NONPARAMETRIC MULTIVARIATE STATISTICAL PROCESS CONTROL USING PRINCIPAL COMPONENT ANALYSIS AND SIMPLICIAL DEPTH

by

LUIS A. BELTRAN
B.A. Mathematics, St. Thomas University, 1989
M.A. Mathematics, University of Miami, 1993

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Industrial Engineering and Management Systems
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Spring Term
2006

Major Professor: Dr. Linda Malone

ABSTRACT

Although there has been progress in the area of Multivariate Statistical Process Control (MSPC), there are numerous limitations as well as unanswered questions with the current techniques. MSPC charts plotting Hotelling's T^2 require the normality assumption for the joint distribution among the process variables, which is not feasible in many industrial settings, hence the motivation to investigate nonparametric techniques for multivariate data in quality control. In this research, the goal will be to create a systematic distribution-free approach by extending current developments and focusing on the dimensionality reduction using Principal Component Analysis. The proposed technique is different from current approaches given that it creates a nonparametric control chart using robust simplicial depth ranks of the first and last set of principal components to improve signal detection in multivariate quality control with no distributional assumptions. The proposed technique has the advantages of ease of use and robustness in MSPC for monitoring variability and correlation shifts. By making the approach simple to use in an industrial setting, the probability of adoption is enhanced. Improved MSPC can result in a cost savings and improved quality.

I dedicate this dissertation to my family, both living and deceased. To my late father, Luis, you always told me how proud you were of me and my sister. To my niece and nephew, Elizabeth and Julian Jr., thanks for understanding when there were times that I could not be with you guys because of school. To my brother-in law Julian, you have always been to me my older brother. To my in-laws, Francisco and Haydee, thanks for your encouragement and meals that provided brain food. To my aunt, Eugenia, you have always been more than an aunt. You are a special lady in my life who has taught me by example. To my mother, Conchita, I have never forgotten when Papi died, you did not allow yourself or us to fall apart. You told Monica and I two things: Do not stop believing in God and Do not stop studying. To my sister Dr. Monica DeZulueta, you have always been my inspiration. Since we were kids, I have always looked up to you. Your completion of your doctoral degree last year, once again gave me that inspiration and motivation to finish. To my late grandmother Isabel, when I was a kid and you would help me with my homework, you always said that when I finished my studies, you would be walking alongside me. Last but certainly not least, to my wife Dulce, you have been cheering me on all the way. I know that we have sacrificed a lot during the last few years, but you would keep reminding me of the prize at the end. You would tell me that I was the little engine that could, but that description fits you even better. Dulce, I love, honor and truly cherish you. This dissertation would not be possible without all the love and continued support that I have always received from you and the rest of the family. I love all of you with all my heart, and I thank God for giving me such a loving family.

ACKNOWLEDGMENTS

I wish to thank Dr. Linda Malone, Advisor and Committee Chair, Dr. Ahmad Elshennawy, Dr. Lucy Morse, Dr. Luis Rabelo and Dr. Nizam Uddin for serving on my dissertation committee. To Dr. Elshennawy, you have always demonstrated a deep appreciation for the concepts and applications of Quality, which gave me the inspiration to conduct research in this area of study. To Dr. Morse, Dr. Rabelo and Dr. Uddin, I appreciate all of your time, assistance and suggestions throughout the entire dissertation process. To my advisor, Dr. Linda Malone, the word thanks cannot begin to describe my gratitude for everything that you have done for me. Linda, I always reminded you that you got me into the program and got me through it all the way to the end. Since day one when we met to discuss my research, you used the word "exciting" to describe what I was researching which set the stage for the rest of our discussions. The weekly discussions that we had were invaluable. Through those discussions, you would prepare me on a weekly basis for what was to come. I also remember one week when you told me, "I get more excited every time you send me something." You would even take time from your personal schedule to meet with me without hesitation and complete enthusiasm. You are a true educator and a wonderful person. I will never forget all that you have done for me. My sincere gratitude to all of you and also to the IEMS staff.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

SPC – Statistical Process Control

MSPC – Multivariate Statistical Process Control

LCL – Lower Control Limit

CL – Center Line

UCL – Upper Control Limit

PCA – Principal Component Analysis

PC – Principal Component

SD – Simplicial Depth

HDS – Historical Data Set

SPE – Squared Prediction Error

λ – Eigenvalue

α – Alpha

β – Beta

H_0 – Null Hypothesis:

H_1 – Alternative hypothesis

T^2 – Hotelling's Test Statistic for Multivariate Normality *r*-chart – Liu's Nonparametric MSPC r chart

Q-chart – Liu's Nonparametric MSPC *Q*-chart

χ^2 – Chi-Squared Test Statistic

D^2 or Δ^2 – Mahalanobis Distance

$\boldsymbol{\mu}$ – Population mean vector

$\bar{\mathbf{x}}$ – Sample mean vector

$\boldsymbol{\Sigma}$ – Population variance covariance matrix

$\mathbf{S}$ – Sample variance covariance matrix

TQC – Total Quality Control

TQM – Total Quality Management

PDCA – Plan-Do-Check-Act

ARL – Average Run Length

EWMA – Exponentially Weighted Moving Average

CUSUM – Cumulative Sum

UCF – University of Central Florida

IEMS – Industrial Engineering and Management Systems

CHAPTER ONE: INTRODUCTION

1.1 Introduction

One of the tools that has been widely recognized in the Quality Control area is Shewhart's Statistical Process Control Chart. During the past eighty years, these control charts have been utilized and gained much popularity largely because of their simplicity and effectiveness. Unfortunately, a significant limitation to these traditional control charts has been that these charts monitored processes that were determined by univariate data. Hence, processes that are determined by multivariate data may not be monitored effectively by these traditional control charts. Additionally, with the multivariate structure, the variables may be correlated. These correlations may exist between the variables, known as inter-correlations, and/or within each variable over time, known as autocorrelation. Given these additional considerations, there is a need for Multivariate Statistical Process Control (MSPC) techniques. Currently, the MSPC technique that is applied and referenced significantly in the multivariate quality control literature is Hotelling's T^2 statistic, developed by Harold Hotelling in the 1940's. We can utilize current statistical software packages with the multivariate quality control option to construct MSPC control charts by computing and plotting T^2 values along with an Upper Control Limit (UCL) obtained from the critical value. Details of this multivariate technique are provided in Chapter Two of this dissertation along with an overview of historical developments and limitations in multivariate quality control. The Literature Review in Chapter Two will provide a historical background of Quality Control and SPC in order to give a perspective into the attempts by past researchers to address MSPC and why these attempts have generated more issues that need to be investigated. Although the research in MSPC began in the 1940's with Hotelling's developments, much of the literature review in this paper covers publications from the late 1980's to present since extensive research in MSPC is relatively new. The techniques discussed have included nonparametric MSPC. Although there are numerous papers in theoretical statistical multivariate analysis, those not cited specifically in this dissertation are not relevant to the scope of this research. The authors from the publications which are not directly referenced in this dissertation have been included in the APPENDIX section.

1.2 Purpose

Based on the previous research and current developments, there has been progress in the area of MSPC. Unfortunately, there are numerous limitations as well as unanswered questions with the current techniques. Hotelling's approach requires the normality assumption for the joint distribution among the process variables, which is not feasible in many industrial settings. The approaches by researchers from the 1950's through the 1980's have introduced nonparametric methods in multivariate analysis, but these techniques are either limited to bivariate data or are lacking the affine variant property. In the 1990's researchers began to explore new nonparametric techniques to monitor multivariate processes more effectively. Many of these techniques are theoretical in nature and have not been fully applied to multivariate quality control.

The motivation to investigate nonparametric techniques in multivariate quality control is that fewer restrictive assumptions of the process data are imposed. In the quality control literature, Coleman (1997) states that the multivariate normality assumption for industrial data is unrealistic and that "distribution-free multivariate SPC is what we need." The nonparametric approach also tends to be less sensitive to outliers, hence more robust and prone to fewer false alarms. Plotting a univariate chart (parametric and nonparametric) for each variable from multivariate data may not necessarily produce results as accurate as monitoring the multivariate distribution as a whole. In multivariate processes, the multivariate decomposition into separate p

univariate processes generates a loss of power in those tests since this decomposition does not consider correlations between these variables. (Fuchs and Kennet, 1998) Since industries, such as the food and chemical industries among others, inherently deal with numerous variables which are highly correlated, there is a tremendous need for improvement in the area of MSPC to monitor processes more efficiently in the industrial setting. (Elsayed, 2000) Traditional techniques also tend to generate higher false alarms in the presence of autocorrelation which is common in industry. In quality control, the demand for a new or "better" approach in industry is satisfied by establishing a technique that is easy to use and capable of detecting shifts in a process without producing high levels of false alarm rates. (Stoumbos, et.al., 2000)

Woodall (1999) has indicated there is a need for research and development of nonparametric techniques that will monitor multivariate data. Due to the fact that univariate control charts for most of the twentieth century had been considered sufficient, the literature in the multivariate quality control area is not as abundant as the univariate quality control literature. However, as the literature review in this paper will demonstrate, more recent publications, many from the 1990's to the present, have stressed a strong need to develop a new approach in MSPC. The research in this paper will explore nonparametric techniques including simplicial depths and Principal Component Analysis (PCA) to develop robust methods in order to lower false alarm rates while avoiding an increase in missed alarms in MSPC. Since increases in false and missed alarms have a negative effect on quality with a serious economic impact, it is imperative to develop techniques aimed at minimizing false and missed alarms.

Applying the philosophy of Genichi Taguchi, when a process experiences shifts in the target value and increases in the variation there would be a "loss to society," which as he described is a loss that would be incurred by someone – namely the person who owns the process and anyone who would be affected by set process. (Kolarik, 1995) Losses in today's struggling economic situation could even be catastrophic. In the current economic and political situation that we are facing, it is imperative for improvement to be sought in industrial settings where correlations among variables and autocorrelations are possible.

This research will address the needs in MSPC by analyzing the limitations of certain historical discoveries and by utilizing robust nonparametric techniques which can be applied to industrial settings. Given the lack of a unified approach in non-normal multivariate quality, this research proposes a new systematic and efficient approach that will address non-normality in MSPC and lower false alarm rates. This dissertation will be divided into the following chapters: Chapter Two – Literature Review [providing historical contributions and limitations in Quality Control and Multivariate Analysis], Chapter Three – Methodology [applying nonparametric techniques for MSPC], Chapter Four – Findings [how did the nonparametric techniques improve the false alarm rates in MSPC] and Chapter Five – Conclusion [expanding the findings to future research in this area].

CHAPTER TWO: LITERATURE REVIEW

2.1 Background on Quality Control

During the first to the middle part of the 20th century, the United States became the leader in mass production. Unfortunately, with a stronger focus on mass production, high quality standards were de-emphasized. This was also the era when Japan was trying to recover from the second world war and began to develop goods that could be exported to eventually rebuild its economy. At the time, Japan had the reputation for developing cheaper goods, therefore the Japanese knew that in order to successfully compete they would have to use a new approach that would place them in a competitive position. It was during this era that terms such as Quality and Quality Control were being circulated amongst businesses. While these Quality concepts were not considered essential in the United States given the strong lead it had at the time in the domestic and world economy, Japan on the other hand needed desperately to improve its image and establish a good reputation. Using developments in Quality Control, the Japanese began to transform the ideologies in their business world, consequently establishing themselves as strong competitors and eventually overtaking the market.

In order to discuss quality we must first define it. As referenced by Kolarik (1995) the following represent various definitions of quality:

- "Characteristic that belongs to a thing's basic nature" *Webster's dictionary*
- Two aspects of quality: objective and subjective. The first defines quality "as an objective reality independent of the existence of man." The second defines quality "with what we think, feel, or sense…this subjective side of quality is closely linked to value." *Walter A. Shewhart*
- "A fitness for use" *Joseph M. Juran*
- "The conformance to requirements" *Philip B. Crosby*
- An aim "at the needs of the consumer, rising and future" *W. Edwards Deming*
- "The total composite product and service characteristics of marketing, engineering, manufacture and maintenance through which the product and service in use will meet the expectations of the customer." *Armand V. Feigenbaum*
- "The loss a product causes to society after being shipped, other than any losses caused by its intrinsic functions" *Genichi Taguchi*
- "The totality of the features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs" *ISO 9000*

Some of the gurus of Quality Control: W. Edwards Deming, Joseph Juran, Kaoru Ishikawa, Philip B. Crosby, Genichi Taguchi and Walter Shewhart were among the greatest contributors to Japan's improvements. Philip B. Crosby stressed that prevention is the key to quality management. With his "zero defects" concept, which he classified as an absolute of quality management, Crosby stressed the need for setting high standards and expectations to achieve quality. Japanese corporations such as Toyota, Honda, and Toshiba among others applied this concept to ensure that goods were conforming to requirements and consequently consumer confidence.

Kaoru Ishikawa's developments such as the "Quality Circles" and the "Cause and Effect" diagrams, also known as Ishikawa or fish-bone diagrams, have provided the Japanese with tools that promote quality awareness by focusing on the causes and effect of a process. Additionally, Ishikawa is one of the individuals responsible for introducing Shewhart's control chart techniques to the Japanese.

Walter Shewhart, who was an employee from the Bell Laboratories during the 1920's, has been considered a 20th century pioneer of quality given his significant contributions in the area of Statistical Process Control (SPC). Shewhart's contributions have included the concept of "assignable causes," the concept of Type I and Type II errors, when conducting hypothesis testing, and the first control charts. Shewhart developed the control chart as a tool to monitor product quality by detecting a shift in the target value as well as the presence of any variation in a given process. Control charts have a center line, where the target value is located, and control limits above and below the center line (CL), namely the Upper Control Limit (UCL) and the Lower Control

Limit (LCL) which allow a process to be monitored. (See Figure 2.1a) The process is monitored by detecting shifts. Once these shifts have been detected and the process is brought to a state of "statistical control," the next phase is to determine what possible factors may have contributed to these deviations and then proceed to improve the process. The tools that have been developed by Shewhart, Ishikawa, Deming, Juran, Crosby and Pareto are statistical tools used to monitor and control processes, a method known as Statistical Process Control (SPC). The following represent various definitions and attributes of SPC:

- The **Statistical** component (or statistics) is the science that collects, analyzes and interprets data. Data points collected from samples are known as statistics. The **Process** is a systematic series of events with inputs (or variables) and outputs (or effects). **Control** is the application of statistical techniques in order to monitor and improve a process. (American Society for Quality)
- "Statistical methods for analyzing and controlling the variation of a process." (DataMyte Corporation)
- A collection of "production methods and management concepts and practices that can be used throughout an organization." (Gerald Smith)
- A technique that provides continuous improvement to a process.
- A proactive approach that yields a consistently higher quality output.
- A technique that reduces rework and fewer errors or false alarm rates

consequently less waste.

- A system that promotes employee participation which leads to an increase in job satisfaction and performance.
- A process that improves competitive position resulting in more jobs and an increase in profits.

Table 2.1a illustrates the Type I and Type II errors and how they apply to Statistical Process Control. (Kolarik, 1995)

Table 2.1a SPC Proper Indicators and Erroneous Conclusions

	In-Control (Stable Process)	**Out-of-Control (Unstable Process)**
In-Control	Correct Conclusion	***Missed Alarm*** Fail to detect the unstable process (Type II error)
Out-of-Control	***False Alarm*** Signal in a stable process (Type I error)	Correct Conclusion

Shewhart's control charts are known as Statistical Process Control Charts. During the past eighty years control charts have been extensively used in industry due to their simplicity and efficiency for monitoring processes. Although there have been cases where these charts have been misused by supervisors, as has been referenced in various case studies by DeVor, Chang and Sutherland (1992), their contributions and applications in numerous industrial settings have been extensive for decades.

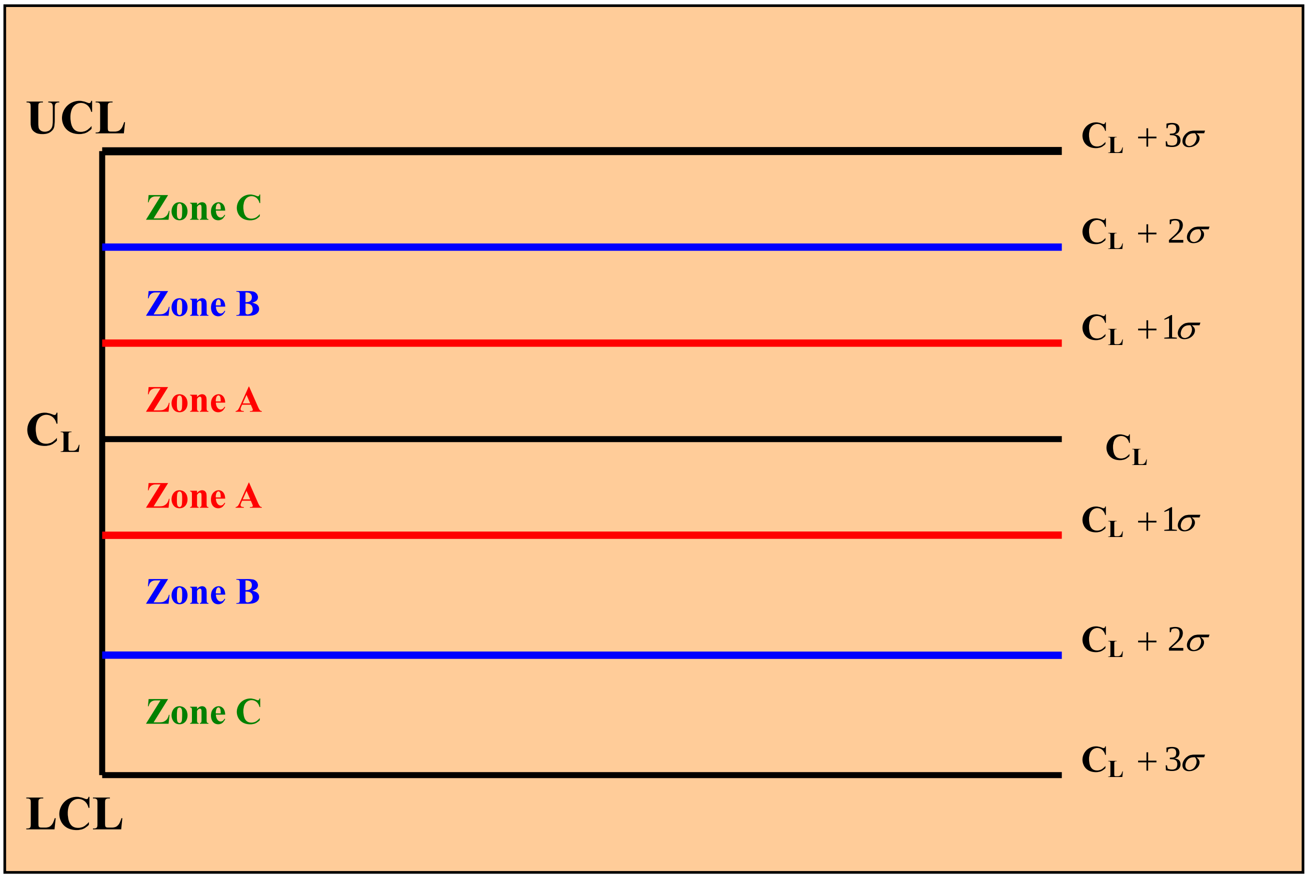

Figure 2.1a Shewhart Control Chart Zones for a 3 sigma X-bar chart. (Shewhart, 1939)

While these charts have been successfully applied to monitor processes in industry, they are limited to univariate data. For MSPC, the power of univariate charts is low, relative to a multivariate approach and may yield an increase in false alarms, whereby a system that was in a state of statistical control was deemed out of control. The correlations that are possible among process variables must be considered when using multivariate control charts as is the case with Hotelling's T^2 control chart. Unfortunately, given the restriction of multivariate normality, Hotelling's T^2 control chart can also generate a higher incidence of false alarms for multivariate processes that are not multivariate normal.

Another pioneer of Quality and Quality Management has been Dr. W. Edwards Deming, whose philosophy is the never–ending cycle of improvement, known as the Plan-Do-Check-Act (PDCA) cycle (see Figure 2.1b). The significance of Deming's contributions to the Japanese is evidenced by the establishment of The Deming Prize in Japan in 1951. Deming was a strong advocate of control chart techniques because of their ability to monitor processes in order to correct and continuously improve the overall systems. Applying the Deming PDCA Cycle, one must seek further improvement within SPC in order to efficiently detect target shifts and the presence of variation for multivariate processes.

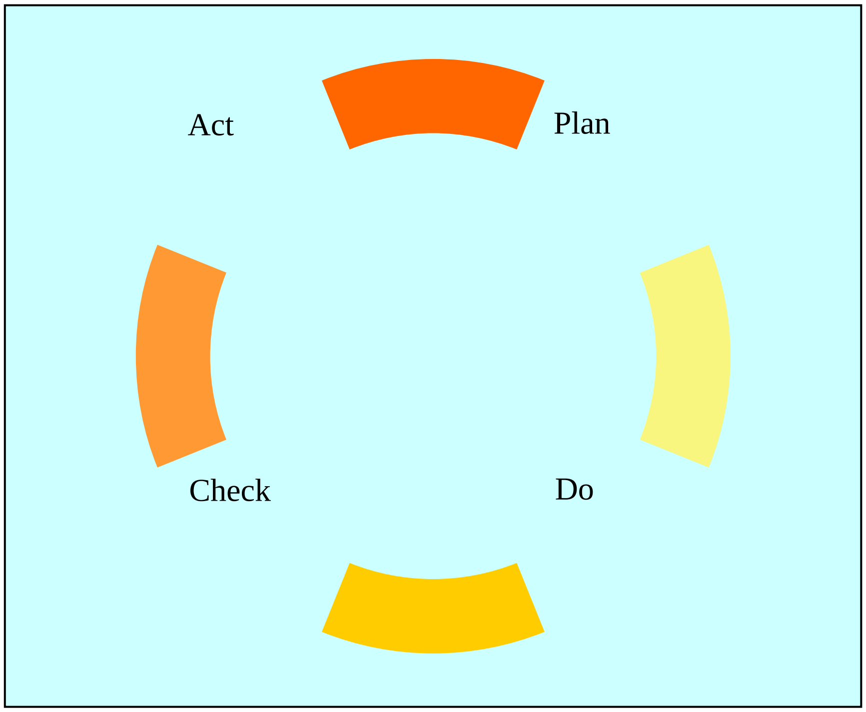

Figure 2.1b Deming's Plan-Do-Check-Act (PDCA) cycle (Kolarik, 1995)

2.2 Technological Influences

The phenomenon of the personal computer in the 1980's has changed the way in which personal and industrial business is conducted. In industrial and social environments the computer has introduced new venues of communication, namely the internet and e-mail transmission. With information so accessible, data availability and industrial interaction have become instantaneous. Businesses now have the ability of gathering and posting information at the click of a button. Software packages such as Minitab, SAS, Statistica, Mathematica, Design Expert, SPSS and others are being utilized to monitor processes through quality tools that have been included within these software packages.

With these computerized tools and the abundance of available data, additional factors or variables affecting individual processes have been discovered. With more factors determining a process, the univariate, or single–variable, approach used in traditional statistical process control cannot effectively monitor a process. These factors create a multivariate setting in which more than one variable will determine the outcome of a process. When the number of variables in a process system increases, the presence of correlation is likely. In addition to the possibility of correlations among variables, there exist correlations within the same variables over time, which is called autocorrelation. In certain industries such as the chemical and food industries, in which outcomes are produced by numerous factors and their interactions, correlation is a common occurrence. Such inter-correlations and/or autocorrelations could not be monitored by traditional control charts resulting in numerous false alarms and possible missed alarms for those processes.

Additionally, with more factors affecting a process and the fact that small samples are gathered, the normality assumption may be less justifiable. Therefore new and creative approaches need to be developed to monitor and improve multivariate processes that are free of the normality assumption. During the past eighty years the "one size fits all” univariate approach worked well, however, in many processes which are multivariate, the univariate approach on each single variable versus a multivariate approach yields a significant loss of power as demonstrated in the multivariate quality control literature (Fuchs and Kennet, 1998) as well as the theoretical multivariate literature (Rencher, 1995). Unfortunately, the vast amount of work in multivariate quality control is based on the assumption that these processes are multivariate normal. As such, many publications in MSPC utilize Hotelling’s T^2 or modified forms proposed by Mason, Young and Tracy (1997), and Kourti and MaGregor (1995, 1996). Details of these MSPC approaches using T^2 as well as new approaches using nonparametric developments will be provided later in this chapter.

According to Sprent (1989), nonparametric methods have played a central role in modern times, because modern channels of communication such as the internet provide an abundance of data with possibly no information as to the distribution of that data. Sprent further stresses how Wilcoxon and other researchers of nonparametric methods have determined that ranks could be utilized effectively with “little loss of information.” There have been limited efforts to utilize nonparametric techniques to address these MSPC situations; however, as the literature has suggested much research is needed in the area of nonparametric MSPC. Additionally, these investigations have been more in the arena of multivariate statistical theory (see APPENDIX A) and not so much in multivariate quality control.

2.3 Historical Developments in Multivariate Quality Control

One of the first researchers in the area of multivariate SPC was Hotelling (1947) whose research explored multivariate quality control. Even before the advent of the modern day personal computer, Hotelling realized that a process was not always determined by univariate measures. With such a discovery Hotelling developed the T^2 statistic to address multivariate process data. The concept behind the T^2 statistic was to develop a model that would test the hypothesis that a multivariate process is in a state of statistical control versus that the multivariate process is not in a state of statistical control. The T^2 statistic would monitor the multivariate structure of the process by observing the mean vector and variance-covariance matrix of the p number of variables that determine the multivariate process.

Some recent articles have referred to affine invariance as a desirable characteristic for process measures. An affine invariant statistic is a statistic which "is invariant under nonsingular linear transformations of the data" which include rotations, reflections, and rescaling. (Kapatou, 1996) The benefit of the rotatability of the data is that it generates the same conclusions gathered from the original data or the principal components of the rotated data. These principal components are those obtained from the original variables using a popular and well established technique from multivariate statistics known as Principal Component Analysis (PCA). Additionally, with the affine invariance property, the statistic manages to maintain the same value under scaling changes. Kapatou, (1996) Johnson and Wichern (1982) have demonstrated that the T^2 statistic developed by Hotelling does display the affine invariance property. The rotatability of the principal components is beneficial particularly in multivariate quality control. Graphical representations of data generate quick information for industrial managers who may not understand the statistical aspects but can gather information from a graph. With rotatability three dimensional graphs would provide consistency and not suffer from a loss of power.

2.4 Multivariate Graphical Tools

Graphical tools are also available measures in Multivariate Analysis. Two, three and even four dimensional graphing schemes are easy to understand and are readily available in many software packages, which are most beneficial in industry. One multivariate graph which is readily available in most statistical software packages is the scree graph which provides a visual test assessing the amount of principal components that should be retained when using Principal Component Analysis, a multivariate dimensional reduction scheme. The Q–Q plot which is another simple graphical tool can be plotted easily to detect trends and assess normality as in the univariate case. The Q–Q plot is also available in statistical software packages. Additional graphical techniques are available in multivariate texts and some software packages however, their usage may not be as reliable as the ones previously described given their subjectivity. Table 2.4a contains a description of various multivariate graphical tools that are available. (Rencher, 1995)

Table 2.4a Multivariate Graphical Tools.

Multivariate Graphical Analysis	Descriptions of Each Graph
Bivariate Scatter Plots	Two dimensional graph with the plots of paired data points. The data points are obtained from 2 variables.
Trivariate Scatter Plots	Three dimensional graph with the plots of data points in 3-space. The data points are obtained from 3 variables.
Four Dimensional Plots	Scatter plots graphed in 2 dimensions with pairs of variables within a larger pair. Each corner for the smaller right angle pair (i.e. the additional pair) would be the scatter for the outer 2 variables.
Q–Q plots	Graphs of quantiles used to assess normality and/or identify trends.
Profiles	Vertical bars with heights representing the values of the variables.
Stars	Rays from the center to the outside of a circle representing the values of the variables. These rays form a polygon. The center of the circle would be the minimal value.
Glyphs	Circles with fixed rays representing the values of the variables.
Faces	Depict each variable as a feature on a face (eyes, nose, mouth, etc.).
Boxes	Each variable is the dimension of a box.
Scree Graphs	Graphical technique to determine the principal components in Principal Component Analysis.

2.5 Principal Component Analysis

Principal Component Analysis (PCA) is a popular multivariate analysis technique that is used to reduce the *p*-dimensionality of a multivariate process. One benefit of this technique is that it is readily available in most software packages and is a well established technique in statistical multivariate analysis. This section will provide insight into PCA as currently used in MSPC.

We begin with a multivariate process having *p* process variables in which $\mathbf{x}_i$ represents each [original] process observation vector with $i = 1,\ldots,n$. Let $\mathbf{z}_i = \mathbf{A}\mathbf{x}_i$ where $\mathbf{A}$ is an orthogonal matrix and $\mathbf{A}'\mathbf{A} = \mathbf{I}$ then $\mathbf{z}_i'\mathbf{z}_i = (\mathbf{A}\mathbf{x}_i)'\mathbf{A}\mathbf{x}_i = \mathbf{x}_i'\mathbf{A}'\mathbf{A}\mathbf{x}_i = \mathbf{x}_i'\mathbf{x}_i$. The $\mathbf{x}_i$'s have been transformed to $\mathbf{z}_i$ in which the axes are rotated, but with the same distance from the origin. Each transformation $\mathbf{z}_i$ is in fact a linear combination of the original variables. The total number of principal components is the same as the original number of variables. The new variables are the principal components $\mathbf{z} = \mathbf{A}\mathbf{x}$ and are not mutually correlated, therefore, the covariances are equal to 0. The sample covariance matrix of $\mathbf{z}$ is $\mathbf{S}_z = \begin{bmatrix} s_{z1} & 0 & \cdots & 0 \\ 0 & s_{z2} & \cdots & 0 \\ \cdots & \cdots & s_{ii} & \cdots \\ 0 & 0 & \cdots & s_{zp} \end{bmatrix}$.

Given that $\mathbf{z} = \mathbf{Ax}$ it follows that $\mathbf{S}_z = \mathbf{ASA}' = \begin{bmatrix} s_{z1} & 0 & \cdots & 0 \\ 0 & s_{z2} & \cdots & 0 \\ \cdots & \cdots & s_{zi} & \cdots \\ 0 & 0 & \cdots & s_{zp} \end{bmatrix}$ with **S** being the sample covariance matrix of **x**. **S** is diagonalized by the orthogonal matrix **A** that is the transpose of matrix **C** with columns representing the normalized eigenvectors of **S**. Each a_i represents the i^{th} normalized eigenvector of **S**.

$\mathbf{C} = \mathbf{A}' = \lfloor \mathbf{a}_1 \quad \mathbf{a}_2 \quad \cdots \quad \mathbf{a}_p \rfloor$ equivalently $\mathbf{C}' = \mathbf{A} = \begin{bmatrix} \mathbf{a}_1' \\ \mathbf{a}_2' \\ \vdots \\ \mathbf{a}_p' \end{bmatrix}$

Since $\mathbf{z} = \mathbf{Ax}$, the principal components are:

$$\mathbf{z}_1 = \mathbf{a}_1'\mathbf{x}_1 = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1p}x_p$$

$$\mathbf{z}_2 = \mathbf{a}_2'\mathbf{x}_2 = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2p}x_p$$

$$\vdots$$

$$\mathbf{z}_n = \mathbf{a}_p'\mathbf{x}_n = a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{np}x_p$$

$\mathbf{D} = \mathbf{C}'\mathbf{SC} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \lambda_i & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{bmatrix}$ where **D** is the resulting diagonal matrix. Thus, the eigenvalues $\lambda_i = s_{zi}$ represent the variances of the principal components. Since the total number of principal components is the same as the original number of variables, the total variance of the principal components account for the total variance of the original variables. Therefore, $\sum_{i=1}^{p} \lambda_i$ the total of the eigenvalues of all principal components equals the total variance. From **S** which is the covariance matrix of **x**, we have the total sample variance is the trace $tr(\mathbf{S}) = \sum_{i=1}^{p} s_{ii}$, the sum of the diagonal elements in the sample covariance matrix. $\therefore \sum_{i=1}^{p} \lambda_i = tr(\mathbf{S}) = \sum_{i=1}^{p} s_{ii}$ The Proportion of Variance is the ratio of the sum of the eigenvalues with k principal components from a total of p principal components. The analysis of eigenvalues, also known as the eigenanalysis, can be summarized by creating a table, such as Table 2.5a.

$$\text{Proportion of Variance} = \frac{\lambda_1 + \lambda_2 + \cdots + \lambda_k}{\lambda_1 + \lambda_2 + \cdots + \lambda_p} = \frac{\lambda_1 + \lambda_2 + \cdots + \lambda_k}{\sum_{i=1}^{p} s_{ii}}$$

Table 2.5a Eigenvalues and Proportion of Variance

Eigenvalue	**Proportion of Variance**	**Cumulative Proportion of Variance**
λ_1	$PV_1 = \dfrac{\lambda_1}{\sum_{i=1}^{p} s_{ii}}$	PV_1
λ_2	$PV_2 = \dfrac{\lambda_2}{\sum_{i=1}^{p} s_{ii}}$	$PV_1 + PV_2$
λ_3	$PV_3 = \dfrac{\lambda_3}{\sum_{i=1}^{p} s_{ii}}$	$PV_1 + PV_2 + PV_3$
$\vdots$	$\vdots$	$\vdots$
λ_k	$PV_k = \dfrac{\lambda_k}{\sum_{i=1}^{p} s_{ii}}$	$PV_1 + PV_2 + \cdots + PV_k = 1$

There are two approaches. The first approach is to standardize the data by subtracting the mean and dividing by the standard deviation. By standardizing the data, all variables have the same standard deviation, namely 1. When using this approach the eigenvalue analysis is on the correlation, thus we analyze the **R** matrix. Here we must specify the center and standardize the variables, which transforms the centroid of the entire data set to 0. If the variables are in different units then we must use the correlation matrix in order to standardize the variables. The second approach is the eigenvalue analysis of the covariance matrix. When using this approach without standardizing the data, we perform and analyze the **S** matrix.

From the eigenanalysis, we move to retaining the principal components. There are three methods for retaining principal components. The first method is to retain those components accounting for a large cumulative proportion or percentage of variation. This approach can be used in either the eigenanalysis of the correlation matrix or the covariance matrix. For our research, we will demonstrate what cumulative percentage of variation would be more robust for our unique distribution free MSPC approach utilizing PCA. The second method for retaining principal components is applicable only to the eigenanalysis of the correlation matrix. In this method, those components with eigenvalues below the average eigenvalue $\frac{\sum_{i=1}^{p} \lambda_i}{p}$ will be excluded. Since the total variance of the correlation matrix equals p and the determination in this method is based on the total variance equal to p components, eigenanalysis is conducted on the correlation matrix. The third method is a graphical approach known as the scree graph analysis. A scree graph is obtained by plotting each i against the eigenvalue λ_i in order to graphically compare large and small eigenvalues. We retain the principal components with the eigenvalues that show a significant steep slope. The cutoff will be the one with the last steepest slope. (see Figure 2.5a)

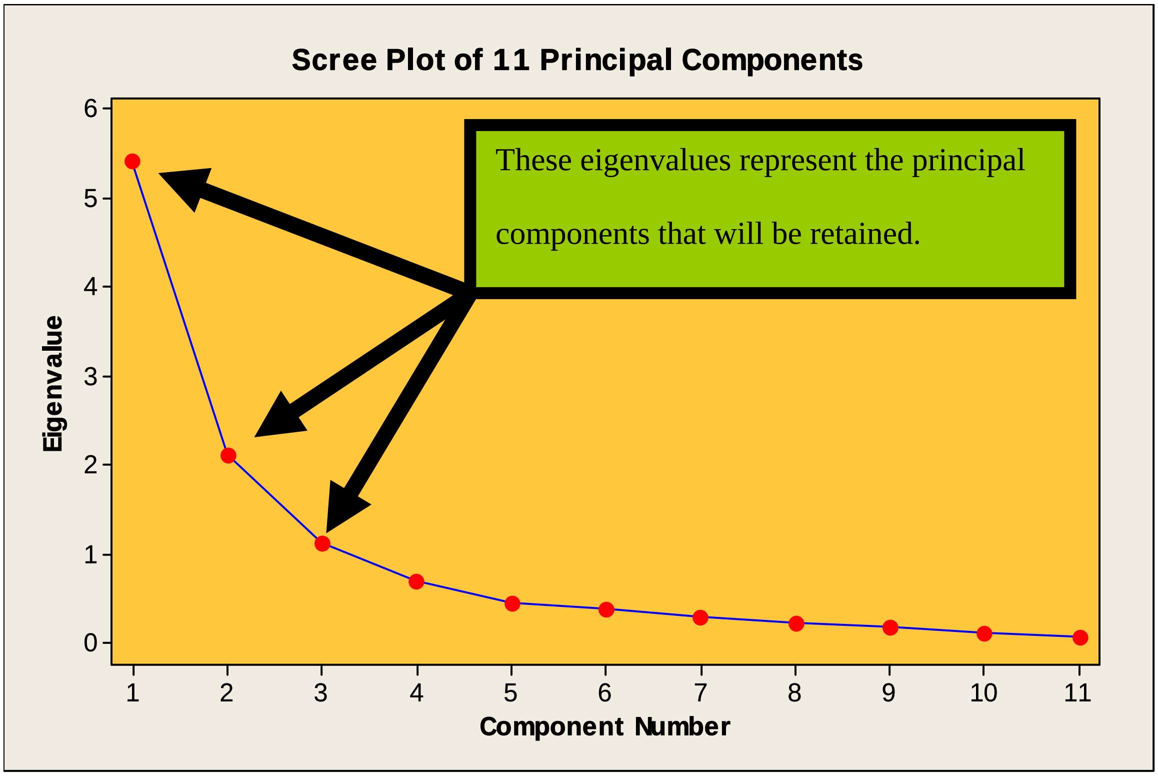

Figure 2.5a. A Sample Scree Graph of all 11 principal components when p =11.

The purpose of the PCA approach is dimensionality reduction that will consider the contributions of all variables and by their principal components indicate which variables in fact represent the largest contribution. The number of p principal components equals the total number of p variables. However, the k principal components that are retained are those that account for the largest variability in the system. One advantage of principal components is that they are mutually independent. If the number of retained principal components equals the number of original process variables then dimensionality reduction is not possible. But as the literature in quality suggests, although a process may be determined by a large number of variables, it is more likely that a subset of all variables will drive the process. (Zhang 2003)

If two components account for the same variation then the data cloud is circular instead of elliptical. If one accounts for higher variation then that elliptical cloud formed is parallel to the axis of the first component. Since the first set of components account for a larger portion of the variability, the PCA axes are obtained from the first two or three components. The method of constructing the first axis PCA1: The line must go through the centroid with minimum squared distances from each point to the line. The method of constructing the second axis PCA2: The line also passes through the centroid, but must be orthogonal to PCA1 as such the components are uncorrelated. In a three dimensional plot, PCA3 would also pass through the centroid and orthogonal to the first two components. Figure 2.5b is a sample representing the elliptic region of a pair of transformed variables.

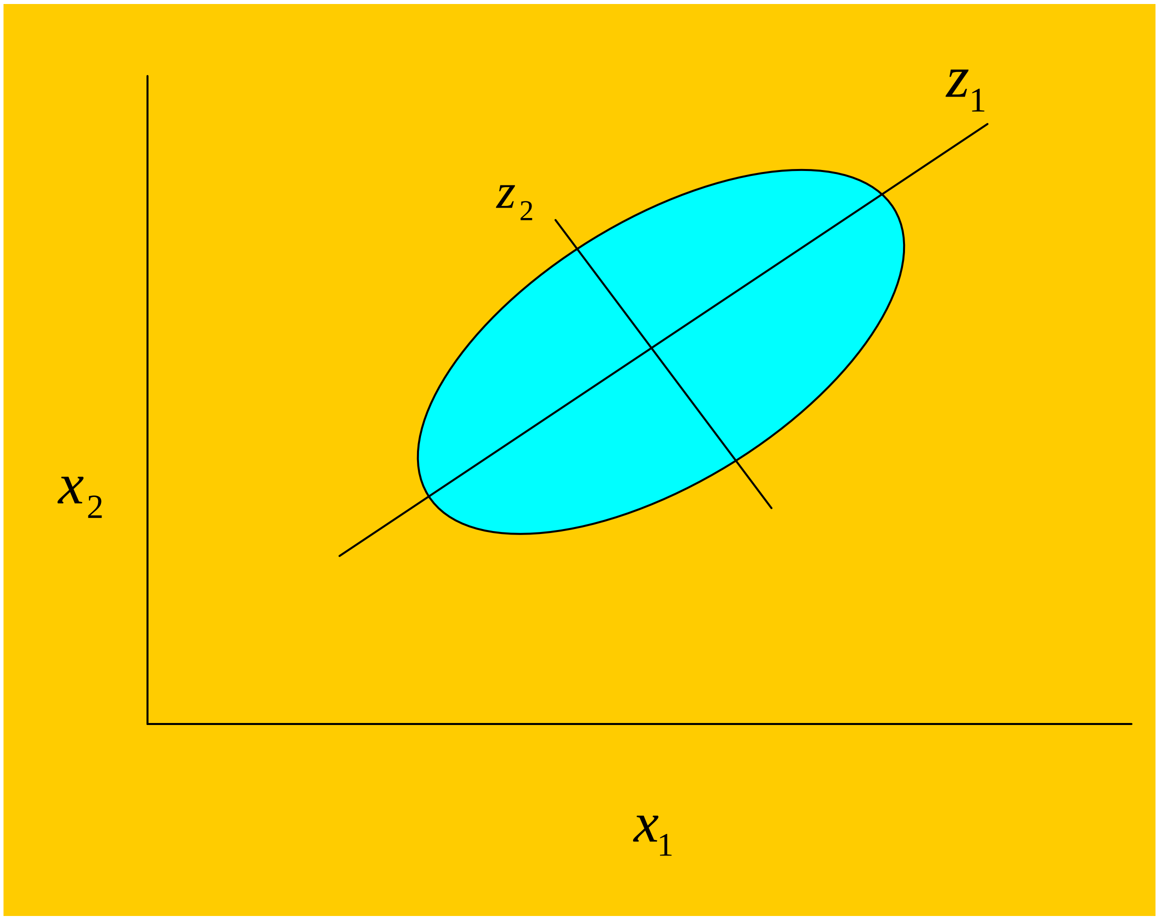

Figure 2.5b. The graph of the first two principal components.

2.6 Construction of a Traditional Multivariate Control Chart

In this section we will discuss the control chart stopping rule in MSPC using the traditional T^2 approach. Given that the focus of this research is in MSPC, we will limit our discussion in this section to the applications of the T^2 statistic in MSPC. We begin with the p process variables identified by the process owner along with the base data where the process was considered to be stable or in control. The base data will be referred to as the historical data set (HDS) which will be used for monitoring new observations. Next, one must determine the test statistic to be used. Under the assumption that the process is multivariate normal, the Hotelling's T^2 statistic would be computed and plotted using the T^2 Control Chart. This approach can be found in numerous publications by Mason and Young (2000, 2002, 2005) and Kourti and MacGregor (1995, 1996) in the multivariate quality literature. T^2 is a univariate measure representing the distances for multidimensional observation vectors. Each observation vector $\mathbf{x}_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{bmatrix}$ is the random vector of p variables with i = 1,…,n observations. $T^2 = (\mathbf{x} - \bar{\mathbf{x}})\mathbf{S}^{-1}(\mathbf{x} - \bar{\mathbf{x}})$ with $\bar{\mathbf{x}}$ and $\mathbf{S}$ representing the mean vector and covariance matrix estimators from a historical data set (HDS). $T^2 = (\mathbf{x} - \bar{\mathbf{x}})\mathbf{S}^{-1}(\mathbf{x} - \bar{\mathbf{x}}) \sim \left[\frac{p(n+1)(n-1)}{n(n-p)}\right]F_{(p,n-p)}$ with p and n – p degrees of freedom. The upper control limit is $\text{UCL} = \left[\frac{p(n+1)(n-1)}{n(n-p)}\right]F_{(\alpha,p,n-p)}$ where $F_{(\alpha,p,n-p)}$ represents the upper α^{th} quantile of $F_{(p,n-p)}$. (Chou, Mason and Young 2001) Since T^2 is a square there are no negative values. The minimum value is zero which is the ideal T^2 value. At zero, the observation vector $\mathbf{x}_i$ is on target i.e. "located at the process center." (Mason and Young, 2000) The following figures illustrate the T^2 Control Charts. The first is a univariate control chart scheme for multivariate observations. (see Figure 2.6a)

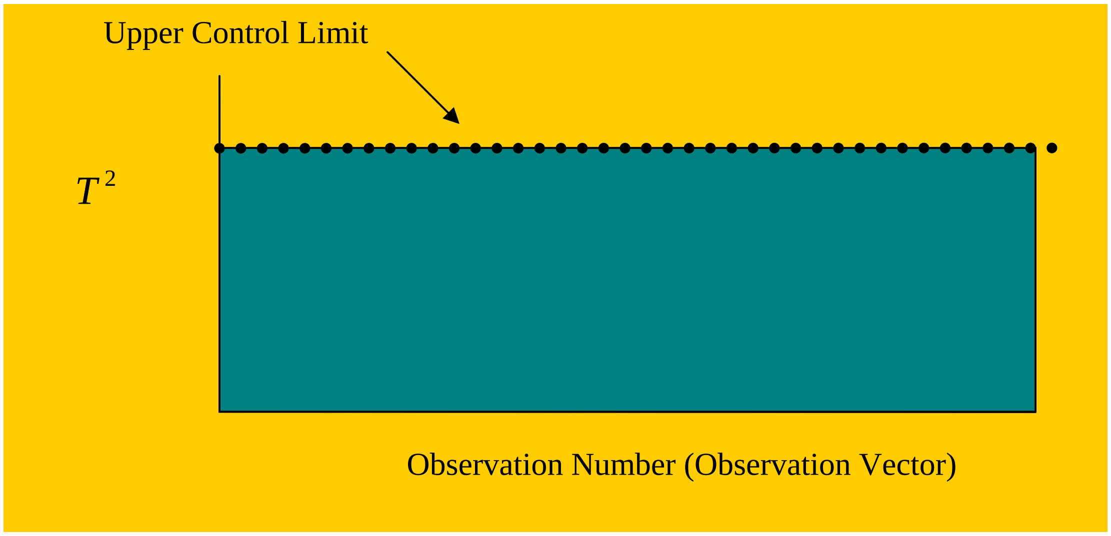

Figure 2.6a Univariate Control Chart scheme for Multivariate Observations

Figures 2.6b and 2.6c are provided as a comparison of bivariate plots of two process variables and two principal components. The elliptical control charts (Figures 2.6b and 2.6c) illustrate the same elliptical UCL for bivariate data. In both cases with p = 2, we see that the UCL is the same for the two original process variables x_1 and x_2 as well the two principal components z_1 and z_2. The same is true for multivariate cases beyond the bivariate when $p > 2$.

