Gest. Prod., São Carlos, v. 23, n. 1, p. 146-164, 2016
http://dx.doi.org/10.1590/0104-530X1445-14

Method for determining the control limits of nonparametric charts for monitoring location and scale

Método de estimativa dos limites da carta de controle não paramétrica que monitora simultaneamente a média e variância

Pedro Carlos Oprime[1]
José Carlos de Toledo[1]
Mario Orestes Aguirre González[2]
S. Chakraborti[3]

Abstract: Classic control charts for continuous variables monitor, separately, the central position and dispersion measures, which are also known as lease and scale measures. The monitoring of these parameters presupposes that the data probability distribution is known and follows the normal pattern, which, in practical situations, does not always occur. For this reason, the so-called nonparametric control graphics have been developed. This work aims to develop a method to determine the limits of statistical control of control charts with unknown probabilistic distribution and, simultaneously, monitor the mean and variance parameters. The proposed method is not exact and allows us to estimate the limits of control charts for combinations of values of *m* and *n*. Control limits were estimated and the properties of the statistics used were analyzed to determine whether they meet the theoretical assumptions; the empirical models obtained were validated by residual analysis. An empirical application of the method was performed to test different combinations of sample sizes (m, n), respectively in phases I and II, with the goal of identifying the best performing combination for detecting special causes acting in the process. Subsequently, we tested the performance of control charts obtained using simulation methods, estimating the ARL and α and β errors. The results were compared with other designs of control charts.

Keywords: Statistical process control; Nonparametric control chart; Control chart limits.

Resumo: *Os gráficos de controle clássicos para variáveis contínuas monitoram, separadamente, as medidas de posição central e de dispersão, conhecidas também como medidas de locação e escala. O monitoramento desses parâmetros tem como pressuposto que a distribuição de probabilidade dos dados seja conhecida e siga o padrão normal, o que, em situações práticas, nem sempre ocorre. Para isso foram desenvolvidos os chamados gráficos de controle não paramétricos. Este trabalho tem como objetivo desenvolver um método para determinar os limites de controle estatístico de gráficos de controle com distribuição de probabilidade desconhecida e que monitore simultaneamente as medidas de locação e escala. O método de pesquisa utilizado integra técnicas de experimentos computacionais com técnicas de planejamento de experimentos. Assim, foi possível: i) determinar os limites de controle de gráficos não paramétricos que monitorem simultaneamente as medidas de posição e escala para situações particulares; ii) a partir dos limites de controle calculados, estimar os erros tipo I e tipo II; e iii) comparar o desempenho desses gráficos com as cartas de controle estatístico de Shewhart para diferentes combinações de amostras (m, n) nas fases I e II. O método proposto foi aplicado em um processo de manufatura com o objetivo de identificar a combinação que minimize os erros tipo I e II. Com base nos resultados, observou-se que o gráfico de controle não paramétrico tem desempenho superior aos gráficos tradicionais de Shewhart quando a distribuição de probabilidade dos dados é assimétrica.*

Palavras-chave: *Controle estatístico de processo; Carta de controle não paramétrica; Determinação de limites de controle.*

[1] Universidade Federal de São Carlos – UFSCar, Rodovia Washington Luís, Km 235, SP-310, CEP 13565-905, São Carlos, SP, Brasil, e-mail: pedro@dep.ufscar.br; toledo@dep.ufscar.br

[2] Universidade Federal do Rio Grande do Norte – UFRN, Campus Universitário Lagoa Nova, CEP 59072-970, Natal, RN, Brasil, e-mail: mario@ct.ufrn.br

[3] Department of Information Systems, Statistics and Management Science, University of Alabama, Tuscaloosa, AL 35487 U.S.A. e-mail: schakrab@cba.ua.edu

Received May 6, 2014 – Accepted Apr. 14, 2015

Financial support: Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP.

1 Introduction

The importance of the statistical process control (SPC) as a research topic can be seen in Figure 1, which shows the growth in the number of indexed publications in the *Web of Science* database, from 1956 to 2013. For over fifty years, SPC has played a key role in monitoring and improving the quality and productivity of industrial processes (Baker & Brobst, 1996; Graves et al., 1999; Duarte & Saraiva, 2008), initially based on the classical Shewhart control chart, which assumes that the statistical parameters of the process, such as mean and standard deviation, are known. The primary issue related to SPC lies on understanding the variability of a quality characteristic, establishing process control and promoting its improvement (Woodall, 2000).

Process parameters are usually unknown, and it affects the efficiency in the use of control charts to detect a special cause, since the control limits are usually calculated based on the estimates of these parameters (Jensen et al., 2006; Castagliola et al., 2009; Castagliola & Maravelakis, 2011). When the process parameters are unknown, they are typically estimated and the control limits are determined from *k* samples of *n* size, obtained from retrospective data called phase I analysis In phase II, *n*-size samples are extracted from the process in order to check whether it is under control. If the plotting statistic is not within the control limits, the process is considered to be out of control, and a probable considerable cause must be identified and corrective actions must be taken to restore the *status quo* (Montgomery, 1992).

Recent studies have evaluated the performance of control charts, in both phase I and II, when the parameters are unknown, proposed in order to establish new procedures to improve the performance of these charts and thus minimize α (type I error) and β risks (type II error) (Chen, 1997; Jones et al., 2001; Epprecht et al., 2005; Chakraborti & Human, 2006; Chakraborti, 2006; Castagliola et al., 2009; Costa et al., 2009; Ozsan et al., 2009; Costa et al., 2010; Trovato et al., 2010; Zhang & Castagliola, 2010; Boone & Chakraborti, 2011; Castagliola & Maravelakis, 2011; Costa & Machado, 2011; Zhang et al., 2011; Castagliola & Wu, 2012; Lee, 2013).

ARL (Average Run Length) is commonly used to measure the control chart performance in phase II and it indicates the mean number of samples required to detect a change in the process parameters. Thus, a control chart type is considered better than the others when it shows lower ARL during the monitoring phase. However, if the process is under control, it is desirable that the ARL is as high as possible. A practical problem in applying the classical Shewhart control charts is that its efficiency (ARL) is affected by the probability distribution governing the process. Non-parametric methods are more efficient when the data distribution is unknown or asymmetrical, (Montgomery, 2004; Chakraborti & Human, 2006). According to Boone & Chakraborti (2011), non-parametric methods have the advantage of requiring fewer statistical assumptions about the data distribution and of being relatively easy to be applied to the shop floor.

Traditional control charts have been designed to monitor two parameters, one, the measure of central tendency and two, the dispersion, usually measured by the mean and the standard deviation. The reasons for monitoring these two parameters are found in Box et al. (1978), Montgomery & Runger (2003), and McCracken & Chakraborti (2013). However, proposals for simultaneous.

y monitoring these two parameters in a single chart, especially the non-parametric control chart, have been highlighted in scientific publications (Mccracken &

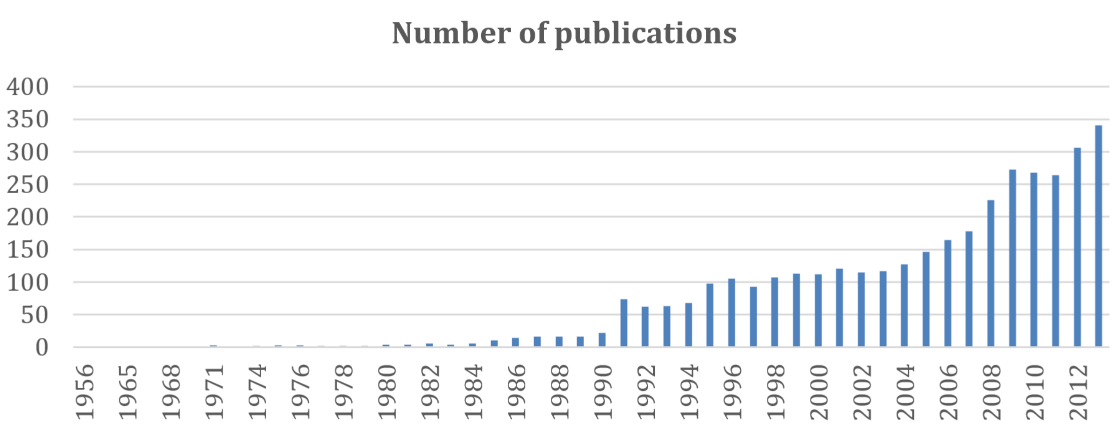

Figure 1. Publications on statistical process control extracted from the *Web of Science* database, of 1956 - 2013 (Thomson Reuters, 2013).

Chakraborti, 2013). This chart is easy to be used by managers and operators on the shop floor, because a single chart is used to identify the presence of special causes in the process.

The combined monitoring of location and scale measurements with a nonparametric chart was analyzed by Mukherjee & Chakraborti (2012), who used computer simulations to find the control limits (H, H_1 and H_2) of a set of sample size combinations for phases I (m) and II (n). However, the results as tabulated by the authors are restricted to a set of values m and n, which restrict its practical use.

A bibliometric research conducted at the *Web of Science* database indicates that there are few studies about the use of non-parametric techniques to monitor processes. Figure 2 shows, by means of cumulative frequency, the records of articles published in the last thirty years. The relationships between keywords relevant to studies about non-parametric methods are found

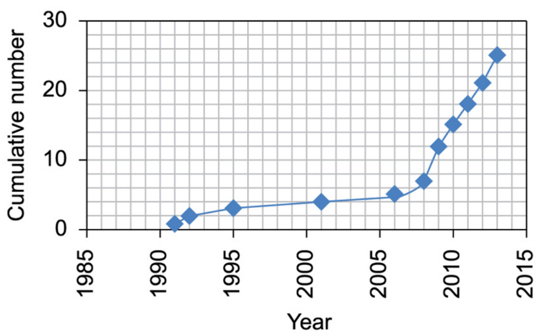

Figure 2. Cumulative number of publications on non-parametric control charts. Source: Research Data.

in Figure 3; for example, the co-occurrence between the keyword "NON-PARAMETRIC" and the words "CUSUM", "RUN LENGTH" and "DISTRIBUTION FREE". Figure 2 shows the increased number of researches on the subject since 2006, thus indicating that it is relatively new in the study about process statistical control. Figure 4 shows the main authors who publish on non-parametric control charts. It is possible to see that Chakraborty is the author nucleating the "non-parametric" theme. The current article relies on the studies by Mukherjee & Chakraborti (2012) in order to develop a framework for the use of non-parametric control charts.

