ACTIVE STATISTICAL PROCESS CONTROL

By

Kamarul 'Asri Ibrahim B.Sc. (Ohio), M.Sc. (Colorado State)

A thesis submitted to the Faculty of Engineering, University of Newcastle upon Tyne, as partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Department of Chemical and Process Engineering
University of Newcastle upon Tyne
March 1996.

Abstract

Most Statistical Process Control (SPC) research has focused on the development of charting techniques for process monitoring. Unfortunately, little attention has been paid to the importance of bringing the process in control automatically via these charting techniques. This thesis shows that by drawing upon concepts from Automatic Process Control (APC), it is possible to devise schemes whereby the process is monitored and automatically controlled via SPC procedures. It is shown that Partial Correlation Analysis (PCorrA) or Principal Component Analysis (PCA) can be used to determine the variables that have to be monitored and manipulated as well as the corresponding control laws.

We call this proposed procedure **Active SPC** and the capabilities of various strategies that arise are demonstrated by application to a simulated reaction process. Reactor product concentration was controlled using different manipulated input configurations *e.g.* manipulating all input variables, manipulating only two input variables, and manipulating only a single input variable. The last two manipulating schemes consider the cases when all input variables can be measured on-line but not all can be manipulated on-line. Different types of control charts are also tested with the new Active SPC method *e.g.* Shewhart chart with action limits; Shewhart chart with action and warning limits for individual observations, and lastly the Exponentially Weighted Moving Average control chart. The effects of calculating control limits on-line to accommodate possible changes in process characteristics were also studied.

The results indicate that the use of the Exponentially Weighted Moving Average control chart, with limits calculated using Partial Correlations, showed the best promise for further development. It is also shown that this particular combination could provide better performance than the common *Proportional Integral* (*PI*) controller when manipulations incur costs.

In the Name of God, the Most Benevolent and the Most Merciful

Acknowledgements

Firstly, I would like to thank my supervisor Dr. Ming T. Tham for all his input, encouragement and guidance during my dissertation work. This includes the original impetus to pursue this line of research, and also for all his valuable suggestion for making my written presentation clearer and concise. His insightful comments kept me going when I thought I was not getting anywhere. Hopefully the student - supervisor relationship will become an everlasting friendship. Not to forget, Dr. Irini Efthimiadu who helped me a lot during the early stages of my work. She told me of the do's and the don'ts and what to expect during the later part of this study. Her experience was really appreciated because she shielded me from some of the potential pitfalls during the research.

To my late mother Kamsiah Md. Taib who could not live to see my success, I express my sincere gratitude for giving me love and encouragement to strive very hard in order to fulfil the basic necessities of life. She thought me the value of knowledge and to take advantage when opportunities come. May God bless her. To my father Mr. Ibrahim Abu Bakar who always provided the support, both financial and emotional, I would like to express my honest appreciation. This accomplishment would not have been achieved without their incredible support.

To my beloved wife, Tini Zainuddin, who believed in me and my goals without reservation, I really want to express my deep heartfelt appreciation. She always encouraged me to reach my highest potential and stood beside me when no one would, I really treasure her sacrifice, courage and support especially during those lonely and difficult periods. To my children Asma, Imran, Iman and Irwan who are always a joy and an inspiration in my life, I would like to express my appreciation for their understanding ot the nature of my work. To all my friends who prayed for my success in this life and the world here after, I sincerely thank them from the bottom of my heart.

I also wish to acknowledge the financial support from the Commonwealth Scholarship Commission and British Council. They generously sponsoring my PhD studentship and also providing maintenance for my family. Last but not least, to Universiti Teknologi Malaysia, my employer, I treasure their support in term of the promotion, paid study leave and housing allowance while in the United Kingdom. Lastly, thank you to those who have contributed directly or indirectly to the completion of my work in the University of Newcastle upon Tyne.

Table of Contents

List of Figures

Figures **page**

List of Tables

Table **page**

Chapter 1

1.1 Introduction

In recent years, the importance of quality has become increasingly apparent. Stiffer competition, tougher environmental and safety regulations, and rapidly changing economic conditions have been key factors in tightening plant product quality. Virtually everyone in industry agrees that consistent high quality is an essential ingredient, and possibly the single most important production performance criterion. Any tool that can provide a measure of performance on a process is appreciated. Subsequently it can become an integral addition towards quality awareness on the production floor. Several methods of "process control" have been implemented in industry to fulfil these quality awareness requirements. There are actually two ways we can describe the terminology of process control. Firstly, from the view point of the manufacturing industry that use Statistical Process Control (SPC) charts to monitor the parts or products that are being produced. Secondly from the process or chemical industries where they employ various forms of feedback and feedforward strategies for process adjustment in what is often called Automatic Process Control (APC). Although the two methods come from different backgrounds, their aims are identical. However, it is not until recently that both methods are used concurrently. This thesis looks at combining traditional SPC procedures and APC strategies via multivariate statistical analysis techniques.

1.2 Overview of the research work

The concepts and methodology of SPC are totally different from those of Automatic Process Control (APC). However, the objective of both methods is the same, attempting to reduce the variability in the process, although the respective procedures for achieving this are quite different. SPC through monitoring the process, seeks the removal of the root cause while APC counteracts variability by adjusting different variables and transfers the variability into these less important manipulated variables.

The SPC methodology has been developed for monitoring the product of manufacturing industries, where the product is in the form of discrete items. When SPC methods give an indication that the process is out-of-control, the machine is stopped. The process of rectifying and identifying is done on the machine to eliminate the root causes of the problem. When all these problems are solved, the machine is started again to produce the discrete items. On the other hand, processes in the chemical industries are continuous systems where they could not afford to shut down their process as often as in the manufacturing industries. There should be better ways to overcome this problem when we want to utilise the SPC methodology on chemical processes. We propose a new scheme whereby the input or manipulated variables are monitored and controlled when they deviate away from their respective targets, instead of just monitoring the quality variables using the SPC control charts. All these control laws are based on using the SPC control charts methodology. This is possible by translating the control limits of quality variables to the prescribed manipulated variables. The statistical relationships between the quality variables and the manipulated variables will provide the basis for developing the control laws. The rationale here is that if all the manipulated variables are in statistical control, then the quality variable will follow suit.

In modern process industries, data loggers and sensors are extensively utilised to monitor the input and the output variables. The data is then passed to the microprocessor based controller at a very short interval of time. Thus, an abundant amount of data is being collected for the process input variables (x_i), for instance process flowrates, temperatures and pressures. Final product quality (x_k) such as concentration and polymer properties are available much less frequently, basically through off-line laboratory analysis. Since classical SPC methodology only monitors few quality variables, the technique is inadequate for chemical processes. They ignore the existence of relationships between the input variables (x_i) and the quality variables (x_k). Certainly, if we are going to design an effective scheme for monitoring and diagnosing the performance of the process, we must be able to use all the data, (x_i) and (x_k) to extract the pertinent information. Nevertheless some of the variables are dependent on each other. Only a few underlying events affect the process at any given time. The rest are simply different reflections of these perturbed events. Consequently, examining the variables' one at a time, as if they are a separate events and independent of each other, makes interpretation and diagnosis difficult. Only multivariate statistical methods can treat the data simultaneously and can extract the information about how the variables are behaving relative to each other. In this work, we utilise two types of multivariate methods, Principal Components Analysis (PCA) and Partial Correlation

Analysis (PCorrA), for process monitoring and diagnosis. Both methods also provide the control laws needed for the new procedure called Active SPC.

1.3 Summary of Research Objective

On the basis of the preceding discussion, the objective of this research is as follows:

1.3.1 Objective

To study in detail and develop a procedure known as **Active SPC** that applies Automatic Process Control (APC) concepts within the realm of Statistical Process Control (SPC). The new procedure can overcome the weaknesses and combines the strengths of APC and SPC procedures. The weakness of APC strategies is due to it continuously making attempts to remove the effects of any disturbance on the output by adjusting the manipulated variables. In contrast, traditional SPC only gives an indication for an action to be taken, when the quality variable exceeds some specified limits on the control chart. However, it is possible to devise a method whereby the process is monitored and automatically controlled while maintaining the SPC policy of non-intervention when the process is in state-of-statistical control. In addition to product quality, input or manipulated variables are also monitored. If these inputs can be kept within their respective control limits, then the quality variable should also be maintained within its control limits. Obviously, the former must be related to the latter limits.

1.3.2 Sub-objectives

The objective above is attained by fulfilling the following sub objectives:

1. By developing a non-linear model of a Continuous Stirred Tank Reactor (CSTR) reversible reaction process. The model is developed from the first principle by using the dynamic mass and heat balance equations. The model is then digitally simulated on the UNIX based workstation utilising PASCAL language. This will be used to evaluate all control strategies.

2. By evaluating the effectiveness of the APC feedback control schemes on the simulated CSTR utilising the Proportional Integral (*PI*) controller when the process is subjected to noise variations. The result will be used as a basis for comparing the APC schemes.

3. Evaluate the use of Partial Correlation coefficients or *correlation* coefficients derived from Principal Component Analysis between the input variables and the quality variable in the Active SPC schemes. The usage of these coefficients serves two purposes, firstly to determine and select the most influential manipulated variables for pairing with controlled or quality variable and secondly to determine and define the limits for the SPC charts used to control the process. They are also used to define the control laws, *i.e.* the manipulations that have to be made to the relevant manipulated variables to keep the quality or controlled variable under statistical control.

4. Evaluate the effectiveness of the various possible Active SPC control schemes, by applying it to the simulated CSTR when the process is subjected to noise variations. The following modified control charts were considered:

   (a) Shewhart individual chart with action limits
   (b) Shewhart individual chart with both action and warning limits
   (c) Exponential Weight Moving Average (EWMA) control chart.

1.4 Contributions

This work presents an overview of both traditional and new statistical process control methods for monitoring and diagnosing process operating performance. Multivariate statistical methods such as Principal Component Analysis (PCA) and Partial Correlation Analysis (PCorrA) form the core technique for this new Active SPC method. The utilisation of input variables (x_i) as well as quality variable (x_k) is illustrated and studied. The methods to integrate the multivariate statistical methods in Active SPC are discussed. The performances of the new strategies is then apply to several types of control charts, *i.e.* Shewhart and EWMA charts.

The combination of classical control systems theory and Statistical Process Control scheme known as Active SPC procedure was developed. This study is the first of its kind. Successful completion will not only meet the objective described in the preceding section, but it could provide an excellent indicator for the future direction of combining SPC methods and APC strategies in the process and manufacturing industries. In some respects, this research provides a starting point from which a control engineer may become fundamentally involved with statisticians in industry.

1.5 Organisation of the thesis

The remainder of this thesis is divided into seven chapters. Chapter 2 introduces Statistical Process Control (SPC) strategies and methodologies. It defines the requirements and the assumptions made in implementing the strategies. It explains the sources and the types of variations in the process. It discusses the fulfilment of the normality assumption and the risks involved when making decisions using SPC control charts. It also elaborates on the two types of control charts that are normally employed in the industries, *i.e.* the Shewhart control chart for individual measurements and the Exponential Weight Moving Average (EWMA) control chart.

Chapter 3 explores the existing literature on the integration of SPC and APC methods. First, we review the concept of SPC and APC methods. Then, the two methodology is compared and lastly the integration of SPC and APC strategies is discussed.

Chapter 4 develops a dynamic mathematical model of a continuous stirred tank reactor (CSTR), by applying the un-steady state mass and energy balances. Control performance using a *Proportional-Integral* (*PI*) controller was assessed by application to simulations on the CSTR. The parameters of the *PI* controller were tuned using the *Process Reaction Curve* technique.

Chapter 5 presents an overview of multivariate statistical techniques. The aim of this section is to present the concepts and methods of multivariate analysis at a level that is understandable to the readers of this thesis. Two types of multivariate statistical tools of multivariate analysis are explored, namely Principal Component Analysis (PCA) and Partial Correlation Analysis (PCorrA). Both methods will be used to calculate the control limits of the Active SPC control charts.

Chapter 6 develops the strategies for integrating SPC and APC methods. It gives an overview of the proposed Active SPC strategies and the details of how they may be realised. The Active SPC schemes are outlined in detail including the procedure of process monitoring, control-rule design and implementation on the simulated CSTR process.

Chapter 7 discusses the results of several rigorous simulations studies of several potential Active SPC strategies. A variety of methods are studied and implemented. The performances of each method are compared so that we can determine the effectiveness of the respective strategies. This includes methods for off-line and on-line updating of the control limits of the SPC charts, the usage of several types of control charts, several configurations of manipulated variables and lastly, the effectiveness of implementing the two multivariate data analysis procedures. Then, we consider the use of different data sets for the design of the control charts. Their effects on control performances are discussed. Lastly the performances of Active SPC methods were compared with APC method.

Chapter 8 summarises the findings and recommends future areas of research in this area.

Chapter 2

Statistical Process Control

2.1 Introduction

This chapter introduces the concept of Statistical Process Control (SPC). It discusses the sources and the types of variations that occur in a process. Then it elaborates on the normality requirements in the data, the SPC tools, *i.e.* the control charts that will be used in this work and finally, hypothesis testing.

In recent years it has been recognised that good quality product is crucial for production because quality and profitability go hand in hand. Poor product quality is expensive, because it has to be de-valued, re-processed or dumped as waste. Thus, it always costs less to produce the product right for the first time. The characteristic of a good process requires that a consistent output is being produced. It would be nice if this could be achieved, conclusively, by carefully adjusting the process equipment and allowing them to run continuously. Unfortunately, this would rarely, if ever result in uniform product because in practice, extraordinary precautions are needed to ensure regularity in the products.

Thus, to attain uniformity in the products, normally, the manufacturing industries employed a technique called Statistical Process Control (SPC) to monitor, detect and eliminate the substandard materials and counter productive operations in the process. By monitoring various steps in the process using SPC methods, abnormal trends can be identified and problems can be solved before they get out of hand. Manufacturers who implement and use SPC techniques on a regular basis can greatly reduce the production costs through avoiding the recycling of chemicals, post production blending, adding additional separation processes and abstaining from dumping the product as waste. Whether the quality of the product is determined by functionality, by durability or by appearance, quality must be built into the product, not added as an afterthought. Thus, the main objective of SPC methodologies is to regulate the process and to maintain the standard of the product to the customers' satisfaction. As a result, it allows manufacturing industries to produce the output at a cost that will fulfil profit objectives.

2.2 Sources of Variations in the Process

As mentioned, poor quality in the product is due to the inconsistency and variability in the process. The variability becomes evident whenever the quality characteristic of the product is measured. The causes of this variability are measurement errors, different methods of sample taking, variations introduced by raw materials, various methods of operation and control in the plant, machinery, equipment, different skill levels of operators, environment, planning and etc. It is impossible to produce an exact replicate of an item, especially, when there is the slightest inconsistency in the process. Clearly, the reason for this inconsistency in the output is due to the discrepancy in the input, not only in materials and parts, but also in assembly and operators.

There are basically three sources of variation as noted by Taguchi (1989). The identification of the sources of variation is important, as we can then formulate and implement the appropriate corrections. The first source is due to "outer noise", referring to external sources of variations that influence the production environment. Examples of outer noise include ambient temperatures, humidity, and vibrations from nearby equipment. Second, is "inner noise" where the internal characteristic of the product is changed because of mechanical wear and ageing in the production equipment. Finally, there is "variational noise", where the product parameter varies from one batch to another as a result of the production process. The forces of outer noise and inner noise can be dealt with effectively through engineering design. However, variational noise is a feature of production imperfection and hence it can be dealt with in part at the process level via statistical process control.

In the SPC literature, variational noise can further be categorised into two classes namely "common cause" and "assignable cause". The former, is also known as random, common, system or chronic cause. It is simply a term used to describe faults inherent in the process. Apart from physically altering the nature of the process, there is nothing else that we can do to remove common causes. A process when affected only by common causes, is normally said to be in a state of *statistical control*, where unnecessary action should not be taken to avoid spending time and money on rectifying a well-behaved system.

The second type of variability manifests as temporary deviations from target, signalled by outliers or unusual patterns of points. This type of variability is sometimes known as assignable or special or sporadic causes and often contributes toward a large part of overall variability in the process. Evidence of this type of variation offers important opportunities for improving the uniformity of product. The quality product *mean* (μ) may change gradually as a result of gradual changes in temperature, reactants concentration, or operator fatigue. This could happen in the case of two operators or equipment performing at different operating conditions (Box, 1993; Deming, 1982; Juran, *et al.*, 1974; Shewhart, 1931).

Assignable causes create abnormal variations in the process. It indicates that something has gone wrong with the system. In this incident, the process is in an *out-of-control* state. Consequently the search and the elimination of assignable cause should then be under-taken. This second type of cause must be studied thoroughly by various techniques of data analysis with the aim of separating it from the common causes.

Table 2.1 summarises the basic nature of process variation and provides some examples in each case. In the case of common causes it is clear by their nature that management must take the responsibility for their removal. For example, only management can take the necessary action to improve training and supervision of the workers. Through the decision of management, new methods or procedures can be established for the process. An understanding about the nature of the variability in the process is important because it will determine the kind of action that is necessary and whom to assign the responsibility for taking actions.

Table 2.1 The nature of faults in the Process

	All processes are subject to two fundamental problems	
Faults as used in the literature.	Local faults Special causes Sporadic problems Assignable causes	System faults Common Cause Chronic problems Chance cause
Examples	Broken tools Equipment malfunction Material contamination	Poor supervision Poor training Inappropriate method
Action by Whom	Correctable locally at the equipment or process by the operator.	Require a change in the system. Only management can specify and implement the change.

2.3 Normal Distribution

The occurrence of variability in the process can be plotted on charts called Frequency Distribution Curves. Frequently, the variable data from most manufacturing and process industries follow a Gaussian or "normal distribution" (Wetherill and Brown, 1991). Two parameters ordinarily define a normally distributed population, namely the *mean* (μ) and the variance (σ^2) and the shorthand notation for a normally distributed population is given by $N(\mu,\sigma^2)$. The properties of the normal distribution can be summarised as follows:

(1) Its probability density function (PDF) is symmetric about the population.
(2) Its *mean*, median and mode are identical.
(3) Its cumulative distribution function (CDF) is completely determined by the *mean* and *standard deviation.*
(4) Any conforming set of data, no matter what is the source, can be translated to a single universal form known as the standard normal distribution (SND).

Mathematically the normal distribution function is defined as:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < \infty \tag{2.1}$$

The visual appearance of this normal distribution is a symmetric, unimodal or bell shaped curve and is shown in Figure 2.1. Many measurements of continuous variables follow this distribution. This is explained by the central limit theorem, which states that the most extreme non-normal distribution of data, will approach a normal distribution as the number of collected samples increases (Jaehn, 1989). Because of this result, we often find good fits of practical data by the normal distribution.

The above function $f(x)$ can be scaled so that the total area under the curve over the full range of x ($-\infty < x < \infty$) equals to 1.0. If the curve is to be divided up according to the number of *standard deviation* from the *mean*, we find that each proportion is actually equivalent to the probability of obtaining some values in each zone when the samples are taken at random from the population.

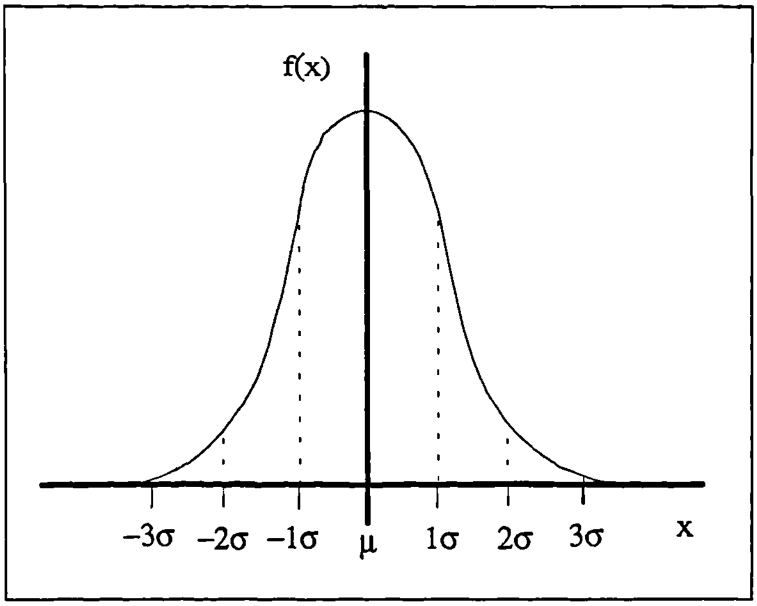

Figure 2.1 Graph of a Cumulative Distribution Function

Table 2.2 The Normal Distribution with Probabilities

In Between	Inside	Outside
±1 σ	68.26 %	31.74 %
±2 σ	95.44 %	4.56 %
±3 σ	99.73 %	0.27 %
±4 σ	99.9937 %	0.0063 %

Table 2.2 summarises the probabilities of obtaining certain values that lie about the *mean*. For example, the probability of obtaining a value lying between ±σ is approximately 68.26%, while the probability of obtaining a value outside these limits is approximately 31.74%. Likewise the probability of obtaining a value between -2σ and +2σ is 0.9544 (95.44%) and between -3σ and +3σ is 0.9973 (99.73%). The probability of a data value falling outside the ±3σ limits is only 0.27%, or 3 in 1000. Since, this occasion rarely occurs, we usually conclude that when data fall outside the ±3σ limits, it indicates that the distribution function has changed, indicating that the process has changed. As a consequence there is some abnormality in the process. Therefore, the area between ±3σ in figure 2.1 is often called a region of *common cause*. On the other hand, the area outside this region is known as *assignable cause* zone.

2.4 Standardised Variable

Since the *mean* (μ) and *standard deviation* (σ) of the normal distribution can take on many different values from situation to situation, it is convenient to define and work with a standardised normal distribution. Such a standard normal distribution for the random variable **z** is defined to a have a *mean* $\mu_z = 0$ and *standard deviation* $\sigma_z = 1$ or *N*(0,1). Partial areas under this standard normal curve have been calculated and tabulated in standard statistical texts. It remains for users to establish the relationship between the particular normal distribution that they are working with the standard normal distribution, so that they may use the tables. In order to translate from the normal distribution *f(x)* to the standard normal distribution *f(z)*, we use the following transformation:

$$z_i = \frac{x_i - \mu}{\sigma} \qquad (2.2)$$

Where z_i is our new value which is in standard deviation unit and x_i is our observation, while μ is the *mean* and σ is the *standard deviation* for variable **x**.

2.5 SPC Control Charts

SPC methodologies are based on the assumption that serially generated data are independent and normally distributed. Based on these assumptions, the probabilities of data values falling in a certain range can therefore be predicted. Using these assumptions, Shewhart (1931) introduced a simple device called the Shewhart ($\overline{X}$) control chart which enables the user to define the state of the process data. It consists of a time plot of data with control limits centred about the target, which is taken from a historical average. The samples are collected in subgroups, randomly from the process, at a regular time interval. A chart is then plotted based on these observations ($\bar{x}$) against time. Then three lines were drawn on the chart. These lines are the grand *mean* ($\bar{\bar{x}}$), the upper control limit line (*UCL*) and the lower control limit line (*LCL*). Basically, the usage of Shewhart control charts are to fulfil the following three distinctly different purposes:

(i) To determine whether a process has achieved a state of statistical process control. For this purpose, appropriate data are collected and tested against the trial control limits.

(ii) To monitor the aim and variability in the process. This is possible by continuously checking the stability of the process. This, in turn will help to assure that the statistical distribution of product characteristic is consistent with quality requirements.

(iii) To maintain current control of a process in which the data is tested against control limits computed from given standards.

In general, a statistically based control chart is a device intended to be used at the point of operation where the process is carried out, and by the operators of the process. In the traditional method, the operators are asked to asses the current situation by taking a sample and plotting the sample results on the control chart, and take action when it is necessary. To elaborate further let us look at the Shewhart control chart for individual measurement and the EWMA control chart, in more detail.

2.5.1 Shewhart Control Chart for Individual Measurements

In certain situations, taking several measurements to form a rational sample size greater than one simply does not make sense because only one measurement is available or meaningful each time that samples are to be taken. For example, certain process characteristics such as temperature and flowrate, will not vary in close proximity in time or space during the period that such sampling normally occurs because the medium is quite homogenous in nature. The apparently different values from the multiple observations of such processes at each sampling are results of reading and measurement error rather than reflections of true process variability.

To initiate the control chart, normally 25 to 50 individual sample measurements are collected when the process is perceived to be in-statistical-control (Oakland and Followell, 1990). Using these samples, we obtain an estimate of the process *mean* (μ) and the process *standard deviation* (σ). Then, lines of *mean* (μ) and control limits are plotted on the chart. The control limits are placed at $\mu \pm 3\sigma$. The process is consider to be in-statistical control when all the data points fall within these control limits.

To improve the sensitivity of the control chart, the previous Shewhart X chart can be modified using several supplemental control rules. The most common is the addition of warning limits. These warning limits are usually set at $\mu \pm 2\sigma$. Two successive points outside these warning limits are usually taken to be a good evidence that assignable causes of variation are present. This chart is known in the literature as the Shewhart X chart with action and warning limits.

The above mentioned control charts are used to present the process data in a framework which will clearly show when an action is necessary, or when further information is required or when no action is necessary. These three kinds of judgement are implemented during process operation using the given sampling technique. In the SPC methodology, inherent variability in the process is inevitable and thus they should be left alone. Figure 2.2 shows an example of a Shewhart X chart for individual measurements with action and warning limits.

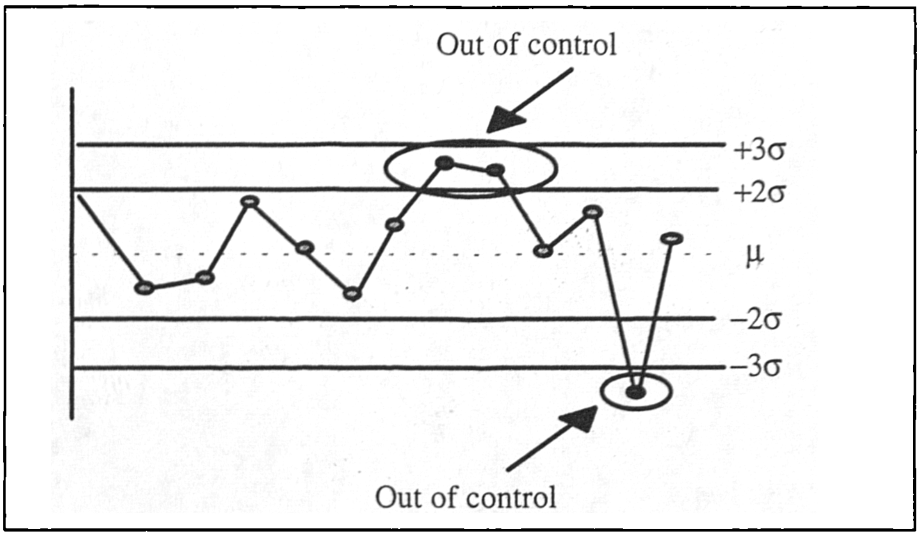

Figure 2.2 Shewhart X Chart with Action and Warning Limits.

Figure 2.2 shows that the control chart has three zones and the required response depends on the zone in which the sample result falls. The possibilities are:

(1) Carry on and do nothing. This is the stable zone.

(2) Be careful and seek information, since the process may require adjustment. This is what we call the warning zone.

(3) Take action, either investigate or where appropriate, adjust the process. This is the action zone.

The procedure is like a set of traffic lights with signals "stop", "caution" and "go". The stable zone is the area between the *mean* (μ) and the $\pm 2\sigma$ limits. The warning zone is indicated by any points falling between $\pm 2\sigma$ and the $\pm 3\sigma$ lines. If two consecutive values fall in this region, then action need to be taken. Lastly the action region is the area outside the $\pm 3\sigma$ lines. Samples in the action region indicates that the process is out-of-control. The search for assignable causes will then be undertaken. Normally, the process will always be in a state of control due to the low probability of out-of-control situation (*i.e.* 3 points in 1000 observations) and unnecessary action on a stable system can be refrained. This will reduce spending time and money on fixing "unbroken" systems (Box, 1993).

2.5.2 EWMA Control Chart

In Shewhart control chart with action limits for individual measurements, the decision signal depends on the last plotted value. There exists another type of control chart called the Exponential Weighted Moving Average control chart. The data plotted is a moving average of previous data in which each data point is assigned a weight. This weight decreases in an exponential decaying fashion from present to remote past. Thus the moving average tends to reflect the most recent performance if most of the weight is allocated to the most recently collected data. The weighting factor θ, given for the process is between 0 an 1.

Basically the EWMA statistic is based on the present predicted value plus θ times the present observed error of prediction. Thus, the EWMA statistic can be mathematically expressed as:

$$\hat{x}_{t+1} = \hat{x}_t + \theta e_t \tag{2.3}$$

$$\hat{x}_{t+1} = \hat{x}_t + \theta(x_t - \hat{x}_t) \tag{2.4}$$

where

$\hat{x}_{t+1}$ = predicted value at time *t+1* (the new EWMA value)
x_t = observed value at time *t*
$\hat{x}_t$ = predicted value at time *t* (the old EWMA value)
e_t = observed error at time *t*

In order to simplify, of the above equation can be written as:

$$\hat{x}_{t+1} = \theta x_t + (1-\theta)\hat{x}_t \quad \text{where } x_0 = \mu \text{ or the target} \tag{2.5}$$

The control limits of the EWMA chart are placed at:

$$\mu \pm K\sigma\sqrt{\frac{\theta}{2-\theta}} \tag{2.6}$$

where μ is the *mean* of the process which is the centre line of the EWMA chart. *K* and θ can be chosen by the user. A single data point outside this limit indicates that the process is out-of-control.

If a shift in *mean* occurs, the EWMA chart will gradually move, depending on the value of θ, to the new *mean* of the process. As mentioned previously, EWMA and Shewhart Control charts differ through the way they handle previous generated data. The traditional Shewhart control chart without the additional control rule may consider the process to be out-of-control when a single point falls outside the $\pm 3\sigma$ limits. This decision is based entirely on the last plotted point. On the other hand, the EWMA chart places weight on the measurements. The weights are given by:

$$w_i = \theta(1-\theta)^{t-i} \qquad (2.7)$$

where w_i is the weight associated with observation x_i, and x_t is the most recent observed data point. When a small value of θ is used, the moving average at any time t carries with it a greater amount of information from the past. Hence, it will be relatively insensitive to recent changes in the process. For control chart applications, where a fast response to process shifts is desired, a relatively large weighting factor of around $\theta = 0.2$ to $\theta = 0.5$ can be used (DeVor *et al.*, 1992). The weightings for the EWMA control chart are displayed in figure 2.3 where time = 10 is the weight for the current observation. The detailed derivations of equations (2.6) to (2.7) are given in Appendix C.

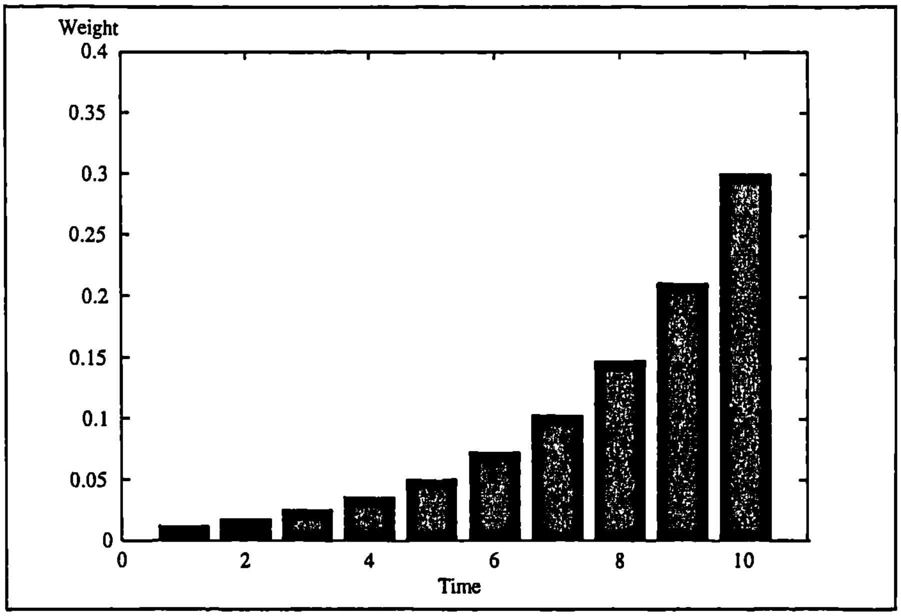

Figure 2.3 Data Weighting for the EWMA Control Chart with $\theta = 0.3$

A plotted point of an EWMA chart can be given a long memory, thus providing a chart similar to the Cumulative Sum (CUSUM) chart (Lucas and Crosier, 1982), or it can be given a short memory and provide a chart analogous to the Shewhart chart. It all depends on the value of θ that we use in the EWMA control chart. When $\theta \rightarrow 1$, $w_i \rightarrow$

1 and $\hat{x}_{t+1}$ practically equals the most recent observation x_t. When the process is under control and $\theta = 1$, points plotted on the classical Shewhart chart and those on an EWMA control chart are therefore identical in their ability to detect signals of departures from assumptions. The EWMA control charts for values of $0 < \theta < 1$ stands between the Shewhart and CUSUM control charts in its use.

One of the advantages of using EWMA charts is that this procedure can also be used in a dynamic process control system because it can provide forecasts of the next value in production data. Through this feature, a future deviation that is too large from target can be avoided through the intervention of feedback control and process operators. Thus, the EWMA chart not only provides the operator with a forecast, but also with control limits to inform them when the forecast is statistically significantly distant from target. Therefore, when an EWMA signal is obtained, appropriate corrective action based on the size of the forecast can be often be devised.

The EWMA chart can be modified to improve its ability to forecast. In situations where the process *mean* steadily trends away from the target value, the EWMA chart can be improved by adding two more terms to the EWMA prediction equation, *i.e.*

$$\hat{x}_{t+1} = \hat{x}_t + \theta_1 e_t + \theta_2 \sum e_t + \theta_3 \nabla e_t \tag{2.8}$$

where the symbol ∇e_t indicates the first difference of the error e_t; that is $\nabla e_t = e_t - e_{t-1}$. The forecast value $\hat{x}_{t+1}$ equals the present predicted value plus three quantities: one proportional to e_t, the second is a function of sum in e_t, and the last one is the function of the first difference in e_t. These terms are sometimes called the "proportional", "integral and "differential" equation in the Automatic Process Control (APC) scheme of proportional, integral and differential (*PID*) control equation. The parameters θ_1, θ_2 and θ_3 can be determined from the historical data using least-squares procedure, in order to give the best forecast (Hunter, 1986; Box and Jenkins, 1976). Appendix C shows how the parameters θ_1, θ_2 and θ_3 are obtained.

2.6 Hypothesis Testing

Fundamentally, the SPC control chart is similar to hypothesis testing in statistical inference. Control limits are analogous to decision rules in statistical hypothesis tests, such that when control limits are exceeded, there is significant evidence to suggest that the process is not in statistical control. It indicates that the variables being monitored by the SPC control chart do not have the same *mean* and *variance* over time.

SPC control charts continuously make statistical hypothesis tests through out the progress of the process. The null hypothesis H_0 for a normally distributed process is that the true process *mean* (μ) is equal to the target process *mean* (μ_0). The alternative hypothesis, H_1, is that μ does not equal μ_0. The interpretation of H_0 is that everything is in good condition and the process should be left alone, whereas H_1 indicates that there is a problem and actions should be taken on the process.

Any statistical hypothesis test associated with SPC control charts will carry two types of risks; producer's risk (α) and the consumer's risk (β). The producer's risk (α) is the chance of making a *type I* error, that is, the null hypothesis is rejected when the process output really follows the distribution of $N(\mu_0,\sigma^2)$. This is the risk of taking action due to a signal given by an extreme observation when in fact the process does not change at all. Hence, this phenomenon is called producer's or manufacturer's risk, or a false alarm. In traditional SPC applications, the probability of *type I* error is usually fixed at a small value. In case of Shewhart (X) with action control limits, α is often fixed at 0.0027. Tests with a small probability of false detection is preferred because of the cost associated with false alarm. If a false alarm occurs, personnel have to search for the source of assignable cause variation and when they find none, effort and time will have been wasted.

The second type of risk, consumer's risk, is the chance of making a *type II* error. It implies accepting H_0 when actually the process output does not follow the distribution $N(\mu_0,\sigma^2)$. It is the chance of not detecting a fault when one occurs or when a disturbance is present and it is not detected. The power of the hypothesis is the probability of signalling a disturbance. If a disturbance is present, the power is one minus the probability of a *type II* error. The *type II* error, which having β risk, imply that the consumer has the chance of accepting bad or unacceptable products because the producer did not detect any malfunction or out-of-control situations in the process. Hence it is termed consumer's risk. Table 2.3 summarises the errors associated with various conclusions of hypothesis tests.

Table 2.3 Decision Error in Hypothesis Testing

Decision	True H_0	False H_0
H_0 not rejected	No error with probability $1-\alpha$	*Type II* error with probability β *Consumer's Risk*
H_0 rejected	*Type I* error with probability α *Producer's Risk*	No error with probability $1-\beta$

2.7 Summary

This chapter discussed the background of Statistical Process Control. We elaborated the sources of variational noise in the process which can be classified into two categories, either inherent to the process or abnormal to the process. We tabulated the characteristics of these basic variations in order to determine what kind of action to take and whom to assign the responsibility for taking control action. Then we discussed the importance of normally distributed data in determining the variability in the process. Two types of control charts were presented, namely the Shewhart control chart for individual measurement and the EWMA control chart. Finally, we discussed hypothesis testing which involves two kinds of risk; either manufacturers' risk or consumers' risk. In the next chapter we will review the development of integrated SPC and APC strategies.

Chapter 3

Literature Review

3.1 Introduction

The literature review is organised in three main parts. First we discuss the basic concepts of Automatic Process Control (APC) and Statistical Process Control (SPC). The second part compares the SPC and APC approaches. Lastly we examine the development of integrated SPC and APC strategies.

3.2 Concepts of APC and SPC

Automatic Process Control (APC) loops are extensively utilised in the chemical industry. In general, chemical industries apply APC methods because the cost of making adjustments are assumed negligible. The philosophy of APC is to minimise the variability by adjusting the manipulating variables in the process. Only conventional feedback control will be elaborated here. For a feedback control system, information is taken from the process variable (Y). This process variable is obtained by a suitable measurement device, which becomes an input to a controller. An error detecting device in the controller compares the input measurement to a signal representing the desired condition, or set point (Y_{SP}), so that any difference between the set point and the input measurement causes the controller to generate a corrective output signal. The signal is then applied to the final control element which in turn adjusts an input variable (X) to the process in the direction which tends to return the controlled variable to the desired condition. This feedback loop operates continuously to maintain the controlled variable at the desired condition. All this action is maintained and supervised by the controller automatically without the intervention of a human operator once the feedback loop is properly tuned and set.

On the other hand, the objective of SPC procedures is to find and to eliminate assignable causes that bring about variability in the process. These assignable causes are undesirable because they prevent the process from maintaining a state of statistical control. The methodology employed to achieve this objective is based on the application of graphical tests of hypotheses. The observations, Y_t (where t is the current observation), are sequentially plotted and compared against control limits. If the plotted point falls outside the limits it indicates that the process is being affected by an assignable cause. The identified assignable cause is to be eliminated from the process thus forcing the process to eventually reach a state of statistical control.

The origin of this graphical technique is usually associated with the publication of Shewhart's book in 1931. It was here that the concept of control charts was first introduced. It was established on the assumption that the measurable characteristics of manufactured products is always subject to some uncontrollable variations. The determinations of these variations are based on appropriate statistics and the help of graphical displays.

The introduction of the control chart aids in determining the capabilities of the production process. Action is taken when these estimated capabilities are unsatisfactory in relation to the design specifications. Furthermore, once the process capabilities have been determined, and are satisfactory, action is taken only when the control chart indicates that the process has fallen out of statistical control (*i.e.* assignable causes of variation have entered). Thus, the control chart has its functions of firstly, to determine whether the process is capable; secondly, to detect and identify the assignable causes of variation, and lastly, to suggest the necessary correction to the process.

Several authors have published reviews on the use of Statistical Process Control charting developments and techniques. The earliest review on the development of control charts was given by Gibra (1975). Later, Vance (1983) published a bibliography on the development of statistical control charting technique until the year of 1980. In other developments, Vardeman and Cornell (1987) compiled a partial inventory of statistical literature on quality and productivity through 1985. They focused on classical statistical quality control and industrial experimental design areas which include a listing of journals, review articles, case studies, books, booklets and some audio-visual materials. Later, Faltin *et al.* (1991) discussed the application of on-line quality control for the process industries for the 1990's and beyond. Saniga and Shirland (1977) made a survey on the usage of quality control methods in industry.

They found that 71% of those sampled utilised the Shewhart $\overline{X}$ Control Chart and 64% used the R (range) chart compared to other types of control chart. Out of all respondents, the largest percentage of industries using SPC methods came from the manufacturing industries, followed by the electrical based industries, automotive industries, chemical industries and lastly the others which comprised the service, aircraft/aerospace, wood fabrications and other industries. This is not surprising as SPC was invented to improve operations in the manufacturing industries.

3.3 Comparison of SPC and APC

In order to compare SPC and APC, first it will be necessary to explore the development of SPC in the chemical industries. As mentioned by Saniga and Shirland (1977), the chemical industries came fourth amongst those using SPC methods. This sector deals with the continuous flow of material like the production of chemicals, petroleum refining and manufacture of synthetic rubbers. Products from such processes are observed using continuous measures such as pound, gallon and litre. Chemical processes are sometimes quite complicated, involving a lot of recycling and many stages of mixing and blending. The application of SPC in this area is quite new, due to less emphasis on quality awareness on the part of the chemical industries. With the increase in demand for speciality chemicals, quality consciousness began to grow. This led to increased implementation of SPC methodologies in the chemical industries (Box and Kramer, 1992).

Although SPC and APC methods came from different backgrounds, the objective of both procedures is the same, *i.e.* to reduce variability in the process. Box and Kramer (1992) mentioned that both methods achieve this objective by the ways they govern their critical variables. SPC looks for signals representing assignable causes, which may indicate an external disturbance that increases process variability. On the other hand, APC actively reverses the effects of process disturbances by making regular adjustments to manipulated variables. Since there are usually several possible variables to manipulate, one has several choices on where to transfer the variation. In view of this APC, is a short term strategy leading to a more capable process. SPC however, attempts to improve the process over the long term by finding and suggesting a removal of root causes behind the variation. For a process with heavy adjusting costs, SPC can be the best choice because it does not penalise any inherent variation in the process which is monitored by the SPC chart. If there is no cost associated with the adjustment of control actions, then APC provides a more powerful approach since it penalises any variation in the process through the process set point.

Vander Wiel *et al.* (1992) provided further the differences between SPC and APC methodologies. In the application context, SPC expects that successive measurements in the process are statistically independent and to have a distribution that does not change over time. This is appropriate for detecting departures from the ideal. However, APC process measurements are expected to be correlated over time, causing the process to wander if adjustments are not made. APC is often used tactically where the system is typically commissioned to maintain the setpoint of important parameters at their desired values. On the other hand, SPC is often considered as a strategic procedure. Only important quality characteristics are charted, allowing SPC to have an immediate impact on the output quality. As far as the target is concern, SPC goes for quality measurement variables whilst APC selects the process parameters that do not necessarily have to be quality measurement variables. As for function, SPC just monitors the process and gives signals in the form of identification and suggestion of root cause removal. However, it does not adjust the process. On the other hand, APC controls the process by adjusting the manipulated variables but does not remove the root or assignable causes. Both strategies are implemented using different ways. SPC is often a downward concept, driven by upper management or customer as part of a company-wide quality improvement exercise while APC is often a upward concept driven by the process control or manufacturing engineer. If all these mentioned philosophies are successful, SPC will lead to process improvement while APC will lead to process optimisation. Table 3.1 summarises the APC and SPC philosophies as discussed.

Comparison of the two methodologies based on Table 3.1 may lead us to believe that the statisticians who deal with SPC and process control engineers who deal with APC have nothing in common in the nature of their work. However, MacGregor (1988) mentioned that although the two groups differ, in reality one very important area of overlap exists between the SPC group and the APC group. In a large number of companies, these two groups are trying to solve the same quality-problem, but by using different techniques. Neither group fully understands the techniques of the other. Normally, the SPC group uses discrete data obtained from infrequent samples, analysed off-line in a laboratory. Although the statisticians are comfortable with the discrete data, they normally have neither a background in process dynamics nor any familiarity with process control. Process control engineers, on the other hand, have a good understanding of process fundamentals, process dynamics, and process control theory using tools such as Laplace Transforms and stability analysis. However, they do not posses a solid background in statistics, and dealing with normal distributed data and hypothesis testing.

Table 3.1 Comparison of SPC and APC strategy.

	Statistical Process Control (SPC)	Automatic Process Control (APC)
Philosophy	Minimise variability by detecting and removing process upset	Minimise variability by adjusting the process to counteract process upset
Application	Expect stationary process	Expect continuous process drift
Deployment		
Level	Strategic	Tactical
Target	Quality characteristics	Process Parameters
Function	Detecting disturbances	Monitoring Set points
Focus	People, method and equipment	Equipment
Function	Monitor the process	Control Process
Autocorrelation	None	Low to high
Implementation	Downward	Upward
Results	Process Improvement	Process Optimisation

APC and SPC strategies are based on highly important concepts and both methods have long and distinguished records of practical achievement. There are some characteristics from both methodologies which are appropriate for integration. If this can be achieved, then it may be possible to devise an effective method that is capable and quick to detect quality changes in the process.

