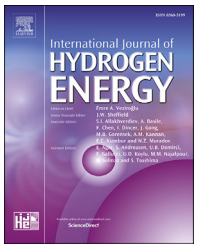
Available online at www.sciencedirect.com

ScienceDirect

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



Determination of alcohols and volatile organic acids in anaerobic bioreactors for H_2 production by near infrared spectroscopy

Maurílio Gustavo Nespeca [a,*], *Caroline Varella Rodrigues* [b], *Kamili Oliveira Santana* [b], *Sandra Imaculada Maintinguer* [c], *José Eduardo de Oliveira* [a]

[a] *Center for Monitoring and Research of the Quality of Fuels, Biofuels, Crude Oil, And Derivatives (Cempeqc), Institute of Chemistry, São Paulo State University (UNESP), Prof. Francisco Degni 55, Zip Code, 14800-060, Araraquara, SP, Brazil*

[b] *Biotechnology Department, Institute of Chemistry, São Paulo State University (UNESP), Prof. Francisco Degni 55, Zip Code, 14800-060, Araraquara, SP, Brazil*

[c] *Bioenergy Research Institute – IPBEN, São Paulo State University (UNESP), Zip Code, 13550-230, Rio Claro, SP, Brazil*

ARTICLE INFO

Article history:
Received 2 January 2017
Received in revised form
6 July 2017
Accepted 7 July 2017
Available online 25 July 2017

Keywords:
Anaerobic fermentation monitoring
Hydrogen bioproduction
Near infrared spectroscopy
Alcohols
Volatile fatty acids
Genetic algorithm

ABSTRACT

In recent years, near infrared (NIR) spectroscopy has been investigated as a tool for monitoring anaerobic digesters, but several adversities in its application have been reported. This study proposes the application of NIR for the determination of alcohols and volatile organic acids from H_2 production bioreactors and evaluates different approaches to optimize the prediction models. Partial least squad (PLS) models were developed using samples from anaerobic batch reactors fed with crude glycerol for wastewater treatment. The analytes predicted were: methanol, ethanol, 1-butanol, acetic, propanoic, butyric, isocaproic and total volatile organic acids (VFA). The optimization of the predictive capacity of the models was achieved through the orthogonal signal correction (OSC) preprocessing and the selection of variables performed by the genetic algorithm (GA). The application of the proposed models were based on the following figures of merit: accuracy, precision, linearity, limits of detection and quantitation, measurement interval, sensitivity, selectivity, signal-to-noise ratio and bias. Despite the low selectivity (maximum of 0.12%), the models presented high sensitivity [$\gamma^{-1} = 0.19$ (mg $L^{-1})^{-1}$], low LOQ (1 mg L^{-1}) and correlation between reference and predicted values (r) at least 0.93, except for propanoic acid ($r_{pred} = 0.85$). The F-test revealed that the selection of variables by GA significantly improved the accuracy and linearity of the prediction models for methanol, acetic acid, isocaproic acid and VFA. NIR spectroscopy has proved to be a powerful tool for monitoring H_2 production bioreactors since provides fast, low cost and multicomponent information.

* *Corresponding author.*
E-mail address: mauriliogn@gmail.com (M.G. Nespeca).

http://dx.doi.org/10.1016/j.ijhydene.2017.07.044

Introduction

The increase of global warming awareness and the pressing need for CO_2 neutral technologies have driven the modern industry towards the use of biomass in sustainable energy production [1,2]. The use of crude glycerol has been investigated in several energy efficient processes (e.g. combustion, composting, anaerobic digestion, and thermochemical or biological conversions to value-added products) and the anaerobic fermentation has gained most intention lately, since it produces hydrogen as a clean energy and ethanol, 1-butanol, and 1,3-propanediol as valuable liquid products [3–5]. High yields of hydrogen have been reported using pure glycerol as feedstock in anaerobic fermentation, however, crude glycerol from biodiesel manufacturing process would be a preferred feedstock for biohydrogen production, since the purification of glycerol is an expensive process [6–8]. In this perspective, a potential application for glycerol appears as an integrated production of biodiesel and biogas in an integrated biorefinery concept [6].

Glycerol may result in different yields of hydrogen per mole of organic substrate, depending on the route of the fermentation (oxidative or reductive) and the composition of the end products [7,9]. According to Rossi et al., the end products of the fermentation process are mainly volatile organic acids (VFA) such as acetic, propionic and butyric acids; and alcohols as ethanol and 1-butanol [10]. Fermentation processes of hydrogen production using anaerobic acidogenic bacteria have been extensively described by several authors [11–16]. In most of the bioconversion of glycerol pathways, H_2 is produced during oxidative metabolism of glycerol, in which pyruvate is formed as an intermediate and it may be metabolized to different end products, such as ethanol, 1-butanol, acetone, acetate, butyrate and lactate [17,18]. Higher yields of H_2 are observed when acetic acid is the major product in the end of the fermentation process, followed by butyric acid, 1-butanol and ethanol [17,19]. This becomes clear through the global bioconversion of glycerol Equations (1)–(4) below.

$$C_3H_8O_3 + H_2O \rightarrow CH_3COOH \text{ (acetic acid)} + CO_2 + 3H_2 \quad (1)$$

$$2C_3H_8O_3 \rightarrow C_4H_8O_2 \text{ (butyric acid)} + 2CO_2 + 4H_2 \quad (2)$$

$$2C_3H_8O_3 \rightarrow C_4H_{10}O \text{ (butanol)} + 2CO_2 + H_2O + 2H_2 \quad (3)$$

$$C_3H_8O_3 \rightarrow C_2H_6O \text{ (ethanol)} + CO_2 + H_2 \quad (4)$$

Thus, the volatile organic compounds present in the liquid medium during the anaerobic fermentation can provide a better understanding of the possible effects that cause increase or decrease of the biohydrogen yield.

Acid concentration levels and their changes over time denote the state of the bioprocess. VFA can accumulate during the anaerobic digestion process and result in an increase that directly reflects the process behaviour and/or imbalances since their accumulation may lead to process failure due to pH-drop they induce [6,20,21]. VFA are excellent compounds for indicating organic overload and the toxic condition which acid consumers and methanogens are inhibited [6]. However, there is no defined threshold inhibitory level of any VFA species that is valid for all reactors [22]. In addition, a kinetic uncoupling between the syntrophic bacteria of acid producers and consumers in these different fermentation steps can be observed by an accumulation of the different VFA and they can reflect the metabolic state of the obligate H_2-producing acetogens and the acetoclastic methanogens [23,24]. Several studies have pointed out that the determination of individual volatile fatty acids will also contribute towards improved understanding of anaerobic digestion and prevent a failure of the process [21,25]. Therefore, concentration and composition of the VFA are the best parameters to reflect the metabolic state of the biochemical process in anaerobic digester and this information would enable the optimal feeding of the biohydrogen or biogas plant [26]. Furthermore, the determination of impurities from crude glycerol such as methanol in the anaerobic digester is relevant as well as the monitoring of VFA due to its inhibitory character in high levels of concentration in biohydrogen production [7,27,28]. The concentration of methanol in the crude glycerol depends on the biodiesel production process and, usually, it is removed by vacuum evaporation and washed with n-hexanol before the use of glycerol as feedstock. However, studies have been carried out without removal of impurities from crude glycerol in order to reduce costs [28–30].

Bioprocess optimization is often limited by its inability to measure biomass, nutrients and product concentrations in a time frame that allows process adjustments [31]. In addition, the production of biogas or biohydrogen by anaerobic digestion is often described as unstable and even a slight variance in the operating conditions during fermentation can decrease the gas production significantly [2,7,32]. Monitoring of anaerobic digesters can take place in the solid, liquid or gas phase. The biochemical analysis of feedstock, digestate characteristics or reactor performance are typically carried out intermittently off-line and measured via wet chemical analysis [33]. Off-line determinations of the VFA and alcohols are usually performed with high performance liquid chromatography (HPLC) or gas chromatography (GC), however, these methods require sample preparation, expensive equipment, maintenance and skilled technicians [2,6,21]. A low cost and quick result of a VFA sum parameter can be achieved by on-line and off-line titrimetric methods, but these ignore the individual composition of the specific VFA [23]. In view of time delay between sampling and acquiring the analytical results as a limiting factor, there is a need for simple, rapid and cost-effective methods that are able to determinate individual components for the routine monitoring and control of anaerobic reactors. All of these requirements can be met through near infrared (NIR) spectroscopy associated with chemometric methods.

NIR spectroscopy satisfy the most significant characteristics of analytical methods, such as: fast (1 min or less per sample); non-destructive; non-invasive; suitable for on-line and at-line use; nearly universal application (any molecule containing C–H, N–H, S–H or O–H bonds); applicable in gaseous, liquid or

Table 1 – Compilation of results referred to the use of multivariate analysis of spectral data in the analysis of anaerobic digesters.

