An Application of Univariate and Multivariate Control Charts in Monitoring Water Quality

Siti Rahayu Mohd Hashim[1], Azwaan Andrew[1] and Wilter Azwal Malandi[2]

[1]*Complex of Science and Technology, Faculty of Science and Natural Resources, University Malaysia Sabah,*
[2]*Complex of Forestry, Faculty of Science and Natural Resources, University Malaysia Sabah*

Control chart is a tool for detecting an out-of-control signal in statistical process control (SPC). It is widely used in process monitoring in order to detect changes in process mean or process dispersion. This study aims to illustrate the application of multivariate control charts in monitoring water quality at one of the water treatments plants in Kota Kinabalu, Sabah. The tested water quality variables in this study are turbidity, pH value, dissolved oxygen (DO) and concentration of ferum. Two multivariate control charts, Hotelling'sT^2 and MCUSUM control charts are constructed under the violation of the multivariate normality assumption. The purpose is to study the effect of non-normal data upon the monitoring process using the selected multivariate control charts. By comparing the monitoring process between the two types of control charts, the consistency of the results is studied. All the univariate and multivariate control charts produced out-of-control signals from different points, hence inconclusive results obtained.

Keywords: Water quality; multivariate control chart; univariate control chart; Hotelling's T2; MCUSUM

I. INTRODUCTION

In many monitoring processes or activities, it is frequently required to monitor several variables and in a number of occasions, the process variables are interrelated. The Hotelling'sT^2 multivariate control chart has always been the most popular one among the researchers and practitioners and it is a widely used statistical tool for monitoring a nominal mean vector and usually by assuming that the process dispersion remains constant. Multivariate HotellingT^2 was first introduced by Hotelling (1947) and has been discussed in detail and used by many researchers since, such as Alt (1977, 1985), Alt and Smith (1988), Ryan (1989, 2000) and Jackson (1991). It is considered as the "natural multivariate extension to the univariate Shewhart chart" (Lowry and Montgomery, 1995). Multivariate HotellingT^2 control chart is said to be insensitive to small and moderate shifts in the mean vector since the procedure is based on only the most recent observation (Lowry & Montgomery, 1995).

There are a few MCUSUM charts proposed by several authors but in this study, the MCUSUM scheme introduced by Crosier (1988) is applied. Lowry and Montgomery (1995) stated that MCUSUM control chart is able to detect small shifts in mean vectors faster than the multivariate HotellingT^2 control chart. Qiu and Hawkins (2001) proposed MCUSUM based on the ranks or the orders of observation in handling normality assumption.

Alt (1985) has defined two phases in constructing a multivariate control chart. Phase I is called retrospective analysis whereas Phase II is known as prospective analysis (Montgomery, 2001; Woodall, 2000). In Phase I, a set of historical data is analysed and plotted to determine whether the process is statistically in control. Any causes of instability in the process will be taken out. The in-control parameters of the process are estimated, and the initial control limits are established in this phase. The control limits are used in Phase II where the data are analysed on-line for detecting any departure from the parameter estimates. It is crucial for the

---

*Corresponding author's e-mail: rahayu@ums.edu.my

researchers as well as the practitioners to be able to distinguish the methods and the applications between these two phases (Woodall, 2004).

The objectives of this study are to illustrate the application of univariate and multivariate control charts and to identify the effect of non-normality in the performance of both types of control chart. Based on these two objectives, this study would like to identify the obstacles and the practicality of applying statistical process control's tool in monitoring water quality such as the effect of imbalance sample size of the variables due to unequal frequency in data collection and time points. Previous studies mainly focus on a dataset with equal number of observations for each quality variables. In this study, the original number of observations is not equal due to different data collection frequency in a week or a month. In this study, the value of the water quality variable is assumed constant and similar to the value from the prior data collection time point whenever observation was not made for any particular time point.

