RT&A, No 1 (82)
Volume 20, March 2025

USE OF MEDIAN BASED ESTIMATOR TO MITIGATE OUTLIER'S EFFECT THROUGH S^2 CHART

Sonam Jaiswal

•

DST-Centre for Interdisciplinary Mathematical Sciences, Banaras Hindu University,
Uttar Pradesh, India
jsonam572@gmail.com

Abstract

In this paper, we consider an upper-sided Phase II variance chart with probability limits in case of unknown parameter because the quality practitioner interested in monitoring increased variance of the process parameter. It is well established that when the Phase I data are contaminated with spurious observations, performance of the chart is suspected to deviate from what is normally expected. Therefore, we propose an improved performance of one-sided variance chart under the exceedance probability criterion for a fixed in-control average run length using the absolute deviation from median estimator. Under the exceedance probability criteria, the chart is designed so that the user can get more confidence in their in-control average run length values. The proposed chart is compared with the existing chart in case of contaminated and non-contaminated observations. Result shows that performance of variance chart shows robust performance when using absolute deviation from median estimator. Finally, an example has been provided in the favour of our proposed study.

Keywords: Average run length, median-based estimator, control chart, in-control and out-of-control performances, process variability.

1. Introduction

S^2 chart is considered to be more useful control chart when the interest of the quality practitioner lies in monitoring variability in the process parameter. As Woodall and Montgomery [24] stated that to maintain a process at a satisfactory level, process variability should be in-control (IC) because a slight change in the process variance could significantly impact the performance of the mean control chart. Therefore, prior to the construction and effective utilization of the mean control chart, it is suggested that a good estimate of the IC process variance must be available so that an effective process monitoring can take place. In this view, S^2 chart is a popular choice to monitor the process variability (Montgomery [17]). However, when the underlying process variance is not known, designing these charts becomes more complex. In this case, the variance is estimated from a Phase I reference sample and perform the Phase I analysis. (Chakraborti, Graham and Human, [2], Jones-Farmer et al. [11]). The estimate is then used to find the control limits which further used in Phase II analysis. For the S^2 chart, several efforts have been made to increase the efficacy of the charts such as: use of memory type control charts (CUSUM and EWMA) (Chang and Gan [3, 4]), use of runs rules (Rakitzis and Antzoulokas [18]), use of some other sampling plans such as repetitive sampling plan (Jaiswal and Kumar [8]), double sampling plan (Khoo [12]), etc.

As we know, in case U, the estimation error can lead to a distorted chart performance. This effect can be reduced by considering a larger number of Phase I samples and, at times, by adjusting the control limits (Saleh et al. [20, 21]). However, when Phase I samples, specially of smaller sizes, contain outliers, it is anticipated that this may have a more severe impact on the chart's performance. Because inclusion of the spurious observations may lead us to the model misspecification, biased parameter estimation and incorrect results. In turn, erroneous parameter estimation may affect the performance of the control chart in Phase II. Consequently, when a control chart indicates an OOC signal, pinpointing the underlying factors responsible for triggering this signal can prove to be a challenging task. Such signals can stem from assignable causes or merely be the result of spurious observations. The primary aim of this article is to recommend an estimator capable of mitigating the influence of outliers on the chart's performance, thereby enabling us to attribute OOC signals to genuine changes in the process.

Recently, Kumar and Jaiswal [15] studied the exponential chart and recommended the median based estimator for estimating the rate parameter so that the chart's performance is robust to the presence of outliers. Schoonhoven, Riaz and Does [23] have discussed different estimators of population variance for the variance chart and recommended the average deviation from median (ADM) estimator which is the function of sample median. They showed that the use of ADM estimator instead of commonly used Pooled estimator helps in minimizing the outliers' impact on the chart's performance. But they adopted the unconditional perspective to assess the chart's properties which mainly considers the mean and standard deviation of the unconditional RL distribution. Please note here that the unconditional run length distribution can be obtained by averaging the conditional run length distribution over the distribution of the estimator (see Chakraborti [1], Kumar and Chakraborti [14]). This method of obtaining results is known as unconditional perspective. This perspective has been criticized by several researchers, for example, Jardim [9, 10], Sarmiento et al. [22], Kumar [13] pointing out that the unconditional perspective does not consider the shape of the RL distribution and ignores the practitioner-to-practitioner variability. In this article, we consider the most recent approach i.e., conditional perspective which is based on the conditional RL distribution (see, Jardim [9, 10], Kumar [13], Gandy and Kvaloy [7], Epprecht et al. [5]). Conditional RL perspective mainly concerned with the exceedance probability criteria (EPC). For a detailed discussion on both perspective, readers are advised to refer some recent papers, for instance, Jardim [10], Sarmiento et al. [22], Kumar [13], Kumar and Jaiswal [16]. The unconditional perspective may lead to a misconception for the user due to the skewed distribution of the IC CARL (CARL(1)). For instance, the unconditional ARL might appear higher than the nominal ARL, suggesting a reduced rate of false alarms compared to the expected one. Nevertheless, examining the percentiles of the CARL(1) may reveal a contrasting narrative.

Hence, the article primarily focuses on the performance in a realistic context based on the percentiles of the CARL distribution and the EPC metric when the parameter is estimated using the ADM estimator. This assessment aims to determine if the chart's performance is suspectable to outliers.

Rest of the article is organized as follows. In section 2, the estimated control limits of the upper-sided S^2-chart has been discussed. In section 3, the IC and OOC performance of plug-in Pooled and ADM chart has been discussed. In section 4, The control limits are adjusted under EPC for the ADM and Pooled chart. In section 5, the IC and OOC performance has been discussed under EPC. In favor of the proposed design, an example has been offered in section 6. Finally, conclusions are offered in section 7.

II. Upper-sided S^2-chart with estimated IC variance

Let $X_1, X_2, \dots, X_n$ be n random samples of size n following a normal distribution with IC process mean μ and process variance $\sigma_0^2 > 0$ i.e., $X \sim N(\mu, \sigma_0^2)$. Traditionally used charting statistic for the S^2-chart is the sample variance, given by $S^2 = \frac{1}{n-1}\sum_{j=1}^{n}\left(X_j - \bar{X}\right)^2, \bar{X} = \frac{1}{n}\sum_{j=1}^{n} X_j$. Let UCL denotes the upper control limit of the S^2-chart which can be obtained by using probability approach, such that $P[S^2 > UCL|IC] = \alpha$, where α is a nominal FAR. It is well known that the statistic $\frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2_{n-1}$. Therefore, UCL is given by.

$$\text{UCL} = \frac{\sigma_0^2}{n-1}\chi^2_{1-\alpha,n-1} \tag{1}$$

where $\chi^2_{1-\alpha,n-1}$ be the $(1-\alpha)$th quantile of the χ^2-distribution with $(n-1)$ degrees of freedom.

