Songklanakarin J. Sci. Technol.
43 (6), 1610-1619, Nov. - Dec. 2021

http://www.sjst.psu.ac.th

Original Article

On the sensitivity of robust control charts in monitoring contaminated data

Kooi Huat Ng[1], Poh Kheng Soon[1], Jeng Young Liew[2*], and Kok Haur Ng[3]

[1] *Department of Mathematical and Actuarial Sciences, Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, Kajang, Selangor, 43000 Malaysia*

[2] *Department of Agro-Based Industry, Faculty of Agro-Based Industry, Universiti Malaysia Kelantan, Jeli Campus, Jeli, Kelantan, 17600 Malaysia*

[3] *Institute of Mathematical Sciences, Faculty of Science, Universiti Malaya, Lembah Pantai, Kuala Lumpur, 50603 Malaysia*

Received: 13 September 2020; Revised: 29 November 2020; Accepted: 29 December 2020

Abstract

Shewhart chart is typically formed under normality assumptions. In reality, much data is contaminated with occasional outliers, which may diminish the Shewhart chart's sensitivity. Hence, robust charts are introduced as outlier-resistant and robust against non-normality. This paper intends to study the robust monitoring of contaminated data using (i) median chart based on median absolute deviation (MAD), and (ii) trimmed mean chart based on winsorized standard deviation. These charts are compared with the conventional Shewhart mean $(\bar{X})$ charts based on standard deviation and range. In general, through extensive simulations, the robust charts are quite comparable with the $\bar{X}$ charts for normal data of small sample size (*n*), but for large *n*, the median chart based on MAD is marginally preferred by all the benchmarks considered. However, when a process is contaminated, both robust charts outshine the $\bar{X}$ charts substantially in a series of investigations.

Keywords: Shewhart chart, robust, outlier, contamination, standard deviation

1. Introduction

Statistical process control (SPC) is a crucial problem-solving tool for improving process productivity through the minimization of variability (Bouslah, Gharbi, & Pellerin, 2018). A control chart is the most powerful tool of SPC (Montgomery, 2012) and it is a graphical display that gives useful information for achieving process stability. Shewhart chart has garnered widespread application since its inception, and a scale estimator such as sample standard deviation is important in SPC as most quality controllers integrate the scale estimators into the Shewhart mean $(\bar{X})$ chart to estimate the process standard deviation for controlling the process variability.

It is common that the Shewhart chart assumes normal data with homoscedasticity (constant variance) (Tereza & Noskievičová, 2017). However, this assumption is difficult to vindicate in some real-life situations (Jacobs, 1990). Besides, the sample standard deviation is also non-robust against slight deviations from normality and outliers. Figueiredo and Gomes (2009) highlighted that even in the potentially normal situations, it is possible to have an underlying non-normal distribution, with a moderate to a strong degree of asymmetry and with its tails heavier than the normal tails, as well as a significant correlation between the observations. In this situation, the Shewhart charts are inappropriate for monitoring the process.

Robust methods are helpful in monitoring the process variability from non-normal or contaminated data. This is because a robust estimator is insensitive to changes in

*Corresponding author
Email address: ljyoung@umk.edu.my

the underlying distribution and outlier-resistant. By applying a robust chart, occurrences of either an extremely high or an extremely low false alarm rate can be deterred if the parameters to be controlled are near the targets, even with departures from normality. Rocke (1989) suggested that a control chart's limits should be formed using robust methods and concurrently, the non-robust methods are graphed on the chart to detect the outliers. Developed by Hampel (1974), the median absolute deviation (MAD) is a robust measure of dispersion. This estimated value is barely affected by the presence of extreme observations and one considers the MAD a robust estimators of the scale (Abu-Shawiesh, 2008). As disclosed by Abu-Shawiesh (2008), the median chart based on MAD bested the $\bar{X}$ chart based on standard deviation for data with heavy-tailed distributions.

Another alternative for replacing the sample mean and sample standard deviation is by using trimmed mean and winsorized standard deviation (WSD), respectively. The trimmed mean approach was presented by Tukey (1948). Compared to mean, trimmed mean shows much higher efficacy when one has a large percentage of trimming from a heavy-tailed symmetric distribution. The estimated standard deviation based on sample standard deviation is affected by extreme outliers. The WSD will compensate this effect by removing a small predetermined percentage of the smallest and largest values. The winsorized method gives the best protection against outliers among all counterparts (Kochelakota & Kocherlakota, 1995). Langenberg and Iglewicz (1986) introduced the $\bar{X}$ chart based on trimmed mean and concluded that it is less affected by outliers than the classical counterparts, at the expense of tighter limits when out-of-control situations occur. Recent notable progress in robust charts has been reported by Chakraborti, Eryilmaz, Human (2009), Hawkins and Deng (2010), Zhou and Geng (2013), Karagöz (2016), and Chiang, Lio, Ng and Tsai (2018).

This paper studies the sensitivity and robust monitoring of contaminated data using two robust charts: (i) median chart based on MAD, and (ii) trimmed mean chart based on WSD. A comparative study is performed to examine the performances of these two charts with the Shewhart charts, i.e., $\bar{X}$ charts based on standard deviation and range for normal and non-normal distributions through extensive simulation. The performances of these charts are assessed in terms of length of control limits, average change point estimates, standard error of change point estimates, and accuracy of estimate of the change point in the process shift.

