embedding dimension the stronger is the resolution. A longer window length produces a detailed decomposition of the signal than a short window length. In some cases large window length may result in noisy components being flagged as significant. Therefore, data that exhibit strong periodic components suggest the first maximum zero of the autocorrelation function as the window length, which guarantees that data points are virtually uncorrelated whilst still being close to one another (Michael, 2005). The effect of using different embedding window sizes *M* is summarized in Table 3 for the fault conditions described above. The best reliability percentage was obtained with M =19, which is also the first maximum decorrelation point of the variables, suggesting adequate separation of the signal from the underlying noise at this choice of *M*. However, other choices

different from the zero of the decorrelation point could give more reliable result. The situation under which this is true remains an open research problem.

Table 3 Reliability (%) of MS-SSA using different embedding dimensions for SSA decomposition

MS-SSA	**Case 0-**		**Case 1**		**Case 2**	
Window Size(M)	T^2	Q	T^2	Q	T^2	Q
3	2.5	5.5	2.1/25.7	2.1/29.3	56.5	50.8
5	2.8	5.3	4.8/31	6.6/29.6	58.6	51.6
10	3.3	5.2	4.1/26.6	7.6/36.1	60.9	54.3
15	4.1	6.4	5.3/29.9	7.6/37.7	62.8	58.6
19	4.5	6.5	5.9/35.9	8.6/33.7	63.5	60.9

4.2 Case Study II: A 2x2 Dynamic Process

In this case study, MS-SSA and conventional PCA are applied to a simple autocorrelated 2x2 process previously studied in Ku et al. (1995) in the context of dynamic PCA. The model is represented by

$$\begin{aligned} \mathbf{x}(t) &= A\mathbf{x}(t-1) + B\mathbf{u}(t-1) \\ \mathbf{u}(t) &= C\mathbf{u}(t-1) + D\mathbf{w}(t-1) \\ \mathbf{y}(t) &= \mathbf{x}(t) + \mathbf{v}(t) \end{aligned} \tag{37}$$

where the coefficient matrices are given by

$$A = \begin{bmatrix} 0.118 & -0.191 \\ 0.847 & 0.264 \end{bmatrix}, \quad B = \begin{bmatrix} 1.0 & 2.0 \\ 3.0 & -4.0 \end{bmatrix}, \quad C = \begin{bmatrix} 0.811 & -0.226 \\ 0.477 & 0.415 \end{bmatrix}, \quad D = \begin{bmatrix} 0.193 & 0.689 \\ -0.320 & -0.749 \end{bmatrix},$$

$\mathbf{x}(t)$ and $\mathbf{u}(t)$ are correlated input at time t, $\mathbf{v}(t)$ and $\mathbf{w}(t)$ are uncorrelated zero mean Gaussian noise, with variances of 0.1 and 1 respectively. The inputs $\mathbf{u}(t)$ and outputs $\mathbf{y}(t)$ are measured and used to monitor the system.

Eight fault conditions were simulated as follows: fault conditions 1-5 were generated by progressively larger shifts in the mean of the first component of the $\mathbf{w}$ vector, while the last three fault conditions were generated by changing the coefficient mapping u_1 to x_2 as shown in Table 4.

Table 4 Settings of fault conditions of 2x2 system

Case	Fault Type	Fault Magnitude
0	Normal Condition	N/A
1	Mean shift of w_1	$0.0 \rightarrow 0.5$
2	Mean shift of w_1	$0.0 \rightarrow 1.0$
3	Mean shift of w_1	$0.0 \rightarrow 1.5$
4	Mean shift of w_1	$0.0 \rightarrow 2.0$
5	Mean shift of w_1	$0.0 \rightarrow 3.0$
6	Change of parameter from u_1 to x_2	$3.0 \rightarrow 2.5$
7	Change of parameter from u_1 to x_2	$3.0 \rightarrow 2.0$
8	Change of parameter from u_1 to x_2	$3.0 \rightarrow 1.0$

Similar to Kano et al., (2002), a 99% confidence limit was empirically determined using Monte Carlo simulation based on 200 realizations of normal operating conditions. Subsequently, the MS-SSA was evaluated for each of the fault conditions and the percentage of samples violating the control limit, or reliability, was computed. An embedding window of size $M = 6$ was determined using the first maximal decorrelation point in the autocorrelation plots of variables (Figure 31). Hence, a six-level multiscale representation was considered, and classical PCA-based statistical process control models were developed at each scale. The retained principal components as well as the corresponding proportion of the total variance explained are given in Table 5. The results reported in Table 6 are averages over 1000 simulations using test samples of size 400. For comparison, the results cPCA method as well as MSPCA method considered in Kano et al., (2002) are also indicated in Table 6.

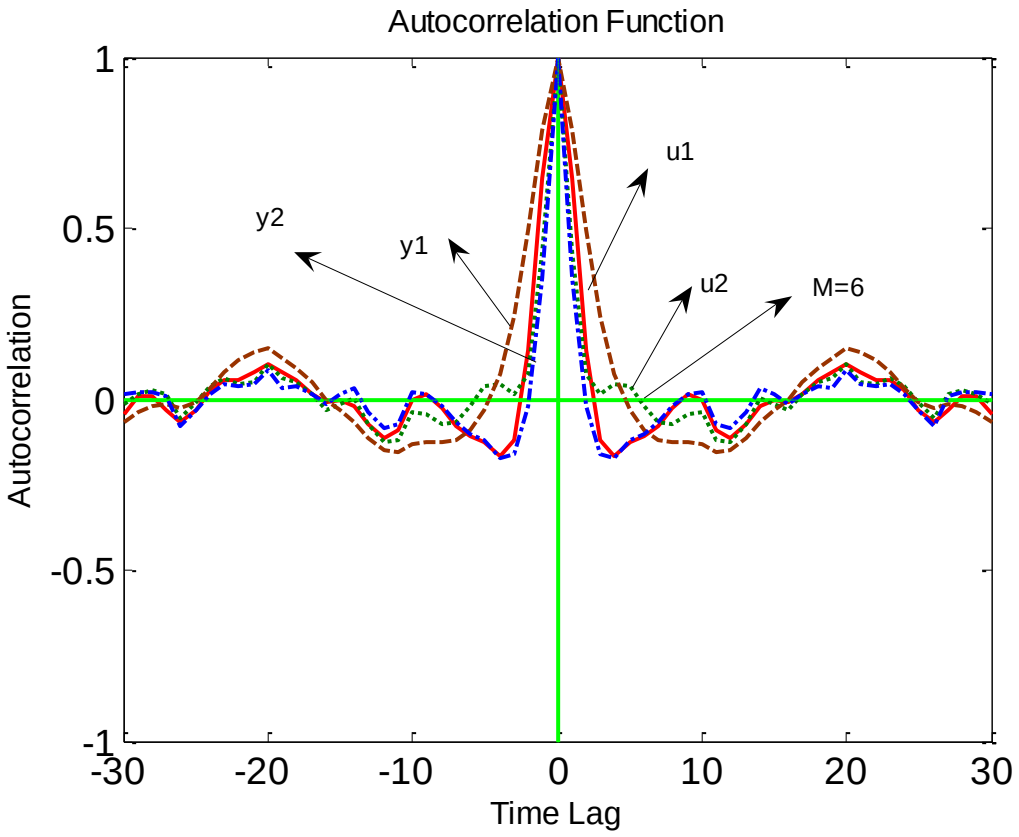

Figure 31 Autocorrelation functions of the variables in 2x2 processes.

Table 5 Percentage of variance captured and number of PCs retained in MS-SSA and PCA approach

	MS-SSA (per scale)						**PCA**
	1	2	3	4	5	6	
% variance captured	99.9	95.39	98.96	96.49	95.88	97.36	98.30
Number of PCs retained	3	2	3	2	2	2	3

Table 6 Reliability (%) of MS-SSA, PCA and MSPCA

Method	**Index**	**Cases**								
		0	1	2	3	4	5	6	7	8
MS-SSA	T^2	1.79	2.14	2.67	4.1	7.23	24.9	1.82	2.04	2.58
	Q	3.04	3.64	3.86	5.84	10.7	32.3	17.0	43.4	68.0
PCA	T^2	1.1	1.3	2.4	4.3	8.5	23	1.2	1.3	2.0
	Q	1.0	1.0	1.1	1.2	1.6	2.3	1.6	3.2	9.5
MS-PCA	T^2	0.7	0.5	0.7	1.3	3.0	15.6	0.6	1.1	4.5
	Q	0.8	0.9	1.5	3.0	7.5	28.8	3.9	19.8	57.0

The results in Table 6 indicate that MS-SSA performs better than PCA, particularly for detecting the parameter changes in the process. For example, in case 7 the reliability of MS-SSA is 43.4%, but that of conventional PCA is 3.2%. The highest reliability for cPCA of 9.5% was obtained in case 8, with a corresponding reliability of 68.0% for MS-SSA. The results clearly show the improved reliability of MS-SSA over PCA for data with significant autocorrelation. It is also to be noted that in most of the cases the reliability of *Q* statistic of MS-SSA is higher than that of PCA, and also higher than the T^2 statistics in the respective cases, possibly because the changes in the mean shift and parameter changes of the abnormal cases change the correlation between the variable rather than the deviations among the variable and this change is captured by the residual space in the PCA model of the corresponding scales. Conventional PCA fails to detect these fault conditions since it ignores auto-correlations in data.

MS-SSA also performs better than MSPCA in all cases which, in turn performs better than conventional PCA in some cases, especially in detecting parameter changes. It can be concluded that multiscale methods outperform single scale methods when correlations amongst variables are affected by parameter changes, with little or no changes in the variation of the variables.

Figures 32 and 33 show the variable contributions to the $T^2$and *Q* statistic values in case 8 using MS-SSA and cPCA for case 8 respectively. The variables that

make the greatest contribution to the deviations in T^2 and Q charts in case 8 can be identified using these plots in both methods. As mentioned earlier, parameter changes to variables in case 8 change the correlation between the variables rather than deviations among the variable, and this change is captured by the residual space in the PCA models of the corresponding scales in MS-SSA as well as the residual space of cPCA. Therefore, the residual contribution of all the variables in both methods have changed significantly in these plots.

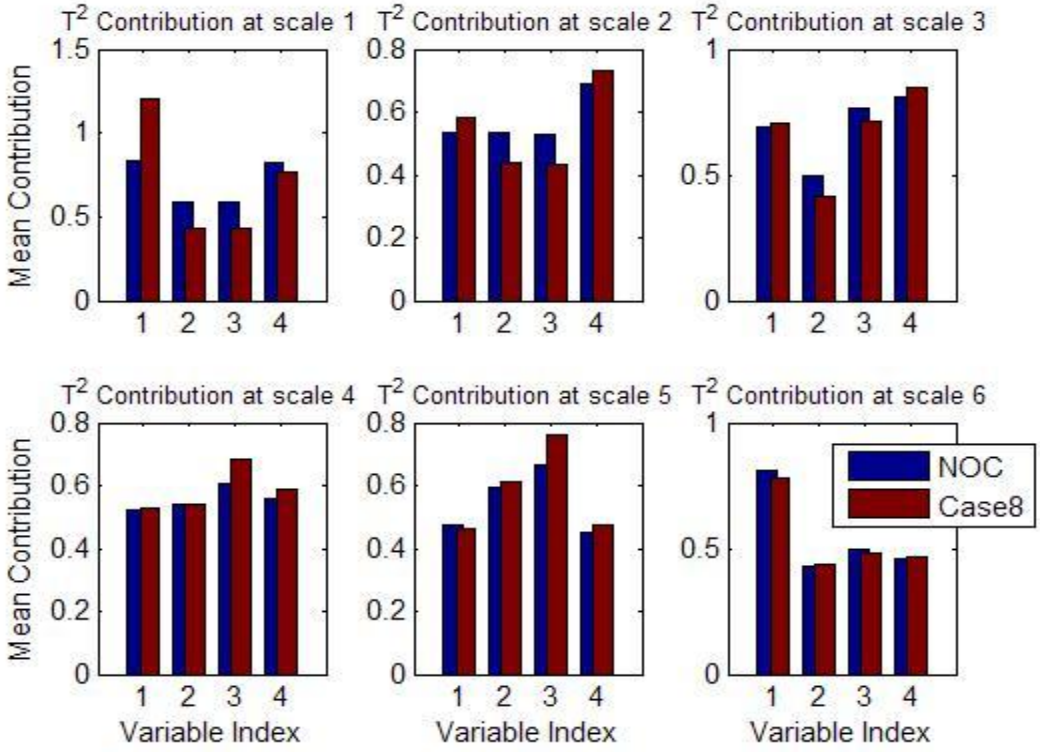

Figure 32 (a) Variable contributions to T^2 statistics in case 8 using MS-SSA.