The next section of the current article presents a literature review on statistical process control and on the non-parametric Shewhart-Lepage (SL) chart of of Mukherjee & Chakraborti (2012). Section 3 presents the search procedure. The fourth section presents an empirical model to estimate the SL control limits and illustrates the application of the non-parametric control chart and discusses the model validation. The following sections compare the performance of the SL control chart, in comparison to the Shewhart charts, and examines the best combinations of m and n through the response surface technique.

2 Theoretical foundation

2.1 Basic concepts of SPC

According to Montgomery (2004), statistical quality control is a set of statistical techniques used to measure, monitor, control, and improve quality. The SPC is one of the classical statistical quality control techniques and it assumes that there is a

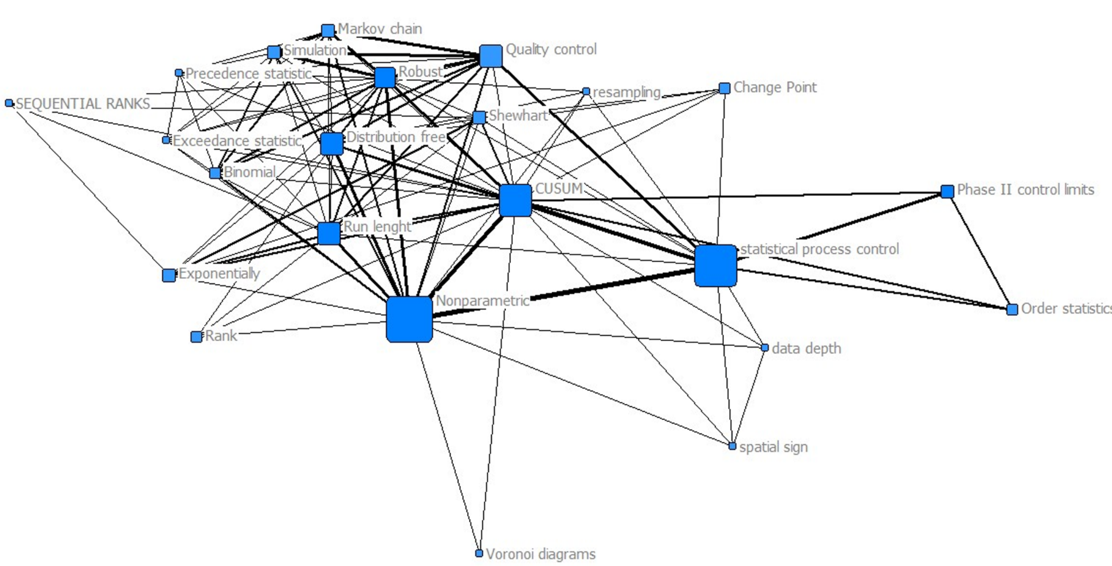

Figure 3. Co-occurrence of keywords on non-parametric statistical control. Source: Research Data.

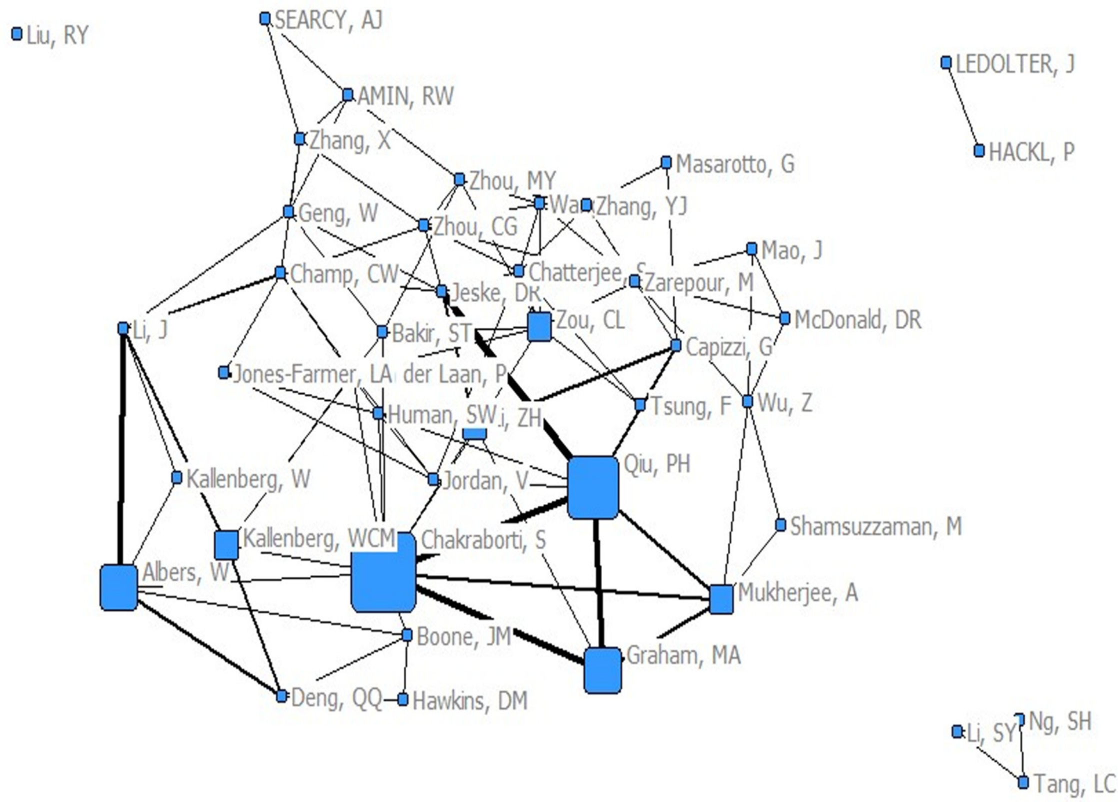

Figure 4. Networks of researchers who publish on non-parametric control charts. Source: Research Data.

process-inherent variation called natural variation, which is usually caused by many variables that individually produce small effects and are difficult to be detected and eliminated. On the other hand, there are special causes that produce great effects, and they are fewer and easier to be detected (Woodall, 2000; Michel & Fogliatto, 2002; Montgomery & Runger, 2003). The distinction between common and special cause is context-dependent - a common cause today may be a common cause tomorrow – and it could affect the sampling process (Woodall, 2000). From a practical perspective, one must act on the cause when it has enough economic impact on the quality (Woodall, 1985, 2000)

A process is considered to be in steady state or under control when only natural variations act on it. On the other hand, when the process is under, in addition to natural variations, the presence of special or assignable causes, it is out of control. The implementation of control charts is done in two phases: phase I, in which the statistical parameters are estimated and the control limits are established; and phase II, in which one monitors the process. In phase II, samples are collected from the process, some plotting statistic is calculated and their values are compared to the control limits set in phase I (Montgomery, 2004).

The control chart performance is generally assessed by different metrics depending on the phase. As it was previously mentioned, the ARL is the metric used to assess the control chart performance in phase II. The ARL value is given by $ARL_0 = \frac{1}{\alpha}$ in a process under control and by $ARL = \frac{1}{1-\beta}$ in a process out of control; wherein α and β are the type I and II errors, respectively (Montgomery, 2004).

2.2 Non-parametric control chart with simultaneous monitoring of location and scale

Having the study by Mukherjee & Chakraborti (2012) as reference and based on the classic WRS (Wilcoxon Rank-Sum) non-parametric test (see, Gibbons & Chakraborti, 2011), which is a defined by a test statistic for the location, T_1, for a sample size m in phase I and size n in phase II, given by Equation 1.

$$T_1 = \sum_{k=1}^{N} kZ_k \quad (1)$$

Wherein Z_k = 1 when the N (where $N = m + n$) data derive from independent samples in phase II; and $Z_k = 0$ when the data derive from independent samples in phase I.

The non-parametric statistical test used to measure the scale is the AB - Freund-Ansari-Bradley-David-Barton, T_2, (see Gibbons & Chakraborti, 2011) and calculated by Equation 2

$$T_2 = \sum_{k=1}^{N} \left| k - \frac{1}{2}(N+1) \right| Z_k \quad (2)$$

A process is said to be under control when $F(x)$ - the probability distribution of phase I - and G(y) - the probability distribution of phase II - are the same (F = G) for the location and scale parameters. Otherwise, the process is said to be out of control. Based on these statistical tests (T1 and T2), Mukherjee & Chakraborti (2012) determined the H, H_1 and H_2. control limits for some value combinations for m and n.

The mathematical expectation and the variance of T_1 statistics for a process under control are obtained by Equations 3 and 4:

$$E(T_1|IC) = \frac{1}{2}n(N+1) \quad (3)$$

$$V(T_1|IC) = \frac{1}{12}mn(N+1) \quad (4)$$

As for T_2 statistics, the mathematical expectation and the variance are give by Equations 5, 6, 7 and 8:

$$E(T_2|IC) = \mu_2 = \frac{nN}{4} \quad \textit{for even } N \quad (5)$$

$$\text{or} \quad E(T_2|IC) = \mu_2 = \frac{n(N^2-1)}{4N} \quad \textit{for odd } N \quad (6)$$

$$V(T_2|IC) = \sigma_2^2 = \frac{1}{48}mn\frac{(N^2-4)}{N-1} \quad \textit{for even } N \quad (7)$$

$$V(T_2|IC) = \sigma_2^2 = \frac{1}{48}\frac{mn(N+1)(N^2-3)}{N^2} \quad \textit{for odd } N \quad (8)$$

IC (*In Control*) indicates that the process is under control.