Ref.	Feedstock	Compound	Range (mg L^{-1})[a]	LV	RMSEP (mg L^{-1})[a]	Correlation
[2]	Pig manure and crop	acetic acid	0–6000	3	913	0.89
		propionic acid	0–2500	3	206	0.91
		butyric acid	1000–13,000	3	400	0.92
		isovaleric acid	100–1100	3	76	0.89
		VFA	1300–22,300	3	1594	0.90
[6]	Glycerol	acetic acid	2500–25,000	3	1476	0.98
		propionic acid	N.I.	1	1364	0.37
		butyric acid	N.I.	5	271	0.74
		isovaleric acid	N.I.	6	55.83	0.97
		VFA	5000–35,000	4	2095	0.98
[22]	Mixture of chicken and pig manure	acetic acid	0–5400	11	800[b]	0.84
		propionic acid	0–3400	10	570[b]	0.83
		VFA	0–9400	10	1530[b]	0.84
[24],[c]	Citrate-producing wastewater	ethanol	87–961	8	40	0.952
		acetic acid	421–1431	7	42	0.930
		propionic acid	318–2020	7	92	0.956
		butyric acid	112–2300	6	93	0.960
[25]	Harvest residues, catch crops and pig manure	acetic acid	37–2906	6	234	0.85
		propionic acid	0–554	6	N.I.	0.78
		butyric acid	0–91	7	N.I.	0.73
		isovaleric acid	0–98	6	N.I.	0.54
		VFA	52.0–3781	6	204	0.84
[26]	Maize silage	acetic acid	70–3080	N.I.	270	0.69
		VFA	310–10,200	N.I.	820	0.94
[32]	Pig slurry with maize Chicken manure	acetic acid	N.I.	N.I.	340[b]	0.919
		butyric acid	N.I.	N.I.	12.8[b]	0.786
[33]	Sewage sludge	VFA	29–1469	7	165	0.71
[36]	Maize silage	VFA	200–13,100	5	880	0.92
[37]	Pond sediments	acetic acid	N.I.	11	630	0.99
[38]	Several silages	acetic acid	100–5710	9	690	0.75
		propionic acid	30–7240	6	570	0.76
		VFA	130–9890	9	810	0.92
[39]	Shredded wheat	acetic acid	580–4820	[d]	330[b]	N.I.
		VFA	750–5490		400[b]	N.I.

RMSEP (root mean square errors of prediction).
N.I. (not informed by the authors).
LV (number of latent variables).
[a] Some values have been approximated to make them in the same unit for comparison purposes.
[b] RMSECV (root mean square errors of cross validation) results.
[c] Anaerobic digester for H_2 production.
[d] Models developed using SVM (support vector machine).

solid samples; and minimal or no sample preparation [26,34]. As a analytical technology, NIR is considered to be a promising method for biological process and it is increasingly being utilized to investigate the performance of anaerobic digesters [35,36]. Studies demonstrate the potential of NIR for VFA monitoring in anaerobic digesters for biogas production, including different feedstocks [2,6,22,25,26,32,33,36–39]. However, only one study was performed with anaerobic digesters for biohydrogen production [24]. The previous studies are summarized in Table 1.

NIR has an almost total dependence on statistics and chemometrics for its application. Partial least square (PLS) regression has been the most widely used multivariate calibration method for multicomponent analysis in anaerobic digesters. Although PLS models are very useful to resolve various calibration problems, the predictive ability of the model is highly influenced by the interferences of background and spectra noise [24,40]. In order to reduce the instrumental and matrix influences, calibration data are often preprocessed prior to data analysis. The baseline shifting is usually corrected by applying the first or second derivative; or by polynomials that correct the displacement based on a standard spectrum [41,42]. When the analytes are solubilized in a solvent with high analytical response, the selectivity of the method is compromised due to the overlap of the analyte bands by the solvent bands. Nonetheless, the selectivity can be improved using multivariate filters such as Orthogonal Signal Correction (OSC) to remove signals from background and interferences through identification of some unwanted covariance structure [24,43,44]. OSC filter uses samples with similar Y-block (concentration of analyte) values to identify the sources of variance in the X-block (spectra) to down-weight and allow a multivariate regression with less latent variables [43]. Another way to improve the selectivity and, consequently, the predictive ability of the chemometric model is by the selection of variables. The removal of variables, in

which the noise dominates over the information related to the analyte, often leads to better accuracy and performance of the analytical method and it is a technique widely accepted [45,46]. The selection of variables can be performed based on the spectral knowledge (manual approach) or through algorithms that seek to minimize the prediction error of the model such as Genetic Algorithm (GA). This algorithm is a popular heuristic optimization technique that employs a probabilistic, non-local search process that manipulates binary strings with the experimental variables [45]. Mixtures with almost identical spectra been successfully calibrated using GA in addition to the better understanding of the chemical system provided by the algorithm [45,47,48]. Furthermore, GA has already been used to optimize the biohydrogen production processes [49–53].

Multivariate calibration methods, such as Partial least squares regression (PLS), rely on accurate and precise reference values to provide appropriate models for prediction. Typically, the reference values of VFA and alcohols are obtained by chromatographic methods [21]. Although these methods may be precise, errors of the reference method are accumulated in the chemometric models. One way to reduce the amount of inserted errors in the model is by constructing a sample set with spiked samples, although this is not possible for all types of matrices due to their complexity [54]. Falk et al. developed an online monitoring of concentration and dynamics of VFA in anaerobic digestion processes with mid-infrared spectroscopy using two calibration sets: one with spiked samples; and other with real samples with reference values determined by HPLC [23]. The most important step for using a spiked sample set is the model validation with real samples. If the figures of merit (FOMs) of the model, such as accuracy, precision, linearity, limit of quantitation, etc, present acceptable values, the model can be applied in routine analyses [55].

Since the anaerobic digester can slow down, be severely disturbed or even break down completely due to the lack of appropriate monitoring, this study aimed to investigate the capacity of NIR spectroscopy combined with GA-PLS to determinate individual VFA and alcohols in anaerobic H_2-producing reactors fed with crude glycerol. The calibration models and the FOMs was performed using samples from previous studies with wastewater and crude glycerol for biohydrogen production [17,28]. In order to provide a substitute for the reference methods, calibration models for methanol, ethanol, acetic acid and butyric acid were also developed using spiked samples and tested with the bioreactor samples.

Materials and methods

Anaerobic batch reactors samples

The development and validation of the predictive models were performed using 100 samples from previous studies of anaerobic batch reactors with wastewater and crude glycerol for H_2 production [17,28]. The sample set consisted of 85 samples from batch reactors inoculated with a granular sludge from a UASB reactor used in the treatment of vinasse (São Martinho, Pradópolis - Brazil) [17]; and 15 samples from the treatment of poultry residues (DAKAR, Tietê - Brazil) [28]. The crude glycerol used in both studies was obtained from a pilot plant of biodiesel production in the Biotechnology Institute of Engineering Renewable Energy of UNIARA (Araraquara - Brazil) through transesterification of used cooking oils. The characterization of the crude glycerol, the growth conditions and the operation of the anaerobic batch reactors are described in detail in Ref. [17].

Reference method

The alcohols and volatile organic acids from the bioreactors were determined by gas chromatography with flame ionization detector (GC-FID). Prior to the GC analyzes, the biomass was removed by centrifugation at 9000 rpm for 10 min, at −4 °C, and then the upper phase was stored at −19° C. The frozen samples were thawed at laboratory temperature (22 °C) on the day of GC analysis and 2 mL aliquots of sample were transferred to 22 mL vials with 1 g of NaCl (to provide salting-out effect). The organic compounds concentrations (including acetone, methanol, ethanol, 1-propanol, 1-butanol, acetic, propionic, butyric, isobutyric, isovaleric, caproic and isocaproic acids) were measured by a GC (Shimadzu, Model 2010) equipped with a headspace sampler (AOC-5000 Plus), split/splitless injector, flame induction detector and a 30 m × 0.32 mm × 3.0 mm fused-silica capillary column (Restek, RTX-1). The samples were heated at 100 °C and shaken for 15 min by the sampler before the injection. Both injector and detector remained at 250 °C. Helium was used as the carrier gas with a flow rate of 1.0 mL min^{-1} (51.6 cm s^{-1}). The oven temperature was programmed initially at 45 °C for 1 min followed with a ramp of 50 °C min^{-1} to the final temperature, 250 °C, maintained for 3 min. Finally, the compounds were quantified by external calibration method.

Spiked samples

A calibration set with spiked samples was prepared to evaluate the possibility of replacing reference methods, e.g. GC analysis. The compounds chosen to prepare the samples were acetic and butyric acids and ethanol because of the relationship with high yields of biohydrogen; and methanol due to its inhibitory character and presence in crude glycerol [7,10]. A total of 100 samples were prepared using analytical grade compounds (Sigma Aldrich and Merk), in range of 150 mg L^{-1} a 8620 mg L^{-1}, into a modified PYG medium composed of glycerin (10.0 g L^{-1}), peptone (5.0 g L^{-1}), yeast extract (5.0 g L^{-1}) and meat extract (5.0 g L^{-1}), at a pH of 7.0. To avoid the covariance between the concentrations of the different compounds, the samples were prepared in a systematic way (Table A1, Appendices A).

Near infrared analysis

NIR spectra of the 100 bioreactors samples, 100 spiked samples and 15 blanks (PYG medium) were obtained by transflectance using 0.5 mL of sample. The Fourier transform spectrometer used was a Nicolet 6700 FT-IR (Thermo Scientific, Waltham, USA) set at 32 scans and 4 cm^{-1} of resolution. The NIR spectra were recorded in range of 10,000–4000 cm^{-1} using a Smart NIR Integrating Sphere accessory coupled to the

spectrometer and a stainless steel sampler of 0.2 mm optical path. The conditions of temperature and relative humidity during the analysis were, respectively, 21.2 ± 0.3 °C and 45 ± 7%.