3.4 Integrating SPC and APC strategies

Several researchers including MacGregor (1988); Vander Weil *et al.* (1992); Tucker *et al.* (1993)· and Faltin *et al.* (1993) suggested that it is essential to integrate the APC and SPC methods. The integrated system utilise APC to reduce the effects of predictable quality variations and SPC to monitor the process for detection of assignable causes. The removal of these assignable causes will result in additional reduction of overall variability.

Until recently, there has been little effort to integrate SPC monitoring and APC strategies. MacGregor (1988) mentioned that the low attention paid by some industrial process control groups in including product quality as part of their control strategy is due to several reasons. Firstly, it is due to undergraduate Chemical Engineering programmes laying heavy emphasis on petrochemical operations where quality has been less of concern than in speciality chemicals, electronics and biomaterials. Secondly, due to the managerial structure of many companies, process control groups have become isolated from the end user. They are rarely able to relate the quality problems that customers experience back to the operation and control of the process. Lastly, it is assumed that process control engineers, with their inadequate background in statistics, are incompetent to handle the noisy, infrequent product quality data that are typically generated off-line in quality control laboratories.

One of the earliest articles that linked APC and SPC methodologies was presented by Box and Jenkins (1962). They presented a concept of adaptive control charts which provide a systematic application of feedback procedures from the data measurement to achieve appropriate adjustment on the process. However, the surge of activity related to SPC-APC in the early 1960s was not sustained. Although the general idea is not new, the idea superimposing statistical process monitoring on a closed-loop system appears to be quite recent and certainly opens a new line of research in the area of quality improvement (Vander Weil *et al.*, 1992).

MacGregor (1988) was apparently the first to revive the idea of integrating SPC and APC in this decade by suggesting to the SPC community that SPC control charts can be used to monitor the performance of a controlled system. Several other researchers emerge later to develop hybrid applications of APC and SPC that they called Algorithmic Statistical Process Control (ASPC) (Vander Weil *et al.*, 1992; Tucker *et al.*, 1993; Faltin *et al.*, 1993). Lately, new interest in applying this method in the process industries has emerged (MacGregor, 1988; Efthimiadu and Tham, 1991; Vander Weil *et al.*, 1992). Thompson and Twig (1994) in an other development use SPC charts to monitor the process output and selectively apply a conventional PID control algorithm. Some other workers have used SPC charts for filtering the product measurements before applying feedback control actions (English and Case 1990; Rhinehart, 1992; Rhinehart, 1995). In other developments, a SPC control chart is used to monitor, detect, and adjust the process by using the relationship between the quality variables and the input variables that they called Active SPC (Efthimiadu and Tham 1991; Efthimiadu *et al.*, 1991; Efthimiadu *et al.*, 1992; Efthimiadu *et al.*, 1993; Ibrahim and Tham 1995).

3.4.1 Algorithmic Statistical Process Control.

Recent work by Vander Wiel *et al.* (1992), Tucker *et al.* (1993) and Faltin *et al.* (1993) led to the development of Algorithmic Statistical Process Control (ASPC). Vander Weil *et al.* (1992) illustrated the methodology by applying it to a batch polymerisation process. Tucker *et al.* (1993) provided technical descriptions of the ASPC concept. Faltin *et al.* (1993) described the steps involved in applying ASPC, and consequently gave the technical and non technical requirements for successfully implementing ASPC. They defined ASPC as an approach that realises quality gains through appropriate process adjustment using APC and through elimination of root causes of variability by using SPC. Thus, ASPC is an integrated approach to quality improvement. Two features are essential for a process to be a good candidate for ASPC. Firstly, it must be possible to use past data to construct a good predictor for future process performance. Secondly, there must be some manipulated variables available for compensatory adjustment. These manipulated variables must affect the performances of quality variables of interest if changes are made.

ASPC succeeds in reducing the variation of quality characteristics through feedback techniques. Then it monitors the entire system to detect and to remove the unpredictable upset variations. If the process possesses some degree of predictability, the APC part will minimise the deviations from target in the short term by adjustment of the process. On the other hand, the SPC part will detect changes from past performance, identify the root problems, and make attempts to remove them. Hence, ASPC aims to reduce both short-term and long-term variability by changing the role of control charts; "monitor, then adjust when out of control" with "adjust optimally and monitor" (Faltin *et al.*, 1993).

Vander Wiel *et al.* (1992) also mentioned that although APC and SPC were developed in isolation from one another, they successfully integrated both fields together so that it is capable of producing quality improvements through two characteristics mentioned earlier. They applied the ASPC technique on a polymerisation reactor process, using viscosity as their quality characteristic and the amount of catalyst as the adjusting compensatory variable. They used the criteria of minimum square error deviation (MSE) control rules to regulate the process. Since the plant personnel were already familiar with Cumulative Sum (CUSUM) monitoring scheme, it was natural to introduce ASPC coupled with this chart. Two main sources of variations were identified in the polymerisation process; one is the effect of raw-materials, and the other is seasonal factors affecting heat-exchanger effectiveness. The net effect of such shifts results in either a sudden change in product viscosity or the nominal catalyst level needed to produce material of the target viscosity. The purpose of the CUSUM chart here was to detect shifts in the system as quickly as possible when they occurred to resolve which mechanisms was in fact responsible. By applying the ASPC methodology, they claimed that the viscosity variability in this reaction process stage has been reduced by 35% and virtually eliminated off-spec material from this source.

Figure 3.1 shows a flow diagram of the ASPC methodology. The top level refers to the conventional APC feedback control loop while the remaining levels form the SPC monitoring scheme. The SPC monitoring scheme will provide signals for locating the assignable cause and provide updates to the APC control algorithm if needed. There are mainly four procedural steps needed for successful implementation of ASPC. First is to develop a time series model for the process output using past record data. Second is to design the control law for the estimated model based on pertinent cost. Third is to place a SPC chart to monitor the progress of the control loop. This SPC chart will signal when the process and the controller are no longer acting as anticipated from the identification and estimation stage. Last is to search for the assignable cause and remove it when SPC chart gives a signal (Vander Weil *et al.*, 1992).

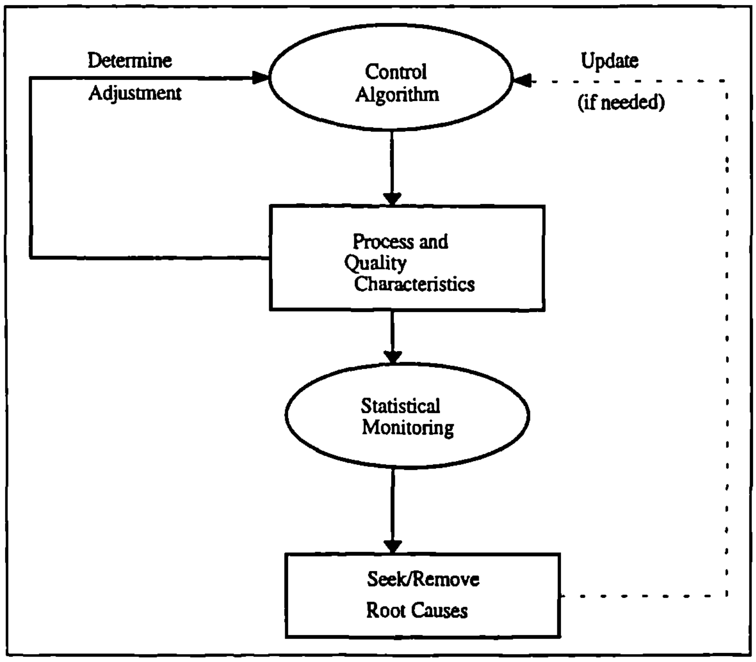

Figure 3.1 ASPC control algorithm based on Faltin *et al.* (1993)

In a recent application, Capilla *et al.* (1995) presented an implementation of integrated SPC and APC to a polymerisation process. Their technique is similar to the ASPC technique developed by Vander *et al.* (1992), Faltin *et al.* (1993) and Tucker *et al.*

(1993). They used the Melt Index (MI), a measure of the polymer viscosity, as their quality variable, and the reactor temperature as the manipulated variable. Both of these variables satisfied the requirements of ASPC methodology. That is the MI data is autocorrelated and could be used to construct a good predictor for future process performance, while the reactor temperature could be utilised as a compensatory variable, that adjustment would have a predictable effect on the performance quality variable. Several types of APC strategies were tested on their work including (i) Minimum Square Error (MSE) controller (ii) Two-Steps Ahead Forecasting Controller (TSAFC) (iii) Minimum Weighted Variance Controller (MWVC). They compared the performance of several strategies of APC by using a stochastic control factor *k*:

$$k = \sqrt{\frac{Mean\ Square\ Error_{APC}}{Mean\ Square\ Error_{No\ APC}}} \tag{3.1}$$

where the denominator of the square root is the deviation of MI from the target (*T*) when APC was not applied and the temperature was fixed at its nominal value.

The APC feedback control algorithm was then integrated with the following monitoring schemes: (i) Shewhart Chart for individual measurements; (ii) CUSUM chart for residuals (*i.e.* the difference between the current value and the predicted performance); *(iii)* Run chart (a Shewhart-type chart without control limits) applied to monitor the deviation of MI from target and to monitor the temperature adjustments. They tested the system performance by injecting several types of assignable causes: a sudden shift of 5σ in white noise, MI measurement error in the laboratory (MI_{lab} = 0.3 + MI_{actual}), and sustained temperature sensor failure (T_{set}=T_{actual} - 1°C). All these assignable causes were introduced into the process at time t = 25. They concluded that the combination of Two-Steps Ahead Forecasting Controller (TSAFC) and SPC shows the best potential to control the polymer viscosity.

In yet another development, Montgomery (1991) proposed a scheme for the integration of APC and SPC similar to the ASPC strategy shown in Figure 3.1. The scheme used feedback control to make adjustments to the manipulated variable x_t so that the controlled variable y_{t+1} is on target. This scheme assumed that we can: (i) predict the next observations on the process; (ii) have some input variables that can be manipulated in order to affect the process output, and (iii) know the effect of this manipulated variable so that we can determine how much control action to apply (*e.g.* we can make adjustments in the manipulated variable at t that is more likely to produce the target value in the process output at period $t+1$). In this case, the SPC chart was used for process monitoring but not for control. The function of this control chart was to search for assignable causes in the process. The control part is executed by the APC scheme.

Later on, Montgomery *et al.* (1994) tested the previous proposed scheme by applying it to MacGregor's (1990) model of a funnel experiment and showed that the combined APC and SPC control strategies reduced the overall variability even though the system experiences certain external assignable causes. Two types of assignable causes were introduced in their work, sudden shifts in the *mean* having the magnitudes of 1, 2, 5, 7.5 and 10 units and trends in the *mean* of 0.05, 0.1, 0.25, 0.5 and 1.0 units/period. Four different types of SPC charts were utilised in the simulation studies, namely Shewhart chart for individuals with 3σ limits, EWMA with $\theta = 0.1$ and 3σ limits, EWMA with $\theta = 0.4$ and 3σ limits, and CUSUM with V-mask using $k = 0.5$ and $h = 5$. Where k is the slope of the V mask arms and the decision intervals is equal to $h\sigma$. Their studies were based on a Performance Measure (PM) which is defined by the average squared deviation from the target (T):

$$\mathrm{PM} = \frac{1}{n}\sum_{t=1}^{n}(Y_t - T)^2 \qquad (3.2)$$

From the result of the simulation studies, they found that the integrated APC and SPC strategy gave better performance compared to the APC strategy alone. They concluded that in chemical process plants and in computer integrated manufacturing environments combining APC and SPC strategy can become an important tool for improving process quality.

3.4.2 SPC as Filtering Device in Control Loop

English and Case (1990) employed the EWMA control chart as a filtering device within the feedback control loop. The function of the filtering device is to determine whether an out of control situation has occurred in the system. If an out of control is not due to inherent noise then the current measurement of output is fed back to the comparator and a PID controller is invoked. However, if the current data indicates that the process is in a state of statistical control, the current setpoint value (Ysp) is returned to the controller to represent the process output. Hence the controller will not take any control action because the error is zero. Figure 3.2 shows the configuration of SPC chart as the filtering device based on their work.

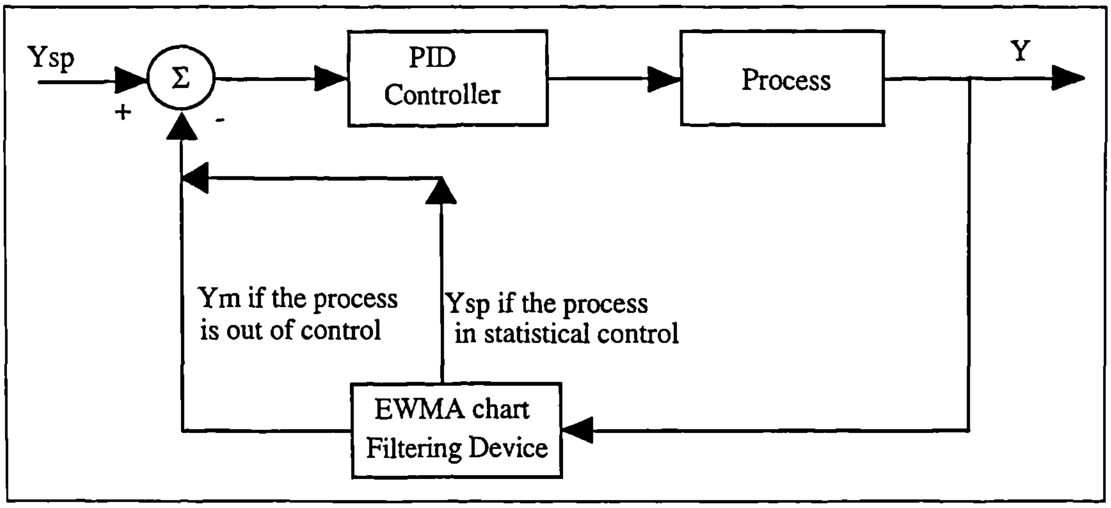

Figure 3.2 SPC chart used as the filtering device in the control loop

Drawing upon similar concepts, Rhinehart (1992) used the CUSUM chart to filter the output of proportional plus integral (*PI*) controller. The advantage of this filtering approach is that it eliminates unnecessary control action on the manipulated variable and yet remains responsive to real process changes. On the other hand, both of these filtering methods will only succeed provided there are no significant deviations from set-point. SPC tools used in this way will create a significant lag in the feedback loop with a trade off between the quietness of the controller actuation and system error.

3.4.3 SPC as Supervisory Unit with PID Feedback Control

Thomson and Twig (1994) developed a method where they employed SPC control charts as a supervisory unit. The chart monitored the process output and selectively applied conventional proportional integral derivative (*PID*) controller. The advantage gain from this method lies in the ability of the control loop to distinguish between inherent process noise and the real error signal. It stopped the controller from acting unnecessarily. As a result it would prevent the actuator from overwork.

Two types of SPC charts were used as the supervisory unit, the *mean* chart ($\bar{X}$) with a group size of four and the exponentially weighted moving average (EWMA) chart with the weight, $\theta = 0.2$. They tested for assignable causes in their system with four kinds of rules. Firstly by using the Action rule, where the process is out of control when a point falls outside the action line. Secondly by using the Warning rule, where two consecutive points falling outside the warning lines indicates that an assignable cause has inhibited the process. Thirdly by using the Run rule, when seven consecutive points fall above or below the target it shows that the *mean* of the process has changed.

Lastly by using the trend rule, whenever seven consecutive points either rise or fall it reveals that the process is being affected by an assignable cause.

If action or warning rules are broken, it indicates that there is a significant error in the process and thus, the supervisory unit switches to *PID* mode immediately. If the run rule or trend rule is broken, it indicates that a gradual drifting of the process average from the set point is occurring. The supervisory unit then switches to integral mode that causes only integral action to be applied. This will eliminate small offsets from the set point. If no rule is broken then the supervisory unit causes the current controller output to be maintained and no changes of control are implemented.

Apart from the four types of tests mentioned previously, the process is also tested based on the distribution of individual data. This test is incorporated to detect sudden error signal immediately without any delay to grouping. The function of this test is to prevent the *PID* and *PI* control actions from returning back to normal before the process has properly settled down from oscillatory behaviour. The limits of the test are set at $Ysp \pm 3\sigma$ on the individual data chart, where *Ysp* is the set point. The process output *Y* is applied directly to this chart without any grouping. If the process output falls outside this limit, *PID* action is summoned. As mentioned before, normal PID control action is invoked when action and warning rules are broken. An integral mode action is only applied when the trend or run rule is violated. Lastly constant control action acts when all rules are obeyed. Figure 3.3 shows the overall picture of this concept.

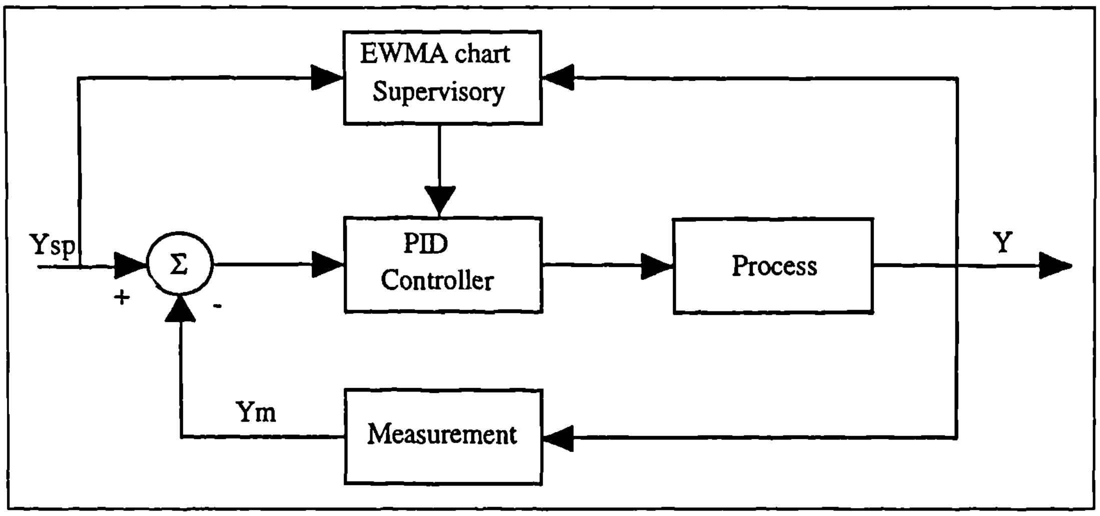

Figure 3.3 The supervisory control using EWMA chart (Thomson and Twig, 1992).

3.4.4 Active SPC Approach

In another important development, Efthimiadu and Tham (1991) managed to change the function of SPC monitoring chart to adjust the process automatically. They avoided the model identification problem by focusing on statistical input-output relationships. The main cause of variation in product quality was identified via multivariate analysis using Principal Component Analysis (PCA). The results of the PCA were then used to identify the manipulated variables and control laws for keeping the quality variable in a state of statistical control. They demonstrated the effectiveness of their procedure by applying it to a simulated CSTR process. It was a reversible exothermic reaction where the heat generated from the process was cooled by the cooling jacket around the CSTR.

Later, Efthimiadu *et al.* (1991) and Efthimiadu *et al.* (1992) extended their work to include the use of another multivariate technique called Partial Correlation Analysis (PCorrA) to analyse the process and to determine the control limits and control laws. They investigated the performance of Active SPC by using Shewhart Charts with Action line, Shewhart Charts with both Action and Warning lines and EWMA control charts. They also studied the effects of off-line and on-line updating of the control limits and control rules for manipulating the input variables. They also considered the case that an input variable that may not be manipulated on-line, by replacing it with another variable. This new variable was selected on the basis that it could compensate for the variation in the latter variable and could affect the quality variable. From their work, they concluded that the application of the proposed procedure for on-line SPC would involve the use of both PCA and PCorrA. The former should be used as a first stage analysis, then, either the PCA or PCorrA technique could be utilised for determining the control laws and control limits. They found out that the method which used PCorrA with EMWA gave the best result. But this better control performance was normally accompanied by a higher level of false alarms.

Later, Efthimiadu *et al.* (1993) called the method that they proposed as Active SPC. Ibrahim and Tham (1995) continued their work by focusing specifically on the utilisation of PCorrA. They investigated the performance of Active SPC by using Shewhart Chart with Action lines and Shewhart Chart with both Action and Warning lines. They also studied the effects of on-line and off-line updating of the corresponding control rules. Additionally, they investigated the outcome of manipulating all or some input variables. Since better control performance is normally accompanied by a higher level of false alarms, they developed a criteria called Index of

Performance (*IP*). The function of this *IP* was to determine which of their methods gave the best control performance. They found out that updating the control rule on-line coupled with manipulating some input variables, provided the best performance. The manipulated variables were the CSTR reaction temperature and the flowrate of the cooling medium.

3.5 Summary

This chapter has made an attempt to review some of the literature published in the domain of Statistical Process Control in chemical industries, focusing on integrated SPC and APC strategies. We discussed briefly, the background and origins of the SPC and APC methodologies and how the methods differ. Then the history of hybrid APC and SPC strategies and the advantage that can be gained by this integration was discussed. Several hybridisation schemes are reviewed namely, Algorithmic Statistical Process Control (ASPC), combination of PID control and SPC charts, SPC charts as filters in a feedback control loop and lastly the Active SPC scheme. This Active SPC scheme will be explored in greater detail in the rest of this thesis. In the next chapter, however, we are going to explore the mathematical modelling and the application of an APC algorithm to a CSTR process.

Chapter 4

CSTR Modelling and Control

4.1 Introduction

This chapter introduces the development of dynamic mathematical modelling, simulation and control on a continuous stirred tank reactor (CSTR). The modelling approach is made possible by applying dynamic mass and energy balances on the CSTR. The simulated CSTR process will then be regarded as the process and will be used through-out the study in this thesis. The *Process Reaction Curve* technique was used to calculate the parameters of the transfer function of the system and used to tune the feedback controller. This was a *Proportional plus Integral* controller and its performance will later be compared with those of Active SPC schemes in Chapter 7.

A good mathematical model is important because it can enhance the understanding of how the process works. Mathematical modelling and simulation is indispensable, particularly in operator training when the new process is still under construction or when new controller modes are being tested on the process. It is usually much cheaper, safer, and faster to conduct this kind of training using a simulator compared to hands on experiment on the operating unit. It is not that the real plant training is not important, but by using simulator a variety of process operating conditions can be tested without fear of losing production. By interfacing process simulator with standard process control equipments, a realistic environment can be created for operator training without the cost and exposure to dangerous conditions that might exist in real plant conditions. The discussion about the importance of simulation would not be complete without descriptions of the processes. In the chemical industries, these may involve reactors, distillation columns, absorption columns and many others. One of the most important processes is the continuous stirred tank reactor (CSTR).

4.2 Continuous Stirred Tank Reactor (CSTR)

This work involves the simulated applications of proposed control strategies to a continuous stirred tank reactor (CSTR). Thus the mathematical modelling and simulation is performed on this process. Two types of generic reactants A and B are fed in excess to the CSTR to produce C via non-linear second-order reversible exothermic reaction kinetics. The energy generated by the reaction process is absorbed by a cooling jacket. The input variables are the concentrations of reactants, *Ain* and *Bin*; the temperature of reactants, *Tin*; the input temperature of the cooling medium, *Tjin*; the flowrate of reactants, *F*, and the flowrate of cooling medium, *Fj*. The output variables are the temperature of the product, *T*; coolant output temperature, *Tjout*; output concentrations of the reactants, *Aout*, *Bout*, and the product concentration, *Cout*. *Cout* is our quality variable of interest, that is, it has to be controlled and kept under statistical control. Figure 4.1 shows the diagram of the CSTR used in this work.

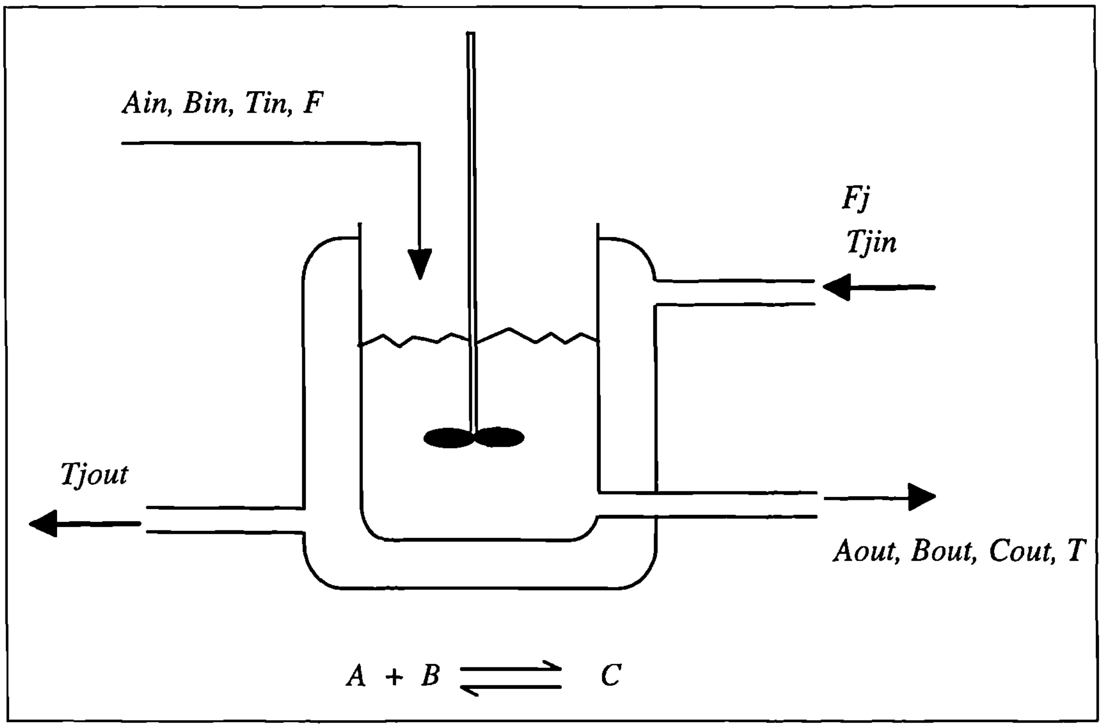

Figure 4.1 Continuous Stirred Tank Reactor (CSTR)

4.3 Modelling the CSTR Equations

Prior to CSTR modelling, certain assumptions have to be made to simplify the process. Obviously, rigorous models may involve microscopic detail and will be time consuming to formulate. It would also be complex and will take an excessive amount of computer utilisation to solve. An engineering approach should compromise between this rigorous description and getting an answer. In practice, normally the optimum approach must correspond to a model that is as complex as the available computing facilities will permit. It indeed requires a lot of skill, ingenuity and practise in order to develop a good model. The assumptions should be carefully considered and listed. The impose limitation of the model should always be kept in mind when calculating the predicted results. Following the above discussion, the derivation of the CSTR mathematical model assumes the following:

- Perfect mixing in the reactor
- Constant heat capacity of the mixture in the reactor.
- Constant volume in the reactor.

Based on the above assumptions, the fundamental equations describing the reversible reaction process can be modelled by using the mass balance and the heat balance. For the rest of the discussion, the inputs to the CSTR are chosen as the independent variables while the outputs are termed as dependent variables. As there are five dependent variables in the CSTR process, obviously there should be five equations to describe the process. These equations are categorised below

4.3.1 Energy Balances

Energy balance in the reactor:

$$\frac{dT}{dt} = Fc_{11}(Tin - T) + c_{21}\, r_1[C_{Aout}][C_{Bout}] - c_{22} r_2 [C_{Cout}]^2 - Uc_{31}(T - Tjout) \qquad (4.1)$$

where:	F	is the flowrate of cooling medium into the jacket	(litre/sec)
	Tin	is the reactant temperature into the jacket	(K)
	$Tjout$	is the cooling medium temperature in the jacket	(K)
	T	is the reactor temperature	(K)
	U	is the overall heat transfer coefficient	(cal/cm^2/K/sec)
	C_{Aout}	concentration of component A in the CSTR	(mole/litre)
	C_{Bout}	concentration of component B in the CSTR	(mole/litre)
	C_{Cout}	concentration of component C in the CSTR	(mole/litre)
	r_i	is the rate of reaction i	(litre/mole/sec)

By applying an energy balance on the cooling jacket of the reactor, the following equation is obtained:

$$\frac{dTjout}{dt} = F_j c_{41}(Tjin - Tjout) + Uc_{51}(T - Tjout) \qquad (4.2)$$

where:	F_j	is the flowrate of cooling medium into the jacket	(litre/sec)
	$Tjin$	is the cooling medium temperature into the jacket	(K)
	$Tjout$	is the cooling medium temperature in the jacket	(K)
	T	is the reactor temperature	(K)
	U	is the overall heat transfer coefficient	(cal/cm^2/K/sec)

4.3.2 Species Material Balances.

The following relationships are obtained when molar balances are performed on the individual components in the reaction system. The mass balance of component A in the reactor is given by:

$$\frac{dC_{Aout}}{dt} = Fc_{11}(C_{Ain} - C_{Aout}) - r_1[C_{Aout}][C_{Bout}] + r_2[C_{Cout}]^2 \qquad (4.3)$$

Mass balance of component B in the reactor:

$$\frac{dC_{Bout}}{dt} = Fc_{11}(C_{Bin} - C_{Bout}) - r_1[C_{Aout}][C_{Bout}] + r_2[C_{Cout}]^2 \qquad (4.4)$$

Mass balance of component C or the product from the reactor:

$$\frac{dC_{Cout}}{dt} = -Fc_{11}C_{Cout} + r_1[C_{Aout}][C_{Bout}] - r_2[C_{Cout}]^2 \qquad (4.5)$$

where:	F	is the flowrate of reactant into the CSTR	(litre/sec)
	C_{Aout}	concentration of A in the CSTR	(mole/litre)
	C_{Bout}	concentration of B in the CSTR	(mole/litre)
	C_{Cout}	concentration of C in the CSTR	(mole/litre)
	C_{Ain}	concentration of A into the CSTR	(mole/litre)
	C_{Bin}	concentration of B into the CSTR	(mole/litre)
	r_i	is the rate of reaction i	(litre/mole/sec)

4.3.3 Constant Parameters.

In order to obtain numerical values later on in the work, the following constant parameters are specified for the CSTR process:

$$c_{11} = \frac{1}{V_{cstr}} = 0.2\,\text{litre}^{-1} \qquad c_{21} = \frac{(-\Delta Hr_1)}{\rho_1 C_p} = 0.1\frac{\text{litre}\cdot\text{K}}{\text{mole}}$$

$$c_{22} = \frac{(-\Delta Hr_2)}{\rho_2 C_p} = 0.01\frac{\text{litre}\cdot\text{K}}{\text{mole}} \qquad c_{31} = \frac{A}{\rho_1 C_p V_{cstr}} = 0.0034\frac{\text{cm}^2\cdot\text{K}}{\text{cal}}$$

$$c_{41} = \frac{1}{V_{jacket}} = 0.36\ \text{litre}^{-1} \qquad c_{51} = \frac{A}{\rho_2 C_p V_{jacket}} = 0.0024\ \frac{\text{cm}^2\cdot\text{K}}{\text{cal}}$$

$$U = 19\frac{\text{cal}}{\text{cm}^2\cdot\text{K}\cdot\text{sec}}$$

where: V_{cstr}	is the reactor volume	(litre)
V_{jacket}	is the jacket volume	(litre)
C_p	is the heat capacity of reaction mixture	(cal/mole/K)
U	is the overall heat transfer coefficient	(cal/cm^2/K/sec)
ρ_i	is the molar density i	(mole/litre)
ΔHr_i	is the heat of reaction for reaction i	(cal/mole)
A	is the surface area for heat transfer	(cm^2)

4.3.4 Arrhenius Equations

The temperature dependence for the reaction rate constant is expressed in the standard Arrhenius equation, *i.e.* $r_i = A_i e^{\frac{-E_i}{RT}}$ where r_i is the reaction rate constant, E_i is the activation energy, A_i is the frequency factor for reaction i, R is the ideal gas constant and T is the reaction temperature. The rate constant for the forward (r_1) and the backward (r_2) reactions are as follows:

$$r_1 = 1200e^{\left(\frac{-13{,}000}{1.987T}\right)}\ \frac{\text{litre}}{\text{mole}\cdot\text{sec}} \quad \text{and} \quad r_2 = 100e^{\left(\frac{-12{,}000}{1.987T}\right)}\ \frac{\text{litre}}{\text{mole}\cdot\text{sec}} \qquad (4.6)$$

4.4 CSTR Simulations

In the last section, we developed the dynamic models of the CSTR based on application of physical and chemical principles. The next step is to discuss how the differential equation models might be solved numerically using a digital computer. There are several numerical algorithms in the literature that can be utilised to solve these problems. If the models are non-linear, the solution of the equations may involve an iterative method. If they are linear than numerical integration of ordinary differential equation can be utilised. Normally, discrete finite difference method is used to approximate the continuous differential equations. In this work, we adopted the Euler algorithm for solving the differential equations. The method uses explicit calculation of the derivatives over small time increments. It was chosen because the method is self-starting and easy to use. For a general integration problem, the Euler Algorithm can be presented by:

$$y_{i+1} = y_i + \Delta h f(x_i, y_i) \qquad (4.7)$$

where y_i is the current value of output, y_{i+1} is the next value of output and Δh is the integration step size and $f(x_i,y_i)$ are the functions given by equations 4.1 to 4.5. y_i are represented by *Aout, Bout, Cout, T* and *Tjout* while x_i refer to *Ain, Bin, Tin, Tjin, F* and *Fj*.

This Euler algorithm is started using the knowledge of the initial conditions of all variables, the upper limit of integration and the step size Δh. If the step size is small, the variables estimation will be more accurate but the solution will require extra steps and excessive amount of computer time. The value of Δh must be chosen such that it is a compromise between the accuracy and the steps needed to solve the differential equation. The details about the Euler algorithm can be found in any Numerical Analysis text book. For the purpose of dynamic studies, the initial operating conditions of the CSTR and Δh are presented in Table 4.1

Table 4.1 CSTR initial condition for dynamic response.

T	382.35 K	*Tjout*	359.71 K
F	0.08 litre/sec	*Fj*	0.04 litre/sec
Aout	294.82 mole/litre	*Bout*	294.82 mole/litre
Tin	450 K	*Tjin*	288 K
Ain	500 mole/litre	*Bin*	500 mole/litre
Cout	205.17 mole/litre	Δh	1

4.5 Graphical Fitting of Step Test Results

Once all process parameters have been set up, the process can be simulated digitally. All simulations were executed on an HP Apollo workstation, running the UNIX operating system. The software for the simulation was written in Pascal. The results from the simulation were used to study the dynamic behaviour of the process, *i.e.* to obtain the parameters that will aid in the selection of controller constants.

To obtain the gain (*Kp*), the process time delay (τ_D), and the process time constant (τ_p) of our process, the CSTR was perturbed using several step inputs of different magnitudes. The magnitude of these step functions were +5%, -5%, +10% and -10% of the steady-state values of the various input variables tabulated in Table 4.1. When performing the step-tests, only one input variable at a time was changed. Plots in figure 4.2 shows the responses of various step changes in *Tin*. Examples of the *Cout* responses resulting from these step changes in the input variables are shown in figures 4.3 to 4.8.

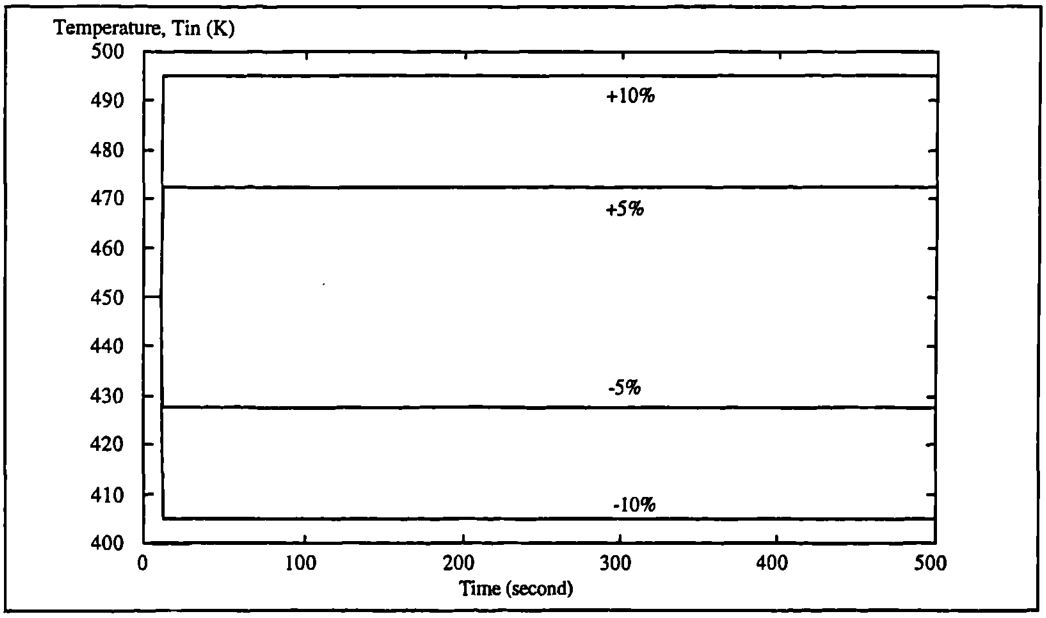

Figure 4.2 Step changes in *Tin*

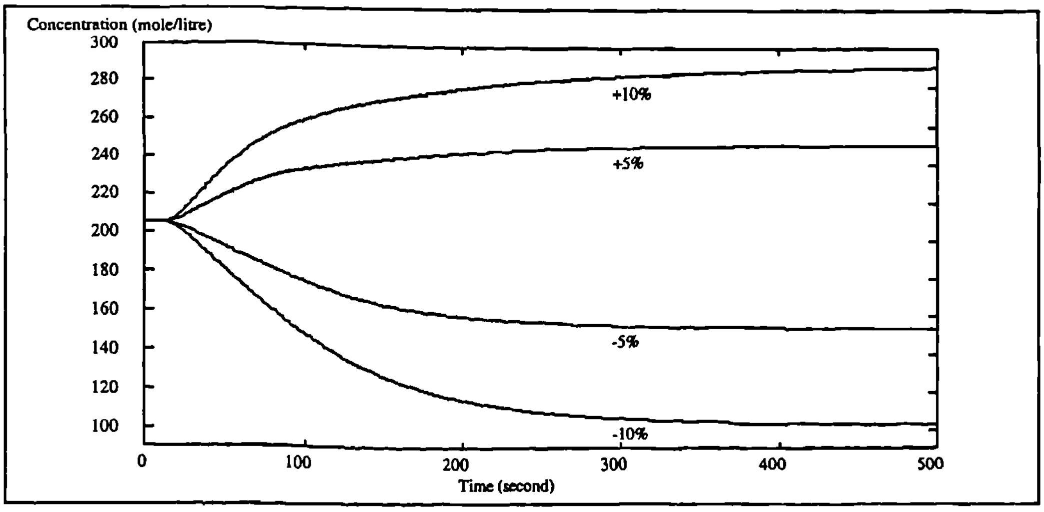

Figure 4.3 Responses of *Cout* to step changes in *Tin*

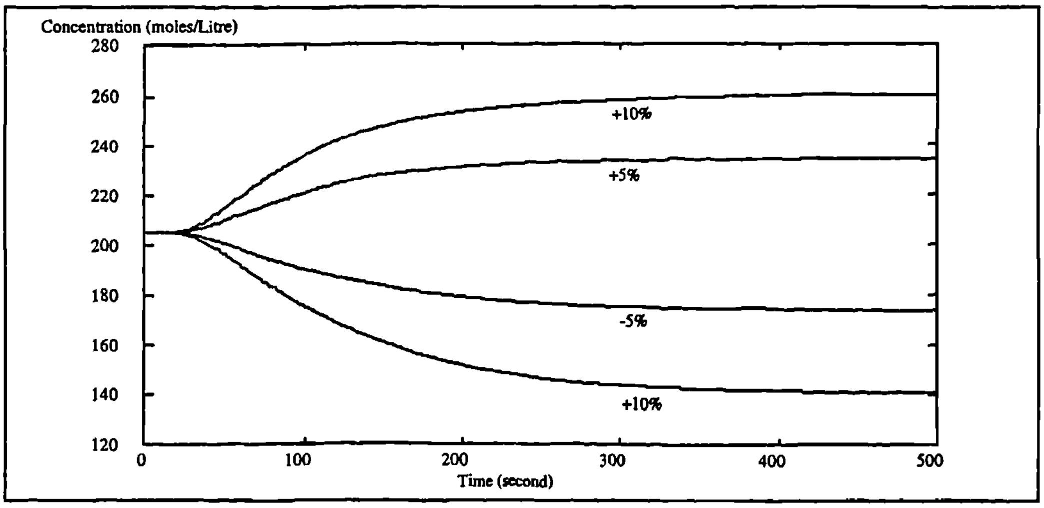

Figure 4.4 Responses of *Cout* to step changes in *Tjin*

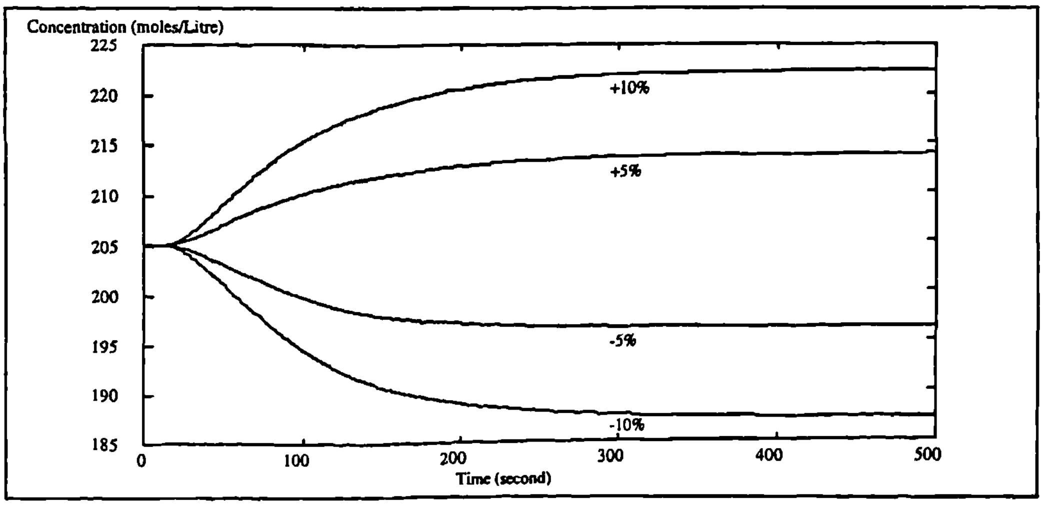

Figure 4.5 Responses of *Cout* to step changes in *Ain*

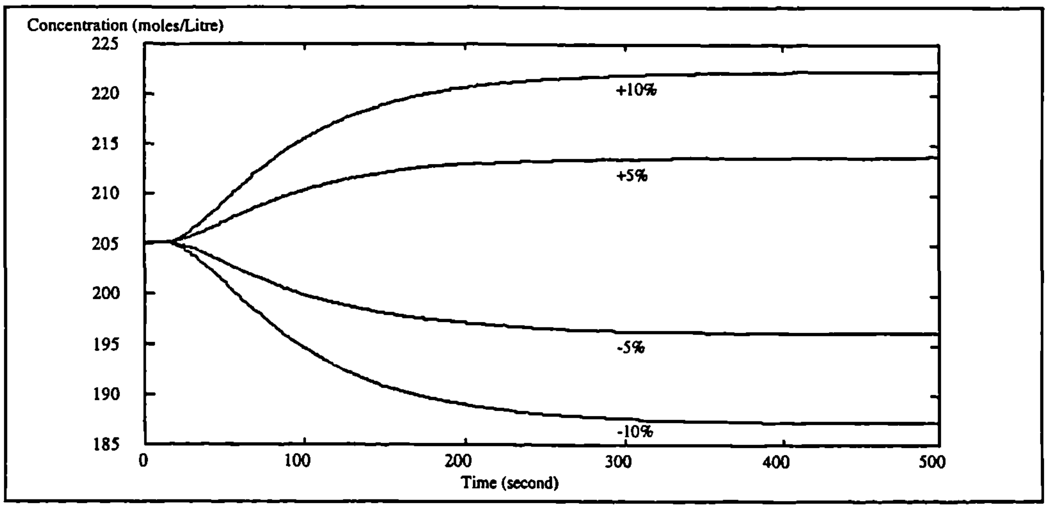

Figure 4.6 Responses of *Cout* to step changes in *Bin*

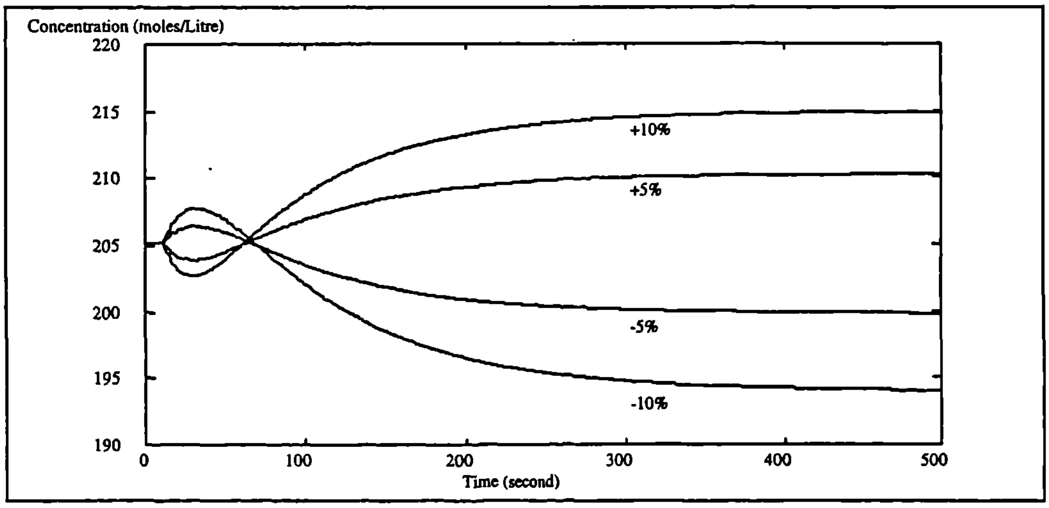

Figure 4.7 Responses of *Cout* to step changes in *F*

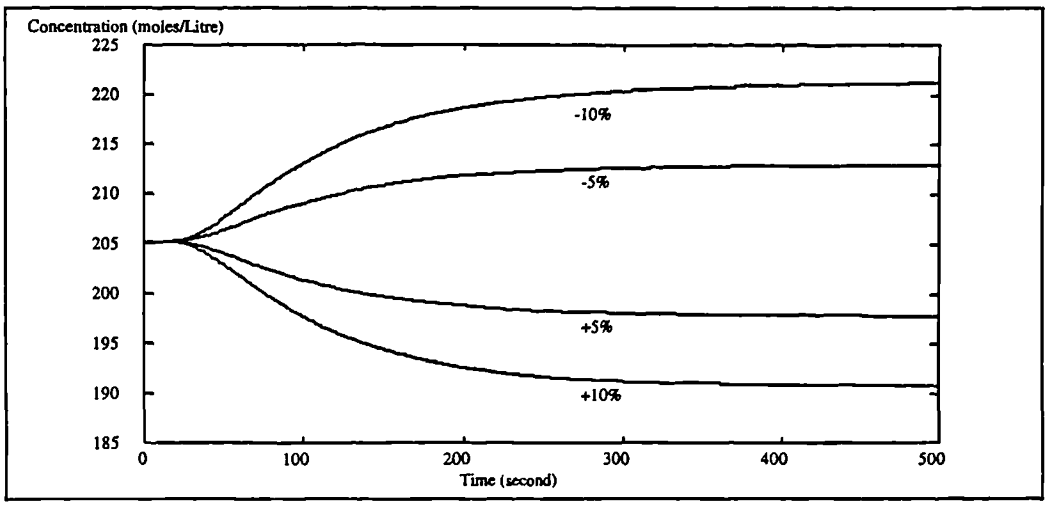

Figure 4.8 Responses of *Cout* to step changes in *Fj*.

