International Seminar on the Application of Science & Mathematics 2011
ISASM 2011

ROBUST HOTELLING'S T^2 CONTROL CHARTING IN SPIKE PRODUCTION PROCESS

Rohayu Mohd Salleh[1,2] and Maman A. Djauhari[1]

[1]Department of Mathematics, Universiti Teknologi Malaysia,
81310, Skudai, Johor, MALAYSIA

[2]Department of Science and Mathematics, Universiti Tun Hussein Onn Malaysia,
86400, Batu Pahat, Johor, MALAYSIA

[1]*rohayusalleh@yahoo.com*
[2]*maman@utm.my*

In this paper, we propose a control charting procedure based on robust Hotelling's T^2. For that purpose, a new stopping rule will be introduced to improve the computational efficieny of data concentration process used in Phase I operation. A simulation study will show the advantage of that robust approach compared to the classical one. A case study in spike production process in a Malaysian company will be reported.

Keywords: Hotelling's T^2; Mahalanobis distance; minimum covariance determinant; minimum vector variance; statistical process control.

1. INTRODUCTION

Each of manufacturers in manufacturing industry is playing a vital role in delivering a high quality of end product and, at the same time should fulfilling the increasing demand. To achieve such quality and productivity, that industry is basically equipped with the online system that can be used to handle large and high dimension data processing. In any production process, the overall quality is defined simultaneously by a number of quality characteristics. One way to monitor the quality is by using control charts which perform as key monitoring and investigating on the comparison of what is happening today with what happened previously. Therefore, the implementation of multivariate approach in process control charting is highly demanded and has showed a great contribution in industrial quality improvement.

To monitor the quality of the process, we may use Hotelling's T^2 control chart to take into account the correlation among those characteristics. This T^2 statistic is a scalar that combines information from the inversion of covariance matrix and mean vector of several variables. To implement the T^2 control chart based on individual observations which is our concern, for each observation i we calculate,

$$T^2(i) = (X_i - \bar{X})' S^{-1} (X_i - \bar{X}) \quad ; i = 1, 2, \ldots, n$$

where $X_1, X_2, \ldots, X_n$ is the reference sample obtained in Phase 1 operation and assumed to follow a p-variate normal distribution. Furthermore, $\bar{X} = n^{-1}\sum_i^{n=1} X_i$ and $S = (n-1)^{-1}\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})'$ are the estimates of the mean and covariance matrix of the process. The details of the implementation of T^2 control chart in practice can be seen in [1] and [2].

In this paper, we report our experience in using robust T^2 control chart in spike production in a Malaysian manufacturing company to ensure the stability of the production process. The name of that company is kept undeliverable due to its confidentiality. The study on the significant role of robust estimators can be easily found in the literature specifically in the construction of robust T^2 control chart in [3], [4] and [5].

By robust T^2 control charting we mean that the location and covariance matrix estimates in Phase I are determined based on robust method while process monitoring in Phase II is conducted in usual manner. Consequently, if we compare this control charting method with the classical (non-robust) T^2 control charting [6], the former is more effective in detecting the shift in mean vector than the later [7]. In this paper, we use the Fast Minimum Covariance Determinant (FMCD) since it ensures that the estimates are of high breakdown point. Since the stopping rule in FMCD makes high computational complexity in data concentration process, we introduce a new stopping rule which will reduce the computational complexity.

The rest of the paper is organized as follows. In the next section we recall and discuss the FMCD algorithm for location and scale estimation and then we define a new stopping rule in data concentration step. In Section 3 we present the result of the implementation of the proposed robust approach in spike production process monitoring. Some interesting results will be highlighted and a conclusion will be given in the last section.

2. PROPOSED ROBUST ESTIMATION METHOD IN PHASE I

The robust estimation of location and scale in Phase I operation based on the same approach as FMCD as discussed in [4]. But in this paper, we propose a new stopping rule to improve the computational efficieny of data concentration process used in Phase I operation. The next sub section will recall on the most popular and widely used robust estimation method.

