Multivariate Exponentially Weighted Moving Average (MEWMA) and Multivariate Exponentially Weighted Moving Variance (MEWMV) Chart Based on Residual XGBoost Regression for Monitoring Water Quality

Nur Sulistiawanti, Muhammad Ahsan, and Hidayatul Khusna

Abstract—**Water Treatment Plant (WTP) in Palopo City, Indonesia, is one of the national companies engaged in the provision of clean water services for the people of Palopo City. There are 5 WTP managed in Palopo City. Testing the quality of production water at each WTP uses several interrelated quality characteristics, including turbidity, pH, and residual chlorine. The data used is obtained from the results of testing the quality of production water at each WTP in July-December 2021 for a daily period. In this study, the method used to control the quality of production water at each WTP is the Multivariate Exponentially Weighted Moving Average (MEWMA) control chart to control the process mean and the Multivariate Exponentially Weighted Moving Variance (MEWMV) control chart to control process variability based on residual XGBoost regression to overcome autocorrelation. In phase I, the process has been statistically controlled for both the variance and the process mean. The weighting values for the process variability are ω = 0.4 and λ = 0.4, and for the process mean is λ = 0.3. In phase 2, based on the data analyzed only WTP V in Batupapan whose process variability has been statistically controlled. The other WTPs in phase 2 have not been statistically controlled for both the variability and the process mean.**

Index Terms— **Autocorrelation, Water Quality, MEWMA, MEWMV, XGBoost Regression**

I. INTRODUCTION

Water is one of the elements of life that has an important role in the life of living things. Water that is safe, clean, healthy, and not polluted is necessary to maintain the health of the general public who consumes the water. Currently, surface water sources such as rivers are mostly polluted due to natural processes (rain) or due to irresponsible human activities, causing a decrease in water quality and ultimately unfit for direct use or consumption. Therefore, before using water, it must first go through a processing process from raw water to become clean water that is feasible and safe.

The Tirta Mangkaluku Drinking Water Company (PAM) in Palopo City is one of the Regional Owned Enterprises (BUMD) owned by the Palopo City government. PAM Tirta Mangkaluku Palopo City is a company engaged in the provision and service of clean water for communities around Palopo City. PAM Tirta Mangkaluku Palopo City currently manages five Drinking Water Treatment Plants (WTP), namely WTP I Abdul Majid, WTP II Magandang, WTP III Battang, WTP IV Latuppa, and WTP V Batupapan, with a total capacity of 550 L/d.

Each WTP PAM Tirta Mangkaluku Palopo City has several parameters or quality characteristics that are used to determine the quality of production water, some of which are turbidity, pH, and residual chlorine. These three quality characteristics need to be monitored every day because they are wary of water pollution. In controlling the quality of production water, the specification limits for turbidity and pH refer to the Regulation of the Minister of Health of the Republic of Indonesia Number 492 of 2010, while the remaining chlorine refers to the Regulation of the Minister of Health of the Republic of Indonesia Number 736 of 2010 [1].

Product quality is a major factor that is non-negotiable by the company so that it becomes a benchmark for an item or service with certain standards that make an item or service recognized as having its characteristics and characteristics. Quality control has an important role in the production process because a good production process will produce good products. In some production processes, it is often found that quality characteristics are more than one and influence each other between quality characteristics.

In general, based on the process used to monitor quality, there are two types of control charts, namely control charts to detect mean shifts and process variability. In some production processes, it is often found that quality characteristics are more than one and influence each other between quality characteristics. The water production process at PAM Tirta Mangkaluku is a company that has more than one quality characteristic and is interconnected.

Control chart can be used to monitor the quality of the process [2]. There are two types of control charts namely,

Manuscript received August 9, 2022; revised April 5, 2023.

N. Sulistiawanti is undergraduate student in Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, 60111 Indonesia. (e-mail: nursululistiawanti.18062@mhs.its.ac.id).

M. Ahsan is Assistant Professor in Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, 60111 Indonesia. (e-mail: muh.ahsan@its.ac.id).

H. Khusna is Assistant Professor in Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, 60111 Indonesia. (e-mail: hidayatul@its.ac.id).

the variable [3]–[5] and attribute charts [6], [7]. The multivariate control chart can be applied [8]. Quality control with conventional control charts on data that has autocorrelation will cause many false alarms to appear [9].

Autocorrelation is a condition in which successive observations have a relationship [9]–[11]. Autocorrelation between observational data can come from various factors, either from the operator or the process itself. If there is significant autocorrelation in a process, conventional control charts with the assumption of IID (Independent, Identic, and Normally Distributed) can still be used but are not effective. The presence of autocorrelation among observations leads to false alarms (type I error) and misleading conclusions about the control state of the process [11], [12].

The water production process at PAM Tirta Mangkaluku is a company that has more than one quality characteristic and is interconnected. Therefore, the study that will be discussed in this research is the control of production water quality in each PAM Tirta Mangkaluku Palopo City using the MEWMA control chart for controlling the process mean and MEWMV for controlling variability based on residual XGBoost regression to overcome autocorrelation, where this chart is used to detect small shifts of the production process with some quality characteristics.

II. Literature Review

A. Statistical Quality Control (SQC)

Quality control is a technique and management to measure the quality characteristics of a product or service, then it is used to compare the measurement results with product specifications that have been determined in order to take appropriate action to improve quality if differences in actual and standard performance are found.

B. Multivariate Normality Test

The multivariate normal distribution test is a normal distribution test with more than one variable number. This distribution is used in a group of data whose variables are mutually dependent. The following is the hypothesis of the multivariate normal distribution test.

Hypothesis:

H_0 : Data is normally distributed multivariate.

H_1 : Data is not normally distributed multivariate.