A. Water Quality Variables

The tested water quality variables are turbidity, pH value, DO and the concentration of ferum. All of the variables are tested for multivariate normality assumption. The Mardia multivariate normality test, Henze-Zirkler test and Royston test were performed on the data. All the p-values of the tests showed a significant shift from multivariate normality.

Table 1. Multivariate normality tests

Normality test		Statistic	p-value
Mardia	Skewness	71.04	< 0.001
	Kurtosis	5.58	< 0.001
Henze-Zirkler		1.40	< 0.001
Royston		83.14	< 0.001

The relationship between the variables is also been investigate and the Spearman correlation test was performed for this purpose. The result of the correlation test is given in Table 2.

The results of the correlation test showed that variable pH and the concentration of iron are significantly correlated in opposite directions.

Table 2. Spearman correlation test

	Turbidity	pH	DO	Ferum
Turbidity	1	-0.191	-0.23	0.053
pH		1	-0.087	-0.317[*]
DO			1	0.076
Iron				1

B. Constructing Hotelling T²Control Charts

Letm is the number of samples with size $n=1$ and p is the number of quality variables. HotellingT² statistic is defined as below.

$$T^2 = (\mathrm{x} - \bar{\mathrm{x}})'\mathbf{S}^{-1}(\mathrm{x} - \bar{\mathrm{x}}) \qquad [1]$$

With $\bar{X}$ is the mean vector and variance-covariance matrix is given as,

$$\boldsymbol{S}_1 = \frac{1}{m-1}\sum_{i=1}^{m}(\mathrm{x_i} - \bar{\mathrm{x}})(\mathrm{x_i} - \bar{\mathrm{x}})' \qquad [2]$$

Where m is the sample size.

1. Phase I

The control limits of multivariate control charts for individual observations is given as

$$UCL = \frac{(m-1)^2}{m}\beta_{\alpha,\frac{p}{2},\frac{m-p-1}{2}}; \quad LCL = 0 \qquad [3]$$

where, $\beta_{\alpha,\frac{p}{2},\frac{m-p-1}{2}}$ is the upper percentage of α for beta distribution with $\frac{p}{2}$ and $\frac{m-p-1}{2}$ as the distribution parameters.

2. Phase II

In Phase II, the control limits are based on the F probability distribution as defined below.

$$UCL = \frac{p(m+1)(m-1)}{m^2-mp}F_{\alpha,p,m-p} \qquad [4]$$

Estimated control limit is suggested for $m>100$ and it is given as,

$$UCL = \frac{p(m-1)}{m-p}F_{\alpha,p,m-p} \qquad [5]$$

Another control limit, $UCL = \chi^2_{\alpha,p}$is proposed if and only if the variance covariance matrix is known (Montgomery, 2009).

C. Constructing MCUSUM Control Chart

Crosier (1988) has proposed two multivariate procedures for control chart. Crosier (1988) scheme is good in identifying out-of-control signal and the statistics is defined as below,

$$T_m^2 = \left[S_i' \left(\frac{\Sigma}{n} \right)^{-1} S_i \right]^{1/2} > h \qquad [6]$$

$$S_i = \begin{cases} 0 & if\ C_i \le k \\ (S_{i-1} + \bar{X}_i - \mu_0)\left(1 - \dfrac{k}{C_i}\right) & if\ C_i > k \end{cases}$$

$S_0 = 0, k > 0$, and

$$C_i = \left[(S_{i-1} + \bar{X}_i - \mu_0)' \left(\frac{\Sigma}{n}\right)^{-1} (S_{i-1} + \bar{X}_i - \mu_0) \right]^{1/2}$$

The upper control limit for Crosier (1988) scheme is, UCL=h. Control limit h can also be expressed as a function of reference value, $h_c(k)$ (Healy, 1987).

