Research Article

Quality and Reliability Engineering International

(wileyonlinelibrary.com) DOI: 10.1002/qre.1655

Published online 10 June 2014 in Wiley Online Library

A Robust Phase I Exponentially Weighted Moving Average Chart for Dispersion

Inez M. Zwetsloot[†], **Marit Schoonhoven**[‡*†] **and Ronald J. M. M. Does**[¶]

A Phase I estimator of the dispersion should be efficient under in-control data and robust against contaminations. Most estimation methods proposed in the literature are either efficient or robust against either sustained shifts or scattered disturbances. In this article, we propose a new estimation method of the dispersion parameter, based on exponentially weighted moving average charting, which is efficient and robust to both types of unacceptable observations in Phase I. We compare the method with various existing estimation methods and show that the proposed method has the best overall performance if it is unknown what type of contaminations are present in Phase I. We also study the effect of the robust estimator from Phase I on the Phase II exponentially weighted moving average control chart performance.

Keywords: changepoint; contaminations; outliers; standard deviation; statistical process control

1. Introduction

Control charts are designed as monitoring tools with the aim of detecting any change in the process characteristic as soon as possible. Implementation of a control chart usually consists of two phases. The purpose of Phase I is to define the 'in-control' state of the process and to estimate the process parameters. These estimates are used to set up the control limits for Phase II, the prospective monitoring stage. The Phase I data set can contain unusual observations, which are problematic as they can influence the parameter estimates, resulting in Phase II control charts with less ability to detect changes in the process characteristic. In our paper, we focus on monitoring the process dispersion and consider the problem of estimating the dispersion parameter when the Phase I data may contain contaminated samples.

One approach to deal with contaminations in Phase I is to use robust point estimators (Jensen *et al.* [1]). There exists a long tradition of using robust estimators in statistical process control. Some references are Rocke,[2] Tatum,[3] and Psarakis *et al.*.[4] Unfortunately, robust estimators are usually not very efficient under in-control Phase I data.

A second approach is to construct Phase I charts which identify potentially contaminated samples, remove these from Phase I, and use the remaining samples to estimate the process parameters. An overview of Phase I charts for univariate data was given by Chakraborti *et al.*[5] and Jones-Farmer *et al.*[6] In addition, Schoonhoven *et al.*[7] and Schoonhoven and Does[8] studied the Shewhart Phase I chart to obtain a robust estimator for the dispersion. Jones-Farmer and Champ[9] and Capizzi and Masarotto[10] studied non-parametric Phase I charts for robustness against disturbances in Phase I.

A third approach is to use changepoint methods. These methods are especially suited to detecting sustained changes in the process parameters. There is a long tradition of testing for sustained shifts in Phase I; for a literature overview see Amiri and Allahyari.[11]

The optimal choice of estimation method requires knowledge of the type of contaminations. Typically, Phase I charts are suitable when outliers are present in Phase I and changepoint methods are suitable if sustained shifts occur. The aim of our paper is to introduce a new Phase I estimation methodology for the dispersion parameter which provides reliable estimates regardless of the pattern of contaminations in Phase I. This is achieved by using an exponentially weighted moving average (EWMA) chart in Phase I. The EWMA control chart is described in Section 2. The new estimation technique and some competing methods mentioned above are described in Section 3. Section 4 compares their performance in terms of efficiency for uncontaminated and various contaminated data sets. In Section 5, we evaluate the effect of the proposed estimators on the Phase II EWMA control chart and Section 6 offers some concluding comments.

Department of Operations Management, Institute for Business and Industrial Statistics, University of Amsterdam, Plantage Muidergracht 12, 1018 TV Amsterdam The Netherlands

**Correspondence to: Marit Schoonhoven, Department of Operations Management, Institute for Business and Industrial Statistics, University of Amsterdam, Plantage Muidergracht 12, 1018 TV, Amsterdam, The Netherlands.*

[†] *E-mail: m.schoonhoven@uva.nl*

[‡] *E-mail: i.m.zwetsloot@uva.nl*

[¶] *E-mail: r.j.m.m.does@uva.nl*

2. The EWMA dispersion control chart

The EWMA control chart, originally proposed by Roberts,[12] can detect small shifts in the process parameters in Phase II more quickly than the traditional Shewhart control chart, by taking into account the information of the current samples as well as the history of the process. The EWMA control chart for location has been thoroughly studied in the literature, some references are Crowder,[13] Lucas and Saccucci,[14] and Jones *et al.*[15] In addition to location EWMA control charts, considerable research has been done to develop EWMA control charts for monitoring process dispersion. Two categories of EWMA dispersion control chart can be distinguished. EWMA control charts in the first category are based on a dispersion statistic, like the sample variance S^2 or sample standard deviation S (e.g. Ng and Case[16]). EWMA control charts in the second category, are based on transformations of these statistics, such as $\log(S^2)$, to correct for the skewness of S and S^2 (e.g. Shu and Jiang[17]).

Knoth[18] studied these competing statistics and compared EWMA dispersion control charts based on R (the range), S^2, S, and $\log(S^2)$ and concluded that "the best performance in terms of the average run length profile is given by the S^2 and S EWMA control charts". In our paper, we therefore consider an EWMA dispersion control chart based on S. We use S rather than S^2 because in practice the process dispersion is most often evaluated in terms of S.

The EWMA statistic is defined as $W_t = (1-\lambda)W_{t-1} + \lambda S_t$, with S_t the sample standard deviation of sample t, and λ a smoothing constant satisfying $0 < \lambda \leq 1$. We set $W_0 = E[S_t] = c_4(n)\sigma$, where n is the sample size and $c_4(n)$ the bias correcting coefficient

$$c_4(n) = \sqrt{\frac{2}{n-1}}\frac{\Gamma(n/2)}{\Gamma((n-1)/2)}.$$

Under the assumption of independently, identically, and normally distributed observations, the mean and variance of W_t are $E[W_t] = c_4(n)\sigma$ and $V[W_t] = \sigma^2\left(1-c_4(n)^2\right)\frac{\lambda}{2-\lambda}\left(1-(1-\lambda)^{2t}\right)$. When monitoring dispersion, an increase in the dispersion indicates some special cause of variation that should be detected and removed, while a decrease in dispersion indicates a process improvement. As we are most interested in detecting increases, we use a one-sided EWMA control chart. Therefor, we reset the EWMA statistic to its expected value whenever it drops below

$$W_t = \max[\ (1-\lambda)W_{t-1} + \lambda S_t\ ,\ c_4(n)\sigma\ ].$$

The EWMA control chart gives an out-of-control signal whenever W_t exceeds the upper control limit UCL_t, where

$$UCL_t = c_4(n)\sigma + L\sigma\sqrt{(1-c_4(n)^2)}\sqrt{\frac{\lambda}{2-\lambda}}\sqrt{1-(1-\lambda)^{2t}}.$$

Here, L is a positive coefficient which, together with λ, determines the in-control performance of the EWMA dispersion control chart. We use the so called time-varying control limits to enhance the charts sensitivity to shifts in the first samples (cf. Steiner[19]).

