

# Fast at-line characterization of solid organic waste: Comparing analytical performance of different compact near infrared spectroscopic systems with different measurement configurations

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

### ▶ To cite this version:

Alexandre Mallet, Margaud Pérémé, Lorraine Awhangbo, Cyrille Charnier, Jean-Michel Roger, et al.. Fast at-line characterization of solid organic waste: Comparing analytical performance of different compact near infrared spectroscopic systems with different measurement configurations. Waste Management, 2021, 126, pp.664-673. 10.1016/j.wasman.2021.03.045 . hal-03204372

### HAL Id: hal-03204372 https://hal.inrae.fr/hal-03204372

Submitted on 1 Jun 2022

**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

- Fast at-line characterization of solid
- <sup>2</sup> organic waste: comparing analytical
- <sup>3</sup> performance of different compact near
- <sup>4</sup> infrared spectroscopic systems with
- <sup>5</sup> different measurement configurations
- 7 Authors: Alexandre Mallet<sup>a,b,c,d,\*</sup>, Margaud Pérémé<sup>a,d</sup>, Lorraine
- 8 Awhangbo<sup>a,d</sup>, Cyrille Charnier<sup>c</sup>, Jean-Michel Roger<sup>b,d</sup>, Jean-
- 9 Philippe Steyer<sup>a</sup>, Éric Latrille<sup>a,d</sup>, Ryad Bendoula<sup>b</sup>
- <sup>a)</sup> INRAE, Univ Montpellier, LBE, 102 Av des Etangs, Narbonne F-11100, France
- 11 <sup>b)</sup> INRAE, UMR ITAP, Montpellier University, Montpellier, France
- 12 c) Bioentech, F-11100 Narbonne, France
- 13 <sup>d)</sup> ChemHouse Research Group, Montpellier, France
- 14

<sup>\*</sup> Corresponding author E-mail address: alexandre.mallet@bioentech.eu Full postal address: INRAE-LBE, 102, Avenue des Etangs, F-11100 Narbonne

## 15 Abstract

16 Fast characterization of solid organic waste using near infrared spectrosco-17 py has been successfully developed in the last decade. However, its adoption in biogas plants for monitoring the feeding substrates remains limited due to the lack 18 of applicability and high costs. Recent evolutions in the technology have given rise 19 20 to both more compact and more modular low-cost near infrared systems which 21 could allow a larger scale deployment. The current study investigates the rele-22 vance of these new systems by evaluating four different Fourier transform near-23 infrared spectroscopic systems with different compactness (laboratory, portable, 24 micro spectrometer) but also different measurement configurations (polarized light, 25 at distance, in contact). Though the conventional laboratory spectrometer showed 26 the best performance on the various biochemical parameters tested (carbohydrates, lipids, nitrogen, chemical oxygen demand, biochemical methane potential), 27 the compact systems provided very close results. Prediction of the biochemical 28 methane potential was possible using a low-cost micro spectrometer with an inde-29 pendent validation set error of only 91 NmL(CH<sub>4</sub>).gTS<sup>-1</sup> compared to 60 30 NmL(CH<sub>4</sub>).gTS<sup>-1</sup> for a laboratory spectrometer. The differences in performance 31 were shown to result mainly from poorer spectral sampling; and not from instru-32 33 ment characteristics such as spectral resolution. Regarding the measurement con-34 figurations, none of the evaluated systems allowed a significant gain in robustness. In particular, the polarized light system provided better results when using its multi-35

36 scattered signal which brings further evidence of the importance of physical light-

37 scattering properties in the success of models built on solid organic waste.

38

### 39 Keywords

40 Near infrared spectroscopy; anaerobic digestion;

41 process monitoring; biochemical methane potential; compact systems; measure-

42 ment modes;

# 43 1. Introduction

In anaerobic digestion processes, different organic waste are often co-digested to 44 enhance the production of both biogas and fertilizers (Hagos et al., 2017). A tre-45 mendous diversity of waste is concerned by these bioprocesses such as agricul-46 47 tural residues (animal manure, crop stems/stalks, silage), food industry waste (brewery, sugar refinery), urban solid waste, meat waste or catering waste. This 48 49 implies that these waste cover a large range of biochemical composition and physical properties. Moreover, such properties may fluctuate according to factors such 50 51 as crop seasonality, transport or storage. This brings important challenges for ensuring the stability of the process and the efficiency of biogas production in digest-52 ers (Wu et al., 2019). To answer this, online monitoring of the feeding substrate 53 guality could allow the direct adaptation of the feeding strategy to the feeding sub-54 55 strate quality (Jacobi et al., 2011). However, up to today, this has only been shown to be possible on digesters fed with a single type of substrate (like maize silage),
while the usefulness of such system appears greater with important variations of
substrate type and quality (Jacobi et al., 2012). In light of this, for co-digestion
plants, there is a need for the development of fast and reliable characterization
methods that are applicable on highly diverse organic waste.

61

Near infrared spectroscopy (NIRS), coupled with multivariate analysis techniques 62 63 (Næs and Martens, 1984), has been successfully used as a fast and robust characterization method of solid organic waste (Skvaril et al., 2017). In the composting 64 process, NIRS was used to monitor the degradation phases of compost (Albrecht 65 66 et al., 2008), or to predict biochemical characteristics such as the carbon/nitrogen ratio (Vergnoux et al., 2009). In the anaerobic digestion process, the technology 67 was initially used for in-situ monitoring of dry solids (DS), volatile solids (VS), 68 chemical oxygen demand (COD) and volatile fatty acids (VFA) in digesters (Jacobi 69 70 et al., 2009; Lomborg et al., 2009; Stockl and Lichti, 2018; Wolf et al., 2011). NIRS 71 was then proposed for the determination of biochemical methane potential (BMP) 72 on municipal solid waste (Lesteur et al., 2011), and has since been extended to other types of waste (Doublet et al., 2013; Fitamo et al., 2017; Godin et al., 2015; 73 74 Triolo et al., 2014; Yang et al., 2021). Today, NIRS appears the most suitable method for predicting BMP on various organic substrates (Rodrigues et al., 2019). 75 76 More recently, NIRS was used to estimate complementary characteristics such as 77 carbohydrates content, lipid content, nitrogen content, COD, and kinetic parame-78 ters (Charnier et al., 2017). In terms of process monitoring, these developments

79 allow time-consuming reference measurements (which last typically one to two 80 months for a characteristic like BMP) to be available in less than five days. Today, what limits the adoption of NIRS in full-scale biogas plants is its low applicability 81 and high costs (Wu et al., 2019). Indeed, freeze-drying and grinding steps are nec-82 essary to avoid water and particle size effects in NIRS (Mallet et al., 2021), which 83 currently limits the online applicability of such system. Moreover, the high costs of 84 85 the spectrometer and the logistics involved in sending the sample at the laboratory still limits a regular and exhaustive analysis of the feeding substrates. Whether 86 NIRS is applied directly on fresh waste or with a prior freeze-drying step, there is a 87 need to develop cheap and reliable instruments which can be used on a wide 88 89 range of substrate types in order to promote a greater adoption of NIRS in co-90 digestion plants. This could be addressed by an at-site use of low-cost and com-91 pact near infrared (NIR) systems.

92

In the past few years, the use of NIRS has developed out of laboratories, 93 94 thanks to important progress in the miniaturization of instruments. In particular, 95 handheld Fourier transform near infrared (FT-NIR) micro spectrometers have appeared in the market, and make use of a micro-electro-mechanical systems-based 96 97 (MEMS) Michelson interferometer (Beć et al., 2021). While conventional Michelson 98 interferometers are made of discrete elements (including the moving mirror actioned by a motor, the fixed mirror, and beam splitter), MEMS technology enables a 99 100 monolithic integration of these elements on a single chip, with the particularity that 101 the moving mirror is operated by an electrical signal. Amongst the spectrometers

102 making use of this technology, the NeoSpectra instrument has shown good analyt-103 ical performance results for soil organic and total carbon content characterization 104 (Sharififar et al., 2019; Tang et al., 2020), or authenticity screening in food (McVey et al., 2021). These compact spectrometers allow the measurements to be per-105 106 formed on site, thanks to their compactness, robustness and cost. However, these 107 compact portable spectrometers tend to have poorer instrument performances than 108 laboratory spectrometer, with lower resolution, spectral range, and signal-to-noise 109 ratio (Beć et al., 2020; Crocombe, 2018). Therefore, the suitability of such systems 110 for the characterization of diverse organic waste still needs to be assessed.

