WILEY

Multivariate process control charts based on the *Lp* **depth**

Giuseppe Pandolfo¹ Carmela Iorio[1](https://orcid.org/0000-0001-5622-4124) Michele Staiano¹ Massimo Aria² Roberta Siciliano¹

¹Department of Industrial Engineering, University of Naples Federico II, Naples, Italy

2Department of Economics and Statistics, University of Naples Federico II, Naples, Italy

Correspondence

Carmela Iorio, Department of Industrial Engineering, University of Naples Federico II, Naples, Italy. Email: carmela.iorio@unina.it

Abstract

Even if large historical dataset could be available for monitoring key quality features of a process via multivariate control charts, previous knowledge may not be enough to reliably identify or adopt a unique model for all the variables. When no specific parametric model turns out to be appropriate, some alternative solutions should be adopted and exploiting non-parametric methods to build a control chart appears a reasonable choice. Among the possible non-parametric statistical techniques, data depth functions are gaining a growing interest in multivariate quality control. Within the literature, several notions of depth are effective for this purpose, even in the case of deviation from the normality assumption. However, the use of the L^p depth for constructing non-parametric multivariate control charts has been surprisingly neglected so far. Hence, the goal of this work is to investigate the behavior the L^p depth in the statistical process control and to compare its performances to those of the Mahalanobis depth, which is often adopted to build depth-based control charts.

KEYWORDS

Q charts, ARL, Mahalanobis depth, non-parametric statistics

1 INTRODUCTION

Several techniques for monitoring production processes can be found within the literature and, among them, control charts play a key role.¹ Generally speaking, they are tools which provide visual information about the process behavior, whose effectiveness relies on very specific models for the data generating process.

Control chart methodology has a very similar approach to the statistical inference procedure.² Indeed, control charts are used to identify non-random behaviors of a production process by monitoring the changes in the distribution of the tested quality characteristics. Shewart's $\left(X,\overline{X}\right)$ and CUSUM charts are very common and widely used for this purpose.

Therefore, multivariate control charts are needed when dealing with more than one variable. Monitoring the process of related variables is usually called a multivariate quality control problem.

Studies of multivariate quality control were first conducted by Hotelling in 1947³ and later many scholars focused on the subject (for more details see the book of Montgomery¹).

Within the literature, control charting process techniques are distinguished into Phase I and Phase II.⁴ The former, also known as retrospective or preliminary phase, employs the charts with the goal of defining if a process is statistically in-control when the first sub-groups were being drawn, while the latter is aimed at testing whether the process is in-control when the future sub-groups were being drawn. In Phase II the process distribution is assumed to be known, and most of the classical applications require the assumption that the process under inspection follows This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](http://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made. © 2021 The Authors. *Applied Stochastic Models in Business and Industry* published by John Wiley & Sons Ltd.

230 ALLER AND CLEPT ALL

a multivariate normal distribution. However, in many cases, such assumption is not valid, especially for multivariate observations. Moreover, even when the process follows the multivariate normal distribution, the mean vector and the covariance matrix of the process are to be replaced with their estimates obtained from an in-control reference sample.

Non-parametric control charts require no distributional assumptions on the process data and generally enjoy more robustness (i.e., less sensitive to outlier) over parametric control schemes. However, the literature on non-parametric control charts is much more sparse than that on parametric charts.

The concept of data depth to produce control charts was originally used by Liu⁵ who introduced depth-based charts (*r*, *Q*, and *S*) for monitoring processes of multivariate quality measurements. The key idea is using ranks of multivariate measurements (obtained by means of a data depth function) rather than the multivariate measurements themselves. Such charts are really effective as they bring out the shift in location and the increase in scale through a two dimensional plot, which makes them easy to interpret. Furthermore, the normality assumption is not required.

The contribution of this work is to investigate how a non-parametric approach based on the notion of the *L^p* depth can be exploited for Statistical Process Control (SPC). Specifically, we focus on the *Q*-type control charts.

This article is organized as follows. Section 2 focuses on the non-parametric multivariate control charting. In Section 3, we give a brief introduction of the notion of data depth. Section 4 offers a basic theory of the depth-based methodology for the construction of control charts. In Section 5 are shown the performances of these charts along with their comparison with the charts proposed by $Liu⁵$ Finally, Section 6 offers some concluding remarks.

