Available online www.ejaet.com

European Journal of Advances in Engineering and Technology, 2022, 9(3):42-49

Research Article ISSN: 2394 - 658X

An Application of Hotelling's T^2 Decomposition in a Multivariate Statistical Process Control

Gulumbe, S. U., Zoramawa, A. B., Umar Usman and Kilishi, M. U.*

Department of Statistics, Usmanu Danfodiyo University, Sokoto, Nigeria
*Corresponding Author's Email: omuhammad7@gmail.com

ABSTRACT
Multivariate Statistical Process Control (MSPC) is the most acceptable monitoring tool for several variables, it is advantageous when compared to the simultaneous use of univariate scheme. However, one of the challenges in this scheme is the identification of influential variable(s). Any signal by this scheme implies one or more variables in the process are out-of-control. Therefore, identification of such variable(s) correctly is very important. The aim of this research is to determine the variable(s) that cause out-of-control signal in cement production process of Cement Company of Northern Nigeria (CCNN), Sokoto. The data used for this research were percentages of clinker minerals (Alite, Bilite, Aluminate and Ferrite), obtained from the Cement Company of Northern Nigeria (CCNN), Sokoto. We achieved our aim through the construction of Hotelling's T^2 phase I and II control charts, presentation of MYT decomposition models and computation of the T^2 decomposition terms. Observations composing of 100 samples regarding cement clinker variables in Phase I were acquired in order to form the reference data set in the application. It was found that 82 of these samples formed the reference data set as a result of 8 steps analyses. In Phase II, T^2 chart was constructed for 30 new observations and it was detected that 7 samples were out-of-control. The MYT decomposition revealed the T^2 values and their corresponding critical values, the T^2 values that were higher than their corresponding critical values were responsible for the out of control signal. We conclude that variable 4 (Ferrite) is the influential variable.

Key words: Hotelling's T^2 control chart, MSQC package, Multivariate Statistical Process Control, MYT decomposition

1. INTRODUCTION

In any production process regardless of how well designed or carefully maintained, a certain amount of inherent or natural variability will always exist. Occasionally, however, assignable causes will occur, seemingly at random, resulting in a "shift" to an out-of-control state where a large proportion of the process output does not conform to requirements. A major objective of statistical process control is to quickly detect the occurrence of assignable causes or process shifts so that investigation of the process and corrective action may be undertaken before many nonconforming units are manufactured. Although it may not be possible to eliminate variability completely, the control chart helps reduce it as much as possible. A typical control chart is a graphical display of a quality characteristic that has been measured or computed from a sample versus the sample number or time. A particular nonrandom pattern usually appears on a control chart for a reason, and if that reason can be found and eliminated, process performance can be improved [1].

Process monitoring in which several variables are of interest is called Multivariate Statistical Process Control (MSPC). Multivariate control charts are widely used in practice to monitor the simultaneous performance of several related quality characteristics [2-3].

The need to execute Multivariate Statistical Process Control (MSPC) in production process for quality improvements increases daily. Statistical methods play a very important role in quality improvement in manufacturing industries [4].

The use of T^2 decomposition proposed by Mason *et al.*, [5] is considered as the most valuable technique. Mason *et al.*, [6] presented a practical application of the T^2 decomposition on a bivariate data (two process variables). They concluded that there is a need for a faster sequential computation scheme for the decomposition.

Ulen and Demir [7] applied the T^2 decomposition using three variables obtained from a pharmaceutical industry to determine variables that contributed to out-of-control signal. They suggested that the calculation of multivariate control methods is not easy when compared to Shewhart charts. Statistical packet programs are needed for these calculations. Graphic and calculation can be performed with the help of software regarding this topic.

Agog *et al*., [5] decomposed the hotelling's T^2 statistic using four variables. They concluded that they have presented a simplified way of decomposing the T^2 statistic by using four process variables where 24 decompositions were obtained. Each of these decompositions can be used to identify variable(s) that significantly contributes to an out-of-control signal.

Agog *et al*., [6] determined out of control variables in a multivariate quality control chart using four variables. They concluded that by removing the effect of the machines that caused the out-of-control signal brought the process to a state of control.