Let σ_1^2 denotes the magnitude of the process variance shift from IC process variance σ_0^2 to the shifted variance $\sigma_1^2 = \delta\sigma_0^2$. A control chart gives an OOC signal when the charting statistic, S^2, falls above the UCL. This event is called signaling event E. And the probability of this signaling event, commonly known as the probability of signal for given shift, δ, denoted by $\beta(\delta)$ is given by.

$$\beta(\delta) = P[\,S^2 > \text{UCL}|\sigma_1^2 = \delta\sigma_0^2] = 1 - F_{\chi^2_{n-1}}\left(\frac{\chi^2_{1-\alpha,n-1}}{\delta}\right) \tag{2}$$

where $F_{\chi^2_{n-1}}(\cdot)$ denotes the CDF of χ^2-distribution with $n-1$ degrees of freedom. Its corresponding ARL is the reciprocal of probability of signal i.e., $\beta(\delta)$, is given by.

$$\text{ARL}(\delta) = \frac{1}{\beta(\delta)} = \frac{1}{1 - F_{\chi^2_{n-1}}\left(\frac{\chi^2_{1-\alpha,n-1}}{\delta}\right)} \tag{3}$$

Clearly, $\delta = 1$ represents the process is IC, otherwise, the process is OOC. On the other hand, an OOC ARL should be as low as possible so that the chart could detect shift in the process through a valid alarming signal as early as possible.

In case U, the process parameters are often unknown, and they need to be estimate using the Phase I samples assuming that the samples are collected from the IC process and they are ready to estimate the unknown process parameter. Let Y_{ij} be the i^{th} Phase I sample of size n. For the S^2 chart, the most prominent unbiased estimator suggested in the literature is sample Pooled estimator is given by.

$$\hat{\sigma}_0^2 = \hat{\sigma}^2_{\text{Pooled}} = \frac{1}{m}\sum_{i=1}^{m} S_i^2 \tag{4}$$

where $S_i^2 = \frac{1}{n-1}\sum_{j=1}^{n}\left(Y_{ij} - \bar{Y}_i\right)^2$ and $\bar{Y}_i = \frac{1}{n}\sum_{j=1}^{n} Y_{ij}$ is the sample variance of i^{th} group of samples. It is a function of the sample mean. On the other hand, the ADM estimator, the function of the sample median is given by.

$$\hat{\sigma}_0^2 = \hat{\sigma}^2_{\text{ADM}} = \overline{\text{ADM}} = \frac{1}{m}\sum_{i=1}^{m} \text{ADM}_i \tag{5}$$

where ADM_i is the average absolute deviation from the median of sample i, which is given by

$$\text{ADM}_i = \frac{1}{n}\sum_{j=1}^{n}\left|Y_{ij} - M_i\right|, \tag{6}$$

where M_i denotes the median of the i^{th} Phase I sample. An unbiased ADM estimator for estimating the sample variance is $\frac{\overline{\text{ADM}}}{t_2(n)}$. Here, $t_2(n)$ is a constant, function of sample size n and defined as $t_2(n) = \frac{2(n-1)}{n\sqrt{2\pi n(n-1)}} + 2\int_{-\infty}^{+\infty} x\Phi\left(\sqrt{n-1}x\right)\phi(x)dx$, where $\Phi(\cdot)$ and $\phi(\cdot)$ are the CDF and PDF of the standard normal distribution (see, Wu et al., 2002). Since the expression of $t_2(n)$ cannot be obtained in the closed form. Therefore, Riaz and Saghir [19] obtained the simulated results of the expression $t_2(n)$ for different values of n and mentioned in Table A1 of the appendix of the paper Riaz and Saghir [19].

Let $\widehat{\text{UCL}}$ denotes the estimated upper control limit which can be obtained by replacing σ_0^2 given in Equation (1) by its estimate $\hat{\sigma}_0^2$ where $\hat{\sigma}_0^2 = \hat{\sigma}^2_{\text{Pooled}}$ or $\hat{\sigma}^2_{\text{ADM}}$ given in Equation (4) and (5), respectively. The $\widehat{\text{UCL}}$ for the upper-sided S^2 chart is given by.

$$\widehat{\text{UCL}} = \frac{\hat{\sigma}_0^2}{n-1}\chi^2_{1-\alpha,n-1} \tag{7}$$

Let E_i be the i[th] event falling outside the $\widehat{\text{UCL}}$. Therefore, its corresponding conditional probability of signal (CPS), denoted by $\hat{\beta}$ for given shift (δ), is given by.

$$\hat{\beta}(\hat{\sigma}_0^2,\delta) = P\left[S^2 > \widehat{UCL}|\hat{\sigma}_0^2\right] = 1 - F_{\chi^2_{n-1}}\left(\frac{\hat{\sigma}_0^2\chi^2_{1-\alpha,n-1}}{\sigma_0^2\delta}\right) \tag{8}$$

Therefore, the conditional average run length (CARL) of the upper-sided S^2 chart can be obtained by using Equation (8), is given by.

$$\text{CARL}(\hat{\sigma}_0^2,\delta) = \frac{1}{\hat{\beta}(w,\delta)} = \left(1 - F_{\chi^2_{n-1}}\left(\frac{\hat{\sigma}_0^2\chi^2_{1-\alpha,n-1}}{\sigma_0^2\delta}\right)\right)^{-1} \tag{9}$$

The unconditional average run length for given shift δ, denoted by $\mu_{\text{CARL}}(\delta)$ is given by

$$\mu_{\text{CARL}}(\delta) = \int_0^\infty \left(1 - F_{\chi^2_{n-1}}\left(\frac{\hat{\sigma}_0^2\chi^2_{1-\alpha,n-1}}{\sigma_0^2\delta}\right)\right)^{-1} f_{\hat{\sigma}_0^2}\, d\hat{\sigma}_0^2 \tag{10}$$

The standard deviation of CARL for given shift (δ) is given by.