2 NON-PARAMETRIC MULTIVARIATE CONTROL CHARTS

The classical SPC approach aims to examine few variables to control the quality of an industry process. Nowadays, the industries collect a large amount of data on a set of several variables to be monitored simultaneously. For this reason, multivariate analysis is becoming increasingly important within the SPC,⁶ while traditional control charts aimed at monitoring a single variable are no longer useful to detect the overall quality of a process, where interaction of multiple correlated variables occurs.5,7 Multivariate control charts turn out to be really useful when dealing with correlated variables since they are able to take into account the association among the components of a multivariate process as all the observations are regarded as *d*-dimensional vectors and then used for detecting possible shifts in the *d*-dimensional distribution of the process. Therefore a multivariate approach turns out to be more effective than a joint monitoring scheme consisting of a set of univariate traditional control charts.⁸

The most popular multivariate control charts are the Hotelling's T^2 charts, which are a multivariate counterparts of the univariate Shewart's charts. Based on the invariance of the covariance matrix, the Hotelling's $T²$ charts are useful to detect (large) shifts in the mean vector of a multivariate process when the normality assumption holds. Indeed, in case of small shifts, adopting either the multivariate version of cumulative sum⁸ control chart or the multivariate extension of the exponentially weighted moving average⁹ control chart appear to be recommended. Even so, it is worth noting that in real applications the independence of measured data does not hold and the assumption that all individual variables (and all subsets of them) are normally distributed distribution is invalid.

As for the univariate counterparts, also the multivariate control charts can be distinguished into parametric and non-parametric. A recent survey on parametric multivariate charts can be found in Bersimis et al.,¹⁰ while the work of Chakraborti and Graham¹¹ reviews non-parametric multivariate control charts.

In this work, we focus on the non-parametric multivariate charts since in many applications there is no enough information to justify the assumption of a specific form for the underlying process distribution. In addition, the distribution of a parametric multivariate control chart is hard to estimate for processes with multiple quality characteristics and thus a non-parametric approach turns out to be a valid tool.

Non-parametric control charts have several advantages over their parametric counterparts. One of them is the flexibility, since the distribution-free multivariate methods do not require the assumptions of any parametric probability distribution for the underlying process. Such advantage guarantees a more robust SPC compared to their parametric version, in which the in-control run length changes as the deviations from distributional hypothesis change, increasing the number of false alarms as well as decreasing the validness of the monitoring process.

In the following, the attention is paid to rank-based non-parametric multivariate control charts. In particular, we refer to longitudinal ranking of the observed data, in which the order could be among observations at different time points. Three types of longitudinal ranking have been discussed in the literature: component-wise longitudinal ranking, spatial longitudinal ranking, and longitudinal ranking by data depth.¹² Unlike the first method, the latter two take into account all components jointly when computing the longitudinal rank.

In SPC, data depth are recognized as an efficient method to deal with multivariate robustness for Phase I analysis.11,13 Moreover, the methods based on data depth allow for a dimensional reduction of the variable space in a non-parametric way differing from methods based on the principal component analysis¹⁴ or partial least squares. The first non-parametric multivariate control chart based on data depth¹⁵ was proposed by Liu.⁵ Statistical data depth have become increasingly researched as a useful tool in non-parametric inference for multivariate data,¹⁶ but it was extended also to functional data¹⁷ and spherical data.¹⁸ The concept of data depth aims to order the observations from the most central point to the most outlying point with respect to the *d*-dimensional distribution of the observations. This goal is achieved by using a specific definition of data depth (e.g., Mahalanobis depth, the half-space depth, the simplicial depth, and so forth).

Three depth-based control charts, namely, *r*, *O*, and *S*, were proposed by Liu⁵ for monitoring processes of multivariate quality characteristics in a non-parametric way. The proposed charts are able to detect shift in location and increase in scale, simultaneously. The *r*, *Q*, and *S* control charts are depth-based multivariate generalizations of the univariate *X*, *X*, and CUSUM charts. Taking advantages of the simplicial depth for multivariate data, Liu et al.¹⁹ proposed a non-parametric moving average chart (DDMA-chart) improving the earlier proposed charts. For elliptical distributions,²⁰ Hamurkaroglu et al.21 developed the *r* and *Q* control charts proposed in the work of Liu et al. by using the Mahalanobis depth. However, the computation of the control limits of the Mahalanobis depth based-control charts has some weak points.7 First, the value of the generalized variance depends on the sample size of the data as well as on the number of quality characteristics. Then, it may happen that two different covariance matrices can give the same generalized variance.

