

Principal component analysis applied to the statistical control of multivariate processes

Análisis de componentes principales aplicados al control estadístico de procesos multivariados

Juan Carlos Herrera Vega



Bruno de Jesús Rahmer



German Herrera Vidal



Fundación Universitaria Tecnológico Comfenalco, Colombia

OPEN  ACCESS

Recibido: 24/08/2021

Aceptado: 22/11/2021

Publicado: 08/02/2022

Correspondencia de autores:

jherrera@tecnocomfenalco.edu.co



Copyright 2020
by Investigación e
Innovación en Ingenierías

Abstract

Objective: Propose the analysis and monitoring of a chemical process sustained in the theoretical principles of a factorial method cataloged as principal component analysis (PCA), whose ultimate objective is to represent the original variables of the system. **Methodology:** Methodologically the data was analyzed in a more compact dimensional space, under the hypothesis of multivariate normality. A control chart based on the error prediction squares is constructed in the later phase to monitor the behavior of the successor variables once the ACP has been applied. **Results:** The results of the univariate control charts constructed from the t-scores of the individual components, in which the behavior of the variables contributing to the generation of alarm signals is exhibited, can be used as the basis for an auxiliary analysis based on the results provided by the univariate control charts. **Conclusion:** The results obtained corroborate that the three retained components explain a majority fraction of the original variability of the data cloud and that the control chart constructed from the sub-dimensions registers the existence of outliers or extreme values.

Keywords: Control chart, Statistical process control, Multivariate analysis, Analysis in main components (ACP).

Resumen

Objetivo: Proponer el análisis y monitoreo de un proceso químico sostenido en los principios teóricos de un método factorial catalogado como análisis de componentes principales (PCA), cuyo objetivo final es representar las variables originales del sistema. **Metodología:** Metodológicamente, los datos se analizaron en un espacio dimensional más compacto, bajo la hipótesis de la normalidad multivariada. una gráfica de control basada en los cuadrados de predicción de error se construye en la fase posterior para monitorear el comportamiento de las variables sucesoras una vez que se ha aplicado el ACP. **Resultados:** Los resultados de los gráficos de control univariados construidos a partir de las puntuaciones t de los componentes individuales, en los que se exhibe el comportamiento de las variables que contribuyen a la generación de señales de alarma, se pueden utilizar como base para el análisis auxiliar basado en los resultados proporcionados por los gráficos de control univariados. **Conclusiones:** Los resultados obtenidos corroboran que los tres componentes retenidos explican una fracción mayoritaria de la variabilidad original de la nube de datos y que la gráfica de control construida a partir de las subdimensiones registra la existencia de valores atípicos o extremos.

Palabras clave: Cuadro de control, control estadístico de procesos, Análisis multivariado, Análisis en componentes principales (ACP).

introduction

The superlative capacity of multivariate analysis for the detection of anomalies in processes where comovements between the analyzed variables occur and that also incorporate huge amounts of information [1], has been patently recognized in multiple domains, including industrial monitoring [2], since the techniques circumscribed within them are theoretically delimited in order to outline monitoring frameworks that aim to stabilize the process, reduce the variability that is natural to it and detect the appearance of background events that negatively affect the performance of production systems. Although, it is undeniable that such techniques do not enjoy a generalized popularity, it is necessary to indicate that one of the inherent advantages to them is that they make possible the reduction of effort quotas to evaluate the state of multiple variables, without assuming the loss of a high fraction of information, and therefore, they avoid the immovable restrictions imposed on the methods of control of processes with high dimensionality. In this sense, multivariate methods for process control are generally concerned with measuring the directionality of observations in a multivariate space, contrary to univariate methods that only monitor the magnitude and variation of individual variables without scrutinizing simultaneous movements between quality characteristics.

The need to apply multivariate statistical tools to simultaneously control two or more variables in the processes becomes more important every day. [3] A very useful multivariate analysis method to explain the sources of variability of a process and reduce the dimensionality of the data, is the Principal Component Analysis (PCA). [4]

ACP is one of the oldest and most popular multivariate statistical techniques in data analysis. Its main objectives are: extract the most important information from a multivariate data set, compress a multivariate data set keeping only the information that is considered important (reduce the dimensionality of the data), simplify the description of a data set and analyze the structure of observations and variables. [5].

