Quality of Water Production Process Using Mixed Multivariate EWMA-CUSUM (MEC) Control Chart

N H D Asmara, Wibawati*, M Ahsan, M Mashuri, H Khusna

Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, 60111, Indonesia

*E-mail: wibawati@statistika.its.ac.id

Abstract. Human life is highly dependent on clean and freshwater. Therefore, the quality of water used by humans must be properly monitored and controlled. Based on the regulations of the Ministry of Health, the quality of clean water can be monitored from the parameters of physics, chemistry, and microbiology. The main characteristics of water quality are turbidity level, organic substance, and chlorine residual which are correlated with each other. This research proposes the combination of MEWMA and MCUSUM charts called MEC control chart for monitoring the process mean of water quality with the variables used are turbidity level, organic substance, and chlorine residual. This charting method is divided into two-part, namely MEC1 and MEC2. Further, the water production process is characterized by dry and rainy seasons. Based on the analysis results, it is found that the optimum $\lambda$ in the production process during the dry and rainy season for the MEC1 control chart is 0.3. Meanwhile, the optimum $\lambda$ in the production process during the dry and rainy season for the MEC2 control chart is 0.1. In this condition, the MEC1 control chart is more sensitive than the MEC2 control chart because of the higher number of out of control observations. Moreover, the organic substance is the main issue causing the out of control observations during the dry and rainy season. In the capability process analysis, the process of water production during the dry and rainy season is said to be capable because the capability index of the turbidity level is larger than the specification limit. On the other hand, the water production process with organic substance and chlorine residual are not capable due to the lower capability index compared to the specification limit.

1. Introduction

Water is a natural material needed to meet the needs of human, plant, and animal life, and also a source of energy. Therefore, a clean water supply is very much needed [1]. The indicator that water has been contaminated is that there are changes that can be observed, such as changes in water color, temperature, pH, smell, and taste [2]. Also, the indicator of other clean water is water free of organic and anorganic pathogens but contains chemicals that are needed by the human body [3]. The need for clean water will increase along with the increase in population in an area.

The quality of drinking water must be considered so it must always be safe and healthy when consumed by the community. Drinking water quality can be controlled using statistical analysis, namely using statistical quality control or statistical process control (SPC). Statistical quality control is a statistical tool that can control products in the production process, one of which can use control graphs. A control chart is a statistical control chart measurement tool that can monitor whether a production process is in control or not [4]. The determining characteristics of water quality include turbidity, organic
substance, and residual chlorine. These three quality characteristics have a relationship with one another. The higher the turbidity, the more chlorine addition will be, likewise when the levels of organic substances are higher, the remaining chlorine will also be lower.

Based on the previous description, in this study, the quality of drinking water will be carried out using the MEC proposed by Ajadi and Riaz in 2016 [5]. The MEC control chart is a combination of MEWMA and MCUSUM control chart used to monitor the process mean.

2. Literature
This section will explain the literature used in this research.

2.1. Shapiro-Wilk’s Test
Before carrying out quality control, it is necessary to test the normal distribution of the data used to determine observational data on whether to follow the normal distribution or not. This study will use Shapiro-Wilk’s test to check the multivariate normal distribution. An observation is said to have a multivariate normal distribution when it has a density function as in equation (1) [6].

\[
f(x) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)}. \tag{1}\]

The \(x\) is the observations data and \(\Sigma\) is the variance-covariance matrix of \(x\). The following hypothesis and test statistics are for the Shapiro-Wilk’s test.

Hypothesis:
\(H_0\): The data is following the multivariate normal distribution
\(H_1\): The data is not following the multivariate normal distribution

Test statistics:
\[W^* = \frac{1}{p} \sum_{j=1}^{p} W_{x_j}, \tag{2}\]

with
\[W_{x_j} = \frac{\left[ \sum_{i=1}^{n} a_i x_{(i)} \right]^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}. \tag{3}\]

Shapiro-Wilk’s statistics is represented by \(W^*\) and the individual statistics is represented by \(W_{x_j}\). The \(H_0\) will be rejected if \(W^* < c_{\alpha, p}\) or p-value < \(\alpha\), it means the observations data is not from a sample of multivariate normal distribution.

2.2. Bartlett Test
The Bartlett test is used to determine whether there is a variance relationship between multivariate variables. If \(X_1, X_2, \ldots, X_k\) are independent or not related, then the correlation matrix between variables is the same as the identity matrix [7]. The Bartlett test is used when the data follows a normal distribution. Bartlett sphericity is defined in the hypothesis and test statistics in equation (4).

