Multi-block SO-PLS approach based on infrared spectroscopy for anaerobic digestion process monitoring

Lorraine Awhangbo, R. Bendoula, Jean-Michel J. M. Roger, Fabrice Béline

Multi-block SO-PLS approach based on infrared spectroscopy for anaerobic digestion process monitoring

L. Awhangbo[1, 4], R. Bendoula[2], J.M. Roger[2, 3] & F. Béline[1]

[1]Irstea, UR OPAALE, 17 av. de Cucillé, CS 64427, F-35044, Rennes, France
[2] ITAP, Univ Montpellier, Irstea, Montpellier SupAgro, 361, rue J.F. Breton, BP 5095, F-34196, Montpellier, France
[3] Chemhouse Research group, Montpellier, France
[4] Univ. Bretagne Loire, France

Awhangbo Lorraine lorraine.awhangbo@irstea.fr
Bendoula Ryad (Corresponding author: ryad.bendoula@irstea.fr)
Roger Jean-Michel jean-michel.roger@irstea.fr
Fabrice Béline fabrice.beline@irstea.fr

Abstract

Near infrared spectroscopy combined with multivariate calibration such as partial least squares regression is a promising technique for on-line monitoring of anaerobic digesters. Different substrates are used in digesters, depending on their availability and their methanogen potential, to optimize the process. In Europe, the feedstock for anaerobic digesters is dominated by slurry and food waste which are respectively highly biodegradable and fat-containing substrates. The monitoring of the anaerobic digestion process based on digestates coming from these substrates presents some difficulties. The digestion of highly biodegradable substrates comes with the presence of water, which hinders spectroscopic calibration. And fat-containing substrates could lead to the accumulation of long chain fatty acids which are quite difficult to detect in the infrared region. While all existing studies have explored adapted spectroscopic measurements to improve the process monitoring, this study investigated the use of NIRS combined with multi-block analysis to track important anaerobic digestion stability parameters. Infrared measurements can come from several sources in the process monitoring. In addition, sequential and orthogonalized partial least squares have proven their ability of exploiting the underlying relation between several data blocks. These multi-block methods are powerful chemometric tools which can be applied in the monitoring of anaerobic digestion. Polarization light spectroscopy which is also known to improve the comprehension of scattering media like the digestate was also studied.

Keywords: Anaerobic digestion monitoring**,** Near InfraRed Spectroscopy, Polarization Light Spectroscopy, SO-PLS.

ABBREVIATIONS

AD Anaerobic Digestion

LCFA Long Chain Fatty Acids

NH_4^+ ammonium

NIRS Near InfraRed Spectroscopy

OLR Organic Load Rate

PoLiS Polarization Light Spectroscopy

R(λ) reflectance from the immersed probe

$R_{bs}(\lambda)$ total backscattered reflectance of the remote probe

$R_{ms}(\lambda)$ multiple scattered reflectance of the remote probe

$R_{ss}(\lambda)$ single scattered reflectance of the remote probe

TS Total Solids

VFA Volatile Fatty Acids

VS Volatile Solids

1. INTRODUCTION

Visible and Near InfraRed Spectroscopy (Vis & NIRS) are noninvasive and nondestructive analytical methods used to assess major compounds in different types of materials. These methods, especially NIRS, have a widespread usage in routine analysis in biomedical, agricultural and environmental fields, with generally no sample preparation. Among environmental applications is the monitoring of anaerobic digestion (AD) or methanization process. AD process is the microbial degradation of organic matter consisting of complex reaction chains, under anaerobic conditions, producing a biogas rich in methane (CH_4) and carbon dioxide (CO_2) and a digestate. AD process is influenced at different levels by environmental factors or inhibitors and requires the monitoring of several state indicators such as volatile fatty acids (VFA), Long chain fatty acids (LCFA) and ammonia/ammonium (NH_3/NH_4^+). These inhibitors can come from the digested substrates. For example, substrates with high lipid content, which are interesting because of their high methanogenic potential, also present the greatest inhibition risks in biological processes. Indeed, lipid degradation leads to LCFA accumulation whose toxicity and inhibitory effect on methanogenic flora have been highlighted by several studies [1–2]. These inhibitions create significant fluctuations in biogas production, sometimes until complete failure of the anaerobic process. While wet chemistry is time-consuming and increases reaction time in case of inhibition in the process, NIRS can overcome these flaws by quickly providing concentration estimation for these multiple parameters once calibrations have been developed for the parameter of interest [3]. Therefore, NIRS can provide an instantaneous response of the process state and give early warnings of instabilities. Calibrations for AD monitoring with NIRS are mainly achieved by the means of chemometric methods such as partial least square (PLS) regression.

There are numerous difficulties regarding PLS calibrations with NIRS for AD process monitoring. The main difficulty is the presence of water in digestate samples especially with highly biodegradable substrates which leads to low total solids (TS) content in the digester. Water is a hindrance that upsets PLS calibrations because its absorption masks a non-negligible part of the chemical information [4]. Indeed, water absorbance in the near infrared region is very important especially at two characteristic bands: 1450nm and 1940nm [5]. Therefore, digestate samples were often dried before infrared measurements [6], which greatly improved the prediction of some parameters [7]. In the particular case of anaerobic digestion, this improvement due to drying is detrimental to the efficiency of the method because drying requires a long preparation of the sample. It is also worth noting that liquid digestate samples come with particles that increase diffusion or light scattering effect which strongly influence the total absorbance of the samples. Diffusion also affects the baseline and the profile of the spectra especially with the lengthening of the average optical path of the light passing through the sample.

Although, in existing literature, no work has been performed for LCFA prediction by NIRS in AD process, several studies have been conducted on raw digestates for VFA and ammonium. PLS models were able to produce interesting prediction results for VFA ($0.84 \leq R^2 \leq 0.94$ and 200 mg/l $\leq$ RMSEP $\leq$ 1590 mg/l) [8-10] and ammonium ($0.91 \leq R^2 \leq 0.97$ and 160 mg/l $\leq$ RMSEP $\leq$ 250 mg/l) [11-12]. However, the substrates used in these studies were municipal solid waste, maize silage, pig or cattle manure sometimes mixed with slurry (liquid manure) leading to a TS varying between 5% and 10% in the digester. In the case of digestate with low TS content (< 5%) as in this

study performed on sewage sludge, the results were different [13]. For example, PLS model results obtained for VFA were less accurate ($0.69 \leq R^2 \leq 0.71$ and 160 mg/l $\leq$ RMSEP $\leq$ 180 mg/l). It was also shown in the study that apparent absorbance of sewage sludge was strongly affected by the distribution of dry matter throughout the sludge. This suggests that PLS models used in these previous studies were based on spectra less impeded by water absorbance. In addition, pig slurry digestion or co-digestion is particularly developed in France and a Europe-wide survey showed that TS content of pig slurry is very low with an average of 5% [14]. Therefore, there is a real interest in modeling AD process stability indicators under diluted conditions and, consequently, additional work is required. To overcome water hindrance, most studies focused on using different measurement systems such as transflectance method with transflexive embedded near infrared sensor (TENIRS) [8, 15]. It was highlighted that transflectance technique outperformed the reflectance technique; however transflectance required the use of a small optical path length (1 mm). This physical limitation makes the technique less attractive as the nature of sludge and digestates will exclude the use of small optical path lengths due to the possibility of fouling [15]. Fourier Transform NIRS is also often used to improve AD process monitoring.

While several studies are focused on improving NIRS measurement system, the regression method is another level of improvement. An unexplored potential solution is performing data fusion analysis by the means of multi-block methods to extract more relevant information from different sources to improve the process monitoring. Combining information from many datasets can improve the interpretation of the trends observed in the studied system [16]. It has been demonstrated, that it was more convenient to extract information from multi-block data sets handling all the blocks at the same time. Several statistical and chemometric multi-block methods are available and used for the purpose of exploring and modeling the relationships between several datasets to be predicted from several other datasets. These methods included: Hierarchical-PLS [17], Multi-Block-PLS (MB-PLS) [18], Sequential and Orthogonalized Partial Least Squares (SO-PLS) [19], Parallel Orthogonalized Partial Least Squares (PO-PLS) [20], Predictive-ComDim [21], Sequential and Orthogonalized multi-way version of PLS (SO-N-PLS) [22], Sequential and Orthogonalized Covariance Selection (SO-CovSel) [23] …. Most of these multivariate linear projection methods are based on PLS regression and are generally used in biological systems as metabolomics, industrial pharmaceutical process or quality control. In all cases, data fusion increased the interpretability of the models and enabled important biological conclusions on the monitored process [16].

