

Integrating Mean and Median Charts for Monitoring an Outlier-Existing Process

Ling Yang*, Suzanne Pai, and Yuh-Rau Wang

Abstract—An effective control scheme can be instrumental in increasing productivity and reducing cost. While facing an outlier-existing process, using the mean (\bar{X}) control chart and the range (R) control chart for monitoring the process mean and variance will lead to high level false alarms. Recently, some median (\tilde{X}) control charts, such as the Shewhart- \tilde{X} control chart, the exponentially weighted moving average (EWMA)- \tilde{X} control chart, and the generally weighted moving average (GWMA)- \tilde{X} median control chart have been developed in succession for monitoring the process mean/median. Although those mentioned \tilde{X} control charts are demonstrated much more robust than the \bar{X} control charts with outliers, the shift-detecting ability is worse than that of \bar{X} control charts without respect to outliers. Davis and Adams proposed a synthesized control scheme (called SS synthesized control scheme for short) which using the Shewhart- \bar{X} and R (denoted as Shewhart- \bar{X}/R) charts and the Shewhart trimmed mean and trimmed range (denoted as Shewhart- \bar{X}^*/R^*) charts for monitoring the process mean and variance with outliers. This study proposes a modified synthesized control scheme which integrating the EWMA- \bar{X}/R control charts and the EWMA- \tilde{X}/R^* control charts (called EE synthesized control scheme for short) for monitoring an outlier-existing process. A diagnostic statistic technique is adopted herein to be a bridge between the \bar{X}/R charts and the \tilde{X} (or \bar{X}^*/R^*) charts. With various shifts of the process sample mean, the average time to work stoppage (ATWS) is evaluated under some contaminated normal distributions. We conclude that the EE synthesized control scheme outperforms the SS synthesized control scheme for monitoring the small shift of the process mean or variance. This result provides a valuable recommendation while facing an outlier-existing process.

Index Terms—control chart, EWMA, median, outliers.

Manuscript received November 24, 2007. This work was supported in part by the National Science Council, Taiwan, ROC, under Grants NSC96-2221-E-129-008.

Ling Yang is with the Department of Industrial Engineering and Management, St. John's University, Tamsui, Taipei 251, Taiwan, ROC (corresponding author, phone: +886-2-28013131 ext. 6583; fax: +886-2-28013131 ext. 6589; e-mail: lgyang@mail.sju.edu.tw).

Suzanne Pai is with the Department of Industrial Engineering and Management, St. John's University, Tamsui, Taipei, Taiwan, ROC (e-mail: sjpai@mail.sju.edu.tw).

Yuh-Rau Wang is with the Department of Computer Science and Information Engineering, St. John's University, Tamsui, Taipei, Taiwan, ROC (e-mail: yrwang@mail.sju.edu.tw).

I. INTRODUCTION

An effective control scheme can be instrumental in increasing productivity and reducing cost. From past experience, there are some certain processes with outliers that happened occasionally. The outliers are the values of observations that are larger or smaller than the majority of the other observations even though the process is in control. Samples containing outliers are said to be contaminated. Using the mean (\bar{X}) control chart and the range (R) control chart for monitoring the process mean and variance will lead to high level false alarms. Recently, some median (\tilde{X}) control charts, such as the EWMA- \tilde{X} control chart (Castagliola, [2]), the Shewhart- \tilde{X} control chart (Khoo, [6]), and the generally weighted moving average (GWMA)- \tilde{X} control chart (Sheu and Yang, [10]), had been developed in succession. As discussed in [10], the \tilde{X} control charts are much more robust than the \bar{X} control charts with outliers-existing process. However the shift-detecting ability of \tilde{X} control charts is worse than that of \bar{X} control charts without respect to outliers.

