DOCTORAT BRETAGNE LOIRE MATHSTIC

UNIVERSITÉ DE NANTES

THESE DE DOCTORAT

L'UNIVERSITE DE NANTES

ECOLE DOCTORALE N° 601
Mathématiques et Sciences et Technologies
de l'Information et de la Communication
Spécialité : *Mathématiques et leurs interactions*

Par
« Fatima Sehar ZAIDI »

« Development of Statistical Monitoring Procedures for Compositional Data»

Thèse présentée et soutenue à « Nantes », le « 05 Oct 2020 »
Unité de recherche : Laboratoire des Sciences du Numérique de Nantes

Rapporteurs avant soutenance :

Stelios PSARAKIS	Professeur des Universités, Athens University of Economics and Business, Athens, Greece
Biagio PALUMBO	Professeur des Universités, University of Naples Federico II, Naples, Italy

Composition du Jury :

Examinateurs :	Fernanda Otília FIGUEIREDO	Maître de Conférence, University of Porto, Porto, Portugal
	Marit SCHOONHOVEN	Maître de Conférence, Amsterdam Business School, Amsterdam, Netherland
	Giovanni CELANO	Maître de Conférence, Università di Catania, Catania, Italy
Dir. de thèse :	Philippe CASTAGLIOLA	Professeur des Universités, Université de Nantes, Nantes, France

Acknowledgments

Firstly, I would like to express my very special appreciation, sincere gratitude and thanks to my adviser Professor Philippe CASTAGLIOLA, for the confidence he gave me by agreeing to supervise me for this thesis and for the continuous support during my PhD study and related research, for his patience, motivation, and immense knowledge. I would also like to thank for his confidence, his encouragement, his dedication, his availability and especially his help in writing this manuscript allowed me to flourish serenely throughout my work. I could not have imagined having a better adviser and mentor for my PhD study.

Besides my advisor, I would also like to give my warm thank to the committee members, Professor Stelios PSARAKIS (Athens University of Economics and Business, Athens, Greece) and Professor Biagio PALUMBO (University of Naples Federico II, Naples, Italy) to show their interest in my work and to agree to be reviewer of my thesis. I would also like to thank Doctor Fernanda Otília FIGUEIREDO (University of Porto, Porto, Portugal), Doctor Marit SCHOONHOVEN (Amsterdam Business School, Amsterdam, Netherlands) and Doctor Giovanni CELANO (Università di Catania, Catania, Italy) for serving as my jury members.

Thank you to all members and ex-members of the SLP team for the support and encouragement. I would like to thank my fellow lab mates Dorra Rahali, Jorge Arturo Garza Venegas, Anan Tang for the stimulating discussions, for all the support and for helping me throughout my PhD. I thank everyone who contributed directly or indirectly to the accomplishment of this thesis.

I cannot finish the acknowledgments without thanking my family. My parents and my siblings supported me a lot during all these years of the thesis. They always believed in me and gave me moral support in doing this work.

I also especially thank my husband for his love, his support, and for being understanding and patient with me during all these years of study, and good luck with his thesis too. I truly thank Imran for sticking by my side, even when I was irritable and depressed. I feel that what we both learned a lot about life and strengthened our commitment and determination.

Finally, I thank my Husband and my Family for letting me through all the difficulties. Also I want to give my heartiest thank to Higher Education commission (HEC) of Pakistan to provide the financial and moral support throughout my PhD.

Acronyms

The reader will find below the main acronyms used throughout of this thesis.

ABBREVIATION	DEFINITION
ARL	Average Run Length
SDRL	Standard deviation of the run length
MCUSUM	Multivariate Cumulative sum
MEWMA	Multivariate Exponentially Weighted Moving Average
SPM	Statistical Process Monitoring
LCL	Lower Control Limit
UCL	Upper Control Limit
CoDa	Compositional data
$DDMA$	Data Depth Moving Average
MNOR	Multivariate normal

Notations	Descriptions
$\mathcal{S}^p$	Simplex sample space
$\mathbb{R}^p$	Real space
$\oplus$	Perturbation operator defined in Aitchison's geometry
$\odot$	Powering operator defined in Aitchison's geometry
a	Perturbation constant
b	Powering constant
$\mathbf{y}_i$	Independent multivariate normal random compositions
$\mathbf{x}_{i,j}$	Observable quality characteristics used to measure $\mathbf{y}_i$
$\varepsilon_{i,j}$	Random error term
$\mathbf{x}^*$	Ilr transformed value of x
μ^*	Composition mean
Σ^*	Composition variance
$\mu^*_{\bar{\mathbf{x}}}$	Composition mean of $\bar{\mathbf{x}}$
$\Sigma^*_{\bar{\mathbf{x}}}$	Composition variance of $\bar{\mathbf{x}}$
σ_{M}	Standard-deviation measurement error common in all directions
Σ^*_{M}	Measurement error variance-covariance matrix
δ	Non centerality parameter without measurement error
δ_{M}	Non centerality parameter in the presence of measurement error
λ	Eigenvalues
Δ	Percentage improvement operator
μ^*_0	Composition mean of $\mathbf{y}_i$ when the process is in-control
μ^*_1	Composition mean when the process is out-of-control
H	Upper control limit
r	Smoothing parameter
ξ	Clr transformed values of $\mathbf{x}$
$\mathcal{C}(\mathbf{x})$	Closure function of $\mathbf{x}$
κ	Constant sum of composition

I

State of the Art

General Introduction

SPM (Statistical Process Monitoring) is a methodology used to improve manufacturing processes by detecting process shifts (due to assignable causes) using control charts and by fixing the problems before a large number of nonconforming units is manufactured. The selection of a suitable control chart depends on the type and distribution of the data. When there are several quality characteristics, multivariate control charts have to be adopted. CoDa (for Compositional Data) is a specific category of multivariate data which are constrained by definition. CoDa are defined as vectors of strictly positive real numbers and they only convey relative information.

There are many contributions concerning CoDa in different fields like, health sciences, geology, statistical analysis and modeling of industrial data etc. Among them, there are several papers discussing SPM using control charts for CoDa. This thesis makes an attempt to study new control charts for CoDa. This thesis is divided into six chapters.

- The Chapter 1 introduces the Statistical process monitoring and defines some of the Univariate, nonparametric and multivariate control charts along with the performance measure of the control charts.
- The Chapter 2 introduces the CoDa along with some previous researches on CoDa.
- The Chapter 3 studies the performance of the Hotelling T^2 control chart for CoDa in the presence of measurement errors.
- The Chapter 4 studies the performance of the MEWMA-CoDa control chart in the presence of measurement errors.
- The Chapter 5 studies the performance of some nonparametric charts for compositional data using data depth.
- Finally the Chapter 6 contains the general conclusions of the thesis along with the perspective for future research for CoDa.

1

Introduction to Statistical Process Control

In this chapter we discuss introduction to statistical process monitoring along with control charts in details. We present some basic univariate, multivariate and nonparametric control charts and also discuss how to check the performance of control charts.

1.1 Introduction

SPM (Statistical Process Monitoring) is a widely used methodology, based on the implementation of *control charts*, for achieving process stability and improving capability through the reduction of the process variability .

The concept of statistical process control was introduced by Shewhart in 1920. Further Shewhart (1930) developed control charts to evaluate whether a process is in control or not. The charts introduced by Shewhart are also known as the Shewhart control charts. In 1946 the American society for quality control was established.

Statistical process control is a method used to differentiate between causes of variation (i.e. assignable or common causes). There are three main phases of statistical process control.

1. Emphasize on the process and find the control limits.
2. Eliminate the variation which are assignable to make the process stable.
3. Use control charts to monitor the production process by detecting changes in the mean or variance.

The basic tools of statistical process control are histograms, scatter diagrams, check sheets, pareto charts, cause and effect diagrams and control charts.

During the last decade, an enormous number of new advanced control charts have been proposed for univariate and multivariate processes with many applications in manufacturing and service sectors. Some of them are listed below.

1.2 Univariate control charts

There are different types of control charts. For example $(\bar{X}, R)$ control chart, $(\bar{X}, S)$ control chart, u chart, p chart etc. The Shewhart control charts are also known as memory less control charts as they only rely on the current information and they are less sensitive to detect small shifts but they are most efficient at detecting the large shifts.

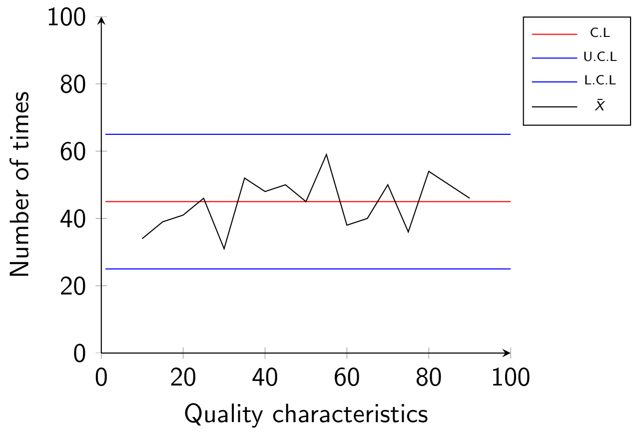

Figure 1.1 – $\bar{X}$ control chart

1.2.1 Control Charts for the Position

Suppose that the quality characteristic X is normally distributed with known mean μ and known standard deviation σ then a sample $\{X_1, \ldots, X_n\}$ of n observations has an average,

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

and its standard deviation is,

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

The confidence interval for the population mean μ can be written as,

$$\bar{X} + Z_{\frac{\alpha}{2}}\sigma_{\bar{X}} \quad \text{and} \quad \bar{X} - Z_{\frac{\alpha}{2}}\sigma_{\bar{X}}$$

where $Z_{\alpha/2}$ is the $\alpha/2$ quantile of the standard normal distribution which can be found in the table of standard deviation or evaluated with a software like R Core Team (2020). The control limits for the sample mean $\bar{X}$ can also be written as,

$$\mu + Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \quad \text{and} \quad \mu + Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \tag{1.1}$$

The confidence interval is also used as upper and lower control limits for the sample means and if the sample values $\bar{X}$ fall within the control limits and do not exhibit any systematic pattern, the process is said to be in-control.

While dealing with real life data the value of the real mean μ and standard deviation σ is unknown then the in-control sample statistics can be used instead of the real ones and the quality characteristics of the data can be studied using a subsample. Suppose there are m samples and each sample have n observations. Table 1.1 on the facing page shows subgroup data of m samples having n observations each. The mean of the sample means $\bar{X}_j$ can be written as,

$$\bar{\bar{X}} = \frac{1}{m}\sum_{j=1}^{m} \bar{X}_j$$

Table 1.1 – Table of $m \times n$ measures of $X_{i,j}$

Samples	Measures			
1	$X_{1,1}$	$X_{1,2}$	$\cdots$	$X_{1,n}$
2	$X_{2,1}$	$X_{2,2}$	$\cdots$	$X_{2,n}$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
m	$X_{m,1}$	$X_{m,2}$	$\cdots$	$X_{m,n}$

The mean of the sample median $\tilde{X}_j$ can be written as,

$$\bar{\tilde{X}} = \frac{1}{m}\sum_{j=1}^{m}\tilde{X}_j$$

The above mentioned limits in Equation 1.1 were the control limits for phase II Shewhart control chart when μ and σ were known. The control limits for sample mean $\bar{X}$ when μ and σ are known for phase I Shewhart control limit can be written as,

$$\begin{aligned}\text{LCL} &= \bar{\bar{X}} - k\hat{\sigma}/\sqrt{n}\\ \text{UCL} &= \bar{\bar{X}} + k\hat{\sigma}/\sqrt{n}\end{aligned}$$

where k is the "distance" of the control limits from the center line, expressed in standard deviation units. In general $k = 3$ has been used. As discussed earlier when the mean and standard deviation are unknown we use the above mean and standard deviations to determine the control limits for the sample mean. The averages can be written as,

$$\begin{aligned}\hat{\mu} &= \bar{\bar{X}}\\ &\text{or}\\ \hat{\mu} &= \bar{\tilde{X}}\end{aligned}$$

There are three different estimators for dispersion,

1. The first estimator of dispersion is the range (R). Suppose $R_1, R_2, \ldots, R_m$ are the ranges of m samples then the mean of R is $\bar{R}$ can be written as,

   $$\bar{R} = \frac{1}{m}\sum_{j=1}^{m}R_j$$

   The control limits for the $(\bar{X}, R)$ chart can be written as,

   $$\begin{aligned}\text{LCL}_{(\bar{X},R)} &= \bar{\bar{X}} - A_{(\bar{X},R)}(n) \times \bar{R}\\ \text{UCL}_{(\bar{X},R)} &= \bar{\bar{X}} + A_{(\bar{X},R)}(n) \times \bar{R}\end{aligned}$$

   where $A_{(\bar{X},R)}(n) = 3/d_2\sqrt{n}$ with d_2 depends on value of n. The values of $A_{(\bar{X},R)}(n)$ for different values of sample size n are listed in Table 1.2 on the next page.
   If we use $\tilde{X}$ instead of $\bar{X}$, then the control limits for the chart $(\tilde{X}, R)$ can be written as,

   $$\begin{aligned}\text{LCL}_{(\tilde{X},R)} &= \bar{\tilde{X}} - A_{(\tilde{X},R)}(n) \times \bar{R}\\ \text{UCL}_{(\tilde{X},R)} &= \bar{\tilde{X}} + A_{(\tilde{X},R)}(n) \times \bar{R}\end{aligned}$$

   where $A_{(\tilde{X},R)}(n)$ is a coefficient listed in Table 1.2 on the following page.

n	$A_{(\bar{X},R)}(n)$	$A_{(\bar{X},S)}(n)$	$A_{(\tilde{X},R)}(n)$	$L_S(n)$	$U_S(n)$	$L_R(n)$	$U_R(n)$	$K_S(n)$	$K_R(n)$
2	1.8800	2.6587		0.0021	4.0171	0.0021	4.0171	0.7979	1.1284
3	1.0233	1.9544	1.1972	0.0415	2.9006	0.0414	2.9247	0.8862	1.6926
4	0.7286	1.6281		0.1080	2.4775	0.1071	2.5257	0.9213	2.0588
5	0.5768	1.4273	0.6962	0.1730	2.2442	0.1705	2.3119	0.9400	2.3259
6	0.4832	1.2871		0.2293	2.0925	0.2245	2.1761	0.9515	2.5344
7	0.4193	1.1819	0.5123	0.2769	1.9841	0.2695	2.0808	0.9594	2.7044
8	0.3725	1.0991		0.3173	1.9019	0.3071	2.0095	0.9650	2.8472
9	0.3367	1.0317	0.4140	0.3519	1.8369	0.3389	1.9538	0.9693	2.9700
10	0.3083	0.9754		0.3818	1.7838	0.3660	1.9087	0.9727	3.0775
11	0.2851	0.9274	0.3519	0.4080	1.7395	0.3895	1.8714	0.9754	3.1729
12	0.2658	0.8859		0.4312	1.7018	0.4100	1.8398	0.9776	3.2585
13	0.2494	0.8495	0.3087	0.4518	1.6691	0.4281	1.8127	0.9794	3.3360
14	0.2354	0.8173		0.4704	1.6406	0.4442	1.7890	0.9810	3.4068
15	0.2231	0.7885	0.2766	0.4872	1.6153	0.4587	1.7681	0.9823	3.4718
16	0.2123	0.7626		0.5024	1.5928	0.4717	1.7495	0.9835	3.5320
17	0.2028	0.7391	0.2518	0.5164	1.5725	0.4836	1.7329	0.9845	3.5879
18	0.1943	0.7176		0.5292	1.5541	0.4944	1.7178	0.9854	3.6401
19	0.1866	0.6979	0.2319	0.5411	1.5374	0.5044	1.7041	0.9862	3.6890
20	0.1796	0.6797		0.5521	1.5220	0.5135	1.6915	0.9869	3.7350
21	0.1733	0.6629	0.2155	0.5623	1.5079	0.5220	1.6800	0.9876	3.7783
22	0.1675	0.6473		0.5719	1.4948	0.5299	1.6693	0.9882	3.8194
23	0.1621	0.6327	0.2018	0.5808	1.4827	0.5373	1.6595	0.9887	3.8583
24	0.1572	0.6191		0.5893	1.4714	0.5442	1.6503	0.9892	3.8953
25	0.1526	0.6063	0.1901	0.5972	1.4609	0.5506	1.6417	0.9896	3.9306

Table 1.2 – Coefficients used to calculate control limits for Shewhart control charts

2. The second estimator of dispersion is the standard deviation. Suppose $S_1, S_2, \ldots S_m$ are the standard deviations of m samples then the mean of S is $\bar{S}$ can be written as,

$$\bar{S} = \frac{1}{m}\sum_{j=1}^{m} S_j$$

The control limits for $(\bar{X}, S)$ can be written as,

$$\begin{aligned} \text{LCL}_{(\bar{X},S)} &= \bar{\bar{X}} - A_{(\bar{X},S)}(n) \times \bar{S} \\ \text{UCL}_{(\bar{X},S)} &= \bar{\bar{X}} + A_{(\bar{X},S)}(n) \times \bar{S} \end{aligned}$$

where $A_{(\bar{X},S)}(n)$ is a coefficient listed in Table 1.2.

3. The third estimator of dispersion is U. The mean of U can be written as,

$$\bar{U} = \sqrt{\frac{1}{m}\sum_{j=1}^{m} S_j^2}$$

The control limits can be written as,

$$\begin{aligned} \text{LCL} &= \bar{\bar{X}} - k\bar{U}/(c_{4,m}\sqrt{n}) \\ \text{UCL} &= \bar{\bar{X}} + k\bar{U}/(c_{4,m}\sqrt{n}) \end{aligned}$$

where $c_{4,m}$ can be found using the given formula,

$$c_{4,m} = \frac{\sqrt{2}\Gamma((m(n-1)+1)/2)}{\sqrt{m(n-1)}\Gamma(m(n-1)/2)}$$

where Γ is the gamma function.

1.2.2 Control Charts for the dispersion

Let a quality characteristics X is normally distributed with known parameter σ, then the control limits for range chart can be written as,

$$\begin{aligned} \text{LCL}_R &= L_R(n) \times \bar{R} \\ \text{UCL}_R &= U_R(n) \times \bar{R} \end{aligned}$$

where $L_R(n)$ and $U_R(n)$ are coefficients listed in Table 1.2 on the facing page.

When the standard deviation is used as an estimator, then the control limits of the S chart can be written as,

$$\begin{aligned} \text{LCL}_S &= L_S(n) \times \bar{S} \\ \text{UCL}_S &= U_S(n) \times \bar{S} \end{aligned}$$

where $L_S(n)$ and $U_S(n)$ are coefficients listed in Table 1.2 on the preceding page.

1.2.3 Control Charts for Nonconformity

There are some specific requirements of every thing being produced. Suppose glass bottles are being produced in a company. The company analyse a sample of the bottle according to the air bubbles on the bottle and they split the final product into two categories, i.e. conforming(good) item, when there is no air buble and nonconforming(defective) item, when there is air buble . This type of data that deals with quality of products is known as attributes data. Control charts for attributes proposed by Yang and Hillier (1970) further Woodall (1997), Stoumbos et al. (2000) and Montgomery (2001) worked on these charts to avoid the problem of nonconforming product and to minimize loss.
The fraction nonconforming is defined as the number of defective items (D) divided by total number of items (n). The probability of defective items is p.The random variable D follows a binomial distribution with parameters n and p.

$$P(D = X) = \binom{n}{X} p^X (1-p)^{n-X}, \qquad X = 0, 1, 2, \ldots, n$$

where $\hat{p}$ is defined as ratio of defectives in the sample (D) to sample size (n).

$$\hat{p} = \frac{D}{n}$$

where the mean and variance of $\hat{p}$ are,

$$\mu_{\hat{p}} = p \quad \text{and} \quad \sigma_{\hat{p}^2} = \frac{p(1-p)}{n}$$

The control limits for the p chart using Shewhart type control limits are,

$$\begin{aligned} \text{LCL} &= \max\left(0, p - 3\sqrt{\frac{p(1-p)}{n}}\right) \\ \text{UCL} &= p + 3\sqrt{\frac{p(1-p)}{n}} \end{aligned}$$

These limits are also known as phase II control limits. As in real life data the value of p is not known so it is estimated from a Phase I sample. Then the average of m samples each containing n observations is,

$$\bar{p} = \frac{\sum_{i=1}^{m} D_i}{mn} = \frac{\sum_{i=1}^{m} \hat{p}_i}{m}$$

The phase I control limits of the p chart are,

$$\begin{aligned} \text{LCL} &= \max\left(0, \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}\right) \\ \text{UCL} &= \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \end{aligned}$$

If the number of nonconforming units is plotted instead of the fraction nonconforming then another chart, the np control chart, is used with the following control limits,

$$\begin{aligned} \text{LCL} &= \max\left(0, np - 3\sqrt{np(1-np)}\right) \\ \text{UCL} &= np + 3\sqrt{np(1-np)} \end{aligned}$$

These limits are known as phase II control limits but if true value of p is not known then phase I limits will be used instead, they are,

$$\begin{aligned} \text{LCL} &= \max\left(0, n\bar{p} - 3\sqrt{n\bar{p}(1-n\bar{p})}\right) \\ \text{UCL} &= n\bar{p} + 3\sqrt{n\bar{p}(1-n\bar{p})} \end{aligned}$$

1.2.4 Control Charts for the number of Defects

The p and np chart plots only the number of defective items without taking into account the number of defects within a item. For example if we take the same example of a company producing glass bottles. A bottle may have one or more than one defects, so it is also important to control the number of defects within the item. The c chart is used to study the number of defects within an item. The number of defects are denoted by "c" and is assumed to follow a Poisson distribution. The c chart is applied only when the sample size is a constant. The average number of defects per item is denoted by $\bar{c}$ and is defined as,

$$\bar{c} = \frac{\text{Sum of defects}}{\text{Total number of samples inspected}} = \frac{1}{k}\sum_{i=1}^{k} c_i$$

where k is the total number of samples inspected. The control limits of the c chart are,

$$\begin{aligned} \text{LCL} &= \max(0, \bar{c} - 3\sqrt{\bar{c}}) \\ \text{CL} &= \bar{c} \\ \text{UCL} &= \bar{c} + 3\sqrt{\bar{c}} \end{aligned}$$

As discussed above that the c chart can only be used when the sample size a constant, but sometimes the sample size varies from one sample to another and, in this case, we use a u chart. The u chart can also be applied on data with constant sample size as well as variable sample size. The major difference between c and u chart is, in the u charts instead of monitoring the number of defects per sample, we monitor the number of defects per item. First we compute the number of defects per sample as,

$$u_1 = \frac{c_1}{n_1}, u_2 = \frac{c_2}{n_2}, \ldots, u_k = \frac{c_k}{n_k}$$

The average number of defects per item is denoted by $\bar{u}$ and is defined as,

$$\bar{u} = \frac{1}{k}\sum_{i=1}^{k} u_i$$

The control limits of the u chart are,

$$\begin{aligned} \text{LCL} &= \max\left(0, \bar{u} - 3\sqrt{\frac{\bar{u}}{\bar{n}}}\right) \\ \text{CL} &= \bar{u} \\ \text{UCL} &= \bar{u} + 3\sqrt{\frac{\bar{u}}{\bar{n}}} \end{aligned}$$

1.2.5 Cumulative Sum Control Charts

Page (1954) introduced Cumulative sum (CUSUM) control charts. Further, many other scientist like Ewan (1963) , Gan (1991) and Adam et al. (1992) worked on these charts. These charts overcome the problem of Shewhart chart as they are not memory less and they can be used to detect small as well as large shift in the values of the variables. As shown by the name "cumulative sum" these charts use successive sums of the observations.

Let $X_{j,k}$ be a independently identically distributed normal random variables with parameters (μ_0, σ_0). Let $\bar{X}_1, \ldots, \bar{X}_k, \ldots$ be the averages calculated for samples of size n. The CUSUM -$\bar{X}$ chart uses the sum of deviations between the averages $\bar{X}_j$ and the target value μ_0, such that if the process deviates significantly from this target value, the cumulative difference will increase, and will detect the shift even if it is very small. Using $\bar{X}_1, \ldots, \bar{X}_k, \ldots$, we get,

$$\begin{aligned} Y_k^+ &= \max\{0, \bar{X}_k - (\mu_0 + H) + Y_{k-1}^+\} \\ Y_k^- &= \max\{0, (\mu_0 - H) - \bar{X}_k + Y_{k-1}^-\} \end{aligned}$$

where $H = H_0\sigma_0$ is a reference value and H_0 is a constant which depends on n. The sequence of random variables $Y_1^+, \ldots, Y_k^+, \ldots$ allows to detect a positive shift in the mean, while the series of random variables $Y_1^-, \ldots, Y_k^-, \ldots$ allows to detect a negative shift. We usually takes $Y_0^- = Y_0^+ = 0$. A shift will be detected if,

$$\max(Y_k^-, Y_k^+) > \text{UCL}$$

where $\text{UCL} = K\sigma_0$ is a threshold and K is a constant that depends on n.

1.2.6 Exponentially weighted moving average control chart

Roberts (1958) introduced the Exponentially weighted moving average(EWMA) in order to improve the sensitivity for small shifts and to overcome the drawbacks of traditional Shewhart control charts. The control limits of the EWMA chart are given by,

$$\begin{aligned} \text{LCL} &= \mu_0 - K\sigma_0, \\ \text{UCL} &= \mu_0 + K\sigma_0, \end{aligned}$$

where $K > 0$. The statistic to be monitored is:

$$Y_i = (1 - \lambda)Y_{i-1} + \lambda\bar{X}_i,$$

where $\lambda \in (0, 1]$ is a smoothing constant.

An EWMA-$\bar{X}$ chart considers the process is in control if $Y_i \in [\text{LCL}, \text{UCL}]$. If we replace Y_{i-1} in function of Y_{i-2}, then Y_{i-2} in function of Y_{i-3}, etc, we get,

$$Y_i = (1 - \lambda)^i Y_0 + \lambda \sum_{j=0}^{i-1} (1 - \lambda)^j \bar{X}_{i-j}.$$

Thus, we can clearly see that Y_i is a linear combination of the initial random variable Y_0 weighted by a coefficient $(1-\lambda)^i$, and random variables $\bar{X}_1,\ldots,\bar{X}_i$ weighted by the coefficients $\lambda(1-\lambda)^{i-1},\lambda(1-\lambda)^{i-2},\ldots,\lambda$. It is the reason that the series $Y_1,\ldots,Y_i,\ldots$ is called EWMA series (Exponentially Weighted Moving Average). When the value of λ is close to 0, the series $Y_1,\ldots,Y_i,\ldots$ is " smoothed " version of the series $\bar{X}_1,\ldots,\bar{X}_i,\ldots$. Ultimately, for $\lambda=0$, we have $Y_i=Y_{i-1}=\cdots=Y_0$. Conversely, when λ is close to 1, the series $Y_1,\ldots,Y_i,\ldots$ looks similar to the series $\bar{X}_1,\ldots,\bar{X}_i,\ldots$. Ultimately, for $\lambda=1$, we have $Y_i=\bar{X}_i$, $i\geq 1$.

When the process is under control, we have $E(\bar{X}_i)=\mu_0$ and therefore the mathematical expectation of the random variable Y_i,$i\geq 1$ is equal to,

$$\begin{aligned} E(Y_i) &= (1-\lambda)^i E(Y_0)+\lambda\mu_0\sum_{j=0}^{i-1}(1-\lambda)^j \\ &= (1-\lambda)^i E(Y_0)+\lambda\mu_0\left(\frac{1-(1-\lambda)^i}{\lambda}\right). \end{aligned}$$

After simplification, we get,

$$E(Y_i)=(1-\lambda)^i E(Y_0)+\mu_0-\mu_0(1-\lambda)^i.$$

If we assume that $E(Y_0)=\mu_0$, then it simply remains,

$$E(Y_i)=\mu_0.$$

We can notice that even if $E(Y_0)\neq\mu_0$, then the asymptotic mathematical expectation from the series $Y_1,\ldots,Y_i,\ldots$ is equal to,

$$\lim_{i\to+\infty}E(Y_i)=\mu_0.$$

When the process is under control, we have $V(\bar{X}_i)=\frac{\sigma_0^2}{n}$ and, the random variables $\bar{X}_1,\bar{X}_2,\ldots$ being independent, the variance of the random variable Y_i is equal to,

$$\begin{aligned} V(Y_i) &= (1-\lambda)^{2i}V(Y_0)+\lambda^2\frac{\sigma_0^2}{n}\sum_{j=0}^{i-1}(1-\lambda)^{2j} \\ &= (1-\lambda)^{2i}V(Y_0)+\lambda^2\frac{\sigma_0^2}{n}\left(\frac{1-(1-\lambda)^{2i}}{\lambda(2-\lambda)}\right). \end{aligned}$$

After simplification, we get,

$$V(Y_i)=(1-\lambda)^{2i}V(Y_0)+\left(\frac{\lambda}{2-\lambda}\right)\frac{\sigma_0^2}{n}(1-(1-\lambda)^{2i}).$$

We can consider that,

1. let $Y_0=\mu_0$ be a constant and therefore $V(Y_0)=0$. In this case, we get

$$V(Y_i)=\left(\frac{\lambda}{2-\lambda}\right)\frac{\sigma_0^2}{n}(1-(1-\lambda)^{2i}).$$

2. either Y_0 is a random variable such that $E(Y_0)=\mu_0$ and $V(Y_0)=\frac{\sigma_0^2}{n}$. In this case, we get

$$V(Y_i)=\left(\frac{\lambda+2(1-\lambda)^{2i+1}}{2-\lambda}\right)\frac{\sigma_0^2}{n}.$$

This expression seems to us the most logical as it implies that $V(Y_i)$ is a decreasing function of i.

We can notice that whatever the value of $V(Y_0)$, the asymptotic variance of the series $Y_1, \ldots, Y_i, \ldots$ is equal to,

$$\lim_{i \to +\infty} V(Y_i) = \left(\frac{\lambda}{2-\lambda}\right)\frac{\sigma_0^2}{n}$$

For this reason, some authors prefer to use the following control limits for the EWMA $-\bar{X}$ chart:

$$\begin{aligned} \text{LCL} &= \mu_0 - K'\sqrt{\frac{\lambda}{n(2-\lambda)}}\sigma_0, \\ \text{UCL} &= \mu_0 + K'\sqrt{\frac{\lambda}{n(2-\lambda)}}\sigma_0. \end{aligned}$$

1.3 Nonparametric control charts

The vast majority of control chart depends on the assumptions of the underlying distribution. In most of the literature about statistical process monitoring the data are assumed to be normally distributed. However in real world situations, many times the data are not normally distributed. To deal with this type of data which does not fulfill the assumptions of normality some nonparametric control charts have been proposed during the last decades. Some of them are discussed below.

1.3.1 Signed-Rank (SR) Charts

Suppose a random sample $(X_{t1}, X_{t2}, \ldots, X_{tm})$ of subgroup size $n > 1$ collected at interval $t = 1, 2, \ldots$ is independent and follows a continuous symmetric distribution with an in-control median θ_0. Then the Shewhart-type chart based on the Signed-Rank (SR) is defined as,

$$\psi = \sum_{j=1}^{n} \text{sign}(X_{tj} - \theta_0) R_{tj}^{+} \qquad t = 1, 2, \ldots$$

where

$$R_{tj}^{+} = 1 + \sum_{j=1}^{n} I(|X_{ti} - \theta_0| < X_{tj} - \theta_0)$$

where R_{tj}^{+} is Wilcoxon signed-rank statistic (see for instance Gibbons and Chakraborti (2010)) and I is the indicator function with

$$\text{sign}(X) = \begin{cases} 1 & \text{if } X > 0, \\ 0 & \text{if } X = 0, \\ -1 & \text{if } X, 0, \end{cases}$$

The test statistics ψ is the the sum of the ranks of absolute values of the deviations corresponding to the negative deviations from the sum of the ranks of the absolute values of the deviations corresponding to the positive deviations.

1.3.2 Mann Whitney control charts

Let us assume that a reference sample $X = (X_1, X_2, \ldots, X_m)$ is an in-control independent random sample of size m, Let us also assume an arbitrary test sample $Y = (Y_1, Y_2, \ldots, Y_n)$ of size n, where X and Y are independent of each other. Then, the Mann Whitney statistics defined by Mann and Whitney (1947) is,

$$M_{XY} = \sum_{i=1}^{m}\sum_{j=1}^{n} I(X_i < Y_j)$$

where $I(X_i < Y_j)$ is the indicator function for the event $X_i < Y_j$ with $0 \leq M_{XY} \leq mn$. The shift is said to be positive when the value of M_{XY} is large, whereas the shift is said to be negative when the value of M_{XY} is small. The nonparametric Shewhart chart for location based on the MW statistics was first proposed by Chakraborti and de Wiel (2008).

1.3.3 Ansari and Bradley control chart

Suppose we have two independent sample $X = (X_1, X_2, \ldots, X_m)$ and $Y = (Y_1, Y_2, \ldots, Y_n)$ with size m and n with continuous distribution function $F_X(u)$, and $F_Y(u)$ respectively with assumption of the known median difference $\mu_X - \mu_Y$. Let us assume $\mu_X = \mu_Y = 0$, thus $F_X(u)$, and $F_Y(u)$ have the same form. Same ranks are assigned to the smallest and the largest values after arranging the sample data in ascending order. So the rank scheme is,

$$\begin{cases} 1, 2, \ldots, \frac{m+n}{2}, \frac{m+n}{2}, \ldots, 2, 1 & \text{for} \quad m+n \quad \text{even} \\ 1, 2, \ldots, \frac{m+n-1}{2}, \frac{m+n-1}{2}, \ldots, 2, 1 & \text{for} \quad m+n \quad \text{odd} \end{cases}$$

The test statistics defined by Ansari and Bradley (1960) is,

$$W = \sum_{i}^{n} R(X_i)$$

where the rank sum is associated with the X sample. The standardized test statistics is defined as,

$$U = \frac{W - E(W)}{\sqrt{V[W]}}$$

where

$$E[W] = \begin{cases} \frac{m(m+n+2)}{4} & \text{for} \quad m+n \quad \text{even} \\ \frac{2m(m+n+1)}{4(m+n)} & \text{for} \quad m+n \quad \text{odd} \end{cases}$$

$$V[W] = \begin{cases} \frac{mn(m+n+2)(m+n+2)}{48(m+n-1)} & \text{for} \quad m+n \quad \text{even} \\ \frac{mn(m+n+1)(3+(m+n)^2)}{48(m+n)^2} & \text{for} \quad m+n \quad \text{odd} \end{cases}$$

The chart based on this statistics follows the basic Shewhart scheme. The central line is equal to $CL = 0$, with the upper and lower control limits being $\text{UCL} = 3$ and $\text{LCL} = -3$ respectively. This chart has been proposed by Das (2008).

1.4 Multivariate Control charts

During the last decade, an enormous number of new advanced control charts has been proposed for univariate as well as for multivariate processes. When there are several quality characteristics, multivariate control charts have to be adopted. Montgomery and Wadsworth (1972), Crosier (1988), Pignatiello and Runger (1990), Lowry et al. (1992a), Lowry and Montgomery (1995) and many other contributed in Multivariate control charts. The selection of a suitable control chart depends on the type and distribution of the data. When there are several quality characteristics, multivariate control charts have to be adopted.

1.4.1 Hotelling T^2 control chart

Suppose that the quality characteristics $\mathbf{X}$ follows multivariate normal distribution $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ in $\mathbb{R}^p$ with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$. By definition, the Mahalanobis distance is

$$T^2 = (\bar{\mathbf{X}} - \boldsymbol{\mu})^\intercal \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu}).$$

This distance quantifies the closeness of the point $\mathbf{X}$ to the means $\boldsymbol{\mu}$ by considering the covariances. The T^2 control chart is a multivariate control chart for the process mean developed by Hotelling (1947).

— For groups of observations, the statistic T_i^2 is equal to,

$$T_i^2 = n(\bar{\mathbf{X}} - \boldsymbol{\mu})^\intercal \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu}),$$

The upper control limit UCL is equal to,

$$\text{UCL} = \frac{p(m+1)(n-1)}{m(n-1)-p+1} F_F^{-1}(1-\alpha|p, m(n-1)-p+1),$$

where, $F_F^{-1}(1-\alpha|p, m(n-1)-p+1)$ is the inverse cumulative function of the F distribution with parameters p and $m(n-1)-p+1$ with sample of size m and p number of variables.

— For individual observations ($n = 1$), the statistic T^2 is equal to,

$$T_i^2 = (\mathbf{X} - \boldsymbol{\mu})^\intercal \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}),$$

The statistic T_i^2 follows χ^2 distribution with p degrees of freedom, if the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are known. Therefore, the upper control limit of the Hotelling T^2 chart is given by,

$$\text{UCL} = F_{\chi^2}^{-1}(1-\alpha|p),$$

where α is the type I error and $F_{\chi^2}^{-1}(\ldots|p)$ is the inverse distribution function of the χ^2 distribution with p degrees of freedom.

1.4.2 Multivariate EWMA control chart

Lowry et al. (1992b) proposed a multivariate EWMA control chart (MEWMA) to monitor the mean vectors. Suppose that the quality measure $\mathbf{X}$ follows a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$. The MEWMA statistics is defined as,

$$Q_i = \mathbf{Y}_i{}^\intercal \boldsymbol{\Sigma}_{Y_i}^{-1} \mathbf{Y}_i,$$

with

$$\boldsymbol{\Sigma}_{Y_i} = \frac{r}{2-r} \boldsymbol{\Sigma},$$

and

$$\mathbf{Y}_i = r(\mathbf{X}_i - \boldsymbol{\mu}) + (1-r)\mathbf{Y}_{i-1},$$

where $r \in (0,1)$ (equivalent to λ for the EWMA chart) and $\mathbf{Y}_0 = \mathbf{0}$. The MEWMA control charts is said to be out of control when $Q_i > H$, where $H > 0$ is the specified control limit.

1.4.3 Multivariate CUSUM control chart

Crosier (1988) developed multivariate cumulative sum (MCUSUM) control charts as a multivariate generalization of the univariate CUSUM chart. Suppose a multivariate variable $\mathbf{X}$ follows multivariate normal distribution $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ in $\mathbb{R}^p$ with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$. The MCUSUM control chart is based on statistic,

$$Y_i = (\mathbf{S}_{H,i}{}^\intercal \boldsymbol{\Sigma}^{-1} \mathbf{S}_{H,i})^{1/2}$$

with

$$\mathbf{S}_{H,i} = \begin{cases} \mathbf{0} & \text{if } C_i \leq k, \\ (\bar{\mathbf{X}}_i - \boldsymbol{\mu} + \mathbf{S}_{H,i-1}) & \text{if } C_i > k, \end{cases}$$

where $C_i = (\bar{\mathbf{X}}_i - \boldsymbol{\mu} + \mathbf{S}_{H,i-1})^\intercal \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{X}}_i - \boldsymbol{\mu} + \mathbf{S}_{H,i-1})$. The MCUSUM control chart is said to be out of control when $Y_i > H$ where $H > 0$ is the specified control limit.

