Talanta 111 (2013) 28–38

Contents lists available at SciVerse ScienceDirect

Talanta

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

Chemometrics and in-line near infrared spectroscopic monitoring of a biopharmaceutical Chinese hamster ovary cell culture: Prediction of multiple cultivation variables

Matthieu Clavaud, Yves Roggo*, Ralph Von Daeniken, André Liebler, Jan-Oliver Schwabe

F. Hoffmann-La Roche Ltd, Wurmisweg 4303, Kaiseraugst, Switzerland

ARTICLE INFO

Article history:
Received 18 December 2012
Received in revised form
14 March 2013
Accepted 16 March 2013
Available online 26 March 2013

Keywords:
Chinese hamster ovary
Near infrared spectroscopy
Cell culture
Bioprocess monitoring
Process analytical technology
Chemometrics

ABSTRACT

In the present study near infrared (NIR) spectroscopy was used to monitor the cultivation of mammalian Chinese hamster ovary (CHO) cells producing a monoclonal antibody in a fed-batch cell culture process. A temperature shift was applied during the cultivation. The cells were incubated at 37 °C and 33 °C. The Fourier transform near infrared (FT-NIR) multiplex process analyzer spectroscopy was investigated to monitor cultivation variables of the CHO cell culture from 10 independent batches using two channels of the FT-NIR. The measurements were performed on production scale bioreactors of 12,500 L. The cell cultures were analyzed with the spectrometer coupled to a transflection sterilizable fiber optic probe inserted into the bioreactors. Multivariate data analysis (MVDA) employing unsupervised principal component analysis (PCA) and partial least squares regression methods (PLS) were applied. PCA demonstrated that 96% of the observed variability was explained by the process trajectory and the inter-batch variability. PCA was found to be a significant tool in identifying batch homogeneity between lots and in detecting abnormal fermentation runs. Seven different cell culture parameters such as osmolality, glucose concentration, product titer, packed cell volume (PCV), integrated viable packed cell volume (ivPCV), viable cell density (VCD), and integrated viable cell count (iVCC) were monitored inline and predicted by NIR. NIR spectra and reference analytics data were computed using control charts to evaluate the monitoring abilities. Control charts of each media component were under control by NIR spectroscopy. The PLS calibration plots offered accurate predictive capabilities for each media. This paper underlines the capability for inline prediction of multiple cultivation variables during bioprocess monitoring.

1. Introduction

Pharmaceutical and biotechnological companies are encouraged to optimize the efficiency of their established processes. For economic, technical, regulatory and quality requirements the optimization of biotechnological processes is essential. It is under these conditions that the food and drug administration (FDA) encourages pharmaceutical firms to set up innovative tools to better understand their processes. For this purpose the FDA has launched a regulatory framework initiative called process analytical technologies (PAT) [1]. In parallel, the international conference of harmonisation (ICH) has suggested that all critical process parameter which act upon critical quality attribute should be monitored or controlled according to the ICH Q8(R2) "Pharmaceutical Development" guideline [2]. In this way, real time process monitoring becomes interesting, to optimize productivity (by controlling, understanding, troubleshooting bioprocesses) and to ensure product quality. Thus, vibrational spectroscopy has emerged as one of the key analytical tools to be used and traditional wet-chemistry analyses tend to be replaced by fast estimated analytics.

Vibrational spectroscopy covers near infrared (NIR), mid-infrared (MIR) and Raman spectroscopy. These tools are suitable for solid, liquid and biotechnological pharmaceutical forms. In this study NIR was used. NIR can be implemented during pharmaceutical development phases, in production for process monitoring or in quality control laboratories for release and stability analysis [3]. NIR is a fast and non-destructive vibrational spectroscopy based on molecular overtone and combination vibrations. The molecular broad bands seen in the NIR lead to more or less complex spectra. In NIR it is difficult to assign specific features to specific chemical components. NIR delivers chemical and physical information. Thus, multivariate calibration algorithm coupled with statistical methods (e.g. Multivariate data analysis or chemometrics) are

* Correspondence to: F. Hoffmann-La Roche Ltd, Building 65, Room 516, Grenzacherstrasse, 4070 Basel, Switzerland. Tel.: +41 61 68 81 336.
E-mail addresses: yves.roggo@roche.com, y_roggo@hotmail.com (Y. Roggo).

0039-9140/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.talanta.2013.03.044

used to create qualitative and quantitative models by extracting the desired information [4,5].

The objective of this study is to apply NIR spectroscopy on a biopharmaceutical product to monitor in-line Chinese hamster ovary (CHO) cell culture and bring new features for quantification of multiple components. NIR spectroscopy has been used to monitor a large number of bioprocesses and not only related to biopharmaceutical products. The aim here is not to present all papers which successfully used these techniques but to show features already done as the basis of our paper. Recently, the application of at-line NIR transmittance spectroscopy on supernatant samples from CHO cells producing a monoclonal antibody has been published [6]. It was demonstrated that parameters of interest like nutrient or product concentrations can be reliably estimated with at least an accuracy of 15% versus the reference method. Moreover, it was concluded that NIR monitoring carried out at-line was effective in replacing in-process monitoring of some critical process parameters by reference methods. Previously, Strother, *T.* has predicted by NIR CHO protein concentrations in a live and dynamic cell culture [7]. A Thermo® Process analyzer was able to predict protein to 0.5 g L^{-1} on a wide range from 0.16 to 5.23 g L^{-1}. In 2008, NIR was used to monitor multiple components in mammalian cell culture of Human Embryonic cells [8]. The authors were able to develop PLS calibration for cell density, pH, and lactate, glucose, ammonia and glutamine concentration. In 2007, a proposal to build NIR calibrations from multiple bioreactors run in parallel was published [9]. The authors observed a slight model degradation in comparison to the conventional single probe models. In 2003, Arnold and Al published results on the use of in-situ NIR to monitor analytes in a CHO-K1 culture (derived as a subclone from the parental CHO) [10]. They were able to develop models for glucose, lactate, glutamine, and ammonia concentration with interesting standard error of prediction results. In 1998, NIR was used to measure the concentration of human antithrombin III and metabolites in a CHO cell culture supernatant [11]. All these papers show that the monitoring of multiple components on cell culture is possible but with differing results. In our study we will apply NIR on a large production scale.

The calibration strategy developed in this paper is presented in Fig. 1. During production runs, offline data are generated to observe culture conditions. By connecting a NIR spectrometer to the production run, spectral information is produced. The aim was to correlate the spectral information with the offline data in order to propose statistical calibration models for inline monitoring.

2. Materials and methods

2.1. Cell culture conditions and detail

In an aerobic cell culture process, CHO cells producing a monoclonal antibody were cultivated in large scale bioreactors. The culture duration was approximately 15 days. During the cultivation a temperature shift was performed on day three from 37 °C to 33 °C until the end. 10 independent batches were monitored.

2.2. Near infrared acquisition

The 10 cell culture batches were monitored using a Fourier transform near infrared (FT-NIR) multiplex process analyzer (Antaris II MX, Thermo Fisher Scientific®, Madison, USA) qualified to be used under good manufacturing practice (GMP). The spectrometer was programmed to take spectra continuously during the whole culture process. The spectrometer was equipped with a sterilizable transflection probe (Helma®, Type 661.687, Germany) connected to fiber optics. The probe with a pathlength distance of 1 mm was inserted into the bioreactor and sterilized prior to running the cell culture process. 256 scans were averaged per spectrum using a 8 cm^{-1} resolution on a spectral range from 4000 to 10,000 cm^{-1}. A delay of 240 s was applied between 2 spectra. Sample spectra were corrected by using the background transfer function TS/TB. This function addresses the difference between background path (intern standard called TB) and sample path (air spectrum called TS). This sample spectra correction is specified by (S/B)/(TS/TB) where (TS/TB) is the background transfer function, S is the sample spectrum (media spectrum) and B the background spectrum (intern standard) [12].

Each batch was measured independently. Two channels and two probes of the same type were used to acquire spectra from two bioreactors in parallel. The chemometric models were developed by computing 3 spectra for each reference value.