Therefore, the objective of this study was to explore the applicability of the SO-PLS method, to predict relevant parameters in AD using two different sources of NIRS measurements on digestates. AD experiments were conducted with highly biodegradable and fat-containing co-substrates, known to induce inhibitions in the digester and to create interferences in infrared measurements. Two infrared probes were evaluated for the prediction of state indicators such as VFA, LCFA and NH_4^+. Focus was put on interpretation as well as prediction ability of these multi-block models. Moreover, one of the probes used in this study was based on polarization light spectroscopy (PoLiS), which can provide a unique contrast mechanism due to its sensitivity to particle morphology and other polarization properties [24-25]. The collected polarized signals were explored and related to the different parameters. They allowed a better comprehension of infrared measurements on digestate in anaerobic digestion process.

2. MATERIALS AND METHODS

2.1. Samples and Reference analysis

Samples used in this study come from a continuously stirred 35-liter tank reactor operating under mesophilic conditions (38 °C). The digester was fed once in the morning and several experiments were conducted with organic load rate (OLR) varying between 1.5 and 5kgCOD.m^{-3}.d^{-1} (Table 1).
All collected digestate samples went through different chemical analysis such as total solids, volatile solids, ammonium (NH_4^+) and pH measured according to standard chemical methods [26]. VFA were determined, on the supernatant after centrifugation of the samples, with high performance liquid chromatography (HPLC, Varian©, U3000). Gas chromatography/mass selective (GC/MS, Agilent Technologies, 7890B/5977A) was used to determine LCFA. Biogas production was automatically determined with a wet tipping bucket flow meter connected to the acquisition program of the digester.

Table 1: Operating conditions of performed co-digestion experiments ($g.l^{-1}$: g of co-substrate per liter of pig slurry)

N°	Substrates Mixture	OLR
1	Pig slurry + Horse feed residues (20$g.l^{-1}$) + Fruit waste (270$g.l^{-1}$) + Food fats (20$g.l^{-1}$)	4.9
2	Pig slurry + Horse feed residues (20$g.l^{-1}$) + Catering waste (200 $g.l^{-1}$)	3.4
3	Pig slurry + Horse feed residues (20$g.l^{-1}$) + Fruit waste (270 $g.l^{-1}$)	2.2
4	Pig slurry + Horse feed residues (40$g.l^{-1}$) + Fruit waste (540 $g.l^{-1}$)	4.2
5	Pig slurry + Horse feed residues (20$g.l^{-1}$) + Fruit waste (270$g.l^{-1}$)	1.4
	Pig slurry + Horse feed residues (40$g.l^{-1}$) + Fruit waste (540$g.l^{-1}$)	3.0
6	Pig slurry + Horse feed residues (20$g.l^{-1}$) + Fruit waste (270$g.l^{-1}$)	1.7
7	Pig slurry + Horse feed residues (20$g.l^{-1}$) + Fruit waste (270$g.l^{-1}$)	1.6
	Pig slurry + Fruit waste (135$g.l^{-1}$) + Protein waste (20$g.l^{-1}$)	3.8
8	Pig slurry + Horse feed residues (20$g.l^{-1}$) + Fruit waste (270 $g.l^{-1}$)	1.7
	Pig slurry + Horse feed residues (20$g.l^{-1}$) + Fruit waste (270 $g.l^{-1}$) + Food fats (20$g.l^{-1}$)	3.4

OLR: organic load rate

2.2. Spectral acquisition

Two spectroscopic measurement systems were tested at-line on the raw digestate samples. Both systems used the same light source, namely a Tungsten-Halogen source (Ocean Optics HL-200-FHSA) and the same spectrometer (LabSpec1, ASD Boulder). The spectral range of measurement extended from 350nm to 2500nm with a step of 1nm, and a resolution of 3nm for the range 350nm - 1000nm and 10nm for the range 1000nm - 2500nm.

2.2.1. Immersed or diffuse reflectance probe

This first probe (figure 1a) consisted of two fibers, one fiber for illumination and a second fiber for signal collection. The core diameter Ø of the two fibers was equal to 1000µm and a Numerical Aperture (NA) equal to 0.39 (BFY1000, Thorlabs). The principle of this probe was based on diffuse optical spectroscopy. For each sample, the intensity of the reflected light ($I(\lambda)$) was measured. Dark current ($I_n(\lambda)$) i.e. signal without light, was recorded from all measured spectra and subtracted. A white reference (SRS99, Spectralon®) ($I_0(\lambda)$) was measured to standardize spectra and prevent nonlinearities of all the instrumentation components (light source, fibers and spectrometer). From

these measurements, a reflectance ($R(\lambda)$) was calculated for each sample, as follows:

$$R(\lambda) = \frac{I(\lambda) - I_n(\lambda)}{I_0(\lambda) - I_n(\lambda)} \quad \text{Eq.1}$$

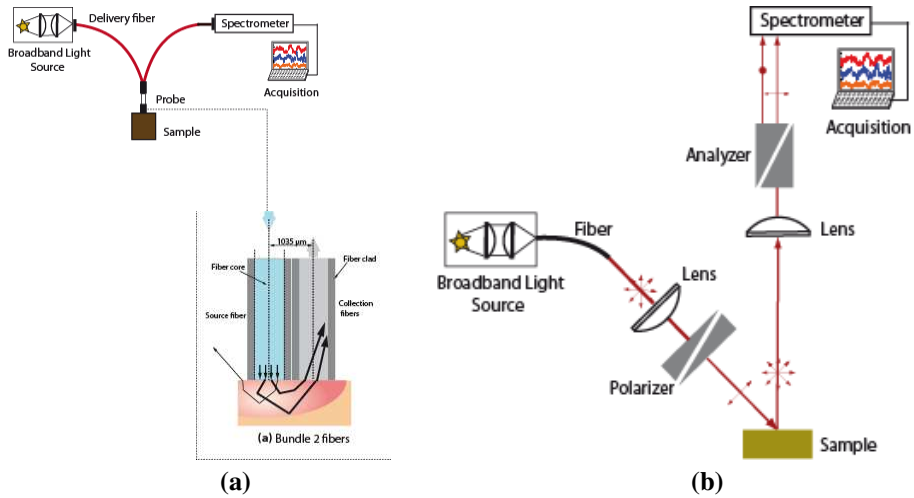
2.2.2. Remote or polarized diffuse reflectance probe

The second spectroscopic system was also based on diffuse reflectance spectroscopy (figure 1b), but the measurement was made at a distance of 5 cm from the sample. The other particularity of this probe was the use of polarization light spectroscopy. The source's emitted light was injected in an optical fiber (Thorlabs FG910LEC, Ø 910µm, NA 0.22) and a biconvex lens (Thorlabs LB1723-B, f= 60mm, d=50.8mm), forming a ≈ 2mm spot on the surface. The incident light cone was s-polarized using a wire-grid polarizer (Thorlabs WP25L-UB) offering a broadband polarization range (250nm - 4µm). The sample backscattered light was collected by a biconvex lens (Thorlabs LB1092, f=15mm, d=12.5 mm), forming a ≈ 1mm image of the lighted spot. This image was then split into an s-polarized image and a p-polarized image with a calcite Wollaston polarizer (Thorlabs WP10P) corresponding respectively to parallel $I_{II}(\lambda)$ and perpendicular $I_{\perp}(\lambda)$ emitted lights picked up by two optical fibers (Thorlabs FG910LEC, Ø 910µm, NA 0.22), allowing easy coupling to a spectrometer. As previously, a dark current $I_n(\lambda)$ was systematically recorded for all measured spectra with the same optical configuration and subtracted to each measurement. A reference measurement (SRS99, Spectralon®) was also taken as $I_0(\lambda) = I_0(\lambda)_{II} + I_0(\lambda)_{\perp}$. From these measurements, the weakly scattered reflectance $R_{ss}(\lambda)$ (Eq.2) and the multiple scattered reflectance $R_{ms}(\lambda)$ (Eq.3) were computed for each sample and summed in the total backscattered reflectance $R_{bs}(\lambda)$ (Eq.4) according to Bendoula et al. (2015) [27] as follows:

$$R_{ss}(\lambda) = \frac{(I_{\Pi}(\lambda) - I_n(\lambda)) - (I_{\perp}(\lambda) - I_n(\lambda))}{I_0(\lambda) - I_n(\lambda)} \quad \text{Eq. 2}$$

$$R_{ms}(\lambda) = 2 \times \frac{I_{\perp}(\lambda) - I_n(\lambda)}{I_0(\lambda) - I_n(\lambda)} \quad \text{Eq.3}$$

$$R_{bs}(\lambda) = \frac{(I_{\Pi}(\lambda) - I_n(\lambda)) + (I_{\perp}(\lambda) - I_n(\lambda))}{I_0(\lambda) - I_n(\lambda)} \quad \text{Eq. 4}$$



(a) **(b)**

Figure 1: Schematic (a) the immersed probe and (b) the remote probe.

2.3. Pre-processing

Absorbance or reflectance spectra were preprocessed depending on the used probe and the parameter to predict. It is customary to calculate absorbance as $A(\lambda) = \log(1/R(\lambda))$ from spectra

acquired in reflectance (R(λ)) to comply with the natural law of Beer Lambert. However, this logarithmic transformation is an intrinsic preprocessing which might not be necessary, especially for poor reflectance spectra. Several preprocessing methods such as smoothing or derivate using Savitsky–Golay (SG) algorithm [28]; detrending [29] and Standard Normal Variate (SNV) [30] were also tested on these spectra. Finally, these spectra were truncated in order to focus on most relevant ranges for each parameter, depending of the baseline deviation effect and noise presence, as indicated in Table 2 below.