1.5 Performance of control charts

There are two objectives for designing control charts: i) the false alarm rate should be minimum in an in-control process and ii) the false alarm rate should be higher in an out-of-control process. Different measures have been proposed to evaluate the performance of control charts. Arorian and Levene (1950) used the Run Length (RL) distribution to study the performance of control chart. The RL is number of points on a chart until a control chart signals. Average run length (ARL) is the average of RL and is commonly used to measure the performance of control charts. When the process is in-control then ARL is denoted by ARL_0 while it is denoted by ARL_1 when the process is out-of-control. To check the performance of a chart ARL_0 is set to be fixed, while ARL_1 is used to detect whether the chart is effective or not. The smaller the value of ARL_1 the more effective is the chart. When the process is in control the ARL is defined as,

$$\mathrm{ARL}_0 = \frac{1}{\alpha}$$

where α is the Type I error and when the process is out of control the ARL is defined as,

$$\mathrm{ARL}_1 = \frac{1}{1-\beta}$$

where β is the Type II error. When we have to estimate process parameters EARL have been considered to study the performance of control charts. In most of the cases the distribution of RL is skewed so the Median Run Length MRL can be preferred. As the mean and the median are measures of central tendency, to evaluate the dispersion of the Run Length, hereafter referred to as SDRL is evaluated as,

$$\mathrm{SDRL} = \frac{\sqrt{\beta}}{1-\beta}.$$

In this chapter we have presented control charts (in general) along with their performance measures. In the next chapter we will introduce the compositional data along with the methods to monitor these kind of data. Also we will discuss the past papers which deal with control charts on compositional data.

2

Compositional Data - Definition and Monitoring

This chapter is divided into three sections: in first section we have the introduction of compositional data along with the Aitchison's geometry and in the second section we discuss the transformations that can be made to transform compositional data and in the third section we have included some previous studies about control charts on compositional data.

2.1 Introduction to Compositional data

In statistics, while dealing with the case of continuous multivariate processes, the vast majority of statistical methods assumes that the data are *unconstrained*. But there is a specific category of multivariate data which are *constrained* by definition. This kind of data is called CoDa (for Compositional Data) is represented by vectors whose *strictly positive* components only convey relative information. CoDa includes measurements in probability, proportions, percentages and parts. Usually the sum of the components of CoDa vector is expressed as some constant κ being equal to 1 if we are working with proportions, 100 if we are working with percentages, 10^6 if we are working with parts per million (ppm), and so on. If we apply statistical methods for unconstrained data on CoDa then we get inappropriate inference. As we will see, the sample space of CoDa is not the same as the Euclidean space for unconstrained data.

Pearson (1897) identified the problem of interpreting correlation between the variables x, y, z that are uncorrelated then we cannot make an inference about correlation between $x/y, y/z$ or z/x, because there is a possibility of correlation between the ratio of variables. This is known as spurious correlation which was further adjusted by Pearson through scaling. But, the scaling ignores a range of data as it is useful only if the scaling variable is either strictly positive or strictly negative. Tanner (1949) suggested a log transformation and to check the normality of the transformed and original values to know whether the transformation is advantageous or not. Later on Chayes (1960) showed that while working with CoDa some of the correlation must be negative because of the constraint that the sum must be equal to unity. But he does not find any model to remove the effect of unity constraint because according to him multivariate tests for unconstrained data should not be used for constrained data.

After the effort of many scientists the first rigorous mathematical definition for CoDa analysis was given by Aitchison (1986). He developed an "adequate" geometry to model and transform such data. It provides only relative information so, we can say that the composition is scale invariant, i.e. if it is multiplied by a positive number it remains the same as it was before.

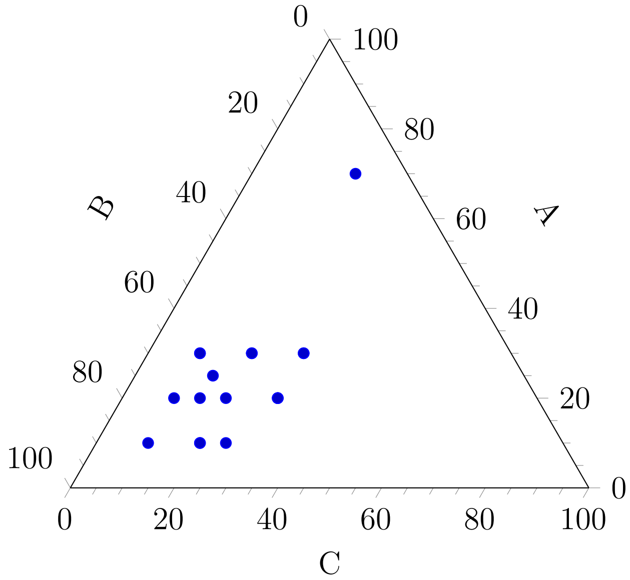

Figure 2.1 – Graphical representation of compositional data

A simple example of CoDa is the composition of chocolate. Talbot (2012) conducted a study on the composition of chocolate using different amount of all the ingredients to make different types of chocolates. The typical plain chocolate is composed of 32.5 % cocoa powder, 15.5% cocoa butter, 51.6% sugar and 0.4% lecithin, hence this composition can be written as [32.5,15.5,51.6,0.4]. A chocolate with low fat contains of 20 % cocoa powder, 18.25% cocoa butter, 15% full cream milk powder, 46.3% sugar and 0.4% lecithin, i.e. [20,18.25,15,46.3,0.4]. While, a high fat milk chocolate is composed of 20 % cocoa powder, 15.75% cocoa butter, 25% full cream milk powder, 38.85% sugar and 0.4% lecithin, i.e. [20,15.75,25,38.85,0.4].

Another example of CoDa is the composition of the earth crust studied by Lutgens and Tarbuck (2000). The earth crust is composed of 46.6% oxygen, 27.7% silicon, 8.1% aluminum, 5% iron, 3.6% potassium and 2.1% magnesium, i.e. [46.6,27.7,8.1,5,3.6,2.8, 2.6,2.1].

2.1.1 Definition

The first mathematical definition for CoDa analysis, given by Aitchison (1986), is as follows,
A p-part *composition* is defined as a *row vector* $\mathbf{x} = (x_1, \dots, x_p)$ that belongs to the simplex $\mathcal{S}^p$ defined as,

$$\mathcal{S}^p = \left\{ \mathbf{x} = (x_1, x_2, \dots, x_p) | x_i > 0, i = 1, 2, \dots, p \text{ and } \sum_{i=1}^{p} x_i = \kappa \right\},$$

where $\kappa > 0$ is a constant (for example, $\kappa = 1$ if components $x_1, x_2, \dots, x_p$ are proportions and $\kappa = 100$ if they are percentages). Important remark: In order to avoid any misunderstanding, all the vectors in this paper are considered as *row* vectors, not as column vectors. By definition, two compositions $\mathbf{x}$ and $\mathbf{y}$ can be different (for instance $\mathbf{x} = (0.2, 0.5, 0.3)$ and $\mathbf{y} = (20, 50, 30)$) but they are *compositionally equivalent* (i.e. they carry the same relative information) if $\mathcal{C}(\mathbf{x}) = \mathcal{C}(\mathbf{y})$, where the *closure* function is defined as,

$$\mathcal{C}(\mathbf{x}) = \left(\frac{\kappa x_1}{\sum_{i=1}^{p} x_i}, \frac{\kappa x_2}{\sum_{i=1}^{p} x_i}, \dots, \frac{\kappa x_p}{\sum_{i=1}^{p} x_i} \right).$$

2.1.2 Principles of Compositional Data Analysis

There are three basic principles that should be kept in mind while dealing with CoDa analysis. These are as follows,

a Scale invariance
b Permutation invariance
c Subcompositional coherence

a. Scale Invariance

If we have two compositions $\mathbf{x} = (0.2, 0.5, 0.3)$ and $\mathbf{y} = (20, 50, 30)$, it is clearly visible that $\mathbf{x} \neq \mathbf{y}$. But, compositionally they are equivalent because they carry the same relative information and can be written as $\mathcal{C}(\mathbf{x}) = \mathcal{C}(\mathbf{y})$.

This characteristic is also applicable on more than two compositions, as if we have compositions of grocery items like lentils, cereals, dairy and fruits etc. whatever will be the unit of measurements of the compositions like g, kg, L, or ounces etc. and it can be 20 grams, 50 grams and 5 liters of the product in a composition, it does not affect the results of the composition. Similarly the proportion of diamond, heart, spade and club cards in a poker hand and bridge can be compared with the help of closure property. If a 3 dimensional positive real sample space $\mathbb{R}^3_+$ have five different compositions such that $a = [4, 2, 9], b = [12, 6, 27], c = [40, 20, 90], d = [4/13, 2/13, 9/13]$ and $e = [8/17, 4/17, 18/17]$, all these results gives same relative information because the ratios between their components are same.

b. Permutation Invariance

A composition is said to be permutation invariant if it gives the same results after changing the order of the parts of the composition. For example if the sandstone grain is composed of only quartz (Q), feldspar (F) and rock fragments (R). If the order of the composition is changed from [Q, F, R] to [F, R, Q] or [R, F, Q], the results will remain the same. While studying the size of grains and its relationship with sediment, it is not very important to have all the components in a fix order, as the order does not affect the percentages of each component in the composition.

c. Subcompositional Coherence

In multivariate data analysis, marginal is used to represent a subspace of the data using fewer values to make analysis easier. Using the same strategy, subcomposition in CoDa analysis are used to make the analysis easier instead of dealing with all data. Subcompositional Coherence have some consequences, some of them are discussed below,
First is known as subcompositional dominance, i.e. if the distance between two subcompositions is measured it has to be less than the distance between two full compositions. The classical Euclidean distance formula is not useful for CoDa, because it does not fulfill the condition of subcompositional coherence as it does not gives larger values for distance between two compositions than the value of the distances between two subcompositions.
Secondly, the value of the dispersion of the subcomposition has to be lower than the value of dispersion of the full composition.
Finally, the results of the subcomposition should remain unchanged after removing any non-informative part from the composition and the result must remain the same if in a p-part composition another non-informative part is included. The results will be same for p-part composition as for $p - 1$ part composition and also for $p + 1$ part composition.

2.1.3 Aitchison Geometry

The standard Euclidean geometry that defines a vector space with a metric structure in $\mathbb{R}^p$ (unconstrained space) cannot be used for compositions in $\mathcal{S}^p$ (constrained space). For instance, if $\mathbf{x} = (0.2, 0.5, 0.3) \in \mathcal{S}^p$ and $\mathbf{y} = (0.3, 0.6, 0.1) \in \mathcal{S}^p$ then $\mathbf{x} + \mathbf{y} = (0.5, 1.1, 0.4) \notin \mathcal{S}^p$ and $2.\mathbf{x} = (0.4, 1, 0.6) \notin \mathcal{S}^p$ (i.e. the traditional operators $+$ and $.$ are useless). Consequently, if we want a vector space with a metric structure for $\mathcal{S}^p$, there is a need to define a geometry. This specific geometry, proposed by Aitchison (2011) (and called the Aitchison's geometry) defines two new operators.

i. The perturbation operator $\oplus$ of $\mathbf{x} \in \mathcal{S}^p$ by $\mathbf{y} \in \mathcal{S}^p$ (analogous to the translation in $\mathbb{R}^p$) defined as,

$$\mathbf{x} \oplus \mathbf{y} = \mathcal{C}(x_1y_1, x_2y_2, \ldots, x_py_p),$$

ii. The powering operator $\odot$ of $\mathbf{x} \in \mathcal{S}^p$ by a constant $a \in \mathbb{R}$ (analogous to the scalar multiplication in $\mathbb{R}^p$ by a constant) defined as,

$$a \odot \mathbf{x} = \mathcal{C}(x_1^a, x_2^a, \ldots, x_p^a).$$

Being a vector space, the simplex with perturbation and powering property should fulfill the following properties.

a. Commutative property of addition

$$\mathbf{x} \oplus \mathbf{y} = \mathbf{y} \oplus \mathbf{x} \quad \text{where} \quad \mathbf{x}, \mathbf{y} \in \mathcal{S}^p$$

b. Associative property of addition

$$(\mathbf{x} \oplus \mathbf{y}) \oplus \mathbf{z} = \mathbf{x} \oplus (\mathbf{y} \oplus \mathbf{z}) \quad \text{where} \quad \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{S}^p$$

c. Neutral element for $\oplus$

$$\mathcal{C}(1, 1, \ldots, 1) = (\frac{1}{p}, \ldots, \frac{1}{p}) = \mathbf{0}_{\mathcal{S}^p}$$

d. Additive inverse of the composition $\mathbf{x}$

$$\mathbf{x}^{-1} = \mathcal{C}(x_1^{-1}, x_2^{-1}, \ldots, x_p^{-1})$$

e. Associative property of multiplication

$$\begin{array}{rcl} \alpha \odot (\beta \odot \mathbf{x}) & = & (\alpha.\beta) \odot \mathbf{x} \\ \text{where} \quad \alpha, \beta \in \mathbb{R} & \text{and} & \mathbf{x}, \mathbf{y} \in \mathcal{S}^p \end{array}$$

f. Distributive property of multiplication over addition

$$\alpha \odot (\mathbf{x} \oplus \mathbf{y}) = (\alpha \odot \mathbf{x}) \oplus (\alpha \odot \mathbf{y})$$

$$(\alpha + \beta) \odot \mathbf{x} = (\alpha \odot \mathbf{x}) \oplus (\beta \odot \mathbf{x})$$

g. Neutral element for $\odot$

$$1 \odot \mathbf{x} = \mathbf{x}$$

2.1.4 Difference between the Euclidean and Aitchison Geometries

a. Inner product

Aitchison's inner product is defined as,

$$\langle \mathbf{x}, \mathbf{y} \rangle_A = \frac{1}{2D} \sum_{i=1}^{D} \sum_{i=1}^{D} \ln \frac{x_i}{x_j} \ln \frac{y_i}{y_j}.$$

where subindex A stands for Aitchison and $\langle \cdot\,, \cdot \rangle$ is the inner product in $\mathbb{R}^p$. $\mathbf{x}$ and $\mathbf{y}$ are compositionally orthogonal if $\langle \mathbf{x}, \mathbf{y} \rangle_A = 0$.

b. Norm

Aitchison's norm is defined as,

$$||\mathbf{x}||_A = \sqrt{\frac{1}{2D}\sum_{i=1}^{D}\sum_{i=1}^{D}\left(\ln\frac{x_i}{x_j}\right)^2}.$$

where $||\cdot||_2$ is the L^2-norm in $\mathbb{R}^p$.

c. Distance

Aitchison's distance is defined as,

$$d_A(\mathbf{x},\mathbf{y}) = ||\mathbf{x}\ominus\mathbf{y}||_a = \sqrt{\frac{1}{2D}\sum_{i=1}^{D}\sum_{i=1}^{D}\left(\ln\frac{x_i}{x_j} - \ln\frac{y_i}{y_j}\right)^2}.$$

2.2 Transformation of Compositional data

CoDa are always in form of part of some whole with the constraint of a constant sum. It is not possible to apply standard statistical procedures on these data because of the constraint. To remove the constraint of constant sum, the data should be transformed using some transformation. Firstly, Aitchison (1986) gives a transformation to deal with CoDa. As the CoDa does not gives us the exact values and only gives us the relative information, then the ratios between the elements of the composition have to be used.

There are three main transformations used to transform CoDa.

i. Centered log-ratio (clr) transformation
ii. Additive log-ratio (alr) transformation
iii. Isometric log-ratio (ilr) transformation

2.2.1 Centered log-ratio transformation

The centered log-ratio (clr) transformation is defined by Aitchison (1986) is defined as,

$$\mathrm{clr}(\mathbf{x}) = \left(\ln\frac{x_1}{\bar{x}_G}, \ln\frac{x_2}{\bar{x}_G}, \ldots, \ln\frac{x_p}{\bar{x}_G}\right),$$

where $\bar{x}_G$ is the componentwise geometric mean of $\mathbf{x}$, i.e.

$$\bar{x}_G = \left(\prod_{i=1}^{p} x_i\right)^{\frac{1}{p}} = \exp\left(\frac{1}{p}\sum_{i=1}^{p}\ln x_i\right).$$

This formula transforms a composition $\mathbf{x}\in\mathcal{S}^p$ into a vector $\mathrm{clr}(\mathbf{x}) = \boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_p) \in \mathbb{R}^p$ satisfying the constraint $\xi_1 + \xi_2 + \cdots + \xi_p = 0$.

2.2.2 Additive log-ratio transformation

The additive log-ratio (alr) transformation was defined by Billheimer et al. (2001), who developed the algebra for compositions and the proof of the Hilbert space structure of the simplex in terms of this transformation. The additive log-ratio alr transformation for a composition in the simplex sample space can be written as,

$$\mathrm{alr}(\mathbf{x}) = \left(\ln \frac{x_1}{x_D}, \ln \frac{x_2}{x_D}, \ldots, \ln \frac{x_{D-1}}{x_D} \right) = \zeta$$

where $x = [x_1, x_2, \ldots, x_D]$ is a composition in simplex sample space and x_D can be any arbitrary value from the composition x.

To find the value of x from the above transformation, the inverse alr transformation is used that is,

$$\mathbf{x} = \mathrm{alr}^{-1}(\zeta) = C[\exp(\zeta_1), \exp(\zeta_2), \ldots, \exp(\zeta_{D-1}), 1]$$

where

$$\zeta = [\zeta_1, \zeta_2, \ldots, \zeta_{D-1}]$$

The alr provides different results for different value of x_D. The alr transformation is also known as the asymmetric transformation.

2.2.3 Isometric log-ratio transformation

The isometric log-ratio (ilr) transformation for a composition $\mathbf{x} \in \mathcal{S}^p$ is defined as,

$$\mathrm{ilr}(\mathbf{x}) = \mathbf{x}^* = \mathrm{clr}(\mathbf{x})\mathbf{B}^\intercal$$

where $\mathbf{B}$ is a $(p-1, p)$ matrix. There are many possible choices for $\mathbf{B}$ (for more explanations, see for instance, Pawlowsky-Glahn et al. (2015, page 40)), one of them is given below:

$$B_{i,j} = \begin{cases} \sqrt{\frac{1}{(p-i)(p-i+1)}} & j \leq p-i \\ -\sqrt{\frac{p-i}{p-i+1}} & j = p-i+1 \\ 0 & j > p-i+1 \end{cases}$$

The ilr transformation is fundamental because it is an isometry that allows a unique transformation of the (constrained) coordinates of a composition $\mathbf{x}$ into the (unconstrained) ilr-coordinates $\mathbf{x}^* \in \mathbb{R}^{p-1}$. Conversely, if we want to obtain the composition coordinates $\mathbf{x}$ from the ilr-coordinates $\mathbf{x}^*$ we have to use the *inverse isometric log-ratio* transformation defined as,

$$\mathrm{ilr}^{-1}(\mathbf{x}^*) = \mathbf{x} = \mathcal{C}(\exp(\mathbf{x}^*\mathbf{B})).$$

There are two ways to deal with CoDa, one is to work with the Aitchison geometry on the simplex and use the power and perturbation operator and the second is to transform the data into the real space by means of log-ratio coordinates, then apply classical statistics and use backward transformation to the simplex for interpretation, if necessary. Finally, by definition (see Pawlowsky-Glahn et al. (2015, page 114)), $\mathbf{x}$ is a *multivariate normal random composition* on the simplex $\mathcal{S}^p$, denoted as $\mathrm{MNOR}_{\mathcal{S}^p}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$, if $\mathbf{x}^* = \mathrm{ilr}(\mathbf{x}) \sim \mathrm{MNOR}_{\mathbb{R}^{p-1}}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$, i.e. $\mathbf{x}^*$ is a multivariate normal random vector on $\mathbb{R}^{p-1}$, where $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}^*$ are the $(1, p-1)$ mean (row) vector and $(p-1, p-1)$ variance-covariance matrix, respectively.

2.3 Control Charts for Compositional Data

In statistical process control literature one can find few contributions of control charts devoted to CoDa. Some of them are listed in the following sections of this chapter.

2.3.1 Chi-square statistic to monitor compositional process data

Boyles (1997) attempted to propose a control chart for CoDa for the first time. He developed a Chi-square control chart to monitor CoDa. This chart has been found to be useful for multinomial data, as the properties of the Dirichlet distribution are very restricted and, sometimes, CoDa arise from a Dirichlet distribution so he found this chart also valid for CoDa. He used the sample means to estimate the process averages $\pi_1, \pi_2, \ldots, \pi_k$ and he determined the statistic X^2 as,

$$X^2 = \frac{(u_1-\pi_1)^2}{\pi_1} + \frac{(u_2-\pi_2)^2}{\pi_2} + \cdots + \frac{(u_k-\pi_k)^2}{\pi_k} \quad \text{where} \quad X^2 \sim \chi^2_{k-1}$$

where X^2 follows a χ^2 distribution with $k-1$ degrees of freedom. This chart have been compared with the T^2 control chart based on log-ratio transformation using the last component of CoDa. To check the performance of the X^2 chart and to compare it with the T^2 chart, Boyles (1997) plotted both charts. He founds that there are few out-of-control points that are detected by the X^2 chart but not by the T^2 chart but also there are some points that are detected by the T^2 chart but not by X^2 chart. He used some examples to show the performance of the X^2 chart and he found it more sensitive in one example but less sensitive in the another one. The main advantage of this study is its simplicity.

2.3.2 T^2 control chart for a $p = 3$ part composition

Vives-Mestres et al. (2014a) studied the out-of-control signal for three part CoDa. Let the row vector $\mathbf{x} = (x_1, \ldots, x_p)$ belong to the simplex $\mathcal{S}^p$ and $\mathbf{z}_i = (z_1, z_2, \ldots, z_{p-1})$ be the corresponding coordinates obtained as $z_i = ilr(\mathbf{x})$ in $\mathbb{R}^{p-1}$ with $\boldsymbol{\mu}_z$ the mean vector and $\boldsymbol{\Sigma}_z$ the variance covariance matrix of $\mathbf{z}_i$. Then the T^2 control chart for $p = 3$ part compostional data is defined as

$$T_c^2 = (\mathbf{z} - \boldsymbol{\mu}_z)^T \boldsymbol{\Sigma}_z^{-1} (\mathbf{z} - \boldsymbol{\mu}_z).$$

They found this chart better than the T^2 chart as it gives a better fit is constructed on z_i that, differently from $\mathbf{x}_i$, tends to be normally distributed, and the control region back transformed into $\mathcal{S}^p$ through ilr is consistent with the compositional (restricted) sample space. It also fulfills the sub-compositional coherence property.

2.3.3 Control chart for individual observations for compositional data

As an extension of the T^2 control chart for a $p = 3$ part composition, Vives-Mestres et al. (2014b) proposed a control chart for individual observations for compositionl data by deleting one observation from the raw CoDa. Then the CoDa T^2 statistic T_c^2 is defined as,

$$(T_c^2)_t = (\mathbf{z}_t - \boldsymbol{\mu}_z)^T \boldsymbol{\Sigma}_z^{-1} (\mathbf{z}_t - \boldsymbol{\mu}_z).$$

If the values of the mean $\boldsymbol{\mu}_z$ and variance $\boldsymbol{\Sigma}_z$ are known then $T_C^2 \sim \chi^2(p-1)$ and the upper control limit will be,

$$\text{UCL} = F^{-1}_{\chi^2}(\alpha|p-1)$$

where α is type I error. If the values of parameters $\boldsymbol{\mu}_z$ and $\boldsymbol{\Sigma}_z$ is not known then the phase I control limits are

$$\text{UCL} = \frac{(m-1)^2}{m} F^{-1}_{\beta}(\alpha|(p-1)/2,(m-p-2)/2)$$

where m is the sample size and $(p-1)$ is the number of variables. F^{-1}_{β} is the inverse distribution function of the Beta distribution with parameters $(p-1)/2$ and $(m-p-2)/2$. The typical T^2 control chart after deleting one variable is not consistent while dealing with CoDa. This method fails when the data sets have some specific shapes. Vives-Mestres et al. (2014b) showed that when the samples are close to a vertex the T^2_C chart has a better performance than the T^2 control charts in terms of in-control ARL. But, when the samples are homogeneous both T^2 and T^2_C gave good performance in terms of in-control ARL.

2.3.4 Phase II MEWMA control chart for compositional data

? monitored CoDa using a Multivariate Exponentially Weighted Moving Average Scheme.

Let $\mathbf{x}_{i,1},\ldots,\mathbf{x}_{i,n}$ be a p-part compositions at specific sampling periods $i = 1,2,\ldots$ and $\mathbf{x}^*_{i,j}$ be the ilr coordinates of $\mathbf{x}_{i,j}$ which are supposed to follow a multivariate normal distribution with mean $\boldsymbol{\mu}^*$ and variance $\boldsymbol{\Sigma}^*$. As $\boldsymbol{\mu}^* = \text{ilr}(\boldsymbol{\mu})$ and $\boldsymbol{\mu}$ is the center of composition, then

$$\bar{\mathbf{x}}_i = \frac{1}{n} \odot (\mathbf{x}_{i,1} \oplus \ldots \oplus \mathbf{x}_{i,n})$$

and

$$\bar{\mathbf{x}}^*_i = \text{ilr}(\bar{\mathbf{x}}_i).$$

Then according to Lowry et al. (1992a),

$$Q_i = \mathbf{y}_i^T \boldsymbol{\Sigma}^{-1}_{Y_i} \mathbf{y}_i$$

where

$$\boldsymbol{\Sigma}_{Y_i} = \frac{r}{2-r}\boldsymbol{\Sigma}$$

The control chart is said to be out of control when $Q_i > H$, where $\text{UCL} = H > 0$. They used a Markov chain method to evaluate the performance of the proposed control chart. They defined the out-of-control ARL of MEWMA-CoDa control chart as,

$$\text{ARL} = \mathbf{s}^{\intercal}(\mathbf{I}-\mathbf{P})^{-1}\mathbf{1}$$

where

$$\mathbf{P} = \mathbf{T}(\alpha,\beta) \circledast \mathbf{P}_2$$

where $\circledast$ is the element wise multiplication of matrices and $\mathbf{T}$ is the $(2m1+1, m2+1)$ matrix defined as,

$$\mathbf{T}(\alpha, \beta) = \begin{cases} 1 & \text{if state } (\alpha, \beta) \text{ is transient} \\ 0 & \text{otherwise} \end{cases}$$

and $\mathbf{s}$ is the $m_1 + 1$ initial probability vector with first element equal to one and zero otherwise. When the number of states is large, the approximation of the ARL is more accurate. As using a large number of states needs more computing resources and time, they selected $m_1 = m_2 = 30$. After selecting a particular value for the in-control $\mathrm{ARL}_0 = 370$, the values of the optimal couples (r, H) have been calculated when there is no shift. Using a particular value of the shift for the mean, they select the optimal couple (r^*, H^*) from the set of designed couples such that the value of the out-of-control ARL is minimum. They found the MEWMA-CoDa chart to be more sensitive than T_C^2 control chart.

In this chapter we discussed the CoDa and their transformations. Also we have presented the previous efforts on development of control charts using CoDa. In the next chapter we will discuss the control charts on CoDa in the presence of measurement error.

II

Contributions

3

Performance of the Hotelling T^2-CoDa Control Chart in the Presence of Measurement Errors

In quality control application, there are two sources of variation: the first one is from the production process and the second one is due to the measuring devices. The difference between the actual values and the observed ones is known as *measurement errors*. Many researchers have already worked on different control charts in the presence of measurement errors. Different types of measurement error models are used by the researcher, such as additive model, multiplicative model, two component model and four component model. Most of the researchers used additive covariate model $y = a+bx+\varepsilon$, where a and b are two constant and ε is the error term that follows normal distribution with zero mean and a given variance. There are also different variance behaviour assumptions for measurement error such as constant, increasing or constant and linearly increasing. Most of the researchers assumed constant variance of measurement error. Bennett (1954) was one of the first to work on the Shewhart $\bar{x}$ chart in the presence of measurement errors. He used the simple model $y = x+\varepsilon$, where ε is the measurement error term. Linna and Woodall (2001) investigated the effect of measurement errors on the Shewhart $\bar{x}$ chart using the additive covariate model with constant error variance. Yang et al. (2007) used additive covariate model with constant error variance to study the performance of EWMA $\bar{x}$ chart in the presence of measurement errors. Hu et al. (2016c) also used additive covariate model with constant error variance to study economic design of the upper-sided synthetic chart with measurement errors. Further Tran et al. (2016) studied the performance of the Shewhart-RZ control chart in the presence of measurement error using the additive covariate model with constant error variance. Hu et al. (2016a) also used the additive model with constant variance of measurement error to study the performance of the variable sampling interval $\bar{x}$ charts in the presence of measurement errors.

There are a few researchers who used linearly increasing variance of measurement error such as Dizabadi et al. (2016) who studied the effect of measurement error with linearly increasing variance on simultaneous monitoring of process mean and variability under additive covariate model. Maleki et al. (2016) also worked on simultaneous monitoring of multivariate process mean and variability in the presence of measurement error with linearly increasing variance under additive covariate model. Tang et al. (2018) also used the additive model with linearly increasing error variance to study the effect of measurement errors on the adaptive EWMA $\bar{x}$ charts.

There are also many contributions where the additive covariate model is used with constant and linearly increasing variance of measurement error. Linna et al. (2001) studied the multivariate χ^2 control chart in the presence of measurement errors using the additive model with constant and linearly increasing error variance. Hu et al. (2016b) studied the performance of the variable sample size $\bar{x}$ charts using the additive model with constant and linearly increasing error variance. Cheng and Wang (2018a) also used the additive model with constant and linearly increasing error variance to analyze the performance of EWMA median and CUSUM median control charts for a normal process. Further, Cheng and Wang (2018b) studied VSSI median control chart with estimated parameters using the additive model with constant and linearly increasing error variance. For a detailed literature review on measurement errors in statistical process monitoring, see Maleki et al. (2017).

Even though many researches have been conducted on the evaluation of control charts in the presence of measurement errors, few of them have been devoted to multivariate data and, as far as we know, none of them have been devoted to CoDa. The goal of this chapter is to fill this gap and, therefore, to study the effect of measurement errors on the T^2_C control chart proposed by Vives-Mestres et al. (2014b). The remainder of this chapter is organized as follows: in Section 3.1, the linearly covariate measurement error model for CoDa is introduced. Section 3.2 details the Hotelling CoDa T^2 control chart in the presence of measurement errors and Section 3.3 investigates the performance of this control chart. Finally, a very detailed illustrative example is provided in Section 3.4 and conclusions and future research directions are presented in Section 3.5.

3.1 Linearly covariate measurement error model for CoDa

Let us assume that, at time $i = 1, 2, \ldots$, the quality characteristic is a p-part composition $\mathbf{y}_i = (y_{i,1}, \ldots, y_{i,p}) \in \mathcal{S}^p$ and $\mathbf{y}_1, \mathbf{y}_2, \ldots$ are independent $\mathrm{MNOR}_{\mathcal{S}^p}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$ random compositions with $p = d+1$. Moreover, let us also assume that the quality characteristic $\mathbf{y}_i$ is not directly observable, but can only be assessed from the results $\mathbf{x}_{i,1}, \ldots, \mathbf{x}_{i,m}$ of a set of $m \geq 1$ *measurement operations* with each $\mathbf{x}_{i,j}$, $j = 1, \ldots, m$, being equal to (linearly covariate measurement error model):

$$\mathbf{x}_{i,j} = \mathbf{a} \oplus (b \odot \mathbf{y}_i) \oplus \varepsilon_{i,j}, \tag{3.1}$$

where $\mathbf{a} \in \mathcal{S}^p$ and $b \in \mathbb{R}$ are known (perturbation and powering) constants and where $\varepsilon_{i,j}$ is a multivariate normal $\mathrm{MNOR}_{\mathcal{S}^p}(\mathbf{0}, \boldsymbol{\Sigma}^*_{\mathrm{M}})$ random error term due to the measurement inaccuracy, which is independent of $\mathbf{y}_i$ and $\boldsymbol{\Sigma}^*_{\mathrm{M}}$ is the unknown measurement error variance-covariance matrix.

This measurement error model for CoDa is inspired by the one proposed by Linna and Woodall (2001) and Linna et al. (2001) for "classical" univariate and multivariate data, respectively. In these papers, these authors advocate taking multiple (m) measurements and for averaging them in order to compensate for the effect of measurement errors and to decrease the variance of the measurement error component. Of course, the quality practitioner will have to find the right balance between the extra costs and time associated with too many measurements and an acceptable level of measurement error. Consequently, similar to Linna and Woodall (2001) and Linna et al. (2001), we suggest to define, at time $i = 1, 2, \ldots$, the composition sample mean as (see (Pawlowsky-Glahn et al., 2015, page 132)):

$$\begin{aligned} \bar{\mathbf{x}}_i &= \frac{1}{m} \odot (\mathbf{x}_{i,1} \oplus \cdots \oplus \mathbf{x}_{i,m}), \\ &= \mathbf{a} \oplus (b \odot \mathbf{y}_i) \oplus \left(\frac{1}{m} \odot (\varepsilon_{i,1} \oplus \cdots \oplus \varepsilon_{i,m}) \right). \end{aligned}$$

Let us define $\mathbf{a}^* = \mathrm{ilr}(\mathbf{a})$. Using theorem 6.20 in (Pawlowsky-Glahn et al., 2015, page 117), we have $\bar{\mathbf{x}}_i \sim \mathrm{MNOR}_{\mathcal{S}^p}(\boldsymbol{\mu}^*_{\bar{\mathbf{x}}}, \boldsymbol{\Sigma}^*_{\bar{\mathbf{x}}})$ with

$$\boldsymbol{\mu}^*_{\bar{\mathbf{x}}} = \mathbf{a}^* + b\boldsymbol{\mu}^*, \tag{3.2}$$

$$\boldsymbol{\Sigma}^*_{\bar{\mathbf{x}}} = b^2\boldsymbol{\Sigma}^* + \frac{1}{m}\boldsymbol{\Sigma}^*_{\mathrm{M}}. \tag{3.3}$$

3.2 Hotelling CoDa T^2 control chart in the presence of measurement errors

Let us assume that at time $i = 1, 2, \ldots$, we have m measures $\mathbf{x}_{i,1}, \ldots, \mathbf{x}_{i,m}$ of the quality characteristic $\mathbf{y}_i$. When the process is in-control, we have $\mathbf{y}_i \sim \mathrm{MNOR}_{\mathcal{S}^p}(\boldsymbol{\mu}_0^*, \boldsymbol{\Sigma}^*)$ and, when the process is out-of-control, we have $\mathbf{y}_i \sim \mathrm{MNOR}_{\mathcal{S}^p}(\boldsymbol{\mu}_1^*, \boldsymbol{\Sigma}^*)$, where $\boldsymbol{\mu}_0^*$ and $\boldsymbol{\mu}_1^*$ are the in- and out-of-control mean vectors, respectively ($\boldsymbol{\Sigma}^*$ is supposed to be unchanged). Let $\bar{\mathbf{x}}_i^* = \mathrm{ilr}(\bar{\mathbf{x}}_i)$. Using a similar approach as in Vives-Mestres et al. (2014b), a Hotelling T^2 control chart for CoDa, which takes into account of measurement errors, can be proposed. According to the results obtained at the end of the previous section, the statistic monitored by the Hotelling CoDa T^2 control chart with measurement errors is (remember that the vectors are considered as row vectors)

$$Z_i = (\bar{\mathbf{x}}_i^* - \mathbf{a}^* - b\boldsymbol{\mu}_0^*)\left(b^2\boldsymbol{\Sigma}^* + \frac{1}{m}\boldsymbol{\Sigma}_{\mathrm{M}}^*\right)^{-1}(\bar{\mathbf{x}}_i^* - \mathbf{a}^* - b\boldsymbol{\mu}_0^*)^{\intercal}. \tag{3.4}$$

When the process is in-control, we have $Z_i \sim \chi^2(p-1)$, i.e. a χ^2 distribution with $p-1$ degrees of freedom. As a consequence, an upper control limit UCL can be defined for this Hotelling T^2 control chart as

$$\mathrm{UCL} = F_{\chi^2}^{-1}(1-\alpha_0|p-1),$$

where $F_{\chi^2}^{-1}(\ldots|p-1)$ is the inverse c.d.f. (cumulative distribution function) of the $\chi^2(p-1)$ distribution and $\alpha_0 = \frac{1}{\mathrm{ARL}_0}$ is the Type I error, ARL_0 being the in-control ARL. When the process is out-of-control, the statistic $Z_i \sim \chi^2(p-1, \delta_{\mathrm{M}})$, i.e. a non-central χ^2 distribution with $p-1$ degrees of freedom and non-centrality parameter δ_{M} equal to

$$\delta_{\mathrm{M}} = b^2(\boldsymbol{\mu}_1^* - \boldsymbol{\mu}_0^*)\left(b^2\boldsymbol{\Sigma}^* + \frac{1}{m}\boldsymbol{\Sigma}_{\mathrm{M}}^*\right)^{-1}(\boldsymbol{\mu}_1^* - \boldsymbol{\mu}_0^*)^{\intercal}.$$

It should be noted that δ_{M} depends on b but it does not depend on $\mathbf{a}^*$. A particular value of δ_{M} is $\delta = (\boldsymbol{\mu}_1^* - \boldsymbol{\mu}_0^*)(\boldsymbol{\Sigma}^*)^{-1}(\boldsymbol{\mu}_1^* - \boldsymbol{\mu}_0^*)^{\intercal}$ corresponding to the "without measurement error" case, i.e. when $\boldsymbol{\Sigma}_{\mathrm{M}}^* = \mathbf{0}$ and $b = 1$

The Type II error of the Hotelling CoDa T^2 control chart with measurement errors is $\beta = F_{\chi^2}(\mathrm{UCL}|p-1, \delta_{\mathrm{M}})$ and the corresponding out-of-control ARL is $\mathrm{ARL}_1 = \frac{1}{1-\beta}$. Ideally we want a high ARL_0 when the process is in control and a low ARL_1 when the process is out of control.

As pointed out in Linna et al. (2001), multivariate control charts based on covariates (i.e. Equation 3.1 on the facing page) do not generally possess the property of directional invariance to shifts in the mean vector (here $\bar{\mathbf{x}}_i^*$) and, in the presence of measurement errors, these multivariate control charts are therefore, more powerful at detecting shifts in some directions than in others. In order to overcome this problem, Linna et al. (2001) suggested to compute $\delta_{\min}$ and $\delta_{\max}$ as the minimum and maximum values of δ_{M} corresponding to a fixed value of δ.