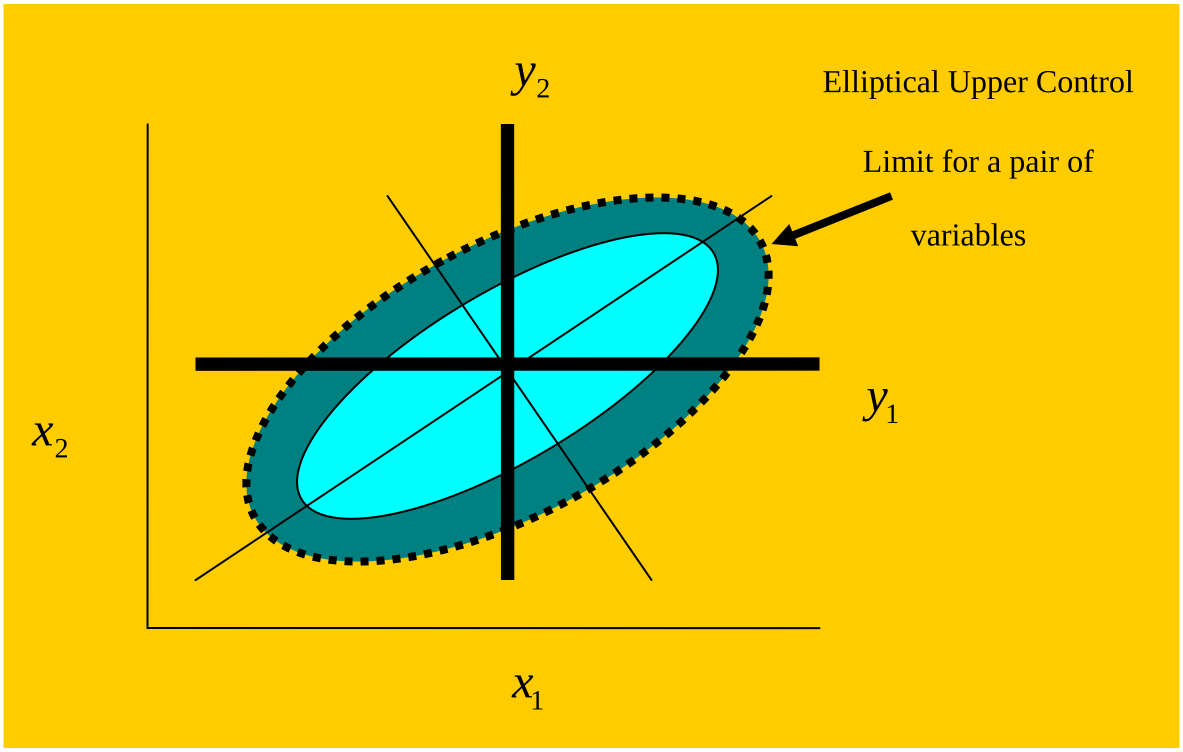

Figure 2.6b Elliptical Control chart scheme for Multivariate Observations

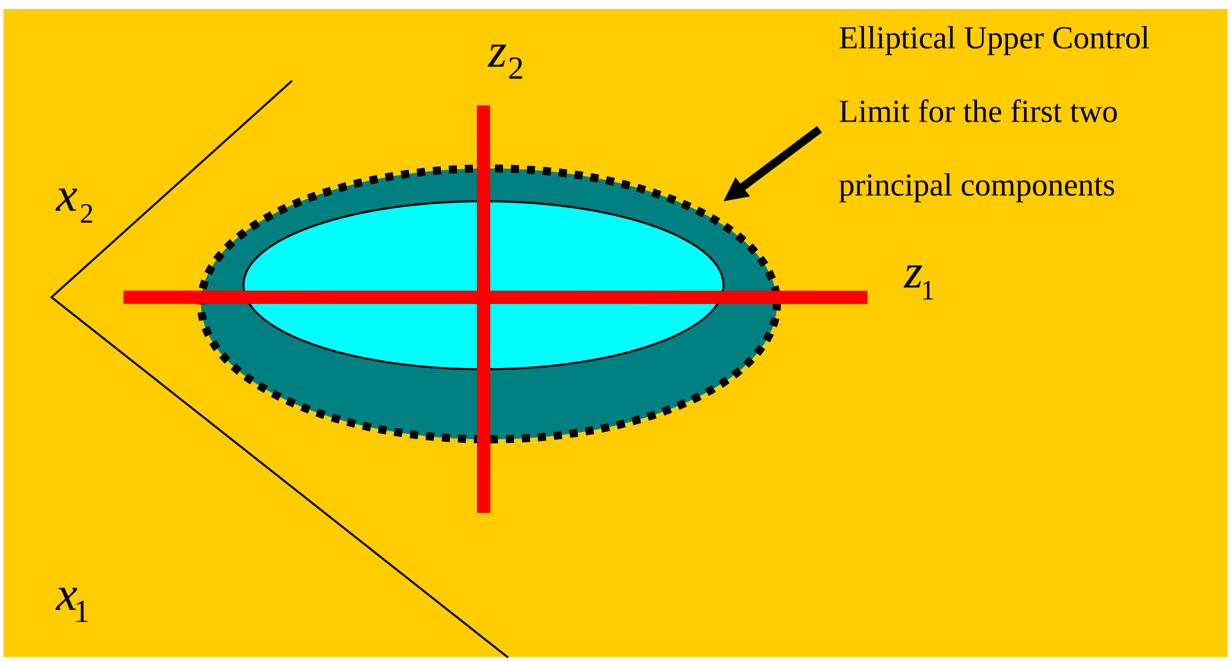

Figure 2.6c Elliptical Control Chart scheme of the first two principal components.

In trivariate cases, the graphs of the clouds would be represented using ellipsoid control limits. However, in higher dimensionality such graphical schemes become very difficult to use and understand. Multidimensional processes that can be monitored using univariate schemes are more desirable given the ease of use and interpretability. Such is the case for the univariate scheme presented in the univariate T^2 control chart (Figure 2.6a) as well as the univariate plots for multivariate observations proposed by Liu (1995). Details of the Liu control charts will be discussed in section 2.7.

Kourti and MacGregor (1995, 1996) have illustrated how projection methods such as PCA can be used to improve signal detection in MSPC. Their methodology is to perform PCA and then use a reduced T^2 scheme with $k < p$ observation vectors. The full p dimensional T^2 statistic can be described as follows with $k < p$ principal components:

$$T^2 = T_k^2 + \sum_{i=k+1}^{p} \frac{z_i^2}{s_{ii}} \text{ with } T_k^2 \text{ denoting the reduced } T^2 \text{ of } k \text{ components}$$

Kourti and MacGregor (1996) compute the UCL based on T_k^2 with $(k, m-k)$ degrees of freedom. The T_k^2 chart follows the same univariate scheme as the full p dimensional T^2 control chart with the focus on the k principal components that drive the process. Additionally, Kourti and MacGregor (1996) suggest analyzing the remaining $p - k$ discarded components with the use of the SPE (squared prediction error), based on the Q statistic by Jackson (1991), described as a "companion statistic" (Milectic, et.al., 2004) to the T_k^2. These variations of the full p dimensional T^2 statistic assume multivariate normality.

The T^2 statistic is most beneficial in industry, given that this statistic is readily available in software packages and is well known. Additionally, as we have illustrated, the T^2 statistic, in full as well as reduced form, can be plotted in a simple univariate chart as well as an elliptical chart (for two variables). Unfortunately, in non-normal settings, the T^2 statistic suffers a loss of power. Since the T^2 statistic requires the assumption that the distribution must be multivariate normal, Mason and Young (2002) suggest using the Kernel smoothing distribution function of T^2 for a multivariate non-normal approximation of the T^2 distribution when the assumption of normality is rejected. The kernel estimate or kernel distribution of T^2 is denoted by $FK(t)$ where $FK(t) = \frac{1}{1000}\sum_{j-1}^{1000}\varphi\left\{\frac{t-T^2(j)}{h}\right\}$, h is a two stage estimate of the bandwidth (Polansky, 1997) and φ represents the standard normal distribution function.

The next step is the stopping rule or the signal. Large values of T^2, those values beyond the Upper Control Limit, are signals indicating that the observation vector $\mathbf{x}_i$ is out-of-control. If the observed $T^2 < \mathrm{UCL}$, the observation vector $\mathbf{x}_i$ is in control. If the process is stopped then a diagnostic is needed to determine what process shift may have contributed most to the signal. There are two ways of producing these signals: “moving a particular variable’s observed value beyond its operational range and/or contaminating a linear relationship between two or more process variables." (Mason and Young, 2000) In order to determine the most likely cause(s) of this process degradation when the signal occurred, the process owner should be consulted at this stage in order to provide good insight into the process itself. Now, we must take corrective measures to lower the variation and/or correlation effect that may have contributed to the signal and proceed to improve the process. Using the Deming philosophy, the process resumes with new improvements and the continuous improvement cycle restarts.

2.7 Nonparametric Multivariate Control Charts

In the literature very little can be found with regards to nonparametric multivariate quality since the emphasis on such an approach is more recent. Since the 1950's the developments in general nonparametric multivariate methods have been more prevalent. (See APPENDIX A) Within the last decade nonparametric schemes have been proposed in MSPC by Hayter and Tsui (1994), Kapatou (1996), Liu (1995, 2003) and Zarate (2003). Hayter and Tsui (1994) proposed the use of a location statistic called the M statistic to monitor process location, however this scheme fails to monitor correlations among multivariate components which is common in multivariate quality control. Kapatou (1996) on the other hand utilizes test statistics from theoretical multivariate analysis whereby vectors of multivariate locations are used to monitor shifts when small samples are collected and normality cannot be satisfied. Kapatou (1996) demonstrates how her method is more robust than Hotelling's T^2 when non-normal multivariate samples are collected. However, her calculations are quite difficult to compute and not readily available in statistical software packages. Additionally, Kapatou's approach suffers from the lack of affine invariance, so rotatability, scaling and transformations using linear combinations could alter the results. The aforementioned properties are affine invariant and essential for performing eigenvalue analysis and utilizing dimensionality reduction schemes such as Principal Component Analysis. Given the lack of affine invariance in the Kapatou (1996) approach and the inability of the Hayter and Tsui (1994) approach to monitor correlations between the components, we will focus on the Liu (1995) approach based on data depth which displays the property of affine invariance. We will now provide details of the Liu approach (1995) as well as an extension to the Liu developments proposed by Zarate (2003).

The Liu (1995) nonparametric control charting scheme is based on ranking data depths of the multivariate observations of the *p* process variables and plotting these ranks using a univariate control chart scheme in order to easily detect multivariate process shifts visually. Liu (1990) provides various definitions of data depths: Mahalanobis depth, Tukey's depth, Majority (or Halfspace Depth) and Simplicial Depth. Each data depth measure represents the geometrical notion of depth of a point within a data cloud. Liu (1995, 2003) states the Mahalanobis depth is the easiest to compute but is not robust given that this measure is computed using "nonrobust statistics: the sample mean and sample dispersion." Liu (1995) uses simplicial depth which is a robust statistic for her proposed nonparametric data depth MSPC control charts.

Liu (1990) defines simplicial depth relative to the probability that a point lies within a random simplex with vertices $p + 1$. These $p + 1$ vertices are independent observations from the distribution F (not to be confused with the F-distribution). Liu (1990) further indicates $D_F(x)$ also denoted $D(x)$ represents the measure of depth of a point $\boldsymbol{x}$ with respect to the continuous distribution F. For some given point $\boldsymbol{x}$ "in the p-dimensional Euclidean space $\mathbf{R}^p$, simplicial depth is a measure of how central" $\boldsymbol{x}$ lies within a random sample $\{X_1, \ldots, X_n\} \subset \mathbf{R}^p$. (Stoumbos , et.al. 2001) The empirical distribution of $D_F(x)$ or $D(x)$ is denoted as $D_{F_n}(x)$ or $D_n(x)$. The empirical distribution $D_n(x)$ "gives rise to a natural ordering of the data points from the center outward." This ordering is an extension of the univariate sample median and L–statistics (linear combinations of order statistics) to the multivariate case, where both are affine invariant statistics. The notion of simplicial data depth is a geometrically affine invariant measure whereby we use the concept of the depth of a point within a simplex. (Liu, 1990)

Let $\{X_1, \ldots, X_n\} \subset \mathbf{R}^p$ denote the set of n observations in p-dimensional space.

For the univariate case where *p* = 1, the simplicial depth of a given point $\boldsymbol{x}$ represents the proportion of closed intervals that contain $\boldsymbol{x}$, among all closed intervals formed by pairs of points from $\{X_1, \ldots, X_n\} \subset \mathbf{R}^1$. Each closed interval represents the line segment $\overline{X_i X_j}$ formed between each pair X_i, X_j.

<u>For the bivariate case where *p* = 2</u>, the simplicial depth of a given point ***x*** represents the proportion of triangles that contain ***x***, among all triangles formed by triple sets of points X_i, X_j, X_k from $\{X_1, \ldots, X_n\} \subset \mathbf{R}^2$.

<u>For the trivariate case where *p* = 3</u>, the simplicial depth of a given point ***x*** represents the proportion of tetrahedrons that contain ***x***, among all tetrahedrons formed by quadruple sets of points $\{X_1, \ldots, X_n\} \subset \mathbf{R}^3$.

The following example for bivariate data is illustrated by Liu (1990) and is presented here in order to provide insight into simplicial data depth. Let $X_1, \ldots, X_n$ represent a bivariate data with *n* observations. Any three data points will form a closed triangle with vertices X_i, X_j, X_k denoted $\Delta(X_i, X_j, X_k)$. Then, there are $\binom{n}{3}$ number of triangles for *n* observations. Define event A as the event where point ***x*** lies within a triangle $\Delta(X_i, X_j, X_k)$ in other words A is defined as $x \in \Delta(X_i, X_j, X_k)$. The indicator function indicates the probability that $x \in \Delta(X_i, X_j, X_k)$ or the probability of event A denoted as $I(A)$. The indicator function is represented as

$$I(A) = I\left(x \in \Delta\left(X_i, X_j, X_k\right)\right) = \begin{cases} 1 & \text{if } x \in \Delta\left(X_i, X_j, X_k\right) \\ 0 & \text{otherwise} \end{cases}.$$

The function $D_n(x) = \binom{n}{3}^{-1} \sum_{1 \le i < j < k \le n} I\left(x \in \Delta\left(X_i, X_j, X_k\right)\right)$ represents the proportion of triangles that contain point ***x***. Liu (1990) presents the following analogy,

> "**imagine placing a layer of clay with thickness** $\binom{n}{3}^{-1}$ **on the region corresponding to each triangle** $\Delta\left(X_i, X_j, X_k\right)$**, one by one until all** $\binom{n}{3}$ **triangles are exhausted**."

The solid formed will represent the exact shape of $D_n(x)$, therefore $D_n(x)$ is the empirical distribution of $D(x)$ in which the X_i's are i.i.d. (independently and identically distributed) with a common distribution *F*. As ***x*** moves closer to the center of the distribution the value of $D(x)$ increases, while the value of $D(x)$ decreases as ***x*** moves away from the center.

We define $D(x)$ where $D(x) = P_F\left(x \in \Delta\left(X_1, X_2, X_3\right)\right)$ as the simplicial depth of ***x*** with respect to *F* in $\mathbf{R}^2$.

We define $D_n(x)$ where $D_n(x) = \binom{n}{3}^{-1} \sum_{1 \le i < j < k \le n} I\left(x \in \Delta\left(X_i, X_j, X_k\right)\right)$ as the sample simplicial depth of ***x*** with respect to the data cloud $X_1, \ldots, X_n$ in $\mathbf{R}^2$.

In contrast, the simplicial depth for the univariate case $\mathbf{R}^1$ from 2 observations is $D(x) = P\left(x \in \overline{X_1 X_2}\right)$ with a cumulative density function F. We have $D(x) = 2F(x)[1 - F(x)]$ when F is continuous. The population median is defined as any point $\boldsymbol{x}$ which maximizes $D(x)$. In the univariate case the maximum value of $D(\cdot)$ is 0.5 and decreases monotonically to 0 as $\boldsymbol{x}$ moves away from the center (median). For the bivariate case $D(\cdot)$ represents the bivariate simplicial median. Liu (1990) uses population median notation μ (not to be confused with the mean), and $\hat{\mu}_n$ for sample bivariate median which represents the point that has the highest SD (simplicial depth). If a maximum occurs at more than one point then $\hat{\mu}_n$ would be obtained by calculating the average of those X's. The following is a heuristic for $\hat{\mu}_n$ as sample median: If $D(\cdot)$ is continuous and μ maximizes it in $\mathbf{R}^2$, then the point estimate for μ would be a point x_0 in the plane which maximizes $D_n(\cdot)$ Liu (1990) stipulates that if the F distribution "has a nonzero density in the neighborhood of μ, we would expect the data point X_{i_0} which maximizes $D_n(\cdot)$ among all the data points to be close to x_0 and, hence, to μ."

<u>For the general multivariate case for all *p* > 1</u>

We define the simplicial depth of $\boldsymbol{x}$ with respect to a continuous distribution F (not the F-distribution) as $D(x) = D_F(x) = P_F\left(x \in s\left\lfloor X_1, \ldots, X_{p+1} \right\rfloor\right)$ where $s\left\lfloor X_1, \ldots, X_{p+1} \right\rfloor$ represents a p dimensional simplex with vertices $X_1, \ldots, X_{p+1}$ which are random observations from *F*. The *s* for the univariate, bivariate and trivariate cases represents the line segment, triangle and tetrahedron, respectively. This measure $D_F(x)$ describes "how central the point ***x*** is within the distribution." (Liu, 1990)

When the distribution is unknown, we use a reference sample $X_1,\ldots,X_n$ to compute the sample simplicial depth. The sample simplicial depth is $D_n(x) = \begin{pmatrix} n \\ p+1 \end{pmatrix}^{-1} \sum_{1\le i<\ldots<p+1\le n} I\left(x \in s\left[X_1,\ldots,X_{p+1}\right]\right)$ in which F_m represents the empirical distribution of $X_1,\ldots,X_n$ with $n \geq (p+1)$ and $I(\cdot)$ is the indicator function $I\left(x \in s\left[X_1,\ldots,X_{p+1}\right]\right) = \begin{cases} 1 & \text{if } x \in s\lfloor X_1,\ldots,X_{p+1} \rfloor \\ 0 & \text{otherwise} \end{cases}$. This measure $D_n(x)$ determines "how central (or outward) the point ***x*** is within the data cloud" of $X_1,\ldots,X_n$. (Liu, 1990)

In lower dimensionality where *p* = 1, 2 or 3 it is easy to verify graphically if point ***x*** lies inside the simplex which is a segment, a triangle or a tetrahedron, respectively. The space $\mathbf{R}^p$ has *p*-dimension denoted $\dim\left(\mathbf{R}^p\right) = p$ which means that there exist *p* linearly independent vectors. (Magaril–Il'yaev and Tikhomirov, 2003) According to Liu (1990), in higher dimensionality, it is a straightforward verification that ***x*** lies inside the simplex $s\lfloor X_1,\ldots,X_{p+1} \rfloor$ which is accomplished by solving this system of linear equations consisting of *p* + 1 unknowns which are scalars:

$$x = \alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_{p+1} X_{p+1} \text{ (or expressed as } x = \sum_{i=1}^{p+1} \alpha_i X_i \text{)}$$

$$\alpha_1 + \alpha_2 + \cdots + \alpha_{p+1} = 1 \text{ (or expressed as } \sum_{i=1}^{p+1} \alpha_i = 1\text{)}$$

Geometrically, the point *x* of the simplex is uniquely defined as the sum of the vertices for that simplex multiplied by a factor of α_i, in other words as a linear combination of its vertices of that simplex. These vertices $X_1, \ldots, X_{p+1}$ define coordinates $\alpha_1, \ldots, \alpha_{p+1}$ which are called the barycentric coordinates of point *x* in the vector space. We can define point *x* in the vector space:

$$(\alpha_1 + \cdots + \alpha_{p+1})x = \alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_{p+1} X_{p+1}$$

such that α_i's where $i = 1, \ldots, p+1$ are the barycentric coordinates of point *x* in the vector space with respect to these vertices $X_1, \ldots, X_{p+1}$. The coordinates of the vertices of a standard simplex with unit distance are $(1,0,...,0)$, $(0,1,...,0)$, $(0,0,1,...,0)$, . . . , $(0,0,...,1)$. In convex geometrical analysis it is illustrated that these barycentric coordinates are not unique such that for any nonzero constant *c*, the following points $(c\alpha_1, c\alpha_2, \ldots, c_{p+1}\alpha_{p+1})$ are also coordinates of *x*. This linear system of equations has a unique solution for a "nondegenerate" simplex and *x* lies within the simplex

$s\lfloor X_1, \ldots, X_{p+1} \rfloor$ if and only if all α_i's where $i = 1, \ldots, p+1$ are positive. (Magaril–Il'yaev and Tikhomirov, 2003)

The following demonstrates the affine invariance of $D_F(x)$. If we let *A* represent a $p \times p$ matrix and *b* a point in $\mathbf{R}^p$, from the linear system of equations above, we obtain $D_{A,b}(Ax+b) = D(x)$. Let $D_{A,b}(y)$ represent the probability that a point *y* in $\mathbf{R}^p$ lies within the simplex with vertices $AX_i + b$ where $i = 1, \ldots, p+1$. Thus after the transformations the simplicial depth $D_F(x)$ is affine invariant. (Liu, 1990)

Liu (1990) demonstrates that if *F* is absolutely continuous when $n \to \infty$, $D_n(x)$ "converges uniformly and strongly to $D(x) = D_F(x)$." Additionally, $D_n(x)$ is affine invariant. The affine invariance property ensures that Liu's "control charts are coordinate free." (Liu, 1995) From an in-control reference sample or historical data set (HDS), simplicial depth from the data is utilized as a measure to monitor if that process data is in control.

Liu (1995) also defines data depth for point ***x*** in the distribution *F* from the mean using Mahalanobis distance. Liu indicates that aside from the simplicial depth, which defines depth using medians, data depth can also be defined as:

$MD_F(x) = \dfrac{1}{\left[1 + (x - \boldsymbol{\mu}_F)' \boldsymbol{\Sigma}_F^{-1} (x - \boldsymbol{\mu}_F)\right]}$ where $\boldsymbol{\mu}_F$ and $\boldsymbol{\Sigma}_F$ represent the mean and covariance matrix of the distribution.

$MD_n(x) = \dfrac{1}{\left[1+(x-\bar{\mathbf{x}})'\mathbf{S}^{-1}(x-\bar{\mathbf{x}})\right]}$ where $\bar{\mathbf{x}}_F$ and $\mathbf{S}_F$ represent the sample mean and sample covariance matrix for the empirical distribution. When the data depth is defined by the Mahalanobis distance, the depth of point ***x*** within the distribution is measured as the reciprocal of the quadratic distance to the mean.

Three charts have been proposed by Liu (1995): the r chart, Q chart [not to be confused with Quesenberry's (1991, 1993) developments nor with the *Q* statistic from Jackson (1991)] and the s chart. These distribution free control charts are multivariate analogues to the *X* chart, $\overline{X}$ chart and CUSUM chart. Liu (1995) describes her control chart scheme as follows,

> "**The main idea behind our control charts is to reduce each multivariate measurement to a univariate index – namely, its relative center-outward ranking by data depth**."

In multivariate data analysis, elliptical and ellipsoid contours may be used to represent visually bivariate and trivariate data, respectively. However, in higher dimensionality, graphical analyses are not feasible, thus a graphical scheme to represent higher dimensionality is desirable, specifically in industrial applications. Data depth is advantageous, since it visually represents depth when *p* is greater than or equal 2 while it can still be plotted as a univariate index. Liu's approach is to convert multivariate data to univariate indices using the ranks of the simplicial data depths. Since the control charts are based on ranking these depths with no distributional assumptions, this approach is nonparametric. Liu's approach is affine invariant which is critical for our proposed distribution free approach given that PCA will be a significant focus, whereby affine invariance will be needed for the transformations. Given the ease of computation, Liu (2003) indicates that for near elliptical distributions, the Mahalanobis data depth measure may be used in her control chart scheme as did Zarate (2003), otherwise simplicial depth will "reflect more accurately the underlying probabilistic geometry" and no moment conditions are required.

Liu's (1995) description of the r-chart scheme begins with letting G denote the prescribed quality distribution with p process variables with $Y_1, \ldots, Y_n$ random observations. In other words for MSPC, $Y_1, \ldots, Y_n$ will represent the reference in-control data set or HDS. Collect a sample of new observations $X_1, \ldots, X_t$, and assume that its distribution is called F (note: not the F-distribution). To determine if the process has gone out of control, compare the sample of new observations $X_1, \ldots, X_t \sim F$ against the reference sample $Y_1, \ldots, Y_n \sim G$.

$H_0 : G = F$ with a false alarm rate of α

H_1 : There is a shift in location and/or scale increase from G to F.

Compute the ranks of the depths $Y \sim G$. As the rank decreases, the more outlying is the point within that distribution.

$$r_G(y) = P\{D_G(Y) \le D_G(y) | Y \sim G\}$$

$$r_{G_n}(y) = \frac{\#\{Y_j | D_{G_n}(Y_j) \le D_{G_n}(y),\ j = 1,\ldots,n\}}{n}$$

The assumption is that distribution $D_G(X)$ is continuous and that $r_G(X)$ converges to a uniform distribution ~U[0,1], as such the expected value is 0.5 which will serve as the CL (center line). The uniform convergence is illustrated by Liu and Singh (1993). (See APPENDIX B) A trend towards degradation occurs when the quality characteristics are not converging to the distribution, in other words when the value starts falling below the expected value of 0.5. As such there is no UCL (Upper Control Limit) but only a LCL (Lower Control Limit). Based on the uniform convergence that Liu stipulates for the ranks of the simplicial depths, the center line of the *r*-chart is 0.5. The r chart values are obtained by plotting each rank $r_G(X_i)$ against time *i* and can be used to monitor both "location shifts and scale increases." The smaller the *r* value, the more outlying is the point from the data cloud which may signal a shift from the reference sample (HDS) distribution. As a result of the uniform convergence of $r_G(X)$, Liu (1995) points out that α, which represents the false alarm rate, is the lower control limit (LCL) for this chart. The process will be declared out-of-control when $r_G(X_i) < \alpha$, in other words when the ranks are below the LCL which is α. Given this plotting scheme the chart is a univariate chart which is easy to read and interpret.

Liu (1995) also proposes the nonparametric multivariate analogue to the univariate $\overline{X}$ chart, which she calls the Q-chart. This chart is also referenced in the literature as the SD or Simplicial Depth Chart. (Stoumbos, et. al. 2001) For consistency, throughout the rest of this paper we will identify this chart as the Q-chart so as to avoid any confusion with Liu's r chart which is also based on simplicial depth. Liu's Q-chart represents the average relative ranks as:

$$Q(G,F) = P\{D_G(Y) \le D_G(X) | Y \sim G, X \sim F\}$$

$= E_F[r_G(X)]$ which is the expected (or average) value of the rank.

$Q(G,F_c) = \dfrac{\sum_{i=1}^{c} r_G(X_i)}{c}$ or the empirical case $Q(G_n,F_c) = \dfrac{\sum_{i=1}^{c} r_{G_n}(X_i)}{c}$ where c represents the number of c samples collected.

The averages $Q(G,F_c^i)$ or $Q(G_n,F_c^i)$ are plotted against c. As with the r chart, the center line of 0.5 is used as a reference line to detect any trends towards quality degradation. The lower control limit is

$$\text{LCL} = 0.5 - z_{\alpha}\frac{1}{\sqrt{12c}} \text{ for } Q\left(G, F_c^i\right) \text{ or}$$

$$\text{LCL} = 0.5 - z_{\alpha}\sqrt{\frac{1}{12}\left(\frac{1}{n}+\frac{1}{c}\right)} \text{ for } Q\left(G_n, F_c^i\right)$$

Unfortunately, the issue of a reference sample size is unresolved. (Stoumbos, et. al. 2001) Liu (1990) claims that this control limit (either case) holds true if $c \geq 5$, if not for c < 5, she recommends LCL is $\frac{(c!\alpha)^{\frac{1}{c}}}{c}$. Liu's Control charts are beneficial in quality control given that the multivariate structure is reduced to univariate index, which is an easy to plot univariate control chart scheme. Since they are based on the ranks of the data depth, there are no distributional assumptions or requirements of independence.

Numerous recent MSPC developments by Dai, Zhou and Wang (2004), Zarate (2003) and Messaoud, Weihs and Hering (2004) utilize data depth to construct nonparametric multivariate control charts. The control charts proposed by Dai, Zhou and Wang (2004) and Messaoud, Weihs and Hering (2004) are multivariate CUSUM and multivariate EWMA respectively, in which the observations are based on data depth measures of all the process variables as is the case with the Liu (1995) control charts. Zarate (2003) extended Liu's work by utilizing PCA first to reduce the dimensionality and focus on the principal components that drove the process. After reducing the dimensionality, Zarate proposed to compute the data depth ranks of the PCs and plot those ranks on the nonparametric r-chart which was presented by Liu (1995). Zarate computed the data depth ranks using the Mahalanobis data depth measure which is easier to compute. Thus, the construction of Zarate's control chart is based on the measure of depth by the mean versus Liu's measure of depth by location or the median using simplices. The Mahalanobis data depth is affine invariant which is necessary for PCA, but as Liu (2003) has indicated is a less robust measure of data depth. Additionally, Rousseeuw and Leroy (1987, 1990) demonstrated that in the presence of more than one outlier, the Mahalanobis distances suffer from the "masking" effect.

2.8 Summary

Hotelling's test statistic is the statistic referenced extensively in the multivariate quality literature. The T^2 statistic is used in multivariate quality to establish the Upper Control Limit for the MSPC control chart. Unfortunately, Hotelling's approach has a significant limitation, in that the underlying assumption is that the distribution of the multivariate process must be normal. Thus, if the distribution is multivariate non-normal, a multivariate control chart based on Hotelling's T^2 would be less powerful and generate higher false alarms. The following flowcharts are provided to illustrate the current normal MSPC approach using T^2. The first, Figure 2.8a, is high level and illustrates how a practitioner would traditionally start approaching a multivariate SPC problem (by determining whether the data are normal or not). Once the decision about normality is made, a distinct path is chosen as demonstrated in Figure 2.8b. Additionally Table 2.8a lists the advantages and disadvantages of the traditional T^2 approach in MSPC.

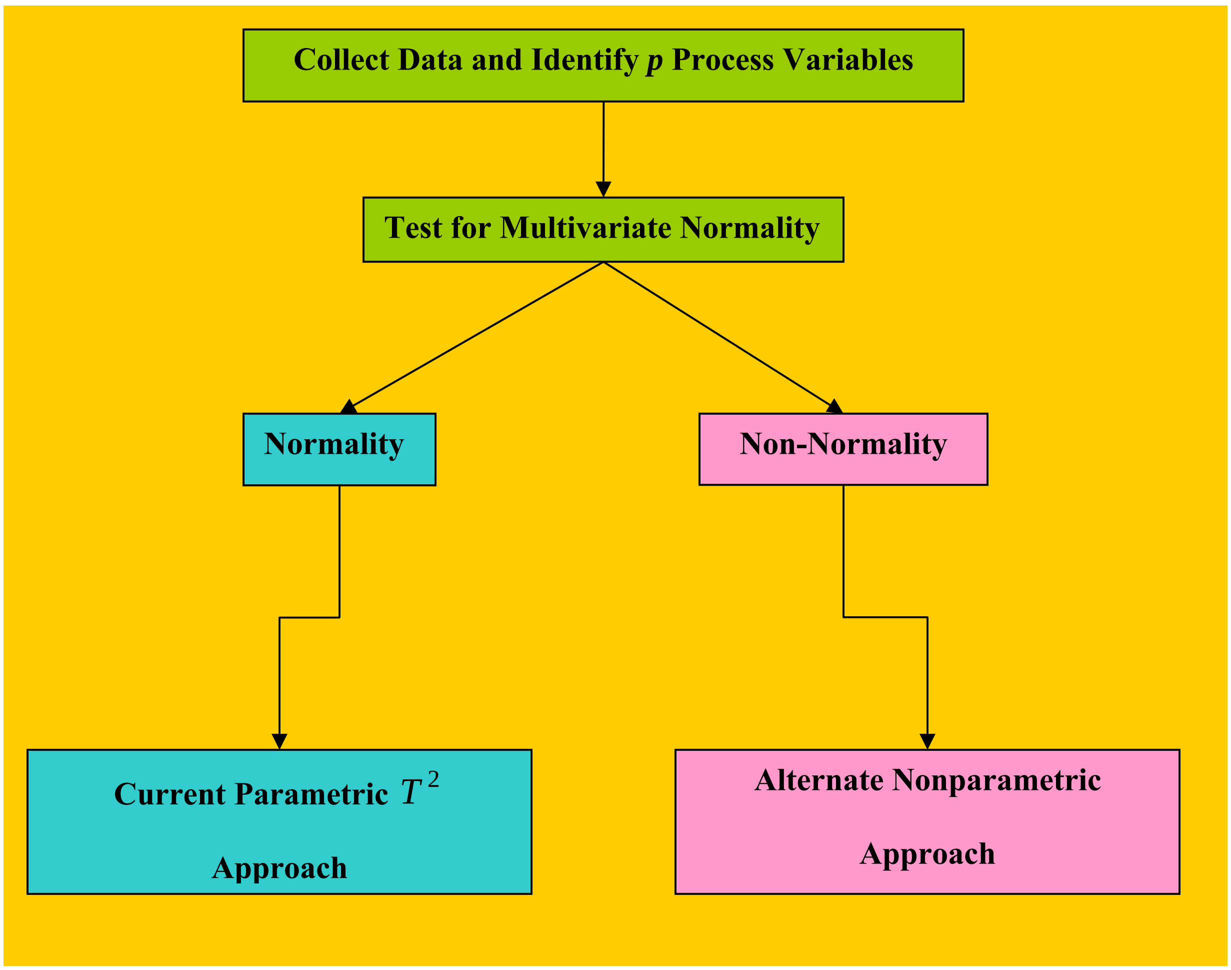

Figure 2.8a Flowchart with the normality path versus the non-normality path.

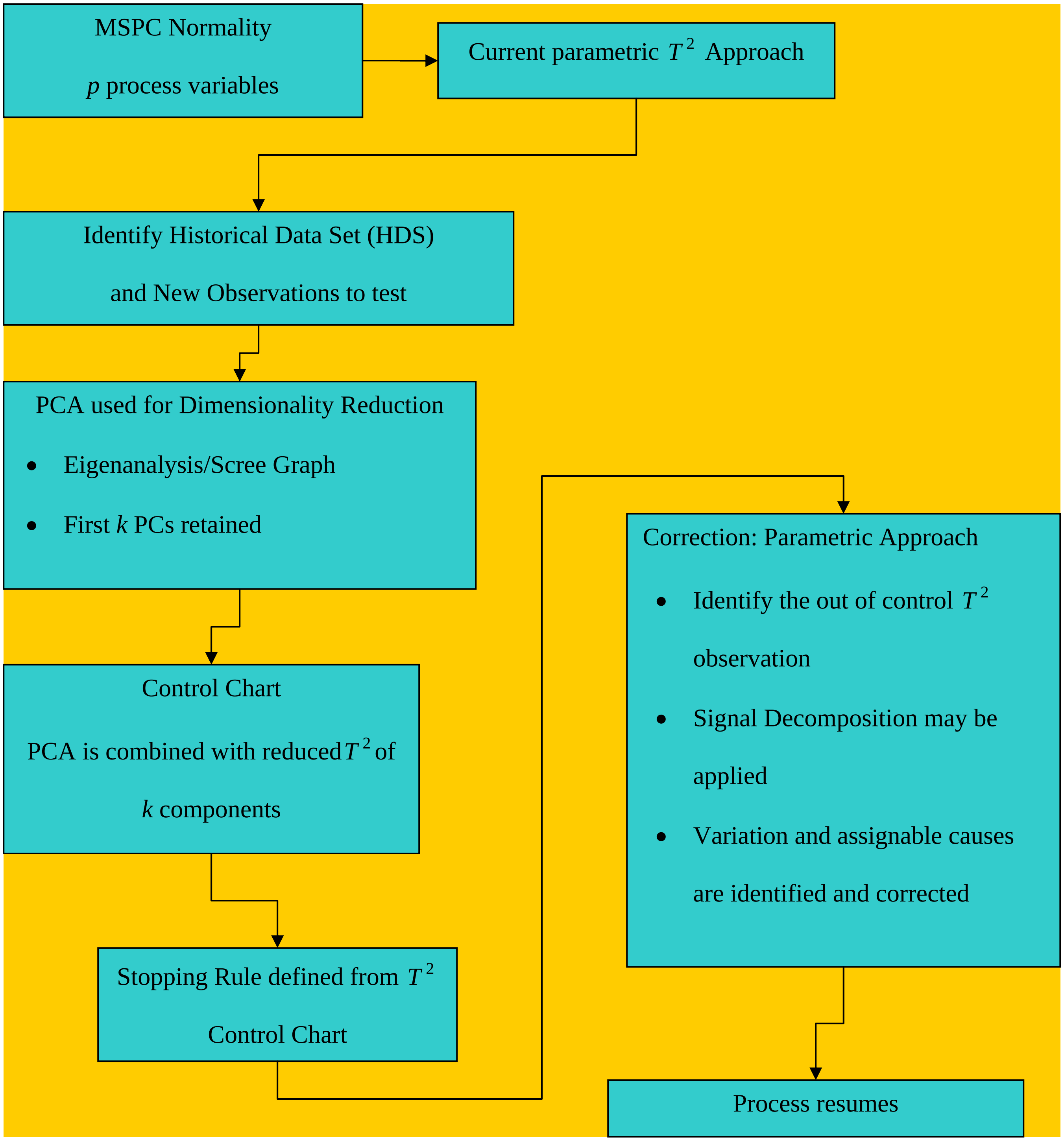

Figure 2.8b Flowchart with the Traditional PCA MSPC Parametric Scheme based on T^2

Traditional T^2 Approach in Multivariate Quality Control

Table 2.8a. Traditional approach in MSPC

Advantages	Disadvantages
• Well-known in the theoretical multivariate analysis area • Well-known in the multivariate quality area; used in industry • Affine-invariant statistic • Can be plotted in univariate style which is easy to visualize and understand • PCA is used to reduce dimensionality and control charts of the components based on the T^2 can be constructed • Signals are detected from the control charts. Assignable causes are identified via a decomposition of T^2 • Stepwise decomposition to detect and "correct" the signal • Readily available in most statistical software packages	• Limited to Multivariate Normal distribution • Assumes independence • Autocorrelation not allowed • Loss of power in the presence of multivariate non-normality • Higher incidence of false alarms in non-normal MSPC • False alarms are costly in industry • Distribution-free property of PCA is not utilized • When PCA is based on Traditional T^2 the approach will be dependent on multivariate normality • Adaptations of T^2 to detect assignable causes depend on normality

In recent publications, it has been the position of various authors such as Coleman (1997), Nester (1996) and Box and Luceño (1997) that normality cannot exist in practice. These authors stress that recent developments on multivariate control charts still have a "potential drawback" due to the fact "that they are based on the assumption of multivariate normal process data." Coleman (1997) strongly believes that in industry the normality assumption is unbelievable, therefore as he has stated "distribution free multivariate SPC is what we need" to remove the normality assumption required in current methods. Given the successful applications throughout the years of control charts assuming normality, we cannot dismiss normality altogether. We do need to recognize how infeasible it is in certain industries with multivariate processes, in which correlations between variables and/or autocorrelations exist as indicated by these authors. Also, in the past there was not such an abundance of available data as there is today with the advent of on-line data which generate more process variables yet with small samples, and as indicated by Kapatou (1996), in quality control it is realistic to deal with small samples. The significance of these authors' contention with normality is the unrealistic assumption that these industrial processes with multiple variables can be classified as normal. As indicated by various researchers, the old fashioned approach served its purpose well and will continue to do so in certain industrial situations, however limiting ourselves to just these methods would allow high levels of false or missed alarms to occur in those industrial settings that are determined by non-normal multivariate data.

As Elsayed (2000) suggested, there is a tremendous need for improvement in the area of SPC in industries such as the food, chemical, automotive and manufacturing industries, given that these industries inherently deal with numerous variables which are highly correlated. An industrial process could easily consist of a substantial number of variables where inter-correlations and autocorrelations are quite common, compounded with possible non-normality. Therefore, current parametric techniques in MSPC would be inappropriate and ineffective in such non-normal, correlated processes. Utilizing MSPC techniques that assume multivariate normality on a non-normal multivariate process would generate false alarms, yielding unnecessary corrections of a process which are costly, or missed alarms which in some of the aforementioned processes could be hazardous. (Mason and Young, 2002)

Stoumbos, et.al. (2001) also stress that there is a strong need to find simple and efficient schemes in quality control that will detect shifts without producing high levels of false alarm rates. Stoumbos, et.al.'s paper(2001), along with Elsayed (2000), have also stressed that the abundance of available data no longer support these assumptions of normality and independence. Multivariate data that is non-normal and highly correlated requires a distribution free approach that will dismiss the normality requirement as well as promote a stronger detection of a correlation shift that might lead to the overall quality loss in the system. The need to research new and efficient processes has generated some development in nonparametric SPC but with limitations. According to Stoumbos, et.al. (2001), there are many "unresolved issues" with some current approaches to nonparametric control charting.

One of these recent developments has been the Hayter and Tsui (1994) proposal called the M procedure, which is a nonparametric scheme. The limitation to this M procedure is that it ignores correlation among multivariate components. The literature in multivariate quality control emphasizes that all, if not most, of the *p* process variables in industrial settings are correlated. As such this procedure will not monitor effectively the intercorrelation of multivariate factors in the food and chemical industries among others.

Kapatou (1996) has developed nonparametric multivariate charts that are quite robust, but as Stoumbos, et. al. (2001) indicated Kapatou's multivariate charts require estimation of nuisance parameters related to process covariance structure. Additionally, Kapatou's nonparametric control charts are based on multivariate test statistics which are not affine invariant. The lack of rotatability implies that charts based on non-affine invariant statistics are powerful at detecting shifts in some directions while not in others, which is further compounded in situations with higher dimensionality. Additionally, with the lack of affine invariance, dimensionality reduction is more challenging. As such the PCA technique which uses scaling and rotatability to reduce dimensionality in multivariate data could not be applied. The Liu approach is significant given its robust distribution free approach in MSPC along with its ease of use. Given the univariate scheme that is used for plotting them, Liu's charts are easy to plot and implement in an industrial setting. Liu's nonparametric approach (1995) of using simplicial depth based on multivariate generalizations of univariate Shewhart control charts still has unresolved the issue of reference sample size requirement. (Stoumbos, et. al., 2001) Additionally, Liu's control charts, along with the nonparametric multivariate CUSUM and multivariate EWMA control charts proposed by Dai, Zhou and Wang (2004) and Messaoud, Weihs and Hering (2004), respectively, are based on full dimensionality, and in the multivariate quality control literature, Zhang (2003) and Zarate (2003) pointed out that typically a process is driven mostly by a subset of all process variables. Also, since Liu's charts are based on the *p* dimensionality, as *p* and/or *n*, the reference sample size, increase, the computation of the simplicial depth is more challenging and time consuming.

Given the simplicity of the univariate scheme used for plotting the Liu charts and the well established dimensionality reduction scheme of PCA, Zarate (2003) extended Liu's work by performing PCA prior to computing the ranks for the r chart. As such the ranks that were plotted by Zarate were the data depth ranks of the principal components that drive the process versus the ranks of all process variables. The limitation to Zarate's approach is that her data depth ranks were computed using the Mahalanobis data depth which is known to be non-robust as indicated by Liu (2003) and Rousseeuw (1990). Given the non-robustness of this approach, the incidence of false alarms may be higher. Overcorrection is a serious problem since it leads to waste which is costly, and lost signals are also extremely costly in terms of hazardous risks. In areas such as the automotive industry and manufacturing among others, a robust approach can tremendously impact and reduce the current waste and inefficiencies caused by false alarms that currently plague these industries. Additionally, higher incidence of missed alarms can prove to be even more costly and dangerous. The aforementioned nonparametric developments as well as their advantages and limitations are listed in Table 2.8b.

Table 2.8b Developments in Nonparametric MSPC

Control Chart	Advantages	Disadvantages
M nonparametric control scheme ***Hayter and Tsui***	Monitors process-location parameter	Ignores correlation among multivariate components
Vector of Winsorized Ranks control chart ***Kapatou***	Utilizes developments from the theoretical multivariate analysis area Robust process in the non-normal multivariate quality area	Multivariate Winsorized Rank Statistics used are not affine-invariant statistics PCA cannot be applied to non-affine invariant statistics Correlation shifts are not detected Not readily available in software packages
Simplicial Data Depth control charts **r, Q and s-charts** ***Liu*** **Nonparametric MCUSUM** ***Dai, Zhou and Wang*** **Nonparametric MEWMA** ***Messaoud, Weihs and Hering***	Reduces each multivariate measurement to a univariate ranking index Easy to visualize Simplicial depth is robust and affine invariant Simultaneously detects changes in process location and/or scale.	Reference sample size requirement is unresolved Full dimensionality Correlation shifts are not detected

PCA Mahalanobis Data Depth r-charts ***Zarate***	Reduces each multivariate measurement to a univariate ranking index Focuses on principal components that drive the process Easy to visualize Simultaneously detects changes in process location and/or scale.	Reference sample size requirement is unresolved Mahalanobis data depth is non-robust Correlation shifts are not detected

In the quality literature numerous advances in nonparametric univariate SPC have been cited by Chakrati, et.al. (2001). However, limited research has been conducted on developing multivariate nonparametric control chart as indicated by experts such as Elsayed, Woodall, Stoumbos, Liu, Jones, Reynolds, Kapatou and more recently Zarate in the MSPC literature. Given the increasing demand to develop distribution free techniques for MSPC and the limitations of previous attempts, multivariate nonparametric approaches need to be further researched.

CHAPTER THREE: METHODOLOGY

3.1 Overall Proposed Plan in MSPC

The goal will be to create a systematic distribution-free approach by extending current developments and focusing on dimensionality reduction using PCA. In multivariate quality control PCA has gained much attention and acceptance in industrial applications within the last decade as suggested by numerous researchers in MSPC such as Kourti and MacGregor (1995, 1996, 2005), Kourti (2003), Zarate (2003) and Milectic (2004). We propose to use PCA which is distribution free in nature to reduce dimensionality by using those principal components that drive the process. The proposed technique is different from current approaches in that it creates a robust affine invariant distribution free approach to improve signal detection for both outliers and shifts in correlation in multivariate quality control. The notion of simplicial data depth will be applied to the principal components and the ranks of these depths will be plotted in a univariate graphing scheme as proposed by Liu (1995).

For our control chart, simplicial depth is more logical approach given its robustness versus the mean which is more sensitive to outliers. The proposed technique is easy to use and robust to outliers. Additionally, our scheme is different given that our approach will focus specifically on the set of components that drive the process and define the correlation structure versus current data depth control charts developed by Liu (1995) Dai, Zhou and Wang (2004) and Messaoud, Weihs and Hering (2004) which are based on all of the multivariate process variables. Our control charts for both the first and last PCs will monitor signals that may be attributed to shifts in the variability and correlation structure, respectively. Our proposed PCA Simplicial Depth r chart is significantly different from the reduced dimensionality data depth scheme developed by Zarate (2003). Our simplicial depth approach is robust, while the Mahalanobis depth approach used by Zarate is non-robust. Also, we propose a correlation monitoring scheme whereas Zarate did not monitor correlation shifts. Table 3.1a illustrates the significant differences between our PCA Simplicial Depth r chart versus current data depth control charts.

Table 3.1a Competing Nonparametric MSPC Data Depth r charts

Liu **Simplicial Depth** **r Control Chart**	**Zarate** **PCA-Mahalanobis Depth** **r Control Chart**	**Proposed** **PCA-Simplicial Depth r** **Control Chart**
Full Dimensionality	**Reduced Dimensionality** **Apply PCA** Determine the k principal components that account for **variability** (at most 85%)	**Reduced Dimensionality** **Apply PCA** Determine the k principal components that account for **variability** (at most 60%) ***AND*** ***last PC(s) which may account for correlation structure*** (at most 0.009)
Computes the ***simplicial data depths*** of each multivariate observation using full dimensionality	Computes the ***Mahalanobis data depths*** of each multivariate observation from the first set of PCs that contribute to the variability	Computes the ***simplicial data depths*** of each multivariate observation from the first set of PCs that contribute to the variability ***AND*** ***the last PC(s) to account for correlation***
Computes the ***ranks of the simplicial data depths*** of the observations	Computes the ***ranks of the Mahalanobis data depths from the first PC(s)***	Computes the ***ranks of the simplicial data depths from the first PC(s)*** ***AND*** ***the last PC(s)***
Plots the ranks of each observation in a univariate scheme	Plots the ranks of each PC in a univariate scheme	Plots the ranks of each PC in a univariate scheme
LCL is α No UCL CL is 0.5 (reference line)	LCL is α No UCL CL is 0. 5 (reference line)	LCL is α No UCL CL is 0. 5 (reference line)

Our goal is that our proposed nonparametric scheme is robust, and consequently there is a single path for all MSPC applications as demonstrated in Figure 3.1a. Another goal is adaptability of future nonparametric developments from the multivariate theoretical statistical area.