Chemometric analysis

The calibration models were carried out using Partial Least Square regression (PLS) by MATLAB software (version R2013a) and PLS Toolbox (version 7.9.3). The bioreactor samples plus the blanks were divided into calibration (76 samples) and prediction (39 samples) sets by the Onion method to select representative samples for both sets [56]. Different data preprocessing were evaluated to provide the best model. The tested preprocessing were: mean center; Savtizky-Golay smoothing and derivatives; standard normal variate (SNV); multiplicative scatter correction (MSC); generalized least squares weighting (GLSW); and orthogonal signal correction (OSC).

After verification of the most appropriate preprocessing for each model, the use of the Genetic algorithm (GA) for variable selection was evaluated. Since GA of the PLS toolbox is limited to 200 generations, the algorithm was executed twice for each model. Both executions were performed with population size of 128 models, initial terms of 30%, mutation rate of 0.5%, double crossover and PLS regression method.

The number of latent variables (LV) was chosen based on root mean square errors of calibration (RMSEC), cross-validation (RMSECV) and validation (RMSEP) values to avoid model overfitting [57].

Validation

The validation was performed using the following FOMs: accuracy; precision; selectivity (SEL); sensitivity (SEN); analytical sensitivity (γ); linearity; measuring interval; bias; limit of detection (LOD); limit of quantitation (LOQ); and signal-to-noise ratio (S/N). As recommended by the analytical division of IUPAC, the figures of merit SEL, SEN, LOD, LOQ and S/N were calculated using the net analyte signal (NAS) theory [58,59]. The NAS is defined as the part of the analytical signal related to the analyte of interest alone and orthogonal to the space containing the interferences [60]. The NAS theory and FOMs equations have been well described and applied in the scientific literature [58,61–64], and thus are not described in detail here.

The accuracy was assessed by the errors values of the model, RMSEC and RMSEP, and the correlation coefficients (r) between the reference and predicted values. The precision of the analytical method was evaluated by the relative standard deviation (RSD) of seven samples analyzed in nine replicates on the same day.

The measuring interval was considered the range between the LOQ and the highest concentration present in the calibration set of the model. Analytical sensitivity was calculated as (SEN/δx), where δx is the approximation of the instrumental noise estimated by the standard deviation of the NAS value for 15 spectra of the reference signal (blank spectrum) [60].

Bias indicates presence of systematic errors in the model. This FOM can be assessed by a t-test for the validation samples at a confidence interval of 95%. The average bias was calculated by summing the differences between the reference value and the predicted value divided by the number of validation samples (Equation (1)) [54].

$$\text{bias} = \sqrt{\frac{\sum_{i=1}^{n_v}(y_i - \hat{y}_i)^2}{n_v}} \quad (1)$$

Then, the standard deviation of validation errors (SDV) was calculated as

$$\text{SDV} = \sqrt{\frac{\sum[(y_i - \hat{y}_i) - \text{bias}]^2}{n_v - 1}} \quad (2)$$

and finally the value of t_{bias} is given by

$$t_{bias} = \frac{|\text{bias}|\sqrt{n_v}}{\text{SDV}} \quad (3)$$

If the value obtained for t_{bias} was greater than the critical value for $n_v - 1$ degrees of freedom, the multivariate model had significant systematic errors.

The calibration models developed with the spiked sample set were validated using all the bioreactors samples and the application of the proposed analytical method was evaluated through RMSEP values and correlation coefficients.

Results and discussion

Characteristics of the bioreactor samples

One of the main impurities of the crude glycerol from transesterification process is methanol. Its presence is considered an inhibitory effect to microbial growth and may interfere in the metabolic pathway to H_2 generation [10]. In this study, the crude glycerol used in the anaerobic batch reactors presented methanol content of 15.84% (w/w) [17]. Santana et al. observed that high yields of H_2 production in the fermentation process was inversely proportional to the methanol content in the digesters [28]. The concentration range of methanol in the analyzed samples was 43 mg L^{-1} to 21,435 mg L^{-1}. The samples with high concentration of methanol were obtained from bioreactors fed with 240 g L^{-1} and 320 g L^{-1} of crude glycerol.

The H_2 production is directly proportional to the generation of volatile organic compounds in liquid medium [19]. The generation of acetic acid is the main indicator of higher yields of H_2 followed by the presence of butyric acid, 1-butanol and ethanol. The concentration range, median and standard deviation of the main alcohol and volatile organic acids present in the bioreactor samples are shown in Table 2. The components with the highest mean of concentration were ethanol, acetic acid and isocaproic acid. Despite the wide range of 1-butanol concentration, most of the samples had low concentration of this analyte. The VFA values were resulted from the sum of the concentration of acetic, propanoic, butyric, isobutyric, valeric, isovaleric, caproic and isocaproic acids. The acids not listed in Table 2 were present in the samples in very low concentrations, below or near the LOQ of the reference method, so no predictive models for these analytes were developed in this study. The complete description of the anaerobic batch reactors can be found in Ref. [28].

Table 2 – Statistic values of bioreactor samples: full, calibration and prediction sets.

Item (mg L^{-1})	Full set			Calibration set[a]			Validation set[a]		
	Range	Median	SD	Range	Median	SD	Range	Median	SD
Methanol	43–21,435	986	2868	0–21,435	1040	3034	0–14,189	972	2546
Ethanol	0–1755	198	319	0–1755	160	317	0–1050	250	324
1-Butanol	0–2428	20	327	0–2428	17	352	0–1346	19	273
Acetic acid	0–931	341	267	0–931	327	270	0–835	386	262
Propanoic acid	0–310	42	63	0–310	42	64	0–279	42	61
Butyric acid	0–820	94	172	0–820	95	176	0–704	86	166
Isocaproic acid	0–511	209	123	0–511	223	128	0–357	191	114
VFA	5–3079	917	653	0–3079	933	655	0–2451	912	659

[a] The calibration and validation sets include the blank samples.

Spectral features

Wide bands that correspond to overtones and combinations of fundamental vibrations characterize NIR spectra. Nonetheless, some bands occur regularly enough to distinguish molecular groups [42]. Even compounds with same functional group have unique spectra in NIR, but when they are diluted in an aqueous medium, the intrinsic characteristics of each analyte are no longer visibly discernible because the vibration of O–H bonds of water are intensity in NIR. The spectra of the bioreactor samples and blanks (PYG medium) are shown in Fig. 1. The intense bands at 5400–4600 cm^{-1} and 7300–6500 cm^{-1} could be attributed mainly to the water because correspond to combination of axial and angular deformations and first overtone of OH group, respectively. Since NIR is dominated by absorptions associated with functional groups with hydrogen atom, the carbonyl group of the organic acids showed a weak band at 5263 cm^{-1} together the O–H stretching and H–O–H deformation combination band at 5200 cm^{-1}. The CH group was observed by the C–H combination band at 4314 cm^{-1} and by the wide band at 8500 cm^{-1} resulted from the third overtone of C–H stretching [65].

The possibility of quantifying alcohols and carboxylic acids in aqueous medium is due to the sensitivity of the NIR region to the hydrogen bonds [65]. The presence of hydrogen bonding increases the length of O–H bands, thereby perturbing the frequency of O–H vibration to shorter frequencies. For the reason that water, alcohols and carboxylic acids molecules participate in multiple hydrogen bonds, these solutions can be considered to consist of equilibrium mixtures of different hydrogen-bonded species [41].

Models development

The range, median and standard deviation of the analytes concentration for the calibration and validation sets are shown in Table 2. The Onion method was effective in the selection of samples for each set of samples, since the values of median and standard deviation were similar in both sets and, in addition, the calibration sets covered broader concentration ranges. Although the MSC and SNV are very useful in NIR analyses to correct the light beam scattering caused by solid particles present in the samples, especially, when much of the signal in a sample is the same in all samples [42], these preprocessing did not provide the best models. All models presented better prediction results using only OSC filter. Zhang et al. [24] obtained the best NIR/PLS models for ethanol, acetate, propionate and butyrate in anaerobic reactors using the OSC with first derivative, however, the use of the latter preprocessing did not generate better prediction results in this

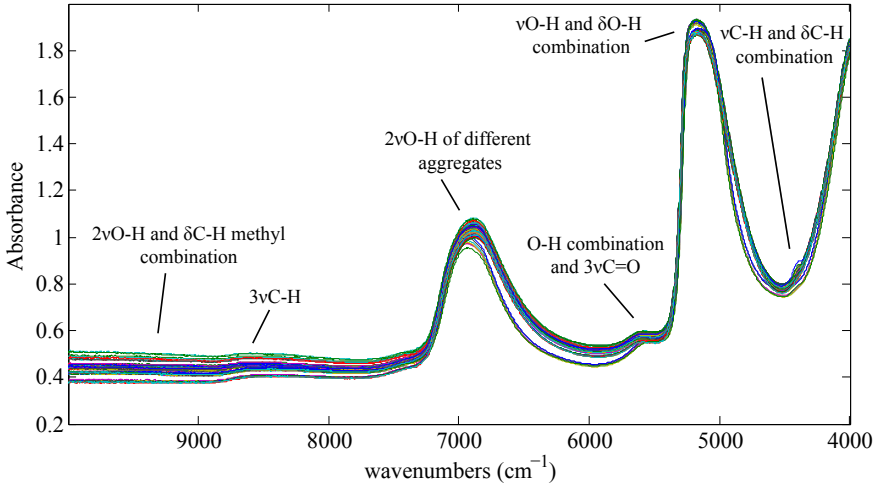


Fig. 1 – Spectra of bioreactor samples and PYG medium solutions.

study. This can be explained by the reduction of the signal-to-noise ratio related to the use of the first derivative [41].