Let $\mathbf{G}$ and $\mathbf{H} > \mathbf{0}$ be a symmetric and a positive definite (p, p) matrix, respectively. Let $\mathbf{x}_{\min}$ be the row $(1, p)$ vector satisfying

$$\mathbf{x}_{\min} = \underset{\mathbf{x}}{\mathrm{argmin}}(\mathbf{x}\mathbf{G}\mathbf{x}^{\intercal}),$$

subject to the constraint

$$\mathbf{x}_{\min}\mathbf{H}\mathbf{x}_{\min}^{\intercal} = c > 0.$$

Based on Theorem A.9.2 in Mardia et al. (1979)(page 479), it can be easily proven that if $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_p$ and $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ are the ordered eigenvalues and the corresponding eigenvectors of $\mathbf{H}^{-1}\mathbf{G}$, respectively, then

$$\mathbf{x}_{\min} = \mathbf{v}_1 \sqrt{\frac{c}{\mathbf{v}_1 \mathbf{H} \mathbf{v}_1^\intercal}},$$

and

$$\min_{\mathbf{x}}(\mathbf{x}\mathbf{G}\mathbf{x}^\intercal) = \mathbf{x}_{\min}\mathbf{G}\mathbf{x}_{\min}^\intercal = c\lambda_1.$$

Mardia et al. (1979)(page 479), $c = 1$. Now, if $\mathbf{x}_{\max}$ is the row vector satisfying

$$\mathbf{x}_{\max} = \operatorname*{argmax}_{\mathbf{x}}(\mathbf{x}\mathbf{G}\mathbf{x}^\intercal),$$

subject to the constraint

$$\mathbf{x}_{\max}\mathbf{H}\mathbf{x}_{\max}^\intercal = c > 0$$

then it can be proven that

$$\mathbf{x}_{\max} = \mathbf{v}_p \sqrt{\frac{c}{\mathbf{v}_p \mathbf{H} \mathbf{v}_p^\intercal}},$$

and

$$\max_{\mathbf{x}}(\mathbf{x}\mathbf{G}\mathbf{x}^\intercal) = \mathbf{x}_{\max}\mathbf{G}\mathbf{x}_{\max}^\intercal = c\lambda_p.$$

So it can be proven that

$$\begin{aligned} \delta_{\min} &= \delta\lambda_1, \\ \delta_{\max} &= \delta\lambda_{p-1}, \end{aligned}$$

where λ_1 and λ_{p-1} are the smallest and the largest eigenvalues of the $(p-1, p-1)$ matrix $b^2\mathbf{\Sigma}^*(b^2\mathbf{\Sigma}^* + \frac{1}{m}\mathbf{\Sigma}_{\mathrm{M}}^*)^{-1}$, respectively. Once $\delta_{\min}$ and $\delta_{\max}$ are evaluated, it is possible to compute the minimum and maximum Type II errors $\beta_{\min} = F_{\chi^2}(\mathrm{UCL}|p-1, \delta_{\min})$ and $\beta_{\max} = F_{\chi^2}(\mathrm{UCL}|p-1, \delta_{\max})$ as well as the corresponding minimum and maximum out-of-control ARL's, $\mathrm{ARL}_{1,\min} = \frac{1}{1-\beta_{\min}}$ and $\mathrm{ARL}_{1,\max} = \frac{1}{1-\beta_{\max}}$. These values have to be considered as being the "best" and "worst" out-of-control ARL values of the Hotelling CoDa T^2 control chart in the presence of measurement errors for a fixed value of δ.

3.3 Performance of the Hotelling CoDa T^2 control chart in the presence of measurement errors

The goal of this section is to investigate the performance of the Hotelling CoDa T^2 control chart in the presence of measurement errors. For simplicity, we assume that $p = 3$ and the measurement errors variance-covariance matrix is $\mathbf{\Sigma}_{\mathrm{M}}^* = \sigma_{\mathrm{M}}^2\mathbf{I}$, where σ_{M} is the standard-deviation measurement error (common for all dimensions) and $\mathbf{I}$ is the $(2, 2)$ identity matrix. Similar to Linna et al. (2001), the two following situations are considered for the CoDa variance-covariance matrix $\mathbf{\Sigma}^*$

Case #1 uncorrelated case

$$\mathbf{\Sigma}^* = \begin{pmatrix} 0.005 & 0 \\ 0 & 0.01 \end{pmatrix},$$

Case #2 correlated case

$$\mathbf{\Sigma}^* = \begin{pmatrix} 0.005 & 0.002 \\ 0.002 & 0.01 \end{pmatrix}.$$

We will now separately investigate the influence of parameters σ_{M}, b and m on the performance of the Hotelling CoDa T^2 control chart in the presence of measurement errors.

3.3.1 Influence of parameter σ_{M}

In this subsection, we investigate the influence of the parameter $\sigma_{\mathrm{M}} \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6\}$ for fixed values $b = 1$ and $m = 3$. For shifts $\delta \in [0, 2]$, values of ARL are given in Table 3.1 on the next page also with the different ARL curves are plotted in Figure 3.1 on page 39 for the Hotelling CoDa T^2 control chart without and in the presence of measurement errors. The ARL curve corresponding to the without measurement error case is shown using a solid line, the ARL curves (best and worst situations) corresponding to the Case #1 are shown using dashed lines and the ones (best and worst situations) corresponding to the Case #2 are shown using dotted lines. From the plots in Figure 3.1 on page 39, we can draw the following conclusions:

1. the out-of-control ARL difference between the best $\mathrm{ARL}_{1,\min}$ and worst $\mathrm{ARL}_{1,\max}$ situations is *always* larger for Case #2 than for Case #1 . For instance, when $\sigma_{\mathrm{M}} = 0.1$ and $\delta = 1.4$, the best and worst ARL values are $(\mathrm{ARL}_{1,\min} = 101.46, \mathrm{ARL}_{1,\max} = 131.54)$ for Case #1, while for Case #2, they are equal to $(\mathrm{ARL}_{1,\min} = 98.56, \mathrm{ARL}_{1,\max} = 137.77)$.
2. as expected, when σ_{M} increases the out-of-control ARL also increases (i.e. an increase in σ_{M} has a *negative* impact on the Hotelling CoDa T^2 control chart). For instance, if $\sigma_{\mathrm{M}} = 0.3$ and $\delta = 1.4$, the best and worst ARL values are $(\mathrm{ARL}_{1,\min} = 147.61, \mathrm{ARL}_{1,\max} = 169.28)$ for Case #1 and $(\mathrm{ARL}_{1,\min} = 145.07, \mathrm{ARL}_{1,\max} = 172.94)$ for Case #2. But, when $\sigma_{\mathrm{M}} = 0.5$ (increase), the best and worst ARL values are $(\mathrm{ARL}_{1,\min} = 164.36, \mathrm{ARL}_{1,\max} = 180.21)$ for Case #1 and $(\mathrm{ARL}_{1,\min} = 162.39, \mathrm{ARL}_{1,\max} = 182.71)$ for Case #2.

3.3.2 Influence of parameter b

In this subsection, we investigate the influence of the parameter $b \in \{0.25, 0.5, 1, 2, 4, 8\}$ for fixed values $\sigma_{\mathrm{M}} = 0.3$ and $m = 3$. All the values of ARL are given in Table 3.2 on page 40 and similar to Figure 3.1 on page 39, we plotted in Figure 3.2 on page 41 several ARL curves for shifts $\delta \in [0, 2]$. From these plots we can conclude that when b increases the out-of-control ARL decreases (i.e. an increase of b has a *positive* impact on the Hotelling CoDa T^2 control chart). For instance, if $\delta = 1.4$ and $b = 1$, the best and worst ARL values are $(\mathrm{ARL}_{1,\min} = 147.61, \mathrm{ARL}_{1,\max} = 169.28)$ for Case #1 and $(\mathrm{ARL}_{1,\min} = 145.07, \mathrm{ARL}_{1,\max} = 172.94)$ for Case #2. But, when $b = 4$ (increase), the best and worst ARL values are $(\mathrm{ARL}_{1,\min} = 48.62, \mathrm{ARL}_{1,\max} = 64.94)$ for Case #1 and $(\mathrm{ARL}_{1,\min} = 47.44, \mathrm{ARL}_{1,\max} = 69.61)$ for Case #2.

3.3.3 Influence of parameter m

In this subsection, we investigate the influence of the parameter $m \in \{1, 2, 3, 4, 5, 6\}$ for fixed values $\sigma_{\mathrm{M}} = 0.3$ and $b = 1$. All the values of ARL are given in Table 3.3 on page 42 and similar to Figures 3.1 and 3.2, we plotted in Figure 3.3 on page 43 several ARL curves for shifts $\delta \in [0, 2]$. From these plots we can see that when m increases the out-of-control ARL slowly decreases (i.e. an increase in m has a *positive* impact on the Hotelling CoDa T^2 control chart). For instance, if $\delta = 1.4$ and $m = 3$, the best and worst ARL values are $(\mathrm{ARL}_{1,\min} = 147.61, \mathrm{ARL}_{1,\max} = 169.28)$ for Case #1 and $(\mathrm{ARL}_{1,\min} = 145.07, \mathrm{ARL}_{1,\max} = 172.94)$ for Case #2. But, when $m = 5$ (increase), the best and worst ARL values are $(\mathrm{ARL}_{1,\min} = 127.08, \mathrm{ARL}_{1,\max} = 154.08)$ for Case #1 and $(\mathrm{ARL}_{1,\min} = 124.16, \mathrm{ARL}_{1,\max} = 159.04)$ for Case #2.

3.4 Illustrative example

A company produces a muesli (for breakfast), where every 100 grams contains: (A) 66% of whole-grain cereals (barley flakes, oat flakes, wheat flakes), (B) 24% of dried fruits (raisin, papaya, banana) and

Table 3.1 – Influence of parameter σ_{M}

Without correlation

		ARL					
delta	without M.E	$\sigma_{\mathrm{M}} = 0.1$	$\sigma_{\mathrm{M}} = 0.2$	$\sigma_{\mathrm{M}} = 0.3$	$\sigma_{\mathrm{M}} = 0.4$	$\sigma_{\mathrm{M}} = 0.5$	$\sigma_{\mathrm{M}} = 0.6$
0.00	200.00	(200.00 , 200.00)	(200.00 , 200.00)	(200.00 , 200.00)	(200.00 , 200.00)	(200.00 , 200.00)	(200.00 , 200.00)
0.10	156.75	(188.37 , 193.29)	(193.29 , 196.36)	(195.28 , 197.50)	(196.36 , 198.10)	(197.04 , 198.47)	(197.50 , 198.71)
0.20	127.08	(177.84 , 186.94)	(186.94 , 192.83)	(190.74 , 195.06)	(192.83 , 196.23)	(194.15 , 196.95)	(195.06 , 197.44)
0.30	105.59	(168.26 , 180.94)	(180.94 , 189.41)	(186.38 , 192.67)	(189.41 , 194.39)	(191.33 , 195.46)	(192.67 , 196.19)
0.40	89.41	(159.51 , 175.25)	(175.25 , 186.08)	(182.19 , 190.32)	(186.08 , 192.58)	(188.58 , 193.98)	(190.32 , 194.94)
0.50	76.86	(151.48 , 169.86)	(169.86 , 182.86)	(178.15 , 188.02)	(182.86 , 190.80)	(185.90 , 192.53)	(188.02 , 193.71)
0.60	66.88	(144.11 , 164.74)	(164.74 , 179.72)	(174.25 , 185.77)	(179.72 , 189.04)	(183.28 , 191.09)	(185.77 , 192.49)
0.70	58.80	(137.30 , 159.87)	(159.87 , 176.68)	(170.50 , 183.57)	(176.68 , 187.31)	(180.72 , 189.67)	(183.57 , 191.29)
0.80	52.15	(131.01 , 155.23)	(155.23 , 173.72)	(166.88 , 181.40)	(173.72 , 185.61)	(178.22 , 188.27)	(181.40 , 190.09)
0.90	46.60	(125.18 , 150.82)	(150.82 , 170.84)	(163.39 , 179.28)	(170.84 , 183.94)	(175.78 , 186.88)	(179.28 , 188.91)
1.00	41.92	(119.76 , 146.60)	(146.60 , 168.04)	(160.01 , 177.21)	(168.04 , 182.29)	(173.39 , 185.51)	(177.21 , 187.75)
1.10	37.92	(114.71 , 142.58)	(142.58 , 165.32)	(156.75 , 175.17)	(165.32 , 180.66)	(171.06 , 184.16)	(175.17 , 186.59)
1.20	34.48	(110.00 , 138.74)	(138.74 , 162.67)	(153.60 , 173.17)	(162.67 , 179.06)	(168.78 , 182.83)	(173.17 , 185.45)
1.30	31.50	(105.59 , 135.06)	(135.06 , 160.09)	(150.56 , 171.20)	(160.09 , 177.48)	(166.54 , 181.51)	(171.20 , 184.32)
1.40	28.90	(101.46 , 131.54)	(131.54 , 157.58)	(147.61 , 169.28)	(157.58 , 175.92)	(164.36 , 180.21)	(169.28 , 183.19)
1.50	26.62	(97.59 , 128.17)	(128.17 , 155.13)	(144.75 , 167.39)	(155.13 , 174.39)	(162.23 , 178.92)	(167.39 , 182.09)
1.60	24.60	(93.94 , 124.94)	(124.94 , 152.74)	(141.99 , 165.53)	(152.74 , 172.88)	(160.14 , 177.65)	(165.53 , 180.99)
1.70	22.81	(90.51 , 121.84)	(121.84 , 150.42)	(139.31 , 163.71)	(150.42 , 171.39)	(158.09 , 176.39)	(163.71 , 179.90)
1.80	21.21	(87.28 , 118.86)	(118.86 , 148.15)	(136.71 , 161.93)	(148.15 , 169.92)	(156.09 , 175.15)	(161.93 , 178.82)
1.90	19.77	(84.22 , 115.99)	(115.99 , 145.94)	(134.19 , 160.17)	(145.94 , 168.48)	(154.13 , 173.92)	(160.17 , 177.76)
2.00	18.48	(81.34 , 113.24)	(113.24 , 143.78)	(131.75 , 158.45)	(143.78 , 167.05)	(152.21 , 172.71)	(158.45 , 176.70)

With correlation

		ARL					
delta	without M.E	$\sigma_{\mathrm{M}} = 0.1$	$\sigma_{\mathrm{M}} = 0.2$	$\sigma_{\mathrm{M}} = 0.3$	$\sigma_{\mathrm{M}} = 0.4$	$\sigma_{\mathrm{M}} = 0.5$	$\sigma_{\mathrm{M}} = 0.6$
0.00	200.00	(200.00 , 200.00)	(200.00 , 200.00)	(200.00 , 200.00)	(200.00 , 200.00)	(200.00 , 200.00)	(200.00 , 200.00)
0.10	156.75	(187.79 , 194.10)	(192.89 , 196.83)	(194.99 , 197.84)	(196.13 , 198.36)	(196.85 , 198.68)	(197.34 , 198.89)
0.20	127.08	(176.78 , 188.49)	(186.20 , 193.75)	(190.18 , 195.71)	(192.38 , 196.74)	(193.77 , 197.37)	(194.74 , 197.79)
0.30	105.59	(166.81 , 183.14)	(179.89 , 190.75)	(185.57 , 193.63)	(188.75 , 195.14)	(190.78 , 196.07)	(192.19 , 196.70)
0.40	89.41	(157.74 , 178.04)	(173.92 , 187.82)	(181.15 , 191.58)	(185.24 , 193.56)	(187.87 , 194.79)	(189.70 , 195.62)
0.50	76.86	(149.47 , 173.18)	(168.28 , 184.97)	(176.90 , 189.56)	(181.83 , 192.01)	(185.03 , 193.52)	(187.27 , 194.55)
0.60	66.88	(141.88 , 168.53)	(162.94 , 182.20)	(172.81 , 187.59)	(178.53 , 190.47)	(182.26 , 192.27)	(184.89 , 193.50)
0.70	58.80	(134.91 , 164.09)	(157.88 , 179.49)	(168.88 , 185.64)	(175.32 , 188.96)	(179.56 , 191.03)	(182.55 , 192.45)
0.80	52.15	(128.48 , 159.84)	(153.07 , 176.84)	(165.09 , 183.73)	(172.22 , 187.46)	(176.93 , 189.80)	(180.27 , 191.41)
0.90	46.60	(122.54 , 155.77)	(148.50 , 174.27)	(161.45 , 181.86)	(169.20 , 185.99)	(174.36 , 188.59)	(178.03 , 190.38)
1.00	41.92	(117.03 , 151.87)	(144.15 , 171.75)	(157.94 , 180.01)	(166.27 , 184.54)	(171.85 , 187.39)	(175.84 , 189.35)
1.10	37.92	(111.92 , 148.13)	(140.01 , 169.29)	(154.55 , 178.20)	(163.42 , 183.10)	(169.40 , 186.20)	(173.70 , 188.34)
1.20	34.48	(107.16 , 144.54)	(136.05 , 166.90)	(151.28 , 176.41)	(160.65 , 181.68)	(167.01 , 185.03)	(171.60 , 187.34)
1.30	31.50	(102.71 , 141.08)	(132.28 , 164.55)	(148.12 , 174.66)	(157.97 , 180.28)	(164.67 , 183.86)	(169.54 , 186.35)
1.40	28.90	(98.56 , 137.77)	(128.68 , 162.27)	(145.07 , 172.94)	(155.35 , 178.90)	(162.39 , 182.71)	(167.52 , 185.36)
1.50	26.62	(94.67 , 134.57)	(125.23 , 160.03)	(142.12 , 171.24)	(152.80 , 177.54)	(160.16 , 181.58)	(165.54 , 184.38)
1.60	24.60	(91.02 , 131.50)	(121.93 , 157.85)	(139.27 , 169.57)	(150.33 , 176.19)	(157.98 , 180.45)	(163.60 , 183.42)
1.70	22.81	(87.58 , 128.54)	(118.78 , 155.71)	(136.52 , 167.93)	(147.92 , 174.86)	(155.85 , 179.34)	(161.69 , 182.46)
1.80	21.21	(84.35 , 125.69)	(115.75 , 153.63)	(133.85 , 166.31)	(145.57 , 173.55)	(153.77 , 178.23)	(159.83 , 181.50)
1.90	19.77	(81.31 , 122.94)	(112.84 , 151.58)	(131.26 , 164.72)	(143.28 , 172.26)	(151.73 , 177.14)	(158.00 , 180.56)
2.00	18.48	(78.44 , 120.29)	(110.05 , 149.59)	(128.76 , 163.16)	(141.05 , 170.98)	(149.74 , 176.06)	(156.20 , 179.63)

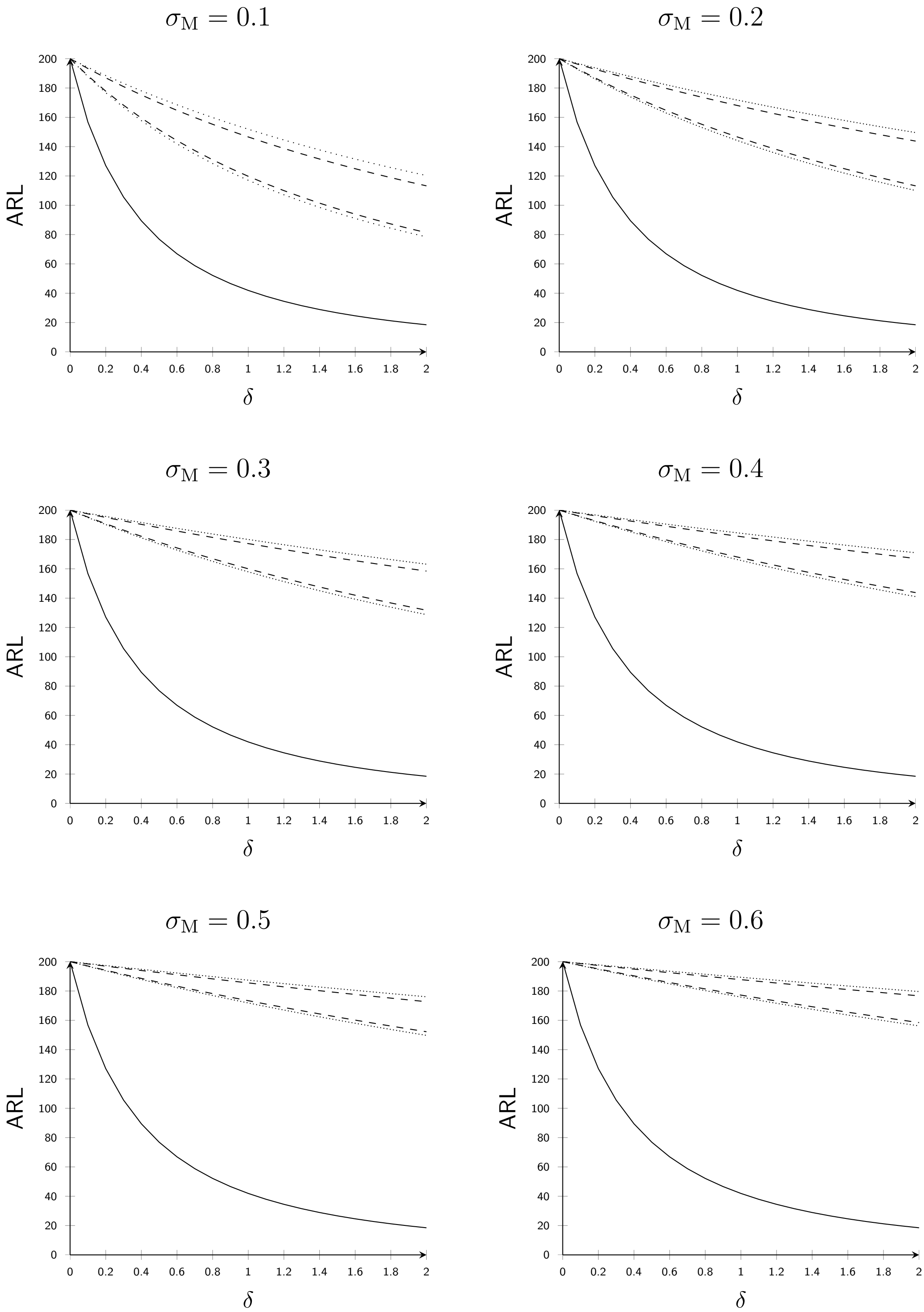

Figure 3.1 – ARL curves corresponding to the without measurement error case (solid line), Case #1 (dashed lines) and Case #2 (dotted lines), for $b = 1$, $m = 3$ and $\sigma_M \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6\}$

Table 3.2 – Influence of parameter b

Without correlation

		ARL					
delta	without M.E	$b = 0.25$	$b = 0.5$	$b = 1$	$b = 2$	$b = 4$	$b = 8$
0	200.00	(200.00 , 200.00)	(200.00 , 200.00)	(200.00 , 200.00)	(200.00 , 200.00)	(200.00 , 200.00)	(200.00 , 200.00)
0.1	156.75	(199.67 , 199.84)	(198.71 , 199.35)	(195.28 , 197.50)	(185.77 , 191.49)	(171.35 , 178.59)	(161.64 , 165.53)
0.2	127.08	(199.34 , 199.67)	(197.44 , 198.70)	(190.74 , 195.06)	(173.17 , 183.57)	(148.96 , 160.75)	(134.13 , 139.94)
0.3	105.59	(199.02 , 199.51)	(196.19 , 198.05)	(186.38 , 192.67)	(161.93 , 176.18)	(131.01 , 145.69)	(113.53 , 120.26)
0.4	89.41	(198.69 , 199.34)	(194.94 , 197.41)	(182.19 , 190.32)	(151.85 , 169.28)	(116.35 , 132.83)	(97.62 , 104.72)
0.5	76.86	(198.37 , 199.18)	(193.71 , 196.77)	(178.15 , 188.02)	(142.77 , 162.82)	(104.19 , 121.72)	(85.01 , 92.18)
0.6	66.88	(198.04 , 199.01)	(192.49 , 196.14)	(174.25 , 185.77)	(134.55 , 156.75)	(93.94 , 112.05)	(74.80 , 81.87)
0.7	58.80	(197.72 , 198.85)	(191.29 , 195.51)	(170.50 , 183.57)	(127.08 , 151.06)	(85.22 , 103.57)	(66.40 , 73.29)
0.8	52.15	(197.40 , 198.69)	(190.09 , 194.88)	(166.88 , 181.40)	(120.26 , 145.69)	(77.72 , 96.08)	(59.40 , 66.04)
0.9	46.60	(197.07 , 198.52)	(188.91 , 194.26)	(163.39 , 179.28)	(114.02 , 140.64)	(71.22 , 89.41)	(53.49 , 59.86)
1.0	41.92	(196.75 , 198.36)	(187.75 , 193.63)	(160.01 , 177.21)	(108.28 , 135.86)	(65.53 , 83.46)	(48.44 , 54.54)
1.1	37.92	(196.43 , 198.20)	(186.59 , 193.02)	(156.75 , 175.17)	(103.00 , 131.35)	(60.53 , 78.11)	(44.10 , 49.92)
1.2	34.48	(196.12 , 198.04)	(185.45 , 192.40)	(153.60 , 173.17)	(98.13 , 127.08)	(56.10 , 73.29)	(40.33 , 45.88)
1.3	31.50	(195.80 , 197.87)	(184.32 , 191.79)	(150.56 , 171.20)	(93.61 , 123.03)	(52.15 , 68.92)	(37.04 , 42.33)
1.4	28.90	(195.48 , 197.71)	(183.19 , 191.18)	(147.61 , 169.28)	(89.41 , 119.18)	(48.62 , 64.94)	(34.14 , 39.18)
1.5	26.62	(195.16 , 197.55)	(182.09 , 190.58)	(144.75 , 167.39)	(85.51 , 115.53)	(45.45 , 61.32)	(31.58 , 36.39)
1.6	24.60	(194.85 , 197.39)	(180.99 , 189.98)	(141.99 , 165.53)	(81.87 , 112.05)	(42.59 , 58.00)	(29.30 , 33.88)
1.7	22.81	(194.53 , 197.23)	(179.90 , 189.38)	(139.31 , 163.71)	(78.48 , 108.74)	(39.99 , 54.96)	(27.27 , 31.64)
1.8	21.21	(194.22 , 197.07)	(178.82 , 188.78)	(136.71 , 161.93)	(75.30 , 105.59)	(37.64 , 52.15)	(25.44 , 29.61)
1.9	19.77	(193.91 , 196.90)	(177.76 , 188.19)	(134.19 , 160.17)	(72.32 , 102.58)	(35.49 , 49.56)	(23.80 , 27.78)
2.0	18.48	(193.60 , 196.74)	(176.70 , 187.60)	(131.75 , 158.45)	(69.52 , 99.71)	(33.52 , 47.17)	(22.31 , 26.11)

With correlation

		ARL					
delta	without M.E	$b = 0.25$	$b = 0.5$	$b = 1$	$b = 2$	$b = 4$	$b = 8$
0	200.00	(200.00 , 200.00)	(200.00 , 200.00)	(200.00 , 200.00)	(200.00 , 200.00)	(200.00 , 200.00)	(200.00 , 200.00)
0.10	156.75	(199.65 , 199.86)	(198.63 , 199.44)	(194.99 , 197.84)	(185.12 , 192.47)	(170.71 , 180.22)	(161.35 , 166.64)
0.20	127.08	(199.30 , 199.72)	(197.27 , 198.88)	(190.18 , 195.71)	(172.01 , 185.41)	(147.93 , 163.51)	(133.71 , 141.62)
0.30	105.59	(198.95 , 199.57)	(195.93 , 198.32)	(185.57 , 193.63)	(160.37 , 178.78)	(129.77 , 149.23)	(113.05 , 122.24)
0.40	89.41	(198.60 , 199.43)	(194.60 , 197.77)	(181.15 , 191.58)	(149.99 , 172.53)	(114.99 , 136.89)	(97.12 , 106.83)
0.50	76.86	(198.25 , 199.29)	(193.29 , 197.22)	(176.90 , 189.56)	(140.68 , 166.63)	(102.76 , 126.15)	(84.50 , 94.33)
0.60	66.88	(197.91 , 199.15)	(192.00 , 196.67)	(172.81 , 187.59)	(132.27 , 161.07)	(92.50 , 116.71)	(74.31 , 84.02)
0.70	58.80	(197.56 , 199.01)	(190.72 , 196.12)	(168.88 , 185.64)	(124.66 , 155.81)	(83.78 , 108.37)	(65.93 , 75.39)
0.80	52.15	(197.22 , 198.87)	(189.45 , 195.58)	(165.09 , 183.73)	(117.74 , 150.82)	(76.30 , 100.96)	(58.94 , 68.09)
0.90	46.60	(196.87 , 198.73)	(188.20 , 195.04)	(161.45 , 181.86)	(111.42 , 146.10)	(69.83 , 94.33)	(53.05 , 61.84)
1.00	41.92	(196.53 , 198.59)	(186.96 , 194.50)	(157.94 , 180.01)	(105.64 , 141.61)	(64.18 , 88.36)	(48.03 , 56.44)
1.10	37.92	(196.19 , 198.45)	(185.73 , 193.96)	(154.55 , 178.20)	(100.32 , 137.35)	(59.21 , 82.98)	(43.71 , 51.75)
1.20	34.48	(195.85 , 198.31)	(184.52 , 193.43)	(151.28 , 176.41)	(95.42 , 133.29)	(54.82 , 78.10)	(39.96 , 47.63)
1.30	31.50	(195.51 , 198.17)	(183.32 , 192.90)	(148.12 , 174.66)	(90.89 , 129.43)	(50.92 , 73.66)	(36.68 , 44.00)
1.40	28.90	(195.17 , 198.03)	(182.13 , 192.37)	(145.07 , 172.94)	(86.70 , 125.74)	(47.44 , 69.61)	(33.80 , 40.78)
1.50	26.62	(194.83 , 197.89)	(180.96 , 191.84)	(142.12 , 171.24)	(82.81 , 122.23)	(44.31 , 65.89)	(31.26 , 37.92)
1.60	24.60	(194.50 , 197.75)	(179.80 , 191.32)	(139.27 , 169.57)	(79.18 , 118.87)	(41.48 , 62.48)	(29.00 , 35.35)
1.70	22.81	(194.16 , 197.61)	(178.65 , 190.80)	(136.52 , 167.93)	(75.81 , 115.65)	(38.93 , 59.34)	(26.98 , 33.04)
1.80	21.21	(193.83 , 197.47)	(177.52 , 190.28)	(133.85 , 166.31)	(72.65 , 112.58)	(36.61 , 56.44)	(25.17 , 30.95)
1.90	19.77	(193.50 , 197.33)	(176.39 , 189.76)	(131.26 , 164.72)	(69.70 , 109.64)	(34.50 , 53.75)	(23.53 , 29.06)
2.00	18.48	(193.16 , 197.19)	(175.28 , 189.25)	(128.76 , 163.16)	(66.93 , 106.82)	(32.57 , 51.26)	(22.06 , 27.34)

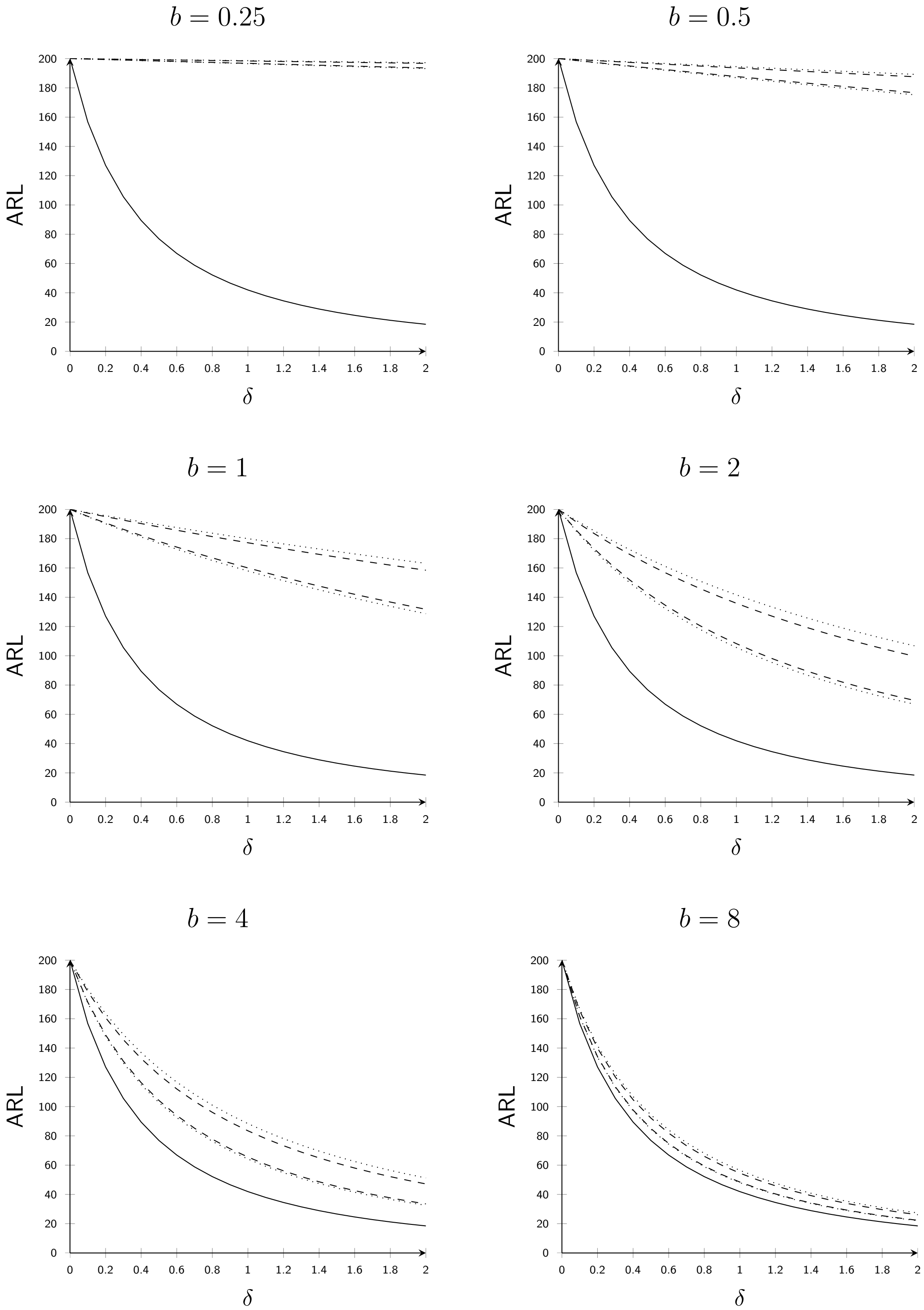

Figure 3.2 – ARL curves corresponding to the without measurement error case (solid line), Case #1 (dashed lines) and Case #2 (dotted lines), for $\sigma_{\mathrm{M}} = 0.3$, $m = 3$ and $b \in \{0.25, 0.5, 1, 2, 4, 8\}$

