A Methodology to Interpret Multivariate *T* Squared Control Chart Signals

Teodoro García[1,3], Maura Vásquez[2], Guillermo Ramírez[2], and Carolina Pérez[1]
[1]Industrial Engineering School, University of Carabobo, Valencia, VENEZUELA
[2]Statistics school, Central University of Venezuela, Caracas, VENEZUELA
[3]Corresponding Author: Teodoro García, email: tgarcia@uc.edu.ve

Abstract

Multivariate control charts based on the use of Hotelling *T* squared Statistic are a very useful tool to be applied on Statistical Production Process Control, where a simultaneous monitoring of quality variables of terminated products should occur. The main advantage of these charts is that they are based on an optimum statistical test that allows detecting changes on the process mean vector for multivariate samples of any size. However, once the alarm signal is detected, additional interpreting criteria should be established in order to determine which variables are producing these changes. This paper proposes a methodology based on the *GH* Biplot Factorization of Hawkins' *Z* scale, which results into a multivariate graphic tool that represents the alarm signal, where the identification of the source of the signal detected by the *T* squared Statistic is possible. Therefore, this paper establishes a criterion to select the *GH* Biplot plane where the signal is best represented. It also suggests the sample statistic to evaluate the significance of the signal projection on the studentized partial residuals (*SPR*) of each variable on the selected plane. This proposal has been validated in an environment of simulated scenarios and proven through its application on regular and real problems in Venezuelan industry. Its application has been effective to detect changes in the process mean vector; being those changes the outliers or on one variable. In addition, as the sample size becomes greater it is more effective to detect changes. On the other hand if the change increases it is easier to detect.
Key words: Alarm signal, studentized partial residuals, *GH* Biplot, plane selections.

1. Introduction

The univariate and multivatiate control charts are applied in *Phase I* and *Phase II*. According to Montgomery (2005), *Phase I* is used to analyze the performance of processes, allowing operating personnel to remove the most important causes of variation and see if the process is under statistical stability. If the process has the mentioned stability, the data gathered during the analysis gives reliable control limits to monitor future production. *Phase II* uses the control chart as a control tool and allows the monitoring of the occurrence of alert signals.

Ryan (1989) comments that the occurrence of alert signals when monitoring a process through a T^2 multivariate control chart has two possible origins: First, when the signal keeps the correlation structure among variables and shows a significant variation respect to the historic data. Secondly, if the signal shows shifts in its behavior pattern in one or several of its variables, it is said that there is a rupture in the correlation structure. However, there is a major inconvenient to establish the origin of an alert signal since the so mentioned signal is produced by several variables.

Hawkins (1991, 1993) and Mason, Tracy and Young (1995, 1997, 2002) (*MTY*) have proposed several ways of decomposing the T^2 statistic to be able to identify the variables which give origin to the alert signal.

The classic *GH* Biplot (Column Metric Preserving) proposed by Gabriel (1971) is a tool that allows simultaneous representation of variables and observations from the factorization of a multivariate data matrix. This tool analyzes the structure of relations among variables and visualizes the effects the variables produce on the individuals.

Lastly, this work proposes to show the information originated by the studentized partial residuals matrix of Hawkins' *Z* scale (1991) as a *GH* Biplot graphic in order to have an improved methodology that allows industry operating personnel to interpret alert signals.

2. Materials and Methods

2.1 *GH* Biplot

According to Gower, Lubbe and le Roux (2011) the term "Biplot" is due to Gabriel (1971) who popularized versions in which the variables are represented by directed vectors.
Classic Biplots select the markers through the decomposition in singular values of a data matrix $X = UDV'$, defining then the factorization matrixes as:

$$G = UD^{\gamma} \qquad H = VD^{1-\gamma}$$

Gabriel (1971), proposed to get the Biplots through the selection of γ values. If $\gamma = 1$, a Biplot is constructed where $G = UD \qquad H = V$

On the other hand any $X_{n\times p}$ matrix of range *r* may always be factorized as:

$$X_{n\times p} = G_{n\times r} * H'_{r\times p}$$

The fundamental property derived from the factorization allows the reconstruction of the data of the matrix as:

$$x_{ij} = g'_i * h_j \quad i=1,2,...n \quad \text{y} \quad j=1,2,...,p$$

The orthogonal projection of a row marker g_i, on a column marker h_j is:

$$Proy\left(g_i/h_j\right) = \left(\frac{g'_i h_j}{h'_j h_j}\right) h_j$$

Additionally, the location of this projection on the axis defined by the marker h_j is obtained by measuring its distance to the origin, as:

$$d\left(Proy\left(g_i/h_j\right),\theta\right) = \frac{g'_i h_j}{\left\|h_j\right\|} = \frac{X_{ij}}{\left\|h_j\right\|} \qquad i = 1,2,\ldots,n; \quad j = 1,2,\ldots,p$$

