



HAL
open science

Near infrared spectroscopy applied to solid organic waste how to avoid water effects?

Alexandre Mallet

► **To cite this version:**

Alexandre Mallet. Near infrared spectroscopy applied to solid organic waste how to avoid water effects?. Chemical and Process Engineering. Montpellier SupAgro, 2021. English. NNT: 2021NSAM0026 . tel-04049594

HAL Id: tel-04049594

<https://theses.hal.science/tel-04049594>

Submitted on 28 Mar 2023

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.

THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE MONTPELLIER SUPAGRO

École doctorale: GAIA – Biodiversité, Agriculture, Alimentation, Environnement, Terre, Eau
Spécialité: Agroressource, Procédés, Aliments, Bioproduits (APAB)

Menée à

l'Institut national de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE) au sein du Laboratoire de Biotechnologie de l'Environnement (LBE) à Narbonne, et de l'unité mixte de recherche Information, Technologie, Agro-Procédés (ITAP) à Montpellier

Caractérisation spectroscopique robuste des déchets organiques: comment s'affranchir des effets de l'eau?

Devant être défendue par Alexandre MALLET
Le 17 septembre 2021

Sous la direction de Dr. Jean-Philippe STEYER et Dr. Jean-Michel ROGER

Devant le jury composé de

Véronique BELLON-MAUREL, Directrice de recherche, INRAE-ITAP, France	[Examinatrice]
Cyrille CHARNIER, Directeur technique, BIOENTECH, France	[Encadrant]
Anna DE JUAN, Professeur associé, Université de Barcelone, Espagne	[Rapportrice]
Jean-Michel ROGER, Ingénieur de recherche, INRAE-ITAP, France	[Co-directeur]
Heinz Wilhelm SIESLER, Professeur, Université de Duisburg-Essen, Allemagne	[Rapporteur]
Jean-Philippe STEYER, Directeur de recherche, INRAE-LBE, France	[Co-directeur]



UNIVERSITÉ
DE MONTPELLIER

THESIS FOR THE OBTENTION OF THE DEGREE OF DOCTOR OF PHILOSOPHY OF MONTPELLIER SUPAGRO

Doctoral School: GAIA – Biodiversity, Agriculture, Food, Environnement, Soil, Water
Specialty: Agroressources, Processes, Food, Bio-products (APAB)

Carried at
the National Research Institute for Agriculture, Food and Environment (INRAE) of France, at the
Laboratory of Biotechnology for Environment (LBE) in Narbonne, and at the mixed research unit
of Information, Technology, Agro-Processes (ITAP) in Montpellier.

Near infrared spectroscopy applied to solid organic waste: how to avoid water effects?

To be defended on the 17th of September 2021 by
Alexandre MALLET

Under the direction of Dr. Jean-Philippe STEYER and Dr. Jean-Michel ROGER

In front of the jury composed by

Véronique BELLON-MAUREL, Research Director, INRAE-ITAP, France	[Referee]
Cyrille CHARNIER, Chief Technical Director, BIOENTECH, France	[Supervisor]
Anna DE JUAN, Associate Professor, University of Barcelona, Spain	[Reviewer]
Jean-Michel ROGER, Research Engineer, INRAE-ITAP, France	[Co-director]
Heinz Wilhelm SIESLER, Professor, University of Duisburg-Essen, Germany	[Reviewer]
Jean-Philippe STEYER, Research Director, INRAE-LBE, France	[Co-director]



UNIVERSITÉ
DE MONTPELLIER

Preface

This thesis is submitted in order to obtain the PhD degree from the GAIA Doctoral school from the University of Montpellier and Montpellier SupAgro/l'institut Agro.

The work results from a joint collaboration between the National Research Institute for Agriculture, Food and Environment (INRAE) and the company BioEnTech, and was financially supported by the National Agency of Technology and Research (ANRT) through its CIFRE program (grant number 2018/0461). Two laboratories from INRAE were involved in this project: the Laboratory of Biotechnology and Environment (LBE) in Narbonne (France) and the joint research unit Informations, Technologies for Agro-Processes (ITAP) in Montpellier (France).

The research was co-directed by Dr. Jean-Philippe STEYER (LBE, INRAE) and Dr. Jean-Michel ROGER (ITAP, INRAE), and was co-supervised by Dr. Cyrille CHARNIER (BioEnTech), Dr. Eric LATRILLE (LBE, INRAE) and Dr. Ryad BENDOULA (ITAP, INRAE).

The project lasted three years, from September 2018 to September 2021.

This PhD work has been awarded by the Agreenium International Research School (EIR-A) label, in result of an international collaboration with Professor Roumiana TSENKOVA from the Laboratory of Bio-measurement of Kobe University (Kobe, Japan).

The present thesis is divided in two parts. In part I, seven chapters are presented: the Chapter I introduces the state of the art and the thesis' research questions, the Chapter II to Chapter VI present the different contributions of this thesis and the Chapter VII provides general conclusions and perspectives. In part II, the associated published papers are presented.

You can't trust water: Even a straight stick turns crooked in it. W. C. Fields

Acknowledgements

First of all, I would like to thank the reviewers, Prof. Heinz Siesler and Prof. Anna De Juan, and other members of the jury, Dr. Véronique Bellon-Maurel, Dr. Jean-Philippe Steyer, Dr. Jean-Michel Roger, and Dr. Cyrille Charnier for accepting to evaluate this work.

Naturally, I would like to warmly thank my supervising team, Dr. Jean-Philippe Steyer, Dr. Jean-Michel Roger, Dr. Ryad Bendoula, Dr. Eric Latrille, and Dr. Cyrille Charnier. What a great team you have been! 5 top experts, 5 ways of doing research, 5 senses of humor, 5 highly dedicated teammates, it has been a highly enriching journey, thank you so much for your confidence, your guidance and your support during these three years.

To the big family of LBE, what a great lab with good and fun people, thank you! So many good memories, among them, I name Eric's regular kayak outings, the traditional Saturday sailing at la Nautique followed by the market's raviolis, the climbing sessions with Ze Clape Team, the beachvolleys at the Chalet-Plage, the traditional welcome party of interns, the ski sessions in the Pyrénées with Roland's papamobile, the Routier for generously taking care of our Summer body, the highly serious sessions at l'Echoppe (sometimes continued at the Live, thanks Kevin), the board games dinners (thanks Ulysse for the legacy), longlasting coffee breaks with Denis my passionate beekeeping mentor, and many more.

Cheers to my colleagues from BioEnTech with whom the beach volleyball games during those three years have always been a refreshment. A special thanks to Morand, my Bioentech partner at the lab.

To all members of the ChemHouse group at Montpellier, thank you for the numerous presentations and discussions around chemometrics; long live this science (and the associated pasta lunches)!

I am particularly grateful towards the aquaphotomics team (Prof. Roumiana Tsenkova and Dr. Jelena Muncan) of Kobe University for warmly welcoming me in their lab and sharing with me the fascinating mysteries of water, as well as introducing me to the Japanese culture.

Finally, I would like to thank all my friends and family who have supported me during these three years. To the one that was by my side throughout this experience, a big thank you.

Eager to find out what comes next!

Publications and communications

1. Papers in international peer-reviewed journals

- ❖ A. Mallet, C. Charnier, E. Latrille, R. Bendoula, J.-P. Steyer, J.-M. Roger, Unveiling non-linear water effects in near infrared spectroscopy: A study on organic wastes during drying using chemometrics, *Waste Manag.* 122 (2021) 36–48. <https://doi.org/10.1016/j.wasman.2020.12.019>.
- ❖ A. Mallet; R. Tsenkova, J. Muncan, C. Charnier, E. Latrille, R. Bendoula, J.-P. Steyer, J.-M. Roger, Relating Near-Infrared Light Path-Length Modifications to the Water Content of Scattering Media in Near-Infrared Spectroscopy: Toward a New Bouguer-Beer-Lambert Law. *Anal. Chem.* 93 (2021), 6817–6823. <https://doi.org/10.1021/acs.analchem.1c00811>.
- ❖ A. Mallet, M. Pérémé, L. Awhangbo, C. Charnier, J. Roger, J. Steyer, E. 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. <https://doi.org/10.1016/j.wasman.2021.03.045>.
- ❖ M. Pérémé, A. Mallet, L. Awhangbo, C. Charnier, J.-M. Roger, J.-P. Steyer, E. Latrille, R. Bendoula, 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 Br.* 36 (2021) 107126. <https://doi.org/10.1016/j.dib.2021.107126>.
- ❖ A. Mallet, B. Zennaro, C. Charnier, E. Latrille, R. Bendoula, J.-P. Steyer, J.-M. Roger, Limitations of global correction approaches for moisture content correction in the context of organic waste characterization (to be submitted shortly)

2. Oral communications

- ❖ A. Mallet, B. Zennaro, J.M. Roger, C. Charnier, R. Bendoula, E. Latrille, J.-P. Steyer, Robust near-infrared spectroscopic characterization of wet organic wastes: how to avoid water interferences ?, in: 19th HelioSPIR Annual Meeting, Montpellier, France, 2018. <https://hal.inrae.fr/hal-03196769>.

Publications and communications

- ❖ A. Mallet, E. Latrille, R. Bendoula, C. Charnier, J.-P. Steyer, J.-M. Roger, Study of rehydration of different organic substrates by means of near-infrared spectroscopy and signal de-mixing techniques, in: Conférence Chimiométrie, SupAgro, Montpellier, 2019. <https://hal.inrae.fr/hal-03196782>.
- ❖ A. Mallet, E. Latrille, R. Bendoula, C. Charnier, J.-P. Steyer, J.-M. Roger, Study of the dehydration of different organic substrates by means of near-infrared spectroscopy and chemometrics: first steps towards a correction strategy, in: NIR 2019 - 19th Biennial Meeting International Council of NIR Spectroscopy (ICNIRS), Brisbane, Australia, 2019. <https://hal.inrae.fr/hal-03196757>.
- ❖ A. Mallet, R. Tsenkova, J. Muncan, C. Charnier, E. Latrille, R. Bendoula, J.-P. Steyer, J.M. Roger, The effects of water on scattering: taking into account path-length modifications, in: 4th Aquaphotomics International Conference (AIC). Exploring Water Molecular Systems in Nature., Kobe, Japan, 2021. <https://hal.inrae.fr/hal-03196791>.
- ❖ A. Mallet, R. Tsenkova, J. Muncan, C. Charnier, E. Latrille, R. Bendoula, J.-P. Steyer, J.M. Roger, Taking into account light path-length modifications induced by water in near infrared spectroscopy, in: 21th HelioSPIR Annual Meeting, Online conference, France, 2021.

Résumé étendu

La décarbonisation de la production d'énergie a été identifiée par le groupe d'experts intergouvernemental sur l'évolution du climat (GIEC) comme un levier d'action clé pour mitiger les risques associés au réchauffement climatique. La valorisation des déchets organiques à travers le procédé de digestion anaérobie (méthanisation) est une des solutions durables envisagées. Ce procédé biologique permet, dans des conditions anaérobies, de dégrader la matière organique et de produire du biogaz (mélange de méthane et gaz carbonique) ainsi qu'une boue résiduelle, appelée digestat, réutilisable en tant que fertilisant dans l'agriculture. Bien que les unités de méthanisation ont historiquement fonctionné avec un seul type de déchets (fumiers et lisiers, ou cultures énergétiques comme le maïs), le développement actuel de la filière passe par la co-digestion d'une plus grande diversité de déchets organiques tels que des déchets urbains (boues de station d'épuration, fraction fermentescible des ordures ménagères), et déchets de l'agro-alimentaire (drèches brassicoles, huiles et graisses). Cette co-digestion permet de fait une plus large valorisation des déchets organiques existants, mais également une meilleure stabilité du procédé et des rendements en biogaz accrus.

Cependant, les digesteurs actuels souffrent d'une sous-optimisation de leur production de gaz, notamment dû à des difficultés pour rationaliser le choix de la recette d'alimentation. En effet, le type et la quantité de déchets organiques co-digérés sur un site peuvent varier quotidiennement en fonction des gisements de déchets locaux, des pratiques culturelles, des périodes de l'année, du transport et du stockage. Or, une mauvaise décision sur la recette d'alimentation peut entraîner des inhibitions du procédé biologique pouvant aller jusqu'à provoquer l'arrêt complet du procédé sur plusieurs mois. Dans ce contexte, l'optimisation et le contrôle du procédé reposent sur trois éléments: un suivi en temps réel des caractéristiques des intrants, un suivi en temps réel de l'état biologique dans le réacteur, et enfin, une modélisation du procédé de digestion anaérobie qui permet de simuler et anticiper les performances en fonction des différents paramètres mesurés. Malheureusement, des solutions sur site qui permettent le suivi en temps réel des caractéristiques des intrants et qui sont adaptées à une diversité grandissante de matrices physiques et biochimiques semblent toujours faire défaut.

Dans ce contexte, des technologies reposant sur l'interaction de la lumière avec le milieu telles que la spectroscopie proche infra-rouge (SPIR) ont été proposées comme solutions rapides et fiables de caractérisation des déchets organiques. En effet, la SPIR est une technique non-destructive particulièrement adaptée à la caractérisation de matrices organiques complexes telles que les végétaux, les sols, les aliments et les déchets. Des développements ont permis d'utiliser la SPIR pour prédire

Résumé étendu

différentes caractéristiques importantes des déchets organiques telles que le potentiel biochimique méthanogène (BMP), les teneurs en lipides, glucides, protéines, la demande chimique en oxygène (DCO), ou encore la cinétique de production de méthane. Le coût d'analyse de l'ensemble de ces caractéristiques était jusqu'alors prohibitif, mais surtout très long (*e.g.*, plus d'un mois pour le BMP). Aujourd'hui, l'analyse complète d'un déchet peut-être faite en quelques jours seulement. Cependant, des étapes de lyophilisation et de broyage sont nécessaires pour réduire les effets de l'eau et de la granulométrie sur le spectre mesuré. En pratique, cette préparation de la matière limite l'adoption de cette technologie sur site ou en ligne. Cela empêche donc de suivre en temps réel la composition des déchets et donc de raisonner le choix de la recette d'alimentation. D'autres contraintes associées à la préparation de matière peuvent être évoquées telles que la perte des composés volatils pendant le séchage (tels que les acides gras volatils (AGV)), ou encore l'échantillonnage et le transport de la matière jusqu'au laboratoire d'analyse. Fort de ce constat, il apparaît nécessaire de trouver des moyens de s'affranchir des effets de l'eau en SPIR et ainsi permettre une caractérisation robuste et sur site des déchets bruts et humides.

Dans ce contexte applicatif, cette thèse s'intéresse à **(i) mieux caractériser les effets de l'eau sur la SPIR** et **(ii) proposer des stratégies de modélisation pour s'en affranchir**.