Table 3.3 – Influence of parameter m

	Without correlation						
		ARL					
delta	without M.E	$m = 1$	$m = 2$	$m = 3$	$m = 4$	$m = 5$	$m = 6$
0	200.00	(200.00 , 200.00)	(200.00 , 200.00)	(200.00 , 200.00)	(200.00 , 200.00)	(200.00 , 200.00)	(200.00 , 200.00)
0.10	156.75	(198.30 , 199.13)	(196.73 , 198.30)	(195.28 , 197.50)	(193.93 , 196.73)	(192.67 , 195.99)	(191.49 , 195.28)
0.20	127.08	(196.63 , 198.28)	(193.56 , 196.63)	(190.74 , 195.06)	(188.16 , 193.56)	(185.77 , 192.12)	(183.57 , 190.74)
0.30	105.59	(194.98 , 197.42)	(190.47 , 194.98)	(186.38 , 192.67)	(182.67 , 190.47)	(179.28 , 188.37)	(176.18 , 186.38)
0.40	89.41	(193.36 , 196.58)	(187.46 , 193.36)	(182.19 , 190.32)	(177.45 , 187.46)	(173.17 , 184.75)	(169.28 , 182.19)
0.50	76.86	(191.75 , 195.74)	(184.53 , 191.75)	(178.15 , 188.02)	(172.47 , 184.53)	(167.39 , 181.24)	(162.82 , 178.15)
0.60	66.88	(190.17 , 194.90)	(181.67 , 190.17)	(174.25 , 185.77)	(167.72 , 181.67)	(161.93 , 177.84)	(156.75 , 174.25)
0.70	58.80	(188.61 , 194.07)	(178.89 , 188.61)	(170.50 , 183.57)	(163.18 , 178.89)	(156.75 , 174.55)	(151.06 , 170.50)
0.80	52.15	(187.07 , 193.25)	(176.18 , 187.07)	(166.88 , 181.40)	(158.85 , 176.18)	(151.85 , 171.35)	(145.69 , 166.88)
0.90	46.60	(185.56 , 192.43)	(173.54 , 185.56)	(163.39 , 179.28)	(154.70 , 173.54)	(147.19 , 168.26)	(140.64 , 163.39)
1.00	41.92	(184.06 , 191.62)	(170.96 , 184.06)	(160.01 , 177.21)	(150.73 , 170.96)	(142.77 , 165.25)	(135.86 , 160.01)
1.10	37.92	(182.59 , 190.82)	(168.45 , 182.59)	(156.75 , 175.17)	(146.93 , 168.45)	(138.56 , 162.34)	(131.35 , 156.75)
1.20	34.48	(181.13 , 190.02)	(165.99 , 181.13)	(153.60 , 173.17)	(143.28 , 165.99)	(134.55 , 159.51)	(127.08 , 153.60)
1.30	31.50	(179.69 , 189.22)	(163.60 , 179.69)	(150.56 , 171.20)	(139.78 , 163.60)	(130.73 , 156.75)	(123.03 , 150.56)
1.40	28.90	(178.27 , 188.43)	(161.27 , 178.27)	(147.61 , 169.28)	(136.41 , 161.27)	(127.08 , 154.08)	(119.18 , 147.61)
1.50	26.62	(176.87 , 187.65)	(158.98 , 176.87)	(144.75 , 167.39)	(133.18 , 158.98)	(123.59 , 151.48)	(115.53 , 144.75)
1.60	24.60	(175.49 , 186.87)	(156.75 , 175.49)	(141.99 , 165.53)	(130.07 , 156.75)	(120.26 , 148.96)	(112.05 , 141.99)
1.70	22.81	(174.13 , 186.10)	(154.58 , 174.13)	(139.31 , 163.71)	(127.08 , 154.58)	(117.07 , 146.50)	(108.74 , 139.31)
1.80	21.21	(172.78 , 185.34)	(152.45 , 172.78)	(136.71 , 161.93)	(124.20 , 152.45)	(114.02 , 144.11)	(105.59 , 136.71)
1.90	19.77	(171.46 , 184.57)	(150.37 , 171.46)	(134.19 , 160.17)	(121.42 , 150.37)	(111.09 , 141.78)	(102.58 , 134.19)
2.00	18.48	(170.14 , 183.82)	(148.34 , 170.14)	(131.75 , 158.45)	(118.74 , 148.34)	(108.28 , 139.51)	(99.71 , 131.75)
	With correlation						
		ARL					
delta	without M.E	$m = 1$	$m = 2$	$m = 3$	$m = 4$	$m = 5$	$m = 6$
0.0	200.00	(200.00 , 200.00)	(200.00 , 200.00)	(200.00 , 200.00)	(200.00 , 200.00)	(200.00 , 200.00)	(200.00 , 200.00)
0.1	156.75	(198.19 , 199.25)	(196.52 , 198.53)	(194.99 , 197.84)	(193.56 , 197.16)	(192.24 , 196.51)	(191.01 , 195.88)
0.2	127.08	(196.41 , 198.51)	(193.15 , 197.08)	(190.18 , 195.71)	(187.47 , 194.39)	(184.98 , 193.12)	(182.69 , 191.90)
0.3	105.59	(194.65 , 197.78)	(189.87 , 195.65)	(185.57 , 193.63)	(181.69 , 191.69)	(178.17 , 189.83)	(174.95 , 188.05)
0.4	89.41	(192.92 , 197.04)	(186.68 , 194.24)	(181.15 , 191.58)	(176.20 , 189.04)	(171.76 , 186.63)	(167.75 , 184.33)
0.5	76.86	(191.21 , 196.32)	(183.58 , 192.84)	(176.90 , 189.56)	(170.99 , 186.46)	(165.73 , 183.52)	(161.03 , 180.74)
0.6	66.88	(189.53 , 195.59)	(180.57 , 191.46)	(172.81 , 187.59)	(166.02 , 183.94)	(160.05 , 180.50)	(154.74 , 177.26)
0.7	58.80	(187.88 , 194.87)	(177.64 , 190.10)	(168.88 , 185.64)	(161.30 , 181.47)	(154.68 , 177.56)	(148.85 , 173.89)
0.8	52.15	(186.25 , 194.16)	(174.79 , 188.75)	(165.09 , 183.73)	(156.79 , 179.06)	(149.60 , 174.70)	(143.32 , 170.62)
0.9	46.60	(184.64 , 193.45)	(172.01 , 187.42)	(161.45 , 181.86)	(152.49 , 176.70)	(144.80 , 171.92)	(138.12 , 167.46)
1.0	41.92	(183.06 , 192.75)	(169.31 , 186.11)	(157.94 , 180.01)	(148.38 , 174.40)	(140.24 , 169.20)	(133.23 , 164.39)
1.1	37.92	(181.50 , 192.04)	(166.67 , 184.81)	(154.55 , 178.20)	(144.45 , 172.14)	(135.91 , 166.56)	(128.61 , 161.42)
1.2	34.48	(179.96 , 191.35)	(164.11 , 183.52)	(151.28 , 176.41)	(140.68 , 169.93)	(131.80 , 163.99)	(124.25 , 158.54)
1.3	31.50	(178.44 , 190.65)	(161.61 , 182.25)	(148.12 , 174.66)	(137.08 , 167.77)	(127.89 , 161.48)	(120.12 , 155.73)
1.4	28.90	(176.95 , 189.97)	(159.17 , 181.00)	(145.07 , 172.94)	(133.63 , 165.65)	(124.16 , 159.04)	(116.21 , 153.01)
1.5	26.62	(175.47 , 189.28)	(156.79 , 179.76)	(142.12 , 171.24)	(130.31 , 163.58)	(120.61 , 156.66)	(112.51 , 150.37)
1.6	24.60	(174.02 , 188.60)	(154.47 , 178.53)	(139.27 , 169.57)	(127.13 , 161.55)	(117.22 , 154.33)	(108.99 , 147.80)
1.7	22.81	(172.58 , 187.93)	(152.21 , 177.32)	(136.52 , 167.93)	(124.07 , 159.56)	(113.98 , 152.06)	(105.65 , 145.31)
1.8	21.21	(171.17 , 187.25)	(150.00 , 176.12)	(133.85 , 166.31)	(121.13 , 157.61)	(110.89 , 149.85)	(102.47 , 142.88)
1.9	19.77	(169.77 , 186.59)	(147.85 , 174.93)	(131.26 , 164.72)	(118.31 , 155.71)	(107.92 , 147.69)	(99.44 , 140.52)
2.0	18.48	(168.39 , 185.92)	(145.74 , 173.76)	(128.76 , 163.16)	(115.58 , 153.84)	(105.09 , 145.58)	(96.55 , 138.22)

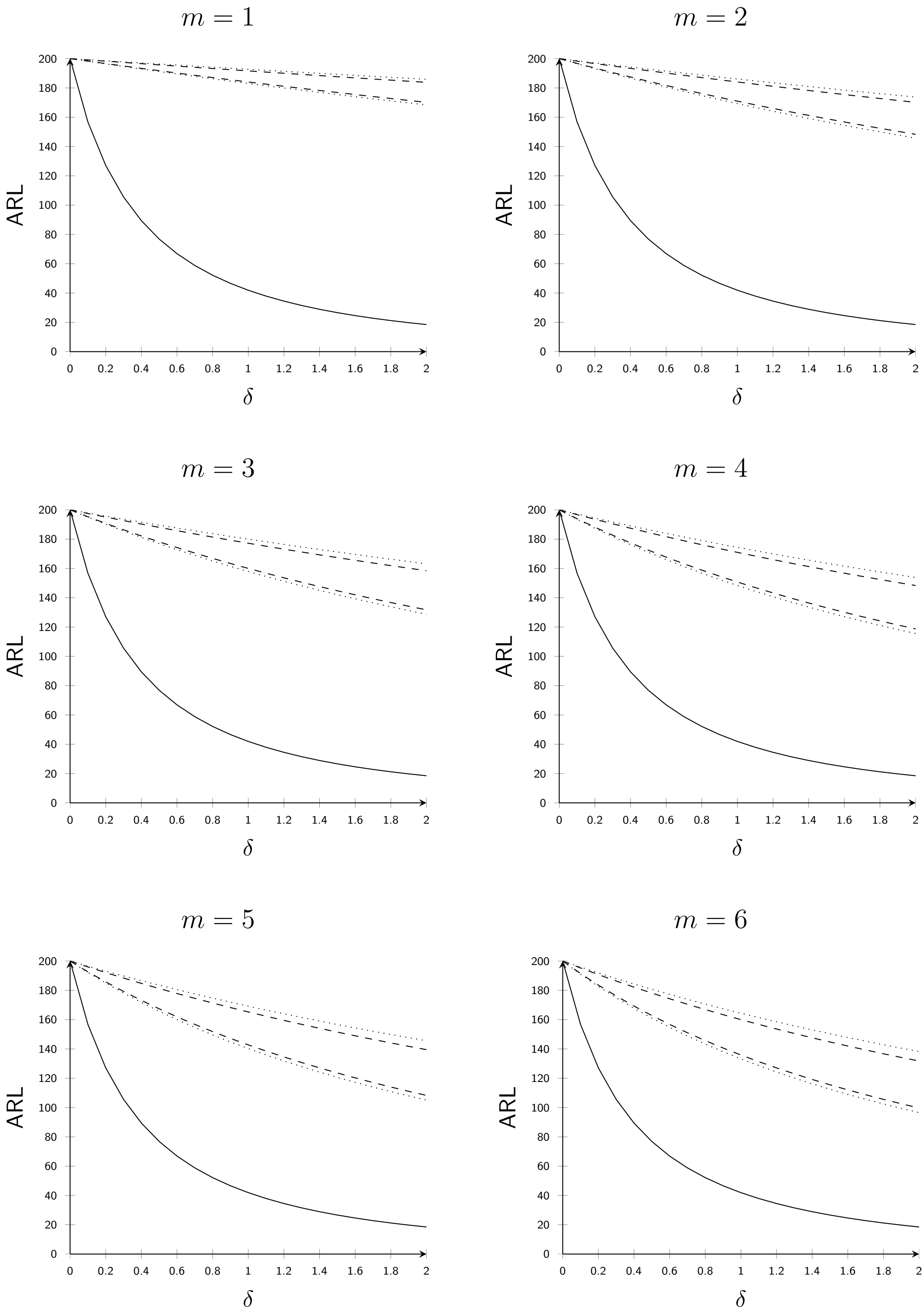

Figure 3.3 – ARL curves corresponding to the without measurement error case (solid line), Case #1 (dashed lines) and Case #2 (dotted lines), for $\sigma_{\mathrm{M}} = 0.3$, $b = 1$ and $m \in \{1, 2, 3, 4, 5, 6\}$

(C) 10% of nuts (almond, hazelnut, coconut). In the first step, the company has decided to calibrate the measurement device in charge of measuring the percentage of each components of the produced muesli (i.e. estimate $\mathbf{a}$, b and $\boldsymbol{\Sigma}^*_{\mathrm{M}}$). In order to do this calibration, the company carefully prepared $k = 4$ samples with percentages for each component *perfectly known* in advance and measured them $n = 7$ times. The results of this calibation procedure are in Table 3.4 on the next page as well as in Figure 3.4 on the facing page. The values $\mathbf{y}_1, \ldots, \mathbf{y}_4$ correspond to the $k = 4$ samples with a known composition (represented by "+" in Figure 3.4 on the next page) and the values $\mathbf{x}_{1,1}, \ldots, \mathbf{x}_{4,7}$ correspond to the $k \times n = 28$ observed values obtained using the measurement device (represented by "∘" in Figure 3.4 on the facing page). In Table 3.4 on the next page, x and y refers to row data. Table 3.4 on the facing page also provides the ilr transformed values $\mathbf{y}^*_i = \mathrm{ilr}(\mathbf{y}_i)$ and $\mathbf{x}^*_{i,j} = \mathrm{ilr}(\mathbf{x}_{i,j})$ (plotted in the right side of Figure 3.4 on the next page). An estimate $\hat{\mathbf{a}}$, $\hat{b}$ and $\hat{\boldsymbol{\Sigma}}^*_{\mathrm{M}}$ of $\mathbf{a}$, b and $\boldsymbol{\Sigma}^*_{\mathrm{M}}$ with $d = 2$, $\mathbf{v}_i = \mathbf{y}^*_i$ and $\mathbf{u}_{i,j} = \mathbf{x}^*_{i,j}$ for $i = 1, \ldots, 4$ and $j = 1, \ldots, 7$, can be obtained using the method given below,

Let $\mathbf{v}_1, \ldots, \mathbf{v}_k$ be k *known* row vectors in $\mathbb{R}^d$. For each vector $\mathbf{v}_i$, $i = 1, \ldots, k$, we observe n random vectors $\mathbf{u}_{i,1}, \ldots, \mathbf{u}_{i,n}$ and we assume that $\mathbf{u}_{i,j} = \mathbf{a} + b\mathbf{v}_i + \boldsymbol{\varepsilon}_{i,j}$, where $\mathbf{a} \in \mathbb{R}^d$ and $b \in \mathbb{R}$ are a constant row vector and a constant scalar, respectively, while $\boldsymbol{\varepsilon}_{i,j}$ is a random error term which follows a multivariate normal distribution $\mathrm{MNOR}_{\mathbb{R}^d}(\mathbf{0}, \boldsymbol{\Sigma}^*_{\mathrm{M}})$. In order to estimate $\mathbf{a}$ and b, based on the k known vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ and on the $k \times n$ observations $\mathbf{u}_{1,1}, \ldots, \mathbf{u}_{1,n}$, $\mathbf{u}_{2,1}, \ldots, \mathbf{u}_{2,n}$, $\ldots, \mathbf{u}_{k,1}, \ldots, \mathbf{u}_{k,n}$, we suggest to define the $(d \times k \times n, 1)$ column vector $\mathbf{u}$ and the $(d \times k \times n, d + 1)$ matrix $\mathbf{V}$ as

$$\mathbf{u} = \begin{pmatrix} \mathbf{u}^\intercal_{1,1} \\ \vdots \\ \mathbf{u}^\intercal_{1,n} \\ \mathbf{u}^\intercal_{2,1} \\ \vdots \\ \mathbf{u}^\intercal_{2,n} \\ \vdots \\ \mathbf{u}^\intercal_{k,1} \\ \vdots \\ \mathbf{u}^\intercal_{k,n} \end{pmatrix} \quad \text{and} \quad \mathbf{V} = \begin{pmatrix} \mathbf{I} & \mathbf{v}^\intercal_1 \\ \vdots & \vdots \\ \mathbf{I} & \mathbf{v}^\intercal_1 \\ \mathbf{I} & \mathbf{v}^\intercal_2 \\ \vdots & \vdots \\ \mathbf{I} & \mathbf{v}^\intercal_2 \\ \vdots & \vdots \\ \mathbf{I} & \mathbf{v}^\intercal_k \\ \vdots & \vdots \\ \mathbf{I} & \mathbf{v}^\intercal_k \end{pmatrix},$$

where $\mathbf{I}$ is the (d, d) identity matrix, we can compute the $(d + 1, 1)$ column vector $\hat{\mathbf{c}}$ using ordinary least square estimation for minimizing the sum of square of residuals $\boldsymbol{\Sigma}$ that is

$$\hat{\mathbf{c}} = (\mathbf{V}^\intercal\mathbf{V})^{-1}\mathbf{V}^\intercal\mathbf{u}.$$

Then, an estimate $\hat{\mathbf{a}}$ of $\mathbf{a}$ can be obtained as the first d components of $\hat{\mathbf{c}} = (c_1, c_2, \ldots, c_d, c_{d+1})^\intercal$ and an estimate $\hat{b}$ of b can be obtained as the $(d + 1)th$(last) component of $\hat{\mathbf{c}}$, i.e. $\hat{\mathbf{a}} = (c_1, \ldots, c_d)$ and $\hat{b} = c_{d+1}$. An estimator $\hat{\boldsymbol{\Sigma}}$ can simply be obtained using

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{k \times n} \sum_{i=1}^{k} \sum_{j=1}^{n} (\mathbf{u}_{i,j} - \hat{\mathbf{a}} - \hat{b}\mathbf{v}_i)^\intercal (\mathbf{u}_{i,j} - \hat{\mathbf{a}} - \hat{b}\mathbf{v}_i).$$

Using this method, we obtain the following estimates:

1. $\hat{\mathbf{a}} = (0.3354, 0.3357, 0.3289)$ or, equivalently, $\hat{\mathbf{a}}^* = (0.0162972, -0.0006318)$,
2. $\hat{b} = 1.1070$,
3. $\hat{\boldsymbol{\Sigma}}^*_{\mathrm{M}} = \begin{pmatrix} 0.0014346 & 0.0007812 \\ 0.0007812 & 0.0102893 \end{pmatrix}$.

As we can see, the value of $\hat{\mathbf{a}}$ is very close from the center $\mathbf{0}_{\mathcal{S}^3} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ of the simplex $\mathcal{S}^3$ but the value of $\hat{b}$ is a bit larger than 1.

Table 3.4 – Data used for calibrating the measurement system

i	j	$\mathbf{y}_i$			$\mathbf{x}_{i,j}$			$\mathbf{y}_i^*$		$\mathbf{x}_{i,j}^*$	
1	1	0.33	0.33	0.33	0.34	0.33	0.33	0.0000	0.0000	0.0122	0.0211
	2				0.32	0.35	0.33			0.0115	-0.0634
	3				0.33	0.35	0.32			0.0491	-0.0416
	4				0.31	0.35	0.34			-0.0259	-0.0858
	5				0.36	0.34	0.30			0.1255	0.0404
	6				0.34	0.32	0.34			-0.0247	0.0429
	7				0.31	0.36	0.33			0.0100	-0.1057
2	1	0.60	0.20	0.20	0.62	0.19	0.19	0.4485	0.7768	0.4828	0.8363
	2				0.65	0.17	0.18			0.5009	0.9484
	3				0.59	0.22	0.19			0.5224	0.6976
	4				0.61	0.20	0.19			0.4971	0.7885
	5				0.67	0.15	0.18			0.4621	1.0583
	6				0.65	0.17	0.18			0.5009	0.9484
	7				0.61	0.19	0.20			0.4343	0.8248
3	1	0.20	0.60	0.20	0.17	0.65	0.18	0.4485	-0.7768	0.5009	-0.9484
	2				0.14	0.69	0.17			0.4926	-1.1279
	3				0.21	0.60	0.19			0.5103	-0.7423
	4				0.19	0.62	0.19			0.4828	-0.8363
	5				0.20	0.62	0.18			0.5479	-0.8000
	6				0.16	0.66	0.18			0.4823	-1.0020
	7				0.20	0.62	0.18			0.5479	-0.8000
4	1	0.20	0.20	0.60	0.18	0.20	0.62	-0.8970	0.0000	-0.9668	-0.0745
	2				0.19	0.19	0.62			-0.9657	0.0000
	3				0.21	0.19	0.60			-0.8980	0.0708
	4				0.21	0.18	0.61			-0.9336	0.1090
	5				0.20	0.18	0.62			-0.9668	0.0745
	6				0.19	0.19	0.62			-0.9657	0.0000
	7				0.21	0.17	0.62			-0.9702	0.1494

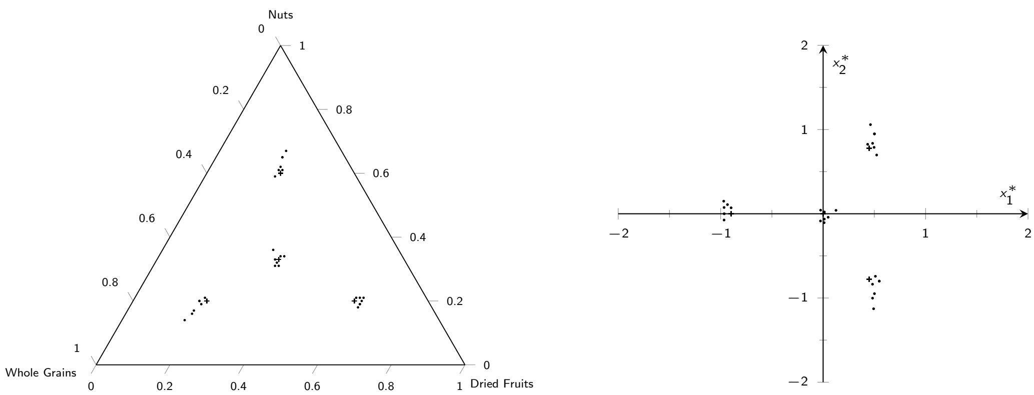

Figure 3.4 – Data used for calibrating the measurement system: $\mathbf{y}_i$ and $\mathbf{x}_{i,j}$ in $\mathcal{S}^p$ (left side), $\mathbf{y}_i^*$ and $\mathbf{x}_{i,j}^*$ in $\mathbb{R}^2$ (right side)

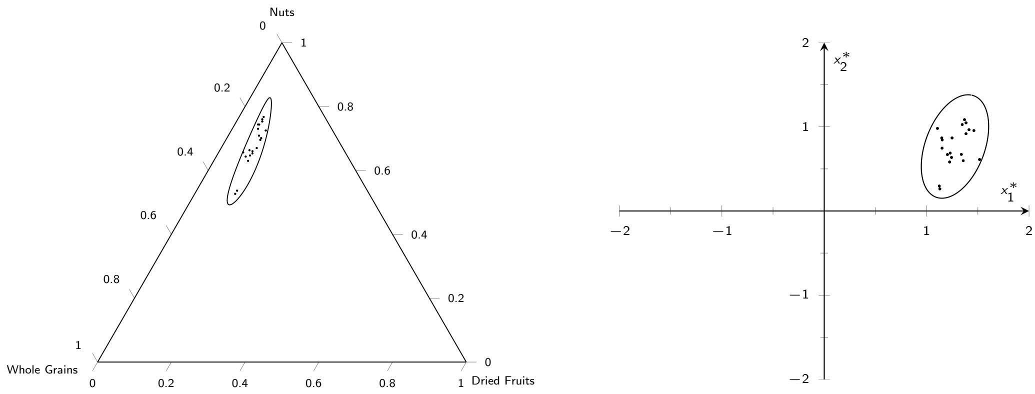

Figure 3.5 – Phase I data for the muesli example: $\bar{\mathbf{x}}_i$ (left side) and $\bar{\mathbf{x}}_i^*$ (right side)

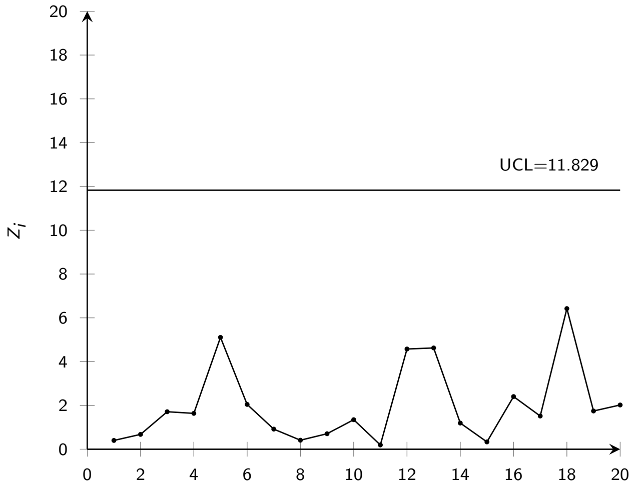

Figure 3.6 – Hotelling CoDa T^2 control chart for muesli Phase I data

In order to obtain an estimate $\hat{\boldsymbol{\mu}}_0^*$ and $\hat{\boldsymbol{\Sigma}}^*$ of the in-control composition parameters $\boldsymbol{\mu}_0^*$ and $\boldsymbol{\Sigma}^*$, $i = 1, \ldots, 20$ batches of muesli have been sampled and measured $m = 3$ times (Phase I). The results are in Table 3.5 on the facing page with the values $\mathbf{x}_{i,j}$, $\bar{\mathbf{x}}_i$ and $\bar{\mathbf{x}}_i^*$, $i = 1, \ldots, 20$. The values $\bar{\mathbf{x}}_i$ and $\bar{\mathbf{x}}_i^*$ are also plotted in Figure 3.5. From the values $\bar{\mathbf{x}}_i^*$, it is easy to obtain

1. $\hat{\boldsymbol{\mu}}_{\bar{\mathbf{x}}}^* = (1.2766, 0.7657)$,
2. $\hat{\boldsymbol{\Sigma}}_{\bar{\mathbf{x}}}^* = \begin{pmatrix} 0.0146362 & 0.0105839 \\ 0.0105839 & 0.0510887 \end{pmatrix}$.

Then solving Equations 3.2 on page 34 and 3.3 on page 34 for $\hat{\boldsymbol{\mu}}_0^*$ and $\hat{\boldsymbol{\Sigma}}^*$, we have

1. $\hat{\boldsymbol{\mu}}_0^* = \frac{1}{\hat{b}}(\hat{\boldsymbol{\mu}}_{\bar{\mathbf{x}}}^* - \hat{\mathbf{a}}^*) = (1.1385, 0.6922)$,
2. $\hat{\boldsymbol{\Sigma}}^* = \frac{1}{\hat{b}^2}\left(\hat{\boldsymbol{\Sigma}}_{\bar{\mathbf{x}}}^* - \frac{1}{m}\hat{\boldsymbol{\Sigma}}_{\mathrm{M}}^*\right) = \begin{pmatrix} 0.0115533 & 0.0084242 \\ 0.0084242 & 0.038891 \end{pmatrix}$.

In Table 3.5 on the facing page, we have also listed the Hotelling CoDa T^2 statistics Z_i, $i = 1, \ldots, 20$, computed using Equation 3.4 on page 35. These values are also plotted in Figure 3.6 with the upper control limit $\mathrm{UCL} = F_{\chi^2}^{-1}(1 - 0.0027|3 - 1) = 11.829$. As it can be noticed, all the values Z_i are smaller than $\mathrm{UCL} = 11.829$, confirming that the data in Table 3.5 on the next page are actually in-control.

Table 3.5 – Phase I data for the muesli example

i	j	$\mathbf{x}_{i,j}$			$\bar{\mathbf{x}}_i$			$\bar{\mathbf{x}}_i^*$		Z_i
1	1	0.77	0.16	0.07	0.7090	0.2078	0.0832	1.2483	0.8678	0.4008
	2	0.67	0.24	0.09						
	3	0.68	0.23	0.09						
2	1	0.64	0.27	0.09	0.6301	0.2767	0.0932	1.2241	0.5819	0.6777
	2	0.63	0.28	0.09						
	3	0.62	0.28	0.10						
3	1	0.76	0.16	0.08	0.6958	0.2108	0.0934	1.1523	0.8445	1.7118
	2	0.65	0.25	0.10						
	3	0.67	0.23	0.10						
4	1	0.65	0.27	0.08	0.6434	0.2766	0.0800	1.3575	0.5968	1.6376
	2	0.64	0.28	0.08						
	3	0.64	0.28	0.08						
5	1	0.50	0.38	0.12	0.5268	0.3634	0.1098	1.1289	0.2625	5.1085
	2	0.54	0.36	0.10						
	3	0.54	0.35	0.11						
6	1	0.80	0.14	0.06	0.7677	0.1657	0.0666	1.3699	1.0843	2.0462
	2	0.74	0.19	0.07						
	3	0.76	0.17	0.07						
7	1	0.75	0.18	0.07	0.7304	0.1995	0.0701	1.3843	0.9177	0.9184
	2	0.71	0.22	0.07						
	3	0.73	0.20	0.07						
8	1	0.65	0.26	0.09	0.6534	0.2533	0.0932	1.2030	0.6700	0.4119
	2	0.65	0.25	0.10						
	3	0.66	0.25	0.09						
9	1	0.65	0.27	0.08	0.6635	0.2565	0.0800	1.3389	0.6721	0.7071
	2	0.66	0.26	0.08						
	3	0.68	0.24	0.08						
10	1	0.76	0.17	0.07	0.7534	0.1766	0.0700	1.3478	1.0257	1.3477
	2	0.75	0.18	0.07						
	3	0.75	0.18	0.07						
11	1	0.67	0.24	0.09	0.6601	0.2499	0.0900	1.2302	0.6868	0.1944
	2	0.65	0.26	0.09						
	3	0.66	0.25	0.09						
12	1	0.53	0.36	0.11	0.5367	0.3533	0.1100	1.1234	0.2956	4.5764
	2	0.54	0.35	0.11						
	3	0.54	0.35	0.11						
13	1	0.75	0.16	0.09	0.7251	0.1813	0.0936	1.1058	0.9803	4.6248
	2	0.67	0.23	0.10						
	3	0.75	0.16	0.09						
14	1	0.67	0.24	0.09	0.6701	0.2333	0.0966	1.1510	0.7460	1.1944
	2	0.67	0.23	0.10						
	3	0.67	0.23	0.10						
15	1	0.64	0.27	0.09	0.6468	0.2632	0.0900	1.2431	0.6357	0.3344
	2	0.64	0.27	0.09						
	3	0.66	0.25	0.09						
16	1	0.72	0.21	0.07	0.7440	0.1928	0.0632	1.4614	0.9548	2.4039
	2	0.74	0.20	0.06						
	3	0.77	0.17	0.06						
17	1	0.73	0.20	0.07	0.7436	0.1899	0.0665	1.4137	0.9652	1.5162
	2	0.75	0.18	0.07						
	3	0.75	0.19	0.06						
18	1	0.65	0.28	0.07	0.6568	0.2767	0.0665	1.5169	0.6112	6.4220
	2	0.65	0.28	0.07						
	3	0.67	0.27	0.06						
19	1	0.76	0.17	0.07	0.7604	0.1730	0.0665	1.3846	1.0468	1.7457
	2	0.74	0.19	0.07						
	3	0.78	0.16	0.06						
20	1	0.67	0.23	0.10	0.7013	0.2055	0.0931	1.1475	0.8678	2.0202
	2	0.68	0.22	0.10						
	3	0.75	0.17	0.08						

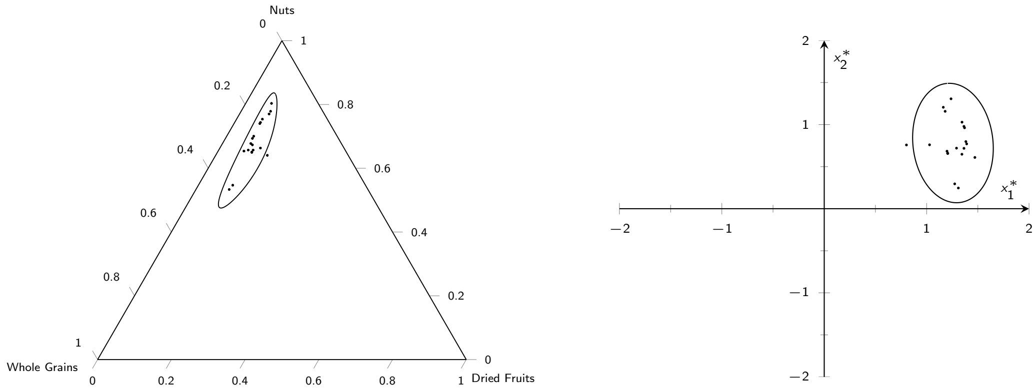

Figure 3.7 – Phase II data for the muesli example: $\bar{\mathbf{x}}_i$ (left side) and $\bar{\mathbf{x}}_i^*$ (right side)

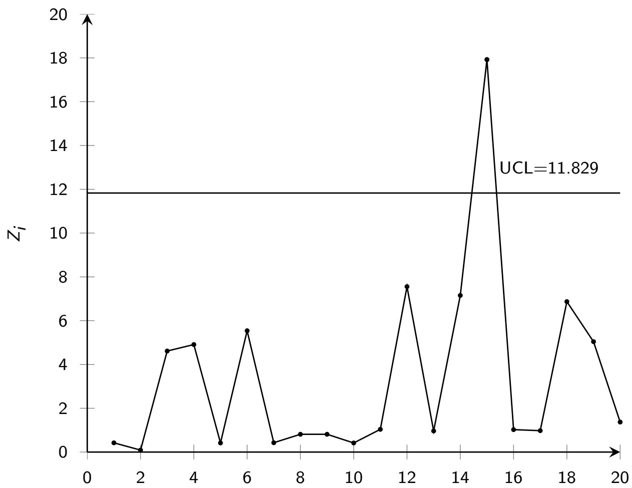

Figure 3.8 – Hotelling CoDa T^2 control chart for muesli Phase II data

During a Phase II of the production, $i = 1, \ldots, 20$ batches of muesli have been sampled and measured $m = 3$ times. The results are in Table 3.6 on the facing page with the values $\mathbf{x}_{i,j}$, $\bar{\mathbf{x}}_i$ and $\bar{\mathbf{x}}_i^*$, $i = 1, \ldots, 20$. The values of $\bar{\mathbf{x}}_i$ and $\bar{\mathbf{x}}_i^*$ are plotted in Figure 3.7 and the values Z_i are also plotted in Figure 3.8. As it can be seen, the process seems to be in-control up to sample #14 but , #15 is clearly out-of-control (see the "•" in Figures 3.7 and 3.8). Investigations showed that the level of whole-grain cereals dropped down suddenly due to a malfunction of the hatch regulating the quantity of whole-grain cereals. After having repaired this hatch, the process restarted without any out-control situations.

3.5 Conclusions

In this chapter, the influence of measurement errors on the Hotelling CoDa T^2 control chart has been investigated. A linearly covariate measurement error model for CoDa has been presented, where the quality characteristics $\mathbf{y}_i$ is not directly observable and can only be assessed using several independent measurements $\mathbf{x}_{i,j}$. Two situations have been considered for the CoDa variance-covariance matrix: a correlated case and an uncorrelated one. Different combinations of the parameters σ_{M}, b and m have been selected in order to study their influence on the performance of the Hotelling CoDa T^2 control chart. The main conclusions drawn from these investigations are: i) if b and m are kept constant, the ARL increases when the value of

Table 3.6 – Phase II data for the muesli example

i	j	$\mathbf{x}_{i,j}$			$\bar{\mathbf{x}}_i$			$\bar{\mathbf{x}}_i^*$		Z_i
1	1	0.68	0.23	0.09	0.6504	0.2563	0.0933	1.2050	0.6586	0.4213
	2	0.63	0.27	0.10						
	3	0.64	0.27	0.09						
2	1	0.68	0.24	0.08	0.6734	0.2433	0.0832	1.2916	0.7198	0.0893
	2	0.66	0.25	0.09						
	3	0.68	0.24	0.08						
3	1	0.62	0.31	0.07	0.6538	0.2761	0.0701	1.4712	0.6096	4.6133
	2	0.67	0.26	0.07						
	3	0.67	0.26	0.07						
4	1	0.67	0.22	0.11	0.6634	0.2265	0.1100	1.0283	0.7598	4.9092
	2	0.65	0.24	0.11						
	3	0.67	0.22	0.11						
5	1	0.63	0.27	0.10	0.6501	0.2566	0.0933	1.2060	0.6573	0.4168
	2	0.66	0.25	0.09						
	3	0.66	0.25	0.09						
6	1	0.77	0.15	0.08	0.7702	0.1498	0.0800	1.1804	1.1577	5.5390
	2	0.76	0.16	0.08						
	3	0.78	0.14	0.08						
7	1	0.67	0.24	0.09	0.6570	0.2498	0.0933	1.1990	0.6838	0.4268
	2	0.66	0.24	0.10						
	3	0.64	0.27	0.09						
8	1	0.64	0.28	0.08	0.6777	0.2456	0.0767	1.3644	0.7176	0.8119
	2	0.72	0.21	0.07						
	3	0.67	0.25	0.08						
9	1	0.72	0.21	0.07	0.7003	0.2265	0.0732	1.3826	0.7982	0.8118
	2	0.68	0.24	0.08						
	3	0.70	0.23	0.07						
10	1	0.64	0.26	0.10	0.6501	0.2567	0.0932	1.2062	0.6571	0.4153
	2	0.66	0.25	0.09						
	3	0.65	0.26	0.09						
11	1	0.73	0.20	0.07	0.7436	0.1863	0.0700	1.3640	0.9786	1.0378
	2	0.74	0.19	0.07						
	3	0.76	0.17	0.07						
12	1	0.81	0.12	0.07	0.8034	0.1266	0.0700	1.2381	1.3067	7.5542
	2	0.80	0.13	0.07						
	3	0.80	0.13	0.07						
13	1	0.64	0.28	0.08	0.6570	0.2630	0.0801	1.3448	0.6474	0.9632
	2	0.68	0.24	0.08						
	3	0.65	0.27	0.08						
14	1	0.82	0.11	0.07	0.7786	0.1415	0.0799	1.1633	1.2057	7.1523
	2	0.74	0.17	0.09						
	3	0.77	0.15	0.08						
15	1	0.61	0.25	0.14	0.6405	0.2193	0.1402	0.8028	0.7580	**17.9220**
	2	0.66	0.20	0.14						
	3	0.65	0.21	0.14						
16	1	0.70	0.23	0.07	0.6936	0.2332	0.0732	1.3906	0.7709	1.0240
	2	0.71	0.22	0.07						
	3	0.67	0.25	0.08						
17	1	0.74	0.19	0.07	0.7401	0.1899	0.0700	1.3700	0.9620	0.9773
	2	0.73	0.20	0.07						
	3	0.75	0.18	0.07						
18	1	0.56	0.35	0.09	0.5335	0.3765	0.0901	1.3099	0.2465	6.8694
	2	0.51	0.40	0.09						
	3	0.53	0.38	0.09						
19	1	0.55	0.35	0.10	0.5468	0.3600	0.0932	1.2736	0.2955	5.0413
	2	0.55	0.36	0.09						
	3	0.54	0.37	0.09						
20	1	0.77	0.16	0.07	0.7539	0.1760	0.0701	1.3460	1.0285	1.3701
	2	0.76	0.17	0.07						
	3	0.73	0.20	0.07						

σ_{M} increases, ii) if σ_{M} and m are kept constant, the ARL decreases as the value of b increases, iii) if σ_{M} and b are kept constant, the ARL decreases (slowly) as the value of m increases.

The work presented in this chapter has been published in Zaidi et al. (2019).

In this chapter we discussed the effect of measurement errors on the T^2_C control chart proposed by Vives-Mestres et al. (2014b). In next chapter we will discuss the effect of measurement errors on the MEWMA-CoDa control chart proposed by TRAN et al. (2018).

4

Performance of the MEWMA-CoDa Control Chart in the Presence of Measurement Errors

During the last decade, an enormous number of new advanced control charts have been proposed for univariate as well as for multivariate processes. Among them, Costa and Rahim (2006) studied a single EWMA (Exponentially Weighted Moving Average) chart for the simultaneous monitoring of the process mean and variance. They are also many researches done on nonparametric EWMA type control charts, such as Graham et al. (2011) who proposed a nonparametric exponentially weighted moving average signed-rank chart for monitoring the process location, and Yang et al. (2011) who studied a new nonparametric EWMA sign chart. The MEWMA control chart was first developed by Lowry et al. (1992b) as an extension of the univariate EWMA control chart developed by Roberts (1958). Many researchers worked on MEWMA control charts for different situations, some of them are discussed in this chapter. Linderman and Love (2000) presented several economic statistical designs for the MEWMA control chart. Khoo (2003), in order to increase the sensitivity of this chart, proposed a MEWMA control chart based on the exact variance-covariance matrix of the MEWMA statistic instead of its asymptotic version. Hawkins et al. (2007) used a full smoothing matrix instead of a diagonal-only one to improve the general performance of the MEWMA control chart and they found that this approach outperforms the diagonal version of the MEWMA chart. Mahmoud and Maravelakis (2010) studied the estimated parameter case for the MEWMA control chart. Wu et al. (2015) proposed various estimators based MEWMA control charts for process monitoring. Bilen et al. (2017) investigated an autoregressive time series model with a time-correlated output variable which depends on many multicorrelated input variables. They used a dual monitoring scheme with EWMA charts for the output variable and a MEWMA chart for the input variables to increase the sensitivity of this chart. Liu et al. (2017) studied the parameter optimization for a modified MEWMA control chart based on a PSO (Particle Swarm Optimization) algorithm. Epprecht et al. (2018) proposed an optimum variable dimension EWMA chart for multivariate statistical process control and Khusna et al. (2018) studied a multi-output least square SVR based MEWMA control chart. Furthermore Abbas et al. (2019) used phase II MEWMA control chart to monitor linear profiles with random explanatory variable under Bayesian framework.

The goal of this chapter is to study the effect of measurement errors on the MEWMA-CoDa control chart proposed by TRAN et al. (2018). The remainder of this chapter is organized as follows: in Section 4.1 details the MEWMA-CoDa control chart in the presence of measurement errors and Section 4.2 investigates the performance of this control chart. Finally, a very detailed illustrative example is provided in Section 4.3 and conclusions and future research directions are presented in Section 4.4.