2.3. Wet chemical analysis used as reference

All reference data were sampled twice per day during the whole cell culture. The cell culture performance in each bioreactor was monitored through in-line and off-line reference data measurements.

The Bioprofile 400 (Nova Biomedical Corporation®, Massachusetts, United States) was used to determine the glucose concentration in g L^{-1}. The measurement of osmolality expressed in mOsm was

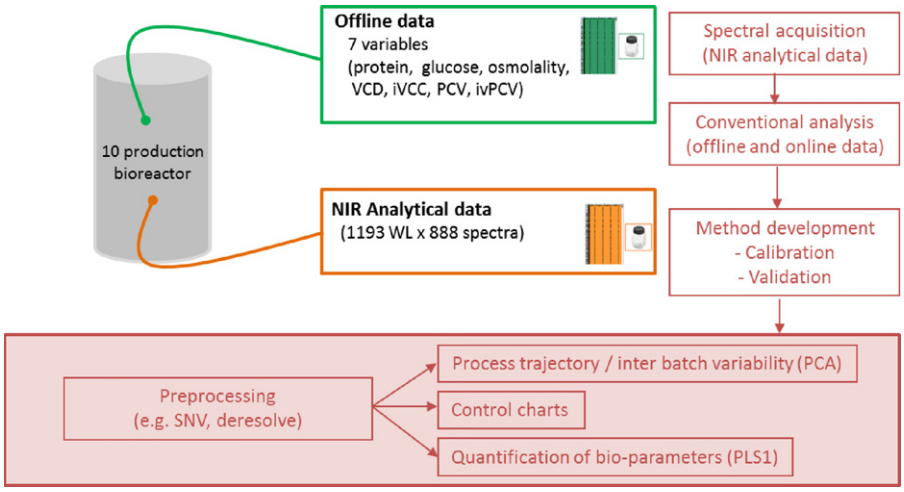

Fig. 1. Calibration strategy applied on the 10 bio-processes runs.

performed using the Micro-osmometer Model 3320 (Advanced Instruments Inc®., Massachusetts, United States).

The protein concentration was determined using the IgG-HPLC-Assay with a Protein A affinity column (POROS-A Applied Biosystems, Applied Biosystems®, California, United States). The data analysis was performed using the Chromeleon Chromatography Management Systems (Dionex Corporation®, California, United States).

A ViCell XR (Beckman–Coulter®, California, United States) was used to determine the viable cell density (VCD) and the viability in percent. This device performs an automated trypan blue exclusion assay. The VCD is expressed in 10 E5 cells mL^{-1}. The packed cell volume (PCV) was measured by centrifugation of a sample in special tubes (6.5 or 10 mL Tube) for 10 min at 830 g. The PCV is expressed in percent. The parameters iVCC (expressed in cells dm L^{-1}) and ivPCV (expressed in mL dL^{-1}) are integrated from measurements of VCD as well as of viability and PCV.

2.4. Calibration strategy and chemometric tools

2.4.1. Calibration strategy

The dataset of ten batches was separated into two independent sets. The calibration set contained 2/3 of the entire dataset and the validation set contained 1/3 of the dataset according to the requirements of the European medicines agency (EMA) guideline for near-infrared spectroscopy [13]. Six batches were used for the calibration set and four batches for the validation set. The batches were randomly distributed among the calibration and the validation sets.

2.4.2. Spectral preprocessing

NIR spectra are subjected to variations due to instrument stability, environmental conditions (e.g. temperature, humidity), characterization of the sample measured (size, forms), which induced broad overlapped peaks making spectral interpretation difficult [14]. Preprocessing techniques were proposed to enhance spectral interpretation [15]. The main interest in applying preprocessing is to display the information of interest relating to the sampling and to highlight the relevant information. This is a necessary step for the removal of background noise, normalization, and other undesirable factors in the matrix.

Several preprocessing techniques have been tested. The preprocessing techniques retained were the standard normal variate (SNV) [20], and the deresolve function.

The SNV was applied to the NIR spectra in order to normalize the baseline and to remove the scattering effects of the cells culture. Each spectrum was mean centered and scaled to unit variance by spectrum using the following calculation:

$$X_{iSNV} = \frac{X_i - \overline{X}}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})/n-1}}$$

where "*Xi*" is the NIR absorbance value at the wavenumber "*i*", "*n*" are the number of wavenumbers, "*X*" is the mean of the NIR absorbances values and "*iSNV*" is the corrected NIR absorbance value at the wavenumber "*i*".

The deresolve function is a tool provided in Unscrambler® version 10.1. It is used to reduce the noise level present in the spectra. A number of channels of 2 were used for convolution.

2.4.3. Principal component analysis (PCA)

Principal component analysis (PCA) forms the basis for multivariate data analysis [16]. This is a statistical technique of changing the variables in a summarized and visual representation. PCA was applied after the spectra had been preprocessed.

This statistical technique uses an orthogonal transformation to convert a set of correlated variables into a set of uncorrelated variables values called principal components (PC). The number of PC has to be less than or equal to the number of variables. This transformation is defined in such way that the first PC has the largest variance. That means the 1st PC (related to a known or unknown variable) explains the maximum percentage of variance observed in the data set matrix.

PCA has been used for visualization of data and for unsupervised monitoring of the cell culture. The data was mean centered and the nonlinear iterative partial least squares (NIPALS) algorithm was used to determine loadings and scores. The NIPALS algorithm [17] calculates one PC at a time and handles missing values. The mean centering ensures that all results will be interpretable in terms of variance around the mean.

Additionally, the Hotelling T^2 test was applied. The Hotelling T^2 is an alternative to plotting sample leverages allowing the detection of abnormal situations. This test is applied during the PCA analysis. A confidence interval of 95% was applied.

2.4.4. Partial least squares statistical (PLS)

Linear regression method used was the partial least squares 1 (PLS-1). PLS is a technique which computes linear combination between NIR spectra and references values [16]. PLS models the reference values and NIR spectra to extract the scores which are most correlated to NIR spectra. Thus, several PLS factors (similar to PC in PCA) were evaluated. Depending on the data matrix, the best PLS model retained will be the one explaining the maximum covariance between NIR spectra and reference values.

The PLS regressions were performed by mean centering data followed by a full cross-validation (FCV) using a NIPALS algorithm. Preprocessing used for each analyte was different.

The retained model were developed with samples of calibration set by using the number of factors with the lowest root mean square error of cross-validation (RMSECV) and the maximum explained variance for cross-validation. The optimum number of terms for the calibration set minimized the over fitting. Due to the complex media fill, the number of factors had to be well selected. Too many factors overfit the model and too few underfit it. This action was done for all of the 7 fermentation variables.

The accuracy of a prediction model was evaluated by a low root mean square error of calibration (RMSEC), a low root mean square error of prediction (RMSEP), a high square correlation coefficient (R^2), regression vectors with a reasonable noise level, and a low bias. The RMSE indicators are compared to Standard Deviation as they explain the dispersion of the results. These criteria were used to judge the performance of each multivariate calibration model.

Modeling accuracy was also evaluated by using the ratio of performance deviation (RPD) and the range error ratio (RER) [18,19]. The RPD evaluates the standard deviation of reference values from the calibration set divided by the RMSEP. The RPD should be greater than 3 and then be discussed. The second additional parameter called RER was the ratio of concentration range of reference values used for calibration set divided by RMSEP. A RER greater than 10 indicates that the model is able to predict the required concentration with an accuracy of at least one tenth of the range.

2.5. Software

The NIR spectra acquisition was obtained using Results NIR Suite 3 software (Thermo Fisher Scientific®, Madison, USA). NIR Spectra were exported and computed with Unscrambler® version 10.1 (32-bit) (CAMO® Software AS, Oslo, Norway). This is a

graphical software for process modeling and multivariate data analysis to extract information within complex data.

