



ISSN 1823-626X

# Journal of Fundamental Sciences

available online at <http://jfs.ibnusina.utm.my>



## Monitoring the variability of beltline moulding process using Wilks's statistic

Aishah Mohd Noor and Maman A. Djauhari,

*Department of Mathematics, Faculty of Science, University Technology Malaysia.*

Received 4 May 2010, Revised 30 June 2010, Accepted 7 July 2010, Available online 26 October 2010

### ABSTRACT

Due to the increase of the complexity of customer demand on products and services, monitoring process quality is becoming multivariate in nature. In this setting there are two important parameters to be monitored, i.e., the mean vector and the covariance structure which determines the variability of the process. This paper deals with process variability monitoring of beltline moulding process at an automotive industry where the process is in multivariate setting and monitoring process is based on individual observations. Our approach is based on Wilks's statistic. A real application will be presented and the strength of that statistic, as well as its limitations, will be discussed.

| Multivariate statistical process control | Process variability | Covariance matrix | Wilks's statistic |

© 2010 Ibnu Sina Institute. All rights reserved.

### 1. INTRODUCTION

The recent growth of research interest and developments in multivariate control charting technique has been significantly contributed to the field of multivariate statistical process control (SPC). In many industrial applications, there is an increasing demand to implement multivariate SPC in a production process for quality improvements. This is because, in many situations quality of a product is determined by two or more correlated process variables. Thus, the multivariate control charting technique plays an important role in any multivariate process quality improvement initiative where monitoring simultaneously several correlated process variables is required. The univariate control chart could still be used for monitoring these process variables individually. However, ignoring the interrelationship will be inappropriate. Although, the multivariate control charting opens up opportunity in considering the correlation among process variables, the complexity in the applications of multivariate SPC charting technique in terms of understanding the multivariate process properties itself is a challenging job to practitioners.

The application of multivariate SPC requires an understanding of two different phases in control charting procedures that is phase I and phase II operations. According to Woodall [1], it is important to distinguish between the applications of those two phases. The objective of phase I is to determine whether a process was in-control when historical data set (HDS) is collected.

In this phase, the initial control limits are established and, if needed, recalculated in order to bring the process to the in-control state. In addition, if there is no information about the in-control process parameters  $\mu_0$  and  $\Sigma_0$ , both parameters could be estimated from the HDS. Meanwhile, in phase II, the objective is to investigate whether the process remains in-control when a future observation is drawn. In this phase, the control charts are used for monitoring and detecting any changes in the process mean vector or covariance matrix from the in-control process parameters.

In this paper we focus only on monitoring the process variability which is, see e.g. Alt [2], equivalent to performing repeatedly the hypothesis testing  $H_0: \Sigma = \Sigma_0$  against  $H_1: \Sigma \neq \Sigma_0$  where  $\Sigma_0$  is the in-control process covariance matrix. More specifically, on monitoring the process variability based on individual observations. In this case, many control charting procedures are available. For example, Khoo and Quah [3] proposed a test statistic which is the successive differences between multivariate observations with an assumption that  $\Sigma_0$  is known. They showed that the test statistic is distributed as  $\chi_p$  when the process is in-control. One of the advantages of their control charting procedure is that it could be readily used since the control limits is easily obtained from that distribution. However, the statistic at one sample point  $M_t$  and that at the next sample point  $M_{t+1}$  are not independent which can lead to inappropriate interpretation of the source of an out-of-control signal.

*Corresponding author at: Department of Mathematics, Faculty of Science  
University Technology Malaysia, UTM 81000 Skudai, Johor, Malaysia.  
E-mail address: aishah.mm@unimap.edu.my (Aishah Mohd Noor)*