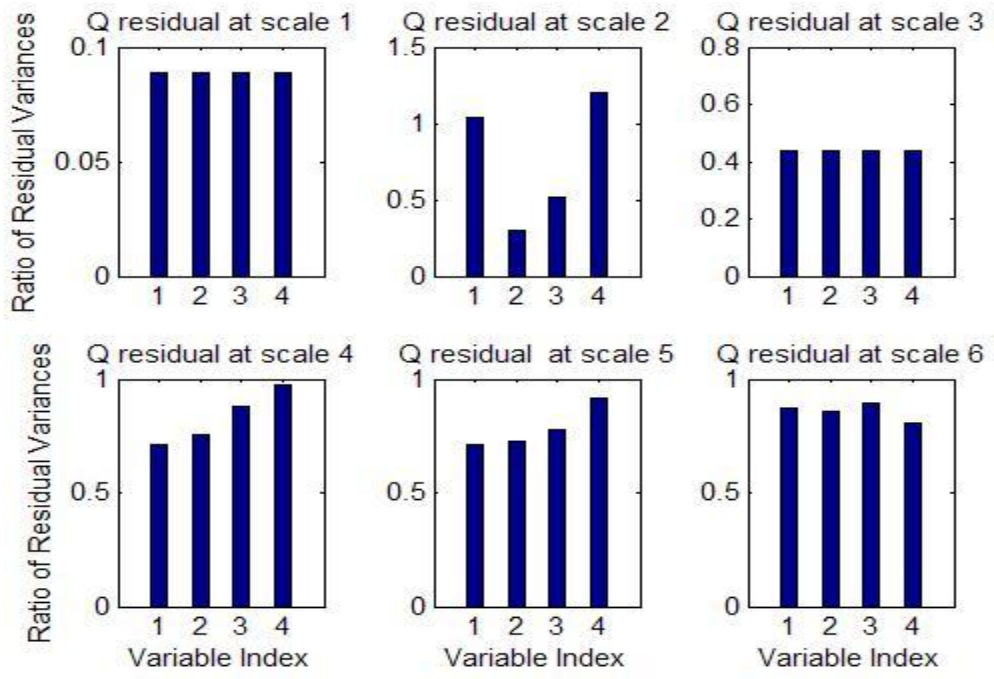

Figure 32 (b) Variable contributions to Q statistics in case8 using MS-SSA.

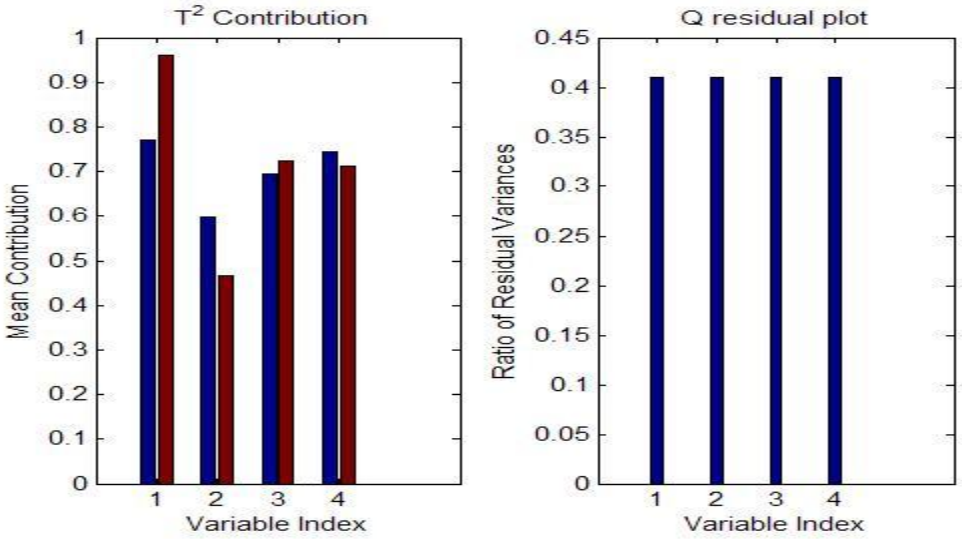

Figure 33 Variable contributions to T^2 and Q statistics in case8 using cPCA.

4.3 Case Study III: Tennessee Eastman Process

The Tennessee Eastman Process (TEP) is a simulation of an actual chemical process developed as a realistic industrial case study useful for plant-wide process control problems including process monitoring and fault diagnosis (Downs and Vogel, 1993; Kano et al., 2002). The process consists of five major units (a reactor, a product condenser, a recycle compressor, a vapor-liquid separator, and product stripper) and involves eight components labeled A, B, C, D, E, F, G and H. Components G and H are liquid products produced from the four gaseous reactants, A, C, D and E. The inert product B is also fed to reactor and the byproduct F is produced. The process flow sheet is shown in the Figure 34.

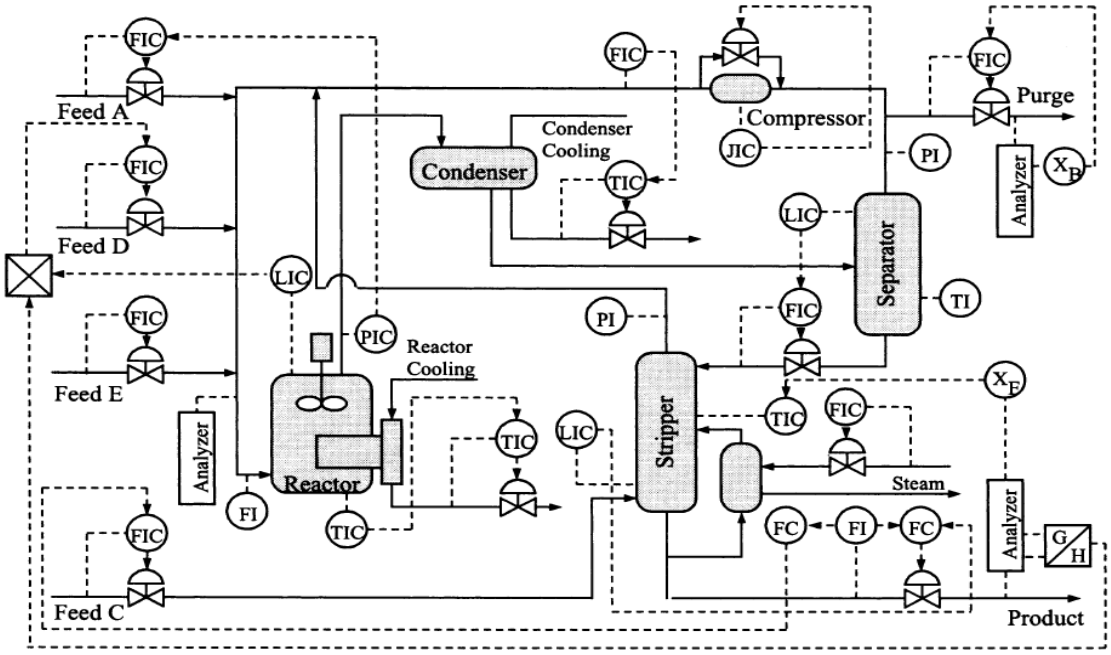

Figure 34 Process flow sheet for TEP(Downs and Vogel, 1993)

The reactions in the reactor are represented by:

$$A(g) + C(g) + D(g) \rightarrow G(liq), product1 \quad (38)$$

$$A(g) + C(g) + E(g) \rightarrow H(liq), product2 \quad (39)$$

$$A(g) + E(g) \rightarrow F(liq), byproduct \quad (40)$$

$$3D(g) \rightarrow 2F(liq), byproduct \quad (41)$$

The reactions in (38)-(41) are exothermic, irreversible and approximately first-order with respect to the reactant concentrations. The reaction resulting in G product formation (38) has a higher activation energy than reaction (39) and, therefore has a high sensitivity to temperature.

The reactant gases after being fed into the reactor form liquid products, which are catalyzed by a non-volatile catalyst dissolved in the liquid. Heat of reaction is reduced by cooling water that is circulated in the reactor. The gaseous products are separated from the reactor while the catalyst is retained. The product gas is cooled by using a condenser and then fed to a vapor-liquid separator. The recycling of the non-condensed vapor from the separator to the reactor is done through a compressor. A portion of the recycled vapor is purged to keep the inert product and byproduct from accumulating in the process using the vapor-liquid separator. The remaining reactants in the condensed stream from the separator are removed in the stripper.

From a data perspective, the process has 12 manipulated or control variables and 41 measured or observed variables. Of the latter, 22 are continuous and 19 are composition measurements. Of the 53 process variables, 16 variables as selected by Chen and McAvoy (1998) are used for monitoring purposes and are listed in Table 7. The data representing normal operating conditions contain 500 samples.

Table 7 Process Variables used for monitoring.

Variable	**Description**
1	A feed
2	D feed
3	E feed
4	A and C feed
5	Recycle flow
6	Reactor feed rate
7	Reactor temperature
8	Purge rate
9	Product separator temperature
10	Product separator pressure
11	Product separator underflow
12	Stripper pressure
13	Stripper temperature
14	Stripper steam flow
15	Reactor cooling water outlet temperature
16	Separator cooling water outlet temperature

The TEP process contains 21 faults listed in Table 8. Of the 21 faults, 16 faults are known and 5 are unknown. The first 7 faults are associated with a step change in a process variable. Faults from 8 to 12 are associated with the increased variability in some of the process variables. Fault 13 is a slow drift in the reaction kinetics while faults 14, 15, and 21 are related to actuator faults such as sticking valves.

Table 8 Process Faults

Case	Disturbance	Type
1	A/C feed ratio, B composition constant	Step
2	B composition, A/C ratio constant	Step
3	D feed temperature	Step
4	Reactor cooling water inlet temperature	Step
5	Condenser cooling water inlet temperature	Step
6	A feed loss	Step
7	C header pressure loss - reduced availability	Step
8	A, B, C feed composition	Random variation
9	D feed temperature	Random variation
10	C feed temperature	Random variation
11	Reactor cooling water inlet temperature	Random variation
12	Condenser cooling water inlet temperature	Random variation
13	Reaction kinetics	Slow drift
14	Reactor cooling water valve	Sticking
15	Condenser cooling water valve	Sticking
16-20	Unknown	Unknown
21	The valve for Stream 4 was fixed at the steady state position	Constant Position

The proposed MS-SSA framework is compared against conventional PCA in the following. As in earlier case studies, 95% control limits are assumed, adjusted for multiple tests in the multiscale case.

Using cPCA, 13 PCs explaining 96% of the variation in the data for the normal case were retained in the PCA model. The estimated reliability of T^2 and Q statistics in each case using cPCA is listed in Table 9.

An embedding window of size $M = 38$ based on the first maximal decorrelation point of monitored variables, leading to a 38-level multiscale representation in MS-SSA case. The PCs retained in the PCA model for each scale explains at least 96% variance in the data are shown in Figure 35. The reliability percent based on MS-SSA are calculated from the samples which violate the 95% confidence limits in the respective scales. In each scale the ratio of the number samples violating the control limits in both T^2 and Q statistics values are calculated, and the maximum detection rate is considered as the reliability percent of the associated fault condition as summarized in Table 9. The monitoring performance of MSPCA method for all fault conditions is also evaluated for analyzing the proficiency of multiscale methods in fault detection. In this application the level of wavelet decomposition is set at 3. The PCs retained in the PCA model of reconstructed signal capture at least 96% variance in the data. The reliability of MSPCA in different fault conditions is also summarized in Table 9.