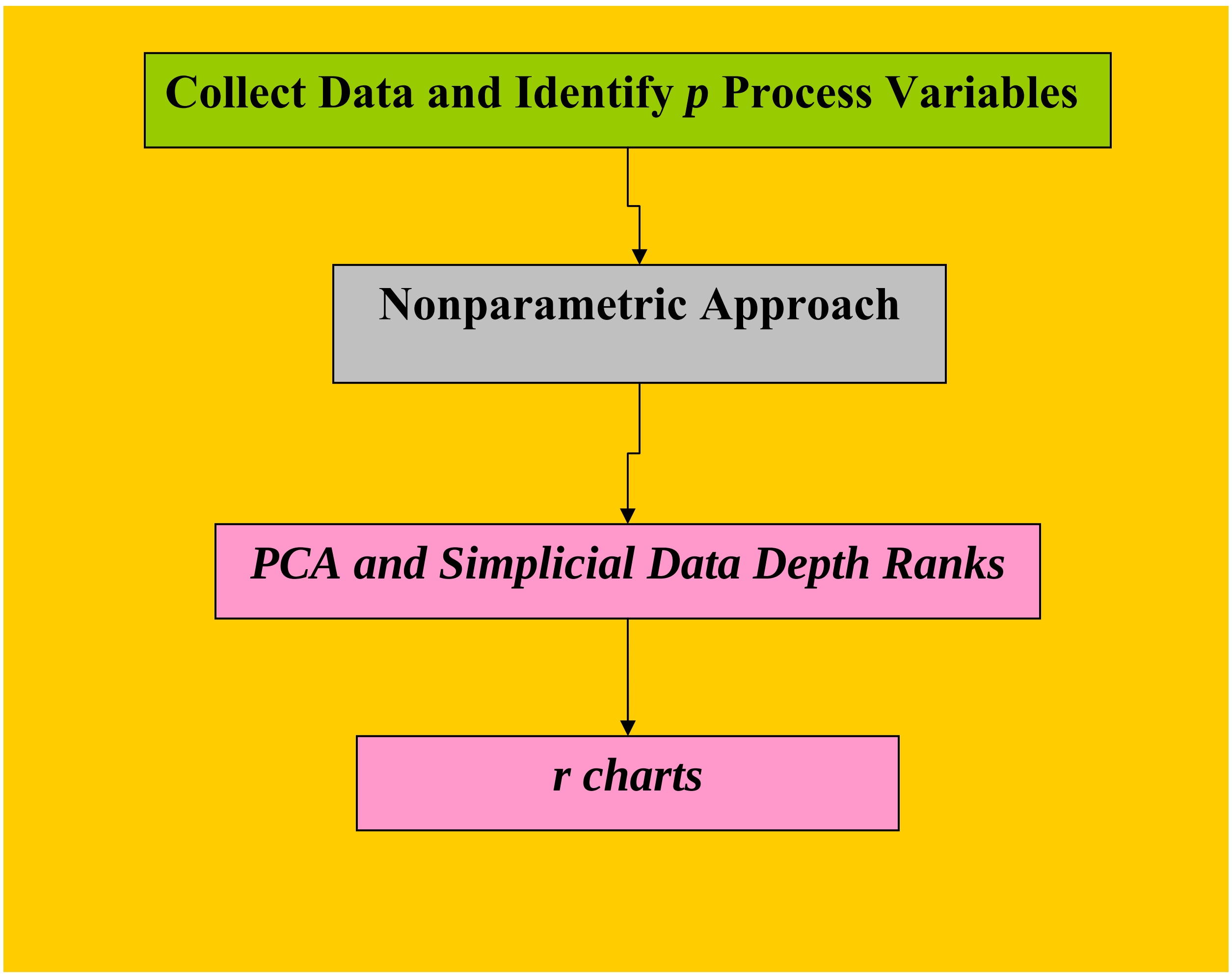

Figure 3.1a Flowchart with the single path implies no distribution is assumed.

3.2 Distribution Free Techniques for Problem Correction

Although there are statistically sound nonparametric techniques that have been developed, there is no multivariate nonparametric integrated approach from start to finish. We apply or adapt proven theoretical multivariate statistical techniques along with established quality techniques to improve the signal identification in multivariate quality control in non-normal processes. A different adaptation of PCA, which is a distribution free dimensionality reduction scheme for multivariate data, along with the robust and geometrically affine invariant simplicial data depth measure will be a major component of our research. We need no distributional or independence assumption as illustrated in Figure 3.1a in the previous section.

PCA has been well received in multivariate industrial applications such as the manufacturing, automotive, chemical and food industries. (Kourti and MacGregor, 2005). PCA is referenced in the multivariate quality literature in conjunction with the T^2 test statistic which assumes multivariate normality. Kourti and MacGregor (1995, 2005) have indicated that along with dimensionality reduction, PCA can in fact be used to provide diagnostics on how to remedy the process. Unfortunately, their projection methods which include PCA are based on adaptations of the T^2 test statistic. We utilize these two distribution free techniques together in a univariate r chart scheme, which is different from the full p dimensional r chart proposed by Liu (1995) and the PCA-Mahalanobis r chart proposed by Zarate (2003). The ease of use and graphical interpretation as well as the availability of PCA transformations in statistical software make this adaptation readily available to any industrial setting.

Simplicial depth, which is a theoretically sound distribution free measure from topology (Munkres, 1975) and convex geometry (Magaril–Il'yaev and Tikhomirov, 2003) (Andersson, et.al., 2004) has been utilized by Liu (1995) in MSPC. In Chapter 2, we described Liu's (1990) definition of data depth. The application of data depth in multivariate data is to measure the depth of a point within a data cloud. The measure represents the number of simplices that contain that specific observation. As we illustrated in Chapter 2, simplicial depth is easy to understand and in low dimensionality is easy to visualize. In Chapter 2, we provided the mathematical definition and formulas for simplicial depth. In this section, we provide a graphical representation of the univariate (p = 1) and bivariate (p = 2) cases as an illustration of the computations that are used in Chapter 4. The n observations must be at least p + 1 with each simplex defined by p + 1 endpoints.

For p = 1, the simplex p + 1 = 2 would be a line segment with $n \geq 2$. The simplicial depth of a point x would be the proportion of segments that contain x. $D_n(x) = \binom{n}{2}^{-1} \sum_{1 \leq i < j < k \leq n} I\left(x \in X_i X_j\right)$ where I represents the indicator function which is the number of segments that contain x.

For our illustration, we will let n = 4, and the points will be A, B, C and X. The number of the possible segments formed would be $\binom{n}{2} = \binom{4}{2} = 6$, and the segments would be formed by AX, BX, CX, AB, BC and AC.

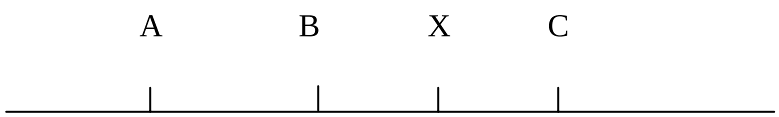

Figure 3.2a Simplicies for p = 1

We see that point X is contained within 2 segments, AC and BC. Therefore, the simplicial depth of X is $\frac{2}{6} \approx 0.333$. Point B is also contained within 2 segments, AX and AC. Therefore, the simplicial depth of B is also $\frac{2}{6} \approx 0.333$. Points A and C are not within any of the segments, therefore the simplicial depths for those two observations are $\frac{0}{6} = 0$.

For p = 2, the simplex p + 1 = 2 would be a triangle with $n \geq 3$. The simplicial depth of a point x would be the proportion of triangles that contain x. The number of the possible triangles formed if n = 4 would be $\binom{n}{3} = \binom{4}{3} = 4$, and the triangles would be formed by AXB, ABX, ABC and ACX.

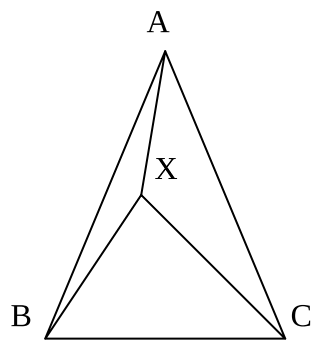

Figure 3.2b Simplicies for p = 2

In this case the three outer points, A, B and C, are not contained within any triangle, therefore those 3 simplicial depths are zero. Point X lies within 1 triangle ABC. Therefore the simplicial depth of X is $\frac{1}{4} = 0.25$.

From the three dimensional case when p = 3, each simplex would be a tetrahedron. Beyond p = 3, it is difficult to visualize the simplicial depth, however, since the depth measure is univariate, it is easy to plot. In MSPC, Liu's control charts plot the ranks of the simplicial depths which are univariate and easy to visualize. From the two previous examples, we see that when a point is located within the center of the simplices of the data cloud, its depth will be higher. We see that location is a distribution free measure. When applying this concept to MSPC, Liu (1995) identified a "deep point" as more in control. This implies that a higher value is better so there is no upper control limit when using a data depth control chart scheme. The out of control observation will have a small or zero simplicial depth.

We use simplicial data depth and PCA, which are both distribution free and affine invariant, to detect the out of control signal(s). Additionally, we investigate if the first few principal components and the last component(s) along with eigenvalue analysis can be used for signal detection and provide insight into both the variability and correlation structure. Since, the first few components account for the majority of the variability, additional analysis of the last few principal components is conducted to gain insight into the correlation structure. (Dunteman, 1989) We provide the cumulative percentages of variation from the eigenanalysis and determine which initial set of PCs will provide the best insight into the variability shift. These comparisons illustrate which set of PCs may be desirable based on the sample sizes of the HDS and the corresponding alpha values. The proposed technique is better based on its ease of use and its robustness to outliers in MSPC.

3.3 Proposed PCA-Simplical Data Depth r-chart

We propose computing the ranks of the data depths of the PCs using the robust simplicial data depth measures described in Section 3.2. As such, our proposed control chart is a PCA-Simplicial Depth r chart which will focus on the principal components that drive the process, and compute robust ranks of those components. These ranks will be plotted using the Liu (1995) and Zarate (2003) univariate style plot for ease of use, but we believe that our approach is better given that the focus will be on the principal components of the process, and the ranks used will be computed from robust measures. As previously stated, a significant difference between our approach and the Zarate (2003) control chart scheme will be in our use of the last PC(s) to identify any process degradation that may have been caused by a shift in the correlation of the process variables. (See Table 3.1a)

In the literature, we also illustrated that current developments in MSPC utilize PCA to reduce dimensionality and proceed to construct a control chart with a parametric approach, namely the T^2 or adaptations of the T^2. The nonparametric approach that is proposed will focus more heavily on PCA and on the robust simplicial data depth measure in a completely different manner than the current literature suggests. One of the challenges of multivariate SPC is that there are two ways of producing signals: “moving a particular variable’s observed value beyond its operational range and/or contaminating a linear relationship between two or more process variables.” (Mason and Young, 2000)

This statement in the literature has motivated us to develop a monitoring scheme that may be used to identify both variability and correlation shifts in multivariate processes. Our robust nonparametric scheme will improve signal detection when there has been process degradation due to a variability and/or correlation shift in MSPC. Once the process has been stopped and the signal has been identified, our distribution free control charts will provide the user a diagnosis of the problem along with corrective measures in order to subsequently restart the process.

Rencher (1995) points out that PCA can be applied to any distribution of the original process variables. Zarate (2003) used PCA for her distribution free approach which utilizes ranks of data depth measures for monitoring the variability, whereas our PCA data depth scheme will utilize the robust simplicial data depth measure and plot a PCA-Simplicial Depth r chart of both the first set of PCs to monitor variability shifts and the last PC(s) as well in order to monitor correlation shifts. Kourti and MacGregor (1995, 1996, 2005) have indicated how projection methods, which utilize PCA for diagnostics in multivariate quality control, are becoming widely used in industrial applications, however their work with PCA continues to use an adaptation of the T^2 statistic. The current PCA approach in multivariate quality is to utilize PCA with the T^2 statistic or some adapted form of the T^2 statistic. Additionally, Mason and Young (1997, 2002) whose contributions to MSPC have been extensive within the last decade, focus on using PCA for their decomposition of the T^2 statistic. PCA based control charting is gaining significant attention in MSPC (Kourti and MacGregor 1995, 1996,

2005) (Kourti, 2003) (Milectic, 2004). As Milectic, et. al. (2004) demonstrated when utilizing MSPC in their study, their challenge has been “to develop a monitoring scheme that would alarm reliably” and would be easy to present to the process operator. Our nonparametric scheme is not based on the T^2 statistic, since we utilize the distribution free property of PCA along with simplicial data depth throughout MSPC to avoid the pitfalls of current normality dependent schemes when the distribution is unknown or when autocorrelation is present. Figure 3.3a illustrates our distribution free MSPC approach from start to finish.

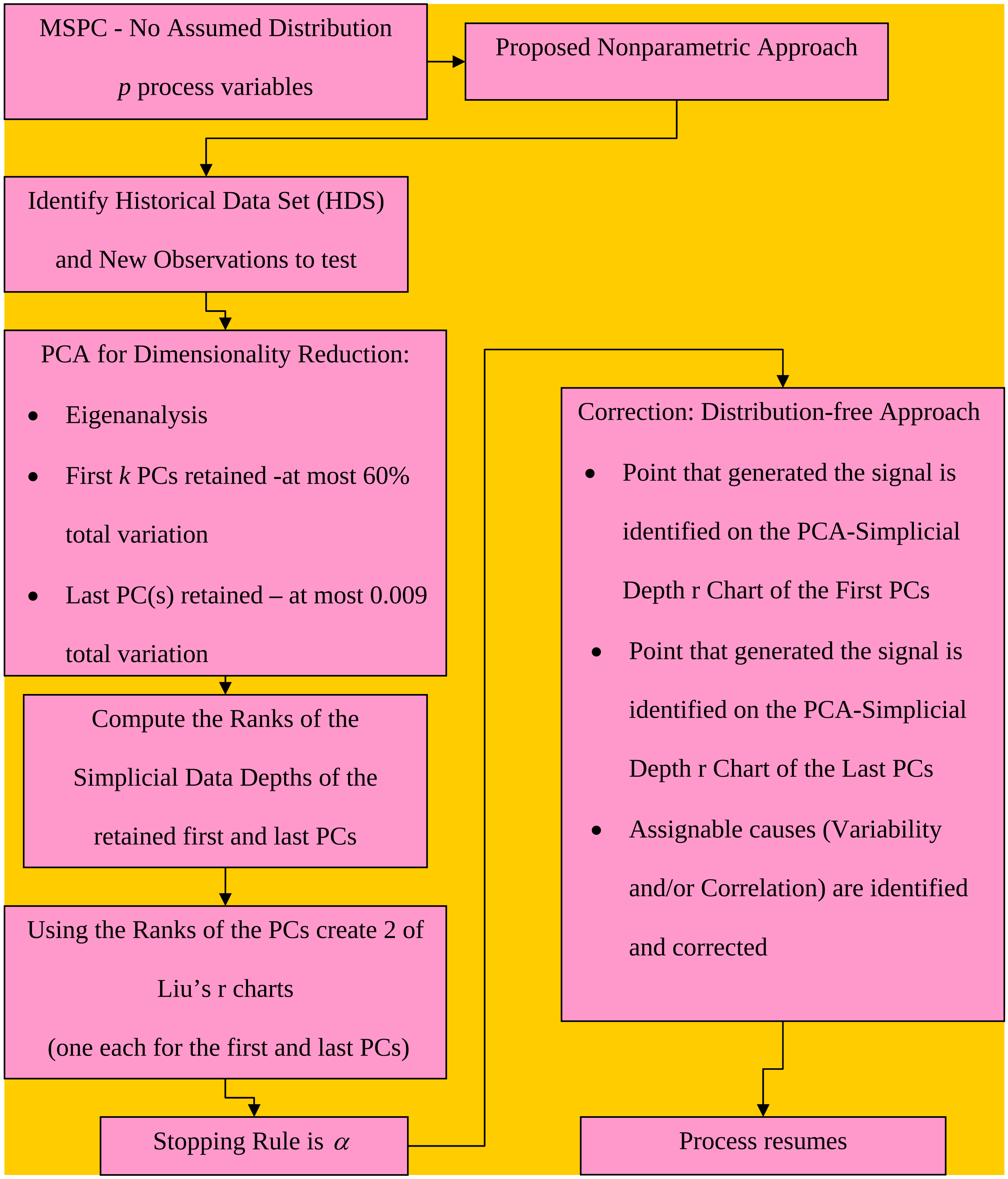

Figure 3.3a Flowchart with the Proposed Nonparametric MSPC Scheme

We begin by collecting the HDS (Historical Data Set) with all process variables identified. We let G denote the prescribed quality distribution with p process variables with $Y_1, \ldots, Y_n$ random observations, and perform PCA on the HDS. An eigenvalue analysis is done to identify the $k < p$ principal components that seem to drive the process variability and correlation, respectively. We use a cutoff of 60% cumulative variation for the initial PCs and at most 0.009 for the final PCs that will be retained. We calculate the simplicial data depths of the retained initial and final PCs. Next, new observations $X_1, \ldots, X_t$ are collected. We denote the distribution of the sample as F (note: not the F-distribution) similar to Liu (1995). We standardize the new observations, and from the eigenvectors of the HDS, we compute the PC score of each new observation. We compute the simplicial depths of the retained PCs of the new observations based on the retained PCs from the HDS. This is followed by the calculations of the ranks of the simplicial depths of these new observations. To determine if the process has gone out of control we will compare the simplicial depths of the PCs of our test sample $X_1, \ldots, X_t \sim F$ against the simplicial depths of the PCs of our reference sample or HDS $Y_1, \ldots, Y_n \sim G$. The ranks of the simplicial depths of the retained first and last PCs of the new observations are computed as follows:

$$r_{G_n}(X_i) = \frac{\#\left\{Y_j \,\middle|\, D_{G_n}(Y_j) \le D_{G_n}(X_i),\ j = 1, \ldots, n\right\}}{n}$$

The ranks of the simplicial depths of the first and last PCs of the new observations are plotted on each of two r charts, each chart following the univariate scheme presented by

Liu (1995). The higher ranks indicate that the observation is deeper within the data cloud. According to Liu (1995), 0.5 may be considered a centerline to identify any trends toward process degradation. The only control limit is the lower control limit, LCL, and is identified as α. Figure 3.3b represents the univariate graphing scheme presented by Liu (1995) which will be used for our PCA-Simplicial Depth r chart whereby the ranks of the first and last PCs from the test observations are plotted against the observations.

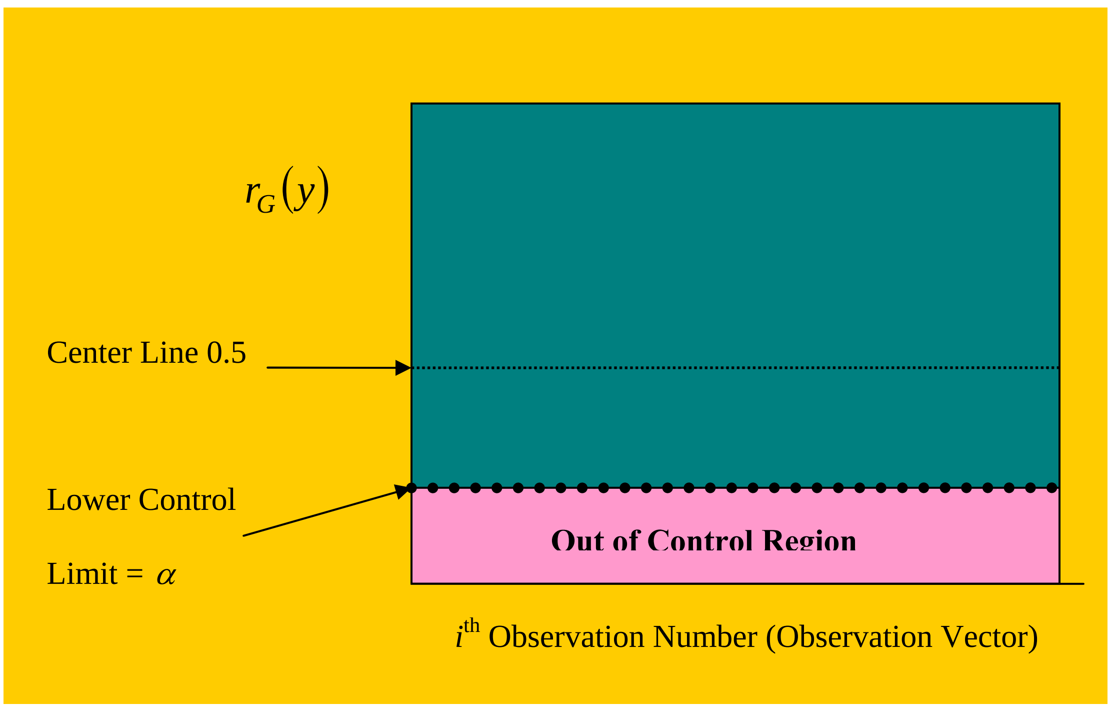

Figure 3.3b Liu r chart (1995)

As the rank decreases, the more outlying is the point within that distribution. Thus, any point lying below the LCL indicates that the observation is an out of control point. For the r chart utilizing the first PC(s) we test:

$H_0 : G = F$ with a false alarm rate of α

$H_1 :$ There is a shift in variability from *G* to *F*.

While for the r chart utilizing the last PC(s), we test:

$H_0 : G = F$ with a false alarm rate of α

$H_1 :$ There is a shift in correlation from *G* to *F*.

In our findings in Chapter 4, we provide comparisons on the effect of choosing different Lower Control Limits determined by either $\alpha = 0.05$ or $\alpha = 0.10$, with various sample sizes. We also discuss the effect of changing the rules governing the retention of PCs.

CHAPTER FOUR: CASE STUDIES AND RESULTS

4.1 MSPC Data Sets

The case studies are based on the following real multivariate data sets from various industries which have been made available to us directly from a company or which are discussed in the literature. The data set titled "Industrial" is made available to us by a former Ph.D. student now working at a company, which will be generically described due to a non-disclosure agreement. The historical data sets are provided as a basis to identify when the process was in a state of statistical control. Additionally, new observations are given to identify when the process was deemed out of control. The following data sets gathered from the multivariate quality literature will be analyzed using our proposed nonparametric scheme by plotting PCA-Simplicial Depth r charts: Steam Turbine Data (Mason and Young, 2002), Fruit Juice Data (Fuchs and Kennet, 1998), Industrial Data (former Ph.D. student), Aluminum Pin Data (Fuchs and Kennet, 1998), Automotive Data (Wade and Woodall, 1993), Electrolyzer Data (Mason and Young, 2002) and Mechanical Part Data (Fuchs and Kennet, 1998). Additionally, we have gathered a bivariate data set from the literature which simulates two of the process variables from the Mechanical Part Data with a random error point identified by Fuchs and Kennet (1998), which will be utilized to test the robustness of our proposed nonparametric scheme in the presence of an outlier that has not been the result of process degradation.

For each data analysis we are including the original observations, both the HDS and the new points, along with the eigenanalysis to provide the cumulative percentage of variation for the PCs from the HDS. The logic behind using PCA is to lower dimensionality and focus on the components that drive the highest variability shift. If too many of the first PCs are selected, a larger cumulative percentage of variability is included, possibly generating too many false alarms. Given that the last PCs account for a very small percentage of variation, we will also analyze the last PC(s) for insight into the correlation structure.

Using the data subsequent to the HDS very conservatively, we held out any point which looked even close to being out of control. We then took the remaining points and added them to the HDS in bunches in order to study the effect of sample size of the HDS on the robustness of the results. We are providing summary tables at the end of each data analysis in order to display the effect of sample size on the first and last set of PCs and the corresponding alpha level. For consistency and ease of understanding, we will follow the same format of tables and figures for all data analyses.

By analyzing case studies from various industries, we were able to develop heuristics that will generate robust results when utilizing our PCA Simplicial Depth r chart. These rules include the suggested total percentage of variability to be considered when selecting which PCs to retain. From the eigenanalysis of each data set, we investigated various scenarios by adjusting the maximum cumulative percentage of variability used as the cutoff for retaining PCs. The different scenarios provided insight into the effect of number of PCs retained, α, and sample size. Our findings indicated that to avoid false alarms and neutralize the effect of autocorrelation, the first set of PCs should account for a maximum of 60% cumulative variability when using the PCA Simplicial Depth r chart for monitoring the variability of the process. For the final PCs, we find that selecting the last PC or the last PCs that account for a maximum 0.009 cumulative variability provide robust results for monitoring correlations. A point that was identified as out of control from the first PCs but not the last could be indicative of a shift in variability. The control chart of the last PC identified points that could represent correlation shifts. A point the signaled on both the first PC control chart and the last PC control chart may indicate that the signal was a result of a shift in both variability and correlation. Given the nature of the bivariate case and based on our findings, the recommendation that we can provide for a bivariate process is to chart the PCA Simplicial Depth r chart of the first and the last PC. At the end of the chapter these recommendations are given along with a complete summary of all data analyses complied into two tables, one with the first PC(s) and the other with last PC, in order to provide insight into our findings.

4.2 Steam Turbine Data Set

The first data set with 6 process variables consists of 44 observations with the first 28 representing the Historical Data Set (HDS) (Mason and Young, 2002) of a Steam Turbine process. Table 4.2a is a list of the 28 observations from the Steam Turbine data set that were identified as the Historical Data Set for this process which is comprised of 6 process variables namely, Fuel, Steam Flow, Steam Temperature, Megawatts, Cooling Temperature and Pressure. (Mason and Young, 2002) The eigenanalysis in Table 4.2b illustrates the cumulative proportion of variation of the PCs from the HDS. The 16 new observations that will be utilized for process monitoring based on the HDS are listed in Table 4.2c, followed by the control charts of the ranks of the Simplicial Depths of the first and last PCs of the new observations. For Table 4.2c, the first column represents the observation number from the set of new observations, while each letter in the second column will be used to name each point to further identify that specific outlying observation when additional runs with a different sample size for the Historical Data Set are used for our proposed control charts.

Historical Data Set (HDS) Original 28 observations

Table 4.2a Historical Data Set of the Steam Turbine Data with n = 28

Obs	Fuel	Steam Flow	Steam Temp	MegaWatts	Cool Temp	Pressure
1	232666	178753	850	20.53	54.1	29.2
2	237813	177645	847	20.55	54.2	29.2
3	240825	177817	848	20.55	54	29.2
4	240244	178839	850	20.57	53.9	29.1
5	239042	177817	849	20.57	53.9	29.2
6	239436	177903	850	20.59	54	29.1
7	234428	177903	848	20.57	53.9	29.2
8	232319	177990	848	20.55	53.7	29.1
9	233370	177903	848	20.48	53.6	29.1
10	237221	178076	850	20.49	53.9	29.1
11	238416	177817	848	20.55	53.9	29.1
12	235607	177817	848	20.55	53.8	29.1
13	241423	177903	847	20.55	53.7	29.1
14	233353	177731	849	20.53	53.6	29.1
15	231324	178753	846	20.64	53.9	29.1
16	243930	187378	844	21.67	53.9	29.1
17	252550	187287	843	21.65	54.2	29.1
18	251166	187745	842	21.67	53.7	29.1
19	252597	188770	841	21.78	53.4	29.1
20	243360	179868	842	20.66	53.7	29.1
21	238771	181389	843	20.81	53.9	29.1
22	239777	181411	841	20.88	54	29.1
23	219664	167330	850	19.08	54.1	29.2
24	228634	176137	846	20.64	54	29.2
25	231514	176029	843	20.24	53.8	29.2
26	235024	176115	846	20.22	53.6	29.2
27	239413	176115	845	20.31	53.7	29.2
28	228795	176201	847	20.24	54.3	29.2

Eigenanalysis of the Steam Turbine Correlation Matrix with *n*=28

Table 4.2b The Eigenanalysis of the HDS with n = 28

	PC1	PC2	PC3	PC4	PC5	PC6
Eigenvalue	3.6939	1.0004	0.7241	0.4045	0.1647	0.0125
Proportion	0.616	0.167	0.121	0.067	0.027	0.002
Cumulative	0.616	0.782	0.903	0.970	0.998	1.000

16 NEW OBSERVATIONS for *n*=28

Table 4.2c The 16 NEW observations for n = 28

Obs	Name	Fuel	SteamFlow	SteamTemp	MegaWatts	CoolTemp	Pressure
1	A1	234953	181678	843	20.84	54.5	29
2	A2	247080	189354	844	20.86	54.4	28.9
3	A3	238323	184419	845	21.1	54.5	28.9
4	A4	248801	189169	843	22.18	54.5	28.9
5	A5	246525	185511	842	21.21	54.6	28.9
6	A6	233215	180409	845	20.75	54.5	29
7	A7	233955	181323	842	20.82	54.6	29
8	A8	238693	181346	844	20.92	54.8	29
9	A9	248048	185307	844	21.15	54.6	29
10	A10	233074	181411	844	20.93	54.5	29
11	A11	242833	186216	844	21.59	54.4	29
12	A12	243950	182147	844	21.37	54.2	29
13	A13	238739	183349	844	21.01	54.3	29
14	A14	251963	188012	850	21.68	54.4	29
15	A15	240058	183372	846	21.15	54.2	29
16	A16	235376	182436	844	20.99	54.3	29

PCA Simplicial Depth r chart for the first PC of the 16 NEW

observations for *n*=28 (α = 0.05 and 0.10)

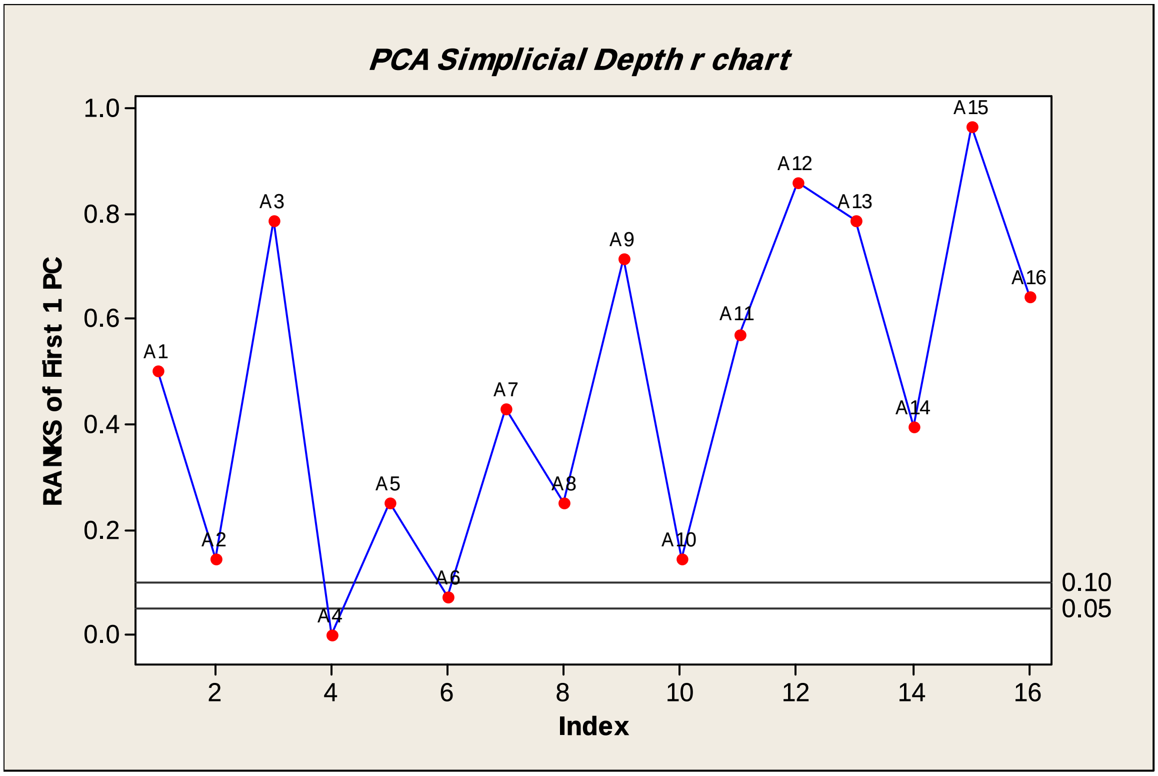

Figure 4.2a PCA Simplicial Depth r chart using the first PC for n = 28.

Using the HDS of n = 28, the following points from the 16 new observations were identified as out of control by the control chart when selecting the first PC:

At α = 0.05: Point A4

At α = 0.10: Points A4 and A6

PCA Simplicial Depth r chart for the last PC of the 16 NEW

observations for *n*=28 (α = 0.05 and 0.10)

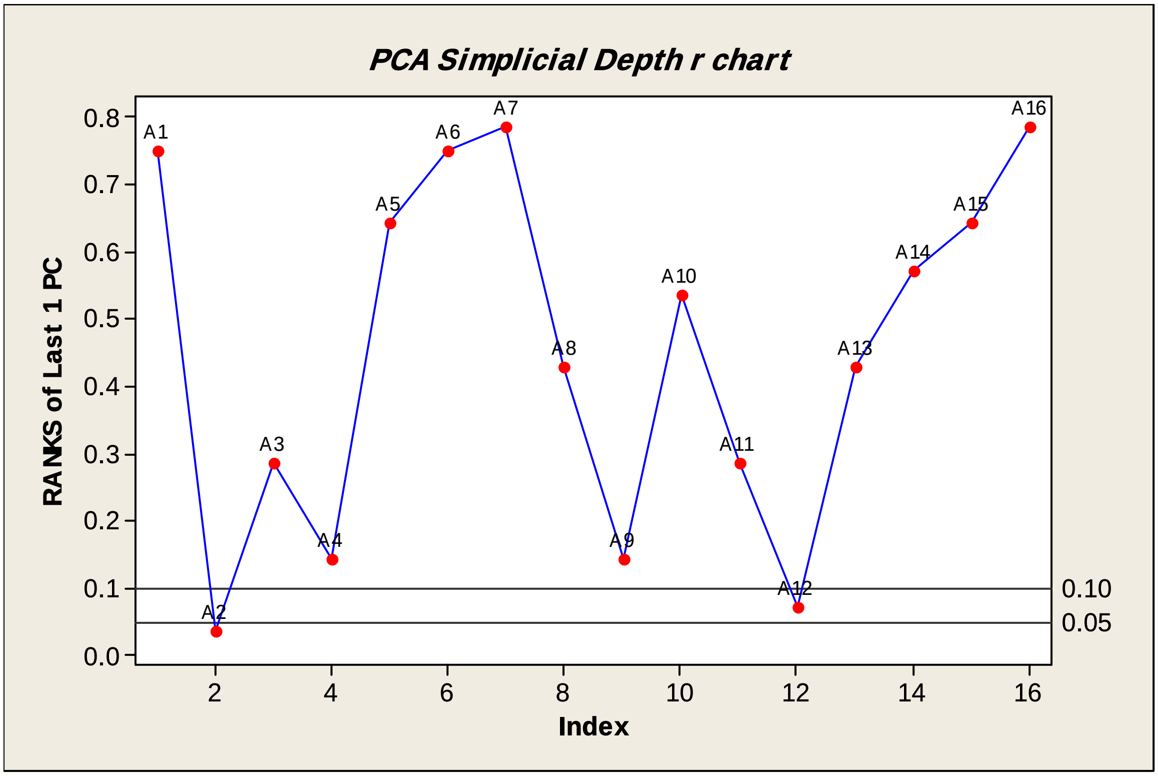

Figure 4.2b PCA Simplicial Depth r chart using the last PC for n = 28.

Using the HDS of n = 28, the following points from the 16 new observations were identified as out of control by the control chart when selecting the last PC:

At α = 0.05: Point A2

At α = 0.10: Points A2 and A12

Points A2, A3, A4, A6, A7, A8, A9, A12, A15 and A16 are out of control under various scenarios. [Some scenarios included more PCs than shown here in generating possible false alarms.] We removed all of these points, because we believe that some of these may be some false alarms. Since observations A1, A5, A10, A11, A13 and A14 are always in control, we will now augment the original HDS using these points in order to determine if the results of the PCA Simplicial Depth r-chart are significantly affected by changes in sample size. Table 4.2d represents the extended HDS with the six additional in control points included followed by the eigenanalysis in Table 4.2e. The 10 out of control points that were removed are listed in Table 4.2f.

Steam Turbine Data – Extended Historical Data Set (HDS) with *n*=34

(28 plus 6 identified as in-control from the control charts with *n*=28)

Table 4.2d Extended Historical Data Set of the Steam Turbine Data with *n* = 34.

Obs	Fuel	Steam Flow	Steam Temp	MegaWatts	Cool Temp	Pressure
1	232666	178753	850	20.53	54.1	29.2
2	237813	177645	847	20.55	54.2	29.2
3	240825	177817	848	20.55	54	29.2
4	240244	178839	850	20.57	53.9	29.1
5	239042	177817	849	20.57	53.9	29.2
6	239436	177903	850	20.59	54	29.1
7	234428	177903	848	20.57	53.9	29.2
8	232319	177990	848	20.55	53.7	29.1
9	233370	177903	848	20.48	53.6	29.1
10	237221	178076	850	20.49	53.9	29.1
11	238416	177817	848	20.55	53.9	29.1
12	235607	177817	848	20.55	53.8	29.1
13	241423	177903	847	20.55	53.7	29.1
14	233353	177731	849	20.53	53.6	29.1
15	231324	178753	846	20.64	53.9	29.1

16	243930	187378	844	21.67	53.9	29.1
17	252550	187287	843	21.65	54.2	29.1
18	251166	187745	842	21.67	53.7	29.1
19	252597	188770	841	21.78	53.4	29.1
20	243360	179868	842	20.66	53.7	29.1
21	238771	181389	843	20.81	53.9	29.1
22	239777	181411	841	20.88	54	29.1
23	219664	167330	850	19.08	54.1	29.2
24	228634	176137	846	20.64	54	29.2
25	231514	176029	843	20.24	53.8	29.2
26	235024	176115	846	20.22	53.6	29.2
27	239413	176115	845	20.31	53.7	29.2
28	228795	176201	847	20.24	54.3	29.2
29	234953	181678	843	20.84	54.5	29
30	246525	185511	842	21.21	54.6	28.9
31	233074	181411	844	20.93	54.5	29
32	242833	186216	844	21.59	54.4	29
33	238739	183349	844	21.01	54.3	29
34	251963	188012	850	21.68	54.4	29

Eigenanalysis of the Steam Turbine Correlation Matrix with *n*=34

Table 4.2e The Eigenanalysis of the extended HDS with $n = 34$.

	PC1	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**
Eigenvalue	3.5924	1.1941	0.6726	0.3755	0.1545	0.0109
Proportion	0.599	0.199	0.112	0.063	0.026	0.002
Cumulative	0.599	0.798	0.910	0.972	0.998	1.000

10 NEW OBSERVATIONS for *n*=34

Table 4.2f The 10 NEW observations for n = 34

Obs	Name	Fuel	SteamFlow	SteamTemp	MegaWatts	CoolTemp	Pressure
1	**A2**	247080	189354	844	20.86	54.4	28.9
2	**A3**	238323	184419	845	21.1	54.5	28.9
3	**A4**	248801	189169	843	22.18	54.5	28.9
4	**A6**	233215	180409	845	20.75	54.5	29
5	**A7**	233955	181323	842	20.82	54.6	29
6	**A8**	238693	181346	844	20.92	54.8	29
7	**A9**	248048	185307	844	21.15	54.6	29
8	**A12**	243950	182147	844	21.37	54.2	29
9	**A15**	240058	183372	846	21.15	54.2	29
10	**A16**	235376	182436	844	20.99	54.3	29

PCA Simplicial Depth r chart for the first PC of the 10 NEW observations for n=34 (α = 0.05 and 0.10)

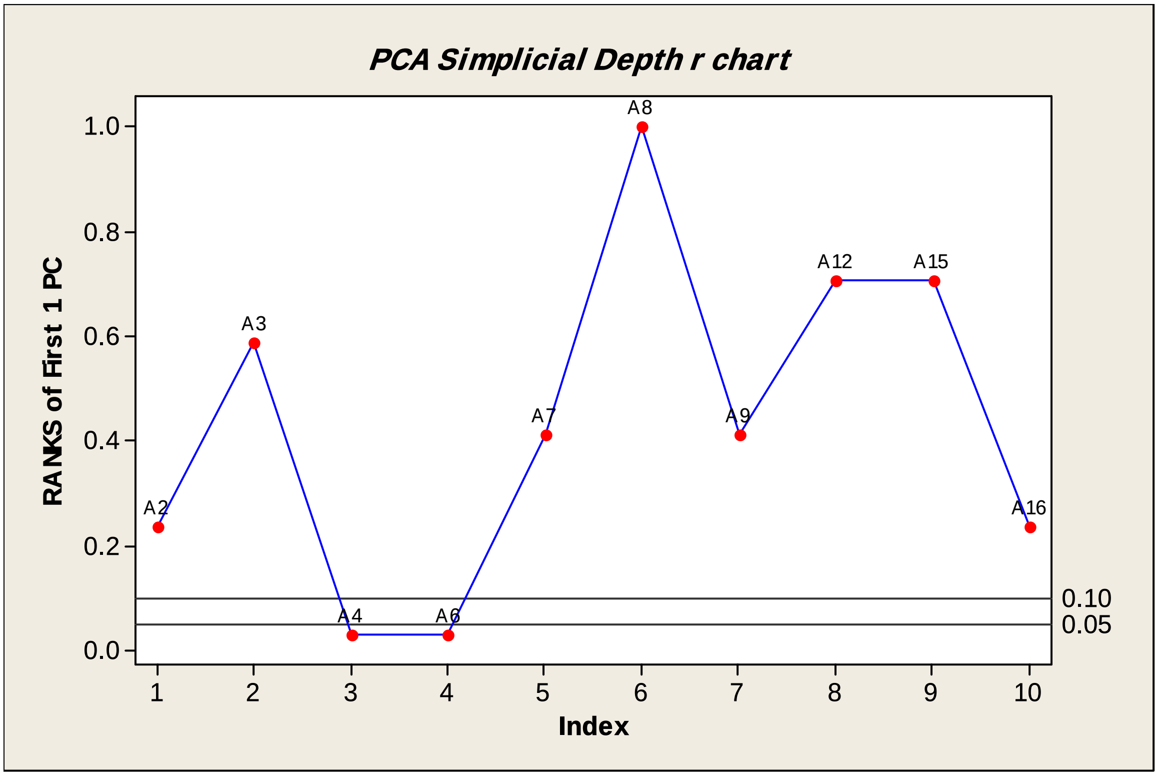

Figure 4.2c PCA Simplicial Depth r chart using the first PC n = 34.

Using the HDS of n = 34, the following points from the 10 new observations were identified as out of control by the control chart when selecting the first PC:

At α = 0.05: Points A4 and A6

At α = 0.10: Points A4 and A6

PCA Simplicial Depth r chart for the last PC of the 10 NEW observations for *n*=28 (α = 0.05 and 0.10)

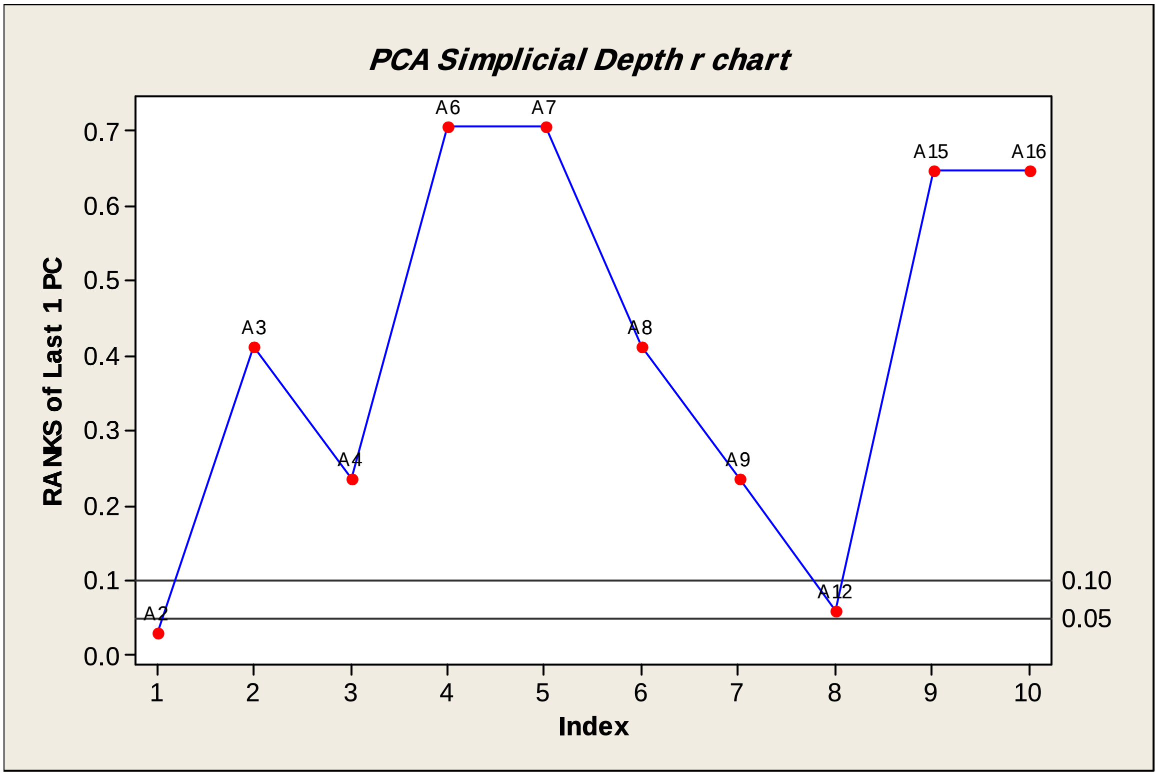

Figure 4.2d PCA Simplicial Depth r chart using the last PC for n = 34.

Using the HDS of n = 34, the following points from the 10 new observations were identified as out of control by the control chart when selecting the last PC:

At α = 0.05: Point A2

At α = 0.10: Points A2 and A12

Table 4.2g summarizes the points that were identified as out of control from our proposed PCA Simplicial Depth r charts with the first and last PCs selected for α = 0.05 and 0.10.

Steam Turbine Summary Table

Table 4.2g   16 NEW points with the out of control observations labeled **X(*n*=28) and Y(*n*=34)**.

	$\alpha = 0.05$		$\alpha = 0.10$	
Name of point	**First PC**	**Last PC**	**First PC**	**Last PC**
A1				
A2		X Y		X Y
A3				
A4	X Y		X Y	
A5				
A6	Y		X Y	
A7				
A8				
A9				
A10				
A11				
A12				X Y
A13				
A14				
A15				
A16				

4.3 Fruit Juice Data Set

Our second data set with 11 process variables consists of 69 observations with the first 36 observations representing the HDS. This is from Fuchs and Kennet (1998) and is a fruit juice process made up of 11 amino acids namely, Lysine (LYS), Arginine (ARG), Aspartic Acid (ASP), Serine (SER), Glutamine Acid (GLU), Proline (PRO), Glycine (GLY), Alanine (ALA) Valine (VAL), Phenyl Alanine (PHA) and Gamma-amino Acid Butric Acid (GABA). Table 4.3a lists the 36 observations of the HDS.