The central idea of the ACP is to achieve the simplification of a set of quantitative data, derived from a set of interrelated variables. This objective is achieved by obtaining, from linear combinations of the variables originally measured, a new set of the same number of variables, called principal components (CP) in which, the variability present in the original data remains and that, when ordering them decreasingly by their variance, allow us to explain the phenomenon of studies with the first PC. [6]

In summary, the PCA is made up of: 1) the W_j vectors that are given by the dominant eigenvectors (those with the largest associated eigenvalue); 2) the sample covariance matrix. [7]

The continuous review of orthodox techniques used to obtain a credible overview of the overall capacity and stability of the processes has served as a bridge for the inclusion of multivariate statistical techniques such as ACP, catalogued as a method of dimensionality reduction when there is an extensive amount of quantitative variables, in order to obtain a transformed system of coordinates called main components, linear combination of the primitives [8] and which are shown to be capable of retaining the global variation exhibited by the original data [9].

These types of methods and their extensions not only allow the projection of data in a more compact dimensional sub-space, but also open the way to the implementation of alternative methodologies for the design of control charts-such as those built for the analysis of latent variables, since the direct uses of conventional multivariate tools for operations control are not solvent for addressing non-abnormal circumstances such as the existence of collinear variables -circumstance that generates that the covariance

matrix becomes an almost singular matrix and difficultly invertible- besides being potentially conducive to the increase of false alarm rates. Therefore, the use of control graphs that integrate information synthesis methodologies such as the one shown in this paper boast flagship properties such as those previously highlighted, and could well be used in a mandatory way in manufacturing environments for the systematic characterization and evaluation of processes in which a multiplicity of highly correlated quality characteristics underlie, as long as the quantity and quality of the available data is at least acceptable.

Methodology

The present investigation is segmented into three sections which will be explained in detail below. As a starting point, we proceed to identify and characterize the behavior of a conglomerate of variables of the productive process submitted to monitoring is the definition of certain measurable properties that favor the elucidation of a general nuance on the general stability of the process and the correlations between them. The set of data obtained via randomized sample selection is carried out to carry out a brief exploratory analysis of them [10].

Subsequently, the analysis of major components is carried out using the originating data. In general terms, the main component analysis involves a set of data with observations on numerical variables for each of the entities or individuals. These values define vectors p n- dimensional $x_1, x_2 \dots x_n$ and searches for a linear combination given by: $\sum_{j=1}^p a_j x_j$ [11], whose coefficients are obtained from the vectors of the covariance matrix of the original data. For the sequential extraction of the main components, the algorithmic procedure catalogued as partially iterative non-linear least squares is available. And the determination of the optimal number of proxies is done through the application of Krzanowski's cross validation method. It is further emphasized that within this analysis the extraction of the major components of typed variables is not disregarded as the metrics of each differ significantly.

Once the analysis in main components is concluded, a multivariate control chart is constructed, this being the neuralgic phase of the present paper, since it detects the occurrence of anomalous events and mobilizations of the process outside the hyper-plane defined by the reference model. Under this same intention, an auxiliary analysis based on the construction of monitoring graphs for each extracted component is drafted. In both cases, the specification limits are computed from the historical information of the process.

Results

For the acquisition of the data of each one of the selected variables, a simple random sampling plan has been applied without replacement, a probabilistic sample selection procedure in which all the subjects have the same probability of choosing and without the opportunity for another possible selection. Initially, there is a gross volume of data in which it is difficult to identify mathematical relationships, so it is necessary to carry out a timely treatment of the information collected so that its use is permissible in last statistical procedures.

Table 1 lists 9 variables corresponding to an expanded polystyrene manufacturing process: Pressure (P), Water vapour temperature (T), Resistance to water vapour diffusion (R), Density (D), Pellet diameter (Diam), Pentane content (PE), Polymer concentration (PO), Oxygen concentration (O).

