



**HAL**  
open science

## On-site substrate characterization in the anaerobic digestion context: A dataset of near infrared spectra acquired with four different optical systems on freeze-dried and ground organic waste

Margaud Pérémé, Alexandre Mallet, Lorraine Awhangbo, Cyrille Charnier, Jean-Michel Roger, Jean-Philippe Steyer, Éric Latrille, Ryad Bendoula

### ► To cite this version:

Margaud Pérémé, Alexandre Mallet, Lorraine Awhangbo, Cyrille Charnier, Jean-Michel Roger, et al.. On-site substrate characterization in the anaerobic digestion context: A dataset of near infrared spectra acquired with four different optical systems on freeze-dried and ground organic waste. *Data in Brief*, 2021, 36, pp.107126. 10.1016/j.dib.2021.107126 . hal-03260913

**HAL Id: hal-03260913**

**<https://hal.inrae.fr/hal-03260913>**

Submitted on 15 Jun 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License



## Data Article

# On-site substrate characterization in the anaerobic digestion context: A dataset of near infrared spectra acquired with four different optical systems on freeze-dried and ground organic waste



Margaud Pérémé<sup>a,d,e</sup>, Alexandre Mallet<sup>a,b,c,e,\*</sup>, Lorraine Awhangbo<sup>a,e</sup>,  
Cyrille Charnier<sup>c</sup>, Jean-michel Roger<sup>b,e</sup>, Jean-philippe Steyer<sup>a</sup>,  
Éric Latrille<sup>a,e</sup>, Ryad Bendoula<sup>b</sup>

<sup>a</sup> INRAE, Univ Montpellier, LBE<sup>1</sup>, 102 Av des Etangs, Narbonne F-11100, France

<sup>b</sup> INRAE, UMR ITAP, Montpellier University, Montpellier F-34000, France

<sup>c</sup> BIOENTECH Company, Narbonne F-11100, France

<sup>d</sup> ENSCM, 240 Av du professeur Emile Jeanbrau, Montpellier F-34090, France

<sup>e</sup> ChemHouse Research Group, Montpellier F-34000, France

## ARTICLE INFO

## Article history:

Received 14 April 2021

Revised 28 April 2021

Accepted 30 April 2021

Available online 11 May 2021

## Keywords:

Near infrared spectroscopy

Chemometrics

Organic waste

Compact spectrometers

Instrument standardization

## ABSTRACT

The near infrared spectra of thirty-three freeze-dried and ground organic waste samples of various biochemical composition were collected on four different optical systems, including a laboratory spectrometer, a transportable spectrometer with two measurement configurations (an immersed probe, and a polarized light system) and a micro-spectrometer. The provided data contains one file per spectroscopic system including the reflectance or absorbance spectra with the corresponding sample name and wavelengths. A reference data file containing carbohydrates, lipid and nitrogen content, biochemical methane potential (BMP) and chemical oxygen demand (COD) for each sample is also provided. This data enables the comparison of the optical systems for predictive model calibration based for example

DOI of original article: [10.1016/j.wasman.2021.03.045](https://doi.org/10.1016/j.wasman.2021.03.045)

\* Corresponding author.

E-mail address: [alexandre.mallet@bioentech.eu](mailto:alexandre.mallet@bioentech.eu) (A. Mallet).

<sup>1</sup> @LBE\_INRAE

<https://doi.org/10.1016/j.dib.2021.107126>

2352-3409/© 2021 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

on Partial Least Squares Regression (PLS-R) [1], but could be used more broadly to test new chemometrics methods. For example, the data could be used to evaluate different transfer functions between spectroscopic systems [2]. This dataset enabled the research work reported by Mallet et al. 2021 [3].

