Performance of robust EWMA control chart for variability process using non-normal data

N A Prabawani¹, M Mashuri² and Irhamah³
¹,²,³ Department of Statistics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember, Surabaya, 60111, Indonesia

Email: ¹nimasayuprabawani@gmail.com, ²m_mashuri@statistika.its.ac.id, and ³irhamah@statistika.its.ac.id

Abstract. The main purpose of quality control is to quickly detect the presence of assignable causes and shifts in the process so that an investigation of the process can be carried out as early as possible. The Shewhart control chart provides good performance when the observation data is normally distributed, whereas when the normality assumption is not met, a Robust control chart is needed. The performance of the control chart depends on the stability of the estimator used to estimate the process parameters and establish control limits in phase I. In this study, a Robust Exponentially Weighted Moving Average (EWMA) control chart will be presented to monitor process variability using one of the estimator Robust scale to estimate standard deviation. This estimator is used to develop robust control limits. Then evaluate the control chart performance using Average Run Length (ARL) and Standard Deviation Run Length (SDRL) with Monte Carlo simulation. Furthermore, the robust chart was applied to monitor the quality characteristics of the number of bacterial colonies in each aquaculture medicine product. The results obtained in this study are formed a control chart that is resistant to the existence of outliers and sensitive to shifts in the process of variability.

1. Introduction
Control charts are the basic tools in statistical process control and widely used to control variations in industrial processes, so it is very important for control charts to quickly detect signs of out of control when there is a shift in process averages [1]. The EWMA control chart can provide good sensitivity to detect small shifts in process averages and process variability compared to Shewhart's X-bar control graph [2]. Some of the expansion of EWMA control charts such as in the research that designed one EWMA chart can monitored process averages and process variability simultaneously using MaxEWMA [2], extending EWMA control charts to Generally EWMA (GWMA) for both GWMA mean and variability [1] and the combination of MaxEWMA control charts with GWMA so that called MaxGWMA control charts [3]. The Shewhart control chart provides good performance when the observation data is normally distributed, whereas when the normality assumption has not met, a Robust control chart is needed. Similarly, when there are extreme values or outliers, averages and standard deviations cannot represent data properly, so it is necessary to use the robust method. Robust estimators are estimators that are not sensitive to changes in the underlying distribution and also resistant to outliers [4]. Quality control using robust estimators for process averages has been carried out to developing robust control charts based on trimmed mean and modification of trimmed standard deviation and using Gini’s mean difference [4].
EWMA control chart based on several robust scale estimators shows that the robust control chart performance is measured using Expected Point Out of Control (EPO) and Expected Width (EW) [5]. Previous research evaluates robust estimators to estimate location in phase I on the EWMA control chart [6]. Other research explain about the use of robust estimators on process variability using Median Absolute Deviation (MAD) to estimate the standard deviation [7]. Whereas the study proposes several robust estimators namely, calculate pooled subgroup standard deviation, calculate residuals in each subgroup, average subgroup range, average median deviation from the median and look at the efficiency of the estimators by using Mean Square Error (MSE) [8]. In this study, the EWMA control chart will be presented to monitor process variability using the Robust estimator on the quality characteristics of the number of bacterial colonies in each aquaculture medicine product that did not meet the assumption of normality. One of the Robust scale estimators used to estimate variance is the $S_0$ estimator [9]. The EWMA Robust control chart performance then evaluated using Average Run Length (ARL) and Standard Deviation Run Length (SDRL) with the Monte Carlo approach.

2. Literature

2.1. EWMA Control Chart for Variability

The statistic that used to monitoring process variability is EWMS [10]. In addition, EWMA statistics have been developed for monitoring process variability as follows [1].

$$V_i = (1 - \lambda) \max\{V_{i-1}, \sigma_i^2\} + \lambda (X_i - \mu_0)^2, i = 1, 2, ..., n$$

where $X_i$ is the $i$-th quality characteristic with $i$ showing subgroup of $n$, $V_0 = \sigma_0^2$ and $0 < \lambda \leq 1$.