Suppose that the occurrence of outliers is due to the common causes and only the assignable causes will lead to permanent shifts. Davis and Adams [5] proposed a synthesized control scheme for the outlier-existing process. In [5], the Shewhart- \bar{X} control chart and the R control chart (denoted as Shewhart- \bar{X}/R) are used for monitoring the process mean and variance first. While the Shewhart- \bar{X} control chart or the R control chart signals, the diagnostic statistic (DS) is compared with the pre-determined decision value (K) to ascertain whether the collected sample is contaminated or not. If it is regarded as a contaminated sample, the Shewhart trimmed mean (\bar{X}^*) control chart and the trimmed range (R^*) control chart (Langenberg and Iglewicz, [7]) (denoted as Shewhart- \bar{X}^*/R^*) are used to carry on the process monitoring. Davis and Adams [5] indeed provided a good choice for the practitioner. However, the fast-detecting technique, e.g. the EWMA- \bar{X} control chart, and those mentioned \tilde{X} control charts were not considered in their study.

In this work, a modified synthesized control system is proposed. The EWMA- \bar{X} control chart and the R control chart (denoted as EWMA- \bar{X}/R) are adopted in the first stage. The EWMA- \tilde{X} control chart and the R^* control chart (denoted as EWMA- \tilde{X}/R^*) are adopted in the second stage.

The corresponding K values are provided. With various shifts of the process mean, the average time to work stoppage (ATWS) are evaluated under some contaminated normal distributions.

II. DESCRIPTION OF SOME CONTROL CHARTS

Suppose that the quality characteristic is a variable and the samples have been collected at each point in time (the size of rational subgroups, denoted as n). Let $\bar{X}_i, \tilde{X}_i, \bar{X}_i^*, R_i,$ and R_i^* be the sample average, sample median, trimmed sample mean, sample range, and trimmed sample range of i th subgroup respectively, which are composed of n independent normal (μ_i, σ_i^2) random variables $X_{i,1}, \dots, X_{i,n}$, where μ_i is the nominal process mean/median and σ_i^2 is the nominal process variance. When the process is in-control, $\mu_i = \mu_0,$ $\sigma_i^2 = \sigma_0^2$ (the target value of the process mean and variance). That is

$$\begin{aligned} \bar{X}_i &= \sum_{j=1}^n X_{i,j} / n, \\ \tilde{X}_i &= X_{[i,(n+1)/2]}, \quad (\text{if } n \text{ is odd}), \\ \bar{X}_i^* &= \sum_{j=1}^{n-1} X_{[i,j]} / (n-1), \\ R_i &= X_{[i,n]} - X_{[i,1]}, \\ R_i^* &= X_{[i,n-1]} - X_{[i,1]}, \end{aligned}$$

where $X_{[i,j]}$ represents the j th order statistic for the i th sample.

In [3], the EWMA- \bar{X} control statistic, Y_i , can be represented as

$$Y_i = \alpha \bar{X}_i + (1-\alpha)Y_{i-1}, \quad \text{for } i=1, 2, \dots \quad (1)$$

where $Y_0 = \mu_0$, the smooth parameter α ($\alpha > 0$) is determined by the practitioner. The central line (CL), the time-varying upper control limit (UCL) and lower control limit (LCL) of the EWMA- \bar{X} control chart can be represented as

$$\begin{aligned} UCL &= \mu_0 + L \sqrt{\frac{\alpha(1-(1-\alpha)^{2i})}{2-\alpha}} \frac{\sigma_0}{\sqrt{n}} \\ CL &= \mu_0 \\ LCL &= \mu_0 - L \sqrt{\frac{\alpha(1-(1-\alpha)^{2i})}{2-\alpha}} \frac{\sigma_0}{\sqrt{n}}, \end{aligned} \quad (2)$$

where L determines the width of the control limits. The process is considered out of control and some actions should be taken whenever Y_i falls outside the range of the control

limits. When $\alpha = 1$, the plotted statistic in (1) will be $Y_i = \bar{X}_i$, and the control limits will be

$$\begin{aligned} UCL &= \mu_0 + L \frac{\sigma_0}{\sqrt{n}} \\ CL &= \mu_0 \\ LCL &= \mu_0 - L \frac{\sigma_0}{\sqrt{n}}. \end{aligned}$$

These are the equations of the Shewhart- \bar{X} control chart. Therefore, when $\alpha = 1$, the EWMA- \bar{X} control chart reduces to the Shewhart- \bar{X} control chart.