Akeem *et al*., [8] decomposed the hotelling's T^2 statistic using five variables. They concluded that they have simplified T^2 decomposition model for five variables from which the contribution of each and every variable can be seen obviously when the process goes out of control.

Akeem and Olatunji [9] revealed the invariant attribute of the MYT decomposition using five variables. They concluded that they were able to reveal the invariant attribute of the MYT decomposition using five variables for illustration and this usually aid correctness and proper identification of the process variable contributing to signal of any process.

Multivariate statistical process control (MSPC) is the most acceptable monitoring tool for several variables, and it is advantageous when compared to the simultaneous use of univariate scheme. However, one of the many challenges in this scheme is the identification of influential variable(s). Any signal by this scheme implies one or more variables in the process has gone out-of-control. Therefore, identification of such variable(s) correctly is very important. The Mason, Young and Tracy (MYT) decomposition diagnosis is one of the approaches commonly used to identify the influential variables. This approach aids the breaking down, the overall T square value and shows the individual variable contribution, while their joint contribution is also revealed [8].

The aim of this research is to determine the variable(s) that cause out-of-control signal in the Cement Production process of Cement Company of Northern Nigeria (CCNN), Sokoto. The objectives are; construction of 1st Phase T^2 Chart, construction of 2nd Phase T^2 Chart and application of MYT Decomposition.

2. METHODLOGY

The MYT Decomposition technique of the Hotelling's T^2 statistic into orthogonal components is used for identifying variables that significantly vary. The Hotelling's T^2 statistic is a common tool used in multivariate process control chart. This method has two distinct phases of control charting, Phase I and Phase II. In phase I, charts are used for retrospectively testing whether the process was in control when the first samples were being drawn. Phase II is used for monitoring future production. This phase is aimed at detecting whether subsequent production is capable of causing any of the observation vectors from the historical dataset to be out-of-control. The two phases of the charts were employed for out-of-control condition.

Data Description

The data (Cement Clinker) used for this work were obtained from the Quality Control Division of the Cement Company of Northern Nigeria (CCNN), Sokoto. Where 100 samples were used in Phase I and 30 samples were used in Phase II. Cement is a fine powder produced by grinding cement clinker. Cement Clinker is nodules (diameters, 5-25mm) of sintered material produced by heating a homogeneous mixture of raw materials in a kiln to a sintering temperature of approximately 1450°C for modern cements. The resulting clinker consists of four principal chemical compounds (Alite, Belite, Aluminate and Ferrite), which are normally referred to as the clinker minerals.

Individual Observations

Most chemical and process industries have a subgroup size of n=1. This is because there are many quality characteristics that must be monitored. Suppose that *m* samples each of size n=1 are to be monitored, and that *p* is the number of quality characteristics observed in each sample. Let $\overline{X}$ be the sample mean vector and ***S*** be the covariance matrix for the individual observations. The T^2 statistic for a *p* dimensional observation vector $X' = (x_1, x_2, ..., x_p)$ can be represented as $T^2 = (X - \overline{X})' S^{-1} (X - \overline{X})$

MYT Decomposition Model

The observation vectors $(x_1, x_2, ..., x_p)$ are disintegrated into *p!* which generates the same overall T^2 statistic. For *p*=4, there are 4!=24 possible decompositions of the Hotelling's T^2 having 96 terms. Agog *et al*., (2014) presented stepwise procedure for T^2 decomposition using four variables. The result is as follows;

