Article

Improvements in the Robustness of Mid-Infrared Spectroscopy Models against Chemical Interferences: Application to Monitoring of Anaerobic Digestion Processes

Magida Zeaiter 1, Éric Latrille 2,3, Pascal Gras 2, Jean-Philippe Steyer 2, Véronique Bellon-Maurel 1 and Jean-Michel Roger 1,3,*

1 ITAP, INRAE, Institut Agro, University Montpellier, 34060 Montpellier, France; magida_z@yahoo.com (M.Z.); veronique.bellon@inrae.fr (V.B.-M.)
2 INRAE, University Montpellier, LBE, 102 Avenue des Etangs, 11100 Narbonne, France; eric.latrille@inrae.fr (É.L.); pascal.gras@inrae.fr (P.G.); jean-philippe.steyer@inrae.fr (J.-P.S.)
3 ChemHouse Research Group, 34196 Montpellier, France

* Correspondence: jean-michel.roger@inrae.fr

Abstract: The monitoring and control of bioprocesses rely on the measurement of the main metabolite concentrations. To this end, infrared spectroscopy (IR) is a good candidate with which to perform rapid and non-destructive measurements. However, IR-based measurements rely on a calibration step linking the measured spectra to the concentrations of the compounds of interest. This calibration may suffer with problems of robustness when the measuring conditions change, such as when some chemicals not present in the calibration spectra are added when using the IR sensor. In this study, a method based on orthogonal projection, dynamic orthogonal projection (DOP), was tested for its ability to cope with the robustness problem caused by the addition of ammonia in a pilot-scale anaerobic digester, whose volatile fatty acid concentrations were monitored by mid-IR spectrometry. The results demonstrate that DOP has significant potential as a form of process analytical technology.

Keywords: bioprocess monitoring; infrared spectrometry; chemometrics; orthogonal projection; robustness

1. Introduction

The advanced monitoring and control of bioprocesses rely on the use of appropriate sensors to measure the concentrations of major metabolites with sufficient accuracy and long-term stability under minimum maintenance. Infrared (IR) spectroscopic sensors offer many advantages in this field. Near-infrared spectrometry (NIRS) is the most frequently used technique online, because the wavelength range (700–2500 nm) allows the light to penetrate deeply into the sample. Wang et al. [1] provide a comprehensive review of NIRS online applications for liquid media. These include alcoholic and non-alcoholic beverages, oils and milk. The objectives of these applications include quality assessment, classification, and adulteration control. However, since NIRS is strongly affected by water, mid-infrared spectrometry (MIRS) offers an interesting alternative to NIRS for aqueous media. Karoui et al. [2] provide a review of MIRS applications in food, including milk, meat, fish, oil, and cereals. They conclude that MIRS is a good candidate to be used online, not only for food processing. In [3], a review of the application of MIRS to monitoring processes reported that MIRS can predict glucose, ethanol, methanol, glucuronic acid, and ammonium concentrations with fair accuracy. Consequently, MIRS is a good potential process analytical technology (PAT).

However, the calibration of IR-based sensors for bioprocess monitoring is still a challenging robustness problem due to the overlapping absorbance features and the metabolic interactions and correlations among the major metabolites. The former problem may cause
some difficulties in retrieving the part of the spectrum that is related to the compound of interest. The latter problem may lead to the building of a calibration of a compound A, on the basis of the spectral response of a compound B, correlated to A. Partial least-squares regression (PLSR) is often used to calibrate spectra with overlapping absorbance features. However, only few methods have pointed out specifications for model calibration to circumvent the correlation problem for reactive mixtures [4,5] and cell-culture reactions [6,7]. The robustness of PLSR models used with infrared spectroscopy data was widely addressed in the literature, from a theoretical point of view [8–13]. Various factors may affect this robustness. Different pre-processing methods were developed to correct physical- and chemical-factor interference. Among these, approaches based on orthogonal projection methods make it possible to subtract the unwanted signal from the spectra before calibration [14,15].

