Mixed Multivariate EWMA-CUSUM (MEC) Chart based on MLS-SVR Model for Monitoring Drinking Water Quality

M Mashuri\textsuperscript{1*}, H Khusna\textsuperscript{1}, Wibawati\textsuperscript{1}, and F D Putri\textsuperscript{1}
\textsuperscript{1}Department of Statistics, Institut Teknologi Sepuluh Nopember, Indonesia

*Email: m_mashuri@statistika.its.ac.id

Abstract. Monitoring the quality of drinking water needs to be conducted considering the important role of water in human life. Mixed Multivariate EWMA-CUSUM (MEC) chart is a multivariate control chart developed for observing the mean process. Based on the previous studies, this chart has better performance in detecting a shift in the process mean. In this research, the MEC is applied to observe the grade of drinking water. However, there is autocorrelation in drinking water data which lead to more false alarm occurred. Therefore, the Multioutput Least Square Support Vector Regression (MLS-SVR) model is employed to reduce or even remove the autocorrelation in the data. Using the optimal hyperparameter, the MLS-SVR algorithm produces the residuals of phase I with no autocorrelation. Those residuals are then used to form the MEC control charts. When the MEC is used to monitor the residual in phase I, there is no signal of out-of-control found. Further, in phase II, out-of-control observations are detected. The MEC chart can detect more signals out of control compared to the conventional Hotelling’s $T^2$ and Multivariate Exponentially Moving Average (MEWMA) charts.

Keywords: Mixed multivariate EWMA-CUSUM; water quality; autocorrelation; MLS-SVR

1. Introduction
The industrial world is increasingly experiencing fierce competition. One indicator of a country's progress can be seen from the development of its industrial sector. One way to improve the industrial sector is to ensure the quality of the items produced. Quality controls are the activities of ensuring whether policies or standards given can be reflected in the final results. In other words, efforts to maintain the quality of the items produced to comply with product specifications that have been established based on company policy [1].

The control chart is a tool to describe a quality characteristic that has been measured from the sample [2]. The multivariate control chart can be characterized into three types, namely Shewhart [3–6], Multivariate Cumulative Sum (MCUSUM) [7–10], and Multivariate Exponentially Weighted Moving Average (MEWMA) [11–13]. Jimoh Olawale Ajadi and Muhammad Riaz [14] proposed a control chart that combines MEWMA and MCUSUM called Mixed Multivariate EWMA-CUSUM (MEC). In its application, the MEC control chart consists of Mixed Multivariate EWMA-CUSUM Control Chart 1 (MEC1) and Mixed Multivariate EWMA-CUSUM Control Chart 2 (MEC2). MEC1 is formed from the MEWMA [15] transformation integrated into MCUSUM [16], while MEC2 is formed from the MEWMA transformation into Multivariate CUSUM 1 (MC1) [17].

In this research, MEC charts are applied to monitor the drinking water grade. However, it is found that the drinking water quality data has autocorrelation. Autocorrelation is a situation where the observations have a relationship [18]. However, the conventional multivariate control chart assumes that the data are uncorrelated [2]. The presence of autocorrelation leads to more false alarms occurring. The
Multi-output Least Square -Support Vector Regression (MLS-SVR) method [19] can be employed in minimizing or removing the presence of autocorrelation. Therefore, in this study, the quality control of the drinking water is performed using the combination between MLS-Modelling and MEC chart. The remaining parts of this paper are composed as follows: Section 2 presents the MLS-SVR algorithms. Section 3 shows the Mixed Multivariate EWMA-CUSUM (MEC) procedures. Section 4 presents the result and discussion. Finally, Section 5 displays a summary of this research.

2. Multi-output least square -support vector regression (MLS-SVR)
In this section, the MLS-SVR algorithms are presented. The basic Least Square SVR algorithm learns the mapping from multiple inputs to a single output. Let \( \mathbf{Y} = [y_i] \in \mathbb{R}^{n \times m} \), where \( i = 1, 2, \ldots, n \) is the sample size, and \( j = 1, 2, \ldots, m \) is the number of output variable. The MLS-SVR algorithm uses a kernel function to perform a nonlinear mapping to a higher dimension, \( h \) is a higher dimension.