$$\sigma_{\text{CARL}}(\delta) = \sqrt{E(\text{CARL}^2(\hat{\sigma}_0^2,\delta)) - [E(\text{CARL}(\hat{\sigma}_0^2,\delta))]^2} \tag{11}$$

where $E(\text{CARL}^2(\hat{\sigma}_0^2,\delta)) = \int_0^\infty \left(1 - F_{\chi^2_{n-1}}\left(\frac{\hat{\sigma}_0^2\chi^2_{1-\alpha,n-1}}{\sigma_0^2\delta}\right)\right)^{-2} f_{\hat{\sigma}_0^2}\, d\hat{\sigma}_0^2$. Please note here $\delta = 1$ represents that the process is IC otherwise, the process is OOC. It is well known that lower values of $\sigma_{\text{CARL}}(1)$ are desirable for a good chart that reflects more confidence of the user in his/her CARL(1) value and hence in adopting the chart. The 100p[th] percentile of the $\text{CARL}(\hat{\sigma}_0^2,\delta)$ distribution denoted by $\text{CARL}(1)_p$, is given by.

$$\text{CARL}(1)_p = inf\{z: F_{\text{CARL}}(z) \geq p\} \tag{12}$$

where inf indicates infimum and $F_{\text{CARL}}(z)$ is the distribution function of the $\text{CARL}(\hat{\sigma}_0^2,\delta)$.
Beside the metrics discussed above, EP is the exceedance probability, denoted by $\pi(1)$ is defined as the chance that a chart will achieve his CARL(1), value at least nominal ARL_0, is given by.

$$\pi(1) = P[\text{CARL}(1) \geq \text{ARL}_0] \tag{13}$$

III. EPC performance of the Pooled and ADM chart with and without outliers

In this section, we examine the effect of upper outliers on the performance of S^2 chart. For this purpose, we have applied the simulation procedure using approximately 1,00,000 replications. With the underlying objective discussed above, the present study undertakes an examination of two different scenarios i.e., 5% and 10% spurious observations in each Phase I sample. In both inspections, the Phase I sample configuration encompasses spurious observations, with specific proportion of 5% and 10% relative to the total Phase I samples. Because the number of outliers, say γ, is an integer, we look at only the integer part of 5% or 10% of the Phase I sample of size m. Following simulation steps are carried out to obtain the performance metrices.

- Generate observations $Y_{i,j}; i = 1,2,\ldots,m\,; j = 1,2,\ldots,n$ from the normal distribution with mean μ and variance σ^2.
- Obtain S_i^2 or ADM_i for $i = 1,2,\ldots,m$.
- Sort them in either ascending or descending order.
- To produce the upper extremes (${S^2}_{(n)}$ or $\text{ADM}_{(n)}$) in the Phase I sample, multiply a constant, c, i.e., $c > 1$ to the first largest γ observation whereas $c = 1$ represents the Phase I sample with no contamination.
- Calculate the control limits with the estimators $\hat{\sigma}^2_{\text{Pooled}}$ or $\hat{\sigma}^2_{\text{ADM}}$ Given in Equation 4 or 5.
- Calculate CARL function using Equation 9 associated with its control limits obtained in previous step.
- Repeat the process at least 100,000 times to get the μ_{CARL}, σ_{CARL}, $\pi(1)$ and $\text{CARL}(1)_p$.

Following Table 1 and 2 represents the IC plug-in performance of the Pooled chart (Pooled estimator based S^2 chart) and ADM chart (ADM estimator based S^2 chart), respectively at $\alpha = 0.0027$,

$n = 5$ with the effect of 5% outliers in the Phase I samples, respectively. For the convenient of the computation of the ADM chart, the values of the constant $t_2(n)$ are taken from Riaz and Saghir [19]. The numerical value of $t_2(n)$ for $n = 5$ is 0.664980 and for $n = 7$ is 0.703800. In the Tables 1 and 2, m represent the sample size, γ represents the number of outliers in the Phase I samples, c is a multiplier to produce the outliers of different sizes in the Phase I sample, $\mu_{\text{CARL}}(1)$ and $\sigma_{\text{CARL}}(1)$ represents the mean and standard deviation of the CARL distribution, $\pi(1)$ is the probability that the CARL(1) is at least ARL_0 and $\text{CARL}(1)_p$ is the different percentiles of the CARL distribution. Moreover, $c = 1$ represents Phase I sample having no oulier.

Table 1: *IC performance of Pooled chart at $\alpha = 0.0027$ with and without outlier at $n = 5$ (with 5% outlier).*

					Percentile				
$m(\gamma)$	c	$\mu_{\text{CARL}}(1)$	$\sigma_{\text{CARL}}(1)$	$\pi(1)$	0.10	0.25	0.50	0.75	0.90
20(1)	1	803.91	2102.16	0.48	90	166	346	776	1687
	1.2	1020.05	2987.47	0.55	106	198	422	962	2126
	1.5	1519.47	5606.73	0.64	132	254	564	1335	3081
	2	2954.40	13393.58	0.77	194	393	921	2347	5807
50(2)	1	490.78	452.67	0.49	149	225	361	594	954
	1.2	536.69	500.14	0.53	160	241	391	650	1050
	1.5	604.64	584.36	0.59	176	267	439	732	1189
	2	748.41	778.05	0.68	208	320	530	901	1494
100(5)	1	425.25	241.23	0.49	194	260	366	519	722
	1.2	440.80	252.84	0.52	201	269	379	540	750
	1.5	465.30	272.85	0.56	210	284	400	570	793
	2	512.15	298.04	0.63	228	308	436	626	880
200(10)	1	397.43	150.72	0.50	235	289	369	473	594
	1.2	402.29	160.39	0.51	237	293	373	478	601
	1.5	414.33	158.54	0.54	244	301	384	492	621
	2	433.39	166.30	0.59	253	315	402	515	650
500(25)	1	380.24	88.77	0.50	277	317	369	431	497
	1.2	383.40	87.19	0.51	278	319	371	435	502
	1.5	387.04	92.19	0.53	281	322	375	440	506
	2	393.92	91.55	0.55	286	328	382	447	514