Anaerobic digestion (AD) is a biological process during which organic carbon is converted through oxidation-reduction reactions to both its most oxidized state (CO$_2$) and its most reduced form (CH$_4$) [16]. The methane produced is a versatile energy carrier that can be used for generating power and heat, injecting into the gas grid, or use as a fuel for vehicles. In Europe, incentive policies have led to the emergence of a biogas industry, mainly based on agricultural feedstocks (agricultural residues, energy crops, and catch crops) which are the largest available deposits of organic matter [17,18]. However, AD processes are very sensitive, and they can face a large number of disturbances due the fluctuating characteristics and quantities of the waste they handle. Thus, the monitoring of AD processes is an imperative task in order to ensure optimized operations and to prevent failures and serious consequences during the running of plants [19].

In this paper, dynamic orthogonal projection [20] is applied to circumvent the problem of overlapping absorbance features in the MIR spectra collected on a pilot-scale anaerobic digester to monitor volatile fatty acid (VFA) concentrations.

2. Materials and Methods
2.1. Experimental Procedure
2.1.1. Process

A 1-cubic-meter pilot-scale AD fixed-bed reactor was used to assess the response to organic overloads and toxicant shock loads. The prime objective was to use advanced online instrumentation available in the process to analyze, in dynamic conditions, the effects of disturbances on the overall efficiency of the process. Among the different possible disturbances (e.g., modification of the feed characteristics, presence of a toxicant, change in process temperature, etc.), only the effect of ammonia addition on the measurement of total volatile fatty acids using MIRS was studied in the present work.

The effluents used were raw industrial wine wastewaters obtained from local wineries in the area of Narbonne, France [21]. The process involved an up-flow anaerobic fixed-bed reactor made of a circular column 3.5 m in height, 0.6 m in diameter, and with a useful volume of 0.948 m$^3$ [22]. The reactor was highly instrumented, with the following measurements available online every two minutes: input and recirculation liquid flow rates, pH of the reactor and of the input wastewater, temperature, biogas output flow rate, CO$_2$, CH$_4$ and H$_2$ composition in the gas phase, and total organic carbon (TOC) in the reactor. Every half hour, measurements were taken using a titrimetric sensor and a MIR spectrometer [23] to measure: total volatile fatty acids (VFA), soluble chemical oxygen demand (COD), bicarbonate concentrations, and total and partial alkalinity in the liquid phase. Since VFA are very important intermediates in the AD reaction scheme and can inhibit the overall process, their measurement was considered in this study, together with their robustness to the ammonia addition.

2.1.2. Addition of Ammonia in the Reactor

Ammonia-addition experiments were performed to assess robustness of the fixed-bed reactor when facing toxicant shock loads. In these experiments, ammonia from NH$_4$Cl was
chosen as the toxic compound, since it is known to be a strong inhibitor of methanogenesis [24–26].

During the process, different additions of ammonia took place at different times (days 0.0, 5.7, and 13.7). Table 1 and Figure 1 show the various events that occurred during this 15-day period.

Table 1. List of the events during the batch process and their impact on the VFA prediction.

| Time (Days) | Addition of Chemicals | Impact on the Predictions |
|-------------|-----------------------|---------------------------|
| 0           | Addition of NH₄Cl (low conc.) | Moderate negative bias |
| 4.0         | Addition of NaOH       | Weak positive bias        |
| 5.7         | Addition of NH₄Cl (medium conc.) | Strong negative bias |
| 9.3         | Discharge and water circulation | None                  |
| 10.0        | Re-charge with wine waste | None                    |
| 11.5        | Stopping of the reactor  | None                    |
| 13.0        | Restart of the reactor  | None                    |
| 13.7        | Addition of NH₄Cl (high conc.) | Strong negative bias |

Figure 1. VFA concentration evolution measured by the reference titrimetric sensor (blue) and predicted by the raw PLSR (red). Vertical lines indicate the main events (Cf Table 1). Horizontal arrows indicate the measurements used for calibration and testing of the PLSR model.