Historical Data Set (HDS) Original 36 observations

Table 4.3a Historical Data Set of the Fruit Juice Data with *n* = 36

Obs	LYS	ARG	ASP	SER	GLU	PRO	GLY	ALA	VAL	PHA	GABA
1	0.48	5.81	2.12	4.68	0.78	12.41	0.31	0.96	0.18	0.2	4.73
2	0.47	5.25	2.75	4.42	0.88	14.72	0.3	1.04	0.19	0.22	3.96
3	0.42	4.98	2.79	3.85	0.75	12.13	0.32	0.99	0.15	0.2	3.94
4	0.35	4.79	2.79	3.39	0.81	12.77	0.25	0.75	0.16	0.15	3.69
5	0.43	4.92	2.88	3.53	0.78	13.11	0.25	0.91	0.16	0.15	4.23
6	0.4	5.61	2.26	3.39	0.69	12.69	0.2	1.06	0.16	0.18	3.76
7	0.35	4.54	2.96	3.89	0.88	14.01	0.24	0.86	0.16	0.12	3.92
8	0.34	3.82	2.86	3.63	0.86	15.73	0.22	1.34	0.14	0.12	2.88
9	0.27	3.42	2.27	4.81	0.9	8.99	0.23	1.43	0.1	0.1	2.68
10	0.39	3.6	2.99	5.03	0.92	13.71	0.28	1.99	0.13	0.1	2.88
11	0.37	3.39	2.78	5.96	0.84	12.92	0.24	1.76	0.12	0.14	3.01
12	0.26	2.72	3.82	6.03	1.17	7.18	0.15	1.3	0.11	0.07	3.4
13	0.24	3.13	3.35	5.76	0.96	6.75	0.21	1.14	0.11	0.08	2.43
14	0.2	2.15	3.28	5.8	1.04	5.34	0.22	1.06	0.12	0.08	2.41
15	0.26	2.89	3.67	6.34	1.22	5.87	0.18	1.1	0.14	0.12	2.4
16	0.52	5.53	2.97	3.37	0.78	10.74	0.24	0.96	0.1	0.16	3.4
17	0.42	5.07	3.06	4.32	0.91	15.37	0.47	1.32	0.16	0.2	3.63
18	0.45	5.46	3.06	4.68	0.84	16.52	0.39	1.35	0.14	0.18	3.89
19	0.47	5.79	2.91	4.44	0.8	16.21	0.35	1.2	0.2	0.18	4.52
20	0.44	2.52	2.4	4.09	0.72	12.81	0.28	0.86	0.18	0.23	4.43
21	0.48	5.14	2.66	4.04	0.94	16.77	0.33	0.97	0.22	0.23	4.9

22	0.49	4.77	2.42	5.92	1	15.62	0.34	1.93	0.5	0.15	4.05
23	0.37	4.35	3.04	5.07	0.87	15.81	0.31	2.08	0.19	0.1	4.17
24	0.36	4.01	2.37	3.93	0.76	11.28	0.22	0.75	0.12	0.12	3.27
25	0.46	4.26	2.51	7.29	1.07	18.57	0.37	2.67	0.19	0.1	2.95
26	0.34	3.46	2.2	3.8	0.93	11.73	0.26	1.4	0.18	0.1	3.06
27	0.34	4.13	2.72	6.01	0.95	13.96	0.34	2.3	0.1	0.08	3.06
28	0.31	3.7	2.77	5.29	0.85	10.8	0.22	1.68	0.1	0.01	2.61
29	0.3	3.18	2.54	5.04	0.95	11.25	0.21	1.84	0.1	0.01	2.48
30	0.3	3.57	2.45	5.7	1.06	12.28	0.26	1.53	0.1	0.1	2.46
31	0.3	3.31	2.53	5.21	0.88	9.1	0.23	1.37	0.08	0.01	2.55
32	0.3	3.13	2.82	5.85	1	10.31	0.21	1.55	0.1	0.08	2.69
33	0.33	3.1	3.01	7.15	1.04	12.71	0.23	1.79	0.09	0.1	3.52
34	0.32	3.84	3.79	6.08	1.01	10.13	0.18	1.3	0.09	0.01	3.67
35	0.3	3.75	2.83	6.24	0.71	6.2	0.16	1.2	0.05	0.08	3.01
36	0.26	3.34	3.46	7.01	1.02	6.68	0.2	1.52	0.1	0.08	2.18

Eigenanalysis of the Fruit Juice Correlation Matrix with *n*=36

Table 4.3b The Eigenanalysis of the HDS with *n* = 36

	PC1	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**	**PC7**	**PC8**
Eigenvalue	5.4394	2.1033	1.1141	0.6935	0.4324	0.3652	0.2942	0.2148
Proportion	0.494	0.191	0.101	0.063	0.039	0.033	0.027	0.020
Cumulative	0.494	0.686	0.787	0.850	0.889	0.923	0.949	0.969

	PC9	**PC10**	**PC11**
Eigenvalue	0.1748	0.1025	0.0658
Proportion	0.016	0.009	0.006
Cumulative	0.985	0.994	1.000

33 NEW OBSERVATIONS for *n*=36

Table 4.3c The 33 NEW observations for *n* = 33

Obs	Name of point	LYS	ARG	ASP	SER	GLU	PRO	GLY	ALA	VAL	PHA	GABA
1	**B1**	0.43	5.84	2.84	3.54	0.8	11.9	0.3	0.86	0.2	0.18	3.88
2	**B2**	0.5	4.61	2.08	5.7	0.71	18.46	0.42	1.91	0.18	0.18	6.14
3	**B3**	0.51	6.19	3.55	4.29	1.16	19.01	0.4	1.2	0.15	0.18	4.72
4	**B4**	0.43	5.44	2.71	4.38	0.79	13.59	0.35	1.23	0.14	0.2	4.08
5	**B5**	0.38	5.22	2.54	3.97	0.73	14.47	0.3	0.98	0.15	0.2	4.18
6	**B6**	0.5	5.19	3.13	4.32	0.9	16.74	0.34	1.09	0.2	0.22	4.85
7	**B7**	0.4	4.68	2.38	3.47	0.68	12.01	0.26	0.92	0.16	0.18	3.95
8	**B8**	0.43	4.99	2.03	3.52	0.63	9.84	0.24	0.71	0.19	0.2	4.06
9	**B9**	0.41	5.33	2.64	4.22	0.81	13.66	0.3	0.86	0.17	0.22	4.52
10	**B10**	0.45	5.42	2.96	4.8	0.91	15.73	0.3	1.09	0.19	0.2	3.8
11	**B11**	0.36	4.83	2.72	3.32	0.75	12.28	0.23	0.71	0.13	0.12	3.63
12	**B12**	0.4	4.34	1.92	4.57	0.74	11.13	0.28	1.63	0.14	0.1	3.34
13	**B13**	0.36	4.41	2.88	3.76	0.89	14.32	0.25	0.89	0.14	0.12	3.35
14	**B14**	0.3	4.14	2.5	5.26	0.86	15.48	0.35	2.34	0.21	0.12	3.02
15	**B15**	0.38	3.91	2.32	5.14	0.82	14.27	0.29	1.87	0.22	0.1	3.98
16	**B16**	0.42	3.9	2.45	5.26	0.94	18.14	0.29	2.03	0.16	0.12	3.65
17	**B17**	0.31	3.56	2.61	5.4	0.97	12.29	0.22	1.59	0.08	0.1	2.82
18	**B18**	0.32	4.18	3.76	5.53	0.98	10.81	0.22	1.32	0.1	0.14	2.91
19	**B19**	0.32	3.05	3.24	6.87	1.43	13.01	0.24	1.81	0.1	0.1	2.91
20	**B20**	0.23	3.13	3.43	6.3	1.15	10.67	0.26	1.67	0.12	0.16	2.86
21	**B21**	0.24	2.85	3.18	4.64	0.86	6.91	0.21	1.08	0.01	0.12	2.75
22	**B22**	0.36	4.31	2.25	3.15	0.65	11.32	0.22	0.83	0.19	0.2	3.66
23	**B23**	0.35	4.62	2.4	2.94	0.71	10.18	0.19	0.89	0.19	0.2	3.01
24	**B24**	0.39	4.51	2.82	4	0.87	13.76	0.27	0.88	0.17	0.12	3.56
25	**B25**	0.41	4.12	2.38	5.14	0.83	11.36	0.26	1.71	0.16	0.08	3.65
26	**B26**	0.33	3.6	2.36	5.07	0.94	13.93	0.3	1.62	0.1	0.08	3.51
27	**B27**	0.43	4.11	2.22	6.86	1.12	14.35	0.27	1.68	0.1	0.12	3.96
28	**B28**	0.31	3.7	2.77	5.44	1.02	12.68	0.32	1.75	0.1	0.1	3.47
29	**B29**	0.36	3.64	2.21	6.56	1.02	15.53	0.39	1.96	0.1	0.1	3.07
30	**B30**	0.27	3.25	2.82	4.92	0.91	8.43	0.2	1.53	0.1	0.01	2.32
31	**B31**	0.28	2.91	3.21	6.41	1.35	9.42	0.22	1.8	0.12	0.01	2.85
32	**B32**	0.3	3.64	2.73	5.76	0.73	5.55	0.2	0.94	0.05	0.05	3.14
33	**B33**	0.28	2.68	3.61	6.38	1.06	6.94	0.22	1.22	0.11	0.11	2.71

PCA Simplicial Depth r chart for the first PC of the 33 NEW observations for n=36 (α = 0.05 and 0.10)

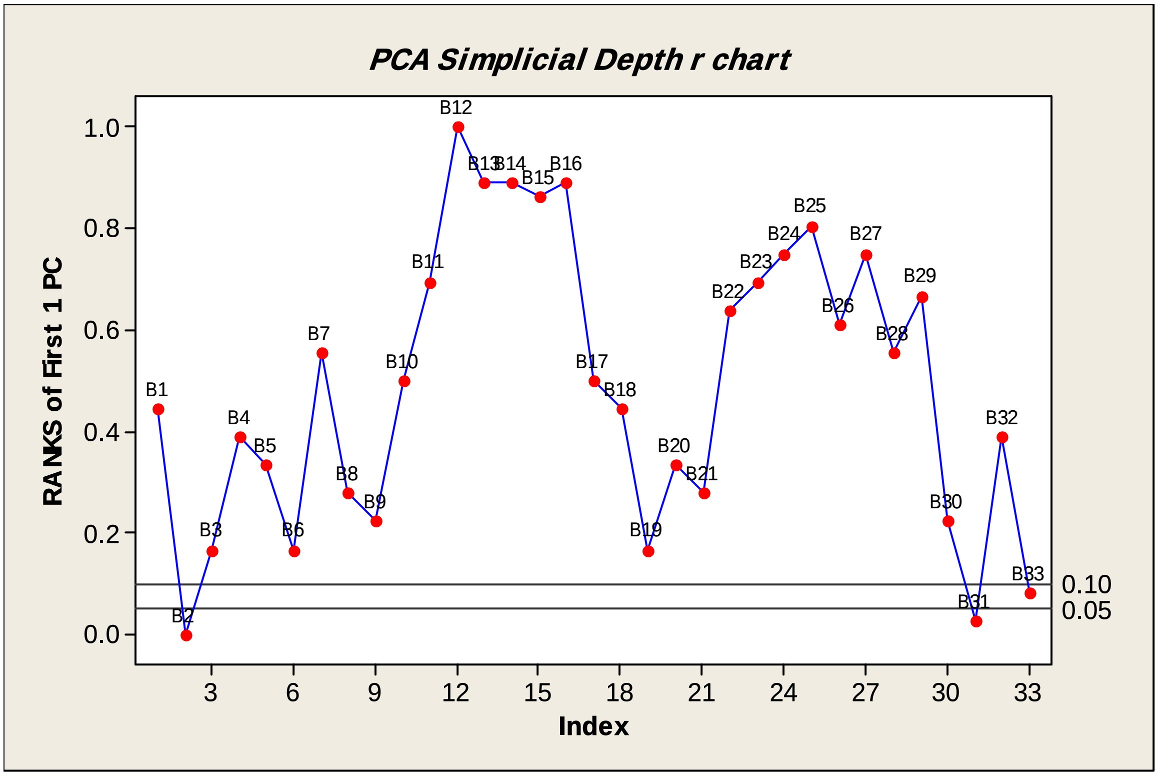

Figure 4.3a PCA Simplicial Depth r chart using the first PC for n=36

Using the HDS of n = 36, the following points from the 33 new observations were identified as out of control by the control chart when selecting the first PC:

At α = 0.05: Points B2 and B31

At α = 0.10: Points B2, B31 and B33

PCA Simplicial Depth r chart for the last PC of the 33 NEW observations for n=36 (α = 0.05 and 0.10)

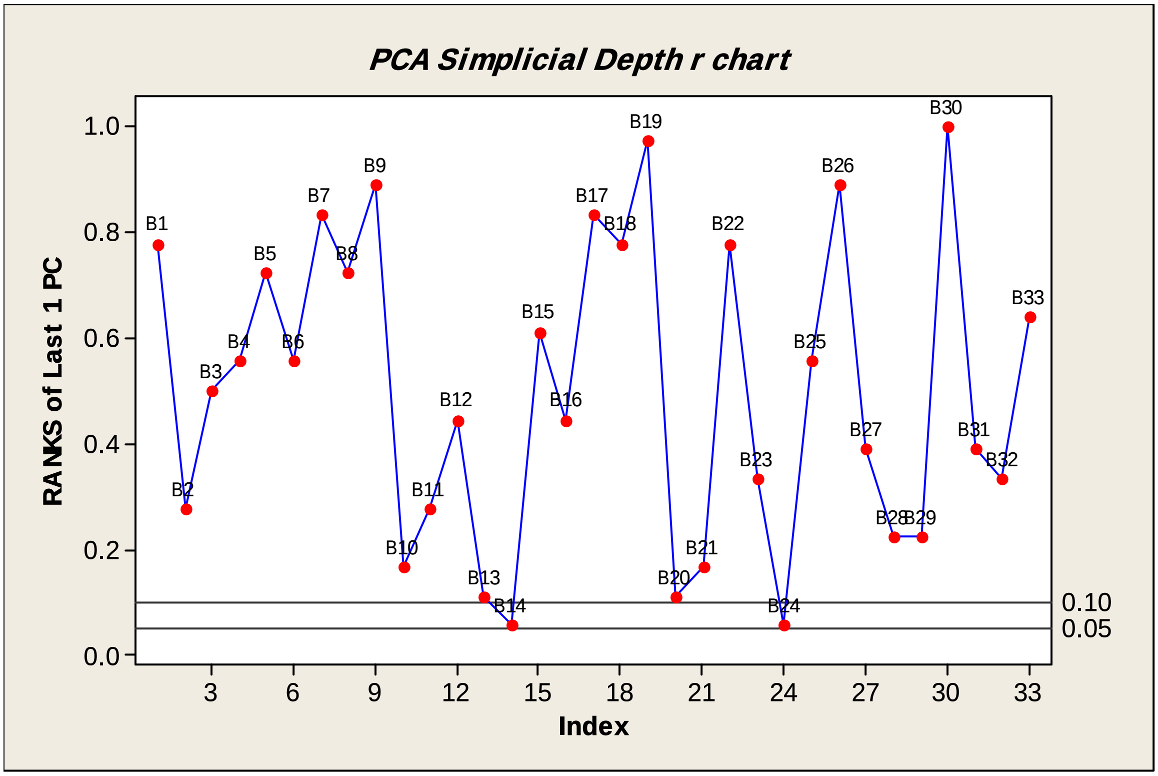

Figure 4.3b PCA Simplicial Depth r chart using the last PC for n = 36

Using the HDS of n = 36, the following points from the 33 new observations were identified as out of control by the control chart when selecting the last PC:

At α = 0.05: No Points

At α = 0.10: Points B14 and B24

The control charts generated a signal for the following 8 points: B1, B2, B3, B8, B14, B24, B31 and B33 for various scenarios. [Some scenarios included more PCs than shown here in generating possible false alarms.] From the remaining 25 points that were determined to be in-control, we will initially select a subset of 11 and augment the HDS to 47 followed by an additional augmentation to 61 from the rest of the in-control points, in order to determine the sensitivity the control chart has with respect to sample size. The following are the 25 in-control points: B4, B5, B6, B7, B9, B10, B11, B12, B13, B15, B16, B17, B18, B19, B20, B21, B22, B23, B25, B26, B27, B28, B29, B30 and B32. From the list, we will choose every other point until we acquire 11 points. The 11 in-control points that will be included in the extended HDS with n = 47 are: B5, B7, B10, B12, B15, B17, B19, B21, B23, B26, B28. For the extended HDS with n = 61, we will include those 11 points in addition to the remaining 14 in-control points namely, B4, B6, B9, B11, B13, B16, B18, B20, B22, B25, B27, B29, B30 and B32. Table 4.3d illustrates the 8 observations that were identified as out of control with n = 36 and will be further monitored using the extended HDS.

8 OBSERVATIONS out of 33 that were identified as out of control from various PCA Simplicial Depth r-charts for *n*=36.

Table 4.3d The 8 NEW points for Extended HDS with n = 47 and with n = 61.

Obs	Name of point	LYS	ARG	ASP	SER	GLU	PRO	GLY	ALA	VAL	PHA	GABA
1	B1	0.43	5.84	2.84	3.54	0.8	11.9	0.3	0.86	0.2	0.18	3.88
2	B2	0.5	4.61	2.08	5.7	0.71	18.46	0.42	1.91	0.18	0.18	6.14
3	B3	0.51	6.19	3.55	4.29	1.16	19.01	0.4	1.2	0.15	0.18	4.72
4	B8	0.43	4.99	2.03	3.52	0.63	9.84	0.24	0.71	0.19	0.2	4.06
5	B14	0.3	4.14	2.5	5.26	0.86	15.48	0.35	2.34	0.21	0.12	3.02
6	B24	0.39	4.51	2.82	4	0.87	13.76	0.27	0.88	0.17	0.12	3.56
7	B31	0.28	2.91	3.21	6.41	1.35	9.42	0.22	1.8	0.12	0.01	2.85
8	B33	0.28	2.68	3.61	6.38	1.06	6.94	0.22	1.22	0.11	0.11	2.71

Since some of these signals may have been false positives, we will monitor these 8 observations using the augmented HDS for n = 47 and n = 61. The data analyses of the 8 observations with the extended HDS for n = 47 and n = 61 are illustrated by the following tables and figures.

Fruit Juice Data – Extended Historical Data Set (HDS) with *n*=47

(36 plus 11 identified as in-control from the control charts with *n*=36)

Table 4.3e Extended Historical Data Set of the Fruit Juice Data with n = 47.

Obs	LYS	ARG	ASP	SER	GLU	PRO	GLY	ALA	VAL	PHA	GABA
1	0.48	5.81	2.12	4.68	0.78	12.41	0.31	0.96	0.18	0.2	4.73
2	0.47	5.25	2.75	4.42	0.88	14.72	0.3	1.04	0.19	0.22	3.96
3	0.42	4.98	2.79	3.85	0.75	12.13	0.32	0.99	0.15	0.2	3.94
4	0.35	4.79	2.79	3.39	0.81	12.77	0.25	0.75	0.16	0.15	3.69
5	0.43	4.92	2.88	3.53	0.78	13.11	0.25	0.91	0.16	0.15	4.23
6	0.4	5.61	2.26	3.39	0.69	12.69	0.2	1.06	0.16	0.18	3.76
7	0.35	4.54	2.96	3.89	0.88	14.01	0.24	0.86	0.16	0.12	3.92
8	0.34	3.82	2.86	3.63	0.86	15.73	0.22	1.34	0.14	0.12	2.88
9	0.27	3.42	2.27	4.81	0.9	8.99	0.23	1.43	0.1	0.1	2.68
10	0.39	3.6	2.99	5.03	0.92	13.71	0.28	1.99	0.13	0.1	2.88
11	0.37	3.39	2.78	5.96	0.84	12.92	0.24	1.76	0.12	0.14	3.01
12	0.26	2.72	3.82	6.03	1.17	7.18	0.15	1.3	0.11	0.07	3.4
13	0.24	3.13	3.35	5.76	0.96	6.75	0.21	1.14	0.11	0.08	2.43
14	0.2	2.15	3.28	5.8	1.04	5.34	0.22	1.06	0.12	0.08	2.41
15	0.26	2.89	3.67	6.34	1.22	5.87	0.18	1.1	0.14	0.12	2.4
16	0.52	5.53	2.97	3.37	0.78	10.74	0.24	0.96	0.1	0.16	3.4
17	0.42	5.07	3.06	4.32	0.91	15.37	0.47	1.32	0.16	0.2	3.63
18	0.45	5.46	3.06	4.68	0.84	16.52	0.39	1.35	0.14	0.18	3.89
19	0.47	5.79	2.91	4.44	0.8	16.21	0.35	1.2	0.2	0.18	4.52
20	0.44	2.52	2.4	4.09	0.72	12.81	0.28	0.86	0.18	0.23	4.43
21	0.48	5.14	2.66	4.04	0.94	16.77	0.33	0.97	0.22	0.23	4.9
22	0.49	4.77	2.42	5.92	1	15.62	0.34	1.93	0.5	0.15	4.05
23	0.37	4.35	3.04	5.07	0.87	15.81	0.31	2.08	0.19	0.1	4.17
24	0.36	4.01	2.37	3.93	0.76	11.28	0.22	0.75	0.12	0.12	3.27
25	0.46	4.26	2.51	7.29	1.07	18.57	0.37	2.67	0.19	0.1	2.95
26	0.34	3.46	2.2	3.8	0.93	11.73	0.26	1.4	0.18	0.1	3.06
27	0.34	4.13	2.72	6.01	0.95	13.96	0.34	2.3	0.1	0.08	3.06
28	0.31	3.7	2.77	5.29	0.85	10.8	0.22	1.68	0.1	0.01	2.61
29	0.3	3.18	2.54	5.04	0.95	11.25	0.21	1.84	0.1	0.01	2.48
30	0.3	3.57	2.45	5.7	1.06	12.28	0.26	1.53	0.1	0.1	2.46
31	0.3	3.31	2.53	5.21	0.88	9.1	0.23	1.37	0.08	0.01	2.55
32	0.3	3.13	2.82	5.85	1	10.31	0.21	1.55	0.1	0.08	2.69
33	0.33	3.1	3.01	7.15	1.04	12.71	0.23	1.79	0.09	0.1	3.52

34	0.32	3.84	3.79	6.08	1.01	10.13	0.18	1.3	0.09	0.01	3.67
35	0.3	3.75	2.83	6.24	0.71	6.2	0.16	1.2	0.05	0.08	3.01
36	0.26	3.34	3.46	7.01	1.02	6.68	0.2	1.52	0.1	0.08	2.18
37	0.38	5.22	2.54	3.97	0.73	14.47	0.3	0.98	0.15	0.2	4.18
38	0.4	4.68	2.38	3.47	0.68	12.01	0.26	0.92	0.16	0.18	3.95
39	0.45	5.42	2.96	4.8	0.91	15.73	0.3	1.09	0.19	0.2	3.8
40	0.4	4.34	1.92	4.57	0.74	11.13	0.28	1.63	0.14	0.1	3.34
41	0.38	3.91	2.32	5.14	0.82	14.27	0.29	1.87	0.22	0.1	3.98
42	0.31	3.56	2.61	5.4	0.97	12.29	0.22	1.59	0.08	0.1	2.82
43	0.32	3.05	3.24	6.87	1.43	13.01	0.24	1.81	0.1	0.1	2.91
44	0.24	2.85	3.18	4.64	0.86	6.91	0.21	1.08	0.01	0.12	2.75
45	0.35	4.62	2.4	2.94	0.71	10.18	0.19	0.89	0.19	0.2	3.01
46	0.33	3.6	2.36	5.07	0.94	13.93	0.3	1.62	0.1	0.08	3.51
47	0.31	3.7	2.77	5.44	1.02	12.68	0.32	1.75	0.1	0.1	3.47

Eigenanalysis of the Correlation Matrix with n = 47

Table 4.3f The Eigenanalysis of the extended HDS with n = 47.

	PC1	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**	**PC7**	**PC8**
Eigenvalue	5.3553	2.2153	1.1227	0.6386	0.4160	0.3133	0.3028	0.2598
Proportion	0.487	0.201	0.102	0.058	0.038	0.028	0.028	0.024
Cumulative	0.487	0.688	0.790	0.848	0.886	0.915	0.942	0.966

	PC9	**PC10**	**PC11**
Eigenvalue	0.1868	0.1173	0.0722
Proportion	0.017	0.011	0.007
Cumulative	0.983	0.993	1.000

PCA Simplicial Depth r chart for the first PC of the 8 NEW observations for *n*=47 (α = 0.05 and 0.10)

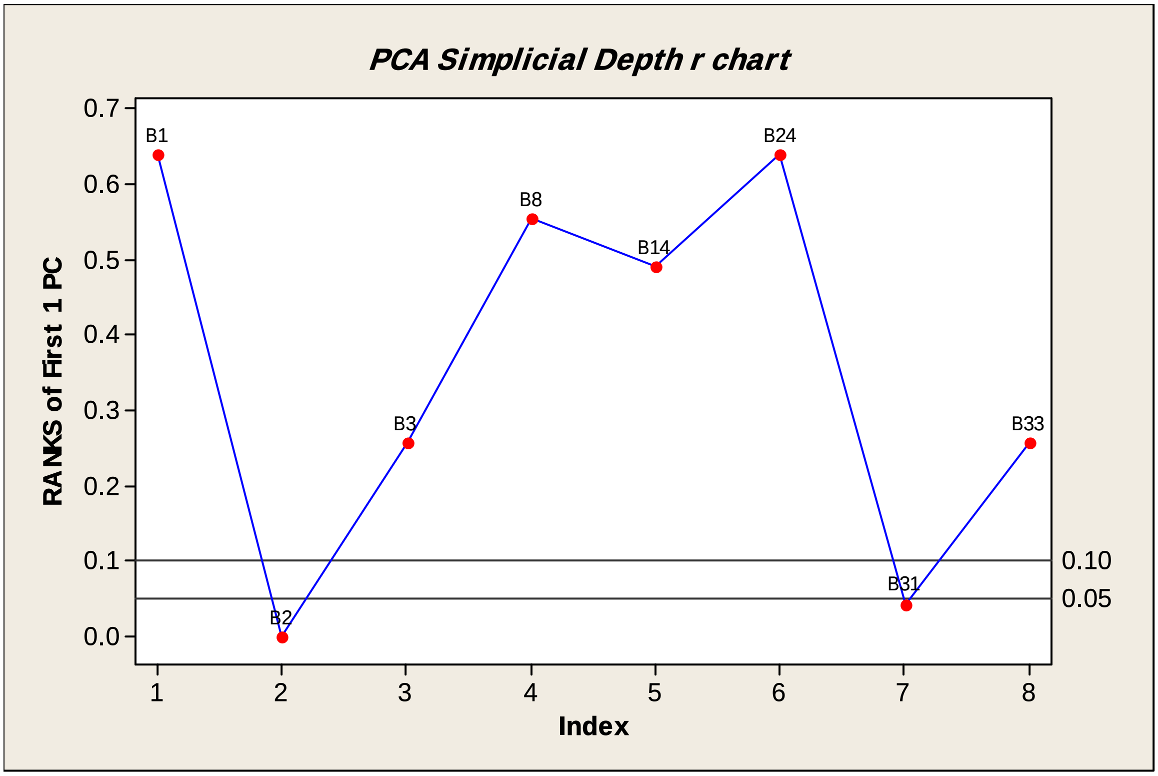

Figure 4.3c PCA Simplicial Depth r chart using the first PC for *n*=47.

Using the HDS of n = 47, the following points from the 8 new observations were identified as out of control by the control chart when selecting the first PC:

At α = 0.05: Points B2 and B31

At α = 0.10: Points B2 and B31

PCA Simplicial Depth r chart for the last PC of the 8 NEW observations

for n=47 (α = 0.05 and 0.10)

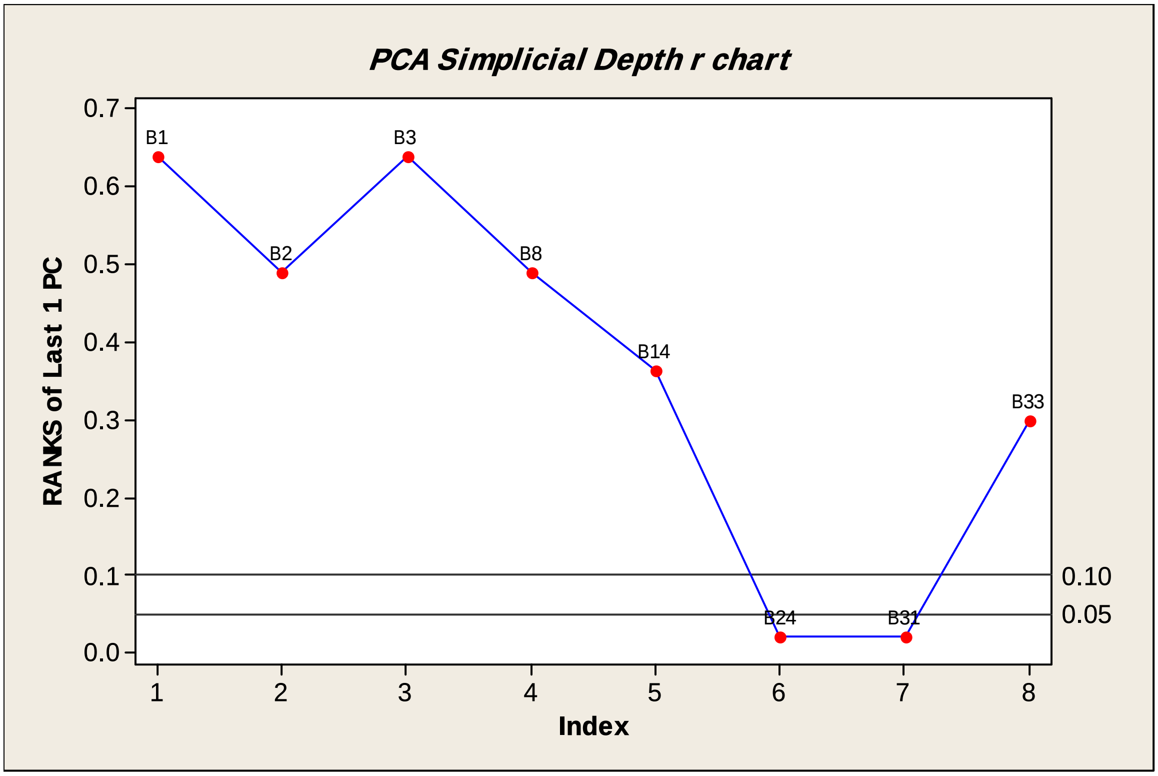

Figure 4.3d PCA Simplicial Depth r chart using the last PC for n = 47.

Using the HDS of n = 47, the following points from the 8 new observations were identified as out of control by the control chart when selecting the last PC:

At α = 0.05: Points B24 and B31

At α = 0.10: Points B24 and B31

Fruit Juice Data – Extended Historical Data Set (HDS) with *n*=61

(36 plus 25 identified as in-control from the control charts with *n*=36)

Table 4.3g Extended Historical Data Set of the Fruit Juice Data with n = 61.

Obs	LYS	ARG	ASP	SER	GLU	PRO	GLY	ALA	VAL	PHA	GABA
1	0.48	5.81	2.12	4.68	0.78	12.41	0.31	0.96	0.18	0.2	4.73
2	0.47	5.25	2.75	4.42	0.88	14.72	0.3	1.04	0.19	0.22	3.96
3	0.42	4.98	2.79	3.85	0.75	12.13	0.32	0.99	0.15	0.2	3.94
4	0.35	4.79	2.79	3.39	0.81	12.77	0.25	0.75	0.16	0.15	3.69
5	0.43	4.92	2.88	3.53	0.78	13.11	0.25	0.91	0.16	0.15	4.23
6	0.4	5.61	2.26	3.39	0.69	12.69	0.2	1.06	0.16	0.18	3.76
7	0.35	4.54	2.96	3.89	0.88	14.01	0.24	0.86	0.16	0.12	3.92
8	0.34	3.82	2.86	3.63	0.86	15.73	0.22	1.34	0.14	0.12	2.88
9	0.27	3.42	2.27	4.81	0.9	8.99	0.23	1.43	0.1	0.1	2.68
10	0.39	3.6	2.99	5.03	0.92	13.71	0.28	1.99	0.13	0.1	2.88
11	0.37	3.39	2.78	5.96	0.84	12.92	0.24	1.76	0.12	0.14	3.01
12	0.26	2.72	3.82	6.03	1.17	7.18	0.15	1.3	0.11	0.07	3.4
13	0.24	3.13	3.35	5.76	0.96	6.75	0.21	1.14	0.11	0.08	2.43
14	0.2	2.15	3.28	5.8	1.04	5.34	0.22	1.06	0.12	0.08	2.41
15	0.26	2.89	3.67	6.34	1.22	5.87	0.18	1.1	0.14	0.12	2.4
16	0.52	5.53	2.97	3.37	0.78	10.74	0.24	0.96	0.1	0.16	3.4
17	0.42	5.07	3.06	4.32	0.91	15.37	0.47	1.32	0.16	0.2	3.63
18	0.45	5.46	3.06	4.68	0.84	16.52	0.39	1.35	0.14	0.18	3.89
19	0.47	5.79	2.91	4.44	0.8	16.21	0.35	1.2	0.2	0.18	4.52
20	0.44	2.52	2.4	4.09	0.72	12.81	0.28	0.86	0.18	0.23	4.43
21	0.48	5.14	2.66	4.04	0.94	16.77	0.33	0.97	0.22	0.23	4.9
22	0.49	4.77	2.42	5.92	1	15.62	0.34	1.93	0.5	0.15	4.05
23	0.37	4.35	3.04	5.07	0.87	15.81	0.31	2.08	0.19	0.1	4.17
24	0.36	4.01	2.37	3.93	0.76	11.28	0.22	0.75	0.12	0.12	3.27
25	0.46	4.26	2.51	7.29	1.07	18.57	0.37	2.67	0.19	0.1	2.95
26	0.34	3.46	2.2	3.8	0.93	11.73	0.26	1.4	0.18	0.1	3.06
27	0.34	4.13	2.72	6.01	0.95	13.96	0.34	2.3	0.1	0.08	3.06
28	0.31	3.7	2.77	5.29	0.85	10.8	0.22	1.68	0.1	0.01	2.61
29	0.3	3.18	2.54	5.04	0.95	11.25	0.21	1.84	0.1	0.01	2.48
30	0.3	3.57	2.45	5.7	1.06	12.28	0.26	1.53	0.1	0.1	2.46
31	0.3	3.31	2.53	5.21	0.88	9.1	0.23	1.37	0.08	0.01	2.55
32	0.3	3.13	2.82	5.85	1	10.31	0.21	1.55	0.1	0.08	2.69
33	0.33	3.1	3.01	7.15	1.04	12.71	0.23	1.79	0.09	0.1	3.52
34	0.32	3.84	3.79	6.08	1.01	10.13	0.18	1.3	0.09	0.01	3.67
35	0.3	3.75	2.83	6.24	0.71	6.2	0.16	1.2	0.05	0.08	3.01

36	0.26	3.34	3.46	7.01	1.02	6.68	0.2	1.52	0.1	0.08	2.18
37	0.38	5.22	2.54	3.97	0.73	14.47	0.3	0.98	0.15	0.2	4.18
38	0.4	4.68	2.38	3.47	0.68	12.01	0.26	0.92	0.16	0.18	3.95
39	0.45	5.42	2.96	4.8	0.91	15.73	0.3	1.09	0.19	0.2	3.8
40	0.4	4.34	1.92	4.57	0.74	11.13	0.28	1.63	0.14	0.1	3.34
41	0.38	3.91	2.32	5.14	0.82	14.27	0.29	1.87	0.22	0.1	3.98
42	0.31	3.56	2.61	5.4	0.97	12.29	0.22	1.59	0.08	0.1	2.82
43	0.32	3.05	3.24	6.87	1.43	13.01	0.24	1.81	0.1	0.1	2.91
44	0.24	2.85	3.18	4.64	0.86	6.91	0.21	1.08	0.01	0.12	2.75
45	0.35	4.62	2.4	2.94	0.71	10.18	0.19	0.89	0.19	0.2	3.01
46	0.33	3.6	2.36	5.07	0.94	13.93	0.3	1.62	0.1	0.08	3.51
47	0.31	3.7	2.77	5.44	1.02	12.68	0.32	1.75	0.1	0.1	3.47
48	0.43	5.44	2.71	4.38	0.79	13.59	0.35	1.23	0.14	0.2	4.08
49	0.5	5.19	3.13	4.32	0.9	16.74	0.34	1.09	0.2	0.22	4.85
50	0.41	5.33	2.64	4.22	0.81	13.66	0.3	0.86	0.17	0.22	4.52
51	0.36	4.83	2.72	3.32	0.75	12.28	0.23	0.71	0.13	0.12	3.63
52	0.36	4.41	2.88	3.76	0.89	14.32	0.25	0.89	0.14	0.12	3.35
53	0.42	3.9	2.45	5.26	0.94	18.14	0.29	2.03	0.16	0.12	3.65
54	0.32	4.18	3.76	5.53	0.98	10.81	0.22	1.32	0.1	0.14	2.91
55	0.23	3.13	3.43	6.3	1.15	10.67	0.26	1.67	0.12	0.16	2.86
56	0.36	4.31	2.25	3.15	0.65	11.32	0.22	0.83	0.19	0.2	3.66
57	0.41	4.12	2.38	5.14	0.83	11.36	0.26	1.71	0.16	0.08	3.65
58	0.43	4.11	2.22	6.86	1.12	14.35	0.27	1.68	0.1	0.12	3.96
59	0.36	3.64	2.21	6.56	1.02	15.53	0.39	1.96	0.1	0.1	3.07
60	0.27	3.25	2.82	4.92	0.91	8.43	0.2	1.53	0.1	0.01	2.32
61	0.3	3.64	2.73	5.76	0.73	5.55	0.2	0.94	0.05	0.05	3.14

Eigenanalysis of the Correlation Matrix with *n* = 61

Table 4.3h The Eigenanalysis of the extended HDS with n = 61.

	PC1	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**	**PC7**	**PC8**
Eigenvalue	5.2257	2.3095	1.1446	0.6261	0.4072	0.3620	0.2981	0.2419
Proportion	0.475	0.210	0.104	0.057	0.037	0.033	0.027	0.022
Cumulative	0.475	0.685	0.789	0.846	0.883	0.916	0.943	0.965

	PC9	**PC10**	**PC11**
Eigenvalue	0.1907	0.1225	0.0717
Proportion	0.017	0.011	0.007
Cumulative	0.982	0.993	1.000

PCA Simplicial Depth r chart for the first PC of the 8 NEW observations for *n*=61 (α = 0.05 and 0.10)

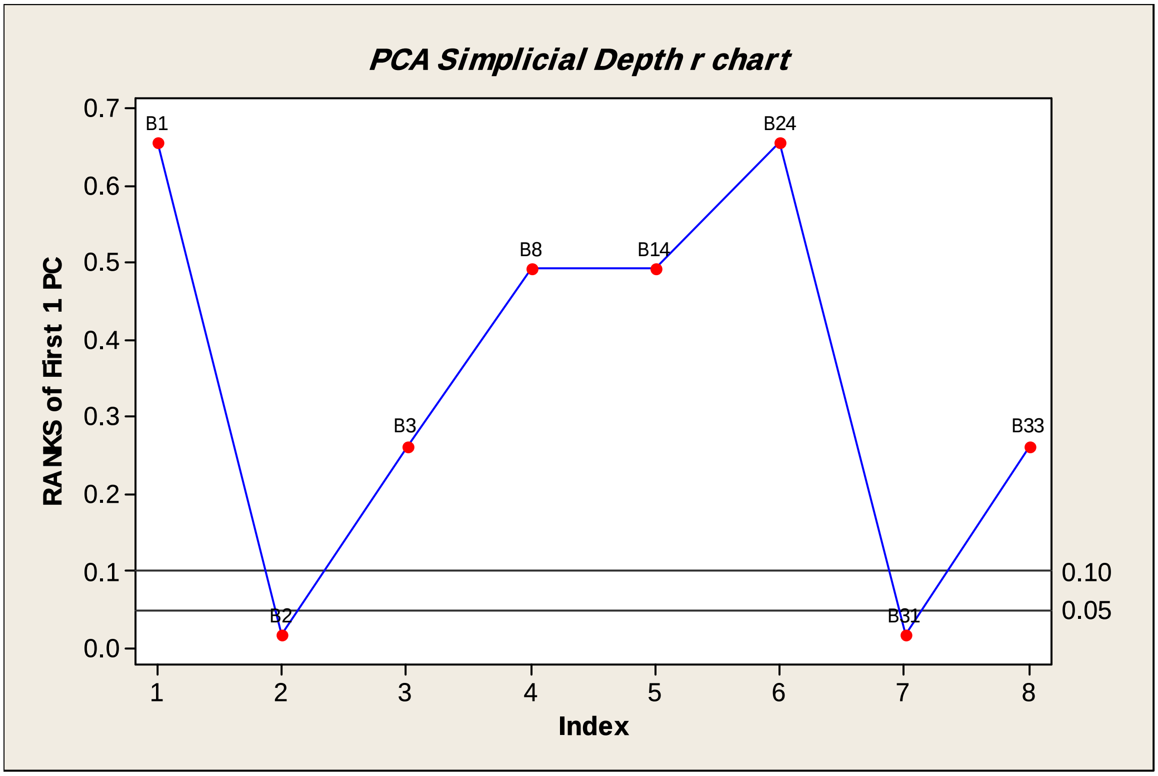

Figure 4.3e PCA Simplicial Depth r chart using the first PC for *n*=61.

Using the HDS of n = 61, the following points from the 8 new observations were identified as out of control by the control chart when selecting the first PC:

At α = 0.05: Points B2 and B31

At α = 0.10: Points B2 and B31

PCA Simplicial Depth r chart for the last PC of the 8 NEW observations for *n*=61 (α = 0.05 and 0.10)

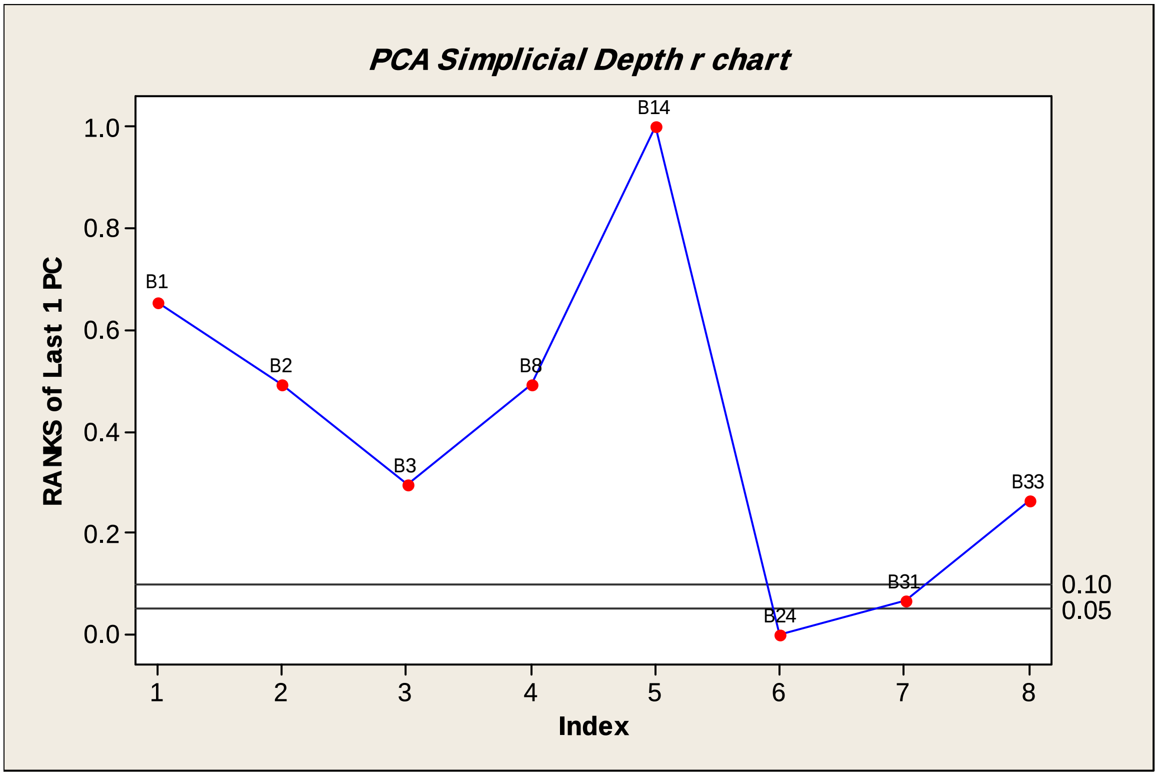

Figure 4.3f PCA Simplicial Depth r chart using the last PC for n = 61.

Using the HDS of n = 61, the following points from the 8 new observations were identified as out of control by the control chart when selecting the last PC:

At α = 0.05: Points B24

At α = 0.10: Points B24 and B31

Fruit Juice Summary Table

Table 4.3i 33 NEW points with the out of control observations labeled **X(*n*=36)**, **Y(*n*=47)** and **Z(*n*=61)**.

	$\alpha = 0.05$		$\alpha = 0.10$	
Name of point	**First PC**	**Last PC**	**First PC**	**Last PC**
B1				
B2	XYZ		XYZ	
B3				
B4				
B5				
B6				
B7				
B8				
B9				
B10				
B11				
B12				
B13				
B14				X
B15				
B16				
B17				
B18				
B19				
B20				
B21				
B22				
B23				
B24		YZ		XYZ
B25				
B26				
B27				
B28				
B29				
B30				
B31	XYZ	Y	XYZ	YZ
B32				
B33			X	

4.4 Industrial Data Set

Our third data set with 8 process variables consists of 68 observations with the first 7 observations representing the HDS. As previously mentioned, this data will be generically described due to a non-disclosure agreement as the Industrial Data Set with process variables A, B, C, D, E, F, G and H. Table 4.4a lists the 7 in control observations that will be used for the HDS.

Historical Data Set (HDS) Original 7 observations

Table 4.4a Historical Data Set of the Industrial Data with n = 7.

Obs	A	B	C	D	E	F	G	H
1	7.613700	0.084000	7.618567	0.090533	0.035167	10.776467	0.018233	0.019233
2	7.615567	0.063000	7.618767	0.069733	0.031700	10.798833	0.019067	0.021533
3	7.616800	0.057433	7.618600	0.082867	0.044667	10.798133	0.019567	0.019333
4	7.617167	0.081600	7.619433	0.080100	0.024033	10.773800	0.018533	0.017533
5	7.616633	0.050733	7.619333	0.081267	0.031400	10.786900	0.019200	0.020967
6	7.617433	0.036033	7.613933	0.057700	0.015133	10.773467	0.017400	0.018167
7	7.617300	0.058567	7.618700	0.089433	0.037467	10.779933	0.016867	0.018467

61 NEW OBSERVATIONS for *n*=7

Table 4.4b The 61 NEW observations for n = 7

Obs	A	B	C	D	E	F	G	H
1	7.622800	0.034933	7.619233	0.043767	0.015967	10.773633	0.016167	0.016967
2	7.620267	0.067133	7.611900	0.063000	0.048633	10.735267	0.020967	0.019600
3	7.620967	0.090533	7.614500	0.095833	0.053167	10.765933	0.019400	0.020133
4	7.619400	0.052767	7.607233	0.067767	0.031933	10.702067	0.019967	0.020700
5	7.617867	0.075567	7.609533	0.033467	0.051167	10.715400	0.019667	0.020767
6	7.618633	0.092600	7.614867	0.111133	0.057233	10.709800	0.018200	0.019667
7	7.614633	0.071733	7.608867	0.055933	0.018400	10.719700	0.019767	0.020167
8	7.617933	0.086933	7.607833	0.076200	0.023567	10.688367	0.019667	0.021233
9	7.616167	0.107667	7.611533	0.074700	0.027200	10.709867	0.021367	0.021167
10	7.612300	0.096167	7.614900	0.069367	0.027467	10.697233	0.019667	0.020433
11	7.613400	0.089067	7.618333	0.108433	0.050333	10.694833	0.020233	0.021167
12	7.614133	0.100833	7.618367	0.097367	0.044067	10.697700	0.018933	0.021133
13	7.620733	0.067367	7.616600	0.076267	0.032000	10.713433	0.022167	0.019200
14	7.621767	0.066400	7.619533	0.084300	0.045633	10.744067	0.019933	0.020933
15	7.619567	0.104600	7.617633	0.073300	0.051300	10.712967	0.020700	0.020867
16	7.618900	0.042300	7.613600	0.052267	0.018333	10.721133	0.020667	0.021000
17	7.615733	0.027967	7.618533	0.035233	0.025667	10.678367	0.020467	0.020433
18	7.619467	0.057033	7.615967	0.036267	0.023433	10.736900	0.019967	0.020733
19	7.618967	0.087500	7.613500	0.056767	0.036300	10.728433	0.019300	0.017900
20	7.618867	0.052933	7.618900	0.065267	0.022967	10.713033	0.018400	0.019033
21	7.617233	0.095967	7.623567	0.098000	0.046000	10.697333	0.016800	0.015367
22	7.617200	0.072933	7.613000	0.071967	0.016567	10.687100	0.018233	0.016833
23	7.613233	0.087067	7.618033	0.075367	0.046200	10.724900	0.017867	0.019033
24	7.619700	0.078700	7.615533	0.048800	0.055667	10.732933	0.018067	0.020367
25	7.616400	0.065100	7.616600	0.076733	0.036233	10.708367	0.018400	0.019900
26	7.622567	0.098667	7.622233	0.102000	0.057333	10.736000	0.020433	0.019800
27	7.627033	0.068000	7.619200	0.070467	0.047467	10.758900	0.016733	0.019400
28	7.628033	0.056167	7.621900	0.111667	0.033467	10.781667	0.019167	0.020300
29	7.628300	0.113400	7.619467	0.074900	0.045733	10.785067	0.019933	0.019833
30	7.626333	0.091500	7.619333	0.060533	0.038667	10.803600	0.021767	0.017433
31	7.624533	0.039567	7.620700	0.071600	0.030000	10.758900	0.019000	0.021333
32	7.626367	0.051067	7.621167	0.048167	0.013933	10.705267	0.018867	0.021933
33	7.627567	0.087333	7.620400	0.081333	0.040233	10.742500	0.018433	0.021600
34	7.621300	0.077567	7.621600	0.110000	0.039833	10.742167	0.018267	0.021133
35	7.619533	0.090667	7.620333	0.068700	0.027033	10.803533	0.018600	0.019100
36	7.619533	0.102233	7.621233	0.117433	0.027067	10.736667	0.019133	0.019500
37	7.618733	0.112133	7.617400	0.093633	0.022500	10.748600	0.019233	0.018200

38	7.621567	0.083733	7.616400	0.100900	0.038100	10.785900	0.018267	0.020533
39	7.622233	0.064000	7.619567	0.071600	0.040833	10.728300	0.019267	0.018800
40	7.624467	0.069900	7.613433	0.052800	0.044367	10.721600	0.019333	0.020800
41	7.618267	0.092100	7.615733	0.047600	0.045500	10.710833	0.019367	0.016100
42	7.619567	0.139700	7.608900	0.028933	0.068167	10.722467	0.019967	0.019333
43	7.622600	0.071900	7.613867	0.045567	0.049767	10.721467	0.020433	0.019333
44	7.619500	0.111467	7.615967	0.081667	0.045053	10.748267	0.020167	0.017567
45	7.615967	0.095867	7.614267	0.117333	0.041933	10.701933	0.021233	0.019600
46	7.620767	0.097867	7.618233	0.100867	0.039400	10.748567	0.020367	0.020233
47	7.617967	0.091400	7.615800	0.114400	0.030633	10.729867	0.020633	0.019433
48	7.617200	0.072900	7.613967	0.096100	0.068700	10.746967	0.019600	0.018333
49	7.619733	0.063000	7.614100	0.105967	0.048900	10.769400	0.020467	0.018533
50	7.619100	0.072700	7.616800	0.097000	0.035367	10.761533	0.018267	0.016767
51	7.622067	0.047033	7.620067	0.110567	0.052467	10.734700	0.018167	0.016233
52	7.623500	0.104867	7.621467	0.100533	0.030400	10.700100	0.018333	0.018033
53	7.617533	0.100433	7.618833	0.082033	0.035933	10.704533	0.016933	0.017533
54	7.617933	0.098200	7.617767	0.091200	0.032267	10.723000	0.019133	0.020800
55	7.615267	0.106500	7.612700	0.113633	0.040500	10.703900	0.016400	0.014433
56	7.618400	0.088967	7.613533	0.063767	0.040567	10.687200	0.017433	0.017700
57	7.622200	0.098033	7.617700	0.094033	0.020433	10.687600	0.020300	0.019300
58	7.620300	0.129400	7.620533	0.106667	0.024033	10.684333	0.019700	0.019333
59	7.619367	0.118567	7.614933	0.103247	0.026133	10.682700	0.019900	0.019533
60	7.624167	0.113600	7.615867	0.051700	0.040300	10.858100	0.019767	0.020433
61	7.617533	0.133567	7.613300	0.095200	0.029667	10.677167	0.020233	0.019167

When applying PCA, we find that A and C are singular as such, we will remove them from the analysis. In the literature when using PCA in multivariate quality it is stated that the variable(s) causing the singularity need to be removed in order to proceed. (Mason and Young 2002) The process variables that will be analyzed are B, D, E, F, G and H. From this point on, any reference to the process variables of the HDS and the new observations will be solely towards these six that are in the analysis which we will identify as X1, X2, X3, X4, X5 and X6.

