CHEMOMETRIC PROCESS ANALYTICAL TECHNOLOGY (PAT) APPLICATIONS IN BIOPROCESS ENGINEERING

João A. Lopes[1], Teresa P. Alves[2], José C. Menezes[1]

[1]Centre for Chemical & Biological Engineering, IST, Technical University of Lisbon, Av. Rovisco Pais, P-1049-001, Lisbon, Portugal
[2]CIPAN, S.A.,Vala-do-Carregado, P-2601-906 Castanheira do Ribatejo, Portugal.

Abstract: The production process of an active pharmaceutical ingredient (API) by fermentation in an industrial environment was analysed using process analytical technologies. The Food and Drug Administration's (FDA) process analytical technology (PAT) initiative is intended to be a collaborative effort with industry to promote the integration of new manufacturing technologies into pharmaceutical production. Within the PAT framework the aim is to design, develop and operate processes consistently to ensure a predefined quality at the end of the manufacturing process. Several case-studies are discussed in which combining chemometrics methods with NIR spectra from samples in different process stages leads to increased process understanding and process control of an API production process.

Keywords: pharmaceutical production; process monitoring, near-infrared spectroscopy; chemometrics; process analytical technology PAT

1. INTRODUCTION

Process analytical technologies (PAT) are "systems for analysis & control of manufacturing processes based on timely measurements of critical quality parameters and performance attributes of raw - materials and in-process products, to assure acceptable end-product quality at the completion of the process" (i.e., quality by design). PATs are a landmark in the acceptance of process systems engineering (PSE) tools in modern pharmaceutical manufacturing and quality assurance of food and drug processes in general. PATs involve the application of process analytical chemistry (i.e., inprocess monitoring techniques), chemometrics (e.g., data-based modelling techniques) and process control techniques (viz., intelligent use of process data with multivariate supervision and diagnosis strategies).

PATs as such were born in XXI century for this century's industries and are more than a simple sum of existing disciples as there is a significant gain in their integrated use. Successful PAT applications are the real driver for developments in some of its supporting sciences and not the other way around. In terms of process monitoring the use of PATs represent a paradigm shift in the sense that sophisticated quality control moves from labbased to process-based (i.e., in-process). Numerous publications on this subject have appeared in the last few years (2000-2003) as a simple web search with the PAT acronym will reveal. In fact, the discipline is going through an exponential growth phase at the moment with dedicated sessions and even entire congresses worldwide.

The pharmaceutical industry of active pharmaceutical ingredients (API) production by fermentation (e.g., antibiotics) is one of the segments in which - due to the intrinsic characteristics of the processes involved – the introduction of PATs is expected to make a significant contribution on improving process understanding, observability and control.

Many factors contribute to the increased system complexity of bioprocesses in terms of monitoring and control, in large-scale industrial conditions. The high cell densities required for an economic bioprocess lead to mass transfer problems that add up to the overall intracellular biochemical complexity of fermentation processes. The raw materials (substrates) that give higher process productivities are natural agro-products subject to significant and uncontrolled variability in composition (e.g., soybean flour), that if anticipated can be partially or sometimes almost entirely compensated for. Strain improvement (microorganism breeding) is carried out by selection of isolates that give higher shake-flask (batch) productivities, thus selecting cultures more adapted to no external controls, while later that capability will work against controllability of the large scale bioprocess. Finally, quality of the seed culture (i.e., inoculum) has a profound effect on the production stage (i.e., fermentation) productivity and should therefore be close monitored and controlled. In all, the apparent irreproducibility of bioprocesses is thus the outcome of many poorly understood and/or monitored and therefore uncontrolled phenomena combined together (Chataway et al., 1993).

In this paper the production process of an API by fermentation in an industrial environment was analysed using process analytical technologies. Using different case-studies the use of PAT both as a process analytical chemistry tool and as a process chemometrics tool, will be illustrated. In the subsequent paper the use of PAT within the process supervision and diagnosis context will be discussed.