Test statistics:

$$d_i^2 = \left(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i\right)' \mathbf{S}^{-1} \left(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i\right),\ i = 1,2,\ldots,n \quad (1)$$

where d_i^2 is the distance value of the square of the i-th observation. The inverse value of the variance-covariance matrix **S** can be shown in the following equation.

C. Dependency Test

The existence of a correlation or relationship between independent quality variables must be met before conducting a multivariate analysis. To test the independence between variables can be done with the following Bartlet test.

Hypothesis:

H_0 : Quality variables are mutually independent.

H_1 : Quality variables are interdependent.

Test Statistics:

$$\chi^2_{hitung} = -\left(n - 1 - \frac{2p+5}{6}\right)\ln|\mathbf{R}|. \quad (2)$$

Information:

$|\mathbf{R}|$ = correlation matrix determinant

n = number of observations

p = number of variables

If $\chi^2_{hitung} \le \chi^2_{\alpha,v}$ or $p-value > \alpha$ with $\alpha = 0,05$ and degrees of freedom $v = p(p-1)/2$ then fail to reject H_0, so the quality variable is independent.

D. XGBoost Regression

Xtreme Gradient Boosting (XGBoost) is a method that combines boosting with gradient boosting. The model is built using the boosting method, namely by creating a new model to predict errors from the previous model. Such an algorithm is called gradient boosting because it uses gradient descent to minimize errors when constructing a new model [13].

The computational process of the XGBoost algorithm is shown in Figure 1 [14].

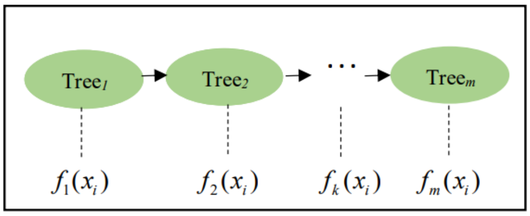

Fig. 1. Schematic Diagram of XGBoost Algorithm

The predicted value at step m is assumed to be $y_i^{(m)}$ with the formula $y_i^{(m)}$ can be seen in the following equation.

$$y_i^{(m)} = \sum_{k=1}^{m} f_k(x_i) \quad (3)$$

$f_k(x_i)$ describe the tree model. For y_i is obtained from the following calculation.

$$y_i^{(0)} = 0$$
$$y_i^{(1)} = f_1(x_1) = y_i^{(0)} + f_1(x_1)$$
$$y_i^{(2)} = f_1(x_1) + f_2(x_2) = y_i^{(1)} + f_2(x_2)$$
$$\vdots$$
$$y_i^{(m)} = y_i^{(m-1)} + f_m(x_i)$$
$$y_i^{(m)} = \sum_{k=1}^{m} f_k(x_i)$$

where:

$y_i^{(m)}$ = final tree model

$y_i^{(m-1)}$ = previously generated tree model

$f_m(x_i)$ = new model built

m = total number of models from base tree models

In the training process, each iteration minimizes the value of the loss function based on the initial function $F_0(x)$. In general, the gradient boosting algorithm has the following equation.

$$\{\gamma_m, h_m\} = \arg\min \sum_{m=1}^{M} L(y_i, f^{(m-1)}(x_i) + \gamma_m h_m(x_i)) \quad (4)$$

Observational data are not independent of each other will cause the conventional control chart to not function properly. Data that has autocorrelation will be overcome using the XGBoost regression model. Defined $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m$ is the output observation vector, which is autocorrelated multivariate data, where $\mathbf{y}_j = (y_{1j}, y_{2j}, \ldots, y_{nj})^T$ with $j = 1, 2, \ldots, m$ is the number of XGBoost outputs. Each output vector is assumed to have a significant Partial Autocorrelation Function (PACF) until $\text{lag}-p_1, p_2, \ldots, p_m$ so the input of the XGBoost model is as follows.

$$\mathbf{x} = (\mathbf{y}_{1,(i-1)}, \ldots, \mathbf{y}_{1,(i-p_1)}, \ldots, \mathbf{y}_{m(i-1)}, \ldots, \mathbf{y}_{m,(i-p_m)}) \quad (5)$$

XGBoost is a version of the Gradient Boosting Method (GBM) that is more efficient and scalable because it can complete various functions such as regression, classification, and ranking. XGBoost was first introduced at the Higgs Boson Competition, where in this competition the XGBoost method became the method most used by most teams. In addition to the competition, the XGBoost method is also a method that is widely used in machine learning competitions organized by Kaggle in 2015. XGBoost is a tree ensembles algorithm consisting of several classification and regression trees (CART). XGBoost algorithm performs optimization 10 times faster than other GBM implementations [14].

E. Multivariate Exponentially Weighted Moving Average (MEWMA) Control Chart

MEWMA control chart is a control chart for detecting the occurrence of small process mean shifts in multivariate ways. In the multivariate case, there was more than one quality characteristic (p=2). One of the advantages of the MEWMA control chart is that it is more robust against the normal distribution, which means that if the data does not meet the assumption of a multivariate normal distribution, the MEWMA control chart can still be used.

