Analytica Chimica Acta 952 (2017) 9–17

Contents lists available at ScienceDirect

Analytica Chimica Acta

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

Multivariate statistical process control (MSPC) using Raman spectroscopy for in-line culture cell monitoring considering time-varying batches synchronized with correlation optimized warping (COW)

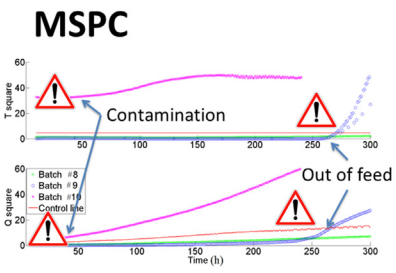
Ya-Juan Liu [a], Silvère André [a], Lydia Saint Cristau [b], Sylvain Lagresle [c], Zahia Hannas [c], Éric Calvosa [b], Olivier Devos [a], Ludovic Duponchel [a, *]

[a] LASIR CNRS UMR 8516, Université de Lille, Sciences et Technologies, 59655 Villeneuve d'Ascq Cedex, France
[b] Sanofi Pasteur, 1541 Avenue Marcel Mérieux, 69280 Marcy-l'Étoile, France
[c] Merial, 29 Avenue Tony Garnier, 69007 Lyon, France

HIGHLIGHTS

- Multi-way principal component analysis (MPCA) is applied on Raman spectra for MSPC monitoring of cell cultures.
- Correlation Optimized Warping is used to synchronize batches with varying durations analyzed with Raman spectroscopy.
- Early contamination is detected with this tool while it is only observed more than 200 h through a visual inspection.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 29 August 2016
Received in revised form
18 November 2016
Accepted 21 November 2016
Available online 2 December 2016

Keywords:
Cell culture
Raman spectroscopy
Chemometrics
Multivariate statistical process control (MSPC)
Correlation optimized warping (COW)
Multi-way principal component analysis (MPCA)

ABSTRACT

Multivariate statistical process control (MSPC) is increasingly popular as the challenge provided by large multivariate datasets from analytical instruments such as Raman spectroscopy for the monitoring of complex cell cultures in the biopharmaceutical industry. However, Raman spectroscopy for in-line monitoring often produces unsynchronized data sets, resulting in time-varying batches. Moreover, unsynchronized data sets are common for cell culture monitoring because spectroscopic measurements are generally recorded in an alternate way, with more than one optical probe parallelly connecting to the same spectrometer. Synchronized batches are prerequisite for the application of multivariate analysis such as multi-way principal component analysis (MPCA) for the MSPC monitoring. Correlation optimized warping (COW) is a popular method for data alignment with satisfactory performance; however, it has never been applied to synchronize acquisition time of spectroscopic datasets in MSPC application before. In this paper we propose, for the first time, to use the method of COW to synchronize batches with varying durations analyzed with Raman spectroscopy. In a second step, we developed MPCA models at different time intervals based on the normal operation condition (NOC) batches synchronized by COW. New batches are finally projected considering the corresponding MPCA model. We monitored the evolution of the batches using two multivariate control charts based on Hotelling's T^2 and Q. As illustrated with results, the MSPC model was able to identify abnormal operation condition including contaminated

* Corresponding author. Address: LASIR CNRS UMR 8516, Université de Lille, Sciences et Technologies, Bât. C5, 59655 Villeneuve d'Ascq Cedex, France.
E-mail address: ludovic.duponchel@univ-lille1.fr (L. Duponchel).

http://dx.doi.org/10.1016/j.aca.2016.11.064
0003-2670/

batches which is of prime importance in cell culture monitoring We proved that Raman-based MSPC monitoring can be used to diagnose batches deviating from the normal condition, with higher efficacy than traditional diagnosis, which would save time and money in the biopharmaceutical industry.

1. Introduction

In biotechnology, in-line monitoring of process batches is important to improve the batch-to-batch reproducibility, as the deteriorated product can be avoided by early detection of batches out of the normal operation condition (NOC) [1,2]. Conventional statistical process control is routinely used in biopharmaceutical processing often based on one variable (or at most a few ones) watched over by plant operators [3,4]. However, univariate statistical process control may have late or wrong alarms because single or a few variables are not enough for complex bioprocess monitoring [2]. Thus, multivariate statistical process control (MSPC) has been developed since acquiring multivariate dataset is common in the biotechnology [5–11]. The objective of MSPC is to monitor the process and diagnose the batch out-of-control based on datasets with multiple variables, by extracting the information from the variables collinear to each other [12,13]. MSPC is generally applied in bioprocess considering two types of datasets. The first one has variables such as temperature, oxygen pressure, pH, etc. available in real-time from various sensors [5–9]. However, such data sets are not enough for monitoring complex bioprocess because they don't reveal variations of different chemical compounds evolved in the cell culture. That is why the second kind of data set generated with in-line spectroscopy is increasingly common in MSPC. Near-infrared (NIRS) [10,11] and Raman spectroscopy [14] are two popular analytical methods used in biotechnology. Indeed they have the advantages of low water interference, be non-destructive, no sample preparation and low reagent consumption to perform in-line monitoring. However compared with NIRS, Raman spectroscopy is often considered to contain more chemical information in its spectrum [15]. Thus in the proposed work we monitor cell cultures for antibody production using an in-line Raman spectrometer with immersion probes in the bioreactor.