2. Methodology

A review of Shewhart $\bar{X}$ charts and the two robust charting methods is given in this section.

2.1 Shewhart $\bar{X}$ chart based on standard deviation

In the assumption of random variable X having normal distribution with unknown mean μ and unknown standard deviation σ, let $\bar{X}_1, \bar{X}_2, ..., \bar{X}_m$ be the means of m subgroups with each subgroup consisting of n observations, the estimated μ is

$$\bar{\bar{X}} = \left(\bar{X}_1 + \bar{X}_2 + ... + \bar{X}_m\right) / m. \quad (1)$$

The j-th subgroup's sample standard deviation is

$$S_j = \sqrt{\sum_{i=1}^{n}\left(X_i - \bar{X}_j\right)^2 / (n-1)}, \quad j = 1, 2, ..., m. \quad (2)$$

Let $S_1, S_2, ..., S_m$ be the standard deviations for m subgroups. The average standard deviation is

$$\bar{S} = \left(S_1 + S_2 + ... + S_m\right) / m \cdot \quad (3)$$

The lower and upper control limits (LCL and UCL), and center line (CL) of the 3-sigma Shewhart $\bar{X}$ chart based on standard deviation are

$$\text{LCL} = \bar{\bar{X}} - 3\bar{S} / (c_4\sqrt{n}), \quad \text{UCL} = \bar{\bar{X}} + 3\bar{S} / (c_4\sqrt{n}), \quad \text{and} \quad \text{CL} = \bar{\bar{X}}, \quad (4)$$

respectively, where c_4 is the control limit factor depending on the sample size (Table 1) and $\bar{S}/c_4$ is an unbiased estimator of σ. A process is deemed to be out-of-control if a point falls beyond LCL and UCL.

2.2 Shewhart $\bar{X}$ chart based on range

The range is the difference between the maximum and minimum values in the observations, that is

$$R = X_{\max} - X_{\min}. \quad (5)$$

Assume that $R_1, R_2, ..., R_m$ are the ranges for *m* subgroups. The average range is

$$\bar{R} = \left(R_1 + R_2 + ... + R_m\right)/m. \quad (6)$$

The LCL, UCL and CL of the 3-sigma Shewhart $\bar{X}$ chart based on range are given by

$$\text{LCL} = \bar{\bar{X}} - 3\bar{R}/(d_2\sqrt{n}), \ \text{UCL} = \bar{\bar{X}} + 3\bar{R}/(d_2\sqrt{n}), \ \text{and} \ \text{CL} = \bar{\bar{X}}, \quad (7)$$

respectively, where d_2 is the control limit factor depending on the sample size (Table 1), and $\bar{R}/d_2$ is an unbiased estimator of σ.

2.3 Median chart based on MAD

The robust chart is robust against outliers. With a robust location estimator, *T* and the corresponding scale estimator, *D*, the 3-sigma robust chart's control limits and CL are

$$\text{LCL} = T - 3D/(A\sqrt{n}), \ \text{UCL} = T + 3D/(A\sqrt{n}), \ \text{and} \ \text{CL} = T, \quad (8)$$

where *A* is the control limit factor depending on the sample size and *D/A* is an unbiased estimator of σ.

The location and scale estimators of *j*-th subgroup for the median chart based on MAD are

$$M_j = \text{median}\left(X_i\right), \ i = 1, 2, ..., n; \ j = 1, 2, ..., m, \quad (9)$$

and

$$S_{Mj} = \text{MAD} = \text{median}\left\{\left|X_i - M_j\right|\right\}, \ i = 1, 2, ..., n; \ j = 1, 2, ..., m, \quad (10)$$

respectively.

Let $M_1, M_2, ..., M_m$ and $S_{M1}, S_{M2}, ..., S_{Mm}$ respectively be the median and MAD estimates for *m* subgroups. Then the average median is $\bar{M} = \sum_{j=1}^{m} M_j / m$ and average MAD is $\bar{S}_M = \sum_{j=1}^{m} S_{Mj} / m$. The control limits and CL for median chart based on MAD are computed by replacing *T* with $\bar{M}$, *D* with $\bar{S}_M$, and *A* with A_1 (Table 1) as given in Equation (8). Reader may refer to Kochelakota and Kocherlakota (1995) for A_1 of different sample sizes.

2.4 Trimmed mean chart based on WSD

The location estimator of *j*-th subgroup for trimmed mean chart based on WSD is

$$W_j = \frac{1}{n\left(1-2p\right)}\left\{\sum_{i=k+1}^{n-k} X_i + \left(k - pn\right)\left[X_k + X_{n-k+1}\right]\right\}, \ j = 1, 2, ..., m, \quad (11)$$

while the scale estimator is

$$S_{Wj}^2 = \frac{1}{n\left(1-2p\right)^2}\left\{\sum_{i=k+1}^{n-k}\left(X_i - W_j\right)^2 + k\left[(X_k - W_j)^2 + (X_{n-k+1} - W_j)^2\right]\right\}, \ j = 1, 2, ..., m, \quad (12)$$

where $k = \left[pn\right] + 1$, $\left[h\right]$ is the greatest integer less than equal to *h* in which *h* is an arbitrary real number, and *p* is the proportion to trim from each end. Let $W_1, W_2, ..., W_m$ and $S_{W1}, S_{W2}, ..., S_{Wm}$ respectively be the trimmed mean and WSD estimates for *m* subgroups. Then the average trimmed mean is $\bar{W} = \sum_{j=1}^{m} W_j / m$ and the average WSD is $\bar{S}_W = \sum_{j=1}^{m} S_{Wj} / m$. The control limits and CL are respectively given by Equation (8) with the replacement of the robust location estimator *T* as $\bar{W}$, scale estimator *D* as $\bar{S}_W$, and *A* with A_2 (Table 1). Reader may also refer to Kochelakota and Kocherlakota (1995) for A_2 of different sample sizes.