4.1 MEWMA-CoDa control charts in the presence of measurement errors

When the process is in-control, we assume that $\mathbf{y}_i \sim \mathrm{MNOR}_{\mathcal{S}^p}(\boldsymbol{\mu}_0^*, \boldsymbol{\Sigma}^*)$ and, when the process is out-of-control, we assume that $\mathbf{y}_i \sim \mathrm{MNOR}_{\mathcal{S}^p}(\boldsymbol{\mu}_1^*, \boldsymbol{\Sigma}^*)$ where $\boldsymbol{\mu}_0^*$ and $\boldsymbol{\mu}_1^*$ are the in- and out-of-control mean vectors, respectively, and $\boldsymbol{\Sigma}^*$ is the variance-covariance matrix assumed to be unchanged in both cases. According to the linearly covariate measurement error model described in the previous section, the statistic monitored by the MEWMA-CoDa control chart with measurement errors is

$$Q_i^* = \mathbf{w}_i^* \boldsymbol{\Sigma}_{\mathbf{w}_i^*}^{-1} \mathbf{w}_i^{*\intercal}, \tag{4.1}$$

with

$$\mathbf{w}_i^* = r(\bar{\mathbf{x}}_i^* - \mathbf{a}^* - b\boldsymbol{\mu}_0^*) + (1-r)\mathbf{w}_{i-1}^*, \tag{4.2}$$

where $\bar{\mathbf{x}}_i^* = \mathrm{ilr}(\bar{\mathbf{x}}_i)$, $\mathbf{w}_0^* = \mathbf{0}$ and $0 < r \leq 1$ is a smoothing parameter. The MEWMA-CoDa control chart with measurement errors issues a warning signal when $Q_i^* > H$ where $H > 0$ is a specified upper control limit and $\boldsymbol{\Sigma}_{\mathbf{w}_i^*}$ is the variance-covariance matrix of $\mathbf{w}_i^*$. The asymptotic variance-covariance matrix proposed by Lowry et al. (1992b) will be used in the rest of chapter, i.e.

$$\boldsymbol{\Sigma}_{\mathbf{w}_i^*} = \frac{r}{(2-r)}\boldsymbol{\Sigma}_{\bar{\mathbf{x}}^*} = \frac{r}{(2-r)}\left(b^2\boldsymbol{\Sigma}^* + \frac{1}{m}\boldsymbol{\Sigma}_{\mathrm{M}}^*\right). \tag{4.3}$$

In practice, the MEWMA-CoDa control chart with measurement errors can be implemented in 3 phases:

1. A "Phase 0" (calibration) where the measurement parameters $\mathbf{a}^*$ (or $\mathbf{a}$), b and $\boldsymbol{\Sigma}_{\mathrm{M}}^*$ have to be estimated from a specific reference sample, measured m times for the purpose of calibration and using the estimation method provided in chapter 3 page 44. The practical implementation of this phase will be illustrated in the Illustrative example section.
2. A Phase I where the in-control chart parameters $\boldsymbol{\mu}_0^*$ and $\boldsymbol{\Sigma}^*$ have to be estimated from a assumed in-control CoDa dataset $\mathbf{x}_{i,j}$. These estimators can be obtained by computing $\bar{\mathbf{x}}_i$ and $\bar{\mathbf{x}}_i^*$, by estimating the mean vector $\hat{\boldsymbol{\mu}}_{\bar{\mathbf{x}}}^*$ and the variance-covariance matrix $\hat{\boldsymbol{\Sigma}}_{\bar{\mathbf{x}}}^*$ of the $\bar{\mathbf{x}}_i^*$ and, finally, by solving Equations 3.2 on page 34 and 3.3 on page 34 for $\hat{\boldsymbol{\mu}}_0^*$ and $\hat{\boldsymbol{\Sigma}}_{\mathrm{M}}^*$, i.e.

$$\hat{\boldsymbol{\mu}}_0^* = \frac{1}{\hat{b}}(\hat{\boldsymbol{\mu}}_{\bar{\mathbf{x}}}^* - \hat{\mathbf{a}}^*), \tag{4.4}$$

$$\hat{\boldsymbol{\Sigma}}^* = \frac{1}{\hat{b}^2}\left(\hat{\boldsymbol{\Sigma}}_{\bar{\mathbf{x}}}^* - \frac{1}{m}\hat{\boldsymbol{\Sigma}}_{\mathrm{M}}^*\right). \tag{4.5}$$

3. A Phase II where an incoming CoDa data set $\mathbf{x}_{i,j}$ (measured m times) must be monitored for detecting possible changes in the process. At each time $i = 1, 2, \ldots$, the monitoring procedure involves computing $\bar{\mathbf{x}}_i$, $\bar{\mathbf{x}}_i^*$, $\mathbf{w}_i^*$ using Equation 4.2 recursively, Q_i^* using Equation 4.1 and an out-of-control signal is triggered at time i if $Q_i^* > H > 0$.

Lowry et al. (1992b) proved that the run length performance of MEWMA charts depends on the non-centrality parameter δ. When there is no measurement error the value of δ is equal to

$$\delta = (\boldsymbol{\mu}_1^* - \boldsymbol{\mu}_0^*)(\boldsymbol{\Sigma}^*)^{-1}(\boldsymbol{\mu}_1^* - \boldsymbol{\mu}_0^*)^\intercal.$$

When there are measurement errors present in the process, the non centrality parameter will be denoted by δ_{M} and it is equal to

$$\delta_{\mathrm{M}} = b^2(\boldsymbol{\mu}_1^* - \boldsymbol{\mu}_0^*)\left(b^2\boldsymbol{\Sigma}^* + \frac{1}{m}\boldsymbol{\Sigma}_{\mathrm{M}}^*\right)^{-1}(\boldsymbol{\mu}_1^* - \boldsymbol{\mu}_0^*)^\intercal. \tag{4.6}$$

As it can be noted, the value of δ_{M} only depends on b but not on $\mathbf{a}^*$. Linna et al. (2001) explained that, in presence of measurement errors, multivariate control charts are unfortunately not equally powerful in detecting shifts δ, in all possible directions. This means that, in presence of measurement errors, for a fixed shift δ, they are an infinite number of possible shifts δ_{M} and, as it is impossible to investigate them all, Linna et al. (2001) suggested to only focus on two extreme values $\delta_{\min}$ (best case) and $\delta_{\max}$ (worst case) of δ_{M}. These values can be obtained using the results given on pages 35-36 of chapter 3 as,

$$\begin{aligned}\delta_{\min} &= \delta\lambda_1,\\ \delta_{\max} &= \delta\lambda_{p-1},\end{aligned}$$

where λ_1 and λ_{p-1} are the smallest and the largest eigenvalues of the $(p-1,p-1)$ matrix $b^2\boldsymbol{\Sigma}^*(b^2\boldsymbol{\Sigma}^* + \frac{1}{m}\boldsymbol{\Sigma}^*_{\mathrm{M}})^{-1}$, respectively.

The ARL of the MEWMA-CoDa control chart (with or without measurement errors) can be obtained using a dedicated Markov chain method. The implementation of this Markov chain has been taken from TRAN et al. (2018) and is explained below.

As explained in Lowry et al. (1992b), the run length performance of the MEWMA-$\bar{X}$ chart is a function of the sample size n, the in-control mean vector $\boldsymbol{\mu}_0^*$, the out-of-control mean vector $\boldsymbol{\mu}_1^*$ and the variance-covariance matrix $\boldsymbol{\Sigma}^*$ only through the non centrality parameter δ defined as

$$\delta = \sqrt{n(\boldsymbol{\mu}_1^* - \boldsymbol{\mu}_0^*)^\intercal(\boldsymbol{\Sigma}^*)^{-1}(\boldsymbol{\mu}_1^* - \boldsymbol{\mu}_0^*)}.$$

Without loss of generality, TRAN et al. (2018) restrict the numerical study to $n = 1$, $\boldsymbol{\mu}_0^* = \mathbf{0}$ (i.e. the in-control mean is centered to the homogeneous composition $\boldsymbol{\mu}_0 = (\frac{1}{p}, \frac{1}{p}, \ldots, \frac{1}{p})$) and $\boldsymbol{\Sigma}^* = \mathbf{I}_{p-1}$ (identity covariance matrix in $\mathbb{R}^{p-1}$). In this *standardized* case, Q_i in Equation 4.1 on the preceding page reduces to $Q_i = b||Y_i||_2^2$ with $b = \frac{2-r}{r}$ and the parameter δ reduces to $\delta = ||\boldsymbol{\mu}_1^*||_2 = ||\,\mathrm{ilr}(\boldsymbol{\mu}_1)||_2$. In order to evaluate the in- and out-of-control run length distributions of the MEWMA-$\bar{X}$ chart, Runger and Prabhu (1996) suggested to approximate its calculation for the statistic $q_i = ||Y_i||_2$ by using the following Markov chain models.

Concerning the *in-control* case, Runger and Prabhu (1996) showed that the run length distribution of q_i can be approximated by using a one dimensional Markov chain in which the interval $[0, \mathrm{UCL}']$, where $\mathrm{UCL}' = (H/b)^{1/2}$, is divided into $m+1$ subintervals/states: the first subinterval/state has length $\frac{g}{2}$ and the others have length g, where the width $g = \frac{2\mathrm{UCL}'}{2m+1}$. Concerning the MEWMA-CoDa control chart, the elements $p(i,j)$ (probability of transition from state i to state j) of the $(m+1, m+1)$ transition matrix $\mathbf{P}_1$ corresponding to the transient states are equal to:

— for $i = 0, 1, \ldots, m$ and $j = 1, 2, \ldots, m$,

$$p(i,j) = P\left(\left(\frac{(j-0.5)g}{r}\right)^2 < \chi^2(p-1,c) < \left(\frac{(j+0.5)g}{r}\right)^2\right)$$

where $\chi^2(p-1,c)$ is a non central chi-square random variable with $p-1$ degrees of freedom and noncentrality parameter $c = \left(\frac{(1-r)ig}{r}\right)^2$ which depends on state i,

— for $j = 0$,

$$p(i,0) = P\left(\chi^2(p-1,c) < \left(\frac{g}{2r}\right)^2\right).$$

The zero-state in-control ARL of the MEWMA-CoDa control chart is equal to

$$\mathrm{ARL} = \mathbf{s}^\intercal(\mathbf{I}_{m+1} - \mathbf{P}_1)^{-1}\mathbf{1}_{m+1},$$

where $\mathbf{1}_{m+1} = (1, 1, \ldots, 1)^\intercal$ is the $m+1$ column vector of 1's and $\mathbf{s}$ is the starting probability vector, i.e. $\mathbf{s} = (1, 0, 0, \ldots, 0)^\intercal$.

Concerning the *out-of-control* case, Runger and Prabhu (1996) proposed to approximate the run length distribution of q_i using a two dimensional Markov chain with $\mathbf{Y}_i \in \mathbb{R}^{p-1}$ partitioned into $Y_{i1} \in \mathbb{R}$ with nonzero mean δ and $\mathbf{Y}_{i2} \in \mathbb{R}^{p-2}$ with zero mean. Then $q_i = ||\mathbf{Y}_i||_2 = (Y_{i1}^2 + \mathbf{Y}_{i2}^\intercal \mathbf{Y}_{i2})^{\frac{1}{2}}$. Like in Runger and Prabhu (1996), they also use for the MEWMA-CoDa chart a two-dimensional Markov chain for Y_{i1}^2 and $||\mathbf{Y}_{i2}||_2$. The component Y_{i1}^2 can be approximated using the Markov chain based approach proposed by Lucas and Saccucci (1990). The values of $||\mathbf{Y}_{i2}||_2$ can be approximated by the same method as that used for the in-control case, described above, with $p-2$ replacing $p-1$, see Runger and Prabhu (1996).

Concerning Y_{i1}, the transition probability $h(i, j)$ from state i to state j is used to analyze the out-of-control component. In this case, the number of states of the Markov chain is chosen to be equal to $2m_1 + 1$. Consequently, for $i = 1, 2, \ldots, 2m_1 + 1$ and $j = 1, 2, \ldots, 2m_1 + 1$ we have

$$h(i,j) = \Phi\left(\frac{-\mathrm{UCL}' + jg_1 - (1-r)c_i}{r} - \delta\right) - \Phi\left(\frac{-\mathrm{UCL}' + (j-1)g_1 - (1-r)c_i}{r} - \delta\right)$$

where Φ represents the cumulative standard normal distribution function and where $c_i = -\mathrm{UCL}' + (i - 0.5)g_1$ is the centerpoint of state i with $g_1 = \frac{2\mathrm{UCL}'}{2m_1+1}$ as the width of each state.

Concerning $\mathbf{Y}_{i2}$, the transition probability $v(i, j)$ from state i to state j is used to analyze the in-control component. In this case, the control region is partitioned into $m_2 + 1$ transient states. The transition probability $v(i, j)$ has the same expression as in the in-control case, except that $p-2$ replaces $p-1$, i.e.

— for $i = 0, 1, 2, \ldots, m_2$ and $j = 1, 2, \ldots, m_2$, we have

$$v(i,j) = P\left(\left(\frac{(j-0.5)g_2}{r}\right)^2 < \chi^2(p-2, c) < \left(\frac{(j+0.5)g_2}{r}\right)^2\right)$$

where $c = \left(\frac{(1-r)ig_2}{r}\right)^2$ and the width of the states is $g_2 = \frac{2\mathrm{UCL}'}{2m_2+1}$.

— for $j = 0$,

$$v(i,0) = P\left(\chi^2(p-2, c) < \left(\frac{g_2}{2r}\right)^2\right)$$

Let $\mathbf{P}_2$ be the transition probability matrix of the two dimensional Markov chain. Since Y_{i1} is independent of $\mathbf{Y}_{i2}$, we have $\mathbf{P}_2 = \mathbf{H} \otimes \mathbf{V}$, where $\mathbf{H}$ is the $(2m_1 + 1, 2m_1 + 1)$ transition probability matrix of Y_{i1} with elements $h(i, j)$, $\mathbf{V}$ is the $(m_2 + 1, m_2 + 1)$ transition probability matrix of $||\mathbf{Y}_{i2}||_2$ with elements $v(i, j)$ and $\otimes$ denotes the Kronecker's matrices product. The transition probability matrix $\mathbf{P}_2$ consists of the transition probabilities between all transient and some absorbing states of the Markov chain.

Let $\mathbf{T}$ be the $(2m_1 + 1) \times (m_2 + 1)$ dimensional matrix defined as

$$\mathbf{T}(\alpha, \beta) = \begin{cases} 1 & \text{if state } (\alpha, \beta) \text{ is transient} \\ 0 & \text{otherwise} \end{cases}$$

Let $\mathbf{P}$ be the transition probability matrix that contains all the transient states of the Markov chain. Then we have $\mathbf{P} = \mathbf{T}(\alpha, \beta) \circledast \mathbf{P}_2$ where the symbol $\circledast$ indicates the elementwise matrices multiplication.

Finally, the out-of-control ARL of the MEWMA-CoDa control chart is defined as $\mathrm{ARL} = \mathbf{s}^\intercal(\mathbf{I} - \mathbf{P})^{-1}\mathbf{1}$ where $\mathbf{s}$ is the initial probability vector with all components equal to zero except the component corresponding to the state $(\alpha, \beta) = (m_1 + 1, 0)$ which is equal to one. In the particular case where $m_1 = m_2 = m$, the component of $\mathbf{s}$ which is equal to 1 is the $(m(m + 1) + 1)$th entry (see Lee and Khoo (2006)). Like in Molnau et al. (2001), the method used for the computation of the ARL of the MEWMA-CoDa control chart depends on the number of states m_1 and m_2 used in the approximation. More specifically, the larger the number of states, the more accurate is the ARL approximation. However, a large number of states will require more computing resources and time. Motivated by the performance of the program used for the computation of the ARL in Molnau et al. (2001), we have decided to use $m_1 = m_2 = 30$.

At time $i = 1, 2, \ldots$, the sample size is assumed to be $n = 1$ and we are only able to monitor shifts in the composition mean vector using the MEWMA approach. Now, if we relax this constraint and we allow $n > p$, then it becomes also possible to monitor (without the MEWMA feature) the composition *variability* using the approach derived from Gnanadesikan and Gupta (1970) who suggest to use the statistic W_i defined as

$$W_i = \frac{(n-1)|\hat{\mathbf{\Sigma}}_i^*|^{1/p}}{|b^2\mathbf{\Sigma}^* + \frac{1}{m}\mathbf{\Sigma}_{\mathrm{M}}^*|^{1/p}}.$$

This statistic is known to exactly follow a gamma distribution of parameters $a = \frac{p(n-p)}{2}$ and $b = \frac{2}{p}\left(1 - \frac{(p-1)(p-2)}{2n}\right)^{-1/p}$ when $p \leq 2$ and, to approximately follow this distribution, when $p > 2$. This allows to define an upper control limit for W_i as $\mathrm{UCL} = F_{\Gamma}^{-1}(1-\alpha_0|a,b)$ where α_0 is the type I error and $F_{\Gamma}^{-1}(\ldots|a,b)$ is the inverse cummulative distribution function of the gamma distribution of parameters a and b and, therefore, this approach allows to monitor the multivariate variability of Coda.

4.2 Performance of the MEWMA-CoDa control chart in the presence of measurement errors

The goal of this section is to investigate the performance of the MEWMA-CoDa control chart in the presence of measurement errors. The measurement error variance-covariance matrix is equal to $\mathbf{\Sigma}_{\mathrm{M}}^* = \sigma_{\mathrm{M}}^2\mathbf{I}$, where σ_{M} is the standard-deviation measurement error (common for all dimensions) and $\mathbf{I}$ is the $(2,2)$ identity matrix. The first step involved in the design of the MEWMA-CoDa control chart is to select the optimal couple (r, H) which minimizes the out-of-control ARL for a fixed value of the shift δ_{M} subject to a constrained value for the in-control ARL. The procedure to obtain the optimal couple is based on the following three steps:

— Select a particular value for the in-control ARL_0. In our case we select $\mathrm{ARL}_0 = 370$.
— Find the set of design couples (r, H) such that $\mathrm{ARL} = \mathrm{ARL}_0$ when there is no shift, i.e. when $\delta = \delta_{\mathrm{M}} = 0$.
— For a particular value of the shift δ for $\boldsymbol{\mu}^*$, select the optimal couple (r^*, H^*) from the set of designed couples such that the value of the out-of-control ARL is minimum (in order to obtain the best statistical performance).

If the value of the smoothing parameter r is too small, the Markov Chain approach is known to diverge leading to unreliable results (see for instance Castagliola et al. (2011) and Tran et al. (2015)). For this reason, we have chosen to constrain the value of $r \geq 0.05$. Smaller values for r can actually be used if the number of states (h_1 and h_2, see TRAN et al. (2018)) of the Markov chain is large enough but, in this case, it will require more time to obtain the optimal couples.

The following four situations for the CoDa variance-covariance matrix $\mathbf{\Sigma}^*$ are considered,

Case #1 uncorrelated case with equal variances

$$\mathbf{\Sigma}^* = \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix},$$

Case #2 negatively correlated case with equal variances

$$\boldsymbol{\Sigma}^* = \begin{pmatrix} 3 & -1/2 \\ -1/2 & 3 \end{pmatrix}.$$

Case #3 uncorrelated case with unequal variances

$$\boldsymbol{\Sigma}^* = \begin{pmatrix} 1.5 & 0 \\ 0 & 3 \end{pmatrix},$$

Case #4 positively correlated case with unequal variances

$$\boldsymbol{\Sigma}^* = \begin{pmatrix} 1.5 & 1/2 \\ 1/2 & 3 \end{pmatrix}.$$

We will now separately investigate the influence of parameters σ_{M}, b, m and p on the performance of the MEWMA-CoDa control chart in the presence of measurement errors.

4.2.1 Influence of parameter σ_{M}

This subsection is further divided into two parts, the first one studies the effect of σ_{M} on the optimal values of r and H and, the second one studies the effect of σ_{M} on the ARL. We have used values of $\sigma_{\mathrm{M}} \in \{0.1, 0.3, 0.6\}$ and fixed values $b = 1$, $m = 3$ and $p = 3$. For different values of the shift $\delta \in [0, 2]$, the values of δ_{M} along with the optimal values of ARL, r and H are listed in Table 4.1 on page 58 for the MEWMA-CoDa control chart with ($\min$ and $\max$) and without the presence of measurement errors. The first rows in Table 4.1 on page 58 shows the values for Case #1, i.e. when the variance-covariance matrix $\boldsymbol{\Sigma}^*$ is uncorrelated with equal variances and the second row represents the Case #2, i.e. when the variance-covariance matrix $\boldsymbol{\Sigma}^*$ is negatively correlated with equal variances. The third row represents the Case #3, i.e. when the variance-covariance matrix $\boldsymbol{\Sigma}^*$ is uncorrelated with unequal variances and the fourth row represents the Case #4, i.e. when the variance-covariance matrix $\boldsymbol{\Sigma}^*$ is positively correlated with unequal variances. From Table 4.1 on page 58, we can conclude that:

— All the out-of-control ARL values of the "best" $\mathrm{ARL}_{\min}$ are larger for Case #4 rather than for all other cases while, the out-of-control ARL values of the "worst" $\mathrm{ARL}_{\max}$ are larger for Case #2 rather than for all other cases. For instance, when $\sigma_{\mathrm{M}} = 0.1$ and $\delta = 1.75$ the "best" and "worst" ARL values are $(\mathrm{ARL}_{\min} = 4.881, \mathrm{ARL}_{\max} = 4.898)$ for Case #1 while, for Case #2, they are equal to $(\mathrm{ARL}_{\min} = 4.873, \mathrm{ARL}_{\max} = 4.912)$. For Case #3, they are equal to $(\mathrm{ARL}_{\min} = 4.962, \mathrm{ARL}_{\max} = 4.878)$ while, for Case #4, they are equal to $(\mathrm{ARL}_{\min} = 4.981, \mathrm{ARL}_{\max} = 4.874)$.

— The out-of-control ARL values also increase with the increase in the value of σ_{M}. For instance, when $\sigma_{\mathrm{M}} = 0.3$ and $\delta = 1.75$ the "best" and "worst" ARL values are $(\mathrm{ARL}_{\min} = 5.049, \mathrm{ARL}_{\max} = 5.099)$ for Case #1, $(\mathrm{ARL}_{\min} = 5.025, \mathrm{ARL}_{\max} = 5.142)$ for Case #2, $(\mathrm{ARL}_{\min} = 5.301, \mathrm{ARL}_{\max} = 5.046)$ and for Case #4 they are equal to $(\mathrm{ARL}_{\min} = 5.359, \mathrm{ARL}_{\max} = 5.034)$ But, when $\sigma_{\mathrm{M}} = 0.6$ (i.e. σ_{M} increased), the "best" and "worst" ARL values are $(\mathrm{ARL}_{\min} = 5.304, \mathrm{ARL}_{\max} = 5.407)$ for Case #1 , $(\mathrm{ARL}_{\min} = 5.256, \mathrm{ARL}_{\max} = 5.496)$ for Case #2 , $(\mathrm{ARL}_{\min} = 5.825, \mathrm{ARL}_{\max} = 5.301)$ for Case #3 and $(\mathrm{ARL}_{\min} = 5.945, \mathrm{ARL}_{\max} = 5.276)$ for Case #4.

— When σ_{M} is small the values of the optimal couples corresponding to the maximum and minimum values of δ_M are the same in first two Cases. For instance, when $\sigma_{\mathrm{M}} = 0.1$ and $\delta = 1.75$ the maximum and minimum values of the optimal couples (r, H) are $(r_{\max} = 0.276, H_{\max} = 11.336)$ and $(r_{\min} = 0.278, H_{\min} = 11.344)$ for the first two cases, while they are equal to $(r_{\max} = 0.276, H_{\max} = 11.33)$ and $(r_{\min} = 0.273, H_{\min} = 11.322)$ for Case #3 and $(r_{\max} = 0.276, H_{\max} = 11.33)$ and $(r_{\min} = 0.271, H_{\min} = 11.314)$ for Case #4. But, when $\sigma_{\mathrm{M}} = 0.6$ (increase) the values of the optimal couples for Case #1 are $(r_{\min} = 0.257, H_{\min} = 11.273)$ and $(r_{\max} = 0.252, H_{\max} = 11.256)$ while, for case #2 their values are $(r_{\min} = 0.259, H_{\min} = 11.281)$ and

$(r_{\max} = 0.25, H_{\max} = 11.274)$, for Case #3 they are equal to $(r_{\min} = 0.235, H_{\min} = 11.182)$ and $(r_{\max} = 0.257, H_{\max} = 11.265)$ and for Case #4 the are equal to $(r_{\min} = 0.231, H_{\min} = 11.161)$ and $(r_{\max} = 0.257, H_{\max} = 11.265)$. So we can conclude that the values of the optimal couple (r, H) decreases with an increase in σ_{M}.

4.2.2 Influence of parameter b

In this subsection, the influence of parameter $b \in \{0.25, 1, 8\}$ for fixed values $\sigma_{\mathrm{M}} = 0.3$, $m = 3$ and $p = 3$ is investigated. Similar to Table 4.1 on the next page, we listed in Table 4.2 on page 61 values for δ_{M}, ARL, r and H. Based on this table, it is concluded that b has a *positive* impact on the MEWMA-CoDa control chart, that is the out-of-control ARL values decrease with an increase in the value of b. For instance, when $\delta = 1.75$ and $b = 1$, the "best" and "worst" ARL values are $(\mathrm{ARL}_{\min} = 5.049, \mathrm{ARL}_{\max} = 5.099)$ for Case #1, $(\mathrm{ARL}_{\min} = 5.025, \mathrm{ARL}_{\max} = 5.142)$ for Case #2, $(\mathrm{ARL}_{\min} = 5.301, \mathrm{ARL}_{\max} = 5.046)$ for Case #3 and $(\mathrm{ARL}_{\min} = 5.359, \mathrm{ARL}_{\max} = 5.034)$ for Case #4 . But, when $b = 8$ (increase), the "best" and "worst" ARL values are $(\mathrm{ARL}_{\min} = 4.802, \mathrm{ARL}_{\max} = 4.801)$ for Case #1, $(\mathrm{ARL}_{\min} = 4.801, \mathrm{ARL}_{\max} = 4.803)$ for Case #2, $(\mathrm{ARL}_{\min} = 4.803, \mathrm{ARL}_{\max} = 4.799)$ for Case #3 and $(\mathrm{ARL}_{\min} = 4.804, \mathrm{ARL}_{\max} = 4.799)$ for Case #4.

The value of the optimal couples (r, H) increases with an increase in the value of b in all cases. For example, when $\delta = 1.75$ and $b = 0.25$ the maximum and minimum values for the optimal couple (r, H) are $(r_{\max} = 0.138, H_{\max} = 10.563)$ and $(r_{\min} = 0.152, H_{\min} = 10.693)$ for Case #1, while we have $(r_{\max} = 0.128, H_{\max} = 10.465)$ and $(r_{\min} = 0.159, H_{\min} = 10.752)$ for Case #2, we have $(r_{\max} = 0.152, H_{\max} = 10.679)$ and $(r_{\min} = 0.098, H_{\min} = 10.036)$ and for Case #3 and we have $(r_{\max} = 0.155, H_{\max} = 10.699)$ and $(r_{\min} = 0.090, H_{\min} = 9.914)$ and for Case #4. But when $b = 8$ (increase) the values for the optimal couple for minimum and maximum values of δ_M also increases, i.e. $(r_{\min} = r_{\max} = 0.280, H_{\min} = H_{\max} = 11.351)$ for Case #1 and Case #2, while $(r_{\min} = r_{\max} = 0.280, H_{\min} = H_{\max} = 11.344)$ for Case #3 and Case #4.

4.2.3 Influence of parameter m

In this subsection, we investigate the influence of the parameter $m \in \{1, 3, 6\}$ for fixed values $\sigma_{\mathrm{M}} = 0.3$, $b = 1$ and $p = 3$. Similar to Tables 4.1 and 4.2, we listed in Table 4.3 on page 65 the ARL values along with the optimal couples (r, H) for shifts $\delta \in [0, 2]$. From this table we can see that when m increases the out-of-control ARL slowly decreases (i.e. an increase in m has a *positive* impact on the MEWMA-CoDa control chart). For instance, if $\delta = 1.75$ and $m = 1$, the "best" and "worst" ARL values are $(\mathrm{ARL}_{\min} = 5.564, \mathrm{ARL}_{\max} = 5.723)$ for Case #1 , $(\mathrm{ARL}_{\min} = 5.491, \mathrm{ARL}_{\max} = 5.858)$ for Case #2 , $(\mathrm{ARL}_{\min} = 6.367, \mathrm{ARL}_{\max} = 5.561)$ for Case #3 and $(\mathrm{ARL}_{\min} = 6.553, \mathrm{ARL}_{\max} = 5.523)$ for Case #4. But, when $m = 6$ (increase), the "best" and "worst" ARL values are $(\mathrm{ARL}_{\min} = 4.923, \mathrm{ARL}_{\max} = 4.948)$ for Case #1 , $(\mathrm{ARL}_{\min} = 4.911, \mathrm{ARL}_{\max} = 4.969)$ for Case #2 , $(\mathrm{ARL}_{\min} = 5.046, \mathrm{ARL}_{\max} = 4.920)$ for Case #3 and $(\mathrm{ARL}_{\min} = 5.074, \mathrm{ARL}_{\max} = 4.914)$ for Case #4.

The value of the optimal couples (r, H) increases with an increase in the value of m for all the cases. When $\delta = 1.75$ and $m = 1$, the maximum and minimum values of the optimal couples (r, H) are $(r_{\max} = 0.240, H_{\max} = 11.210)$ and $(r_{\min} = 0.245, H_{\min} = 11.229)$ for Case #1, while they are $(r_{\max} = 0.235, H_{\max} = 11.190)$ and $(r_{\min} = 0.250, H_{\min} = 11.247)$ for Case #2. For Case #3, they are equal to $(r_{\max} = 0.245, H_{\max} = 11.221)$ and $(r_{\min} = 0.216, H_{\min} = 11.095)$ and, for Case #4, they are $(r_{\max} = 0.247, H_{\max} = 11.230)$ and $(r_{\min} = 0.212, H_{\min} = 11.072)$.

But when $m = 6$ (increase) the values of the optimal couples for minimum and maximum values of δ_M also increase, i.e. $(r_{\max} = 0.273, H_{\max} = 11.329)$ and $(r_{\min} = 0.276, H_{\min} = 11.337)$ for the first two

Table 4.1 – Influence of parameter σ_{M}

$\sigma_{\mathrm{M}} = 0.1$

	δ_{M}		ARL			r			H		
δ	min	max	without	min	max	without	min	max	without	min	max
0	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
0.25	0.247	0.247	91.212	92.682	92.976	0.050	0.050	0.050	8.895	8.895	8.895
	0.246	0.247	91.212	92.625	92.976	0.050	0.050	0.050	8.895	8.895	8.895
	0.245	0.247	91.212	93.462	92.007	0.050	0.050	0.050	8.895	8.895	8.895
	0.244	0.247	91.212	93.789	91.937	0.050	0.050	0.050	8.895	8.895	8.895
0.5	0.495	0.493	32.162	32.686	32.792	0.050	0.050	0.050	8.895	8.895	8.895
	0.495	0.493	32.162	32.637	32.881	0.050	0.050	0.050	8.895	8.895	8.895
	0.489	0.495	32.162	33.077	32.552	0.050	0.050	0.050	8.895	8.895	8.895
	0.488	0.495	32.162	33.196	32.527	0.050	0.050	0.050	8.895	8.895	8.895
0.75	0.742	0.740	17.720	18.016	18.075	0.083	0.081	0.081	9.802	9.754	9.754
	0.743	0.739	17.720	17.988	18.126	0.083	0.081	0.081	9.802	9.754	9.754
	0.734	0.742	17.720	18.268	17.972	0.083	0.079	0.081	9.802	9.679	9.730
	0.732	0.742	17.720	18.335	17.957	0.083	0.079	0.081	9.802	9.679	9.730
1	0.989	0.987	11.437	11.633	11.672	0.126	0.124	0.124	10.438	10.412	10.412
	0.990	0.985	11.437	11.614	11.705	0.126	0.124	0.124	10.438	10.412	10.412
	0.978	0.989	11.437	11.811	11.614	0.126	0.121	0.124	10.438	10.366	10.394
	0.976	0.990	11.437	11.855	11.605	0.126	0.121	0.124	10.438	10.366	10.394
1.25	1.236	1.234	8.099	8.239	8.267	0.174	0.171	0.171	10.857	10.841	10.841
	1.238	1.231	8.099	8.226	8.291	0.174	0.171	0.171	10.857	10.841	10.841
	1.223	1.236	8.099	8.371	8.230	0.174	0.169	0.171	10.857	10.811	10.828
	1.220	1.237	8.099	8.403	8.224	0.174	0.166	0.171	10.857	10.793	10.828
1.5	1.484	1.480	6.098	6.204	6.225	0.226	0.223	0.221	11.149	11.138	11.128
	1.485	1.478	6.098	6.194	6.243	0.226	0.223	0.221	11.149	11.138	11.128
	1.467	1.484	6.098	6.306	6.200	0.226	0.219	0.223	11.149	11.107	11.129
	1.464	1.484	6.098	6.330	6.195	0.226	0.219	0.223	11.149	11.107	11.129
1.75	1.731	1.727	4.798	4.881	4.898	0.280	0.278	0.276	11.351	11.344	11.337
	1.733	1.724	4.798	4.873	4.912	0.280	0.278	0.276	11.351	11.344	11.337
	1.712	1.731	4.798	4.962	4.878	0.280	0.273	0.276	11.351	11.322	11.330
	1.708	1.732	4.798	4.981	4.874	0.280	0.271	0.276	11.351	11.314	11.330
2	1.978	1.974	3.902	3.969	3.983	0.337	0.333	0.333	11.495	11.485	11.485
	1.980	1.970	3.902	3.963	3.994	0.337	0.333	0.330	11.495	11.485	11.480
	1.957	1.978	3.902	4.035	3.968	0.337	0.328	0.333	11.495	11.469	11.480
	1.952	1.979	3.902	4.050	3.964	0.337	0.326	0.333	11.495	11.464	11.480

$\sigma_M = 0.3$

δ	δ_M min	δ_M max	ARL without	ARL min	ARL max	r without	r min	r max	H without	H min	H max
0	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
0.25	0.242	0.240	91.212	95.626	96.510	0.050	0.050	0.050	8.895	8.895	8.895
	0.243	0.239	91.212	95.213	97.258	0.050	0.050	0.050	8.895	8.895	8.895
	0.234	0.242	91.212	99.291	94.918	0.050	0.050	0.050	8.895	8.895	8.895
	0.233	0.242	91.212	100.273	94.708	0.050	0.050	0.050	8.895	8.895	8.895
0.5	0.484	0.481	32.162	33.747	34.068	0.050	0.050	0.050	8.895	8.895	8.895
	0.485	0.478	32.162	33.598	34.341	0.050	0.050	0.050	8.895	8.895	8.895
	0.469	0.484	32.162	35.216	33.606	0.050	0.050	0.050	8.895	8.895	8.895
	0.465	0.485	32.162	35.582	33.530	0.050	0.050	0.050	8.895	8.895	8.895
0.75	0.726	0.721	17.720	18.613	18.793	0.083	0.079	0.079	9.802	9.704	9.704
	0.728	0.717	17.720	18.529	18.946	0.083	0.079	0.076	9.802	9.704	9.652
	0.703	0.726	17.720	19.467	18.566	0.083	0.074	0.079	9.802	9.572	9.679
	0.698	0.727	17.720	19.670	18.523	0.083	0.074	0.079	9.802	9.572	9.679
1	0.968	0.962	11.437	12.028	12.147	0.126	0.119	0.119	10.438	10.355	10.355
	0.971	0.956	11.437	11.972	12.248	0.126	0.121	0.117	10.438	10.384	10.326
	0.938	0.968	11.437	12.606	12.008	0.126	0.114	0.119	10.438	10.277	10.338
	0.931	0.969	11.437	12.742	11.980	0.126	0.112	0.119	10.438	10.246	10.338
1.25	1.210	1.202	8.099	8.522	8.607	0.174	0.166	0.164	10.857	10.806	10.788
	1.213	1.195	8.099	8.482	8.680	0.174	0.166	0.162	10.857	10.806	10.770
	1.172	1.210	8.099	8.941	8.512	0.174	0.157	0.166	10.857	10.719	10.793
	1.164	1.212	8.099	9.039	8.492	0.174	0.157	0.166	10.857	10.719	10.793
1.5	1.452	1.442	6.098	6.418	6.482	0.226	0.216	0.214	11.149	11.105	11.093
	1.456	1.435	6.098	6.388	6.537	0.226	0.216	0.212	11.149	11.105	11.082
	1.406	1.452	6.098	6.738	6.413	0.226	0.207	0.216	11.149	11.047	11.095
	1.396	1.454	6.098	6.811	6.397	0.226	0.204	0.216	11.149	11.035	11.095
1.75	1.694	1.683	4.798	5.049	5.099	0.280	0.269	0.266	11.351	11.314	11.306
	1.699	1.674	4.798	5.025	5.142	0.280	0.271	0.264	11.351	11.321	11.298
	1.641	1.694	4.798	5.301	5.046	0.280	0.257	0.269	11.351	11.265	11.307
	1.629	1.696	4.798	5.359	5.034	0.280	0.254	0.269	11.351	11.257	11.307
2	1.935	1.923	3.902	4.105	4.146	0.337	0.323	0.321	11.495	11.464	11.458
	1.941	1.913	3.902	4.086	4.180	0.337	0.323	0.318	11.495	11.464	11.453
	1.875	1.935	3.902	4.309	4.103	0.337	0.309	0.323	11.495	11.424	11.458
	1.862	1.938	3.902	4.356	4.093	0.337	0.307	0.323	11.495	11.418	11.458

