

Statistical Process Control for a Limited Amount of Data

José Gomes Requeijo¹, António Abreu^{2,3} and Ana Sofia Matos¹

¹UNIDEMI, Departamento de Engenharia Mecânica e Industrial, Faculdade de Ciências e Tecnologia,
Universidade Nova de Lisboa, 2829-516 Caparica, Portugal

²ISEL/IPL – Instituto Politécnico de Lisboa, Lisboa, Portugal

³CTS – Uninova, Lisboa, Portugal

Keywords: SPC (Statistical Process Control), *Q* Control Charts, *MQ* Control Charts, Process Capability.

Abstract: Some production systems control many quality characteristics with a restricted amount of data, not allowing a convenient estimation of the process parameters (mean and variance), thereby creating a difficulty in implementing the traditional Statistical Process Control (SPC). In order to address this question, the approach suggested is to adopt the developments proposed by Charles Quesenberry, which consists in the statistics sample transformation at time i . This transformation is based on a parameter estimation at time $(i - 1)$. This paper addresses two situations, the univariate and multivariate SPC, with the use of Q dimensionless statistics. Both univariate (Q) and multivariate (MQ) statistics are distributed according to standard Normal distribution. It is also suggested the application of new capability indices Q_L and Q_U to study the univariate process capability, which are represented in the mean Q control chart to evaluate in real time the performance of the various processes and predict the possibility of production of nonconforming product, which will increase customer satisfaction. The methodology is applicable to different production systems, both for industry and services. Based on a methodology developed, a case study is presented and discussed.

1 INTRODUCTION

Initially the SPC was presented by Shewhart (1931) in Bell Telephone Laboratories. The Shewhart's developments were a valuable contribution to the continuous improvement of quality.

However, the traditional Shewhart control charts have some limitations such as: it considers only a characteristic of quality control, and its implementation in a real process requires a large amount of data. In order to implement these control charts, it is necessary to ensure the following principles:

- 1) The data is grouped in rational subgroups.
- 2) The sampling frequency must maximize the possibility of variation between samples.
- 3) The data should be independent, that is, $x_{ik} = \mu + \varepsilon_{ik}$ ($i = 1, \dots, n$; $k = 1, \dots, m$), where $\varepsilon \sim N(0, \sigma^2)$ is a random variable, called white noise.
- 4) The control limits are distant from the central line (mean of the control variable) ± 3 standard

deviations of this variable, it means the significant level is 0.27%.

- 5) The data has a Normal distribution with mean μ and variance σ^2 ($X \sim N(\mu, \sigma^2)$).

The current productive systems (lean production) have distinct characteristics from those of the Shewhart period (mass production). Nowadays, firms have new constraints, such as: they produce simultaneously a large number of different products in small amounts and need to control, at the same time, several quality characteristics. Consequently, it is necessary to develop other approaches that meet these new constraints. This topic has been the subject of study by several researchers, such as Bothe (1988), Wheeler (1991), Pyzdek (1993), Quesenberry (1997), Montgomery (2012) and, Pereira and Requeijo (2012), among other authors.

This approach is commonly known as the Short Run SPC. When there is a restricted amount of data, the estimation of the process parameters is not possible. Thus, Quesenberry (1997) suggested the application of a new kind of control charts, named Q charts (univariate approach) and MQ charts