MLS-SVR solves the problem by looking for parameters and \( \mathbf{W} = (w_1, w_2, \ldots, w_m) \in \mathbb{R}^{h \times m} \)
\[ \mathbf{b} = (b_1, b_2, \ldots, b_n)^T \in \mathbb{R}^n \] which can minimize the objective function with the following constraint:
\[
\begin{align*}
\text{minimize} & \quad J(\mathbf{w}_0, \mathbf{V}, \Xi) = \frac{1}{2} \left( \mathbf{w}_0^T \mathbf{w}_0 \right) + \frac{\gamma^*}{2m} \text{trace} \left( \mathbf{V}^T \mathbf{V} \right) + \frac{\gamma^*}{2} \text{trace} \left( \Xi^T \Xi \right), \\
\text{with constrain} & \quad \mathbf{Y} = \mathbf{Z}^T \mathbf{W} + \text{repmat}(\mathbf{b}^T, n, 1) + \Xi.
\end{align*}
\]

The optimization in the previous equation can be made into a Lagrange function as follows:
\[
L(\mathbf{w}_0, \mathbf{V}, \Xi, \mathbf{A}) = J(\mathbf{w}_0, \mathbf{V}, \Xi) - \text{trace} \left( \mathbf{A}^T \left[ \mathbf{Z}^T \mathbf{W} + \text{repmat}(\mathbf{b}^T, n, 1) + \Xi - \mathbf{Y} \right] \right).
\]

The function of the MLS-SVR decision can be obtained as follows:
\[
\hat{y}(\mathbf{x}) = \varphi(\mathbf{x})^T \mathbf{W} + \left( \mathbf{b} \right)^T = \varphi(\mathbf{x})^T \text{repmat}(\mathbf{w}_0, 1, m) + \varphi(\mathbf{x})^T \mathbf{V} + \left( \mathbf{b} \right)^T \\
= \varphi(\mathbf{x})^T \text{repmat} \left( \sum_{i=1}^{m} \mathbf{a}_i, 1, m \right) + \frac{m}{\gamma} \varphi(\mathbf{x})^T \mathbf{Z} \left( \mathbf{a} \right) + \hat{b}^T \\
= \text{repmat} \left( \sum_{i=1}^{m} \sum_{j=1}^{n} \mathbf{a}_i \mathbf{k} (x, x_j), 1, m \right) + \frac{m}{\gamma} \sum_{i=1}^{m} \mathbf{a}_i \mathbf{k} (x, x_j) + \hat{b}^T.
\]

The optimal hyperparameters (\( \gamma^*, \gamma^+, \sigma \)) are selected using the grid search method with Mean Square Error (MSE) criterion. The kernel function used in this study is the Radial Basis Function (RBF).

3. Mixed multivariate EWMA-CUSUM (MEC)

3.1. Mixed multivariate exponentially weighted moving average – cumulative sum 1 (MEC1)
Let \( \mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_p \) is the sample quality characteristics to be monitored. MEC1 chart is developed by transforming the sample into MEWMA statistics [15] then integrating it with MCUSUM statistics [16] so that it can be expressed as follows:
\[
\text{MEC}_1 = \max \left( 0, \text{MEC}_{1,1} + (Z_t - \mu_\alpha) - k^* \right).
\]

where \( k^* \) is defined as follows:
\[
k^* = k \sqrt{\frac{\text{MEC}_{1,1} + (Z_t - \mu_\alpha)}{\left( \text{MEC}_{1,1} + Z_t - \mu_\alpha \right)^T \Sigma^{-1} (\text{MEC}_{1,1} + Z_t - \mu_\alpha)}}^2;
\]

\[
k \in \mathbb{R}^+, k > 0.
\]

If \( k \geq \text{MEC}_{1,1} + (Z_t - \mu_\alpha) \), the MEC1 will have a value of zero or negative. Thus, for simplification, the MEC1 statistics can be written as follows:
\[
\text{MEC}_1 = \text{MEC}_{1,1} + (Z_t - \mu_\alpha) - k^*.
\]
Finally, the statistics $MEC_1$ plotted on the control chart is calculated from the following equation:

$$MEC_1 = MEC_i \cdot \Sigma n_i MEC_i.$$  

The process is declared statistically uncontrolled if the value of $MEC_1 > h_i$. The $h_i$ is the control limit obtained from the Monte Carlo simulation for the certain ARL$_0$ value and it is related to the number of quality characteristics $p$ [14].