Table 2: Preprocessing performed on the spectra with respect to the parameters and on each probe.

Signals	Pre-processing
Total VFA	
$R(\lambda)$	Reflectance, 1st Derivative SG 71pt window, SNV, 450–1700nm
$R_{bs}(\lambda)$	Reflectance, Smoothing SG 71pt window, 450–1700nm
$R_{ms}(\lambda)$	Reflectance, Smoothing SG 71pt window, 450–1700nm
$R_{ss}(\lambda)$	Reflectance, Smoothing SG 71pt window, 450–1700nm
Total LCFA	
$R(\lambda)$	Reflectance, 2nd Derivative SG 71pt window, 450–1700nm
$R_{bs}(\lambda)$	Reflectance, 1st Derivative SG 71pt window, 2-order Detrend, 450–1700nm
$R_{ms}(\lambda)$	Reflectance, 1st Derivative SG 71pt window, 2-order Detrend, 450–1700nm
$R_{ss}(\lambda)$	Reflectance, 1st Derivative SG 71pt window, 2-order Detrend, 450–1700nm
NH_4^+	
$R(\lambda)$	Absorbance, Smoothing SG 71pt window, 450–1800nm
$R_{bs}(\lambda)$	Absorbance, Smoothing SG 71pt window, SNV, 450–1800nm
$R_{ms}(\lambda)$	Absorbance, Smoothing SG 71pt window, 2-order Detrend, 450–1800nm
$R_{ss}(\lambda)$	Reflectance, 1st Derivative SG 71pt window, 450–1800nm

2.4. Data sets

166 samples were used in this study and analyzed by each spectroscopic technique. In order to validate the models, the samples were split in a training set with spectra collected from experiments N°1, 2, 3, 4 & 7 (107 spectra) and a test set with spectra collected from experiments N°5, 6 & 8 (59 spectra). Exploratory data analysis was performed to check that both sets spanned the whole variability domain. Moreover, as each experiment was independent, validation samples allowed the monitoring of entire AD process experiments with spectra having different characteristics. Cross-validation (CV) procedures were first performed on the training set to deduce the size of the model also corresponding to the number of latent variables (LV). The same CV blocks (i.e. 7±1 consecutive spectra with respect the experimentation duration, thus making 15 blocks) were used for all models computations (PLS and SO-PLS models) made in this study.

Hence, PLS models were calculated for each acquired data set ($R_{ms}(\lambda)$, $R_{ss}(\lambda)$, $R_{bs}(\lambda)$ & $R(\lambda)$) and validated on their respective test set. SO-PLS was next performed on different infrared data combinations. All computations and multivariate data analysis were performed with Matlab software v. R2013b (The Mathworks Inc., USA).

3. THEORY

3.1. PLS regression

PLS links a block of X_i descriptors with a block of responses Y. The general idea of PLS is to extract the latent variables (LVs) T and U, by simultaneous factorization of the independent and dependent blocks into their respective scores (T & U) and loadings vectors (P & Q) as (equation 5):

$$X = TP^T + E \ \& \ Y = UQ^T + F \qquad \text{Eqs.5}$$

E and F are residual matrices. PLS assume a linear relationship between X and Y so that:

$$\hat{Y} = TQ^T \qquad \text{Eq.6}$$

. In this study, four predictor blocks were considered which are: X_1, X_2, X_3 and X_4 respectively from the multiple scattered reflectance $R_{ms}(\lambda)$, the weakly scattered reflectance $R_{ss}(\lambda)$ and the total backscattered reflectance $R_{bs}(\lambda)$ of the polarized probe and the reflectance $R(\lambda)$ from the immersed probe. The response block Y consists of state indicators VFA, LCFA and ammonium. A Partial Least Square (PLS) algorithm was used to model each X_i block for each response Y. As mentioned above, standard validation methods were next used to determine the number of components to incorporate in each regression model and to assess the quality of the predictor obtained. These models feature performance parameters such as the coefficient of determination (R^2), the Root Mean Squared Error (RMSE) of Cross-validation or Prediction (RMSECV/RMSEP).

3.2. SO-PLS regression

The SO-PLS regression method [19, 31] is developed for estimating regression equations with N blocks of independent variables, i.e.

$$Y = X_1B_1 + X_2B_2 + ... + X_NB_N + E \qquad \text{Eq.7}$$

where $X_1(A\times J)$, $X_2(A\times M)$ and $X_N(A\times N)$ are the predictors blocks with the same number of observation A; $B_1(J\times R)$, $B_2(M\times R)$ and $B_N(N\times R)$ are the regression coefficients; $E(A\times R)$ is the residual matrix and $Y(A\times R)$ the response block. The SO-PLS method is assumed to be linear and the formula with two blocks of independent variables can be represented by the equation:

$$Y = X_1B_1 + X_2B_2 + E \qquad \text{Eq.8}$$

The algorithm is simple and requires four steps after centering and possibly scaling the data:

- Y is fitted to X_1 by PLS-regression, giving PLS scores T_{X_1},
- X_2 is orthogonalized (obtaining X_2^{orth}) with respect to T_{X_1}: $X_2^{orth} = X_2 - T_{X_1}(T_{X_1}{}^T T_{X_1})^{-1} T_{X_1}{}^T X_2$
- The estimated Y residuals from the first PLS are fitted to X_2^{orth} using PLS regression, giving PLS scores $T_{X_2}{}^{orth}$. The original Y could also have been fitted to the deflated X_2 without changing the results.
- The full predictive model is then computed as the ordinary least squares fit of Y to T_{X_1} and $T_{X_2}{}^{orth}$ used as independent variables and can be written as:

$$\hat{Y} = T_{X_1}\hat{Q}_{X_1} + T_{X_2}{}^{orth}\hat{Q}_{X_2} = X_1\hat{V}_{X_1}\hat{Q}_{X_1} + X_2^{orth}\hat{V}_{X_2}\hat{Q}_{X_2} = X_1B_1^* + X_2^{orth}B_2^* \qquad \text{Eq.9}$$

where $\hat{V}_{X_i}$ are the weight matrices needed for the PLS scores calculation and $\hat{Q}_{X_i}$ are the associated loading vectors. As shown, SO-PLS is a sequential use of PLS regression in combination with orthogonalization which focuses on the incremental contributions of each new block. SO-PLS and especially orthogonalization allows (*i*) the independence of the relative scaling of the blocks, (*ii*) scale or dimensionality invariance and (*iii*) a non-iterative estimation procedure [19, 31]. The optimal number of components in the model can be defined for each block (independently on the

others) by: incremental or global estimation of components giving the lowest RMSECV errors [19, 31]. In the incremental strategy, the number of latent variables to be used is separately optimized for each regression; with the best number of LVs for the regression model between X_1 and Y as the start point. And, the number of components for next blocks is successively chosen, for separate optimization. In the global approach, the best possible combination is determined by evaluating a graph called Måge plot [19, 31] which shows all the possible combinations of LVs reporting the RMSECVs as a function of the total number of components. Both approaches were tested in this study. RMSECV reduction is regularly used to qualify multi-block models [32]. However, it is still recommended to test the method with the selected number of LV with an independent dataset.

4. RESULTS AND DISCUSSION

4.1. Digestate characteristics

In AD, load rates influence the process operation. Increases in the OLR can improve biogas production but can also induce failures of the digester in case of overload. Moreover, in the particular case of the digestion of substrates such as fats and proteins, there can be a rapid accumulation of the associated inhibitory compounds. In this study, various failures occurred in the digester as illustrated on figure 2, showing biogas production with VFA, LCFA and NH_4^+ evolution over time.