In addition, as highlighted in Lange et al., 22 the Mahalanobis depth depends on the first two moments of a distribution and does not reflect any asymmetries of the data. Departing from these weakness, this article will be focused on *L^p* data depth to build a control chart. The *L^p* depth has additional advantages over other existing depth-based control charts already introduced in a multivariate SPC. First of all, the *L^p* depth may be more desirable because it does not require moment conditions. Then, they ensure an ease of computation even in high dimensions. The use of *L^p* depth function in a control charting has been neglected so far even though it is well-known and adopted in robust statistics and in many fields of application (see for instance the work of Pandolfo et al.²³).

3 DATA DEPTH CONCEPT

Statistical depth functions are largely used in non-parametric statistics for the analysis of multivariate data. A depth function aims at providing the degree of centrality of a point *x* with respect to a distribution *F* in \mathbb{R}^d , and it is denoted by $D(x, F)$. Higher depth values correspond to deeper (more central) points while smaller values indicate less central points (i.e., further away from the center of *F*). Hence, a center-outward ranking of the data is provided. More in detail, consider a random sample X_1, \ldots, X_n distributed as F, then the center-outward ranking X_i is $\{ \# X_j \in \{X_1, \ldots, X_n\} : D(X_j, F) \geq D(X_i, F) \}$, with F_n denoting the sample distribution function. If $X_{[j]}$ denotes the sample point associated with the *j*-th smallest depth value, then $X_{[1]}$, ..., $X_{[n]}$ are the ascending order statistic of X_i 's, with $X_{[n]}$ being the most central point so that larger ranks are assigned to more central points while the outlying points are associated with smaller ranks with respect to the underlying distribution *F*.

As stated by Zuo and Serfling,¹⁶ a statistical depth function should possess the four desirable properties listed in the following. Let *F* be a distribution on \mathbb{R}^d and *X* a random variable distributed as *F*.

- **P1.** Affine invariance: for any regular $d \times d$ matrix *A* and any *d*-dimensional vector *b* it holds $D(x, F) =$ $D(Ax + b, F_{AX+b});$
- **P2.** *Maximality at center:* if θ is the center of symmetry of *F*, then it holds $(\theta, F) = \sup(x, F)$; *x*∈R*^d*
- **P3.** *Monotonicity relative to the deepest point:* for any *F* with deepest point θ , then $D(x, F) \leq D((1 \lambda)\theta + \lambda x, F)$ for any $x \in \mathbb{R}^d$ and any $\lambda \in [0, 1]$;
- **P4.** *Vanishing at infinity:* for any *F*, then $D(x, F) \to 0$ as $||x|| \to \infty$.

The sample version of a depth function, denoted by $D(\cdot, F_n)$, is obtained by replacing *F* with its sample distribution *Fn*. Note that these properties are not satisfied by all data depths, but are rather desirable.

232 ALLER AND CLEPTER ALL

There are several different definitions of data depth function in the literature. The most popular are the halfspace, the simplicial, and the Mahalanobis depth function. The first two are based on geometrical structures while the latter on a measure of distance (i.e., the Mahalanobis distance). Their definition are given as follows.

Definition 1. The halfspace depth²⁴ of $x \in \mathbb{R}^d$ with respect to *F* is defined as

HD = $\inf_H \{ P_F(H) : H \text{ is a closed halfspace in } \mathbb{R}^d \text{ and } x \in H \}.$

The sample version of $HD(\cdot, \cdot)$ is obtained by evaluating $P_F(H)$ by means of the empirical distribution F_n . The halfspace of a point *x* in \mathbb{R}^d with respect to a dataset $X_1, \ldots, X_n \in \mathbb{R}^d$ is the smallest proportion of observations in any closed halfspace that contains *x*. It can be defined as

$$
HD(x, F_n) = \frac{1}{n} \min_{\|u\|=1} \# \{i : u^T X_i \ge u^T x\},
$$

where *u* ranges over all vectors in \mathbb{R}^d with $||u||=1$.

Definition 2. The simplicial depth¹⁵ of $x \in \mathbb{R}^d$ with respect to *F* is defined as

$$
SD(x, F) = P_F(x \in S[X_1, ..., X_{d+1}]),
$$

where $S[X_1, \ldots, X_{d+1}]$ denotes a closed *d*-dimensional simplex having $d+1$ random observations X_1, \ldots, X_{d+1} as vertices.

The sample version of the simplicial depth $SD(\cdot, \cdot)$ is defined as $SD(x, F_n) = \binom{n}{d}$ *d*+1 $\sum_{1 \le i_1 < i_1 \le i_{d+1} < n} I(x \in S[X_{i_1}, \ldots, X_{i_{d+1}}]),$ where X_1, \ldots, X_n is a random sample from F, F_n denotes its empirical distribution $, X_1, \ldots, X_{d+1}$ are all possible subsets from the sample and *I*(⋅) is the indicator function. For example, the bivariate $SD(x, F_m)$ relative to X_1, \ldots, X_n is equal to the proportion of closed triangles with vertices X_i , X_j , X_k that contain x , $1 \le i < j < k \le n$.