2. PROBLEM DESCRIPTION

Figure 1 illustrates the stages involved in the selected biochemical process: the API's production phase (inoculum production and fermentation) and the API's isolation phase (filtration, extraction and precipitation). Several variables are monitored across the process. The most important are the liquid and gas phase variables monitored in the inoculum and fermentation stages by standard reference methods (Costa, 2000; Neves et al, 2001). Table 1 describes the potential of application of the NIR monitoring technique over several stages of a biochemical process. Near infrared (NIR) spectroscopy is an established analytical technique, widely used to predict the chemical composition of agricultural raw-materials and in the last decade pharmaceutical industry (Chalmers, 2000; Costa, 2000). NIR has many advantages over other analytical techniques. It can be used at-line or on-line given that appropriate process are available, to measure chemical and physical properties of a sample. The low cost of NIR operation and simplicity of sample preparation and analysis speed are also very attractive factors. Therefore, NIR spectroscopy is especially suited to build PAT monitoring strategy upon especially in the pharmaceutical industry. In this paper we have investigated the feasibility and benefits of using NIR spectroscopy in the above context, in several process stages: quality assessment of fermentation raw-materials, fermentation process monitoring, monitoring the downstream API purification process and moisture content in solid API samples.

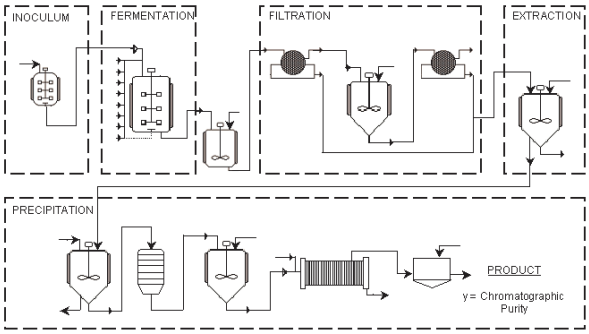


Fig. 1. Schematic representation of the industrial pharmaceutical API production process.

Table 1. Application of FT-NIR spectroscopy in biochemical processing

	Fermentation	Extrac-tion	Chemical synthesis	Formu-lation
Sample type	Solid; Liquid	Liquid	Solid; Liquid	Solid
Principle	R ; T	T	R; T	R
Analysis type	QT ; QL	QT	QT	QT ; QL

Legend: R-Reflectance; T-Transmitance, QT-Quantitative; QL-Qualitative

2.1 Chemometric Modelling

Chemometric tools are methods to establish relationships between different measurements from a chemical system or process with the state of the system through the application of mathematical or statistical methods. Typically, a chemometric problem might be the definition of a relationship between properties of interest (e.g., difficult to measure in the lab) based on knowledge of other properties easily obtained and which affect the property of interest (Otto, 1998). The values of these variables are generally obtained by doing experiments and in general no first principles assumption are made. Therefore, a set of data must be selected from process measurements preferentially coming from designed experiment because the training set should cover the space spanned by the process. The next step is to build and validate a model either by using multivariate regression or multivariate classification methods depending on the purpose of the model. There are many methods to validate a model. In this paper we used two different approaches depending on the type of the samples. If samples are independent (that is they are not related in time) the leave-one-out method (LOO) was adopted. LOO omits only one sample, uses the rest to create the model, and tests the one remaining sample.

The procedure is repeated for each sample separately. If samples are related in time (e.g., taking spectra during a fermentation) then entire batches are left out for validation to avoid over fitting.
Regression methods include multiple linear regression (MLR), principle component regression (PCR), and partial least squares (PLS) (Martens and Martens, 2001). Classification methods include discriminant linear analysis, principal component analysis (PCA), factor analysis (FA), cluster analysis (CA) (Joliffe, 1986). Non-linear techniques such as neural networks for calibration and Kohonen networks for classification are often used but are in general less robust than linear techniques because of lack of extrapolation capacity. Partial least squares were adopted to build calibration models and Kohonen networks to build classification models.

2.2 Measure for model goodness of fit

The amount of variance predicted Q^2_Y is the goodness of fit measure used throughout this paper. Model predictions ($\hat{y}$) used to estimate Q^2_Y are always obtained for samples not used in calibration (e.g., cross-validation).

$$Q_Y^2 = \left(1 - \frac{(y-\hat{y})^T (y-\hat{y})}{y^T y}\right) \cdot 100 \qquad (1)$$

3. EXPERIMENTAL

The process used in this work was the industrial production of an antibiotic-like molecule (API – active ingredient product) by fermentation (Neves et al., 2001). Fed-batch cultivation of a Streptomycetes strain was carried out using a nondefined (complex) medium containing soybean flour and a carbon source. Soybean flour is a typical protein-based solid substrate commonly used in industrial fermentation media, which contains a non-negligible amount of insoluble materials. The conditions used were typical of those employed routinely in industry for aerobic microbial growth (Strohl, 1997). Depending on the experiment, 24 and 48 hours old inocula were used to inoculate the production fermenter and different fermenter sizes were used. The fermentation process lasts approximately 140 hours. After the production fermentation step the culture media was transferred for downstream processing. The latter includes a filtration, an extraction and a precipitation stage after which the product is obtained with the required chromatographic purity. The main stages of the active product ingredient (API) production process are represented in Fig. 1. NIR spectra were collected for different sample types, with a Bomem FT-NIR spectrophotometer (12,000 – 4,000 cm^{-1}) series MB 160, with a tungsten-halogen light source and an InGaAS detector, with accessories for diffuse reflectance and transmission measurements both by static accessories and fiber optics (Axiom's reflectance probe and Solvias AG transflectance probe).