© 2021 The Author(s). Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

## Specifications Table

Subject	VIS-NIR Spectroscopy
Specific subject area	Optical system comparison, organic waste characterization
Type of data	Table Figure Script
How data were acquired	Data was acquired on the following FT-NIR spectrometers: <ul style="list-style-type: none"> <li>- NIRFlex N-500 FT-NIR (BUCHI, France)</li> <li>- Immersed probe consisting of two fibers (one for illumination, the other for signal collection) plugged to a Rocket FTNIR-L1-025-2TE (Arcoptics, Switzerland)</li> <li>- Polarized light system (PoLiS) plugged to a Rocket FTNIR-L1-025-2TE (Arcoptics, Switzerland)</li> <li>- Neospectra-micro + Raspberry-Pi (Si-Ware, Egypt)</li> </ul>
Data format	Raw Analyzed Presented as .tab files (table) and .py files (script)
Parameters for data collection	33 solid organic waste substrates of different biochemical composition were analyzed on four different optical systems: a benchtop laboratory spectrometer, a compact spectrometer with two measurement configurations (contact immersed probe and polarized light system), and a micro spectrometer.
Description of data collection	33 solid organic waste samples were freeze-dried and ground, and scanned on four different optical systems.  Depending on the optical system, absorbance or reflectance values are provided. For each sample, triplicate (x3) spectra were acquired to enhance spectral representativeness.  In parallel, a characterization method based on NIRS allowed to obtained reference values for each of these samples: carbohydrates content, lipids content, nitrogen content, chemical oxygen demand (COD), and biochemical methane potential (BMP)
Data source location	Institution: LBE, INRAE City/Town/Region: Narbonne Country: France
Data accessibility	Repository name: Data Inrae (Dataverse) Dataset name: On-site substrate characterization in the anaerobic digestion context: a dataset of near infrared spectra acquired with four different optical systems on freeze-dried and ground organic wastes Data identification number: 10.15454/SQQTUU Direct URL to data: <a href="https://doi.org/10.15454/SQQTUU">https://doi.org/10.15454/SQQTUU</a>
Related research article	A. Mallet, M. Pérémé, L. Awhangbo, C. Charnier, J.-M. Roger, J.-P. Steyer, É. Latrille, R. Bendoula, Fast at-line characterization of solid organic waste: Comparing analytical performance of different compact near infrared spectroscopic systems with different measurement configurations, <i>Waste Manag.</i> 126 (2021) 664–673. <a href="https://doi.org/10.1016/j.wasman.2021.03.045">https://doi.org/10.1016/j.wasman.2021.03.045</a> .

## Value of the Data

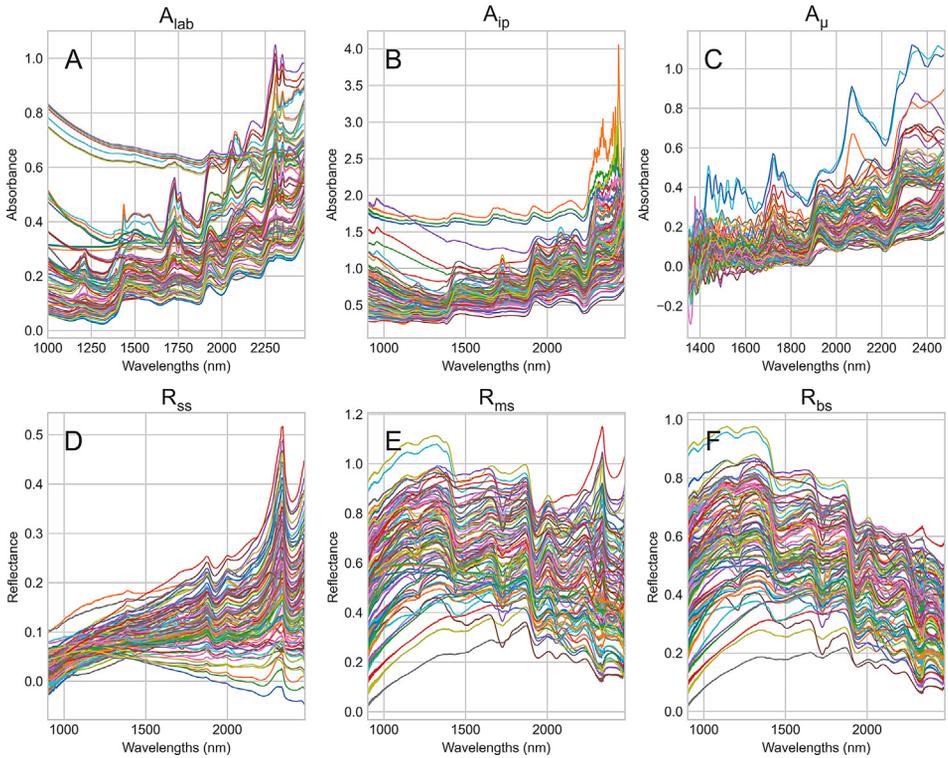
- Near infrared spectroscopy provides a fast and non-destructive methodology to characterize solid organic waste substrates involved in bioprocesses such as anaerobic digestion or composting.
- This unique dataset allows to compare the analytical performance of different compact spectroscopic systems for organic waste characterization including a handheld micro spectrometer, and two potential at-line systems with an immersed probe or polarized light system [4]. These spectroscopic systems can be compared to a laboratory spectrometer considered as the standard instrument of reference.
- The dataset could be used as well for instrument standardization [2] to test different strategies for building transfer models between instruments such as piecewise direct standardization (PDS) [5], or transfer by orthogonal projection (TOP) [6]. The instrument standardization from benchtop spectrometers to portable spectrometers is of high interest for researchers wishing to use calibration models on the field [7].
- Researchers in chemometrics or bioprocesses can benefit from this data, as it allows to build predictive models on organic waste. These models could find their application in anaerobic digestion monitoring.

