STATISTICAL PROCESS CONTROL IN THE PRESENCE OF WITHIN-SUBGROUP CORRELATION WITH APPLICATIONS TO THE PRODUCTION OF INTEGRATED CIRCUITS

by

Douha Felfli

A thesis submitted to the School of Graduate Studies and Research

in partial fulfilment of the requirements for the

degree of

MASTER OF APPLIED SCIENCE

in the Department of Chemical Engineering

University of Ottawa

May 1993

ISBN 0-315-89672-8

ABSTRACT

Most standard control charting schemes for Statistical Process Control are based on the assumption that measurements of the product quality variable are independent within and between subgroups. When this assumption is violated, the performance of the control procedures deteriorates. While most studies in the recent literature have concentrated on resolving the problems associated with between-subgroup correlation, this thesis explores the effects of within-subgroup correlation. Such a situation occurs when multiple measurements of the same variable are taken from a single item and treated as a rational subgroup. Past attempts to use standard control charts resulted in numerous frustrating false alarms.

In this study, various methods for determining the existence and magnitude of within-subgroup correlation were evaluated. Using computer simulations to generate run length distributions, it was found that the performance of the traditional $\overline{X}$, Cumulative Sum (CUSUM) and Exponentially Weighted Moving Average (EWMA) charts is very poor in the presence of within-subgroup correlation. Three alternate approaches were proposed and evaluated: the use of modified univariate charts, principal components charts and multivariate versions of the $\overline{X}$, CUSUM and EWMA charts. Based on the comparison of the average and median run length performance of each chart, the use of a modified EWMA chart is recommended. The effectiveness of the proposed methods was demonstrated with applications to chemical processing steps in the semiconductor manufacturing industry.

ACKNOWLEDGEMENTS

I wish to express my gratitude to my two supervisors, Dr. David McLean and Mr. Nicholas Burn, for their constant support and guidance throughout this project.

I would also like to thank Northern Telecom Electronics Ltd. for the financial and technical support and the use of the computing and manufacturing facilities that made this research possible.

TABLE OF CONTENTS

ABSTRACT ii

ACKNOWLEDGEMENTS iii

TABLE OF CONTENTS iv

LIST OF TABLES vi

LIST OF FIGURES viii

NOMENCLATURE x

ACRONYMS xi

Chapter 1: INTRODUCTION 1

Chapter 2: THEORETICAL BACKGROUND AND LITERATURE REVIEW . . 6
2.1 On-Line Statistical Process Control Charts 6
2.1.1 Univariate Control Charts 9
2.1.2 Multivariate Control Charts 22
2.2 Testing the Correlation Structure for Independence 28
2.2.1 The Generalized Likelihood Ratio Test 30
2.2.2 The Partitioning Method 31
2.2.3 The Pairwise Comparisons Approach 33
2.3 Proposed Solutions 35
2.3.1 Modified Univariate Charts 36
2.3.2 Principal Components Approach 41
2.3.3 Multivariate Control Charts 43

Chapter 3: METHODOLOGY AND EXPERIMENTAL ASPECTS 44
3.1 Tests of Independence 45
3.2 Proposed SPC Charts 49

Chapter 4: RESULTS AND DISCUSSION 56
4.1 Tests of Independence 56
4.2 Proposed SPC Approaches 62
4.2.1 Modified Univariate Charts 65
4.2.2 Principal Components Approach 79
4.2.3 Multivariate Control Charts 79
4.3 Comparison of the Proposed Charts 87

Chapter 5: SEMICONDUCTOR FABRICATION 92
5.1 Process Overview 92
5.2 Application of SPC 103
5.2.1 Example 1. Photoresist Spin-Coating 104
5.2.2 Example 2. Oxidation Furnaces 110

Chapter 6: CONCLUSIONS AND RECOMMENDATIONS 115
6.1 Conclusions 115
6.2 Recommendations for Future Work 116

References 117

Appendix A: Generation of Correlated Data 120

Appendix B: Computer Simulation Programs 121

Appendix C: Additional Simulation Results 140

Appendix D: Sample Calculations 149

LIST OF TABLES

1. Sample table of data for a Shewhart chart 8

2. Shewhart control charts and their control limits 17

3. Summary of parameter settings for testing independence 47

4. SPC procedures used in the experiments 49

5. Summary of factors involved in SPC chart simulations 54

6. Percentage of acceptance of the generalized likelihood method for H_0: $\rho=I$ . . 58

7. Percentage of acceptance of the partitioning method for H_0: $\rho=I$ 58

8. Percentage of acceptance of the pairwise method using the t distribution for H_0: $\rho=I$ 59

9. Percentage of acceptance of the confidence intervals method for H_0: $\rho=I$ . . . 59

10. Percentage error involved in (a) ARL values and (b) MRL values 64

11. Simulated results for the $\overline{X}$ chart with Western Electric rules and $k \rightarrow \infty$ . . . 66

12. Simulated results for the $\overline{X}$ chart with Western Electric rules and $k = 20$ . . 67

13. Simulated results for the $\overline{X}$ chart with Western Electric rules and $k = 50$ . . 68

14. Simulated results of the 2-sided CUSUM (k_c=0.5, h=4) for $k \rightarrow \infty$ 71

15. Simulated results of the 2-sided CUSUM (k_c=0.5, h=4) for $k = 20$ 72

16. Simulated results of the 2-sided CUSUM (k_c=0.5, h=4) for $k = 50$ 73

17. Simulated results of the EWMA chart (λ=0.06, L=1.954) for $k \rightarrow \infty$ 76

18. Simulated results of the EWMA chart (λ=0.06, L=1.954) for k =20 77

19. Simulated results of the EWMA chart (λ=0.06, L=1.954) for k =50 78

20. Simulated results of Hotelling's T^2 for $k \rightarrow \infty$ 81

21. Simulated results of Hotelling's T^2 for $k = 20$ 81

22. Simulated results of Hotelling's T^2 for $k = 50$ 81

23. Simulated results of MEWMA ($r=0.06$, $h_4=13.5$) for $k \to \infty$ 83

24. Simulated results of MEWMA ($r=0.06$, $h_4=13.5$) for $k = 20$ 83

25. Simulated results of MEWMA ($r=0.06$, $h_4=13.5$) for $k = 50$ 83

26. Simulated results of MC1 ($k_1 = 0.5$, $h_2 = 9.5$) for $k \to \infty$ 85

27. Simulated results of MC1 ($k_1 = 0.5$, $h_2 = 9.5$) for $k = 20$ 85

28. Simulated results of MC1 ($k_1 = 0.5$, $h_2 = 9.5$) for $k = 50$ 85

29. Simulated results of MCUSUM ($k_2 = 0.5$, $h_3 = 6.82$) for $k \to \infty$ 86

30. Simulated results of MCUSUM ($k_2 = 0.5$, $h_3 = 6.82$) for $k = 20$ 86

31. Simulated results of MCUSUM ($k_2 = 0.5$, $h_3 = 6.82$) for $k = 50$ 86

32. Univariate control chart comparison based on ARLs for $p = 5$ 88

33. Multivariate control chart comparison based on ARLs for $p = 5$ 89

C1. Variability involved in the simulations of the tests of independence 140

C2. Simulated results of the 2-sided CUSUM ($k_c=0.5$, $h=5$) for $k \to \infty$ 141

C3. Simulated results of the 2-sided CUSUM ($k_c=0.5$, $h=5$) for $k = 20$ 142

C4. Simulated results of the 2-sided CUSUM ($k_c=0.5$, $h=5$) for $k = 50$ 143

C5. Simulated results of the EWMA chart ($\lambda=0.05$, $L=2.399$) for $k \to \infty$ 144

C6. Simulated results of the EWMA chart ($\lambda=0.05$, $L=2.399$) for $k =20$ 145

C7. Simulated results of the EWMA chart ($\lambda=0.05$, $L=2.399$) for $k =50$ 146

C8. Univariate control chart comparison based on MRLs for $p = 5$ 147

C9. Multivariate control chart comparison based on MRLs for $p = 5$ 148

D1. Training set data for example 2 149

LIST OF FIGURES

1. An illustration of (a) rational and (b) correlated subgroups 2

2. A Shewhart's $\overline{X}$ chart for an etch process with correlated subgroups 3

3. Sampling and data collection scheme 7

4. Shewhart's $\overline{X}$ chart for an in-control process with target 0 10

5. Zones of the $\overline{X}$ chart 14

6. Two-sided CUSUM chart with 0.5σ shift at subgroup 30 18

7. EWMA control chart with an 0.5σ upward shift after the 30'th subgroup . . . 21

8. Flowchart for simulations of the tests of independence 48

9. A typical run length distribution of a simulated chart 54

10. Flowchart for simulations of the control charts 55

11. The performance of the tests of independence for p=4 and k=20 61

12. Variability trends in (a) ARL and (b) MRL values 63

13. A planar silicon wafer 93

14. Cross sections of (a) a transistor, (b) a resistor and (c) a capacitor 94

15. Basic steps of IC fabrication for creating a p region 95

16. Basic steps of lithography 97

17. The difference between positive and negative photoresist 98

18. Possible locations for the measurements of the PQV on a wafer 104

19. The photoresist spin-coating process 105

20. The modified $\overline{X}$ chart for the training set of example 1 106

21. Correlation matrix of the training set of example 1 106

22. The modified (a) $\overline{X}$, (b) CUSUM and (c) EWMA charts for example 1 107

23. The standard (a) $\overline{X}$, (b) CUSUM and (c) EWMA charts for example 1 109

24. A schematic diagram of an oxidation furnace 111

25. Correlation matrix of the training set of example 2 111

26. The modified $\overline{X}$ chart for the training set of example 2 112

27. The modified (a) $\overline{X}$, (b) CUSUM and (c) EWMA charts for example 2 114

28. The standard (a) $\overline{X}$, (b) CUSUM and (c) EWMA charts for example 2 114

NOMENCLATURE

A_2, A_3	parameters of the standard $\overline{X}$ chart
B_3, B_4	parameters of the s chart
C	correction factor
D_3, D_4	parameters of the R chart
$\mathbf{I}$	identity matrix
k_1, h_2	parameters of the MCUSUM procedure
k_2, h_3	parameters of the MC1 procedure
k_c, h	parameters of the CUSUM chart
k	training set size
n	number of subgroups in a data set
p	subgroup size
q	number of multivariate quality characteristics
r, h_4	parameters of the MEWMA procedure
r_{ij}	sample correlation coefficient
R	range statistic
$\mathbf{R}$	sample correlation matrix
s	sample standard deviation statistic
$\mathbf{S}$	sample covariance matrix
T^2	Hotelling's T^2 statistic
X	random variable representing the PQV of interest
$\mathbf{X}$	vector of several random variables
$\overline{X}$	subgroup average
$\overline{\mathbf{X}}$	vector of subgroup averages
z	the standardized Normal/CUSUM statistic
Z	the EWMA statistic
$\mathbf{Z}$	the MEWMA statistic
Y_i	i'th principal component
Δ	magnitude of shift
λ_i, e_i	eigenvalue-eigenvector pair of the covariance matrix
λ, L	parameters of the EWMA chart
Λ	likelihood ratio statistic
ρ_{ij}	correlation coefficient of variables X_i and X_j
$\boldsymbol{\rho}$	population correlation matrix
σ	population standard deviation
$\boldsymbol{\Sigma}$	population covariance matrix
μ	population mean
$\boldsymbol{\mu}$	vector of population means

ACRONYMS

ARL	Average Run Length
CL	Center Line
CVD	Chemical Vapor Deposition
CUSUM	Cumulative Sum
EWMA	Exponentially Weighted Moving Average
IC	Integrated Circuit
LCL	Lower Control Limit
MRL	Median Run Length
PCA	Principal Component Analysis
PQV	Product Quality Variable
RL	Run Length
SPC	Statistical Process Control
UCL	Upper Control Limit

Chapter 1

INTRODUCTION

Statistical process control (SPC) is an important tool for improving process productivity and product quality. As industries become more concerned about delivering products of high quality to their customers while maintaining a high level of production, SPC has become popular and indispensable in the manufacturing environment. However, SPC has limitations which are founded primarily in the underlying assumptions about the input data. The main thrust of this thesis was to study the performance of SPC methods when the assumption of within-subgroup independence is violated.

SPC is a set of statistical tools that are essential for experimental design and data analysis in the performance of on-line quality control. Throughout this thesis, SPC will be studied relevant to on-line quality control and, unless otherwise stated, the term "SPC" will refer to this context. The primary objective of on-line SPC is to detect any shifts in the level or variability of a product quality variable (PQV). When a shift is detected, a flag or alarm is signalled so that corrective action can be initiated to remove the cause of the shift. A false alarm is expected occasionally as part of the process natural variation. However, false alarms occurring more frequently indicate that SPC has not been implemented properly and the validity of the underlying SPC assumptions should be re-assessed.

Conventional SPC methods are based on the assumption that observations of the PQV arise from identical, independent, Normal populations having constant variance.

Another important assumption lies in the notion of rational subgroups. A subgroup consists of a fixed number of observations which are monitored with time. In a rational subgroup, the within subgroup variation is random and small when compared to the variation between subgroups (Holmes and Mergen, 1989). The difference in variation enables the detection of shifts from one subgroup to another. Violation of one or more of these assumptions leads to confusion and misinterpretation of the results of SPC.

The assumption of within-subgroup independence is the main concern of this thesis. This assumption is violated when the observations constituting a subgroup have a fixed correlation structure. Usually, this situation occurs when multiple measurements of the same variable are taken from a single item or part, as shown in Figure 1.

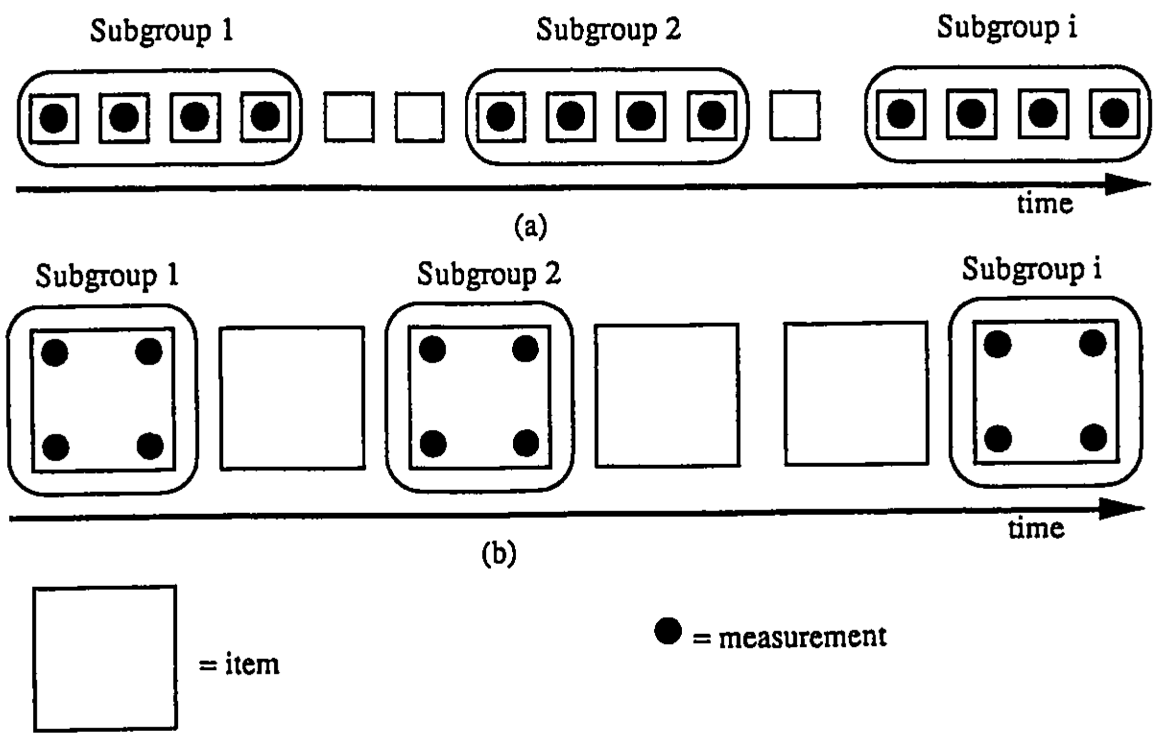


Figure 1. An illustration of a) rational subgroups and b) correlated subgroups.

The incentive for the study arises from the semiconductor manufacturing industry. Fabrication of integrated circuits takes place by a series of planar processes on the surface of a flat silicon wafer. A typical subgroup consists of five or six measurements of a critical dimension taken at specific locations on a single wafer. The ultimate objective of such processing is to achieve perfect uniformity both across the wafer and across a batch of wafers. This choice of subgroup violates the rational subgroup criterion because the within-subgroup variation is not random but has a specific correlation structure. Past attempts to implement the widely used Shewhart control charts resulted in frequent false alarms leading to time and production losses as non-existent causes for the alarms were sought. One of these charts is displayed in Figure 2 which shows that most of the data points lie outside the control limits incorrectly indicating that the process is constantly out of control.

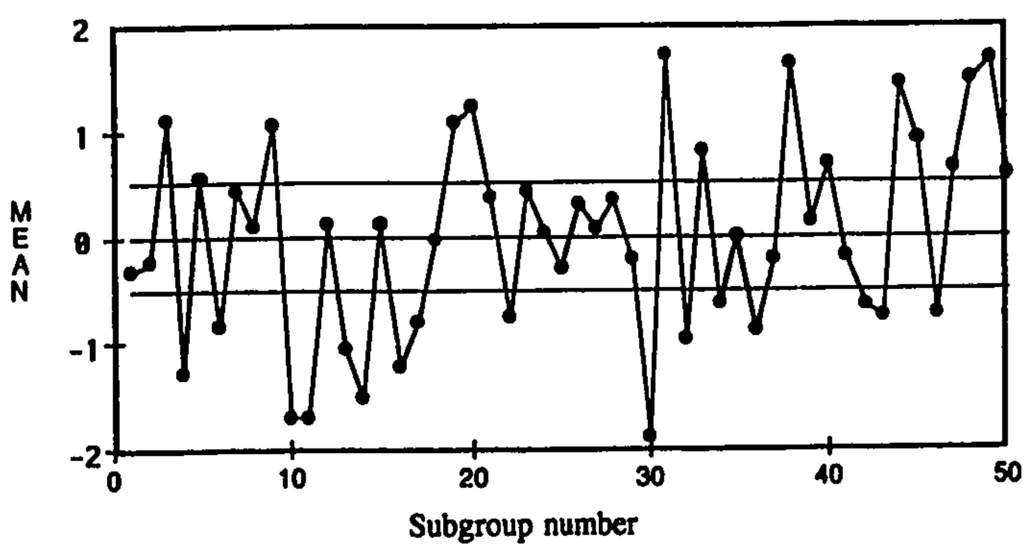


Figure 2. A Shewhart $\overline{X}$ chart for an etch process with correlated subgroups.

While the motivation for this research originates from the semiconductor manufacturing industry, correlated subgroup data is likely to occur in other industries such as in parts manufacturing where product consistency on the same unit is important or in batch processing where batch uniformity is desired (Burn, 1988). Therefore, the results of this study can be applied in a variety of production environments.

The existence of a correlation structure in SPC subgroups has not been addressed extensively in the scientific literature. Neuhardt (1987) suggested the use of a correction factor that modifies the control limits in Shewhart's $\overline{X}$ chart in order to reduce the frequency of false alarms. Burn (1988) attempted to use the corrected Shewhart charts in semiconductor fabrication and succeeded in decreasing the number of false alarms. However, both papers did not discuss the sensitivity of the corrected charts to shifts in the process. Another approach to the problem is to use multivariate control charts that simultaneously control several correlated process variables or characteristics such as temperature and pressure inside a closed container. Multivariate charts can monitor correlated subgroups if each element of the subgroup is considered as a separate variable and the subgroup is treated as a multivariate vector of observations. Nevertheless, this approach has never been applied on correlated subgroups.

The overall goal of this research was to find appropriate SPC methods that effectively deal with the existence of within-subgroup correlation. The first specific objective was to evaluate the effectiveness of several methods that test for the presence of within-subgroup correlation. An effective test would provide a way to determine when alternate monitoring procedures need to be employed to deal with the correlation.

The second objective was to assess three approaches to SPC that have the potential of treating within-subgroup correlation. The first used Neuhardt's correction factor to modify three standard SPC procedures: Shewhart's $\overline{X}$ chart, the cumulative sum (CUSUM) scheme and the exponentially weighted moving average (EWMA) chart. In the second approach, each subgroup was treated as a multivariate observation and SPC charts for the principal components of the multivariate observations were attempted. The third approach also treated each subgroup as a multivariate vector of observations and multivariate control schemes, such as Hotelling's T^2, multivariate EWMA (MEWMA) and multivariate CUSUM (MCUSUM), were used. The performance of each of these charts was evaluated on the basis of the frequency of false alarms as well as the speed of detecting a shift in the level of the PQV by means of computer simulations. A third objective was to assess the performance of the best procedures by implementation in the integrated circuit fabrication industry where this problem is prevalent.

This thesis is divided into six chapters. Chapter 2 explains the theoretical background of the research and gives a review of the proposed methods. The methodology used in solving the problem is discussed in Chapter 3. Chapter 4 includes the results of the study along with a discussion of their significance. In Chapter 5, a brief overview of the integrated circuit fabrication process is given, followed by illustrations of the application of the proposed methods to real data from that industry. Finally, the conclusions and recommendations are stated in Chapter 6.

Chapter 2

THEORETICAL BACKGROUND AND LITERATURE REVIEW

This chapter discusses the theoretical background for the research. It is divided into three main sections. The first section reviews several SPC charting techniques including univariate and multivariate charts. The second section is a review of statistical methods that might be used to identify the existence of within subgroup correlation. This section also briefly discusses theoretical methods for recognizing equal correlation structures when the correlation is known to exist. The third and last section of this chapter covers theoretical aspects of SPC charts that have the potential to handle correlated subgroups.

2.1 On-Line Statistical Process Control Charts

On-line SPC charts are designed to detect significant changes in the operation of a process. The ultimate goal of all these charts is to signal an alarm when a shift due to an "assignable" or "special" cause occurs. The level of operation is usually reflected in one or more product quality variables or characteristics. For example, in a process that manufactures screws, one product quality variable is the diameter of the screw because it predicts whether the screw is functional or not. The process is said to be in a state of "statistical control" when the measured values of the PQV vary randomly about a target or mean value. "Common cause" variation refers to this type of variability and is considered a normal part of the process. When the bounds of normal variation are exceeded, an "assignable cause" is considered to have produced a shift in the level or variability of the PQV away from its target value. Examples of an assignable cause would be a new operator who lacks training, a change in the properties of the raw materials or a drift in the equipment setting.

An important factor in the evaluation of a control charting procedure is its average run length (ARL). The run length of a chart is the number of points plotted on the chart before an alarm is signalled. The average run length (ARL) is the expected value of the run length. It is desirable to obtain a high ARL when the process is "in control" to avoid numerous false alarms, and a low ARL when a small shift occurs in order to have sensitive shift detection. This concept is explained in detail in Chapter 3.

The sampling scheme, data collection and notation for the construction of simple SPC charts is now described. Let X be the random variable that represents the product quality variable (PQV) of interest. The true mean and standard deviation of X are μ and σ respectively. Subgroups of size p are considered over time and measurements of X for each member of each subgroup are taken. Let X_{ij} denote the j'th observed value of X in the i'th subgroup, as illustrated in Figure 3.

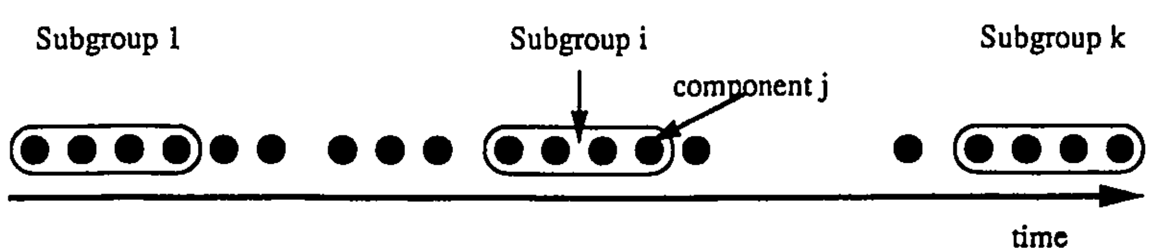


Figure 3. Sampling and data collection scheme

The data are reported in a form similar to that shown in Table 1 where each row represents the measurements taken from one subgroup. The average and standard deviation of the i'th subgroup, $\overline{X}_i$ and s_i, are given by

$$\overline{X}_i = \frac{\sum_{j=1}^{p} X_{ij}}{p} \quad \text{and} \quad s_i = \frac{\sqrt{\sum_{j=1}^{p} (X_{ij} - \overline{X}_i)^2}}{p-1}.$$

The subgroup average and standard deviation are indicative of the PQV level and variability respectively. This notation will be used throughout the thesis whenever univariate charts are referenced.

Table 1. Sample table of data for a Shewhart chart

subgroup	X_1	X_2	.	.	X_p	$\overline{X}_i$	s_i
1	X_{11}	X_{12}	.	.	X_{1p}	$\overline{X}_1$	s_1
2	X_{21}	X_{22}	.	.	X_{2p}	$\overline{X}_2$	s_2
.	.	.	.	.	.	.	.
.	.	.	.		.	.	.
k	X_{k1}	X_{k2}	.	.	X_{kp}	$\overline{X}_k$	s_k

There are two types of SPC charts: univariate and multivariate. Univariate charts, as the name indicates, monitor one quality characteristic only. On the other hand, multivariate charts simultaneously monitor several quality characteristics which are highly correlated. In such a case, the use of univariate charts for each quality characteristic can give misleading results because of the correlation between these characteristics (Ryan, 1989).

When multiple measurements of the same PQV are sampled from a single part, it is logical to use univariate control charts to monitor the process over time since all the measurements are representative of the same variable. However, because of the correlation that exists among the measurements, an alternative approach is to treat each measurement as a separate quality characteristic and to use multivariate control charts. This section reviews the traditional SPC schemes which are used in this thesis to study the effect of correlated subgroups.

2.1.1 Univariate Control Charts

The univariate charts reviewed in this section are the Shewhart, Cumulative Sum (CUSUM) and Exponentially Weighted Moving Average (EWMA) charts. All of these schemes monitor one quality characteristic and are based on common assumptions that the measurements are Normally, independently, identically distributed.

Shewhart Control Charts

These charts were first introduced by W. Shewhart in 1931. His main objective was to provide a historical summary of process performance that would detect any disturbances and allow for their removal, thereby improving the stability of the process. Today, Shewhart charts are the most common among all SPC charts. They are easily implemented, detect shifts and provide a historical summary of the process, the latter being a significant advantage over other charting schemes (Hoerl and Palm, 1992).

The Shewhart control charts involve plotting measures of both the dispersion and level of the PQV versus time or order of production. Control lines are constructed on the chart to show the bounds of the natural variation and the expected level of that statistic. The control lines consist of a center line (CL), an upper control limit (UCL) and a lower control limit (LCL). When the process is in control, the plotted points should vary around the center line while remaining within the region defined by the upper and lower control limits. Figure 4 shows a typical Shewhart $\overline{X}$ chart with its control limits.

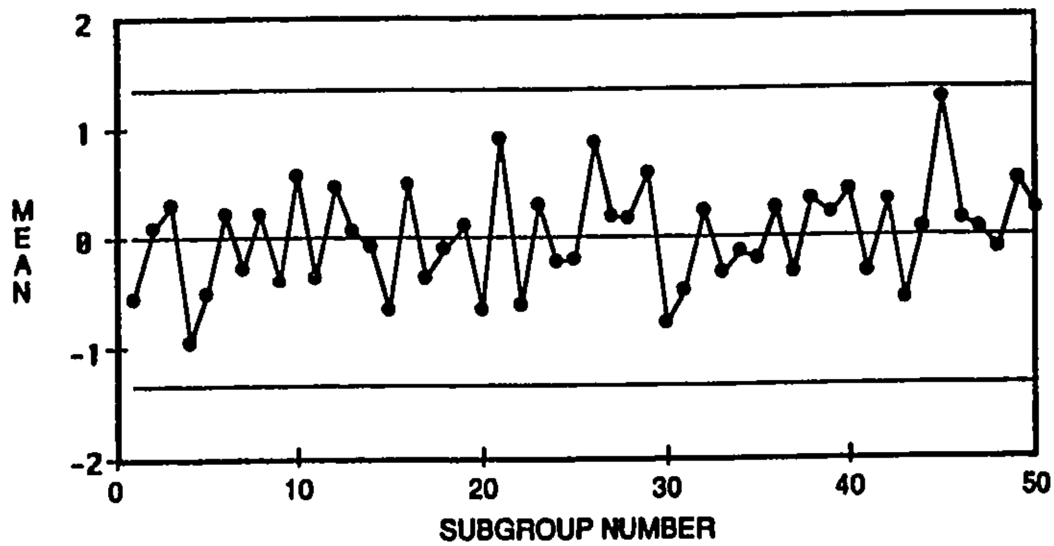


Figure 4. A Shewhart $\overline{X}$ chart for an in-control process with target 0.

Shewhart charts consist of either an R or s chart coupled with an $\overline{X}$ chart. The R and s charts monitor the variation <u>within</u> the subgroups while the $\overline{X}$ chart monitors the level of operation as well as the variation <u>between</u> the subgroups. Hence, a combination of the $\overline{X}$/R or $\overline{X}$/s charts is a good indicator of the level of operation as well as the within and between subgroup variability. The R and s charts should be examined prior to the $\overline{X}$ chart because the distribution of the measurements would not be stable unless the variability is in a state of statistical control. Individual observations can also be plotted rather than the subgroup averages resulting in the so-called I or X chart.

R and s Charts The R and s charts monitor the within subgroup variability. Define the range, R, statistic as

$$R_i = X_{i,max} - X_{i,min}$$

where $X_{i,max}$ and $X_{i,min}$ are the maximum and minimum values of X within the i'th subgroup. The s statistic is given by

$$s_i = \sqrt{\frac{\sum_{i=1}^{p} (X_{ij} - \overline{X}_i)^2}{p-1}}$$

where $\overline{X}_i$ is the subgroup average. The R and s statistics are estimators of the true population standard deviation σ.

For the R chart, the control limits are given as follows:

$$CL = \overline{R} = \frac{R_1 + R_2 + \ldots + R_k}{k}$$
$$UCL = \overline{R} + 3\hat{\sigma}_R$$
$$LCL = \overline{R} - 3\hat{\sigma}_R$$

or $\overline{R} \pm 3\hat{\sigma}_R$. Similarly, the control limits for the s chart are $\overline{s} \pm 3\hat{\sigma}_s$, where $\hat{\sigma}_R$ and $\hat{\sigma}_s$ are the estimates of the standard deviation of R and s respectively.

$\overline{R}$ and $\overline{s}$ are biased estimators of σ, the population standard deviation. The bias is compensated for by applying correction factors d_2 and c_4 to $\overline{R}$ and $\overline{s}$, where d_2 and c_4

are tabulated values that are a function of the subgroup size, p. $\overline{R}/d_2$ and $\overline{s}/c_4$ constitute unbiased estimators of σ when the data is Normally distributed. Following the same line of reasoning, the estimated standard deviations of R and s are given by

$$\hat{\sigma}_R = d_3\hat{\sigma} = \frac{d_3\overline{R}}{d_2} \quad \text{and} \quad \hat{\sigma}_s = \sqrt{1-c_4^2}\,\hat{\sigma} = \frac{\sqrt{1-c_4^2}\,\overline{s}}{c_4}.$$

Therefore, the upper and lower control limits for R and s charts can be written as

$$\overline{R} \pm 3\frac{d_3}{d_2}\overline{R} \quad \text{and} \quad \overline{s} \pm 3\frac{\sqrt{1-c_4^2}}{c_4}\overline{s}.$$

Alternatively, the control limits can be expressed for R charts as

$$\begin{aligned} LCL &= D_3\overline{R} \\ UCL &= D_4\overline{R} \end{aligned}$$

$$\text{where} \quad D_3 = 1-3\frac{d_3}{d_2} \quad \text{and} \quad D_4 = 1+3\frac{d_3}{d_2}.$$

For s charts,

$$\begin{aligned} LCL &= B_3\overline{s} \\ UCL &= B_4\overline{s} \end{aligned}$$

$$\text{where} \quad B_3 = 1-3\frac{\sqrt{1-c_4^2}}{c_4} \quad \text{and} \quad B_4 = 1+3\frac{\sqrt{1-c_4^2}}{c_4}.$$

Values for d_2, d_3, c_4, D_3, D_4, B_3 and B_4 are tabulated as functions of p and can be found in most quality control texts or manuals (such as "ASTM Manual on Quality Control of Materials").

The $\overline{X}$ Chart For a subgroup of size p, the i'th subgroup average, $\overline{X}_i$, is an unbiased estimator of the true population mean, μ. The $\overline{X}$ chart is a plot of the subgroup averages versus time that monitors the variation between subgroups. The center line of the $\overline{X}$ chart is a desired target value or the average of k subgroup averages,

$$\overline{\overline{X}} = \frac{\sum_{i=1}^{k} \overline{X}_i}{k}.$$

The upper and lower control limits are $\overline{\overline{X}} \pm 3\hat{\sigma}_{\overline{X}}$. When the distribution of the data is Normal, 99.7% of the $\overline{X}$'s are expected to lie within the control limits during normal operating conditions. The estimated standard deviation of $\overline{X}$, $\hat{\sigma}_{\overline{X}}$, can be calculated from the standard deviation of the population since

$$\hat{\sigma}_{\overline{X}} = \frac{\hat{\sigma}}{\sqrt{p}}$$

provided that the observations are <u>independent</u>.

When the sample subgroup ranges are used to estimate σ, $\hat{\sigma} = \overline{R}/d_2$ and consequently

$$\hat{\sigma}_{\overline{X}} = \frac{\hat{\sigma}}{\sqrt{p}} = \frac{\overline{R}}{d_2\sqrt{p}}.$$

Then, the control limits are

$$\overline{\overline{X}} \pm A_2\overline{R} \quad \text{where } A_2 = \frac{3}{d_2\sqrt{p}}.$$

Alternatively, when the sample standard deviations are used to monitor σ, $\hat{\sigma} = \overline{s}/c_4$ and

$$\hat{\sigma}_{\bar{X}} = \frac{\bar{s}}{c_4\sqrt{p}}.$$

The control limits are

$$\bar{\bar{X}} \pm A_3\bar{s} \quad \text{where } A_3 = \frac{3}{c_4\sqrt{p}}.$$

As in the case of s or R charts, A_2, A_3, d_2 and c_4 are tabulated values as a function of p, the subgroup size.

Many rules have been developed to detect non-random, out of control behaviour of an $\bar{X}$ chart. The use of the concept of a "run" in the chart becomes necessary here. A run is defined in one of two ways (Burn and McLean, 1990):

- A set of consecutive points that lie either above or below the average.
- A set of consecutive points either increasing or decreasing in value.

Another useful concept in detecting out of control behaviour in $\bar{X}$ charts is based on dividing the control chart into zones. Division of the space between the control limits into 6 equal areas creates the zones shown in Figure 5. Each zone occupies one $\hat{\sigma}_{\bar{X}}$.

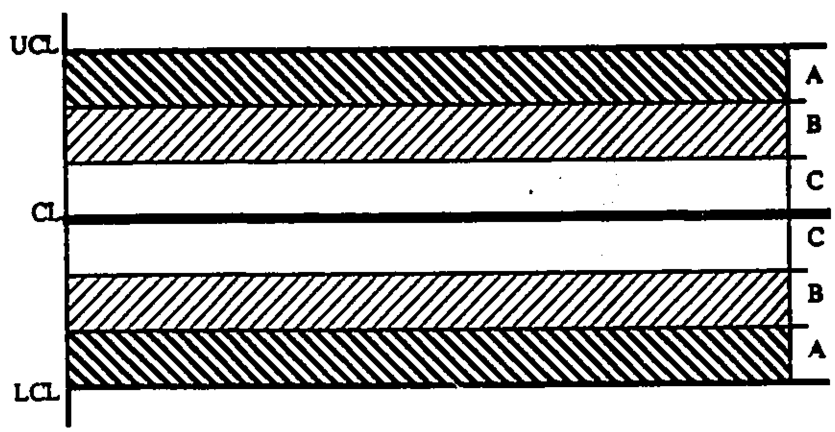


Figure 5. Zones of the $\bar{X}$ chart

There are two standard sets of rules which make use of the control chart zones to identify non random behaviour of the process. A process in a state of statistical process control should not exhibit the behavior described by these rules. The first set was produced by Western Electric (1956) and is summarized below:

(i) 1 data point beyond zone A.

(ii) 2 out of 3 data points in zone A or beyond on one side of the center line.

(iii) 4 out of 5 data points in zone B or beyond on one side of the center line.

(iv) 8 consecutive data points on one side of the center line.

(v) 15 consecutive data points inside zone C on either side of the center line.

(vi) 8 consecutive data points outside zone C on either side of the center line.

The other set was provided by Nelson (1985) and is listed below:

(i) 1 data point beyond zone A.

(ii) 9 consecutive points in zone C or beyond.

(iii) 6 consecutive points steadily increasing or decreasing.

(iv) 14 consecutive points alternating up and down.

(v) 2 out of 3 points in zone A or beyond.

(vi) 4 out 5 points in zone B or beyond.

(vii) 15 consecutive points in zone C on one side of the centerline.

(viii) 8 consecutive points outside zone C on both sides of the centerline.

The Individuals Chart The I or X chart is a control chart for individual observations instead of subgroup averages. The use of this chart may be unavoidable in some situations, such as a slow process where considerable time is required to produce one measurement. The major restriction of this chart is that the raw data must be Normally distributed. The control limits are $\hat{\mu} \pm 3\hat{\sigma}$ where $\hat{\mu}$ and $\hat{\sigma}$ are estimates of the population mean and standard deviation respectively. The sample average, $\overline{X}$, and standard deviation, s, of k observations are good estimates of μ and σ. Ryan (1989) recommends the use of at least 50 observations to obtain these estimates.