2.1 Recall on FMCD

FMCD is one of the most popular and widely used robust estimation method of location and scale with high breakdown point. It was introduced by [8]. This algorithm consists of two main objectives, i.e., to order the data points in p dimensional space, $p > 1$, and to select the most concentrated data subset. The first objective is materialized by using Mahalanobis distance and the second is by minimizing the covariance determinant. More specifically, let $X_1, X_2, \ldots, X_n$ be a random sample from a p-variate normal distribution $N_p(\mu, \Sigma)$. To reach the first objective, FMCD consists of the following four steps:

Step 1. Select arbitrarily a subset H_{old} containing h data points, $h = \left\lceil \frac{n+p+1}{2} \right\rceil$, where $\lceil x \rceil$ represents the smallest integer greater than x.

Step 2. Compute the mean vector $\bar{X}_{H_{old}}$ and covariance matrix $S_{H_{old}}$ of all observations belonging to H_{old}.

Step 3. Compute $d^2_{H_{old}}(i) = \left(X_i - \bar{X}_{H_{old}}\right)^t S^{-1}_{H_{old}} \left(X_i - \bar{X}_{H_{old}}\right)$ for $i = 1, 2, ..., n$.

Step 4. Sort $d^2_{H_{old}}(i)$ for $i = 1, 2, ..., n$ in increasing order and let us write the result as $d^2_{H_{old}}(\pi(1)) \leq d^2_{H_{old}}(\pi(2)) \leq ... \leq d^2_{H_{old}}(\pi(n))$ where π is a permutation on $\{1, 2, ..., n\}$.

The most concentrated data subset is given by FMCD as follows:

Step 5. Construct $H_{new} = \{X_{\pi(1)}, X_{\pi(2)}, ..., X_{\pi(h)}\}$ and then calculate $\bar{X}_{H_{new}}$, $S_{H_{new}}$ and $d^2_{H_{new}}(i)$ for $i = 1, 2, ..., n$ using the same manner as in *Step 2*.

Step 6. If $\det(S_{H_{new}})$ = 0, repeat *Step 1* – *Step 5*. Otherwise, if $\det(S_{H_{new}}) \neq \det(S_{H_{old}})$, let $H_{old} := H_{new}$, $\bar{X}_{H_{old}} := \bar{X}_{H_{new}}$, and $S_{H_{old}} := S_{H_{new}}$. Then go to *Step 3*. Otherwise, the process is stopped and we get $\det(S_{H_{k-1}}) = \det(S_{H_k})$.

2.2 Discussion

In that algorithm, the stopping rule in *Step 6* is to minimize covariance determinant (MCD). The computational complexity of this stopping rule is of order $O(p^3)$ where p is the number of variables. This shows that *Step 6* needs a lot of operations and thus a lot of time to run. To reduce this time complexity, [9] proposed to use minimum vector variance (MVV) where the order of computational complexity is $O(p^2)$ which is far less than $O(p^3)$ especially for large p. In MVV algorithm, the *Step 6* is modified by the following:

*Step 6**. If $Tr(S^2_{H_{new}}) = 0$, repeat points *Step 1* – *Step 5*. Otherwise, if $Tr(S^2_{H_{new}}) \neq Tr(S^2_{H_{old}})$, let $H_{old} := H_{new}$, $\bar{X}_{H_{old}} := \bar{X}_{H_{new}}$, and $S_{H_{old}} := S_{H_{new}}$. Then go to *Step 3*. Otherwise, the process is stopped and we get $Tr(S^2_{k-1}) = Tr(S^2_k)$.

The *Step 6** is a significant improvement from *Step 6* in terms of computational complexity. However, both stopping rules in *Step 6* and *Step 6** are actually not necessary because n is finite and by construction of $H_{new} = X_{\pi 1}, X_{\pi 2},..., X_{\pi h}$, the number of iterations until the sequence $H_{old}, H_{new}, H_{old}, H_{new}, \ldots$ is convergent is finite as long as $S_{H_{new}}$ is non-singular. Therefore, in the next sub section we propose a new stopping rule in data concentration process.

