A chemometric approach to minimise the diffusive effect of the biomass when using near infrared spectroscopic measurements for the monitoring of bioprocesses

M. Lesteur, E. Latrille, J.M. Roger, Véronique Bellon Maurel, C. Gonzalez, G. Junqua, J.P. Steyer

Author-produced version
11th International Conference on Chemometrics in Analytical Chemistry. Montpellier France 30 Juin - 4 Juillet 2008.

A chemometric approach to minimise the diffusive effect of the biomass when using near infrared spectroscopic measurements for the monitoring of bioprocesses.

Lesteur, M.[1], Latrille, E[1]., Roger, J-M[2]., Bellon-Maurel, V[2]., Gonzalez, C[3]., Junqua, G[3]., Steyer, J-P[1].

[1] INRA, UR50, Laboratoire de Biotechnologie de l'Environnement F-11000 Narbonne (*email : lesteur@supagro.inra.fr*).
[2]CEMAGREF Unite Mixte de Recherche Cemagref - ENSAM – CIRAD, Information et Technologie pour les Agro-Procedes, Cemagref - BP 5095 34196 MONTPELLIER Cedex 1 France 3
[3]Ecole des mines d'Ales – EMA, centre LGEI, équipe Biodiagnostics et métrologie – 6, avenue de Clavières 30319 ALES Cedex France.

Keywords : PLS calibration, SNV, biomass scattering effect, bio-hydrogen.

Abstract

This article focuses the problem of the light absorption and diffusion of the bacterial concentration on the calibration of near infrared (NIR) spectroscopic methods for the monitoring of bioprocesses. The biomass or bacterial concentration produces multiplicative and additive effects on the NIR spectral measurements. The process monitored is a batch anaerobic production of bio-hydrogen from molasses in a small-scale reactor. This monitoring consisted on the prediction of the main compounds of the process i.e. the concentrations of acetate, butyrate, sucrose and biomass. The system used was a near-infrared spectrometer scanning the range 1000-2500 nm with a 3 nm resolution, connected to the bioreactor by an in-situ optical fibre device. The work shows that the measurements were heavily influenced by the presence of biomass in the liquid medium. Dedicated pre-treatments like second derivative or Standard Normal Deviate (SNV) and the selection of wavelengths were used in order to improve the prediction accuracy of the compound concentrations. The Partial Least Square (PLS) regression associated with SNV pre-treatment of the transmittance values gave the best results leading to a prediction error of 0.56, 0.12 and 2.5 $g.L^{-1}$ for butyrate, acetate and sucrose concentrations, respectively for compound concentrations ranges comprised between 0 and 6.65, 1.7 and 15.7 $g.L^{-1}$, respectively. The biomass concentration had a prediction error of 75 NTU (Nephelometric Turbidity Unit) without pre-treatment for a biomass range comprised between 0 and 680 NTU.

1 Introduction

Biological processes are increasingly used for the treatment of wastewaters produced by farms and food industries. These processes require frequent, fast and reliable analyses to improve their control [1]. These measurements are generally performed manually are long time consuming. Different types of measurements are performed to monitor the process. Measurements of COD (Chemical Oxygen Demand) and TOC (Total Organic Carbon) are commonly carried out. The reference method for COD measurement is a total chemical oxidation of the sample and requires at least 2 hours. Recent studies show the possibility to use UV spectroscopic methods [2-3] to determine COD values from the spectral measurement of the sample. Moreover, to get a more accurate control of the process, the concentrations of the molecules of interest must be determinated. The use of mid-infrared spectroscopy has shown its interest but required an ultrafiltration device for the physical treatment of the sample in order to remove biomass and suspended matter [4]. The monitoring of biological process by near infrared spectroscopy is more and more studied [5-7] according to its advantages: fast and non-destructive measurements, no need of preparation of samples and method cheaper than reference methods such as titrimetry or gas chromatography analysis.

In our study, the monitored process used molasses as raw material, which contains mainly sucrose. The fermentation of sucrose by Clostridii bacteria leads to the production of hydrogen and volatile fatty acids

like acetate and butyrate. The aim of this study is the use of chemometric methods to remove the diffusive effect of biomass suspension for the calibration of a prediction model of the biochemical compound concentrations.

2 Theory

Biomass is composed of particules of 1 to 10 µm of diameter and has an effect on spectral measurement due to its scattering effect of the electromagnetic radiations. The measurement of the molecule absorbance is based on the Beer Lambert's law. In a complex media, the absorbance at each wavelength is the sum of the absorbance of each compound. Biomass, du to the induced light diffusion, influences the value of absorbance two ways: a multiplicative effect mainly caused by the rise of the optical path length and an additive effect that leads to the decrease of the intensity of the incident radiation reaching the photosensitive cell. Both effects can be lowered by using dedicated mathematical pre-treatments: the second derivative to reduce the additive effects and SNV correction to reduce the multiplicative ones [8-9].