Another method for estimating σ involves the computation of the differences between successive observations. These differences are known as the moving ranges (MR). The average moving range, $\overline{MR}$, constitutes a good estimate of σ when divided by d_2, as in the case of subgroup sample ranges. Therefore, the control limits of the I chart are $\overline{X} \pm 3\overline{MR}/d_2$ when $\overline{MR}$ is used to estimate σ or $\overline{X} \pm 3s/c_4$ when s is used instead.

Note that the I chart is not as sensitive as the $\overline{X}$ chart for detecting shifts in the mean (Ryan, 1989) and that its applicability is dependent on the Normality and independence of the population. The I chart is usually coupled with a Moving Range (MR) chart which monitors the variability of the observations. The MR chart is not critical for this research and hence will not be discussed. A summary of all the Shewhart charts and their control limits is given in Table 2.

Table 2. Shewhart Control charts and their control limits

Type of chart	Method of Estimating σ	CL	UCL	LCL
$\overline{X}$	true σ s R	$\hat{\mu}$ $\overline{\overline{X}}$ $\overline{\overline{X}}$	$\hat{\mu} + 3\sigma/\sqrt{p}$ $\overline{\overline{X}} + A_3\overline{s}$ $\overline{\overline{X}} + A_2\overline{R}$	$\hat{\mu} - 3\sigma/\sqrt{p}$ $\overline{\overline{X}} - A_3\overline{s}$ $\overline{\overline{X}} - A_2\overline{R}$
R		$\overline{R}$	$D_4\overline{R}$	$D_3\overline{R}$
s		$\overline{s}$	$B_4\overline{s}$	$B_3\overline{s}$
I or X	true σ s MR	μ $\overline{X}$ $\overline{X}$	$\mu + 3\sigma$ $\overline{X} + 3s/c_4$ $\overline{X} + 3\overline{MR}/d_2$	$\mu - 3\sigma$ $\overline{X} - 3s/c_4$ $\overline{X} - 3\overline{MR}/d_2$
MR		$\overline{MR}$	$D_4\overline{MR}$	$D_3\overline{MR}$

The CUSUM Chart

The cumulative sum (CUSUM) procedures are control schemes that involve the accumulation of the sum of a sample statistic. The first CUSUM chart was introduced by E.S. Page (1954); however, many modifications have been made to that chart resulting in a variety of CUSUM procedures. In this thesis, only the two-sided CUSUM procedure (Lucas, 1982) will be considered due to its simplicity and ease of application in computer simulations.

The two-sided CUSUM consists of two cumulative sums: one monitors upward trends while the other monitors downward trends. These two sums are denoted by S_H and S_L respectively. Suppose that the subgroup average, $\overline{X}$, is the statistic to be monitored by the chart. Let $\overline{\overline{X}}$ and $\hat{\sigma}_{\overline{X}}$ be the average and sample standard deviation of k subgroup averages. The i'th subgroup average, $\overline{X}_i$, is standardized to its z value,

$$z_i = \frac{\bar{X}_i - \bar{\bar{X}}}{\sigma_{\bar{X}}}$$

The two cumulative sums S_{Hi} and S_{Li} are then formed.

$$S_{Hi} = \max[0, (z_i - k_c) + S_{Hi-1}]$$
$$S_{Li} = \max[0, (-z_i - k_c) + S_{Li-1}]$$

where k_c is one half the value of the shift to be detected by the chart in units of z. The two sums are started with S_{H0} and S_{L0} both set to zero. The procedure ensures that the standardized i'th subgroup average is accumulated only if it diverts from the process mean by more than k_c units in the upward or downward directions. An out of control signal is given when either $S_{Hi} > h$ or $S_{Li} > h$ where h is a threshold value typically chosen as 4 or 5.

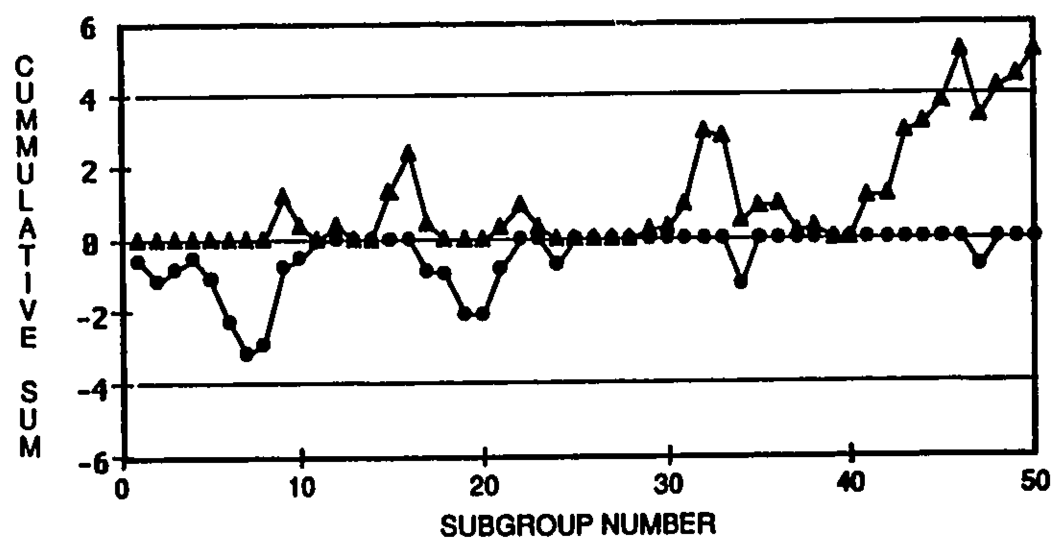


Figure 6. A Two-sided CUSUM chart (k_c = 0.5, h = 4) with 0.5σ shift at subgroup 30

The two sums S_H and S_L can be monitored in tabular or graphical form. In either case, different criteria can be applied to the procedure in order to detect out of control situations. This is achieved by varying the parameters k_c and h to obtain the desired ARL. A two-sided CUSUM chart is shown in Figure 6. Note that the values of S_L are negated and compared to -h for simplicity.

The Exponentially Weighted Moving Average (EWMA) Chart

The exponentially weighted moving average (EWMA) chart was first introduced by Roberts (1959). It is a quality control tool that utilizes the statistic Z_i which is given by the expression

$$Z_i = (1-\lambda) Z_{i-1} + \lambda \overline{X}_i$$

where λ is a constant that determines the relative weighting of the current and past values of $\overline{X}$ such that $0 < \lambda < 1$. The EWMA statistic can also be applied using the individual observations X_i instead of $\overline{X}_i$, in the same situations when an I chart is used instead of an $\overline{X}$ chart.

For the (i-1)'th subgroup, the EWMA statistic is

$$Z_{i-1} = (1-\lambda) Z_{i-2} + \lambda \overline{X}_{i-1}$$

Substitution of Z_{i-1} in the expression for Z_i gives

$$Z_i = (1-\lambda)^2 Z_{i-2} + (1-\lambda) \lambda \overline{X}_{i-1} + \lambda \overline{X}_i$$

Recursive substitution for Z_{i-j} until i=j yields the expression

$$Z_i = (1-\lambda)^i Z_0 + (1-\lambda)^{i-1}\lambda\bar{X}_1 + (1-\lambda)^{i-2}\lambda\bar{X}_2 + \ldots + (1-\lambda)\lambda\bar{X}_{i-1} + \lambda\bar{X}_i$$

$$\text{or} \quad Z_i = (1-\lambda)^i Z_0 + \lambda\sum_{j=0}^{i-1}(1-\lambda)^j\bar{X}_{i-j}$$

Since $0 < (1-\lambda) < 1$, $(1-\lambda)^j$ decreases exponentially as j increases. This implies that as i, the number of subgroups, increases, the weight put on past observations decreases exponentially. It is noted here that Shewhart and CUSUM control procedures employ different weighting schemes on current and past data. While Shewhart control charts consider only the current subgroup in detecting out of control behaviour, the two-sided CUSUM puts an equal weight on every past data point which is accumulated either in S_H or S_L.

The EWMA chart, like the Shewhart chart, consists of a center line and upper and lower control limits. The center line is set at the target, τ, of the process or alternatively at $\bar{\bar{X}}$. The upper and lower control limits are LCL $= \tau - L_1\hat{\sigma}_{Zi}$ and UCL $= \tau + L_2\hat{\sigma}_{Zi}$, where L_1 and L_2 are chosen to provide the desired average run length. $\hat{\sigma}_{Zi}$ is the estimated standard deviation of the Z statistic and is determined as

$$\hat{\sigma}_{Zi} = \sqrt{(1-\lambda)^{2i}\sigma_{Z0}^2 + \left(\frac{\lambda}{2-\lambda}\right)[1-(1-\lambda)^{2i}]\,\sigma_{\bar{X}}^2}.$$

For large i,

$$\hat{\sigma}_{Z_i} = \sqrt{\left(\frac{\lambda}{2-\lambda}\right)}\,\hat{\sigma}_{\bar{X}} = \sqrt{\left(\frac{\lambda}{2-\lambda}\right)}\frac{\hat{\sigma}}{\sqrt{p}}$$

where $\hat{\sigma}$ is an estimate of the population standard deviation. Note that the final equality assumes that the observations within the subgroup are independent.

It is conventional to equate L_1 and L_2 to one value L so that the control limits are symmetrical about the target. The parameters of the EWMA chart are λ and L. Both parameters can be chosen to provide a desirable ARL. Optimal values of the parameters have been tabulated along with their ARLs for shifts of various magnitudes (Lucas and Saccucci, 1990; Robinson and Ho, 1978). Numerical simulation also allows the determination of optimal parameter settings for a given situation. Figure 7 shows a EWMA control chart with a 0.5σ shift in the PQV occurring after the 30'th subgroup.

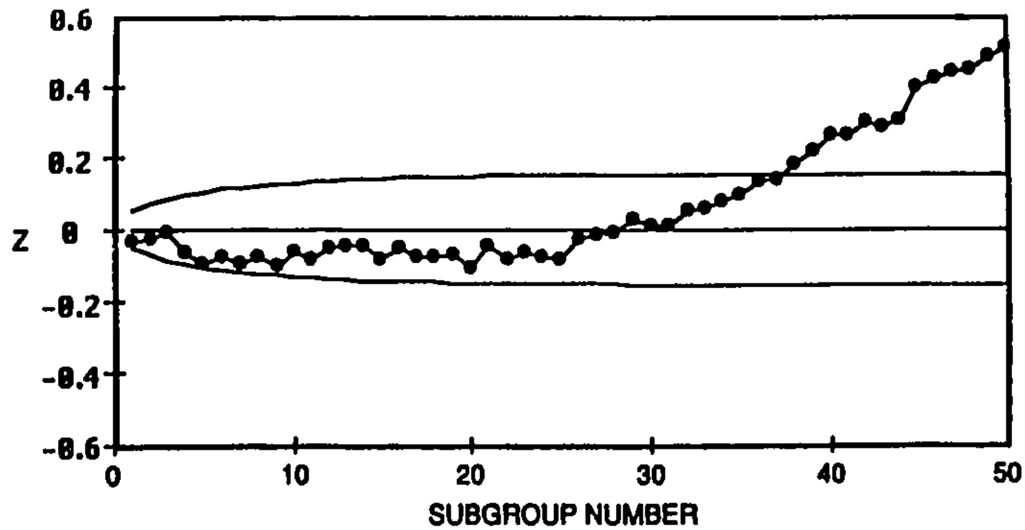


Figure 7. A EWMA control chart with a 0.5σ upward shift after the 30'th subgroup

When any control chart is to be implemented, the true mean and standard deviation of a process are often unknown. However, a training set can be used to obtain estimates of μ and σ. The training set usually consists of 20 to 50 subgroups, measured when the process can be considered in control. Based on the information given by the training set, the control limits for the charts are established. Whenever permanent change occurs to the process, a new training set is considered and the control limits are reset based on this change.

As indicated above, all univariate control charts are based on the assumption that the data points are identically independently Normally distributed. Control charts that monitor subgroup averages function properly even if the distribution is not exactly Normal due to the central limit theorem that states that the distribution of $\overline{X}$ is Normal if the data set is large enough regardless of the distribution of X. However, violation of the independence assumption affects the performance of control charts. The existence of within subgroup correlation is of particular interest in this thesis and will be discussed later. In the following subsection, multivariate control charts are addressed.

2.1.2 Multivariate Control Charts

Multivariate control charts are used whenever there are several quality characteristics or variables that need to be controlled simultaneously. Maintaining univariate charts for each quality variable would give misleading results when these variables are correlated (Ryan, 1989). The multivariate charts include Hotelling's T^2, multivariate CUSUM (MCUSUM) and EWMA (MEWMA).

Let X_1, X_2, ..., X_q be normal random variables representing the q quality characteristics that are to be monitored. Assume that each random variable, X_u, is Normally distributed with mean μ_u and standard deviation σ_u. Let **X** be a qx1 random vector representing these quality characteristics such that

$$\mathbf{X}^T = [X_1, \ X_2, \ . \ . \ . \ , \ X_q].$$

The vectors $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_3$, ... denote the values of $\mathbf{X}$ when observed over time, that is

$$\mathbf{X}_i = \begin{bmatrix} X_{1i} \\ X_{2i} \\ \cdot \\ \cdot \\ X_{qi} \end{bmatrix}$$

Let μ be the mean vector and Σ the variance-covariance matrix of $\mathbf{X}$. Then,

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \cdot \\ \cdot \\ \cdot \\ \mu_q \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & . & . & \sigma_{1q} \\ \sigma_{21} & \sigma_{22} & . & . & \sigma_{2q} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \sigma_{q1} & \sigma_{q2} & . & . & \sigma_{qq} \end{bmatrix}$$

where σ_{uv} is the covariance of X_u and X_v. A set of n observations can be considered as a multivariate pxn data set consisting of the vectors $\mathbf{X}_1$, $\mathbf{X}_2$, ..., $\mathbf{X}_n$. The estimates of μ_u and σ_{uv} are $\overline{X}_u$ and s_{uv} respectively such that

$$\overline{X}_u = \frac{\sum_{v=1}^{n} X_{uv}}{n-1} \quad \text{and} \quad s_{uv} = \frac{\sum_{w=1}^{n} (X_{uw}-\overline{X}_u)\ (X_{vw}-\overline{X}_v)}{n-1}$$

Similarly, the best estimates of μ and Σ are $\overline{\mathbf{X}}$ and $\mathbf{S}$, where:

$$\overline{\mathbf{X}}_{qx1} = \begin{bmatrix} \overline{X}_1 \\ \overline{X}_2 \\ \cdot \\ \cdot \\ \cdot \\ \overline{X}_q \end{bmatrix} \quad \text{and} \quad \mathbf{S}_{qxq} = \begin{bmatrix} s_{11} & s_{12} & . & . & s_{1q} \\ s_{21} & s_{22} & . & . & s_{2q} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ s_{q1} & s_{q2} & . & . & s_{qq} \end{bmatrix}$$

Using these multivariate statistical concepts, several multivariate charts will be discussed. It will be assumed that the process is in control when the mean is μ_0.

Hotelling's T^2 Procedures

These procedures are based on the T^2 statistic which was first introduced by Hotelling (1947). The T^2 statistic is given by

$$T^2 = n\,(\mathbf{\bar{X}} - \boldsymbol{\mu}_0)^T \mathbf{S}^{-1} (\mathbf{\bar{X}} - \boldsymbol{\mu}_0)$$

where $\mathbf{S}^{-1}$ is the inverse of $\mathbf{S}$. It is desirable to test the hypothesis H_0 that $\mu = \mu_0$. When $\mu = \mu_0$, Hotelling's T^2 is approximately distributed as $q(n-1)/(n-q)\ F_{\alpha,(q,n-q)}$ (Ryan, 1989). Therefore, H_0 is accepted whenever $T^2 < q(n-1)/(n-q)\ F_{\alpha,(q,n-q)}$ at a level of significance α.

This concept is utilized to construct multivariate control charts. As in the case of univariate charts, μ is usually unknown and has to be estimated. When there are k subgroups, $\bar{\bar{\mathbf{X}}}$ is an estimate of μ such that

$$\bar{\bar{\mathbf{X}}} = \begin{bmatrix} \bar{\bar{X}}_1 \\ \bar{\bar{X}}_2 \\ \cdot \\ \cdot \\ \cdot \\ \bar{\bar{X}}_q \end{bmatrix} \quad \text{and} \quad \bar{\bar{X}}_u = \frac{\sum_{v=1}^{k} \bar{X}_{uv}}{k}$$

where $\bar{X}_{uv}$ is the average of variable v for the u'th subgroup.

For the u'th subgroup, Hotelling's T^2 is computed as

$$T_u^2 = n\,(\mathbf{\bar{X}}_u - \bar{\bar{\mathbf{X}}})^T \mathbf{S}_p^{-1} (\mathbf{\bar{X}}_u - \bar{\bar{\mathbf{X}}})$$

where $\mathbf{S}_p$ is the pooled variance-covariance matrix over k subgroups. The values of T^2 are plotted versus time. Alt (1982) showed the upper control limit for the chart is

$$UCL = \left(\frac{knq - kq - nq + q}{kn - k - q - 1}\right) F_{\alpha,\,(q,\,kn-k-q+1)}\,.$$

It is noted that there is no lower control limit for this chart.

Jackson (1959) extended this T^2 concept to provide a multivariate analogue of the I chart. Here, multivariate individual observations are used instead of subgroup averages. A training set of k individual multivariate observations provides estimates of the mean vector and the variance covariance matrix. Let $\overline{\mathbf{X}}_k$ and $\mathbf{S}_k$ denote these estimates respectively. For the i'th vector of observations, $\mathbf{X}_i$, the quantity

$$T_i^2 = (\boldsymbol{X}_i - \overline{\boldsymbol{X}}_k)^T \boldsymbol{S}^{-1} (\boldsymbol{X}_i - \overline{\boldsymbol{X}}_k)$$

is computed and plotted on the control chart. The upper control limit for this chart is

$$UCL = \frac{q(k+1)(k-1)}{k^2 - kq} F_{\alpha, (q, k-q)}$$

where k is the size of the training set and q is the number of the characteristics that are to be monitored. This control chart is effective for large values of k and cannot be used to test whether the observations in the training set are in control (Ryan, 1989).

Multivariate CUSUM Procedures

Several attempts have been made to develop CUSUM procedures for multivariate data. Only two of these procedures will be considered in this thesis due to their relatively good ARL performance (Lowry et. al, 1992).

The first multivariate CUSUM that will be discussed was proposed by Crosier (1988). It will be denoted by MC1. The control scheme for the i'th point involves the following statistics:

$$C_i = [(\boldsymbol{S}_{i-1} + \boldsymbol{X}_i)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{S}_{i-1} + \boldsymbol{X}_i)]^{\frac{1}{2}}$$

where S_{i-1} is a vector of CUSUMs and C_i is a scalar.

$$\begin{aligned} \boldsymbol{S}_i &= \boldsymbol{0} && \text{if } C_i \le k_1 \\ &= (\boldsymbol{S}_{i-1}+\boldsymbol{X}_i)\left(1-\frac{k_1}{C_i}\right) && \text{if } C_i > k_1 \end{aligned}$$

where $\mathbf{S}_0 = \mathbf{0}$ and k_1 is a positive number. The MC1 procedure signals an out of control alarm whenever

$$T_i = [\boldsymbol{S}_i^T\boldsymbol{\Sigma}^{-1}\boldsymbol{S}_i]^{\frac{1}{2}} > h_2 \text{ such that } h_2 > 0.$$

k_1 and h_2 are the parameters of the chart and the ARL performance of the chart depends on their choice.

The other multivariate CUSUM scheme was introduced by Pignatiello and Runger (1990) and will be denoted here by MCUSUM. For the i'th observation, the procedure uses the quantities

$$\boldsymbol{D}_i = \sum_{j=i-l_i+1}^{i} \boldsymbol{X}_j$$

and

$$MC_i = \max[0, (\boldsymbol{D}_i^T\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_i)^{\frac{1}{2}} - k_2 l_i] \quad \text{where} \quad \begin{aligned} l_i &= l_{i-1}+1 && \text{if } MC_{i-1} > 0 \\ l_i &= 1 && \text{if } MC_{i-1} \le 0 \end{aligned}$$

with $k_2 > 0$. MCUSUM gives an out of control signal when $MC_i > h_3$, a chosen positive number. As in the case of MC1, k_2 and h_3 are the parameters of the chart and are optimized to yield the desired ARL performance.

It is noted that the vector $\mathbf{X}_i$ in multivariate CUSUM or EWMA schemes may represent either individual observations or a sample mean vector if subgrouping is used. The variance-covariance matrix Σ is estimated in the same way as discussed for the Hotelling's T^2 chart; that is, S_p is used when there are subgroups and S_m is used otherwise.

Multivariate EWMA Procedures

Lowry et al. (1992) proposed a multivariate exponentially weighted moving average control procedure which is denoted by MEWMA. It is an extension of the univariate EWMA. Let $\mathbf{Z}_i$ represent the i'th vector of EWMAs

$$\mathbf{Z}_i = \mathbf{R}\mathbf{X}_i + (\mathbf{I}-\mathbf{R})\mathbf{Z}_{i-1} \quad \text{where } \mathbf{Z}_0 = \mathbf{0}.$$

$\mathbf{R}$ is a diagonal matrix of weighting constants. $\mathbf{R} = \text{diag}(r_1, r_2, \ldots, r_p)$ such that $0 < r_j \leq 1$, $j = 1,2,\ldots,p$. Note that r_j is analogous to λ in the univariate EWMA and that $\mathbf{X}_i$ represents either individual observations or sample means as discussed in the case of multivariate CUSUM. An out of control signal is given when

$$T_i^2 = \mathbf{Z}_i^T \Sigma_{Z_i}^{-1} \mathbf{Z}_i > h_4$$

where $h_4 > 0$ is a parameter chosen to provide a specified ARL. Σ_{Zi} is the variance covariance matrix of $\mathbf{Z}_i$. It is common to equate the weighting coefficients to each other.

$$r_1 = r_2 = \ldots = r_p = r$$

The MEWMA vectors become

$$\mathbf{Z}_i = r\mathbf{X}_i + (1-r)\mathbf{Z}_{i-1}$$

and the covariance matrix is given by

$$\Sigma_{z_i} = \frac{r[1-(1-r)^{2i}]}{2-r}\,\Sigma .$$

For large i, Lowry et al. (1992) showed that Σ_{zi} becomes

$$\Sigma_{z_i} = \frac{r}{2-r}\,\Sigma .$$

2.2 Testing the Correlation Structure for Independence

Before the implementation or use of any scientific method, whether it is in Statistical Process Control (SPC) or any other field, it is important to consider and check the validity of the underlying assumptions. In most SPC schemes, one of the most significant assumptions is that of within-subgroup independence. Therefore, it is essential to test the independence of the observations in order to decide whether the charts are appropriate or not. If lack of independence is detected, the traditional SPC methods, such as the Shewhart, CUSUM and EWMA charts, cannot be used directly and certain modifications have to be introduced.

In this section, detection and evaluation of any correlation structure within the sample subgroups is of interest. It is convenient to consider the subgroups as a multivariate data set which consists of p variables and n observations (Johnson and Wichern, 1988). The correlation matrix is an indicator of the extent of linear association between the variables. The correlation matrix, $\boldsymbol{\rho}$, is a pxp matrix that are defined as

$$\rho_{ij} = 1, \quad i = j$$

$$\rho_{ij} = \frac{cov(i,j)}{\sqrt{var(i)}\sqrt{var(j)}}, \quad i \neq j .$$

From the above, it is evident that the correlation matrix consists of diagonal elements of 1 and off diagonal elements that vary between -1 and 1. When ρ_{ij} $(i \neq j)$ is close to 0, it indicates that variables i and j have very slight linear association. On the other hand, when the absolute value of ρ_{ij} $(i \neq j)$ is close to 1, there is a high association between these two variables. If the p observations that constitute a subgroup were independent, $\boldsymbol{\rho}$ would have the structure:

$$\boldsymbol{\rho}_{pxp} = \begin{bmatrix} 1 & 0 & . & . & . & 0 \\ 0 & 1 & . & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & 1 \end{bmatrix}$$

This section mainly deals with statistical methods that could be used to test for the correlation structure shown above. The goal is to examine the possibility of the existence of within-subgroup correlation. An estimate, **R**, of the correlation matrix is required to perform the tests. **R** consists of sample correlation coefficients, r_{ij}, which are calculated as

$$r_{ij} = 1, \quad i = j$$

$$r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}}\sqrt{s_{jj}}} = \frac{\sum_{k=1}^{n} (X_{ik}-\bar{X}_i)(X_{jk}-\bar{X}_j)}{\sqrt{\sum_{k=1}^{n} (X_{ik}-\bar{X}_i)^2}\sqrt{\sum_{k=1}^{n} (X_{jk}-\bar{X}_j)^2}}, \quad i \neq j$$

Several statistical tests have been proposed to assess the correlation structure of a set of data. Three tests will be discussed here: the generalized likelihood ratio test, the partitioning method and the pairwise comparison method. The first two are based on a likelihood ratio approach; however, the partitioning method excels over the first method in that it allows the comparison of a cluster of variables with another cluster. The third method tests for pairwise independence by considering each correlation coefficient individually.

Before the first two tests are addressed, the statistical reasoning behind the likelihood ratio test is explained. The general likelihood ratio test involves the calculation of the maximum of the likelihood function under two different hypotheses and computation of the ratio of the two values. Let θ be a vector whose elements are the unknown population parameters. The likelihood function, $L(\theta)$, is the joint density of the vectors of observations $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ at their observed values $x_1, x_2, \ldots, x_n$. Let Θ indicate the parameter set where θ takes its value. Under the null hypothesis H_0: $\theta = \theta_0$, θ lies in a subset Θ_0 of Θ. The alternate hypothesis is H_1: $\theta \notin \Theta_0$. The likelihood ratio is calculated as

$$\Lambda = \frac{\max_{\theta \in \Theta_0} L(\theta)}{\max_{\theta \in \Theta} L(\theta)}$$

where max $L(\theta)$ denotes the maximum of the likelihood function under the null hypothesis that θ lies in Θ_0. The distribution of Λ can be determined and the calculated value of Λ can be compared to that distribution at a chosen level of significance.

2.2.1 The Generalized Likelihood Ratio Test

For the given problem, the null hypothesis H_0 and the alternate hypothesis H_1 are

$$H_0 : \boldsymbol{\rho} = \boldsymbol{I} = \begin{bmatrix} 1 & 0 & . & . & . & 0 \\ 0 & 1 & . & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & 1 \end{bmatrix} \quad \text{and} \quad H_1 : \boldsymbol{\rho} \neq \boldsymbol{I}.$$

Johnson and Wichern (1988) showed that the result of applying the likelihood ratio test gives

$$\Lambda = |\boldsymbol{R}|^{n/2}$$

It can be shown that when n is large, -2 ln Λ has the approximate distribution $\chi^2_{p(p-1)/2,\,\alpha}$. Thus, at the level of significance α, the null hypothesis H_0 is rejected in favour of H_1 if -2 ln $\Lambda > \chi^2_{p(p-1)/2,\,\alpha}$. Bartlett (1954) suggested the use of an alternate statistic of -2[1-(2p+11)/6n] ln Λ instead of -2 ln Λ for moderate values of n.

2.2.2 The Partitioning Method

In this method, the vector of random variables, $\mathbf{X}_{(px1)}$, is partitioned into q subsets: $\mathbf{X}_{1(p1x1)}$, $\mathbf{X}_{2(p2x1)}$, ..., $\mathbf{X}_{q(pqx1)}$, where

$$\sum_{i=1}^{q} p_i = p.$$

It is desired to test whether the variables in each subset are mutually independent from the variables in the other subsets.

Following the suggested partitioning, the sample mean vector **X**, the covariance matrix **S** as well as the correlation matrix **R** can be partitioned accordingly. For example, **R** can be arranged as

$$
\boldsymbol{R} = \begin{bmatrix} R_{11} & R_{12} & \cdot & \cdot & R_{1q} \\ R_{21} & R_{22} & \cdot & \cdot & R_{2q} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ R_{q1} & R_{q2} & \cdot & \cdot & R_{qq} \end{bmatrix}
$$

where $\mathbf{R}_{ij}$ is a (p_i x p_j) matrix.

The null hypothesis H_0 is that all the subsets are independent.

$$
H_0 : \ \Sigma_{ij} = \mathbf{0}_{(p_i x q_j)}, \ \ i \neq j
$$

Substitution of the maximum likelihood estimators into the appropriate likelihood functions yields the likelihood ratio (Karson ,1982)

$$
\Lambda = \frac{|\boldsymbol{R}|^{n/2}}{\prod_{i=1}^{q} |\boldsymbol{R}_{ii}|^{n/2}}
$$

where $\mathbf{R}_{ii}$ is the sample correlation matrix of the i'th subset. Define the number m as

$$
m = (n - \frac{3}{2}) - \frac{\left(p^3 - \sum_{i=1}^{q} p_i^3\right)}{3\left(p^2 - \sum_{i=1}^{q} p_i^2\right)}
$$

The statistic $-m \ln \Lambda^{2/n}$ is approximately distributed as $\chi^2_{v,\alpha}$ where

$$
v = \frac{\left(p^2 - \sum_{i=1}^{q} p_i^2\right)}{2}
$$

The null hypothesis is rejected if $-m \ln \Lambda^{2/n} > \chi^2_{v,\alpha}$. In this thesis, it is of interest to test whether each variable is independent from all the others to establish the existence of within subgroup correlation. Therefore, this test is appropriate when p = q and all the subsets consist of one variable only.

2.2.3 The Pairwise Comparisons Approach

This approach involves testing the independence of two variables at a time. Instead of examining the correlation matrix as a whole, each of its off diagonal elements is studied independently from the others. There are two methods that can be used and these are discussed below.

Bickel and Doksum (1965) proved that when $\rho_{ij} = 0$ $(i \neq j)$, the statistic

$$T_n = \frac{\sqrt{n-2}\,\hat{\rho}_{ij}}{\sqrt{1-\hat{\rho}_{ij}^2}}$$

has a t distribution with n-2 degrees of freedom. The null hypothesis is H_0: $\rho_{ij} = 0$ if $i \neq j$. H_0 is rejected if $T_n > t_{n-2,(1-\alpha)}$. Note that $\hat{\rho}_{ij}$ is the maximum likelihood estimator of ρ_{ij} which is given by the sample correlation coefficient, r_{ij}.

There is also another method that can be used to test for the significance of the correlation coefficients. Through this method, it is possible to construct confidence intervals for all of the pairwise correlation coefficients.

Let $h(\rho)$ denote Fisher's transformation which is given by the expression

$$h(\rho_{ij}) = \frac{1}{2}\ln\left(\frac{1+\rho_{ij}}{1-\rho_{ij}}\right), \qquad (i \neq j).$$

The quantity $(n-3)^{1/2}[h(\hat{\rho}_{ij}) - h(\rho_{ij})]$ has an approximate distribution of N(0,1). For large n,

$$P[\sqrt{n-3}\,|h(\hat{\rho}_{ij})-h(\rho_{ij})| \le z(1-\tfrac{1}{2}\alpha)] = 1-\alpha$$

This yields a confidence interval for ρ_{ij} at the 1-α level given by

$$\left[\tanh\left(h(\hat{\rho}_{ij})-\frac{z(1-\frac{1}{2}\alpha)}{\sqrt{n-3}}\right),\ \tanh\left(h(\hat{\rho}_{ij})+\frac{z(1-\frac{1}{2}\alpha)}{\sqrt{n-3}}\right)\right]$$

This method can be used for different purposes. First, it is a test of independence. If all the confidence intervals contain 0, then the null hypothesis that all the variables are mutually independent is not rejected. Second, this procedure can be used to test whether the equal correlation structure exists. This is achieved by observing whether all the confidence intervals overlap. However, this method is not expected to be rigorous due to the high level of uncertainty that is involved. Each confidence interval has an associated probability of 1-α. Since there are p(p-1)/2 such intervals, the combined probability becomes $(1-\alpha)^{p(p-1)/2}$, which decreases as p increases. For example, if there are 5 variables (p = 5) and the desired level of significance is 0.05, the combined probability of 5 intervals is $(1-0.05)^{(5)(5-1)/2} = 0.59$ while each of the individual confidence intervals has an associated probability of 0.95.

In practical situations, it may also be of interest to recognize if an equal correlation structure exists. In such a structure, all the off-diagonal elements of $\boldsymbol{\rho}$ are equal to a constant value ρ_0, such that

$$H_0 : \boldsymbol{\rho}_{(pxp)} = \boldsymbol{\rho}_0 = \begin{bmatrix} 1 & \rho_0 & . & . & \rho_0 \\ \rho_0 & 1 & . & . & \rho_0 \\ . & . & . & . & . \\ . & . & . & . & . \\ \rho_0 & \rho_0 & . & . & 1 \end{bmatrix}$$

Although this structure is not of critical importance in this thesis, three methods are cited if the need for their use arises. They are: Lawley's method (Lawley, 1963), Jennrich's method (Karson, 1982) and Wilk's Lambda test (Johnson and Wichern, 1988).

This concludes the review of the methods that examine the correlation structure within the subgroups. The next section explores the theoretical aspects of the proposed methods that accommodate the existence of within subgroup correlation.

2.3 Proposed Solutions

There are a variety of SPC charts that can be used when within subgroup correlation exists. Multivariate control charts without subgrouping can be employed with no modifications since they are designed to monitor correlated quality variables. However, univariate control charts must be adjusted to compensate for the subgroup correlation structure. Determination of the principal components that describe the variability of the elements of the subgroup and implementing SPC charts on these variables is yet another approach.

In general, within-subgroup correlation arises when one product quality variable is measured at several fixed locations on the same item. One way of handling the problem is to construct p univariate schemes that monitor measurements at each individual location such as I charts. This approach is sensitive to changes in only a subset of the p measurements. However, since the combined probability of the charts for detecting a shift is $(1-\alpha)^p$, this approach is not effective in detecting shifts on the whole item and may give misleading results. A better alternative is to use subgrouping and utilize the averages of the subgroups to monitor the PQV in a univariate chart which is adjusted to accommodate the existing correlation structure. This approach is explained below.

2.3.1 Modified Univariate Control Charts

The importance of the assumption of within subgroup independence is manifested in the evaluation of $\sigma_{\bar{X}}$ from an estimate of σ, the population standard deviation. To demonstrate the effect of subgroup correlation on $\sigma_{\bar{X}}$, the true expression for $\sigma_{\bar{X}}$ is derived below.

Let Y be a linear combination of two variables, X_1 and X_2 with means μ_1 and μ_2 and standard deviations σ_1 and σ_2 respectively.

$$Y = a_1 X_1 + a_2 X_2$$

Let E(Y) denote the expected value of the random variable Y. The mean of Y is given by

$$\begin{aligned} \mu_Y = E(Y) &= E(a_1 X_1 + a_2 X_2) \\ &= a_1 E(X_1) + a_2 E(X_2) \\ &= a_1 \mu_1 + a_2 \mu_2 \end{aligned}$$

By definition, the variance of Y is

$$Var(Y) = E(Y - \mu_Y)^2$$

Substituting for Y and μ_Y,

$$
\begin{aligned}
Var(Y) &= E[\,a_1X_1 + a_2X_2 - a_1\mu_1 + a_2\mu_2]^2 \\
&= E[\,a_1\,(X_1 - \mu_1) + a_2\,(X_2 - \mu_2)\,]^2
\end{aligned}
$$

Expanding the terms,

$$
\begin{aligned}
Var(Y) &= E[a_1^2(X_1-\mu_1)^2 + a_2^2(X_2-\mu_2)^2 + 2a_1a_2(X_1-\mu_1)(X_2-\mu_2)] \\
&= a_1^2E[X_1-\mu_1]^2 + a_2^2E[X_2-\mu_2]^2 + 2a_1a_2E[(X_1-\mu_1)(X_2-\mu_2)] \\
&= a_1^2\,Var(X_1) + a_2\,Var(X_2) + 2\,a_1a_2\,Cov(X_1,X_2)\,.
\end{aligned}
$$

Similarly, if Y is a linear combination of p variables, X_1, X_2, ...X_p, with means μ_1, μ_2, ..., μ_p and standard deviations σ_1, σ_2, ..., σ_p respectively, the variance of Y is given by

$$
\begin{aligned}
Var(Y) = {} & a_1^2\,Var(X_1) + a_2^2\,Var(X_2) + \ldots + a_p^2\,Var(X_p) + \\
& 2a_1a_2\,Cov(X_1,X_2) + 2a_1a_3\,Cov(X_1,X_3) + \ldots + \\
& 2a_{p-1}a_p\,Cov(X_{p-1},X_p)
\end{aligned}
$$

The subgroup mean $\overline{X}$ is a linear combination of p variables such that $a_1 = a_2 = \ldots = a_p = 1/p$.

$$
\overline{X} = \frac{X_1 + X_2 + \ldots + X_p}{p} = \left(\frac{1}{p}\right)X_1 + \left(\frac{1}{p}\right)X_2 + \ldots + \left(\frac{1}{p}\right)X_p
$$

Hence,

$$
\begin{aligned}
Var(\overline{X}) &= \left(\frac{1}{p}\right)^2 Var(X_1) + \left(\frac{1}{p}\right)^2 Var(X_2) + \ldots + \left(\frac{1}{p}\right)^2 Var(X_p) \\
&\quad + 2\left(\frac{1}{p}\right)^2 Cov(X_1,X_2) + 2\left(\frac{1}{p}\right)^2 Cov(X_1,X_3) + \ldots \\
&\quad + 2\left(\frac{1}{p}\right)^2 Cov(X_{p-1},X_p) \\
&= \left(\frac{1}{p}\right)^2 (Var(X_1) + Var(X_2) + \ldots + Var(X_p)) \\
&\quad + 2\left(\frac{1}{p}\right)^2 [Cov(X_1,X_2) + Cov(X_1,X_3) + \ldots + Cov(X_{p-1},X_p)] \\
&= \frac{1}{p^2}\,(\sigma_{11} + \sigma_{22} + \ldots + \sigma_{pp} + 2\,[\sigma_{12} + \sigma_{13} + \ldots + \sigma_{p-1,p}])
\end{aligned}
$$

where σ_{ii} is the variance of variable X_i and σ_{ij} is the covariance of the random variables X_i and X_j for $i = 1,2,\ldots,p$ and $j = i+1,\ldots,p$.

