Revista Colombiana de Estadística
ISSN: 0120-1751
revcoles_fcbog@unal.edu.co
Universidad Nacional de Colombia
Colombia

Guevara-González, Rubén Darío; Vargas-Navas, José Alberto; Linero-Segrera, Dorian Luis
Profile Monitoring for Compositional Data
Revista Colombiana de Estadística, vol. 37, núm. 1, junio, 2014, pp. 159-181
Universidad Nacional de Colombia
Bogotá, Colombia

Revista Colombiana de Estadística
Junio 2014, volumen 37, no. 1, pp. 159 a 181

Profile Monitoring for Compositional Data

Monitoreo de perfiles para datos composicionales

Rubén Darío Guevara-González[1,a], José Alberto Vargas-Navas[1,b], Dorian Luis Linero-Segrera[2,c]

[1]Department of Statistics, National University of Colombia, Bogotá, Colombia
[2]Engineering School, National University of Colombia, Bogotá, Colombia

Abstract

In a growing number of quality control applications, the quality of a product or process is best characterized and summarized by a functional relationship between a response variable and one or more explanatory variables. Profile monitoring is used to understand and to check the stability of this relationship over time. In some applications with compositional data, the relationship can be characterized by a Dirichlet regression model. We evaluate five T^2 control charts for monitoring these profiles in Phase I. A real example from production of concrete is given.

***Key words*:** Control chart, Dirichlet distribution, Statistical process control.

Resumen

En un gran número de aplicaciones la calidad de un producto o proceso está mejor representada por una relación funcional entre una variable de respuesta y una o más variables explicatorias. El monitoreo de perfiles permite entender y chequear la estabilidad de esta relación funcional a través del tiempo. En algunas aplicaciones con datos composicionales, la relación puede ser representada por un modelo de regresión Dirichlet. En este artículo nosotros evaluamos cinco cartas de control T^2 para monitorear estos perfiles en Fase I. Un ejemplo real asociado a la producción de concreto es presentado.

***Palabras clave*:** carta de control, control estadístico de procesos, distribución Dirichlet.

[a]Assistant professor. E-mail: rdguevarag@unal.edu.co
[b]Professor. E-mail: javargasn@unal.edu.co
[c]Assistant professor. E-mail: dllineros@unal.edu.co

1. Introduction

In most of the statistical process control (SPC) applications, the quality of a process or product is represented by the distribution of a univariate or multivariate quality characteristic. However, in other applications, process quality is better characterized by a relationship between a response variable and one or more explanatory variables. This relationship is usually known as a profile. In these situations, the focus of the SPC lies on the parameters of the profile monitoring rather than on the monitoring of the univariate or multivariate characteristics. Such profiles can be represented using linear or nonlinear models. Some discussion of the general issues involving profile monitoring can be found in Woodall, Spitzner, Montgomery & Gupta (2004), Woodall (2007), Noorossana, Saghaei & Amiri (2012) and Qiu (2013). Profile practical applications have been reported by many researchers, including Stover & Brill (1998), Kang & Albin (2000), Mahmoud & Woodall (2004), Wang & Tsung (2005) and Kusiak, Zheng & Song (2009). Several control chart approaches for monitoring simple linear profiles have been developed by Kang & Albin (2000), Kim, Mahmoud & Woodall (2003), Zou, Zhang & Wang (2006), Zou, Zhou, Wang & Tsung (2007), Mahmoud, Parker, Woodall & Hawkins (2007), Soleimani, Narvand & Raissi (2013), Zhang, He, Zhang & Woodall (2013), Yeh & Zerehsaz (2013) and Amiri, Zou & Doroudyan (2014). Proposals for monitoring multivariate linear profiles (simple and/or multiple) have been developed by Mahmoud (2008), Noorossana, Eyvazian & Vaghefi (2010), Noorossana, Eyvazian, Amiri & Mahmoud (2010), Eyvazian, Noorossana, Saghaei & Amiri (2011) and Zou, Ning & Tsung (2012).

The linear regression model is commonly used for monitoring profiles. However, it is not appropriate for situations where the response is restricted to the interval $(0, 1)$ since it may yield fitted values in the variable of interest that exceed its lower and upper bounds. Ferrari & Cribari-Neto (2004) proposed a regression model that is tailored for situations where the dependent variable Y is measured continuously on the standard unit interval, i.e. $0 < Y < 1$. The proposed model is based on the assumption that the response is Beta distributed. The Beta distribution is very flexible for modeling proportions since its density can have quite different shapes depending on the values of the two parameters that index the distribution. Vasconcellos & Cribari-Neto (2005) proposed a class of regression models where the response is Beta distributed and the two parameters that index this distribution are related to covariates and regression parameters. However, the proposed regression models are restricted to the univariate case and cannot be applied in many practical situations where data consist of multivariate positive observations summing to one, that is, the study of compositional data, see Aitchison (1986) and Aitchison (2003). Melo, Vasconcellos & Lemonte (2009) proposed a particular structure for compositional data regression, based on the Dirichlet distribution, which is a generalization of the Beta distribution for the simplex sample space. A profile application in a concrete manufacturing plant, which after a preliminary study was found to fit appropriately this structure motivated this paper.

Compositional data are frequently encountered in industries such as the chemical, pharmaceutical, textil, plastic, concrete, steel, asphalt, among other. Sev-

eral statistical methods for monitoring processes characterized by compositional data have been studied. See for example, Sullivan & Woodall (1996), Boyles (1997), Yang, Cline, Lytton & Little (2004) and Vives-Mestres, Daunis-i Estadella & Martín-Fernández (2013). However, there are not methods for monitoring these processes when the random vectors associated to the compositional data present a functional relationship with a set of explanatory variables.

In this paper, the control charting mechanisms discussed by Williams, Woodall & Birch (2007) and Yeh, Huwang & Li (2009) are extended for monitoring functional relationships in Phase I characterized by a Dirichlet regression model using a regression structure that allows the modeling of relationships between random vectors with Dirichlet distribution and a set of explanatory variables.

The structure of this paper is outlined as follows: In Section 2, we show the Dirichlet regression model for compositional data and the estimation of the model parameters. Five T^2 control charts approaches used for monitoring linear profiles in Phase I with compositional data are presented in Section 3. In Section 4, the performance of the proposed approaches is evaluated through simulation studies. A real example is given in Section 5. In the last section we conclude the paper.

2. Dirichlet Regression

Compositional data are used to indicate how parts contribute to the whole. In most cases they are recorded as closed data, i.e. data summing to a constant, such as 100%. Compositional data occupy a restricted space where variables can vary only from 0 to 100, or any other given constant. Such a restricted space is known formally as a simplex, see Pawlowsky-Glahn & Egozcue (2006).

Let c be a positive number. The p-dimensional closed simplex in $\mathbb{R}^n$ and $(p-1)$-dimensional open simplex in $\mathbb{R}^{p-1}$ are defined by

$$\mathbb{T}_p(c) = \left\{ (y_1, \ldots, y_p)^t : y_j > 0, 1 \leq j \leq p, \sum_{j=1}^{p} y_j = c \right\}$$

and

$$\mathbb{V}_{p-1}(c) = \left\{ (y_1, \ldots, y_{p-1})^t : y_j > 0, 1 \leq j \leq p-1, \sum_{j=1}^{p-1} y_j < c \right\}$$

respectively, where the superscript t means the function transpose. Furthermore, let $\mathbb{T}_p = \mathbb{T}_p(1)$ and $\mathbb{V}_{p-1} = \mathbb{V}_{p-1}(1)$.

A random vector $\mathbf{Y} = (Y_1, \ldots, Y_p)^t \in \mathbb{T}_p$ is said to have a Dirichlet distribution if the density function of $\mathbf{Y_{-p}} = (Y_1, \ldots, Y_{p-1})^t$ is

$$f\left(\mathbf{Y_{-p}}|\mathbf{a}\right) = \frac{\Gamma\left(\sum_{j=1}^{p} a_j\right)}{\prod_{j=1}^{p} \Gamma\left(a_j\right)} \prod_{j=1}^{p} y_j^{a_j - 1}, \qquad (y_1, \ldots, y_{p-1}) \in \mathbb{V}_{p-1}, \tag{1}$$

where $\mathbf{a} = (a_1, \ldots, a_p)^t$ and $a_j > 0$, $j = 1 \ldots p$. We will write $\mathbf{Y} \sim Dirichlet_p(a_1, \ldots, a_p)$, see Ng, Tian & Tang (2011).

