

**ON ENHANCED PROCESS MONITORING USING
MEMORY SCHEMES FOR HIGH-DIMENSIONAL DATA**

BY

OSAMA ALHADI

A Thesis Presented to the
COLLEGE OF GRADUATE AND INTERDISCIPLINARY STUDIES
KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

In

APPLIED STATISTICS

May 2025

This thesis, written by **OSAMA ALHADI** under the direction of his thesis advisor and approved by his thesis committee, has been presented and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree **MASTER OF SCIENCE IN APPLIED STATISTICS**.

Thesis Committee



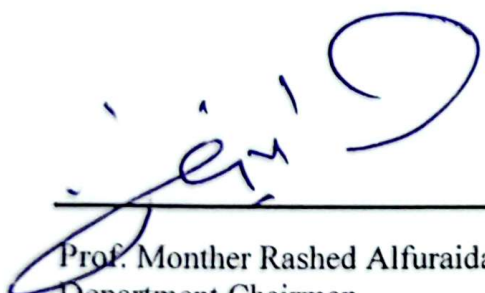
Prof. Muhammad Riaz
(Advisor)



Dr. Emmanuel Osita Afuecheta
(Member)



Dr. Ridwan A. Sanusi
(Member)



Prof. Monther Rashed Alfuraidean
Department Chairman



Prof. Suliman S. Al-Homidan
Dean of Graduate Studies

Date



© Osama Alhadi 2025

To my parents

ACKNOWLEDGMENTS

First and foremost, I offer my deepest gratitude to Almighty Allah for his countless blessings and guidance throughout my journey. I am sincerely thankful to my family for their unwavering encouragement and support, which have been instrumental in my academic and personal growth. I am profoundly grateful to King Fahd University of Petroleum and Minerals (KFUPM) for awarding me the prestigious scholarship, which played a vital role in enabling my studies. My sincere appreciation extends to the Department of Mathematics at KFUPM for their outstanding academic and administrative support. I would also like to express my heartfelt thanks to our department chairman, professor Monther Rashed Alfuraidean, for his continuous encouragement and invaluable support. I am also deeply thankful to my adviser, Prof. Muhammad Riaz, for his invaluable guidance since my first semester at KFUPM; he knows how to bring out the best in a student. I am proud to work under his supervision and appreciate his unlimited support and guidance. I am deeply grateful to Dr. Nasir Abbas for his support and for his assistance in ensuring the high-dimensional data simulation was performed smoothly. He is perhaps the best teacher I have ever had, and I have learned so much from such a distinguished individual. I would also like to sincerely thank Dr. Tahir Mahmood for his support and insightful advice throughout this work. Additionally, I thank Dr. Emmanuel Afuecheta for teaching sampling methods and non-parametric techniques and for all his advice and support. I must thank Dr. Mohammad Hafidz Omar for teaching us scientific writing during my first term paper. I also thank Dr. Ridwan A. Sanusi for his time and support. I have been fortunate to receive support from the Sudanese student community at KFUPM. Finally, I thank everyone who has helped and supported me.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	V
TABLE OF CONTENTS	VI
LIST OF TABLES	IXX
LIST OF FIGURES	XI
LIST OF ABBREVIATIONS	XII
ABSTRACT.....	XIII
ملخص الرسالة.....	XIV
CHAPTER 1 INTRODUCTION.....	1
1.1. Quality and Statistical Process Control	1
1.2. Basic Concepts of Control Charts.....	2
1.2.1. Classification of Control Charts Based on Their Structure	3
1.3. The High-Dimensional Data.....	3
1.3.1. Available High-Dimensional Control Charts and Their Issues	4
1.4. Problem Statement.....	5
1.5. Objectives of the Study	6
1.6. The Thesis Framework	6
CHAPTER 2 CUSUM CONTROL CHARTS FOR TWO-SAMPLED HIGH-DIMENSIONAL DATA	7
2.1. Introduction.....	7
2.2. High-Dimensional Charting Methodologies	12
2.2.1. The Structure of Two-Sample Test for High-Dimensional Data.....	12

2.2.2. Existing Shewhart Control Charts for Two-Sample High-Dimensional Data	13
2.2.3. Proposed CUSUM Charts for Two-Sample High-Dimensional Data	14
2.3. Design of The Simulation Study	16
2.4. Performance and Comparative Analysis	20
2.5. Real-Life Applications	25
 CHAPTER 3 CSEWMA CONTROL CHARTS FOR THE MEANS OF TWO SAMPLED HIGH-DIMENSIONAL DATA.....	32
3.1. Introduction.....	32
3.2. High-Dimensional Charting Methodologies	37
3.2.1. The Structure of Two-Sample High-Dimensional Data	37
3.2.2. Conventional Shewhart Charts for Two-Sample High-Dimensional Data	38
3.2.3. Conventional EWMA Charts for Two-Sample High-Dimensional Data	39
3.2.4. Proposed CSEWMA Charts for Two Sample High-Dimensional Data.....	41
3.3. Design of the Simulation Study.....	44
3.4. Performance and Comparative Analysis	50
3.5. Real-Life Application.....	63
 CHAPTER 4 CSCUSUM CONTROL CHARTS FOR THE MEANS OF TWO-SAMPLED HIGH-DIMENSIONAL DATA.....	70
4.1. Introduction.....	70
4.2. Development of High-Dimensional Charting Techniques.....	75
4.2.1. Design Structure of High-Dimensional Data Two Sample Testing Procedure	75
4.2.2. Classical Shewhart Setup Charts for Two-Sample High-Dimensional Data	76
4.2.3. Classical CUSUM Setup Charts for Two-Sample High-Dimensional Data	77

4.2.4. Proposed CUSUM Setup Charts for Two-Sample High-Dimensional Data	79
4.3. Simulation Study Framework.....	83
4.4. Performance Evaluation and Comparative Analysis.....	88
4.5. Illustrative Application Example.....	98
CHAPTER 5 CONCLUSION, REMARKS, AND RECOMMENDATIONS	106
5.1. Introduction.....	106
5.1.1. Concluding Remarks Based on CUSUM Charts for Two-Sample High-Dimensional Data	106
5.1.2. Concluding Remarks Based on CSEWMA Charts for Two-Sample High-Dimensional Data	106
5.1.3. Concluding Remarks Based on CSCUSUM Charts for Two-Sample High-Dimensional Data	109
REFERENCES.....	111
VITAE	119

LIST OF TABLES

Table 1	The mean, standard deviation, and coefficient values of the proposed charts at fixed $ARL_0 = 200$	19
Table 2	RL profile of proposed and competing charts for equal sample sizes $m = n = p \times 30\%$	20
Table 3	RL profile of proposed and competing charts for unequal sample sizes $m = p \times 30\%, n = p \times 45\%$	21
Table 4	The EQL values of proposed and competing charts	22
Table 5	Statistics and control limit coefficients at $ARL_0 = 200$ for proposed charts used in real application 1.....	26
Table 6	Statistics and control limit coefficients at $ARL_0 = 200$ for proposed charts used in real application 2	30
Table 7	The mean, standard deviation, and the coefficient values of the proposed control charts at $ARL_0 = 200$	46
Table 8	ARL(SDRL) and 25th, 50th, and 75th percentiles of the proposed DRCSEWMA and competing charts with $m = 30\%; n = 30\%$ sample sizes.....	51
Table 9	ARL(SDRL) and 25th, 50th, and 75th percentiles of the proposed DRCSEWMA and competing charts with $m = 30\%; n = 45\%$ sample sizes.....	52
Table 10	ARL(SDRL) and 25th, 50th, and 75th percentiles of the proposed BSCSEWMA and competing charts with $m = 30\%; n = 30\%$ sample sizes.....	53
Table 11	ARL(SDRL) and 25th, 50th, and 75th percentiles of the proposed BSCSEWMA and competing charts with $m = 30\%; n = 45\%$ sample sizes.....	54
Table 12	ARL(SDRL) and 25th, 50th, and 75th percentiles of the proposed SDCSEWMA and competing charts with $m = 30\%; n = 30\%$ sample sizes.....	56
Table 13	ARL(SDRL) and 25th, 50th, and 75th percentiles of the proposed SDCSEWMA and competing charts with $m = 30\%; n = 45\%$ sample sizes.....	57
Table 14	The mean, standard deviation, and coefficient limits of the proposed and competing charts for Gas Sensor Array Monitoring in Flow Modulation Systems at a fixed $ARL_0 = 200$	65
Table 15	Mean, standard deviation, and the coefficients for the proposed control charts at fixed $ARL_0 = 200$	85
Table 16	RL profile for the proposed DRCSUSUM and competing charts with ($m = 30\%; n = 30\%$) sample sizes	88
Table 17	RL profile for the proposed DRCSUSUM and competing charts with ($m = 30\%; n = 45\%$) sample sizes	90
Table 18	RL profile for the proposed BSCSUSUM and competing charts with ($m = 30\%; n = 30\%$) sample sizes	91

Table 19 RL profile for the proposed BSCSCUSUM and competing charts with ($m = 30\%$; $n = 45\%$) sample sizes.	92
Table 20 RL profile for the proposed SDCSCUSUM and competing charts with ($m = 30\%$; $n = 30\%$) sample sizes.	93
Table 21 RL profile for the proposed SDCSCUSUM and competing charts with ($m = 30\%$; $n = 45\%$) sample sizes.	95
Table 22 The mean, standard deviation, and coefficient limits of the proposed and competing charts for Gas Sensor Array Monitoring in Flow Modulation Systems at a fixed $ARL_0 = 200$	101

LIST OF FIGURES

Figure 1 Showing the number of dimensions (genes) for a given individual where $P > n$	4
Figure 2 The schematic diagram for the implementation of proposed control charts.....	18
Figure 3 Effect of the K values on the performance of the DRCUSUM chart.....	25
Figure 4 DRCUSUM and SDCUSUM charts for real application 1	28
Figure 5 DRCUSUM and SDCUSUM charts for real application 2	31
Figure 6 The schematic diagram for the implementation of proposed control charts.....	49
Figure 7 Effect of the λ values on the performance of the SDCSEWMA chart.....	62
Figure 8 Conventional DR, BS and SD charts for Gas Sensor Array Monitoring in Flow Modulation Systems	66
Figure 9 Conventional DREWMA, BSEWMA and SDEWMA charts for Gas Sensor Array Monitoring in Flow Modulation Systems.....	67
Figure 10 Proposed DRCSEWMA, BSCSEWMA and SDCSEWMA charts for Gas Sensor Array Monitoring in Flow Modulation Systems	68
Figure 11 Framework diagram for the implementation of the proposed control charts.	87
Figure 12 Conventional DR, BS and SD charts for Gas Sensor Array Monitoring in Flow Modulation Systems	102
Figure 13 Conventional DCUSUM, BSCUSUM, and SDCUSUM charts for Gas Sensor Array Monitoring in Flow Modulation Systems	103
Figure 14 Proposed DRCSCUSUM, BSCSCUSUM and SDCSCUSUM charts for Gas Sensor Array Monitoring in Flow Modulation Systems.....	104

LIST OF ABBREVIATIONS

SPC	:	Statistical Process Control
IC	:	In Control
OOC	:	Out Of Control
CL	:	Center Line
UCL	:	Upper Control Limit
LCL	:	Lower Control Limit
RL	:	Run Length
ARL	:	Average Run Length
SDRL	:	Standard Deviation of Run Length
EQL	:	Extra Quadratic Loss
CUSUM	:	Cumulative Sum
EWMA	:	Exponentially Weighted Moving Average
PCA	:	Principal Component Analysis
KPCA	:	Kernel Principal Component Analysis
LASSO	:	Least Absolute Shrinkage and Selection Operator
LAR	:	Least Angle Regression
CSEWMA	:	Combined Shewhart EWMA
CSCUSUM	:	Combined Shewhart CUSUM
KNN	:	K-nearest neighbours
SVDD	:	Support Vector Data Description
LMNN	:	Latent Mapping Embedding Neural Network
MVV	:	Minimum Vector Variance
GMM	:	Gaussian Mixture Model
MS-MSPS	:	Multistate Multivariate Statistical Process Control

ABSTRACT

Full Name : OSAMA ALHADI

Thesis Title : ON ENHANCED PROCESS MONITORING USING MEMORY
SCHEMES FOR HIGH-DIMENSIONAL DATA

Major Field : APPLIED STATISTICS

Date of Degree : APRIL, 2025

Monitoring process involving high-dimensional data where the number of the features exceeds the number of observations poses a significant challenge in the domain of Statistical Process Control (SPC). In high-dimensional data environments, traditional control charts face challenges such as the "curse of dimensionality," which may adversely affect process monitoring accuracy and raise false alarm rates. Hence, improved SPC methodologies are essential to effectively address the curse of high dimensionality. The Memoryless and memory-type schemes discussed in the preceding literature review mostly focus on monitoring single-sample high-dimensional data; nevertheless, there is an absence of research concerning two-sample high-dimensional data. One paper introduced Shewhart setup charts for two-sample high-dimensional data, but it is insensitive for the small to moderate shifts. Hence, in this thesis, we propose memory-setup, combined memory and memoryless setup charts based on the Dempster (DR), Bai and Saranadasa (BS), and Srivastava-Du (SD) statistics, respectively without using dimensional reduction or variable selection techniques. These charts are optimized to detect small to moderate process shifts. The performance of the proposed charts is systematically benchmarked against each other and conventional Shewhart setup charts across high-dimensional environments, exemplified by multivariate normal, t , and gamma data structures.

ملخص الدراسة

الاسم كامل: أسامة الهادي إسماعيل هرون

عنوان البحث: تحسين مراقبة العمليات باستخدام المخططات ذات الذاكرة للبيانات عالية الأبعاد

التخصص: الإحصاء التطبيقي

تاريخ الدرجة العلمية: أبريل 2025

تعد مراقبة العمليات ذات البيانات عالية الأبعاد التي تعرف بالبيانات التي فيها عدد الأبعاد أكثر من عدد المشاهدات أمراً مهماً في مجال ضبط الجودة الإحصائية. في بيئات البيانات عالية الأبعاد تواجه مخططات ضبط الجودة الإحصائية التقليدية عدد من المشكلات والتحديات مثل مشكلة الارتباطات المتعددة ومشكلة عدم المقدرة على إيجاد مصفوفة التباينات والتغايرات فيما يعرف "بمشكلة الأبعاد" هذه المشكلات قد تؤدي بدورها إلى تقليل جودة أداء مخططات ضبط الجودة الإحصائية والتي بدورها تؤدي إلى زيادة معدل الإنذارات غير الحقيقية. لذا يعد تحسين مخططات ضبط الجودة الإحصائية أمراً ضرورياً لمعالجة مشكلة البيانات ذات الأبعاد العالية بفاعلية. تركز مخططات ضبط الجودة الإحصائية عديمة الذاكرة ومخططات ضبط الجودة الإحصائية ذات الذاكرة التي نوقشت في مراجعة الأدبيات السابقة بشكل رئيسي على مراقبة عينة واحدة ذات أبعاد عالية الأمر الذي جعلنا نلاحظ عدم وجود أبحاث سابقة لمراقبة ضبط الجودة الإحصائية لعينتين من البيانات ذات الأبعاد العالية باستثناء دراسة بحثية واحدة تمت في العام 2024 قدمت مخططات من النوع عديم الذاكرة لمراقبة ضبط الجودة الإحصائية للبيانات عالية الأبعاد لكن لوحظ أن هذا النوع من المخططات عديم الذاكرة غير حساس للتغيرات الصغيرة والمتوسطة التي تحدث في العمليات الإحصائية. لذلك في هذا البحث قمنا باقتراح مخططات ضبط جودة إحصائية مركبة من المخططات ذات الذاكرة والمخططات عديمة الذاكرة بالاعتماد على إحصاءات مخصصة للبيانات عالية الأبعاد دون استخدام تقنيات أو طرق الاختزال البعدي أو اختيار المتغيرات. هذه المخططات المحسنة لمراقبة ضبط الجودة في العمليات ذات البيانات عالية الأبعاد المخصصة لتتبع التغيرات الصغيرة والمتوسطة تم تقييمها ومقارنة أداءها بشكل منهجي مع بعضها البعض ومع المخططات عديمة الذاكرة التقليدية بالاعتماد على تصورات مختلفة وتوزيعات إحصائية مختلفة وباستخدام طرق مختلفة لقياس جودة أداء مخططات ضبط الجودة الإحصائية.

CHAPTER 1

Introduction

This chapter elucidates the concept of quality and the function of Statistical Process Control (SPC) in its assurance. It elucidates how SPC facilitates the monitoring and preserving process quality through statistical techniques that identify deviations and ensure processes stay stable and predictable. It comprehensively elucidates various control charts, encompassing both memoryless and memory-type charts. And high-dimensional data with a summary of available high-dimensional control charts. Furthermore, the chapter gives a problem statement and thesis objectives. And outlines the thesis's organization, summarizing the topics addressed in each section.

1.1. Quality and Statistical Process Control

Quality is an essential element in the current manufacturing and service sectors, verifying that products and processes fulfill customer expectations, regulatory requirements, and industry standards. SPC comprises a robust array of instruments and techniques employed to monitor, regulate, and enhance processes via statistical analysis [1]. Quality has consistently been a fundamental element of all products and services. Statisticians and production managers possess divergent viewpoints regarding quality enhancement. Production managers equate the production of homogeneous commodities with quality, but statisticians acknowledge that variability is inevitable. Consequently, from a statistical perspective, enhancing quality necessitates minimizing variability in the process. SPC is an approach that distinguishes between common reasons and unusual cause variations within a process. Common cause variation requires systemic

changes, while special cause variation requires the identification and elimination of external factors. SPC utilizes control charts to observe process behavior and detect changes, thereby directing practical enhancement initiatives. [2]. Comprehending the difference between common cause and special cause variation is essential for efficient process control. Control charts are vital instruments in this distinction, enabling firms to maintain process stability and quality. A process is deemed in-control (IC) when it displays solely common cause variation. Conversely, the existence of a particular cause variation signifies instability, rendering the process out-of-control (OOC) and requiring prompt corrections. SPC includes numerous approaches that are essential for improving quality by detecting and Reducing process variability and involves essential tools such as control charts, histograms, and Pareto charts. provide efficient monitoring and management of processes across various industry sectors [1].

1.2. Basic Concepts of Control Charts

Control charts are the key instrument in SPC, offering a graphical representation of process performance throughout time [1, 3]. A control chart typically includes a center line (CL), an upper control limit (UCL), and a lower control limit (LCL). The central line represents the average or expected value of the process, while the control limits provide the parameters of permissible variance. The data points displayed in the figure signify measurements obtained from the process at consistent intervals. By analyzing the patterns and trends of these data points concerning the control limits, one can evaluate process stability and identify the existence of specific reasons [4]. Since its introduction, control schemes have been extensively utilized for process checking across the manufacturing, healthcare, business, and finance sectors. This extensive applicability underscores the necessity to devise novel methodologies and improve current frameworks to meet the distinct difficulties and demands of each sector. The efficacy of control charts is typically

assessed via the average run length (ARL), defined as the average number of samples needed before an OOC condition occurs. Control charts can be categorized based on the structure of control schemes into two memoryless and memory-type control charts. This aids in selecting the most suitable chart for diverse process monitoring requirements and particular applications.

1.2.1. Classification of Control Charts Based on Their Structure

Control charts can also be categorized into two based on their structure according to how historical information is handled. There are memoryless and memory-based charts that are vital tools in SPC, each fulfilling specific roles in the assessment of process performance. Memoryless charts, shown by traditional Shewhart charts, are simple to construct and interpret, making them suitable for various applications. They depend exclusively on the most recent data points. Consequently, they are proficient at identifying significant, unexpected changes in the process location or dispersion parameters. For example [5, 6]. Conversely, memory-setup schemes, such as Cumulative Sum (CUSUM) [7]. And Exponentially Weighted Moving Average (EWMA) charts [8], are designed to be more precision to small, gradual shifts in the process mean compared with Shewhart schemes. These charts incorporate historical data to detect subtle changes that might not be immediately apparent in a Shewhart chart.

1.3. The High-Dimensional Data

High-dimensional data is defined as the datasets that consist of a substantially larger number of dimensions p than observations N , providing distinct difficulties and opportunities, High-dimensional data is prevalent throughout nearly all sectors, including industrial data [9, 10], Medical Science data [11, 12], image Data [13, 14]. The complexity of high-dimensional data requires sophisticated analytical approaches to derive significant insights, as conventional statistical procedures frequently falter because of the “curse of dimensionality”.

	Gene feature 1	Gene feature 2	Gene feature 3	Gene feature 4	...	...
Individual 1						
Individual 2						
Individual 3						

Figure 1 Showing the number of dimensions (genes) for a given individual where $P > n$

1.3.1. Available High-Dimensional Control Charts and Their Issues

Control charts are visual instruments utilized to observe the efficiency of processes over time. They help detect deviations from the normal process, enabling timely corrective actions. However, in high-dimensional data environments, traditional control charts face challenges such as the "curse of dimensionality," which can lead to reduced performance and increased false alarms. High-dimensional data analysis is generally classified into two fundamental categories, high-dimensional approaches and analysis using dimensionality reduction techniques. Dimension reduction techniques are categorized into two main groups dimension reduction approaches and variable selection techniques.

Dimension reduction methods minimize variables while retaining critical information for monitoring process performance, thereby enhancing control chart efficiency by eliminating redundant or less useful features in high-dimensional data. There are different types of dimension reduction approaches, for instance, Principal Component Analysis (PCA) [15, 16], Kernel Principal Component Analysis (KPCA) [17].

Variable selection techniques, such as least angle regression (LAR), are employed for selecting the most informative variables within high-dimensional datasets [18]. These strategies improve the efficacy of control charts by concentrating on pertinent variables [19]. There are different types

of variable selection methods, for example, the Least Absolute Shrinkage and Selection Operator (LASSO) based procedure [20], forward-selection technique [21].

1.4. Problem Statement

Utilizing conventional control schemes in high-dimensional data settings presents substantial methodological difficulties because high dimensionality introduces complexities such as multicollinearity which hinder parameter estimation and the identification of runaway conditions. Although dimensionality reduction and variable selection techniques have been introduced to address these problems, they often result in a loss of information and are not always effective [22]. Monitoring shifts between two samples of high-dimensional data is largely neglected in the current literature, which primarily concentrates on the creation of control schemes for single-sample high-dimensional data. Mahmood et al. [23]. attempted to address this gap by developing control charts (DR, BS, SD) for detecting mean variations in two-sample high-dimensional data based on three tests Srivastava and Du [24], Bai and Saranadasa [25], and Dempster [26, 27]. To monitor the means between two high-dimensional data sets without exclusively depending on dimensionality reduction techniques, more control charts must be developed. The need for robust control charts capable of handling high-dimensional, dual-sample data is critical to ensure effective process monitoring and maintaining quality in various industries where high-dimensional data is prevalent. This thesis aims to address this gap by developing more comprehensive control charts that can monitor high-dimensional two-sample processes, thereby improving sensitivity and performance in shift detection.

1.5. Objectives of the study

To deal with high-dimensional processes, the existing monitoring schemes are memoryless and don't count the history of the data. A significant amount of information is present in previous data points that may be capitalized for effective tracking of the parameters of the process. The same is the target of this thesis. We intend to develop memory structures for high-dimensional data using CUSUM and EWMA schemes. The following are the specific objectives that will be covered in this thesis.

- Advancing various memory structures using CUSUM and EWMA and extending them to the combined and mixed schemes for two sample high-dimensional data
- Examining the run length characteristics of the suggested methods over several distributional frameworks, including multivariate normal, t, and gamma distributions.
- Implementing the proposed schemes for real datasets from different disciplines such as business and engineering.

1.6. The Thesis Framework

The structure of this thesis comprises five main chapters. Chapter 1 discusses the SPC tools and their significance in maintaining quality throughout the production process, while also offering a brief overview of the high-dimensional concept of SPC. Chapter 2 introduces CUSUM control charts for the means of two sampled high-dimensional data. Chapter 3 introduces the combined Shewhart-EWMA (CSEWMA) control schemes for the means of two sampled high-dimensional data. Chapter 4 presents the combined Shewhart-CUSUM (CSCUSUM) control schemes for the means of two sampled high-dimensional data. Chapter 5 provides a summary, concluding observations, future work, and research opportunities.

CHAPTER 2

CUSUM CONTROL CHARTS FOR THE TWO SAMPLED HIGH-DIMENSIONAL DATA

2.1. Introduction

The term high-dimensional data typically denotes scenarios where the number of the variables significantly exceeds the number of observations. Due to the revolution in sensor technology, high-dimensional data recorded through multiple sensors can be found in almost all sectors, including industry [28, 29], medical sciences [30-33], networks systems [34, 35], financial areas [36, 37], science and engineering [38, 39]. Usually, high-dimensional data is recorded for a single sample population. For example, the semiconductor manufacturing process dataset [40], the Aluminium smelter company dataset [41], Solar flare detection dataset [42]. However, some high-dimensional processes produce high-dimensional data for a two-sample population. For example, the benchmark data sets [43], and the grease damage dataset [23].

Control charts serve as essential instruments in SPC, facilitating real-time surveillance of high-dimensional processes. Control charts are generally categorized into memory-type and memoryless-type schemes due to their structure and efficiency in monitoring the strength of shifts. The importance of using memory control charts over memoryless charts to monitor the shifts in high-dimensional processes is significant in ensuring that historical information is incorporated by enabling the proper monitoring of data behaviour over time [44]. The conventional memoryless

chart under the Shewhart setup is suitable for lower dimensions and higher shifts [23]. When the dimension increases, and the target is to detect anomalies due to small to moderate shifts, the Cumulative Sum (CUSUM), which is a type of memory chart, is a more suitable tool [45].

In SPC literature, the proposal of the CUSUM schemes for high-dimensional single-sample data is very common. This literature is divided into four different aspects [44] such as (i) CUSUM schemes incorporating dimension reduction procedures, (ii) CUSUM schemes incorporating variable selection procedures, (iii) CUSUM schemes incorporating high-dimensional approaches, and (iv) CUSUM schemes incorporating advanced projection-based and machine learning based techniques.

The dimension reduction method involves decreasing the number of dimensions in a dataset while preserving as much pertinent information as possible; subsequent to the application of dimension reduction techniques, the CUSUM chart was utilized to identify changes in high-dimensional processes. For instance, reduced-dimension control charts using spatial scanning are proposed by Chen, et al. [46] to monitor the locations of high-dimensional data during the data-gathering stage. Chen, et al. [47] offered a unique multivariate CUSUM scheme that efficiently handles high-dimensional data by using principal components for dimension reduction. This approach posits that the weighted sum of the primary directions derived from the population covariance matrix signifies the expected mean shift. Especially for large-dimensional datasets, the suggested chart illustrates the multivariate charts that are currently in use, including Crosier's benchmark. Li, et al. [48] developed a nonparametric monitoring approach for multivariate data utilizing the K-nearest neighbours (KNN) algorithm. This method regards KNN as a dimensionality reduction tool, enabling the monitoring of high-dimensional data through the integration of machine learning and conventional control charts. Wang, et al. [49] developed a nonparametric approach for the

surveillance of high-dimensional and high-frequency data streams. Their technique utilizes a CUSUM process on converted univariate data, adeptly addressing the complexities associated with high-dimensional environments.

CUSUM chart has also been frequently applied after the application of variable selection techniques that can decrease the size of data with high dimensionality. Variable selection approaches seek to identify a subset of significant variables from a high-dimensional dataset including numerous variables. These methods are utilized to select relevant variables effectively, and then the CUSUM scheme is utilized to oversee high-dimensional processes. For instance, variable Selection Techniques like Least Absolute Shrinkage and Selection Operator LASSO [20], and Principal Component Analysis PCA [50-52] are commonly integrated with CUSUM charts to address dimensionality challenges. Moreover, Zhang, et al. [53] enhanced the identification of sparse mean shifts in high-dimensional processes, and a sensitive variable selection control chart integrating classification algorithms was presented. Variable selection techniques are applied to multivariate CUSUM schemes to improve the sensitivity of OOC setting [54]; the chart was proved superior to the usual MCUSUM and Hotelling T^2 charts. Ullah, et al. [55] suggested a technique that uses a leave-one-out resampling approach and regularizes the covariance matrix to enhance the identification of anomalies in high-dimensional setup. This method outperforms the traditional Hotelling T^2 method in terms of power while preserving false positive rates. Ahmad and Ahmed [56] investigated how T^2 statistics are distributed in high-dimensional data and how this affects statistical process monitoring. Because of overfitting and noise accumulation, they demonstrated that conventional multivariate techniques are insufficient in high-dimensional environments. Through the integration of variable selection techniques, the work enhanced the efficacy of T^2 statistics for identifying changes in high-dimensional process parameters. Li, et al. [57] suggested

an adaptive penalized weighted mean regression technique that uses the adaptive Huber loss function to improve outlier detection in high-dimensional data, shows resilience to outliers in response variables and covariates and guarantees efficient variable selection.

High-dimensional control charts employ advanced approaches such as regularization, robust covariance estimation, and nonparametric techniques to directly monitor process fluctuations for data with high-dimensionality, without any dimensionality reduction techniques. These methodologies address issues like computational complexity and multicollinearity while preserving the integrity of the complete data structure [44]. Kim, et al. [58] introduced a Bayesian methodology for the high-speed monitoring of multidimensional processes. Their strategy utilizes Bayesian updates to attain real-time adaptation to changes in the process. This approach is very efficacious in high-dimensional setups since it incorporates data fusion techniques, hence guaranteeing computing efficiency. Ye, et al. [59] established a monitoring framework for high-dimensional, asynchronous, and heterogeneous data streams. Their approach directly models high-dimensional data streams using statistical frameworks capable of accommodating both location and scale shifts rather than employing dimension reduction techniques. The approach facilitates the examination of intricate process interactions in their authentic high-dimensional state. Moala, et al. [60] evaluated MCUSUM and MEWMA control charts for high-dimensional data to track the monthly incidence of vehicle thefts in São Paulo. They employed multivariate techniques to monitor regionally scattered data over time and highlighted the significance of high-dimensional technologies in simultaneously observing numerous regions.

Monitoring high-dimensional processes poses considerable problems because of the complexity and magnitude of the data involved. Conventional control charts frequently prove insufficient for efficiently analyzing such data and necessitate the use of advanced machine-learning techniques

[61]. Yu, et al. [62] suggested a latent mapping embedding neural network (LMNN) technique that effectively monitors nonlinear dynamic processes by extracting features from high-dimensional process data using a deep neural network and mapping those features to a low-dimensional subspace. Shah, et al. [63] introduced a new multivariate shrinkage-based diagonal T^2 control chart specifically designed for high-dimensional data. This method utilizes a shrinkage estimator to create a diagonal covariance matrix that is invertible and well-conditioned, thus improving the efficiency of the sample covariance matrix estimation. Support Vector Data Description (SVDD) is a machine learning method employed for single-class classification and anomaly detection. In high-dimensional process monitoring, SVDD-based control charts, like the K-chart, have been developed to oversee multivariate processes without considering normality in the underlying data distribution [64]. Bodnar and Bodnar [65] also discussed MCUSUM control charts based on Projection Pursuit in the monitoring procedure setup. Furthermore, for non-normal distributions and sparse OOC scenarios, random projection-based monitoring based on spatial rank has been proposed by Zhang, et al. [66].

The above-stated CUSUM-based control charts predominantly focus on monitoring single-sample high-dimensional data; however, a few studies are available for two-sample cases. Recently, Mahmood, et al. [23] proposed Shewhart-type control charts named DR, BS, and SD charts based on the Dempster (DR) [27], Bai and Saranadasa (BS) [67] and Srivastava and Du (SD) [68] tests for surveillance the means of two-sampled high-dimensional data. Their findings depicted that these high-dimensional charts not only outperformed the conventional Hotelling's T^2 chart under lower dimensions ($p \leq 50$), but they have shown good detection ability under high dimensions ($p \geq 50$). Notably, the suggested schemes were based on the Shewhart structure, which only considered the current information to monitor; Consequently, they are less proficient at identifying

small to moderate variations in means of two-sample high-dimensional data. Hence, to surveillance the small to moderate shifts effectively in two-sampled high-dimensional data, this study intends to propose CUSUM schemes. The presented CUSUM schemes are assessed based on the run length metrics under different distributional setups, dimensions, and sample sizes. Further, the demonstration of these charts is also shown using real-life data related to the finance and engineering sectors.

The subsequent sections of this chapter are as follows. Section 2.2 discusses the mathematical details of two sample high-dimensional tests and suggested charts. Section 2.3 outlined the design requirements for the simulation study. Section 2.4 highlights the performance and comparison of the suggested charts, while Section 2.5 offers illustrated examples. The final remarks are presented in Section 2.6.

2.2. High-Dimensional Charting Methodologies

This section describes the control charting methodologies used for high-dimensional data under two-sample setups. This section will also cover both existing Shewhart and proposed CUSUM control charts.

2.2.1. The Structure of Two Sample Test for High-Dimensional Data

For the two-sample high-dimensional case, x_{ijlt} , $i = 1, 2, \dots, m$ or $i = 1, 2, \dots, n$, $j = 1, 2, \dots, p$, $l = 1, 2$ and $t = 1, 2, 3, \dots$, is the p -dimensional t^{th} multivariate normal random vectors with mean vectors $\mu_{0l} = (\mu_1, \dots, \mu_p)'$; $k = 1, 2$ and unknown common positive variance-covariance matrices Σ_{0l} ; $l = 1, 2$. The sizes of the first and second samples are denoted by m and n , respectively; p is the size of the dimension, l is the sample number, and t represents time. The t^{th} statistics T_{1t} , T_{2t} and T_{3t} suggested by Dempster [27], Bai and Saranadasa [67] and Srivastava

and Du [68], individually, for the means testing of two-sampled high-dimensional data are defined as follows:

$$T_{1t} = \frac{\frac{mn}{m+n}(\bar{x}_{1t} - \bar{x}_{2t})'(\bar{x}_{1t} - \bar{x}_{2t})}{tr S_t}, \quad (2.1)$$

$$T_{2t} = \frac{\frac{mn}{m+n}(\bar{x}_{1t} - \bar{x}_{2t})'(\bar{x}_{1t} - \bar{x}_{2t}) - tr S_t}{\left[\frac{2(m+n)(m+n+1)}{(m+n-1)(m+n+2)} \left(tr S_t^2 - \frac{(tr S_t)^2}{m+n} \right) \right]^{\frac{1}{2}}}, \quad (2.2)$$

$$T_{3t} = \frac{\frac{mn}{m+n}(\bar{x}_{1t} - \bar{x}_{2t})' D_{S_t}^{-1} (\bar{x}_{1t} - \bar{x}_{2t}) - \frac{(m+n)p}{m+n-2}}{\left[2 \left(tr R_t^2 - \frac{p^2}{m+n} \right) c_{p,m+n} \right]^{\frac{1}{2}}}. \quad (2.3)$$

Where the t^{th} sample means the vector of the first sample is $\bar{x}_{1t} = \sum_{i=1}^m x_{ij1t}/m$ and t^{th} sample means the vector of the second sample is $\bar{x}_{2t} = \sum_{i=1}^n x_{ij1t}/n$. The t^{th} diagonal matrix (D_{S_t}) of sample variance-covariance matrix (S) is defined by $D_{S_t} = \text{diag} (s_{11}, \dots, s_{pp})$, the t^{th} sample correlation matrix (R_t) is defined by $R_t = D_{S_t}^{-\frac{1}{2}} S_t D_{S_t}^{-\frac{1}{2}}$ and the adjustment coefficient $c_{p,m+n}$ can be obtained by $c_{p,m+n} = 1 + \frac{\text{tr } R_t^2}{p^{3/2}}$.

2.2.2. Existing Shewhart Control Charts for Two-Sample High-Dimensional Data

Mahmood, et al. [23] developed DR, BS, and SD schemes to surveillance means of the two-sample high-dimensional data. For the DR chart, the t^{th} statistics T_{1t} presented in equation (2.1) are considered as the plotting statistics against the probability limits, which are obtained as follows on a given false alarm rate α :

$$\begin{cases} P(T_{1t} \leq LCL) = \int_{t_1=0}^{LCL} P(T_{1t} = t_1) \leq \frac{\alpha}{2}, \\ P(T_{1t} \geq UCL) = \int_{t_1=UCL}^{\infty} P(T_{1t} = t_1) \leq \frac{\alpha}{2}. \end{cases} \quad (2.4)$$

For the BS chart, T_{2t} presented in equation (2.2) is considered as the plotting statistics against the probability limits as follows:

$$\begin{cases} P(T_{2t} \leq LCL) = \int_{t_2=0}^{LCL} P(T_{2t} = t_2) \leq \frac{\alpha}{2}, \\ P(T_{2t} \geq UCL) = \int_{t_2=UCL}^{\infty} P(T_{2t} = t_2) \leq \frac{\alpha}{2}. \end{cases} \quad (2.5)$$

For the SD chart, T_{3t} presented in equation (2.3) is considered as the plotting statistics against the probability limits as follows:

$$\begin{cases} P(T_3 \leq LCL) = \int_{t_3=0}^{LCL} P(T_3 = t_3) \leq \frac{\alpha}{2}, \\ P(T_3 \geq UCL) = \int_{t_3=UCL}^{\infty} P(T_3 = t_3) \leq \frac{\alpha}{2}. \end{cases} \quad (2.6)$$

Notably, the suggested schemes, titled DR, BS, and SD, gives an OOC signal when T_{1t} , T_{2t} and T_{3t} falls outside their corresponding control limits given in equations (2.4) - (2.6). Otherwise, the process is considered an IC situation.