Tracy et al. [4] proposed a multivariate control chart based on Hotelling  $T^2$  statistic for monitoring process in start-up stage when only individual observations are available. The proposed statistic  $Q_i$  is used to detect any shift in mean vector. This statistic is also sensitive to the change in process covariance matrix. The  $Q_i$  statistic (times a constant) follows a Beta distribution with  $p/2$  and  $(n-p-1)/2$  degrees of freedom if the observation vector  $\mathbf{x}$  is not independent of  $\bar{\mathbf{x}}$  and  $\mathbf{S}$  in which  $\bar{\mathbf{x}}$  and  $\mathbf{S}$  are the estimated values of parameters  $\boldsymbol{\mu}_0$  and  $\Sigma_0$  respectively. For phase II operation where the future observations are independent from phase I, the  $Q_i$  statistic (times a constant) follows an  $F$  distribution with  $p$  and  $(n-p)$  degrees of freedom. Sullivan and Woodall [5] demonstrated comparison of multivariate control charts for individual observations case. Several alternative approaches for estimating  $\Sigma$  are presented. They observed that the control chart employing the familiar covariance matrix estimator  $\mathbf{S}_1$ , that is obtained from the pooled observations is ineffective in detecting shift in the mean vector. They showed that control chart based on differences between each successive pair of observation vectors to estimate  $\Sigma$  performs much better than that when estimator  $\mathbf{S}_1$  is employed. The recent technology in monitoring process variability based on individual observations is presented by Mason et al. [6]. The methodology is based on original work by Wilks [7,8] in the detection of multivariate outliers. The control statistic,  $W$ , is the ratio between the determinant of sample covariance matrices issued from the augmented data set (ADS) in phase II and that of phase I (HDS). The ADS refers to a set of data consisting of HDS and a new observation obtained during phase II. In this paper, we discuss the used of the statistic  $W$  in order to monitor changes in the process variability characterized by the covariance matrix  $\Sigma$  of size  $(p \times p)$ , where  $p$  is the number of process variables.

This paper is organized as follows. In Section 2, we describe the procedure to develop a control procedure based on Wilk's statistic. The control chart based on  $W$  statistic is presented in Section 3. An example based on data from automotive industry is demonstrated in Section 4 to illustrate its application. In the last section, the conclusions are given.

## 2. CONTROL PROCEDURE BASED ON WILK'S STATISTIC

Suppose that we have a data matrix  $\mathbf{X}$  of size  $(n \times p)$  that consists of  $n$  random  $p$ -dimensional observations from an in-control process in phase I. We will regard this data matrix as HDS and we assume that those observations follow a multivariate normal distribution with mean vector  $\boldsymbol{\mu}_0$  and covariance matrix  $\Sigma_0$ ,  $N_p(\boldsymbol{\mu}_0, \Sigma_0)$ . For simplicity, we write the data matrix  $\mathbf{X}$  in matrix notation:

$$\mathbf{X}_{HDS} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & \ddots & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & \dots & \dots & x_{np} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_n' \end{bmatrix} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]' \quad (1)$$

where  $\mathbf{x}_i' = [x_{i1}, x_{i2}, \dots, x_{ip}]$ .

Based on HDS, the unknown parameters  $\boldsymbol{\mu}_0$  and  $\Sigma_0$  are estimated, respectively, as follows

$$\mathbf{S}_{HDS} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}}_{HDS})(\mathbf{x}_i - \bar{\mathbf{x}}_{HDS})' \quad (2)$$

where  $\bar{\mathbf{x}}_{HDS} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix}$  is the sample mean vector issued from

HDS with  $\bar{x}_i = \frac{1}{n} \sum_{r=1}^n x_{ri}$  as  $i$ th variable.

In order to perform the phase II operation, where monitoring any departure from in-control process variability is conducted based on individual observations and  $\Sigma_0$  is estimated by  $\mathbf{S}_{HDS}$ , the methodology proposed by Mason et al. [6] for monitoring a shift in a covariance matrix is as follows.

Let us now consider the data matrix ADS

$\mathbf{X}_{ADS} = [\mathbf{x}_1', \mathbf{x}_2', \dots, \mathbf{x}_{n+1}']'$ . The sample covariance matrix of ADS is

$$\mathbf{S}_{ADS} = \frac{1}{n} \sum_{i=1}^{n+1} (\mathbf{x}_i - \bar{\mathbf{x}}_{ADS})(\mathbf{x}_i - \bar{\mathbf{x}}_{ADS})' \quad (3)$$

where  $\bar{\mathbf{x}}_{ADS} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix}$  is the sample mean vector issued from

ADS and here  $\bar{x}_i = \frac{1}{n+1} \sum_{r=1}^{n+1} x_{ri}$ .

To monitor a shift in process variability, they used the following Wilk's statistic [7,8].

$$W = \frac{|\mathbf{SS}_{HDS}|}{|\mathbf{SS}_{ADS}|} = \left( \frac{n-1}{n} \right)^p \frac{|\mathbf{S}_{HDS}|}{|\mathbf{S}_{ADS}|} \quad (4)$$

where  $SS_{HDS}$  and  $SS_{ADS}$  represent the scatter matrix of HDS and ADS.