Table 1. Control limit factors of the corresponding control charts for $n \in \{5,10,20\}$.

Sample size, *n*	Factor			
	c_4	d_2	A_1	A_2
n = 5	0.940406	2.327529	0.548118	0.759179
n = 10	0.966257	3.056314	0.613456	0.923293
n = 20	0.984210	3.718258	0.646097	0.968217

2.5 Simulation study

The performances of the robust charts versus the $\bar{X}$ charts are compared for two process states suggested by Davis and Adams (2005) (Table 2).

Table 2. Process states for analysis.

Process states	Standard normal distribution
	Standard normal distribution with contamination

The first simulation results (Table 3) are conducted for $n \in \{5,10,20\}$ with m = 25 under (i) standard normal distribution, N(0,1) that represents the non-contaminated data, (ii) contamination model with 1% observations from N(3.5,1) and the remaining 99% from N(0,1) and (iii) contamination model with 5% observations from N(3.5,1) and the remaining 95% observations from N(0,1) for the performance evaluation.

For the second simulation, the results (Tables 4 to 6) are obtained using Samuel, Pignatiello and Calvin's (1998) procedure. The observations are first randomly simulated from a standard normal distribution for subgroups 1 to 100. Then, the observations are simulated from a contaminated distribution where the mean now is shifted to δ starting from subgroup 101 onwards. As the process change occurred at subgroup 101, the estimated time for detecting the change point should be close to 101. More specifically, the simulation from subgroups 1 to 100 is generated from $X_{ij} \sim \mathrm{N}\left(\mu,\sigma^2\right)$, for $i = 1, 2, ..., n,$ $j = 1, 2, ..., m.$ As for subgroups 101 through 10,000, the data are generated from

$$X_{ij} \sim \mathrm{N}\left(\mu,\sigma^2\right)\mathrm{I}_{(\beta,1)}\left(p_{ij}\right) + \mathrm{N}\left(\mu+\delta,\sigma^2\right)\mathrm{I}_{(0,\beta)}\left(p_{ij}\right),\ i = 1, 2, ..., n;\ j = 101, 102, ..., 10000, \quad (14)$$

where $\mu = 0,$ $\sigma = 1,$ δ is the process shift magnitude, and $\mathrm{I}_{(a,b)}\left(p\right)$ represents an indicator function with

$$\mathrm{I}_{(a,b)}\left(p\right) = \begin{cases} 1, & a < p < b, \\ 0, & \text{elsewhere}, \end{cases} \quad (15)$$

and β denotes the proportion of contaminated data. Note that *a* and *b* are the lower and upper limits of *p* which represents a random probabilistic value under the indicator function, I.

Let $\hat{t}_1, \hat{t}_2, ..., \hat{t}_N$ be the estimated times of change point detected for *N* trials, the average change point estimates of the process change is

$$\bar{\hat{t}} = \frac{1}{N}\sum_{i=1}^{N}\hat{t}_i, \quad i = 1, 2, ..., N, \quad (16)$$

and the corresponding standard error is

$$\text{standard error}\left(\hat{t}\right) = \frac{1}{N}\sqrt{\sum_{i=1}^{N}(\hat{t}_i - \bar{\hat{t}})^2}. \quad (17)$$

Simulation algorithms written in R programming language were used to compute the charts' (i) length of control limits, (ii) average change point estimates, (iii) standard error of the change point estimates, and (iv) percentages of detecting correctly the change point, and the simulation process was repeated for *N* = 10000 trials.

For ease of elaboration, the $\bar{X}$ charts based on standard deviation and range, median chart based on MAD, and trimmed mean chart based on WSD are denoted as E_1, E_2, E_3, and E_4, respectively, in the corresponding tables presented in Section 3. Meanwhile, data without contamination, 1% of contamination, and 5% of contamination are indicated by J_0, J_1, and J_5, respectively, as shown in Tables 3 to 6.

3. Results and Discussion

3.1 Length of control limits

The $\bar{X}$ charts have the shortest control limit range when compared to the robust charts under standard normal distribution for n = 5 (Table 3). However, the opposite is true when n increases to 10. Note that the length is the deviation between LCL and UCL.

When n = 20, the median chart defeats all charts under comparison by having the shortest range of control limits. The four charts are affected by 1% of contaminated data as the charts' control limits are wider in contrast to the normal data. In this case, the trimmed mean chart has the shortest length while the $\bar{X}$ charts possess the largest length of range. When the proportion of contamination is 5%, the lengths of control limits of the four charts become even much larger relative to the 1% contamination case. The $\bar{X}$ charts are significantly impacted by the 5% of contamination in comparison to its corresponding control limits without contamination. The robust charts, meanwhile, exhibit high robustness against contamination, so they have a narrower range within control limits than the $\bar{X}$ charts.

Table 3. Length of control limits for the $\bar{X}$ charts and robust charts for $n \in \{5, 10, 20\}$ and $m = 25$ computed under 10,000 trials for normal and non-normal data with contamination.