4. RESULTS AND DISCUSSION

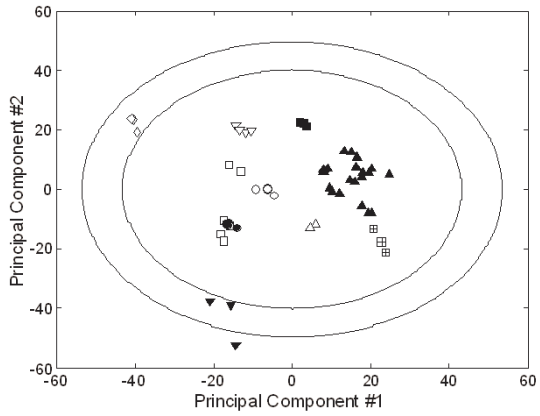
4.1 Raw-materials quality diagnosis

One of the most important substrates in this process is soybean flour, which is the principal nitrogen source. Experience proved that the simple change in the batch or supplier of the flour could be disastrous in the antibiotic productivity, even when it's chemical characterisation, made by usual methods, gives the same results. We intended to prove that, under more or less established operating conditions, is possible to predict in advance the final quality of fermentation using data from the flour used for inoculum growth and fermentation (Welsh et al., 1996). A set of 45 soybean flours samples from 10 different lots was characterisation by FT-NIR reflectance spectroscopy. The flour discrimination obtained with PCA model for the two first principal components is given in Fig. 2a. Since NIR was able to discriminate between different lots, a set of 25 soybean flour samples (used in 25 fermentation runs) was monitored with NIR spectroscopy. Spectra PCA scores were classified with a Kohonen network with 4 output nodes (target classes). The obtained classification was compared with experimental values of the corresponding fermentation final API concentrations (Fig. 2b). The analysis shows that the Kohonen network did predict successfully approximately 70% of the examples. In cases where model fails, the prediction is always in first neighbourhood of the real value. Therefore we can conclude that such model can identify what we should expect at the end of the fermentation given the soybean flour used.

4.2 Inoculum/fermentation processes ab-initio diagnosis

It is known that product formation can also be associated with the inoculum quality and the fermentation operation. A different type of PLS model, the multiblock PLS model (Westerhuis and Smilde, 2001), was applied to regress the API concentration obtained at the end of the fermentation stage with the inoculum and fermentation data (Lopes et al., 2002). The aim is to inspect the relative importance's of the inoculum and fermentation stages in the API productivity. From the model block-weights, the influences of each stage in the overall production process can be established (Brás et al., 2004). The variables were classified in two groups: manipulated and quality variables. Manipulated variables are used to control the process (inputs to the process).

(a)



(b)

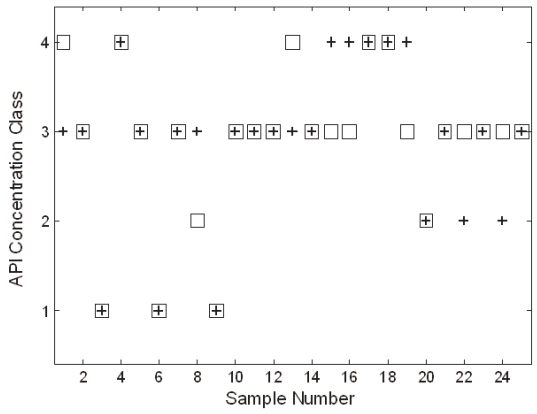


Fig. 2. a) MPCA scores for sample mode extracted from 45 soybean flour samples NIR spectra of ten different manufacturers (89.3% of captured variance in the two components). b) Kohonen network classification based on raw-materials quality (NIR spectra) of the final API concentration compared with experimentally obtained values for 25 pilot-scale fermentations classified as: 1) very good, 2) good, 3) low, 4) poor; (□ experimental value, + model classification).