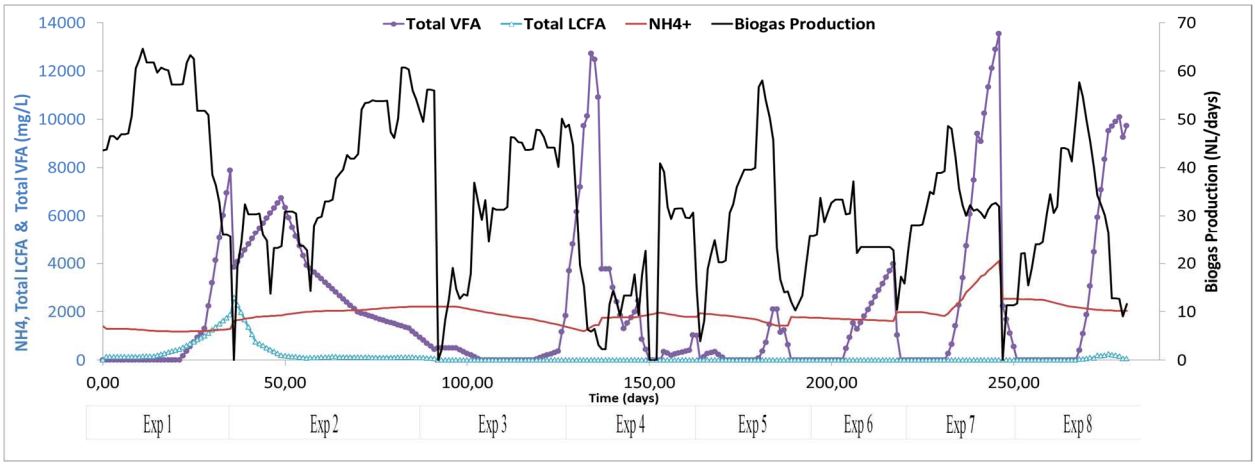


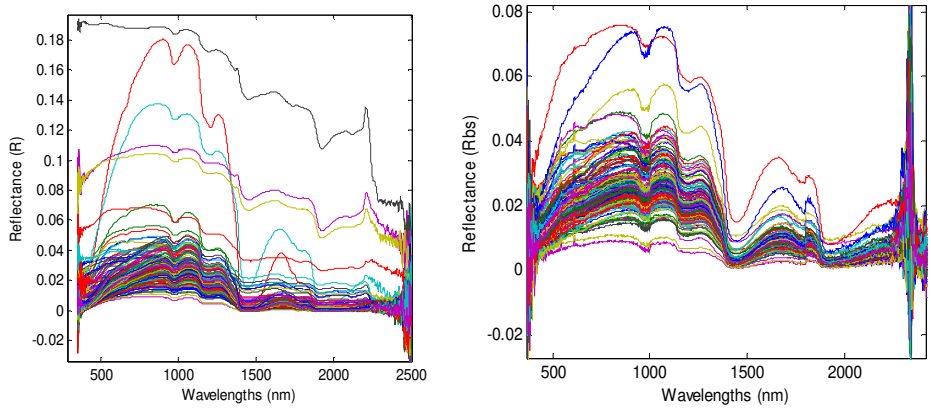
Figure 2: Evolution of the biogas production and VFA/LCFA/ NH_4^+ concentrations over time

In parallel with increases in the OLR, failures were mainly caused by inhibition due to protein, lipids or carbohydrates co-substrates. LCFA and ammonium accumulation occurred especially in co-digestion experiments of fats and protein waste substrate, respectively. VFA accumulation was common to all failures in the digester. Total and volatile solids varied little in all experiments due the highly biodegradable co-substrates added. TS content varied between 1.4% and 2.6% in all experiments. Different states of the process were observed in the digester. The digester went from steady state without VFA and LCFA accumulation to imbalance state with LCFA and VFA accumulation up to 2200 mg/l and 13500 mg/l respectively and often until acidosis state. These different states created wide ranges for the performance parameters. Particularly for VFA and LCFA, in steady state, reference analysis values were null. In the whole dataset, 75% of the reference values were null for LCFA and 46% for total VFA. Subsequently, for VFA and LCFA prediction with infrared spectra, negative values were systematically corrected to null for all

calculations performed in the present study.

4.2. Spectral Features

The collected digestate samples contain less than 3% of dry matter (total solids). As expected, all spectra of digestate have water spectral signature with absorption bands at 970nm, 1200nm, 1440nm and 1940nm, regardless of the used probe (immersed or remote probe). The difference between various spectra is a function of the reflectance intensity as a result of a combination of digestate chemical composition and the light's optical path, related to the light scattering, in the digestate. Indeed, all collected spectra had a reflectance intensity varying between 1% and 5% for the two probes except for some samples where it had reached 7% for the remote probe and between 10% and 18% for the immersed probe. These levels of intensity were coherent considering the digestate samples and their very liquid appearance. The signal of the remote probe was noisier than the signal of immersed probe, especially between 2000nm and 2500nm. Another characteristic of the immersed probe is the presence in some reflectance spectra of a feature at 2200nm, which is a peak between two absorption bands of water. This peak-like feature might be due to the appearance of another water adsorption peak at 2250nm characterizing a change in the optical path of the light. These characteristics are mostly related to physical particles in the digestate. This suggests that the physical aspect and light scattering of the digestate have a great influence on the infrared signal. However, changes in absorption bands of water could bring some non-negligible chemical information.

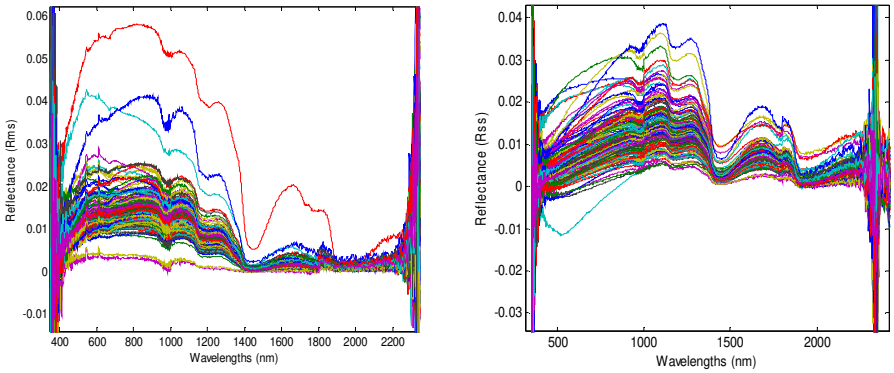


(a) **(b)**

Figure 3: Collected spectra for (a) the immersed probe and (b) the remote probe

As previously stated, the remote probe is based on polarization light spectroscopy which is unexplored among NIRS measurement systems used in AD process monitoring. Indeed, light scattering in biological media such as digestates leads to deformations of the measured spectra. These deformations result in nonlinearities which can degrade the quality of calibrations and lead to lack of robustness of PLS models. The use of polarization can reduce these distortions. There has been an increasing interest toward propagation of polarized light in highly scattering media, especially biological materials [33]. Mueller matrix, often used to completely characterize sample polarization properties [33], allowed decomposition of the total backscattered light of the remote probe into multiple scattered reflectance signal $R_{ms}(\lambda)$ and weakly scattered reflectance signal $R_{ss}(\lambda)$ (Figure 4). Globally, $R_{ms}(\lambda)$ and $R_{ss}(\lambda)$ had the same levels of reflectance intensity which varied between 1% and 4% except for some $R_{ms}(\lambda)$ spectra where the intensity reached 5%. $R_{ss}(\lambda)$ detected

photons which have low penetration in the digestate and corresponded to single scattering. This is confirmed by the collected $R_{ss}(\lambda)$ spectra as water absorption bands were less pronounced or nonexistent (at 970nm and 1200nm), characteristic of a short optical path in the media. $R_{ms}(\lambda)$ spectra detected photons having a longer optical path in the digestate. Indeed, in contrast to $R_{ss}(\lambda)$, these spectra showed defined absorption bands of water with shapes closer to those of the bands in the initial signal $R_{bs}(\lambda)$. The main advantage of light polarization in the present study is that the method provided several polarization components of the backscattered light related to the studied media.



(a) **(b)**

Figure 4: Collected spectra for (a) $R_{ms}(\lambda)$ and (b) $R_{ss}(\lambda)$ signals

4.3. Mono-block PLS model results

PLS models results on the training set as well as on the test set (LV, RMSECV, RMSEP and R^2), for mono-block regression, are shown in Table 3 for each response Y.

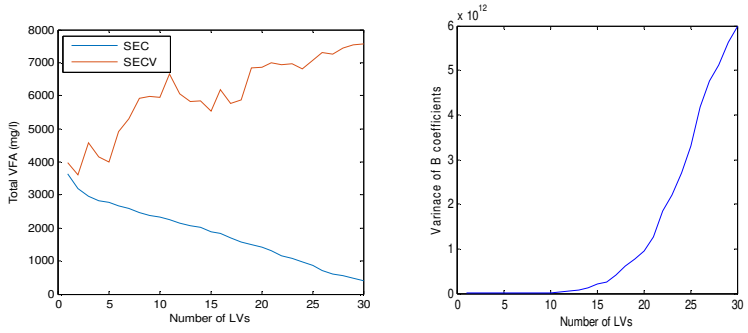
Table 3: PLS models results based each infrared signal $R_{ms}(\lambda)$, $R_{ss}(\lambda)$, $R_{bs}(\lambda)$ and $R(\lambda)$

Training	Range (mg/l)	$R_{ms}(\lambda)$ (X_1)			$R_{ss}(\lambda)$ (X_2)			$R_{bs}(\lambda)$ (X_3)			$R(\lambda)$ (X_4)		
		LV	R^2	RMSECV (mg/l)	LV	R^2	RMSECV (mg/l)	LV	R^2	RMSECV (mg/l)	LV	R^2	RMSECV (mg/l)
VFA	0 – 13548	5	0.05	3904	3	0.01	4520	11	0.54	2500	16	0.81	1711
LCFA	0 – 2269	4	0.42	353	7	0.19	412	7	0.31	391	4	0.39	357
NH_4^+	1180–4090	11	0.27	541	6	0.22	592	11	0.44	474	10	0.43	472
Test	Range (mg/l)	LV	R^2	RMSEP (mg/l)	LV	R^2	RMSEP (mg/l)	LV	R^2	RMSEP (mg/l)	LV	R^2	RMSEP (mg/l)
VFA	0 – 10096	5	0.73	1570	3	0.00	2982	11	0.55	2737	16	0.68	2366
LCFA	0 – 251	4	0.52	275	7	0.00	223	7	0.62	188	4	0.44	110
NH_4^+	1420–2530	11	0.43	407	6	0.52	313	11	0.57	409	10	0.63	516

From these PLS model results, it is noted that all RMSEP errors were lower than in training set. In the case of LCFA prediction, the test set was really shortened because LCFA is inhibitory at very low concentrations in the digester [34] as shown for example in experiment 8. It is therefore relevant to predict these low concentrations.