Pour répondre à ces objectifs, un dispositif expérimental a été conçu pour collecter de manière dynamique les spectres de réflectance ainsi que la teneur en eau de 90 substrats variés pendant leur séchage dans des conditions standard de température et de pression. Cela a permis de constituer une base de données unique de plus de 150 000 spectres proche-infrarouge qui couvrent une large gamme de matrices biochimiques (gras, sucres, fibres), et de teneurs en eau (de 1% à 99%).

En exploitant ces données, cette thèse propose une évaluation des méthodes de chimiométrie qui permettent de rendre un modèle prédictif existant plus robuste par rapport à une grandeur d'influence (*e.g.*, température, instrumentation, granulométrie, teneur en eau). Une synthèse des méthodes est ainsi présentée, en classant les méthodes selon l'élément de la chaîne de modélisation concerné (les spectres, le modèle, les prédictions) et selon si les méthodes reposent sur des données supplémentaires sur la grandeur d'influence ou non. Plusieurs études montrent que ces méthodes ont permis d'améliorer avec succès la robustesse de modèles prédictifs dans différents contextes (sur du sol ou des végétaux), mais toujours sur des gammes restreintes de teneur en eau (<20%) et pour des types biochimiques homogènes. Diverses méthodes (projections orthogonales et obliques, normalisation, sélection de variable, enrichissement, pondération, et superposition de modèles) ont donc été évaluées dans le cadre de la caractérisation de déchets organiques humides et variés. Le but était de corriger un modèle de prédiction

Résumé étendu

du potentiel biochimique méthanogène (BMP) fonctionnant sur différents déchets organiques secs pour pouvoir l'appliquer sur des déchets humides couvrant une large gamme de teneur en eau (1% à 99%). Les résultats montrent que le modèle d'origine présente une erreur (RMSEP) de 65 mL(CH₄).gTS⁻¹ sur les substrats secs, et une erreur de 192 mL(CH₄).gTS⁻¹ sur les substrats humides. La plupart des méthodes employées ont permis de réduire fortement cette erreur sur les substrats humides jusqu'à une erreur de 93 mL(CH₄).gTS⁻¹ pour une méthode simple d'enrichissement de la base. Bien que cela soit encourageant, ces erreurs restent importantes et soulignent les limites d'une approche globale de correction des effets de l'eau dans le contexte de la caractérisation de déchets organiques variés. Par ailleurs, pour une même gamme de teneur en eau, l'erreur sur les prédictions diffère fortement selon le substrat. Cela nous indique que la sensibilité aux variations de teneur en eau et/ou le type d'effet de l'eau dépend de la nature du substrat, et donc qu'un modèle linéaire global ne permet pas d'aboutir à une robustesse suffisante. L'ensemble de ces résultats fait l'objet d'un article en cours de soumission.

A travers l'analyse en composantes principales (ACP), les effets complexes de l'eau sur les spectres ont été détaillés, avec la mise en évidence d'effets à la fois physiques et chimiques. Une des conséquences de la grande diversité biochimique retrouvée dans les déchets organiques est que les propriétés physiques, et donc l'interaction de la lumière avec la matière, diffèrent fortement entre substrats. Ainsi, par exemple, des substrats liquides ou transparents comme de l'eau sucrée, du poisson ou du beurre, vont avoir tendance à présenter des niveaux de pseudo-absorbance globale (non spécifique à des régions spectrales) plus élevés (~2.5) que des substrats solides comme de la paille, du fumier, ou de la poudre de lait (~1.5). Ceci s'explique par des différences de niveaux de diffusion dans la matière. En observant l'évolution des scores sur la première composante de l'ACP, on a pu mettre en évidence que l'eau, en modifiant les propriétés physiques des substrats, modifiait ces niveaux globaux de pseudo-absorbance. La plupart des substrats présentent une augmentation de la pseudo-absorbance en fonction de la teneur en eau : au fur et à mesure du séchage, l'eau est remplacée par l'air ce qui provoque une augmentation des différences d'indices de réfraction et donc de la réflexion diffuse. Pourtant, certains substrats à forte teneur en lipides, présentent un comportement inverse avec une diminution de la pseudo-absorbance en fonction de la teneur en eau. Dans ces systèmes, au cours du séchage, l'eau est remplacée par des éléments lipidiques et non de l'air, ce qui entraîne au contraire une augmentation des indices de réfraction. Un premier comportement non-linéaire a pu donc être mis en évidence. De plus, l'évolution de ces changements de diffusion induits par l'eau n'est pas linéaire avec la teneur en eau. Par ailleurs, il a pu être mis en évidence que l'eau avait un effet chimique sur le spectre. En effet, le niveau d'absorption de la lumière par les liaisons OH (notamment à 1209 nm, 1456 nm et 1933 nm) est directement affecté par la présence d'eau dans le milieu. Ainsi, l'évolution de la signature spectrale des autres constituants

Résumé étendu

(lipides, sucres, protéines) évoluent directement avec la teneur en eau. Par ailleurs, certaines composantes de l'ACP nous ont permis de mettre en évidence que la teneur en eau modifie non seulement la quantité d'eau dans le système mais également l'état de l'eau. Ainsi, l'eau, selon le nombre de liaisons hydrogènes qu'elle forme, ne présente pas les mêmes signatures spectrales d'absorption. Des décalages de pics autour de 1430 nm peuvent notamment être observés. A travers cette étude, nous avons donc pu mettre en évidence la non-linéarité des effets de l'eau en SPIR, avec une dépendance selon le type biochimique, les propriétés physiques et la gamme de teneur en eau. L'ensemble de ces résultats a fait l'objet d'une publication dans le journal *Waste Management* [Mallet, A.; Charnier, C.; Latrille, É.; Bendoula, R.; Steyer, J. P.; Roger, J. M. *Unveiling Non-Linear Water Effects in near Infrared Spectroscopy: A Study on Organic Wastes during Drying Using Chemometrics. Waste Manag.* **2021**, *122*, 36–48. <https://doi.org/10.1016/j.wasman.2020.12.019>].

Afin de mieux comprendre les effets physiques de l'eau sur la diffusion, un milieu modèle simple (composé d'aluminium mélangé à de l'eau) a été étudié au cours de son séchage. Le fait que l'aluminium réfléchit quasiment entièrement la lumière nous a permis d'étudier les effets de l'eau indépendamment de toute interaction chimique liée aux absorbants de la matière sèche. Dans le cadre de cette expérience, une modification de la loi de Bouguer-Beer-Lambert a été proposée pour les milieux humides diffusants où le trajet optique est directement relié à la teneur en eau par une fonction simple de puissance. Ainsi, plus on a d'eau dans le système, plus le trajet optique est allongé, et plus la probabilité d'absorption de la lumière par les molécules d'eau est importante. Différentes implications de cette loi sont discutées. Tout d'abord, ce modèle empirique est particulièrement intéressant dans le cadre d'études spectroscopiques de l'eau dans des milieux diffusants où la position et l'intensité relative des pics sont recherchées. En effet, cette loi nous permet d'identifier les coefficients purs d'absorptivité. Cela offre par exemple de nouvelles perspectives pour les recherches qui portent sur la compréhension de la structure de l'eau dans les milieux diffusants. Par ailleurs, cette loi nous confirme que la relation entre le signal mesuré et la teneur en eau est non-linéaire, et qu'elle est modélisable sur des milieux simples par une loi de puissance. Tout d'abord, cela permet de mieux comprendre les limites observées des méthodes d'algèbre linéaire classiquement utilisées en SPIR (telle que la régression par les moindres carrés partiels (PLS)) pour des milieux humides diffusants. Mais surtout, dans le cadre de la prédiction de matière sèche, nous montrons qu'en prenant en compte cette relation de puissance (*e.g.*, en passant au logarithme), nous pourrions par exemple construire des modèles plus simples et donc plus robustes. L'ensemble de ces résultats a fait l'objet d'une publication dans le journal de *Analytical Chemistry* [Mallet, A.; Tsenkova, R.; Muncan, J.; Charnier, C.; Latrille, É.; Bendoula, R.; Steyer, J. P.; Roger, J. M. *Relating Near-Infrared Light Path-Length Modifications to the Water Content of Scattering Media*

Résumé étendu

in Near-Infrared Spectroscopy: Toward a New Bouguer-Beer-Lambert Law. Anal. Chem. **2021**, *93* (17), 6817–6823. <https://doi.org/10.1021/acs.analchem.1c00811>].

En conclusion, nous évaluons une approche qui consiste à construire des modèles locaux basés sur des groupes d'échantillons aux comportements similaires vis-à-vis de l'eau. En particulier, la résolution de courbes multivariées (MCR-ALS) est proposée comme un outil pertinent pour identifier ces groupes.

Dans le contexte applicatif de cette thèse, des travaux complémentaires ont porté sur la comparaison des performances de différents spectromètres portables appliqués aux déchets organiques, avec des modes de mesures innovants tels que la polarisation. Les résultats montrent le potentiel de ces outils pour la caractérisation de déchets organiques sur site. L'ensemble de ces résultats a fait l'objet d'une publication dans le journal *Waste Management* [Mallet, A.; Péréomé, M.; Awhangbo, L.; Charnier, C.; Roger, J. M.; Steyer, J. P.; Latrille, É.; Bendoula, R. *Fast At-Line Characterization of Solid Organic Waste: Comparing Analytical Performance of Different Compact near Infrared Spectroscopic Systems with Different Measurement Configurations. Waste Manag.* **2021**, *126*, 664–673. <https://doi.org/10.1016/j.wasman.2021.03.045>] ainsi que la publication d'un data paper dans le journal *Data in Brief* [Péréomé, M.; Mallet, A.; Awhangbo, L.; Charnier, C.; Roger, J. M.; Steyer, J. P.; Latrille, É.; Bendoula, R. *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 Br.* **2021**, *36*, 107126. <https://doi.org/10.1016/j.dib.2021.107126>].

Abstract (Français)

La spectroscopie proche infra-rouge (SPIR) est aujourd'hui une technologie mature et fiable qui permet la caractérisation de matières organiques complexes tels que les végétaux, le sol ou les déchets organiques. En particulier, la SPIR a pu être utilisée pour caractériser des déchets organiques complexes et divers (tels que les fumiers/lisiers, les résidus de culture, les déchets de l'agro-alimentaire, ou les boues de stations d'épuration) qui peuvent être utilisés dans les procédés de digestion anaérobie (méthanisation). En revanche, des étapes de lyophilisation et de broyage sont aujourd'hui nécessaires pour éviter les effets de l'eau et de la granulométrie en SPIR, ce qui empêche toute utilisation sur site ou en ligne. Dans ce contexte, cette thèse s'intéresse aux effets de l'eau sur la SPIR. L'objectif est de mieux comprendre ces effets et de proposer des stratégies pour s'en affranchir. Différents aspects de l'étalonnage multivarié et des questions de robustesse sont ainsi discutés. Un dispositif expérimental a été conçu pour collecter de manière dynamique les spectres de réflectance ainsi que la teneur en eau de 90 substrats variés pendant leur séchage. Cela a permis de constituer une base de données importante de 150 000 spectres qui couvrent une large gamme de matrices biochimiques (gras, sucres, fibres), et de teneurs en eau (de 1% à 99%). A travers l'analyse en composantes principales, les effets complexes de l'eau ont été détaillés, avec des effets physiques et chimiques. La non-linéarité de ces effets a été mise en évidence, avec une dépendance selon le type biochimique, les propriétés physiques et la gamme de teneur en eau. Dans une deuxième étude, les effets physiques de l'eau sur la diffusion ont été étudiés. En se basant sur un système modèle (de l'aluminium mélangé à de l'eau), il a été démontré que les variations de teneur en eau provoquent des modifications du trajet optique de la lumière. Une modification de la loi de Bouguer-Beer-Lambert a ainsi été proposée pour les milieux humides diffusants où le trajet optique est directement relié à la teneur en eau par une fonction de puissance. Les implications sur les recherches menées sur la structure de l'eau dans la matière, ainsi que l'étalonnage quantitatif pour prédire la matière sèche par SPIR sont discutées. La dernière partie du travail consiste à présenter le potentiel de cette nouvelle loi sur des milieux complexes que sont les déchets organiques. La résolution de courbes multivariées (MCR-ALS) est notamment proposée comme un outil pertinent pour identifier différents groupes de substrats aux comportements similaires vis-à-vis de l'eau.

Mots-clés : Digestion anaérobie ; Spectroscopie proche infra-rouge ; Effets de l'eau ; Robustesse.

Abstract (English)

Near infrared spectroscopy (NIRS) is today a mature and reliable technology allowing the characterization of complex organic materials such as crops, soil, and organic waste. In particular, NIRS allows to characterize highly diverse and complex organic waste (such as animal manure, crop residues, food and catering waste, wastewater treatment sludge, etc.) that are used in the anaerobic digestion processes. However, the current freeze-drying and grinding steps that are required today to avoid water and particle size effects impede any on-site and online applications. In light of this, this thesis tackles the topic of water effects in NIRS. The objective is to better develop our understanding of these water effects in order to develop new strategies for building robust calibrations. Various aspects of multivariate modeling and robustness issues are therefore discussed. A dedicated experimental set-up was designed to collect dynamically and simultaneously the reflectance spectra of 90 various substrates during their drying with its estimated water content. This allowed us to constitute a unique dataset of 160 000 spectra covering a wide range of biochemical compositions (fat, carbohydrates, fibrous), physical properties (crystalline/amorphous, solid/liquid) and a wide range of moisture content (1% to 99%). The effects of water on the collected spectra were analyzed using principal components analysis (PCA) and shown to be both chemical and physical effects. This first study allowed us to unveil the complexity and non-linearity of these effects, and to demonstrate its dependence on the biochemical composition, on the physical properties and on the range of moisture content. A local approach is proposed to account for these effects. In a second study, the physical effects of water on scattering were further investigated. Using a model system (aluminum pellets mixed with water), water variations were shown to induce path-length modifications. A modification of the Bouguer-Beer-Lambert law was thus proposed for wet scattering media where path-length was shown to be directly related to a power function of the moisture content. The implications of these results for research on water structure using NIRS, and moisture content determination using NIRS are discussed. The last part of the work consists in further investigating how this new law behaves in more complex systems such as organic waste. Multivariate curve resolution (MCR-ALS) is proposed as an appropriate methodology to explore the water effects in these different systems; and further define the classes of substrates with similar behaviors regarding water.

Keywords : Anaerobic digestion ; Near infrared spectroscopy ; Water effects ; Robustness.