## 1. Data Description

Data provided in this article consists of near infrared spectra of 33 organic waste samples of various biochemical composition, acquired on four different optical systems: an FT-NIR laboratory spectrometer (NIRFlex N-500, Buchi), a FT-NIR compact spectrometer (Rocket, Arcoptics) with two configurations (a contact immersed probe and a polarized light system PoLiS), and an FT-NIR micro spectrometer (Neospectra-micro, Si-Ware) based on MEMS FT-NIR technology. Each sample was scanned in triplicates on each optical system. Depending on the system, absorbance or reflectance spectra were obtained (Fig. 1). Each collected signals type was put into a separate file: the signals from the laboratory spectrometer ("*lab\_spectrometer\_spectra\_absorbance\_nm.tab*"), the signals from the compact spectrometer with the immersed probe ("*immersed\_probe\_spectra\_absorbance\_nm.tab*"), the three signals from the compact spectrometer with the PoLiS system ("*Polis\_Rbs\_spectra\_reflectance\_nm.tab*", "*Polis\_Rss\_spectra\_reflectance\_nm.tab*", "*Polis\_Rms\_spectra\_reflectance\_nm.tab*"), and the signals from the micro-spectrometer ("*microspectrometer\_spectra\_absorbance\_nm.tab*"). In each of these files, the first column corresponds to the name of the substrate, and the first row header corresponds to the wavelengths in nm.

In addition, a reference data file containing chemical information about each sample (carbohydrate, lipid and Nitrogen content, biochemical methane potential and chemical oxygen demand) is provided, as shown in Fig. 2. The reference data is consolidated into one file called "*reference\_data.tab*", with the first column that corresponds to substrate names, and the first row header that corresponds to the reference variable name (including its unit).

Finally, a python script is provided to show how a partial least squares regression (PLS-R) can be applied on this data. The script runs a train/test random split, followed by a k-fold cross-validation on train dataset to find optimal latent variable number of the PLS model and finally plots predictions and observed values on test set to evaluate obtained model performance. The python script name is "*data\_usage\_example.py*", and a documentation on how to use this script can be found in the text file "*README.md*".



**Fig. 1.** Raw spectra obtained on the 33 organic waste samples with A) the laboratory spectrometer (absorbance, in nm, referred as  $A_{lab}$ ), B) the compact spectrometer with an immersed probe (absorbance, in nm, referred as  $A_{ip}$ ), C) the micro spectrometer (absorbance, in nm, referred as  $A_{\mu}$ ), and the spectra obtained using the compact spectrometer plugged to the polarized light system PoLiS with D) the single scattered signals (reflectance, in nm, referred as  $R_{ss}$ ), E) the multi-scattered signal (reflectance, in nm, referred as  $R_{ms}$ ) and F) the total backscattered signal (reflectance, in nm, referred as  $R_{bs}$ ).