When all $a_j \to 0$, the distribution becomes noninformative. When $p = 2$, the Dirichlet distribution $Dirichlet_2(a_1, a_2)$ reduces to the $Beta(a_1, a_2)$ distribution. The marginal distributions of the components of $\mathbf{Y}$, $Y_j, j = 1, 2, \ldots, p$, are distributed as $Beta(a_j, \phi - a_j)$, where $\phi = \sum_{j=1}^{p} a_j$. In this sense, the Dirichlet distribution can be seen as a multivariate extension of the Beta distribution. Therefore, we have

$$E(Y_j) = \frac{a_j}{\phi}, \qquad j = 1, \ldots, p \tag{2}$$

$$Var(Y_j) = \frac{a_j(\phi - a_j)}{\phi^2(\phi + 1)}, \qquad j = 1, \ldots, p \tag{3}$$

$$Cov(Y_j, Y_l) = -\frac{a_j a_l}{\phi^2(\phi + 1)} < 0, \qquad j \neq l; j, l = 1, \ldots, p \tag{4}$$

The Dirichlet distribution is widely used to model data in the form of proportions, where each observation is a vector of positive numbers summing to one. It allows great flexibility of modeling, provided by the appropriate choice of its parameters. See Ng et al. (2011) and Melo et al. (2009).

Gueorguieva, Rosenheck & Zelterman (2008) described a Dirichlet multivariate regression method which is useful for modeling data representing components as a percentage of a total. They described each $\log(a_j)$ as a separate linear function of covariates and regression coefficients. That is, for each component $j = 1, \ldots, p$ they used a log-link with

$$\log a_{ij} = \boldsymbol{\beta}_j^t \mathbf{X}_i \tag{5}$$

for covariates X_i recorded on the ith individual ($i = 1, \ldots, n$) and regression coefficients $\boldsymbol{\beta}_j$ to be estimated using maximum likelihood. These estimates are denoted $\widehat{\boldsymbol{\beta}}_j$. The estimates $\widehat{\boldsymbol{a}}_j = \{\widehat{a}_{ij}\}$ of $\boldsymbol{a}_j = \{a_{ij}\}$ are defined by

$$\widehat{a}_{ij} = \exp(\widehat{\boldsymbol{\beta}}_j^t \mathbf{X}_i)$$

Gueorguieva et al. (2008) refer to the $\{\boldsymbol{a}_j\}$ as $meta - parameters$ because they combine the effects of the covariates $\mathbf{X}_i$ using regression parameters $\{\boldsymbol{\beta}_j\}$.

Melo et al. (2009) proposed a generalization of this model. The proposed model is defined by establishing relationships between the parameters that index the Dirichlet distribution and linear predictors on the explanatory variables. They assume a set of independent vector observations $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$, where $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{ip})$ with $Y_{i1} + \cdots + Y_{ip} = 1$, for each i. They suppose that $\mathbf{Y}_i \sim Dirichlet_p(a_{i1}, \ldots, a_{ip})$ with

$$a_{ij} = g_j(\beta_{1j} x_{i1} + \cdots + \beta_{kj} x_{ik}) \tag{6}$$

where each function $g_j : \mathbb{R} \to (0, \infty)$ is three times differentiable, injective and known, $x_{i1}, \ldots, x_{ik}$ are the values corresponding to the ith observation for k explanatory variables and $\beta_{1j}, \ldots, \beta_{kj}$ are k unknown parameters corresponding to

the jth component. The model, therefore, has kp unknown parameters, which can be estimated through maximum likelihood (See Melo et al. 2009).

The covariates of this regression model affect the vector mean, the variance covariance structure of the distribution of the observations and the higher-order moments. The functions g_j play a similar role to the link functions of generalized linear models, in the sense that they specifically define how the parameters of the distribution of interest are linked to linear combinations of the covariates. The coefficients of this linear combination are unknown. The regression parameters are identifiable if the link functions are injective and the covariates are linearly independent (See Melo et al. 2009).

In the Dirichlet regression model, if $p = 2$ we have the Beta regression model described in Vasconcellos & Cribari-Neto (2005), Gueorguieva et al. (2008) and Melo et al. (2009).

The regression coefficients can be estimated using maximum likelihood. Let $\mathbf{B}$ the $k \times p$ matrix with the β_{hj}'s, $h = 1, 2 \ldots, k$ and $j = 1, 2, \ldots, p$. The log-likelihood function is given by

$$l(\mathbf{B}) = \sum_{i=1}^{n} \left\{ \log[\Gamma(\phi_i)] - \sum_{j=1}^{p} \log[\Gamma(a_{ij})] + \sum_{j=1}^{p} a_{ij} \log(Y_{ij}) \right\} \tag{7}$$

where $\phi_i = a_{i1} + \cdots + a_{ip}$ for each $i = 1, 2, \ldots, n$.

If $\widehat{\mathbf{B}}$ is the maximum likelihood estimator for $\mathbf{B}$, under some regularity conditions, $\sqrt{n}vec(\widehat{\mathbf{B}} - \mathbf{B}) \overset{a}{\sim} N_{kp}(\mathbf{0}, \mathbf{K}(\mathbf{B})^{-1})$, when n is large, with $\overset{a}{\sim}$ denoting asymptotically distributed, N_{kp} representing a kp-variate normal distribution and $\mathbf{K}(\mathbf{B})$ representing the $kp \times kp$ information matrix for the vector version of $\mathbf{B}$ (See Melo et al. 2009). The matrix $\mathbf{K}(\mathbf{B})$ can be obtained as

$$\mathbf{K}(\mathbf{B}) = \left(\mathbf{I}_p \bigotimes \mathbf{X} \right)^t \mathbf{L} \left(\mathbf{I}_p \bigotimes \mathbf{X} \right) \tag{8}$$

where $\bigotimes$ represents the Kronecker product, $\mathbf{L}$ is an $np \times np$ matrix defined in partitioned form as

$$\mathbf{L} = \begin{pmatrix} \mathbf{L}_{11} & \cdots & \mathbf{L}_{1p} \\ \vdots & \ddots & \vdots \\ \mathbf{L}_{p1} & \cdots & \mathbf{L}_{pp} \end{pmatrix} \tag{9}$$

where each $\mathbf{L}_{cd}$, with $c = 1, 2, \ldots, p$ and $d = 1, 2, \ldots, p$, is a diagonal matrix having ith element in the diagonal given by

$$l_i^{(cd)} = \begin{cases} -g_c'(\eta_{ic})^2[\psi'(\phi_i) - \psi'(a_{ic})], & c = d \\ -g_c'(\eta_{ic}) g_d'(\eta_{id}) \psi'(\phi_i), & c \neq d \end{cases} \tag{10}$$

where $\phi_i = a_{i1} + \cdots + a_{ip}$, for each $i = 1, 2, \ldots, n$, $\eta_{ij} = \beta_{1j} X_{i1} + \cdots + \beta_{kj} X_{ik}$ with $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, p$, g' is the first-order derivative of g with respect to its argument and ψ (digamma function) is the first derivative of the log of the gamma function.

3. Profile Monitoring Control Charts

In this section, we propose and study five Hotelling T^2 control charts for monitoring linear profiles with compositional data using the Dirichlet regression model described in the previous section. The study is limited for Phase I. We consider the same charts analyzed by Yeh et al. (2009) in their study for profile monitoring in binary responses: T^2 based on the sample mean vector and covariance matrix (T^2_{Usual}), T^2 based on the sample average and successive differences estimator (T^2_{SD}) proposed by Sullivan & Woodall (1996), T^2 based on the sample average and intra−profile pooling (T^2_{Int}) Williams et al. (2007), T^2 based on the Minimum Volume Ellipsoid (T^2_{MVE}) and T^2 based on the Minimum Covariance Determinant (T^2_{MCD}) studied by Vargas (2003) and Jensen, Birch & Woodall (2007).

We assume that when the process is in control, the matrix of model parameters is $\mathbf{B}_0$. In Phase I control, m independent samples are taken. In each sample r, $r = 1, \ldots, m$, there are a set of n_r independent vector observations $\mathbf{Y}_{1r}, \ldots, \mathbf{Y}_{n_r r}$, where $\mathbf{Y}_{ir} = (Y_{ir1}, \ldots, Y_{irp})$ with $Y_{ir1} + \cdots + Y_{irp} = 1$, for each $i = 1, \ldots, n_r$. We suppose that $\mathbf{Y}_{ir} \sim Dirichlet_p(a_{i1}, \ldots, a_{ip})$. We assume that the relationship between the parameters that index the Dirichlet distribution and k explanatory variables $(X_1, \ldots, X_k)$ given in equation (6) is $g_j = exp(\cdot)$.

