RESEARCH ARTICLE | MARCH 17 2017

Control of wastewater using multivariate control chart FREE

Jaka Nugraha; Is Fatimah; Rino Galang Prabowo

AIP Conf. Proc. 1823, 020126 (2017)
https://doi.org/10.1063/1.4978199

Control Of Wastewater Using Multivariate Control Chart

Jaka Nugraha[1*], Is Fatimah[2], Rino Galang Prabowo[1]

[1] *Statistics Department, Universitas Islam Indonesia*
[2]*Chemistry Department, Universitas Islam Indonesia*

*email : jnugraha@uii.ac.id

Abstract. Wastewater treatment is a crucial process in industry cause untreated or improper treatment of wastewater may leads some problems affecting to the other parts of environmental aspects. For many kinds of wastewater treatments, the parameters of Biological Oxygen Demand (BOD), Chemical Oxygen Demand (COD), and the Total Suspend Solid (TSS) are usual parameters to be controlled as a standard. In this paper, the application of multivariate Hotteling T^2 Individual was reported to control wastewater treatment. By using wastewater treatment data from PT. ICBP, east Java branch, while the fulfillment of quality standards are based on East Java Governor Regulation No. 72 Year 2013 on Standards of Quality of Waste Water Industry and / or Other Business Activities. The obtained results are COD and TSS has a correlation with BOD values with the correlation coefficient higher than 50%, and it is is also found that influence of the COD and TSS to BOD values are 82% and 1.9% respectively. Based on Multivariate control chart Individual T^2 Hotteling, it is found that BOD-COD and BOD-TSS are each one subgroup that are outside the control limits. Thus, it can be said there is a process that is not multivariate controlled, but univariately the variables of BOD, COD and TSS are within specification (standard quality) that has been determined.

INTRODUCTION

Waste treatment is a crucial problem in the industrial world. Industrial waste that is not properly managed would leads to damage for human being. In order to control the effects of industrial waste, some regulations from the government are applied as well as the Government of East Java who issued the regulation number 71 of 2013 on the standards of quality of waste water from industry and / or other business activities. As many regulations related to waste water treatment, some parameters such as biological oxygen demand (BOD), chemical oxygen demand (COD) and total suspended solid (TSS) are the main indicators. Since the main content of the waste water persistence and strongly affect to the environment are organic contaminant, both COD and BOD are the representing values for the water quality. The difference between COD and BOD lays on the source of organic compound as well as the possible degradation mechanism required to reduce in that COD refers to the amount of organic contaminant destroyable by organic oxidation while BOD refers to that of destroyable by biological oxidation. In correlation for both parameters, TSS; a parameter reflecting the substance and consisting of organic and inorganic substances floating in the water is theoretically close related to the presence of the contaminants as well as organic contaminants in water. Based on that indicators refer to, COD and BOD values are usually in the same trends with TSS value. The difference between COD and BOD provides the big picture of organic materials that are difficult explained in the waters, so that there is a relationship between both parameters. Some previous invetigations provides data and information on the correlation between the TSS and BOD, alkalinity and and Oil / fats content in the liquid waste mills[1,2]. There are also some statistical and mathematical modeling provide information for predicting parameters for a certain quality of wastewater treatment.

Control of waste production is a very important stage in the processing of waste. Statistical Process Control (SPC) is a method to monitor whether a process is going well including waste production control or not. The production process is better when the running is stable and produce quality products according to the specifications expected. Multivariate quality control needs to be done in the observation that involves more than two quality

International Conference on Chemistry, Chemical Process and Engineering (IC3PE) 2017
AIP Conf. Proc. 1823, 020126-1–020126-11; doi: 10.1063/1.4978199
Published by AIP Publishing. 978-0-7354-1491-4/$30.00

characteristics. One of the tools can be adopted in the quality control is a control chart and process capability analysis. Hotelling T^2 control chart can be utilized to detect mean shift process by using the sample average vector and the covariance matrix. Process capability analysis is a statistical technique that aims to analyze the variability of the specifications which can then be used to reduce such variability[3].

Jackson (1991) stated that the analysis of multivariate process control (multivariate process control) includes four aspects: a) to answer the question "Is the process in control?" b) Is the "Procedure diagnoses an out-of-control state erroneously" must be specified; c) searching for relationships between variables / attributes. d) "If the process is out-of-control, what is the problem?". Those basic aspects of the multivariate process control may applied for sewage treatment which involves several variables to be controlled[4].