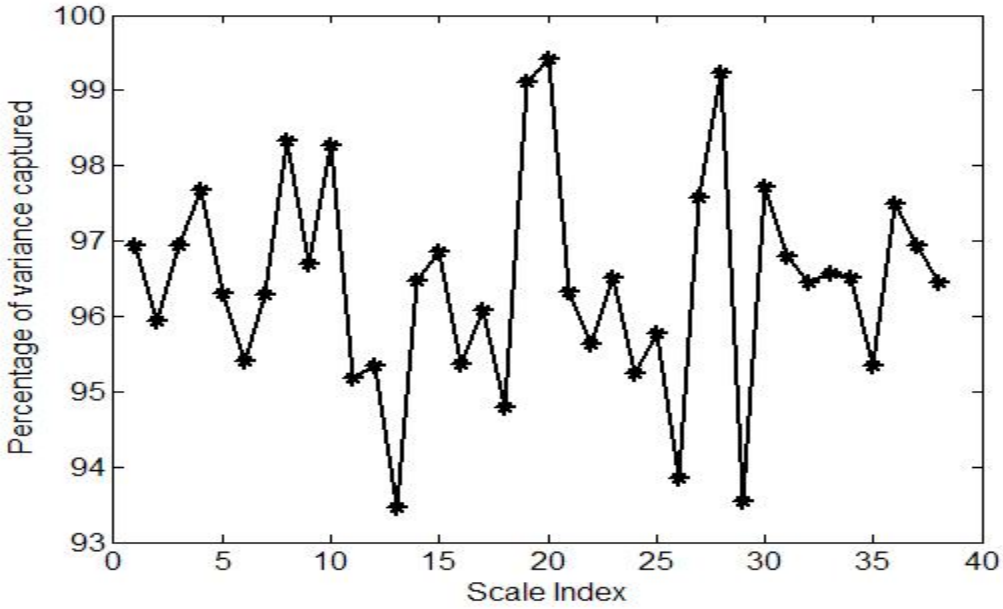

Figure 35 Percentage of variance captured by PCs in all scales in MS-SSA.

Table 9 Reliability (%) of MS-SSA, PCA and MSPCA.

	Process Monitoring Method					
	cPCA		**MS-SSA**		**MSPCA**	
	T^2	Q	T^2	Q	T^2	Q
Case 1	100	100	100	100	100	100
Case 2	100	93	100	100	100	96
Case 3	60(1.7)	16(18.9)	48	51	40(19.6)	23(10.7)
Case 4	29	9	31	79	35	100
Case 5	100	81	100	100	100	93
Case 6	100	100	100	100	100	100
Case 7	100	87	100	100	100	100
Case 8	98(84.6)	90(88.2)	100	100	100(86.2)	97(87.5)
Case 9	11(3.1)	10(14.9)	31	46	1(15)	27(13.8)
Case 10	84(76.5)	86(81.2)	93	95	90(78.5)	57(81.3)
Case 11	67	14(20.4)	93	86	0(4)	41(47.9)
Case 12	100	89	100	100	100	90
Case 13	83	88	100	100	75	81
Case 14	100(6.2)	93(70.8)	98	65	49(65.5)	30(94.5)
Case 15	8	7	37	49	30	10
Case 16	28(9)	12(21.9)	55	87	31(23.4)	26(39)
Case 17	99	95	100	100	97	98
Case 18	52	48	68	87	51	40
Case 19	40(10.3)	14(11.3)	72	91	2(29.5)	10(67.1)
Case 20	45(67.4)	32(69.1)	54	81	51(67.1)	40(68.5)
Case 21	9	10	20	100	49	51

The reliability of MS-SSA is considerably better than that of cPCA for fault conditions 4, 9, 11, 15, 16, 18, 19, 20 and 21. Case 4 is a step change in the reactor cooling water inlet temperature. Hence small variations in the measurements compared to the normal operating condition may occur as a result of a change in the temperature of the reactor. MS-SSA is able to detect these small events better than what cPCA does in these cases. Similarly, fault condition 11 is associated with random variation in the reactor cooling water inlet temperature. Hence, large oscillations in the reactor cooling water flow rate

induce fluctuations in the reactor temperature. The unknown faults 16, 18, 19, 20 and 21 are also detected better by MS-SSA than cPCA.

In cases 2, 5, 7, 8, 12, 13, and 17, MS-SSA outperformed cPCA with 100% reliability in both T^2 and Q statistics. In those cases the reliability of cPCA was above 80%, and that of MSPCA was above 90%, except for case 3 where the reliability of T^2 statistics for cPCA was higher than that of MS-SSA and MSPCA. However, the variables contributing to this change could not be identified from the T^2 contribution plots shown in Figure 36. The step change in D feed temperature causes a mean shift in flow and outlet temperature of the reactor cooling water via a cascade control system for reactor temperature (Kano et al., 2002). This small mean shift in the measurement caused a change in the relationship between the process variables as defined under normal operating conditions. This possibly caused an increase in the values of the Q statistics in case 3 for MS-SSA. This indicates the relatively improved potential of MS-SSA in detecting mean shifts in the process compared to conventional PCA. As indicated in Table 9, MS-SSA performed at least as well as the MSPCA. In fact in most of the cases except in Case 4 MS-SSA gave better performance in detecting the corresponding faults in those cases. Generally, MS-SSA performed better than other methods in detecting faults in most cases. The slight differences in the results obtained with cPCA and MSPCA in some faults and those reported by Kano et al., (2002) as shown in brackets in Table 9 could have been caused by procedural differences; Kano et al., (2002) results are based on 10 realizations while in this study the % reliability is calculated based on the first 100 samples after the occurrence of a fault, with a single data set being used in each case.

The contribution plots in Figure 37(a) and (b) show the variable contribution to the Q and T^2 statistics. The performance of MSPCA in detecting mean shift was not satisfactory in this case. Generally MS-SSA outperformed the other techniques in detecting faults especially in case 9 and 11.

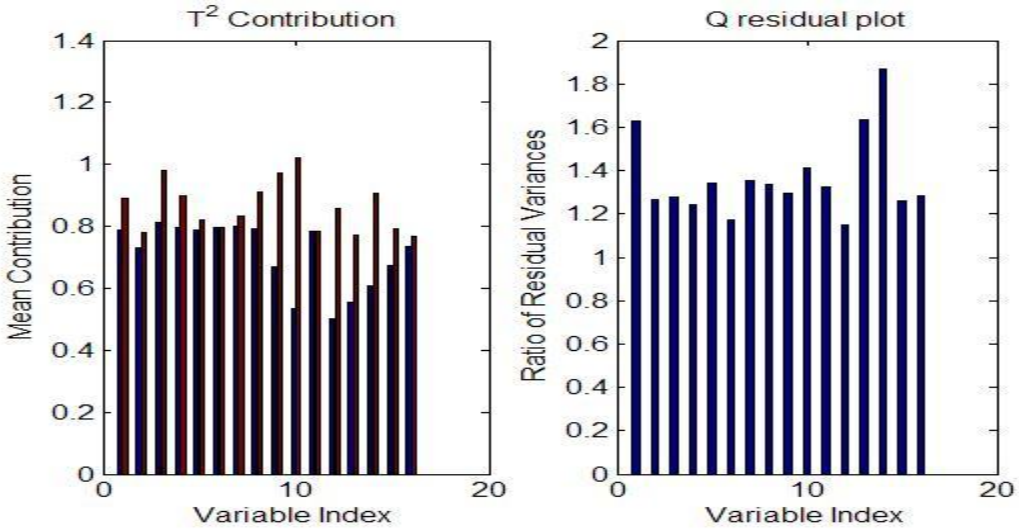

Figure 36 Variable contributions to T^2 and Q statistics in case3 using cPCA. The Q residual plots shows variables that are contributing the fault condition in this case.

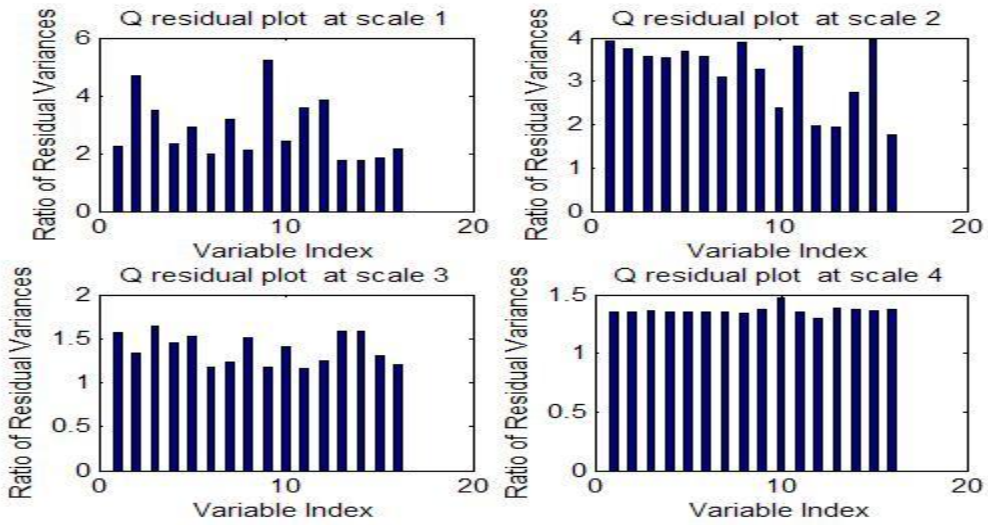

Figure 37(a) Variable contributions to Q statistics in case3 using MS-SSA. Variables that are violating the control limits of Q residual plot in scale 1 shows the variables that are contributing to this fault condition.

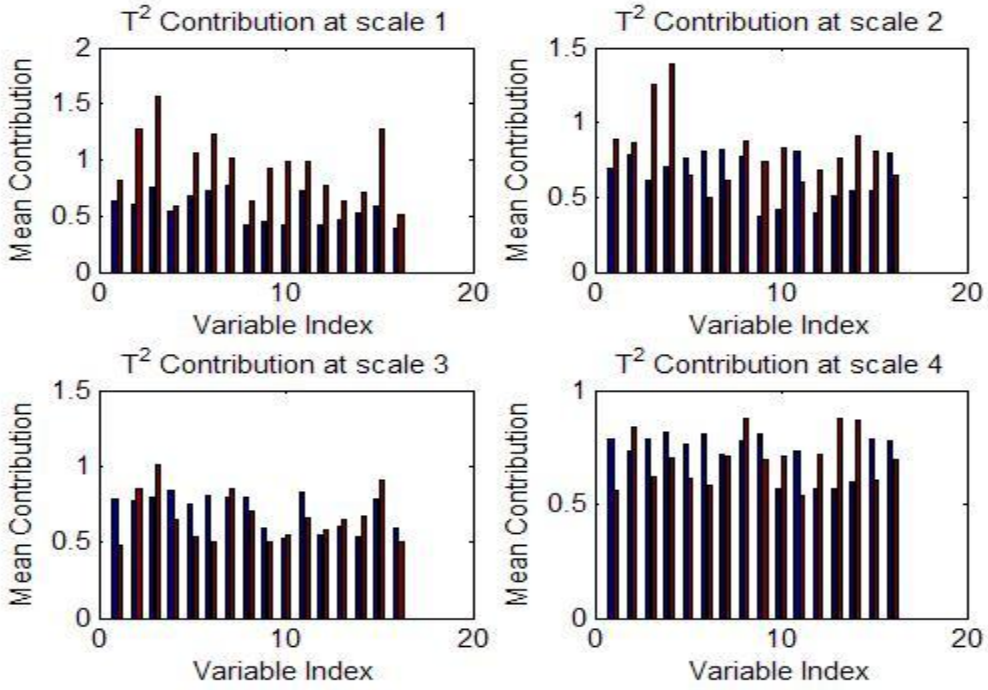

Figure 37(b) Variable contributions to T^2 statistics in case3 using MS-SSA. Variables having largest contribution in the mean contribution to T^2 statistics are contributing to the fault condition in this case.

4.4 Case Study IV: PGM Milling Circuit

In the final case study, the proposed MS-SSA monitoring is applied to industrial data from a milling circuit of a precious group metals plant. In this context, MS-SSA can be used for early detection of any abnormal event that could potentially lead to sudden changes in a pre-specified key performance indicator (KPI) measure against which process stability is tracked against. This is achieved by identifying process variables that significantly deviate from set points and, thereby inducing perturbations in the KPI index. Variable identification is useful in tracing the cause or origin of a fault in troubleshooting efforts. Instead of haphazard efforts, focus is directed on subsystem(s) of the PGM milling circuit where the faults are most likely to have originated. Such troubleshooting assistance can significantly reduce plant recovery of operations when a fault results in plant downtime. Also, fault propagation to other subsystems is avoided when proper and immediate diagnostic actions are taken (Chiang and Braatz, 2003).

Unlike data sets used in the other case studies investigated earlier, prior assumptions with regard to process operations are difficult to make with respect to industrial data. Specifically, quality of data is unknown and it is possible that data are corrupted with unknown faults and/or gross errors. It can be conjectured that multiscale methods will perform better than classical multivariate approaches because of their improved ability to separate deterministic and stochastic components in data, with a consequent improvement in safety and productivity of an operation.

4.4.1 Data Description

The PGM milling circuit consists of a few integrated unit operations, namely a crushing circuit, a grinding circuit, and an array of flotation banks as depicted in the process flow sheet diagram in Figure 38. Data are available for 11 monitored variables and two dependent variables as listed in Table 10. Manipulated variables are not included in the analysis. Variables 13 and 5 are dependent on

the values of variables 1, 2 and 3. Variables 3 and 4 are related to the final products and variables 1 and 2 related to the feed.