Definition 3. The Mahalanobis depth²⁵ of $x \in \mathbb{R}^d$ with respect to *F* is defined as

$$
MhD(x, F) = \frac{1}{1 + (x - \mu)^T \Sigma^{-1} (x - \mu)},
$$

where μ and Σ are the mean vector and the covariance matrix of the distribution *F*, respectively.

The sample version of *MD*(⋅ , ⋅) is obtained by using the sample mean *x* and the estimated covariance matrix Σ*̂* , and can be defined as

$$
MhD(x, F_n) = \frac{1}{1 + (x - \bar{x})^T \hat{\Sigma}^{-1} (x - \bar{x})}.
$$

Definition 4. Taking the L^p ($p > 0$) norm as measure of distance of a point *x* from a random sample $X = \{X_1, \ldots, X_n\}$ the corresponding L^p depth is defined as

$$
L^{p}D(x, F) = \frac{1}{1 + E(|x - X||_{p})},
$$

where $X \sim F$, $\|\cdot\|$ denotes the L^p -norm (when $p=2$ the Euclidean norm is derived) and $E(\cdot)$ is its expected value.

The sample version L^p is obtained by replacing *F* with its empirical counterpart F_n while evaluating $E\left(\|\mathbf{x} - X\|_p\right)$. The first two notions are based on geometrical structures while the latter two on metric distances. Several other depths are available in the literature and each notion is able to capture different features of the underlying distribution and presents specific advantages and drawbacks. Hence, there is no optimal choice to be made in advance, and this holds true also when using depths for statistical process control.

However, it is worth stressing that when dealing with multivariate data, the computational feasibility turns out to be a really desirable characteristic, and in this regard the *L^p* depth is particularly appealing because of its computational ease even in high dimensional spaces. This stands as the key reason why we propose to use this notion of depth that has been neglected in this framework so far. For a detailed discussion of the properties of the above listed depth and many others we refer the reader to the paper of Zuo and Serfling.¹⁶

4 DEPTH-BASED CONTROL CHARTS

Depth-based methodology to construct control charts can be interpreted as a multivariate generalization of standard univariate rank methods because they are both based on the idea of ranks. However, an important difference between the two methods exists in the way of ranking the data: in the univariate case, it is a linear ranking from the smallest to the largest, while the multivariate one (provided by data depth functions) is a center-outward ranking. Hence, once each observation in \mathbb{R}^d (for $d>1$) can be represented by its corresponding depth rank. Then, control charts based on these ranks can be built by following the criteria for the univariate control charts. This approach is fully non-parametric, meaning that the obtained charts are valid without parametric assumptions on the process distribution. In addition, such charts allow for the simultaneous detection of both location and scale changes in a process.

4.1 Statistics based on data depths to construct control charts

Within the literature, data depth function has been mainly used to construct multivariate control charts. Let *F* denote a *d*-dimensional distribution (if the measurements follow the distribution *F*, the process is considered to be in control). The distribution *F* can be either known or it can be estimated, and F_n denotes its empirical version. Let $X = \{x_1, x_2, ..., x_n\}$ denote a random sample taken from a population having distribution of F while $Y=\{y_1,y_2\,\ldots\, ,y_q\}$ is a new sample of observations taken from a population having distribution *G* (that is, new measurements taken from the process). Based on observations y_i , it will be determined whether the product quality is deteriorating or whether the process is out of control. Hence, *F* with *G* are to be compared. Thus, if the y_i 's do not approach the distribution *F*, it means that the quality of product has deteriorated. Then, the control charts based on depth functions as proposed by Liu⁵ are rooted in the multivariate range induced by a depth function *D*(⋅ , ⋅) which is defined as

$$
r(y, F) = P_F\{x : D(x, F) \le D(y, F)\}.
$$

From this, since multivariate data have been mapped in the univariate space, the control charts listed below can be derived.

• *Control charts for r instead of X-Charts*

$$
r(y_j, F_n) = #\{X_i : D(X_i, F_n) \le D(y_j, F_n)\}/n \quad \text{for} \quad i = 1, 2, ..., n \quad \text{and} \quad j = 1, 2, ..., q
$$
 (1)

• *Control charts for Q, average range (instead of X*-*Charts)*

$$
Q(F, G) = P\{D(X, F) \le D(y, F)\} \quad (= E_G[r(y, F)]). \tag{2}
$$

• *Control charts for S, accumulated range (instead of CUSUM charts)*

$$
S(Y, F_n) = \sum_{i=1}^{q} \left(r(y_i, F_n) - \frac{1}{2} \right)
$$

The following section provides more details about each of the above-mentioned types of control charts based on depths.