2.1.3. MIR Spectra Collected on the Bioreactor

MIR spectrometer (Nicolet AVATAR 360) was used to collect transmission spectra using a recirculation loop and a 50-micrometer flow cell. The resolution was set to 4 cm⁻¹. The fingerprint range (1572–1000 cm⁻¹) was kept, corresponding to 150 wavenumbers. A total of 616 spectra and corresponding VFA concentrations (mg/L) were measured online on equally spaced instants, over the course of 15 days.
2.2. Chemometrics

2.2.1. Notations

Capital bold characters are used for matrices, e.g., $X$; small bold characters for column vectors, e.g., $x_j$ denote the $j$th column of $X$; row vectors are denoted by the transpose notation, e.g., $x_i^T$ denotes the $i$th row of $X$; non-bold characters are used for scalars, e.g., indexes, as $i$. When needed for clarity purposes, matrix dimensions are indicated as $X_{(n,p)}$, where $n$ is the number of rows and $p$ the number of columns.

Model calibration

From the acquired dataset, two subsets were drawn (see Figure 1):

- A calibration set $(X_0, y_0)$ made up of 100 spectra and VFA concentrations obtained during a sequence where the process was in a standard mode, i.e., between approx. days 9 and 13 (see Table 1 and Figure 1). This period corresponded to the restart of the reactor, so that the $y_0$ values covered a wide range of VFA concentrations.
- A test set $(X_1, y_1)$ was made up of all the acquired samples, containing 616 couples of spectra and VFA concentrations. This test set included various states of functioning, including normal states and ammonia-addition events.

VFA estimation from MIR spectra was carried out by means of a PLSR [27], without any preprocessing of the spectra apart from column mean centering. PLSR estimates a response $y$ by a linear combination of the columns of $X$: $\hat{y} = Xb + r$. Since columns of $X$ are highly correlated, an ordinary least-squares regression cannot be used. Instead, PLSR proceeds by first calculating $a$ latent variables, which are linear combinations $\{u_1, u_2, \ldots, u_a\}$ of the columns of $X$, which maximize $\text{cov}^2(Xu_i, y)$ with the constraint $Xu_i \perp Xu_j$, for $i \neq j$. Next, a linear regression is calculated between $(Xu_i)$ and $y$. Subsequently, the vector $b$ is obtained by merging the latent variable loadings $U$ and the regression coefficients of the linear regression. To set the number of latent variables $a$, a 2-random-blocks cross-validation was performed and repeated 10 times. The value of $a$ was set as the one yielding the lower RMSECV. The raw PLSR model was thus built on 5 latent variables and yielded the predictions reported in Figure 1.

2.2.2. Dynamic Orthogonal Projection

When a perturbation $\delta x$ is added to a spectrum $x$, an error $\delta \hat{y} = \delta x^T b$ is added to the estimation of the response $\hat{y}$. Orthogonal projection (OP) methods aim to reduce this error by constraining the PLSR to build a $b$ orthogonal to $\delta x$, thus lowering $\delta \hat{y}$. OP methods start by identifying the subspace spanned by the $\delta x$ (detrimental subspace). An orthonormal basis $P_{(p)}$ of this subspace is calculated, and then $X_{(np)}$ is projected onto the subspace orthogonal to $P$, resulting in the corrected spectra:

$$\tilde{X}_{(np)} = X - XPP^T$$

This projection acts by removing linear combinations of columns of $X$, i.e., it removes a spectral subspace from the space spanned by $X$. The estimation of $P$ is the key to orthogonal correction. This estimation commonly uses a set of dedicated experiments, incorporating systematic variations. Let $D_{(np)}$, be a matrix of $m$ spectra generated by the detrimental variations only. A PCA is carried out on $D$ and the first $k$ loadings are inserted in $P$. Different methods are used to build the $D$ matrix. In the IIR method [28], the matrix $D$ is built with samples that do not contain the component of interest but include variations due to the influencing factors. Its main disadvantage is that it requires two matrices to correct the spectra, which are often not available. In the EPO method [12], the matrix $D$ is built using a small set of appropriate samples measured at different levels of the influence factor. This method does not require the $y$ reference measurements. However, only known influence factors (temperature, scale, instruments, etc.) can be considered, while unknown factors cannot.
The DOP method [20] was developed to cope with variations due to unknown factors. The specific advantage of DOP over other OP methods is that it does not require standard samples, nor experimental designs. It uses a few control points, for which the true value of \( y \) is known. Let \((X_0, y_0)\) be a calibration set containing \( n_0 \) samples. Let \( y_r \) be the values of \( y \) measured at \( n_r \) control points and \( X_r \) the corresponding acquired spectra. The DOP method runs as follows:

- First, estimate the ideal spectra \( Z_r \), which should be measured in the absence of influencing factors. This produces two matrices \((X_r, Z_r)\) similar to those measured in the case of standard samples used for calibration transfer or calibration qualification test, except that these samples are rarely available for online applications.