In practice, σ is unknown and has to be estimated. We denote the estimate of σ, which is based on the Phase I data, by $\hat{\sigma}$. Many authors have studied the design of control charts based on estimated parameters, see for example Psarakis *et al.*[4] and Saleh *et al.*[20] For the effect of estimation on the EWMA control chart see Jones *et al.*[15] and Jones.[21]

3. Phase I estimators of dispersion

Below, we describe various estimation methods that can be used within Phase I of the control charting process to obtain an estimate of σ. We consider efficient and robust point estimators, a changepoint method and we present the new estimation method based on EWMA Phase I control charting.

Let X_{it}, with $i = 1, 2, ..., n$ and $t = 1, 2, ..., k$, be the Phase I observations of the process characteristic. Assume X_{it} to be independent $N(\mu, \sigma^2)$ distributed with mean μ and standard deviation σ assumed stable if the process is in control. We set $k = 50$ and $n = 5$.

3.1. Point estimators

Traditionally, σ is estimated with the pooled sample standard deviation:

$$S_p = \frac{1}{c_4(k(n-1)+1)}\sqrt{\frac{1}{k}\sum_{t=1}^{k} S_t^2}.$$

Mahmoud *et al.*[22] showed that this estimator is more efficient than the mean of the sample standard deviations or the mean of the sample ranges when data are normally distributed. The pooled sample standard deviation provides a basis for comparison as it is the most efficient unbiased estimator under uncontaminated normal data.

Next, we include two robust point estimators in our comparison. The first is based on the sample interquartile ranges, which are defined as $IQR_t = X_{(b)t} - X_{(a)t}$, where $X_{(o)t}$ denotes the o-ordered value in sample t, $a = \lceil n/4 \rceil$, and $b = n - a + 1$. The ceiling function $\lceil z \rceil$ denotes the smallest integer not less than z. We consider a trimmed version of the mean of the sample IQRs as proposed by Rocke[2];

$$\overline{IQR}_\alpha = \frac{1}{k - 2\lceil k\alpha \rceil} \sum_{o=\lceil k\alpha \rceil + 1}^{k - \lceil k\alpha \rceil} IQR_{(o)},$$

where $IQR_{(o)}$ denotes the o-th ordered value of the samples IQRs. We take $\alpha = 20\%$. An unbiased estimate of σ is given by dividing $\overline{IQR}_{20}$ by 0.9261 (obtained through 100,000 Monte Carlo simulations).

The second robust point estimator we consider is an estimator proposed by Tatum.[3] It is based on a variant of the biweight A estimator. First, it centers each observation on its sample median M_t, creating residuals $e_{it} = X_{it} - M_t$. If n is odd, each sample contains one residual equal to zero, which is dropped. The resulting $n'k$ residuals, with $n' = n - 1$ when n is odd and $n' = n$ if n is even, are weighted by $u_{it} = \frac{h_t e_{it}}{cM^*}$, where M^* is the median of the absolute values of the $n'k$ residuals,

$$h_t = \begin{cases} 1 & E_t \le 4.5, \\ E_t - 3.5 & 4.5 < E_t \le 7.5, \\ c & E_t > 7.5, \end{cases}$$

$E_t = IQR_t / M^*$, and c is a tuning constant. To estimate σ, only the residuals that are small, i.e. for which $|u_{it}| \le 1$, are used

$$S^* = \frac{n'k}{\sqrt{n'k - 1}} \frac{\sqrt{\sum_{t=1}^{k} \sum_{i:|u_{it}|<1} e_{it}^2 (1 - u_{it}^2)^4}}{|\sum_{t=1}^{k} \sum_{i:|u_{it}|<1} (1 - u_{it}^2)(1 - 5u_{it}^2)|}.$$

Tatum[3] showed that for $c = 7$ the estimator is robust against various contaminations. An unbiased estimator of σ is given by $S^*/d(n, k, c)$, where $d(5, 50, 7) = 1.0677$ (obtained through 100,000 Monte Carlo simulations). The resulting estimator is denoted by $D7$ as in Tatum.[3]

3.2. *Changepoint method*

Changepoint methods are based on the log likelihood of the observations in Phase I. Sullivan and Woodall[23] showed that the change-point method outperforms the Shewhart chart in detecting sustained shifts in Phase I. We apply a modified version of the estimator they proposed.

Let $\hat{\sigma}_{j:l}^2$ be the maximum likelihood estimator of the variance of samples j through l

$$\hat{\sigma}_{j:l}^2 = \frac{1}{n(l - j + 1)} \sum_{t=j}^{l} \sum_{i=1}^{n} (X_{it} - \overline{\overline{X}}_{j:l})^2,$$

where $\overline{\overline{X}}_{j:l}$ is the overall mean of all observations in samples j through l. To test for the existence of a sustained shift at sample τ, Sullivan and Woodall[23] computed the likelihood ratio statistic as $LRT[\tau] = nk \ln\left[\hat{\sigma}_{1:k}^2\right] - n\tau \ln\left[\hat{\sigma}_{1:\tau}^2\right] - n(k - \tau) \ln\left[\hat{\sigma}_{\tau+1:k}^2\right]$. Because the expected value of $LRT(\tau)$ varies with τ, we first standardize $LRT(\tau)$ by the expected value under normally distributed data $E(LRT(\tau))$. These expected values were determined through 100,000 simulations and are presented in Table I. The standardized values are denoted by $LRT'(\tau)$. A chart can be constructed by plotting $LRT'(\tau)$ versus τ and an out-of-control signal occurs if $LRT'(\tau)$ exceeds the upper control limit UCL_{CP}. Every out-of-control signal indicates a possible sustained shift in the process, i.e. the process parameters in samples $1, .., \tau$ are different from samples $\tau + 1, .., k$. When multiple signals are given, we set the estimated changepoint ($\hat{\tau}$) equal to the τ for which $LRT'(\tau)$ is largest. If there is no out-of-control signal we set $\hat{\tau} = k$.

To determine UCL_{CP}, we have set the overall in-control false alarm rate - the number of observations deemed out of control when all data are in fact in control - at 1 percent. Using 100,000 simulations, we found that $UCL_{CP} = 5.92$.

When the changepoint $\hat{\tau}$ is estimated, we can determine which samples are out of control. In practice, knowledge of the process would be used to determine whether the data before or after the estimated $\hat{\tau}$ are in control. In our paper, we use the following decision rule as a surrogate for process knowledge: 'the majority of the samples represent the in-control process.' This implies that if $\hat{\tau} \le k/2$ we delete samples 1 up to $\hat{\tau}$ from Phase I. If $\hat{\tau} > k/2$ we delete samples $\hat{\tau} + 1$ up to k from Phase I. The remaining samples are used to compute the pooled standard deviation, yielding an estimator of σ based on changepoint analysis, which we denote by CP. This

Table I. $E(LRT(\tau))$ for different values of τ

τ	2	3	4	5	6	7	8	9-10	11-17	18-32
$E(LRT(\tau))$	2.21	2.14	2.10	2.08	2.07	2.06	2.05	2.04	2.03	2.02
τ	33-39	40-41	42	43	44	45	46	47	48	
$E(LRT(\tau))$	2.03	2.04	2.05	2.06	2.07	2.08	2.10	2.13	2.21	

Table II. Phase I dispersion estimators based on EWMA charts

Estimator	Description	L_I
$sS_{0.5}$	Phase I EWMA screening estimator with $\hat{\sigma}_I = S_p$ and $\lambda_I = 0.5$	2.553
$s\overline{IQR}_{0.3}$	Phase I EWMA screening estimator with $\hat{\sigma}_I = \overline{IQR}_{20}$ and $\lambda_I = 0.3$	2.970
$s\overline{IQR}_{0.5}$	Phase I EWMA screening estimator with $\hat{\sigma}_I = \overline{IQR}_{20}$ and $\lambda_I = 0.5$	2.900
$s\overline{IQR}_{1}$	Phase I EWMA screening estimator with $\hat{\sigma}_I = \overline{IQR}_{20}$ and $\lambda_I = 1$	2.755

changepoint method is designed to detect a single changepoint $\hat{\tau}$ and is at a disadvantage if multiple step changes occur in Phase I. Alternative changepoint methods can be designed based on recursive testing for step changes.