It is well known that estimation error exerts bad impact on the performance of the chart. Moreover, the sample Pooled estimator is a function of mean whereas ADM estimator is a function of sample median. Therefore, effect of outliers on the performance of the chart can be visualize from these tables. For instance, when $m = 20$, 5% of the Phase I sample (m) produces 1 outlier. It can be observed that when the Phase I sample is free from the outliers, its $\mu_{\text{CARL}}(1)$ and $\sigma_{\text{CARL}}(1)$ is 803.91 and 2102.16 whereas after including outliers say, for $c = 1.5$, its $\mu_{\text{CARL}}(1)$ and $\sigma_{\text{CARL}}(1)$ is 1519.47 5606.73, respectively which is approximately 98% larger than the 803.91 and much far than the nominal 370. On the other hand, using the ADM estimator, when the Phase I sample is free from the outliers i.e., $c = 1$, its $\mu_{\text{CARL}}(1)$ and $\sigma_{\text{CARL}}(1)$ is 438.79 and 301.61 whereas after including outliers, say, for $c = 1.5$, its $\mu_{\text{CARL}}(1)$ and $\sigma_{\text{CARL}}(1)$ is 613.67 and 449.80,respectively which is approximately 40% larger than the 438. These results shows that the Pooled chart deviated more from its nominal performance in case U than the ADM chart. Moreover, $\pi(1)$ values are showing less confidence in the values of CARL(1) which is only 50% even for the large sample sizes. And 10^{th} percentile is 90 for the Pooled chart and 168 for the ADM chart when $(m, n) = (20,5)$ which shows that there is 90% chance that CARL(1) of a conditional chart will be greater than or equal to 90 and 168 respectively, which is very low even for the larger Phase I samples. From these tables, it can be seen that the ADM chart puts on a guard against the outliers in the Phase I samples. The study shows that more than 500 Phase I samples of the size $n = 5$ are required to attain the control chart's performance close to

Table 2: *IC performance of ADM chart at $\alpha = 0.0027$ with and without outlier at $n = 5$ (with 5% outlier).*

					Percentile				
$m(\gamma)$	c	$\mu_{\text{CARL}}(1)$	$\sigma_{\text{CARL}}(1)$	$\pi(1)$	0.10	0.25	0.50	0.75	0.90
20(1)	1	438.79	301.61	0.48	168	239	358	544	792
	1.2	502.61	357.73	0.56	189	271	408	621	916
	1.5	613.67	449.80	0.68	226	326	493	759	1130
	2	864.80	670.03	0.83	299	440	684	1072	1620
50(2)	1	391.16	158.09	0.47	223	280	361	469	595
	1.2	411.79	170.20	0.53	234	294	380	493	627
	1.5	445.03	185.12	0.60	251	316	409	534	681
	2	507.83	210.98	0.72	284	358	466	611	782
100(5)	1	377.24	103.07	0.47	257	302	362	435	515
	1.2	386.62	109.04	0.51	263	309	371	446	529
	1.5	401.25	112.89	0.56	272	321	385	463	549
	2	426.49	122.37	0.64	289	341	409	493	583
200(10)	1	369.90	71.51	0.46	284	318	362	412	464
	1.2	374.82	72.03	0.48	287	323	367	418	470
	1.5	381.49	70.57	0.52	293	328	373	425	479
	2	393.20	79.18	0.58	301	338	385	439	494
500(25)	1	370.00	45.95	0.44	311	335	363	393	423
	1.2	371.80	43.17	0.45	312	336	365	396	426
	1.5	377.95	50.65	0.47	314	338	366	398	429
	2	379.42	50.81	0.51	318	342	371	402	434

the case K. Such a large amount of data is not easily available in real practice. Thus, it needs an adjustment in the control limits so that desired IC performance of the chart can be achieved with the available Phase I samples at hand. Therefore, to improve the performance, specially, for small sample sizes, we adjust the UCL of the chart so that higher chance of occurrence can be achieved.

IV. Adjusted control limit of the upper-sided S^2 chart under the EPC

In light of the limited availability of the extensive dataset, we have designed the control limit of the Pooled and ADM chart using the EPC approach. As discussed earlier, EPC approach ensures the high chance of occurrence, say 0.90, of the CARL(1) at least a nominal value such as 370.4. Formally, the condition of EPC approach can be written in terms of following equation as follows.

$$P[\text{CARL}(1) \geq \text{ARL}_0] = 1 - p; 0 < p < 1 \quad (14)$$

The values of the design constants are obtained at $p = 0.10$ i.e., EPC = 0.90 . The control limits of the proposed ADM chart under the EPC can be obtained by using the following simulation study.

- Fix the value of p, ARL_0, m, n and U where $U = \chi^2_{1-\alpha,n-1}$ is a design parameter.
- Generate observations $X_{i,j}; i = 1,2, \dots, m; j = 1,2, \dots, n$ from the normal distribution with mean μ and variance σ^2.
- Sort the subgroup data of size n in ascending or descending order and obtain the median (M_i) of the i^{th} sample of size n and calculate the ADM estimator for estimating the sample variance using $\frac{\overline{\text{ADM}}}{t_2(n)}$.

- Calculate the conditional control limit using ADM estimator and obtain the empirical distribution of the CARL(1) function using Equation (9).
- Repeat the process atleast 1,00,000 times to obtain the pth percentile i.e., $\mathrm{CARL}(1)_p$ of the CARL distribution, say $p = 0.10$.
- If $\mathrm{CARL}(1)_p > ARL0$, stop the loop and use the current value of $\widehat{UCL}$ otherwise increase the value of $\widehat{UCL}$ until the $CARL(1)_p > ARL0$ occur and return to previous step.

In order to obtain the control limits for the Pooled chart under the EPC, please follow Faraz et al. [6].

Table 3: *Design parameter of upper-sided Pooled and ADM chart with estimated parameter at $n = 5,7$ $p = 0.10$, $ARL_0 = 370.4$.*

	$n = 5$		$n = 7$	
m	Pooled chart	ADM chart	Pooled chart	ADM chart
20	20.2264	18.2357	23.9253	21.9893
50	18.5905	17.4718	22.3684	21.2454
75	18.1196	17.2454	21.9121	21.0200
100	17.8485	17.1122	21.6475	20.8898
200	17.3536	16.8654	21.1614	20.6434
500	16.9338	16.6506	20.7455	20.4313

V. IC and OOC performance of the Pooled and ADM chart with or without contamination under the EPC

I. IC performance with and without outliers

In this section, we are analyzing the IC performance of the upper-sided S^2 chart in the presence of some contaminated or spurious observations using the Pooled and ADM estimator under the EPC.