3. Results and discussion

3.1. Explorative analysis

3.1.1. Near infrared spectral selection

Fig. 2 plots raw NIR spectra showing variability observed in the cell culture of batch B81 from 4000 to 10,000 cm^{-1}. This figure highlights two important elements. The strong water band at 5130 cm^{-1} (combination OH region due to water in the medium) shows a peak saturation commonly observed for this kind of study. Additionally, the noise of optic fibers over 4300 cm^{-1} is well seen. Consequently, the water band region and the optic fiber noise region were removed from the analysis. At these wavelengths the absorbance exceeds the linear range of the detector/probe combination and the peaks are saturated.

Therefore, the region of interest for modeling was between 5400 to 10,000 cm^{-1} in the first and second overtone regions in order to focus on the water band at 7000 cm^{-1}. At this wavelength region, the biomass absorption was acquired.

By focusing on the data set of spectra, the spectral quality acquisition decreases in the last days of the bioprocesses. Due to the large amount of cells and proteins in the cell cultures the noise in the spectra increases.

Fig. 3 displays the spectral evolution during the cell culture of batch B81 over the two-week period in the region of interest (5400–10,000 cm^{-1}). During the first 64 h of the bioprocess the cell culture was cultivated at 37 °C. This first step is well observed by an exponential cell growth. During the second incubation phase a temperature of 33 °C was applied. Cells cultivated at this temperature deviate from the original temperature condition. The cells experience a growth lag observed when the temperature changed. Thus, the two temperature phases of the bioprocess are well observed.

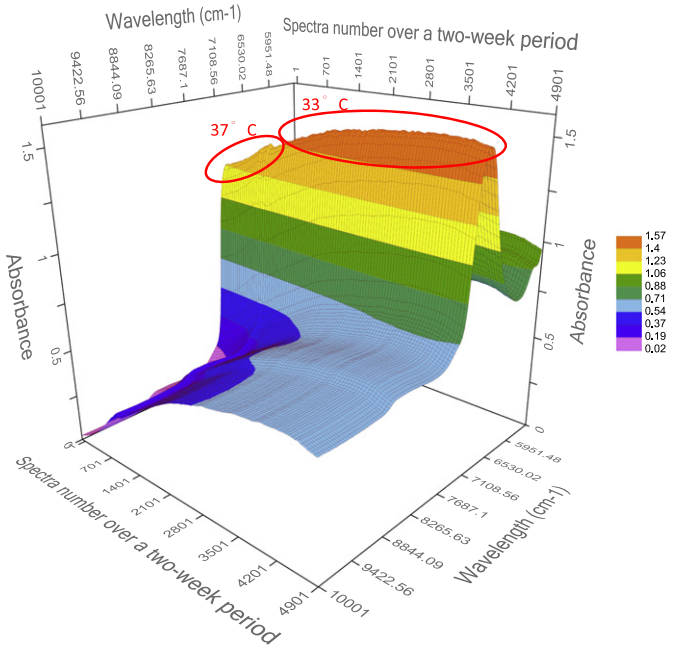

Fig. 3. NIR raw spectra of batch B81 from 5400 to 10,000 cm^{-1} showing the cell culture process monitoring over the time during a two-week period. The legend color represents the absorbance units. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

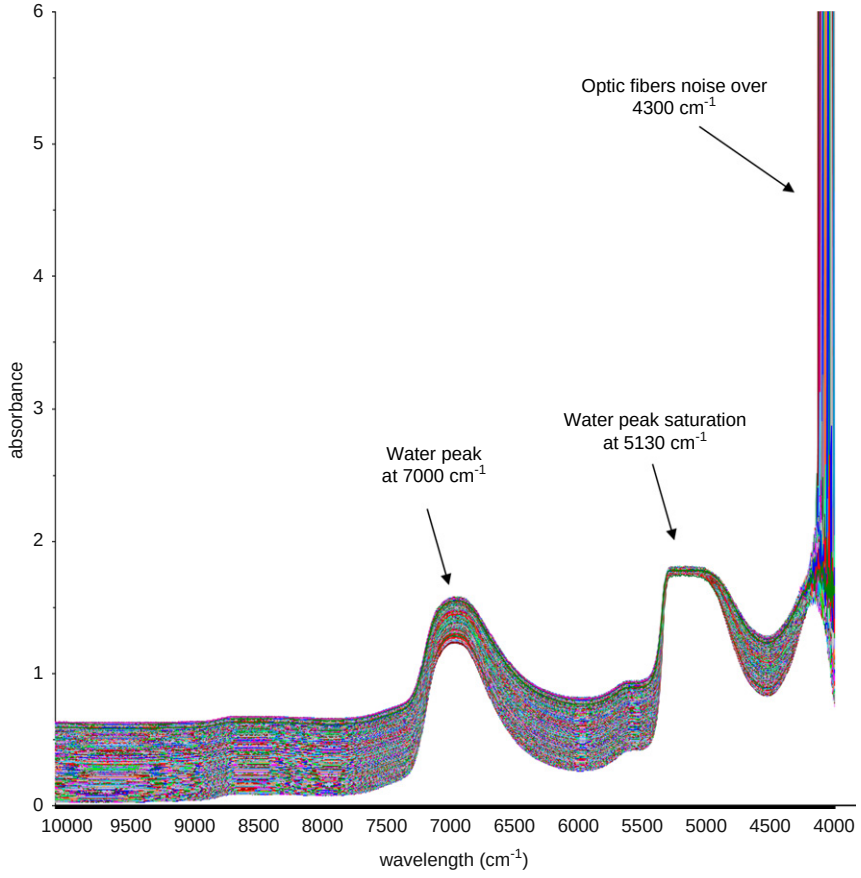

Fig. 2. Representative raw NIR spectra showing variability observed in the cell culture of batch B81 from 4000 to 10,000 cm^{-1}.

The region of interest was preprocessed by normalizing the baseline with a SNV. The preprocessing removes the scattering effects of the cells culture and highlights chemical absorbance.

3.1.2. Process trajectory and inter-batch variability

PCA model on the preprocessed SNV spectra from 5400 to 10,000 cm^{-1} was applied and gave results illustrated in Fig. 4. The mathematical data set matrix consisted of 1193 variables (wavelength) and 888 spectra (spectra of the dataset of 10 batches). The PCA was performed by mean centering data followed by a FCV using a NIPALS algorithm.

The plot shows that the 1st PC axis explained at 78% the variance of the process trajectory and the 2nd PC axis at 18% the inter-batch variability variance. In the 1st PC axis the process trajectory focuses on the real time culture run (batches follow the same trending over the time) and on the temperature changing. The two temperature cultivation phases are easily distinguished. In the 2nd PC axis the inter-batch variability indicates the batch profile situations and if there are no outliers. The applied 95% confidence Hotelling T^2 ellipse do not highlight any abnormal situations.

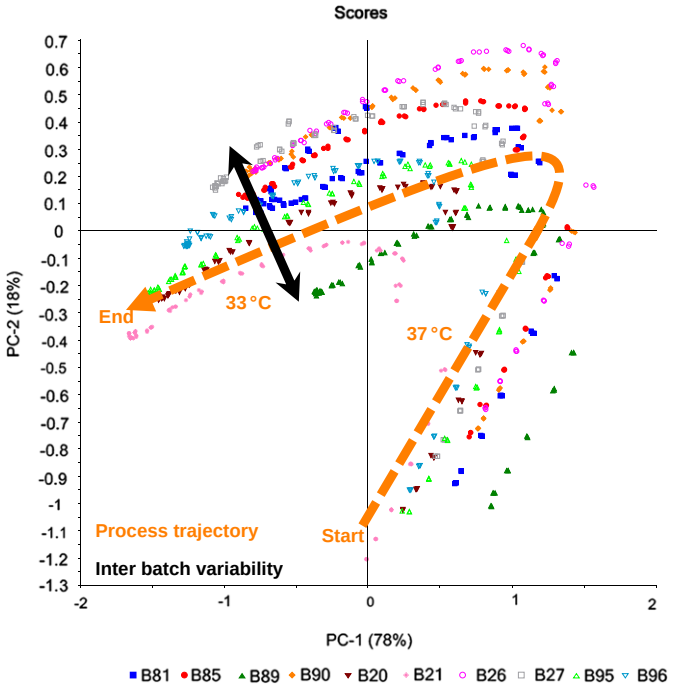

Fig. 4. PCA on the preprocessed SNV spectra from 5400 to 10,000 cm^{-1} (1193 variables and 888 spectra) showing process trajectory and inter batch variability. The 95% confidence Hotelling T2 ellipse covers the majority of all spectra.