3.3. The proposed estimation method

In this section, we propose an estimation method for the dispersion based on EWMA charting in Phase I. The new method provides a robust estimate of the dispersion when it is unknown what type of contaminations are present in Phase I. The EWMA chart can be viewed as a compromise between the Shewhart chart and methods with a memory like the CUSUM chart and the changepoint method. The proposed estimation method consists of the following steps:

1. *Use all observations in Phase I and compute an initial (robust) estimate of the dispersion*. This estimate is denoted by $\hat{\sigma}_I$. Note that the subscript '*I*' denotes that the parameter is associated with Phase I charting.
2. *Set up a Phase I EWMA chart,* using $\hat{\sigma}_I$. This chart plots the EWMA statistic $W_t = \max[(1-\lambda_I)W_{t-1} + \lambda_I S_t\ ,\ c_4(n)\hat{\sigma}_I]$ together with the upper control limit $\widehat{UCL}_t = c_4(n)\hat{\sigma}_I + L_I\hat{\sigma}_I\sqrt{(1-c_4(n)^2)}\sqrt{\frac{\lambda_I}{2-\lambda_I}}\sqrt{1-(1-\lambda_I)^{2t}}$, and with $W_0 = c_4(n)\hat{\sigma}_I$.
3. *Delete from Phase I all samples for which the corresponding EWMA statistic gives an out-of-control signal*.
4. *Compute an efficient unbiased estimator of the standard deviation, S_p, based on the remaining samples.*

The resulting estimator is denoted by $s\hat{\sigma}_{I,\lambda_I}$, where '*s*' indicates that we use a screening Phase I chart, $\hat{\sigma}_I$ stands for the initial dispersion estimator chosen in step 1 and the subscript λ_I denotes the value selected for the smoothing constant in step 2. To operationalize this screening estimator, we need to select an estimator for $\hat{\sigma}_I$ and values for λ_I and L_I.

In step 1, we select the efficient estimator S_p and the robust estimator $\overline{IQR}_{20}$ as initial estimator. This provides a comparison with an efficient estimator when no contaminations are present and with a robust estimator when contaminations occur.

In step 2, small values of λ_I enable quick detection of sustained shifts, because we use the memory property of the EWMA chart, while larger values of λ_I enables quick detection of outliers more effectively. To assess the trade-off between high and low values for λ_I, we set λ_I equal to 0.3, 0.5 and 1. For $\lambda_I = 1$, the chart is equivalent to the Phase I Shewhart chart.

We obtained values for L_I by setting the false alarm rate in Phase I at 1 percent, thereby following Chakraborti *et al.*[5] Table II gives an overview of the screening estimators considered and the corresponding values of L_I (obtained through 100,000 Monte Carlo simulations).

4. Comparison phase I estimation methods

In this section, we evaluate the performance of the proposed dispersion estimation methods when the Phase I data are in control as well as when the Phase I data contain contaminations. Recall that the Phase I data are $N(\mu, \sigma^2)$ distributed if the process is in control. Without loss of generality, we set $\mu = 0$ and $\sigma = 1$. In Section 4.1 we describe the data scenarios considered in Phase I. In Section 4.2 we define the performance measures, while in Section 4.3 we outline the simulation procedure. Section 4.4 contains the Phase I results.

4.1. Contamination scenarios

Many different contaminations scenarios are studied in the literature. In our paper, we distinguish between *scattered* and *sustained* special causes of variation. We evaluate two scattered scenarios - localized and diffuse - based on Tatum[3] and Schoonhoven and Does[8] and two sustained shift scenarios - single and multiple step shifts - based on Chen and Elsayed[24] and Amiri and Allahyari.[11] These four scenarios were also used by Zwetsloot *et al.*[25] except that they applied them to the location parameter.

1. A model with *localized* variance disturbances in which all observations in a sample have a probability $1-p$ of being drawn from the $N(0,1)$ distribution and a probability p of being drawn from the $N\left(0, \delta_I^2\right)$ distribution, with $\delta_I = 1, 1.5, .., 3.5, 4$.
2. A model with *diffuse* variance disturbances in which each observation is drawn from the $N(0,1)$ distribution and has a probability p of having a multiple of a χ_1^2 variable added to it, with multiplier γ_I, with $\gamma_I = 0, 0.5, 1, .., 2.5, 3$.
3. A model with a *single step* shift in the variance. All observations in the last $\lceil p * k \rceil$ Phase I samples, are drawn from the $N(0, \delta_I^2)$ distribution, with $\delta_I = 1, 1.5, .., 3.5, 4$.
4. A model with *multiple step* shifts in the variance. At each time point, the sample has a probability op q of being the first of $\lceil p * k \rceil$ consecutive samples drawn from the $N(0, \delta_I^2)$ distribution, with $\delta_I = 1, 1.5, .., 3.5, 4$. After any such step shift, each sample again has a probability of q of being the start of another step shift. If a shift occurs at the end of the Phase I data set, say at $k-1$, only samples $k-1$ and k will be contaminated, even if the length of the shift ($\lceil p * k \rceil$) should be longer.

Table III. Phase I contamination scenarios

Scenario	Prob.	Description	Shift size
In-control	0	All observations from $N(0,1)$	n/a
Localized	0.05	95 − 5 random mixture of $N(0,1)$ and $N(0,\delta_l^2)$	$\delta_l = 1.25, 1.5, ..., 4$
	0.10	90 − 10 random mixture of $N(0,1)$ and $N(0,\delta_l^2)$	$\delta_l = 1.25, 1.5, ..., 4$
Diffuse	0.05	95 − 5 random mixture of $N(0,1)$ and $N(0,1) + \gamma_l\chi_1^2$	$\gamma_l = 0.25, 0.5, ..., 3$
	0.10	90 − 10 random mixture of $N(0,1)$ and $N(0,1) + \gamma_l\chi_1^2$	$\gamma_l = 0.25, 0.5, ..., 3$
Single step	0.05	Samples 1- 47 are $N(0,1)$ and samples 48-50 are $N(0,\delta_l^2)$	$\delta_l = 1.25, 1.5, ..., 4$
	0.10	Samples 1- 45 are $N(0,1)$ and samples 46-50 are $N(0,\delta_l^2)$	$\delta_l = 1.25, 1.5, ..., 4$
Multiple steps	0.05	Shifts of length 3 from $N(0,\delta_l)$ occurring with probability 0.018	$\delta_l = 1.25, 1.5, ..., 4$
	0.10	Shifts of length 5 from $N(0,\delta_l)$ occurring with probability 0.023	$\delta_l = 1.25, 1.5, ..., 4$

All four scenarios are modeled such that they have an (approximate) contamination rate of $p * 100$ percent. We study both $p = 0.05$ and $p = 0.1$, and set $q = 0.018$ and $q = 0.023$ respectively in the multiple shift scenario. These values for q were determined through 100,000 Monte Carlo simulations. An overview of the contamination scenarios is provided in Table III.