Data	Chart	n = 5			n = 10			n = 20		
		LCL	UCL	Length	UCL	UCL	Length	LCL	UCL	Length
J_0	E_1	-1.3427	1.3429	2.6856	-0.9537	0.9551	1.9088	-0.6731	0.6719	1.3450
	E_2	-1.3423	1.3425	**2.6848**	-0.9538	0.9552	1.9090	-0.6745	0.6734	1.3479
	E_3	-1.3571	1.3575	2.7146	-0.9510	0.9519	1.9029	-0.6723	0.6703	**1.3426**
	E_4	-1.3511	1.3521	2.7032	-0.9498	0.9509	**1.9007**	-0.6932	0.6913	1.3845
J_1	E_1	-1.3769	1.3785	2.7554	-1.0009	0.9986	1.9995	-0.7029	0.7019	1.4048
	E_2	-1.3777	1.3792	2.7569	-1.0107	1.0085	2.0192	-0.7206	0.7196	1.4402
	E_3	-1.3703	1.3712	2.7415	-0.9638	0.9616	1.9254	-0.6784	0.6773	**1.3557**
	E_4	-1.3578	1.3591	**2.7169**	-0.9614	0.9594	**1.9208**	-0.6999	0.6990	1.3989
J_5	E_1	-1.5659	1.5644	3.1303	-1.1457	1.1466	2.2923	-0.8164	0.8162	1.6326
	E_2	-1.5704	1.5688	3.1392	-1.1843	1.1852	2.3695	-0.8937	0.8935	1.7872
	E_3	-1.4521	1.4517	2.9038	-1.0049	1.0041	2.0090	-0.7025	0.7036	**1.4061**
	E_4	-1.4272	1.4261	**2.8533**	-1.0015	1.0010	**2.0025**	-0.7254	0.7257	1.4511

Boldfaced letters denote the shortest range within control limits, within a distribution for each sample size considered in the study.

3.2 Detection of change point in monitoring process

For comparison purposes, the performance of the four charts in detecting the change point in the monitoring process is assessed based on standard normal distribution N(0,1), and distributions with 1% and 5% of contamination. As the actual change point is at the time t = 101, the average change point estimate, $\bar{\hat{t}}$, should be close to 101 in the existence of shift.

For data without contamination, when n = 5 and $\delta = 0.5$, all four charts are insensitive enough to detect the change point at the exact change time of t = 101 but the $\bar{X}$ charts based on standard deviation and range exhibits a much quicker change point detection capability than the median and trimmed mean charts (Table 4). As the process shift δ increases, the four charts' average change point estimates decrease considerably. A similar trend is observed for $n \in \{10, 20\}$. In the case of the standard normal distribution N(0,1), for a large shift, says $\delta = 3$, the four charts detect the change point at approximately 102 irrespective of n values. The results in Table 4 reveal that both $\bar{X}$ charts have the best performance in estimating the change point when n = 5. In this case, the $\bar{X}$ chart based on range is somewhat outperforming the $\bar{X}$ chart based on standard deviation. There is a rapid change in trend in terms of performance in estimating the change point when n increases. When n = 10, the trimmed chart surpasses all charts under comparison, while as n = 20, the median chart rules for normal data.

The performance of the $\bar{X}$ charts' average change point estimate had deteriorated sharply for any n when contamination occurs, especially when δ is small. To illustrate, for 1% and 5% of contamination, both $\bar{X}$ charts' average change point estimates when n = 10 and $\delta = 0.5$ were over 300 and 1200, respectively, which fall far apart from the correct t = 101. For the non-normal data with 1% of contamination, the performance of the trimmed mean chart surpasses other charts in terms of average change point estimate when $n \in \{5, 10\}$. This is untrue for the case n = 20, where the median chart gives the best average change point estimate. Both robust charts are rather robust to non-normality when contamination is raised to 5% (Table 4). This is because both robust charts are insensitive against outliers and hence outshine the $\bar{X}$ charts in detecting the change point.

3.3 Standard error of the change point estimates

In this section, comprehensive comparisons between the robust charts versus the $\bar{X}$ charts in terms of standard errors of the change point estimates are made for $n \in \{5, 10, 20\}$ and $\delta \in \{0.5, 1, 1.5, 2, 3\}$ under normal and contaminated non-normal distributions.

Table 4. Average change point estimates for the $\bar{X}$ charts and robust charts, for normal and non-normal data with contamination when $n \in \{5,10,20\}$ and $\delta \in \{0.5,1,1.5,2,3\}$.