By using the Shewhart-Lepage control chart, Mukherjee & Chakraborti (2012) proposed an eight-step procedure to build a non-parametric control chart. This procedure uses the standardized statistics of the WRS and AB (Equations 9, 10 and 11) tests as well as the S_i^2 statistics (Equation 12):

$$S_{1i} = \frac{T_{1i} - \frac{1}{2}n(N+1)}{\sqrt{\frac{1}{12}mn(N+1)}} \quad (9)$$

$$S_{2i} = \frac{T_{2i} - \frac{nN}{4}}{\sqrt{\frac{1}{48}mn\frac{(N^2-4)}{N-1}}} \quad \textit{when N is even} \quad (10)$$

$$S_{2i} = \frac{T_{2i} - \frac{n(N^2-1)}{4N}}{\sqrt{\frac{1}{48}\frac{mn(N+1)(N^2-3)}{N^2}}} \quad \textit{when N is odd} \quad (11)$$

$$S_i^2 = S_{1i}^2 + S_{2i}^2 \quad (12)$$

The S_i^2 statistics is plotted and compared to the control limit H. If it is below the control limit, the process is considered to be under control; if it is above the control limit, the process is considered to be out of control, and the $S1i$ and $S2i$ statistics are compared to the $H1$ location and $H2$ scale limits, respectively. If both statistics are above the control limits, the process is considered to be out of control for both location and scale. If it is above one of the limits ($H1$ or $H2$), the process is considered to be out of control for location ($S1i2 > H1$) or for scale ($S1i2 > H2$) .

mits were determined by Mukherjee & Chakraborti (2012) for $ARL0$=500 and with different values (m, n) through computer simulation methods. Table 1 shows the limits found by the authors for some (m, n) value combinations.

The $H = H1 + H2$ relationship is a feature of these limits. Another feature is that $P(S2i > H|IC) = \alpha\ 0.0027$, which is partitioned into three exclusive events for a process under control, namely: A- Probability of the $S1i2 > H1$ location and of the $S1i2 > H2$ scale; B -Probability of the $S1i2 > H1$ location and of the $S1i2 > H2$ scale; C - Probability of the $S2i2 > H2$ location and of the $S_{2i}^2 > H_2$ scale. Thus, the probability of a false alarm α follows the following relationship between these events: $\gamma_1 + \gamma_2 - \gamma_1\gamma_2 = \alpha$, wherein γ_1 is the probability of a false positive for location, γ_2 is the probability of a false positive for scale, and $\gamma_1\gamma_2$ is the probability of a false positive for location and scale, simultaneously.

Table 1. Location and scale control combination (m, n) and limits.

m	n	H	H1	H2
30	5	9.4	5.75	3.65
30	11	9.24	5	4.24
30	25	8.4	4.3	4.1
50	5	10.32	6.52	3.8
50	11	10.1	6.1	4
50	25	9.5	5	4.5
100	5	11.25	7.25	4
100	11	11.07	6.35	4.72
100	25	10.74	5.4	5.34
150	5	11.5	7.65	3.85
150	11	11.45	6.8	4.65
150	25	11.17	5.61	5.56

Source: Mukherjee & Chakraborti (2012).

3 Research procedure

The research procedure shown in Figure 5 was followed in order to develop the application of non-parametric control charts. Stage 1 begins after the definition of the product or process features and quality parameters; this stage defines the test statistics used to measure location and scale, in the current case, the WRS and AB tests presented in Section 2.

Stage 2 proposes and analyzes a multiple regression model type $y = \beta_0 + \beta_1 m + \beta_2 n + \beta_{11} m^2 + \beta_{22} n^2 + \beta_{12} mn$ and stage 3 estimates the H, H_1 and H_2 control limits. Stage 4 estimates the statistical control limits for different (m, n) values in order to enlarge the set of sample combination options in phases I and II, when the proposed non-parametric control chart is deployed. Stage 5 validates the proposed empirical model that estimates the control limits through residuals analysis; evaluates the control chart performance by ARL, determined by simulation methods; and compares the performance of this chart to Shewhart charts with normal and exponential distributions in order to identify possible advantages over other control chart types.

The best combination of samples from phases I and II were obtained in stage 6 by using response surface techniques. The goal was to adjust the parameters (m, n) that reflect the best ARL values in terms of m and n. Simulation methods (Maple Software) are also used in this stage to obtain the ARL value with different m and n values. Stage 7 estimated the type I and II errors (α, β) and the ARL around the optimum solution obtained in stage 6. Stage 8 analyzed and compared the chart performance in terms of ARL, m and n, in order to find a solution that combines good statistical properties and lower sampling cost (m, n). Finally, stage 9 defined the sample sizes in phase I (m) and in phase II (n) for the non-parametric control chart.

4 Estimates of the control limits: H, H1 and H2

4.1 Control limits estimates

By fitting a multiple linear regression model (Equation 13) to the data in Table 1 using the least squares method, it is possible to establish a relationship between the (m, n) parameters and the $H = H_1 + H_2$ control limits. The following model was tested in the present study:

$$y = \beta_0 + \beta_1 m + \beta_2 n + \beta_{11} m^2 + \beta_{22} n^2 + \beta_{12} mn + \varepsilon \qquad (13)$$

The control limit H has statistically significant relationship only to β_1, β_{11} and β_{12}, according to the results in Table 2. However, it is possible to see that H is significantly dependent on the sample size in phase I. The residuals analysis and the $R2$ value are described in section five and indicate the suitability of the proposed model to the data in Table 1.

A similar approach showed that $\beta 1$, $\beta 11$, and $\beta 2$ were statistically significant for $H1$. Regarding this limit, m is significant in its two parameters - simple linear and linear quadratic - and n is significant in the simple linear term. The results are shown in Table 3.

As for $H2$, the parameters of m were not identified as statistically significant (according to Table 4).

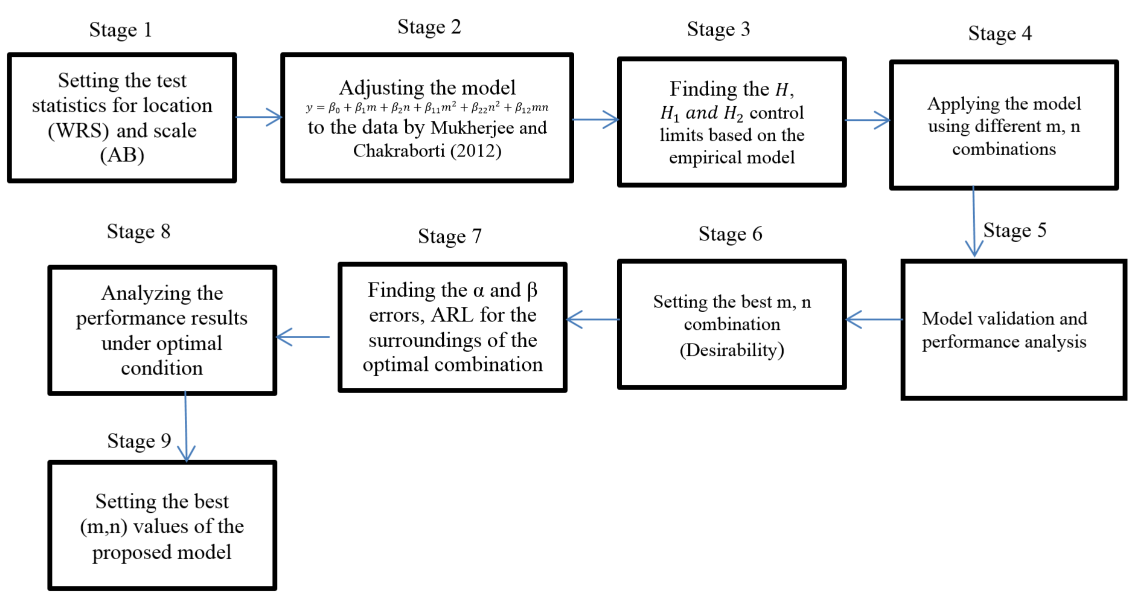

Figure 5. Research procedure. Source: Research Data.

The parameters of n and those of the interaction between m and n were identified as significant.

The estimates of the second-order regression parameters are provided in Table 2 in order to determine the H limit. Table 2 also shows the 95% confidence interval for these parameters. The $H1$ and $H2$ control limit estimates were obtained by using the same procedure, and the results are presented in Tables 3 and 4.

In the case of the H, H_1 and H_2 estimates, the following regression models were found:

$$\hat{H} = 8.332 + 0.0500m - 0.000195m^2 - 0.0399n - 0.000560n^2 + 0.000284mn \quad (14)$$

$$\hat{H}_1 = 5.4997 + 0.03833m - 0.000125m^2 - 0.1423n + 0.002565n^2 - 0.000247mn \quad (15)$$

$$\hat{H}_2 = 2.8325 + 0.01170m - 0.00007m^2 + 0.1024n - 0.003125n^2 + 0.000531mn \quad (16)$$

The results of the estimates of the H, $H1$ and $H2$ control limits, according to the proposed model, are shown in Table 5. It appears that the proposed method of estimating the control limits can be quite satisfactory in practice.

4.2 Illustrating the use of non-parametric control chart

A real case comprised a sample of 125 rubber products used in automotive components manufactured by hot forming process. The thickness of the pieces was measured, whose specification rate is 1.17 to 1.37 mm with tolerance of ± 0.10 mm regarding the nominal value of 1.26. The goal is to apply the mathematical models obtained from the results found by Mukherjee & Chakraborti (2012) (Equations 14, 15 and 16) and to determine the sample size in phase I (m) in order to build a non-parametric control chart to simultaneously monitor the location and scale measures.

The literature (Mukherjee & Chakraborti, 2012) and the results in Tables 2, 3 and 4 indicate that m is the most important parameter used to estimate the control limits in phase I, and n is important in phase II. Thus, four statistical process control strategies - for the use non-parametric control charts - were tested for the following m values (5, 14, 25 and 50), setting $n = 5$. The H, H_1 and H_2 control limits were estimated from the proposed regression model. Next, these m and n combinations for the four strategies will be analyzed.

a) Combinations (***m*** = 5, ***n*** = 5) **and** (***m*** = 14, ***n*** = 5)

A five-size sample (m = 5) was taken in phase I. Subsequently, fourteen five-size samples (n = 5) were taken in phase II. The eight-step procedure by Mukherjee & Chakraborti (2012) was applied. The control limits were calculated from the proposed mathematical model.

The results are shown in Figure 6. The dashed line refers to the control limit H estimated by the mathematical model. The results of the S_i^2 statistics obtained for each of the fifteen samples in phase II were plotted in the charts of Figure 6a. The S_i^2 statistics

Table 2. Estimate of the regression model parameters for H of the (m,n) combination.