Table 4: *IC performance of Pooled chart with estimated parameter at $n = 5, p = 0.10, ARL_0 = 370.4$ with 5% outlier*

					Percentile				
$m(\gamma)$	c	$\mu_{\mathrm{CARL}}(1)$	$\sigma_{\mathrm{CARL}}(1)$	$\pi(1)$	0.10	0.25	0.50	0.75	0.90
20(1)	1	8205.59	41339.33	0.90	368	803	2047	5659	15046
	1.2	11064.39	115454.45	0.93	453	1004	2621	7479	20589
	1.5	20390.54	118299.76	0.95	604	1396	3805	11369	32884
	2	50015.21	1083023.37	0.98	965	2389	7078	23277	74305
50(2)	1	1544.50	1783.65	0.90	371	592	1031	1839	3179
	1.2	1718.68	2012.15	0.92	400	645	1125	2021	3522
	1.5	1982.94	2392.56	0.94	448	728	1286	2330	4112
	2	2549.14	3313.11	0.96	541	895	1610	2981	5328
100(5)	1	894.92	583.27	0.90	368	512	747	1100	1587
	1.2	933.95	621.40	0.91	384	532	777	1149	1658
	1.5	988.92	652.95	0.93	404	563	822	1218	1750
	2	1099.09	732.12	0.95	442	618	909	1352	1961
200(10)	1	655.63	267.27	0.90	369	464	602	787	1006
	1.2	667.19	276.57	0.91	376	472	612	800	1023
	1.5	688.39	283.47	0.92	386	486	631	826	1058
	2	720.99	301.97	0.94	403	508	661	866	1109

500(25)	1	517.06	128.03	0.90	371	426	500	589	684
	1.2	521.37	127.48	0.91	373	429	505	594	690
	1.5	526.59	129.90	0.91	376	433	510	600	697
	2	536.57	132.00	0.93	384	442	519	611	711

The performance of the chart can be obtained by using the design parameters provided in Table 3 under EPC. Following Table 4 - 5 represents the IC performance of the Pooled and ADM chart under the EPC for $n = 5$ having 5% spurious observation in the Phase I samples. Further, Table 6 - 7 represents the IC performance of the Pooled and ADM charts for $n = 5$, respectively having 10% spurious observation in the Phase I samples. It can be observed from Table 4 that the Pooled chart performance when the Phase I sample having no outlier i.e., $c = 1$, its $\pi(1)$ value 0.90 and $\left(\mu_{\text{CARL}}(1), \sigma_{\text{CARL}}(1)\right) = (8205.59, 41339.33)$ when $(m, n) = (20,5)$ while when the Phase I sample having outlier i.e., $c = 1.5$, its $\pi(1)$ value is 0.95 and $\left(\mu_{\text{CARL}}(1), \sigma_{\text{CARL}}(1)\right) = (20390.54, 118299.76)$. On the other hand, performance of the ADM chart from Table 5 informs us that when Phase I sample having no outlier i.e., $c = 1$, its $\pi(1)$ value is also 0.90 and $\left(\mu_{\text{CARL}}(1), \sigma_{\text{CARL}}(1)\right) = (1126.81, 943.21)$ when $(m, n) = (20,5)$ while when the Phase I sample having outlier i.e., $c = 1.5$, its $\pi(1)$ value is 0.96 and $\left(\mu_{\text{CARL}}(1), \sigma_{\text{CARL}}(1)\right) = (1656.19, 1450.44)$. Study reflect that both the charts are reflecting confidence in the values of $\text{CARL}(1)$ by the metric $\pi(1)$ i.e., $\pi(1) = 0.90$ when Phase I sample having no outliers. But the performance of the Pooled chart is deviated more in the presence of outliers than the ADM chart. Moreover, the 10th percentile of the Pooled and ADM chart is approaching 370 which shows more confidence in the values of the $\text{CARL}(1)$. For instance, when $(m, n) = (20,5)$, the 75th percentile of the Pooled chart is 5659 when $c = 1$ and 11369 at $c = 1.5$ whereas the 75th percentile of the ADM chart is 1393 when $c = 1$ and 2044 when $c = 1.5$. It means there is approximately 25% chance that the $\text{CARL}(1)$ of the chart may occur greater that the 5659 for the Pooled chart and 1393 and 2044 for the ADM chart, respectively. All these information about the ADM chart are appearing more closer to the desired performance and less deviated from the nominal performance. Therefore, ADM chart outperforms the Pooled chart when we consider the estimation of the parameter with contaminated data.

Table 5: *IC performance of ADM chart with estimated parameter at* $n = 5, p = 0.10, ARL_0 = 370.4$ *with 5% outlier*

					Percentile				
$m(\gamma)$	c	$\mu_{\text{CARL}}(1)$	$\sigma_{\text{CARL}}(1)$	$\pi(1)$	0.10	0.25	0.50	0.75	0.90
20(1)	1	1126.81	943.21	0.90	369	551	871	1393	2154
	1.2	1313.90	1104.80	0.93	425	635	1005	1620	2520
	1.5	1656.19	1450.44	0.96	511	777	1251	2044	3220
	2	2435.42	2244.63	0.99	709	1094	1800	2994	4812
50(2)	1	683.84	301.52	0.90	370	472	622	826	1071
	1.2	723.41	316.94	0.92	388	497	658	875	1134
	1.5	787.24	354.23	0.95	420	540	715	952	1239
	2	905.31	418.50	0.97	479	617	820	1097	1431
100(5)	1	555.97	164.89	0.90	370	439	531	645	771
	1.2	570.33	170.97	0.91	379	450	546	663	791
	1.5	593.08	176.35	0.93	393	467	567	690	823
	2	634.21	186.53	0.96	418	498	606	739	884
200(10)	1	486.70	99.42	0.90	369	417	476	545	616
	1.2	493.73	96.19	0.91	374	422	483	553	625
	1.5	502.32	105.94	0.93	381	430	491	562	635
	2	519.21	101.24	0.94	393	443	507	582	659

500(25)	1	436.91	60.92	0.90	369	398	432	471	508
	1.2	439.70	55.21	0.91	372	401	436	473	511
	1.5	442.78	53.19	0.92	374	403	439	477	516
	2	448.32	54.98	0.93	379	408	444	483	523

Similarly, Tables 6-7 which entails us about the study of 10% contaminations in the Phase I samples of size $n = 5$, respectively. The comprehensive study of the 10% contaminations also suspected to deviate from its nominal than expected. For example, when we consider $m = 50$ Phase I observations each of size $n = 5$, then 10% contamination produces $\gamma = 5$ outliers. For the Pooled chart, $\left(c, \mu_{\text{CARL}}(1), \sigma_{\text{CARL}}(1)\right) = (1{,}1554.08, 1766.69)$ and $\left(c, \mu_{\text{CARL}}(1), \sigma_{\text{CARL}}(1)\right) = (1.5, 1865.62, 2213.95)$. Similarly, when using ADM estimator, $(c, \mu_{\text{CARL}}(1), \sigma_{\text{CARL}}(1))$ is (1, 683.67, 302.38) and $(c, \mu_{\text{CARL}}(1), \sigma_{\text{CARL}}(1))$ is $(1.5, 770.89, 344.03)$ with high probability. Therefore, we recommend our proposed ADM chart under the EPC when the Phase I samples having spurious observations.