Historical Data Set (HDS) Original 7 observations with 6 process variables

Table 4.4c Historical Data Set with the six variables of the Industrial Data with *n* = 7.

Obs	X1	X2	X3	X4	X5	X6
1	0.084000	0.090533	0.035167	10.776467	0.018233	0.019233
2	0.063000	0.069733	0.031700	10.798833	0.019067	0.021533
3	0.057433	0.082867	0.044667	10.798133	0.019567	0.019333
4	0.081600	0.080100	0.024033	10.773800	0.018533	0.017533
5	0.050733	0.081267	0.031400	10.786900	0.019200	0.020967
6	0.036033	0.057700	0.015133	10.773467	0.017400	0.018167
7	0.058567	0.089433	0.037467	10.779933	0.016867	0.018467

Eigenanalysis of the Industrial Correlation Matrix with *n*=7

Table 4.4d The Eigenanalysis of the HDS with *n* = 7

	PC1	PC2	PC3	PC4	PC5	PC6
Eigenvalue	2.7983	1.9425	0.7109	0.3590	0.1773	0.0120
Proportion	0.466	0.324	0.118	0.060	0.030	0.002
Cumulative	0.466	0.790	0.909	0.968	0.998	1.000

61 NEW OBSERVATIONS for *n*=7

Table 4.4e The 61 NEW observations for $n = 7$

Obs	Name of each point	X1	X2	X3	X4	X5	X6
1	C1	0.034933	0.043767	0.015967	10.773633	0.016167	0.016967
2	C2	0.067133	0.063000	0.048633	10.735267	0.020967	0.019600
3	C3	0.090533	0.095833	0.053167	10.765933	0.019400	0.020133
4	C4	0.052767	0.067767	0.031933	10.702067	0.019967	0.020700
5	C5	0.075567	0.033467	0.051167	10.715400	0.019667	0.020767
6	C6	0.092600	0.111133	0.057233	10.709800	0.018200	0.019667
7	C7	0.071733	0.055933	0.018400	10.719700	0.019767	0.020167
8	C8	0.086933	0.076200	0.023567	10.688367	0.019667	0.021233
9	C9	0.107667	0.074700	0.027200	10.709867	0.021367	0.021167
10	C10	0.096167	0.069367	0.027467	10.697233	0.019667	0.020433
11	C11	0.089067	0.108433	0.050333	10.694833	0.020233	0.021167
12	C12	0.100833	0.097367	0.044067	10.697700	0.018933	0.021133
13	C13	0.067367	0.076267	0.032000	10.713433	0.022167	0.019200
14	C14	0.066400	0.084300	0.045633	10.744067	0.019933	0.020933
15	C15	0.104600	0.073300	0.051300	10.712967	0.020700	0.020867
16	C16	0.042300	0.052267	0.018333	10.721133	0.020667	0.021000
17	C17	0.027967	0.035233	0.025667	10.678367	0.020467	0.020433
18	C18	0.057033	0.036267	0.023433	10.736900	0.019967	0.020733
19	C19	0.087500	0.056767	0.036300	10.728433	0.019300	0.017900
20	C20	0.052933	0.065267	0.022967	10.713033	0.018400	0.019033
21	C21	0.095967	0.098000	0.046000	10.697333	0.016800	0.015367
22	C22	0.072933	0.071967	0.016567	10.687100	0.018233	0.016833
23	C23	0.087067	0.075367	0.046200	10.724900	0.017867	0.019033
24	C24	0.078700	0.048800	0.055667	10.732933	0.018067	0.020367
25	C25	0.065100	0.076733	0.036233	10.708367	0.018400	0.019900
26	C26	0.098667	0.102000	0.057333	10.736000	0.020433	0.019800
27	C27	0.068000	0.070467	0.047467	10.758900	0.016733	0.019400
28	C28	0.056167	0.111667	0.033467	10.781667	0.019167	0.020300
29	C29	0.113400	0.074900	0.045733	10.785067	0.019933	0.019833
30	C30	0.091500	0.060533	0.038667	10.803600	0.021767	0.017433
31	C31	0.039567	0.071600	0.030000	10.758900	0.019000	0.021333
32	C32	0.051067	0.048167	0.013933	10.705267	0.018867	0.021933
33	C33	0.087333	0.081333	0.040233	10.742500	0.018433	0.021600
34	C34	0.077567	0.110000	0.039833	10.742167	0.018267	0.021133
35	C35	0.090667	0.068700	0.027033	10.803533	0.018600	0.019100

36	**C36**	0.102233	0.117433	0.027067	10.736667	0.019133	0.019500
37	**C37**	0.112133	0.093633	0.022500	10.748600	0.019233	0.018200
38	**C38**	0.083733	0.100900	0.038100	10.785900	0.018267	0.020533
39	**C39**	0.064000	0.071600	0.040833	10.728300	0.019267	0.018800
40	**C40**	0.069900	0.052800	0.044367	10.721600	0.019333	0.020800
41	**C41**	0.092100	0.047600	0.045500	10.710833	0.019367	0.016100
42	**C42**	0.139700	0.028933	0.068167	10.722467	0.019967	0.019333
43	**C43**	0.071900	0.045567	0.049767	10.721467	0.020433	0.019333
44	**C44**	0.111467	0.081667	0.045053	10.748267	0.020167	0.017567
45	**C45**	0.095867	0.117333	0.041933	10.701933	0.021233	0.019600
46	**C46**	0.097867	0.100867	0.039400	10.748567	0.020367	0.020233
47	**C47**	0.091400	0.114400	0.030633	10.729867	0.020633	0.019433
48	**C48**	0.072900	0.096100	0.068700	10.746967	0.019600	0.018333
49	**C49**	0.063000	0.105967	0.048900	10.769400	0.020467	0.018533
50	**C50**	0.072700	0.097000	0.035367	10.761533	0.018267	0.016767
51	**C51**	0.047033	0.110567	0.052467	10.734700	0.018167	0.016233
52	**C52**	0.104867	0.100533	0.030400	10.700100	0.018333	0.018033
53	**C53**	0.100433	0.082033	0.035933	10.704533	0.016933	0.017533
54	**C54**	0.098200	0.091200	0.032267	10.723000	0.019133	0.020800
55	**C55**	0.106500	0.113633	0.040500	10.703900	0.016400	0.014433
56	**C56**	0.088967	0.063767	0.040567	10.687200	0.017433	0.017700
57	**C57**	0.098033	0.094033	0.020433	10.687600	0.020300	0.019300
58	**C58**	0.129400	0.106667	0.024033	10.684333	0.019700	0.019333
59	**C59**	0.118567	0.103247	0.026133	10.682700	0.019900	0.019533
60	**C60**	0.113600	0.051700	0.040300	10.858100	0.019767	0.020433
61	**C61**	0.133567	0.095200	0.029667	10.677167	0.020233	0.019167

PCA Simplicial Depth r chart for the first PC of the 61 NEW observations for *n*=7 (α = 0.05 and 0.10)

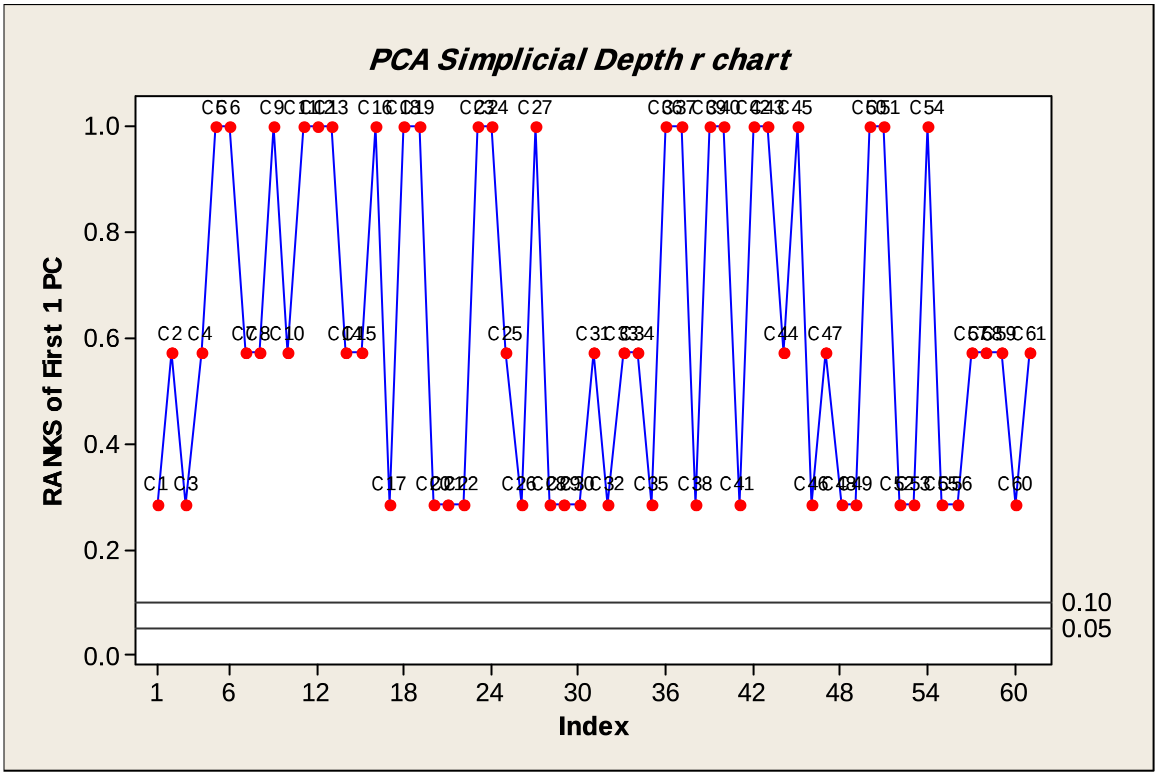

Figure 4.4a PCA Simplicial Depth r chart using the first PC *n* = 7.

Using the HDS of *n* = 7, the following points from the 61 new observations were identified as out of control by the control chart when selecting the first PC:

At α = 0.05: No Points

At α = 0.10: No Points

PCA Simplicial Depth r chart for the last PC of the 61 NEW observations for n=7 (α = 0.05 and 0.10)

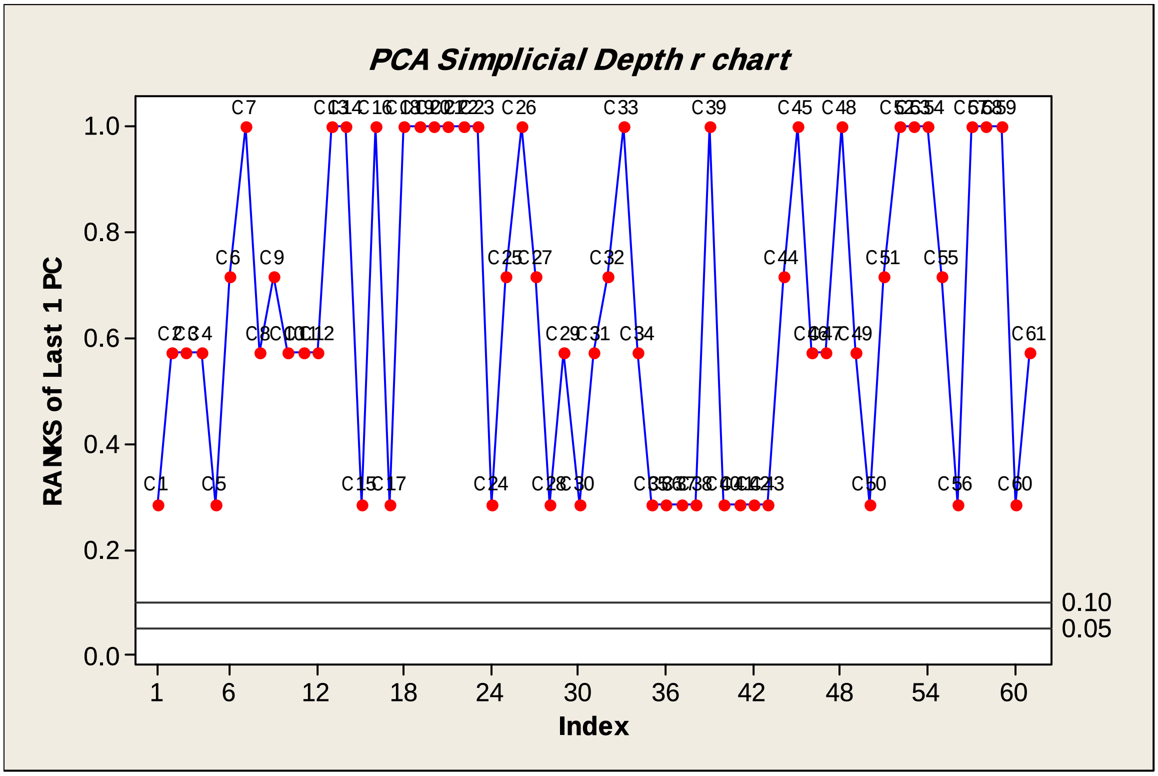

Figure 4.4b PCA Simplicial Depth r chart using the last PC for n = 7.

For the HDS of n = 7, the following points from the 61 new observations were identified as out of control by the control chart when selecting the last PC:

At α = 0.05: No Points

At α = 0.10: No Points

Since there were no signals from any of the charts, yet we know a priori that at some point the process was out of control, we will augment the HDS by selecting 20 points from the 61 by picking every third point starting with C3. To determine if this set is “good” we will now run the original HDS of $n = 7$ against these 20: C3, C6, C9, C12, C15, C18, C21, C24, C27, C30, C33, C36, C39, C42, C45, C48, C51, C54, C57, C60 which are listed in Table 4.4f. Any points identified from this subset as out of control will be discarded, and we will run more points until we find 20 in control points. The 20 that we find in control will be used to augment the HDS from 7 to 27.

20 OBSERVATIONS out of 61 that were selected by choosing every third point for *n*=7.

Table 4.4f The 20 NEW points for HDS with *n* = 7.

Obs	Name of each point	X1	X2	X3	X4	X5	X6
1	C3	0.090533	0.095833	0.053167	10.765933	0.019400	0.020133
2	C6	0.092600	0.111133	0.057233	10.709800	0.018200	0.019667
3	C9	0.107667	0.074700	0.027200	10.709867	0.021367	0.021167
4	C12	0.100833	0.097367	0.044067	10.697700	0.018933	0.021133
5	C15	0.104600	0.073300	0.051300	10.712967	0.020700	0.020867
6	C18	0.057033	0.036267	0.023433	10.736900	0.019967	0.020733
7	C21	0.095967	0.098000	0.046000	10.697333	0.016800	0.015367
8	C24	0.078700	0.048800	0.055667	10.732933	0.018067	0.020367
9	C27	0.068000	0.070467	0.047467	10.758900	0.016733	0.019400
10	C30	0.091500	0.060533	0.038667	10.803600	0.021767	0.017433
11	C33	0.087333	0.081333	0.040233	10.742500	0.018433	0.021600
12	C36	0.102233	0.117433	0.027067	10.736667	0.019133	0.019500
13	C39	0.064000	0.071600	0.040833	10.728300	0.019267	0.018800
14	C42	0.139700	0.028933	0.068167	10.722467	0.019967	0.019333
15	C45	0.095867	0.117333	0.041933	10.701933	0.021233	0.019600
16	C48	0.072900	0.096100	0.068700	10.746967	0.019600	0.018333
17	C51	0.047033	0.110567	0.052467	10.734700	0.018167	0.016233
18	C54	0.098200	0.091200	0.032267	10.723000	0.019133	0.020800
19	C57	0.098033	0.094033	0.020433	10.687600	0.020300	0.019300
20	C60	0.113600	0.051700	0.040300	10.858100	0.019767	0.020433

PCA Simplicial Depth r chart for the first PC of the 20 NEW observations for *n*=7 (α = 0.05 and 0.10)

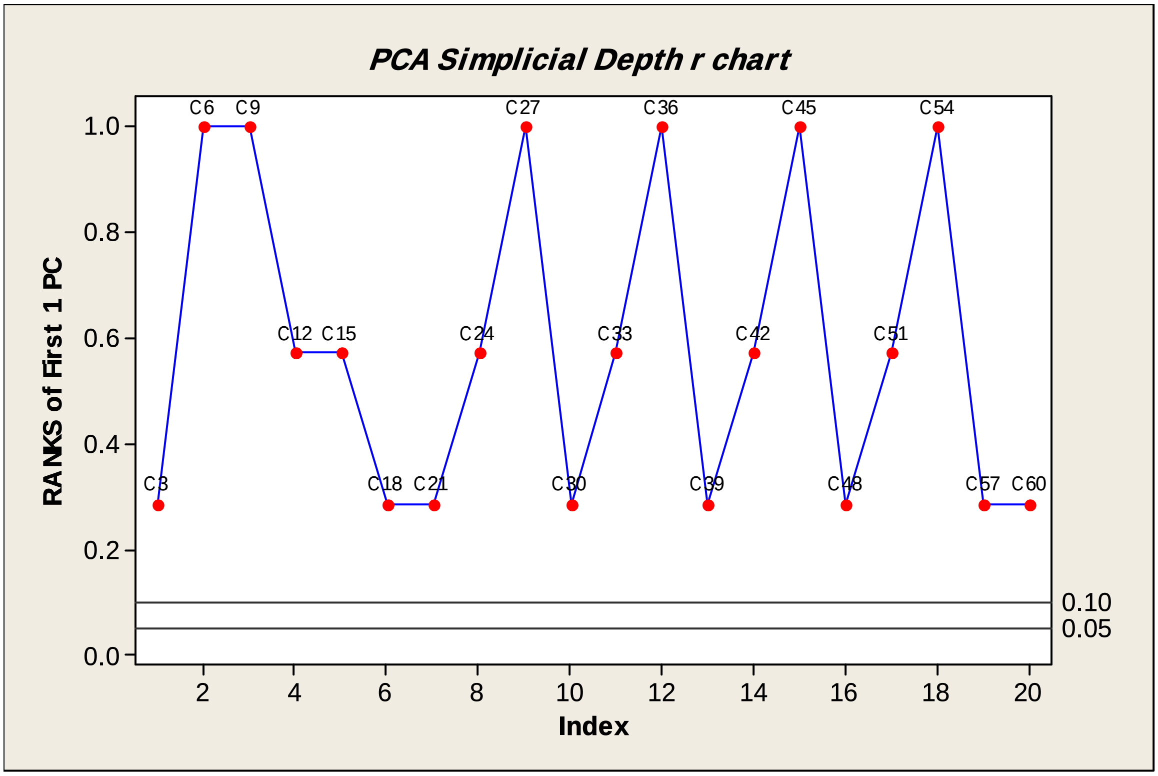

Figure 4.4c PCA Simplicial Depth r chart using the first PC for *n*=7.

Using the HDS of n = 7, the following points from the 20 new observations were identified as out of control by the control chart when selecting the first PC:

At α = 0.05: No Points

At α = 0.10: No Points

PCA Simplicial Depth r chart for the last PC of the 20 NEW observations for *n*=7 (α = 0.05 and 0.10)

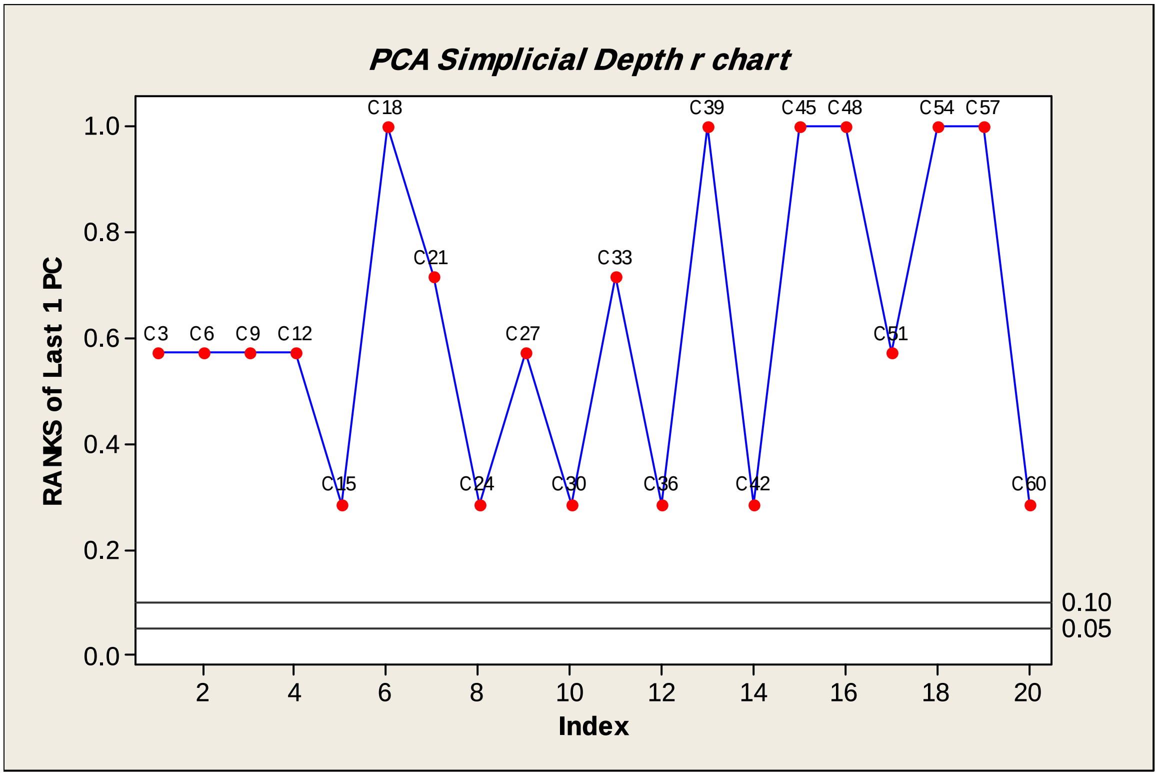

Figure 4.4d PCA Simplicial Depth r chart using the last PC for *n*=7.

For the HDS of n = 7, the following points from the 20 new observations were identified as out of control by the control chart when selecting the last PC:

At α = 0.05: No Points

At α = 0.10: No Points

Since there were no out of control points from these 20 points, we will augment the HDS by including these 20 points in order to identify the signal(s) that is/are missed as a result of the small HDS that was made available to us. (see Table 4.4g)

Industrial Data – Extended Historical Data Set (HDS) with *n*=27 (7 plus 20 identified as in-control from the control charts with *n*=7)

Table 4.4g Extended Historical Data Set of the Industrial Data with *n* = 27.

Obs	X1	X2	X3	X4	X5	X6
1	0.084000	0.090533	0.035167	10.776467	0.018233	0.019233
2	0.063000	0.069733	0.031700	10.798833	0.019067	0.021533
3	0.057433	0.082867	0.044667	10.798133	0.019567	0.019333
4	0.081600	0.080100	0.024033	10.773800	0.018533	0.017533
5	0.050733	0.081267	0.031400	10.786900	0.019200	0.020967
6	0.036033	0.057700	0.015133	10.773467	0.017400	0.018167
7	0.058567	0.089433	0.037467	10.779933	0.016867	0.018467
8	0.090533	0.095833	0.053167	10.765933	0.019400	0.020133
9	0.092600	0.111133	0.057233	10.709800	0.018200	0.019667
10	0.107667	0.074700	0.027200	10.709867	0.021367	0.021167
11	0.100833	0.097367	0.044067	10.697700	0.018933	0.021133
12	0.104600	0.073300	0.051300	10.712967	0.020700	0.020867
13	0.057033	0.036267	0.023433	10.736900	0.019967	0.020733
14	0.095967	0.098000	0.046000	10.697333	0.016800	0.015367
15	0.078700	0.048800	0.055667	10.732933	0.018067	0.020367
16	0.068000	0.070467	0.047467	10.758900	0.016733	0.019400
17	0.091500	0.060533	0.038667	10.803600	0.021767	0.017433
18	0.087333	0.081333	0.040233	10.742500	0.018433	0.021600
19	0.102233	0.117433	0.027067	10.736667	0.019133	0.019500
20	0.064000	0.071600	0.040833	10.728300	0.019267	0.018800
21	0.139700	0.028933	0.068167	10.722467	0.019967	0.019333
22	0.095867	0.117333	0.041933	10.701933	0.021233	0.019600
23	0.072900	0.096100	0.068700	10.746967	0.019600	0.018333
24	0.047033	0.110567	0.052467	10.734700	0.018167	0.016233
25	0.098200	0.091200	0.032267	10.723000	0.019133	0.020800
26	0.098033	0.094033	0.020433	10.687600	0.020300	0.019300
27	0.113600	0.051700	0.040300	10.858100	0.019767	0.020433

Eigenanalysis of the Industrial Correlation Matrix with *n*=27

Table 4.4h The Eigenanalysis of the extended HDS with n = 27.

	PC1	PC2	PC3	PC4	PC5	PC6
Eigenvalue	1.7478	1.5328	1.0189	0.6808	0.5871	0.4324
Proportion	0.291	0.255	0.170	0.113	0.098	0.072
Cumulative	0.291	0.547	0.717	0.830	0.928	1.000

41 NEW OBSERVATIONS for *n*=27

Table 4.4i The 41 NEW observations for n = 27

Obs	Name of each point	X1	X2	X3	X4	X5	X6
1	C1	0.034933	0.043767	0.015967	10.773633	0.016167	0.016967
2	C2	0.067133	0.063000	0.048633	10.735267	0.020967	0.019600
3	C4	0.052767	0.067767	0.031933	10.702067	0.019967	0.020700
4	C5	0.075567	0.033467	0.051167	10.715400	0.019667	0.020767
5	C7	0.071733	0.055933	0.018400	10.719700	0.019767	0.020167
6	C8	0.086933	0.076200	0.023567	10.688367	0.019667	0.021233
7	C10	0.096167	0.069367	0.027467	10.697233	0.019667	0.020433
8	C11	0.089067	0.108433	0.050333	10.694833	0.020233	0.021167
9	C13	0.067367	0.076267	0.032000	10.713433	0.022167	0.019200
10	C14	0.066400	0.084300	0.045633	10.744067	0.019933	0.020933
11	C16	0.042300	0.052267	0.018333	10.721133	0.020667	0.021000
12	C17	0.027967	0.035233	0.025667	10.678367	0.020467	0.020433
13	C19	0.087500	0.056767	0.036300	10.728433	0.019300	0.017900
14	C20	0.052933	0.065267	0.022967	10.713033	0.018400	0.019033
15	C22	0.072933	0.071967	0.016567	10.687100	0.018233	0.016833
16	C23	0.087067	0.075367	0.046200	10.724900	0.017867	0.019033
17	C25	0.065100	0.076733	0.036233	10.708367	0.018400	0.019900
18	C26	0.098667	0.102000	0.057333	10.736000	0.020433	0.019800
19	C28	0.056167	0.111667	0.033467	10.781667	0.019167	0.020300
20	C29	0.113400	0.074900	0.045733	10.785067	0.019933	0.019833
21	C31	0.039567	0.071600	0.030000	10.758900	0.019000	0.021333
22	C32	0.051067	0.048167	0.013933	10.705267	0.018867	0.021933

23	**C34**	0.077567	0.110000	0.039833	10.742167	0.018267	0.021133
24	**C35**	0.090667	0.068700	0.027033	10.803533	0.018600	0.019100
25	**C37**	0.112133	0.093633	0.022500	10.748600	0.019233	0.018200
26	**C38**	0.083733	0.100900	0.038100	10.785900	0.018267	0.020533
27	**C40**	0.069900	0.052800	0.044367	10.721600	0.019333	0.020800
28	**C41**	0.092100	0.047600	0.045500	10.710833	0.019367	0.016100
29	**C43**	0.071900	0.045567	0.049767	10.721467	0.020433	0.019333
30	**C44**	0.111467	0.081667	0.045053	10.748267	0.020167	0.017567
31	**C46**	0.097867	0.100867	0.039400	10.748567	0.020367	0.020233
32	**C47**	0.091400	0.114400	0.030633	10.729867	0.020633	0.019433
33	**C49**	0.063000	0.105967	0.048900	10.769400	0.020467	0.018533
34	**C50**	0.072700	0.097000	0.035367	10.761533	0.018267	0.016767
35	**C52**	0.104867	0.100533	0.030400	10.700100	0.018333	0.018033
36	**C53**	0.100433	0.082033	0.035933	10.704533	0.016933	0.017533
37	**C55**	0.106500	0.113633	0.040500	10.703900	0.016400	0.014433
38	**C56**	0.088967	0.063767	0.040567	10.687200	0.017433	0.017700
39	**C58**	0.129400	0.106667	0.024033	10.684333	0.019700	0.019333
40	**C59**	0.118567	0.103247	0.026133	10.682700	0.019900	0.019533
41	**C61**	0.133567	0.095200	0.029667	10.677167	0.020233	0.019167

PCA Simplicial Depth r chart for the first 2 PCs of the 41 NEW observations for *n*=27 (α = 0.05 and 0.10)

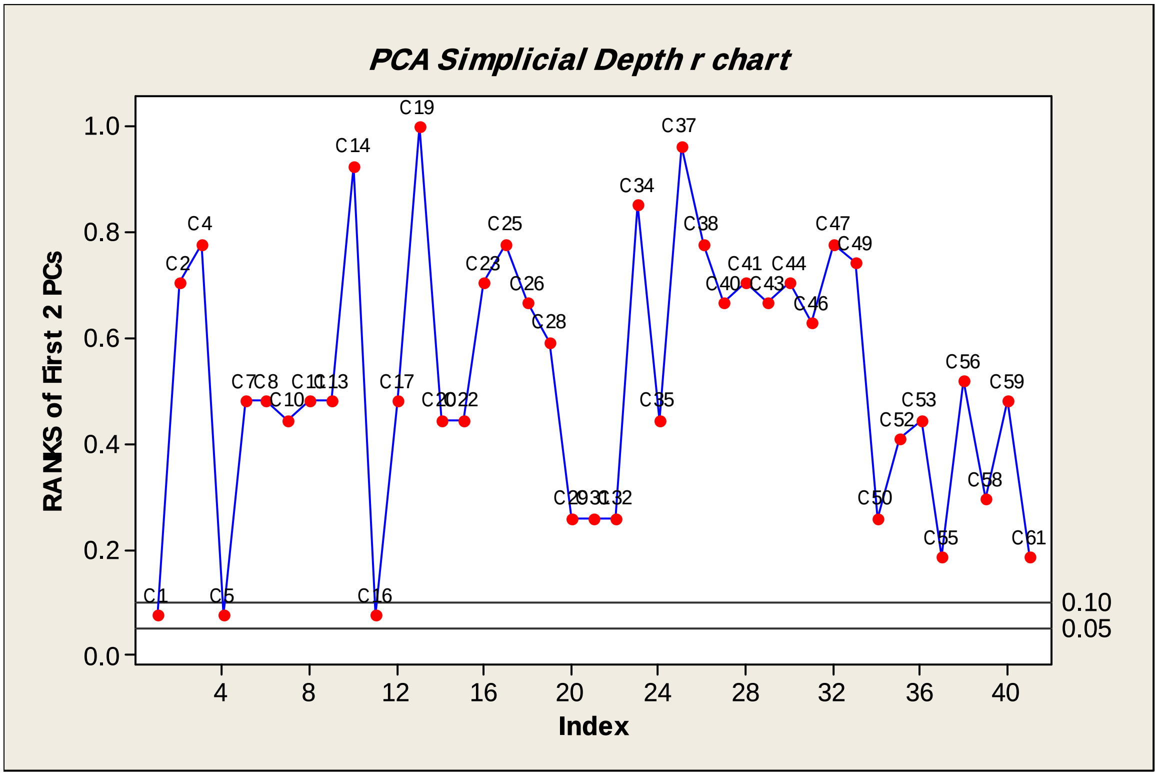

Figure 4.4e PCA Simplicial Depth r chart using the first 2 PCs for *n*=27.

Using the HDS of n = 27, the following points from the 41 new observations were identified as out of control by the control chart when selecting the first 2 PCs:

At α = 0.05: No Points

At α = 0.10: Points C1, C5 and C16

PCA Simplicial Depth r chart for the last PC of the 41 NEW observations for *n*=27 (α = 0.05 and 0.10)

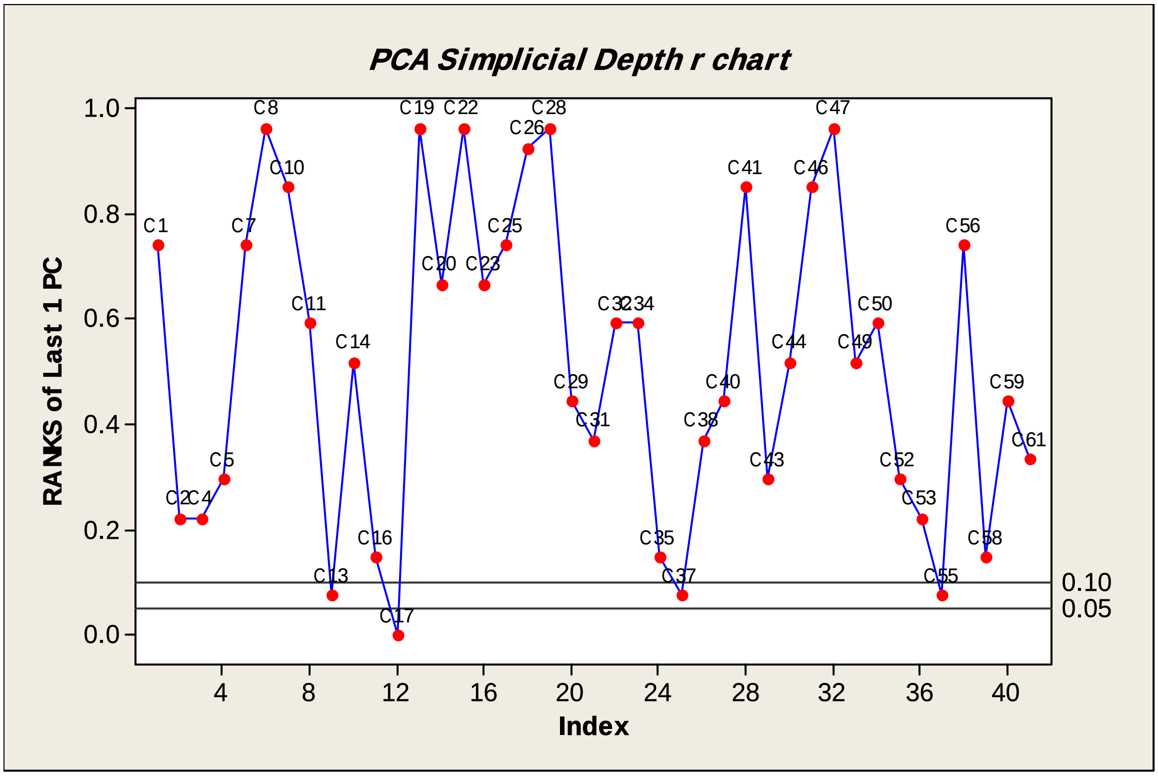

Figure 4.4f PCA Simplicial Depth r chart using the last PC for *n*=27.

Using the HDS of n = 27, the following points from the 41 new observations were identified as out of control by the control chart when selecting the last PC:

At α = 0.05: Point C17

At α = 0.10: Points C13, C17, C37 and C55

From the 29 points that have been identified as in control from the control charts, we will choose the first 20 and augment the HDS to 47. [Some scenarios included more PCs than shown here in generating possible false alarms.] These 20 points are C2, C4, C7, C8, C10, C11, C14, C19, C20, C23, C25, C26, C28, C29, C31, C32, C34, C35, C40 and C41. Table 4.4j illustrates the extended HDS with n = 47 with these additional 20 points.

Industrial Data – Extended Historical Data Set (HDS) with *n*=47 (27 plus 20 identified as in-control from the control charts with *n*=27)

Table 4.4j Extended Historical Data Set of the Industrial Data with n = 47.

Obs	X1	X2	X3	X4	X5	X6
1	0.084000	0.090533	0.035167	10.776467	0.018233	0.019233
2	0.063000	0.069733	0.031700	10.798833	0.019067	0.021533
3	0.057433	0.082867	0.044667	10.798133	0.019567	0.019333
4	0.081600	0.080100	0.024033	10.773800	0.018533	0.017533
5	0.050733	0.081267	0.031400	10.786900	0.019200	0.020967
6	0.036033	0.057700	0.015133	10.773467	0.017400	0.018167
7	0.058567	0.089433	0.037467	10.779933	0.016867	0.018467
8	0.090533	0.095833	0.053167	10.765933	0.019400	0.020133
9	0.092600	0.111133	0.057233	10.709800	0.018200	0.019667
10	0.107667	0.074700	0.027200	10.709867	0.021367	0.021167
11	0.100833	0.097367	0.044067	10.697700	0.018933	0.021133
12	0.104600	0.073300	0.051300	10.712967	0.020700	0.020867
13	0.057033	0.036267	0.023433	10.736900	0.019967	0.020733
14	0.095967	0.098000	0.046000	10.697333	0.016800	0.015367
15	0.078700	0.048800	0.055667	10.732933	0.018067	0.020367
16	0.068000	0.070467	0.047467	10.758900	0.016733	0.019400
17	0.091500	0.060533	0.038667	10.803600	0.021767	0.017433
18	0.087333	0.081333	0.040233	10.742500	0.018433	0.021600
19	0.102233	0.117433	0.027067	10.736667	0.019133	0.019500
20	0.064000	0.071600	0.040833	10.728300	0.019267	0.018800
21	0.139700	0.028933	0.068167	10.722467	0.019967	0.019333

22	0.095867	0.117333	0.041933	10.701933	0.021233	0.019600
23	0.072900	0.096100	0.068700	10.746967	0.019600	0.018333
24	0.047033	0.110567	0.052467	10.734700	0.018167	0.016233
25	0.098200	0.091200	0.032267	10.723000	0.019133	0.020800
26	0.098033	0.094033	0.020433	10.687600	0.020300	0.019300
27	0.113600	0.051700	0.040300	10.858100	0.019767	0.020433
28	0.067133	0.063000	0.048633	10.735267	0.020967	0.019600
29	0.052767	0.067767	0.031933	10.702067	0.019967	0.020700
30	0.071733	0.055933	0.018400	10.719700	0.019767	0.020167
31	0.086933	0.076200	0.023567	10.688367	0.019667	0.021233
32	0.096167	0.069367	0.027467	10.697233	0.019667	0.020433
33	0.089067	0.108433	0.050333	10.694833	0.020233	0.021167
34	0.066400	0.084300	0.045633	10.744067	0.019933	0.020933
35	0.087500	0.056767	0.036300	10.728433	0.019300	0.017900
36	0.052933	0.065267	0.022967	10.713033	0.018400	0.019033
37	0.087067	0.075367	0.046200	10.724900	0.017867	0.019033
38	0.065100	0.076733	0.036233	10.708367	0.018400	0.019900
39	0.098667	0.102000	0.057333	10.736000	0.020433	0.019800
40	0.056167	0.111667	0.033467	10.781667	0.019167	0.020300
41	0.113400	0.074900	0.045733	10.785067	0.019933	0.019833
42	0.039567	0.071600	0.030000	10.758900	0.019000	0.021333
43	0.051067	0.048167	0.013933	10.705267	0.018867	0.021933
44	0.077567	0.110000	0.039833	10.742167	0.018267	0.021133
45	0.090667	0.068700	0.027033	10.803533	0.018600	0.019100
46	0.069900	0.052800	0.044367	10.721600	0.019333	0.020800
47	0.092100	0.047600	0.045500	10.710833	0.019367	0.016100

Eigenanalysis of the Industrial Correlation Matrix with *n*=47

Table 4.4k The Eigenanalysis of the extended HDS with $n = 47$.

	PC1	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**
Eigenvalue	1.5768	1.3874	1.0484	0.8645	0.6095	0.5134
Proportion	0.263	0.231	0.175	0.144	0.102	0.086
Cumulative	0.263	0.494	0.669	0.813	0.914	1.000

12 NEW OBSERVATIONS for *n*=47

Table 4.4l The 12 NEW observations for n = 47

Obs	Name of each point	X1	X2	X3	X4	X5	X6
1	C1	0.034933	0.043767	0.015967	10.773633	0.016167	0.016967
2	C5	0.075567	0.033467	0.051167	10.715400	0.019667	0.020767
3	C13	0.067367	0.076267	0.032000	10.713433	0.022167	0.019200
4	C16	0.042300	0.052267	0.018333	10.721133	0.020667	0.021000
5	C17	0.027967	0.035233	0.025667	10.678367	0.020467	0.020433
6	C22	0.072933	0.071967	0.016567	10.687100	0.018233	0.016833
7	C37	0.112133	0.093633	0.022500	10.748600	0.019233	0.018200
8	C38	0.083733	0.100900	0.038100	10.785900	0.018267	0.020533
9	C49	0.063000	0.105967	0.048900	10.769400	0.020467	0.018533
10	C55	0.106500	0.113633	0.040500	10.703900	0.016400	0.014433
11	C58	0.129400	0.106667	0.024033	10.684333	0.019700	0.019333
12	C61	0.133567	0.095200	0.029667	10.677167	0.020233	0.019167

PCA Simplicial Depth r chart for the first 2 PCs of the 12 NEW observations for *n*=47 (α = 0.05 and 0.10)

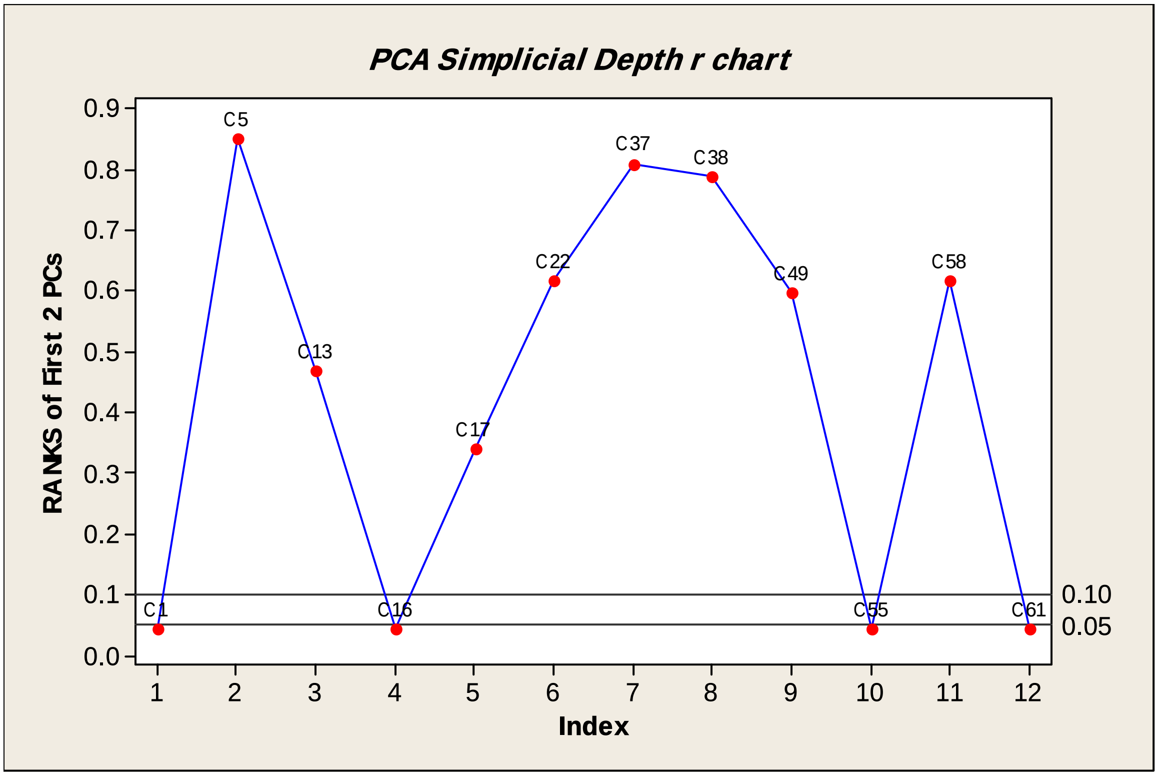

Figure 4.4g PCA Simplicial Depth r chart using the first 2 PCs for *n*=47.

Using the HDS of n = 47, the following points from the 12 new observations were identified as out of control by the control chart when selecting the first 2 PCs:

At α = 0.05: Points C1, C16, C55 and C61

At α = 0.10: Points C1, C16, C55 and C61

PCA Simplicial Depth r chart for the last PC of the 12 NEW observations for *n*=47 (α = 0.05 and 0.10)

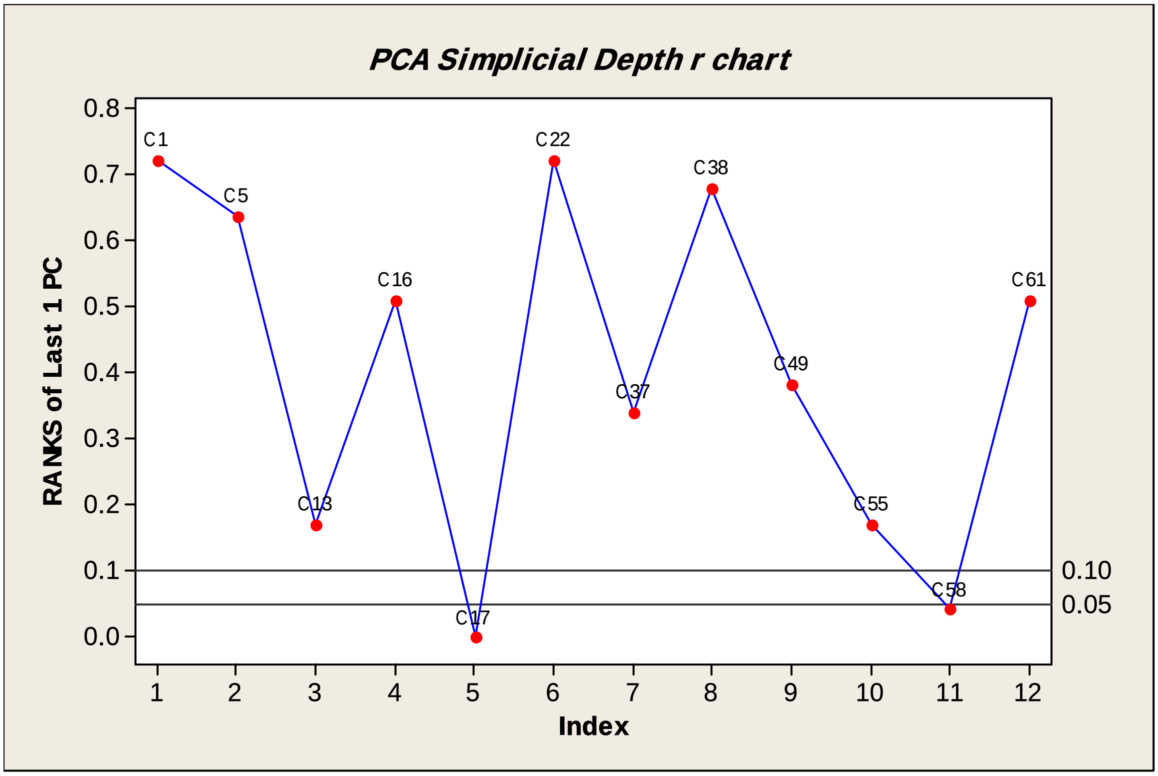

Figure 4.4h PCA Simplicial Depth r chart using the last PC for n=47.

For the HDS of n = 47, the following points from the 12 new observations were identified as out of control by the control chart when selecting the last PC:

At α = 0.05: Points C17 and C58

At α = 0.10: Points C17 and C58

As we can see in this case study, like in the others, as n increases we begin to detect signals. The final PC generates signals with one point specifically signaling consistently at alpha of 0.05. At alpha of 0.10, point C55 signals at the last PC along with the first PC. Given the low percentage variability accounted by the first PC, it is possible that this correlation shift detected by the last PCs could have resulted in a small variability shift at that point as well. It is interesting to find that the signals are detected with small alpha including C1 and C61 in the first PCs, which may be indicative of a variability shift, and C17 in the last PC, possibly due to a correlation shift. This discovery is most beneficial in MSPC given that with a fairly small sample size for the HDS, we may be able to detect both a shift in variability and a correlation using our proposed control chart scheme for the first PCs and last PCs, respectively. In real MSPC data, such as this one, we may not have the availability of large historical base data sets, and we can see that if the HDS is too small, as would be expected, we cannot detect out of control points.