Moreover it was observed that the process trajectory of batches B81 and B27 follows three phases: the two temperature phases and a third one. This third phase appeared in the last two days. Moreover, these runs were longer than all the others. Thus, the third phase observed is not attributed to a temperature phase but to the starting death phase of the cell culture. This demonstrates that the temperature phases are correlated with cell growth. The changes observed are attributed to the viability of cells due to the long cell culture. This death phase is expressed by noisy spectra related to a high concentration of both protein and cells. Low viability at the end of cultivation is more a sign of a bad culture performance and therefore not desired.

These two batches were not removed from the study because even if they are atypical they are not considered as outliers. In real time monitoring there is a possibility to get batches with different duration times.

Finally 96% of the variance observed on the dataset of the ten batches was explained by the culture process itself including mainly the temperature, the time duration and parameters linked to cell growth. The influence of the probes, channels, fiber optics are directly linked to the culture process and have no significant observable impact. In 96% of the variance observed these were included. The PCA is a significant tool for cell culture process understanding especially for process trajectory and inter-batch variability but also for batch homogeneity and identification of abnormal fermentation runs.

Additionally to PCA, a control chart analysis was applied to provide another data visualization by computing reference values against NIR spectra.

3.1.3. Control charts—process monitoring

In this chapter, control charts are intended to determine whether variables of the cell culture process are in a state of control. For each variable, a control chart representing the ten batches was created. If the analysis of the control chart indicates that a variable is under control then it means NIR is suitable for the analyzed variable. The control charts are used to evaluate the performance of the calibration model of the intended variable. The control chart is seen as a disciplined approach which enables decisions regarding control of the process by NIR. The control

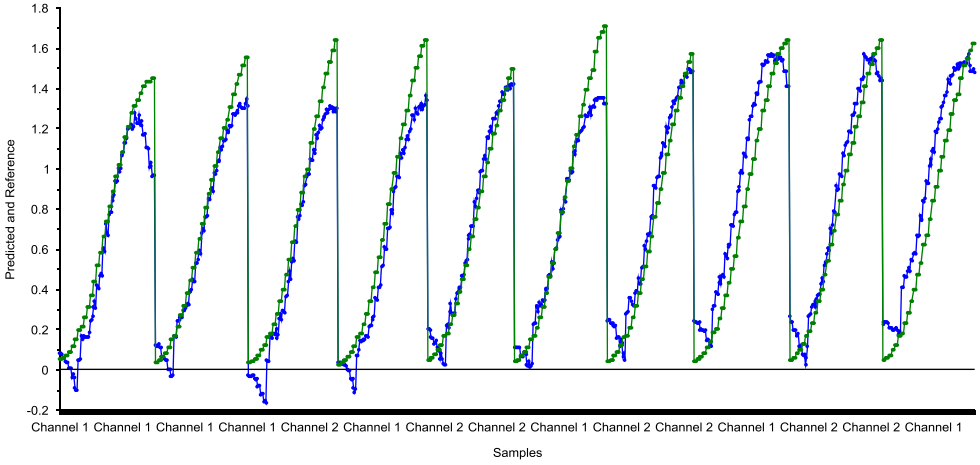

Fig. 5. Control chart of predicted NIR protein values (in blue) superimposed over lab data (in green) after SNV and deresolve on 10 batches (4 factors used). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

charts give additional information on the processes as well as abnormal NIR spectra (e.g. air bubbles or accumulation of biomass in the pathlength of the probe) or abnormal reference data.

The 10 independent batches acquired during the whole cell culture process were displayed in control chart plots. For each control chart, NIR spectra and references values were computed over time. The accumulation of biomass, related cell growth variability and other components were plotted. Figs. 5 and 6 display respectively the main control chart for titer and iVCC.

For the titer and iVCC control chart few differences between reference analytics data and NIR data at the end of the processes were observed. It showed that there was a lack of precision in the reference analytics at the end of the process. Two hypothesis were investigated. This lack of precision could come from the "noisy" spectra observed in the spectral plot. Indeed, spectra begin to be altered which induce that spectra cannot predict values over a certain limit. The probe is clogged by the medium which is more and more dense at the end of the processes. The second hypothesis comes from the accuracy of reference analytics. At a high concentration of analyte the reference method is not as precise as in a low limit.

3.2. Predictive abilities

Previous results showed that the 10 independent batches were well distributed in terms of process trajectory and inter batch variability. The control charts have demonstrated that NIR is suitable to monitor in-line product titer and iVCC. The next step was to evaluate the predictive abilities of each retained variable. Table 1 and Figs. 7–13 summarize all results discussed in the following chapters. The following calibration models used a SNV preprocessing with a deresolve.

3.2.1. Quantification of protein content

The best modeling result was acquired with a goodness for R^2 of 0.95 using 4 components.

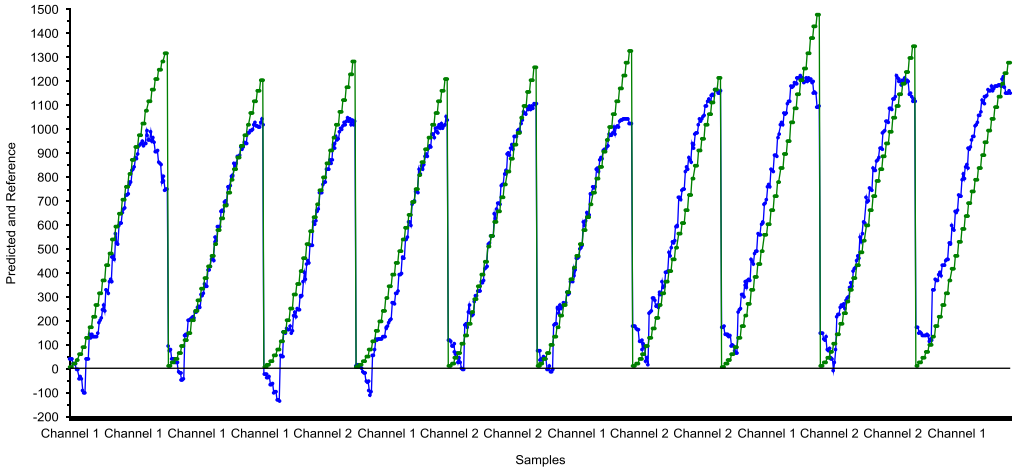

Fig. 6. Control chart of predicted NIR iVCC values (in blue) superimposed over lab data (in green) after SNV and deresolve on 10 batches (4 factors used). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
PLS-1 Statistical results and calibration, prediction, cross-validation performances for retained fermentation variables. SNV = standard normal variate; DER2 = deresolve with 2 numbers of channels used for convolution.