4.2 *r***-type data depth control chart**

An *r*-type control chart is the multivariate extension of the univariate Shewart *X*-chart for individual measurements. This latter is very simple and efficient, however it cannot be easily generalized to a multivariate process.

From (1) it follows that $r(\cdot, \cdot)$ denotes the fraction of observations within the reference sample $X = \{X_1, \ldots, X_n\}$ which have a depth value that does not exceed $D(Y_i, F_n)$, and thus $r(\cdot, \cdot) \in [0, 1]$. Small values of $r(\cdot, \cdot)$ indicates that a small proportion of *Y_j*'s is more outlying than X_i . Thus X_i is at the "outskirt" of $\{Y_1, \ldots, Y_q\}$ and is not conforming to most of the central part of the good dataset. Here we recall that an *r*-type control chart with $LCL = \alpha$ theoretically corresponds to the iterative application of an α -level test of the following hypothesis (so that α is the false alarm rate):

 H_0 : $F = G$ vs. *H*₁: There is a location shift and/or a scale increase from *F* to *G*.

When the value of the statistic $r(\cdot, \cdot)$ is smaller than α the above mentioned test reject the null hypothesis H_0 and the process is declared to be out of control. The center line (*CL*) is a reference line that enables the observation of a possible trend or non-random pattern. This is because *r*(⋅ , ⋅)∼ *U*[0, 1], and thus its expected value is 0.5. Furthermore, values which are larger than 0.5 could indicate a decrease in scale or a neglectable shift in location, hence the process is not considered out of control. Therefore, in a *r*-type control chart we have only a lower control limit (*LCL*). The center line (*CL*) and the lower control limit (*LCL*) of the chart are defined as

$$
CL = 0.5
$$
 and $LCL = \alpha$,

where the crossing of the line at *LCL* is considered the statistical signal of a loss in the quality (i.e., "out of control") of the process.

4.3 *Q***-type data depth control chart**

The *Q*-type control chart is the multivariate analogue of the average univariate chart \overline{X} . This type of chart is based on the charting statistic defined in (2). The idea is to plot the averages of subgroups (each one of size *q*) of $r(Y_i, F_j)$'s (or $r(Y_i, F_n)$) that are given as follows:

$$
Q(F, G_q) = \frac{1}{q} \sum_{j=1}^q r(Y_j, F)
$$
 and $Q(F_n, G_q) = \frac{1}{q} \sum_{j=1}^q r(Y_j, F_n)$.

The values of the central line (CL) and the lower control limit (*LCL*) of a *Q*-type control chart depend on the the size *q* of the subgroups. When *q* is moderately large, *CL* and *LCL* are evaluated as:

$$
CL = 0.5
$$
 and $LCL = 0.5 - z_{\alpha} \sqrt{\frac{1}{12} \left[\frac{1}{n} + \frac{1}{q} \right]}$,

with α be the required false alarm probability and z_α be the upper α -th quantile of the standard normal. Instead, for subgroups having small size ($q \leq 5$) the *LCL* is calculated as:

$$
LCL = 0.5 - z_{\alpha} (12q)^{-\frac{1}{2}}.
$$

However, when *q* is even small (*q* = 3 or 4) and $\alpha \leq 1/q!$, then LCL is defined as $\frac{(q! \alpha)^{\frac{1}{3}}}{q}$ $\frac{a}{q}$. Note that the data must be centered if the *Q*-type chart is used to detect changes in scale only.

4.4 *S***-type data depth control chart**

The *S*-type control chart takes its motivation from the univariate CUSUM charts, which are the plot of $\sum_{i=1}^{n}(X_i - \mu)$; this indicates the total deviation from the mean in the univariate case. The CUSUM chart is effective when the task is to detect

TABLE 1 Distributions used in the simulation study

small changes in a process. Moving to the multivariate case, the charting statistic defined by Liu²⁵ are:

$$
S_n(F) = \sum_{j=1}^m \left(r(Y_j, F) - \frac{1}{2} \right) \text{ and } S_n(F_n) = \sum_{j=1}^q \left(r(Y_j, F_n) - \frac{1}{2} \right).
$$

A standardized version of the *S*-type chart, denoted $S_n^*(F)$ and $S_n^*(F_n)$, respectively, can be obtained as follows:

$$
S_n^*(F) = \frac{S_n(F)}{\sqrt{n/12}}
$$
 and $S_n^*(F_n) = \frac{S_n(F_n)}{\sqrt{q^2(1/q + 1/n)/12}}$.