Table of Contents

Preface	v
Acknowledgements	vii
Publications and communications	ix
1. Papers in international peer-reviewed journals	ix
2. Oral communications	ix
Résumé étendu	xi
Abstract (Français)	xvii
Abstract (English)	xix
Table of Contents	xxi
List of Figures	xxv
List of Tables	xxix
List of symbols and abbreviations	xxxi
Chapter I. Introduction	1
1. Motivation	1
1.1 Valorization of organic waste through anaerobic digestion (AD).....	1
1.2 Near infrared spectroscopy (NIRS) and chemometrics for bioprocess monitoring.....	4
1.3 The effects of water on NIRS	17
2. Objectives.....	20
2.1 Identified levers for building models robust to water effects	20
2.2 Research questions, contributions and outline of the thesis	26
Chapter II. Materials & Methods: an experimental set-up for collecting spectral variations related to moisture content variations	29
1. Introduction	29
2. Materials & Methods.....	29
3. Opportunities	35

Table of Contents

Chapter III. Limitations of global correction approaches for moisture content correction	37
1. Introduction	37
2. Materials & Methods	38
3. Results & Discussion	41
4. Concluding remarks	44
Chapter IV. Unveiling the non-linearity of moisture content effects in organic waste using principal components analysis (PCA)	45
1. Introduction	45
2. Materials & Methods	45
3. Results & Discussion	46
4. Concluding remarks	52
Chapter V. Modeling the influence of water content on scattering: towards a new Bouguer-Beer-Lambert law for wet scattering samples	53
1. Introduction	53
2. Results	53
3. Concluding remarks	59
Chapter VI. A knowledge-based local approach to account for water effects in NIRS	61
1. Introduction	61
2. Building predictive models on a local subset of substrates with homogeneous moisture content effects	61
2.1 Defining the local subset	61
2.2 Fitting the local model	63
3. Leveraging the model of moisture content effects on scattering to define classes of substrates with homogeneous effects	66
3.1 Introduction	66
3.2 Materials & Methods	66

Table of Contents

3.3 Results & Discussion.....	68
4. Concluding remarks	71
Chapter VII. General concluding remarks	73
1. Summary of results.....	73
2. Topics for future research.....	74
Included papers	77
Paper I – Limitations of global correction approaches for moisture content correction in the context of organic waste characterization.....	78
Paper II – Unveiling non-linear water effects in near infrared spectroscopy: A study on organic wastes during drying using chemometrics.....	79
Paper III – Relating Near-Infrared Light Path-Length Modifications to the Water Content of Scattering Media in Near-Infrared Spectroscopy: Toward a New Bouguer – Beer – Lambert Law	80
Paper IV – Fast at-line characterization of solid organic waste : Comparing analytical performance of different compact near infrared spectroscopic systems with different measurement configurations	81
Paper V – 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	82
All references	83

List of Figures

Figure 1. The anaerobic digestion process monitoring. Process monitoring (modeling and control) are calibrated based on both feedstock characteristics and process state variables.....	3
Figure 2. Modes of measurements in NIRS with the corresponding light trajectories. I_{incident} represents the incident light source beam. In transmittance mode ($I_{\text{transmittance}}$), the direct and diffused transmitted photons are collected. In reflectance mode ($I_{\text{reflectance}}$), both the direct (specular) and diffused reflected photons are collected. In transreflectance mode ($I_{\text{transreflectance}}$), both the direct and diffused reflected photons are collected but a light reflector allows light to travel double more than in simple reflectance mode. In interactance mode ($I_{\text{interactance}}$), a light barrier allows to collect mainly diffused reflected photons.	5
Figure 3. The Bouguer-Beer-Lambert law set-up.	5
Figure 4. From a substrate to its predicted composition: the analytical pathway. For each steps of the analytical methodology, the state of the sample is provided, and the risks in terms of representativity errors are provided.....	16
Figure 5. Strategies for robustifying a NIRS system.....	20
Figure 6. Experimental set-up for the collection of NIR spectral variations related to moisture content variations. The samples are put into a rotating quartz cup over an FT-NIR spectrometer (Buchi N-500) which automatically collects reflectance spectra (1000-2500 nm). The sample cup is plugged to a closed tube loop connected to a strong dessicant (sodium hydroxide) which captures the water present in the gas phase. A peristaltic pump generates an internal circulation of the air, leading to a progressive drying of the substrate. To evaluate the loss of water during drying (i.e., therefore the moisture content MC of the sample), the desiccant is weighed continuously using a precision balance.....	30
Figure 7. The sample characteristics. Histograms of the reference values for the dried samples.	31
Figure 8. The range of moisture content covered for all substrates.....	32
Figure 9. The raw pseudo-absorbance spectra colored by moisture content (blue for high MC, red for low MC). The spectral variations are presented for nine substrates representative of the diversity of biochemical compositions and physical properties (poultry manure, ramial chipped wood / rwc, fish, cooked steak, sugar, syrup, sour cream, aluminum and plastic bag).....	34
Figure 10. The preprocessed spectra (Savitzky-Golay second derivation with polynomial order of three and window length of 21) colored by moisture content (blue for high MC, red for low MC). The spectral variations are presented for nine substrates representative of the diversity of biochemical compositions and physical properties (poultry manure, ramial chipped wood / rwc, fish, cooked steak, sugar, syrup, sour cream, aluminum and plastic bag).....	35

List of Figures

Figure 11. Histograms of BMP values for the train and test set (respectively in blue and orange).	39
Figure 12. Flowchart for the comparison of moisture content correction strategies.	40
Figure 13. Predicted vs. observed values for the ten correction methods. The dry calibration train set is plot in blue, the dry calibration test set is plot in orange, the nuisance train set is plot in green, the nuisance test set is plot in red. The statistics (RMSE, MAE, R^2 , r^2 , Bias) are provided for the dry calibration test set (p) and the wet nuisance test set (pn).	42
Figure 14. The evolution of loadings (left subplot) and the evolutions of cumulative explained variance and the Durbin-Watson (DW) criterion, according to the principal component number (PC#). The blue curve is the DW criterion and the green curve is the cumulative explained variance.	47
Figure 15. Visual aspect of substrates with low and high levels of pseudo-absorbance. Pictures were taken from above the NIRS measurement cell.	48
Figure 16. Loadings of the first eight components. Explained variance percentage of each principal component is given in the title. For each component, the corresponding eigenvector of the between-substrate variance-covariance matrix (see Eq. 8 in Paper I) is plot (in dashed black line), as well as the corresponding eigenvector of the within-substrate variance-covariance matrix (Eq. 9 in Paper I) (in dashed red line).	50
Figure 17. Evolutions of the first eight components' scores along drying for nine representative substrates. Other substrates are colored in grey.	51
Figure 18. Evolutions with water content % of (A) raw absorbance values, (B) corrected absorbance values, and (C) log-transformed corrected absorbance values are provided. The latter log-transformed corrected absorbance values are plot with log-transformed water content %, In the (C) subplot, the OLS regression line is plot in red, with the slope, intercept and coefficient of determination (R^2).	56
Figure 19. The exponential of the fitted intercept values ($\epsilon\lambda \cdot \mathbf{10}$) in eq.(9) (in green) and the corresponding Savitzky-Golay second derivative (in orange).	57
Figure 20. RMSEC curves obtained for four different models: $X \sim y$ (in blue), $\text{EMSC_additive}(X) \sim y$ (in orange), $\text{EMSC_complete}(X) \sim y$ (in green), and $\log(\text{EMSC_additive}(X)) \sim \log(y)$ (in red).	58
Figure 21. Scores kinetics for the first principal component of the identified subset with the fitted power laws in red dotted lines (left subplot). Boxplot of the R^2 obtained for each fitted power law.....	62
Figure 22. Predictions vs observed values for models (PLS-R or updated PLS-R) built on the group with homogeneous moisture content effects, with four moisture content ranges considered: $\text{MC}<99\%$ (first line), $\text{MC}<60\%$ (second line), $\text{MC}<40\%$ (third line), $\text{MC}<20\%$ (fourth line). The dry calibration train set is plot in blue, the dry calibration test set is plot in orange, the nuisance train set is plot in green, the nuisance test set is plot in red. The statistics (RMSE, MAE, R^2 , r^2 , Bias) are provided for the dry calibration test set (p) and the wet nuisance test set (pn).	64

List of Figures

Figure 23. MCR-ALS results for the drying of ligno-cellulosic substrate (wheat chaff). The top subplot corresponds to the evolution of the reconstruction error (MSE) over the 40 iterations of MCR-ALS. The middle subplots corresponds to the concentration profiles (C_{optimal}) and the pure spectra (S_{optimal}) of the optimal solution found by MCR-ALS. The lower subplots corresponds to the observed raw spectra (X_{observed}) and the reconstructed spectra from MCR-ALS decomposition ($X_{\text{reconstructed}}$). The red box presents the different errors of reconstruction (MSE, Frobenius Norm Ratio, Lack of Fit % (LOF%), and Explained Variance %).	69
Figure 24. MCR-ALS results for the drying of substrate with soluble carbohydrates (apple). The top subplot corresponds to the evolution of the reconstruction error (MSE) over the 40 iterations of MCR-ALS. The middle subplots corresponds to the concentration profiles (C_{optimal}) and the pure spectra (S_{optimal}) of the optimal solution found by MCR-ALS. The lower subplots corresponds to the observed raw spectra (X_{observed}) and the reconstructed spectra from MCR-ALS decomposition ($X_{\text{reconstructed}}$). The red box presents the different errors of reconstruction (MSE, Frobenius Norm Ratio, Lack of Fit % (LOF%), and Explained Variance %).	70

List of Tables

Table 1. A review of NIRS-based models for the characterization of solid organic waste.....	13
Table 2. Methodology for building robust models in regards to external factor influences.....	25
Table 3. Quality of models corrected from moisture content. The RMSE, MAE and R^2 are provided for cross-validation (CV), calibration (C), prediction (P) (calibration data), prediction (PN) (nuisance data).	43

List of symbols and abbreviations

2D-COS: Two-Dimensional Correlation Spectroscopy	DWT: Discrete Wavelet Transform	MCR-ALS: Multivariate Curve Resolution Alternating Least Squares	R²: coefficient of determination
AcoD: Anaerobic co-Digestion	EBA: European Biogas Association	MIR: Mid-infrared	r²: Squared Pearson's correlation coefficient
AD: Anaerobic Digestion	EMSC: Extended Multiplicative Scatter Correction	MIRS: Mid-infrared spectroscopy	RANSAC: RANdom SAMple Consensus
ADF: Acid Detergent Fiber	EPO: External Parameter Orthogonalization	MSC: Multiplicative Scatter Correction	REPFILe: Repeatability file
ADM1: Anaerobic Digestion Model N ^o 1	EROS: Error Removal by Orthogonal Substraction	MSE: Mean Squared Error	RMSE: Root Mean Squared Error
ALS: Asymmetric Least Squares	FT: Fourier Transform	NDF: Neutral Detergent Fiber	RNV: Robust Normal Variate
BBL: Bouguer-Beer-Lambert	GLSW: Generalized Least Squares Weighing	NIR: Near infrared	SG: Savitzky-Golay
BMP: Biochemical Methane Potential	ICA: Independent Component Analysis	NIRS: Near infrared spectroscopy	SOC: Soil Organic Carbon
CCS: Carbon Capture & Storage	IIR: Independent Interference Reduction	NNSC: Nearest Neighbor Spectral Correction	SNV: Standard Normal Variate
COD: Chemical Oxygen Demand	IPCC: Intergovernmental Panel on Climate Change	NSR: Normalized Spectral Ratio	SPORT: Sequential Pre-processing through ORThogonalization
CR: Continuum Removal	knn: K Nearest Neighbours	OSC: Orthogonal Signal Correction	SVD: Singular Value Decomposition
CSTR: Continuously Stirred Tank Reactor	LCFA: Long Chain Fatty Acid	P2G: Power-to-Gas	TOP: Transfer by Orthogonal Projection
CV: Cross-Validation	LV: Latent Variable	PCA: Principal Components Analysis	TS: Total solids
CWT: Continuous Wavelet Transform	LW: Locally Weighted	PDS: Piecewise Direct Standardization	VFA: Volatile Fatty Acids
DO: Direct Orthogonalization	MAE: Mean Absolute Error	PLFA: Phospholipid Fatty Acid	VM: Volatile matter
DOP: Dynamic Orthogonal Projection	MC: Moisture content (%)	PLS-R: Partial Least Squares Regression	VS: Volatile solids
DM: Dry matter		PQN: Probabilistic Quotient Normalization	VSN: Variable Sorting for Normalization
DT: Detrend			WLS: Weighted Least Squares
DW: Durbin-Watson			WWTP: Wastewater Treatment Plant

Chapter I. Introduction

1. Motivation

1.1 Valorization of organic waste through anaerobic digestion (AD)

1.1.1 AD processes as part of a global energy production decarbonization strategy

The last synthesis report consolidated by the Intergovernmental Panel on Climate Change (IPCC) confirms the unequivocal warming of the global climate system. It points out the recent anthropogenic emissions of greenhouse gases that have led to atmospheric concentrations of carbon dioxide, methane and nitrous oxide to unprecedented values in the last 800,000 years.¹ Among other, the decarbonization of energy supply is a key response option for risk mitigation. Indeed, in the low-concentration stabilization scenarios, the share of low-carbon electricity supply (including renewable energy, nuclear energy and bio-energy) is expected to increase from the current share of 30% to 80% by 2050. While nuclear energy could appear today as the most decarbonized solution, the high safety risks and the high costs of radioactive waste management should make renewable energy and bio-energy part of the future energy supply. In renewable electrical energy (coming from solar panels, wind turbines, and hydrolic turbines), the energy storage is of high importance to balance the intermittent supply and demand. Today, both battery storage and the power-to-gas (P2G) concept are envisioned.² The latter solution consists in producing hydrogen gas using electricity through the electrolysis of water; which can then be integrated in the natural gas grid directly as hydrogen gas (with some limitations in terms of concentration) or with a prior transformation in methane using carbon dioxide through the methanation process³. With this vision of gas as the future building block of energy storage, the development of low-carbon gas production technologies is foreseen. Such processes include power-to-gas (P2G), carbon capture & storage (CCS), gasification, and the most mature in terms of industrialization, the anaerobic digestion (AD).⁴

1.1.2 The AD process: principles

The anaerobic digestion^{5,6} consists in a biological organic matter degradation process, in oxygen-free conditions, that produces both biogas and digestate. The biogas is primarily composed of methane (CH₄) (from 50% to 75%) but also contains carbon dioxide (CO₂), hydrogen sulfide (H₂S) and water vapor. The biogas can be used to generate electricity and heat through combined heat and power (CHP) units,

Chapter I

but can also be purified using membrane technologies⁷ to generate natural gas that can be directly injected into the natural gas grid, compressed for vehicle fuel production⁸ or be used as chemical feedstock in the manufacture of plastics⁹ or other biochemicals. The digestate refers to the solid and liquid degraded materials which come out of the process. Under certain conditions, this co-product can be used as a soil amendment and fertilizer, or animal bedding.¹⁰ The biological process involves four main different steps (hydrolysis, acidogenesis, acetogenesis, and methanogenesis), each involving different microbial communities.^{11,12}

In terms of process design, though high-solid systems exist (with an influent substrate concentration of 25% of total solids (TS)), the most conventional process design is the continuously stirred tank reactor (CSTR) where an influent substrate concentration of 1-15% total solids (TS) is added daily in a closed tank. Factors to be considered in the design of anaerobic digestion plants include the nature of digester, temperature, pH value, composition of feeding substrate, organic loading rate, retention time, mixing, waste particle size, carbon/nitrogen ratio (C/N), costs, moisture content.¹³

A variety of organic materials can be integrated into the anaerobic digestion process: agricultural waste (animal manure, crop residues, energy crops), wastewater biosolids (municipal sewage sludge), food waste (household, restaurant, grocery, food production)^{14,15}, or agro-industrial waste (spent brewery grain, fat/oil/grease¹⁶, crude glycerol).