- Second, compute the \( D \) matrix as the difference between the measured spectra and the ideal spectra:
  \[
  D = X_r - Z_r \tag{2}
  \]

- Third, extract the \( k \) first loadings of a PCA computed on \( D \) and insert them in \( P \).

- Project \( X_0 \) orthogonal to \( P \) and recalibrate the model.

Estimating \( Z_r \) is the cornerstone of the DOP approach. This estimation must be performed using only the calibration set \((X_0, y_0)\), and the control points \( y_r \). The most straightforward method consists in computing a linear combination \( A_{(n_r,n_0)} \) of \( y_0 \) that estimates \( y_r \) and then to applying it to \( X_0 \):

- Calculate \( A \) so that:
  \[
  y_r \approx Ay_0 \tag{3}
  \]

- Apply \( A \) on \( X_0 \):
  \[
  Z_r = AX_0 \tag{4}
  \]

The calculation of \( A \) can be performed in different ways. The kNN (nearest-neighbors) method consists of calculating the distance between the sample to be estimated and all the samples from the calibration database to select the \( k \) nearest ones used for estimation. Different variants of this method can be used, such as giving more weight to the closest samples. This method is commonly used in classification. This is a simple method that makes no statistical assumption about the normality of distribution. The kernel methods generalize the kNN; they use a kernel function to estimate the density function of a population, which can then be used to weigh an estimate. Different kernel functions can be used: uniform, triangular, Gaussian, etc. Gaussian kernels are commonly used.

In this study, a Gaussian kernel was used, so that the \( A \) matrix was calculated as follows:

\[
A_{ij} = \frac{1}{\varepsilon \sigma_0 \sqrt{2\pi}} e^{-\frac{(y_{0i} - y_{rj})^2}{2\varepsilon^2 \sigma_0^2}} \tag{5}
\]

where \( \sigma_0 \) is the standard deviation of \( y_0 \) and \( \varepsilon \) is a parameter adjusting the width of the kernel. In [20], it is theoretically justified that a good value for \( \varepsilon \) is \( 1/n_0 \), if \( y_0 \) is uniformly distributed. It was verified that this value is adapted to most distributions of \( y_0 \) and it was adopted as a rule of thumb. The same approach was therefore applied in this study.

Since the original calibration database is transformed by means of an orthogonal projection, the correction is embedded into the model [14,15]. Consequently, there is no need to correct new spectra when using the model.

2.2.3. Figures of Merit

The quality of prediction was assessed by means of the following figures of merit:

- Root Mean Square Error of Calibration:
  \[
  RMSEC = \sqrt{\frac{\sum (\hat{y}_i - y_i)^2}{N - LV - 1}} \tag{6}
  \]
• Root Mean Square Error of Cross-Validation:

$$\text{RMSECV} = \sqrt{\frac{\sum (\hat{y}_i - y_i)^2}{N}}$$  \hspace{1cm} (7)

• Root Mean Square Error of Prediction:

$$\text{RMSEP} = \sqrt{\frac{\sum (\hat{y}_i - y_i)^2}{N}}$$  \hspace{1cm} (8)

• Bias:

$$\text{Bias} = \frac{\sum (\hat{y}_i - y_i)}{N}$$  \hspace{1cm} (9)

3. Results and Discussion

The cross-validation of the raw PLSR model exhibited a RMSECV of about 58 mg/L for five latent variables. In [29], tests using regular conditions yielded RMSEP values of 156 mg/L and 235 mg/L for the acetic acid and propionic acid, respectively. In [30], a RMSEP of 372 mg/L was exhibited for the VFA prediction. Considering that the RMSEP is usually higher than the RMSECV, and considering that our samples were dependent, we consider that the performance we observed was in accordance with the literature.