Based on the background, this paper discuss the water quality standard control using univariate and multivariate approaches. Full Map Multivariate Hotteling Individual T^2 was used to see the characteristics of the waste simultaneously.

Statistical Process Control (SPC)

SPC is a statistical tool that is able to control and monitor the production process. SPC is a collection of methods to identify special causes and bring the process into a controlled state, and reducing variation. According to the two types of process variation, i.e

a) Variations random (common cause variation)

Random variation; a variation that is inevitable and occurs due to factors that can not or difficult to control. Normal random variation occurs, then when the variation of a process included in this type, will fall within the limits of statistical control.

b) Variation assignable (assignable causes variation)

Assignable variation is the variation that can be avoided. This variation is caused by factors that can be controlled. If a process is classified into these variations, the process is said to be beyond the control of the statistics (out of statistical control).

There are three types of control lines, namely

a) The diameter or Center Line (CL), which is a line showing the average of a certain quality characteristic that is plotted on the control chart.

b) The upper limit of control or Upper Control Limit (UCL), which is a line which is the upper limit control to make the decision process.

c) The lower limit of control or Lower Control Limit (LCL), which is a line which is the lower limit control to make the decision process.

To detect the uncontrolled process is

a) The observations are outside the control limit is above the UCL or below the LCL.

b) There are nine points sequentially fall on the same side of the CL.

c) There are six points in a row up or down.

d) There are fourteen points which bergaitian heaving

Based on the number of quality characteristics to be measured, the control chart is divided into two types namely univariate control charts (Shewhart) and multivariate control chart (Hotelling T^2). Hotelling's T^2 control chart has two versions, namely for data subgroup and for individual data. Multivariate Statistics Processing Control (MSPC) is also intended to look at the stability of a production process in order to be effective, but it MSPC is a first quality control methods used in the field of modern industry [3].

In the case of univariate control chart that gives a signal out-of-control, researchers can easily deduce what the problem and provide solutions related to the observed variables. In multivariate control charts, making the conclusion that the process is out-of-control is not easy because it involves many variables that are likely to influence each other. Methods to detect for the multivariate case, Mason et al. (1995) proposed the use Hotelling T^2 decomposition. The main idea of this method is to describe the statistical T^2 into independent parts, each of which reflects the contribution of individual variables. The main problem in this method is a statistical decomposition T^2 be T^2 independent components is not unique. Furthermore, Mason et al. (1997) provide appropriate computational scheme which can reduce the computational effort. Mason et al. (1996) provide an alternative control procedures to monitor the processes that are based on double decomposition T^2 Hotelling statistics[5–7]. Mason and Young (1999) showed that by increasing the model specifications at the time of the data set is built, it is possible to increase the sensitivity of statistical T^2 for signal detection[8].

T^2 Hotteling Chart

Hotteling T^2 control chart is a control chart to control the mean vector in a multivariate process. There are two types of control chart Hotelling T^2 by the user, ie for observation of grouped data and individual data that has n = 1, m is the number of samples taken for the subgroup data. Map control with a number of subgroups (n) = 1 is used Hotteling Individual T^2 control chart. Meanwhile, if the subgroup> 1 then use the T2 control chart Hotteling in general.

In the T^2 control chart Individual Hotteling necessary assumption that the data are normal p-variat, multivariate normal kepadaatan function denoted as follows X ~ Np (μ, Σ).

Vectors average sample of i characteristics:

$$\bar{X}_j = \begin{bmatrix} \bar{X}_{1j} \\ \bar{X}_{2j} \\ \vdots \\ \bar{X}_{pj} \end{bmatrix} \quad (1)$$

with $\bar{X}_{ij}$ is the mean of characteristic -i on the sample number of -j, j =1,2,...,m. Mean of observation for –i characteristic of sample-j is

$$\bar{X}_{ij} = \frac{\sum_{k=1}^{n} x_{ijk}}{n} \quad (2)$$

dengan i = 1,2,...,p dan j =1,2,...,m serta nomor observasi ke-k untuk k=1,2,...,n. Vector average of each characteristic for m samples is

$$\bar{\bar{X}}_i = \frac{\sum_{k=1}^{m} \bar{X}_{ij}}{m} \quad (3)$$

Variance of the i^{th} sample characteristics of the sample j is

$$s_{ij}^2 = \frac{1}{(n-1)} \sum_{k=1}^{n} (X_{ijk} - \bar{X}_{ij})^2 \quad (4)$$