Data	n	Chart	Average change point estimates				
			$\delta = 0.5$	$\delta = 1$	$\delta = 1.5$	$\delta = 2$	$\delta = 3$
J_0	5	E_1	255.3456	143.9117	115.4165	106.2829	102.0259
		E_2	254.9885	143.8246	115.3857	106.2725	102.0248
		E_3	270.9050	147.3100	116.4211	106.6225	102.0769
		E_4	262.2711	146.0682	116.0169	106.4805	102.0566
	10	E_1	262.8621	145.8465	115.5606	106.4023	102.0506
		E_2	263.0213	145.8780	115.5751	106.4056	102.0520
		E_3	258.3066	144.6666	115.2744	106.3011	102.0325
		E_4	257.2095	144.3966	115.1983	106.2717	102.0278
	20	E_1	256.6290	144.2126	114.9380	106.3603	102.0007
		E_2	259.4345	144.9152	115.1385	106.4331	102.0092
		E_3	253.4546	143.4942	114.7324	106.2856	101.9885
		E_4	300.8532	154.5131	117.7602	107.281	102.1491
J_1	5	E_1	296.2518	153.8858	117.6338	107.2503	102.1369
		E_2	297.6695	154.0400	117.6914	107.2752	102.1400
		E_3	286.7039	151.6860	117.0536	107.0753	102.1094
		E_4	272.9958	148.6470	116.2168	106.7741	102.0672
	10	E_1	344.8088	163.4998	120.5119	108.0844	102.2934
		E_2	368.2659	169.1627	122.0891	108.5153	102.3627
		E_3	274.2057	147.5370	116.0777	106.7125	102.0673
		E_4	270.2467	146.7834	115.8531	106.6452	102.0576
	20	E_1	328.5640	161.8112	120.0435	107.7708	102.2395
		E_2	386.2722	174.8746	123.5364	108.8299	102.4098
		E_3	268.4445	146.8232	115.9807	106.5427	102.0308
		E_4	320.0970	159.6444	119.4842	107.6022	102.2143
J_5	5	E_1	830.1699	260.4282	143.7540	115.0267	103.2139
		E_2	855.0729	264.8079	144.8352	115.2835	103.2405
		E_3	422.8699	180.8393	124.3253	109.3447	102.4655
		E_4	369.2588	170.3996	121.6551	108.4156	102.3339
	10	E_1	1203.651	333.1635	158.7129	119.0098	103.7800
		E_2	1782.368	435.8703	179.8675	124.5061	104.4435
		E_3	356.4580	166.7855	120.9853	108.3895	102.3407
		E_4	349.3804	165.2202	120.6184	108.2434	102.3235
	20	E_1	1296.793	349.1050	163.0229	120.2658	103.9010
		E_2	4272.271	823.6488	258.1678	143.7664	106.3743
		E_3	336.7357	163.7403	120.3871	108.0358	102.2591
		E_4	418.6906	181.5689	125.0002	109.4342	102.4923

From Table 5, the larger the process shift magnitude δ, the lower the standard error of the change point estimate for the four charts. This is contributed by the higher sensitivity of a control chart in responding to a shift in the process mean when the process shift magnitude δ turns larger. When n = 5, the $\bar{X}$ charts have the lowest standard error of the change point estimates for all δ considered under normal data. It is worth pointing out that the $\bar{X}$ chart based on range is slightly superior to the $\bar{X}$ chart based on standard deviation for this case, but the triumph of the former chart against the latter chart turns to be rather trivial from $\delta = 2$ onwards. When n = 10, the trimmed mean chart surpasses all other charts under comparison except when $\delta = 2$ where both median and trimmed mean charts have similar performance. For n = 20, the median chart is preferable as it possesses the least standard error of change point estimates.

The presence of contaminated data has inflated the overall standard errors of change point estimates of a control chart. Of note, the overall standard error of change point estimates is evidently higher than that of the results for normal data (Table 5). The higher the proportion of contamination, the larger are the charts' standard errors in change point estimates. It is found that the $\bar{X}$ chart based on standard deviation performs better than the $\bar{X}$ chart based on range when the underlying data are contaminated. Nevertheless, both $\bar{X}$ charts fall short in comparison with the robust charts regardless of the contamination severity. This trend holds for any n and δ considered in this study. For 1% of contamination, the trimmed mean chart has lower standard errors compared to the median chart when $n \in \{5,10\}$, but for n = 20, the reverse is true. An analogous trend is observed for the case with 5% of contamination where the trimmed chart always bested the median chart in terms of standard error; but when n = 20, the median chart takes the crown. Generally, the robust charts have a much better standard error performance when dealing with contaminated data, relative to the $\bar{X}$ charts. However, for normal data, the $\bar{X}$ charts are more desirable when n is small.

3.4 Accuracy of the estimate of change point in process shift

Accuracy of the estimate of change point in process shift was also studied by evaluating the robust charts' and the

Table 5. Standard error of change point estimates from the correct t = 101 for the $\bar{X}$ charts and robust charts, for normal and non-normal data with contamination when $n \in \{5, 10, 20\}$ and $\delta \in \{0.5, 1, 1.5, 2, 3\}$.