According to Castagliola [1], the distribution of the sample median \tilde{X}_i is very close to the $(\mu_i, \tilde{\sigma}_i^2)$ normal distribution, where $\tilde{\sigma}_i^2$ is the variance of \tilde{X}_i . If $\tilde{\sigma}_{0,1}$ is the standard deviation of normal (0,1) sample median, we have $\tilde{\sigma}_i = \sigma_i \times \tilde{\sigma}_{0,1}$. The values of $\tilde{\sigma}_{0,1}$ had been derived by [1]. When the process is in-control, $\mu_i = \mu_0,$ $\sigma_i^2 = \sigma_0^2$. The EWMA- \tilde{X} control statistic, Z_i , can be represented as

$$Z_i = \beta \tilde{X}_i + (1-\beta)Z_{i-1}, \quad \text{for } i=1, 2, \dots, \quad (3)$$

where $Z_0 = \mu_0$, the smooth parameter β ($\beta > 0$) is determined by the practitioner. The time-varying control limits of the EWMA- \tilde{X} control chart can be represented as

$$\begin{aligned} UCL &= \mu_0 + \eta \sqrt{\frac{\beta(1-(1-\beta)^{2i})}{2-\beta}} \times \sigma_0 \tilde{\sigma}_{0,1} \\ CL &= \mu_0 \\ LCL &= \mu_0 - \eta \sqrt{\frac{\beta(1-(1-\beta)^{2i})}{2-\beta}} \times \sigma_0 \tilde{\sigma}_{0,1}, \end{aligned} \quad (4)$$

where η determines the width of the control limits. The process is considered out of control and some actions should be taken whenever Z_i falls outside the range of the control limits. When $\beta = 1$, the plotted statistic in (3) will be $Z_i = \tilde{X}_i$, and the control limits will be

$$\begin{aligned} UCL &= \mu_0 + \eta \sigma_0 \tilde{\sigma}_{0,1} \\ CL &= \mu_0 \\ LCL &= \mu_0 - \eta \sigma_0 \tilde{\sigma}_{0,1}. \end{aligned}$$

These are the equations of the Shewhart- \tilde{X} control chart. Therefore, when $\beta = 1$, the EWMA- \tilde{X} control chart reduces to the Shewhart- \tilde{X} control chart.

The control scheme of the R control chart refers to [8] for

details. The control schemes of the trimmed mean (\bar{X}^*) and trimmed range (R^*) control charts refer to [7].

III. MODIFIED SYNTHESIZED CONTROL SCHEME

In Davis and Adams' synthesized control scheme [5], several diagnostic statistics (DS) were used to recognize whether the collected sample contained contaminated data. Through simulation, the maximum-median DS , combined with Shewhart- \bar{X}/R control charts and Shewhart- \bar{X}^*/R^* control charts (called SS synthesized control scheme for short), performed best. The definition of DS_i is

$$DS_i = X_{[i,n]} - X_{[i,(n+1)/2]}, \text{ for } i = 1, 2, \dots \quad (5)$$

The Shewhart- \bar{X}^*/R^* control charts were used only when the Shewhart- \bar{X} control chart and/or R control chart signaled and the value of DS figured out that the sample contains outliers (i.e. $DS_i > K$). As mentioned in the first section, Davis and Adams [5] indeed provided a good choice for the practitioner with outliers consideration. However, the fast-detecting technique (e.g. the EWMA- \bar{X} control chart) and the robust control technique (e.g. the EWMA- \tilde{X} control chart) were not considered in their study.