2.3 New Stopping Rule in Data Concentration Process

Since the number of iterations in those algorithms is finite, then, to stop the computation process it is sufficient to test whether $H_{new} = H_{old}$. In other words, if I_{new} and I_{old} are the index set of all observations in H_{new} and H_{old} , respectively, then, it is sufficient to test whether $I_{new} = I_{old}$. With this new stopping rule, we propose the following data concentration process.

*Step 5**. Construct $H_{new} = X_{\pi 1}, X_{\pi 2},..., X_{\pi h}$.

*Step 6***. Let $I_{old} = \pi_{(1)}^{old}, \pi_{(2)}^{old}, \ldots, \pi_{(h)}^{old}$ and $I_{new} = \pi_{(1)}^{new}, \pi_{(2)}^{new}, \ldots, \pi_{(h)}^{new}$ be the index sets that correspond to the observations in H_{old} and H_{new}, respectively. If $I_{new} \neq I_{old}$, let $H_{old} := H_{new}$ and then go to *Step 2*. Otherwise, the process is stopped.

If in FMCD and MVV we calculate the determinant of covariance matrix and the vector variance, respectively, in the proposed stopping rule, there is nothing to calculate. Here, we only compare the index sets I_{new} and I_{old} . This is a non negligible advantage. Besides that, according to simulation experiments by using Matlab R2008a, here we report the other advantages of the proposed stopping rule in data concentration process.

(1) In the first experiment, we use n = 100, p = 2, 5, 7, 10 and 50 based on the distribution $N_p\ 0, I_p$ suggested in [10]. We find that the proposed data concentration gives the same robust Mahalanobis distance as FMCD and MVV.

(2) In term of computational efficiency in data concentration process, measured by running time allocation, the ratio IS: MCD: MVV is 1: 649: 52 where IS refers to the comparison of index sets. This means that if IS needs 1 second, MCD and MVV need 649 second and 52 second, respectively. This result is obtained from the second simulation experiment with p = 1000 and n = 5000. For p = 500 and n = 2500, the ratio is IS : MCD : MVV = 1: 210 : 13.

Another simulation experiment has conducted to evaluate the performance between classical T^2 control chart and robust T^2 control chart using the proposed stopping rule

(6**). The performance will be compared and measured by the percentage of out of control signal occurred in Phase II. First, for Phase I operation we generate n = 500 observations for each p = 2, 3, 5, and 10 from p-variate normal distribution $N_p\left(0, I_p\right)$ suggested in [10]. Then, we estimate the parameter based on both methods. Second, for Phase II monitoring operation, n = 50000 observations are generated from a p-variate normal mixture model $\left(1-\varepsilon\right) N_p\left(\vec{\mu}_1, I_p\right) + \varepsilon N_p\left(\vec{\mu}_2, I_p\right)$ for each p = 2, 3, 5, and 10 where $\varepsilon = 0.2$, $\vec{\mu}_1 = \vec{0}$, $\vec{\mu}_2 = 5\vec{e}$, and $\vec{e}$ is a vector of p dimension. Based on those experiments, the performance of robust T^2 can be summarized as follows:

(1) In terms of the percentage of out of control signals, the robust T^2 control chart detects 100% while the classical one detects around 99%.

(2) In terms of the percentage of false negative, robust T^2 does not show any false negative for each p. Meanwhile, classical T^2 gives false negative more than 15% for all p.

(3) In terms of false positive produce by robust T^2 approach is less than 5% but the classical gives around 1%.

We conclude that the proposed robust estimation method presented in this section is as effective as FMCD and MVV methods but with lower computational complexity in terms of running time. Due to those commendable properties, in the next section, we implement the proposed method in monitoring the production of spike.

3. INDUSTRIAL EXAMPLE: SPIKE PRODUCTION PROCESS CONTROL

Spike is one of the components of intravenous drip and blood transfusion set in medical devices industry. The quality of this product is basically based on the behavior of a set of p = 2 interrelated CTQ (critical to quality), namely, (i) inner diameter of the hole as can be seen in Fig. 1 and (ii) length of the product from the top to bottom as can be seen in Fig. 2.