1.1.3 A clear trend towards anaerobic co-digestion (AcoD) processes calls for the development of fast and robust feedstock characterization methods

According to the European Biogas Association (EBA)¹⁷, at the end of 2019, Europe was the largest biogas producer, with 18,943 anaerobic digestion plants in operation for a production of 167 TWh, and it is projected to increase to up to 467 TWh by 2030. In terms of biomethane production, France is leading the market development with over 1,000 biomethane injection projects. According to the latest report¹⁸, a clear trend can be observed in the feedstock usage starting 2013: energy crops are progressively abandoned towards the use of agricultural residues, municipal waste and sewage sludge. While mono-substrate digestion plants focusing on one type of organic waste is still common, especially in Germany with energy crops (such as maize)^{19,20}, a growing number of AD installations mix two or more substrates together, a process known as anaerobic co-digestion (AcoD)^{21,22}. Such co-digestion offers numerous advantages including a greater stability to the process^{23,24}, with greater obtained methane yields ranging from 13% to 176%²⁵. However, AcoD comes with new challenges in terms of process monitoring and optimization. These include questions on how to optimize the feeding substrate

Chapter I

recipe since random decisions often lead to process disturbances such as inhibitions by ammonia²⁶, or acidification by volatile fatty acids (VFA). The choice of the feeding substrates proportions is mostly reasoned by the biochemical methane potential (BMP), but must also take into account factors such as the carbon/nitrogen ratio (C/N)^{27,28} to foster development of microbial communities²⁹, the macro- and micro-nutrients²², the pH and alkalinity^{30,31}, the inhibitors and toxic compounds³², and the biodegradable organic and dry matter^{33,34}. In order to design the best feeding strategies and optimize AcoD plants, approaches involving the combination of feedstock characterization^{35,36} with process modeling³⁷⁻³⁹ have been proposed^{40,41} (Figure 1). By knowing the feedstock characteristics, and being able to simulate the AD process empirical performances, the choice of recipe can be optimized.

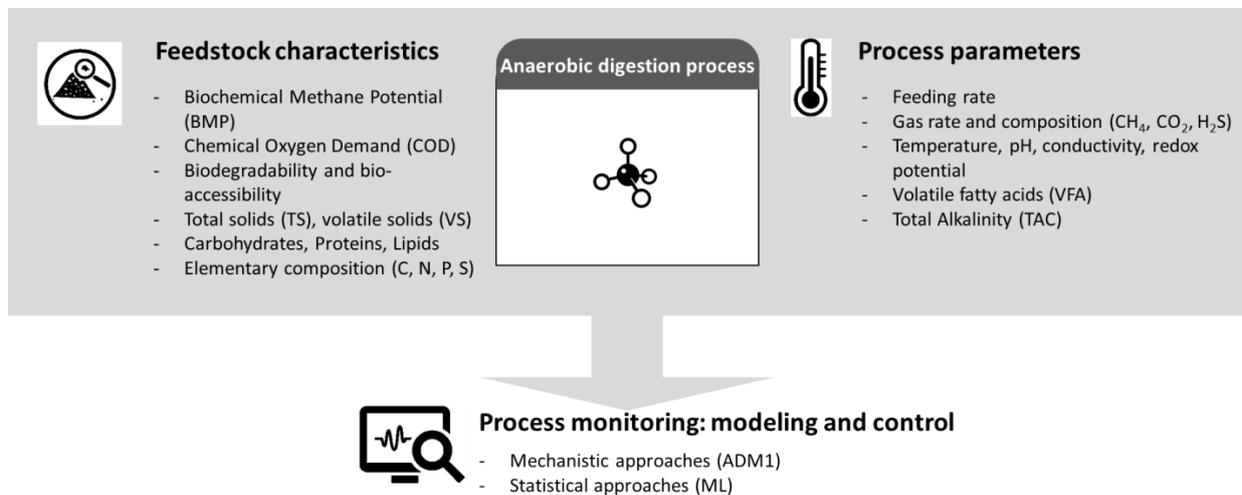


Figure 1. The anaerobic digestion process monitoring. Process monitoring (modeling and control) are calibrated based on both feedstock characteristics and process state variables.

Feedstock characteristics can include the determination of the biochemical methane potential (BMP)³⁶, the chemical oxygen demand (COD), the biodegradability and bio-accessibility⁴², the total solids (TS) and volatile solids (VS), the biochemical composition (carbohydrates content, proteins content and lipids content), but also elementary composition (carbon (C), nitrogen (N), phosphore (P) and sulfur (S)). Not only these feedstock characteristics vary according to the substrate type, but these properties may also vary along the year according to factors such as crop seasonality, transport or storage. Therefore, the main condition to being able to run a substrate feed control of anaerobic digestion plants is the development of fast and reliable characterization methods that are applicable on highly diverse organic waste⁴³ and that can be deployed on-site for high-frequency measurements. Amongst the wide range of instrumentation that can benefit to the monitoring of anaerobic digestion process⁴⁴, near infrared spectroscopy has appeared as the most promising technology for developing such robust and low-maintenance online measurement systems.

1.2 Near infrared spectroscopy (NIRS) and chemometrics for bioprocess monitoring

1.2.1 Introduction to NIRS and chemometrics

1.2.1.1 NIRS and its founding principles

Near infrared spectroscopy (NIRS) is a vibrational spectroscopy technique based on the interaction of near-infrared light with matter. While the existence of NIR light was reported by William Herschel in 1800, its use for analytical applications was first explored by Karl Norris in the 1960s for the determination of moisture content in wheat flour, or fat in meat; then by Thomas Hirschfeld who developed more theoretical aspects on NIRS, and Philip Williams who used NIRS for protein determination in wheat to replace the conventional Kjeldahl method^{45,46}.

The principle of vibrational spectroscopy is that the different vibrational modes of a molecule (ways in which atoms vibrate together with the same frequency) may absorb light at specific wavelengths (corresponding to different energy levels). By quantifying this absorption, quantification of analytes in the material is therefore possible. In the NIR region, which is situated between 750 to 2,500 nm (13,300 cm^{-1} to 4,000 cm^{-1}), the photons energy is higher than the energy required to promote molecules to their fundamental vibrational transition mode (like in the mid infrared region), but lower than the energy that allows electron excitation in molecules (like in the visible, and ultraviolet region). The NIR region hosts the overtones and combination bands, which are in a first approximation, respectively multiples and sum/differences of fundamental vibration frequencies (mostly of O-H, C-H, N-H, and S-H bonds)⁴⁷.

In practice, NIRS consists in sending light photons at different wavelengths onto a sample, and measuring the transmitted or reflected light intensity in order to evaluate absorption intensity levels (Figure 2). The various modes of measurements allow the collection of different types of photons, which may lead to very different spectral information⁴⁸. In the NIR region, the absorption intensities of overtones and combinations bands are between 10 to 1000 times lower than the absorption intensities of the corresponding fundamental bands (present in mid-infrared (MIR) region)⁴⁹. While this results in highly convoluted spectra with broad features, this also allows light to go deeper into the materials than with mid-infrared (MIR) light.

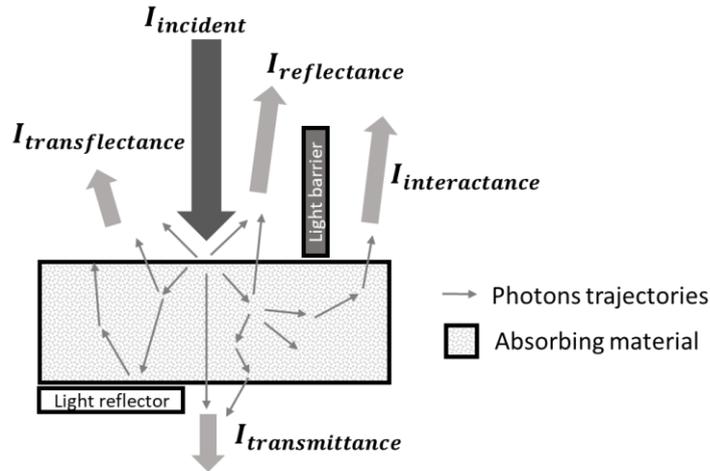


Figure 2. Modes of measurements in NIRS with the corresponding light trajectories. $I_{incident}$ represents the incident light source beam. In transmittance mode ($I_{transmittance}$), the direct and diffused transmitted photons are collected. In reflectance mode ($I_{reflectance}$), both the direct (specular) and diffused reflected photons are collected. In transflectance mode ($I_{transflectance}$), both the direct and diffused reflected photons are collected but a light reflector allows light to travel double more than in simple reflectance mode. In interactance mode ($I_{interactance}$), a light barrier allows to collect mainly diffused reflected photons.

The way light absorption is related to the analyte concentration has been first analyzed for simple systems that are homogeneous, isotropic and do not contain particles: this is known as the Bouguer-Beer-Lambert (BBL) law⁵⁰. By measuring such system in transmission through a distance L , the absorbance $A(\lambda)$ is linearly related to the concentration c of the attenuating specie with a specific extinction coefficient $\epsilon(\lambda)$ (Eq. 1) (Figure 3).

$$A(\lambda) = -\log(T(\lambda)) = -\log(I_1(\lambda)/I_0(\lambda)) = \epsilon(\lambda) \cdot L \cdot c \quad \text{(Eq. 1)}$$

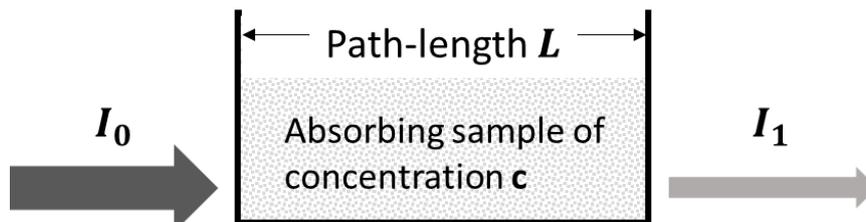


Figure 3. The Bouguer-Beer-Lambert law set-up.

It has been suggested that this law should be named “ideal absorption law”⁵¹ (in analogy with the “ideal gas law”), highlighting the fact that this law holds in very restrictive conditions. Indeed, in most systems where NIRS is applied, the BBL law does not strictly hold. For example, when measuring in reflectance mode (Figure 2), a pseudo-absorbance $A(\lambda) = -\log(R(\lambda))$ is usually calculated, and supposed to be linear with concentration, but it is known that this reflectance signal $R(\lambda)$ does not account exclusively for the pure chemical absorption but depends strongly on the scattering within the material. Indeed, from

the initial light that is sent on the material, part of it is directly reflected to the detector (the specular reflected light), part of it refracts multiple times before coming back to the detector (diffused reflected light), part of it may be transmitted and never come back to the detector (direct and diffused transmitted light), and of course, part of it may be absorbed (absorbed light). While the BBL law does not strictly hold, chemometrics allowed to leverage these measurements and develop quantitative calibrations using sound multivariate statistics.

1.2.1.2 Chemometrics, or how to leverage multivariate spectroscopic data

Linear algebra for multi-dimensional and collinear data

As defined by Svante Wold, one of its founding member, chemometrics is “how to get chemically relevant information out of measured chemical data, how to represent and display this information, and how to get such information into data”⁵¹. In fact, such measured chemical data is in most cases complex and thus multivariate. As mentioned, the NIR region hosts the absorption related to overtones and combinations, which leads to measured spectra with very broad peaks, and with individual wavelength absorptions of poor chemical specificity and sensitivity. Therefore, in order to develop accurate quantitative calibrations, individual absorption variables from the whole spectral range are used in combination to build (latent) variables that are highly sensitive and specific to the predicted characteristic/phenomena. In order to build these combinations of highly collinear variables, multivariate statistics (linear algebra) are used and are at the basis of chemometrics science⁵².

Dimension reduction and matrix decomposition, the building modeling block of chemometrics

In chemometrics, the spectral data is represented as a matrix $\mathbf{X}_{(n,p)}$, with \mathbf{n} rows representing the individual spectra (observations) and \mathbf{p} columns representing each wavelength absorptions (variables). Similarly, the reference data (characteristic to predict, group label) is represented as a matrix $\mathbf{Y}_{(n,q)}$ with \mathbf{n} rows representing the individual observations and \mathbf{q} columns representing each characteristic to be predicted. For quantitative purposes, models that use \mathbf{X} to predict \mathbf{Y} are built, which is known as the multivariate calibration process⁵³. In most techniques, the aim is to decompose the $\mathbf{X}_{(n,p)}$ matrix into a product of matrices with \mathbf{k} components (called principal components, latent variables, pure components) which can be interpreted:

$$\mathbf{X}_{(n,p)} = \mathbf{T}_{(n,k)} \cdot \mathbf{P}^T_{(k,p)} + \mathbf{E}_{(n,p)}, \quad (\text{Eq. 2})$$

Chapter I

with $\mathbf{T}_{(n,k)}$ matrix of scores/concentrations (projections of each spectra on these components), $\mathbf{P}_{(p,k)}$ matrix of loadings/weights/pure spectra (contributions of original variables to the component), and $\mathbf{E}_{(n,p)}$ the residuals.

What changes in the different methods is the way and the purpose for which these matrix decompositions are found. In a principal components analysis (PCA)⁵⁴, for exploratory data analysis, the principal components correspond to orthogonal variables that explain most of the variance in \mathbf{X} . In a partial-least square regression (PLS-R)⁵⁵, the aim is to build a predictive model based on \mathbf{X} to predict \mathbf{Y} variables: the latent variables are found so that they explain the most variance of \mathbf{X} while maximizing the correlation with the \mathbf{Y} matrix to be predicted. In multivariate curve resolution alternating least squares (MCR-ALS)⁵⁶, the aim is to find physically meaningful components (pure spectra and concentrations) that explain the most variance of \mathbf{X} using a variety of constraints (non-negativity, closure, unimodality, selectivity, correlation, asymmetry, kinetic modeling, smoothing, multi-set, etc.). In independent component analysis (ICA), for signal unmixing purposes, the components are found so that the loadings/signals have non-Gaussian distributions, assuming independence of these signals⁵⁷. What is core to all these methods used in chemometrics is that a big place is given to the interpretability of the found components.