VFA prediction from $R_{ms}(\lambda)$ also provides an interesting result as prediction results were way better

than cross-validation results. By looking at the convergence curves of calibration and CV errors (Figure 5), 2 or 5 LVs could have been used in the model. However, there is potentially an outlier in the CV block used because RMSECV increased between these two points. This could explain why in CV, model results had less accurate performances than in prediction with 5 LVs. Moreover, the CV models b-coefficients variance did not increase until LV number equal to 10 showing that the chosen number of LVs is valid [35].



(a) **(b)**

Figure 5: (a) Cross-validation and calibration errors for VFA prediction by $R_{ms}(\lambda)$ (b) Evolution of the variance of the values of the regression vector, b, as a function of the number of latent variables in the PLS model

From models with polarized spectra (Table 3), it was shown that VFA and LCFA parameters were better modeled by $R_{ms}(\lambda)$ and NH_4^+ by $R_{ss}(\lambda)$. $R_{ms}(\lambda)$ provided higher R^2 for VFA and LCFA, both in training and test, with similar RMSECV and lower RMSEP than with $R_{ss}(\lambda)$. This is a bit surprising, knowing that for strongly scattering samples, multiple scattering will generally give an incoherent contribution to the scattering pattern and standardized data analysis tools cannot be applied [36]. In contrast, single scattering can provide a straightforward result which will simply link the chemical composition to the spectrum as shown for NH_4^+ prediction by $R_{ss}(\lambda)$. However, multiple scattering was found to significantly contribute to VFA and LCFA interpretations. This can be explained by the particularities of these parameters. For example, LCFA inhibition is accompanied by their absorption in the biomass, and flotation phenomena [37-38]. Therefore, LCFA are more present in the solid phase of the digestate due to the formation of fats aggregates. As multiple scattering is related to photons having a longer optical path and is affected by the presence of particles and dispersion in the digestate, $R_{ms}(\lambda)$ will more likely collect more information about this parameter. In the case of VFA, there is no aggregation due to their accumulation in the digestate however, there are also fatty acids. The total backscattered reflectance $R_{bs}(\lambda)$ provided similar results to the best models from $R_{ss}(\lambda)$ and $R_{ms}(\lambda)$. For LCFA, prediction results were better with $R_{bs}(\lambda)$ (R^2 of 0,62 and RMSEP of 188 mg/l).

Model results were globally similar between the two probes tested, except for LCFA. As explained above LCFA inhibition results in flotation phenomena which can be fully captured by the polarized probe. This probe also has a remote architecture in contrast with the immersed probe which might not be able to perceive these compounds due to immersion.

4.4. SO-PLS model results

Multi-block models were performed on different combinations of the available signals. The results

of these SO-PLS models are summarized in Tables 4, 5, 6 and 7. Both global and incremental approaches were tested. In cross-validation, it is worth noting that the combination of components giving the lowest RMSECV is not always the best solution. Therefore, to avoid over-fitted models, parsimony testing was used, and the best models were chosen between the different RMSECVs. The models were then validated on the test set with these chosen combinations of LVs.
The chosen order of the blocks can influence the final result in SO-PLS models. In this study, this influence was not very noticeable. For example, the prediction of LCFA with either $R_{bs}(\lambda)$ (X_3) and $R(\lambda)$ (X_4) or $R(\lambda)$ (X_4) and $R_{bs}(\lambda)$ (X_3) provided similar results. With X_3 & X_4, the number of LVs was 7-3 with RMSECV of 309 mg/l, RMSEP of 111 mg/l and R^2 test of 0.53. While, with X_4 & X_3, in this order, the number of LVs was 4-6 with RMSECV of 284 mg/l, RMSEP of 102 mg/l and R^2 test of 0.57. These results are very similar, especially with RMSEP errors. Therefore, the order of the blocks was not further studied and the used order in this study rather privilege the remote probe, which gives the possibility of several infrared signals.

4.4.1. SO-PLS models with the polarized probe

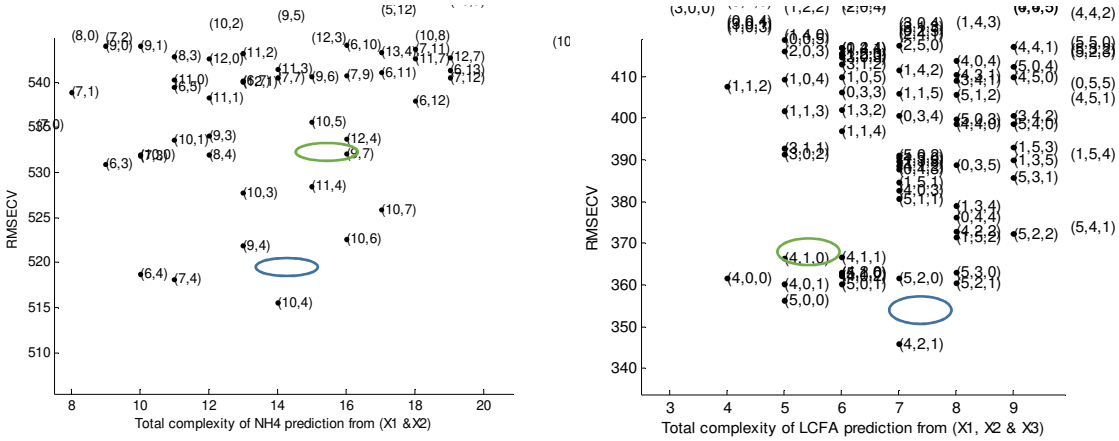
First, multi-blocks models were performed on signals from the polarized probe for all parameters. SO-PLS models were made on $R_{ss}(\lambda)$ and $R_{ms}(\lambda)$ (X_1 & X_2) in order to compare the results with their respective mono-block model results and also with $R_{bs}(\lambda)$. A second analysis was performed on the three signals (X_1 & X_2 & X_3) of the polarized probe as listed in Table 4.

Table 4: SO-PLS models results based on infrared signals from the polarized probe

	Global approach			**Sequential approach**			**Global approach**			**Sequential approach**		
Training	X_1 & X_2			X_1 & X_2			X_1 & X_2 & X_3			X_1 & X_2 & X_3		
	LV	R^2	RMSECV (mg/l)	LV	R^2	RMSECV (mg/l)	LV	R^2	RMSECV (mg/l)	LV	R^2	RMSECV (mg/l)
VFA	4-11	0.35	3015	5-3	0.08	3800	1-1-14	0.57	2410	5-3-0	0.08	3800
LCFA	4-0	0.42	353	4-0	0.42	353	4-2-1	0.48	332	4-0-1	0.44	350
NH_4^+	10-4	0.34	515	11-4	0.31	528	2-0-10	0.51	438	11-4-0	0.31	528
Test	LV	R^2	RMSEP (mg/l)	LV	R^2	RMSEP (mg/l)	LV	R^2	RMSEP (mg/l)	LV	R^2	RMSEP (mg/l)
VFA	4-11	0.64	1694	5-3	0.74	1481	1-1-14	0.56	2674	5-3-0	0.74	1481
LCFA	4-0	0.52	275	4-0	0.52	275	4-2-1	0.42	317	4-0-1	0.52	293
NH_4^+	10-4	0.29	401	11-4	0.4	370	2-0-10	0.55	338	11-4-0	0.4	370

In table 4, for all parameters, the model results showed that the global approach always provided the best models in cross-validation. However, in prediction, the sequential approach provided better prediction for all parameters, having higher R^2 and smaller errors, with one exception. Both approaches came together when models were parsimonious as in the case of LCFA prediction by X_1 & X_2 blocks. By analyzing Måge plots of some of these models (Figure 6), it was noted that the resulting LVs in the sequential approach was never very far from the combination of LVs selected in the global approach. It could be that, the combination with the lowest RMSECV, highlighted by the global approach, is not always the best solution. It may be necessary to explore a set of possible

minimums.