Table 1. Descriptive statistics and scaling of variables

| VARIABLE | Mean | Standard deviation | Scale factor | Lower limit | Upper limit |
|----------------------------|--------|--------------------|--------------|---------------------|----------------------|
| Pressure (P) | 19,510 | 9,659 | 9,659 | 20 Kpa | 22 Kpa |
| Temperature (T) | 81,899 | 11,030 | 11,030 | 80°C | 100 °C |
| Diffusion resistance (R) | 22,366 | 5,620 | 5,620 | 20 | 40 |
| Density Nominal (D) | 9,086 | 5,332 | 5,332 | 9 Kg/m ³ | 11 Kg/m ³ |
| Pellet diameter (Diam) | 4,967 | 1,305 | 1,305 | 0,2 mm | 1 mm |
| Pentane content (PE) | 5,668 | 1,037 | 1,037 | 5% | 7% |
| Polymer concentration (PO) | 0,233 | 0,050 | 0,050 | 0,03 | 0,07 |
| Oxygen concentration (O) | 0,967 | 0,328 | 0,328 | 92% | 98% |

Source: Own elaboration

Since the metrics of the observed variables differ significantly, this would suppose, a priori, an imminent problem for the purposes of estimating the main components, since they are not comparable to each other and those whose variance is high will irremediably dominate the first sub-dimensions. In order to avoid this undesirable circumstance, the original variables are standardized, so that each one of them will be equipped at the moment of starting the analysis. Note that in addition to certain descriptive statistics and the corresponding scale factors, the specification limits of the characteristics of interest are expressed.

Extraction of Main Components

Intuitively, the essential purpose in the first section is to find a new set of orthogonal directions that define the maximum variability in terms of the variance-covariance structure of the original variables, in such a way that the information contained in the complete set of components found is the exact equivalent of the original data information, since it is conjectured that there are redundant elements that only add dimensionality to the problem studied. [12]. Through the ACP, the original data is projected in a much more compact and parsimonious dimensional representation [13], which enables the analysis and monitoring of multiple variables under a slightly simplified approach.

Then be a multivariate data matrix $X_{n,p} \in \mathbb{R}^{n \times p}$, with $X_{n,p} = (x_1, \dots, x_n)'$

whose columns contain the variables and the rows, the elements. The mean is given by (1):

$$\bar{x} = \left(\frac{1}{n}\right) X'_{n,p} \quad (1)$$

and the covariance matrix by (2):

$$S = \left[\frac{1}{(n-1)}\right] (X_{n,p} - 1_n \bar{x}')' (X_{n,p} - 1_n \bar{x}') \quad (2)$$

The vectors p-dimensional x are transformed into a vector of "scores" denoted by (3):

$$y = P' (x - \bar{x}) \quad (3)$$

being P the load matrix.

These scores can be found by decomposing the characteristic values in the sample covariance matrix $S = P\Lambda P'$. The own values of S are contained, in descending order, in $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ [14,15].

It is also postulated that the sum of the variances of the variables or total inertia of the point cloud is equivalent to the sum of the variances of the main components. In this sense, the percentage of inertia explained by an i-th component is (4):

$$\lambda_i * 1 / \sum_{i=1}^p \lambda_i = \lambda_i / \sum_{i=1}^p V(x_i) \quad (4)$$

Being $\sum_{i=1}^p V(x_i) = \text{traza}(v)$, an expression of the measure of variability associated with the original variables [16].

Component extraction and cross validation

For the sequential extraction of the main components, the algorithmic procedure has been used *partially iterative non-linear least squares*. In each iteration there is a linear adjustment of the columns of X on a vector of scores to obtain a vector of loads p followed by a linear regression. Instantly, the rows of X are returned on the load vector for re-estimation until the predefined convergence criterion is met. This algorithmic procedure can be synthesized in the following steps [17]:

- Select a column from the data matrix and make it equal to a vector t_h
- The vector t_h is used to predict the matrix X with the regression model (5):

$$X = t_h b_h^T + U \quad (5)$$

$$\text{Where,} \quad \hat{b}_h^T = \frac{t_h^T X}{t_h^T t_h} \quad (6)$$

the minimum quadratic estimator of b_h^T and constitutes the projection of the columns of X about the address of t_h , defined in the space of the N observations.

- The vector is defined with the equation (7):

$$p_h = \hat{b}_h \quad (7)$$

- The vector p_h is normalised, so that its length is made equal to the unit.
- The vector p_h is used to predict the matrix X^T from the model seen in (8):

$$X^T = p_h b_h^T + F, \quad (8)$$

where the minimum quadratic estimator b_h^T is given by (9):

$$\hat{b}_h^T = \frac{t_h^T X^T}{p_h^T p_h}. \quad (9)$$

Projection of the rows of the matrix X on the direction of the vector p_h defined over the space of the K variables is obtained in this way.