Next, the dynamic response were parameterised using the *Process Reaction Curve* technique. The technique, first developed by Cohen and Coon in 1953, is based on the observation that the response of output variable of most open loop systems to a step change in the input variable, has a sigmoidal shape, which can be adequately approximated by a first or second order differential equation with dead time. This is possible because almost all physical processes encountered in a chemical plant are simple first order or multicapacity processes whose responses have the general over-damped shaped (Stephanopoulos, 1984). The transfer function of a first order process with dead time can be mathematically presented using Laplace Transforms as:

$$G_p(s) \approx \frac{K_p e^{-\tau_D s}}{\tau_p s + 1} \tag{4.8}$$

where *Gp* is the process transfer function that relate the input variable to the output variable, *Kp* is the process gain, τ_p is the process time constant and τ_D is the process time delay. Equation 4.8 can be transformed back to time domain and can be presented as:

$$y(t) = K_p(1 - e^{\frac{-t}{\tau_p}})u(t - \tau_D) \tag{4.9}$$

All the figures in the previous section are accompanied by tables 4.2 to 4.7. These tables are tabulated with the size of output response (Bu), the size of step change (M), process gain (*Kp*), process time delay (τ_D), and process time constant (τ_p) for the particular step response in the process.

Table 4.2 Transfer function parameters relating *Cout* to *Tin*

Step Function	Bu (mole/litre)	M (K)	Kp (mole/litre/K)	τ_D (second)	τ_p (second)
+10%	87.35	45	1.9411	10	93.92
+5%	45.68	22.5	2.0302	9	91.89
-5%	-51.51	-22.5	2.2893	10	102.90
-10%	-102.93	-45	2.2873	10	109.38

Table 4.3 Transfer function parameters relating *Cout* to *Tjin*

Step Function	Bu (mole/litre)	M (K)	Kp (mole/litre/K)	τ_D (second)	τ_P (second)
+10%	55.6	28.8	1.931	19	107.52
+5%	29.26	14.4	2.033	19.5	112.74
-5%	-31.43	-14.4	2.183	20.5	126.66
-10%	-64.24	-28.8	2.231	21	130.68

Table 4.4 Transfer function parameters relating *Cout* to *Ain*

Step Function	Bu (mole/litre)	M (mole/litre)	Kp	τ_D (second)	τ_P (second)
+10%	17.41	55	0.348	12	95.34
+5%	8.89	25	0.356	12	95.04
-5%	-8.73	-25	0.349	12	98.52
-10%	-17.83	-50	0.357	12	98.28

Table 4.5 Transfer function parameters relating *Cout* to *Bin*

Step Function	Bu (mole/litre).	M (mole/litre)	Kp	τ_D (second)	τ_P (second)
+10%	17.41	55	0.348	12	95.34
+5%	8.89	25	0.356	12	95.04
-5%	-8.73	-25	0.349	12	98.52
-10%	-17.83	-50	0.356	12	98.28

Table 4.6 Transfer function parameters relating *Cout* to *F*

Step Function	Bu (mole/litre)	M (litre/second)	Kp (mole/second)	τ_D (second)	τ_P (second)
+10%	9.93	0.008	1241.1	28	142.98
+5%	5.21	0.004	1302.5	29	144.60
-5%	-5.21	-0.004	1302.5	31	159.72
-10%	-10.99	-0.008	1373.8	32	162.19

Table 4.7 Transfer function parameters relating *Cout* to *Fj*

Step Function	Bu (mole/litre)	M (litre/second)	Kp (mole/second)	τ_D (second)	τ_P (second)
+10%	-14.2	0.004	-3550	19.5	117.72
+5%	-7.21	0.002	-3605	20	120.84
-5%	7.98	-0.002	-3989	20	119.52
-10%	16.22	-0.004	-4055	20	118.56

4.6 Cross-Correlation Function

The time delays (τ_D) were determined using the *Cross-Correlation* technique because these values were very small. This method measures the correlation between the input variable (x_t) and the output variable (y_t). To elaborate, suppose we have N observations on two variables, x and y. The observation of the bivariate process can be denoted as $(x_1,y_1),\ldots,(x_N,y_N)$. These observations may be regarded as a finite realisation of a discrete stochastic process (x_t,y_t).

To describe the properties of bivariate process, it is useful to know the moments of the process up to second order. For a univariate process, the first two moments are the *mean* and *auto covariance* function. For a bivariate process, the moments up to second order consists of the *mean*, *auto covariance* functions for each of the two components plus a new function, called the cross-covariance function, which is given by:

$$\gamma_{xy}(t,k) = cov(x_t, y_{t+k}) \tag{4.10}$$

The size of the *cross covariance* coefficients depends on the units in which x_t and y_t is measured. To avoid bias estimates, it is advisable to standardise the cross covariance function so that it will produce a function called the cross-correlation function $\rho_{xy}(k)$, which is defined by:

$$\rho_{xy}(k) = \frac{\gamma_{xy}(k)}{\sqrt{\left(\gamma_{xx}(0)\gamma_{yy}(0)\right)}} \tag{4.11}$$

where:

$$\gamma_{xx}(k) = cov(x_t, x_{t+k}) \tag{4.12}$$

$$\gamma_{yy}(k) = cov(y_t, y_{t+k}) \tag{4.13}$$

The *cross correlation* function has the following properties

(a) $\rho_{xy}(k) = \rho_{yx}(-k)$

(b) $\left|\rho_{xy}\right| \leq 1$

The above function will be used to measure the *correlation* between the input variable, (x_t) and output variable *Cout*, (y_{t+k}). A time delay exists when the maximum value of the cross-correlation coefficient is not at $\rho_{xy}(0)$. The time-delay value is determined by how far the maximum value has been shifted to the right. Figure 4.9 shows the plot of the cross-correlation between *Cout* and *Tin*. The results were generated from our simulation and analysed by using a program written using MATLAB. From the figure, we see that the maximum value of the cross-correlation coefficient occurs at *lag k* equal to 9. Since the sampling time use for this step input test is equal to one second, the τ_D value for this case will be 9 seconds. All the τ_D values in table 4.2 to 4.7 were obtained using the same technique.

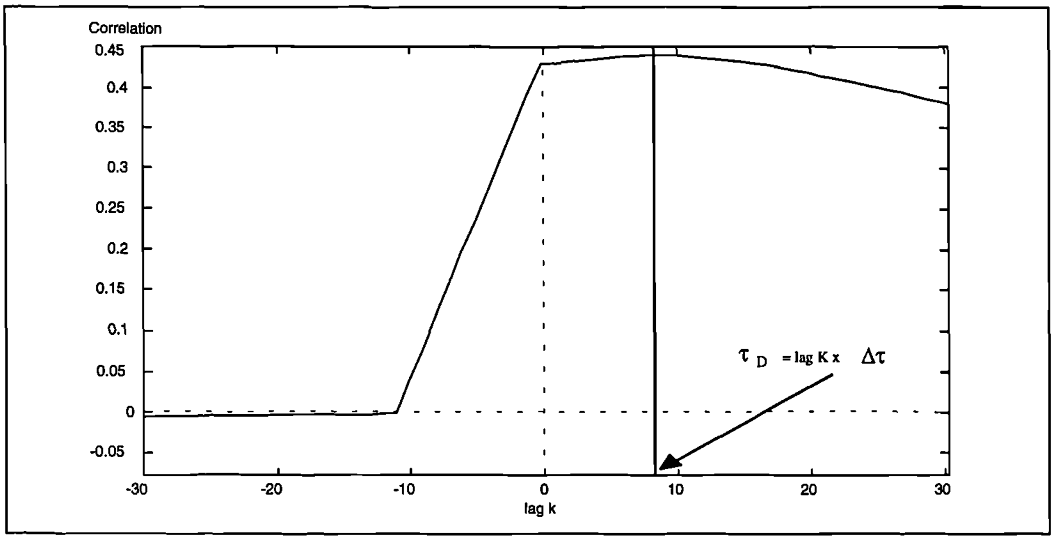

Figure 4.9 Cross-correlation using 5% step increase in initial condition of *Tin*.

4.7 Input-Output Selection

Several judgements could be made based on the input-output responses from figures 4.3 to 4.8. Since it is preferable to choose the input that brings maximum change to the target output, the input that provide the largest steady state gain *Kp* is normally selected. On the other hand we favour input changes that do not exhibit undesirable characteristics such as a large time delay and inverse response in our output variables. At the same time, the influence of control must be quick and thus the process time constant must be small. These qualitative observations are principally based on engineering common sense and can be used as quick guidelines when needed. In addition, they can be used as supporting evidence to verify the results obtained by quantitative analysis.

Our output of interest in this study is the concentration of the product, *Cout*. Figure 4.7 shows that the step changes in reactant flowrate *F*, resulted in inverse responses in *Cout*. To monitor the concentration *Ain* and *Bin* in practise may incur a high cost, because they may require special analytical equipment, *e.g.* gas chromatography. Thus these three variables were eliminated from the selection process. From tables 4.2, 4.3 and 4.7 we found that *Tin* has the smallest values of τ_p and τ_D compared to *Tjin* and *Fj*, indicating a quick response to process change. Although the values of the process gains (*Kp*) involving *Fj* are bigger compared to those associated with *Tin* and *Tjin*, the units of the process gains are not the same. Since *Cout* is strongly dependent on temperature, we choose *Tin* as our manipulated variable to control the CSTR.

4.8 Proportional Integral (*PI*) Feedback Controller

In the previous section we have the selection of the manipulated variable to control CSTR product concentration. This section elaborates the proportional integral (*PI*) controller mode that will be used to control the process. Only this type of feedback controller mode will be used to compare with the SPC methods in Chapter 7. Moreover, this feedback controller is popular in industry.

4.8.1 Digital *PI* Controller

Normally the conventional feedback controllers are analogue devices. The characteristic of this analogue controller is that its input and output signals are in continuous form. Recent advances in the performance and cost of digital equipment such as minicomputers, microcomputers and corresponding digital interface elements have made digital control systems generally preferred over conventional analogue controllers. The advantages of digital control include increased flexibility and accuracy. Moreover it improves plant monitoring through data acquisition, storage and analysis.

A straight forward way of deriving a digital version of the ideal *PI* control law is to replace the integral mode by their discrete equivalents. Thus, by approximating the integral by summation gives:

$$p_n = \bar{p} + K_c \left[e_n + \frac{\Delta t}{\tau_i} \sum_{k=1}^{n} e_k \right] \tag{4.14}$$

where Δt = sampling period of the controlled variable.

p_n = controller output at the *n th* sampling instant, $n = 1, 2, \ldots$

$\bar{p}$ = controller output at steady state

e_n = error at the *n th* sampling instant.

The above equation is referred to as the position form of *PI* control algorithm since the actual controller output is calculated. An alternative approach is to use a velocity form of the algorithm in which the change in controller output is calculated. It is derived by extracting the *(n-1)* sampling instant of the position form:

$$\Delta p_n = p_n - p_{n-1} = K_c\left[(e_n - e_{n-1}) + \frac{\Delta t}{\tau_i} e_n\right] \tag{4.15}$$

There are several advantages to using the velocity form. Firstly, it contains the provision for antireset windup because the summation of errors is not explicitly calculated. Secondly, the output term Δp_n can be utilised directly by final control elements that require an input change in position, such as the valve driven by a pulsed stepping motor. Thirdly, if the final control element has been placed in the appropriate position during the start up procedure it does not require any initialisation when switching back from manual mode to automatic mode. However, if the actual output is needed we can re arrange the above equation to solve for p_n:

$$p_n = p_{n-1} + K_c\left[(e_n - e_{n-1}) + \frac{\Delta t}{\tau_i} e_n\right] \tag{4.16}$$

The purpose of the feedback control system is to ensure that the closed loop system has a desired dynamic characteristics. Hence, it must satisfy certain performance criteria. For example, the closed-loop system must be stable; the effects of disturbances are minimised; it must give rapid, smooth responses to set point-changes; it must also eliminate offset; and it must avoid excessive control actions. In typical control problems, it is not possible to achieve all of these goals since they involve inherent conflicts and trade offs. Thus, the best way is to tune the control parameters using certain criteria.

4.8.2 Controller Tuning

When the control system is installed, the controller settings must usually be adjusted until the control system performance is considered to be satisfactory. Thus, it is desirable to have good preliminary estimates of satisfactory controller settings. Among the earlier tuning method was the Ziegler and Nichols *Ultimate Gain* method (Ziegler and Nichols, 1942) and *Process Reaction Curve* method (Cohen and Coon, 1953). The former devise a trial and error method based on sustained oscillation while the later utilise a step function in the input variables. These two methods are quite expedient and hence find widespread application in the process industries. An alternative approach is to develop a controller design based on a performance index that consider the entire loop performance. Typical performance criteria for this are Integral Absolute Error (*IAE*), Integral Squared Error (*ISE*) and Integral Time Absolute Error (*ITAE*) .

Apart from giving the value of K_p, τ_D, and τ_p the *Process Reaction Curve* technique that we have used before can also be utilised to find the controller setting for the process. Assuming that equation (4.8) is the true process transfer function, Cohen and Coon (1953) derived the theoretical values for the controller settings which will fulfil the criteria of responses having one quarter decay ratio. The decay ratio is the ratio of two successive peaks of under damped oscillation when the process is subjected to step response.

4.9 Results and Discussions

Table 4.8 tabulates the values of the controller settings for the proportional integral (*PI*) control mode using the Cohen Coon tuning method. The manipulated variable as previously mentioned is the input temperature (*Tin*) and the controlled variable is the concentration of C (*Cout*). The sampling time, Δt for the process is roughly one tenth of the time constant tabulated in table 4.2, *i.e.* 10 seconds.

Table 4.8 Controller setting based on Cohen Coon Method

Step Function	Proportional Integral (PI)	
	Kc (mole/litre/K)	τ_i (second)
+10%	4.52	27.24
+5%	4.60	24.87
-5%	4.07	27.68
-10%	4.34	27.95

The integral control mode normally makes the system more sensitive. A high gain (*Kc*) in the *PI* controller will cause the process response to become oscillatory and possibly unstable. Because of this, we prefer to have *Kc* small and τ_i large for the *PI* controller. Based on the above limitation, the settings of the *PI* controller were chosen to be *Kc* = 4.07 and τ_i = 27.68 (see table 4.8) to control the CSTR process.

After the controller settings have been determined the process once again were simulated. This time, all the input variables were injected with white noise, $N(\mu,\sigma^2)$, to introduce some disturbances to the process. The values of μ were 0 while the value of

σ^2 can be specified by the user. The response of the controlled variable (*Cout*) using *PI* controller is depicted in figure 4.10. Figure 4.11 shows the behaviour of manipulated variable (*Tin*). The Cohen-Coon controller design tends to yield oscillatory closed loop response since the design objective is 1/4 decay ratio. If less oscillatory responses are desired, *Kc* should be reduced and τ_i should be increased for *PI* controller setting. It seems that no general conclusions about the relative merits can be drawn from figures 4.10 and 4.11. For comparison with the SPC methods in section 7.5.6 of chapter 7, we would like to determine several parameters: (i) The utilisation of control energy in relation to the nominal value. Further details about this scheme will be elaborated later in section 7.5.6; (ii) The number of control action for which the controller has to take so that the process will not deviate from the setpoint (*Ysp*). The maximum number of control actions are 5000. Table 4.9 summarises the result from this simulation. From the table we can see that the *PI* control mode take control actions all the time to avoid deviation from setpoint.

Table 4.9 The merits of *PI* controller settings.

Controller	*Kp* (mole/Litre/K)	τ_i (second)	Energy Utilisation	Number of Control Action
PI	4.07	27.68	38218.32	5000

4.10 Summary

In this chapter we discussed the basic mathematical modelling of a CSTR process with reversible exothermic reaction. We set-up the equations using chemical and physical principles. Then we proceed with the dynamic studies of the process. We tested the system with a step function in the input variables and observed the process response. Then we used a graphical fitting technique referred to as the *Process Reaction Curve* method to formulate an approximate process transfer function and calculated the values of process parameters. Since the values of time delays, τ_D are small and very difficult to discern from the graphical technique, we used the Cross-Correlation technique. The transfer function model permits prediction of how the CSTR process will react to other types of disturbances or input changes. From there, we proceeded with the tuning and implementation of a *Proportional Integral* (*PI*) feedback controller on the process. In the next chapter we are going to explore the multivariate techniques namely the *Partial Correlations Analysis* and *Principal Component Analysis.* Both of these techniques will play an important role in Active SPC scheme.

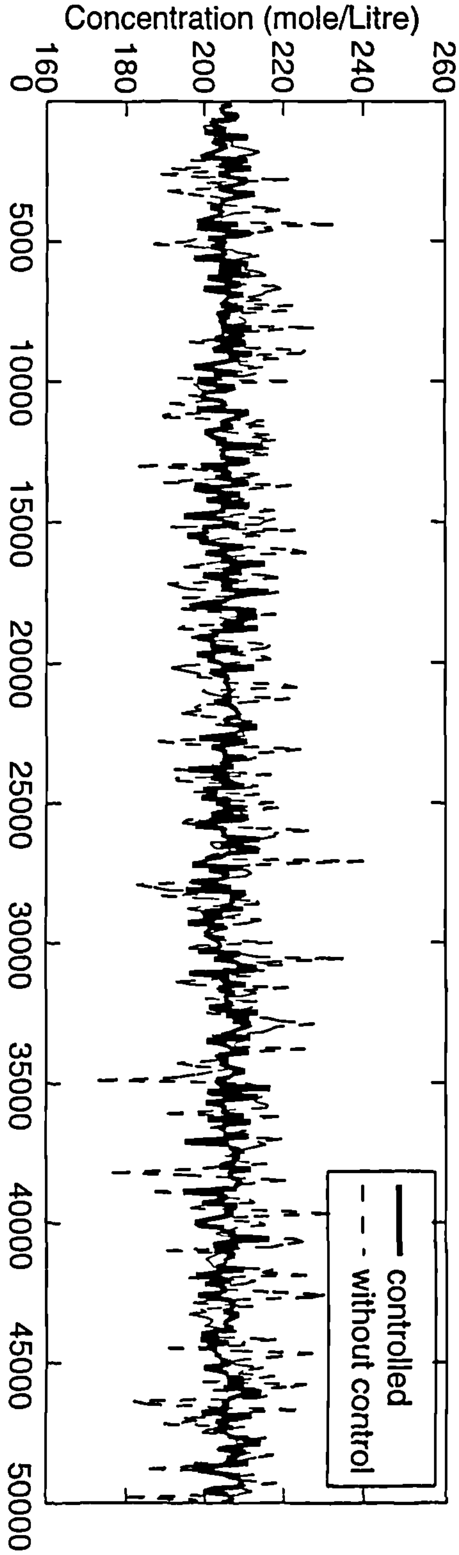

Figure 4.10 Performance of Cout with PI controller

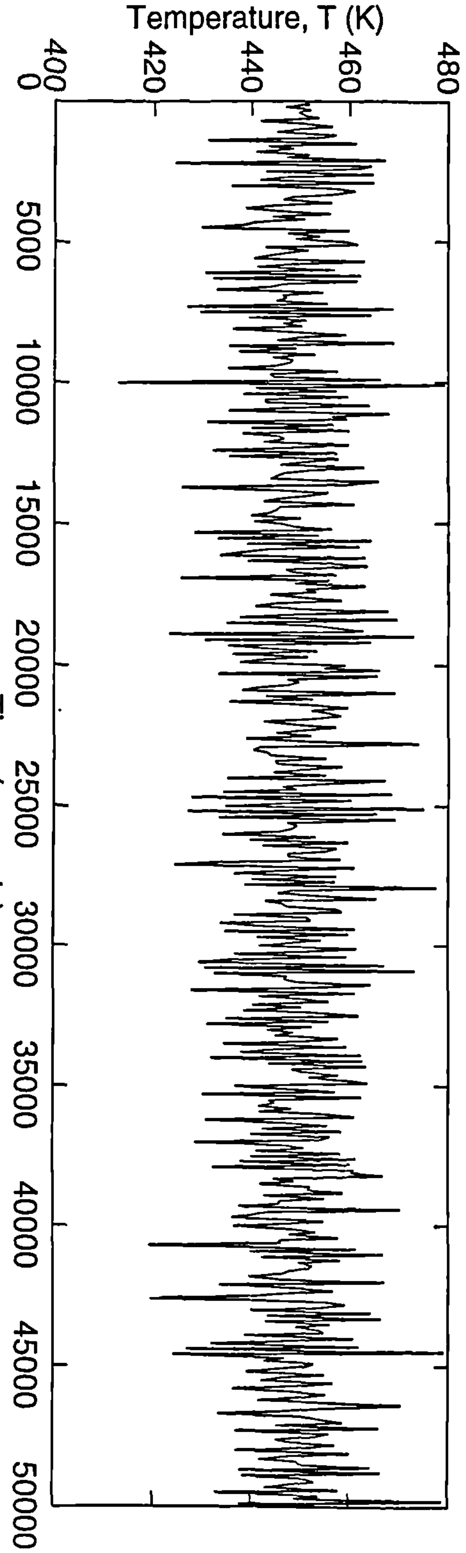

Figure 4.11 Performance of manipulated variable (Tin) due to PI controller

Chapter 5

Multivariate Data Analysis

5.1 Introduction

The advantages of using multivariate analysis compared to univariate analysis lie in the ability of the former to treat all the data simultaneously and to extract information about the variations in the process. It can also give the relevant information on how the variables are behaving relative to one another. This chapter will introduce the basics of multivariate analysis. It focuses on the multivariate analysis tools involved in this work.

Before proceeding further, let us exam the basic problems that normally occur when analysing multiple input and multiple output processes. When faced by a multivariate data set, the analyst often feel overwhelmed by the sheer amount of numbers it contains. An engineer will not consider it excessive to measure twenty different variables on each unit operation. Yet with only ten unit operations there will be 200 variables being recorded. At this stage, the problem becomes apparent, namely that these are not 200 separate unrelated variables. However, when we examine them carefully, some relationships often exist between the 200 variables. These relationships arise as a consequence of all variables being measured on an individual basis. Each piece of equipment will then contribute an $n \times 20$ data matrix, where n is the instance of observation. The natural inclination when presented with this set of numbers, is to scan through them, in the hope of detecting some interesting features or patterns in the data.

The above example implies that any single number in the data matrix must be judged in relation to all the other numbers in the same row as well as the same column of the data matrix. As a result, simple visual inspection of the data matrix is unlikely to show up any immediate patterns that may exist in the numbers. The problems will intensify when more variables and readings are to be measured and analysed. Consequently, formal procedures are needed to help engineers and plant operators to search for this pattern. To overcome all these problems, this part of the thesis will elaborate in detail how a data set is transformed to facilitate analysis. Some of the information may be fundamental, but to assure comprehension and the flow of the thesis, this basic knowledge is included.

This study considers the application of two multivariate statistical methods, Partial Correlation Analysis (PCorrA) and Principal Component Analysis (PCA). Both techniques will be used to analyse and to interpret the data from the CSTR described previously. During the investigation, PCA and PCorrA were used to determine the correct *correlations* between the input and the quality variables. These *correlations* are then used in the calculation of control limits for the Active SPC charts.

5.2 Basic Matrix Properties.

Throughout this thesis, we will be concerned with analysing measurements made on several variables. These measurements, commonly called data, must frequently be arranged and displayed in various ways. To clear the obstacle for later derivation, this section introduces the preliminary concepts of data organisation.

Consider the hypothetical data, representing *m* measurements and *n* variables. The notation x_{ij} denotes the value of the *j*' th variable that is observed at the *i*' th instant. That is:

$x_{ij} =$ *i*' th observation of the *j*' th variable.

Consequently, with *m* number of measurements with *n* variables we can arrange it in matrix form with *m* rows and *n* columns:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{ij} & x_{12} & \cdots & x_{ij} & \cdots & x_{in} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mj} & \cdots & x_{mn} \end{bmatrix} = \begin{bmatrix} \mathbf{x_1} & \mathbf{x_2} & \cdots & \mathbf{x_j} & \cdots & \mathbf{x_n} \end{bmatrix}^{\mathbf{T}} \quad (5.1)$$

The above array or matrix **X** contains all the data in which each column represent a certain variable and each row represent an instance of observations. There are certain advantages when we arrange the data in the form of array. It facilitates exposition and allows numerical calculations to be performed in an orderly and efficient manner. The efficiency can be gained through (1) describing numerical calculations as operations of arrays; (2) implementation on computers, where there are many programming languages and statistical packages that can readily perform array operations.

Throughout the rest of this thesis, lower case bold alphabets will be used to denote a column vector; capital bold alphabets will indicate a matrix while italicised alphabets represent scalars. The first subscript of an entry in an array indicates the row while the second subscript denote the column.

As mentioned previously, a large data set is bulky and can pose a serious obstacle to any attempt at visualisation and extraction of any pertinent information. This problem can be partially avoided through assessing the information in the data by calculating certain summary numbers, known as "descriptive statistics". Some of the descriptive statistics normally used are *sample mean, sample variance, sample covariance* and *sample correlation coefficient.* All these simple methods of analysis will be elaborated in detail below.

Descriptive statistics are used to measure location, spread, and linear association in the data set. The arithmetic average or *sample mean* provides a measure of location of central value for a set of numbers. The *sample mean* can be computed from the m measurements on each n variables, so that in general there are n sample *means*:

$$\bar{x}_j = \frac{1}{m}\sum_{i=1}^{m} x_{ij} \qquad j = 1,2,\ldots,n \qquad (5.2)$$

In vector notation, the sample *means*, will be presented by a vector with n elements:

$$\bar{\mathbf{x}}^{\mathrm{T}} = \begin{bmatrix} \bar{x}_i & \ldots & \bar{x}_j & \ldots & \bar{x}_n \end{bmatrix} \qquad (5.3)$$

A measure of spread is provided by the *sample variance*, where for m measurements on the j' th variable is defined as:

$$\mathrm{var}(\mathbf{x}_j) = s_{jj}^2 = \frac{1}{m-1}\sum_{i=1}^{m}(x_{ij} - \bar{x}_j)^2 \qquad j = 1,2,\ldots,n \qquad (5.4)$$

The square root of the *sample variance*, $\sqrt{s_{jj}^2}$ is known as the sample *standard deviation*. To measure the linear association between the measurements of variable *j* and *k*, we need to find the *sample covariance*. Association here signifies that the values of x_{ij} bear some clear relationship to the corresponding values x_{ik}. Since there are so many possible varieties of non-linear association, the term "association" is restricted almost exclusively to indicate linear association. The fundamental measure of linear association between the variables *j* and *k* is defined as:

$$\text{cov}(\mathbf{x_j}, \mathbf{x_k}) = s_{jk}^2 = \frac{1}{m-1}\sum_{i=1}^{m}(x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k) \tag{5.5}$$

The *sample covariance* reduces to the *sample variance* when $j = k$. Moreover, $s_{jk}^2 = s_{kj}^2$ for all *j* and *k*. The sample covariance s_{jk}^2 can either be positive or negative. If there is no apparent linear trend in the data, the s_{jk}^2 will give a *covariance* near zero. In matrix notation the *sample covariance* will be defined as:

$$\mathbf{S}^2 = \begin{bmatrix} s_{11}^2 & s_{12}^2 & \cdots & s_{1n}^2 \\ s_{21}^2 & s_{22}^2 & \cdots & s_{2n}^2 \\ \vdots & \vdots & \ddots & \vdots \\ s_{n1}^2 & s_{n2}^2 & \cdots & s_{nn}^2 \end{bmatrix} \tag{5.6}$$

The unsatisfactory feature of *sample covariance* as a measure of association is that it is scale dependent. It is influenced by the spread of values in the variable concerned. However, this can be corrected by "standardising" the variable, in which the mean of a variable is subtracted from all measurements of that variable and the result is divided by its standard deviation.

The final descriptive statistic considered is the *sample correlation coefficient*. It is a measure of thc linear association between two variables, even though both variables may have different units. The *sample correlation coefficient*, for the *j*' th and *k*' th variable is defined by:

$$corr(\mathbf{x_j}, \mathbf{x_k}) = r_{jk} = \frac{s_{jk}^2}{\sqrt{s_{jj}^2 s_{kk}^2}} = \frac{\sum_{i=1}^{m}(x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)}{\sqrt{\sum_{i=1}^{m}(x_{ij} - \bar{x}_j)^2}\sqrt{\sum_{i=1}^{m}(x_{ik} - \bar{x}_k)^2}} \tag{5.7}$$

for $j = 1,2,...,n$ and $k = 1,2,...,n$. Note that $r_{jk} = r_{kj}$ for all j and k. In matrix notation the *sample correlation* matrix **R** is :

$$\mathbf{R} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1n} \\ r_{21} & 1 & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & 1 \end{bmatrix} \tag{5.8}$$

Suppose the original values of x_{ij} and x_{ik} are replaced by standardise value defined as:

$$z_{ij} = \frac{(x_{ij} - \bar{x}_j)}{s_{jj}} \quad \text{and} \quad z_{ik} = \frac{(x_{ik} - \bar{x}_k)}{s_{kk}}$$

The standardised values are more favourable than the original variables since both sets are centred to zero mean and expressed in standard deviation units. Thus the *sample correlation coefficient*, r_{jk}, is just the *sample covariance* of the standardised observations.

It is easy to verify that **SRS** = $\mathbf{S}^2$ and **R** = $(\mathbf{S})^{-1}\mathbf{S}^2(\mathbf{S})^{-1}$ and that $\mathbf{S}^2$ can be obtained from **S** and **R**, while **R** can be obtained from $\mathbf{S}^2$. The expression of these relationships in terms of matrix operations allows the calculations to be conveniently implemented on a computer especially using the MATLAB programming environment.

Although the sign of the *sample correlation coefficient* and the *sample covariance* are the same, the *sample correlation coefficient* is ordinarily easier to interpret because its magnitude is bounded and has the following properties:

(1) The value of the *sample correlation coefficient*, r_{jk} is between -1 and +1.

(2) The *sample correlation coefficient*, r_{jk} measures the strength of the linear association. A value of +1 implies a perfect positive linear association between x_j and x_k, a value of -1 implies a perfect negative linear association between x_j and x_k, while a value of zero implies no association between x_j and x_k. It is important to remember, though, that "no association" here signifies no linear association.

The quantities s_{jk}^2 and r_{jk} can convey false association between the two variables especially when an outlier observation and non linear pattern occur in the data set. In spite of the this weakness, *sample covariance* and *sample correlation coefficients* are routinely calculated and analysed. They provide numerical summaries of association when the data do not exhibit obvious non-linear patterns and wild observations (outliers) are not present.

5.3 Partial Correlation Analysis (PCorrA)

Large *correlations* are often picked out from a *correlation* matrix as being of special interest. The variables that exhibit such *correlation* become the focus of attention. Much effort is often expanded on explaining, interpreting and investigating the causes of these large *correlations*. However, picking out isolated entries from a *correlation* matrix can be misleading and sometimes promote incorrect inferences and conclusions. This happens because some or all the entries in the matrix may be interrelated.

In attempting to interpret or explain a high *correlation* between two variables x_i and x_k, therefore, we must be aware that this high *correlation* could be due to mutual association of x_i and x_k with some other variables. If the *correlation* between x_i and x_k is intrinsic, it should remain high when the effects of extraneous or other variables has been removed.

If a variable x_j induces a high *correlation* between x_i and x_k, it must be because there is a considerable variation in the observed values of x_j, and throughout the whole range of data there are strong relationships between x_j and each of x_i and x_k. To examine the "true" association between x_i and x_k, consequently, we must therefore compute the *correlation* between x_i and x_k with the value of x_j held fixed. That way, x_j has no means of influencing the *correlation*. Similarly, if a set of variables $x_j,...,x_n$ is thought to be inducing a high *correlation* between x_i and x_k, then the "true" association is obtained by computing the *correlation* between x_i and x_k when all the values of all variables $x_j,...,x_n$ are held fixed.

To overcome the above problem, a procedure called Partial Correlation Analysis (PCorrA) is suggested (*e.g.* Graybill, 1976). Partial Correlations describe the relative influence of x_i on the variation in x_k when all the variables, $x_j,...,x_n$ are held fixed. The interpretation of partial *correlation* here, is that it is a measure of linear association between x_i and x_k when both variables have been adjusted for their linear association with the remaining variables.

Thus, to gain a better insight into the relationship amongst variables in a multivariate data set, it is desirable to stratify the population into sub-populations in which one or more random variables are held constant and to determine the *correlation* among the other random variables. Consider a $(n\times 1)$ vector **x**, partitioned it into two portions, where $\mathbf{x}^{(1)}$ has q elements and $\mathbf{x}^{(2)}$ has the remaining $(n-q)$ elements. Partition **x** and $\overline{\mathbf{x}}$ as shown below:

$$\mathbf{x}=\begin{bmatrix} x_1 \\ \vdots \\ x_q \\ \hline x_{q+1} \\ \vdots \\ x_n \end{bmatrix}=\begin{bmatrix} \mathbf{x}^{(1)} \\ \hline \mathbf{x}^{(2)} \end{bmatrix} \tag{5.9}$$

$$\overline{\mathbf{x}}=\begin{bmatrix} \overline{\mathbf{x}}^{(1)} \\ \hline \overline{\mathbf{x}}^{(2)} \end{bmatrix} \tag{5.10}$$

where $\overline{\mathbf{x}}^{(1)}$ and $\overline{\mathbf{x}}^{(2)}$ are the sample mean vectors constructed from the observations $\mathbf{x}^{(1)}=\begin{bmatrix} x_1 & \dots & x_q \end{bmatrix}^T$ and $\mathbf{x}^{(2)}=\begin{bmatrix} x_{q+1} & \dots & x_n \end{bmatrix}^T$ respectively. Consequently the *sample covariance* matrix of the partitioned matrix is :

$$\mathbf{S}^2=\left[\begin{array}{c|c} \mathbf{S}_{11}^2 & \mathbf{S}_{12}^2 \\ \hline \mathbf{S}_{21}^2 & \mathbf{S}_{22}^2 \end{array}\right]=\left[\begin{array}{ccc|ccc} s_{1,1}^2 & \cdots & s_{1,q}^2 & s_{1,q+1}^2 & \cdots & s_{1,n}^2 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ s_{q,1}^2 & \cdots & s_{q,q}^2 & s_{q,q+1}^2 & \cdots & s_{q,n}^2 \\ \hline s_{q,n}^2 & \cdots & s_{q+1,q}^2 & s_{q+1,q+1}^2 & \cdots & s_{q+1,n}^2 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ s_{n,1}^2 & \cdots & s_{n,q}^2 & s_{n,q+1}^2 & \cdots & s_{n,n}^2 \end{array}\right] \tag{5.11}$$

The *sample covariance* matrix of $\mathbf{x}^{(1)}$ is $\mathbf{S}_{11}^2$, that of $\mathbf{x}^{(2)}$ is $\mathbf{S}_{22}^2$ and that of elements from $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ is $\mathbf{S}_{12}^2$ or $\mathbf{S}_{21}^2$. The next step is to calculate the partial *covariance* matrix $\mathbf{S}_{11.2}^2$ which is the *sample covariance* matrix of $\mathbf{x}^{(1)}$ after the dependence of $\mathbf{x}^{(2)}$ has been removed and is given by:

$$\mathbf{S}_{11.2}^2=\mathbf{S}_{11}^2-\mathbf{S}_{12}^2\left(\mathbf{S}_{22}^2\right)^{-1}\mathbf{S}_{21}^2 \tag{5.12}$$

Writing $\mathbf{D}_{11.2}$ as the diagonal matrix containing the square roots of diagonal elements of $\mathbf{S}^2_{11.2}$ and standardising $\mathbf{S}^2_{11.2}$ to $\mathbf{R}_{11.2} = \mathbf{D}^{-1}_{11.2}\mathbf{S}^2_{11.2}\mathbf{D}^{-1}_{11.2}$, we obtain the matrix of sample Partial Correlations. The (i,k) th element of this matrix is an estimate of the *correlation* between the i' th and k' th variates of $\mathbf{x}^{(1)}$ when the value of variates $\mathbf{x}^{(2)}$ is fixed. This removes the effect of the variables of $\mathbf{x}^{(2)}$ on the relationship between x_i and x_k, and gives an estimate of the intrinsic association between x_i and x_k.

From the above method, the partial correlation coefficient between x_i and x_k which are in $\mathbf{x}^{(1)}$ given a constant $\mathbf{x}^{(2)}$ is:

$$r_{ik.(q+1,\ldots n)} = \frac{s^2_{ik.(q+1,\ldots,n)}}{\sqrt{s^2_{ii.(q+1,\ldots,n)}\, s^2_{kk.(q+1,\ldots,n)}}} \tag{5.13}$$

where $s^2_{ik.(q+1,\ldots,n)}$ is the ik' th element of $\mathbf{S}^2_{11.2}$.

Alternatively, the partial correlation can be built up by using a sequence of recurrent relationships. Denote by $r_{ik.l}$ the "first order" partial correlation between x_i and x_k when fixing x_l. Next continue with $r_{ik.ln}$ the "second order" partial correlation between x_i and x_k on fixing both x_l and x_n, and so on. The sequential application of this method starts from the *correlation* matrix. The relationship yields:

$$r_{ik.l} = \frac{(r_{ik} - r_{il}r_{kl})}{\sqrt{\{(1 - r_{il}^2)(1 - r_{kl}^2)\}}} \tag{5.14}$$

$$r_{ik.ln} = \frac{(r_{ik.l} - r_{in.l}\, r_{kn.l})}{\sqrt{\{(1 - r_{in.l}^2)(1 - r_{kn.l}^2)\}}} \tag{5.15}$$

This recursion pattern can be used to calculate higher order partial correlations and enables a q' th order partial correlation to be obtained from those of $(q-1)$ th order partial correlations. This method is most useful when low order partial correlations are required.

5.4 Principal Component Analysis (PCA)

When a large number of variables are monitored, it is natural to enquire whether they could be replaced by a fewer number of variables or appropriate functions provided the original information is retained. This is made possible by applying a technique known as Principal Component Analysis (PCA). PCA is able to look at a single set of variables and attempts to access the structure of the variables in this set, independently of any relationship it may have to variables outside the set (Wold *et al.*, 1987; Jackson, 1980; Jackson, 1991). The technique can be used to reduce the dimensionality of a set of variables, that is described by equation (5.1) or matrix **X**, with little loss of information. This is achieved via the definition of a new set of variables.

The new variables derived by PCA called Principal Components, are simply weighted sums of the original variables. The weights are given by the eigenvector of $\mathbf{X}^T\mathbf{X}$ where **X** is the original data matrix. The elements of each eigenvector actually define the orientation of a particular Principal Component line in the co-ordinates of the original data space. Each of the Principal Component lines describes the maximum data variation in their respective dimensions and the variance of each Principal Components is given by the corresponding eigenvalues of matrix $\mathbf{X}^T\mathbf{X}$ (Tham, 1995). The first Principal Component is a linear combination of the original variables which gives the largest *variance* of the original data. It is therefore relevant to examine in what sense PCA can provide a reduction of data without losing the information that we are seeking. The premier study on PCA can probably be traced back in the work of Pearson (1901), while the statistical properties of PCA were investigated in detail by Hotelling (1933).

The objective of Principal Components analysis is to decompose the total variation of a set of original variables into new linearly independent composite variables, so that each Principal Component successively accounts for the maximal variability in the data. Unique linear combinations from the original variables is achieved through the computations of eigenvalues and eigenvectors of the characteristic equations for the *covariance* or *correlation* matrix. The transformations are performed such that each principal components have the maximum variation in that dimension of the transformed space. Further, we require that the principal components are pairwise uncorrelated.

Only a few Principal Components are needed to summarise the data adequately since the original variables are generally intercorrelated with each other to a certain degrees. In practice, we usually retain only the first few principal components that account for the major pattern of variation. This is because each successive component accounts for a smaller amount of *variance* in the sample. Thus, we are assured that the total variance described by these first few axes is maximal for the chosen dimensionality. Due to this characteristic, Principal Component analysis is a widely utilised method for summarising data in few dimensions while retaining most of the essential information from the original data.

5.4.1 Geometrical Concepts of PCA

To give the real flavour of Principal Components Analysis let us turn to the object space, and consider the representation of a two dimensional sample of *m* individuals. In order to label the axes and to avoid confusion, the variates or axes are denoted by bold lower case alphabet and the individual value by italicised alphabet of lower case. The two dimensional sample, after mean centring can be presented by a scatter plot given by Figure 5.1.

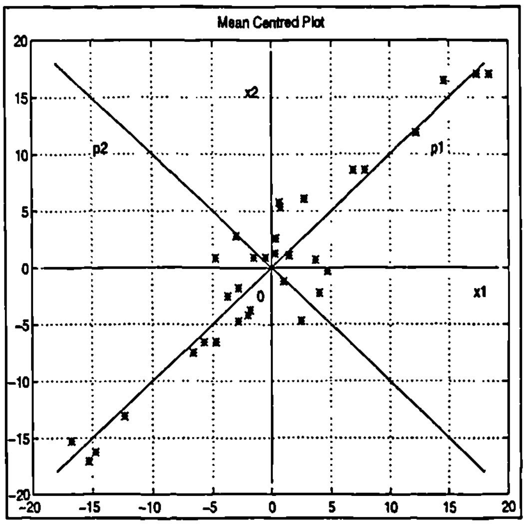

Figure 5.1 Mean Centring Scatter Plot

Variable $\mathbf{x}_1$ and $\mathbf{x}_2$ is represented by axes $\mathbf{ox}_1$ and $\mathbf{ox}_2$ respectively. When performing the data analysis, the reference point of the data configuration is important. For example, in Figure 5.1 we can rotate the axes of $\mathbf{ox}_1$ and $\mathbf{ox}_2$ to new positions $\mathbf{op}_1$ and $\mathbf{op}_2$ without altering any data points. The points are then related to the new axes for future analysis, and may actually contain some useful meaning to the investigator, indeed sometimes even more meaningful than the original data set.