Hypothesis:
\(H_0\) : \(R = I\) (there is no correlation between variables).
\(H_1\) : \(R \neq I\) (there is a correlation between variables).

Test statistics:
\[\chi^2 = -\left[n - 1 - \frac{2p + 5}{6}\right] \ln |R|, \tag{4}\]
Where \( n \) is the number of observations and \( \mathbf{R} \) is the correlation matrix for each variable. The \( H_0 \) will be rejected if \( \chi^2 > \chi^2_{\alpha, \frac{1}{2}(n-1)} \) or if p-value < \( \alpha \).

### 2.3. Mixed Multivariate Exponentially Weighted Moving Average – Cumulative Sum (MEC) Control Chart

Quality control or statistical process control (SPC) is an activity to solve a problem to get a stable process and able to reduce the variability in the process. Also, quality control is used to describe controllable and uncontrollable variability. There are several tools to control quality, one of them is a control chart [4]. Quality control that involves two or more interrelated characteristics can be monitored using multivariate control charts, one of the multivariate control chart is the MEC control chart. The MEC control chart is a combination of the MEWMA and MCUSUM control charts and is divided into two, namely MEC1 and MEC2. The MEWMA control chart is written in equation (5), the covariance matrix in equation (6), and statistics in equation (7).

\[
\mathbf{Z}_i = \lambda \mathbf{X}_i + (1-\lambda) \mathbf{Z}_{i-1}, \quad (5)
\]

\[
\Sigma_z = \frac{\lambda}{2-\lambda} \Sigma, \quad (6)
\]

and

\[
T_{i}^2 = \mathbf{Z}_i^T \Sigma_z^{-1} \mathbf{Z}_i. \quad (7)
\]

The \( T_i^2 \) statistics is the point plotted on the control chart. If \( T_i^2 \) is higher than the upper control limit \( h \) then the process is said to be not statistically controlled. MCUSUM control chart has statistics value as in equation (8). MCUSUM is developed into an MC1 control chart based on the mean square as in equation (9) and the value of the statistic as in equation (10).

\[
C_i = \left( \mathbf{S}_{n+1} + \mathbf{X}_i \right)^T \Sigma^{-1} \left( \mathbf{S}_{n+1} + \mathbf{X}_i \right)^{\frac{1}{2}}, \quad (8)
\]

\[
\mathbf{S}_i = \sum_{j=n_1+1}^{i} \left( \mathbf{X}_j - \mu_0 \right), \quad (9)
\]

and

\[
V_i = \max \left\{ 0, \left( \mathbf{S}_i^T \Sigma^{-1} \mathbf{S}_i \right)^{\frac{1}{2}} - kn \right\}. \quad (10)
\]

MEC1 control chart is a transformation of the MEWMA and MCUSUM control chart as in equation (11). The value of \( k^* \) can be calculated with equation (12) and the MEC1 can be obtained in equation (13).

\[
\text{MEC}_i = \max \left( 0, \text{MEC}_{i-1} + (\mathbf{Z}_i - \mu_0) - k^* \right) \quad (11)
\]

\[
k^* = k \frac{\text{MEC}_{i-1} + \mathbf{Z}_i - \mu_0}{\left[ \left( \text{MEC}_{i-1} + \mathbf{Z}_i - \mu_0 \right)^T \Sigma_z^{-1} \left( \text{MEC}_{i-1} + \mathbf{Z}_i - \mu_0 \right) \right]^{\frac{1}{2}}}, \quad (12)
\]

and

\[
\text{MEC}_i = \text{MEC}_{i-1} + (\mathbf{Z}_i - \mu_0) - k^*. \quad (13)
\]

After getting the MEC\(_i\), the next step is to find the \( \text{MEC}_l \) statistics with equation (14). The production process is said to be statistically uncontrolled if there is an observation that higher than the \( h \) control limit.

\[
\text{MEC}_l = \text{MEC}_l^T \Sigma_z^{-1} \text{MEC}_l. \quad (14)
\]

The second control chart is the MEC2 control chart, which is a combination of the MEWMA control chart with MC1. In this approach, the MEWMA statistics will be transformed into the cumulative sums vector of MC1. Furthermore, the cumulative sums variable is defined in equation (15) which will be entered into the \( \text{MEC}_2 \) statistic in equation (16) [5].
The production process is said to be not statistically controlled if the value of \( MEC^2 \), higher than the \( h \) control limit.