II. RESULTS AND DISCUSSIONS

Two type of control charts, univariate and multivariate have been constructed to illustrate the application of the methods and to compare the outcome of the monitoring process between the two types of control chart. Figures 1 to 4 showed the stability of the process for each water quality variable separately. The number of observations used in constructing univariate control are varies between water quality variables depending on the availability of the data.

Whereas Figures 5 to 8 showed the multivariate HotellingT^2 and MCUSUM control charts in both phases. Twenty-five observations were used in Phase I in order to establish the control limits for Phase II. Once the control limits have been established, it will be used in Phase II for monitoring purposes. A new set of data consists of 43 observations is monitored in Phase II.

A. Univariate Control Charts

Figure 1 shows the average values of pH throughout the study are statistically in control since all the points are within the control limits.
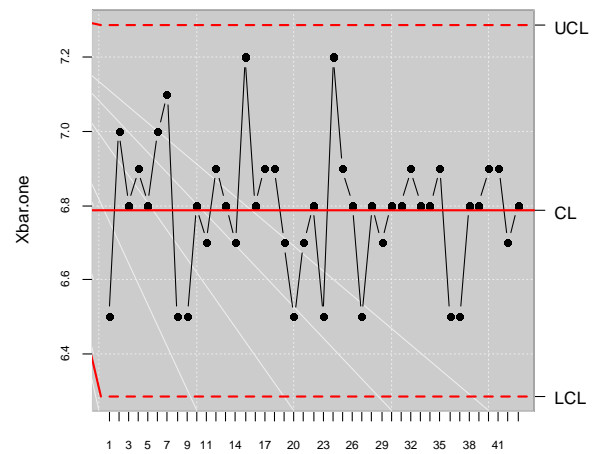

Figure 1. $\bar{X}$ control chart for pH

The same thing is observed in Figure 2. The average values of DO are found to be statistically in-control and none of the points located outside the in-control region.
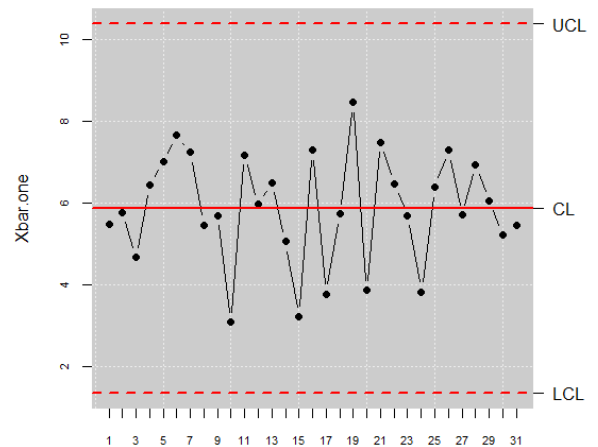

Figure 2. $\bar{X}$ control chart for dissolved oxygen (DO)

On the other hand, Figure 3 shows one observation above the UCL. This indicates that the process is statistically instable. The average value of ferum at that particular time is found to be higher than the maximum level allowed.
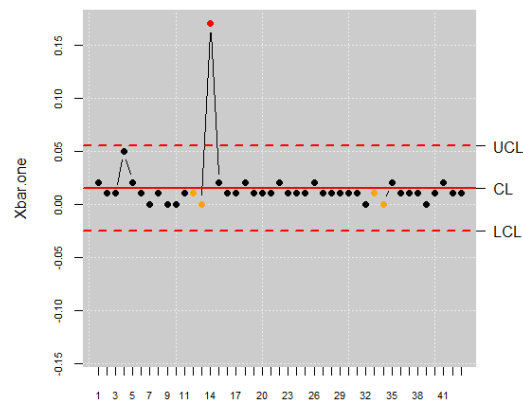