$$T^2 = T_1^2 + T_{2.1}^2 + T_{3.1,2}^2 + T_{4.1,2,3}^2$$
$$T^2 = T_1^2 + T_{3.1}^2 + T_{4.1,3}^2 + T_{2.1,3,4}^2$$
$$T^2 = T_1^2 + T_{4.1}^2 + T_{2.1,4}^2 + T_{3.1,2,4}^2$$
$$T^2 = T_1^2 + T_{2.1}^2 + T_{4.1,2}^2 + T_{3.1,2,4}^2$$
$$T^2 = T_1^2 + T_{3.1}^2 + T_{2.1,3}^2 + T_{4.1,2,3}^2$$
$$T^2 = T_1^2 + T_{4.1}^2 + T_{3.1,4}^2 + T_{2.1,3,4}^2$$
$$T^2 = T_2^2 + T_{1.2}^2 + T_{3.1,2}^2 + T_{4.1,2,3}^2$$
$$T^2 = T_2^2 + T_{3.2}^2 + T_{4.2,3}^2 + T_{1.2,3,4}^2$$
$$T^2 = T_2^2 + T_{4.2}^2 + T_{1.2,3}^2 + T_{3.1,2,4}^2$$
$$T^2 = T_2^2 + T_{1.2}^2 + T_{4.1,2}^2 + T_{3.1,2,4}^2$$
$$T^2 = T_2^2 + T_{3.2}^2 + T_{1.2,3}^2 + T_{4.1,2,3}^2$$
$$T^2 = T_2^2 + T_{4.2}^2 + T_{3.2,4}^2 + T_{1.2,3,4}^2$$
$$T^2 = T_3^2 + T_{1.3}^2 + T_{2.1,3}^2 + T_{4.1,2,3}^2$$
$$T^2 = T_3^2 + T_{2.3}^2 + T_{4.2,3}^2 + T_{1.2,3,4}^2$$
$$T^2 = T_3^2 + T_{4.3}^2 + T_{1.3,4}^2 + T_{2.1,3,4}^2$$
$$T^2 = T_3^2 + T_{1.3}^2 + T_{4.1,3}^2 + T_{2.1,3,4}^2$$
$$T^2 = T_3^2 + T_{2.3}^2 + T_{1.2,3}^2 + T_{4.1,2,3}^2$$
$$T^2 = T_3^2 + T_{4.3}^2 + T_{2.3,4}^2 + T_{1.2,3,4}^2$$
$$T^2 = T_4^2 + T_{1.4}^2 + T_{2.1,4}^2 + T_{3.1,2,4}^2$$
$$T^2 = T_4^2 + T_{2.4}^2 + T_{3.2,4}^2 + T_{1.2,3,4}^2$$
$$T^2 = T_4^2 + T_{3.4}^2 + T_{1.3,4}^2 + T_{2.1,3,4}^2$$
$$T^2 = T_4^2 + T_{1.4}^2 + T_{3.1,4}^2 + T_{2.1,3,4}^2$$
$$T^2 = T_4^2 + T_{2.4}^2 + T_{1.2,4}^2 + T_{3.1,2,4}^2$$
$$T^2 = T_4^2 + T_{3.4}^2 + T_{2.3,4}^2 + T_{1.2,3,4}^2$$

Computing the T^2 Decomposition Terms

One of the decomposition for the observation vector (x_1, x_2, x_3, x_4) is given as

$$T^2_{(x_1,x_2,x_3,x_4)} = T_1^2 + T_{3.1}^2 + T_{2.1,3}^2 + T_{4.1,2,3}^2 \quad (1)$$

The computation starts by first determining the value of the conditional term, $T^2_{4.1,2,3}$. From equation (4),

$$T^2_{4.1,2,3} = T^2_{(x_1,x_2,x_3,x_4)} - T^2_{(x_1,x_2,x_3)} \quad (2)$$

The variance-covariance matrix and the mean vector for the observation vector (x_1, x_2, x_3, x_4) is given as;

$$S_{44}=\begin{pmatrix} S_1^2 & S_{12} & S_{13} & S_{14} \\ S_{21} & S_2^2 & S_{23} & S_{24} \\ S_{31} & S_{32} & S_3^2 & S_{34} \\ S_{41} & S_{42} & S_{43} & S_4^2 \end{pmatrix} \text{ and } \overline{X}^{(4)}=\begin{pmatrix} \overline{x}_1 \\ \overline{x}_2 \\ \overline{x}_3 \\ \overline{x}_4 \end{pmatrix} \quad (3)$$

Thus the computation of $T^2_{(x_1,x_2,x_3,x_4)}$ is as follows;

$$T^2_{(x_1,x_2,x_3,x_4)}=(X^{(4)}-\overline{X}^{(4)})' S_{44}^{-1}(X^{(4)}-\overline{X}^{(4)}) \quad (4)$$