For any given sample r, $r = 1, 2, \ldots, m$, $\widehat{\mathbf{B}}_r$ is the maximum likelihood estimator of $\mathbf{B}$. Let $\widehat{\boldsymbol{\beta}}_r = vec(\widehat{\mathbf{B}}_r) = (\widehat{\beta}_{11_r}, \widehat{\beta}_{21_r}, \ldots, \widehat{\beta}_{k1_r}, \widehat{\beta}_{12_r}, \widehat{\beta}_{22_r}, \ldots, \widehat{\beta}_{k2_r}, \ldots, \widehat{\beta}_{1p_r}, \widehat{\beta}_{2p_r}, \ldots, \widehat{\beta}_{kp_r})$ $\widehat{\boldsymbol{\beta}}_r$ is a multivariate random vector, where each $\widehat{\beta}_{sj_r}$ represents the estimator of the parameter corresponding to the explanatory variable X_s, $s = 1, \ldots, k$, applied on the j components of $\mathbf{Y}_{ir}$.

The Hotelling's T^2 statistic measures the Mahalanobis distance of the corresponding vector from the sample mean vector. The general form of the statistic is

$$T^2_r = (\widehat{\boldsymbol{\beta}}_r - \boldsymbol{\beta}_0)^t \boldsymbol{\Sigma}_0^{-1} (\widehat{\boldsymbol{\beta}}_r - \boldsymbol{\beta}_0)$$

where $\boldsymbol{\beta}_0$ is the expected value of $\widehat{\boldsymbol{\beta}}_r$ when the process is in control, and $\boldsymbol{\Sigma}_0$ is the in-control covariance matrix of $\widehat{\boldsymbol{\beta}}_r$.

In Phase I control, $\boldsymbol{\beta}_0$ and $\boldsymbol{\Sigma}_0$ both need to be estimated and the performance of the control chart depends on the types of estimates being used. The T^2 statistics for the proposed control charts are calculated by:

$$T^2_{Usual,r} = (\widehat{\boldsymbol{\beta}}_r - \overline{\boldsymbol{\beta}})^t \mathbf{S}^{-1}_{Usual} (\widehat{\boldsymbol{\beta}}_r - \overline{\boldsymbol{\beta}}) \tag{11}$$

where $\overline{\boldsymbol{\beta}} = \frac{1}{m} \sum_{r=1}^{m} \widehat{\boldsymbol{\beta}}_r$ and $\mathbf{S}_{Usual} = \frac{1}{m-1} \sum_{r=1}^{m} (\widehat{\boldsymbol{\beta}}_r - \overline{\boldsymbol{\beta}})(\widehat{\boldsymbol{\beta}}_r - \overline{\boldsymbol{\beta}})^t$

$$T^2_{SD,r} = (\widehat{\boldsymbol{\beta}}_r - \overline{\boldsymbol{\beta}})^t \mathbf{S}^{-1}_{SD} (\widehat{\boldsymbol{\beta}}_r - \overline{\boldsymbol{\beta}}) \tag{12}$$

where $\mathbf{S}_{SD} = \frac{1}{2(m-1)} \sum_{r=1}^{m-1} (\widehat{\boldsymbol{\beta}}_{r+1} - \widehat{\boldsymbol{\beta}}_r)(\widehat{\boldsymbol{\beta}}_{r+1} - \widehat{\boldsymbol{\beta}}_r)^t$

$$T^2_{Int,r} = (\widehat{\boldsymbol{\beta}}_r - \overline{\boldsymbol{\beta}})^t \mathbf{S}^{-1}_{Int} (\widehat{\boldsymbol{\beta}}_r - \overline{\boldsymbol{\beta}}) \tag{13}$$

where $\mathbf{S}_{Int} = \frac{1}{m}\sum_{r=1}^{m} \widehat{var}(\widehat{\beta}_r)$, which is calculated using the observed information matrix,

$$T^2_{MVE,r} = (\widehat{\boldsymbol{\beta}}_r - \widehat{\boldsymbol{\beta}}_{MVE})^t \mathbf{S}^{-1}_{MVE}(\widehat{\boldsymbol{\beta}}_r - \widehat{\boldsymbol{\beta}}_{MVE}), \tag{14}$$

where $\widehat{\boldsymbol{\beta}}_{MVE}$ and $\mathbf{S}_{MVE}$ are estimates of $\boldsymbol{\beta_0}$ and Σ_0, respectively, based on the MVE method (See Rousseeuw & Van Zomeren 1990), and

$$T^2_{MCD,r} = (\widehat{\boldsymbol{\beta}}_r - \widehat{\boldsymbol{\beta}}_{MCD})^t \mathbf{S}^{-1}_{MCD}(\widehat{\boldsymbol{\beta}}_r - \widehat{\boldsymbol{\beta}}_{MCD}), \tag{15}$$

where $\widehat{\boldsymbol{\beta}}_{MCD}$ and $\mathbf{S}_{MCD}$ are estimates of $\boldsymbol{\beta_0}$ and Σ_0, respectively, based on the MCD method (See Rousseeuw & Van Zomeren 1990).

Although $\widehat{\boldsymbol{\beta}}_r$ is distributed asymptotically normal we do not know its sampling distribution. Therefore, we used simulations to approximate the upper control limit (UCL). For simplicity we consider that the number of components is $p = 2, 3, 4, 5$ and 8. For a chart given we generated m independent samples. For each sample we generated a set of n independent vector observations $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$, where $\mathbf{Y}_i \sim Dirichlet_p(a_{i1}, a_{i2}, \ldots, a_{ip})$, $i = 1 \ldots, n$. The parameters of the Dirichlet distribution, a_{ij} with $j = 1, 2, \ldots, p$, are described by $a_{ij} = exp(\beta_{0j} + \beta_{1j} X_i)$ with $\beta_{01} = 2$, $\beta_{11} = 3$, $\beta_{02} = 1$, $\beta_{12} = 4$, $\beta_{03} = 3$, $\beta_{13} = -2$, $\beta_{04} = 0$, $\beta_{14} = 2$, $\beta_{05} = -0.1$, $\beta_{15} = 2.5$, $\beta_{06} = 1$, $\beta_{16} = 2$, $\beta_{07} = 3$, $\beta_{17} = 2$, $\beta_{08} = 1$ and $\beta_{18} = 2.5$. The values of the regressor variable (X) can be random but we have assumed that X takes fixed values, $X = 0.1, 0.2, \ldots, 0.9$. For the m samples generated we calculate the maximum T^2, denoted by $T^2_{\max}$. This process was then repeated 10,000 times which resulted in 10,000 $T^2_{\max}$ values. The 95th quantile of these $T^2_{\max}$ values was then taken as an estimate of the UCL for that chart.

For each of the proposed charts, we ran the simulations for $m = 30, 60$ and 90 samples with a prespecified type I error probability $\alpha = 0.05$. The UCLs obtained are shown in Table 1. We used the R language to run the simulations, in particular we used the DirichletReg-package written by Maier (2011) to calculate the estimates of $\boldsymbol{\beta}_r$. We also used the functions cov.mve and cov.mcd from the MASS-package to calculate $\widehat{\boldsymbol{\beta}}_{MVE}$ and $\mathbf{S}_{MVE}$ in equation (14), and $\widehat{\boldsymbol{\beta}}_{MCD}$ and $\mathbf{S}_{MCD}$ in equation (15).

If a process is modeled using the multivariate normal regression, the response variables can take any real value, (Noorossana, Eyvazian, Amiri & Mahmoud 2010). However, this assumption is not met here, because the response variables for compositional data are always positive and range only from 0 to 100, or any other constant. Therefore, the use of the multivariate normal regression in this kind of processes can produce invalid results. For more details see Aitchison (2003) and Pawlowsky-Glahn & Egozcue (2006).

4. The Performance Evaluation

In this section we compare the performance of the proposed methods for Phase I, monitoring of compositional data, through linear regression profiles in terms of the overall probability of a signal under step and drift shift and outliers. The

TABLE 1: Values of simulated UCL for the proposed control charts with $\alpha = 0.05$

Number of components (p)	Total samples (m)	T^2_{Usual}	T^2_{SD}	T^2_{Int}	T^2_{MVE}	T^2_{MCD}
2	30	20.941	24.610	87.627	212.263	246.696
	60	35.680	37.925	117.805	154.041	164.521
	90	47.359	48.993	139.634	163.259	162.749
3	30	19.505	26.127	56.683	161.556	362.290
	60	30.644	34.035	75.480	102.306	124.628
	90	38.699	41.184	87.004	96.878	103.855
4	30	19.905	30.427	49.128	186.489	799.903
	60	28.324	32.960	58.086	87.104	139.786
	90	32.124	35.336	60.897	70.180	81.265
5	30	20.815	37.366	47.412	253.064	1820.518
	60	28.870	35.098	55.731	95.451	204.833
	90	32.973	37.308	60.638	72.088	92.750
8	30	24.550	74.952	54.258	710.379	11972.690
	60	33.487	47.861	61.769	175.620	521.925
	90	38.155	46.609	65.442	103.688	187.424

signal probability is defined as the probability that at least one sample, of a total of m samples, is considered to be out of control. When the process is out of control, a large signal probability indicates better ability of a control chart to detect the out-of-control process. However, when the process is in control, a large signal probability actually works against a control chart since it gives a higher false alarm rate (See Yeh et al. 2009).

