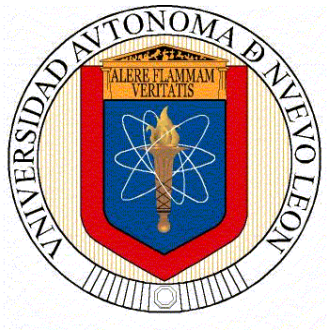
UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS



TESIS

NONPARAMETRIC MULTIVARIATE PROCESS MONITORING WITH GUARANTEED IN CONTROL PERFORMANCE FOR CHANGES IN LOCATION

PRESENTADA POR

JORGE LUIS MERLO MEJÍA

EN OPCIÓN AL GRADO DE

DOCTOR EN CIENCIAS CON ORIENTACIÓN EN MATEMÁTICAS

2026

Universidad Autónoma de Nuevo León

Facultad de Ciencias Físico Matemáticas

Posgrado en Ciencias con Orientación en Matemáticas

Nonparametric Multivariate Process Monitoring with Guaranteed In Control Performance for Changes in Location

por

Jorge Luis Merlo Mejía

como requisito parcial para obtener el grado de

DOCTOR EN CIENCIAS CON ORIENTACIÓN EN MATEMÁTICAS

Universidad Autónoma de Nuevo León
Facultad de Ciencias Físico Matemáticas
Posgrado en Ciencias con Orientación en Matemáticas

Los miembros del Comité de Tesis recomendamos que la Tesis "Nonparametric Multivariate Process Monitoring with Guaranteed In Control Performance for Changes in Location", realizada por el alumno Jorge Luis Merlo Mejía, con número de matrícula 1487895, sea aceptada para su defensa como opción al grado de Doctor en Ciencias con Orientación en Matemáticas.

El Comité de Tesis

Dr. Álvaro Eduardo Cordero Franco
Asesor

Dr. Víctor Tercero Gómez
Co-Director

Dr. Jorge Arturo Garza Venegas
Revisor

Dr. Omar Jorge Ibarra Rojas
Revisor

Dra. Perla Marlene Viera Gonzalez
Revisor

Dr. Álvaro Eduardo Cordero Franco
Subdirector de Estudios de Posgrado

Dedication

Dedico esta tesis a la memoria de mi padre, Jorge Merlo Cuadros, a mi madre, Eneida Mejía Castro, y a mi hermana, Melissa Merlo, por haberme dado siempre su amor, apoyo y motivación para seguir adelante.

Asimismo, dedico este trabajo a mi pareja, Celeste, por su amor incondicional, apoyo constante, compañía durante todo este proceso.

Acknowledgments

Expreso mi más sincero agradecimiento al Dr. Eduardo Cordero Franco, por la confianza depositada, su orientación y su apoyo continuo como director de esta tesis. Aprecio profundamente su disposición para compartir su conocimiento, su paciencia a lo largo de este proyecto y el tiempo dedicado a la guía y desarrollo de mi trabajo.

Asimismo, agradezco al Dr. Víctor Tercero Gómez por la confianza otorgada a esta investigación, así como por su apoyo académico, sus valiosos consejos y su constante aliento durante este proceso.

En particular, expreso mi gratitud al *Profe Soto*, quien continúa presente a través de sus enseñanzas, los cuales marcaron de manera significativa mi formación.

Agradezco a la Facultad de Ciencias Físico Matemáticas y a la Universidad Autónoma de Nuevo León por proporcionar el entorno académico y las oportunidades que hicieron posible mi desarrollo profesional y personal.

Finalmente, expreso mi reconocimiento al CONACYT por el apoyo financiero otorgado mediante la beca de manutención durante mis estudios. Sin este respaldo, la culminación de este trayecto académico no habría sido posible.

Contents

Abstract

Multivariate statistical process control (MSPC) addresses the concurrent monitoring of several measurements. Since multivariate normality is rare in practice, nonparametric schemes become useful alternatives. Inspired by the development of distribution free Shewhart-type control charts to monitor multivariate data streams as a mean to address high-dimensional processes where distance measures are ranked and compared using two-sample test statistics, a novel distribution-free multivariate chart for location that combines Mahalanobis distances with the Mann-Whitney statistic is proposed. By considering the practitioner-to-practitioner variation, probably for the first time in nonparametric MSPC, we account for the uncertainty of the run length distribution, where computational complexity is addressed using the cumulative conditional false alarm probability as a proxy of the conditional average run length. This approach enables faster determination of control limits that guarantee an in-control performance given a Phase-I sample, a common situation in parametric approaches but scarcely discussed in nonparametric monitoring. Since no direct alternative exists, results are compared with the traditional approach to show the benefits of a guarantee in-control performance with comparable sensitivity when dealing with medium to large shifts. An implementation is illustrated in a real scenario related to Wine Quality Data.

Keywords: Multivariate Control Chart, Distribution-free, Practitioner-to-Practitioner Variation, Guaranteed Control Limits

Chapter 1

Introduction

Due to the rapid technological advances of data acquisition tools, we can collect a large amount of information in real time. This, along with the development of adequate statistical methodologies, have allowed us to control and monitor production process quality in a way that satisfies the requirements established by the consumers. These methodologies are under the label of Statistical Process Control where the control charts are mainly used for monitoring. A control chart monitors sequentially the quality characteristics of a product, ensuring the stability of a production process. If any deviations from the established requirements are presented an alarm will be issued the soonest possible. To decide whether a process is conforming or not, SPC is divided into two different phases: Phase I and Phase II. Objective of Phase I is to set up a process to make it run properly, guaranteeing its stability. Since this is typically the first thing that is done when SPC is applied, not much knowledge about the process is available. Therefore, a retrospective analysis is made when establishing Phase I conditions, a control chart analyzes different datasets obtained iteratively under different process parameters. In each iteration unusual patterns are detected by the control chart and adjustments are made to the process in order to increase its stability. The objective is to obtain a final dataset and parameters characteristic of a process that works stably. This final dataset obtained is called "in control" (IC) sample, by this stage enough observations have been collected to establish the IC distribution. In Phase II analysis, once the IC conditions have been defined an online sequential monitoring of the process is performed. The objective is to ensure that the process to be monitored remains IC, and a signal will be triggered as soon as possible by the correct detection of any process variation that leads to "out of control" (OC) conditions. This translates in an adequate detection of any potential shift in mean, scale or both of the IC process distribution.

Consider a process whose IC observations follow an F distribution with mean μ_0 and standard deviation σ_0. When monitoring the mean of a process, it is natural to consider

the following hypothesis test:

$$H_0 : \mu_i = \mu_0 \quad \text{versus} \quad H_1 : \mu_i \neq \mu_0, \forall i$$

where μ_i is the real process mean at time i. Common monitoring schemes are useful when dealing with normal and univariate data. That is the case of the $\bar{X}$ control chart which is designed assuming that IC μ_0 and σ_0 of the process are both known. Then, at each monitoring time point i we collect a sample of n observations $x_1, x_2, \ldots, x_n$ and compute its average $\bar{X}_i = (\sum_{j=1}^{n} x_j)/n$, since we know that $\bar{X} \sim N(\mu_0, \sigma_0/n)$, a good test statistic for testing the previous hypothesis is

$$Z^* = \frac{\bar{X}_i - \mu_0}{\sigma/\sqrt{n}} \overset{H_0}{\sim} N(0,1) \tag{1.1}$$

H_0 is rejected in a pre-specified level of significance α if $|Z^*| > Z_{1-\alpha/2}$ where $Z_{1-\alpha/2}$ is the $(1-\alpha/2)$ quantile of standard normal distribution. Therefore, the process will be declared as OC at the $i-$th time point if

$$\bar{X}_i < LCL = \mu_0 - Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \quad \text{or} \quad \bar{X}_i > UCL = \mu_0 + Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \tag{1.2}$$

where LCL and UCL stand for lower and upper control limits, respectively. Usually a value $\alpha = 0.0027$ is established, therefore the interval $[LCL, UCL]$ covers approximately 99.73% of the population. Figure 1.1 illustrates an implementation of this control chart for a simulated process with a $N(\mu_0, \sigma_0/n)$ distribution where $\mu_0 = 10$, $\sigma_0 = 1$ and a subgroup size $n = 5$.

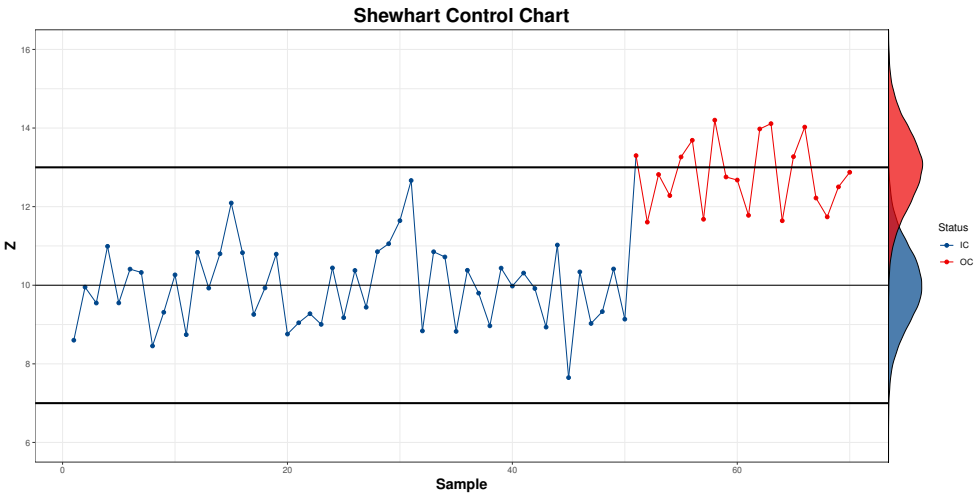


Figure 1.1: Classical univariate $\bar{X}$ control chart used in the detection of a mean shift process that follows a normal distribution. The two black width lines indicate the control limits, stated as $\pm 3\sigma$ from the in control mean.

The previous control chart is an example of a Shewhart's chart; where each monitored sample is compared with the IC distribution by using a test statistic. It is known that these charts are good at detecting large to medium changes in the process being monitored. Other commonly used approaches are the Cumulative Sum (CUSUM), Page (1954) and the Exponentially Weighted Moving Average (EWMA), Roberts (1959) control charts, which are efficient alternatives to overcome the problem of detecting small to moderate shifts in the process. In Qiu (2014), we can found a detailed description of the univariate monitoring schemes mentioned above, as well as the implications that need to be taken into account when the parameters of the distribution are unknown, a very common scenario in practice. In most SPC applications quality tends to be evaluated upon many characteristics, hence, the need for multivariate statistical process control (MSPC). In practice, several process quality characteristics are related to another. Although we could apply a control chart for each quality characteristic, this can be very misleading since we are not considering the correlation between them. To illustrate this, an example with 30 simulated bivariate observations (x_1, x_2) is presented in Figure 1.2. We consider a bivariate normal distribution with mean vector $\boldsymbol{\mu}_0 = [0, 0]$ for the first 29 observations and a mean vector $\boldsymbol{\mu}_1 = [2, 0]$ for the remaining one. As for the scale parameters, we set a matrix $\boldsymbol{\Sigma}_0 = \left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)$ where the correlation factor is $\rho = 0.5$.

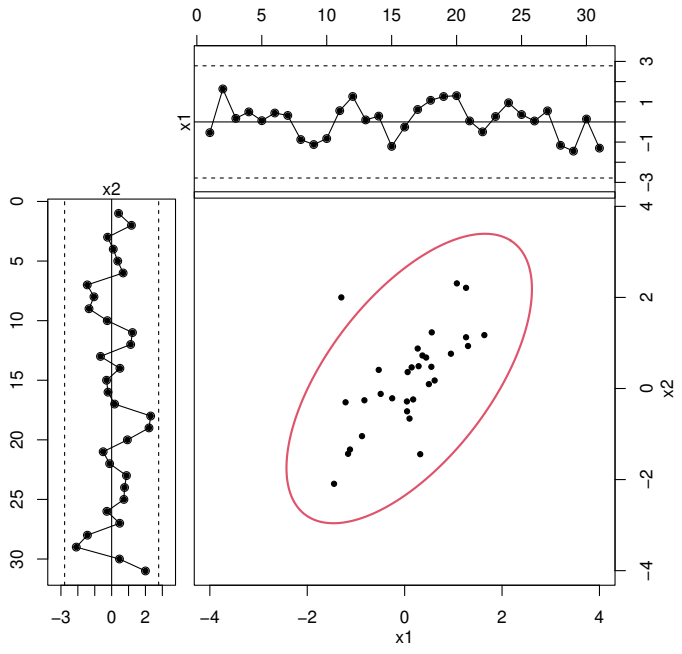


Figure 1.2: Individual control charts for x_1 and x_2 variables, black dashed lines indicate the control limits. Scatter plot for the joint distribution (x_1, x_2), solid red line ellipse indicates the 95 % confidence interval for a bivariate normal distribution with mean $\boldsymbol{\mu}_0 = [0, 0]$.

Since x_1 and x_2 are individual measurements, each can be monitored separately using a classic $\bar{X}$-chart, as defined in Equation 1.1. The process is deemed "in control" (IC) only if both $\bar{X}_1$ and $\bar{X}_2$ statistics fall within their respective control limits. This approach is visualized in Figure 1.2, where all monitoring points are classified as IC.

However, when examining the scatter plot of the joint distribution x_1 and x_2, it becomes evident that an outlier exists, one that is not identified by the individual control charts. This highlights the limitations of univariate monitoring. Consequently, it is more effective to treat all multivariate observations as p-dimensional vectors, enabling a more comprehensive analysis of the data.

1.1 Problem Statement

As a Phase II approach, lets assume there are m reference observations independent and identically distributed (iid), $\boldsymbol{X}_{-m+1}, \ldots, \boldsymbol{X}_0 \in \mathbb{R}^p$, for some integer $p \geq 1$, and $\boldsymbol{X}_i$ is the i-th observation which is collected over time in the following model,

$$\boldsymbol{X}_i \overset{iid}{\sim} \begin{cases} F_0(\boldsymbol{X}; \boldsymbol{\mu}_0) & \text{para} \quad i = -m+1, \ldots, 0, 1, \ldots, \tau, \\ F_1(\boldsymbol{X}; \boldsymbol{\mu}_1) & \text{para} \quad i = \tau+1, \ldots, \end{cases} \tag{1.3}$$

τ is an unknown change point and $\boldsymbol{\mu}_0 \neq \boldsymbol{\mu}_1$. F_0 and F_1 are the distribution functions IC and OC, respectively. From the Phase-I sample the mean vector and covariance matrix can be estimated as

$$\bar{\boldsymbol{X}} = \frac{1}{m} \sum_{i=1}^{m} \boldsymbol{X}_i \tag{1.4}$$

and

$$\boldsymbol{S} = \frac{1}{m-1} \sum_{i=1}^{m} (\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})', \tag{1.5}$$

respectively. Hotelling's T^2 test statistic, is built under the assumption of normality. This control chart monitors the mean vector of a multivariate process and has an excellent performance as long as the normality assumption is met. According to Equation 1.3, the T^2 control chart is defined as follows:

$$T^2 = (\boldsymbol{X}_i - \bar{\boldsymbol{X}})' \boldsymbol{S}^{-1} (\boldsymbol{X}_i - \bar{\boldsymbol{X}}) \tag{1.6}$$

Assuming $\boldsymbol{X}_i$ follows a multivariate normal distribution, then T^2 follows a $\{p(m+1)(m-1)/m(m-p)\}F_{p,m-p}$ distribution, see Mason et al. (2001). Therefore, a control limit h can be established as a quantile that ensures a significance level α such that, if

$T^2 > h$ an alarm is issued. It is well known that T^2 chart can effectively control both Type I and II error rates when process is multivariate normal. In this case $P(T^2 > p(m+1)(m-1)/m(m-p)F_{p,m-p,1-\alpha}|F_0) = \alpha$, it can be noted that in each observation the probability of a false alarm remains, and the first false alarm follows a geometric distribution with probability α.

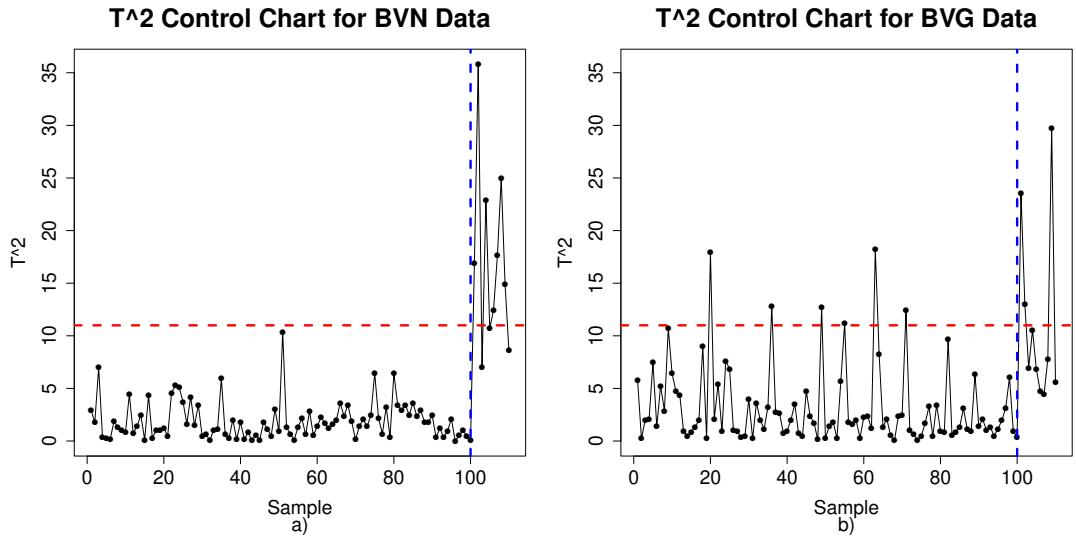


Figure 1.3: Classical Hotelling's T^2 control chart used in the detection of mean shifts for bivariate processes that follows a bivariate normal(BVN) and bivariate gamma (BVG) distribution. The red dashed line represents the control limit and the blue dashed lines indicates the point where the OC conditions start, we can appreciate the presence of false alarms when bivariate data is not normally distributed.