The Beer's law equation gives:

$$A = \varepsilon(\lambda) \cdot \ell \cdot C \cdot k + a\lambda + b \qquad (1)$$

With:

- $A(\lambda)$: absorbance.
- $\varepsilon(\lambda)$: molar absorptivities.
- l : the path length, here a constant.
- C : concentrations.
- k : multiplicative effect, which acts at the same level for each wavelength.
- $a\lambda+b$: additive effect depending on the wavelength.

If the multiplicative effect is negligible compared to the additive effect, equation (1) becomes:

$$A = \varepsilon(\lambda) \cdot \ell \cdot C + a\lambda + b = A_{mol} + A_b \qquad (2)$$

With:

- $A_{mol}(\lambda)$: absorbance dues to all molecular compounds.
- $A_b(\lambda)$: absorbance dues to biomass.

Transforming absorbance in transmittance leads to equation (3):

$$T = T_{mol} \cdot T_b \qquad (3)$$

With:

- T: transmittance.
- T_{mol}: transmittance dues to the molecular compounds,
- T_b: transmittance dues to the biomass.

By this way, applying the SNV correction to the transmittance spectrum should reduce the biomass effect.

3 Material and methods

The substrate used in the present study was molasses, a residue from industrial processing of sugar beet. Molasses contain mainly sucrose. The process used was a 2 L reactor with a working volume of 1.27 L. The agitation was maintained at 300 rpm. The pH was kept constant at 5.5 with an addition of NaOH (2

mol.L^{-1}). The reactor temperature was controlled at 37 ° C. Four different batch cultures were carried out with different initial concentrations of sucrose, acetate, butyrate and biomass by centrifuging and diluting the media between the batchs in order to keep uncorrelated these initial concentrations. An additional specific experiment was performed and consisted in addition of know biomass concentration to the media while keeping constant the concentration of the other biochemical compounds. Samples were taken during the fermentations and analysed by Gas Chromatography (GC) to determine acetate and butyrate concentrations, by High Performance Liquid Chromatography (HPLC) for sucrose concentration and by turbidity measurement at 875 nm for the biomass determination.
Spectral measurements were carried out in parallel with a near infrared spectrometer (NIRVIS 91,BUCHI, Switzerland), equipped with an optic fibre transmission probe and with an acquisition software (Nircal v4.21). The spectral range used covered wavelengths from 1000 to 2500 nm with a resolution of 3 nm. The Scilab (v4.1) and Unscrambler (v9.6) softwares were used to determine the model calibration with PLS method associated with cross-validation. Pre-treatments SNV and Savitzky-Golay second derivative were applied and compared.

4 Results and discussion

The results of sample measurements are shown in Figure 1 for the 4 batch cultures and during the biomass addition phase. During the cultures, sucrose was consumed and acetate, butyrate and biomass were produced. During the biomass addition phase, the biomass concentration was increased and the other compound concentrations were maintained constant. This phase was conducted to see the effect of biomass on the spectral measurement. At the beginning of this phase, 0 NTU was present in the reactor because biomass has been harvested by centrifugation. The supernatant was put in the reactor and biomass added step by step.

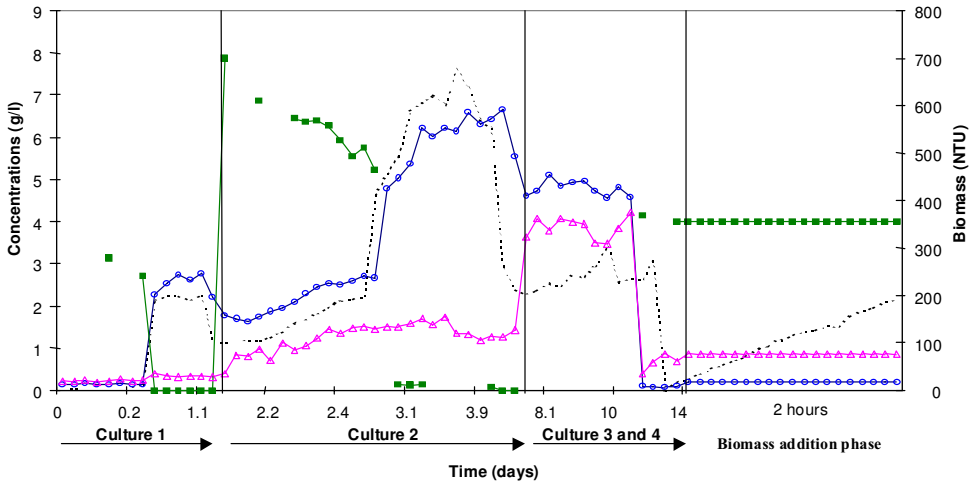

Figure 1 – Evolution of compound concentrations and biomass during the cultures and the biomass addition phase. -- : Biomass ; o : Butyrate ; Δ : Acetate ; ■ : Sucrose.