To obtain $T^2_{(x_1,x_2,x_3)}$, the original estimates of the mean vector and covariance structure is partitioned to obtain the mean vector and covariance matrix of the sub vector $X^{(3)}=(x_1,x_2,x_3)$. The corresponding partition is given as;

$$S_{33}=\begin{pmatrix} S_1^2 & S_{12} & S_{13} \\ S_{21} & S_2^2 & S_{23} \\ S_{31} & S_{32} & S_3^2 \end{pmatrix} \text{ and } \overline{X}=\begin{pmatrix} \overline{x}_1 \\ \overline{x}_2 \\ \overline{x}_3 \end{pmatrix} \quad (5)$$

Thus the computation $T^2_{(x_1,x_2,x_3)}$ is as follows;

$$T^2_{(x_1,x_2,x_3)}=(X^{(3)}-\overline{X}^{(3)})' S_{33}^{-1}(X^{(3)}-\overline{X}^{(3)}) \quad (6)$$

Also the decomposition of $T^2_{(x_1,x_2,x_3)}$ is given by

$$T^2_{(x_1,x_2,x_3)}=T_1^2+T_{3.1}^2+T_{2.1,3}^2 \quad (7)$$

Equation (10) above can be obtained by first computing the conditional term $T^2_{2.1,3}$ as follows;

$$T^2_{2.1,3}=T^2_{(x_1,x_2,x_3)}-T^2_{(x_1,x_3)} \quad (8)$$

To obtain the term $T^2_{(x_1,x_3)}$, the original estimates of the mean vector and the covariance structure is partitioned to obtain the mean vector and covariance matrix of the sub vector $X^{(2)}=(x_1,x_3)$. The corresponding partition is given as;

$$S_{22}=\begin{pmatrix} S_1^2 & S_{13} \\ S_{31} & S_3^2 \end{pmatrix} \text{ and } \overline{X}^{(2)}=\begin{pmatrix} \overline{x}_1 \\ \overline{x}_3 \end{pmatrix} \quad (9)$$

Hence, the computation of the term $T^2_{(x_1,x_3)}$ is as follows;

$$T^2_{(x_1,x_3)}=(X^{(2)}-\overline{X}^{(2)})' S_{22}^{-1}(X^{(2)}-\overline{X}^{(2)}) \quad (10)$$

Also, the decomposition for $T^2_{(x_1,x_3)}$ is given by

$$T^2_{(x_1,x_3)}=T_1^2+T_{3.1}^2 \quad (11)$$

The term $T^2_{3.1}$ is obtained by computing the T^2 value of the sub vector $X^{(1)}=(x_1)$. Hence, the unconditional term T_1^2 is computed by

$$T_1^2=\frac{(x_1-\overline{x}_1)^2}{s_1^2} \quad (12)$$

Thus, $T^2_{3.1}$ is computed as;

$$T^2_{3.1}=T^2_{(x_1,x_3)}-T^2_{(x_1)} \quad (13)$$

$$T_j^2 = \frac{(x_j - \bar{x}_j)^2}{s_j^2} \tag{14}$$

$j = 1, 2, \dots, p.$ is the square of a univariate *t* statistic for the observed value of the j^{th} variable of vector X.

3. RESULTS

Correlation Matrix

Before multivariate quality control chart can be implemented, the quality characteristics must be related. Thus, the first step in this analysis is the calculation of the correlation matrix between the four clinker minerals. The correlation matrix R below indicates that there is a strong relationship between the four clinker minerals and thus the multivariate technique is required.

$$R = \begin{pmatrix} 1 & -0.8171 & -0.3054 & -0.4214 \\ -0.8171 & 1 & 0.2141 & 0.7501 \\ -0.3054 & 0.2141 & 1 & 0.0190 \\ -0.4214 & 0.7501 & 0.0190 & 1 \end{pmatrix}$$

PHASE I

First phase process was applied for 100 samples in order to acquire reference data set in our study. UCL and the T^2 values were compared, observations with $T^2 \geq UCL$ were extracted from the data group. The reference data set was formed after an 8-step analysis.

