Journal of Applied Mathematics & Bioinformatics, vol.8, no.2, 2018, 83-97
ISSN: 1792-6602 (print), 1792-6939 (online)
Scienpress Ltd, 2018

A simulation study of distribution-free Phase II control charts: advances and applications

Ioannis S. Triantafyllou[1] and Nikolaos I. Panayiotou[2]

Abstract

In this article, we consider adding runs-rules to enhance the performance of the nonparametric Phase II Shewhart-type chart based on the Wilcoxon–type rank–sum statistic proposed by Balakrishnan *et al*. (2009). The average run length performance and several design aspects of the proposed chart are studied through extensive simulations. A comparison between the proposed scheme and other competitive charts already introduced in the literature reveals the robustness of the new monitoring scheme.

Keywords: Average run length; Distribution-free control charts; Run rules; Nonparametric methods; Shewhart $\overline{X}$ – chart; Statistical process control.

---

[1] Department of Computer Science and Biomedical Informatics, University of Thessaly, Lamia, Greece.
[2] Department of Computer Science and Biomedical Informatics, University of Thessaly, Lamia, Greece.

Article Info: *Received* : January 7, 2018. *Revised* : March 2, 2018.
Published online : September 15, 2018.

1. Introduction

Statistical quality control charts were introduced in the early work of Shewhart (1926) and since then several variations of them have been proposed for monitoring continuous characteristics. Most of the control charts are distribution–based procedures in the sense that the process output is assumed to follow a specified probability distribution. However, this assumption is not always satisfied in practice and therefore the resulting control charts may not be reliable. To overcome this issue and yet keep the traditional structure of a monitoring scheme, several nonparametric (or distribution-free) charts have been suggested in the literature. For example, Bakir (2006) has constructed a Shewhart–type control chart using a signed–rank statistic, Chakraborti and Eryilmaz (2007) have considered an alternative class of charts based on the same statistic, while Chakraborti *et al.* (2004) introduced a family of nonparametric control charts based on a test sample quantile with limits calculated from an in–control (or a reference) sample. Moreover, Balakrishnan *et al.* (2010) constructed a distribution-free control chart based on the location of a single order statistic of the test sample (such as the median) as well as the number of observations in that test sample that lie between the control limits. In addition, Balakrishnan *et al.* (2009) exploited run and Wilcoxon–type rank sum statistics to detect possible shifts of a monitored process. Some recent advances on this topic can be also found in Malela-Majika *et al.* (2016a, 2016b), Mukherjee *et al.* (2013) or Triantafyllou (2017). For an overview of distribution–free control charts for continuous variables, interested readers are referred to Chakraborti and Graham (2007) or Chakraborti (2011).

To increase the sensitivity of control charts in detecting small process shifts sensitizing rules based on runs and scans are often used in practice. Klein (2000) suggested two alternatives to the standard Shewhart $\overline{X}$ control chart: the two of two (2/2) and the two of three (2/3) control charts which have symmetric upper and lower control limits. Both control charts are easily implemented and have better *ARL* performance than the standard Shewhart $\overline{X}$ control chart for process average shifts up to 2.6 standard deviation. Khoo (2004) extended the work of Klein (2000) by proceeding a simulation study of the *ARL* performance of the 2/2, 2/3, 2/4, 3/3, and 3/4 control charts. Moreover, two new Shewhart-type nonparametric control charts are proposed therein for monitoring the unknown location parameter of a continuous population in Phase II (prospective) applications. The charts are based on control limits given by two specified order statistics from a reference sample and using some runs-type signalling rules.

In the present article, we implement the runs rule 2–out–of–3 to the distribution–free Phase II Shewhart–type control chart of Balakrishnan *et al*. (2009), which uses the Wilcoxon–type rank sum statistic. In Section 2, the setup of the proposed chart is described and the simulation algorithm is presented in detail. In Section 3, the False Alarm Rate (*FAR)* values are computed for several design parameters. The out–of–control performance of the proposed control scheme is investigated under two different underlying distributions. Finally, the proposed chart is compared to other antagonistic charts already introduced in the literature.

2. Main results

Generally speaking, the control limits of a distribution–free control chart are established based on a reference sample $X_1, X_2, ..., X_m$ drawn from an in-control (*IC*) process. Let us first denote by $F_X(x) = F(x)$ the in–control (cumulative) distribution and assume that two specific order statistics, say $X_{a:m}, X_{b:m}$, $1 \le a < b \le m$, are picked out. The integers *a*, *b* are design parameters and are appropriately selected so that a specific false alarm rate or an in–control average run length (ARL_{in}) value is achieved.

