Journal of Process Control 69 (2018) 97–102

Contents lists available at ScienceDirect

Journal of Process Control

journal homepage: www.elsevier.com/locate/jprocont

Multivariate statistical process control in annual pharmaceutical product review

Adamastor R. Tôrres[a], Antônio Diógenes P. de Oliveira[b], Severino Grangeiro Júnior[b], Wallace D. Fragoso[a,*]

[a] Chemistry Department, Federal University of Paraíba, João Pessoa, PB, Brazil
[b] Pharmaceutical Laboratory in the state of Pernambuco, Governor Miguel Arraes LAFEPE, Recife, PE, Brazil

ARTICLE INFO

Article history:
Received 3 February 2016
Received in revised form 29 March 2018
Accepted 5 June 2018

Keywords:
Annual product review
Multivariate statistical process control
Principal component analysis
Reconstruction-based contributions
Hydrochlorothiazide

ABSTRACT

Annual Product Review recommending by the Food and Drug Administration and the International Conference on Harmonization is a periodic assessment of products quality to verifying the consistency of the production process and the quality of the finished product. In this work it was performed an annual product review using multivariate statistical process control. Multivariate control charts was used for the periodic evaluation of the Hydrochlorothiazide 25 mg tablet, produced between 2009 and 2013 in the Pharmaceutical Laboratory of Pernambuco (Brazil). The procedure has allowed one to monitor and diagnose deviations in the production process. A corrective action was taken in the process by the Quality System in 2011 and the methodology applied was used to evaluate this action's effectiveness. Through the Multivariate Statistical Process Control it was possible to make an annual product review in accordance with the pharmaceutical industry regulatory requirements.

1. Introduction

The Statistical Process Control (SPC) is the monitoring of the stability of a process or product from a variable of a process or product. A control chart is constructed following the characteristic variable of the process, or a statistic calculated from it over time. The use of SPC in the pharmaceutical industry has increasingly found applications in the search for continuous improvement of the quality management system [1,2]. The SPC is a valuable tool that incorporates the concepts of Good Manufacturing Practices and provides essential information for monitoring processes, signaling the possible sources of quality deviations.

The Annual Product Review (APR) is a periodic assessment of the quality of products, with the goal of verifying the consistency of the production process and the quality of raw material and finished product. According to the guidance documents published by the Food and Drug Administration (FDA) [3] and by the International Conference on Harmonization (ICH) [4], pharmaceutical companies must use the APR as a strategy of global analysis of performance of production and commercialization of their products. This strategy should include a continuous assessment of the manufacturing process to ensure that the product remains in statistical control during commercialization. Furthermore, it should allow the identification of trends in the process and the possible need for intervention to make improvements in the final product [5].

The assessment of the production process in the APR may be performed through the SPC. The univariate SPC, however, does not take into account possible synergistic relationships between various variables of the products which must be monitored in a medicine. This involves the design and evaluation of a large number of control charts, one for each variable. This procedure makes the interpretation harder and favors unnecessary control actions. Consequently, it causes false alarms in the process. The results of the quality characteristics of the finished product can be combined and analyzed together in the APR through the Multivariate Statistical Process Control (MSPC) [6].

The Principal Component Analysis (PCA) is the method of projection of latent structures most commonly used in multivariate data analysis. When PCA is used in MSPC two control charts are built. The first one is the T^2 chart, designed with the PCA scores, that monitors if a specific sample has a systematic difference regarding the samples considered under statistical control. The second one is the SPE chart, or Q chart, which monitors if the residual of a new sample, when projected into the model, is under statistical control. [7]. A more recent strategy simplifies the MSPC by com-

* Corresponding author at: Universidade Federal da Paraíba, Centro de Ciências Exatas e da Natureza (CCEN), Departamento de Química. Campus I, Jardim Cidade Universitária s/n, João Pessoa, PB, Postal Code: 58051-900, Brazil.
E-mail address: wallace.fragoso@pq.cnpq.br (W.D. Fragoso).

https://doi.org/10.1016/j.jprocont.2018.06.001
0959-1524/

bining the T^2 and SPE statistics, normalized by their control limits, into a new statistic, the combined index φ, allowing the monitoring of the process with a single control chart. [8–11]. These methods have been applied in the statistical control of pharmaceutical processes through the monitoring of chromatographic [7,12,13], spectroscopic [14–16] and physical parameters [17,18].

A problem in the interpretation of an out of control multivariate signal is to determine which variable or group of variables are contributing to the fault control. There are some methods that help identify the variables that influence the stability of the process [19–24]. When the φ statistic is used, the Reconstruction-based Contributions (RBC) [8,25,26] is the appropriate way to identify the origin of the faults.

In this work is proposed the use of MSPC, the combined index φ and the RBC to evaluate and diagnose the consistency of the production process during the APR of Hydrochlorothiazide 25 mg tablets produced over the past five years by the "Laboratório Farmacêutico do Estado de Pernambuco Governador Miguel Arraes" (LAFEPE), Brazil.