In this work, a modified synthesized control scheme is proposed. An EWMA- \bar{X} control chart and a R control chart (denoted as EWMA- \bar{X}/R) are used in the first stage. An EWMA- \tilde{X} control chart and a R^* control chart (denoted as EWMA- \tilde{X}/R^*) are used in the second stage. The proposed control scheme is called EE synthesized control scheme for short. The similar computation of DS (as shown in (5)) is used. Because the EWMA- \bar{X} control chart and the EWMA- \tilde{X} control chart need to collect successive data, a modified synthesized control scheme is shown in Fig. 1. In the first stage, if the EWMA- \bar{X}/R control chart(s) signal(s) and $DS_i \leq K$ (i.e. the sample does not contain outliers), the process is assumed to be out of control and a search for an assignable cause is initiated. If the EWMA- \bar{X}/R control chart(s) signal(s) and $DS_i > K$ (i.e. the sample contains outliers), the second stage will be started. In the second stage, if the EWMA- \tilde{X}/R^* control chart(s) signal(s), the process is assumed to be out of control and a search for an assignable cause is initiated.

As mentioned in [5], the diagnostic statistic technique uses a decision value (K) from the conditional distribution of DS given a signal of the first stage control chart, i.e.

$$P(DS_i > K \mid \text{first stage control charts signal(s)}) \leq 0.0027$$

Without loss of generality, we assume that the in-control $\mu_0 = 0$ and variance $\sigma_0^2 = 1$. Due to the complex region of integration and the accuracy of simulation, [5] elected to use the simulated decision values [4] (as listed in the first row of Table 1, for $n = 5$). In this study, we provide the decision values for the EE synthesized control scheme via simulation too (as listed in the second row of Table 1, for $n = 5$).

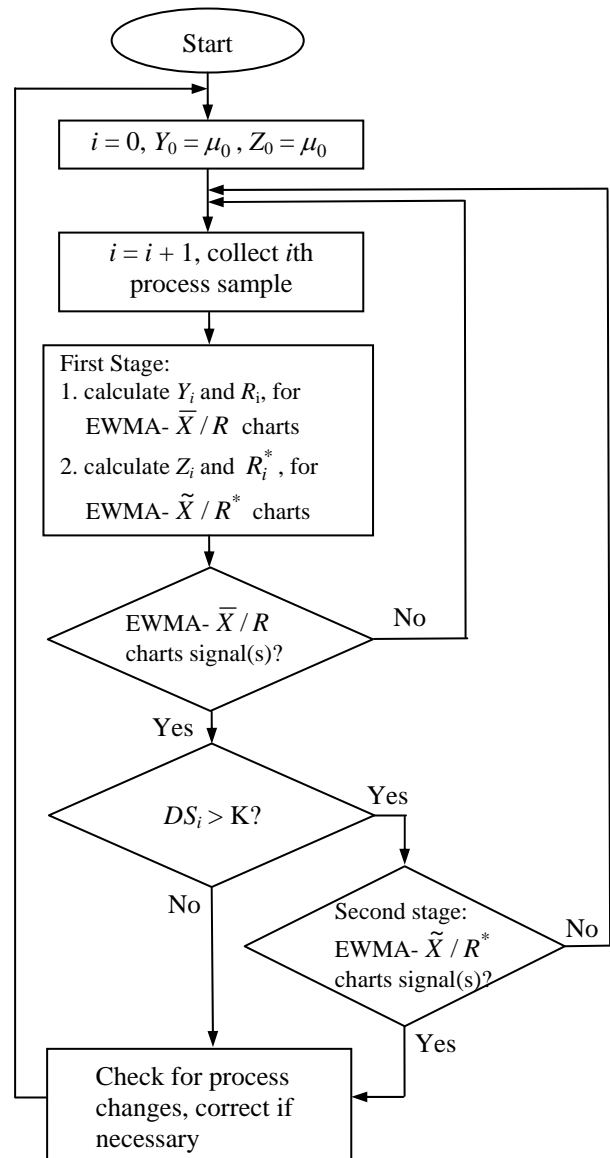


Fig. 1. EE synthesized control scheme

Table 1. Decision values

Synthesized control scheme	Control limits	Decision value (K) ($n = 5$)
SS	Shewhart- \bar{X} chart: (-1.3416, 1.3416) R chart: (0, 4.918)	2.059
EE	EWMA- \bar{X} chart: $\alpha = 0.1, L = 2.703$ R chart: (0, 4.918)	1.465