Suppose next that test samples are drawn independently of each other (and also of the reference sample) and that we are interested in checking whether the process is still in-control or not. In statistical terms, if $Y_1, Y_2, ..., Y_n$ denotes the test sample and $F_Y(x) = G(x)$ the corresponding cumulative distribution function, our aim is to detect a possible shift in the underlying distribution from $F(x)$ to $G(x)$, i.e., to test the null hypothesis $H_0 : F(x) = G(x)$ against the one-sided alternative $H_1 : F(x) < G(x)$. Note that some specific alternatives such as the location-shift alternative and the Lehmann alternative are subclasses of the general alternative mentioned above.

The proposed nonparametric control charts make use of the length of runs of test sample observations that lie between successive failures of the reference sample. More specifically, let us denote by M_i, i=1,2,…,m, the number of test sample observations Y_j that fall between the (i–1)–*th* and i–*th* order statistic of the X–sample (with the convention: $X_{(0)} = -\infty$). Clearly, M_i provide the lengths of runs of Y–observations between successive X–observations. Summing up the ranks (in the joint X and Y sample) of the Y–observations that lie between the control limits, we deduce the Wilcoxon–type rank sum statistic

$$W = \sum_{i=a+1}^{b} W_i ,$$

where W_i stands for the sum of ranks of the Y–observations falling between $X_{(i-1)}$ and $X_{(i)}$. It is noteworthy that the Wilcoxon–type rank sum statistic can also be expressed in terms of M_i, on observing that

$$W_i = \sum_{j=1}^{M_i} ((i-1) + \sum_{r=1}^{i-1} M_r + j) = M_i ((i-1) + \sum_{r=1}^{i-1} M_r) + \frac{M_i (M_i + 1)}{2} .$$

Carrying out some direct algebraic manipulations, the following expression may be established

$$W = \frac{1}{2}\left(\sum_{i=a+1}^{b} M_i\right)^2 + \sum_{i=a+1}^{b} iM_i + (M_0 + a - \frac{3}{2})\sum_{i=a+1}^{b} M_i \, .$$

Please note that Balakrishnan *et al.* (2009) introduced three new distribution–free Shewhart–type control charts that exploit run and Wilcoxon–type rank sum statistics to detect possible shifts of a monitored process.

The class of nonparametric control charts, introduced in the present paper, makes use of an 2–out–of–3 runs rule. More specifically, the process is declared to be out-of-control if at least two out of three successive test samples do not meet at least one of the following two conditions:

Condition 1. The number of observations of the Y–sample before the observation $X_{a:m}$ is less or equal to c_1, namely $M_0 = \sum_{i=1}^{a} M_i \le c_1$, where c_1 is a positive-valued parameter.

Condition 2. The Wilcoxon–type rank sum statistic W does not exceed a threshold, namely $W \le c_2$, where c_2 is a positive-valued parameter.

Each one of the conditions stated above, defines a separate plotting statistic. In words, the proposed distribution-free control scheme requires the construction of two parallel one-sided charts. The first one, should depict the test statistic M_0, while in the second chart the function W shall be plotted. Note that constants c_1 and c_2 are design parameters, playing the role of the corresponding upper control limits of the aforementioned charts. It is goes without saying that, if at least one of the conditions is violated, the corresponding test sample is characterized as out–of–control.

The performance of a control chart is usually adjudged by its run length distribution and/or its potential (not) to produce an alarm signal when the process is out–of–control (in–control, respectively). The term “run length” refers to the distribution of the waiting time random variable, which accounts for the number of samples until we get the first out–of–control signal from the chart. In what follows, we shall take into account the mean of the abovementioned distribution, called *Average Run Length*, for measuring the performance of the proposed monitoring scheme.

Throughout this paper, we present a simulation study of the distribution-free control chart introduced earlier. More specifically, we apply appropriate Monte Carlo simulation scheme measuring the performance of the proposed chart in order to determine numerically its robustness and effectiveness in detecting possible shifts of the underlying distribution. The simulation algorithm is based on the assumption that whenever the process is in-control both reference and test samples are derived from the same distribution (or alternatively we say that there is no shift in process underlying distribution, namely $\delta = 0$). On the other hand, when the process has shifted to an out-of-control state (*OOC* case, hereafter), test samples are drawn from an alternative distribution of the same shape compared to the in-control state, but having a shifted location parameter. In words, the *Y*–observations are distributed as $(X-\delta)$–observations, namely the *Y*–population is supposed to be shifted to the right (left) if $\delta > 0\,(\delta < 0)$. A brief description of the algorithm for simulating the performance of the proposed nonparametric control chart is given as follows.