As an example, we apply Hotelling's T^2control chart to monitor the mean vector under bivariate normal (BVN) and bivariate gamma (BVG) scenarios. An in-control (IC) reference sample consisting of 200 observations is simulated for each distribution. After that, a Phase II analysis is carried by monitoring 110 simulated observations where the first 100 ones are IC observations whereas the last 10 observations are considered as OC since they are simulated with a mean shift of 3 standard deviations in the first component of the bivariate data. The objective of this simulation is to illustrate how the Hotelling's T^2 chart can generate false alarms when the assumption of normality is not fulfilled. The expected performance is not to issue any (false) alarm when the process is considered IC and detect as soon as possible any real shift in the process mean that leads to OC conditions (which is the case in Figure 1.3 (a)). As it can be seen from Figure 1.3 (b), the widely used T^2 control chart may lead to wrong conclusions about the real state of the process when departures from normality are presented. Even though the latter might not be representative of a general case, there are several studies that have shown, through a proper statistical analysis, the performance degradation of the T^2 control chart

when the assumption of normality is violated, see Qiu (2008), Phaladiganon et al. (2011), Mukherjee et al. (2017), among others.

Control charts performance are constantly evaluated using the run length (RL) and its average (ARL). RL is defined as the number of consecutive observations or samples that a process takes before it produces a point outside the control limits in a control chart. If the independent observations assumption is met, the RL has a geometric distribution and, therefore, the IC ARL can be calculated as $1/\alpha$ where α is the false alarm rate. In case of OC scenarios a control chart with a lower OC ARL (ARL_1) demonstrates a superior ability to detect shifts.

So far, we have emphasized the importance of developing non-parametric or distribution-free control charts, as these are highly relevant in practical scenarios where non-normality frequently arises across various domains. For instance, they have proven useful in monitoring semiconductor manufacturing processes Chen et al. (2015), analyzing wine quality data Yue and Liu (2017), and controlling data from aluminum electrolytic capacitor production Zou and Tsung (2011). These examples highlight how non-parametric methods address the distributional challenges often encountered in industrial applications. In this research, however, we go further by considering the inherent variability among practitioners, a topic extensively explored in univariate cases but scarcely addressed in MSPC since there is an intrinsic complexity that earlier studies might not have considered.

Whenever a monitoring procedure depends on observations from a Phase-I sample, performance is subject to sample variation. Therefore, when dealing with this analysis it is more appropriate to consider the Conditional Average Run Length, $CARL(x)$, the mean of the RL distribution conditioned to the Phase-I sample $\mathcal{X} = x$. For illustration, we simulate 1000 Phase-I samples of sizes $m = 300, 500$ to later obtain the $CARL(x)$ performance for each of them, this by using the Hotelling's T^2 control chart defined in Equation 1.6 where its control limit is established according to an $ARL_0 = 200$. Results are displayed in Figure 1.4, and as we can see, more than half of simulated Phase-I samples (or practitioners) would get a false alarm before the preespecified threshold ARL_0 for both scenarios, reducing this negative effect when Phase-I sample size is increased. This lead to performance not being guaranteed and practitioners with their corresponding Phase-I samples might obtain a performance way below the expected. Although the example shown here corresponds to a parametric control chart where distribution is assumed as known, this similar behaviour is present in nonparametric approaches, see Celano and Chakraborti (2020) and Goedhart et al. (2020) for univariate cases. Therefore, in the present research, we also address this issue by determining control limits that guarantee in-control performance, the so called guaranteed control limits.

This work addresses the challenge of designing a distribution-free tool for multivariate process monitoring, as reliance on distributional assumptions can significantly degrade the performance of control charts. Additionally, this model addresses practitioner variability a critical factor that, if not accounted for, can lead to performance outcomes as unpredictable as random events. By focusing on these aspects, which frequently arise in real world applications the proposed tool seeks to enhance control reliability and robustness, providing a more consistent approach to managing non-normal multivariate processes across different practitioners.

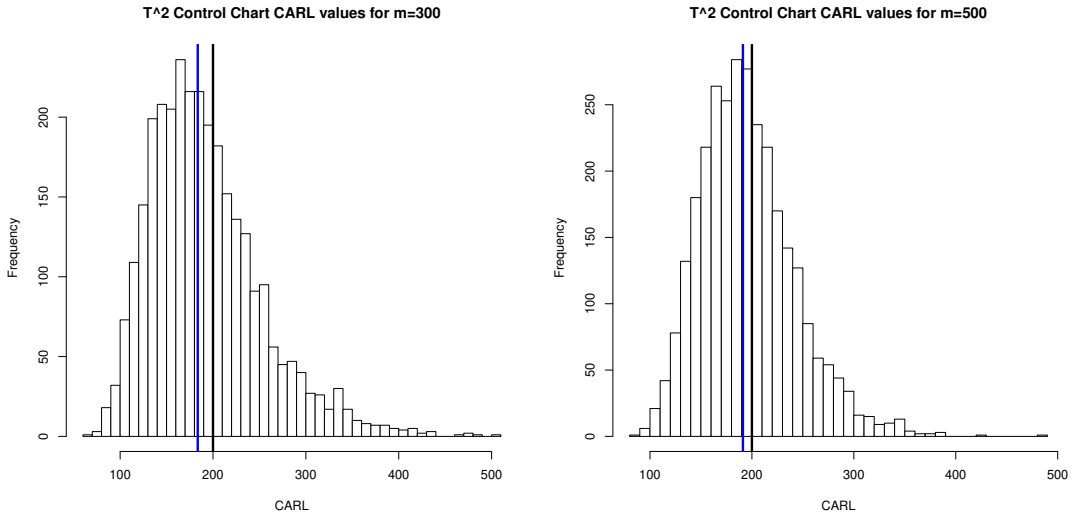


Figure 1.4: $CARL$ distribution for the classical Hotelling's T^2 control chart defined in Equation 1.6. A $CARL$ value was obtained for each one of the 1000 Phase-I samples simulated from a 5-dimensional normal process. Black line represents the preespecified ARL_0 and blue line indicates the median of $CARL$ distribution.

1.2 Research Framework and Objectives

A new multivariate nonparametric Phase II control chart for location is proposed in this research, combining Mahalanobis distances with the Mann-Whitney statistic. The Mahalanobis distance is employed to assess each observation's distance from the mean while accounting for correlations between variables. This approach is particularly effective in transforming multivariate data into a single dimension, facilitating more straightforward monitoring of process deviations. Meanwhile, the Mann-Whitney statistic is selected for its robustness as a nonparametric method, enabling the detection of shifts in central tendency without relying on normality assumptions. By integrating these techniques, it is expected that any changes in the underlying multivariate distribution will be reflected in the Mann-Whitney statistic, effectively signaling potential deviations in process behavior.

Furthermore, the calibration of control limits considers practitioner-to-practitioner variation. Additional details about the proposed methodology and the resulting properties will be discussed in the following sections.

Research Questions

1. What is the effectiveness, in terms of the CARL of the proposed control chart for monitoring the location of multivariate processes whose distribution is unknown?

2. For multivariate processes, can we obtain control limits that guarantee a minimum in-control behaviour in terms of the average run length ARL? In other words, Can the proposed control chart guarantee a desirable rate of false alarms with the "amount of data at hand"?

3. When addressing the practitioner-to-practitioner variation, is the conditional perspective a better alternative in terms of the IC performance than the control charts based on the unconditional perspective?

Research Hypothesis

The hypotheses related to the previous questions are:

- The IC performance of the proposed control chart is not affected by any underlying distribution. Therefore, the proposed control chart can be considered as distribution-free.
- The control limit (singular) of the proposed control chart considers the practitioner-to-practitioner variability and is able to guarantee an IC performance regardless the reference sample at hand.
- Under the conditional perspective, the proposed control chart is a better alternative than the unconditional perspective when the objective is to guarantee an IC performance.

Research Objectives

In order to evaluate the previous hypotheses, we do the following:

- Design a Phase II distribution free multivariate control chart based on Mahalanobis distances ranking. Mean vector and covariance matrix of the process are not to be known a priori.
- Design control limits to obtain a guaranteed performance through Monte Carlo simulations, using an approach similar to Celano and Chakraborti (2020).
- Perform a series of simulations to evaluate the effectiveness of the proposed control chart under various IC and OC scenarios, using the median of the CARL as the performance evaluation criterion

1.3 Scope and limitations

The proposed chart is designed for a Phase II control scheme, namely an online monitoring, or assumes that an IC reference sample has already been obtained from a proper Phase I analysis. It is designed to monitor subgrouped data and to detect a potential shift in the mean vector of the multivariate IC distribution. The information to decide whether the process is IC or OC is obtained from the most recent sample, ignoring the information contained in previous samples. As we will see in section 2.2 nonparametric MSPC seems not to have studied yet the use of guaranteed control limits, being our proposal the first distribution-free multivariate control chart addressing this topic. Since no precedent is available we focus on the advantages and disadvantages of our proposal under both unconditional and conditional perspectives as well as its robustness to any multivariate distribution. The remainder of this document is organized as follows. A detailed description of the latest work in nonparametric MSPC is given in section 2, followed by section 3, where we show detailed description of the chart statistic and how the control limits is determined. In section 4, the performance of our proposed chart is studied. A real data scenario is tested to illustrate the usefulness of our proposed scheme in section 5. At the end, comments and recommendations for practitioners are given in section 6.

Chapter 2

Literature Review

The potential success of any product or service generated through an industrial process relates to accomplishing the requirements established by the consumers. To meet these specifications, we need to evaluate the process quality by monitoring its underlying distribution in real-time, looking for the detection, as soon as possible, of any potential mean or scale shift attributed to special causes of variation. Since quality is a multifaceted concept, most of the SPC applications are concerned with multiple characteristics, mainly nowadays when we count with the necessary tools to recollect a huge amount of data sequentially. Therefore, as Qiu (2018) suggested, SPC research should focus on developing statistical tools for multivariate cases, considering the great frequency with which these scenarios occur in practice.

For each control chart, there is a mathematical model that considers certain assumptions about the process to be monitored, due to this, its performance is closely related to their fulfillment. Two common assumptions are the normality of the process and that control chart performance is uniform across the different practitioners who can use it. In the following two subsections we will delve into the aspects that need to be taken into account when the previous assumptions are considered or not. The goal is to emphasize how frequently these assumptions are not accomplished in practice and the different methodologies that have been proposed in the SPC area in order to be able to monitor processes. This will allow us to define areas of opportunity in SPC research, in which the proposed methodology developed here will be a contribution.

2.1 Non-normality in multivariate statistical process control

Nowadays, traditional Phase-II control charts such as the Hotelling's T^2 Hotelling (1947) and time-dependent approaches such as the multivariate CUSUM Crosier (1988) and Healy (1987) and multivariate EWMA Lowry et al. (1992) charts are popular when it

comes to monitoring the location of multivariate data. The former is better suited for medium to large changes, and it is usually recommended during an early stage of a monitoring. The latter approach deals better with small and sustained shifts, ideally used in a process that have reached stability where frequent and isolated causes of variation are no longer a concern, Montgomery (2020). Although the MSPC area has been explored to a great extent, most of the works such as those mentioned above and many others that can be consulted in Qiu (2014) are built under the assumption that the process follows a multivariate normal distribution.

It is well known that in practice normality assumption is rarely fulfilled. Actually, fulfillment of normality in multivariate processes entails much greater complexity than the univariate case since this implies that the dataset under analysis must have normally distributed marginal distributions and its marginals to be jointly normally distributed. There is a quote cited in Stoumbos and Sullivan (2002) that largely represents what often happens in reality: Coleman (1997) noted that he "... would never believe the multivariate normal assumption for industrial data," and that even if he wanted to, he "...cannot believe that there are tests for multivariate normality with sufficient power for practical sample sizes ..." that he "... would even bother to use them." He concluded that "... distribution-free multivariate SPC is what we need."

In the statistical literature, existing methods for addressing multivariate non-normal data are limited. If a control chart scheme based on the normality assumption is used where the assumption is invalid, this may lead to wrong conclusions in the determination of whether the process state is IC or OC and then the actual performance of the control chart deteriorates severely, see for example Chou et al. (2001), Phaladiganon et al. (2011) and Mukherjee et al. (2017). Therefore, development of multivariate non-parametric control charts is exceptionally important.

Recently, there has been an increase in research related to the development of non-parametric control charts that allow the monitoring of processes whose distribution is unknown. Although the vast majority of proposals to solve this problem address the univariate case (Chakraborti & Graham, 2019; Qiu, 2014), there is some literature dedicated to the multivariate counterpart. Qiu and Hawkins (2001) suggested a nonparametric control approach based on the cross-component ranking of the data. Stoumbos and Sullivan (2002) and Testik et al. (2003) reported that small values of the parameter lambda result in a multivariate EWMA that is robust to nonnormality. Qiu (2008) proposed a multivariate CUSUM chart that takes advantage of the categorization of data based on a log-linear model. A bootstrap Hotelling's T^2 control chart was proposed by Phaladiganon et al. (2011). As an extension of Das (2009) who proposed a control chart scheme based on the bivariate sign test, Boone and Chakraborti (2012) developed two multivariate

Shewhart-type control charts based on component-wise signs and signed-ranks sums. Zou and Tsung (2011) adapted the spatial sign test to a monitoring scheme. The proposed method by Chen et al. (2015) uses the Wilcoxon rank sum statistic in an EWMA monitoring scheme whose control limits are determined online. Yue and Liu (2017) proposed an adaptive multivariate EWMA control chart with variable sampling interval. Mukherjee et al. (2017) developed a nonparametric Shewhart type scheme for monitoring bivariate processes based on The Mathur statistic. Huang et al. (2018) proposed two bivariate nonparametric EWMA monitoring schemes to monitor time between events, which are based on the statistics developed by Jurečková and Kalina (2012) and Mathur (2009). W. Li et al. (2020) take advantage on the availability of historical IC and OC data to propose a multivariate nonparametric CUSUM approach based on the $k-$nearest neighbors algorithm. Zhang et al. (2020) designed a multivariate nonparametric EWMA control chart based on random projections to decompose high dimensional data into several subspaces that are then used in a spatial rank test. A self-starting EWMA control chart based on run tests is proposed by Y. Li et al. (2020). Cabana and Lillo (2021) proposed a multivariate control chart based on robust shrinkage estimators. Recently, Mukherjee and Marozzi (2020) presented five Shewhart type control charts for different purposes in the monitoring of high dimensional processes. The novelty in these charts lies in the use of multivariate distance measures whose ranks are distribution free, or, at least, distribution robust, and can be used to construct a two sample test in the sense of Jurečková and Kalina (2012). Their chart statistics based on Euclidean distances showed favorable results, leaving as a future research topic the construction of monitoring procedures based on their scheme.

Literature review and open problems in nonparametric MSPC can be found in Qiu (2018) and Chakraborti and Graham (2019). Chakraborti and Graham (2019) concluded that is important to consider the effect of parameter estimation in the performance of Phase-II nonparametric control charts, as demonstrated in the MSEWMA control chart by Dovoedo and Chakraborti (2017), which uses the spatial sign statistic in a Phase-II MEWMA control chart and adjust the control limits to obtain a nominal IC average run length ARL_0. To address the challenge of monitoring processes with unknown distributions, we draw inspiration from the work of Mukherjee and Marozzi (2020), who proposed transforming multivariate observations into Mahalanobis distances and analyzing them with a Mann-Whitney monitoring statistic. Their methodology is particularly suited for subgroup data with an available Phase-I in-control sample, where the distribution, mean vector, and covariance matrix are unknown.

However, a critical gap remains in the literature: the inherent variability between practitioners in nonparametric MSPC has not been adequately explored. Existing method-

ologies, including those mentioned above, do not account for how differences among practitioners may influence the performance and implementation of control charts. Addressing this issue is crucial to enhance the robustness and practical applicability of nonparametric MSPC methods.

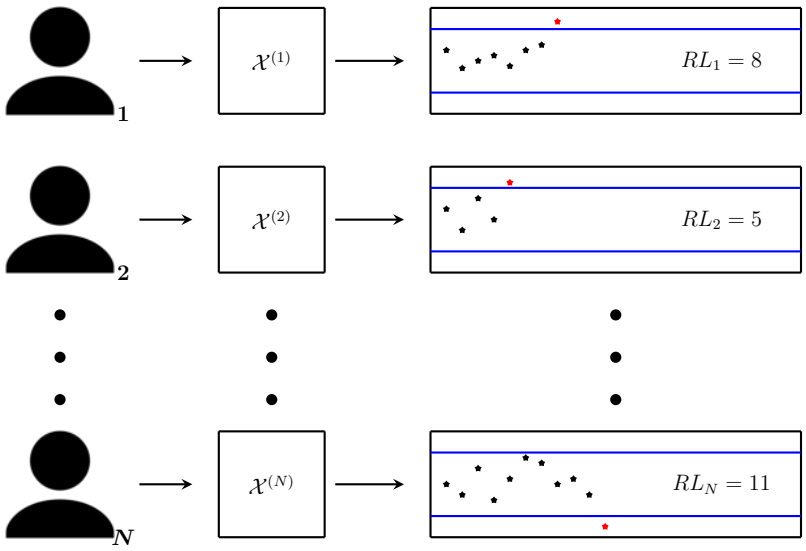


Figure 2.1: An overview of the classical procedure to estimate the ARL. When using this performance assessment, the inherent variability that exists between practitioners (and their corresponding reference samples) is not taken into account. Therefore, control limits estimated under this approach cannot guarantee an IC performance for the practitioners with their corresponding Phase-I sample $\mathcal{X}$.

2.2 The practitioner-to-practitioner variation

2.2.1 The traditional approach for control limits

As we briefly discussed in section 1.1, traditionally researchers have used the ARL as a criterion for performance evaluation. If the process distribution is known we could obtain the distribution of the chart statistic and then define its control limits to accomplish a desired ARL_0, recalling that $ARL_0 = 1/\alpha$, where α is a pre-specified false alarm rate. For instance, if we define $\alpha = 0.005$, then we would expect an alarm every 200 monitoring points when the process is in an IC state, in other words $ARL_0 = 200$. On the other hand, when the process is OC, the lower the OC ARL or ARL_1 the better the sensitivity of control chart to detect a parameter shift in the distribution process. However, sometimes it is not straightforward to obtain the exact null distribution for the chart statistic. To overcome this situation we can rely on Monte Carlo simulations to generate the performance results, once the control limits have been defined. For this, a series of processes

with their corresponding Phase-I samples are simulated to obtain the RL values according to the control chart to be evaluated. On each simulation we use N replicates to then get the estimated ARL value, $ARL = \sum_{i=1}^{N} RL_i/N$. When the simulation is an IC process we would expect to have a simulated ARL values close to the nominal ARL_0, as well as we expect to obtain lower and lower ARL values as the shift in the mean distribution process to be analyzed increases. A graphic representation of this simulation is presented in Figure 2.1.