Figure 3: $\bar{X}$ control chart for concentration of ferum

Figure 4 shows two observations above the UCL which indicates an instability in the process.
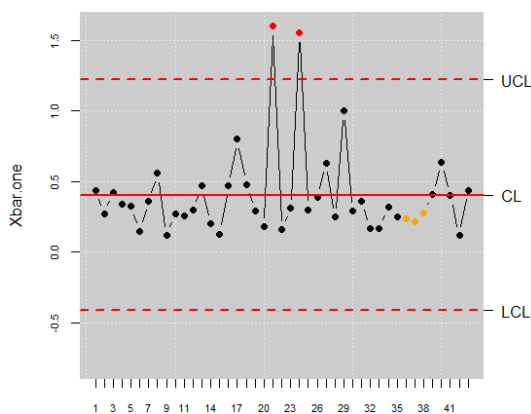

Figure 4: $\bar{X}$ control chart for turbidity

The direction of the shifts in the mean is also depicted by Figures 3 and 4 where the two variables, turbidity as well as the concentration of ferum have increased and surpassed the maximum allowable level.

B. Multivariate Control Charts

In this study, two multivariate control charts have been constructed in order to monitor the water quality variables by computing two types of statistics as defined in equations [1] and [5].

The mean vector and the variance covariance matrix obtained from the data set are shown below. It will be used in [1] and [5] for the computation of the statistics value of HotellingT^2 and MCUSUM.

$$\bar{X} = \begin{bmatrix} 7.31 \\ 6.87 \\ 0.02 \\ 0.02 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 1.1 \times 10^{-2} & 0.00620 & -8.5 \times 10^{-4} & 3.2 \times 10^{-5} \\ 6.2 \times 10^{-3} & 0.27000 & 2.8 \times 10^{-3} & -7.2 \times 10^{-4} \\ -8.5 \times 10^{-4} & 0.00280 & 5.3 \times 10^{-4} & 4.4 \times 10^{-5} \\ 3.2 \times 10^{-5} & -0.00072 & 4.4 \times 10^{-5} & 2.0 \times 10^{-4} \end{bmatrix}$$

i. Phase I

The upper control limit for the multivariate HotellingT^2 control chart is computed based on equation [3] and the value obtained is 10.61. The Hotelling's T^2 statistics in Table 3 are computed using equation [1].

Table 3. HotellingT^2 statistics in Phase I

k	$T2$	k	$T2$
1	0.55	14	1.56
2	0.82	15	5.88
3	7.76	16	5.32
4	0.67	17	1.72
5	13.45	18	0.48
6	0.63	19	8.22
7	0.71	20	0.85
8	9.89	21	0.82
9	2.19	22	1.57
10	2.67	23	1.28
11	0.91	24	6.03
12	2.26	25	12.87
13	6.88		

The values are then plotted as shown in Figure 5. Two out-of-control signals are triggered by the chart indicating that the multivariate process is not statistically stable. The two points need to be taken out and the parameters, mean vector as well variance covariance matrix need to be re-evaluated.
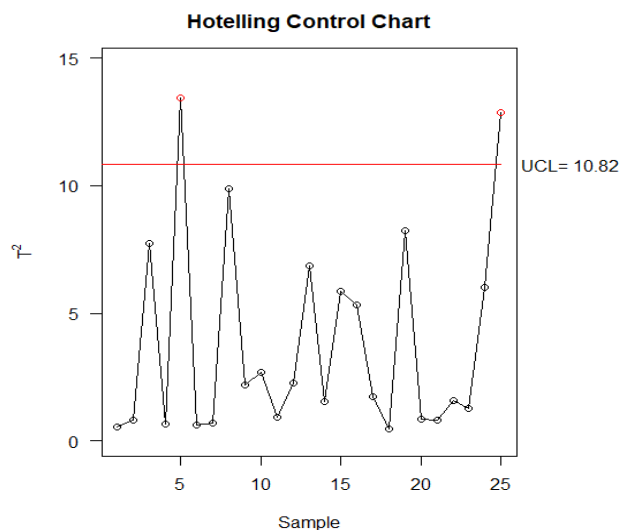