MEWMA is an extension of the univariate EWMA which is defined by the following equation.

$$\mathbf{Z}_i = \lambda\mathbf{X} + (1-\lambda)\mathbf{Z}_{i-1},\ 0 < \lambda \le 1 \text{ dan } i = 1, 2, ..., n \quad (6)$$

Information:

$\mathbf{Z}_i$ = weighted mean of all sample means before, with $\mathbf{Z}_0 = \mathbf{0}$

λ = weighting value, $0 < \lambda \le 1$

n = the number of samples of observations made

A small value is more effective for detecting a small shift in the process mean and a large value is more effective for detecting a large shift in the process mean. The percentage that is out of control is influenced by the selection of the Average Run Length (ARL) value. In detecting out-of-control signals on the MEWMA control chart, the following equation is used.

$$T_i^2 = \mathbf{Z}_i' \textstyle\sum_{\mathbf{Z}_i}^{-1} \mathbf{Z}_i > H \quad (7)$$

where H > 0 with the covariance matrix of $\mathbf{Z}_i$ as follows [4].

$$\textstyle\sum_{\mathbf{Z}_i} = \frac{\lambda}{2-\lambda}[1-(1-\lambda)^{2i}]\sum \quad (8)$$

The plot of control chart T_i^2 is a control chart with the upper control limit expressed by the value of H taken from the simulation for a certain ARL_0 [15]. While the lower control limit for the MEWMA control chart is equal to 0 because the value of T_i^2 which is always positive so that the minimum lower control limit of a positive value is 0. The process is said to be uncontrolled if there is a value of $T_i^2 > H$ [2].

F. Control Chart Multivariate Exponentially Weighted Moving Variance (MEWMV)

MEWMV control chart is a control chart that is used to control process variability with the advantage that it is more sensitive to detect small shifts in process variability. The following is the formula for the MEWMV control chart.

$$\mathbf{V}_n = \omega(\mathbf{x}_n - \mathbf{y}_n)(\mathbf{x}_n - \mathbf{y}_n)' + (1-\omega)\mathbf{V}_{n-1} \quad (9)$$

where is the weighting value of smoothing constant (ω) is $0 < \omega < 1$, with value $\mathbf{V}_0 = (\mathbf{x}_1 - \mathbf{y}_1)(\mathbf{x}_1 - \mathbf{y}_1)'$. The estimated value of $\mathbf{y}_i$ can be calculated using the following formula.

$$\mathbf{y}_n = \lambda\mathbf{x}_n + (1-\lambda)\mathbf{y}_{n-1} \quad (10)$$

where λ is the weighting value ranged $0 < \lambda \le 1$, with $\mathbf{y}_n$ is a natural estimate for the process mean at time n from the MEWMA control chart with a value of $\mathbf{y}_0 = \mathbf{0}$ [8].

In general, equations (7) and (8) can be written as follows.

$$\mathbf{V}_n = \sum_{i=1}^{n} \omega(1-\omega)^{n-1}(\mathbf{x}_i - \mathbf{y}_i)(\mathbf{x}_i - \mathbf{y}_i)' + (1-\omega)^n \mathbf{V}_0 \quad (11)$$

$$\mathbf{y}_n = \sum_{i=1}^{n} \lambda(1-\lambda)^{n-1}\mathbf{x}_i \quad (12)$$

The statistics $\text{tr}(\mathbf{V}_n)$ can be obtained by the following equation.

$$\begin{aligned}\text{tr}(\mathbf{V}_n) &= \text{tr}(\mathbf{X}'\mathbf{Q}\mathbf{X}) \\ &= \text{tr}(\mathbf{Q}\mathbf{X}\mathbf{X}')\end{aligned} \quad (13)$$

so that obtained

$$\begin{aligned}\text{tr}(\mathbf{V}_n) &= \sum_{j=1}^{n} q_{1j}\left(\sum_{k=1}^{p} x_{1k}x_{jk}\right) + \sum_{j=1}^{n} q_{2j}\left(\sum_{k=1}^{p} x_{2k}x_{jk}\right) + \cdots + \sum_{j=1}^{n} q_{nj}\left(\sum_{k=1}^{p} x_{nk}x_{jk}\right) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} q_{ij}\left(\sum_{k=1}^{p} x_{ik}x_{jk}\right)\end{aligned} \quad (14)$$

When $p = 1$, equality $\text{tr}(\mathbf{V}_n)$ will be the EWMA equation, which is the control chart for univariate data. When the process is in control, Equation 15 can be used to estimate E[$\text{tr}(\mathbf{V}_n)$].

$$\begin{aligned}\text{E}\left[\text{tr}(\mathbf{V}_n)\right] &= \sum_{i=1}^{n} q_{ii}\text{E}\left(\sum_{k=1}^{p} x_{ik}^2\right) + \sum_{i=1}^{n}\sum_{j\neq 1}^{n} q_{ij}\text{E}\left(\sum_{k=1}^{p} x_{ik}x_{jk}\right) \\ &= p\sum_{i=1}^{n} q_{ii} = p\ \text{tr(Q)}\end{aligned} \quad (15)$$

So, it can be shown the control limit for each n based on $\text{tr}(\mathbf{V}_n)$ which is given by the following equation.

$$\text{E}\left[\text{tr}(\mathbf{V}_n)\right] \pm L\sqrt{\text{Var}\left[\text{tr}(\mathbf{V}_n)\right]} = p\ \text{tr}(\mathbf{Q}) \pm L\sqrt{2p\sum_{i=1}^{n}\sum_{j=1}^{n} q_{ij}^2} \quad (16)$$

L is a constant whose value is obtained based on the number of observation variables or p, the weighting value of ω and λ which has been determined using a Monte-Carlo simulation to get the desired average run length (ARL0), which is 370 [9].

III. METHODOLOGY

A. Data source

The data used in this study is secondary data from the Process Control Section at the PAM Tirta Mangkaluku Office, Palopo City. Secondary data was obtained from the results of testing the quality of production water at each WTP in PAM Tirta Mangkaluku, Palopo City in July-December 2021 for a daily period.

B. Research variable

The variables used in this study consisted of three quality characteristics, namely turbidity, residual chlorine, and pH with the following specifications as shown in Table 1.