Obtaining control limits it is not straightforward when process distribution to be monitored is unknown, for these cases, Monte Carlo simulations and a search algorithm are used to calibrate the control limits according to a desired ARL_0 value. For illustration purposes, we introduce the following pseudo code as presented in Qiu (2014), to calibrate control limits using a classical searching algorithm as the bisection method:

Pseudo Code to Search for a Control Limit *h*

Let A_0 be the pre-specified ARL_0 value, $[h_L, h_U]$ be the interval from which h is searched, $\rho > 0$ be a small number denoting the required estimation accuracy, and M be the maximum number of iterations involved in the search. Then, the search is performed iteratively as follows. In the $j-$th iteration, for $1 \leq j \leq M$, do the following steps:

Step 1. Let $h = (h_L + h_U)/2$. Compute the ARL_0 value of a **given chart** with this h value.
Step 2. If the ARL_0 value computed in Step 1 is included in $[A_0 - \rho, A_0 + \rho]$, then the algorithm is stopped, and the searched value of h is set to be the one used in Step 1. Otherwise, define

$$\begin{cases} h_L = h_L, h_U = (h_L + h_U)/2 & \text{if } ARL_0 > A_0, \\ h_L = (h_L + h_U)/2, h_U = h_U & \text{if } ARL_0 < A_0, \end{cases}$$

and return to Step 1.

If the above algorithm does not stop before or at the $M-th$ iteration and the ARL_0 value computed in the $M-$th iteration is still outside the interval $[A_0 - \rho, A_0 + \rho]$, print a statement that the estimation accuracy specified by ρ cannot be reached, and set the h value defined in the $M-$th iteration as the searched value of h.

Following the same purpose, several authors have proposed different methodologies to estimate control limits, for instance, Zou and Tsung (2011) and Zou et al. (2012)

applied a Markhov chain approach; Chakraborti and Van de Wiel (2008) and Celano and Chakraborti (2020) compute control limits using Lugannani-Rice-saddlepoint; Capizzi and Masarotto (2016) proposed an stochastic approximation to find control limits of a control chart, this was later adapted to estimate guaranteed control limits in Capizzi and Masarotto (2020). Once the control limits have been defined, we can evaluate control chart performance over different scenarios either for analyzing its robustness to non-normality or its sensitivity to shift locations. Actually, most of the papers presented in section 2.1 use similar steps when determining control limits and evaluating their corresponding proposals. However, when working with this type of approaches, there are some points that we need to highlight. Based on what Figure 2.1 shows, for the simulated practitioners with their corresponding Phase-I sample $\mathcal{X}$, a RL value is obtained instead of a conditional performance. One of the major objectives of any statistical tool, such as a control chart, is that it can be used by a large number of practitioners. Therefore, it is crucial to design tools that allow a great robustness to the different scenarios that may occur in practice, such as the inherent variability that exists between practitioners. For this reason, lately, SPC research has focused some of its efforts to develop control charts that take into account this variability, an overview of these works is presented in next subsection.

2.2.2 Guaranteed Performance Control Limits

Whenever a monitoring procedure depends on observations from a Phase-I sample, as our proposal, performance is subject to sample variation. If we define performance using the average run length (ARL), different Phase-I samples generate different ARL values, this is called the practitioner-to-practitioner variation. Sometimes these values will fall below pre-specified in-control $ARL(ARL_0)$, which creates an undesirable increase in false alarms that bring distrust in our application. Since performance of a control chart will vary among Phase-I samples, it is appropriate to consider the Conditional Average Run Length, $CARL(x)$, the mean of the distribution of run length conditioned to the Phase-I sample $\mathcal{X} = x$. The distribution of the $CARL(x)$ is, in most of the cases, unknown, and for this reason in our proposal we use Monte Carlo simulation to estimate this distribution through the empirical distribution, as in Celano and Chakraborti (2020). For this, we generate T Phase-I samples $x^{(t)}$, with $t = 1, \ldots, T$, and for each Phase-I sample we obtain its $CARL(x^{(t)})$, this is $CARL(x^{(t)}) = \sum_{i=1}^{N} RL_i(x^{(t)})/N$. This set of T random variables $CARL(x^{(t)})$ is then used to obtain the empirical distribution of the $CARL(x)$. A graphic representation of this procedure is shown in Figure 2.2.

Several authors have shown that in-control performance of Phase-II charts depends largely on the parameters estimated from Phase-I sample. Literature reviews of Jensen

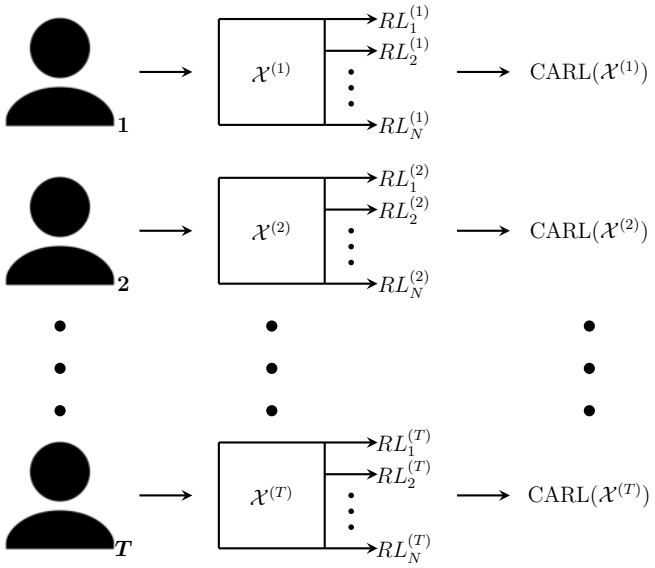


Figure 2.2: Overview of the procedure used to estimate the empirical distribution of $CARL$, where performance is conditioned on the Phase-I sample $\mathcal{X}$. Since different practitioners obtain different Phase-I samples $\mathcal{X}$, practitioner-to-practitioner variation naturally arises. Consequently, the control limits estimated under this approach guarantee the performance of the control chart for the specific Phase-I sample under consideration.

et al. (2006) and Psarakis et al. (2014) offer an extensive and well-documented discussion addressing this topic. It should be noted that most of the literature presented in these papers evaluates performance averaging over the effects of Phase-I sample parameter estimates. This is known as the unconditional perspective, and was widely used in earlier studies over the effects of parameter estimates. Generally they concluded that Phase-I sample size should be increased until the unconditional ARL achieves a pre-specified ARL_0, see Quesenberry (1993), Chakraborti (2000), L. A. Jones et al. (2004). However, one should notice that the unconditional ARL does not guarantee an in-control performance for the chart with its corresponding estimated parameters, therefore there could still be a high probability for a practitioner to get a false alarm, Jardim et al. (2019).

Later, researchers accounted for the inherent variability between practitioners and their respective Phase-I samples. This variability in Phase-I samples introduces variation in estimated parameters and control chart performance. Therefore, some studies on the effects of parameter estimation were expanded to include what is known as the conditional perspective, which focuses on practitioner-to-practitioner variation. An example of this is found in Saleh, Mahmoud, Jones-Farmer, et al. (2015) who extended the work of L. A. Jones et al. (2001) studying the effects of estimated parameters in univariate EWMA. To evaluate under the conditional perspective M. A. Jones and Steiner (2012) proposed to use the standard deviation of ARL as a performance metric. Other authors that have

used this metric are Saleh, Mahmoud, Keefe, et al. (2015), Epprecht et al. (2015) and Saleh et al. (2016) which recommend to increase the amount of Phase-I sample until a certain $SDARL$ value is achieved, for example 10 % of the ARL_0 as in Saleh et al. (2016). As can be appreciated, both perspectives reach similar conclusions: increase the Phase I sample size, where the increase for the unconditional perspective is lower than for the conditional perspective, although the latter guaranteeing an in-control performance for the practitioners and their corresponding Phase-I samples. However, despite advancements in technology that enable the collection of large amounts of data, it remains challenging to obtain the recommended Phase-I sample sizes for both perspectives. This difficulty often arises from practical limitations, making the proposed or required Phase-I sample sizes impractical to achieve in many cases.

Because of the effect of the Phase-I sample in the performance of the control charts, it is necessary to adjust the control limits to achieve an in-control performance. Some authors have adjusted control limits considering the unconditional perspective, see Chakraborti (2006), Goedhart et al. (2016) and Diko et al. (2017), looking for a desired average in the $CARL$. However, as we have previously mentioned, although this scheme may seem appropriate it does not guarantee the performance of a control chart for a specific Phase-I sample since it is an average over the performance using estimated parameters and there could still be a high probability for a practitioner to get, in average, false alarms below the desired ARL_0. Recently the exceedance probability criterion (EPC) proposed by Albers and Kallenberg (2005) has been a more accepted approach to guarantee an in-control performance for the practitioner, Does et al. (2020). This criterion is used to obtain control limits for which just a predetermined proportion of the practitioners have an average performance below the expected. EPC has become increasingly important in SPC research as its implementation has been recommended by several authors such as Gandy and Kvaløy (2013), Goedhart et al. (2017), Diko et al. (2019), Jardim et al. (2019) for univariate control charts. Although practitioner-to-practitioner variation has gained a lot of attention recently, all the papers consulted above referring to EPC assume a known distribution of the process. However, data distribution is generally unknown. The use of the EPC in nonparametric charts can be seen in Goedhart et al. (2020) for a classical Phase-II approach and Celano and Chakraborti (2020) for finite horizon production (FHP) processes.

As for MSPC, Aly et al. (2016) and Saleh and Mahmoud (2017) studied, under the conditional perspective, the effects of estimated parameters on the Multivariate Adaptive EWMA chart and MEWMA chart respectively. Both researches concluded that in-control performance is severely affected by the variation between practitioners needing larger Phase-I samples to achieve a desired ARL_0. In order to overcome this situation, both

recommend to use the bootstrapp methodology proposed for Gandy and Kvaløy (2013) to meet EPC, however, they do not provide any experimentation implementing their suggestion.

Because of the above discussion and the literature review consulted, nonparametric MSPC seems not to have studied yet the use of guaranteed control limits. It is to our knowledge that up to the moment this manuscript was written, no previous work has been published addressing guaranteed in-control performance in nonparametric MSPC, being this the main contribution of our study.

Building on the foundational ideas of Mukherjee and Marozzi (2020), we propose a Phase-II control chart that retains the transformation of multivariate observations into Mahalanobis distances while incorporating a novel consideration of practitioner-to-practitioner variability. By explicitly addressing this variability, our methodology aims to bridge the gap in existing research and contribute to the development of more resilient and adaptable nonparametric MSPC tools. Given the lack of precedents, this study emphasizes evaluating the advantages and limitations of our proposal from both unconditional and conditional perspectives, as well as examining its robustness across various multivariate distributions. In order to reduce the added complexity required to perform Monte Carlo simulations to estimate the conditional ARL_0, we use the cumulative conditional false alarm probability, $CCFAP$, based on runs truncated at a finite point L. More details on the methodology implemented in this research are provided in Chapter 3.

Chapter 3

Methodology

3.1 Proposal

The proposed monitoring scheme is based on two types of multivariate samples. The Phase-I observations, assumed to be in-control with mean vector $\boldsymbol{\mu}_0$ and covariance matrix $\boldsymbol{\Sigma}_0$, correspond to the first sample set. The second set is defined by Phase-II observations that arrive in form of batches. The Phase-I in-control (IC) sample is denoted by $\mathcal{X} = \{\boldsymbol{X}_1, \boldsymbol{X}_2 \ldots, \boldsymbol{X}_m\}$, where each $\boldsymbol{X}_i$ for $i = 1, \ldots m$ is characterized by p variables. The Phase-II observations $\mathcal{Y}_k = \{\boldsymbol{Y}_{1k}, \boldsymbol{Y}_{2k}, \ldots, \boldsymbol{Y}_{nk}\}$ denote the p dimensional k^{th} batch of size n, where $\boldsymbol{Y}'_{jk} = (Y_{1jk}, Y_{2jk}, \ldots, Y_{pjk})'$ for $j = 1, 2, \ldots, n$ and $k = 1, 2, 3, \ldots$, represent the j^{th} observation of the k^{th} batch.

From the Phase-I sample the mean vector and covariance matrix can be estimated as

$$\bar{\boldsymbol{X}} = \frac{1}{m} \sum_{i=1}^{m} \boldsymbol{X}_i \tag{3.1}$$

and

$$\boldsymbol{S} = \frac{1}{m-1} \sum_{i=1}^{m} (\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})', \tag{3.2}$$

respectively. A distance between each observation $\boldsymbol{Y}_{jk}$ in a Phase-II batch from the Phase-I sample can be estimated with the Mahalanobis distance

$$D^2_{jk} = (\boldsymbol{Y}_{jk} - \bar{\boldsymbol{X}})' \boldsymbol{S}^{-1} (\boldsymbol{Y}_{jk} - \bar{\boldsymbol{X}}). \tag{3.3}$$

Independence between $\boldsymbol{Y}_{jk}$ observations translates to independent D^2_{jk} variables. A rank transformation over these variables leads to distribution free statistics when the process is in-control. We can also use the Mahalanobis distance over each $\boldsymbol{X}_i$ observation from the Phase-I observations as

$$R^2_i = (\boldsymbol{X}_i - \bar{\boldsymbol{X}})' \boldsymbol{S}^{-1} (\boldsymbol{X}_i - \bar{\boldsymbol{X}}) \tag{3.4}$$

and effectively reduce the problem to a single dimension. Nevertheless equation (3.4) leads to observations that are not independent, as $\bar{\boldsymbol{X}}$ and $\boldsymbol{S}$ are estimated from the same set. A rank transformation over these variables is not distribution free, however, dependence is reduced as the sample gets larger, and as observed in section 4.1, statistics based on these observations are distribution robust. Following the approach of Mukherjee and Marozzi (2020), multivariate observations are transformed to univariate distance measures that, at the same time, can be transformed into ranks whose null distribution is independent of the distribution of random vectors.

To evaluate whether Phase-II observations are in-control, each batch is evaluated against the Phase-I sample using a Mann-Whitney statistic over previously calculated distance measures. Using equation (3.4) we define $\boldsymbol{R}^2 = (R_1^2, R_2^2, \ldots, R_m^2)'$ as the Phase-I sample of size m of Mahalanobis distances. From the k^{th} batch, we obtain its corresponding Mahalanobis distances $\boldsymbol{D}_k^2 = (D_{1k}^2, D_{2k}^2, \ldots, D_{nk}^2)'$ using equation (3.3). Any change in the location of the process distribution is likely to be reflected in the elements of $\boldsymbol{D}_k^2$, otherwise we would expect values close to those of $\boldsymbol{R}^2$. In our case, we can use the ranks $\boldsymbol{D}_k^2$ with respect to $\boldsymbol{R}^2$ to compare a Phase-II batch with the Phase-I sample using a Mann-Whitney statistic

$$U_k = \sum_{i=1}^{m} \sum_{j=1}^{n} I(D_{jk}^2 > R_i^2), \tag{3.5}$$

where $I(D_{jk}^2 > R_i^2)$ is the indicator function:

$$I(D_{jk}^2 > R_i^2) = \begin{cases} 1 & \text{if } D_{jk}^2 > R_i^2, \\ 0 & \text{if } D_{jk}^2 \leq R_i^2. \end{cases}$$

If the process is in-control then, for all k,

$$E(U_k|IC) = \mu = \frac{mn}{2} \tag{3.6}$$

and

$$Var(U_k|IC) = \sigma^2 = \frac{mn(m+n+1)}{12}. \tag{3.7}$$

When the process is IC, U_k is symmetric about μ, and

$$MW_k = \frac{U_k - \mu}{\sigma} \tag{3.8}$$

will approximate a $N(0,1)$ as the Phase-I sample gets larger. We can use MW_k as the plotting statistic. A process is said to be out-of-control if $MW_k > UCL$, where UCL

stands for upper control limit. Even though we can also define a lower control limit for MW_k, changes in location only require an upper bound.

Alternatively, we can assure independence of $\boldsymbol{R}^2$ elements by randomly splitting the Phase-I sample $\mathcal{X}$ in two groups. The first group is used to estimate a mean vector and a covariance matrix to replace in equations (3.3) and (3.4), and the second group is transformed to Phase-I Mahalanobis distances using equation (3.4) with the new mean vector and covariance matrix. The remain of the process is left as usual. Formally, the first group is determined arbitrarily and defined as $\mathcal{V} = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_v\}$, where $1 < v < m$, where its corresponding sample mean vector and sample covariance matrix are denoted as $\bar{\boldsymbol{X}}_{\mathcal{V}}$ and $\boldsymbol{S}_{\mathcal{V}}$. The remaining $m - v$ observations create a new $\boldsymbol{R}^2$ vector set where each element is defined as $R_i^2 = (\boldsymbol{X}_{v+i} - \bar{\boldsymbol{X}}_{\mathcal{V}})' \boldsymbol{S}_{\mathcal{V}}^{-1} (\boldsymbol{X}_{v+i} - \bar{\boldsymbol{X}}_{\mathcal{V}})$ for $i = 1, 2, \ldots, (m - v)$. By doing this, independence of the elements of $\boldsymbol{R}^2$ is assured. The value of v can be defined as $v = m(1 - b)$ where b is the proportion of $\mathcal{X}$ that will form the new set of $\boldsymbol{R}^2$. For example, assume there are 500 $p-$variate observations in the Phase-I sample $\mathcal{X}$ and choose a proportion $b = 0.7$, then we will have 150 $p-$variate observations to generate $\bar{\boldsymbol{X}}_{\mathcal{V}}$ and $\boldsymbol{S}_{\mathcal{V}}$ and $\boldsymbol{R}^2$ will be formed of 350 independent R_i^2 Mahalanobis distances.