Through this rotation of axes (Figure 5.1), the new individual's co-ordinates on $\mathbf{op}_1$ and $\mathbf{op}_2$ are given by :

$$\mathbf{p}_1 = \mathbf{x}_1 \cos\alpha + \mathbf{x}_2 \sin\alpha = v_{11}\mathbf{x}_1 + v_{12}\mathbf{x}_2 \quad (5.16)$$

$$\mathbf{p}_2 = -\mathbf{x}_1 \sin\alpha + \mathbf{x}_2 \cos\alpha = v_{21}\mathbf{x}_1 + v_{22}\mathbf{x}_2 \quad (5.17)$$

Where α is the angle between $\mathbf{ox_1}$ and $\mathbf{op_1}$ or $\mathbf{ox_2}$ and $\mathbf{op_2}$. Hence from the equations, $\mathbf{p_1}$ and $\mathbf{p_2}$ are linear combinations of $\mathbf{x_1}$ and $\mathbf{x_2}$. If all the coefficients of v_{ij} are collected to form a new matrix $\mathbf{V}$, this new matrix is orthogonal, *i.e.* it's inverse is equal to its transpose $\mathbf{V^T}$. This can be proved by:

$$\mathbf{V^T V} = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix}$$

$$= \begin{bmatrix} \cos^2\alpha + \sin^2\alpha & 0 \\ 0 & \cos^2\alpha + \sin^2\alpha \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\mathbf{V^T V} = \mathbf{I} \qquad (5.18)$$

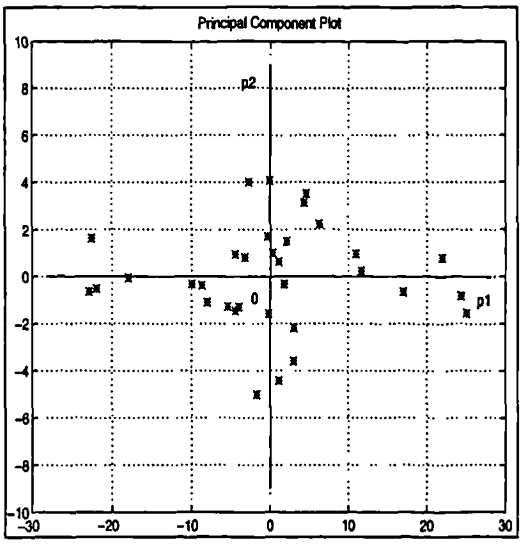

Figure 5.2 Principal Component Plot

When the axes of $\mathbf{p_1}$ and $\mathbf{p_2}$ are re-drawn (Figure 5.2), there is a wide spread of values along the $\mathbf{p_1}$ axis. In contrast, the spread of values along the $\mathbf{p_2}$ axes is relatively small. Thus, we can conclude that it is possible to approximate the two dimensional system using only one dimension. Thus, we can characterise the *n* individuals sufficiently well, by quoting the co-ordinates in terms of $\mathbf{x_1}$ and $\mathbf{x_2}$ for each point, by simply referring to $p_{11} = x_{11} \cos\alpha + x_{12} \sin\alpha$. Replacing the two original variables or co-ordinates by a single derived variable $\mathbf{p_1}$ in this way effects a reduction in dimensionality from 2 to 1. This is because we can represent the sampled data by plotting the individuals to their $\mathbf{p_1}$ values.

Accordingly, different values of α will give different axes of $\mathbf{p_1}$ and hence different plots. Amongst all these plots, there will be one that is deemed to be the best that can represent the "truest" impression of the relationship that exist between the *n* points in the two dimensional space. The "truest" impression of all the relationships will be provided by that value of α that gives rise to the smallest displacement of all points from the original position.

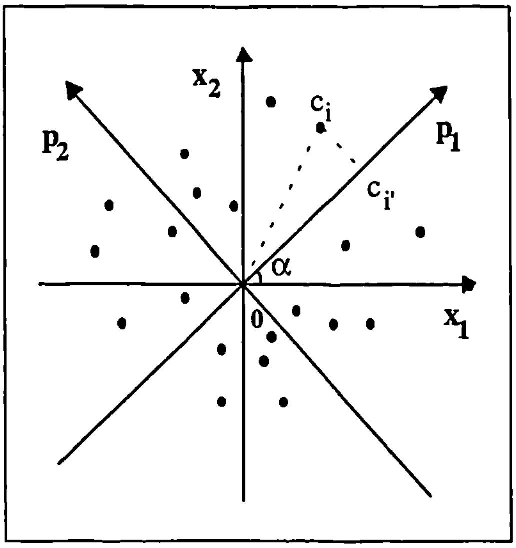

Figure 5.3 Geometric perspective of Principal Components

From Figure 5.3, the co-ordinate value of c_i can be projected orthogonally to axes $\mathbf{op_1}$ and indicated by c_i'. In order to obtain the line $\mathbf{op_1}$ that fits best all the points, Pearson (1901) minimised the distance between c_ic_i' through minimising $\sum_{i=1}^{m}(c_ic_i')^2$. In contrast, regression lines minimise the sum of square of either horizontal or vertical displacements. Applying Pythagoras' Theorem to the triangle oc_ic_i', we obtain:

$$(oc_i)^2 = (oc_i')^2 + (c_ic_i')^2 \qquad (5.19)$$

Considering all data points,

$$\sum_{i=1}^{m}(oc_i)^2 = \sum_{i=1}^{m}(oc_i')^2 + \sum_{i=1}^{m}(c_ic_i')^2 \qquad (5.20)$$

The value on the left hand side of the above equation is fixed, irrespective of the co-ordinate system employed. Thus, choosing $\mathbf{op_1}$ to minimise the value of $\sum_{i=1}^{m}(c_ic_i')^2$ is equivalent to maximising the value of $\sum_{i=1}^{m}(oc_i')^2$. The latter maximisation finds $\mathbf{op_1}$ such that the projections of the points on it have maximum *variance*. This was how

Hotelling (1933) approached the derivation of Principal Components. Choosing $\mathbf{op_1}$ to ensure the smallest possible perpendicular deviation of all the points is equivalent to the choice of rectangular axes that gives the smallest spread of projections on $\mathbf{op_2}$ and hence the largest spread on $\mathbf{op_1}$.

For an n-dimensional data set, with an associated ($m \times n$) data matrix, a similar sequence of steps is adopted. The data is modelled as usual by a swarm of m points in n dimensions, each corresponding to a measured variable. First, the principal axis $\mathbf{op_1}$ is found such that the spread of m points when projected into it is maximum. This operation defines a derived variable of the form:

$$\mathbf{p_1} = v_{11}\mathbf{x_1} + v_{21}\mathbf{x_2} + \ldots + v_{n1}\mathbf{x_n} \qquad (5.21)$$

with $\sum_{i=1}^{n}(v_{i1})^2 = 1$ and that the *variance* of $\mathbf{p_1}$ is maximised.

Having obtained $\mathbf{op_1}$, look for the next principal axes or line $\mathbf{op_2}$ orthogonal to $\mathbf{op_1}$ such that the spread of points in the remaining (n-1) dimensional subspace when projected to this line is maximum. However this spread would not be greater than the spread on $\mathbf{op_1}$. This equivalently indicates finding a line that is at right angles to $\mathbf{op_1}$. The next task is to find the (n-2) dimensional space that is orthogonal to $\mathbf{op_1}$ and $\mathbf{op_2}$, *i.e.* a line that is at right angles to both $\mathbf{op_1}$ and $\mathbf{op_2}$ such that the *variance* or spread is maximum after the spreads of $\mathbf{op_1}$ and $\mathbf{op_2}$ have been accounted for. The process is continued until all n mutually orthogonal lines of $\mathbf{op_i}$ (i = $1,\ldots,n$) have been determined. Each has a derived variable of the form:

$$\mathbf{p_i} = v_{1i}\mathbf{x_1} + v_{2i}\mathbf{x_2} + \ldots + v_{ni}\mathbf{x_n} \qquad (5.22)$$

where the constants v_{ij} are determined by the requirement that the *variance* of $\mathbf{p_i}$ is maximum but subject to certain orthogonal constraints, that is:

$$\begin{array}{ll} \textit{maximise} & var(\mathbf{p_i}) \\ \textit{subject to} & \mathbf{v_i^T v_i} = 1 \\ & \mathbf{v_i^T v_k} = 0 \text{ where } i \neq k \end{array}$$

The $\mathbf{p_i}$ obtained from this procedure are called the Principal Components of the system and analysis based on these new variables is called Principal Components Analysis.

5.4.2 PCA Mathematical Details

The previous section introduced Principal Components geometrically as those variates corresponding to the principal axes of the scatter diagram. In order to understand the subject thoroughly, this section provides the mathematical interpretation of the new variables. It will show that the Principal Components are just linear combinations of the original variables which explain progressively smaller portions of the total *sample variance.*

Consider, a set of data in the form of equation (5.1), comprising m measurements and n variables. The definition of Principal Components indicates that we can generate n new variables from the original n variables. Each of this new variables is a linear combination of the original variables such that:

$$\mathbf{p_i} = v_{1i}\mathbf{x_1} + v_{2i}\mathbf{x_2} + \dots + v_{ni}\mathbf{x_n}. \qquad i = 1,2,\dots,n$$

The Principal Component matrix is defined by:

$$\mathbf{P} = \mathbf{XV} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{n1} & v_{n2} & \cdots & v_{nn} \end{bmatrix} = \sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij}v_{jk} \qquad (5.23)$$

where $k = 1,2,\dots,n$ and each column of matrix $\mathbf{V}$ contains the coefficients for the Principal Components. The first Principal Component, is expressed as:

$$\mathbf{p_1} = v_{11}\mathbf{x_1} + v_{21}\mathbf{x_2} + \dots + v_{n1}\mathbf{x_n}.$$

that is:

$$\mathbf{p_1} = \begin{bmatrix} p_{11} \\ p_{21} \\ \vdots \\ p_{m1} \end{bmatrix} = \mathbf{Xv_1} = \begin{bmatrix} x_{11}v_{11} + x_{12}v_{21} + \dots + x_{1n}v_{n1} \\ x_{21}v_{11} + x_{22}v_{21} + \dots + x_{2n}v_{n1} \\ \vdots \\ x_{m1}v_{11} + x_{m2}v_{21} + \dots + x_{mn}v_{n1} \end{bmatrix} = \sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij}v_{j1}$$

The above equation shows how the original matrix $\mathbf{X}$ can be converted to the first Principal Component. To obtain the relationship between the original variables and the Principal Components, here we will elaborate in detail the method for finding the covariance of Principal Components in terms of the original variable $\mathbf{X}$. This derivation is important for the constraint optimisation problem to be considered later on. Before we proceed, let us express the *mean* and *covariance* matrix **X** via matrix operations. The *mean* of matrix **X** can be written as follows:

$$\bar{\mathbf{x}}^{\mathbf{T}} = \frac{1}{m}\mathbf{1}^{\mathbf{T}}\mathbf{X} = \frac{1}{m}\underset{(m\times 1)}{[1 \ \cdots \ 1]}\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{12} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mm} \end{bmatrix}$$

$$= \left[\bar{x}_1 \ \ldots \ \bar{x}_j \ \ldots \ \bar{x}_n\right] \tag{5.24}$$

In order to convert the vector $\bar{\mathbf{x}}^{\mathbf{T}}$ into a matrix, the above equation is pre-multiplied by a $(m\times 1)$ vector of 1s.

$$\mathbf{1}\bar{\mathbf{x}}^{\mathbf{T}} = \frac{1}{m}\mathbf{1}\mathbf{1}^{\mathbf{T}}\mathbf{X} = \frac{1}{m}\underbrace{\begin{bmatrix}1\\ \vdots \\ 1\end{bmatrix}}_{(1\times m)}\underset{(m\times 1)}{[1 \ \cdots \ 1]}\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{12} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mm} \end{bmatrix}$$

$$= \begin{bmatrix} \bar{x}_1 & \bar{x}_2 & \ldots & \bar{x}_n \\ \bar{x}_1 & \bar{x}_2 & \ldots & \bar{x}_n \\ \vdots & \vdots & \ddots & \vdots \\ \bar{x}_1 & \bar{x}_2 & \ldots & \bar{x}_n \end{bmatrix} \tag{5.25}$$

The *mean* centred matrix is obtained by :

$$\mathbf{X} - \frac{1}{m}\mathbf{1}\mathbf{1}^{\mathbf{T}}\mathbf{X} \tag{5.26}$$

The *sample covariance* matrix is given by:

$$\mathbf{S}^2 = \frac{1}{m-1}\left(\mathbf{X} - \frac{1}{m}\mathbf{1}\mathbf{1}^{\mathbf{T}}\mathbf{X}\right)^{\mathbf{T}}\left(\mathbf{X} - \frac{1}{m}\mathbf{1}\mathbf{1}^{\mathbf{T}}\mathbf{X}\right) \approx \frac{1}{m-1}\mathbf{X}^{\mathbf{T}}\left(\mathbf{I} - \frac{1}{m}\mathbf{1}\mathbf{1}^{\mathbf{T}}\right)\mathbf{X} \tag{5.27}$$

The simplification arises because:

$$\left(\mathbf{I} - \frac{1}{m}\mathbf{1}\mathbf{1}^{\mathbf{T}}\right)^{\mathbf{T}}\left(\mathbf{I} - \frac{1}{m}\mathbf{1}\mathbf{1}^{\mathbf{T}}\right) = \mathbf{I} - \frac{2}{m}\mathbf{1}\mathbf{1}^{\mathbf{T}} + \frac{1}{m^2}\mathbf{1}\mathbf{1}^{\mathbf{T}}\mathbf{1}\mathbf{1}^{\mathbf{T}} = \mathbf{I} - \frac{1}{m}\mathbf{1}\mathbf{1}^{\mathbf{T}} \tag{5.28}$$

To find the *sample covariance* of matrix **P**, we can again use equation (5.27):

$$\mathbf{S}^2_{(\mathbf{p})} = cov(\mathbf{P}) = \frac{1}{m-1}\mathbf{P}^{\mathrm{T}}\left(\mathbf{I}-\frac{1}{m}\mathbf{1}\mathbf{1}^{\mathrm{T}}\right)\mathbf{P} \tag{5.29}$$

However from equation (5.23), **P** = **XV**. The above equation then becomes:

$$\begin{aligned} cov(\mathbf{P}) &= \frac{1}{m-1}(\mathbf{XV})^{\mathrm{T}}\left(\mathbf{I}-\frac{1}{m}\mathbf{1}\mathbf{1}^{\mathrm{T}}\right)\mathbf{XV} \\ &= \frac{1}{m-1}\mathbf{V}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\left(\mathbf{I}-\frac{1}{m}\mathbf{1}\mathbf{1}^{\mathrm{T}}\right)\mathbf{XV} \\ &= \mathbf{V}^{\mathrm{T}}\left\{\frac{1}{m-1}\mathbf{X}^{\mathrm{T}}\left(\mathbf{I}-\frac{1}{m}\mathbf{1}\mathbf{1}^{\mathrm{T}}\right)\mathbf{X}\right\}\mathbf{V} \end{aligned} \tag{5.30}$$

From equation (5.27), equation (5.30) becomes:

$$\mathbf{S}^2_{(\mathbf{p})} = cov(\mathbf{P}) = \mathbf{V}^{\mathrm{T}}\mathbf{S}^2\mathbf{V} = \mathbf{L} \tag{5.31}$$

From the above equation, we can see that Principal Components is based upon a *key result* from matrix algebra. An ($n \times n$) symmetric, non singular matrix, such as the *covariance* matrix $\mathbf{S}^2$, can be reduced to a diagonal matrix **L** by pre-multiplying and post-multiplying it with a particular orthonormal matrix **V**. The diagonal elements of **L**, $l_1, l_2, \ldots, l_n$ are called *characteristics roots, latent roots* or *eigenvalues* of $\mathbf{S}^2$. The columns of **V**, $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are called the characteristic vectors, principal axes vectors or eigenvectors of $\mathbf{S}^2$.

To elaborate further, let us consider the derivation of the first Principal Component. As defined, the first Principal Component, $\mathbf{p}_1$ is the linear combination of the original variables $\mathbf{X}\mathbf{v}_1$. It will give rise to the maximum variation of $var(\mathbf{p}_1)$ when the value of $\mathbf{v}_1^{\mathrm{T}}\mathbf{S}^2\mathbf{v}_1$ is maximised subject to the constraint $\mathbf{v}_1^{\mathrm{T}}\mathbf{v}_1 = 1$. This leads to the constraint optimisation problem where:

$$\begin{aligned} &\textit{maximise} \quad && (\mathbf{S}^2_{(\mathbf{p})1}) = var(\mathbf{p}_1) = \mathbf{v}_1^{\mathrm{T}}\mathbf{S}^2\mathbf{v}_1 \\ &\textit{subject to} \quad && \mathbf{v}_1^{\mathrm{T}}\mathbf{v}_1 = 1 \end{aligned}$$

Using the Lagrange multiplier λ_1, the $\mathbf{v_1}$ that maximises $var(\mathbf{p_1})$ is the value that sets $\frac{\partial Y1}{\partial v_1} = 0$, where :

$$Y_1 = v_1^T S^2 v_1 - \lambda_l (v_1^T v_1 - 1) \tag{5.32}$$

and hence,

$$\frac{\partial Y_1}{dv_1} = 2(S^2 - \lambda_l I) v_1 = 0 \tag{5.33}$$

$$S^2 = \lambda_1 I \tag{5.34}$$

For a non trivial solution, the elements of $\mathbf{v_1}$ cannot be a zero or null vector. Thus:

$$\left| S^2 - \lambda_l I \right| = 0 \tag{5.35}$$

To continue, consider the second Principal Component. The linear combination for this step is $\mathbf{p_2} = \mathbf{Xv_2}$. It will have the same condition as before, *i.e.* $v_2^T v_2 = 1$, where $v_2^T =$ $[v_{21}, v_{22}, ..., v_{2n}]$. Since the principal components are uncorrelated with each other, it must fulfil another criteria which is $v_2^T v_1 = v_1^T v_2 = 0$. The variance of $\mathbf{p_2}$ is $v_2^T S^2 v_2$. In order to maximise the *variance* subject to the above two constraints, we again have to introduce Lagrange Multipliers. With two constraints, we need to use two such multipliers, λ_2 and θ, and the maximisation criteria are as follows:

maximise	$(S^2_{(p)2}) = var(\mathbf{p_2}) = v_2^T S^2 v_2$
subject to	$v_2^T v_2 = 1$
	$v_2^T v_1 = v_1^T v_2 = 0$

$$Y_2 = v_2^T S^2 v_2 - \lambda_2 (v_2^T v_2 - 1) - \theta v_2^T v_1 \tag{5.36}$$

Thus to maximise $\mathbf{Y_2}$, we set

$$\frac{\partial Y_2}{\partial v_2} = 2(S^2 - \lambda_2 I) v_2 - \theta v_1 \tag{5.37}$$

to zero, yielding:

$$2(S^2 - \lambda_2 I)v_2 - \theta v_1 = 0 \qquad (5.38)$$

Pre-multiplying the above equation by v_1^T and since $v_1^T v_1 = 1$ while $v_1^T v_2 = 0$ we arrive at:

$$2v_1^T(S^2 - \lambda_2 I)v_2 = \theta v_1^T v_1$$

$$2v_1^T S^2 v_2 = \theta \qquad (5.39)$$

From the derivation of p_1 in equation (5.33)

$$2(S^2 - \lambda_1 I)v_1 = 0$$

Pre multiplying this equation by v_2^T and noting that $v_2^T v_1 = 0$

$$2v_2^T(S^2 - \lambda_1 I)v_1 = 0$$

$$2v_2^T S^2 v_1 = 0 \qquad (5.40)$$

Since $v_1^T S^2 v_2$ is a scalar quantity and S^2 is symmetric matrix then $v_1^T v_2 = v_2^T v_1 = 0$. Substituting this into equation (5.39) leads to $\theta = 0$ and hence from equation (5.38) we can conclude that the coefficients v_2 of the second Principal Component also satisfy:

$$(S^2 - \lambda_2 I)v_2 = 0 \qquad (5.41)$$

Using a similar reasoning used in the derivation of p_1, the second Principal Component's *variance*, λ_2, must be the second largest after the first Principal Component has been accounted for. Thus the coefficients of the second Principal Component are given by the elements of column vector v_2, the eigenvector corresponding to the second largest eigenvalue λ_2 of S^2.

The previous procedure could be extended to calculate the remaining Principal Components. Although the formal algebra is tedious, the results are essentially generalisations of the above procedure which can be summarised as:

maximise $\quad var(p_i) = \lambda_i = v_i^T S^2 v_i$ where $\lambda_1 > \lambda_2 > \dots > \lambda_i > \dots > \lambda_n$

subject to $\quad v_i^T v_i = 1$

$$v_i^T v_k = v_k^T v_i = 0 \quad k \neq i \quad k = 1,2,\ldots,n$$

5.4.3 Principal Component via Singular Value Decomposition

Principal Components can be derived using several methods. All these methods lead to the solution of an eigenvalue-eigenvector problem. There is a powerful technique that can cope with equations or matrices that are either singular or numerically very close to singular. This technique is known as Singular Value Decomposition (SVD) (Rawlings, 1988; Press, 1989). Any $(m \times n)$ matrix of variables **X**, whose number of rows m is greater than or equals to its number of columns n, can be written as the product of an $(m \times m)$ column-orthogonal matrix **U**, a diagonal matrix $\mathbf{L}^{1/2}$ with positive elements, and the transpose of an $(n \times n)$ orthogonal matrix **V**,

$$\mathbf{X} = \mathbf{U}\,\mathbf{L}^{1/2}\,\mathbf{V}^T \tag{5.42}$$

U is the eigenvector matrix of $\mathbf{XX}^T$ while **V** is the eigenvector matrix of $\mathbf{X}^T\mathbf{X}$. $\mathbf{L}^{1/2}$ is a diagonal matrix whose elements are the positive square roots of the eigenvalues, λj $(j = 1,\ldots,n)$ of $\mathbf{X}^T\mathbf{X}$ and are called *singular values*. The matrices **U** and **V** have the following properties:

$$\mathbf{UU}^T = \mathbf{U}^T\mathbf{U} = \mathbf{I} \tag{5.43}$$

$$\mathbf{VV}^T = \mathbf{V}^T\mathbf{V} = \mathbf{I} \tag{5.44}$$

The Principal Components of the data matrix **X** are given by the columns of the $(m \times n)$ matrix **P** :

$$\mathbf{P} = \mathbf{XV} = \mathbf{UL}^{1/2} \tag{5.45}$$

5.4.4 Correlation based on Principal Component calculations

The new transformed variables are used to reduce the dimensional of the original variables. Now we want to show how these new variables can be utilised to correlate between the input variables (x_i) and the quality variable of interest (x_k). Before we proceed, it is advisable to standardise the original matrix **X**, because the data from the process normally come with different units. Then, to determine the *correlation* between the x_i and x_k, we can utilise the equation below:

$$corr(\mathbf{x_i},\mathbf{x_k}) = \frac{cov(\mathbf{x_i},\mathbf{x_k})}{\sqrt{var\ \mathbf{x_i}}\sqrt{var\,\mathbf{x_k}}} \tag{5.46}$$

Since our matrix **X** is in standardised form, $var(\mathbf{x_i}) = 1$ and $var(\mathbf{x_k}) = 1$. The $corr(\mathbf{x_i},\mathbf{x_k})$ is therefore:

$$corr(\mathbf{x_i},\mathbf{x_k}) = cov(\mathbf{x_i},\mathbf{x_k}) = \mathbf{x_i^T x_k} \tag{5.47}$$

The values of $\mathbf{x_i}$ and $\mathbf{x_k}$ can be written in term of Principal Components by using equation (5.45). Multiplying both sides of equation (5.45) with $\mathbf{V^T}$, the relationship becomes:

$$\mathbf{X} = \mathbf{PV^T} \tag{5.48}$$

Let $\quad \mathbf{V^T} = \tilde{\mathbf{V}} = [\tilde{\mathbf{v}}_1 \quad \tilde{\mathbf{v}}_2 \quad \dots \quad \tilde{\mathbf{v}}_n] \qquad (5.49)$

where $\tilde{\mathbf{v}}_i^T = [v_{i1} \quad v_{i2} \quad \cdots \quad v_{in}] \qquad (5.50)$

but $\quad \tilde{\mathbf{v}}_i \neq \mathbf{v}_i^T \qquad (5.51)$

where $\tilde{\mathbf{v}}_i^T \tilde{\mathbf{v}}_i = 1$

$\tilde{\mathbf{v}}_i^T \tilde{\mathbf{v}}_j = 0$

Then since from equation (5.48) $\qquad \mathbf{X} = \mathbf{PV^T}$

then $\quad \mathbf{x_i} = \mathbf{P}\tilde{\mathbf{v}}_i \qquad (5.52)$

and $\quad \mathbf{x_k} = \mathbf{P}\tilde{\mathbf{v}}_k \qquad (5.53)$

Substitute back equations (5.52) and (5.53) into equation (5.47)

$$cov(\mathbf{x_i},\mathbf{x_k}) = (\mathbf{P}\tilde{\mathbf{v}}_i)^\mathbf{T}\,\mathbf{P}\tilde{\mathbf{v}}_k \tag{5.54}$$

and rearrange:

$$cov(\mathbf{x_i},\mathbf{x_k}) = \tilde{\mathbf{v}}_i^\mathbf{T}\mathbf{P^T P}\tilde{\mathbf{v}}_k \tag{5.55}$$

From equation (5.45), $\mathbf{P} = \mathbf{UL}^{1/2}$ thus:

$$\mathbf{P}^T\mathbf{P} = (\mathbf{UL}^{1/2})^T(\mathbf{UL}^{1/2}) \tag{5.56}$$

rearrange, where $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ then:

$$\mathbf{P}^T\mathbf{P} = (\mathbf{L}^{1/2})^T\mathbf{U}^T\mathbf{UL}^{1/2} = \mathbf{L} \tag{5.57}$$

Inserting equation (5.57) into equation (5.55)

$$cov(\mathbf{x}_i, \mathbf{x}_k) = \tilde{\mathbf{v}}_i^T \mathbf{L} \tilde{\mathbf{v}}_k \tag{5.58}$$

The correlation between $\mathbf{x}_i$ and $\mathbf{x}_k$ is possible by inserting the above equation back to equation (5.47), leading to:

$$corr(\mathbf{x}_i, \mathbf{x}_k) = \tilde{\mathbf{v}}_i^T \mathbf{L} \tilde{\mathbf{v}}_k \tag{5.59}$$

Thus the *correlation* between the original variables $\mathbf{x}_i$ and $\mathbf{x}_k$ based on the above equation can be written in form of the transformed variables from PCA as:

$$corr(\mathbf{x}_i, \mathbf{x}_k) = \sum_{j=1}^{n} v_{ij} v_{kj} l_j \tag{5.60}$$

If some of our Principal Components ($\mathbf{p}_j$) are discarded then the value of *correlation* in equation (5.60) will be based on reduced space. The detail about how we can reduce the dimension will be elaborated in chapter 7.3. The function of the above equation can therefore be used to determine the *correlation* between the quality variable and the input variables as will be discussed in the next chapter. It forms the basis for translating the control limits imposed on the quality variable to corresponding limits for all input variables.

5.5 Summary

This chapter has examined, elaborated and mathematically defined two types multivariate statistical methods, Partial Correlation Analysis (PCorrA) and Principal Component Analysis (PCA). The PCA method was discussed from both geometrical and the mathematical views. Apart from the ability to reduce the dimension of the original data via the use of new transformed variables, the PCA method can be used to calculate the relationship between the quality variable and the input variables. Equations (5.13) and (5.60) will be used extensively throughout the rest of this thesis. In the next chapter we will explore the use of PCorrA and PCA and how they contribute to the new proposed Active SPC method.

Chapter 6

The Active SPC Methodology

6.1 Introduction

This chapter discussed in detail the concept of Active SPC. First, we will examine the requirements of Active SPC methodologies. Second, we will elaborate how the multivariate data analysis techniques discussed previously can be used to design Active SPC strategies. In particular, we will show how the results from these multivariate data analyses can be used to define appropriate control laws. Here, the traditional function of SPC passive process monitoring is changed to provide an active role, *i.e.* adjusting the manipulated variables automatically to keep the process under statistical control.

We have seen in Chapter 4 that Automatic Process Control (APC) strategies provide continuous corrective actions whenever the controlled variable deviates away from set point by adjusting the manipulated variables. It does not matter whether the deviations are inherent to the process. In some cases, the effort is pointless and may even be detrimental to the process. Hence, it has been argued that APC does not improve the process since it doesn't eliminate the root cause of the problem. In contrast, traditional Statistical Process Control (SPC) charts only give indications on when action should be taken, namely when the quality variable exceeds some specified limits on the control chart. However, it is possible to devise a method whereby the process is monitored and automatically controlled whilst retaining the SPC policy of the non-intervention when the process is in a state of statistical control. In addition to product quality, input variables are also monitored. If these inputs can be kept within their respective control limits, then the quality variable should also be maintained within its control limits. Obviously, the former must be related to the latter limits. These quantifying relationships are provided directly from Partial Correlation Analysis (PCorrA) and *correlations* derived from Principal Component Analysis (PCA).

The role of PCorrA and PCA in this work is to describe the relationship between the quality variable with the input variables. Through this relationship we can work out which input variables has the greatest influence on the quality variable. Hence, the influential input variables can be chosen as our manipulated variables. This is similar to the practical problem of how the quality variables and the manipulated variables should be paired in a multiloop control scheme. Incorrect pairing will lead to poor control system performance. PCorrA and PCA will give recommendations concerning the most effective pairing between the quality variable and the manipulated variables. Both techniques also avoid the model identification problem by focusing attention on statistical input-output relationships. The results from the analyses can be utilised directly to provide guidelines on the required magnitudes of manipulative changes. The strategy can thus be applied not only for on-line process monitoring but also for the on-line control of the quality variable by making the appropriate adjustments on the manipulated variables.

6.2 Preliminary Procedure

One of the functions of a control chart is to monitor the process by periodically making observations of the production process. But before the charts can be used to identify the presence of assignable causes in the process, the process must be brought into a state of statistical control; that is, the data being used to determine the initial control limits must be subject only to a constant system of common causes. This requirement is essential so that the limits on the control chart will not be biased. Once the initial control limits are established, the charts may be employed to monitor the behaviour of the process.

Additionally, there are certain considerations concerning the way the data from the process should be collected. The first step in setting up control charts is the selection of the sample size. Since we are concerned with continuous processes, which involves variables measurement in continuous form, the sample size used for the SPC chart is one; that is, the sample consists of an individual observation. The second step is to determine the sampling time (Δt). The sampling time is chosen based on the autocorrelation behaviour of the signal and will discussed in detail in section 6.2.1 and Chapter 7. To provide a basis for the initiation of SPC charts, we should use 25 to 50 samples to estimate the process capability (Wetherill and Brown, 1991). Then, information about the process is assessed by calculating the process *mean* and the process *standard deviation.*

The next step is to examine whether the process data is normally distributed. The normally distributed data assumption will help us determine the regions of common causes and assignable causes. For an assignable cause to occur is highly unlikely to be due to chance, because only 0.27% of normally distributed data should fall in this category. This procedure is fundamental, since it can predict the probability of occurrence within a certain range of values. The process data is incapable if it is not normal. As a consequence, SPC chart procedure is not suitable for solving the problem of quality in the process.

6.2.1 Autocorrelation

Autocorrelation normally plays an important role in determining the properties of time series data. It gives a series of quantities called sample autocorrelation coefficients, which are measures of the *correlation* between observations at different instances in time. It is similar to the ordinary *correlation* analysis, but instead of correlating with different variables, the analysis is performed on an observation and successive observations of the same variables. Mathematically, the autocorrelation is represented by:

$$r_k = \frac{\sum_{t=1}^{N-k}(x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{N}(x_t - \bar{x})^2} \tag{6.1}$$

Where r_k is the auto-correlation coefficient at lag k, x_t is the current observation, $\bar{x}$ is the average of the process and x_{t+k} is the observation at $t+k$. The above equation shows how we can determine the autocorrelation between the current and observations at k samples apart. It is also possible to calculate the autocorrelation coefficient by using the autocovariance coefficient, c_k which is defined as:

$$c_k = \frac{1}{N}\sum_{t=1}^{N-k}(x_t - \bar{x})(x_{t+k} - \bar{x}) \tag{6.2}$$

The above equation represents the autocovariance coefficient at lag k. We then compute the autocorrelation coefficient by :

$$r_k = c_k / c_o \tag{6.3}$$

where c_0 is the variance of the time series. A useful aid in interpreting a set of autocorrelation coefficients is the correlogram, which is a graph of r_k plotted against the lag k. Nevertheless, interpreting the meaning of a set of autocorrelation coefficients is not always easy. The correlogram often initially exhibit a fairly large value of r_1 followed by few further coefficients which, while greater than zero, tend to get successively smaller. It will then oscillate about zero and finally, values of r_k for longer lags tend to be approximately zero.

The data is often not normally distributed when autocorrelation exists in the process. Autocorrelation is particularly common in the process industries because disturbances tend to have immediate, as well as lasting effects on the process. Autocorrelations is not necessarily bad. It does indicate that the process is somewhat predictable and suggests the possibility of compensation in the process. Nevertheless, it will violate the assumption associated with the use of SPC charts; that is the data generated by the process when in control are normally and independently distributed with *mean* (μ) and *standard deviation* (σ). During in-control situations, both μ and σ are considered fixed. An out-of-control condition occurs when there is a significant change in μ or σ or both. When autocorrelation exists in the process, SPC charts will therefore give misleading results in the form of false alarms. However, the autocorrelation behaviour can be removed from the data by selecting an appropriate sampling time (Δt) and this will be elaborated in Chapter 7.

6.2.2 Transformation and Standardisation

After the process data is successfully made free from autocorrelation behaviour, and the sampled data is normally distributed, we should then consider the possibility of non-linear relationships in the variables. Principal Component Analysis (PCA) and Partial Correlation Analysis (PCorrA) procedure are used only for investigating linear association among the variables. Non-linear affiliations between variables should be taken into account prior to application of these techniques. Therefore, any non-linear relationships between the variables have to be linearised before the analyses is performed. For example, in the CSTR, temperature relationships are non-linear due to Arrhenius reaction kinetics. Therefore the temperature of the reactants (*Tin*) and the temperature of reaction mass (*T*) were linearised by using the following transformations:

$$Trtin = \exp(-1/Tin) \quad \text{and} \quad Tr = \exp(-1/T) \tag{6.4}$$

After the non-linear variables have been linearised, we should then treat all the variables in the process so that they have the same units. This is accomplished by standardising the data. By doing so, the data will be dimensionless and in the form of *standard deviation* unit. This conversion actually allows all normal distributed data from each variable to be related in one form. The probabilities or areas under the curve then can be extracted from a standard table which can be found in any statistic text book. The data is standardised by subtracting its *mean* from each observation and dividing by its *standard deviation.*

6.3 PCorrA and PCA in Active SPC

Having completed all the necessary data conditioning, we can now apply PCorrA and PCA to design Active SPC methodologies. Initially, the partial *correlations* and *correlation* derived from PCA between inputs and outputs are calculated using historical process operating records. Similar to a process operability study or an APC loop input-output pairing exercise, the *correlations* are used to determine which variables are influential in keeping the process under control. Once the variables have been identified, the appropriate *correlations* are used to determine the relevant limits for the control charts as well as the corresponding control laws for maintaining output quality. The details of this procedure are described below.

6.3.1 Determination of Control Limits

Normally, the individual SPC charts monitor the products or quality variables by imposing a control limits at $\mu_0 \pm 3\sigma$. The process is considered to be in statistical control when it is kept within these limits. If it is out of statistical control, then the cause is located and rectified. Contrary to this conservative approach, however, we propose an alternative monitoring and manipulation strategy. Firstly, the PCA or PCorrA is used to determine the strength of association between the quality variable and the input variables. PCorrA and the *correlation* derived from PCA can be used to correlate x_k, the quality variable of interest and x_i, the inputs variables. When PCorrA and PCA have been calculated based on standardised variables, the control rule needed to moderate the variation of the quality variable x_k, is to maintain the following *correlation* between x_k and x_i:

$$x_k^s = C_{ik} x_i^s \qquad (6.5)$$

where $x_k^s = (x_k - \mu_k)/\sigma_k$, $x_i^s = (x_i - \mu_i)/\sigma_i$, and C_{ik} are the *correlation* coefficient between x_k^s and x_i^s using standardised variables. The values of C_{ik} are given by the relationship below:

$$C_{ik} = r_{ik.(q+1,...,n)} \text{ or } C_{ik} = \sum_{j=1}^{n} v_{ij} v_{kj} l_j \tag{6.6}$$

where $r_{ik.(q+1,...,n)}$ is the partial *correlation* coefficient between x_i and x_k (equation 5.13) and $\sum_{j=1}^{n} v_{ij} v_{kj} l_j$ is the *correlation* based on PCA (equation 5.60). Given the *correlation* C_{ik}, between the inputs and the quality variable, the limits on the quality variable can be translated to corresponding limits on input variables. When $\mu_o \pm 3\sigma$ output control limits are used, then since x_k^s is standardised, the limits becomes:

$$-3 < x_k^s < 3 \tag{6.7}$$

The corresponding limits on the input variable are:

$$-3 / C_{ik} < x_i^s < 3 / C_{ik} \tag{6.8}$$

Figures 6.1 shows how the Shewhart Chart control limits lines for the quality variable are translated into limits for the input variables.

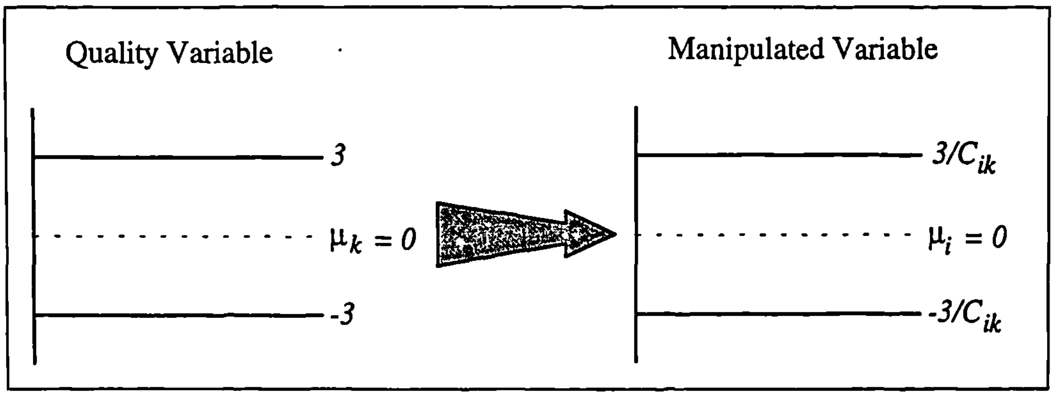

Figure 6.1 Translated control limits for standardised variables.

If the relationships between input and output variables do not change, then by maintaining x_i between the limits given by equation (6.8), the output should also be kept within the respective desired control limits, equation (6.7). If any manipulated variables fall outside these control limits, control action is invoked on that particular manipulated variables by bringing it back to the appropriate control limit.

The limits for Shewhart Charts with both action and warning lines, and EWMA charts can be developed in similar manner. Deviations in input variables from their respective limits indicate that the quality variable will suffer an out-of-control situation. If the monitored input variable is also the manipulated variable, adjustment will be made if an out-of-control situation occurs. As a consequence, the quality variable should not affected and will be on target. The relationship from the quality variable x_k and input variables x_i is given by C_{ik} from equation (6.5). Figure 6.2 shows how the Active SPC scheme can be implemented on the EWMA control chart.

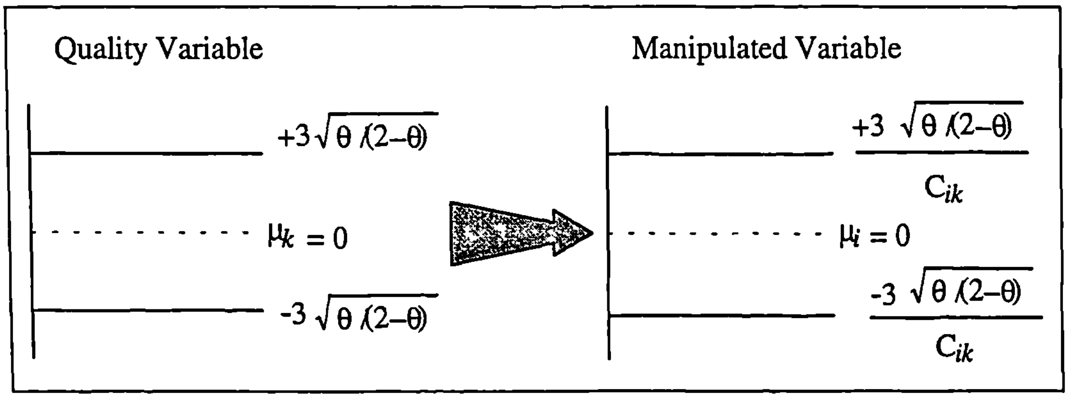

Figure 6.2 Translated control limits for EWMA chart for standardised variable

6.3.2 Measured Disturbances

In some situations, not all input variables are measured on-line, either because of economic constraints or because suitable instrumentation does not exist. In other cases, some input variables cannot be manipulated, *e.g.* the quality of purchased raw materials. The proposed technique assumes that all the input variables can be measured on-line, but does not require that all inputs can be manipulated. If this is the case, an alternative manipulative input variable, x_m, which can affect the quality variable must be determined. When the monitored variable x_i exceeds specified limits, x_m must then be manipulated to compensate for the deviations in x_i so that the quality variable x_k remains within its target limits. This alternative control law can be expressed as follows. The relationship between x_k and x_m is:

$$x_k^s = C_{mk} x_m^s \tag{6.9}$$

The predicted value of x_k^s from the monitored value of x_i^s is given by equation (6.5). Thus, the difference between this value (x_i^s) and a target limit l_i^s will be:

$$\Delta x_{ik}^s = C_{ik}\left(x_i^s - \ell_i^s\right) \qquad (6.10)$$

Δx_{ik}^s can be expressed in term of Δx_{im}^s by combining equations. (6.9) and (6.10):

$$\Delta x_{im}^s = \frac{\Delta x_{ik}^s}{C_{mk}} = \frac{\left(x_i^s - \ell_i^s\right)C_{ik}}{C_{mk}} \qquad (6.11)$$

The amount of x_m has to be adjusted to compensate for the predicted deviations of measured disturbances variables, will be:

$$x_m^{s,new} = x_m^s - \sum_{i=1}^{ni} \Delta x_{im}^s \qquad i \neq m \qquad (6.12)$$

where ni is the number of measured disturbances variables whose deviations that x_m has to be adjusted for. When two manipulated variables are selected, the reactant temperature (*Trtin*) will be used to compensate for out-of-control situations in the reactants molar concentrations *Ain*, and *Bin*; and the flowrate of the reactants, *F*. While input coolant temperature (*Tjin*) will compensate for out-of-control variations in cooling medium flowrate *Fj*. For manipulating a single variable, the out-of-control variations in all measured disturbances variables will be compensated by adjusting *Trtin*. However, the new value of x_m resulting from equation (6.12) is also compared against its control limits. If these limits are exceeded, then x_m is brought back to the appropriate control limit. The detail of how the manipulated variables are selected are given in section 7.3.

6.3.3 On-line Calculation of Control Limits

There are two ways to implement Active SPC. One is to calculate the relationship between the quality variable and the input variables off-line using good operating records of data. Alternatively, the *correlations* may be determined on-line. The off-line analysis is applied first to select the manipulated variables for controlling the quality variable before the on-line technique is implemented. It will also provide excellent initial control limits for the SPC charts and hence the degree of correction needed on the manipulated variables.

Calculating the *correlations* continuously as the technique is being implemented significantly increases the computational requirements. However, an advantage is that changing relationships can be tracked. If the values of C_{ik} are calculated on-line, then they should be smoothed to reduce the possibility of abnormal changes due to the out-of-control situations using the following expression:

$$\tilde{C}_{ik}(t) = (1-\lambda)\tilde{C}_{ik}(t-1) + \lambda C_{ik}(t) \qquad (6.13)$$

$\tilde{C}_{ik}(t)$ is the exponentially smoothed value of $C_{ik}(t)$ at time t while λ is the smoothing constant. $\tilde{C}_{ik}(t)$ is then used in place of C_{ik} in the equations above.

The above development concerns limits for Shewhart charts with action lines and EWMA charts. Limits for other SPC charts can be developed in a similar manner. Inputs which show relatively high *correlation* coefficients with the quality variable are monitored and subsequently be chosen as the manipulated variables. If Shewhart charts are used, the limits are given by equation (6.8). They also define the control laws, *i.e.* the manipulations that have to be made to the relevant manipulated variables to keep the quality variable under statistical control. Manipulations are executed by bringing deviant back to their allowable limits. Since the most likely causes of process upsets have been pre-determined via PCorrA and *correlation* based on PCA, undesirable deviations due to disturbances can therefore be automatically contained and out-of-control situations of the quality variable will be minimised. The strategy therefore anticipate and compensate against potential process upsets. Compared to traditional practice, it is therefore an *active* approach to SPC.

In monitoring the important variables, the charting techniques employed the familiar Shewhart and EWMA control chart. There is no need to resort to multivariate charts because the inputs and outputs of the process have effectively been decomposed to smaller, independent sub-system. Since the probable causes of the process deviations have been predetermined, on-line SPC reduces the need for expensive and time consuming experimentation after the incidence of out-of-control situations. Moreover, abnormal variations due to disturbances can be corrected before they affect output quality. Control is therefore achieved in anticipatory manner.

6.4 New Development in this Work

This work follows the preliminary studies of Efthimiadu *et al.* (1991), Efthimiadu *et al.* (1992), and Efthimiadu *et al.* (1993) who attempted to devise a new, pro-active approach to achieve SPC of continuous process. Detailed evaluations and some modifications were performed in the current work to assess the applicability of the various Active SPC strategies that might evolve.