4.2. *Performance measures*

One of the requirements of Phase I is to deliver an accurate parameter estimate of σ, even if Phase I contains contaminated observations. In order to evaluate the accuracy of the dispersion estimators, we determine their mean squared error (MSE), which is computed as

$$MSE = \frac{1}{R}\sum_{r=1}^{R}\left(\frac{\hat{\sigma}^r - \sigma}{\sigma}\right)^2 = \frac{1}{R}\sum_{r=1}^{R}(\hat{\sigma}^r - 1)^2.$$

Here $\hat{\sigma}^r$ denotes one of the proposed estimators presented in Section 3.1, calculated in the rth simulation run, and R is the total number of simulation runs.

The proposed estimators are also evaluated with two additional quality characteristics: the true alarm percentage (TAP) and the false alarm percentage (FAP). These additional performance measures reflect the ability of the screening estimators to detect unacceptable observations without triggering false alarms for acceptable observations. Related measures were presented by Fraker *et al.*,[26] Chakraborti *et al.*,[5] and Frisén.[27] The TAP and FAP are calculated as

$$TAP = \frac{1}{R}\sum_{r=1}^{R}\frac{(\text{\#correctly identified unacceptable observations})^r}{(\text{\#unacceptable observations})^r} * 100\%$$

and

$$FAP = \frac{1}{R}\sum_{r=1}^{R}\frac{(\text{\#incorrectly identified unacceptable observations})^r}{(\text{\#acceptable observations})^r} * 100\%,$$

where the superscript r denotes the r-th simulation run.

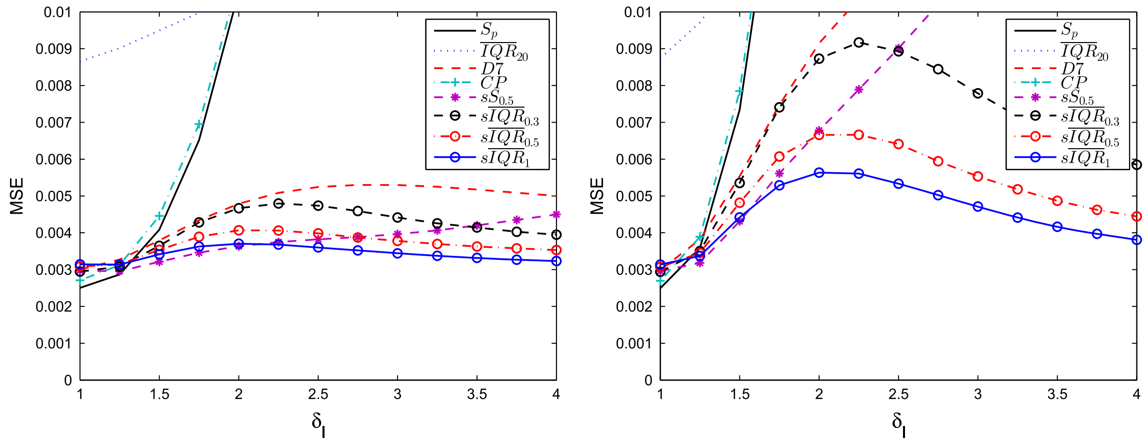
4.3. *Simulation procedure*

A Monte Carlo simulation study was conducted. We drew R Phase I data consisting of $k = 50$ samples of size $n = 5$, for each of the four contamination scenarios, each contamination rate, and each value of δ_l or γ_l (see Table III). The proposed eight dispersion estimators were calculated for each simulation run, and the three performance measures were computed based on the R runs. The relative simulation error is defined as the standard deviation of the MSE expressed as a percentage of the MSE. We set the value of R equal to 100,000, so that this error was never larger than 0.5%. The MSE results are presented in Figures 1 through 4 and the TAP and FAP metrics are shown in Table IV.

4.4. *Phase I results*

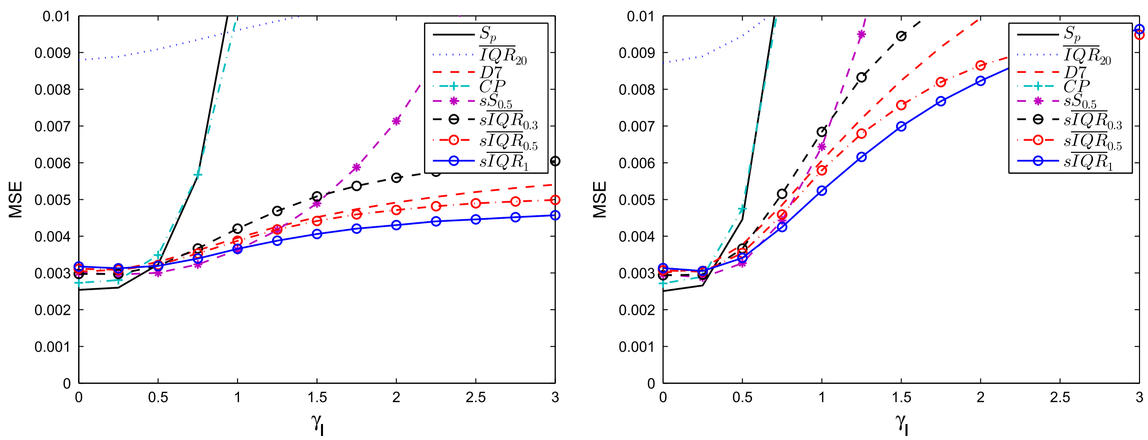
First consider the situation when the Phase I data are in control. In Table IV, the column corresponding to FAP and $\delta_l = 1$ or $\gamma_l = 0$ shows the false alarm percentage if the Phase I data are in control. This percentage was set equal to 1% for all estimators. In Figures 1 through 4, the y-intercept represents the MSE of the estimators based on uncontaminated Phase I data. As expected, the pooled sample standard deviation S_p shows the smallest MSE followed by the CP estimator, the screening estimators, and $D7$. The robust point estimator $\overline{IQR}_{20}$ has a very large MSE.

Furthermore, for uncontaminated Phase I data, the four screening estimators show an approximately equal MSE level. This implies that, from the perspective of efficiency under uncontaminated Phase I data, it does not matter whether we use an efficient ($sS_{0.5}$) or



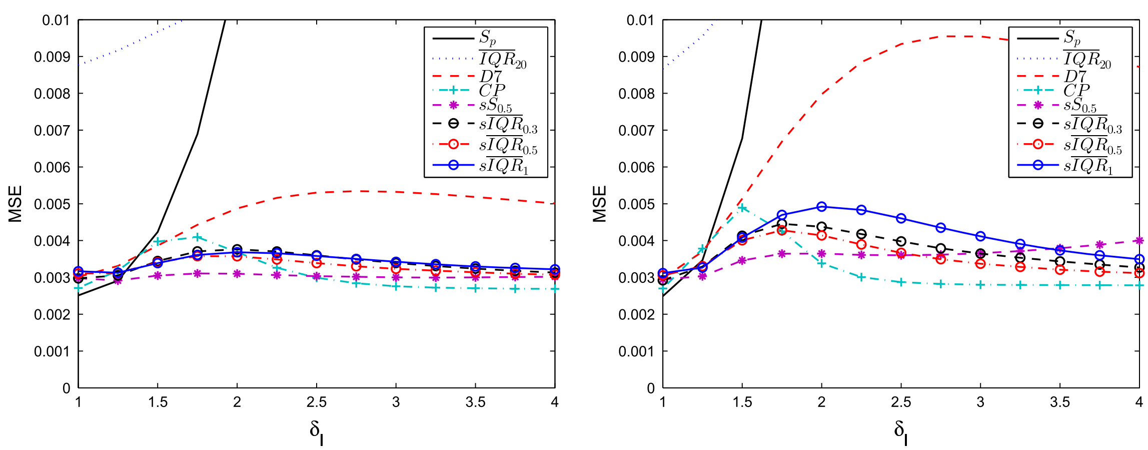
(a) 5% contamination (b) 10% contamination