2.4. Process Capability Analysis

Capability analysis is performed to estimate the capability of a process. A capability index is needed to determine the capability of a process. A production process is said to be capable if it has a capability index of more than 1 [8]. The capability index that is used in this study is \( P_{pk} \) for the individual observations and \( MP_{pk} \) for the multivariate observations because the production process is not statistically controlled. The \( P_{pk} \) and \( MP_{pk} \) can be seen in equations (17) and (18).

\[
P_{pk} = \min \left( \frac{UCL - \mu}{3\sigma}, \frac{\mu - LCL}{3\sigma} \right),
\]

and

\[
MP_{pk} = \sum_{k=1}^{p} W_k \sum_{k=1}^{p} X_{pk}.
\]

\( W_k \) shows the weights of the \( k \)-th quality characteristics, provided by \( \sum_{k=1}^{p} W_k = 1 \) [9].

3. Result

This section will present the result of the water production monitoring process.

3.1. Multivariate Normal Test

Three characteristics that determine the quality of production water used are turbidity, organic substance, and chlorine residual, so it is necessary to test the multivariate normal assumptions using the Shapiro-Wilk’s normality test. There are two data categories in this research, the dry season which has 242 observations, and the rainy season which has 122 observations. Using the \( \alpha = 0.05 \) and p-value = \( 3.800 \times 10^{-14} \) for the dry season and \( 0.1144 \) for the rainy season, so the decision taken is to reject \( H_0 \) for dry season data and accept \( H_0 \) for rainy season data. This means that the water production data during the dry season is not following a multivariate normal distribution, but the water production data during the rainy season is following the multivariate normal distribution.

3.2. Bartlett Test

Bartlett test is used to determine is there a correlation between variables used in this case. In testing the dependency assumption using the Bartlett test with a significant level of \( \alpha = 0.05 \) and the degrees of freedom (df) = 3 shows the results of the chi-square value (\( \chi^2 \)) = 38.447 for dry season data, \( \chi^2 = 26.597 \) during the rainy season data, and the p-value = 0.000. Because of the p-value < \( \alpha \), the decision taken is to reject \( H_0 \). It can be concluded that the three variables of water quality characteristics used has correlations to one another.

3.3. MEC Control Chart

Statistical quality control in this study is used to improve the quality of the water production process, one of the tools is the control chart. The recent research in 2016 about the control chart is proposed by Ajadi and Riaz which combines two control charts namely MEC control chart. MEC control chart is
used to monitor the shifts in the process mean. this control graph has two types, namely MEC1 and MEC2 control chart.

3.3.1. MEC1 Control Chart
MEC1 control chart is the combination of MCUSUM and MC1 control chart used to monitor the shifts in the process mean. In this research, the monitoring process using MEC1 control chart is separated for the dry season and rainy season data. The $\lambda$ used in this case are 0.1, 0.2, and 0.3 for each season, because if the weight value is greater than the result will be biased [10]. Figure 1 is the result of monitoring the water production process using MEC1 during the dry season.

![Figure 1](image1.png)

(a) (b) (c)

**Figure 1.** MEC1 Control Chart for Dry Season Data using (a) $\lambda = 0.1$, (b) $\lambda = 0.2$, and (c) $\lambda = 0.3$

Based on Figure 1, monitoring using MEC1 control chart and three values of $\lambda$ shows that the water production process is not statistically controlled because some observations are out of the upper control limit for each $\lambda$. For the values of $\lambda = 0.1$, $\lambda = 0.2$, and $\lambda = 0.3$, it can be found sequentially that there are 96, 56, and 126 out of control observations. The optimum $\lambda$ is chosen by a large number of out of control observations. Using the value of $\lambda = 0.3$ has the most number of out of control, so $\lambda = 0.3$ was chosen as the optimal $\lambda$ because it is more sensitive in detecting small shifts in the process mean. After the monitoring process for dry season data, the monitoring process using MEC1 control chart with a weighted value 0.1, 0.2, and 0.3 for rainy season data is in Figure 2.
Figure 2. MEC1 Control Chart for Rainy Season Data using (a) $\lambda = 0.1$, (b) $\lambda = 0.2$, and (c) $\lambda = 0.3$

Figure 3. MEC2 Control Chart for Dry Season Data using (a) $\lambda = 0.1$, (b) $\lambda = 0.2$, and (c) $\lambda = 0.3$
According to the result of monitoring the water production process for rainy season data as in Figure 2, there are some observations out of the upper control limit. The weighted values of 0.1, 0.2, and 0.3 have the out-of-control observations as 40, 37, and 46 observations, so that can be concluded that the water production process in the rainy season is not statistically controlled. The monitoring using weighted value 0.3 has the highest number of out of control observations, which means $\lambda = 0.3$ is chosen as the optimum weighted value for monitoring the water production process using MEC1 during the rainy season.

