



Dataset of organic sample near infrared spectra acquired on different spectrometers

Céline Chauvergne, Laurent Bonnal, Denis Bastianelli, Hélène Carrère, Yves Griveau, Marie-Pierre Jacquemot, Matthieu Reymond, Valérie Méchin, Virginie Rossard, Éric Latrille

► To cite this version:

Céline Chauvergne, Laurent Bonnal, Denis Bastianelli, Hélène Carrère, Yves Griveau, et al.. Dataset of organic sample near infrared spectra acquired on different spectrometers. *Data in Brief*, 2020, 32, pp.106264. 10.1016/j.dib.2020.106264 . hal-02934745

HAL Id: hal-02934745

<https://hal.inrae.fr/hal-02934745>

Submitted on 9 Sep 2020

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

Dataset of organic sample near infrared spectra acquired on different spectrometers



Céline Chauvergne , Laurent Bonnal , Denis Bastianelli ,
Hélène Carrère , Yves Griveau , Marie-Pierre Jacquemot ,
Matthieu Reymond , Valérie Méchin , Virginie Rossard ,
Éric Latrille

PII: S2352-3409(20)31158-6
DOI: <https://doi.org/10.1016/j.dib.2020.106264>
Reference: DIB 106264

To appear in: *Data in Brief*

Received date: 8 July 2020
Revised date: 18 August 2020
Accepted date: 27 August 2020

Please cite this article as: Céline Chauvergne , Laurent Bonnal , Denis Bastianelli , Hélène Carrère , Yves Griveau , Marie-Pierre Jacquemot , Matthieu Reymond , Valérie Méchin , Virginie Rossard , Éric Latrille , Dataset of organic sample near infrared spectra acquired on different spectrometers, *Data in Brief* (2020), doi: <https://doi.org/10.1016/j.dib.2020.106264>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 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/>)

Dataset of organic sample near infrared spectra acquired on different spectrometers

Céline Chauvergne¹, Laurent Bonnal^{2,3}, Denis Bastianelli^{2,3}, Hélène Carrère¹, Yves Griveau⁴, Marie-Pierre Jacquemot⁴, Matthieu Reymond⁴, Valérie Méchin⁴, Virginie Rossard¹, Éric Latrille^{1*}

1. INRAE, Montpellier Univ, LBE (102 Avenue des Étangs, 11100, Narbonne, France)
2. CIRAD, UMR SELMET, F-34398 Montpellier, France
3. SELMET, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France
4. Institut Jean-Pierre Bourgin, INRAE, AgroParisTech, Université Paris-Saclay, 78000, Versailles, France

Corresponding author

Éric Latrille (eric.latrille@inrae.fr)

Abstract

This dataset presents 127 raw near infrared spectra of different organic samples acquired on three different spectrometers in three different labs. An example of data processing is shown to create six spectra transfer models between the three spectrometers (two by two). In order to build and validate these transfer models, the dataset was split into two sets of spectra: a first set was used to compute six spectra transfer models thanks to the Piecewise Direct Standardisation function (PDS). A second set of spectra, independent of the first one was used to validate transfer models. Spectrum treatments and models were created on ChemFlow (<https://vm-chemflow-francegrille.eu/>), a free online chemometric software that includes all the necessary functions.

Keywords

Near infrared spectroscopy; spectrum transfer model; chemometrics; organic samples

Specifications Table

Subject	Spectroscopy
Specific subject area	Near infrared spectroscopy, chemometrics
Type of data	Table Figure
How data were acquired	Sample spectra were acquired using three different near infrared spectrometers: NIRFlex N-500 FT-NIR (BUCHI, France, serial number 1000001305) NIR-System5000 (FOSS, Denmark, serial number 94630632) ANTARIS II FT-NIR (Thermo Fisher Scientific, USA, serial number AHY0800337)
Data format	Raw