(a) **(b)**

Figure 6: Måge plots (a) for NH_4^+ prediction by blocks X_1 & X_2 (b) and LCFA prediction by blocks X_1 & X_2 & X_3.
(Blue: global approach result. Green: sequential approach result)

Independently of the approach used, the best SO-PLS models were discussed compared to mono-block PLS models.

For VFA, the 2-block (X_1 & X_2) model performances slightly improved compared to the best model mono-block model with $R_{ms}(\lambda)$ (Table 4). Although R^2 was similar (0.73 and 0.74), RMSEP errors improved by 7%, from 1570mg/l to 1481mg/l. For LCFA, the model was parsimonious and did not involve $R_{ss}(\lambda)$; which is coherent with the previous mono-block results where there was no correlation. NH_4^+ prediction did not improve compared to the previous mono-block models.

While the 2-block (X_1 & X_2) model performed well (Table 4) compare to X_3 for VFA prediction (Table 3), it was not the case for LCFA and NH_4^+. Indeed, VFA prediction with the two decomposed polarized signals improved in comparison with its prediction with the total signal $R_{bs}(\lambda)$; with R^2 from 0.55 to 0.74 and RMSEP from 2737mg/l to 1481mg/l (improvement by 46%). However, predictions of NH_4^+ and LCFA with $R_{bs}(\lambda)$ were still more accurate than predictions with these 2-block models.

SO-PLS models were also performed with the three signals (X_1 & X_2 & X_3) of the polarized probe. The general observation of the results of these 3-block models is that they were parsimonious and less interesting in prediction than the previous 2-block models. However, their cross-validation provided a better result which is interesting because the same CV-blocks were used in all models. Depending on the parameters, LVs of the best models were selected either from blocks X_1 & X_2 or from blocks X_1 & X_3. A multi-block model with all three signals might reproduce this same parsimonious pattern. It would be more interesting to focus on synergies between pairs of signals from the polarized probe.

Therefore, several 2-block models were performed with different combinations of the signals from the polarized probe: X_1 & X_3 and X_2 & X_3. The decomposition of the signal allowed focusing on combination of light scattering that can be related to variations of each parameter in the infrared

spectra. SO-PLS model results of these 2-block combinations are summarized in Table 5. Again, both approaches (global and sequential) were tested.

Table 5: SO-PLS models results based on infrared signals from the polarized probe

	Global approach			**Sequential approach**			**Global approach**			**Sequential approach**		
Training	X_1 & X_3			X_1 & X_3			X_2 & X_3			X_2 & X_3		
	LV	R^2	RMSECV (mg/l)	LV	R^2	RMSECV (mg/l)	LV	R^2	RMSECV (mg/l)	LV	R^2	RMSECV (mg/l)
VFA	1-12	0.59	2351	5-3	0.08	3810	1-11	0.59	2361	3-4	0.1	3700
LCFA	4-1	0.44	350	4-1	0.44	350	4-4	0.38	362	7-5	0.41	352
NH_4^+	3-10	0.48	451	11-8	0.43	471	2-10	0.45	466	6-10	0.42	488
Test	LV	R^2	RMSEP (mg/l)	LV	R^2	RMSEP (mg/l)	LV	R^2	RMSEP (mg/l)	LV	R^2	RMSEP (mg/l)
VFA	1-12	0.55	2675	5-3	0.74	1481	1-11	0.54	2711	3-4	0.75	1655
LCFA	4-1	0.52	293	4-1	0.52	293	4-4	0.46	257	7-5	0.42	199
NH_4^+	3-10	0.58	381	11-8	0.43	454	2-10	0.46	400	6-10	0.52	340

In these results, the sequential approach again produced better models than the global approach except for NH_4^+ prediction with X_1 & X_3. There were no parsimonious models in any approaches. Regardless of the approach and the parameter to be predicted, results were very similar for X_1 & X_3 and X_2 & X_3 (Table 5) and also with X_1 & X_2 (Table 4). For example, VFA prediction with either X_1 & X_3 or X_1 & X_2 provided the same results with the same number of LVs. Only slight differences were observed mainly in the RMSEP errors. This showed that only two signals from the polarized probe were needed to capture relevant information on the variations of these parameters. At this point, from all models developed with the polarized probe (including mono-block models), NH_4^+ prediction was better with X_1 & X_3 (Table 5), VFA prediction was better with X_1 & X_3 (Table 5) or X_1 & X_2 (Table 4) and LCFA with X_3 (Table 3).

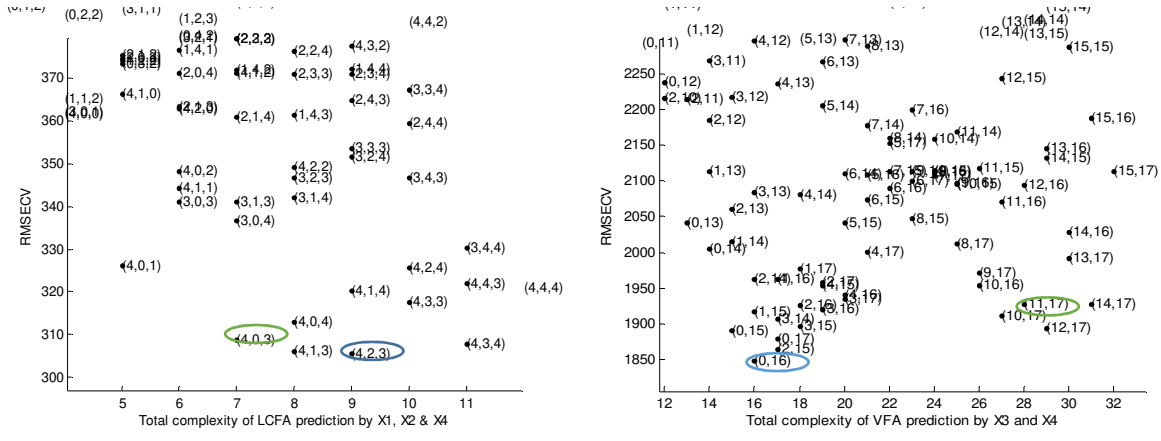
4.4.2. SO-PLS models with the polarized probe and the immersed probe

SO-PLS models were next developed on combinations of signals from the two probes to analyze their joint contribution in the prediction of these stability indicators as follows: $R_{ms}(\lambda)$, $R_{ss}(\lambda)$ and $R(\lambda)$ (X_1 & X_2 & X_4) and, $R_{bs}(\lambda)$ and $R(\lambda)$, (X_3 & X_4). The results of these models are summarized in Table 6 below.

Table 6: SO-PLS models results based on infrared signals the two probes

	Global approach			**Sequential approach**			**Global approach**			**Sequential approach**		
Training	X_1 & X_2 & X_4			X_1 & X_2 & X_4			X_3 & X_4			X_3 & X_4		
	LV	R^2	RMSECV (mg/l)	LV	R^2	RMSECV (mg/l)	LV	R^2	RMSECV (mg/l)	LV	R^2	RMSECV (mg/l)
VFA	0-2-16	0.81	1690	5-3-15	0.74	1900	0-16	0.81	1711	11-17	0.76	1814
LCFA	4-2-3	0.58	297	4-0-3	0.56	304	8-5	0.61	295	7-3	0.56	309
NH_4^+	2-2-10	0.45	464	11-4-11	0.46	460	10-9	0.63	382	12-10	0.50	446
Test	LV	R^2	RMSEP (mg/l)	LV	R^2	RMSEP (mg/l)	LV	R^2	RMSEP (mg/l)	LV	R^2	RMSEP (mg/l)
VFA	0-2-16	0.68	2420	5-3-15	0.67	2577	0-16	0.68	2366	11-17	0.70	2188
LCFA	4-2-3	0.49	244	4-0-3	0.44	201	8-5	0.37	126	7-3	0.53	111
NH_4^+	2-2-10	0.68	496	11-4-11	0.52	482	10-9	0.65	516	12-10	0.65	558

In this second part, the global approach also provided the best models in cross-validation. However, the trend observed in prediction was different than before. Predictions were better in the global approach for the 3-block models (X_1 & X_2 & X_4) while the sequential approach performed well with the 2-block models (X_3 & X_4). In the sequential approach, the optimization is performed one block after the other. Therefore, it might be difficult to take into consideration the pairing effect of the additional blocks. This was not previously spotted because of parsimonious models obtained with the previous 3-block models. This effect does not appear on the two-block models, hence the results obtained for X_3 & X_4 with the sequential approach**.** As before, both solutions were fairly close of each other as shown on some Måge plots from these models (Figure 7). In the particular case of VFA, Måge plots showed two minimums and the chosen number of LVs respectively corresponded to these two minimums (Figure 7b). The data structure and the synergy of the blocks used in the models can influence the choice of one approach compared to the other.