## 2. Experimental Design, Materials and Methods

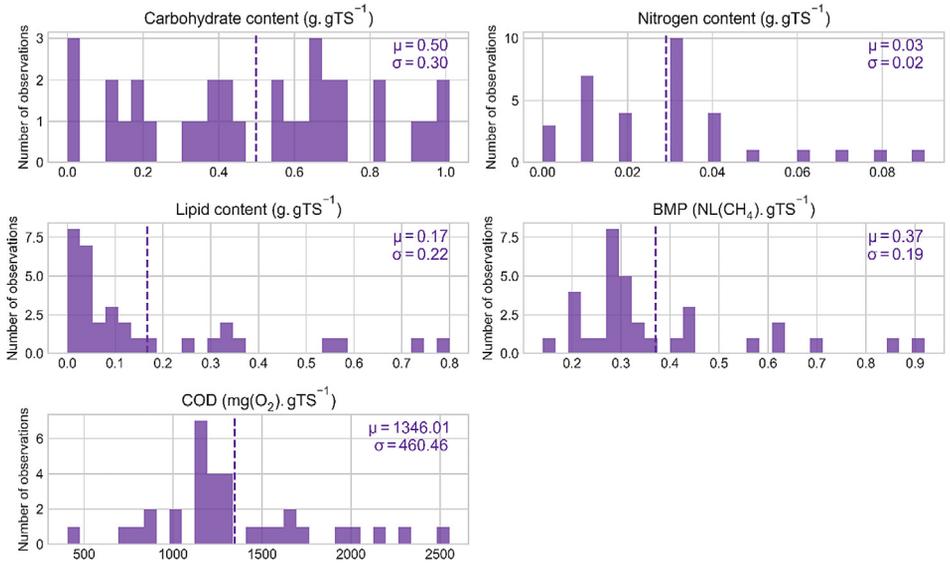
### 2.1. Samples and reference data

The 33 freeze-dried and ground samples were gathered from different anaerobic digestion plant in France (Fig. 3). The substrates cover a wide range of waste types including animal manure, animal waste, crop residues, food waste, and wastewater treatment plant waste (Table 1). Reference chemical data were obtained using a characterization method based on NIRS as described in Charnier et al. [8]. The standard errors of prediction evaluated on an independent test set were  $53 \text{ mgO}_2\text{.gTS}^{-1}$  for carbohydrates content,  $3.2 \times 10^{-2} \text{ g.gTS}^{-1}$  for lipid content,  $8.6 \times 10^{-3} \text{ g.gTS}^{-1}$  for nitrogen content and  $83 \text{ mgO}_2\text{.gTS}^{-1}$  for chemical oxygen demand.

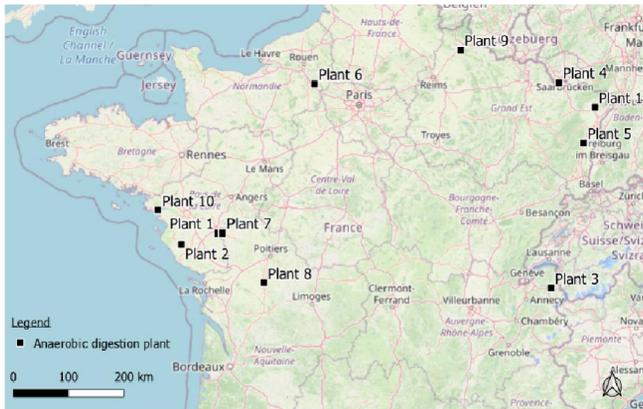
### 2.2. Near infrared spectra acquisition

#### 2.2.1. Laboratory spectrometer

For the laboratory spectrometer reference, freeze-dried and ground samples were scanned in reflectance over  $12,500 - 4000 \text{ cm}^{-1}$  (1000–2500 nm), with a resolution of  $8 \text{ cm}^{-1}$ , using a BUCHI NIR-Flex N-500 solids spectrophotometer fitted with a vial accessory (Buchi, Flawil,



**Fig. 2.** Histograms of reference values obtained on the 33 organic waste samples using NIRS prediction model. For each subplot, both the mean value ( $\mu$ ) and standard deviation value ( $\sigma$ ) are provided; and the dashed line represents the mean.



**Fig. 3.** Map (focused on France) showing the positions of the anaerobic digestion plants from where the substrates were collected.

Switzerland). Each sample was measured three times and shaken between each replicate. Furthermore, each measurement consisted of an average of 96 scans performed during vial rotation to obtain a representative measurement. White reference background signal  $I_0(\lambda)$  was collected on a Spectralon® (99% reflectance). Measurements  $I(\lambda)$  were obtained and absorbance  $A_{lab}(\lambda)$  was calculated.

$$A_{lab}(\lambda) = -\log_{10}\left(\frac{I(\lambda)}{I_0(\lambda)}\right). \quad (1)$$

### 2.2.2. Compact spectrometer with immersed probe

Both the immersed probe (Section 2.2.2) and PoLiS (Section 2.2.3) measurements were acquired using a compact Arcoptix FT-NIR Rocket spectrometer and an Ocean Optics' HL-