Fermentation variables		**Protein (g L^{-1})**	**iVCC (10^5 cells dmL^{-1})**		**Glucose (g L^{-1})**	**PCV (%)**	**VCD (10^5 cells mL^{-1})**	**ivPCV (mL dL^{-1})**	**Osmolality (mOsm)**
			PLS1	**PLS37**					
Preprocessing		SNV + DER2							
Calibration set	*N*	525	507	90	525	525	525	507	525
	Slope	0.95	0.95	0.97	0.80	0.91	0.89	0.93	0.78
	offset	0.04	31.84	1.45	1.55	0.28	9.56	12.18	74.80
	Factors	4	4	4	4	4	5	4	4
	R^2	0.95	0.95	0.97	0.80	0.91	0.89	0.93	0.78
	RMSEC	0.12	95.35	5.54	1.72	0.50	9.60	37.85	15.45
	Yexpl val	94.69	94.54	96.82	80.04	94.66	89.38	38.28	78.44
	RMSECV	0.12	96.33	6.29	1.74	0.50	9.82	38.45	15.61
Validation set	*N*	363	351	60	363	363	363	351	363
	R^2	0.86	0.82	0.47	0.64	0.89	0.74	0.77	0.62
	RMSEP (% of the calibration range)	0.20 (11.98%)	176.86 (13.19%)	20.96 (21.46%)	2.22 (14.45%)	0.54 (10.84%)	15.27 (11.56%)	68.25 (14.96%)	22.24 (17.65%)
	RPD (> 3)	2.65	2.31	1.49	1.73	3.11	1.93	2.04	1.50
	RER (> 10)	8.35	7.58	4.66	6.92	9.22	8.65	6.69	5.67
	RMSEP / RMSECV	*1.67*	*1.84*	*3.33*	*1.28*	*1.08*	*1.55*	*1.78*	*1.42*
Calibration set	Average	0.73	583.63	45.70	7.75	3.24	90.05	164.24	347.51
Reference data	Min-Max	0.04–1.71	5.43–1345.95	5.43–103.10	0.50–15.86	0.22–5.20	12.98–145.12	0.84–457.12	288–414
	Stdev	0.53	408.61	31.23	3.85	1.68	29.49	139.12	33.31

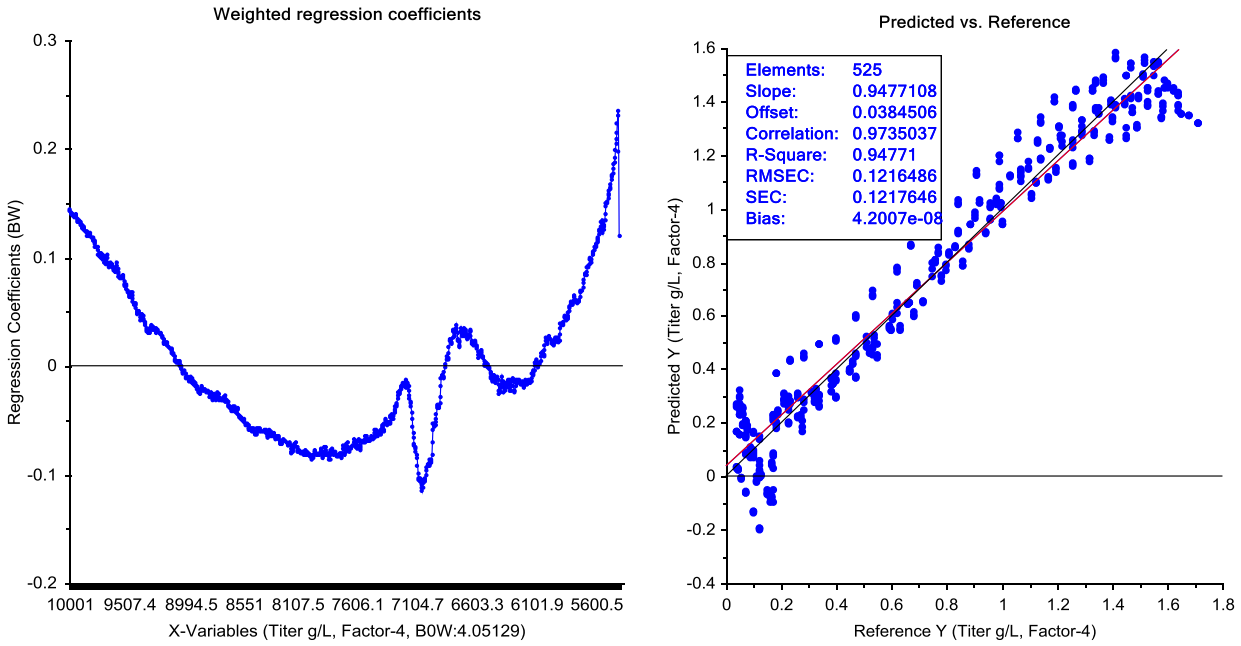

Fig. 7. Protein PLS-1 calibration, prediction statistics and regression vectors plots which compare calculated values (NIR) versus wet-chemistry reference values. A low prediction bias is expressed by a tilting regression line (in red) compared to the target line (in black). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

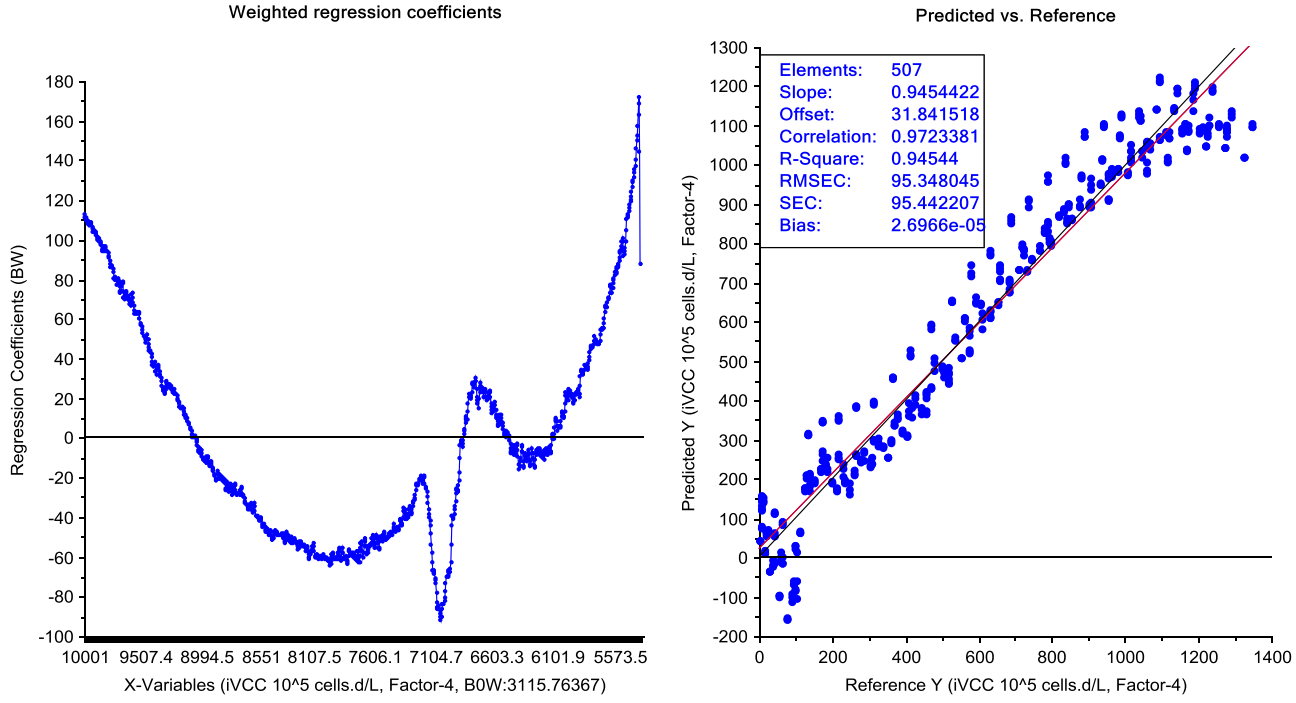

Fig. 8. iVCC PLS-1 calibration, prediction statistics and regression vectors plots which compare calculated values (NIR) versus wet-chemistry reference values. A low prediction bias is expressed by a tilting regression line (in red) compared to the target line (in black). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 7 which shows the PLS calibration regression plot reveals excellent predictive capabilities within the range of 0.04–1.71 g L^{-1}. A slope of 0.95 on the calibration plot is expressed by a tilting regression line compared to the target line. The related errors RMSEC of 0.12, RMSECV of 0.12 and RMSEP of 0.20 g L^{-1} indicate that the protein concentration can be predicted to 0.20 g L^{-1} or less. The RMSEP error represents 12% of the calibration range. The RPD and RER values of 2.65 and 8.35 are good enough for prediction. As previously shown in the control charts the protein content can be monitored and accurately predicted by NIR.