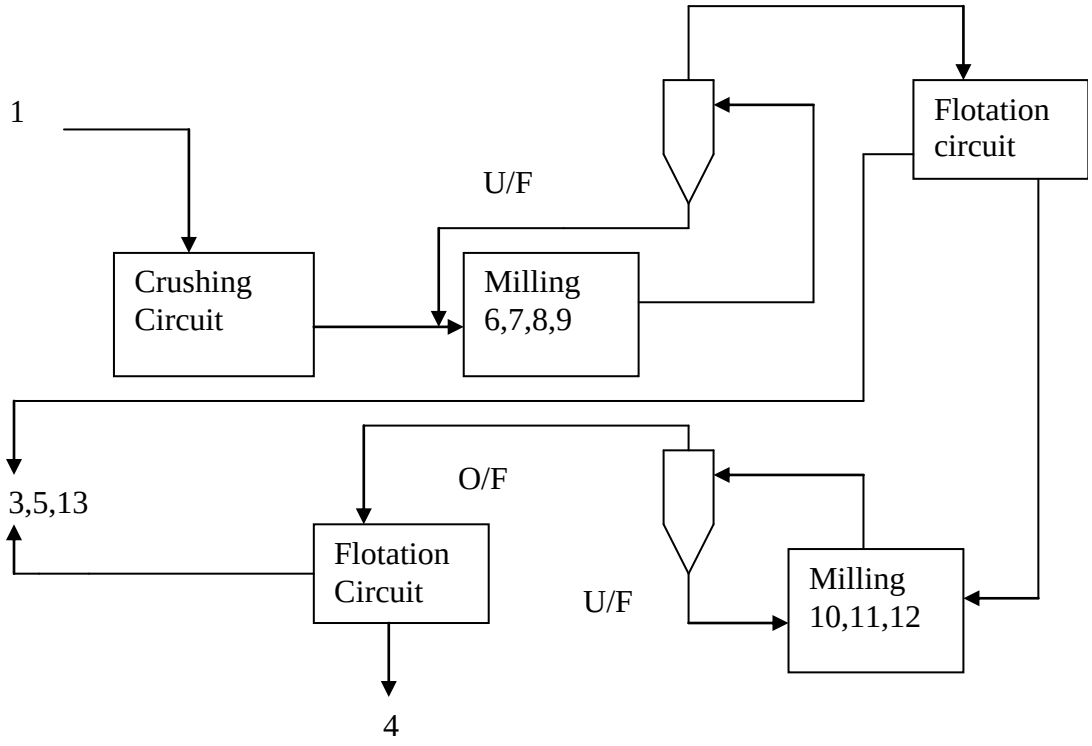

Figure 38 Flow sheet diagram of PGM milling circuit.

Table 10 Process Variables for PGM milling circuit

Process Variable	Description
1	Sample head grade
2	Tons treated
3	Concentrate grade
4	Final tails grade
5	Mass pull
6	Primary mill power
7	Primary grind –75□m
8	Primary grind +75□m
9	Primary mill discharge density
10	Secondary grind –75□m
11	Secondary mill discharge density
12	Cyclone Pressure
13	Recovery

4.4.2 Problem Description

To demonstrate the effectiveness of the MS-SSA approach over the PCA approach, two sets of analysis are conducted on two sets of data, designated as fault condition 1 and fault condition 2. These are obtained based on abnormal variations in the mean and standard deviations of variable 13 during certain periods of process operation when compared to a period considered as representative of normal operating conditions. The objective in this case study is to determine the causes of variation in the values of variable 13 during the respective periods of abnormal behaviour. Table 11 shows the mean and standard deviation of variable 13 for the different phases, from which it can be seen that fault condition 1 and 2 periods showed a larger deviation compared to normal operating condition.

Table 11 Mean and Standard Deviation of Variable 13 in normal and fault conditions

	Mean	Standard Deviation
Fault condition 1	1.84	0.0294
Normal condition	1.85	0.0098
Fault condition 2	1.83	0.0251

Sample data with 12 process variables obtained during the normal operating condition are used as a training data set to build reference monitoring models and a validation set to adjust the control limits of the monitoring charts. The fault conditions 1 and 2 are monitored based on the MS-SSA and PCA model built with the 28 data points and 12 measured variables as indicated in the Table 10 using data collected in normal condition. Data sample sizes were 174 and 119 for fault condition 1 and 2 respectively. MSPCA model could not be used in this study, since the data points obtained in the normal operating condition was not having sufficient dyadic length.

4.4.3 Results

Using autocorrelation functions of the variables, the embedding window was selected as *M*=9. The percentage of variance captured and the number of PCs retained for both approaches are given in Table 12

Table 12 Percentage of variance captured and the number of PCs retained in MS-SSA and PCA approach.

	MS-SSA (per-scale)									**PCA**
	1	2	3	4	5	6	7	8	9	
% variance captured	97.7	98.1	96.6	96.5	97.5	96.6	96.2	96.0	97.7	96.7
Number of PCs retained	4	8	9	9	10	10	9	9	9	9

The T^2 and Q statistics plots obtained by PCA and MS-SSA with 95% confidence limit are shown in Figures 39-40.

In fault condition 2, Figure 41(b) and 41(d) show that both T^2 and Q values violate the 95% limit after time=75. These values increase monotonically until reaching a maximum value at time 82 and, thereafter, decrease slowly in all scales, indicating that a step change in the process may have occurred. A similar disturbance is also observed in the plots of fault condition 1 in Figure 41(a) and 41(c) around the time=31 and time=182. In fault condition 2 in Figure 41(c) the spike violating the Q statistics limit at time=114 in the high frequency scales

arises from spiky disturbances (disturbances that are strong in high frequency scales, but they becomes weaker in the low frequency scales). However this disturbance is detected only in the last few scales. There are also some spiky disturbances which are detected in fault condition 1 as well in the high frequency scales. These spiky disturbances could be due to local disturbances in some of the variables. These disturbances with large spikes are also detected in the PCA plots, consistently violating the limits in both T^2 and Q charts. This can be considered a false alarm due to abrupt changes in the process operations or localized events that are not persistent. In both fault conditions the disturbances detected in PCA is hard to interpret properly since PCA uses information with noise in the data and also PCA is based on the statistical distribution in the time-domain. This once again confirms the better detection of the occurrence of small disturbance in the signal using methods based on time-frequency domain.

In general multiscale methods perform better than conventional multivariate SPC methods for the investigated dynamic process. Fault detection rates shown in Table13 (a) and (b) also indicate that MS-SSA gives good results compared to PCA due to its ability to capture the deterministic characteristics in the data.

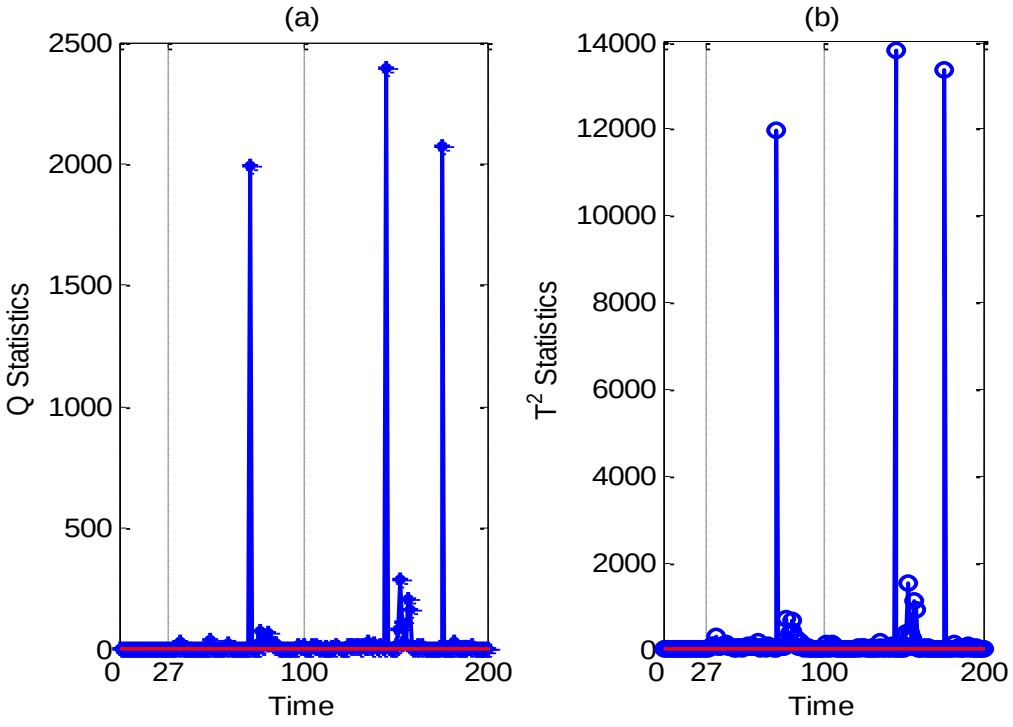

Figure 39 (a) Q statistics chart using cPCA in normal condition and fault condition 1 (b) T^2 statistics chart using cPCA in normal condition and fault condition 1.

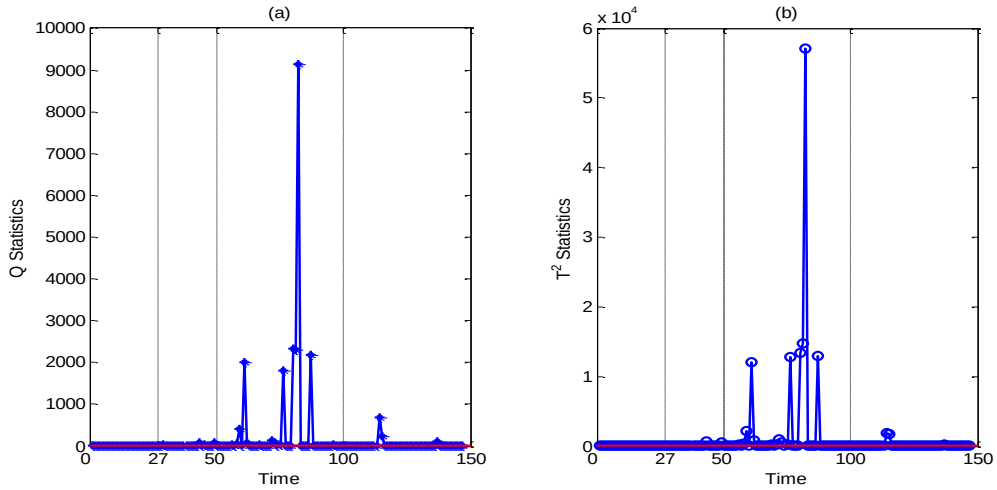

Figure 40(a). Q statistics chart using cPCA in normal condition and fault condition 2 (b) T^2 statistics chart using cPCA in normal condition and fault condition 2.

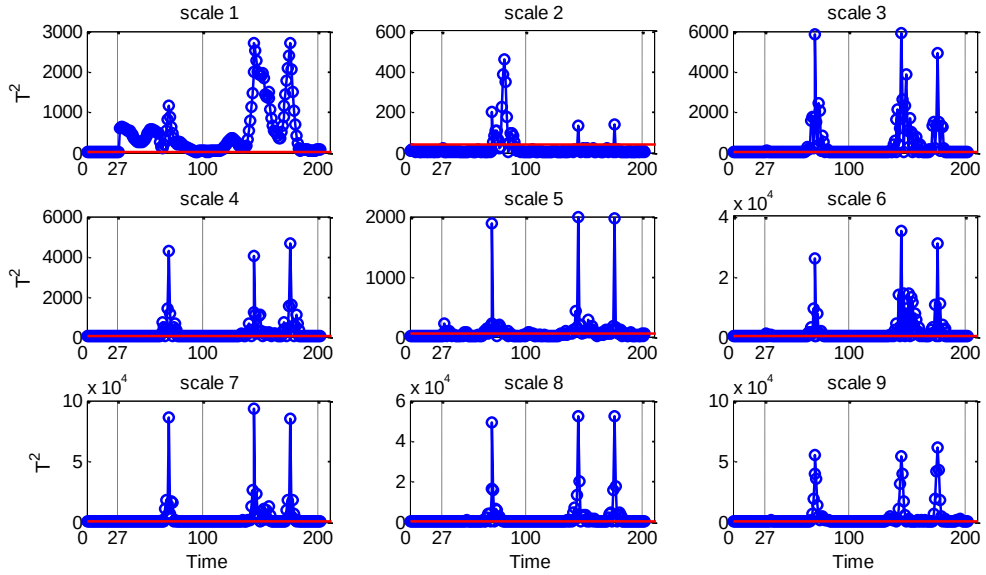

Figure 41(a) T^2 statistics chart using MS-SSA in normal condition and fault condition 1.