Parameters for data collection	Near infrared spectra of various organic samples were acquired on the three spectrometers in the standard using conditions. Samples were analysed in either vials or cups.
Description of data collection	Organic samples were dried and ground (<1 mm) before acquiring near infrared spectra.
Data source location	LBE, INRAE, Narbonne, France SELMET, CIRAD, Montpellier, France IJPB, INRAE, Versailles, France
Data accessibility	Data are available on public repositories: Repository name: Data INRAE (https://data.inrae.fr/) Data identification number: doi:10.15454/XFTLV4 Direct URL to data: https://doi.org/10.15454/XFTLV4 Repository name: Zenodo Data identification number: doi:10.15454/XFTLV4 Direct URL to data: https://zenodo.org/record/3908210#.Xvmx_igzZPY

Value of the Data

- These data can be used as supplements for the description of organic matter samples and can be compared to other studies.
- Spectra were acquired using three specific NIR spectrometers and they could be compared to spectra provided by other devices. Any researcher can use these data to perform chemometric analyses.
- This dataset shows the value of near infrared spectroscopy coupled with chemometrics for inter-laboratory works.

Data Description

Data provided in this article are 127 raw near infrared spectra of different organic samples acquired on three different spectrometers in three different labs.

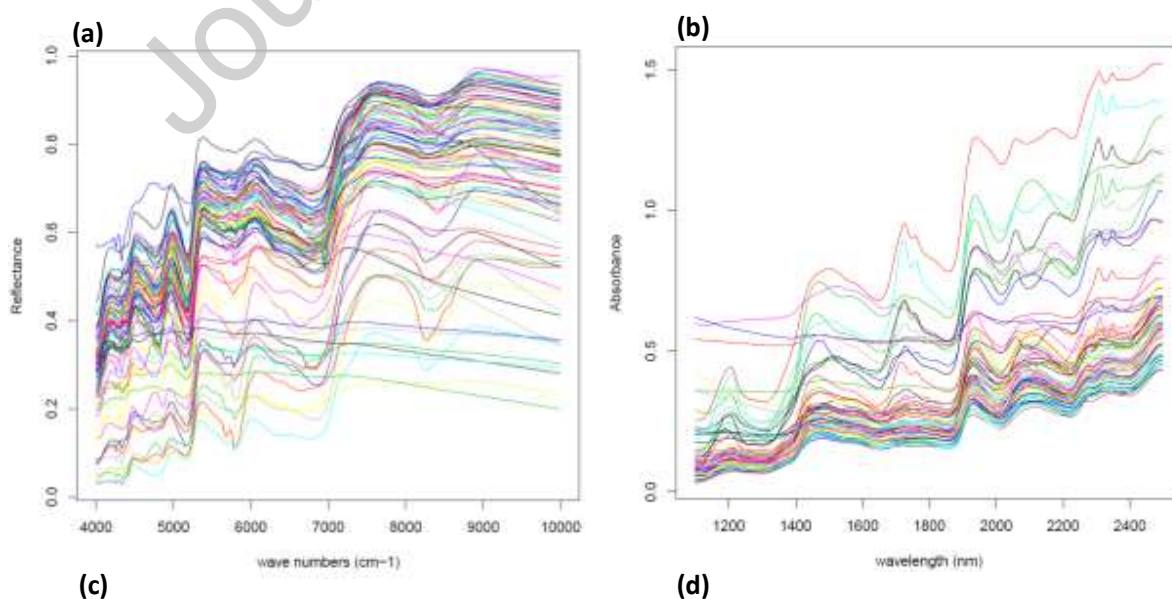
In order to build and validate transfer models, the dataset was split into two sets of spectra. A first set of 92 organic samples, listed in Appendix 1, was selected to cover a wide range of variation for cell wall composition to promote a wide range of spectra, in order to create spectra transfer models. First, spectra were acquired on the BUCHI NIRFlex N-500 device, in triplicate and then averaged (Fig.1a). Secondly, sample spectra were acquired on the FOSS NIRSystem5000 spectrometer in duplicate and

were automatically averaged (Fig.1b). Finally, spectra were acquired on the ThermoScientific ANTARIS II device ten times and then averaged (Fig.1c). Fig.1d and Fig.1e represents BUCHI NIRFlex N-500 spectra and ThermoScientific ANTARIS II spectra converted in nm as usually represented in the NIR field.