The PLSR model was applied to the entire test set. Its global bias was $-626$ mg/L and the RMSEP was 1310 mg/L. The predictions are reported as a function of time in Figure 1. In the interval between days 9.3 and 13.0, which corresponds more or less to the calibration data, the predictions fit the reference value very well, despite small variations. However, Figure 1 shows that all the ammonia additions (days 0, 5.7 and 13.7) had a very detrimental effect on the model. After each addition, the VFA prediction was strongly negatively biased. The intensity of the bias was related to the amount of ammonia added. The effect on the model was immediate, showing that the model was directly affected by the ammonia spectra.

Figure 2 shows the MIR spectra, as a function of time. The effect of ammonia additions on the spectra is clearly visible. After the addition of ammonia, the intensity of the spectral deformation slowly decreased as the spectral ammonia contribution disappeared. The ammonia had an effect on the model because the ammonia spectrum had peaks in areas where the b-coefficients of the PLSR were non-null. In [31], MIR spectra of various forms of ammonia are analyzed. A peak at $1450 \text{ cm}^{-1}$ is reported as being linked to the symmetric deformation mode $\nu_4$ of $\text{NH}_4^+$. This peak corresponds to the main peak appearing in Figure 2 at each ammonia addition.

In order to correct the PLSR model for the effect of the ammonia absorbance in the spectral region of interest, the DOP method was used. Two correction points were chosen at day 0.2 (spectrum number 10) and time 6.2 (spectrum number 270) to perform the DOP with $\varepsilon = 0.01$ and using the maximal number of components for the orthogonalization, i.e., $k = 1$ and $k = 2$, respectively. The number of latent variables of the DOP-corrected models was set to 4, for both correction points. Figure 3 shows the prediction after DOP application, clearly demonstrating the improvement in the PLSR model predictions resulting from the DOP. After the first DOP correction, it can be observed that the negative bias of the raw model was almost completely corrected by the DOP. After the large addition of ammonia, on day 5.7, it can be observed that the DOP corrected model still suffered from a negative bias, albeit much less intense than in the raw model. After the second DOP correction, the model became completely insensitive to ammonia additions, including a very significant addition of ammonia at day 13.7. It is also noticeable that the model continued to work in the absence of ammonia, i.e., from day 9 to day 13. This shows that the DOP correction does not simply compensate for the influence, as would be the case in a bias/slope correction, but also makes the model independent of the influence, whatever its level.
Figure 2. The MIR spectra collected with the addition of ammonia.

Figure 3. Prediction of the raw PLSR model (red) and of the DOP PLSR model (blue) after correction with DOP at point 1 after first addition of ammonia and point 2 after the second, more concentrated, addition of ammonia. Points A, B, and C correspond to low VFA/low ammonia; medium VFA/low ammonia; and high VFA/high ammonia. Dotted vertical lines show the moments of ammonia addition (cf. Figure 1 and Table 1). Dashed vertical lines show the control points used by DOP.

The implementation of the DOP requires the setting of the value of two parameters: \( \varepsilon \), related to the width of the kernel, and \( k \), the dimension of the subspace removed by the orthogonalization step. In this study, the value of \( \varepsilon \) was fixed following the rule of thumb proposed in [20], and the value of \( k \) was chosen as its maximum value, i.e., the number of available control points. In real industrial applications, these two parameters must be set...
According to each particular situation. The value of \( \epsilon \) depends only on the \( y \)-distribution of the calibration set and can be adjusted when setting up the model. The value of \( k \) must be adjusted each time the DOP correction is performed. This can be achieved through various generic orthogonal projection methods, as described in [15]. The goal of this adjustment is to remove enough dimensions to correct the problem found by the control points, while preserving the predictive power of the model. A method well suited to DOP is to observe the cross-validation error of the PLS as a function of \( k \), and then adopt the largest value of \( k \) that does not alter the model.

After DOP correction, the PLSR model presented a bias of \(-19.1\) mg/L and an RMSEP of \(127\) mg/L that were 30 and 10 times less than those of the raw model, respectively. The effect was spectacular, since DOP correction reduced the bias by a factor of approximately 30 and the RMSEP by a factor of approximately 10. The RMSEP of the corrected model was lower than the values reported in the literature, discussed above.