2.2.3. Proposed CUSUM Charts for Two-Sample High-Dimensional Data

Shewhart charts effectively detect large shifts in the means of two-sample high-dimensional data [23]. When the objective is to monitor small to moderate shifts, Shewhart charts are ineffective in such a situation. Hence, the CUSUM chart proposed by Page [69] is an alternative solution. Therefore, based on T_{1t} , T_{2t} and T_{3t} Statistics given in equations (2.1) -(2.3), CUSUM charts were proposed and named by DRCUSUM, BSCUSUM, and SDCUSUM charts. The following presents the plotting statistics used for DRCUSUM scheme:

$$\begin{aligned} C_{t(DRCUSUM)}^- &= \min [0, (T_{1t} - \mu_{T_1}) + K \sigma_{T_1} + C_{t-1(DRCUSUM)}^-], \\ C_{t(DRCUSUM)}^+ &= \max [0, (T_{1t} - \mu_{T_1}) - K \sigma_{T_1} + C_{t-1(DRCUSUM)}^+], \end{aligned} \quad (2.7)$$

the following expressions define the control limits of the DRCUSUM scheme:

$$\begin{aligned} LCL_{(DRCUSUM)} &= -h_1 \sigma_{T_1}, \\ UCL_{(DRCUSUM)} &= h_2 \sigma_{T_1}. \end{aligned} \quad (2.8)$$

Notably, μ_{T_1} and σ_{T_1} are the mean and standard deviation of T_{1t} statistic, while h_1 and h_2 the control limit coefficient for the DRCUSUM scheme, and K is the reference value, which is usually chosen as half of the mean shift of process. The values of these parameters are reported in Table

1. The plotting statistics of the BSCUSUM chart are presented as follows:

$$\begin{aligned} C_{t(BSCUSUM)}^- &= \min [0, (T_{2t} - \mu_{T_2}) + K \sigma_{T_2} + C_{t-1(BSCUSUM)}^-], \\ C_{t(BSCUSUM)}^+ &= \max [0, (T_{2t} - \mu_{T_2}) - K \sigma_{T_2} + C_{t-1(BSCUSUM)}^+], \end{aligned} \quad (2.9)$$

where μ_{T_2} and σ_{T_2} are the mean and standard deviation of T_{2t} statistic, while based on control limit coefficients h_3 and h_4 , control limits can be obtained as follows:

$$\begin{aligned} LCL_{(BSCUSUM)} &= -h_3 \sigma_{T_2}, \\ UCL_{(BSCUSUM)} &= h_4 \sigma_{T_2}. \end{aligned} \quad (2.10)$$

The SDCUSUM plotting statistics based on the mean and standard deviation of T_{3t} statistic (μ_{T_3} and σ_{T_3}) (cf. Table 1) are explained as follows:

$$\begin{aligned} C_{t(SDCUSUM)}^- &= \min [0, (T_{3t} - \mu_{T_3}) + K \sigma_{T_3} + C_{t-1(SDCUSUM)}^-], \\ C_{t(SDCUSUM)}^+ &= \max [0, (T_{3t} - \mu_{T_3}) - K \sigma_{T_3} + C_{t-1(SDCUSUM)}^+], \end{aligned} \quad (2.11)$$

and control limits of the SDCUSUM chart based on control limit coefficients h_5 and h_6 (cf. Table 1) can be obtained as follows:

$$\begin{aligned} LCL_{(SDCUSUM)} &= -h_5 \sigma_{T_3}, \\ UCL_{(SDCUSUM)} &= h_6 \sigma_{T_3}. \end{aligned} \quad (2.12)$$

It is to be noted that the initial value of all CUSUM statistics provided in equations (2.7, 2.9, and 2.11) are set equal to zero, i.e. $C_{0(DRCUSUM)}^{\pm} = C_{0(BSCUSUM)}^{\pm} = C_{0(SDCUSUM)}^{\pm} = 0$. For decision rules, a particular chart, such as DRCUSUM, BSCUSUM, and SDCUSUM, is known to be in an IC state if both statistics (either equation (2.7) or (2.9) or (2.11)) fall within the limits (either equation (2.8) or (2.10) or (2.12)); otherwise, it is declared OOC. Further, it is to be mentioned here that for testing the means of two-sample high-dimensional data, Bai and Saranadasa [67] showed that the T_{1t} and T_{2t} statistics have similar performance, which is further highlighted in SPC literature by Mahmood, et al. [23]. Therefore, for brevity, we have excluded the results of the BSCUSUM chart.

2.3. Design of The Simulation Study

The simulation study is designed based on the Monte Carlo simulations using the R language program. The performance of control charts, including proposed and existing, is assessed using the Run Length (RL) profile, which measures the number of samples or time points until a signal is triggered. Key metrics derived from RL include: (a) Average Run Length (ARL); ARL represents the average number of points needed to detect an out-of-control (OOC) signal. It is calculated as: $\frac{1}{ARL} = \frac{1}{ARL^+} + \frac{1}{ARL^-}$, where ARL^+ and ARL^- are the ARLs for upper and lower one-sided CUSUM charts, respectively [70]. When the process is stable, ARL is denoted ARL_0 ; for an OOC process, it is ARL_1 . It is to be noted that a lower ARL_1 indicates faster detection of shifts. (b) Standard Deviation of Run Length (SDRL); SDRL measures variability in RL, providing insight into the consistency of chart performance. It is calculated by $SDRL = \sqrt{E(RL - ARL)^2}$ and the lower SDRL indicates more consistent detection. (c) percentile; The 25th, 50th (median), and 75th percentiles of RL highlight the distribution of run lengths, indicating where the shortest 25%,

middle 50%, and longest 25% of RL values lie. (d) Extra Quadratic Loss (EQL); EQL evaluates overall chart performance across all shift magnitudes using a weighted average of ARL over the entire shift domain ($\delta_{min} < \delta < \delta_{max}$) using the square of shift (δ^2) as the weight. The EQL is calculated by $EQL = \int_{\delta_{min}}^{\delta_{max}} \delta^2 \text{ARL}(\delta) d\delta / (\delta_{max} - \delta_{min})$ where lower EQL values indicate better overall chart performance.

By following Stoumbos and Sullivan [71] and Mahmood, et al. [23], two independent high-dimensional samples (represented by $l = 1, 2$) are generated through four distributional environments: (i) multivariate normal with mean vector ($\mu_{0l} = 0; l = 1, 2$) and variance-covariance matrix ($\Sigma_{0l}; l = 1, 2$), where unit variances and covariances are equal to 0, denoted as MVN_0 ; (ii) multivariate normal with mean vector ($\mu_{0l} = 0; l = 1, 2$) and variance-covariance matrix ($\Sigma_{0l}; l = 1, 2$), where unit variances and covariances are equal to 0.5, denoted as MVN_1 ; (iii) multivariate t with 5 degrees of freedom, denoted as MVT; (iv) multivariate gamma with shape parameter 5 and scale parameter 1, denoted as MVG. For MVT and MVG, the variance-covariance matrix ($\Sigma_{0l}; l = 1, 2$) considered with unit variances and covariances are equal to 0.5. The size of the dimensions is set equal to $p = 100$ and $p = 300$. The sample size for the two samples is considered as $(m = p \times 30\%, n = p \times 30\%)$ and $(m = p \times 30\%, n = p \times 45\%)$.

The simulation for this study is conducted in a three-step process. In the first step, T_{1t} , T_{2t} and T_{3t} are computed a large number of times say $t = 20,000,000$ for each distributional environment (i-iv), dimensions and sample sizes explained above. Based on these 20,000,000 statistics, means ($\mu_{T_1}, \mu_{T_2}, \mu_{T_3}$) and variances ($\sigma_{T_1}, \sigma_{T_2}, \sigma_{T_3}$) are estimated, which are presented in Table 1. In the second step, the IC Average Run Length (ARL_0) is considered equal to 200 (i.e., $ARL_0 = 200$)

under different distributions and dependency structures, values of control limit coefficients h_1 h_2 , h_3 , h_4 , h_5 and h_6 are found using 10,000 iterations and steps discussed in **Figure 2**.

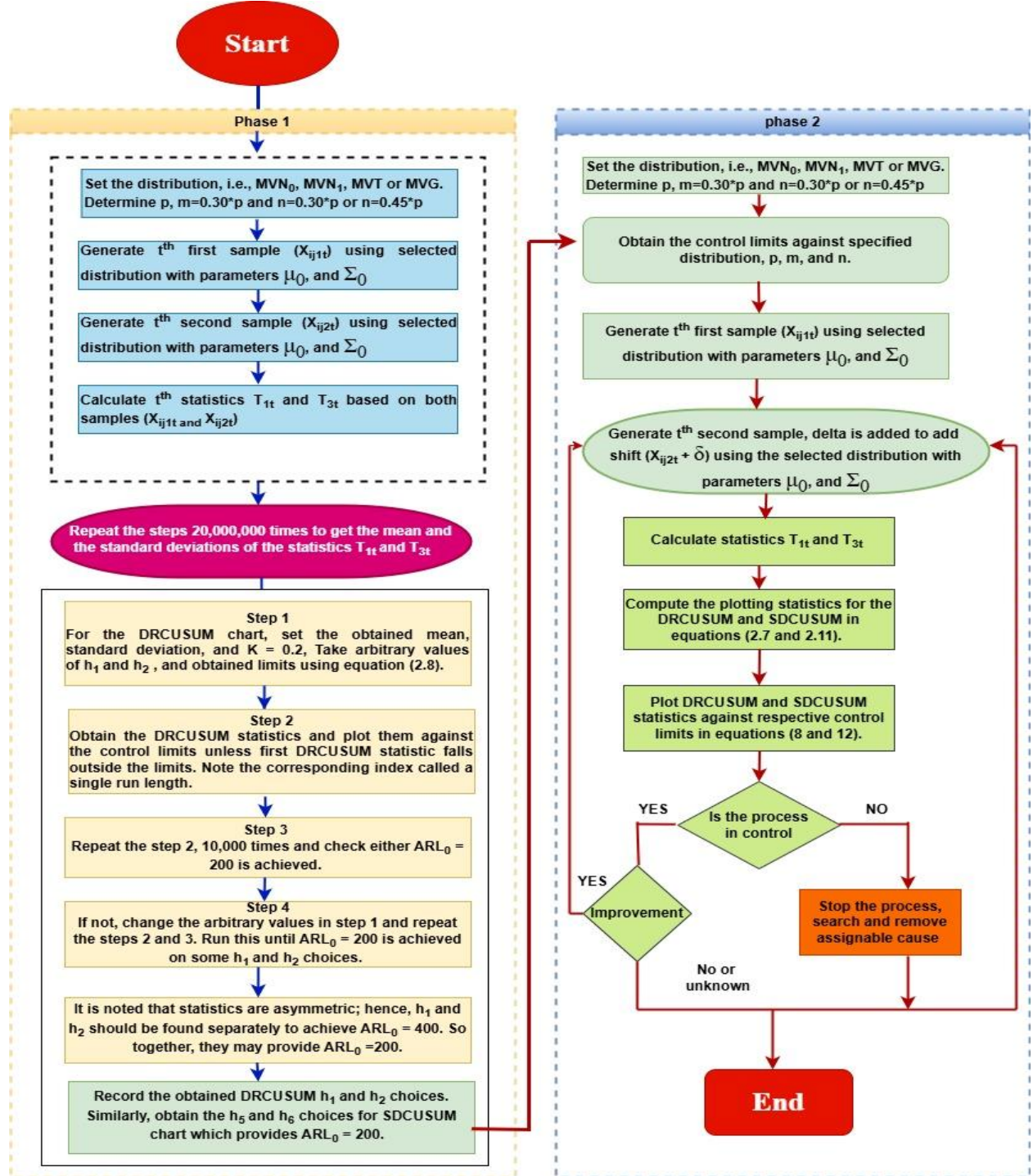


Figure 2. The schematic diagram for the implementation of proposed control charts.

The obtained control limit coefficients are provided in **Table 1**. In the last step, at a fixed $ARL_0 = 200$, run length (RL) performance of the proposed charts is calculated by adding the shifts in the second sample (i.e., $x_{ij2t} + \delta$), where shifts are considered equal to 0.1, 0.2, 0.3, 0.4, 0.5, 0.75 and 1. The detection ability of the proposed charts under the above-mentioned shifts is observed under different distributional environments (i-iv), dimensions, and sample sizes using 10,000 iterations of Monte Carlo simulations in the R language program. Finally, the results are reported in the form of ARL, SDRL, Quartiles of RL, and EQL are provided in **Tables 2-4**.

Table 1 The mean, standard deviation, and coefficient values of the proposed charts at fixed $ARL_0 = 200$

p	Sample Size	Distribution	DRCUSUM				SDCUSUM			
			μ_{T_1}	σ_{T_1}	h_1	h_2	μ_{T_3}	σ_{T_3}	h_5	h_6
100	$m = p * 30\%$, $n = p * 30\%$	MVN_0	1.000040309	0.14255512	3.836	4.544	-0.0029238	0.951397836	3.49	4.179
		MVN_1	1.000187538	0.183380737	3.606	4.803	-0.00144175	0.928676369	3.116	4.23
		MVT	0.999864439	0.180988125	1.892	2.809	-0.00678564	0.655156075	1.57	2.3653
		MVG	0.999702409	0.175502044	3.278	4.357	-0.01587835	0.886857558	2.975	3.785
	$m = p * 30\%$, $n = p * 45\%$	MVN_0	0.999691981	0.142201644	3.846	4.519	-0.00459206	0.949071417	3.491	4.14
		MVN_1	1.000204864	0.183483473	3.606	4.781	-0.00069258	0.928764712	3.107	4.196
		MVT	0.999312465	0.189333435	2.144	3.139	-0.00613944	0.700265498	1.764	2.64
		MVG	1.000392461	0.177122464	3.357	4.458	-0.00691915	0.901689453	3.00	3.888
300	$m = p * 30\%$, $n = p * 30\%$	MVN_0	0.999657982	0.081928849	3.977	4.358	-0.00463596	0.951974795	3.622	3.985
		MVN_1	0.999674135	0.10552098	3.858	4.488	-0.00349928	0.935901972	3.399	4.08
		MVT	1.00021901	0.105306809	1.871	2.355	0.000133486	0.629888276	1.575	2.014
		MVG	0.999803006	0.101147208	3.762	4.328	-0.00509719	0.921848875	3.332	3.865
	$m = p * 30\%$, $n = p * 45\%$	MVN_0	0.999757358	0.081894671	3.965	4.358	-0.00356889	0.956090162	3.608	4.105
		MVN_1	0.99991906	0.106006896	3.848	4.502	-0.00114039	0.944174072	3.422	4.05
		MVT	1.000442512	0.110259723	2.169	2.728	0.002107492	0.682944042	1.837	2.348
		MVG	1.000122774	0.101540601	3.762	4.368	-0.00112509	0.933384975	3.395	3.907

2.4. Performance and Comparative Analysis

The RL profile of the proposed charts is based on $ARL_0 = 200$, shifts ($\delta = 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.75, 1$) have been evaluated under MVN_0 , MVN_1 , MVT , and MVG distribution. We have considered two combinations of sample sizes for both equal ($m = 30\%$; $n = 30\%$) and unequal cases ($m = 30\%$; $n = 45\%$) covering $p = 100$ and 300 dimensions. We have computed RL-based measures, including ARL, SDRL, Quartiles of RL, and EQL. All results of all these measures are provided in **Tables 2-4** for the suggested schemes. The presented schemes have been contrasted with the DR and SD schemes offered by Mahmood, et al. [23]; their results are also included in **Tables 2-4**.

Table 2 RL profile of proposed and competing charts for equal sample sizes ($m = n = p \times 30\%$)

Distribution	δ	$p = 100$				$p = 300$			
		DR	SD	DRCUSUM	SDCUSUM	DR	SD	DRCUSUM	SDCUSUM
MVN_0	0	199.8(196.1)	198.8(194.7)	199.2 (190.8)	200.5 (195.0)	200.5(199.7)	200.6(198.9)	200.4 (196.7)	199.6 (195.0)
	0.1	202.7 (203.4)	201 (200.2)	192.6 (183.9)	192.0 (185.6)	170.8(172.2)	171.6(171.5)	114.4 (108.5)	116.8 (111.1)
	0.2	180.6 (180.4)	181.2 (179.9)	143.7 (135.6)	149.5 (141.6)	51.5 (50.1)	52.1 (51.3)	15.5 (11.2)	16.1 (12.0)
	0.3	136.0 (138.4)	139.8 (140.4)	67.8 (61.5)	72.3 (67.6)	9.0 (8.6)	9.4 (8.9)	4.6 (2.0)	4.7 (2.0)
	0.4	80.8 (79.5)	85.1 (85.1)	27.3 (22.4)	29.9 (25.2)	2.3 (1.7)	2.4 (1.8)	2.3 (0.7)	2.4 (0.8)
	0.5	42.0 (42.2)	45.9 (46.1)	13.2 (8.9)	14.3 (10.1)	1.2 (0.4)	1.2 (0.5)	1.5 (0.5)	1.6 (0.5)
	0.75	7.6 (7)	9.1 (8.6)	4.3 (1.8)	4.6 (2.1)	1.0 (0)	1.0 (0)	1.0 (0)	1.0 (0)
	1	2.1 (1.5)	2.6 (2.1)	2.4 (0.76)	2.5 (0.88)	1.0 (0)	1.0 (0)	1.0 (0)	1.0 (0)
MVN_1	0	197.6 (197.3)	199 (199.2)	199.7(197.13)	199.8 (198.2)	197.5(193.8)	197.1(194.4)	199.6 (194.2)	200.2(195.8)
	0.1	199.1 (195.2)	199.8 (198.6)	201.2 (198.9)	198.9 (196.9)	188.2(190.2)	186.7(189.8)	137.8 (131.0)	148.8 (144.5)
	0.2	190.8 (190.2)	189.7 (192)	170.1 (161.9)	174.4 (171.5)	86.4 (84.4)	87.1 (86.6)	24.0 (19.4)	26.6 (22.0)
	0.3	168.7 (164.7)	170.6 (165.5)	100.5 (94.7)	106.7 (100.7)	20.1 (19.3)	20.5 (19.7)	6.5 (3.2)	6.8 (3.4)
	0.4	116.8 (118)	121 (119.9)	46.4 (41.6)	50.9 (47.1)	4.9 (4.3)	5.1 (4.5)	3.2 (1.1)	3.3 (1.2)
	0.5	72.8 (73.7)	76.6 (77.7)	21.4 (16.6)	23.8 (19.3)	1.7 (1.1)	1.8 (1.2)	2.0 (0.6)	2.1 (0.6)
	0.75	17.5 (17)	20.2 (19.7)	6.2 (2.9)	6.6 (3.4)	1.0 (0)	1.0 (0)	1.0 (0)	1.0 (0.1)
	1	4.6 (4.1)	5.8 (5.3)	3.2 (1.1)	3.4 (1.3)	1.0 (0)	1.0 (0)	1.0 (0)	1.0 (0)
	0	198.5 (197.6)	199.1 (198.7)	201.1 (201.8)	200.9 (198.6)	199.5(197.4)	201.3(205.4)	199.4 (192.8)	199.7 (195.2)

MVT	0.1	195.9 (195.9)	195.6 (196.9)	200.4 (200.4)	200.7 (200.2)	199.5(199.9)	196.7(196.7)	184.6 (180.4)	185.3 (183.5)
	0.2	191.2 (189.7)	190.2 (188.8)	193.1 (192.9)	193.8 (193.4)	137.3(138.7)	133.4(133.1)	73.5 (68.3)	77.3 (72.3)
	0.3	185.5 (185.4)	182.5 (182.2)	159.4 (154.2)	161.7 (160.3)	54.7 (53.3)	53.3 (52)	17.8 (14.8)	18.8 (15.9)
	0.4	154.3 (151.5)	151.9 (149)	106.1 (101.0)	110.6 (109.4)	17.0 (16.4)	17.0 (16.1)	6.1 (3.7)	6.2 (3.9)
	0.5	119.2 (116.8)	117.3 (116.5)	60.3 (56.3)	63.6 (60.0)	5.5 (4.9)	5.7 (5.2)	3.2 (1.4)	3.3 (1.5)
	0.75	42.0 (41.2)	42.7 (42.4)	13.9 (11.0)	15.0 (12.4)	1.1(0.4)	1.2 (0.5)	1.0 (0.5)	1.3 (0.5)
	1	12.8 (12.1)	14.2 (13.6)	5.3 (2.9)	5.6 (3.3)	1.0 (0)	1.0 (0.1)	1.0 (0)	1.0 (0)
MVG	0	195.0 (196.9)	197.1 (197.4)	200.2 (192.9)	199.4 (195.7)	200.5(201.4)	196.6(195.6)	200.8 (197.9)	199.5 (197.4)
	0.1	198.6 (197.7)	198.5 (198.1)	200.2 (193.6)	199.1 (194.7)	187.0(187.6)	182.9(181.7)	143.6 (135.9)	139.5 (133.7)
	0.2	188.8 (191.5)	184.7 (184.1)	169.6 (164.0)	166.0 (160.6)	78.9 (78.1)	73.3 (72.4)	25.0(23.0)	22.7 (18.2)
	0.3	163.6 (164.1)	155.4 (154.3)	98.3 (92.7)	89.3 (84.2)	17.7 (17.2)	16.4 (15.9)	6.8 (3.9)	5.9 (2.9)
	0.4	115.3 (115.2)	101.8 (101.5)	43.4 (38.3)	38.6 (33.9)	4.3 (3.8)	3.9 (3.3)	3.4 (1.5)	2.9 (1.0)
	0.5	68.8 (68.2)	58.7 (58.3)	20.1 (15.6)	17.9 (13.9)	1.6 (1)	1.5 (0.9)	2.2 (0.9)	1.9 (0.5)
	0.75	15.2 (14.9)	13.2 (12.7)	5.7 (2.7)	5.2 (2.5)	1.0 (0)	1.0 (0)	1.3 (0.4)	1.0 (0)
	1	3.9 (3.4)	3.7 (3.2)	2.9 (1.0)	2.7 (1.0)	1.0 (0)	1.0 (0)	1.0 (0)	1.0 (0.1)

The first row of each cell shows ARL (SDRL), and the second row of each cell shows the values of 25th, 50th, and 75th percentiles, respectively. Bold numbers mean the smallest ARL for each shift in different dimension sizes.

Table 3 RL profile of proposed and competing charts for unequal sample sizes ($m = p \times 30\%$, $n = p \times 45\%$)

Distribution	δ	$p = 100$				$p = 300$			
		DR	SD	DRCUSUM	SDCUSUM	DR	SD	DRCUSUM	SDCUSUM
MVN₀	0	204.3(203.1)	199.7(201.7)	201.4(195.2)	200.9 (195.7)	199.8(206.2)	199.4(197.2)	199.2 (193.7)	201.1(194.9)
	0.1	196.6(194.7)	196.4 (195.3)	198.6(188.7)	198.1 (191.6)	182.7(181.5)	177.4(176.6)	142.9 (136.6)	151.8(146.8)
	0.2	183.6(182.2)	182.8(181.7)	168.2 (164.6)	169.8 (163.0)	78.9 (79.5)	79.5 (79.4)	26.3 (21.2)	27.9 (22.8)
	0.3	155.2(155.1)	157 (156.8)	101.6 (96.6)	106.4 (103.0)	18.6 (18.3)	19.2 (18.9)	7.1 (3.6)	7.2 (3.9)
	0.4	111.6(110.2)	116.3(116.5)	48.7 (42.9)	52.3 (48.1)	4.6 (4.1)	4.7 (4.2)	3.4 (1.2)	3.5 (1.3)
	0.5	70.3 (70.1)	73.4 (73.9)	23.6 (18.7)	25.4 (20.8)	1.7 (1.1)	1.8 (1.2)	2.1 (0.6)	2.2 (0.6)
	0.75	16.9 (16.1)	18.9 (18.7)	6.7 (3.4)	7.1 (3.7)	1.0 (0)	1.0 (0)	1.0 (0.2)	1.0 (0.2)
	1	4.7 (4.1)	5.7 (5.1)	3.4 (1.3)	3.6 (1.4)	1.0 (0)	1.0 (0)	1.0 (0)	1.0 (0)
MVN₁	0	196.3(198.3)	195.2(195.3)	200.1(194.6)	201.2 (195.4)	198.7(195.7)	197.8(195.5)	199.4 (192.1)	200.5
	0.1	199.5(196.3)	200.4	199.4(194.2)	196.7 (190.7)	191.7(189.8)	189.9(188.8)	162.4 (155.5)	164.0(156.1)
	0.2	190.0(189.6)	189.5 (189)	181.1(175.5)	179.6 (173.2)	109.8(108.2)	109.1(108.3)	41.7 (36.0)	43.4 (38.5)
	0.3	177.7(175.5)	171.3 (170)	129.7 (125.5)	133.3 (129.5)	34.9 (34.3)	35.0 (34)	10.7 (6.6)	10.9 (7.0)

	0.4	138.7 (139)	139.3	74.1 (68.5)	78.0 (72.9)	9.8 (9.3)	10.1 (9.5)	4.5 (2.0)	4.7 (2.1)
	0.5	99.2 (100)	101.3	38.0 (32.5)	41.4 (36.2)	3.4 (2.9)	3.4 (2.9)	2.1 (0.6)	2.8 (1.0)
	0.75	31.7 (31.4)	34.5 (34.1)	10.0 (5.9)	10.8 (6.8)	1.0 (0.2)	1.0 (0.2)	1.2 (0.3)	1.3 (0.4)
	1	9.7 (9.2)	11.6 (11.3)	4.7 (2.1)	4.9 (2.3)	1.0 (0)	1.0 (0)	1.0 (0)	1.0 (0)
MVT	0	196.0(194.3)	194.8(193.2)	199.4(195.1)	199.6 (196.7)	201.3(200.8)	200.5(202.5)	199.7 (193.3)	200.1
	0.1	195.6(194.1)	195.0 (194.2)	198.4(195.1)	202.4 (198.6)	198.2(198.7)	199.9(195.7)	192.6 (190.8)	193.9
	0.2	195.3(195.3)	194.8(193.4)	194.6 (193.9)	197.5 (190.8)	159.9(160.4)	156.9(159.9)	104.3 (100.7)	107.6
	0.3	190.9(191.1)	187.7(192.7)	172.7 (168.2)	175.1 (172.0)	86.9 (85.3)	81.5 (81.9)	31.4 (28.2)	33.1 (30.1)
	0.4	170.6(170.4)	167.1(163.7)	133.6 (125.8)	134.8 (128.5)	34.6 (34.2)	32.3 (32.1)	10.7 (7.5)	11.0 (8.0)
	0.5	144.4(142.3)	142.3(140.1)	89.4 (84.5)	92.3 (87.1)	13.0 (12.6)	12.2 (11.8)	5.1 (2.7)	5.2 (2.9)
	0.75	68.8 (67.5)	67.4 (66.5)	26.5 (22.7)	28.0 (24.8)	1.9 (1.3)	2.0 (1.4)	1.8 (0.7)	2.0 (0.7)
	1	27.6 (26.8)	28.2 (27.3)	9.6 (6.6)	10.0 (7.3)	1.0 (0.2)	1.1 (0.3)	1.1 (0.3)	1.2 (0.4)
MVG	0	199.7(196)	201.7(202.5)	201.7(196.0)	201.8 (198.4)	204.3(203.2)	201.3(202.1)	200.3 (196.9)	199.3(195.0)
	0.1	204.2 (202.1)	204.4 (202)	207.6(200.6)	209.8 (208.8)	202.5(203.6)	200.5(201.3)	167.6 (165.9)	165.2 (161.2)
	0.2	211.9(209.1)	209.9(209.8)	193.8 (188.3)	194.8 (189.3)	119.2(119.1)	114.4 (116)	39.0(33.0)	38.2 (32.7)
	0.3	200.6(201.6)	189.9(188.4)	139.9(134.0)	135.0 (130.8)	34.6 (34)	32.3 (32.4)	9.9 (5.9)	9.5 (5.7)
	0.4	162.4(162.3)	152.7(152.9)	77.7 (71.3)	72.3 (67.6)	9.3 (8.8)	8.5 (8.0)	4.4 (1.8)	4.2 (1.8)
	0.5	116.9(116.7)	103.8(103.2)	37.6 (32.6)	34.7 (30.3)	3 (2.5)	2.9 (2.3)	2.7 (0.9)	2.6 (0.8)
	0.75	34.3 (34)	29 (28.8)	9.4 (5.3)	8.7 (5.1)	1.0 (0.1)	1.0 (0)	1.3 (0.4)	1.2 (0.4)
	1	9.6 (9)	8.7 (8.3)	4.4 (1.8)	4.2 (1.8)	1.0 (0)	1.0 (0)	1.0 (0)	1.0 (0)

The first row of each cell shows ARL (SDRL), and the second row of each cell shows the values of 25th, 50th, and 75th percentiles, respectively. Bold numbers mean the smallest ARL for each shift in different dimension sizes.

Table 4 The EQL values of proposed and competing charts

Sample Size		$m = n = p \times 30\%$				$m = p \times 30\%, n = p \times 45\%$			
p	Distribution	DR	SD	DRCUSUM	SDCUSUM	DR	SD	DRCUSUM	SDCUSUM
100	MVN₀	6.611	7.158	3.297	3.504	10.153	10.783	4.965	5.231
	MVN₁	10.570	36.913	4.737	5.066	14.788	15.462	6.933	7.305
	MVT	17.820	17.941	9.360	9.793	24.867	24.566	13.506	13.951
	MVG	9.906	8.852	4.501	4.136	16.593	14.903	7.002	6.627
300	MVN₀	0.813	0.821	0.586	0.598	1.079	1.088	0.724	0.743
	MVN₁	1.133	1.146	0.697	0.727	1.513	1.517	0.941	1.750

	MVT	2.019	2.025	1.142	1.215	3.134	3.018	1.691	1.757
	MVG	1.066	1.017	1.314	0.679	1.536	1.477	0.909	0.878

The results of the proposed and competing charts given in **Tables 2-4** advocate the following:

- For all the distribution structures, with $p = 100$ and $p = 300$, for both combinations of sample sizes (i.e., $(m = n = p \times 30\%)$ and $(m = p \times 30\%, n = p \times 45\%)$), the proposed charts have shown superior detection ability than the DR and SD (cf. **Tables 2-3**), as their values of ARL_1 for different shifts are less than that of the conventional charts.
- Under the distribution structures (i.e., MVN_0 , MVN_1 , and MVT) with $p = 100$ and $p = 300$ for the different combinations of sample sizes (i.e., $(m = n = p \times 30\%)$ and $(m = p \times 30\%, n = p \times 45\%)$), the proposed DRCUSUM chart has shown better detection ability than the SDCUSUM chart (cf. **Tables 2-3**), as it offers the lower ARL_1 values for various shifts. While for the MVG distribution structure with $p = 100$ and $p = 300$, for the different combinations of sample sizes, i.e., $(m = n = p \times 30\%)$ and $(m = p \times 30\%, n = p \times 45\%)$, the proposed SDCUSUM chart has shown better detection ability than the DRCUSUM chart (cf. **Tables 2-3**), as it offers the lowest ARL_1 values for various shifts.
- Under all distribution environments with $p = 100$ and $p = 300$ and equal sample size choice $(m = n = p \times 30\%)$, the proposed control charts have shown superior detection ability compared to their performance under unequal sample size $(m = p \times 30\%, n = p \times 45\%)$ (cf. **Tables 2-3**).
- The proposed memory control charts with $p = 100$ and $p = 300$ for both combinations of sample sizes (i.e., $(m = n = p \times 30\%)$ and $(m = p \times 30\%, n = p \times 45\%)$),

generally showed the lowest SDRL values compared to the memoryless control chart for various shifts. This indicates that the proposed charts are more sensitive to changes under different distribution structures. (cf. **Tables 2-3**).

- For overall performance, under the $p = 300$ with $(m = n = p \times 30\%)$ or $(m = p \times 30\%, n = p \times 45\%)$ combination sample size, the proposed control charts have shown superior detection ability compared to their performance under $p = 100$ with all combinations of sample sizes (cf. **Tables 4**). Because they offer the lowest EQL values for all the distribution structures.
- The reference value K in the proposed control charts plays a crucial role in detecting shifts in the process mean. The increasing values of K have an inverse relation to the performance of the proposed charts, making it less sensitive to small shifts in the process mean. For example, the proposed DRCUSUM chart with equal sample sizes $(m = n = p \times 30\%)$, $p = 100$ and under MVN_1 structure provides lowest ARL_1 values when $K = 0.25$ compared to the performance of the same chart with the same parameters for different K values 0.50, 0.75, and 1.0 (shown in **FIGURE 3**). Hence, the lower choice of K , makes both DRCUSUM and BSCUSUM charts more efficient in detecting the small shifts.
- Overall, for detecting shifts in the second sample of the two-sample high-dimensional data, the proposed memory control charts (i.e., DRCUSUM and SDCUSUM) outperform the competing memoryless control charts (i.e., DR and SD) regardless of the distribution structures (See *EQL* in Error! Reference source not found.).

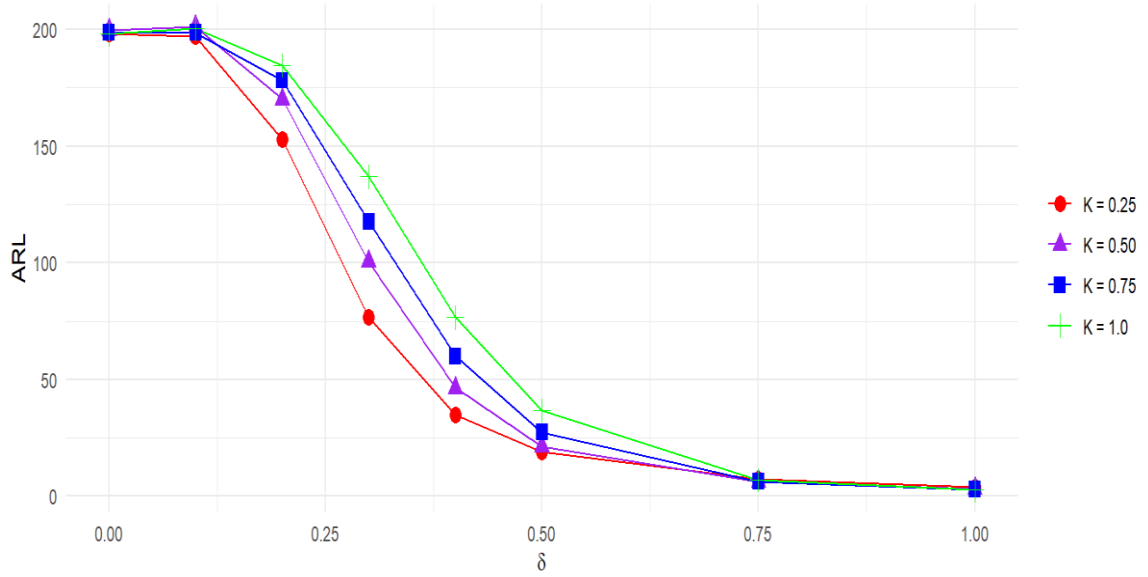


Figure 3 Effect of the K values on the performance of the DRCUSUM chart

2.5. Real-Life Applications

In this section, we provide two real-life applications to show the implementation of the proposed charts. In these two applications, we demonstrate the practical use of these charts in tracking daily stock price fluctuations and monitoring gas sensor arrays under flow modulation, where real-time reactions of high-dimensional data are critical.

2.5.1. An Application for Daily Stock Prices (Application 1)

Daily stock prices are influenced by a variety of factors, including company-specific events, macroeconomic conditions, and investor sentiment. Understanding these prices requires analyzing historical transaction data, which can reveal trends and predict future movements. The daily stock price data provided by Pun is used. The datasets can be found in Mendeley Data; see Pun [72]. The dataset contains 3180 rows (daily stock market prices) and 319 columns (features or

dimensions). Each sample's raw data consists of a 30-time series, demonstrating the daily stock prices for 30 days. We categorized the data into two groups. Group 1 had 1590 samples (stock prices per day during the initial 1590 days) within 53 data matrices, each consisting of 30 stock prices with 319 dimensions. Group 2 with 1590 samples (stock prices per day during the last 1590 days) within 53 data matrices, each consisting of 30 stock prices with 319 dimensions.

Using the 3180 samples, 1590 for group 1 and 1590 for group 2, will give us 53 data matrices, each with 30 samples (stock prices per day for one month) and 319 dimensions. These two high-dimensional samples are used to track changes in the means of two-sample high-dimensional data. The implementation of these proposed charts (DRCUSUM and SDCUSUM charts) was conducted in three steps in the initial step using 20,000,000 iterations, the values of the mean and the standard deviations of T_{1t} and T_{3t} statistics are obtained from our daily stock price data. Subsequently, in the second step, using predefined $ARL_0 = 200$ the values of h_1 , h_2 , h_5 and h_6 are found using 10,000 iterations and the pre-obtained means and standard deviations. Finally, considering the values of means, standard deviations, and the values of h_1 , h_2 , h_5 and h_6 shown in Error! Reference source not found.. We have calculated the plotting statistics for the proposed charts DRCUSUM and SDCUSUM using equations (2.7) and (2.11). This will give us 53 plotting statistics for each proposed chart from our daily stock price data.

Table 5 Statistics and control limit coefficients at $ARL_0 = 200$ for proposed charts used in real application 1

DRCUSUM				SDCUSUM			
Mean	Standard deviation	h_1	h_2	Mean	Standard deviation	h_5	h_6
1740.554	1666.598	1.499	6.970	7645.033	7359.26	1.774	6.535

Bootstrapping is a statistical resampling technique that involves repeatedly sampling with replacement from the original data to estimate the distribution of a sample statistic. When the sample size is small, or the population distribution is unknown, it is beneficial for determining the reliability of an estimate (e.g. mean, median, or variance) [73]. The bootstrap technique is used to generate 200 IC data samples. We start with our original 53 samples and then randomly select 53 samples with replacements from the original sample and repeat this resampling process 200 times. This procedure gives us 200 samples. For the IC data, plotting statistics are calculated for each set of samples. Similarly, we used the same technique to generate 100 samples for the OOC data by introducing a 0.02 shift amount in the second sample (i.e., $x_{ij2t} + \delta$) and plotting statistics for DRCUSUM and SDCUSUM charts are calculated for each set of samples. **Figure 4.** It shows 200 IC and 100 OOC and upper and lower-side test statistic values of the charts and their control limits. The IC and OOC values are separated by a vertical line in each of the two plots, where the IC region is presented in pink color while the white area is used as the OOC region, and the blue points represent the upper-sided statistics for our proposed charts and the black points represent the lower-sided statistics for our proposed charts. Two horizontal green lines in each plot represent UCL and LCL for the corresponding charts. The plotting statistics of the proposed charts are shown in blue or black points when they are in IC state and red points when they are in OOC state.

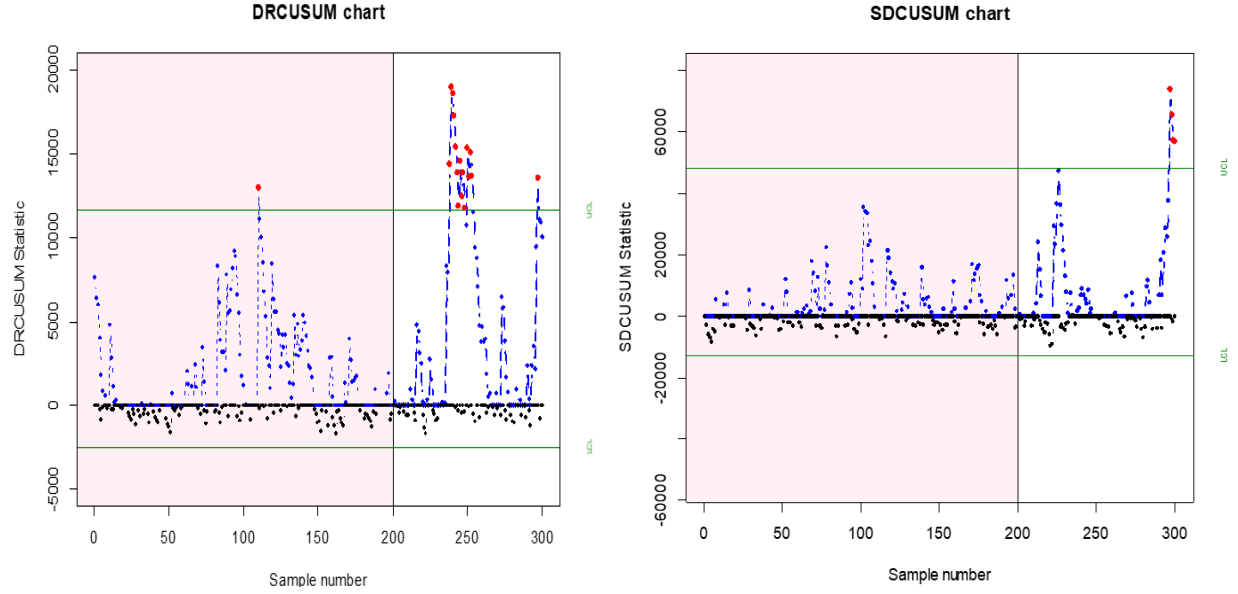


Figure 4 . DRCUSUM and SDCUSUM charts for real application 1

For the DRCUSUM chart, when the process is IC, we notice that all the points are within limits, except one point that is out of control limits, and we consider it a false alarm that has been detected. The performance of DRCUSUM is better than that of SDCUSUM, as the DRCUSUM chart detected 16 OOC values. In contrast, SDCUSUM detected only 4 OOC values. Hence, the DRCUSUM chart is more efficient for studying and monitoring the behavior of daily stock prices.