Spectrometer acquisition modalities were specific to each equipment: absorbance spectra in wavelengths (nm) for FOSS spectrometer, reflectance spectra in wave numbers (cm^{-1}) for BUCHI spectrometer and absorbance spectra in wave numbers (cm^{-1}) for ANTARIS spectrometer. Six spectra transfer models were created based on this first dataset with Piecewise Direct Standardization (PDS) [1]: transfer model from BUCHI to FOSS and vice versa, transfer model from BUCHI to ANTARIS and vice versa and transfer model from FOSS to ANTARIS and vice versa. Spectrum treatments and models were created on ChemFlow (<https://vm-chemflow-francegrille.eu/>) [2], a free online chemometric software that includes all the necessary functions.

Fig.2 is a schematic representation of the main functions used for the transfer model from a FOSS spectrum to a BUCHI spectrum. The resulting spectra transfer model is a transform matrix (Appendix 2), allowing transforming FOSS spectra into BUCHI spectra by matrix multiplication. The other five transfer models were established in a same way and transform matrices were used alike (data not shown in the article but available in the repository).

Then, a second set, of 11 sorghum and 24 miscanthus samples, listed in Appendix 3, was selected to validate transfer models and estimate the quality of these transfers. Sample spectra of validation set were acquired on the three devices with the same modalities as before (Figs. 3a, 3b, 3c, 3d and 3e) and transferred with previous transfer models (data not shown in the article but available in the repository).