Figure 1. Localized shifts in Phase I



(a) 5% contamination (b) 10% contamination

Figure 2. Diffuse shifts in Phase I

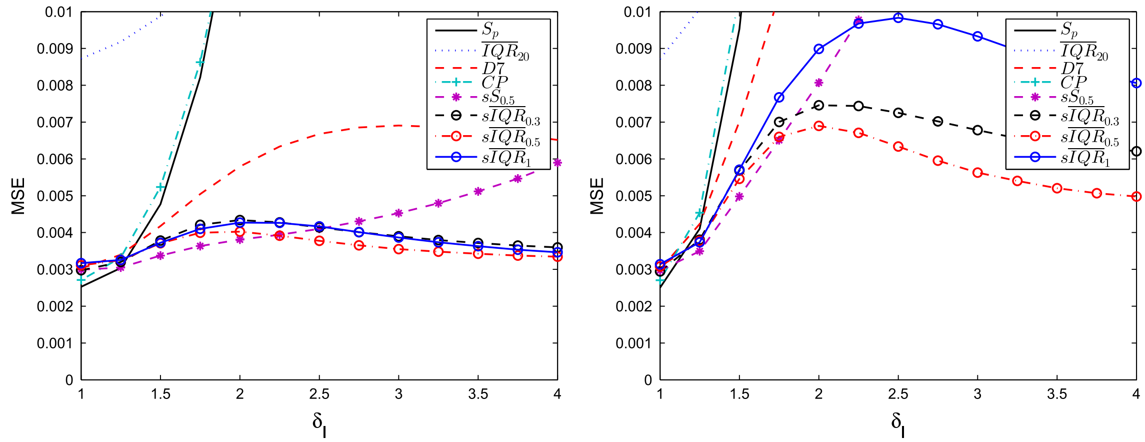


(a) 5% contamination (b) 10% contamination

Figure 3. A single step shift in Phase I

robust ($s\overline{IQR}_{0.5}$) initial estimator for $\hat{\sigma}_I$. Furthermore, it also does not matter from the perspective of efficiency under uncontaminated Phase I data which λ_I we use.

Next, we consider the situation when contaminations are present in the Phase I data ($\delta_I > 1$ or $\gamma_I > 0$). Overall, the EWMA-based methods perform reasonably well for any type of contamination, while the point estimator S_p is most sensitive to contaminations.



(a) 5% contamination (b) 10% contamination

Figure 4. Multiple step shifts in Phase I

When localized or diffuse shifts are present in Phase I, the screening estimators based on a robust initial estimator show the lowest MSE over all shifts sizes (see Figures 1 and 2). For the 10% contamination scenario, the screening estimators show diverging MSE levels, with $s\overline{IQR}_1$ and $s\overline{IQR}_{0.5}$ having the lowest MSE levels. The difference between 5% and 10% contamination rates is also shown in Table IV. The FAP of the estimator $s\overline{IQR}_{0.3}$ is the highest among the screening estimators. This is due to the low value of λ_I, suggesting that it is undesirable to set λ_I too low.

When a single step shift is present in Phase I (Figure 3), the estimator CP shows a low MSE, as was to be expected, because CP is specifically designed for sustained shifts. However, for small shift sizes ($1 < \delta_I < 2$), its MSE is higher than the MSE of the screening estimators. The TAP and FAP values show the same pattern: CP has the highest TAP followed by $s\overline{IQR}_{0.5}$ and it has the lowest FAP for both a 5% and 10% contamination rate.

When multiple step shifts are present in Phase I (Figure 4), the screening estimators based on $\overline{IQR}_{20}$ have the lowest MSE. For a 5% contamination rate, it does not really matter which λ_I we select. However, if 10% of the observations are contaminated and show multiple step shifts, the estimator $s\overline{IQR}_{0.5}$ shows the best MSE performance. It is able to detect most of the contaminated samples with TAP values up to 94.7%, and incorrectly deletes no more than 3% of the observations.

Overall, the best estimator is the screening estimator based on $\overline{IQR}_{20}$ and $\lambda_I = 0.5$. Irrespective of the contaminations in Phase I, it always has low MSE levels.

5. Application to the phase II EWMA control chart

Phase I estimation methods are used to design the Phase II control chart. In this section, we consider the effect of estimating σ on the EWMA control chart in Phase II when the Phase I data may or may not be contaminated.

Let Y_{it}, with $i = 1, .., n$ and $t = 1, 2, ..$, be the Phase II data which are independently and identically $N\left(\mu, \delta_{II}^2\sigma^2\right)$ distributed, where δ_{II} is a constant and the index 'II' indicates Phase II data. If $\delta_{II} = 1$, the data are in control and, if $\delta_{II} > 1$, the data are out of control. We only consider $\delta_{II} \geq 1$ as we study the one-sided EWMA dispersion control chart. We set $\mu = 0$ and $\sigma = 1$ without loss of generality.

5.1. *Design of the phase II EWMA control chart*

The Phase II EWMA control chart consists of the EWMA statistic

$$W_t = \max[\ (1-\lambda_{II})W_{t-1} + \lambda_{II}S_t\ ,\ c_4(n)\hat{\sigma}\ ], \tag{1}$$

with $W_0 = c_4(n)\hat{\sigma}$, and upper control limit

$$\widehat{UCL}_t = c_4(n)\hat{\sigma} + L_{II}\hat{\sigma}\sqrt{\frac{\lambda_{II}}{2-\lambda_{II}}}\sqrt{1-(1-\lambda_{II})^{2t}}\sqrt{(1-c_4(n)^2)}, \tag{2}$$

where $\hat{\sigma}$ is the Phase I estimate of σ. We consider the eight estimation methods described in Section 3.1, which results in eight Phase II EWMA control charts. Furthermore, λ_{II} is set equal to 0.3. This value differs from the value of λ_I (0.5) chosen in Phase I, as Phase I is used for exploratory purposes. The values of L_{II}, presented in Table V, were determined such that all EWMA control charts have an approximate in-control average run length performance of 200, whereby we followed the design procedure originally proposed by Jones[21].