To distinguish between transforming the full set $\mathcal{X}$ to generate the vector $\boldsymbol{R}^2$, and using a subset to assure independent measures, we define as the $MW^{(\mathcal{D})}$ chart to indicate the former, and the $MW^{(\mathcal{I})b}$ chart for the latter. Both methodologies are built as multivariate Shewhart-type monitoring schemes which can be implemented as follows:

1. Set a Phase-I sample $\mathcal{X}$, that for the purposes of this study it will be assumed as IC. For more details on how to obtain a Phase-I sample through proper analysis, see Cheng and Shiau (2015) and Capizzi and Masarotto (2017) regarding to nonparametric MSPC.

2. Collect the Phase-II sample $\mathcal{Y}_k$.

3. Using samples $\mathcal{X}$ and $\mathcal{Y}_k$, calculate the plotting statistic of either the $MW^{(\mathcal{D})}$ or $MW^{(\mathcal{I})b}$ control charts, depending on the selected monitoring scheme.

4. If the selected plotting statistic exceeds its corresponding UCL, then an alarm will be issued. Otherwise monitoring continues and we go to Step 2.

In next section we give a brief review and analyze the performance of one of the high-dimensional Shewhart-type schemes (HDS) developed by Mukherjee and Marozzi (2020). This control chart uses Euclidean distances of each observation in the reference sample from the origin (OR) in the Wilcoxon statistic (W). This will be referenced as the HDSOR-W control chart, which was designed for location monitoring.

3.2 Validation of our proposal

3.2.1 The HDSOR-W control chart

The HDSOR-W control chart is based on Euclidean distances from the origin. Let ED_i denote the Euclidean distance of the ith reference sample $\boldsymbol{X}_i$ from the origin $\mathbf{0} \in \mathbb{R}$ as

$$ED_i^{\mathcal{X}} = ED(0, \boldsymbol{X}_i) = ||\boldsymbol{X}_i|| \text{ for } i = 1, 2, \ldots m. \tag{3.9}$$

Further, let the Euclidean of the j^{th} member of k^{th} batch sample $\mathcal{Y}$ from the origin be $ED_k^{\mathcal{Y}}$. So,

$$ED_{jk}^{\mathcal{Y}} = ED(0, \boldsymbol{Y}_{jk}) = ||\boldsymbol{Y}_{jk}|| \text{ for } j = 1, 2, \ldots n, k = 1, 2, \ldots \tag{3.10}$$

Now, $\boldsymbol{ED}^{\mathcal{X}} = (ED_1^{\mathcal{X}}, ED_2^{\mathcal{X}}, \ldots, ED_m^{\mathcal{X}})'$ is the reference sample of size m of the Euclidean distances from origin and $\boldsymbol{ED}_k^{\mathcal{Y}} = (ED_{1k}^{\mathcal{Y}}, ED_{2k}^2, \ldots, ED_{nk}^2)'$ is the k^{th} batch sample of size n of the Euclidean distances from the origin. Once the multivariate observations are transformed to Euclidean distances measures Mukherjee and Marozzi (2020) designs a Shewhart-type control chart based on the ranks of $(\boldsymbol{ED}^{\mathcal{X}}, \boldsymbol{ED}_k^{\mathcal{Y}})$ and using the Wilcoxon statistic, defined as

$$W_k = \sum_{i=1}^{N} iZ_i \tag{3.11}$$

where $Z_i = 1$ when the i^{th} order statistic of $(\boldsymbol{ED}^{\mathcal{X}}, \boldsymbol{ED}_k^{\mathcal{Y}})$ belongs to the test sample, $Z_i = 0$ otherwise. If the process is IC then, for all k,

$$E(W_k|IC) = \mu = \frac{n(m+n+1)}{2} \tag{3.12}$$

with the same variance as defined in equation 3.7. Since W_k is symmetric about μ_1 in an IC set-up, and a large value of $W_{1k} - \mu_W$ indicates a process shift, we consider a Shewhart-type chart based on the plotting statistic

$$W_k' = \frac{W_k - \mu}{\sigma} \tag{3.13}$$

where W_k' denotes the standardized value of W_k. For OC observations we would expect Euclidean distances in $\mathcal{Y}_k$ to be larger than those present in $\mathcal{X}$ and this will be reflected in larger values of W_k. Therefore, an alarm is issued if $W_k > UCL$. Mukherjee and Marozzi (2020) noted that Euclidean distances are affected by the variability of different components. To avoid this problem, they suggest to consider the transformations $\mathcal{X}\boldsymbol{S}_D^{1/2}$

and $\mathcal{Y}_k \boldsymbol{S}_D^{1/2}$ $k = 1, 2, ...$, where $\boldsymbol{S}_D$ is the diagonal matrix of $\boldsymbol{S}$. With this operation, the reference sample and test sample will be standardized without altering the covariance structure and without masking the effect of the shift. The HDSOR-W control chart can be implemented following the same steps previously defined for the $MW^{(\mathcal{D})}$ control chart, given their similarities they are suitable for a performance comparison and to demonstrate the validity of our proposed statistic.

3.3 Comparison versus other Charts

The HDSOR-W control chart was not designed to have a guaranteed performance. Therefore, for a fair comparison, both control charts must be evaluated under same circumstances. For this, we assess the $MW^{(\mathcal{D})}$ control chart under the conditions in which the analysis was carried out in Mukherjee and Marozzi (2020). In this approach, the performance is not intended to be guaranteed, thus practitioner-to-practitioner variation is not taken into account. Following this scheme, control limits of $MW^{(\mathcal{D})}$ control chart were adjusted to accomplish a nominal in control median run length, MRL_0, equal to 250. A classical searching algorithm, such as the bisection method, was used to determine the control limits. We recommend consulting section 2.2.1, where a pseudocode similar to that of Qiu (2014) is presented, to delve into the subject.

Having established the parameters of this analysis, we study the IC and OC performance of HDSOR-W and $MW^{(\mathcal{D})}$ control charts. The control limits used for HDSOR-W chart were taken from Mukherjee and Marozzi (2020), for reference and subgroup sizes of $m = 100, 300$ and $n = 5, 10$ respectively. A total of 10,000 simulations were conducted for each scenario to evaluate the performance of the control charts. We consider the following three distributions: for the symmetric case the multivariate normal ($Norm$) and multivariate t with 5 degrees of freedom (t), that were generated using the R package *mvtnorm*, and for the asymmetric case we choose the multivariate gamma (Gam) with shape parameter 3 and scale parameter 1 presented in the Appendix A-II of Zhang et al. (2016). We selected the last two distributions since they are often used to test the robustness to non-normality, see Stoumbos and Sullivan (2002), Zou and Tsung (2011), Chen et al. (2015). We consider dimensions of $p = \{5, 10, 20\}$, the necessary adjustments for each distribution were implemented to generate an in-control multivariate process mean vector $\boldsymbol{\mu_0} = \mathbf{0_{p \times p}}$, and a covariance matrix $\boldsymbol{\Sigma}_0 = (\sigma_{ij})$ so that $\sigma_{ii} = 1$ and $\sigma_{ij} = 0.5^{|i-j|}$, for $i, j = 1, 2, \ldots p$. For sensitivity assessment, we set the OC mean vector $\boldsymbol{\mu_1} = \mathbf{1}'\delta_{p \times 1}$ for simulation of test samples $\mathcal{Y}_k$ and set the IC mean vector $\boldsymbol{\mu_0}$ for simulation of reference sample $\mathcal{X}$. In both samples a covariance matrix $\boldsymbol{\Sigma}_0$ is considered.

Table 3.1: Median run length values for HDSOR-W and $MW^{(\mathcal{D})}$ control charts under selected IC ($\delta = 0$) and OC ($\delta > 0$) scenarios. For $\delta = 0$, boldface indicates the value closer to the nominal in-control MRL of 250; Δ is omitted ("–"). For $\delta > 0$, $\Delta = MRL_{MW^{(\mathcal{D})}} - MRL_{HDSOR-W}$; negative values indicate faster detection by $MW^{(\mathcal{D})}$.

		m=100, n=5			m=100, n=10			m=300, n=5			m=300, n=10		
Distribution	δ	HDSOR	MW	Δ	HDSOR	MW	Δ	HDSOR	MW	Δ	HDSOR	MW	Δ
	0	243	**251.5**	–	**249**	246	–	**252**	**252**	–	246	**248**	–
	0.5	52.5	31	-21.5	29	17	-12	48	29	-19	25	15	-10
$Norm_5$	1	3	2	-1	1	1	0	2	2	0	1	1	0
	1.5	1	1	0	1	1	0	1	1	0	1	1	0
	2	1	1	0	1	1	0	1	1	0	1	1	0
	0	**243.5**	260	–	**249**	257	–	**249**	248	–	242	**250.5**	–
	0.5	25	18	-7	12	8	-4	24	15	-9	11	7	-4
$Norm_{10}$	1	1	1	0	1	1	0	1	1	0	1	1	0
	1.5	1	1	0	1	1	0	1	1	0	1	1	0
	2	1	1	0	1	1	0	1	1	0	1	1	0
	0	**246**	**246**	–	**248**	260	–	240	**245.5**	–	255	**250**	–
	0.5	11	10	-1	4	4	0	10	7	-3	4	3	-1
$Norm_{20}$	1	1	1	0	1	1	0	1	1	0	1	1	0
	1.5	1	1	0	1	1	0	1	1	0	1	1	0
	2	1	1	0	1	1	0	1	1	0	1	1	0
	0	**242**	241	–	255	**246**	–	238	**249.5**	–	247	**251**	–
	0.5	96	37	-59	21	10	-11	87	40	-47	18	9	-9
t_5	1	2	1	-1	1	1	0	2	1	-1	1	1	0
	1.5	1	1	0	1	1	0	1	1	0	1	1	0
	2	1	1	0	1	1	0	1	1	0	1	1	0
	0	242	**251.5**	–	**251**	246.5	–	247	254	–	246	**248**	–
	0.5	97	40	-57	17	8	-9	86	38	-48	14	7	-7
t_{10}	1	2	1	-1	1	1	0	1	1	0	1	1	0
	1.5	1	1	0	1	1	0	1	1	0	1	1	0
	2	1	1	0	1	1	0	1	1	0	1	1	0
	0	**241**	259.5	–	**245**	262	–	**249**	250	–	**250.5**	241	–
	0.5	97	37	-60	15	7	-8	90	37	-53	12	6	-6
t_{20}	1	1	1	0	1	1	0	1	1	0	1	1	0
	1.5	1	1	0	1	1	0	1	1	0	1	1	0
	2	1	1	0	1	1	0	1	1	0	1	1	0
	0	240	**250.5**	–	259.5	**251**	–	238	**251**	–	247	**253.5**	–
	0.5	17	45	28	4	43	39	15	48	33	3	45	42
Gam_5	1	1	4	3	1	2	1	1	4	3	1	2	1
	1.5	1	1	0	1	1	0	1	1	0	1	1	0
	2	1	1	0	1	1	0	1	1	0	1	1	0
	0	**243**	261	–	**253**	257	–	**247**	252	–	249	**247**	–
	0.5	7	37	30	2	22	20	7	35	28	2	22	20
Gam_{10}	1	1	3	2	1	1	0	1	2	1	1	1	0
	1.5	1	1	0	1	1	0	1	1	0	1	1	0
	2	1	1	0	1	1	0	1	1	0	1	1	0
	0	**249**	254	–	**255**	259	–	246	254	–	**248**	246	–
	0.5	3	25	22	1	12	11	2	24	22	1	12	11
Gam_{20}	1	1	1	0	1	1	0	1	1	0	1	1	0
	1.5	1	1	0	1	1	0	1	1	0	1	1	0
	2	1	1	0	1	1	0	1	1	0	1	1	0

Table 3.1 compares the performance of HDSOR-W control chart vs. the $MW^{(\mathcal{D})}$ control chart considering different shift sizes $\delta = \{0, 0.5, 1, 1.5, 2\}$. To facilitate the analysis, we add the subscript p to indicate the dimension of multivariate distribution, this is $Norm_p$, t_p and Gam_p. Both control charts have the distribution-free property, as can be seen in in Table 3.1 for $\delta = 0$, presenting a very similar IC performance. As for the OC performance, we can clearly see how the $MW^{(\mathcal{D})}$ outperforms the HDSOR-W control chart under the normal and t distributions for $\delta = 0.5$ but being equivalent sensitive for shifts $\delta \geq 1$. On the other hand, when dealing with the multivariate gamma distribution, the HDSOR-W has a superior performance with respect to $MW^{(\mathcal{D})}$ control chart for $\delta = 0.5$. However, as it was presented in previous case, both control charts can be considered as having an equivalent OC performance after shift $\delta \geq 1$. Given the practicality of $MW^{(\mathcal{D})}$ control chart, and that can be considered as superior or equivalent in most of the scenarios under analysis to other recent monitoring statistics, such as the HDSOR-W control chart, we can confirm the validity of our proposed methodology. In next section we describe the methodology for adjusting control limits for our proposed control chart taking into account the practitioner-to-practitioner variation.

3.4 Proposed Guaranteed Control Limits

The determination of control limits to address the effects of Phase I sample variation on control chart performance has been studied through two approaches: the unconditional and conditional perspectives, see Jardim et al. (2020). In the unconditional perspective, control limits are obtained looking for a desired average of the distribution of the $CARL(x)$, known as the unconditional ARL, or $UARL$. However, under this perspective several practitioners might have an undesired performance, i.e., the rate of the practitioners where their $CARL(x)$ falls below the desired in-control ARL_0 is large.

To mitigate this issue, the conditional perspective offers an alternative. In this approach, control limits are determined using a criterion that minimizes the number of practitioners with undesired performance by accounting for the quantiles of the $CARL(x)$ distribution. This ensures that only a limited number of practitioners experience deviations from the desired performance.

This criterion known as the exceedance probability criterion (EPC) was introduced by Albers and Kallenberg (2005) and has become increasingly important in SPC research. For the conditional perspective the quantiles of the empirical distribution of the $CARL(x)$

are used to adjust the control limits according to

$$P(CARL(x^{(t)}) < ARL_0) = 1 - \frac{q}{100}, \tag{3.14}$$

where q is a desired confidence level. If we assume $\frac{q}{100} = 0.95$, then the so-called guaranteed control limits are obtained when the 95th quantile of the empirical distribution of $CARL(x)$ meets the nominal ARL_0; $\widehat{CARL}_{95}(x^{(t)}) = \text{ARL}_0$. For comparison purposes and to highlight the pros and cons of using the conditional approach, we also consider the unconditional perspective. In this case the control limits are selected such that the mean of the $CARL(x^{(t)})$, the $\widehat{UARL}$, meets the nominal ARL_0:

$$\widehat{UARL} = \frac{1}{T}\sum_{t=1}^{T} CARL(x^{(t)}) = ARL_0. \tag{3.15}$$

Given the t-th Phase-I sample, the estimation of the $CARL(x^{(t)})$ by means of Monte Carlo simulation can be computationally demanding, and hence the empirical distribution of the $CARL(x)$ cannot be addressed in a reasonable computational time. Because of this, we propose to use an estimation to the $CARL(x^{(t)})$ based on runs truncated at a finite point L. For a given Phase-I sample $\mathcal{X} = x$ we use the cumulative conditional false alarm probability ($CCFAP$), denoted as $CCFAP(x|L) = P(RL \leq L|\text{IC})$ to obtain the $CARL(x)$. In L independent trials, with a false alarm probability of p_F, $CCFAP(x|L)$ can be calculated as:

$$CCFAP(x|L) = 1 - \binom{L}{0} p_F^0 (1 - p_F)^L \tag{3.16}$$

then

$$p_F = 1 - (1 - CCFAP(x|L))^{(1/L)}. \tag{3.17}$$

Since in a Shewhart control chart the relationship between $CARL$ and probability false alarm p_F is given by $CARL_0 = 1/p_F$, then

$$CARL(x) = 1/(1 - (1 - CCFAP(x|L))^{1/L}) \tag{3.18}$$

and since runs are truncated, the computational time can be reduced significantly. The $CARL(x^{(t)})$ is estimated using equation (3.18) and the estimation of $CCFAP(x|L)$ ob-

tained with Monte Carlo simulation,

$$\widehat{CCFAP}(x^{(t)}|L) = \sum_{i=1}^{N} I(RL_i(x^{(t)}) \leq L)/N \tag{3.19}$$

where $I(RL_i(x^{(t)}) \leq L)$ is an indicator function:

$$I(RL_i(x^{(t)}) \leq L) = \begin{cases} 1 & \text{if } RL_i(x^{(t)}) \leq L, \\ 0 & \text{if } RL_i(x^{(t)}) > L, \end{cases}$$

and N is the total number of RL values estimated from the Phase-I sample $\mathcal{X} = x$.

The in-control performance of a distribution-free control chart must remain unaffected by the underlying distribution. Control limits for different m and n are obtained by Monte Carlo simulation using a bivariate normal process $N(\mathbf{0}, \mathbf{\Sigma}_0)$ with $p = 2$, simulations showed that control limits estimated with this dimension are robust and applicable to higher dimensions. Guaranteed control limits are set by using the following algorithm:

Algorithm 1 Procedure to Calculate Control Limits

Require: Initial upper control limit (UCL), confidence level (q), Phase-I sample size (m), Phase-II sample size (n), finite point (L), number of RL values (N), total Phase-I samples (T), nominal average run length (ARL_0), tolerance level (ϵ)

Ensure: Optimal UCL satisfying the specified perspective constraint (conditional or unconditional)

1: **Initialization:** Set input parameters: UCL, q, m, n, L, N, T, ARL_0, and ϵ.
2: **Generate Phase-I samples:** Simulate T Phase-I samples of size m from a 2-dimensional multivariate normal distribution $N(\mathbf{0}, \mathbf{\Sigma}_0)$ using the procedure described in section 3.3.
3: **Compute $\widehat{\text{CARL}}(x^{(t)})$:** For the current UCL, calculate the $\widehat{\text{CARL}}(x^{(t)})$ distribution using equation (3.18) for all $t = 1, \ldots, T$.
4: **Evaluate constraint:**
5: **if** Conditional perspective is selected **and** equation (3.14) is satisfied substituting ARL_0 by the interval $[ARL_0, (1+\epsilon)ARL_0]$ **then**
6: Select the current UCL as the optimal value.
7: **Stop.**
8: **else if** Unconditional perspective is selected **and** equation (3.15) is satisfied substituting ARL_0 by the interval $[ARL_0, (1+\epsilon)ARL_0]$ **then**
9: Select the current UCL as the optimal value.
10: **Stop.**
11: **else**
12: Update UCL using the bisection method.
13: **Go to Step 2.**
14: **end if**

3.4.1 On the selection of the L and N parameters

In this section an experimental analysis is presented to validate the $CFAP$-$CARL$ relationship previously described. For this we use the univariate MW control chart (Chakraborti and Van de Wiel (2008)), given that is equivalent to the statistic presented in equation (3.8). As a first step, control limits are calibrated using $CFAP(x|L = 50)$ for reference and test sample sizes of $m = 500, 300$ and $n = 5, 10$ respectively. The value $L = 50$ was chosen after performing simulations with different values of L, including $L = 20, 50, 100, 500, 1000$. The results indicated that $L = 50$ provides a good trade-off between the approximation of the $CFAP$-$CARL$ relationship and computational efficiency, as it requires less time while maintaining reliable results.