Quality variables are descriptive of the process state (process outputs). 30 runs were considered (NOC and designed experiments). A five components model produced an amount of variance predicted of 80.2%. Table 2 shows that the inoculum stage plays an important role on the multiblock model. The average importance of the inoculum stage on the model in two PLS components was about 30%. This clearly indicates that in the future it is necessary to increment the quality control of the inoculum growth stage. A model of this type could be used in the early detection of a poor quality inoculum (or the selection of the best inoculum among several possible candidates) leading to improved fermentation performance and reduced process variability. Even if the inoculum growth duration parameter does not seems to affect the amount of API obtained it is important because it reduces the production fermentation stage variability in a significant way. In the following sections we report the application of at-line NIR spectroscopy to monitor the API production process in the fermentation and extraction stages and to monitor humidity in solid samples of different compounds (intermediate and final product).

Table 2. Importances of the inoculum production and fermentation stages in the final API productivity obtained for two latent variables of a multiblock-PLS model. Variables were blocked in operation (feed-rates, agitation, aeration, temperature) and quality (pH, concentrations, carbon evolution rate, oxygen uptake rate) variables.

	Inoculum Stage			Fermentation Stage		
lv	O	Q	T	O	Q	T
1	11 %	16 %	27 %	46 %	27 %	73 %
2	32 %	10 %	42 %	27 %	31 %	58 %

Legend: O (operation); Q (quality); T (total).

4.3 Fermentation process monitoring

Analytical methods to determine residual concentrations on typical fermentation media take from 30 minutes to several hours to give a result. The process engineer cannot therefore rely on these measures availability to act on the process if a quick corrective measure is needed. To evaluate the applicability of the NIR spectroscopy on fermentation complex media monitoring, 7 pilotscale fermentation runs were used to identify models for the viscosity (an indirect measure of biomass) and API concentration. Diffuse reflectance spectra were obtained at-line from fermentation broth samples. Fig 3a and Fig. 3b show the raw and processed diffuse reflectance NIR spectra respectively of a selected fermentation recorded over time.

A PLS model was calibrated against experimental data obtained with reference methods. Fig. 4a and 4b present the results obtained for the viscosity and API models (calibration and validation) respectively. The relative mean errors were 5% and 6% for the viscosity and API contents respectively. From these results, industrial pilot-plant trials using an in-situ reflectance probe are being conducted to monitor on-line the fermentation media constituents. From the reflectance NIR spectra real-time information on the four key fermentation analytes and biomass (calibrated against viscosity) is being obtained and will hopefully be integrated into an advisory systems for advanced process control.

4.4 Downstream process monitoring

API purification in the studied process involves one ionic exchange column. The established chemical assay method at the plant takes approximately 10 minutes to provide a measure.

This is inadequate for optimal process operation and leads to a very conservative scheme of fraction collection. With a NIR spectrometer an estimate of the API content can potentially be made in 30 seconds. Thus, the transmitance NIR spectra of samples taken at-line throughout the elution process in several batches were used to develop an in-process

monitoring alternative. Spectra preprocessing included: the subtraction of the spectra taken at the beginning of the elution cycle (zero time) to reduce the changing effects due to the matrix, baseline correction, computation of the second derivative with a Savitsky-Golay filter and selection of appropriate wavenumber regions in the spectra (using cross-validation).