Industrial Data Summary Table

Table 4.4m 61 NEW points with the out of control observations labeled **X(*n*=7)**, **Y(*n*=27)** and **Z(*n*=47)**.

	$\alpha = 0.05$		$\alpha = 0.10$	
Name of point	**First 2 PCs**	**Last PC**	**First 2 PCs**	**Last PC**
C1	Z		Y Z	
C2				
C3				
C4				
C5			Y	
C6				
C7				
C8				
C9				
C10				
C11				
C12				
C13				Y
C14				
C15				
C16	Z		Y Z	
C17		Y Z		Y Z
C18				
C19				
C20				
C21				
C22				
C23				
C24				
C25				
C26				
C27				
C28				
C29				

C30					
C31					
C32					
C33					
C34					
C35					
C36					
C37					y
C38					
C39					
C40					
C41					
C42					
C43					
C44					
C45					
C46					
C47					
C48					
C49					
C50					
C51					
C52					
C53					
C54					
C55	z			z	y
C56					
C57					
C58		z			z
C59					
C60					
C61	z			z	

4.5 Aluminum Pin Data Set

Our next data set with 6 process variables consists of 70 observations with the first 30 observations representing the HDS. The 6 process variables are Diameter 1, Diameter 2, Diameter 3, Diameter 4, Length 1 and Length 1. Table 4.5a lists the 30 in control observations that will be used for the HDS.

Historical Data Set (HDS) Original 30 observations

Table 4.5a Historical Data Set of the Aluminum Pin Data with n = 30.

Obs	Diameter1	Diameter2	Diameter3	Diameter4	Length1	Length2
1	9.99	9.97	9.96	14.97	49.89	60.02
2	9.96	9.96	9.95	14.94	49.84	60.02
3	9.97	9.96	9.95	14.95	49.85	60
4	10	9.99	9.99	14.99	49.89	60.06
5	10	9.99	9.99	14.99	49.91	60.09
6	9.99	9.99	9.98	14.99	49.91	60.08
7	10	9.99	9.99	14.98	49.91	60.08
8	10	9.99	9.99	14.99	49.89	60.09
9	9.96	9.95	9.95	14.95	50	60.15
10	9.99	9.98	9.98	14.99	49.86	60.06
11	10	9.99	9.98	14.99	49.94	60.08
12	10	9.99	9.99	14.99	49.92	60.05
13	9.97	9.96	9.96	14.96	49.9	60.02
14	9.97	9.96	9.96	14.96	49.91	60.02
15	9.97	9.97	9.96	14.97	49.9	60.01
16	9.97	9.97	9.96	14.97	49.89	60.04
17	9.98	9.97	9.96	14.96	50.01	60.13
18	9.98	9.97	9.97	14.96	49.93	60.06
19	9.98	9.98	9.97	14.98	49.93	60.02
20	9.98	9.97	9.97	14.97	49.94	60.06
21	9.98	9.97	9.97	14.97	49.93	60.06

22	9.98	9.97	9.97	14.97	49.91	60.02
23	9.98	9.97	9.96	14.98	49.92	60.06
24	10	9.99	9.98	14.98	49.88	60
25	9.99	9.99	9.99	14.98	49.91	60.04
26	10	9.99	9.99	14.99	49.85	60.01
27	10	10	9.99	14.99	49.91	60.05
28	10	9.99	9.99	15	49.92	60.04
29	10	9.99	9.99	14.99	49.89	60.01
30	10	10	9.99	14.99	49.88	60

Eigenanalysis of the Aluminum Pin Correlation Matrix with *n*=30

Table 4.5b The Eigenanalysis of the HDS with $n = 30$

	PC1	PC2	PC3	PC4	PC5	PC6
Eigenvalue	3.7746	1.6948	0.2694	0.1282	0.0755	0.0574
Proportion	0.629	0.282	0.045	0.021	0.013	0.010
Cumulative	0.629	0.912	0.956	0.978	0.990	1.000

40 NEW OBSERVATIONS for *n*=30

Table 4.5c The 40 NEW observations for $n = 30$

Obs	Name	Diameter1	Diameter2	Diameter3	Diameter4	Length1	Length2
1	**D1**	10	9.99	9.99	14.99	49.92	60.03
2	**D2**	10	9.99	9.99	15	49.93	60.03
3	**D3**	10	10	9.99	14.99	49.91	60.02
4	**D4**	10	9.99	9.99	14.99	49.92	60.02
5	**D5**	10	9.99	9.99	14.99	49.92	60
6	**D6**	10	10	9.99	15	49.94	60.05

7	**D7**	10	9.99	9.99	15	49.89	59.98
8	**D8**	10	10	9.99	14.99	49.93	60.01
9	**D9**	10	10	9.99	14.99	49.94	60.02
10	**D10**	10	10	9.99	15	49.86	59.96
11	**D11**	10	9.99	9.99	14.99	49.9	59.97
12	**D12**	10	10	10	14.99	49.92	60
13	**D13**	10	10	9.99	14.98	49.91	60
14	**D14**	10	10	10	15	49.93	59.98
15	**D15**	10	9.99	9.98	14.98	49.9	59.99
16	**D16**	9.99	9.99	9.99	14.99	49.88	59.98
17	**D17**	10.01	10.01	10.01	15.01	49.87	59.97
18	**D18**	10	10	9.99	14.99	49.81	59.91
19	**D19**	10.01	10	10	15.01	50.07	60.13
20	**D20**	10.01	10	10	15	49.93	60
21	**D21**	10	10	10	14.99	49.9	59.96
22	**D22**	10.01	10.01	10.01	15	49.85	59.93
23	**D23**	10	9.99	9.99	15	49.83	59.98
24	**D24**	10.01	10.01	10	14.99	49.9	59.98
25	**D25**	10.01	10.01	10	15	49.87	59.96
26	**D26**	10	9.99	9.99	15	49.87	60.02
27	**D27**	9.99	9.99	9.99	14.98	49.92	60.03
28	**D28**	9.99	9.98	9.98	14.99	49.93	60.03
29	**D29**	9.99	9.99	9.98	14.99	49.89	60.01
30	**D30**	10	10	9.99	14.99	49.89	60.01
31	**D31**	9.99	9.99	9.99	15	50.04	60.15
32	**D32**	10	10	10	14.99	49.84	60.03
33	**D33**	10	10	9.99	14.99	49.89	60.01
34	**D34**	10	9.99	9.99	15	49.88	60.01
35	**D35**	10	10	9.99	14.99	49.9	60.04
36	**D36**	9.9	9.89	9.91	14.88	49.99	60.14
37	**D37**	10	9.99	9.99	15	49.91	60.04
38	**D38**	9.99	9.99	9.99	14.98	49.92	60.04
39	**D39**	10.01	10.01	10	15	49.88	60
40	**D40**	10	9.99	9.99	14.99	49.95	60.01

PCA Simplicial Depth r chart for the first PC of the 40 NEW observations for n=30 (α = 0.05 and 0.10)

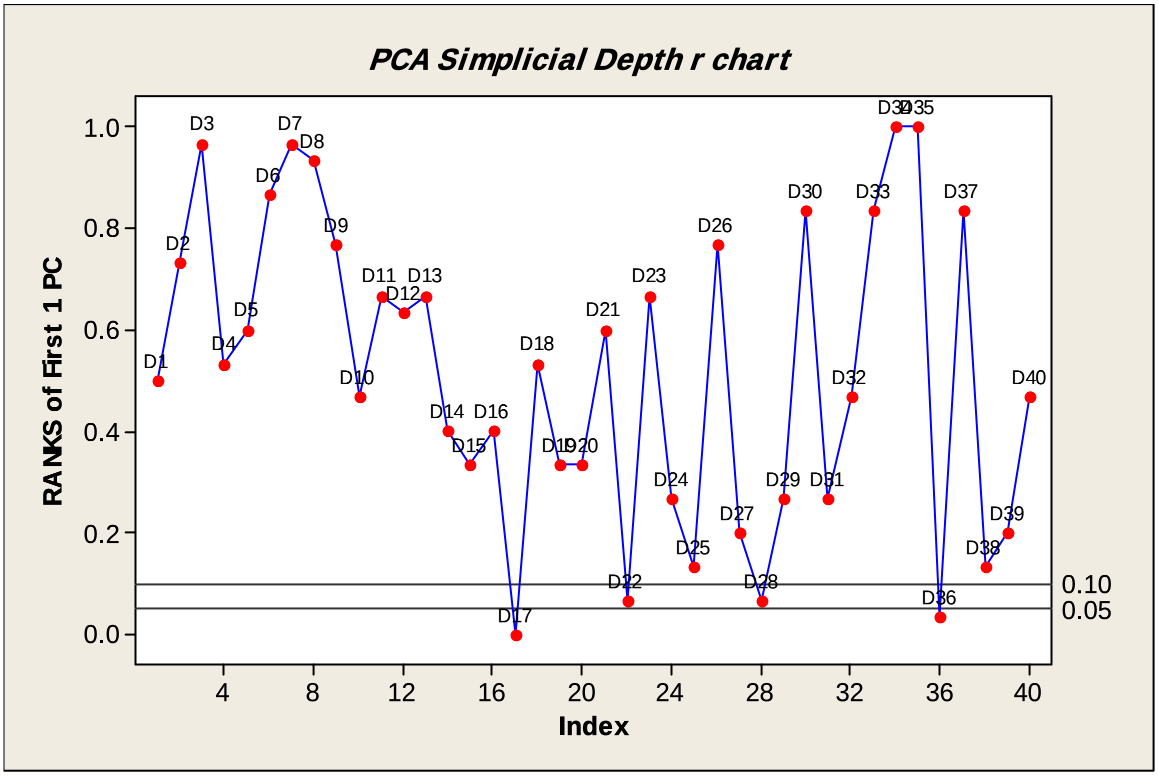

Figure 4.5a PCA Simplicial Depth r chart using the first PC n = 30.

Using the HDS of n = 30, the following points from the 40 new observations were identified as out of control by the control chart when selecting the first PC:

At α = 0.05: Points D17 and D36

At α = 0.10: Points D17, D22, D28 and D36

PCA Simplicial Depth r chart for the last PC of the 40 NEW observations for *n*=30 (α = 0.05 and 0.10)

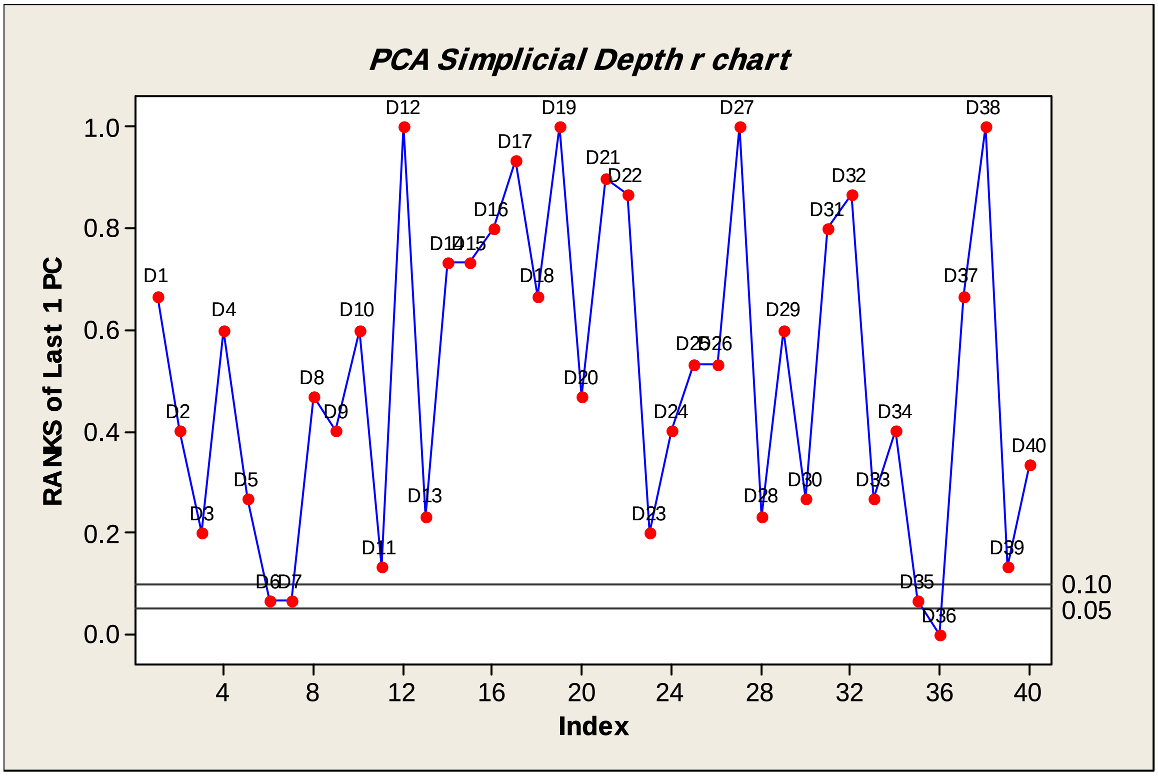

Figure 4.5b PCA Simplicial Depth r chart using the last PC for *n* = 30.

Using the HDS of *n* = 30, the following points from the 40 new observations were identified as out of control by the control chart when selecting the last PC:

At α = 0.05: Point D36

At α = 0.10: Points D6, D7, D35 and D36

7 OBSERVATIONS out of 40 that were identified as out of control from various PCA Simplicial Depth r-charts for *n*=30.

Table 4.5d   The 7 NEW points for Extended HDS with n = 40, n = 50 and n = 63.

Obs	Name	Diameter1	Diameter2	Diameter3	Diameter4	Length1	Length2
1	**D6**	10	10	9.99	15	49.94	60.05
2	**D7**	10	9.99	9.99	15	49.89	59.98
3	**D17**	10.01	10.01	10.01	15.01	49.87	59.97
4	**D22**	10.01	10.01	10.01	15	49.85	59.93
5	**D28**	9.99	9.98	9.98	14.99	49.93	60.03
6	**D35**	10	10	9.99	14.99	49.9	60.04
7	**D36**	9.9	9.89	9.91	14.88	49.99	60.14

The table above illustrates the seven points D6, D7, D17, D22, D28, D35 and 36 from the 40 new observations that were identified as out of control from the control charts of various scenarios with the HDS of n = 30. (see Table 4.5d)   From the remaining 33 points that were identified as in control, we will choose every third point to get an initial extra 10 points to augment the HDS. The 10 additional points are: D3, D8, D11, D14, D18, D21, D25, D29, D32 and D37. An additional set of 10 will be added to augment the HDS to n = 50. These additional 10 points are selected by choosing every other point of those in control that were left: D2, D5, D10, D13, D16, D20, D24, D27, D31 and D34. The remaining 13 points: D1, D4, D9, D12, D15, D19, D23, D26, D30, D33, D38, D39 and D40 will be added to augment the HDS to n = 63. Tables 4.5e, 4.5g and 4.5i illustrate the extended HDS for n = 40, n = 50 and n = 63.

Aluminum Pin Data – Extended Historical Data Set (HDS) with *n*=40

(30 plus 10 identified as in-control from the control charts with *n*=30)

Table 4.5e Extended Historical Data Set of the Aluminum Pin Data with n = 40.

Obs	Diameter1	Diameter2	Diameter3	Diameter4	Length1	Length2
1	9.99	9.97	9.96	14.97	49.89	60.02
2	9.96	9.96	9.95	14.94	49.84	60.02
3	9.97	9.96	9.95	14.95	49.85	60
4	10	9.99	9.99	14.99	49.89	60.06
5	10	9.99	9.99	14.99	49.91	60.09
6	9.99	9.99	9.98	14.99	49.91	60.08
7	10	9.99	9.99	14.98	49.91	60.08
8	10	9.99	9.99	14.99	49.89	60.09
9	9.96	9.95	9.95	14.95	50	60.15
10	9.99	9.98	9.98	14.99	49.86	60.06
11	10	9.99	9.98	14.99	49.94	60.08
12	10	9.99	9.99	14.99	49.92	60.05
13	9.97	9.96	9.96	14.96	49.9	60.02
14	9.97	9.96	9.96	14.96	49.91	60.02
15	9.97	9.97	9.96	14.97	49.9	60.01
16	9.97	9.97	9.96	14.97	49.89	60.04
17	9.98	9.97	9.96	14.96	50.01	60.13
18	9.98	9.97	9.97	14.96	49.93	60.06
19	9.98	9.98	9.97	14.98	49.93	60.02
20	9.98	9.97	9.97	14.97	49.94	60.06
21	9.98	9.97	9.97	14.97	49.93	60.06
22	9.98	9.97	9.97	14.97	49.91	60.02
23	9.98	9.97	9.96	14.98	49.92	60.06
24	10	9.99	9.98	14.98	49.88	60
25	9.99	9.99	9.99	14.98	49.91	60.04
26	10	9.99	9.99	14.99	49.85	60.01
27	10	10	9.99	14.99	49.91	60.05
28	10	9.99	9.99	15	49.92	60.04
29	10	9.99	9.99	14.99	49.89	60.01
30	10	10	9.99	14.99	49.88	60
31	10	10	9.99	14.99	49.91	60.02
32	10	10	9.99	14.99	49.93	60.01
33	10	9.99	9.99	14.99	49.9	59.97

34	10	10	10	15	49.93	59.98
35	10	10	9.99	14.99	49.81	59.91
36	10	10	10	14.99	49.9	59.96
37	10.01	10.01	10	15	49.87	59.96
38	9.99	9.99	9.98	14.99	49.89	60.01
39	10	10	10	14.99	49.84	60.03
40	10	9.99	9.99	15	49.91	60.04

Eigenanalysis of the Aluminum Pin Correlation Matrix with *n* = 40

Table 4.5f The Eigenanalysis of the extended HDS with n = 40.

	PC1	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**
Eigenvalue	4.0045	1.4095	0.3478	0.1180	0.0698	0.0504
Proportion	0.667	0.235	0.058	0.020	0.012	0.008
Cumulative	0.667	0.902	0.960	0.980	0.992	1.000

PCA Simplicial Depth r chart for the first PC of the 7 NEW observations for *n*=40 (α = 0.05 and 0.10)

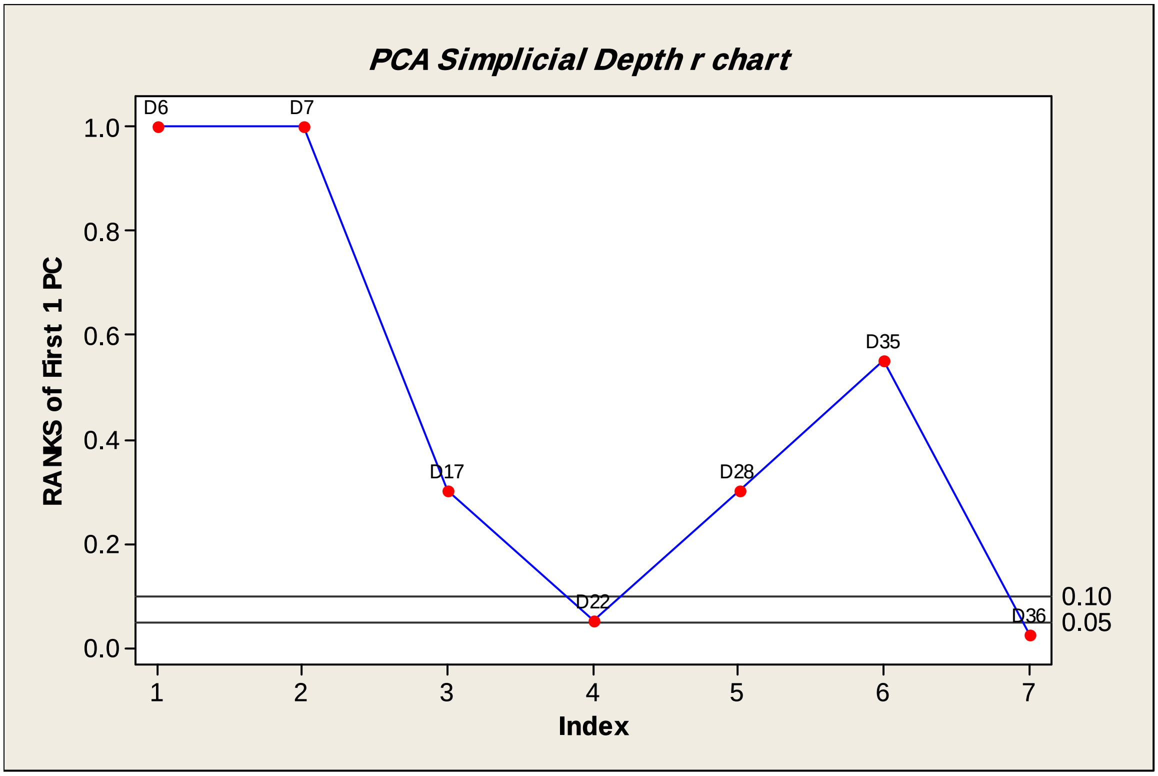

Figure 4.5c PCA Simplicial Depth r chart using the first PC n = 40.

Using the HDS of n = 40, the following points from the 7 new observations were identified as out of control by the control chart when selecting the first PC:

At α = 0.05: Point D36

At α = 0.10: Points D22 and D36

PCA Simplicial Depth r chart for the last PC of the 7 NEW observations for *n*=40 (α = 0.05 and 0.10)

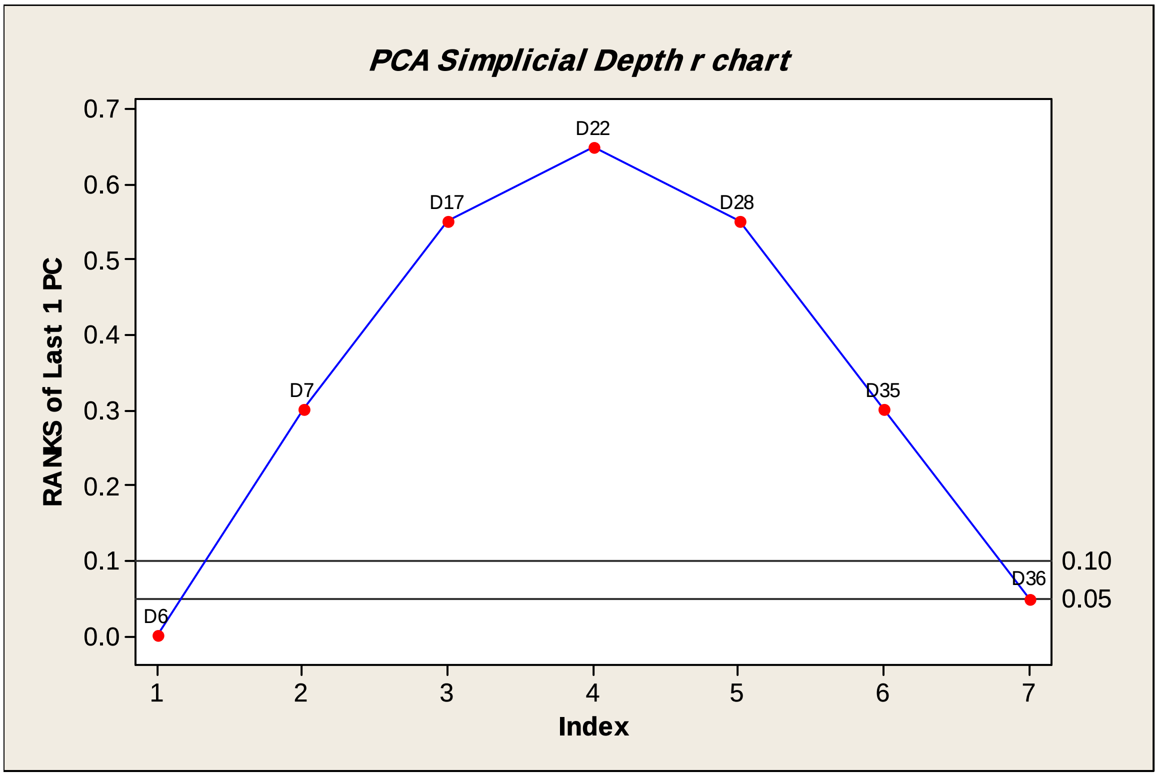

Figure 4.5d PCA Simplicial Depth r chart using the last PC for n = 40.

Using the HDS of n = 40, the following points from the 7 new observations were identified as out of control by the control chart when selecting the last PC:

At α = 0.05: Points D6 and D36

At α = 0.10: Points D6 and D36

Aluminum Pin Data – Extended Historical Data Set (HDS) with *n*=50

(30 plus 20 identified as in-control from the control charts with *n*=30)

Table 4.5g Extended Historical Data Set of the Aluminum Pin Data with n = 50.

Obs	Diameter1	Diameter2	Diameter3	Diameter4	Length1	Length2
1	9.99	9.97	9.96	14.97	49.89	60.02
2	9.96	9.96	9.95	14.94	49.84	60.02
3	9.97	9.96	9.95	14.95	49.85	60
4	10	9.99	9.99	14.99	49.89	60.06
5	10	9.99	9.99	14.99	49.91	60.09
6	9.99	9.99	9.98	14.99	49.91	60.08
7	10	9.99	9.99	14.98	49.91	60.08
8	10	9.99	9.99	14.99	49.89	60.09
9	9.96	9.95	9.95	14.95	50	60.15
10	9.99	9.98	9.98	14.99	49.86	60.06
11	10	9.99	9.98	14.99	49.94	60.08
12	10	9.99	9.99	14.99	49.92	60.05
13	9.97	9.96	9.96	14.96	49.9	60.02
14	9.97	9.96	9.96	14.96	49.91	60.02
15	9.97	9.97	9.96	14.97	49.9	60.01
16	9.97	9.97	9.96	14.97	49.89	60.04
17	9.98	9.97	9.96	14.96	50.01	60.13
18	9.98	9.97	9.97	14.96	49.93	60.06
19	9.98	9.98	9.97	14.98	49.93	60.02
20	9.98	9.97	9.97	14.97	49.94	60.06
21	9.98	9.97	9.97	14.97	49.93	60.06
22	9.98	9.97	9.97	14.97	49.91	60.02
23	9.98	9.97	9.96	14.98	49.92	60.06
24	10	9.99	9.98	14.98	49.88	60
25	9.99	9.99	9.99	14.98	49.91	60.04
26	10	9.99	9.99	14.99	49.85	60.01
27	10	10	9.99	14.99	49.91	60.05
28	10	9.99	9.99	15	49.92	60.04
29	10	9.99	9.99	14.99	49.89	60.01
30	10	10	9.99	14.99	49.88	60
31	10	10	9.99	14.99	49.91	60.02
32	10	10	9.99	14.99	49.93	60.01
33	10	9.99	9.99	14.99	49.9	59.97

34	10	10	10	15	49.93	59.98
35	10	10	9.99	14.99	49.81	59.91
36	10	10	10	14.99	49.9	59.96
37	10.01	10.01	10	15	49.87	59.96
38	9.99	9.99	9.98	14.99	49.89	60.01
39	10	10	10	14.99	49.84	60.03
40	10	9.99	9.99	15	49.91	60.04
41	10	9.99	9.99	15	49.93	60.03
42	10	9.99	9.99	14.99	49.92	60
43	10	10	9.99	15	49.86	59.96
44	10	10	9.99	14.98	49.91	60
45	9.99	9.99	9.99	14.99	49.88	59.98
46	10.01	10	10	15	49.93	60
47	10.01	10.01	10	14.99	49.9	59.98
48	9.99	9.99	9.99	14.98	49.92	60.03
49	9.99	9.99	9.99	15	50.04	60.15
50	10	9.99	9.99	15	49.88	60.01

Eigenanalysis of the Aluminum Pin Correlation Matrix with n = 50

Table 4.5h The Eigenanalysis of the extended HDS with n = 50.

	PC1	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**
Eigenvalue	3.9183	1.5037	0.2970	0.1512	0.0778	0.0521
Proportion	0.653	0.251	0.049	0.025	0.013	0.009
Cumulative	0.653	0.904	0.953	0.978	0.991	1.000

PCA Simplicial Depth r chart for the first PC of the 7 NEW observations for *n*=50 (α = 0.05 and 0.10)

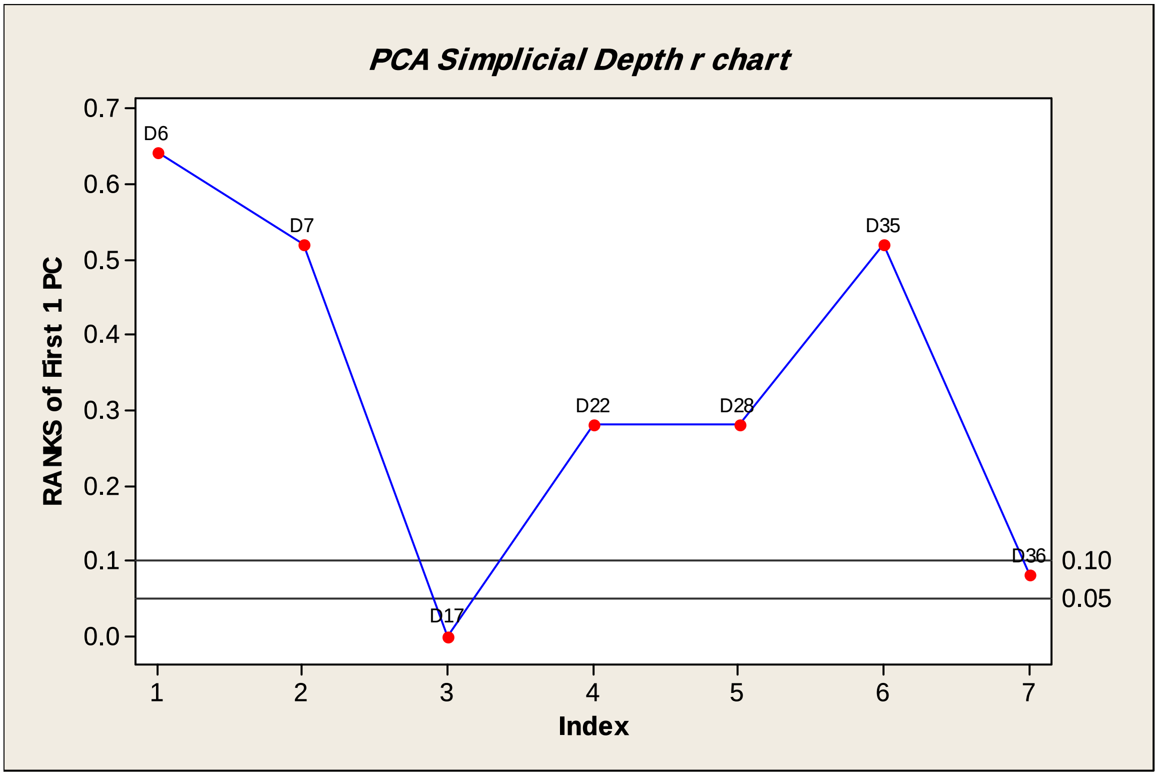

Figure 4.5e PCA Simplicial Depth r chart using the first PC n = 50.

Using the HDS of n = 50, the following points from the 7 new observations were identified as out of control by the control chart when selecting the first PC:

At α = 0.05: Points D17

At α = 0.10: Points D17 and D36

PCA Simplicial Depth r chart for the last PC of the 7 NEW observations for *n*=50 (α = 0.05 and 0.10)

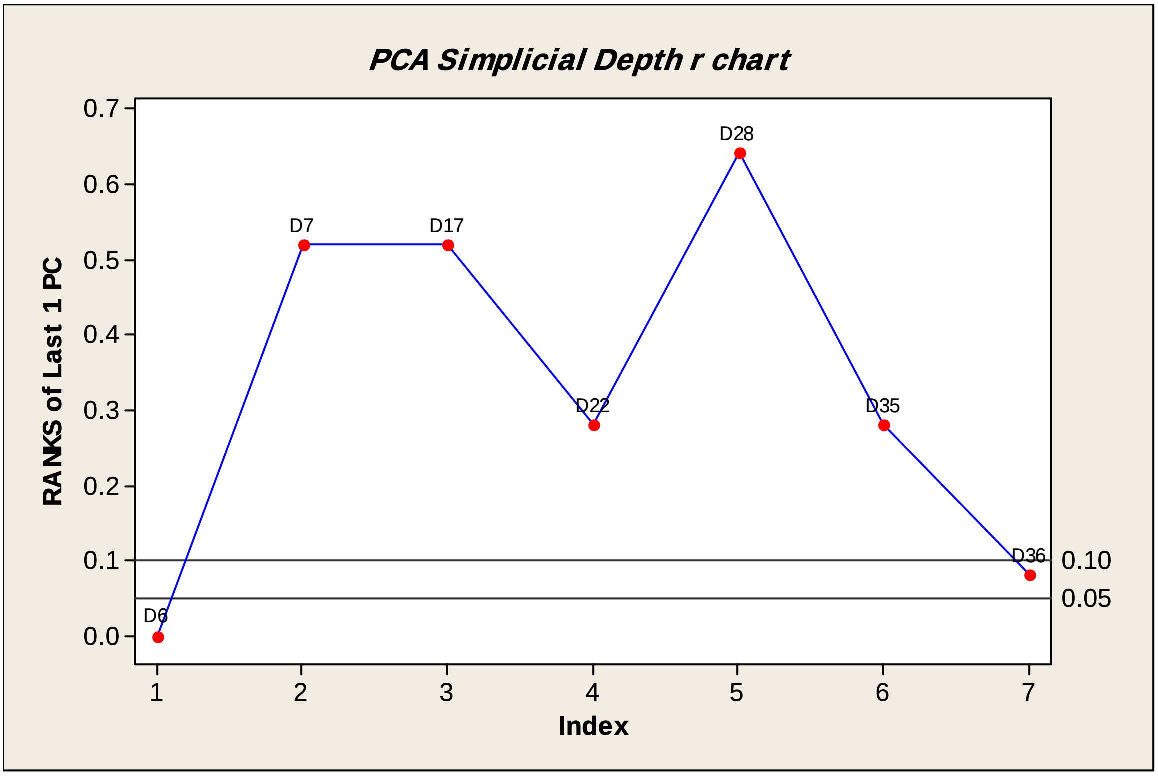

Figure 4.5f PCA Simplicial Depth r chart using the last PC for n = 50.

Using the HDS of n = 50, the following points from the 7 new observations were identified as out of control by the control chart when selecting the last PC:

At α = 0.05: Points D6

At α = 0.10: Points D6 and D36

Aluminum Pin Data – Extended Historical Data Set (HDS) with *n*=63

(30 plus 33 identified as in-control from the control charts with *n*=30)

Table 4.5i Extended Historical Data Set of the Aluminum Pin Data with *n* = 63.

Obs	Diameter1	Diameter2	Diameter3	Diameter4	Length1	Length2
1	9.99	9.97	9.96	14.97	49.89	60.02
2	9.96	9.96	9.95	14.94	49.84	60.02
3	9.97	9.96	9.95	14.95	49.85	60
4	10	9.99	9.99	14.99	49.89	60.06
5	10	9.99	9.99	14.99	49.91	60.09
6	9.99	9.99	9.98	14.99	49.91	60.08
7	10	9.99	9.99	14.98	49.91	60.08
8	10	9.99	9.99	14.99	49.89	60.09
9	9.96	9.95	9.95	14.95	50	60.15
10	9.99	9.98	9.98	14.99	49.86	60.06
11	10	9.99	9.98	14.99	49.94	60.08
12	10	9.99	9.99	14.99	49.92	60.05
13	9.97	9.96	9.96	14.96	49.9	60.02
14	9.97	9.96	9.96	14.96	49.91	60.02
15	9.97	9.97	9.96	14.97	49.9	60.01
16	9.97	9.97	9.96	14.97	49.89	60.04
17	9.98	9.97	9.96	14.96	50.01	60.13
18	9.98	9.97	9.97	14.96	49.93	60.06
19	9.98	9.98	9.97	14.98	49.93	60.02
20	9.98	9.97	9.97	14.97	49.94	60.06
21	9.98	9.97	9.97	14.97	49.93	60.06
22	9.98	9.97	9.97	14.97	49.91	60.02
23	9.98	9.97	9.96	14.98	49.92	60.06
24	10	9.99	9.98	14.98	49.88	60
25	9.99	9.99	9.99	14.98	49.91	60.04
26	10	9.99	9.99	14.99	49.85	60.01
27	10	10	9.99	14.99	49.91	60.05
28	10	9.99	9.99	15	49.92	60.04
29	10	9.99	9.99	14.99	49.89	60.01
30	10	10	9.99	14.99	49.88	60
31	10	10	9.99	14.99	49.91	60.02
32	10	10	9.99	14.99	49.93	60.01
33	10	9.99	9.99	14.99	49.9	59.97

34	10	10	10	15	49.93	59.98
35	10	10	9.99	14.99	49.81	59.91
36	10	10	10	14.99	49.9	59.96
37	10.01	10.01	10	15	49.87	59.96
38	9.99	9.99	9.98	14.99	49.89	60.01
39	10	10	10	14.99	49.84	60.03
40	10	9.99	9.99	15	49.91	60.04
41	10	9.99	9.99	15	49.93	60.03
42	10	9.99	9.99	14.99	49.92	60
43	10	10	9.99	15	49.86	59.96
44	10	10	9.99	14.98	49.91	60
45	9.99	9.99	9.99	14.99	49.88	59.98
46	10.01	10	10	15	49.93	60
47	10.01	10.01	10	14.99	49.9	59.98
48	9.99	9.99	9.99	14.98	49.92	60.03
49	9.99	9.99	9.99	15	50.04	60.15
50	10	9.99	9.99	15	49.88	60.01
51	10	9.99	9.99	14.99	49.92	60.03
52	10	9.99	9.99	14.99	49.92	60.02
53	10	10	9.99	14.99	49.94	60.02
54	10	10	10	14.99	49.92	60
55	10	9.99	9.98	14.98	49.9	59.99
56	10.01	10	10	15.01	50.07	60.13
57	10	9.99	9.99	15	49.83	59.98
58	10	9.99	9.99	15	49.87	60.02
59	10	10	9.99	14.99	49.89	60.01
60	10	10	9.99	14.99	49.89	60.01
61	9.99	9.99	9.99	14.98	49.92	60.04
62	10.01	10.01	10	15	49.88	60
63	10	9.99	9.99	14.99	49.95	60.01

Eigenanalysis of the Aluminum Pin Correlation Matrix with *n* = 63

Table 4.5j The Eigenanalysis of the extended HDS with $n = 63$.

	PC1	PC2	PC3	PC4	PC5	PC6
Eigenvalue	3.8207	1.6032	0.2810	0.1567	0.0824	0.0561
Proportion	0.637	0.267	0.047	0.026	0.014	0.009
Cumulative	0.637	0.904	0.951	0.977	0.991	1.000

PCA Simplicial Depth r chart for the first PC of the 7 NEW observations for *n*=63 (α = 0.05 and 0.10)

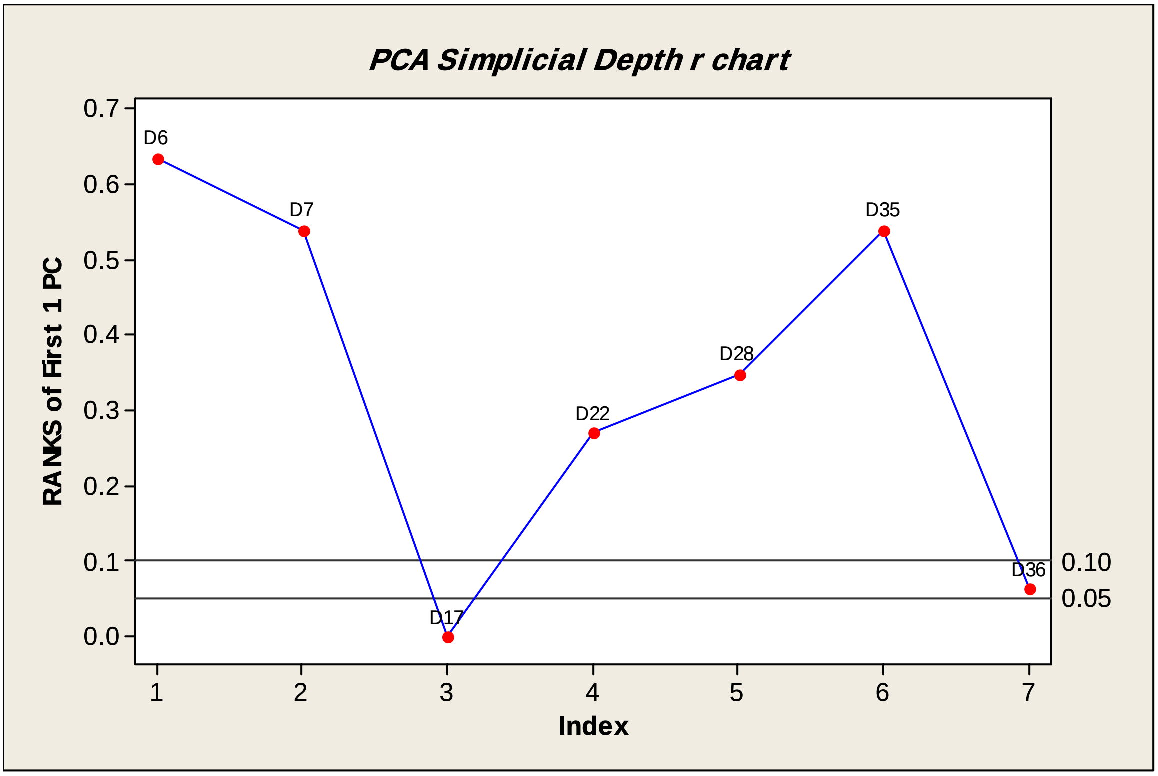

Figure 4.5g PCA Simplicial Depth r chart using the first PC n = 63.

Using the HDS of n = 63, the following points from the 7 new observations were identified as out of control by the control chart when selecting the first PC:

At α = 0.05: Points D17

At α = 0.10: Points D17 and D36

PCA Simplicial Depth r chart for the last PC of the 7 NEW observations for *n*=63 (α = 0.05 and 0.10)

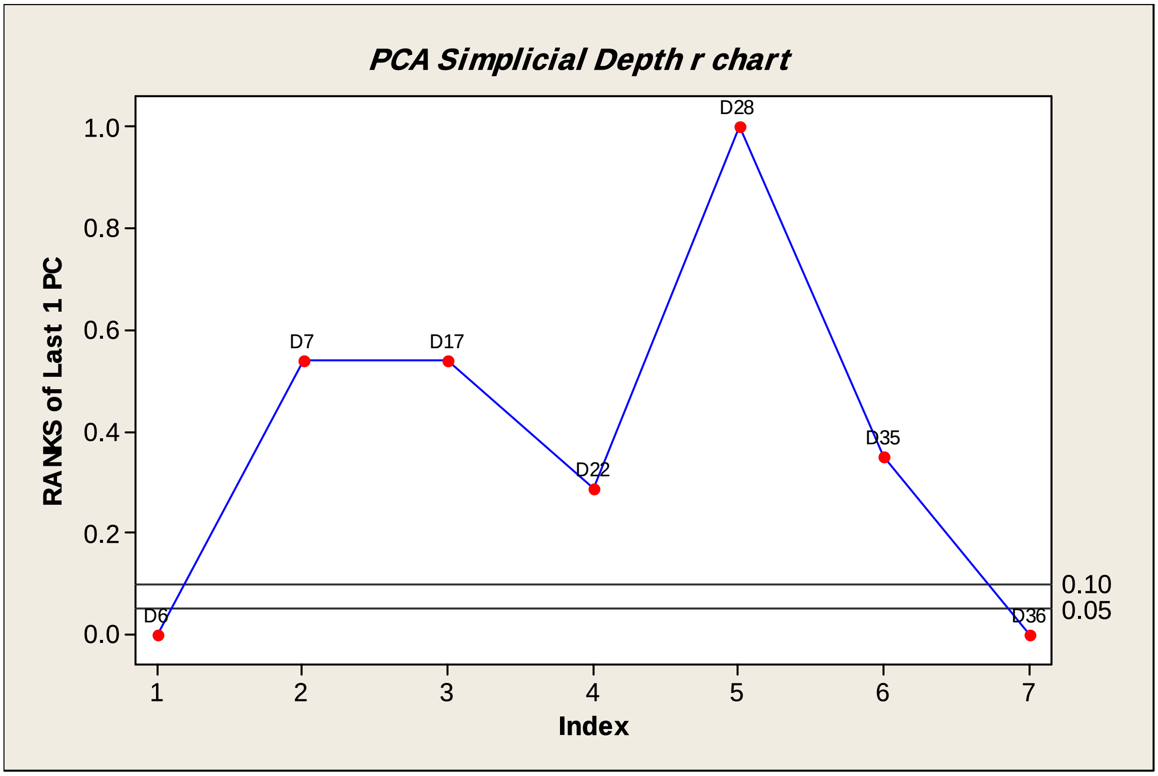

Figure 4.5h PCA Simplicial Depth r chart using the last PC for n = 63.

Using the HDS of n = 63, the following points from the 7 new observations were identified as out of control by the control chart when selecting the last PC:

At α = 0.05: Points D6 and D36

At α = 0.10: Points D6 and D36

Aluminum Pin Data Summary Table

Table 4.5k 40 NEW points with the out of control observations labeled X(*n*=30), Y(*n*=40), Z(*n*=50) and U(*n*=63).

	$\alpha = 0.05$		$\alpha = 0.10$	
Name of point	**First PC**	**Last PC**	**First PC**	**Last PC**
D1				
D2				
D3				
D4				
D5				
D6		Y Z U		X Y Z U
D7				X
D8				
D9				
D10				
D11				
D12				
D13				
D14				
D15				
D16				
D17	X Z U		X Z U	
D18				
D19				
D20				
D21				
D22			X Y	
D23				
D24				
D25				
D26				
D27				
D28			X	
D29				
D30				

D31					
D32					
D33					
D34					
D35					X
D36	XY	XYU		XYZU	XYZU
D37					
D38					
D39					
D40					

4.6 Automotive Data Set

Our fourth data set with 2 process variables consists of 70 observations with the first 45 observations representing the HDS. In numerous studies of MSPC, Mason and Young (2003) and Wade and Woodall (1993), Liu (1995) bivariate data has been used to gain insight into various techniques. For example, Mason and Young (2002) utilize PCA on a bivariate data set to easily illustrate the scaling and the transformations. This bivariate data set, which Wade and Woodall (1993) gathered from Constable, et.al. (1988), represents measurements of a component part for the breaking system for an automobile, and the two process variables measured are rollweight (RWT) and brakeweight (BWT). Table 4.6a contains the 45 points for the HDS.

Historical Data Set (HDS) Original 45 observations

Table 4.6a Historical Data Set of the Automotive Data with n = 45.