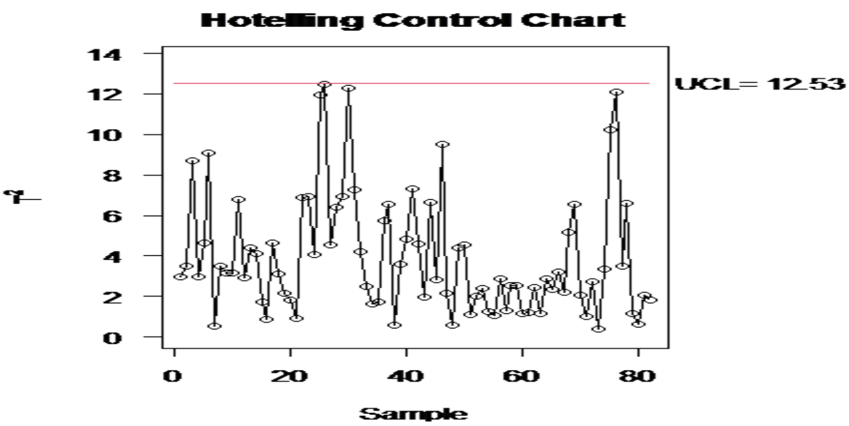

Fig. 1 Phase I Hotelling's T^2 control chart for 82 observations

Figure 1 Indicates that all 82 samples are smaller than the Upper Control Limit (UCL), that is, there is no uncontrolled situation and all samples take lower values than UCL. Depending on the other steps and especially on the results of the 8^{th} step, it is observed that it has reached a homogenous data set as reference data set. In this case, the first phase is completed. With the remaining 82 samples, we have the mean vector $\bar{X}$ and variance-covariance matrix which are estimated for the 2^{nd} Phase procedures.

PHASE II

After forming reference data set, the 2^{nd} Phase is applied to confirm the result obtained in Phase I analysis. New samples (30 samples) were applied in order to control the process and variance-covariance matrix coming from the reference data set and process average vector are used in acquiring T^2 statistics of these new samples.

Process Average Vector of reference data set is $\overline{X} = (55.75, 17.37, 7.28, 12.79)$ and the Covariance Matrix is

$$S = \begin{bmatrix} 39.00 & -36.000 & -1.30 & 0.120 \\ -36.00 & 34.000 & 1.20 & -0.086 \\ -1.30 & 1.200 & 0.60 & -0.380 \\ 0.12 & -0.086 & -0.38 & 0.440 \end{bmatrix}$$

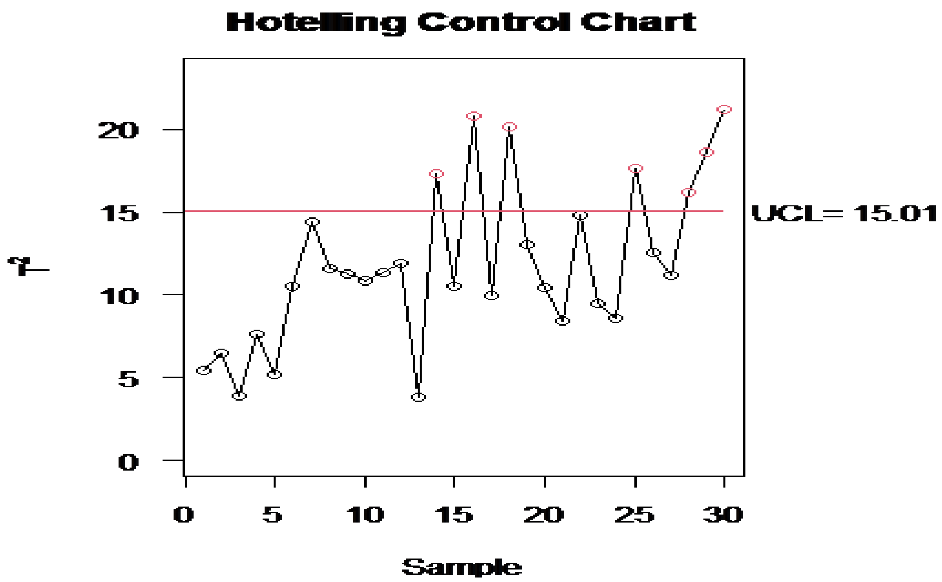

Fig. 2 Phase II Hotelling's T^2 control chart

Figure 2 shows that samples 14, 16, 18, 25, 28, 29 and 30 are out of control.