Since $X_1, X_2, \ldots, X_p$ are measurements of the same variable on the same surface, it is reasonable to assume that they have the same variance σ^2, such that $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_p^2 = \sigma^2$. The variance of $\overline{X}$ thus becomes:

$$\begin{aligned} Var(\overline{X}) &= \frac{1}{p^2}\left[(\sigma^2+\sigma^2+\ldots+\sigma^2)+2(\sigma_{12}+\sigma_{13}+\ldots+\sigma_{p-1,p})\right] \\ &= \frac{1}{p^2}\left[p\sigma^2+2(\sigma_{12}+\sigma_{13}+\ldots+\sigma_{p-1,p})\right] \\ &= \frac{1}{p}\sigma^2+\frac{2}{p^2}(\sigma_{12}+\sigma_{13}+\ldots+\sigma_{p-1,p}) \end{aligned}$$

When $X_1, X_2, \ldots, X_p$ are independent, $\sigma_{ij} = 0$ where $i = 1, 2, \ldots, p$ and $j = 1,2,\ldots,p$ such that $i \neq j$. Then,

$$Var(\overline{X}) = \frac{1}{p}\sigma^2$$

and $$\sigma_{\overline{X}} = \sqrt{Var(\overline{X})} = \frac{1}{\sqrt{p}}\sigma.$$

This expression is used by all the univariate charts discussed in section 2.1, under the assumption of within-subgroup independence. When this assumption is violated, $\sigma_{\overline{X}}^2$ is under-estimated by the term $2(\sigma_{12} + \sigma_{13} + \ldots + \sigma_{p-1,p})/p^2$. Therefore, the true estimate of $\sigma_{\overline{X}}$ should be used to correct for the presence of within subgroup correlation. The expression for the variance of $\overline{X}$ can be simplified further as

$$Var(\overline{X}) = \sigma^2\left[\frac{1}{p}+\frac{2}{p^2}\left(\frac{\sigma_{12}}{\sigma^2}+\frac{\sigma_{13}}{\sigma^2}+\ldots+\frac{\sigma_{(p-1),p}}{\sigma^2}\right)\right]$$

Recall that the correlation, ρ_{ij}, between the variables X_i and X_j is

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{Var(X_i)}\sqrt{Var(X_j)}}.$$

For the given situation, $Var(X_i) = Var(X_j) = \sigma^2$ for $i = 1, \ldots, p$ and $j = 1, \ldots, p$. Thus, $\rho_{ij} = \sigma_{ij}/\sigma^2$. Substitution for ρ_{ij} in $Var(\overline{X})$ gives

$$Var(\overline{X}) = \sigma^2\left[\frac{1}{p} + \frac{2}{p^2}(\rho_{12} + \rho_{13} + \ldots + \rho_{p-1,p})\right].$$

Let $\overline{\rho}$ be the average of all the off-diagonal elements of the correlation matrix of X_1, X_2, $\ldots$, X_p.

$$\overline{\rho} = \frac{\sum_{i=1}^{p}\sum_{\substack{j=1\\ i\neq j}}^{p}\rho_{ij}}{\frac{1}{2}p(p-1)}$$

The variance of $\overline{X}$ can be rearranged as

$$\begin{aligned} Var(\overline{X}) &= \sigma^2\left(\frac{1}{p} + \frac{2}{p}\cdot\frac{1}{2}p(p-1)\left[\frac{\rho_{12} + \rho_{13} + \ldots + \rho_{p-1,p}}{\frac{1}{2}p(p-1)}\right]\right) \\ &= \frac{\sigma^2}{p}[1 + (p-1)\overline{\rho}] \end{aligned}$$

$$\text{or} \quad \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{p}}[1 + (p-1)\overline{\rho}]^{\frac{1}{2}}.$$

Neuhardt (1986) reported that the control limits of the traditional $\overline{X}$ chart are under-estimated whenever the observations within a subgroup are correlated with $\overline{\rho} > 0$, and suggested the use of the true standard deviation of $\overline{X}$, $\sigma_{\overline{X}}$, to correct the performance of these charts. Burn (1988) gave a similar recommendation. If the correct estimate of $\sigma_{\overline{X}}$ is used, then the corrected limits of the $\overline{X}$ charts are

$$\mu \pm 3\sigma\left[\frac{1+(p-1)\overline{\rho}}{p}\right]^{\frac{1}{2}}$$

if the true μ, σ and ρ_{ij} are known. When a training set is used to estimate the parameters of the population, $\sigma_{\overline{X}}$ can be estimated by one of two methods.

1) The average, $\overline{r}$, of the off-diagonal elements of the sample correlation matrix is used as an estimator of $\overline{\rho}$ and a correction factor C is applied to the sample standard deviation s of the population:

$$\hat{\sigma}_{\overline{X}} = \frac{s}{\sqrt{p}}C \quad \text{where} \quad C = [1+(p-1)\overline{r}]^{\frac{1}{2}}.$$

2) The sample standard deviation of the subgroup averages, $s_{\overline{X}}$, is simply used to estimate $\sigma_{\overline{X}}$.

Using the correction factor C to compensate for within subgroup correlation, the control limits of the various univariate charts are determined as follows:

- The respective control limits of the $\overline{X}$ chart when R and s are used to estimate σ are:

$$\overline{\overline{X}} \pm 3\frac{s}{\sqrt{p}}C \quad \text{or} \quad \overline{\overline{X}} \pm 3\, s_{\overline{X}}.$$

Note that the factors such as A_2 and A_3 can no longer be applied because they are based on the assumption of independent subgroups.

- In the CUSUM procedure, the $\overline{X}$'s are standardized as:

$$z_i = \frac{\sqrt{p}}{C}\cdot\frac{\overline{X}_i-\overline{\overline{X}}}{s} \quad \text{or} \quad z_i = \frac{\overline{X}_i-\overline{\overline{X}}}{s_{\overline{X}}}$$

- The control limits of the EWMA chart are expanded by applying the correction to the estimate of σ_{z_i} as follows

$$\hat{\sigma}_{z_i} = \frac{C}{\sqrt{\bar{p}}}\sqrt{(1-\lambda)^{2i}\sigma_{z_0}^{2} + \left(\frac{\lambda}{2-\lambda}\right)[1-(1-\lambda)^{2i}]}\ s$$

$$\text{or} \qquad \hat{\sigma}_{z_i} = \sqrt{(1-\lambda)^{2i}\sigma_{z_0}^{2} + \left(\frac{\lambda}{2-\lambda}\right)[1-(1-\lambda)^{2i}]}\ s_{\bar{X}}.$$

Note that for $\bar{\rho} > 0$, the control limits of the $\bar{X}$ and EWMA charts are expanded around the center line. This expansion is expected to reduce the number of false alarms that are typical of such situations. The R and s charts are not affected by within subgroup correlation (Burn, 1988) because the R and s statistics do not involve any estimation of $\sigma_{\bar{X}}$. Therefore, no correction is required for these charts.

2.3.2 Principal Components Approach

This approach is based on the concept that if the variability conveyed by the p components of the subgroup can be compressed into new fewer independent variables, then these variables can be used to monitor the level and variability of the product quality variable. This approach has the potential of reducing the dimensionality of the data when there is a correlation structure. Principal Components Analysis (PCA) is a data reduction technique that is capable of finding the new variables. Note that the nomenclature and data structure in this section is analogous to that defined for multivariate control charts.

Principal Components Analysis (PCA) is a multivariate procedure that endeavours to explain the variability within a multivariate data set arranged in a matrix $\mathbf{X}_{p \times n}$. PCA determines the first principal component, Y_1, which is a new variable that describes the direction of the greatest variability. The direction of the second principal component,

Y_2, is orthogonal to that of Y_1 and explains the greatest part of the remaining variability. Proceeding in this manner, p principal components can be obtained. These principal components rotate the axes of **X** into a new orthogonal frame of reference. Usually, q principal components explain most of the variability in the data set where $q << p$. Therefore, the dimensionality of the original data set is reduced from p to a much smaller q (Johnson and Wichern, 1988).

Let $\mathbf{\Sigma}$ be the variance-covariance matrix associated with **X**. Let $\mathbf{\Sigma}$ have the eigenvalue-eigenvector pairs (λ_1, e_1), (λ_2, e_2), ..., (λ_p, e_p) such that $\lambda_1 > \lambda_2 > \ldots > \lambda_p > 0$. The i'th principal component, Y_i, is:

$$Y_i = \boldsymbol{e}_i^T \boldsymbol{X} = e_{1i}X_1 + e_{2i}X_2 + \ldots + e_{pi}X_p \qquad (i = 1, 2, \ldots, p).$$

Each "loading", e_{ij} ($j = 1,2, \ldots, p$), is a measure of the importance of variable X_j to Y_i. The principal components have the following characteristics:

$$Var(Y_i) = \lambda_i \quad \text{and} \quad Cov(Y_i, Y_j) = 0 \quad (i, j = 1, 2, \ldots, p, \quad i \neq j)$$

The proportion of the total variance due to the i'th principal component is:

$$\frac{\lambda_i}{\sum_{j=1}^{p} \lambda_j}, \qquad i = 1, 2, \ldots, p$$

Based on this proportion, it is possible to retain the few principal components which capture a specified percentage of the variation in the original data set. Thus, the chosen principal components can be used to transmit, in a lower dimensional space, most of the information available from the original data block.

Kresta et al. (1991) suggested the use of PCA to compress vast amounts of process information into lower dimensional spaces. Multivariate SPC procedures, similar to univariate Shewhart charts, are then implemented using the retained principal components. This concept can be utilized in this research for the case of correlated subgroups. If one principal component captures most of the variability conveyed by the p subgroup components, then an I chart or its CUSUM/EWMA analogue can be implemented to monitor the PQV. If more than one principal component are required, then multivariate charts are used on these new variables.

2.3.3 Multivariate Control Charts

The use of multivariate control charts is proposed to monitor processes which involve correlated subgroups. Each component of the previously defined univariate subgroup is treated as a separate quality characteristic and the px1 vector, $\mathbf{X}_i$, is used to represent the measurements of the i'th subgroup. A multivariate control chart such as Hotelling's T^2, MCUSUM, MC1 and MEWMA can monitor the level of the PQV. These charts are used as multivariate analogues of Shewhart's I chart since there are no multivariate subgroups in the sampling scheme shown in Chapter 1.

This chapter reviewed the existing SPC charts and potential methods for identifying and handling the existence of within-subgroup correlation. Since the theoretical aspects of the research have been established, the next chapter discusses the methodology and the experimental considerations for examining the applicability and effectiveness of the proposed methods.

Chapter 3

METHODOLOGY AND EXPERIMENTAL ASPECTS

This chapter describes the methodology and the experimental strategy of the research. The basic objectives of this thesis are: 1) to identify the existence of within subgroup correlation and 2) to find adequate SPC schemes that account for this correlation. Sections 2.2 and 2.3 reviewed several methods that have the potential of achieving these objectives.

In order to evaluate the effectiveness of these methods when used in real processes, their performance should be measured on an appropriate scale. In this research, the methods are applied to simulated data to determine their power and characteristics. Once their properties are known, they are then applied to real process data from the semiconductor manufacturing industry to confirm the validity of the results obtained from the computer simulations. Simulated data has the advantage of having known characteristics that can be adjusted easily and the derived results involve small amounts of experimental error. The use of real process data will demonstrate the usefulness and power of the proposed methods in the real manufacturing environment.

This chapter is divided into two sections. The first section deals with the experimental considerations for the tests that identify the presence of within-subgroup correlation. The second section is concerned with the experimental aspects associated with the various SPC procedures that handle such a correlation.

It is noted that all the methods discussed in section 2.2 as well as all SPC charts are essentially statistical tests of hypotheses. An important concept in such tests is that of type I and type II errors. If H_0 denotes the null hypothesis to be tested versus the alternate hypothesis H_1, then type I error refers to the error of accepting H_0 while, in fact, it is not true. Type II error is associated with the reverse situation where H_0 is rejected when in reality it is true. For an ideal test of hypothesis, the probabilities of having type I and type II errors are zero. In this thesis, the performance of the various tests and methods will be assessed using these concepts, as discussed below.

3.1 Tests of Independence

Given a pxn multivariate data set representing n observations of p random variables, which, in this research, denote the p measurements that constitute a univariate sample. It is desired to determine whether these variables are independently distributed. The correlation matrix, $\boldsymbol{\rho}$, is an indicator of the pairwise linear association between the variables as discussed previously in section 2.2. When the p random variables are independent, all of the off-diagonal elements of $\boldsymbol{\rho}$ are zero.

If the true $\boldsymbol{\rho}$ is known, then it would be obvious if its off-diagonal elements are equal to zero, or equivalently whether the variables are independent or correlated. In practice, the true correlation matrix is usually unknown and is determined from a set of observations. The sample correlation matrix **R** is an estimate of $\boldsymbol{\rho}$. Section 2.2.1 lists four methods that test for the independence of the p random variables based on **R**. Namely, these methods are: the generalized likelihood ratio test, the partitioning method, the pairwise t-test and the confidence interval method.

One of the objectives of this research was to identify the existence of within subgroup correlation. Theoretically, the above mentioned methods are capable of testing for independence. However, the literature sources for these methods reveal no indication of their power when they are applied to real data sets of varying sizes and degrees of correlation. It is of interest to determine their potential and estimate their performance at varying conditions. The probability of obtaining type I and type II errors is a good indicator of the performance of these tests.

The experimental strategy for estimating the power of these tests and comparing their performance involves the use of computer simulations. A pxn multivariate data set consisting of p identically, independently and Normally distributed random variables was simulated. The Normal distributions were generated with mean 0 and standard deviation 1. A specified correlation structure was injected into the data set as described in Appendix A. The four tests of independence were performed and their success or failure in accepting the null hypothesis was recorded. A new data set having the same characteristics was then generated, the four tests of independence were again applied and their success or failure was recorded. This procedure was repeated 5000 times. After the 5000 trials, the number of successes for each test was counted and divided by 5000 to give the probability of obtaining type I or type II errors. If the magnitude of the injected correlation is 0, then the calculated probability would be (1 - probability of type I error); otherwise, it would be that of type II. If both of these probabilities are close to 0, then the test under study is very powerful in recognizing within subgroup independence.

The parameters varied in the experiments described above were the number of variables or the subgroup size, p, the degree of correlation, ρ_0, and the number of observations,n, which was used to compute **R**. In practice, **R** is estimated from the same training set that is used to determine the control limits of the SPC chart that is to be implemented, that is, $n = k$. Thus, in this study, it is sufficient to compare the performance of the independence tests at $k = 20$ and $k = 50$ because these are the typical training set sizes. With respect to p, most univariate SPC charts use a subgroup size of 4 to 6 observations and hence p was set to 4, 5 or 6. Furthermore, the pairwise correlation coefficient, ρ_{ij} $(i \neq j)$, is a number between -1 and 1. In the situation of concern in this thesis, where multiple measurements of the same variable are taken from a single item, negative correlations are very unlikely to occur because that item undergoes identical processing conditions. Thus, ρ_{ij} was varied between 0 and 1 in the experiments. Table 3 displays a summary of the experimental conditions for measuring the power of the independence tests. The flowchart for the computer simulations is shown in Figure 8.

Table 3. Summary of parameter settings for testing independence

Parameter	values
$n = k$	20, 50
p	4, 5, 6
ρ_{ij}	0, 0.05, 0.1, 0.3, 0.5, 0.9

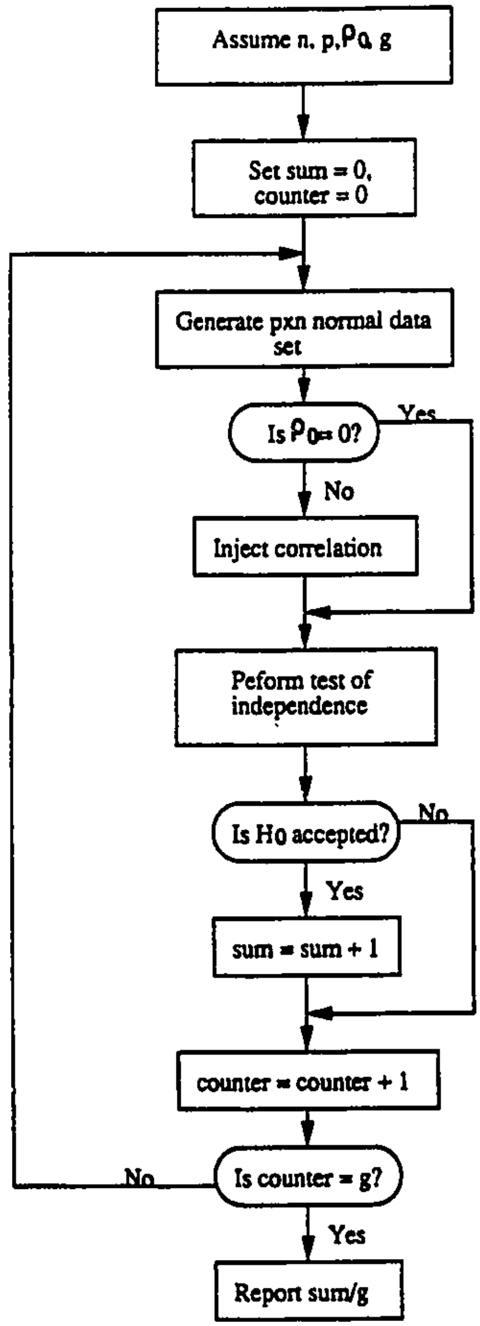


Figure 8. Flowchart for simulations of the tests of independence

3.2 Proposed SPC Charts

Section 2.3 proposed three approaches to implement statistical process control on correlated samples. These approaches are: modified univariate charts, principal components charts and multivariate charts. Table 4 lists these approaches and the associated procedures. One of the goals of this research was to select the most suitable chart for use in the presence of within subgroup correlation. Therefore, an appropriate criteria had to be chosen to analyze the performance of each chart.

Table 4. SPC procedures used in the experiments

Proposed approach	Procedure
Modified univariate charts	$\overline{X}$, CUSUM, EWMA
Principal components	(depends on number of components)
Multivariate charts	Hotelling's T^2, MC1, MCUSUM, MEWMA

As mentioned earlier, all SPC charts are repeated tests of hypotheses about the level of a process. When a point is plotted on a chart, a test of hypothesis is performed to determine if it lies within the expected bounds of variation. The null hypothesis, H_0, is that the monitored process is in a state of statistical control. When H_0 is rejected, an out of control flag is raised. Type I error occurs when the process goes out of control and the chart fails to signal an alarm. Similarly, type II error takes place when the chart gives a false alarm, indicating that a significant event has occured when in fact the process is in control.

Determination of the probability of type I and type II errors is a good indicator of the efficiency and usefulness of a chart; however, it is does not give a complete profile of its performance. While the probability of type II error is sufficient to determine the expected number of false alarms, it does not reveal any information with regards to the magnitude of the shift that occurs and the speed of detecting that shift. A more meaningful alternative to evaluate a chart is the use of run lengths.

Recall from Chapter 2 that the run length of a chart is the number of points plotted before an alarm is signalled. The average run length (ARL) is the expected value of the run length, which is considered as a random variable with its own probability distribution. ARL(0) denotes the ARL when the process is in control and is an indicator of the frequency of false alarms. On the other hand, ARL(Δ) is the ARL when a shift of size Δ has occurred. The "perfect" control chart has an ARL(0) of ∞ and ARL(Δ) of 0 where $\Delta \neq 0$, that is, it does not have errors of types I and II. Such a chart does not exist in practice; however, other charts can be compared to it in order to assess their efficiency. Thus, a long ARL(0) and a short ARL(Δ) are desirable in a control chart when Δ is small. Unfortunately, ARL(0) and ARL(Δ) are correlated and a chart that is sensitive to shifts is usually prone to frequent false alarms. As the distribution of run lengths tend to be positively skewed, it is appropriate to report the median run length (MRL) as an additional performance criteria. Although the ARL and the MRL are expected to follow the same trends, both of them were recorded in this study to avoid any misinterpretations. Note that any subsequent discussion involving the ARLs of a chart in this chapter is equally valid for the MRLs.

Based on the above discussion, the average run length and the median run length of a chart will be used in the research to evaluate the various charts. The chart with the longest in control ARL and shortest ARL(Δ) where Δ is a specified shift, is deemed to have the best performance. It is conventional in SPC to measure Δ in terms of σ, the standard deviation of the population. The magnitude of Δ depends on the nature of the monitored process, the sensitivity it requires as well as the speed of detecting the shift. Therefore, the effect of Δ on the ARL was studied with $\Delta = 0, 0.5\sigma, 1.0\sigma$ and 1.5σ. These values of Δ represent small to moderate magnitudes of the shift. They are reported in terms of σ, the standard deviation of the expected random error.

The basic experimental strategy for assessing the various SPC schemes was similar to that described in the previous section. Simulated data was used to study the characteristics of each chart as well as the factors that affect its performance. Then, the charts were implemented on real process data to verify the results derived from the simulations. Computer simulations are valuable experimental tools in SPC mainly because they permit easy manipulation of the characteristics of the problem. In addition, while simulations account for the random error that is introduced by the materials and equipment, they are free of biased errors, such as those associated with malfunctioning or breakdown in the experimental equipment.

There are a variety of factors that have an impact on the run length of a control chart. If a control scheme is to be implemented on a process, its control limits have to be determined based on the distribution of the product quality variable. Two scenarios are possible: 1) the true parameters that describe the distribution, such as μ and σ, are known exactly, or 2) these parameters are unknown and have to be estimated from a training set of size k. As k increases, the estimates of these parameters approach their true values. In fact, the first case corresponds to an infinitely large training set ($k \rightarrow \infty$). Hence, k is expected to influence the run length of a control chart since it has an impact on its control limits. In this study, the effect of the choice of k on the performance of a chart was investigated. The values of k studied were 20, 50 and ∞, the latter being the case where the true parameters of the distribution of the PQV are known. Another factor related to the estimation of these parameters is p, the univariate subgroup size or the multivariate number of characteristics. In practical terms, p is the number of measurements per item as described in Chapter 1. Typical values of p = 4, 5 and 6 were studied in this research.

Section 2.3.1 reveals that when the existence of within subgroup correlation is ignored, $\sigma_{\bar{X}}$ is underestimated by a factor of $[1 + (p-1)\bar{\rho}]^{0.5}$. This means that the uncorrected control limits of univariate charts will be tighter relatively as the magnitude of the correlation increases. Consequently, the ARL(0) of these charts will be shorter and the frequency of false alarms increases with correlation. It was of interest to study the effect of the degree of correlation on the ARLs of the various proposed control schemes as well as the traditionally used charts. The goal of this investigation was to select a chart that is robust to within-subgroup correlation.

For simplicity, the pairwise correlation coefficients, ρ_{ij}, between the p measurements will be assumed equal to a common positive value, ρ_0. This assumption is justified by the fact that it often holds true in real situations where within subgroup correlation exits, such as a wafer surface in semiconductor fabrication. Furthermore, the proposed correction for univariate charts depends solely on the average of the correlation coefficients, $\overline{\rho}$, and not on the variability between the individual coefficients. Therefore, an additional factor to be examined in this research is the extent of correlation, ρ_0, which will be varied between 0 and 1.

There is one extra factor that was worth investigating when modified univariate charts are studied. This factor is the correction method. Recall from section 2.3.1 that there are two methods to estimate $\sigma_{\bar{x}}$ when a training set is used. The first one involves the calculation of a correction factor C which is a function of $\overline{r}$, the average of the off-diagonal elements of the sample correlation matrix, **R**. The second method employs an estimate of $\sigma_{\bar{x}}$, $s_{\bar{x}}$, calculated from the k subgroup averages in the training set. The use of each of these methods in correcting univariate charts or no correction at all was examined in this research.

Experimental computer simulations were performed in a similar fashion as described for the tests of independence. A pxn multivariate normal set of data was generated and a uniform correlation of magnitude ρ_0 was injected into the data set. A specific control scheme was adopted to monitor the hypothetical process. When an alarm was signalled, the run length was recorded. The same procedure was repeated 5000 times so that the run length distribution generated precise average run lengths (ARL) and median run lengths (MRL). Figure 9 shows a typical run lengths distribution of a simulated chart. A list of the factors that were considered in these simulations is given in Table 5 and the experimental procedure for the simulations is outlined in Figure 10.

All of the computer simulations were performed using the RS1 statistical package on a VAX/VMS mainframe system. The simulation programs are listed in Appendix B.

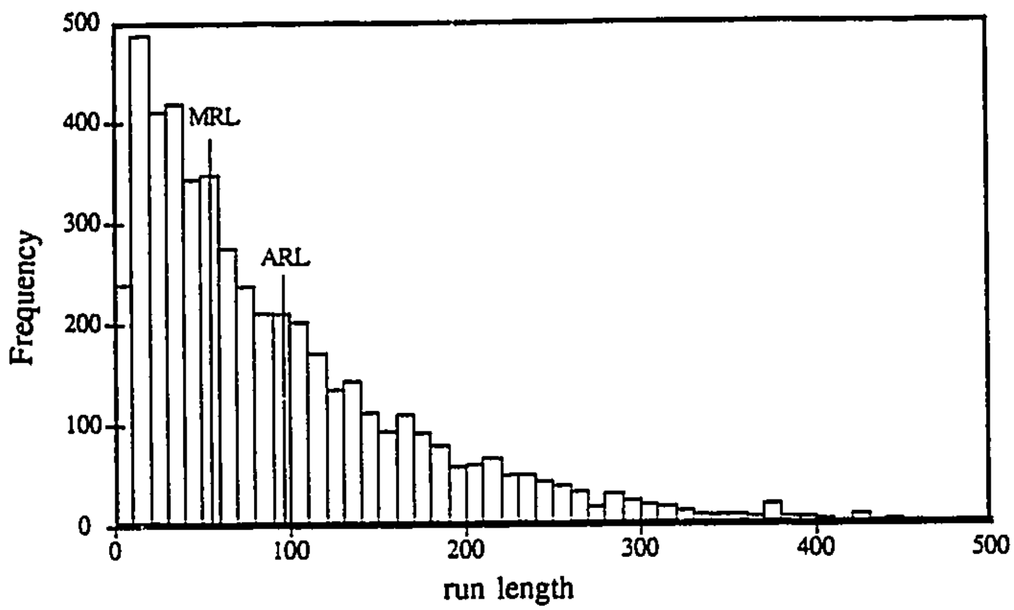


Figure 9. A typical run length distribution of a simulated chart

Table 5. Summary of factors involved in SPC chart simulations

Factor/parameter	Value
Training set size, k	0, 20, 50
sample/subgroup size, p	4, 5, 6
degree of correlation, ρ_o	0, 0.3, 0.5, 0.95
magnitude of shift, Δ	0, 0.5σ, 1.0σ, 1.5σ
method of correction (univariate charts only)	none, C, $s_{\bar{x}}$

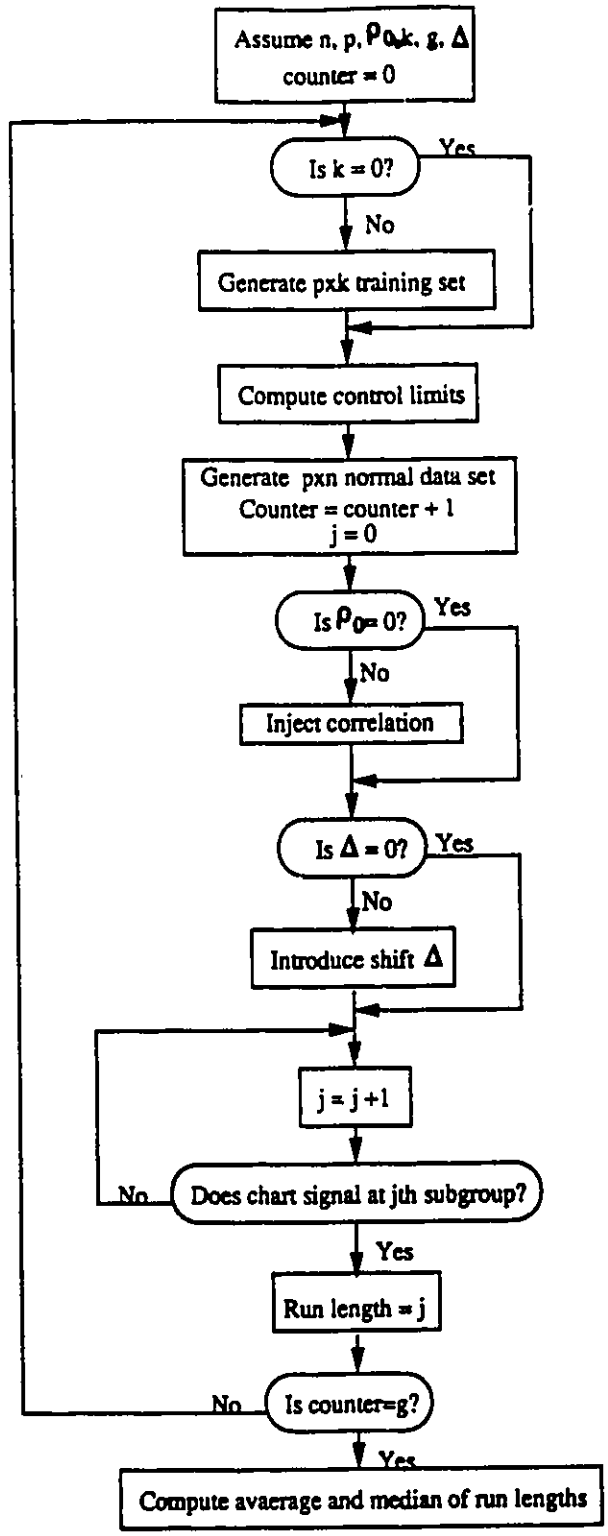


Figure 10. Flowchart for the simulations of the control charts

Chapter 4

RESULTS AND DISCUSSION

In this chapter, the results of the research are discussed and evaluated. Section 4.1 addresses the tests of independence and their power in detecting the existence of within sample correlation. Section 4.2 examines the performance of the various SPC schemes that are proposed to account for such a correlation. Finally, section 4.3 compares the power of these schemes in order to determine the best chart.

4.1 Tests of Independence

The results of simulating the tests of independence are shown in Tables 6 through 9. The values displayed in these tables are the calculated percentage of accepting the null hypothesis that the observations constituting a sample are independent of each other. These values are representative of the estimated probability of type I and type II errors for each test according to whether ρ_0 is 0 or not. As discussed in Chapter 3, for the "ideal" test, the probability of accepting H_0 is 100% at $\rho_0=0$ and 0% elsewhere, corresponding to 0% type I and type II errors respectively.

Table 6 shows the results of the generalized likelihood ratio test for training sets of 20, 50 and 100. In general, the probability of accepting H_0 is low for this test when $\rho_0 = 0$ for training sets of different sizes. As the size of the subgroup, p, increases, the probability of accepting H_0 increases at the various ρ_0 values. This implies that the probability of type I error drops while that of type II rises. The same effect is observed when the training set size, k, increases. Increasing the size of the pxk data set which estimates the sample correlation matrix, **R**, is expected to improve the performance of the test since **R** approaches ρ_0. However, the results show that a decrease in the probability of type I error is accompanied by an increased occurrence of type II error.

The simulated results of the partitioning method are displayed in Table 7. This test is likely to verify that the observations within a sample are independent at various degrees of correlation. Although the probability of having type I error is 0, that of type II error is extremely high even if the true magnitude of the correlation coefficient is as big as 0.5. As in the generalized likelihood method, the percentage of accepting H_0 increases with increasing p and with increasing k if p is 5 or 6. When p = 4, some improvement in the performance of this test is achieved upon increasing the training set size at higher correlations. For example, when p = 4, Table 7 shows that while the percentage of acceptance of H_0 at $\rho_0 = 0$ remains 100% at k=20, 50 and 100, this value drops when $\rho_0=0.5$ from 17.9% at k=20 to 5.2% at k=100. In general, the results of the simulation show that the partitioning method is very likely to indicate that a number of variables are independent at low to moderately high values of correlation. In other words, this test will establish the existence of within-subgroup correlation only if ρ_0 is approximately 0.8 or higher.

Tables 8 and 9 show the results of the simulated pairwise methods using the t distribution and confidence intervals respectively. Both of these tables demonstrate that as p increases, the probability of accepting H_0 decreases at a variety of ρ_0 values, the opposite effect of what is observed for the generalized likelihood and partitioning tests. Also, as k increases, the probability of type II error decreases at the higher values of ρ_0

Table 6. Percentage of acceptance of the Generalized Likelihood Method for H_0: $\boldsymbol{\rho} = \mathbf{I}$

(a) $k = 20$

ρ_0	$p = 4$	$p = 5$	$p = 6$
0.0	21.2	31.78	44.1
0.05	20.3	30.14	41.3
0.10	16.7	24.3	31.4
0.30	2.44	3.28	3.98
0.50	0.08	0.10	0.06
0.90	0.00	0.00	0.00

(b) $k = 50$

ρ_0	$p = 4$	$p = 5$	$p = 6$
0.0	25.6	41.7	57.1
0.05	20.9	33.9	47.9
0.10	12.7	19.1	27.9
0.30	0.04	0.00	0.04
0.50	0.00	0.00	0.00
0.90	0.00	0.00	0.00

(c) $k = 100$

ρ_0	$p = 4$	$p = 5$	$p = 6$
0.0	28.1	44.2	63.2
0.10	6.2	8.3	11.2
0.30	0.0	0.0	0.0
0.50	0.0	0.0	0.0

Table 7. Percentage of acceptance of the Partitioning Method for H_0: $\boldsymbol{\rho} = \mathbf{I}$

(a) $k = 20$

ρ_0	$p = 4$	$p = 5$	$p = 6$
0.0	100	100	100
0.05	100	100	100
0.10	100	100	100
0.30	74.4	97.6	100
0.50	17.9	58.4	91.7
0.75	0.1	1.4	9.28
0.90	0.0	0.0	0.0

(b) $k = 50$

ρ_0	$p = 4$	$p = 5$	$p = 6$
0.0	100	100	100
0.05	100	100	100
0.10	100	100	100
0.30	91.8	99.9	100
0.50	11.7	70.3	99.0
0.75	0.0	0.02	1.44
0.90	0.0	0.00	0.00

(c) $k = 100$

ρ_0	$p = 4$	$p = 5$	$p = 6$
0.0	100	100	100
0.05	100	100	100
0.30	98.4	100	100
0.50	5.2	81.0	99.9

Table 8. Percentage of acceptance of the Pairwise Method using the t distribtuion for H_0: $\boldsymbol{\rho} = \mathbf{I}$

(a)
$k = 20$

ρ_0	p = 4	p = 5	p = 6
0.0	98.5	95.1	91.2
0.05	98.4	96.9	92.1
0.10	99.2	97.6	94.5
0.50	78.8	86.7	91.1
0.75	6.2	8.2	11.3
0.90	0.02	0.02	0.02

(b)
$k = 50$

ρ_0	p = 4	p = 5	p = 6
0.0	98.2	95.4	90.9
0.05	98.9	96.7	92.6
0.10	99.4	98.2	96.1
0.30	88.5	95.2	97.9
0.50	13.2	19.9	26.9
0.90	0.0	0.0	0.0

(c)
$k = 100$

ρ_0	p = 4	p = 5	p = 6
0.0	98.5	95.7	89.5
0.30	99.7	99.1	98.0
0.50	0.04	0.30	0.30
0.75	0.00	0.00	0.00

Table 9. Percentage of acceptance of the Confidence Intervals Method for H_0: $\boldsymbol{\rho} = \mathbf{I}$

(a)
$k = 20$

ρ_0	p = 4	p = 5	p = 6
0.0	73.5	58.4	45.3
0.05	71.2	57.9	42.6
0.10	65.5	50.1	34.6
0.30	25.4	12.5	5.8
0.50	2.5	0.7	0.2
0.90	0.0	0.0	0.0

(b)
$k = 50$

ρ_0	p = 4	p = 5	p = 6
0.0	74.4	59.2	46.6
0.05	69.4	52.1	38.7
0.10	53.3	36.9	21.7
0.50	0.0	0.0	0.0
0.90	0.0	0.0	0.0

(c)
$k = 100$

ρ_0	p = 4	p = 5	p = 6
0.0	73.5	59.7	45.8
0.05	62.4	45.8	30.7
0.10	35.6	19.2	8.9
0.50	0.0	0.0	0.0

while that of type I error remains virtually unchanged. For instance, at $p = 5$, $\rho_0 = .5$ and $k = 20$, the percentage of accepting H_0 is 86.7% for the t-distribution method. When $k = 100$, the above percentage drops to 0.3%. Therefore, some improvement in terms of performance is achieved by increasing the training set size that estimates **R**. The confidence interval method has lower rates for accepting H_0 than the t-distribution method.

In order to simplify the comparison between the various methods, a plot of the percentage of acceptance of H_0 versus ρ_0 for the various methods at $p = 5$ and $k = 20$ in Figure 11. The partitioning and the t-distribution methods have similar behaviour with high acceptance rate of H_0 at low to moderately high values of ρ_0. In other words, these two methods do not differentiate between a correlation of 0 (independence) and 0.4. On the other hand, both the confidence interval and the generalized likelihood methods have low acceptance rates which decrease with increasing ρ_0. Both methods have a high probability of not recognizing a set of independent variables. In general, the power of the four methods in testing for independence is unsatisfactory and too far from being ideal. However, the partitioning and t-distribution methods can still be used at high degrees of correlation, as discussed below.