Non-linear methods for heterogeneous and complex data

While in most cases, linear approaches are sufficient in order to deal with spectroscopic data, the current explosion of spectroscopic measurements has led to the emergence of highly complex and heterogeneous datasets. This brings non-linearities in the dataset including both clusters (groups of spectra with very different characteristics) and curvatures (spectra which do not show a linear relationship with the variable of interest \mathbf{Y}). For such cases, non-linear methods have been proposed.

Non-linear methods: local strategies

A first category of non-linear methods corresponds to local strategies, where models are built based on only a subset of the dataset. This subset can be chosen based on spectral characteristics, or using expert knowledge and metadata. In local methods based on spectra, there is the clustered PLS approach where a clustering method (k-means⁵⁸, hierarchical clustering⁵⁹, decision tree⁶⁰) is used to identify groups on which to train PLS models. When predicting a new observation, it is assigned to its cluster based on a spectral distance (Euclidean or Mahalanobis), and the corresponding cluster's model is used for prediction. In strictly speaking local approaches, such as the k-nearest neighbours locally-weighted PLS (knn-LW-PLSR)⁶¹, the subset is selected on-the-fly based on a spectral distance and the model based on

Chapter I

this subset is built on-the-fly. These local approaches are very suitable to account for non-linearities in data. However, the implementation of such methods can bring some challenges:

- Expensive computation time. In knn-LW-PLSR, the hyperparameter tuning (number of neighbors and number of PLS latent variables) can be demanding in terms of calculations as there are as many models to build as there are predictions to make. But these limits can be overcome by optimizing the PLS model calculation⁶² (using the kernel trick), calculating the spectral distance on global PLS scores⁶³, or optimizing the spectral distance calculation step using random projections and indexing⁶⁴.
- The nature of the spectral distance, or the criteria on which the clusters are made is of utmost importance and can be a difficult choice⁶⁵. The Euclidean distance in high-dimensional data such as spectral data may be insufficient for finding the nearest neighbor; in fact, it can provide instead the farthest neighbors in high-dimensional data⁶⁶. In NIR, this is partly resolved by the Euclidean distance on the scores of a PCA or PLS (Mahalanobis distance), with a more limited number of dimensions (usually between 5 and 15). However, in NIR data of complex matter such as waste, soil, food and crops, most of the information in the spectra comes from scattering and not absorption. Possibly, the local subset is found based on the physical properties of the samples instead of their chemical properties, which may lead to irrelevant localities. One way to overcome this is to use pre-processing steps to remove the variations related to scattering and therefore emphasize the variations related to chemical absorption⁶⁷.

Non-linear methods from the machine learning and deep learning communities

A second category of non-linear methods suitable for highly complex and heterogeneous datasets comes from the machine learning and deep learning communities. These include regression trees^{68,69} and random forests (RF)⁷⁰⁻⁷², support vector machines (SVM)^{73,74}, and of course neural networks (NN)⁷⁵ including convolutional structures (CNN)^{76,77}. All these methods have been shown to be highly flexible and suitable for modeling non-linear relationships, but appear difficult to train and prone to overfitting, especially with small datasets (<100 samples) which are usually obtained in NIRS due to expensive reference measurements.

Pre-processing to account for scattering effects in NIR spectra of complex matter

As already mentioned, in complex matter (such as crops, food, waste, soil), the measured NIR spectra are seldom only related to chemical absorption and usually contain information related to physical scattering. This results from the interaction of different factors such as the measurement mode (the way light is sent, filtered and collected) and the physical properties of the measured material (state of matter,

granulometry, structure). In order to enhance the sensitivity of the measured signal to the chemical absorption information and thus develop more predictive quantitative calibrations, a series of methods have been developed to transform \mathbf{X} , a step referred as pre-processing^{78,79}. A first set of pre-processing step includes simple transformations such as centering and scaling⁸⁰, which can be both column-wise or row-wise as in the standard normal variate (SNV)⁸¹. Other transformations concentrate on baseline removal such as detrend (DT)⁸¹ which allows the removal of polynomial baselines of different orders, continuum removal (CR)⁸² which removes the convex hull baseline, or asymmetric least squares (ALS)^{83,84} which identifies more complex baselines using the Whittaker smoother and regularization⁸⁵. Other pre-processing techniques include derivation to deconvolute the signal such as Savitzky-Golay derivation (SG)⁸⁶ or discrete wavelet transform (DWT)^{87,88}.

Other techniques are based on a physical model of scattering effects: the hypothesis is that scattering effects are made of both multiplicative effects \mathbf{a} (due to the modification of light path-length L) and additive effects \mathbf{b} (due to the photons loss: photons which are transmitted and never come back to the detector) (Eq. 3).

$$\mathbf{A}_{observed}(\lambda) = \varepsilon(\lambda) \cdot (\mathbf{a} \cdot L) \cdot \mathbf{c} + \mathbf{b} \quad (\text{Eq. 3})$$

These physical-model-based techniques attempt to identify these additive and multiplicative coefficients (\mathbf{a} and \mathbf{b}) to correct the spectra, such as optical path-length estimation and correction (OPLEC)^{89,90} or multiplicative scatter correction (MSC)^{91–93} and its extended versions (EMSC)^{94–96}. Different extensions of MSC were proposed. A first approach consists in adding the modeling of quadratic terms in addition to the MSC's linear terms (referred to as the basic EMSC model^{95,97}). This approach successfully corrected for light scattering differences in MIRS reflectance measurements of soil⁹⁸, and in NIRS transmittance measurements of powder mixtures⁹⁷. In particular, it allowed to account for light scattering differences of sorghum grain samples due to weathering damages⁹⁹. This has been extended with higher-order polynomials than quadratic terms. For example, up to sixth order polynomial EMSC models have been used in Raman spectroscopy to correct baselines due to fluorescence phenomena^{94,100,101}. In the same idea, a logarithmic wavelength-dependent term has been proposed¹⁰². A second approach relates to the addition of a constituent spectra other than the reference spectrum. This approach was successfully used to remove the influence of water on FT-IR data of meat products¹⁰³: in the EMSC model, the difference spectra of meat product taken by FT-IR (MIRS) at low and high humidity levels was added, and successfully reduced the variations in the O-H stretching region between 3500 cm^{-1} and 3400 cm^{-1} , and at the O-H bending vibration around 1640 cm^{-1} . A broader extension of EMSC consists in adding a nuisance or constituent orthogonal subspace model. For example, between-replicates variations have

Chapter I

been reduced using this method. A principal components analysis is applied on mean-centered replicate spectra and a few components are selected and introduced into the EMSC model. Though this approach has originally been developed to reduce the between-replicate variation in FT-IR spectra of microorganisms¹⁰⁴, it can be used to model any unwanted variations. Applications include blood glucose NIRS measurements¹⁰⁵, modeling of Mie scattering for synchrotron based microscopy¹⁰⁶, or Mie scattering for heterogeneous samples with cylindrical domains¹⁰⁷. Recently, several improvements have been made to the Mie extinction EMSC¹⁰⁸ and open-source codes have been made available¹⁰⁹. A slight variant of this extension consists in having multiple reference spectra instead^{110,111}. The only constraint regarding EMSC is that full rank of the model needs to be respected for the least squares estimation of the EMSC parameters⁹⁴. In the framework with multiple reference spectra, QR-factorization is used for this purpose.

A great advantage of the EMSC preprocessing is the fact that it is a model-based preprocessing method which implies that its parameters can be interpreted physically and chemically. This was first demonstrated for *in vivo* monitoring of a biological cell culture, where the dynamics of correction factors were interpreted and provided new insights on physical and chemical properties of the system. Later, these correction factors have been used on NIRS spectra for the prediction of ash content in woodchips and pellet samples respectively for the energy and feed sector. These factors proved to be as much predictive as EMSC corrected spectra, therefore showing how much prediction of ash content depended on physical properties (light scattering) and to a lesser extent on chemical information¹¹².

To avoid wavelength regions where the chemical absorbance variations might influence the estimation of the parameters, a weighted least-squares (WLS) estimation has been proposed⁹⁵. Authors have proposed different methodologies to determine these weights, including the simple inverse of the mean spectra or, in a more sophisticated way, using RANSAC regression as in VSN¹¹³. In the same objective, a window-based implementation has been developed (PMSC¹¹⁴). Less elegant but simpler solutions have been proposed, where the “interference dominant region” (*i.e.*, the spectral region where additive and multiplicative effects prevail) is found by applying an MSC or SNV based on different spectral regions and selecting the region with best predictive performance¹¹⁵.

Another important aspect of EMSC framework is that the spectra should be colinear to the reference spectrum. Such issue can appear when shifting phenomena appear, for which a shift correction methodology has been proposed¹¹⁶. Another solution to treat this problem is to apply MSC multiple times until stabilization of the reference spectrum (Loopy MSC¹¹⁷). As mentioned before, in the case of very different spectra, a framework with multiple reference spectra has been proposed⁹⁴.

Other preprocessing techniques make use of decomposition techniques, potentially followed by filtering of certain unwanted components. This is the case in continuous wavelet transform (CWT)^{118,119} where a multi-scale decomposition is done, followed by a potential threshold filtering to remove certain features of given frequency. This filtering can be done quite similarly with components found by PCA^{120,121} or ICA^{122,123}. Let \mathbf{Q} represent a signal (a continuous wavelet, a loading from a PCA, an independent component from ICA, a pure spectrum from MCR-ALS) representing the unwanted information, the removal of variations related to this signal can be done by simply projecting \mathbf{X} orthogonally to this signal (Eq. 4):

$$\mathbf{X}_{corrected} = \mathbf{X} \cdot (\mathbf{I} - \mathbf{Q} \cdot (\mathbf{Q}^T \mathbf{Q})^{-1} \mathbf{Q}^T) \quad (\text{Eq. 4})$$

However, the risk of removing such components is that these components rarely only contain unwanted variations, which may lead to a loss of information.

Another noteworthy approach that is developed further (section 2.1.3) is to use external data that contains the unwanted variations in order to remove these from \mathbf{X} . These approaches may downweigh the variables (also referred as pre-whitening or shrinkage) using generalized least square weighing (GLSW)¹²⁴ or remove the variations through orthogonal projections like in external parameter orthogonalization (EPO)¹²⁵, transfer by orthogonal projections (TOP)¹²⁶ or dynamic orthogonal projection (DOP)¹²⁷. These techniques use external data (for example an experimental design where spectra are collected at different levels of an influencing factor) to run the pre-processing. These have been successfully applied on soils to deal with moisture effects¹²⁸ or sucrose solutions to deal with temperature effects¹²⁹.

Finally, all the pre-processing techniques may be combined together in a multi-block structure as in sequential pre-processing through orthogonalization (SPORT)¹³⁰⁻¹³² or the parallel pre-processing through orthogonalization (PORTO)¹³² in order to leverage the complementary information provided by all the different techniques. These two approaches are compared on pharmaceutical tablets and both show improved prediction results¹³³.

While the choice of pre-processing technique should ideally be supported by theoretical grounds, in most cases, it is usually more pragmatically based on the model performances obtained after using cross-validation for tuning the pre-processing technique hyperparameters. However, when the objective is to interpret the models and the spectral regions which are into play, the way pre-processing affects the signal needs to be carefully accounted for in order to make the correct spectral assignments¹³⁴.

Although pre-processing appears as an important step in analyzing NIRS, this may also lead to the degradation of models due to a loss of information¹³⁵. Some authors have argued that preprocessing is only here to compensate for small sample size in NIRS datasets¹³⁶. Moreover, in some cases, the scattering information is highly correlated to the chemical information of interest¹¹². For example, for dry matter content estimation, the baseline variations related to scattering have been found in fact very informative¹³⁷.

1.2.2 Application of NIRS to AD processes: where are we heading?

1.2.2.1 NIRS for the fast characterization of feeding substrates

The applications of NIRS to the monitoring of AD processes can be separated in two complementary approaches¹³⁸: a NIR spectroscopic characterization of the process itself through the analysis of the digestate, or a NIR spectroscopic characterization of the input feedstock (the substrates) of the process.

The use of NIRS to monitor the AD process was first demonstrated in an experiment with a reactor fed with a mixture of cellulose, albumin and minerals and exposed to an overload of glucose¹³⁹. NIR spectra collected *in-situ* were shown adequate for the prediction of different constituents such as acetate (RMSEP=0.20 g.L⁻¹, R²=0.88), propionate (RMSEP=0.21 g.L⁻¹, R²=0.93), glucose (RMSEP=0.6 g.L⁻¹, R²=0.99) and phospholipid fatty acids (PLFAs)(RMSEP=9 nmol.mL⁻¹, R²=0.98). A variety of studies^{138,140} further reported the use of NIRS to predict process variables including volatile fatty acids (VFA)¹⁴¹⁻¹⁴⁷, total and volatile solids (TS and VS)^{142-144,147}, nitrogen content (ammonia NH₃ and ammonium NH₄⁺)^{142-144,146}, alkalinity or total organic/inorganic carbon (TOC, TIC)^{144,145,147}. Some developments aiming at optimizing the prediction performance include the use of different measurement modes (transflection¹⁴² or polarization¹⁴⁶), or new modeling approaches (multi-block¹⁴⁸). It appears from these studies that NIRS is a suitable technique for providing information on process state and potential instabilities. However, the quality of the prediction models (RMSE and R²) appears to vary significantly depending on the type of fed materials (municipal solid waste, maize silage, solid or liquid manure), the range of TS (*i.e.*, below 5% or between 5% to 10%), and the range of the predicted parameter (*i.e.*, VFA from 0.1-10 g.L⁻¹ or 0.5-20 g.L⁻¹). Therefore, the robustness of these systems still needs to be assessed¹⁴⁴ in AcoD processes where digestates are likely to cover a wider range of physical and biochemical states.

As mentioned, another more indirect, or at least more upstream, monitoring approach of AD process consists in the use of NIRS for the characterization of the input feedstock. Such usage was largely inspired from the livestock community who used NIRS to assess the quality of plant/forage/feed tissue and predict diet quality from the corresponding feces¹⁴⁹. In AD process, the biochemical methane

Chapter I

potential (BMP) is a crucial characteristic of a substrate: it corresponds to the volume of methane CH_4 gas that could be released in non-limiting AD conditions (usually expressed in $\text{mL}(\text{CH}_4).\text{gVS}^{-1}$). Depending on the substrate¹⁵⁰, the reference measurement may last between 20 to 50 days as it consists in measuring the gas production during the substrate's degradation¹⁵¹. Such cumbersome and expensive BMP assay led to the development of fast alternative methods¹⁵² including the use of NIRS-based models to assess BMP¹⁵³ (see Table 1). Today, this allows to measure a BMP in less than 5 days (including transport, sample preparation and analysis). Though the benefits of using NIRS is clearer for BMP due to the time-consuming reference measurement, the use of NIRS has been extended to the prediction of biochemical composition components such as carbohydrates content, lipid content, nitrogen content, COD¹⁵⁴, as well as methane fermentation kinetics^{41,155,156}. Indeed, the biochemical composition has been shown to be a good indicator of the biodegradability and the methane potential of organic waste^{157,158}.