2.5.2. A Case Study on Gas Sensor Arrays Under Flow Modulation (Application 2)

A gas sensor, known as an electronic nose, is a possible solution for a robotic olfactory system that can detect and distinguish a variety of gas molecules. Gas sensors under flow modulation are often employed in safety-critical industries, such as chemical plants, where accurate detection of toxic or flammable gases is crucial, see Liu, et al. [74] and Shinde and Rout [75]. The most challenging issue in chemical sensing is still sensor drift. A gas sensor array is an advanced detection system comprising several gas sensors; each intended to identify a particular gas or vapor in a specific

environment. The number of sensors in the sensor array varies based on how a case needs to be analyzed and generated [76]. This section describes how our suggested DRCUSUM and SDCUSUM charts are used in practice to track changes in the means of two-sample high-dimensional data. For this purpose, a gas sensor array under flow modulation obtained by Ziyatdinov, et al. [77] is used. The dataset contains 896 rows (a measurement per sensor), 100 columns (features or dimensions), and 896 rows (one measurement per sensor). Each sample's raw data consists of 16 time series, one for each sensor. The twelve tested gas classes formed a relatively wide range of two analytes, acetone, and ethanol, in binary mixtures. For five minutes, each time series was recorded at a 25 Hz (samples per second) sampling rate, producing 100 data points for each time series. We categorized the four gases, acetone, air, ethanol, and a mixture, into two groups according to their chemical nature; group 1 with 448 samples for the organic solvents (Acetone, Ethanol) because both acetone and ethanol are organic solvents, volatile, polar, and flammable. They are frequently used in chemical reactions and as cleaning agents. They are similar in forming vapors that can mix with air, creating flammable mixtures. Group 2 has 448 samples for non-flammable gases (air, mixture) because the air itself is not flammable but can support combustion. Although potentially flammable, a mixture of acetone and ethanol with air contains air as a key component. This group focuses on the presence of air, which distinguishes it from the pure organic solvents in group 1. Using the 896 samples 448 for group 1 and 448 for group 2, each sample contains 16 sensors with 100 features per sensor. We have calculated the plotting statistics for the proposed charts DRCUSUM and SDCUSUM; this will give us 28 plotting statistics for each one of the proposed charts. and we calculate the mean, standard division, and the h_1 , h_2 , h_5 and h_6 coefficient values of the proposed charts, as shown in Error! Reference source not found..

Table 6 . Statistics and control limit coefficients at $ARL_0 = 200$ for proposed charts used in real application 2

DRCUSUM				SDCUSUM			
Mean	Standard deviation	h_1	h_2	Mean	Standard deviation	h_5	h_6
24.78713	23.44102	2.275	5.498	5.144355	5.085552	2.2803	5.4839

The bootstrap resampling technique is used to generate 100 samples. For the in-control data, plotting statistics are calculated for each set of samples. Similarly, we used the same technique for the OOC data control data to generate 50 samples OOC by introducing 0.05 amount of shift in the second sample (i.e., $x_{ij2t} + \delta$). Plotting statistics for DRCUSUM and SDCUSUM charts were calculated for each set of samples. **Figure 5.** Shows the charts' 100 IC and 50 OOC test statistics value, along with their control limits. The IC and OOC values are separated by a vertical line in each of the three plots, where the IC region is presented in pink color while the white area is used as the OOC region, and the blue points represent the upper-sided statistics for our proposed charts and the black points represent the lower-sided statistics for our proposed charts. Two horizontal green lines in each plot represent UCL and LCL for the corresponding charts. The plotting statistics of the proposed charts are shown in blue or black points when they are in IC state and red points when they are OOC state.

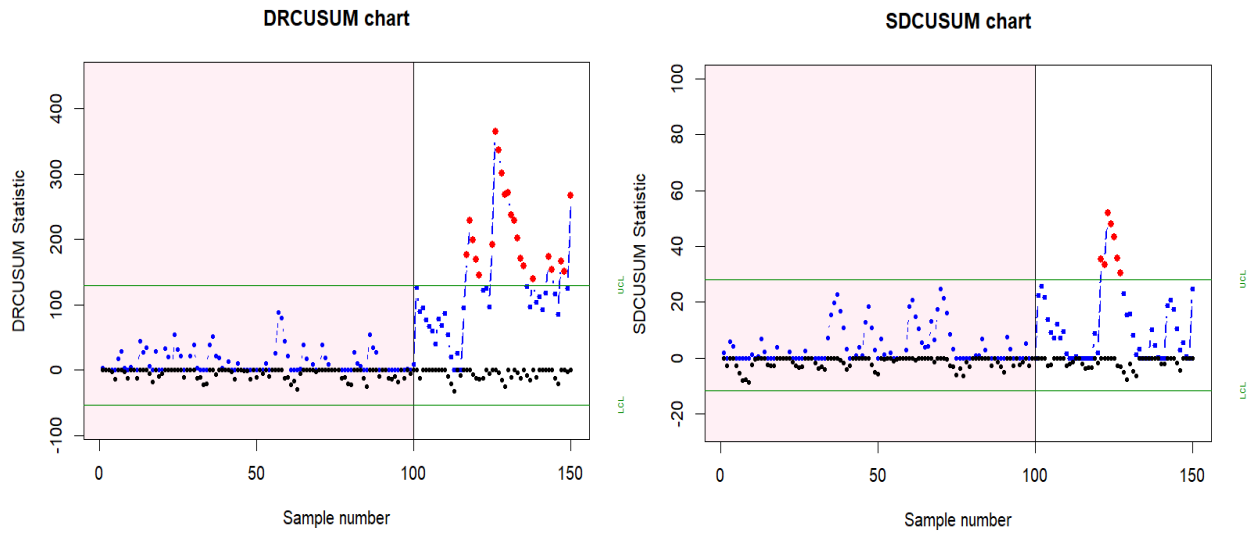


Figure 5 DRCUSUM and SDCUSUM charts for real application 2

For each of the charts, when the process is in control, all the points are within limits, and we notice that no false alarms have been detected. The performance of DRCUSUM is better than SDCUSUM as the DRCUSUM chart detected 22 OOC points. At the same time, SDCUSUM detected only seven OOC points. Hence, the DRCUSUM chart is more efficient in studying and monitoring the behavior of the gas sensor arrays under flow modulation.

CHAPTER 3

CSEWMA CONTROL CHARTS FOR THE MEANS OF TWO

SAMPLED HIGH-DIMENSIONAL DATA

3.1. Introduction

High-dimensional data refers to the dataset with many features or dimensions relative to the number of observations. These datasets are common in many fields [78], such as industrial data [79, 80], Medical Science Data [11, 81], Image Data [13], and Solar Data [82, 83]. High-dimensional data usually relates to a single sample population. For instance, the Lung Cancer dataset [84], Water vapor flux divergence dataset [85]. For a population with two samples, some high-dimensional processes produce high-dimensional data. For instance, the grease damage dataset [23].

SPC remains a foundation of quality management because it makes it possible to identify deviations in technological, medical, and industrial processes. While conventional control charts effectively monitor low-dimensional data, their efficacy diminishes significantly when confronted with the "curse of dimensionality" inherent in high-dimensional datasets. This phenomenon arises because the computational complexity and the probability of false alarms increase exponentially with the number of variables [78]. Control charts are divided due to their structure and efficiency in monitoring the strength of the shifts into memoryless and memory control charts [86]. If relatively large mean shifts in the process are of interest, memoryless control charts like Shewhart

are an easy-to-use and efficient tool for monitoring production processes. However, because the chart only utilizes the process information collected from the most recent sample observation, it is insensitive to small shifts [1]. Therefore, for the small shifts in the process, the Exponentially Weighted Moving Average (EWMA) proposed by [8] is often used. If the purpose is to monitor moderate to large sizes of shifts in a process, Lucas and Saccucci [87] proposed a Combined Shewhart-EWMA (CSEWMA) control chart, which is very effective in monitoring the shifts.

In SPC literature, the proposal of memoryless and memory-type control charts for high-dimensional single-sample data is very common. Mainly multivariate charts such as Hotelling's T^2 and MEWMA are commonly used as alternatives to their univariate charts Shewhart and EWMA charts. This literature is separated into three categories based on the control chart structure [78], such as control charts based on (i) dimension reduction techniques, (ii) variable selection methods, and (iii) high-dimensional approaches, which are thoroughly discussed as follows.

In statistical control charts, dimension reduction techniques refer to transforming high-dimensional process data into a low-dimensional space while simultaneously obtaining crucial information for anomaly detection and monitoring. Ray, et al. [88] reviewed various dimension reduction techniques for high-dimensional data analysis. It categorizes these techniques into feature extraction and feature selection, providing a systematic comparison to address the dimensionality problem. Several methods for reducing the dimension have been effectively incorporated into Shewhart charts to mitigate the difficulties associated with high-dimensional data. The principal component analysis (PCA) comes as a widespread linear method that reduces dimensionality while maintaining a significant part of the variance of the data [89]. However, the PCA linearity assumption can be restricted if data with complex, non-linear relationships occur. To address this restriction, the kernel principal component analysis (KPCA) has turned out to be a

powerful non-linear expansion of PCA [90]. Prasetya, et al. [91] enhanced Hotelling's T^2 control charts by incorporating robust statistical estimators, namely Fast Minimum Covariance Determinant (Fast-MCD) and Minimum Regularized Covariance Determinant (MRCD), to improve detection accuracy against outlier contamination. The research demonstrated the effectiveness of these enhanced charts using the UNSW-NB15 dataset, showing superior performance compared to conventional methods in detecting intrusions despite increased execution time. Oyegoke, et al. [92] proposed a multivariate Hotelling's T^2 chart based on minimum vector variance (MVV) for monitoring high-dimensional correlated processes. Further, other dimension reduction methods have been explored for use with Shewhart charts by the studies [93-95]. In a high dimensional context, the estimation of the covariance matrix may become unstable, which leads to poor performance of the EWMA chart. The approaches to reducing the dimensions increase the stability of the estimate and thus improve the reliability of the chart [78].

. Similar to Shewhart charts, PCA and its variations are often used to reduce the dimensionality of the data before the MEWMA algorithm is used [78, 96]. Epprecht, et al. [97] developed a variable-dimension EWMA T^2 chart (VDEWMA T^2 chart) for monitoring multivariate processes. It combines cost reduction from variable dimension sampling with the enhanced sensitivity of EWMA too small to moderate shifts in the process mean.

Variable selection techniques in statistical control charts refer to ways to find and record the most pertinent process variables while eliminating redundant or unnecessary ones to increase control charts' effectiveness, precision, and readability. A range of variable selection methods were used in process monitoring, which was broadly divided into three categories: filter methods, wrapper methods, and embedded methods [98]. Zhang, et al. [53] developed a new variable selection control chart that incorporates a classification algorithm; it improves sensitivity in identifying

changes in high-dimensional processes, outperforming conventional Shewhart charts in settings with large data. Yan, et al. [99] developed variable-selection (VS) algorithms using penalized likelihood for Shewhart charts in high-dimensional data, emphasizing a Gaussian mixture model (GMM) to address multimodal distributions, enhancing process monitoring and fault diagnosis capabilities effectively. Penalized likelihood approaches [81, 99] have become very popular because they can effectively handle high-dimensional data. Methods include the LASSO and ridge regression. Wang and Jiang [100] proposed a variable-selection-based multivariate statistical process control procedure utilizing a forward-selection algorithm alongside a multivariate control chart specifically designed for monitoring and diagnosing faults in high-dimensional processes, enhancing the effectiveness of Shewhart charts. Sangahn [101] discussed variable selection methods for EWMA charts in high-dimensional multistage processes, emphasizing the importance of selecting relevant variables to enhance process monitoring and control, thereby improving the effectiveness of SPC procedures.

High-dimensional approaches in control charts are related to statistical process monitoring techniques used to process data with a high number of variables relative to the number of observations. These approaches address issues like the curse of dimensionality and an inaccurate covariance estimate without utilizing variable selection or dimensional reduction techniques. Recent studies have focused on high-dimensional techniques that maintain the integrity of the data without reducing dimensions or selecting variables. Ullah, et al. [102] developed a regularized covariance matrix estimation technique to stabilize the estimation of the baseline parameters, and they used a leave-one-out resampling approach to estimate the empirical distribution of future observations. El-Sheikh, et al. [103] introduced two innovative techniques that combine Random Forest with the least absolute shrinkage and selection operator and elastic net, effectively

addressing high-dimensional challenges, including multicollinearity and outliers, while outperforming traditional methods without solely relying on dimension reduction or variable selection. Fang, et al. [85] developed a novel pairwise method for online monitoring of high-dimensional binary data streams, specifically for monitoring extreme weather conditions. This method tackles the difficulties of contemporary sensor technology and the requirement for timely data analysis. Qi et al. (2015) developed a conditional false discovery rate-adjusted scheme for a one-sided MEWMA chart to track the data quality of high-dimensional data streams to improve detection power and reduce false alarms.

These Memoryless (Hotelling's T^2) and memory type (MEWMA) control charts in the aforementioned literature review mainly concentrate on the surveillance of single-sample high-dimensional data; however, limited research exists for two sample scenarios. Recently, Mahmood, et al. [23] proposed Shewhart-type control charts to monitor the means of two-sampled high-dimensional data. As their proposed charts are based on memoryless (Shewhart-type) control charts, which only consider the current information to monitor; therefore, they are less effective in detecting the small shifts in means of two-sample high-dimensional data. Hence, [104] in his master thesis has developed memory-type control charts to monitor the means of two-sampled high-dimensional data based on the EWMA chart structure, which was able to consider the current and past information in the process and effective in detecting the shifts in the means of two-sampled high-dimensional data. However, there are no existing charts for monitoring the moderate to large shifts in the means of two-sampled high-dimensional data. Hence, this chapter intends to propose combination Shewhart and EWMA (CSEWMA) charts to monitor the small to large shifts in the means of two-sampled high-dimensional data. The proposed DRCSEWMA, BSCSEWMA, and SDCSEWMA charts are designed based on the two-sample high-dimensional means tests

named Dempster (DR) [27], Bai & Saranadasa (BS) [67], and Srivastava & Du (SD) [68]. These proposed charts are evaluated based on the run length metrics under different distributional setups, dimensions, and sample sizes. Further, the demonstration of these charts is also shown using gas sensor arrays under flow modulation data.

The remaining part of the chapter is as follows. In Section 3.2, the proposed charts are developed. Section 3.3 describes the designs of simulation study. In Section 3.4, the performance and the comparison of the proposed and the competing charts are presented, and an illustrative example is provided in Section 3.5. The concluding remarks are reported in Section 3.6.

3.2. High-Dimensional Charting Methodologies

This section describes the two-sample high-dimensional mean tests and control charting methodologies used for high-dimensional data under two-sample setups.

3.2.1. The Structure of Two Sample High-Dimensional Data

For the two-sample high-dimensional case, x_{ijlt} , $i = 1, 2, \dots, m$ or $i = 1, 2, \dots, n$, $j = 1, 2, \dots, p$, $l = 1, 2$ and $t = 1, 2, 3, \dots$, is the p -dimensional, t^{th} multivariate normal random vectors with mean vectors $\mu_{0l} = (\mu_1, \dots, \mu_p)'$; $l = 1, 2$ and unknown common positive variance-covariance matrices Σ_{0l} ; $l = 1, 2$. The sizes of the first and second samples are denoted by m and n , respectively; p is the size of the dimension, l is the sample number, and t represents time. The t^{th} statistics T_{1t} , T_{2t} and T_{3t} proposed by Dempster [27], Bai and Saranadasa [67] and Srivastava and Du [68], respectively, for the means testing of two-sample high-dimensional data are defined as follows:

$$T_{1t} = \frac{\frac{mn}{m+n}(\bar{x}_{1t}-\bar{x}_{2t})'(\bar{x}_{1t}-\bar{x}_{2t})}{tr S_t}, \quad (3.1)$$

$$T_{2t} = \frac{\frac{mn}{m+n}(\bar{x}_{1t}-\bar{x}_{2t})'(\bar{x}_{1t}-\bar{x}_{2t})-tr S_t}{\left[\frac{2(m+n)(m+n+1)}{(m+n-1)(m+n+2)} \left(tr S_t^2 - \frac{(tr S_t)^2}{m+n} \right) \right]^{\frac{1}{2}}}, \quad (3.2)$$

$$T_{3t} = \frac{\frac{mn}{m+n}(\bar{x}_{1t}-\bar{x}_{2t})' D_{S_t}^{-1} (\bar{x}_{1t}-\bar{x}_{2t}) - \frac{(m+n)p}{m+n-2}}{\left[2 \left(tr R_t^2 - \frac{p^2}{m+n} \right) c_{p,m+n} \right]^{\frac{1}{2}}}. \quad (3.3)$$

Where the t^{th} sample means the vector of the first sample is $\bar{x}_{1t} = \sum_{i=1}^m x_{ij1t}/m$ and t^{th} sample means the vector of the second sample is $\bar{x}_{2t} = \sum_{i=1}^n x_{ij2t}/n$. The t^{th} diagonal matrix (D_{S_t}) of sample variance-covariance matrix (S) is defined by $D_{S_t} = diag(s_{11}, ..., s_{pp})$, the t^{th} sample correlation matrix (R_t) is defined by $R_t = D_{S_t}^{-\frac{1}{2}} S_t D_{S_t}^{-\frac{1}{2}}$ and the adjustment coefficient $c_{p,m+n}$ can be obtained by $c_{p,m+n} = 1 + \frac{tr R_t^2}{p^{3/2}}$.

3.2.2. Conventional Shewhart Charts for Two Sample High-Dimensional Data

Mahmood, et al. [23] proposed DR, BS, and SD charts to monitor the means of the two-sample high-dimensional data. For the DR chart, the t^{th} statistic T_{1t} presented in equation (3.1) is considered as the plotting statistic against the limits, which are obtained as follows on a given false alarm rate α :

$$\begin{cases} LCL_{(DR)} = \mu_{0T_{1t}} - L_{s1} \sigma_{0T_{1t}} \\ UCL_{(DR)} = \mu_{0T_{1t}} + L_{s2} \sigma_{0T_{1t}} \end{cases} \quad (3.4)$$

For the BS chart, the t^{th} statistic T_{2t} presented in equation (3.2) is considered as the plotting statistic against the limits as follows:

$$\begin{cases} LCL_{(BS)} = \mu_{0T_{2t}} - L_{s3} \sigma_{0T_{2t}} \\ UCL_{(BS)} = \mu_{0T_{2t}} + L_{s4} \sigma_{0T_{2t}} \end{cases} \quad (3.5)$$

For the SD chart, the t^{th} statistic T_{3t} presented in equation (3.3) is considered as the plotting statistic against the limits as follows:

$$\begin{cases} LCL_{(SD)} = \mu_{0T_{3t}} - L_{s5} \sigma_{0T_{3t}} \\ UCL_{(SD)} = \mu_{0T_{3t}} + L_{s6} \sigma_{0T_{3t}} \end{cases} \quad (3.6)$$

It is to be noted that all three charts, named DR, BS, and SD charts, signal an out-of-control (OOC) point when T_{1t} , T_{2t} and T_{3t} falls outside their corresponding control limits given in equations (3.4), (3.5), and (3.6). Otherwise, the process is considered to be an in-control (IC) situation.

3.2.3. Conventional EWMA Charts for Two Sample High-Dimensional Data

For the t^{th} statistic T_{1t} presented in equation (3.1) obtained from the two-sample high-dimensional data is integrated into EWMA statistics as follows:

$$Z_{t(DREWMA)} = \lambda T_{1t} + (1 - \lambda)Z_{t-1(DREWMA)} \quad (3.7)$$

The lower control limit (LCL) and upper control limit (UCL) for the statistic $Z_{t(DREWMA)}$, can be obtained from the following limits:

$$\begin{cases} LCL_{(DREWMA)} = \mu_{0T_{1t}} - L_{e1} \sqrt{\frac{\lambda}{2-\lambda}} \sigma_{0T_{1t}} \\ UCL_{(DREWMA)} = \mu_{0T_{1t}} + L_{e2} \sqrt{\frac{\lambda}{2-\lambda}} \sigma_{0T_{1t}} \end{cases} \quad (3.8)$$

where, $\mu_{0T_{1t}}$ and $\sigma_{0T_{1t}}$ are the mean and standard deviation of the statistic T_{1t} . The values of L_{e1} and L_{e2} can be determined empirically, and $0 < \lambda \leq 1$. It is to be noted that the test statistic T_{1t}

was proposed by Dempster [27]. Therefore, the control chart based on the statistic T_{1t} and EWMA is named the DREWMA chart.

For the t^{th} statistic T_{2t} presented in equation (3.2) obtained from the two-sample high-dimensional data is integrated into EWMA statistics as follows:

$$Z_{t(BSEWMA)} = \lambda T_{2t} + (1 - \lambda)Z_{t-1(BSEWMA)} \quad (3.9)$$

The lower control limit (LCL) and upper control limit (UCL) for the statistic $Z_{t(BSEWMA)}$, can be obtained from the following limits:

$$\begin{cases} LCL_{(BSEWMA)} = \mu_{0T_{2t}} - L_{e3} \sqrt{\frac{\lambda}{2-\lambda}} \sigma_{0T_{2t}} \\ UCL_{(BSEWMA)} = \mu_{0T_{2t}} + L_{e4} \sqrt{\frac{\lambda}{2-\lambda}} \sigma_{0T_{2t}} \end{cases} \quad (3.10)$$

where, $\mu_{0T_{2t}}$ and $\sigma_{0T_{2t}}$ are the mean and standard deviation of the statistic T_{2t} . The values of L_{e3} and L_{e4} can be determined empirically, and $0 < \lambda \leq 1$. It is to be noted that the test statistic T_{2t} was proposed by Bai and Saranadasa [67]. Therefore, the control chart based on the statistic T_{2t} and EWMA is named as BSEWMA chart.

For the t^{th} statistic T_{3t} presented in equation (3.3) obtained from the two-sample high-dimensional data is integrated into EWMA statistics as follows:

$$Z_{t(SDEWMA)} = \lambda T_{3t} + (1 - \lambda)Z_{t-1(SDEWMA)} \quad (3.11)$$

The lower control limit (LCL) and upper control limit (UCL) for the statistic $Z_{t(SDEWMA)}$, can be obtained from the following limits:

$$\begin{cases} LCL_{(SDEWMA)} = \mu_{0T_{3t}} - L_{e5} \sqrt{\frac{\lambda}{2-\lambda}} \sigma_{0T_{3t}} \\ UCL_{(SDEWMA)} = \mu_{0T_{3t}} + L_{e6} \sqrt{\frac{\lambda}{2-\lambda}} \sigma_{0T_{3t}} \end{cases} \quad (3.12)$$

where, $\mu_{0T_{3t}}$ and $\sigma_{0T_{3t}}$ are the mean and standard deviation of the statistic T_{3t} . The values of L_{e5} and L_{e6} can be determined empirically, and $0 < \lambda \leq 1$. It is to be noted that the test statistic T_{3t} was proposed by Srivastava and Du [68]. Therefore, the control chart based on the statistic T_{3t} and EWMA is named as SDEWMA chart.

It is to be noted that all three charts, named DREWMA, BSEWMA, and SDEWMA charts, signal an OOC point when $Z_{t(DREWMA)}$, $Z_{t(BSEWMA)}$, and $Z_{t(SDEWMA)}$ given in equations (3.7), (3.9), and (3.11) respectively fall outside their corresponding control limits given in equations (3.8), (3.10), and (3.12) respectively. Otherwise, the process is considered to be an IC situation. The starting values of $Z_{0(DREWMA)}$, $Z_{0(BSEWMA)}$, and $Z_{0(SDEWMA)}$ before the first sample is taken, is set to equal $\mu_{0T_{1t}}$, $\mu_{0T_{2t}}$ and $\mu_{0T_{3t}}$ respectively.

3.2.4. Proposed CSEWMA Charts for Two Sample High-Dimensional Data

The CSEWMA control chart is a hybrid statistical process monitoring tool that integrates Shewhart's strengths and the sensitivity of EWMA control charts. Shewhart charts efficiently detect large shifts in the means of two-sample high-dimensional data [23]. When the objective is to monitor moderate to large shifts at the same time, combining Shewhart with EWMA charts is effective in such a situation [105, 106]. Hence, the CSEWMA chart proposed by Lucas and Saccucci [87] is an alternate method. Therefore, based on T_{1t} , T_{2t} and T_{3t} Statistics given in equations (3.1 – 3.3), CSEWMA charts were proposed and named by DRCSEWMA,

BSCSEWMA, and SDCSEWMA charts. The plotting statistics of the DRCSEWMA chart are provided as follows:

For the proposed CSEWMA based on the statistic T_{1t} equations (3.1) obtained from the two-sample high-dimensional data are integrated into EWMA statistics. And for the DR chart, the t^{th} statistic T_{1t} presented in equation (3.1) are considered as the two plotting statistics for our proposed chart.

$$\begin{cases} DR_t = T_{1t} \\ Z_{t(DREWMA)} = \lambda T_{1t} + (1 - \lambda)Z_{t-1(DREWMA)} \end{cases} \quad (3.13)$$

The lower control limits (LCL) and upper control limits (UCL) for the proposed chart are the combined limits of the $Z_{t(SDEWMA)}$ statistic, and the DR_t statistic, for a given false alarm α , can be obtained as follows:

$$\begin{cases} LCL_{(DR)} = \mu_{0T_{1t}} - L_{s1} \sigma_{0T_{1t}} \\ UCL_{(DR)} = \mu_{0T_{1t}} + L_{s2} \sigma_{0T_{1t}} \\ LCL_{(DREWMA)} = \mu_{0T_{1t}} - L_{e1} \sqrt{\frac{\lambda}{2-\lambda}} \sigma_{0T_{1t}} \\ UCL_{(DREWMA)} = \mu_{0T_{1t}} + L_{e2} \sqrt{\frac{\lambda}{2-\lambda}} \sigma_{0T_{1t}} \end{cases} \quad (3.14)$$

where, $\mu_{0T_{1t}}$ and $\sigma_{0T_{1t}}$ are the mean and standard deviation of the statistic T_{1t} . The values of L_{s1} , L_{s2} , L_{e1} , and L_{e2} can be determined empirically (cf. Table 1), and $0 < \lambda \leq 1$. It is to be noted that the test statistic T_{1t} was proposed by Dempster [27]. Therefore, the control chart based on the statistic T_{1t} and CSEWMA is named as DRCSEWMA chart.

For the proposed CSEWMA based on the statistic T_{2t} presented in equation (3.2) obtained from the two-sample high-dimensional data is integrated into EWMA statistics. And for the BS chart,

the t^{th} statistic T_{2t} presented in equation (3.2) are considered as the two plotting statistics for our proposed chart.

$$\begin{cases} BS_t = T_{2t} \\ Z_{t(BSEWMA)} = \lambda T_{2t} + (1 - \lambda)Z_{t-1(BSEWMA)} \end{cases} \quad (3.15)$$

The lower control limits (LCL) and upper control limits (UCL) for the proposed chart are the combined limits of the $Z_{t(BSEWMA)}$ statistic, and the BS_t statistic, for a given false alarm α , can be obtained as follows:

$$\begin{cases} LCL_{(BS)} = \mu_{0T_{2t}} - L_{s3} \sigma_{0T_{2t}} \\ UCL_{(BS)} = \mu_{0T_{2t}} + L_{s4} \sigma_{0T_{2t}} \\ LCL_{(BSEWMA)} = \mu_{0T_{2t}} - L_{e3} \sqrt{\frac{\lambda}{2-\lambda}} \sigma_{0T_{2t}} \\ UCL_{(BSEWMA)} = \mu_{0T_{2t}} + L_{e4} \sqrt{\frac{\lambda}{2-\lambda}} \sigma_{0T_{2t}} \end{cases} \quad (3.16)$$

where, $\mu_{0T_{2t}}$ and $\sigma_{0T_{2t}}$ are the mean and standard deviation of the statistic T_{2t} . The values of L_{s3} , L_{s4} , L_{e3} , and L_{e4} can be determined empirically (cf. Table 1), and $0 < \lambda \leq 1$. It is to be noted that the test statistic T_{2t} was proposed by Bai and Saranadasa [67]. Therefore, the control chart based on the statistic T_{2t} and CSEWMA is named as BSCSEWMA chart.

For the proposed CSEWMA based on the statistic T_{3t} presented in equation (3.3) obtained from the two-sample high-dimensional data is integrated into EWMA statistics. And for the SD chart, the t^{th} statistic T_{3t} presented in equation (3.3) are considered as the two plotting statistics for our proposed chart.

$$\begin{cases} SD_t = T_{3t} \\ Z_{t(SDEWMA)} = \lambda T_{3t} + (1 - \lambda)Z_{t-1(SDEWMA)} \end{cases} \quad (3.17)$$

The lower control limits (LCL) and upper control limits (UCL) for the proposed chart are the combined limits of the $Z_{t(SDEWMA)}$ statistic, and the SD_t statistic, for a given false alarm α , can be obtained as follows:

$$\left\{ \begin{array}{l} \left\{ \begin{array}{l} LCL_{(SD)} = \mu_{0T_{3t}} - L_{s5} \sigma_{0T_{3t}} \\ UCL_{(SD)} = \mu_{0T_{3t}} + L_{s6} \sigma_{0T_{3t}} \end{array} \right. \\ \left\{ \begin{array}{l} LCL_{(SDEWMA)} = \mu_{0T_{3t}} - L_{e5} \sqrt{\frac{\lambda}{2-\lambda}} \sigma_{0T_{3t}} \\ UCL_{(SDEWMA)} = \mu_{0T_{3t}} + L_{e6} \sqrt{\frac{\lambda}{2-\lambda}} \sigma_{0T_{3t}} \end{array} \right. \end{array} \right. \quad (3.18)$$

where, $\mu_{0T_{3t}}$ and $\sigma_{0T_{3t}}$ are the mean and standard deviation of the statistic T_{3t} . The values of L_{s5} , L_{s6} , L_{e5} , and L_{e6} can be determined empirically (cf. Table 1), and $0 < \lambda \leq 1$. It is to be noted that the test statistic T_{3t} was proposed by Srivastava and Du [68]. Therefore, the control chart based on the statistic T_{3t} and CSEWMA is named as SDCSEWMA chart.

It is noted that all proposed three control charts, names DRCSEWMA, BSCSEWMA, and SDCSEWMA charts signal an OOC point when DR_t , BS_t , and SD_t falls outside of their control limits given in the equations (3.14), (3.16), and (3.18) respectively. Or $Z_{t(DrEWMA)}$, $Z_{t(BSEWMA)}$, and $Z_{t(SDEWMA)}$ falls outside of their control limits given in the equations (3.14), (3.16), and (3.18) respectively. Otherwise, the process is considered an IC situation.

3.3. Design of The Simulation Study

The simulation study is designed based on the Monte Carlo simulations using the R language program. The performance of control charts is assessed using the Run Length (RL) profile, which measures the number of samples or time points until a signal is triggered. Key metrics derived from RL include: (a) Average Run Length (ARL); ARL represents the average number of points

needed to detect an OOC signal. It is classified into two types as follows: when the process is stable, ARL is denoted ARL_0 is the expected number of samples or observations taken before a false alarm occurs when the process is IC. For an OOC process, it is denoted as ARL_1 is the expected number of samples or observations taken before detecting a process shift when the process is OOC. It is to be noted that a lower ARL_1 indicates faster detection of shifts. (b) Standard Deviation of Run Length (SDRL): SDRL measures variability in RL, providing insight into the consistency of chart performance. It is calculated by $SDRL = \sqrt{E(RL - ARL)^2}$ and the lower SDRL indicates more consistent detection. (c) percentile; The 25th, 50th (median), and 75th percentiles of RL highlight the distribution of run lengths, indicating where the shortest 25%, middle 50%, and longest 25% of RL values lie. (d) Extra Quadratic Loss (EQL); EQL evaluates overall chart performance across all shift magnitudes using a weighted average of ARL over the entire shift domain ($\delta_{min} < \delta < \delta_{max}$) using the square of shift (δ^2) as the weight. The EQL is calculated by $EQL = \int_{\delta_{min}}^{\delta_{max}} \delta^2 ARL(\delta) d\delta / \delta_{max} - \delta_{min}$, where lower EQL values indicate better overall chart performance. By following [71] and Mahmood, et al. [23], two independent high-dimensional samples (represented by $l = 1, 2$) are generated through four distributional environments: (i) multivariate normal with mean vector ($\mu_{0l} = 0; l = 1, 2$) and variance-covariance matrix ($\Sigma_{0l}; l = 1, 2$), where unit variances and covariances are equal to 0, denoted as MVN_0 ; (ii) multivariate normal with mean vector ($\mu_{0l} = 0; l = 1, 2$) and variance-covariance matrix ($\Sigma_{0l}; l = 1, 2$), where unit variances and covariances are equal to 0.5, denoted as MVN_1 ; (iii) multivariate t with 5 degrees of freedom, denoted as MVT; (iv) multivariate gamma with shape parameter 5 and scale parameter 1, denoted as MVG. For MVT and MVG, the variance-covariance matrix ($\Sigma_{0l}; l = 1, 2$) considered with unit variances and covariances are equal to 0.5. The size of the dimensions is set equal to $p = 100$ and $p = 300$. The sample size for the two

samples is considered as $(m = p \times 30\%, n = p \times 30\%)$ and $(m = p \times 30\%, n = p \times 45\%)$.

The simulation for this study is conducted in a three-step process. In the first step, T_{1t} , T_{2t} and T_{3t} are computed a large number of times say $t = 20,000,000$ for each distributional environment (i-iv), dimensions and sample sizes are explained above. Based on these 20,000,000 statistics means $(\mu_{T_1}, \mu_{T_2}, \mu_{T_3})$ and variances $(\sigma_{T_1}, \sigma_{T_2}, \sigma_{T_3})$ are estimated, which are presented in **Table 7**. In the second step, the IC Average Run Length (ARL_0) is considered equal to 200 (i.e., $ARL_0 = 200$) Under different distributions and dependency structures, values of control limit coefficients for CSEWMA-based control charts $L_{S1}, L_{S2}, L_{S3}, L_{S4}, L_{S5}, L_{S6}$ and $L_{e1}, L_{e2}, L_{e3}, L_{e4}, L_{e5}, L_{e6}$ are found using 10,000 iterations, the Shewhart and EWMA-based charts are running simultaneously to obtain the coefficients at $ARL_0 = 800$ on one side and then combined Shewhart-based coefficients at $ARL_0 = 400$ and combined EWMA-based coefficients at $ARL_0 = 400$ and together CSEWMA-based coefficients at $ARL_0 = 200$. The steps are discussed in [Figure 6](#). The obtained control limit coefficients are provided in **Table 7**.

Table 7 The mean, standard deviation, and the coefficient values of the proposed control charts at $ARL_0 = 200$

Sample Size		$m = p * 30\%, n = p * 30\%$				$m = p * 30\%, n = p * 45\%$			
p	Dist.	MVN_0	MVN_1	MVT	MVG	MVN_0	MVN_1	MVT	MVG
100	μ_{T_1}	1.000040309	1.000187538	0.999864439	0.999702409	0.999691981	1.000204864	0.999312465	1.000392461
	σ_{T_1}	0.14255512	0.183380737	0.180988125	0.175502044	0.142201644	0.183483473	0.189333435	0.177122464
	LS_1	2.581	2.388	2.408	2.422	2.577	2.385	2.353	2.378
	LS_2	3.489	3.727	3.719	3.693	3.448	3.722	3.748	3.718
	Le_1	2.644	2.575	2.572	2.598	2.672	2.585	2.577	2.565
	Le_2	3.117	3.181	3.167	3.178	3.068	3.168	3.199	3.228
	μ_{T_1}	0.999657982	0.999674135	1.00021901	0.999803006	0.999757358	0.99991906	1.000442512	1.000122774

DRCSEWMA	300	σ_{T_1}	0.081928849	0.10552098	0.105306809	0.101147208	0.081894671	0.106006896	0.110259723	0.101540601
		Ls_1	2.735	2.628	2.633	2.648	2.729	2.598	2.612	2.635
		Ls_2	3.251	3.432	3.392	3.372	3.252	3.412	3.448	3.408
		Le_1	2.739	2.667	2.708	2.691	2.759	2.695	2.681	2.685
		Le_2	2.999	3.042	3.038	3.031	2.996	3.025	3.041	3.021
BSCSEWMA	100	μ_{T_2}	0.000201319	0.001997831	0.000232192	-0.00057593	-0.002088601	0.001937466	-0.001671207	0.002934977
		σ_{T_2}	1.000052183	1.000572346	0.721134763	0.947208446	0.999206217	1.002318981	0.773646987	0.966375533
		Ls_3	2.595	2.391	2.502	2.409	2.588	2.365	2.448	2.372
		Ls_4	3.424	3.696	3.898	3.706	3.444	3.768	3.858	3.739
		Le_3	2.691	2.601	2.601	2.614	2.671	2.555	2.559	2.582
		Le_4	3.063	3.188	3.238	3.198	3.069	3.239	3.228	3.196
	300	μ_{T_2}	-0.004178261	-0.002917642	0.001482241	-0.001723809	-0.002941502	-0.00059306	0.003920036	0.001377601
		σ_{T_2}	1.000670089	0.999327802	0.68770646	0.979975699	1.000811142	1.004497301	0.742891509	0.988042273
		Ls_3	2.756	2.608	2.730	2.635	2.737	2.595	2.675	2.674
		Ls_4	3.241	3.428	3.536	3.376	3.272	3.408	3.515	3.415
		Le_3	2.716	2.701	2.710	2.692	2.726	2.677	2.728	2.728
		Le_4	3.01	3.058	3.082	3.029	2.998	3.051	3.068	3.024
SDCSEWMA	100	μ_{T_3}	-0.0029238	-0.00144175	-0.00678564	-0.015878346	-0.004592064	-0.000692584	-0.006139438	-0.006919155
		σ_{T_3}	0.951397836	0.928676369	0.655156075	0.886857558	0.949071417	0.928764712	0.700265498	0.901689453
		Ls_5	2.584	2.365	2.495	2.433	2.585	2.345	2.461	2.408
		Ls_6	3.466	3.765	3.898	3.679	3.462	3.788	3.855	3.676
		Le_5	2.665	2.566	2.594	2.595	2.655	2.565	2.561	2.588
		Le_6	3.078	3.196	3.242	3.176	3.078	3.253	3.225	3.138
	300	μ_{T_3}	-0.004635957	-0.003499283	0.000133486	-0.005097199	-0.003568896	-0.001140389	0.002107492	-0.001125099
		σ_{T_3}	0.951974795	0.935901972	0.629888276	0.921848875	0.956090162	0.944174072	0.682944042	0.933384975
		Ls_5	2.742	2.625	2.745	2.652	2.764	2.595	2.709	2.647
		Ls_6	3.255	3.438	3.543	3.382	3.216	3.418	3.508	3.415
		Le_5	2.726	2.673	2.694	2.682	2.691	2.669	2.704	2.707
		Le_6	2.994	3.052	3.075	3.027	3.012	3.046	3.075	2.995

In the last step, at a fixed $ARL_0 = 200$, the run length (RL) performance of the proposed charts is calculated by adding the shifts in the second sample (i.e., $x_{ij2t} + \delta$), where shifts are considered equal to 0.1, 0.2, 0.3, 0.4, 0.5, 0.75 and 1. The detection ability of the proposed charts under the above-mentioned shifts is observed under different distributional environments (i-iv), dimensions, and sample sizes using 10,000 iterations of Monte Carlo simulations using the R language program.