111

112 Another aspect of these compact systems concerns their modularity and the 113 possibility of testing different measurement configurations, in order to enhance the 114 measured signal. Indeed, in complex matter such as solid organic waste, the 115 Bouguer-Beer-Lambert law does not hold due to important light scattering (Dahm 116 and Dahm, 2004). To answer this, spectral pre-processing has been proposed to 117 remove both additive and multiplicative effects (Rabatel et al., 2020; Rinnan et al., 118 2009; Zeaiter et al., 2005) and thus, make the problem linear again. However, other developments have rather focused on enhancing the measured signal directly. A 119 120 promising optical pre-processing method, based on polarized light spectroscopy 121 (Backman et al., 1999) has been proposed to improve the absorbance signal 122 measurement on such scattering samples (Bendoula et al., 2015; Gobrecht et al., 123 2015; Xu et al., 2019). Such system has shown analytical performance improve-124 ments for soils (Gobrecht et al., 2016), and more recently, for digestate (Awhangbo

125 et al., 2020), but has never been evaluated on solid organic waste. Still in the aim 126 of enhancing the measured spectra, time-resolved spectroscopy shows a promis-127 ing future : applied to pharmaceutical tablets, collected photons with a particular propagation time were shown to be most informative for guantification (Alayed and 128 Deen, 2017; Johansson et al., 2002). However, the cost of this technology still re-129 mains prohibitive for the organic waste management sector. Finally, the measure-130 131 ment mode (at distance or in contact, in reflectance or in interactance) also plays 132 an important role in the final accuracy for estimating biochemical properties 133 (Hemrattrakun et al., 2021; Khodabakhshian et al., 2019; Schaare and Fraser, 2000). Though current used laboratory spectrometers make use of a distance re-134 135 flectance measurement, a contact immerged probe measurement has been shown 136 to be useful for prediction of parameters on digestates (Awhangbo et al., 2020). 137 Authors observe higher reflectance levels with less noise in the collected spectra, 138 as well as new chemical features which were not apparent in a remote probe con-139 figuration. In light of this, it appears that the use of different measurement configu-140 rations could enable the calibration of more accurate and more robust NIRS mod-141 els on diverse solid organic waste.

142

As mentioned, the applicability of compact and low-cost spectroscopic systems remains to be assessed for biochemical characterization of highly diverse solid organic waste. Moreover, the modularity offered by such compact systems is a unique opportunity to evaluate whether the use of different measurement configurations can help build more robust models. This study aims to assess these two

148	matters by comparing the analytical performances of four different NIRS systems:
149	a standard laboratory spectrometer, a portable spectrometer with two measure-
150	ment configurations (contact mode and polarized mode), and a micro-
151	spectrometer. For this purpose, measurements were acquired with each system on
152	a selection of solid organic waste. Then, for each spectroscopic system, prediction
153	models for five biochemical characteristics (carbohydrates, lipids, nitrogen, COD
154	and BMP) were calibrated and their performances were compared.

# 155 2. Materials and Methods

### 156 2.1. Sample preparation and reference analyses

Thirty-three substrates were selected amongst various waste types that have been 157 collected in rural, territorial and industrial anaerobic digestion plants in France. 158 159 These substrates cover a wide range of biochemical and physical properties: solid 160 cellulosic waste (like silage, cereals and corn cobs), liquid cellulosic suspensions 161 (such as manure), liquid fat suspensions (catering waste or biowaste), sweet emulsions (such as lactoserum), or protein and fat solid pastes (such as egg waste or 162 cacao butter). The visual aspect of some of these substrates in raw form is pre-163 164 sented in Appendix A. 165 For spectral measurements, each substrate sample was freeze-dried and ground 166 to 1 mm. The dataset is fully described in a data paper [On-site substrate charac-167 terization in the anaerobic digestion context: a dataset of NIR spectra acquired with 168 four different optical systems on freeze-dried and ground organic waste – this data

- 169 paper is submitted jointly with this article, reference will need an update after re-
- 170 viewing] and available online [https://doi.org/10.15454/SQQTUU].
- 171 Biochemical characterization of substrates was obtained by using a NIRS calibrat-
- ed model (Charnier et al., 2017), with errors on independent test sets of 53
- 173 mg(O<sub>2</sub>).gTS<sup>-1</sup> for carbohydrates content, 3.2\*10<sup>-2</sup> g.gTS<sup>-1</sup> for lipids content, 8.6
- mg.gTS<sup>-1</sup> for nitrogen content, and 83 mg(O<sub>2</sub>).gTS<sup>-1</sup> for COD. The histograms of
- 175 obtained prediction values are presented in Figure 1.

### 176 2.2. Spectroscopic systems

The four spectroscopic systems compared in this study are presented below. Inaddition, spectral measurement protocols are compared in Table 1.

#### 179 2.2.1. Laboratory spectroscopic system

The laboratory spectroscopic system consists of a NIR-Flex N-500 solids FT-NIR spectrophotometer with a vial accessory (Buchi, Flawil, Switzerland), scanning in reflectance mode with a spectral range of 4 000 cm<sup>-1</sup> to 10 000 cm<sup>-1</sup> (1000-2500 nm) and a resolution of 4 cm<sup>-1</sup>. An external white reference (Spectralon<sup>®</sup>) signal  $I_0(\lambda)$  is automatically taken every 10 minutes. For each sample, an intensity signal  $I(\lambda)$  was collected, and the pseudo-absorbance signal  $A_{lab}(\lambda)$  was computed:

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

#### 188 2.2.2. Portable spectrometer with immersed contact probe

189 The immersed contact probe system consists of a FT-NIR Rocket spectrometer

190 (Arcoptix, Neuchatel, Switzerland) scanning in reflectance mode with a spectral

191 range of 3 800 cm<sup>-1</sup> to 11 000 cm<sup>-1</sup> (900-2500 nm) and a resolution of 4 cm<sup>-1</sup>. The

192 spectrometer was connected to two optical fibers (for illumination and signal collec-

tion) of 1000 μm core diameter and numerical aperture of 0.39 (BFY1000,

194 Thorlabs). A tungsten-halogen source (Ocean Optics HL-200-FHSA) was used for

195 illumination. For each sample, the intensity  $I(\lambda)$  was collected. A white reference

197 in  $I_0(\lambda)$ . Finally, a dark current signal  $I_n(\lambda)$  corresponding to the instrumental noise

(SRS99, Spectralon®) was scanned every hour during the measurements resulting

198 was recorded and subtracted to all spectra. A pseudo-absorbance signal  $A_{ip}(\lambda)$ 199 was thus calculated:

200

196

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

201

#### 202 2.2.3. Portable spectrometer with polarized light spectroscopy

203

The polarized light system consists of the same elements (spectrometer, light source, optical fibers) as the immersed contact probe system, however, measure-

206 ments were made at a distance of 5 cm from the samples, and a polarized light 207 component (Awhangbo et al., 2020) was connected to the spectrometer. This 208 component consisted in a wire-grid polarizer (Thorlabs WP25L-UB) to s-polarize the incident light; and a calcite Wolaston polarizer (Thorlabs WP10P) to split the 209 210 reflected light in an s-polarized and p-polarized image, corresponding to parallel 211  $I_{\parallel}(\lambda)$  and perpendicular  $I_{\perp}(\lambda)$  light signals. As in the previous system, both the dark current signal  $I_n(\lambda)$ , and a white reference signal  $I_0(\lambda) = I_{\parallel}(\lambda)_0 + I_{\perp}(\lambda)_0$  were col-212 213 lected. Three signals were then calculated following Bendoula et al. (2015) formula (Bendoula et al., 2015): the single scattering reflectance  $R_{ss}(\lambda)$ , the multiple scat-214 tering reflectance  $R_{ms}(\lambda)$ , and the total backscattering reflectance  $R_{bs}(\lambda)$ : 215

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

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

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

216

217 2.2.4. Handheld micro spectrometer

218 The micro spectrometer system consists of a MEMS FT-NIR NeoSpectra spec-

trometer (Si-Ware, Cairo, Egypt) scanning in reflectance mode with a spectral

220 range of 3 921 cm<sup>-1</sup> to 7 407 cm<sup>-1</sup> (1350-2550 nm) and a resolution of 66 cm<sup>-1</sup>. A

white reference (SRS99, Spectralon®) signal  $I_0(\lambda)$  was collected before each

measurement. For each sample, an intensity signal  $I(\lambda)$  was collected and the pseudo-absorbance signal  $A_{\mu}(\lambda)$  was computed:

$$A_{\mu}(\lambda) = -log_{10}\left(R_{\mu}(\lambda)\right) = -log_{10}\left(\frac{I(\lambda)}{I_{0}(\lambda)}\right).$$
 (Eq. 6)

224

### 225 2.3. Data analysis: model calibration

All the data analysis was performed using Python 3.6.5: data wrangling with Pan-

das 0.25.1, NumPy 1.17.3, SciPy 1.3.1, Scikit-learn 0.21.3, and plotting with Mat-

plotlib 2.2.2 (Harris et al., 2020; Hunter, 2007; McKinney, 2010; Pedregosa et al.,

229 2015; van Rossum and Drake, 2009; Virtanen et al., 2020).