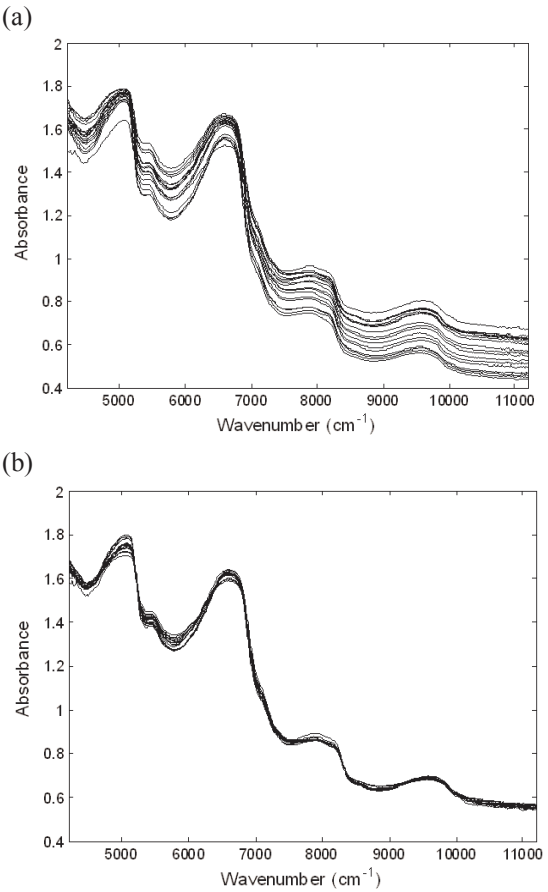


Fig. 3. Diffuse reflectance spectra of the fermentation media recorded over time. NIR spectra was measured at-line using samples collected over time (a) raw spectra; b) processed spectra by multiplicative scatter correction).

A PLS model was calibrated with samples of known API concentration obtained with the reference method. Fig. 5 illustrates the predictions of the API concentration for nine elution cycles (validation experiments) not used in calibration development. The mean relative error for this model is 6.4%, which is in good agreement with the experimental error of the reference method (~5%).

4.5 Moisture Content Monitoring

Measurement of moisture content of API samples taken from dryers or blenders is one of the most significant PAT applications that may be established in a pharmaceutical specialities environment, as materializes most of the benefits that characterize PATs - viz., reducing off-line expensive and time-consuming analysis on specialized quality control labs, better use of available production equipment with faster monitoring techniques, consistent process operation, reduction on product variability and consequent higher quality assurance.

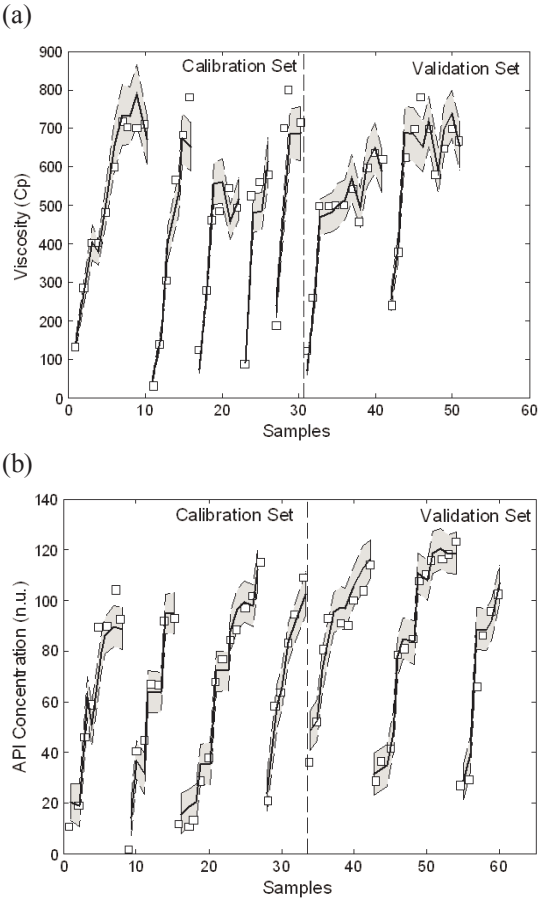


Fig. 4. Predictions of viscosity (a) and API concentration (b) from NIR spectra for selected calibration and validation datasets. The 95% confidence bands were drawn (□ experimental, ▬ FT-NIR based PLS model predictions).

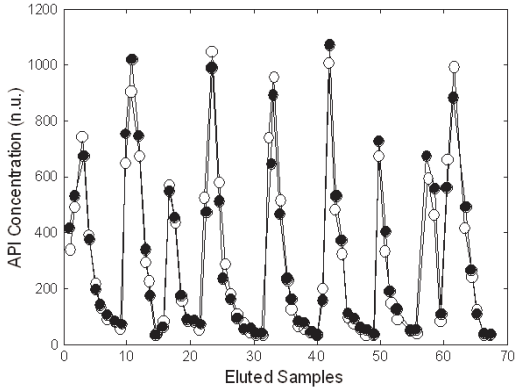


Fig. 5. API concentration predictions by PLS using FT-NIR spectra and the corresponding reference analytical method results in nine elution cycles not used in calibration development (○ reference method, ● FT-NIR based PLS calibration method).