Figure 5. Multivariate HotellingT^2 control chart in Phase I

The new mean vector and variance covariance matrix are given below. These parameters will be used in computing the Hotelling T^2 statistics in for Phase II.

$$\bar{X} = \begin{bmatrix} 7.31 \\ 6.95 \\ 0.02 \\ 0.02 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 1.2 \times 10^{-2} & 0.0058 & -9.4 \times 10^{-4} & 6.3 \times 10^{-5} \\ 5.8 \times 10^{-3} & 0.1300 & 1.2 \times 10^{-3} & -1.0 \times 10^{-3} \\ -9.4 \times 10^{-4} & 0.0012 & 5.6 \times 10^{-4} & 5.4 \times 10^{-5} \\ 6.3 \times 10^{-5} & -0.0010 & 5.4 \times 10^{-5} & 1.0 \times 10^{-4} \end{bmatrix}$$

ASM Science Journal, Volume 13, 2020

MCUSUM statistics in Table 4 are computed using equation [5]. Unlike multivariate HotellingT^2 control chart, there is no out-of-control signal has been triggered by MCUSUM control chart. All of the points located below the upper control limit, UCL = 5.5.

Table 4. MCUSUM statistics in Phase I

k	MCUSUM	k	MCUSUM
1	0.24	14	2.87
2	0.30	15	4.06
3	2.29	16	3.29
4	1.66	17	3.20
5	3.17	18	3.13
6	3.33	19	2.89
7	2.67	20	2.54
8	4.24	21	2.69
9	2.75	22	1.47
10	1.75	23	1.08
11	1.08	24	1.81
12	1.39	25	4.32
13	2.49		

Figure 6. MCUSUM control chart in Phase I

ii. Phase II

In Phase II, the new mean vector and variance covariance matrix are used in computing the statistics values.

Table 5. Hotelling T^2 statistics in Phase II

k	T2	k	T2
1	1.19	22	2.07
2	1.70	23	6.92
3	7.44	24	1.00
4	1.72	25	7.19
5	1.41	26	2.87
6	0.86	27	24.39
7	10.64	28	3.37
8	4.76	29	0.86
9	2.30	30	4.43
10	1.07	31	2.33
11	3.31	32	3.38
12	9.88	33	0.25
13	1.54	34	9.91
14	9.35	35	11.78
15	5.44	36	0.43
16	3.59	37	4.98
17	0.55	38	4.83
18	9.38	39	6.84
19	1.16	40	0.23
20	0.70	41	2.60
21	3.67		

Table 5 and Table 6 show the new statistics values using the new observations for the process monitoring.
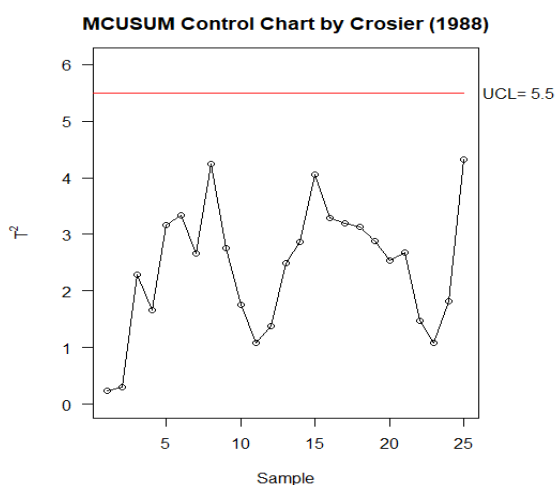