	$\sigma_{\mathrm{M}} = 0.6$										
	δ_{M}		ARL			r			H		
δ	min	max	without	min	max	without	min	max	without	min	max
0	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
0.25	0.234	0.231	91.212	100.049	101.819	0.050	0.050	0.050	8.895	8.895	8.895
	0.236	0.229	91.212	99.222	103.317	0.050	0.050	0.050	8.895	8.895	8.895
	0.221	0.234	91.212	108.032	99.291	0.050	0.050	0.050	8.895	8.895	8.895
	0.218	0.235	91.212	109.989	98.871	0.050	0.050	0.050	8.895	8.895	8.895
0.5	0.469	0.463	32.162	35.368	36.026	0.050	0.050	0.050	8.895	8.895	8.895
	0.472	0.458	32.162	35.062	36.587	0.050	0.050	0.050	8.895	8.895	8.895
	0.441	0.469	32.162	38.537	35.216	0.050	0.050	0.050	8.895	8.895	8.895
	0.435	0.470	32.162	39.301	35.060	0.050	0.050	0.050	8.895	8.895	8.895
0.75	0.703	0.694	17.720	19.518	19.883	0.083	0.074	0.074	9.802	9.598	9.598
	0.707	0.687	17.720	19.348	20.194	0.083	0.076	0.071	9.802	9.652	9.542
	0.662	0.703	17.720	21.299	19.467	0.083	0.069	0.074	9.802	9.455	9.572
	0.653	0.705	17.720	21.716	19.380	0.083	0.067	0.074	9.802	9.393	9.572
1	0.938	0.926	11.437	12.628	12.870	0.126	0.114	0.112	10.438	10.296	10.265
	0.943	0.916	11.437	12.515	13.076	0.126	0.114	0.109	10.438	10.296	10.233
	0.882	0.938	11.437	13.828	12.606	0.126	0.105	0.114	10.438	10.146	10.277
	0.871	0.940	11.437	14.106	12.548	0.126	0.102	0.114	10.438	10.110	10.277
1.25	1.172	1.157	8.099	8.952	9.126	0.174	0.157	0.155	10.857	10.733	10.713
	1.179	1.145	8.099	8.871	9.274	0.174	0.159	0.152	10.857	10.752	10.693
	1.103	1.172	8.099	9.820	8.941	0.174	0.145	0.157	10.857	10.616	10.719
	1.089	1.175	8.099	10.021	8.900	0.174	0.143	0.159	10.857	10.594	10.738
1.5	1.406	1.389	6.098	6.743	6.875	0.226	0.207	0.202	11.149	11.057	11.032
	1.415	1.375	6.098	6.682	6.987	0.226	0.207	0.200	11.149	11.057	11.019
	1.324	1.406	6.098	7.403	6.738	0.226	0.188	0.207	11.149	10.939	11.047
	1.306	1.410	6.098	7.556	6.706	0.226	0.185	0.207	11.149	10.924	11.047
1.75	1.641	1.620	4.798	5.304	5.408	0.280	0.257	0.252	11.351	11.273	11.256
	1.650	1.604	4.798	5.256	5.496	0.280	0.259	0.250	11.351	11.281	11.247
	1.544	1.641	4.798	5.825	5.301	0.280	0.235	0.257	11.351	11.182	11.265
	1.524	1.646	4.798	5.945	5.276	0.280	0.231	0.257	11.351	11.161	11.265
2	1.875	1.852	3.902	4.311	4.395	0.337	0.309	0.304	11.495	11.430	11.418
	1.886	1.833	3.902	4.272	4.466	0.337	0.311	0.299	11.495	11.435	11.405
	1.765	1.875	3.902	4.733	4.309	0.337	0.285	0.309	11.495	11.359	11.424
	1.742	1.881	3.902	4.831	4.289	0.337	0.278	0.311	11.495	1.337	11.430

Table 4.2 – Influence of parameter b

$b = 0.25$

	δ_{M}		ARL			r			H		
δ	min	max	without	min	max	without	min	max	without	min	max
0	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
0.25	0.163	0.152	91.212	159.506	171.770	0.050	0.050	0.050	8.895	8.895	8.895
	0.169	0.144	91.212	153.570	181.647	0.050	0.050	0.050	8.895	8.895	8.895
	0.121	0.163	91.212	211.959	157.991	0.050	0.050	0.050	8.895	8.895	8.895
	0.114	0.166	91.212	221.875	154.996	0.050	0.050	0.050	8.895	8.895	8.895
0.5	0.326	0.305	32.162	61.295	67.898	0.050	0.050	0.050	8.895	8.895	8.895
	0.337	0.289	32.162	58.281	73.620	0.050	0.050	0.050	8.895	8.895	8.895
	0.242	0.326	32.162	94.918	60.937	0.050	0.050	0.050	8.895	8.895	8.895
	0.229	0.332	32.162	102.773	59.395	0.050	0.050	0.050	8.895	8.895	8.895
0.75	0.489	0.457	17.720	33.215	36.689	0.083	0.050	0.050	9.802	8.894	8.894
	0.506	0.433	17.720	31.647	39.750	0.083	0.050	0.050	9.802	8.894	8.894
	0.363	0.489	17.720	51.775	33.077	0.083	0.050	0.050	9.802	8.894	8.894
	0.343	0.497	17.720	56.415	32.272	0.083	0.050	0.050	9.802	8.894	8.894
1	0.652	0.610	11.437	21.827	24.103	0.126	0.067	0.062	10.438	9.421	9.289
	0.674	0.578	11.437	20.785	26.067	0.126	0.071	0.057	10.438	9.542	9.144
	0.484	0.652	11.437	33.606	21.763	0.126	0.050	0.067	10.438	8.894	9.393
	0.457	0.663	11.437	36.522	21.228	0.126	0.050	0.069	10.438	8.894	9.455
1.25	0.815	0.762	8.099	15.624	17.295	0.174	0.093	0.083	10.857	9.978	9.802
	0.843	0.722	8.099	14.861	18.743	0.174	0.098	0.079	10.857	10.057	9.704
	0.605	0.815	8.099	24.311	15.591	0.174	0.060	0.093	10.857	9.188	9.956
	0.572	0.829	8.099	26.389	15.199	0.174	0.055	0.095	10.857	9.033	9.996
1.5	0.978	0.915	6.098	11.830	13.114	0.226	0.121	0.109	11.149	10.384	10.233
	1.011	0.867	6.098	11.244	14.231	0.226	0.128	0.102	11.149	10.465	10.130
	0.726	0.978	6.098	18.566	11.811	0.226	0.079	0.121	11.149	9.679	10.360
	0.686	0.995	6.098	20.191	11.509	0.226	0.071	0.124	11.149	9.515	10.394
1.75	1.141	1.067	4.798	9.326	10.348	0.280	0.152	0.138	11.351	10.693	10.563
	1.180	1.011	4.798	8.861	11.239	0.280	0.159	0.128	11.351	10.752	10.465
	0.847	1.141	4.798	14.720	9.315	0.280	0.098	0.152	11.351	10.036	10.679
	0.800	1.161	4.798	16.029	9.075	0.280	0.090	0.155	11.351	9.914	10.699
2	1.304	1.220	3.902	7.580	8.415	0.337	0.185	0.169	11.495	10.935	10.824
	1.348	1.156	3.902	7.201	9.144	0.337	0.195	0.155	11.495	10.992	10.713
	0.968	1.304	3.902	12.008	7.573	0.337	0.119	0.185	11.495	10.338	10.924
	0.915	1.327	3.902	13.088	7.378	0.337	0.109	0.190	11.495	10.214	10.953

$b = 1$

	δ_M		ARL			r			H		
δ	min	max	without	min	max	without	min	max	without	min	max
0	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
0.25	0.242	0.240	91.212	95.626	96.510	0.050	0.050	0.050	8.895	8.895	8.895
	0.243	0.239	91.212	95.213	97.258	0.050	0.050	0.050	8.895	8.895	8.895
	0.234	0.242	91.212	99.291	94.918	0.050	0.050	0.050	8.895	8.895	8.895
	0.233	0.242	91.212	100.273	94.708	0.050	0.050	0.050	8.895	8.895	8.895
0.5	0.484	0.481	32.162	33.747	34.068	0.050	0.050	0.050	8.895	8.895	8.895
	0.485	0.478	32.162	33.598	34.341	0.050	0.050	0.050	8.895	8.895	8.895
	0.469	0.484	32.162	35.216	33.606	0.050	0.050	0.050	8.895	8.895	8.895
	0.465	0.485	32.162	35.582	33.530	0.050	0.050	0.050	8.895	8.895	8.895
0.75	0.726	0.721	17.720	18.613	18.793	0.083	0.079	0.079	9.802	9.704	9.704
	0.728	0.717	17.720	18.529	18.946	0.083	0.079	0.076	9.802	9.704	9.652
	0.703	0.726	17.720	19.467	18.566	0.083	0.074	0.079	9.802	9.572	9.679
	0.698	0.727	17.720	19.670	18.523	0.083	0.074	0.079	9.802	9.572	9.679
1	0.968	0.962	11.437	12.028	12.147	0.126	0.119	0.119	10.438	10.355	10.355
	0.971	0.956	11.437	11.972	12.248	0.126	0.121	0.117	10.438	10.384	10.326
	0.938	0.968	11.437	12.606	12.008	0.126	0.114	0.119	10.438	10.277	10.338
	0.931	0.969	11.437	12.742	11.980	0.126	0.112	0.119	10.438	10.246	10.338
1.25	1.210	1.202	8.099	8.522	8.607	0.174	0.166	0.164	10.857	10.806	10.788
	1.213	1.195	8.099	8.482	8.680	0.174	0.166	0.162	10.857	10.806	10.770
	1.172	1.210	8.099	8.941	8.512	0.174	0.157	0.166	10.857	10.719	10.793
	1.164	1.212	8.099	9.039	8.492	0.174	0.157	0.166	10.857	10.719	10.793
1.5	1.452	1.442	6.098	6.418	6.482	0.226	0.216	0.214	11.149	11.105	11.093
	1.456	1.435	6.098	6.388	6.537	0.226	0.216	0.212	11.149	11.105	11.082
	1.406	1.452	6.098	6.738	6.413	0.226	0.207	0.216	11.149	11.047	11.095
	1.396	1.454	6.098	6.811	6.397	0.226	0.204	0.216	11.149	11.035	11.095
1.75	1.694	1.683	4.798	5.049	5.099	0.280	0.269	0.266	11.351	11.314	11.306
	1.699	1.674	4.798	5.025	5.142	0.280	0.271	0.264	11.351	11.321	11.298
	1.641	1.694	4.798	5.301	5.046	0.280	0.257	0.269	11.351	11.265	11.307
	1.629	1.696	4.798	5.359	5.034	0.280	0.254	0.269	11.351	11.257	11.307
2	1.935	1.923	3.902	4.105	4.146	0.337	0.323	0.321	11.495	11.464	11.458
	1.941	1.913	3.902	4.086	4.180	0.337	0.323	0.318	11.495	11.464	11.453
	1.875	1.935	3.902	4.309	4.103	0.337	0.309	0.323	11.495	11.424	11.458
	1.862	1.938	3.902	4.356	4.093	0.337	0.307	0.323	11.495	11.418	11.458

$b = 8$

δ	δ_M min	δ_M max	ARL without	ARL min	ARL max	r without	r min	r max	H without	H min	H max
0	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
0.25	0.250	0.250	91.212	91.281	91.295	0.050	0.050	0.050	8.895	8.895	8.895
	0.250	0.250	91.212	91.275	91.307	0.050	0.050	0.050	8.895	8.895	8.895
	0.250	0.250	91.212	90.689	90.621	0.050	0.050	0.050	8.895	8.895	8.895
	0.250	0.250	91.212	90.705	90.618	0.050	0.050	0.050	8.895	8.895	8.895
0.5	0.500	0.500	32.162	32.186	32.191	0.050	0.050	0.050	8.895	8.895	8.895
	0.500	0.500	32.162	32.184	32.195	0.050	0.050	0.050	8.895	8.895	8.895
	0.499	0.500	32.162	32.080	32.055	0.050	0.050	0.050	8.895	8.895	8.895
	0.499	0.500	32.162	32.085	32.054	0.050	0.050	0.050	8.895	8.895	8.895
0.75	0.750	0.750	17.720	17.734	17.736	0.083	0.083	0.083	9.802	9.802	9.802
	0.750	0.749	17.720	17.732	17.739	0.083	0.083	0.083	9.802	9.802	9.802
	0.749	0.750	17.720	17.704	17.691	0.083	0.081	0.081	9.802	9.730	9.730
	0.749	0.750	17.720	17.708	17.690	0.083	0.081	0.081	9.802	9.730	9.730
1	0.999	0.999	11.437	11.446	11.448	0.126	0.126	0.126	10.438	10.438	10.438
	1.000	0.999	11.437	11.445	11.449	0.126	0.126	0.126	10.438	10.438	10.438
	0.999	0.999	11.437	11.438	11.428	0.126	0.126	0.126	10.438	10.422	10.422
	0.999	1.000	11.437	11.440	11.428	0.126	0.126	0.126	10.438	10.422	10.422
1.25	1.249	1.249	8.099	8.105	8.107	0.174	0.174	0.174	10.857	10.857	10.857
	1.249	1.249	8.099	8.105	8.108	0.174	0.174	0.174	10.857	10.857	10.857
	1.249	1.249	8.099	8.103	8.097	0.174	0.174	0.174	10.857	10.845	10.845
	1.249	1.249	8.099	8.105	8.097	0.174	0.174	0.174	10.857	10.845	10.845
1.5	1.499	1.499	6.098	6.103	6.104	0.226	0.226	0.226	11.149	11.149	11.149
	1.499	1.499	6.098	6.103	6.105	0.226	0.226	0.226	11.149	11.149	11.149
	1.498	1.499	6.098	6.104	6.099	0.226	0.226	0.226	11.149	11.140	11.140
	1.498	1.499	6.098	6.105	6.099	0.226	0.226	0.226	11.149	11.140	11.140
1.75	1.749	1.749	4.798	4.802	4.802	0.280	0.280	0.280	11.351	11.351	11.351
	1.749	1.749	4.798	4.801	4.803	0.280	0.280	0.280	11.351	11.351	11.351
	1.748	1.749	4.798	4.803	4.799	0.280	0.280	0.280	11.351	11.344	11.344
	1.748	1.749	4.798	4.804	4.799	0.280	0.280	0.280	11.351	11.344	11.344
2	1.999	1.999	3.902	3.905	3.906	0.337	0.337	0.337	11.495	11.495	11.495
	1.999	1.999	3.902	3.905	3.906	0.337	0.337	0.337	11.495	11.495	11.495
	1.998	1.999	3.902	3.905	3.906	0.337	0.337	0.337	11.495	11.490	11.490
	1.998	1.999	3.902	3.908	3.904	0.337	0.337	0.337	11.495	11.490	11.490

cases, while they are $(r_{\max} = 0.269, H_{\max} = 11.330)$ and $(r_{\min} = 0.269, H_{\min} = 11.307)$ for Case #3 and $(r_{\max} = 0.276, H_{\max} = 11.330)$ and $(r_{\min} = 0.266, H_{\min} = 11.299)$ for Case #4.

4.2.4 Influence of number of variables p

In this subsection, we investigate the influence of the parameter $p \in \{3, 5, 7, 9\}$ for fixed values $\sigma_{\mathrm{M}} = 0.3$, $b = 1$ and $m = 3$. Here we have only considered Case #1 for the variance-covariance matrix $\mathbf{\Sigma}^*$, i.e when $\mathbf{\Sigma}^*$ is uncorrelated with equal variances. For different values of the shift $\delta \in [0, 2]$, the values of δ_{M} along with the optimal values of ARL, r and H are listed in Table 4.4 on page 68 for the MEWMA-CoDa control chart with ($\min$ and $\max$) and without the presence of measurement errors. The first rows in Table 4.4 on page 68 show the values when $p = 3$, the second row represents the values for $p = 5$, the third row represents the values when $p = 7$ and the forth row shows the values when $p = 9$. From this table we can see that when p increases the out-of-control ARL also increases. (i.e. an increase in p has a *negative* impact on the MEWMA-CoDa control chart). For instance, if $\delta = 1.75$ and $p = 3$, the "best" and "worst" ARL values are $(\mathrm{ARL}_{\min} = 5.049, \mathrm{ARL}_{\max} = 5.099)$. But, when $p = 5$ (increase), the "best" and "worst" ARL values are $(\mathrm{ARL}_{\min} = 7.319, \mathrm{ARL}_{\max} = 7.319)$.

The value of the optimal couples (r, H) decreases with an increase in the value of p. When $\delta = 1.75$ and $p = 3$, the maximum and minimum values of the optimal couple (r, H) are $(r_{\max} = 0.266, H_{\max} = 11.306)$ and $(r_{\min} = 0.269, H_{\min} = 11.314)$. But when $p = 9$ (increase) the values of the optimal couple for minimum and maximum decreases, i.e $(r_{\min} = r_{\max} = 0.207, H_{\min} = H_{\max} = 22.726)$.

4.2.5 Comparison between Hotelling CoDa T^2 and MEWMA-CoDa control charts in the presence of measurement errors

In this section first we fix the in-control ARL of the Hotelling CoDa T^2 and MEWMA-CoDa control charts in the presence of measurement errors to be 370.In Table 4.5 on page 70, we compare the out-of-control performances of the Hotelling CoDa T^2 and MEWMA-CoDa control charts in the presence of measurement errors for several combinations of $\sigma_{\mathrm{M}} \in \{0.1, 0.3, 0.6\}$, $b \in \{0.25, 1, 8\}$, $m \in \{1, 3, 6\}$ (one of these parameters varies and the others remain fixed) and for several values of the shift $\delta \in \{0.25, 0.75, 2\}$. For both Hotelling CoDa T^2 and MEWMA-CoDa control charts, the best $(\mathrm{ARL}_{\min})$ and worst $(\mathrm{ARL}_{\max})$ ARL values are provided for both case #1 (uncorrelated) and case #2 (correlated) as well as their percentage improvement indicators $\Delta_{\min} = \frac{100(\mathrm{ARL}_{\min}^{(T^2)} - \mathrm{ARL}_{\min}^{(\mathrm{MEWMA})})}{\mathrm{ARL}_{\min}^{(T^2)}}$ and $\Delta_{\max} = \frac{100(\mathrm{ARL}_{\max}^{(T^2)} - \mathrm{ARL}_{\max}^{(\mathrm{MEWMA})})}{\mathrm{ARL}_{\max}^{(T^2)}}$. The ARL values for the Hotelling T^2 CoDa control chart have already been obtained in Zaidi et al. (2019). From Table 4.5 on page 70, we can draw the following conclusions:

1. No matter the case (correlated or uncorrelated), the MEWMA-CoDa control chart clearly has the smaller best $\mathrm{ARL}_{\min}$ and the smaller worst $\mathrm{ARL}_{\max}$ values compared to the Hotelling CoDa T^2 control chart. For instance, when $\sigma_{\mathrm{M}} = 0.3$, $b = 1$, $m = 3$ and $\delta = 0.25$ the best and worst ARL values for the Hotelling CoDa T^2 control chart are $(\mathrm{ARL}_{\min} = 346.46, \mathrm{ARL}_{\max} = 357.35)$ for Case #1 and $(\mathrm{ARL}_{\min} = 345.05, \mathrm{ARL}_{\max} = 359.01)$ for Case #2 while, for the MEWMA-CoDa control chart, these values are $(\mathrm{ARL}_{\min} = 95.65, \mathrm{ARL}_{\max} = 96.51)$ for Case #1 and $(\mathrm{ARL}_{\min} = 95.21, \mathrm{ARL}_{\max} = 97.26)$ for Case #2.

2. In terms of their percentage improvement indicators we can see that, depending on the parameters σ_{M}, b, m and the level of shift δ, the MEWMA-CoDa control chart is between 70% to 90% more efficient than the Hotelling CoDa T^2 control chart. More precisely (see the last row of 4.5), for Case #1, the MEWMA-CoDa control chart is 86.59% in average more efficient than the Hotelling CoDa T^2 chart for $\mathrm{ARL}_{\min}$ and 87.01% more efficient for $\mathrm{ARL}_{\max}$ and, for Case #2, the latter is 86.19% in average more efficient than the former for $\mathrm{ARL}_{\min}$ and 86.5% more efficient for $\mathrm{ARL}_{\max}$ (i.e.

Table 4.3 – Influence of parameter m

$m = 1$

	δ_M		ARL			r			H		
δ	min	max	without	min	max	without	min	max	without	min	max
0	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
0.25	0.227	0.223	91.212	104.473	107.126	0.050	0.050	0.050	8.895	8.895	8.895
	0.229	0.220	91.212	103.234	109.368	0.050	0.050	0.050	8.895	8.895	8.895
	0.208	0.227	91.212	116.721	103.665	0.050	0.050	0.050	8.895	8.895	8.895
	0.205	0.228	91.212	119.627	103.035	0.050	0.050	0.050	8.895	8.895	8.895
0.5	0.455	0.446	32.162	37.023	38.032	0.050	0.050	0.050	8.895	8.895	8.895
	0.458	0.440	32.162	36.556	38.896	0.050	0.050	0.050	8.895	8.895	8.895
	0.417	0.455	32.162	41.989	36.860	0.050	0.050	0.050	8.895	8.895	8.895
	0.409	0.457	32.162	43.179	36.621	0.050	0.050	0.050	8.895	8.895	8.895
0.75	0.682	0.670	17.720	20.434	20.989	0.083	0.071	0.069	9.802	9.542	9.483
	0.688	0.660	17.720	20.177	21.461	0.083	0.071	0.069	9.802	9.542	8.894
	0.625	0.682	17.720	23.171	20.377	0.083	0.062	0.071	9.802	9.259	9.515
	0.614	0.685	17.720	23.809	20.245	0.083	0.062	0.071	9.802	9.259	9.515
1	0.909	0.893	11.437	13.236	13.606	0.126	0.109	0.107	10.438	10.233	10.199
	0.917	0.880	11.437	13.065	13.921	0.126	0.109	0.105	10.438	10.233	10.165
	0.833	0.909	11.437	15.081	13.213	0.126	0.095	0.109	10.438	9.996	10.214
	0.818	0.913	11.437	15.510	13.125	0.126	0.093	0.109	10.438	9.956	10.214
1.25	1.136	1.116	8.099	9.389	9.654	0.174	0.152	0.147	10.857	10.693	10.652
	1.146	1.099	8.099	9.266	9.881	0.174	0.152	0.145	10.857	10.693	10.631
	1.042	1.136	8.099	10.724	9.377	0.174	0.133	0.150	10.857	10.499	10.659
	1.023	1.141	8.099	11.035	9.314	0.174	0.128	0.152	10.857	10.448	10.679
1.5	1.364	1.339	6.098	7.074	7.275	0.226	0.197	0.193	11.149	11.006	10.978
	1.375	1.319	6.098	6.981	7.447	0.226	0.200	0.188	11.149	11.019	10.950
	1.250	1.364	6.098	8.090	7.068	0.226	0.174	0.197	11.149	10.845	10.995
	1.227	1.370	6.098	8.326	7.020	0.226	0.169	0.197	11.149	10.811	10.995
1.75	1.591	1.563	4.798	5.564	5.723	0.280	0.245	0.240	11.351	11.229	11.210
	1.605	1.539	4.798	5.491	5.858	0.280	0.250	0.235	11.351	11.247	11.190
	1.458	1.591	4.798	6.367	5.561	0.280	0.216	0.245	11.351	11.095	11.221
	1.432	1.598	4.798	6.553	5.523	0.280	0.212	0.247	11.351	11.072	11.230
2	1.818	1.786	3.902	4.522	4.650	0.337	0.297	0.290	11.495	11.399	11.379
	1.834	1.759	3.902	4.462	4.759	0.337	0.299	0.283	11.495	11.405	11.358
	1.667	1.818	3.902	5.173	4.519	0.337	0.261	0.297	11.495	11.282	11.393
	1.636	1.826	3.902	5.324	4.489	0.337	0.257	0.299	11.495	11.265	11.399

$m = 3$

δ	δ_M min	δ_M max	ARL without	ARL min	ARL max	r without	r min	r max	H without	H min	H max
0	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
0.25	0.242	0.240	91.212	95.626	96.510	0.050	0.050	0.050	8.895	8.895	8.895
	0.243	0.239	91.212	95.213	97.258	0.050	0.050	0.050	8.895	8.895	8.895
	0.234	0.242	91.212	99.291	94.918	0.050	0.050	0.050	8.895	8.895	8.895
	0.233	0.242	91.212	100.273	94.708	0.050	0.050	0.050	8.895	8.895	8.895
0.5	0.484	0.481	32.162	33.747	34.068	0.050	0.050	0.050	8.895	8.895	8.895
	0.485	0.478	32.162	33.598	34.341	0.050	0.050	0.050	8.895	8.895	8.895
	0.469	0.484	32.162	35.216	33.606	0.050	0.050	0.050	8.895	8.895	8.895
	0.465	0.485	32.162	35.582	33.530	0.050	0.050	0.050	8.895	8.895	8.895
0.75	0.726	0.721	17.720	18.613	18.793	0.083	0.079	0.079	9.802	9.704	9.704
	0.728	0.717	17.720	18.529	18.946	0.083	0.079	0.076	9.802	9.704	9.652
	0.703	0.726	17.720	19.467	18.566	0.083	0.074	0.079	9.802	9.572	9.679
	0.698	0.727	17.720	19.670	18.523	0.083	0.074	0.079	9.802	9.572	9.679
1	0.968	0.962	11.437	12.028	12.147	0.126	0.119	0.119	10.438	10.355	10.355
	0.971	0.956	11.437	11.972	12.248	0.126	0.121	0.117	10.438	10.384	10.326
	0.938	0.968	11.437	12.606	12.008	0.126	0.114	0.119	10.438	10.277	10.338
	0.931	0.969	11.437	12.742	11.980	0.126	0.112	0.119	10.438	10.246	10.338
1.25	1.210	1.202	8.099	8.522	8.607	0.174	0.166	0.164	10.857	10.806	10.788
	1.213	1.195	8.099	8.482	8.680	0.174	0.166	0.162	10.857	10.806	10.770
	1.172	1.210	8.099	8.941	8.512	0.174	0.157	0.166	10.857	10.719	10.793
	1.164	1.212	8.099	9.039	8.492	0.174	0.157	0.166	10.857	10.719	10.793
1.5	1.452	1.442	6.098	6.418	6.482	0.226	0.216	0.214	11.149	11.105	11.093
	1.456	1.435	6.098	6.388	6.537	0.226	0.216	0.212	11.149	11.105	11.082
	1.406	1.452	6.098	6.738	6.413	0.226	0.207	0.216	11.149	11.047	11.095
	1.396	1.454	6.098	6.811	6.397	0.226	0.204	0.216	11.149	11.035	11.095
1.75	1.694	1.683	4.798	5.049	5.099	0.280	0.269	0.266	11.351	11.314	11.306
	1.699	1.674	4.798	5.025	5.142	0.280	0.271	0.264	11.351	11.321	11.298
	1.641	1.694	4.798	5.301	5.046	0.280	0.257	0.269	11.351	11.265	11.307
	1.629	1.696	4.798	5.359	5.034	0.280	0.254	0.269	11.351	11.257	11.307
2	1.935	1.923	3.902	4.105	4.146	0.337	0.323	0.321	11.495	11.464	11.458
	1.941	1.913	3.902	4.086	4.180	0.337	0.323	0.318	11.495	11.464	11.453
	1.875	1.935	3.902	4.309	4.103	0.337	0.309	0.323	11.495	11.424	11.458
	1.862	1.938	3.902	4.356	4.093	0.337	0.307	0.323	11.495	11.418	11.458

$m = 6$

	δ_M		ARL			r			H		
δ	min	max	without	min	max	without	min	max	without	min	max
0	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
0.25	0.246	0.245	91.212	93.418	93.859	0.050	0.050	0.050	8.895	8.895	8.895
	0.246	0.244	91.212	93.211	94.233	0.050	0.050	0.050	8.895	8.895	8.895
	0.242	0.246	91.212	94.918	92.734	0.050	0.050	0.050	8.895	8.895	8.895
	0.241	0.246	91.212	95.409	92.629	0.050	0.050	0.050	8.895	8.895	8.895
0.5	0.492	0.490	32.162	32.950	33.109	0.050	0.050	0.050	8.895	8.895	8.895
	0.493	0.489	32.162	32.876	33.243	0.050	0.050	0.050	8.895	8.895	8.895
	0.484	0.492	32.162	33.606	32.814	0.050	0.050	0.050	8.895	8.895	8.895
	0.482	0.492	32.162	33.785	32.776	0.050	0.050	0.050	8.895	8.895	8.895
0.75	0.738	0.735	17.720	18.165	18.255	0.083	0.081	0.081	9.802	9.754	9.754
	0.739	0.733	17.720	18.123	18.330	0.083	0.081	0.079	9.802	9.754	9.704
	0.726	0.738	17.720	18.566	18.120	0.083	0.079	0.079	9.802	9.679	9.730
	0.723	0.738	17.720	18.667	18.099	0.083	0.079	0.081	9.802	9.679	9.730
1	0.984	0.980	11.437	11.731	11.790	0.126	0.121	0.121	10.438	10.384	10.384
	0.985	0.978	11.437	11.704	11.840	0.126	0.124	0.121	10.438	10.412	10.384
	0.968	0.984	11.437	12.008	11.713	0.126	0.119	0.119	10.438	10.338	10.366
	0.964	0.984	11.437	12.075	11.699	0.126	0.119	0.121	10.438	10.338	10.366
1.25	1.230	1.225	8.099	8.309	8.352	0.174	0.169	0.169	10.857	10.824	10.824
	1.231	1.222	8.099	8.290	8.388	0.174	0.171	0.169	10.857	10.841	10.824
	1.210	1.230	8.099	8.512	8.300	0.174	0.166	0.166	10.857	10.793	10.811
	1.205	1.230	8.099	8.560	8.290	0.174	0.164	0.169	10.857	10.776	10.811
1.5	1.475	1.471	6.098	6.257	6.289	0.226	0.221	0.219	11.149	11.128	11.116
	1.478	1.467	6.098	6.242	6.316	0.226	0.221	0.219	11.149	11.128	11.116
	1.452	1.475	6.098	6.413	6.253	0.226	0.216	0.216	11.149	11.095	11.118
	1.446	1.477	6.098	6.449	6.245	0.226	0.214	0.221	11.149	11.084	11.118
1.75	1.721	1.716	4.798	4.923	4.948	0.280	0.276	0.273	11.351	11.337	11.329
	1.724	1.711	4.798	4.911	4.969	0.280	0.276	0.273	11.351	11.337	11.329
	1.694	1.721	4.798	5.046	4.920	0.280	0.269	0.269	11.351	11.307	11.330
	1.687	1.723	4.798	5.074	4.914	0.280	0.266	0.276	11.351	11.299	11.330
2	1.967	1.961	3.902	4.003	4.023	0.337	0.330	0.328	11.495	11.480	11.474
	1.970	1.955	3.902	3.993	4.040	0.337	0.330	0.328	11.495	11.480	11.474
	1.935	1.967	3.902	4.103	4.001	0.337	0.323	0.323	11.495	11.458	11.474
	1.929	1.969	3.902	4.126	3.996	0.337	0.321	0.330	11.495	11.453	11.474

Table 4.4 – Influence of parameter p

	δ_{M}		ARL			r			H		
δ	min	max	without	min	max	without	min	max	without	min	max
0	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
	0.000	0.000	370.000	370.000	370.000	0.050	0.050	0.050	8.895	8.895	8.895
0.25	0.242	0.240	91.212	95.626	96.510	0.050	0.050	0.050	8.895	8.895	8.895
	0.242	0.242	91.212	117.688	117.688	0.050	0.050	0.050	8.895	13.011	13.011
	0.242	0.242	91.212	133.422	133.422	0.050	0.050	0.050	8.895	16.599	16.599
	0.242	0.242	91.212	145.456	145.456	0.050	0.050	0.050	8.895	19.915	19.915
0.5	0.484	0.481	32.162	33.747	34.068	0.050	0.050	0.050	8.895	8.895	8.895
	0.484	0.484	32.162	41.166	41.166	0.050	0.050	0.050	8.895	13.011	13.011
	0.484	0.484	32.162	46.800	46.800	0.050	0.050	0.050	8.895	16.599	16.599
	0.484	0.484	32.162	51.498	51.498	0.050	0.050	0.050	8.895	19.915	19.915
0.75	0.726	0.721	17.720	18.613	18.793	0.083	0.079	0.079	9.802	9.704	9.704
	0.726	0.726	17.720	22.508	22.508	0.083	0.069	0.069	9.802	13.693	13.693
	0.726	0.726	17.720	25.401	25.401	0.083	0.064	0.064	9.802	17.187	17.187
	0.726	0.726	17.720	27.784	27.784	0.083	0.060	0.060	9.802	20.360	20.360
1	0.968	0.962	11.437	12.028	12.147	0.126	0.119	0.119	10.438	10.355	10.355
	0.968	0.968	11.437	14.489	14.489	0.126	0.105	0.105	10.438	14.468	14.468
	0.968	0.968	11.437	16.324	16.324	0.126	0.098	0.098	10.438	18.054	18.054
	0.968	0.968	11.437	17.845	17.845	0.126	0.090	0.090	10.438	21.318	21.318
1.25	1.210	1.202	8.099	8.522	8.607	0.174	0.166	0.164	10.857	10.806	10.788
	1.210	1.210	8.099	10.221	10.221	0.174	0.145	0.145	10.857	14.986	14.986
	1.210	1.210	8.099	11.492	11.492	0.174	0.136	0.136	10.857	18.635	18.635
	1.210	1.210	8.099	12.548	12.548	0.174	0.126	0.126	10.857	21.964	21.964
1.5	1.452	1.442	6.098	6.418	6.482	0.226	0.216	0.214	11.149	11.105	11.093
	1.452	1.452	6.098	7.665	7.665	0.226	0.190	0.190	11.149	15.352	15.352
	1.452	1.452	6.098	8.599	8.599	0.226	0.176	0.176	11.149	19.027	19.027
	1.452	1.452	6.098	9.376	9.376	0.226	0.166	0.166	11.149	22.422	22.422
1.75	1.694	1.683	4.798	5.049	5.099	0.280	0.269	0.266	11.351	11.314	11.306
	1.694	1.694	4.798	6.004	6.004	0.280	0.238	0.238	11.351	15.606	15.606
	1.694	1.694	4.798	6.721	6.721	0.280	0.219	0.219	11.351	19.306	19.306
	1.694	1.694	4.798	7.319	7.319	0.280	0.207	0.207	11.351	22.726	22.726
2	1.935	1.923	3.902	4.105	4.146	0.337	0.323	0.321	11.495	11.464	11.458
	1.935	1.935	3.902	4.860	4.860	0.337	0.288	0.288	11.495	15.789	15.789
	1.935	1.935	3.902	5.430	5.430	0.337	0.264	0.264	11.495	19.510	19.510
	1.935	1.935	3.902	5.904	5.904	0.337	0.250	0.250	11.495	22.950	22.950

MEWMA-CoDa control chart is about 86%-87% more efficient than the Hotelling CoDa T^2 chart in the presence of measurement errors).

4.3 Illustrative example

Let us take the example that we have discussed in chapter 3 of a company produces muesli (for breakfast), where every 100 grams contain: (A) 66% of whole-grain cereals (barley flakes, oat flakes, wheat flakes), (B) 24% of dried fruits (raisin, papaya, banana) and (C) 10% of nuts (almond, hazelnut, coconut).(Note that the data used in this example are simulated realistic data depicting a practical situation). Firstly, the measurement device used to measure the percentages of the components in the produced muesli, is calibrated by the company to estimate the values of parameters $\mathbf{a}$, b and $\mathbf{\Sigma}^*_{\mathrm{M}}$. In order to do that, $k = 4$ reference samples of muesli have been prepared by the company with known percentages of each component and they have been measured $m = 7$ times for the purpose of calibration. Table 3.4 on page 45 shows the results obtained by this calibration along with the ilr transformed values $\mathbf{y}^*_i = \mathrm{ilr}(\mathbf{y}_i)$ and $\mathbf{x}^*_{i,j} = \mathrm{ilr}(\mathbf{x}_{i,j})$. In Figure 3.4 on page 45, the values of known composition $\mathbf{y}_1, \ldots, \mathbf{y}_4$ are marked with a "+". The observed values of the compositions $\mathbf{x}_{1,1}, \ldots, \mathbf{x}_{4,7}$ are plotted on the left side of Figure 3.4 on page 45 and their corresponding ilr transformed values $\mathbf{y}^*_i$ and $\mathbf{x}^*_{i,j}$ are plotted on the right side of Figure 3.4 on page 45.

Using the estimation method provided in chapter 3 on page 44 with $d = 2$, $\mathbf{v}_i = \mathbf{y}^*_i$ and $\mathbf{u}_{i,j} = \mathbf{x}^*_{i,j}$ for $i = 1, \ldots, 4$ and $j = 1, \ldots, 7$, As we saw in chapter 3 page 46, we get the following estimators $\hat{\mathbf{a}}$, $\hat{b}$ and $\hat{\mathbf{\Sigma}}^*_{\mathrm{M}}$ for $\mathbf{a}$, b and $\mathbf{\Sigma}^*_{\mathrm{M}}$:

1. $\hat{\mathbf{a}} = (0.3354, 0.3357, 0.3289)$ or, equivalently, $\hat{\mathbf{a}}^* = (0.0162972, -0.0006318)$,
2. $\hat{b} = 1.1070$,
3. $\hat{\mathbf{\Sigma}}^*_{\mathrm{M}} = \begin{pmatrix} 0.0014346 & 0.0007812 \\ 0.0007812 & 0.0102893 \end{pmatrix}$.
4. $\hat{\boldsymbol{\mu}}^*_{\bar{\mathbf{x}}} = (1.2766, 0.7657)$,
5. $\hat{\mathbf{\Sigma}}^*_{\bar{\mathbf{x}}} = \begin{pmatrix} 0.0146362 & 0.0105839 \\ 0.0105839 & 0.0510887 \end{pmatrix}$.
6. $\hat{\boldsymbol{\mu}}^*_0 = \frac{1}{\hat{b}}(\hat{\boldsymbol{\mu}}^*_{\bar{\mathbf{x}}} - \hat{\mathbf{a}}^*) = (1.1385, 0.6922)$,
7. $\hat{\mathbf{\Sigma}}^* = \frac{1}{\hat{b}^2}\left(\hat{\mathbf{\Sigma}}^*_{\bar{\mathbf{x}}} - \frac{1}{m}\hat{\mathbf{\Sigma}}^*_{\mathrm{M}}\right) = \begin{pmatrix} 0.0115533 & 0.0084242 \\ 0.0084242 & 0.038891 \end{pmatrix}$.

We have computed the MEWMA-CoDa statistics Q^*_i for $i = 1, \ldots, 20$ using Equation 4.1 on page 52. The optimal couple for $\delta = 1.5$ is $(r = 0.226, H = 11.149)$. The values of the MEWMA-CoDa statistics Q^*_i are listed in Table 4.6 on page 71 and also plotted in Figure 4.1 on page 72 with the upper control limit $\mathrm{UCL} = H = 11.149$. One can see that the process is in control, as all the values in Figure 4.1 on page 72 are smaller than the upper control limit.

Concerning the Phase II, $i = 1, \ldots, 20$ batches of muesli have also been measured $m = 3$ times. Table 4.7 on page 73 shows the results with the values $\mathbf{x}_{i,j}$, $\bar{\mathbf{x}}_i$ and $\bar{\mathbf{x}}^*_i$, $i = 1, \ldots, 20$. The values of the MEWMA-CoDa statistics Q^*_i in the presence of measurement error are listed in Table 4.7 on page 73 and also plotted in Figure 4.2 on page 72 with the upper control limit $\mathrm{UCL} = H = 11.149$ obtained in Phase I.