The control limits on the EWMA control chart for process variability are as follows.

$$UCL = \sigma_0^2 + L\sigma_0^2 \sqrt{\frac{\lambda}{(2 - \lambda)}}$$

In the case of subgroup data with unit subgroup size $k > 1$, so change $X_i$ to be $\overline{X}_i$ and $\sigma_0$ become $\sigma_\tau = \sigma_0 \sqrt{k}^{1/2}$, so the EWMA statistics for process variability are defined as follows.

$$V_i = (1 - \lambda) \max\{V_{i-1}, \sigma_\tau^2\} + \lambda (\overline{X}_i - \mu_0)^2, i = 1, 2, ..., n; j = 1, 2, ..., k$$

2.2. Robust Estimator for Variance

Suppose that $X_{ij}$ is the quality characteristics of the $i$-th subgroup ($i = 1, 2, ..., n$) and the $j$-th unit subgroup ($j = 1, 2, ..., k$), using a robust estimator for standard deviations by calculating residuals in each subgroup. This approach applies to subgroup size $k \geq 4$. The residual calculation involves subtracting each observation value ($X_{ij}$) with the median subgroup ($M_j$) as follows.

$$res_{ij} = X_{ij} - M_j$$

If the unit subgroup size $(k)$ is odd, then the number of residual is $k^* = (k - 1)n$ and $k^* = kn$ when the unit subgroup size $(k)$ is even. Furthermore, the standard deviation estimator $(S)$ is defined as follows.

$$S = \frac{k^*}{\sqrt{(k^* - 1)}} \left( \sum_{j=1}^{k^*} \sum_{|u_j| \leq 1} res_{ij}^2(1 - u_j^2)^4 \right)^{1/2}$$

where $u_{ij} = h_{ij}e_k$, $c^\text{H}$ and $M^*$ is the median of the absolute values of all residuals, with $h_j$ as follows.
and \( E_i = \text{IQR}_i / M^* \). IQR\(_i\) is the interquartile range in the \( i\)-th subgroup and is defined as the difference between the second smallest observation with the second largest for the unit subgroup sizes \( 4 \leq k \leq 7 \) and the difference between the third smallest observation and the third largest for the unit subgroup sizes \( 8 \leq k \leq 11 \). Each constant \( c \) leads to a different estimator. Estimators with \( c = 7 \) providing less efficient estimators when there is no interference, but estimators are more efficient when there are disturbances [9].

### 2.3. EWMA control chart for variability

The performance of the control chart depends on the stability of the estimator used to estimate the process parameters and establish control limits in phase I. That is because if the estimator used to determine the control limits is affected by extreme values, the control limits obtained will be biased and can affect the performance of the control chart. One of them can cause the control limit on the control chart to widen, so the sensitivity level of the control chart decreases [11]. To overcome this, you can use Robust estimators in making control charts, namely estimators that are not sensitive to changes in the underlying distribution and resistant to outliers. Previous research mentioned several scale estimators that are commonly used and have good performance for one or more distributions including [12], \( S \) (sample standard deviation), Strim (trimmed standard deviation), MAD [13], SG (Gaussian skip), \( M^* \)-estimate, and \( A^*-estimate \). In the case of an outlier the estimator for the standard deviation namely \( S \), is sufficient to be considered because it can be screened in each subgroup with IQR [8]. The first step to obtaining plot statistics in equation (1) is to substitute robust estimators for standard deviations as in equation (5) and apply also to control limits in equation (2). The control limits (\( h \)) developed using the estimator are as follows.

\[
h = S_c^2 + L S_c^2 \sqrt{\frac{\lambda}{(2-\lambda)}}
\]

### 3. Real Data Example

The data used in this study are secondary data sourced from the Quality Control division of a company engaged in the production and distribution of veterinary medicines and pesticides. All products are produced under a Good Manufacturing Practice (GMP) based management system, from receiving raw materials to final delivery. Eco-friendly products at this company are intended for animal health (aquaculture and poultry) and public health (mosquito control). The research variable used in this study is the number of bacterial colonies in the product which is a quality characteristic of aquaculture medicine. Calculation number of product colonies based on the Standar Nasional Indonesia (SNI) 01-2332.3-2006. Data on product samples was taken for 9 samples in each test the number of bacterial colonies to obtain residual values as in Table 1.