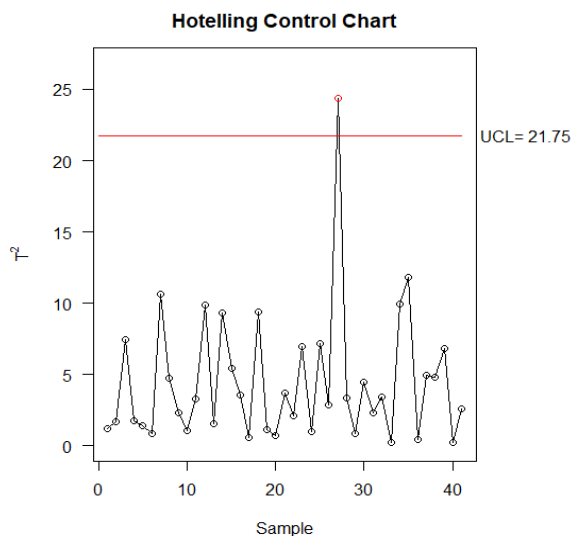Figure 7. Hotelling T^2 control chart in Phase II

Figure 7 shows that the multivariate Hotelling T2 triggered an out-of-control signal with the point located above the UCL = 21.75. Whereas Figure 8 shows a subsequent out-of-control signals starting from point k=32.

5

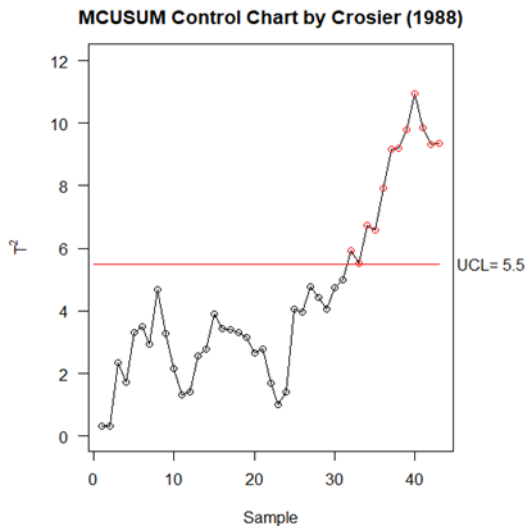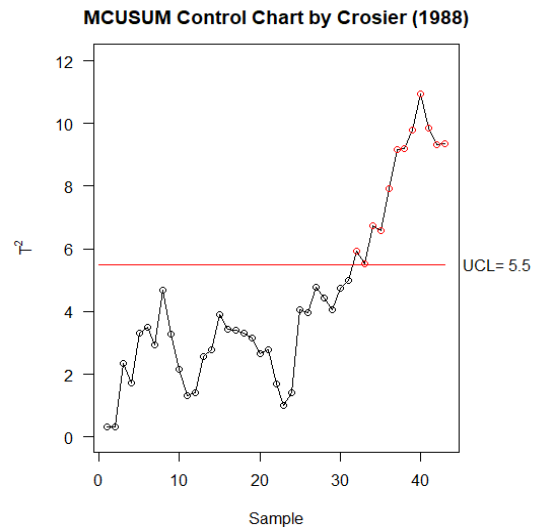Figure 8. MCUSUM control chart in Phase II

Table 6. MCUSUM statistics in Phase II

k	MCUSUM	k	MCUSUM
1	0.31	22	1.69
2	0.31	23	1.00
3	2.35	24	1.42
4	1.72	25	4.06
5	3.30	26	3.97
6	3.49	27	4.78
7	2.95	28	4.44
8	4.67	29	4.07
9	3.28	30	4.75
10	2.17	31	4.98
11	1.31	32	5.93
12	1.40	33	5.51
13	2.55	34	6.73
14	2.77	35	6.59
15	3.90	36	7.92
16	3.43	37	9.16
17	3.40	38	9.21
18	3.31	39	9.78
19	3.15	40	10.95
20	2.66	41	9.85
21	2.79		