3.3.2. MEC2 Control Chart

After the monitoring process using MEC1 control chart for both dry and rainy season data, this step is monitoring the water production process data using MEC2 control chart using the same weighted value as MEC1 control chart. Figure 3 below is the result of monitoring the water production process during the dry season using MEC2 for a weighted value of 0.1, 0.2, and 0.3.

Figure 3 shows that there are some observations out of the upper control limit. Knowing that the monitoring process for dry season data using MEC2 with $\lambda = 0.1$ has 79 out of control observations, and using the weighted value of 0.2 and 0.3 has 54 and 17 out of control observations. It can be concluded that monitoring using MEC2 for the dry season is not statistically controlled. The MEC2 control chart for dry season with $\lambda = 0.1$ has the highest number of out of control observations, so it has chosen as the optimal weighted value because it indicates the more sensitive performance. The monitoring of the water production process using MEC2 for rainy season data can be seen in Figure 4.

![Figure 4. MEC2 Control Chart for Rainy Season Data using (a) $\lambda = 0.1$, (b) $\lambda = 0.2$, and (c) $\lambda = 0.3$](image)

Based on the result in Figure 4 can be seen that the monitoring with $\lambda = 0.1$ has 35 observations as the highest number of out of control observations, so $\lambda = 0.1$ has chosen as the optimal weighted value for the monitoring process of water production during the rainy season using MEC2 control chart.
After knowing the monitoring result using MEC1 and MEC2 control charts for dry and rainy season data, the next step is to compare the MEC1 and MEC2 control charts to find the appropriate control chart for the water production data in this study.

### 3.4. Comparison between MEC1 and MEC2

According to the result of monitoring the water production process using MEC1 and MEC2, Table 1 below shows the result comparison to select the more sensitive control chart between MEC1 and MEC2 by the number of out of control observations.

**Table 1. Comparison of the Number of Out of Control Observations**

| Season | $\lambda$ | Number of Out of Control Observations |
|--------|-----------|---------------------------------------|
|        |           | MEC1 | MEC2 |
| Dry    | 0.1       | 96   | 79   |
|        | 0.2       | 56   | 54   |
|        | 0.3       | 126  | 17   |
| Rainy  | 0.1       | 40   | 35   |
|        | 0.2       | 37   | 30   |
|        | 0.3       | 46   | 2    |

![MEC1 Control Chart for Dry Season Data](image)

![MEC1 Control Chart for Rainy Season Data](image)

**Figure 5.** MEC1 Control Chart for Dry Season Data for a combination of variables Turbidity and Organic Substance (a), Turbidity and Chlorine Residual (b), Organic Substance and Chlorine Residual (c)
Based on Table 1, shows that in the same weighted values in the dry season and rainy season, the MEC1 control chart has more out of control observations than the MEC2 control chart. Therefore, in the case of the water production data, the MEC1 control chart is more sensitive if it is compared to the MEC2 control chart. As mentioned by Ajadi and Riaz, the MEC1 control chart will be better than MEC2 control chart. So, the next step in this study will use MEC1 control chart to identify which variable causes the most out of control observations.

3.5. Identify Out Of Control
After the monitoring process, there are still many out of control observations that cause the production process is not statistically controlled. So, in this section, the variable which causes the out of control observation will be identified. The identity process of out-of-control observations using MEC1 control chart with the optimal weighted value $\lambda = 0.3$ for every combination of variables such as turbidity and organic substances, turbidity and chlorine residual, organic substances, and chlorine residual. The result of the identified process is shown in Figure 5 below.

Figure 5 shows that the monitoring process using $\lambda = 0.3$ and upper control limit 18.720 has two combinations of variables that have the highest number of out of control, namely the combination of variables turbidity and organic substance with 35 out of control observations and also the combinations of variables organic substance and chlorine residual with 84 out of control observations. Based on the results of the identification with the three combinations, it can be concluded that organic substances are a variable that contributes the most cause of out of control observations for dry season data. Figure 6 below is the result of the identification of the most cause of out-of-control observations of the water production process during the rainy season.

Figure 6. MEC1 Control Chart for Rainy Season Data for a combination of variables Turbidity and Organic Substance (a), Turbidity and Chlorine Residual (b), Organic Substance and Chlorine Residual (c).