Table IV. True Alarm Percentage (TAP) and False Alarm Percentage (FAP)

Scenario	perc.	$\hat{\sigma}$	TAP			FAP			
				δ_l			δ_l		
			2	3	4	1	2	3	4
Localized	5	CP	9.6	29.8	39.4	1.0	3.9	13.3	20.0
		$sS_{0.5}$	39.0	68.3	80.1	1.0	0.8	1.1	1.4
		$s\overline{IQR}_{0.3}$	31.3	69.0	86.5	1.0	1.8	4.4	7.5
		$s\overline{IQR}_{0.5}$	38.6	74.9	89.5	1.0	1.2	2.6	4.3
		$s\overline{IQR}_{1}$	43.1	77.7	90.8	1.0	0.7	0.7	0.6
	10	CP	10.7	26.4	30.0	1.0	6.4	19.0	24.5
		$sS_{0.5}$	33.0	57.7	68.4	1.0	0.6	0.8	1.1
		$s\overline{IQR}_{0.3}$	30.1	68.0	86.0	1.0	2.3	7.1	12.7
		$s\overline{IQR}_{0.5}$	36.4	73.0	88.4	1.0	1.3	3.7	6.8
		$s\overline{IQR}_{1}$	39.9	74.8	89.1	1.0	0.5	0.4	0.4
				γ_l			γ_l		
			1	2	3	0	1	2	3
Diffuse	5	CP	5.5	18.1	25.0	1.0	4.4	14.8	21.2
		$sS_{0.5}$	7.7	16.5	21.5	1.0	2.0	3.7	4.8
		$s\overline{IQR}_{0.3}$	7.8	23.0	35.7	1.0	3.0	8.8	15.0
		$s\overline{IQR}_{0.5}$	8.5	22.7	34.0	1.0	2.6	6.6	10.6
		$s\overline{IQR}_{1}$	8.7	21.4	30.9	1.0	2.2	4.2	5.9
	10	CP	7.9	21.6	25.9	1.0	6.9	20.1	24.9
		$sS_{0.5}$	7.4	14.6	17.9	1.0	2.6	5.0	6.1
		$s\overline{IQR}_{0.3}$	9.4	28.8	44.7	1.0	4.8	15.7	26.8
		$s\overline{IQR}_{0.5}$	9.6	26.2	39.5	1.0	4.0	11.4	18.7
		$s\overline{IQR}_{1}$	9.2	22.7	32.7	1.0	3.2	7.1	10.1
				δ_l			δ_l		
			2	3	4	1	2	3	4
Single step	5	CP	74.4	97.9	99.5	0.9	1.4	0.3	0.1
		$sS_{0.5}$	57.4	85.6	92.6	1.0	0.3	0.1	0.0
		$s\overline{IQR}_{0.3}$	52.5	85.2	94.2	1.0	0.6	0.4	0.4
		$s\overline{IQR}_{0.5}$	55.0	87.6	95.6	1.0	0.6	0.5	0.5
		$s\overline{IQR}_{1}$	43.4	78.3	91.1	1.0	0.7	0.6	0.6
	10	CP	90.0	98.6	99.5	0.9	1.1	0.2	0.1
		$sS_{0.5}$	58.6	85.0	91.4	1.0	0.1	0.0	0.0
		$s\overline{IQR}_{0.3}$	62.1	89.6	95.9	1.0	0.4	0.2	0.2
		$s\overline{IQR}_{0.5}$	59.6	90.6	96.9	1.0	0.4	0.3	0.3
		$s\overline{IQR}_{1}$	41.1	76.2	89.9	1.0	0.5	0.4	0.4
Multiple steps	5	CP	29.7	57.3	62.4	1.0	6.2	13.6	15.0
		$sS_{0.5}$	50.2	77.0	85.0	1.0	0.8	1.0	1.3
		$s\overline{IQR}_{0.3}$	49.8	83.7	93.4	1.0	1.9	3.9	5.5
		$s\overline{IQR}_{0.5}$	51.7	85.7	94.8	1.0	1.1	2.0	2.8
		$s\overline{IQR}_{1}$	40.8	75.7	89.5	1.0	0.8	0.7	0.7
	10	CP	39.5	53.3	53.7	1.0	11.5	17.0	17.4
		$sS_{0.5}$	45.3	69.1	76.5	1.0	0.6	0.8	0.9
		$s\overline{IQR}_{0.3}$	54.8	85.9	94.0	1.0	2.1	4.6	6.4
		$s\overline{IQR}_{0.5}$	51.7	85.9	94.7	1.0	1.0	1.9	2.8
		$s\overline{IQR}_{1}$	35.4	69.6	85.2	1.0	0.6	0.5	0.5

5.2. Performance measures

The performance of a control chart in Phase II is usually expressed in terms of the distribution function of the run length (RL) where RL is a random variable that represents the number of samples before a signal occurs, and a realization of *RL* will be denoted by *rl*.

For control charts with estimated parameters, a distinction is made between the conditional and unconditional run length distribution. The conditional distribution is the distribution of RL given the Phase I parameter estimate $\hat{\sigma}$. This reflects that the run length has a different distribution for each value of $\hat{\sigma}$. Similarly, percentiles and moments of the conditional distribution take different values for

Table V. L_{II} for the Phase II EWMA control chart based on $\hat{\sigma}$

$\hat{\sigma}$:	S_p	$\overline{IQR}_{20}$	$D7$	CP	$sS_{0.5}$	$s\overline{IQR}_{0.3}$	$s\overline{IQR}_{0.5}$	$s\overline{IQR}_{1}$
L_{II} :	2.607	2.210	2.570	2.580	2.680	2.650	2.660	2.677

Table VI. Performance of the Phase II EWMA control chart when 5% of the data in Phase I are contaminated