III. SUMMARY

The out-of-control signals produced by both univariate and multivariate control charts in the previous section did not share similar points. The points which fell beyond the UCL for both univariate control charts are within the control limits in both multivariate control charts. Whereas the out-of-control signals triggered by the multivariate control charts produced by the points which are located between UCL and LCL in univariate control charts. The MCUSUM control chart for some reasons continue to produce out-of-control signal starting from the point $k=32$. However, it did not happen to multivariate Hotelling T^2 control chart. This shows how sensitive MCUSUM control chart when the MVN is violated and the MCUSUM's lack of robustness was discussed by Nidsunkid *et al.* (2018). Both types of control chart produced out-of-control signals from different points, hence inconclusive results obtained. This study proposes the application of non-parametric multivariate T^2 control chart as an alternative tool for process monitoring whenever the multivariate normality assumption is violated (Mostajeran & Iranpanah, 2018; Boone & Chakraborti, 2012).

IV. ACKNOWLEDGEMENTS

V.   REFERENCES

Alt, F. B., 1977, Economic Design of Control Charts for Correlated, Multivariate Observations. PhD Dissertation, Georgia Institute of Technology, Atlanta, GA.

Alt, F.B., 1985, Multivariate Quality Control. Encyclopedia of Statistical Science. Kotz and Johnson Eds, Vol. 6, John Wiley and Sons, New York, pp.110 -122.

Alt, F. B. and Smith, N.D., 1988, Multivariate Process Control. Handbook of Statistics, Vol. 7, Krishnaiah PR, Rao CR Eds. Elsevier: Amsterdam, pp. 333-351.

Boone, J., & Chakraborti, S. 2012, Two simple Shewhart-type multivariate nonparametric control charts. Applied Stochastic Models in Business and Industry, 28(2), 130-140.

Crosier, R. B. 1988, Multivariate Generalizations of Cumulative Sum Quality Control Schemes. Technnometrics, 30: 219-303.

Hotelling, H., 1947, Multivariate Quality Control. Techniques of Statistical Analysis. In Eisenhart, C., Hastay M.W. and Wallis W.A. McGraw Hill, New York, pp. 111-184.

Jackson, J. E, 1991, Series in Probability and Mathematical Statistics. John Wiley and Sons.

Lowry, C.A. and Montgomery, D.C. 1995, A Review of Multivariate Control Charts. IIE Transactions, 27, pp. 800-810.

Nidsunkid et al. 2018, The Performance of MCUSUM Control Charts when the Multivariate Normality Assumption Is Violated. Thailand Statistician. 16(2): 140-155.

Montgomery, D. C 2001, Introduction to Statistical Quality Control. 4th Edition, New York: John Wiley & Sons, Inc.

Montgomery, D. C 2009, Introduction to Statistical Quality Control. Ed. ke-6. John Wiley & Sons, Inc.

Mostajeran, A., Iranpanah, N., & Noorossana, R. 2018, An explanatory study on the non-parametric multivariate T2 control chart. Journal of Modern Applied Statistical Methods, 17(1), eP2559.

Nidsunkid et al. 2018, The Performance of MCUSUM Control Charts when the Multivariate Normality Assumption Is Violated. Thailand Statistician, 16(2): 140-155.

Ryan, T.P 1989, Statistical Methods for Quality Improvement, John Wiley and Sons, New York.

Ryan, T.P 1989, Statistical Methods for Quality Improvement, John Wiley and Sons, New York.

Tianhua Wang & Shuguang Huang 2016. An Adaptive Multivariate CUSUM Control Chart for Signaling a Range of Location Shifts. Communications in Statistics - Theory and Methods, 45:16, 4673-4691.

Qiu, P.H. and Hawkins D 2001, A Rank Based Multivariate CUSUM Procedure. Technometrics, 43, pp.120-132.

Woodall, W. H. &Ncube, M. N. 1985, Multivariate CUSUM Quality Control Process. Technometrics, **27**(3): 285-292.

Woodhall, W. H 2000, Controversies and Contradiction in Statistical Process Control. Journal of Quality Technology, 32, pp. 341-378.