The objective is to analyze how well the estimated $CFAP$ and its equivalent $CARL$ approach to an independently simulated $CARL$ (estimated using the procedure shown in Figure 2.2, for a single reference sample). Both metrics are obtained from the same reference sample and these results are illustrated in Table 3.2 where we can appreciate that for values from $L = 50$ and $N = 1000$, the $CFAP$-$CARL$ relationship gets a good estimation, and as expected, it improves as the value of N increases. Based on the results presented in Table 3.2, we use $L = 50$, and calibrate control limits with $CCFAP_0(x|50) = 0.2216$, which is equivalent to an in-control $CARL_0(x) \approx 200$.

Table 3.2: Performance of univariate MW control chart for a single reference sample using $CFAP$ and $CARL$ estimated independently. Upper control limits (UCL) were calibrated under conditional perspective and using a $CCFAP$ criterion with $L = 50$. The last column reports the agreement between $CCFAP$-$CARL$ and $\widehat{CARL}$ (smaller is better).

N	m	n	UCL	$CCFAP$	$CCFAP$-$CARL$	$\widehat{CARL}$	|Gap|(%)
1000	300	5	2.4591	0.2151	207.06	201.95	2.53
		10	2.5049	0.1542	299.48	308.90	3.05
	500	5	2.4664	0.1684	272.36	270.64	0.64
		10	2.5092	0.2962	142.97	154.73	7.60
5000	300	5	2.4591	0.2868	148.44	146.47	1.34
		10	2.5049	0.0896	533.15	562.31	5.19
	500	5	2.4664	0.3656	110.38	112.19	1.61
		10	2.5092	0.1404	331.00	334.02	0.90
10000	300	5	2.4591	0.2829	150.86	149.85	0.67
		10	2.5049	0.2275	194.21	197.55	1.69
	500	5	2.4664	0.1958	229.96	227.75	0.97
		10	2.5092	0.1555	296.34	289.03	2.53
20000	300	5	2.4591	0.4154	93.66	93.96	0.32
		10	2.5049	0.1861	243.47	237.22	2.63
	500	5	2.4664	0.2329	189.09	189.82	0.38
		10	2.5092	0.2225	199.18	196.75	1.24

Chapter 4

Performance Evaluation

The performance of the proposed monitoring schemes is evaluated using Monte Carlo simulation under the conditional and unconditional approaches for both control charts $MW^{(\mathcal{I})b}$ with $b = 0.5$ and $MW^{(\mathcal{D})}$. As for the reference sample and subgroup sizes, we have taken $m = \{100, 300, 500\}$ and $n = \{5, 10\}$, respectively. The parameters considered in the simulation were $L = 50$, $q = 95$, $T = 1000$, $N = 1000$ and $\text{ARL}_0 = 200$. As for the multivariate distributions we used the multivariate normal ($Norm_p$), multivariate t (t_p) and multivariate gamma (Gam_p), all simulations in this chapter follow this design, except for the OC performance, where $\delta = 0.5$, 1, 1.5, and 2 were used.

4.1 In-control performance

Because of the distribution-free property of the used statistics, the control limits were obtained following Algorithm 1. For each multivariate distribution and combination of m, n and p considered in this assessment, we simulate $T = 1000$ reference samples and obtain its empirical distribution of $CARL(x^{(t)})$ using the corresponding control limits calibrated previously. Recall that in order to study if the in-control control chart performance is guaranteed, we check the proportion of $CARL(x^{(t)}) > ARL_0$. The IC performance for both $MW^{(\mathcal{I})0.5}$ and $MW^{(\mathcal{D})}$ charts under the conditional and unconditional perspective is shown in Tables 4.1 and 4.2, respectively. These tables present the control limits, the $SDARL$, and the proportion of $CARL(x^{(t)})$, values greater than the nominal $ARL_0 = 200$ value (> 200). The first notable result about the proposed monitoring schemes is that, since they are distribution-free, their performance is not affected by the specific characteristics of the studied multivariate distributions, such as heavy tails or skewness. This aligns with findings from studies like those by Mukherjee and Marozzi (2020), which explored similar methodologies under such conditions. Also it can be noted that the UCL for the conditional perspective is always greater than the for the unconditional perspective. From Tables 4.1 and 4.2 we can clearly see the IC performance is guaranteed under the conditional perspective. We conclude this because having set a $q = 95$, for all the

scenarios simulated under this perspective the proportion of reference samples that lead to a $CARL(x^{(t)}) > ARL_0$ is very close to the desired level, which is 0.95. On the other hand we only get about a 0.55 proportion of reference samples that meet this criterion under the unconditional perspective. In other words, under the unconditional perspective a slightly less than half of simulated reference samples (or practitioners) would get a false alarm before the preespecified threshold ARL_0 since we are not taking into account the practitioner-to-practitioner variation. This contrasts with the conditional perspective in which we ensure with a great probability the absence of false alarms for any given reference sample $\mathcal{X} = x$. We can also notice for both $MW^{(\mathcal{I})0.5}$ and $MW^{(\mathcal{D})}$ control charts, how variability of $CARL(x^{(t)})$ decreases as the m value gets smaller, and increases with the batch size n. Similar behaviour was presented in Celano and Chakraborti (2020) when using the MW control chart for FHP processes. This counter-intuitive behavior is explained by the increase in sensitivity of the batch size. As an attempt to provide an intuitive explanation, a reduced spread of the Phase-I increases the chance of having an out-of-control observation, and the bigger the batch size, the greater the probability. The opposite happens when the Phase-I samples overestimate the spread of the data.

Figure 4.1 and Figure 4.2 display the boxplots with the 5th, 25th, 50th, 75th and 95th percentiles depicted in Tables 4.1 and 4.2. For each multivariate distribution the boxplots compare the empirical distribution of $CARL(x^{(t)})$ corresponding to the values of n. The dotted line represents the prespecified ARL_0. These figures allow us to graphically study the effects of considering either the conditional or unconditional perspective and to compare the IC performance of both $MW^{(\mathcal{D})}$ and $MW^{(\mathcal{I})0.5}$ control charts. From these plots is clear that the conditional perspective guarantees an in-control performance since the dotted line largely agrees with the 5th percentile of $CARL(x^{(t)})$. Also, from Figure 4.2 we can notice the great variability caused by larger values of $CARL$. This variation decreases from the $MW^{(\mathcal{I})0.5}$ chart to the $MW^{(\mathcal{D})}$ chart which can be explained by the size of the Phase-I sample of Mahalanobis distances. On the other hand this variability is less under the unconditional perspective but there is still a large proportion of Phase-I samples whose $CARL$ is less than the expected.

4.2 Out-of-control performance

In addition to the control of false alarms, a main objective in MSPC is the detection as soon as possible of any potential shift in the parameter to be monitored. In order to evaluate the power of our proposed control charts, we simulate out-of-control scenarios

Table 4.1: In-control $CARL$ for the $MW^{(\mathcal{I})0.5}$ chart under different (m,n). Standard deviation ($SDARL$) and proportion of in-control $CARL$s greater than target ARL_0 are displayed.

Sample sizes	m=100, n=5		m=100, n=10		m=300, n=5		m=300, n=10		m=500, n=5		m=500, n=10	
$MW^{(\mathcal{I})0.5}$ under conditional perspective												
UCL	2.781556		2.92572		2.696419		2.828441		2.653714		2.782467	
Distribution	$SDARL$	> 200	$SDARL$	> 200	$SDARL$	> 200	$SDARL$	> 200	$SDARL$	> 200	$SDARL$	> 200
$Norm_5$	7973.33	0.957	11824.68	0.950	1075.37	0.959	2320.48	0.965	343.47	0.954	794.97	0.953
t_5	8606.17	0.975	10780.95	0.952	932.26	0.953	2662.48	0.955	350.60	0.952	684.35	0.952
Gam_5	8496.57	0.950	11871.93	0.963	827.65	0.960	2809.08	0.957	326.25	0.947	656.16	0.957
$Norm_{10}$	7728.89	0.957	11229.93	0.956	925.00	0.955	2401.97	0.964	336.66	0.949	592.83	0.942
t_{10}	7990.65	0.951	11866.53	0.940	840.94	0.951	2186.81	0.946	374.26	0.946	717.36	0.946
Gam_{10}	8115.96	0.946	11628.81	0.948	693.20	0.968	2405.79	0.961	333.51	0.945	653.63	0.963
$Norm_{20}$	7294.01	0.961	11678.38	0.952	874.73	0.964	2297.20	0.965	360.34	0.935	598.67	0.958
t_{20}	8075.65	0.958	11489.94	0.956	714.37	0.965	3179.90	0.955	332.25	0.952	680.19	0.951
Gam_{20}	7790.37	0.960	11209.73	0.960	848.20	0.96	2458.41	0.950	323.88	0.946	777.49	0.951
$MW^{(\mathcal{I})0.5}$ under unconditional perspective												
UCL	2.430041		2.454062		2.459058		2.491407		2.459058		2.492662	
Distribution	$SDARL$	> 200	$SDARL$	> 200	$SDARL$	> 200	$SDARL$	> 200	$SDARL$	> 200	$SDARL$	> 200
$Norm_5$	1357.13	0.642	3592.82	0.638	223.35	0.556	447.95	0.600	125.29	0.571	195.72	0.539
t_5	1902.07	0.639	3971.57	0.662	195.39	0.542	353.58	0.604	117.28	0.575	199.03	0.562
Gam_5	1311.15	0.625	4264.03	0.665	209.60	0.568	359.82	0.588	119.7	0.528	214.72	0.523
$Norm_{10}$	1381.50	0.638	3843.35	0.62	207.58	0.571	543.69	0.565	135.86	0.535	194.65	0.551
t_{10}	1521.46	0.635	4608.29	0.639	206.3	0.582	383.66	0.604	126.73	0.524	183.78	0.558
Gam_{10}	2698.67	0.639	4942.26	0.657	191.32	0.601	350.12	0.559	112.5	0.529	210.48	0.527
$Norm_{20}$	2014.79	0.605	4552.65	0.668	177.71	0.568	510.26	0.608	128.19	0.545	216.57	0.547
t_{20}	1397.18	0.643	3309.23	0.617	186.90	0.558	394.29	0.585	132.58	0.566	222.13	0.543
Gam_{20}	2691.09	0.624	4158.85	0.650	184.60	0.558	534.14	0.593	129.94	0.563	199.03	0.547

Table 4.2: In-control $CARL$ for the $MW^{(\mathcal{D})}$ chart under different (m,n). Standard deviation ($SDARL$) and proportion of in-control $CARL$s greater than target ARL_0 are displayed.

Sample sizes	m=100, n=5		m=100, n=10		m=300, n=5		m=300, n=10		m=500, n=5		m=500, n=10	
$MW^{(\mathcal{D})}$ under conditional perspective												
UCL	2.70928		2.86175		2.63442		2.770437		2.60124		2.722521	
Distribution	$SDARL$	> 200	$SDARL$	> 200	$SDARL$	> 200	$SDARL$	> 200	$SDARL$	> 200	$SDARL$	> 200
$Norm_5$	1731.40	0.948	5102.66	0.956	246.25	0.949	599.53	0.953	161.78	0.949	272.33	0.955
t_5	2649.48	0.956	6088.97	0.949	265.70	0.941	545.18	0.955	142.62	0.947	243.93	0.948
Gam_5	1760.43	0.950	6970.09	0.939	246.15	0.950	539.25	0.952	147.42	0.947	257.77	0.950
$Norm_{10}$	2156.18	0.929	6918.96	0.944	247.96	0.955	560.41	0.96	154.51	0.929	356.98	0.967
t_{10}	2424.53	0.944	5821.36	0.949	272.95	0.952	488.58	0.946	150.09	0.955	267.10	0.955
Gam_{10}	2885.71	0.952	5679.76	0.944	278.61	0.960	807.98	0.959	140.06	0.942	255.54	0.953
$Norm_{20}$	3354.21	0.924	11209.73	0.960	270.00	0.953	579.57	0.950	143.47	0.947	267.69	0.963
t_{20}	3017.83	0.940	5931.12	0.951	267.91	0.955	490.43	0.955	145.43	0.931	287.84	0.954
Gam_{20}	2554.85	0.939	5712.24	0.950	270.31	0.942	522.01	0.960	154.44	0.952	262.08	0.960
$MW^{(\mathcal{D})}$ under unconditional perspective												
UCL	2.43890		2.45665		2.463875		2.514911		2.468782		2.519348	
Distribution	$SDARL$	> 200	$SDARL$	> 200	$SDARL$	> 200	$SDARL$	> 200	$SDARL$	> 200	$SDARL$	> 200
$Norm_5$	355.12	0.593	665.61	0.614	114.44	0.506	181.34	0.586	82.98	0.553	112.07	0.544
t_5	339.56	0.581	1352.80	0.577	113.60	0.542	190.64	0.568	78.95	0.540	117.23	0.535
Gam_5	385.25	0.573	948.76	0.591	109.99	0.559	186.67	0.561	79.69	0.539	113.08	0.544
$Norm_{10}$	390.86	0.589	1283.97	0.602	112.84	0.543	177.48	0.565	76.47	0.531	108.04	0.536
t_{10}	431.52	0.609	1102.69	0.617	114.67	0.541	168.06	0.604	83.63	0.572	113.76	0.548
Gam_{10}	515.43	0.589	1771.39	0.601	103.30	0.556	189.34	0.559	77.58	0.541	103.74	0.563
$Norm_{20}$	501.25	0.586	1979.37	0.605	116.21	0.549	179.44	0.559	76.84	0.544	109.74	0.545
t_{20}	307.26	0.573	1794.35	0.582	116.31	0.519	169.87	0.536	79.50	0.552	116.65	0.533
Gam_{20}	350.85	0.567	1794.35	0.582	113.42	0.558	168.91	0.565	76.14	0.540	109.02	0.555

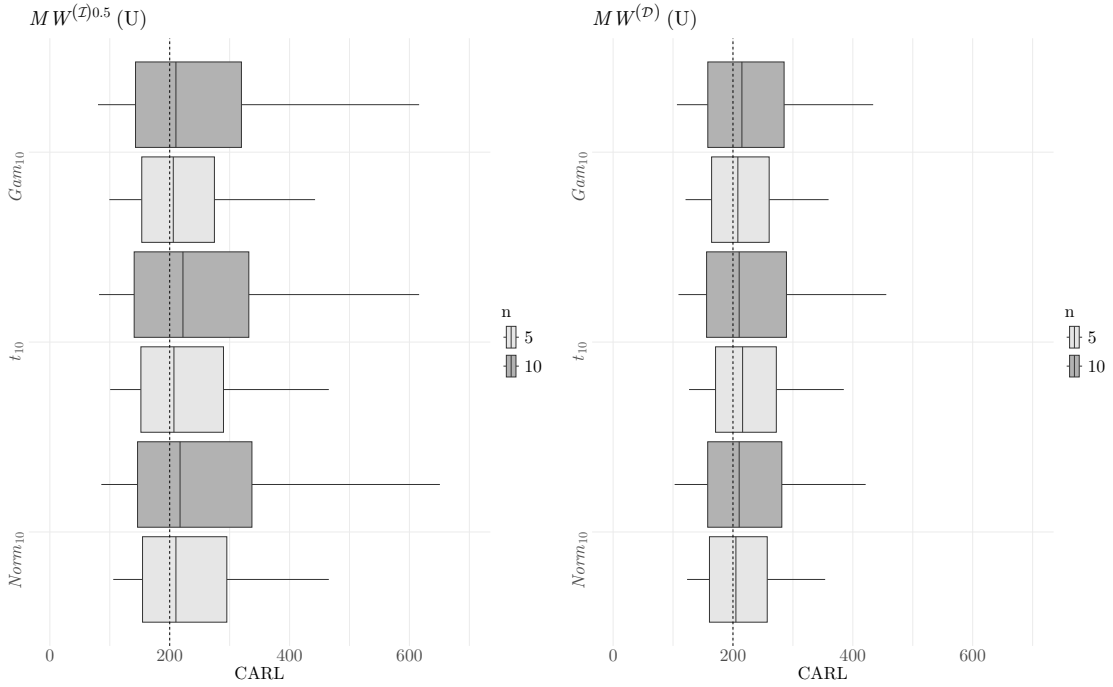


Figure 4.1: In-control performance of $MW^{(\mathcal{I})0.5}$ and $MW^{(\mathcal{D})}$ control charts under unconditional (U) perspective. A dimension of $p = 10$ and reference sample $m = 500$ is considered. Boxplots indicate the 5th, 25th, 50th, 75th and 95th percentiles.