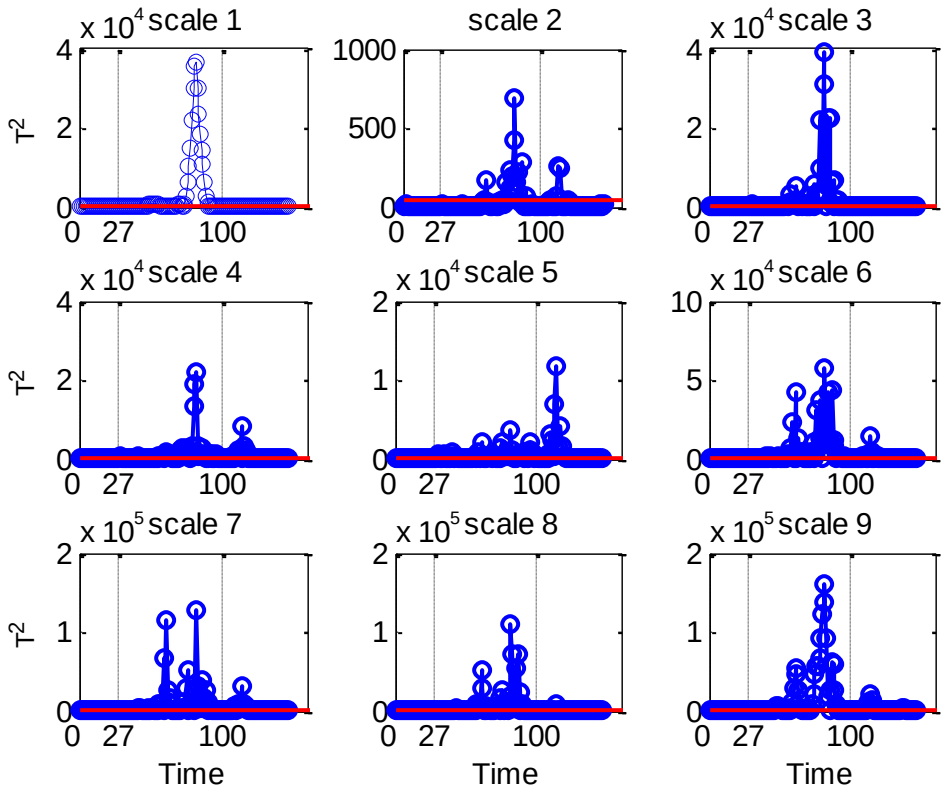

Figure 41(b) T^2 *statistics chart using MS-SSA in normal condition and fault condition 2.*

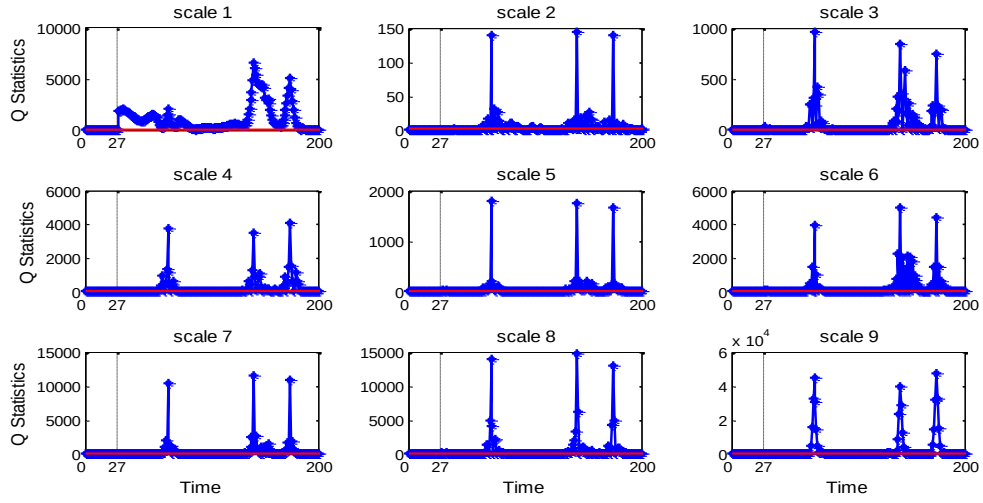

Figure 41(c) Q *statistics chart using MS-SSA in normal condition and fault condition 1.*

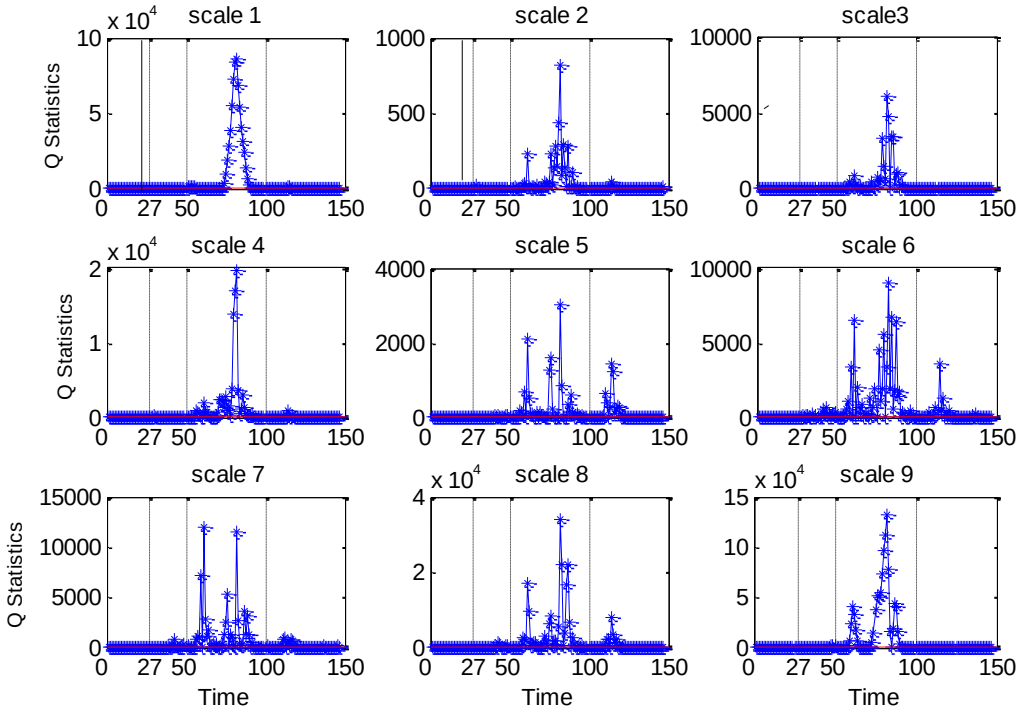

Figure 41 (d) Q statistics chart using MS-SSA in normal condition and fault condition 2.

Table 13(a) The reliability(%) of MS-SSA and PCA approach in fault condition 1.

	MS-SSA (per scale)									**PCA**
	1	2	3	4	5	6	7	8	9	
T^2	100	10.9	34.5	35.1	40.8	58.6	73.6	81	94.8	60.9
Q	100	40.8	57.5	69.5	56.3	82.8	89.1	94.8	99.4	94.3

Table 13(b) The reliability (%) of MS-SSA and PCA approach in fault condition 2.

	MS-SSA (per scale)									**PCA**
	1	2	3	4	5	6	7	8	9	
T^2	98.3	17.7	56.3	70.6	70.6	86.6	90.8	94.9	99.2	70.6
Q	100	61.3	71.4	94.9	85.7	93.3	98.3	99.2	100	80.7

To identify the variables responsible for the deviations in the recovery (variable 13) for fault condition 1 and fault condition 2 cases, the mean contribution of T^2 values and the ratio of residual variances are evaluated as in previous case studies. These contribution plots are shown in Figures 42-44. In fault condition1 and fault condition 2 variable 11 (secondary mill discharge density) provides the largest contribution to the T^2 values in most of the scales with MS-SSA analysis.

Significant contributions for each of the fault conditions were observed from different variables. Variable 11 had the largest contribution among these variables to the fault conditions. The change in variable 11 occurs because of a change in variables in the primary milling circuit. Hence the change in variable 11 caused other variables in the secondary mill to deviate from their normal operating values, thereby affecting variable 13 as well. Since the operations are performed in a closed loop control system the analyzer controller detects these changes in the primary mill variables and makes the corresponding adjustments to other variables in the same unit to counter effects of those changes. Eventually, this results in changes to variable 11 in secondary milling unit so as to facilitate the changes in primary mill variables.

In PCA analysis the T^2 contribution plots of fault conditions 1 and 2 also show variable 11 as the faulty variable with the highest value which contributes most to the observed process shift.

Although both methods identified variable 11 as a faulty variable in the two fault conditions, MS-SSA method performed better than PCA, which can be attributed to the fact that MS-SSA accounted for serial correlation in the variables. It is difficult to draw any conclusive inferences based on the Q contribution plots as all the variables were flagged as faulty in both cases using either MS-SSA (at the fine scales) or PCA approaches. It can be considered that any sudden change that is not explained in the reference model, potentially changes the nature of the relationship between the process variables, as observed by an increase in the value of the Q statistics (Kresta et al., 1991). The contributions of variables in the first scale in both classes also confirm the deviations in the secondary mill

variable as already confirmed in the T^2 contributions. Using PCA, most of the variables were found to be giving mis-identification of fault. One should also bear in mind that contribution plots are not reliable especially when the number of faulty variables is large due to the “smearing” effect (Qin, 2003). Generally the performance of MS-SSA in this study was found to be better than that of cPCA for all fault conditions. .

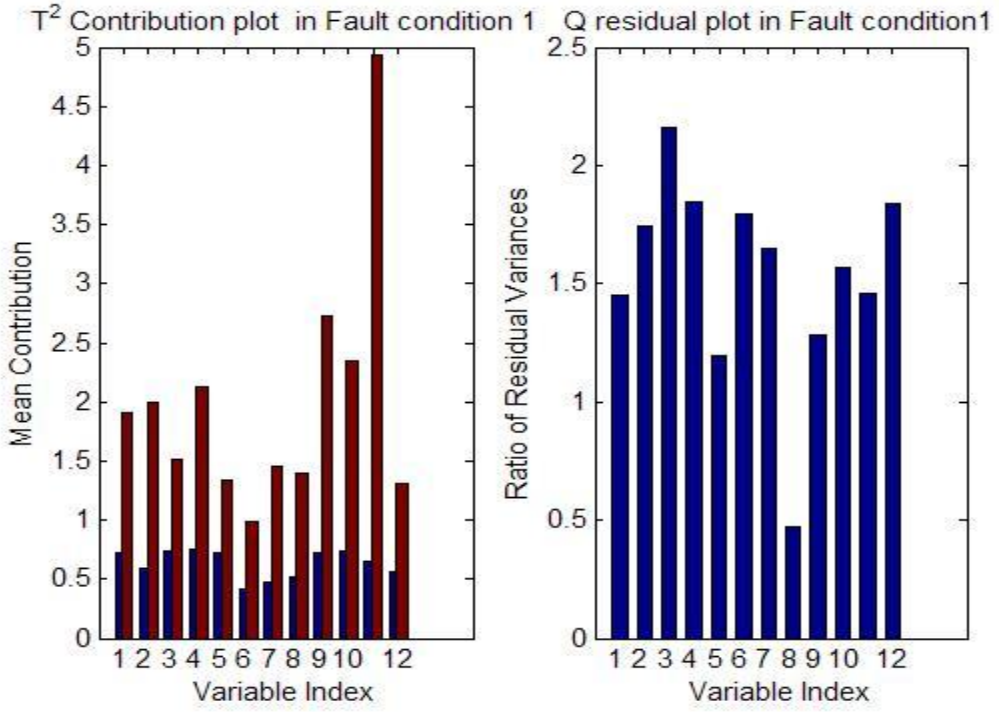

Figure 42 Variables contributing to the deviations in T^2 and Q statistics in fault condition 1 using cPCA.