3.2.2. *Quantification of glucose concentration and osmolality content*

The calibration model for glucose and osmolality used 4 factors. PLS regression reveals good predictive capabilities within the range of 0.50–15.86 g L^{-1}. The related errors RMSEC of 1.72,

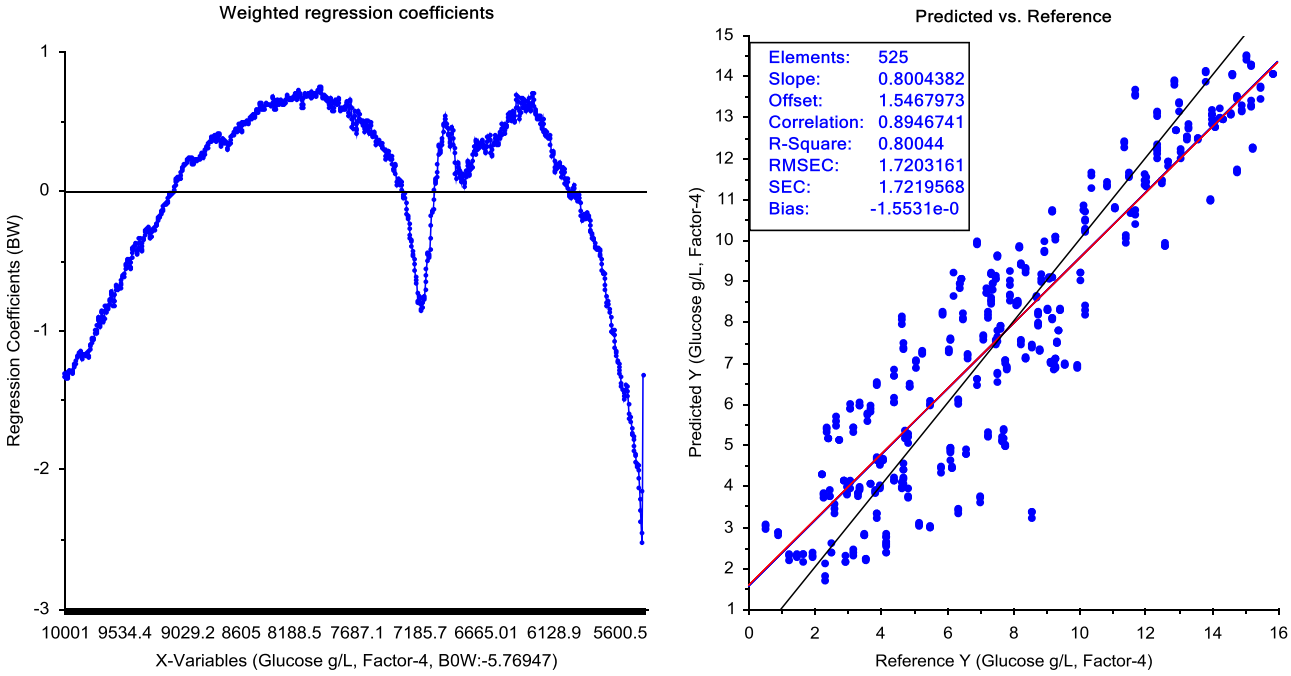

Fig. 9. Glucose PLS-1 calibration, prediction statistics and regression vectors plots which compare calculated values (NIR) versus wet-chemistry reference values. A low prediction bias is expressed by a tilting regression line (in red) compared to the target line (in black). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

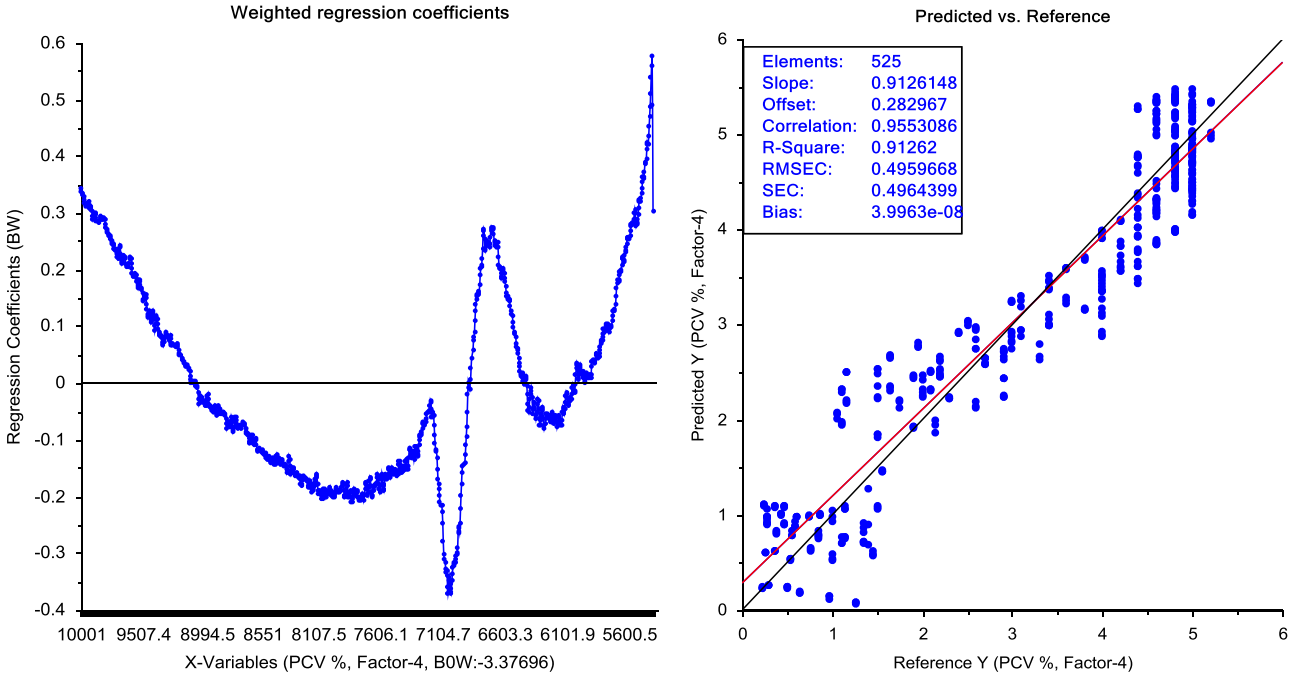

Fig. 10. PCV PLS-1 calibration, prediction statistics and regression vectors plots which compare calculated values (NIR) versus wet-chemistry reference values. A low prediction bias is expressed by a tilting regression line (in red) compared to the target line (in black). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

RMSECV of 1.74 and RMSEP of 2.22 indicate that the glucose concentration can be accurately predicted.

Osmolality also reveals good predictive capabilities within the range of 288–414 mOsm. The related errors are RMSEC of 15.45, RMSECV of 15.61 and RMSEP of 22.24 mOsm. RPD and RER values for glucose concentration (1.73 and 6.92 respectively) and osmolality (1.50 and 5.67) enable prediction abilities.