2.2 Hawkins' Z scale

Hawkins' Z scale (1991) may be expressed as:

$$Z = X_c S^{-1}\left(diag\left(S^{-1}\right)\right)^{-1/2}$$

Being:

$$\left(diag\left(S^{-1}\right)\right)^{-1/2} = diag\left(S_{1.(-1)}, S_{2.(-2)}, \ldots, S_{p.(-p)}\right)$$

where X_c is the centered matrix.
Therefore, the generic element z_{ij} of the transformed matrix coincides with the studentized partial residual (*SPR*) of the *i*th observation in the adjustment of the x_j regression in terms of $x_{(-j)}$. The generic element z_{ij} is expressed as:

$$z_{ij} = \frac{(x_{ij} - \tilde{x}_{ij})}{s_{j.(-j)}}$$

2.3 Critical values

Supposing hypothesis $H_o: X_j / X_{(-j)} \sim N\left(\mu_{j.(-j)}, \sigma^2_{j.(-j)}\right)$ then $\mu_j = \mu_{j.(-j)}$, $\sigma^2 = \sigma^2_{j.(-j)}$, therefore the resulting conditional variance $S^2_{j.(-j)}$ could be estimated using $\tilde{\sigma}^2 = \frac{e'e}{n-p-1}$. The confident limits for the $T_{j.(-j)}$ terms of the *MTY* decomposition when the process control is carried out through individual observations is:

$$T_{j.(-j)} = \frac{x_{oj} - \tilde{E}(x_{oj})}{s_{j.(-j)}} \sim \sqrt{\frac{(n-1)(n+1)}{(n-p-1)n}}\, t_{n-p-1}$$

If the sample size *m* is greater than 1, then:

$$T_{j.(-j)} = \frac{\overline{x}_{(.)j} - \tilde{E}(\overline{x}_{(.)j})}{s_{j(-j)}} \sim \sqrt{\frac{(n-1)(m+n)}{m\,n\,(n-p-1)}}\, t_{n-p-1}$$

2.4 Selection of the best representing plane for a signal in the *z* scale.

There are as many Biplot representing planes as combinations of C^p_2, hence a criterion is required to select a plane with the best representation of the signal. If the observations in the *Z* scale are represented on the *GH* Biplot then:

- When considering the signal $z_{i(\alpha,\beta)} = (z'_i v^\alpha, z'_i v^\beta)$ projected on the plane (α,β), the length of the vector representing the signal can be measured and an approximation to the norm of the signal in the original space will be obtained.
- If vector $z_{i(\alpha,\beta)}$ is projected on a vector hj, $Proy(z_{i(\alpha,\beta)} / h_j) = \tilde{z}_{ij(\alpha,\beta)}$ the approximate representation of the studentized partial residual $T_{j.(-j)}$ is obtained. Consequently, the squared norm of the vector projected on h_j should always be less than or equal to the squared norm of $z_{i(\alpha,\beta)}$:

$$\left\|\tilde{z}_{ij(\alpha,\beta)}\right\|^2 < \left\|z_{i(\alpha,\beta)}\right\|^2$$

From this perspective, this paper proposes to analyze the plane with the best approximation to the signal in terms of $T_{j.(-j)}$.

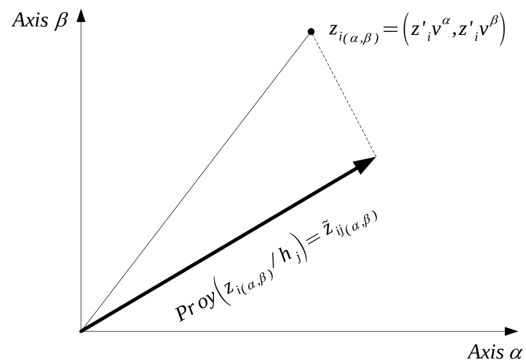

Figure1. Reconstruction of signal through a projection on the h_j vector

In conclusion, the plane with the best representation of the signal should coincide with the one that has the best approximations for the studentized partial residuals of the analyzed observation.

3. Methodology to explain the origin of an alert signal in a T^2 control chart.

The data involved to apply the T^2 control chart must follow a multivariate normal distribution. Additionally this paper requires the mentioned data to be standardized to avoid scale problems.

This process has two matrixes: Matrix *X* which contains the data corresponding to *Phase I* and from where parameters μ and Σ have been estimated, and matrix *XI* which contains the elements of matrix *X* as well as the observation that identifies the signal alert in *Phase II*.

Next step is to standardize matrix *XI* to transform it into matrix *ZI* through the application of Z scale. Hence, the basic structure of matrixes *XI* and *ZI* are going to be affected by the alert signal.

Then the factorization of *GH* - Biplot of matrix *ZI* is constructed and approximations of the vector z_i representing the alert signal on each of the possible planes are calculated. Lastly, the plane with the best approximation to the absolute value of the alert signal is selected.

A procedure to analyze the characteristics of the signal is proposed as follows:

- Projections of the alarm signal on the *SPR* vectors of the variables are calculated and plotted.
- If all projections are below the significant established limit, the point is an outlier.
- If one or several of the projections are above the limit, the signal represents a breaking in the correlation structure.
- If the projection on the vector *SPR* of a variable is significant and follows the direction of growth, the existence of an increase of the mean of the corresponding variable is suggested. If it follows an opposite direction, then a decrease of the mean value is suggested.