Measurements of the 33 substrates on the four spectroscopic configurations yielded six different matrices : the absorbance signal  $A_{lab}$  from the laboratory spectrometer, the absorbance signal  $A_{ip}$  from the immerged contact probe system, the three reflectance signals  $R_{ss}$ ,  $R_{ms}$  and  $R_{bs}$  (respectively single scattered, multiple scattered, total back-scattered) from the polarized system, and finally the absorbance signal  $A_{\mu}$  from the micro spectrometer system.

For noise reduction and baseline correction, a selection of seven pretreatments that have proven to be efficient in previous studies on organic waste (Charnier et al., 2017; Lesteur et al., 2011) have been used: the standard normal variate (Barnes et al., 1989) (SNV), the first-order detrend (Barnes et al., 1989) (DT1), the first-order Savitzky-Golay (Savitzky and Golay, 1964) derivation (SG1), the second-order Savitzky-Golay derivation (SG2), combinations of SNV and first-order
Savitzky-Golay derivation (SNV+SG1 or SG1+SNV) and finally, a weighted EMSC
with variable sorting for normalization (VSN) (Rabatel et al., 2020). The raw signal
was used directly as well, which resulted overall in testing eight different preprocessing conditions.

246 In order to evaluate the models built on each spectroscopic system, a validation 247 test set was constituted. With the aim of producing a representative validation test 248 set, the Duplex algorithm (Snee, 1977) was run for each reference characteristic 249 (carbohydrates content, lipid content, total nitrogen content, COD, BMP). This resulted in a training set of 22 substrates, and a validation test set of 11 substrates. 250 251 To assess the representativeness of the validation test set in terms of spectral var-252 iability, a principal components analysis (Cordella, 2012) was done, and obtained 253 scores were plot in Figure 2 and Appendix C.

254 Models were built using a partial least squares regression (PLS1-R) with NI-255 PALS algorithm (Næs and Martens, 1984; WOLD, 1973). To determine the number 256 of latent variables, a cross-validation was done using a repeated randomized 257 group-k-fold cross-validation with k = 5 the fold number and n repeats = 30 the repetition number. Sample triplicates were always kept within one fold to ensure 258 independence. For each cross-validation run, various metrics were then calculated: 259 260 the root-mean-square error (RMSE), the mean absolute error (MAE) (Willmott and 261 Matsuura, 2005), the coefficient of determination (R<sup>2</sup>), and B-coefficients metrics 262 which are the Durbin-Watson statistic (DW) and the variance (Rutledge and 263 Barros, 2002). The choice of the number of latent variables was made by analyzing

all these metrics together (i.e. choosing the minimal number of latent variables
 while minimizing RMSE and MAE, maximizing R<sup>2</sup>, and detecting rate increase of

266 DW and variance of B-coefficients).

267 Spectral range was also optimized for each of the signal types. This was done

by calibrating a first model, analyzing its B-coefficients (available in Appendix E),

and shrinking the spectral range adequately before recalibrating the model.

270 The final performances of the obtained models were evaluated on the validation

test set, based on the root-mean-square error (RMSE) and the coefficient of determination ( $R^2$ ).

To assess prediction repeatability of a given model, the variance of each sample's triplicate spectra predictions was calculated:

275

$$s_r^2 = \sum_{i=1}^3 (\hat{y}_i - \bar{\hat{y}})^2,$$
 (Eq. 7)

276

and the global prediction repeatability standard deviation was calculated as the quadratic mean of each  $s_r$ :

279

$$S_r = \sqrt{\sum_{i=1}^{n\_test\_samples} s_{r_i}^2}.$$
 (Eq. 8)

## 281 3. Results & Discussion

### 282 3.1. Data overview

#### 283 3.1.1. Training set and validation test set

284 For each reference characteristic, train and test set distributions (respectively in blue and orange) are presented as histograms in Figure 1. Very similar distribu-285 286 tions (same mean, same standard deviation) for all characteristics show that the Duplex algorithm succeeded in obtaining a representative test set in terms of bio-287 288 chemical composition. To complete this analysis, the representativeness in terms of signal is assessed by looking at the scores of train and test sets for each signal 289 290 type in Figure 2 (only the first and second component scores are displayed, but 291 scores up to the tenth component were checked). As shown, the test set signals (in 292 orange) cover most of the range covered by train set signals. However, in some 293 cases, the variability of train set signals is not fully well represented in the test set. 294 For example, in the upper left score plot representing the laboratory spectrometer 295 configuration signal A<sub>lab</sub>, no test set signal (orange square) is found in the far right 296 plot where there are two train set signals (blue triangles). This is consistent with the 297 fact that the Duplex algorithm was run on the reference values and not on the 298 spectral values, so there is no guarantee for test set spectral representativeness. 299 Although this is not optimal for evaluating calibration models alone, such method-

300 ology appeared to be the best to compare different spectroscopic systems on iden-301 tical samples without bias.

302

303 3.1.2. Raw spectra analysis

304

305 Figure 3 presents the raw reflectance signals obtained with each spectroscopic 306 system. In all signals, the main peaks found in organic waste were apparent: the 307 CH, CH<sub>2</sub> and CH<sub>3</sub> combination bands particularly present in fat (1731 nm, 1764 308 nm, 2310 nm, 2350 nm), the OH bands present in simple sugars (1436 nm, 1932 309 nm), the OH combination bands in starch or cellulose (2092 nm), and the NH com-310 bination bands present in proteins (2180 nm) (Williams and Antoniszyn, 2019; 311 Workman Jr. and Weyer, 2012). However, the relative amplitude of these peaks 312 seems to differ. For example, in the micro spectrometer signal ( $R_{\mu}$ ) the CH<sub>2</sub> combi-313 nation bands at 2310 nm and 2350 nm seem much less sharp than in the laborato-314 ry spectrometer signal (R<sub>lab</sub>). This can be well explained by the lower resolution of 315 the micro-spectrometer (66 cm<sup>-1</sup>) compared to the laboratory spectrometer (4 cm<sup>-1</sup>) 316 <sup>1</sup>). A consequence of this is that the models built on compact systems such as the 317 micro spectrometer will be based on more simple features, which could lead them 318 to be less accurate models but potentially also more robust. 319 In addition, the sensitivity with respect to the spectral range appears to differ from 320 one spectrometer to another. For the micro spectrometer, the measured signal be-

low 1600 nm seems much noisier than in other systems. Such sharp peaks are

not, *a priori*, expected in NIR spectra of complex matter. Similarly, for the immersed probe contact (R<sub>ip</sub>) or polarized signals (R<sub>ss</sub>, R<sub>ms</sub>, R<sub>bs</sub>), it seems the measured signals below 1200 nm and over 2240 nm are as well very noisy. For this reason, these spectral regions were later removed from the calibration of the built
models.