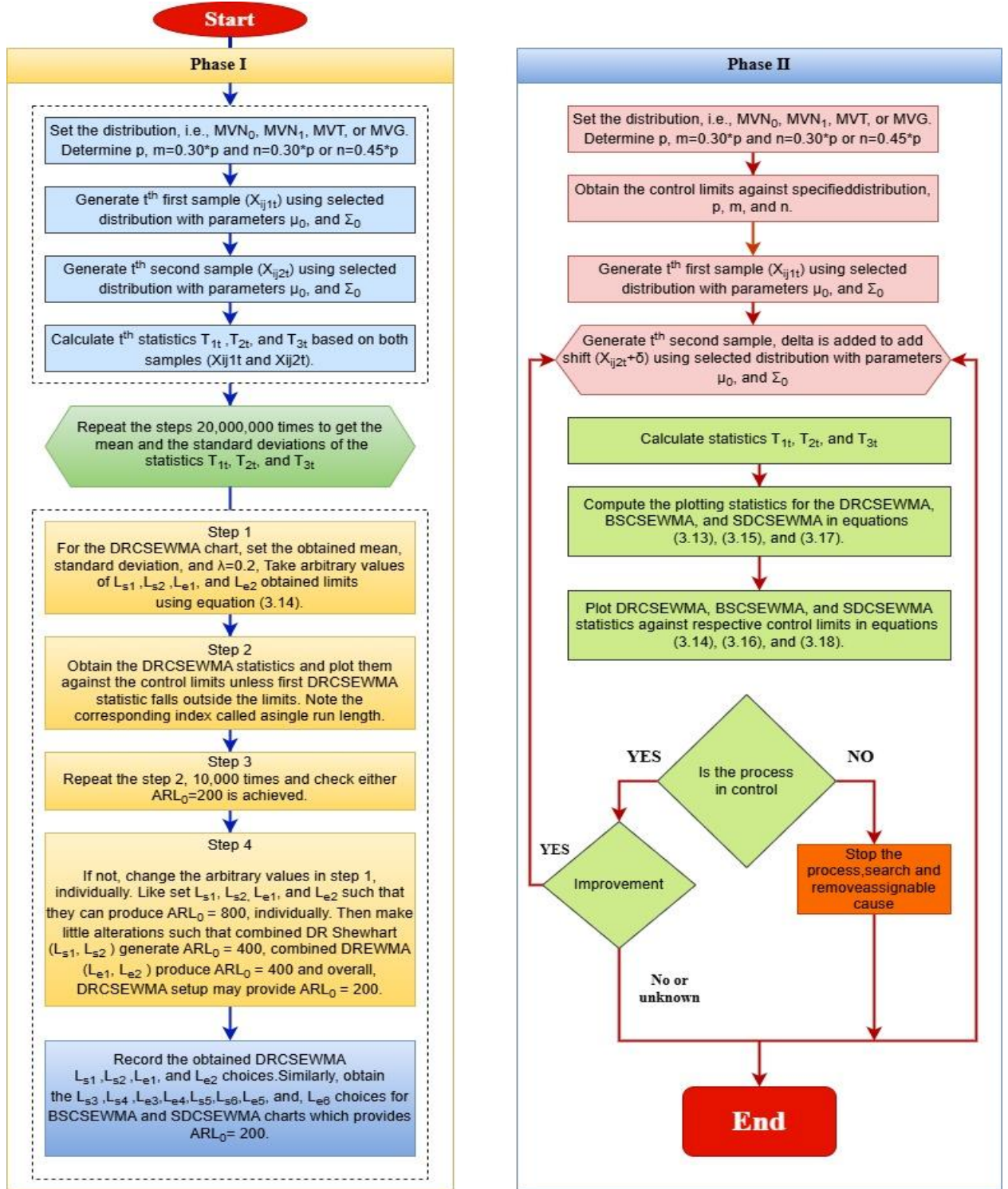


Figure 6 The schematic diagram for the implementation of proposed control charts.

3.4. Performance and Comparative Analysis

Average Run Length (ARL) is the mean number of points required to identify an OOC signal in a process. Whereas the term run length indicates the consecutive points observed before detecting the OOC state of a process. When the process is in a stable state the ARL can be termed as ARL_0 , while the term ARL is denoted by ARL_1 if the process is not in a stable state. Commencing with a predefined value of ARL_0 , lower values of ARL_1 is an indication that a particular control chart can detect shifts early. The RL profile of the suggested charts is based on $ARL_0 = 200$, shifts ($\delta = 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.75, 1$) analyzed under MVN_0 , MVN_1 , MVT , and MVG distribution. We investigated two sample size combinations for both equal ($m = 30\%$; $n = 30\%$) and unequal cases ($m = 30\%$; $n = 45\%$) encompassing $p = 100$ and 300 features. We have calculated RL-based metrics, including ARL, SDRL, Quartiles of RL, and EQL. The outcomes of all these measures are presented in **Tables 8-13** for the suggested charts. The proposed charts have been contrasted with the DR, BS, and SD chars proposed by Mahmood, et al. [23] and with DREWMA, BSEWMA, and SDEWMA charts proposed by [104]; their results have been incorporated in **Tables 8-13**. The focus is to check if the proposed DRCSEWMA, BSCSEWMA, and SDCSEWMA charts outperform the conventional Shewhart and EWMA based on two sampled high-dimensional data. Hence, the run length profile performance is compared to the proposed and the computing charts.

Table 8 ARL(SDRL) and 25th, 50th, and 75th percentiles of the proposed DRCSEWMA and competing charts with m = 30%; n= 30% sample sizes.

Dist.	δ	p = 100			p = 300		
		DR	DREWMA	DRCSEWMA	DR	DREWMA	DRCSEWMA
MVN ₀	0	199.8 (196.1)	197.2 (194.3)	200.2 (195.1)	200.5 (199.7)	197.0 (197.2)	199.7 (200.6)
	0.1	202.7 (203.4)	196.9 (192.6)	198.7 (196.0)	170.8 (172.2)	109.4 (105.4)	123.9 (128.8)
	0.2	180.6 (180.4)	144.2 (139.1)	149.7 (150.8)	51.5 (50.1)	14.8 (10.8)	16.6 (14.9)
	0.3	136.0 (138.4)	64.7 (59.6)	79.1 (76.4)	9.0 (8.6)	4.4 (2.0)	4.0 (1.8)
	0.4	80.8 (79.5)	27.0 (22.6)	34.8 (32.8)	2.3 (1.7)	2.3 (0.8)	1.9 (0.8)
	0.5	42.0 (42.2)	12.7 (8.9)	12.0 (11.1)	1.2 (0.4)	1.6 (0.5)	1.2 (0.5)
	0.75	7.6 (7)	4.1 (1.8)	3.3 (2.2)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
	1	2.1 (1.5)	2.2 (0.8)	2.0 (1.0)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
	EQL	6.610	3.195	3.305	0.812	0.580	0.574
MVN ₁	0	197.6 (197.3)	198.8 (197.3)	199.8 (197.5)	197.5 (193.8)	196.7 (191.8)	201.5 (195.7)
	0.1	199.1 (195.2)	202.3 (198.9)	201.1 (201.3)	188.2 (190.2)	137.8 (130.9)	151.7 (160.7)
	0.2	190.8 (190.2)	169.8 (165.7)	178.4 (177.2)	86.4 (84.4)	24.3 (19.9)	31.6 (28.5)
	0.3	168.7 (164.7)	97.4 (92)	115.0 (119.6)	20.1 (19.3)	6.5 (3.5)	6.7 (3.9)
	0.4	116.8 (118)	45.2 (40.1)	51.3 (56.7)	4.9 (4.3)	3.1 (1.2)	2.8 (1.4)
	0.5	72.8 (73.7)	21.7 (17)	19.5 (23.3)	1.7 (1.1)	2.0 (0.6)	1.6 (0.8)
	0.75	17.5 (17)	5.9 (3.1)	5.4 (2.9)	1.0 (0.0)	1.0 (0.1)	1.0 (0.0)
	1	4.6 (4.1)	3.0 (1.1)	2.6 (1.3)	1.0 (0.0)	1.0 (0)	1.0 (0.0)
	EQL	10.570	4.635	4.707	1.133	0.696	0.7188
MVT	0	198.5 (197.6)	196.0 (195.1)	199.5 (196.9)	199.5 (197.4)	200.6 (195.3)	199.2 (195.0)
	0.1	195.9 (195.9)	193.8 (186.6)	201.4 (201.1)	199.5 (199.9)	175.4 (171.3)	183.5 (191.5)
	0.2	191.2 (189.7)	181.5 (179.6)	188.1 (187.8)	137.3 (138.7)	55.2 (49.8)	61.1 (67.9)
	0.3	185.5 (185.4)	139.0 (133.4)	156.2 (164.4)	54.7 (53.3)	14.5 (10.4)	15.2 (13.4)
	0.4	154.3 (151.5)	83.2 (78.1)	90.5 (104.9)	17.0 (16.4)	6.0 (3.1)	4.8 (3.5)
	0.5	119.2 (116.8)	45.8 (41.2)	45.5 (58.0)	5.5 (4.9)	3.4 (1.4)	2.0 (1.2)
	0.75	42.0 (41.2)	12.0 (8.3)	11.4 (10.5)	1.1 (0.4)	1.5 (0.5)	1.0 (0.5)
	1	12.8 (12.1)	5.2 (2.7)	4.2 (3.0)	1.0 (0.0)	1.1 (0.2)	1.0 (0.0)
	EQL	17.820	7.843	7.926	2.033	1.119	0.994
	0	195.0 (196.9)	196.1 (189.7)	199.7 (196.9)	200.5 (201.4)	205.1 (202.5)	199.2 (199.8)
	0.1	198.6 (197.7)	200.4 (198.8)	205.5 (203.3)	187.0 (187.6)	139.9 (135.3)	143.2 (147.7)
	0.2	188.8 (191.5)	167.1 (164.6)	175.0 (183.1)	78.9 (78.1)	22.7 (18.0)	23.1 (21.2)
	0.3	163.6 (164.1)	91.8 (87.4)	110.3 (117.8)	17.7 (17.2)	6.1 (3.2)	5.7 (3.5)

MVG	0.4	115.3 (115.2)	41.1 (36.7)	45.7 (43.8)	4.3 (3.8)	3.0 (1.1)	2.3 (1.3)
	0.5	68.8 (68.2)	19.5 (15.2)	18.8 (16.6)	1.6 (1.0)	1.9 (0.6)	1.1 (0.7)
	0.75	15.2 (14.9)	5.5 (2.8)	4.7 (2.4)	1.0 (0.0)	1.0 (0.1)	1.0 (0.0)
	1	3.9 (3.4)	2.8 (1.0)	2.0 (1.0)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
	EQL	9.905	4.329	4.362	1.066	0.682	0.637

The first row of each cell shows ARL (SDRL), the second row of each cell shows the value of 25th, 50th, and 75th percentiles, respectively and the last cell in each chart shows the value of the EQL. Bold numbers mean the smallest ARL for each shift in different dimension sizes.

Table 9 ARL(SDRL) and 25th, 50th, and 75th percentiles of the proposed DRCSEWMA and competing charts with m = 30%; n= 45% sample sizes.

Dist.	δ	p = 100			p = 300		
		DR	DREWMA	DRCSEWMA	DR	DREWMA	DRCSEWMA
MVN₀	0	204.3 (203.1)	194.3 (187.5)	199.0 (192.1)	199.8	199.0 (191.1)	200.6 (191.9)
	0.1	196.6 (194.7)	192.6 (184.6)	198.4 (192.8)	182.7	144.1 (139.3)	153.5 (155.6)
	0.2	183.6 (182.2)	162.5 (156.7)	175.8 (175.3)	78.9 (79.5)	25.3 (20.6)	29.3 (28.2)
	0.3	155.2 (155.1)	94.5 (91.1)	113.3 (118.6)	18.6 (18.3)	6.8 (3.7)	6.1 (3.5)
	0.4	111.6 (110.2)	46.0 (41.6)	53.6 (56.6)	4.6 (4.1)	3.3 (1.3)	3.2 (1.6)
	0.5	70.3 (70.1)	22.2 (18.4)	21.3 (17.3)	1.7 (1.1)	2.1 (0.6)	1.4 (0.8)
	0.75	16.9 (16.1)	6.4 (3.5)	5.6 (3.1)	1.0 (0.0)	1.0 (0.2)	1.0 (0.0)
	1	4.7 (4.1)	3.2 (1.3)	2.8 (1.1)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
	EQL	7.653	4.700	4.848	1.079	0.716	0.7036
MVN₁	0	196.3 (198.3)	194.1 (190.4)	201.1 (196.6)	198.7	201.5 (197.7)	201.8 (201.5)
	0.1	199.5 (196.3)	194.7 (191.7)	204.1 (201.3)	191.7	163.1 (161.7)	174.3 (171.0)
	0.2	190.0 (189.6)	174.8 (170.6)	183.1 (182.2)	109.8	39.0 (34.9)	43.0 (37.9)
	0.3	177.7 (175.5)	124.8 (121.5)	140.1 (141.6)	34.9 (34.3)	10.1 (6.5)	10.6 (5.8)
	0.4	138.7 (139.0)	67.7 (63.0)	89.0 (90.3)	9.8 (9.3)	4.5 (2.1)	4.0 (1.9)
	0.5	99.2 (100)	35.9 (31.3)	43.2 (38.3)	3.4 (2.9)	2.7 (1.0)	1.9 (0.5)
	0.75	31.7 (31.4)	9.7 (6.3)	9.1 (6.9)	1.0 (0.2)	1.2 (0.4)	1.0 (0.1)
	1	9.7 (9.2)	4.4 (2.2)	4.0 (2.2)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
	EQL	14.789	6.584	7.291	1.516	0.893	0.854
	0	196.0(194.3)	199.2 (194.2)	199.0 (194.8)	201.3	197.6 (194)	201.2 (202.0)
	0.1	195.6(194.1)	202.3 (198.2)	197.5 (196.5)	198.2	183.8 (182)	191.1 (196.5)
	0.2	195.3(195.3)	195.2 (191.5)	191.0 (192.1)	159.9	86.4 (82.6)	94.7 (95.6)
	0.3	190.9(191.1)	166.4 (161.1)	169.4 (175.4)	86.9 (85.3)	25.8 (21.4)	29.9 (28.5)
	0.4	170.6(170.4)	121.6 (117.4)	118.0 (117.9)	34.6 (34.2)	9.9 (6.4)	8.4 (6.8)

MVT	0.5	144.4(142.3)	78.2 (73.5)	77.1 (75.5)	13.0 (12.6)	5.1 (2.6)	4.1 (3.0)
	0.75	68.8 (67.5)	22.6 (18.6)	20.8 (21.4)	1.9 (1.3)	2.1 (0.7)	1.4 (0.6)
	1	27.6 (26.8)	9.0 (5.8)	8.1 (6.3)	1.0 (0.2)	1.3 (0.4)	1.0 (0.2)
	EQL	24.867	12.150	11.684	3.134	1.600	1.474
MVG	0	199.7(196.0)	196.8 (192.1)	201.8 (198.0)	204.3	203.9 (203.9)	201.4 (200.2)
	0.1	204.2 (202.1)	204.2 (201.2)	205.5 (203.2)	202.5	162.2(158.1)	174.3 (172.6)
	0.2	211.9 (209.1)	192.6 (188)	201.4 (202.0)	119.2	38.2 (33.3)	49.5 (43.0)
	0.3	200.6 (201.6)	134.5 (130.8)	164.3 (168.0)	34.6 (34)	9.5 (5.9)	9.0 (5.6)
	0.4	162.4 (162.3)	71.0 (66.4)	81.6 (86.6)	9.3 (8.8)	4.2 (1.8)	3.7 (1.7)
	0.5	116.9 (116.7)	35.5 (31.1)	41.8 (39.9)	3.0 (2.5)	2.6 (0.9)	1.9 (0.8)
	0.75	34.3 (34)	9.2 (5.8)	8.7 (5.7)	1.0 (0.1)	1.2 (0.4)	1.0 (0.0)
	1	9.6 (9)	4.2 (1.9)	3.6 (2.1)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
	EQL	16.593	6.692	7.297	1.536	0.875	0.861

The first row of each cell shows ARL (SDRL), the second row of each cell shows the value of 25th, 50th, and 75th percentiles, respectively and the last cell in each chart shows the value of the EQL. Bold numbers mean the smallest ARL for each shift in different dimension sizes.

Table 10 ARL(SDRL) and 25th, 50th, and 75th percentiles of the proposed BSCSEWMA and competing charts with m = 30%; n= 30% sample sizes.

Dist.	δ	p = 100			p = 300		
		BS	BSEWMA	BSCSEWMA	BS	BSEWMA	BSCSEWMA
MVN₀	0	199.8 (196.1)	197.2 (194.3)	199.1 (195.7)	200.5 (199.7)	197 (197.2)	200.1 (199.2)
	0.1	202.7 (203.4)	196.9 (192.6)	195.9 (189.4)	170.8 (172.2)	109.4 (105.4)	121.9 (120.1)
	0.2	180.6 (180.4)	144.2 (139.1)	148.3 (151.4)	51.5 (50.1)	14.8 (10.8)	15.6 (14.9)
	0.3	136.0 (138.4)	64.7 (59.6)	78.4 (78.1)	9.0 (8.6)	4.4 (2)	4.0 (2.4)
	0.4	80.8 (79.5)	27.0 (22.6)	33.7 (32.0)	2.3 (1.7)	2.3 (0.8)	1.9 (1.0)
	0.5	42.0 (42.2)	12.7 (8.9)	11.3 (11.5)	1.2 (0.4)	1.6 (0.5)	1.1 (0.4)
	0.75	7.6 (7)	4.1 (1.8)	3.2 (2.0)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
	1	2.1 (1.5)	2.2 (0.8)	2.0 (1.0)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
	EQL	6.610	3.195	3.228	0.812	0.580	0.564
	0	197.6 (197.3)	198.8 (197.3)	200.2 (200.9)	197.5 (193.8)	196.7 (191.8)	201.6 (196.7)
	0.1	199.1 (195.2)	202.3 (198.9)	201.4 (202.7)	188.2 (190.2)	137.8 (130.9)	154.2 (161.3)
	0.2	190.8 (190.2)	169.8 (165.7)	177.9 (176.3)	86.4 (84.4)	24.3 (19.9)	28.6 (26.8)
	0.3	168.7 (164.7)	97.4 (92)	113.5 (118.6)	20.1 (19.3)	6.5 (3.5)	6.8 (4.1)
	0.4	116.8 (118)	45.2 (40.1)	42.1 (40.5)	4.9 (4.3)	3.1 (1.2)	2.5 (1.2)
	0.5	72.8 (73.7)	21.7 (17)	20.2 (23.2)	1.7 (1.1)	2.0 (0.6)	1.4 (0.8)

	0.75	17.5 (17)	5.9 (3.1)	4.9 (2.7)	1.0 (0.0)	1.0 (0.1)	1.0 (0.0)
	1	4.6 (4.1)	3.0 (1.1)	2.3 (1.4)	1.0 (0.0)	1.0 (0)	1.0 (0.0)
	EQL	10.570	4.635	4.458	1.113	0.696	0.697
MVT	0	198.5 (197.6)	196.0 (195.1)	201.2 (196.1)	199.5 (197.4)	200.6 (195.3)	201.8 (200.9)
	0.1	195.9 (195.9)	193.8 (186.6)	204.6 (202.6)	199.5 (199.9)	175.4 (171.3)	185.4 (186.1)
	0.2	191.2 (189.7)	181.5 (179.6)	194.2 (192.6)	137.3 (138.7)	55.2 (49.8)	65.4 (69.1)
	0.3	185.5 (185.4)	139.0 (133.4)	159.5 (161.9)	54.7 (53.3)	14.5 (10.4)	16.2 (14.0)
	0.4	154.3 (151.5)	83.2 (78.1)	104.2 (102.5)	17.0 (16.4)	6.0 (3.1)	5.3 (3.0)
	0.5	119.2 (116.8)	45.8 (41.2)	54.6 (52.6)	5.5 (4.9)	3.4 (1.4)	2.7 (1.6)
	0.75	42.0 (41.2)	12.0 (8.3)	11.9 (11.4)	1.1 (0.4)	1.5 (0.5)	1.1 (0.4)
	1	12.8 (12.1)	5.2 (2.7)	4.8 (3.1)	1.0 (0.0)	1.1 (0.2)	1.0 (0.0)
	EQL	17.820	7.843	8.746	2.033	1.119	1.075
MVG	0	195.0 (196.9)	196.1 (189.7)	200.9 (199.4)	200.5 (201.4)	205.1 (202.5)	199.7 (199.5)
	0.1	198.6 (197.7)	200.4 (198.8)	204.7 (205.9)	187.0 (187.6)	139.9 (135.3)	143.9 (142.2)
	0.2	188.8 (191.5)	167.1 (164.6)	177.9 (185.1)	78.9 (78.1)	22.7 (18.0)	25.9 (23.9)
	0.3	163.6 (164.1)	91.8 (87.4)	111.3 (117.9)	17.7 (17.2)	6.1 (3.2)	6.0 (3.0)
	0.4	115.3 (115.2)	41.1 (36.7)	46.2 (44.2)	4.3 (3.8)	3.0 (1.1)	2.6 (1.0)
	0.5	68.8 (68.2)	19.5 (15.2)	20.1 (20.8)	1.6 (1.0)	1.9 (0.6)	1.4 (0.7)
	0.75	15.2 (14.9)	5.5 (2.8)	5.2 (2.4)	1.0 (0.0)	1.0 (0.1)	1.0 (0.0)
	1	3.9 (3.4)	2.8 (1.0)	2.1 (0.8)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
	EQL	9.905	4.329	3.630	1.066	0.682	0.684

The first row of each cell shows ARL (SDRL), the second row of each cell shows the value of 25th, 50th, and 75th percentiles, respectively and the last cell in each chart shows the value of the EQL. Bold numbers mean the smallest ARL for each shift in different dimension sizes.

Table 11 ARL(SDRL) and 25th, 50th, and 75th percentiles of the proposed BSCSEWMA and competing charts with m = 30%; n= 45% sample sizes.

Dist.	δ	p = 100			p = 300		
		BS	BSEWMA	BSCSEWMA	BS	BSEWMA	BSCSEWMA
	0	204.3 (203.1)	194.3 (187.5)	199.6 (194.6)	199.8 (206.2)	199 (191.1)	199.6 (194.2)
	0.1	196.6 (194.7)	192.6 (184.6)	200.7 (193.9)	182.7 (181.5)	144.1 (139.3)	152.9 (148.6)
	0.2	183.6 (182.2)	162.5 (156.7)	173.5 (176.2)	78.9 (79.5)	25.3 (20.6)	28.4 (24.3)
	0.3	155.2 (155.1)	94.5 (91.1)	112.8 (118.9)	18.6 (18.3)	6.8 (3.7)	5.9 (3.6)
	0.4	111.6 (110.2)	46.0 (41.6)	52.6 (57.3)	4.6 (4.1)	3.3 (1.3)	2.7 (1.1)

MVN₀	0.5	70.3 (70.1)	22.2 (18.4)	20.8 (16.7)	1.7 (1.1)	2.1 (0.6)	1.2 (0.4)
	0.75	16.9 (16.1)	6.4 (3.5)	5.1 (3.2)	1.0 (0.0)	1.0 (0.2)	1.0 (0.0)
	1	4.7 (4.1)	3.2 (1.3)	2.4 (1.1)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
	EQL	7.653	4.700	4.676	1.079	0.716	0.680
MVN₁	0	196.3 (198.3)	194.1 (190.4)	200.4 (197.8)	198.7 (195.7)	201.5 (197.7)	201.6 (201.8)
	0.1	199.5 (196.3)	194.7 (191.7)	204.7 (200.8)	191.7 (189.8)	163.1 (161.7)	169.7 (167.5)
	0.2	190.0 (189.6)	174.8 (170.6)	179.8 (180.7)	109.8 (108.2)	39.0 (34.9)	42.8 (38.3)
	0.3	177.7 (175.5)	124.8 (121.5)	139.0 (137.0)	34.9 (34.3)	10.1 (6.5)	10.0 (7.1)
	0.4	138.7 (139.0)	67.7 (63.0)	87.3 (86.7)	9.8 (9.3)	4.5 (2.1)	3.8 (3.6)
	0.5	99.2 (100)	35.9 (31.3)	40.7 (36.0)	3.4 (2.9)	2.7 (1.0)	1.7 (0.8)
	0.75	31.7 (31.4)	9.7 (6.3)	8.6 (8.2)	1.0 (0.2)	1.2 (0.4)	1.0 (0.1)
	1	9.7 (9.2)	4.4 (2.2)	3.7 (2.1)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
	EQL	14.789	6.584	7.024	1.516	0.893	0.831
MVT	0	196.0 (194.3)	199.2 (194.2)	201.6 (199.9)	201.3 (200.8)	197.6 (194)	199.9 (198.3)
	0.1	195.6(194.1)	202.3 (198.2)	202.0 (202.7)	198.2 (198.7)	183.8 (182)	186.8 (185.7)
	0.2	195.3 (195.3)	195.2 (191.5)	195.4 (201.2)	159.9 (160.4)	86.4 (82.6)	98.3 (96.5)
	0.3	190.9 (191.1)	166.4 (161.1)	175.6 (183.6)	86.9 (85.3)	25.8 (21.4)	30.2 (25.6)
	0.4	170.6 (170.4)	121.6 (117.4)	137.9 (138.0)	34.6 (34.2)	9.9 (6.4)	8.7 (7.2)
	0.5	144.4 (142.3)	78.2 (73.5)	86.9 (94.9)	13.0 (12.6)	5.1 (2.6)	4.6 (3.0)
	0.75	68.8 (67.5)	22.6 (18.6)	22.6 (21.7)	1.9 (1.3)	2.1 (0.7)	1.8 (0.9)
	1	27.6 (26.8)	9.0 (5.8)	8.4 (6.7)	1.0 (0.2)	1.3 (0.4)	1.0 (0.2)
	EQL	24.867	12.150	12.800	3.134	1.600	1.570
MVG	0	199.7 (196.0)	196.8 (192.1)	201.5 (197.6)	204.3 (203.2)	203.9 (203.9)	201.8 (200.6)
	0.1	204.2 (202.1)	204.2 (201.2)	206.5 (204.9)	202.5 (203.6)	162.2 (158.1)	174.9 (182.8)
	0.2	211.9 (209.1)	192.6 (188)	196.8 (191.0)	119.2 (119.1)	38.2 (33.3)	45.6 (40.4)
	0.3	200.6 (201.6)	134.5 (130.8)	164.7 (160.1)	34.6 (34)	9.5 (5.9)	9.1 (5.6)
	0.4	162.4 (162.3)	71.0 (66.4)	104.7 (102.9)	9.3 (8.8)	4.2 (1.8)	3.8 (1.3)
	0.5	116.9 (116.7)	35.5 (31.1)	42.0 (37.2)	3 (2.5)	2.6 (0.9)	2.0 (0.8)
	0.75	34.3 (34.0)	9.2 (5.8)	8.8 (6.3)	1.0 (0.1)	1.2 (0.4)	1.0 (0.0)
	1	9.6 (9.0)	4.2 (1.9)	3.8 (1.7)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
	EQL	16.593	6.692	7.701	1.536	0.875	0.853

The first row of each cell shows ARL (SDRL), the second row of each cell shows the value of 25th, 50th, and 75th percentiles, respectively and the last cell in each chart shows the value of the EQL. Bold numbers mean the smallest ARL for each shift in different dimension sizes.

Table 12 ARL(SDRL) and 25th, 50th, and 75th percentiles of the proposed SDCSEWMA and competing charts with m = 30%; n= 30% sample sizes.

Dist.	δ	p = 100			p = 300		
		SD	SDEWMA	SDCSEWMA	SD	SDEWMA	SDCSEWMA
MVN ₀	0	198.8 (194.7)	191.5 (188.4)	199.4 (195.4)	200.6 (198.9)	197.6 (197.9)	200.3 (199.7)
	0.1	201.0 (200.2)	189.2 (185.8)	198.3 (191.7)	171.6 (171.5)	109.9 (104.7)	132.8 (127.6)
	0.2	181.2 (179.9)	139.4 (132.6)	154.4 (154.7)	52.1 (51.3)	15.0 (11.0)	18.4 (14.8)
	0.3	139.8 (140.4)	63.9 (59.3)	84.3 (82.4)	9.4 (8.9)	4.5 (2.1)	4.2 (2.5)
	0.4	85.1 (85.1)	27.2 (23.2)	35.5 (31.7)	2.4 (1.8)	2.4 (0.8)	2.0 (1.0)
	0.5	45.9 (46.1)	13.4 (9.6)	12.2 (11.6)	1.2 (0.5)	1.6 (0.5)	1.1 (0.4)
	0.75	9.1 (8.6)	4.3 (2)	3.8 (2.5)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
	1	2.6 (2.1)	2.4 (0.9)	2.2 (0.6)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
	EQL	7.158	3.248	3.485	0.821	0.584	0.589
MVN ₁	0	199.0 (199.2)	196.9 (194.6)	199.2 (198.0)	197.1 (194.4)	204.0 (199.7)	199.3 (194.0)
	0.1	199.8 (198.6)	198.6 (195.7)	202.9 (199.2)	186.7 (189.8)	144.9 (137.3)	158.7 (156.1)
	0.2	189.7 (192)	169.5 (166.6)	184.5 (183.6)	87.1 (86.6)	25.4 (21.0)	33.2 (26.4)
	0.3	170.6 (165.5)	98.5 (92.5)	117.6 (124.6)	20.5 (19.7)	6.7 (3.7)	7.2 (4.1)
	0.4	121.0 (119.9)	46.9 (42)	53.8 (59.6)	5.1 (4.5)	3.2 (1.2)	3.0 (1.5)
	0.5	76.6 (77.7)	22.7 (18)	21.0 (18.3)	1.8 (1.2)	2.1 (0.6)	1.5 (0.8)
	0.75	20.2 (19.7)	6.3 (3.4)	5.8 (3.1)	1.0 (0.0)	1.0 (0.2)	1.0 (0.0)
	1	5.8 (5.3)	3.2 (1.3)	2.8 (1.1)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
	EQL	11.346	4.792	4.944	1.145	0.715	0.735
MVT	0	199.1 (198.7)	211.8 (209.2)	199.6 (197.3)	201.3 (205.4)	206.9 (203.3)	199.1 (198.5)
	0.1	195.6 (196.9)	209.3 (204.5)	202.6 (201.3)	196.7 (196.7)	186.4 (183.4)	191.6 (188.2)
	0.2	190.2 (188.8)	194.3 (194.2)	192.9 (191.3)	133.4 (133.1)	57.5 (52.5)	62.4 (61.7)
	0.3	182.5 (182.2)	145.6 (139.1)	157.6 (161.5)	53.3 (52.0)	14.7 (10.9)	15.7 (13.2)
	0.4	151.9 (149)	86.9 (82.1)	97.3 (107.6)	17.0 (16.1)	6.0 (3.1)	5.1 (3.6)
	0.5	117.3 (116.5)	45.9 (41.5)	52.5 (53.0)	5.7 (5.2)	3.4 (1.4)	2.1 (1.6)
	0.75	42.7 (42.4)	12.3 (8.5)	11.6 (11.3)	1.2 (0.5)	1.6 (0.6)	1.0 (0.0)
	1	14.2 (13.6)	5.4 (2.8)	4.4 (3.1)	1.0 (0.1)	1.0 (0.2)	1.0 (0.0)
	EQL	17.940	8.100	8.427	2.025	2.482	1.021
	0	197.1 (197.4)	209.6 (210.6)	199.0 (196.1)	196.6 (195.6)	203.2 (200.6)	201.7 (200.8)
	0.1	198.5 (198.1)	175.9 (167.1)	203.4 (205.1)	182.9 (181.7)	140.4 (136.7)	143.5 (142.8)
	0.2	184.7 (184.1)	132.0 (133.4)	158.1 (156.4)	73.3 (72.4)	21.7 (17.2)	22.9 (23.1)
	0.3	155.4 (154.3)	66.7 (63)	95.6 (93.1)	16.4 (15.9)	5.9 (3.0)	5.4 (3.7)

MVG	0.4	101.8 (101.5)	28.6 (24.3)	36.1 (34.5)	3.9 (3.3)	2.9 (1.0)	2.1 (0.8)
	0.5	58.7 (58.3)	14.9 (11.5)	16.3 (13.2)	1.5 (0.9)	1.9 (0.6)	1.0 (0.5)
	0.75	13.2 (12.7)	4.7 (2.3)	4.5 (2.2)	1.0 (0.0)	1.0 (0.1)	1.0 (0.0)
	1	3.7 (3.2)	2.6 (0.9)	1.8 (0.8)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
	EQL	8.851	3.399	3.844	1.017	0.675	0.626

The first row of each cell shows ARL (SDRL), the second row of each cell shows the value of 25th, 50th, and 75th percentiles, respectively and the last cell in each chart shows the value of the EQL. Bold numbers mean the smallest ARL for each shift in different dimension sizes.

Table 13 ARL(SDRL) and 25th, 50th, and 75th percentiles of the proposed SDCSEWMA and competing charts with m = 30%; n= 45% sample sizes.

Dist.	δ	p = 100			p = 300		
		SD	SDEWMA	SDCSEWMA	SD	SDEWMA	SDCSEWMA
MVN₀	0	199.7 (201.7)	195.3 (190.8)	200.3 (195.2)	199.4 (197.2)	196.4 (190.1)	199.4 (193.6)
	0.1	196.4 (195.3)	193.6 (184.8)	202.8 (194.2)	177.4 (176.6)	139.2 (135.3)	161.3 (155.0)
	0.2	182.8 (181.7)	165.1 (160.1)	178.7 (178.3)	79.5 (79.4)	25.0 (20.7)	33.1 (28.9)
	0.3	157.0 (156.8)	97.5 (94.3)	109.9 (119.8)	19.2 (18.9)	6.8 (3.8)	6.6 (3.7)
	0.4	116.3 (116.5)	48.1 (43.5)	56.3 (60.0)	4.7 (4.2)	3.3 (1.3)	3.4 (1.6)
	0.5	73.4 (73.9)	23.2 (19.2)	22.7 (18.8)	1.8 (1.2)	2.1 (0.7)	1.8 (0.9)
	0.75	18.9 (18.7)	6.9 (4.0)	6.5 (3.8)	1.0 (0.0)	1.0 (0.2)	1.0 (0.0)
	1	5.7 (5.1)	3.5 (1.4)	3.0 (1.1)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
	EQL	10.782	4.923	5.089	1.087	0.710	0.751
MVN₁	0	195.2 (195.3)	197.4 (193.6)	199.8 (197.7)	197.8 (195.5)	192.2 (189.9)	199.2 (200.0)
	0.1	200.4 (198.3)	197.0 (193.7)	204.7 (201.0)	189.9 (188.8)	160.5 (155.6)	175.8 (170.0)
	0.2	189.5 (189)	178.8 (174.7)	188.7 (189.0)	109.1 (108.3)	40.1 (35.8)	51.2 (42.5)
	0.3	171.3 (170)	127.7 (121.7)	145.4 (146.1)	35.0 (34)	10.3 (6.7)	12.1 (8.3)
	0.4	139.3 (136.2)	69.9 (65.6)	102.5 (101.5)	10.1 (9.5)	4.5 (2.1)	4.1 (2.0)
	0.5	101.3 (100.2)	37.9 (33.5)	46.3 (51.8)	3.4 (2.9)	2.7 (1)	2.3 (1.2)
	0.75	34.5 (34.1)	10.4 (7.1)	8.9 (8.3)	1.0 (0.2)	1.3 (0.4)	1.0 (0.2)
	1	11.6 (11.3)	4.8 (2.4)	4.2 (2.1)	1.0 (0)	1.0 (0.0)	1.0 (0.0)
	EQL	15.462	6.900	7.710	1.517	0.911	0.921
	0	194.8 (193.2)	197.4 (192.0)	199.6 (195.4)	200.5 (202.5)	195.3 (191)	201.2 (198.9)
	0.1	195.0 (194.2)	198.3 (193.1)	198.4 (195.9)	199.9 (195.7)	181.3 (177.1)	194.8 (192.5)
	0.2	194.8 (193.4)	192.3 (189.6)	192.7 (191.5)	156.9 (159.9)	83.5 (79.4)	104.8 (102.5)
	0.3	187.7 (192.7)	158.9 (154.8)	173.2 (170.1)	81.5 (81.9)	25.1 (20.9)	32.2 (27.3)
	0.4	167.1(163.7)	115.3 (111.3)	135.3 (132.8)	32.3 (32.1)	9.5 (6.2)	9.1 (6.6)

MVT	0.5	142.3 (140.1)	73.9 (69.8)	83.9 (82.3)	12.2 (11.8)	5.1 (2.6)	4.5 (2.4)
	0.75	67.4 (66.5)	21.8 (17.7)	21.5 (20.4)	2.0 (1.4)	2.1 (0.7)	1.6 (0.7)
	1	28.2 (27.3)	8.9 (5.6)	8.3 (6.6)	1.1 (0.3)	1.3 (0.4)	1.0 (0.2)
	EQL	24.565	11.653	12.424	3.030	1.574	1.596
MVG	0	201.7 (202.5)	195.4 (188.9)	200.3 (197.0)	201.3 (202.1)	196.6 (196.1)	199.6 (195.4)
	0.1	204.4 (202)	199.1 (197.8)	204.6 (203.9)	200.5 (201.3)	157.6 (151.4)	169.3 (178.6)
	0.2	209.9 (209.8)	183.3 (177.5)	196.4 (194.4)	114.4 (116)	35.6 (31.1)	44.2 (40.5)
	0.3	189.9 (188.4)	119.2 (117.1)	155.5 (152.6)	32.3 (32.4)	9.1 (5.7)	8.7 (7.0)
	0.4	152.7 (152.9)	61.5 (56.8)	77.4 (83.2)	8.5 (8.0)	4.1 (1.8)	3.4 (1.6)
	0.5	103.8 (103.2)	30.1 (25.6)	32.2 (37.3)	2.9 (2.3)	2.5 (0.9)	1.7 (1.1)
	0.75	29.0 (28.8)	8.3 (5.2)	7.8 (5.4)	1.0 (0)	1.2 (0.4)	1.0 (0.0)
	1	8.7 (8.3)	4.0 (1.8)	3.3 (1.4)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
	EQL	14.903	5.973	6.546	1.477	0.850	0.818

The first row of each cell shows ARL (SDRL), the second row of each cell shows the value of 25th, 50th, and 75th percentiles, respectively and the last cell in each chart shows the value of the EQL. Bold numbers mean the smallest ARL for each shift in different dimension sizes.

The results of the proposed charts and the competing charts are given in Tables 8-0 indicate the following:

- **Comparative Analysis with the conventional Shewhart charts.**

Under all the distribution structures MVN_0 , MVN_1 , MVT , and MVG , with $p = 100$ and $p = 300$, for both combinations of sample sizes (i.e., $(m = n = p \times 30\%)$ and $(m = p \times 30\%, n = p \times 45\%)$), the proposed DRCSEWMA, BSCSEWMA, and SDCSEWMA charts have shown superior detection ability compared to the DR, BS, and SD (cf. **Tables. 8-13**), as their values of ARL_1 for all the shift types are less than that of the conventional charts. For example, to detect a 0.2 shift under the MVN_0 structure with $(m = n = p \times 30\%)$ sample size and $p = 100$ the ARL_1 for the DR chart is 180.6 samples. In contrast, the ARL_1 for DRCSEWMA it is only 149.7 samples.