The process seems to be in-control up to sample #14 but sample #15 is clearly out-of-control. Then, in samples #15 and #16 we found that the level of whole-grain cereals dropped down suddenly due to a malfunction of the hatch regulating the quantity of whole grain cereals causing a shift from $\hat{\boldsymbol{\mu}}^*_0 = (1.1385, 0.6922)$ to $\hat{\boldsymbol{\mu}}^*_{\bar{\mathbf{x}}} = (1.2766, 0.7657)$. Using Equation 4.6 on page 52, we found that $\delta = 1.62$. Here, a shift of size $\delta = 1.5$ in $\hat{\boldsymbol{\mu}}^*_{\bar{\mathbf{x}}}$ has been interpreted. For $m = 3$ and $\delta = 1.5$, the optimal parameters of MEWMA-CoDa chart are $r = 0.226$, $H = 11.149$.

Table 4.5 – Comparison in terms of out-of-control ARLs between the Hotelling CoDa T^2 and MEWMA-CoDa control charts in the presence of measurement error

δ	Parameters			Hotelling CoDa T^2				MEWMA-CoDa				Performance percentage			
	σ_M	b	m	Case#1		Case#2		Case#1		Case#2		% Case#1		% Case#2	
				ARL_{min}	ARL_{max}	ARL_{min}	ARL_{max}	ARL_{min}	ARL_{max}	ARL_{min}	ARL_{max}	Δ_{min}	Δ_{max}	Δ_{min}	Δ_{max}
0.25	0.1	1	3	314.78	336.99	312.24	340.83	92.69	92.98	93.23	92.98	70.56	72.41	70.15	72.72
0.25	0.3	1	3	346.46	357.35	345.06	359.01	95.63	96.52	95.22	97.26	72.4	73	72.41	72.91
0.25	0.6	1	3	357.35	363.43	356.54	364.32	100.05	101.82	99.23	103.32	72.01	71.99	72.18	71.65
0.25	0.3	0.25	3	368.31	369.15	368.2	369.27	159.51	171.77	153.57	181.65	56.7	53.47	58.3	50.81
0.25	0.3	1	3	346.46	357.35	345.06	359.01	95.63	96.52	95.22	97.26	72.4	73	72.41	72.91
0.25	0.3	8	3	216.63	228.9	215.75	232.5	91.29	91.3	91.28	91.31	57.87	60.12	57.7	60.73
0.25	0.3	1	1	361.35	365.57	360.78	366.17	104.48	107.13	103.24	109.37	71.09	70.7	71.39	70.14
0.25	0.3	1	3	346.46	357.35	345.06	359.01	95.63	96.52	95.22	97.26	72.4	73	72.41	72.91
0.25	0.3	1	6	328.68	346.46	326.53	349.36	32.95	33.11	93.22	94.24	89.98	90.45	71.46	73.03
0.75	0.1	1	3	237.85	283.92	233.05	292.73	18.02	18.08	17.99	18.13	92.43	93.64	92.29	93.81
0.75	0.3	1	3	306.2	334.11	302.78	338.59	18.62	18.8	11.98	12.25	93.93	94.38	96.05	96.39
0.75	0.6	1	3	334.11	350.86	331.92	353.39	19.52	19.89	19.35	20.2	94.16	94.34	94.18	94.29
0.75	0.3	0.25	3	364.97	367.46	364.62	367.82	33.22	36.69	31.65	39.76	90.9	90.02	91.33	89.2
0.75	0.3	1	3	306.2	334.11	302.78	338.59	18.62	18.8	11.98	12.25	93.93	94.38	96.05	96.39
0.75	0.3	8	3	104.1	116.31	103.26	120.08	17.74	17.74	17.74	17.74	82.97	84.75	82.83	85.23
0.75	0.3	1	1	345.03	356.95	343.44	358.71	20.44	20.99	20.18	21.47	94.08	94.12	94.13	94.02
0.75	0.3	1	3	306.2	334.11	302.78	338.59	18.62	18.8	11.98	12.25	93.93	94.38	96.05	96.39
0.75	0.3	1	6	265.73	306.2	261.21	313.38	18.17	18.26	18.13	18.34	93.17	94.04	93.07	94.16
2	0.1	1	3	137.9	197.68	132.58	211.16	3.97	3.99	3.97	4	97.13	97.99	97.02	98.11
2	0.3	1	3	233.29	285.76	227.48	295.15	4.11	4.15	4.09	4.19	98.25	98.55	98.21	98.59
2	0.6	1	3	285.76	322.39	281.29	328.31	4.32	4.4	4.28	4.47	98.5	98.64	98.49	98.64
2	0.3	0.25	3	356.82	363.29	355.93	364.22	7.59	8.42	7.21	9.15	97.88	97.69	97.98	97.49
2	0.3	1	3	233.29	285.76	227.48	295.15	4.11	4.15	4.09	4.19	98.25	98.55	98.21	98.59
2	0.3	8	3	33.99	40.31	33.58	42.37	3.91	3.91	3.91	3.91	88.51	90.31	88.37	90.78
2	0.3	1	1	309.16	336.83	305.64	341.12	4.53	4.65	4.47	4.76	98.54	98.62	98.55	98.61
2	0.3	1	3	233.29	285.76	227.48	295.15	4.11	4.15	4.09	4.19	98.25	98.55	98.21	98.59
2	0.3	1	6	172.08	233.29	166.14	245.88	4.01	4.03	4	4.05	97.68	98.28	97.6	98.36
												86.59	87.01	86.19	86.5

Table 4.6 – Phase I data for the muesli example

i	j	$\mathbf{x}_{i,j}$			$\bar{\mathbf{x}}_i$			$\bar{\mathbf{x}}_i^*$		Q_i
1	1	0.77	0.16	0.07	0.7090	0.2078	0.0832	1.2483	0.8678	0.1530
	2	0.67	0.24	0.09						
	3	0.68	0.23	0.09						
2	1	0.64	0.27	0.09	0.6301	0.2767	0.0932	1.2241	0.5819	0.1672
	2	0.63	0.28	0.09						
	3	0.62	0.28	0.10						
3	1	0.76	0.16	0.08	0.6958	0.2108	0.0934	1.1523	0.8445	1.0106
	2	0.65	0.25	0.10						
	3	0.67	0.23	0.10						
4	1	0.65	0.27	0.08	0.6434	0.2766	0.0800	1.3575	0.5968	0.2361
	2	0.64	0.28	0.08						
	3	0.64	0.28	0.08						
5	1	0.50	0.38	0.12	0.5268	0.3634	0.1098	1.1289	0.2625	3.1357
	2	0.54	0.36	0.10						
	3	0.54	0.35	0.11						
6	1	0.80	0.14	0.06	0.7677	0.1657	0.0666	1.3699	1.0843	0.2382
	2	0.74	0.19	0.07						
	3	0.76	0.17	0.07						
7	1	0.75	0.18	0.07	0.7304	0.1995	0.0701	1.3843	0.9177	0.1118
	2	0.71	0.22	0.07						
	3	0.73	0.20	0.07						
8	1	0.65	0.26	0.09	0.6534	0.2533	0.0932	1.2030	0.6700	0.0518
	2	0.65	0.25	0.10						
	3	0.66	0.25	0.09						
9	1	0.65	0.27	0.08	0.6635	0.2565	0.0800	1.3389	0.6721	0.3602
	2	0.66	0.26	0.08						
	3	0.68	0.24	0.08						
10	1	0.76	0.17	0.07	0.7534	0.1766	0.0700	1.3478	1.0257	0.3204
	2	0.75	0.18	0.07						
	3	0.75	0.18	0.07						
11	1	0.67	0.24	0.09	0.6601	0.2499	0.0900	1.2302	0.6868	0.0314
	2	0.65	0.26	0.09						
	3	0.66	0.25	0.09						
12	1	0.53	0.36	0.11	0.5367	0.3533	0.1100	1.1234	0.2956	1.5351
	2	0.54	0.35	0.11						
	3	0.54	0.35	0.11						
13	1	0.75	0.16	0.09	0.7251	0.1813	0.0936	1.1058	0.9803	1.9162
	2	0.67	0.23	0.10						
	3	0.75	0.16	0.09						
14	1	0.67	0.24	0.09	0.6701	0.2333	0.0966	1.1510	0.7460	3.0274
	2	0.67	0.23	0.10						
	3	0.67	0.23	0.10						
15	1	0.64	0.27	0.09	0.6468	0.2632	0.0900	1.2431	0.6357	2.2233
	2	0.64	0.27	0.09						
	3	0.66	0.25	0.09						
16	1	0.72	0.21	0.07	0.7440	0.1928	0.0632	1.4614	0.9548	0.0621
	2	0.74	0.20	0.06						
	3	0.77	0.17	0.06						
17	1	0.73	0.20	0.07	0.7436	0.1899	0.0665	1.4137	0.9652	0.4494
	2	0.75	0.18	0.07						
	3	0.75	0.19	0.06						
18	1	0.65	0.28	0.07	0.6568	0.2767	0.0665	1.5169	0.6112	3.1374
	2	0.65	0.28	0.07						
	3	0.67	0.27	0.06						
19	1	0.76	0.17	0.07	0.7604	0.1730	0.0665	1.3846	1.0468	3.3321
	2	0.74	0.19	0.07						
	3	0.78	0.16	0.06						
20	1	0.67	0.23	0.10	0.7013	0.2055	0.0931	1.1475	0.8678	0.9865
	2	0.68	0.22	0.10						
	3	0.75	0.17	0.08						

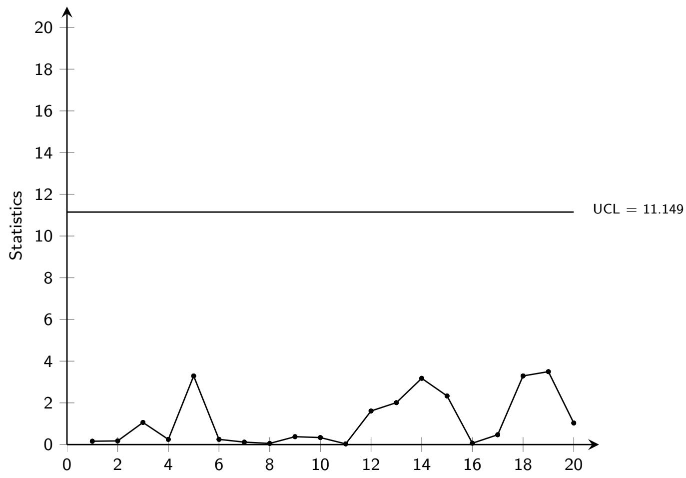

Figure 4.1 – MEWMA-CoDa control chart for muesli Phase I data

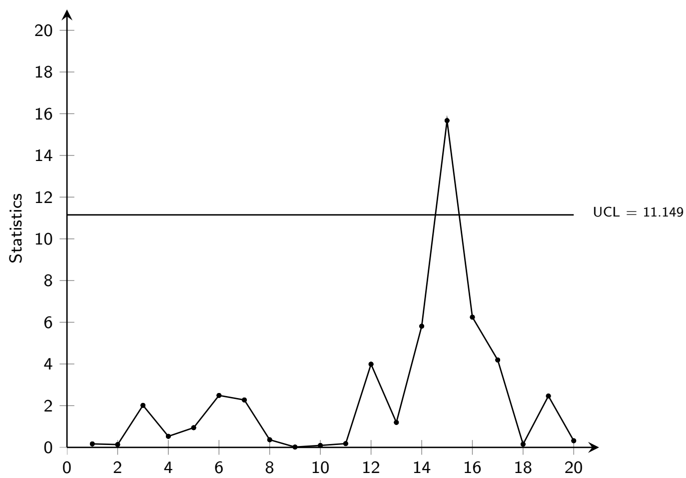

Figure 4.2 – MEWMA-CoDa control chart for muesli Phase II data

Table 4.7 – Phase II data for the muesli example

i	j	$\mathbf{x}_{i,j}$			$\bar{\mathbf{x}}_i$			$\bar{\mathbf{x}}_i^*$		Q_i
1	1	0.68	0.23	0.09	0.6504	0.2563	0.0933	1.2050	0.6586	0.1608
	2	0.63	0.27	0.10						
	3	0.64	0.27	0.09						
2	1	0.68	0.24	0.08	0.6734	0.2433	0.0832	1.2916	0.7198	0.1297
	2	0.66	0.25	0.09						
	3	0.68	0.24	0.08						
3	1	0.62	0.31	0.07	0.6538	0.2761	0.0701	1.4712	0.6096	1.9213
	2	0.67	0.26	0.07						
	3	0.67	0.26	0.07						
4	1	0.67	0.22	0.11	0.6634	0.2265	0.1100	1.0283	0.7598	0.5056
	2	0.65	0.24	0.11						
	3	0.67	0.22	0.11						
5	1	0.63	0.27	0.10	0.6501	0.2566	0.0933	1.2060	0.6573	0.9005
	2	0.66	0.25	0.09						
	3	0.66	0.25	0.09						
6	1	0.77	0.15	0.08	0.7702	0.1498	0.0800	1.1804	1.1577	2.3750
	2	0.76	0.16	0.08						
	3	0.78	0.14	0.08						
7	1	0.67	0.24	0.09	0.6570	0.2498	0.0933	1.1990	0.6838	2.1677
	2	0.66	0.24	0.10						
	3	0.64	0.27	0.09						
8	1	0.64	0.28	0.08	0.6777	0.2456	0.0767	1.3644	0.7176	0.3496
	2	0.72	0.21	0.07						
	3	0.67	0.25	0.08						
9	1	0.72	0.21	0.07	0.7003	0.2265	0.0732	1.3826	0.7982	0.0164
	2	0.68	0.24	0.08						
	3	0.70	0.23	0.07						
10	1	0.64	0.26	0.10	0.6501	0.2567	0.0932	1.2062	0.6571	0.0896
	2	0.66	0.25	0.09						
	3	0.65	0.26	0.09						
11	1	0.73	0.20	0.07	0.7436	0.1863	0.0700	1.3640	0.9786	0.1728
	2	0.74	0.19	0.07						
	3	0.76	0.17	0.07						
12	1	0.81	0.12	0.07	0.8034	0.1266	0.0700	1.2381	1.3067	3.8026
	2	0.80	0.13	0.07						
	3	0.80	0.13	0.07						
13	1	0.64	0.28	0.08	0.6570	0.2630	0.0801	1.3448	0.6474	1.1412
	2	0.68	0.24	0.08						
	3	0.65	0.27	0.08						
14	1	0.82	0.11	0.07	0.7786	0.1415	0.0799	1.1633	1.2057	5.5389
	2	0.74	0.17	0.09						
	3	0.77	0.15	0.08						
15	1	0.61	0.25	0.14	0.6405	0.2193	0.1402	0.8028	0.7580	**14.9341**
	2	0.66	0.20	0.14						
	3	0.65	0.21	0.14						
16	1	0.70	0.23	0.07	0.6936	0.2332	0.0732	1.3906	0.7709	5.9509
	2	0.71	0.22	0.07						
	3	0.67	0.25	0.08						
17	1	0.74	0.19	0.07	0.7401	0.1899	0.0700	1.3700	0.9620	3.9943
	2	0.73	0.20	0.07						
	3	0.75	0.18	0.07						
18	1	0.56	0.35	0.09	0.5335	0.3765	0.0901	1.3099	0.2465	0.1454
	2	0.51	0.40	0.09						
	3	0.53	0.38	0.09						
19	1	0.55	0.35	0.10	0.5468	0.3600	0.0932	1.2736	0.2955	2.3477
	2	0.55	0.36	0.09						
	3	0.54	0.37	0.09						
20	1	0.77	0.16	0.07	0.7539	0.1760	0.0701	1.3460	1.0285	0.3038
	2	0.76	0.17	0.07						
	3	0.73	0.20	0.07						

4.4 Conclusions

Measurement errors have a large impact on control charts. This chapter deals with the MEWMA-CoDa control chart in the presence of measurement errors. We assumed that the quality characteristics Y_i is not directly observable but can be measured through an observable quantity $X_{i,j}$. An additive model has been used to study the influence of measurement errors on the MEWMA-CoDa control chart. For the variance covariance matrix of CoDa, four cases have been taken into account. To study the influence of the parameters σ_{M}, b , m and p on the MEWMA-CoDa control chart, different values of each parameter have been studied. The main conclusions are: i) σ_{M} has a negative impact on the ARL performance of the MEWMA-CoDa control chart keeping b , m and p constant, ii) b has a positive impact on the ARL performance of the MEWMA-CoDa control chart keeping σ_{M} , m and p constant, iii) m has a positive but mild impact on the ARL performance of the MEWMA-CoDa control chart keeping σ_{M} , b and p constant, iv) p has a negative impact on the ARL performance of the MEWMA-CoDa control chart keeping σ_{M} ,b and m constant. A comparison with the Hotelling CoDa T^2 control chart also proved that the MEWMA-CoDa control chart is between 70% to 90% more efficient than the Hotelling CoDa T^2 control chart in the presence of measurement errors. Future research about control charts for monitoring CoDa can focus on studying the effect of the estimation of the parameters on the statistical properties of CoDa type control charts.

The work presented in this chapter has been accepted for publication in Journal of Quality and Reliability Engineering International.

In this chapter we discussed the effect of measurement errors on the MEWMA-CoDa control chart proposed by TRAN et al. (2018). In next chapter we will discuss some nonparametric control charts for compositional data using data depth.

5

Nonparametric control charts for Compositional Data

In Multivariate situations, a data depth is used to measure how deep or how centred is a given point in a given data cloud. While dealing with data depth, there is no distributional requirement of any kind concerning the data. There are many studies that deals with control charts using data depth to overcome the problem of non normality of the data. Among them Liu et al. (2004) studied a nonparametric multivariate moving average control charts based on data depth. Bae et al. (2016) studied multivariate control charts for depth-based data that provide dimension reduction for high-dimensional data in a completely nonparametric way. Also Yue and Liu (2017) proposed an adaptive multivariate nonparametric exponentially weighted moving average control chart with variable sampling interval using the Mahalanobis depth. Further Idris et al. (2019) studied a control chart using data depth based on an influence function of a variance vector and they found that the disadvantage of the Mahalanobis distance lies in the generalized variance and using a vector variance is proved to be a better option instead of using the generalized variance. Recently, Barale and Shirke (2019) proposed a nonparametric chart based on a data depth for location parameter.

The goal of this chapter is to study some nonparametric control charts for compositional data using data depth. The remainder of this chapter is organized as follows: in Section 5.1, an Introduction to Data Depth is briefly presented. Section 5.2 details the nonparametric control charts and Section 5.3 investigates the performance of the r chart, Q chart and DDMA chart applied to Compositional Data and Section 5.4 describes an application of the three charts. Finally, an illustrative example is provided in Section 5.5 and conclusions and future research directions are presented in Section 5.6.

5.1 Introduction to Data Depth

The notion of data depth was first introduced by Mahalanobis (1936), then Tukey (1975) defined the half space depth as a graphical representation of bivariate data sets. Further Donoho and Gasko (1992) extended the concept of depth to multivariate data sets. A data depth is described as how centered or deep is a given point from a multivariate distribution. The data depth gives the outward ordering of the data set from the center of the distribution. While dealing with data depth there is no need to fulfill any assumption of statistical distribution, hence it can be considered as a nonparametric statistical data analysis tool. Data depth is suitable for high dimensional data with outliers as depth functions are known to be robust.

Let us assume that $\mathbf{x}_i = (\mathbf{x}_{i,1}, \ldots, \mathbf{x}_{i,p}) \in \mathcal{S}^p$ are p-part compositions at time $i = 1, 2, \ldots$ and $\mathbf{x}_1, \mathbf{x}_2, \ldots$ are independent multivariate random compositions with mean $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$, Let us also assume another sample $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m$ be m compositions where $\mathbf{y}_i$ follows certain distribution G and is referred to as the reference sample. Based on the observations $\mathbf{x}_i$, we have to find whether it is meeting the properties of the prescribed distribution G or not. Assuming F as the distribution of $\mathbf{x}_i$, the objective is to see, if any differences exist between the distributions F and G. In this situation, the notion of data depth can be used.

Some of the depth functions are discussed below,

1. The Mahalanobis depth was introduced by Mahalanobis (1936). Let a point $\mathbf{x}_i = (\mathbf{x}_{i,1}, \ldots, \mathbf{x}_{i,p}) \in \mathcal{S}_n \subset \mathbb{R}^p$, then the Mahalanobis depth of $\mathbf{x}$ having p-dimensional data set $\mathcal{S}_n$ will be

$$M_h D(\mathbf{x}; \mathcal{S}_n) = [1 + (\mathbf{x} - \bar{\mathbf{x}})^\intercal \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}})]^{-1}$$

where $\bar{\mathbf{x}}$ and $\mathbf{S}$ are the mean vector and dispersion matrix of $\mathcal{S}_n$, respectively.

Unfortunately, this function depends on non robust measures such as the mean and dispersion matrix. Also, this function depends on the second moment, so it makes no sense to compute it in the case when the second moment does not exist.

2. The location depth, also known as the Tukey's depth or half space depth has been defined by Tukey (1975). If a point $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p) \in \mathcal{S}_n$, then its location depth will be the smallest number of data points in a closed halfspace with boundary through the point $\mathbf{x}$.

The depth of a point $\mathbf{x}$ can easily be found using the median in case of univariate data. In case of multivariate data the generalized form of the median is used as the point with maximal depth. Therefore, the formula to find Tukey's depth is,

$$TD(F, \mathbf{x}) = \inf_{\mathbf{H}} F(\mathbf{H}) : \mathbf{H} \text{ is a closed half space containing } \mathbf{x}$$

3. The simplicial depth was defined by Liu (1990). Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{p+1}$ be $(p+1)$ iid observations from F. The simplicial depth at point $\mathbf{x}$ is,

$$D_m(\mathbf{x}) = \binom{m}{p+1}^{-1} \sum_{1 \leq i_1 < \ldots < i_{p+1} \leq m} I(\mathbf{x} \in S[\mathbf{x}_{i_1}, \ldots, \mathbf{x}_{i_{p+1}}]) \tag{5.1}$$

4. Projection depth was introduced by Zuo (2003) as,

$$PD(F, \mathbf{x}) = [1 + O_p(F, \mathbf{x})]^{-1}$$

where $O_p(F, \mathbf{x}) = \sup_{\|\mathbf{u}\|=1} \frac{\|\mathbf{u}'\mathbf{x} - \mu F \mathbf{u}\|}{\sigma F \mathbf{u}}$, $F_{\mathbf{u}}$ denotes the distribution of $\mathbf{u}'\mathbf{x}$

5. Spatial depth was defined by Gao (2003) as,

$$PD(F, \mathbf{x}) = 1 - \| \; E(S(\mathbf{x} - \mathbf{y})) \; \|^2$$

where $\mathbf{y} \sim F$ and $S(\mathbf{x}) = \begin{cases} \frac{\mathbf{x}}{\|\mathbf{x}\|} & \mathbf{x} \neq 0, \\ 0 & \mathbf{x} \leqslant 0. \end{cases}$

5.2 Nonparametric Control charts

In this section we will discuss the statistics that can be derived from data depth which are suitable for plotting control charts. Let us assume a sample $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m$ be m random compositions from certain distribution G then the relative rank of $\mathbf{y}$ is,

$$r_G(\mathbf{y}) = P\{D_G(\mathbf{y}) < D_G(\mathbf{y}) \mid \mathbf{y} \sim G\}$$

where $D_G(\mathbf{y})$ is the depth of $\mathbf{y}$ under certain distribution G.

$$r_{G_m}(\mathbf{y}) = \#\{\mathbf{y}_j \mid D_{G_m}(\mathbf{y}_j) < D_{G_m}(\mathbf{y}), \quad j = 1, 2, \ldots, m\}/m \tag{5.2}$$

where $r_G(\mathbf{y})$ is an estimation of the probability of points that are more outlying than the given point $\mathbf{y}$ in the space $\mathbb{R}^k$ under a certain distribution G. It gives the relative outward ranks of $\mathbf{y}$ and $r_{G_m}(\mathbf{y})$ is used for the sample points. If the values of $r_{G_m}(\mathbf{y})$ are very small, it shows that $\mathbf{y}$ comes from the same distribution like $\mathbf{y}_i$'s. After representing all the observations by their relative ranks $r_{G_m}(\mathbf{y})$, we can construct control charts based on these ranks.

5.2.1 r charts

The r charts deals with the relative ranks of $\{\mathbf{x}_1, \mathbf{x}_2, \ldots\}$ instead of dealing directly with the values of $\mathbf{x}_i$'s. Here $r_{G_m}(\mathbf{y})$ will be plotted against the time $i = 1, 2, \ldots$ with a center line $\text{CL} = 0.5$ and the lower control limit $\text{LCL} = \alpha$. If the value of $r_{G_m}(\mathbf{y})$ is less than LCL it means that the process is out-of-control and the distribution of $\mathbf{x}$ is not same as the distribution of reference sample $\mathbf{y}$. If the values of $r_{G_m}(\mathbf{y})$ are larger than CL then we can conclude that the distribution of $\mathbf{x}$ is the same as the distribution of $\mathbf{y}$

5.2.2 Q charts

According to Liu (1995) the multivariate chart Q is the equivalent of the univariate $\bar{\mathbf{x}}$ chart, in which the test statistic is the subgroup average of the statistics $r_{G_m}(\mathbf{x})_i$ taking a fixed subgroup size. Let us assume q subgroup size. The average of $r_{G_m}(\mathbf{x})_i$ is denoted by $Q(G_m, F_q^j)$ where F_q^j is the empirical distribution of $\mathbf{x}_i$ in the jth subgroup with $j = 1, 2, \ldots$. The Q statistics can be defined as

$$Q(G, F) = P\{D_G(\mathbf{y}) \leq D_G(\mathbf{x}) | \mathbf{y} \sim G, \mathbf{x} \sim F\} \tag{5.3}$$

with sample approximation of Q being equals to,

$$Q(G_m, F_q) = \frac{1}{q} \sum_{i=1}^{q} r_{G_m}(\mathbf{x}_i) \tag{5.4}$$

For the Q chart, we have $\text{CL} = 0.5$ and,

$$\text{LCL} = \frac{(\alpha q!)^{1/q}}{q} \tag{5.5}$$

5.2.3 DDMA charts

Liu et al. (2004) defined the DDMA chart. Let us assume that a reference sample $\mathbf{y} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m\}$ follows a distribution G and another sample of new observations $\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ follows another distribution F. So the DDMA charts can be used to deal with moving averages of the new observed sample, where q is the length of the moving average, such that $\widetilde{\mathbf{x}}_q = (\mathbf{x}_1 + \ldots + \mathbf{x}_q)/q, \quad \widetilde{\mathbf{x}}_{q+1} = (\mathbf{x}_2 + \ldots + \mathbf{x}_{q+1})/q, \quad \ldots, \quad \widetilde{\mathbf{x}}_n = (\mathbf{x}_{n-q+1} + \ldots + \mathbf{x}_n)/q$. Let $\widetilde{\mathbf{x}} = (\widetilde{\mathbf{x}}_q, \widetilde{\mathbf{x}}_{q+1}, \ldots, \widetilde{\mathbf{x}}_n)$. Then the corresponding reference sample will be $\widetilde{\mathbf{y}}_q = (\mathbf{y}_1 + \ldots + \mathbf{y}_q)/q, \quad \widetilde{\mathbf{y}}_{q+1} = (\mathbf{y}_2 + \ldots + \mathbf{y}_{q+1})/q, \quad \ldots, \quad \widetilde{\mathbf{y}}_m = (\mathbf{y}_{m-q+1} + \ldots + y_m)/q$. Let $\widetilde{\mathbf{y}} = (\widetilde{\mathbf{y}}_q, \widetilde{\mathbf{y}}_{q+1}, \ldots, \widetilde{\mathbf{y}}_m)$. For each $\widetilde{\mathbf{x}}_i \in \widetilde{\mathbf{x}}$, we calculate the relative ranks $(\widetilde{\mathbf{y}}_q, \widetilde{\mathbf{y}}_{q+1}, \ldots, \widetilde{\mathbf{y}}_m)$, that is

$$r_{\widetilde{G}_{m-q+1}}(\widetilde{\mathbf{x}}_i) = \frac{\#\{\widetilde{\mathbf{y}}_j | D_{\widetilde{G}_{m-q+1}}(\mathbf{y}_j) < D_{\widetilde{G}_{m-q+1}}(\mathbf{x}_i), \quad j = q, \ldots, m\}}{m - q + 1} \tag{5.6}$$

where $\widetilde{G}_{m-q+1}$ is the empirical distribution of $\widetilde{\mathbf{y}}$ and $D_{\widetilde{G}_{m-q+1}}(\cdot)$ is the empirical depth computed with respect to $\widetilde{G}_{m-q+1}$. The DDMA chart is simply the plot of $r_{\widetilde{G}_{m-q+1}}(\widetilde{\mathbf{x}}_i)$ along with the indices $i = q, \ldots, n$. Similarly to the r chart, the LCL of the DDMA control chart is equal to α.

5.3 Performance of the control charts

The goal of this section is to investigate the performance of all the three control charts defined above for the specific case of CoDa. As for the other chapters, we assume a $p = 3$ part composition. Three different multivariate distributions are taken into consideration for the ilr coordinates i.e the multivariate normal , exponential and Cauchy distributions. The pdf of these distributions are as follows,

1. The bivariate normal distribution

$$f(x_1, x_2) = \frac{1}{(2\pi)\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}\right]\right)$$

where ρ is the correlation between x_1 and x_2 and where $\sigma_1 > 0$ and $\sigma_2 > 0$.

2. The bivariate exponential distribution

$$f(x_1, x_2) = \left((1-\theta)\lambda_1\lambda_2 + \theta{\lambda_1}^2\lambda_2\left(x_1 + \frac{1}{\lambda_1}\right) + \theta\lambda_1{\lambda_2}^2\left(x_2 + \frac{1}{\lambda_2}\right) + \theta^2{\lambda_1}^2{\lambda_2}^2\left(x_1 + \frac{1}{\lambda_1}\right)\left(x_2 + \frac{1}{\lambda_2}\right)\right)$$
$$\exp\left(\lambda_1\left(x_1 + \frac{1}{\lambda_1}\right) + \lambda_2\left(x_2 + \frac{1}{\lambda_2}\right) + \theta\lambda_1\lambda_2\left(x_1 + \frac{1}{\lambda_1}\right)\left(x_2 + \frac{1}{\lambda_2}\right)\right)$$

3. The bivariate Cauchy distribution

$$f(x_1, x_2) = \frac{1}{2\pi}\left[\frac{\gamma}{((x_1 - x_{1,0})^2 + (x_2 - x_{2,0})^2 + \gamma^2)^{1.5}}\right]$$

where $x_{1,0}$ and $x_{2,0}$ are the location parameters and γ is the scale parameter.

First 1000 random samples of $p = 3$ part CoDa are generated using the above mentioned distributions with mean $\boldsymbol{\mu}^* = [0, 0]$ and, similar to Zaidi et al. (2019), the following two situations are considered for the ilr CoDa variance-covariance matrix $\boldsymbol{\Sigma}^*$

Case #1 uncorrelated case

$$\boldsymbol{\Sigma}^* = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

Case #2 correlated case

$$\boldsymbol{\Sigma}^* = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}.$$

These 1000 samples are being considered as the reference sample to find the data depth. To simulate the ARL of the control charts 10000 new samples of size 40 are being generated, among each sample 20 observations are from the original distribution and the other 20 observations are from a shifted distribution. Then the simplicial depths of all the samples have been calculated with respect to the reference sample and then using the depth values of the ilr transformed values of CoDa instead of the original values we can simulate the performance of the above mention charts.

To find the ARL we first select the LCL of all the three charts. For the r and DDMA charts, we have $\mathrm{LCL} = 0.0027 = \alpha$ while, for the Q chart using Equation 5.5 we have $\mathrm{LCL} = 0.01596$. Using these control limits we can simulate the ARL of all the three charts. For the Q chart $q = 5$ has been taken, also for the DDMA chart $q = 5$ has been studied.

The different shifts in the means being considered are as follows,

Table 5.1 – ARL values

	With Correlation								
Shifted mean	r chart			Q chart			DDMA chart		
	Normal	Exponential	Cauchy	Normal	Exponential	Cauchy	Normal	Exponential	Cauchy
$[0,0]$	19.23	17.98	19.42	7.93	7.93	7.93	14.09	9.44	15.11
$[0.5,0.5]$	21.16	10.42	19.63	7.93	7.40	7.93	10.38	10.60	17.71
$[0.5,0.0]$	21.18	17.38	19.15	7.94	7.89	7.93	10.09	9.77	16.01
$[1.0,0.5]$	22.39	10.48	19.14	7.90	7.32	7.93	10.76	9.30	16.39
$[0.0,1.0]$	22.55	17.55	19.02	7.78	7.86	7.93	10.43	9.32	16.01
$[1.0,1.0]$	22.14	9.55	19.60	7.90	7.08	7.93	10.83	9.39	18.07
	Without Correlation								
$[0,0]$	19.78	18.56	19.30	7.95	7.93	7.93	12.20	10.90	20.97
$[0.5,0.5]$	22.02	10.18	19.66	7.95	7.32	7.93	10.90	9.98	22.04
$[0.5,0.0]$	20.73	9.91	19.27	7.97	7.07	7.93	11.23	10.53	20.40
$[1.0,0.5]$	23.43	15.24	19.61	7.93	7.82	7.93	10.87	10.05	23.19
$[0.0,1.0]$	24.39	9.52	19.63	7.85	6.87	7.93	10.92	10.03	22.78
$[1.0,1.0]$	24.36	9.63	19.64	7.92	6.94	7.94	10.32	9.28	21.85

1. From $\boldsymbol{\mu}_0^* = [0,0]$ to $\boldsymbol{\mu}_1^* = [0.5,0.5]$,
2. From $\boldsymbol{\mu}_0^* = [0,0]$ to $\boldsymbol{\mu}_2^* = [0.5,0]$,
3. From $\boldsymbol{\mu}_0^* = [0,0]$ to $\boldsymbol{\mu}_3^* = [1,0.5]$,
4. From $\boldsymbol{\mu}_0^* = [0,0]$ to $\boldsymbol{\mu}_4^* = [0,1]$,
5. From $\boldsymbol{\mu}_0^* = [0,0]$ to $\boldsymbol{\mu}_5^* = [1,1]$.

The values of ARL are listed in Table 5.1

From Table 5.1 we can see that, when the sample is normally distributed the values of ARL for the r chart increases with an increase in the values of the shift so, we can say that in both Cases #1 and #2 the r charts do not give a better performance when the sample is normally distributed. While the ARL for both the Q and DDMA charts decreases with an increase in the values of the shift. So it is clearly visible from Table 5.1 that the Q and DDMA charts performs better in case of normal distribution. When the sample is exponentially distributed the ARL of the r, Q and DDMA charts decreases with an increase in the values of the shift. So it is clearly visible from Table 5.1 that the three charts performs better in both Cases #1 and #2 with an increase in shift.

When the sample follows the Cauchy distribution the ARL of the r, Q and DDMA charts increases with an increase in the values of the shift. So it is clearly visible from Table 5.1 that all the three charts do not performs better in both Cases #1 and #2 with an increase in the values of shift.

5.4 Application of the chart

The goal of this section is to study the application of the control charts defined above using data depth for compositional data. We assume that at time $i = 1, 2, \dots,$ we collect a sample of size $n = 1000$ with $p = 3$ part composition from a multivariate normal distribution with mean $\boldsymbol{\mu}^* = [0,0]$ and variance-covariance $\boldsymbol{\Sigma}^* = (1, 0.05; 0.05, 1)$. We use this sample as reference sample. Then we generate another sample of size $n = 40$ with $p = 3$ from a multivariate normal distribution, among these $n = 40$ values 20 are with same mean and variance like the reference sample while the other 20 are with a shift in the mean from $\boldsymbol{\mu}^* = [0,0]$ to $\boldsymbol{\mu}^* = [1,1]$ and same variance. The first step is to transform the data from the simplex sample space $\mathcal{S}^p$ to the real space $\mathbb{R}^{p-1}$ using the ilr transformation, then the data depth of the samples have been computed using the simplicial depth given in Equation 5.1. Then using Equation 5.2 values of the r statistic have been obtained.

Firstly for the r charts, the statistics r is directly being plotted. Secondly for the Q chart we use the Equation 5.4. We have presented the Q chart with four different possibilities for average $q = 2, 3, 5$ and 10. Similarly we have also plotted the DDMA charts with four different possibilities for the moving average parameter $q = 2, 3, 5$ and 10. For the sake of comparison of the Q and DDMA charts we have plotted both the charts in Figure 5.1 on the next page using different values of averages together to see the performance of these ones more clearly.

From Figure 5.1 on the facing page we can clearly see that when the value of the subgroup size $q = 2$ and the length of the moving average $q = 2$ is selected, the DDMA charts shows more out of control points than the Q charts. The Q chart shows the first out of control point on the 40th observation while the DDMA chart shows it on the 17th observation. On total there are five out of control points in the DDMA chart while just one out of control point in the Q chart. Similarly when the value of the subgroup size $q = 3$ and the length of the moving average $q = 3$ are selected the Q chart gives no out of control point, while the DDMA chart shows six out of control points with the first point on the 18th observation. Also for $q = 10$, the Q chart gives no out of control point, while the DDMA chart shows six out of control points with the first point on the 25th observation. So we can conclude that the DDMA charts are more efficient in detecting out of control points than the Q charts.

Next we plotted the r chart along with the DDMA chart with the length of the moving average $q = 2$ to see the efficiency of the DDMA chart over the r chart and we can see in Figure 5.2 on page 82 that the r chart shows two out of control points with the first point at the 21st observation while the DDMA chart shows five out of control points with the first one on the 17th observation.

5.5 Illustrative Example

A pathologist is interested in introducing a new method to determine the composition of white blood cells in his laboratory (i.e granulocytes, lymphocytes, monocytes). First he used a predetermined method known as microscopic inspection for 30 samples of white blood cells and he recorded the 3-part compositions and then he used a quick new method known as image analysis on another set of 30 samples. The data has been taken from Aitchison (1986, page 366, Data 11) . Here we used the first sample of 30 compositions as the reference sample as it is known to be an accurate procedure to study the composition of white blood cells and the other samples of 30 compositions as a new sample under study. First, we will apply the normality test on the sample data to see if the data is normally distributed or not. For this purpose we have applied the Henze-Zirkler test for multivariate normality on the new data set and we have found that the p-value of Henze-Zirkler test comes out to be 0.000080 , so we can say that the data are not normally distributed as the p-values is less than the minimum threshold (i.e 0.05). In this case, the Hotelling T^2 chart is not reliable as the data are not normally distributed so we have all the three charts, with the Q chart using the subgroup sizes 2 and 5 as well as the DDMA chart using the length of the moving average to be 2 and 5 along with the values of the Hotelling T^2 statistics.