	β_{ij}	STD Error	t(6)	p	-95.%	+95.%
mean	8.332189	0.251530	33.12608	0.000000	7.716718	8.947660
m	0.050031	0.004590	10.89993	0.000035	0.038800	0.061263
m^2	-0.000195	0.000024	-8.06600	0.000194	-0.000254	-0.000136
n	-0.039910	0.029886	-1.33539	0.230175	-0.113039	0.033219
n^2	-0.000560	0.000921	-0.60742	0.565836	-0.002813	0.001694
m x n	0.000284	0.000091	3.12055	0.020571	0.000061	0.000507

Source: Results obtained using Statistica 11 software (StatSoft, 2013).

Table 3. Estimate of the regression model parameters for H_1 of the (m, n) combination.

	β_{ij}	STD Error	t(6)	p	-95.%	+95.%
mean	5.499655	0.468467	11.73968	0.000023	4.353357	6.645953
m	0.038328	0.008549	4.48344	0.004177	0.017410	0.059247
m^2	-0.000125	0.000045	-2.77765	0.032097	-0.000235	-0.000015
n	-0.142341	0.055662	-2.55722	0.043070	-0.278543	-0.006140
n^2	0.002565	0.001716	1.49537	0.185447	-0.001632	0.006763
m x n	-0.000247	0.000170	-1.45590	0.195669	-0.000662	0.000168

Source: Research Data.

Table 4. Estimate of the regression model parameters for H_2 of the (m, n) combination.

	β_{ij}	STD Error	t(6)	p	-95.%	+95.%
mean	2.832534	0.312140	9.07455	0.000100	2.068754	3.596314
m	0.011703	0.005696	2.05455	0.085700	-0.002235	0.025641
m²	-0.000070	0.000030	-2.33101	0.058560	-0.000143	0.000003
n	0.102431	0.037088	2.76185	0.032774	0.011680	0.193182
n²	-0.003125	0.001143	-2.73375	0.034016	-0.005922	-0.000328
m x n	0.000531	0.000113	4.69966	0.003327	0.000255	0.000808

Source: Research Data.

Table 5. Comparing the estimated results and the exact values.

		Obtained from Table 1			Estimates			Error		
m	n	H	H1	H2	H	H1	H2	H	H1	H2
50	5	10.32	6.52	3.80	10.205	6.395	3.810	-0.115	-0.125	0.01
50	11	10.10	6.10	4.00	9.997	5.713	4.284	-0.103	-0.387	0.284
50	25	9.50	5.00	4.50	9.355	4.840	4.515	-0.145	-0.16	0.015
100	5	11.25	7.25	4.00	11.319	7.314	4.005	0.069	0.064	0.005
100	11	11.07	6.35	4.72	11.196	6.558	4.638	0.126	0.208	-0.082
100	25	10.74	5.40	5.34	10.753	5.512	5.241	0.013	0.112	-0.099
150	5	11.50	7.65	3.85	11.459	7.609	3.851	-0.041	-0.041	0.001
150	11	11.45	6.80	4.65	11.422	6.779	4.643	-0.028	-0.021	-0.007
150	25	11.17	5.61	5.56	11.178	5.560	5.618	0.008	-0.05	0.058
30	5	9.40	5.75	3.65	9.487	5.853	3.635	0.087	0.103	-0.015
30	11	9.24	5.00	4.24	9.245	5.200	4.045	0.005	0.2	-0.195
30	25	8.40	4.30	4.10	8.524	4.397	4.127	0.124	0.097	0.027

Source: Research Data.

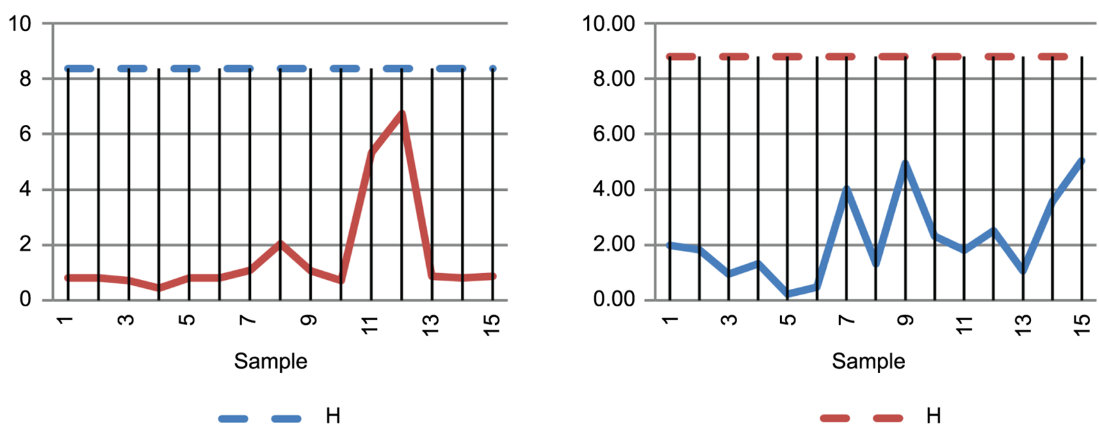

Figure 6. a) S_i^2 statistics obtained by the (m = 5, n = 5) combination; b) S_i^2 statistics obtained by the (m = 14, n = 5) combination. Source: Research Data.

shown in Figure 6b was obtained by increasing the sample size of phase I to m = 14. From a theoretical perspective, the size of the largest sample in phase I improves the detection capability in phase II.

b) Combinations (**m = 25, n = 5) and (m = 50, n = 5)**

The S_i^2 statistics results of m = 25 are presented in Figure 7a and those of m = 50 are shown in Figure 7b. The last configuration detects one point out of control, which could indicate better capability to detect an unstable process, i.e., the capability to detect special causes in the control chart when m increases. It would meet the theory, which, by mathematical means, shows the effects of increasing the number of samples in phase I on the control charts' performance in phase II.

The analysis of the data frequency distribution in phases I and II of the (m = 50, n = 5) combination was performed, as shown in Figure 8. Figure 8a refers to the data distribution obtained in phase I and Figure 8b shows the data obtained in phase II. It is possible to see in phase II that the data are distributed in a more dispersed and less symmetrical way in comparison to the data of phase I. This behavior shows that this process is not under control, as shown by the control chart of Figure 7b.

The same sampled data (125) shown in Appendix A were used to build the Shewhart charts for mean and range; 25 samples of size n = 5 were extracted. These charts, shown in Figure 9, correspond to the phase I of the classical procedure used to build control charts. There was increased dispersion, which may

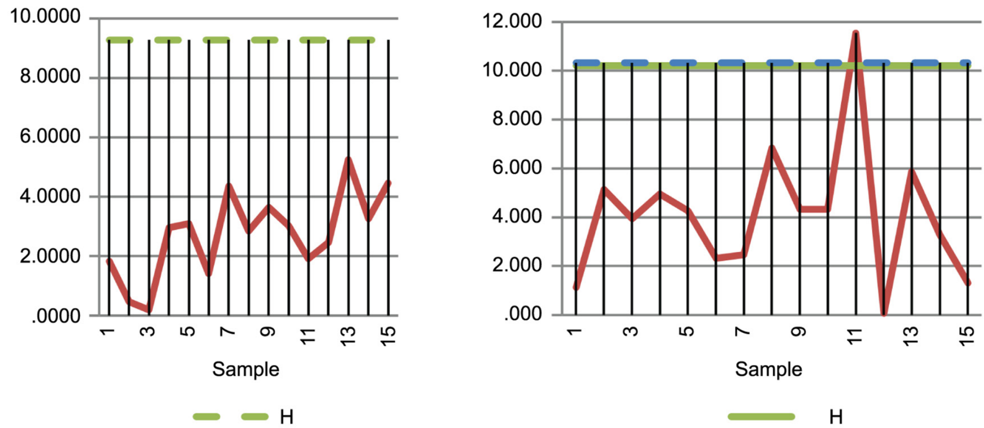

Figure 7. a) S_i^2 statistics obtained by the (m = 25, n = 5) combination; b) S_i^2 statistics obtained by the (m = 50, n = 5) combination. Source: Research Data.

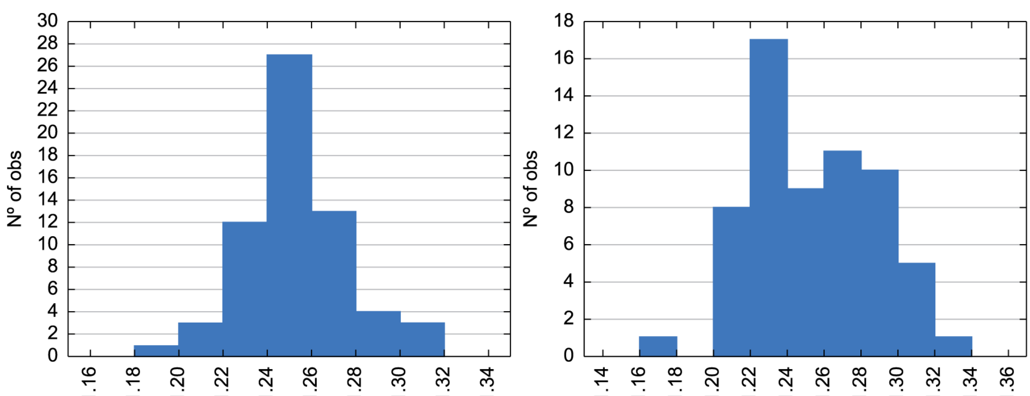

Figure 8. (a) Histogram of the sample of phase I; (b) Histogram of the samples of phase II. Source: Research Data.