Five hundred and fifty solid samples of four API intermediates and the final API product were

monitored at-line with reflectance NIR for moisture. Four PLS models were built to model moisture content of the four different solids. Since samples are independent the number of latent variables selected for each model were optimised using the LOO validation strategy. Table 3 depicts the results obtained for each calibration model. The validation correlation coefficient (Q^2_Y) was in all cases higher than 90%. For three products the Q^2_Y was higher than 99%, which reflects that indeed NIR spectra captured the information required for PLS to model this quality parameter. Relative errors for the FT-NIR based model were higher than the reference method relative errors as expected. However, they are acceptable from the point of view of the quality control of the company. Therefore, the reference method was replaced by the NIR spectrometer for routine analysis. Reference method is currently used only to validate humidity content of the final product.

Table 3. PLS model statistical parameters for the determination of humidity in different solid product samples (P#) based on FT-NIR reflectance spectra.

	Number of samples	LV	Q^2_Y (LOO)	Reproducibility error (%)	Model prediction error (%)
P1	70	5	0.991	0.1	0.4
P2	300	9	0.993	0.3	0.4
P3	135	7	0.996	0.2	0.5
P4	45	14	0.920	--	--

5. CONCLUSIONS

In this paper we reported on process analytical technology applications on an industrial multistage API production process. NIR spectroscopy was selected as the technique to monitor several production steps. From raw-material qualification with process analytical techniques based on NIR, to the analysis of the main antibiotic production process step (fermentation) and downstream operations, with a variety of exploratory data, modelling and multivariate monitoring techniques, we have demonstrated the significant role that chemometrics can play in bioprocessing and bioprocess development. The combination of different types of process information was demonstrated in the ab-initio prediction of the final titter of a nominal antibiotic fermentation. NIR spectroscopy was applied to monitor at-line different quality variables for process intermediates solid and liquid samples. The potential impact of the obtained results are obvious, in terms of re-designing both monitoring and sampling schemes to be used either during the process R&D stages as well as in routine process production, to improve process analysis, supervision and diagnosis. The combined application of different techniques was successfully demonstrated for the studied API production process. Nevertheless, a more integrated use of chemometrics and its embedding right from the beginning into the R&D process is desirable. Namely if process analysis and optimisation from an overall process perspective (i.e., considering all relevant process stages, operating variables, raw-materials, microorganism strain) is attempted as a faster process prototyping strategy, chemometrics will certainly become a significant enabling-science in bioprocess engineering.

ACKNOWLEDGEMENTS

The authors would like to thank Companhia Industrial Produtora de Antibióticos SA (CIPAN) in Portugal for authorizing this publication and co-financing the work together with The National Innovation Agency (ADI). The financial support from the Foundation for Science and Technology (SFRH/BPD/7194/2001) is gratefully acknowledged by JAL

REFERENCES

Brás, L.P., J.A. Lopes, C.R. Santos, J.P. Cardoso, J.C. Menezes (2004) Modelling and Identification of Individual Stage Contributions in an Industrial Pharmaceutical Process by Multiblock PLS, European Symposium on Computer Aided Process Engineering (ESCAPE-14).

Chalmers, J. (2000) Spectroscopy in Process Analysis, Sheffield Academic Press, UK, p. 380.

Chattaway, T., G.A. Montague, and A.J. Morris (1993) Fermentation Monitoring and Control., in Bioprocessing., vol. 3, pg. 319-354, 2nd ed., Ed. G. Stephanopoulos in "Biotechnology", Eds. H.J. Rehm, G. Reed, VCH, Germany.

Costa, P.F. (2000) “Aplicações na Monitorização de Processos Farmacêuticos da Espectroscopia FT-NIR.” , MSc Thesis (138 pp, Portuguese), Technical University of Lisbon.

Jolliffe, I. (1986) Principal Components Analysis, Springer-Verlag, New York, USA, p.502.

Lopes J.A., J.C., Menezes, J.A. Westerhuis and A.K. Smilde (2002) Biotechnol. Bioeng., **80**, 419-427.

Martens, H. and M. Martens (2001) Multivariate Analysis of Quality : An Introduction., John Wiley & Sons, New-York.

Neves, A., L. Vieira, and J. Menezes (2001) Biotechnol. Bioeng., **72**, 628-633.

Otto, M. (1998) Chemometrics: Statistics and Computer Application in Analytical Chemistry, Wiley-VCH.

Strohl, W. (1997) in W.Strohl (Eds.), Biotechnology of Antibiotics, Marcel Dekker Inc., New York, USA, p. 842.

Welsh, W. , W. Lin, S. Tersigni, E. Collantes, R. Duta and M. Carey (1996) Anal. Chem., **68**, 3473-3482.

Westerhuis, J. and A. Smilde (2001) J. Chemometrics, **15**, 485-493.