Another point of comparison concerns the observed reflectance levels (Figure 4). 327 The reflectance levels of  $R_{lab}$ ,  $R_{bs}$ , and  $R_{\mu}$  are much higher (with 75% of the values 328 that range respectively between 0.46 and 0.73, 0.43 and 0.66, and 0.60 and 0.90), 329 330 than the reflectance levels of  $R_{ip}$  (with 75% of the values that range between 0.15 331 and 0.35). This is mostly related to the way the signal is acquired (i.e. the meas-332 urement configuration). Indeed, reflectance levels are the result of both the absorp-333 tion level (dependent of chemical composition) and the scattering level (modifications of light optical path-length, and photon leakage (Gobrecht et al., 2014)). 334 Therefore, the chosen measurement configuration might favour one or the other, 335 336 leading to differences in the measured reflectance levels. Results show here that 337 the distant mode systems (i.e. R<sub>lab</sub>, R<sub>bs</sub>, and R<sub>µ</sub>) collect much more scattering photons than the contact mode system (i.e. R<sub>ip</sub>). Regarding the advantage of one 338 339 measurement mode over the other, this will mostly be dependent on the character-340 istic to be predicted, and its dependency on physical properties. 341 In the polarized light system, a clear difference of reflectance level can also be ob-342 served: 75% of the values of the multiple scattering signal R<sub>ms</sub> range between 0.52 and 0.76, against 0.05 and 0.11 for the single scattering signal R<sub>ss</sub>. This is con-343

sistent with the sole principle of polarized spectroscopy where R<sub>ss</sub> corresponds to
single scattering photons with low penetration in the media while R<sub>ms</sub> corresponds
to multiple scattering photons with a longer optical path length in the media due to
refraction events. This further confirms the efficiency of polarized spectroscopy as
an optical method to remove the scattering effects in the measured signal
(Bendoula et al., 2015; Gobrecht et al., 2016). However, the impacts on the subsequent models built on such signals remain to be studied.

### 351 3.2. Model performances

For the five reference characteristics that were studied, the best selected modelsobtained on each of the six signals are presented in Table 2.

354 For prediction models built using the laboratory spectrometer system, the errors

obtained on the test set (RMSEP) were of 0.108 g.gTS<sup>-1</sup> for carbohydrates content,

5.8 mg.gTS<sup>-1</sup> for nitrogen content, 0.034 g.gTS<sup>-1</sup> for lipids content, 0.060

357 NL(CH<sub>4</sub>).gTS<sup>-1</sup> for BMP, and 136.4 mg(O<sub>2</sub>).gTS<sup>-1</sup> for COD. These are all con-

358 sistent to the performances of reference models (Charnier et al., 2017). The slightly

lower performances obtained can be explained by the more limited number of

360 samples on which these models were built (22 samples) compared to the original

361 models (about 80 samples).

362 For all predicted characteristics, the laboratory system (A<sub>lab</sub>) showed better analyti-

363 cal performance results than the compact systems ( $A_{ip}$ ,  $A_{\mu}$ ,  $R_{ss}$ ,  $R_{ms}$ ,  $R_{bs}$ ). Howev-

364 er, in many cases, these latter systems showed similar performances to the labora-

365 tory system. For example, for BMP prediction, the prediction error (RMSEP) of the 366 model obtained with the laboratory spectrometer signal A<sub>lab</sub> was 60 mL(CH<sub>4</sub>).gTS<sup>-1</sup>. 367 In comparison, for the polarized system signals R<sub>ss</sub>, R<sub>ms</sub>, R<sub>bs</sub>, the model prediction errors were of respectively 115 mL(CH<sub>4</sub>).gTS<sup>-1</sup>, 111 mL(CH<sub>4</sub>).gTS<sup>-1</sup> and 100 368 mL(CH<sub>4</sub>).gTS<sup>-1</sup>, while for the micro spectrometer signal  $A_{\mu}$  the model's error was of 369 370 only 91 mL(CH<sub>4</sub>).gTS<sup>-1</sup>. Similarly, for carbohydrates content prediction, while the 371 prediction error for the laboratory spectrometer was 0.108 g.gTS<sup>-1</sup>, the errors for the micro-spectrometer and the immersed probe system were of only 0.134 g.gTS<sup>-</sup> 372 373 <sup>1</sup> and 0.104 g.gTS<sup>-1</sup>. Such observation can be made for all the other characteristics 374 studied. This is very promising because these models have acceptable errors 375 compared to the laboratory spectrometer. This means that the models built on or-376 ganic waste rely on sufficiently simple features so that the lower spectral resolution of compact spectrometers does not affect too much the performances. Such result 377 is consistent with similar studies on low-cost compact NIR spectrometers with lim-378 379 ited spectral range that are applied to herbaceous feedstock such as corn stover or 380 sorghum (Wolfrum et al., 2020). Knowing that these spectrometers are low-cost 381 (while the laboratory spectrometer costs about 50 000€, the immersed contact 382 probe and polarized systems each cost about 20 000€ and the micro spectrometer 383 only costs about 3 000 €), and can be used at-site, this holds great promises re-384 garding an increased adoption of NIRS for robust solid organic waste characteriza-385 tion in anaerobic digestion plants.

386

387 Models built on the three signals R<sub>ss</sub>, R<sub>ms</sub>, R<sub>bs</sub> obtained using the polarized spec-388 troscopy system show different performances. While Rbs is a signal very similar to 389 those obtained from the other set-ups because it includes information from all the backscattering light, R<sub>ss</sub> and R<sub>ms</sub> differ in terms of type of photons that are captured 390 391 by the spectrometer (respectively single scattering photons, and multiple scattering 392 photons). For all characteristics, models built on R<sub>bs</sub> signal show better perfor-393 mances than models built on the R<sub>ss</sub> and R<sub>ms</sub> signals. For example, for COD pre-394 diction, while the prediction error (RMSEP) of the model using the total back-395 scattering signal (R<sub>bs</sub>) is of 129.2 mg(O<sub>2</sub>).gTS<sup>-1</sup>, it is of 147.8 mg(O<sub>2</sub>).gTS<sup>-1</sup> using  $R_{ms}$  and 273.9 mg(O<sub>2</sub>).gTS<sup>-1</sup> using  $R_{ss}$ . This suggests that the use of polarized 396 397 spectroscopy for predicting these characteristics on organic waste may not be par-398 ticularly recommended.

399 In fact, for lipids content, COD and nitrogen content, the models built with Rss signals show much greater errors than models built with R<sub>ms</sub> signals (respectively 400 401 0.124 g.gTS<sup>-1</sup> greater than 0.068 g.gTS<sup>-1</sup>, 273.9 mg(O<sub>2</sub>).gTS<sup>-1</sup> greater than 147.8 402  $mg(O_2).gTS^{-1}$ , and 21.4 mg.gTS<sup>-1</sup> greater than 12.1 mg.gTS<sup>-1</sup>). This came as a 403 surprise, as the single scattering signal is theoretically supposed to be more related to absorbing constituents and less impacted by scattering effects (Gobrecht et 404 405 al., 2015). This can be explained by the fact that for a dataset with such diverse 406 solid organic substrates, the biochemical composition is closely related to the phys-407 ical properties. For example, high lipid content substrates (which also correspond 408 to substrates with high COD) tend to form liquid transparent solutions (like oil), 409 which transmit light much more than low lipid content substrates which usually form

410 highly scattering porous media. This relatively poor performance obtained with the 411 single scattering signal is consistent with results obtained on digestates where 412 physical structure appeared determinant (Awhangbo et al., 2020). One additional reason can be found in the measurement technique itself: the intensity captured for 413 414 the single scattering signal is much more limited than a classical total backscatter-415 ing signal, leading to higher signal-to-noise ratios. Further investigations could 416 concentrate on the use of a multi-block approach combining these three comple-417 mentary signals, as proposed on digestates (Awhangbo et al., 2020).