Here, we would like to highlight some of the modifications involved in the current work compared to the previous work of Efthimiadu *et al.* (1991), (1992) and (1993). The control limits that they used to determine the out-of-control situations is based on the unstandardised variables. Thus, instead of the limits that we used in equations (6.7) and (6.8), they utilised the equations below:

$$\mu_k - 3\sigma_k < x_k < \mu_k + 3\sigma_k \tag{6.14}$$

for the quality variable and the translated version for the inputs control limits will be:

$$\mu_i - \frac{3\sigma_i}{C_{ik}} < x_i < \mu_i + \frac{3\sigma_i}{C_{ik}} \tag{6.15}$$

The ambiguous aspect is that the values of C_{ik} that they used are based on standardised variables. If C_{ik} are calculated using PCorrA this will not affect the control strategy, but when we use PCA which is scale dependent, the results will be dubious.

The magnitude of control actions are also different between the current work and the previous works. If any manipulated variables falls outside the control limits, control action is invoked on that particular manipulated variable by bringing it back to the appropriate control limit. The previous work adjusts it back to the *mean*. The main reason for this new approach is to avoid over control of the process. Moreover, the inherent variations in the process is still within the control limits.

Efthimiadu *et al.* (1992) only considered manipulating all variables and two variables. For the two manipulated variables case, they utilised *Trtin* and *Fj* as the manipulated variables, which were not based on whether the manipulated variables have high influence on the quality variable. In the current work, we utilise *Trtin* and *Tjin* which have high influence on the quality variable. We also investigated the use of only one manipulated variable to control the process.

To quantify the effectiveness of each strategy, we also defined an "Index of Performance" (*IP*) (Ibrahim and Tham, 1995). It is a measure which penalises the out-of-control situations as well as false alarms. By doing so the Active SPC scheme is judged, not only by how well it can prevent the out-of-control situations, but also its propensity to limit the number of unnecessary control actions. That is, the *IP* provides a measure of overall control performance.

6.5 The Active SPC Procedure

The proposed technique can be viewed as a form of static feedforward control with dead zones. Instead of deterministic steady state models used in the design of static feedforward compensation strategies, the proposed scheme makes use of statistical relationships. Nevertheless, it is suggested that the proposed technique offers better flexibility. Apart from being able to compensate for deviations in inputs that cannot be manipulated, the control framework can also be easily extended to provide the equivalent of feedback-feedforward control (Ibrahim and Tham, 1995).

The implementation of the Active SPC strategy can be divided into five stages: observation, evaluation, diagnosis, decision, and lastly, the application stage. The observation stage considers which variables are to be collected, how many sample measurements to take and how often should they be taken, *i.e.* sampling time. The sampling time is chosen so that the samples are not autocorrelated. Next, the necessary conditioning on the data is performed, *i.e.* linearise the non linear variables and transform the variables to standardised variables.

The evaluation stage involves analysis of the observed data. After the *means* and *standard deviations* have been determined, the data is checked to assess whether the normality assumption holds. Based on this, control limits are determined, then the control charts are plotted to implement control on-line.

The diagnosis stage is perhaps the most crucial part in this work. Here we applied two types of multivariate statistical analysis, the Principal Component Analysis (PCA) and Partial Correlation Analysis (PCorrA) to identify the *correlations* that might exist between the quality variables and the input variables. If *correlations* exist between the quality variable x_k and the input variable x_i, the indications is that the values of x_i will affect the value of x_k. These input variables will be selected as the manipulated variables.

Once the appropriate manipulated variables have been decided, it is necessary to specify the tool for making the correction in the implementation stage. Here, we utilised traditional control charts namely the Shewhart chart with action lines (ShewA), Shewhart chart with both action and warning lines (ShewAW), and the EWMA chart to rectify the process if it departs from intended operating conditions.

Apart from being used in the diagnosing stage, both PCA and PCorrA are also extensively used in the decision stage. The correlations derived from both methods are used to modify the control limits of the traditional SPC charts. They are used to formulate the control law and the magnitude of control actions for adjusting the process when assignable causes occur. This is made possible by translating the control limits on the quality variable x_k to limits on the input variable x_i through the *correlation* procedure. The input variables selected for monitoring and manipulation are based on how they influence the quality characteristics. By calculating the *correlations* on-line, changing relationships between the quality variable x_k and the input variables x_i can be tracked.

6.6 Summary

The procedural steps to follow for applying Active SPC are summarised below:

1. Choose a sampling time (Δt) such that there is no autocorrelation behaviour in the sampled data.
2. Collect the data during a period when the process is perceived to be in statistical control.
3. Calculate the *mean* and the *standard deviation* for all variables in the process.
4. Check whether the data fulfils the normal distributed data assumption. If it fulfils this criterion proceed to step 5, if not go back to step 1.
5. Transform all non-linear variables in the process.
6. Standardise the data so that it will be dimensionless and in *standard deviation* units.
7. Calculate the *correlation* coefficient based on either Partial Correlation Analysis (equation 5.13) or Principal Component Analysis (equation 5.60).
8. Based on this result, select the most influential manipulated variables for controlling the quality variables.
9. Use the *correlation* between the quality variable and the input variables to determine the control limits and the control laws for the manipulated variables.

10. Translate the given control limits and control laws to an existing control chart *e.g.* Shewhart Charts or EWMA chart.
11. If the input variables cannot be manipulated then use equation (6.11) and equation (6.12), to calculate a relationship based on an alternative manipulated variable.
12. For on-line *correlations* updating, use equation (6.13) to smooth the value of C_{ik}.

In the next chapter we will evaluate the performance of this Active SPC strategy by application to a simulated CSTR where a reversible exothermic reaction takes place.

Chapter 7

Evaluation of Active SPC by application to a CSTR

7.1 Introduction

This chapter applies the proposed Active SPC methodology to the non-linear simulation of a continuous stirred tank reactor (CSTR) described previously. The input variables for the simulated process are: the concentration of *Ain*; concentration of *Bin*; the temperature of reactants, *Tin*; temperature of the cooling medium, *Tjin*; the flowrate of reactants, *F*; and the flowrate of the cooling medium, *Fj*. Meanwhile the output variables are: the temperature of the reaction mass, *T*; cooling medium output temperature, *Tjout*; concentration of *Aout* and *Bout*; and finally *Cout*, the product concentration which was chosen as the quality variable and has to be kept under statistical control.

Several Active SPC schemes were tested on the process to evaluate the effectiveness of the respective configurations. Four situations were considered: (i) using on-line and off-line calculation of control rules; (ii) applying different types of multivariate techniques to calculate the correlations between the quality variable and the input variables, *i.e.* using either PCA or PCorrA; (iii) using different manipulating strategies to control the process, by manipulating all input variables, by using two manipulated variables and lastly by utilising a single manipulated variable; (iv) utilising different types of SPC charts, particularly, Shewhart charts with action limits (ShewA), Shewhart chart with both action and warning limits (ShewAW), and Exponential Weight Moving Average (EWMA) charts. We also discuss the importance of historical data on the performance of Active SPC techniques. Lastly, we compare the performance of these Active SPC methods with the APC strategy, specifically *Proportional Integral* (*PI*) feedback control.

7.2 Prerequisites Before Plotting Control Chart

Before process data is collected, there are several requirements that have to be fulfilled as mentioned in chapter 6. The sampling interval (Δt) for collecting, monitoring and controlling the process must be determined. This has to fulfil the criterion that there must be no autocorrelation in the sampled data. Then, we collect the data using this sampling time when the process is perceived to be in statistical control. The *means* and the *standard deviations* were then calculated from the historical data. Later, the data is checked for normality, so that the probabilities of data values falling within certain ranges can be predicted. Non-linear behaviour in the data were also linearised before PCorrA or PCA is applied to determine the relationship between the quality variable and the input variables. This requirement is essential since both of these multivariate methods are intended for determining linear association in the data. These requirements have to be fulfilled before Active SPC can be implemented on the process.

7.2.1 Determining the Sampling Interval Δt

As mentioned previously, the collection of measurements from the process is based on individual observations. The task now, is to determine the sampling interval, Δt for collecting the data. In the APC scheme of chapter 4, the sampling interval, Δt was chosen to be one tenth of the process time constant (τ_p) which was determined by using the *Process Reaction Curve* method. For SPC schemes, the sampling interval Δt is chosen such that there is no autocorrelation in the data. To do this, 50 sample measurements (m), were collected from the process with a sample interval of 10 seconds. Wetherill and Brown (1991) mentioned that any autocorrelation coefficient falling outside the range of $\pm 2 / \sqrt{m}$ can be regarded as significant indicating that the data is autocorrelated. Generally, we expect higher correlations for initial values of lag k. The autocorrelation then starts to subside, finally oscillating near zero for large lags. Figure 7.1 shows the autocorrelation for all input variables while Figure 7.2 shows the autocorrelation for *Cout*, the quality variable. From Figure 7.2, the autocorrelation line crosses the $\pm 2 / \sqrt{m}$ at lag $k \approx 10$. Since, the sampling interval for collecting the measurements was 10 seconds, the appropriate sampling time Δt for variable *Cout* was therefore determined to be 100 seconds. The same principle can be used to determine the sampling interval for other variables. Since the autocorrelation line cross the $\pm 2 / \sqrt{m}$ line at lag $k \approx 1$ for all input variables in figure 7.1, the sampling interval Δt for these input variables were chosen to be 10 seconds. Thus, the sampling time (Δt) for the quality variable will be 100 seconds, while the input variables will be sampled, monitored and controlled if necessary every 10 seconds.

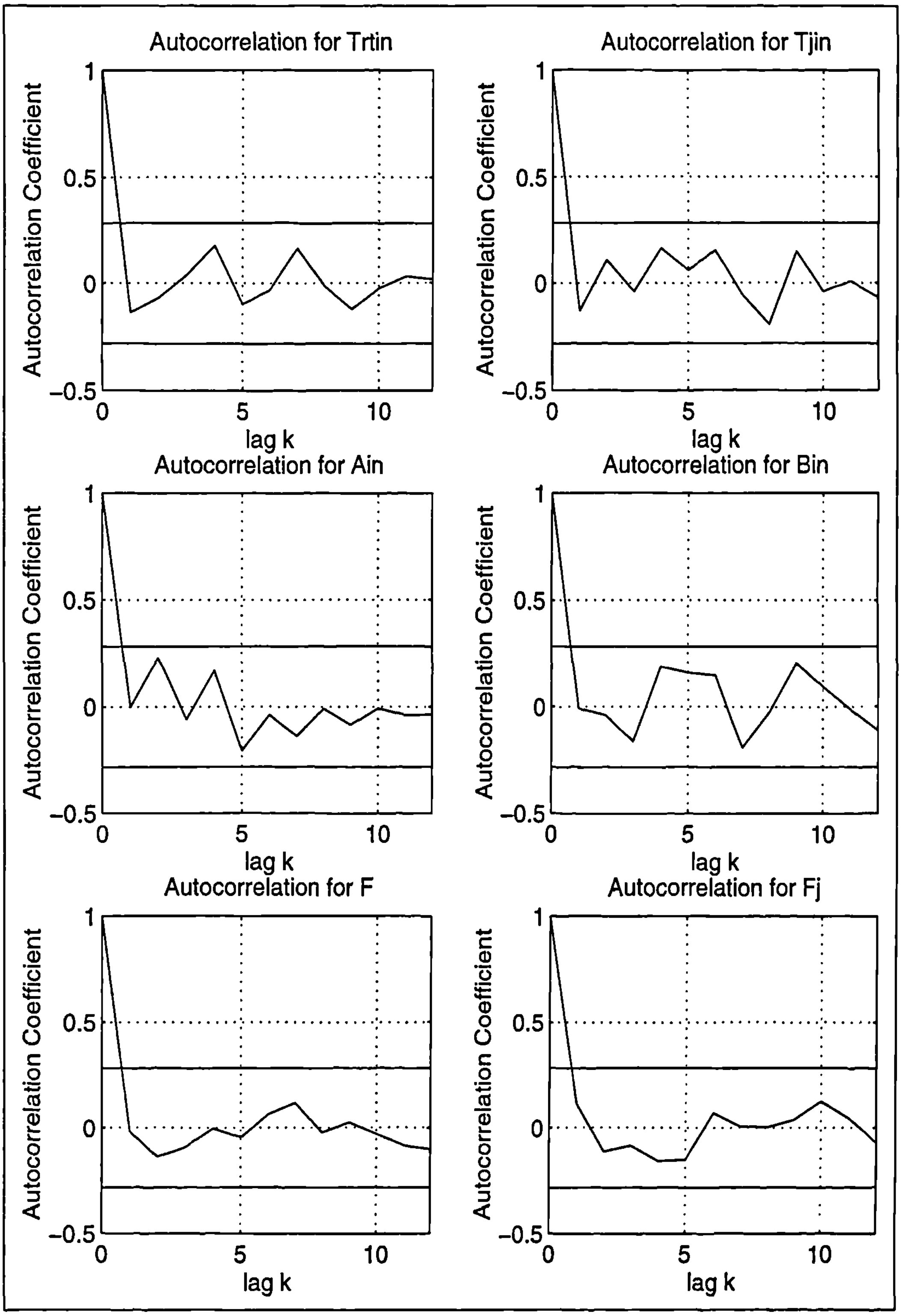

Figure 7.1 Autocorrelation functions for all manipulated variables

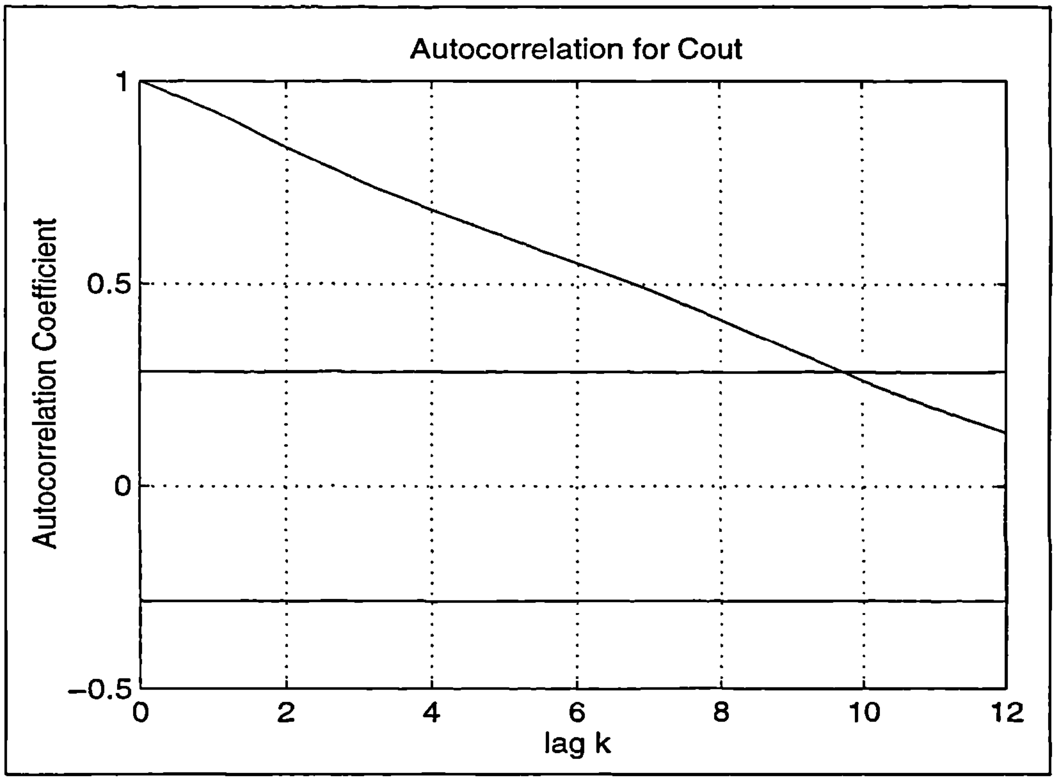

Figure 7.2 Autocorrelation function for *Cout*

7.2.2 Calculation of the Initial Control Limits

After the sampling time, Δt, has been determined, 100 data records were collected when the process was perceived to be in statistical control. The *means* and *standard deviations* of all input variables, output variables and quality variable were calculated and the results are given in table 7.1.

Table 7.1 Means and standard deviations based on historical data

Variables	Mean (μ)	Standard Deviation (σ)
Trtin	0.99778 K	1.5802e-05 K
Tjin	288 K	1.0716 K
Ain	500 mole/litre	8.7009 mole/litre
Bin	500 mole/litre	2.9797 mole/litre
F	0.08 litre/sec	0.0022 litre/sec
Fj	0.04 litre/sec	0.0007 litre/sec
Cout	205 mole/litre	10.0217 mole/litre
Tr	0.99738 K	2.4874e-05 K
Tjout	359.79 K	2.9313 K
Aout	294.90 mole/litre	11.9016 mole/litre
Bout	294.53 mole/litre	9.6629 mole/litre

The *means* and the *standard deviations* for the input and the quality variables were used for the design of the SPC charts. The *mean* (μ) was the target while the *standard deviation*, (σ) was used to determine the control limits on the SPC charts.

7.2.3 Normality Distributed in the Data

As discussed in the previous chapter, the data from the process must be normally distributed before SPC charts can be plotted. Normally distributed properties play an important role in SPC control chart because the probability of data falling within certain ranges can then be predicted. We can then determine the range in which the process is considered to be in statistical control.

To test whether our data from the CSTR process satisfy this assumption, normal probability plots are constructed for all input variables and the quality variable. These normal probability plots have cumulative probabilities scale on the vertical axis and values of the variable on the horizontal axis. Then a line of normal probability is drawn at the centre of the plot passing through the *mean* (μ) of the variable and the 50% cumulative probability. If most of the data points lie on this line, the data may be considered to be normally distributed. Conversely, if the points appear to be an S-shape, then the indication is that the data are not normally distributed. Sometimes, the line of normal probability describes the majority of the data but does not characterise a few extreme points. These extreme points are outliers that may have arisen from a different distribution. If a part or all of the data naturally cluster about a straight line and does not go through the intersection of the *mean* and 50% of cumulative probability, it indicates that there is an error in the estimated *mean*. We can also obtain a graphical estimate of the variable's *standard deviation* (σ) by finding the associated value measured along the horizontal axis corresponding to $P_{15.87\%}$ on the vertical axis. (DeVor *et al.,* 1992)

Figure 7.3 shows the normal probability plots for all input variables while figure 7.4 shows the normal probability plot for the quality variable, *Cout.* The majority of the data points of all variables are close to the line except for a few at either end of the line. Moreover, all the normal distributed lines pass through the intersection of the *mean* tabulated in table 7.1 and its 50% cumulative probability. Thus we can conclude that all variables approximately follow the normal distribution, and that the respective sampling intervals have been appropriately chosen.

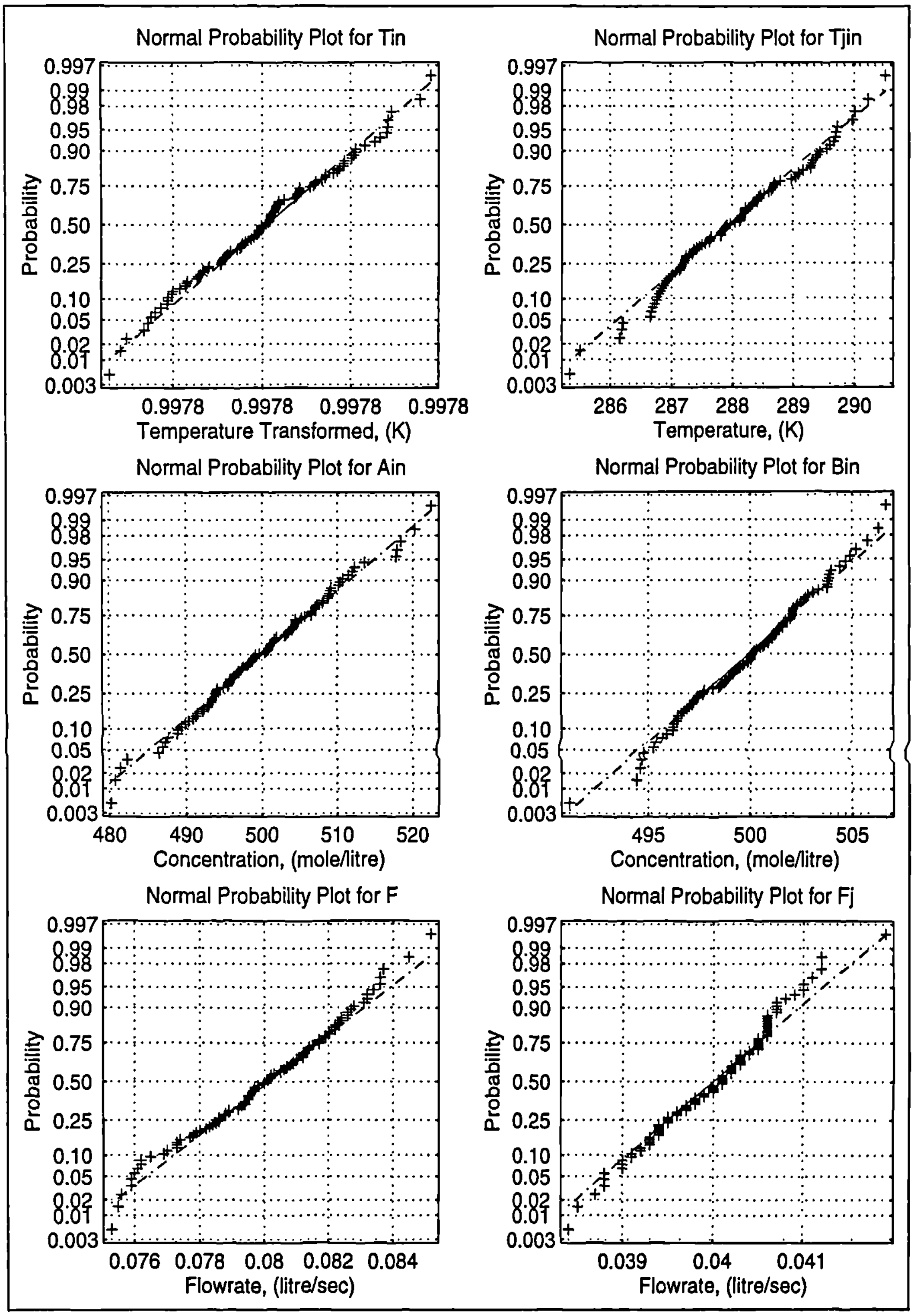

Figure 7.3 Normal probability plots for all input variables.

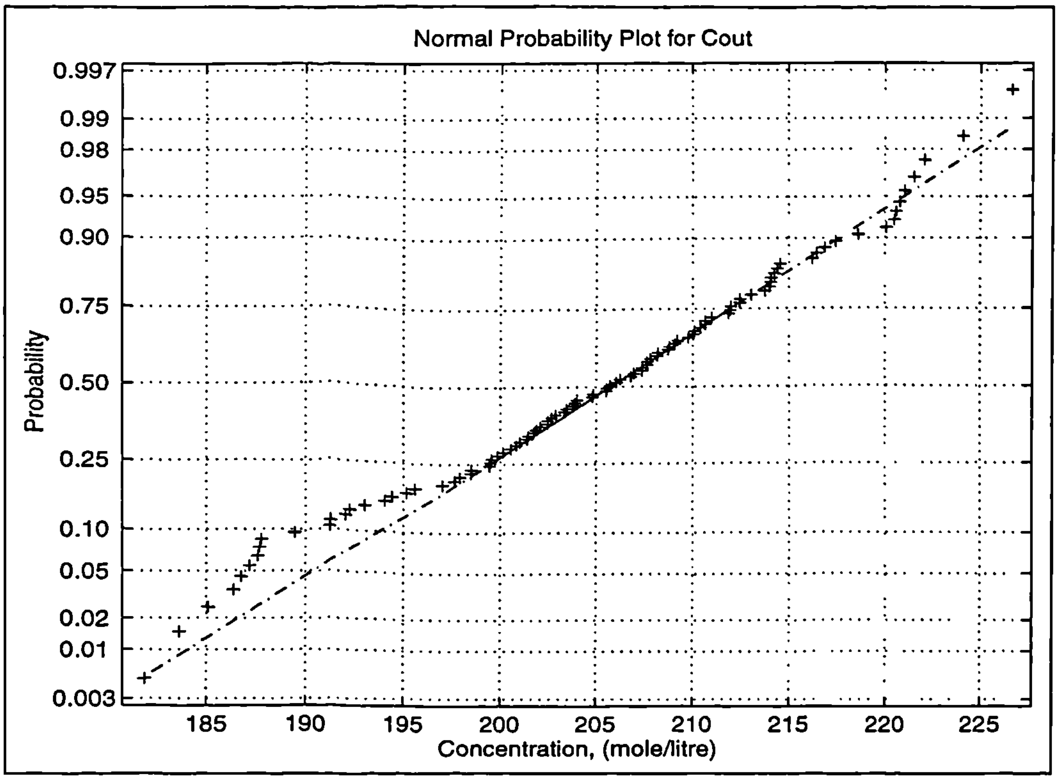

Figure 7.4 Normal probability plot for *Cout.*

7.2.4 Effect of Transformation on the CSTR Temperature.

Figure 7.5 shows *the variation of quality product concentration (Cout) due to variation* in CSTR temperature (*T*). From the figure, we see the non-linear behaviour of product concentration with temperature. As mentioned previously this non-linear behaviour is due to the Arrhenius reaction kinetics. To enable analysis using PCA or PCorrA, we have to linearise the effects of CSTR temperature (*T*). This was achieved by applying exp(-1/*T*) (e.g. equation 6.4). Figure 7.6 shows the relationship between this transformed temperature on *Cout.* Here, we see that we have a linear function between the transformed CSTR temperature (*Tr*) and *Cout.*

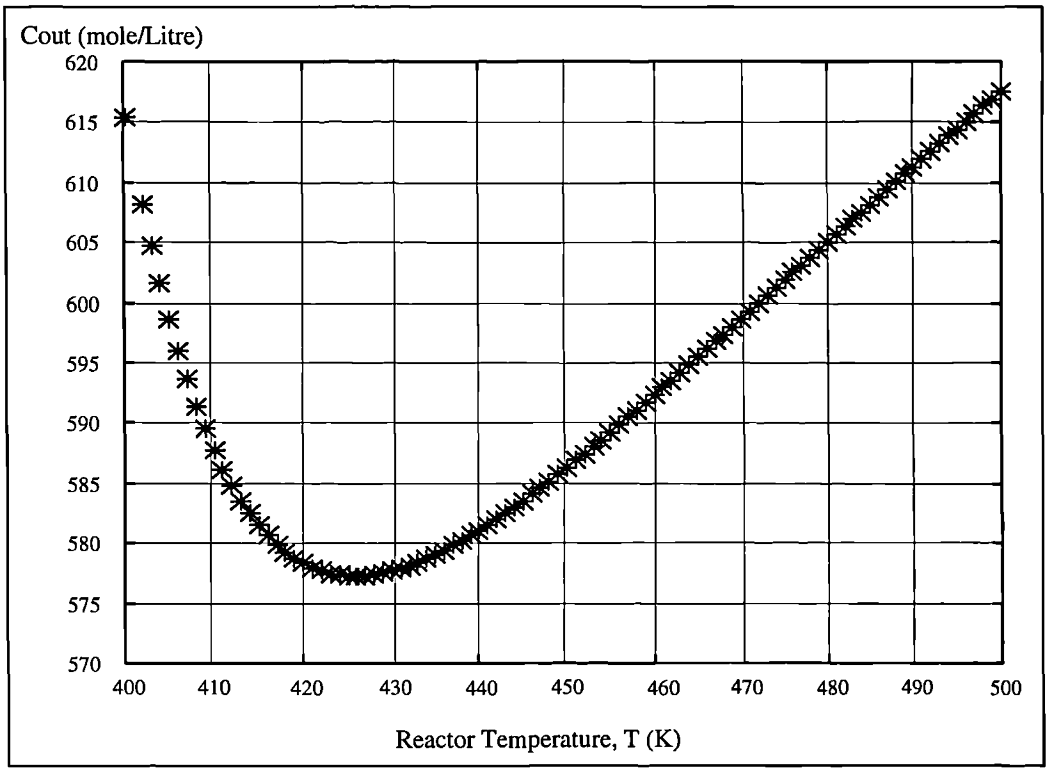

Figure 7.5 Effect of variation in CSTR temperature (*T*) on the product quality concentration (*Cout*)

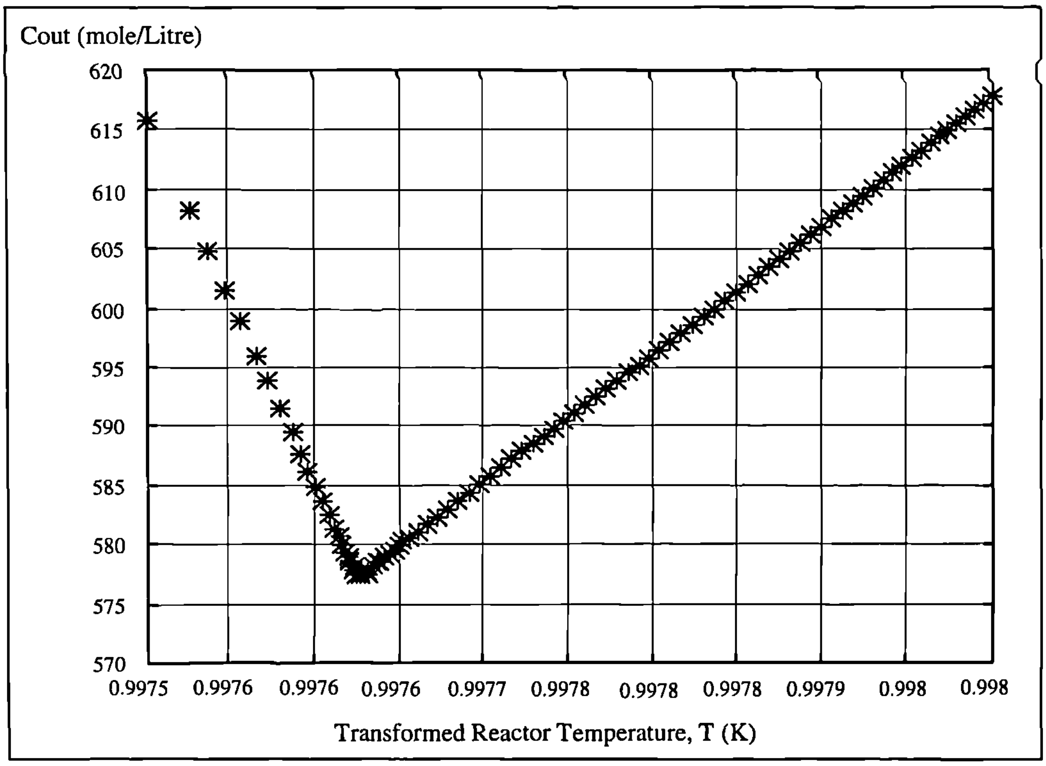

Figure 7.6 Effect of variation of transformed CSTR temperature (*Tr*) on the product quality concentration (*Cout*)

7.3 Calculation of PCorrA and PCA from the preliminary data.

The 100 data records collected were then used in the Principal Component Analysis (PCA) and the Partial Correlations Analysis (PCorrA). The resulting *correlations* (C_{ik}) are used to quantify the relationships between the quality variable and the manipulated variables. Before we proceed with the calculations, we know that the temperature of the reaction mixture is governed by the non-linear Arrhenius equation. Since both multivariate methods are only applicable for linear systems, the values of the input temperature (*Tin*) and the reactor temperature, (*T*) were linearised using the exponential transformation of equation (6.4). Then all input and output variables were standardised before PCA and PCorrA analyses were performed.

The results from the PCA and PCorrA calculations are tabulated in Tables 7.2 and 7.3. Table 7.2 shows the percentage of contribution of each Principal Component to the total variation in the data. They are arranged according to the corresponding eigenvalues in descending order. The results reveal that the first seven principal components account for more than 98% of the total variation in the data. It verifies that most of the variations in the process is due to the six input variables and the quality variable which describe the variation in the outputs. The four remaining principal components do not offer any significant explanation about the data variation and hence can be neglected from further analysis. Thus the reduction of dimension in the data space *n* in equation (5.60) for this case will be to 7 instead of 11.

Table 7.2 Percentage variation due to each Principal Components

p_j	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9-p_{11}
l_j^2	5.279	1.403	1.108	0.952	0.919	0.622	0.495	0.193	0.024
% var	47.99	12.76	10.08	8.661	8.357	5.662	4.502	1.761	0.227
Σ %	47.99	60.75	70.83	79.49	87.84	93.51	98.01	99.77	100

Table 7.3 shows the result of applying equations (5.13) and (5.60) to calculate the C_{ik} values. The results show that *Fj* has a negative *correlation* with *Cout*. The negative *correlation* between *Cout* and *Fj* is significant because it matches the analysis based on the *Process Reaction Curve*. From figure (4.8), we can see that increasing the flowrate of the cooling medium decreases the concentration of *Cout*. The results also show that *Cout* have positive *correlations* with *Trtin*, *Tjin*, *Ain*, *Bin* and *F*. From the same table

we also found that *Trtin* is the most influential input variable. This is shown by its highest C_{ik} value.

Table 7.3 C_{ik} based on PCA and PCorrA

	var	*Trtin*	*Tjin*	*Ain*	*Bin*	*F*	*Fj*
PCA	*Cout*	0.8075	0.6487	0.1066	0.2214	0.1665	-0.0136
PCorrA	*Cout*	0.7837	0.4525	0.3072	0.2848	0.2292	-0.1922

These *correlations* between the input variables and the quality variables are used to select the manipulated variables to control the process. Three manipulating strategies will be implemented on the system. Firstly, we are going to use all the inputs as manipulated variables. For cases where some of the input variables cannot be manipulated, we are going to implement two kinds of manipulating schemes, *e.g.* either using two or a single manipulated variable to control the process. For the two manipulated variables option, appropriate adjustments of *Trtin* could compensate for variations in *Ain, Bin* and *F*, while appropriate adjustment in *Tjin* could compensate for variations in *Fj*, thus maintaining control of *Cout*. These two variables were selected because of their high influence towards the quality variable compared to the other input variables. Since *Trtin* is the most influential variable compared to the others it will be a good candidate as the manipulated variable for a single variable manipulation scheme.

Comparison of the coefficients in table 7.3 shows that although the signs of *correlations* of the inputs with the quality output are the same in both techniques, the magnitudes of PCA *correlations* are higher in general for *Trtin*, and *Tjin* in particular. The anticipated consequence is that the PCorrA based charts will be less sensitive, compared to PCA based chart. This is because the limits of PCorrA based SPC charts will be wider than those of PCA based charts.

7.4 Performance Evaluation

To evaluate the effectiveness of the various Active SPC methodologies that are being considered, white noise was added to the inputs to simulate out of statistical control situations. Table 7.4 shows the standard deviations of the white noise used for the input variables. Active SPC charts with limits determined as described in the previous chapter were then applied to monitor those input variables indicated by the PCA or the PCorrA as being significantly correlated with the product concentration. When out-of-control cases were detected, adjustments were made on the chosen manipulated variables. The control actions on the manipulated variable were to simply bring them back within the particular inputs' limits.

Table 7.4 Standard deviations for white noise

Input variable	Standard Deviation
Trtin	0.8487 Kelvin
Tjin	6.0 Kelvin
Ain	18 mole/litre
Bin	30 mole/litre
F	0.05 litre/second
Fj	0.02 litre/second

7.4.1 Cases Considered

Initially, it was assumed that all input variables affecting the quality variable *Cout* could be measured and manipulated on-line. However, there are situations where some of the input variables are not measured was also considered. For example, the concentration of *Ain* and *Bin* may need to be measured using gas chromatography (GC) which is very expensive. In other situations, sometimes the input variable cannot be manipulated on-line. In this work we assume that all the input variables can be measured on-line, but some variables cannot be manipulated. The manipulation of all inputs, the manipulation of only some inputs, and the manipulation of a single input were then considered. In the strategy involving only some of the inputs, it was assumed that the input concentrations of the reactants, the flowrate of reactants and the flowrate of the coolant cannot be adjusted on-line leaving only *Trtin* and *Tjin* as the manipulated variables. The last scheme assumes that only *Trtin* can be manipulated.

This study also considered the use of Shewhart charts with action limits (ShewA); Shewhart charts with both action and warning limits (ShewAW), and lastly Exponential Weight Moving Average (EWMA) charts. The ShewA chart is a time plot of data with action limits at $\pm 3\sigma$ centred about the *mean* and the process is considered to be out-of-control when a single observation falls outside these limits. The ShewAW chart will be more sensitive to variation, because this chart is equipped with two control rules. The first rule is adopted from the ShewA control rule while the second one considers the process to be out-of-control when two out of three consecutive observations lie beyond the *mean* $\pm$ 2σ warning lines. On the hand, using the EWMA chart, if an observation falls outside the *mean* $\pm\ 3\sigma\sqrt{\frac{\theta}{(2-\theta)}}$ limits, the process is considered to be out-of-control, where θ is the weight assigned to the current sample observation. Here, θ was chosen to be 0.3.

The use of fixed as well as on-line calculated PCA or PCorrA coefficients were also investigated. The coded manipulation strategies for the above cases are described below and have the following meanings:

1st digit:	1) Off-line correlations calculation
	2) On-line correlations calculation
2nd digit:	1) Utilising Partial Correlation Analysis (PCorrA)
	2) Utilising Principal Component Analysis (PCA)
3rd digit:	1) Shewhart chart with action limits (ShewA)
	2) Shewhart chart with action and warning limits (ShewAW)
	3) EWMA control Chart (EWMA)
4th digit:	1) Manipulate all input variables
	2) Manipulate *Trtin* and *Tjin* only
	3) Manipulate *Trtin* only

Thus, the code 2112 denotes results obtained using Shewhart charts with action limits (ShewA), where the limits are based on Partial Correlations Analysis (PCorrA) calculated on-line using *Trtin* and *Tjin* as manipulated variables.

7.4.2 Block Diagrams of Active SPC schemes

Figures 7.7 to 7.12 show the block diagrams for several configurations of Active SPC schemes. Figure 7.7 and 7.8 show the Active SPC schemes for on-line calculated control limits for PCA and PCorrA, utilising all input variables as manipulated variables. Figure 7.9 and 7.10 show the corresponding control schemes when only two manipulated variables are used. Finally figures 7.11 and 7.12 show the Active SPC scheme utilising only one manipulated variable to control the quality variable. When fixed control limits are used, the PCorrA block and PCA block with the block referring to equation (5.60) are calculated off-line. Additionally equation (6.13) that is used to smooth the C_{ik} values is no longer necessary.

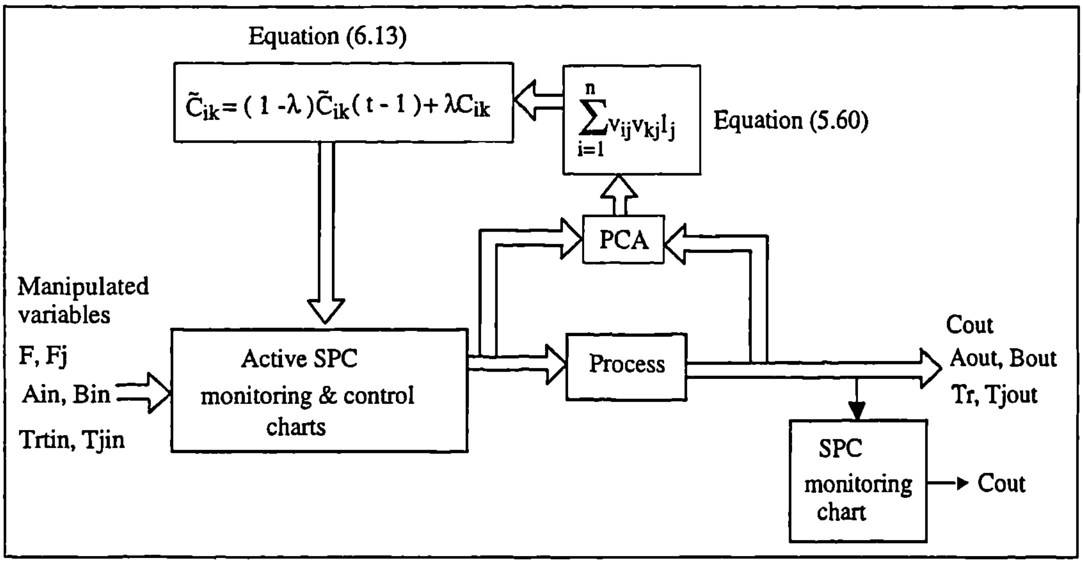

Figure 7.7 Active SPC scheme for on-line PCA calculated control limits using six manipulated variables

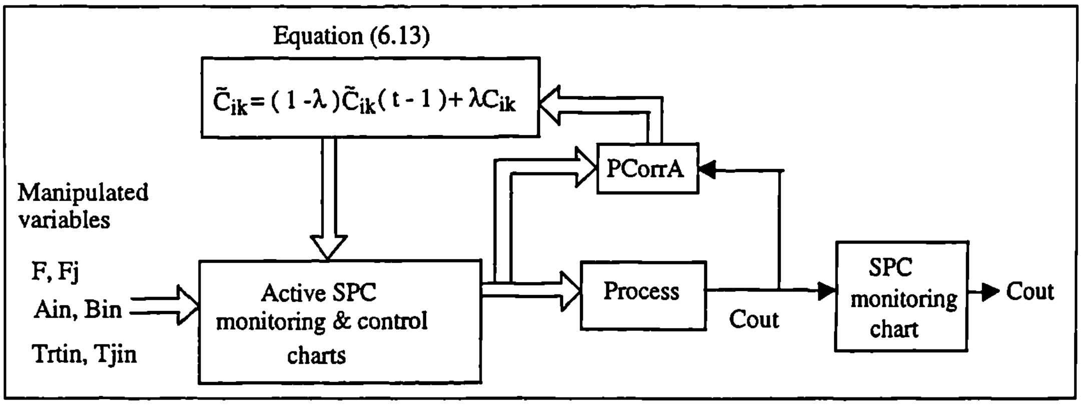

Figure 7.8 Active SPC schemes for on-line PCorrA calculated control limits using six manipulated variables

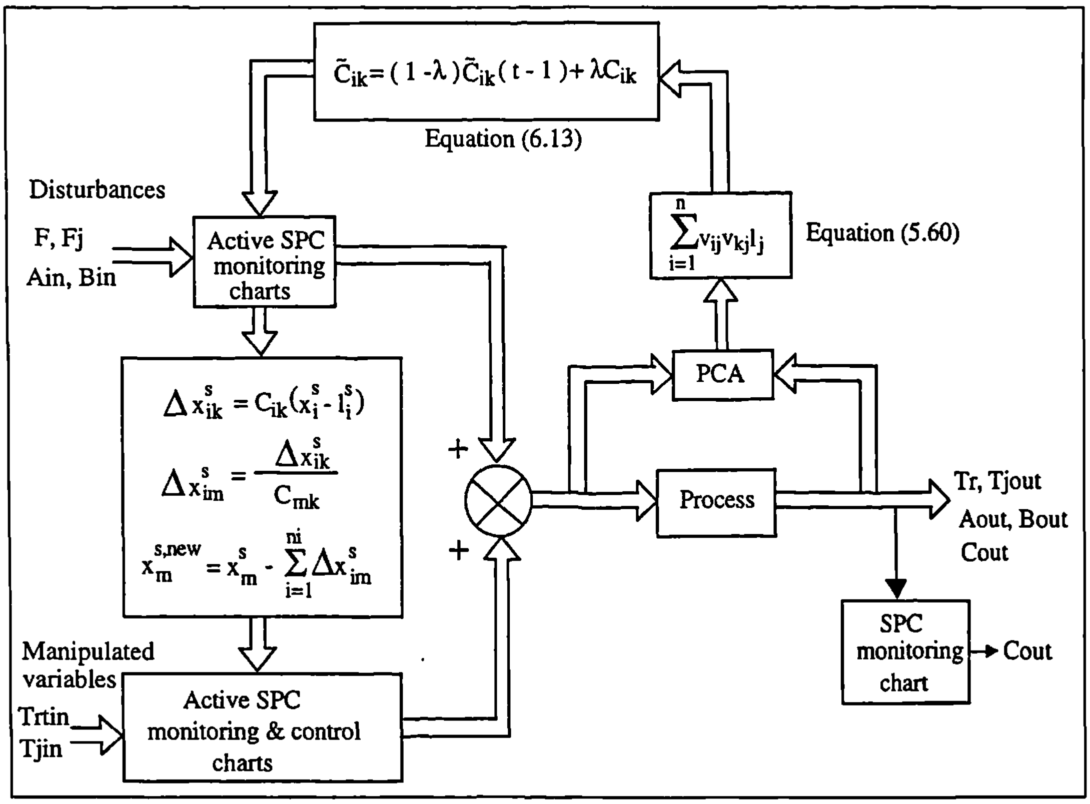

Figure 7.9 Active SPC scheme for on-line PCA calculated control limits using only two manipulated variables