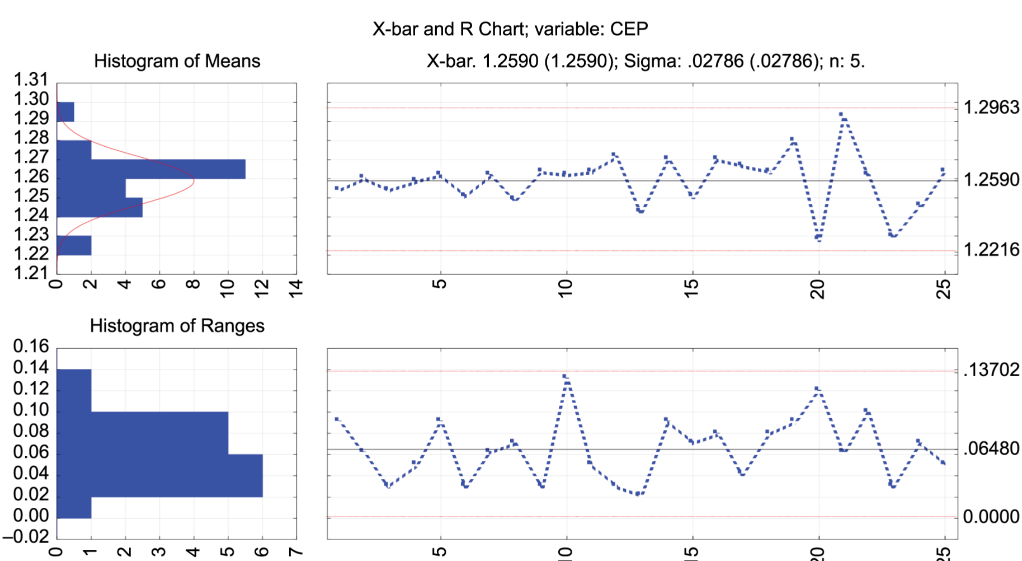

Figure 9. Shewhart-type control chart considering 25 samples of size 5. Source: Research Data.

be seen in the middle of the chart from sample 13. However, no point was detected outside the control limits, differently from what was observed in the chart of Figure 7b.

4.3 Statistical validation of the model and optimal combination of m, n

According to Gibbons & Chakraborti (2011), regarding large samples subjected to certain conditions, the statistic $[T_N - E(T_N)/\sigma(T_N)$ has an approximate standard normal probability distribution (it is the standardization used to calculate S_1 and S_2). Figures 10 and 11 show the probability distribution of these statistics, which have roughly symmetrical distributions (Gibbons & Chakraborti, 2011). Figure 12 shows the residuals analysis of the model that estimates *H*. Figure 12 shows that the residuals are stable and follow the normal probability distribution; this result is required to validate the estimation model of the proposed control limit.

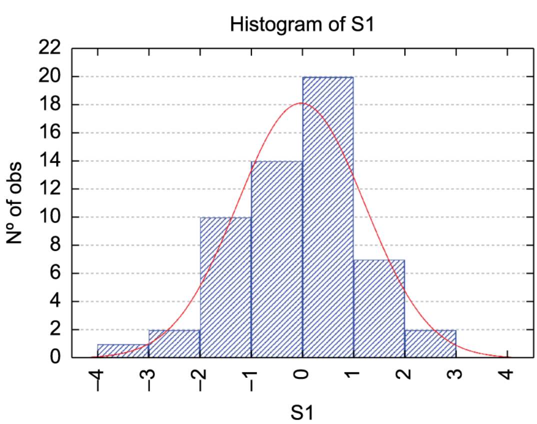

Figure 10. Probability distribution of S_1 statistics. Source: Research Data.

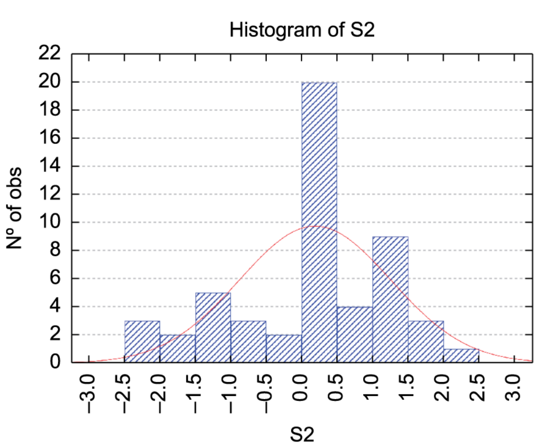

Figure 11. Probability distribution of S_2 statistics. Source: Research Data.

One of the important aspects of the response surface analysis technique consists of finding the optimal value of *m,n* to find the best *H* estimate. The herein found value was $m = 82$ and $n = 12$. These values are shown in Figure 13. Results similar to $m = 82$ and $n = 12$ were obtained for H1 and H2. These results are shown in Figures 14 and 15.

The residual analyses for H1 - shown in Figure 16 - and for H2 - shown in Figure 17 - indicate slight deviation in the residuals normality, especially for H2. Unlike H, in which the residuals showed symmetrical behavior in the normal probability distribution, the limit estimation methods for location (H1) and scale (H2) should be carefully analyzed to assess the impact of these deviations on the control chart performance. It is worth emphasizing that the estimated H limit helps monitor the location and scale parameters in process monitoring. In addition, H1 and H2 are objects of analysis for the effects on the central position or dispersion measurements. Then, studying the performance is important to check the performance level obtained in this type of chart, and it is presented in the next section.

5 Analysis of the non-parametric control chart performance obtained by the proposed model and compared to Shewhart charts

The performance analysis of different types of control charts is traditionally based on the ARL parameter. Table 6 and Figure 18 show the ARL results for different combinations of (*m,n*) values.

Computer simulation – in MAPLE – was used to estimate the ARL values for $\tau = 0.01$ to 0.07. Fifty thousand (50,000) control chart simulations were performed for the combinations shown in Table 6. The results show that the ARL decreases as *m* increases. For example, for $\tau = 0.01$ the (m = 14, n = 5) combination needs, on average, 50.84 samples to detect one point out of control; whereas 32.26 samples are necessary for the (m = 30, n = 5) combination, which means a better performance for a sample size $m = 30$ for phase I in comparison to $m = 14$.

The results shown in Table 6 and in Figure 18 indicate that, regarding an out-of-control process, the non-parametric control chart performance improves as the sample size (*m*) increases. The performance of the non-parametric control charts is worse than that of the Shewhart-type charts with normal distribution, which means higher α and β errors. However, they perform better than the Shewhart-type control charts with exponential distribution. Therefore, the non-parametric control chart performs better than the classic control chart when the data probability distribution is unknown or when there is no normal probability distribution.

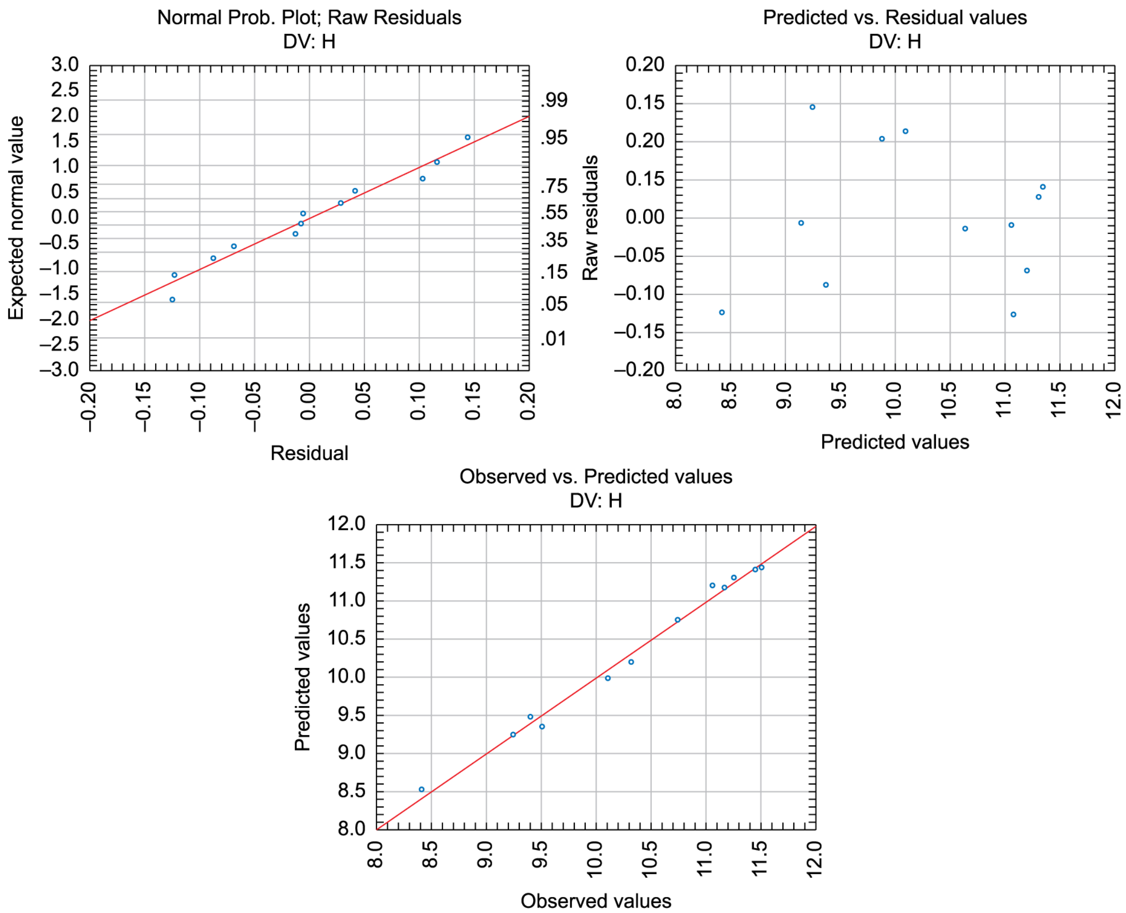

Figure 12. Analysis of H residuals. Source: Research Data.