- **Comparative Analysis with the conventional EWMA charts.**

For all the distribution structures, with $p = 100$ for both combinations of sample sizes (i.e., $(m = n = p \times 30\%)$ and $(m = p \times 30\%, n = p \times 45\%)$), the proposed DRCSEWMA, BSCSEWMA, and SDCSEWMA charts have shown better detection ability than the DREWMA, BSEWMA, and SDEWMA for the moderate to large shifts (i.e., 0.5, 0.75, and 1.0) (cf. **Tables. 8-13**), as their values of ARL_1 are less than those in the DREWMA, BSEWMA, and SDEWMA charts; otherwise, the DREWMA, BSEWMA, and SDEWMA have better detection ability than the proposed charts. For example, to detect a 0.75 shift under the MVN_0 structure with $(m = n = p \times 30\%)$ sample size and $p = 100$ the ARL_1 for the DREWMA chart is 4.1 samples. In contrast, the ARL_1 for the DRCSEWMA chart is only 3.3 samples. But when we compare them under the $p = 300$ for both combinations of sample sizes (i.e., $(m = n = p \times 30\%)$ and $(m = p \times 30\%, n = p \times 45\%)$), the proposed DRCSEWMA, BSCSEWMA, and SDCSEWMA charts have shown better detection ability than the DREWMA, BSEWMA, and SDEWMA for the moderate to large shifts (i.e., 0.3, 0.4, 0.5, 0.75, and 1.0) (cf. **Tables. 8-13**), as their values of ARL_1 are less than those in the DREWMA, BSEWMA, and SDEWMA charts. For example, to detect a 0.3 shift under the MVN_0 structure with $(m = n = p \times 30\%)$ sample size and $p = 300$ the ARL_1 for the DREWMA chart is 4.4 samples. In contrast, the ARL_1 for the DRCSEWMA chart, it is only 4.0 samples.

- **Comparative Analysis for the proposed charts based on different sample sizes.**

Under all distribution environments with $p = 100$ and $p = 300$ and equal sample size choice $(m = n = p \times 30\%)$, the proposed DRCSEWMA, BSCSEWMA, and SDCSEWMA charts have shown superior detection ability compared to their performance under unequal sample size $(m = p \times 30\%, n = p \times 45\%)$ (cf. **Tables. 8-13**). as their

values of ARL_1 with equal sample size choice ($m = n = p \times 30\%$) are less than the values of the ARL_1 under unequal sample size ($m = p \times 30\%, n = p \times 45\%$). For example, to detect a 0.2 shift under the MVN_0 structure with ($m = n = p \times 30\%$) sample size and $p = 100$ the ARL_1 for the DRCSEWMA chart it is 149.7 samples. In contrast, to detect the same amount of the shift with unequal sample size ($m = p \times 30\%, n = p \times 45\%$) the ARL_1 for DRCSEWMA chart is 175.8 samples.

- **Comparative Analysis for the proposed charts based on different dimension values.**

Under the $p = 300$ with ($m = n = p \times 30\%$) or ($m = p \times 30\%, n = p \times 45\%$) combination sample sizes, the proposed DRCSEWMA, BSCSEWMA, and SDCSEWMA control charts have shown superior detection ability compared to their performance under $p = 100$. As their values of ARL_1 under $p = 300$ choice are less than the values of the ARL_1 under $p = 300$ choice. For example, to detect a 0.2 shift under the MVN_0 structure with ($m = n = p \times 30\%$) sample size and $p = 100$ the ARL_1 for the DRCSEWMA chart, it is 149.7 samples. In contrast, to detect the same amount of the shift with the same sample size ($m = n = p \times 30\%$) and with $p = 300$ the ARL_1 for the DRCSEWMA chart it is only 16.6 samples. (cf. **Tables. 8-13**).

- **Comparative Analysis between the proposed control charts.**

Under the distribution structures (i.e., MVN_0 and MVN_1) with $p = 100$ and $p = 300$ for the different combinations of sample sizes (i.e., ($m = n = p \times 30\%$) and ($m = p \times 30\%, n = p \times 45\%$)), the proposed BSCSEWMA chart exhibits super detection performance compared to the assessed charts. Furthermore, the DRCSEWMA chart surpasses the SDCSEWMA chart. (cf. **Tables. 8-13**), as it offers lower ARL_1 values for various shifts. For example, to detect a 0.2 shift under the MVN_0 structure with

($m = n = p \times 30\%$) sample size and $p = 100$ the ARL_1 for the SDCSEWMA chart it is 154.4 samples. And the ARL_1 for the DRCSEWMA chart it is 149.7 samples, while the ARL_1 for the BSCSEWMA chart is only 148.3 samples. While for the *MVT* distribution structure with $p = 100$ and $p = 300$, for the different combinations of sample sizes, i. e., ($m = n = p \times 30\%$) and ($m = p \times 30\%, n = p \times 45\%$), the proposed DRCSEWMA chart demonstrates better detection performance compared to the assessed charts. Furthermore, the SDCSEWMA chart exceeds the BSCSEWMA chart. For example, to detect a 0.2 shift under the *MVT* structure with ($m = n = p \times 30\%$) sample size and $p = 100$ the ARL_1 for the BSCSEWMA chart is 194.2 samples. And the ARL_1 for SDCSEWMA chart is 192.9 samples, while the ARL_1 for DRCSEWMA chart is only 188.1 samples. (cf. **Tables. 8-13**). For the *MVG* distribution structure, the proposed SDCSEWMA exhibits the better detection ability among the assessed charts. Furthermore, the DRCSEWMA chart surpasses the BSCSEWMA chart. For example, to detect a 0.2 shift under the *MVG* structure with ($m = n = p \times 30\%$) sample size and $p = 100$ the ARL_1 for the BSCSEWMA chart it is 177.9 samples. And the ARL_1 for DRCSEWMA chart is 175.0 samples, while the ARL_1 for the SDCSEWMA chart is only 158.1 samples. (cf. **Tables. 8-13**).

▪ **Comparative Analysis for the effect of the parameter λ on the proposed charts.**

The parameter value λ in the proposed control charts plays a crucial role in detecting shifts in the process means. The increasing values of λ has an inverse relation to the performance of the proposed charts, making it less sensitive to small shifts in the process mean. For example, the proposed SDCSEWMA chart with equal sample sizes ($m = n = p \times 30\%$), $p = 100$ and under MVN_0 structure provides lowest ARL_1 values when $\lambda =$

0.25 compared to the performance of the same chart with the same parameters for different λ values 0.50, 0.75, and 1.0. **Figure 7** illustrates the effect of varying the λ values on the ARL_1 for the proposed charts. Hence, the lower choice of λ , makes the proposed control charts more efficient in detecting the different types of shifts. For example, to detect a 0.2 shift with parameter $\lambda = 0.25$ the ARL_1 for the SDCSEWMA chart is 155.5 samples, with parameter $\lambda = 0.50$ the ARL_1 for the SDCSEWMA chart is 176.6 samples, with parameter $\lambda = 0.75$ the ARL_1 for the SDCSEWMA chart is 181.9 samples. And with parameter $\lambda = 1.0$ the ARL_1 for the SDCSEWMA chart is 180.6 samples.

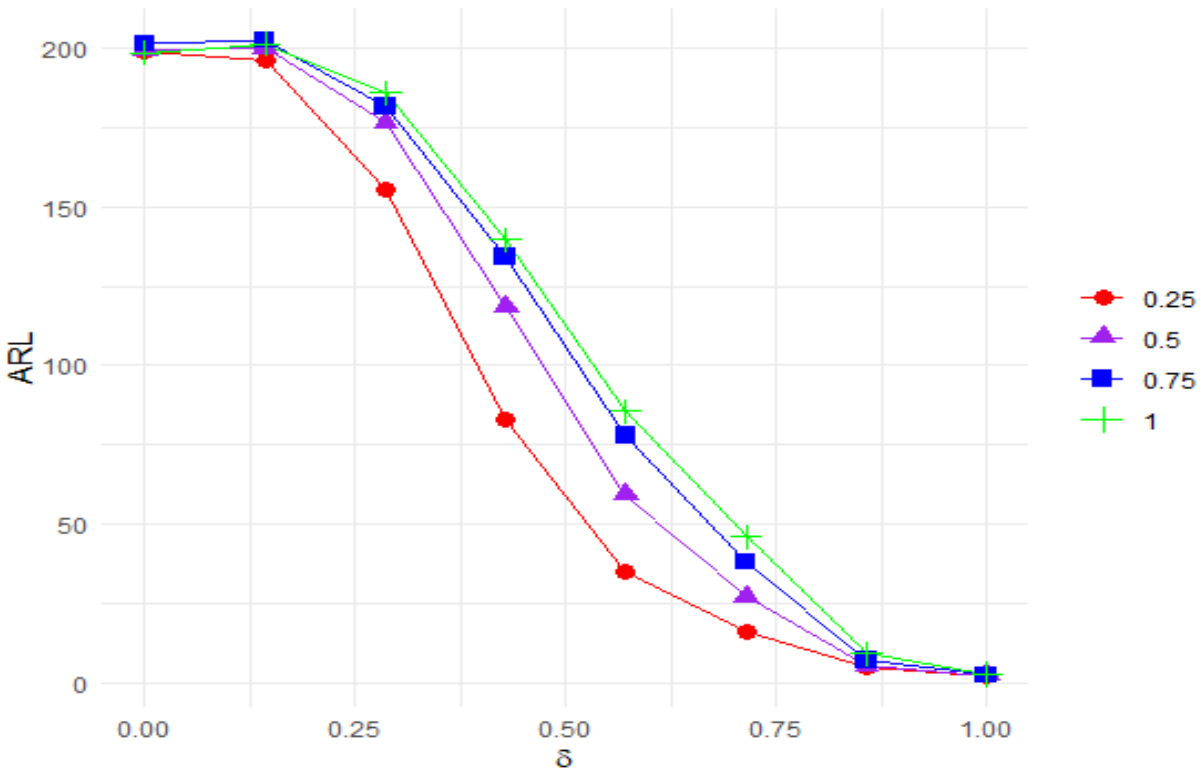


Figure 7 Effect of the λ values on the performance of the SDCSEWMA chart

3.5. Real-Life Application

This section presents a practical application demonstrating the execution of the proposed charts. This application illustrates the actual utilization of these charts for monitoring gas sensor arrays under flow modulation, where real-time responses from high-dimensional data are essential.

Gas Sensor Array Monitoring in Flow Modulation Systems

An electronic nose, a type of gas sensor, may serve as a viable option for a robotic olfactory system capable of detecting and differentiating various gas molecules. Gas sensors utilizing flow modulation are frequently utilized in safety-critical sectors, such as chemical manufacturing, where precise identification of hazardous or combustible gases is essential, see [74] and [75]. The most challenging issue in chemical sensing is still sensor drift. A gas sensor array is an advanced detection system comprising several gas sensors; each intended to identify a particular gas or vapor in a specific environment. The number of sensors in the sensor array varies based on how a case needs to be analyzed and generated [76]. This section describes how our suggested DRCSEWMA, BSCSEWMA, SDCSEWMA, and the competing charts are used in practice to track changes in the means of two-sample high-dimensional data. For this purpose, a gas sensor array under flow modulation obtained by [77] is used. The dataset contains 896 rows (a measurement per sensor), 100 (we consider only the last 100 columns with non-zero values) columns (features or dimensions), and 896 rows (one measurement per sensor). Raw data on each sample consists of 16 time series, one for each sensor. The twelve tested gas classes formed a relatively wide range of two analytes, acetone, and ethanol, in binary mixtures. For five minutes, each time series was recorded at a 25 Hz (samples per second) sampling rate, producing 100 data points for each time series. We categorized the four gases, acetone, air, ethanol, and a mixture, into two groups

according to their chemical nature; group 1 with 448 samples for the organic solvents (Acetone, Ethanol) because both acetone and ethanol are organic solvents, volatile, polar, and flammable. They are frequently used in chemical reactions and as cleaning agents. They are similar in forming vapors that can mix with air, creating flammable mixtures. Group 2 has 448 samples for non-flammable gases (air, mixture) because the air itself is not flammable but can support combustion. Although potentially flammable, a mixture of acetone and ethanol with air contains air as a key component. This group focuses on the presence of air, which distinguishes it from the pure organic solvents in group 1. Using the 896 samples 448 for group 1 and 448 for group 2, each sample contains 16 sensors with 100 features per sensor. The implementation of these proposed DRCSEWMA, BSCSEWMA, and SDCSEWMA charts was conducted in three steps in the initial step using 20,000,000 iterations, the values of the mean and the standard deviations of T_{1t} , T_{2t} , and T_{3t} statistics are obtained from our gas sensor array monitoring in flow modulation systems data. Subsequently, in the second step, using predefined $ARL_0 = 200$ the $L_{S1}, L_{S2}, L_{S3}, L_{S4}, L_{S5}, L_{S6}, L_{e1}, L_{e2}, L_{e3}, L_{e4}, L_{e5}$, and L_{S6} coefficient values are found using 10,000 iterations and the pre-obtained means and standard deviations. Finally, considering the values of means, standard deviations, and the values of $L_{S1}, L_{S2}, L_{S3}, L_{S4}, L_{S5}, L_{S6}, L_{e1}, L_{e2}, L_{e3}, L_{e4}, L_{e5}$, and L_{S6} shown in Error! Reference source not found.. We have calculated the plotting statistics for the proposed DRCSEWMA, BSCSEWMA, and SDCSEWMA charts using equations (3.13), (3.15), and (3.17). This will give us 28 plotting statistics, Shewhart-based and 28 plotting statistics EWMA-based for each proposed chart from our gas sensor array monitoring in flow modulation systems data.

Table 14 The mean, standard deviation, and coefficient limits of the proposed and competing charts for Gas Sensor Array Monitoring in Flow Modulation Systems at a fixed $ARL_0 = 200$

Classical Shewhart charts for two-sample high-dimensional data					
DR		BS		SD	
μ_{T_1}	19.11305	μ_{T_2}	13.07246	μ_{T_3}	3.940614
σ_{T_1}	23.02136	σ_{T_2}	16.65367	σ_{T_3}	5.050718
L_{s_1}	0.82994857	L_{s_3}	0.82787	L_{s_5}	0.82608
L_{s_2}	3.15889785	L_{s_4}	3.2610	L_{s_6}	3.32930
Classical EWMA charts for two-sample high-dimensional data					
DREWMA		BSEWMA		SDEWMA	
μ_{T_1}	19.11305	μ_{T_2}	13.07246	μ_{T_3}	3.940614
σ_{T_1}	23.02136	σ_{T_2}	16.65367	σ_{T_3}	5.050718
L_{e_1}	1.72701	L_{e_3}	1.72640	L_{e_5}	1.74650
L_{e_2}	3.12580	L_{e_4}	3.22579	L_{e_6}	3.22577
Proposed CSEWMA charts for two-sample high-dimensional data					
DRCSEWMA		BSCSEWMA		SDCSEWMA	
μ_{T_1}	19.11305	μ_{T_2}	13.07246	μ_{T_3}	3.940614
σ_{T_1}	23.02136	σ_{T_2}	16.65367	σ_{T_3}	5.050718
L_{s_1}	0.829950	L_{s_3}	0.82790	L_{s_5}	0.82610
L_{s_2}	3.291377	L_{s_4}	3.34901	L_{s_6}	3.33540
L_{e_1}	1.80805	L_{e_3}	1.8050	L_{e_5}	1.8030
L_{e_2}	3.53901	L_{e_4}	3.5482	L_{e_6}	3.5380

Bootstrapping is a statistical resampling technique that involves repeatedly sampling with replacement from the original data to estimate the distribution of a sample statistic. When the

sample size is small, or the population distribution is unknown, it is beneficial for determining the reliability of an estimate (e.g. mean, median, or variance) [73]. The bootstrap technique is used to generate 150 IC data samples. We start with our original 28 samples and then randomly select 28 samples with replacements from the original sample and repeat this resampling process 150 times. This procedure gives us 150 samples. For the IC data, plotting statistics are calculated for each set of samples. Similarly, we used the same technique to generate 50 samples for the OOC data by introducing a 0.05 shift amount in the second sample (i.e., $x_{ij2t} + \delta$) and plotting statistics for the proposed DRCSEWMA, BSCSEWMA, and SDCSEWMA, and the competing charts were calculated for each set of samples. **Figure 8-10** shows the 100 IC and 50 OOC test statistic values of the proposed and the competing charts, and their control limits.

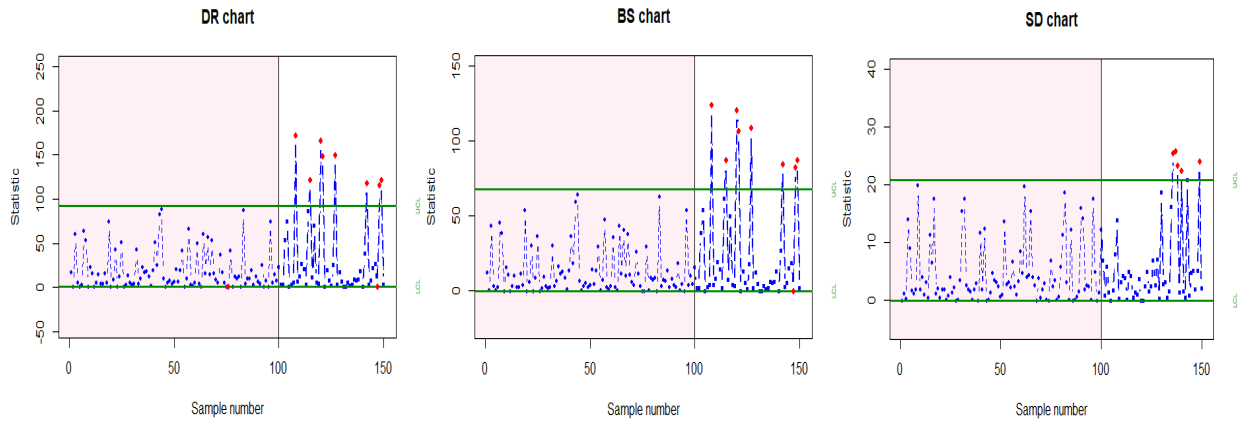


Figure 8 Conventional DR, BS and SD charts for Gas Sensor Array Monitoring in Flow Modulation Systems

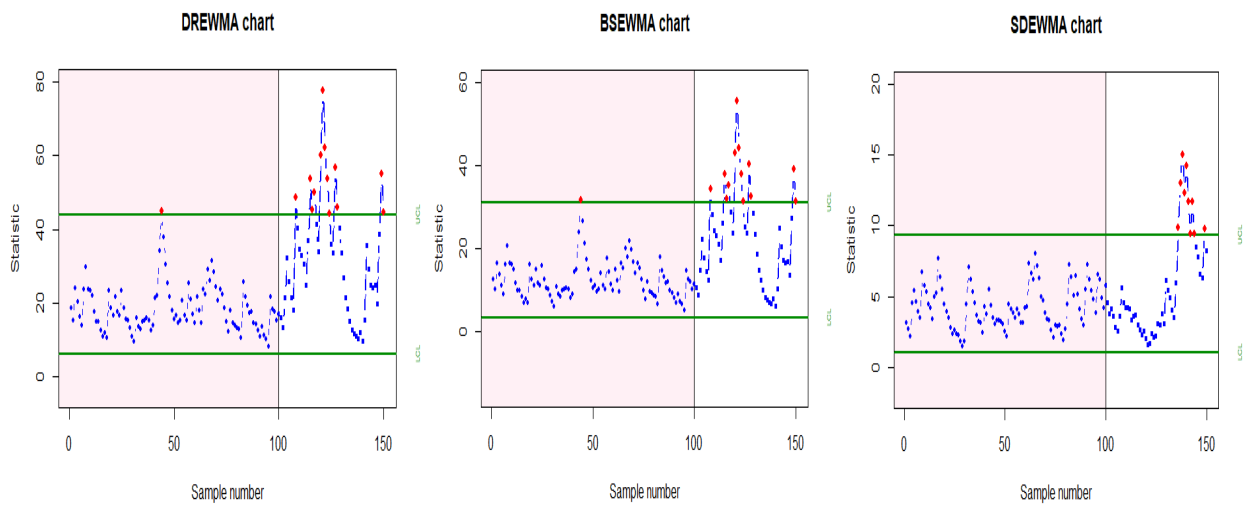
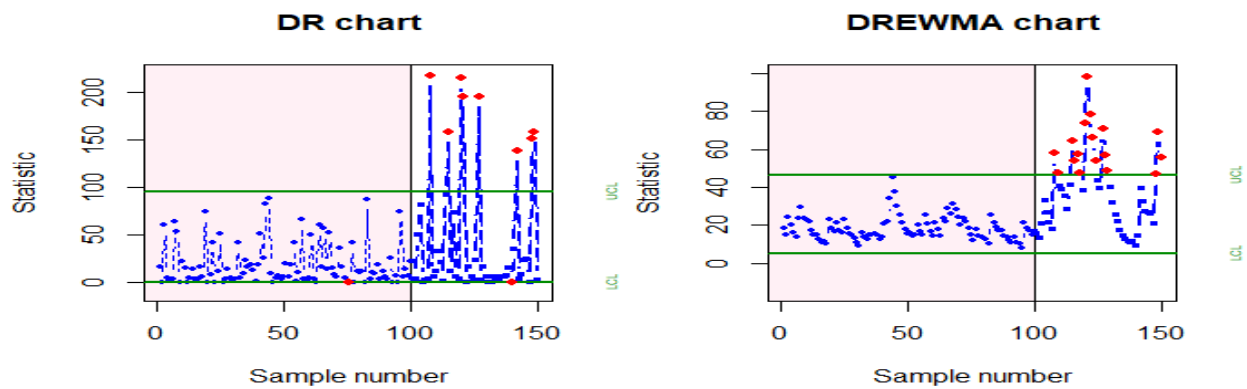


Figure 9 Conventional DREWMA, BSEWMA and SDEWMA charts for Gas Sensor Array Monitoring in Flow Modulation Systems

DRCSEWMA Chart



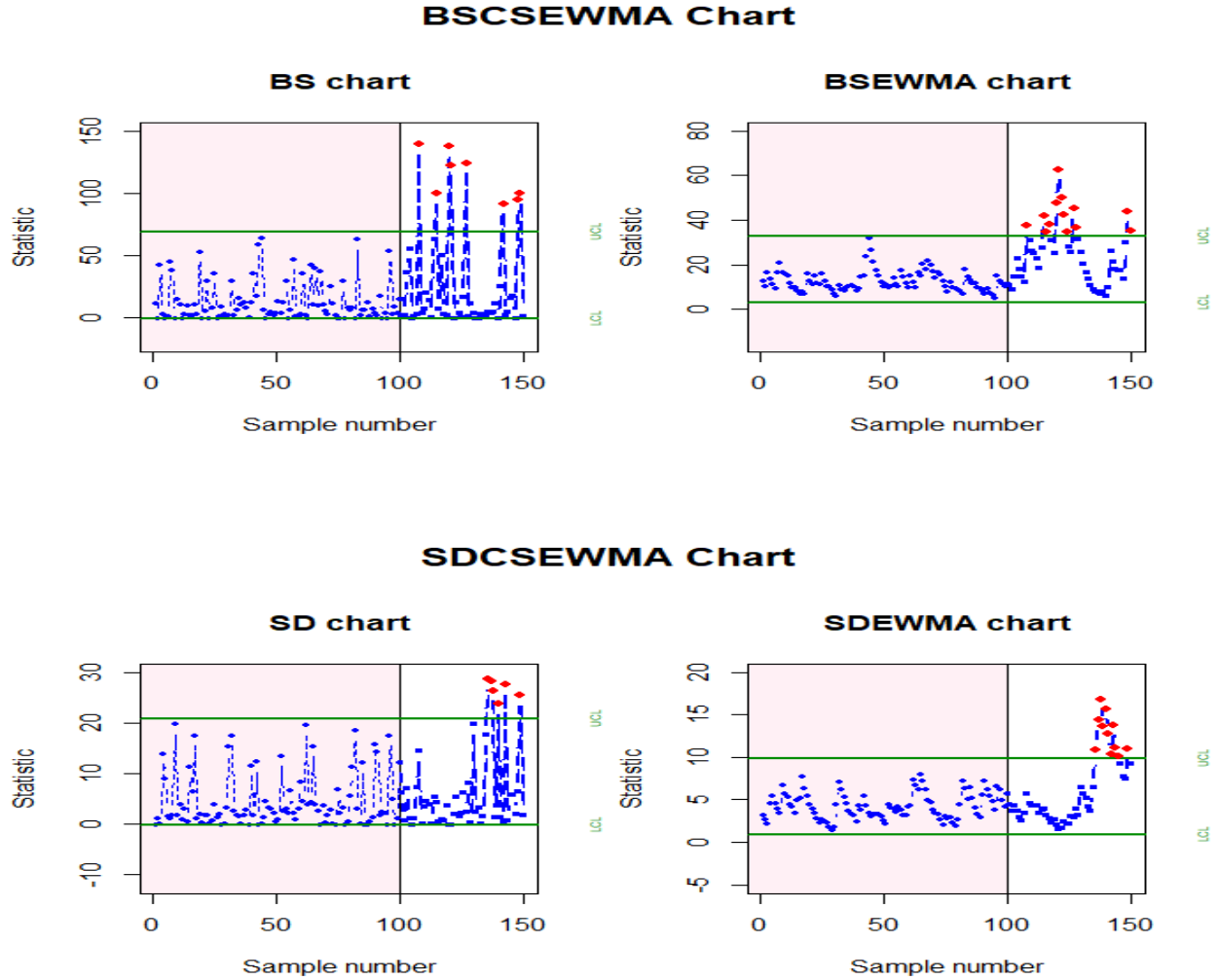


Figure 10 Proposed DRCSEWMA, BSCSEWMA and SDCSEWMA charts for Gas Sensor Array Monitoring in Flow Modulation Systems

The IC and OOC values are separated by a vertical line in each of the plots, where the IC region is presented in pink color while the white area is used as the OOC region, and the blue points represent the plotting statistics for the proposed and competing charts. Two horizontal green lines in each plot represent UCL and LCL for the corresponding charts. The plotting statistics of the proposed and competing charts are shown in blue when they are in IC state and red points when they are in OOC state. For each competing chart, we can note that DR and BS charts have the same performance as it was also proved by (Mahmood et al., 2024) that the DR and BS have the same average run length

profile. And both of them outperformed the SD. We note the same things for the second competing chart, which shows that DREWMA and BSEWMA have the same performance. [104] also proved in his thesis that DREWMA and BSEWMA have the same average run length profile, and both of them have better detection ability. If we compare our proposed charts with competing charts, for example, the DR chart detected only 8 points OOC while the DREWMA chart detected 13 OOC points; at the same time, our proposed DRCSEWMA chart outperformed both of them by detecting 26 OOC points. For the comparison between the three proposed charts, the DRCSEWMA chart demonstrates better detection performance compared to the assessed charts; when the process is under control, there is only one point that is out of control limits, and this point, we can consider it as false alarm has been detected. However, the DRCSEWMA detects 26 OOC points when the process is OOC. Furthermore, the BSCSEWMA chart exceeds the SDCSEWMA chart since it detected 21 OOC points. At the same time, SDCUSUM detected only 17 OOC points. Hence, the DRCSEWMA chart is more efficient in studying and monitoring the behavior of the gas sensor arrays under flow modulation.

CHAPTER 4

CSCUSUM CONTROL CHARTS FOR THE MEANS OF TWO

SAMPLED HIGH-DIMENSIONAL DATA

4.1. Introduction

High-dimensional data refers to datasets that consist of a substantially larger number of features p than observations N , providing distinct difficulties and opportunities, high-dimensional data can be found in almost all sectors. The complexity of high-dimensional data requires sophisticated analytical approaches to derive significant insights, as conventional statistical procedures frequently falter because of the “curse of dimensionality” [78]. High-dimensional data often refers to a singular sample population. For example, Industrial Datasets, Medical Science Datasets, Chemical Datasets, Image Datasets, Solar Datasets, and Miscellaneous datasets [78]. In a population with two samples, certain high-dimensional processes generate high-dimensional data. For example, grease damage in bearings of wind turbines dataset [23].

Control charts are an essential tool in SPC, providing a graphical depiction of process performance across time. Control charts have been extensively utilized for process monitoring across the manufacturing, healthcare, business, and finance sectors. This extensive applicability underlines the necessity to devise novel methodologies and improve current frameworks to meet the distinct difficulties and demands of each sector. However, the control charts can be categorized based on the structure of control charts into two memoryless and memory-type control charts. This aids in

selecting the most suitable chart for diverse process monitoring requirements and particular applications [1]. Control charts can also be categorized into two based on their structure according to how historical information is handled. There are memoryless and memory-based charts that are vital tools in SPC, each fulfilling specific roles in the assessment of process performance. Memoryless charts, shown by traditional Shewhart charts, are simple to construct and interpret, making them suitable for various applications. They depend exclusively on the latest data points. Hence, they are effective at detecting large, abrupt shifts in the process mean [5, 6]. Conversely, memory-based charts, such as Cumulative Sum (CUSUM) [7], are designed to be more sensitive to small, amounts of the shifts in the process means than Shewhart charts. These charts incorporate historical data to detect subtle changes that might not be immediately apparent in a Shewhart chart. For the large shifts and lower dimensions the conventional memoryless charts based on the Shewhart structure are the suitable option [23]. As dimensionality rises, and if we are interested in detecting the small shifts the CUSUM charts appear as an effective tool for detecting anomalies resulting from minor modifications in the process. A Shewhart control chart is proficient in identifying significant process adjustments, however it exhibits diminished sensitivity to minor shifts. A CUSUM chart is more effective than a Shewhart chart in identifying minor process shifts, however it is less adept at recognizing significant shifts. While the Combined Shewhart-CUSUM (CSCUSUM) effectively monitors both minor and significant alterations [107-109]. In the literature on SPC, the presence of memoryless control charts, such as the Shewhart control chart, and memory control charts, such as CUSUM, for high-dimensional single-sample data is prevalent. Therefore, we divided the literature review for the single-sample high-dimensional data based on the control charts a structure to three parts [78].

Control charts (i) utilizing dimension reduction techniques, (ii) utilizing variable selection techniques, and (iii) utilizing high-dimensional approaches are comprehensively reviewed as follows.

Dimension reduction techniques are the methods used to reduce the features or the dimensions while retaining critical information for monitoring process performance, thereby enhancing control chart efficiency by eliminating redundant or less useful features in high-dimensional data. There are different types of dimension reduction approaches, for instance, Principal Component Analysis (PCA) [15, 16], Kernel Principal Component Analysis (KPCA) [17]. [110] They developed a distribution-free CUSUM approach for monitoring location and scale in high-dimensional processes. This approach addresses the limitations presented by conventional parametric methods, especially in situations where the underlying distribution remains unidentified. This novel method utilizes nonparametric statistics to efficiently identify changes in both location and scale, exhibiting strong performance across many dimensions and distributions. [51] assessed the implementation of multistate multivariate statistical process control (MS-MSPC) as an advanced method for monitoring and regulating processes across many states. This approach improves conventional SPC by including multivariate data analysis, facilitating a deeper understanding of process variability and quality monitoring. [111] investigated the utilization of variational autoencoders (VAEs) for the surveillance of high-dimensional nonlinear processes, highlighting their efficacy in collecting intricate data patterns. This method is especially pertinent in industrial environments where conventional monitoring strategies may fail because of the complexities of the data. [112] studied a Variational Autoencoder-Long Short Term Memory (VAE-LSTM) model for fault detection in high-dimensional, time-dependent processes, overcoming the constraints of current methods that presume independence and normal distribution, thereby enhancing

monitoring efficacy and minimizing false alarms. [46] presented reduced-dimension control charts for spatio-temporal surveillance, tackling challenges associated with high-dimensional data by using dimension reduction during data collecting. It illustrates efficacy by quantitative outcomes and a case study on water quality monitoring, emphasizing the advantages of implementation.

Variable selection techniques are the methods that are used for selecting the most informative variables within high-dimensional data. These strategies improve the efficacy of control charts by concentrating on pertinent variables [19]. There are different types of variable selection methods, for example, the Least Absolute Shrinkage and Selection Operator based (LASSO-based) procedure [20], and the forward-selection technique [21]. [113] introduced an innovative approach for detecting fault variables in high-dimensional processes. It presents an adaptive step-down technique using the conditional T^2 statistic, concentrating on the selection of variables exhibiting the least indication of change at each stage. This method efficiently utilizes the relationship between fault and non-fault variables, improving diagnostic performance relative to current techniques. The research illustrates enhanced recognition of altered components in the mean vector. [54] introduced a variable selection-based multivariate cumulative sum (VS-MCUSUM) control chart to enhance the identification of minor variations in high-dimensional process parameters. This method overcomes the limitations of conventional statistical techniques, surpassing traditional MCUSUM and VS-multivariate moving average control charts in simulation and numerical analyses. [99] developed a variable-selection control chart using a Gaussian mixture model to effectively monitor multimodal processes, enhancing fault diagnosis and process monitoring efficiency.

High-dimensional control charts utilize advanced techniques to monitor the process changes in high-dimensional data setups without the need for dimensionality reduction or variable selection

techniques. For instance, [102] addressed the issue of identifying anomalous observations in high-dimensional situations, when conventional techniques such as Hotelling's T^2 statistic prove to be ineffective. They introduced an innovative method that regularizes the covariance matrix and employs a leave-one-out resampling technique to enhance detection power while preserving acceptable false positive rates. [114] proposed a subdivided approach for tracking high-dimensional normally distributed multi-stream processes, involving a single Hotelling chi-square chart in the initial phase, followed by a second stage that augments monitoring efficiency in complex data contexts. [115] presented an adaptive sampling technique for the online monitoring of high-dimensional streaming data, emphasizing the collection of partial information. It incorporates recursive projection for data imputation and an adaptive framework to improve data collecting, hence improving change detection capabilities. [12] introduced an innovative dynamic network technique for monitoring high-dimensional picture profiles, tackling the limitations of conventional SPC methods. It illustrates efficient online control charts for monitoring complex picture structures in biomanufacturing and machining concepts.

The Memoryless and memory type control charts discussed in the preceding literature review mostly focus on monitoring single-sample high-dimensional data; nevertheless, there is an absence of research concerning two-sample high-dimensional data. Mahmood, et al. [23] recently introduced Shewhart-type control charts for monitoring the means of two-sampled high-dimensional data. Their proposed charts, grounded in the Shewhart framework, simply consider current process information; hence, they are less adept at identifying small variations in the means of two-sample high-dimensional data. Consequently, [104] in his master's thesis has devised memory setup control charts for tracking the means of two-sampled high-dimensional data, utilizing the CUSUM chart framework, which effectively incorporates both current and historical

information in the process and is proficient in identifying shifts in the means of two-sampled high-dimensional data. However, there are currently no available charts to track the small to large variations in the means of two-sampled high-dimensional data at the same time. This article aims to introduce a combination of Shewhart and CUSUM (CSEWMA) charts for monitoring small to large shifts in the means of two-sampled high-dimensional data.

The proposed DRCSCUSUM, BSCSCUSUM, and SDCSCUSUM charts are designed based on the two-sample high-dimensional means tests named [27], [67], and [68]. The proposed charts are assessed according to run length metrics across various distributional structures, dimensions, and sample sizes. The real-life applications of these charts are additionally illustrated by a case study on daily stock prices.

The remaining part of the chapter is as follows. In Section 4.2, it describes the development of high-dimensional charting techniques. Section 4.3 describes the framework of the simulation study. In Section 4.4, the performance evaluation and comparative analysis of the proposed and the competing charts are presented, and an illustrative application example is provided in Section 4.5. The concluding remarks are reported in Section 4.6.

4.2. Development of High-Dimensional Charting Techniques

This section overviews the two-sample high-dimensional mean tests and control charting procedures applied to high-dimensional data. These procedures are utilized for two-sample setups.

4.2.1. Design Structure of High-Dimensional Data Two Sample Testing Procedure

For the two-sample high-dimensional case, x_{ijlt} , $i = 1, 2, \dots, m$ or $i = 1, 2, \dots, n$, $j = 1, 2, \dots, p$, $l = 1, 2$ and $t = 1, 2, 3, \dots$, is the p -dimensional, t^{th} multivariate normal random

vectors with mean vectors $\mu_{0l} = (\mu_1, \dots, \mu_p)'$; $l = 1, 2$ and unknown common positive variance-covariance matrices Σ_{0l} ; $l = 1, 2$. The sizes of the first and second samples are denoted by m and n , respectively; p is the size of the dimension, l is the sample number, and t represents time. The t^{th} statistics T_{1t} , T_{2t} and T_{3t} proposed by Dempster [27], Bai and Saranadasa [67] and Srivastava and Du [68], respectively, for the means testing of two-sample high-dimensional data are defined as follows:

$$T_{1t} = \frac{\frac{mn}{m+n}(\bar{x}_{1t} - \bar{x}_{2t})'(\bar{x}_{1t} - \bar{x}_{2t})}{tr S_t}, \quad (4.1)$$

$$T_{2t} = \frac{\frac{mn}{m+n}(\bar{x}_{1t} - \bar{x}_{2t})'(\bar{x}_{1t} - \bar{x}_{2t}) - tr S_t}{\left[\frac{2(m+n)(m+n+1)}{(m+n-1)(m+n+2)} \left(tr S_t^2 - \frac{(tr S_t)^2}{m+n} \right) \right]^{\frac{1}{2}}}, \quad (4.2)$$

$$T_{3t} = \frac{\frac{mn}{m+n}(\bar{x}_{1t} - \bar{x}_{2t})' D_{S_t}^{-1} (\bar{x}_{1t} - \bar{x}_{2t}) - \frac{(m+n)p}{m+n-2}}{\left[2 \left(tr R_t^2 - \frac{p^2}{m+n} \right) c_{p,m+n} \right]^{\frac{1}{2}}}. \quad (4.3)$$

Where the t^{th} sample means the vector of the first sample is $\bar{x}_{1t} = \sum_{i=1}^m x_{ij1t}/m$ and t^{th} sample means the vector of the second sample is $\bar{x}_{2t} = \sum_{i=1}^n x_{ij2t}/n$. The t^{th} diagonal matrix (D_{S_t}) of sample variance-covariance matrix (S) is defined by $D_{S_t} = diag(s_{11}, \dots, s_{pp})$, the t^{th} sample correlation matrix (R_t) is defined by $R_t = D_{S_t}^{-\frac{1}{2}} S_t D_{S_t}^{-\frac{1}{2}}$ and the adjustment coefficient $c_{p,m+n}$ can be obtained by $c_{p,m+n} = 1 + \frac{tr R_t^2}{p^{3/2}}$.

4.2.2. Classical Shewhart Setup Charts for Two-Sample High-Dimensional Data

To track the means of the two-sample high-dimensional data, Mahmood, et al. [23] developed DR, BS, and SD charts to monitor the means of the two-sample high-dimensional data. When referring to the DR chart, the t^{th} statistic T_{1t} , which are described in equation (4.1), are taken into

consideration as the plotting statistic against the limitations. These limits are obtained by using the following procedure, based on a certain false alarm rate α :

$$\begin{cases} LCL_{(DR)} = \mu_{0T_{1t}} - L_{s1} \sigma_{0T_{1t}} \\ UCL_{(DR)} = \mu_{0T_{1t}} + L_{s2} \sigma_{0T_{1t}} \end{cases} \quad (4.4)$$

For the BS chart, the t^{th} statistic T_{2t} presented in equation (4.2) is considered as the plotting statistic against the limits as follows:

$$\begin{cases} LCL_{(BS)} = \mu_{0T_{2t}} - L_{s3} \sigma_{0T_{2t}} \\ UCL_{(BS)} = \mu_{0T_{2t}} + L_{s4} \sigma_{0T_{2t}} \end{cases} \quad (4.5)$$

For the SD chart, the t^{th} statistic T_{3t} presented in equation (4.3) is considered as the plotting statistic against the limits as follows:

$$\begin{cases} LCL_{(SD)} = \mu_{0T_{3t}} - L_{s5} \sigma_{0T_{3t}} \\ UCL_{(SD)} = \mu_{0T_{3t}} + L_{s6} \sigma_{0T_{3t}} \end{cases} \quad (4.6)$$

It is important to note that all three charts, which are referred to as DR charts, BS charts, and SD charts, indicate an out-of-control (OOC) point when the values of T_{1t} , T_{2t} and T_{3t} fall outside of the control limits that are specified in equations (4.4), (4.5), and (4.6) respectively. Otherwise, the process is considered to be an in-control (IC) state.