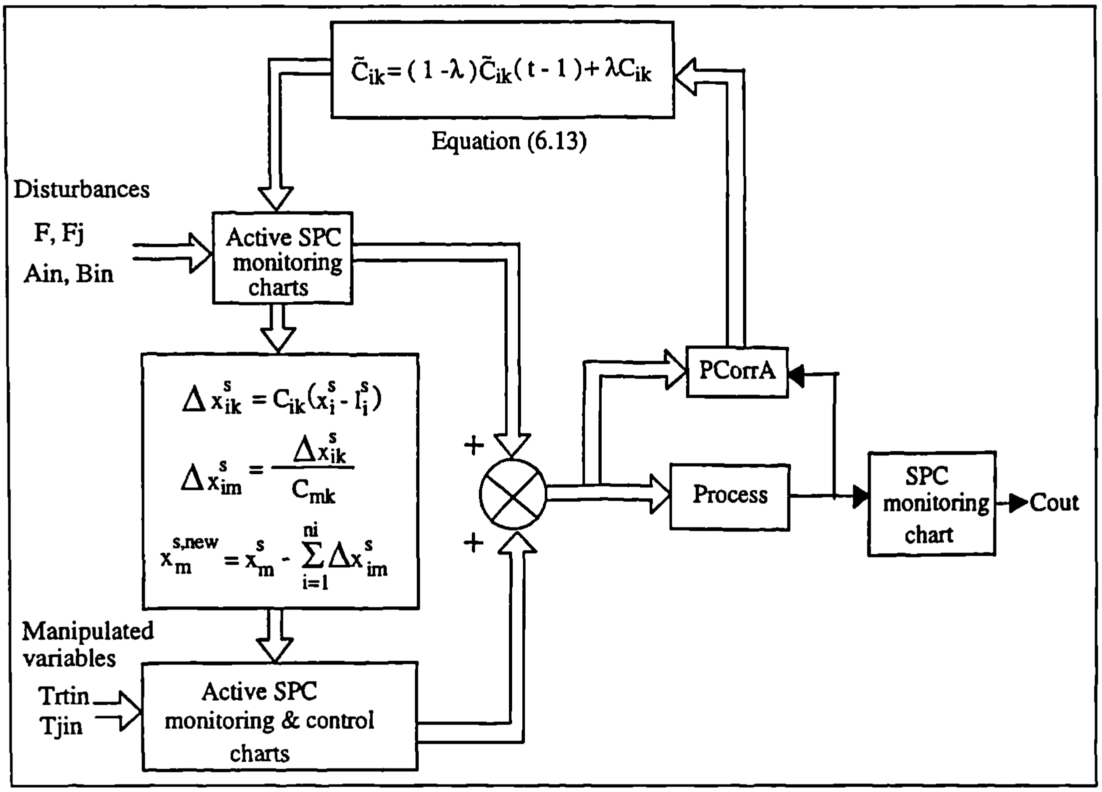

Figure 7.10 Active SPC scheme for on-line PCorrA calculated control limits using only two manipulated variables

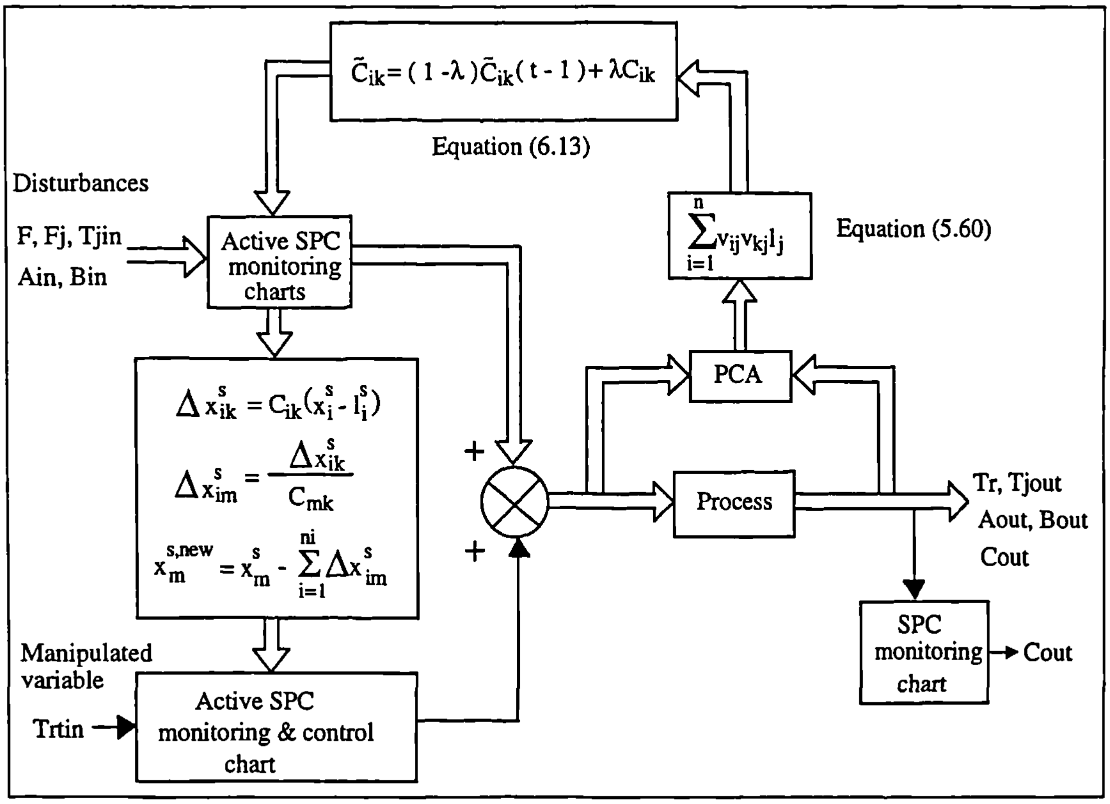

Figure 7.11 Active SPC scheme for on-line PCA calculated control limits using only a single manipulated variable

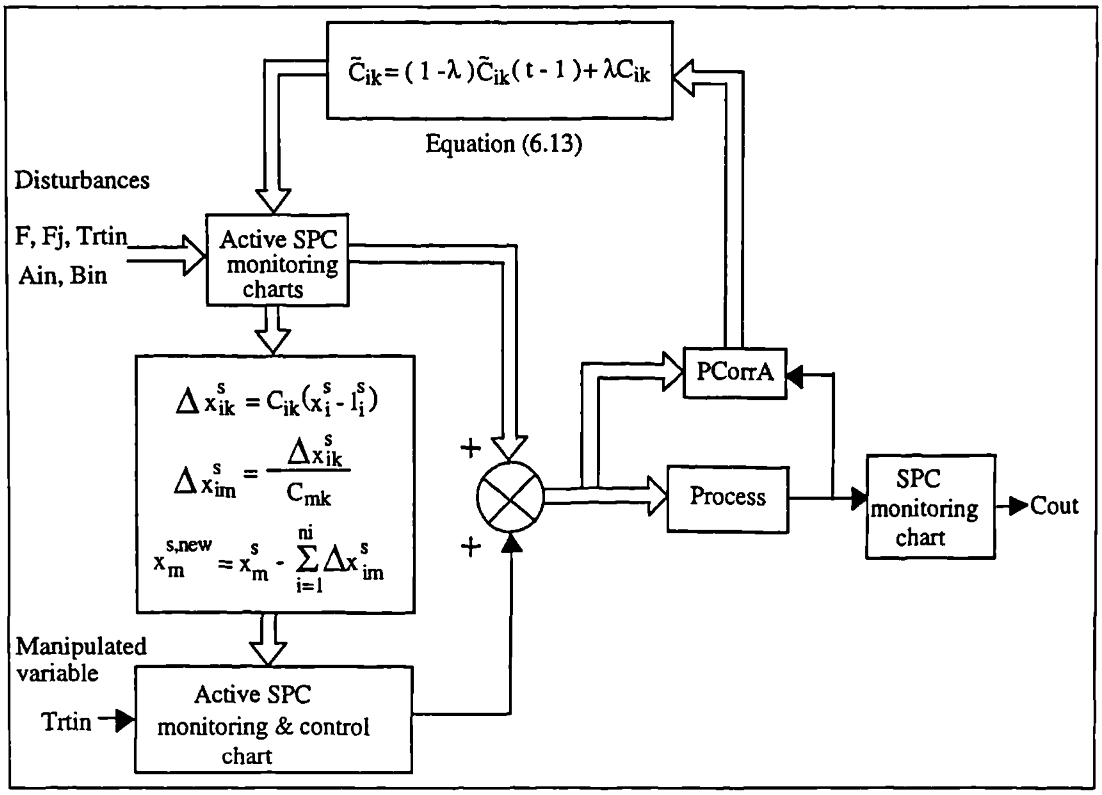

Figure 7.12 Active SPC scheme for on-line PCorrA calculated control limits using only a single manipulated variable

7.4.3 Application of PCA and PCorrA

The 100 data records, as described previously, were used to determine the control limits used in implementing the Active SPC methodology. When on-line calculated coefficient were employed, the off-line calculated C_{ik} were used for the first 50 observations. Thereafter, the C_{ik} were calculated every 5th. sample instant using a window of 50 previous samples. These were then smoothed using equation (6.13) with $\lambda = 0.3$ and the smoothed values were used to update the respective control limits and the manipulated variables. When PCA was used to determine the C_{ik} coefficients, only those principal components which contributed to 95% of the cumulative variation were considered.

7.5 Results and Discussions

The performance of the various strategies were evaluated by observing the percentage of out-of-control (NOC), controlled (NUC) and false alarms (NFA) associated with the quality variable over 500 observations when the Active SPC methods were applied to the CSTR process. The percentage of NOC, NUC and NFA were determined based on the number of out-of-control points that would have occurred if the process was not controlled. Table 7.5 *shows the maximum number of out-of-control situations* and the maximum number of false alarms in each type of control charts if the process was not controlled.

Table 7.5 Maximum NOC and NFA in different control charts

Type of Charts	Maximum number of Out of Control Points	Maximum number of False Alarms
ShewA	11	489
ShewAW	19	481
EWMA	38	462

Table 7.5 also shows the sensitivity of the respective SPC charts. The ShewAW chart have eight extra out-of-control situations compared to the ShewA chart. These eight extra out-of-control points were due to the violation of two out of three consecutive point falling between *mean* $\pm 2\sigma$ to *mean* $\pm 3\sigma$ limit lines. The EWMA chart is even more sensitive having thirty eight out-of-control situations. The number of false alarm for each SPC charts were calculated by taking away the maximum number of out-of-control points from the maximum number of observations in the process. The rest of

the results from this work is tabulated in Appendix A. Selected features from these results will be discussed in the following sections.

7.5.1 Using Fixed Control Limits from off-line Multivariate Analysis

Figure 7.13 shows the performances of the various Active SPC strategies that manipulated all variables when the coefficients were calculated using off-line analysis. As mentioned before, the PCorrA based charts (x1xx) are conservative compared to the PCA based charts (x2xx). For example, this can be seen by comparing the performance of run 1111 with run 1211. Run 1211 resulted in less out-of-control instances compared to run 1111.

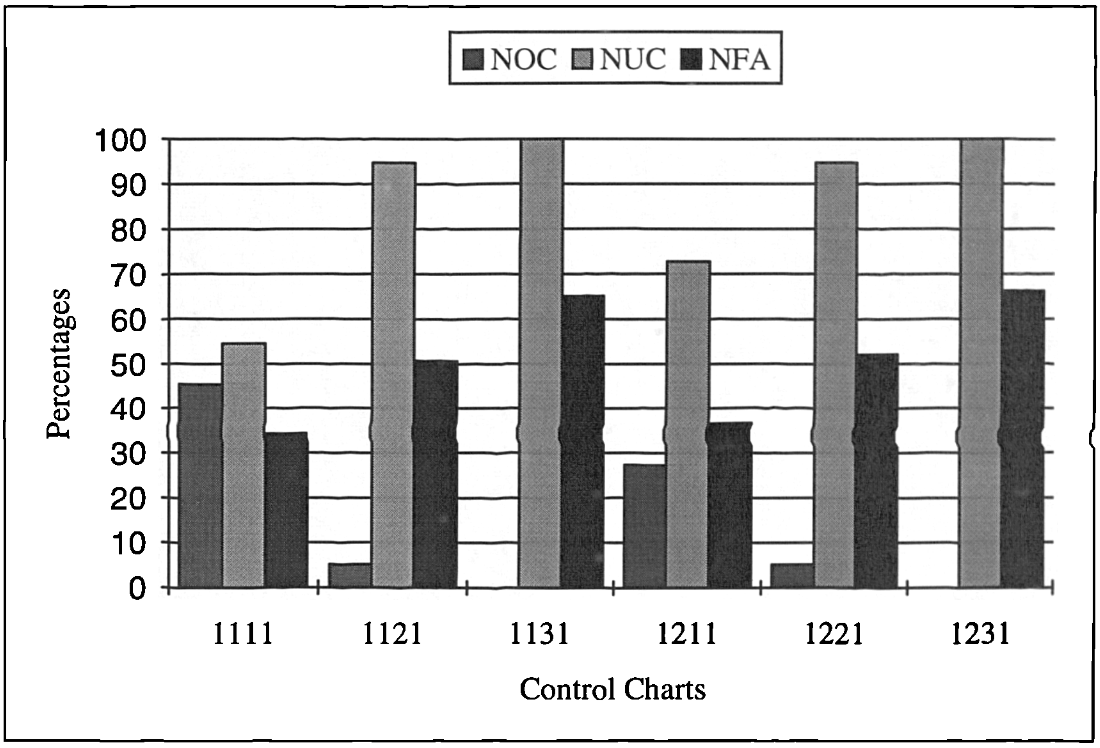

Figure 7.13 Performances of off-line analysis by manipulating all variables.

Comparing the performances of all the SPC charts considered, we are able to assess the sensitivity of the control charts involved. ShewAW charts (xx2x) only allow a low percentage of out-of-control points to occur. Meanwhile, the EWMA charts (xx3x) are the most sensitive. It does not allow any out-of-control situations to take place in the process. Since the EWMA charts are sensitive to small shifts in the process, it will always be alert and take control action when the process show signs of change. This is also reflected in the higher number of false alarms recorded. The main reason for this is that it has the smallest band of control limits compared to the other two charts. As a result the control is tighter but at the expense of a higher number of false alarms. This behaviour of high false alarms can also be seen in the ShewAW charts' performance, which was caused by the implementation of extra control rules.

Figure 7.14 shows the performances of the Active SPC strategies, manipulating only *Trtin* and *Tjin* using off-line calculated coefficients. The EWMA and ShewAW charts maintained their tight control performance by not allowing out-of-control points in the process. Looking at the figure, we can again see that in general, tight control performance is accompanied by a high number of false alarms.

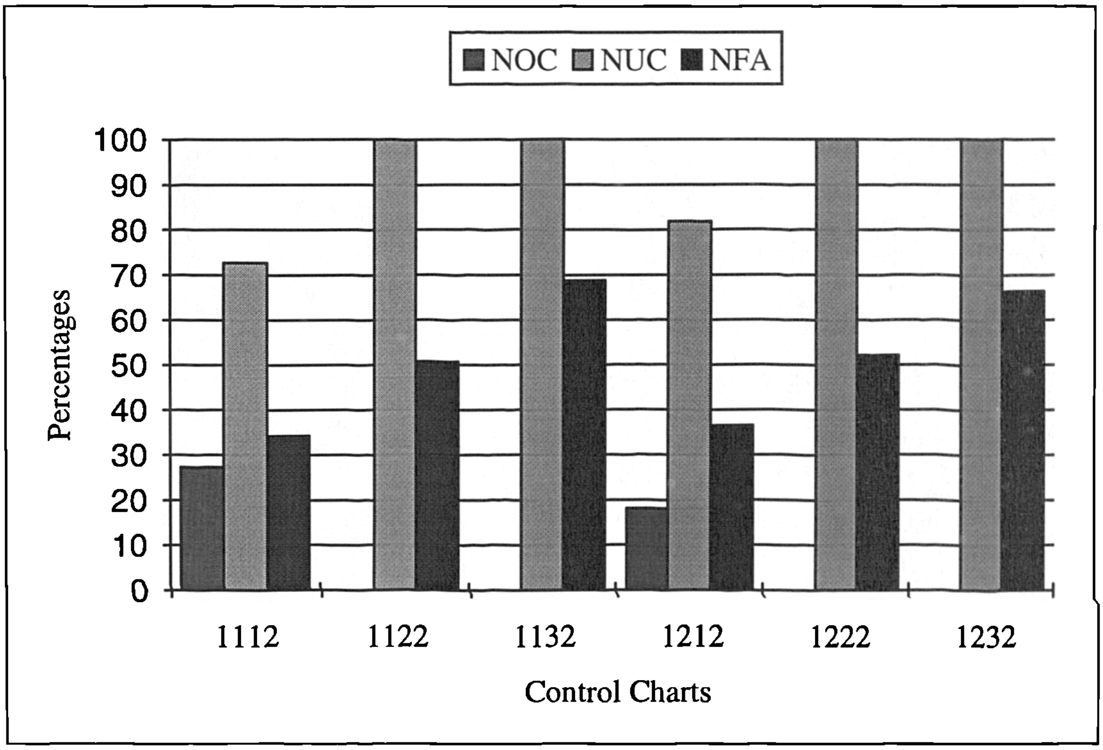

Figure 7.14 Performances of off-line analysis by manipulating *Trtin* and *Tjin*

The performances of manipulating only two variables are better compared to those manipulating all variables. This can be seen by comparing the performance of runs 1111 and 1211 from figure 7.13 with their counterparts 1112 and 1212 in figure 7.14. By changing the manipulation strategy from all input variables to two input variables the percentage of out-of-control using scheme 1112 was reduced by 18% and for the 1212 chart, were reduced by 9%. The possibility of this good performance is because the out-of-control situations of the other measured disturbances variables were compensated by utilising *Trtin* and *Tjin* as the manipulated variables, both of which have the highest correlation between the input and the quality variables (Table 7.3). Thus these manipulated variables will be very sensitive to changes in the process. Even though *Trtin* and *Tjin* were also manipulated in the (xxx1) schemes, the effect of trying to compensate for deviations in measured disturbance variables in (xxx2) schemes, ensures that *Trtin* and *Tjin* are always affected when we have out-of-control situations in these measured disturbances variables (see equation 6.12). However, the number of false alarms remain approximately the same as in the previous case.

Figure 7.15 shows the performances of the SPC charts, manipulating only *Trtin*, using off-line calculated control limits. The EWMA and ShewAW strategies still maintain their tight control performance by not allowing any out-of-control situations. Overall, better control is achieved using this single manipulated variable. The percentage out-of-control in run 1113 is further reduced by 18% using this manipulation strategy. On the other hand, there are no out-of-control points for run 1213 (see runs 1211 and 1212 in figures 7.13 and 7.14).

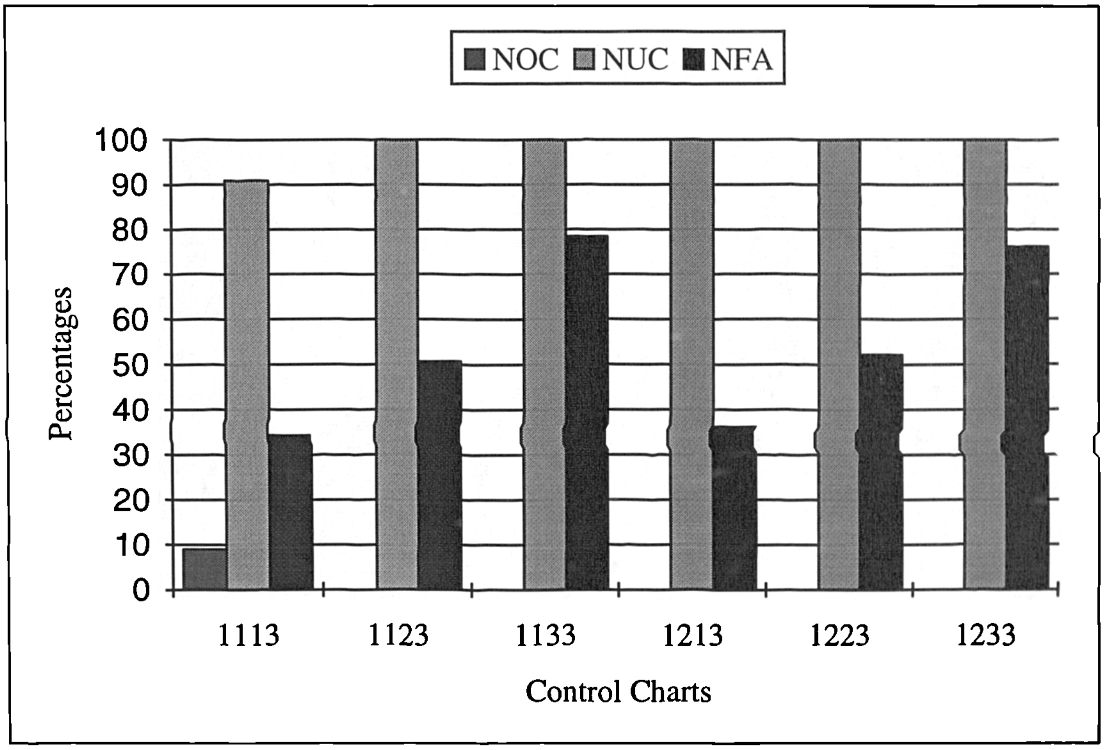

Figure 7.15 Performances of off-line analysis by manipulating only *Trtin*

From the preceding discussion we can see that changing the manipulation strategy from all variables to two and finally to one variable reduces the percentage of out-of-control points for the ShewA and ShewAW charts. For ShewAW charts, both multivariate methods reduce the out-of-control points by 5% when we reduce the number of manipulated variable from all to two. However, each change of manipulation strategy reduces the percentage of out-of-control cases in PCorrA based ShewA charts by 18%. For PCA based ShewA charts, changing the manipulation strategy from manipulating all inputs to two reduces the percentage of out-of-control point by 9%. However, reducing the manipulated variable from two to one for PCA based charts, a further reduction of 18% in the percentage of out-of-control points is achieved. A possible reason for this good performance when manipulating a single variable is because we used *Trtin* as the manipulated variable which is sensitive to changes in the process (*e.g.* having the highest C_{ik} value, see table 7.3). It seems that the summation effects of out-of-control situations from the rest of the measured disturbances render control using *Trtin* more sensitive.

All EWMA and ShewAW charts with off-line calculated control rules have identical percentages of controlled points but differ slightly in the percentage of false alarms, except for runs 1121 and 1221. If the performances of both multivariate methods are compared using the preceding mentioned strategies, we see that PCorrA based charts generally have a lower percentages of false alarm compared to PCA based charts. This is because the *correlation* based on PCorrA is smaller compared to PCA (see table 7.3) and thus it will take a slightly more conservative approach towards controlling the process.

From this set of results we can observe that the control performance generally becomes better as we change the type of control charts that are used. This can be seen by low percentages of out-of-control (NOC) situations when ShewAW charts were used and no out-of-control points when EWMA were utilised. This is expected, since tighter control limits and more control rules on the SPC charts will suppress out-of-control situations, but at the expense of an increase in false alarms.

7.5.2 Results using on-line Calculated Control Limits

Figure 7.16 shows the performances of the Active SPC charts, manipulating all variables, and using on-line updating of $\tilde{C}_{ik}$. As illustrated in the figure, PCorrA based SPC charts (2111) resulted in 18% out-of-control points. This is a 40% improvement over the results obtained using the (1111) scheme. Meanwhile, the PCA based SPC chart (2211) allowed about 45% out-of-control cases to pass through the process. This is 18% more out-of-control points compared to the performance of run 1211.

The values of $\tilde{C}_{ik}$ calculated on-line using PCorrA started to increase compared to the off-line method. Meanwhile, the values of $\tilde{C}_{ik}$ based on PCA started to decrease. This can be seen in figure 7.17, where the $\tilde{C}_{ik}$ for *Trtin, Tjin, Ain, Bin, F* and *Fj* are seen to be increasing for run 2111 and decreasing for run 2211. As a result the PCorrA based charts become more sensitive to changes in the process, while the PCA based charts become conservative, *i.e.* have wider control limits. Thus, better control is achieved using PCorrA based charts. On the other hand, the performance of PCA based charts degraded when the $\tilde{C}_{ik}$ coefficients are calculated on-line. Nevertheless, the method performed well when fixed C_{ik} values were used because the data is free from out-of-control situations. During on-line updating, the calculation of $\tilde{C}_{ik}$ is based on the data as the process being controlled, and this included data containing several out-of-control situations.

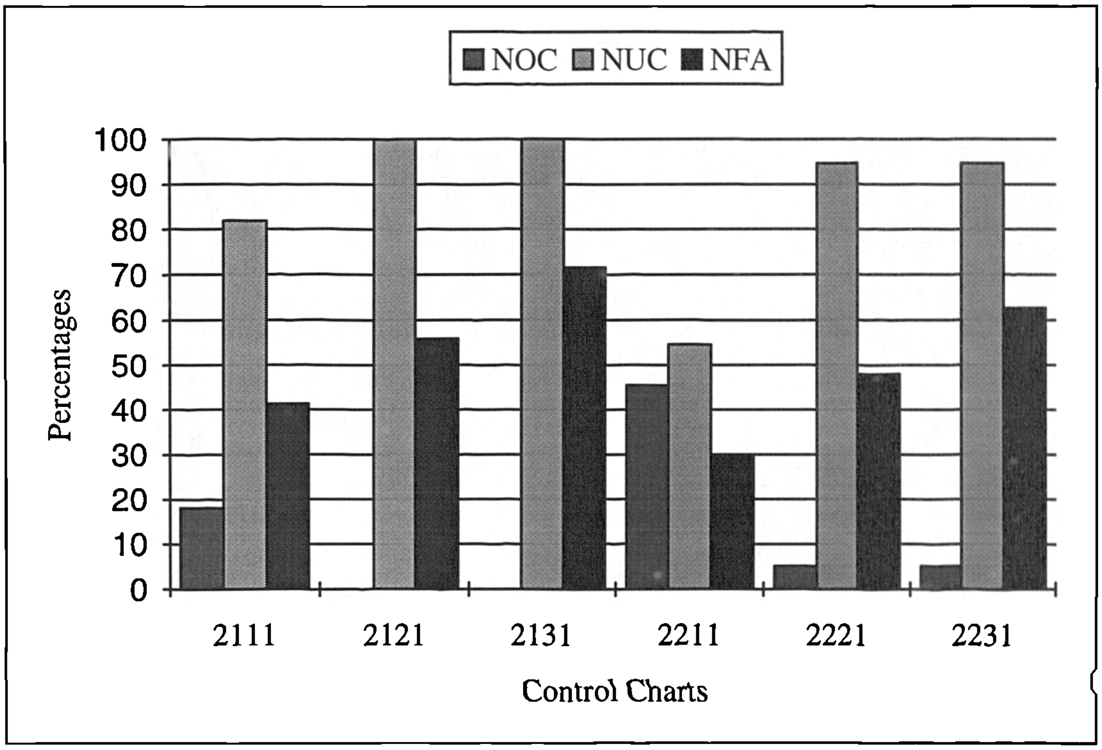

Figure 7.16 Performances based on on-line analysis by manipulating all variables

The same trend of degrading performances can also be seen for the on-line calculated control limits for the PCA based ShewAW and EWMA charts. There are around 5% out-of-control situations for these two charts. Both of these phenomenon will be discussed later in section 7.5.3. Here, the observations indicate that we can reduce the percentages of out-of-control points in the process by changing the types of PCA based charts. The 45% out-of-control points using ShewA chart (2211) can be reduced to roughly 5% by using the ShewAW (2221) and EWMA (2231) charts.

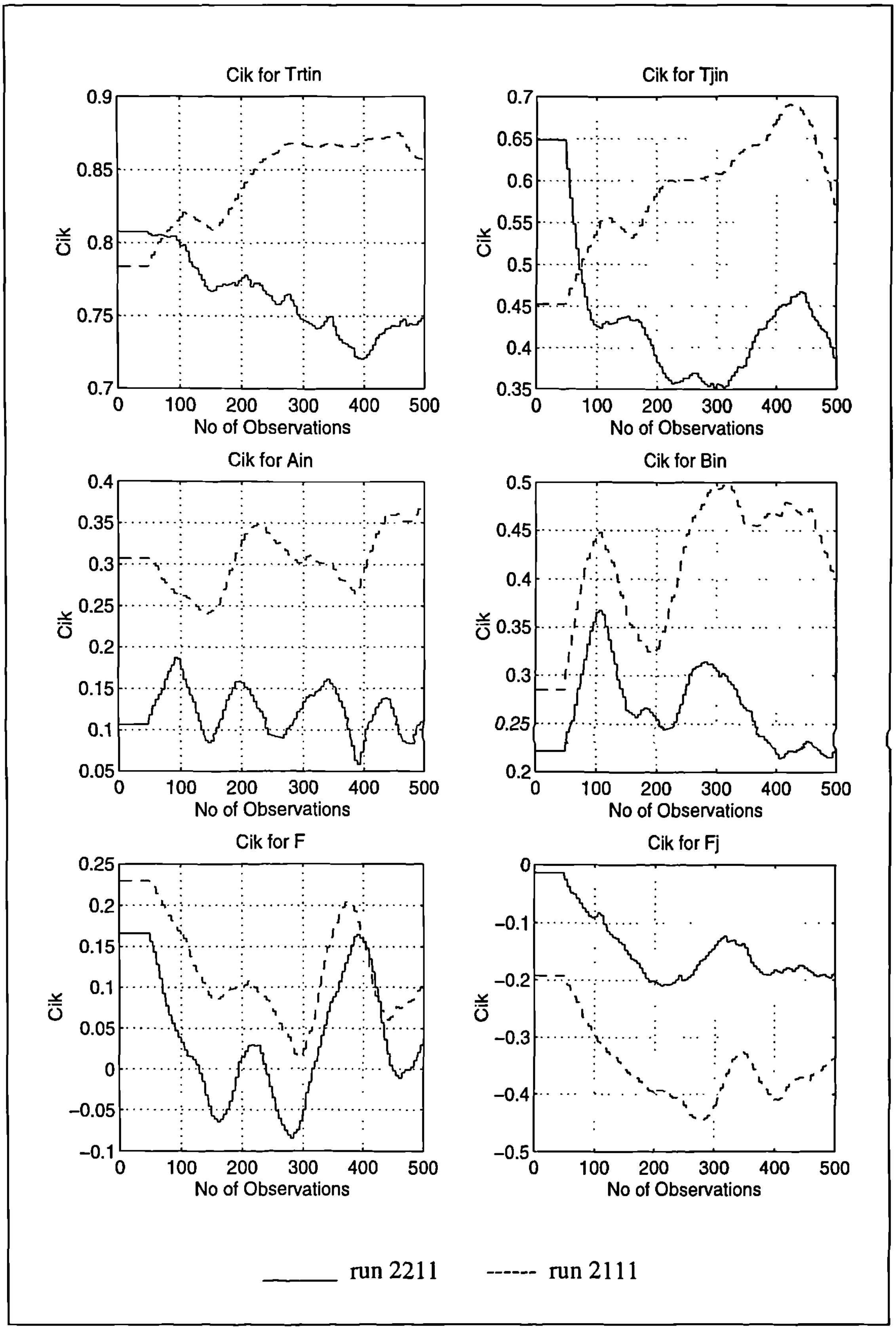

Figure 7.17 $\tilde{C}_{ik}$ values for runs 2111 and 2211

Figure 7.18 shows the performances of the SPC charts that update $\tilde{C}_{ik}$ on-line, and manipulating *Trtin* and *Tjin*. The on-line updated $\tilde{C}_{ik}$ via PCorrA reduces the out-of-control points in the ShewA SPC chart (2112) by 9% compared to the 2111 chart. Meanwhile the rest of PCorrA based charts maintained the good control performances by not allowing any out-of-control situations to pass through the process.

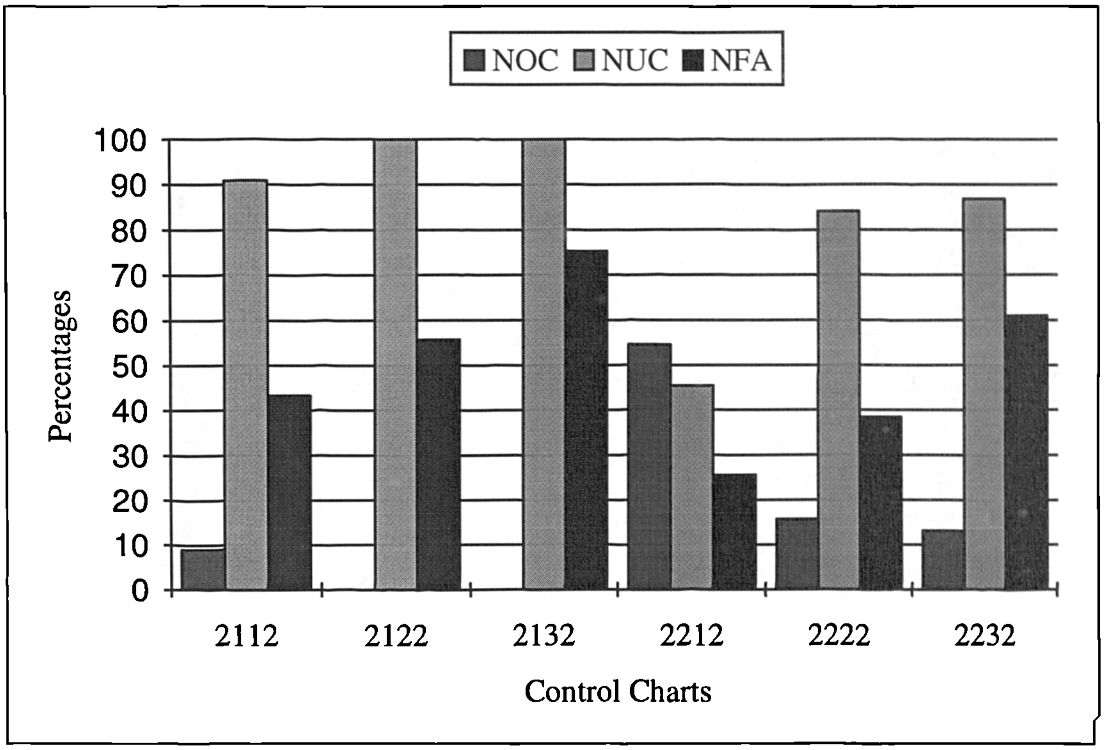

Figure 7.18 Performances based on on-line analysis and manipulating *Trtin* and *Tjin*

On the other hand, on-line updating of $\tilde{C}_{ik}$ using PCA resulted in higher percentages of out-of-control situations compared to the previous manipulating strategy. This can be explained by looking at figure 7.19, where the $\tilde{C}_{ik}$ values of runs 2211, 2212 and 2213 are illustrated as the process is being controlled. From the figure we can see that the $\tilde{C}_{ik}$ values for *Trtin* and *Bin* in run 2212 are smaller than run 2211. At the same time the other $\tilde{C}_{ik}$ values are similar. Since, for this control schemes we are using just two variables to control the process, the control charts for *Trtin* will become conservative and have wider limits. This will allow more variations to enter the process. Nevertheless the percentages of out-of-control situation reduce as we change the type of control chart to govern the process, *e.g.* from ShewA to ShewAW and lastly to EWMA.

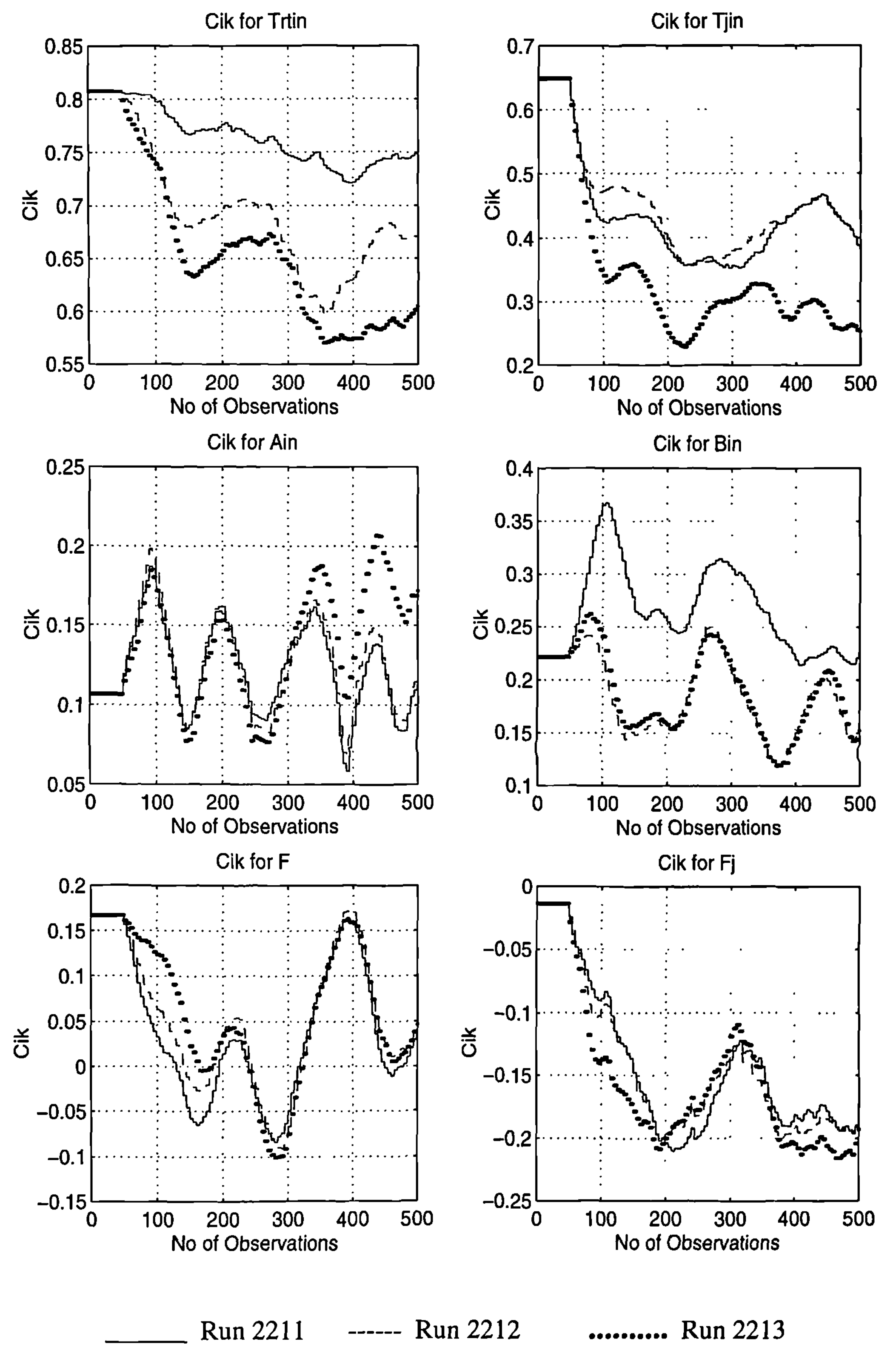

Figure 7.19 $\tilde{C}_{ik}$ values for runs 2211, 2212 and 2213

Figure 7.20 shows the performances of the last set of SPC charts with $\tilde{C}_{ik}$ updated on-line, and manipulating only *Trtin*. There are no out-of-control points when the 2113 chart was used, an improvement over the previous chart of 2112. Here we see that changing the strategy of PCorrA based ShewA charts from manipulating all to two and finally to a single variable decreases the percentage out-of-control cases by 9% each time, (compare runs 2111, 2112 and 2113). The above situations can be explained by looking at how the $1/\tilde{C}_{ik}$ values based on PCorrA changes during the control run as shown in figure 7.21. In this figure, the $1/\tilde{C}_{ik}$ values of run 2111, 2112 and 2113 are plotted for comparison. From the figure we can see that the $1/\tilde{C}_{ik}$ values for *Trtin* and *Fj* are similar for all runs. The $1/\tilde{C}_{ik}$ values for *Bin* and *F* are similar for runs 2112 and 2113, but they are smaller than that for run 2111. On the other hand, the $1/\tilde{C}_{ik}$ values for *Tjin* are similar for run 2111 and 2112, but they are bigger than that for run 2113. Lastly when we compare the $1/\tilde{C}_{ik}$ value for *Ain*, we see that run 2111 has a larger $1/\tilde{C}_{ik}$ than 2112, and 2112 has a larger $1/\tilde{C}_{ik}$ value than 2113. From this, through the combined effects of all variables, we can deduce why the control performance of manipulating a single variable (*Trtin*) is better than manipulating two variables and all variables for PCorrA based ShewA charts. Even though we are using only *Trtin* to manipulate the process, in run 2113 we still have to take into account the effect of the rest of the measured disturbances variables in the above discussion. This is because they were the basis of intended control limit by which *Trtin* is used to compensate for variations *in the other inputs to the process (refer to equations 6.11 and 6.12).*

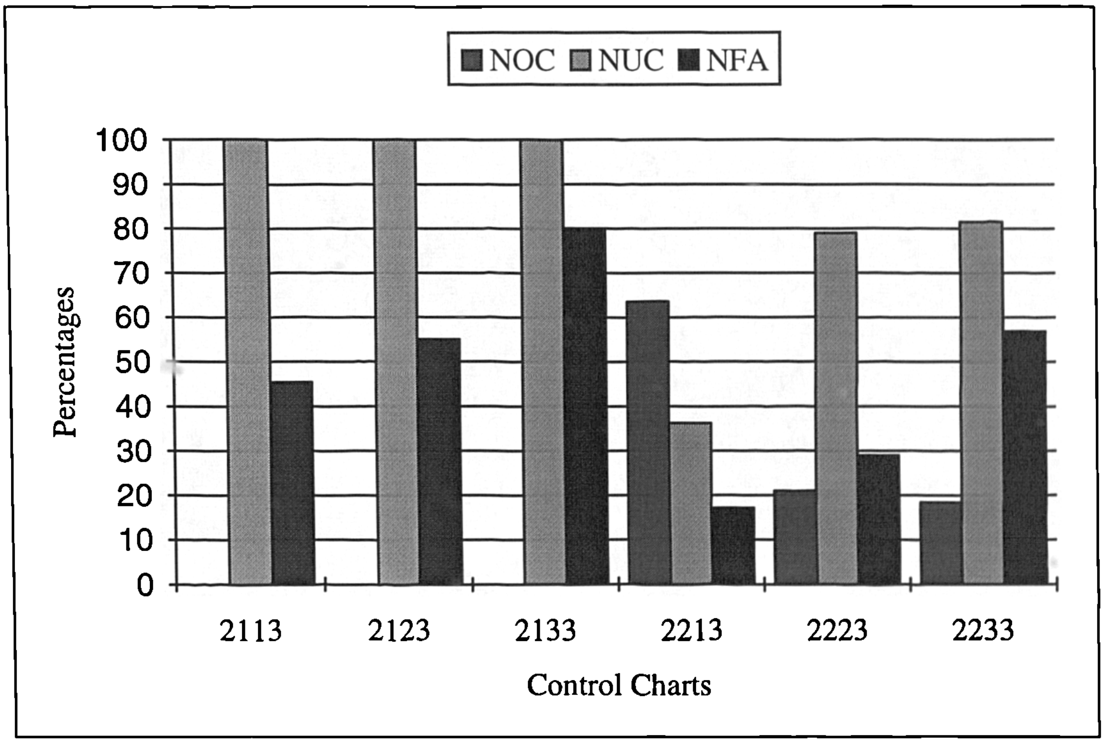

Figure 7.20 Performances based on on-line analysis and manipulating only *Trtin*

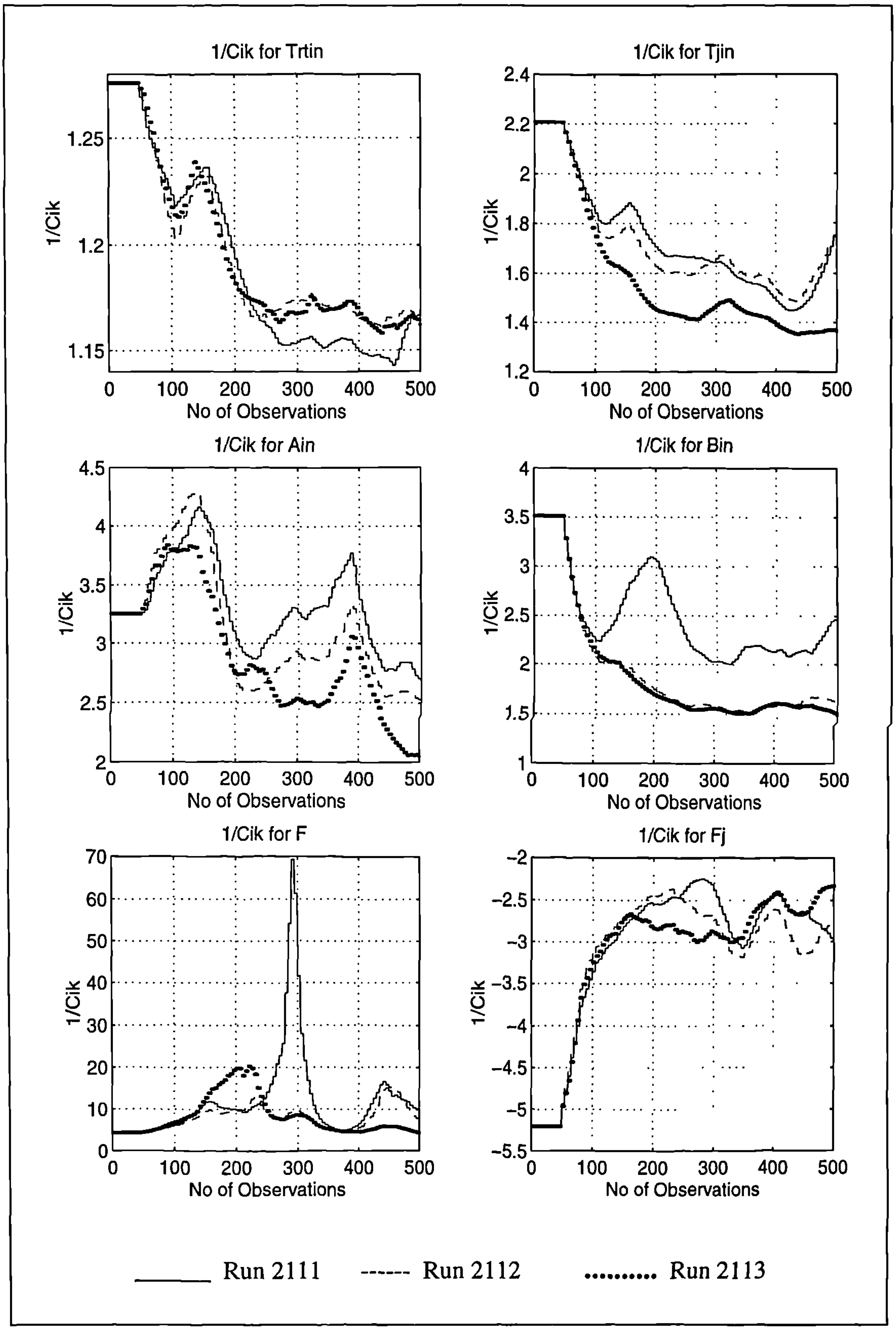

Figure 7.21 $1/\tilde{C}_{ik}$ values for runs 2111, 2112 and 2113

However, the ShewA charts based on PCA show increases of 9% in out-of-control situations as we change the manipulation strategy from manipulating all to two and finally to a single variable, (compare the runs of 2211, 2212 and 2213). This phenomenon can be seen by looking at the $\tilde{C}_{ik}$ values of *Trtin* and *Tjin* of run 2213, which is smaller than run 2212 in figure 7.19. Due to this phenomenon, the 2213 control scheme, (*i.e.* only *Trtin* is the manipulated variable for this case) is no longer sensitive compared to run 2212.