3.2. Mixed multivariate exponentially weighted moving average – cumulative sum 2 ($MEC2$)

The $MEC2$ chart is a combination of MEWMA statistics [15] with MC1 statistics [17]. The statistics of $MEC2$ can be written as follows:

$$MEC2 = \max \left\{ 0, (S_i^T \Sigma^{-1} S_i)^{1/2} - k_i n_i \right\},$$

where $S_i = \sum_{m=i-n_i+1}^i (Z_i - \mu_0)$ and $n_i = \begin{cases} n_{i-1} + 1, & \text{if } MEC2_{i-1} > 0 \\ 1, & \text{if } MEC2_{i-1} \leq 0. \end{cases}$

Furthermore, the $k_i$ is calculated using the following expression:

$$k_i = k \left( (\bar{x} - \mu_0)^T \Sigma^{-1} (\bar{x} - \mu_0) \right)^{1/2},$$

The process is declared statistically uncontrolled if the value of $MEC2 > h_i$, where the value of $h_i$ is the control limit obtained from the Monte-Carlo simulation for the certain ARL$_0$ value and it is related to the number of quality characteristics $p$ [14].

4. Result and discussion

4.1. Drinking-Water data

The data used is obtained from the drinking water company from August 1, 2020, to January 24, 2021. The daily data used will be classified into two phases. Phase I is on August 1 – October 31, 2020, and phase II is on November 1, 2020 – January 24, 2021. There are three variables analyzed are acidity, chlorine residual, and turbidity. From Figure 1, it can be seen that autocorrelation occurs in all the variables studied. Therefore, to reduce or remove the autocorrelation, the MLS-SVR is conducted in phase I.
Figure 1. ACF plot of full data for: a) acidity, b) chlorine residual, and c) turbidity

Figure 2. ACF plot of phase I residual for: a) acidity, b) chlorine residual, and c) turbidity
Figure 2 depicts the ACF plot of phase I residual of the MLS-SVR model. From all variables studied, there is no autocorrelation found. Thus, using the optimum hyperparameter of the MLS-SVR model, $\gamma^* = 2^{15}$, $\gamma^* = 2^4$, and $\sigma = 2^{-11}$ with an MSE of 0.0107, the autocorrelation can be removed, and the residual can be used in MEC charts.

4.2. Monitoring results for Phase I
The phase I data was taken from August 1 – October 31, 2020, and treated as the training data to find the optimal hyperparameter. Figure 3 presents the monitoring results in phase I for $MEC_1$ and $MEC_2$ charts. The $MEC_1$ and $MEC_2$ charts are constructed using $k=0.1$ and $\lambda=0$. From the results, it can be seen that all observations are statistically controlled.

![Figure 3](image)

**Figure 3.** Monitoring results for phase I from: a) $MEC_1$ chart and b) $MEC_2$ chart

4.3. Monitoring results for phase II
The monitoring results of the phase II control chart are displayed in this subsection. The phase II data was taken from November 1, 2020 – January 24, 2021, and treated as the testing data. Figure 4 shows the monitoring results for phase II using $MEC_1$ and $MEC_2$ charts. Similar to phase I, the $MEC_1$ and $MEC_2$ charts are constructed using $k=0.1$ and $\lambda=0.5$. From the monitoring results, it can be seen that all of the control charts send out-of-control signals.

![Figure 4](image)

**Figure 4.** Monitoring results for phase II from: a) $MEC_1$ chart and b) $MEC_2$ chart
4.4. Comparison with the other multivariate charts

In this subsection, the performance of the MEC charts is compared with the Hotelling’s $T^2$ and MEWMA charts. Figure 5 shows the monitoring results for Hotelling’s $T^2$ and MEWMA charts in phase II. According to the results, it can be seen that the Hotelling’s $T^2$ and MEWMA charts can detect the out-of-control signal. The comparison of out-of-control signals in phase II is displayed in Table 1. From the table, it can be concluded that both $MEC_1$ dan $MEC_2$ charts can detect more out-of-control signals compared to the other charts. It also can be concluded that the $MEC_1$ chart detects more out-of-control observations than the $MEC_2$ chart.