Phase I		ARL and percentiles of the unconditional run length distribution															
		In-control Phase II data				Out-of-control Phase II data											
		$\delta_{II}=1$				$\delta_{II}=1.1$				$\delta_{II}=1.2$				$\delta_{II}=1.4$			
Scenario	$\hat{\sigma}$	10th	50th	90th	ARL	10th	50th	90th	ARL	10th	50th	90th	ARL	10th	50th	90th	ARL
In-control	S_p	10	86	467	201	3	23	100	42	1	9	36	15	0	3	11	5
	$D7$	9	79	458	200	3	21	98	42	1	9	35	15	0	3	11	5
	$\overline{IQR}_{20}$	3	36	348	207	1	12	79	39	0	6	29	13	0	2	9	4
	CP	10	84	461	204	3	22	99	42	1	9	36	15	0	3	11	5
	$sS_{0.5}$	10	82	469	204	3	22	99	42	1	9	35	15	0	3	11	5
	$s\overline{IQR}_{0.3}$	10	82	468	202	3	22	99	42	1	9	36	15	0	3	11	5
	$s\overline{IQR}_{0.5}$	10	82	471	204	3	22	99	42	1	9	35	15	0	3	11	5
	$s\overline{IQR}_{1}$	9	79	460	200	3	21	97	41	1	9	35	15	0	3	11	5
Localized ($\delta_I=2.5$)	S_p	42	673	22854	4881	9	104	1745	1101	3	30	285	213	1	7	33	17
	$D7$	15	139	1063	515	4	32	177	80	2	12	54	23	0	4	14	6
	$\overline{IQR}_{20}$	5	60	723	458	2	18	133	76	1	8	43	21	0	3	11	5
	CP	23	332	15809	3895	6	62	1273	1283	2	20	222	394	0	5	29	32
	$sS_{0.5}$	14	124	898	446	4	30	155	72	2	11	49	22	0	4	13	6
	$s\overline{IQR}_{0.3}$	12	117	954	506	4	29	161	79	1	11	50	22	0	4	13	6
	$s\overline{IQR}_{0.5}$	12	106	771	382	3	27	138	63	1	11	45	20	0	3	12	5
	$s\overline{IQR}_{1}$	12	103	707	335	3	26	130	58	1	10	43	19	0	3	12	5
Diffuse ($\gamma_I=1.5$)	S_p	48	833	30000	6649	10	125	5664	2599	4	35	679	1020	1	7	53	168
	$D7$	14	131	886	406	4	31	158	68	2	12	50	21	0	4	13	6
	$\overline{IQR}_{20}$	5	52	580	345	2	16	113	59	1	7	38	17	0	3	11	4
	CP	25	347	14544	3800	6	65	1210	1515	2	21	217	688	1	5	29	157
	$sS_{0.5}$	18	170	1336	663	5	37	206	98	2	14	61	27	0	4	15	6
	$s\overline{IQR}_{0.3}$	14	141	1119	550	4	33	184	84	2	12	56	24	0	4	14	6
	$s\overline{IQR}_{0.5}$	13	130	976	466	4	31	166	74	2	12	51	22	0	4	13	6
	$s\overline{IQR}_{1}$	14	128	913	429	4	30	159	70	2	12	50	22	0	4	13	6
Single step ($\delta_I=2.5$)	S_p	95	1288	19996	5278	16	167	1602	838	5	42	273	138	1	8	34	15
	$D7$	17	157	1132	513	4	35	186	80	2	13	56	24	0	4	14	6
	$\overline{IQR}_{20}$	5	66	794	481	2	19	143	78	1	8	46	21	0	3	12	5
	CP	10	82	479	223	3	22	100	46	1	9	36	16	0	3	11	5
	$sS_{0.5}$	13	115	734	322	4	28	134	58	2	11	44	19	0	4	12	5
	$s\overline{IQR}_{0.3}$	12	111	763	344	4	27	137	59	1	11	45	19	0	4	12	5
	$s\overline{IQR}_{0.5}$	12	102	679	307	3	26	126	55	1	10	42	18	0	3	12	5
	$s\overline{IQR}_{1}$	12	108	747	338	4	27	135	59	1	11	44	19	0	3	12	5
Multiple steps ($\delta_I=2.5$)	S_p	25	494	30000	6280	7	83	4611	2231	3	26	579	643	1	6	47	42
	$D7$	14	138	1353	778	4	33	207	121	2	12	59	30	0	4	14	6
	$\overline{IQR}_{20}$	5	60	840	596	2	18	150	100	1	8	46	25	0	3	12	5
	CP	13	140	23776	3708	4	34	1735	2047	2	13	257	1051	0	4	29	178
	$sS_{0.5}$	12	111	840	498	4	28	147	88	1	11	47	24	0	4	13	6
	$s\overline{IQR}_{0.3}$	11	101	745	413	3	26	136	71	1	10	44	21	0	3	12	5
	$s\overline{IQR}_{0.5}$	11	96	669	346	3	25	126	61	1	10	42	19	0	3	12	5
	$s\overline{IQR}_{1}$	11	105	776	425	3	26	140	73	1	11	45	21	0	3	12	5

each value of $\hat{\sigma}$. We denote the conditional distribution by $F_{RL|\Sigma}(rl|\hat{\sigma}) = P\left[RL \leq rl|\hat{\sigma}\right]$. Here, $\hat{\sigma}$ is a realization of the random variable Σ, which is a function (one of the eight proposed estimators) of the random Phase I data set. In order to evaluate the overall behavior of the Phase II EWMA control chart, we consider the unconditional distribution of the run length $F_{RL}(rl)$, which takes account of the variability of $\hat{\sigma}$, i.e., $F_{RL}(rl) = \int_0^{\infty} F_{RL|\Sigma}(rl|\hat{\sigma}) f_{\Sigma}(\hat{\sigma}) d\hat{\sigma}$.

A common measure of the performance of a control chart is the expected value of the *RL*, the average run length (ARL). It is desirable to have a large ARL when the process is in control and a small ARL when the process is out of control. We evaluate the ARL as well as the 10th, 50th, and 90th percentiles of the unconditional run length distribution as performance measures of the Phase II EWMA control chart for dispersion.

5.3. Simulation procedure

A Monte Carlo simulation study was conducted to evaluate the unconditional run length distribution of the Phase II EWMA dispersion control chart based on the eight Phase I dispersion estimators presented in Section 3. We used the following simulation procedure: first $k = 50$ samples of size $n = 5$ are drawn from $N(0, 1)$. We calculated $\hat{\sigma}$ using the eight proposed estimators. These estimates were used to set up eight Phase II EWMA control charts according to equations 1 and 2. Next, observations from $N\left(0, \delta_{II}^2\right)$ are drawn until the associated W_t falls above the control limit. The corresponding run length equals $t-1$. The calculations are made for $\delta_{II} = 1, 1.1, 1.2, 1.4$. The entire procedure was repeated for $R = 100,000$ simulation runs. This gives 100,000 realizations of $f_{RL}(rl)$ for each value of δ_{II} and for each estimator $\hat{\sigma}$. The unconditional ARL is computed by averaging over all *rl*s and the percentiles are taken by sorting the 100,000 run lengths and selecting the $10,000th$, the average of the $50,000th$ and $50,001st$, and the $90,000th$ run lengths.

This whole procedure was repeated for each of the four Phase I contamination scenarios, where we set $\delta_I = 2.5$ or $\gamma_I = 1.5$, and for simplicity, we only considered the Phase I scenarios with a 5% contamination rate. The results are presented in Table VI. The first part of the table shows the performance of the EWMA control charts when the Phase I data are in control, followed by the results when Phase I contains out-of-control observations as defined in the four Phase I contamination scenarios.

For computational convenience and speed, we truncated the simulation at $rl = 30,000$ and set a run length greater than 30,000 equal to this value. These values are therefore an underestimate of the actual ARL or percentile.

5.4. Phase II results

First, consider the situation where the Phase I data are uncontaminated (first part of Table VI). A general observation is that under in-control data the EWMA control charts show similar performance across all estimators (i.e. they have in-control ARLs of around 200 and they show similar out-of-control ARLs and percentiles). One exception is the EWMA control chart based on $\overline{IQR}_{20}$, which has the worst performance.

Next, consider the situation where the Phase I data are contaminated. A general effect of contamination in Phase I is an increased in-control ARL as well as an increased out-of-control ARL. This is of course undesirable. The control chart based on the traditional estimator S_p is less effective if any type of contamination is present in Phase I. The control chart based on *CP* shows good performance for the single step shift scenario but does not work very well for the other contamination scenarios. The control chart based on $s\overline{IQR}_1$ shows the best performance for localized and diffuse shifts. The estimator $s\overline{IQR}_{0.5}$ has the second-best performance for these scenarios. If there are multiple step shifts present, the best performance is given by the control chart based on $s\overline{IQR}_{0.5}$.

To summarize, the type of disturbance and estimation methods used in Phase I strongly determines the performance of the Phase II EWMA control chart. When it is unknown which type of contaminations are present in Phase I, we recommend the use of the EWMA chart in Phase I based on a robust initial estimator and $\lambda_I = 0.5$ ($s\overline{IQR}_{0.5}$).