| Subgroup | Residuals |
|----------|-----------|
| 1        | -0.3      | -0.4 | 0.2  | -0.91 | 0.2 | 0.2 | 0.7 | -0.2 |
| 2        | -1.78     | 0.3  | 1.2  | -1.93 | 0.3 | 0.9 | -0.4 | -1.2 |
| 3        | -0.29     | 0.8  | 1.1  | 0     | 0   | 12.9 | 13.9 | 0   |
| 4        | 0.4       | 0.3  | 1.17 | -0.5  | 0.1 | -0.5 | -0.5 | 0.1 |
| 5        | -0.76     | 0.6  | 1.8  | -0.53 | 0   | -0.52 | 0.5 | 1   |
| 6        | -0.74     | 0.6  | 1.8  | -0.37 | 0   | -0.3 | 0.5 | 1   |
Using this data, some control limits are obtained in Table 2. Based on the results obtained, it shows that the control limit values are getting narrower when the parameter values \( L \) and \( \lambda \) are small. Therefore the number of out of control (ooc) points also increases with the narrowing of the control limits. The control chart with increasing out of control points is shown in Figure 1.

| Subgroup | Residuals |
|----------|-----------|
| 7        | -0.3      | 0.3 | 0.2 | 0.2 | -0.3 | -0.5 | 0.5 | -0.1 |
| 8        | -0.3      | 1.1 | -0.2 | -0.1 | 0.6 | -0.3 | 0.5 | 1.4 |
| 9        | -0.5      | 0.7 | -0.8 | 0.3 | 1.5 | -0.74 | -0.1 | 1.3 |
| 10       | -0.6      | 0.2 | -0.6 | -0.1 | 0.1 | -0.4 | 0.6 | 0.1 |
| 11       | -0.8      | 0.1 | -0.3 | -1 | 0.3 | -0.4 | 7.8 | 0 |
| 12       | -1.1      | -0.2 | 0.3 | -1.1 | -0.4 | 0.2 | 0.1 | 0.2 |
| 13       | -0.6      | 0.2 | 0.4 | -0.4 | -0.3 | -0.1 | 0.3 | 0.4 |
| 14       | -0.5      | 0.8 | 1.1 | -0.6 | 0.8 | 1.2 | -0.6 | -0.2 |
| 15       | -0.3      | 0.2 | 0.1 | -1 | -0.1 | 0.1 | 0.3 | -0.7 |
| 16       | 0.8       | 1 | -0.1 | 0.3 | -0.1 | -0.7 | 0.5 | -0.2 |
| 17       | 1.2       | 0.2 | -0.4 | -0.4 | -0.74 | 0.5 | 0.4 | -0.6 |
| 18       | 0.2       | -0.3 | 0.3 | -0.2 | -0.5 | 0.6 | 0.3 | -0.2 |
| 19       | -0.3      | 0.1 | -0.4 | 0.1 | 0.4 | -0.2 | 0.1 | -0.5 |
| 20       | 0.1       | 0.1 | -0.5 | -0.1 | -0.5 | 0.3 | 0.2 | -0.3 |

Using this data, some control limits are obtained in Table 2. Based on the results obtained, it shows that the control limit values are getting narrower when the parameter values \( L \) and \( \lambda \) are small. Therefore the number of out of control (ooc) points also increases with the narrowing of the control limits. The control chart with increasing out of control points is shown in Figure 1.

| Table 2. Control limits with some values \( L \) and \( \lambda \) |
|-----------------|-------|------|------|------|-------|
| No. | \( L \) | \( \lambda \) | UCL  | LCL  | Number of ooc |
|-----|------|------|------|------|----------------|
| 1   | 3.090 | 1.00 | 1.3055 | -0.6671 | 1               |
| 2   | 3.087 | 0.75 | 1.0825 | -0.4441 | 1               |
| 3   | 3.071 | 0.50 | 0.8852 | -0.2468 | 1               |
| 4   | 3.054 | 0.40 | 0.8066 | -0.1682 | 2               |
| 5   | 3.023 | 0.30 | 0.7246 | -0.0862 | 2               |
| 6   | 2.998 | 0.25 | 0.6809 | -0.0425 | 2               |
| 7   | 2.962 | 0.20 | 0.6344 | 0.0040 | 2               |
| 8   | 2.814 | 0.10 | 0.5253 | 0.1131 | 2               |
| 9   | 2.615 | 0.05 | 0.4529 | 0.1855 | 3               |
| 10  | 2.437 | 0.03 | 0.4152 | 0.2232 | 4               |

Based on Figure 1 shows that the presence of outliers does not affect the control limits formed or control limits do not widen, it means that the robust estimator used is not sensitive to outliers. In Figure 1.a, there is 1 out of control points in 3rd subgroup, while in Figure 1.b there are 2 out of control points, namely in 3rd and 4th subgroups. This show the smaller value of the parameter \( L \) and \( \lambda \) the control chart becomes more sensitive.
4. Performance of Control Chart Based on Simulation Procedure