4.2.3. Classical CUSUM Setup Charts for Two-Sample High-Dimensional Data

The classical CUSUM control charts are developed based on T_{1t} , T_{2t} and T_{3t} Statistics are given in equations (4.1) -(4.3), and named DRCUSUM, BSCUSUM, and SDCUSUM charts respectfully. For the plotting statistics of the DRCUSUM chart, we will integrate the test statistics given in equation (4.1) into CUSUM statistics as follows:

$$\begin{aligned} C_{t(DRCUSUM)}^- &= \min [0, (T_{1t} - \mu_{T_1}) + K \sigma_{T_1} + C_{t-1(DRCUSUM)}^-], \\ C_{t(DRCUSUM)}^+ &= \max [0, (T_{1t} - \mu_{T_1}) - K \sigma_{T_1} + C_{t-1(DRCUSUM)}^+], \end{aligned} \quad (4.7)$$

and control limits of the DRCUSUM chart can be obtained as follows:

$$\begin{aligned} LCL_{(DRCUSUM)} &= -h_1 \sigma_{T_1}, \\ UCL_{(DRCUSUM)} &= h_2 \sigma_{T_1}. \end{aligned} \quad (4.8)$$

It is to be noted that μ_{T_1} and σ_{T_1} are the mean and standard deviation of T_{1t} statistic, while h_1 and h_2 the control limit coefficient for the DRCUSUM chart. For the plotting statistics of the BSCUSUM chart, we will integrate the test statistics given in equation (4.2) into CUSUM statistics as follows:

$$\begin{aligned} C_{t(BSCUSUM)}^- &= \min [0, (T_{2t} - \mu_{T_2}) + K \sigma_{T_2} + C_{t-1(BSCUSUM)}^-], \\ C_{t(BSCUSUM)}^+ &= \max [0, (T_{2t} - \mu_{T_2}) - K \sigma_{T_2} + C_{t-1(BSCUSUM)}^+], \end{aligned} \quad (4.9)$$

where μ_{T_2} and σ_{T_2} are the mean and standard deviation of T_{2t} statistic, while h_3 and h_4 the control limit coefficient for the BSCUSUM chart. The control limits can be obtained as follows:

$$\begin{aligned} LCL_{(BSCUSUM)} &= -h_3 \sigma_{T_2}, \\ UCL_{(BSCUSUM)} &= h_4 \sigma_{T_2}. \end{aligned} \quad (4.10)$$

For the plotting statistics of the SDCUSUM chart, we will integrate the test statistics given in equation (4.3) into CUSUM statistics as follows:

$$\begin{aligned} C_{t(SDCUSUM)}^- &= \min [0, (T_{3t} - \mu_{T_3}) + K \sigma_{T_3} + C_{t-1(SDCUSUM)}^-], \\ C_{t(SDCUSUM)}^+ &= \max [0, (T_{3t} - \mu_{T_3}) - K \sigma_{T_3} + C_{t-1(SDCUSUM)}^+], \end{aligned} \quad (4.11)$$

where μ_{T_3} and σ_{T_3} are the mean and standard deviation of T_{3t} statistic, while h_5 and h_6 the control limit coefficient for the SDCUSUM chart. The control limits can be obtained as follows:

$$\begin{aligned} LCL_{(SDCUSUM)} &= -h_5 \sigma_{T_3}, \\ UCL_{(SDCUSUM)} &= h_6 \sigma_{T_3}. \end{aligned} \quad (4.12)$$

The initial values of all CUSUM statistics in equations (4.7), (4.9), and (4.11) are set equal to zero, i.e. $C_{0(DRCUSUM)}^\pm = C_{0(BSCUSUM)}^\pm, C_{0(SDCUSUM)}^\pm = 0$, and K is the reference value, which is usually chosen as half of the mean shift of the process. In the context of decision rules, a specific chart, such as DRCUSUM, BSCUSUM, and SDCUSUM, is classified as being in an IC situation if both chart statistics in equations (4.7), (4.9), and (4.11) remain within the specified limits outlined in equations (4.8), (4.10), and (4.12); otherwise, it will be considered to be in an OOC situation.

4.2.4. Proposed CUSUM Setup Charts for Two-Sample High-Dimensional Data

The CSCUSUM control charts are powerful statistical techniques utilized for process monitoring to identify shifts in process parameters efficiently. These charts combine the benefits of both Shewhart and CUSUM charts, rendering them extremely effective in detecting small to large variations in the process parameters [108, 116, 117]. The Shewhart chart is recognized for its rapid identification of large shifts because it takes care only of the current information about the process, but the CUSUM chart is proficient at detecting minor and moderate shifts because it takes care of the current and the past information about the process [1]. The integration of these two methodologies results in CSCUSUM charts, which offer an effective solution for process monitoring, guaranteeing the timely detection of shifts for small to large-magnitude [107]. The Shewhart charts efficiently detect large shifts in the means of two-sample high-dimensional data [23]. Since our objective is to monitor the small to large shifts in the means of two-sample high-dimensional data at the same time, therefore, CSCUSUM is a good option.

For the suggested CSCUSUM based on the statistic T_{1t} in equations (4.1) calculated from two-sample high-dimensional data is firstly integrated into CUSUM statistics and secondly for the DR chart, the t^{th} statistic T_{1t} identified in equation (4.1) is considered as a plotting statistic for the DR chart. These test statistics are considered plotting statistics for our proposed CSCUSUM chart.

$$\begin{cases} \{DR_t = T_{1t} \\ C_{t(DRCUSUM)}^- = \min [0, (T_{1t} - \mu_{T_1}) + K \sigma_{T_1} + C_{t-1(DRCUSUM)}^-], \\ C_{t(DRCUSUM)}^+ = \max [0, (T_{1t} - \mu_{T_1}) - K \sigma_{T_1} + C_{t-1(DRCUSUM)}^+]. \end{cases} \quad (4.13)$$

The lower control limits (LCL) and upper control limits (UCL) for the proposed chart are the combined limits of the $C_{t(DRCUSUM)}^-$, $C_{t(DRCUSUM)}^+$, and the DR_t statistics, for a given false alarm α , can be obtained as follows:

$$\begin{cases} \begin{cases} LCL_{(DR)} = \mu_{0T_{1t}} - L_{s1} \sigma_{0T_{1t}}, \\ UCL_{(DR)} = \mu_{0T_{1t}} + L_{s2} \sigma_{0T_{1t}}. \end{cases} \\ \begin{cases} LCL_{(DRCUSUM)} = -h_1 \sigma_{T_1}, \\ UCL_{(DRCUSUM)} = h_2 \sigma_{T_1}. \end{cases} \end{cases} \quad (4.14)$$

where, $\mu_{0T_{1t}}$ and $\sigma_{0T_{1t}}$ are the mean and standard deviation of the statistic T_{1t} . The values of L_{s1} , L_{s2} , h_1 and h_2 can be determined empirically (cf. Table 15), and K is the reference value, which is usually chosen as half of the mean shift of the process. It is to be noted that the test statistic T_{1t} was proposed by Dempster [27]. Therefore, the control chart based on the statistic T_{1t} and CSCUSUM is named as DRCSCUSUM chart.

For the suggested CSCUSUM based on the statistic T_{2t} in equations (4.2) calculated from two-sample high-dimensional data is firstly integrated into CUSUM statistics and secondly for the BS chart, the t^{th} statistic T_{2t} presented in equation (4.2) is considered as a plotting statistic for the BS chart. These test statistics are considered plotting statistics for our proposed CSCUSUM chart.

$$\left\{ \begin{array}{l} BS_t = T_{2t} \\ C_{t(BSCUSUM)}^- = \min [0, (T_{2t} - \mu_{T_2}) + K \sigma_{T_2} + C_{t-1(BSCUSUM)}^-] \\ C_{t(BSCUSUM)}^+ = \max [0, (T_{2t} - \mu_{T_2}) - K \sigma_{T_2} + C_{t-1(BSCUSUM)}^+] \end{array} \right. \quad (4.15)$$

The lower control limits (LCL) and upper control limits (UCL) for the proposed chart are the combined limits of the $C_{t(BSCUSUM)}^-$, $C_{t(BSCUSUM)}^+$, and the BS_t statistics, for a given false alarm α , can be obtained as follows:

$$\left\{ \begin{array}{l} LCL_{(BS)} = \mu_{0T_{2t}} - L_{s3} \sigma_{0T_{2t}}, \\ UCL_{(BS)} = \mu_{0T_{2t}} + L_{s4} \sigma_{0T_{2t}}, \\ \left\{ \begin{array}{l} LCL_{(BSCUSUM)} = -h_3 \sigma_{T_2}, \\ UCL_{(BSCUSUM)} = h_4 \sigma_{T_2}. \end{array} \right. \end{array} \right. \quad (4.16)$$

where, $\mu_{0T_{2t}}$ and $\sigma_{0T_{2t}}$ are the mean and standard deviation of the statistic T_{2t} . The values of L_{s3} , L_{s4} , h_3 and h_4 can be determined empirically (cf. Table 15), and K is the reference value, which is usually chosen as half of the mean shift of the process. It is to be noted that the test statistic T_{2t} was proposed by Bai and Saranadasa [67]. Therefore, the control chart based on the statistic T_{2t} and CSCUSUM is named the BSCSCUSUM chart.

For the suggested CSCUSUM based on the statistic T_{3t} in equations (4.3) calculated from two-sample high-dimensional data is firstly integrated into CUSUM statistics and secondly for the SD chart, the t^{th} statistic T_{3t} presented in equation (4.3) is considered as a plotting statistic for the SD chart. These test statistics are considered as plotting statistics for our proposed CSCUSUM chart.

$$\left\{ \begin{array}{l} SD_t = T_{3t} \\ C_{t(SDCUSUM)}^- = \min [0, (T_{3t} - \mu_{T_3}) + K \sigma_{T_3} + C_{t-1(SDCUSUM)}^-] \\ C_{t(SDCUSUM)}^+ = \max [0, (T_{3t} - \mu_{T_3}) - K \sigma_{T_3} + C_{t-1(SDCUSUM)}^+] \end{array} \right. \quad (4.17)$$

The lower control limits (LCL) and upper control limits (UCL) for the proposed chart are the combined limits of the $C_{t(SDCUSUM)}^-$, $C_{t(SDCUSUM)}^+$, and the SD_t statistics, for a given false alarm α , can be obtained as follows:

$$\left\{ \begin{array}{l} LCL_{(SD)} = \mu_{0T_{3t}} - L_{s5} \sigma_{0T_{3t}}, \\ UCL_{(SD)} = \mu_{0T_{3t}} + L_{s6} \sigma_{0T_{3t}}, \\ \left\{ \begin{array}{l} LCL_{(SDCUSUM)} = -h_5 \sigma_{T_3}, \\ UCL_{(SDCUSUM)} = h_6 \sigma_{T_2}. \end{array} \right. \end{array} \right. \quad (4.18)$$

where, $\mu_{0T_{3t}}$ and $\sigma_{0T_{3t}}$ are the mean and standard deviation of the statistic T_{3t} . The values of L_{s5} , L_{s6} , h_5 and h_6 can be determined empirically (cf. Table 15), and K is the reference value, which is usually chosen as half of the mean shift of the process. It is to be noted that the test statistic T_{3t} was proposed by Srivastava and Du [68]. Therefore, the control chart based on the statistic T_{3t} and CSCUSUM is named the SDCSCUSUM chart.

It is noted that all proposed three control charts, names DRCSCUSUM, BSCSCUSUM, and SDCSCUSUM charts signal an OOC point when DR_t , BS_t , and SD_t fall outside of their control limits given in the equations (4.14), (4.16), and (4.18) respectively. Or $C_{t(DRCUSUM)}^-$, $C_{t(DRCUSUM)}^+$, $C_{t(BSCUSUM)}^-$, $C_{t(BSCUSUM)}^+$, $C_{t(SDCUSUM)}^-$ and $C_{t(SDCUSUM)}^+$ fall outside of their control limits given in the equations (4.14), (4.16), and (4.18) respectively. Otherwise, the process is considered an IC situation.

4.3. Simulation Study Framework

This section explores the evaluation criteria utilized to assess the efficiency of the proposed schemes and provides comparisons with existing schemes and outlining the simulation design.

The measures are assessed based on the properties of the Run Length (RL) profile, which is defined as the number of subgroups that must be taken before an OOC signal is given by the scheme. There are two types of measures based on the method of evaluation. The first category involves individual performance measures, which assess the control charts performance to detect specific shifts in the process parameters. The most used measures of individual performance are (a) Average Run Length (ARL); ARL represents the average number of points needed to detect an OOC signal. It is classified into two types as follows: when the process is stable, ARL is denoted as ARL_0 it refers to the zero-state condition which represents the expected number of samples taken before a false alarm occurs when the process is IC. While the ARL_1 refers to the steady state which represents the expected number of samples taken before detecting a process shift when the process is OOC. It is to be noted that a lower ARL_1 indicates faster detection of shifts. (b) Standard Deviation of Run Length (SDRL): SDRL measures variability in RL, providing insight into the consistency of chart performance. It is calculated by $SDRL = \sqrt{E(RL - ARL)^2}$ and the lower SDRL indicates more consistent detection. (c) percentile; The 25th, 50th (median), and 75th percentiles of RL highlight the distribution of run lengths, indicating where the shortest 25%, middle 50%, and longest 25% of RL values lie. The second category, overall performance measures these measures offer a summary of the overall ability of the control chart. For example, Extra Quadratic Loss (EQL); EQL evaluates overall chart performance across all shift magnitudes using a weighted average of ARL over the entire shift domain ($\delta_{min} < \delta < \delta_{max}$) using the square of shift (δ^2) as

the weight. The EQL is calculated by $EQL = \int_{\delta_{min}}^{\delta_{max}} \delta^2 ARL(\delta) d\delta / (\delta_{max} - \delta_{min})$, where lower EQL values indicate better overall chart performance. By following [71] and Mahmood, et al. [23], two independent high-dimensional samples (represented by $l = 1, 2$) are generated through four distributional environments: (i) multivariate normal with mean vector ($\mu_{0l} = 0; l = 1, 2$) and variance-covariance matrix ($\Sigma_{0l}; l = 1, 2$), where unit variances and covariances are equal to 0, denoted as MVN_0 ; (ii) multivariate normal with mean vector ($\mu_{0l} = 0; l = 1, 2$) and variance-covariance matrix ($\Sigma_{0l}; l = 1, 2$), where unit variances and covariances are equal to 0.5, denoted as MVN_1 ; (iii) multivariate t with 5 degrees of freedom, denoted as MVT; (iv) multivariate gamma with shape parameter 5 and scale parameter 1, denoted as MVG. For MVT and MVG, the variance-covariance matrix ($\Sigma_{0l}; l = 1, 2$) considered with unit variances and covariances are equal to 0.5. The size of the dimensions is set equal to $p = 100$ and $p = 300$. The sample size for the two samples is considered as $(m = p \times 30\%, n = p \times 30\%)$ and $(m = p \times 30\%, n = p \times 45\%)$.

The simulation study is designed based on the Monte Carlo simulations using the R language program. The simulation for this study is conducted in a three-step process. In the first step, T_{1t} , T_{2t} and T_{3t} are computed a large number of times say $t = 20,000,000$ for each distributional environment (i-iv), dimensions and sample sizes are explained above. Based on these 20,000,000 statistics means $(\mu_{T_1}, \mu_{T_2}, \mu_{T_3})$ and variances $(\sigma_{T_1}, \sigma_{T_2}, \sigma_{T_3})$ are estimated, which are presented in **Table 15**. In the second step, the IC Average Run Length (ARL_0) is considered equal to 200 (i.e., $ARL_0 = 200$) Under different distributions and dependency structures, values of control limit coefficients for CSCUSUM-based control charts $L_{S1}, L_{S2}, L_{S3}, L_{S4}, L_{S5}, L_{S6}, h_1, h_2, h_3, h_4, h_5$ and h_6 are found using 10,000 iterations, the Shewhart and CUSUM-based charts are running

individually to obtain the coefficients at $ARL_0 = 800$ on one side and then combined Shewhart setup coefficients generate $ARL_0 = 400$ and combined CUSUM setup coefficients generate $ARL_0 = 400$ and overall, CSCUSUM setup coefficients may provide $ARL_0 = 200$. These steps are discussed in **Figure 11**. The obtained control limit coefficients are provided in **Table 15**.

Table 15 Mean, standard deviation, and the coefficients for the proposed control charts at fixed $ARL_0 = 200$

DRSCUSUM	Sample Size		$m = p * 30\%, n = p * 30\%$				$m = p * 30\%, n = p * 45\%$			
	p	Dist.	MVN ₀	MVN ₁	MVT	MVG	MVN ₀	MVN ₁	MVT	MVG
100	μ_{T_1}	Dist.	1.000040309	1.000187538	0.999864439	0.999702409	0.999691981	1.000204864	0.999312465	1.000392461
			0.14255512	0.183380737	0.180988125	0.175502044	0.142201644	0.183483473	0.189333435	0.177122464
			2.592	2.409	2.416	2.427	2.612	2.402	2.378	2.431
			3.393	3.692	3.663	3.654	3.395	3.683	3.711	3.621
			4.370	4.124	4.128	4.171	4.345	4.112	4.064	4.103
			5.211	5.484	5.480	5.491	5.192	5.442	5.516	5.381
	σ_{T_1}	Dist.	0.999657982	0.999674135	1.00021901	0.999803006	0.999757358	0.99991906	1.000442512	1.000122774
			0.081928849	0.10552098	0.105306809	0.101147208	0.081894671	0.106006896	0.110259723	0.101540601
			2.771	2.648	2.656	2.658	2.771	2.618	2.628	2.652
			3.215	3.388	3.358	3.368	3.238	3.365	3.391	3.363
			4.496	4.371	4.418	4.395	4.469	4.357	4.388	4.382
			5.057	5.155	5.112	5.142	5.046	5.135	5.135	5.095
100	μ_{T_2}	Dist.	0.000201319	0.001997831	0.000232192	-0.00057593	-0.002088601	0.001937466	-0.001671207	0.002934977
			1.000052183	1.000572346	0.721134763	0.947208446	0.999206217	1.002318981	0.773646987	0.966375533
			2.621	2.395	2.548	2.432	2.612	2.398	2.465	2.422
			3.392	3.688	3.857	3.667	3.404	3.685	3.805	3.648
			4.371	4.107	4.164	4.165	4.358	4.101	4.082	4.112
			5.209	5.512	5.584	5.507	5.198	5.440	5.572	5.388
	σ_{T_2}	Dist.	-0.004178261	-0.002917642	0.001482241	-0.001723809	-0.002941502	-0.00059306	0.003920036	0.001377601

BSCSCUSUM	300	σ_{T_2}	1.000670089	0.999327802	0.68770646	0.979975699	1.000811142	1.004497301	0.742891509	0.988042273
		LS ₃	2.775	2.675	2.771	2.664	2.775	2.688	2.721	2.642
		LS ₄	3.155	3.391	3.481	3.354	3.228	3.368	3.474	3.365
		h ₃	4.485	4.375	4.442	4.394	4.532	4.342	4.396	4.378
		h ₄	5.028	5.156	5.168	5.122	5.053	5.113	5.167	5.105
SDCSCUSUM	100	μ_{T_3}	-0.0029238	-0.00144175	-0.00678564	-0.015878346	-0.004592064	-0.000692584	-0.006139438	-0.006919155
		σ_{T_3}	0.951397836	0.928676369	0.655156075	0.886857558	0.949071417	0.928764712	0.700265498	0.901689453
		LS ₅	2.606	2.380	2.533	2.442	2.605	2.386	2.475	2.425
		LS ₆	3.441	3.727	3.840	3.645	3.406	3.704	3.806	3.639
		h ₅	4.368	4.094	4.140	4.178	4.338	4.096	4.092	4.452
		h ₆	5.232	5.517	5.581	5.456	5.232	5.440	5.576	5.352
	300	μ_{T_3}	-0.004635957	-0.003499283	0.000133486	-0.005097199	-0.003568896	-0.001140389	0.002107492	-0.001125099
		σ_{T_3}	0.951974795	0.935901972	0.629888276	0.921848875	0.956090162	0.944174072	0.682944042	0.933384975
		LS ₅	2.782	2.628	2.775	2.675	2.762	2.588	2.728	2.658
		LS ₆	3.225	3.356	3.485	3.362	3.215	3.378	3.472	3.346
		h ₅	4.496	4.365	4.345	4.408	4.487	4.349	4.436	4.409
		h ₆	5.042	5.158	5.178	5.128	5.035	5.138	5.161	5.078

In the last step, at a fixed $ARL_0 = 200$, the run length (RL) performance of the proposed charts is calculated by adding the shifts in the second sample (i.e., $x_{ij2t} + \delta$), where shifts are considered equal to 0.1, 0.2, 0.3, 0.4, 0.5, 0.75 and 1. The detection ability of the proposed charts under the above-mentioned shifts is observed under different distributional environments (i-iv), dimensions, and sample sizes using 10,000 iterations.

zes using 10,000 iterations.

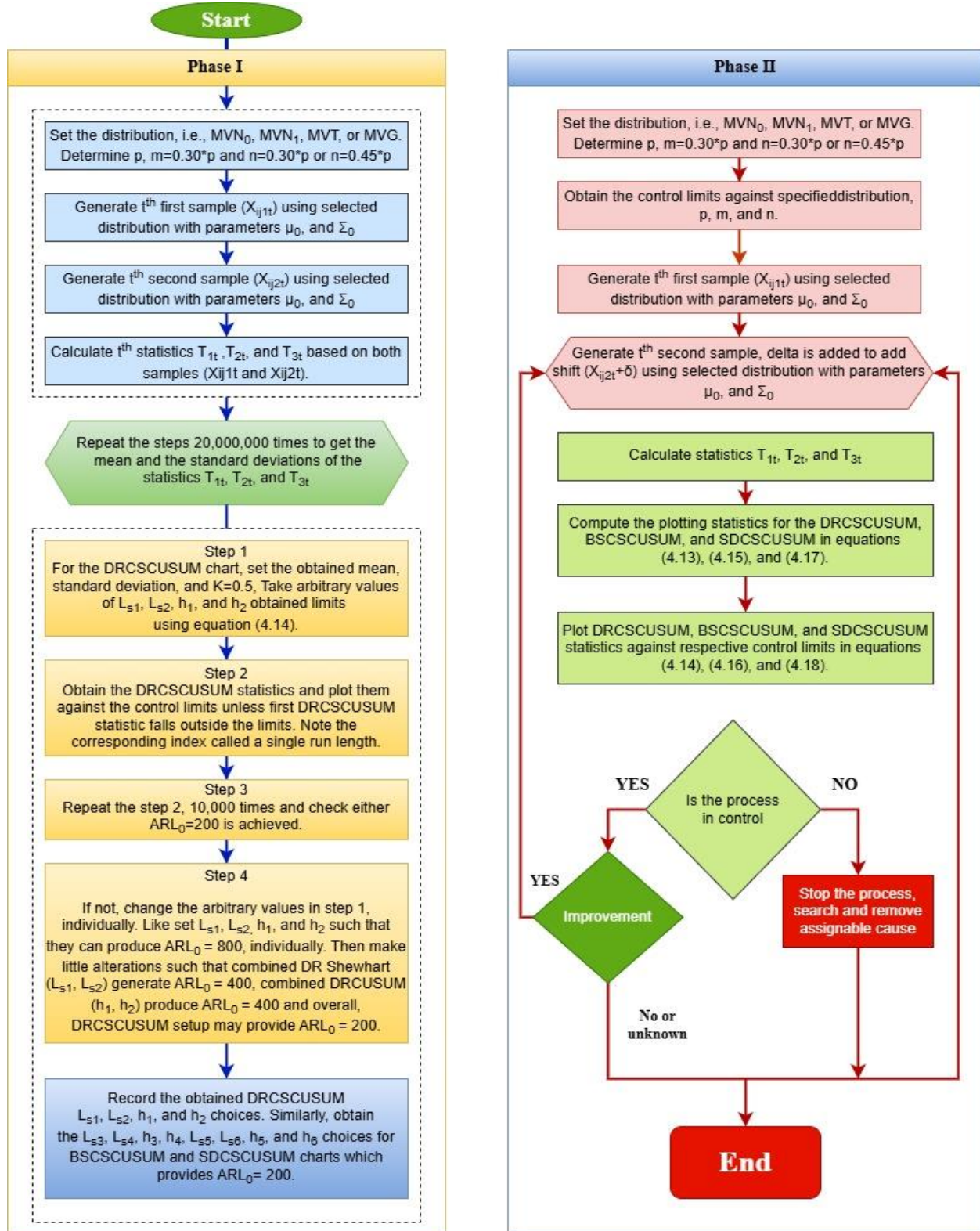


Figure 11 Framework diagram for the implementation of the proposed control charts.

4.4. Performance Evaluation and Comparative Analysis

To assess the efficacy or sensitivity of our proposed and competing control charts, we utilized Monte Carlo simulation for the first run-rule (a point exceeding the control limits). The statistical process monitoring literature extensively advocates for the utilization of characteristics derived from the run-length profile, such as the Average Run Length (ARL), Standard Deviation of Run Length (SDRL), and the percentiles of the run-length variable, which indicates the number of samples required before the scheme generates an OOC signal. The results of all these measures are displayed in **Tables 16-21** for the proposed charts. The proposed charts have been compared with the DR, BS, and SD charts introduced by Mahmood, et al. [23] and the DRCUSUM, BSCUSUM, and SDCUSUM charts presented by [104] their findings have been included in **Tables 16-21**. The objective is to determine whether the suggested DRCSCUSUM, BSCSCUSUM, and SDCSCUSUM charts surpass the traditional Shewhart and EWMA setups using two sampled high-dimensional data. Consequently, the performance of the run length profile is evaluated against the proposed and the computing charts.

Table 16 RL profile for the proposed DRCSCUSUM and competing charts with (m = 30%; n= 30%) sample sizes

Dist.	δ	p = 100			p = 300		
		DR	DRCUSUM	DRCSCUSUM	DR	DRCUSUM	DRCSCUSUM
MVN ₀	0	199.8 (196.1)	199.2 (190.8)	198.8 (192.1)	200.5 (199.7)	200.4 (196.7)	200.5 (198.6)
	0.1	202.7 (203.4)	192.6 (183.9)	191.0 (183.2)	170.8 (172.2)	114.4 (108.5)	123.4 (120.2)
	0.2	180.6 (180.4)	143.7 (135.6)	148.3 (144.1)	51.5 (50.1)	15.5 (11.2)	13.4 (12.7)
	0.3	136.0 (138.4)	67.8 (61.5)	73.9 (71.5)	9.0 (8.6)	4.6 (2.0)	4.1 (2.4)
	0.4	80.8 (79.5)	27.3 (22.4)	29.4 (26.2)	2.3 (1.7)	2.3 (0.7)	2.0 (1.1)
	0.5	42.0 (42.2)	13.2 (8.9)	11.2 (9.8)	1.2 (0.4)	1.5 (0.5)	1.2 (0.4)
	0.75	7.6 (7)	4.3 (1.8)	3.7 (2.2)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
	1	2.1 (1.5)	2.4 (0.76)	2.0 (1.0)	1.0 (0.0)	1.0 (0)	1.0 (0.0)

	EQL	6.610	3.297	3.179	0.812	0.586	0.564
MVN₁	0	197.6 (197.3)	199.7(197.13)	201.2 (198.3)	197.5 (193.8)	199.6 (194.2)	200.7 (194.8)
	0.1	199.1 (195.2)	201.2 (198.9)	202.8 (199.4)	188.2 (190.2)	137.8 (131.0)	144.2 (142.4)
	0.2	190.8 (190.2)	170.1 (161.9)	176.3 (175.9)	86.4 (84.4)	24.0 (19.4)	23.7 (21.0)
	0.3	168.7 (164.7)	100.5 (94.7)	105.6 (101.0)	20.1 (19.3)	6.5 (3.2)	6.0 (3.3)
	0.4	116.8 (118)	46.4 (41.6)	48.5 (44.0)	4.9 (4.3)	3.2 (1.1)	3.0 (1.4)
	0.5	72.8 (73.7)	21.4 (16.6)	20.3 (18.9)	1.7 (1.1)	2.0 (0.6)	1.5 (0.7)
	0.75	17.5 (17)	6.2 (2.9)	6.0 (3.3)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
	1	4.6 (4.1)	3.2 (1.1)	2.8 (1.4)	1.0 (0.0)	1.0 (0)	1.0 (0.0)
	EQL	10.570	4.737	4.716	1.133	0.697	0.672
MVT	0	198.5 (197.6)	201.1 (201.8)	199.2 (196.7)	199.5 (197.4)	199.4 (192.8)	201.8 (200.7)
	0.1	195.9 (195.9)	200.4 (200.4)	201.0 (203.8)	199.5 (199.9)	184.6 (180.4)	184.7 (180.1)
	0.2	191.2 (189.7)	193.1 (192.9)	190.0 (192.8)	137.3 (138.7)	73.5 (68.3)	67.7 (61.2)
	0.3	185.5 (185.4)	159.4 (154.2)	157.1 (153.8)	54.7 (53.3)	17.8 (14.8)	16.7 (12.9)
	0.4	154.3 (151.5)	106.1 (101.0)	99.9 (93.2)	17.0 (16.4)	6.1 (3.7)	6.4 (3.5)
	0.5	119.2 (116.8)	60.3 (56.3)	55.9 (49.7)	5.5 (4.9)	3.2 (1.4)	3.1 (1.5)
	0.75	42.0 (41.2)	13.9 (11.0)	12.1 (9.0)	1.1 (0.4)	1.0 (0.5)	1.0 (0.6)
	1	12.8 (12.1)	5.3 (2.9)	4.9 (2.9)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
	EQL	17.820	9.360	8.732	2.033	1.142	1.126
MVG	0	195.0 (196.9)	200.2 (192.9)	198.9 (195.7)	200.5 (201.4)	200.8 (197.9)	201.4 (198.5)
	0.1	198.6 (197.7)	200.2 (193.6)	203.2 (196.8)	187.0 (187.6)	143.6 (135.9)	148.2 (145.8)
	0.2	188.8 (191.5)	169.6 (164.0)	172.9 (171.0)	78.9 (78.1)	25.0 (23.0)	26.2 (22.9)
	0.3	163.6 (164.1)	98.3 (92.7)	103.6 (101.1)	17.7 (17.2)	6.8 (3.9)	6.1 (3.5)
	0.4	115.3 (115.2)	43.4 (38.3)	44.5 (43.7)	4.3 (3.8)	3.4 (1.5)	3.7 (1.4)
	0.5	68.8 (68.2)	20.1 (15.6)	19.4 (16.8)	1.6 (1.0)	2.2 (0.9)	1.9 (0.8)
	0.75	15.2 (14.9)	5.7 (2.7)	5.5 (3.0)	1.0 (0.0)	1.3 (0.4)	1.0 (0.0)
	1	3.9 (3.4)	2.9 (1.0)	2.5 (1.3)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
	EQL	9.905	4.501	4.473	1.066	1.314	0.715

The first row of each cell shows ARL (SDRL), the second row of each cell shows the value of 25th, 50th, and 75th percentiles, respectively and the last cell in each chart shows the value of the EQL. Bold numbers mean the smallest ARL for each shift in different dimension sizes.

Table 17 RL profile for the proposed DRCSCUSUM and competing charts with (m = 30%; n = 45%) sample sizes

Dist.	δ	p = 100			p = 300		
		DR	DRCUSUM	DRCSCUSUM	DR	DRCUSUM	DRCSCUSUM
MVN ₀	0	204.3 (203.1)	201.4(195.2)	199.7 (195.8)	199.8 (206.2)	199.2 (193.7)	198.8 (192.4)
	0.1	196.6 (194.7)	198.6(188.7)	199.3 (193.8)	182.7 (181.5)	142.9 (136.6)	151.4 (147.5)
	0.2	183.6 (182.2)	168.2 (164.6)	173.1 (169.6)	78.9 (79.5)	26.3 (21.2)	27.2 (24.6)
	0.3	155.2 (155.1)	101.6 (96.6)	110.1 (105.4)	18.6 (18.3)	7.1 (3.6)	6.7 (4.4)
	0.4	111.6 (110.2)	48.7 (42.9)	51.6 (44.8)	4.6 (4.1)	3.4 (1.2)	3.2 (1.1)
	0.5	70.3 (70.1)	23.6 (18.7)	23.1 (20.9)	1.7 (1.1)	2.1 (0.6)	1.6 (0.9)
	0.75	16.9 (16.1)	6.7 (3.4)	6.3 (3.8)	1.0 (0.0)	1.0 (0.2)	1.0 (0.0)
	1	4.7 (4.1)	3.4 (1.3)	3.3 (1.6)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
	EQL	10.153	4.965	5.017	1.079	0.724	0.707
MVN ₁	0	196.3 (198.3)	200.1 (194.6)	201.7 (198.2)	198.7 (195.7)	199.4 (192.1)	201.8 (201.5)
	0.1	199.5 (196.3)	199.4 (194.2)	200.6 (195.5)	191.7 (189.8)	162.4 (155.5)	170.3 (167.0)
	0.2	190.0 (189.6)	181.1 (175.5)	186.0 (182.9)	109.8 (108.2)	41.7 (36.0)	43.0 (37.9)
	0.3	177.7 (175.5)	129.7 (125.5)	136.9 (134.0)	34.9 (34.3)	10.7 (6.6)	10.2 (5.8)
	0.4	138.7 (139.0)	74.1 (68.5)	72.2 (68.2)	9.8 (9.3)	4.5 (2.0)	4.0 (1.9)
	0.5	99.2 (100)	38.0 (32.5)	36.6 (34.9)	3.4 (2.9)	2.1 (0.6)	1.9 (0.5)
	0.75	31.7 (31.4)	10.0 (5.9)	9.4 (6.8)	1.0 (0.2)	1.2 (0.3)	1.0 (0.1)
	1	9.7 (9.2)	4.7 (2.1)	4.4 (2.4)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
	EQL	14.789	6.933	6.805	1.516	0.941	0.846
MVT	0	196.0(194.3)	199.4 (195.1)	201.3 (199.8)	201.3 (200.8)	199.7 (193.3)	199.7 (195.1)
	0.1	195.6 (194.1)	198.4 (195.1)	205.5 (203.1)	198.2 (198.7)	192.6 (190.8)	191.7 (189.7)
	0.2	195.3 (195.3)	194.6 (193.9)	194.8 (192.1)	159.9 (160.4)	104.3 (100.7)	101.7 (98.5)
	0.3	190.9 (191.1)	172.7 (168.2)	176.0 (171.8)	86.9 (85.3)	31.4 (28.2)	29.9 (24.4)
	0.4	170.6 (170.4)	133.6 (125.8)	132.8 (133.6)	34.6 (34.2)	10.7 (7.5)	10.0 (6.9)
	0.5	144.4 (142.3)	89.4 (85.6)	87.4 (84.9)	13.0 (12.6)	5.1 (2.7)	4.8 (2.8)
	0.75	68.8 (67.5)	26.5 (22.7)	25.1 (21.8)	1.9 (1.3)	1.8 (0.7)	1.6 (0.9)
	1	27.6 (26.8)	9.6 (6.6)	9.4 (6.3)	1.0 (0.2)	1.1 (0.3)	1.0 (0.2)
	EQL	24.867	13.506	13.221	3.134	1.691	1.671
MVG	0	199.7(196.0)	201.7(196.0)	201.7 (196.2)	204.3 (203.2)	200.3 (196.9)	200.3 (196.6)
	0.1	204.2 (202.1)	207.6(200.6)	204.5 (200.8)	202.5 (203.6)	167.6 (165.9)	173.7 (170.6)
	0.2	211.9 (209.1)	193.8 (188.3)	198.1 (195.2)	119.2 (119.1)	39.0 (33.0)	41.8 (38.7)
	0.3	200.6 (201.6)	139.9 (134.0)	141.5 (138.2)	34.6 (34)	9.9 (5.9)	9.6 (6.6)
	0.4	162.4 (162.3)	77.7 (71.3)	79.7 (77.4)	9.3 (8.8)	4.4 (1.8)	4.0 (2.2)

	0.5	116.9 (116.7)	37.6 (32.6)	36.8 (34.3)	3.0 (2.5)	2.7 (0.9)	2.1 (1.2)
	0.75	34.3 (34)	9.4 (5.3)	8.6 (6.0)	1.0 (0.1)	1.3 (0.4)	1.0 (0.1)
	1	9.6 (9)	4.4 (1.8)	4.2 (2.2)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
	EQL	16.593	7.002	6.889	1.536	0.909	0.848

The first row of each cell shows ARL (SDRL), the second row of each cell shows the value of 25th, 50th, and 75th percentiles, respectively and the last cell in each chart shows the value of the EQL. Bold numbers mean the smallest ARL for each shift in different dimension sizes.

Table 18 RL profile for the proposed BSCSCUSUM and competing charts with (m = 30%; n = 30%) sample sizes.