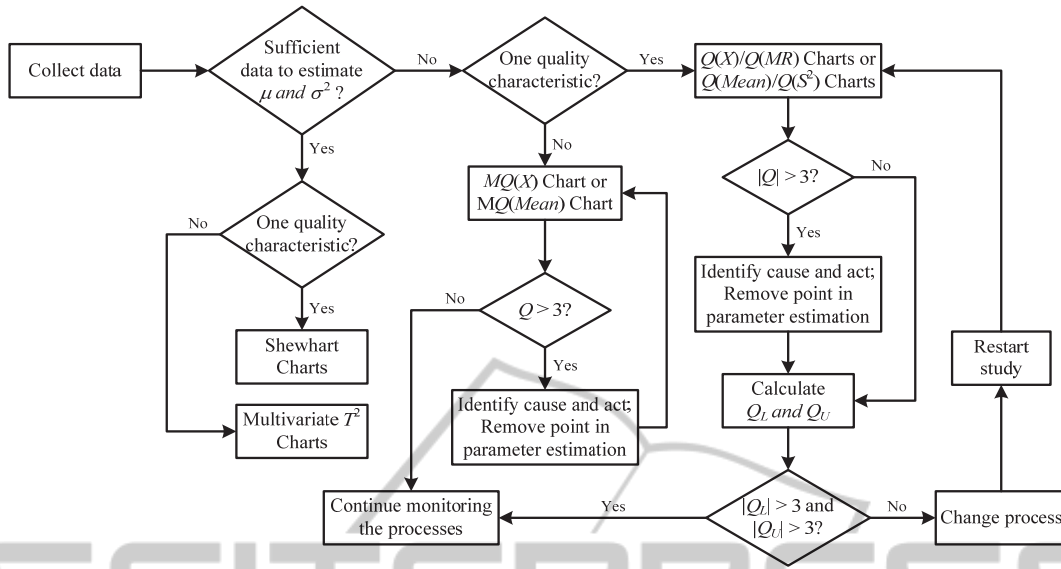


Figure 1: Methodology for implementation of SPC for a limited amount of data.

(multivariate approach). The implementation of these control charts is done in order to include in the same chart all products/quality characteristics, and based on this approach it is possible to control the stability of all processes.

In this paper the Short Run SPC approach is discussed for continuous variables, when a restricted amount of data in two distinct control contexts (univariate and multivariate) exists.

In order to apply correctly the univariate and multivariate SPC, the data must be Normality distributed for all characteristics. To support a continuous evaluation of the process capability, the application of new indices Q_L and Q_U developed by Pereira and Requeijo (2012) is suggested. The implementation of these two indices allows the capability of the process to be analyzed in real time.

In order to demonstrate the importance of these charts and validate the methodology proposed in this paper, a case study is presented and discussed.

2 METHODOLOGY

When there is a limited amount of data, that is, when there is not enough information to properly estimate the processes parameters, the authors of this paper suggest the methodology presented in Figure 1.

3 SHORT RUN SPC TO A LIMITED AMOUNT OF DATA

3.1 Univariate SPC

In some cases, due to the existence of a restricted amount of data, process parameters cannot be properly estimated, creating an obvious difficulty in the construction of traditional Shewhart, Z and W control charts. To answer this question, the adoption of the methodology developed by Charles Quesenberry is suggested, which consists in transforming the quality characteristic X in a variable Q , where X is a continuous variable. The variable Q follows approximately the standard Normal distribution, with zero mean and variance 1, if the variable X is independent and approximately Normal $N(\mu, \sigma^2)$. This method transforms the statistics calculated at time i by estimating the process parameters at time $(i-1)$, considering all data prior to this time.

There are two situations to consider, the data are samples or individual measurements.

3.1.1 Q Control Charts for Individual Data

The statistic Q at time i is calculated from the measurement X at that moment, and it is defined by

$$Q_i(X_i) = \Phi^{-1} \left(G_{i-2} \left(\sqrt{\frac{i-1}{i}} \left(\frac{X_i - \bar{X}_{i-1}}{S_{i-1}} \right) \right) \right) \quad (1)$$

$i = 3, 4, \dots$

The variability process should be calculated based on the moving range of two consecutive observations. Thus, the statistic Q at time i , which is obtained from the statistics at that moment MR , is defined by

$$Q_i(MR_i) = \Phi^{-1} \left(F_{1,\nu} \left(\frac{\nu (MR_i)^2}{(MR_2)^2 + \dots + (MR_{i-2})^2} \right) \right) \quad (2)$$

$i = 4, 6, \dots; \nu = i/2 - 1$

In equations (1) and (2) the following variables denote:

- X_i – individual measurement at time i
- \bar{X}_{i-1} – average of $(i-1)$ observations
- S_{i-1} – standard deviation of the sample of $(i-1)$ observations,
- MR_i – moving range at time i
- $\Phi^{-1}(\bullet)$ – inverse of the standard Normal distribution function
- $G_\nu(\bullet)$ – Student-t distribution function,
- $F_{\nu_1, \nu_2}(\bullet)$ – Fisher distribution function with ν_1 and ν_2 degree of freedom.