Step 1. Determine the reference and test sample size (*m*, *n*) and the number of repetitions (*k*).

Step 2. Choose appropriate design parameters a, b, r_0, w.

Step 3. Generate a random reference sample from the pre-specified distribution *F*. The upper and lower control limit of the control chart are determined by using two order statistics of the reference sample, namely

$$LCL = X_{a:m}, UCL = X_{b:m}, \quad 1 \le a < b \le m \ .$$

Step 4. Produce a test sample from the same distribution *F*. Calculate the number of observations of the test sample, say M_0, that are smaller than $LCL = X_{a:m}$ and the number M_i, $i = a+1, a+2, ..., b$, of test sample observations that fall between the (*i*–1)–*th* and *i*–*th* order statistic of the reference sample. Check whether the conditions 1 and 2, stated earlier in the manuscript, are satisfied or not and fulfill appropriately the entries of Table *T* (1 and 0 respectively).

Step 5. Repeat Step 4 for *k* times.

Step 6. Determine the number of consecutive (uninterrupted) triplets consisting of *T's* elements, namely ($T(i)$, $T(i+1)$, $T(i+2)$), $i = 1, ..., k-2$, that do not satisfy the 2–out–of-3 rule. These numbers, say RL_i, are associated with the "run length" of the chart.

Step 7. Calculate the False Alarm Rate as $FAR = 1 - \frac{\sum RL_i}{k-2}$ and the value of the in-control Average Run Length according to $ARL_{in} = \frac{1}{length(RL)} \sum RL_i$.

Please note that in order to determine the out-of-control performance of the proposed scheme, namely to calculate the out-of-control *Average Run Length* or *Alarm Rate*, the main difference constitutes that test samples are now drawn from a shifted distribution, meaning from a distribution with location parameter shifted by δ units compared to distribution *F*.

For the implementation of the simulation algorithm, MATLAB software was used. We first apply a slight modification of the abovementioned algorithm in order to calculate *FAR* values of the well-known nonparametric control chart established by Balakrishnan *et al.* (2009). Table 1 reveals that the proposed simulation scheme produces adequate approximations of the corresponding theoretical results. More specifically, the first two columns of Table 1 consist of the theoretical False Alarm Rate values of the monitoring scheme proposed by Balakrishnan *et al.* (2009) for reference sample size m = 40 and 100 respectively (denoted by *B-T-K*). The second part of Table 1 includes *FAR* results, for the same design parameters, produced by adopting a modified algorithm that turns off the 2–out–of–3 rule. It is worth mentioning that for each result shown in the following table, the simulation scheme produced 1,000 random reference samples, each one of them combined with 100,000 test samples (k=100,000 repetitions). It goes without saying that simulated results are quite near to the corresponding theoretical values.

Table 1: *FAR* values for several designs of the nonparametric chart by Balakrishnan *et al*. (2009)

		B-T-K						***Results of the simulation algorithm***					
		***m* = 40**			***m* = 100**			***m* = 40**			***m* =100**		
		(*a*,*b*)	***w***	***FAR***	**(*a*,*b*)**	***w***	***FAR***	**(*a*,*b*)**	***w***	***FAR***	**(*a*,*b***	***w***	***FAR***
5	2	(7,11)	200	0.0471	(15,19)	15	0.0292	(7,11)	200	0.0480	(15,19)	150	0.0297
			40	0.0597		50	0.0464		40	0.0582		50	0.0465
			33	0.0900		33	0.0696		33	0.0889		33	0.0711
	3	(13,17)	90	0.0499	(15,19)	50	0.0201	(13,17)	90	0.0505	(15,19)	50	0.0201
			63	0.0601		33	0.0451		63	0.0590		33	0.0445
			58	0.0923		32	0.0857		58	0.0916		32	0.0850
	4	(16,20)	100	0.0249	(25,29)	60	0.0185	(16,20)	100	0.0246	(25,29)	60	0.0180
			72	0.0430		54	0.0323		72	0.0443		54	0.0316
			65	0.0987		53	0.0604		65	0.0998		53	0.0612

3. Numerical experimentation

In this section, we carry out an extensive numerical experimentation to illustrate the efficacy of the new control charts and their robustness features under both in-control and out–of–control situations. All numerical results displayed in this section, are produced based on the algorithm mentioned previously. From this point forward the proposed control chart will be denoted by $B\text{-}T\text{-}K\text{-}RR_{2\text{-}of\text{-}3}$.