TABLE I
RESEARCH VARIABLE

Variable	Variable Name	Unit	Specification
X1	Turbidity	NTU	maximum 5
X2	Residual Chlor	mg/L	0.2 – 1
X3	pH		6.5 – 8.5

C. Analysis Step

The steps of analysis in this study are as follows.

1. Conduct literature studies to assist in the research process.
2. Formulate the problem so that it can answer the existing problems.
3. Defines the variables used.
4. Perform data retrieval from the variables used
5. Perform descriptive analysis to determine the characteristics of the data.
6. Perform a dependency test using the Bartlet Test. If there is no correlation between quality characteristics, it will be continued by making a univariate control chart.
7. Select data for phase I by looking at the most stable time series plot of all WTP and data for phase II.
8. Check autocorrelation using ACF and PACF graphs for each variable.
9. If there is a significant autocorrelation, it will be continued by looking for residual data using the XGBoost regression method to reduce or overcome the autocorrelation.
10. Perform residual multivariate normality test.
11. Create a MEWMV control chart with the following steps.
    a. Create a matrix **C** with the main diagonal being the weighted values ω and the matrix M in the form of a lower triangle with its elements being the weighting value of λ.
    b. Calculate the value of Q.
    c. Calculate value $\text{tr}(\mathbf{V}_n)$ and $\text{E}\left[\text{tr}(\mathbf{V}_n)\right]$.
    d. Calculate BKA and BKB for the MEWMV control chart.
    e. Create plots $\text{tr}(\mathbf{V}_n)$ with UCL and LCL based on each value of λ, ω, and L.
    f. Find the maximum observation point of $\text{tr}(\mathbf{V}_n)$.
    g. Select the optimum value of λ, ω, and L based on the difference $|\max(\text{tr}(\mathbf{V}_n)) - UCL|$ and the smallest difference in control limits.
12. Create a MEWMA control chart with the following steps.
    a. Calculate the $\boldsymbol{Z}_i$ vector.
    b. Calculate the covariance variance matrix based on $\boldsymbol{Z}_i$.
    c. Calculate the value T_i^2 of each observation.
    d. Create plots T_i^2 with UCL=H and LCL=0 based on each λ.
    e. Select the optimum value of λ.
13. Conclude and provide suggestions.

IV. ANALYSIS AND DISCUSSIONS

A. Data Characteristics

The characteristics of the results of testing the quality of production water in each WTP daily in July-December 2021 based on the three quality characteristics are the quality of the production water in WTP 1 which meets the requirements for the predetermined levels, only residual chlorine, the quality of the production water in WTP 2 which meets the requirements of the predetermined levels. only turbidity is determined, for the quality of production water at WTP 3 that meets the requirements, the levels that have been set are turbidity and pH, for the quality of production water at WTP 4 that meets the requirements, the levels that have been set are residual chlorine and pH, and the quality of production water at WTP 5 based on turbidity, pH, and organic substances have met the requirements of the specified levels.

B. Dependency Test

The dependency test was conducted to determine whether the quality characteristic variables were related or not. The significant level used is 0.05 and H_0 will be rejected if $\chi^2_{\text{hitung}} \leq \chi^2_{\text{tabel}}$ or *p-value* is less than, with a value of $\chi^2_{\text{tabel}} = 7.814$. Table 2 shows that the results of the dependency test between quality characteristics in each WTP obtained a *P-value* which is smaller than the value of and the value of χ^2_{hitung} which is larger than χ^2_{tabel} so that it can be concluded that the quality characteristics of each WTP have met the assumption of dependencies, which means that there is a relationship between the quality characteristic variables.

TABLE 2
DEPENDENCY TEST BETWEEN VARIABLES OF EACH WTP

WTP	*Chi-Square*	*P-value*	Conclusion
WTP 1 Abdul Majid	11.267	0.010	Dependent
WTP 2 Magandang	30.653	0.000	Dependent
WTP 3 Battang	30.419	0.000	Dependent
WTP 4 Latuppa	41.336	0.000	Dependent
WTP 5 Batupapan	11.106	0.011	Dependent

C. Autocorrelation Checking

To check that each quality characteristic data has no autocorrelation, it can be checked through the Auto Correlation Functions (ACF) plot of each production water quality characteristic. The process of purification and water treatment that is carried out continuously makes the water characteristics have autocorrelation as evidenced by the results of significant autocorrelation values in each WTP. Therefore, modeling with XGBoost regression can be used to overcome the autocorrelation in the production of water quality data. The use of MEWMV and MEWMA control charts based on the residual XGBoost model is the right choice for controlling the quality of the water production process at PAM Tirta Mangkaluku, Palopo City.

D. Residual Normality Test

The multivariate normality test aims to determine whether the observational data in Phases 1 and 2 follow a normal distribution or not. Phase 1 data uses residual data from July-September 2021 taken from several WTPs which are assumed to be in control processes based on the time series plot, where pH quality characteristics are taken from WTP 4, and residual quality characteristics of chlorine and turbidity are taken from WTP 2. Table 4 is a table of the results of the analysis of the multivariate normality test.

TABLE 3
MULTIVARIATE NORMALITY TEST

Phase	WTP	$d_i^2 < \chi^2_{p,\alpha}$	Information
Phase 1	Selected WTP	70%	Not normal
Phase 2	WTP I Abdul Majid	75%	Not normal
	WTP II Magandang	94%	Not normal
	WTP III Battang	86%	Not normal
	WTP IV Latuppa	65%	Not normal
	WTP V Batupapan	52%	Not normal

Based on Table 3, the distance $d_i^2 < \chi^2_{p,\alpha}$ which is more than 50%, it can be concluded that the multivariate normal distribution is not. Therefore, MEWMA and MEWMV control charts will be used which have robust properties against data abnormalities so that the research can be continued.