Mason et al. [6] pointed out that the new observation  $\mathbf{x}_{n+1}$  is said to be similar to the observations in HDS if  $\mathbf{x}_{n+1}$  does not increase the volume of space provided by HDS regardless additional new observation is included. However, if the volume of space provided by HDS is increased after a new observation  $\mathbf{x}_{n+1}$  is added, we conclude that the new observation is different than HDS. Thus, by using the relationship between the determinant of a sample covariance matrix and the volume of a  $p$ -dimensional parallelotope, one could measure the change in the volume of space provided by HDS when a new observation is drawn. For example, if  $|S_{ADS}|$  is less than  $|S_{HDS}|$ , then the values of  $W$  will be large. On the other hand, if  $|S_{ADS}|$  is greater than  $|S_{HDS}|$ , the values of  $W$  will be small.

Thus, if  $\mathbf{x}_{n+1}$  does not increase the volume of the space provided by HDS,  $W$  is large. In contrast,  $\mathbf{x}_{n+1}$  which increases that volume will give small  $W$ . Hence, in order to detect the shift in a covariance matrix when a new observation  $\mathbf{x}_{n+1}$  is added, the ratio between the

determinant of  $S_{HDS}$  and that of  $S_{ADS}$  that is  $\frac{|S_{HDS}|}{|S_{ADS}|}$  could be used. The influence of  $\mathbf{x}_{n+1}$  on the covariance structure will be reflected by the  $W$  statistic. Thus, the shift in the process variability could be monitored by constructing a control chart based on  $W$  statistic.

Wilks showed that the possible value of  $W$  is between zero and one. If  $|S_{HDS}|$  is similar to  $|S_{ADS}|$ , the values of  $W$  is close to one. Otherwise if the values of  $W$  is close to zero, these two determinants are said to be dissimilar and  $|S_{ADS}| > |S_{HDS}|$ . We conclude that the  $S_{ADS}$  has different variability than that given by  $S_{HDS}$ . In other words, the process variability is shifted when new observation is available.

Control procedure based on Wilk's statistic is to plot the values of  $W$  together with the values of its upper control limit (UCL) and lower control limit (LCL). These control limits will be discussed in the next section.

### 3. CONTROL LIMITS

The upper control limit (UCL) and the lower control limit (LCL) are calculated using the distribution of  $W$ . Since the values of  $W$  are between zero and one, while values near zero indicate an existence of the shift in process variability, the out of control region is at the left region of that distribution. Thus, an out-of-control signal will occur as soon as the value of  $W$  is less than LCL. Consequently, the

UCL is not required. To calculate the LCL, we need the distribution of  $W$  under in-control state.

Consider again the  $W$  statistic in (4). To determine the distribution of  $W$  under in-control state, Mason et al. [6] used the relationship between  $W$  with Hotelling's  $T^2$  statistic. It can be seen, see Rencher [9], that  $W$  can be written as a function of  $T^2$  statistic and can be expressed as

$$W = 1 - \frac{(n+1)}{n^2} T^2 \tag{5}$$

where

$$T^2 = (\mathbf{x}_{n+1} - \bar{\mathbf{x}}_{ADS})' S_{ADS}^{-1} (\mathbf{x}_{n+1} - \bar{\mathbf{x}}_{ADS}) \tag{6}$$

Since  $\mathbf{x}_{n+1}$  is included in the calculation of  $\bar{\mathbf{x}}_{ADS}$  and  $S_{ADS}$ , thus,  $\mathbf{x}_{n+1}$  is not independent of both  $\bar{\mathbf{x}}_{ADS}$  and  $S_{ADS}$ . Mason et al. [6] used equation (5) in order to derive the distribution for  $W$ . Let us first consider the  $T^2$  in phase I. If the true population parameters  $\boldsymbol{\mu}_0$  and  $\Sigma_0$  are known, for all  $i = 1, 2, \dots, n$ ,

$$T^2 = (\mathbf{x}_i - \boldsymbol{\mu}_0)' \Sigma_0^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_0) \tag{7}$$

is distributed as  $\chi_p$  (e.g. see Mason and Young, [10] or Seber, [11]). Otherwise, if those parameters are unknown, they must be estimated by  $\bar{\mathbf{x}}_{HDS}$  and  $S_{HDS}$  respectively, the  $T^2$  statistic in (7) can be expressed as

$$T^2 = (\mathbf{x}_i - \bar{\mathbf{x}}_{HDS})' S_{HDS}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}_{HDS}) \tag{8}$$

The distribution of  $T^2$  statistic in (8), (see Tracy and Young, [4]), is a scalar multiplication of Beta distribution with degrees of freedom,  $p/2$  and  $(n-p-1)/2$ ,

$$T^2 = (\mathbf{x}_i - \bar{\mathbf{x}}_{HDS})' S_{HDS}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}_{HDS}) \sim \frac{(n-1)^2}{n} \text{Beta}_{(p/2, (n-p-1)/2)} \tag{9}$$