Updating the $\tilde{C}_{ik}$ values on-line for PCorrA based charts improved the control performance compared to the off-line method. This can be seen by comparing the performances of charts 2111, 2112 and 2113 with their off-line counterparts, 1111, 1112 and 1113. For the first case, comparing 2111 and 1111, there were 27% less out-of-control situations, for the second case, comparing 2112 and 1112, there were 18% less out-of-control situations and finally when comparing 2113 and 1113, there were 9% less out-of-control situations when $\tilde{C}_{ik}$ was calculated on-line.

On the other hand, the converse was true when PCA was used to calculate $\tilde{C}_{ik}$. Here the use of on-line calculated control limits degraded control performance and several out-of-control points were allowed to pass through the process.

7.5.3 Index of Performance

From the above discussions, it is difficult to discern which configuration gave the best overall control performance. Figures 7.13 to 7.16, 7.18 and 7.20 show that high percentage of points under control (NUC) is normally accompanied by a high rate of false alarms, *i.e.* increased producer's risk, which is undesirable. In judging which of the Active SPC configuration provided the best overall performance, a compromise is therefore necessary to balance reduced consumer risk against increased producer risk. In this investigation, the following heuristically derived *index of performance* (*IP*) was used to identify the configuration that provided the best control:

$$IP = 1 - 0.5\left(\frac{\%NOC + \%NFA}{\%NUC + \%NFA}\right) \qquad (7.1)$$

This index penalises the percentage of out-of-control points as well as the percentage of false alarms. It is scaled so that the best control strategy, *i.e.* no out-of-control points and no false alarms would have an *IP* equal to one. The worst case which corresponds to all points being out-of-control (*e.g.* NOC=100%, NUC=0%) and 100 percent false alarms would lead to an *IP* of zero. The *IPs* for the various configurations tested previously are plotted in figures 7.22 to 7.23 and 7.27.

From figure 7.22, which shows the *IPs* of ShewA charts, we see that changing the number of manipulating variables from all to two and finally to one generally increases the *IP* for SPC charts using fixed control limits. The use of on-line calculated $\tilde{C}_{ik}$ using PCorrA also shows the same trend. In contrast, the use of on-line calculated $\tilde{C}_{ik}$ via PCA resulted in decreasing *IP* values as the number of manipulated inputs were decreased.

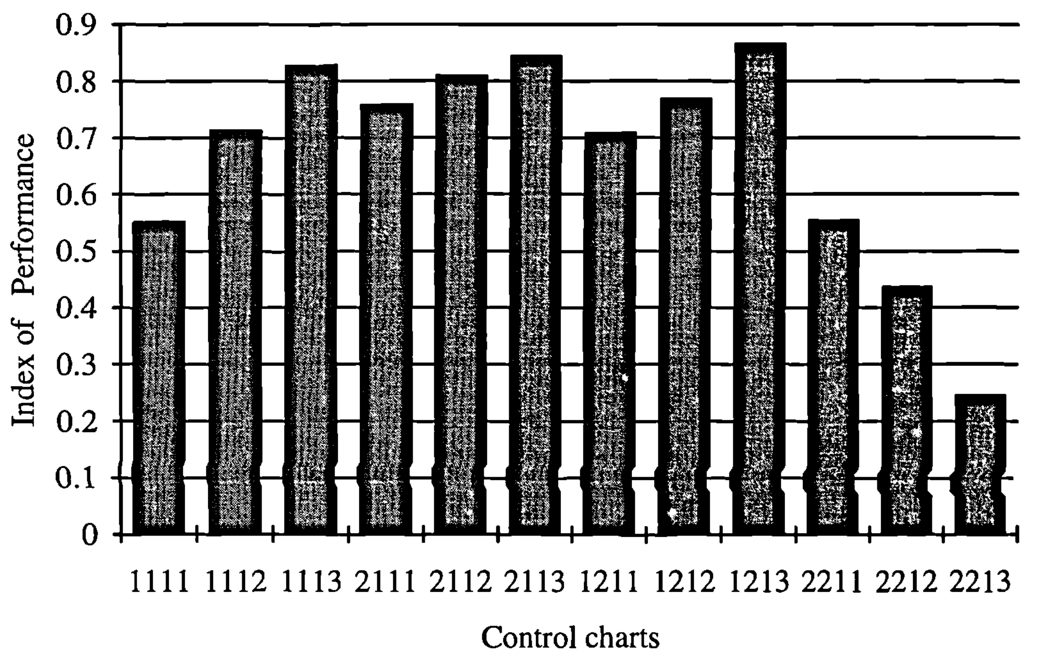

Figure 7.22 *IPs* for ShewA charts

The best *IP* performance for the ShewA charts was recorded by the 1213, followed by the 2113 control scheme. Both of these runs did not allow any out-of-control points to pass through the process. In fact these two control schemes have the highest *IP* values when we compare all the control schemes considered in this study. A possible reason why the 1213 and 1113 control schemes have the best *IP* is because the manipulated variable that we chose, *Trtin*, was very sensitive to changes in the process (see table 7.3).

Figure 7.22 also show that the *IPs* of PCA based ShewA charts where C_{ik} is fixed, are generally better compared to its multivariate analysis counterpart, PCorrA based ShewA charts using the same C_{ik} calculation scheme (compare 1211 with 1111, 1212 with 1112, and 1213 with 1113). This is also because the former is more sensitive to changes compared to the latter as shown in table 7.3.

The results in figure 7.22 also shows that the use of PCorrA to update the limits for ShewA charts on-line generally leads to better control performances than when fixed limits were used. This implementation deal with changes in process characteristics by changing the control limits.

Figure 7.23 shows the *IPs* for the ShewAW charts. The ShewAW charts maintained the same trend of *IPs* as the ShewA charts, where charts using off-line calculated limits and limits calculated on-line via PCorrA show better control performance when the number of manipulating variables was reduced. This is because as we reduce the number of manipulating variables from all to two and finally to one, for the off-line calculated control limits, the combination of manipulative variables that we chose as the substitute become more sensitive to changes in the process (*i.e.* have high C_{ik} values). On the other hand, the on-line updating scheme for PCA based ShewAW charts show a contrasting behaviour, having a downward trend in *IPs*. This is because the $\tilde{C}_{ik}$ values for the chosen manipulated variables such as *Trtin* and *Tjin* started to decrease as we reduce the number of manipulated variables for on-line PCA updating schemes. This phenomenon can be seen in figure 7.24 where the $\tilde{C}_{ik}$ values for all input variables are plotted for runs *2221, 2222 and 2223.*

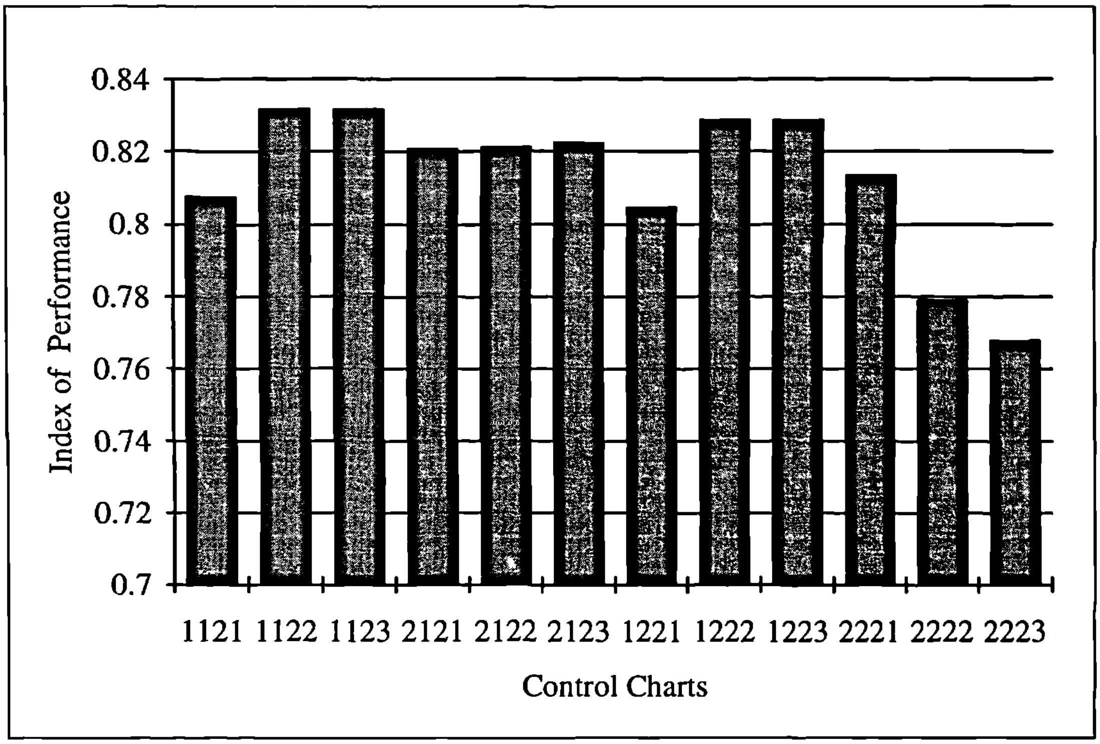

Figure 7.23 *IPs* for ShewAW charts

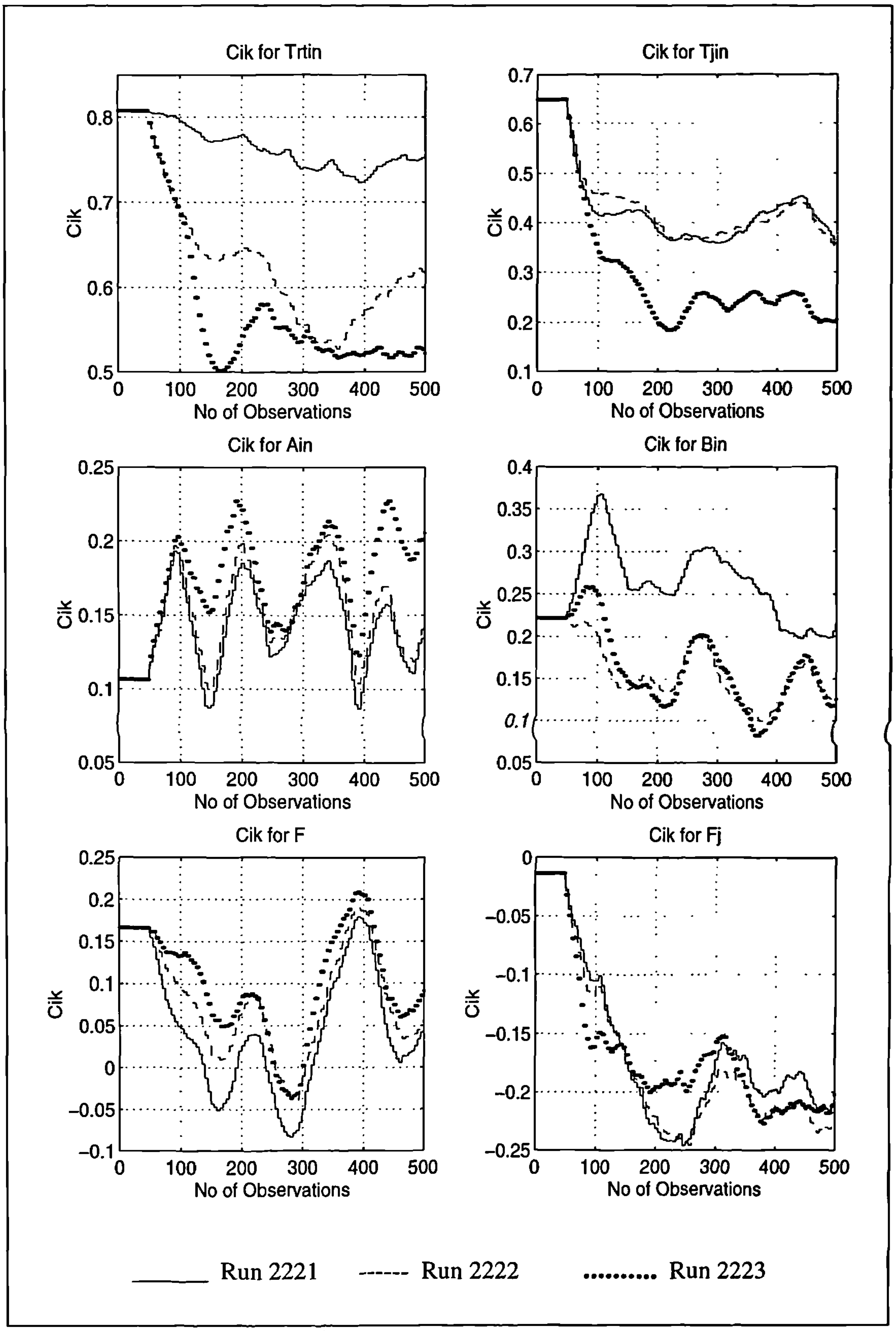

Figure 7.24 $\bar{C}_{ik}$ values for runs 2221, 2222 and 2223

Comparing the *IPs* for all PCorrA based ShewAW charts, we see that the *IP* values of off-line designs are generally better than the on-line methods. It seems that changing the specification of control limits from off-line design to on-line updating deteriorates the overall performance. Nevertheless, all the charts (x12x) do not allow any out-of-control point to pass through the process except for run 1121. Thus the NOC = 0%, NUC = 100%. The only difference between all these charts is their values of NFA. If we arrange the *IPs* of charts, *i.e.* 1121,1122,1123,2123,2122, and 2121; we see from figure 7.25 that changing control limits calculation strategy from off-line to on-line, the number of false alarms is increased.

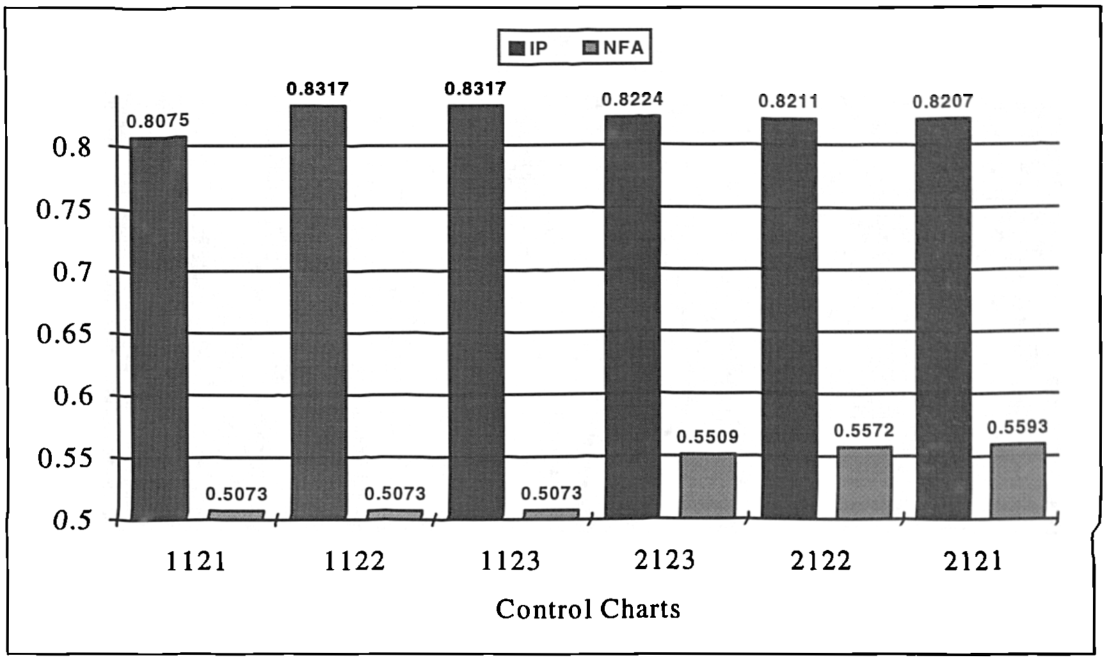

Figure 7.25 *IPs* for PCorrA ShewAW charts

The $\tilde{C}_{ik}$ values for PCorrA on-line updating schemes are illustrated in figure 7.26. The $\tilde{C}_{ik}$ value of *Trtin* for a single manipulated variable is the most conservative, compared to the other two manipulating strategies. Nevertheless, through the combined effects of other input variables, the performance of run 2123 is a bit conservative compared to run 2122 and run 2121. The result is that the percentage of NFA in run 2123 is slightly less compared to runs 2122 and run 2121. They make the *IP* of run 2123 is better than run 2122 and run 2121. The on-line updated scheme PCorrA based ShewAW charts are more sensitive compared to their off-line updated counterparts. Since, the NUC for the on-line updating scheme is zero, this sensitive behaviour increased the percentages of false alarms. As a result some of the off-line updating schemes shows better overall performance than the on-line designs (*e.g.* 1123 compared to 2123 and 1122 compared to 2122).

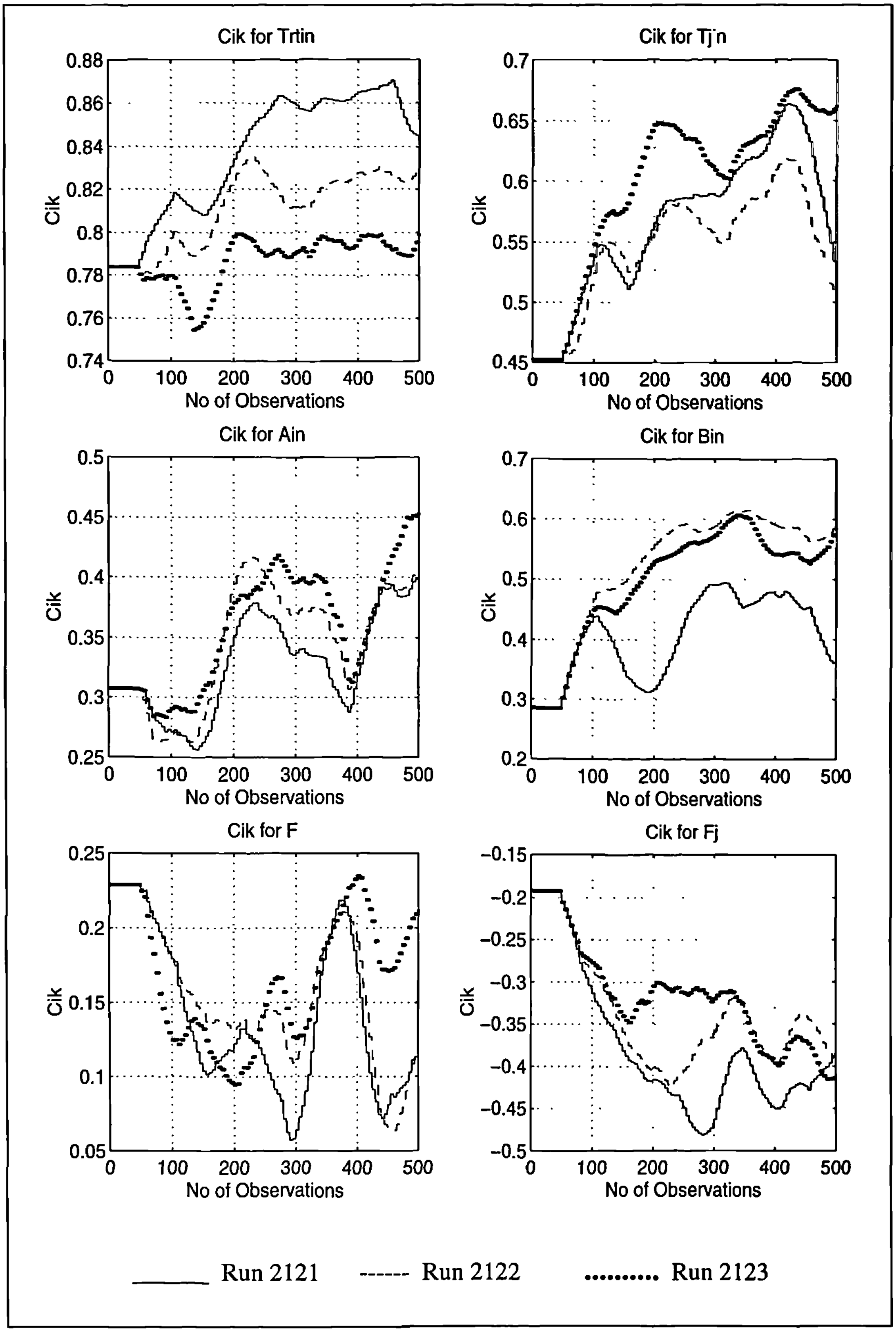

Figure 7.26 $\tilde{C}_{ik}$ values for runs 2121, 2122 and 2123

The best overall *IP* for the ShewAW charts is that obtained using PCorrA off-line design by manipulating two and a single manipulated variable (1122 and 1123). Both of these charts have the same *IP* values. These are followed by the PCA off-line updating schemes utilising the same manipulating strategies (1222 and 1223). Almost all the ShewAW charts used in this study do not violate the warning rules, except for runs 2212 and 2213. Furthermore all the ShewAW charts used in this study have *IPs* greater than 0.75.

Figure 7.27 shows the *IPs* for EWMA charts. For PCorrA based SPC charts, there is a downward trend in *IP* from 1131 to 1133 and 2131 to 2133. However there are no out-of-control situations in all these PCorrA based charts. In general, it is rare to have data points fall outside the limits on EWMA charts because of the tighter limits that these charts are using. The fall in *IPs* is due to the increase in the number of false alarms.

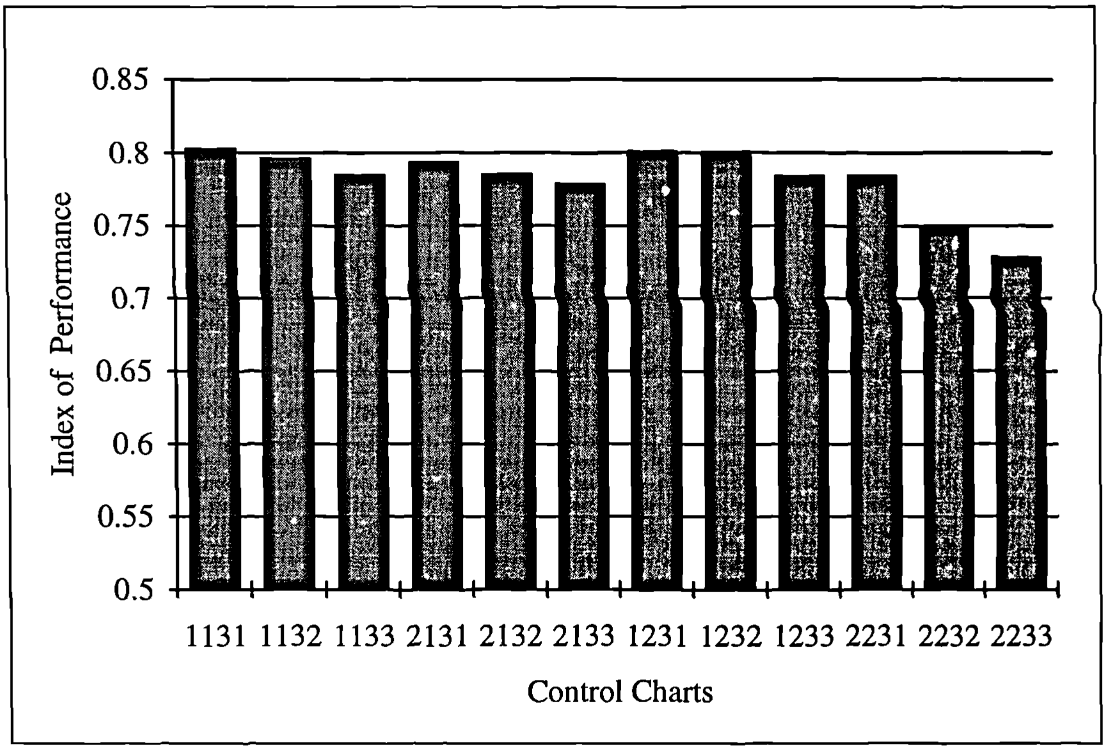

Figure 7.27 *IPs* for EWMA control charts

For the EWMA charts based on PCA, the same circumstances prevail. There is a downward trend all the way from the off-line to on-line calculations of control limits. However, the downward trend of *IPs* for the PCA on-line calculated control limits included cases where there are several out-of-control points. This is because the chosen manipulated variables that are suppose to be sensitive became conservative during the run. This event can be traced to the downward trend in $\tilde{C}_{ik}$ value for *Trtin* as shown in figure 7.28.

In general, the *IPs* for the EWMA control charts decrease as we reduce the number of manipulated variables. Although we have this decreasing trend in *IP* for the EWMA charts, there is no out-of-control incident for the PCorrA based charts and PCA off-line calculated control limits. They only differ because of the percentages of false alarms, which increase as we reduce the number of manipulated variables. This indicates that these control schemes remain sensitive to changes in the process.

Generally, the EWMA charts gave very good control performance by not allowing any out-of-control points to pass through the process except when PCA is used to update the control limits on-line. Nevertheless all the EWMA charts used in this study have *IP* values greater than 0.72. The weakness of this chart is due to the associated high rate of false alarms. From Chapter Two we mentioned that the EWMA chart is equivalent to *Proportional* feedback control. Thus it has some characteristic of feedback control where it takes a lot of control action when the process deviates away from the limits.

From figure 7.22, 7.23 and 7.27 we see that the descending order of *IPs* for PCA on-line updating schemes goes from 22x1 to 22x2 and 22x3. For all these manipulating strategies, the percentages of out-of-control points increase as we change the manipulating strategies from all to two to a single manipulated variable. The main reason for this as mentioned is due to the *correlation* between the quality variable and the chosen manipulated variables become smaller. As a result, the chosen manipulated variables control charts become conservative and can tolerate wider control limits. Due to this effect much variations can enter the process.

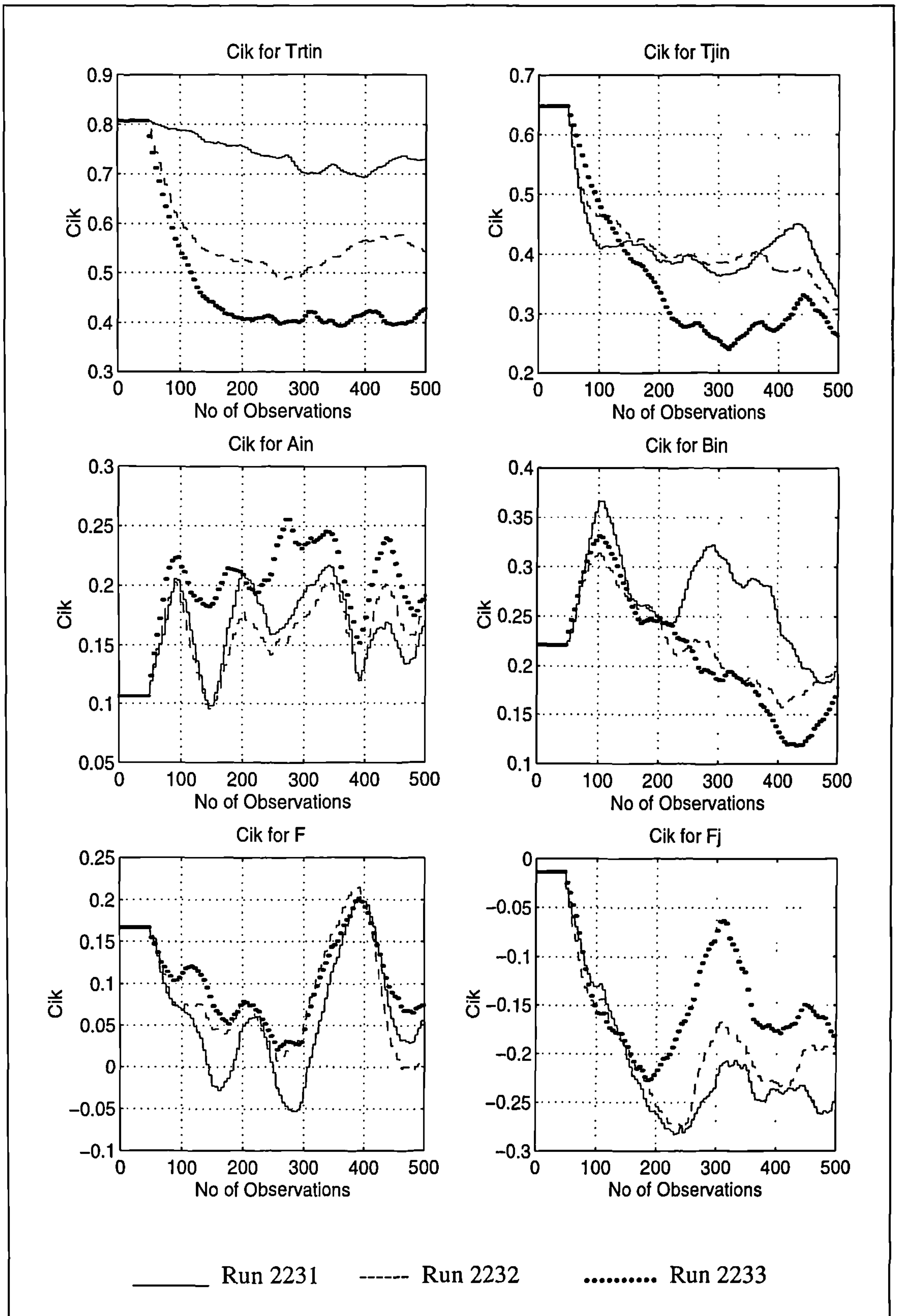

Figure 7.28 $\tilde{C}_{ik}$ values for runs 2231, 2232 and 2233

7.5.4 The Effects of Historical Data in Designing Active SPC Strategies

As mentioned, the data used to design the Active SPC strategies and with their effects on their performances will be discussed here. Two sets of historical data were taken from the process when it is perceived to be in statistical control. These two set of historical data have the same *means* but differ in *standard deviations*, and hence yields different initial values of C_{ik}. The performances of charts designed based on each of these data sets are compared in figures 7.29-7.31. The charts without asterisk is based on historical data that we have used and discussed in previous sections, while the charts coded with asterisk (*) indicates the use of the new batch of historical data. The latter data set will be referred to as data set "b", while the former will be referred to as history data set "a". Only PCorrA based charts will be discussed because they showed more consistent results as indicated by the analyses in section 7.5.1 to 7.5.3. The results for data set "b" and their initial values of C_{ik} are tabulated in Appendix B.

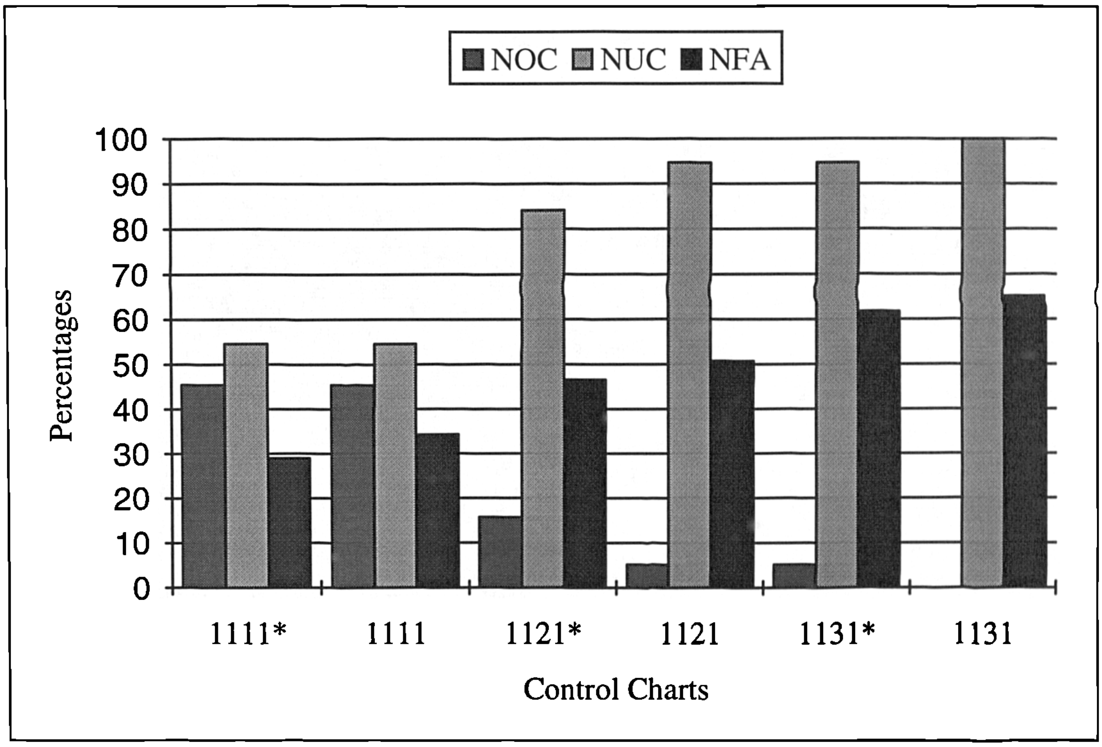

Figure 7.29 Effects of different data sets for off-line PCorrA, manipulating all inputs

Figure 7.29 shows the performances of charts with limits determined off-line and manipulating all variables. From this figure we see that the performance of charts based on data set "a" is superior to those based on data set "b" as the latter strategy allow out-of-control points to pass through the process (*e.g.* comparing 1121* and 1121; and 1131* and 1131). However, when both data sets resulted in charts having the same number of NUC, the use of data set "b" gave better performance than "a", due to less false alarms. This phenomenon can be seen by comparing the runs 1111* and 1111. From this discussion we can see that data set "b" results in more conservative charts compared to data set "a".

To elaborate further, let us look at charts with fixed limits that manipulate only *Trtin*. From figure 7.30 we see the same effects. Whenever out-of-control situations occur, and where the NUC are not the same, the use of data set "a" leads to better control performances than data set "b" and vice versa if there are no out-of-control situations and NUC are the same. By comparing the *IPs* for all these charts in figure 7.31, the same trend is observed. Based on these results we can conclude that although the history data "b" is conservative compared to "a", whenever the control charts are designed based on both data sets can eliminate out-of-control situations from the process, those based on data set "b" will perform better than those based on data set "a". This is because having smaller band of control limits on the charts can only improve the control performance to a certain extent. Beyond this, it will just increase the percentages of false alarm.

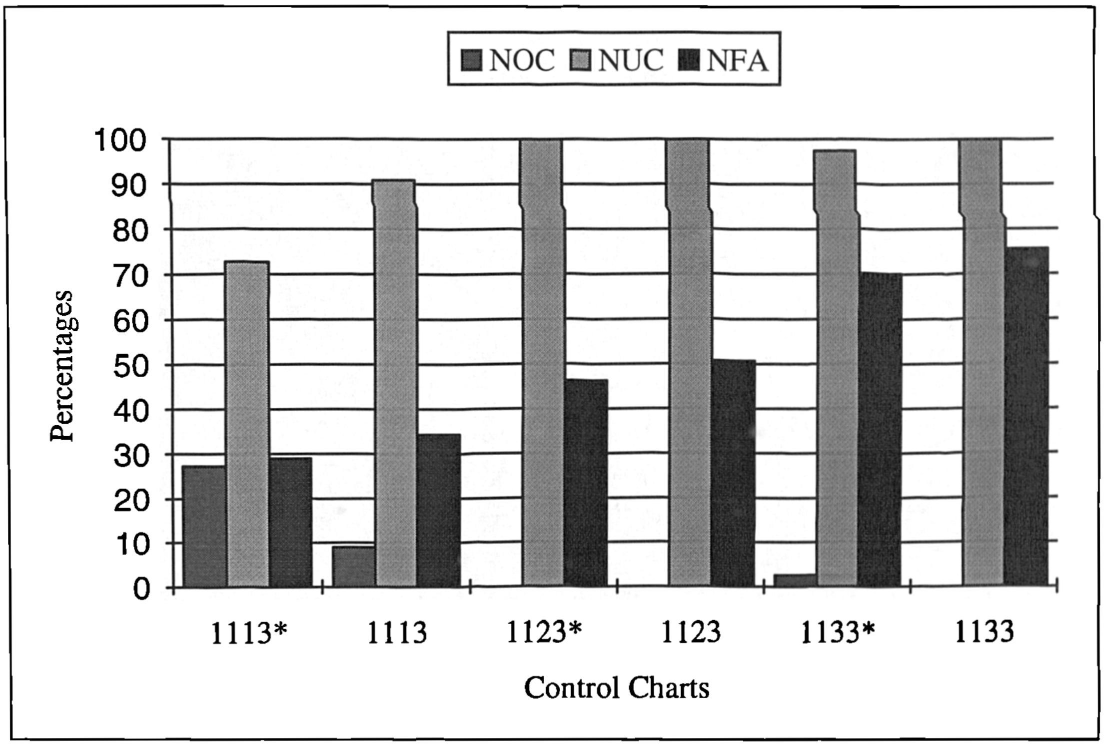

Figure 7.30 Effects of different data sets for off-line PCorrA charts, manipulating *Trtin*

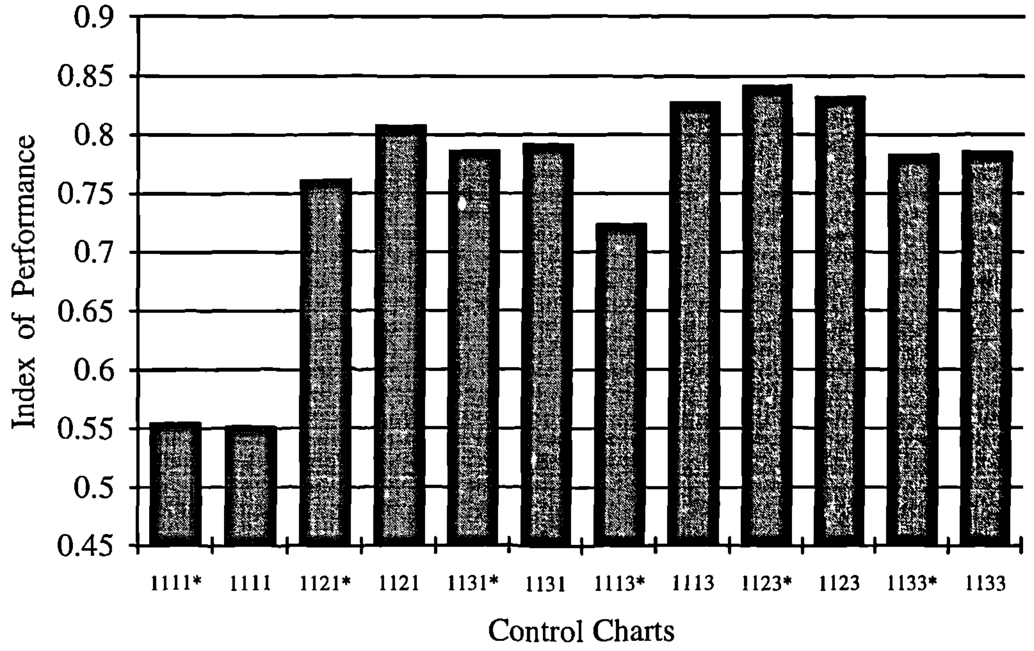

Figure 7.31 Comparison of *IP* between histories data

7.5.5 SPC Charts that being used in the Analysis

Figures 7.32 to 7.34 show examples of SPC charts that have been discussed in this chapter. Figure 7.32 and figure 7.33 show the reactants input temperature (*Trtin*) and the coolant temperature (*Tjin*) for the PCA based ShewA chart using on-line calculated control limits, manipulating only two input variables (*i.e.* run 2212). For the first 50 samples, both manipulated variables ShewA charts control limits were kept constant. Then as the limits were updated on-line based on the calculated PCA *correlation* coefficients, significant changes were observed after the 50'th sample in the ShewA charts for *Trtin* and *Tjin*. The limits on both charts started to widen, rendering the control charts more conservative. Due to this, we have several out-of-control points in the quality variable as shown in the ShewA chart of figure 7.34.