2. Theory

2.1. Principal component analysis

PCA is the decomposition of a data matrix **X** into two others, the matrix of scores **T** and the matrix of loadings **P,** plus a matrix of residue **E,** as shown in Eq. (1):

$$\mathbf{X}_{(nxp)} = \mathbf{T}_{(nxh)}\mathbf{P}_{(pxh)}{}^{T} + \mathbf{E}_{(nxp)} \quad (1)$$

where h indicates the number of principal components (PCs) retained in the model. Each principal component is associated with an amount of explained variance λ so that the first principal component (PC1) is the direction in multivariate space that explains the greatest amount of variance in the database. The second principal component (PC2) is the direction that explains the greatest amount of variance that is not explained by PC1 being orthogonal to it, and so on. The loadings correspond to the cosines of the angles between the original variables and the PCs, thus representing how much each original variable contributes to a specific PC. The scores represent the coordinates of the samples in the axis system formed by the Principal Components [27,28].

After performing the PCA, the physical interpretation of the Principal Components is desirable. This interpretation, however, is not always obvious or even possible. This is because each Principal Component is influenced by a portion of the information from each variable.

2.2. Multivariate statistical process control charts

Kourti and MacGregor [29] and Kourti, Lee and MacGregor [30] described the application of multivariate projection methods such as PCA for monitoring and diagnosis of faults in continuous processes and in batches. These methods are used in combination with Hotelling's T^2 statistic in multivariate quality control. In MSPC a PCA is performed on a set of historical data of the process where only common causes of variation are present (ïn-controlšet). The behavior of future observations is referenced with respect to this ïn-controlšet [31].

The scores of the most significant components are used to calculate the T^2 statistic according to Eq. (2):

$$T_i^2 = \sum_{k=1}^{h} \frac{t_{ik}^2}{\sigma_{t_{ik}}^2} \quad (2)$$

where t_{ik} is the score of the k^{th} Principal component, $\sigma_{t_{ik}}^2$ is the variance of t_k and h denotes the number of Principal Components retained in the PCA model [31]. The T^2 chart constructed with the PCA scores measures the distance from an observation to the center of the ïn-controlšet and determines whether a future observation has a systematic deviation in relation to the samples considered in statistical control [29]. ∴ There are typically two distinct phases in the construction of the T^2 chart [32]. Phase I corresponds to the retrospective study of the case, in order to verify if the process was in statistical control when the first observations were being collected. In Phase II control charts are used to test whether the process remains in control when future observations are monitored. The control limit of the T^2 chart for individual observations in Phase I is:

$$CL_{T^2}^{I} = \frac{(n-1)^2}{n}\beta_{1-\left(\frac{\alpha}{2}\right)\left(h/2,\ \frac{(n-h-1)}{2}\right)} \quad (3)$$

where $\beta_{1-\left(\frac{\alpha}{2}\right)\left(h/2,\ \frac{(n-h-1)}{2}\right)}$ is the percentile 100(*1*-α) of the beta distribution with parameters *h/2* e *(n-h-1)/2*, *n* is the number of samples and *h* is the number of principal components retained in the model. For Phase II, the new limit to be used is given by:

$$CL_{T^2}^{II} = \frac{h(n+1)(n-1)}{n(n-h)}F_{1-\alpha;h,n-h} \quad (4)$$

where $F_{1-\alpha;h,n-h}$ is the percentile of the F distribution being *h* and *n-h* degrees of freedom [32,33].

On other hand, the PCA matrix of residue **E** shall be used to build an additional control chart, the SPE chart. The SPE statistic is calculated by:

$$SPE_i = \sum_{p=1}^{P}\left(x_{ip} - \hat{x}_{ip}\right)^2 \quad (5)$$

where x_p and $\hat{x}_p$ are the value of p^{th} variable and its predict value for the i^{th} observation, respectively. The control limits for the SPE chart are given by [32]:

$$CL_{SPE} = \theta_1 - \left[1 - \theta_2 h_0\left(\frac{1-h_0}{\theta_1^2}\right) + \frac{\sqrt{z_\alpha\left(2\theta_2 h_0^2\right)}}{\theta_1}\right]^{1/h_0} \quad (6)$$

where **V** is the covariance matrix of **E**, θ_1 is the trace of **V**, θ_2 the trace of $\mathbf{V}^2$, θ_3 the trace of $\mathbf{V}^3$, $h_0 = 1-(2\,\theta_1\,\theta_3/\,3\,\theta_3^2)$ e z_α is standardized normal variable with (*1*-α) reliability level. SPE statistic is the square Euclidean distance perpendicular to an observation from the subspace defined by PCA, and gives a measure of how close this observation is the h-dimensional subspace.

The T^2 and SPE statistics can be combined in a single index in order to simplify the monitoring and fault detection task. [8, 25] The combined index (Φ) incorporates as two statistics according to the Eq. (7):

$$\phi = \frac{SPE_i}{CL_{SPE}} + \frac{T_i^2}{CL_{T^2}^{II}} \quad (7)$$

The control limit for Φ is calculed by [8]:

$$CL_\varphi = g^\varphi \chi_\alpha^2(h^\varphi) \quad (8)$$

with (1-α) x 100% confidence level, where

$$g^\varphi = \left(\frac{pc}{(CL_{T^2}^{II})^2} + \frac{\theta_2}{(CL_{SPE})^2}\right) / \left(\frac{pc}{(CL_{T^2}^{II})} + \frac{\theta_1}{(CL_{SPE})}\right) \quad (9)$$