Table 1. A review of NIRS-based models for the characterization of solid organic waste.

Type of sample	Sample preparation	Predicted characteristics	Modeling approach	Reference
Municipal solid waste (MSW), green waste	Freeze-drying + grinding (<2mm)	BMP	PLS	M. Lesteur et al., 2011 ¹⁵⁹
Fermented corn	Oven-drying (55°C) + grinding (<1mm)	BMP, kinetics	PLS	C. Grieder et al., 2011 ¹⁵⁵
Meadow grasses	Oven-drying (60°C/24h) + grinding (<0.8mm)	BMP	PLS	C.S. Raju et al., 2011 ¹⁶⁰
Energy crops	Raw	VS, COD	PLS	D. Wolf et al., 2011 ¹⁶¹
Maize silage	Raw	BMP	PLS	H.F. Jacobi et al., 2012 ¹⁶²
Agro-industrial waste, MSW, Green waste, Wastewater Sewage sludge, Biowaste	Oven-drying (40°C/48h) or Freeze-dried + grinding (<1mm)	BMP	PLS	J. Doublet et al., 2013 ¹⁶³
Reed canary grass	Oven-drying (?°C) + grinding (<1mm)	BMP, kinetics	PLS	T.P. Kandel et al., 2013 ¹⁵⁶
Maize silage	Oven-drying (60°C/72h) + grinding (<1mm) Raw	BMP	PLS	F. Mayer et al., 2013 ¹⁶⁴
Plant biomasses	Oven-drying (60°C) + grinding (<1mm)	BMP	PLS, PLS (on selection)	J.M. Triolo et al., 2014 ¹⁶⁵

Chapter I

Grass, cereal residues (raw or silage)	Oven-drying (60°C/72h) + grinding (<1mm) Raw	BMP	PLS, Local PLS	B. Godin et al., 2015 ¹⁶⁶
Crop residues, Grass	Oven-drying (60°C/48h) + grinding (<0.8mm)	BMP, Protein, Lignin, Cellulose, hemicellulose, NDF, ADF, carbon	PLS	R. Wahid et al., 2015 ¹⁶⁷
Agro-industrial waste, MSW, Green waste, Wastewater Sewage sludge, Biowaste	Freeze-drying + grinding (<1mm)	BMP	PLS	S. Preys et al., 2015 ¹⁶⁸
Urban organic waste, Plant biomasses	Freeze-drying + grinding (<1mm)	BMP	PLS, PLS (on selection)	T. Fitamo et al., 2017 ¹⁶⁹
Agro-industrial waste, MSW, Green waste, Wastewater Sewage sludge, Biowaste	Freeze-drying + grinding (<1mm)	Kinetics	PLS	C. Charnier et al., 2017 ⁴¹
Agro-industrial waste, MSW, Green waste, Wastewater Sewage sludge, Biowaste	Freeze-drying + grinding (<1mm)	Carbohydrates, Lipids, Nitrogen content, COD	PLS	C. Charnier et al., 2017 ¹⁵⁴
Agricultural residues, animal waste, MSW, agro-industrial waste	Oven-drying (60°C) + grinding (<1mm)	BMP	PLS, PLS (on selection)	P. Mortreuil et al., 2018 ¹⁷⁰
Corn stover, livestock manure	Oven-drying (60°C) + grinding (<0.85mm)	BMP	PLS, feature selection	G. Yang et al., 2021 ¹⁷¹
Corn stover, livestock manure	Oven-drying (60°C) + grinding (<0.85mm)	Cellulose, Hemi-cellulose, Lignin	PLS, SVM	J. Liu et al., 2021 ¹⁷²
Agro-industrial waste, MSW, Green waste, Wastewater Sewage sludge, Biowaste	Freeze-drying + grinding (<1mm)	BMP, Carbohydrates, Nitrogen, Lipids, COD	PLS	A. Mallet et al., 2021 ¹⁷³

In most cases, the samples are dried and ground to avoid the effects of water and particle size on NIRS (see section 1.3.1). Such preparation steps also reduce the high physical and chemical heterogeneity of most raw organic waste samples. The drying step is done in two ways: a freeze-drying process (freezing of the sample, lowering of pressure, then sublimation of ice into water vapor) or simple dehydration process (using oven-drying at about 60°C for about 48 hours). Furthermore, the grinding is mostly done on the dried sample (with the mesh size usually below 1 mm).

Chapter I

Some studies show consistent BMP prediction results using NIR spectra taken directly on raw samples: using only corn silage¹⁶² or only plant biomass silage¹⁶⁶. These studies offer great promises for the online supervision of input feedstock in mono-digestion plants¹⁷⁴. However, these models appear to be built on substrates with limited biochemical composition ranges (only corn silage or only plant biomass silage), while the advantage of supervising the feeding substrates has been shown to be interesting mostly when the variability of biochemical composition is high¹⁶².

It must be noted that other fast alternative methodologies to determine organic waste composition have also been developed using mid-infrared spectroscopy (MIRS)¹⁷⁵ and in particular, for BMP determination¹⁷⁶, using a photoacoustic measurement mode (FTIR-PAS) which allows measurements on dark and opaque samples¹⁷⁷. The performances appear to be similar to NIRS, but MIRS suffers from lower spectral sampling capabilities which require cumbersome sample preparation steps (penetration of MIR light in organic materials is at the micron scale while the NIR light is at the millimeter/centimeter scale).

1.2.2.2 Current limitations and opportunities

Limitations of sample preparation to avoid water effects

Today, NIRS-based characterization of feeding substrates appears as the most accurate and cost-effective methodology¹⁷⁸. However, the adoption of NIRS for on-line or at-site industrial applications is still very limited¹⁷⁹. Different reasons may account for this low adoption. A first general reason is that NIRS involves a complex modeling and calibration process, with a regular maintenance of both the instrument and the calibrated model. But most importantly, as mentioned above, monitoring the feeding substrates appears most valuable when there is a high variability in the substrate type and quality¹⁶². Therefore, the system should be applicable on a wide variety of substrates. Currently, this has been shown possible, but still involves cumbersome, but necessary, sample preparation steps (drying and grinding) which makes the methodology unsuitable for on-line and at-site uses. Therefore, the samples are sent to the laboratory for analysis, which involves costs in logistics and precludes any real-time analysis. Another constraint related to the drying step is the fact that during this process, there is a loss of volatile compounds¹⁸⁰ such as ammonia (NH₃) or VFA, which may account for a high proportion of the methane potential (*e.g.*, up to 30% of COD is lost in the case of spent apples)¹⁸¹. In order to compensate for this loss in some of the substrates, complementary reference analyses (COD, total nitrogen content) made on the raw substrate are sometimes required. Another constraint related to the current way substrates are analyzed (sent to a laboratory for analysis) concerns the potential errors that may arise from sampling and the lack of representativity (Figure 4), as described by the Theory of

Sampling (TOS)¹⁸². In fact, while it is important to reduce the analytical error itself (the NIRS system), the greatest variability of samples is often found in the first steps of the process, and may amount to 100-1000 times this analytical error if not well controlled¹⁸².

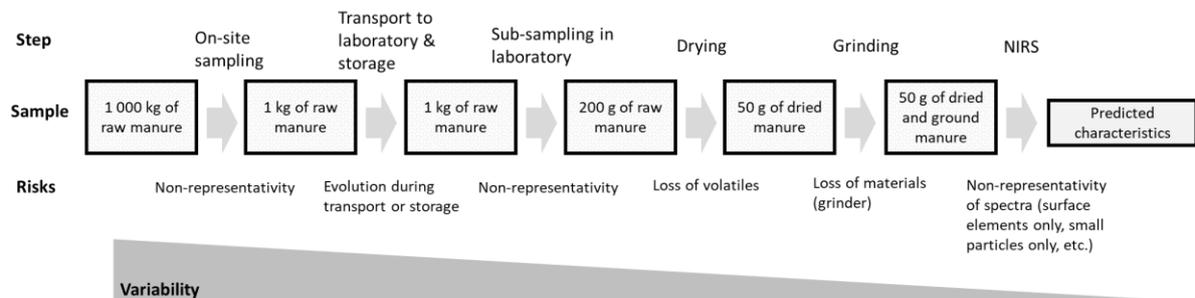


Figure 4. From a substrate to its predicted composition: the analytical pathway. For each steps of the analytical methodology, the state of the sample is provided, and the risks in terms of representativity errors are provided.

Reducing the preparation steps and bringing the NIRS analysis directly online should certainly reduce the sampling steps and therefore mitigate the risk of obtaining biased and non-representative results due to unrepresentative sampling.

It appears from this analysis, that while NIRS already provides tremendous advantages for the optimization of AD plants, there is a great need for the development of NIRS systems that enable an at-site or online characterization of highly diverse organic waste types with cost-effective preparation steps.

Opportunities provided by portable and low-cost spectrometers: an at-site and online use

In the past few years, the use of NIRS has developed out of laboratories, thanks to important progress in the miniaturization of instruments¹⁸³⁻¹⁸⁵. In particular, handheld Fourier transform near infrared (FT-NIR) micro spectrometers have appeared in the market, and make use of a micro-electro-mechanical systems-based (MEMS) Michelson interferometer¹⁸⁶. While conventional Michelson interferometers are made of discrete elements (including the moving mirror actioned by a motor, the fixed mirror, and beam splitter), MEMS technology enables a monolithic integration of these elements on a single chip, with the particularity that the moving mirror is operated by an electrical signal. Amongst the spectrometers making use of this technology, the NeoSpectra instrument has shown good analytical performance results for soil organic and total carbon content characterization^{187,188}, or authenticity screening in food¹⁸⁹. These compact spectrometers allow the measurements to be performed on site, thanks to their compactness, robustness and cost. However, these compact portable spectrometers tend to have poorer instrument performances than laboratory spectrometer, with lower resolution, spectral range, and signal-to-noise ratio^{184,185}. Therefore, the suitability of such systems for the characterization of diverse organic waste still needs to be assessed.

Chapter I

Investigating the performances of low-cost, portable spectrometers, and the use of different measurement configurations including polarization, has been the object of an experiment presented in Paper IV – Fast at-line characterization of solid organic waste : Comparing analytical performance of different compact near infrared spectroscopic systems with different measurement configurations (published in Waste Management journal). In addition, the collected data has been published in an open access data paper published in Data in Brief journal (Paper V – 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). Refer to the last section ‘Included Papers’ for the published versions and the full references. Results from this study demonstrated the suitability of low-cost systems applied to dry organic waste samples. This makes one foresee even more an at-site and online use of NIRS on wet organic waste samples and further highlights the crucial need of finding ways to account for water effects.

1.3 The effects of water on NIRS

1.3.1 Moisture content effects on NIR spectra

The effects of moisture content on near infrared spectra have been described for a wide variety of different matter types including soil^{190–196}, crops^{197–201}, food²⁰², plants²⁰³, wood²⁰⁴, pharmaceuticals²⁰⁵, object models^{206–208}, and water-dominant systems²⁰⁹. In addition, though not focused on the analysis of moisture content effects in NIRS, some studies used NIRS to monitor drying or hydration processes where moisture content varies^{210,211}.

Three broad OH absorbance bands

The main water effect on NIRS that is described relates to the apparition of the three broad OH absorbance bands of water molecules (centered at about 1210 nm, 1450 nm and 1940 nm). These are attributed respectively to the combination of the first overtone of the O-H stretching and O-H bending band, the first overtone of the O-H stretching band and the combination of the O-H stretching band and O-H bending band of water^{209,212}.

Modifications due to changes in physical properties

Other important effects described relate to changes of physical properties: pH, ionic strength, and differences in physical state (crystallinity/amorphicity) were all shown to change the observed spectra.²⁰⁸ For example, while the spectra of dry crystalline solids (of single compounds such as sugars, amino acids) show sharp peaks, the wet amorphous solutions of these single compounds show broader peaks

like in polysaccharides such as cellulose and starch²¹³. In addition to the modification of peak shapes, changes in moisture content have been shown to induce baseline modifications.¹⁹¹

Modifications of hydrogen bonding and hydration state

Finally, changes in moisture content modify the hydrogen bonding in the system, and have been shown to provoke peak shifts in the observed spectra.^{214–216}

1.3.2 Water, its model and current questions related to NIRS

Although apparently a simple molecule (H_2O), water shows complex behaviors and presents many physical anomalies compared to other liquids^{217,218}. This is mostly explained by its dominating inter-molecular hydrogen-bonding. However, as surprising as it may seem, the structure model of water is far from having reached a broad consensus²¹⁹. In the past years, NIRS has been used to reveal details of the water structure and its functionalities in aqueous solutions, but very different models and interpretations of what constitutes the first overtone OH absorbance region in NIRS have been proposed. Using second derivative, up to 6 underlying water species were detected, and said to correspond to different water species : protonated water (Sr), and water with no, one, two, three, and four hydrogen bonds (S0-S4)²²⁰. Using non-linear fitting procedures, these authors fit six Gaussian peaks to reconstruct the observed spectra. This was later confirmed by other authors studying water-glucose solutions²²¹. Though good reconstruction errors were obtained, these non-linear fitting procedures may suffer from user-guided hypothesis and initialization biases. In another study, using multivariate curve-resolution (MCR-ALS), three components were found best to explain water spectra at different temperatures²²². Furthermore, it was shown that though water with salt showed the same trends, the position of these three components depended on the ionic strength of salts. The fact that the components were not forced to be Gaussian in this study could explain why a lower number of components was obtained. A two-state model (water species with weaker and stronger hydrogen bonds) has also been outlined using second derivatives, two-dimensional correlation spectroscopy (2D-COS), and principal component analysis (PCA) applied on temperature-dependent NIRS²²³. Recently, this model was supported by new temperature-dependent data that suggests the presence of a coherent state²²⁴. In this experiment, an isosbestic point has been observed around 1438 nm which suggests the existence of an equilibrium between two populations/states of water, referred as the ‘mixture-model’²²⁵. Using second derivation, authors have found two negative subpeaks with stable positions under temperature variations (positioned at 1412 nm and at 1462 nm). However, the authors pointed out that the isosbestic point was not perfectly constant and could be dependent of temperature, as previously also observed by Gowen et al²²², which seems to

be still an unresolved question today: is there a third population of vibration²²³? An interaction term? Baseline artefacts? A continuous distribution of hydrogen bond geometries²²⁶?