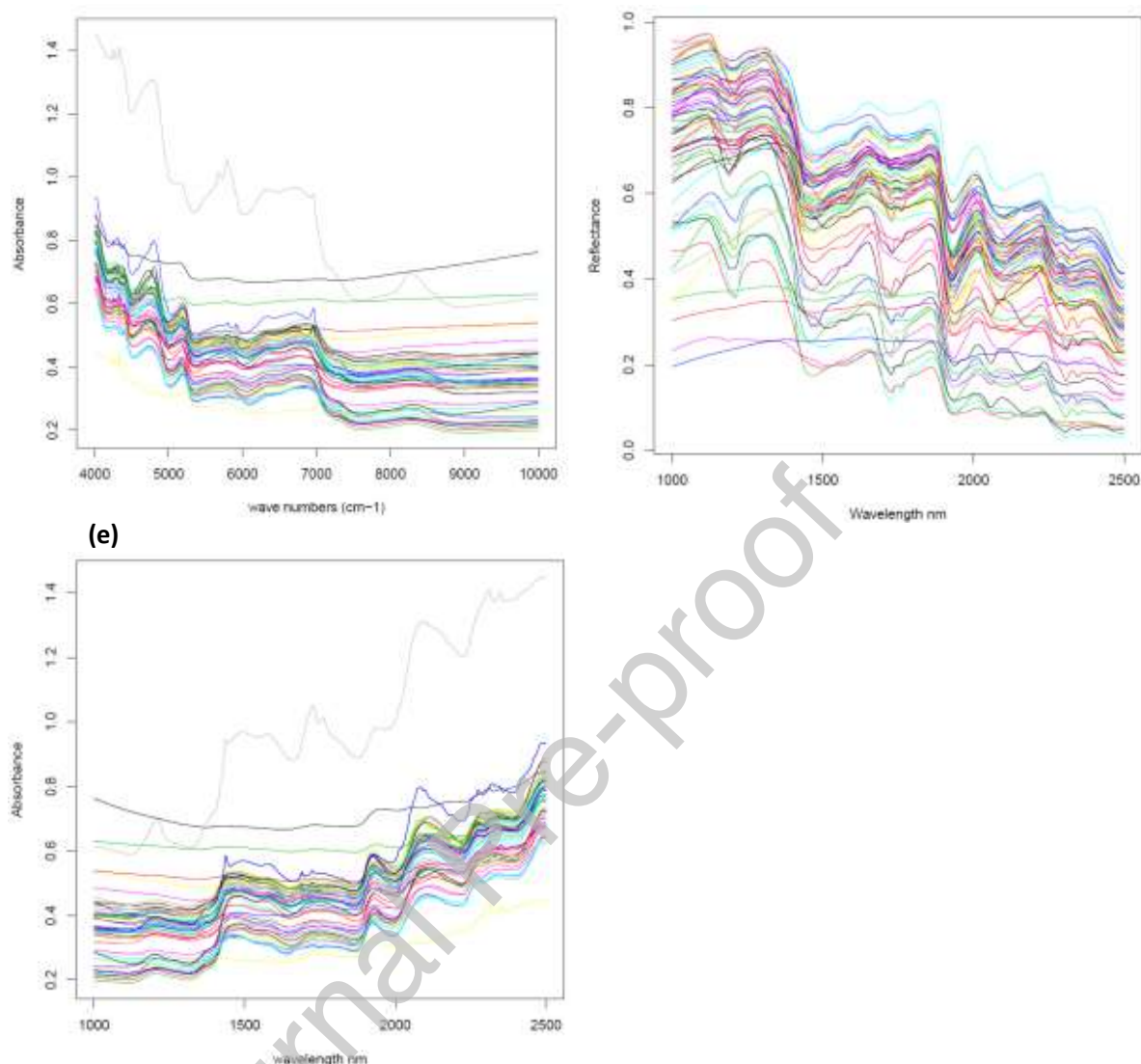


Fig. 1. **a.** 92 averaged BUCHI NIRFlex N-500 spectra of organic samples of the transfer set (raw spectra in reflectance, cm⁻¹)

b. 71 averaged FOSS NIRSystem5000 spectra of organic samples of the transfer set (raw spectra in absorbance, nm)

c. 53 averaged ThermoScientific ANTARIS II spectra of organic samples of the transfer set (raw spectra in absorbance, cm⁻¹)

d. 92 averaged BUCHI NIRFlex N-500 spectra of organic samples of the transfer set (converted spectra in reflectance, nm)

e. 53 averaged ThermoScientific ANTARIS II spectra of organic samples of the transfer set (converted spectra in absorbance, nm)

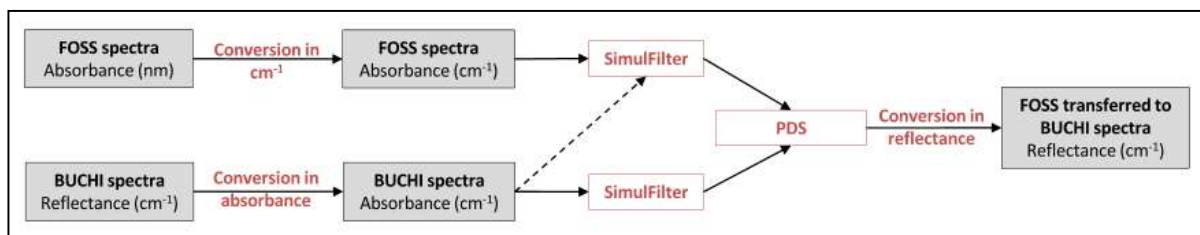


Fig.2. Schematic representation of the transfer model from a FOSS spectrum to a BUCHI spectrum.
 SimulFilter: algorithm for interpolations, PDS: Piecewise Direct Standardization