IV. PERFORMANCE MEASUREMENT AND COMPARISON

In [5], the average time to work stoppage (*ATWS*) is used for the performance measurement of the synthesized control scheme. *ATWS* is defined as the average number of points plotted before a signal of second stage control chart. In this work, the simulation [9] is used to estimate the *ATWS* of the synthesized control schemes. The simulation program is written in BASIC language. Analysis begins with 20,000 EWMA (or Shewhart)- \bar{X}/R control charts signal(s) from subgroup of samples size $n = 5$.

In order to evaluate the *ATWS* of the synthesized control schemes in the presence of outliers, the contaminated normal distribution used in [6] is adopted. A contaminated normal distribution is that the observations $(100 - \theta)\%$ come from $(N(0, 1))$ normal distribution and $\theta\%$ come from $(N(C, 1))$ normal distribution, where θ denotes the level of contamination, C denotes the mean of an outlier. We assumed that the outliers occur due to the common causes of variation and lead to a temporary shift. Only the assignable causes will make the process a permanent shift. We are interested in detecting a permanent shift. Three kind of contaminated normal distributions $(C, \theta) \in \{(1, 1), (2, 1), (3, 1)\}$ are used herein to evaluate the *ATWS* of the synthesized control schemes. The control limits for the synthesized control schemes are based on the data which 100% come from the $(N(0, 1))$ normal distribution (i.e. $(C, \theta) = (0, 0)$) with a desired in-control *ATWS* (denoted $ATWS_0 \cong 139$). This is the in-control average run length (ARL_0) of the Shewhart- \bar{X}/R charts and the EWMA- \bar{X}/R charts in the first stage. These two synthesized control schemes are standardized to this ARL_0 for fair comparison. Table 2 presents the simulation result of $ATWS_0$. Tables 3 and 4 present the simulation result of out-of-control *ATWS* (denoted $ATWS_1$) without and with contaminated data, respectively.

In Tables 2–4, various combinations of (C, θ) denote various contaminated normal distributions. For example, when $(C, \theta) = (0, 0)$, the data 100% come from the normal distribution $N(0, 1)$. When $(C, \theta) = (2, 1)$, the data 99% come from the normal distribution $N(0, 1)$ and 1% come from the normal distribution $N(2, 1)$. In Table 2, when the process is in-control (i.e. $(\mu, \sigma^2) = (1, 0)$), $(C, \theta) \in \{(0, 0), (1, 1), (2, 1), (3, 1)\}$, the $ATWS_0$ of the EE synthesized control scheme are similar to that of the SS synthesized control scheme. It means that both the EE synthesized control scheme and the SS synthesized control scheme are outliers-resistant. However, when the mean shift is small, the EE synthesized control scheme outperforms the SS synthesized control scheme for both the data is contaminated or not. For instance, in Table 3, when $(\mu, \sigma^2) = (0.3, 1)$ and $(C, \theta) = (0, 0)$, the $ATWS_1$ of the EE synthesized control scheme (= 15.33) is less than that of the SS synthesized control scheme (= 58.19). In Table 4, when $(\mu, \sigma^2) = (0.3, 1)$ and $(C, \theta) = (2, 1)$, the $ATWS_1$ of the EE synthesized control scheme (= 13.18) is less than that of the SS synthesized control scheme (= 49.15). It means that

the EE synthesized control scheme can detect the process small shift more quickly than the SS synthesized control scheme.