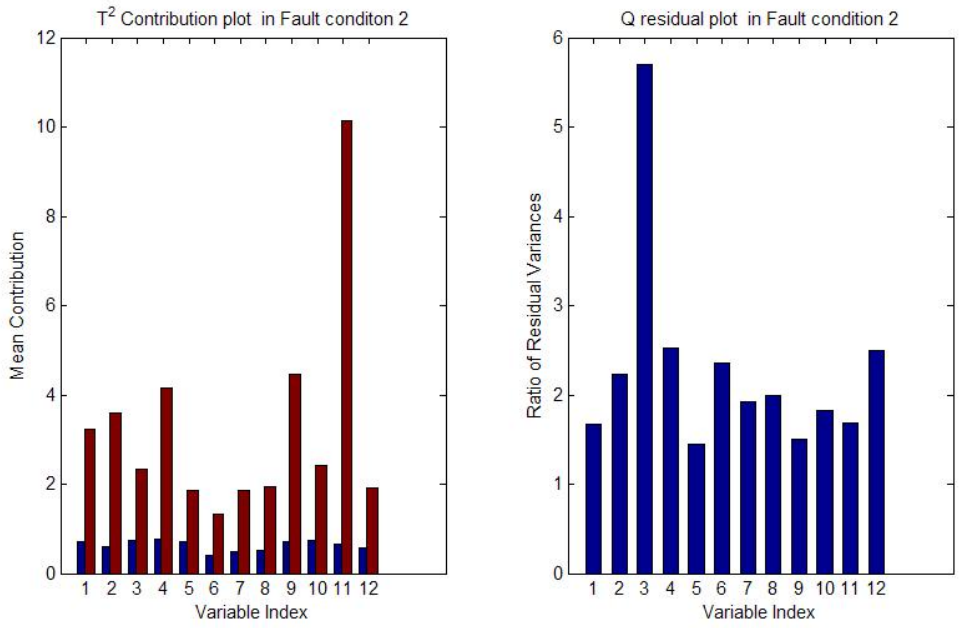

Figure 43 Variables contributing to the deviations in T^2 and Q statistics in fault condition 2 using cPCA.

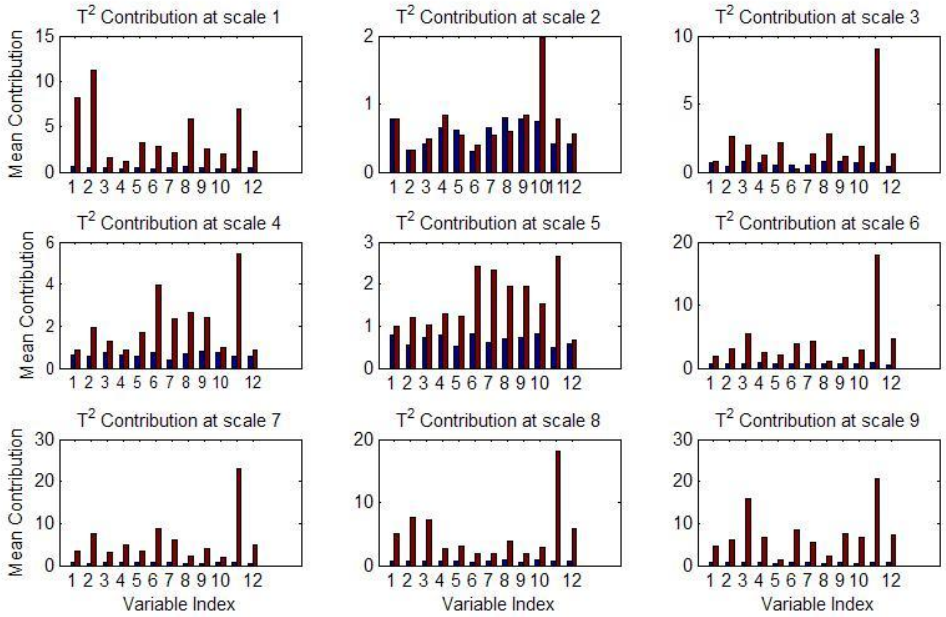

Figure 44(a) Variables contributing to T^2 statistics in fault condition 1 using MS-SSA

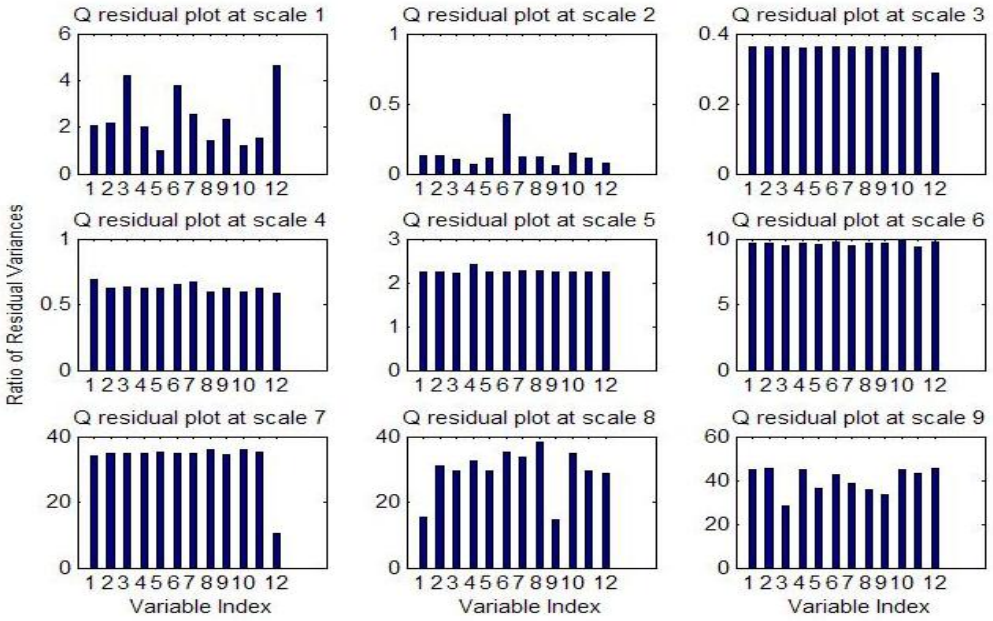

Figure 44(b) Variables contributing to Q statistics in fault condition 1 using MS-SSA

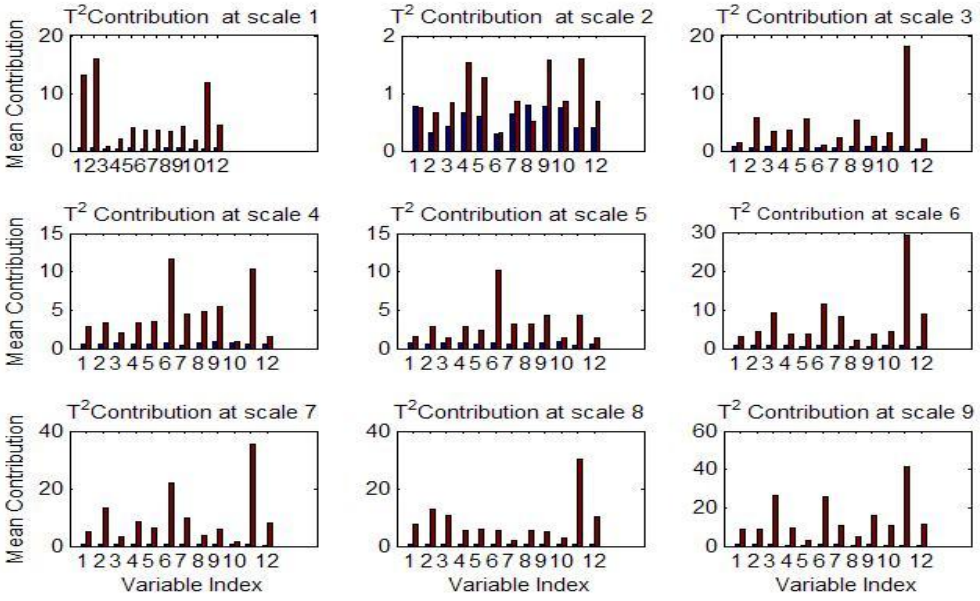

Figure 44(c) Variables contributing to T^2 statistics in fault condition 2 using MS-SSA

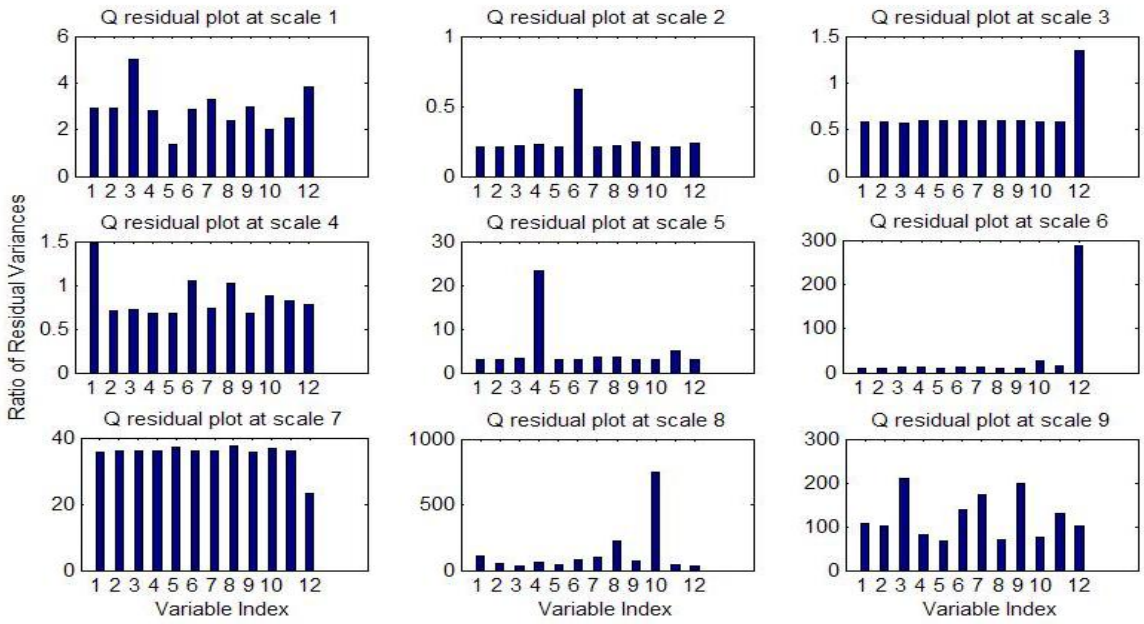

Figure 44(d) Variables contributing to Q statistics in fault condition 2 using MS-SSA.

Chapter 5 Conclusions

In this thesis, a multiscale process monitoring framework using singular spectrum analysis for signal decomposition was proposed as an alternative to existing wavelet-based approaches. The main advantage of the use of singular spectrum analysis is their data adaptive properties. Compared to wavelet techniques, the basis functions are not pre-specified but determined from the data. Decomposing signals on the basis of their singular spectra is equivalent to using wavelets that are constructed from the data themselves, and their shape (basis function) adapted to fit these data accurately. However, SSA explicitly accounts for the autocorrelation in observed process data, and is particularly well-suited for handling short time series as well.

From the simulated and industrial applications MS-SSA, the results show that the proposed MS-SSA algorithm performs better than cPCA and MSPCA in most of the fault conditions investigated, making it a useful addition to SPC process monitoring tools. This method can handle slow and feeble changes in the process signals and, therefore, early detection of faults in monitoring of chemical process in practical situations. With respect to the detection of disturbances and parameter changes in the data, the performance of MS-SSA was not only more successful in early detection of faults but also in the accurate detection of process faults than cPCA. Moreover, being a multiscale approach, MS-SSA provide for accurate fault diagnosis based on the information on which scale the fault is detected.

The multiple scales of the signal obtained in MS-SSA consist of slow varying trend components represented by first few scales and high frequency content represented by last few scales. In most of the applications in this study it was found that the monitoring results obtained by applying PCA in first few scales could identify the process faults since the first few scales carries deterministic components and thus they preserve main information content of the original signal. The high frequency scales carry stochastic components or noisy signals,

which are related to irregular variations in the data. In most of the case studies the variables were observed to violate control limits of the monitored statistics in the last few scales. The multiscale method based on SSA provides good reliability results when it is applied to processes with autocorrelated and/or cross-correlated variables as shown in case study 1 and case study 2.

Unlike other multiscale methods based on wavelets, a key advantage of this method is that it requires the selection of only one parameter, namely the window length. The method is very intuitive in the sense that the data adaptive nature of the basis functions allows a better separation of the signal from noise. From the case studies it was shown that this method also works well for non stationary short time series with autocorrelated measurements.

Finally, another advantage of decomposing the data based on the singular spectra of the signals is that it readily lends itself to non-linear extensions. In principle at least, the effectiveness with which the proposed method detects process deviations can be improved by evaluating the data at the respective scales with nonlinear singular spectrum analysis rather than linear PCA as was the case in this study. The usage of generalized principal components could also be considered to obtain orthonormal bases of functions with multiscale compact supports.