The transmittance spectra used for the model calibration are presented in Figure 2. A change in the baseline is visible and due to the scattering effect generated by the biomass. Wavelengths higher than 2125 nm were omitted for the biomass calibration (table 1).
The regression with low values of turbidity is better (SECV : 25.4 NTU) than the one with high values (SECV : 75.14 NTU). The reason is that the high value samples (turbidity higher than 260 NTU) confer a non-linearity in the regression. It seems that the linearity hypothesis is not true for NTU values higher than 260 NTU.

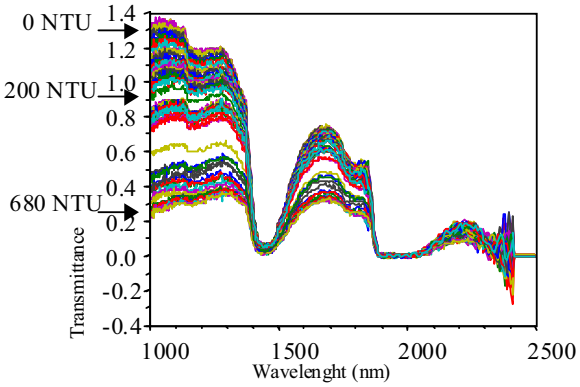

Figure 2 – Transmittance spectra of the whole samples.
Left arrows indicate corresponding biomass concentrations.

	SECV (NTU)	R	Min (NTU)	Max (NTU)
66 samples (>260 NTU)	75.14	0.90	0	676
56 samples (<260 NTU)	25.4	0.94		

Table 1 – Results of biomass calibration with or without values higher than 260 NTU).

The influence of biomass is also clearly visible in Figure 3. The left plot indeed shows the prediction of butyrate by PLS regression in cross-validation without any pre-treatment applied. From the 41st sample, the validation's curve fits the shape of the curve of biomass during the biomass addition phase (Figure 1). But at this stage, the molecular compound concentrations are constant (see the measure points in figure 3). The baseline transformation influences significantly the molecule concentration predictions. The application of the SNV correction on the transmittance values allows the best correction of the effect, without a significant improvement in the basic model without pre-treatment (See Figure 3 right and Table 2). The improvement is mostly visible during the biomass addition phase because the butyrate prediction does not increase and fluctuates around a constant value.

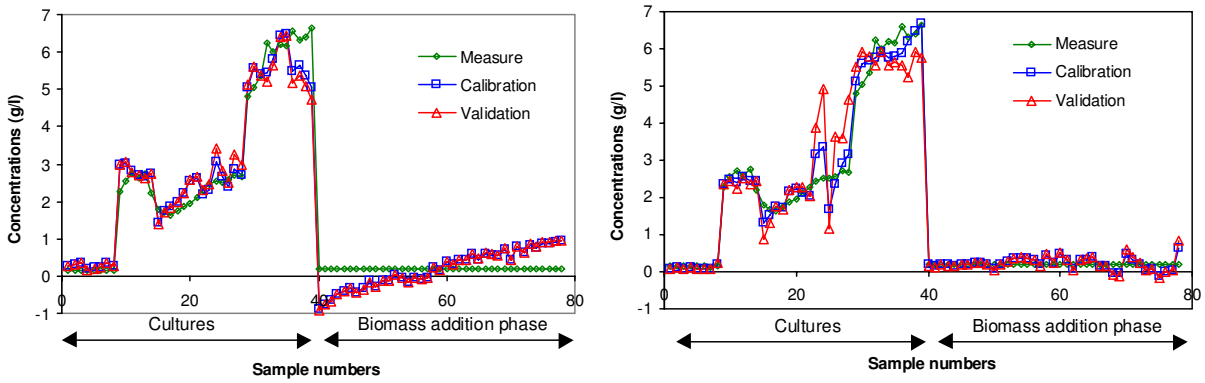

Figure 3 – Butyrate's prediction of PLS regression with cross-validation in the reactor. Left : Without pre-traitments. Right : Use of SNV with transmittance value then the log has been applied.