(a) **(b)**

Figure 7: Måge plots (a) for LCFA prediction by blocks X_1 & X_2 & X_4 (b) and VFA prediction by blocks X_3 & X_4. (Blue: global approach result. Green: sequential approach result)

The best models were also discussed independently of the approach used. For all parameters, predictions with X_3 & X_4, $R_{bs}(\lambda)$ and $R(\lambda)$, were better than predictions with X_1 & X_2 & X_4 including the decomposed polarized signal $R_{ss}(\lambda)$ and $R_{ms}(\lambda)$.

The 2-block SO-PLS models of $R_{bs}(\lambda)$ and $R(\lambda)$ have really improved compared to their mono-block models. For VFA, R^2 improved from 0.55 and 0.68 to 0.70. RMSEP went from 2737 mg/l and 2366 mg/l in mono-block models to 2188 mg/l in the 2-block model, representing an improvement of 20% and 7% for $R_{bs}(\lambda)$ and $R(\lambda)$ respectively. NH_4^+ was previously better predicted by the immersed probe with R^2 of 0.63 and RMSEP of 516 mg/l. The 2-block model provided R^2 of 0.68 and a lower RMSEP of 496mg/l. For LCFA, the prediction was less accurate with the 2-block models (R^2 of 0.53) compared to the best mono-block model (R^2 of 0.62). However, the 2-block model provided a lower RMSEP (111mg/l) than the best mono-block model (188mg/l).

To go further in the exploration, 3-block models were developed with the other combinations of the polarized probe (X_1 & X_3 and X_2 & X_3) and the immersed probe. The results obtained from these models, namely X_1 & X_3 & X_4 and X_2 & X_3 & X_4, are summarized in Table 7 below. Except for VFA, the global approach provided better predictions of the parameters. The results were also similar with prediction with the three blocks X_1 & X_2 & X_4 (Table 6) and some models were parsimonious.

Table 7: SO-PLS models results based on infrared signals of the two probes

	Global approach			**Sequential approach**			**Global approach**			**Sequential approach**		
Training	X_1 & X_3 & X_4			X_1 & X_3 & X_4			X_2 & X_3 & X_4			X_2 & X_3 & X_4		
	LV	R^2	RMSECV (mg/l)	LV	R^2	RMSECV (mg/l)	LV	R^2	RMSECV (mg/l)	LV	R^2	RMSECV (mg/l)
VFA	0-0-16	0.81	1711	5-3-0	0.08	3810	2-0-16	0.81	1690	3-4-2	0.12	3931
LCFA	4-2-3	0.63	280	4-1-3	0.60	290	4-4-3	0.58	300	7-5-0	0.41	352
NH_4^+	3-9-6	0.55	418	11-8-10	0.44	466	3-5-11	0.57	415	6-10-8	0.57	415
Test	LV	R^2	RMSEP (mg/l)	LV	R^2	RMSEP (mg/l)	LV	R^2	RMSEP (mg/l)	LV	R^2	RMSEP (mg/l)
VFA	0-0-16	0.68	2366	5-3-0	0.74	1481	2-0-16	0.68	2420	3-4-2	0.78	1343
LCFA	4-2-3	0.51	245	4-1-3	0.47	229	4-4-3	0.45	192	7-5-0	0.42	199
NH_4^+	3-9-6	0.62	431	11-8-10	0.47	496	3-5-11	0.63	397	6-10-8	0.54	367

Predictions of some parameters have improved by these 3-block models, using the two probes:

- For VFA, the most interesting model is a 3-block model performed with X_2 & X_3 & X_4 (Table 7), corresponding to polarized signals $R_{ss}(\lambda)$ and $R_{bs}(\lambda)$ and the immersed probe signal $R(\lambda)$.

VFA was predicted with a R^2 of 0.78 and a RMSEP of 1343mg/l. These results represented an improvement of 38.6% compared to the best 2-block model with X_3 & X_4 (R^2 of 0.70 and RMSEP 2188mg/l). Compared to results obtained in similarly diluted conditions (low TS contents), these multi-block models also presented an improvement. Indeed, with raw sewage sludge digestate (with TS < 5%), the models obtained $0.69 \leq R^2 \leq 0.71$ and 160 mg/l ≤ RMSEP ≤ 180 mg/l with VFA ranging from 24mg/l to 1500mg/l [13]. The range of validation for VFA in the present study is 0-10096mg/l. For a similar VFA range (200-13100 mg/l), R^2 of 0.85 and RMSEP of 900mg/l were obtained with a digester fed with maize silage (high TS contents) [39]. Therefore, SO-PLS multi-block analysis has somehow helped to fill the gap between the monitoring of digestion with high TS contents and digestion with low TS contents.

- For NH_4^+, the 3-block model with X_1 & X_2 & X_4 ($R_{ms}(\lambda)$, $R_{ss}(\lambda)$ and $R(\lambda)$) in Table 6 also produced better results compared to all models tested for this parameter.
- For LCFA, only $R_{bs}(\lambda)$ signal provided the most interesting model (Table 3). No combination was able to improve this parameter prediction. There are no studies on digestate matrices for LCFA prediction by NIRS. However, the results obtained in this study were coherent with studies performed on raw sheep milk which has a consistence similar to the digestate's ($0.60 \leq R^2 \leq 0.76$) [40].

These results collectively show the usefulness of multi-block methods for anaerobic digestion process monitoring.

5. CONCLUSION

The application of SO-PLS improved the monitoring of AD process through the prediction of state indicators such as of VFA, LCFA and ammonium. The obtained results support the finding that combining several sources can achieve synergies for an optimized monitoring of the process. With regard to each infrared probe, models were also promising especially for LCFA monitoring with the polarized probe. Polarization light spectroscopy has helped to improve the understanding of the digestate media related to scattering effect. Moreover, the remote probe compared to the immersed has the capacity to avoid saturation and fouling problems which is quite frequent in anaerobic digestion. However, the complementarity of these spectroscopic techniques was highlighted through this study. As multi-block methods are becoming more common in many fields, several possibilities can be considered in AD process monitoring. The monitoring of anaerobic digestion could be improved by integrating in these multi-bock data, chemical data from routine analysis performed on digesters.

Funding source

L. Awhangbo is the beneficiary of a PhD scholarship funded by Irstea and the Bretagne Region. This project was also funded by the French Environment & Energy Management Agency (ADEME) (COMET project N°1606C0010).

References

[1] Kurade, M.B., Saha, S., Salama, E., Patil, S.M., Govindwar, S.P., Jeon B. Acetoclastic methanogenesis led by *Methanosarcina*in anaerobic co-digestion of fats, oil and grease for enhanced production of methane. *Bioresource Technology* 272(2019), 351-359

https://doi.org/10.1016/j.biortech.2018.10.047

[2] Ma, J., Zhao, Q.B., Laurens, L.L., Jarvis, E.E., Nagle, N.J., Chen, S., Frear, C.S. Mechanism, kinetics and microbiology of inhibition caused by long-chain fatty acids in anaerobic digestion of algal biomass. *Biotechnology for biofuels* 8:141 (2015). https://doi.org/10.1186/s13068-015-0322-z

[3] Reeves, J.B., Van Kessel, J.S. Near-infrared spectroscopic determination of carbon, total nitrogen, and ammonium-n in dairy manures. Journal of Dairy Science 83 (8), (2000), 1829-1836. https://doi.org/10.3168/jds.S0022-0302(00)75053-3

[4] Xie, L., Xingqian, Y., Liu, D., Ying, Y. Quantification of glucose, fructose and sucrose in bayberry juice by NIR and PLS. *Food Chemistry*, 114(2009), 1135-1140. https://doi.org/10.1016/j.foodchem.2008.10.076

[5] Roggo, Y., Chalus, P., Maurer, L., Lema-Martinez, C., Edmond, A., & Jent, N. A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. *Journal of Pharmaceutical and Biomedical Analysis, 44*(3), (2007), 683-700. https://doi.org/10.1016/j.jpba.2007.03.023

[6] Lovett, D.K., Deaville, E.R., Givens, D.I., Finlay, M., Owen, E. Near infrared reflectance spectroscopy (NIRS) to predict biological parameters of maize silage: effects of particle comminution, oven drying temperature and the presence of residual moisture. *Animal Feed Science and Technology*, 120(3-4), (2005), 323–332. https://doi.org/10.1016/j.anifeedsci.2005.02.001

[7] Liu, X., Han, L., Yang, Z., Xu, C. Prediction of silage digestibility by near infrared reflectance spectroscopy. *Journal of Animal and Feed Sciences, 17*(4), (2008), 631-639. https://doi.org/10.22358/jafs/66691/2008

[8] Holm-Nielsen, J.B., Andree, H., Lindorfer, H., Esbensen, K.H. Transflexive embedded near infrared monitoring for key process intermediates in anaerobic digestion/biogas production. *Journal of Near Infrared Spectroscopy* 15(2), (2007),123–135. https://doi.org/10.1255/jnirs.719