The main drawback is the lack of general theoretical guidelines or fixed rules for selecting the window length. Nevertheless, it was shown that the results were not very sensitive to arbitrarily chosen window lengths below the first point of decorrelation (case study 1). As a recommendation, future studies need to be focused on methods for optimal selection of the window length for SSA decomposition. Also, it is not clear which significant scales should be monitored.

References

Abarbanel, H. D. I. (1997). *Analysis of Observed Chaotic Data* New-York, Springer-Verlag

Aldrich, C. and. Barkhuizen., M. (2003). "Process system identification strategies based on the use of singular spectrum analysis." *Minerals Engineering* 16 : 815-826

Allen, M. R., and Robertson, A.W. (1996). "Distinguishing modulated oscillations from colored noise in multivariate data sets." *Climate Dynamics* 12: 775-784

Alonso, F. J., Castillo Del, J.M., and Pintado, P (2005). "Application of singular spectrum analysis to the smoothing of raw kinematic signals." *Journal of Biomechanics* 38:1085-1092

Alonso, F. J., and Salgado, D.R., (2005). Application of Singular spectrum analysis to tool wear detection using sound signals. *IMechE, Journal of Engineering Manufacture*.219:703-710

Alsberg, B. K., Woodward, A.M., and Kell, D.B. (1997). "An introduction to wavelet transforms for chemometricians: A time - frequency approach." *Chemometrics and Intelligent Laboratory Systems* 37: 215-239.

Alt, F. B. (1982). Multivariate Quality Control: State of the Art. *Transactions of the 1982 ASQC Quality Congress*, Detroit, MI.

Alt, F. B. (1984). Multivariate Quality Control. *The Encyclopedia of Statistical Sciences*: 110-122.

Alt, F. B., and Smith, N.D. (1988). Multivariate Process Control. *Krishnaish and C.R.Rao, Eds*, North-Holland, Amsterdam.

Aradhye, H. B., Davis J. F.,and Bakshi, B. R. (2003). "Multiscale SPC using wavelet-theoretical analysis and properties." *AIChE Journal* 49: 939-958.

Bailey, S.J. (1984). "^{c} rom desktop to plant floor, a C^{o}T is the control operators window on the process." *Control Engineering* 31(6): 86-90

Bakshi, B. R., and Stephanopolos, G. (1994). "Representation of process trends III. Multi-scale extraction of trends from process data." *Computers and Chemical Engineering* 4: 267-302.

Bakshi, B. R. (1998). "Mutiscale PCA with applications to multivariate statistical process monitoring." *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 Nounou,M.N, (2000). Multiscale methods for denoising and compression. in *Wavelets in Chemistry*, Walcza, B (ed), Elsevier, Amsterdam, The Netherlands.

Barkhuizen, M. (2003). Analysis of process data with singular spectrum methods. Process Engineering, *Unpublished Masters Thesis*, University of Stellenbosch, South Africa.

Barkhuizen, M., and Aldrich, C (2003). Dynamic process analysis with non-linear Monte Carlo singular spectrum analysis. *Proceedings of the First African Control Conference*, Cape Town, South Africa.

Bilancia, M., and Stea, G. (2007). "Timescale effect estimation in time-series studies of air pollution and health: a singular spectrum analysis approach." *Electronic Journal of Statistics* 2: 432-453.

Botha, P. J. (2006). Detecting Change in Complex Process Systems with Phase Space Methods. *Masters Thesis*, University of Stellenbosch, South Africa.

Bray, M. A., and Wikswo,J.P. (2002). "Considerations in phase plane analysis for nonstationary reentrant cardiac behavior." *Physical Review E* 65(5):1902-1909.

Broomhead, D. S., and King, G.P (1986). "Extracting qualitative dynamics from experimental data." *Physica D* 20: 217-236.

Carniel, R., Barazza,F.,and Pascolo,P. (2006). "Improvement of Nakamura technique by singular spectrum analysis." *Soil Dynamics and Earthquake Engineering* 26: 55-63.

Chen, G., and McAvovy,T.J. (1998). "Predictive on-line monitoring of continuous processes." *Journal of Process Control* 8(5-6): 409-420.

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

Chiang, H. L., and Braatz, D. R (2003). "Process monitoring using causal map and multivariate statistics: Fault detection and identification." *Chemometrics and Intelligence Laboratory Systems* 65:159-178.

Choi, S. W., Morris, J., and Lee,I. (2008). "Nonlinear multiscale modeling for fault detection and identification." *Chemical Engineering Science* 63: 2252-2266.

Casdagli, M., Eubank, S., Farmer, J D and Gibson, J (1991). "State space reconstruction in the presence of noise." *Physica D* 51(1-3):52-98.

Daubechies, I. (1992). Ten Lectures in Wavelets, *SIAM*, Philadelphia, PA.

Dayal, B. S., and MacGregor, J.F. (1997). "Recursive exponentially weighted PLS and its application to adaptive control and prediction." *Journal of Process Control* 7(3): 169-179.

Dong, D., and McAvoy, T.J (1996). "Nonlinear Principal Component Analysis - Based on principal curves and neural networks." *Computers and Chemical Engineering* 20(1): 65-78.

Donoho, D. L., Johnstone, I.M., Kerkyacharian, G., and Picard,D. (1995). "Wavelet Shrinkage: Asymptopia?" *Journal of the Royal Statistical Society*, B 57(41):301-369.

Downs, J. J., and Vogel,E.F. (1993). "A Plant-wide Industrial Process Control Problem." *Computers and Chemical Engineering* 17(3): 245-255.

Engel, J. (1994). "A simple wavelet approach to nonparametric regression from recursive partitioning" *Journal of Multivariate Analysis* 49(2): 242-254.

Fourie, S. H., and de Vaal, P (2000). "Advanced process monitoring using an on-line non-linear multiscale principal component analysis methodology." *Computers and Chemical Engineering* 24(2-7): 755-760.

Frank, P. M. (1993). Robust model-based fault detection i dynamic systems. *IFAC Symposium Series* Pergamon Press, Oxford.

Ganesan, R., Das, T. K , and Venkataramanan, V (2004). "Wavelet-based multiscale process monitoring: A literature review." *IIE Transactions* 36(20): 787-806.

Ghil, M., and Yiou,P. (1996). Spectral methods: What they can and cannot do for climatic time series In Decadal Climate Variability. *Dynamics and Predictability (Eds,Anderson,D.and Willebrand,J.)* Elsevier, Amsterdam, 445-482.

Ghil,M.,Allen,M.R.,Dettiner,M.D.,Ide,K.,Kondrashov,D.,Mann,M.E.,Robertson, A.W.,Saunders, A.,Tian,Y.,Varadi,F.,Yiou,P.(2002). "Advanced Spectral Methods for Climatic Time Series." *Reviews of Geophysics* 40(1): 1003-1043.

Golyandina, N., Nekrutkin, V, and Zhigljavsky, A (2001). Analysis of time series structure - SSA and related techniques, *Chapman & Hall/CRC Press*, Boca Raton, Florida

Grossman, A., and Morlet, J. (1984). "Decomposition of Hardy functions into square integrable wavelet of constant shape." *SIAM Journal on Mathematical Analysis* 15: 723-736.

Horn, J. L. (1965). "A rationale and test for the number of factors in factor analysis." *Psychometrika* 30: 179-185.

Hunter, J. S. (1986). "The exponentially weighted moving average." *Journal of Quality Technology* 18(4): 203 -209.

Jackson, J. E. (1985). "Multivariate Quality Control." *Communications in Statistics-Theory and Methods* 14(110): 2657-2688.

Jackson, J. E. (1991). A User's Guide to Principal Components, *Wiley*, New York.

Jemwa, G. T., and Aldrich, C (2006). "Classification of process dynamics with Monte Carlo singular spectrum analysis." *Computers and Chemical Engineering* 30: 816-831.

Johnson, R. A., and Wichern, D.W. (1992). Applied Multivariate Statistical Analysis, *Prentice-Hall*, London.

Jollife, I. T. (1986). Principal component Analysis. *Springer*, New York.

Kano, M., Nagao, K., Hasebe, S., Hashimoto, I., Ohno, H., Strauss, R., and Bakshi, B.R. (2000). "Comparison of statistical process monitoring methods." *Computers and Chemical Engineering* 24: 175-181.

Kano, M., Nagao, K., Hasebe, S., Hashimoto, I., Ohno, H., Strauss, R., and Bakshi, B.R. (2002). "Comparison of multivariate statistical process monitoring methods with applications to the Estaman challenge problem." Computers and Chemical Engineering 26: 161-174.

Kano, M., and Nakagawa, Y. (2008). "Data-based process monitoring, process control, and quality improvement: Recent developments and applications in steel industry." *Computers and Chemical Engineering* 32(1-2): 12-24.

Kepenne, C. L. (1995). "An ENSO signal in soybean future prices." *Journal of Climate* 8: 1685-1689.

Kosanovich, K. A., and Piovoso, M. (1997) “PCA of wavelet transformed process data for monitoring.” *Intelligent Data Analysis Journal* 1: 85-99

Kourti, T., Lee, J., and MacGregor, J.F.(1996). "Experiences with industrial applications of projection methods for multivariate statistical process control." *Computers and Chemical Engineering* 20(Suppl): 745-750.

Kourti, T. (2002). Process Analysis and Abnormal Situation Detection: From Theory to Practice. *IEEE Control Systems Magazine* 22(5): 10-25.

Kourti, T. (2005). "Application of latent variable methods to process control and multivariate statistical process control in industry." *International Journal of Adaptive Control and Signal Processing* 19: 213-243.

Kramer, M. A. (1991). "Nonlinear Principal Component Analysis using autoassociative neural networks." *AIChE Journal* 37(2): 233-243.

Kresta, J. V., MacGregor, J. F, and Marlin, T. E (1991). "Multivariate Statistical Monitoring of Process Operating Performance." *The Canadian Journal of Chemical Engineering* 69(1): 35-47.

Kruger, U., Kumar, S., and Littler, T. (2007). "Improved principal component monitoring using the local approach." *Automatica* 43: 1532-1542.

Ku, W., Storer, R. H , and Geogakis, C (1995). "Disturbance Detection and Isolation by Dynamic Principal Component Analysis." *Chemometrics and Intelligent Laboratory Systems* 30(1): 179-196.

Kumar, V., Sundararaj, U., Shah, S.L., Hair, D., and Griend, I.J. (2003). "Multivariate Statistical Monitoring of a High-Pressure Polymerization Process." *Polymer Reaction Engineering* 11 (4):1017-1052.

Laser, M. (2000). “Recent Safety and Environmental Legislation ”, *Process Safety and Environmental Protection*, Transactions of the Institution of Chemical Engineers. 78B: 419-422.

Lee, D. S., Park, J.M., and Vanrolleghem, P.A (2005). "Adaptive multiscale principal component analysis for on-line monitoring of a sequencing batch reactor." *Journal of Biotechnology* 116: 195-210.

Lees, F. P. (1996). Loss prevention in process industries: hazard identification, assessment and control. *Butterworth-Heinmann*. London.

Li, X. (2002). "A brief review: Acoustic emission method for tool wear monitoring during turning." *International Journal of Machine Tools &Manufacture* 42: 157-165.

Lin, W., Qian,Y, and Li,X (2000). "Non-linear dynamic principal component analysis for on-line process monitoring and diagnosis." *Computers and Chemical Engineering* 24(2-7): 423-429.

Luo, R., Misra, M, and Himmelblan ,D .M (1999). "Sensor fault detection via multiscale analysis and dynamic PCA." *Industrial and Engineering Chemistry Research* 38(4):1489-1495.

MacGregor, J.F., Kourti, T., and Kresta, J.V (1991). Multivariate Identification: A Study of Several Methods. *Proceedings of IFAC Symposium ADCHEM*-91. Toulouse, France.

MacGregor, J. F., Jaekle, C, Kiparissides, C, and Koutoudi, C (1994). "Process monitoring and diagnosis by multiblock PLS methods." *AIChE Journal* 40(5):826-838.

MacGregor, J. F., and Kourti, T (1995). "Statistical Process Control of Multivariate Processes." *Control Engineering Practice* 3(3): 403-414.

Malinowski, E. R. (1991). Factor Analysis in Chemistry, *Wiley*, New York

Mallat, S. G. (1989). "A theory for Multiresolution Signal Decomposition: The Wavelet Representation." *IEEE Transaction on Pattern Analysis and Machine Intelligence* 11(7): 674-693.

Mallat, S. G., and Zhong,S. (1992). "Characterizations of signals from multiscale edges." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14: 710-732.

Mamou, J., and Feleppa, E.J (2007). "Singular spectrum analysis applied to ultrasonic detection and imaging of brachytherapy seeds." *Journal of Acoustical Society of America* 121(3): 1790-1801.