Data	n	Chart	Standard error of change point estimates				
			$\delta = 0.5$	$\delta = 1$	$\delta = 1.5$	$\delta = 2$	$\delta = 3$
J_0	5	E_1	1.5372	0.4350	0.1466	0.0578	0.0143
		E_2	1.5338	0.4344	0.1463	0.0576	0.0143
		E_3	1.7002	0.4714	0.1572	0.0610	0.0148
		E_4	1.6393	0.4580	0.1520	0.0596	0.0145
	10	E_1	1.6454	0.4553	0.1526	0.0592	0.0145
		E_2	1.6464	0.4553	0.1528	0.0592	0.0146
		E_3	1.6011	0.4416	0.1499	0.0582	0.0143
		E_4	1.5915	0.4387	0.1488	0.0578	0.0143
	20	E_1	1.5720	0.4347	0.1464	0.0593	0.0142
		E_2	1.5952	0.4413	0.1479	0.0600	0.0143
		E_3	1.5364	0.4284	0.1443	0.0584	0.0141
		E_4	1.9992	0.5380	0.1742	0.0682	0.0159
J_1	5	E_1	1.9382	0.5357	0.1711	0.0676	0.0156
		E_2	1.9566	0.5369	0.1716	0.0679	0.0156
		E_3	1.8428	0.5140	0.1660	0.0657	0.0153
		E_4	1.7113	0.4859	0.1578	0.0631	0.0149
	10	E_1	2.3820	0.6411	0.1982	0.0765	0.0170
		E_2	2.6642	0.6942	0.2151	0.0808	0.0178
		E_3	1.7119	0.4755	0.1549	0.0625	0.0146
		E_4	1.6653	0.4687	0.1524	0.0619	0.0145
	20	E_1	2.2925	0.6111	0.1927	0.0723	0.0167
		E_2	2.8509	0.7380	0.2257	0.0829	0.0186
		E_3	1.6981	0.4591	0.1543	0.0602	0.0144
		E_4	2.2061	0.5877	0.1867	0.0709	0.0165
J_5	5	E_1	7.2240	1.5833	0.4283	0.1446	0.0270
		E_2	7.5501	1.6283	0.4394	0.1470	0.0272
		E_3	3.2340	0.8066	0.2375	0.0880	0.0189
		E_4	2.6784	0.7013	0.2114	0.0786	0.0176
	10	E_1	11.0430	2.3372	0.5723	0.1860	0.0323
		E_2	16.6967	3.3304	0.7907	0.2409	0.0388
		E_3	2.5839	0.6521	0.2042	0.0792	0.0179
		E_4	2.5246	0.6381	0.2013	0.0779	0.0178
	20	E_1	11.9271	2.4671	0.6185	0.1953	0.0338
		E_2	40.6804	7.1937	1.5851	0.4295	0.0586
		E_3	2.4142	0.6386	0.1998	0.0744	0.0167
		E_4	3.2503	0.8133	0.2466	0.0879	0.0192

Shewhart $\bar{X}$ charts' percentages of detecting correctly the change point for normal and non-normal data with contamination. To be more specific, the percentage of detecting correctly the change point t = 101 is assessed. Likewise, $n \in \{5, 10, 20\}$ and $\delta \in \{0.5, 1, 1.5, 2, 3\}$ are used for the comparative study for the four charts.

The probability of detecting correctly the change point in the process mean for all charts becomes higher as δ increases (Table 6). The $\bar{X}$ chart based on standard deviation is generally comparable to the corresponding $\bar{X}$ chart based on range, though the former chart transcends occasionally by a narrow margin in detecting the change point correctly at time 101. Both $\bar{X}$ charts outrival the performance of the robust charts in detecting change point correctly if n is small, i.e., n = 5 under a normal condition. This scenario changes completely for n = 10, where both robust charts provide better estimates compared to the $\bar{X}$ charts. For n = 20, median chart displays the best performance in the presence of contamination-free data.

When contamination occurs, all charts' overall percentages of estimating the change point correctly in the process are generally lesser than with the normal data. For instance, considering $\delta = 1$ and n = 5, the estimated percentage of identifying correctly the change point for the $\bar{X}$ chart based on standard deviation is 2.36% and the estimated percentage decreases considerably to 1.83% when the process contains 1% of contaminated data. An even more serious loss of the percentage of estimating change point correctly for all charts is observed in Table 6 when the contamination is raised to 5%. The precisions of $\bar{X}$ charts in detecting the change point correctly have been significantly influenced by the contamination level. That is, the performance of the $\bar{X}$ charts is becoming much inferior if the proportion of contamination increases. The robust charts, in contrast, are insensitive to the contaminated data. It can be seen that the percentages for both robust charts of detecting correctly the change point are always higher than those of the $\bar{X}$ charts irrespective of the values of n and δ (Table 6). Despite showing inferiority to the robust charts for contaminated data, both $\bar{X}$ charts generally have about similar performances in detecting the change point for small n. For example, for n = 5 and $\delta = 1$ the percentages of estimating correctly the time of a step change

Table 6. Percentages of detecting correctly the change point for the $\bar{X}$ charts and robust charts, for data without contamination and data with contamination when $n \in \{5,10,20\}$ and $\delta \in \{0.5,1,1.5,2,3\}$.