- The vector is defined $t_h = \hat{b}_h$
- The quadratic norm is calculated from the difference between the vectors obtained in the antecedent step and in the initial step.
- The quadratic norm is compared with the tolerance value set in 10^{-6} . If the difference is less than this level, the i th component has been obtained. In the opposite case, return to step 2.

The procedure described above is repeatedly applied to obtain n -components, however, to retain an appropriate amount of sub-dimensions and avoid the phenomenon of over-adjustment of the PCA model is applied a cross validation method such as Krzanowski, alternative scheme based on the algorithm of decomposition of singular values, whose purpose is to determine an optimal number of components and facilitates the identification of redundant information in the data matrix [18]. For such purpose, a test of statistical significance is applied to each extracted component in order to determine which of them merit being incorporated into the model. The proposed method assumes that it is intended to predict the elements x_{ij} of the matrix X through model (10):

$$x_{ij} = \sum_{t=i}^m v_{it} \gamma_t u_{tj} + e_{ij}. \quad (10)$$

In this sense, the value $\hat{x}_{ij}^{(m)}$ de x_{ij} ($i = 1, \dots, g; j = 1 \dots e$) for the selection of a quantity of components m . The sum of the squares of the difference between current values and estimates is quantified through the expression (11):

$$PRESS(m) = \frac{1}{np} \sum_i^n \sum_j^p [\hat{x}_{ij}^{(m)} - x_{ij}]^2 \quad (11)$$

How to avoid predictive biases, the x_{ij} should not be used to calculate $\hat{x}_{ij}^{(m)}$ for each $i \in J$. Krzanowski's method starts from the assumption that the decomposition of singular values of the matrix X can be expressed as $X = UDV^T$. You subsequently define the statistician (12):

$$W_m = \frac{PRESS(m-1)PRESS(m)}{D_m} + \frac{PRESS(m)}{D_r} \quad (12)$$

Where D_m the degrees of freedom required to adjust the most-third components and D_r the degrees of freedom remaining once the umpteenth components have been adjusted W_m represents the increase in predictive information in each remaining component. The standard cross validation procedure consists of subdividing X in several groups, remove each group from the data, evaluate the predictor parameters from the remaining data and predict the deleted values [19].

Following an orthodox and subtly restrictive criterion, the components with the greatest statistical significance should have the following values W_m to the unit. Unquestionably, it would be inappropriate to cease adding major components as soon as possible W_m , on the first occasion, is located below the unit, because it is a function of m decreasing non-monotonic [20]. This suggests that there are no clear canons and formal criteria to determine what proportion of variability should be explained by successor variables [21]. However, for monitoring purposes, this empirical pattern will be followed.

Table 2 compiles a summary of the results of the main component analysis. El parameter R^2X quantifies the goodness-of-fit of the model by accounting for the proportion of variability explained by the model. Note that the three main components extracted are capable of explaining up to approximately 93% of the total variation.

On the other side, Q^2X , which is an expression similar to R^2X -except that it evidences a less inflationary behavior as the complexity of the model increases-, it allows evaluating the predictive capacity of the model, indicating that such is acceptable, since it assumes a value close to the unit.

The vector of eigenvalues, on the other hand, can be conceptualized as the magnitude of the variance of the observations along the direction of its corresponding autovector. It is a reliable fact that the first factor explains a majority fraction of the total dispersion of the cloud and the successive ones explain minuscule portions of it, called residual variance.

Table 2. Summary of the extraction of the main components

| COMPONENT | R^2X | R^2X ACUM | EIGEN VALOR | Q^2X | Q^2X ACUM | LIMIT |
|-----------|--------|-------------|-------------|--------|-------------|--------|
| 1 | 0,7085 | 0,7085 | 5,6633 | 0,6051 | 0,6051 | 0,1347 |
| 2 | 0,1306 | 0,8391 | 1,0435 | 0,2059 | 0,6864 | 0,1525 |
| 3 | 0,0849 | 0,9240 | 0,6820 | 0,3362 | 0,7919 | 0,1763 |

Source: Own elaboration

Next, Table 3 lists the saturations or factorial loads that determine the orientation of the new axes of the main components with respect to the initial coordinate system. From a strictly analytical point of view, this matrix contains the saturations, which are calculated from the product between the weightings of the primitive variables in the extracted components and the square root of the auto-values, that is, the following equation (13):

$$r_{jh} = r_{ih} \sqrt{\lambda_h} \quad (13)$$

Positive values close to 1 indicate a strong correlation between a component and any variable as happens with the variables Pressure and Pentane Content in the first factor; on the contrary, negative saturations close to the unit reveal an accentuated degree of negative association between both.