Mastrangelo, C. M., Runger, G.C, and Montgomery, D.C (1996). "Statistical process monitoring with principal components." *Quality and Reliability Engineering International* 12: 203-210.

Maulud A, W. D., and Romagnoli J A. (2006). "A multi-scale orthogonal nonlinear strategy for multi-variate statistical process monitoring." *Journal of Process Control* 16(7): 671-683.

Meyer.Y. (1992). Wavelets and Operators, *Cambridge University Press* New York, Cambridge.

Michael, S. (2005). Applied Nonlinear Time Series Analysis: Applications in Physics, Physiology and Finance, *World Scientific Publishing Company*, Singapore.

Mineva, A., and Popivanov, D (1996). "Method for single-trial readiness potential identification, based on singular spectrum analysis." *Journal of Neuroscience Methods* 68: 91-99.

Misra M, Y. H. H., Qin S J, and Ling C (2002). "Multivariate process monitoring and fault diagnosis by multiscale PCA." *Computers and Chemical Engineering* 26(9): 1281-1293.

Montgomery, D C. (1996). Introduction to Statistical Control, *John Wiley and Sons, Inc.*

National Safety Council (1999), *Chemical Engineering Progress* 91(9): 36-45.

Negiz, A., and Cinar,A. (1997). "Statistical monitoring of multivariable dynamic processes with state-space models." *AIChE Journal* 43(8): 2002-2012.

Nimmo, I. (1995). "Adequately address abnormal situation operations." *Chemical Engineering Progress* 91(9): 36-45.

Nomikos, P., and MacGregor, J.F (1995). "Multivariate SPC charts for Monitoring Batch Processes." *Technometrics* 37(1): 41-59.

Ogden, T., and Parzen, E. (1996). "Data dependent wavelet thresholding in non-parametric regression with change point application." *Computational Statistical Data Analysis* 22: 53 - 70.

Ogunnaik, B. A., and Ray,W.H (1994). Process Dynamics, Modeling, and Control, *Oxford University Press*, New York.

Ormerod, P., and Campbell,M (1997). Predictability and economic time series in C.Heij, J.M. Schumacher, B. Hanzon, C. Praagman, eds., *System Dynamics in Economic and Financial Models, John Wiley& Sons.*

Palomo, M. J., Sanchis,R., Verdu,G., and Ginestar,D (2003). "Analysis of pressure signals using a singular spectrum analysis (SSA) methodology." *Progress in Nuclear Energy* 43(1-4):329-336.

Pereira, W. C., and Maciel, C.D (2001). "Performance of ultrasound echo decomposition using singular spectrum analysis." *Ultrasound in Medicine and Biology* 27(9):1231-1238.

Pereira, W. C., Bridal, S.L., Coron,A., and Laugier, P (2004). "Singular spectrum analysis applied to backscattered ultrasound signals from in vitro human cancerous bone specimens". *IEEE Transaction on Ultrason Ferroelectrics Frequency Control 51(3):302-312.*

Plaut, G., and Vautard,R. (1994). "Spells of Low-Frequency Oscillations and Weather Regimes in the Northern Hemisphere." *Journal of the Atmospheric Sciences* 51(2):210-236.

Polikar, R. (1996). The Wavelet Tutorials Part 111 Multiresolution Analysis & the Continuous Wavelet Transform, *https://ccrm.stanford.edu/WTpart/ html*.

Qin, S.J. (2003). "Statistical Process monitoring: basics and beyond". *Journal of Chemometrics* 17 480-502.

Ralston, P., DePuy, G., and Graham, J.H., (2004). "Graphical enhancement to support PCA-based process monitoring and fault diagnosis." *ISA Transactions* 43: 639-653.

Reis, M. S., and Saraiva, P.M (2006). "Multiscale Statistical Process Control with Multiresolution Data." *AIChE Journal* 52(6): 2107-2119.

Rosen, C., and Lennox, J.A (2001). "Multivariate and multiscale monitoring of waste water treatment operation." *Water Res*. 35(14): 3402-3410.

Rozynski, G., Larson,M., and Pruszak, Z (2001). "Forced and self- organized shoreline response for a beach in the southern Baltic Sea determined through singular spectrum analysis." *Coastal Engineering* 43(1): 41-58.

Sadler, B. M., and Swami,A. (1999). "Analysis of Multiscale Products for Step Detection and Estimation." *IEEE Transaction on Information Theory* 45: 1043-1051.

Salgado, D. R., and Alonso, F.J (2006). "Tool wear detection in turning operations using singular spectrum analysis." *Journal of Materials Processing Technology* 171(3): 451-458.

Salgado, D. R., and Alonso, F.J (2007). "An approach based on current and sound signals for i-process tool wear monitoring." *International Journal of Machine Tools &Manufacture* 47(14): 2140-2152.

Saucier, A. (2005). "Construction of data-adaptive wavelet bases with an extension of principal component analysis." *Applied and Computational Harmonic Analysis* 18: 300-328.

Schoellhamer, D. H. (2001). "Singular spectrum analysis for time series with missing data." *Geophysical Research Letters* 28(16): 3187-3190.

Schuster, H. G. (1988). Deterministic Chaos: an Introduction. *Weinheim, Physik Verlag.*

Shao, R., Jia, F, Martin, E. B, and Morris, A. J (1999). "Wavelets and nonlinear principal component analysis for process monitoring." *Control Engineering Practice* 7(7): 865-879.

Shaw, R. S. (1984). The dripping faucet as a model chaotic system, Science Frontier Express Series, *Aerial Press* Santa Cruz, CA.

Sjoberg,.A.,Zhang,Q.H.,Ljung,L.,Benvensite,A.,Delyon,B.,Glorence,P.Y.,Hjalmarsson,H.,and Juditsky,A. (1995). "Non-linear black box modeling in system - identification: A unified overview." *Automatica* 31: 1691-1724.

Song, F., and Jutamulia, S. (2000). Wavelet transform and its use in edge detection. *In Proceedings of SPIE-The international Society for Optical Engineering*, SPIE, Bellingham, WA, Wash, B(ed).

United States Bureau of Labor Statistics (1998). Occupational injuries and illnesses in the United States by industry. *Government Printing Office,* Washington DC.

Stephanopoulos, G., Karsligil, O., Dyer, M. S. (2008). "Multiscale theory for linear dynamic processes Part 1. Foundations." *Computers and Chemical Engineering* 32: 875-884.

Stockwell, D. R. B. (2007). Niche Modeling Predictions from Statistical Distributions. *Chapman&Hall/CRC,* New York.

Sun, W., Palazoglu, A., and Romagnoli, J.A (2003). "Detecting abnormal process trends by wavelet-domain hidden Markov models." *AIChE Journal* 49(1): 140-150.

Tan, S., and Mavrovouniotis, M. L (1995). "Reducing data dimensionality through optimizing neural network inputs." *AIChE Journal* 41(6): 1471-1480.

Tangirala, A. K. (2001). Multirate Control and Multiscale Monitoring of Chemical Processes, *PhD. Thesis*, University of Alberta, Canada.

Teppola, P., Mujunen, S., Minkkinrn, P., Puijola, T., and Pursiheimo, P. (1998). "Principal component analysis, contribution plots and feature weights in the monitoring of sequential data from a paper machine's wet end." *Chemometrics and Intelligent Laboratory Systems* 44(1-2): 307-317.

Teppola, P., and Minkkinen, P. (2000). "Wavelet-PLS regression models for both exploratory data analysis and process monitoring." *Journal of Chemometrics* 14: 383-399.

Thomakos, D. D., Wang, T., Wille, L.T. (2002). "Modeling daily realized futures volatility with singular spectrum analysis." *Physica A* 312(3-4): 505-519.

Top, S., and Bakshi, B.R. (1998). Improved statistical process control using wavelets. *Proceedings of the Foundations of Computer Aided Process Operation*. Snowbird, UT.

Tracy, N. D., Young ,J .C , and Mason, R .L (1992). "Multivariate control charts for individual observations." *Journal of Quality Technology* 24(2): 88-95.

Trygg, J., Kettanech-Wold, N.,and Wallbacks, L (2001). "2D wavelet analysis and compression of on-line industrial process data." *Journal of Chemometrics* 15: 299-319.

Tsonis, A. A. (1992). Chaos: from Theory to Applications, *Plenium Press*, New York.

Tzagkarakis, G., Papadopouli,M.,Tsakalides,P. (2007). Singular Spectrum Analysis of Traffic Workload in a Large-Scale Wireless LAN. *MSWIM'07*. Chania, Crete Island, Greece.

Uraikul, V., Chan, C W., and Toniwachwuthikul, P. (2007). “Artificial intelligence for monitoring and supervisory control of process systems.” *Engineering Applications of Artificial Intelligence*.20: 115-131

Vautard, R., and Ghil, M (1989). "Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series." *Physica D* 35: 395-424.

Vautard, R., Yiou, P, and Ghil,M (1992). "Singular -spectrum analysis : A toolkit for short, noisy chaotic signals." *Physica D* 58(1-4): 95-126.

Venkatasubramanian, V. (2005). "Prognostic and diagnostic monitoring of complex systems for product lifecycle management: Challenges and opportunities." *Computers and Chemical Engineering* 29(6): 1253-1263.

Vitanov,N.K.,Sakai,K.,and Dimitrova, Z.I.(2008). "SSA,PCA,TDPSC,ACFA:Useful combination of methods for analysis of short and nonstationary time series." *Chaos Soliton and Fractals* 37(1):187-202.

Wang, D., and Romagnoli, J.A (2005). "Robust multi-scale principal component analysis with applications to process monitoring." *Journal of Process Control* 15: 869-882.

Wang, W. J., Chen, J., and Wu, X.K (2001). "The application of some non-linear methods in rotating machinery fault diagnosis." *Mechanical Systems and Signal Processing* 15(4).

Wang, Y. (1995). "Jump and sharp cusp detection by wavelets." *Biometrika 82 (2)*: 385-397.

Western Electric Company (1956). Statistical Quality Control *Handbook*, Western Electric Corp., Indianapolis, IN: 25-28.

Wise, B. M., Veltkamp, D.J., Ricker,N.L., Kowalski, B.R., Barnes, S.M., and Arakali, V(1991). Application of Multivariate Statistical Process control (MSPC) to the West Valley Slurry-Red Ceramic Melter Process, *Proceedings of .Waste Management.* Tuscon, AZ.

Wold,S.,Albano,C.,DunnIII,W.J.,Edlund,O.,Esbensen,K.,Geladi,P.,Hellberg,S.,Johansson, E.,Lindberg,W.,and Sjostrom,M (1984). "Multivariate Data Analysis in Chemistry." *Chemometrics; Mathematics and Statistics in Chemistry*.17-95.

Wold,S.,(1994). “Exponentially weighted moving principal component analysis and projection to latent structures.” *Chemometrics Intelligent Laboratory Systems* 23(1): 149-161.

Wold, S., Kettane, H.N, and Tjessem, K (1996). "Hierarchical multiblock PLS and PC model for easier model interpretation and as an alternative to variable selection." *Journal of Chemometrics* 10(5-6): 463-482.

Wu, H., and Gong,J. (2000). Forecast of Network Behavior Based on Singular - Spectrum Analysis. *IEEE Proceedings 1: 702-706.*

Yang, M., Yao, T., Wang,H., and Gou, X. (2006). "Climatic oscillations over the past 120kyr recorded in the Guliya ice core, China." *Quaternary International* 1154-155: 11-18.

Yiou, P., Baert, E and Loutre, M.F (1996). "Spectral Analysis of Climate Data." *Surveys in Geophysics* 17(6): 619-663.

Yiou, P., Sonette, D and Ghil, M (2000). "Data-adaptive wavelets and multiscale singular spectrum analysis." *Physica D* 142(3-4): 254-290.

Yoon, S., and MacGregor, J.F (2004). "Principal -Component Analysis of Multiscale Data for Process Moitoring and Fault Diagnosis." *AIChE Journal* 50(11): 2891-2983.

Zhang, J., Martin, E, and Morris, A.J (1995). Fault detection and classification through multivariate statistical techniques. In Proceedings of American Control conference, Piscataway, N J, *IEEE Press*.

Zhang, Y., and Dudzic, M.S. (2006). "Online monitoring of steel casting processes using multivariate statistical technologies: From continuous to transitional operations." *Journal of Process Control* 16: 819-829.

Zhinqiang, G., and Qunxiong, Z (2005). "Multiscale Nonlinear Principal Component Analysis (NLPCA) and its Application for Chemical Process Monitoring." *Industrial & Engineering Chemistry Research* 44(10): 3585-3593.