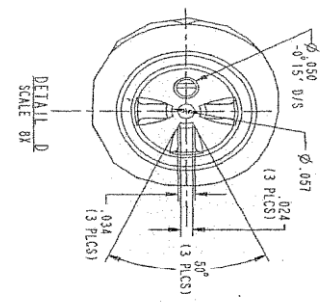

Fig. 1 Technical drawing (cross section hole view)

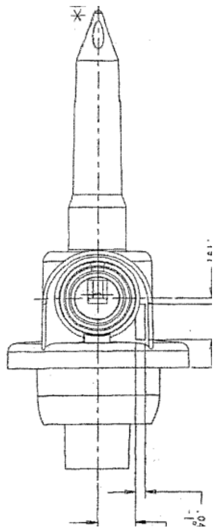

Fig. 2 Technical drawing (side view)

The historical data set (HDS) for Phase I operation consists of 47 observations. By using the algorithm in the previous section with IS as the stopping rule, we find that seven observation is outlier. Therefore, the reference sample consists of n = 40 observations. From this reference sample we get the following estimates of process mean vector and covariance matrix,

$$\bar{X} = \begin{pmatrix} 1.9576 \\ 0.5698 \end{pmatrix} \text{ and, } S = \begin{pmatrix} 2.04E-06 & -2.5E-07 \\ -2.5E-07 & 2.71E-07 \end{pmatrix}$$

It is important to note that based on classical Phase I operation, there is only one outlier found. The above results on parameter estimates are used in Phase II operation and the monitoring of process mean will be based on 282 individual observations. The T^2 statistic and the corresponding control chart are visualized in Fig. 3. The upper control limit is UCL = 14.5983 at the probability of false alarm α = 0.0027.

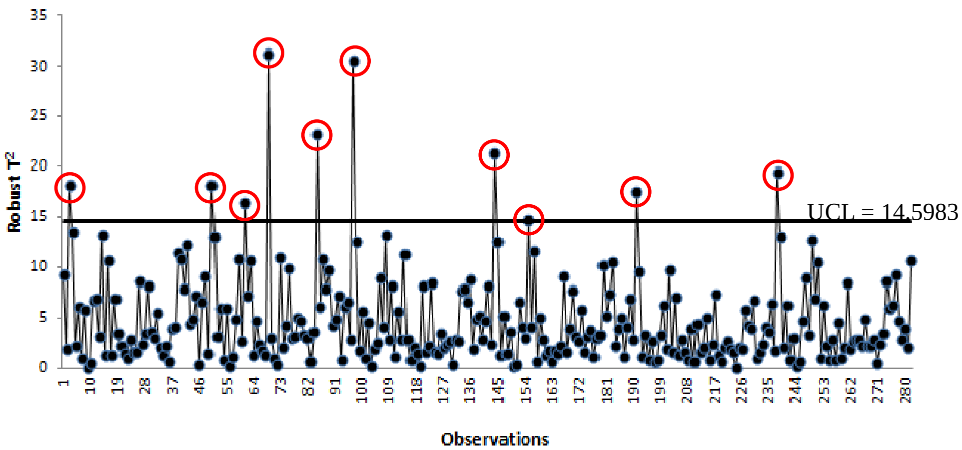

Fig. 3 Robust T^2 control chart

In Fig. 3, an out of control signal occurs at ten observations which are indicated by red circle. However, classical T^2 control chart in Fig. 4 give the different message.