A concept that has been developed in the recent years, called aquaphotomics²²⁷, deals with the water-light interaction over the whole electromagnetic spectrum. It has been found that the first overtone OH region consists in 12 to 14 different water absorbance bands, each related to a given water structure²⁰⁹. Various water structures have been experimentally identified in various systems containing water and later found to be in agreement with theoretical calculations, such as, water solvation shells at 1364-1384 nm, water molecules confined by ions with no hydrogen bonding at 1396-1403 nm²²⁸, free OH water with no hydrogen bonding like in water-vapor at 1403-1418 nm (S0), protein hydration shells at 1418-1430 nm, water molecules with 1, 2, 3 or 4 hydrogen bonds (S1-S4) respectively at 1432-1454 nm, 1458-1468 nm, 1472-1482 nm, 1482-1495 nm, and strongly bound water at 1506-1516 nm. Shao et al.⁸⁸ on the other hand, in the same region found 10 spectral components (Gaussian peaks) corresponding to 9 different water molecular structures, using knowledge-based genetic algorithm.

It appears from all these studies that the NIR spectra of wet materials hold information on the water state (which could allow to differentiate between substrates). However, the correct assignment of these bands to water states implies that the peaks truly correspond to absorbing species and are not due to scattering artefacts as introduced in Section 1.3.1. This makes it difficult to consolidate water state specific assignments in scattering materials. At least, the effects of scattering need to be taken into account before being able to do it.

1.3.3 Moisture content effects on NIRS quantitative calibrations

The various effects of moisture content on the NIR spectra that have been described result in a deleterious effect of moisture content variations on the quality of quantitative calibrations if not accounted for. For example, the accuracy of soil organic carbon (SOC) content NIRS-based predictions was evaluated with an R^2 of 0.82 and RMSE of 1.65 g.100g⁻¹ on dried samples (MC%<5%), while it declined to an R^2 of 0.52 and RMSE of 7.75 g.100g⁻¹ on wet samples (MC%=18%)¹²⁸. For sorghum grain protein content NIRS-based predictions, the accuracy shifted from an R^2 of 0.90 and an RMSE of 0.42 g.100g⁻¹ on dry samples (7.5%<MC%<12%) to an R^2 of 0.72 and RMSE of 0.87 g.100g⁻¹ on wet samples (7.5%<MC%<18%)²²⁹. This highlights the need to account for moisture content variations in the NIRS calibrations. However, even when models are built on wet samples, it appears the quality of the models remain low: for example, for silage analysis, the quality of models (evaluated by RMSE and

R²) built on intact fresh samples was shown not to match the quality of models built on dried samples^{166,230,231}.

This suggests there is a need to develop new methods that specifically take into account moisture content effects in order to build reliable quantitative calibrations which are robust to moisture content variations.

2. Objectives

2.1 Identified levers for building models robust to water effects

Any given analytical system has a scope of action (*i.e.*, conditions of functioning). While here the effect of moisture content will be studied, a variety of other influencing factors may influence the result of the analysis (material heterogeneity, temperature, humidity, operator, seasonality, instrument, etc.). In order to make the system robust towards this influencing factor (*i.e.*, the results provided by the analysis do not vary according to the influencing factor), many different methods can be employed as illustrated (Figure 5).

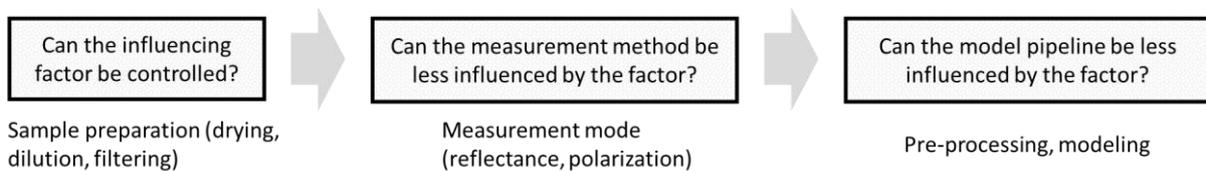


Figure 5. Strategies for robustifying a NIRS system.

2.1.1 Sample preparation

Drying steps

Certainly the most easiest yet very time-consuming strategy to reduce water effects on NIRS is to control the moisture content range under a certain level (0%-5%) through a drying process. As mentioned earlier, this can be done through freeze-drying or simple dehydration (oven-drying at 60°C for 48h). This way of proceeding is today the most common way NIRS is applied in laboratories on crops, soils, food, waste, etc. The freeze-drying process appears as most suitable for drying organic waste which are thermally sensitive and prone to oxidation (Maillard reactions) since it operates at low temperatures and under high vacuum. Though the drying process allows to reduce strongly the moisture content range, residual moisture remains and may still have a significant impact on the multivariate calibrations^{232,233}.

Grinding / Homogenizing / Filtering / Diluting / Centrifuging

The choice of sample preparation for NIRS analysis should not be targeted to reduce exclusively one influencing factor, as there may be strong interactions with other factors. For example, the moisture content effects will themselves be affected by particle size or temperature. This is why, while a drying step will by definition reduce the moisture content effects, a grinding step may also reduce the moisture content effects by uniformizing the particle size. Indeed, particle size effects (grinding at 1 mm or 0.5 mm) were evaluated on powders²³³, but also on wet substrates (homogenizing at 50 μm or 100 μm)²³⁴. Grinding complex and diverse physical matrices can be complicated, one way of doing is to use freeze-grinding (dry ice grinding), as previously tested on silage samples²³¹, but this process still appears cumbersome.

Another strategy may consist in reducing the moisture content range (but not necessarily to very low levels like in drying), through filtration. Filtration through a 1 mm mesh filter applied to livestock slurry and digestate samples was shown to enhance the prediction of TKN, TAN, and VFA, thanks to a reduction of scattering effects²³⁴. Similarly, meat samples were centrifugated and filtered before NIRS analysis to discriminate frozen and unfrozen meat samples²³⁵. Of course, for quantification purposes, this requires to measure the filtered/unfiltered quantities to express the final quantity on a raw basis, just like dry matter content is required when drying.

Dilution may also be used to reduce the moisture content range. For example, rewetting of dry soil samples was reported to enhance the SOC calibration in soil: one of the reason could be that it reduces the over-estimations of samples rich in sand²³⁶. The same results were obtained for clay content determination.²³⁷

Additives: drying salts

One way of “drying” a sample, is using drying solvents such as sulfate magnesium. These drying agents are used to purify alcoholic solutions for example. Unfortunately, it seems adapted to liquids only, as the subsequent separation of the dried organic solution from the drying salt is usually done by gravity filtration or decanting.

Additives: plating agents

One way to “structure” a sample is to use plating agents where water is encapsulated by nano-particles which gives the sample a powder aspect while being at high moisture content levels. This is the concept of dry water^{238,239}. This has never been studied using NIRS, but it is highly probable that this structuring allows to modify the scattering. This could transform liquid samples into new samples that have

comparable levels of scattering to solid powders. Again, for quantification purposes, the mixing proportions of the plating agent and the substrate would need to be evaluated and accounted for.

Fast drying techniques

In order to reduce the time-consuming freeze-drying step, fast drying systems have been proposed: DESIR system^{240–248} and SPOT²⁴⁴, with applications such as fungicide residues detection on intact lettuces²⁴⁹. It seems the biggest constraint for the system to work is that the sample should be very homogeneous, as there is little sampling of the matter since the quantity is of 0.5 mL to 1 mL. There is apparently also little reproducibility regarding the preparation and drying step. Moreover, the loss of volatiles pertains, as any drying step. Authors state the drying is fast enough for Maillard reactions not to take place; but nitrogen saturated environment could ensure this more. For all these reasons, the system seems very well adapted to liquid and homogeneous matrices such as milk, fruit juices, beer, wine, etc. For solid organic wastes, a centrifugation, followed by the characterization of the liquid part could be imagined, but, this means precisely quantifying solid/liquid proportions. Another possibility, is to make sure the sample is sufficiently homogenized, which is the purpose of SPOT system, where manual grinding of the sample is done in a mortar.

In general, finding a common preparation methodology applicable to the diversity of organic waste is difficult. The way a straw, a milk solution, a glycerol solution or a syrup should be grinded, or dried will be different. It seems the solution can be found by using a combination of techniques, but this inevitably makes the preparation cumbersome.

2.1.2 Measurement mode

Other development efforts to increase the accuracy and robustness of NIR systems have rather focused on enhancing the measured signal directly. A promising ‘optical pre-processing’ method, based on polarized light spectroscopy²⁵⁰ has been proposed to improve the absorbance signal measurement on such scattering samples^{251–253}. Such system has shown analytical performance improvements for soils²⁵⁴, and more recently, for digestate¹⁴⁶, but has never been evaluated on solid organic waste. Still in the aim of enhancing the measured spectra, time-resolved spectroscopy shows a promising future: applied to pharmaceutical tablets, collected photons with a particular propagation time were shown to be most informative for quantification^{255,256}. However, the cost of this technology still remains prohibitive for the organic waste management sector. Finally, the measurement mode (at distance or in contact, in reflectance or in interactance) also plays an important role in the final accuracy for estimating biochemical properties^{48,257,258}. Though current used laboratory spectrometers make use of a distance

reflectance measurement, a contact immersed probe measurement has been shown to be useful for prediction of parameters on digestates¹⁴⁶. Authors observe higher reflectance levels with less noise in the collected spectra, as well as new chemical features which were not apparent in a remote probe configuration. In light of this, it appears that the use of different measurement configurations could enable the calibration of more accurate and more robust NIRS models on diverse solid organic waste.

2.1.3 Building robust model pipelines

The last way of robustifying the NIRS-based analytical systems lies in the modeling pipeline itself. A summary of the methodologic levers allowing the robustification of a model pipeline in regards to external factor influences (moisture, temperature, seasonality, spectroscopic instrument, etc.) is provided in Table 2. This further consolidates the schema provided by Bellon-Maurel et al²⁵⁹. Robustification can happen at three levels: transforming the input spectra, transforming the model itself or transforming the output predictions. For each of these three levels, methods will differ depending on whether external data related to the external factor influence (noted as **G**) is used or not. For simplification, **G** can correspond to the level of the external factor (temperature level, moisture content, spectrometer type) and/or to the spectra which span the external factor influence range (spectra acquired at different temperatures, spectra acquired at different moisture content levels, spectra acquired using different spectrometers). Indeed, methods may use only the external factor, only the external spectra, or both together.

Without external (nuisance) data

When not using external data, the robustification strategy consists in applying a classical chemometrics approach. If all the steps (pre-processing, modeling) are well optimized, the found model should be made simple and robust. This approach has already been detailed above (section 1.2.1.2). However, in most cases, the dataset does not contain the full range of levels that the external factor may cover in practice. Therefore, without the use of external data, it is most probable that the obtained model will be affected by the influencing factor. Amongst the simple but noteworthy approaches which are directly targeted to moisture content correction, there is the principle of manually cutting the water spectral regions (1300-1600 nm and 1800-2000 nm) which are expected not to hold any useful information^{196,260}.

With external (nuisance) data

When using external (nuisance) data, it appears the most simple approach is to add the external nuisance data to the calibration model, a method referred as the exhaustive model, model augmentation or model update strategy^{229,261-263}. Within these approaches, some methods make use of the quantified levels

Chapter I

(clustering approach based on the nuisance levels, like moisture-explicit DS²⁶⁴), and others don't (clustering approach using a humidity index^{265–267}).

Most methods use a separate nuisance dataset with quantified levels of the interfering factor. Amongst these, the external orthogonalization parameter pre-processing (EPO)^{125,128,204,268–273} requires a targeted experimental design to identify the nuisance spectral subspace and then makes use of an orthogonal projection to this subspace to remove the related information. This method can be potentially combined with a prior OSC to remove in the nuisance data any variations related to the reference data \mathbf{y} ²⁷⁴. Other *a priori* methods try to model the transformation between the nuisance spectra and the reference spectra, like direct standardization (DS^{273,275,276} and its piecewise version PDS^{277,278}), latent variable modeling such as PLS²⁷⁹ or local standardization approaches such as nearest neighbor spectral correction (NNSC)²⁷⁹. These methods are interesting due to the fact that they provide the dry spectra that can be studied for other purposes, but tend to be very sensible to noise (DS, PDS), or depend on a similarity distance (NNSC) which has unpredictable or biased behaviors in high-dimensional data such as NIRS. In addition, more simple methods make use of humidity indexes to correct the spectra from moisture effect¹⁹³. Of course, *a posteriori* methods have also been proposed to correct the predictions using bias-slope correction^{261,264}, but this seems mostly suitable for simple interferences.

Still in the case of a separately acquired nuisance dataset, some methods can cope with unquantified levels of the interfering factor. Amongst these, as reviewed by some authors^{78,280}, there are orthogonal net analyte signal (NAS) pre-processing approaches (IIR²⁸¹, DO²⁸², EROS²⁸³, SBC²⁸⁴), generalized Tikhonov regularization (GTR) and augmentation approaches (repeatability file^{260,285–287}, counter-balanced distortions²⁸⁸) and generalized least squares (GLS) pre-processing approaches (GLSW^{124,289,290}). More recently, domain adaptation methods such as domain adaptive PLS (DA-PLS)²⁹¹ and domain invariant PLS (DI-PLS)²⁹² make use of regularization to make sure that the differences between the projected X and the projected nuisance X are low. Compensation of instrumental/seasonal and sensor temperature changes has been evaluated for NIR-based fruit quality assessment^{292,293}.

In all these approaches, a trade-off needs to be found between the complete removal of the interfering spectral subspace and the preservation of enough analyte information in the calibration dataset.

Chapter I

Table 2. Methodology for building robust models in regards to external factor influences.