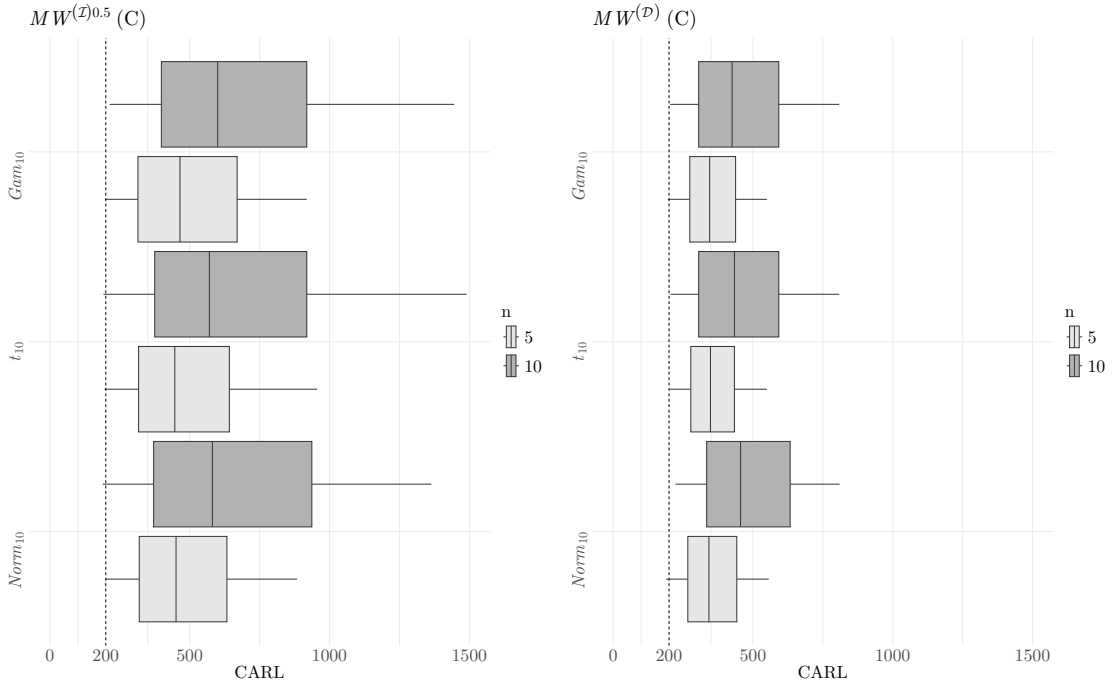


Figure 4.2: In-control performance of $MW^{(\mathcal{I})0.5}$ and $MW^{(\mathcal{D})}$ control charts under conditional (C) perspective. A dimension of $p = 10$ and reference sample $m = 500$ is considered. Boxplots indicate the 5th, 25th, 50th, 75th and 95th percentiles.

by considering a shift of size δ in the multivariate process location. To assess the OC performance we simulate $T = 1000$ reference samples to get the empirical distribution of the OC $CARL$ $(x^{(t)})$, $t = 1, \ldots, T$ by considering the different shift sizes $\delta = 0.5, 1, 1.5, 2$. Then we obtain the median of the OC $CARL$ $(x^{(t)})$ distribution to evaluate the OC performance of both $MW^{(\mathcal{D})}$ and $MW^{(\mathcal{I})0.5}$ control charts, each under conditional and unconditional perspectives. Similar to Celano and Chakraborti (2020), we select the median of the $CARL$ to evaluate performance. The results are displayed through Figures 4.3-4.8, plotting the median of the OC $CARL$ $(x^{(t)})$ against shift size δ.

Under the conditional perspective control limits tend to be wider, and although we guarantee an in-control performance, there is more variation with respect to the one under the unconditional perspective. Therefore, this will cause a degradation in the power to detect location shifts being the "price to pay" when using guaranteed control limits, see Saleh et al. (2016). In Figures 4.3-4.8 we can see that for a small shift $\delta = 0.5$ there is a large difference between the performance of both perspectives, the approaches built under the unconditional perspective being much more sensitive. Nevertheless both perspectives are not as effective in detecting this small shift as they would be expected with Shewhart type control charts. However, according to the OC performance results sensitivity can be considered as equivalent after some shift δ. In case of $Norm_p$ we can clearly see that OC performance of both perspectives are practically the same for a shift $\delta \geq 1$. For t_p and Gam_p we present a similar result for $\delta \geq 1$ although requiring a batch size of $n = 10$ for the $MW^{(\mathcal{I})0.5}$ and $MW^{(\mathcal{D})}$ control charts under the conditional perspective. Also, for the t_p distribution and under the conditional perspective, we observed that OC performance got worse when increasing the dimension p, this happens only for a batch size $n = 5$. However, this unique behaviour does not occur when we consider a value $n = 10$. Apart from this, no similar scenario was presented in the OC performance analysis.

When we compare Figure 4.3 versus Figure 4.5 it can be noticed that as more data is used the difference between perspectives becomes smaller. A similar conclusion can be obtained when comparing Figure 4.6 versus Figure 4.8. As for the comparison between the $MW^{(\mathcal{I})0.5}$ and $MW^{(\mathcal{D})}$ control charts, it is pretty clear that the latter has a better sensitivity to mean shifts. This is expected as we are using more data in the reference sample when implementing this approach. Now, we have seen that dependence between Mahalanobis distances does not affect the IC performance of $MW^{(\mathcal{D})}$ control chart as presented in section 4.1. Based on these results, we recommend using the $MW^{(\mathcal{D})}$ chart instead of the $MW^{(\mathcal{I})b}$ as it eliminates the need to split the reference sample-a process that, as observed, reduces the chart's sensitivity.

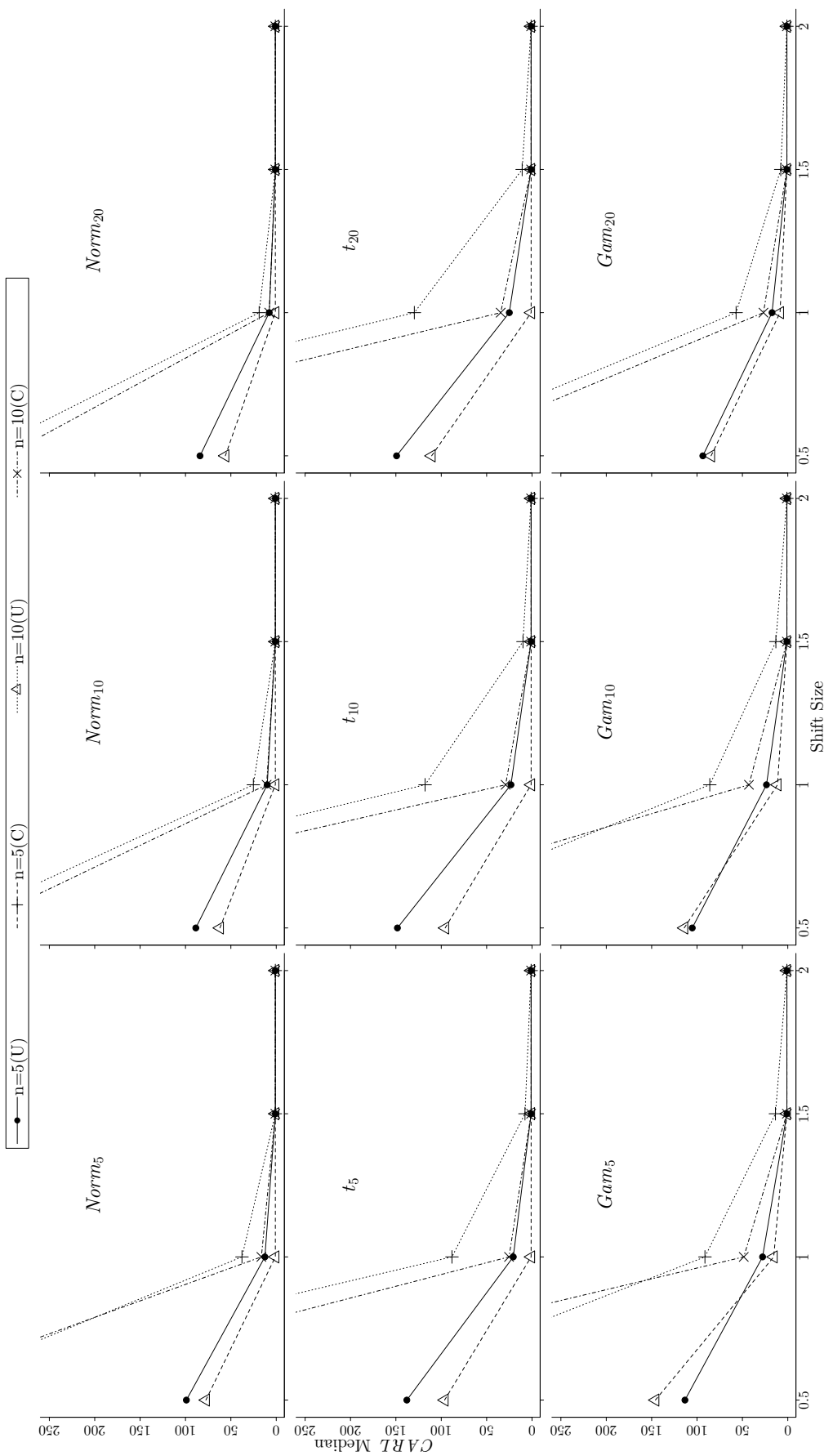


Figure 4.3: OC performance of $MW^{(\mathcal{I})0.5}$ control chart under conditional (C) and unconditional (U) perspectives. Reference sample $m = 100$, batch size $n = \{5, 10\}$ and dimension $p = \{5, 10, 20\}$ are considered.

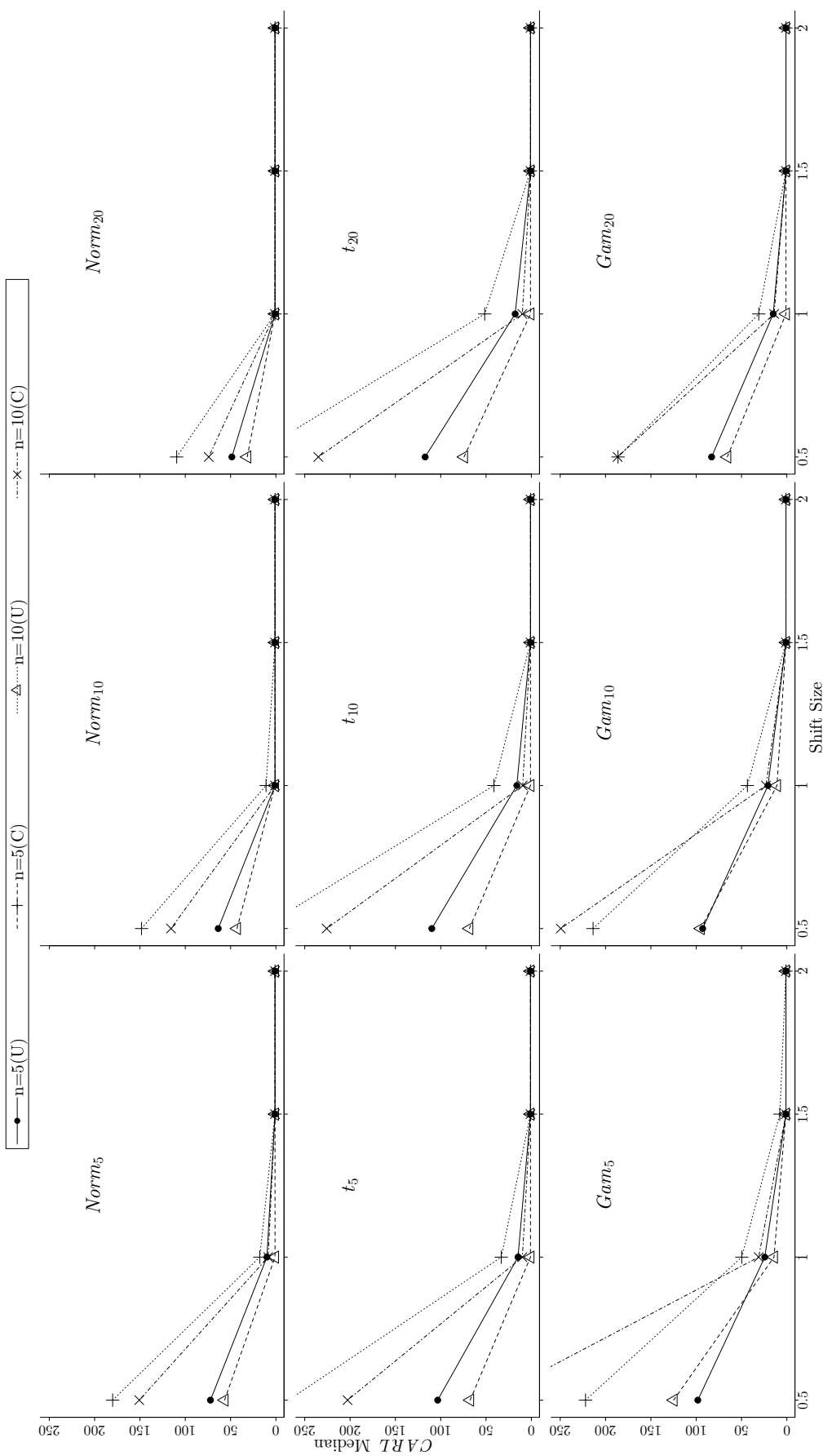


Figure 4.4: OC performance of $MW^{(\mathcal{I})0.5}$ control chart under conditional (C) and unconditional (U) perspectives. Reference sample $m = 300$, batch size $n = \{5, 10\}$ and dimension $p = \{5, 10, 20\}$ are considered.

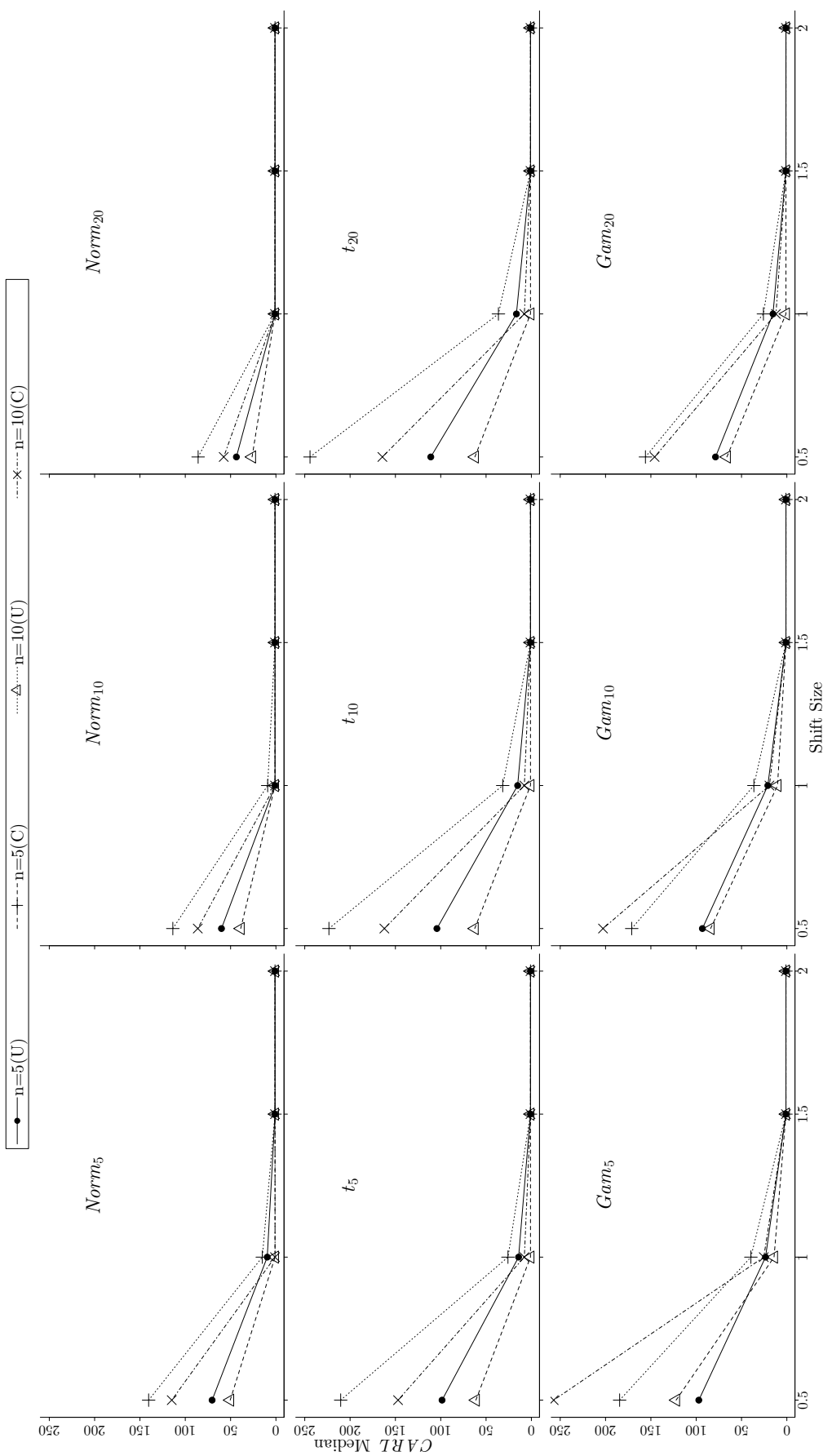


Figure 4.5: OC performance of $MW^{(\mathcal{I})0.5}$ control chart under conditional (C) and unconditional (U) perspectives. Reference sample $m = 500$, batch size $n = \{5, 10\}$ and dimension $p = \{5, 10, 20\}$ are considered.