MYT Decomposition

Table -1 Unique Terms of the T^2 Decomposition for Phase II Chart

S/N	Components of Decompos	T^2 Values of the Sample Observations							Critical Value
	1-2-3-4	14	16	18	25	28	29	30	
1	1-0-0-0	3.7231	4.7775	0.1939	0.9078	1.7031	0.9078	1.6206	7.0434
2	2-0-0-0	3.2860	3.9372	0.2881	0.8481	1.3480	0.6692	1.0483	7.0434
3	3-0-0-0	2.4807	2.4807	**8.9707***	3.8507	6.8007	**7.4907***	3.8507	7.0434
4	4-0-0-0	**11.9184***	**12.9820***	**18.9820***	**15.2457***	**12.9820***	**16.4457***	**15.2457***	7.0434
5	1-2-0-0	4.1208	6.1567	0.6488	0.9272	2.4404	1.5860	4.0551	10.0040
6	1-3-0-0	8.4473	9.8175	9.1139	6.2120	**11.1374***	**10.5629***	7.3443	10.0040
7	1-4-0-0	**16.0409***	**18.2311***	**19.0808***	**16.3827***	**14.9702***	**17.5921***	**17.1686***	10.0040
8	2-3-0-0	7.8370	8.6921	9.0428	6.0889	**10.4986***	**10.0597***	6.4197	10.0040
9	2-4-0-0	**15.4904***	**17.2457***	**19.1757***	**16.2618***	**14.5233***	**17.2709***	**16.4799***	10.0040
10	3-4-0-0	**14.0307***	**15.6032***	**19.0959***	**17.1360***	**12.9892***	**16.5976***	**17.1360***	10.0040
11	1-2-3-0	8.8450	11.1967	9.5688	6.2314	11.8747	11.2411	9.7788	12.5773
12	1-2-4-0	**16.6506***	**20.0049***	**19.8210***	**16.4771***	**15.9998***	**18.5889***	**20.1609***	12.5773
13	1-3-4-0	**16.6189***	**18.9288***	**19.3180***	**17.4998***	**15.1877***	**17.5921***	**18.0341***	12.5773
14	2-3-4-0	**16.1481***	**18.0914***	**19.4659***	**17.4243***	**14.6865***	**17.2742***	**17.5556***	12.5773
15	1-2-3-4	**17.2908***	**20.8199***	**20.1025***	**17.6297***	**16.1731***	**18.5910***	**21.1968***	15.0085

Table 1 shows that the T^2 values in Bold and with Asterisk (*) are higher than their Critical Values. Hence, they have significantly contributed to the out-of-control signal of the samples they belong.