418

419 Apart from the single scattering signal, all signals allowed to build satisfactory 420 models for the biochemical characterization of organic waste. However, no spec-421 troscopic system allowed to surpass the analytical performance of the laboratory 422 spectrometer system. Figure 5 presents the observed and prediction plots for each 423 signal for the prediction of biochemical potential. In A<sub>lab</sub> (upper left subplot), the 424 predictions for each of the three replicate spectra do not differ (for one observed value, the prediction values are overlaid on the graph); while for the other spectro-425 scopic systems, the predictions for each of the three replicate spectra are very dif-426 427 ferent. For example, the sample with a BMP of 0.63 NL(CH<sub>4</sub>)g.TS<sup>-1</sup> has predictions 428 that vary for A<sub>lab</sub> between 0.869 NL(CH<sub>4</sub>)g.TS<sup>-1</sup> and 0.870 NL(CH<sub>4</sub>)g.TS<sup>-1</sup>, while the predictions for A<sub>ip</sub> and A<sub>µ</sub> vary respectively between 0.863 NL(CH<sub>4</sub>)g.TS<sup>-1</sup> and 429 430 0.940 NL(CH<sub>4</sub>)g.TS<sup>-1</sup> and between 0.709 NL(CH<sub>4</sub>)g.TS<sup>-1</sup> and 0.771 NL(CH<sub>4</sub>)g.TS<sup>-1</sup>. 431 This is observed for all the other characteristics as shown in Appendix D. It ap-432 pears that in all the compact systems, the replicate spectra vary much more from

433 each other than with the laboratory spectrometer. This could explain the greater 434 errors obtained using the compact systems. The following result may be investigated more quantitatively by calculating the global repeatability standard deviations 435 436 as presented in Figure 6. Indeed, for all characteristics, the compact systems show 437 much higher repeatability standard deviations than the laboratory spectrometer (in 438 red). Such differences are due to the way the spectral measurements are acquired. 439 As detailed in Table 1, the systems do not have the same number of scans and 440 sampling surface. While for one measurement, the laboratory spectrometer collects 441 scans during a full rotation of the sampling cup, the other systems only collect 442 scans on a fixed point of the sample's surface. This means that in 443 compact systems the spectral measurement is much less representative of the to-444 tal sample. It appears that the performance of compact systems could be en-445 hanced by optimizing the way the spectra are taken: increasing the number of 446 scans and the number of replicates to ensure a better stability of the measure-447 ments.

448

While the suitability of the compact and low-cost spectrometers has been demonstrated, some challenges remain. Though models could be calibrated on the compact systems' signals directly as in this study, it is most probable that models will remain being built and maintained on standard laboratory spectrometers, with transfer functions being built between the laboratory spectrometer (referred as the "master" spectrometer) and the compact spectrometers (referred as the "slave" spectrometers). This transfer approach has already been proven to be successful

between a laboratory spectrometer and an online spectrometer for in situ monitoring of anaerobic digestion (Krapf et al., 2013). However, the robustness of these
transfer functions applied to compact systems still needs to be assessed.

## 459 4. Conclusions

Results have shown that compact and low-cost systems including a hand-held mi-460 461 cro spectrometer are suitable for online characterization of diverse solid organic 462 waste. However, the use of new measurement configurations such as the polarized 463 mode was not shown to be an effective way to enhance the quality of predictive 464 models. This suggests that the physical scattering properties of the substrates are 465 the main determinant of analytical performance of NIRS calibration models built on such highly diverse solid organic waste. Keys for the improvement of the compact 466 467 systems appear to lie in further optimization of the sampling protocol. These results 468 set the path to a new era of low-cost and on-site NIRS analysis of the feeding sub-469 strates in co-digestion plants.

## 470 Acknowledgements

Financial support from the National Research Institute for Agriculture, Food and
Environment (INRAE), the French Agency of National Research and Technology
(ANRT) [grant number 2018/0461] and the Biogaz-RIO platform [FEDER-FSE
Languedoc Roussillon 2014-2020] is hereby acknowledged.

## 475 **References**

- 476 Alayed, M., Deen, M.J., 2017. Time-resolved diffuse optical spectroscopy and imaging
- 477 using solid-state detectors: Characteristics, present status, and research challenges.
- 478 Sensors (Switzerland) 17, 2115. https://doi.org/10.3390/s17092115
- 479 Albrecht, R., Joffre, R., Gros, R., Le Petit, J., Terrom, G., Périssol, C., 2008. Efficiency of
- 480 near-infrared reflectance spectroscopy to assess and predict the stage of
- 481 transformation of organic matter in the composting process. Bioresour. Technol. 99,
- 482 448–455. https://doi.org/10.1016/j.biortech.2006.12.019
- 483 Awhangbo, L., Bendoula, R., Roger, J.M., Béline, F., 2020. Multi-block SO-PLS approach
- 484 based on infrared spectroscopy for anaerobic digestion process monitoring. Chemom.
- 485 Intell. Lab. Syst. 196, 103905. https://doi.org/10.1016/j.chemolab.2019.103905
- 486 Backman, V., Gurjar, R., Badizadegan, K., Itzkan, I., Dasari, R.R., Perelman, L.T., Feld,
- 487 M.S., 1999. Polarized light scattering spectroscopy for quantitative measurement of
- 488 epithelial cellular structures in situ. IEEE J. Sel. Top. Quantum Electron. 5, 1019–
- 489 1026. https://doi.org/10.1109/2944.796325
- 490 Barnes, R.J., Dhanoa, M.S., Lister, S.J., 1989. Standard normal variate transformation and
- 491 de-trending of near-infrared diffuse reflectance spectra. Appl. Spectrosc. 43, 772–
- 492 777. https://doi.org/10.1366/0003702894202201
- 493 Beć, K.B., Grabska, J., Huck, C.W., 2021. Principles and Applications of Miniaturized
- 494 Near-Infrared (NIR) Spectrometers. Chem. A Eur. J. 27, 1514–1532.
- 495 https://doi.org/10.1002/chem.202002838
- 496 Beć, K.B., Grabska, J., Siesler, H.W., Huck, C.W., 2020. Handheld near-infrared
- 497 spectrometers: Where are we heading? NIR news 31, 28–35.
- 498 https://doi.org/10.1177/0960336020916815

- 499 Bendoula, R., Gobrecht, A., Moulin, B., Roger, J.M., Bellon-Maurel, V., 2015. Improvement
- 500 of the chemical content prediction of a model powder system by reducing multiple
- 501 scattering using polarized light spectroscopy. Appl. Spectrosc. 69, 95–102.
- 502 https://doi.org/10.1366/14-07539
- 503 Charnier, C., Latrille, E., Jimenez, J., Lemoine, M., Boulet, J.C., Miroux, J., Steyer, J.P.,
- 504 2017. Fast characterization of solid organic waste content with near infrared
- 505 spectroscopy in anaerobic digestion. Waste Manag. 59, 140–148.
- 506 https://doi.org/10.1016/j.wasman.2016.10.029
- 507 Cordella, C.B.Y., 2012. PCA: The Basic Building Block of Chemometrics, in: Krull, I.S.
- 508 (Ed.), Analytical Chemistry. IntechOpen, pp. 1–46. https://doi.org/10.5772/51429
- 509 Crocombe, R.A., 2018. Portable Spectroscopy. Appl. Spectrosc. 72, 1701–1751.
- 510 https://doi.org/10.1177/0003702818809719
- 511 Dahm, K.D., Dahm, D.J., 2004. Relation of representative layer theory to other theories of
- 512 diffuse reflection. J. Near Infrared Spectrosc. 12, 189–198.
- 513 https://doi.org/10.1255/jnirs.426
- 514 Doublet, J., Boulanger, A., Ponthieux, A., Laroche, C., Poitrenaud, M., Cacho Rivero, J.A.,
- 515 2013. Predicting the biochemical methane potential of wide range of organic
- 516 substrates by near infrared spectroscopy. Bioresour. Technol. 128, 252–258.
- 517 https://doi.org/10.1016/j.biortech.2012.10.044
- 518 Fitamo, T., Triolo, J.M., Boldrin, A., Scheutz, C., 2017. Rapid biochemical methane
- 519 potential prediction of urban organic waste with near-infrared reflectance
- 520 spectroscopy. Water Res. 119, 242–251. https://doi.org/10.1016/j.watres.2017.04.051
- 521 Gobrecht, A., Bendoula, R., Roger, J.M., Bellon-Maurel, V., 2016. A new optical method
- 522 coupling light polarization and Vis-NIR spectroscopy to improve the measurement of 523 soil carbon content. Soil Tillage Res. 155, 461–470.