Figure 4 shows the effect of the DOP correction on the three spectra corresponding to points A, B, and C on Figure 3. These three points correspond to VFA amounts of 623, 1476, and 2060 mg/L, respectively, i.e., to three levels in proportions of approximately 1, 2, and 3. The spectra of points A and B were not affected by the ammonia, while the spectrum of point C was. In Figure 4a, which represents the raw spectra, we can observe that there was no proportionality at all between the three spectra, especially in the range of \(1300-1500\) cm\(^{-1}\). In Figure 4b, we can observe that the three spectra are properly ordered, in a proportional scale, over the whole wavenumber range after the DOP correction. More precisely, the three spectra coincide over a wide range, from \(1000\) to \(1350\) cm\(^{-1}\) and separate in two peaks in the zone of \(1350-1700\) cm\(^{-1}\). It is clear from these figures that the DOP removed all the features that were affected by the ammonia, thus producing very meaningful corrected spectra.

A point of discussion must be addressed regarding spectra pretreatments. First, it should be noted that DOP is not really a form of preprocessing. It is a method that changes the calibration of the PLS, forcing it to produce a model that is insensitive to the problems that arise. Second, preprocessing is generally used to improve the performance of the calibration by removing detrimental information contained in the spectra, such as additive or multiplicative effects. Some authors also claim that pretreatments increase the robustness of the models [32]. This is only true if the effect removed by the preprocessing is the one added by the event causing the robustness problem. For example, if the robustness problem is caused by a change in light scattering, it is likely that a baseline reduction will have a beneficial effect on the robustness. Therefore, preprocessing can address the robustness problems that are known when the calibration is performed. This article addresses a case in which the events causing the robustness problem were not known when the calibration was performed, i.e., ammonia addition. This is why it was decided to not preprocess the spectra at all.

In this study, the power of the DOP correction was demonstrated and described, on the basis on two well-chosen control points. Indeed, they were chosen at two instants just after ammonia addition (days 0 and 5.7). In order to test the influence of the number of control points and of their placements on the process, the following procedure was carried out: \( N \) points were drawn at random across the whole time range; next, the DOP correction was applied using 1, ..., \( N \) points, simulating the real process; and subsequently, the RMSEP was calculated. These steps were repeated 100 times, for \( N \) ranging from 1 to 15. Figure 5 reports the results of these simulations, using boxplots. It can be observed that even for a low number of control points, the DOP produced efficient corrections in the majority of cases. Until \( N = 3 \), some DOP-corrected models were worse than the raw model, indicating that an appropriate choice of control points is important when they are few. It is possible to imagine that if the control points for calculating DOP are determined when there is no difference between the online and calibration conditions, the correction does not work correctly. This possibility becomes less and less likely as \( N \) increases. This confirms, as seen previously, that DOP correction is effective with a small number of points that are
chosen when disturbances are present. This highlights the potential of integrating DOP in the toolbox of process analytical technology [33].

Figure 4. Illustration of the DOP correction. Spectra of the three points, A, B, and C, shown in Figure 3, corresponding to VFA = 623, 1476, and 2060 mg/L. (a) Before DOP correction. (b) After DOP correction.

Figure 5. Summaries of RMSEP obtained after DOP correction, as a function of the number of control points used. Each boxplot represents the distribution of the RMSEP values obtained after 100 repetitions with randomly chosen control points. Horizontal red line indicates the median, while the edges of the boxes are the first and third quartiles. Dashed whiskers extend to the most extreme data not considered as outliers. Horizontal dashed line indicates the RMSEP of the raw PLSR.
4. Conclusions

This study highlighted the potential of infrared spectroscopy for the advanced monitoring of complex biological processes, such as anaerobic digestion. In particular, the proposed usage of the dynamic orthogonal projection method significantly improved the robustness of the calibration model and can circumvent the problem of unexpected changes in process operations (e.g., the presence of reactive compounds not accounted for during the calibration phase). This was demonstrated on a 1-cubic-meter pilot-scale anaerobic digester by online measurements of volatile fatty acid concentrations, which were still accurate despite the presence of ammonia. This creates significant possibilities for the optimization of anaerobic digestion processes. Indeed, organic overloads and toxicant shock loads can damage the performances of digesters and strongly affect biogas production. The proposed approach, by efficiently handling these disturbances, paves the way to industrial applications.