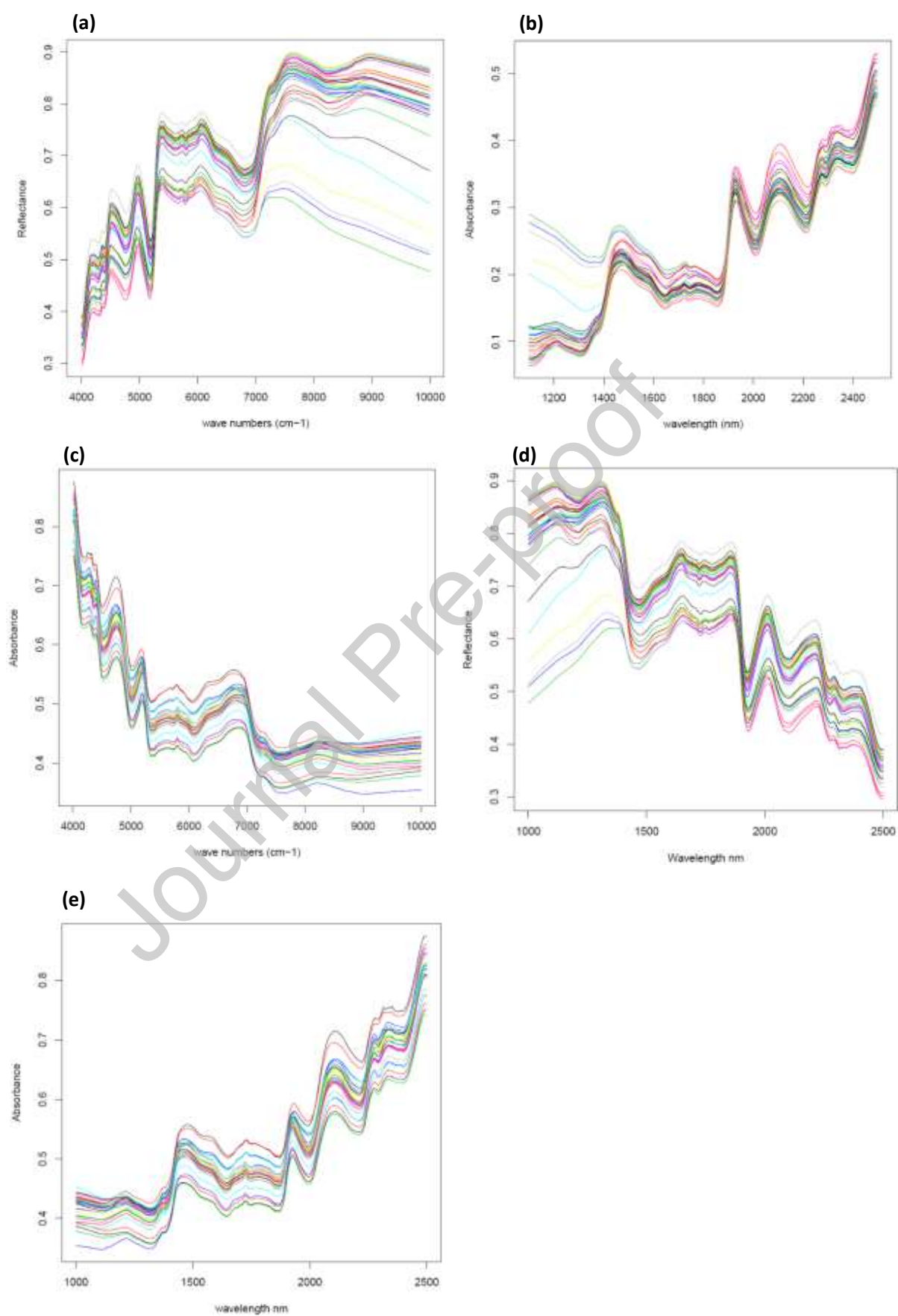


Fig. 3. **a.** 35 averaged BUCHI NIRFlex N-500 spectra of organic samples of the transfer validation set (raw spectra in reflectance, cm^{-1})
b. 35 averaged FOSS NIRSystem5000 spectra of organic samples of the transfer validation set (raw spectra in absorbance, nm)
c. 29 averaged ThermoScientific ANTARIS II spectra of organic samples of the transfer validation set (raw spectra in absorbance, cm^{-1})
d. 35 averaged BUCHI NIRFlex N-500 spectra of organic samples of the transfer validation set (converted spectra in reflectance, nm)
e. 29 averaged ThermoScientific ANTARIS II spectra of organic samples of the transfer validation set (converted spectra in absorbance, nm)

Experimental Design, Materials, and Methods

1. Sample preparation

The 127 raw samples were dried either by freeze-drying or by oven drying and then ground to 1 mm size before being packaged in hermetic bottles or bags. All samples were either shared between laboratories quickly so that the spectra could be acquired on similar conditions (no material degradation) or shared with a longer time frame.