524 https://doi.org/10.1016/j.still.2015.06.003

- 525 Gobrecht, A., Bendoula, R., Roger, J.M., Bellon-Maurel, V., 2015. Combining linear
- 526 polarization spectroscopy and the Representative Layer Theory to measure the Beer-
- 527 Lambert law absorbance of highly scattering materials. Anal. Chim. Acta 853, 486-
- 528 494. https://doi.org/10.1016/j.aca.2014.10.014
- 529 Gobrecht, A., Roger, J.M., Bellon-Maurel, V., 2014. Major Issues of Diffuse Reflectance
- 530 NIR Spectroscopy in the Specific Context of Soil Carbon Content Estimation. A
- 531 Review., Advances in Agronomy. https://doi.org/10.1016/B978-0-12-420225-2.00004-2
- 532
- 533 Godin, B., Mayer, F., Agneessens, R., Gerin, P., Dardenne, P., Delfosse, P., Delcarte, J.,
- 534 2015. Biochemical methane potential prediction of plant biomasses: Comparing
- 535 chemical composition versus near infrared methods and linear versus non-linear
- 536 models. Bioresour. Technol. 175, 382–390.
- 537 https://doi.org/10.1016/j.biortech.2014.10.115
- 538 Hagos, K., Zong, J., Li, D., Liu, C., Lu, X., 2017. Anaerobic co-digestion process for biogas
- 539 production: Progress, challenges and perspectives. Renew. Sustain. Energy Rev. 76,
- 540 1485-1496. https://doi.org/10.1016/j.rser.2016.11.184
- 541 Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D.,
- 542 Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., van
- 543 Kerkwijk, M.H., Brett, M., Haldane, A., del Río, J.F., Wiebe, M., Peterson, P., Gérard-
- 544 Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C.,
- 545 Oliphant, T.E., 2020. Array programming with NumPy. Nature 585, 357–362.
- 546 https://doi.org/10.1038/s41586-020-2649-2
- 547 Hemrattrakun, P., Nakano, K., Boonyakiat, D., Ohashi, S., Maniwara, P., Theanjumpol, P.,
- Seehanam, P., 2021. Comparison of Reflectance and Interactance Modes of Visible 548

- and Near-Infrared Spectroscopy for Predicting Persimmon Fruit Quality. Food Anal.
- 550 Methods 14, 117–126. https://doi.org/10.1007/s12161-020-01853-w
- 551 Hunter, J.D., 2007. Matplotlib: A 2D graphics environment. Comput. Sci. Eng. 9, 90–95.
- 552 https://doi.org/10.1109/MCSE.2007.55
- Jacobi, H.F., Moschner, C.R., Hartung, E., 2011. Use of near infrared spectroscopy in
- online-monitoring of feeding substrate quality in anaerobic digestion. Bioresour.

555 Technol. 102, 4688–4696. https://doi.org/10.1016/j.biortech.2011.01.035

- Jacobi, H.F., Moschner, C.R., Hartung, E., 2009. Use of near infrared spectroscopy in
- 557 monitoring of volatile fatty acids in anaerobic digestion. Water Sci. Technol. 60, 339–
- 558 346. https://doi.org/10.2166/wst.2009.345
- Jacobi, H.F., Ohl, S., Thiessen, E., Hartung, E., 2012. NIRS-aided monitoring and
- 560 prediction of biogas yields from maize silage at a full-scale biogas plant applying
- 561 lumped kinetics. Bioresour. Technol. 103, 162–172.
- 562 https://doi.org/10.1016/j.biortech.2011.10.012
- Johansson, J., Folestad, S., Josefson, M., Sparen, A., Abrahamsson, C., Andersson-
- 564 Engels, S., Svanberg, S., 2002. Time-resolved NIR/vis spectroscopy for analysis of
- solids: Pharmaceutical tablets. Appl. Spectrosc. 56, 725–731.
- 566 https://doi.org/10.1366/000370202760077676
- 567 Khodabakhshian, R., Emadi, B., Khojastehpour, M., Golzarian, M.R., 2019. A comparative
- 568 study of reflectance and transmittance modes of Vis/NIR spectroscopy used in
- 569 determining internal quality attributes in pomegranate fruits. J. Food Meas. Charact.
- 570 13, 3130–3139. https://doi.org/10.1007/s11694-019-00235-z
- 571 Krapf, L.C., Nast, D., Gronauer, A., Schmidhalter, U., Heuwinkel, H., 2013. Transfer of a
- 572 near infrared spectroscopy laboratory application to an online process analyser for in
- 573 situ monitoring of anaerobic digestion. Bioresour. Technol. 129, 39–50.

574 https://doi.org/10.1016/j.biortech.2012.11.027

- 575 Lesteur, M., Latrille, E., Maurel, V.B., Roger, J.M., Gonzalez, C., Junqua, G., Steyer, J.P.,
- 576 2011. First step towards a fast analytical method for the determination of Biochemical
- 577 Methane Potential of solid wastes by near infrared spectroscopy. Bioresour. Technol.
- 578 102, 2280–2288. https://doi.org/10.1016/j.biortech.2010.10.044
- 579 Lomborg, C.J., Holm-Nielsen, J.B., Oleskowicz-Popiel, P., Esbensen, K.H., 2009. Near
- 580 infrared and acoustic chemometrics monitoring of volatile fatty acids and dry matter
- 581 during co-digestion of manure and maize silage. Bioresour. Technol. 100, 1711–
- 582 1719. https://doi.org/10.1016/j.biortech.2008.09.043
- 583 Mallet, A., Charnier, C., Latrille, É., Bendoula, R., Steyer, J.-P., Roger, J.-M., 2021.
- 584 Unveiling non-linear water effects in near infrared spectroscopy: A study on organic
- 585 wastes during drying using chemometrics. Waste Manag. 122, 36–48.
- 586 https://doi.org/10.1016/j.wasman.2020.12.019
- 587 McKinney, W., 2010. Data Structures for Statistical Computing in Python, in: Proceedings
- of the 9th Python in Science Conference. pp. 56–61. https://doi.org/10.25080/majora92bf1922-00a
- 590 McVey, C., McGrath, T.F., Haughey, S.A., Elliott, C.T., 2021. A rapid food chain approach
- 591 for authenticity screening: The development, validation and transferability of a
- 592 chemometric model using two handheld near infrared spectroscopy (NIRS) devices.
- 593 Talanta 222, 121533. https://doi.org/10.1016/j.talanta.2020.121533
- Næs, T., Martens, H., 1984. Multivariate calibration. II. Chemometric methods. Trends
- 595 Anal. Chem. 3, 266–271. https://doi.org/10.1016/0165-9936(84)80044-8
- 596 Pedregosa, F., Varoquaux, G., Buitinck, L., Louppe, G., Grisel, O., Mueller, A., 2015.
- 597 Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 19, 29–33.
- 598 Rabatel, G., Marini, F., Walczak, B., Roger, J.M., 2020. VSN: Variable sorting for

- 599 normalization. J. Chemom. 34, 1–16. https://doi.org/10.1002/cem.3164
- 600 Rinnan, Å., Berg, F. van den, Engelsen, S.B., 2009. Review of the most common pre-
- 601 processing techniques for near-infrared spectra. TrAC Trends Anal. Chem. 28,
- 602 1201–1222. https://doi.org/10.1016/j.trac.2009.07.007
- Rodrigues, R.P., Rodrigues, D.P., Klepacz-Smolka, A., Martins, R.C., Quina, M.J., 2019.
- 604 Comparative analysis of methods and models for predicting biochemical methane
- potential of various organic substrates. Sci. Total Environ. 649, 1599–1608.
- 606 https://doi.org/10.1016/j.scitotenv.2018.08.270
- 607 Rutledge, D.N., Barros, A.S., 2002. Durbin-Watson statistic as a morphological estimator
- of information content. Anal. Chim. Acta 454, 277–295.
- 609 https://doi.org/10.1016/S0003-2670(01)01555-0
- 610 Savitzky, A., Golay, M.J.E., 1964. Smoothing and Differentiation of Data by Simplified
- 611 Least Squares Procedures. Anal. Chem. 36, 1627–1639.
- 612 https://doi.org/10.1021/ac60214a047
- 613 Schaare, P.N., Fraser, D.G., 2000. Comparison of reflectance, interactance and
- 614 transmission modes of visible-near infrared spectroscopy for measuring internal
- 615 properties of kiwifruit (Actinidia chinensis). Postharvest Biol. Technol. 20, 175–184.
- 616 https://doi.org/10.1016/S0925-5214(00)00130-7
- 617 Sharififar, A., Singh, K., Jones, E., Ginting, F.I., Minasny, B., 2019. Evaluating a low-cost
- 618 portable NIR spectrometer for the prediction of soil organic and total carbon using
- 619 different calibration models. Soil Use Manag. 35, 607–616.
- 620 https://doi.org/10.1111/sum.12537
- 621 Skvaril, J., Kyprianidis, K.G., Dahlquist, E., 2017. Applications of near-infrared
- 622 spectroscopy (NIRS) in biomass energy conversion processes: A review. Appl.
- 623 Spectrosc. Rev. 52, 675–728. https://doi.org/10.1080/05704928.2017.1289471