E. Modeling Using XGBoost Regression

Modeling using XGBoost is carried out on water production quality data in each phase to overcome autocorrelation. After modeling with XGBoost on the observation data, the results of the predicted Y values can be compared with the actual data values.

1) Modeling Using XGBoost Phase I Regression

In doing the modeling using XGBoost regression, first look for a significant lag in the PACF plot of each quality characteristic variable to determine the input data on the XGBoost model. In phase I, the data used is taken from several WTPs whose conditions are the most stable. Figure 2 is a time series plot of each quality characteristic taken from several WTPs whose conditions are the most stable.

Figure 2 shows a time series plot of each production water quality characteristic taken from several WTPs with the most stable conditions, namely pH in WTP IV, residual chlorine in WTP II, and turbidity in WTP II. Furthermore, the three characteristics in Phase I will be modeled using XGBoost to obtain the residual value. Figure 2 is a residual model from phase 1 data for each quality characteristic.

Based on Figure 3, it can be seen that the predicted value for each phase I variable from modeling with XGBoost shows a pattern similar to the actual data value of phase 1 using the maximum depth value of 3 for each variable with the number of iterations for the variable. pH and residual chlorine as much as 50 iterations, and 500 iterations for the turbidity variable. This means that the XGBoost model obtained can already be used on production water quality data. After all, it can follow the actual data pattern well, as evidenced by the RMSE values for the variables pH, residual chlorine, and turbidity of 0.113, 0.072, and 0.305, which means the prediction model is quite good because the RMSE value is close to 0.

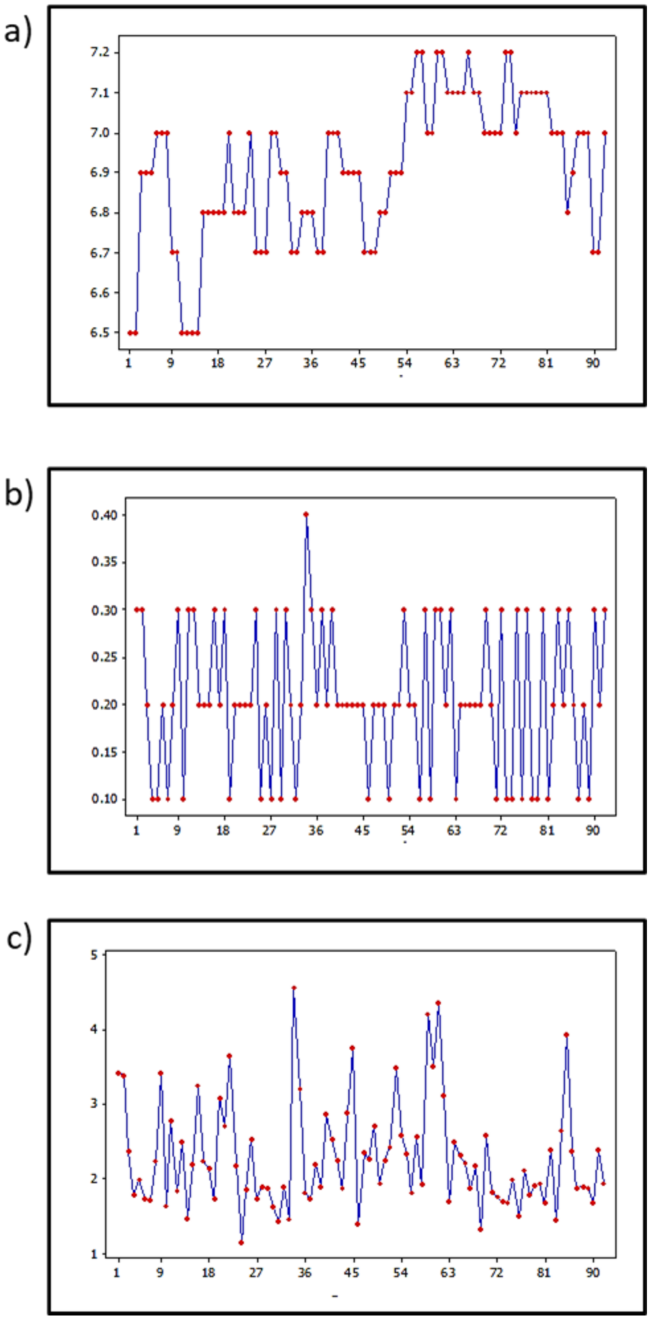

Fig. 2. Plot Time Series of Data Phase I
(a) pH, (b) Residual Chlorine, (c) Turbidity