Covariance between the characteristics of the i-th and all of the samples h j is

$$s_{ihj} = \frac{1}{(n-1)} \sum_{k=1}^{n} (X_{ijk} - \bar{X}_{ij})(X_{hjk} - \bar{X}_{hj}) \quad (5)$$

for i≠h. Elements of the variance matrix-covariance S is

$$s_i^2 = \frac{1}{m} \sum_{j=1}^{m} s_{ij}^2 \quad (6)$$

and

$$s_{ih} = \frac{1}{m} \sum_{j=1}^{m} s_{ihj} \quad (7)$$

with i≠h. The covariance matrix is

$$S = \begin{bmatrix} s_1^2 & s_{12} & \cdots & s_{1p} \\ s_{12} & s_2^2 & & \cdots \\ \cdots & \cdots. & & \\ s_{1p} & \cdots & & s_p^2 \end{bmatrix} \quad (8)$$

Statistics Hotelling's T ^ 2 is the observation group

$$T^2 = n(\bar{x} - \bar{\bar{x}})' S^{-1} (\bar{x} - \bar{\bar{x}}) \quad (9)$$

T^2 is the statistical value of T^2 Hotelling having distributed F (n, n-p). Upper Control Limit (UCL) it is

$$UCL = \left(\frac{mnp - mp - np + p}{mn - m - p + 1}\right) F_{(\alpha, p(mn-m-p+1)} \quad (10)$$

In order to look for covariance matrix, Holmes and Mergen (1993) proposed the use of the method of sucessive difference. According to Sullivan and Woodall (1995), this method is more efficient in calculating the covariance matrix using equation (8). This method uses the difference of two vectors observations sequentially.

$$V = \begin{bmatrix} v_1' \\ v_2' \\ \vdots \\ v_{m-1}' \end{bmatrix} = \begin{bmatrix} (x_2 - x_1)' \\ (x_3 - x_2)' \\ \vdots \\ (x_m - x_{m-1})' \end{bmatrix} \quad (11)$$

Where V_1' is the value of the difference between vector and vector i i + 1, and n is the number of data samples used. Next calculate the variance covariance matrix.

$$S = \frac{1}{2}\frac{V'V}{(m-1)} \quad (12)$$

V variable is normal multivariate distributed. multivariate normal, $V \sim N_p(0,2\Sigma)$. V variation is estimated by using S . The statistics for each characteristic is

$$T_i^2 = n(x_i - \bar{x})' S^{-1}(x_i - \bar{x}) \quad (13)$$

If mean (μ) and variance (Σ) are known so the statistic of T_i^2 is Chi square (χ^2) distributed. As mean (μ) and variance (Σ) are not known,so

$$T_i^2 \frac{m}{(m-1)^2} \sim B\left(\frac{p}{2}, \frac{f-p-2}{2}\right) \quad (14)$$

with $= \frac{2(m-1)^2}{3m-4}$, f is the size of the sub-group, m is the number of sub groups and p is the number of characteristics that are controlled. Limit control is

$$UCL = \frac{(m-1)^2}{m} \beta_{\alpha, \frac{p}{2}, \frac{(f-p-2)}{2}} \qquad \text{dan } LCL = 0 \quad (15)$$

where $\beta_{\alpha, \frac{p}{2}, \frac{(f-p-2)}{2}}$ is beta distribution and the parameter are $\frac{p}{2}$ and $\frac{(p-p-2)}{2}$

Multivariate Normal Test

q-q plot can be used for the evaluation of the normal distribution with following stages :

1. Creating an average vector of each variable
2. Determining the value of variance covariance matrix
3. Sort the value of d_i^2 from small to large: $d_{(1)}^2 \le d_{(2)}^2 \le d_{(3)}^2 \le \ldots \le d_{(n)}^2$ and determine the value of $p_i = \frac{i-1/2}{n}, i = 1,\ldots,n$, which is the value of general quadratic distance *mahalanobis* or *square* of each point of observation with the average vector.
4. Determine the q_i value such that $\int_{-\infty}^{q_i} f(\chi^2) d\chi^2 = p_i$

   Which is the density function of the chi-square opportunities.
5. Create a scatter-plot of $d_{(i)}^2$ with q_i
6. If the scatter-plot is likely to form a straight line and more than 50% of the value of $d_i^2 \le \chi_{k,0.50}^2$ a multivariate normal distribution of data.