- 624 Snee, R.D., 1977. Validation of Regression Models: Methods and Examples.
- 625 Technometrics 19, 415–428. https://doi.org/10.1080/00401706.1977.10489581
- 626 Stockl, A., Lichti, F., 2018. Near-infrared spectroscopy (NIRS) for a real time monitoring of
- 627 the biogas process. Bioresour. Technol. 247, 1249–1252.
- 628 https://doi.org/10.1016/j.biortech.2017.09.173
- Tang, Y., Jones, E., Minasny, B., 2020. Evaluating low-cost portable near infrared sensors
  for rapid analysis of soils from South Eastern Australia. Geoderma Reg. 20, e00240.
- 631 https://doi.org/10.1016/j.geodrs.2019.e00240
- Triolo, J.M., Ward, A.J., Pedersen, L., Løkke, M.M., Qu, H., Sommer, S.G., 2014. Near
- 633 Infrared Reflectance Spectroscopy (NIRS) for rapid determination of biochemical
- methane potential of plant biomass. Appl. Energy 116, 52–57.
- 635 https://doi.org/10.1016/j.apenergy.2013.11.006
- van Rossum, G., Drake, F.L., 2009. Python 3 Reference Manual, CreateSpace. Scotts
  Valley, CA. https://doi.org/10.5555/1593511
- 638 Vergnoux, A., Guiliano, M., Le Dréau, Y., Kister, J., Dupuy, N., Doumenq, P., 2009.
- 639 Monitoring of the evolution of an industrial compost and prediction of some compost
- 640 properties by NIR spectroscopy. Sci. Total Environ. 407, 2390–2403.
- 641 https://doi.org/10.1016/j.scitotenv.2008.12.033
- Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D.,
- Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M.,
- 644 Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E.,
- 645 Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J.,
- 646 Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro,
- 647 A.H., Pedregosa, F., van Mulbregt, P., Vijaykumar, A., Bardelli, A. Pietro, Rothberg,
- 648 A., Hilboll, A., Kloeckner, A., Scopatz, A., Lee, A., Rokem, A., Woods, C.N., Fulton,

- 649 C., Masson, C., Häggström, C., Fitzgerald, C., Nicholson, D.A., Hagen, D.R.,
- 650 Pasechnik, D. V., Olivetti, E., Martin, E., Wieser, E., Silva, F., Lenders, F., Wilhelm,
- F., Young, G., Price, G.A., Ingold, G.L., Allen, G.E., Lee, G.R., Audren, H., Probst, I.,
- Dietrich, J.P., Silterra, J., Webber, J.T., Slavič, J., Nothman, J., Buchner, J., Kulick, J.,
- 653 Schönberger, J.L., de Miranda Cardoso, J.V., Reimer, J., Harrington, J., Rodríguez,
- 654 J.L.C., Nunez-Iglesias, J., Kuczynski, J., Tritz, K., Thoma, M., Newville, M.,
- 655 Kümmerer, M., Bolingbroke, M., Tartre, M., Pak, M., Smith, N.J., Nowaczyk, N.,
- 656 Shebanov, N., Pavlyk, O., Brodtkorb, P.A., Lee, P., McGibbon, R.T., Feldbauer, R.,
- 657 Lewis, S., Tygier, S., Sievert, S., Vigna, S., Peterson, S., More, S., Pudlik, T.,
- 658 Oshima, T., Pingel, T.J., Robitaille, T.P., Spura, T., Jones, T.R., Cera, T., Leslie, T.,
- Zito, T., Krauss, T., Upadhyay, U., Halchenko, Y.O., Vázquez-Baeza, Y., 2020. SciPy
- 660 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods 17, 261–
- 661 272. https://doi.org/10.1038/s41592-019-0686-2
- Williams, P., Antoniszyn, J., 2019. Near-infrared Technology: Getting the best out of light.
  African Sun Media. https://doi.org/10.18820/9781928480310
- 664 Willmott, C.J., Matsuura, K., 2005. Advantages of the mean absolute error (MAE) over the
- root mean square error (RMSE) in assessing average model performance. Clim. Res.
- 666 30, 79–82. https://doi.org/10.3354/cr030079
- 667 WOLD, H., 1973. Nonlinear Iterative Partial Least Squares (NIPALS) Modelling: Some
- 668 Current Developments, Multivariate Analysis–III. Academic press, Inc.
- 669 https://doi.org/10.1016/b978-0-12-426653-7.50032-6
- Wolf, D., von Canstein, H., Schröder, C., 2011. Optimisation of biogas production by
- 671 infrared spectroscopy-based process control. J. Nat. Gas Sci. Eng. 3, 625–632.
- 672 https://doi.org/10.1016/j.jngse.2011.07.006
- 673 Wolfrum, E.J., Payne, C., Schwartz, A., Jacobs, J., Kressin, R.W., 2020. A Performance

- 674 Comparison of Low-Cost Near-Infrared (NIR) Spectrometers to a Conventional
- 675 Laboratory Spectrometer for Rapid Biomass Compositional Analysis. Bioenergy Res.
- 676 https://doi.org/10.1007/s12155-020-10135-6
- 677 Workman Jr., J., Weyer, L., 2012. Practical Guide and Spectral Atlas for Interpretive Near-
- 678 Infrared Spectroscopy. CRC press. https://doi.org/10.1201/b11894
- Wu, D., Li, L., Zhao, X., Peng, Y., Yang, P., Peng, X., 2019. Anaerobic digestion: A review
  on process monitoring. Renew. Sustain. Energy Rev. 103, 1–12.
- 681 https://doi.org/10.1016/j.rser.2018.12.039
- Ku, J.L., Gobrecht, A., Gorretta, N., Héran, D., Gowen, A.A., Bendoula, R., 2019.
- 683 Development of a polarized hyperspectral imaging system for investigation of
- absorption and scattering properties. J. Near Infrared Spectrosc. 27, 314–329.
- 685 https://doi.org/10.1177/0967033519857732
- 686 Yang, G., Li, Y., Zhen, F., Xu, Y., Liu, J., Li, N., Sun, Y., Luo, L., Wang, M., Zhang, L.,
- 687 2021. Biochemical methane potential prediction for mixed feedstocks of straw and
- 688 manure in anaerobic co-digestion. Bioresour. Technol. 326, 124745.
- 689 https://doi.org/10.1016/j.biortech.2021.124745
- 690 Zeaiter, M., Roger, J.M., Bellon-Maurel, V., 2005. Robustness of models developed by
- 691 multivariate calibration. Part II: The influence of pre-processing methods. TrAC -
- 692 Trends Anal. Chem. 24, 437–445. https://doi.org/10.1016/j.trac.2004.11.023

# 694 Tables

### 695 Table 1. Spectral measurement protocol characteristics.

	iiiiiiei seu	FUIdi IzaliUli	Micro spec-	
spectrometer	contact	system	trometer	
	probe			
3	3	3	3	
96 scans	10 scans	10 scans	28s scan time	
360° rotation	Fixed point	Fixed point	Fixed point	
(~5 cm²)	(~0.05 cm <sup>2</sup> )	(~1 cm <sup>2</sup> )	(~25 cm <sup>2</sup> )	
Mix the whole	Change the	Change the	Change the	
sample	measured	measured	measured	
	surface posi-	surface posi-	surface posi-	
	tion	tion	tion	
	spectrometer 3 96 scans 360° rotation (~5 cm <sup>2</sup> ) Mix the whole sample	spectrometercontact probe33310 scans360° rotationFixed point (~0.05 cm²)4000 rotationFixed point (~0.05 cm²)400 rotationFixed point 	spectrometercontact probesystem333310 scans10 scans96 scans10 scans10 scans860° rotationFixed point (~0.05 cm²)Fixed point (~1 cm²)Wix the wholeChange the measured surface posi- tionChange the measured surface posi- tion	

Table 2. Descriptive statistics of the calibrated models. For each predicted reference measurement (carbohydrates, nitrogen, lipids, BMP, COD), the retained
model for each spectroscopic system signal (A<sub>lab</sub>, A<sub>µ</sub>, R<sub>ss</sub>, R<sub>ms</sub>, R<sub>bs</sub>) is presented.
The optimized parameters are provided (the spectral range, the pretreatment and
the number of latent variables) along with the different performance metrics
(RMSEC, RMSECV, RMSEP, R<sup>2</sup>C, R<sup>2</sup>P).