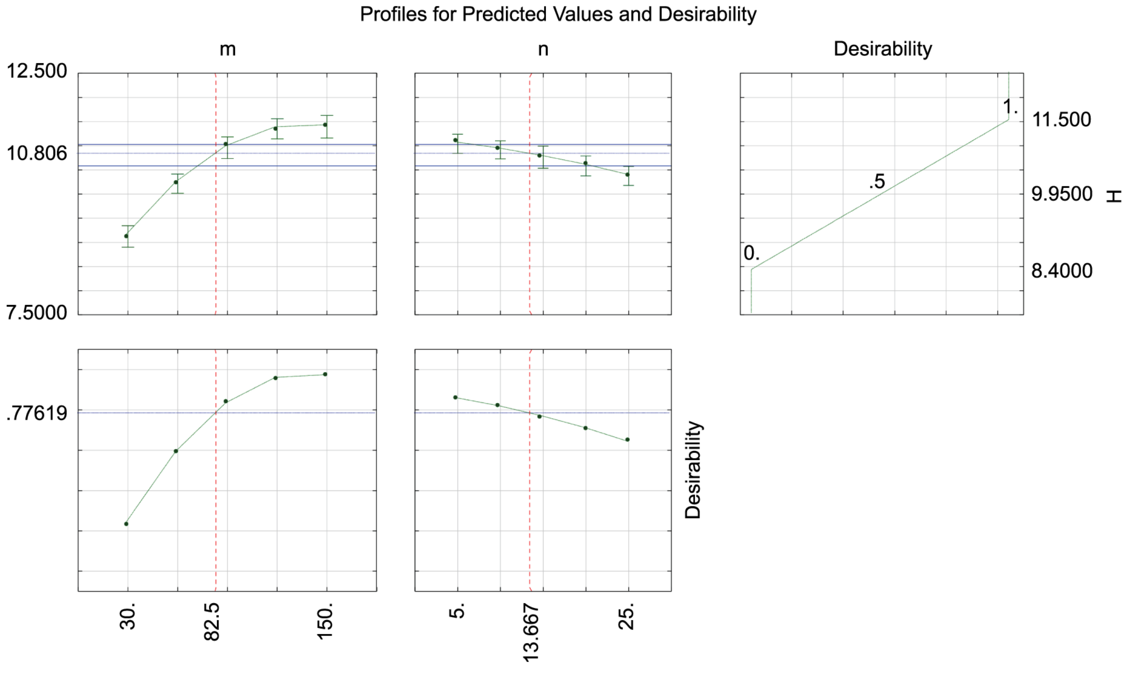

Figure 13. Optimal values of *m*, *n* for H.

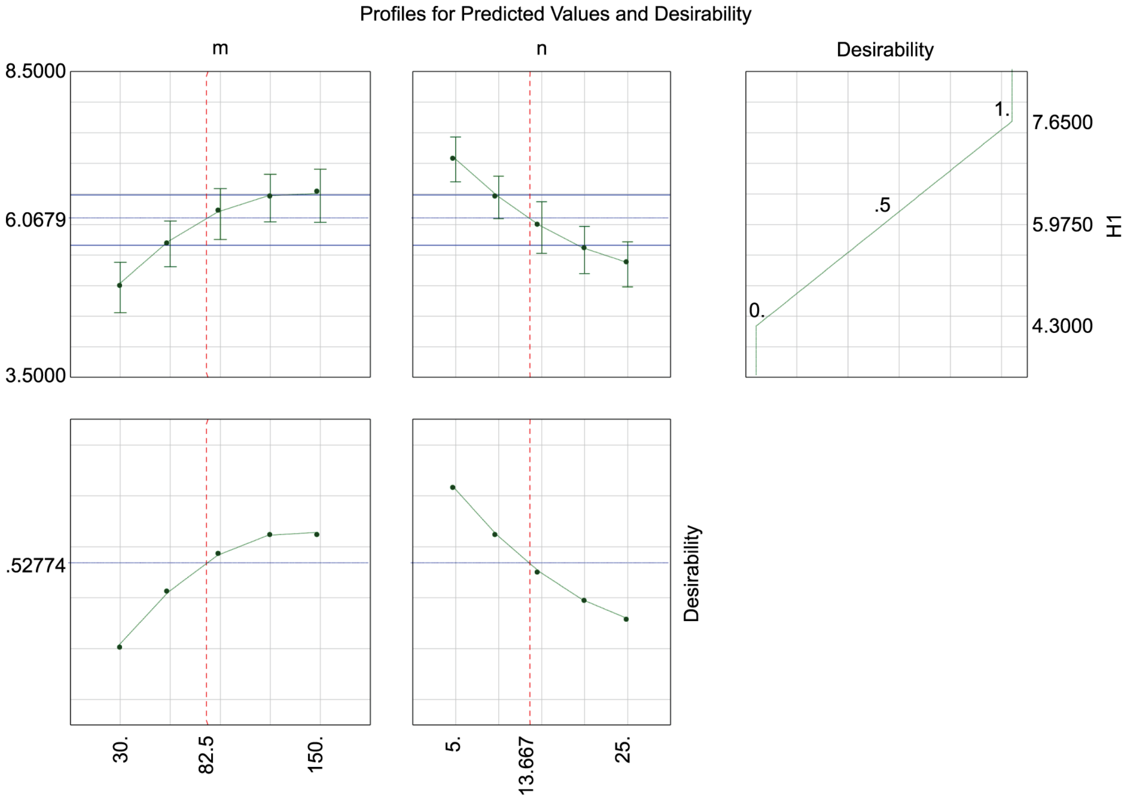

Figure 14. Optimal values of *m*, *n* for H1. Source: Research Data.

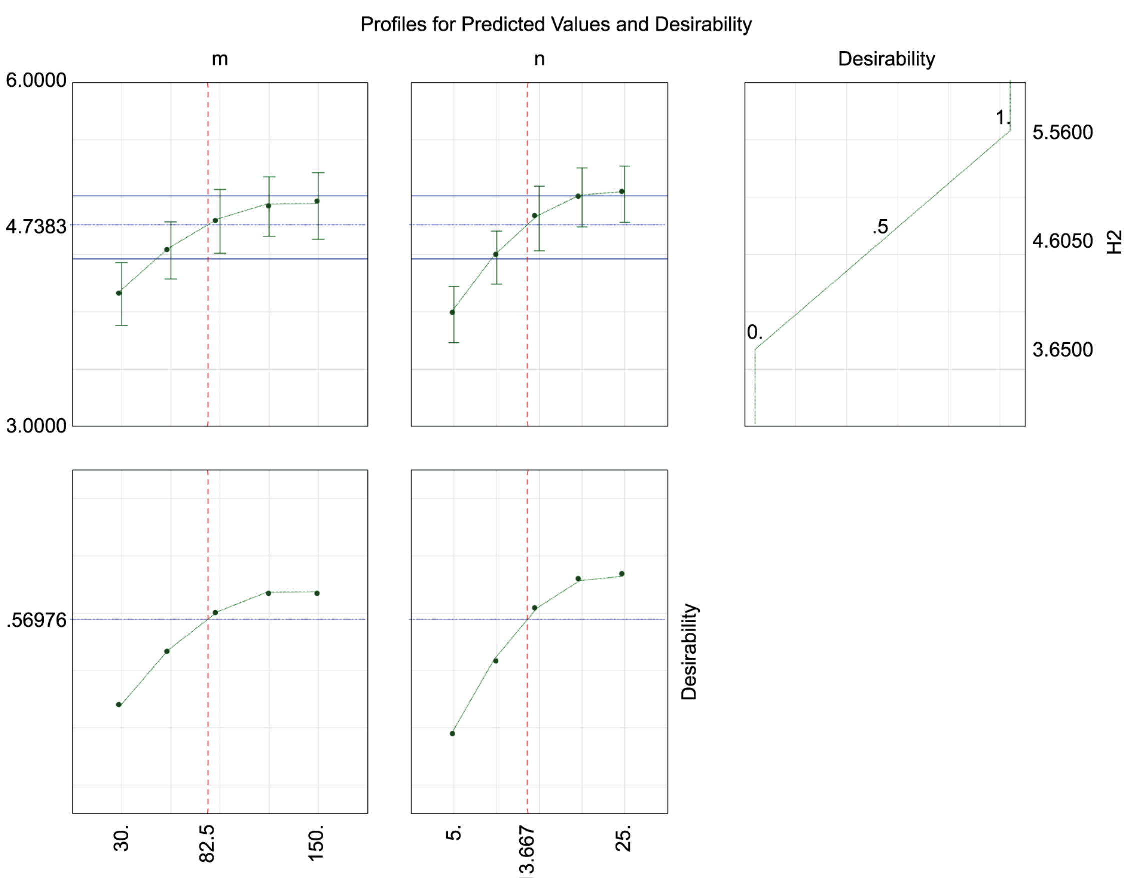

Figure 15. Optimal values of **m*, *n* for H2. Source: Research Data.

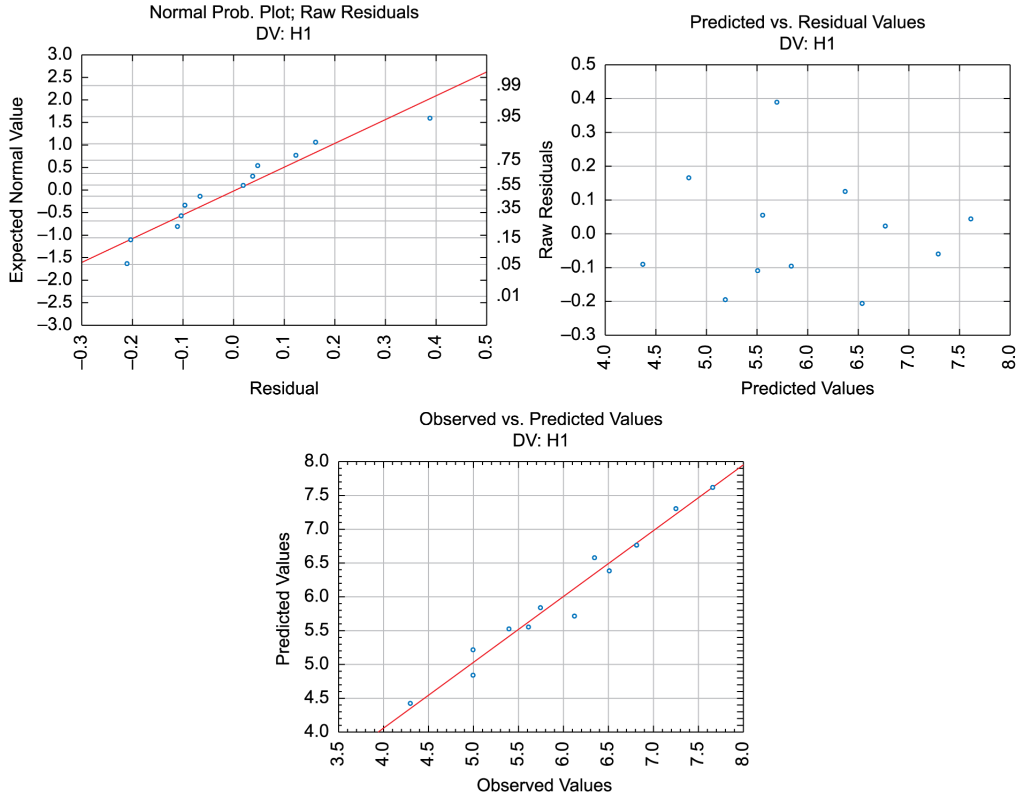

Figure 16. Analysis of H1 residuals. Source: Research Data.