![Figure 5. Monitoring results for phase II from: a) hotelling’s $T^2$ chart and b) MEWMA chart](image)

**Table 1. Comparison of out-of-control signal in phase II**

| Control Charts   | Number of out-of-control signals |
|------------------|----------------------------------|
| Hotelling’s $T^2$ | 2                                |
| MEWMA            | 5                                |
| $MEC_1$          | 11                               |
| $MEC_2$          | 10                               |

5. Conclusion and future works

In this work, the drinking water quality data is monitored using the Mixed Multivariate EWMA-CUSUM (MEC) chart. However, due to the presence of autocorrelation on the drinking water data, the conventional multivariate charts are not appropriate to be employed. The MLS-SVR method is employed to handle the data with autocorrelation. After finding the optimal hyperparameter of the RBF kernel ($\gamma' = 2^{15}$, $\gamma'' = 2^4$, and $\sigma = 2^{-11}$ with an MSE at about 0.0107), we calculate the residual of the drinking quality data for phases I and II. In phase I, both $MEC_1$ and $MEC_2$ charts cannot detect any out-of-control observation indicating that the processes are statistically in-control. Furthermore, in phase II, the $MEC_1$ and $MEC_2$ charts find some out-of-control observations. Compared to MEWMA and Hotelling’s $T^2$ charts, the proposed MEC charts detect more out-of-control samples. Furthermore, for this case, it also can be deduced that the $MEC_1$ chart detects more out-of-control observations than the $MEC_2$ chart.

6. References

[1] Heizer J 2016 *Operations management* (Pearson Education India)
[2] Montgomery D 2009 *Introduction to statistical quality control* (New York)
[3] Ahsan M, Mashuri M, Kuswanto H and Prastyo D D 2018 Intrusion Detection System using
Multivariate Control Chart Hotelling’s T2 based on PCA Int. J. Adv. Sci. Eng. Inf. Technol. 8 1905–11

[4] Mashuri M, Haryono H, Aksioma D F, Wibawati W, Ahsan M and Khusna H 2019 Tr (R2) control charts based on kernel density estimation for monitoring multivariate variability process Cogent Eng. 6 1665949

[5] Ahsan M, Mashuri M, Kuswanto H, Prastyo D D and Khusna H 2018 T2 Control Chart based on Successive Difference Covariance Matrix for Intrusion Detection System Journal of Physics: Conference Series vol 1028 (IOP Publishing) p 12220

[6] Ahsan M, Mashuri M and Khusna H 2018 Intrusion Detection System Using Bootstrap Resampling Approach Of T2 Control Chart Based On Successive Difference Covariance Matrix J. Theor. Appl. Inf. Technol. 96 2128–38

[7] Woodall W and Ncube M 1985 Multivariate CUSUM Quality-Control Procedures Technometrics 27 285–92

[8] Khusna H, Mashuri M, Ahsan M, Suhartono S and Prastyo D D 2018 Bootstrap Based Maximum Multivariate CUSUM Control Chart Qual. Technol. Quant. Manag.

[9] Khusna H, Mashuri M, Suhartono, Prastyo D D, Lee M H and Ahsan M 2019 Residual-based maximum MCUSUM control chart for joint monitoring the mean and variability of multivariate autocorrelated processes Prod. Manuf. Res. 7 364–94

[10] Arkat J, Niaki S T A and Abbasi B 2007 Artificial neural networks in applying MCUSUM residuals charts for AR(1) processes Appl. Math. Comput. 189 1889–901

[11] Haq A and Khoo M B C 2019 An adaptive multivariate EWMA chart Comput. Ind. Eng. 127 549–57

[12] Khusna H, Mashuri M, Suhartono S, Prastyo D D and Ahsan M 2018 Multioutput least square SVR based multivariate EWMA control chart: The performance evaluation and application Cogent Eng.

[13] Zaidi F S, Castagliola P, Tran K P and Khoo M B C 2020 Performance of the MEWMA-CoDa control chart in the presence of measurement errors Qual. Reliab. Eng. Int.

[14] Ajadi J O and Riaz M 2017 Mixed multivariate EWMA-CUSUM control charts for an improved process monitoring Commun. Stat. Methods 46 6980–93

[15] Lowry C A, Woodall W H, Champ C W and Rigdon S E 1992 A Multivariate Exponentially Weighted Moving Average Control Chart Technometrics 34 46–53

[16] Crosier R B 1988 Multivariate generalizations of cumulative sum quality-control schemes Technometrics 30 291–303

[17] Pignatiello J and Runger J 1990 Comparison of multivariate CUSUM charts J. Qual. Technol. 22 173–8

[18] Karaoğlan A D 2010 Regression control chart for autocorrelated data

[19] Xu S, An X, Qiao X, Zhu L and Li L 2013 Multi-output least-squares support vector regression machines Pattern Recognit. Lett. 34 1078–84