Table 6: *IC performance of* Pooled *chart with estimated parameter at* $n = 5, p = 0.10, ARL_0 = 370.4$ *with* 10% *outlier*

					Percentile				
$m(\gamma)$	c	$\mu_{\text{CARL}}(1)$	$\sigma_{\text{CARL}}(1)$	$\pi(1)$	0.10	0.25	0.50	0.75	0.90
20(2)	1	8144.06	50446.91	0.90	371	804	2054	5691	15341
	1.2	10477.68	141575.40	0.92	429	951	2465	6943	19201
	1.5	15791.01	114907.40	0.94	545	1245	3363	9865	27930
	2	31623.24	208322.50	0.97	797	1889	5364	16780	50540
50(5)	1	1554.08	1766.69	0.90	370	593	1031	1844	3194
	1.2	1657.69	2010.15	0.91	391	630	1096	1962	3393
	1.5	1865.62	2213.95	0.93	429	696	1215	2196	3852
	2	2266.70	2916.76	0.95	497	811	1447	2640	4699
100(10)	1	897.21	589.07	0.90	371	514	747	1101	1589
	1.2	928.81	606.99	0.91	381	528	774	1144	1647
	1.5	977.27	632.49	0.92	398	556	811	1201	1734
	2	1058.88	702.13	0.94	426	595	875	1299	1894
200(20)	1	653.83	279.71	0.90	37	463	601	783	1001
	1.2	665.24	275.90	0.91	376	472	612	797	1016
	1.5	681.55	280.48	0.92	383	482	625	817	1046
	2	709.17	295.85	0.93	397	499	650	851	1092
500(50)	1	515.81	126.99	0.90	368	425	498	588	686
	1.2	520.15	128.88	0.90	371	428	504	593	688
	1.5	523.62	132.36	0.91	374	431	506	596	693
	2	532.06	129.50	0.92	380	439	514	608	704

II. OOC performance of the chart without contamination

As for as OOC performance concern, following Table 8 represents the $\mu_{\text{CARL}}(\delta)$ and $\sigma_{\text{CARL}}(\delta)$ metrics for both the charts having different shift parameter, δ. These values are obtained using the expressions given in Equations (9) and (10) for nominal $\text{ARL}_0 = 370.4$, $p = 0.10$ and $\delta = 1{,}1.2, 1.5, 2$. The value $\delta > 1$ corresponds to the OOC situation when the process is deteriorated. We mention here that the ADM chart outperforms the Pooled chart under the EPC in terms of lower $\mu_{\text{CARL}}(\delta)$ and $\sigma_{\text{CARL}}(\delta)$ values for $\delta > 1$. Please note here that $\delta = 1$ represents IC performance of the Pooled and

ADM chart, respectively. For instance, when $(m, n, \delta) = (20,5,1.2)$ the $\mu_{\text{CARL}}(\delta)$ and $\sigma_{\text{CARL}}(\delta)$ is 1132.83 and 3197.63 respectively for the Pooled chart whereas $\mu_{\text{CARL}}(\delta)$ and $\sigma_{\text{CARL}}(\delta)$ is 268.91 and 171.92, respectively for the ADM chart. It implies that the ADM chart under the EPC takes less time to detect an OOC signal than the Pooled chart when the process deteriorates due to an increase in the rate parameter. Hence, ADM chart shows better performance in the OOC scenario and Pooled chart missed the signal.

Table 7: *IC performance of ADM chart with estimated parameter at* $n = 5, p = 0.10, ARL_0 = 370.4$ *with* 10% *outlier.*

					Percentile				
$m(\gamma)$	c	$\mu_{\text{CARL}}(1)$	$\sigma_{\text{CARL}}(1)$	$\pi(1)$	0.10	0.25	0.50	0.75	0.90
20(2)	1	1126.13	924.36	0.90	367	550	871	1395	2155
	1.2	1292.61	1088.71	0.93	418	623	990	1596	2481
	1.5	1583.65	1379.11	0.96	493	747	1199	1951	3079
	2	2228.73	2030.10	0.98	655	1009	1651	2740	4387
50(5)	1	683.67	302.38	0.90	371	475	623	824	1065
	1.2	718.32	315.86	0.92	385	494	653	868	1128
	1.5	770.89	344.03	0.94	413	529	701	934	1208
	2	872.96	397.46	0.97	463	595	789	1058	1380
100 (10)	1	555.51	165.33	0.90	370	438	531	645	770
	1.2	568.37	168.81	0.91	378	448	544	660	788
	1.5	588.76	175.02	0.93	391	464	563	683	818
	2	625.25	182.07	0.95	414	492	597	727	870
200 (20)	1	487.90	93.55	0.90	370	417	477	547	617
	1.2	492.67	101.49	0.91	374	421	482	552	623
	1.5	500.96	100.47	0.92	380	427	489	562	636
	2	515.31	108.45	0.94	391	440	504	578	653
500 (50)	1	437.49	52.58	0.90	396	39	434	572	510
	1.2	439.02	52.32	0.90	371	400	435	473	511
	1.5	442.09	58.80	0.91	374	403	438	476	514
	2	446.95	59.63	0.92	378	407	442	482	520

Table 8: *The OOC performance metrics* $\mu_{CARL}(\delta)$ *and* $\sigma_{CARL}(\delta)$ *of the Pooled chart and ADM chart for* $p = 0.10$ *and* $ARL_0 = 370.4$ *and shift size* $\delta = 1,1.2,1.5,2$ *at* $n = 5$.