For this type of chart $CL = 0$ and $LCL = -z_a$. Thus, the lower control limit does not depend neither on *n* nor on *q*.

5 SIMULATION STUDY

This section is aimed at showing how the control charts based on the *L^p* depth perform in comparison to those based on the Mahalanobis depth (which is often adopted in SPC) when analyzing multivariate process data. We do not consider the simplicial and halfspace depth functions because of their infeasibility when the dimension and the sample size increase.

The simulation study was conceived as an analysis to screen the chart performances under multiple settings, defined with regard to the number of variables to be monitored (i.e., the dimension *d*), the size (*n*) of the reference sample, the size (*q*) of the sub-group, and by considering different distributional settings.

In accordance with Bae et al.,26 we do not consider the *S*-type control chart in our evaluation study as the cumulated deviations are tiny in the simulation setting and thus only the *Q*-type control charts were evaluated.

5.1 Study design

The performances of the *Q*-type charts based on the *Lp* depth function were evaluated through a Monte Carlo simulation study and compared to those based of the Mahalanobis depth-based *Q*-type charts in terms of average run length (ARL), which is one of the performance measures used for comparing the control charts. ARL is defined as the expected number of samples required to get a first out-of-control signal, and it can be obtained by taking reciprocal of false alarm probability. Both the in-control (IC) and out-of-control (OOC) cases were considered. Distributions used in the simulation study are listed in Table 1 below.

The parameter μ_0 denotes the location, Σ_0 denotes the scatter matrix and ξ denotes the shape parameter (or skewness parameter) of the skew Normal distribution.

Without loss of generality, we set the location parameter $\mu_0 = 0$, the covariance matrix $\Sigma_0 = I_d$ (i.e., uncorrelated unit-variance components) and the skewness parameter of the skew normal distribution $\xi = 0$. For each parameters' combination, 10,000 replications were carried out, with an expected rate of the type I error, or false alarm, $\alpha = 0.0027$ (hence the in-control ARL₀ = 1/0.0027 \approx 370), reference samples of size *n* ∈ {100,250, 500} and sub-groups of size *q* ∈ {5, 10, 15}. We set $p = 2$ for the computation of the L^p depth. The average run length and its standard deviation were computed.

Three out-of-control scenarios including shift in the mean vector, change in the variance and a combination of both variance change and shift in the mean were considered. Specifically:

Setup 1: Shift in location $\delta \in \{0.15, 0.30\}$ applied to the location parameter of out-of-control process distributions,

PANDOLFO ET AL. **235 235**

- *Setup 2*: Change in variance of magnitude λ by multiplying a scalar λ to only one randomly chosen diagonal element of the scatter matrix Σ_0 , with $\lambda \in \{0.15, 0.30\}$,
- *Setup 3*: Shift in location ($\delta \in \{0.15, 0.30\}$) and change in variance ($\lambda \in \{0.15, 0.30\}$) as done in setup 1 and setup 2, respectively.

Data were generated in dimensions *d*∈{5, 10, 25}. We did not consider higher dimensions because when the reference sample size *n* is small, as the dimension increases, the Mahalanobis depth may not work well. Indeed, when using the Mahalanobis distance, the contamination bias in estimating the covariance matrix grows rapidly with *d*. This drawback has been revealed by some recent study on high dimensional tests.^{27,28}

5.2 In-control case performances

Results obtained from the in-control setting for each size of the considered reference sample are reported in Tables 2 to 4. As one can see, ARL values indicate that the *Q*-type charts based on *L*²*D* and *MhD* perform quite similar regardless of the process distribution or the dimensionality of both the reference and the sub-group samples. Indeed, it can be observed that the ARL values approach the nominal value $ARL₀ = 370$, and the gap between actual and nominal rate seems not to be affected by the sample size or the distribution. Even no particular difference can noticed in terms of standard deviation, and thus no recommendation about which depth function should be preferred can be done for the considered in-control setting.

5.3 Out-control case performances

The out-of-control average run lengths (*ARL*1) for the for considered setups are reported in Tables 5 to 13.