Data source

Data obtained in this research is secondary data from the analysis of waste from the Environment Agency (Balai Lingkungan Hidup/BLH) Pasuruan. This data is taken from a tub Wastewater Treatment Plant (WWTP) and is taken at the outlet. At the outlet will drain the waste out of the factory.

The study was conducted on data from January 2012 to April 2014. The variables used in this study are BOD, COD and TSS. Following the observed data are presented in Table 1.

TABLE 1. Data from the Environment Agency

No	BOD (mg/L)	COD (mg/L)	TSS (mg/L)	No	BOD (mg/L)	COD (mg/L)	TSS (mg/L)
1	8.65	17.90	10.00	15	6.73	16.30	19.60
2	18.40	47.20	7.00	16	12.40	33.60	6.40
3	34.70	81.60	18.40	17	19.80	49.30	9.60
4	15.80	38.40	20.80	18	4.33	13.12	8.67
5	19.00	43.20	7.20	19	29.73	73.28	48.67
6	10.50	25.60	13.20	20	12.38	21.44	13.83
7	10.10	25.60	8.40	21	15.12	23.83	12.22
8	14.50	42.40	8.80	22	16.13	20.07	7.67
9	12.30	30.40	11.20	23	7.17	38.97	13.00
10	11.30	27.20	12.00	24	6.32	19.96	9.30
11	4.30	9.90	6.00	25	10.23	14.40	5.40
12	28.40	68.80	16.80	26	10.11	26.88	10.40
13	8.60	20.20	16.40	27	12.12	14.56	10.00
14	8.19	20.20	7.20	28	12.24	18.56	6.83

RESULTS AND DISCUSSION

Based on the data in Table 1 can be summarized with a confidence interval 95% confidence level for each of the variables are presented in Table 2.

TABLE 2. Confidence interval

Variable	Sample size sample	Minimum	Maximum	Mean-rata	95% CI
BOD	28	4.30	34.70	13.56	(10.68; 16.43)
COD	28	9.90	81.60	31.53	(24.32; 38.74)
TSS	28	5.40	48.67	12.32	(9.11; 15.53)

Effluent standards refer to East Java Governor Regulation No. 72 Year 2013 is as follows

a. Quality Standard BOD is 0-50.

b. Quality Standard COD is 0-100

c. TSS Quality Standards is 0-50

Based on the confidence interval as Table 2 it can be concluded that the average of the three variables still meet water quality standards.
The data in Table 1is presented in Box Plot graph, as in Fig. 1.

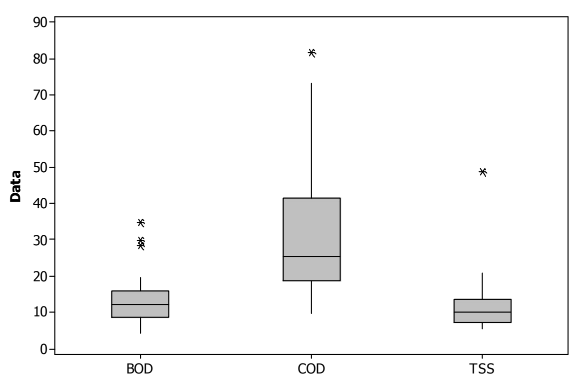

FIGURE 1. Box Plot OF BOD, COD AND TSS

Based on Fig. 1, it is known that the BOD of data there are three outliers of data that is data-3, 12, 19, whereas the COD are the outliers of data that is data to 3rd, and the TSS are the outliers of data that is data-19th.

The absence of data outlier observations made because the existence of which are beyond the control map, as can be witnessed by Fig. 2, Fig. 3 and Fig. 4.