When there is insufficient selectivity in the variables, multivariate filters as GLSW and OSC may remove signals from background and interferences through identification of some unwanted covariance structure (i.e., how variables change together) [43]. GLSW uses samples with similar Y-block (component concentration) values to identify the sources of variance in the X-block (spectra) to down-weight and allows a regression using few latent variables [43]. OSC is a filter that removes from the X-block only the variables that are unrelated to Y. This is made by ensuring that the removed variables are mathematically orthogonal to Y. Normally, two OSC components are used for NIR spectra, where the first component generally resembles a "baseline correction" and the second component may correct multiplicative effects [66]. Here we use only one OSC component, since the multiplicative effects are not intense in NIR analyzes by transflectance [65].

During the evaluation of the most suitable preprocessing for the PLS models with the full spectra, no samples were considered as outlier, therefore, the genetic algorithm was executed with all samples and OSC preprocessing. As each execution of the algorithm was performed with 30% of the initial terms and the full spectra had 3112 variables, the first execution resulted in models with approximately 934 variables and the second execution with number of variables close to 280. Each run took about 2 h with the GA settings used and a computer with Intel Core i7 3632QM 2.20 GHz processor and 8.0 Gb of RAM. The number of variables and the values of RMSEC, RMSEP and correlation coefficients of each model are available in Table 3. The RMSEP values and correlation coefficients of the prediction sets showed that the use of GA reduced the prediction errors and improved the correlation between the reference values and the predicted values for all the analytes. In order to compare the performance of PLS models with full spectra and GA variable selection, a *F*-test was performed. The *F*-values were obtained by the ratio of the variances (RMSEP squared) as in Equation (4).

$$F = \frac{RMSEP_1^2}{RMSEP_2^2} \quad (4)$$

where $RMSEP_1 > RMSEP_2$. If the calculated F value was greater than the tabulated F value, there was a significant improvement in the fit of the model using GA. Although the selection of variables improved the fit of all models, the reduction of RMSEP values was significant for the prediction of methanol, acetic acid, isocaproic acid and VFA. The GA-PLS model for prediction of propanoic acid was an exception in relation to the number of variables. After the first execution, there was a model overfitting and, consequently, the prediction errors increased. Thus, the final model for propanoic acid presented 924 variables.

Validation

The main FOMs of the final GA-PLS models are presented in Table 4. Despite the significant reduction of X block variables using GA, the prediction models required a large number of LVs. This may be justified by the complexity of the matrix and similarity between the compounds of interest. As the number of LVs was also defined by the lower RMSEP value, there was no overfitting of the calibration model.

The correlation coefficients, r_{cal} and r_{pred}, of the models were above 0.93, which represents the good fit between the values determined by the reference method and the values

Table 3 – Calibration and prediction results for the models with full spectra and GA variable selection.

Model	Number of variables	r (cal)	r (pred)	RMSEC (mg L^{-1})	RMSEP (mg L^{-1})	F value
Methanol						
Full spectra	3112	0.9722	0.9432	706	867	1.77
GA	237	0.9884	0.9658	457	652	
Ethanol						
Full spectra	3112	0.9840	0.8807	56	152	1.58
GA	296	0.9840	0.9273	56	121	
1-Butanol						
Full spectra	3112	0.8705	0.9355	172	140	1.08
GA	263	0.9970	0.9357	27	135	
Acetic acid						
Full spectra	3112	0.9909	0.9278	36	103	3.33
GA	210	0.9934	0.9762	31	57	
Propanoic acid						
Full spectra	3112	0.9416	0.7703	21	39	1.45
GA	924	0.9380	0.8529	22	32	
Butyric acid						
Full spectra	3112	0.9720	0.8799	41	80	1.70
GA	284	0.9985	0.9319	10	61	
Isocaproic acid						
Full spectra	3112	0.9849	0.9263	22	46	2.08
GA	307	0.9964	0.9632	11	32	
VFA						
Full spectra	3112	0.9697	0.9136	159	269	2.14
GA	249	0.9965	0.9620	55	184	

*F tabulated (95% confidence level; 38 degrees of freedom) = 1.72.

Table 4 – Validation results of the GA-PLS models.

Component	Accuracy					Linearity				Bias	
	LVs	r (cal)	r (pred)	RMSEC (mg L^{-1})	RMSEP (mg L^{-1})	R^2 (cal)	R^2 (pred)	Intercept	Inclination	Prediction bias	t_{bias}
Methanol	8	0.988	0.966	457	652	0.977	0.933	-4.9×10^{-2}	5.6×10^{5}	8	0.08
Ethanol	10	0.984	0.927	56	121	0.968	0.860	-3.1×10^{-2}	2.0×10^{5}	1	0.07
1-Butanol	13	0.997	0.936	27	135	0.994	0.875	-1.5×10^{-2}	3.2×10^{5}	226	2.80
Acetic acid	10	0.993	0.976	31	57	0.987	0.953	2.7×10^{-4}	1.3×10^{5}	−6	0.69
Propanoic acid	8	0.938	0.853	22	32	0.880	0.728	8.1×10^{-3}	2.5×10^{4}	4	0.77
Butyric acid	15	0.998	0.932	10	61	0.997	0.868	4.7×10^{-4}	2.7×10^{5}	15	1.46
Isocaproic acid	13	0.996	0.963	11	32	0.993	0.928	7.7×10^{-4}	9.3×10^{4}	−3	0.49
VFA	13	0.996	0.962	55	184	0.993	0.925	-1.6×10^{-2}	5.9×10^{5}	−46	1.54

Component	Selectivity (%)	S/N		Analytical sensitivity (mg L^{-1})	γ^{-1} (mg L^{-1})$^{-1}$	LOD (mg L^{-1})	LOQ (mg L^{-1})	RSD (%)			Measuring interval (mg L^{-1})
		mean	maximum					lowest	highest	mean	
Methanol	0.11	8	50	7.0	0.14	0.5	1.4	8	26	13	1.4–21,435
Ethanol	0.06	6	29	5.4	0.19	0.6	1.9	5	17	11	1.9–1755
1-Butanol	0.03	10	72	9.7	0.10	0.3	1.0	7	16	12	1.0–2428
Acetic acid	0.12	14	41	11.8	0.08	0.3	0.8	4	27	12	0.8–931
Propanoic acid	0.06	9	35	7.6	0.13	0.4	1.3	15	58	34	1.3–310
Butyric acid	0.03	31	129	27.5	0.04	0.1	0.4	7	49	20	0.4–820
Isocaproic acid	0.03	16	64	13.7	0.07	0.2	0.7	4	16	8	0.7–511
VFA	0.01	8	32	7.1	0.14	0.5	1.4	7	40	15	1.4–3079

t critical (95% confidence level; 38 degrees of freedom) = 1.69.

predicted by the model. The only exception was the propanoic acid prediction model, which presented r_{pred} equal to 0.85. The fit of the models can be observed by the plot of reference versus predicted values for the calibration (black circles) and prediction (white circles) sets in Fig. 2. The RMSEC and RMSEP values were proportional to the range of the measurement interval. The GA-PLS models for acetic and isocaproic acids presented the smallest prediction errors when compared to the median concentrations of the prediction sets.

The models that had a significant reduction in RMSEP values through the selection of variables showed the highest values of R^2_{pred}, i.e. the use of GA also improved the linearity of the calibration models. Although the values of R^2_{cal} and R^2_{pred} were below 0.90 for prediction model of propanoic acid, there was no significant bias in predicted values, as could be verified by the t_{bias} value. Only the model for 1-butanol showed a significant bias in predicted concentrations. The presence of an anomalous sample in the prediction set of 1-butanol is observed in the plot of Fig. 2, which justifies the high positive bias in the prediction set.

The selectivity indicates the part of the measured signal that is unique to the analyte of interest [60]. The results presented in Table 3 shows that the better selectivity was only 0.12% of the signal (calibration of acetic acid) and the worst was 0.01% (calibration of VFA). In addition to the complexity of the matrix and the similarity between the compounds of interest, the fact that the analytes are diluted in aqueous medium reduces the selectivity of the method due to the high absorption of the NIR radiation by the water molecules [65]. However, as discussed in the literature, this fact does not preclude the determination of the analytes in question, even with low selectivity, since the hydrogen bonds between different groups cause shifting in the bands related to OH group [41].