The partitioning and t-distribution methods reject H_0 only if ρ_0 is large. At small values of ρ_0, the probability of accepting H_0 is very high, in fact 100% for the partitioning method and in the 90-99% range for the t-distribution method, as shown in Tables 7 and 9. As ρ_0 increases, this probability decreases to become essentially 0% beyond certain values of ρ_0. Thus, these two methods can be used to identify the existence of relatively significant correlations. However, is such a situation, the correlation structure can be identified easily by inspection of the correlation matrix. Therefore, there is no significant advantage for the use of any of the tests of independence.

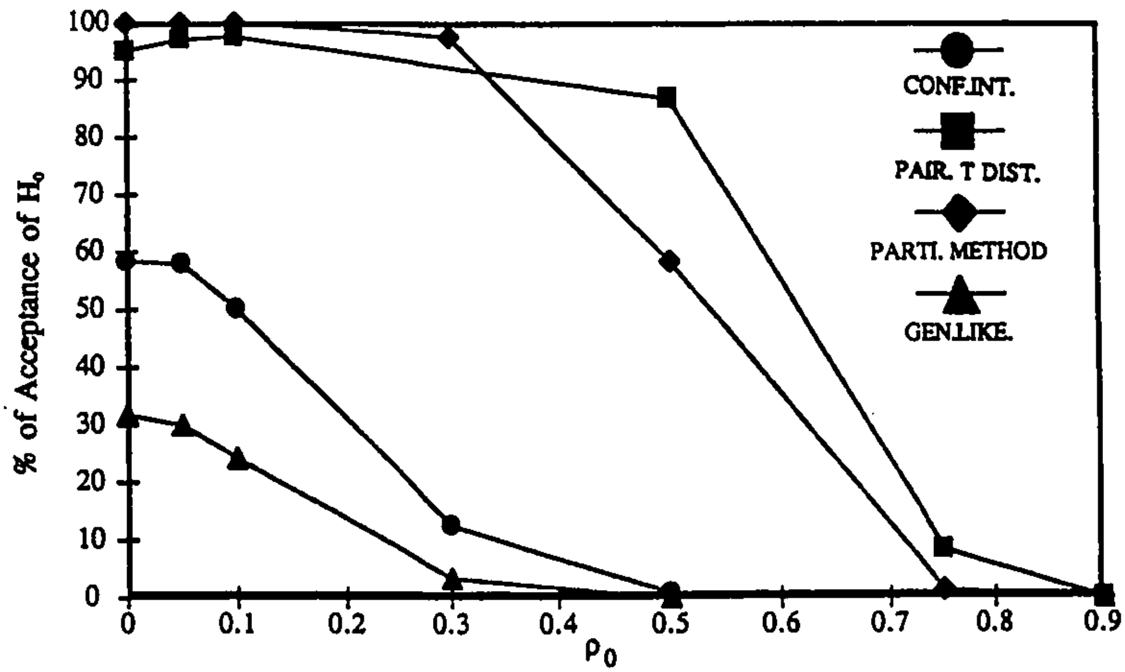


Figure 11. The performance of the tests of independence for p=5 and k=20.

Note that the above conclusions are valid only in the experimental range that was studied, that is, with p varying between 4 and 6 and k between 20 and 100. This experimental range serves the purpose of this study when the above tests of independence are used in the traditional SPC sample and training set sizes.

With respect to the error involved in the reported values, several simulations were replicated and the mean and standard deviation of each set of replicates were recorded. The results are shown in Table C1 in Appendix C. The variability in the results does not show any particular trends. Although the results in Tables 6, 7, 8 and 9 vary between 0 and 100%, none of the associated standard deviations exceeds 0.75%. For the generalized likelihood method, the standard deviation ranges between 0.036 and 0.75%, while the range for the partitioning method is 0 to 0.57%. Similarly, the standard deviation varies between 0.12 and 0.65% for the confidence interval method and between 0.01 and 0.45% for the t distribution method. The difference between the magnitudes of these numbers and those in Tables 6 to 9 confirms that the trends discussed in this section are true and not due to experimental error.

4.2 Proposed SPC Approaches

This section evaluates the various SPC methods that are proposed to deal with within-subgroup correlation. As discussed in Chapter 3, the average and median run lengths of each chart will be used in the evaluation. The objective of this section is to establish general trends in the behaviour of these charts at optimal parameter settings with varying degrees of correlation, subgroup and training set sizes when shifts of different magnitudes occur. After these trends are verified, the next step is to compare the performance of the proposed schemes in order to determine which is the best chart.

Before any discussion can be based on the simulated results, it is essential to consider the magnitude of the experimental error. For each chart, the simulations at a particular set of conditions were replicated and the averages and standard deviations of the ARLs and MRLs were recorded (denoted by $\overline{ARL}$,s_{ARL}, $\overline{MRL}$ and s_{MRL} respectively).

The procedure was repeated at several sets of conditions and the results were plotted in Figure 12. Examination of Figures 12(a) and 12(b) shows that, as the average ARL or MRL increases, the variability in the reported ARL and MRL values increases. The various charts display the same trend and no clear distinction can be made among them especially at lower run length values. This indicates that the experimental error can be considered as a percentage. Confidence intervals for each of the average ARL or MRL can be constructed at the 99.7% significance level as $\overline{ARL} \pm 3s_{\overline{ARL}}$ and $\overline{MRL} \pm 3s_{\overline{MRL}}$. A practical representation of the error is then $e_1 = 3s_{\overline{ARL}}/\overline{ARL}$ and $e_2 = 3s_{\overline{MRL}}/\overline{MRL}$. These values are shown in Table 10.

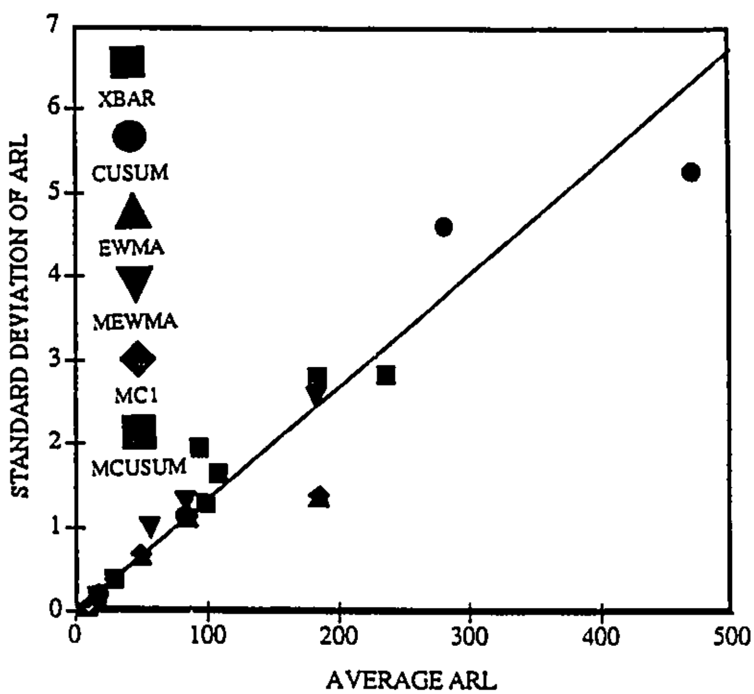


Figure 12(a). Variability trends in the ARL values.

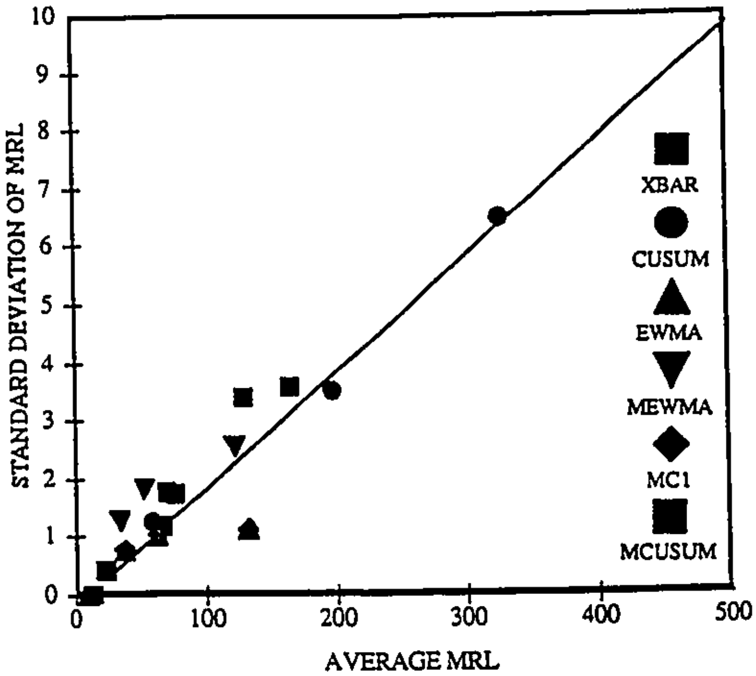


Figure 12(b). Variability trends in the MRL values.

Table 10(a) shows that the magnitude of error in the ARL values does not vary significantly and ranges between 1.6% and 5.8% with an average of 3.8%. On the other hand, Table 10(b) demonstrates that more variability is involved in e_2 that varies between 0.0% and 8.8% with an average at 4.3%. Both e_1 and e_2 are not chart dependent. In other words, the use of one chart does not imply more variability in the percentage of error in the run lengths, and e_1 and e_2 reflect only the error incurred by the simulations and not the control chart itself.

Table 10 (a) Percentage of error involved in ARL values.

X		EWMA		CUSUM		MEWMA		MCI		MCUSUM	
$\overline{ARL}$	$e_1(\%)$	$\overline{ARL}$	$e_1(\%)$	$\overline{ARL}$	$e_1(\%)$	$\overline{ARL}$	$e_1(\%)$	$\overline{ARL}$	$e_1(\%)$	$\overline{ARL}$	$e_1(\%)$
16.7	3.1	3.8	3.8	19.3	2.8	56.5	5.3	9.5	1.6	29.8	3.7
99.4	5.8	83.7	4.9	82.8	5.1	83.3	4.6	49.2	4.1	184.8	4.5
99.8	1.3	144.0	4.8	470.1	3.3	182.8	4.2	85.4	3.9	236.3	3.5
		271.7	4.4					184.7	2.4		
		307.0	4.1								

Table 10 (b) Percentage of error involved in MRL values.

X		EWMA		CUSUM		MEWMA		MCI		MCUSUM	
$\overline{MRL}$	$e_2(\%)$	$\overline{MRL}$	$e_2(\%)$	$\overline{MRL}$	$e_2(\%)$	$\overline{MRL}$	$e_2(\%)$	$\overline{MRL}$	$e_2(\%)$	$\overline{MRL}$	$e_2(\%)$
13.0	0.0	3.0	0.0	14.0	0.0	33.5	0.0	9.0	0.0	22.8	5.5
66.9	5.5	52.2	8.8	59.0	6.7	51.8	7.1	36.8	6.4	129.4	7.8
70.8	7.5	94.7	5.9	326.5	5.9	70.8	7.4	62.8	4.9	165.4	6.2
		182.9	7.5					132	2.9		
		208.0	7.8								

In brief, the above discussion shows that the experimental error involved in the values that will be discussed below can be considered as a percentage of the reported value. This means that the larger the true run length value, the more variability is expected in the observed values (ARL or MRL). Furthermore, the magnitude of error does not depend on the nature of the chart.

4.2.1 Modified Univariate Charts

Section 2.3.1 proposed modifying the univariate $\overline{X}$, CUSUM and EWMA charts to handle the existence of within-subgroup correlation. The modification that was suggested involves the use of a corrected estimate of the variability of the subgroup average, $s_{\overline{X}}^2$. The objective of this section is to determine the trends in the performance of each chart at varying conditions in order to evaluate its potential for use in situations where within-subgroup correlation exists.

The $\overline{X}$ Chart

The simulation results of the modified $\overline{X}$ chart are shown in Tables 11, 12 and 13. Note that the Western Electric rules were applied in the simulations in order to make the control scheme as sensitive to shifts as the CUSUM and EWMA procedures. Table 11(a) shows that when the process is in control, the run length of the uncorrected $\overline{X}$ chart drops sharply with increasing correlation. For example, when 4 independent observations constitute a subgroup ($p = 4$, $\rho_0 = 0$), the ARL(0) is 84. The ARL(0) becomes 6.5 when ρ_0 is 0.95, indicating that, on average, a false alarm is expected after the 6'th or 7'th subgroup is plotted. When the modified chart is used, the ARL(0) for the same point ($p = 4$, $\rho_0 = 0.95$) increases to 99, a significant improvement over the

Table 11. Simulated results for the $\overline{X}$ chart with Western Electric rules and $k \to \infty$ (no training set)

(a)

$\Delta = 0$, no correction

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	84	58	83	57	84	58
0.30	21	15	15.0	10	12.4	9
0.50	11.9	8	8.9	6	7.3	5
0.95	6.5	5	5.0	4	4.2	3

(b)

$\Delta = 0.5\sigma$, no correction

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	10.4	8	8.7	8	7.5	7
0.30	7.9	6	6.3	5	5.3	4
0.50	6.4	5	5.3	4	4.5	3
0.95	4.7	4	3.9	3	3.3	3

(c)

$\Delta = 1.5\sigma$, no correction

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	1.8	1	1.5	1	1.3	1
0.30	1.8	1	1.6	1	1.5	1
0.50	1.8	2	1.7	1	1.5	1
0.95	1.9	1	1.7	1	1.6	1

(d)

$\Delta = 0$, correction with $\sigma_{\overline{X}}$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	100	70	100	73	100	71
0.50	101	72	100	71	100	70
0.95	99	71	101	72	99	71

(e)

$\Delta = 0.5\sigma$, correction with $\sigma_{\overline{X}}$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	18	13	17	13	16	12
0.50	22	16	21	16	21	15
0.95	30	22	30	22	30	22

(f)

$\Delta = 1.5\sigma$, correction with $\sigma_{\overline{X}}$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	3.0	3	2.9	3	2.7	3
0.50	3.7	3	3.7	3	2.6	3
0.95	5.3	5	5.3	5	5.4	5

Table 12. Simulated results for the $\overline{X}$ chart with Western Electric rules and k = 20

(a)
$\Delta = 0$, no correction

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	71	46	72	48	73	48
0.30	19	13	14.2	10	11.8	8
0.50	11.7	8	8.4	6	6.9	5
0.95	6.1	4	4.8	3	4.0	3

(b)
$\Delta = 0.5\sigma$, no correction

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	11.9	8	9.6	8	8.2	7
0.30	8.6	6	6.7	5	5.6	4
0.50	6.7	5	5.3	4	4.5	3
0.95	4.7	3	3.8	3	3.3	3

(c)
$\Delta = 1.5\sigma$, no correction

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	1.8	2	1.5	1	1.3	1
0.30	1.8	1	1.6	1	1.5	1
0.50	1.8	1	1.6	1	1.5	1
0.95	1.9	1	1.7	1	1.6	1

(d)
$\Delta = 0$, correction with C

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	66	40	64	41	64	39
0.50	67	42	66	41	66	40
0.95	62	37	64	39	64	39

(e)
$\Delta = 0.5\sigma$, correction with C

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	20	12	19	12	18	11
0.50	25	15	24	15	24	14
0.95	32	18	32	18	32	18

(f)
$\Delta = 1.5\sigma$, correction with C

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	3.1	3	2.9	3	2.8	3
0.50	3.8	3	3.7	3	3.6	3
0.95	5.3	5	5.3	5	5.3	4

(g)
$\Delta = 0$, correction with $s_{\overline{X}}$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	66	40	67	42	65	40
0.50	65	40	66	40	65	40
0.95	66	41	65	40	68	41

(h)
$\Delta = 0.5\sigma$, correction with $s_{\overline{X}}$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	20	12	19	12	18	11
0.50	25	15	24	15	24	14
0.95	34	19	33	19	32	19

(i)
$\Delta = 1.5\sigma$, correction with $s_{\overline{X}}$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	3.1	3	2.9	3	2.8	3
0.50	3.9	3	3.7	3	3.5	3
0.95	5.4	5	5.5	5	5.4	5

Table 13. Simulated results for the $\overline{X}$ chart with Western Electric rules and k = 50

(a)
$\Delta = 0$, no correction

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	83	56	82	57	83	58
0.30	20	14	14.9	11	12.2	9
0.50	11.9	8	8.8	6	7.8	5
0.95	6.2	4	4.8	3	4.1	3

(b)
$\Delta = 0.5\sigma$, no correction

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	11.0	8	9.2	8	7.7	7
0.30	8.2	6	6.5	5	5.5	4
0.50	6.6	5	5.2	4	4.4	3
0.95	4.6	3	3.8	3	3.3	3

(c)
$\Delta = 1.5\sigma$, no correction

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	1.8	1	1.5	1	1.3	1
0.30	1.8	1	1.6	1	1.5	1
0.50	1.9	1	1.7	1	1.5	1
0.95	1.9	1	1.7	1	1.6	1

(d)
$\Delta = 0$, correction with C

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	81	53	83	54	81	53
0.50	81	53	81	54	80	53
0.95	77	51	77	51	81	53

(e)
$\Delta = 0.5\sigma$, correction with C

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	21	12	19	12	18	11
0.50	25	14	24	14	22	14
0.95	32	18	32	18	31	18

(f)
$\Delta = 1.5\sigma$, correction C

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	3.1	3	2.9	3	2.8	3
0.50	3.9	3	3.7	3	3.7	3
0.95	5.3	5	5.3	4	5.3	4

(g)
$\Delta = 0$, correction with $s_{\overline{X}}$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	82	54	82	55	83	55
0.50	82	54	82	55	80	53
0.95	81	54	80	54	83	56

(h)
$\Delta = 0.5\sigma$, correction with $s_{\overline{X}}$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	19	13	18	13	17	12
0.50	24	16	23	16	23	15
0.95	33	20	32	21	32	21

(i)
$\Delta = 1.5\sigma$, correction with $s_{\overline{X}}$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	3.1	3	2.9	3	2.8	3
0.50	3.9	3	3.7	3	3.6	3
0.95	5.4	5	5.4	5	5.3	5

uncorrected chart even when $\rho_0 = 0$.

Having established that the use of the modified $\overline{X}$ chart reduces the number of false alarms, it is essential to consider its effectiveness when a shift in the level occurs. Tables 11(b) and (e) illustrate that as ρ_0 increases, the ARL(0.5σ) and MRL(0.5σ) of the traditional $\overline{X}$ chart decrease while those of the modified chart increase. However, when Δ, the shift size, increases to 1.5σ, the run lengths of both charts increase with ρ_0. Comparison of the values in Tables 11(b) and (c) versus (e) and (f) demonstrates that the run lengths of the modified chart are longer than those of the traditional chart. This means that it takes longer for the modified chart to detect a shift than the standard chart. The difference in run lengths is lessened as Δ increases.

In brief, Table 11 shows that, when no training set is used, the modified $\overline{X}$ chart is less effective in detecting shifts than the regular $\overline{X}$ chart in the presence of within-subgroup correlation. While false alarms are reduced, the chart becomes insensitive to small shifts in the level of the PQV. Consequently, the modified chart offers a better control scheme than the traditional chart only if it is desired to detect large shifts.

It is now desired to observe the effect of using a training set. Tables 12 and 13 display the simulated results when $k = 20$ and $k = 50$ respectively. The use of a training set produces a decrease in the in-control run length of the $\overline{X}$ chart while the run lengths are not affected significantly when a shift occurs. The same trends in the uncorrected $\overline{X}$ chart are observed as in the case when no training set is used where the run length decreases with increasing ρ_0. The modified chart corrects the occurrence of the false alarms; however, the achieved improvement is not as good as that when no training set is used ($k \rightarrow \infty$). While the values in Table 11(e) are not significantly different from those in Tables 12(e) or 13(e) where $\Delta > 0$, the in-control run lengths in Table 11(d) ($k \rightarrow \infty$) are definitely higher than those in Table 13 (c) ($k = 50$) which are in turn higher than those in Table 12 (c) ($k = 20$). Therefore, increasing the training set size improves the performance of the modified $\overline{X}$ chart in terms of reducing the number of false alarms while no change is observed in its power in detecting shifts.

With respect to the method of correction, no significant difference can be detected. Slightly longer run lengths are observed when $s_{\overline{X}}$ is used as can be seen from pairwise comparison of Tables 13 (d), (e) and (f) with (g), (h) and (i) respectively. Correcting with $s_{\overline{X}}$ also produces run length values closer to those corrected with $\sigma_{\overline{X}}$. This effect is not noticeable for the smaller training set size in Table 12.

The subgroup size, p, does not have any notable effects on the modified $\overline{X}$ chart disregarding the shift or training set sizes. However, as p increases, the run lengths of the regular $\overline{X}$ chart decrease especially when their values are relatively small.

The CUSUM Chart

The simulated behaviour of the 2-sided CUSUM procedure for $h = 4$ is shown in Tables 14, 15 and 16 corresponding to values of k set to ∞, 20 and 50 respectively.

Tables 14(a), (b) and (c) display the run length values of the regular CUSUM chart in the presence of shifts of various sizes. As ρ_0 increases, the frequency of false alarms increases significantly. For $p = 4$, a false alarm is expected every 166 subgroups when $\rho_0 = 0$ and every 9 subgroups when $\rho_0 = 0.95$ (Table 14(a)). The speed of detecting a 0.5σ shift is slightly improved for correlated subgroups. However, it does

Table 14. Simulated results of the 2-sided CUSUM (k_c=0.5, h=4) for k → ∞ (no training set).

(a)

$\Delta = 0$, no correction

	p = 4		p = 5		p = 6	
ρ_0	ARL	MRL	ARL	MRL	ARL	MRL
0.0	166	115	167	119	167	117
0.30	28	21	22	16	17 3	13
0.50	17.9	13	13.1	10	10.7	8
0.95	9.3	7	7.4	6	6.1	5

(b)

$\Delta = 0.5\sigma$, no correction

	p = 4		p = 5		p = 6	
ρ_0	ARL	MRL	ARL	MRL	ARL	MRL
0.0	8.4	7	7.0	6	6.2	6
0.30	7.5	6	6.4	5	5.8	5
0.50	7.1	6	6.1	5	5.3	4
0.95	6.0	5	5.1	4	4.4	4

(c)

$\Delta = 1.5\sigma$, no correction

	p = 4		p = 5		p = 6	
ρ_0	ARL	MRL	ARL	MRL	ARL	MRL
0.0	2.2	2	1.9	2	1.8	2
0.30	2.2	2	2.0	2	1.8	2
0.50	2.3	2	2.1	2	1.9	2
0.95	2.3	2	2.1	2	1.9	2

(d)

$\Delta = 0$, correction with $\sigma_{\bar{X}}$

	p = 4		p = 5		p = 6	
ρ_0	ARL	MRL	ARL	MRL	ARL	MRL
0.30	171	120	169	118	168	119
0.50	172	122	167	118	168	117
0.95	167	117	173	125	173	122

(e)

$\Delta = 0.5\sigma$, correction with $\sigma_{\bar{X}}$

	p = 4		p = 5		p = 6	
ρ_0	ARL	MRL	ARL	MRL	ARL	MRL
0.30	14.0	11	13.2	11	12.7	10
0.50	18	14	17.4	14	16.6	13
0.95	26	20	25	20	27	19

(f)

$\Delta = 1.5\sigma$, correction with $\sigma_{\bar{X}}$

	p = 4		p = 5		p = 6	
ρ_0	ARL	MRL	ARL	MRL	ARL	MRL
0.30	3.0	3	2.9	3	2.8	3
0.50	3.5	3	3.4	3	3.4	3
0.95	4.6	4	4.7	4	4.6	4

Table 15. Simulated results of the 2-sided CUSUM (k_c=0.5, h=4) for k = 20.

(a)

$\Delta = 0$, no correction

	p = 4		p = 5		p = 6	
ρ_0	ARL	MRL	ARL	MRL	ARL	MRL
0.0	129	66	117	63	121	69
0.30	24	16	18.1	12	14.9	10
0.50	15.3	10	11.6	8	9.8	7
0.95	8.7	6	6.7	5	5.6	4

(b)

$\Delta = 0.5\sigma$, no correction

	p = 4		p = 5		p = 6	
ρ_0	ARL	MRL	ARL	MRL	ARL	MRL
0.0	9.6	7	8.0	6	6.8	6
0.30	8.5	6	7.1	5	6.2	5
0.50	7.6	6	6.4	5	5.6	4
0.95	6.1	4	4.9	4	4.4	3

(c)

$\Delta = 1.5\sigma$, no correction

	p = 4		p = 5		p = 6	
ρ_0	ARL	MRL	ARL	MRL	ARL	MRL
0.0	2.2	2	2.0	2	1.8	2
0.30	2.4	2	2.0	2	1.8	2
0.50	2.3	2	2.0	2	1.9	2
0.95	2.3	2	2.1	2	1.9	2

(d)

$\Delta = 0$, correction with C

	p = 4		p = 5		p = 6	
ρ_0	ARL	MRL	ARL	MRL	ARL	MRL
0.30	210	54	214	57	216	57
0.50	215	63	214	58	225	58
0.95	214	53	206	55	196	52

(e)

$\Delta = 0.5\sigma$, correction with C

	p = 4		p = 5		p = 6	
ρ_0	ARL	MRL	ARL	MRL	ARL	MRL
0.30	21	11	21	10	19	10
0.50	31	13	29	13	29	13
0.95	49	17	55	17	44	17

(f)

$\Delta = 1.5\sigma$, correction with C

	p = 4		p = 5		p = 6	
ρ_0	ARL	MRL	ARL	MRL	ARL	MRL
0.30	3.0	3	2.9	3	2.8	3
0.50	3.5	3	3.5	3	3.4	3
0.95	4.6	4	4.6	4	4.6	4

(g)

$\Delta = 0$, correction with $s_{\bar{x}}$

	p = 4		p = 5		p = 6	
ρ_0	ARL	MRL	ARL	MRL	ARL	MRL
0.30	226	60	236	58	242	61
0.50	222	58	224	60	225	57
0.95	231	60	237	63	215	58

(h)

$\Delta = 0.5\sigma$, correction with $s_{\bar{x}}$

	p = 4		p = 5		p = 6	
ρ_0	ARL	MRL	ARL	MRL	ARL	MRL
0.30	25	11	19	10	19	10
0.50	32	13	31	13	33	13
0.95	61	18	55	18	53	18

(i)

$\Delta = 1.5\sigma$, correction with $s_{\bar{x}}$

	p = 4		p = 5		p = 6	
ρ_0	ARL	MRL	ARL	MRL	ARL	MRL
0.30	3.1	3	2.9	3	2.8	3
0.50	3.7	3	3.5	3	3.5	3
0.95	4.8	4	4.8	4	4.8	4

Table 16. Simulated results of the 2-sided CUSUM (k_r=0.5, h=4) for k = 50.

(a)
$\Delta = 0$, no correction

	p = 4		p = 5		p = 6	
ρ_0	ARL	MRL	ARL	MRL	ARL	MRL
0.0	139	87	140	86	137	86
0.30	26	18	19.9	14	16.3	12
0.50	16.2	12	12.4	9	10.0	7
0.95	9.2	7	7.2	6	5.8	5

(b)
$\Delta = 0.5\sigma$, no correction

	p = 4		p = 5		p = 6	
ρ_0	ARL	MRL	ARL	MRL	ARL	MRL
0.0	8.7	7	7.4	6	6.4	6
0.30	7.9	6	6.8	5	6.0	5
0.50	7.1	6	6.2	5	5.5	4
0.95	6.1	5	5.1	4	4.4	4

(c)
$\Delta = 1.5\sigma$, no correction

	p = 4		p = 5		p = 6	
ρ_0	ARL	MRL	ARL	MRL	ARL	MRL
0.0	2.2	2	1.9	2	1.8	2
0.30	2.2	2	2.0	2	1.8	2
0.50	2.3	2	2.0	2	1.9	2
0.95	2.3	2	2.1	2	1.9	2

(d)
$\Delta = 0$, correction with C

	p = 4		p = 5		p = 6	
ρ_0	ARL	MRL	ARL	MRL	ARL	MRL
0.30	176	88	178	84	184	87
0.50	170	80	165	83	176	83
0.95	168	80	170	81	164	82

(e)
$\Delta = 0.5\sigma$, correction with C

	p = 4		p = 5		p = 6	
ρ_0	ARL	MRL	ARL	MRL	ARL	MRL
0.30	16	11	15	11	14	10
0.50	22	14	20	13	19	13
0.95	33	19	32	19	32	19

(f)
$\Delta = 1.5\sigma$, correction with C

	p = 4		p = 5		p = 6	
ρ_0	ARL	MRL	ARL	MRL	ARL	MRL
0.30	3.0	3	2.9	3	2.9	3
0.50	3.6	3	3.5	3	3.4	3
0.95	4.7	4	4.6	4	4.6	4

(g)
$\Delta = 0$, correction with $s_{\bar{x}}$

	p = 4		p = 5		p = 6	
ρ_0	ARL	MRL	ARL	MRL	ARL	MRL
0.30	179	83	177	83	180	86
0.50	177	85	174	85	177	83
0.95	170	80	179	89	178	84

(h)
$\Delta = 0.5\sigma$, correction with $s_{\bar{x}}$

	p = 4		p = 5		p = 6	
ρ_0	ARL	MRL	ARL	MRL	ARL	MRL
0.30	16	11	15	11	14	10
0.50	21	14	20	14	20	13
0.95	32	18	32	19	33	19

(i)
$\Delta = 1.5\sigma$, correction with $s_{\bar{x}}$

	p = 4		p = 5		p = 6	
ρ_0	ARL	MRL	ARL	MRL	ARL	MRL
0.30	3.1	3	2.9	3	2.8	3
0.50	3.6	3	3.5	3	3.5	3
0.95	4.7	4	4.6	4	4.7	4

not make sense to use a chart that gives a false alarm every 9 points and requires 6 points to detect a small shift because it would not be possible to distinguish between false and true alarms.

The use of a modified CUSUM for correlated subgroups increases the ARL(0) with the values of the regular scheme when $\rho_0 = 0$. As in the modified $\overline{X}$ chart, the corrected CUSUM chart takes longer to detect small shifts than the regular chart. Again, this effect is diminished as the size of the shift becomes larger, as the comparison between Tables 14(c) and (f) demonstrates.

Tables 15 and 16 illustrate the effect of using a training set on the CUSUM chart. Comparison with Table 14 shows that the in-control run length of the uncorrected CUSUM chart decreases with decreasing k. The run lengths of this chart are not affected by k when a 0.5σ or 1.5σ occurs. When a modified CUSUM procedure is used, the ARL(0) increases with decreasing k while the MRL(0) decreases. For example, when $k = 20$, $p = 4$ and $\rho_0 = 0.95$, the ARL(0) and MRL(0) are 214 and 53 respectively versus 167 and 117 when $k \rightarrow \infty$ (Tables 14(d) and 15(d)). The increased difference between the averages and medians of the run lengths indicates that the run length distribution becomes more skewed as k decreases. The same trend is observed for $\Delta = 0.5\sigma$ but the effects are not as pronounced as the case where $\Delta = 0$. As Δ increases, this effect gradually decreases.

As far as the method of correction is concerned, the use of $s_{\overline{X}}$ produces slightly longer run lengths than the use of C. This effect can be observed in Table 15 where $k = 20$. As k increases, the difference between the two methods is minimized as expected since both estimates of $\sigma_{\bar{X}}$ approach the true $\sigma_{\bar{X}}$ value.

The ARL and MRL of the regular CUSUM decreases with increasing p when $\rho_0 > 0$ and $\Delta = 0$. On the other hand, p does not have any effect on the performance of the modified CUSUM procedure. With respect to the degree of correlation, as ρ_0 increases, the run length of the regular CUSUM decreases. When the modified chart is used instead, the same effect is observed only when $\Delta \neq 0$. This means that the performance of the regular chart deteriorates with increasing correlation and the speed of the modified chart in detecting a shift decreases.

The simulations of the 2-sided CUSUM procedure were repeated for $h = 5$ and produces longer run lengths. The results are displayed in Appendix C. The same trends are observed as in the case where $h = 4$. This confirms that the above discussion and conclusions are not scale dependent.

The EWMA Chart

The simulated results of using the EWMA chart are shown in Tables 17, 18 and 19. The parameters λ and L were set to 0.06 and 1.954 respectively; these values were reported as being optimal by Lucas and Saccucci (1990). The EWMA chart was also simulated for λ and L values of 0.05 and 2.399 respectively, which are supposed to produce longer run lengths than the first set. These results are shown in Appendix C.

Examination of the results shows similar trends to the CUSUM chart discussed above. The regular EWMA generates a large number of false alarms in the presence of within subgroup correlation. When no training set is used, the modified chart has essentially the same in-control run lengths as the regular chart when the subgroups are

Table 17. Simulated results of the EWMA chart ($\lambda = 0.06$, $L = 1.954$) for $k \rightarrow \infty$

(a)
$\Delta = 0$, no correction

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0	79	48	79	48	80	50
0.3	20	9	16	7	12	5
0.5	12	5	9	4	7	3
0.95	7	3	5	2	4	2

(b)
$\Delta = 0.5\sigma$, no correction

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	4.4	3	3.8	3	3.3	3
0.30	4.3	3	3.7	2	3.4	2
0.50	4.2	3	3.5	2	3.1	2
0.95	3.6	2	2.9	2	2.6	2

(c)
$\Delta = 1.5\sigma$, no correction

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0	1.1	1	1.1	1	1.0	1
0.30	1.2	1	1.2	1	1.2	1
0.50	1.3	1	1.3	1	1.2	1
0.95	1.4	1	1.4	1	1.3	1

(d)
$\Delta = 0$, correction with $\sigma_{\bar{x}}$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.3	80	51	79	49	82	51
0.5	78	47	80	50	82	52
0.95	79	48	81	50	80	48

(e)
$\Delta = 0.5\sigma$, correction with $\sigma_{\bar{x}}$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	7.1	6	6.6	5	6.4	5
0.50	8.8	7	8.4	6	8.3	6
0.95	12.1	9	12.0	9	11.9	9

(f)
$\Delta = 1.5\sigma$, correction with $\sigma_{\bar{x}}$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	1.5	1	1.5	1	1.4	1
0.50	1.8	2	1.8	2	1.8	2
0.95	2.4	2	2.4	2	2.4	2

Table 18. Simulated results of the EWMA chart ($\lambda = 0.06$, L = 1.954) for k = 20

(a)

$\Delta = 0$, no correction

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	83	46	82	49	79	46
0.30	20	8	16	6	12	5
0.50	13	5	9.2	4	7.2	3
0.95	6.4	3	4.7	2	3.7	2

(b)

$\Delta = 0.5\sigma$, no correction

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	4.4	3	3.8	3	3.3	3
0.30	4.2	3	3.6	2	3.2	2
0.50	4.0	3	3.4	2	3.1	2
0.95	3.5	2	2.9	2	2.5	2

(c)

$\Delta = 1.5\sigma$, no correction

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	1.1	1	1.1	1	1.0	1
0.30	1.2	1	1.2	1	1.1	1
0.50	1.3	1	1.2	1	1.2	1
0.95	1.4	1	1.3	1	1.3	1

(d)

$\Delta = 0$, correction with C

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	94	38	97	41	99	44
0.50	98	43	97	42	99	44
0.95	90	41	91	40	88	39

(e)

$\Delta = 0.5\sigma$, correction with C

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	7.1	5	6.8	5	6.5	5
0.50	8.7	6	8.4	6	8.2	6
0.95	11.8	8	11.6	8	11.9	8

(f)

$\Delta = 1.5\sigma$, correction with C

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	1.5	1	1.5	1	1.5	1
0.50	1.8	2	1.8	1	1.7	1
0.95	2.4	2	2.4	2	2.4	2

(g)

$\Delta = 0$, correction with $s_{\bar{x}}$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	102	45	103	47	100	44
0.50	99	44	104	44	98	40
0.95	101	44	103	44	103	47

(h)

$\Delta = 0.5\sigma$, correction with $s_{\bar{x}}$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	7.2	5	6.9	5	6.5	5
0.50	8.7	6	8.6	6	8.4	6
0.95	12.0	9	12.2	9	11.9	8

(i)

$\Delta = 1.5\sigma$, correction with $s_{\bar{x}}$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	1.6	1	1.5	1	1.4	1
0.50	1.8	2	1.8	2	1.8	1
0.95	2.4	2	2.4	2	2.4	2

Table 19. Simulated results of the EWMA chart ($\lambda = 0.06$, $L = 1.954$) for $k = 50$

(a)

$\Delta = 0$, no correction

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	82	49	83	51	82	52
0.30	20	8	15.2	6	12.5	5
0.50	13	5	9.2	4	7.4	3
0.95	6.4	3	4.8	2	3.9	2

(b)

$\Delta = 0.5\sigma$, no correction

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	4.5	4	3.8	3	3.3	3
0.30	4.2	3	3.7	3	3.3	2
0.50	4.1	3	3.6	2	3.1	2
0.95	3.6	2	3.0	2	2.6	2

(c)

$\Delta = 1.5\sigma$, no correction

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	1.0	1	1.1	1	1.0	1
0.30	1.3	1	1.2	1	1.2	1
0.50	1.3	1	1.2	1	1.2	1
0.95	1.4	1	1.3	1	1.3	1

(d)

$\Delta = 0$, correction with C

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	87	48	87	46	84	45
0.50	84	45	85	46	85	46
0.95	83	47	81	45	82	45

(e)

$\Delta = 0.5\sigma$, correction with C

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	7.2	6	6.7	5	6.4	5
0.50	8.6	7	8.4	6	8.3	6
0.95	11.9	9	11.6	9	11.9	9

(f)

$\Delta = 1.5\sigma$, correction with C

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	1.5	1	1.5	1	1.4	1
0.50	1.8	2	1.8	2	1.8	1
0.95	2.4	2	2.4	2	2.3	2

(g)

$\Delta = 0$, correction with $s_{\bar{x}}$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	86	46	86	45	85	46
0.50	82	44	90	49	86	47
0.95	88	45	88	48	87	49

(h)

$\Delta = 0.5\sigma$, correction with $s_{\bar{x}}$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	7.2	5	6.9	5	6.5	5
0.50	8.8	7	8.5	7	8.0	6
0.95	11.8	9	12.7	9	11.9	9

(i)

$\Delta = 1.5\sigma$, correction with $s_{\bar{x}}$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	1.5	1	1.5	1	1.4	1
0.50	1.8	2	1.8	2	1.7	1
0.95	2.4	2	2.4	2	2.4	2

independent. However, the run lengths when $\Delta \neq 0$ become longer when the correction is applied and this effect diminishes as Δ increases to become negligible at $\Delta = 1.5\sigma$.