3.2.3. *Quantification of cell density (PCV, VCD)*

The calibration model for PCV and VCD used a SNV with 4 and 5 factors respectively. PLS regression of PCV revealing interesting predictive capabilities within the range 0.22–5.20% with RMSEC of 0.50, RMSECV of 0.50 and RMSEP of 0.54. Moreover, VCD and ivPCV also reveal interesting predictive capabilities within respectively the range 12.98–145.12 10 E5 cells mL^{-1} and

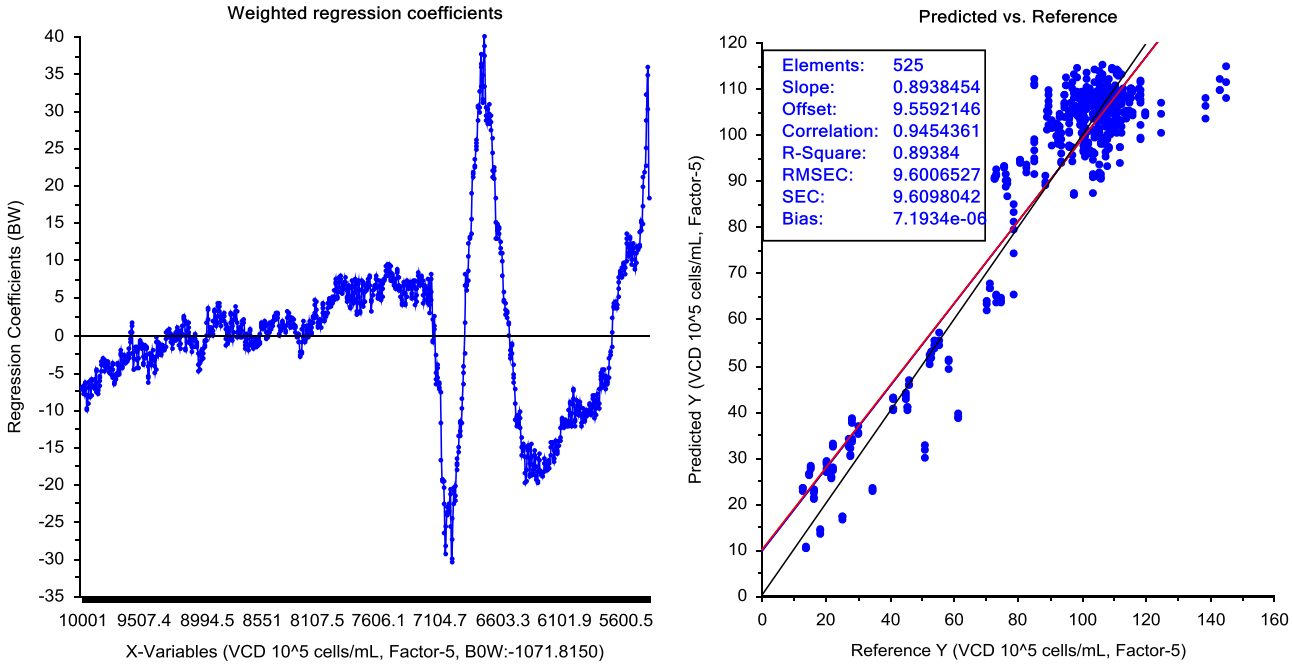

Fig. 11. VCD PLS-1 calibration, prediction statistics and regression vectors plots which compare calculated values (NIR) versus wet-chemistry reference values. A low prediction bias is expressed by a tilting regression line (in red) compared to the target line (in black). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

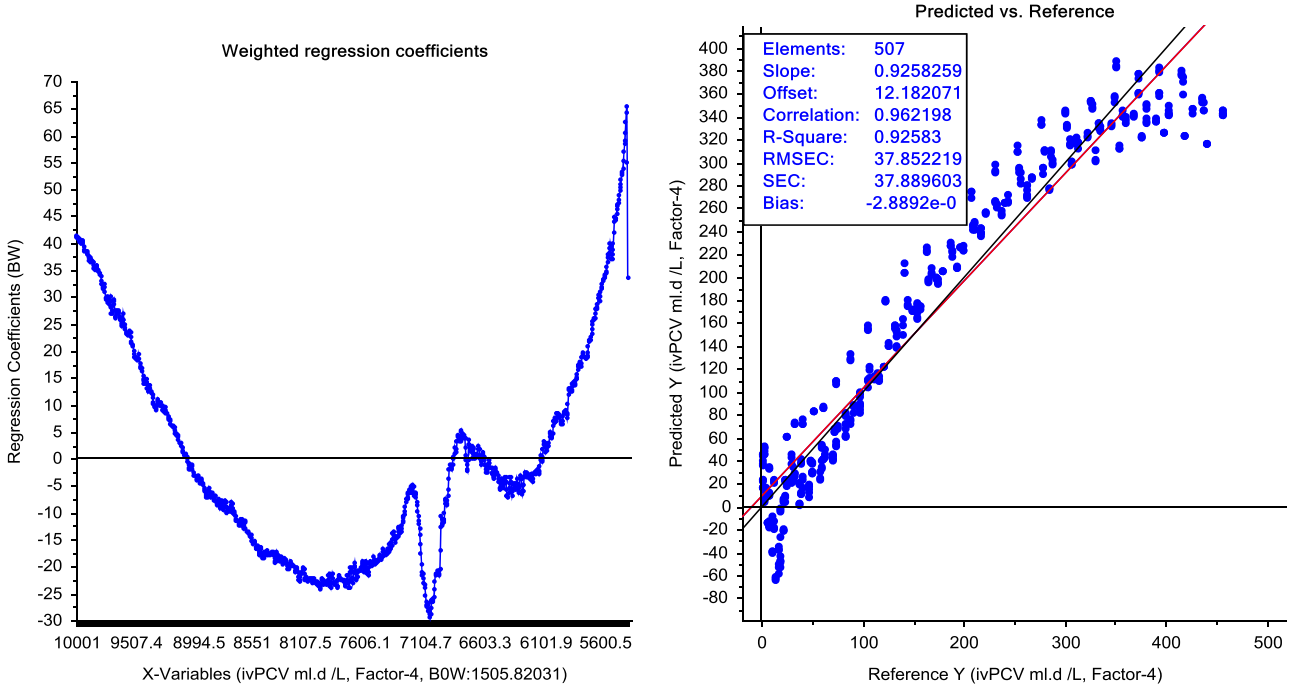

Fig. 12. ivPCV PLS-1 calibration, prediction statistics and regression vectors plots which compare calculated values (NIR) versus wet-chemistry reference values. A low prediction bias is expressed by a tilting regression line (in red) compared to the target line (in black). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

0.84–457.12 mL dL^{-1}. The RMSEP were respectively 15.27 cells mL^{-1} and 68.25 mL dL^{-1}.

3.2.4. *Quantification of iVCC coupled with temperature variable*

The calibration model used 4 factors. This variable was modeled with a R^2 of 0.95, RMSEC of 95.35, RMSECV of 96.33 and RMSEP of 176.86.

Two tendencies have been visually observed on the calibration and validation plots of Fig. 8. The first tendency appeared during the cultivation phase at 37 °C. Samples are linear expressed by a tilting evolution. The second tendency appeared by the passage from 37 °C to 33 °C till the end of the process. The slowing down of cell growth imposed by temperature changes has an impact on the viability of cells and consequently on the PLS model.

The iVCC variable has been evaluated by creating one PLS model (PLS37 in Table 1) representing only the temperature phase at 37 °C of the bioprocess.

The results show a gain for RMSE in the beginning of the cell culture (10 times less compared to the global model). The model