Table 3. Load Matrix

| | Component 1 | Component 2 | Component 3 |
|-----------------------------|-------------|-------------|-------------|
| Pressure (P) | 0,803 | 0,587 | 0,066 |
| Temperature (T) | 0,726 | 0,743 | 0,043 |
| Resistance to diffusion (R) | -0,895 | 0,186 | 0,847 |
| Nominal Density (D) | -0,846 | 0,154 | 0,464 |
| Diameter of pellets (Diam) | 0,876 | -0,090 | 0,825 |
| Pentane content (PE) | 0,883 | -0,211 | 0,321 |
| Polymer Concentration (PO) | -0,783 | 0,709 | -0,303 |
| Oxygen concentration (O) | 0,877 | -0,087 | -0,200 |

Source: Own elaboration

Design of the control chart

Multivariate control diagrams such as the one made in the present case are based on the transformation of a vector \mathbb{R}^P to an escalar through a quadratic function that detects the existence of outliers or extreme directions. When the presence of a serial correlation structure in the observations has been discarded and the process verifies the condition of seasonality, control charts of this nature can be implemented for adjustment effects in the monitoring of processes exhibiting high dimensionality. Unlike the control diagrams of T_{PCA}^2 these are apt to discover a change in the mean of the process if the displacement is orthogonal to the first own vectors of the covariance matrix S [22].

The generation of the control chart from the main components model is based on the squares of the prediction errors (SPE) given by the expression (14):

$$SPE_y = \sum_{k=1}^i (y_{nueva,i} - \hat{y}_{nueva,i})^2 \quad (14)$$

The statistician Q represents the quadratic orthogonal distance of a new multivariate observation from the multivariate subspace of the main components. In other words, it is a measure that quantifies the lack of adjustment of the new sample with respect to the model that includes the retained components, by

detecting the projected data that are not represented by the model. For a sample x_i the residuals denoted by r_i are given by (15):

$$r_i = x_i - \hat{x}_i = x_i(I - P_i P_i^T). \quad (15)$$

While the magnitude of the waste is equivalent to [23] to the expression given by (16):

$$Q = \|r_i\| = r_i r_i^T = x_i(I - P_i P_i^T)x_i^T. \quad (16)$$

Where, r_i the i -th row of the residual matrix representing the prediction errors of the model, $P_i = [p_1, \dots, p_n]$ e I the identity or unit matrix.

The approximate control limits for Q can be determined as long as the data can be described by a normal multivariate distribution and the auto-values of the covariance matrix are known. So, the threshold Q_α responds to the following expression (17):

$$Q_\alpha = \theta_1 \left(\frac{z_\alpha \sqrt{2\theta_2 h_0}}{\theta_1} + 1 + \frac{\theta_2 h_0 (1 - h_0)}{\theta_1^2} \right)^{\frac{1}{h_0}} \quad (17)$$