Table 2 displays the False Alarm Rate (*FAR*) for several design parameters of B-T-K-$RR_{2\text{-}of\text{-}3}$ control chart.

Table 2: *FAR* values for given design parameters

		Reference Sample size m								
		40			**60**			**100**		
n	r_0	***(a,b)***	***w***	***FAR***	***(a,b)***	***w***	***FAR***	***(a,b)***	***w***	***FAR***
5	2	(7,11)	80	0.011	(15,18)	170	0.040	(23,27)	125	0.025
			30	0.054		60	0.056		48	0,087
	4	(13,17)	58	0.012	(18,23)	60	0.018	(18,22)	38	0.029
			30	0.067		40	0,064		30	0.093
10	3	(5,9)	100	0.011	(13,17)	180	0,083	(17,20)	120	0.024
			33	0.063		100	0,095		40	0.048
	6	(16,20)	210	0.021	(16,20)	90	0.010	(26,30)	90	0.020
			90	0.084		71	0.048		57	0.052

According to the above table, one may easily construct distribution-free control charts that achieve a pre-determined level pf *FAR*. The large amount of design parameters of the new control scheme, e.g. m, n, a, b, r_0, w gives the practitioner a notable flexibility for achieving a pre-specified level of in-control or out-of-control performance of the resulted chart. For example, if m = 40 and n = 5, then *FAR* equal to $f = 0.01$ (almost) can be achieved using:

(a) the 7th and 11th observed value of the reference sample $(a = 7, b = 11)$, $w = 80$ and $r_0 = 2$ (with *FAR* = 0.011) or,

(b) the 13th and 17th observed value of the reference sample, $(a = 13, b = 17)$, $w = 58$ and $r_0 = 4$ (with *FAR* = 0.012).

The simulation algorithm has been applied to study the performance of the proposed control chart for *OOC* process shifts. We next compute the Average Run Length (*ARL*) for small mean shifts of two different underlying distributions: (a) the standard Normal distribution *N* (0,1) and (b) the *Gamma* distribution with scale parameter equal to 3 and shape parameter equal to 1 (see Table 3).

Table 3: *ARL* values of the proposed chart under the *N*(0,1) and *Gamma* (3,1) distribution

***Shift* (*δ*)**	***N*(0,1)**	***Gamma* (3,1)**
0.0	506.09	499.78
0.1	245.17	263.02
0.2	105.07	167.72
0.3	54.11	84.34
0.4	29.59	57.34
0.5	18.25	35.30
0.6	10.74	23.17
0.7	6.61	15.93
0.8	4.74	10.91
0.9	3.50	7.19
1.0	2.78	5.45

We next evaluate the capability of the $B\text{-}T\text{-}K\text{-}RR_{2\text{-}of\text{-}3}$ control chart to detect the offset of the process mean that extends up to 1 unit ($\delta = 0\ (0.1)\ 1$). The in-control *ARL's* are almost equal to 500 for both underlying distributions. For example, whenever the process shifts by $\delta = 0.4$ units, the $B\text{-}T\text{-}K\text{-}RR_{2\text{-}of\text{-}3}$ scheme is expected to produce an out–of–control alarm after testing almost 30 samples (or more precisely 29.59 samples). Figure 1 illustrates the numerical results displayed in Table 3.

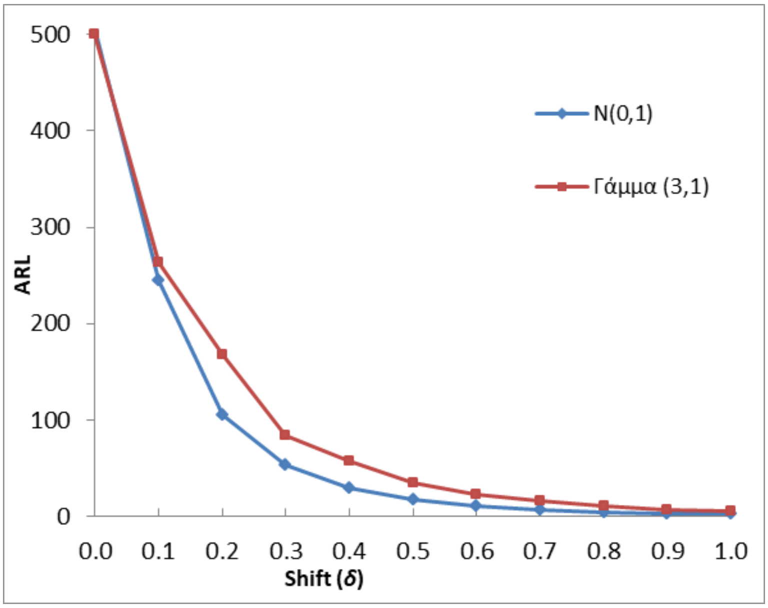

Figure 1: *ARL* values of the proposed chart under the *N*(0,1) and *Gamma* (3,1) distribution