ARL definition is the average number of points that must be plotted before a point indicates conditions beyond control [10]. ARL is one of several ways to suspect signs of a process not being controlled or outside the control limits. ARL is also called the waiting time for out of control sign, because ARL shows how long the average time is needed to plot points in the control chart before an out of control point is detected. In general, there are three procedures used to obtain the ARL distribution, namely integral equations, Markov Chain approach, and Monte Carlo simulation. Suppose $\hat{v}_i$ is a plot statistic estimated from the $i$-th subgroup of $X_{ij}$, $i = 1, 2, ..., n$. Then $F$ is event $\hat{v}_i$ greater than UCL or lower than LCL. Defined $P(F_i | \hat{v}_i)$ as a probability that subgroup $i$ produces the signal, given as follows.

$$ P(F_i | \hat{v}_i) = P(\hat{v}_i < LCL \ or \ \hat{v}_i > UCL | \hat{v}_i) \quad (7) $$

Conditional ARL is $E(ARL | \hat{v}_i) = 1 / P(F_i | \hat{v}_i)$, the expectations of $X_{ij}$ then obtained unconditional ARL and SDRL as follows [8].

$$ ARL = E \left( \frac{1}{P(F_i | \hat{v}_i)} \right) $$

$$ SDRL = \sqrt{\text{var}(ARL)} = \sqrt{E(\text{var}(ARL | \hat{v}_i)) + \text{var}(E(ARL | \hat{v}_i))} \quad (8) $$

Expectation values are obtained using a dataset simulation based on a normal or contaminated normal distribution. Each dataset obtained the value of $E(ARL | \hat{v}_i)$ and iteration 1000 times with the results of the ARL and SDRL values using several combinations of process variability shifts listed in Table 3.
Table 3. ARL and SDRL with $n = 500$, some value of parameter $L$ and $\lambda$, then some shifts in variability process $\theta = 0.0, 0.25, 0.50, 0.75, 1.00, 1.50$

| Parameter L and $\lambda$ | Shift | 0 | 0.25 | 0.5 | 0.75 | 1 | 1.5 |
|---------------------------|-------|----|------|-----|------|---|-----|
| $L=3.090, \lambda=1.00$  | ARL   | 44.028 | 20.272 | 11.874 | 8.086 | 5.658 | 3.914 |
|                           | SDRL  | 64.783 | 25.286 | 15.207 | 9.123 | 6.009 | 4.076 |
| $L=3.087, \lambda=0.75$  | ARL   | 15.42 | 9.374 | 4.498 | 3.836 | 2.704 | 1.944 |
|                           | SDRL  | 25.036 | 14.199 | 11.352 | 4.707 | 2.694 | 1.722 |
| $L=3.071, \lambda=0.50$  | ARL   | 3.23 | 2.008 | 1.532 | 1.472 | 1.154 | 1.106 |
|                           | SDRL  | 8.069 | 4.452 | 1.638 | 1.895 | 0.824 | 0.988 |
| $L=3.054, \lambda=0.40$  | ARL   | 2.388 | 1.66 | 1.28 | 1.248 | 1.086 | 1.026 |
|                           | SDRL  | 7.725 | 2.796 | 1.466 | 1.421 | 0.865 | 0.271 |
| $L=3.023, \lambda=0.30$  | ARL   | 1.538 | 1.276 | 1.102 | 1.12 | 1.062 | 1.01 |
|                           | SDRL  | 4.232 | 1.659 | 0.87 | 1.06 | 0.832 | 0.184 |
| $L=2.998, \lambda=0.25$  | ARL   | 1.876 | 1.206 | 1.104 | 1.034 | 1.054 | 1.008 |
|                           | SDRL  | 8.379 | 1.665 | 1.939 | 0.396 | 0.544 | 0.179 |
| $L=2.962, \lambda=0.20$  | ARL   | 2.134 | 1.106 | 1.09 | 1.002 | 1.028 | 1.008 |
|                           | SDRL  | 12.039 | 1.207 | 1.925 | 0.045 | 0.384 | 0.179 |
| $L=2.814, \lambda=0.10$  | ARL   | 1.536 | 1.242 | 1.05 | 1 | 1.008 | 1 |
|                           | SDRL  | 7.137 | 3.661 | 1.074 | 0 | 0.141 | 0 |
| $L=2.615, \lambda=0.05$  | ARL   | 1.316 | 1.328 | 1.034 | 1 | 1.002 | 1 |
|                           | SDRL  | 4.136 | 5.008 | 0.76 | 0 | 0.045 | 0 |
| $L=2.437, \lambda=0.03$  | ARL   | 1.106 | 1.754 | 1.092 | 1 | 1 | 1 |
|                           | SDRL  | 1.44 | 13.135 | 1.886 | 0 | 0 | 0 |