Author Contributions: Data curation, M.Z., É.L., P.G., J.-P.S. and J.-M.R.; formal analysis, M.Z. and J.-M.R.; investigation, J.-P.S.; methodology, J.-M.R.; software, J.-M.R.; supervision, J.-P.S., V.B.-M. and J.-M.R.; validation, É.L., J.-P.S. and V.B.-M.; writing—original draft, M.Z.; writing—review and editing, J.-P.S. and J.-M.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wang, L.; Sun, D.W.; Pu, H.; Cheng, J.H. Quality analysis, classification, and authentication of liquid foods by near-infrared spectroscopy: A review of recent research developments. Crit. Rev. Food Sci. Nutr. 2017, 57, 1524–1538. [CrossRef] [PubMed]
2. Karoui, R.; Downey, G.; Blecker, C. Mid-infrared spectroscopy coupled with chemometrics: A tool for the analysis of intact food systems and the exploration of their molecular Structure–Quality relationships—A review. Chem. Rev. 2010, 110, 6144–6168. [CrossRef] [PubMed]
3. Kornmann, H.; Valentinotti, S.; Duboc, P.; Marison, I.; Von Stockar, U. Monitoring and control of Gluconacetobacter xylinus fed-batch cultures using in situ mid-IR spectroscopy. J. Biotechnol. 2004, 113, 231–245. [CrossRef] [PubMed]
4. Amrhein, M. Reaction and Flow Variants/Invariants for the Analysis of Chemical Reaction Data (No. THESIS); EPFL: Lausanne, Switzerland, 1998.
5. Amrhein, M.; Srinivasan, B.; Bonvin, D.; Schumacher, M.M. Calibration of spectral reaction data. Chemom. Intell. Lab. Syst. 1999, 46, 249–264. [CrossRef]
6. Chung, H.; Arnold, M.A.; Rhiel, M.; Murhammer, D.W. Simultaneous measurements of glucose, glutamine, ammonia, lactate, and glutamate in aqueous solutions by near-infrared spectroscopy. Appl. Spectrosc. 1996, 50, 270–276. [CrossRef]
7. Rhiel, M.; Cannizzaro, C.; Valentinotti, S.; Marison, I.; von Stockar, U. Comprehensive In-Situ Bioreactor Monitoring and Control Based on a Mid-Infrared Spectroscopic Sensor System; EPFL Scientific Publications: Lausanne, Switzerland, 2000.
8. Brown, S.D.; Huwei, T.; Feudale, R. Improving the robustness of multivariate calibrations. In ACS Symposium Series; Oxford University Press: Oxford, UK, 2005; Volume 894, pp. 15–30.
9. Dabros, M. Robust Model Development and Enhancement Techniques for Improved On-Line Spectroscopic Monitoring of Bioprocesses. Ph.D. Thesis, EPFL Scientific Publications, Lausanne, Switzerland, 2008.
10. Zeaiter, M.; Roger, J.M.; Bellon-Maurel, V.; Rutledge, D.N. Robustness of models developed by multivariate calibration. Part I: The assessment of robustness. TrAC Trends Anal. Chem. 2004, 23, 157–170. [CrossRef]
11. 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. 2005, 24, 437–445. [CrossRef]
12. Roger, J.M.; Chauchard, F.; Bellon-Maurel, V. EPO–PLS external parameter orthogonalisation of PLS application to temperature-independent measurement of sugar content of intact fruits. Chemom. Intell. Lab. Syst. 2003, 66, 191–204. [CrossRef]
13. Chauchard, F.; Roger, J.M.; Bellon-Maurel, V. Correction of the temperature effect on near infrared calibration—Application to soluble solid content prediction. J. Near Infrared Spectrosc. 2004, 12, 199–205. [CrossRef]
14. Boulet, J.C.; Roger, J.M. Pretreatments by means of orthogonal projections. Chemom. Intell. Lab. Syst. 2012, 117, 61–69. [CrossRef]
15. Roger, J.M.; Boulet, J.C. A review of orthogonal projections for calibration. J. Chemom. 2018, 32, e3045. [CrossRef]
16. Brémond, U.; de Buyer, R.; Steyer, J.P.; Bernet, N.; Carrere, H. Biological pretreatments of biomass for improving biogas production: An overview from lab scale to full-scale. Renew. Sustain. Energy Rev. 2018, 90, 583–604. [CrossRef]