6. Concluding remarks

In this article, we have proposed a new Phase I estimation method for the dispersion of a process. This method is based on Phase I EWMA charting and provides an efficient estimator of the dispersion for in-control Phase I data and a robust estimate of the dispersion if contaminations are present in Phase I. We have compared this new method with several estimation methods from the literature, in terms of their accuracy (MSE) and the percentage of successfully identified samples (TAP and FAP). Moreover, we have investigated the impact of data contaminations in Phase I on the performance of the Phase II EWMA control chart based on the various dispersion estimators.

In our study we show that the existing Phase I estimation methods provide robust estimates for *specific* patterns of disturbances in Phase I. In particular, estimators based on Phase I Shewhart charts are robust primarily to outliers in Phase I and changepoint methods are robust to sustained shifts in Phase I. The new method, based on Phase I EWMA charting, shows MSE levels which are comparable to the MSE of the Phase I Shewhart chart estimator if outliers are present and which are also comparable to the MSE of the changepoint methods if sustained shifts are present. Thus, the proposed method provides a robust estimate for *any* pattern of disturbances in Phase I.

The choice of the smoothing constant of Phase I EWMA chart is important as it influences robustness against the various patterns of contaminations. By studying the TAP and FAP, we have discovered that for small values of λ (0.3) the estimator deletes too many in-control samples from Phase I and for larger values of λ (1) the estimator does not identify sustained shifts in Phase I optimally. Therefore, we recommend estimating the process dispersion by means of a Phase I EWMA chart with a smoothing constant of 0.5.

Furthermore, we recommend the use of a two-step procedure, namely a robust estimator to estimate the dispersion and construct the Phase I chart, and an efficient estimator for post-screening estimation. For practitioners, the Phase I EWMA chart is easy to implement.

Acknowledgement

The authors are very grateful to prof. W.H. Woodall for his very helpfull comments.

References

1. Jensen WA, Jones-Farmer LA, Champ CW, Woodall WH. Effects of Parameter Estimation on Control Chart Properties: a Literature Review. *Journal of Quality Technology* 2006; **38**(4):349–364.
2. Rocke DM. Robust Control Charts. *Technometrics* 1989; **31**(2):173–184.
3. Tatum LG. Robust Estimation of the Process Standard Deviation for Control Charts. *Technometrics* 1997; **39**(2):127–141.
4. Psarakis S, Vyniou AK, Castagliola P. Some Recent Developments on the Effects of Parameter Estimation on Control Charts. *Quality and Reliability Engineering International* 2013; **29**(7):951–970.
5. Chakraborti S, Human SW, Graham MA, Phase I. Statistical Process Control Charts: An Overview and Some Results. *Quality Engineering* 2008; **21**(1): 52–62.
6. Jones-Farmer LA, Woodall WH, Steiner SH, Champ CW. An Overview of Phase I Analysis for Process Improvement and Monitoring, Accepted for publication in *Journal of Quality Technology* 2014.
7. Schoonhoven M, Riaz M, Does RJMM. Design and Analysis of Control Charts for Standard Deviation with Estimated Parameters. *Journal of Quality Technology* 2011; **43**(4):307–333.
8. Schoonhoven M, Does RJMM. A Robust Standard Deviation Control Chart. *Technometrics* 2012; **54**(1):73–82.
9. Jones-Farmer LA, Champ CW. A Distribution-Free Phase I Control Chart for Subgroup Scale. *Journal of Quality Technology* 2010; **42**(4):373–387.
10. Capizzi G, Masarotto G. Phase I Distribution-Free Analysis of Univariate Data. *Journal of Quality Technology* 2013; **45**(3):273–284.
11. Amiri A, Allahyari S. Change Point Estimation Methods for Control Chart Postsignal Diagnostics: A Literature Review. *Quality and Reliability Engineering International* 2012; **28**(7):673–685.
12. Roberts SW. Control Chart Tests Based on Geometric Moving Averages. *Technometrics* 1959; **1**(3):239–250.
13. Crowder SV. A Simple Method for Studying Run–Length Distributions of Exponentially Weighted Moving Average Charts. *Technometrics* 1987; **29**(4):401–407.
14. Lucas JM, Saccucci MS. Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements. *Technometrics* 1990; **32**(1):1–12.
15. Jones LA, Champ CW, Rigdon SE. The Performance of Exponentially Weighted Moving Average Charts with Estimated Parameters. *Technometrics* 2001; **43**(2):156–167.
16. Ng CH, Case KE. Development and Evaluation of Control Charts using Exponentially Weighted Moving Averages. *Journal of Quality Technology* 1989; **21**(4):242–250.
17. Shu L, Jiang W. A New EWMA Chart for Monitoring Process Dispersion. *Journal of Quality Technology* 2008; **40**(3):319–331.
18. Knoth S. Control Charting Normal Variance–Reflections, Curiosities, and Recommendations, in Frontiers in Statistical Quality Control 9. Springer-Verlag: Berlin Heidelberg, 2010; 3–18.
19. Steiner SH. EWMA Control Charts with Time-Varying Control Limits and Fast Initial Response. *Journal of Quality Technology* 1999; **31**(1):75–86.
20. Saleh NA, Mahmoud AM, Keefe MJ, Woodall WH. The Difficulty in Designing Shewhart $\overline{X}$ and X Control Charts with Estimated Parameters. Accepted for publication in *Journal of Quality Technology* 2014.
21. Jones LA. The Statistical Design of EWMA Control Charts with Estimated Parameters. *Journal of Quality Technology* 2002; **34**(3):277–288.
22. Mahmoud MA, Henderson GR, Epprecht EK, Woodall WH. Estimating the Standard Deviation in Quality-Control Applications. *Journal of Quality Technology* 2012; **42**(4):348–357.
23. Sullivan JH, Woodall WH. A Control Chart for Preliminary Analysis of Individual Observations. *Journal of Quality Technology* 1996; **28**(3):265–278.
24. Chen A, Elsayed EA. Design and Performance Analysis of the Exponentially Weighted Moving Average Mean Estimate for Processes Subject to Random Step Changes. *Technometrics* 2002; **44**(4):379–389.
25. Zwetsloot IM, Schoonhoven M, Does RJMM. A Robust Estimator for Location in Phase I Based on an EWMA Chart. Accepted for publication in *Journal of Quality Technology* 2014.
26. Fraker SE, Woodall WH, Mousavi S. Performance Metrics for Surveillance Schemes. *Quality Engineering* 2008; **20**(4):451–464.
27. Frisén M. Optimal Sequential Surveillance for Finance Public Health, and Other Areas (with discussion). *Sequential Analysis: Design Methods and Applications* 2009; **28**(3):310–337.

Authors' biographies

Inez M. Zwetsloot is a consultant of statistics at IBIS UvA and a PhD student at the Department of Operations Management at the University of Amsterdam. She is working on her PhD on exponentially weighted moving average control charting techniques. Her email address is i.m.zwetsloot@uva.nl.

Marit Schoonhoven is a senior consultant of statistics at IBIS UvA. Her current research focuses on the design of control charts for nonstandard situations. Her email address is m.schoonhoven@uva.nl.

Ronald J.M.M. Does is a professor of Industrial Statistics at the University of Amsterdam, a managing director of IBIS UvA, and a director of the Institute of Executive Programmes at the University of Amsterdam. He is a fellow of ASQ and ASA. His current research activities include the design of control charts for nonstandard situations, health care engineering, and Lean Six Sigma methods. His email address is r.j.m.m.does@uva.nl.