			δ			
m	Estimator	PM	1	1.2	1.5	2
20	Pooled	$\mu_{\text{CARL}}(\delta)$	8205.59	1132.83	181.13	33.52
		$\sigma_{\text{CARL}}(\delta)$	41339.33	3197.63	290.07	30.38
	ADM	$\mu_{\text{CARL}}(\delta)$	1126.90	268.91	67.14	17.91
		$\sigma_{\text{CARL}}(\delta)$	932.39	171.92	31.84	5.82
50	Pooled	$\mu_{\text{CARL}}(\delta)$	1547.32	341.11	79.51	20.01
		$\sigma_{\text{CARL}}(\delta)$	1783.65	289.74	49.41	8.58
	ADM	$\mu_{\text{CARL}}(\delta)$	683.78	182.86	53.50	15.27
		$\sigma_{\text{CARL}}(\delta)$	301.52	65.28	24.33	4.75
100	Pooled	$\mu_{\text{CARL}}(\delta)$	894.92	224.90	59.10	16.44
		$\sigma_{\text{CARL}}(\delta)$	583.27	115.03	22.31	4.39
	ADM	$\mu_{\text{CARL}}(\delta)$	555.76	155.56	44.98	13.65
		$\sigma_{\text{CARL}}(\delta)$	164.89	36.28	8.13	1.85

200	Pooled	$\mu_{\text{CARL}}(\delta)$	655.51	177.15	49.53	14.59
		$\sigma_{\text{CARL}}(\delta)$	268.56	57.81	12.32	2.53
	ADM	$\mu_{\text{CARL}}(\delta)$	487.00	139.80	41.55	12.93
		$\sigma_{\text{CARL}}(\delta)$	97.79	24.23	5.34	1.13
500	Pooled	$\mu_{\text{CARL}}(\delta)$	517.07	146.66	43.07	13.25
		$\sigma_{\text{CARL}}(\delta)$	126.33	29.87	6.50	1.53
	ADM	$\mu_{\text{CARL}}(\delta)$	436.93	128.37	38.88	12.35
		$\sigma_{\text{CARL}}(\delta)$	55.07	13.49	3.38	0.73

VI. Simulated Example

In this section, we provide an illustration of simulated data set to show the findings. Following Table 9 represents a simulated dataset for $m = 20$ and $n = 5$ which are generated from the IC normal process i.e., $N(2,5)$. And the next 20 samples are generated from the $N(2,6.5)$. The fist 20 samples are used to estimate the parameter, $\hat{\sigma}^2$, for the S^2 chart and obtained as 21.96003. Design parameter of the Pooled chart (black dashed lines) and ADM chart (red dashed lines) for $m = 20$ are 20.22638 and

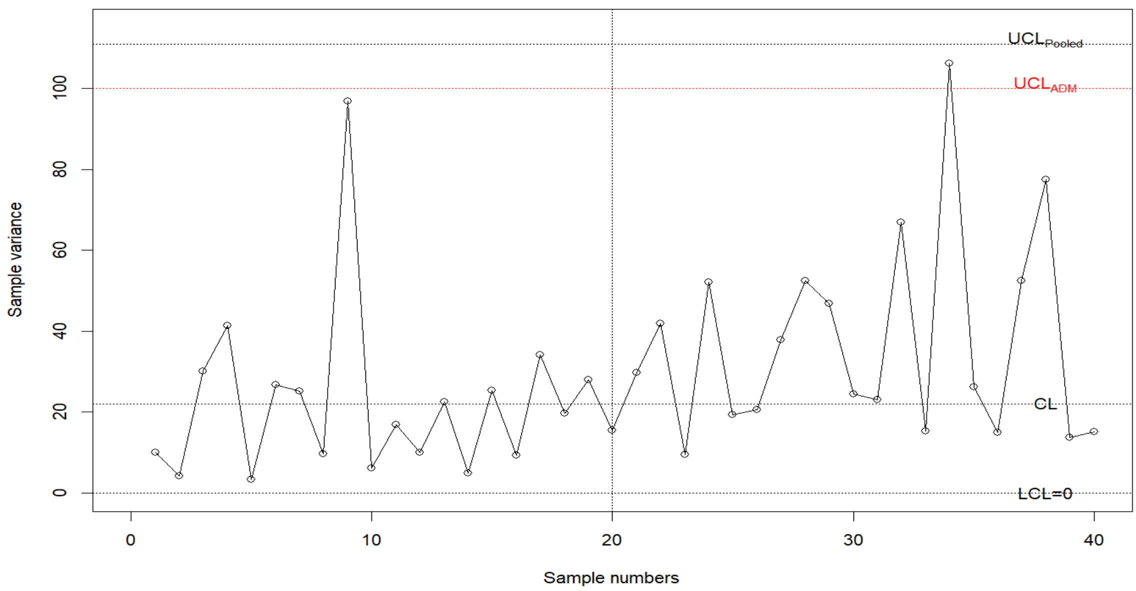

Figure 1: *Phase II control limits of the ADM chart and Pooled chart.*

Table 9: *Simulated dataset of $n = 5$ and $m = 40$.*

1	2	3	4	5	1	2	3	4	5
4.390	-1.876	-1.711	-3.963	-1.936	8.423	6.150	-2.161	4.727	-4.142
3.674	-0.163	0.309	2.008	4.467	1.211	10.332	14.511	9.105	-0.894
2.583	0.067	-5.437	8.495	-3.553	0.831	-1.156	-6.280	-2.271	1.441
4.661	-0.749	0.524	1.378	15.113	2.818	11.521	-7.932	-2.342	-0.992
-0.806	1.587	-0.575	2.097	3.476	4.079	14.962	7.891	4.323	7.178
-6.971	-3.373	3.723	0.242	5.745	4.776	-1.385	11.043	5.391	2.579
-2.447	3.949	5.299	11.462	6.507	-6.480	8.580	3.875	-1.578	-4.306
0.130	6.252	-1.676	-0.7694	0.807	0.028	0.937	1.473	-4.090	15.039
5.019	11.502	-6.013	-10.249	10.601	-12.609	-10.569	0.404	-0.301	2.081
4.051	4.236	8.168	1.918	6.835	2.105	4.565	-1.102	11.162	-0.366

8.699	10.217	4.429	3.398	0.069	-0.325	5.689	4.338	-0.056	-6.540
7.738	7.231	-0.079	4.510	6.362	4.326	10.523	-11.248	2.647	5.855
-0.001	9.725	-1.881	2.948	-1.181	1.383	9.274	0.665	0.427	-0.051
3.424	7.605	4.941	7.390	2.832	14.697	0.805	-12.239	-4.933	6.146
4.851	-3.772	9.981	2.367	5.381	-2.900	6.981	-3.778	-2.365	5.480
5.434	-0.973	1.394	1.425	6.247	7.268	5.136	-0.783	9.026	2.670
4.859	5.758	-5.486	9.878	6.982	-5.437	12.501	-4.984	1.411	1.063
13.135	3.751	4.430	1.668	4.289	5.986	-2.271	12.744	-9.001	8.882
-5.015	3.971	-0.611	6.145	8.011	9.175	6.264	13.228	7.132	3.265
4.019	-3.330	-2.007	0.802	5.971	3.048	13.574	7.352	9.829	7.016

18.2357, respectively. Estimated upper control limit of the Pooled chart is 111.043 whereas for the ADM chart is 100.1141. It can be observed from the Figure 1 that ADM-chart detects the OOC signal first at the 34th sample point under the EPC whereas the Pooled chart missed the OOC signal.