Now, we consider the  $T^2$  in phase II. In this case, the distribution of  $T^2$  statistic in (6) is

$$T^2 = (\mathbf{x}_{n+1} - \bar{\mathbf{x}}_{ADS})' S_{ADS}^{-1} (\mathbf{x}_{n+1} - \bar{\mathbf{x}}_{ADS}) \sim \frac{n^2}{n+1} \text{Beta}_{(p/2, (n-p)/2)} \tag{10}$$

It follows that

$$\frac{n+1}{n^2} T^2 \sim \text{Beta}_{(p/2, (n-p)/2)} \tag{11}$$

In order to determine the distribution of  $W$ , the following property of beta distribution is applied (see Mason et al., [12]). If the variable  $y$  is distributed as  $Beta_{(a,b)}$ , then the variable  $x = 1 - y$  is distributed as  $Beta_{(b,a)}$ . This implies that,

$$W = 1 - \frac{(n+1)}{n^2} T^2 \sim Beta_{((n-p)/2, p/2)} \tag{12}$$

Based on this distribution, the lower control limit of  $W$  control charting is

$$LCL = Beta_{((n-p)/2, p/2)} \tag{13}$$

#### 4. AN ILLUSTRATIVE EXAMPLE

In this section, we consider a set of data from the beltline moulding process consisting of 57 individual observations with  $p = 8$  process variables collected at certain time interval from an automotive industry obtained from Bon [13]. We used this data set to illustrate the application of control chart based on  $W$  statistic for monitoring multivariate process variability. We first established the phase I in-control process and obtained the estimates of unknown process parameters of mean vector  $\mu_0$  and covariance matrix  $\Sigma_0$ . For this purpose, we use the first 30 observations as HDS,  $n = 30$  and the last 27 observations will be used in phase II. We assume that these observations are random and follow a multivariate normal distribution.

From the phase I operation, we calculated the sample covariance matrix of HDS,  $S_{HDS}$ . We get the following in-control sample covariance matrix

$$S_{HDS} = \begin{pmatrix} 0.0545 & -0.0155 & 0.1147 & 0.0497 & 0.0189 & 0.0614 & -0.0063 & -0.0034 \\ & 0.0857 & 0.0047 & 0.0564 & 0.0039 & 0.0434 & -0.0026 & -0.0193 \\ & & 0.4482 & 0.2331 & 0.0577 & 0.2789 & -0.0248 & -0.0286 \\ & & & 0.5459 & -0.0009 & 0.1052 & -0.0115 & -0.0508 \\ & & & & 0.0399 & 0.0590 & -0.0055 & -0.0085 \\ & & & & & 0.5284 & -0.0261 & -0.0423 \\ & & & & & & 0.0362 & 0.0135 \\ & & & & & & & 0.0271 \end{pmatrix}$$

and the determinant of  $S_{HDS}$  is  $|S_{HDS}| = 1.4201 \times 10^{-9}$ .

Now, we perform the phase II operation. We want to assess whether the process variability is still in-control when new observation is available. In this phase, the aim is to detect a shift in the process covariance matrix. Table 1 shows the values of  $|S_{ADS}|$  for each of the 27 observations.

Table 2 presents the corresponding values of  $W$  statistic for each ADS. The values of  $W$  are calculated based on equation (4) for  $n = 30, p = 8, |S_{HDS}| = 1.4201 \times 10^{-9}$  and substituting  $|S_{ADS}|$  for each  $k$  where  $k = 1, 2, \dots, 27$ . The

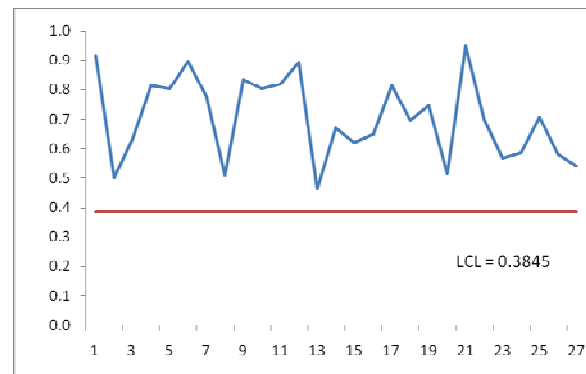
change in the process variability is then monitored by plotting the values of  $W$  in Table 2 on the control chart together with the value of LCL. According to Equation (13), for  $\alpha = 0.0027$  we get  $LCL = 0.3845$ . It is the  $\alpha$ th quantile of  $Beta_{(11,4)}$ .