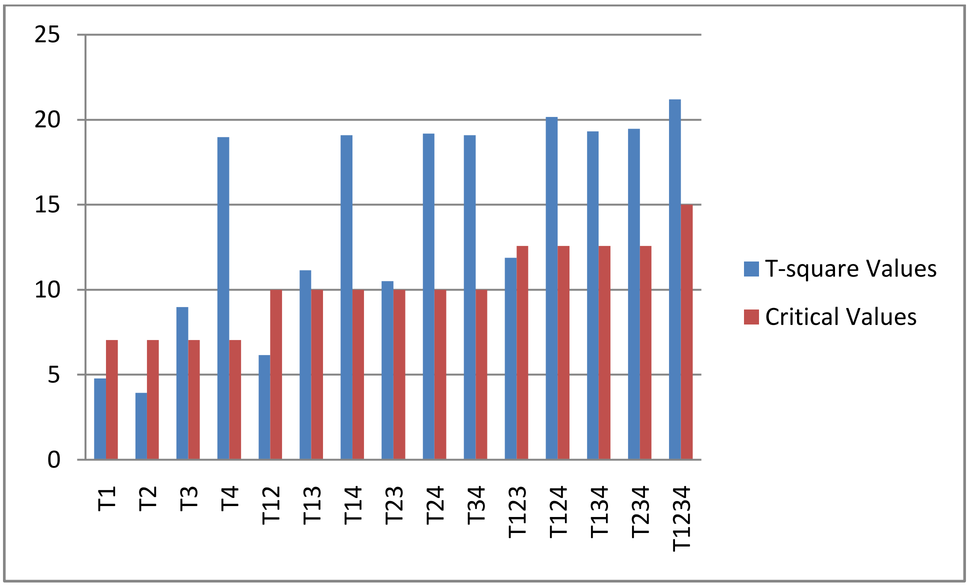

Fig. 3 Decomposition Bar Chart

Figure 3 shows that variables 3 (Aluminate) and 4 (Ferrite) are significant but variable 4 (Ferrite) is the influential variable because the bar of the T^2 value of variable 4 rise above the bar of the corresponding critical value more than any other variable. Hence, this idea is applicable to the bivariate relationship and the multivariate relationship too.

4. CONCLUSION

Observations composing of 100 samples regarding cement clinker variables in Phase I were acquired in order to form the reference data set in the application. It was found that 82 of these samples formed the reference data set as a result of 8 steps analyses. In Phase II, T^2 chart was constructed for 30 new observations and it was detected that 7 samples were out-of-control.

MYT decomposition method was used to calculate the unconditional and conditional terms, the variable(s) and the relationship among variables that caused or contributed to the out-of-control points were revealed. We found variable 4 (Ferrite) to be the influential variable.

When it is taken into consideration that many of the production processes or products are composed of more than one variable in real life, drawing only univariate $\overline{X}$ graphics would be insufficient. With these developments, process engineers will get the chance to keep lots of variables and observations under control during production. In statistical quality control studies, it must be determined which kind of quality control chart should be used. If there is more than one quality characteristic, control of these variables separately is wrong, multivariate quality control techniques which can enable the control of the variables together should be used. Therefore, we recommend the use of the Hotelling's T^2 decomposition in the control of cement production process, especially when we are dealing with many variables. R software (version 4.1.2) was used for the analysis with the aid of MSQC (Multivariate Statistical Quality Control) package downloaded through the Comprehensive R Archive Network (CRAN).

REFERENCES

[1]. Montgomery, D. C., and Runger G. C. (2003). Applied Statistics and Probability for Engineers. New York: John Wiley and Sons.

[2]. Agog, S. N., Dikko, H. G. and Asiribo, O. E. (2014). Decomposing Hotelling's T^2 statistic using four variables. *International Journal of Innovative Research in Science, Engineering and Technology,* 3(4), pp. 11449-11454.

[3]. Agog, S. N., Dikko, H. G. and Asiribo, O. E. (2014). Determining Out-of-Control Variable(s) in a Multivariate Quality Control Chart. *Scientia Africana*,13(2), pp. 266-280.

[4]. Woodall W.H. (2000). Controversies and contradictions in statistical process control. *Journal of Quality Technology*, 32(4), pp. 341-350.

[5]. Mason, R.L., Tracy, N.D. and Young, J.C. (1995). Decomposition of T^2 for multivariate control chart interpretation. *Journal of Quality Technology*, 27(2), pp. 99-108.
[6]. Mason, R.L., Tracy, N.D. and Young, J.C. (1997). Monitoring a multivariate step process. *Journal of Quality Technology*, 28(1), pp. 39-50.
[7]. Ulen, M. and Demir, I. (2013). Application of multivariate statistical quality control in pharmaceutical industry. *Balkan Journal of Mathematics*, 1(8), pp. 93-105.
[8]. Akeem, A. A., Yahaya, A. and Asiribo, O. E. (2015). Hotelling's T^2 Decomposition: Approach for Five Process Characteristics in a Multivariate Statistical Process Control. *American Journal of Theoretical and Applied Statistics*, 4(6), pp. 432-437.
[9]. Akeem, A. A., and Olatunji, I. O. (2017). MYT Decomposition and Its Invariant Attribute. *Research Journal of Mathematical and Statistical Sciences*, 5(2), pp. 14-22.