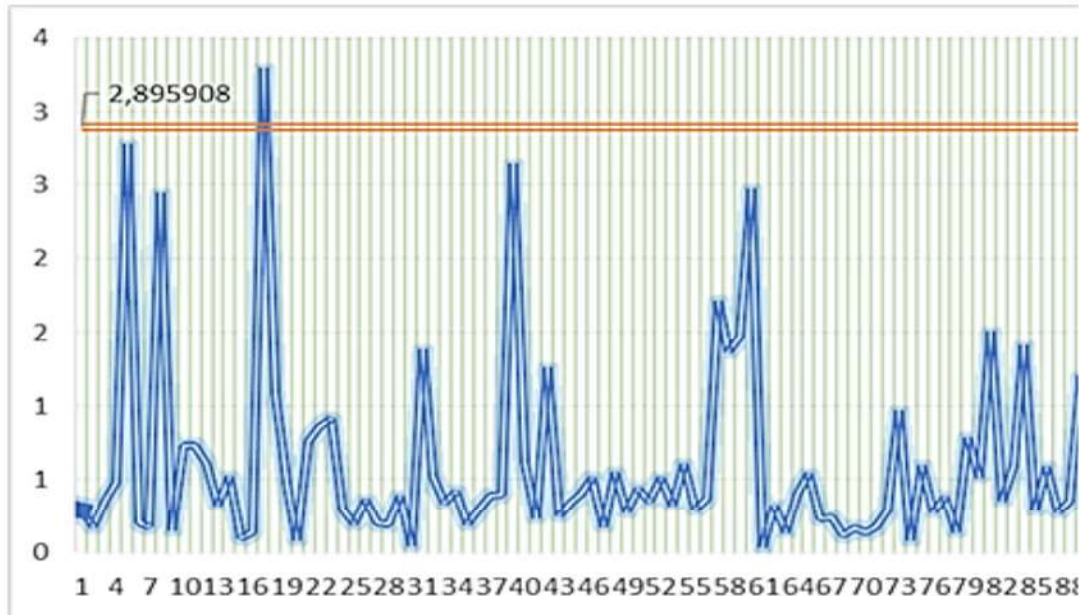
Where Q_α the upper confidence limit for Q with a level of significance α y z_α the statistic of the standardized normal distribution corresponding to the upper percentile $(1 - \alpha)$.

Meanwhile,
$$h_0 = 1 - \frac{2\theta_1 \theta_3}{3\theta_2^2} \quad (18)$$

y,
$$\theta_1 = \sum_{j=k+1}^p \lambda_j^i \quad (19)$$

and λ_j is the proper value of the covariance matrix. Errors between target and actual values can be measured to determine their statistical significance and to identify current operating conditions. In such a case that Q exceeds the threshold, that is, $Q > Q_\alpha$ then there are indications of the presence of events that affect the covariance structure of X , as shown in Figure 1, and therefore it is argued that there are anomalous fluctuations in the process that are not explained by the model.

Figure 1. SPE Control Chart



Source: Own elaboration

The presence of samples outside the limits of the control diagram constructed from component scores or the existence of systematic trends and behaviors in the scores graph shows with factual certainty that the process is "out of control", i.e., that a break has occurred in the correlation structure of the estimated model [23].

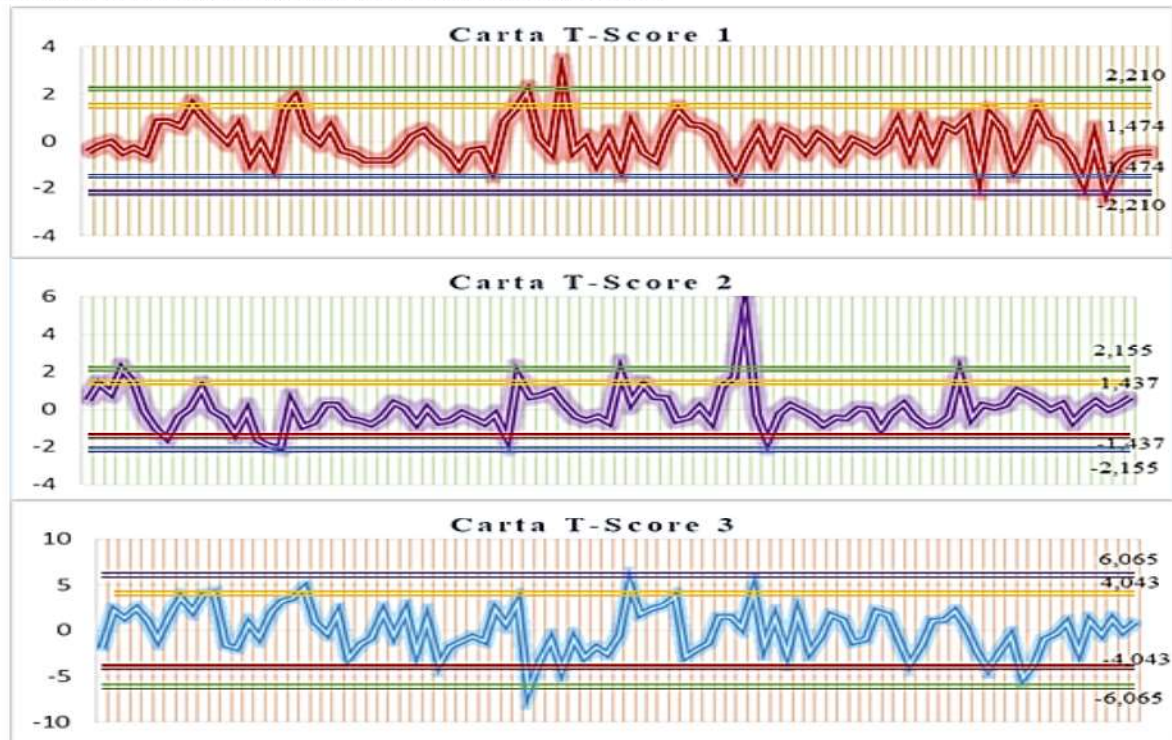
Meanwhile, the results of the univariate control charts constructed from the t-scores of the individual components, in which the behavior of the variables contributing to the generation of alarm signals is exhibited, can be used as the basis for an auxiliary analysis based on the results provided by the univariate control charts. This is shown in Figure 2.