First, we describe the performances of the control charts under evaluation when only a shift in location is considered (Tables 5 to 7). Charts appear to be sensitive to small shift in process location vector and thus effective to detect the changes in all the considered dimensions since the ARL decrease both for $\delta = 0.15$ and $\delta = 0.30$ regardless of the size of the sub-groups *q*. Both *L*²*D* and *MhD*-based charts behave about the same for all three distributions. However, it is worth highlighting that the *MhD*-based charts present some ARL₁ values that exceed the in-control ARL while those based on L^2D show constantly ARL₁s smaller than 370. Moreover, when data are normally distributed in dimension $d=25$ for $n = 100$ and $q = 5$, the out-of-control ARL of the *MhD*-based chart is equal to 140.6, which is a quite low value indicating that too many false out-of-control signals were detected. It is presumably the effect of the reference sample size *n*, since the performance of the *MhD*-based charts improve when *n* increase (i.e., 250 and 500). Hence, we suggest the use of the *LpD* for better performance for this scenario.

Now consider the case of change in variance, that is, Setup 2, whose results are reported in Tables 8 to 10. Here again, it can be seen that ARLs of the L^2D -based charts are stably below 370 for the two shift magnitudes in Covariance matrix $(\lambda = 0.15$ and 0.30), and there seems to be no effect of dimensionality, reference sample size *n*, sub-group size *q*, and distribution on the charts' performances. In general, also in this case it is observed that the *Q*-type charts based on *L*²*D* perform generally better than those based on *MhD*, which even shows ARLs greater than the in-control ARL in a few cases.

Tables 11 to 13 give ARL values when there is a shift in location and scale both. In such scenario, the out-of-control data were generated by applying a shift $\delta \in \{0.15, 0.30\}$ to the location parameter of the in-control process and a change ∈ {0*.*15*,* 0*.*30} to the covariance matrix of the in-control process, simultaneously. The charts under evaluation appear to be sensitive to the considered shifts even though, here again, some more attention must be paid to the $ARL₁$ values of the *MhD*-based charts. More in detail, when data follow a multivariate Normal and a Skew-Normal distribution in *d*=25 dimensions for *n*=100 and *q*=5, the *Q*-type control charts based on the Mahalanobis depth show ARLs which appear to be to much small, especially in relation to the small magnitude of the considered shifts, as that means that the *MhD*-based *Q*-type charts produced too many false out-of-control signals in such scenario. It is no longer occurring when *n*=250 and 500. *MhD*-based charts show also some ARLs exceeding ARL₀ also in this Setup, where, once again *Q*-type charts based on *L*²*D* seem to perform better.

(ARL) and the corresponding standard deviation between brackets for $n = 100$

RIGHTSLINK[®]

PANDOLFO ET AL. **237 237**

TABLE 3 In-control average run length (ARL) and the corresponding standard deviation between brackets for *n*=250

RIGHTS LINK()

TABLE 4 In-control average run length (ARL) and the corresponding standard deviation between brackets for $n = 500$

TABLE 5 Setup 1: out-control average run length (ARL) and the corresponding standard deviation between brackets for

TA B L E 11 Setup 3: out-control average run length (ARL) and the corresponding standard deviation between brackets for *n*=100

TA B L E 13 Setup 3: out-control average run length (ARL) and the corresponding standard deviation between brackets for *n*=500

PANDOLFO ET AL. **249** DOLFO ET AL.

6 CONCLUDING REMARKS

The purpose of this article is proposing the adoption of the *L^p* depth function for the construction of non-parametric multivariate process control charts. The performances of the proposed method, which do not require any distributional assumption, are investigated through simulations by considering different distributional settings, that is, the multivariate Normal, Cauchy and Skew-Normal distributions. The *L^p* depth-based control charts are compared to those based on the Mahalanobis depth.

Results suggest that the *Q*-type charts based on *L*²*D* perform better than those based on the Mahalanobis depth regardless of the process distribution, the dimensionality and the size of both the reference and the sub-group samples. Moreover, it can be highlighted that with regard to the process distribution of the reference sample, the leptoKurtosys (introduced via the Cauchy sampling distribution) is not more challenging than asymmetry for both the depth notions even if the *L*²*D*-based charts show more stable performances.

Hence, the use of the *LP* depth in SPC deserves attention and it can be considered a valid alternative to the well-known Mahalanobis depth as well as other geometrically defined data depths which are unfeasible in high multidimensional space.

ACKNOWLEDGEMENT

The authors would like to thank the Editor and the anonymous reviewer for their thoughtful comments.