3.1.2 Q Charts for Subgrouped Data

The $Q(\bar{X})$ statistic at time i is calculated from the sample average \bar{X} , defined by

$$Q_i(\bar{X}_i) = \Phi^{-1} \left(G_{\nu_i + \dots + \nu_i}(\omega_i) \right), \quad i = 2, 3, \dots$$

$$\omega_i = \sqrt{\frac{n_i(n_1 + \dots + n_{i-1})}{n_1 + \dots + n_i}} \left(\frac{\bar{X}_i - \bar{\bar{X}}_{i-1}}{S_{p,i-1}} \right) \quad (3)$$

$$\nu_i = n_{i-1}; S_{p,i}^2 = \sum_{j=1}^i \nu_j S_j^2 / \sum_{j=1}^i \nu_j$$

Also, the sample variance is transformed in the $Q(S^2)$ statistic, defined by

$$Q_i(S_i^2) = \Phi^{-1} \left(F_{\nu_i, \nu_1 + \dots + \nu_{i-1}}(\theta_i) \right), \quad i = 2, 3, \dots$$

$$\theta_i = S_i^2 / S_{p,i-1}^2 \quad (4)$$

In equations (3) and (4) the following variables denote:

- n_i – sample size at time i
- \bar{X}_i – average of the sample i
- $\bar{\bar{X}}_i$ – grand average of i samples
- $S_{p,i}^2$ – pooled variance of i samples

3.1.3 Control Limits

The Q statistics are random variables with standard

Normal distribution. Thus, the control limits are ± 3 ($\alpha = 0.27\%$) for $Q(X)$, $Q(MR)$, $Q(\bar{X})$ and $Q(S^2)$ charts.

One of the serious problems of the Q charts is its poor sensitivity in the detection of special cause of variation. In order to solve this drawback, it is important to apply rules to increase the sensitivity of the charts. The application of Rule 1 (a point outside the control limits) and Rule 6 (four out of five consecutive points in zone B or A on the same side of the center line) referred to in ISO 7870-2:2013 is suggested.

3.1.4 Process Capability

The application of the Q_L and Q_U capability indices developed by Pereira and Requeijo (2012) is advisable; these indices are defined at time i by

$$(\hat{Q}_L)_i = \frac{LSL - \hat{\mu}_i}{k \hat{\sigma}_i} \quad (5)$$

$$(\hat{Q}_U)_i = \frac{USL - \hat{\mu}_i}{k \hat{\sigma}_i} \quad (6)$$

In equations (5) and (6) USL and LSL are respectively the upper and lower specification limits and k is the minimum value to processes capability.

3.2 Multivariate SPC

Similarly to univariate SPC, when there is a limited amount of data, the best solution to implement the multivariate SPC is to use the Q statistics. The present developments are based on the assumption that the joint distribution of p variables is a multivariate Normal distribution $N_p(\mu, \Sigma)$. The control of the mean vector is made by MQ_X or $MQ_{\bar{X}}$ charts.

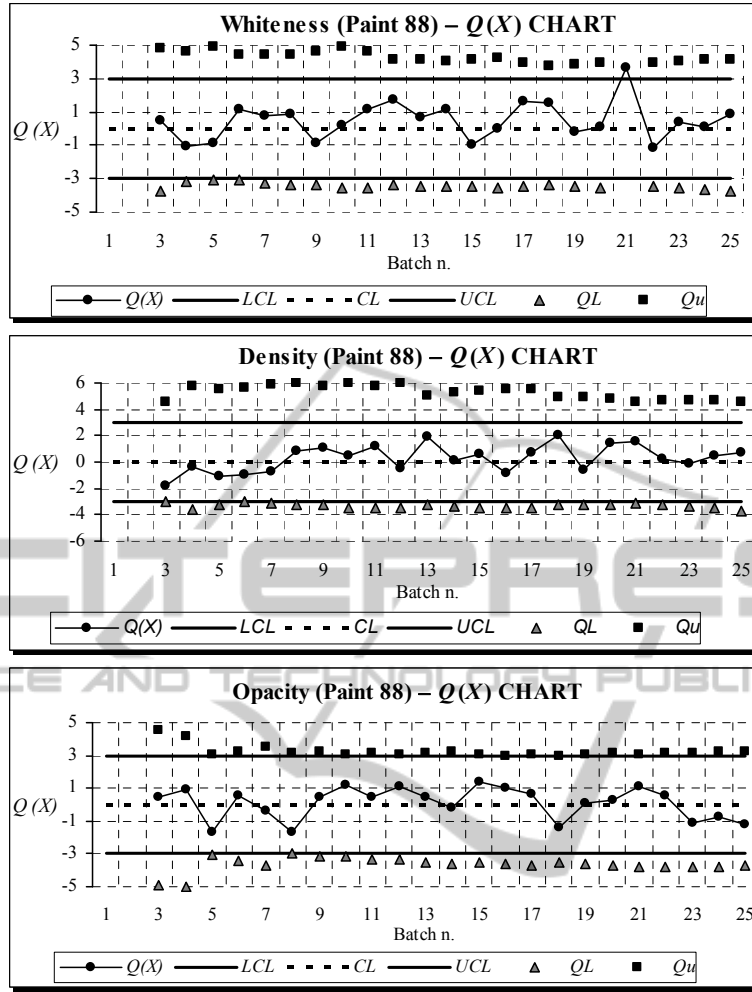
3.2.1 MQ Control Charts for Individual Data