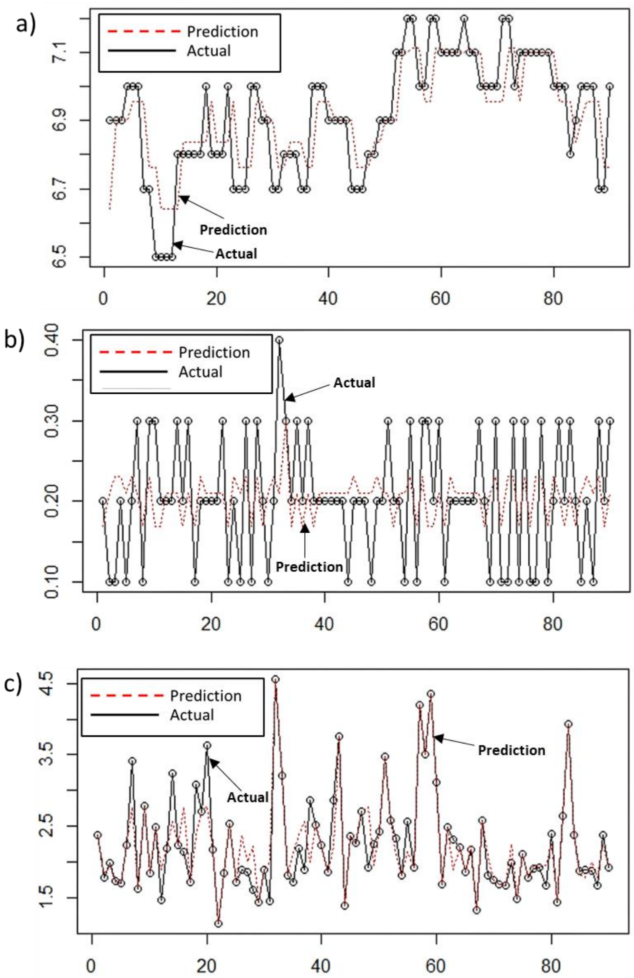

Fig. 3. Time Series Plot Actual Value and Predicted Value Phase I
(a) pH, (b) Residual Chlorine, (c) Turbidity

2) Modeling Using XGBoost Phase II Regression

In the same way, using XGBoost regression modeling on phase 2 data, the prediction value for each variable from the XGBoost model has a pattern similar to the actual data value for phase 2 in each WTP. Based on the XGBoost model that has been obtained, the residual value can already be used in water quality data because it can follow the actual data pattern well.

F. MEWMV Control Chart Application

MEWMV control chart for process variability control. In the application of the MEWMV control chart, the weighting values of ω and λ are less than 0.4, respectively, because according to Huwang [16] the weighting value can provide better performance in controlling process variability.

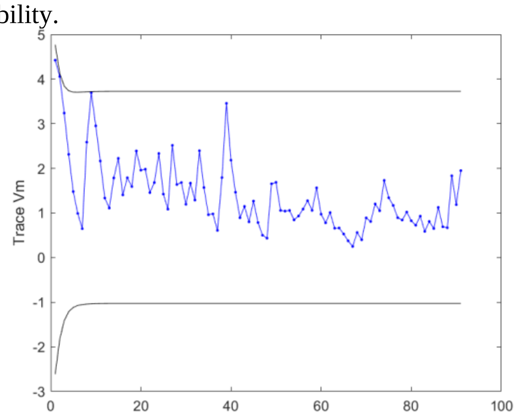

Fig. 4. MEWMV Control Chart of Phase 1 with ω=0.4 and λ=0.4

Based on Figure 4, information is obtained that controlling process variability with the MEWMV control chart using the values of ω=0.4, λ=0.4, and L=4.1875, the upper control limit is 3.7081 and the lower control limit is 1.0275. Based on the weighted value, it can be seen that all observation points are within the control limits, meaning that the variability of the production water process for phase 1 data with weights ω = 0.4 and λ =0.4 has been statistically controlled. Further, it will be continued with controlling process variability for phase 2 data in each WTP using the same weighting as phase 1. Figure 5 is a MEWMV control chart for production water quality in phase 2 data in each WTP with ω = 0.4 and λ = 0.4.

The process stability of the production water quality test results in each WTP is for process variability in phase 2 data obtained only WTP V Batupapan which has been statistically controlled, with weights ω = 0.4 and λ = 0.4.

Furthermore, it will be continued with controlling the process mean for phase 1 and phase 2 data. Table 5 shows the number of observations that are out of control limits in phase 2 in each WTP.

TABLE 4
NUMBER OF OUT-OF-CONTROL (OOC) OBSERVATIONS IN PHASE 2 OF THE MEWMV CHART

No	WTP	Number of OOC
1	Abdul Majid	4
2	Magandang	5
3	Battang	5
4	Latuppa	2
5	Batupapan	0

Table 4 shows the number of observations that are out of control in phase 2 in each WTP. Based on the table, it can be seen that only WTP V Batupapan does not have out-of-control data, so for process variability in phase 2, only WTP V Batupapan is statistically controlled.

G. MEWMA Control Chart Application

The MEWMA control chart has the main objective of controlling the process mean by taking individual samples. In general, the selection of the best weight for the MEWMA control chart is based on the minimum difference between the maximum observation point value and UCL and the width of the control limit. However, in this study, only 3 weights were compared, namely 0.1, 0.2, and 0.3. Of the three weights, the best weighting is determined, which is 0.3 because if you use a smaller weight, it will be oversensitive. If the chart is oversensitive, it can cause losses for the company. Thus, the best weighting is 0.3 which is thought to be neither sensitive nor oversensitive. Figure 6 is a MEWMA control chart on Phase 1 data with a weighting of 0.3.

The results of controlling the process mean using the MEWMA control chart at λ = 0.3 with UCL = 13.79, the results of the plot pattern T_i^2 there are no observation points that are outside the upper control limit (in control) so it can be concluded that the mean data processing in Phase 1 has been statistically controlled. Next will be controlling the Phase 2 data. Figure 6 is a MEWMA control chart for the quality of production water in phase 2 data in each WTP with a weight of 0.3.