We have that $\sqrt{n}vec(\widehat{\mathbf{B}} - \mathbf{B}) \overset{a}{\sim} N_{kp}(\mathbf{0}, \mathbf{K}(\mathbf{B})^{-1})$. Following equations (6), (8), (9) and (10), the information matrix $\mathbf{K}(\mathbf{B})$ depends on the unknown parameters of the regression and of the values assigned to the regressor variable. For simplicity, we consider that $p = 2$. So, for $\boldsymbol{\beta}_0 = c(2, 3, 1, 4)$ and $X = 0.1, 0.2, \ldots, 0.9$ we have that

$$\Sigma_0 = \mathbf{K}(\mathbf{B})^{-1} = \begin{pmatrix} \sigma^2_{\beta_{01}} & \sigma_{\beta_{01}\beta_{11}} & \sigma_{\beta_{01}\beta_{02}} & \sigma_{\beta_{01}\beta_{12}} \\ \sigma_{\beta_{11}\beta_{01}} & \sigma^2_{\beta_{11}} & \sigma_{\beta_{11}\beta_{02}} & \sigma_{\beta_{11}\beta_{12}} \\ \sigma_{\beta_{02}\beta_{01}} & \sigma_{\beta_{02}\beta_{11}} & \sigma^2_{\beta_{02}} & \sigma_{\beta_{02}\beta_{12}} \\ \sigma_{\beta_{12}\beta_{01}} & \sigma_{\beta_{12}\beta_{11}} & \sigma_{\beta_{12}\beta_{02}} & \sigma^2_{\beta_{12}} \end{pmatrix}$$
$$= \begin{pmatrix} 1.0322 & -1.6290 & 0.9807 & -1.5621 \\ -1.6290 & 3.2702 & -1.5615 & 3.1763 \\ 0.9807 & -1.5615 & 1.0041 & -1.5926 \\ -1.5621 & 3.1763 & -1.5926 & 3.2218 \end{pmatrix}$$

Let $\Delta = (\delta_1\sigma_{\beta_{01}}, \delta_2\sigma_{\beta_{11}}, \delta_3\sigma_{\beta_{02}}, \delta_4\sigma_{\beta_{12}})$, where $\delta_j = 0, 1, 2, 3$, $j = 1, 2, 3, 4$. If $\boldsymbol{\beta}_r$ changes from $\boldsymbol{\beta}_0$ to $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_0 + \Delta$, with $\Delta \neq \mathbf{0}$, the process is out-of-control. The level of shifts in $\boldsymbol{\beta}_r$ is described by the non-centrality parameter (ncp). The non-centrality parameter measures the severity of a shift to the out-of-control vector $\boldsymbol{\beta}_1$ from the in-control vector $\boldsymbol{\beta}_0$ and is defined by $ncp = \Delta^t\Sigma_0^{-1}\Delta = (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)^t\Sigma_0^{-1}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)$ (See Vargas (2003) and Yeh et al. (2009)).

The out-of-control signal probabilities were calculated based on 5,000 simulations, as the percentage of times the $T^2_{\max}$ exceeds the corresponding UCL.

For step shift or sustained shift, we generate a shift in the vector of parameters of the regressions, $\boldsymbol{\beta}$, from $\boldsymbol{\beta}_0$ to $\boldsymbol{\beta}_1$. The shift starts from the sample l, for $l = [m * k] + 1$, where $[x]$ denotes the largest integer which is less or equal than x and $k = \frac{1}{4}, \frac{1}{2}$ and $\frac{3}{4}$. So, for $k = \frac{1}{2}$ the first half of the samples is in-control, while the second half is in the out-of-control state. The signal probabilities for each control chart, when $m = 30$, were calculated through simulation.

Table 2: Signal probabilities when the intercept and the slope of the profile corresponding to the second component have not changed.

ncp	δ_1	δ_2	δ_3	δ_4	T^2_{Usual}	T^2_{SD}	T^2_{Int}	T^2_{MVE}	T^2_{MCD}
0	0	0	0	0	0.0492	0.0494	0.0464	0.0452	0.0488
175.0568	1	0	0	0	0.0114	0.8052	0.8966	0.0312	0.0346
700.2274	2	0	0	0	0.0134	0.8976	1	0.0264	0.032
1575.512	3	0	0	0	0.0146	0.9078	1	0.0294	0.0328
297.7102	0	1	0	0	0.0086	0.843	0.9912	0.0234	0.031
910.8436	1	1	0	0	0.0102	0.8974	1	0.0284	0.031
1874.091	2	1	0	0	0.0092	0.8996	1	0.0256	0.0344
3187.452	3	1	0	0	0.0096	0.9144	1	0.025	0.0336
1190.841	0	2	0	0	0.009	0.9004	1	0.019	0.028
2242.051	1	2	0	0	0.0072	0.9142	1	0.0268	0.0322
3643.375	2	2	0	0	0.0056	0.9124	1	0.0222	0.029
5394.812	3	2	0	0	0.006	0.9058	1	0.03	0.0284
2679.391	0	3	0	0	0.0096	0.907	1	0.026	0.0328
4168.678	1	3	0	0	0.007	0.9064	1	0.0242	0.0294
6008.079	2	3	0	0	0.0046	0.8998	1	0.0256	0.0322
8197.593	3	3	0	0	0.0072	0.9016	1	0.0268	0.0338
6008.079	2	3	0	0	0.006	0.901	1	0.0262	0.033

Table 2 shows the signal probabilities of the five control charts considered for a step shift occuring in $l = [m/2] + 1$ when the intercept and the slope of the profile corresponding to the second component have not changed. When $ncp = 0$ the signal probabilities for the T^2_{Usual}, T^2_{SD}, T^2_{Int}, T^2_{MVE} and T^2_{MCD} control charts are close to 0.05. For other values of ncp, the signal probabilities of the T^2_{Int} control chart are 1 o very near 1, showing an excellent performance to detect step shifts.

Figures 1 to 5 describe the signal probabilities of the T^2_{Usual}, T^2_{SD}, T^2_{Int}, T^2_{MVE} and T^2_{MCD} control charts for a step shift occurring in three scenarios: the last three quarters, the second half, and the last quarter of the 30 samples considered. With exception of the T^2_{Int} control chart, the location of the step shift affects the performance of the T^2 control charts considered. For example, the signal probabilities decrease considerably when the shift stars in the half of the samples. The effect is greater in the T^2_{MVE} and T^2_{MCD} control charts. These charts are more powerful when the shifts occur at $k = \frac{1}{4}$ and $k = \frac{3}{4}$.

For drift shifts, the first sample generated was in control, $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_0$, and the process parameter vector started to change from the second sample to $\boldsymbol{\beta}_1$, where $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_0 + \frac{r-1}{m-1}(\Delta)$, with $r = 2 \ldots, m$ and $m = 30$. Figure 6 shows the signal probabilities found by simulation for the 256 possible values of ncp. We observe

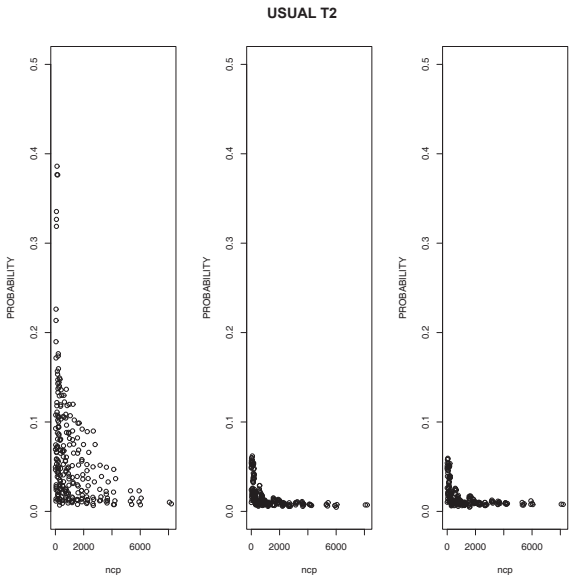

FIGURE 1: Signal probabilities of the T^2_{Usual} control chart for a step shift occurring in $l = [m * k] + 1$, with $k = \frac{1}{4}$ (on the left), $k = \frac{2}{4}$ (on the middle), and $k = \frac{3}{4}$ (on the right), when the number of samples is $m = 30$.

that the T^2_{SD}, T^2_{Int} control charts have a good performance for detecting shifts with trend.