The \mathbf{X} vector is transformed in the MQ statistic. Considering the intention of detecting shifts between \mathbf{X} vector and μ vector and the structure of the statistic, MQ at time i is

$$MQ_i(\mathbf{X}_i) = \max \left(0; \Phi^{-1} \left(F_{p, i-1-p}(A_i) \right) \right)$$

$$A_i = f \cdot (\mathbf{X}_i - \bar{\mathbf{X}}_{i-1})' \mathbf{S}_{i-1}^{-1} (\mathbf{X}_i - \bar{\mathbf{X}}_{i-1}) \quad (7)$$

$$f = \left(\frac{(i-1)(i-1-p)}{ip(i-2)} \right), \quad i = p+2, p+3, \dots$$

Figure 2: Q_X Charts for Whiteness, Density and Opacity of the Paint 88.

where

$$\mathbf{X}_i = (X_{i1}, X_{i2}, \dots, X_{ip})' \quad (8)$$

$$\bar{\mathbf{X}}_i = \frac{1}{i}((i-1)\bar{\mathbf{X}}_{i-1} + \mathbf{X}_i) \quad (9)$$

$$\mathbf{S}_i = \left(\frac{i-2}{i-1}\right)\mathbf{S}_{i-1} + \frac{1}{i}(\mathbf{X}_i - \bar{\mathbf{X}}_{i-1})(\mathbf{X}_i - \bar{\mathbf{X}}_{i-1})' \quad (10)$$

3.2.2 MQ Charts for Subgrouped Data

The MQ statistic at time i is calculated from $\bar{\mathbf{X}}$ vector at that time by

$$\begin{aligned} MQ_i(\bar{\mathbf{X}}_i) &= \max\{0, \Phi^{-1}(F_{p, N_i-i-p+1}(A_i))\} \\ A_r &= g \cdot (\bar{\mathbf{X}}_i - \bar{\bar{\mathbf{X}}}_{i-1})' \mathbf{S}_{pool,i}^{-1} (\bar{\mathbf{X}}_i - \bar{\bar{\mathbf{X}}}_{i-1}) \\ g &= \left(\frac{n_i N_{i-1} (N_i - i + 1 - p)}{N_i p (N_i - i)} \right), i = 2, 3, \dots \end{aligned} \quad (11)$$

where

$$\bar{\mathbf{X}}_i = (\bar{X}_{i1}, \bar{X}_{i2}, \dots, \bar{X}_{ip})' \quad (12)$$

$$\bar{\bar{\mathbf{X}}}_i = \frac{1}{N_i} (N_{i-1} \bar{\bar{\mathbf{X}}}_{i-1} + n_i \bar{\mathbf{X}}_i); \quad N_i = n_1 + \dots + n_i \quad (13)$$

$$\begin{aligned} \mathbf{S}_{pool,i} &= \frac{1}{N_i - 1} ((N_{i-1} - i + 1) \mathbf{S}_{pool,i-1} + (n_i - 1) \mathbf{S}_i) \\ \mathbf{S}_{pool,0} &= \mathbf{0} \end{aligned} \quad (14)$$

3.2.3 Control Limits

As the T^2 charts, the MQ charts only contemplate the upper control limit (UCL). The lower control limit (LCL) is zero and the upper is + 3.

4 CASE STUDY

To illustrate the potential applicability of the Short Run SPC approach for continuous variables, when a restricted amount of data exists, two distinct control charts (univariate and multivariate) will be applied.

The study focuses on production of paint, named "Paint 88." With the purpose of controlling three quality characteristics: whiteness, density and opacity, 25 sample batches of this paint were analyzed.

The univariate approach contemplates the construction of Q_X chart for each one of three characteristics, as shown in Figure 2. In order to calculate the Q statistic, equation (1) was used. However, to calculate indices Q_L and Q_U , equations (5) and (6) were used.

Looking at univariate Q_X chart (figure 2) it is possible to detect a special cause of variation related to whiteness characteristic (see batch 21). However, all characteristics have capability since $Q_L < -3$ and $Q_U > 3$ occurs in all batches.

The multivariate approach contemplates the construction of only one MQ_X chart, as shown in Figure 3, and the multivariate statistics of the 25 sample batches were calculated based on equations from (7) to (10).