Obs	RWT	BWT
1	210	200
2	211	200
3	208	199
4	208	200
5	209	203
6	210	203
7	211	202
8	211	201
9	210	201
10	213	203
11	210	200
12	211	203
13	210	201
14	210	201
15	209	202
16	211	202
17	211	201
18	211	202
19	212	202
20	208	200
21	212	202
22	209	201
23	210	202
24	210	201
25	211	201
26	210	200
27	210	200
28	210	200
29	211	201
30	211	202
31	211	201
32	210	201
33	209	201
34	212	203
35	209	200
36	209	199
37	208	201

38	210	202
39	210	200
40	212	200
41	209	201
42	212	203
43	211	201
44	212	204
45	209	200

Eigenanalysis of the Automotive Correlation Matrix with *n*=45

Table 4.6b The Eigenanalysis of the HDS with $n = 45$

	PC1	**PC2**
Eigenvalue	1.5372	0.4628
Proportion	0.769	0.231
Cumulative	0.769	1.000

25 NEW OBSERVATIONS for *n*=45

Table 4.6c The 25 NEW observations for *n* = 45

Obs	Name of point	RWT	BWT
1	E1	209	201
2	E2	206	200
3	E3	210	200
4	E4	208	199
5	E5	208	198
6	E6	208	200
7	E7	206	197
8	E8	208	199
9	E9	211	201
10	E10	214	204
11	E11	212	203
12	E12	209	200
13	E13	209	204
14	E14	206	201
15	E15	214	202
16	E16	211	202
17	E17	212	202
18	E18	210	203
19	E19	214	201
20	E20	209	200
21	E21	210	201
22	E22	212	201
23	E23	214	206
24	E24	214	204
25	E25	212	203

PCA Simplicial Depth r chart for the first PC of the 25 NEW observations for *n*=45 (α = 0.05 and 0.10)

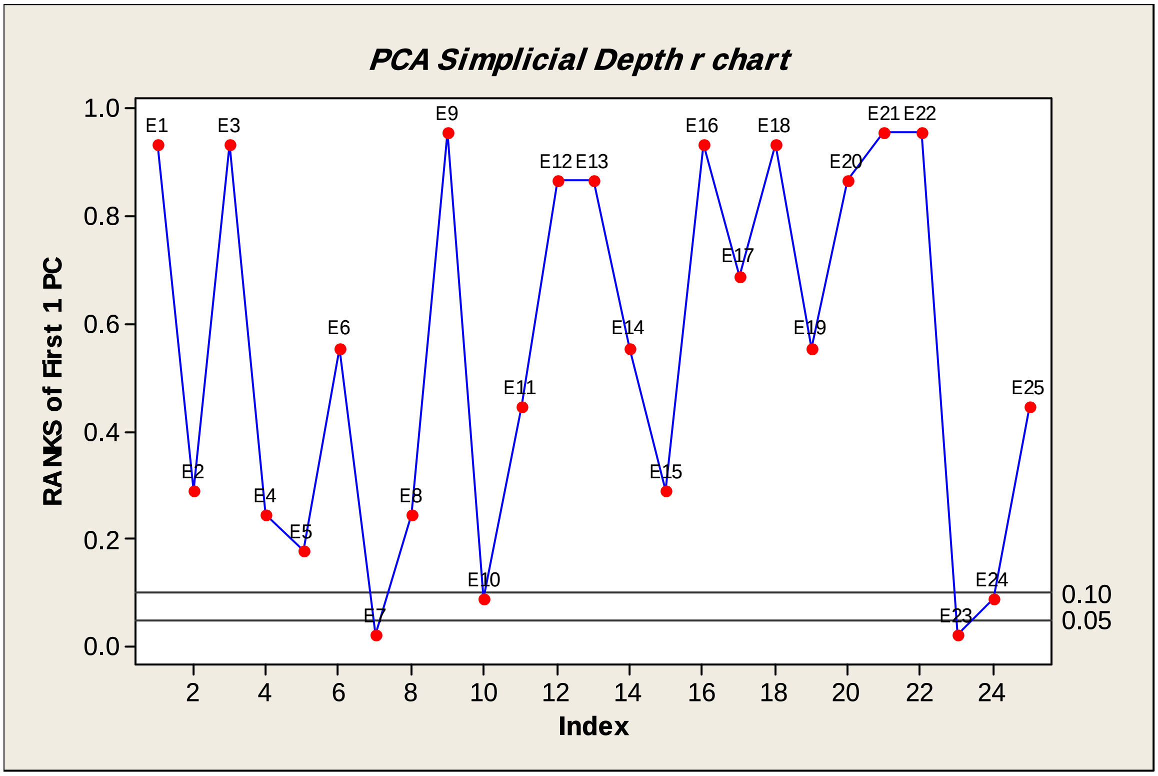

Figure 4.6a PCA Simplicial Depth r chart using the first PC *n* = 45.

Using the HDS of *n* = 45, the following points from the 25 new observations were identified as out of control by the control chart when selecting the first PC:

At α = 0.05: Points E7 and E23

At α = 0.10: Points E7, E10, E23 and E24

PCA Simplicial Depth r chart for the last PC of the 25 NEW observations for n=45 (α = 0.05 and 0.10)

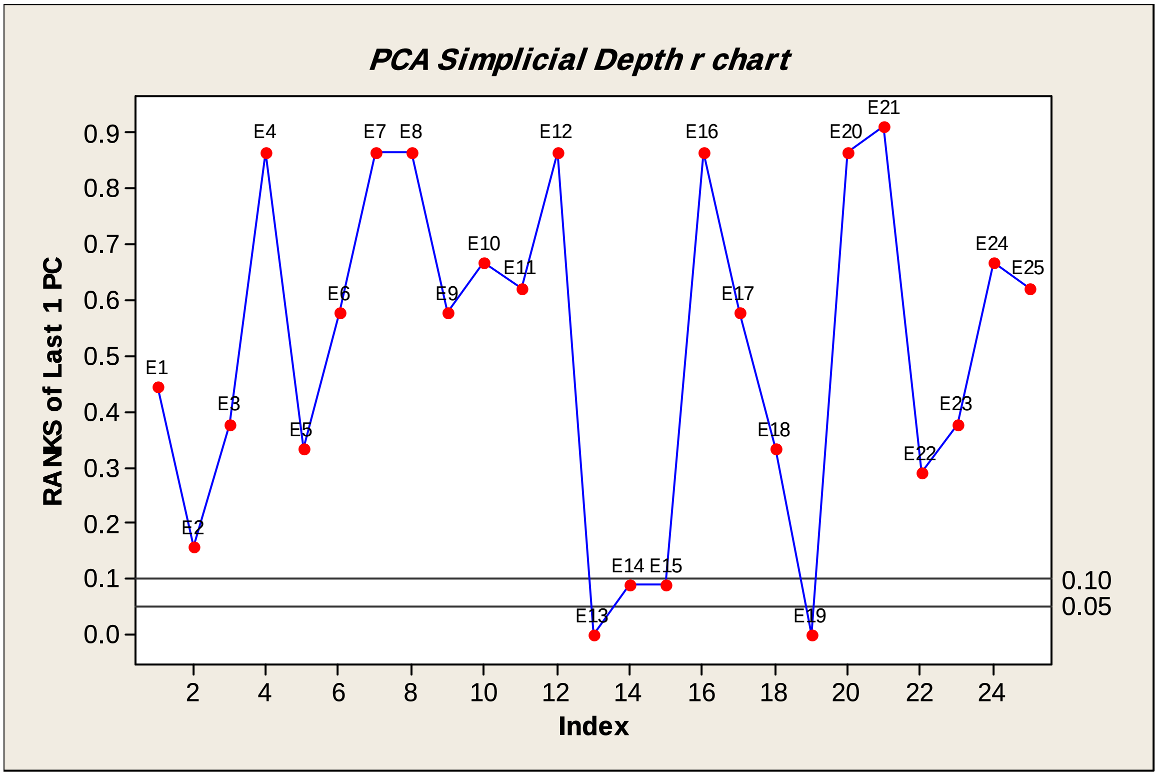

Figure 4.6b PCA Simplicial Depth r chart using the last PC for n = 45.

Using the HDS of n = 45, the following points from the 25 new observations were identified as out of control by the control chart when selecting the last PC:

At α = 0.05: Points E13 and E19

At α = 0.10: Points E13, E14, E15 and E19

From the 25 points which were monitored using the different PCs, the control chart generated signaled for the following 8 points: E7, E10, E13, E14, E15, E19, E23 and E24. [The third scenario included both PCs not shown here in generating possible false alarms.] Using the 17 in-control points: E1, E2, E3, E4, E5, E6, E8, E9, E11, E12, E16, E17, E18, E20, B21, B22 and E25, we will augment the HDS to 62 in order to again determine the sensitivity of the control chart with respect to sample size. (see Table 4.6d)

Automotive Data – Extended Historical Data Set (HDS) with 62 observations (45 plus 17 identified as in-control from the control charts with *n*=45)

Table 4.6d Extended Historical Data Set of the Automotive Data with n = 62.

Obs	RWT	BWT
1	210	200
2	211	200
3	208	199
4	208	200
5	209	203
6	210	203
7	211	202
8	211	201
9	210	201
10	213	203
11	210	200
12	211	203
13	210	201
14	210	201
15	209	202
16	211	202
17	211	201
18	211	202
19	212	202
20	208	200
21	212	202
22	209	201
23	210	202

24	210	201
25	211	201
26	210	200
27	210	200
28	210	200
29	211	201
30	211	202
31	211	201
32	210	201
33	209	201
34	212	203
35	209	200
36	209	199
37	208	201
38	210	202
39	210	200
40	212	200
41	209	201
42	212	203
43	211	201
44	212	204
45	209	200
46	209	201
47	206	200
48	210	200
49	208	199
50	208	198
51	208	200
52	208	199
53	211	201
54	212	203
55	209	200
56	211	202
57	212	202
58	210	203
59	209	200
60	210	201
61	212	201
62	212	203

Eigenanalysis of the Automotive Correlation Matrix with *n* = 62

Table 4.6e The Eigenanalysis of the extended HDS with *n* = 62.

	PC1	PC2
Eigenvalue	1.6298	0.3702
Proportion	0.815	0.185
Cumulative	0.815	1.000

8 OBSERVATIONS out of 25 that were identified as out of control from the PCA Simplicial Depth r-charts for *n*=45.

Table 4.6f   The 8 NEW points for Extended HDS with *n* = 62.

Obs	Name of point	RWT	BWT
1	**E7**	206	197
2	**E10**	214	204
3	**E13**	209	204
4	**E14**	206	201
5	**E15**	214	202
6	**E19**	214	201
7	**E23**	214	206
8	**E24**	214	204

PCA Simplicial Depth r chart for the first PC of the 8 NEW observations for *n*=62 (α = 0.05 and 0.10)

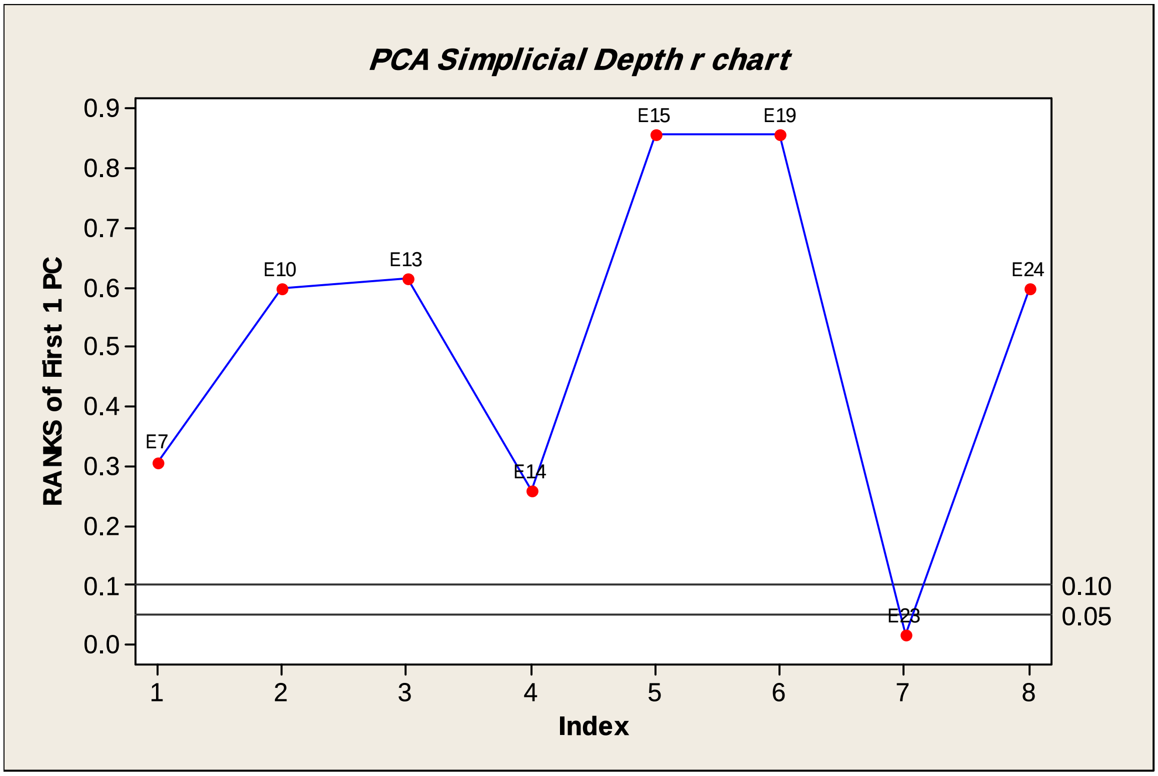

Figure 4.6c PCA Simplicial Depth r chart using the first PC for *n*=62.

Using the HDS of n = 62, the following points from the 8 new observations were identified as out of control by the control chart when selecting the first PC:

At α = 0.05: Point E23

At α = 0.10: Point E23

PCA Simplicial Depth r chart for the last PC of the 8 NEW observations for n=62 (α = 0.05 and 0.10)

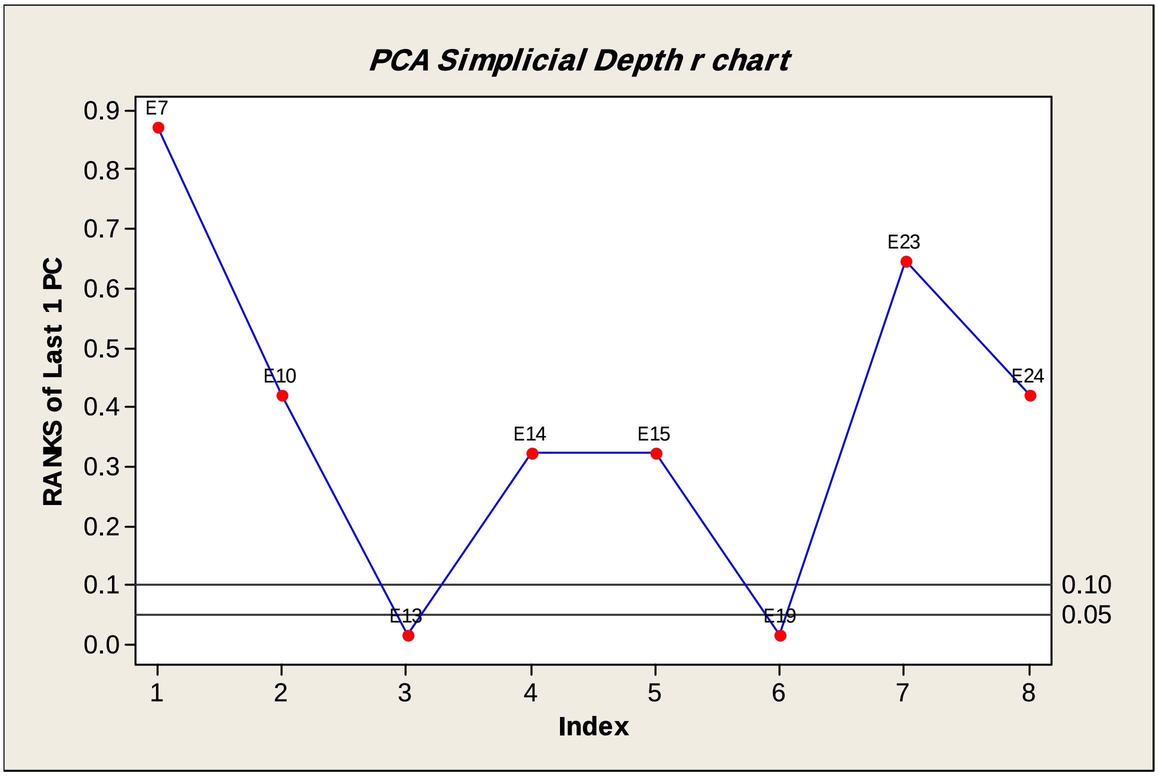

Figure 4.6d PCA Simplicial Depth r chart using the last PC for n =62.

Using the HDS of n = 62, the following points from the 8 new observations were identified as out of control by the control chart when selecting the last PC:

At α = 0.05: Points E13 and E19

At α = 0.10: Points E13 and E19

For a bivariate process, a PCA Simplicial Depth r chart for each PC may unmask the cause(s) of the signal. In this case study, the eigenanalysis for each HDS n= 45 and n=62 illustrates how sample size clearly affects the percentages of variability in this bivariate data set. However, for both sample sizes the possible correlation signals of E13 and E19 were identified with alpha at 0.05.

Automotive Data Summary Table

Table 4.6g 25 NEW points with the out of control observations labeled **X(*n*=45)** and **Y(*n*=62)**.

	$\alpha = 0.05$			$\alpha = 0.10$	
Name of point	**First PC**	**Last PC**		**First PC**	**Last PC**
E1					
E2					
E3					
E4					
E5					
E6					
E7	X			X	
E8					
E9					
E10				X	
E11					
E12					
E13		X Y			X Y
E14					X
E15					X
E16					
E17					
E18					
E19		X Y			X Y
E20					
E21					
E22					
E23	X Y			X Y	
E24				X	
E25					

4.7 Electrolyzer Data Set

Our sixth data set with 6 process variables consists of 21 observations representing the HDS and 6 additional observations which represent mean vectors of sets of 4 observations.(Mason and Young 2002) The six process variables, namely NaOH, NaCl, I_1, I_2, O_2 and Cl_2, are the chemicals used in a electrolyzer process. Table 4.7a lists the 21 in control observations that will be used for the HDS.

Historical Data Set (HDS) Original 21 observations

Table 4.7a Historical Data Set of the Electrolyzer Data with n = 21.

Obs	NaOH	NaCl	I_1	I_2	O_2	Cl_2
1	134.89	203	0.05	4	98.37	1.17
2	129.3	203.1	0.06	1.9	98.37	1.17
3	145.5	208.6	0.17	6.1	98.23	1.42
4	143.8	188.1	0.11	0.4	98.44	1.12
5	146.3	189.1	0.22	0.5	98.44	1.11
6	141.5	196.19	0.16	3.5	98.26	1.35
7	157.3	185.3	0.09	2.9	98.23	1.4
8	141.1	209.1	0.16	0.5	98.69	0.86
9	131.3	200.8	0.17	3.8	97.95	1.64
10	156.6	189	0.19	0.5	97.97	1.62
11	135.6	192.8	0.26	0.5	97.65	1.94
12	128.39	213.1	0.07	3.6	98.43	1.23
13	138.1	198.3	0.15	2.7	98.12	1.36
14	140.5	186.1	0.3	0.3	98.15	1.37
15	139.3	204	0.25	3.8	98.02	1.54
16	152.39	176.3	0.19	0.9	98.22	1.3
17	139.69	186.1	0.15	1.6	98.3	1.25
18	130.3	190.5	0.23	2.6	98.08	1.37
19	132.19	198.6	0.09	5.7	98.3	1.16
20	134.8	196.1	0.17	4.9	97.98	1.5
21	142.3	198.8	0.09	0.3	98.41	1

Eigenanalysis of the Electrolyzer Correlation Matrix with *n*=21

Table 4.7b The Eigenanalysis of the HDS with $n = 21$

	PC1	PC2	PC3	PC4	PC5	PC6
Eigenvalue	2.5666	1.9017	0.7369	0.4338	0.3348	0.0261
Proportion	0.428	0.317	0.123	0.072	0.056	0.004
Cumulative	0.428	0.745	0.868	0.940	0.996	1.000

6 NEW OBSERVATIONS for *n*=21

Table 4.7c The 6 NEW observations for $n = 21$

Obs	Electrolyzer Number	NaOH	NaCl	I_1	I_2	O_2	Cl_2
1	**573**	130.04	192.26	0.26	10.93	98.1	1.47
2	**372**	131.77	189.75	0.16	0.94	98.36	1.21
3	**834**	134.37	184.99	0.14	1.1	98.43	1.12
4	**1021**	140.47	195.84	0.14	4.51	97.91	1.64
5	**963**	129.68	198.03	0.23	8.52	97.82	1.71
6	**622**	137.96	196.98	0.2	1.74	98.18	1.38

PCA Simplicial Depth r chart for the first PC of the 6 NEW observations for *n*=21 (α = 0.05 and 0.10)

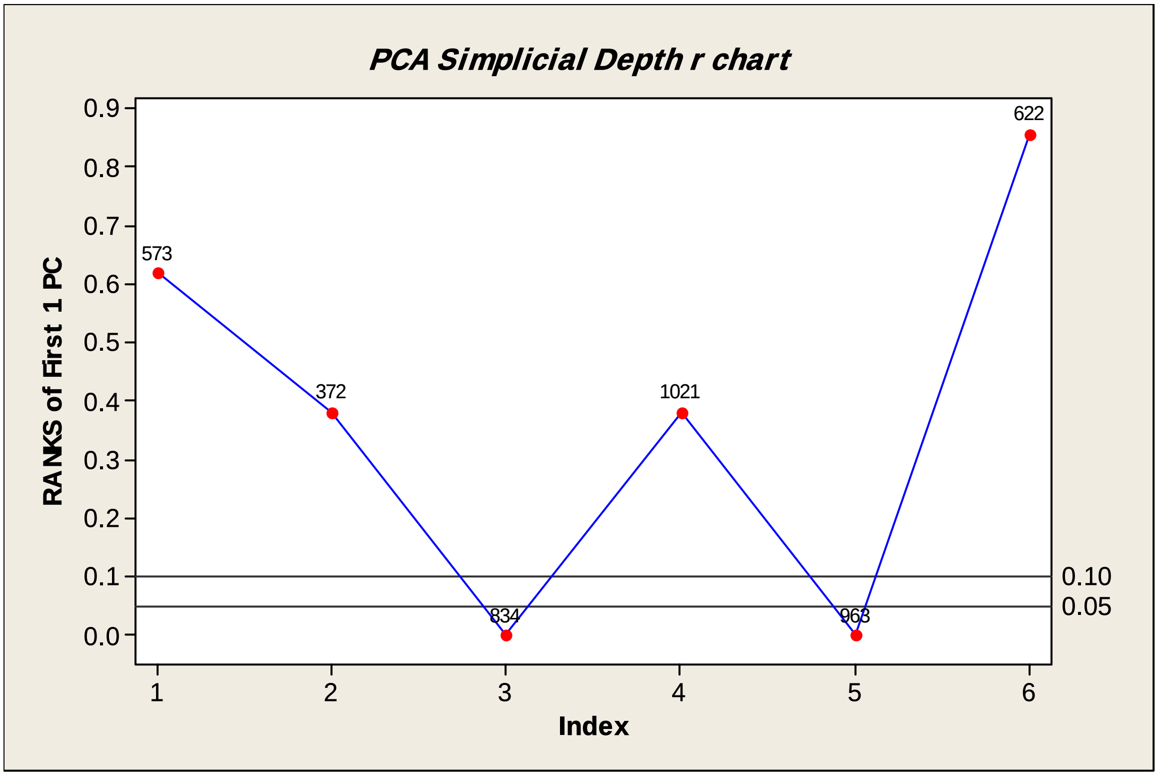

Figure 4.7a PCA Simplicial Depth r chart using the first PC n = 21.

Using the HDS of n = 21, the following points from the 6 new observations were identified as out of control by the control chart when selecting the first PC:

At α = 0.05: Points 834 and 963

At α = 0.10: Points 834 and 963

PCA Simplicial Depth r chart for the last PC of the 6 NEW observations for n=21 (α = 0.05 and 0.10)

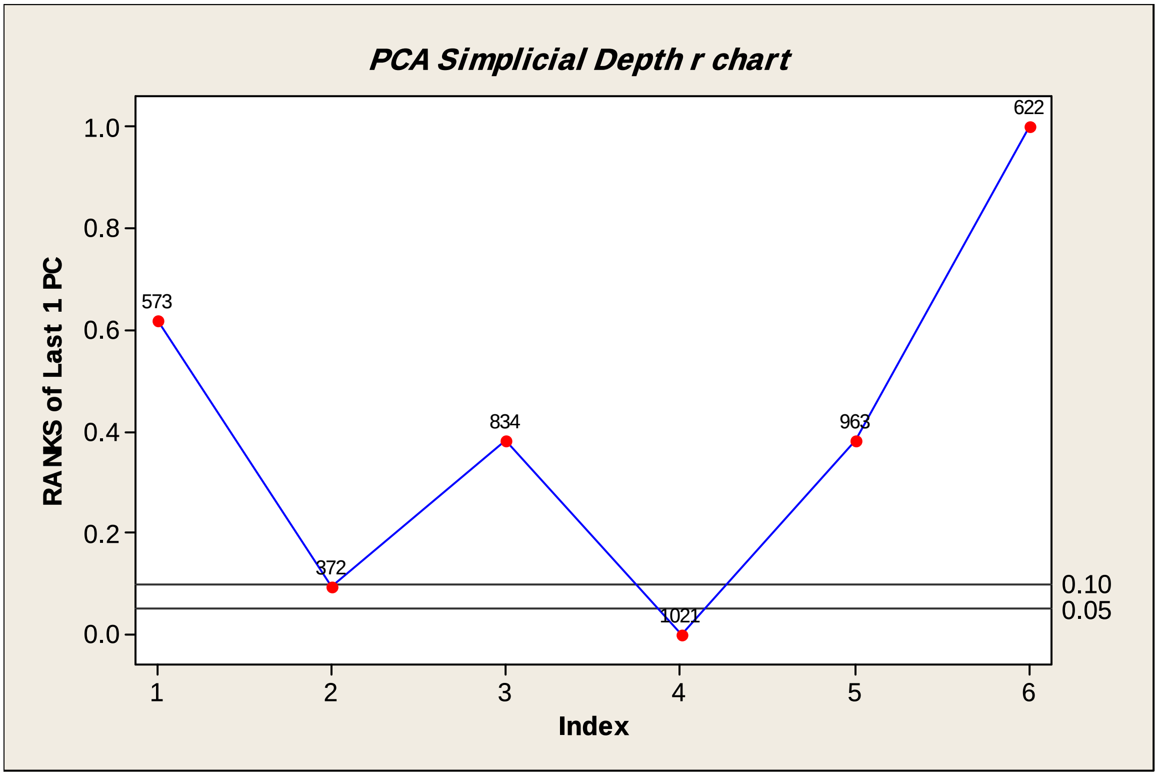

Figure 4.7b PCA Simplicial Depth r chart using the last PC for n = 21.

For the HDS of n = 21, the following points from the 6 new observations were identified as out of control by the control chart when selecting the last PC:

At α = 0.05: Point 1021

At α = 0.10: Points 372 and 1021

Electrolyzer Data Summary Table

Table 4.7d 6 NEW points with the out of control observations labeled **✗(*n*=21)**.

	α = 0.05		α = 0.10	
Name of point	**First PC**	**Last PC**	**First PC**	**Last PC**
573				
372				✗
834	✗		✗	
1021		✗		✗
963	✗		✗	
622				

4.8 Mechanical Data Set

Our next data set with 7 process variables consists of 36 observations of a complex mechanical part. (Fuchs and Kennet 1998) The process variables are: X1, X2 (both represent two diameter measurements on another cylindrical segment of that part), X3, X4, X5 (all three represent lengths of various portions of the part), X6 and X7 (both represent two different tall width measures). Table 4.8a lists the 27 in control observations that will be used for the HDS.

Historical Data Set (HDS) 27 observations

Table 4.8a Historical Data Set of the Mechanical Data with *n* = 27.

Obs	X1	X2	X3	X4	X5	X6	X7
1	74.95	74.95	10	26.22	29	0.07	0.12
2	74.95	75	10	26.2	29	0.08	0.13
3	74.93	75	10	26.18	29	0.05	0.12
4	74.93	75	10	26.18	29	0.08	0.15
5	74.95	75	10	26.2	29	0.09	0.14
6	74.95	75	10	26.2	29	0.1	0.18
7	74.94	75	10	26.2	29	0.13	0.25
8	74.02	75	10	26.3	29	0.05	0.15
9	74.95	75	10	26.22	29	0.07	0.16
10	74.94	75	10.04	26.22	29	0.05	0.12
11	74.95	74.97	10.06	26.3	29	0.05	0.13
12	74.95	74.75	10.04	26.2	29	0.06	0.05
13	74.96	74.75	10.04	26.2	29	0.1	0.2
14	74.94	74.99	10.06	26.2	29	0.12	0.16
15	74.93	74.98	10.06	26.2	29	0.12	0.18
16	74.92	74.99	10.08	26.2	29	0.05	0.25
17	74.93	74.98	10.06	26.18	29	0.05	0.05
18	74.92	74.96	10.06	26.24	29.04	0.06	0.04
19	74.93	74.99	10.08	26.2	29.04	0.03	0.12
20	74.92	74.99	10.1	26.1	29.08	0.12	0.17
21	74.93	74.98	10.1	26.16	29.06	0.08	0.06
22	74.93	74.99	10.04	26.16	29.06	0.08	0.1
23	74.92	74.97	10.06	26.16	29.06	0.1	0.05
24	74.92	74.97	10.04	26.22	29.08	0.04	0.09
25	74.93	74.99	10.06	26.14	28.98	0.15	0.15
26	74.92	74.98	10.06	26.16	28.98	0.1	0.05
27	74.91	74.98	10.04	26.2	29.02	0.04	0.1

Eigenanalysis of the Mechanical Correlation Matrix with *n*=27

Table 4.8b The Eigenanalysis of the HDS with *n* = 27

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Eigenvalue	2.0940	1.6740	1.1028	0.7646	0.6035	0.4984	0.2627
Proportion	0.299	0.239	0.158	0.109	0.086	0.071	0.038
Cumulative	0.299	0.538	0.696	0.805	0.891	0.962	1.000

9 NEW OBSERVATIONS for *n*=27

Table 4.8c The 9 NEW observations for *n* = 27

Obs	Name	X1	X2	X3	X4	X5	X6
1	**H1**	74.92	74.99	10.06	26.21	29.04	0.07
2	**H2**	74.93	74.98	10.12	26.28	28.98	0.11
3	**H3**	74.94	75	10.08	26.22	28.98	0.05
4	**H4**	74.93	74.99	10.06	26.2	29.04	0.09
5	**H5**	74.94	75	10.06	26.22	29.04	0.08
6	**H6**	74.93	75	10.04	26.22	29.04	0.04
7	**H7**	74.93	74.99	10	26.2	28.98	0.07
8	**H8**	74.93	74.99	10.06	26.28	29.04	0.13
9	**H9**	74.94	74.99	10.04	26.25	28.98	0.13

PCA Simplicial Depth r chart for the first 2 PCs of the 9 NEW observations for *n*=27 (α = 0.05 and 0.10)

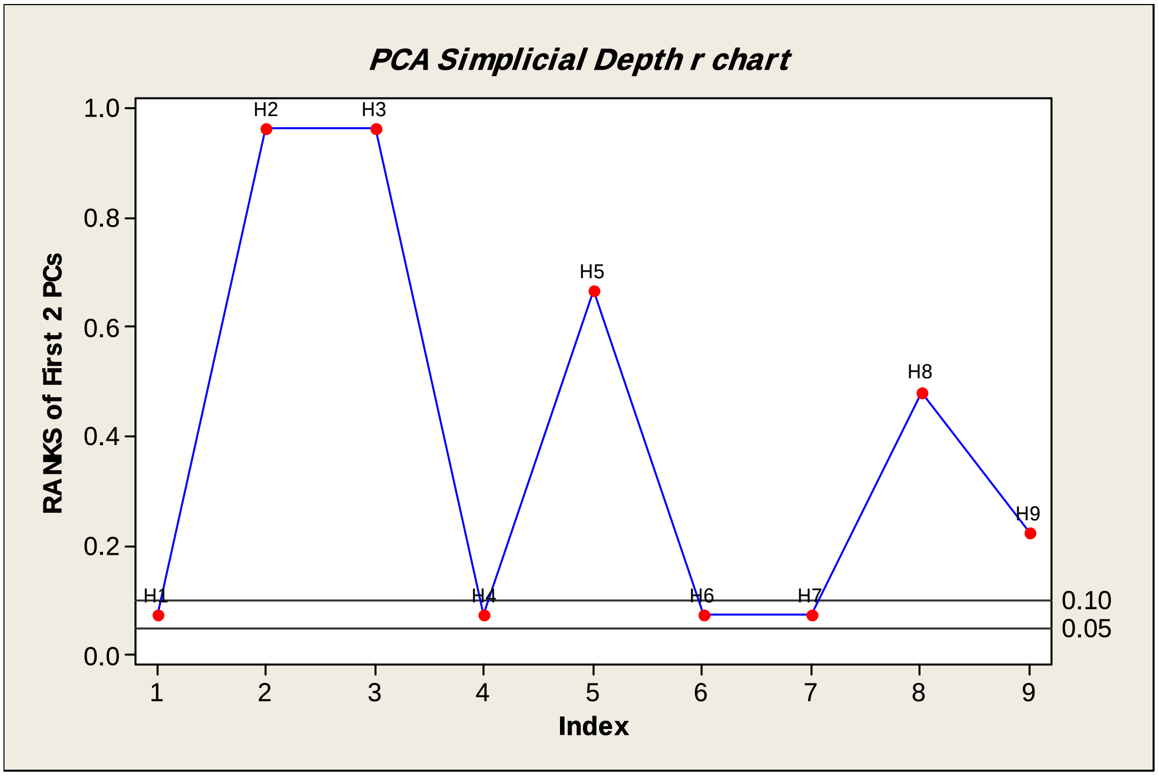

Figure 4.8a PCA Simplicial Depth r chart using the first 2 PCs n = 27.

Using the HDS of n = 27, the following points from the 9 new observations were identified as out of control by the control chart when selecting the first 2 PCs:

At α = 0.05: No Points

At α = 0.10: Points H1, H4, H6 and H7

PCA Simplicial Depth r chart for the last PC of the 9 NEW observations for *n*=27 (α = 0.05 and 0.10)

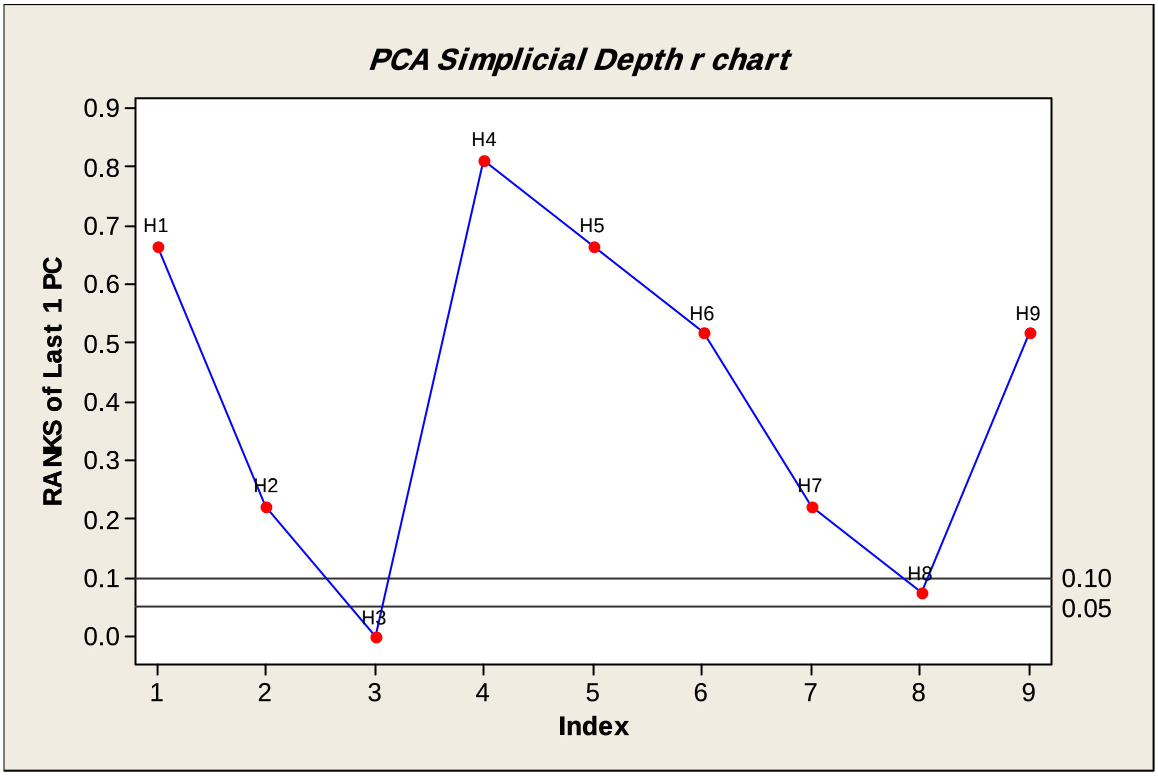

Figure 4.8b PCA Simplicial Depth r chart using the last PC for n = 27.

Using the HDS of n = 27, the following points from the 40 new observations were identified as out of control by the control chart when selecting the last PC:

At α = 0.05: Point H3

At α = 0.10: Points H3 and H8

Mechanical Data Summary Table

Table 4.8d 9 NEW points with the out of control observations labeled ✗(***n***=**27**).

	$\alpha = 0.05$		$\alpha = 0.10$	
Name of point	**First 2 PCs**	**Last PC**	**First 2 PCs**	**Last PC**
H1			✗	
H2				
H3		✗		✗
H4			✗	
H5				
H6			✗	
H7			✗	
H8				✗
H9				

4.9 Robustness to Random Error in the HDS

The following data set contains an observation generated as a random error in the HDS (Fuchs and Kennet 1998). Fuchs and Kennet (1998) used a simulated data run of the Aluminum Pin Data and generated a bivariate data set from two of the variables, specifically Length 1 and Length 2, and titled them VAR 1 and VAR 2. Table 4.9a displays the HDS with n = 50 with observation 23 identified as the random error in the bivariate dataset with process variables VAR1 and VAR2. (Fuchs and Kennet 1998)

Historical Data Set (HDS) Original 50 observations with one random error

Table 4.9a Historical Data Set of the Bivariate Data with n = 50 including the random error point (Observation 23)

Obs	VAR1	VAR2
1	49.8585	60.0008
2	49.8768	59.9865
3	49.8706	60.0055
4	49.9117	60.0126
5	49.847	60.0165
6	49.8883	60.0216
7	49.9158	60.0517
8	49.9152	60.0673
9	49.9055	60.0726
10	49.8969	60.0208
11	49.9137	60.0928
12	49.8586	59.9823
13	49.9514	60.0866
14	49.8988	60.0402
15	49.8894	60.072
16	49.9403	60.0681
17	49.9132	60.035
18	49.8546	60.0145
19	49.8815	59.9982
20	49.8311	59.9963
21	49.8816	60.0457
22	49.8501	59.986
23*	49.9778*	60.0875*
24	49.869	60.0159
25	49.8779	60.0055
26	49.868	60.0088
27	49.9388	60.0711
28	49.9133	60.0634
29	49.912	60.056
30	49.925	60.0749
31	49.9442	60.11
32	49.8386	59.9725

33	49.9492	60.1014
34	49.9204	60.0803
35	49.8994	60.0625
36	49.8703	60.0219
37	49.8846	60.0271
38	49.958	60.0878
39	49.8985	60.0329
40	49.9397	60.0826
41	49.8741	60.0061
42	49.914	60.0401
43	49.9501	60.081
44	49.8865	60.0169
45	49.8912	60.0406
46	49.9252	60.0532
47	49.9326	60.0741
48	49.968	60.1219
49	49.9289	60.0709
50	49.9233	60.0632

Eigenanalysis of the Bivariate Correlation Matrix with *n*=50 including the Random Error Observation

Table 4.9b The Eigenanalysis of the HDS with n = 50

	PC1	PC2
Eigenvalue	1.8827	0.1173
Proportion	0.941	0.059
Cumulative	0.941	1.000

25 NEW OBSERVATIONS for *n*=50 and *n*=49

Table 4.9c The 25 NEW observations for $n = 50$ and for $n = 49$

Obs	Name of point	VAR1	VAR2
1	G1	49.8798	60.0417
2	G2	49.9208	60.0292
3	G3	49.9606	60.1172
4	G4	49.9498	60.0543
5	G5	49.839	59.9665
6	G6	49.9284	60.0079
7	G7	49.9648	60.0482
8	G8	49.978	60.0186
9	G9	50.0218	60.0854
10	G10	50.0606	60.1399
11	G11	50.0365	60.1005
12	G12	49.9756	60.0387
13	G13	49.984	60.0857
14	G14	50.0028	60.0482
15	G15	49.977	60.0278
16	G16	49.8579	60.0588
17	G17	49.8997	60.082
18	G18	49.9156	60.1415
19	G19	49.9258	60.1132
20	G20	49.8384	60.0449
21	G21	49.8937	60.0893
22	G22	49.8631	60.0757
23	G23	49.9406	60.1298
24	G24	49.9046	60.0739
25	G25	49.8718	60.0676

The 25 new observations in Table 4.9c will be tested with the HDS for $n = 50$ which includes the random error and $n = 49$ with the random error point deleted in order to test the robustness of our proposed control chart scheme when the HDS has been contaminated by a point that was not caused by process degradation but rather a random error reading.

PCA Simplicial Depth r chart for the first PC of the 25 NEW observations for *n*=50 (α = 0.05 and 0.10)

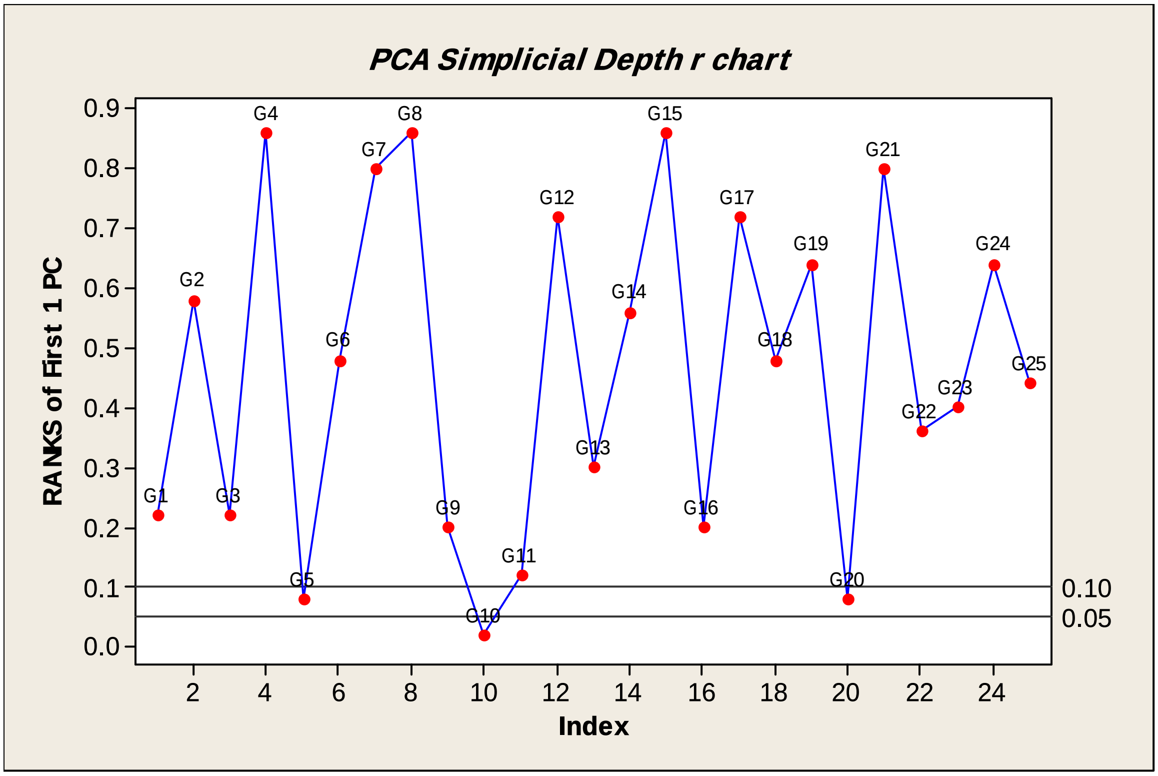

Figure 4.9a PCA Simplicial Depth r chart using the first PC n = 50.

Using the HDS of n = 50, the following points from the 25 new observations were identified as out of control by the control chart when selecting the first PC:

At α = 0.05: Point G10

At α = 0.10: Points G5, G10 and G20.

PCA Simplicial Depth r chart for the last PC of the 25 NEW observations for *n*=50 (α = 0.05 and 0.10)

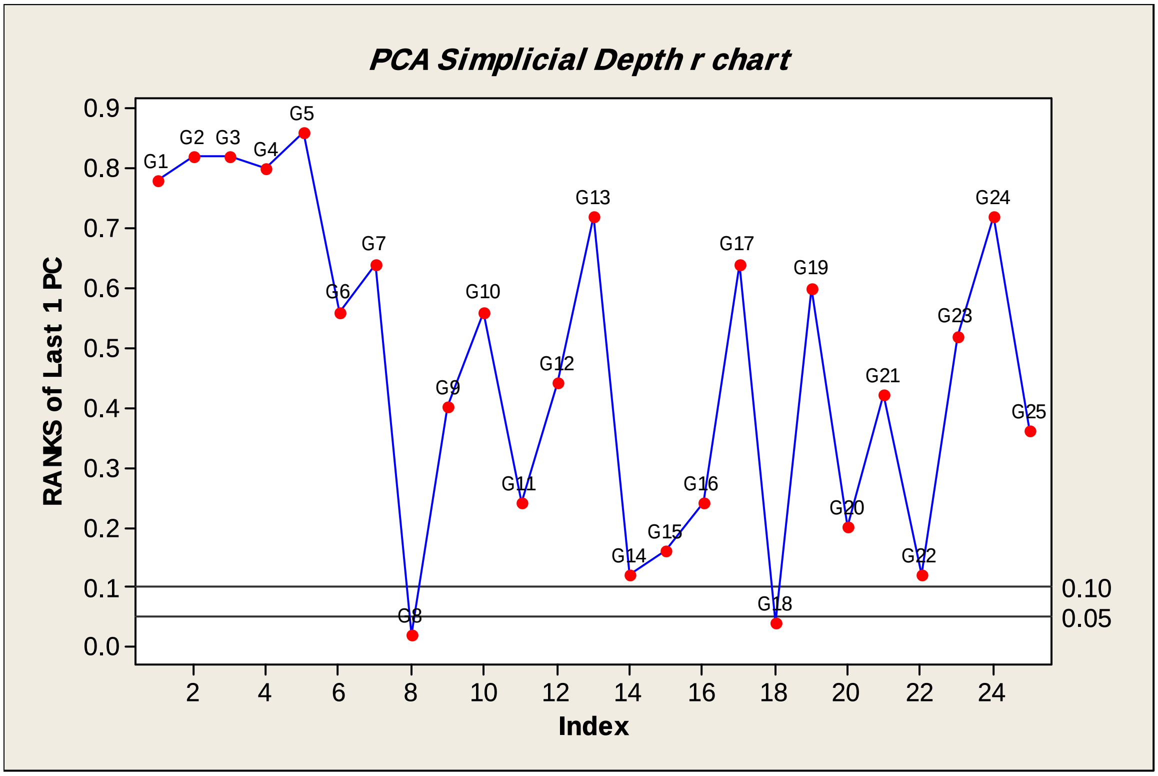

Figure 4.9b PCA Simplicial Depth r chart using the last PC for *n* = 50.

Using the HDS of *n* = 50, the following points from the 25 new observations were identified as out of control by the control chart when selecting the last PC:

At α = 0.05: Points G8 and G18

At α = 0.10: Points G8 and G18

Historical Data Set (HDS) Original 49 observations without the random error (Corrected)

Table 4.9d Historical Data Set of the Bivariate Data with n = 49 without the random error

Obs	VAR1	VAR2
1	49.8585	60.0008
2	49.8768	59.9865
3	49.8706	60.0055
4	49.9117	60.0126
5	49.847	60.0165
6	49.8883	60.0216
7	49.9158	60.0517
8	49.9152	60.0673
9	49.9055	60.0726
10	49.8969	60.0208
11	49.9137	60.0928
12	49.8586	59.9823
13	49.9514	60.0866
14	49.8988	60.0402
15	49.8894	60.072
16	49.9403	60.0681
17	49.9132	60.035
18	49.8546	60.0145
19	49.8815	59.9982
20	49.8311	59.9963
21	49.8816	60.0457
22	49.8501	59.986
23	49.869	60.0159
24	49.8779	60.0055
25	49.868	60.0088
26	49.9388	60.0711
27	49.9133	60.0634
28	49.912	60.056
29	49.925	60.0749
30	49.9442	60.11
31	49.8386	59.9725
32	49.9492	60.1014
33	49.9204	60.0803

34	49.8994	60.0625
35	49.8703	60.0219
36	49.8846	60.0271
37	49.958	60.0878
38	49.8985	60.0329
39	49.9397	60.0826
40	49.8741	60.0061
41	49.914	60.0401
42	49.9501	60.081
43	49.8865	60.0169
44	49.8912	60.0406
45	49.9252	60.0532
46	49.9326	60.0741
47	49.968	60.1219
48	49.9289	60.0709
49	49.9233	60.0632

Eigenanalysis of the Bivariate Correlation Matrix with *n*=49 Corrected

Table 4.9e The Eigenanalysis of the corrected HDS with $n = 49$

	PC1	**PC2**
Eigenvalue	1.8861	0.1139
Proportion	0.943	0.057
Cumulative	0.943	1.000

PCA Simplicial Depth r chart for the first PC of the 25 NEW observations for *n*=49 (α = 0.05 and 0.10)

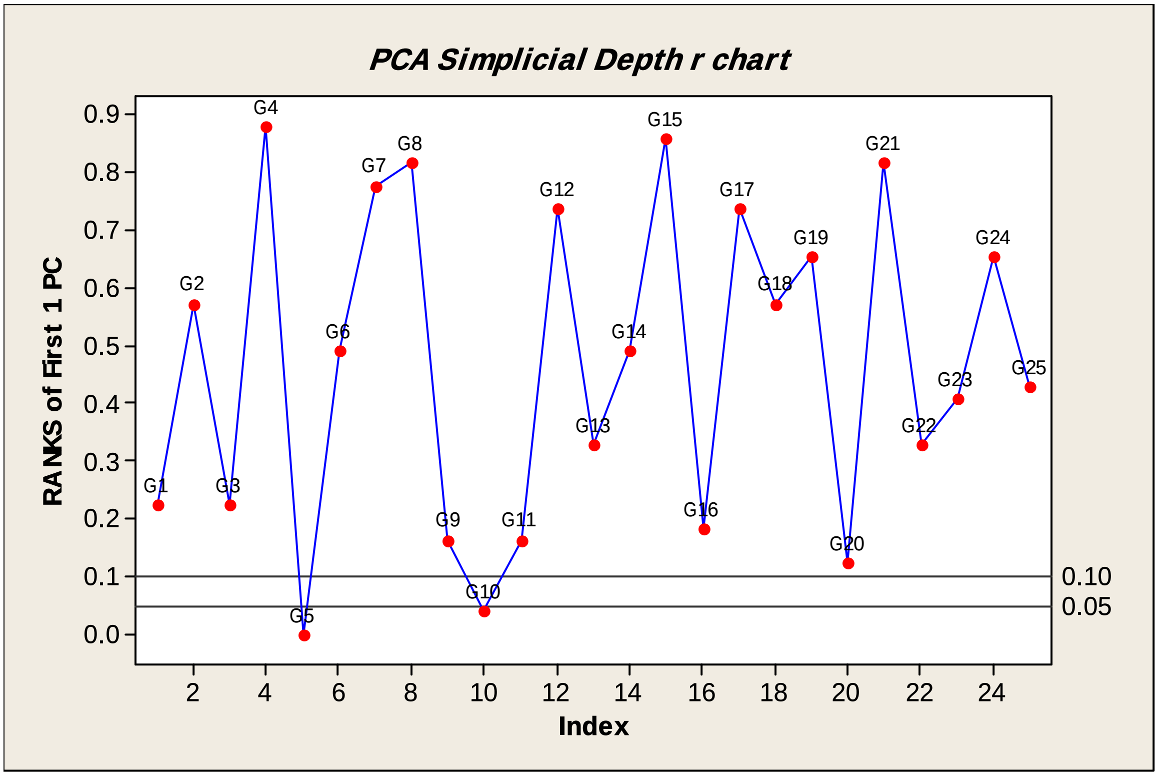

Figure 4.9c PCA Simplicial Depth r chart using the first PC *n* = 49.