The models with higher signal-to-noise ratio are more sensitive because the analytical signal of the compound of interest is relatively more intense. Therefore, the predictive models of acetic, butyric and caproic acids were the models with the highest analytical sensitivity (γ) and, consequently, lower values of γ^{-1}, which indicates the smallest difference in concentration that is discernible by the analytical method [60]. As results of γ^{-1}, all the models were very sensitive to small differences of concentration, values between 0.04 and 0.19 mg L^{-1} (Table 4). The LOD ad LOQ were directly related to the standard deviation of the blank (PYG medium) and inversely to the sensitivity of the model. The low values of LOD and LOQ were consequences of the low standard deviation of the blank resulted from multiple scans (32 scans each spectrum). The LOQ values provided comprehensive measurement ranges for all analytes (Table 4).

The precision of the models was evaluated by the relative standard deviation (RSD) of nine replicates of seven samples that were randomly chosen. Table 4 shows the values of the lowest, highest and mean RSD for each analyte. The precision evaluation resulted in high RSD values for predictive models of propanoic acid, butyric acid and VFA. However, only the propanoic acid model showed the mean RSD above 20%.

Models with spiked samples

Prior to the development of calibration models with spiked samples, the spectral differences between these samples and the bioreactor samples were evaluated. The spectra of the bioreactor samples presented less intense OH bands than the spiked samples, which may be the result of several compounds doing hydrogen bonds and presence of dry matter in solution. In order to correct the differences between the sets, SNV preprocessing was applied to normalize the spectra and

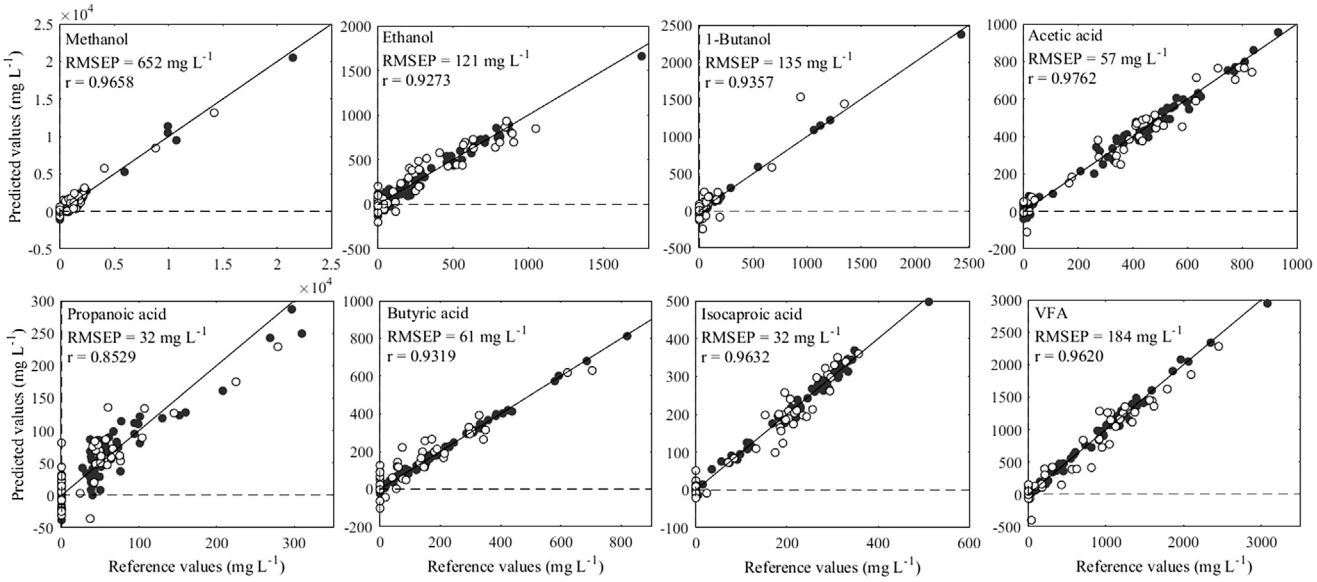


Fig. 2 – Plot of reference versus predicted values for the calibration (black circles) and validation (white circles) of spiked samples.

first derivative to correct the baseline shifting. The scores plot (Fig. 3) generated by Principal Component Analysis (PCA) of the spectra clearly shows the discrimination between the spiked and the bioreactor samples prior to preprocessing and the reduction of dissimilarity after preprocessing.

After normalization and baseline correction, the calibration models were developed using full spectra and GA variables selection, and then validated with the bioreactor samples. The results of calibration and validation of the models are present in Table 5. Despite the attenuation of the spectral differences of the two sets by the use of preprocessing, the models developed with spiked samples did not provide good prediction results for bioreactor samples. Differently from the models developed with the bioreactor samples, the use of GA increased RMSEP values and decreased correlation coefficients, except for prediction of methanol. The selection of variables only accentuated the differences between the two sets leading to larger errors of prediction.

As the spiked samples were prepared with only four compounds on PYG media, the amount of other by-products may have influenced the analytical signal of the analytes of interest. Some bioreactor samples had considerable concentration of 1-butanol and isocaproic acids. Since these compounds can make hydrogen bonds, the equilibrium between compounds with hydroxyl groups is different in the two sets.

The potential of NIR for monitoring anaerobic bioreactors

Some of the studies have demonstrated the effectiveness of NIR in the scanning of fermentation process through different approaches [2,6,26,33,36,39,67]. Unlike the analyses performed by conventional methods, NIR analysis is rapid, relatively low cost and allows the acquisition of many real-time data, therefore, it provides an efficient monitoring of the anaerobic digester. Besides the ease of performing NIR analysis, the method developed in offline mode can be transferred to online mode even using different equipment. Krapf et al. [67] demonstrated the possibility of transferring an offline NIR application to an online process analyzer for in situ monitoring of anaerobic digestion.

Online monitoring is problematic due to the complexity of the matrix. Some systems have been reported for online measurement of individual VFA using GC, however the main problem is the saturation of the automatic filtration system [68]. Boe et al. (2007) developed a system for online analysis of volatile compounds in anaerobic digesters based on headspace extraction without filtration. This same extraction concept for GC measurement could be applied to NIR since the heating temperature is constant and the measurement can be performed in the gas phase, and, consequently, reduce most of the interferences related to the matrix.

The validation of the developed models provided significant information about the NIR application for monitoring of main constituents in H_2-production bioreactors. Although the analytes have individual spectral characteristics, the intense bands of water overlap the intrinsic bands of each analyte, which results in low signal-to-noise ratio and low selectivity. However, the use of multivariate filters such as OSC has been effective in increasing prediction ability in aqueous media, as demonstrated in this study and by Zhang et al. [24].

Besides the water content, other problematic factors cited by authors were: (i) distribution of dry matter [33]; (ii) poor laboratory analysis quality (Y data) [25]; and (iii) changes in the physical or chemical constitution [23,67]. (i) The heterogeneous dry matter distribution can be problematic for the online monitoring of bioreactors, as it will increase the amount of noise in the models, however this can be overcome by preprocessing such as MSC and SNV, which compensate differences in specular reflectance and diffuse reflectance [41]; and by extensive replication and averaging sample spectra

[25]. (ii) Multivariate calibration is highly dependent on accurate and precise reference values (Y data). The concentrations of components determined by poorly calibrated analytical methods will result in more residues in the model, lower explained variance, and need for more latent variables, which may lead to an overfitting [41]. (iii) According to Krapf et al. [67], variations in pH, temperature, and analytes concentration of the slurry can change the hydrogen bonding within anaerobic digester samples and then may invalidate the NIR model. It is possible to control these variations through pH adjustment and temperature control in off-line analysis, but the only strategy to avoid erroneous estimates in online