**Table 1**  
Substrate characteristics (name, waste type and origin).

Substrate name	Waste type	Origin
goat_manure	Animal manure	Plant 7
chicken_manure	Animal manure	Plant 7
cow_manure	Animal manure	Plant 10
pig_slurry	Animal manure	Plant 8
horse_manure	Animal manure	Plant 5
dairy_sludge	Animal waste	Plant 1
duck_slurry	Animal waste	Plant 1
charcuterie_grease_tank	Animal waste	Plant 1
slaughterhouse_sludge	Animal waste	Plant 1
gelatinous_water	Animal waste	Plant 1
corn_waste	Crop residues	Plant 5
grape_marc	Crop residues	Plant 5
wheat_derivative	Crop residues	Plant 8
corn_derivative	Crop residues	Plant 8
sunflower_derivative	Crop residues	Plant 8
clover_silage	Crop residues	Plant 9
straw	Crop residues	Plant 9
overpressed_beet_pulp	Crop residues	Plant 9
egg	Food waste	Plant 6
food_industrie_waste	Food waste	Plant 11
ready_meal_grease	Food waste	Plant 1
biscuit_dough	Food waste	Plant 1
ready_meal_waste	Food waste	Plant 1
lactoserum	Food waste	Plant 5
vegetables	Food waste	Plant 4
chocolate	Food waste	Plant 4
vegetables_hydrolysis	Food waste	Plant 3
lemon_pulp	Food waste	Plant 1
water_treatment_plant_grease_tank	Wastewater treatment	Plant 1
sewage_sludge	Wastewater treatment	Plant 2
paper_mill_waste	Wastewater treatment	Plant 9
water_treatment_plant_flotting_sludge	Wastewater treatment	Plant 3

2000 halogen lamp as light source. Samples were scanned in reflectance mode over 3800 to 11,000  $\text{cm}^{-1}$  (900–2500 nm) with a resolution of 4  $\text{cm}^{-1}$ . Each sample was measured on three different spots, and each measurement is an average of ten scans. The dark current (signal recovered without light)  $I_n(\lambda)$  is recorded at the beginning of the measurement session and automatically subtracted from the measured intensity. A white reference (SRS99, Spectralon®) was used as a reference  $I_0(\lambda)$  to standardize spectra from non-uniformities of all components of the instrumentation (light source, fibers, spectrometer) every hour.

The reflected light intensity  $I(\lambda)$  was recorded and the absorbance signal  $A_{ip}(\lambda)$  was computed:

$$A_{ip}(\lambda) = -\log_{10} \left( \frac{I(\lambda) - I_n(\lambda)}{I_0(\lambda) - I_n(\lambda)} \right). \tag{2}$$

**2.2.3. Compact spectrometer with polis system**

For the PoLiS measurements, the incident light cone was s-polarized using a wire-grid polarizer (Thorlabs WP25L-UB). The reflected light was split into an s-polarized image and a p-polarized image with a calcite Wollaston polarizer (Thorlabs WP10P) providing respectively parallel  $I_{\parallel}(\lambda)$  and perpendicular  $I_{\perp}(\lambda)$  light intensities.

The system resulted in three different signals: the single scattering reflectance  $R_{ss}(\lambda)$ , the multiple scattering reflectance  $R_{ms}(\lambda)$ , and the total backscattering reflectance  $R_{bs}(\lambda)$ :

$$R_{ss}(\lambda) = \frac{(I_{\parallel}(\lambda) - I_n(\lambda)) - (I_{\perp}(\lambda) - I_n(\lambda))}{I_0(\lambda) - I_n(\lambda)}. \tag{3}$$

$$R_{ms}(\lambda) = \frac{2(I_{\perp}(\lambda) - I_n(\lambda))}{I_0(\lambda) - I_n(\lambda)}. \quad (4)$$

$$R_{bs}(\lambda) = \frac{(I_{\parallel}(\lambda) - I_n(\lambda)) + (I_{\perp}(\lambda) - I_n(\lambda))}{I_0(\lambda) - I_n(\lambda)} \quad (5)$$

While the immersed probe is a contact probe (sample is placed directly in contact with probe), the PoLiS system is a distant measurement system placed at 5 cm from the sample.