Under the assumption of normality, the control limits for a new \bar{t} -score in a time interval and at a level of significance are given by the expression (20):

$$\pm t_{n-1, \frac{\alpha}{2}} S_{ref} \left(1 + \frac{1}{n}\right)^{1/2} \quad (20)$$

Where S_{ref} the estimated standard deviation of the sample of \bar{t} -scores in an interval of time denoted by k . The expression $t_{n-1, \frac{\alpha}{2}}$ is the critical value of the variable studied with $n - 1$ degrees of freedom and level of significance $\frac{\alpha}{2}$ [24].

Figure 2. Combined scoring chart for individual components.



Source: Own elaboration

It is possible to detect in Figure 2 that out of control signals emerge in the monitoring charts associated with components 1 and 2 and 3 and, consequently, it is stated that the process is contaminated by the presence of outliers, causing the temporary breakdown of the stability of the process. Despite the fact that after a time interval the process returns to a state in which only common causes of variability operate, the existence of anomalous behaviors continues to be persistent, which allows us to conclude that the comovements of the variables contained in the sub-dimensions found transgress the natural limits of variation for all the cases evaluated.

Conclusions

Multivariate control charts, extensions of their univariate analogues, can become canonical tools of regular use in the field of statistical process control to identify situations in which manufacturing systems with multiple variables - deviate from their typical behavior.

In the present case study, a monitoring framework was proposed for industrial processes with multiple quality characteristics, endorsing the fact that the control charts built from the analysis of main components meet the objective concerning the exercise of control over multidimensional production systems under a simple pragmatism clause. Avoiding theoretical lucubrations, it is possible to affirm that the finding of these components induces a substantial reduction of the dimensionality, as well as, in the creation of structures of interdependence that involve the observed variables.

The results obtained corroborate that the three retained components explain a majority fraction of the original variability of the data cloud and that the control chart constructed from the sub-dimensions registers the existence of outliers or extreme values. The premise that the process is not in a state of statistical control is then derived, a judgement revalidated by meticulously inspecting the behaviour shown by the control charts associated with the individual components. For this reason, it is necessary to direct actions aimed at improving the consistency of the productive process so that the quality characteristics exteriorize a relatively homogeneous behavior and without exceeding the limits of natural variation. In a case such as this, the additional application of a univariate monitoring tool would be beneficial, since one limitation of the proposed methodology is its inability to identify which variables substantially contribute to the generation of out of control signals in the short term.

References

1. Y. A. Siddiqui, A. W. A. Saif, L. Cheded, M. Elshafei, A. Rahim, Integration of multivariate statistical process control and engineering process control: a novel framework. *International Journal of Advanced Manufacturing Technology*. 78, 259–268 (2015).
2. J. Camacho, A. Pérez-Villegas, P. García-Teodoro, G. Maciá-Fernández, PCA-based multivariate statistical network monitoring for anomaly detection. *Computers and Security*. 59, 118–137 (2016)
3. Control estadístico multivariante de proceso aplicado en la industria. *Cultura Científica y Tecnológica*. 0 (2018).
4. O. Gozá, M. Fernández, R. Rodríguez, E. Ojito, VacciMonitor, in press (available at <https://www.redalyc.org/jatsRepo/2034/203462663003/html/index.html>).
5. J. M. Polanco Martínez, El papel del análisis por componentes principales en la evaluación de redes de control de la calidad del aire. *Comunicaciones en Estadística*. 9, 271 (2016).
6. J. Rodríguez, E. Giménez, Aplicación del análisis de componentes principales en la investigación de aguas de pozo para el consumo humano. *Población y Desarrollo*. 23, 38–52 (2017).
7. E. D. L. H. Domínguez, T. F. Herrera y A. Mendoza, Métodos multivariantes aplicados a la efectividad de los disolventes en la industria farmacéutica, INGE CUC, 2018. <https://doi.org/10.17981/ingecuc.14.1.2018.12>