Using the HDS of *n* = 49, the following points from the 25 new observations were identified as out of control by the control chart when selecting the first PC:

At α = 0.05: Points G5 and G10

At α = 0.10: Points G5 and G10

PCA Simplicial Depth r chart for the last PC of the 25 NEW observations for *n*=49 (α = 0.05 and 0.10)

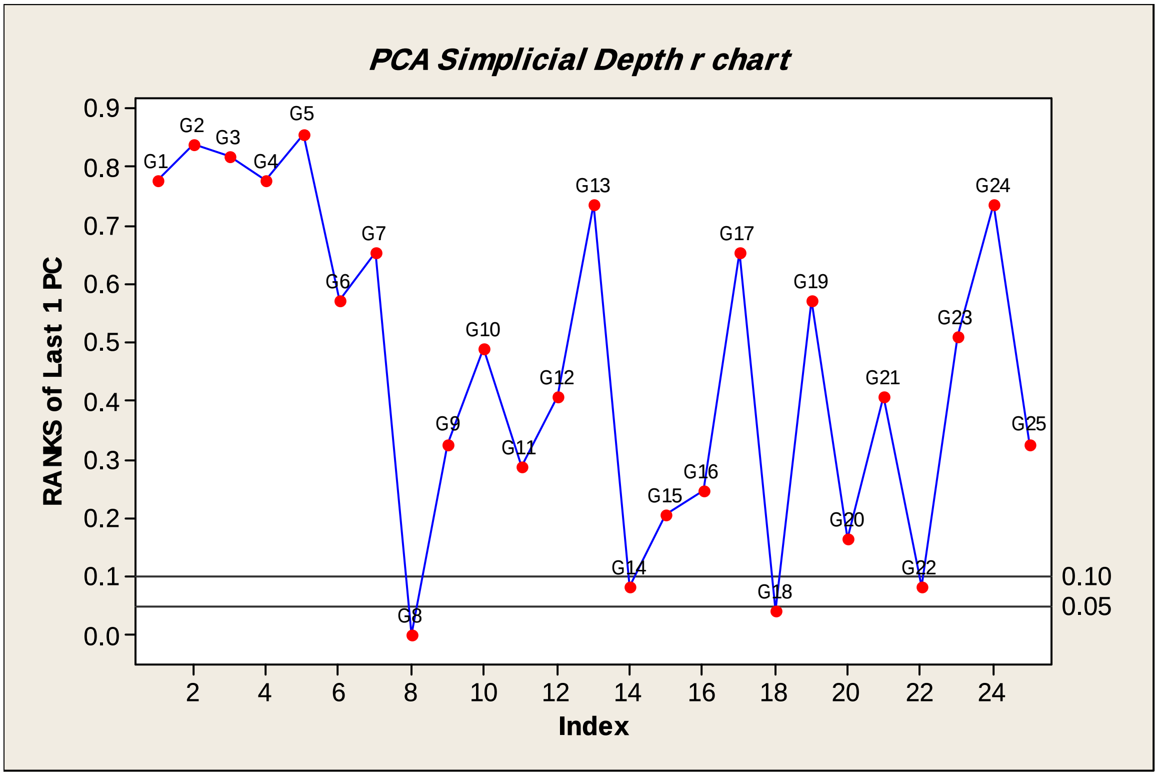

Figure 4.9d PCA Simplicial Depth r chart using the last PC for n = 49.

Using the HDS of n = 49, the following points from the 25 new observations were identified as out of control by the control chart when selecting the last PC:

At α = 0.05: Points G8 and G18

At α = 0.10: Points G8, G14 G18 and G22

Bivariate Data with Outlier Summary Table

Table 4.9f 25 NEW points with the out of control observations labeled **R(*n*=50 with one Random Error)** and **C(*n*=49 Corrected without the random error)**.

	α = 0.05		α = 0.10	
Name of point	**First PC**	**Last PC**	**First PC**	**Last PC**
G1				
G2				
G3				
G4				
G5	C		R C	
G6				
G7				
G8		R C		R C
G9				
G10	R C		R C	
G11				
G12				
G13				
G14				C
G15				
G16				
G17				
G18		R C		R C
G19				
G20			R	
G21				
G22				C
G23				
G24				
G25				

4.10 Summary and Contributions to Multivariate Quality Control

This section will provide a summary of our analyses, the outcomes of our findings with recommendations and their contributions to MSPC. In this research we have utilized theoretically sound distribution free techniques, namely simplicial depth and PCA, which have been proven to be robust (Liu, 2003) and adaptable to different industrial settings. Given the distribution free nature of the techniques that were utilized in this research, our proposed scheme can be implemented without having to perform diagnostics on distributional assumptions such as normality. Additionally, autocorrelation does not violate any assumptions for the validity of our procedure. We found that by unifying these nonparametric techniques, our approach is likely to be robust when the HDS is contaminated with a multivariate outlier that was a random error and not the result of process degradation. In Table 4.9f we find that the robustness appears more significant for the last PC which is used to monitor possible correlation shifts. We found that points G8 and G18 were both detected with small alpha values for both the random error (R) and corrected (C) data sets. Given that a random error reading may occur, our proposed control chart will likely generate few false alarms and not erroneously stop the process when it is still in control. This is most beneficial given that it will prevent an unnecessary interruption in the process that can be costly.

A significant contribution from this research has been to devise a detection scheme that will identify possible shifts in the correlation structure as well as the variability shifts. In his Principal Component Analysis text, Dunteman (1989) has stated that the last PC(s) may provide insight into the correlation structure between the variables. In multivariate quality control, Kourti and MacGregor (1995, 1996) and Milectic (2004) discuss the benefits of monitoring the remaining set of un-retained principal components that contributed to a small percentage of variability. However, their applications were based on normality, whereas our proposed PCA Simplicial Depth r chart for both the first and last PCs is nonparametric. We find that our proposed distribution free scheme for the correlation structure will detect signals with fairly low sample sizes. According to our comparisons, the control chart of the last PC appears to detect signals robustly, and slight improvements occur at times with small increases in sample size. Our findings indicate that at a low alpha, we can identify out of control points which may have been attributed to a correlation shift. We recommend setting the alpha level at 0.05 since signal detection is possible even with small samples. For very small samples, it may be necessary to raise alpha to 0.10 to avoid increases in missed alarms depending on the type of industrial process that is being monitored. As Mason and Young (2002) have suggested, in certain processes such as a chemical process, in order to avoid hazardous situations caused by missed alarms, it may be desirable to increase alpha. In our case studies, the final PC appears to generate robust signals. If too many of the last PCs are used in the analysis, we believe that some correlation signals may be masked by the added variability. We recommend that the last PC alone or the last PCs for which the cumulative variability is less than or equal to 0.009 be used to monitor correlation shifts. Table 4.10a illustrates the effect of sample size on the HDS when selecting the last PC. We find that a signal may be detected with sample sizes between 20 and 30 as was the case with the Steam Turbine, Industrial and Electrolyzer data sets. For the Industrial data set that had an HDS of n = 7, it is not surprising that signal detection was not possible. As such, we recommend that in situations with sample sizes smaller than 20, the HDS should be augmented to at least 20 since we found that n = 20 seemed to provide fairly good results. From the results in Table 4.10a, we find that our charts are capable of identifying highly significant points for correlation shifts such as A2, C17, 1021 and H3 with a 0.05 alpha level and sample sizes between 20 and 40. It appears that there are some false alarms and some missed points until the sample size gets close to 50, when it seems to converge to a particular set of points chosen. Additional points such as B24, D6 and D36 were detected at the alpha level of 0.10 for the smaller sample sizes, and subsequently at a lower alpha value as n increases. Hence, adjusting α depending on n as well as the consequences of missed versus false alarms may be a reasonable approach.

Table 4.10a Signals from the **Last 1 PC** and the effect of sample size

	<20	**20 – 29**	**30 – 39**	**40 – 49**	**50 – 59**	**60 or >**
Steam Turbine $\alpha = 0.05$		1 pt A2	1 pt A2			
Fruit Juice $\alpha = 0.05$			NONE	2 pts B24, B31		1 pt B24
$\alpha = 0.10$			2 pts B14, B24	2 pts B24, B31		2 pts B24, B31
Industrial $\alpha = 0.05$	NONE	1 pt C17		2 pts C17, C58		
$\alpha = 0.10$	NONE	4 pts C13, C17 C37, C55		2 pts C17, C58		
Aluminum Pin $\alpha = 0.05$			1 pt D36	2 pts D6, D36	1 pt D6	2 pts D6, D36
$\alpha = 0.10$			4 pts D6, D7, D35, D36	2 pts D6, D36	2 pts D6, D36	2 pts D6, D36
Electrolyzer $\alpha = 0.05$		1021				
Mechanical $\alpha = 0.05$			H3			

Our proposed control chart scheme is twofold. We have discovered that it appears possible to utilize the proposed control chart scheme to detect two different types of signals, correlation and variability shifts. We have just offered our recommendations based on our findings of the last PC(s), however, the challenge in our findings remains at the selection of the first set of PCs to monitor the possible shifts caused by variability. In our analyses, we find that the signal detection becomes extremely erratic to the point where true signal detection is lost when high cumulative proportions of variability are used as the cutoff for the number of first PCs to be retained. We investigated the possibility of choosing a cumulative percentage variability cutoff of 80% or 70% from the eigenanalysis for each data set and found that we risk either the loss of signals which could lead to hazardous conditions or excessive false alarms which are costly. From our control charts, the cutoff of 60% cumulative variability appears to generate the best compromise to provide a reasonable risk of false alarms with possibly minimal missed alarms. Using different sample sizes, we find that the cutoff of 60% cumulative variability will provide the most consistent signal detection. Additionally, we find that using a cumulative percent of variability explained by less than or equal to 60% neutralizes the effect of the autocorrelation. Table 4.10b provides a comparison of the results of selecting the first PCs and the corresponding points that signal using our recommended 60% variability cutoff with changes in sample size. When plotting the control chart of the first PCs that have been retained using the 60% suggested rule, the points that appeared to be highly significant such as A4, B2, B31, D17, D36, 834 and 963 were detected with low alpha levels even with sample sizes below 40. Additional points that were not detected at 0.05 such as C1 and C16 may be unmasked with the 0.10 alpha. Subsequently, these additional points will be detected with low alpha levels as the sample size increases.

Table.4.10b Signals from the first PCs with **less than 60%** total variation as *n* changes

	<20	20 – 29	30 – 39	40 – 49	50 – 59	60 or >
Steam Turbine $\alpha = 0.05$		**1 PC 1 pt A4	1 PC 2 pts A4, A6			
$\alpha = 0.10$		**1 PC 2 pts A4, A6	1 PC 2 pts A4, A6			
Fruit Juice $\alpha = 0.05$			1 PC 2 pts B2, B31	1 PC 2 pts B2, B31		1 PC 2 pts B2, B31
Industrial $\alpha = 0.05$	1 PC NONE	2 PCs NONE		2 PCs 4 pts C1, C16, C55, C61		
$\alpha = 0.10$	1 PC NONE	2 PCs 3 pts C1, C5, C16		2PCs 4 pts C1, C16, C55, C61		
Aluminum Pin $\alpha = 0.05$			**1 PC 2 pts D17, D36	**1 PC 1 pt D36	**1 PC 1 pt D17	**1 PC 1 pt D17
$\alpha = 0.10$			**1 PC 4 pts D17, D22, D28, D36	**1 PC 2 pts D22, D36	**1 PC 2 pts D17, D36	**1 PC 2 pts D17, D36
Electrolyzer $\alpha = 0.05$		1 PC 2 pts 834, 963				
Mechanical $\alpha = 0.05$			2 PCs NONE			
$\alpha = 0.10$			2 PCs 4 pts H1, H4, H6, H7			

**For these cases, the first PC was above the 60% cutoff.

In our comparisons illustrated in Tables 4.10a and 4.10b, we did not include the bivariate analyses given their unique structure. In those cases, our suggested 60% cutoff percentage of cumulative variability would not be feasible. For the bivariate cases, we propose using the PCA simplicial depth r chart for each PC individually regardless of the proportion of the variability explained by each. We believe this is the only way to address both, a detection of shift in correlation structure as well as increase in variability. The literature in MSPC has indicated the benefits of using PCA in an industrial setting. Applications of projection methods such as Principal Components Analysis (PCA), Partial Least Squares and Residual Analysis have been investigated and determined to be beneficial in different industrial settings that consist of multivariate processes. Industries, in which projection methods have been well received given their ease of use and interpretability, have included manufacturing, automotive, chemical and food industries. (Kourti and MacGregor 1995, 1996, 2005) However, in many MSPC applications of these projection methods, the use of PCA has been in conjunction with the T^2 statistic, either by using the MYT decomposition (Mason, Young and Tracy, 1997) or the Kourti and MacGregor (1996) adaptations of the T^2, thus depending on multivariate normality. Our approach utilizes the distribution-free property of projection methods such as PCA and simplicial depth, thus multivariate normality will not be required and hence independence is likewise unnecessary. This contribution is significant in industry given the need to develop nonparametric techniques in MSPC to eliminate the need to assume multivariate normality (Coleman, 1997) and given the presence of autocorrelation in many cases.

An additional advantage to our approach is the computation and computer aspect for implementation. In terms of software use, most statistical software packages, including Minitab and Statistica among others, include the multivariate analysis feature, and perform PCA with the eigenanalysis, eigenvectors, scree graphs and PC scores generated instantly. We found that computing simplicial depths was feasible within an efficient timeframe which would be beneficial to the user. Rousseeuw and Ruts (1996) designed an algorithm for an efficient computation in the bivariate case for extremely large values of *n*. In higher dimensionality, the computation of simplicial depth is even more challenging which has motivated research for efficient algorithms. (Mustafa, 2004) However, we utilized Mathematica and by solving systems of linear equations for our scheme with reduced dimensionality, the computations were obtained rather quickly.

Additionally, although we investigated using multiple PCs with high proportions and low proportions of variability explained, we found that usually we only needed one or two of the first set of PCs and one PC on the back end with low variability. This also simplified the computations. Once the simplicial depths and ranks are computed, the univariate control chart scheme for plotting the ranks can be generated using any statistical software package with the points labeled clearly as we illustrated. This provides the process user a chart that is easy to use and easy to interpret particularly since it is so close in appearance to the old univariate control chart based on normality. Additionally, our proposed PCA-Simplicial Depth r control chart is based on theoretically sound nonparametric multivariate techniques to identify signals for monitoring the ranks of the simplicial depths of the PCs. Other adaptations may be possible as new developments are found in the theoretical nonparametric multivariate statistical field. Table 4.10c summarizes the outcomes from our proposed PCA Simplicial Depth r-control chart scheme.

Table 4.10c Outcomes from our proposed PCA Simplicial r-chart.

Advantages of this Approach in MSPC
Integrated theoretically sound techniques in a unified nonparametric MSPC approach • *No distribution is assumed; normality is not required* • *Autocorrelation of observations is not a problem* • *Adaptation of existing techniques that are easy to use and recognized in Multivariate Quality Control Literature*
Monitors and Identifies Variability and Correlations Shifts • *First PC(s) for variability* • *Last PC(s) for correlation*
Robustness in MSPC • *Will not stop the process unnecessarily* • *Low sensitivity to outliers which are not caused by process degradation* • *Low false alarm rates*
Easy to use in different industrial settings that consist of multivariate processes • *Adaptation of existing techniques that are easy to use and recognized in Quality Control Literature* • *PCA-readily available in software and recognized in MSPC* • *Simplicial Depth computations of reduced dimensionality are rapid* • *Graphed in an easy to use univariate scheme*
Foster quality improvement in an industrial setting
Adaptability • *Can be incorporated into new multivariate nonparametric test statistics that may be developed in the future*

CHAPTER FIVE: CONCLUSION

5.1 Summary

Based on our findings, we have developed a unified distribution free scheme in MSPC that has proven to be robust, efficient and easy to understand. This contribution is significant in industry given that it provides an alternative approach to MSPC that will be free of distributional assumptions, such as normality and the requirement of independence of errors. As we found consistently in the quality literature, there have been significant developments in univariate nonparametric SPC (Chakrati, Van Der Laan, and Bakir , 2001), however, the research and development of nonparametric multivariate quality schemes has been limited. (Stoumbos, 2001) We sought to develop a technique utilizing multivariate nonparametric methods that would prove to be easy to use and interpret in MSPC.

The nonparametric data depth charts proposed by Liu (1995) have been well received in the multivariate quality literature given their affine invariance and their ease of use. As with the T^2 control chart, the control chart scheme by Liu is univariate which is simple to plot and interpret. The readability of a univariate plot provides the process owner a quick and efficient graphing scheme such that the out of control signal can be quickly detected. Additionally, we found that the computations needed for our proposed nonparametric Simplicial Depth r chart will be easy to implement.

With the affine invariance property satisfied, the statistic utilized by Liu is invariant to transformations, scaling and rotations which are necessary when utilizing multivariate dimensionality reduction schemes including Principal Component Analysis. In the multivariate quality literature, it has been stipulated that a subset of all variables will drive a multivariate process. Thus, it is best to utilize a dimensionality reduction scheme such as Principal Component Analysis (PCA) that will identify the components that drive the process. PCA is a technique that is well known and established in both the theoretical multivariate statistics field as well as in the industrial multivariate quality area. Given its recognition and availability in software, different industries that are driven by multivariate processes have acknowledged the benefit and adaptability of PCA. According to the literature in MPSC, PCA has traditionally been used in conjunction with the T^2 statistic (Mason and Young, 2002) or modified forms of the T^2 statistic (Kourti and MacGregor, 1995) thus imposing normality. In our proposed scheme, we found that PCA, given its distribution free nature, can be utilized beyond traditional parametric approaches and incorporated into nonparametric MSPC.

In order to test our proposed scheme and its adoptability in different industrial processes, we gathered numerous real multivariate data sets from the chemical, food and manufacturing industries. The Historical Data Sets (HDS) and new observations were available in the literature and from an undisclosed company data that we analyzed. We performed additional runs with different sample sizes on the HDS in order to test the effect of sample size on our proposed scheme. In our different analyses, we found that increases in sample size lowered false alarm rates. An important finding for the small sample case was that signal detection was still possible. We found that by retaining the initial PCs that explain no more than 60% total variation, the autocorrelation effect is neutralized. We plotted ranks of the Simplicial Depths of the first PC(s) and found that retaining the first PC(s) that account for at most 60% cumulative variability provide robust results even with small samples. The exception to this rule is when the first PC accounts for more than 60% variability in which case the only possibility is to retain the first PC. This holds true for the bivariate case where the first PC is the one that would be retained for detecting significant shifts in the process due to variability..

Additionally, we find that applying our proposed nonparametric scheme to the last PC(s) provides insight into the correlation structure of the multivariate process, which has been a challenge. We recommend selecting either the last PC, or using a cutoff of 0.009 for the maximum cumulative variability if more than one of the final PCs will be retained for the correlation charts. For the bivariate processes, we recommend utilizing the last PC for correlation signal identification. This is different from current applications of projection methods (Kourti and MacGregor, 2003) which utilize the remaining $p - k$ PC(s) with the Squared Prediction Error, or Q-statistic, which is dependent on normality. As with the PCA Simplicial Depth r-chart of the first PC(s), our proposed scheme for the last PC(s) is distribution free. Our proposed PCA Simplicial Depth r-chart of the last

PC(s) identified signals with small samples usually at the 0.05 alpha level. This implies that with relatively low false alarm rates a signal generated by a correlation shift can be detected even in the small sample cases.

5.2 Future Research

Our findings are based on utilizing nonparametric methods such as PCA and the ranks of the robust Simplicial Depths for an improved signal detection scheme in MSPC. Our research focuses on the plots of the ranks of the PC scores of the observations on the r control chart, which is a nonparametric multivariate analog of the Shewhart univariate X chart as Liu (1995) and Zarate have indicated. We recommend extending our findings in several ways. One may be to devise a scheme to monitor the expected value which could represent an analog to the Shewhart univariate $\overline{X}$ chart. Additional control charts to complement PCA and robust simplicial depths may be developed to assist in extending beyond our variation and/or correlation detection in an attempt to analyze the variables themselves. There is also an opportunity to fine tune the work on sample size requirements. Also, given the numerous applications of projection methods such as PCA, it is possible to investigate some of the other projection methods that represent adaptations of PCA including ICA (Independent Component Analysis) and DPCA

(Dynamic Principal Component Analysis).

Currently, PCA is the technique that is available in most statistical software packages which facilitates implementation into an industrial application. However, as software developers continue to enhance these statistical programs to include more multivariate techniques, these additional projection methods may be utilized more in MSPC and may be used to extend our research. With the computational power of modern computers and research that is currently done in computational geometry by Mustafa (2004) and Miller and Rousseeuw (2003), we believe that these control chart schemes will eventually be programmed in a statistical software package in the same manner as control charts such as the X, $\overline{X}$ and T^2 which are generated effortlessly. This research provides those in the computer programming field and the quality control area the opportunity for joint research and development. We find that our proposed PCA Simplicial Depth r-chart scheme has created an alternative to current developments and also generates opportunities for additional research in various fields, most significantly in nonparametric multivariate quality control.

APPENDIX A
NONPARAMETRIC MULTIVARIATE STATISTICS

During the last six decades, a group of researchers have developed multivariate nonparametric statistics but these have not been adopted by the SPC users. In some cases this may be due to the fact that some of the statistics have properties which are considered undesirable (such as not being affine invariant). In other cases, the statistics may be difficult to use or limited to bivariate data. Given the need in industry to develop an approach which is easy to use and interpret, for our research we have chosen not to utilize these theoretical developments. These statistics are enumerated and their advantages and disadvantages from a theoretical perspective are given in Table A.

Nonparametric Statistics from Multivariate Analysis

which have not been fully adopted in MSPC

Table A Multivariate Analysis Developments

Statistic	Advantages	Disadvantages
Bennet's multivariate sign test with an approximate chi-squared test statistic	Provided information for 2, 3 and in certain situations 4 variables	Property of affine invariance is not satisfied. Based on properties of multivariate normal distributions
Bennet's bivariate signed rank test	Monitors location	Property of affine invariance is not satisfied. Could only monitor bivariate data
Bickel's vector of medians	Monitors vector of medians	Property of affine invariance is not satisfied. Efficiency decreases as the correlation increases
Chatterjee's bivariate sign test	Monitors location	Property of affine invariance is not satisfied.
Gower's Mediancentre	Bivariate location measure to minimize the sum of absolute distances to observations	Property of affine invariance is not satisfied. Could only monitor bivariate data

Utts and Hettmansperger vector of Winsorized rank statistics	Robust nonparametric approach to outliers More efficient than Hotelling's T^2 statistic in non-normal distributions Kapatou Nonparametric Control Chart scheme	Property of affine invariance is not satisfied PCA cannot be applied in control chart scheme due to lack of affine invariance
Brown's spatial median for spatial data	Spatial median according to Brown was more efficient than the mean	Property of affine invariance is not satisfied. Could only monitor bivariate data
Oja's Generalized multivariate median	Measure location, scatter, skewness and kurtosis Property of affine invariance is satisfied	Computation is highly intensive
Brown and Hettmansperger's analogues to the Wilcoxon signed rank sum and the Mann-Whitney-Wilcoxon rank sum test.	Property of affine invariance is satisfied	Limited to bivariate data
Randles, multivariate analog to the sign test.	Property of affine invariance is satisfied More efficient than Hotelling's T^2 statistic in heavy-tailed distribution Small sample distribution-free approach	Sign Test are not as powerful as other nonparametric statistics
Peters and Randles' multivariate analog to the signed rank test	Property of affine invariance is satisfied Performs well for a light-tailed distribution As efficient as T^2 statistic in multivariate normal distributions	Performs poorly in heavy-tailed situations Less efficient than Randles sign test

APPENDIX B
CONVERGENCE OF THE UNIFORM DISTRIBUTION

The following is an overview of the convergence of the simplicial depth rank as demonstrated by Liu and Singh (1993)

Definition Monotonicity Property for the data depth function $D(F;\ \cdot)$ from Liu and Singh (1993).

$D(F;\ \cdot)$ is monotonically decreasing along any fixed ray from the center θ_0 of the distribution. The distribution *F* is symmetric around point θ_0. Given the symmetry around θ_0 if *F* is the distribution of X, then $(X-\theta_0)$ and $(\theta_0 - X)$ will have the same distribution. If $D(F;\theta_0+\alpha(x-\theta_0)) \geq D(F;x)$ for every x and for α such that $0 \leq \alpha \leq 1$.

Theorem: If $D(F;X)$ has a continuous distribution under *F* and assume that $\lim_{m\to\infty} \sup_{x\in\mathbf{R}^p} |D(F_m;x) - D(F;x)| = 0$ assuming the distributions *F* = *G* conditionally on X, we have that as $m \to \infty$, the rank statistic $R(F_m;Y_1) \to U[0,1]$ along almost all X sequences. (Liu and Singh, 1993)

Proof:

According to Liu and Singh (1993), to prove this theorem "it suffices to show that $R(F_m;y_1)$ converges to $R(F;y)$ for almost all fixed *y* (with respect to *F*) along almost all sequences **X**."

“Fix a sequence **X**” such that the condition $\lim_{m\to\infty} \sup_{x\in\mathbf{R}^p} \left|D(F_m;x)-D(F;x)\right| = 0$ holds.

We have that for any $\varepsilon > 0$, $\sup_{y\in\mathbf{R}^p} \left|D(F_m;y)-D(F;y)\right| \le \frac{\varepsilon}{2}$ such that every $m \ge m_0$ for some m_0.

Therefore for all $m \ge m_0$, if $\varepsilon \to 0$ we can deduce the following:

$$\{Y : D(F;Y) \ge D(F;y)+\varepsilon\} \subseteq \{Y : D(F_m;Y) \ge D(F_m;y)\}$$
$$\subseteq \{Y : D(F;Y) \ge D(F;y)-\varepsilon\}$$

Theorem: Let $\{f_n(x)\}_{n=1}^{\infty}$ be a sequence of functions defined on a set $A \subset R$ and converges to $f(x)$.

Define λ_n:

$$\lambda_n = \sup_{x\in A} \left| f_n(x) - f(x) \right|$$

The sequence $\{f_n(x)\}_{n=1}^{\infty}$ converges uniformly to $f(x)$ on A if and only if $\lambda_n \to 0$ as $n \to \infty$.

Proof: Two parts: sufficiency and necessity. (Khuri,1993)

Sufficiency: Prove: If $\lambda_n \to 0$ as $n \to \infty$, then the sequence $\{f_n(x)\}_{n=1}^{\infty}$ converges uniformly to $f(x)$ on A.

Suppose $\lambda_n \to 0$ as $n \to \infty$. Let $\varepsilon > 0$ be given. Then there exists some integer N such that $\lambda_n < \varepsilon$ for $n > N$. Therefore, we get

$$|f_n(x) - f(x)| \leq \lambda_n < \varepsilon \quad \text{for every } x \in A.$$

Given that N depends only on ε, the sequence $\{f_n(x)\}_{n=1}^{\infty}$ converges uniformly to $f(x)$ on A.

Necessity: If the sequence $\{f_n(x)\}_{n=1}^{\infty}$ converges uniformly to $f(x)$ on A, then $\lambda_n \to 0$ as $n \to \infty$.

Suppose that sequence $\{f_n(x)\}_{n=1}^{\infty}$ converges uniformly to $f(x)$ on A. Let $\varepsilon > 0$ be given. Then there exists some integer N depends only on ε such that for $n > N$,

$$|f_n(x) - f(x)| < \frac{\varepsilon}{2} \text{ for every } x \in A.$$

Consequently, we get $\lambda_n \leq \dfrac{\varepsilon}{2}$

In other words, $\lambda_n = \sup\limits_{x \in A} |f_n(x) - f(x)| \leq \dfrac{\varepsilon}{2}$. Therefore, $\lambda_n \to 0$ as $n \to \infty$.

For the data depth functions we have $\sup\limits_{y \in \mathbf{R}^p} |D(F_m; y) - D(F; y)| \leq \dfrac{\varepsilon}{2}$ and similarly $R(F_m; y_1)$ converges uniformly to $R(F; y)$ as such the rank statistic $R(F_m; Y_1) \to U[0,1]$ along almost all X sequences. (Liu and Singh, 1993)

LIST OF REFERENCES

Alloway, J.A., Raghavachari, M. (1991) "An Introduction to Multivariate Control Charts" *45th Annual Quality Congress Transactions, American Society for Quality Control*, Milwaukee, Wisconsin, pp. 773 – 783

Alwan, L.C., Roberts, H.V., (1995) "The Problem of Misplaced Control Limits" *Applied Statistics* 44, pp. 269 – 278

Andersson, M., Passare, M., Sigurdsson, R. (2004) *Complex Convexity and Analytic Functionals*, Basel: Birkhauser Verlag

Bennett, B. M. (1962) "On Multivariate Sign Tests" *Journal of the Royal Statistical Society, Series B: Methodological* 24, pp. 159 – 161

Bennett, B. M. (1964) "A Bivariate Signed Rank Test" *Journal of the Royal Statistical Society, Series B: Methodological* 26(3), pp. 457 – 461

Bickel, P. J. (1964) "On Some Alternative Estimates for Shift in the P–Variate One Sample Problem" *Annals of Mathematical Statistics* 35, pp. 1079 – 1090

Bickel, P. J. (1964) "On Asymptotically Nonparametric Competitors of Hotelling's T^2" *Annals of Mathematical Statistics* 36, pp. 160 – 173

Box, G.E.P, Luceño, A. (1997) *Statistical Control by Monitoring and Feedback Adjustment*, New York: Wiley

Brown, B. M. (1983) "Statistical Uses of the Spatial Median" *Journal of the Royal Statistical Society Series B: Methodological* 45(1), pp. 25 – 30

Brown, B. M., Hettmansperger, T. P. (1987) "Affine Invariant Rank Methods in the Bivariate Location Model" *Journal of the Royal Statistical Society Series B: Methodological* 49(3), pp. 301 – 310

Brown, B. M., Hettmansperger, T. P. (1989) "An Affine Invariant Version of the Sign Test" *Journal of the Royal Statistical Society Series B: Methodological* 51(1), pp. 117 – 125

Calzada, M.E., Scariano, S.M. (2001) "The Robustness of the Synthetic Control Chart to Non-normality" *Communications in Statistics: Simulation and Computation* 30, pp. 311 – 326

Chakrati, S. Van Der Laan, P., Bakir, S.T. (2001) "Nonparametric Control Charts: An Overview and Some Results" *Journal of Quality Technology* 33, No. 3, pp. 304 – 315

Champ, C. W., and Woodall, W. H. (1987) "Exact Results for Shewhart Control Charts with Supplementary Runs Rules," *Technometrics* 29, pp. 393 – 399

Chatterjee, S. K. (1966) "A Bivariate Sign Test for Location" *Annals of Mathematical Statistics* 37, pp. 1771 – 1782

Chou, Y. M., Mason, R.L, Young, J.C. (2001) "The Control Chart for Individual Observations from a Multivariate Non-Normal Distribution" *Communication in Statistics* 30, pp. 1937 – 1949

Coleman, D.E. (1997) "Individual Contributions" in "A Discussion on Statistically–Based Process Monitoring and Control" edited by D.C. Montgomery and W.H. Woodall. *Journal of Quality Technology* 29, pp. 148 – 149

Dai, Y., Zhou, C., Wang, Z. (2004) "Multivariate CUSUM Control Charts Based on Data Depth For Preliminary Analysis" The Natural Sciences Foundation of Tianjin 033603111

Davis, R.B., Woodall W.H. (2002) "Evaluating and Improving the Synthetic Control Chart" *Journal of Quality Technology* 34, pp. 200 – 208

DeVor, R. E., Chang, T., Sutherland, J.W. (1992) *Statistical Quality Design and Control: Contemporary Concepts and Methods*, New York: Prentice Hall

Dietz, E.J. (1982) "Bivariate Nonparametric Tests for One–Sample Location Problem" *Journal of the American Statistical Association* 77, pp. 163 – 169

Eakin, B. K., McMillen, D.P., Buono, M.J., (1990) "Constructing Confidence Intervals using the Bootstrap: An Application to a Multi–Product Cost Function" *Review of Economics and Statistics* 72, pp. 339 – 344

Dunteman, G. H. (1989) *Principal Components Analysis*, Newbury Park: Sage

Easton, G. McCulloch, R. (1990) "A Multivariate Generalization of Quantile-Quantile Plots" *Journal of American Statistical Association* 85, pp. 376 – 386

Elsayed, E.A. (2000) “Perspectives and Challenges for Research in Quality and Reliability” *International Journal of Production Research* 38, pp. 1953 – 1976

Faltin, F.W., Mastrangelo, C.M., Runger, G.C., Ryan, T.P. (1997) “Considerations in the Monitoring of Autocorrelated and Independent Data” *Journal of Quality Technology* 29, pp. 131 – 133

Fuchs, C., Kenett, R. (1998) *Multivariate Quality Control*, New York: Marcel Dekker

Goetsch, D. L., Davis S.B. (2000) *Quality Management: Introduction to Total Quality Management for Production, Processing and Services,* Columbus: Prentice–Hall

Hackl, P., Ledolter, J. (1992) “A New Nonparametric Quality Control Technique” *Communications in Statistics – Simulation and Computation* 21, pp. 423 – 443

Hawkins, D.M. (1993) “Regression Adjustment for Variables in Multivariate Quality Control” *Journal of Quality Technology* 25, pp. 170 – 182

Hawkins, D.M. (1993) “Robustification of Cumulative Sum Charts by Winsorization” *Journal of Quality Technology* 25, pp. 248 – 261

Hayter, A.J., Tsui, K.L. (1994) “Identification and Quantification in Multivariate Quality Control Problems” *Journal of Quality Technology* 29, pp. 131 – 133

Hoerl, R. (2000) “Discussion on ‘Controversies and Contradictions in Statistical Process Control’” *Journal of Quality Technology* 32, pp. 351 – 355

Hotelling, H. (1947) “Multivariate Quality Control” *Techniques of Statistical Analysis*, eds. C. Eisenhart, M.W. Hasty, W.A. Wallis, New York: McGraw Hill, pp. 111 – 184

Jackson, J. (1959) “Quality Control Methods for Several Related Variables” *Technometrics* 1, pp. 359 – 377

Jackson, J. (1991) *A User’s Guide to Principal Components*, New York: Wiley & Sons

Janacek, G.J., Meikle, S.E. (1997) “Control Charts Based on Medians” *The Statistician* 46, pp. 19 – 31

Johnson, R.A., Wichern D.W. (1982) *Applied Mathematics Statistical Analysis*, Englewood Cliffs, New Jersey: Prentice Hall

Jolliffe, I.T. (1986) *Principal Component Analysis*, New York: Springer-Verlag

Jones, L.A., Woodall, W.H. (1998) "The Performance of Bootstrap Control Charts" *Journal of Quality Technology* 30, pp. 363 – 375

Justel, A., Pena, D., Zamar, R. (1997) "A Multivariate Kolmogrov-Smirnov Test of Goodness of Fit" Statistics and Probability Letters 35, pp. 251 – 259

Juran, J. (Sept 1997), "Early SQC: A Historical Supplement" *Quality Progress* 30pt2, pp 73 – 81

Kapatou, A. (1996) "Multivariate Nonparametric Control Charts using Small Samples" PhD DISSERTATION, Department of Statistics, Virginia Polytechnical Institute and State University , Blacksburg, Va.

Khattree, R., Naik, D. (2000) *Multivariate Data Reduction and Discrimination*, Cary, North Carolina: Wiley & Sons

Khuri, A. (1993) *Advanced Calculus with Applications in Statistics*, New York: Wiley & Sons

Kirkwood, J. (1989) *An Introduction to Analysis*, Boston: PWS-Kent

Klein, M. (2000) "Two Alternatives to the Shewart X-bar Control Chart" *Journal of Quality Technology* 32, pp. 427 – 431

Kolarik, William J. (1995) *Creating Quality: Concepts, Systems, Strategies, and Tools* New York: McGraw Hill

Kourti, T. (2003) "Abnormal Situation Detection, Three-way Data and Projection Methods; Robust Data Archiving and Modeling for Industrial Applications" *Annual Reviews in Control* 27 (2), pp. 131 – 138

Kourti, T. (2005) "Application of Latent Variable Methods to Process Control and Statistical Process Control in Industry" (special issue on condition monitoring) *International Journal of Adaptive Control and Signal Processing*, 19, pp. 213 – 246

Kourti, T., Lee, J., MacGregor, J.F. (1996) "Experiences with Industrial Applications of Projection Methods for Multivariate Statistical Process Control" *Computers Chemical Engineering* 20 (supplement), pp. S745 – S750

Kourti, T., MacGregor, J.F. (1995) "Process Analysis, Monitoring and Diagnosis Using Multivariate Projection Methods-A Tutorial" *Chemometrics and Intelligent Laboratory Systems* 28, pp. 3 – 21

Kourti, T., MacGregor, J.F. (1995) “Statistical Process Control of Multivariate Processes” *Chemometrics and Intelligent Laboratory Systems Control Engineering Practice* 3(3), pp. 403 – 414

Kourti, T., MacGregor, J.F. (1996) “Recent Developments in Multivariate SPC Methods for Monitoring and Diagnosing Process & Product Performance” *Journal of Quality Technology* 28 (4), pp. 409 – 428

Larson, R., Edwards, B. H., Falvo, D. C. (2004) *Elementary Linear Algebra* 5th ed., Boston: Houghton Mifflin

Ledolter, J., Swersey, A. (1997) “An Evaluation of Pre-Control” *Journal of Quality Technology* 29, pp. 163 – 171

Linna, K.W., Woodall, W.H. (2001) “Effect of Measurement Error on Shewart Control Charts” *Journal of Quality Technology* 33, pp. 213 – 222

Linna, K.W., Woodall, W.H. (2001) “The Performance of Multivariate Control Charts in the Presence of Measurement Error” *Journal of Quality Technology* 33, pp. 349 – 355

Liu, R.Y. (1990) “On a Notion of Data Depth Based on Random Simplices” *Annals of Statistics* 18, pp. 405 – 414

Liu, R.Y. (1995) “Control Charts for Multivariate Processes” *Journal of the American Statistical Association*, 90pt2, pp. 1380 – 1387

Liu, R.Y. (2003) “Data Depth: Center-Outward Ordering of Multivariate Data and Nonparametric Multivariate Statistics” *Recent Advancements and Trends in Nonparametric Statistics* eds. M. Akritas and D. Politis, Amsterdam, Netherlands: Elsevier, pp. 155 – 167

Liu, R., Singh, K. (1993) “A Quality Index Based on Data Depth and Multivariate Rank Test” *Journal of the American Statistical Association*, 88, pp. 257 – 260

Lowry, C.A., Montgomery, D.C. (1995) “A Review of Multivariate Control Charts” *IIE Transactions* 27, pp. 800 – 810

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

Mahalanobis, P. C. (1936) “On the Generalized Distance in Statistics” *Proceeding of the National Academy India*, Vol. 12, 49-55

Martens, H., Martens, M. (2001) *Multivariate Analysis of Quality: An Introduction*, New York: Wiley and Sons Publishing

Mason, R.L., Young, J.C. (2005) "Multivariate Tools: Principal Component Analysis" *Quality Progress*, pp. 83 – 85

Mason, R. L., Young, J.C. (2002) "Variation in SPC: How the Two Types of Variation Can Each Affect a Control Procedure" *Quality Progress*, pp. 83 – 85

Mason, R.L., Young, J.C. (2002) *Multivariate Statistical Process Control with Industrial Applications*, Philadelphia: ASA-Siam

Mason, R.L., Young, J.C. (1998) "Hotelling's T^2: A Multivariate Statistic for Industrial Process Control" *ASQ 's 52nd Annual Quality Congress Proceedings*, pp. 78 – 85

Mason, R.L., Champ, C.W., Tracy, N.D., Wierda, S.J., Young, J.C. (1995) "Decomposition of T^2 for Multivariate Control Interpretation" *Journal of Quality Technology* 27, pp. 99 – 108

Mason, R.L., Champ, C.W., Tracy, N.D., Wierda, S.J., Young, J.C. (1995) "A Practical Approach for Interpreting Multivariate T^2 Control Chart Signals" *Journal of Quality Technology* 27, pp. 99 – 108

Mason, R.L., Champ, C.W., Tracy, N.D., Wierda, S.J., Young, J.C. (1997) "Assessment of Multivariate Control Techniques" *Journal of Quality Technology* 29, pp. 140 – 143

Messaoud, A., Weihs, C., Hering, F. (2004) "A Nonparametric Multivariate Control Chart Based on Data Depth" Technical Report 61, Sonderforschungsbereich 475, University of Dortmund, Dortmund, German

Magaril–Il'yaev, G. G., Tikhomirov, V. M. (2003) *Convex Analysis*, Providence: American Mathematical Society

Milectic , I., Quinn, S., Dudzic, M. , Vaculik, V., Champagne, M. (2004) "An Industrial Perspective on Implementing On-line Applications of Multivariate Statistics" *Journal of Process Control* 14, pp. 821 – 836

Miller, K., Ramaswami, S., Rousseeuw, P., et. al. (2003) "Efficient Computation of Location Depth Contours by Methods of Computational Geometry" *Statistics and Computing*, Vol. 13

Munkres, J. (1975) *Topology A First Course*, Englewood Cliffs: Prentice-Hall

Mustafa, N. H. (2004) “Simplification, Estimation and Classification of Geometric Objects” PhD DISSERTATION, Department of Computer Science, Duke University, Durham, North Carolina

Oja, H. (1983) “Descriptive Statistics for Multivariate Distributions” *Statistics and Probability Letters* 1, pp. 372 – 377

Oja, H., Niinimaa, A. (1985) “Asymptotic Properties of the Generalized Median in the Case of Multivariate Normality” *Journal of the Royal Statistical Society Series B: Methodological* 47(2), pp. 372 – 377

Peters, D. (1991) “An Adaptive Multivariate Signed–Rank Test for the One–Sample Location Problem” *Nonparametric Statistics* 1, pp. 157 – 163

Peters, D., Randles, R. H. (1990) “An Multivariate Signed–Rank Test for the One–Sample Location Problem” *Journal of the American Statistical Association* 85, pp. 552 – 557

Quesenberry, C.P. (1991) “SPC charts for Start – up Processes and Short or Long Runs” *Journal of Quality Technology* 23, pp. 213 – 224

Quesenberry, C.P. (1993) “The Effect of Sample Size on Estimated Limits for $\overline{X}$ and X Control Charts” *Journal of Quality Technology* 23, pp. 296 – 303

Randles, R.H. (1989) “A Distribution–Free Multivariate Sign Test Based on Interdirections” *Journal of the American Statistical Association* 84, pp. 1045 – 1050

Rencher, A. (1995) *Applied Nonparametric Methods of Multivariate Analysis*, New York: Wiley & Sons

Rousseeuw, P., Leroy, A. (1987) *Robust Regression and Outlier Detection*, New York: Wiley & Sons

Rousseeuw, P. (1990) “Unmasking Multivariate Outliers and Leverage Points” *Journal of the American Statistical Association* 85, pp. 633 – 639

Rousseeuw, P., Ruts, I. (1996) “Bivariate Location Depth” *Applied Statistics* 45 (1996), pp. 516-526

Runger, G. C., Willemain, T. R. (1995) “Model-Based and Model-Free Control of Autocorrelated Processes” *Journal of Quality Technology* 27, pp. 283 – 292

Schilling, E.G., Nelson, P.R. (1976) "The Effect of Non-normality on the Control Limits of X charts" *Journal of Quality Technology* 8, pp. 183 – 188

Shewhart, W. A. (1931). *Economic Control of Quality of Manufactured Product* New York: Van Nostrand

Shewhart, W. A. (1939) *Statistical Method from the Viewpoint of Quality Control*, New York: Dover reprinted 1986

Sprent, P. (1989) *Applied Nonparametric Methods*, New York: Chapman and Hall

Stoumbos, Z.G., Jones, L.A, Woodall, W.H., and Reynolds, M.R., Jr. (2001) "On Shewart – Type Nonparametric Multivariate Control Charts Based on Data Depth" in *Frontiers in Statistical Quality Control 6*, eds. H.-J. Lenz and P.-Th. Wilrich, Heidelberg, Germany: Springer – Verlag, pp. 208 – 228

Stoumbos, Z.G., Reynolds, M.R., Jr., Ryan, T.P. and Woodall, W.H. (2000) "The State of Statistical Process Control in the 21st Century" *Journal of the American Statistical Association*, 95, pp. 992 – 998

Sullivan, J. H., and Woodall, W. H. (1996) "A Comparison of Multivariate Quality Control Charts for Individual Observations" *Journal of Quality Technology* 28, pp. 398 – 408

Thode, Jr., H. C. (2002), *Testing for Normality*, number 164 *in* 'Statistics: Textbooks and Monographs', Marcel Dekker, New York

Wade, M.R, and Woodall, W.H (1993) "A Review and Analysis of Cause-Selecting Control Charts" *Journal of Quality Technology* 25, pp. 161 – 169

Wierda, S.J. (1994) "Multivariate Statistical Process Control – Recent Results and Directions for Future Research" *Statistica Neerlandica* 48, pp. 147 – 168

Woodall, W. H. (1997) "Control Charting Based on Attribute Data: Bibliography and Review" *Journal of Quality Technology* 29, pp. 172 – 183

Woodall, W.H. (2000) "Controversies and Contradictions in Statistical Process Control" *Journal of Quality Technology* 32, pp. 341 – 350

Woodall, W.H., Montgomery, D.C. (1999) "Research Issues and Ideas in Statistical Process Control" *Journal of Quality Technology* 31, pp. 376 – 387

Woodall, W.H., Thomas, E. (1995) "Statistical Process Control with Several Components of Common Cause Variability" *IIE Transactions* 27, pp. 757 – 764

Wu, Z., Spedding, T.A. (2000a) "A Synthetic Control Chart for Detecting Small Shifts in the Process Mean" *Journal of Quality Technology* 32, pp. 32 – 38

Wu, Z., Spedding, T.A. (2000b) "Implementing Synthetic Control Charts" *Journal of Quality Technology* 32, pp. 75 – 78

Wu, Z., Yeo, S.H. (2001) "Implementing Synthetic Control Charts for Attributes" *Journal of Quality Technology* 33, pp. 112 – 114

Wu, Z., Yeo, S.H., Spedding, T.A. (2001) "A Synthetic Control Chart for Detecting Fraction Nonconforming Increases" *Journal of Quality Technology* 33, pp. 104 – 111

Yang, K., Trewn. J. (2004) *Multivariate Statistical Methods in Quality Management*, New York: McGraw Hill

Yourstone, S. A., Zimmer, W. J. (1992) "Non–normality and the Design of Control Charts for Averages" *Decision Science*s 23, pp. 1099 – 1113

Zarate, P. (2003) "Design of Nonparametric Control Chart for Monitoring Multivariate Processes Using Principal Components Analysis and Data Depth" PhD DISSERTATION, Department of Industrial And Management Systems Engineering, University of South Florida, Tampa, Fl.

Zhang, C., Lalor, G. (2003) "Multivariate Relationships and Spatial Distribution of Geochemical Features of Soils in Jamaica" *Chemical Speciation and Bioavailability* 14, pp. 57 – 65