In the scenario considering the presence of outliers, 5 of them were inserted randomly in the m samples, with $m = 30$. They were generated from $\boldsymbol{\beta}_1$, where $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_0 + (\Delta)$. Figure 7 presents the signal probabilities calculated by simulations. T^2_{Int}, T^2_{MVE} and T^2_{MCD} control charts have the best performance for detecting outliers.

5. Example of Application

The concrete is a composite material that essentially consists of a mixture of cement, water and aggregates, which is regularly used in infrastructure and buildings construction (Li 2011). The aggregates are rock fragments named coarse aggregate and sand particles called fine aggregate, which be derived from land- or sea-based deposits, from gravel pits or hard-rock quarries, from sand dunes or river courses. The aggregate occupies between 70% and 75% of the concrete volume and affect its strength, durability, workability and cohesiveness. One aspect of interest in the quality of the aggregate is the particle size distribution known as gradation (Alexander & Mindess 2005).

In order to obtain the gradation of the aggregate, a series of standard sieves are nested or stacked, one on top of another, with increasing aperture size from bottom to top, and through which a aggregate sample is passed from top, usually

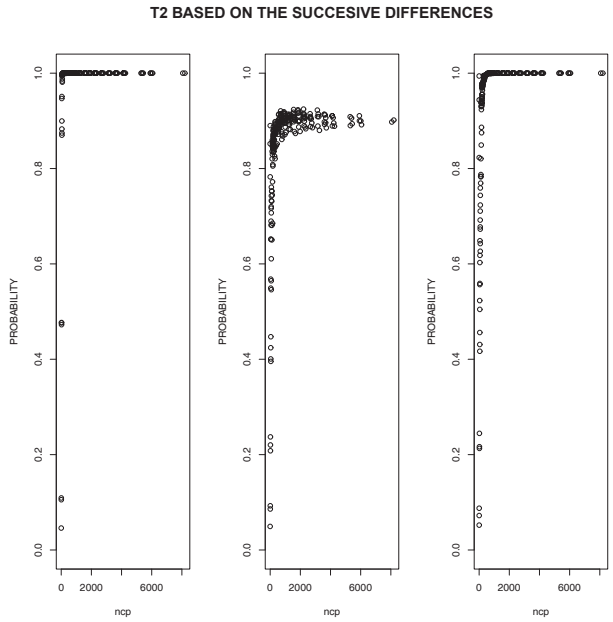

FIGURE 2: Signal probabilities of the T^2_{SD} control chart for a step shift occurring in $l = [m * k] + 1$, with $k = \frac{1}{4}$ (on the left), $k = \frac{2}{4}$ (on the middle), and $k = \frac{3}{4}$ (on the right), when the number of samples is $m = 30$.

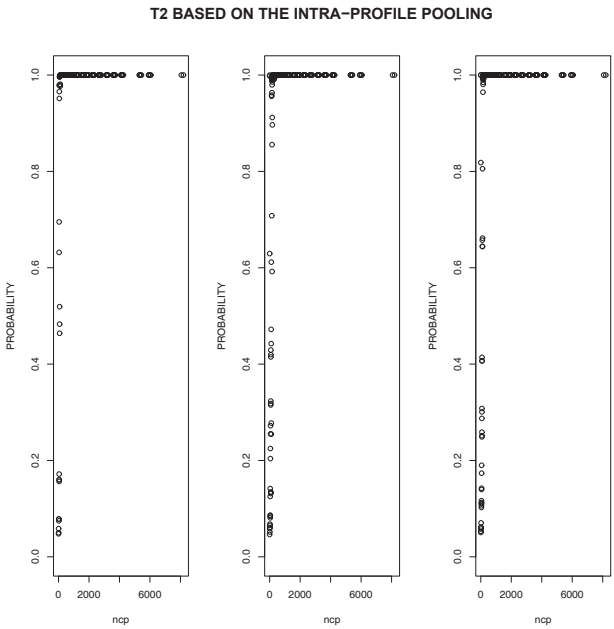

FIGURE 3: Signal probabilities of the T^2_{Int} control chart for a step shift occurring in $l = [m * k] + 1$, with $k = \frac{1}{4}$ (on the left), $k = \frac{2}{4}$ (on the middle), and $k = \frac{3}{4}$ (on the right), when the number of samples is $m = 30$.

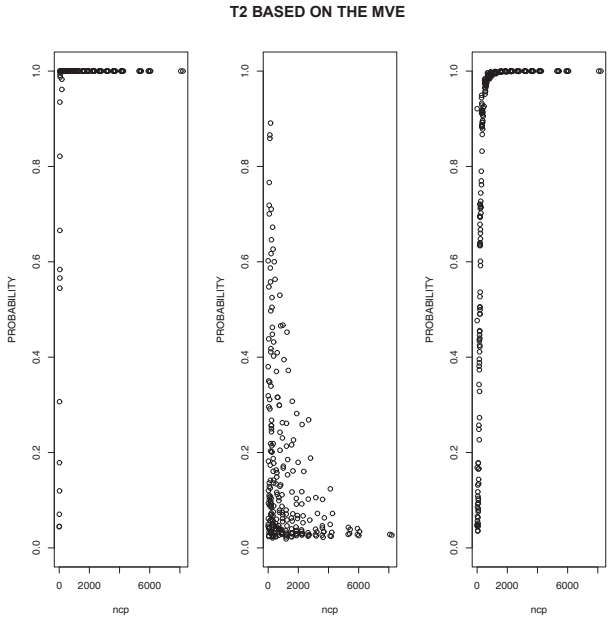

FIGURE 4: Signal probabilities of the T^2_{MVE} control chart for a step shift occurring in $l = [m * k] + 1$, with $k = \frac{1}{4}$ (on the left), $k = \frac{2}{4}$ (on the middle), and $k = \frac{3}{4}$ (on the right), when the number of samples is $m = 30$.

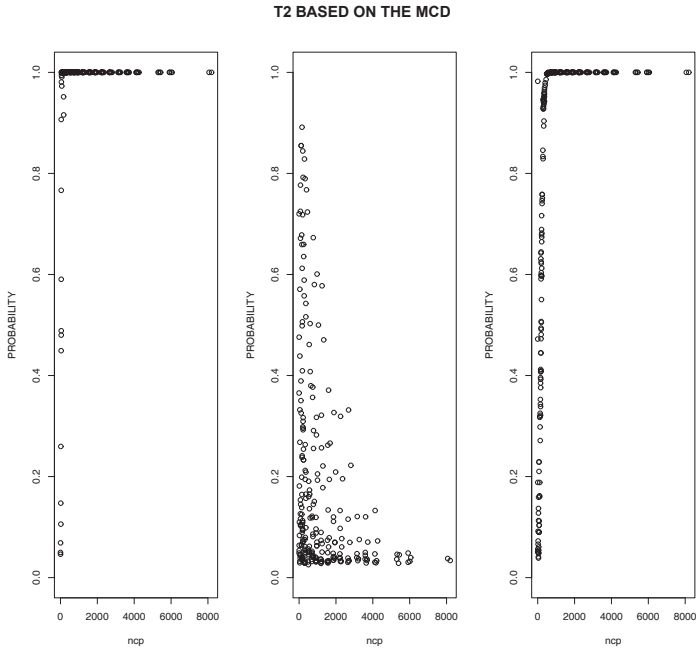

FIGURE 5: Signal probabilities of the T^2_{MCD} control chart for a step shift occurring in $l = [m * k] + 1$, with $k = \frac{1}{4}$ (on the left), $k = \frac{2}{4}$ (on the middle), and $k = \frac{3}{4}$ (on the right), when the number of samples is $m = 30$.