We fixed $\alpha = 0.05$ as, if we take smaller than that, the lower control limit will be very small for the DDMA charts LCL $= \alpha$, so by selecting LCL $= 0.05$ we can clearly see from Table 5.2 on page 82 that the r chart shows six out of control points having the first point on the first observation, while for the Q chart , using Equation 5.5, hence LCL $= 0.158$ (when $q = 2$) and LCL $= 0.286$ (when $q = 5$). From Table 5.2 on page 82 we can see that the Q chart shows only two out of control points for both $q = 2$ and $q = 5$ with the first out of control point on the 10th observation. In case of the DDMA charts, when $q = 2$ the chart shows a total of sixteen out of control points with the first one on the first observation. Also when $q = 5$ has been selected the DDMA chart gives fourteen out of control points with the first one on the first observation. This can also be seen in Figure 5.3 on page 84. For the sake of comparison we have also plotted the Hotelling T^2

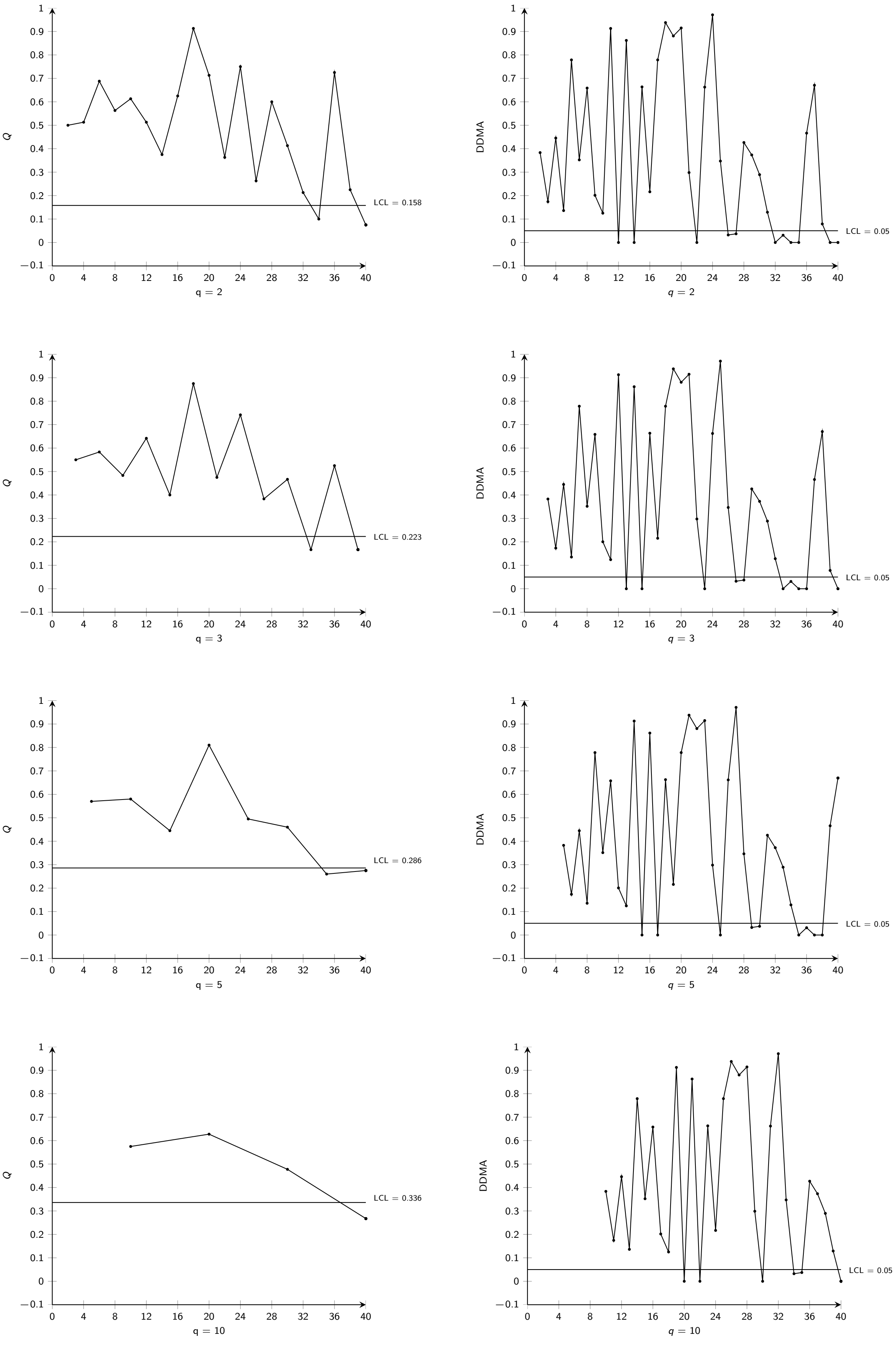

Figure 5.1 – Q chart and DDMA charts with different values of q

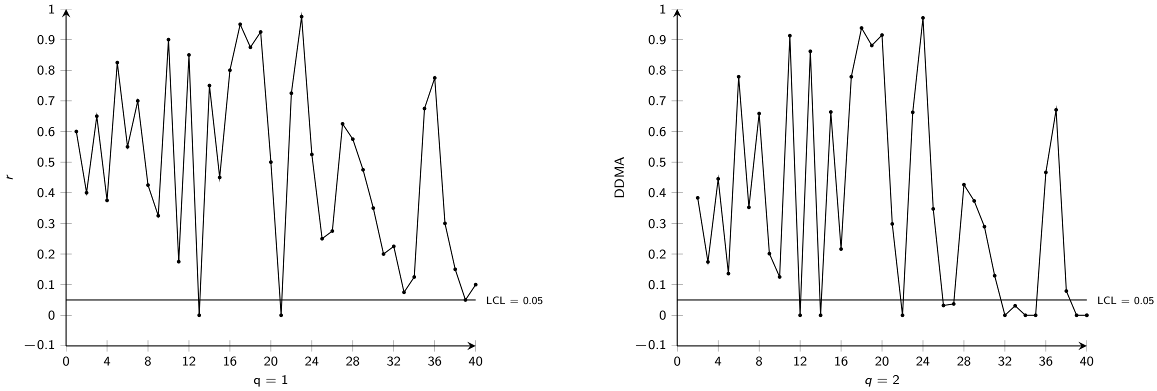

Figure 5.2 – r chart and DDMA chart

Table 5.2 – White cell compositions of 30 blood cells

i	$\mathbf{x}_i$			$\mathbf{x}_i^*$		r	$Q(2)$	$Q(5)$	DDMA(2)	DDMA(5)	T^2
1	0.73	0.26	0.01	2.93	0.74	0.00	-	-	-	-	1.79
2	0.66	0.28	0.06	1.67	0.61	0.90	0.62	-	0.00	-	0.01
3	0.73	0.21	0.07	1.45	0.86	0.80	-	-	0.69	-	0.10
4	0.81	0.18	0.02	2.44	1.08	0.43	0.67	-	0.66	-	0.80
5	0.62	0.35	0.03	2.27	0.40	0.70	-	0.63	0.31	0.00	0.55
6	0.86	0.11	0.03	1.88	1.43	0.93	0.62	-	0.34	0.65	0.70
7	0.96	0.03	0.01	2.10	2.45	0.33	-	-	0.34	0.62	4.05
8	0.93	0.05	0.02	1.96	2.02	0.53	0.18	-	0.00	0.31	2.29
9	0.90	0.07	0.03	1.90	1.79	0.63	-	-	0.00	0.35	1.54
10	0.94	0.06	0.01	2.64	2.00	0.00	0.13	0.25	0.00	0.35	3.08
11	0.87	0.11	0.02	2.49	1.44	0.37	-	-	0.00	0.00	1.36
12	0.45	0.52	0.03	2.22	-0.11	0.57	0.67	-	0.00	0.00	1.28
13	0.24	0.74	0.02	2.34	-0.79	0.00	-	-	0.69	0.00	3.59
14	0.48	0.47	0.05	1.79	0.00	0.67	0.60	-	0.00	0.00	0.68
15	0.32	0.66	0.02	2.60	-0.52	0.00	-	0.55	0.69	0.00	3.11
16	0.46	0.52	0.02	2.53	-0.08	0.20	0.40	-	0.00	0.65	1.74
17	0.38	0.25	0.37	-0.15	0.28	0.30	-	-	0.31	0.00	4.08
18	0.44	0.24	0.32	0.01	0.43	0.43	0.22	-	0.00	0.65	3.30
19	0.58	0.14	0.28	0.04	1.00	0.50	-	-	0.00	0.00	3.33
20	0.40	0.17	0.43	-0.42	0.61	0.27	0.00	0.20	0.00	0.31	5.12
21	0.80	0.15	0.04	1.69	1.18	0.83	-	-	0.00	0.00	0.28
22	0.66	0.26	0.08	1.32	0.65	0.73	0.82	-	0.69	0.00	0.15
23	0.73	0.22	0.06	1.59	0.85	0.77	-	-	0.59	0.00	0.04
24	0.65	0.30	0.05	1.74	0.55	0.97	0.87	-	0.66	0.00	0.04
25	0.37	0.17	0.46	-0.51	0.57	0.00	-	0.67	0.72	0.65	5.60
26	0.18	0.80	0.02	2.28	-1.08	0.00	0.23	-	0.00	0.58	4.74
27	0.33	0.63	0.05	1.89	-0.46	0.40	-	-	0.00	0.62	1.87
28	0.43	0.51	0.06	1.65	-0.13	0.60	0.53	-	0.38	0.69	0.92
29	0.94	0.05	0.01	2.40	2.14	0.23	-	-	0.00	0.00	3.12
30	0.86	0.11	0.03	1.84	1.47	0.87	0.45	0.49	0.00	0.00	0.76

chart to see if nonparametric control charts performs better than when the data is not normally distributed. For UCL of Hotelling T^2 chart we know that $\mathrm{UCL} = F_{\chi^2}^{-1}(1-\alpha_0|p-1)$ so $\mathrm{UCL} = 5.99$ using $\alpha = 0.05$. We can see from Table 5.2 on the preceding page that there is no out of control point in the data. From this study we can conclude that the new method introduced by the the pathologist does not seems to be very effective as there are many out of control points using the sample with predetermined accurate observations as the reference sample. In conclusion, there is no evidence of substituting the microscopic inspection with image analysis.

5.6 Conclusions

In this chapter, nonparametric control charts has been investigated for compositional data. These charts have been previously used on regular multivariate data. This chapter makes an attempt to use these non-parametric control charts on compositional data. To study the performance of all the three charts along with their efficiency, three different multivariate distributions i.e. normal, exponential and Cauchy have been studied with two different cases for the variance-covariance matrix, one uncorrelated and the other one correlated. Different combinations of the mean shift have been selected in order to study the performance of these charts. The main conclusion drawn from the performance study is that the DDMA and Q charts perform better in the case of a normal distribution, while in the case of the Cauchy distribution all the three charts does not perform very well while, in the case of the exponential distribution, all the three charts perform better when the shift in mean increases. In case of a Cauchy distribution none of all the three charts under study gives a good performance with different values of the shift. Future research about control charts monitoring CoDa should be focused on another new nonparametric chart that can perform better in case of Cauchy distribution, on studying the control charts on compositional data when some of the information is missing or, on studying the effect of the estimation of parameters on the statistical properties of control charts.

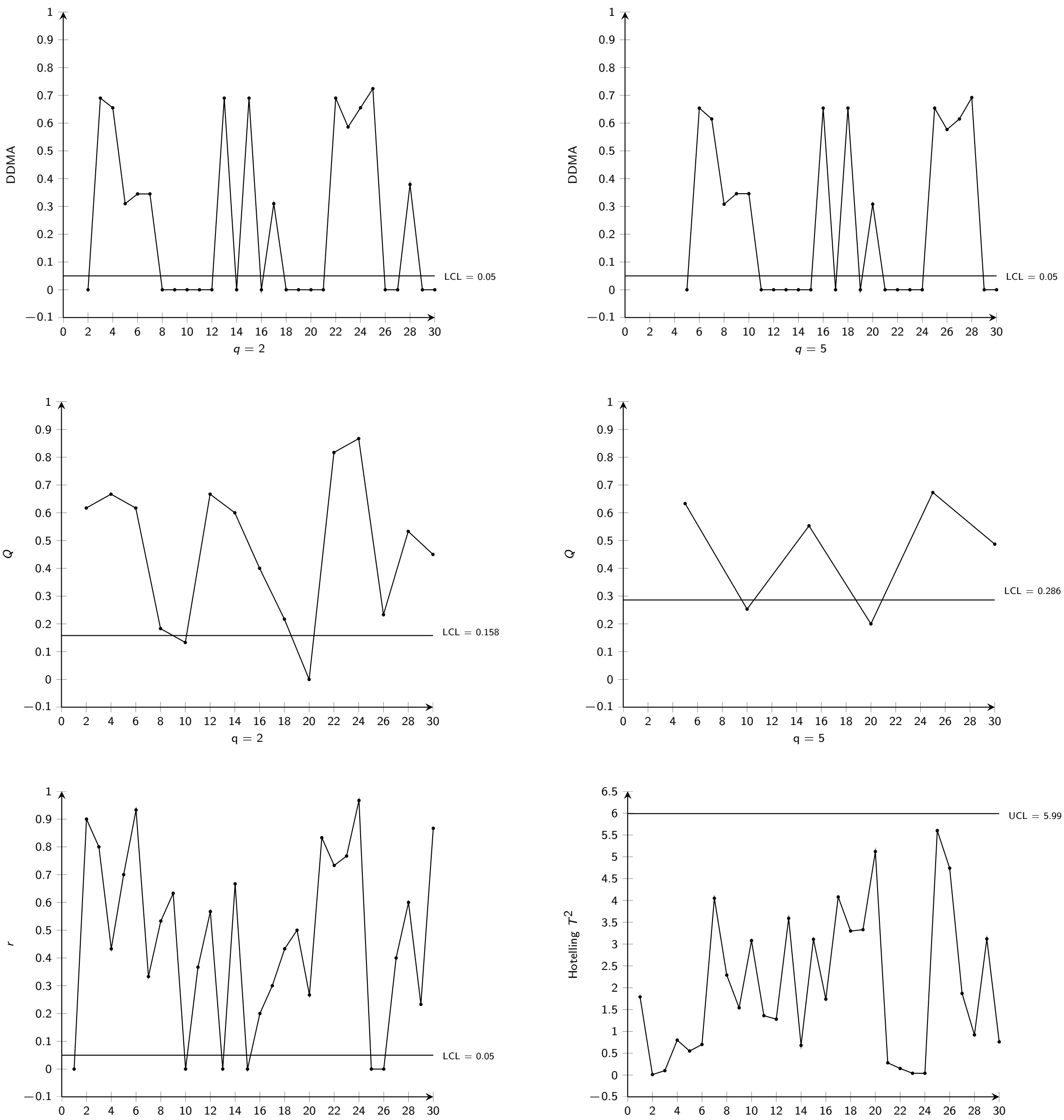

Figure 5.3 – DDMA ,Q , r and Hotelling T^2 charts

6

General Conclusions & Perspectives

6.1 General Conclusions

During the last decade, an enormous number of new advanced control charts have been proposed for univariate and multivariate processes with many applications in manufacturing and service sectors. Chapter 1 of this thesis is divided into three major parts. Part one consists of Univariate control charts, in the second part we discussed some nonparametric control charts and the third part describes some multivariate control charts.

In the case of continuous multivariate processes, the vast majority of control charts assumes that the data are unconstrained. But there is a specific category of multivariate data which are constrained by definition. This kind of data is called CoDa and they are represented by vectors whose strictly positive components only convey relative information. Chapter 2 deals with the literature of CoDa. This chapter is also divided into three major parts: in first part we have an introduction of compositional data along with principles and geometry of CoDa , in second part we discuss the transformations of CoDa from the simplex sample space to the real space, all the three main types of transformation have been discussed, while in third part we have discussed some of the previous researches about control charts on compositional data.

Concerning the statistical process monitoring of CoDa, there are few contributions by now. Even though many researches have been conducted on the evaluation of control charts in the presence of measurement errors, few of them have been devoted to multivariate data and, as far as we know, none of them have been devoted to CoDa. Chapter 3 deals with the effect of measurement errors on control chart for CoDa. The goal of this chapter is to impose measurement error on the T^2_C control chart proposed by Vives-Mestres et al. (2014b). A linearly covariate measurement error model for CoDa has been presented, where the quality characteristics $\mathbf{y}_i$ is not directly observable and can only be assessed using several independent measurements $\mathbf{x}_{i,j}$. Two situations have been considered for the CoDa variance-covariance matrix: a correlated case and an uncorrelated one. Different combinations of the parameters involved in the model have been selected in order to study their influence on the performance of the Hotelling CoDa T^2 control chart. The main conclusions drawn from these investigations are if b and m are kept constant, the ARL increases when the value of σ_{M} increases, and if σ_{M} and m are kept constant, the ARL decreases as the value of b increases, if σ_{M} and b are kept constant, the ARL decreases (slowly) as the value of m increases. At the end of the Chapter 3 a realistic illustrative example based on the production of muesli is used to illustrate the estimation of the measurement device parameters and the in-control process parameters, as well as to demonstrate the ability of this control chart to efficiently detect changes in the muesli composition.

The Chapter 4 can be considered as an extension of the work of TRAN et al. (2018) (for monitoring compositional data using a multivariate EWMA, i.e. a MEWMA-CoDa chart) by taking into account potential measurement errors that are known to highly affect production processes. Similar to Chapter 3 a linearly covariate error model with a constant error variance is also used in this chapter to study the impact of measurement errors on the MEWMA-CoDa control chart. For the variance covariance matrix of CoDa, four cases have been taken into account to study the influence of the involved parameters. The main conclusions are, σ_{M} has negative impact on the performance of the MEWMA-CoDa control chart keeping b and m constant, b has positive impact on the performance of the MEWMA-CoDa control chart keeping σ_{M} and m constant, m has positive but mild impact on the performance of MEWMA-CoDa control chart keeping σ_{M} and b constant. A comparison with the Hotelling CoDa T^2 control chart also proved that the MEWMA-CoDa control chart is between 70% to 90% more efficient than the Hotelling CoDa T^2 control chart in the presence of measurement errors. At the end of Chapter 4 a real life example of muesli production, using multiple measurements for each composition is presented in order to estimate the parameters and also to demonstrate how the MEWMA-CoDa can handle measurement errors to detect shifts in the process.

As discussed above, there are very few contributions that deals with control charts for CoDa. A vast majority of control charts deals with the assumption of multivariate-normality of the data, but it is not the case in many real world examples. Most of the time the data does not fulfill the assumption of normality, and there is a need to investigate some non parametric charts, in which the knowledge of the underlying distribution of the data is not required. In Chapter 5 we have investigated the performance of some nonparametric charts on compositional data using data depth. These charts have been previously used on multivariate data. In this chapter we have made an attempt to use the predefined nonparametric control charts on compositional data. Three charts have been used, the r chart, Q chart and DDMA charts. To study the performance of all the three charts along with their efficiency, three different multivariate distributions normal, exponential and Cauchy have been studied with two different cases for the variance-covariance similar to the previous chapters. Different combinations of the mean shift have been selected in order to study the effect of the shift on the performance of these charts. The main conclusion drawn from the performance study is that DDMA and Q charts perform better in the case of the normal distribution, while in case of the Cauchy distribution all the three charts does not perform very well. While in the case of the exponential distribution, all the three charts perform better when the shift in mean increases. In case of the Cauchy distribution none of the charts under study give a good performance no matter the values of shift.

6.2 Perspectives

As discussed in the previous chapters there is a big gap for control charts for CoDa in the literature so there are many different possibilities of carrying several research works on control charts for CoDa. Some of the good ideas to work on this aspect can be,

1. MCUSUM type chart for CoDa.
2. MCUSUM type chart for CoDa in presence of measurement error.
3. Study of the effect of the estimation of parameters on the statistical properties of control charts.
4. An extension to T^2 chart with subgroup $(n > 1)$.
5. Charts for missing values in CoDa.

Furthermore, in this thesis we have discussed just three types of nonparametric control charts for CoDa using data depth. Another idea can be to develop nonparametric control charts for CoDa directly without using the data depth. Also there is possibility to study the same charts by using CoDa itself instead of using the transformation of data.

References

T. Abbas, S. A. Abbasi, M. Riaz, and Z. Qian. Phase II Monitoring of Linear Profiles with Random Explanatory Variable under Bayesian Framework. *Computers & Industrial Engineering*, 127:1115 – 1129, 2019. 51

B. M. Adam, C. Lowry, and W. H. Woodall. *The Use (and Misuse) of False Alarm Probabilities in Control Chart Design in Frontier in Statistical Qaulity Control*. Physica Verlag, Heidelberg, 1992. 15

J. Aitchison. *The Statistical Analysis of Compositional Data. Monographs on Statistics and Applied Probability*. Reprinted 2003 with additional material by Blackburn Press. Chapman and Hall Ltd, London(UK), 1986. 21, 22, 25, 80

J. Aitchison. *The Statistical Analysis of Compositional Data*. Monographs on Statistics and Applied Probability. Springer Netherlands, 2011. 24

A. R. Ansari and R. A. Bradley. Rank Sum Test for Dispersion. *Annals of Mathematical Statistics*, 31: 1174–1189, 1960. 18

L. A. Arorian and H. Levene. The Effectivenes of Quality Control Charts. *Journal of American Statistical Association*, 45:520–529, 1950. 20

S. J. Bae, G. Do, and Paul P. Kvam. On Data Depth and the Application of Nonparametric Multivariate Statistical Process Control Charts. *Applied Stochastic Models in Business and Industry*, 32(5):660–676, 2016. 75

S. M. Barale and D. T. Shirke. Nonparametric Control Charts Based on Data Depth for Location Parameter. *J Stat Theory*, 13(41), 2019. 75

C.A. Bennett. Effect of Measurement Error on Chemical Process Control. *Industrial Quality Control*, 10 (4):17–20, 1954. 33

C. Bilen, A. Khan, and W. Chattinnawat. Dual-Monitoring Scheme for Multivariate Autocorrelated Cascade Processes with EWMA and MEWMA Charts. *Quality Technology & Quantitative Management*, 14(2): 156–177, 2017. 51

D. Billheimer, P. Guttorp, and W.F. Fagan. Statistical Interpretation of Species Composition. *Journal of the American Statistical Association*, 96(456):1205–1214, 2001. 26

R.A. Boyles. Using the Chi-Square Statistic to Monitor Compositional Process Data. *Journal of Applied Statistics*, 24(5):589–602, 1997. 27

P. Castagliola, G. Celano, and S. Psarakis. Monitoring the Coefficient of Variation Using EWMA Charts. *Journal of Quality Technology*, 43(3):249–265, 2011. 55

S. Chakraborti and M. A. Van de Wiel. A New Nonparametric Control Chart Based on the MannâĂŞWhitney Statistic. *IMS Collections*, 1:156âĂŞ–172, 2008. 18

F. Chayes. On Correlation Between Variables of Constant Sum. *Journal of Geophysical Research*, 65(12): 4185–4193, 1960. 21

X. B. Cheng and F. K. Wang. The Performance of EWMA Median and CUSUM Median Control Charts for a Normal Process with Measurement Errors. *Quality and Reliability Engineering International*, 34 (2):203–213, 2018a. 33

X. B. Cheng and F. W. Wang. VSSI Median Control Chart with Estimated Parameters and Measurement Errors. *Quality and Reliability Engineering International*, 34(5):867–881, 2018b. 34

A. F. B. Costa and M. A. Rahim. A Single EWMA Chart for Monitoring Process Mean and Process Variance. *Quality Technology & Quantitative Management*, 3(3):295–305, 2006. 51

R. B. Crosier. Multivariate Generalizations of Cumulative Sum Quality Control Schemes. *Technometrics*, 30(3):291–303, 1988. 18, 19

N. Das. A Note Efficiency of Nonparametric Control Chart for Monitoring Process Variablility. *Economic Quality Control*, 23(1):85–93, 2008. 18

A. K. Dizabadi, M. Shahrokhi, and M. R. Maleki. On the Effect of Measurement Error with Linearly Increasing-Type Variance on Simultaneous Monitoring of Process Mean and Variability. *Quality and Reliability Engineering International*, 32(5):1693–1705, 2016. 33

D. L. Donoho and M. Gasko. Breakdown Properties of Location Estimates Based on Halfspace Depth and Projected Outlyingness. *The Annals of Statistics*, 20(4):1803–1827, 12 1992. 75

E.K. Epprecht, F. Aparisi, and O. Ruiz. Optimum Variable-Dimension EWMA Chart for Multivariate Statistical Process Control. *Quality Engineering*, 30(2):268–282, 2018. 51

W. D. Ewan. When and How to Use Cusum Charts. *Technometrics*, 5(1):1–22, 1963. 15

F. F. Gan. An Optimal Design of CUSUM Quality Control Charts. *Journal of Quality Technology*, 23(4): 279–286, 1991. 15

Y. Gao. Data Depth Based on Spatial Rank. *Statistics & Probability Letters*, 65(3):217–225, 2003. 76

J. D. Gibbons and S. Chakraborti. *Nonparametric Statistical Inference*. CRC Press, Boca Raton, FL., 5th edition, 2010. 17

M. Gnanadesikan and S. S. Gupta. A Selection Procedure for Multivariate Normal Distributions in Terms of the Generalized Variances. *Technometrics*, 12(1):103–117, 1970. 55

M. A. Graham, S. Chakraborti, and S. W. Human. A Nonparametric Exponentially Weighted Moving Average Signed-Rank Chart for Monitoring Location. *Computational Statistics & Data Analysis*, 55(8): 2490 – 2503, 2011. ISSN 0167-9473. 51

D. M. Hawkins, S. Choi, and S. Lee. A General Multivariate Exponentially Weighted Moving-Average Control Chart. *Journal of Quality Technology*, 39(2):118–125, 2007. 51

H. Hotelling. *Multivariate Quality Control, Techniques of Statistical Analysis*. McGraw-Hill, 1947. 19

X. Hu, P. Castagliola, J. Sun, and M. B. C. Khoo. The Performance of Variable Sample Size $\bar{X}$ Chart with Measurement Errors. *Quality and Reliability Engineering International*, 32(3):969–983, 2016a. 33

X. Hu, P. Castagliola, J. Sun, and M. B. C. Khoo. Effect of Measurement Errors on the VSI $\bar{X}$ Chart. *European Journal of Industrial Engineering*, 10(2):224–242, 2016b. 33

X. L. Hu, P. Castagliola, J. S. Sun, and M. B. C. Khoo. Economic Design of the Upper-sided Synthetic Chart with Measurement Errors. *International Journal of Production Research*, 54(19):5651–5670, 2016c. 33

S. Idris, L. Wachidah, T. Sofiyayanti, and E. Harahap. The Control Chart of Data Depth Based on Influence Function of Variance Vector. *Journal of Physics: Conference Series*, 1366:012125, 2019. 75

M. B. C. Khoo. Increasing the Sensitivity of Multivariate EWMA Control Chart. *Quality Engineering*, 16 (1):75–85, 2003. 51

H. Khusna, M. Mashuri, Suhartono, D. D. Prastyo, and M. Ahsan. Multioutput Least Square SVR Based Multivariate EWMA Control Chart. *Journal of Physics: Conference Series*, 1028(1):012221, 2018. 51

M. H. Lee and M. B. C. Khoo. Optimal Statistical Design of a Multivariate EWMA Chart Based on ARL and MRL. *Communications in Statistics-Simulation and Computation*, 35(3):831–847, 2006. 54

K. Linderman and T. E. Love. Economic and Economic Statistical Designs for MEWMA Control Charts. *Journal of Quality Technology*, 32(4):410–417, 2000. 51

K. W. Linna and W. H. Woodall. Effect of Measurement Error on Shewhart Control Charts. *Journal of Quality Technology*, 33(2):213–222, 2001. 33, 34

K. W. Linna, W. H. Woodall, and K. L. Busby. The Performance of Multivariate Control Charts in the Presence of Measurement Error. *Journal of Quality Technology*, 33(3):349–355, 2001. 33, 34, 35, 36, 53

J. Liu, C. Tan, and C. Zeng. Parameter Optimization for a Modified MEWMA Control Chart Based on a PSO Algorithm. *Communications in Statistics - Simulation and Computation*, 46(5):4140–4153, 2017. 51

R. Y. Liu. On a Notion of Data Depth Based on Random Simplices. *Annals of Statistics*, 18(1):405–414, 03 1990. 76

R. Y. Liu. Control Charts for Multivariate Processes. *Journal of the American Statistical Association*, 90 (432):1380–1387, 1995. 77

R. Y. Liu, K. Singh, and J. H. Teng. DDMA Charts: Nonparametric Multivariate Moving Average Control Charts Based on Data Depth. *Allgemeines Statistisches Archiv*, 88(2):235–258, 2004. 75, 77

C. A. Lowry and D. C. Montgomery. A Review of Multivariate Control Charts. *IIE Transactions*, 27(6): 800–810, 1995. 18

C. A. Lowry, W. H. Woodall, C. W. Champ, and S. E. Rigdon. A Multivariate Exponentially Weighted Moving Average Control Chart. *Technometrics*, 34(1):46–53, 1992a. 18, 28

C.A. Lowry, W.H. Woodall, C.W. Champ, and S.E. Rigdon. A Multivariate Exponentially Weighted Moving Average Control Chart. *Technometrics*, 34(1):46–53, 1992b. 19, 51, 52, 53

J. M. Lucas and M. S. Saccucci. Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements. *Technometrics*, 32(1):1–12, 1990. 54

F. K. Lutgens and E. J. Tarbuck. *Essentials of Geology*. Upper Saddle River, NJ : Prentice Hall, 7th ed edition, 2000. 22

P. C. Mahalanobis. On The Generalized Distance in Statistics. *Proceedings of the National Academy India*, 12:49–55, 1936. 75, 76

M. A. Mahmoud and P. E. Maravelakis. The Performance of the MEWMA Control Chart When Parameters are Estimated. *Communications in Statistics - Simulation and Computation*, 39(9):1803–1817, 2010. 51

M. R. Maleki, A. Amiri, and R. Ghashghaei. Simultaneous Monitoring of Multivariate Process Mean and Variability in the Presence of Measurement Error with Linearly Increasing Variance under Additive Covariate Model. *International Journal of Engineering Transactions A: Basics*, 29(4):471–480, 2016. 33

M. R. Maleki, A. Amiri, and P. Castagliola. Measurement Errors in Statistical Process Monitoring: a Literature Review. *Computers & Industrial Engineering*, 103:316–329, 2017. 34

H. B. Mann and D. R. Whitney. On a Test of Whether One of Two Random Variables is Stochastically Larger than the Other. *Annals of Mathematical Statistics*, 18(1):50–60, 1947. 17

K. V. Mardia, J. T. Kent, and J. M. Bibby. *Multivariate Analysis*. Probability and Mathematical Statistics. Academic Press, London, 1979. 35, 36

W. E. Molnau, G. C. Runger, D. C. Montgomery, K. R. Skinner, E. N. Loredo, and S. S. Prabhu. A Program of ARL Calculation for Multivariate EWMA Charts. *Journal of Quality Technology*, 33(4):515, 2001. 55

D. C. Montgomery. *Introduction to Statistical Quality Control*. John Wily and Sons, 4th edition, 2001. 13

D. C. Montgomery and H. M. Wadsworth. Some Techniques for Multivariate Quality Control Applications. *ASQC Technical Conference Transactions Washington*, pages 427–435, 1972. 18

E. S. Page. Continuous Inspection Schemes. *Biometrics*, 41(1):100–115, 1954. 15

V. Pawlowsky-Glahn, J. J. Egozcue, and R. Tolosana-Delgado. *Modeling and Analysis of Compositional Data*. John Wiley & Sons, 2015. 26, 34

K. Pearson. Mathematical Contributions to the Theory of Evulation. On a Form of a Spurious Correlation Which May Arise When Indices are Used in the Measurement of Organs. *Proceedings of the royal society of London*, pages 489–498, 1897. 21

J. J. Pignatiello and G. C. Runger. Comparison of Multivariate CUSUM Charts. *Journal of Quality Technology*, 22(3):173–186, 1990. 18

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. URL `https://www.R-project.org/`. 10

S. W. Roberts. Properties of Control Chart Zone Tests. *Bell System Technical Journal*, 37:83–114, 1958. 15, 51

G. C. Runger and S. S. Prabhu. A Markov Chain Model for the Multivariate Exponentially Weighted Moving Averages Control Chart. *Journal of the American Statistical Association*, 91(436):1701–1706, 1996. 53, 54

W. A. Shewhart. Economic Control of Qality of Manfactured Product. *Bell system technical journal*, 9(2): 364–389, 1930. 9

Z. G. Stoumbos, M. R. Reynolds, T. P. Ryan, and W. H. Woodall. The State of Statistical Process Control as We Proceed into the 21st Century. *Journal of the American Statistical Association*, 95(451):992–998, 2000. 13

G. Talbot. Chocolate and Cocoa Butter Structure and Composition. In Nissim Garti and Neil R. Widlak, editors, *Cocoa Butter and Related Compounds*, pages 1 – 33. AOCS Press, 2012. 22

A. Tang, P. Castagliola, J. Sun, and X. Hu. The Effect of Measurement Errors on the Adaptive EWMA $\bar{X}$ Chart. *Quality and Reliability Engineering International*, 34(4):609–630, 2018. 33

J. Tanner. Fallacy of Per Weight and Per Surface Area Standards, and Their Relation to Spurious Correlation. *Journal of applied phisiology*, 2(1):1–15, 1949. 21

K. P. Tran, P. Castagliola, and G. Celano. Monitoring the Ratio of Two Normal Variables Using EWMA Type Control Charts. *Quality and Reliability Engineering International*, 32(5):1853–1869, 2015. 55

K. P. Tran, P. Castagliola, and G. Celano. The Performance of the Shewhart-RZ Control Chart in the Presence of Measurement Error. *International Journal of Production Research*, 54(24):7504–7522, 2016. 33

K. P. TRAN, H. D. Nguyen, T. Nguyen, and W. Chattinnawat. One-Sided Synthetic Control Charts for Monitoring the Coefficient of Variation with Measurement Errors. In *IEEE International Conference on Industrial Engineering and Engineering Management (IEEM2018)*, 2018. 50, 51, 53, 55, 74, 86

J. W. W. Tukey. Mathematics and The Picturing of Data. *Proceedings of the international congress of mathematicians*, 2:523–531, 1975. 75, 76

M. Vives-Mestres, J. Daunis-I-Estadella, and J.A. Martín-Fernández. Out-of-Control Signals in Three-Part Compositional T^2 Control Chart. *Quality and Reliability Engineering International*, 30(3):337–346, 2014a. 27

M. Vives-Mestres, J. Daunis-I-Estadella, and J.A. Martín-Fernández. Individual T^2 Control Chart for Compositional Data. *Journal of Quality Technology*, 46(2):127–139, 2014b. 27, 28, 34, 35, 50, 85

W. H. Woodall. Control Charts Based on Attribute Data: Bibliography and Review. *Journal of quality technology*, 29(2):172, 1997. 13

X. Wu, R. Miao, Z. Li, J. Ren, J. Zhang, Z. Jiang, and X. Chu. Process Monitoring Research with Various Estimator-Based MEWMA Control Charts. *International Journal of Production Research*, 53(14):4337–4350, 2015. 51

C. H. Yang and F. S. Hillier. *Mean and Varince Control Charts Limit*, volume 2. American Society for Quality Control, 1970. 13

S. F. Yang, H. W. Ho, and M. A. Rahim. Effects of Measurement Error on Controlling Two Dependent Process Steps. *Economic Quality Control*, 22(1):127–139, 2007. 33

Su-Fen Yang, Jheng-Sian Lin, and Smiley W. Cheng. A New Nonparametric EWMA Sign Control Chart. *Expert Systems with Applications*, 38(5):6239 – 6243, 2011. 51

J. Yue and L. Liu. Multivariate Nonparametric Control Chart with Variable Sampling Interval. *Applied Mathematical Modelling*, 52:603 – 612, 2017. ISSN 0307-904X. 75

F. S. Zaidi, P. Castagliola, K. P. Tran, and M. B. C. Khoo. Performance of the Hotelling T^2 Control Chart for Compositional Data in the Presence of Measurement Errors. *Journal of Applied Statistics*, 46(14): 2583–2602, 2019. 50, 64, 78

Y. Zuo. Projection Based Depth Functions and Associated Medians. *Annals of Statistics*, 31(5):1460–1490, 10 2003. 76

Contents

List of Tables

List of Figures

DOCTORAT
BRETAGNE
LOIRE / MATHSTIC

UNIVERSITÉ DE NANTES

Titre : Développement de procédures de suivi statistique de données compositionnelles

Mots clés : Maîtrise statistique des procédés, Carte de contrôle, données compositionnelles, erreur de mesure, nonparamétrique.

Résumé : La Maîtrise Statistique des Procédés (MSP) est une méthodologie largement utilisée, basée sur la mise en œuvre des cartes de contrôle, permettant de s'assurer de la stabilité du processus et d'améliorer sa capabilité grâce à la réduction de la variabilité du processus. La sélection d'une carte de contrôle appropriée dépend du type et de la distribution des données. Lorsqu'il existe plusieurs caractéristiques de qualité, des cartes de contrôle multivariées doivent être adoptées. Mais il existe une catégorie spécifique de données multivariées qui sont contraintes par définition et connues sous le nom de données compositionnelles (CoDa). Le but de cette thèse est de proposer et d'étudiersystématiquement de nouvelles cartes de contrôle pour les données compositionnelles qui n'ont pas encore été proposées jusqu'à présent dans la littérature. La carte de contrôle de Hotelling T2-CoDa en présence d'erreur de mesure et la carte de contrôle MEWMA-CoDa en présence d'erreur de mesure ont été proposées surveiller des données compositionnelles. En outre, certaines méthodes non paramétriques pour la surveillace de données compositionnelles ont également été proposés. Les performances de chaque carte de contrôle ont été étudiées et les paramètres optimaux ont été systématiquement évalués. Des exemples de données compositionnelles réelles ont été utilisés afin d'étudier les performances des cartes proposées.

Title : Development of statistical monitoring procedures for compositional data

Keywords : Statistical control of processes, control chart, composition data, measurement error, nonparametric.

Abstract : Statistical Process Monitoring (SPM) is a widely used methodology, based on the implementation of control charts, for achieving process stability and improving capability through the reduction of the process variability. The selection of a suitable control chart depends on the type and distribution of he data. When there are several quality characteristics, multivariate control charts have to be adopted. But there is a specific category of multivariate data which are constrained by definition and known as Compositional Data (CoDa). This thesis makes an attempt to systematically propose new control charts for the for compositional data that have not yet been proposed so far in the literature. Hotelling T2-CoDa control chart in the presence of measurement error and MEWMA-CoDa control chart in the presence of measurement error has been proposed for compositional data. Furthermore, some nonparametric charts to monitor compositional data has also been proposed. The performance of each control chart has been studied and the optimal parameters have systematically been evaluated. Real life compositional data examples have been used in order to study the performances of the proposed charts.