Dist.	δ	p = 100			p = 300		
		BS	BSCUSUM	BSCSCUSUM	BS	BSCUSUM	BSCSCUSUM
MVN ₀	0	199.8 (196.1)	199.2 (190.8)	198.9 (192.4)	200.5 (199.7)	200.4 (196.7)	200.7 (195.8)
	0.1	202.7 (203.4)	192.6 (183.9)	188.6 (183.6)	170.8 (172.2)	114.4 (108.5)	122.1 (119.1)
	0.2	180.6 (180.4)	143.7 (135.6)	145.2 (142.7)	51.5 (50.1)	15.5 (11.2)	13.2 (11.5)
	0.3	136.0 (138.4)	67.8 (61.5)	72.7 (67.7)	9.0 (8.6)	4.6 (2.0)	4.1 (2.4)
	0.4	80.8 (79.5)	27.3 (22.4)	28.3 (26.1)	2.3 (1.7)	2.3 (0.7)	1.8 (0.8)
	0.5	42.0 (42.2)	13.2 (8.9)	10.4 (9.2)	1.2 (0.4)	1.5 (0.5)	1.1 (0.1)
	0.75	7.6 (7)	4.3 (1.8)	3.6 (2.2)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
	1	2.1 (1.5)	2.4 (0.76)	1.9 (1.0)	1.0 (0.0)	1.0 (0)	1.0 (0.0)
	EQL	6.610	3.297	3.075	0.812	0.586	0.553
MVN ₁	0	197.6 (197.3)	199.7(197.13)	198.8 (195.7)	197.5 (193.8)	199.6 (194.2)	199.4 (194.1)
	0.1	199.1 (195.2)	201.2 (198.9)	200.7 (198.4)	188.2 (190.2)	137.8 (131.0)	143.4 (138.4)
	0.2	190.8 (190.2)	170.1 (161.9)	175.9 (172.4)	86.4 (84.4)	24.0 (19.4)	23.1 (21.7)
	0.3	168.7 (164.7)	100.5 (94.7)	104.9 (99.4)	20.1 (19.3)	6.5 (3.2)	5.8 (3.1)
	0.4	116.8 (118)	46.4 (41.6)	47.5 (43.1)	4.9 (4.3)	3.2 (1.1)	2.8 (1.2)
	0.5	72.8 (73.7)	21.4 (16.6)	20.1 (15.1)	1.7 (1.1)	2.0 (0.6)	1.4 (0.6)
	0.75	17.5 (17)	6.2 (2.9)	5.7 (3.6)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
	1	4.6 (4.1)	3.2 (1.1)	2.6 (1.4)	1.0 (0.0)	1.0 (0)	1.0 (0.0)
	EQL	10.570	4.737	4.614	1.133	0.697	0.659
MVT	0	198.5 (197.6)	201.1 (201.8)	200.3 (196.5)	199.5 (197.4)	199.4 (192.8)	201.4 (199.6)
	0.1	195.9 (195.9)	200.4 (200.4)	199.7 (197.6)	199.5 (199.9)	184.6 (180.4)	184.1 (180.6)
	0.2	191.2 (189.7)	193.1 (192.9)	193.5 (196.6)	137.3 (138.7)	73.5 (68.3)	66.4 (60.2)
	0.3	185.5 (185.4)	159.4 (154.2)	156.2 (152.0)	54.7 (53.3)	17.8 (14.8)	14.5 (11.6)
	0.4	154.3 (151.5)	106.1 (101.0)	99.5 (95.4)	17.0 (16.4)	6.1 (3.7)	5.6 (3.3)
	0.5	119.2 (116.8)	60.3 (56.3)	55.5 (49.5)	5.5 (4.9)	3.2 (1.4)	2.9 (1.6)

	0.75	42.0 (41.2)	13.9 (11.0)	12.6 (9.1)	1.1 (0.4)	1.0 (0.5)	1.0 (0.5)
	1	12.8 (12.1)	5.3 (2.9)	4.8 (2.8)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
	EQL	17.820	9.360	8.771	2.033	1.142	1.062
MVG	0	195.0 (196.9)	200.2 (192.9)	199.4 (197.6)	200.5 (201.4)	200.8 (197.9)	200.5 (196.6)
	0.1	198.6 (197.7)	200.2 (193.6)	203.1 (198.1)	187.0 (187.6)	143.6 (135.9)	147.5 (142.3)
	0.2	188.8 (191.5)	169.6 (164.0)	172.5 (170.5)	78.9 (78.1)	25.0 (23.0)	24.6 (20.4)
	0.3	163.6 (164.1)	98.3 (92.7)	103.1 (96.3)	17.7 (17.2)	6.8 (3.9)	5.8 (3.1)
	0.4	115.3 (115.2)	43.4 (38.3)	42.6 (40.4)	4.3 (3.8)	3.4 (1.5)	2.6 (1.1)
	0.5	68.8 (68.2)	20.1 (15.6)	19.2 (16.8)	1.6 (1.0)	2.2 (0.9)	1.2 (0.4)
	0.75	15.2 (14.9)	5.7 (2.7)	5.2 (3.0)	1.0 (0.0)	1.3 (0.4)	1.0 (0.0)
	1	3.9 (3.4)	2.9 (1.0)	2.3 (1.3)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
	EQL	9.905	4.501	4.361	1.066	1.314	0.657

The first row of each cell shows ARL (SDRL), the second row of each cell shows the value of 25th, 50th, and 75th percentiles, respectively and the last cell in each chart shows the value of the EQL. Bold numbers mean the smallest ARL for each shift in different dimension sizes.

Table 19 RL profile for the proposed BSCSCUSUM and competing charts with (m = 30%; n= 45%) sample sizes.

Dist.	δ	p = 100			p = 300		
		BS	BSCUSUM	BSCSCUSUM	BS	BSCUSUM	BSCSCUSUM
MVN₀	0	204.3 (203.1)	201.4(195.2)	201.1 (196.7)	199.8 (206.2)	199.2 (193.7)	201.3 (193.9)
	0.1	196.6 (194.7)	198.6 (188.7)	198.2 (193.4)	182.7 (181.5)	142.9 (136.6)	148.9 (145.4)
	0.2	183.6 (182.2)	168.2 (164.6)	172.4 (171.3)	78.9 (79.5)	26.3 (21.2)	25.8 (22.5)
	0.3	155.2 (155.1)	101.6 (96.6)	105.6 (102.0)	18.6 (18.3)	7.1 (3.6)	6.5 (3.2)
	0.4	111.6 (110.2)	48.7 (42.9)	50.1 (48.9)	4.6 (4.1)	3.4 (1.2)	3.1 (1.6)
	0.5	70.3 (70.1)	23.6 (18.7)	22.2 (17.0)	1.7 (1.1)	2.1 (0.6)	1.5 (0.9)
	0.75	16.9 (16.1)	6.7 (3.4)	6.1 (3.8)	1.0 (0.0)	1.0 (0.2)	1.0 (0.0)
	1	4.7 (4.1)	3.4 (1.3)	3.0 (1.6)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
	EQL	10.153	4.965	4.843	1.079	0.724	0.691
MVN₁	0	196.3 (198.3)	200.1 (194.6)	201.1 (197.1)	198.7 (195.7)	199.4 (192.1)	201.8 (200.8)
	0.1	199.5 (196.3)	199.4 (194.2)	199.8 (197.8)	191.7 (189.8)	162.4 (155.5)	168.7 (164.5)
	0.2	190.0 (189.6)	181.1 (175.5)	185.7 (180.5)	109.8 (108.2)	41.7 (36.0)	42.8 (38.3)
	0.3	177.7 (175.5)	129.7 (125.5)	134.1 (131.1)	34.9 (34.3)	10.7 (6.6)	9.6 (7.1)
	0.4	138.7 (139.0)	74.1 (68.5)	71.1 (67.9)	9.8 (9.3)	4.5 (2.0)	3.8 (3.6)
	0.5	99.2 (100)	38.0 (32.5)	37.7 (32.2)	3.4 (2.9)	2.1 (0.6)	1.7 (0.8)
	0.75	31.7 (31.4)	10.0 (5.9)	9.1 (6.6)	1.0 (0.2)	1.2 (0.3)	1.0 (0.1)

	1	9.7 (9.2)	4.7 (2.1)	4.0 (2.2)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
	EQL	14.789	6.933	6.715	1.516	0.941	0.827
MVT	0	196.0 (194.3)	199.4 (195.1)	200.8 (198.0)	201.3 (200.8)	199.7 (193.3)	200.7 (199.1)
	0.1	195.6(194.1)	198.4 (195.1)	203.1 (197.8)	198.2 (198.7)	192.6 (190.8)	191.1 (188.6)
	0.2	195.3 (195.3)	194.6 (193.9)	194.3 (189.0)	159.9 (160.4)	104.3 (100.7)	101.0 (95.2)
	0.3	190.9 (191.1)	172.7 (168.2)	174.7 (168.8)	86.9 (85.3)	31.4 (28.2)	29.4 (24.0)
	0.4	170.6 (170.4)	133.6 (125.8)	132.6 (130.6)	34.6 (34.2)	10.7 (7.5)	9.8 (6.6)
	0.5	144.4 (142.3)	89.4 (84.5)	88.0 (83.7)	13.0 (12.6)	5.1 (2.7)	4.4 (2.8)
	0.75	68.8 (67.5)	26.5 (22.7)	24.9 (20.8)	1.9 (1.3)	1.8 (0.7)	2.1 (0.9)
	1	27.6 (26.8)	9.6 (6.6)	9.1 (6.2)	1.0 (0.2)	1.1 (0.3)	1.0 (0.3)
	EQL	24.867	13.506	13.163	3.134	1.691	1.629
MVG	0	199.7 (196.0)	201.7 (196.0)	200.0 (192.3)	204.3 (203.2)	200.3 (196.9)	199.2 (195.9)
	0.1	204.2 (202.1)	207.6 (200.6)	204.5 (201.7)	202.5 (203.6)	167.6 (165.9)	172.0 (168.6)
	0.2	211.9 (209.1)	193.8 (188.3)	197.7 (195.3)	119.2 (119.1)	39.0 (33.0)	41.2 (37.6)
	0.3	200.6 (201.6)	139.9 (134.0)	136.8 (135.8)	34.6 (34)	9.9 (5.9)	9.2 (5.6)
	0.4	162.4 (162.3)	77.7 (71.3)	76.6 (74.6)	9.3 (8.8)	4.4 (1.8)	3.8 (2.2)
	0.5	116.9 (116.7)	37.6 (32.6)	35.7 (32.4)	3 (2.5)	2.7 (0.9)	1.9 (0.8)
	0.75	34.3 (34.0)	9.4 (5.3)	8.4 (5.2)	1.0 (0.1)	1.3 (0.4)	1.0 (0.1)
	1	9.6 (9.0)	4.4 (1.8)	4.1 (1.7)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
	EQL	16.593	7.002	6.707	1.536	0.909	0.829

The first row of each cell shows ARL (SDRL), the second row of each cell shows the value of 25th, 50th, and 75th percentiles, respectively and the last cell in each chart shows the value of the EQL. Bold numbers mean the smallest ARL for each shift in different dimension sizes.

Table 20 RL profile for the proposed SDCSCUSUM and competing charts with (m = 30%; n= 30%) sample sizes.

Dist.	δ	p = 100			p = 300		
		SD	SDCUSUM	SDCSCUSUM	SD	SDCUSUM	SDCSCUSUM
MVN₀	0	198.8 (194.7)	200.5 (195.0)	199.7 (194.3)	200.6 (198.9)	199.6 (195.0)	201.4 (200.1)
	0.1	201.0 (200.2)	192.0 (185.6)	194.6 (185.6)	171.6 (171.5)	116.8 (111.1)	124.7 (120.1)
	0.2	181.2 (179.9)	149.5 (141.6)	152.7 (146.0)	52.1 (51.3)	16.1 (12.0)	15.4 (11.9)
	0.3	139.8 (140.4)	72.3 (67.6)	74.4 (70.6)	9.4 (8.9)	4.7 (2.0)	4.3 (2.4)
	0.4	85.1 (85.1)	29.9 (25.2)	30.1 (26.7)	2.4 (1.8)	2.4 (0.8)	2.1 (1.0)
	0.5	45.9 (46.1)	14.3 (10.1)	13.5 (10.7)	1.2 (0.5)	1.6 (0.5)	1.3 (0.3)
	0.75	9.1 (8.6)	4.6 (2.1)	4.1 (2.0)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
	1	2.6 (2.1)	2.5 (0.8)	2.1 (0.7)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
	EQL	7.158	3.504	3.386	0.821	0.586	0.581

MVN ₁	0	199.0 (199.2)	199.8 (198.2)	199.7 (199.4)	197.1 (194.4)	200.2(195.8)	200.4 (195.7)
	0.1	199.8 (198.6)	198.9 (196.9)	200.5 (201.0)	186.7 (189.8)	148.8 (144.5)	151.1 (148.4)
	0.2	189.7 (192)	174.4 (171.5)	181.3 (180.4)	87.1 (86.6)	26.6 (22.0)	25.4 (21.6)
	0.3	170.6 (165.5)	106.7 (100.7)	111.8 (105.6)	20.5 (19.7)	6.8 (3.4)	6.2 (3.2)
	0.4	121.0 (119.9)	50.9 (47.1)	52.7 (48.1)	5.1 (4.5)	3.3 (1.2)	3.2 (1.1)
	0.5	76.6 (77.7)	23.8 (19.3)	22.9 (18.9)	1.8 (1.2)	2.1 (0.6)	1.7 (0.8)
	0.75	20.2 (19.7)	6.6 (3.4)	6.2 (3.2)	1.0 (0.0)	1.0 (0.1)	1.0 (0.0)
	1	5.8 (5.3)	3.4 (1.3)	3.0 (1.2)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
	EQL	11.346	5.066	5.023	1.145	0.727	0.699
MVT	0	199.1 (198.7)	200.9 (198.6)	201.1 (196.3)	201.3 (205.4)	199.7 (195.2)	198.2 (195.4)
	0.1	195.6 (196.9)	200.7 (200.2)	199.1 (198.3)	196.7 (196.7)	185.3 (183.5)	182.8 (179.6)
	0.2	190.2 (188.8)	193.8 (193.4)	189.3 (190.0)	133.4 (133.1)	77.3 (72.3)	66.1 (59.3)
	0.3	182.5 (182.2)	161.7 (160.3)	155.8 (153.6)	53.3 (52.0)	18.8 (15.9)	14.1 (11.2)
	0.4	151.9 (149)	110.6 (109.4)	98.8 (93.6)	17.0 (16.1)	6.2 (3.9)	5.5 (3.3)
	0.5	117.3 (116.5)	63.6 (60.0)	54.5 (49.3)	5.7 (5.2)	3.3 (1.5)	2.8 (1.2)
	0.75	42.7 (42.4)	15.0 (12.4)	11.7 (8.9)	1.2 (0.5)	1.3 (0.5)	1.0 (0.5)
	1	14.2 (13.6)	5.6 (3.3)	4.6 (2.7)	1.0 (0.1)	1.0 (0.0)	1.0 (0.0)
	EQL	17.940	9.793	8.543	2.025	1.215	1.050
MVG	0	197.1 (197.4)	199.4 (195.7)	198.3 (194.6)	196.6 (195.6)	199.5 (197.4)	200.5 (196.6)
	0.1	198.5 (198.1)	199.1 (194.7)	199.6 (199.2)	182.9 (181.7)	139.5 (133.7)	146.6 (137.1)
	0.2	184.7 (184.1)	166.0 (160.6)	169.5 (164.1)	73.3 (72.4)	22.7 (18.2)	24.1 (20.1)
	0.3	155.4 (154.3)	89.3 (84.2)	87.9 (82.5)	16.4 (15.9)	5.9 (2.9)	5.2 (2.7)
	0.4	101.8 (101.5)	38.6 (33.9)	37.0 (32.7)	3.9 (3.3)	2.9 (1.0)	2.1 (1.0)
	0.5	58.7 (58.3)	17.9 (13.9)	16.7 (13.7)	1.5 (0.9)	1.9 (0.5)	1.0 (0.2)
	0.75	13.2 (12.7)	5.2 (2.5)	4.6 (2.8)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
	1	3.7 (3.2)	2.7 (1.0)	2.0 (0.8)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
	EQL	8.851	4.136	3.888	1.017	0.679	0.632

The first row of each cell shows ARL (SDRL), the second row of each cell shows the value of 25th, 50th, and 75th percentiles, respectively and the last cell in each chart shows the value of the EQL. Bold numbers mean the smallest ARL for each shift in different dimension sizes.

Table 21 RL profile for the proposed SDCSCUSUM and competing charts with (m = 30%; n= 45%) sample sizes.

Dist.	δ	p = 100			p = 300		
		SD	SDCUSUM	SDCSCUSUM	SD	SDCUSUM	SDCSCUSUM
MVN ₀	0	199.7 (201.7)	200.9 (195.7)	200.0 (196.1)	199.4 (197.2)	201.1 (194.9)	200.2 (194.1)
	0.1	196.4 (195.3)	198.1 (191.6)	198.7 (188.5)	177.4 (176.6)	151.8 (146.8)	160.4 (155.1)
	0.2	182.8 (181.7)	169.8 (163.0)	174.0 (170.9)	79.5 (79.4)	27.9 (22.8)	29.2 (25.6)
	0.3	157.0 (156.8)	106.4 (103.0)	111.8 (109.6)	19.2 (18.9)	7.2 (3.9)	6.9 (3.6)
	0.4	116.3 (116.5)	52.3 (48.1)	51.8 (47.3)	4.7 (4.2)	3.5 (1.3)	3.3 (1.2)
	0.5	73.4 (73.9)	25.4 (20.8)	24.7 (20.5)	1.8 (1.2)	2.2 (0.6)	1.9 (0.9)
	0.75	18.9 (18.7)	7.1 (3.7)	6.7 (3.1)	1.0 (0.0)	1.0 (0.2)	1.0 (0.0)
	1	5.7 (5.1)	3.6 (1.4)	3.5 (1.1)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
	EQL	10.782	5.231	5.189	1.087	0.743	0.740
MVN ₁	0	195.2 (195.3)	201.2 (195.4)	200.7 (195.6)	197.8 (195.5)	200.5 (194.1)	199.2 (200.0)
	0.1	200.4 (198.3)	196.7 (190.7)	200.8 (193.8)	189.9 (188.8)	164.0 (156.1)	171.8 (167.9)
	0.2	189.5 (189)	179.6 (173.2)	187.9 (176.1)	109.1 (108.3)	43.4 (38.5)	45.2 (40.5)
	0.3	171.3 (170)	133.3 (129.5)	137.4 (134.8)	35.0 (34)	10.9 (7.0)	10.5 (8.3)
	0.4	139.3 (136.2)	78.0 (72.9)	77.6 (74.2)	10.1 (9.5)	4.7 (2.1)	4.3 (2.0)
	0.5	101.3 (100.2)	41.4 (36.2)	40.7 (35.0)	3.4 (2.9)	2.8 (1.0)	2.1 (1.2)
	0.75	34.5 (34.1)	10.8 (6.8)	9.7 (6.2)	1.0 (0.2)	1.3 (0.4)	1.0 (0.2)
	1	11.6 (11.3)	4.9 (2.3)	4.8 (2.0)	1.0 (0)	1.0 (0.0)	1.0 (0.0)
	EQL	15.462	7.305	7.175	1.517	1.750	0.873
MVT	0	194.8 (193.2)	199.6 (196.7)	201.0 (197.3)	200.5 (202.5)	200.1 (199.0)	200.6 (199.6)
	0.1	195.0 (194.2)	202.4 (198.6)	202.5 (199.3)	199.9 (195.7)	193.9 (190.9)	189.5 (186.8)
	0.2	194.8 (193.4)	197.5 (190.8)	193.1 (191.0)	156.9 (159.9)	107.6 (104.5)	98.9 (94.2)
	0.3	187.7 (192.7)	175.1 (172.0)	171.0 (165.6)	81.5 (81.9)	33.1 (30.1)	28.6 (22.8)
	0.4	167.1(163.7)	134.8 (128.5)	131.8 (124.6)	32.3 (32.1)	11.0 (8.0)	9.6 (6.4)
	0.5	142.3 (140.1)	92.3 (87.1)	87.2 (80.9)	12.2 (11.8)	5.2 (2.9)	4.1 (2.8)
	0.75	67.4 (66.5)	28.0 (24.8)	23.6 (20.5)	2.0 (1.4)	2.0 (0.7)	1.8 (0.8)
	1	28.2 (27.3)	10.0 (7.3)	8.7 (6.1)	1.1 (0.3)	1.2 (0.4)	1.0 (0.0)
	EQL	24.565	13.951	12.843	3.030	1.691	1.553
MVG	0	201.7 (202.5)	201.8 (198.4)	201.8 (197.1)	201.3 (202.1)	199.3 (195.0)	201.6 (197.2)
	0.1	204.4 (202)	209.8 (208.8)	203.6 (201.1)	200.5 (201.3)	165.2 (161.2)	176.2 (175.1)
	0.2	209.9 (209.8)	194.8 (189.3)	195.8 (187.5)	114.4 (116)	38.2 (32.7)	39.3 (36.4)
	0.3	189.9 (188.4)	135.0 (130.8)	136.7 (132.7)	32.3 (32.4)	9.5 (5.7)	8.6 (5.2)
	0.4	152.7 (152.9)	72.3 (67.6)	71.3 (67.3)	8.5 (8.0)	4.2 (1.8)	3.2 (1.3)

	0.5	103.8 (103.2)	34.7 (30.3)	33.6 (31.2)	2.9 (2.3)	2.6 (0.8)	1.6 (0.7)
	0.75	29.0 (28.8)	8.7 (5.1)	8.2 (4.3)	1.0 (0)	1.2 (0.4)	1.0 (0.2)
	1	8.7 (8.3)	4.2 (1.8)	3.8 (1.3)	1.0 (0.0)	1.0 (0.0)	1.0 (0.0)
	EQL	14.903	6.627	6.452	1.477	0.878	0.797

The first row of each cell shows ARL (SDRL), the second row of each cell shows the value of 25th, 50th, and 75th percentiles, respectively and the last cell in each chart shows the value of the EQL. Bold numbers mean the smallest ARL for each shift in different dimension sizes.

Presented in **Tables 16-21** are the findings of the proposed charts as well as the charts that competed with them, we discuss the results based on different criteria. These results indicate the following:

- **Assessment Versus Conventional Shewhart Setup charts**

Under all distribution structures MVN_0 , MVN_1 , MVT , and MVG , with $p = 100$ and $p = 300$, across both sample size combinations (i.e., $(m = n = p \times 30\%)$ and $(m = p \times 30\%, n = p \times 45\%)$), the proposed DRCSCUSUM, BSCSCUSUM, and SDCSCUSUM charts have shown superior detection ability compared to the DR, BS, and SD (**cf. Tables 16-21**), as their values of ARL_1 for all the shift types were lower than those of the conventional Shewhart setup charts. For instance, to identify a 0.2 shift within the MVN_1 structure, utilizing a sample size of $(m = n = p \times 30\%)$ with $p = 100$, the ARL_1 for DR chart was 190.8 samples. In contrast, the ARL_1 for DRCSEWMA was only 176.3 samples.

- **Assessment versus conventional CUSUM setup charts**

For all the distribution structures, with $p = 100$ for both combinations of sample sizes (i.e., $(m = n = p \times 30\%)$ and $(m = p \times 30\%, n = p \times 45\%)$), the proposed DRCSCUSUM, BSCSCUSUM, and SDCSCUSUM charts have shown better detection ability than the DRCUSUM, BSCUSUM, and SDCUSUM only for the moderate to large shifts (i.e., 0.5, 0.75, and 1.0) (**cf. Tables 16-21**), as their values of ARL_1 are less than those in the

DRCUSUM, BSCUSUM, and SDCUSUM charts; otherwise, the DRCUSUM, BSCUSUM, and SDCUSUM have better detection ability than the proposed charts (**cf. Tables 16-21**). While with $p = 30$ for both combinations of sample sizes (i.e., $(m = n = p \times 30\%)$ and $(m = p \times 30\%, n = p \times 45\%)$), the proposed DRCSCUSUM, BSCSCUSUM, and SDCSCUSUM charts have shown better detection ability than the DRCUSUM, BSCUSUM, and SDCUSUM, except for the very small to moderate shifts (i.e., 0.1 and 0.2). Otherwise, DRCUSUM, BSCUSUM, and SDCUSUM have better detection ability than the proposed charts (**cf. Tables 16-21**). As their values of ARL_1 are less.

- **Assessment of the proposed charts based on different sample sizes**

Under all distribution environments with $p = 100$ or $p = 300$ and equal sample size choice $(m = n = p \times 30\%)$, the proposed DRCSCUSUM, BSCSCUSUM, and SDCSCUSUM charts have shown better detection ability compared to their performance under unequal sample size $(m = p \times 30\%, n = p \times 45\%)$. as their values of ARL_1 with equal sample size choice $(m = n = p \times 30\%)$ are less than the values of the ARL_1 with unequal sample size $(m = p \times 30\%, n = p \times 45\%)$ (**cf. Tables 16-21**).

- **Assessment of the proposed charts based on different dimensions values**

Under the $p = 300$ with $(m = n = p \times 30\%)$ or $(m = p \times 30\%, n = p \times 45\%)$ combination sample sizes, the proposed DRCSCUSUM, BSCSCUSUM, and SDCSCUSUM charts have shown superior detection ability compared to their performance under $p = 100$. As their values of ARL_1 under $p = 300$ choice are less than the values of the ARL_1 under $p = 100$ choice (**cf. Tables 16-21**).

- **Assessment analysis between the proposed charts**

Under the distribution structures (i.e., MVN_0 and MVN_1) with $p = 100$ and $p = 300$ with different combinations of the sample sizes (i.e., $(m = n = p \times 30\%)$ and $(m = p \times 30\%, n = p \times 45\%)$), the proposed BSCCUSUM chart exhibits superior detection performance compared to the assessed charts. Furthermore, the DRCCUSUM chart surpasses the SDCCUSUM chart. As it offers lower ARL_1 values for various shifts compared to the other two charts (cf. **Tables 16-21**). While for the (MVT and MVG) distributions structure with $p = 100$ and $p = 300$, for the different combinations of sample sizes, i.e., $(m = n = p \times 30\%)$ and $(m = p \times 30\%, n = p \times 45\%)$, the proposed SDCCUSUM chart demonstrates better detection performance compared to the assessed charts. Furthermore, the BSCCUSUM chart exceeds the DRCCUSUM chart. As it offers lower ARL_1 values for various shifts compared to the other two charts (cf. **Tables 16-21**).

- **Assessment analysis based on the EQL values for the overall performance**

For overall performance, under the $p = 300$ with $(m = n = p \times 30\%)$ or $(m = p \times 30\%, n = p \times 45\%)$ combination sample size, the proposed control charts have shown superior detection ability compared to their performance under $p = 100$ with all combinations of sample sizes (cf. **Tables 16-21**). Because they offer the lowest EQL values for all the distribution structures. Based on the overall performance, the proposed classical Shewhart setup and the classical CUSUM setup charts give the lower EQL values compared to the assessed charts. Except for the MVN_0 under unequal sample size $(m = p \times 30\%, n = p \times 45\%)$ the classical DRCUSUM had lowest EQL value (cf. **Tables 16-21**).

4.5. Illustrative Application Example

This section presents a practical application demonstrating the execution of the proposed charts. This application illustrates the actual utilization of these charts for monitoring gas sensor arrays under flow modulation, where real-time responses from high-dimensional data are essential.

Gas Sensor Array Monitoring in Flow Modulation Systems

An electronic nose, a type of gas sensor, may serve as a viable option for a robotic olfactory system capable of detecting and differentiating various gas molecules. Gas sensors utilizing flow modulation are frequently utilized in safety-critical sectors, such as chemical manufacturing, where precise identification of hazardous or combustible gases is essential, see (R. Liu et al., 2023) and (Shinde & Rout, 2021). The most challenging issue in chemical sensing is still sensor drift. A gas sensor array is an advanced detection system comprising several gas sensors; each intended to identify a particular gas or vapor in a specific environment. The number of sensors in the sensor array varies based on how a case needs to be analyzed and generated (Wibowo, 2021). This section describes how our suggested DRCSCUSUM, BSCSCUSUM, SDCSCUSUM, and the competing charts are used in practice to track changes in the means of two-sample high-dimensional data. For this purpose, a gas sensor array under flow modulation obtained by (Ziyatdinov et al., 2015) is used. The dataset contains 896 rows (a measurement per sensor), 100 (we consider only the last 100 columns with non-zero values) columns (features or dimensions), and 896 rows (one measurement per sensor). Raw data on each sample consists of 16 time series, one for each sensor. The twelve tested gas classes formed a relatively wide range of two analytes, acetone, and ethanol, in binary mixtures. For five minutes, each time series was recorded at a 25 Hz (samples per second) sampling rate, producing 100 data points for each time series. We categorized the four gases,

acetone, air, ethanol, and a mixture, into two groups according to their chemical nature; group 1 with 448 samples for the organic solvents (Acetone, Ethanol) because both acetone and ethanol are organic solvents, volatile, polar, and flammable. They are frequently used in chemical reactions and as cleaning agents. They are similar in forming vapors that can mix with air, creating flammable mixtures. Group 2 has 448 samples for non-flammable gases (air, mixture) because the air itself is not flammable but can support combustion. Although potentially flammable, a mixture of acetone and ethanol with air contains air as a key component. This group focuses on the presence of air, which distinguishes it from the pure organic solvents in group 1. Using the 896 samples 448 for group 1 and 448 for group 2, each sample contains 16 sensors with 100 features per sensor. The implementation of these proposed charts DRCSCUSUM, BSCSCUSUM, and SDCSCUSUM charts was conducted in three steps in the initial step using the 20,000,000 iterations, the values of means $(\mu_{T_1}, \mu_{T_2}, \mu_{T_3})$ and the variances $(\sigma_{T_1}, \sigma_{T_2}, \sigma_{T_3})$ for the T_{1t} , T_{2t} , and T_{3t} statistics are obtained from our daily stock price data. Subsequently, in the second step, and then the $L_{S1}, L_{S2}, L_{S3}, L_{S4}, L_{S5}, L_{S6}, h_1, h_2, h_3, h_4, h_5$ and h_6 are found using 10,000 iterations, the Shewhart and CUSUM-based charts are running individually to obtain the coefficients at $ARL_0 = 800$ on one side and then combined Shewhart setup coefficients generate $ARL_0 = 400$ and combined CUSUM setup coefficients generate $ARL_0 = 400$ and overall, CSCUSUM setup coefficients may provide $ARL_0 = 200$. Finally, considering the values of the pre-obtained means $(\mu_{T_1}, \mu_{T_2}, \mu_{T_3})$ and standard deviations $(\sigma_{T_1}, \sigma_{T_2}, \sigma_{T_3})$, and the values of $L_{S1}, L_{S2}, L_{S3}, L_{S4}, L_{S5}, L_{S6}, h_1, h_2, h_3, h_4, h_5$ and h_6 shown in (cf. **Tables 22**) We have calculated the plotting statistics for the proposed charts DRCSCUSUM, BSCSCUSUM, and SDCSCUSUM using equations (4.13), (4.15), and (4.17). This will give us 53 plotting statistics for each proposed chart from our daily stock price data.

Table 22 The mean, standard deviation, and coefficient limits of the proposed and competing charts for Gas Sensor Array Monitoring in Flow Modulation Systems at a fixed $ARL_0 = 200$

μ_{T_1}	1737.795	μ_{T_2}	1631.290	μ_{T_3}	7628.229
σ_{T_1}	1670.113	σ_{T_2}	1790.623	σ_{T_3}	7300.836
Classical Shewhart charts for two-sample high-dimensional data					
DR		BS		SD	
L_{s_1}	0.85864	L_{s_3}	0.8468	L_{s_5}	0.8765
L_{s_2}	3.571	L_{s_4}	3.471	L_{s_6}	4.671
Classical CUSUM charts for two-sample high-dimensional data					
DRCUSUM		BSCUSUM		SDCUSUM	
h_1	1.552	h_3	1.505	h_5	1.451
h_2	6.805	h_4	6.392	h_6	6.537
Proposed CSCUSUM charts for two-sample high-dimensional data					
DRCSCUSUM		BSCSCUSUM		SDCSCUSUM	
L_{s_1}	0.858667	L_{s_3}	0.846802	L_{s_5}	0.8766
L_{s_2}	3.605	L_{s_4}	3.493	L_{s_6}	4.693
h_1	1.758	h_3	1.709	h_5	1.655
h_2	6.808	h_4	7.168	h_6	7.165

Bootstrapping is a statistical resampling technique that involves repeatedly sampling with replacement from the original data to estimate the distribution of a sample statistic. When the sample size is small, or the population distribution is unknown, it is beneficial for determining the reliability of an estimate (e.g. mean, median, or variance) [73]. The bootstrap technique is used to generate 200 IC data samples. We start with our original 53 samples and then randomly select 53 samples with replacements from the original sample and repeat this resampling process 200 times.

This procedure gives us 200 samples. For the IC data, plotting statistics are calculated for each set of samples. Similarly, we used the same technique to generate 100 samples for the OOC data by introducing a 0.02 shift amount in the second sample (i.e., $x_{ij2t} + \delta$) and plotting statistics for DRCSCUSUM, BSCSCUSUM, and SDCSCUSUM and the competing charts were calculated for each set of samples. **Figures 12-14.** It shows 200 IC and 100 OOC test statistic values of the proposed and the competing charts, and their control limits. The IC and OOC values are separated by a vertical line in each of the two plots, where the IC region is presented in pink color while the white area is used as the OOC region, and for the DRCUSUM, BSCUSUM, and SDCUSUM chart the blue points represent the upper-sided statistics for the charts and the black points represent the lower-sided statistics for the charts. Two horizontal green lines in each plot represent UCL and LCL for the corresponding charts

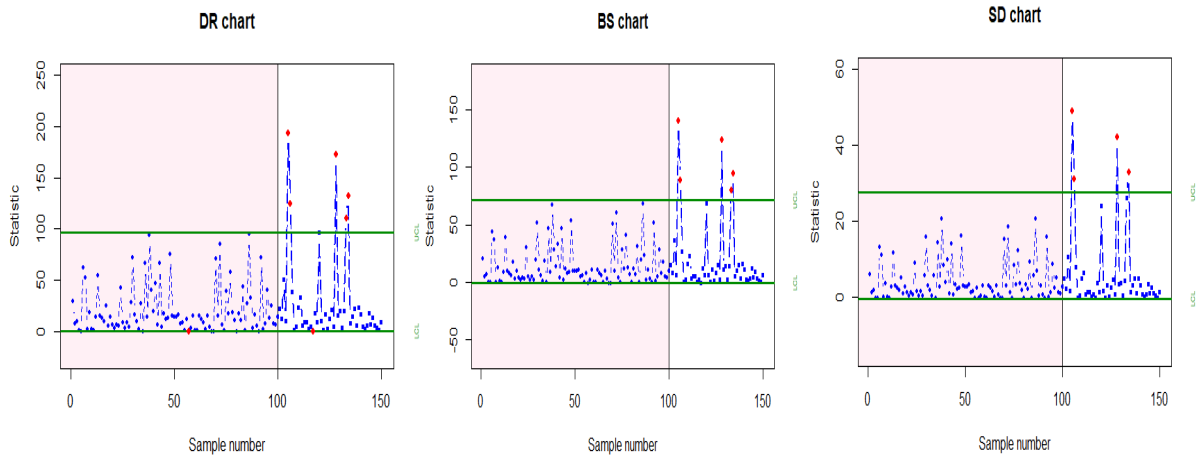


Figure 12 Conventional DR, BS and SD charts for Gas Sensor Array Monitoring in Flow Modulation Systems

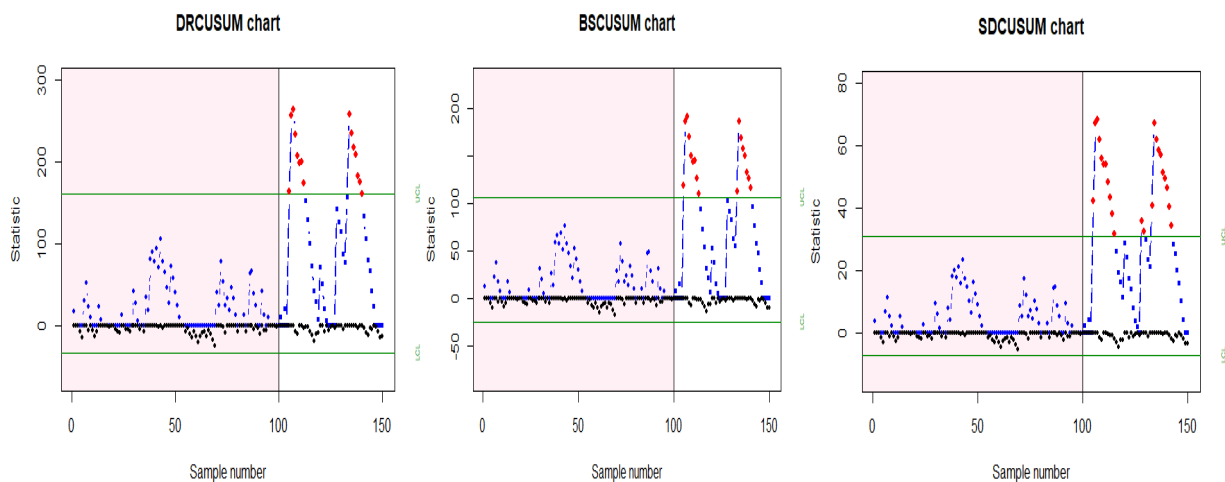
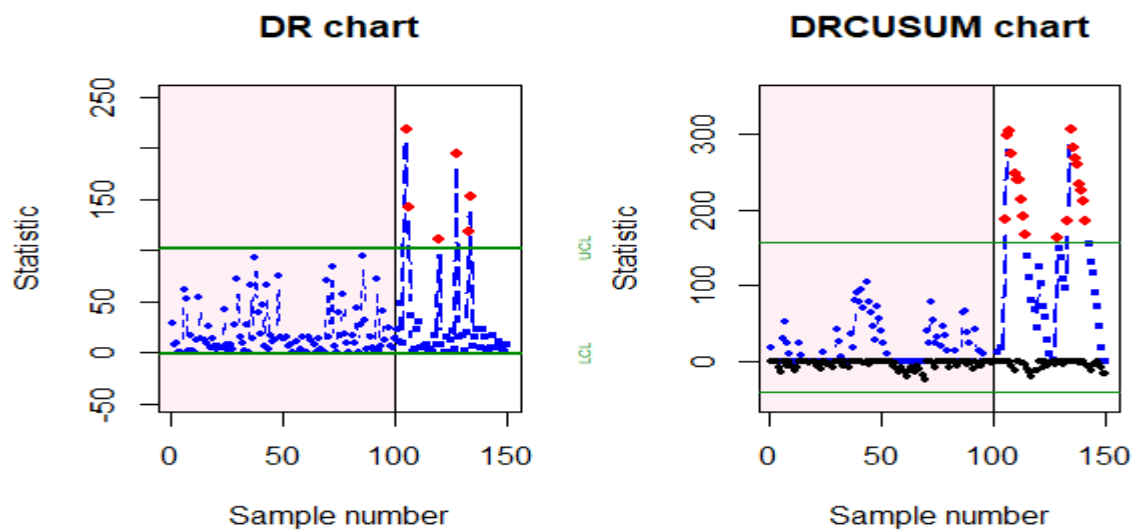
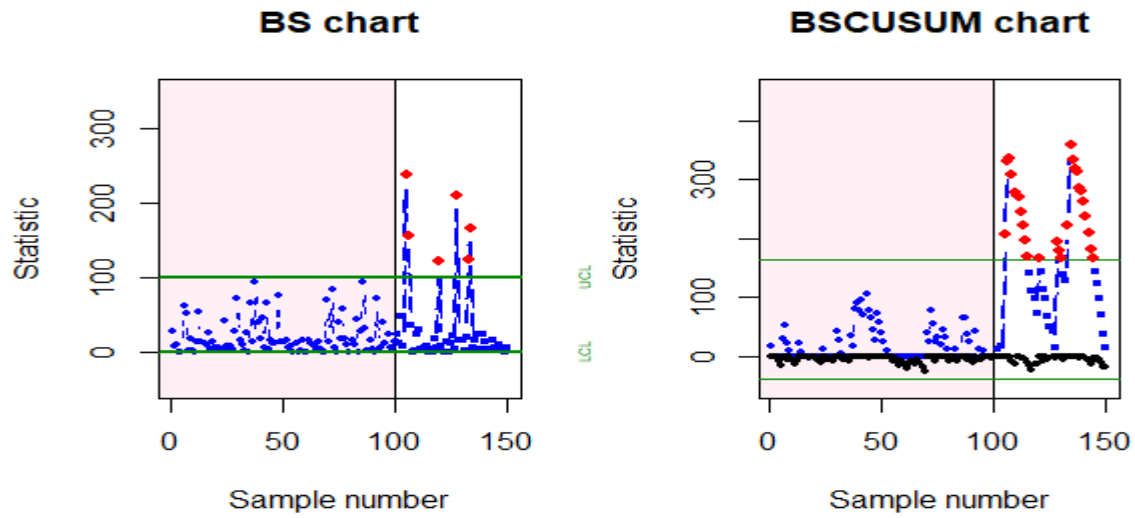


Figure 13 Conventional DCUSUM, BSCUSUM, and SDCUSUM charts for Gas Sensor Array Monitoring in Flow Modulation Systems

DRCSCUSUM Chart



BSCSCUSUM Chart



SDCSCUSUM Chart

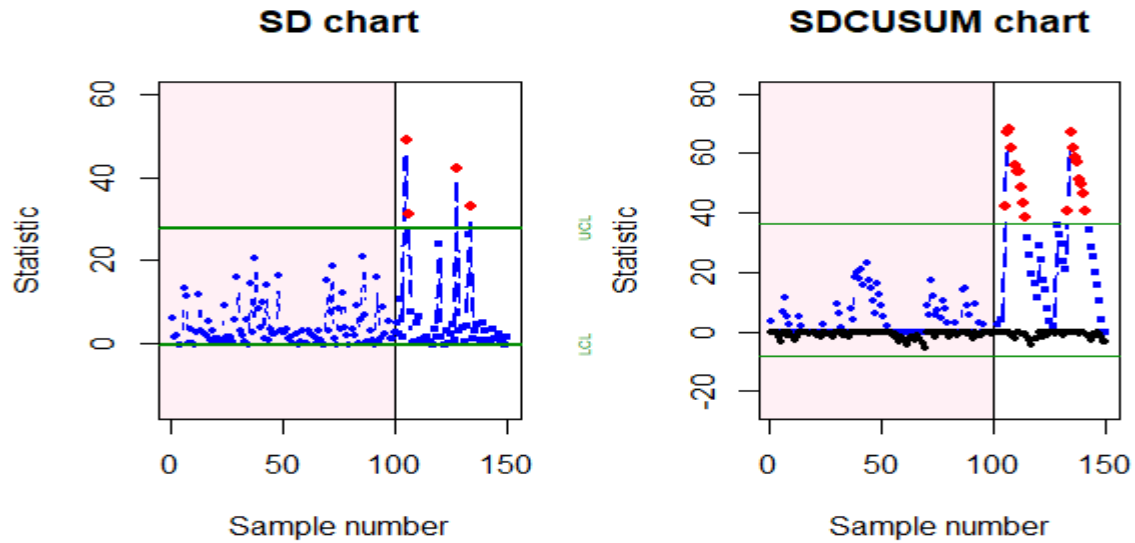


Figure 14 Proposed DRCSCUSUM, BSCSCUSUM and SDCSCUSUM charts for Gas Sensor Array Monitoring in Flow Modulation Systems

The IC and OOC values are separated by a vertical line in each of the plots, where the IC region is presented in pink color while the white area is used as the OOC region, and the blue points represent the plotting statistics for the proposed and competing charts. Two horizontal green lines in each plot represent UCL and LCL for the corresponding charts. The plotting statistics of the proposed and competing charts are shown in blue when they are in IC state and red points when they are in OOC state. For each competing chart, we can note that DR and BS charts have the same performance as it was also proved by (Mahmood et al., 2024) that the DR and BS have the same average run length profile. And both of them outperformed the SD. Our proposed DRCSCUSUM chart outperformed both of the classical charts by detecting 25 OOC points. For the comparison between the three proposed charts, the BSCSCUSUM chart demonstrates better detection performance compared to the assessed charts. However, the BSCSCUSUM detects 32 OOC points when the process is OOC. Furthermore, the DRCSCUSUM chart exceeds the SDCSEWMA chart since it detected 25 OOC points. At the same time, SDCSCUSUM detected only 22 OOC points. Hence, the BSCSCUSUM chart is more efficient in studying and monitoring the behavior of the gas sensor arrays under flow modulation.