ORCID

Carmela Iorio <https://orcid.org/0000-0001-5622-4124>

REFERENCES

- 1. Montgomery DC. *Introduction to Statistical Quality Control*. Hoboken, NJ: John Wiley & Sons; 2007.
- 2. Barale MS, Shirke DT. Nonparametric control charts based on data depth for location parameter. *J Stat Theory Pract*. 2019;13(3):41.
- 3. Hotteling H. Multivariate quality control, illustrated by the air testing of sample bombsights. In: Eisenhart C, Hastay MW Wallis WA, Eds. *Techniques of Statistical Analysis*. New York: McGraw Hill; 1947;111-184.
- 4. Woodall WH. Controversies and contradictions in statistical process control. *J Qual Technol*. 2000;32(4):341-350.
- 5. Liu RY. Control charts for multivariate processes. *J Amer Stat Assoc*. 1995;90(432):1380-1387.
- 6. Woodall WH, Montgomery DC. Research issues and ideas in statistical process control. *J Qual Technol*. 1999;31(4):376-386.
- 7. Idris S, Wachidah L, Sofiyayanti T, Harahap E. The control chart of data depth based on influence function of variance vector. *J Phys Conf Ser*. 2019;1366:012125.
- 8. Crosier RB. Multivariate generalizations of cumulative sum quality-control schemes. *Technometrics*. 1988;30(3):291-303.
- 9. Lowry CA, Woodall WH, Champ CW, Rigdon SE. A multivariate exponentially weighted moving average control chart. *Technometrics*. 1992;34(1):46-53.
- 10. Bersimis S, Psarakis S, Panaretos J. Multivariate statistical process control charts: an overview. *Qual Reliab Eng Int*. 2007;23(5):517-543.
- 11. Chakraborti S, Graham MA. Nonparametric (distribution-free) control charts: an updated overview and some results. *Qual Eng*. 2019;31(4):523-544.
- 12. Qiu P. Some perspectives on nonparametric statistical process control. *J Quality Technology.* 2018;50(1):49-65.
- 13. Li Z, Dai Y, Wang Z. Multivariate change point control chart based on data depth for phase I analysis. *Commun Stat Simul Comput*. 2014;43(6):1490-1507.
- 14. Scepi G. *Parametric and Non Parametric Multivariate Quality Control Charts*. New York, NY: Springer; 2002:163-189.
- 15. Liu RY. On a notion of data depth based on random simplices. *Ann Stat*. 1990;18(1):405-414.
- 16. Zuo Y, Serfling R. General notions of statistical depth function. *Ann Stat*. 2000;28(2):461-482.
- 17. López-Pintado S, Romo J. On the concept of depth for functional data. *J Am Stat Assoc*. 2009;104(486):718-734.
- 18. Pandolfo G, Paindaveine D, Porzio GC. Distance-based depths for directional data. *Canad J Stat*. 2018;46(4):593-609.
- 19. Liu RY, Singh K, Teng JH. DDMA-charts: nonparametric multivariate moving average control charts based on data depth. *Allgemeines Statistisches Archiv*. 2004;88(2):235-258.
- 20. Alt FB, Smith ND. Multivariate process control. *Handbook of Statistics*. Vol 7. North-Holland, Amsterdam: Elsevier; 1988:333-351.
- 21. Hamurkaroglu C, Mert M, Saykan Y. Nonparametric control charts based on Mahalanobis depth. *Qual Control Appl Stat*. 2006;51(1):21-26.
- 22. Lange T, Mosler K, Mozharovskyi P. Fast nonparametric classification based on data depth. *Stat Pap*. 2014;55(1):49-69.
- 23. Pandolfo G, Iorio C, Siciliano R, D'Ambrosio A. Robust mean-variance portfolio through the weighted *Lp* depth function. *Ann Operat Res*. 2020;292(1):519-531.
- 24. Tukey JW. Mathematics and the picturing of data. In: James RD (Ed.), Proceedings of the International Congress of Mathematicians, Vancouver; 1975;2:523-531.
- 25. Liu RY, Singh K. A quality index based on data depth and multivariate rank tests. *J Am Stat Assoc*. 1993;88(421):252-260.
- 26. Bae SJ, Do G, Kvam P. On data depth and the application of nonparametric multivariate statistical process control charts. *Appl Stoch Models Bus Ind*. 2016;32(5):660-676.
- 27. Bai Z, Saranadasa H. Effect of high dimension: by an example of a two sample problem. *Stat Sin*. 1996;6(2):311-329.
- 28. Feng L, Zou C, Wang Z, Chen B. Rank-based score tests for high-dimensional regression coefficients. *Electron J Stat*. 2013;7:2131-2149.

How to cite this article: Pandolfo G, Iorio C, Staiano M, Aria M, Siciliano R. Multivariate process control charts based on the *Lp* depth. *Appl Stochastic Models Bus Ind*. 2021;37:229–250. https://doi.org/10.1002/asmb.2616