[9] Lomborg, C.J., Holm-Nielsen, J.B., Oleskowicz-Popiel, P., Esbensen, K.H. Near infrared and acoustic chemometrics monitoring of volatile fatty acids and dry matter during co-digestion of manure and maize silage. *Bioresour. Technol*, 100(5), (2009), 1711-1719. https://doi.org/10.1016/j.biortech.2008.09.043

[10] Jacobi, H.F., Moschner, C.R. Hartung, E. Use of near infrared spectroscopy in monitoring of volatile fatty acids in anaerobic digestion. *Water Sci. Technol*, 60(2), (2009), 339-346. https://doi.org/10.2166/wst.2009.345

[11] Krapf, L. C., 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. *Journal of Near Infrared Spectroscopy, 19*(6), (2011), 479-493. https://doi.org/10.1255/jnirs.960

[12] Krapf, L. C., 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, 129*(2013), 39-50. https://doi.org/10.1016/j.biortech.2012.11.027

[13] Reed, J.P., Devlin, D., Esteves, S.R.R., Dinsdale, R., Guwy, A.J. Performance parameter prediction for sewage sludge digesters using reflectance FT-NIR spectroscopy. *Water Research*, 45(8), (2011), 2463-2472. https://doi.org/10.1016/j.watres.2011.01.027

[14] Menzi, H., Manure management in Europe: Results of a recent survey. In: Martinez, J. (Ed.), Proceeding of the 10th International conference of the RAMIRAN network - Recycling of agricultural, municipal and industrial residues in agriculture, (2002), Slovak Republic, 93-102

[15] W. Saeys, J. Xing, J. De Baerdemaeker, H. Ramon. Comparison of transflectance and reflectance to analyse hog manures. Journal of Near Infrared Spectroscopy, 13(2005), 99-107. https://doi.org/10.1255/jnirs.462

[16] Surowiec, I., Skotare, T., Sjögren, R., Gouveia-Figueira, S., Orikiiriza, J., Bergström, S., . . . Trygg, J. Joint and unique multiblock analysis of biological data – multiomics malaria study. *Faraday Discussions*, 218 (2019), 268-283. https://doi.ord/10.1039/C8FD00243F

[17] Wold, S., Kettaneh, N., Tjessem, K. Hierarchical multiblock PLS and PC models for easier model interpretation and as an alternative to variable selection. *Journal of chemometrics, 10*(5-6), (1996), 463-482. https://doi.org/10.1002/(SICI)1099-128X(199609)10:5/6<463::AID-CEM445>3.0.CO;2-L

[18] Westerhuis, J. A., Kourti, T., MacGregor, J.F. Analysis of multiblock and hierarchical PCA and PLS models. *Journal of Chemometrics: A Journal of the Chemometrics Society* 12.5(1998), 301-321. https://doi.org/10.1002/(SICI)1099-128X(199809/10)12:5<301::AID-CEM515>3.0.CO;2-S

[19] Næs, T., Tomic, O., Mevik, B.H., Martens, H. Path modelling by sequential PLS regression. *Journal of Chemometrics*, *25*(1), (2011), 28-40. https://doi.org/10.1002/cem.1357

[20] Måge, I., Menichelli, E., Næs, T. Preference mapping by PO-PLS: Separating common and unique information in several data blocks, *Food Qual. Pref.* 24(2012), 8–16. https://doi.org/10.1016/j.foodqual.2011.08.003

[21] El Ghaziri, A., Cariou, V., Rutledge, D.N., Qannari, E.M. Analysis of multiblock datasets using ComDim: Overview and extension to the analysis of (K+1) datasets. *Journal of Chemometrics*, *30*(8), (2016), 420-429. https://doi.org/10.1002/cem.2810

[22] Biancolillo, A., Næs, T. Bro, R. Måge, I. Extension of SO-PLS to multi-way arrays: SO-N-PLS, Chemometr. Intell. Lab. Syst. 164 (2017) 113–126. https://doi.org/10.1002/cem.3120 https://doi.org/10.1016/j.chemolab.2017.03.002

[23] Biancolillo, A, Marini, F, Roger, J-M. SO-CovSel: A novel method for variable selection in a

multiblock framework. Journal of Chemometrics. (2019); e3120. https://doi.org/10.1002/cem.3120

[24] Mourant, J.R., Johnson, T.M., Carpenter, S., Guerra, A., Aida, T., Freyer, J.P. Polarized angular-dependent spectroscopy of epithelial cells and epithelial cell nuclei to determine the size scale of scattering structures. *Journal of Biomedical Optics*, *7*(3), (2002), 378-388. https://doi.org/10.1117/1.1483317

[25] Jacques, S.L., Roman, J.R. and Lee, K. Imaging superficial tissues with polarized light. *Lasers in Surgery and Medicine: The Official Journal of the American Society for Laser Medicine and Surgery* 26.2 (2000), 119-129. https://doi.org/10.1002/(SICI)1096-9101(2000)26:2<119::AID-LSM3>3.0.CO;2-Y

[26] APHA Standard Methods for the Examination of Water and Wastewater (22nded.), American Public Health Association, American Water Works Association, Water Environment Federation. 2012

[27] Bendoula, R., Gobrecht, A., Moulin, B., Roger, J.M., Bellon-Maurel, V. Improvement of the chemical content prediction of a model powder system by reducing multiple scattering using polarized light spectroscopy. *Applied Spectroscopy,* 69(1), (2015), 95-102 https://doi.org/10.1366/14-07539

[28] Savitzky, A. & Golay, M. J. Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry*, 36(8), (1964), 1627–1639. https://doi.org/10.1021/ac60214a047

[29] Zeaiter, M., Roger, J.M., Bellon-Maurel, V., Robustness of models developed by multivariate calibration. Part II: the influence of pre-processing methods. *TrAC,Trends Anal. Chem*. 24 (2005), 437–445. https://doi.org/10.1016/j.trac.2004.11.023

[30] Barnes, R., Dhanoa, M., Lister, J.,. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Appl. Spectrosc*. 43(1989), 772–777. https://doi.org/10.1366/0003702894202201

[31] Biancolillo, A. Naes, T. Chapter 6 - The sequential and orthogonalised PLS regression (SO-PLS) for multi-block regression: theory, examples and extensions, in: M. Cocchi (Ed.), Data Handling in Science and Technology, Vol 31, Elsevier, Amsterdam, (2019), pp. 157-177. https://doi.org/10.1016/B978-0-444-63984-4.00006-5

[32] Niimi, J., Tomic, O., Næs, T., Jeffery, D. W., Bastian, S.E.P., Boss, P.K. Application of sequential and orthogonalised-partial least squares (SO-PLS) regression to predict sensory properties of Cabernet Sauvignon wines from grape chemical composition. Food Chemistry, 256(2018), 195-202. https://doi.org/10.1016/j.foodchem.2018.02.120

[33] Hielscher, A.H., Eick, A.A., Mourant, J.R., Shen, D., Freyer, J.P., Bigio, I.J. Diffuse backscattering Mueller matrices of highly scattering media. Opt Express, 1(13), (1997), 441-453. https://doi.org/10.1364/OE.1.000441

[34] Lalman, J., Bagley, D.M. Effects of C18 long chain fatty acids on glucose, butyrate and hydrogen degradation. *Water Res.*, 36(13), (2002), 3307–3313 https://doi.org/10.1016/S0043-1354(02)00014-3

[35] Rutledge, D. N., Barros, A. S. The Durbin-Watson statistic as a morphological estimator of information content, Analytica Chimica Acta, (2002) 446, 279-294. https://doi.org/10.1016/S0003-2670(01)01555-0

[36] Schelten, J., Schmatz, W. Multiple-scattering treatment for small-angle scattering problems. *Journal of Applied Crystallography, 13*(4), (1980), 385-390. https://doi.org/10.1107/S0021889880012356

[37] Palatsi, J., Affes, R., Fernandez, B., Pereira, M.A., Alves, M.M., Flotats, X. Influence of adsorption and anaerobic granular sludge characteristics on long chain fatty acids inhibition process. *Water Res*, 46(16), (2012), 5268-5278. https://doi.org/10.1016/j.watres.2012.07.008

[38] Pitk, P., Palatsi, J., Kaparaju, P., Fernández, B., Vilu, R. Mesophilic co-digestion of dairy manure and lipid rich solid slaughterhouse wastes: Process efficiency, limitations and floating granules formation. *Bioresour. Technol,* 166(2014), 168-177. https://doi.org/10.1016/j.biortech.2014.05.033

[39] Krapf, L.C., Heuwinkel, H., Schmidhalter, U., Gronauer, A. The potential for online monitoring of short-term process dynamics in anaerobic digestion using near-infrared spectroscopy. *Biomass and Bioenergy, 48*(2013), 224-230

[40] Lužová, T., Šustová, K., KuchtÍk, J., Mlček, J., Vorlová, L., Sumczynski, D. Determination of fatty acid content in sheep milk by means of near infrared spectroscopy. *Acta Veterinaria Brno*, 83(2014), S27-S34. https://doi.org/10.2754/avb201483S10S27