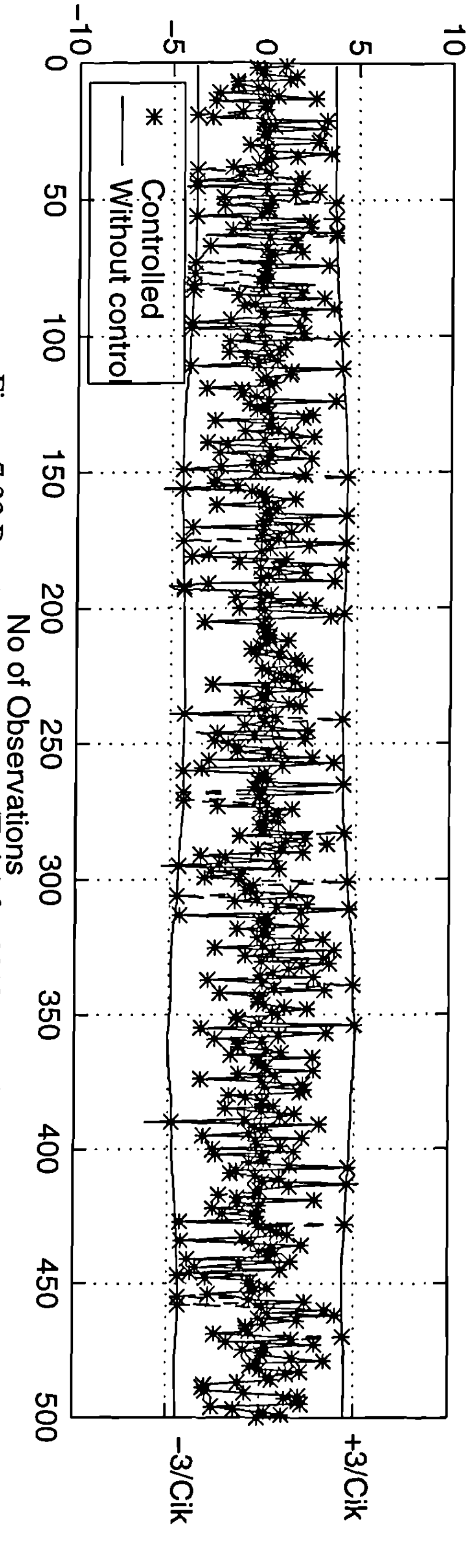

Figure 7.32 Reactants temperature (Trtin) for 2212 control scheme

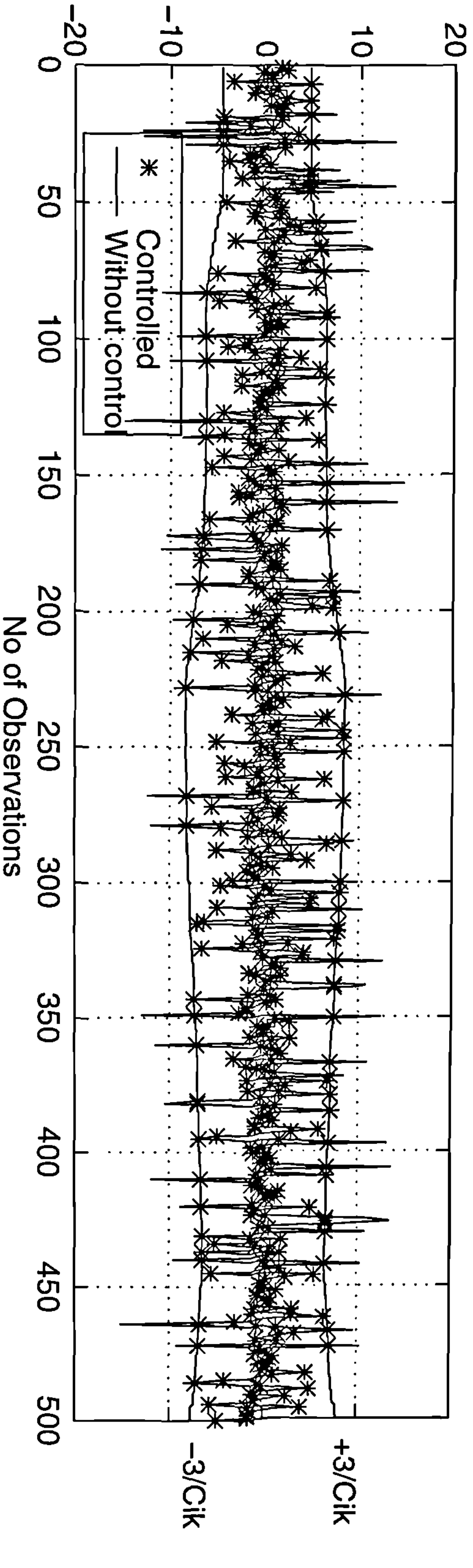

Figure 7.33 Coolant temperature (Tjin) for 2212 control scheme

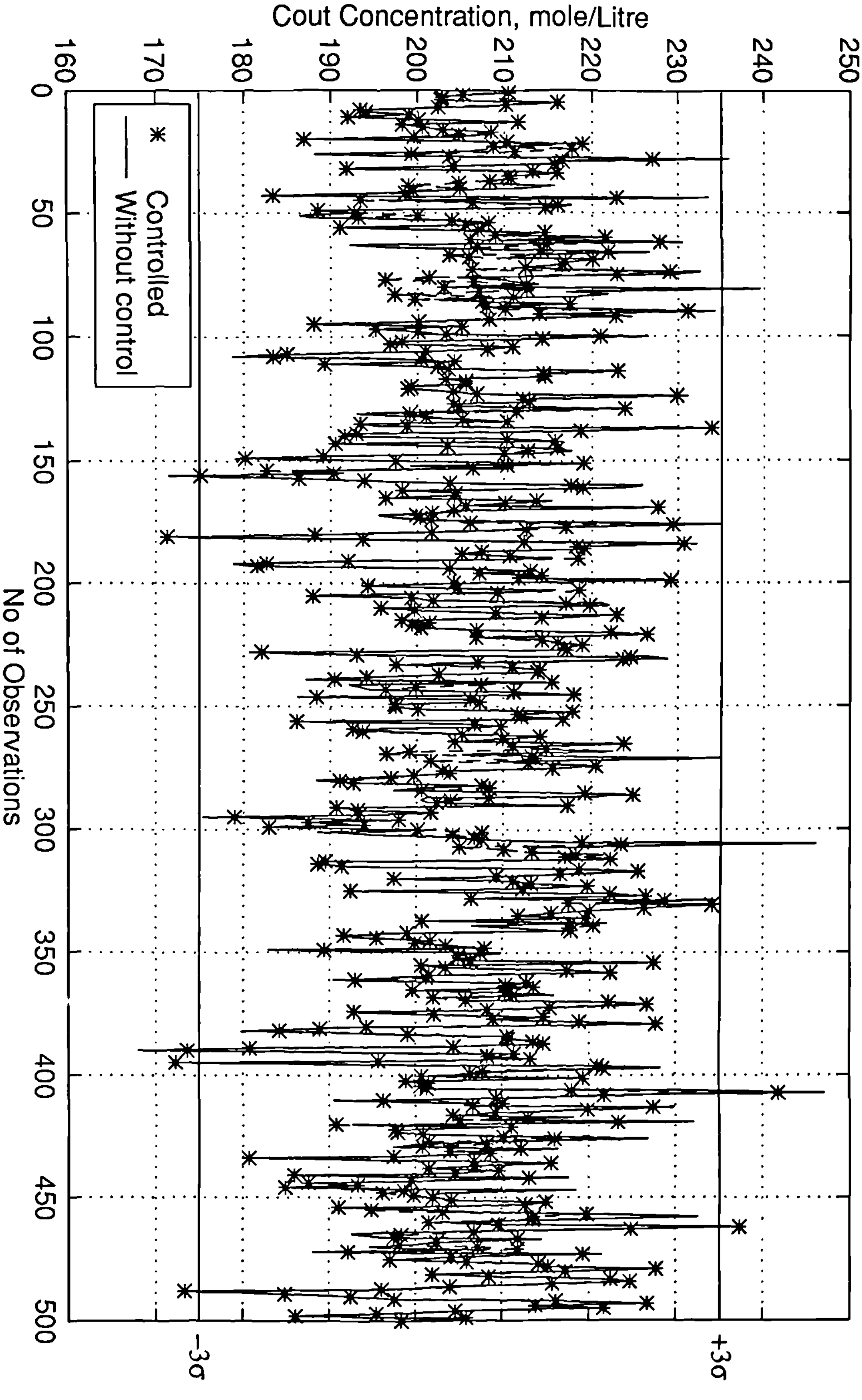

Figure 7.34 Quality variable concentration (Cout) for 2212 control scheme

7.5.6 Comparison between Active SPC and APC

In this section, we will compare the performances of Active SPC and the APC method. With the Active SPC methods, only those charts that do not allow out-of-control points by manipulating *Trtin* will be used. Figure 7.35 shows the percentages of control actions for both methods. The APC method recorded the highest percentage of control actions followed by the EWMA charts, ShewAW charts and lastly the ShewA charts. The high percentages of control action in EWMA charts is expected because it has the smallest control limits compared to the rest of the SPC charts used in this study. The ShewAW chart has high percentage of control actions because of additional control rules. From figure 7.35 we see that the 1213 chart have the least percentages of control action followed by charts 1123 and 2123.

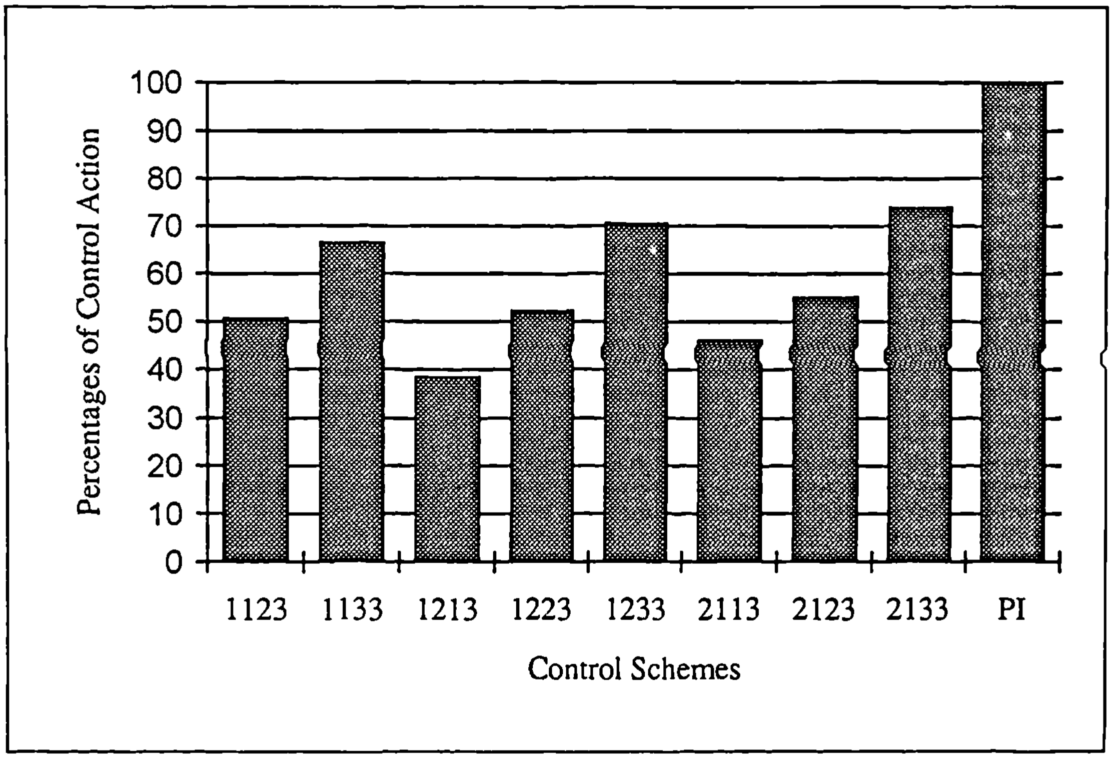

Figure 7.35 Comparison of control action percentages between Active SPC and APC

Regarding the percentages of control action shown in figure 7.35, it is clear that APC is the most effective method for maintaining set point because it is always constantly taking control action to make sure that the process is on target. However, since APC systems are continually making physical adjustments to a process, there can be increased in wear on final control elements. This, together with the maintenance requirements of the control equipment itself, can substantially increase the costs.

Figure 7.36 shows the amount of energy used in relation to the nominal level. The nominal value is the minimum energy needed to maintain the process in control when there are no out-of-control situations in the process. The nominal level value is calculated by summing the deviations of the manipulated variable, (*Trtin*) from the *mean* reactant temperature which is 450 Kelvin. In the case of our CSTR example, the utilisation of *Trtin* as the manipulated variable incurs energy costs. For example, higher costs are incurred if *Trtin* has to be raised to satisfy control objectives. During the calculation of this base line value, the process is in statistical control.

To evaluate the effectiveness of the Active SPC and the APC method, white noise was added again to the inputs to simulate out-of-control situations. Then, the sum of reactant temperature deviation from the *mean* was calculated. From this new value, we take away the nominal value of energy used. The results are illustrated in figure 7.36. From the figure, we see that the EWMA based chart (xx33) utilised the least energy to control the process. The main reason for this is because it has the smallest band of control limits compared to ShewAW and ShewA charts. On the other hand, the APC method consumed the highest amount of energy in maintaining the process at set-point.

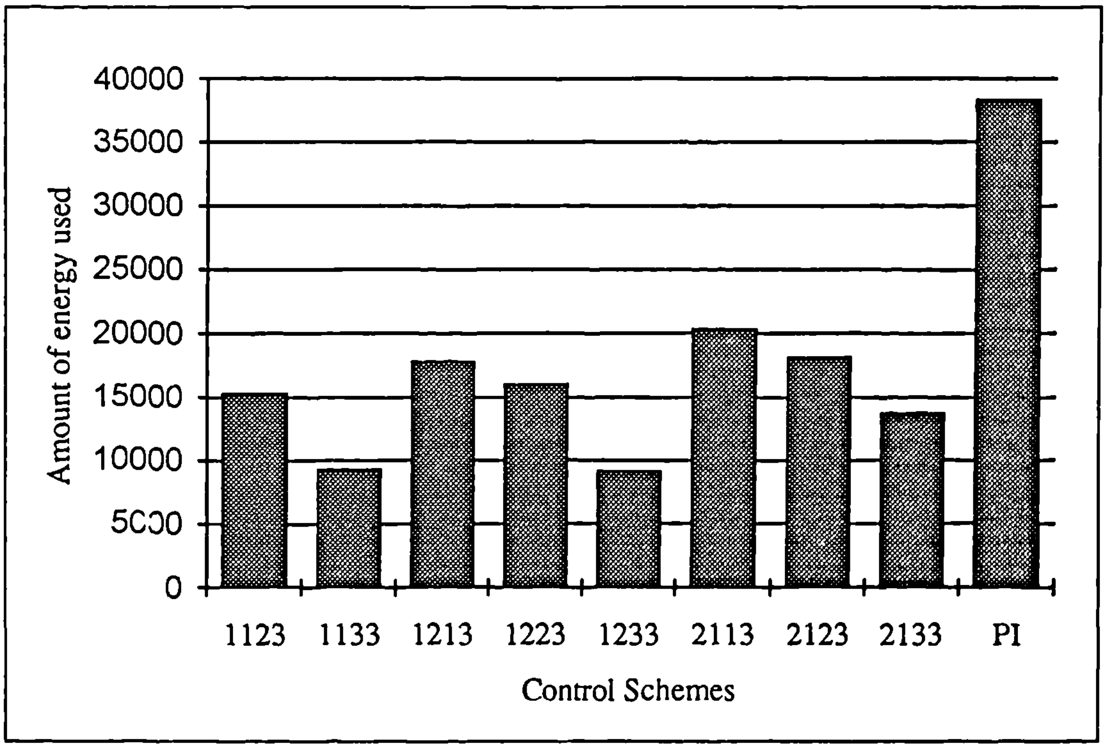

Figure 7.36 The amount of energy used in relation to nominal level

From the above result we can conclude that Active SPC using EWMA control charts can out perform the APC method and the rest of the SPC charts. Although the percentages of control action in EWMA charts is higher compared to the ShewAW and

ShewA charts, savings can be gained through less energy utilisation for process control. In addition, based on the performances of Active SPC schemes in section 7.5.1 and 7.52, we observed that there are no out-of-control situations for all PCorrA based charts and PCA based EWMA chart for the off-line updating control limits schemes. Moreover, as discussed in Chapter Two, the EWMA statistics can be used for forecasting future data. Thus it has an added advantage compared to the other two types of control charts used in this study.

7.6 Summary

This chapter discussed the results of applying Active SPC methods on the simulated CSTR. A procedure for determining an appropriate sampling time is given and elaborated, and it was shown that the procedure leads to data that fulfils normality assumptions. Off-line multivariate analyses were performed on historical data to select relevant manipulated variables for controlling the process. From there, we control the process by using several Active SPC methods. An "Index of Performance" (*IP*) was also defined to enable assessment of the overall performance of the Active SPC schemes. We also discussed the effects of the historical data used to provide the initial values of C_{ik} on the performance of the Active SPC techniques. Lastly, we compare the performance of these Active SPC methods with the APC strategy, specifically *PI* feedback control.

Based on the results, we found out that reducing the number of manipulated variables will increase the *IP* for the PCorrA based ShewA and ShewAW charts. The same phenomenon is observed for the off-line calculated limits for the PCA based ShewA and ShewAW charts. However, the converse is true for on-line calculated limits for PCA based charts. For the EWMA control charts, the *IP* decreases as we reduce the number of manipulated variables. Although we have this decreasing trend in *IPs* for the EWMA charts, there are no out-of-control incidents for the PCorrA based charts and PCA off-line calculated control limits. They only differ in the number of false alarms, which increases as we reduce the number of manipulated variables. It indicates that these control schemes remain sensitive to changes in the process. Since the NOC is zero and the NFA is increasing, the *IP* for these manipulation strategies will show a downward trend because we penalise the number of false alarms in calculating the *IP*.

All charts with limits calculated on-line using PCA shows a decreasing trend in *IP* as we reduce the number of manipulated variables. The number of out-of-control points are generally larger than when off-line calculated control limits were used. For off-line cases, the data used to calculate the limits do not have any out-of-control points, but the converse may not be true. Thus, when off-line calculated limits are used, the chosen manipulated variables are sensitive to changes in the process but when we switch to the on-line updating scheme, the chosen manipulated variables may no longer be sensitive.

Comparing the control performances of the types of control charts used in this study, we found that EWMA charts are the most sensitive, followed by the ShewAW and lastly the ShewA charts. The main reason for this sensitive nature of the EWMA charts is that they have tighter control limits compared to the other two charts, while, the ShewAW charts have an added control rule compared to ShewA chart. Nevertheless, the best overall *IP* for the Active SPC schemes for this study is run 1213 (PCA based ShewA with off-line calculated limits).

This study also highlighted the importance of the data used to determine the state of the process. Only PCorrA based charts were considered because they appear to give the most consistent results. From our comparison of two sets of historical data that differ only in standard deviations and hence C_{ik} values, we observed some obvious trends in the results for the PCorrA based SPC charts. Whenever we want a tighter control on the process (*e.g.* NOC = 0) it is preferably to choose the data sets that have smaller values of standard deviations. On the other hand, by selecting a data set with smaller standard deviations, we may risk a higher number of false alarms. If we want to avoid excessive incidents of false alarms or if control actions may incur cost, we would choose data sets that contain larger values of standard deviations. However, there may be several out-of-control situations.

Comparing the Active SPC and the APC method, we conclude that the EWMA based charts generally perform better. In our example, it utilised the least energy in relation to the nominal level value even though it makes more control actions compared to the ShewA and ShewAW charts. However, the EWMA charts has an added advantage compared to the other two charts because it can be utilised to forecast the data for a drifting process. Thus, the EWMA charts may be viewed as a better alternative to either the ShewA and ShewAW control charts because it can be utilised in both SPC and APC methodologies.

Chapter 8

Conclusions and Future Research

This chapter summarises the important findings in this thesis and suggests new areas where this research can be extended. The section on future research has two major thrusts (i) discussion on the enhancements that can be attempted, and (ii) exploration of new areas pertaining to Active SPC schemes.

8.1 Conclusions

This work follows the preliminary studies of Efthimiadu *et al.* (1991), Efthimiadu *et al.* (1992), Efthimiadu *et al.* (1993) who attempted to devise a new, pro-active approach to SPC of continuous process. Detailed evaluations and some modifications were performed in the current work to assess the applicability of the various Active SPC strategies that might evolve.

Procedures for designing and implementing automatic Statistical Process Control (SPC) charts were described. Traditional charts are still used for monitoring purposes. In addition to the quality variable, those inputs identified using Partial Correlation Analysis (PCorrA) or Principal Components Analysis (PCA) that can potentially cause process upsets are also monitored. The limits for the monitoring charts and more significantly, the manipulation rules to keep the process under statistical control arise naturally as part of the analysis. Since the potential causes of process deviations have been pre-determined, abnormal variations in the input variables can be corrected automatically before they affect output quality. Thus, unlike traditional SPC strategies, control is achieved in an anticipatory manner. The need for expensive and time consuming experimentation after out-of-control incidences is therefore reduced.

Several control strategies were applied to a simulated CSTR process to evaluate the performances of various Active SPC strategies. These arise by using two types of multivariate analysis, PCorrA and PCA; use of different types of traditional SPC charts, (*e.g.* ShewA and ShewAW for individual observation and EWMA charts), using off-line and on-line correlation updating; and finally the use of different manipulation strategies to control product quality (*e.g.* manipulating all input variables, manipulating two input variables and lastly manipulating only a single input variable). The effects of historical data on the performance of the Active SPC were also discussed and compared. Lastly, this work compared the Active SPC method with the APC feedback *Proportional Integral* (*PI*) controller. The results obtained are promising and point towards potential applications of Active SPC to continuous process engineering systems. A summary of some important findings are listed below:

(i) The work detailed in this thesis shows that it is possible to design and successfully implement Active SPC strategies using the correlations calculated via either PCA or PCorrA. In all cases, performances were better than the uncontrolled system, but strategies involving EWMA charts for monitoring and control provided the best performances.

(ii) The historical data used in the design of Active SPC methods plays an important role in determining the performance of the respective schemes. It determines the efficiencies of manipulating the process when fixed control rules are used and for the first 50 observations in schemes where control rules are updated on-line. Only PCorrA based charts were used in this study because they gave consistent results through out the analyses in section 7.5.1 to 7.5.3. From the comparison of two sets of historical data that differ in standard deviations and the C_{ik} values, we observed some trends in the results of PCorrA based SPC charts. If we want a control performance with a low number of out-of-control points and can tolerate high percentage of false alarms, historical data sets with small standard deviations are preferable. In other situations, if we are concerned with the cost of taking control actions on the process, then historical data sets with higher values of standard deviations can be chosen. For this case, however, we may have several out-of-control situations.

(iii) Comparing the performance of charts with limits calculated off-line using both multivariate analyses employed in this study, we found that the performance of PCA based charts are more sensitive than the PCorrA based charts. This can be seen by the higher correlation (C_{ik}) coefficients between the inputs reactant temperature (*Trtin*) and cooling medium temperature (*Tjin*) and the quality variable as shown in table 7.3. Thus, when we compare the performance of charts with off-line calculated control limits, the PCA based charts always recorded higher percentages of false alarms. It indicates that the PCA based charts always take more control action on the process compared to the PCorrA based charts. If both multivariate based charts have different percentages of out-of-control (NOC) points in the process, the PCA based charts will always have a lower percentage of NOC due to its sensitivity. Thus its index of performance (*IP*) will be higher than PCorrA based charts for such cases. The index of performance (*IP*) is a measure, designed to identify which configurations of Active SPC schemes can provide better overall control performance. It penalises the percentage of out-of-control points as well as the percentage of false alarms. However, when both multivariate charts have the same percentage of NOC, the *IP* of PCorrA based charts will be higher than PCA based charts. This is due to the former having a more conservative approach in taking control policy and, thus the PCorrA based chart will have lower percentage of false alarms.

(iv) For on-line calculated control limits, only one trend was observed when we compared the performance of both multivariate methods. Here, PCorrA based charts are superior than the PCA based charts. If we compare the charts and the strategy used, generally the *IP* of the PCorrA based charts are higher than those of the PCA based charts. This is because the charts with PCorrA on-line calculated limits show more sensitivity towards changing conditions in the process. On the other hand the sensitivity of PCA on-line calculated based charts started to degrade as the process is being controlled. This can be explained by comparing the $\tilde{C}_{ik}$ values of on-line calculated Shewhart Action limits for both multivariate methods in figure 7.17, and the $\tilde{C}_{ik}$ for PCA based Shewhart action and warning limits (ShewAW) and Exponential Weight Moving Average (EWMA) charts in figure 7.24 and figure 7.28. When the $\tilde{C}_{ik}$ in PCA based charts start to decrease, a more conservative approach will be adopted on the process and the systems tolerates larger deviations.

(v) Generally when there is good statistical control, *i.e.* the percentage of out-of-control points decrease, and the percentages of false alarm will increase. This is clearly illustrated when the type of control chart is changed from ShewA to ShewAW and lastly to EWMA. This is because the EWMA charts have the smallest band of control limits, while ShewAW charts have additional control rules. Although, out-of-control situations will be reduced the false alarm rate is increased. From the analyses of section 7.5.3, we observed that the *IP* for all ShewAW charts are greater than 0.75, while the *IP* of all EWMA charts are greater than 0.72. Nevertheless there are no out-of-control points when using the EWMA charts compared to the ShewAW charts, except when their control limits were updated on-line using the PCA method. On this basis, we can conclude that the EWMA control chart is a good candidate for future use with Active SPC schemes.

(vi) Different manipulation strategies were applied to account for practical cases where not all inputs could be manipulated. Comparing the performance of different manipulation strategies (*i.e.* configurations of manipulated variables) for controlling the product quality, we found out that whenever the input variables and the given manipulated variables are sensitive to the changes in the process, that particular manipulation strategy will perform better in terms of the Index of Performance (*IP*). For all PCorrA based ShewA and ShewAW charts and ShewA and ShewAW charts using PCA off-line calculated control limits, we discovered that when changing the manipulation strategies from all input variables to two input variables and finally to a single input variable, the *IP* will increase. This is because the selected manipulated variables that we chose for two and single manipulated variables have the highest C_{ik} values compared to the rest of the input variables. For these manipulation strategies, we chose the temperature of input reactants (*Trtin*) and temperature of cooling medium (*Tjin*) as the manipulated variables for the two manipulated variable control scheme, while *Trtin* was chosen for the single manipulated input control strategy. From table 7.3 we see that *Trtin* has the highest C_{ik} values followed by *Tjin*. However, for on-line calculated control limits, the *IP* of PCA based charts are degraded as we reduce the number of manipulated variables to control the quality variables. For these cases, the manipulated variables that we have chosen, especially *Trtin* and *Tjin*, are no longer sensitive to changes in the process. As the process is controlled, the value of $\tilde{C}_{ik}$ for *Trtin* and *Tjin* started to fall. This phenomenon can be seen in figures 7.19, 7.24 and 7.28. Although the PCorrA based EWMA charts showed a downward trend in *IP*, there were no out-of-control incidences. They only differ because of the number of false alarms which showed an increasing trend when the number of manipulated variables is reduced. It shows that these on-line updated PCorrA based EWMA charts are still sensitive to changes in the process.

(vii) Comparing the Active SPC strategies method and the APC method in this study, we found that some of the Active SPC strategies is superior to the APC *Proportional Integral* (*PI*) controller. Some of the Active SPC strategies achieve control with fewer control actions on the process compared to the APC method. The main reason for this is that Active SPC methods do not penalise any disturbances that are inherent to the process. Meanwhile the APC method always takes control action on the process whenever the controlled variable deviates away from set-point. In terms of the amount of energy used in relation to the nominal level required to maintain the process in control, the Active SPC control schemes considered for this comparison, utilised the least amount of energy compared to the APC method. This is because whenever a disturbance is perceived to be out-of-control in the Active SPC, the control action is to bring the disturbance back to its allowable limits. Since the EWMA control charts have the smallest band of control limits compared to the other two control charts used in this study, the EWMA charts consume the least energy. Although the percentages of control actions in EWMA charts are greater than the other two SPC charts, this can be compensated by the energy saved. Moreover, as mentioned in chapter two, the EWMA charts have an added advantage compared to the other two charts because apart from monitoring and giving an indication that the process is out-of-control, it also can be used to forecast future data for a drifting process. Thus, it will be a very good candidate for future use in Active SPC strategies.

8.2 Future Research

The principal objective of this research is to study in detail the potential of Active SPC schemes and to develop an Active SPC chart which can monitor and automatically adjust the process when out-of-control situations occurred. This has been made possible by monitoring the quality and the input variables, and manipulating the input variables. The techniques retain the SPC chart characteristic of non-intervention when the process is in a state of statistical control. However, there are still some intriguing questions about the behaviour of some of the Active SPC methods considered that should be answered.

8.2.1 Enhancements of Current Research

Section 6.3.1 of this thesis showed how the ShewA, ShewAW and EWMA control charts could be adapted to form Active SPC methodologies. The methodologies are designed via the use of correlational procedures based on either the Principal Components Analysis (PCA) or Partial Correlation Analysis (PCorrA). These correlations (C_{ik}) are used to translate the control limits on the quality variable to the control limits on input variables. Examination of these strategies showed that in general, the false alarm rate increases when the assignable causes were eliminated. To make the control strategy more efficient to use in practice, this effect should be reduced or eliminated. The following aspects should be considered in future work.

(i) The control performances of PCA based charts with fixed control rules are better compared to its on-line updated schemes. The off-line schemes have less number of out-of-control points compared to the on-line schemes. Fixed control rules are derived off-line using statistically in-control data. With on-line updating schemes the control rules are determined from data collected from the process as it is being controlled. On the basis of the results in sections 7.5.2 and 7.5.3 we observed that once we opt for a PCA on-line control limit updating scheme, the $\tilde{C}_{ik}$ values for the input variables started to decrease. As a result, the performance of charts updated on-line using PCA becomes more conservative (*i.e.* the control limits become wider). As a consequence, the percentages of out-of-control situations in these schemes are greater than its off-line counterparts. A study should be performed to look into the properties of PCA, to explain why $\tilde{C}_{ik}$ decreases when out-of-control points occur.

(ii) The use of PCorrA with Active SPC methodologies yielded good performances. The on-line updated charts are generally better compared to the off-line designed charts because they reduced the percentage of out-of-control situations. These feature were not observed with the PCA based methods. Thus, it appears that PCorrA is a good tool for designing Active SPC strategies. When on-line updated control rules are being used, it seems that the PCorrA based methods can cope with the changes in the process by tightening the control limits of the input variables. As a consequence, the on-line calculated control limits for PCorrA based charts become more sensitive. To further enhance this study, we should therefore look at the properties of PCorrA to investigate why this multivariate technique tightens up the control limits (increases $\tilde{C}_{ik}$ values) when the on-line updating scheme is used.

8.2.2 New Areas of Investigations

There are also several new areas that merit investigation to test the capability of Active SPC charts to remove assignable causes. It is suggested that the most important areas to explore, in terms of providing Active SPC schemes that would be most beneficial to the process industries, are as follows:

(i) The discussion in this thesis has centred on the implementation of Active SPC to multiple input and single output (MISO) and single input and single output (SISO) systems. Almost all modern process industries are multiple input and multiple output (MIMO) systems. Therefore, an obvious area of development of Active SPC is to extend it to the MIMO case. In order to test the capability of this new procedure, a holistic approach to implement Active SPC would be needed so that the problems of constraints could also be addressed, *e.g.* the excess reactants from the CSTR can be recycled or prescribed control limits on some of the output variables.

(ii) In chapter 4 we see that our simulated process has time-delays associated with various input-output relationships. Fortunately, these were insignificant. There are many processes where time-delays are large and may be time varying. It would be a challenging problem to devise an Active SPC scheme that could accommodate these dynamics characteristics.

(iii) The Active SPC strategy that we proposed resembles feedforward control. It can achieve perfect control when all the disturbances in the process can be accounted for. However, when we have unmeasured disturbances affecting the process, the proposed Active SPC may fail to provide effective control. Thus, to counter the effects of such disturbances, the incorporation of some feedback method is necessary. Future work on Active SPC should address this problem.

References

Box, G.E.P., and Jenkins, G.M. (1962) "Some Statistical Aspects of Adaptive Optimization and Control." *Journal of the Royal Statistical Society*, (Series B), 24(2), pp 297-331.

Box, G.E.P., and Jenkins, G.M. (1976) *Time Series Analysis: Forecasting and Control*, Holden Day, San Francisco, California.

Box, G.E.P., and Kramer, T. (1992) "Statistical Process Monitoring and Feedback Adjustment - A discussion ." *Technometrics*, 34(3), August, pp 251-267.

Box, G.E.P. (1993) "Quality Improvement-The New Industrial Revolution", *International Statistical Review*, 61, pp 3-19.

Capilla, C., Ferrer, A., Romero, R., and Zunica, L. (1995) "Integrating SPC and EPC to Control a Continuous Industrial Process." *Conference on Control of Industrial Processes: Synthesising the Statistical and Engineering Approaches*, Hertfordshire, September 19-20, pp 77-84.

Cohen, G.H. and Coon, G.A. (1953) "Theoretical Considerations of Retarded Control." *ASME Transactions*, 75, pp 827-834.

Deming, W.E. (1982) *Out of Crisis*, Massachusetts Institute of Technology, Center for Advanced Engineering Study, Cambridge University Press, Massachusetts.

DeVor, R.E., Chang, T.S., and Sutherland, J.W. (1992) *Statistical Quality Design and Control-Contemporary Concepts and Methods*, Maxwell Macmillan International Editions, Singapore.

Efthimiadu, I., and Tham, M.T. (1991) "On-Line Statistical Process Control of Chemical Processes." *IEE Conference Control '91*, March 25-28, Heriot-Watt University, Edinburgh.

Efthimiadu, I., Kaluluma, J., and Tham, M.T. (1991) "On-Line Statistical Process Control." Internal Report: DCPE/EFT1/91, December, Department of Chemical and Process Engineering, University Newcastle upon Tyne.

Efthimiadu, I., Kaluluma, J., and Tham, M.T. (1992) "An SPC Strategy for Chemical Processes." *IChemE Conference-Advances in Process Control*, September 23-24, York.

Efthimiadu, I., Tham, M.T., and Willis, M.J. (1993) "Engineering Control and Product Quality Assurance in the Process Industries." *Quality and its Applications*, 1993, September 1-3, University Newcastle upon Tyne, pp 593-598.

English, J.R., and Case, K.E. (1990) "Control Charts Applied as Filtering Devices Within a Feedback Control Loop." *IIE Transactions*, (Institute of Industrial Engineers), 22, September, pp 255-269.

Faltin, F.W., and Tucker, W.T. (1991) "On-Line Quality Control for the Factory of the 1990's and Beyond." *Statistical Process Control in Automated Manufacturing*, (edited by Keats, J.B, and Montgomery, D.C.) Marcel Dekker, New York. pp 331-354.

Faltin, F.W., Hahn, G.J., Tucker, W.T., and Vander Weil, S.A. (1993) "Algorithmic Statistical Process Control: Some Practical Observations." *International Statistical Review*, 61, pp 67-80.

Gibra, I.N. (1975) "Recent Developments in Control Chart Techniques." *Journal of Quality Technology*, 7(4), October, pp 183-192.

Graybill, F.A. (1976) *Theory and Application of Linear Model*, Duxbury Press, Wadsworth Publishing Co, Massachusetts.

Hotteling, H. (1933) "Analysis of a Complex of Statistical Variables into Principal Components." *Journal of Education Psychology*, 24(6), pp 417-441.

Hunter, J.S. (1986) "The Exponentially Weighted Moving Average", *Journal of Quality Technology*, 18(4), October, pp 203-210.

Ibrahim, K.A., and Tham, M.T. (1995) "Towards Active Statistical Process Control." *Proceedings of American Control Conference*, Seattle, June 21-23, vol. 3, pp 2234-2238.

Jaehn, A.H. (1989) "Zone Control Charts Find New Application." *43rd Annual Quality Congress Transactions,* American Society for Quality Control, Milwaukee, Wisconsin, pp 890-895.

Jackson, J.E. (1980) "Principal Components and Factor Analysis: Part I-Principal Components." *Journal of Quality,* 12(4), October, pp 210-213.

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

Juran, J.M., Gryna, F.M., and Bingham, R.S. (1974) *Quality Control Handbook,* McGraw-Hill, New York.

Lucas, J.M., and Crossier, R.B. (1982) "Robust CUSUM: A Robustness Study for CUSUM Quality Control Schemes." *Communications in Statistics-Theory and Methods,* 11, pp 2669-2687.

MacGregor, J.F. (1988) "On-Line Statistical Process Control." *Chemical Engineering Progress,* October, pp 21-31.

MacGregor, J.F. (1990) "A Different View of the Funnel Experiment." *Journal Quality Technology,* 22(4), October, pp 255-259.

Montgomery, D.C. (1991) *Introduction to Statistical Quality Control,* John Wiley and Sons, Singapore.

Montgomery, D.C., Bert Keats, J., Runger, G.C., and Messina, W.S. (1994) "Integrating Statistical Process Control and Engineering Process Control." *Journal of Quality Technology,* 26(2), April, pp 79-87.

Oakland, J.S., and Followell, R.F. (1990) *Statistical Process Control,* Heunemann Newnes, Oxford.

Pearson, K. (1901) "On Lines and Planes of Closest Fit to Systems of Points in Space." *Phil. Mag.,* Series 6, 2(11), pp 559-572.

Press, W.H. (1989) *Numerical Recipes in Pascal: The Art of Scientific Computing,* Cambridge University Press, Cambridge, New York.

Rawlings, J.O. (1988) *Applied Regression Analysis: A Research Tool'*, Wadsworth and Brooks, California.

Rhinehart, R.R. (1992) "CUSUM Type on-line Filter." *Process Control and Quality*, 2, pp 169-175.

Rhinehart, R.R. (1995) "A Watchdog for Controller Performing Monitoring." *Proceedings of American Control Conference*, Seattle, June 21-23, vol. 3, pp 2239-2240.

Saniga, E.M., and Shirland, L.E. (1977) "Quality Control in Practice: A Survey." *Quality Progress*, May, pp 30-33.

Shewhart, W.A. (1931) "Economic Control of Quality of Manufactured Products", D. Van Nostrand, Princeton, New Jersey. Reprinted by the *American Society for Quality Control*, Milwaukee, Wisconsin.

Stephanopoulos, G. (1984) *Chemical Process Control: An Introduction to Theory and Practice*, Prentice Hall, New Jersey.

Taguchi, G., Elsayed, A., and Hsiang, T. (1989) *Quality Engineering in Production System*, McGraw-Hill, New York.

Tham, M.T. (1995) "Principal Components Analysis." Unpublished Notes.

Tucker, W.T., Faltin, F.W., and Vander Weil, S.A. (1993) "Algorithmic Statistical Process Control: an Elaboration." *Technometrics*, 35(4), November, pp 363-375.

Thomson, M., and Twigg, P.M. (1994) "Statistical Process Control Methods in the Supervision and Selective Application of PID Control." *Control '94*, 21-24 March. Conference Publication No. 389, pp 612-617.

Vance, L. (1983) "A Bibliography of Statistical Quality Control Chart Techniques, 1970-1980." *Journal of Quality Technology*, 15(2), April, pp 59-62.

Vander Weil, S.A., Tucker, W.T., Faltin, F.W., and Doganaksoy, N. (1992) "Algorithmic Statistical Process Control: Concepts and Application." *Technometrics*, 34(3), November, pp 286-297.

Vardeman, S., and Cornell, J.A. (1987) "A Partial Inventory of Statistical Literature on Quality and Productivity through 1985." *Journal of Quality Technology*, 19(2), April, pp 90-97.

Wetherill, B.G., and Brown, D.W. (1991) *Statistical Process Control, Theory and Practice*, Chapman and Hall, London.

Wold, S., Esbensen, K., and Geladi, P. (1987) "Principal Component Analysis." *Chemometrics and Intelligent Laboratory Systems*, 2, pp 37-52.

Ziegler, J.G., and Nichols, N.B. (1942). "Optimum Settings for Automatic Controllers." *ASME Transactions*, 64, pp 759-768.

Appendix A

INVESTIGATION INTO ON-LINE SPC

Date : February 5, 1996 Monday   Time : 11.00 am
Name of Historical Data File: jan31.dat $\Delta t =$ 100 sec
Maximum Number of Observations : 500

ShewA	Maximum Number of NFA: 489	Maximum Number of NOC : 11
ShewAW	Maximum Number of NFA: 481	Maximum Number of NOC : 19
EWMA	Maximum Number of NFA: 462	Maximum Number of NOC : 38

Runs	% NOC	% NUC	% NFA	*IP*	EC	# CA
1111	45.45	54.55	34.36	0.5511		
1112	27.27	72.73	34.36	0.7122		
1113	9.09	90.91	34.36	0.8266	16625.5488	1734
1121	5.26	94.74	50.73	0.8075		
1122	00.00	100.00	50.73	0.8317		
1123	00.00	100.00	50.73	0.8317	15285.9434	2529
1131	00.00	100.00	65.15	0.8028		
1132	00.00	100.00	68.83	0.7962		
1133	00.00	100.00	75.54	0.7848	9276.7539	3320
1211	27.27	72.73	36.61	0.7079		
1212	18.18	81.82	36.61	0.7687		
1213	00.00	100.00	36.61	0.8660	17829.1289	1913
1221	5.26	94.74	52.18	0.8045		
1222	00.00	100.00	52.18	0.8286		
1223	00.00	100.00	52.18	0.8286	15961.8633	2612
1231	00.00	100.00	66.23	0.8008		
1232	00.00	100.00	66.45	0.8004		
1233	00.00	100.00	76.19	0.7838	9104.2441	3523
2111	18.18	81.82	41.31	0.7584		
2112	9.09	90.91	43.35	0.8047		
2113	00.00	100.00	45.40	0.8439	20348.8359	2302
2121	00.00	100.00	55.93	0.8207		
2122	00.00	100.00	55.72	0.8211		
2123	00.00	100.00	55.09	0.8224	18124.6328	2753
2131	00.00	100.00	70.35	0.7935		
2132	00.00	100.00	75.32	0.7852		
2133	00.00	100.00	79.65	0.7783	13732.9688	3690
2211	45.45	54.55	30.06	0.5537		
2212	54.55	45.45	25.56	0.4360		
2213	63.64	36.36	17.18	0.2453	13879.6895	904
2221	5.26	94.74	48.02	0.8134		
2222	15.79	84.21	38.46	0.7789		
2223	21.05	89.50	29.11	0.7679	14114.8340	1451
2231	5.26	94.74	62.77	0.7840		
2232	13.16	86.84	61.04	0.7491		
2233	18.42	81.58	56.93	0.7280	17403.6152	2352

NOC : Out-of-control points   NUC : Controlled Points   NFA : Number of False Alarm
IP : Index of Performance   CA : Control Actions   EC : Energy consumption

First Digit	(1) Off-line	(2) On-line	
Second Digit	(1) PCorrA	(2) PCA	
Third Digit	(1) Shewhart Action	(2) Shewhart Action & Warning	(3) EWMA
Fourth Digit	(1) All Variables	(2) *Trtin* & *Tjin* only	(3) *Trtin* only

Appendix B

INVESTIGATION INTO ON-LINE SPC

Date : January 24, 1996 Wednesday   Time : 11.10 am
Name of Historical Data File: jan29.dat $\Delta t =$ 100 sec
Maximum Number of Observations : 500
ShewA   Maximum Number of NFA: 489   Maximum Number of NOC : 11
ShewAW   Maximum Number of NFA: 481   Maximum Number of NOC : 19
EWMA   Maximum Number of NFA: 462   Maximum Number of NOC : 38

Runs	% NOC	% NUC	% NFA	*IP*
1111	45.45	54.55	29.04	0.5544
1112	36.36	63.64	29.04	0.6471
1113	27.27	72.73	29.04	0.7233
1121	15.79	84.21	46.57	0.7616
1122	10.53	89.47	46.36	0.7906
1123	00.00	100.00	46.36	0.8416
1131	5.26	94.74	61.69	0.7860
1132	5.26	94.74	65.80	0.7790
1133	2.63	97.37	70.13	0.7828

NOC : Out-of-control points   NUC : Controlled Points
NFA : Number of False Alarm   *IP* : Index of Performance

First Digit	. (1) Off-line	(2) On-line	
Second Digit	(1) PCorrA	(2) PCA	
Third Digit	(1) Shewhart Action	(2) Shewhart Action & Warning	(3) EWMA
Fourth Digit	(1) All Variables	(2) *Trtin* & *Tjin* only	(3) *Trtin* only

Table B1 C_{ik} based on PCA and PCorrA

	var	*Trtin*	*Tjin*	*Ain*	*Bin*	*F*	*Fj*
PCA	*Cout*	0.8073	0.6487	0.1066	0.2215	0.1666	-0.0135
PCorrA	*Cout*	0.7832	0.4521	0.3069	0.2840	0.2284	-0.1913

Appendix C

Equation (2.5) is given by:

$$\hat{x}_{t+1} = \theta x_t + (1-\theta)\hat{x}_t \tag{2.5}$$

Time-shifting one step back, equation (2.5) becomes

$$\hat{x}_t = \theta x_{t-1} + (1-\theta)\hat{x}_{t-1} \tag{c.1}$$

Substituting equation (c.1) into equation (2.5) gives the EWMA

$$\hat{x}_{t+1} = \theta x_t + \theta(1-\theta)x_{t-1} + (1-\theta)^2 \hat{x}_{t-1} \tag{c.2}$$

but

$$\hat{x}_{t-1} = \theta x_{t-2} + (1-\theta)\hat{x}_{t-2} \tag{c.3}$$

Thus

$$\hat{x}_{t+1} = \theta x_t + \theta(1-\theta)x_{t-1} + \theta(1-\theta)^2 x_{t-2} + (1-\theta)^3 \hat{x}_{t-2} \tag{c.4}$$

By repeating the substitution, we find that the EWMA is a linear combination of x

$$\hat{x}_{t+1} = \sum_{i=0}^{n} w_i x_{(t-i)} \tag{c.5}$$

where the weights are given by equation (2.7)

$$w_i = \theta(1-\theta)^{t-i} \tag{2.7}$$

The variance of the EWMA is given by

$$\text{var}(\hat{x}) = \left(\sum_{i=0}^{t} w_i^2\right)\sigma^2 \qquad \text{where } \sigma^2 = \text{var}(x) \tag{c.6}$$

By putting equation (2.7) into equation (c.6) we get

$$\text{var}(\hat{x}) = \left[\theta^2 + \theta^2(1-\theta)^2 + \theta^2(1-\theta)^4 + \ldots\right]\sigma^2 \qquad \text{(c.7)}$$

$$\text{var}(\hat{x}) = \theta^2\sigma^2\left[1 + (1-\theta)^2 + (1-\theta)^4 + \ldots\right] \qquad \text{(c.8)}$$

Simplifying the above equation we get

$$\text{var}(\hat{x}) = \frac{\theta^2\sigma^2}{\left[1-(1-\theta)^2\right]} \qquad \text{(c.9)}$$

$$\text{var}(\hat{x}) = \frac{\theta\sigma^2}{(2-\theta)} \qquad \text{(c.10)}$$

Thus the standard deviation for the EWMA is given by

$$\sigma_{\hat{x}} = \sigma\sqrt{\frac{\theta}{(2-\theta)}} \qquad \text{(c.11)}$$

The corresponding standard deviation from equation (c.11) can be used to construct control limits at $\mu \pm K\sigma_{\hat{x}}$ limits. This is then given by equation (2.6). If 3σ control limits are used, then the control limits for the EWMA chart will $\mu \pm 3\sigma_{\hat{x}}$.

In cases where the process *mean* steadily trend away from the target, the EWMA chart can be improved by adding extra terms (equation 2.8) (Hunter, 1986)

$$\hat{x}_{t+1} = \hat{x}_t + \theta_1 e_t + \theta_2 \sum e_t + \theta_3 \nabla e_t \qquad \text{(2.8)}$$

Where

$$e_t = x_t - \hat{x}_t$$

$$\sum e_t = \sum(x_t - \hat{x}_t)$$

$$\nabla e_t = (x_t - \hat{x}_t) - (x_{t-1} - \hat{x}_{t-1})$$

The coefficients θ_1, θ_2 and θ_3 can be estimated from the historical data by using least squares. Rearranging equation (2.8)

$$\hat{x}_{t+1} - \hat{x}_t = \theta_1 e_t + \sum e_t + \theta_3 \nabla e_t \qquad \text{(c.12)}$$

Next define the following:

$$\phi = [\theta_1 \quad \theta_2 \quad \theta_3]^T$$

$$\mathbf{a} = \begin{bmatrix} \hat{x}_{t+1} - \hat{x}_t \\ \vdots \\ \hat{x}_N - \hat{x}_{N-1} \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} e_t & \sum e_t & \nabla e_t \\ \vdots & \vdots & \vdots \\ e_N & \sum e_N & \nabla e_N \end{bmatrix}$$

The least-squares solution of ϕ is:

$$\hat{\phi} = \left(\mathbf{B}^{\mathbf{T}}\mathbf{B}\right)^{-1}\mathbf{B}^{\mathbf{T}}\mathbf{a} \qquad \text{(c.13)}$$