17. Weiland, P. Biogas production: Current state and perspectives. Appl. Microbiol. Biotechnol. 2010, 85, 849–860. [CrossRef] [PubMed]

18. Brémond, U.; Bertrandias, A.; Steyer, J.P.; Bernet, N.; Carrere, H. A vision of European biogas sector development towards 2030: Trends and challenges. J. Clean. Prod. 2021, 287, 125065. [CrossRef]

19. Jimenez, J.; Latrille, E.; Harmand, J.; Robles, A.; Ferrer, J.; Gaida, D.; Wolf, C.; Mairet, F.; Bernard, O.; Alcaraz-Gonzalez, V.; et al. Instrumentation and control of anaerobic digestion processes: A review and some research challenges. Rev. Environ. Sci. Bio/Technol. 2015, 14, 615–648. [CrossRef]

20. Zeaiter, M.; Roger, J.M.; Bellon-Maurel, V. Dynamic orthogonal projection. A new method to maintain the on-line robustness of multivariate calibrations. Application to NIR-based monitoring of wine fermentations. Chemom. Intell. Lab. Syst. 2006, 80, 227–235. [CrossRef]

21. Batstone, D.J.; Steyer, J.P. Use of modelling to evaluate best control practice for winery-type wastewaters. Water Sci. Technol. 2007, 56, 147–152. [CrossRef]

22. Steyer, J.P.; Bouvier, J.C.; Conte, T.; Gras, P.; Soubie, P. Evaluation of a four year experience with a fully instrumented anaerobic digestion process. Water Sci. Technol. 2002, 45, 495–502. [CrossRef]

23. Steyer, J.P.; Bouvier, J.C.; Conte, T.; Gras, P.; Harmand, J.; Delgenes, J.P. On-line measurements of COD, TOC, VFA, total and partial alkalinity in anaerobic digestion processes using infra-red spectrometry. Water Sci. Technol. 2002, 45, 133–138. [CrossRef]

24. Koster, I.W.; Lettinga, G. The influence of ammonium-nitrogen on the specific activity of pelletized methanogenic sludge. Agric. Wastes 1984, 9, 205–216. [CrossRef]

25. Koster, I.W.; Lettinga, G. Anaerobic digestion at extreme ammonia concentrations. Biol. Wastes 1988, 25, 51–59. [CrossRef]

26. Hansen, K.H.; Angelidaki, I.; Ahring, B.K. Anaerobic digestion of swine manure: Inhibition by ammonia. Water Res. 1998, 32, 5–12. [CrossRef]

27. Geladi, P.; Kowalski, B.R. Partial least-squares regression: A tutorial. Anal. Chim. Acta 1986, 185, 1–17. [CrossRef]

28. Hansen, P.W. Pre-processing method minimizing the need for reference analyses. J. Chemom. A J. Chemom. Soc. 2001, 15, 123–131. [CrossRef]

29. Falk, H.M.; 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, 237–249. [CrossRef]

30. Wolf, C.; Gaida, D.; Rehorek, A.; Bongards, M. Online monitoring of AD processes using a fully automated, low maintenance mid-infrared (MIR) measurement system. In Proceeding of the 2nd International Conference Biogas Science 2014, Vienna, Austria, 26–30 October 2014; Volume 2014, pp. 24–25.

31. Max, J.J.; Chapados, C. Aqueous ammonia and ammonium chloride hydrates: Principal infrared spectra. J. Mol. Struct. 2013, 1046, 124–135. [CrossRef]

32. de Noord, O.E. The influence of data preprocessing on the robustness and parsimony of multivariate calibration models. Chemom. Intell. Lab. Syst. 1994, 23, 65–70. [CrossRef]

33. Challa, S.; Potumarthi, R. Chemometrics-based process analytical technology (PAT) tools: Applications and adaptation in pharmaceutical and biopharmaceutical industries. Appl. Biochem. Biotechnol. 2013, 169, 66–76. [CrossRef]