The use of a training set does not have as large an effect on the behaviour of the EWMA procedure as it does for the CUSUM. Nevertheless, the same trends are observed. The regular EWMA chart is hardly affected by the training set size as shown by the comparison of Tables 17(a),(b) and (c) with Tables 18(a), (b) and (c) respectively. The modified chart has longer ARL(0)'s and shorter MRL(0)'s than the chart without a training set. Furthermore, the use of an estimate, $s_{\bar{X}}$, of $\sigma_{\bar{X}}$ from the training set in the correction yields longer run lengths which are closer to the true values that are displayed in Tables 17(c), (d) and (e). Therefore, the application of $s_{\bar{X}}$ in the correction of the EWMA chart is preferred over the factor $C = [1+(p-1)\bar{r}]^{0.5}$.

The EWMA chart is affected by p and ρ_0 in the same manner as the $\bar{X}$ and CUSUM charts. Thus, these trends will not be discussed again.

4.2.2 Principal Components Approach

The principal components were determined from sets of simulated data. In all situations, one principal component accounted for almost all of the variability. This principal component was determined to closely estimate the subgroup average, $\bar{X}_i$, as demonstrated in the sample calculations in Appendix D. Therefore, one of the univariate charts discussed above should be used to monitor the process. The results of this approach coincide with those in the previous section.

4.2.3 Multivariate Control Charts

This section addresses the results of the simulations performed on multivariate charts. The effects of p, ρ_0 and k on the Hotelling's T^2, MEWMA, MC1 and MCUSUM charts are discussed. Recall from Chapter 3 that the results obtained for these charts are limited to the situation with an equal correlation structure among the variables, a case which is prevalent in the semiconductor industry. Correlation structures cannot occur randomly since they have several constraints and difficulty was encountered in simulating correlated data with unequal pairwise correlations.

Hotelling's T^2 Chart

Section 2.1.2 describes a Hotelling's T^2 procedure proposed by Jackson (1959). The results of simulating this procedure are displayed in Tables 20, 21 and 22. Table 20(a) shows that this procedure has a relatively low in-control run length when no training set is used. The ARL(0) is approximately 21 while the MRL(0) is 15. When a 0.5σ shift occurs, the run length changes insignificantly. For example, when $p = 4$ and $\rho_0 = 0.95$, the ARL(0.5) drops to 16.8 and the MRL(0.5) to 13. This means that a false alarm is expected every 21 points and a 0.5σ shift is detected after 17 points. The use of a training set produces longer run lengths when $\Delta = 0$ and 0.5σ as shown in Tables 21 and 22 corresponding to $k = 20$ and $k = 50$ respectively. Again, it is not possible to distinguish between a false and a true alarm.

For this procedure, there are no parameters that influence the magnitude of the run lengths such as λ and L for EWMA or h and k for CUSUM. Therefore, the behaviour discussed above is optimal as well as the only one available. Based on the above observations, it was concluded that this procedure is not an effective SPC tool that can be used on correlated samples.

Table 20. Simulated results of Hotelling's T^2 for $k \to \infty$.

(a) $\Delta = 0$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	21	15	21	15	21	15
0.30	21	15	21	15	22	15
0.50	21	15	21	15	21	15
0.95	21	15	21	14	21	15

(b) $\Delta = 0.5\sigma$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	9.7	7	9.0	6	8.5	6
0.30	13.2	9	13.2	9	13.9	10
0.50	14.4	10	15.2	11	15.2	11
0.95	16.8	12	17.4	12	17.4	12

Table 21. Simulated results of Hotelling's T^2 chart for $k = 20$.

(a) $\Delta = 0$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	50	18	53	19	55	20
0.30	47	18	49	19	52	19
0.50	47	19	53	19	56	19
0.95	50	18	49	19	58	19

(b) $\Delta = 0.5\sigma$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	22	10	22	10	24	10
0.30	31	13	33	14	39	14
0.50	33	11	37	15	43	15
0.95	38	15	42	16	45	17

Table 22. Simulated results of Hotelling's T^2 chart for $k = 50$

(a) $\Delta = 0$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	27	16	26	16	26	16
0.30	26	16	26	16	26	16
0.50	26	16	26	16	27	16
0.95	27	16	27	16	27	16

(b) $\Delta = 0.5\sigma$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	12.5	8	11.5	7	11.4	7
0.30	17	10	17.3	11	17.9	11
0.50	19	11	19	12	20	12
0.95	21	12.5	21	13	22	14

The MEWMA Procedure

The MEWMA chart parameters r and h_4 are set at the optimal settings of 0.06 and 13.5 respectively (Lowry et al., 1992). The results of the simulations are shown in Tables 23 through 25. The situation where no training set is used and $\Delta = 0$ is displayed in Table 23(a). This table shows that the ρ_0 does not have a significant effect of the in-control run length of the chart since the observed variability can be explained by the experimental error of the simulations as discussed previously. However, ρ_0 exhibits an important influence on the run length of the chart when a shift occurs. Tables 23(b) and (c) demonstrate that as ρ_0 increases, the chart requires more time to detect the shift. For instance, for $p = 5$ and $\rho_0 = 0$, the ARL(0.5) is 7.4 while this value rises to 27 when the correlation is 0.95. This demonstrates that while the number of false alarms for the MEWMA is not affected by the degree of correlation among the variables, it takes longer for the chart to detect a shift when the degree of correlation is increased.

One major misgiving of this chart is that its performance is dependent on the training set that is used. This effect can be seen clearly from Tables 23 and 24. If a training set of 20 subgroups is used, the ARL(0) drops from 365 to 135 when $p = 4$ and $\rho_0 = 0$. When $k = 50$, the same value becomes 238, which is still significantly different from the 365 value. Thus, a large training set is required to bring the run lengths of this chart to their asymptotic values. This trend is not as drastic on the run length of the MEWMA when a shift occurs.

The number of variables, p, in each **X** vector or alternatively the subgroup size,

Table 23. Simulated results of MEWMA ($r=0.06$, $h_4=13.5$) for $k \rightarrow \infty$.

(a) $\Delta = 0$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	365	253	185	125.5	102	64
0.30	363	251	179	119	101	65
0.50	357	243	185	126	103	70
0.95	355	232	183	123	102	68

(b) $\Delta = 0.5\sigma$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	10.1	9	7.4	7	5.7	5
0.30	17.6	15	14.4	12	11.5	9
0.50	22	19	18.4	15	14.8	12
0.95	33	27.5	27	23	22	17

(c) $\Delta = 1.0\sigma$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	3.2	3	2.5	2	1.2	1
0.30	5.4	5	4.6	2	2.2	2
0.50	6.8	6	5.9	5	2.8	2
0.95	10.0	9	8.7	8	4.1	4

Table 24. Simulated results of MEWMA ($r=0.06$, $h_4=13.5$) for $k=20$.

(a) $\Delta = 0$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	135	76	54	27	24	10
0.30	144	81	56	28	25	10
0.50	161	80	62	27	28	10
0.95	283	81	121	27	49	10

(b) $\Delta = 0.5\sigma$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	8.4	7	5.6	4	3.9	3
0.30	14.2	12	9.7	7	6.6	4
0.50	17.2	14	12.2	9	8.1	9
0.95	24	19	17.8	12	12.1	7

(c) $\Delta = 1.0\sigma$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	2.9	3	2.1	2	1.2	1
0.30	4.7	4	3.6	4	1.8	1
0.50	5.8	5	4.6	4	2.2	2
0.95	8.6	7	6.6	5	3.1	3

Table 25. Simulated results of MEWMA ($r=0.06$, $h_4=13.5$) for $k = 50$.

(a) $\Delta = 0$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	238	159	110	69	55	30
0.30	243	154	112	67	56	30
0.50	253	150	116	67	61	33
0.95	333	154	153	69	78	32

(b) $\Delta = 0.5\sigma$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	9.4	8	6.8	6	4.9	4
0.30	16.2	14	12.2	10	9.4	7
0.50	20	17	16.1	13	11.9	9
0.95	30	25	23	18	17.4	12

(c) $\Delta = 1.0\sigma$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	3.1	3	2.4	2	1.2	1
0.30	5.1	5	4.2	4	2.0	2
0.50	6.6	6	5.4	5	2.6	2
0.95	9.3	8	7.9	7	3.7	3

has a serious effect on the run lengths of the MEWMA. Table 23(a) shows that as p increases from 4 to 6 with $\rho_0 = 0$, the ARL(0) falls from 365 to 102. The same trend is observed in all other cases as Δ, k and ρ_0 vary.

The MCUSUM Procedures

The performance of two multivariate CUSUM procedures was simulated. Tables 26 through 28 show the results of Crosier's MC1 procedure while those of Pignatiello and Runger's MCUSUM are displayed in Tables 29, 30 and 31. Optimal parameter settings of each chart are chosen from the values provided in the literature (Crosier (1988); Pignatiello and Runger (1990)). Note that these are not the only optimal parameter values but the chosen values will be sufficient to study the behavior of each chart.

The trends for both MC1 and MCUSUM are very similar to those of the MEWMA procedure. When no training set is used, the in-control run length of the charts is not dependent on ρ_0 while the run length when a shift occurs increases with the degree of correlation. When a training set of size 20 or 50 is used, the number of false alarms increases notably whereas the run length is not affected very much relatively. As p increases, the run lengths of both charts is diminished.

In brief, the above discussion shows that the proposed Hotelling's T^2 is not an efficient alternative to monitor correlated samples. The three other charts, MEWMA, MC1 and MCUSUM, display the same trends. While the presence of within-sample correlation does not affect the number of false alarms, the speed of detecting a shift decreases with increasing ρ_0. Furthermore, a very large training set is required in order

Table 26. Simulated results of MC1 (k_1=0.5, h_2 = 9.5) for k → ∞.

(a) $\Delta = 0$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	475	336	217	156	116	86
0.30	492	350	213	151	114	82
0.50	481	333	214	150	112	82
0.95	493	348	215	151	115	84

(b) $\Delta = 0.5\sigma$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	14.8	14	11.9	11	10.1	10
0.30	24	21	19.9	18	17.3	16
0.50	30	26	25	22	21	19
0.95	43	35	35	30	29	25

(c) $\Delta = 1.0\sigma$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	6.4	6	5.1	5	4.8	5
0.30	9.2	9	8.5	8	7.8	8
0.50	10.9	10	10.1	10	9.4	9
0.95	14.5	14	13.3	12	12.5	12

Table 27. Simulated results of MC1 (k_1=0.5, h_2 = 9.5) for k = 20.

(a) $\Delta = 0$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	185	114	75	50	40	29
0.30	200	117	77	52	41	29
0.50	224	114	86	51	43	29
0.95	466	121	173	51	68	28

(b) $\Delta = 0.5\sigma$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	13.2	12	10.3	10	8.3	8
0.30	21	18	16.1	14	13.1	12
0.50	25	21	19.4	17	15.6	14
0.95	35	28	27	22	21	17

(c) $\Delta = 1.0\sigma$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	5.9	6	4.9	5	4.2	4
0.30	8.5	8	7.5	7	6.7	6
0.50	10.1	10	9.0	9	8.1	8
0.95	13.4	13	12.1	11	10.8	10

Table 28. Simulated results of MC1 (k_1=0.5, h_2 = 9.5) for k = 50.

(a) $\Delta = 0$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	315	212	133	95	72	53
0.30	330	212	142	98	75	54
0.50	353	217	148	98	72	53
0.95	514	220	206	97	95	54

(b) $\Delta = 0.5\sigma$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	14	13	11.3	11	9.5	9
0.30	22	20	18.4	17	15.9	14
0.50	28	24	23	20	19.2	17
0.95	40	30	32	27	26	22

(c) $\Delta = 1.0\sigma$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	6.2	6	5.3	5	4.6	5
0.30	9.0	9	8.1	8	7.5	7
0.50	10.7	10	9.7	9	8.9	9
0.95	14.1	13	13.0	12	11.9	11

Table 29. Simulated results of MCUSUM (k_2=0.5, h_3 = 6.82) for k -▸ ∞.

(a) $\Delta = 0$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	336	242	204	139	132	93
0.30	337	233	205	146	135	94
0.50	336	233	206	144	133	94
0.95	338	230	200	143	131	93

(b) $\Delta = 0.5\sigma$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	11.9	11	9.4	9	7.7	7
0.30	21	17	17.8	14	15.2	12
0.50	28	22	24	18	21	16
0.95	42	32	37	27	32	23

(c) $\Delta = 1.0\sigma$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	4.7	5	4.1	4	3.6	4
0.30	6.9	7	6.3	6	5.9	6
0.50	8.4	8	7.7	7	7.2	7
0.95	11.5	10	10.8	10	10.0	9

Table 30. Simulated results of MCUSUM (k_2=0.5, h_3 = 6.82) for k = 20.

(a) $\Delta = 0$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	128	78	64	41	37	23
0.30	136	80	68	42	39	24
0.50	141	78	72	43	41	24
0.95	246	82	112	42	63	24

(b) $\Delta = 0.5\sigma$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	10.5	9	7.9	7	6.2	6
0.30	17.8	14	13.8	11	10.9	9
0.50	22	17	17.8	14	13.7	11
0.95	35	25	28	19	20	13

(c) $\Delta = 1.0\sigma$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	4.4	4	3.7	4	3.1	3
0.30	6.4	6	5.6	5	4.9	5
0.50	7.7	7	6.8	6	5.9	5
0.95	10.6	9	9.5	8	8.2	7

Table 31. Simulated results of MCUSUM (k_2=0.5, h_3 = 6.82) for k = 50.

(a) $\Delta = 0$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	225	149	124	82	77	53
0.30	218	142	129	85	78	51
0.50	233	146	131	82	81	53
0.95	297	154	168	84	104	54

(b) $\Delta = 0.5\sigma$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	11.5	8	8.9	8	7.2	6
0.30	19	16	16.5	13	13.8	11
0.50	26	21	22	17	17.9	14
0.95	39	30	33	24	27	19

(c) $\Delta = 1.0\sigma$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	4.6	4	3.9	4	3.4	3
0.30	6.8	6	6.1	6	5.6	5
0.50	8.3	8	7.4	7	6.7	o
0.95	11.2	10	10.3	9	9.4	8

for these charts to achieve their best performance.

After the general trends of each of the proposed charts have been established, the next step was to compare the performance of each of: modified $\overline{X}$, CUSUM, EWMA, MEWMA, MC1 and MCUSUM. The results of the comparison are shown and addressed below in section 4.3.

4.3 Comparison of the Proposed Charts

The objective of this section is to find the chart with the optimal behavior in the range of values that are studied in this thesis. The best chart will have the longest RL(0) and shortest RL(Δ) where $\Delta \neq 0$. Since the behaviour of all of the charts except the $\overline{X}$ chart can be optimized at several parameter settings, a common basis for the comparison had to be established. In the SPC literature, the basis is often chosen to be the ARL(0) and the best chart will detect a shift faster than others, that is, it has the lowest ARL(Δ) value when $\Delta \neq 0$. This method was used here as well as a similar one based on MRLs. In both cases, p is chosen at its middle value, 5, while ρ_0 is set to two extreme values, 0 and 0.95.

Since the ARL(0) of the $\overline{X}$ chart cannot be adjusted by parameters such as λ and L for the EWMA chart, the ARL(0)s of the other charts were matched with the ARL(0) for the $\overline{X}$ chart with $p = 5$ and $\rho_0 = 0$. The values of the respective parameters are as follows: for CUSUM, $k_c = 0.5$ and $h = 3.32$, for EWMA, $\lambda=0.06$ and $L=1.975$, for MEWMA, $r=0.06$ and $h_4=11.4$, for MC1, $k_1=0.5$ and $h_2=7.8$ and for MCUSUM, $k_2=0.5$ and $h_3=5.9$. The results for the univariate charts are shown in Table 32 while those of the multivariate charts in Table 33.

Table 32. Univariate charts comparison based on average run lengths for p = 5.

			CUSUM (k_c = 0.5, h = 3.32)			EWMA (λ = 0.06, L = 1.954)			X		
		method of	shift=0	shift=0.5σ	shift=1.5σ	shift=0	shift=0.5σ	shift=1.5σ	shift=0	shift=0.5σ	shift=1.5σ
ρ_0	training set	correction	ARL MRL	ARL MRL	ARL MRL	ARL MRL	ARL MRL	ARL MRL	ARL MRL	ARL MRL	ARL MRL
0	none	none	83 58	6.1 5	1.7 2	84 52	3.8 3	1.1 1	83 57	8.7 8	1.5 1
0	20	none	64 38	6.6 5	1.7 2	89 53	3.8 3	1.1 1	72 48	9.6 8	1.5 1
0	50	none	73 46	6.2 5	1.7 2	84 51	3.8 3	1.4 1	82 57	9.2 8	1.5 1
0.95	none	none	5.4 4	4.2 3	1.8 2	4.9 2	3.1 2	1.4 1	5.0 4	3.9 3	1.7 1
0.95	none	ρ_0	83 59	19.5 15	3.9 4	88 55	12.2 9	2.5 2	101 72	30 22	5.3 5
0.95	20	none	5.2 4	4.0 3	1.8 2	4.8 2	3.1 2	1.3 1	4.8 3	3.8 3	1.7 1
0.95	20	C	84 32	28 13	4.0 3	111 47	12.3 9	2.5 2	64 39	32 18	5.3 5
0.95	20	$s_{\bar{x}}$	103 35.5	34 14	4.1 4	96 41	11.8 8	2.4 2	65 40	33 19	5.5 5
0.95	50	none	5.4 4	4.1 3	1.9 2	4.8 2	3.0 2	1.4 1	4.7 3	3.8 3	1.7 1
0.95	50	C	79 43	22 14	3.9 4	90 49	12.6 9	2.4 2	77 51	32 18	5.3 4
0.95	50	$s_{\bar{x}}$	82 46	23 14	4.0 4	85 47	12.4 9	2.4 2	80 54	32 21	5.4 5

Table 33. Multivariate charts comparison based on average run lengths for $p = 5$.

		MEWMA ($r = 0.06$, $h_4 = 11.4$)						MCI ($k_1 = 0.5$, $h_2 = 7.85$)						MCUSUM ($k_2 = 0.5$, $h_3 = 5.67$)					
		shift=0		shift=0.5σ		shift=1.5σ		shift=0		shift=0.5σ		shift=1.5σ		shift=0		shift=0.5σ		shift=1.5σ	
ρ_0	training set	ARL	MRL	ARL	MRL	ARL	MRL	ARL	MRL	ARL	MRL	ARL	MRL	ARL	MRL	ARL	MRL	ARL	MRL
0	none	83	49	5.8	5	1.3	1	83	61	9.6	9	3.1	3	83	58	7.6	7	2.3	2
0	20	29	13	4.4	3	1.2	1	38	28	8.2	8	2.8	3	35	23	6.3	5	2.2	2
0	50	55	31	5.3	4	1.2	1	60	43	9.1	8	2.9	3	58	39	7.1	6	2.3	2
0.95	none	84	53	18.3	14	3.7	3	84	63	24	20	6.8	6	82	57	25	13	5.1	5
0.95	20	58	13	13.1	8	3.1	3	63	27	19.0	15	6.3	6	51	22	18.3	12	4.7	4
0.95	50	70	31	15.7	11	3.5	3	76	44	22	18	6.6	6	69	39	23	16	5.0	5

The first row of Table 32 shows that for independent subgroups, EWMA detects 0.5σ and 1.5σ shifts faster than both the CUSUM and the $\overline{X}$ charts with ARL(0.5) values of 3.8 versus 6.1 and 8.7 respectively. The use of a training set does not affect this choice because the EWMA chart is hardly affected by the use of training set at this range of ARL values. This offers an advantage over the other two charts that require large training sets to estimate $\overline{\overline{X}}$ and $s_{\overline{X}}$. A similar comparison can be performed on the multivariate charts when $\rho_0 = 0$ and MEWMA is deemed to be a better chart than both MC1 and MCUSUM. However, the univariate EWMA detects shifts faster than the MEWMA procedure and is not affected by the size of the training set. Therefore, when there is no within-subgroup correlation, the unmodified EWMA chart is the best alternative among all the considered charts.

When the observations in the subgroup are highly correlated, with $\rho_0 = 0.95$, the modified EWMA chart has the highest speed among the univariate charts for detecting a shift in the level of the PQV, disregarding the size of the shift, the method of correction or the training set size. Furthermore, no significant difference is observed between the use of C and $s_{\overline{X}}$ for the correction. The use of $s_{\overline{X}}$ may be preferred since C can be employed only if the members of the subgroup have equal variances, as discussed in section 2.3.1. Among the multivariate charts, MEWMA has the best performance. Nevertheless, the modified EWMA is preferred over MEWMA for the same reasons cited when $\rho_0 = 0$.

Based on comparison of the ARLs, the results indicate that the traditional EWMA chart is the best chart to use when $\rho_0 = 0$ and its modified version with $s_{\overline{X}}$ as a correction when $\rho_0 = 0.95$. The same conclusions are obtained when the comparison is based on the MRLs. These values are displayed in Appendix C.

The next chapter describes the processes involved in the fabrication of integrated circuits and where the results of this research fit in. The use of these results is demonstrated on two real processes and their usefulness is discussed.

Chapter 5

SEMICONDUCTOR FABRICATION

Recent advances in technology owe their success to progress in the development of semiconductor devices. Integrated circuit technology makes it feasible to incorporate complex electronic functions on a small chip which has dimensions in the order of a few millimeters. This chapter presents an overview of the fabrication procedure of integrated circuits and describes two situations where the results of this research were applied.

5.1 Process Overview

Integrated circuits (ICs) consist of small silicon chips on which numerous devices are fabricated. Each IC contains hundreds of thousands of devices such as transistors, capacitors, resistors and inductors. The typical IC size is approximately 100-200 mm^2 whereas that of a transistor is 0.001 mm^2.

Fabrication of integrated circuits takes place on the surface of a planar silicon crystal which is called a wafer or substrate. Hundreds of ICs or die are fabricated on a single wafer (Figure 13). After fabrication is completed, the circuits are tested and the die separated from the wafer. Then, they are assembled on a die deck and placed in a sealed package for installation on a circuit board.

The wafers or substrates are made of single crystal silicon. This material is prepared from electronic grade silicon, a highly pure polycrystalline substance. The objective of IC fabrication is to introduce different impurities to the substrate in order to obtain certain electrical properties. This process is called doping. There are two types of semiconducting materials, n and p, which are produced when an n-type or p-type dopant is introduced to the substrate. The n-type dopant is a group V element and acts as an electron donor. The p-type dopant belongs to group III elements and is an electron acceptor. Phosphorous and arsenic are examples of the n-type dopants, while boron and gallium are of the p-type.

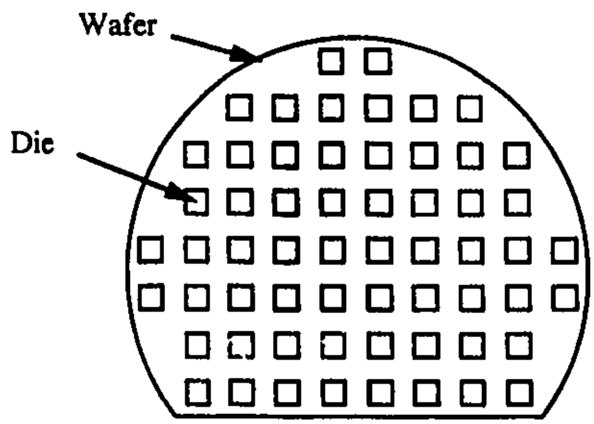


Figure 13. A planar silicon wafer

The fabrication procedure involves the creation of transistors, capacitors and resistors on the surface of the wafer. Each one of these devices uses the features of the two types of semiconductors with metal connections to the outside world. A transistor is basically a junction between an n and a p region and a metal layer (Figure 14(a)). The IC resistor consists of a slab of doped silicon (Figure 14(b)). A capacitor consists of a heavily doped region that acts as one plate, a silicon oxide layer as the dielectric and a metal layer as the second plate (Figure 14(c)).

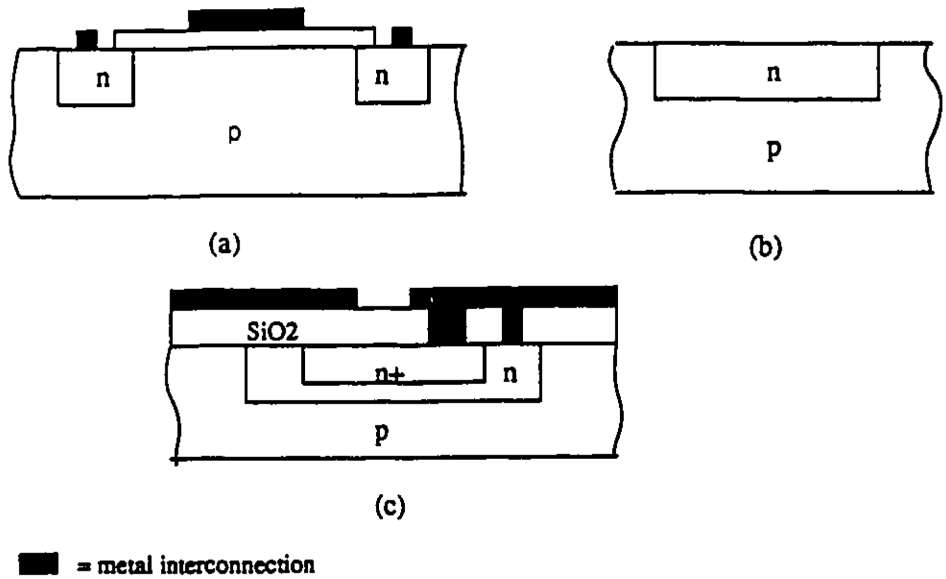


Figure 14. Cross sections of (a) a transistor, (b) a resistor and (c) a capacitor

The above devices are built on the surface of the wafer one layer at a time. Each layer has its particular pattern. As an illustration, the construction of a p region is described in Figure 15. A p region is created by the doping process which involves either one of two techniques: diffusion or ion implantation. The regions of the surface which do not require doping have to be protected with a coat of dopant resisting material. Typically, the resistant material consists of SiO_2 for diffusion and Si_3N_4 for ion implantation. A process called etching removes the unwanted parts of the coat which covers the whole surface of the wafer. However, before etching can be performed, the pattern which defines the p region must be transferred to the surface. A photosensitive layer of etch resistant material, known as photoresist, is applied over the SiO_2 or Si_3N_4 layer. A mask with opaque and transparent parts that define the required pattern is placed over the wafer that is covered with photoresist. When ultraviolet light strikes this layer, it polymerizes and the unpolymerized regions are washed away with the appropriate solvent. Etching removes of the SiO_2 or Si_3N_4 regions which are not covered with polymerized photoresist. Consequently, when the remaining photoresist is stripped, the surface becomes ready for doping in the desired regions. Figure 15 gives a logical summary of the sequence of these steps.

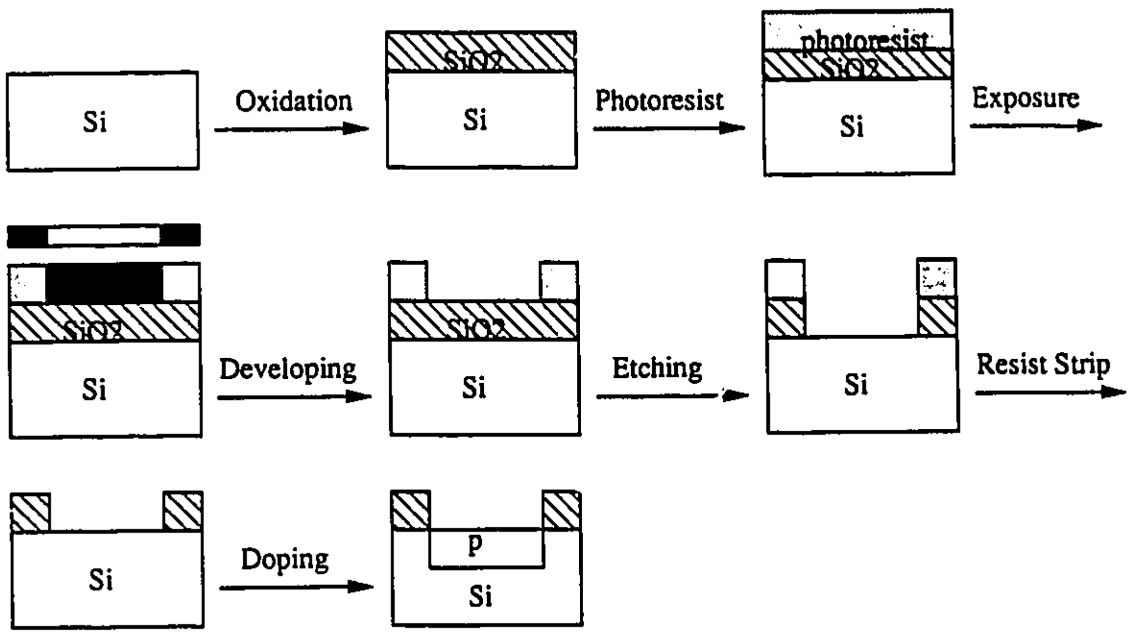


Figure 15. Basic steps of IC fabrication for creating a p region

During fabrication, the combination of the steps that are described above are repeated. The basic operations will be discussed below. They are: oxidation, lithography, etching, doping, chemical vapor deposition and metallization.

Oxidation. When silicon is exposed to an oxidizing atmosphere, such as oxygen, it oxidizes to form silicon dioxide according to the following reaction:

$$Si + O_2 \dashrightarrow SiO_2$$

The rate of this reaction is high at room temperature. However, the produced SiO_2 has a high density that prevents oxygen from diffusing any further into the bulk of the substrate to create thick layers of oxide. Amorphous SiO_2, the appropriate medium for oxygen diffusion, forms at temperatures ranging between 600°C and 1000°C. Usually, this type of SiO_2 is thermodynamically unstable below 1710°C and changes into the dense crystalline type when it reaches equilibrium. Nevertheless, the rate of change is negligible below 1000°C and at room temperature the SiO_2 is essentially stable provided no contaminants exist in the structure (Browman et al.,1990).

In IC fabrication, SiO_2 has several uses. SiO_2 serves as a mask against diffusion or as an oxide pad under Si_3N_4 films when ion implantation is used. It also isolates individual devices in the IC where it is desired to insulate the active electrical elements. The oxide is used as an insulator for the silicon from thin film devices and interconnections that are formed above the SiO_2 layer. Furthermore, SiO_2 provides surface protection and passivation once the fabrication is completed.

In the semiconductor industry, silicon is oxidized by one of three methods. The conventional method is thermal oxidation. Here, the temperature ranges between 600°C and 1200°C and the oxidizing species is either dry or wet oxygen, which is saturated with steam. The second method is high pressure oxidation (HPO). The idea behind this method is that as pressure increases, the flux of oxidizing species to the surface and hence the oxide growth rate both increase. The oxidation time decreases as well. The range of temperature is between 600°C and 1100°C, while the gauge pressure varies between 57 and 110 psi (Browman et al., 1990). The third method is chemical vapor deposition and will be discussed in a subsequent section.

Lithography. Lithographic processing involves the transferring of a pattern onto a surface using ultraviolet light. The devices that are to be created and duplicated on the wafer are built one layer at a time. Lithography transfers the pattern of each layer from a master pattern to a photosensitive material on the surface of the wafer. The main goal of lithography is to perform this task as accurately as possible.

The flowsheet of a typical photolithographic process is shown in Figure 16. Before coating the substrate with photoresist, the surface is primed with an adhesion promoter which is usually hexamethyldisilazane (HMDS). A uniform layer of resist is then applied to the wafer by a spin coating process.

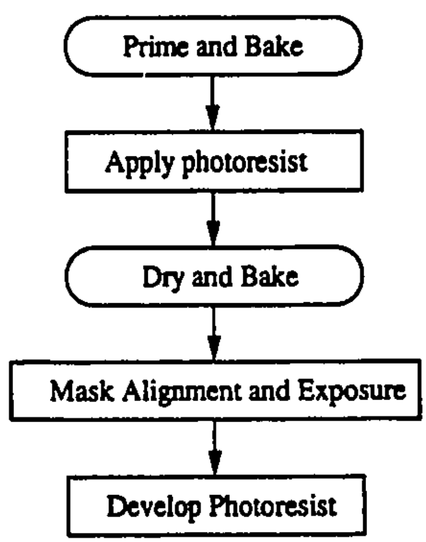


Figure 16. Basic steps of lithography

Since the photoresist is sensitive to white light, the lithographic process is performed in a yellow lighted area to prevent premature development. The photoresist material should have two basic properties. First, it should react with light in the exposed areas; and second, it must protect the underlying materials during etching and doping. Furthermore, photoresist has two forms: negative and positive as shown in Figure 17. Negative resists polymerize during light exposure and the unexposed parts are washed away. Thus, with negative resist, the inverse image of the mask is engraved. The opposite happens with positive resists, where the exposed areas break down and are washed away easily during development.

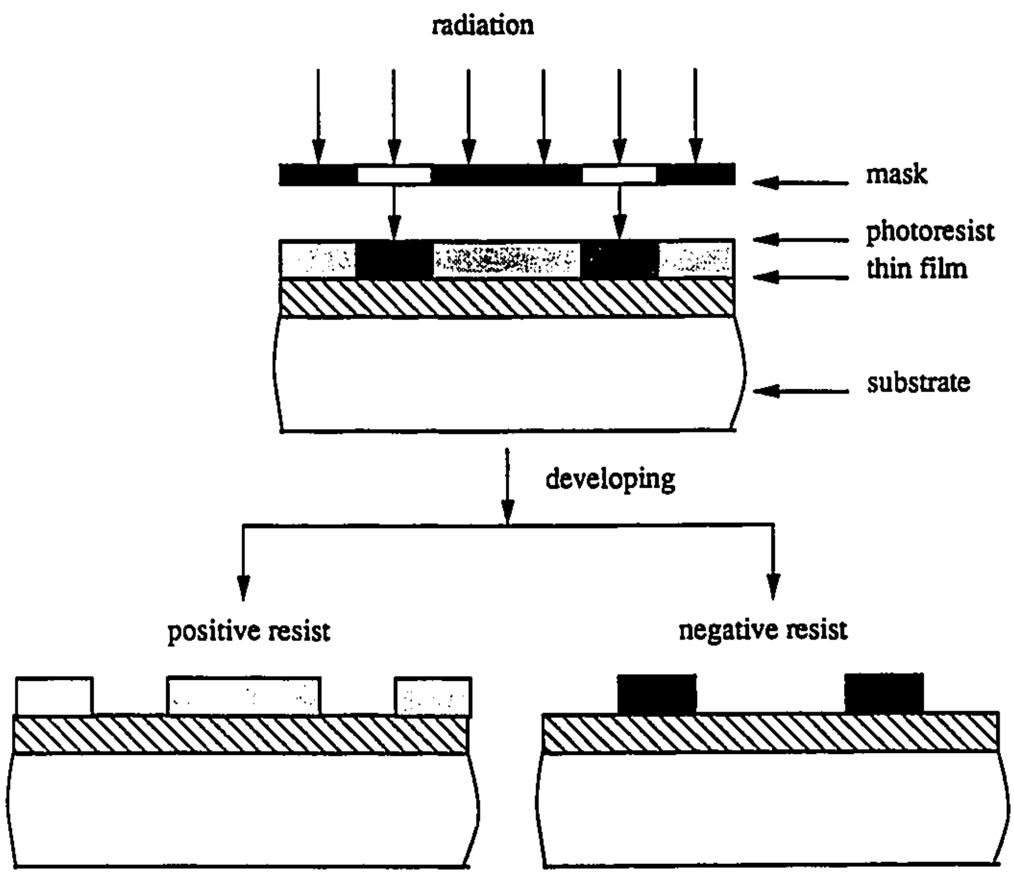


Figure 17. The difference between positive and negative resist

Next, the resist is soft baked within a temperature range of 85°C to 120°C. Soft baking reduces the solvent content in the resist, making it easier to handle. It also promotes the adhesion of the resist to the surface.

There are several techniques for transferring the mask pattern to the wafer. One technique is contact printing, where the mask is placed in contact with the wafer. However, there are a lot of problems associated with this method. The mask consists of clear and opaque areas that define the circuit pattern. The diffraction of light at the edges of the opaque areas may produce a blurred resolution of the image. Another technique is proximity printing which relies on the existence of a gap between the mask and the wafer. This technique requires a very small and constant spacing. It is also incapable of maintaining low defect levels when wafers become larger and layer to layer registration requirements become tighter (Browman et al., 1990).

A third technique provides better resolution and fewer limitations. It is called projection printing. Here, several lens elements focus the mask image on the substrate which is several centimeters away. Two mechanisms, stepping and scanning, generate the patterns. A stepper exposes small groups of device images on the wafer and repeats this action until all the die on the wafer are covered. On the other hand, a scanner projects the image of several and sometimes all the die at the same time.