Data	n	Chart	Percentages of detecting correctly the time of a step change				
			$\delta = 0.5$	$\delta = 1$	$\delta = 1.5$	$\delta = 2$	$\delta = 3$
J_0	5	E_1	0.64%	2.36%	5.99%	15.48%	49.17%
		E_2	0.64%	2.36%	5.99%	15.49%	49.19%
		E_3	0.52%	2.20%	5.61%	14.50%	47.85%
		E_4	0.54%	2.27%	5.73%	14.88%	48.31%
	10	E_1	0.51%	2.34%	6.60%	15.65%	48.43%
		E_2	0.51%	2.34%	6.60%	15.65%	48.39%
		E_3	0.55%	2.38%	6.81%	15.89%	48.89%
		E_4	0.56%	2.39%	6.87%	15.94%	48.96%
	20	E_1	0.52%	2.18%	6.70%	16.27%	49.96%
		E_2	0.49%	2.15%	6.62%	16.08%	49.74%
		E_3	0.53%	2.18%	6.77%	16.46%	50.34%
		E_4	0.38%	1.80%	5.72%	14.16%	46.76%
J_1	5	E_1	0.43%	1.83%	5.77%	13.87%	47.31%
		E_2	0.43%	1.82%	5.77%	13.81%	47.26%
		E_3	0.47%	1.93%	5.96%	14.20%	47.96%
		E_4	0.51%	2.09%	6.23%	14.76%	49.07%
	10	E_1	0.38%	1.54%	4.97%	12.43%	43.25%
		E_2	0.36%	1.43%	4.64%	11.77%	42.07%
		E_3	0.52%	2.06%	6.33%	15.22%	47.96%
		E_4	0.54%	2.14%	6.37%	15.33%	48.21%
	20	E_1	0.37%	1.74%	5.22%	12.82%	45.11%
		E_2	0.32%	1.33%	4.40%	11.24%	42.03%
		E_3	0.52%	2.38%	6.45%	15.22%	49.78%
		E_4	0.38%	1.79%	5.35%	13.10%	45.69%
J_5	5	E_1	0.10%	0.62%	1.99%	6.62%	31.24%
		E_2	0.10%	0.62%	1.91%	6.52%	30.99%
		E_3	0.32%	1.36%	4.16%	10.61%	40.27%
		E_4	0.37%	1.61%	4.69%	11.78%	42.43%
	10	E_1	0.07%	0.39%	1.63%	5.32%	26.59%
		E_2	0.04%	0.25%	1.20%	4.17%	22.62%
		E_3	0.46%	1.49%	4.59%	12.11%	42.89%
		E_4	0.47%	1.53%	4.64%	12.34%	43.26%
	20	E_1	0.10%	0.41%	1.82%	4.82%	26.03%
		E_2	0.03%	0.12%	0.72%	2.37%	15.77%
		E_3	0.53%	1.82%	5.14%	12.60%	44.07%
		E_4	0.37%	1.40%	4.02%	10.71%	40.14%

for $\bar{X}$ charts based on standard deviation and range are 1.83% and 1.82%, respectively, where the disparity is merely 0.01%. However, the $\bar{X}$ chart based on standard deviation is gradually gaining the upper hand over the $\bar{X}$ chart based on range by demonstrating higher accuracy in terms of change point detection for large n. For 1% and 5% of contamination, the trimmed mean chart outperforms the median chart when $n \in \{5,10\}$ but it turns invalid for n = 20, where the median chart dominates. Overall, the robust charts are shown to be more effective over the $\bar{X}$ charts when contamination exists. The median chart based on MAD has the highest accuracy to detect the change point correctly for a contaminated process, particularly when a larger sample size is considered. It is advisable for the quality practitioners to adopt the trimmed mean chart based on WSD to monitor a non-normal process when a small sample size is of interest.

3.5. Empirical example

A data set that describes the filling process of dry dog food with 3% contamination is used to construct the four charts so that the robustness of the robust charts over the Shewhart charts in guarding against contaminated data can be shown. The data set has sample size n = 5 and subgroup count m = 45. The observations in the data set represent the net weights (in pounds) for 20-pound bags of dry dog food with 5 consecutive bags collected at 30-minute intervals. Both $\bar{X}$ charts are in-control though there exists 3% contamination in the process (Figure 1). However, the median chart based on MAD, and the trimmed mean chart based on WSD are able to detect the contaminated points, in which, the former chart signals at points 17 and 31 while the latter at points 7, 9, 12, 17, 21, and 31, respectively. Those out of limit points may in fact signal a change that requires rectification. This example shows that the robust charts are robust against contamination and they outperform the $\bar{X}$ charts in detecting outliers.

Both robust charts have a shorter range length as compared with the $\bar{X}$ charts (Table 7). The trimmed mean chart has the shortest length followed by the median chart and the $\bar{X}$ charts. Note that the $\bar{X}$ charts based on standard deviation and range have approximately the same lengths of range within the limits. Besides, the number of points that fall outside the control limits for the median chart and the trimmed