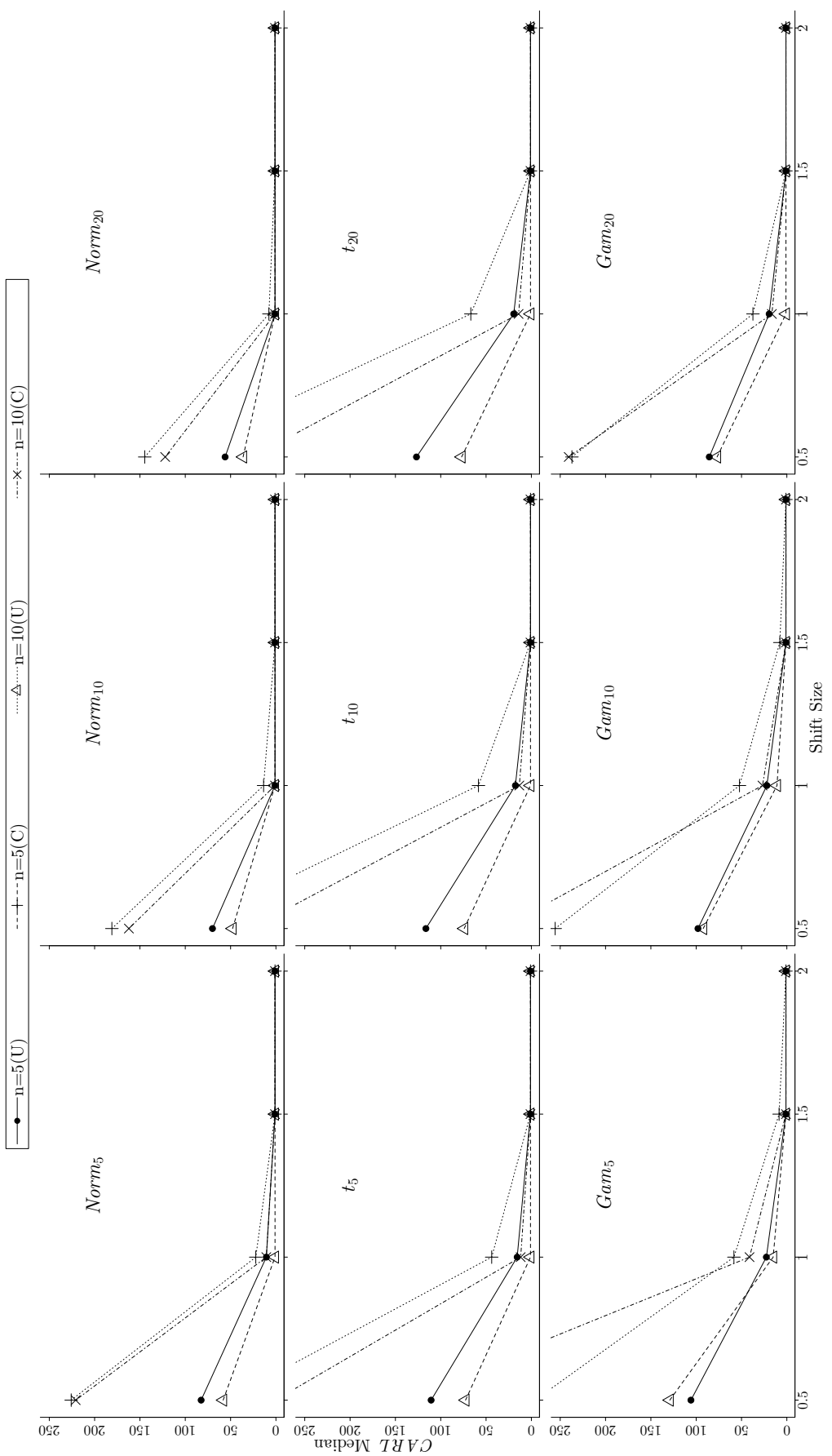


Figure 4.6: OC performance of $MW^{(\mathcal{D})}$ control chart under conditional (C) and unconditional (U) perspectives. Reference sample $m = 100$, batch size $n = \{5, 10\}$ and dimension $p = \{5, 10, 20\}$ are considered.

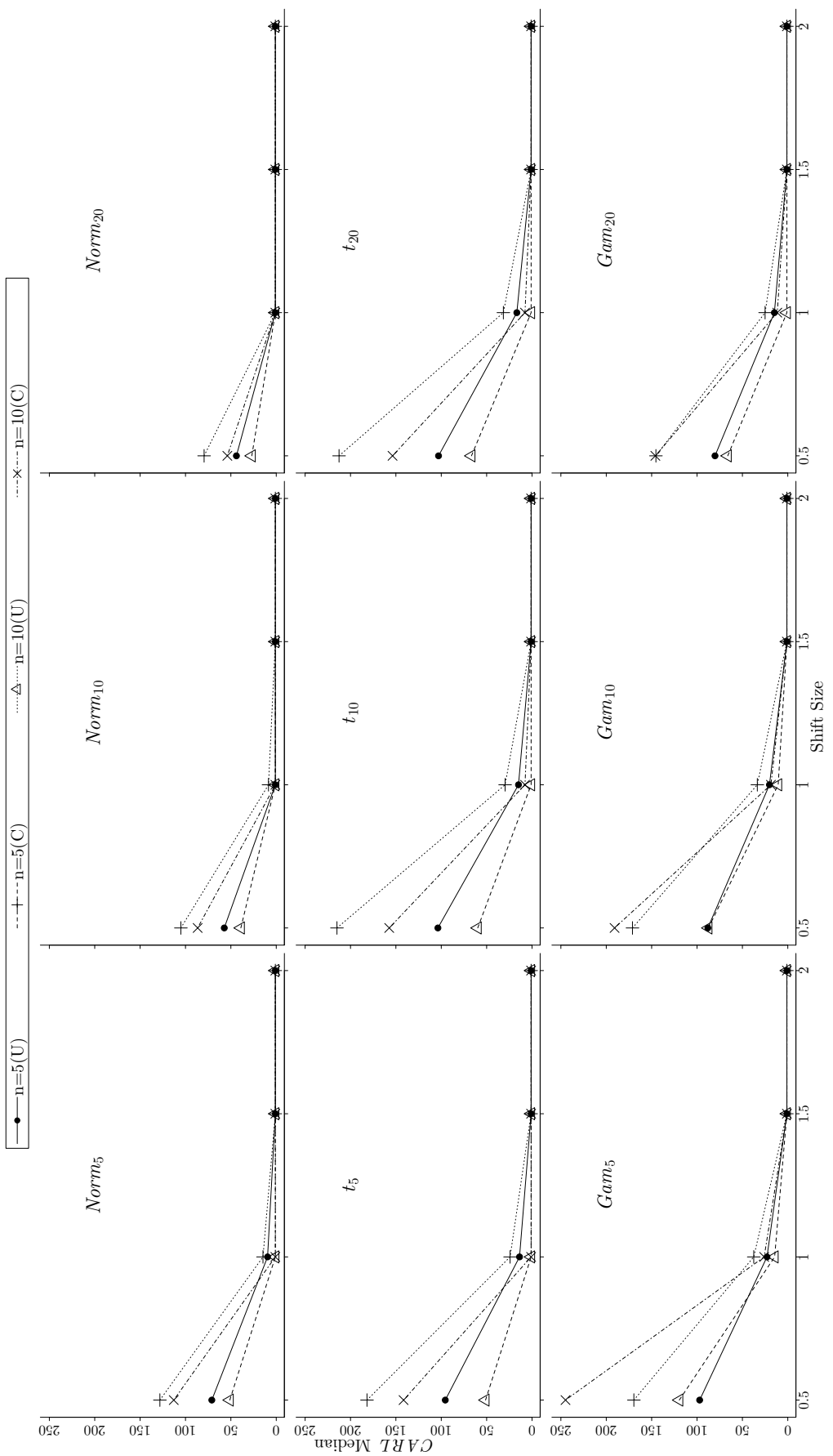


Figure 4.7: OC performance of $MW^{(\mathcal{D})}$ control chart under conditional (C) and unconditional (U) perspectives. Reference sample $m = 300$, batch size $n = \{5, 10\}$ and dimension $p = \{5, 10, 20\}$ are considered.

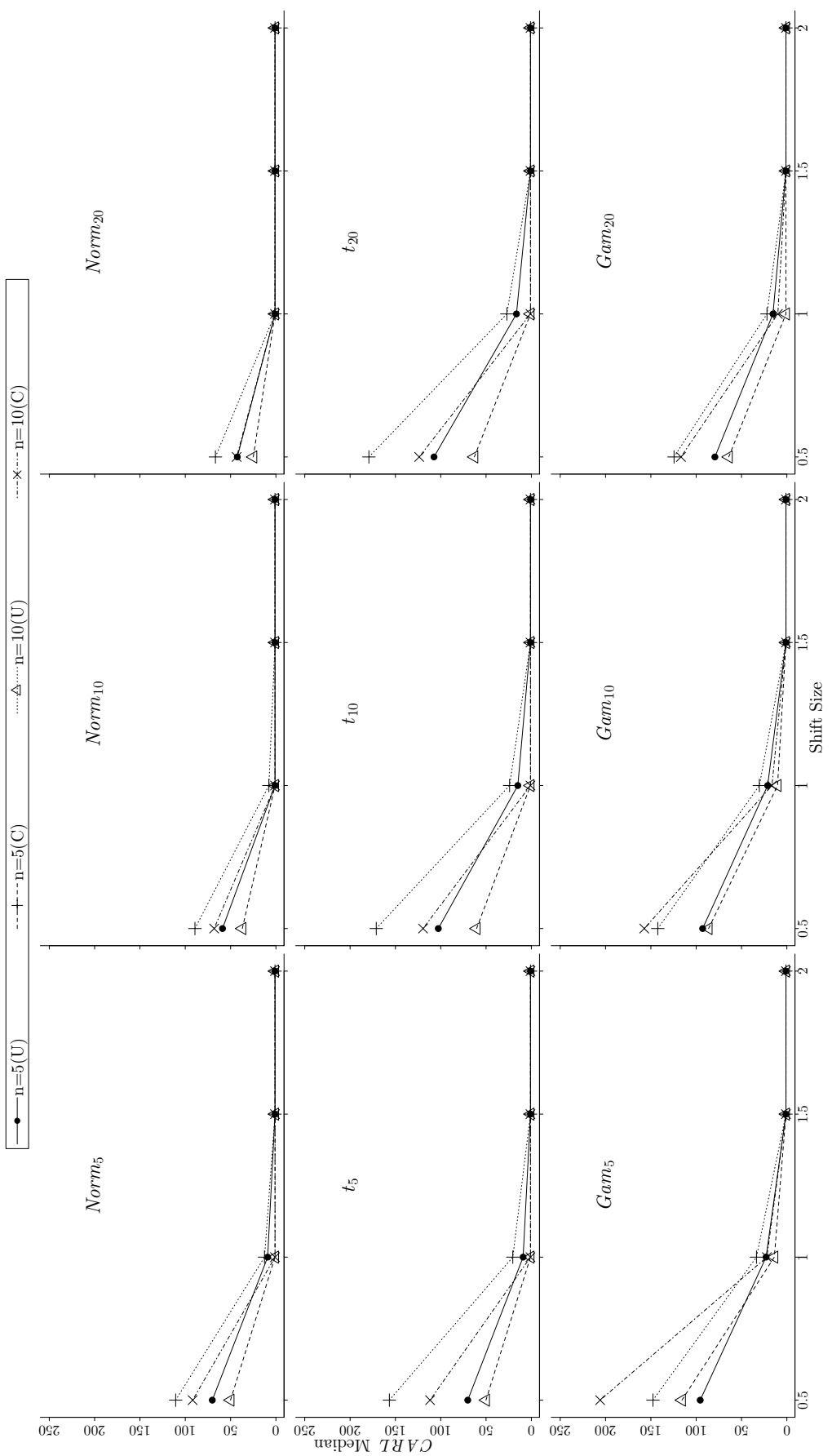


Figure 4.8: OC performance of $MW^{(\mathcal{D})}$ control chart under conditional (C) and unconditional (U) perspectives. Reference sample $m = 500$, batch size $n = \{5, 10\}$ and dimension $p = \{5, 10, 20\}$ are considered.

4.2.1 A Real Data Application

In this section we implement our proposed control chart in a real scenario related to the production of Portuguese *Vinho Verde* white wine. This dataset has been previously used under the context of MSPC to test the usefulness of different multivariate nonparametric control charts, for example consult Zou et al. (2011), Koutras and Sofikitou (2017), Yue and Liu (2017), Y. Li et al. (2020). This example has been chosen due to its recurrent use in the literature, making it a practical reference for researchers and practitioners to replicate and apply the proposed methodology. Provided by Cortez et al. (2009) we make use of Wine Quality Data (WQD), available in http://archive.ics.uci.edu/ml/machine-learning-databases/wine-quality/winequality-white.csv, which consists of 4898 observations collected from May 2004 to February 2007. There are 11 continuous attributes derived from physiochemical measurements that are possible indicators of wine quality (fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, alcohol) and an ordinal variable ranging from 0 (very bad) to 10 (excellent) which determines the quality of wine.

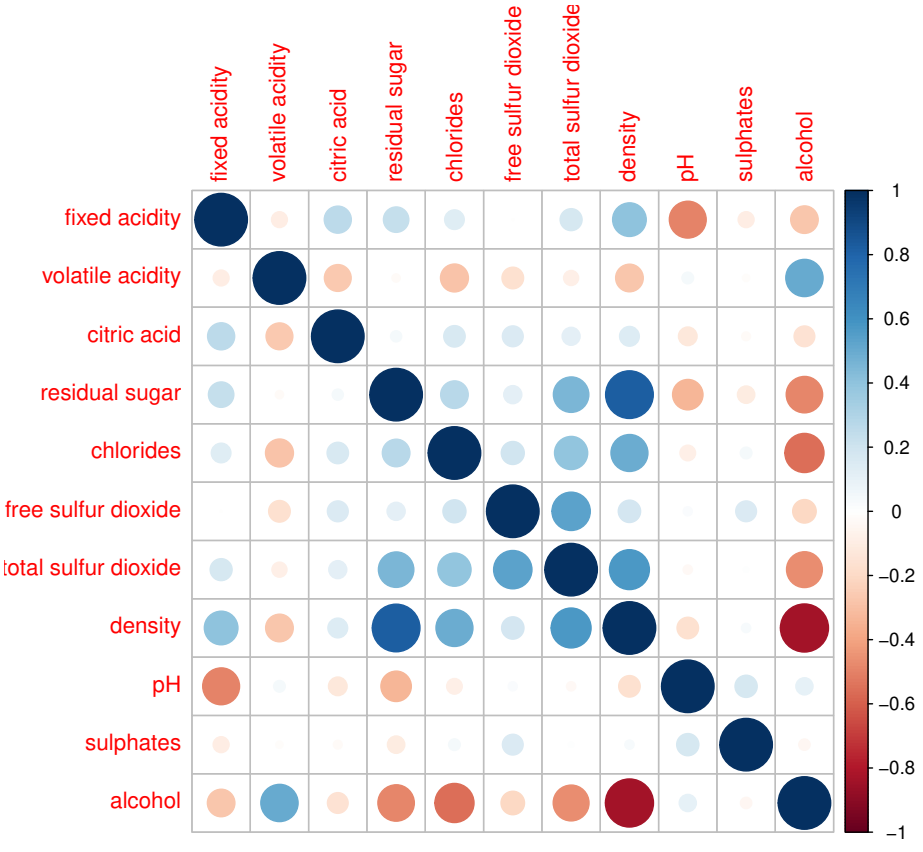


Figure 4.9: Sample correlation matrix for Wine Quality Data, as we can see, there are some variables that are strongly correlated.

As suggested by Cortez et al. (2009), we take level (LVL) 7 as the standard wine quality level. Figure 4.9 shows the sample correlation matrix of this dataset which demonstrates a considerable interrelationships between the variables. Therefore, a multivariate control chart is more appropriate to address this type of problem. With respect to normality, the analysis of the three variables volatile acidity, citric acid and residual sugar made by the scatterplots depicted in Figure 4.10 (a-c) concludes that bivariate data is not normally distributed. Moreover, the QQ-plots in Figures 4.10(d-f) indicates that the marginals are not normal either. In addition to the graph exploration, we also run the Shapiro-Wilk normality test for all the variables as well as the Mardia's multivariate normality test, all the tests performed with a p-value < 0.001. Recall that a dataset, in order to be considered to have a multivariate normal distribution, must have normally distributed marginals as well as be jointly normally distributed. Therefore, with all the previous information together with the plots shown in Figure 4.10 we can suggest that the multivariate normality assumption is invalid making it a very good option for the implementation of our proposal.

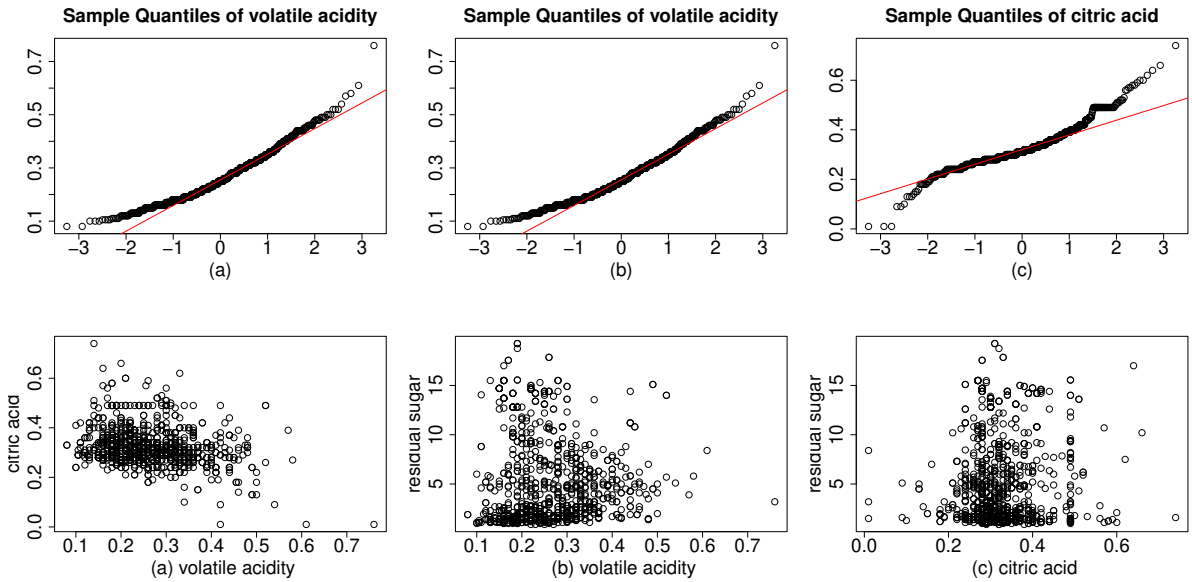


Figure 4.10: Scatter plots for three variables of Wine Quality Data (a-c). Normal Q-Q plots for the volatile acitidy, citric acid and residual sugar respectively (d-f).