V. CONCLUSION

This study proposes a modified synthesized control scheme to retain the fast-shift-detecting ability (the EWMA- \bar{X} control chart) and the robustness to outliers (the EWMA- \tilde{X} control chart) simultaneously. In the EE synthesized control scheme, the diagnostic statistic technique is adopted and the decision values are provided. Under several contaminated normal distributions, the performance of $ATWS_1$ of the EE synthesized control scheme outperforms that of the SS synthesized control scheme with various shifts of the process sample mean and variance. We conclude that the modified synthesized control scheme which combined the EWMA- \bar{X} control chart with the EWMA- \tilde{X} control chart is outliers-resistant and more sensitive in monitoring the small shift of the process mean. This result provides a useful recommendation while facing an outlier-existing process.

Table 2. $ATWS_0$ of the synthesized control schemes with a desired $ATWS_0 \cong 139$

In-control $(\mu, \sigma^2) = (0, 1)$	(C, θ)	SS	EE
no contaminated data	(0, 0)	139.3	139.1
contaminated data	(1, 1)	133.9	134.1
	(2, 1)	121.8	121.0
	(3, 1)	114.3	100.5

Table 3. $ATWS_1$ for no contaminated data with a desired $ATWS_0 \cong 139$

Out-of-control, no contaminated data $(C, \theta) = (0, 0)$	(μ, σ^2)	SS	EE
mean shift	(0.3, 1)	58.19	15.33
	(0.5, 1)	22.83	6.44
	(1.0, 1)	3.89	2.19
	(1.5, 1)	1.53	1.05
	(2.0, 1)	1.10	1.04
variance shift	(0, 1.5)	25.67	24.86
	(0, 2.0)	10.58	9.95
	(0, 2.5)	6.19	5.42
	(0, 3.0)	4.30	3.70

Table 4. $ATWS_1$ for contaminated data with a desired
 $ATWS_0 \cong 139$

Out-of-control, contaminated data $(C, \theta) = (2, 1)$	(μ, σ^2)	SS	EE
mean shift	(0.3, 1)	49.15	13.18
	(0.5, 1)	20.71	5.51
	(1.0, 1)	3.77	2.02
	(1.5, 1)	1.53	1.04
	(2.0, 1)	1.10	1.04
variance shift	(0, 1.5)	24.38	23.48
	(0, 2.0)	10.43	9.22
	(0, 2.5)	6.15	5.21
	(0, 3.0)	4.30	3.71

REFERENCES

- [1] P. Castagliola, "Approximation of the normal sample median distribution using symmetrical Johnson S_U distributions: application to quality control," *Communications in Statistics: Simulation and Computation*, vol. 27, 1998, pp. 289–301.
- [2] P. Castagliola, "An (\bar{X}/R) -EWMA control chart for monitoring the process sample median," *International Journal of Reliability, Quality and Safety Engineering*, vol. 8(2), 2001, pp. 123–135.
- [3] S. V. Crowder, "Design of exponentially weighted moving average schemes," *Journal of Quality Technology*, vol. 21, 1989, pp. 155–162.
- [4] C. M. Davis, *A Diagnostic Tool for Traditional Mean and Range Control Charts*. Ph.D. Dissertation, Department of Information Systems, Statistics, and Management Science, University of Alabama, Tuscaloosa, AL, USA, 2002.
- [5] C. M. Davis and B. M. Adams, "Robust monitoring of contaminated data," *Journal of Quality Technology*, vol. 37(2), 2005, pp. 163–174.
- [6] M. B. C. Khoo, "A Control chart based on sample median for the detection of a permanent shift in the process mean," *Quality Engineering*, vol. 17, 2005, pp. 243–257.
- [7] P. Langenberg and B. Iglewicz, "Trimmed mean \bar{X} and R charts," *Journal of Quality Technology*, vol. 18, 1986, pp. 152–161.
- [8] D. C. Montgomery, *Introduction to Statistical Quality Control*. New York: Wiley, USA, 2005.
- [9] S. M. Ross, *A Course in Simulations*. New York: Macmillan, USA, 1990.
- [10] S. H. Sheu and L. Yang, "The generally weighted moving average median control chart," *Quality Technology and Quantitative Management*, vol. 3(4), 2006, pp. 455–471.