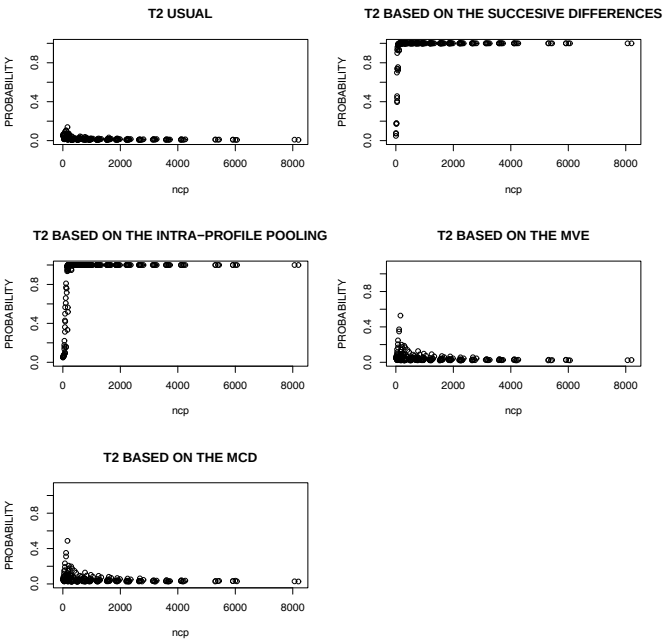

FIGURE 6: Signal probabilities of drift shifts for five control charts: usual, successive differences, intra-profile, MVE and MCD.

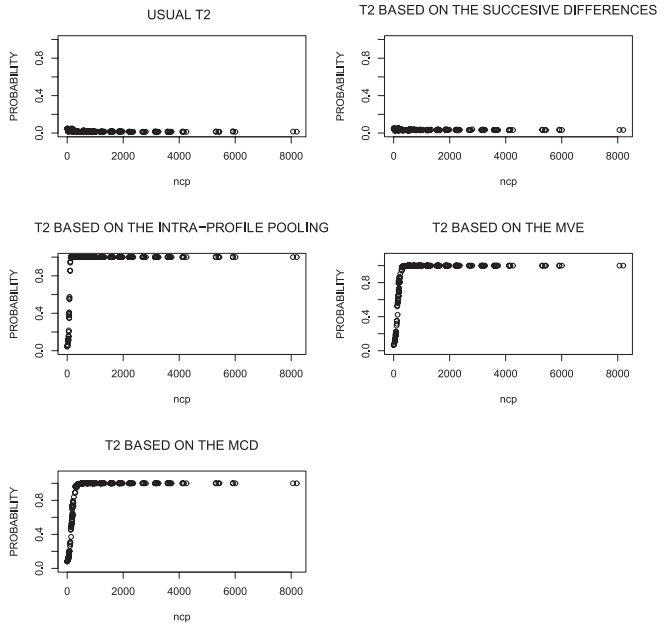

FIGURE 7: Signal probabilities of outliers for five control charts: usual, successive differences, intra-profile, MVE and MCD.

aided by shaking or vibrating the sieves (Alexander & Mindess 2005, Lyons 2008). Figure 8 shows a kind of machine used in the gradation process. The sieves labeled as 200, 100, 50, 30, 16, 8, 4, and P3, have the hole sizes of 0.075 mm, 0.149 mm, 0.297 mm, 0.595 mm, 1.19 mm, 2.38 mm, 4.75 mm and 9.5 mm, respectively. The gradation results are the percent of aggregate retained on each sieve and the fineness modulus, which measures the average particle size. This dimensionless parameter is equal to sum of the percent of aggregate retained on each of sieve divided by 100. A smaller fineness modulus indicates a finer aggregate and a higher value represents a courser aggregate.

FIGURE 8: Series of sieves placed one above the other in order of size with the largest sieve at the top.

In this work, 217 aggregate samples from a concrete manufacturing plant were studied. The aggregate samples were daily tested during 31 weeks. The set of daily observations obtained in a week is named weekly sample, therefore, 31 weekly samples were considered. The proportion passing through of each sieve and the fineness modulus were measured in each aggregate sample.

The components $j = 1, 2, \ldots, 8$ are defined by the aggregate size retained by each sieve. The proportion passing through the sieve j, $j = 1, \ldots, 8$, is the variable Y_j. Each component corresponds to an aggregate with constant size and the proportion of aggregate passing trough them is identified by the vector $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_8)$. Figure 9 shows plots of the marginal frequencies of each component for the aggregate samples. We observe that Y_4, Y_5 and Y_7 are skewed, which implies that $\mathbf{Y}$ does not have a multivariate normal distribution.

The weekly sample r, $r = 1, \ldots, 31$, contains the daily observations $(x_{ir}, \mathbf{Y}_{ir})$, $i = 1, \ldots, n$, with $n = 7$. The vectors $\mathbf{Y}_{1r}, \ldots, \mathbf{Y}_{nr}$ are independent and $\mathbf{Y}_{ir} \sim Dirichlet_8(a_{i1}, \ldots, a_{i8})$. There is a relationship between the fineness modulus and the proportion of aggregate passing through each sieve. Figures 10 and Figures 11 show these relationships for the components $j = 1, 2, 3, 4, 5, 6, 7, 8$ associated to the sieves 200, 100, 50, 30, 16, 8, 4, and P3, respectively. A likelihood ratio test (LRT) for each sample r, shows that the Dirichlet regression models are significant at the 10%, so the null hypothesis $H_0 : \beta_{11} = \beta_{12} = \cdots = \beta_{18} = 0$ is rejected.

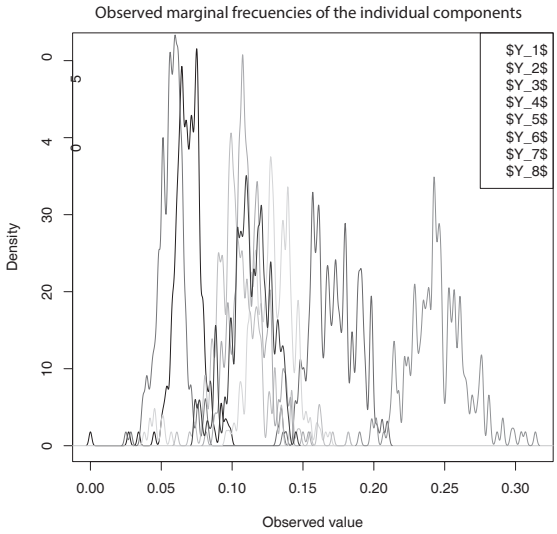

FIGURE 9: Observed marginal frequencies $Y_1, Y_2, \ldots, Y_8$ of the individual components of the aggregate gradings.

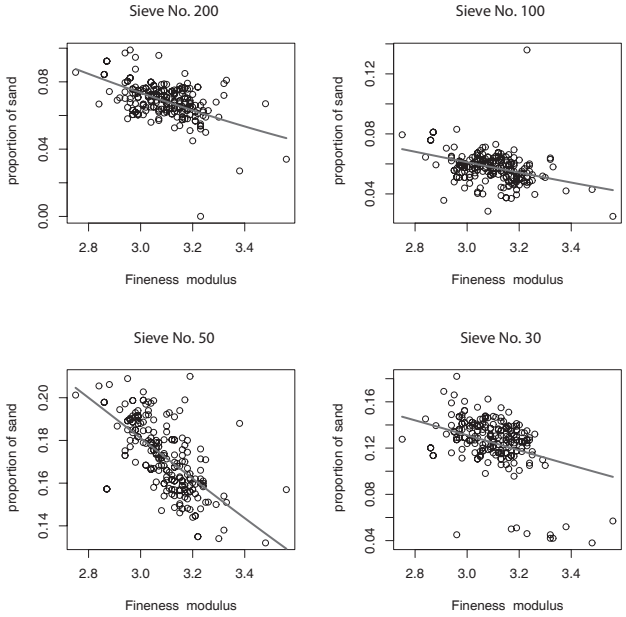

FIGURE 10: Linear relationship between fineness modulus and the proportion of sand passing through the sieves No 200, 100, 50 and 30.

Through simulations we found the upper control limit for each T^2 control chart proposed in the section (3). The T^2 control chart based on successive differences suggests that the process does not present step and drift shifts, but the control chart based on MVE detects the presence of outliers, see Figures 12 and 13.

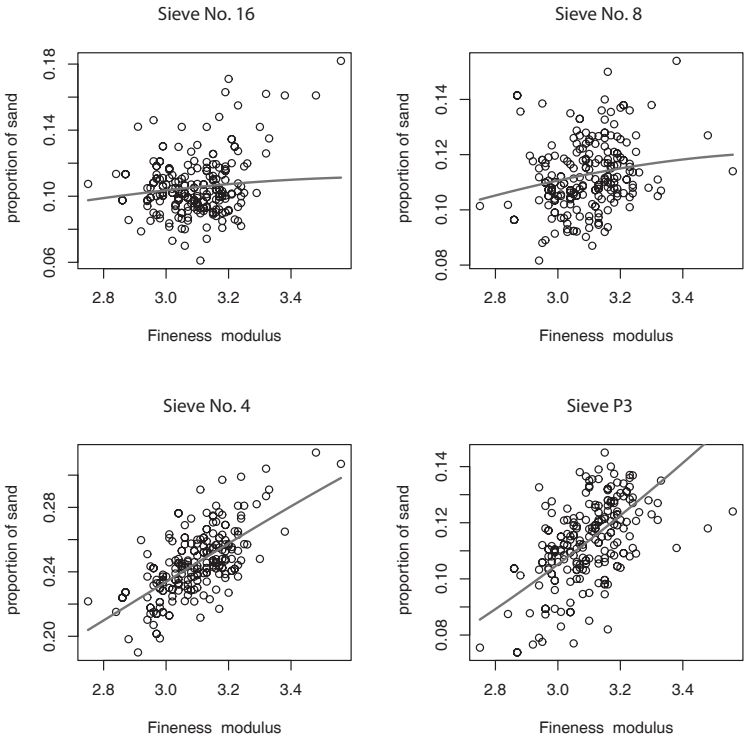

FIGURE 11: Linear relationship between fineness modulus and the proportion of sand passing through the sieves No 16, 8, 4 and P3.