Improvement of the results can be seen when looking at the SECV values. SNV on the absorbance values (SECV = 0.75 $g.L^{-1}$) does not improve the model, it degrades it compared to the basic model (SECV = 0.55 $g.L^{-1}$). Second derivative does not improve neither the results. Predictions without pre-treatment give similar results for acetate and butyrate. For sucrose, it is the opposite effect, because the predicted values decrease during the biomass addition phase. The models without pre-treatment predict an increase of the products of the fermentation and decrease of the substrate (sucrose) when the biomass increases. Table 3

presents the results of PLS regression in cross-validation of sucrose and acetate with SNV on transmittance values. The last 375 wavelengths were deleted too.

	SECV ($g.L^{-1}$)	R	Factors	SECV 5th phase	Max ($g.L^{-1}$)	Min ($g.L^{-1}$)
Abs. (375 last wavelengths deleted)	0.55	0.96	2	0.52	6.65	0.138
Abs-SNV. (375 last wavelengths deleted)	0.75	0.92	2	0.47		
Abs-2nd derivate (375 last wavelengths deleted)	1.95	0.32	1	1.23		
SNV-Abs (without lambda removal)	0.56	0.96	5	0.19		

Table 2 – Butyrate's results of PLS calibrations.

	SECV ($g.L^{-1}$)	r	Factors	SECV 5th phase	Max ($g.L^{-1}$)	Min ($g.L^{-1}$)
Acetate	0.12	0.95	7	0.06	0.203	1.763
Sucrose	2.5	0.72	7	1.04	15.75	0

Table 3 – PLS calibrations results, with SNV on transmittance values.

5 Conclusion

In the light of these results, we confirmed that the biomass has an effect on the near infrared spectroscopic measurements. The biomass creates some scattering effects which increase the absorbance values. Besides the predictions of biomass is not linear for values higher than 260 NTU. The use of non linear PLS methods as Spline-PLS should be a way to improve the biomass calibration. Moreover, despite the strong influence of biomass, SNV correction improved the molecular compound predictions. The best PLS regression in cross-validation was obtained when SNV correction was applied on the transmittance values rather than on the absorbance values. Moreover, the second derivative did not improve the predictions. That suggests that the multiplicative effects are higher than the additive effects. An other explanation of the bad results observed with the second derivative pre-treatment should be the addition of noise with the second derivative operation which disturbs the measurements.

6 Acknowledgments

The authors thank Gaël Rolland from the society BUCHI for his precious availability and Thierry Conte for his help during the experience phases.

7 Références

[1] J-P. Steyer, O. Bernard, D. Batstone, I. Angelidaki: *Lessons learnt from 15 years of ICA in anaerobic digestion processes, Water Science & Technology,* 53(4-5): 25-33, 2006..

[2] S. Fogelman, H. Zhao, M. Blumenstein : *A rapid analytical method for predicting the oxygen demand of wastewater. Anal Bioanal Chem* 386 :1773-1779, 2006.

[3] S. Fogelman, H. Zhao, M. Blumenstein : *Estimation of chemical oxygen demand by ultraviolet spectroscopic profiling and artificial neural networks. Neural Comput & Applic*. 15 : 197-203, 2006.

[4] J.-P. Steyer, J.C. Bouvier, T. Conte, P. Gras, J. Harmand, J._P. Delgenès : O*n-line measurements of COD, TOC, VFA, total and partial alkalinity in anaerobic digestion processes using infra-red spectrometry. Water Science and Technology,* 45(10) : 133-138, 2002.

[5] J.-M. Gonzalez-Saiz, C. Pizarro, I. Esteban-Diez, O. Ramirez, C. Javier Gonzalez Navarro, M.-J. Saiz-Abajo, R. Itoiz : *Monitoring of Alcoholic Fermentation of Onion Juice by NIR : Valorization of worthless onions. J. Agric. Food Chem*. 55 : 2930 – 2936*,* 2007.

[6] Y. Zhang, Z. Zhang, N. Sugiura, T. Maekawa : *Monitoring of methanogen density using near-infrared spectroscopy. Biomass and Bioenergy* 22 : 489 – 495, 2002.

[7] D. Cozzolino, M. Parker, R.-G. Dambergs, M. Herderich, M. Gishen : *Chemometrics and Visible-Near Infrared Spectroscopic Monitoring of Red Wine Fermentation in a pilot scale. Biotechnology and Bioengineering.* 95 (6) : 1101-1107, 2006.

[8] V.-M. Fernandez-Cabanas, A. Garrido-Varo, J. Garcia Olmo, E. De Pedro, P. Dardenne : *Optimisation of the spectral pre-traitement. Used for Iberian pig fat NIR Calibrations. Chemometrics and Intelligent laboratory systems.* 87 : 104-112, 2004.

[9] A.M.C. Davies : Spectroscopy Europe 19 (2 ; 4 ; 6), 2007.