CHAPTER 5

CONCLUSION, REMARKS, AND RECOMMENDATIONS

5.1. Introduction

The term “high-dimensional data” refers to a dataset that contains a large number of features compared to the number of the observations. The complexity of high-dimensional data requires sophisticated analytical approaches to derive significant insights, as conventional statistical procedures frequently falter because of the “curse of dimensionality”. The existing literature review about high-dimensional data mostly focuses on monitoring single-sample high-dimensional data; nevertheless, there is an absence of research concerning two-sample high-dimensional data. Except there is one paper recently published in (2024) introduced a memoryless setup control charts for monitoring the means of two-sampled high-dimensional data. The Shewhart setup proposed charts simply consider current process information; hence, they are less adept at identifying small variations in the means of two-sample high-dimensional data. However, there are currently no available charts to monitor the small variations in the means of two-sampled high-dimensional data. Therefore, in this thesis we introduced a memory setup, and a combined memoryless and memory control chart, which effectively incorporates both current and historical information in the process and is proficient in identifying small to large shifts in the means of two-sampled high-dimensional data at the same time. These proposed charts are designed based on the two-sample high-dimensional means tests named [27], [67], and [68]. The proposed charts are assessed according to run length metrics across various distributional structures, dimensions, and

sample sizes. The real-life applications of these charts are additionally illustrated by using real life datasets.

5.1.1. Concluding Remarks Based on CUSUM Charts for Two-sample High-Dimensional Data

The application of memory-based control charts to deal with high-dimensional data, especially for two-sample high-dimensional data, can play a significant role by addressing the challenges associated with such data to ensure quality control across different domains, including the field of manufacturing. To construct a robust high-dimensional control chart that can handle two-sample high-dimensional data efficiently, previously Mahmood, et al. [23] proposed two Shewhart-type control charts, DR and SD charts. Due to the limitation of these memoryless charts for small shifts, we proposed two more memory charts named DRCUSUM and SDCUSUM. The prevalence of high-dimensional data, resulting from excess features over observations, is rising in contemporary manufacturing processes. This can be attributed to data collection capabilities, automation, and advancements in sensing technology. High-dimensional single-sample data is the subject of a large portion of the SPC literature currently in publication. However, the analysis gets harder as manufacturing processes get more intricate and data-intensive, producing high-dimensional two-sample data. No research has utilized control charts to study the behavior of two-sample high-dimensional data except in one paper by Mahmood, et al. [23]. This chapter discusses the development of new control charts for two-sample high-dimensional data under the CUSUM setup. The focus is to check if the CUSUM integration into the DRCUSUM and SDCUSUM charts outperforms the DR and SD charts. Hence, the run length profile performance is compared for the proposed DRCUSUM and SDCUSUM charts with DR and SD charts under a higher-dimension

environment $p = 100$ and 300 with different combinations of sample sizes ($m = n = p \times 30\%$) or ($m = p \times 30\%, n = p \times 45\%$).

The scope of the chapter may be extended by constructing combined and mixed structure schemes based on Shewhart, EWMA, and CUSUM charts to detect small and moderate shifts in the location parameter. Furthermore, future studies might investigate the monitoring of the covariance parameter of high-dimensional data following the proposed methodology.

5.1.2. Concluding Remarks Based on CSEWMA Charts for Two-sample High-Dimensional Data

Memory-based control charts are essential for handling high-dimensional data in two sample scenarios, particularly in chemical and environmental engineering. Mahmood et al. (2024) proposed Shewhart-type control charts, but these are memoryless and less effective in detecting small to moderate shifts. Ahmed (2023) developed memory-type control charts based on EWMA, which was effective in detecting small to moderate shifts. This study proposes DRCSEWMA, BSCSEWMA, and SDCSEWMA combined memoryless and memory charts to effectively monitor moderate and large shifts in two-sampled high-dimensional data. The growing prevalence of high-dimensional data in modern chemical processes is due to advancements in sensor technologies, automation, and extensive data acquisition. Current SPC research focuses on high-dimensional single-sample data, but the analysis becomes more complex when examining high-dimensional two-sample data due to wide correlation patterns and numerous variables. The DRCSEWMA, BSCSEWMA and SDCSEWMA charts were specifically designed to address these challenges, with their effectiveness evaluated through a comprehensive analysis of run length profiles under various distribution structures (MVN_0 , MVN_1 , MVT , MVG).

This chapter scope may be broadened by developing integrated structural schemes utilizing combined Shewhart-based and CUSUM-based charts to identify very small to large shifts in the location parameter. Moreover, subsequent research may explore the surveillance of the covariance parameter in high-dimensional data or the joint monitoring of location and scale parameters within the high-dimensional framework utilizing the proposed methodology.

5.1.3. Concluding Remarks Based on CSCUSUM Charts for Two-sample High-Dimensional Data

Memory-based control charts are crucial for managing high-dimensional data in two-sample situations, in different industrial areas. [23] introduced Shewhart-setup charts for monitoring the means of two sampled high-dimensional datasets; nevertheless, these charts are devoid of memory and are less proficient in identifying tiny to moderate alterations. [104] devised memory-type control charts utilizing the CUSUM framework, which proved useful in identifying minor alterations. This study introduced DRCSCUSUM, BSCSCUSUM, and SDCSCUSUM, which integrate memoryless (Shewhart-setup) and memory-based (CUSUM-setup) control charts to efficiently monitor both small and large shifts in two-sample high-dimensional data at the same time. The increasing prevalence of high-dimensional data in contemporary chemical processes and other industrial sectors is attributable to developments in sensor technologies, automation, and comprehensive data collecting. Current SPC research concentrates on high-dimensional single-sample data; however, the analysis gets increasingly intricate when addressing high-dimensional two-sample data due to extensive correlation patterns and a multitude of factors. The DRCSCUSUM, BSCSCUSUM and SDCSCUSUM charts were specifically designed to address these challenges, with their effectiveness evaluated through a comprehensive analysis of run length profiles under various distribution structures (MVN_0 , MVN_1 , MVT, MVG). Results demonstrated

that Under the distribution structures (i.e., MVN_0 and MVN_1) with $p = 100$ and $p = 300$ for the different combinations of sample sizes (i.e., $(m = n = p \times 30\%)$ and $(m = p \times 30\%, n = p \times 45\%)$), the proposed BSCSCUSUM chart exhibits superior detection performance compared to the assessed charts. Furthermore, the DRCSCUSUM chart surpasses the SDCSCUSUM chart, as it offers a lower ARL_1 values for various shifts. While for the (MVT, MVG) distributions structure with $p = 100$ and $p = 300$, for the different combinations of sample sizes, i.e., $(m = n = p \times 30\%)$ and $(m = p \times 30\%, n = p \times 45\%)$, the proposed SDCSCUSUM chart demonstrates better detection performance compared to the assessed charts. Furthermore, the BSCSCUSUM chart exceeds the DRCSCUSUM chart.

The study's scope can be expanded by using integrated structural frameworks by mixing the EWMA setup control charts with CUSUM setup control charts to detect the very small alterations effectively in the location parameter. Furthermore, future research may investigate the monitoring of the covariance parameter in high-dimensional data or the concurrent observation of location and scale parameters within the high-dimensional context using the proposed methodology.

References

- [1] D. C. Montgomery, *Introduction to statistical quality control*. John Wiley & Sons, 2020.
- [2] E. H. S. Oberoi, M. Parmar, H. Kaur, and R. Mehra, "SPC (statistical process control): a quality control technique for confirmation to ability of process," *International research journal of engineering and technology (IRJET)*, vol. 3, no. 06, pp. 666-672, 2016.
- [3] W. A. Shewhart, "Quality control charts," *The Bell System Technical Journal*, vol. 5, no. 4, pp. 593-603, 1926.
- [4] P. Qiu, *Introduction to statistical process control*. CRC Press, 2013.
- [5] L. S. Nelson, "The Shewhart control chart—tests for special causes," *Journal of quality technology*, vol. 16, no. 4, pp. 237-239, 1984.
- [6] S. W. Human, S. Chakraborti, and C. F. Smit, "Shewhart-type control charts for variation in phase I data analysis," *Computational Statistics & Data Analysis*, vol. 54, no. 4, pp. 863-874, 2010, doi: 10.1016/j.csda.2009.09.026.
- [7] E. Page, "Control charts for the mean of a normal population," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 16, no. 1, pp. 131-135, 1954.
- [8] S. W. Roberts, "Control Chart Tests Based on Geometric Moving Averages," *Technometrics*, vol. 42, no. 1, pp. 97-101, 2000, doi: 10.1080/00401706.2000.10485986.
- [9] K. Hu and J. Yuan, "Multivariate statistical process control based on multiway locality preserving projections," *Journal of Process Control*, vol. 18, no. 7-8, pp. 797-807, 2008, doi: 10.1016/j.jprocont.2007.11.002.
- [10] Z. Wang, Y. Li, and X. Zhou, "A Statistical Control Chart for Monitoring High-dimensional Poisson Data Streams," *Quality and Reliability Engineering International*, vol. 33, no. 2, pp. 307-321, 2017, doi: 10.1002/qre.2005.
- [11] J.-M. Kim, N. Wang, Y. Liu, and K. Park, "Residual Control Chart for Binary Response with Multicollinearity Covariates by Neural Network Model," *Symmetry*, vol. 12, no. 3, 2020, doi: 10.3390/sym12030381.
- [12] C. Kan and H. Yang, "Network models for monitoring high-dimensional image profiles," in *2015 IEEE International Conference on Automation Science and Engineering (CASE)*, 2015: IEEE, pp. 1078-1083.
- [13] S. Jang, S. H. Park, and J.-G. Baek, "Real-time contrasts control chart using random forests with weighted voting," *Expert Systems with Applications*, vol. 71, pp. 358-369, 2017, doi: 10.1016/j.eswa.2016.12.002.
- [14] S. He, W. Jiang, and H. Deng, "A distance-based control chart for monitoring multivariate processes using support vector machines," *Annals of Operations Research*, vol. 263, no. 1-2, pp. 191-207, 2016, doi: 10.1007/s10479-016-2186-4.
- [15] I. B. Khediri and C. Weihs, "Process monitoring using an online nonlinear data reduction based control chart," in *Frontiers in Statistical Quality Control 10*: Springer, 2012, pp. 97-107.
- [16] P. Phaladiganon, S. B. Kim, V. C. Chen, and W. Jiang, "Principal component analysis-based control charts for multivariate nonnormal distributions," *Expert Systems with Applications*, vol. 40, no. 8, pp. 3044-3054, 2013.

- [17] M. Ahsan, M. Mashuri, Wibawati, H. Khusna, and M. H. Lee, "Multivariate control chart based on kernel PCA for monitoring mixed variable and attribute quality characteristics," *Symmetry*, vol. 12, no. 11, p. 1838, 2020.
- [18] G. Capizzi and G. Masarotto, "A least angle regression control chart for multidimensional data," *Technometrics*, vol. 53, no. 3, pp. 285-296, 2011.
- [19] F. A. P. Peres and F. S. Fogliatto, "Variable selection methods in multivariate statistical process control: A systematic literature review," *Computers & Industrial Engineering*, vol. 115, pp. 603-619, 2018, doi: 10.1016/j.cie.2017.12.006.
- [20] C. Zou, W. Jiang, and F. Tsung, "A LASSO-based diagnostic framework for multivariate statistical process control," *Technometrics*, vol. 53, no. 3, pp. 297-309, 2011.
- [21] K. Wang and W. Jiang, "High-dimensional process monitoring and fault isolation via variable selection," *Journal of Quality Technology*, vol. 41, no. 3, pp. 247-258, 2009.
- [22] J. R. G. Lansangan and E. B. Barrios, "Simultaneous dimension reduction and variable selection in modeling high dimensional data," *Computational Statistics & Data Analysis*, vol. 112, pp. 242-256, 2017.
- [23] T. Mahmood, F. Ahmed, M. Riaz, and N. Abbas, "High-dimensional control charts with application to surveillance of grease damage in bearings of wind turbines," *Production & Manufacturing Research*, vol. 12, no. 1, 2024, doi: 10.1080/21693277.2024.2377739.
- [24] M. S. Srivastava and M. J. J. o. M. A. Du, "A test for the mean vector with fewer observations than the dimension," vol. 99, no. 3, pp. 386-402, 2008.
- [25] Z. Bai and H. Saranadasa, "Effect of high dimension: by an example of a two sample problem," *Statistica Sinica*, pp. 311-329, 1996.
- [26] A. P. Dempster, "A high dimensional two sample significance test," *The Annals of Mathematical Statistics*, pp. 995-1010, 1958.
- [27] A. P. Dempster, "A significance test for the separation of two highly multivariate small samples," *Biometrics*, vol. 16, no. 1, pp. 41-50, 1960.
- [28] Z. Tian, X. Cui, D. Wang, D. Niu, and H. Zhang, "Development of High-dimensional Data Sparse Modeling in Data Space and Its Application in Manufacturing Multi-value Chain Collaboration," in *2022 IEEE 25th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, 2022: IEEE, pp. 71-76.
- [29] K. Huang, Y. Wu, C. Wang, Y. Xie, C. Yang, and W. Gui, "A projective and discriminative dictionary learning for high-dimensional process monitoring with industrial applications," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 1, pp. 558-568, 2020.
- [30] S. Cherlin *et al.*, "Utilising high-dimensional data in randomised clinical trials: A review of methods and practice," *Research Methods in Medicine & Health Sciences*, vol. 5, no. 4, pp. 110-124, 2024.
- [31] H. Momeni and A. Ebrahimkhanlou, "High-dimensional data analytics in structural health monitoring and non-destructive evaluation: A review paper," *Smart Materials and Structures*, vol. 31, no. 4, p. 043001, 2022.
- [32] B. Robson, S. Boray, and J. Weisman, "Mining real-world high dimensional structured data in medicine and its use in decision support. Some different perspectives on unknowns, interdependency, and distinguishability," *Computers in Biology and Medicine*, vol. 141, p. 105118, 2022.
- [33] E. Maboudou-Tchao, C. W. Harrison, and S. Sen, "A comparison study of penalized likelihood via regularization and support vector-based control charts," *Quality Technology & Quantitative Management*, vol. 20, no. 2, pp. 147-167, 2023.

- [34] J. Yu, Y. Mo, and S. J. Qin, "Latent Dynamic Networked System Identification with High-Dimensional Networked Data," in *2023 62nd IEEE Conference on Decision and Control (CDC)*, 2023: IEEE, pp. 461-466.
- [35] S. Zhang, Y. Zhou, P. Geng, and Q. Lu, "Functional Neural Networks for High-Dimensional Genetic Data Analysis," *IEEE/ACM Trans Comput Biol Bioinform*, vol. 21, no. 3, pp. 383-393, May-Jun 2024, doi: 10.1109/TCBB.2024.3364614.
- [36] T. Ando and J. Bai, "Clustering huge number of financial time series: A panel data approach with high-dimensional predictors and factor structures," *Journal of the American Statistical Association*, vol. 112, no. 519, pp. 1182-1198, 2017.
- [37] S. Basu, S. Das, G. Michailidis, and A. Purnanandam, "A high-dimensional approach to measure connectivity in the financial sector," *The Annals of Applied Statistics*, vol. 18, no. 2, pp. 922-945, 2024.
- [38] Y. Wang, G. Chen, and M. Maggioni, "High-dimensional data modeling techniques for detection of chemical plumes and anomalies in hyperspectral images and movies," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 9, pp. 4316-4324, 2016.
- [39] J. O. Hendrickx, J. van Gastel, H. Leysen, B. Martin, and S. Maudsley, "High-dimensionality data analysis of pharmacological systems associated with complex diseases," *Pharmacological reviews*, vol. 72, no. 1, pp. 191-217, 2020.
- [40] A. Mukherjee and M. Marozzi, "Nonparametric Phase-II control charts for monitoring high-dimensional processes with unknown parameters," *Journal of Quality Technology*, vol. 54, no. 1, pp. 44-64, 2021.
- [41] W. Zhao, Z. Wang, and C. Wu, "Adaptive multivariate EWMA charts for monitoring sparse mean shifts based on parameter optimization design," *Journal of Statistical Computation and Simulation*, vol. 91, no. 13, pp. 2670-2683, 2021/09/02 2021, doi: 10.1080/00949655.2021.1904242.
- [42] J. Guo, H. Yan, and C. Zhang, "A Bayesian partially observable online change detection approach with Thompson sampling," *Technometrics*, vol. 65, no. 2, pp. 179-191, 2023.
- [43] M. Biswas and A. K. Ghosh, "A nonparametric two-sample test applicable to high dimensional data," *Journal of Multivariate Analysis*, vol. 123, pp. 160-171, 2014.
- [44] F. Ahmed, T. Mahmood, M. Riaz, and N. Abbas, "Comprehensive review of high-dimensional monitoring methods: trends, insights, and interconnections," *Quality Technology & Quantitative Management*, 2024, doi: 10.1080/16843703.2024.2395745.
- [45] A. Erem and T. Mahmood, "A bivariate CUSUM control chart based on exceedance statistics," *Quality and Reliability Engineering International*, vol. 39, no. 4, pp. 1172-1191, 2023, doi: 10.1002/qre.3285.
- [46] J. Chen, C. Park, S.-H. Kim, and Y. Xie, "To reduce or not to reduce: a study on spatio-temporal surveillance," *Environmental and Ecological Statistics*, vol. 26, no. 3, pp. 217-238, 2019, doi: 10.1007/s10651-019-00425-4.
- [47] J. Chen, H. Yang, and J. Yao, "A new multivariate CUSUM chart using principal components with a revision of Crosier's chart," *Communications in Statistics-Simulation and Computation*, vol. 47, no. 2, pp. 464-476, 2018.
- [48] W. Li, C. Zhang, F. Tsung, and Y. Mei, "Nonparametric monitoring of multivariate data via KNN learning," *International Journal of Production Research*, vol. 59, no. 20, pp. 6311-6326, 2021.

- [49] Z. Wang, X. Li, Y. Wang, Y. Ma, and L. Xue, "Monitoring of high-dimensional and high-frequency data streams: A nonparametric approach," *Quality Technology & Quantitative Management*, pp. 1-20, 2024, doi: 10.1080/16843703.2024.2356970.
- [50] M. Shaohui, G. Tuerhong, M. Wushouer, and T. Yibulayin, "PCA mix-based Hotelling's T2 multivariate control charts for intrusion detection system," *IET Information Security*, vol. 16, no. 3, pp. 161-177, 2022.
- [51] G. J. Odom, K. B. Newhart, T. Y. Cath, and A. S. Hering, "Multistate multivariate statistical process control," *Applied Stochastic Models in Business and Industry*, vol. 34, no. 6, pp. 880-892, 2018.
- [52] D. Sun, G. Lu, H. Zhou, and Y. Yan, "Condition Monitoring of Combustion Processes Through Flame Imaging and Kernel Principal Component Analysis," *Combustion Science and Technology*, vol. 185, no. 9, pp. 1400-1413, 2013/09/02 2013, doi: 10.1080/00102202.2013.798316.
- [53] S. Zhang, L. Xue, Z. He, Y. Liu, and Z. Xin, "A sensitized variable selection control chart based on a classification algorithm for monitoring high-dimensional processes," *Quality and Reliability Engineering International*, vol. 39, no. 7, pp. 2837-2850, 2023, doi: 10.1002/qre.3393.
- [54] G. M. Abdella, K. N. Al-Khalifa, S. Kim, M. K. Jeong, E. A. Elsayed, and A. M. Hamouda, "Variable selection-based multivariate cumulative sum control chart," *Quality and Reliability Engineering International*, vol. 33, no. 3, pp. 565-578, 2017, doi: 10.1002/qre.2041.
- [55] I. Ullah, M. D. Pawley, A. N. Smith, and B. Jones, "Improving the detection of unusual observations in high-dimensional settings," *Australian & New Zealand Journal of Statistics*, vol. 59, no. 4, pp. 449-462, 2017.
- [56] M. R. Ahmad and S. E. Ahmed, "On the distribution of the T2 statistic, used in statistical process monitoring, for high-dimensional data," *Statistics & Probability Letters*, vol. 168, p. 108919, 2021.
- [57] J. Li, L. Kong, B. Jiang, and W. Tu, "High-dimensional outlier detection and variable selection via adaptive weighted mean regression," *arXiv preprint arXiv:2306.13850*, 2023.
- [58] S. Kim, M. Turkoz, and J. W. Baek, "High-Speed Monitoring of Multidimensional Processes Using Bayesian Updates," *IEEE Access*, vol. 10, pp. 97450-97464, 2022.
- [59] H. Ye, Z. Zheng, J.-R. C. Cheng, B. Hable, and K. Liu, "Online monitoring of high-dimensional asynchronous and heterogeneous data streams for shifts in location and scale," *International Journal of Production Research*, vol. 62, no. 3, pp. 720-736, 2024.
- [60] A. B. Moala, L. L. Ho, and R. C. Quinino, "Multivariate control charts to monitor the monthly frequency of vehicle robberies in São Paulo city," *Spatial statistics*, vol. 29, pp. 49-65, 2019.
- [61] H. Ali, Z. Zhang, and F. Gao, "Multiscale monitoring of industrial chemical process using wavelet-entropy aided machine learning approach," *Process Safety and Environmental Protection*, vol. 180, pp. 1053-1075, 2023.
- [62] Z. Yu, W. Wang, X. Wang, Q. Jiang, and G. Wang, "Nonlinear dynamic process monitoring based on latent mapping embedding deep neural networks," *The Canadian Journal of Chemical Engineering*, 2024, doi: 10.1002/cjce.25461.

- [63] E. A. Shah, M. Xiao, R. Ming, and A. Saghir, "A shrinkage based diagonal T2 control chart for high dimensional data," *Journal of Statistical Computation and Simulation*, pp. 1-30, 2024, doi: 10.1080/00949655.2024.2424344.
- [64] D. Kakde, S. Peredriy, and A. Chaudhuri, "A non-parametric control chart for high frequency multivariate data," in *2017 Annual Reliability and Maintainability Symposium (RAMS)*, 2017: IEEE, pp. 1-6.
- [65] O. Bodnar and T. Bodnar, "CUSUM Schemes for Stationary and Nonstationary Gaussian Processes," in *Advanced Statistical Methods in Process Monitoring, Finance, and Environmental Science: Essays in Honour of Wolfgang Schmid*: Springer, 2024, pp. 39-57.
- [66] C. Zhang, N. Chen, and J. Wu, "Spatial rank-based high-dimensional monitoring through random projection," *Journal of Quality Technology*, vol. 52, no. 2, pp. 111-127, 2020/04/02 2020, doi: 10.1080/00224065.2019.1571336.
- [67] Z. Bai and H. Saranadasa, "Effect of high dimension: by an example of a two sample problem," *Statistica Sinica*, vol. 6, no. 2, pp. 311-329, 1996.
- [68] M. S. Srivastava and M. Du, "A test for the mean vector with fewer observations than the dimension," *Journal of Multivariate Analysis*, vol. 99, no. 3, pp. 386-402, 2008.
- [69] E. S. Page, "Controlling the Standard Deviation by Cusums and Warning Lines," *Technometrics*, vol. 5, pp. 307-315, 1963, doi: 10.1080/00401706.1963.10490100.
- [70] D. C. Montgomery, *Introduction to statistical quality control*, Sixth ed. John Wiley & Sons, 2009.
- [71] Z. G. Stoumbos and J. H. Sullivan, "Robustness to Non-Normality of the Multivariate EWMA Control Chart," *Journal of Quality Technology*, vol. 34, no. 3, pp. 260-276, 2018, doi: 10.1080/00224065.2002.11980157.
- [72] C. S. Pun, "Low-and high-dimensional stock price data (Mendeley data, v3)," ed, 2017.
- [73] C. Z. Mooney, R. D. Duval, and R. Duvall, *Bootstrapping: A nonparametric approach to statistical inference* (no. 95). sage, 1993.
- [74] R. Liu *et al.*, "A self-temperature-modulated quadrilateral gas sensor for gas identification," *IEEE Sensors Journal*, vol. 23, no. 8, pp. 8017-8027, 2023.
- [75] P. V. Shinde and C. S. Rout, "Magnetic gas sensing: Working principles and recent developments," *Nanoscale Advances*, vol. 3, no. 6, pp. 1551-1568, 2021.
- [76] F. W. Wibowo, "Electronic Nose in Classification of Gas Sensor Array Detection Through Flow Modulation," in *2021 4th International Conference on Information and Communications Technology (ICOIACT)*, 2021: IEEE, pp. 243-247.
- [77] A. Ziyatdinov, J. Fonollosa, L. Fernández, A. Gutiérrez-Gálvez, S. Marco, and A. Perera, "Data set from gas sensor array under flow modulation," *Data in Brief*, vol. 3, pp. 131-136, 2015.
- [78] F. Ahmed, T. Mahmood, M. Riaz, and N. Abbas, "Comprehensive review of high-dimensional monitoring methods: trends, insights, and interconnections," *Quality Technology & Quantitative Management*, pp. 1-25, 2024, doi: 10.1080/16843703.2024.2395745.
- [79] G. M. Abdella, M. R. Maleki, S. Kim, K. N. Al-Khalifa, and A. M. S. Hamouda, "Phase-I monitoring of high-dimensional covariance matrix using an adaptive thresholding LASSO rule," *Computers & Industrial Engineering*, vol. 144, 2020, doi: 10.1016/j.cie.2020.106465.

- [80] A. Mukherjee and M. Marozzi, "Nonparametric Phase-II control charts for monitoring high-dimensional processes with unknown parameters," *Journal of Quality Technology*, vol. 54, no. 1, pp. 44-64, 2020, doi: 10.1080/00224065.2020.1805378.
- [81] E. Maboudou-Tchao, C. W. Harrison, and S. Sen, "A comparison study of penalized likelihood via regularization and support vector-based control charts," *Quality Technology & Quantitative Management*, vol. 20, no. 2, pp. 147-167, 2022, doi: 10.1080/16843703.2022.2096198.
- [82] A. M. E. Gómez, D. Li, and K. Paynabar, "An Adaptive Sampling Strategy for Online Monitoring and Diagnosis of High-Dimensional Streaming Data," *Technometrics*, vol. 64, no. 2, pp. 253-269, 2021, doi: 10.1080/00401706.2021.1967198.
- [83] H. Yan, K. Paynabar, and J. Shi, "Real-Time Monitoring of High-Dimensional Functional Data Streams via Spatio-Temporal Smooth Sparse Decomposition," *Technometrics*, vol. 60, no. 2, pp. 181-197, 2018/04/03 2018, doi: 10.1080/00401706.2017.1346522.
- [84] E. M. Maboudou-Tchao, "High-dimensional data monitoring using support machines," *Communications in Statistics - Simulation and Computation*, vol. 50, no. 7, pp. 1927-1942, 2019, doi: 10.1080/03610918.2019.1588312.
- [85] Z. Fang, W. Li, X. Liu, X. Pu, and D. Xiang, "Online monitoring of high-dimensional binary data streams with application to extreme weather surveillance," *J Appl Stat*, vol. 49, no. 16, pp. 4122-4136, 2022, doi: 10.1080/02664763.2021.1971633.
- [86] J. Olawale Ajadi and M. Riaz, "New Memory-Type Control Charts for Monitoring Process Mean and Dispersion," *Scientia Iranica*, vol. 0, no. 0, pp. 0-0, 2017, doi: 10.24200/sci.2017.4483.
- [87] J. M. Lucas and M. S. Saccucci, "Exponentially weighted moving average control schemes: properties and enhancements," *Technometrics*, vol. 32, no. 1, pp. 1-12, 1990.
- [88] P. Ray, S. S. Reddy, and T. Banerjee, "Various dimension reduction techniques for high dimensional data analysis: a review," *Artificial Intelligence Review*, vol. 54, no. 5, pp. 3473-3515, 2021, doi: 10.1007/s10462-020-09928-0.
- [89] J. P. George, Z. Chen, and P. Shaw, "Fault detection of drinking water treatment process using PCA and Hotelling's T2 chart," *International Journal of Computer and Information Engineering*, vol. 3, no. 2, pp. 430-435, 2009.
- [90] F. Shahzad, Z. Huang, and W. H. Memon, "Process Monitoring Using Kernel PCA and Kernel Density Estimation-Based SSGLR Method for Nonlinear Fault Detection," *Applied Sciences*, vol. 12, no. 6, 2022, doi: 10.3390/app12062981.
- [91] I. K. Prasetya, M. Ahsan, M. Mashuri, and M. H. Lee, "Bootstrap and MRCD Estimators in Hotelling's T2 Control Charts for Precise Intrusion Detection," *Applied Sciences*, vol. 14, no. 17, 2024, doi: 10.3390/app14177948.
- [92] O. A. Oyegoke, K. S. Adekeye, J. O. Olaomi, and J. C. Malela-Majika, "Hotelling T2 control chart based on minimum vector variance for monitoring high-dimensional correlated multivariate process," *Quality and Reliability Engineering International*, 2024, doi: 10.1002/qre.3687.
- [93] M. Shaohui, G. Tuerhong, M. Wushouer, and T. Yibulayin, "PCA mix-based Hotelling's T2multivariate control charts for intrusion detection system," *IET Information Security*, vol. 16, no. 3, pp. 161-177, 2021, doi: 10.1049/ise2.12051.
- [94] J. Yu and C. Zhang, "Manifold regularized stacked autoencoders-based feature learning for fault detection in industrial processes," *Journal of Process Control*, vol. 92, pp. 119-136, 2020/08/01/ 2020, doi: <https://doi.org/10.1016/j.procont.2020.06.001>.

- [95] J. Yu and X. Liu, "One-dimensional residual convolutional auto-encoder for fault detection in complex industrial processes," *International Journal of Production Research*, vol. 60, no. 18, pp. 5655-5674, 2021, doi: 10.1080/00207543.2021.1968061.
- [96] M. Riaz, B. Zaman, I. A. Raji, M. H. Omar, R. Mehmood, and N. Abbas, "An Adaptive EWMA Control Chart Based on Principal Component Method to Monitor Process Mean Vector," *Mathematics*, vol. 10, no. 12, 2022, doi: 10.3390/math10122025.
- [97] E. K. Epprecht, F. Aparisi, and O. Ruiz, "Optimum variable-dimension EWMA chart for multivariate statistical process control," *Quality Engineering*, vol. 30, no. 2, pp. 268-282, 2017, doi: 10.1080/08982112.2017.1358367.
- [98] W. Jiang, K. Wang, and F. Tsung, "A Variable-Selection-Based Multivariate EWMA Chart for Process Monitoring and Diagnosis," *Journal of Quality Technology*, vol. 44, no. 3, pp. 209-230, 2017, doi: 10.1080/00224065.2012.11917896.
- [99] D. Yan, S. Zhang, and U. Jung, "A variable-selection control chart via penalized likelihood and Gaussian mixture model for multimodal and high-dimensional processes," *Quality and Reliability Engineering International*, vol. 35, no. 4, pp. 1263-1275, 2019, doi: 10.1002/qre.2458.
- [100] K. Wang and W. Jiang, "High-Dimensional Process Monitoring and Fault Isolation via Variable Selection," *Journal of Quality Technology*, vol. 41, no. 3, pp. 247-258, 2017, doi: 10.1080/00224065.2009.11917780.
- [101] K. Sangahn, "Variable selection-based SPC procedures for high-dimensional multistage processes," *Journal of Systems Engineering and Electronics*, vol. 30, no. 1, pp. 144-153, 2019, doi: 10.21629/JSEE.2019.01.14.
- [102] I. Ullah, M. D. M. Pawley, A. N. H. Smith, and B. Jones, "Improving the detection of unusual observations in high-dimensional settings," *Australian & New Zealand Journal of Statistics*, vol. 59, no. 4, pp. 449-462, 2017, doi: 10.1111/anzs.12210.
- [103] A. A. El-Sheikh, M. C. Ali, and M. R. Abonazel, "Development of Two Methods for Estimating High-Dimensional Data in the Case of Multicollinearity and Outliers," *International Journal of Analysis and Applications*, vol. 22, 2024, doi: 10.28924/2291-8639-22-2024-187.
- [104] F. Ahmed, "On statistical surveillance of two-sampled high-dimensional data. Master's Thesis, Department of Mathematics, King Fahd University of Petroleum and Minerals, KSA., 2023.
- [105] M. a. R. Abujiya, M. Riaz, and M. H. Lee, "Enhancing the Performance of Combined Shewhart-EWMA Charts," *Quality and Reliability Engineering International*, vol. 29, no. 8, pp. 1093-1106, 2012, doi: 10.1002/qre.1461.
- [106] G. Capizzi and G. Masarotto, "Combined Shewhart–EWMA control charts with estimated parameters," *Journal of Statistical Computation and Simulation*, vol. 80, no. 7, pp. 793-807, 2010, doi: 10.1080/00949650902773585.
- [107] R. A. Sanusi, M. a. R. Abujiya, M. Riaz, and N. Abbas, "Combined Shewhart CUSUM charts using auxiliary variable," *Computers & Industrial Engineering*, vol. 105, pp. 329-337, 2017, doi: 10.1016/j.cie.2017.01.018.
- [108] J. M. Lucas, "Combined Shewhart-CUSUM quality control schemes," *Journal of quality technology*, vol. 14, no. 2, pp. 51-59, 1982, doi: <https://doi.org/10.1080/00224065.1982.11978790>.

- [109] D. G. Godase and S. B. Mahadik, "The combined Shewhart-CUSUM sign charts," *Communications in Statistics - Simulation and Computation*, vol. 53, no. 1, pp. 357-366, 2021, doi: 10.1080/03610918.2021.2020287.
- [110] A. Tang, Y. Wang, X. Wang, and X. Hu, "A Distribution-Free CUSUM Scheme for Monitoring Location and Scale in High-Dimensional Processes," presented at the 2022 34th Chinese Control and Decision Conference (CCDC), 2022.
- [111] S. Lee, M. Kwak, K.-L. Tsui, and S. B. Kim, "Process monitoring using variational autoencoder for high-dimensional nonlinear processes," *Engineering Applications of Artificial Intelligence*, vol. 83, pp. 13-27, 2019, doi: 10.1016/j.engappai.2019.04.013.
- [112] A. Maged, C. F. Lui, S. Haridy, and M. Xie, "Variational AutoEncoders-LSTM based fault detection of time-dependent high dimensional processes," *International Journal of Production Research*, vol. 62, no. 4, pp. 1092-1107, 2023, doi: 10.1080/00207543.2023.2175591.
- [113] J. Kim, M. K. Jeong, E. A. Elsayed, K. N. Al-Khalifa, and A. M. S. Hamouda, "An adaptive step-down procedure for fault variable identification," *International Journal of Production Research*, vol. 54, no. 11, pp. 3187-3200, 2015, doi: 10.1080/00207543.2015.1076948.
- [114] A. Ahmadi-Javid and M. Ebadi, "A two-step method for monitoring normally distributed multi-stream processes in high dimensions," *Quality Engineering*, vol. 33, no. 1, pp. 143-155, 2020, doi: 10.1080/08982112.2020.1786118.
- [115] K. Liu, Y. Mei, and J. Shi, "An Adaptive Sampling Strategy for Online High-Dimensional Process Monitoring," *Technometrics*, vol. 57, no. 3, pp. 305-319, 2014, doi: 10.1080/00401706.2014.947005.
- [116] M. a. R. Abujiya, M. Riaz, and M. H. Lee, "A New Combined Shewhart–Cumulative Sum S Chart for Monitoring Process Standard Deviation," *Quality and Reliability Engineering International*, vol. 32, no. 3, pp. 1149-1165, 2015, doi: 10.1002/qre.1822.
- [117] Z. Wu, M. Yang, W. Jiang, and M. B. C. Khoo, "Optimization designs of the combined Shewhart-CUSUM control charts," *Computational Statistics & Data Analysis*, vol. 53, no. 2, pp. 496-506, 2008, doi: 10.1016/j.csda.2008.08.032.
- [118] C. S. Pun and H. Y. Wong, "A linear programming model for selection of sparse high-dimensional multiperiod portfolios," *European Journal of Operational Research*, vol. 273, no. 2, pp. 754-771, 2019, doi: 10.1016/j.ejor.2018.08.025.

Vitae

Name : OSAMA ALHADI

Nationality : SUDAN

Date of Birth : 1/1/1996

Address : Om Bada, Omdurman, P.O. Box: 11121, Khartoum, Sudan

Academic Background : B.Sc. in Statistics and Computation, Al Neelain University, Sudan

Master of Science in Statistics, Al Neelain University, Sudan

Professional Experience: Teacher Assistant of Statistics, Al Neelain University, Sudan

Research :

- **High-Dimensional Cumulative Sum Control Charts for Monitoring Industrial and Chemical Processes. (Research paper under review)**
- **Combined Shewhart-EWMA (CSEWMA) Control Charts for Two-sample High-Dimensional Data. (Research paper under review)**
- **Combined Shewhart-CUSUM (CSCUSUM) Control Charts for Two-sample High-Dimensional Data. (To be submitted)**