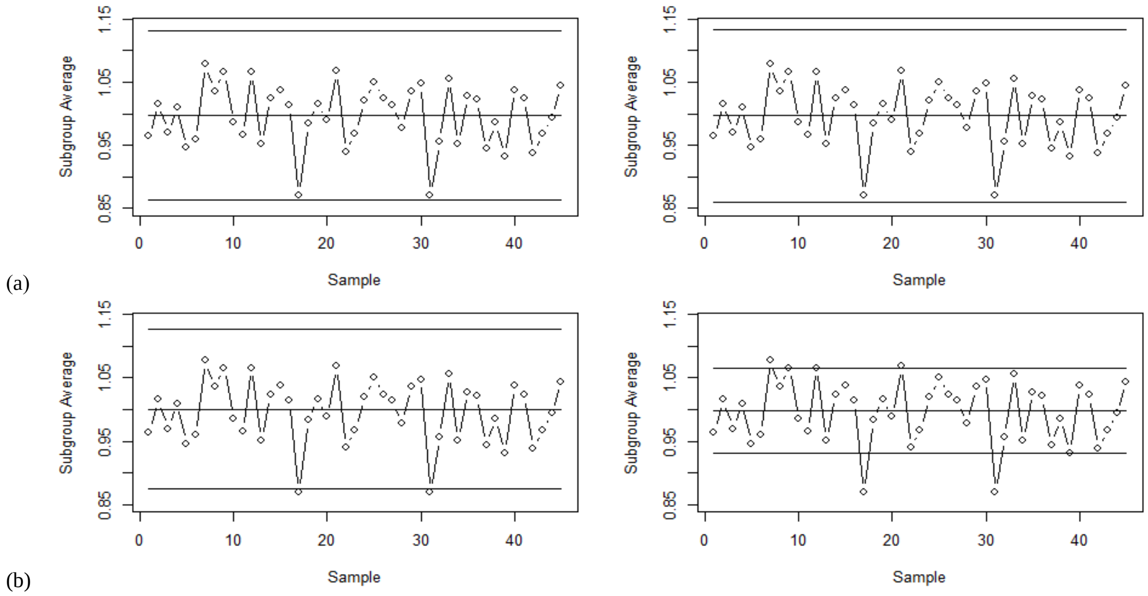

Figure 1. Plots of contaminated filling process of dry dog food for n = 5 and m = 45. (a) Constructed $\bar{X}$ charts based on standard deviation and range. (b) Constructed median chart based on MAD and trimmed mean chart based on WSD.

Table 7. Performance comparison of the $\bar{X}$ charts and the robust charts based on 3% of contamination in a filling process of dry dog food for n = 5 and m = 45.

	E_1	E_2	E_3	E_4
LCL	0.863	0.861	0.875	0.932
CL	0.997	0.997	1.001	0.999
UCL	1.131	1.133	1.127	1.066
Length of the control limits	0.268	0.272	0.252	0.134
Number of points detected by the control limits	0	0	2	6

mean chart are 2 and 6, respectively. All points are found to be inside the control limits for both $\bar{X}$ charts.

4. Conclusions

To sum up, in this study, the performances of the median chart based on MAD and trimmed mean chart based on WSD were compared to the Shewhart $\bar{X}$ charts based on standard deviation and range under normal and non-normal distributions. It was found that both robust charts yield a more favourable performance compared to the $\bar{X}$ charts when the process contains some portions of contamination. This is because the robust charts are relatively insensitive to contamination. When the distribution is normal, the robust charts and the $\bar{X}$ charts have virtually comparable performances under small n, but for large n, the median chart based on MAD prevailed. It is to be highlighted that only process shift in the mean was considered in this paper, hence other types of process shifts like a shift in the process variance or the combination of a process shift with contaminated data can be investigated in future analyses. Moreover, comparing the performances of the median chart and trimmed mean chart in combination with the trimmed range with the existing charts will also be an interesting topic.

Acknowledgements

This research was supported by a grant from the Malaysia Fundamental Research Grant Scheme (FRGS) (Grant No.: R/FRGS/A0700/01450A/001/2018/00560).

References

Abu-Shawiesh, M. O. A. (2008). A simple robust control chart based on MAD. *Journal of Mathematics and Statistics*, *4*(2), 102-107.

Bouslah, B., Gharbi, A., & Pellerin, R. (2018). Joint production, quality and maintenance control of a two-machine line subject to operation-dependent and quality-dependent failures. *International Journal of Production Economics*, *195*, 210-226.

Chakraborti, S., Eryilmaz, S., & Human, S.W. (2009). A phase II nonparametric control chart based on precedence statistics with runs-type signaling rules. *Computational Statistics & Data Analysis*, *53*(4), 1054-1065.

Chiang, J. Y., Lio, Y. L., Ng, H. K. T., & Tsai, T. R. (2018). Robust control charts for percentiles based on location-scale family of distributions. *Quality and Reliability Engineering International*, *34*(6), 1059-1078.

Davis, C. M., & Adams, B. M. (2005). Robust monitoring of contaminated data. *Journal of Quality Technology*, *37*(2), 163-174.

Figueiredo, F., & Gomes, M. I. (2009). Monitoring industrial processes with robust control charts. *REVSTAT-Statistical Journal*, *7*(2), 151-170.

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, *69*(346), 383-393.

Hawkins, D. M., & Deng, Q. (2010). A nonparametric change-point control chart. *Journal of Quality Technology*, *42*(2), 165-173.

Karagöz, D. (2016). Robust X control chart for monitoring the skewed and contaminated process. *Hacettepe Journal of Mathematics and Statistics*, *47*(1), 223-242.

Jacobs, D. C. (1990). Watch out for nonnormal distributions. *Chemical Engineering Progress*, *86*(11), 19-27.

Kocherlakota, K., & Kocherlakota, S. (1995). Control charts using robust estimators. *Total Quality Management*, *6*(1), 79-90.

Montgomery, D. C. (2012). *Statistical Quality Control: A Modern Introduction* (7th ed.). New York: John Wiley & Sons, Inc.

Langenberg, P., & Iglewicz, B. (1986). Trimmed mean $\bar{X}$ and *R* charts. *Journal of Quality Technology*, *18*(3), 152-161.

Rocke, D. M. (1989). Robust control charts. *Technometrics*, *31*(2), 173-184.

Samuel, T. R., Pignatiello, Jr. & Calvin, J. A. (1998). Identifying the time of a step change with $\bar{X}$ control charts. *Quality Engineering*, *10*(3), 521-527.

Tereza, S., & Noskievičová, D. (2017). Nontraditional methods of statistical process control. *Hutnické listy*, *2*, 43-48.

Tukey, J. W. (1948). Some elementary problems of importance to small sample practice. *Human Biology*, *20*, 205-214.

Zhou, M., & Geng, W. (2013). A robust control chart for monitoring dispersion. *Journal of Applied Mathematics*, *2013*, ID 279203.