After the mask is applied, the resist must be developed to remove it from the areas that are to be etched. There are three methods for developing resists: immersion, spray and puddle developing. During immersion developing, a batch of wafers is immersed in the developing chemical and then rinsed immediately. Spray developing involves spraying one wafer or a batch of wafers with developer solution while they are spinning. Puddle developers resemble spray developers when one wafer is processed at a time, except that the developer solution is dispensed on the wafer when it is static.

The last step in lithography is hard baking the resist. This enhances the adhesion of the resist to the wafer and removes residual solvents that remain from the previous steps. Hard baking also increases the etch resistant properties of the photoresist layer.

Etching. Once the lithographic operation is complete, etching is performed to permanently transfer the pattern in the photoresist to the underlying substrate. Typically, it is required to etch silicon, silicon dioxide, silicon nitride and aluminum. There are two types of etching techniques: wet etching and dry or plasma etching.

Wet etching involves the immersion of a batch of wafers in an etchant solution. A mixture of HNO_3 and HF acts as a an etchant for silicon layers, while a buffered HF solution is the most widely used etchant for SiO_2 layers. Hot H_3PO_4 with refluxing or aqueous HF are employed to etch Si_3N_4 regions. CH_3COOH and H_2O provide the best etchant solution for aluminum metallization layers (Browman et al., 1990). Wet etching is highly selective and is useful for removing thin films. However, it has a major drawback in that the process is isotropic, that is, it proceeds at the same rate in all directions. This implies that etching occurs in undesirable regions such as under the photoresist layers in the case of SiO_2 etching. This is known as undercut.

Dry or plasma etching provides a better alternative than wet etching. Etching occurs primarily in the vertical direction; therefore, little or no undercut will be present. The basic principles of operation of dry etching follow. When an electric field of sufficient magnitude is applied to a gas, it breaks down and becomes ionized. A plasma is a fully or partially ionized gas which consists of ions, electrons and neutrons (Sze, 1988). The gas should be able to ionize and form free radicals that will diffuse and react with the exposed regions that are to be etched. The products of the reaction are compounds that will volatilize from the substrate surface. Typical plasmas are: SF_3, Cl_2, CF_4, H_2, BC_3, CCl_4 and O_2.

Doping. Doping is the process where impurities are introduced into the silicon substrate in order to modify its electrical characteristics. Doping is performed by two different operations: diffusion and ion implantation.

During diffusion, the substrate is placed in a gaseous atmosphere which contains the impurities that are to be introduced into the substrate. The impurities dissolve at the surface of the solid, where it is not masked with SiO_2, and then diffuse into the bulk of the material driven by the concentration gradient. This operation takes place at elevated temperatures in a controlled electric furnace called a diffusion furnace. The same furnace can also be used for oxidation and chemical vapor deposition. Diffusion is performed at temperatures between 900°C and 1200°C (Chirlian, 1981).

Ion implantation is another means for doping the substrate. During this operation, dopant ions are electrically accelerated and used to bombard target regions. An analyzing magnet selects the specific dopant ions by bending the ion paths according to their atomic weights. The ion energy controls the dopant penetration depth. Ion implantation can penetrate through SiO_2. On one hand, this is an advantage over diffusion because the size of the doped region, and therefore the voltage of a transistor, can be adjusted after the oxide layer is grown. On the other hand, doping through an oxide is a disadvantage because a non-oxide layer that resists ion implantation should be applied to protect the regions which do not require doping. This layer is typically a film of Si_3N_4 applied over the SiO_2 layer. Lithography and etching can be used as before to tailor the respective patterns.

Chemical Vapor Deposition. The chemical vapor deposition (CVD) process in IC fabrication is used to deposit thin films which can serve as doping masks, interconnections, insulation between conducting layers and passivation coating over finished devices. Typical film materials are SiO_2, Si_3N_4, polysilicon and metals. During this process, vaporized chemicals react to form a solid film on the surface of the wafer. The deposited films are characterized by controlled composition, uniformity and structure, as well as good adhesion and step coverage.

A typical CVD reactor consists of a reaction chamber, wafer carriers and gas inlet systems with mass flow and temperature controllers. There are three methods for CVD: atmospheric pressure (APCVD), low pressure (LPCVD) and plasma enhanced (PECVD). These methods will not be discussed here due to their complexity.

Metallization. The process of metallization involves the formation of metal lines that act as electrical interconnections between the various elements of the IC. A thin metal layer is deposited on the entire surface of the wafer and interconnections between devices are formed by etching the metal parts that are not to be used. The metal itself must meet several requirements such as providing low resistance and strong adhesion to the underlying surface. Aluminum, copper and gold are commonly used in the metallization process.

Physical vapor deposition processes are used to grow the metal film on the surface of the wafer. Deposition takes place under vacuum by one of three mechanisms. The first one is filament evaporation where the metal to be evaporated is placed on a filament which is heated to above the evaporation temperature. The metal evaporates and condenses on the wafers which are kept at a lower temperature. The second mechanism is electron beam evaporation. Here, the metal is placed in a crucible and heated with an electron beam of approximately 10 kilovolts. The metal evaporates and forms a highly pure layer on the cooled wafers. The third mechanism is sputtering where argon gas is ionized by an electric field and accelerated into a negatively charged target plate. This causes the atoms of the target to be rejected and to condense on the ambient surfaces of the wafers.

5.2 Application of SPC

SPC charts are used widely in all of the operations mentioned above. Wafers are processed in batches of 12-24 wafers each. The objective of using SPC is to monitor the within-wafer variability as well as wafer-to-wafer and batch-to-batch variability. A typical PQV consists of either a film thickness or the width of the lines or gaps that are patterned on the wafer. Five or six measurements of the PQV on the same wafer constitute a subgroup. The orientation of these measurements is either vertical, horizontal (relative to the flat edge of the wafer) or circular, as shown in Figure 18. The location of measurements depends on the nature of the operation and the direction in which the greatest variability across the wafer is expected to occur. A high within-subgroup correlation in this case is desirable since it is an indicator of uniform processing.

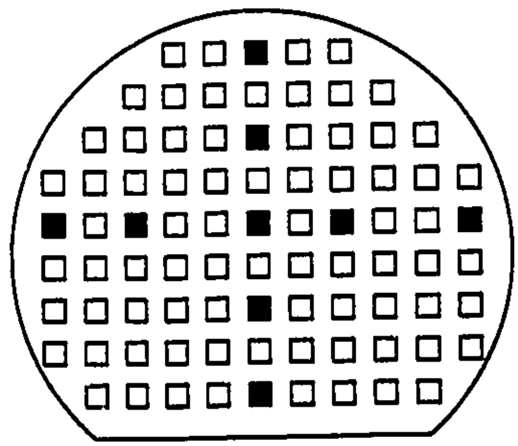


Figure 18. Possible locations for the measurements of the PQV on a wafer

Two examples where the results of this research were applied will be discussed below. The goal of these examples was to demonstrate the power and usefulness of the charts in monitoring the PQV.

5.2.1 Example 1. Photoresist Spin-Coating

This example is taken from the lithography area and involves coating the wafer with a thin film of photoresist. A diagram of this process is displayed in Figure 19. The liquid photoresist is dispensed at the center of the wafer which is placed on a spinning disk. The photoresist is spread evenly on the whole wafer under the action of the centrifugal force. The PQV is the film thickness which is adjusted by changing the spinning speed of the wafer or the spinning time. In this case, the spinning time is held constant and the spinning speed is the only adjustable variable with direct impact on the film thickness.

As far as sampling and data collection are concerned, each subgroup consists of 6 thickness measurements taken radially on the wafer because any variation of the film thickness is expected to be radial due to the spinning effects. The data shown in this example comes from a daily monitoring routine for the thickness.

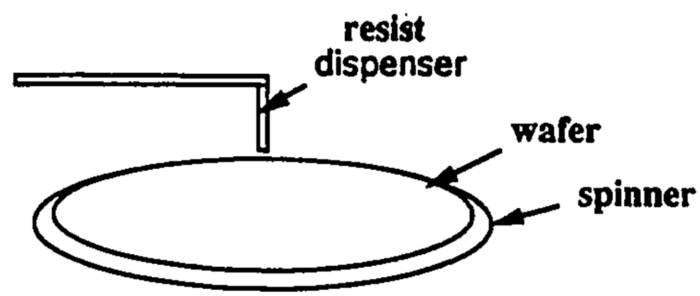


Figure 19. The photoresist spin-coating process.

A training set of 20 subgroups was collected over a three week period when the process was in a stable state (Figure 20). From the 20 subgroups, the sample correlation matrix was computed and confirmed the existence of within-subgroup correlation (Figure 21). The $\bar{\bar{X}}$ and $s_{\bar{x}}$ of the training set were determined and used to construct the modified univariate charts.

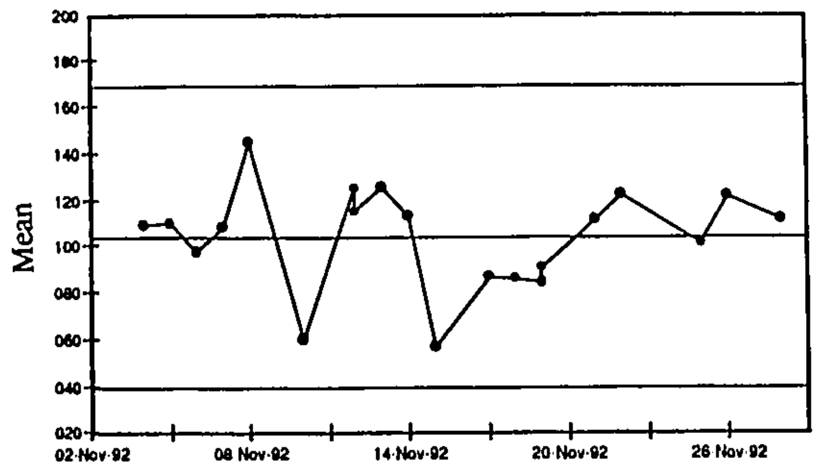
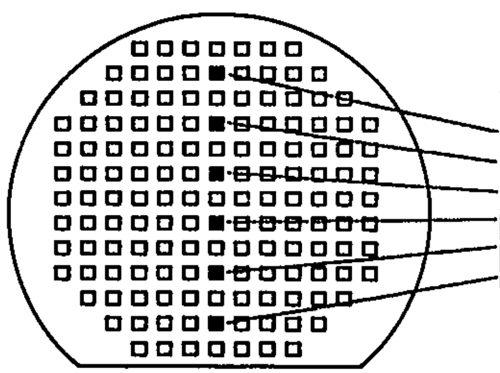


Figure 20. The modified $\overline{X}$ chart of the training set of example 1

Sample Plan



Correlation Matrix

	pos 1	pos 2	pos 3	pos 4	pos 5	pos 6
pos 1	1.00	0.96	0.91	0.72	0.70	0.57
pos 2	0.96	1.00	0.97	0.80	0.79	0.65
pos 3	0.91	0.97	1.00	0.88	0.89	0.78
pos 4	0.72	0.80	0.88	1.00	0.91	0.88
pos 5	0.70	0.79	0.89	0.91	1.00	0.96
pos 6	0.57	0.65	0.78	0.88	0.96	1.00

Figure 21. Correlation matrix and sampling scheme of the training set of example 1

Figure 22 shows the modified univariate charts for the subsequent 6 production weeks. Note that the same set of data is shown by all the charts. The 3 charts show that

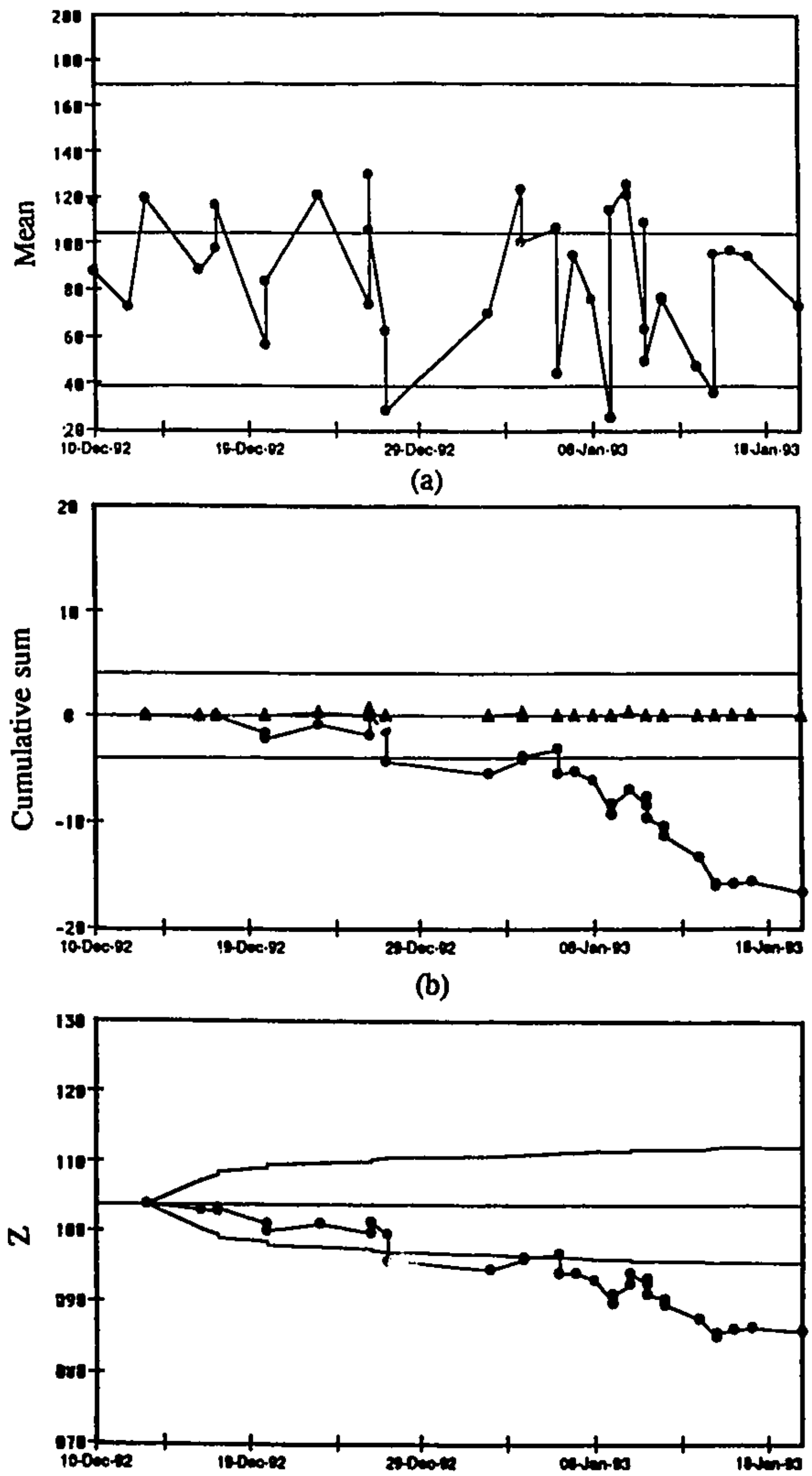


Figure 22. The modified (a) $\overline{X}$ (b) CUSUM and (c) EWMA charts for example 1

the process was in-control for the first three weeks. However, a downward shift was detected by all the charts after the third week, with the first out-of-control point being outside the corrected control limits as well as the specification limits (that is, the product has to be reworked). At that point, no corrective action was taken. After that, the charts continued giving an out-of-control signal for the subsequent weeks. Most of the points plotted on the modified $\overline{X}$ chart were centered below the center line indicating that a step change had occurred while the modified CUSUM and EWMA charts continued trending downwards.

Investigation of the situation revealed that the operators had changed the spinning speed of the wafer when it was not necessary to do so. Originally, all the operators had been given the authority to change the speed when the process starts drifting away from the mean. This led to constant interference with the spinning speed and it become a habit for the operators to introduce the changes. The first out-of-control signal led to more changes in the spinning speed in order to bring the process back on target and caused more variability in the process. Based on this investigation, it was enforced among the operators that the spinning speed should not be altered <u>unless</u> the thickness goes outside the control limits. Subsequently, the spinning speed was adjusted to an appropriate setting that brought the level of the PQV to its target value. The process remained in a stable state for the following few weeks, after which another shift was detected and was linked to some changes in the equipment which were never previously suspected to have an impact on the film thickness.

Figure 23 shows the standard univariate charts of the same data. The $\overline{X}$ chart has

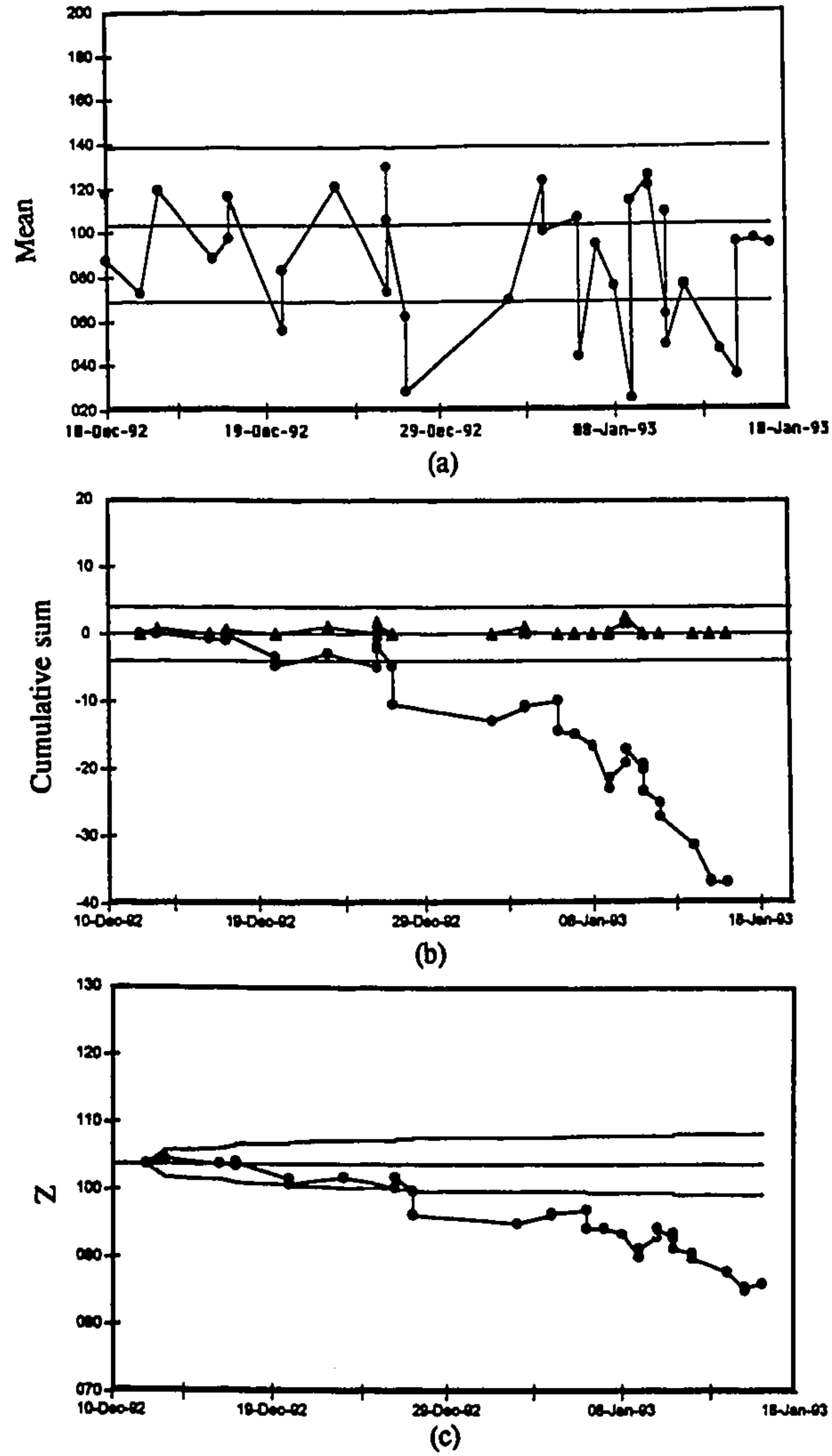


Figure 23. The standard (a) $\overline{X}$ (b) CUSUM and (c) EWMA charts for example 1

tighter control limits than the modified chart. It signals a false alarm at the 7'th point and detects the shift one point before the modified chart. This is expected since the computer simulations discussed in Chapter 4 show that the standard charts generally detect a shift faster than the modified charts; however, they produce too many false alarms in the presence of within-subgroup correlation. The standard CUSUM and EWMA charts exhibit the same behavior as the $\overline{X}$ chart. The frequency of false alarms would have been demonstrated more clearly had the shift not occur as early as it did.

This example demonstrates the usefulness of the modified charts in detecting a shift. In this case, the three modified charts detected the shift at the same time since it was a large shift. Therefore, there is no advantage for using one chart over the other in terms of shift detection. However, visually it is easier to observe the shift when the CUSUM or EWMA charts are used. This example also illustrates the usefulness of a powerful monitoring tool, such as the modified charts, in learning more about the process and the factors that affect it. If the standard charts were used instead, at least one false alarm would have been produced before the detection of the shift.

5.2.2 Example 2. Oxidation Furnaces

The second example involves the oxidation process. A schematic diagram of an oxidation furnace is shown in Figure 24. Each furnace consists of four sections, labelled as "Load", "C1", "C2" and "Source". The oxidizing gases enter at one end of the furnace and a diffusion reaction takes place on the surface of the wafers which are placed vertically in any of the sections. The gases then exit from the other end of the furnace. In this example, the PQV is the thickness of the SiO_2 film which is formed on the wafers in the "C2" section. The sampling plan consists of 5 thickness measurements taken at the center and the peripheries of the wafer as shown in Figure 25.

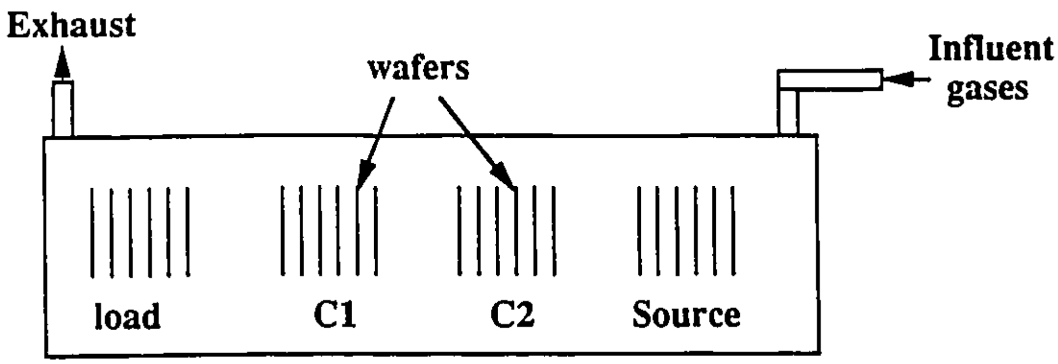
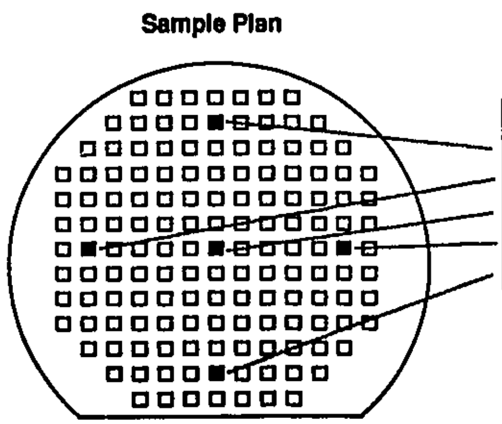


Figure 24. A schematic diagram of an oxidation furnace



Correlation Matrix

	C2_1	C2_2	C2_3	C2_4	C2_5
C2_1	1.00	0.89	0.86	0.76	0.91
C2_2	0.89	1.00	0.78	0.55	0.89
C2_3	0.86	0.78	1.00	0.78	0.80
C2_4	0.76	0.55	0.78	1.00	0.78
C2_5	0.91	0.89	0.80	0.78	1.00

Figure 25. Correlation matrix and sampling scheme of the training set for example 2

As in example 1, a training set of 20 subgroups was considered and the correlation matrix of the training set is displayed in Figure 25. The off-diagonal elements of **R** are all significantly different from 0 establishing the existence of within-subgroup correlation. The $\overline{X}$ chart of the training set is shown in Figure 26.

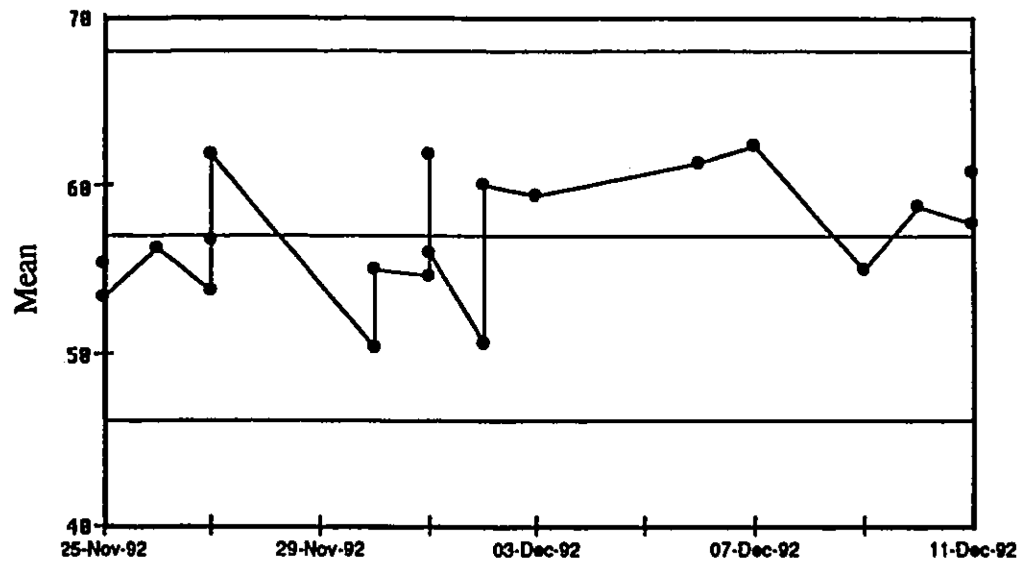


Figure 26. The modified $\overline{X}$ chart for the training set of example 2

The modified charts for the subsequent period are shown in Figure 27. The second point plotted on each of these charts triggered an alarm by the $\overline{X}$ and EWMA charts but not by the CUSUM chart which required an additional 9 points to signal its first alarm. The three charts showed an increasing trend in the PQV followed by a decreasing trend. Investigation of the cause of these alarms showed that they were linked to a drift in the setting of the mass flow controllers that regulate the incoming gases to the furnace. The increasing trend in the charts was reversed to a decreasing one when the monthly maintenance routine was performed revealing that the controllers were, in fact, miscalibrated. Subsequent calibration resolved the problem and brought the process back to target.

The standard charts are displayed in Figure 28. These charts have tighter control limits than the modified charts and most of the data points fall outside these limits. The shift is detected at the same time as the modifed charts since it was a large shift that occurred at the second point. More false alarms were expected if the process had been stable.

Again, this example demonstrates the value of the modified charts. Had the recommended EWMA chart been used in the monitoring of this process, the drift in the setting of the equipment would have been detected and corrected after the first alarm which corresponds to the second point on the chart, approximately 20 days prior to the date of the actual correction. This can be interpreted in terms of losses in production time and yield which, in turn, can be translated in terms of dollar savings.

Note that both the CUSUM and EWMA charts should be reset once a signal has been detected and removed. In both examples, the objective was to monitor the processes without any interference with production in order to demonstrate the benefits resulting from the use of the charts.

This chapter covered the basic operations of the fabrication of integrated circuits and illustrated the usefulness of the methods proposed in the research. The next chapter states the conclusions and recommendations of this study.

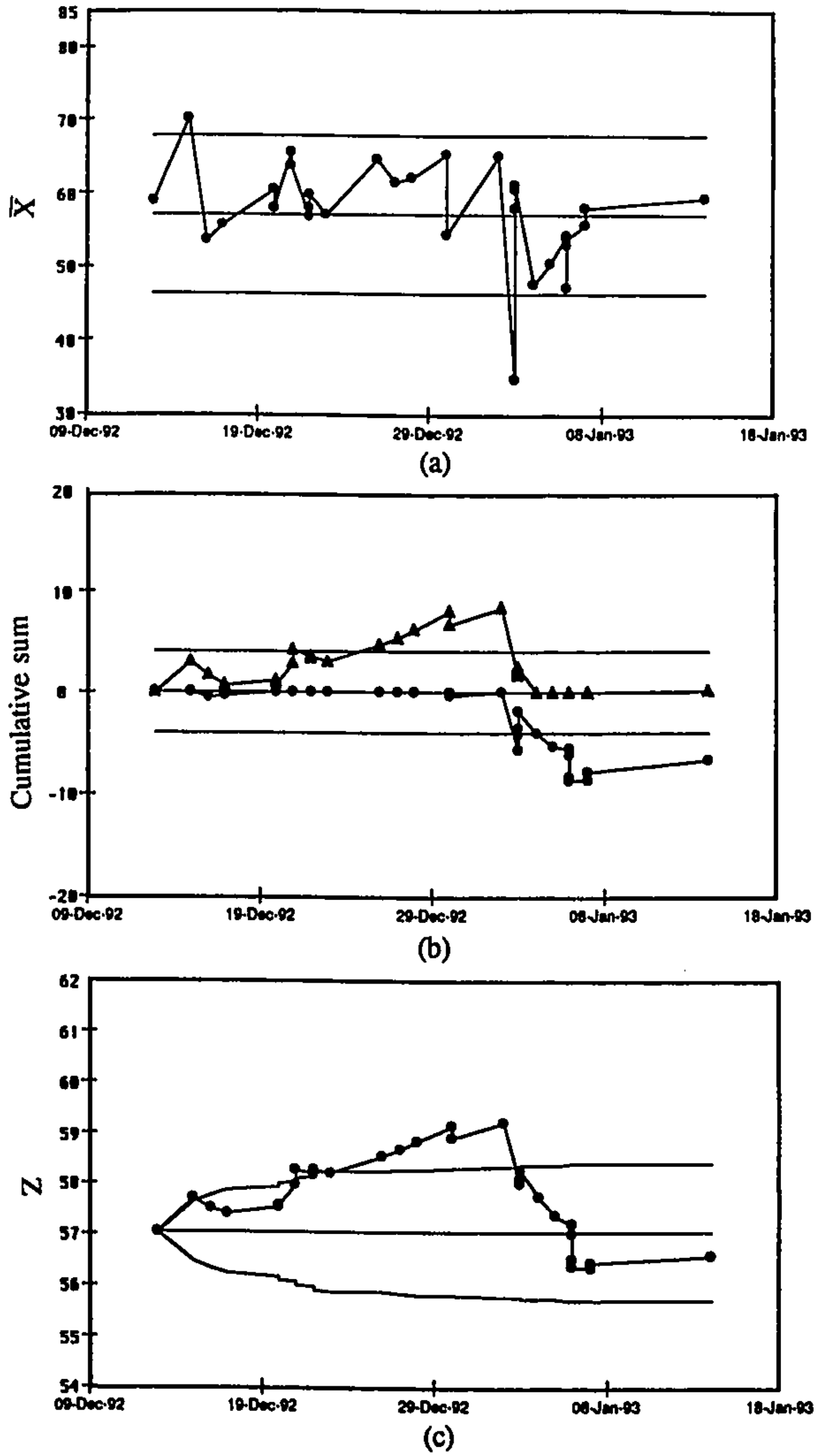


Figure 27. The modified (a) $\overline{X}$ (b) CUSUM and (c) EWMA charts for example 2

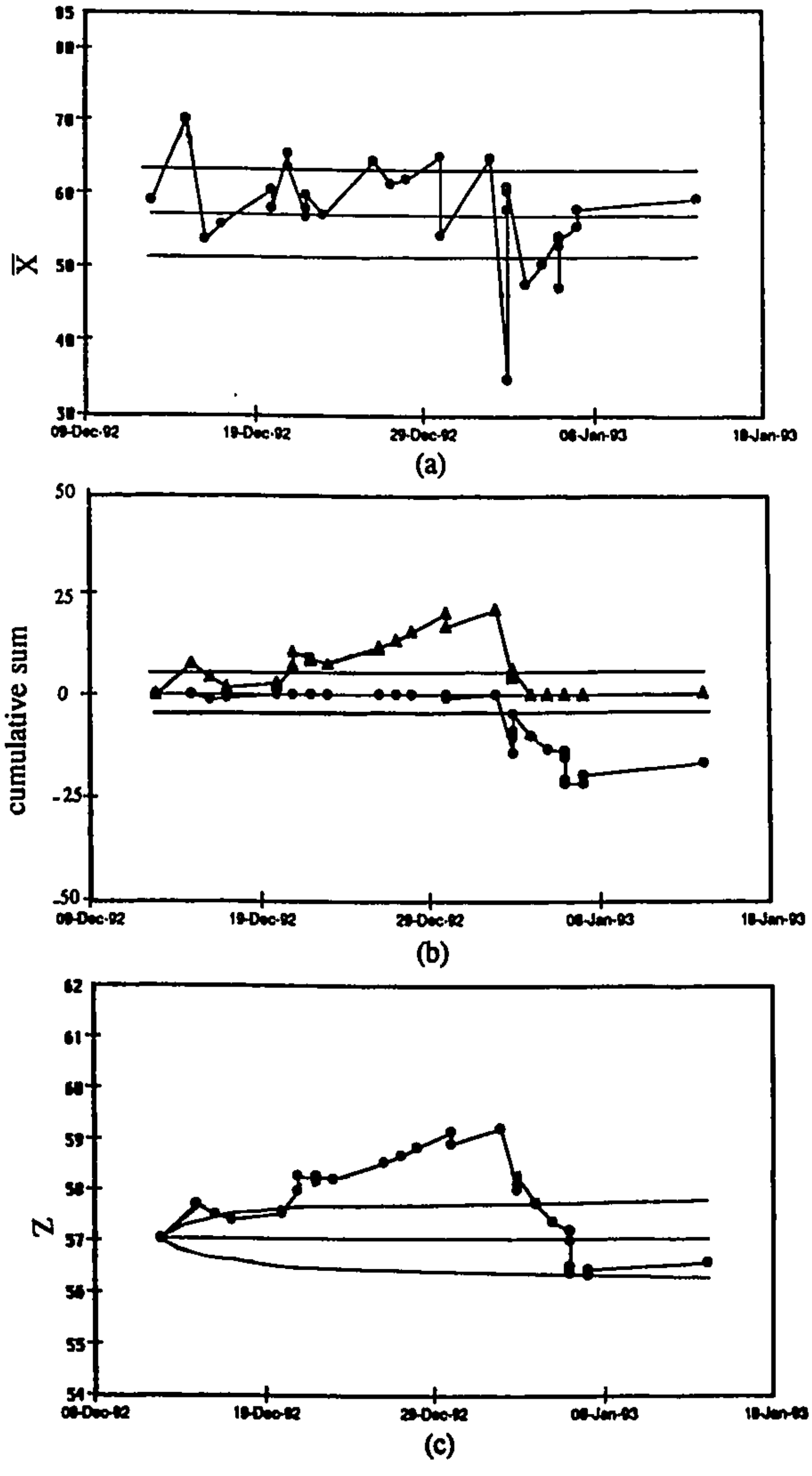


Figure 28. The standard (a) $\overline{X}$ (b) CUSUM and (c) EWMA charts for example 2

Chapter 6

CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

1. Computer simulations as well as actual processing data reveal that the existence of within-subgroup correlation degrades the performance of the standard SPC charts. The frequency of false alarms increases with increasing extent of correlation.

2. The tests of independence that were studied are inefficient. However, the partitioning or t distribution methods may be used to test for the existence of large correlations. Inspection of the sample correlation matrix of the measurements constituting the subgroups often gives an adequate idea of the degree of correlation within the subgroups.

3. The modified univariate charts reduce the number of false alarms that the within-subgroup correlation produces in the standard charts but the modified charts are not as sensitive to small shifts in the mean of the product quality variable.

4. Within-subgroup correlation does not affect the in-control run length of multivariate charts. Nevertheless, these charts lose their sensitivity to small shifts as the extent of correlation increases.

5. The use of a training set deteriorates the performance of most SPC charts. The effect is reduced as the training set size is increased.

6. Multivariate charts are affected more than univariate charts by the subgroup and training set sizes.

7. The use of $s_{\bar{x}}$ as a correction factor for univariate charts is preferred over C since it

yields run lengths which are closer to the true values. In addition, the use of $s_{\bar{x}}$ does not require that the members of the univariate subgroup to have equal variances.

8. Among the proposed SPC charts that were evaluated, the modified univariate EWMA chart has the best performance and its use is recommended.

9. The principal component of a subgroup with equal pairwise correlations is the subgroup mean.

10. The implementation of the modified charts in the semiconductor fabrication industry proved to be effective in detecting shifts and improving the monitored processes.

6.2 Recommendations for Future Work

1. A rigorous test of independence should be developed to detect the existence of correlations of small magnitudes.

2. Tables with optimal parameter values for the modified EWMA chart should be generated. The tables would provide λ and L values that correspond to desired ARL(0) and ARL(Δ) for a given degree of correlation.

3. A chart should be designed to monitor correlation structure, in order to observe any shifts in the processing conditions which may not be recognized otherwise.

4. The situation where both <u>within</u> and <u>between</u> subgroup correlation exists should be examined. So far, each of these cases has been explored separately but not simultaneously.

5. An appropriate method should be determined to simulate correlation matrices with off-diagonal elements that are not equal.