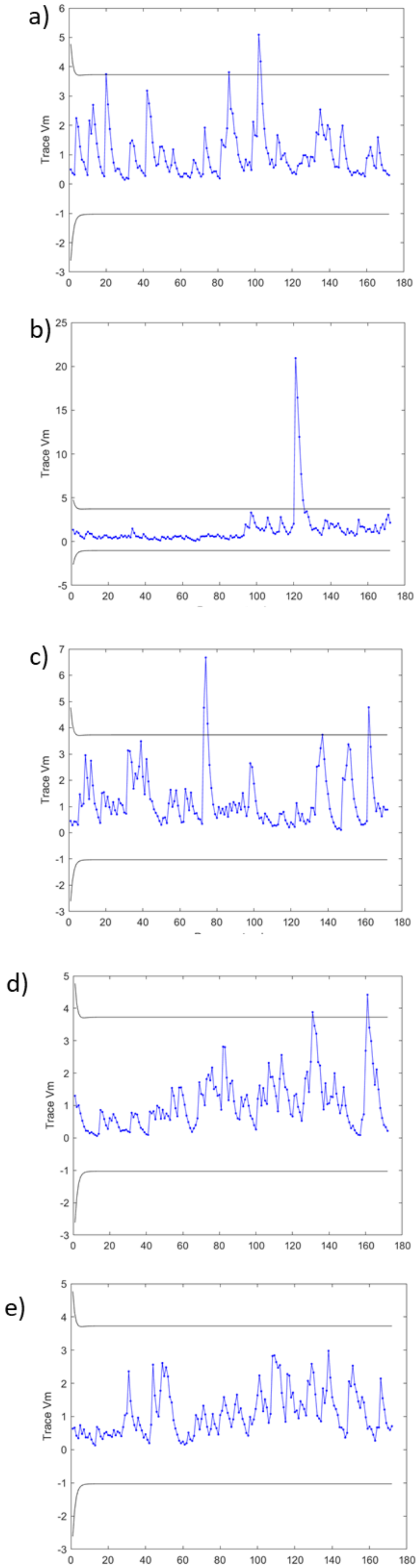

Fig. 5. MEWMV of Data Phase 2 Kontrol Control Chart
(a) WTP I, (b) WTP II, (c) WTP III, (d) WTP IV, (e) WTP V

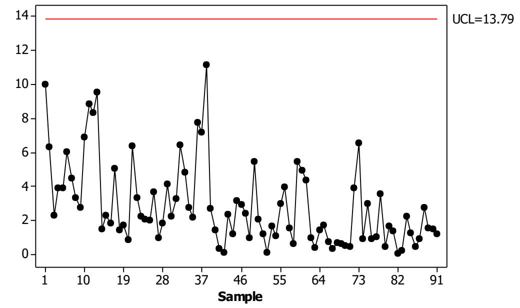

Fig. 6. MEWMA Control Chart of Phase 1 with λ =0.3

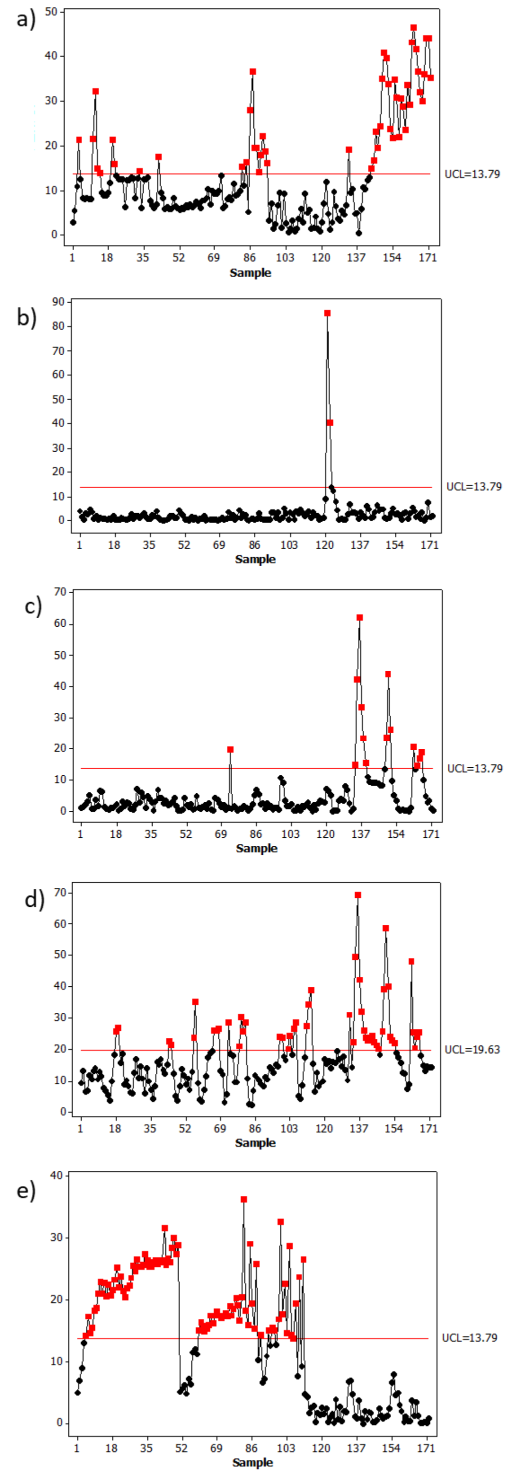

Fig. 7. MEWMA Control Chart of Data Phase 2
(a) WTP I, (b) WTP II, (c) WTP III, (d) WTP IV, (e) WTP V

The process stability of the production water quality test results is that the mean process in phase 2 data in all WTPs has not been statistically controlled with the optimum weighting used in the mean process analysis is $\lambda = 0.3$.

TABLE 5
NUMBER OF OUT-OF-CONTROL (OOC) OBSERVATIONS IN PHASE 2 OF THE MEWMA CHART

No	WTP	Number of OOC
1	Abdul Majid	50
2	Magandang	2
3	Battang	14
4	Latuppa	45
5	Batupapan	90

Table 5 shows the number of observations that went out of control in phase 2 in each WTP. Based on the table, it can be seen that WTP V Batupapan is the WTP that has the most out-of-control data compared to other WTPs.

V. CONCLUSIONS AND RECOMMENDATIONS

A. Conclusion

Based on the results of the analysis and discussion that has been carried out, the conclusions obtained are as follows.