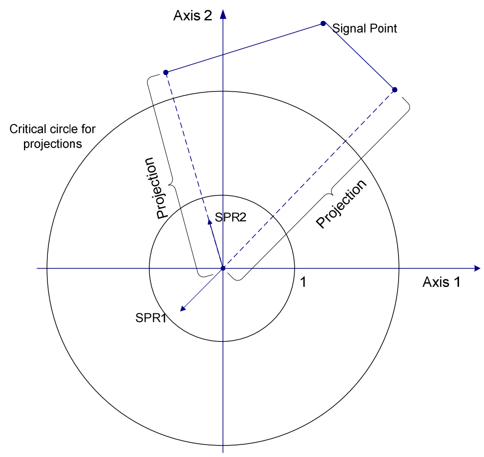

Figure 2. Illustration of an alert signal on a *GH* – Biplot plane.

4. Validation

Each case has 1000 replications of historic matrixes giving origin to 100 alert signals under combinations of the following conditions:

Historic matrix: 50 samples of 5 variables

Sample size: 1, 5 and 10

Correlation between variables: 0.95, 0.85, 0.65 and 0.45

Shifting magnitudes: 3σ, 2σ and 1σ

Types of shifting: 1variable, 2 variables and all variables.

The obtained results were compared with Aparisi et al. (2006) and Mason, Tracy and Young (1995).

On the other hand, this methodology has been highly effective when applied to real problems in Venezuelan industries.

5. Example

This paper reproduces an example carried out by Hawkins (1991) and Mason et al. (1995,1997). This example has 35 observations as historical data and then introduces changes in two variables: variable X_1 increases its standard deviation in 50% and in variable X_5 the estimated mean increases a 25% of its standard deviation. After these alterations 15 additional data are simulated and the 48th datum gives origin to an alert signal in the T squared chart.

The proposed methodology gave plane 2-3 as selected and the resulted *GH* Biplot with a critical value of 1.8585 was plotted.

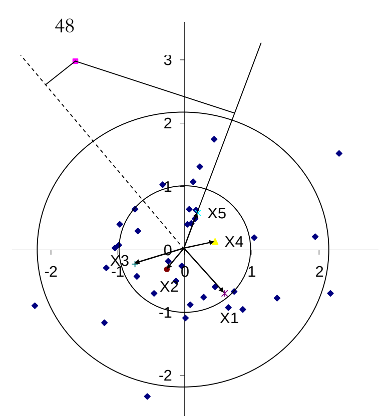

Figure 3. *GH* Biplot representation, plane 2-3, Obs. 48

6. Conclusions

The results of this investigation have yielded into a proposal to interpret alert signals coming from multivariate T^2 control charts, establishing a methodology effective for detecting signals characterized by shifts in only one variable and shifts in all the variables according to the structure of an outlier. These detections are easier as the shifts of the signals are greater. The bigger the sample also increases the probability of detection of the mentioned shifts.

On the other hand:

i. Deviations of $\pm 1\sigma$ are not easily detected. Furthermore, the smaller the shift the more difficult its detection is.
ii. The bigger the sample size the easier to detect and locate the signal is.
iii. The nature of the Biplot representation makes the method work for discrete or continuous variables apart from making it independent of the distribution of probabilities. When particularly working on T^2 charts, limits of criticism for projection vectors under an assumed multivariate normality were needed.
iv. The criterion to select the GH Biplot representation plane that leads to the best signal representation turned out to be highly effective when applied.

References

Aparisi, F. Avendaño, G. y Sanz, J. (2006), "Interpreting T^2 Control Chart Signals: Effectiveness of MTY Decomposition vs. A Neural Network". *IIE T ransactions*, issue 38(8), August.

Flury, B. Riedwyl, H. (1988), "*Multivariate Statistics: A Practical Approach*". p. 151, Chapman & Hall, London.

Gabriel, K.R. (1971), "The Biplot Graphic Display of Matrices with Application to Principal Components Analysis". *Biometrika*, 58, pp. 453.467.

Gower, J. Lubbe, S. le Roux, N. (2011), *"Understanding Biplots"*. John Wiley & Sons. Ltd.

Hawkins, D. (1991), "Multivariate Quality Control based on Regression-Adjusted Variables". *Technometrics*. 33, 61-75

Hawkins, D. (1993), "Regression adjustment for variables in multivariate quality control". *Journal of quality technology*, 25, 170-182.

Mason, R. Tracy, N. Young, J. (1995), "Decomposition of T^2 for Multivariate Control Chart Interpretation". *Journal of Quality Technology* 27, N°2 April. 99-108.

Mason, R. Tracy, N. Young, J. (1997), "A Practical Approach for Interpreting Multivariate T^2 Control Chart Signals". *Journal of Quality Technology*. Vol. 29, N°4 Oct. 396-406.

Mason, R. Young, J. (2002), "*Multivariate Statistical Process Control with Industrial Applications*". ASA-SIAM Series on statistics and Applied Probability.

Montgomery, D.(2001). "*Introduction to Statistical Quality Control*", Wiley, New York.

Ryan, T. (1989), "*Statistical Methods for Quality Improvement*". John Wiley. New York.