Experiment		Optimized parameters			Performance metrics				
Reference	Sig	Spectral	Pretreat-	#LV	RMSEc	RMSEcv	RMSE <sub>P</sub>	<b>R</b> ²c	R² <sub>₽</sub>
	nai	I range (nm)	ment						
Carbohydrates	Alab	1400-2240	SNV+SG1	2	0.143	0.192	0.108	0.78	0.83
g.gTS <sup>-1</sup>	<b>A</b> ip	1450-2240	SNV	2	0.188	0.169	0.104	0.82	0.84
	Au	1400-2500	SNV+SG1	3	0.146	0.137	0.134	0.75	0.88
	$R_{ss}$	1200-2240	SNV	2	0.259	0.296	0.207	0.28	0.37
	R <sub>ms</sub>	1200-2240	Raw	4	0.121	0.185	0.112	0.82	0.82
	Rbs	1300-2300	SNV+SG1	3	0.119	0.158	0.130	0.74	0.91
Nitrogen	Alab	1200-2240	SNV+SG1	10	3.6	15.0	5.8	0.98	0.89
mg.gTS⁻¹	A <sub>ip</sub>	1200-2240	SNV+SG1	5	9.9	18.4	7.5	0.87	0.83
	Au	1400-2500	SNV	5	10.1	20.7	13.1	0.85	0.51
	R <sub>ss</sub>	1200-2240	SG2	3	17.2	23.1	21.4	0.62	-0.4
	R <sub>ms</sub>	1200-2240	SNV+SG1	5	12.6	22.6	12.1	0.80	0.54
	R <sub>bs</sub>	1300-2300	SG1+SNV	4	10.7	17.6	11.7	0.83	0.61
Lipids	Alab	1400-2240	SNV	6	0.025	0.056	0.034	0.99	0.98
g.gTS <sup>-1</sup>	A <sub>ip</sub>	1200-2240	VSN	9	0.057	0.066	0.066	0.99	0.93
	Au	1400-2500	SG1+SNV	6	0.039	0.081	0.067	0.98	0.92
	R <sub>ss</sub>	1200-2240	Raw	5	0.084	0.110	0.124	0.91	0.74
	R <sub>ms</sub>	1200-2240	SG2	5	0.029	0.059	0.068	0.99	0.92

	<b>R</b> <sub>bs</sub> 1200-2480	SG2	6	0.029	0.066	0.066	0.99	0.93
BMP	<b>A</b> <sub>lab</sub> 1400-2240	SNV+SG1	9	0.020	0.078	0.060	0.99	0.90
NL(CH₄).gTS⁻¹	<b>A</b> <sub>ip</sub> 1200-2240	SG1	2	0.088	0.115	0.110	0.77	0.73
	<b>A</b> <sub>u</sub> 1600-2500	SG2	2	0.102	0.109	0.091	0.74	0.82
	<b>R</b> <sub>ss</sub> 1200-2240	SG1	4	0.085	0.115	0.115	0.79	0.71
	<b>R</b> <sub>ms</sub> 1350-2240	SG1	3	0.087	0.115	0.111	0.78	0.73
	<b>R</b> <sub>bs</sub> 1300-2300	SG1	6	0.046	0.120	0.100	0.94	0.78
COD	<b>A</b> <sub>lab</sub> 1400-2240	SNV	6	63.0	186.8	136.4	0.97	0.94
mg(O <sub>2</sub> ).gTS <sup>-1</sup>	<b>A</b> <sub>ip</sub> 1450-2240	SNV+SG1	7	106.2	228.3	160.2	0.93	0.90
	<b>A</b> <sub>u</sub> 1400-2500	SG1+SNV	3	180.4	303.1	196.9	0.81	0.85
	<b>R</b> <sub>ss</sub> 1200-2240	SG2	4	192.0	247.4	273.9	0.78	0.72
	<b>R</b> <sub>ms</sub> 1350-2240	SG1	3	112.5	151.4	147.8	0.92	0.92
	<b>R</b> <sub>bs</sub> 1200-2480	SG1+SNV	4	128.1	237.6	129.2	0.90	0.94





Figure 1. Histograms of reference characteristics. Both train and test sets (respectively in blue and in orange) obtained by a Duplex split are presented. Respective mean (labeled as  $\mu$ ) and standard deviation (labeled as  $\sigma$ ) are presented for train and test sets (respectively in blue, and in orange). Dotted lines represent respective mean values.



Figure 2. Principal component analysis (PCA) score plots of train and test sets (respectively in blue triangles, and in orange squares) for each signal type (from the four spectroscopic configurations). Each subplot represents the scores of the first and second principal components (PC1 and PC2). The percentage of explained variance is provided in the labels. The train and test split was obtained by a Duplex split based on the carbohydrates content levels. Score plots for other reference characteristics are shown in Appendix C.



Figure 3. Raw reflectance spectra of each spectroscopic system (R<sub>lab</sub>: laboratory spectrometer, R<sub>ip</sub>: immersed probe system, R<sub>µ</sub>: micro-spectrometer, R<sub>ss</sub>: single scattered signal
of polarized system, R<sub>ms</sub>: multiple scattered signal of polarized system, R<sub>bs</sub>: total back-

- scattered signal of polarized system). Each spectrum corresponds to the mean of the trip-
- 725 licate measurements.
- 726



727

728 Figure 4. Boxplots of raw reflectance spectra values for each signal type (R<sub>lab</sub>: laboratory 729 spectrometer,  $R_{ip}$ : immersed probe system,  $R_{\mu}$ : micro-spectrometer,  $R_{ss}$ : single scattered 730 signal of polarized system, R<sub>ms</sub>: multiple scattered signal of polarized system, R<sub>bs</sub>: total 731 back-scattered signal of polarized system). Each boxplot was obtained on the flattened 732 matrix (reflectance values for all samples and for all wavelengths). Median values are pre-733 sented with orange lines. The box limits represent the first and third quartile values (re-734 spectively Q1 and Q3), and the lines that extend from the box show the lowest and largest 735 data points excluding any outliers (respectively  $Q1 - 1.5 \times (Q3 - Q1)$ ) and  $Q1 - 1.5 \times (Q3 - Q1)$ 736 *Q*1)). Outliers are presented in empty black circles.



Figure 5. Scatter plots of predicted and observed values for the selected prediction models
of biochemical methane potential (BMP). Each subplot corresponds to the best model selected for a signal type (from the four spectroscopic configurations). Values for train and
test sets are respectively presented in blue triangles and orange squares. The spectral
range, pre-processing type and latent variable number (Iv) of the models are provided in
the titles. Results for other reference characteristics are provided in Appendix D.



Figure 6. Global prediction repeatability standard deviations as calculated with (Eq. 7) and (Eq. 8). For each reference characteristic (carbohydrates, nitrogen, lipids, BMP, COD), a bar plot colored by signal type is provided ( $A_{lab}$  in red,  $A_{ip}$  in blue,  $A_{\mu}$  in green,  $R_{ss}$  in violet,

750  $R_{ms}$  in orange, and  $R_{bs}$  in yellow).

# 1 Graphical abstract



Graphical Abstract - Various organic wastes substrates were collected (1), and spectra
were acquired on four different spectroscopic systems (2) with different compactness
and measurement types. Then, calibration models were built on each of these spectroscopic systems (3) to predict five different biochemical characteristics (BMP, DCO,
Sugar, Nitrogen, Lipids). The spectroscopic systems were compared and evaluated
based on the performances of these models.