1. The characteristics of the results of testing the quality of production water in each WTP daily in July-December 2021 based on the three quality characteristics are the quality of the production water in WTP 1 which meets the requirements for the predetermined levels, only residual chlorine, the quality of the production water in WTP 2 which meets the requirements of the predetermined levels. only turbidity is determined, for the quality of production water at WTP 3 that meets the requirements, the levels that have been set are turbidity and pH, for the quality of production water at WTP 4 that meets the requirements, the levels that have been set are residual chlorine and pH, and the quality of production water at WTP 5 based on turbidity, pH, and organic substances have met the requirements of the specified levels.
2. The process stability of the production water quality test results in each WTP is that the process variability in phase 1 data has been statistically controlled with the weights used in the process variability analysis is ω=0.4 and λ=0.4. In phase 2 data, only WTP V Batupapan was statistically controlled, with a weighting of ω=0.4 and λ=0.4. The mean process in phase 1 data has been statistically controlled, and for phase 2 data in all WTPs, it has not been statistically controlled with the optimum weighting used in the process mean analysis is λ =0.3.

B. Recommendations

Suggestions that can be given based on the conclusions above are as follows.

1. In further research, the data from the results of testing the quality of production water in each WTP are complete, because in this study there are still some missing or empty data.
2. Future research can use mixed-quality characteristics that have not been used in this study [17].
3. The implementation of the MEWMV and MEWMA control charts is expected to be considered for statistical control of the production water quality in addition to only paying attention to the characteristics of the production water quality that meet the specified specification limits.
4. In future research, the multi-response XGBoost regression approach can be used.

REFERENCES

[1] E. P. Melati, “Pengendalian Kualitas Air Produksi di PDAM Surya Sembada Kota Surabaya.” Institut Teknologi Sepuluh Nopember, 2017.

[2] D. C. Montgomery, *Introduction to statistical quality control*. John Wiley & Sons, 2020.

[3] M. Ahsan, M. Mashuri, H. Kuswanto, D. D. Prastyo, and H. Khusna, “T2 Control Chart based on Successive Difference Covariance Matrix for Intrusion Detection System,” in *Journal of Physics: Conference Series*, 2018, vol. 1028, no. 1, p. 12220.

[4] M. Ahsan, M. Mashuri, and H. Khusna, “Hybrid James-Stein and successive difference covariance matrix estimators based hotelling’s T2 chart for network anomaly detection using bootstrap,” *J. Theor. Appl. Inf. Technol.*, vol. 96, no. 20, pp. 6828–6841, 2018.

[5] M. Ahsan, M. Mashuri, H. Kuswanto, D. D. Prastyo, and H. Khusna, “Multivariate T2 Control Chart Based on James-Stein and Successive Difference Covariance Matrix Estimators for Intrusion Detection,” *MJS*, vol. 38, no. Sp2, pp. 23–35, 2019.

[6] Wibawati, M. Mashuri, Purhadi, Irhamah, and M. Ahsan, “Performance Fuzzy Multinomial Control Chart,” *J. Phys. Conf. Ser.*, vol. 1028, no. 1, p. 12120, 2018, [Online]. Available: http://stacks.iop.org/1742-6596/1028/i=1/a=012120.

[7] M. Ahsan, M. Mashuri, and H. Khusna, “Evaluation of Laney p’ Chart Performance,” *Int. J. Appl. Eng. Res.*, vol. 12, no. 24, pp. 14208–14217, 2017.

[8] D. D. Prastyo, H. Khusna, M. Mashuri, S. Suhartono, and M. Ahsan, “Multivariate CUSUM control chart based on the residuals of multioutput least squares SVR for monitoring water quality,” *MJS*, vol. 38, no. Sp2, pp. 73–83, 2019.

[9] H. Khusna, M. Mashuri, Suhartono, D. D. Prastyo, M. H. Lee, and M. Ahsan, “Residual-based maximum MCUSUM control chart for joint monitoring the mean and variability of multivariate autocorrelated processes,” *Prod. Manuf. Res.*, vol. 7, no. 1, pp. 364–394, 2019.

[10] A. D. Karaoğlan, “Regression control chart for autocorrelated data.” DEÜ Fen Bilimleri Enstitüsü, 2010.

[11] H. Khusna, M. Mashuri, S. Suhartono, D. D. Prastyo, and M. Ahsan, “Multioutput least square SVR based multivariate EWMA control chart: The performance evaluation and application,” *Cogent Eng.*, Oct. 2018, doi: 10.1080/23311916.2018.1531456.

[12] M. Ahsan and T. R. Aulia, “Comparing the Performance of Several Multivariate Control Charts Based on Residual of Multioutput Least Square SVR (MLS-SVR) Model in Monitoring Water Production Process,” in *Journal of Physics: Conference Series*, 2021, vol. 2123, no. 1, p. 12018.

[13] H. Mo, H. Sun, J. Liu, and S. Wei, “Developing window behavior models for residential buildings using XGBoost algorithm,” *Energy Build.*, vol. 205, p. 109564, 2019.

[14] T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794.

[15] S. S. Prabhu and G. C. Runger, “Designing a multivariate EWMA control chart,” *J. Qual. Technol.*, vol. 29, no. 1, p. 8, 1997.

[16] L. Huwang, A. B. Yeh, and C.-W. Wu, “Monitoring multivariate process variability for individual observations,” *J. Qual. Technol.*, vol. 39, no. 3, pp. 258–278, 2007.

[17] M. Ahsan, M. Mashuri, H. Kuswanto, D. D. Prastyo, and H. Khusna, “Outlier detection using PCA mix based T2 control chart for continuous and categorical data,” *Commun. Stat. - Simul. Comput.*, pp. 1–28, Apr. 2019, doi: 10.1080/03610918.2019.1586921.