Based on the result in Figure 6 known that only two combinations of variables have out of control observations, there are combinations of turbidity and organic substance with 53 out of control
observations, and combinations of organic substance and chlorine residual with 40 out of control observations, so it can be concluded that the variable of organic substances is a variable that has a big contribution to the out of control observations in the rainy season. Both dry and rainy season has organic substance as the cause of out of control observations. According to the discussion with the quality control specialist in the water production company, it was caused by the water content and the condition of the filter which is quite old.

3.6. Process Capability Analysis

The process capability analysis in this case is used to measure the performance capability of the water production process. A process is said to be capable when the entire product produced is within the specification limit, in this study the specification limit is 1. In this study, the index used to measure the capability of the multivariate process is $M_{Ppk}$ with the weight used is 0.333 for each variable because it has the same contribution, while univariate uses the index value of $P_k$. The $P_k$ and $M_{Ppk}$ index is used due to the data were not statistically controlled. Table 2 below is the results of the process capability analysis, both univariate and multivariate observations.

| Table 2. Process Capability Analysis |
|--------------------------------------|
| Variable                           | Dry Season | Rainy Season |
|                                    | $P_{pk}$   | $MP_{pk}$    | $P_{pk}$   | $MP_{pk}$ |
| Turbidity                          | 1.18       | 1.05         |
| Organic Substance                  | 0.43       | 0.69         | 0.25       | 0.51      |
| Chlorine Residual                  | 0.47       | 0.24         |

Based on Table 1, the $P_{pk}$ index for both dry and rainy seasons with the variable of turbidity is more than 1. It means that the water production process with a variable of turbidity is capable because it has an index within the specification limits. On other hand, the water production process with the variable of organic substance and chlorine residual is said to be not capable because the $P_{pk}$ index for both dry and rainy season is less than 1. In the multivariate process using all the variables, based on $M_{Ppk}$ which has a value of less than 1, so the water production process is not capable according to the specification limit.

4. Conclusion

Based on the monitoring process of water production using MEC1 and MEC2 control chart has not been statistically controlled, but MEC1 is more sensitive than MEC2 because MEC1 has more out of control observations. Based on the monitoring using MEC1 control chart with optimum $\lambda=0.3$ has the most out-of-control observations, so in the identification, the process used $\lambda=0.3$. In the identification process using MEC1 control chart, the main characteristics cause the process has not been statistically controlled during dry and the rainy season is an organic substance. Based on the result of process capability analysis the water production process is not capable because the water production process with organic substance and chlorine residual are not capable due to the lower capability index compared to the specification limit, but the water production process is capable with the turbidity level because of the capability index upper the specification limit.

References

[1] Suseno N V and Widyastuti M 2017 Analisis Kualitas Air PDAM Tirta Manggar Kota Balikpapan J. Bumi Indonesia 6 pp. 1-8.

[2] Indarsih W, Suprayogi S and Widyastuti M 2011 Kajian Kualitas Air Sungai Bedog Akibat Pembuangan Limbah Cair Sentra Indonesia Batik Desa Wijirejo J. Majalah Geografi Indonesia, 25 pp. 40-54 2011.
[3] Sumantri B and Parwiyanto H 2017 *Kualitas Pelayanan Perusahaan Daerah Air Minum (PDAM) Kabupaten Sragen J. Wacana Publik* 1 pp. 11-24.

[4] Heizer J, Render and Chuck M 2017 *Principles of Operation Management*, New York: Pearson Education Limited.

[5] Ajadi J O and Riaz M 2016 *Mixed Multivariate Ewma-Cusum Control Charts for an Improved Process Monitoring* J. Communications in Statistics - Theory and Methods pp. 1-37.

[6] Johnson R A and Wichern D 2007 *Applied Multivariate Statistical Analysis* United States of America: Prentice Hall.

[7] Morrison D F 2005 *Multivariate Statistical Methods* United States of America: McGraw-Hill, Inc.

[8] Montgomery D C 2013 *Introduction to Statistical Quality Control* United State of America: John Wiley & Sons, Inc.

[9] Raissi S 2009 *Multivariate Process Capability Indices on The Presence of Priority for Quality Characteristics* Journal of Industrial Engineering International pp. 27-36.

[10] Hawkins D and Maboudou-Tchao E 2008 *Multivariate Exponentially Weighted Moving Covariance Matrix* J. Technometrics pp. 55-166.