8. D. Marcondes Filho, A. M. O. Sant'Anna, Principal component regression-based control charts for monitoring count data. *International Journal of Advanced Manufacturing Technology*. 85, 1565–1574 (2016).
9. J. Owen, I. Demirkiran, Principal Component Analysis : Data Reduction and Simplification. *McNair Scholars Research Journal*. 1, 1–23 (2014).
10. J. Herrera Vega, G. Herrera Vidal, B. De Jesús Rahmer, in *Proceedings of the International Conference on Industrial Engineering and Operations Management (IEOM Society, 2017)*, vol. 2017, pp. 891–904.
11. I. T. Jolliffe, J. Cadima, Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. 374 (2016), , Doi:10.1098/rsta.2015.0202.
12. D. Montgomery. *Statistical Quality Control*. (2012).
13. E. Vanhatalo, M. Kulahci, B. Bergquist, On the structure of dynamic principal component analysis used in statistical process monitoring. *Chemometrics and Intelligent Laboratory Systems*. 167, 1–11 (2017).
14. B. De Ketelaere, M. Hubert, E. Schmitt, Overview of PCA-based statistical process-monitoring methods for time-dependent, high-dimensional data. *Journal of Quality Technology*. 47 (2015), pp. 318–335.
15. K. A. Severson, M. C. Molaro, R. D. Braatz, Principal component analysis of process datasets with missing values. *Processes*. 5 (2017), Doi:10.3390/pr5030038.
16. C. Pérez, Técnicas de análisis multivariante de datos (2004; <http://bit.ly/1JzSD8y>).
17. M. B. Quaglino, M. S. Vitelleschi, Comparación de Métodos para el Tratamiento de Información Faltante en un Análisis de Componentes Principales sobre Datos Biológicos. *FABICIB*. 13, 115–124 (2009).
18. J. Camacho, A. Ferrer, Cross-validation in PCA models with the element-wise k-fold (ekf) algorithm: Practical aspects. *Chemometrics and Intelligent Laboratory Systems*. 131, 37–50 (2014)
19. C. T. Dos Santos Dias, W. J. Krzanowski, Choosing components in the additive main effect and multiplicative interaction (AMMI) models. *Scientia Agrícola*. 63, 169–175 (2006).
20. Bógalo Román J. Componentes subyacentes comunes en series temporales. 2012. Obtenido de <http://espacio.uned.es/fez/eserv/bibliuned:masterMatavanz-jvbogalo/Documento.pdf>.
21. E. Lazzarotto, L. M. Gramani, A. C. Neto, L. A. Teixeira Junior, PRINCIPAL COMPONENTS IN MULTIVARIATE CONTROL CHARTS APPLIED TO DATA INSTRUMENTATION OF DAMS. *Independent Journal of Management & Production*. 7 (2016), Doi:10.14807/ijmp.v7i1.369.
22. P. Phaladiganon, S. B. Kim, V. C. P. Chen, W. Jiang, Principal component analysis-based control charts for multivariate nonnormal distributions. *Expert Systems with Applications*. 40, 3044–3054 (2013).
23. D. Slišković, R. Grbić, Ž. Hocenski, Multivariate statistical process monitoring. *Tehnicki Vjesnik-Technical Gazette*. 19, 33–41 (2012).
24. F. S. L. Costa, R. H. P. Pedroza, D. L. Porto, M. V. P. Amorimb, K. M. G. Lima, Multivariate control charts for simultaneous quality monitoring of isoniazid and rifampicin in a pharmaceutical formulation using a portable near infrared spectrometer. *Journal of the Brazilian Chemical Society*. 26, 64–73 (2015).
25. M. Zude, *Optical monitoring of fresh and processed agricultural crops* (CRC Press, 2008).