2. Data collection

FOSS spectrometer NIRSystem5000 gave absorbance spectra in wavelength with a range of 1100 nm to 2498 nm and a resolution of 2 nm. Each spectrum was an average of 32 spectral readings of the same cup. All of the 106 (71+35) samples were poured into 3.5 cm diameter cups (about 4 g) and spectra were acquired in duplicate and automatically averaged. The output data format was .csv, compatible with ChemFlow, so data were directly imported into the software and converted in .tsv format.

BUCHI N-500 spectrometer gave reflectance spectra in wave numbers, with a range of 4000 cm^{-1} to 10000 cm^{-1} and with a resolution of 4 cm^{-1} . Each spectrum was an average of 96 spectral readings of the same cup or vial. All of the 127 (92+35) samples were poured either into 10 cm diameter cups (about 30 g) or into 1.3 cm diameter vials (about 3 g) and spectra were acquired in triplicate and then averaged.

ThermoScientific ANTARIS II spectrometer gave absorbance spectra in wave numbers with a range of 4000 cm^{-1} to 10001 cm^{-1} and with a resolution of 3.86 cm^{-1} . All of the 82 (53+29) samples were divided into 3 or 10 subsamples (aliquots) and poured into 0.8 cm diameter vials (about 2 ml) which was scanned once. For each subsample, an average (spectrum) of 64 spectral readings was recorded.

The output data format for BUCHI and ANTARIS spectrometers was .jdx, compatible with ChemFlow, so data were directly imported into the software and converted in .tsv format.

3. Transfer model creation

In order to create FOSS to BUCHI spectra transfer model, spectra of the transfer set needed pretreatments to match the spectral modalities. To use the PDS function, the spectral modalities used were absorbance and cm^{-1} unit. Therefore, the first step was to transform BUCHI spectra in absorbance with the logarithm function and FOSS spectra in cm^{-1} unit with a simple unit conversion. Then, as spectra did not have the same unit values, a SimulFilter function (extrapolation function with unit width of 2) was applied to match FOSS spectra wave numbers with BUCHI spectra values. The transfer model was created with the PDS function, with a number of channels in each window equal to 7. Finally, FOSS output spectra were transformed in reflectance spectra with power function (10^x) to fit with BUCHI spectral modalities.

Acknowledgments

This work was supported by the Biomass For the Future project from the French National Research Agency (ANR, grant ANR-11-BTBR-0006-BFF). The authors acknowledge the contribution of the ChemHouse Research Group, Montpellier, France. This work has benefited from the support of IJPB's Plant Observatory technological platforms. The IJPB benefits from the support of Saclay Plant Sciences-SPS (ANR-17-EUR-0007). The authors acknowledge the contribution of Bio2E platform, INRAE, Environmental Biotechnology and Biorefinery Platform (DOI: 10.15454/1.557234103446854E12).

Competing Interests

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.

Appendix

Appendix 1 is available at this link: <https://doi.org/10.15454/XFTLV4/GCPQ0Z>

Appendix 2 is available at this link: <https://doi.org/10.15454/XFTLV4/ZMTH60>

Appendix 3 is available at this link: <https://doi.org/10.15454/XFTLV4/89FTMC>

References

- [1] Y. Wang, D.J. Velthkamp, B.R. Kowalski, Multivariate instrument standardization, *Anal. Chem.*, 63, 23, (1991) 2750-2756.
- [2] V. Rossard, J.C. Boulet, F. Gogé, E. Latrille, J.M. Roger, ChemFlow, chemometrics using Galaxy. (2016) Presented at Galaxy Community Conference - GCC2016, Bloomington, USA (2016-06-24 - 2016-07-29). <https://doi.org/10.7490/f1000research.1112573.1>