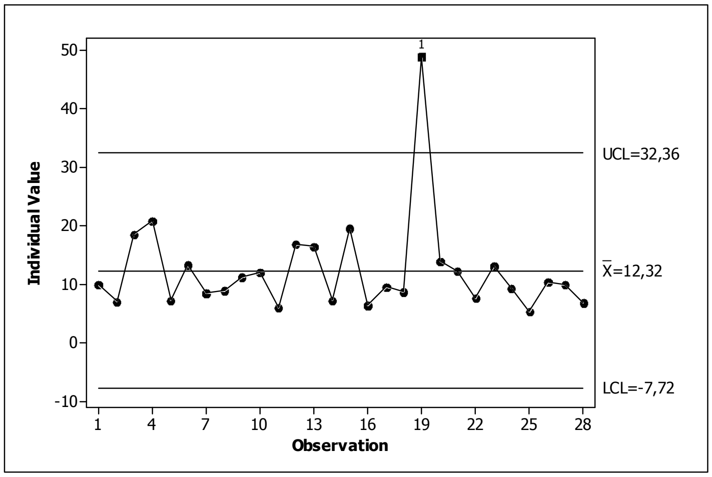

FIGURE 2. Control chart of TSS variable

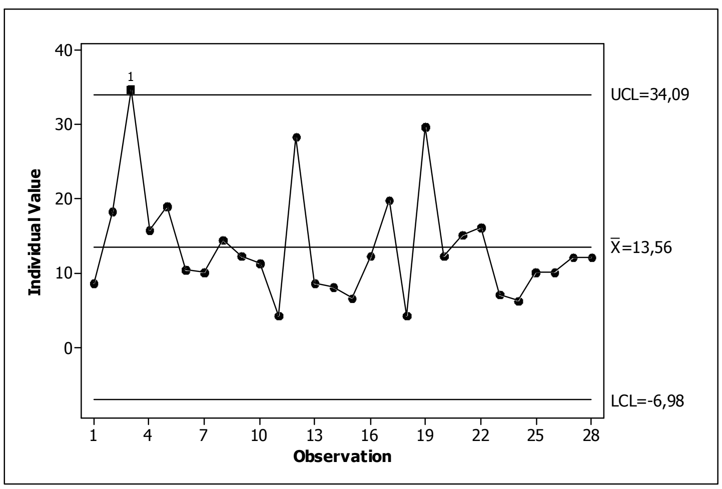

FIGURE 3. Control chart of BOD variable

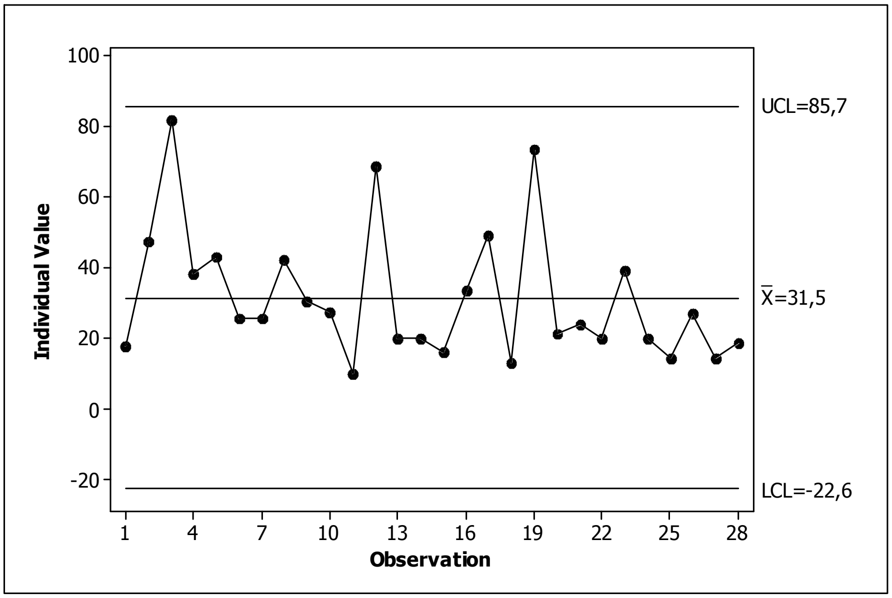

FIGURE 4. Control chart of COD variable

The relationship of all variables is represented by the plot in Fig.5.