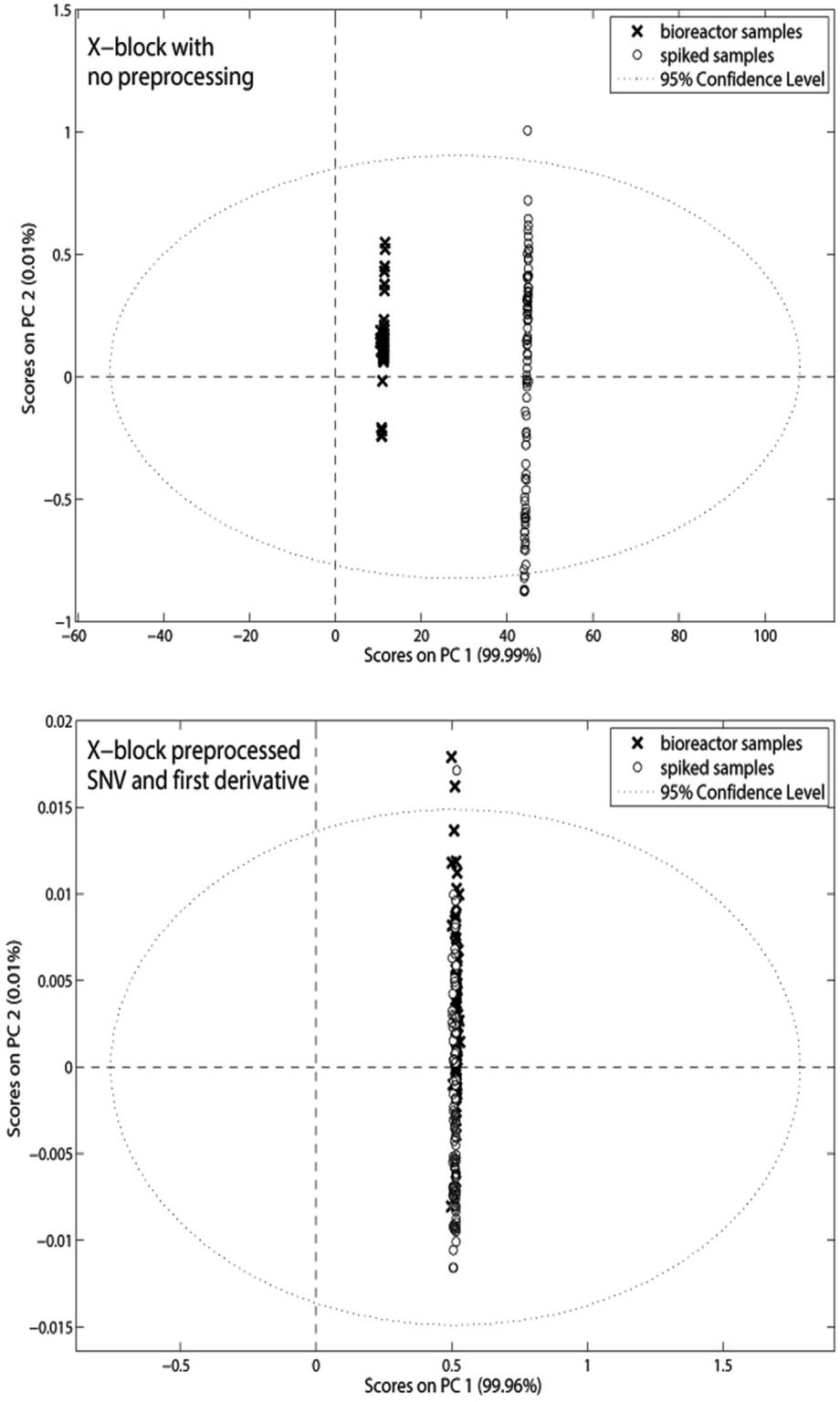


Fig. 3 – PCA of bioreactor and spiked samples with and without prepreprocessed X-block.

Table 5 – Calibration and validation of the models developed with spiked samples.

Model	Number of Variables	Correlation (cal)	Correlation (val)	RMSEC (mg L^{-1})	RMSEP (mg L^{-1})
Methanol					
Full spectra	3112	0.9590	0.8395	545	3221
GA	983	0.9992	0.8687	77	1555
Ethanol					
Full spectra	3112	0.9953	0.5714	179	973
GA	928	0.9641	0.3630	489	3864
Acetic acid					
Full spectra	3112	0.9906	0.3361	254	702
GA	886	0.9942	0.2357	200	4483
Butyric acid					
Full spectra	3112	0.9873	0.0911	298	1251
GA	880	0.9651	0.0793	490	3689

analysis is to incorporate these expected variations in calibration.

All studies found in the literature, except Ref [24], showed the same problems: high RMSEP; or low correlation; or both (Table 1). Even the models generated through support vector machine regression (SVM-r), useful for non-linearity problems, resulted in high RMSEP values [39]. The literature data show that the obstacles to NIR application in the monitoring of anaerobic digesters mentioned above are not easily overcome. Therefore, one feasible way to provide significantly more accurate predictions and better signal-to-noise ratios is through the use of multivariate filters such as OSC and variable selection algorithms such as GA. Accordingly, the methods will be more sensitive to small variations in concentration and more selective for the compounds of interest.

Conclusions

In this study, the validation of the developed models proved that NIR spectroscopy is a rapid and effective technique for the monitoring of alcohols and volatile organic acids in anaerobic bioreactors for H_2 production. The orthogonal signal correction filter provided better prediction capabilities to the models and the genetic algorithm significantly improved the accuracy and linearity of the models. The figures of merit obtained by the NAS concept revealed that the GA-PLS models were sensitive to small variations in analyte concentration and had low limits of quantification (approximately 1 mg L^{-1}). Therefore, the use of a multivariate filter and variable selection has made NIR spectroscopy a more promising technique for multicomponent analysis in anaerobic bioreactors, which provides a fast and low cost monitoring of fermentation process.

Acknowledgments

The authors would like to thank Capes for providing academic scholarships, Fundunesp for financial support and Cempeqc for providing equipments for analyses.

APPENDICES A

Table A1 – Methanol, ethanol, acetic and butyric acids concentrations of spiked samples.

Concentration (mg L^{-1})									
Sample	methanol	ethanol	acetic acid	butyric acid	Sample	methanol	ethanol	acetic acid	butyric acid
SS 1	8000	2000	4000	6000	SS 51	2049	2971	4016	963
SS 2	6000	8000	2000	4000	SS 52	205	389	598	807
SS 3	4000	6000	8000	2000	SS 53	410	779	1197	1615
SS 4	2000	4000	6000	8000	SS 54	615	1168	1795	2422
SS 5	8197	1844	3934	6025	SS 55	820	1557	2393	3230
SS 6	6148	7992	1885	3975	SS 56	1025	1947	2992	4037
SS 7	4098	5943	8033	1926	SS 57	840	168	387	605
SS 8	2049	3893	5984	8074	SS 58	1681	336	773	1210
SS 9	8403	1681	3866	6050	SS 59	2521	504	1160	1815
SS 10	6303	7983	1765	3950	SS 60	3361	672	1546	2420
SS 11	4202	5882	8067	1849	SS 61	4202	840	1933	3025
SS 12	2101	3782	5966	8151	SS 62	630	798	176	395
SS 13	8621	1509	3793	6078	SS 63	1261	1597	353	790
SS 14	6466	7974	1638	3922	SS 64	1891	2395	529	1185
SS 15	4310	5819	8103	1767	SS 65	2521	3193	706	1580
SS 16	2155	3664	5948	8233	SS 66	3151	3992	882	1975

Table A1 – (*continued*)

Concentration (mg L^{-1})									
Sample	methanol	ethanol	acetic acid	butyric acid	Sample	methanol	ethanol	acetic acid	butyric acid
SS 17	800	200	400	600	SS 67	420	588	807	185
SS 18	1600	400	800	1200	SS 68	840	1176	1613	370
SS 19	2400	600	1200	1800	SS 69	1261	1765	2420	555
SS 20	3200	800	1600	2400	SS 70	1681	2353	3227	739
SS 21	4000	1000	2000	3000	SS 71	2101	2941	4034	924
SS 22	600	800	200	400	SS 72	210	378	597	815
SS 23	1200	1600	400	800	SS 73	420	756	1193	1630
SS 24	1800	2400	600	1200	SS 74	630	1134	1790	2445
SS 25	2400	3200	800	1600	SS 75	840	1513	2387	3261
SS 26	3000	4000	1000	2000	SS 76	1050	1891	2983	4076
SS 27	400	600	800	200	SS 77	862	151	379	608
SS 28	800	1200	1600	400	SS 78	1724	302	759	1216
SS 29	1200	1800	2400	600	SS 79	2586	453	1138	1823
SS 30	1600	2400	3200	800	SS 80	3448	603	1517	2431
SS 31	2000	3000	4000	1000	SS 81	4310	754	1897	3039
SS 32	200	400	600	800	SS 82	647	797	164	392
SS 33	400	800	1200	1600	SS 83	1293	1595	328	784
SS 34	600	1200	1800	2400	SS 84	1940	2392	491	1177
SS 35	800	1600	2400	3200	SS 85	2586	3190	655	1569
SS 36	1000	2000	3000	4000	SS 86	3233	3987	819	1961
SS 37	820	184	393	602	SS 87	431	582	810	177
SS 38	1639	369	787	1205	SS 88	862	1164	1621	353
SS 39	2459	553	1180	1807	SS 89	1293	1746	2431	530
SS 40	3279	738	1574	2410	SS 90	1724	2328	3241	707
SS 41	4098	922	1967	3012	SS 91	2155	2909	4052	884
SS 42	615	799	189	398	SS 92	216	366	595	823
SS 43	1230	1598	377	795	SS 93	431	733	1190	1647
SS 44	1844	2398	566	1193	SS 94	647	1099	1784	2470
SS 45	2459	3197	754	1590	SS 95	862	1466	2379	3293
SS 46	3074	3996	943	1988	SS 96	1078	1832	2974	4116
SS 47	410	594	803	193	SS 97	769	1538	2308	3077
SS 48	820	1189	1607	385	SS 98	3415	768	1639	2510
SS 49	1230	1783	2410	578	SS 99	2521	3193	706	1580
SS 50	1639	2377	3213	770	SS 100	2155	2909	4052	884

REFERENCES

[1] Olabi AG. Hydrogen and fuel cell developments: an introduction to the special issue on the 8th international conference on sustainable energy and environmental protection (SEEP 2015), 11-14 August 2015, paisley, Scotland, UK. Int J Hydrogen Energy 2016;41:16323–9. http://dx.doi.org/10.1016/j.ijhydene.2016.07.235.

[2] Lomborg CJ, Holm-Nielsen JB, Oleskowicz-Popiel P, Esbensen KH. Near infrared and acoustic chemometrics monitoring of volatile fatty acids and dry matter during co-digestion of manure and maize silage. Bioresour Technol 2009;100:1711–9. http://dx.doi.org/10.1016/j.biortech.2008.09.043.