Analysing the MQ_X chart for paint 88, it is possible to verify that there are no problematic situations, and therefore, we might conclude that the process is stabilized for the three quality control characteristics.

5 CONCLUSIONS

When there is a restricted amount of data, it is

impossible to implement the traditional SPC, univariate or multivariate, because in these scenarios it is necessary to estimate the process parameters.

The methodology presented in Figure 1, has several advantages over traditional approaches such as:

- 1) Q charts allow the statistical control of all products/quality characteristics in the same document (control chart), even when there is insufficient data to conveniently estimate the process parameters (mean and variance);
- 2) MQ charts allow the simultaneous control of several statistical quality characteristics of various products in the same document, even when there is insufficient data to favorably estimate process parameters (mean vector and covariance matrix);
- 3) it allows the study of different characteristics in the same chart;
- 4) it simplifies the processes analysis;
- 5) based on application of the capability indices Q_L and Q_U in the univariate approach ($Q(X)$ charts or $Q(\bar{X})$) it allows the monitoring of the process capability, revealing an important added value;
- 6) Using Q_L and Q_U indices, permits the study of the process capability in real-time, and reduces the probability of producing nonconforming units.

On the other hand, one disadvantage of Q control charts is the difficulty in analyzing the existence of non-random patterns; consequently, this limitation increases with the number of products/quality characteristics.

Another disadvantage of the Q and MQ control charts is the poor sensitivity in detecting special cause of variation, especially at the beginning of the study. Although the lack of sensitivity in detecting

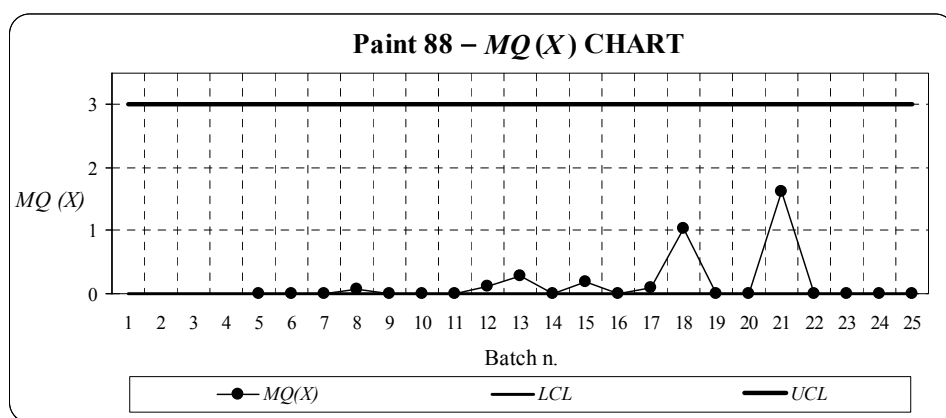


Figure 3: MQ_X Chart for the characteristics Whiteness, Density and Opacity of the Paint 88.

non-random patterns is a limitation of the \bar{Q} and \bar{MQ} control charts, its use is of great importance and utility, since traditional approaches based on univariate and multivariate analysis cannot be implemented when there is a restricted amount of data.

One way to increase the sensitivity of these techniques is to use memory charts, such as Cumulative Sum charts (e.g. $CUSUMQ$ and $MCUSUMQ$) and Exponentially Weighted Moving Average charts (e.g. $EWMAQ$ and $MEWMAQ$), as mentioned in Requeijo and Pereira (2012).

REFERENCES

- Bothe, D. R., 1988. SPC for Short Production Runs, *Quality*, Vol. 27, pp. 58-59, 1988.
- Montgomery, D. C., 2012. *Introduction to Statistical Quality Control*, 7th Edition, John Wiley & Sons, Inc., New York.
- Pereira, Z. L. e Requeijo, J. G., 2012, *Quality: Statistical Process Control and Planning*, 2th Edition, Fundation of FCT/UNL, Lisbon (in portuguese).
- Pyzdek, T., 1993. Process Control for Short and Small Runs, *Quality Progress*, Vol. 26(4), pp. 51-60, 1993.
- Quesenberry, C. P., 1997. *SPC Methods for Quality Improvement*, John Wiley & Sons, Inc., New York.
- Shewhart, W. A., 1931. *Economic Control of Quality of Manufactured Product*, D. Van Nostrand Company, Inc., New York.
- Wheeler, D. J., 1991. *Short Run SPC*, SPC Press, Knoxville, Tennessee.