VII. Conclusion

In this paper, we have considered the estimation of Phase II control limits of an upper sided S^2 chart. The present study shows the adjustment of Phase II control limits of S^2 chart using the EPC approach. Undoubtedly, the EPC criterion degrades the chart's OOC performance but, on the other hand, it provides better IC performance for the pooled chart and ADM chart with desired IC performance for the smaller Phase I samples (say, $m < 100$). The study suggests that the ADM chart outperform in the presence of outliers.

As far as OOC performance is concern, the performance of the charts are also evaluated based on the EPC criterion. Study show that the ADM estimator takes less time, on average, to trigger a signal than the sample pooled estimator when $\delta \geq 1$. Hence, we recommend that the user must ensure that the Phase I sample is free from upper outliers and if in doubt then the ADM chart must be used to construct Phase II control limits. Finally, all the calculations are performed using the R statistical software and the programs are available from the authors on request.

Before closing, we mention that the present study considers an upper-sided S^2 chart. However, we have carried out the study of two-sided S^2 chart. Study shows that there was no any significant difference have been found in the case of contaminated observations in the values of ARL for improvement case because of the skewed nature of the run length distribution. Moreover, as a future direction, the present study can also be extended to examine the effect of presence of the spurious observations in the Phase I data on the EWMA and CUSUM charts.

References

[1] Chakraborti, S. (2000). Run length, average run length and false alarm rate of Shewhart X-bar chart: Exact derivations by conditioning. *Communications in Statistics, Simulation and Computation*, 29: 61–81.

[2] Chakraborti, S., M. A. Graham, and S. W. Human. (2009). Phase I statistical process control charts: An overview and some results. *Quality Engineering,* 21:52–62.

[3] Chang, T. C., and Gan, F. F. (1994). Optimal designs of one-sided EWMA charts for monitoring a process variance. *Journal of statistical Computation and Simulation*, 49:33-48.

[4] Chang, T. C., and Gan, F. F. (1995). A cumulative sum control chart for monitoring process variance. *Journal of Quality Technology*, 27:109-119.

[5] Epprecht, E. K., Loureiro, L. D., and Chakraborti, S. (2015). Effect of the amount of Phase I data on the Phase II performance of S^2 and S control charts. *Journal of Quality Technology*, 47:139–155.

[6] Faraz, A., Woodall, W. H., and Heuchenne, C. (2015). Guaranteed conditional performance of the S^2 control chart with estimated parameters. *International Journal of Production Research*, 53:4405-4413.

[7] Gandy, A., and Kvaløy, J. T. (2013). Guaranteed conditional performance of control charts via bootstrap methods. *Scandinavian Journal of Statistics*, 40:647–668.

[8] Jaiswal, S., and Kumar, N. (2023). Designing of Phase II ARL-unbiased S^2-chart with improved performance using repetitive sampling. *Reliability: Theory & Applications*, 4: 850-860.

[9] Jardim, F., S. Chakraborti, and E. K. Epprecht. (2019). $\bar{X}$ chart with estimated parameters: The conditional ARL distribution and new insights. *Production and Operations Management*, 28:1545–57.

[10] Jardim, F., S. Chakraborti, and E. K. Epprecht. (2020). Two perspectives for designing a Phase II control chart with estimated parameters: The case of the Shewhart $\bar{X}$ Chart. *Journal of Quality Technology* 52:198–217.

[11] Jones-Farmer, L. A., W. H. Woodall, S. H. Steiner, and C. W. Champ. 2014. An overview of Phase I nalysis for process improvement and monitoring. Journal of Quality Technology 46 (3):265–80.

[12] Khoo, M. B. (2004). $S2$ control chart based on double sampling. *International Journal of Pure*

[13] Kumar, N. (2022). Statistical design of Phase II exponential chart with estimated parameters under the unconditional and conditional perspectives using exact distribution of median run length. *Quality Technology & Quantitative Management*, 19:1-18.

[14] Kumar, N., and Chakraborti, S. (2016). Phase II Shewhart-type control charts for monitoring times between events and effects of parameter estimation. *Quality and Reliability Engineering International*, 32:315–328.

[15] Kumar, N., and Jaiswal, S. (2020). Use of median-based estimator to construct Phase II exponential chart. *Quality and Reliability Engineering International*, 36:2044-2056.

[16] Kumar, N., and Jaiswal, S. (2023). The design of two-sided S^2-chart with estimated in-control process variance based on conditional median run length. Quality Technology & Quantitative Management, 0:1-18.

[17] Montgomery, D. C. (2020). Introduction to statistical quality control. *John Wiley & Sons*.

[18] Rakitzis, A. C., and Antzoulakos, D. L. (2011). Chi-square control charts with runs rules. *Methodology and Computing in Applied Probability*, 13:657-669.

[19] Riaz, M. and Saghir, A. (2009). A Mean Deviation-Based Approach to Monitor Process Variability. Journal of Statistical Computation and Simulation 79, 1173‑1193.

[20] Saleh, N. A., Mahmoud, M. A., Jones-Farmer, L. A., Zwetsloot, I. N. E. Z., and Woodall, W. H. (2015). Another look at the EWMA control chart with estimated parameters. *Journal of Quality Technology*, *47*(4), 363–382.

[21] Saleh, N. A., Mahmoud, M. A., Keefe, M. J., and Woodall, W. H. (2015). The difficulty in designing Shewhart *X*– and X control charts with estimated parameters. *Journal of Quality Technology*, *47*(2), 127–138.

[22] Sarmiento, M. G. C., Jardim, F. S., Chakraborti, S., and Epprecht, E. K. (2020). Design of variance control charts with estimated parameters: A head-to-head comparison between two perspectives. *Journal of Quality Technology*, 54:249–268.

[23] Schoonhoven, M., Riaz, M., and Does, R. J. (2011). Design and analysis of control charts for standard deviation with estimated parameters. *Journal of Quality Technology*, 43:307-333.

[24] Woodall, W. H., and D. C. Montgomery. (1999). Research issues and ideas in statistical process control. *Journal of Quality Technology*, 31:376–86.