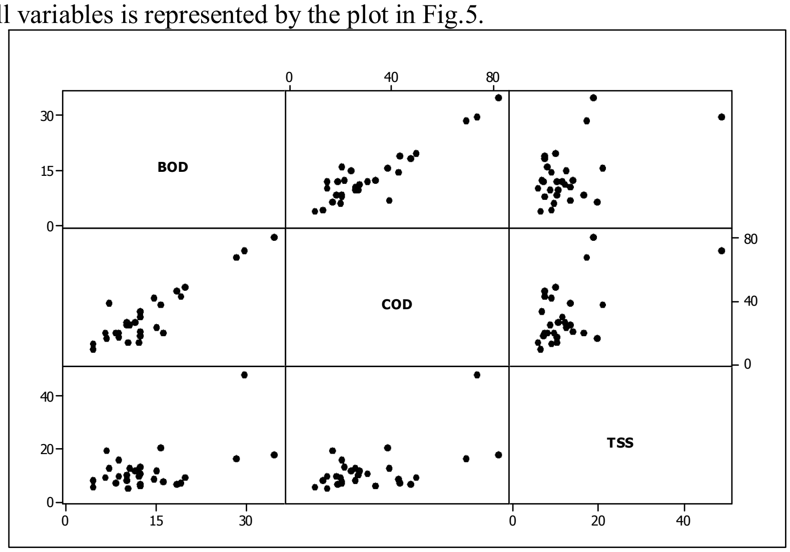

FIGURE 5. Matrix plot of BOD, COD and TSS

Based on Fig. 5, it is clear there is a linear relationship between the variables of BOD and COD. Pearson correlation three variables are as follows:

a. The correlation between COD and BOD of 0.912
b. The correlation between COD and TSS of 0.548
c. The correlation between BOD and TSS of 0.515

Normal Mutivariat Test

In order to analyze using multivariate control map, the data must have a normal distribution. Therefore, it will be tested whether the data were normally distributed. Measures as described in previous section is as follows:

a. *scatter-ploting* $d_{(i)}^2$ with q_i

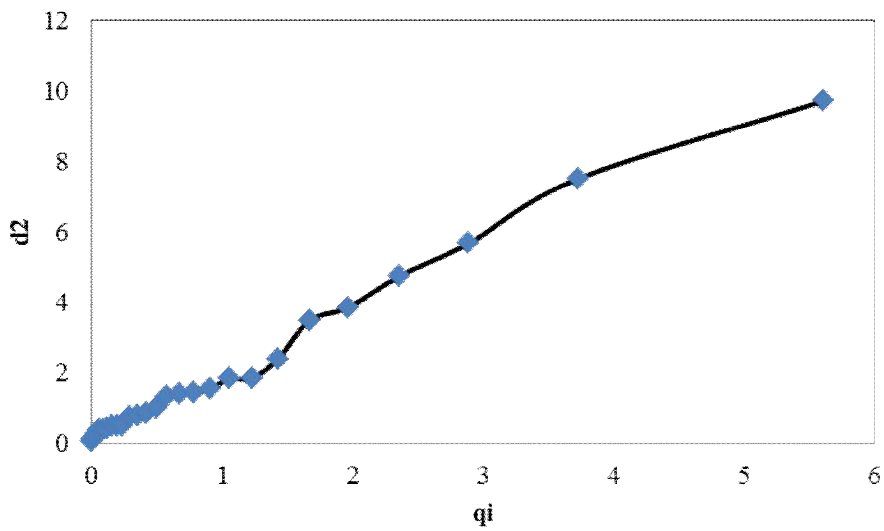

FIGURE 6. scatterplot BOD-COD

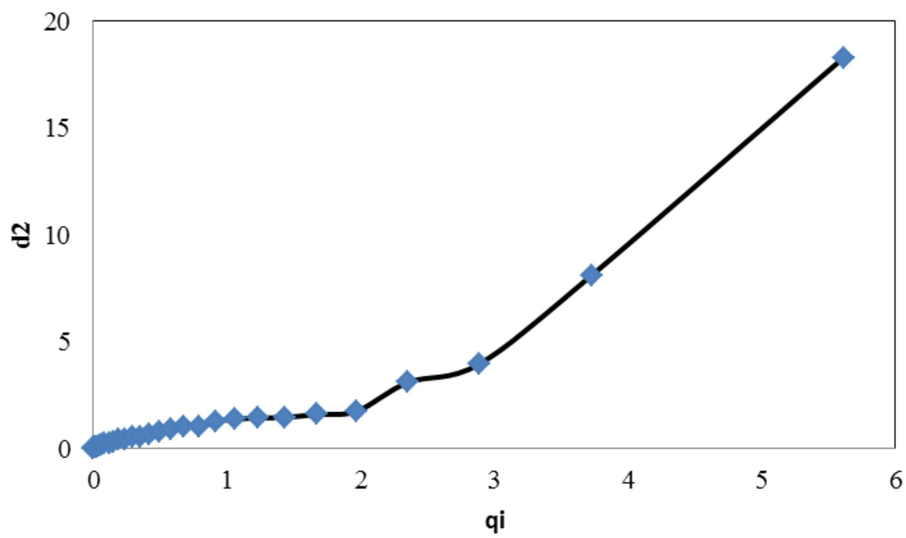

FIGURE 7. scatterplot BOD-TSS

As the scatter-plot is likely to form a straight line and more than 50% of the value of the data is multivariate normal distributed. Based on the Fig. 6 and Fig. 7, the value $d_i^2 < 1.39$ as many as 20 pieces or 80%, so it can be concluded that the data meets the multivariate normal properties.

Multivariate Control Chart

Based on the equation (13) obtained by value T^2 Hotelling and the upper control limit (UCL) is calculated using equation 15. The value of T^2 and UCL-COD BOD variables are presented in Table 3.

TABLE 3. T^2 values and UCL of BOD-COD

Subgroup	T^2	*UCL1*	*Subgrup*	T^2	*UCL1*
1	0.50	9.00	15	0.78	9.00
2	0.77	9.00	16	0.43	9.00
3	7.51	9.00	17	0.86	9.00
4	0.14	9.00	18	1.57	9.00
5	0.51	9.00	19	4.74	9.00
6	0.18	9.00	20	1.03	9.00
7	0.28	9.00	21	2.40	9.00
8	1.41	9.00	22	5.68	9.00
9	0.08	9.00	23	9.72	9.00
10	0.10	9.00	24	1.35	9.00
11	1.43	9.00	25	1.84	9.00
12	3.86	9.00	26	0.40	9.00
13	0.41	9.00	27	3.48	9.00
14	0.50	9.00	28	1.85	9.00

The values of T^2 and UCL from BOD-TSS variables are presented in Table 4.

TABLE 4. T^2 and UCL values of -TSS

Subgroup	T^2	*UCL*	*Subgrup*	T^2	*UCL*
1	0.41	9.00	15	3.11	9.00
2	1.61	9.00	16	0.54	9.00
3	8.12	9.00	17	1.38	9.00
4	1.06	9.00	18	1.47	9.00
5	1.76	9.00	19	18.30	9.00
6	0.28	9.00	20	0.11	9.00
7	0.27	9.00	21	0.06	9.00
8	0.32	9.00	22	0.81	9.00
9	0.03	9.00	23	1.03	9.00
10	0.10	9.00	24	0.90	9.00
11	1.45	9.00	25	0.66	9.00
12	3.97	9.00	26	0.20	9.00
13	1.30	9.00	27	0.08	9.00
14	0.56	9.00	28	0.45	9.00

The graph of plot of T^2 *Hotelling* values are described in Fig. 7 and Fig.8.