Although the engineers believed that the process was in-control, the intra-profile control chart shows a lot of points outside the upper control limit, see Figure 15. The usual T^2 control chart detects some of these points, see Figure 16. Figure 17 describes the behaviour of the linear regressions associated with the first sieve from the sample. This graph shows that the profile is not stable. The other sieves have a similar behaviour. As a first result of this application, engineers are reviewing and adjusting the process to ensure that the linear relationship associated with each sieve is in-control.

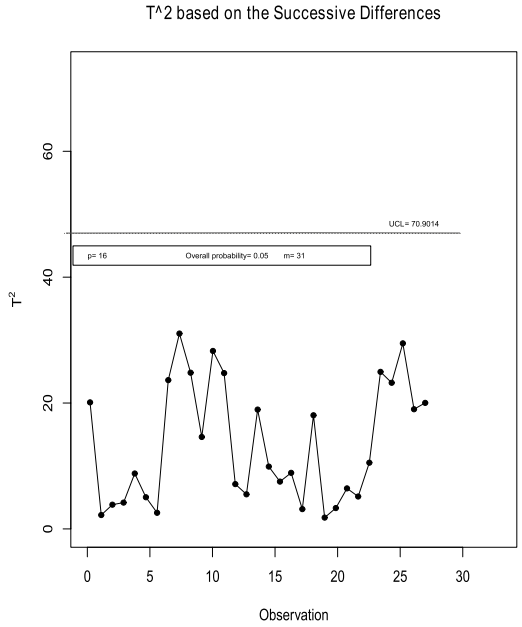

FIGURE 12: T^2 Control chart based on successive differences for the process of grading of sand in a mine of a concrete manufacturing plant.

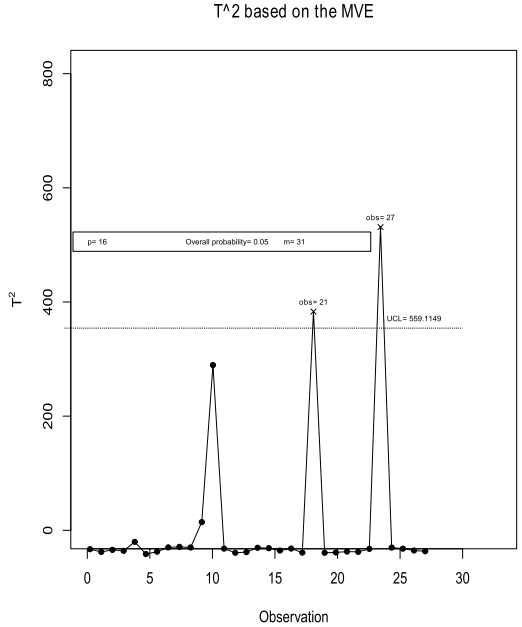

FIGURE 13: T^2 Control chart based on Minimum Volume Ellipsoid (MVE) for the process of grading of sand in a mine of a concrete manufacturing plant.

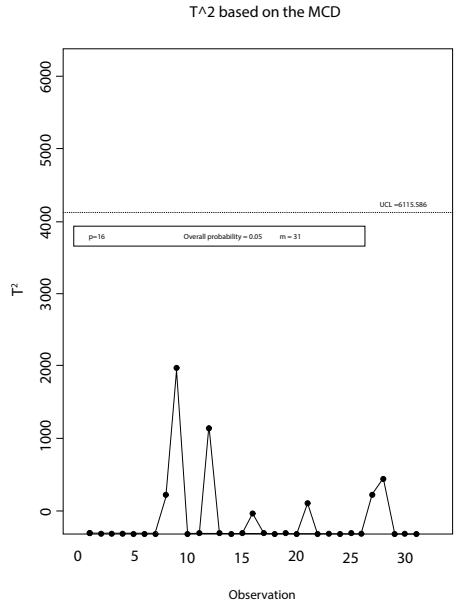

FIGURE 14: T^2_{MCD} for the process of grading of sand in a mine of a concrete manufacturing plant.

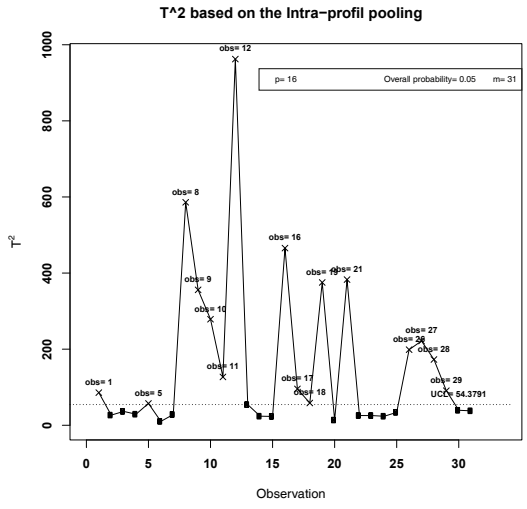

FIGURE 15: T^2 Control chart based on intra-profile pooling for the process of grading of sand in a mine of a concrete manufacturing plant.

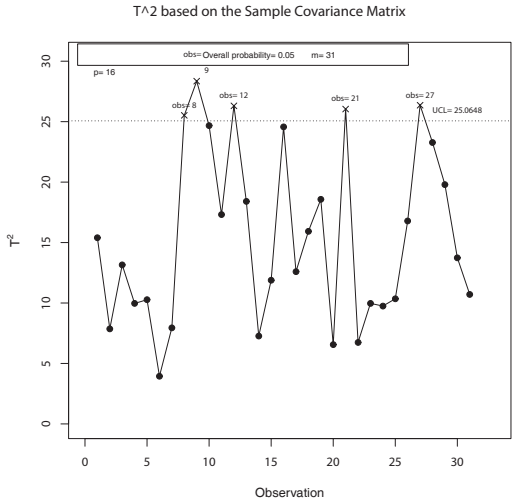

FIGURE 16: Usual T^2 control chart for the process of grading of sand in a mine of a concrete manufacturing plant.

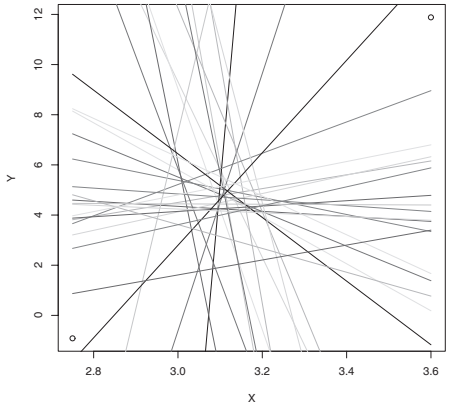

FIGURE 17: Linear regressions for the first component of the sand gradation.

6. Conclusions

In this paper the control charting mechanisms discussed by Williams et al. (2007) and Yeh et al. (2009) have been extended for monitoring compositional data profiles in Phase I processes, whose response variable follows a Dirichlet distribution. This methodology allows us monitoring the linear relationship between the parameters of a Dirichlet distribution and a set of explanatory variables, and assess the stability of the parameters that characterize the studied regression model.

We used five Hotelling's type T^2 control charts and compared their performance for detecting step and drift shifts in the process parameters and outliers in the studied profiles. Simulation procedures suggest that the T^2 control chart called intra-profile pooling is an excellent tool in order to detect outliers in the profiles and step and drift shifts in the parameters of the compositional data profile. The intra-profile pooling T^2 control chart is based on the average of the sample covariance matrices of the estimates of the parameters $\boldsymbol{\beta}$ characterizing the profile. The T^2 control chart based on the vector of successive differences of parameter estimates is a good alternative for detecting step and drift shifts; while the T^2 control charts based on robust estimates for the mean and covariance matrix, minimum volume ellipsoid (MVE) and minimum covariance determinant (MCD) methods, are a good option for detecting outliers.