**Table 1:** The sample generalized variances of each ADS.

$k$	$ S_{ADS} $	$k$	$ S_{ADS} $	$k$	$ S_{ADS} $
1	1.1806E-09	10	1.3480E-09	19	1.4500E-09
2	2.1515E-09	11	1.3213E-09	20	2.1044E-09
3	1.7092E-09	12	1.2132E-09	21	1.1429E-09
4	1.3315E-09	13	2.3436E-09	22	1.5523E-09
5	1.3469E-09	14	1.6186E-09	23	1.9087E-09
6	1.2104E-09	15	1.7442E-09	24	1.8483E-09
7	1.3952E-09	16	1.6706E-09	25	1.5324E-09
8	2.1258E-09	17	1.3307E-09	26	1.8615E-09
9	1.2958E-09	18	1.5578E-09	27	1.9987E-09

**Table 2:** The values of  $W$  statistic.

$k$	$W$	$k$	$W$	$k$	$W$
1	0.9171	10	0.8032	19	0.7467
2	0.5032	11	0.8194	20	0.5145
3	0.6335	12	0.8924	21	0.9474
4	0.8132	13	0.4620	22	0.6975
5	0.8039	14	0.6689	23	0.5673
6	0.8946	15	0.6208	24	0.5858
7	0.7761	16	0.6481	25	0.7066
8	0.5093	17	0.8136	26	0.5817
9	0.8356	18	0.6950	27	0.5417



**Figure 1:** Control chart of  $W$  statistic for 27 individual observations

Figure 1 shows that there is no signal of out-of-control state. Thus, we conclude that the process variability remains in-control.

#### 5. CONCLUSION

In this paper, we discuss a multivariate process variability monitoring based on individual observations using Wilk's statistic. The control statistic  $W$  is defined as the ratio between two sample generalized variances given

by HDS and that given by ADS. The advantage of  $W$  statistic is due to user's familiarity with sample generalized variance. Moreover, the exact distribution of this statistic is known. Thus, the control limit can be exactly determined. However, the statistic is not free from limitations. For example, the value of sample generalized variance can be possibly zero in which case there exist a variable that is a linear combination of other variables. In another example, the two sample generalized variances can give the same

value but they are different to each other in their covariance structures.

## **ACKNOWLEDGEMENT**

We acknowledge financial support from the Ministry of Science, Technology and Innovation of Malaysia via FRGS grant number 78484.

## **REFERENCES**

- [1] W.H. Woodall, Controversies and contradictions in statistical process control, *J. Qual. Technol.* 32 (2000) pp. 341-350.
- [2] F.B. Alt, Multivariate quality control, *The Encyclopaedia of Statistical Sci.*, John Wiley & Sons, New York, NY, (1985) pp. 110-122.
- [3] M.B. Khoo and S.H. Quah, Multivariate control chart for process dispersion based on individual observations, *Qual. Engineering*, 15 (2003) pp. 639-642.
- [4] N.D. Tracy and J.C. Young, Multivariate control charts for individual observations, *J. Qual. Technol.* 24 (1992) pp. 88-95.
- [5] J.H. Sullivan and W.H. Woodall, A comparison of multivariate control chart for individual observation, *J. Qual. Technol.* 28 (1996) pp. 398-408.
- [6] R.L. Mason, Y.M. Chou and J.C. Young, Monitoring variation in a multivariate process when the dimension is large relative to sample size, *Commun. Statist. Theor. and Meth.* 38 (2009) pp. 939-951.
- [7] S.S. Wilks, *Mathematical statistics*, John Wiley & Sons, New York, NY, (1962).
- [8] S.S. Wilks, Multivariate statistical outliers, *Sankhya A*, 25 (1963) pp. 407-426.
- [9] A.C. Rencher, *Methods of multivariate analysis*, 2nd ed. John Wiley & Sons, New York, NY, (2002).
- [10] R.L. Mason and J.C. Young, *Multivariate statistical process control with industrial applications*, ASA-SIAM, Philadelphia, (2002).
- [11] G.A.F. Seber, *Multivariate observations*, John Wiley & Sons, New York, NY, (1984).
- [12] R.L. Mason, Y.M. Chou and J.C. Young, Identifying variables contributing to outliers in phase I, *Commun. Statist. Theor. and Meth.* 37 (2008) pp. 1103-1118.
- [13] A.T. Bon, *Process Quality Improvement on Beltline Moulding Manufacturing*, PhD Thesis (2008).