REFERENCES

Alt, F.B., "Multivariate Quality Control: State of the Art", Quality Congress Transactions, 886-893 (1982).

Alt, F.B., "Multivariate Quality Control", The Encyclopedia of Statistical Sciences, **6**, 110-122 (1984).

American Society for Testing Materials, ASTM Manual on Quality Control of Materials, ASTM, Philadelphia (1987).

Bard Y., Nonlinear Parameter Estimation, Academic Press, New York, 316-317 (1974).

Bartlett M. S., "A Note on the Multiplying Factors for Various χ^2 Approximations", Journal of the Royal Statistical Society, **16**, 196-298 (1954).

Bickel P. and K. Doksum, Mathematical Statistical Inference and Its Applications, Wiley, New York, 219-223 (1965).

Browman, R., G. Fry, J. Griffin, D. Potter and R. Skinner, eds., Practical VLSI Fabrication For the 90s, Integrated Circuit Engineering Corporation, Scottsdale, Arizona, (1990).

Burn, N.J., "The Effect of Correlated Sub-Groups on Shewhart X-bar/r and X-bar/S Control Charts", Internal Report, Northern Telecom Electronics, Ottawa (1988).

Burn, N.J. and D.D. McLean, Statistical Process Control, Class Notes, Department of Chemical Engineering, The University of Ottawa (1990).

Chirlian, P.M., Analysis and Design of Integrated Electronic Circuits, Harper & Row, New York, 84-98 (1981).

Crosier, R.B., "Multivariate Generalizations of Cumulative Sum Quality-Control Schemes", Technometrics, **30**, 291-303 (1988).

Hoerl, R.W. and A.C. Palm, "Discussion: Integrating SPC with APC", Technometrics, **34**, 268-272 (1992).

Holmes, D.S. and A.E. Mergen, "Testing Control Chart Subgroups For Rationality", Quality and Reliability Engineering International, **5**, 143-147 (1989).

Hotelling, H., "Techniques of Statistical Analysis", C. Eisenhart, M.W. Hastay and W.A. Wallis, Eds., McGraw-Hill, New York, 111-184 (1947).

Jackson, J.E., "Quality Control Methods for Several Related Variables", Technometrics, **1**, 359-377 (1959).

Johnson R. and D. Wichern, Applied Multivariate Statistical Analysis, Prentice Hall, Englewood Cliffs, New Jersey (1988).

Karson M.J., Multivariate Statistical Methods: An Introduction, Iowa State University Press, Ames, Iowa, 135-141 (1982).

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

Lawley, D.N., "On Testing a Set of Correlation Coefficients for Equality", Annals of Mathematical Statistics, **34**, 149-151 (1963).

Lowry, C.A., W.H. Woodall, C.W. Champ and S.E. Rigdon, "A Multivariate Exponentially Weighted Moving Average Control Chart", Technometrics, **34**, 46-53 (1992).

Lucas, J.M., "Combined Shewhart-CUSUM Quality Control Schemes", Journal of Quality Technology, **14**, 51-59 (1982).

Lucas, J.M. and M.S. Saccucci, "Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements", Technometrics, **32**, 1-12 (1990).

Nelson, L.S., "Interpreting Shewhart $\overline{X}$ Control Charts", Journal of Quality Technology, **17**, 114-116 (1985).

Neuhardt, J.B., "Effects of Correlated Sub-Samples on Statistical Process Control", IIE Transactions, **19**, 208-214 (1987).

Page, E.S., "Continuous Inspection Schemes", Biometrika, **41**, 100-115 (1954).

Pignatiello, J.J.Jr. and G.C. Runger, "Comparisons of Multivariate CUSUM Charts", Journal of Quality Technology, **22**, 173-186 (1990).

Roberts, S.W., "Control Chart Tests Based on Geometrics Moving Averages", Technometrics, **1**, 239-250 (1959).

Robinson, P.B. and T.Y. Ho, "Average Run Lengths of Geometric Moving Average Charts by Numerical Methods", Technometrics, **20**, 85-93 (1978).

Ryan, T.P., Statistical Methods for Quality Improvement, Wiley, New York (1989).

Sze, J., VLSI Technology, McGraw-Hill, New York, 184-193 (1988).

Western Electric Company, "Statistical Quality Control Handbook", New York (1956).

Woodall, W.H. and N.M. Ncube, "Multivariate CUSUM Quality Control Procedures", Technometrics, **27**, 285-292 (1985).

Appendix A
GENERATION OF CORRELATED DATA

Let Σ be the true covariance matrix of the measurements that constitute a subgroup.

$$\Sigma_{pxp} = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 & . & . & . & \sigma_{1p}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & . & . & . & \sigma_{2p}^2 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \sigma_{p1}^2 & \sigma_{p2}^2 & . & . & . & \sigma_{pp}^2 \end{bmatrix}$$

(Σ has the desired correlation structure that is to be simulated).
The eigenvalues and eigenvectors of Σ satisfy the relationship

$$\mathbf{U}^T\Sigma^T = \Lambda\mathbf{U}^T$$

where $\mathbf{U}^T$ is a matrix whose rows are the normalized eigenvectors

$$\mathbf{U}^T = \begin{bmatrix} \mathbf{U}_1^T \\ \mathbf{U}_2^T \\ . \\ . \\ . \\ \mathbf{U}_p^T \end{bmatrix} \quad \text{where } \mathbf{U}_i^T = [U_{i1}\ U_{i2}\ .\ .\ .\ U_{ip}] \quad \text{and } \mathbf{U}_i^T\mathbf{U}_i = 1.0$$

and Λ is the diagonal matrix of eigenvalues

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & . & . & . & 0 \\ 0 & \lambda_2 & . & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & \lambda_p \end{bmatrix}.$$

Let $N_{(nxp)}$ be a vector of random Normal numbers with mean 0 and variance 1. We can define the nxp matrix $\mathbf{E}$ such that $\mathbf{E} = (\mathbf{R}\ \Lambda^{0.5})U$. $\mathbf{E}$ is a matrix of p Normally distributed variables with mean 0 and covariance Σ . More details can be found in Bard (1974), Appendix D.

Appendix B
COMPUTER SIMULATION PROGRAMS

The computer simulations were performed using the RS1 software on a VAX mainframe. The following computer listings are written in RPL: RS1's programming language. The commands that are used can be found in any RS1 user manuals. Sample simulation programs are listed below. The complete set of programs can be found in "Within-Subgroup Correlations: Simulation Programs" which is available upon request from Dr. D. McLean, Department of Chemical Engineering.

TESTS OF INDEPENDENCE

Generalized Likelihood Ratio Test

```
PROCEDURE(N,p,r,g);
/*  CALLING PROCEDURE TO SIMULATE THE GENERALIZED LIKELIHOOD RATIO TEST
    VARIABLES     N=DATA SET SIZE
                  P=SUBGROUP SIZE
                  R=CORRELATION COEFFICIENT
                  g=NUMBER OF ITERATIONS
*/
cou=lastrow("@simulate@pres3");
res=vec$alloc(5);ssum=0;
type nocr "p= ",p, " r=",r;
/*  PERFORM TEST OF INDEPENDENCE AND COUNT THE NUMBER OF SUCCESSES */
do 13=1 to g;
     call GEN(n,p,r,flagg);
     ssum=ssum+flagg;
end;
r1=ssum*100/g;
res[1]=n;res[2]=p;res[3]=r;res[4]=g;res[5]=r1;
cou=cou+1;
/*  COPY RESULTS TO A TABLE */
call vec$v2trc(res,"@simulate@pres3",cou,false);
end;
```

Procedure GEN

```
procedure(n,p,rho,flagg);
/* The following procedure is designed to TEST FOR INDEPENDENCE
OF CORRELATION MATRIX-GENERAL LIKELIHOOD RATIO TEST (JOHNSON AND WICHERN)*/
alpha=0.05;sig=mat$alloc(p,p);
tnale=mat$alloc(p,p);
/* GENERATE CORRELATION MATRIX  */
do i=1 to p;
   do j=1 to p;
      if i<>j then sig[i,j]=rho;
      if i=j then sig[i,j]=1;
   end;
end;
/* GENERATE NORMAL DATA SET */
do k=1 to n;
   do i=1 to p;
      s=0;
      do j=1 to 12;
      s=s+ran();
      end;
      tnale[k,i]=s-6;
   end;
end;
call $svd(sig,u,s,v);
s=vec$sqrt(s);
s=mv$mdiag(s);
matt=mat$mxm(s,mat$transpose(u));
e=mat$mxm(tnale,matt);
if tableexists("@simulate@cor") then del "@simulate@cor";
call $mvcorr(mat$m2tt(e),true,"@simulate@cor",0,false);
r=mat$t2m("@simulate@cor",true);
/* PERFORM TEST OF INDEPENDENCE */
RDET=SVD$DET(R);
LAM=-N*LOG10(RDET);
IF P=4 THEN CHI=1.635;
IF P=5 THEN CHI=3.940;
IF P=6 THEN CHI=7.26;
if LAM>CHI then
   {
   flagg=0;return;
   }
else flagg=1;
/* RECORD SUCCESS OR FAILURE AND RETURN TO MAIN PROCEDURE */
end;
```

Partitioning Method

```
PROCEDURE(N,p,r,g);
/* CALLING PROCEDURE TO SIMULATE THE PARTITIONING METHOD
        N=DATA SET SIZE
        P=SUBGROUP SIZE
        R=CORRELATION COEFFICIENT
        G=NUMBER OF ITERATIONS*/
cou=lastrow("@simulate@cres");
res=vec$alloc(5);ssum=0;
/*PERFORM TEST G TIMES AND COUNT NUMBER OF SUCCESSES*/
do l3=1 to g;
      call cluscom(n,p,r,flagg);
      ssum=ssum+flagg;
end;
rl=ssum*100/g;
/* REPORT RESULTS AS A PERCENTAGE */
res[1]=n;res[2]=p;res[3]=r;res[4]=g;res[5]=rl;
cou=cou+1;
call vec$v2trc(res,"@simulate@cres",cou,false);
end;
```

Procedure CLUSCOM

```
procedure(n,p,rho,flagg);
/* The following procedure is designed to test for independence of p variables
by THE PARTITIONING METHOD*/;
sig=mat$alloc(p,p);
tnale=mat$alloc(p,p);
/* SET UP CORRELATION MATRIX */
do i=1 to p;
   do j=1 to p;
      if i<>j then sig[i,j]=rho;
      if i=j then sig[i,j]=1;
   end;
end;
/* GENERATE NORMAL DATA SET AND INDUCE CORRELATION */
do k=1 to n;
   do i=1 to p;
      s=0;
      do j=1 to 12;
      s=s+ran();
      end;
      tnale[k,i]=s-6;
   end;
end;
call $svd(sig,u,s,v);
s=vec$sqrt(s);
s=mv$mdiag(s);
matt=mat$mxm(s,mat$transpose(u));
e=mat$mxm(tnale,matt);
if tableexists("@simulate@cor") then del "@simulate@cor";
call $mvcorr(mat$m2tt(e),true,"@simulate@cor",0,false);
r=mat$t2m("@simulate@cor",true);
rdet=svd$det(r);
/* PERFORM TEST OF INDEPENDENCE */
lamst=rdet**(2/n);
m=(n-1.5)-((((P**3)-p)/(3*((p**2)-p)));
k=-m*log(lamst);
if p=4 then chi=1.635;
if p=5 then chi=3.940;
if p=6 then chi=7.261;
if k<chi then flagg=1;
else flagg=0;
/* RECORD FAILURE OR SUCCESS AND RETURN TO CALLING PROCEDURE */
end;
```

Pairwise t-distribtution Method

```
PROCEDURE(N,p,r,g);
/* CALLING PROCEDURE FOR THE T DISTRIBUTION METHOD
        N=DATA SET SIZE
        P=SUBGROUP SIZE
        R=CORRELATION
        G=NUMBER OF ITERATIONS */
cou=lastrow("@simulate@pres2");
res=vec$alloc(5);ssum=0;
type nocr "p= ",p, " r=",r;
/* PERFORM TEST OF INDEPENDENCE AND COUNT NUMBER OF SUCCESSES */
do 13=1 to g;
    call pair2COM(n,p,r,flagg);
    ssum=ssum+flagg;
end;
r1=ssum*100/g;
/* DOCUMENT RESULTS */
res[1]=n;res[2]=p;res[3]=r;res[4]=g;res[5]=r1;
cou=cou+1;
call vec$v2trc(res,"@simulate@pres2",cou,false);
end;
```

Procedure PAIR2COM

```
procedure(n,p,rho,flagg);
/* The following procedure is TEST FOR INDEPENDENCE USING THE T DISTRIBUTION
METHOD for paiwise correlation coefficients
*/;
alpha=0.05;sig=mat$alloc(p,p);
tnale=mat$alloc(p,p);
/* SET UP CORRELATION MATRIX */
do i=1 to p;
   do j=1 to p;
      if i<>j then sig[i,j]=rho;
      if i=j then sig[i,j]=1;
   end;
end;
/* GENERATE DATA SET WITH CORRELATION */
do k=1 to n;
   do i=1 to p;
      s=0;
      do j=1 to 12;
      s=s+ran();
      end;
      tnale[k,i]=s-6;
   end;
end;
call $svd(sig,u,s,v);
s=vec$sqrt(s);
s=mv$mdiag(s);
matt=mat$mxm(s,mat$transpose(u));
e=mat$mxm(tnale,matt);
if tableexists("@simulate@cor") then del "@simulate@cor";
call $mvcorr(mat$m2tt(e),true,"@simulate@cor",0,false);
r=mat$t2m("@simulate@cor",true);
```

```
/*  find max and min corr coefficients*/;
/*******************************************/;
max=-1.0;min=1.0;
do i=1 to p;
   do j=1 to i;
      if i <> j then
         {
         if r[i,j] > max then
            {
            max = r[i,j];
            imax=i;jmax=j;
            }
         if r[i,j] < min then
            {
            min = r[i,j];
            imin=i;jmin=j;
            }
         }
     end;
end;
/**********************************/;
h1=(((N-2)**0.5)*r[imax,jmax])/((1-(r[imax,jmax]**2))**.5);
h2=(((N-2)**0.5)*r[imin,jmin])/((1-(r[imin,jmin]**2))**0.5);
C=$PROBT_INV(1-(ALPHA*.5),N-2);
CC=$PROBT_INV(ALPHA*.5,N-2);
/*********check min and max intervals***********/;
if (((H1>C) OR (H1<CC)) AND (H2>C OR H2<CC)) then
   {
   flagg=0;
   }
else flagg=1;
/**********************************************/;
/* RECORD SUCCESS OR FAILURE AND RETURN TO MAIN PROCEDURE */

end;
```

Confidence Intervals Method

```
PROCEDURE(N,p,r,g);
/* CALLING PROCEDURE FOR CONFIDENCE INTERVALS METHOD */
cou=lastrow("@simulate@pres");
res=vec$alloc(5);ssum=0;
type nocr "p= ",p, " r=",r;
/* REPEAT TEST OF INDEPENDENCE G TIMES AND REPORT RESULTS */
Do l3=1 to g;
  call pair1(n,p,r,flagg);
  ssum=ssum+flagg;
End;
r1=ssum*100/g;
res[1]=n;res[2]=p;res[3]=r;res[4]=g;res[5]=r1;
cou=cou+1;
call vec$v2trc(res,"@simulate@pres",cou,false);
end;
```

Procedure PAIR1

```
procedure(n,p,rho,flagg);
/* The following procedure is designed to obtain confidence intervals
for paiwise correlation coefficients */;
alpha=0.05;sig=mat$alloc(p,p);
tnale=mat$alloc(p,p);
/* SET UP CORRELATION MATRIX */
do i=1 to p;
   do j=1 to p;
      if i<>j then sig[i,j]=rho;
      if i=j then sig[i,j]=1;
   end;
end;
/* GENERATE DATA SET */
do k=1 to n;
   do i=1 to p;
      s=0;
      do j=1 to 12;
      s=s+ran();
      end;
      tnale[k,i]=s-6;
   end;
end;
call $svd(sig,u,s,v);
s=vec$sqrt(s);
s=mv$mdiag(s);
matt=mat$mxm(s,mat$transpose(u));
e=mat$mxm(tnale,matt);
if tableexists("@simulate@cor") then del "@simulate@cor";
call $mvcorr(mat$m2tt(e),true,"@simulate@cor",0,false);
r=mat$t2m("@simulate@cor",true);
/*  find max and min corr coefficients*/;
/*****************************************/;
max=-1.0;min=1.0;
do i=1 to p;
   do j=1 to i;
      if i <> j then
         {
         if r[i,j] > max then
            {
            max = r[i,j];
            imax=i;jmax=j;
            }
```

```
          if r[i,j] < min then
             {
             min = r[i,j];
             imin=i;jmin=j;
             }
          }
      end;
end;
/************************************/;
/* PERFORM TEST OF INDEPENDENCE */
h1=0.5*log((1+r[imax,jmax])/(1-r[imax,jmax]));
h2=0.5*log((1+r[imin,jmin])/(1-r[imin,jmin]));
c=(1-(0.5*alpha));
d=probnorm_inv(c)/sqrt(n-3);
a1=tanh(h1-d);
a2=tanh(h2-d);
b1=tanh(h1+d);
b2=tanh(h2+d);
/*********check min and max intervals***********/;
if ((0<b1 and 0>a1) and (0<b2 and 0>a2)) then
   {
   flagg=1;return;

   }
else flagg=0;
/* RETURN TO MAIN PROCEDURE */
/**********************************************/;
end;
```

Univariate Control Charts

$\overline{X}$ chart without a training set

```
PROCEDURE(N,p,r,g,sh,code);
/*CALLING PROCEDURE FOR XBAR CHART WITH NO TRAINING SET
        N=MAX NUMBER FOR RUN LENGTHS
        P=SUBGROUP SIZE
        R=MAGNITUDE OF CORRELATION
        G=NUMBER OF SIMULATIONS
        SH=MAGNITUDE OF SHIFT
        CODE=CORRECTION CODE
            =1 FOR NO CORRECTION
            =2 FOR CORRECTION*/
cou=lastrow("@simulate@xres");
res=vec$alloc(5);x=vec$alloc(i);out=0;y=vec$alloc(i);
sum1=0;sum2=0;sum3=0;sum4=0;sum5=0;sum6=0;
type nocr "p= ",p," r=",r;
      do l3=1 to g;
/* COMPUTE RUN LENGTH AND REPORT WHICH RULE WAS VIOLATED*/
         call xbar(n,p,r,rl,sh,flagg,code);
         if flagg=0 then {rl=n;type l3;out=out+1;}
         if flagg=1 then sum1=sum1+1;
         if flagg=2 then sum2=sum2+1;
         if flagg=3 then sum3=sum3+1;
         if flagg=4 then sum4=sum4+1;
         if flagg=5 then sum5=sum5+1;
         if flagg=6 then sum6=sum6+1;
         x[l3]=rl;
      end;
arl=vec$mean(x);mrl=vec$median(x);
res[1]=n;res[2]=p;res[3]=r;res[4]=g;res[5]=arl;res[6]=mrl;res[7]=out;
res[8]=sh; res[9]=code;
cou=cou+1;
call vec$v2trc(res,"@simulate@xres",cou,false);
end;
```

Procedure XBAR

```
procedure(n,p,rho,rl,sh,flagg,code);
y=mat$alloc(n,p);z=vec$alloc(n);m1=vec$alloc(50);tscale=mat$alloc(p,p);
e=mat$alloc(20,p);tnale=mat$alloc(n,p);e1=mat$alloc(50,p);
sig=mat$alloc(p,p);c=vec$alloc(100);b=vec$alloc(100);a=vec$alloc(100);
d=vec$alloc(100);g=vec$alloc(100);
do i=1 to p;
   do j=1 to p;
      if i<>j then
         sig[i,j]=rho;
      if i=j then
         sig[i,j]=1.0;
   end;
end;
/**************************************************************
random number generator*/
call $svd(sig,u,s,v);
s=vec$sqrt(s);
s=mv$mdiag(s);
matt=mat$mxm(s,mat$transpose(u));
/* PERFORM NECESSARY CORRECTIONS */
if code= 1 then num=1;
if code = 2 then num=sqrt(1+((p-1)*rHO));
stdv=NUM/sqrt(p);xxbar=0;
```

```
/* DETERMINE THE ZONES OF THE CHART */
r1l=xxbar-stdv;
r1h=xxbar+stdv;
r2l=xxbar-(2*stdv);
r2h=xxbar+(2*stdv);
r3l=xxbar-(3*stdv);
r3h=xxbar+(3*stdv);
flagg=0;
counter=0;
/* GENERATE DATA POINTS*/
do ii=1 to n by 50;
   do k=1 to 50;
      do i=1 to p;
         ss=0;
         do j=1 to 12;
            ss=ss+ran();
         end;
         tscale[k,i]=ss-6;
      end;
   end;
/* INTRODUCE SHIFT*/
   e1=mat$mxm(tscale,matt);
   if sh><0 then{
   do k=1 to 50;
      do i=1 to p;
        e1[k,i]=e1[k,i]+sh;
      end;
   end;}
   m1=mat$collapse(e1,true,true);
   jj=0;
 do i=ii to (ii+49);
   jj=jj+1;
/*PLACE EACH POINT IN THE RESPECTIVE ZONE*/
   c[i]=0;b[i]=0;a[i]=0;d[i]=0;g[i]=-1;
   if (((m1[jj]<r1h) and (m1[jj]>xxbar)) or
       ((m1[jj]<xxbar) and (m1[jj]>r1l))) then c[i]=1;
   if ((m1[jj]<r2h) and (m1[jj]>r1h)) then b[i]=1;
   if ((m1[jj]<r1l) and (m1[jj]>r2l)) then b[i]=-1;
   if ((m1[jj]<r3h) and (m1[jj]>r2h)) then a[i]=1;
   if ((m1[jj]<r2l) and (m1[jj]>r3l)) then a[i]=-1;
   if ((m1[jj]<r3l) or (m1[jj]>r3h)) then d[i]=1;
   if (m1[jj]>xxbar) then g[i]=1;
/**********************************************************/
   counter=counter+1;
/* USE THE WESTERN ELECTRIC RULES TO DETERMINE RUN LENGTHS*/
   if d[i]=1 then {rl=counter; flagg=1; }
   if i>=3 then {sum1=abs(a[i]+a[i-1]+a[i-2]);
                if sum1>=2 then {rl=i;flagg=2;}}
   if i>=5 then {sum1=abs(b[i]+b[i-1]+b[i-2]+b[i-3]+b[i-4]);
                if sum1>=4 then {rl=i;flagg=3;}}
   if i>=8 then {sum1=abs(g[i]+g[i-1]+g[i-2]+g[i-3]+g[i-4]+g[i-5]+g[i-6]+g[i-7]
                );
                if sum1>=8 then {rl=i;flagg=4;}}
   if i>=15 then {sum1=c[i]+c[i-1]+c[i-2]+c[i-3]+c[i-4]+c[i-5]+c[i-6]+
                           c[i-7]+c[i-8]+c[i-9]+c[i-10]+c[i-11]+c[i-12]+
                           c[i-13]+c[i-14];
                if sum1>=15 then {rl=i;flagg=5;}}
   if i>=8 then {sum2=c[i]+c[i-1]+c[i-2]+c[i-3]+c[i-4]+c[i-5]+c[i-6]+c[i-7];
                if sum2=0 then {rl=i;flagg=6;}}
   if flagg>0 then return;
 end;
/*REPORT RUN LENGTH AND RETURN TO CALLING PROCEDURE*/
end;
end;
```

2-sided CUSUM chart with training set of size 20

```
PROCEDURE(N,p,r,g,kc,h,SH,CODE);
/* CALLING PROCEDURE FOR CUSUM CHART
        N= MAX. NUMBER OF POINTS TO BE SIMULATED
        P= SUBGROUP SIZE
        R= MAGNITUDE OF CORRELATION
        G= NUMBER OF ITERATIONS
        KC, H = CHART PARAMETERS
        SH= MAGNITUDE OF SHIFT
        CODE = FLAG FOR CORRECTION METHOD,
                IF CODE = 1 NO CORRECTION
                IF CODE = 2 CORRECTION APPLIED USING C
                IF CODE = 3 CORRECTION APPLIED USING S_XBAR
*/
cou=lastrow("@simulate@CUSRES");
res=vec$alloc(5);x=vec$alloc(g);out=0;
do 13=1 to g;
   call CUSCOM(n,p,r,kc,h,flagg,SH,CODE);
   if flagg=0 then {flagg=n;type 13;out=out+1;}
   x[13]=flagg;
end;
arl=vec$mean(x);mrl=vec$median(x);
res[1]=n;res[2]=p;res[3]=r;res[4]=g;res[5]=arl;res[6]=mrl;res[7]=out;
res[8]=h;RES[9]=SH;RES[10]=CODE;
cou=cou+1;
call vec$v2trc(res,"@simulate@cusres",cou,false);
end;
```

Procedure CUSCOM

```
procedure(n,p,rho,kc,h,flagg,SH,CODE);
/*PROCEDURE TO PERFORM CUSUM CHART SIMULATIONS WITH A TRAINING SET */
y=mat$alloc(n,p);z=vec$alloc(n);m1=vec$alloc(50);tscale=mat$alloc(p,p);
e=mat$alloc(20,p);tnale=mat$alloc(n,p);e1=mat$alloc(50,p);
sig=mat$alloc(p,p);temp=vec$alloc(n*p);
/*SETUP CORRELATION MATRIX*/
do i=1 to p;
   do j=1 to p;
      if i<>j then
         sig[i,j]=rho;
      if i=j then
         sig[i,j]=1.0;
   end;
end;
/*****************************************************************
GENERATE NORMAL RANDOM NUMBERS FOR TRAINING SET*/
tnale=mat$alloc(n,p);
do k=1 to 20;
   do i=1 to p;
      s=0;
      do j=1 to 12;
         s=s+ran();
      end;
      tnale[k,i]=s-6;
     end;
end;
call $svd(sig,u,s,v);
s=vec$sqrt(s);
s=mv$mdiag(s);
matt=mat$mxm(s,mat$transpose(u));
e=mat$mxm(tnale,matt);
st=stdev of cols 1 to p of table(mat$m2tt(e));
```

```
/* APPLY THE APPROPRIATE CORRECTION */
IF CODE=1 THEN STDV=(ST**2)/P;
IF CODE=2 THEN {
   DUM=mat$collapse(e,true,true);
   SXBAR=VEC$STDEV(DUM);
   STDV=SXBAR**2;
   }
IF CODE=3 THEN {
   if tableexists("@simulate@cor") then del "@simulate@cor";
   call $mvcorr(mat$m2tt(e),true,"@simulate@cor",0,false);
   IF RHO<>0 THEN
      {
      sum10=0;
      do i=1 to p;
         do j=1 to p;
            if i>< j then sum10=sum10+row i col j of table("@simulate@cor");
         end;
      end;
      rave=sum10/(p*(p-1));
      }
      IF RHO=0 THEN RAVE=0;
      num=sqrt(1+((p-1)*rave));
      stdv=(st*NUM/sqrt(p))**2;
    }
flagg=0;
counter=0;
/* GENERATE THE DATA SET 50 POINTS AT A TIME */
do ii=1 to n by 50;
   do k=1 to 50;
      do i=1 to p;
         ss=0;
         do j=1 to 12;
            ss=ss+ran();
         end;
         tscale[k,i]=ss-6;
      end;
   end;
   e1=mat$mxm(tscale,matt);
/*INTRODUCE SHIFT*/
   IF SH>0 THEN {
   do k=1 to 50;
      do i=1 to p;
        e1[k,i]=e1[k,i]+SH;
      end;
   end;}
   m1=mat$collapse(e1,true,true);
   jj=0;
/*CHECK CONTROL CRITERIA*/
 do i=ii to (ii+49);
   jj=jj+1;
   counter=counter+1;
   z[i]=(m1[jj]-xxbar)/(stdv**0.5);
   val=(z[i]-kc)+savex;
   val1=(-z[i]-kc)+savey;
   if val>0 then
      shi=val;
   else shi=0;
   if val1>0 then
      sli=val1;
   else sli=0;
   savey=sli;
   savex=shi;
   if shi > h then {flagg=counter;return;}
   if sli > h then {flagg=counter;return;}
 end;
end;
/*COMPUTE RUN LENGTH AND RETURN TO CALLING PROCEDURE*/
end;
```

EWMA chart with training set of size 50

```
PROCEDURE(N,p,r,g,lambda,l,SH,CODE);
/* CALLING PROCEDURE FOR EWMA CHART
        N= MAX. NUMBER OF POINTS TO BE SIMULATED
        P= SUBGROUP SIZE
        R= MAGNITUDE OF CORRELATION
        G= NUMBER OF ITERATIONS
        LAMBDA, L = CHART PARAMETERS
        SH= MAGNITUDE OF SHIFT
        CODE = FLAG FOR CORRECTION METHOD,
                IF CODE = 1 NO CORRECTION
                IF CODE = 2 CORRECTION APPLIED USING C
                IF CODE = 3 CORRECTION APPLIED USING S_XBAR
*/
cou=lastrow("@simulate@ewres");
res=vec$alloc(5);x=vec$alloc(g);out=0;
do l3=1 to g;
   call ewcom(n,p,r,lam,l,flagg,SH,CODE);
   if flagg=0 then {flagg=n;type l3;out=out+1;}
   x[l3]=flagg;
end;
arl=vec$mean(x);mrl=vec$median(x);
res[1]=n;res[2]=p;res[3]=r;res[4]=g;res[5]=arl;res[6]=mrl;res[7]=out;
res[8]=l;RES[9]=SH;RES[10]=CODE;
cou=cou+1;
call vec$v2trc(res,"@simulate@ewres",cou,false);
end;
```

Procedure EWCOM

```
procedure(n,p,rho,lamda,l,flagg,SH,CODE);
/*PROCEDURE TO PERFORM EWMA CHART SIMULATIONS WITH A TRAINING SET */
y=mat$alloc(n,p);z=vec$alloc(n);ml=vec$alloc(50);tscale=mat$alloc(p,p);
e=mat$alloc(20,p);tnale=mat$alloc(n,p);el=mat$alloc(50,p);
sig=mat$alloc(p,p);temp=vec$alloc(n*p);
/*SETUP CORRELATION MATRIX*/
do i=1 to p;
   do j=1 to p;
      if i<>j then
         sig[i,j]=rho;
      if i=j then
         sig[i,j]=1.0;
   end;
end;
/***************************************************************
GENERATE NORMAL RANDOM NUMBERS FOR TRAINING SET*/
tnale=mat$alloc(n,p);
do k=1 to 50;
   do i=1 to p;
     s=0;
     do j=1 to 12;
        s=s+ran();
     end;
     tnale[k,i]=s-6;
    end;
end;
call $svd(sig,u,s,v);
s=vec$sqrt(s);
s=mv$mdiag(s);
matt=mat$mxm(s,mat$transpose(u));
e=mat$mxm(tnale,matt);
st=stdev of cols 1 to p of table(mat$m2tt(e));
```

```
/* APPLY THE APPROPRIATE CORRECTION */
IF CODE=1 THEN STDV=(ST**2)/P;
IF CODE=2 THEN {
   DUM=mat$collapse(e,true,true);
   SXBAR=VEC$STDEV(DUM);
   STDV=SXBAR**2;
   }
IF CODE=3 THEN {
   if tableexists("@simulate@cor") then del "@simulate@cor";
   call $mvcorr(mat$m2tt(e),true,"@simulate@cor",0,false);
   IF RHO<>0 THEN
      {
      sum10=0;
      do i=1 to p;
         do j=1 to p;
            if i>< j then sum10=sum10+row i col j of table("@simulate@cor");
         end;
      end;
      rave=sum10/(p*(p-1));
      }
      IF RHO=0 THEN RAVE=0;
      num=sqrt(1+((p-1)*rave));
      stdv=(st*NUM/sqrt(p))**2;
    }
flagg=0;
counter=0;
/* GENERATE THE DATA SET 50 POINTS AT A TIME */
do ii=1 to n by 50;
   do k=1 to 50;
      do i=1 to p;
         ss=0;
         do j=1 to 12;
            ss=ss+ran();
         end;
         tscale[k,i]=ss-6;
      end;
   end;
   e1=mat$mxm(tscale,matt);
   IF SH>0 THEN {
   do k=1 to 50;
      do i=1 to p;
        e1[k,i]=e1[k,i]+SH;
      end;
   end;}
   m1=mat$collapse(e1,true,true);
   jj=0;
/*CHECK CONTROL CRITERIA*/
 do i=ii to (ii+49);
   jj=jj+1;
   counter=counter+1;
   szi=sqrt((lamda/(2-lamda))*(1-((1-lamda)**(2*i)))*stdv);
   if i=1 then z[i]=(lamda*m1[jj]);
   else z[i]=(lamda*m1[jj])+((1-lamda)*z[i-1]);
   lcl=-1*szi;
   ucl=+1*szi;
   if z[i] > ucl then {flagg=counter;return;}
   if z[i] < lcl then {flagg=counter;return;}
 end;
end;
/*COMPUTE RUN LENGTH AND RETURN TO CALLING PROCEDURE*/
end;
```

MEWMA with training set of size 20

```
PROCEDURE(N,p,r,i,lam,h4,SH);
/*CALLING PROCEDURE FOR MEWMA
        N=MAX. RUN LENGTH VALUE
        P=SUBGROUP SIZE
        R=MAGNITUDE OF CORRELATION
        I=NUMBER OF ITERATIONS
        LAM,H4 = PARAMETERS OF CHART
        SH=MAGNITUDE OF SHIFT*/
cou=lastrow("@simulate@mewres");
res=vec$alloc(5);x=vec$alloc(i);out=0;
/*CALL PROCEDURE I TIMES AND REPORT RESULTS*/
do 13=1 to i;
         call mewcom(n,p,r,lam,h4,flagg,SH);
         if flagg=0 then {flagg=n;type 13;out=out+1;}
         x[13]=flagg;
         collect;
end;
arl=vec$mean(x);mrl=vec$median(x);
res[1]=n;res[2]=p;res[3]=r;res[4]=i;res[5]=arl;res[6]=mrl;res[7]=out;
res[8]=lam;res[9]=h4;RES[10]=SH;
cou=cou+1;
call vec$v2trc(res,"@simulate@mewres",cou,false);
end;
```

Procedure MEWCOM

```
procedure(n,p,rho,lamda,h4,flagg,SH);
/*PROCEDURE FOR SIMULATING THE MEWMA CHART WITH TRAINING SET= 20*/
y=mat$alloc(n,p);z=vec$alloc(n);m1=vec$alloc(50);tscale=mat$alloc(p,p);
e=mat$alloc(20,p);tnale=mat$alloc(n,p);e1=mat$alloc(50,p);
sig=mat$alloc(p,p);z=vec$alloc(p);zz=mat$alloc(p,p);
/*SETUP CORRELATION MATRIX*/
do i=1 to p;
   do j=1 to p;
      if i<>j then
         sig[i,j]=rho;
      if i=j then
         sig[i,j]=1.0;
   end;
end;
/*GENERATE TRAINING SET*/
DO K=1 TO 20;
   DO I=1 TO P;
      S=0;
      DO J=1 TO 12;
         S=S+RAN();
      END;
      TNALE[K,I]=S-6;
   END;
END;
call $svd(sig,u,s,v);
s=vec$sqrt(s);
s=mv$mdiag(s);
matt=mat$mxm(s,mat$transpose(u));
e=mat$mxm(tnale,matt);
if tableexists("@simulate@cor") then del "@simulate@cor";
call $mvcorr(mat$m2tt(e),true,"@simulate@cor",0,false);
```

```
/****************************************************************
random number generator*/
flagg=0;
counter=0;
/*GENERATE DATA POINTS*/
do ii=1 to n by 50;
   do k=1 to 50;
      do i=1 to p;
         ss=0;
         do j=1 to 12;
            ss=ss+ran();
         end;
         tscale[k,i]=ss-6;
      end;
   end;
   e1=mat$mxm(tscale,matt);
/*INTRODUCE SHIFT*/
   IF SH >0 THEN
   {
   do k=1 to 50;
      do i=1 to p;
       e1[k,i]=e1[k,i]+SH;
      end;
   end;
   }
   jj=0;
/*CHECK CONTROL CRITERIA AND REPORT RUN LENGTHS*/
 do i=ii to (ii+49);
   jj=jj+1;
   counter=counter+1;
   szi=Mat$scale(mat$t2m("@simulate@cor"),(lamda/(2-lamda))*(
          1-((1-lamda)**(2*i))));
   sinv=svd$inverse(szi);

   x=mv$m2v(e1,false,jj);
   if i=1 then z=vec$mul(x,lamda);
   else z=vec$add(vec$mul(x,lamda),vec$mul(zi,(1-lamda)));
   zz=mv$v2m(z);
   zzt=mat$transpose(zz);
   tt=mat$mxm(mat$mxm(zzt,sinv),zz);
   t=tt[1,1];
   zi=vec$mul(z,1);
   if t > h4 then {flagg=counter;return;}
  end;
 end;collect;
 end;
```

MC1 with training set of size 20

```
PROCEDURE(N,p,r,g,K1,h2,sh);
/*CALLING PROCEDURE FOR SIMULATING THE MC1 CHART
        N=MAXIMUM VALUE FOR THE RUN LENGTH
        P=NUMBER OF QUALITY CHARACTERISTICS OR SUBGROUP SIZE
        R=MAGNITUDE OF CORRELATION
        G=NUMBER OF ITERATIONS
        K1,H2=CHART PARAMETERS
        SH=MAGNITUDE OF SHIFT*/
cou=lastrow("@simulate@mcres");
res=vec$alloc(5);x=vec$alloc(i);out=0;
/*COMPUTE RUN LENGTH G TIMES AND DETERMINE ARL AND MRL */
do l3=1 to g;
         call msum(n,p,r,k1,h2,flagg,sh);
         if flagg=0 then {flagg=n;type l3;out=out+1;}
         x[l3]=flagg;
         collect;
end;
arl=vec$mean(x);mrl=vec$median(x);
/* REPORT VALUES OF ARL AND MRL*/
res[1]=n;res[2]=p;res[3]=r;res[4]=g;res[5]=arl;res[6]=mrl;res[7]=out;
res[8]=k1;res[9]=h2;res[10]=sh;
cou=cou+1;
call vec$v2trc(res,"@simulate@mcres",cou,false);
end;
```