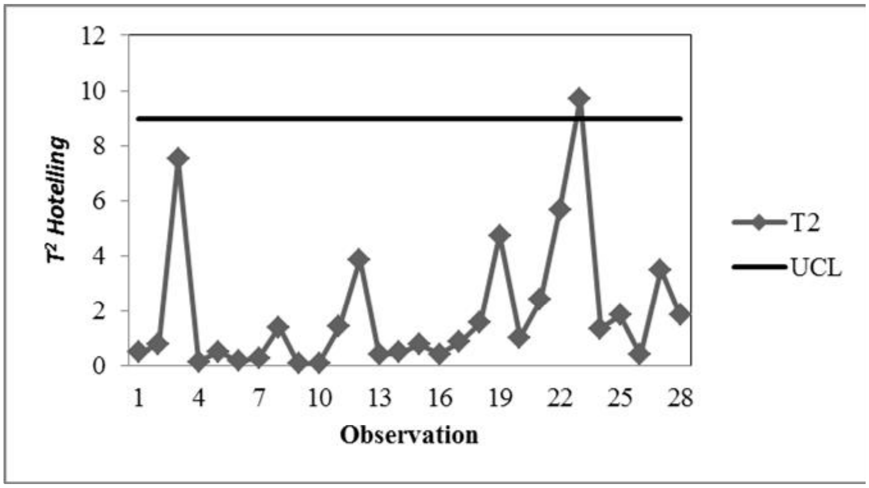

FIGURE 8. T^2 Hotelling plot for BOD-COD control chart

On Individual Hotelling T2 control chart in Fig. 8, it is known that the subgroups out of control is a subgroup of 23, it is caused by the content in wastewater affect the size of BOD and COD. As it is known that BOD and COD are lineary-correlated so that BOD will be increases as COD value increases. COD has an influence on the BOD values by 91.162%. But it is also possible due to other factors such as the weather at the time and conditions of the Wastewater Treatment Plant (WWTP).

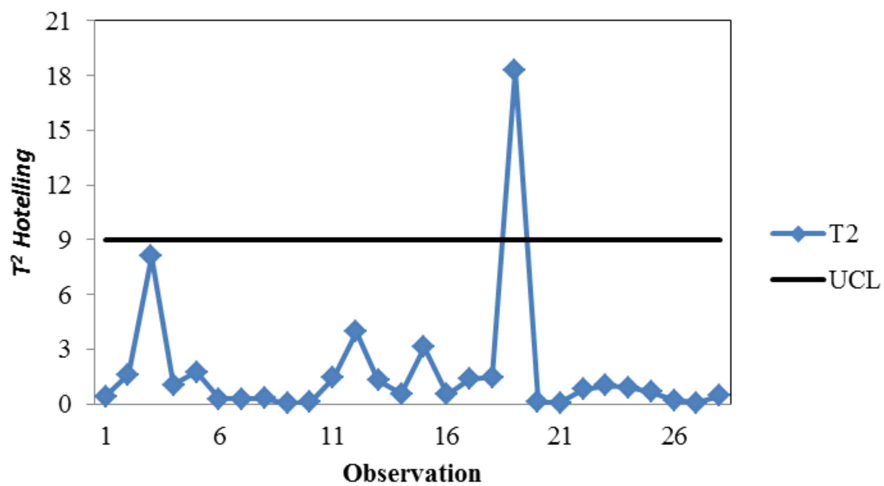

FIGURE 9. T^2 Hotelling value of BOD-TSS control chart

On Individual Hotelling T2 control chart (Fig.9), it is known that subgroups out of control is a subgroup of the 19th representing that it is caused by the content in the wastewater that affect the size of BOD and TSS. As well as the interaction of BOD-COD, BOD and TSS are also lineary-associated.

CONCLUSION

Based on the results, there are some concluding remarks:

1. Variables of BOD, COD and TSS are in compliance with the Quality Standard that has been set by the Government through the East Java gubernatorial. No. 72 Year 2013.

2. Based on Multivariate control chart Individual T^2 Hotteling, it is found that there are observations that are outside the control limits are observations Observation of the 23rd and 19th. Therefore it can be concluded that the process of uncontrolled waste management.

REFERENCES

1. K. Zaher and G. Hammam, International Journal of Sciences: Basic and Applied Research **13,** 42–48 (2014).
2. A. Vijayan and G. S. Mohan, International Journal of Science and Research **5,** 2013–2016 (2016).
3. D. Montgomery, *Statistical Processing Control* (John Wiley & Sons, New Jersey, 2009).
4. J. E. Jackson, A user's guide to principal components (John Wiley & Sons, New York, 1991).
5. M. R. L., T. N.D., Y. J.C., Journal of Quality Technology, 39–50 (1996).
6. R. L. Mason, N. D. Tracy and J. C. Young, Journal of Quality Technology **27**, 99–108 (1995).
7. R. L. Mason and J. C. Young, Journal of Quality Technology **31**, 155–165 (1999).
8. R. L. Mason, N. D. Tracy and J. C. Young, Journal of Quality Technology **29**, 396–406 (1997).