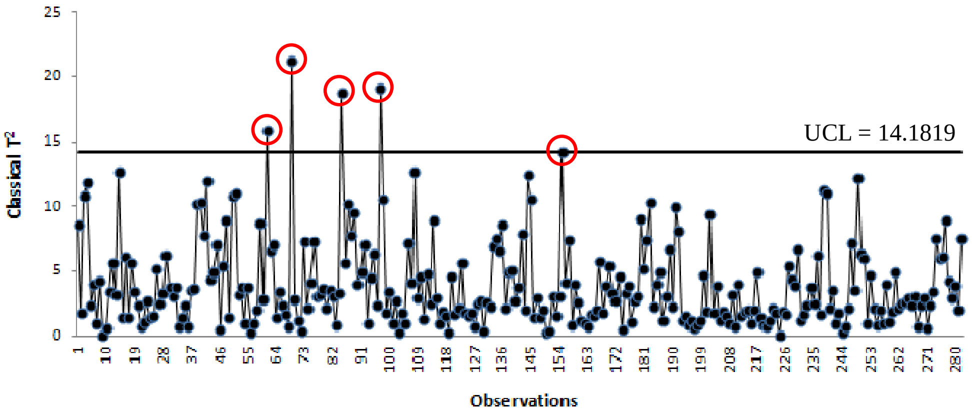

Fig. 4 Classical T^2 control chart

This real industrial examples shows that the implementation of robust T^2 control chart in monitoring the process mean have a significant with early findings in performance simulation. The classical control chart is unable to detect a few number of out of control observations which is equivalent to the false negative. Thus, the robust T^2 control chart is more powerful to detect the shifts than the classical approach.

4. CONCLUSION

If in the original stopping rule needs to compute the determinant of covariance matrix in each iteration, and MVV needs the computation of vector variance, in the proposed stopping rule there is no arithmetical computation. All we have to do is to test the equality of two index sets. This is only a logical comparison.

In terms of running time, the proposed stopping rule is quite promising. For example, for p = 500 and n = 2500, the ratio of the running time (in unit time) IS : FMCD : MVV is 1 : 210 : 13.

Based on simulation experiments, classical T^2 control chart is not good as robust T^2 in detecting out of control signal. This justifies the statement in [7] that T^2 statistic is not effective in detecting sustained step changes in the mean vector.

A case study on spike production process gives a significant result by implemented the proposed robust Phase I operation.

Acknowledgement

We acknowledge financial support from the Government of Malaysia via Ministry of Higher Education through Fundamental Research Grant Schemes vote number 4F013 and Research University Grant Q.J13000.7126.02H18. The authors would like to thank Universiti Teknologi Malaysia and Universiti Tun Hussein Onn Malaysia for the opportunity to do this research. Special thanks go to Ms Basar for the permission to use her data in this research.

References

[1] Tracy, N. D.; Young, J. C.; and Mason, R. L. (1992). Multivariate Control Charts for Individual Observations. *Journal of Quality Technology*, 24: 88-95.

[2] Montgomery, D. C. (2005). Introduction to Statistical Quality Control. 5th Ed, New York: John Wiley and Sons.

[3] Vargas, J. A. (2003). Robust Estimation in Multivariate Control Charts for Individual Observations. *Journal of Quality Technology*, 35: 367–376.

[4] Jensen, W. A.; Birch, J. B.; and Woodall, W. H. (2007). High Breakdown Point Estimation Methods for Phase I Multivariate Control Charts. *Quality and Reliability Engineering International*, 23: 615 – 629.

[5] Chenouri, S.; Steiner, S. H.; and Mulayath, A. (2009). A Multivariate Robust Control Chart for Individual Observations. *Journal of Quality Technology*, 41: 259 – 271.

[6] Yanez, S.; Gonzalez, N.; and Vargas, J. A. (2010). Hotelling's T^2 Control Charts Based on Robust Estimators. *Dyna*, 7: 239 – 247.

[7] Sullivan, J. H. & Woodall, W. H. (1996). A comparison of multivariate control charts for individual observation. *Journal of Quality Technology*, 28: 398-408.

[8] Rousseeuw, P. J. & Van Driessen, K. (1999). A Fast Algorithm for the Minimum Covariance Determinant Estimator. *Technometrics*, 41: 212 – 223.

[9] Djauhari, M. A.; Mashuri, M.; and Herwindiati, D. E. (2008). Multivariate Process Variability Monitoring. *Journal of Communication and Statistics – Theory and Methods*, 37: 1742-1754.

[10] Mason, R. L.; Chou, Y. M.; and Young, J. C. (2009). Monitoring Variation in a Multivariate Process When the Dimension is Large Relative to the Sample Size. *Journal of Communication in Statistics – Theory and Methods*, 38: 939-951.