[3] Pachapur VL, Sarma SJ, Brar SK, Le Bihan Y, Buelna G, Soccol CR. Evidence of metabolic shift on hydrogen, ethanol and 1,3-propanediol production from crude glycerol by nitrogen sparging under micro-aerobic conditions using co-culture of Enterobacter aerogenes and Clostridium butyricum. Int J Hydrogen Energy 2015;40:8669–76. http://dx.doi.org/10.1016/j.ijhydene.2015.05.024.

[4] Liu B, Christiansen K, Parnas R, Xu Z, Li B. Optimizing the production of hydrogen and 1,3-propanediol in anaerobic fermentation of biodiesel glycerol. Int J Hydrogen Energy 2013;38:3196–205. http://dx.doi.org/10.1016/j.ijhydene.2012.12.135.

[5] Rossi DM, da Costa JB, de Souza EA, Peralba M do CR, Ayub MAZ. Bioconversion of residual glycerol from biodiesel synthesis into 1,3-propanediol and ethanol by isolated bacteria from environmental consortia. Renew Energy 2012;39:223–7. http://dx.doi.org/10.1016/j.renene.2011.08.005.

[6] Holm-Nielsen JB, Lomborg CJ, Oleskowicz-Popiel P, Esbensen KH. On-line near infrared monitoring of glycerol-boosted anaerobic digestion processes: evaluation of process analytical technologies. Biotechnol Bioeng 2008;99:302–13. http://dx.doi.org/10.1002/bit.21571.

[7] Sarma SJ, Brar SK, Sydney EB, Le Bihan Y, Buelna G, Soccol CR. Microbial hydrogen production by bioconversion of crude glycerol: a review. Int J Hydrogen Energy 2012;37:6473–90. http://dx.doi.org/10.1016/j.ijhydene.2012.01.050.

[8] Escapa A, Manuel M-F, Moran a, Gomez X, Guiot SRR, Tartakovsky B, et al. Hydrogen production from glycerol in a membraneless microbial electrolysis cell. Energy Fuels 2009;23:4612–8. http://dx.doi.org/10.1021/ef900357y.

[9] Maintinguer SI, Hatanaka RR, de Oliveira JE. Glycerol as a raw material for hydrogen production. Biofuels - Status Perspect InTech 2015:580. http://dx.doi.org/10.5772/60013.

[10] Rossi DM, Berne Da Costa J, Aquino De Souza E, Peralba MDCR, Samios D, Záchia Ayub MA. Comparison of

different pretreatment methods for hydrogen production using environmental microbial consortia on residual glycerol from biodiesel. Int J Hydrogen Energy 2011;36:4814–9. http://dx.doi.org/10.1016/j.ijhydene.2011.01.005.
[11] Kotay SM, Das D. Microbial hydrogen production with Bacillus coagulans IIT-BT S1 isolated from anaerobic sewage sludge. Bioresour Technol 2007;98:1183–90. http://dx.doi.org/10.1016/j.biortech.2006.05.009.
[12] Fan Y, Li C, Lay JJ, Hou H, Zhang G. Optimization of initial substrate and pH levels for germination of sporing hydrogen-producing anaerobes in cow dung compost. Bioresour Technol 2004;91:189–93. http://dx.doi.org/10.1016/S0960-8524(03)00175-5.
[13] Khanal SK, Chen WH, Li L, Sung S. Biological hydrogen production: effects of pH and intermediate products. Int J Hydrogen Energy 2004;29:1123–31. http://dx.doi.org/10.1016/j.ijhydene.2003.11.002.
[14] Lin CN, Wu SY, Lee KS, Lin PJ, Lin CY, Chang JS. Integration of fermentative hydrogen process and fuel cell for on-line electricity generation. Int J Hydrogen Energy 2007;32:802–8. http://dx.doi.org/10.1016/j.ijhydene.2006.09.047.
[15] Hawkes FR, Hussy I, Kyazze G, Dinsdale R, Hawkes DL. Continuous dark fermentative hydrogen production by mesophilic microflora: principles and progress. Int J Hydrogen Energy 2007;32:172–84. http://dx.doi.org/10.1016/j.ijhydene.2006.08.014.
[16] Ito T, Nakashimada Y, Senba K, Matsui T, Nishio N. Hydrogen and ethanol production from glycerol-containing wastes discharged after biodiesel manufacturing process. J Biosci Bioeng 2005;100:260–5. http://dx.doi.org/10.1263/jbb.100.260.
[17] Rodrigues CV, Santana KO, Nespeca MG, de Oliveira JE, Maintinguer SI. Crude glycerol by transesterification process from used cooking oils: characterization and potentialities on hydrogen bioproduction. Int J Hydrogen Energy 2016;41. http://dx.doi.org/10.1016/j.ijhydene.2016.06.209.
[18] Sarma SJ, Brar SK, Le Bihan Y, Buelna G, Rabeb L, Soccol CR, et al. Evaluation of different supplementary nutrients for enhanced biohydrogen production by Enterobacter aerogenes NRRL B 407 using waste derived crude glycerol. Int J Hydrogen Energy 2013;38:2191–8. http://dx.doi.org/10.1016/j.ijhydene.2012.11.110.
[19] Maintinguer SI, Sakamoto IK, Adorno MAT, Varesche MBA. Bacterial diversity from environmental sample applied to bio-hydrogen production. Int J Hydrogen Energy 2015;40:3180–90. http://dx.doi.org/10.1016/j.ijhydene.2014.12.118.
[20] Vanrolleghem PA, Lee DS. On-line monitoring equipment for wastewater treatment processes: state of the art. Water Sci Technol 2003;47:1–34.
[21] Adorno MAT, Hirasawa JS, Varesche MBA. Development and validation of two methods to quantify volatile acids (C2-C6) by GC/FID: headspace (automatic and manual) and liquid-liquid extraction (LLE). Am J Anal Chem 2014;5:406–14. http://dx.doi.org/10.4236/ajac.2014.57049.
[22] Ward AJ, Bruni E, Lykkegaard MK, Feilberg A, Adamsen APS, Jensen AP, et al. Real time monitoring of a biogas digester with gas chromatography, near-infrared spectroscopy, and membrane-inlet mass spectrometry. Bioresour Technol 2011;102:4098–103. http://dx.doi.org/10.1016/j.biortech.2010.12.052.
[23] Falk HM, Reichling P, Andersen C, Benz R. Online monitoring of concentration and dynamics of volatile fatty acids in anaerobic digestion processes with mid-infrared spectroscopy. Bioprocess Biosyst Eng 2015;38:237–49. http://dx.doi.org/10.1007/s00449-014-1263-9.
[24] Zhang ML, Sheng GP, Mu Y, Li WH, Yu HQ, Harada H, et al. Rapid and accurate determination of VFAs and ethanol in the effluent of an anaerobic H2-producing bioreactor using near-infrared spectroscopy. Water Res 2009;43:1823–30. http://dx.doi.org/10.1016/j.watres.2009.01.018.
[25] Holm-Nielsen JB, Andrée H, Lindorfer H, Esbensen KH. Transflexive embedded near infrared monitoring for key process intermediates in anaerobic digestion/biogas production. J Near Infrared Spectrosc 2007;15:123–35. http://dx.doi.org/10.1255/jnirs.719.
[26] Jacobi HF, Moschner CR, Hartung E. Use of near infrared spectroscopy in monitoring of volatile fatty acids in anaerobic digestion. Water Sci Technol 2009;60:339. http://dx.doi.org/10.2166/wst.2009.345.
[27] Sharma Y, Parnas R, Li B. Bioenergy production from glycerol in hydrogen producing bioreactors (HPBs) and microbial fuel cells (MFCs). Int J Hydrogen Energy 2011;36:3853–61. http://dx.doi.org/10.1016/j.ijhydene.2010.12.040.
[28] Santana KO, Rodrigues CV, Nespeca MG, de Oliveira JE, Maintinguer SI. Hydrogen bioprodution from crude glycerol and wastewater. Ciência Tecnol Fatec JB 2016;8:1–14.
[29] Kivistö A, Santala V, Karp M. Non-sterile process for biohydrogen and 1,3-propanediol production from raw glycerol. Int J Hydrogen Energy 2013;38:11749–55. http://dx.doi.org/10.1016/j.ijhydene.2013.06.119.
[30] Cofré O, Ramírez M, Gómez JM, Cantero D. Pilot scale fed-batch fermentation in a closed loop mixed reactor for the biotransformation of crude glycerol into ethanol and hydrogen by Escherichia coli MG1655. Biomass Bioenergy 2016;91:37–47. http://dx.doi.org/10.1016/j.biombioe.2016.04.015.
[31] Sakhamuri S, Bober J, Irudayaraj J, Demirci A. Simultaneous determination of multiple components in nisin fermentation using FTIR spectroscopy. Process Biochem 2001;37:371–8.
[32] Ward a J, Møller HB, Raju CS. Using near infrared spectroscopy to predict volatile fatty acids in biogas processes utilising different feedstocks. In: Cláudia SC, Marques dos Santos Cordovil L, Ferreira, editors. 14th Ramiran int. conf., Lisboa; 2010. http://dx.doi.org/10.2166/wst.2009.345.
[33] Reed JP, Devlin D, Esteves SRR, Dinsdale R, Guwy AJ. Performance parameter prediction for sewage sludge digesters using reflectance FT-NIR spectroscopy. Water Res 2011;45:2463–72. http://dx.doi.org/10.1016/j.watres.2011.01.027.
[34] Pasquini C. Near infrared spectroscopy: fundamentals, practical aspects and analytical applications. J Braz Chem Soc 2003;14:198–219. http://dx.doi.org/10.1590/S0103-50532003000200006.
[35] Reed JP, Devlin D, Esteves SRR, Dinsdale R, Guwy AJ. Integration of NIRS and PCA techniques for the process monitoring of a sewage sludge anaerobic digester. Bioresour Technol 2013;133:398–404. http://dx.doi.org/10.1016/j.biortech.2013.01.083.
[36] Krapf LC, Heuwinkel H, Schmidhalter U, Gronauer A. The potential for online monitoring of short-term process dynamics in anaerobic digestion using near-infrared spectroscopy. Biomass Bioenergy 2013;48:224–30. http://dx.doi.org/10.1016/j.biombioe.2012.10.027.
[37] Zhang Y. Monitoring of methanogen density using near-infrared spectroscopy. Biomass Bioenergy 2002;22:489–95. http://dx.doi.org/10.1016/S0961-9534(02)00019-3.
[38] Krapf LC, Gronauer A, Schmidhalter U, Heuwinkel H. Near infrared spectroscopy calibrations for the estimation of process parameters of anaerobic digestion of energy crops and livestock residues. J Near Infrared Spectrosc 2011;19:479–93. http://dx.doi.org/10.1255/jnirs.960.
[39] Stockl A, Loeffler D, Oechsner H, Jungbluth T, Fischer K, Kranert M. Near-infrared-reflection spectroscopy as measuring method to determine the state of the process for automatic control of anaerobic digestion. Int J Agric