Now, we randomly take $m = 500$ observations corresponding to LVL 7 which will form our in-control Phase-I sample. In order to implement a Phase-II monitoring, we randomly chose 100 observations from LVL 7 followed by 50 LVL 5 observations. Then these 150 observations are grouped into batches of size $n = 5$. This leads to a monitoring of 30 samples of which 20 are considered as in-control and 10 as out-of-control. The selected control chart, $MW^{(\mathcal{D})}$, is implemented following the steps outlined in Algorithm 1. Its

upper control limits (UCL), presented in Table 4.2, are 2.60124 and 2.468782 under the conditional and unconditional perspectives, respectively. Figure 4.11 displays the resulting process monitoring of WQD using the $MW^{(\mathcal{D})}$ control chart. When the process is considered as IC, the plotting statistics are below the UCL under both perspectives. Once the OC state starts it can be seen that control chart issues an alarm right after the first OC batch under both unconditional and conditional UCLs. By successfully implementing the $MW^{(\mathcal{D})}$ statistic in both IC and OC states of WQD, we have demonstrated the practical application of our proposal for monitoring a real-world non-normal multivariate process.

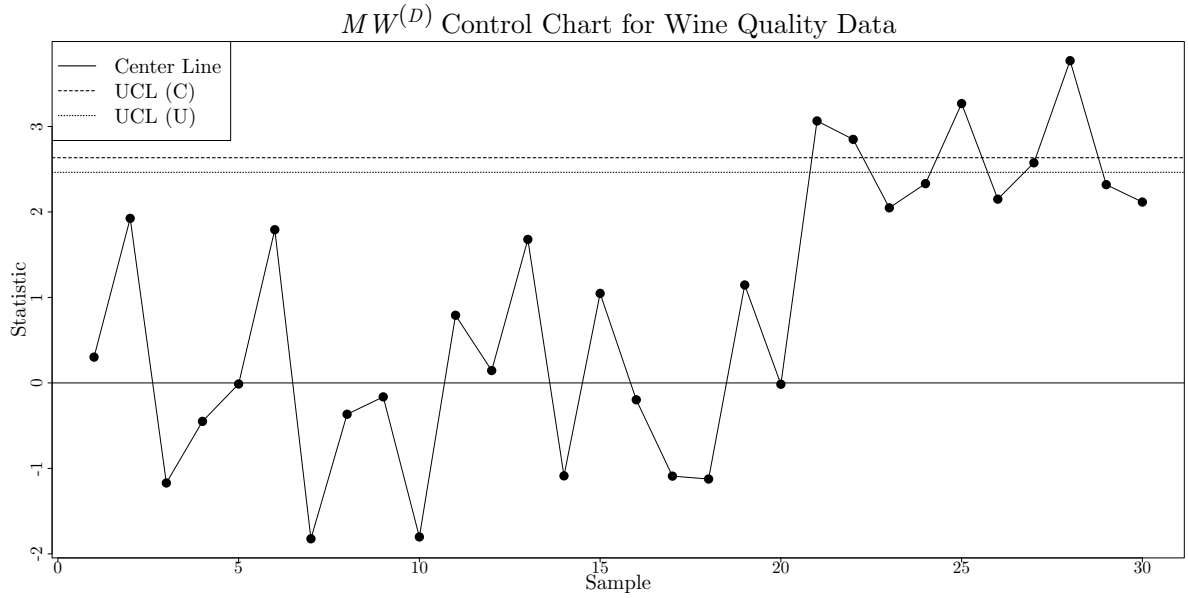


Figure 4.11: Monitoring of 11 quality attributes from the Wine Quality Data (Cortez et al. (2009)) through the proposed multivariate $MW^{(\mathcal{D})}$ plotting statistic with guaranteed control limits [UCL(C)] and compared with traditional limits [UCL (U)].

Chapter 5
Conclusions

A Shewhart type multivariate Phase-II control chart is proposed in this research. It is based on the ranking of Mahalanobis distances and no assumption about the mean vector or covariance matrix is made for the multivariate distribution to be monitored. This monitoring scheme is implemented in two different approaches which are identified as $MW^{(\mathcal{D})}$ and $MW^{(\mathcal{I})b}$ charts. Both are very similar but the former uses the whole reference sample to construct dependent Mahalanobis distances while the latter splits the reference sample to assure independence of these distance measures. In addition, since the performance of Phase-II control charts is conditional to the Phase-I sample, the control limits of both proposals were adjusted under the unconditional and conditional perspectives, studying the scopes and limitations of each one. Under the unconditional perspective we found that, although we achieved in average a desired IC performance, there could still be a large proportion of practitioners that have an IC performance below the expected. On the other hand, the conditional perspective takes into account the practitioner-to-practitioner variation to guarantee a minimum performance for the practitioners with their corresponding selected Phase-I sample. One drawback is that the sensitivity of these approaches is affected since guaranteed control limits tend to be wider than the ones obtained under the unconditional perspective. However, difference in mean shifts detection becomes equivalent under both perspectives for shifts $\delta \geq 1$ and as we acquire more information of the process. Because of the results consulted in the performance assessment, we recommend the use of $MW^{(\mathcal{D})}$ chart whose inherent dependence does not affect its IC performance and has a better sensitivity than the $MW^{(\mathcal{I})b}$ chart, although one drawback is that the former is more demanding computationally. In general, both proposals can be considered as simple tools that are very easy to implement so they might result very useful for the practitioner in multivariate processes where non-normal conditions are presented.

5.1 Future Work

Since the proposal presented in this research is innovative, there are several research lines that can be followed as a future work:

- The proposed charting statistic was implemented in Shewhart-type control chart, but this can be extended to a CUSUM or EWMA monitoring scheme in order to detect small and sustained shifts.
- Once the previous point is addressed, guaranteed control limits can be implemented in order to take into account the practitioner-to-practitioner variation.
- The extension of this work to monitor other parameter distribution such as the variance, either individually or a jointly monitoring with the mean making use of a Lepage-type statistic.
- The $CCFAP$-$CARL$ relationship is presented as a very good option to reduce the computational effort when an ARL simulation is addressed, mainly for the calibration of guaranteed control limits. Therefore, an interesting research topic is a whole study on the performance of this relationship under different scenarios.

Bibliography

Albers, W., & Kallenberg, W. C. M. (2005). New corrections for old control charts. *Quality Engineering*, *17*(3), 467–473.

Aly, A. A., Mahmoud, M. A., & Hamed, R. (2016). The performance of the multivariate adaptive exponentially weighted moving average control chart with estimated parameters. *Quality and Reliability Engineering International*, *32*(3), 957–967.

Boone, J., & Chakraborti, S. (2012). Two simple Shewhart-type multivariate nonparametric control charts. *Applied Stochastic Models in Business and Industry*, *28*, 130–140.

Cabana, E., & Lillo, R. E. (2021). Robust multivariate control chart based on shrinkage for individual observations. *Journal of Quality Technology*, 1–26.

Capizzi, G., & Masarotto, G. (2016). Efficient control chart calibration by simulated stochastic approximation. *IIE Transactions*, *48*(1), 57–65.

Capizzi, G., & Masarotto, G. (2017). Phase I Distribution-Free Analysis of Multivariate Data. *Technometrics*, *59*(4), 484–495.

Capizzi, G., & Masarotto, G. (2020). Guaranteed in-control control chart performance with cautious parameter learning. *Journal of Quality Technology*, *52*(4), 385–403.

Celano, G., & Chakraborti, S. (2020). A distribution-free Shewhart-type Mann-Whitney control chart for monitoring finite horizon productions. *International Journal of Production Research*, *0*(0), 1–18.

Chakraborti, S. (2000). Run length, average run length and false alarm rate of shewhart $\bar{X}$ chart: Exact derivations by conditioning. *Communications in Statistics - Simulation and Computation*, *29*(1), 61–81.

Chakraborti, S., & Graham, M. A. (2019). Nonparametric (distribution-free) control charts: An updated overview and some results. *Quality Engineering*, *31*(4), 523–544.

Chakraborti, S. (2006). Parameter estimation and design considerations in prospective applications of the $\bar{X}$ chart. *Journal of Applied Statistics*, *33*(4), 439–459.

Chakraborti, S., & Van de Wiel, M. A. (2008). A nonparametric control chart based on the Mann-Whitney statistic, 156–172.

Chen, N., Zi, X., & Zou, C. (2015). A Distribution-Free Multivariate Control Chart. *Technometrics*, *58*(4), 448–459.

Cheng, C.-R., & Shiau, J.-J. H. (2015). A Distribution-Free Multivariate Control Chart for Phase I Applications. *Quality and Reliability Engineering International*, *31*(1), 97–111.

Chou, Y.-M., Mason, R. L., & Young, J. C. (2001). The control chart for individual observations from a multivariate non-normal distribution. *Communications in statistics-Theory and methods*, *30*(8-9), 1937–1949.

Cortez, P., Cerdeira, A., Almeida, F., Matos, T., & Reis, J. (2009). Modeling wine preferences by data mining from physicochemical properties [Smart Business Networks: Concepts and Empirical Evidence]. *Decision Support Systems*, *47*(4), 547–553.

Crosier, R. B. (1988). Multivariate generalizations of cumulative sum quality-control schemes. *Technometrics*, *30*(3), 291–303.

Das, N. (2009). A new multivariate non-parametric control chart based on sign test. *Quality Technology & Quantitative Management*, *6*(2), 155–169.

Diko, M. D., Goedhart, R., Chakraborti, S., Does, R. J. M. M., & Epprecht, E. K. (2017). Phase II control charts for monitoring dispersion when parameters are estimated. *Quality Engineering*, *29*(4), 605–622.

Diko, M. D., Chakraborti, S., & Does, R. J. (2019). Guaranteed in-control performance of the EWMA chart for monitoring the mean. *Quality and Reliability Engineering International*, *35*(4), 1144–1160.

Does, R. J., Goedhart, R., & Woodall, W. H. (2020). On the design of control charts with guaranteed conditional performance under estimated parameters. *Quality and Reliability Engineering International*, *36*(8), 2610–2620.

Dovoedo, Y., & Chakraborti, S. (2017). Effects of parameter Estimation on the multivariate distribution-free Phase II Sign EWMA chart. *Quality and Reliability Engineering International*, *33*(2), 431–449.

Epprecht, E. K., Loureiro, L. D., & Chakraborti, S. (2015). Effect of the Amount of Phase I Data on the Phase II Performance of S^2 and S Control Charts. *Journal of Quality Technology*, *47*(2), 139–155.

Gandy, A., & Kvaløy, J. T. (2013). Guaranteed conditional performance of control charts via bootstrap methods. *Scandinavian Journal of Statistics*, *40*(4), 647–668.

Goedhart, R., Schoonhoven, M., & Does, R. J. M. M. (2017). Guaranteed In-Control Performance for the Shewhart X and $\bar{X}$ Control Charts. *Journal of Quality Technology*, *49*(2), 155–171.

Goedhart, R., Schoonhoven, M., & Does, R. J. (2016). Correction factors for Shewhart X and $\bar{X}$ control charts to achieve desired unconditional ARL. *International Journal of Production Research*, *54*(24), 7464–7479.

Goedhart, R., Schoonhoven, M., & Does, R. J. (2020). Nonparametric control of the conditional performance in statistical process monitoring. *Journal of Quality Technology*, *52*(4), 355–369.

Healy, J. (1987). A note on multivariate CUSUM procedures. *Technometrics,* 409–412.

Hotelling, H. (1947). Multivariate quality control. *Techniques of Statistical Analysis (C. Eisenhart, M. Hastay, and W.A. Wallis, eds.),* 111–184.

Huang, S., Yang, J., & Mukherjee, A. (2018). Distribution-free EWMA schemes for simultaneous monitoring of time between events and event magnitude. *Computers & Industrial Engineering*, *126*, 317–336.

Jardim, F. S., Chakraborti, S., & Epprecht, E. K. (2019). $\bar{X}$ Chart with Estimated Parameters: The Conditional ARL Distribution and New Insights. *Production and Operations Management*, *28*(6), 1545–1557.

Jardim, F. S., Chakraborti, S., & Epprecht, E. K. (2020). Two perspectives for designing Phase II control chart with estimated parameters: The case of the Shewhart $\bar{X}$ Chart. *Journal of Quality Technology*, *52*(2), 198–217.

Jensen, W. A., Jones-Farmer, L. A., Champ, C. W., & Woodall, W. H. (2006). Effects of parameter estimation on control chart properties: A literature review. *Journal of Quality Technology*, *38*(4), 349–364.

Jones, L. A., Champ, C. W., & Rigdon, S. E. (2001). The performance of exponentially weighted moving average charts with estimated parameters. *Technometrics*, *43*(2), 156–167.

Jones, L. A., Champ, C. W., & Rigdon, S. E. (2004). The Run Length Distribution of the CUSUM with Estimated Parameters. *Journal of Quality Technology*, *36*(1), 95–108.

Jones, M. A., & Steiner, S. H. (2012). Assessing the effect of estimation error on risk-adjusted cusum chart performance. *International journal for quality in health care*, *24*(2), 176–181.

Jurečková, J., & Kalina, J. (2012). Nonparametric multivariate rank tests and their unbiasedness. *Bernoulli*, *18*(1), 229–251.

Koutras, M. V., & Sofikitou, E. M. (2017). A bivariate semiparametric control chart based on order statistics. *Quality and Reliability Engineering International*, *33*(1), 183–202.

Li, W., Zhang, C., Tsung, F., & Mei, Y. (2020). Nonparametric monitoring of multivariate data via KNN learning. *International Journal of Production Research*, 1–16.

Li, Y., Pei, D., & Wu, Z. (2020). A multivariate non-parametric control chart based on run test. *Computers & Industrial Engineering*, *149*, 106839.

Lowry, C. A., Woodall, W. H., Champ, C. W., & Rigdon, S. E. (1992). A multivariate exponentially weighted moving average control chart. *Technometrics*, *34*(1), 46–53.

Mason, R. L., Chou, Y.-M., & Young, J. C. (2001). Applying Hotelling's T^2 Statistic to Batch Processes. *Journal of Quality Technology*, *33*(4), 466–479.

Mathur, S. K. (2009). A new nonparametric bivariate test for two sample location problem. *Statistical Methods and Applications*, *18*(3), 375.

Montgomery, D. C. (2020). *Introduction to statistical quality control*. John Wiley & Sons.

Mukherjee, A., Cheng, Y., & Gong, M. (2017). A new nonparametric scheme for simultaneous monitoring of bivariate processes and its application in monitoring service quality. *Quality Technology & Quantitative Management*.

Mukherjee, A., & Marozzi, M. (2020). Nonparametric Phase-II control charts for monitoring high-dimensional processes with unknown parameters. *Journal of Quality Technology*, *0*(0), 1–21.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, *41*(1/2), 100–115.

Phaladiganon, P., Kim, S. B., Chen, V. C. P., Baek, J.-G., & Park, S.-K. (2011). Bootstrap-based t^2 multivariate control charts. *Communications in Statistics - Simulation and Computation*, *40*(5), 645–662.

Psarakis, S., Vyniou, A. K., & Castagliola, P. (2014). Some recent developments on the effects of parameter estimation on control charts. *Quality and Reliability Engineering International*, *30*(8), 1113–1129.

Qiu, P. (2014). *Introduction to statistical process control*. CRC Press.

Qiu, P. (2008). Distribution-free multivariate process control based on log-linear modeling. *IIE Transactions*, *40*(7), 664–677.

Qiu, P. (2018). Some perspectives on nonparametric statistical process control. *Journal of Quality Technology*, *50*(1), 49–65.

Qiu, P., & Hawkins, D. (2001). A rank-based multivariate cusum procedure. *Technometrics*, *43*(2), 120–132.

Quesenberry, C. P. (1993). The Effect of Sample Size on Estimated Limits for $\bar{X}$ and X Control Charts. *Journal of Quality Technology*, *25*(4), 237–247.

Roberts, S. W. (1959). Control chart tests based on geometric moving averages. *Technometrics*, *1*(3), 239–250.

Saleh, N. A., & Mahmoud, M. A. (2017). Accounting for Phase I sampling variability in the performance of the MEWMA control chart with estimated parameters. *Communications in Statistics - Simulation and Computation*, *46*(6), 4333–4347.

Saleh, N. A., Mahmoud, M. A., Jones-Farmer, L. A., Zwetsloot, I., & Woodall, W. H. (2015). Another Look at the EWMA Control Chart with Estimated Parameters. *Journal of Quality Technology*, *47*(4), 363–382.

Saleh, N. A., Mahmoud, M. A., Keefe, M. J., & Woodall, W. H. (2015). The Difficulty in Designing Shewhart $\bar{X}$ and X Control Charts with Estimated Parameters. *Journal of Quality Technology*, *47*(2), 127–138.

Saleh, N. A., Zwetsloot, I. M., Mahmoud, M. A., & Woodall, W. H. (2016). Cusum charts with controlled conditional performance under estimated parameters. *Quality Engineering*, *28*(4), 402–415.

Stoumbos, Z. G., & Sullivan, J. H. (2002). Robustness to non-normality of the multivariate ewma control chart. *Journal of Quality Technology*, *34*(3), 260–276.

Testik, M., Runger, G., & Borror, C. (2003). Robustness properties of multivariate EWMA control charts. *Quality and Reliability Engineering International*, *19*(1), 31–38.

Yue, J., & Liu, L. (2017). Multivariate nonparametric control chart with variable sampling interval. *Applied Mathematical Modelling*, *52*, 603–612.

Zhang, C., Chen, N., & Zou, C. (2016). Robust multivariate control chart based on goodness-of-fit test. *Journal of Quality Technology*, *48*(2), 139–161.

Zhang, C., Chen, N., & Wu, J. (2020). Spatial rank-based high-dimensional monitoring through random projection. *Journal of Quality Technology*, *52*(2), 111–127.

Zou, C., Jiang, W., & Tsung, F. (2011). A LASSO-Based Diagnostic Framework for Multivariate Statistical Process Control. *Technometrics*, *53*(3), 297–309.

Zou, C., & Tsung, F. (2011). A Multivariate Sign EWMA Control Chart. *Technometrics*, *53*(1), 84–97.

Zou, C., Wang, Z., & Tsung, F. (2012). A spatial rank-based multivariate EWMA control chart. *Naval Research Logistics (NRL)*, *59*(2), 91–110.

List of Figures

List of Tables