Table 4: ARL_{out}'s of 5 competitive control schemes under *Normal distribution N* (δ, 1)

Shift (δ)	***B-T-K-RR$_{2\text{-}of\text{-}3}$***	***W_{min}***	***MW***	***W-CUSUM***	***W-EWMA* (λ=0.1)**
0.0	506.09	501.66	502.48	498.64	502.94
0.25	79.11	100.82	428.03	333.45	321.52
0.50	18.25	25.59	292.77	107.19	103.15
1.00	2.78	3.52	86.57	13.04	14.29
1.50	1.00	1.33	28.52	6.25	7.52
2.00	1.00	1.03	11.08	4.32	5.30
3.00	1.00	1.00	2.93	2.90	3.64

The proposed control chart *B-T-K-RR$_{2\text{-}of\text{-}3}$* is compared to the $W_{\min}$ –chart introduced by Koutras and Triantafyllou (2017), to the *W-CUSUM* and *W-EWMA* control charts established by Li *et al*. (2010) and the *Mann-Whitney* chart (*MW* chart) proposed by Chakraborti and van de Wiel (2008). It is worth noting that the $W_{\min}$ – chart has been compared with the corresponding nonparametric chart of Balakrishnan *et al.* (2009) and has been verified to outperform. Tables 4 and 5 depict the out–of–control (ARL_{out}) performance of the proposed chart against four other antagonistic control charts.

Table 5: ARL_{out}‘s of 5 competitive control charts under the *Gamma* (3,1) distribution

Shift (δ)	***B-T-K-RR$_{2\text{-}of\text{-}3}$***	***W_{min}***	***MW***	***W-CUSUM***	***W-EWMA* (*λ*=0.1)**
0.0	499.78	501.66	498.25	502.17	501.13
0.25	121.2	140.79	601.43	381.02	340.50
0.50	35.3	43.72	427.29	109.54	103.00
1.00	5.45	6.21	148.28	10.34	11.79
1.50	2.03	1.74	48.16	5.34	6.50
2.00	1.00	1.06	18.22	3.81	4.75
3.00	1.00	1.00	3.35	2.72	3.41

Figures 2 and 3 illustrate the numerical results of Tables 4 and 5. One may readily observe that the proposed control chart is superior to the other competitive charts, in almost all cases considered, as it detects faster the distribution’s displacement.

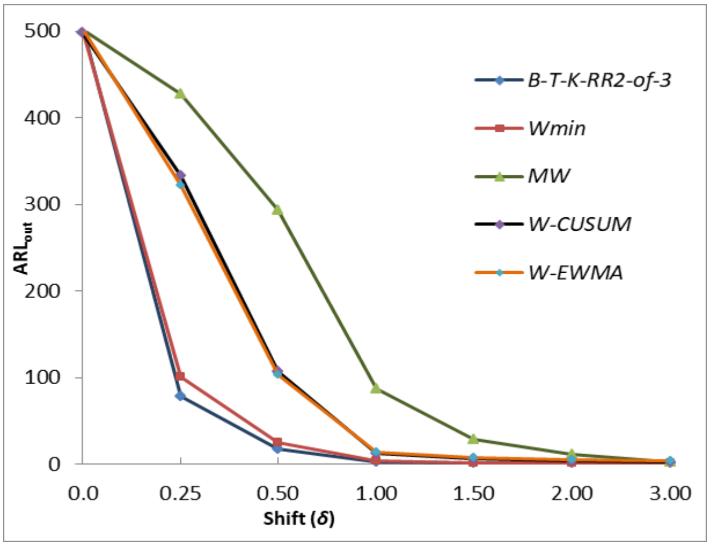

Figure 2: ARL_{out}'s versus shift of 5 competitive control charts under the $N(\delta,1)$ distribution