#### 2.2.4. Micro spectrometer

Finally, the samples were analyzed using a micro spectrometer Neospectra-micro (Si-Ware, Egypt) mounted on a Raspberry Pi single-board computer. The instrument was controlled using a connection to a PC and the SpectraMOST software was used. Samples were scanned in reflectance mode over  $3921 \text{ cm}^{-1}$  to  $7407 \text{ cm}^{-1}$  (1350–2550 nm) and a resolution of  $66 \text{ cm}^{-1}$ . Each sample was measured on three different spots, and each measurement was obtained with a scan time of 28 s. A white reference (SRS99, Spectralon®) signal  $I_0(\lambda)$  was collected before each measurement. For each sample, the intensity signal  $I(\lambda)$  was collected and the absorbance signal  $A_{\mu}(\lambda)$  was computed:

$$A_{\mu}(\lambda) = -\log_{10}\left(\frac{I(\lambda)}{I_0(\lambda)}\right). \quad (6)$$

### Ethics Statement

Authors declare that here is no ethical issues regarding this dataset.

### CRedit Author Statement

**Margaud Pérémé:** Data curation, Writing – original draft, Methodology, Investigation; **Alexandre Mallet:** Conceptualization, Methodology, Software, Resource, Writing – review & editing, Supervision; **Lorraine Awhangbo:** Conceptualization, Methodology, Supervision; **Cyrille Charnier:** Writing – review & editing, Supervision; **Jean-michel Roger:** Writing – review & editing, Supervision; **Jean-philippe Steyer:** Writing – review & editing, Supervision; **Éric Latrille:** Conceptualization, Methodology, Resource, Data curation, Writing – review & editing, Supervision; **Ryad Bendoula:** Conceptualization, Methodology, Resource, Writing – review & editing, Supervision.

### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have or could be perceived to have influenced the work reported in this article.

### Acknowledgments

Financial support from the National Research Institute for Agriculture, Food and Environment (INRAE), and the French Agency of National Research and Technology (ANRT) [grant number 2018/0461] is hereby acknowledged. Authors would also like to acknowledge the technical support brought by the Bio2E platform [9].

## References

- [1] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: a basic tool of chemometrics, *Chemom. Intell. Lab. Syst.* 58 (2001) 109–130, doi:[10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1).
- [2] J.J. Workman, A review of calibration transfer practices and instrument differences in spectroscopy, *Appl. Spectrosc.* 72 (2018) 340–365, doi:[10.1177/0003702817736064](https://doi.org/10.1177/0003702817736064).
- [3] A. Mallet, M. Pérémé, L. Awhangbo, C. Charnier, J.-M. Roger, J.-P. Steyer, É. Latrille, R. Bendoula, Fast at-line characterization of solid organic waste: comparing analytical performance of different compact near infrared spectroscopic systems with different measurement configurations, *Waste Manag.* 126 (2021) 664–673, doi:[10.1016/j.wasman.2021.03.045](https://doi.org/10.1016/j.wasman.2021.03.045).
- [4] L. Awhangbo, R. Bendoula, J.M. Roger, F. Béline, Multi-block SO-PLS approach based on infrared spectroscopy for anaerobic digestion process monitoring, *Chemom. Intell. Lab. Syst.* 196 (2020) 103905, doi:[10.1016/j.chemolab.2019.103905](https://doi.org/10.1016/j.chemolab.2019.103905).
- [5] Y. Wang, D.J. Veltkamp, B.R. Kowalski, Multivariate instrument standardization, *Anal. Chem.* 63 (1991) 2750–2756, doi:[10.1021/ac00023a016](https://doi.org/10.1021/ac00023a016).
- [6] A. Andrew, T. Fearn, Transfer by orthogonal projection: making near-infrared calibrations robust to between-instrument variation, *Chemom. Intell. Lab. Syst.* 72 (2004) 51–56, doi:[10.1016/j.chemolab.2004.02.004](https://doi.org/10.1016/j.chemolab.2004.02.004).
- [7] L.C. Krapf, D. Nast, A. Gronauer, U. Schmidhalter, H. Heuwinkel, 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, doi:[10.1016/j.biortech.2012.11.027](https://doi.org/10.1016/j.biortech.2012.11.027).
- [8] C. Charnier, E. Latrille, J. Jimenez, M. Lemoine, J.C. Boulet, J. Miroux, J.P. Steyer, Fast characterization of solid organic waste content with near infrared spectroscopy in anaerobic digestion, *Waste Manag* 59 (2017) 140–148, doi:[10.1016/j.wasman.2016.10.029](https://doi.org/10.1016/j.wasman.2016.10.029).
- [9] Bio2E, INRAE, 2020. Environmental Biotechnology and Biorefinery Platform, (2020). doi:[10.15454/1.557234103446854E12](https://doi.org/10.15454/1.557234103446854E12).