There are a number of multivariate methods to implement MSPC and multi-way principal components analysis (MPCA) is one of the most commonly used unsupervised methods [12]. MPCA decomposes the matrix which is unfolded from the three-way dataset collected from different batches with a number of variables (e.g. wavenumber in spectroscopic dataset) in time series. MPCA is expected to describe the dataset with a few orthogonal principal components (PCs), keeping the useful information and diminishing the influence of noise, coherence, and outliers. In addition to this, three-way methods such as parallel factor analysis (PARAFAC) have also been applied in MSPC analysis because of the natural three-way structure (batch × variable × time) of the dataset [16]. PARAFAC is expected to obtain the pure profiles of different components based on the trilinear decomposition of the three-way array. The latent variable resulting from PARAFAC means the pure profile for a certain compound when latent variables in PCA represent the directions of highest variances. Thus, for the dataset without trilinear structure, which is common in biotechnology owing to the noise, outliers, and the coherence between different compounds, PARAFAC decomposition would lead to wrong actions for the bioprocess, with the meaningless latent variable. On its part, MPCA has a lower requirement for the linear structure of the dataset. Thus, we choose MPCA in this work for MSPC modeling to monitor the cell culture with different batches based on Raman spectroscopy.

However, it is common that the spectroscopic profiles for different batches are unsynchronized in the time series with the addition of different time lengths. Indeed this kind of data structure is generated when multiple immersion probes are connected to one spectrometer inducing spectroscopic acquisition in an alternate way. Yet most of the multivariate methods, including MPCA, are applied on batches with even length. Except for this, unsynchronized batches destroy the linear structure of datasets, leading to biased results of MSPC. Fortunately, Dynamic time warping (DTW) [17] and correlation optimized warping (COW) [18] have been applied to synchronize the dataset with varying batch-to-batch process times. Both of these two methods synchronize the trajectory by stretching or compressing the sample vector according to a reference. DTW uses the principle of dynamic programming and non-linear warps for the two signals with the minimum distance between them as a criterion. Differently, COW aligns the signal in a piecewise way, and the optimal alignment is determined by the correlation between the aligned fragments of signal and reference [19]. The introduction of segment strategy and the application of correlation replacing distance as the criterion result in a better performance of COW as a linear method. The most popular application of COW is to align peaks in chromatography [19]. Recently it has been used in MSPC analysis for synchronization based on datasets with variables such as temperature, pressure, flow, pH and weight from plant sensors [18]. However, COW has never been used for batch monitoring based on spectroscopic dataset to synchronize the time-varying data sets. In this paper, for the first time, we used COW to synchronize the acquisition times of Raman spectroscopy for MSPC modeling.

After synchronization by the COW method, we analyzed a set of NOC batches using MPCA at different time intervals and projected new batches on the MPCA models to estimate their potential deviations. According to the MPCA models, we monitored the batches in a view of control charts based on T^2 and Q values. In control charts, control limits has been estimated based on the confidence intervals of the T^2 and Q values of the NOC batches. New batches with T^2 and Q values within the control limit are considered as normal batches and vise versa. In this work we monitored cell cultures and detected 'abnormal' batches using the control charts.

2. Material and methods

2.1. Cell cultures

IgG antibody productions were conducted using 7 L Sartorius bioreactors (4 L working volume) and a CHO (GS) cell line at Sanofi Pasteur Company. We monitored pH by a control probe (Mettler Toledo, Columbus, OH, USA) to maintain its value to 7.0 ± 0.2. Dissolved oxygen with 40% of air saturation was monitored by a DO sensor (Mettler Toledo). The process for the production of antibody took 14 days considering continuous feeding with glucose and glutamate. Ten batches were monitored and they were numbered from Batch #1 to #10 in this work. Batches #1 to #7 were considered as historical batches. The first six had satisfactory antibody