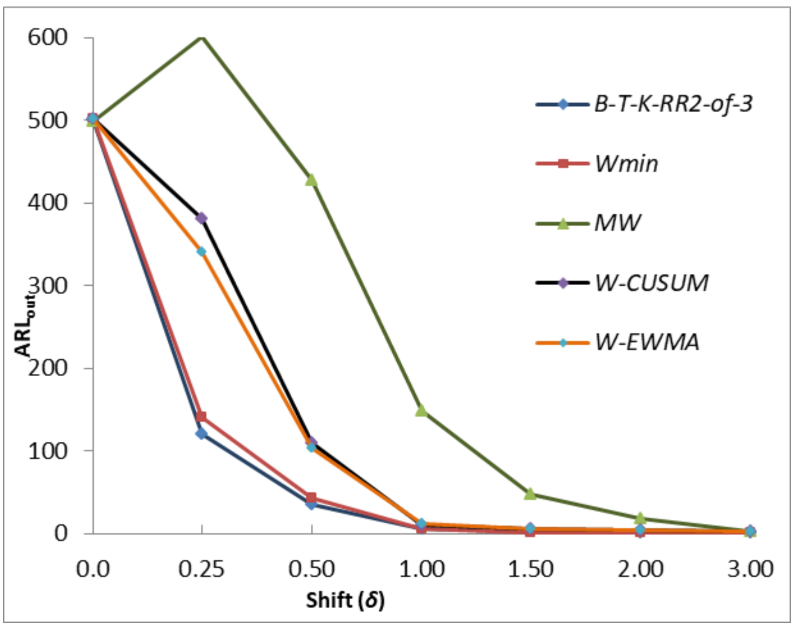

Figure 3: ARL_{out} 's versus shift of 5 competitive control charts under the *Gamma* (3,1) distribution

References

[1] Bakir, S.T. (2006). Distribution−free quality control charts based on signed–rank–like statistics, *Communications in Statistics: Theory & Methods*, **35**, 743–757.

[2] Balakrishnan, N., Triantafyllou, I.S. & Koutras, M.V,. Nonparametric control charts based on runs and Wilcoxon–type rank–sum statistics, *Journal of Statistical Planning and Inference*, **139**, (2009), 3177-3192.

[3] Balakrishnan, N., Triantafyllou, I.S. & Koutras, M.V., A distribution-free control chart based on order statistics, *Communication in Statistics: Theory & Methods*, **39**, (2010), 3652-3677.

[4] Chakraborti, S., Nonparametric (Distribution-free) Quality Control Charts, *Encyclopedia of Statistical Sciences*, (2011),1–27.

[5] Chakraborti, S. and Eryilmaz, S., A nonparametric Shewhart–type signed–rank control chart based on runs, *Communications in Statistics: Simulation & Computation*, **36**, (2007),335–356.

[6] Chakraborti, S. and Graham, M.A., Nonparametric control charts, *Encyclopedia of Quality and Reliability*, New York, John Wiley, 2007.

[7] Chakraborti, S., van der Laan, P. and van der Wiel, M.A., A class of distribution-free control charts, *Journal of the Royal Statistical Society, Series C-Applied Statistics*, **53**, (2004), 443–462.

[8] Khoo, M.B.C., Design of runs rules schemes, *Quality Engineering*, **16**, (2004), 27-43.

[9] Klein, M., Two alternatives to the Shewhart $\overline{X}$ control chart, *Journal of Quality Technology*, **32**, (2000), 427-431.

[10] Koutras, M.V. and Triantafyllou, I.S., A general class of nonparametric control charts, *Quality and Reliability Engineering International, to appear*, (2017).

[11] Malela-Majika, J.C., Chakraborti, S. and Graham, M.A., Distribution-free Phase II Mann–Whitney control charts with runs-rules, *International Journal Advanced Manufacturing Technology*, **86**, (2016a), 723-735.

[12] Malela-Majika, J.C., Chakraborti, S. and Graham, M.A., Distribution-free precedence control charts with improved runs rules, *Applied Stochastic Models in Business and Industry*, **32**, (2016b), 423-439.

[13] Mukherjee, A., Graham, M.A. and Chakraborti, S., Distribution-Free Exceedance CUSUM Control Charts for Location, *Communication in Statistics*: *Simulation & Computation*, **42**, (2013), 1153-1187.

[14] Triantafyllou, I.S., Nonparametric control charts based on order statistics: some advances, *Communication in Statistics*: *Simulation and Computation*, (2017), DOI: 10.180/03610918.2017.1359283.