Procedure MSUM

```
procedure(n,p,rho,K1,h2,flagg,sh);
/*PROCEDURE FOR SIMULATING THE MULTIVARIATE CUSUM CHART (MC1) WITH
TRAINING SET SIZE=20*/
y=mat$alloc(n,p);z=vec$alloc(n);ml=vec$alloc(50);tscale=mat$alloc(p,p);
e=mat$alloc(20,p);tnale=mat$alloc(n,p);el=mat$alloc(50,p);
sig=mat$alloc(p,p);z=vec$alloc(p);zz=mat$alloc(p,p);
/*SETUP CORRELATION MATRIX*/
do i=1 to p;
   do j=1 to p;
      if i<>j then
         sig[i,j]=rho;
      if i=j then
         sig[i,j]=1.0;
   end;
end;
call $svd(sig,u,s,v);
s=vec$sqrt(s);
s=mv$mdiag(s);
matt=mat$mxm(s,mat$transpose(u));
tnale=mat$alloc(p,p);
/*GENERATE TRAINING SET OF SIZE 20*/
do k=1 to 20;
   do i=1 to p;
      s=0;
      do j= 1 to 12;
         s=s+ran();
      end;
      tnale[k,i]=s-6;
   end;
end;
e=mat$mxm(tnale,matt);
if tableexists("@simulate@cor") then del "@simulate@cor";
call $mvcorr(mat$m2tt(e),true,"@simulate@cor",0,false);
sigm=mat$t2m("@simulate@cor",true);
```

```
/*****************************************************************
random number generator*/
flagg=0;
counter=0;
SAVE=VEC$ALLOC(P);
DO I= 1 TO P;
   SAVE[I]=0;
END;
SINV=SVD$INVERSE(SIGm);
do ii=1 to n by 50;
   do k=1 to 50;
      do i=1 to p;
         ss=0;
         do j=1 to 12;
            ss=ss+ran();
         end;
         tscale[k,i]=ss-6;
      end;
   end;
/*INTRODUCE SHIFT*/
   e1=mat$mxm(tscale,matt);
  if sh>0 then {
  do k=1 to 50;
      do i=1 to p;
       e1[k,i]=e1[k,i]+sh;
      end;
   end;}
   jj=0;
/*GENERATE DATA POINTS AND CHECK FOR CONTROL CRITERIA*/
 do i=ii to (ii+49);

 jj=jj+1;
 counter=counter+1;
 x=mv$m2v(e1,false,jj);
 IM=VEC$ADD(SAVE,X);
 INN=MV$V2M(IM);
 INNT=MAT$TRANSPOSE(INN);
 C=MAT$MXM(MAT$MXM(INNT,SINV),INN);
 CI=SQRT(C[1,1]);
 IF CI<= K1 THEN SI=VEC$MUL(SAVE,0);
 ELSE SI=VEC$MUL(VEC$ADD(SAVE,X),(1-(K1/CI)));
 SIM=MV$V2M(SI);
 SIMT=MAT$TRANSPOSE(SIM);
 YI=MAT$MXM(MAT$MXM(SIMT,SINV),SIM);
 YIV=SQRT(YI[1,1]);
 IF YIV > H2 THEN {FLAGG=COUNTER;RETURN;}
 ELSE SAVE=VEC$MUL(SI,1);
end;
end;
/*IF CONTROL CRITERIA MET,REPORT VALUE AND RETURN TO MAIN PROCEDURE*/
collect;
end;
```

MCUSUM with training set of size 50

```
PROCEDURE(N,p,r,i,k2,h3,SH);
/*CALLING PROCEDURE FOR THE MCUSUM CHART
        N=MAX.RUN LENGTH VALUE EXPECTED
        P=NUMBER OF QUALITY CHARACTERISTICS
        R=MAGNITUDE OF CORRELATION
        I=NUMBER OF ITERATIONS
        K2,H3=CHART PARAMETERS
        SH=MAGNITUDE OF SHIFT*/
cou=lastrow("@simulate@m2res");
res=vec$alloc(5);x=vec$alloc(i);out=0;
/*COMPUTE RUN LENGTH DISTRIBUTION AND DETERMINE ARL AND MRL*/
do 13=1 to i;
         call mcr(n,p,r,k2,h3,flagg,SH);
         if flagg=0 then {flagg=n;type 13;out=out+1;}
         x[13]=flagg;
         collect;
end;
arl=vec$mean(x);mrl=vec$median(x);
res[1]=n;res[2]=p;res[3]=r;res[4]=i;res[5]=arl;res[6]=mrl;res[7]=out;
res[8]=k2;res[9]=h3;RES[10]=SH;
cou=cou+1;
call vec$v2trc(res,"@simulate@m2res",cou,false);
end;
```

Procedure MCR

```
procedure(n,p,rho,K2,h3,flagg,SH);
/* PROCEDURE FOR SIMULATING THE MCUSUM CHART WITH A TRAINING SET OF 50*/
y=mat$alloc(n,p);z=vec$alloc(n);m1=vec$alloc(50);tscale=mat$alloc(p,p);
e=mat$alloc(20,p);tnale=mat$alloc(n,p);e1=mat$alloc(50,p);vv=vec$alloc(p);
vtemp=vec$alloc(p);l=vec$alloc(p);d=vec$alloc(p);
sig=mat$alloc(p,p);z=vec$alloc(p);zz=mat$alloc(p,p);
xmat=mat$alloc(p,p);xmat[1,1]=0;
/*SETUP CORRELATION MATRIX*/
do i=1 to p;
   do j=1 to p;
      if i<>j then
         sig[i,j]=rho;
      if i=j then
         sig[i,j]=1.0;
   end;
end;
call $svd(sig,u,s,v);
s=vec$sqrt(s);
s=mv$mdiag(s);
matt=mat$mxm(s,mat$transpose(u));
tnale=mat$alloc(p,p);
/*GENERATE TRAINING SET */
do k=1 to 50;
   do i=1 to p;
      s=0;
      do j=1 to 12;
         s=s+ran();
      end;
      tnale[k,i]=s-6;
   end;
end;
e=mat$mxm(tnale,matt);
if tableexists("@simulate@cor") then del "@simulate@cor";
call $mvcorr(mat$m2tt(e),true,"@simulate@cor",0,false);
sigm=mat$t2m("@simulate@cor",true);
```

```
/*****************************************************************
random number generator*/
flagg=0;
counter=0;
SINV=SVD$INVERSE(SIGm);
save=0;
/* GENERATE DATA POINTS*/
do ii=1 to n by 50;
   do k=1 to 50;
      do i=1 to p;
         ss=0;
         do j=1 to 12;
            ss=ss+ran();
         end;
         tscale[k,i]=ss-6;
      end;
   end;
/*INTRODUCE SHIFT*/
   e1=mat$mxm(tscale,matt);
  if sh>0 then {
  do k=1 to 50;
      do i=1 to p;
       e1[k,i]=e1[k,i]+SH;
      end;
   end;}
   jj=0;
 do i=ii to (ii+49);
   jj=jj+1;
   counter=counter+1;
   x=mv$m2v(e1,false,jj);
   if save>0 then l[counter]=l[(counter-1)]+1;
   else l[counter]=1;
   call mv$v2mrc(xmat,x,counter,true);
   length=counter-l[counter]+1;
   do mm=1 to p;
      d[mm]=0;
      do nn=length to counter;
         d[mm]=d[mm]+xmat[mm,nn];
      end;
   end;
   INN=MV$V2M(d);
   INNT=MAT$TRANSPOSE(INN);
   C=MAT$MXM(MAT$MXM(INNT,SINV),INN);
   ci=sqrt(c[1,1])-(k2*l[counter]);
   if ci>0 then mc=ci;
   else mc=0;
   IF mc> H3 THEN {FLAGG=COUNTER;RETURN;}
   ELSE SAVE=mc;
 end;
end;
collect;
end;
```

Appendix C

Additional Simulation Results

This appendix contains additional simulation results that complement the results discussed in Chapter 4.

Table C1. Variability involved in the simulations of the tests of independence

Generalized Likelihood Ratio Test		Partitioning Method		t-Distribution Method		Confidence Intervals method	
Average	St.deviation	Average	St.deviation	Average	St.deviation	Average	St.deviation
0.09	0.036	1.42	0.208	0.020	0.011	0.59	0.122
2.96	0.260	24.6	0.519	8.64	0.331	59.4	0.650
31.9	0.750	58.1	0.576	86.3	0.453		
		99.9	0.0006	95.6	0.325		

Table C2. Simulated results of 2-sided CUSUM Chart ($k_c = 0.5$, $h = 5$) for $k \to \infty$

(a)
$\Delta = 0$, No correction

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	465	317	464	324	465	337
0.30	53	39	37	27	29	22
0.50	29	21	21	15	16.5	13
0.95	14.3	11	10.6	8	8.6	7

(b)
$\Delta = 0.5\sigma$, No correction

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	10.3	9	8.6	8	7.7	7
0.30	9.6	8	8.0	7	7.1	6
0.50	8.8	7	7.6	6	6.7	6
0.95	7.6	6	6.5	5	5.7	5

(c)
$\Delta = 1.5\sigma$, No correction

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	2.6	3	2.3	2	2.1	2
0.30	2.6	3	2.4	2	2.2	2
0.50	2.7	2	2.4	2	2.2	2
0.95	2.7	2	2.4	2	2.3	2

(d)
$\Delta = 0$, Correction with σ_X

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	462	322	474	328	475	330
0.50	469	326	472	316	473	333
0.95	471	327	483	338	474	327

(e)
$\Delta = 0.5\sigma$, Correction with σ_X

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	18.1	15	16.5	14	15.8	13
0.50	24	19	23	18	22	18
0.95	36	28	36	27	36	28

(f)
$\Delta = 1.5\sigma$, Correction with σ_X

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	3.6	3	3.5	3	3.4	3
0.50	4.2	4	4.2	4	4.0	4
0.95	5.6	5	5.6	5	5.6	5

Table C3. Simulated results of 2-sided CUSUM Chart ($k_c = 0.5$, $h = 5$) for $k = 20$

(a)
$\Delta = 0$, No correction

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	337	135	310	136	305	140
0.30	40	25	29	20	23	17
0.50	24	16	18.0	12	14.3	10
0.95	12.2	9	9.9	7	7.6	6

(b)
$\Delta = 0.5\sigma$, No correction

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	9.5	7	8.0	6	6.8	6
0.30	8.5	6	7.1	5	6.2	5
0.50	7.6	6	6.4	5	5.6	4
0.95	6.1	4	4.9	4	4.4	3

(c)
$\Delta = 1.5\sigma$, No correction

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	2.6	2	2.3	2	2.1	2
0.30	2.6	2	2.4	2	2.2	2
0.50	2.6	2	2.4	2	2.2	2
0.95	2.7	2	2.4	2	2.2	2

(d)
$\Delta = 0$, Correction with C

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	797	122	799	121	754	116
0.50	805	121	701	116	766	121
0.95	594	106	725	106	688	108

(e)
$\Delta = 0.5\sigma$, Correction with C

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	21	11	21	10	20	10
0.50	31	13	29	13	29	13
0.95	50	17	55	17	44	17

(f)
$\Delta = 1.5\sigma$, Correction with C

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	3.6	3	3.5	3	3.4	3
0.50	4.3	4	4.2	4	4.1	4
0.95	5.8	5	5.7	5	5.6	5

(g)
$\Delta = 0$, Correction with $s_{\bar{x}}$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	888	118	759	121	795	121
0.50	742	122	913	126	776	124
0.95	756	117	775	120	846	122

(h)
$\Delta = 0.5\sigma$, Correction with $s_{\bar{x}}$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	25	11	20	10	19	10
0.50	32	13	31	13	33	13
0.95	61	18	55	18	53	18

(i)
$\Delta = 1.5\sigma$, Correction with $s_{\bar{x}}$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	3.7	3	3.5	3	3.4	3
0.50	4.3	4	4.2	4	4.2	4
0.95	5.8	5	5.9	5	5.9	5

Table C4. Simulated results of 2-sided CUSUM Chart ($k_c = 0.5$, $h = 5$) for $k = 50$

(a)

$\Delta = 0$, No correction

	p = 4		p = 5		p = 6	
ρ_0	ARL	MRL	ARL	MRL	ARL	MRL
0.0	368	203	347	204	362	217
0.30	47	32	33	23	26	19
0.50	26	19	19.4	14	15.2	11
0.95	13	10	9.8	7	8.1	6

(b)

$\Delta = 0.5\sigma$, No correction

	p = 4		p = 5		p = 6	
ρ_0	ARL	MRL	ARL	MRL	ARL	MRL
0.0	11.0	9	9.1	8	7.9	7
0.30	10.0	8	8.5	7	7.4	6
0.50	9.4	7	7.9	6	6.8	5
0.95	7.9	6	6.5	5	5.6	5

(c)

$\Delta = 1.5\sigma$, No correction

	p = 4		p = 5		p = 6	
ρ_0	ARL	MRL	ARL	MRL	ARL	MRL
0.0	2.5	2	2.3	2	2.1	2
0.30	2.6	2	2.4	2	2.2	2
0.50	2.7	2	2.4	2	2.2	2
0.95	2.7	2	2.4	2	2.3	2

(d)

$\Delta = 0$, Correction with C

	p = 4		p = 5		p = 6	
ρ_0	ARL	MRL	ARL	MRL	ARL	MRL
0.30	532	188	504	190	478	188
0.50	484	182	532	193	506	187
0.95	486	183	477	182	473	186

(e)

$\Delta = 0.5\sigma$, Correction with C

	p = 4		p = 5		p = 6	
ρ_0	ARL	MRL	ARL	MRL	ARL	MRL
0.30	21	15	19	14	18	13
0.50	31	18	28	18	28	18
0.95	52	26	53	26	52	26

(f)

$\Delta = 1.5\sigma$, Correction with C

	p = 4		p = 5		p = 6	
ρ_0	ARL	MRL	ARL	MRL	ARL	MRL
0.30	3.6	3	3.5	3	3.4	3
0.50	4.3	4	4.2	4	4.1	4
0.95	5.6	5	5.6	5	5.5	5

(g)

$\Delta = 0$, Correction with $s_{\bar{x}}$

	p = 4		p = 5		p = 6	
ρ_0	ARL	MRL	ARL	MRL	ARL	MRL
0.30	542	194	549	203	492	190
0.50	533	201	529	201	509	198
0.95	532	194	546	192	515	199

(h)

$\Delta = 0.5\sigma$, Correction with $s_{\bar{x}}$

	p = 4		p = 5		p = 6	
ρ_0	ARL	MRL	ARL	MRL	ARL	MRL
0.30	22	15	20	14	18.6	13
0.50	30	15	29	17	28	17
0.95	54	28	54	27	59	28

(i)

$\Delta = 1.5\sigma$, Correction with $s_{\bar{x}}$

	p = 4		p = 5		p = 6	
ρ_0	ARL	MRL	ARL	MRL	ARL	MRL
0.30	3.7	3	3.5	3	3.4	3
0.50	4.3	4	4.2	4	4.2	4
0.95	5.7	5	5.7	5	5.7	5

Table C5. Simulated results of the EWMA chart ($\lambda = 0.05$, L $= 2.399$) for k -> ∞

(a)

$\Delta = 0$, no correction

ρ_0	p = 4		p = 5		p = 6	
	ARL	MRL	ARL	MRL	ARL	MRL
0.0	270	185	274	191	273	177
0.30	53	30	37	18	29	13
0.50	30	13	20	8	14.9	6
0.95	12.7	5	8.9	3	7.0	3

(b)

$\Delta = 0.5\sigma$, no correction

ρ_0	p = 4		p = 5		p = 6	
	ARL	MRL	ARL	MRL	ARL	MRL
0.0	6.2	5	5.1	4	4.5	4
0.30	5.9	4	5.1	4	4.5	3
0.50	5.8	4	4.8	3	4.1	3
0.95	5.0	3	4.2	2	3.8	2

(c)

$\Delta = 1.5\sigma$, no correction

ρ_0	p = 4		p = 5		p = 6	
	ARL	MRL	ARL	MRL	ARL	MRL
0.0	1.3	1	1.2	1	1.1	1
0.30	1.4	1	1.3	1	1.2	1
0.50	1.5	1	1.4	1	1.3	1
0.95	1.6	1	1.5	1	1.4	1

(d)

$\Delta = 0$, correction with σ_X

ρ_0	p = 4		p = 5		p = 6	
	ARL	MRL	ARL	MRL	ARL	MRL
0.30	274	186	274	184	278	191
0.50	265	178	273	183	274	185
0.95	269	176	270	185	276	185

(e)

$\Delta = 0.5\sigma$, correction with σ_X

ρ_0	p = 4		p = 5		p = 6	
	ARL	MRL	ARL	MRL	ARL	MRL
0.30	10.3	9	9.8	8	9.4	8
0.50	12.9	11	12.9	11	12.3	10
0.95	18.9	15	18.7	16	18.7	15

(f)

$\Delta = 1.5\sigma$, correction with σ_X

ρ_0	p = 4		p = 5		p = 6	
	ARL	MRL	ARL	MRL	ARL	MRL
0.30	1.9	2	1.8	2	1.8	2
0.50	2.3	2	2.2	2	2.2	2
0.95	1.9	2	3.1	3	3.2	3

Table C6. Simulated results of the EWMA chart ($\lambda = 0.05$, $L = 2.399$) for $k = 20$

(a)

$\Delta = 0$, no correction

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	309	188	292	172	283	172
0.30	55	27	39	18	30	12
0.50	30	12	20	7	15.2	5
0.95	13.2	4	9.1	3	6.9	3

(b)

$\Delta = 0.5\sigma$, no correction

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	6.1	5	5.2	4	4.4	4
0.30	5.9	4	4.9	4	4.4	3
0.50	5.6	4	4.7	3	4.1	3
0.95	4.9	3	4.2	2	3.6	2

(c)

$\Delta = 1.5\sigma$, no correction

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	1.3	1	1.2	1	1.1	1
0.30	1.4	1	1.3	1	1.2	1
0.50	1.5	1	1.4	1	1.3	1
0.95	1.6	1	1.5	1	1.4	1

(d)

$\Delta = 0$, correction with C

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	425	162	434	160	385	145
0.50	440	163	413	155	390	150
0.95	429	162	445	162	392	154

(e)

$\Delta = 0.5\sigma$, correction with C

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	10.5	8	9.8	8	9.5	8
0.50	12.9	10	12.7	10	12.3	10
0.95	18.5	14	18.5	14	17.6	13

(f)

$\Delta = 1.5\sigma$, correction with C

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	1.9	2	1.7	2	1.7	2
0.50	2.3	2	2.2	2	2.2	2
0.95	3.2	3	3.0	3	3.1	3

(g)

$\Delta = 0$, correction with $s_{\bar{x}}$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	432	164	436	156	425	168
0.50	430	164	439	156	446	156
0.95	413	155	438	159	439	159

(h)

$\Delta = 0.5\sigma$, correction with $s_{\bar{x}}$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	10.5	9	10.0	8	9.3	7
0.50	13.2	10	12.5	10	12.3	10
0.95	18.7	15	19.1	15	18.6	14

(i)

$\Delta = 1.5\sigma$, correction with $s_{\bar{x}}$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	1.9	2	1.7	2	1.7	2
0.50	2.3	2	2.2	2	2.2	2
0.95	3.1	3	3.1	3	3.2	3

Table C7. Simulated results of the EWMA chart ($\lambda = 0.05$, $L = 2.399$) for $k = 50$

(a)

$\Delta = 0$, no correction

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	276	177	282	181	280	178
0.30	53	53	40	19	30	13
0.50	30	13	21.2	8	15.4	6
0.95	13.2	4	9.1	3	6.8	3

(b)

$\Delta = 0.5\sigma$, no correction

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	6.1	5	5.2	4	4.5	4
0.30	5.9	4	5.0	4	4.3	3
0.50	5.6	4	4.8	3	4.1	3
0.95	5.0	3	4.2	2	3.6	2

(c)

$\Delta = 1.5\sigma$, no correction

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.0	1.3	1	1.2	1	1.1	1
0.30	1.4	1	1.3	1	1.2	1
0.50	1.5	1	1.4	1	1.3	1
0.95	1.6	1	1.5	1	1.4	1

(d)

$\Delta = 0$, correction with C

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	314	169	322	174	306	168
0.50	319	164	315	169	317	168
0.95	314	167	307	163	288	156

(e)

$\Delta = 0.5\sigma$, correction with C

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	10.4	9	9.9	8	9.2	8
0.50	13.1	11	12.7	10	12.1	10
0.95	18.5	15	18.4	15	18.7	15

(f)

$\Delta = 1.5\sigma$, correction with C

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	1.9	2	1.8	2	1.7	2
0.50	2.3	2	2.3	2	2.2	2
0.95	3.1	3	3.1	3	3.1	3

(g)

$\Delta = 0$, correction with $s_{\bar{x}}$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	323	176	326	174	326	182
0.50	327	179	329	171	321	176
0.95	313	167	310	168	320	174

(h)

$\Delta = 0.5\sigma$, correction with $s_{\bar{x}}$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	10.7	9	9.8	8	9.4	8
0.50	13.1	10	12.6	10	12.5	10
0.95	18.8	15	18.1	14	18.4	15

(i)

$\Delta = 1.5\sigma$, correction with $s_{\bar{x}}$

ρ_0	p = 4 ARL	p = 4 MRL	p = 5 ARL	p = 5 MRL	p = 6 ARL	p = 6 MRL
0.30	1.9	2	1.8	2	1.7	2
0.50	2.3	2	2.3	2	2.2	2
0.95	3.1	3	3.2	3	3.1	3

Table C8. Univariate charts comparison based on median run lengths for p = 5.

			CUSUM (k_c = 0.5, h = 3.32)						EWMA (λ = 0.06, L = 2.00)						X					
		method of	shift=0		shift=0.5σ		shift=1.5σ		shift=0		shift=0.5σ		shift=1.5σ		shift=0		shift=0.5σ		shift=1.5σ	
ρ_0	training set	correction	ARL	MRL	ARL	MRL	ARL	MRL	ARL	MRL	ARL	MRL	ARL	MRL	ARL	MRL	ARL	MRL	ARL	MRL
0	none	none	83	58	6.1	5	1.7	2	90	58	3.9	3	1.1	1	83	57	8.7	8	1.5	1
0	20	none	64	38	6.6	5	1.7	2	92	53	3.9	3	1.1	1	72	48	9.6	8	1.5	1
0	50	none	73	46	6.2	5	1.7	2	92	58	3.9	3	1.1	1	82	57	9.2	8	1.5	1
0.95	none	none	5.4	4	4.2	3	1.8	2	4.8	2	3.3	2	1.5	1	5.0	4	3.9	3	1.7	1
0.95	none	ρ_0	83	59	19.5	15	3.9	4	93	55	12.0	9	2.6	2	101	72	30	22	5.3	5
0.95	20	none	5.2	4	4.0	3	1.8	2	4.9	2	3.4	2	1.4	1	4.8	3	3.8	3	1.7	1
0.95	20	C	84	32	28	13	4.0	3	113	47	12.6	10	2.5	2	64	39	32	18	5.3	5
0.95	20	*X	103	35.5	34	14	4.1	4	105	41	12.0	9	2.5	2	65	40	33	19	5.5	5
0.95	50	none	5.4	4	4.1	3	1.9	2	5.0	2	3.0	2	1.6	1	4.7	3	3.8	3	1.7	1
0.95	50	C	79	43	22	14	3.9	4	96	55	12.8	10	2.5	2	77	51	32	18	5.3	4
0.95	50	*X	82	46	23	14	4.0	4	94	50	12.5	10	2.6	2	80	54	32	21	5.4	5

Table C9. Multivariate charts comparison based on median run lengths for p = 5.

		MEWMA ($r = 0.06$, $h_4 = 11.55$)						MC1 ($k_1 = 0.5$, $h_2 = 7.75$)						MCUSUM ($k_2 = 0.5$, $h_3 = 5.67$)					
		shift=0		shift=0.5σ		shift=1.5σ		shift=0		shift=0.5σ		shift=1.5σ		shift=0		shift=0.5σ		shift=1.5σ	
ρ_0	training set	ARL	MRL	ARL	MRL	ARL	MRL	ARL	MRL	ARL	MRL	ARL	MRL	ARL	MRL	ARL	MRL	ARL	MRL
0	none	89	57	5.9	5	1.3	1	80	58	9.4	9	3.0	3	83	58	7.6	7	2.3	2
0	20	31	14	4.5	3	1.2	1	36	26	8.2	8	2.8	3	35	23	6.3	5	2.2	2
0	50	57	34	5.4	4	1.3	1	56	42	8.9	8	2.9	3	58	39	7.1	6	2.3	2
0.95	none	89	57	19.0	15	3.8	3	80	58	24	20	6.7	6	82	57	25	13	5.1	5
0.95	20	58	13	12.9	7	3.1	3	61	27	18.5	15	6.2	6	51	22	18.3	12	4.7	4
0.95	50	75	33	16.5	12	3.5	3	72	42	21	18	6.5	6	69	39	23	16	5.0	5

Appendix D

SAMPLE CALCULATIONS

A sample calculation of each of the tests of independence as well as the control charts is shown below using the training set of example 2 in section 5.2. The data for the training set is listed in the following table.

Table D1. Training set data for example 2.

subgroup number	C2_1 X_1	C2_2 X_2	C2_3 X_3	C2_4 X_4	C2_5 X_5	MEAN X
1	56	55	58	55	53	55.4
2	56	53	55	53	50	53.4
3	57	55	60	56	53	56.2
4	55	54	56	53	51	53.8
5	60	62	61	46	55	56.8
6	63	61	65	61	59	61.8
7	51	52	51	49	49	50.4
8	55	54	60	55	51	55.0
9	57	54	57	53	52	54.6
10	63	59	63	63	61	61.8
11	57	55	59	56	53	56.0
12	52	50	56	47	48	50.6
13	60	61	60	59	60	60.0
14	61	57	63	61	55	59.4
15	62	61	65	61	57	61.2
16	61	59	69	64	58	62.2
17	57	54	58	55	51	55.0
18	58	58	63	59	56	58.8
19	59	56	60	58	56	57.8
20	62	60	64	60	58	60.8

Thus, $p = 5$ and $k = 20$.

The sample covariance matrix, **S**, is

$$\mathbf{S} = \begin{bmatrix} 11.77 & 10.47 & 12.24 & 13.13 & 11.81 \\ 10.47 & 11.84 & 11.13 & 9.63 & 11.57 \\ 12.24 & 11.13 & 17.39 & 16.39 & 12.53 \\ 13.13 & 9.63 & 16.39 & 25.53 & 14.83 \\ 11.81 & 11.57 & 12.53 & 14.83 & 14.22 \end{bmatrix}.$$

Consequently, the sample correlation matrix is calculated as

$$\mathbf{R} = \begin{bmatrix} 1.00 & 0.88 & 0.85 & 0.75 & 0.91 \\ 0.88 & 1.00 & 0.77 & 0.55 & 0.89 \\ 0.85 & 0.77 & 1.00 & 0.77 & 0.79 \\ 0.75 & 0.55 & 0.77 & 1.00 & 0.77 \\ 0.91 & 0.89 & 0.79 & 0.77 & 1.00 \end{bmatrix}.$$

TESTS OF INDEPENDENCE

H_0: $\boldsymbol{\rho} = \mathbf{I}$ $\qquad$ H_1: $\boldsymbol{\rho} \neq \mathbf{I}$ $\qquad$ $n = k = 20$

Generalized Likelihood Ratio Test

$\Lambda = |\mathbf{R}|^{n/2} = |0.256|^{10} = 1.20\text{x}10^{-6}$

$-2\ln\Lambda = 27.25$

$\chi^2_{(5)(4)/2,0.05} = 18.31$

Since $-2\ln\Lambda > 18.31$, the null hypothesis that X_1, X_2, X_3, X_4 and X_5 are independent can be rejected.

Partitioning Method

$q = p = 5$ $\qquad$ $p_i = 1$

$\mathbf{R}_{ii} = 1, \ i = 1, 2, \ldots, 5$

$$\Lambda = \frac{|\mathbf{R}|^{n/2}}{|\mathbf{R}_{11}|\,|\mathbf{R}_{22}|\,|\mathbf{R}_{33}|\,|\mathbf{R}_{44}|\,|\mathbf{R}_{55}|} = \frac{|0.256|^{10}}{(1)(1)(1)(1)(1)} = (0.265)^{10}$$

$$m = (20-1.5) - \frac{(125-5)}{3(25-5)} = 16.5$$

$$\nu = \frac{25-5}{2} = 10$$

$-m\ln(\Lambda)^{2/n} = -16.5\ln(0.256) = 22.4$

$\chi^2_{10,0.05} = 18.31$

Since $-m\ln v^{2/n} > \chi^2_{v,\alpha}$, the null hypothesis that the variables are independent is rejected.

Pairwise Comparisons Using the t Distribution

The highest and lowest off-diagonal correlation coefficients are 0.91 and 0.55 respectively. It is sufficient to test these two values for independence to confirm the existence of within subgroup correlation.

$$T_{max} = \frac{(20-5)^{0.5}0.91}{(1-0.91^2)^{0.5}} = 9.31$$

$$T_{min} = \frac{(20-5)^{0.5}0.55}{(1-0.55^2)^{0.5}} = 2.79$$

$$t_{(20-2),0.05} = 2.101$$

Since both T_{max} and T_{min} are greater than 2.101, the null hypothesis that the measurements are independent can be rejected.

Confidence Intervals Method

Fisher's transformations for r_{max} and r_{min} are

$$h(r_{max}) = \left(\frac{1}{2}\right)\ln\left[\frac{1+0.91}{1-0.91}\right] = 1.52$$

$$h(r_{min}) = \left(\frac{1}{2}\right)\ln\left[\frac{1+0.55}{1-0.55}\right] = 0.61.$$

The corresponding 95% confidence intervals are [0.78,0.99] and [0.13,0.79] respectively.

Both of these intervals do not contain 0. Thus, the null hypothesis that X_1, X_2, X_3, X_4 and X_5 are independent can be rejected at the level of significance 0.05.

PROPOSED SPC METHODS

From the column of means in Table D1, $\overline{\overline{X}} = 57.05$ and $s_{\overline{X}} = 3.62$. The standard deviation, s, of the X_1, X_2, X_3, X_4 and X_5 measurements is 4.405. The average of the off-diagonal elements of **R** is $\overline{r} = 0.79$.

Modified Univariate Charts

The correction factor is calculated as

$$C = [1+(5-1)0.79/5]^{0.5} = 0.912.$$

The use of C can be justified only if the variances of X_1, X_2, X_3, X_4 and X_5 are equal. An F-test is performed on the pairs of variables and the null hypothesis that the variances are equal is not rejected. The algorithm of this test is a standard statistical procedure. For illustration, this test is performed for X_1 and X_2 whose variances, s_1^2 and s_2^2, are 11.77 and 11.84 respectively with degrees of freedom (20-1) = 19. If $\sigma_2^2=\sigma_1^2$, the ratio $\frac{s_1^2/v_1}{s_2^2/v_2}$ has an F distribution with v_1, v_2 degrees of freedom. $\frac{s_1^2/v_1}{s_2^2/v_2} = 0.994$ and $F_{19,19,0.05} = 2.17 > 0.994$. Therefore, the null hypothesis that $\sigma_2^2 = \sigma_1^2$ cannot be rejected. Similarly, the remaining variables are proven to have equal variances. Thus, the use of C as a correction factor for the modified univariate charts is justified.

Modified $\bar{X}$ chart control limits:	$\bar{\bar{X}} \pm 3s_{\bar{X}}$ or $\bar{\bar{X}} \pm 3Cs$ 57.05 ± 10.87 or 57.05 ± 12.05.
Modified CUSUM chart:	$z_i = \dfrac{\bar{X}_i - 57.05}{3.62}$ or $\dfrac{\bar{X}_i - 57.05}{4.01}$ $S_{Hi} = \max[0,(z_i-0.5)+S_{Hi-1}]$ $S_{Li} = \max[0,(-z_i-0.5)+S_{Li-1}]$ Compare S_{Hi} and S_{Li} versus $h = 4$.
Modified EWMA chart:	Choose $\lambda = 0.05$ and $L = 2.399$ $Z_i = 0.05\bar{X}_i + 0.95Z_{i-1}$ $Var[Z_i] = (0.05/1.95)[1-0.95^{2i}]3.62^2$ $= 0.336[1-0.95^{2i}]$ Control limits: $57.05 \pm 1.39[1-0.95^{2i}]^{0.5}$

Principal Components Approach

The eigenvalues of the covariance matrix, **S**, are: 4.204, 0.448, 0.197, 0.076 and 0.037. The percentages of the total variability explained by the associated principal components are 84.09%, 9.39%, 4.14%, 1.59% and 0.79% respectively (for instance, for the first principal component explains (4.204x100%)/(4.204+0.448+0.197+0.076+0.037) = 84.09%). Therefore, only the first principal component should be retained since it accounts for the largest percentage of the variability.
The normalized eigenvector associated with the 4.204 eigenvalue is e_i^T = [0.47, 0.44, 0.45, 0.41, 0.47] which can be scaled to become e_i^T = [0.20, 0.19, 0.19, 0.17, 0.20]. The first principal component is $Y_1 = 0.20X_1 + 0.19X_2 + 0.19X_3 + 0.17X_4 + 0.20X_5$. The value of Y_1 is very similar to $\bar{X} = (X_1 + X_2 + X_3 + X_4 + X_5)/5 = 0.20X_1 + 0.20X_2 + 0.20X_3 + 0.20X_4 + 0.20X_5$. Therefore, the use of Y_1 has no advantage over $\bar{X}$ and the principal components approach is analogous to the modified univariate charts.

Multivariate Control Charts

From the 5 columns of measurements in Table D1, the vector of averages, $\bar{\mathbf{X}}$, is $\bar{\mathbf{X}}^T$ = [56.5, 56.5, 60.8, 56.2, 54.3]. The $\mathbf{X}_i$ vector consists of the measurements of one subgroup, $\mathbf{X}_i^T = [X_{i1}, X_{i2}, X_{i3}, X_{i4}, X_{i5}]$. For example, for the first subgroup, $\mathbf{X}_1^T$ = [56, 55, 58, 55, 53].

Hotelling's T^2 Chart
Using Jackson's approach(1956), the statistic plotted is $T_i^2 = (\mathbf{X}_i - \bar{\mathbf{X}})^T\mathbf{S}^{-1}(\mathbf{X}_i - \bar{\mathbf{X}})$. For this case,

$$T_i^2 = \begin{bmatrix} X_{i1} - 56.5 \\ X_{i2} - 56.5 \\ X_{i3} - 60.8 \\ X_{i4} - 56.2 \\ X_{i5} - 54.3 \end{bmatrix}^T \begin{bmatrix} 11.77 & 10.47 & 12.24 & 13.13 & 11.81 \\ 10.47 & 11.84 & 11.13 & 9.63 & 11.57 \\ 12.24 & 11.13 & 17.39 & 16.38 & 12.53 \\ 13.13 & 9.63 & 16.39 & 25.53 & 14.83 \\ 11.81 & 11.57 & 12.53 & 14.83 & 14.22 \end{bmatrix}^{-1} \begin{bmatrix} X_{i1} - 56.5 \\ X_{i2} - 56.5 \\ X_{i3} - 60.8 \\ X_{i4} - 56.2 \\ X_{i5} - 54.3 \end{bmatrix}$$

The upper control limit is: $UCL = \frac{5(20+1)(20-1)}{20^2-(20)(5)} F_{5,20-5,0.05} = 19.28.$

MC1

For $k_1 = 0.5$ and $h_2 = 9.5$, the MC1 procedure involves the following calculations. Compute the statistic $C_i = [(\mathbf{S}_{i-1} - \mathbf{X}_i)^T S^{-1}(\mathbf{S}_{i-1} - \mathbf{X}_i)]^{0.5}$ where $\mathbf{S}_i$ is a 5x1 vector such that

$\mathbf{S}_i = \mathbf{0}$ if $C_i \leq 0.5$ or

$\mathbf{S}_i = (\mathbf{S}_{i-1} + \mathbf{X}_i)(1-0.5/C_i)$ if $C_i > 0.5$.

An out-of-control signal is given if $Y_i = [\mathbf{S}_i^T S^{-1} \mathbf{S}_i]^{0.5} > 9.5$.

MCUSUM

For $k_2 = 0.5$ and $h_3 = 6.8$, the MCUSUM procedure uses the following quantities:

$$\mathbf{D}_i = \sum_{j=i-l_i+1}^{i} \mathbf{X}_j$$

and $MC_i = \max[0, (\mathbf{D}_i^T S^{-1} \mathbf{D}_i)^{0.5} - 0.5l_i]$ where $l_i = l_{i-1} + 1$ if $MC_i > 0$ and $l_i = 1$ otherwise. MC_i is plotted versus time or order of production and the chart gives an alarm whenever $MC_i > 6.8$

MEWMA

For $r = 0.06$ and $h_4 = 13.5$, the MEWMA procedure involves the following calculations: $\mathbf{Z}_i = 0.06\mathbf{X}_i + 0.94\mathbf{Z}_{i-1}$. The covariance of $\mathbf{Z}_i$ is $\mathbf{S}_{zi} = 0.031[1-0.94^{2i}]\mathbf{S}$. The statistic $T_i^2 = \mathbf{Z}_i^T \mathbf{S}_{zi}^{-1} \mathbf{Z}_i$ is plotted on the chart which gives an out-of-control signal if $T_i^2 > 13.5$.