In the state of out of control or there is a process shift, it is expected to detect changes in the standard deviation as soon as possible, so the ARL value must be as small as possible. Based on the results obtained in Table 3, the ARL value tend to be smaller than the SDRL value for all shifts and changes in the parameters $L$ and $\lambda$. The overall ARL value gets smaller when the shift in process variability gets larger, so does the parameter $L$ and $\lambda$ value get smaller.

5. Conclusion
In this study, robust estimators which has been given that are not sensitive to the presence of outliers and some value shifting process variability. The Robust control chart is applied to the quality characteristic data of the number of bacterial colonies contained in the outlier data. The EWMA control chart is a control chart that is sensitive to small shifts. In this Robust EWMA control chart, it can be proven that in the ARL or SDRL values, the greater the process progression value, the control graph will detect the shift faster, so the ARL and SDRL values are getting smaller. Likewise, the value of the parameter $L$ and $\lambda$ that affects the width of the control boundary, so that the number of out of control points increases with smaller parameter values. Furthermore, it can be stated the value of the parameter $L$ and $\lambda$ the optimum in some process shifts, both the average shift and process variability. For further research can use control chart performance measures with other approaches such as the Markov chain approach.
Acknowledgment
Thank you to reviewers who have provided useful advice, which led to improvements in this paper.

References
[1] Sheu SH, Tai SH, Hsieh YT and Lin TC 2009 Monitoring Process Mean and Variability with Generally Weighted Moving Average Control Charts *Computers & Industrial Engineering* 57(1) 401-407
[2] Chen G, Cheng SW and Xie H 2001 Monitoring Process Mean and Variability with One EWMA Chart *Journal of Quality Technology* 33(2) 223-233
[3] Sheu SH, Huang CJ and Hsu TS 2012 Extended Maximum Generally Weighted Moving Average Control Chart for Monitoring Process Mean and Variability *Computers & Industrial Engineering* 62(1) 216-225
[4] Sindhumol MR, Srinivasan MR and Gallo M 2016 Robust Control Charts Based on Modified Trimmed Standard Deviation and Gini’s Mean Difference *Journal of Applied Quantitative Methods* 11(3) 18-30
[5] Saeed N and Kamal S 2016 The EWMA Control Chart Based on Robust Scale Estimators *Pakistan Journal of Statistics and Operation Research* 12(4) 659-672
[6] Zwetsloot IM, Schoonhoven M and Does RJMM 2014 A Robust Estimator for Location in Phase I Based on an EWMA Chart *Journal of Quality Technology* 46(4) 302-315
[7] Abu-Shawiesh MO 2008 A Simple Robust Control Chart Based on MAD *Journal of Mathematics and Statistics* 4(2) 102-107
[8] Schoonhoven M and Does RJMM 2012 A Robust Standard Deviation Control Chart *Technometrics* 54(1) 73-82
[9] Tatum LG 1997 Robust Estimation of the Process Standard Deviation for Control Chart *Technometrics* 39 127-141
[10] Montgomery DC and Runger GC 2014 *Applied Statistics and Probability for Engineers 6th Edition* Singapore: John Wiley and Sons
[11] Rocke DM 1989 Robust Control Chart. *Technometrics* 31(2) 173-184
[12] Lax DA 1985 Robust of Scale: Finite-Sample Performance in Long-Tailed Symmetric Distributions *Journal of the American Statistical Association* 80(391) 736-741
[13] Rousseeuw PJ and Croux C 1993 Alternatives to the Median Absolute Deviation *Journal of the American Statistical Association* 88(424) 1273-1283
[14] SNI 01-2332.3-2006 2006 *Penentuan Angka Lempeng Total (ALT) pada Produk Perikanan* Badan Standardisasi Nasional Jakarta