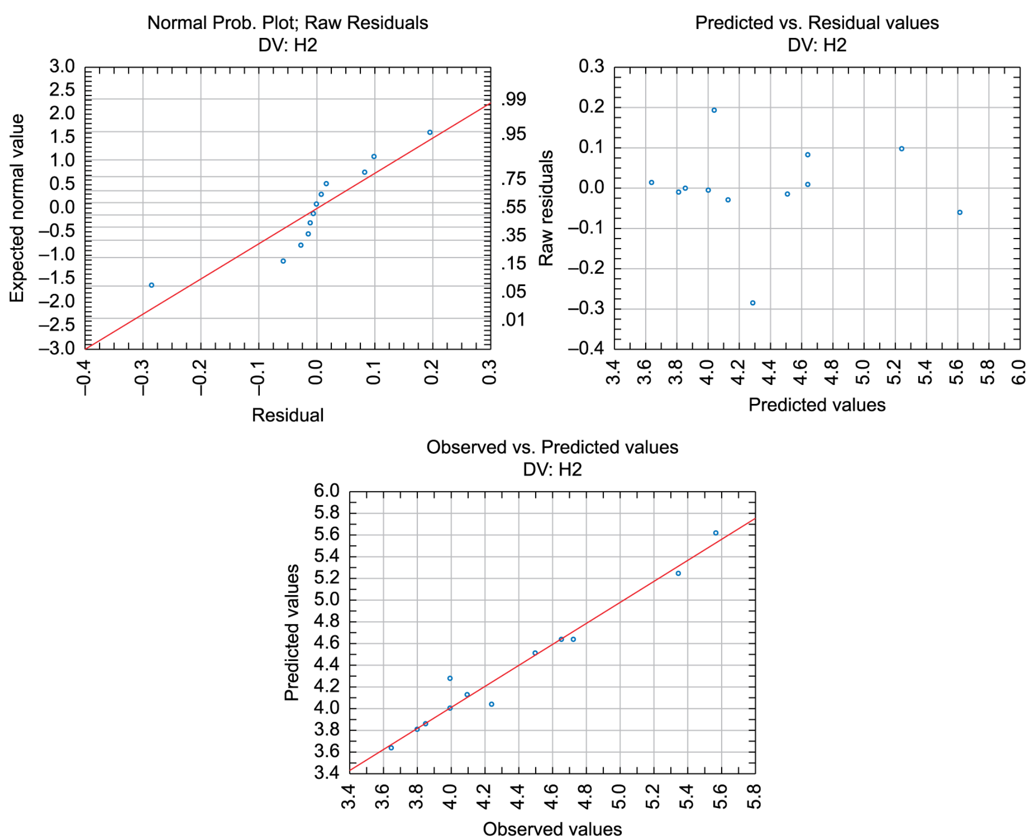

Figure 17. Analysis of H2 residuals. Source: Research Data

6 Analyzing the best conditions of (m, n) variables

This section evaluates the type I and II errors (α and β) and the ARL performance of the non-parametric control chart, whose control limits were obtained by Equations 14, 15 and 16. The results of the analysis are shown in Table 7, whose values were obtained by simulating the industrial process analyzed in the previous sections. Ten thousand (10,000) cycles were performed for each combination shown in Table 7 - $\tau = 0$, when the process was under control, and $\tau = 0.01$ ..0.07 when the process was out of control.

As for the condition of the ($m = 80$, $n = 10$) variables obtained in the previous section, a good control chart performance was obtained when $\tau \geq 0.03$; for example, for τ = 0.03, the *ARL* = 1.39; however, when the

Table 6. Performance between the non-parametric control chart and the Shewhart-type control chart (T-S).

Combination	ARL_0	ARL						
		Proposed non-parametric chart						
(m,n)	$\tau = 0$	0.01	0.02	0.03	0.04	0.05	0.06	0.07
(10.5)	175.41	68.97	21.28	6.62	3.26	1.92	1.32	1.14
(14.5)	161.80	50.84	12.58	4.28	2.16	1.37	1.11	1.03
(20.5)	145.25	46.51	11.30	3.77	1.83	1.25	1.07	1.01
(30.5)	209.02	32.26	10.00	3.13	1.61	1.16	1.03	1.00
		T-S chart (Normal)						
(10.5)	250.67	41.20	7.33	2.48	1.36	1.06	1.01	1.00
(14.5)	363.83	54.40	7.82	2.53	1.35	1.06	1.01	1.00
(20.5)	387.18	53.73	8.52	2.40	1.32	1.07	1.01	1.00
(30.5)	466.07	58.36	8.37	2.32	1.34	1.06	1.01	1.00
		T-S chart (Exponential)						
(10.5)	77.52	73.53	69.93	62.23	62.50	58.64	68.49	51.81
(14.5)	101.01	92.59	75.75	74.07	86.58	76.63	65.36	61.54
(20.5)	129.87	92.59	105.26	89.29	81.97	72.46	76.92	89.69
(30.5)	117.00	97.09	95.24	89.69	84.75	83.33	76.34	80.00

Source: Research Data.

Table 7. Type (I, II) errors and ARL for the non-parametric control chart whose limits were obtained from mathematical models.

Combination	α error/ ARL_0	β error / ARL								
		Synthetic Chart								
(m,n)	τ = 0	0.01	0.02	0.03	0.04	0.045	0.048	0.05	0.06	0.07
(40. 5)	0.005	0.99	0.88	0.61	0.33	0.32	0.16	0.18	0.03	0.00
	200.000	100.00	8.33	2.59	1.48	1.46	1.19	1.22	1.03	1.00
(50. 5)	0.007	0.99	0.87	0.62	0.39	0.21	0.13	0.09	0.05	0.00
	140.845	100.00	7.58	2.60	1.65	1.27	1.15	1.10	1.05	1.00
(50. 10)	0.003	0.96	0.74	0.31	0.05	0.03	0.02	0.00	0.00	0.00
	322.581	24.39	3.82	1.45	1.05	1.03	1.02	1.00	1.00	1.00
(50. 20)	0.003	0.88	0.56	0.04	0.00	0.00	0.00	0.00	0.00	0.00
	333.333	8.26	2.25	1.04	1.00	1.00	1.00	1.00	1.00	1.00
(60. 5)	0.004	0.98	0.93	0.68	0.35	0.25	0.19	0.13	0.02	0.00
	250.000	50.00	14.29	3.08	1.55	1.33	1.24	1.15	1.02	1.00
(60. 10)	0.002	0.97	0.73	0.34	0.03	0.01	0.00	0.00	0.00	0.00
	500.000	33.33	3.68	1.52	1.03	1.01	1.00	1.00	1.00	1.00
(70. 5)	0.003	0.97	0.94	0.73	0.35	0.24	0.16	0.08	0.03	0.00
	333.333	33.33	16.13	3.65	1.54	1.32	1.19	1.09	1.03	1.00
(70. 10)	0.007	0.99	0.98	0.78	0.37	0.16	0.06	0.06	0.02	0.00
	142.857	100.00	50.00	4.46	1.60	1.19	1.07	1.06	1.02	1.00
(80. 5)	0.002	0.95	0.89	0.71	0.31	0.19	0.11	0.12	0.02	0.00
	500.000	19.61	9.01	3.41	1.44	1.24	1.12	1.14	1.02	1.00
(80. 10)	0.006	0.98	0.71	0.28	0.06	0.02	0.01	0.00	0.00	0.00
	166.667	50.00	3.44	1.39	1.06	1.02	1.01	1.00	1.00	1.00

Source: Research Data.

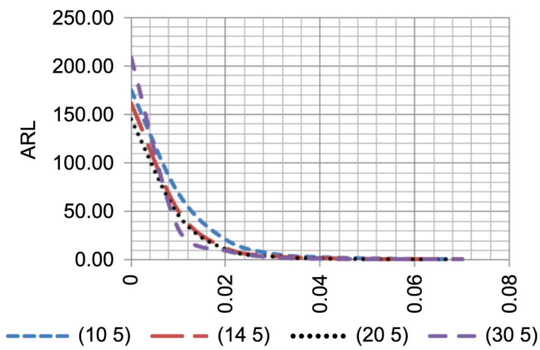

Figure 18. Assessing the performance of the non-parametric control chart. Source: Research Data.

(m = 50, n = 20) combination was used, the result was more interesting, ARL = 1.04. Table 7 shows different results obtained by the H, H_1, H_2 estimation method proposed in the current study. Therefore, the best combination found for these control limits was m = 50 in phase I, and n =20 in phase II.

By analyzing the results, it was observed that, the higher n is, the smaller the β error, and therefore, the better was the capability to detect a special cause. For example, for m =50 and n = (5,10,20), we found ARL = (2.60; 1.45; 1.04 when τ = 0.03. The α error (left side of Table 7), which was found through simulation, is α = 0.002 to 0.007, for ARL = 140.8 to 500.0.

7 Conclusion

The non-parametric control chart, with simultaneous monitoring of location and scale measures, is an alternative to the classical statistical control methods. The advantages of this type of chart are described as follows: it allows evaluating the control state of the variance and mean of a product or process feature using a single parameter; it is more robust because it performs better in terms of ARL than the Shewhart-type charts for asymmetric distributions.

Overall and especially for this type of control chart, phase I is very important to the phase II of the SPC implementation. Table 6 and Figure 18 show the better performance of the control chart in the phase II for relatively higher m values. Therefore, the largest sample size in phase I is, the better the control chart performance, measured by the ARL.

The proposed model estimates the control limits of the non-parametric synthetic charts. From a practical perspective, the multiple linear regression model, which was adjusted to the data in Table 1, allows estimating the control limits with (m, n) combinations different from those presented by Mukherjee & Chakraborti (2012).