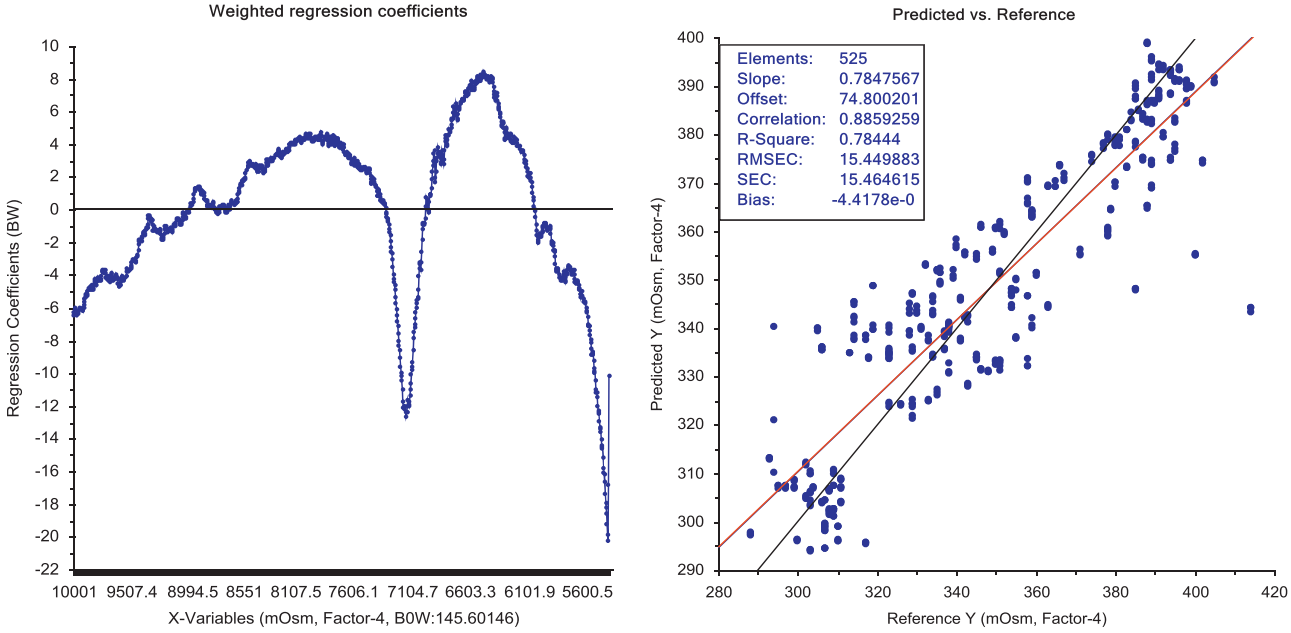

Fig. 13. Osmolality PLS-1 calibration, prediction statistics and regression vectors plots which compare calculated values (NIR) versus wet-chemistry reference values. A low prediction bias is expressed by a tilting regression line (in red) compared to the target line (in black). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

reveals accurate predictive capabilities within the range 5.43–103.10 10^5 cells dmL^{-1} with a R^2 of 0.97 on the calibration set. The related errors RMSEC of 5.54, RMSECV of 6.29 and RMSEP of 20.96 indicate better predicted abilities compared to the global model. The validation samples used to validate this calibration have been predicted by the global model. A RMSEP of 144.20 has been determined indicating a relative error of 85.47% between the two RMSEP's. This indicates that the PLS model at 37 °C offers better predictive abilities in the first hours. This improvement of precision can determine the future fermentation variables if an abnormal situation is detected at this step. This parameter is determinant. The measure of viable cell density is an important criteria to evaluate the future batch profile and related fermentation variables.

3.3. Discussion

The physiological and metabolic changes imposed in the last third of the processes (e.g. death of cells, increase of protein titer) influences the accuracy of the prediction. A difference of 1 $g\,L^{-1}$ is observed between target line and predicted line of the protein variable. This observation is considered as a "normal" situation. On routine analysis reference data of batches are monitored for the full two week period by applying a systematic error of 3 times the standard deviation. Accordingly, the possibility to obtain reference data with a large variability has an influence on PLS models depending on the reference method. For cell count the increased number of cell debris and aggregates makes the analysis difficult. Also manual dilution of samples increase the error rate of the result. The tolerance gap and the possibility of obtaining "abnormal" reference data at the end are significantly increased. Additionally, there is a significant possibility of getting noisy spectra at the end of the process. This is due to high concentration in protein and cell materials. Therefore, all PLS models lose in prediction at the end of processes. Nevertheless, the PLS models appeared rather well modeled.

From a general point of view the absorption peaks of the complex media fill are overlapped. That means the variance observed on an analyte influences the prediction of all the others. The best predictive model was attributed to the product titer content. Thus, this variable does include all variances from each of the other analytes. The predictive tendency observed provides an idea for the other variables. As the protein content appeared to be correlated by VCD, PCV, ivPCV and iVCC, the PLS models observed for these variables were accurately modeled.

4. Conclusion

Obviously, in a PAT scope application modeling should be subjected to ongoing evaluation and optimization in order to integrate variability over the years by adding suitable batches or removing abnormal values. The main difficulty in monitoring life cell culture online is to be able to develop and validate accurate models including cells and NIR instrumentation variability. These strict conditions are imposed to characterize the homogeneity of batches by a PCA, to include the maximum of variability as well as time duration, cell line stain, type of vessel or atypical runs (but within the specification). PCA allows us to obtain an overview of batches and explains potential future results in regression models as well as protein content and cell growth. This study has demonstrated that it is possible to simultaneously quantifying protein content, iVCC, Glucose concentration, PCV, VCD, ivPCV and osmolality on a wide concentration range with a high degree of accuracy.

The use of NIR for monitoring large scale cell culture processes for manufacturing of biopharmaceuticals is possible. This paper underlines the advantages of chemometric tools to predict inline seven independent fermentation variables occurred during a bioprocess run. This encouraging approach demonstrates that in process monitoring NIR carried out in-line can be used in replacing/reducing of reference methods or as an additional analytical tool to process monitoring. The possibility of surveying processes in real time allows the acquisition of more data than conventional techniques and also allow the rapid detection of process deviations.

References

[1] FDA, Guidance for Industry PAT—A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance. Available

from: ⟨http://www.fda.gov/downloads/Drugs/…/Guidances/ucm070305.pdf,2004⟩. (accessed 29.10.12).
[2] International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human use. ICH Harmonised Tripartite Guideline Pharmaceutical Development Q8(R2). Available from: ⟨http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q8_R1/Step4/Q8_R2_Guideline.pdf,August2009⟩. (accessed 14.03.13).
[3] Y. Roggo, P. Chalus, L. Maurer, C. Lema-Martinez, A. Edmond, N. Jent, J. Pharm. Biomed. Anal. 44 (2007) 683–700.
[4] H. Mark, J. Workman Jr., Statistics in Spectroscopy, 2nd Edition, Elsevier, Amsterdam, 2003.
[5] H. Martens, T. Naes, J. Chemometrics 4 (1992) 217–225.
[6] C. Hakemeyer, U. Strauss, S. Werz, G.E. Joseb, F. Folque, J.C. Menezes, Talanta 90 (2012) 12–21.
[7] T. Strother, Spectrosc. Europe 23 (2011) 18–21.
[8] C. Card, B. Hunsaker, T. Smith, J. Hirsch, BioProcess Int. 6 (2008) 58–67.
[9] P. Roychoudhury, R. O'Kennedy, B. McNeil, L.M. Harvey, Anal. Chim. Acta 590 (2007) 110–117.
[10] S.A. Arnold, J. Crowley, N. Woods, L.M. Harvey, B. McNeil, Biotechnol. Bioeng. 84 (2009) 13–19.
[11] S. Harthun, K. Matischak, P. Friedl, Biotechnol. Tech. 12 (1998) 393–398.
[12] J. Hirsch, Thermo Fisher Scientific Madison, WI, USA, 2008.
[13] Note for guidance on the Use of NIR spectroscopy by Pharmaceutical industry and the data requirements for new submissions and variations. EMEA/CVMP/961/01. Available from: ⟨http://www.emea.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003331.pdf,2003⟩. (accessed 14.03.13).
[14] A. Candolfi, R. De Maesschalck, D.L. Massart, P.A. Hailey, A.C.E. Harrington, J. Pharm. Biomed. Anal. 19 (1999) 923–935.
[15] A. Rinnan, F.V.D. Berg, S. Balling Engelsen, TrAC Trends Anal Chem 28 (2009) 1201–1222.
[16] T. Rajalahti, O.M. Kvalheim, Int. J. Pharm. 417 (2011) 280–290.
[17] M. Yoshikatsu, I. Toshiaki, K. Hiroyuki, S. Shin-Ichi, J. Chemometrics 4 (1990) 97–100.
[18] P.C. Williams, D.C. Sobering, in: A.M.C. Davies, P.C. Williams (Eds.), How do we do it: a Brief Summary of the Methods we use in Developing Near Infrared Calibrations—Near Infrared Spectroscopy: The Future Waves, NIR Publications, Chichester, 1996, pp. 185–188.
[19] T. Fearn, NIR News 13 (2002) 12–14.
[20] R.J. Barnes, M.S. Dhanoa, S.J. Lister, Appl. Spectrosc. 43 (1989) 772–777.