Biol Eng 2013;6:63–72. http://dx.doi.org/10.3965/j.ijabe.20130602.00.
[40] Bosch-Reig F, Gimeno-Adelantado JV, Bosch-Mossi F, Doménech-Carbó A. Quantification of minerals from ATR-FTIR spectra with spectral interferences using the MRC method. Spectrochim Acta Part A Mol Biomol Spectrosc 2017;181:7–12. http://dx.doi.org/10.1016/j.saa.2017.02.012.
[41] Gemperline P. Practical guide to chemometrics. 2nd ed. 2006. http://dx.doi.org/10.1201/9781420018301.
[42] Burns DA, Ciurczak EW. Handbook of near-infrared analysis, 3rd ed. Anal Bioanal Chem 2009;393:1387–9. http://dx.doi.org/10.1021/ja015320c.
[43] Eigenvector Research. Advanced preprocessing: multivariate filtering. 2013. In: http://wiki.eigenvector.com/index.php?title=Advanced_Preprocessing:_Multivariate_Filtering.
[44] Laghi L, Versari A, Parpinello GP, Nakaji DY, Boulton RB. FTIR spectroscopy and direct orthogonal signal correction preprocessing applied to selected phenolic compounds in red wines. Food Anal Methods 2011;4:619–25. http://dx.doi.org/10.1007/s12161-011-9240-2.
[45] Xiaobo Z, Jiewen Z, Povey MJW, Holmes M, Hanpin M. Variables selection methods in near-infrared spectroscopy. Anal Chim Acta 2010;667:14–32. http://dx.doi.org/10.1016/j.aca.2010.03.048.
[46] Mehmood T, Liland KH, Snipen L, Sæbø S. A review of variable selection methods in partial least squares regression. Chemom Intell Lab Syst 2012;118:62–9. http://dx.doi.org/10.1016/j.chemolab.2012.07.010.
[47] Vohland M, Ludwig M, Thiele-Bruhn S, Ludwig B. Determination of soil properties with visible to near- and mid-infrared spectroscopy: effects of spectral variable selection. Geoderma 2014;223–225:88–96. http://dx.doi.org/10.1016/j.geoderma.2014.01.013.
[48] Wang L, Liu E, Cheng Y, Bekele DN, Lamb D, Chen Z, et al. Novel methodologies for automatically and simultaneously determining BTEX components using FTIR spectra. Talanta 2015;144:1104–10. http://dx.doi.org/10.1016/j.talanta.2015.07.044.
[49] Brka A, Al-Abdeli YM, Kothapalli G. Predictive power management strategies for stand-alone hydrogen systems: operational impact. Int J Hydrogen Energy 2016;41:6685–98. http://dx.doi.org/10.1016/j.ijhydene.2016.03.085.
[50] Mu Y, Yu H-Q. Simulation of biological hydrogen production in a UASB reactor using neural network and genetic algorithm. Int J Hydrogen Energy 2007;32:3308–14. http://dx.doi.org/10.1016/j.ijhydene.2007.05.021.
[51] Huang SR, Lin CY, Wu CC, Liu ST, Lin LC, Yun SJ. Optimizing circuit parameters of the biohydrogen real-time power generating system by genetic algorithm. Int J Hydrogen Energy 2008;33:1598–606. http://dx.doi.org/10.1016/j.ijhydene.2007.10.051.
[52] Wang J, Wan W. Optimization of fermentative hydrogen production process using genetic algorithm based on neural network and response surface methodology. Int J Hydrogen Energy 2009;34:255–61. http://dx.doi.org/10.1016/j.ijhydene.2008.10.010.
[53] Wang J, Wan W. Application of desirability function based on neural network for optimizing biohydrogen production process. Int J Hydrogen Energy 2009;34:1253–9. http://dx.doi.org/10.1016/j.ijhydene.2008.11.055.
[54] ASTM E1655-05. Standard practices for infrared multivariate quantitative analysis. ASTM Int 2012;5:29. http://dx.doi.org/10.1520/E1655-05R12.2.
[55] Norris KH, Ritchie GE. Assuring specificity for a multivariate near-infrared (NIR) calibration: the example of the Chambersburg Shoot-out 2002 data set. J Pharm Biomed Anal 2008;48:1037–41. http://dx.doi.org/10.1016/j.jpba.2008.07.021.
[56] Eigenvector Research. Automatic sample selection, vol. 1; 2014. http://wiki.eigenvector.com/index.php?title=Automatic_sample_selection.
[57] Hawkins DM. The problem of overfitting. J Chem Inf Comput Sci 2004;44:1–12. http://dx.doi.org/10.1021/ci0342472.
[58] Olivieri AC, Faber NKM, Ferre J, Boque R, Kalivas JH, Mark H. Uncertainty estimation and figures of merit for multivariate calibration. Pure Appl Chem 2006;78:633–61. http://dx.doi.org/10.1351/pac200678030633.
[59] Ferré J, Brown SD, Xavier Rius F. Improved calculation of the net analyte signal in inverse multivariate calibration. J Chemom 2001;15:537–53. http://dx.doi.org/10.1002/cem.647.
[60] Hatanaka RR, Sequinel R, Gualtieri CE, Tercini ACB, Flumignan DL, De Oliveira JE. Development and validation of an environmentally friendly attenuated total reflectance in the mid-infrared region method for the determination of ethanol content in used engine lubrication oil. Talanta 2013;109:191–6. http://dx.doi.org/10.1016/j.talanta.2013.02.047.
[61] Silva MAM, Ferreira MH, Braga JWB, Sena MM. Development and analytical validation of a multivariate calibration method for determination of amoxicillin in suspension formulations by near infrared spectroscopy. Talanta 2012;89:342–51. http://dx.doi.org/10.1016/j.talanta.2011.12.039.
[62] Sarraguça MC, Lopes JA. The use of net analyte signal (NAS) in near infrared spectroscopy pharmaceutical applications: interpretability and figures of merit. Anal Chim Acta 2009;642:179–85. http://dx.doi.org/10.1016/j.aca.2008.10.006.
[63] Ferré J, Faber NM. Net analyte signal calculation for multivariate calibration. Chemom Intell Lab Syst 2003;69:123–36. http://dx.doi.org/10.1016/S0169-7439(03)00118-7.
[64] Mirmohseni A, Abdollahi H, Rostamizadeh K. Net analyte signal-based simultaneous determination of ethanol and water by quartz crystal nanobalance sensor. Anal Chim Acta 2007;585:179–84. http://dx.doi.org/10.1016/j.aca.2006.11.082.
[65] Workman J, Weyer L. Practical guide to interpretive near-infrared spectroscopy. 2007. http://dx.doi.org/10.1002/anie.200885575.
[66] Wold S, Antti H, Lindgren F, Öhman J. Orthogonal signal correction of near-infrared spectra. Chemom Intell Lab Syst 1998;44:175–85. http://dx.doi.org/10.1016/S0169-7439(98)00109-9.
[67] Krapf LC, Nast D, Gronauer A, Schmidhalter U, Heuwinkel H. Transfer of a near infrared spectroscopy laboratory application to an online process analyser for in situ monitoring of anaerobic digestion. Bioresour Technol 2013;129:39–50. http://dx.doi.org/10.1016/j.biortech.2012.11.027.
[68] Boe K, Batstone DJ, Angelidaki I. An innovative online VFA monitoring system for the anaerobic process, based on headspace gas chromatography. Biotechnol Bioeng 2007;96:712–21. http://dx.doi.org/10.1002/bit.21131.