We presented an example of application with real data of the proposed methodology, in order to control the quality of the aggregate gradation in the concrete. The T^2 control chart based on successive differences suggests that the process is in-control and does not present step and drift shifts, the control chart based on MVE detects the presence of some outliers, and the intra-profile and usual T^2 control charts show that the process is out-control. This methodology can be extended to other processes with compositional data.

This paper constitutes an initial solution for monitoring compositional data profiles. It would be worthwhile to study and compare the performance of other control charts like the change point approach. Since the performance of the T^2 control charts deteriorates when number of parameters increases, it is needed more research when the number of components in the response variable increase and/or when the number of covariates increases. Reduction methods for multivariate data or high dimensional methods need also future research.

When the process is out-of-control is important to identify the causes of the anomaly in order to apply appropriate remedial measures. A future work can implement diagnostic aids such as determining the parameters responsible for out-of-control signal.

Finally, some methods for monitoring Dirichlet regression profiles in Phase II can be developed.

Acknowledgments

The authors want to thank Professor William H. Woodall from Virgina Tech, the Editor, and two referees for their helpful comments that have strengthened this paper.

[Recibido: octubre de 2013 — Aceptado: marzo de 2014]

References

Aitchison, J. (1986), *The Statistical Analysis of Compositional Data*, Chapman & Hall.

Aitchison, J. (2003), *The Statistical Analysis of Compositional Data*, Blackburn Press.

Alexander, M. & Mindess, S. (2005), *Aggregates in Concrete*, Taylor & Francis.

Amiri, A., Zou, C. & Doroudyan, M. H. (2014), 'Monitoring correlated profile and multivariate quality characteristics', *Quality and Reliability Engineering International* **30**(1), 133–142.

Boyles, R. A. (1997), 'Using the chi-square statistic to monitor compositional process data', *Journal of Applied Statistics* **24**(5), 589–602.

Eyvazian, M., Noorossana, R., Saghaei, A. & Amiri, A. (2011), 'Phase II monitoring of multivariate multiple linear regression profiles', *Quality and Reliability Engineering Intenational* **27**(3), 281–296.

Ferrari, S. L. & Cribari-Neto, F. (2004), 'Beta regression for modelling rates and proportions', *Journal of Applied Statistics* **31**(7), 799–815.

Gueorguieva, R., Rosenheck, R. & Zelterman, D. (2008), 'Dirichlet component regression and its applications to psychiatric data', *Computational Statistics and Data Analysis* **52**(12), 5344–5355.

Jensen, W. A., Birch, J. B. & Woodall, W. H. (2007), 'High breakdown estimation methods for phase I multivariate control charts', *Quality and Reliability Engineering International* **23**(5), 615–629.

Kang, L. & Albin, S. L. (2000), 'On-line monitoring when the process yields a linear profile', *Journal of Quality Technology* **32**(4), 418–426.

Kim, K., Mahmoud, M. A. & Woodall, W. H. (2003), 'On the monitoring of linear profiles', *Journal of Quality Technology* **35**(3), 317–328.

Kusiak, A., Zheng, H. & Song, Z. (2009), 'Models for monitoring wind farm power', *Renewable Energy* **34**(3), 583–590.

Li, Z. (2011), *Advanced Concrete Technology*, John Wiley & Sons.

Lyons, A. (2008), *Materials for Architects and Builders*, fourth edn, Elsevier.

Mahmoud, M. A. (2008), 'Phase I analysis of multiple linear regression profiles', *Communications in Statistics-Simulation and Computation* **37**(10), 2106–2130.

Mahmoud, M. A., Parker, P. A., Woodall, W. H. & Hawkins, D. M. (2007), 'A change point method for linear profile data', *Quality and Reliability Engineering International* **23**(2), 247–268.

Mahmoud, M. A. & Woodall, W. H. (2004), 'Phase I analysis of linear profiles with calibration applications', *Technometrics* **46**(4), 380–391.

Maier, M. J. (2011), Dirichletreg: Dirichlet regression in R. R package version 0.3-0.

Melo, T. F., Vasconcellos, K. L. & Lemonte, A. (2009), 'Some restriction tests in a new class of regression models for proportions', *Computational Statistics and Data Analysis* **53**(12), 3972–3979.

Ng, K. W., Tian, G.-L. & Tang, M.-L. (2011), *Dirichlet and related distributions: Theory, methods and applications.*, John Wiley & Sons.

Noorossana, R., Eyvazian, M., Amiri, A. & Mahmoud, M. A. (2010), 'Statistical monitoring of multivariate multiple linear regression profiles in phase I with calibration application', *Quality and Reliability Engineering International* **26**(3), 291–303.

Noorossana, R., Eyvazian, M. & Vaghefi, A. (2010), 'Phase II monitoring of multivariate simple linear profiles', *Computers & Industrial Engineering* **58**(4), 563–570.

Noorossana, R., Saghaei, A. & Amiri, A. (2012), *Statistical Analysis of Profile Monitoring*, John Wiley & Sons.

Pawlowsky-Glahn, V. & Egozcue, J. (2006), *Compositional Data Analysis in the Geosciences: From Theory to Practice*, Geological Society, chapter Compositional data and their analysis: An introduction, pp. 1–10.

Qiu, P. (2013), *Introduction to Statistical Process Control*, CRC Press.

Rousseeuw, P. & Van Zomeren, B. (1990), 'Unmasking multivariate outlier and leverage points', *Journal of the American Statistical Association* **85**(411), 633–639.

Soleimani, P., Narvand, A. & Raissi, S. (2013), 'Online monitoring of autocorrelated linear profiles via mixed model', *International Journal of Manufacturing Technology and Management* **27**(4), 238–250.

Stover, F. & Brill, R. (1998), 'Statistical quality control applied to ion chromatography calibrations', *Journal of Chromatography* **804**(1-2), 37–43.

Sullivan, J. H. & Woodall, W. H. (1996), 'A comparison of multivariate control charts for individual observations', *Journal of Quality Technology* **28**(4), 398–408.

Vargas, J. A. (2003), 'Robust estimation in multivariate control charts for individual observations', *Journal of Quality Technology* **35**(4), 367–376.

Vasconcellos, K. L. & Cribari-Neto, F. (2005), 'Improved maximum likelihood estimation in a new class of Beta regression models', *Brazilian Journal of Probability and Statistics* **19**(1), 13–31.

Vives-Mestres, M., Daunis-i Estadella, J. & Martín-Fernández, J. A. (2013), 'Out of control signals in three part compositional t^2 control chart', *Quality and Reliability Engineering Intenational* **30**(3), 337–346.

Wang, K. & Tsung, F. (2005), 'Using profile monitoring techniques for a data-rich environment with huge sample size', *Quality and Reliability Engineering International* **21**(7), 677–688.

Williams, J. D., Woodall, W. H. & Birch, J. B. (2007), 'Statistical monitoring of nonlinear product and process quality profiles', *Quality and Reliability Engineering International* **23**(8), 925–941.

Woodall, W. H. (2007), 'Current research in profile monitoring', *Producao* **17**(3), 420–425.

Woodall, W. H., Spitzner, D. J., Montgomery, D. C. & Gupta, S. (2004), 'Using control charts to monitor process and product quality profiles', *Journal of Quality Technology* **36**(3), 309–320.

Yang, G., Cline, D. B. H., Lytton, R. L. & Little, D. N. (2004), 'Ternary and multivariate quality control charts of aggregate gradation for hot mix asphalt', *Journal of Materials in Civil Engineering* **16**(1), 28–34.

Yeh, A. B., Huwang, L. & Li, Y.-M. (2009), 'Profile monitoring for a binary response', *IIE Transactions* **41**(11), 931–941.

Yeh, A. & Zerehsaz, Y. (2013), 'Phase I control of simple linear profiles with individual observations', *Quality and Reliability Engineering Intenational* **29**(6), 829–840.

Zhang, Y., He, Z., Zhang, C. & Woodall, W. H. (2013), 'Control charts for monitoring linear profiles with within profile correlation using gaussian process models', *Quality and Reliability Engineering Intenational* . DOI: 10.1002/qre.1502.

Zou, C., Ning, X. & Tsung, F. (2012), 'LASSO-based multivariate linear profile monitoring', *Annals of Operations Research* **192**(1), 3–19.

Zou, C., Zhang, Y. & Wang, Z. (2006), 'A control chart based on a change-point model for monitoring linear profiles', *IIE Transactions* **38**(12), 1093–1103.

Zou, C., Zhou, C., Wang, Z. & Tsung, F. (2007), 'A self-starting control chart for linear profiles', *Journal of Quality Technology* **39**(4), 364–375.