By comparing the performance of the non-parametric synthetic control chart - with the estimated control limits - and the classical Shewhart control chart, it is possible to see the better performance of the latter. However, the results showed better performance of the non-parametric synthetic chart when the data distribution was asymmetric.

The results also show that the m parameter is more important in the non-parametric control chart performance, as shown in Table 7. The search methods applied for optimal solutions indicate m = 82 and n = 12; however, by simulating several combinations, it was possible to find a satisfactory control chart performance for m = 50 and n = 20, which was suggested for this type of chart.

The literature shows that there is a theoretical rule for the use of control charts, which consists in phases I and II, in which hypothesis tests are associated as an essential ingredient for the successful application of these charts. According to Woodall (2000), the form of the underlying distribution and the data autocorrelation degree have become an important component in the interpretation of control charts, in phase I, when the control limits are estimated, and in Phase II, when their performance is evaluated. Thus, studying the control chart performance is important as an insight of how control charts behave in practice.

Traditional control chart methods are still applicable to many industrial practical situations; however, it is worth considering new developments of control chart methods that suit the new environmental conditions of the manufacturing industry.

Acknowledgements

This work was supported by Foundation of Research of São Paulo State, who allowed the visit to Brazil of Professor S. Chakraborty, the University of Alabama, USA.

References

Baker, R. C., & Brobst, R. W. (1996). Conditional double sampling. *Journal of Quality Technology*, 10, 150-154.

Boone, J. M., & Chakraborti, S. (2011). Two Simple Shewhart-type multivariate nonparametrics control charts. *Applied Stochastic Models in Business and Industry*, 28(2), 130-140. http://dx.doi.org/10.1002/asmb.900.

Box, G. E. P., Hunter, W. G., & Hunter, J. S. (1978). *Statistics for experimenters*. New York: Wiley.

Castagliola, P., & Maravelakis, A. (2011). CUSUM Control Chart for monitoring the variance when parameters are estimated. *Journal of Statistical Planning and Inference*, 141(4), 1463-1478. http://dx.doi.org/10.1016/j.jspi.2010.10.013.

Castagliola, P., & Wu, S. (2012). Design of the c and np Charts when the Parameters are Estimated. *International Journal of Reliability Quality and Safety Engineering*, 19(2), 125-138. http://dx.doi.org/10.1142/S0218539312500106.

Castagliola, P., Celano, G., & Chen, G. (2009). The exact run length distribution and design of the S2 Chart when the In-Control Variance is estimated. *International Journal of Reliability Quality and Safety Engineering*, 16(1), 23-38. http://dx.doi.org/10.1142/S0218539309003277.

Chakraborti, S. (2006). Parameter estimation and design considerations in prospective applications chart. *Journal of Applied Statistics*, 33(4), 439-459. http://dx.doi.org/10.1080/02664760500163516.

Chakraborti, S., & Human, S. W. (2006). Parameter estimation and performance of the p-chart for attributes data. *IEEE Transactions on Reliability*, 55(3), 559-566. http://dx.doi.org/10.1109/TR.2006.879662.

Chen, G. (1997). The mean and standard deviation of the run length distribution of Charts when control limits are estimated. *Statistica Sinica*, 7, 789-798.

Costa, A. F. B., & Machado, M. A. G. (2011). A control chart based on sample ranges for monitoring the covariance matrix of the multivariate processes. *Journal of Applied Statistics*, 38(2), 233-245. http://dx.doi.org/10.1080/02664760903406413.

Costa, A. F. B., Machado, M. A. G., & Claro, F. A. (2010). Gráfico de controle MCMAX para o monitoramento simultâneo do vetor de médias e da matriz de covariâncias. *Gestão & Produção*, 17(1), 149-156. http://dx.doi.org/10.1590/S0104-530X2010000100012.

Costa, A. F. B., Magalhães, M. S., & Epprecht, E. K. (2009). Monitoring the process mean and variance using synthetic control chart with two-stage testing. *International Journal of Production Research*, 47(18), 5067-5086. http://dx.doi.org/10.1080/00207540802047098.

Duarte, B. P. M., & Saraiva, P. M. (2008). An optimization-based approach for designing attribute acceptance sampling plan. *International Journal of Quality & Reliability*, 25(2), 824-841. http://dx.doi.org/10.1108/02656710810898630.

Epprecht, E. K., Costa, A. F. B., & Mendes, F. C. T. (2005). Gráficos adaptativos de controle por atributos e seu projeto na prática. *Pesquisa Operacional*, 25(1), 113-134. http://dx.doi.org/10.1590/S0101-74382005000100007.

Gibbons, J. D., & Chakraborti, S. (2011). *Nonparametric statistical inference* (5 ed.). New York: Taylor & Francis. p. 630.

Graves, S. B., Murphy, D. C., & Ringuest, J. L. (1999). Acceptance sampling versus redundancy as alternative means to achieving goals for system reliability. *International Journal of Quality & Reliability Management*, 16(4), 362-370. http://dx.doi.org/10.1108/02656719910263724.

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

Jones, L. A., Champ, C. W., & Rigdon, S. E. (2001). The performance of exponentially weighted moving average charts with estimated parameters. *Technometrics*, 43(2), 156-167. http://dx.doi.org/10.1198/004017001750386279.

Lee, P. H. (2013). Joint statistical design of and S charts with combined double sampling and variable sampling interval. *European Journal of Operational Research*, 225(2), 285-297.

McCracken, A. K., & Chakraborti, S. (2013). Control Chart for joint monitoring of mean and variance: an overview. *Quality Technology & Quantitative Management*, 10(1), 17-37.

Michel, R., & Fogliatto, F. S. (2002). Projeto econômico de cartas adaptativas para monitoramento de processos. *Gestão & Produção*, 9(1), 17-31. http://dx.doi.org/10.1590/S0104-530X2002000100003.

Montgomery, C. D., & Runger, G. C. (2003). *Estatística aplicada e probabilidade para engenheiros* (2 ed.). São Paulo: LTC.

Montgomery, D. C. (1992). *Introduction to statistical quality control*. (2 ed.). New York: John Wiley & Sons. 674 p.

Montgomery, D. C. (2004). *Introdução ao controle estatístico da qualidade* (4 ed.). São Paulo: LTC.

Mukherjee, A., & Chakraborti, S. (2012). A distribution free CONTROL Chart for joint monitoring of location and Scale. *Quality and Reliability Engineering International*, 28(3), 335-352. http://dx.doi.org/10.1002/qre.1249.

Ozsan, G., Testik, M. C., & Weib, C. H. (2009). Properties of the exponential EWMA Chart with parameter estimation. *Quality and Reliability Engineering International*, 26(6), 555-569. http://dx.doi.org/10.1002/qre.1079.

StatSoft. (2013). *Statistica 11.0*. Tulsa: StatSoft Inc.

Thomson Reuters. (2013). *Web of Science*. Recuperado em 16 de dezembro de 2013, de http://wokinfo.com.

Trovato, E. A., Castagliola, P., Celano, G., & Fichera; S. (2010). Economic design of inspection strategies to monitor dispersion. *Computers & Industrial Engineering*, 59(4), 887-897. http://dx.doi.org/10.1016/j.cie.2010.08.019.

Woodall, W. H. (1985). The statistical design of quality control charts. *Journal of the Royal Statistical Society. Series D (The tatistician)*, 34(2), 155-160.

Woodall, W. H. (2000). Controversies and contradictions in statistical process control. *Journal of Quality Technology*, 32(4), 341-350.

Zhang, Y., & Castagliola, P. (2010). Run rules xbar charts when process parameters are unknown. *International Journal of Reliability Quality and Safety Engineering*, 17(4), 381-399. http://dx.doi.org/10.1142/S0218539310003858.

Zhang, Y., Castagliola, P., Wu, M. B., & Khoo, C. (2011). The synthetic xbar Chart with estimated parameters. *IIE Transactions*, 43(9), 676-687. http://dx.doi.org/10.1080/0740817X.2010.549547.

Appendix A. Sampled data in a real case: 25 samples of size n = 5.

Cavity	Value	Cavity	Value
1	1.31	**64**	1.23
2	1.26	**65**	1.25
3	1.22	**66**	1.28
4	1.26	**67**	1.31
5	1.22	**68**	1.30
6	1.25	**69**	1.24
7	1.24	**70**	1.22
8	1.30	**71**	1.28
9	1.25	**72**	1.23
10	1.26	**73**	1.23
11	1.24	**74**	1.29
12	1.25	**75**	1.22
13	1.25	**76**	1.24
14	1.27	**77**	1.32
15	1.26	**78**	1.27
16	1.26	**79**	1.28
17	1.28	**80**	1.24
18	1.24	**81**	1.24
19	1.28	**82**	1.27
20	1.23	**83**	1.27
21	1.27	**84**	1.27
22	1.23	**85**	1.28
23	1.32	**86**	1.28
24	1.24	**87**	1.22
25	1.25	**88**	1.30
26	1.23	**89**	1.30
27	1.25	**90**	1.22
28	1.26	**91**	1.32
29	1.26	**92**	1.30
30	1.25	**93**	1.23
31	1.30	**94**	1.25
32	1.28	**95**	1.30
33	1.25	**96**	1.25
34	1.24	**97**	1.22
35	1.24	**98**	1.17
36	1.26	**99**	1.29
37	1.28	**100**	1.21
38	1.25	**101**	1.32
39	1.24	**102**	1.29
40	1.21	**103**	1.26
41	1.26	**104**	1.31
42	1.25	**105**	1.28
43	1.28	**106**	1.26
44	1.26	**107**	1.33
45	1.27	**108**	1.23
46	1.28	**109**	1.24
47	1.26	**110**	1.25
48	1.19	**111**	1.23
49	1.32	**112**	1.22
50	1.26	**113**	1.25
51	1.27	**114**	1.23
52	1.24	**115**	1.22

Source: Research Data.

Appendix A. Continued...

Cavity	Value	Cavity	Value
53	1.25	**116**	1.30
54	1.29	**117**	1.24
55	1.27	**118**	1.23
56	1.28	**119**	1.23
57	1.29	**120**	1.23
58	1.26	**121**	1.29
59	1.27	**122**	1.26
60	1.26	**123**	1.24
61	1.23	**124**	1.28
62	1.25	**125**	1.25
63	1.25		

Source: Research Data.