Model pipeline components/levels	Use of external data (G)	
	No	Yes
Input spectra $\mathbf{Y}^* = \mathbf{X}^* \cdot \mathbf{b}$	$\mathbf{X}^* = \mathbf{f}(\mathbf{X})$ <ul style="list-style-type: none"> - Centering/scaling (SNV) - Derivation (SG, DWT) - Baseline correction (DT, ALS, CR, MSC, EMSC, OPLEC, VSN) - Feature engineering (CWT, SPORT) - Feature selection (cut water bands) 	$\mathbf{X}^* = \mathbf{f}(\mathbf{X}, \mathbf{G})$ <ul style="list-style-type: none"> - Orthogonal projections (EPO, TOP, DOP) - Transfer methods (PDS) - Add metadata (G) to X
Model $\mathbf{Y}^* = \mathbf{X} \cdot \mathbf{b}^*$	$\mathbf{b}^* = \mathbf{f}(\mathbf{X}, \mathbf{Y})$ <ul style="list-style-type: none"> - Feature selection (CovSel, VIP, iPLS) - Regularization (Tikhonov regularization, sparse methods) - Local approach (based on spectral distance) - Non-linear methods (RF, CNN, SVM) 	$\mathbf{b}^* = \mathbf{f}(\mathbf{X}, \mathbf{Y}, \mathbf{G})$ <ul style="list-style-type: none"> - Model update (exhaustive model, conjoint model, repfile) - OSC-EPO-PLS - Local approach (based on expert knowledge, or on G) - Regularization (domain invariant methods) and skewing (GLSW)
Output predictions $\mathbf{Y}^* = \mathbf{f}^*(\mathbf{X}, \mathbf{b})$	$\mathbf{Y}^* = \mathbf{f}(\mathbf{Y})$ <ul style="list-style-type: none"> - Deterministic transforms (log/power) 	$\mathbf{Y}^* = \mathbf{f}(\mathbf{Y}, \mathbf{G})$ <ul style="list-style-type: none"> - Bias-Slope correction - More complex functions (quadratic, non-linear)

2.1.4 Current progress and limitations in regards to correction of water effects

Several methods have been used and compared to deal with water effects on simple materials such as pharmaceuticals²⁰⁵ and complex materials such as crops^{260,286,287}, soil^{279,294–296}, and food^{129,202,297,298}. From these studies, it seems that EPO, GLSW and *a priori* PDS provide the best results for dealing with water's interfering effects. Of course, extending the calibration model to wider conditions of applicability usually leads to poorer prediction results of the spectra for samples at original conditions (dry samples). But, in many cases, the model performance on the original conditions has been improved as well (meaning it allowed a better estimation of the net analyte signal).

However, in most of the successful cases, the moisture content range for which water effects are corrected remains very limited (moisture contents from 1% to 20%); and most importantly, the type of matter is very homogeneous (similar physical structure and biochemical type). The most similar field related to applying NIRS on organic waste is soil spectroscopy, where highly heterogeneous physical matrices may be found, and difficulties have been shown for applying correction methods globally^{237,299}.

2.2 Research questions, contributions and outline of the thesis

2.2.1 Research questions

As presented, in the last ten years, NIRS has proven to be a fast and reliable characterization system of organic waste materials with a wide range of biochemical and physical types. This general applicability has made it the perfect tool for characterizing the feeding substrates in bioprocesses such as anaerobic digestion. However, to avoid moisture content effects on NIRS, a drying step is today necessary. Unfortunately, this sample preparation step comes with several drawbacks. Chemical changes need to be taken into account such as the disappearance of volatile compounds during drying. Moreover, the cumbersome drying step limits any on-site and on-line application.

As reviewed in Section 2.1, the most appealing strategy to avoid moisture content effects would rather consist in changing the way the models are built so that they become insensitive to moisture content variations. While some robustification methods have already been shown successful, it appears that these methods have been studied in restricted conditions with a limited range of moisture content levels and a limited range of biochemical composition type. This thesis work aims to tackle this challenge by **(i) developing a better understanding of the moisture content effects on NIRS applied to a wide range of organic materials**, and **(ii) finding new ways of building models that are robust to moisture content effects**. The underlying questions of each of these two topics can be further detailed as:

- 1) **Analysis of the complexity of moisture content effects**: How does water affect the spectra? What are the chemical and physical effects? How do these effects differ from one substrate to another? How do these effects depend on moisture content? Can these effects be modeled?
- 2) **Evaluation and development of modeling strategies to deal with these moisture content effects on highly heterogeneous databases** (organic waste): How do existing global correction methods deal with these effects? What strategy appears most suitable?

2.2.2 Contributions and outline of the thesis

The next chapters are structured around the major contributions of the thesis work. These contributions can be summarized as:

An experimental set-up for collecting the moisture content effects on NIRS on organic waste (presented in Chapter II)

A comprehensive experimentation was designed to study the moisture content on NIRS at ambient conditions. This involves the design of a system that can acquire NIR spectra of various organic waste during their drying while estimating the moisture content. A unique database was constituted during the thesis and includes substrates that cover a very wide range of biochemical composition and physical properties.

An evaluation of chemometrics global correction techniques (presented in Chapter III)

An evaluation of global robustification techniques was done in the aim of avoiding moisture content effects in existing NIR models built on dry samples. A wide range of techniques were evaluated.

A comprehensive analysis of water effects on NIRS (presented in Chapter IV)

This thesis allowed to draw a generic picture of water effects on NIRS, thanks to the study of complex organic matrices covering a wide range of biochemical composition (sugar, proteins, carbohydrates) and physical properties (physical state, crystallinity, granulometry). The non-linearity of water effects is demonstrated.

A new approach for modeling water effects on scattering media (presented in Chapter V)

Through the study of a simple system composed of water and aluminum, the Bouguer-Beer-Lambert theoretical framework was modified to take into account the effects of moisture content on scattering. This holds promises for the spectral studies of wet systems, including the prediction of dry matter content.

An advocacy for a knowledge-based local approach (presented in Chapter VI)

This thesis allowed to identify groups of substrates with common moisture content influences. These groups can serve as a basis to build local predictive models, a strategy that is proposed to account for the non-linearity in water effects.

Chapter II. Materials & Methods: an experimental set-up for collecting spectral variations related to moisture content variations

This chapter covers the Materials & Methods presented in the Paper II – Unveiling non-linear water effects in near infrared spectroscopy: A study on organic wastes during drying using chemometrics (published in Waste Management journal). See section “Included Papers” for the full reference and published article.

1. Introduction

In order to better understand the water effects on NIRS, a spectral dataset consisting of organic waste at various moisture content levels was constituted. For this, a customized air-drying system with dynamic NIR acquisition and moisture content estimation was designed. This chapter presents the system and the dataset that was constituted. The reader is invited to refer to Paper I for full details on the Materials & Methods.

2. Materials & Methods

Dehydration vs. rehydration

Most studies aiming at studying the moisture content effects on NIRS have operated by adding various amounts of water to a previously dried sample. On the opposite, here, we acquired spectral variations during drying experiments instead. Indeed, it is known that there is an hysteresis effect in organic matter: for example, the evolutions of water activity with dry matter content will be different during dehydrating (water desorption) or during hydrating (water absorption). In other words, drying a raw salad and rewetting it will probably not result in the same salad and therefore not the same measured NIR spectrum³⁰⁰.

Dynamic NIR spectra acquisition and moisture content determination during drying

For this purpose, a customized near infrared acquisition system was developed (Figure 6, detailed in Paper I), with dynamic highly-resolved simultaneous scanning of near infrared spectra and estimation of dry matter content during a drying process at ambient temperature. 89 organic waste substrates were measured using the system. Each drying experiment lasted between 12 hours and 72 hours depending on the substrate type. The substrates were chosen to represent a wide range of organic waste with

Chapter II

different chemical composition: fruits (banana, apple), vegetables (carrots, onions, salads, potato), farm residues (manure, silage, soya meal, grass), dairy products (cream, yoghurt, butter), meat products (beef, grilled/fresh meat, fish), food industry materials (sugar, sauces, fried potatoes, wheat flour) and various AD organic waste (water treatment sludge, digestate, dairy sludge). In order to provide control samples with simplified water effects due to limited water chemical interactions, a selection of packaging materials were also measured (wood, paper, aluminum, plastic). The obtained dataset consists of 116,000 spectra of 89 substrates covering a wide range of moisture content (5% to 95%). In parallel, all the substrates were freeze-dried and ground to 1 mm, and the biochemical composition was determined using a previously calibrated NIRS model¹⁵⁴.

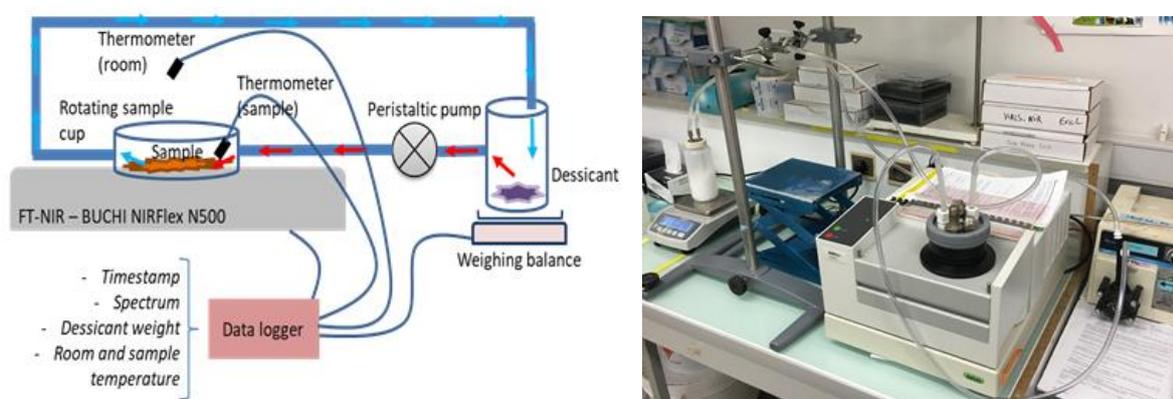


Figure 6. Experimental set-up for the collection of NIR spectral variations related to moisture content variations. The samples are put into a rotating quartz cup over an FT-NIR spectrometer (Buchi N-500) which automatically collects reflectance spectra (1000-2500 nm). The sample cup is plugged to a closed tube loop connected to a strong desiccant (sodium hydroxide) which captures the water present in the gas phase. A peristaltic pump generates an internal circulation of the air, leading to a progressive drying of the substrate. To evaluate the loss of water during drying (i.e., therefore the moisture content MC of the sample), the desiccant is weighed continuously using a precision balance.

A dataset of highly diverse biochemical types which complexifies data structure

Figure 7 presents the biochemical composition of the samples. All characteristics show non-Gaussian distributions, which illustrates the high diversity of biochemical types. This distribution impacts the structure of the data, and a special care must be taken when analyzing data. Indeed, for example, the fat content histogram clearly highlights two populations: one population with no or very low fat content levels ($<0.2 \text{ g.gTS}^{-1}$) and another population with very high fat content levels ($>0.7 \text{ g.gTS}^{-1}$). Unfortunately, such structuring is difficult to avoid, as intermediate compositions with 0.5 g.gTS^{-1} of fat content level results in biphasic systems.

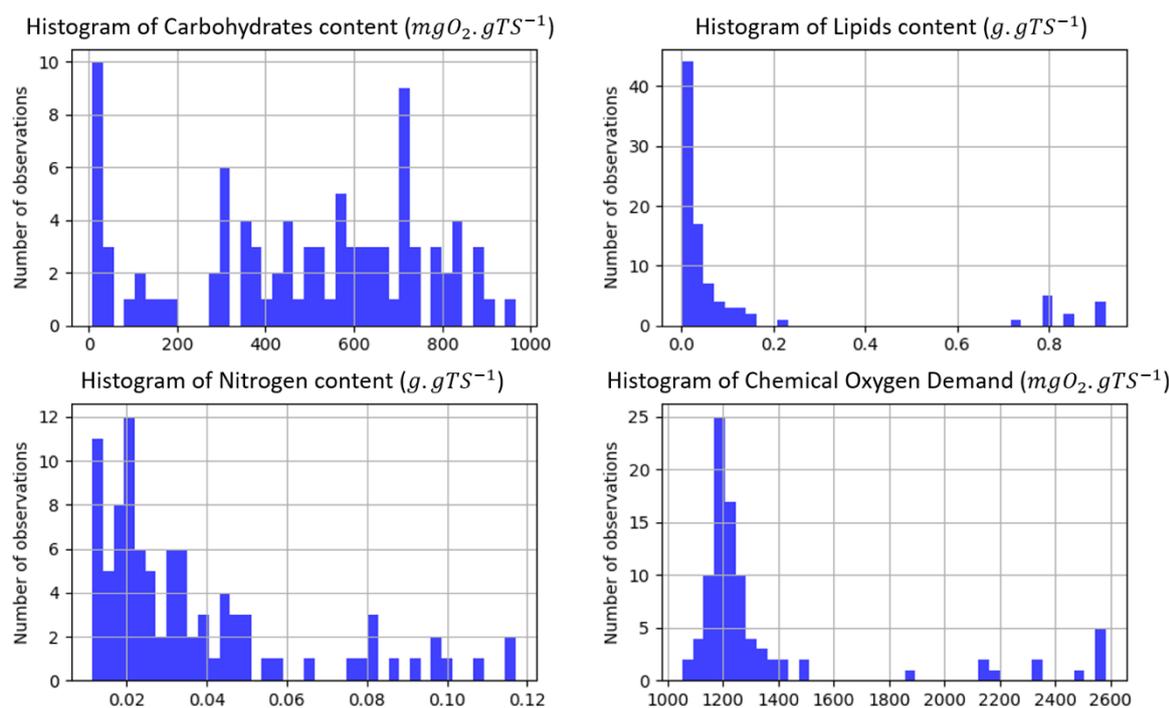


Figure 7. The sample characteristics. Histograms of the reference values for the dried samples.

A very wide range of moisture content

Figure 8 presents for each substrate the range of moisture content over which spectra were obtained. The accuracy of moisture content determination was evaluated (detailed in Paper I) and a mean over-estimation of +0.21% and a standard deviation of $\pm 2.30\%$ were obtained. This appears marginal compared to the large range of moisture content studied. However, this does imply that drawing conclusions on spectral effects due to water over 2-3% of moisture content differences should be done carefully. Contrarily to many studies that focused on limited moisture content ranges (1% to 25%), a very wide range of moisture content was obtained (5% to 95%). However, substrates were not all measured along the same moisture content range (Figure 8). Several reasons explain this including differences in the initial moisture content (very high moisture contents like *salad_1* or *digestate_1*, and very low moisture contents like *butter_2*, *mayonnaise_1*), drying inefficiency related to highly bound water or intra-cellular water (*syrup_1*, *ketchup_1*, *banana_2*, *orange pulp_1*) as well as simple experimental drying interruptions due mostly to electric failures (*banana_1*, *crustbread_1*, *sunflowermeal_1*, *grass_1*, *weeds_3*). These latter samples were still kept in the dataset because they still represented useful spectral variance related to moisture content variations. Two families of substrates can already be defined from these drying behaviors: hydrophobic substrates for which high moisture content levels are difficult to obtain (without having a biphasic sample) but are easily dried (like *butter*, *sour cream*, *mayonnaise*), and hydrophilic substrates in which water is more difficult to extract (like *syrup*, *ketchup*, *banana*, *orange pulp*). Within hydrophilic substrates, the final moisture content to which the substrate was dried relates to numerous factors and their complex interaction such

Chapter II

as the presence of gelling agents like pectin, or water soluble molecules like saccharides, as well as the interaction of proteins and starch controlling viscosity and swelling characteristics³⁰¹.

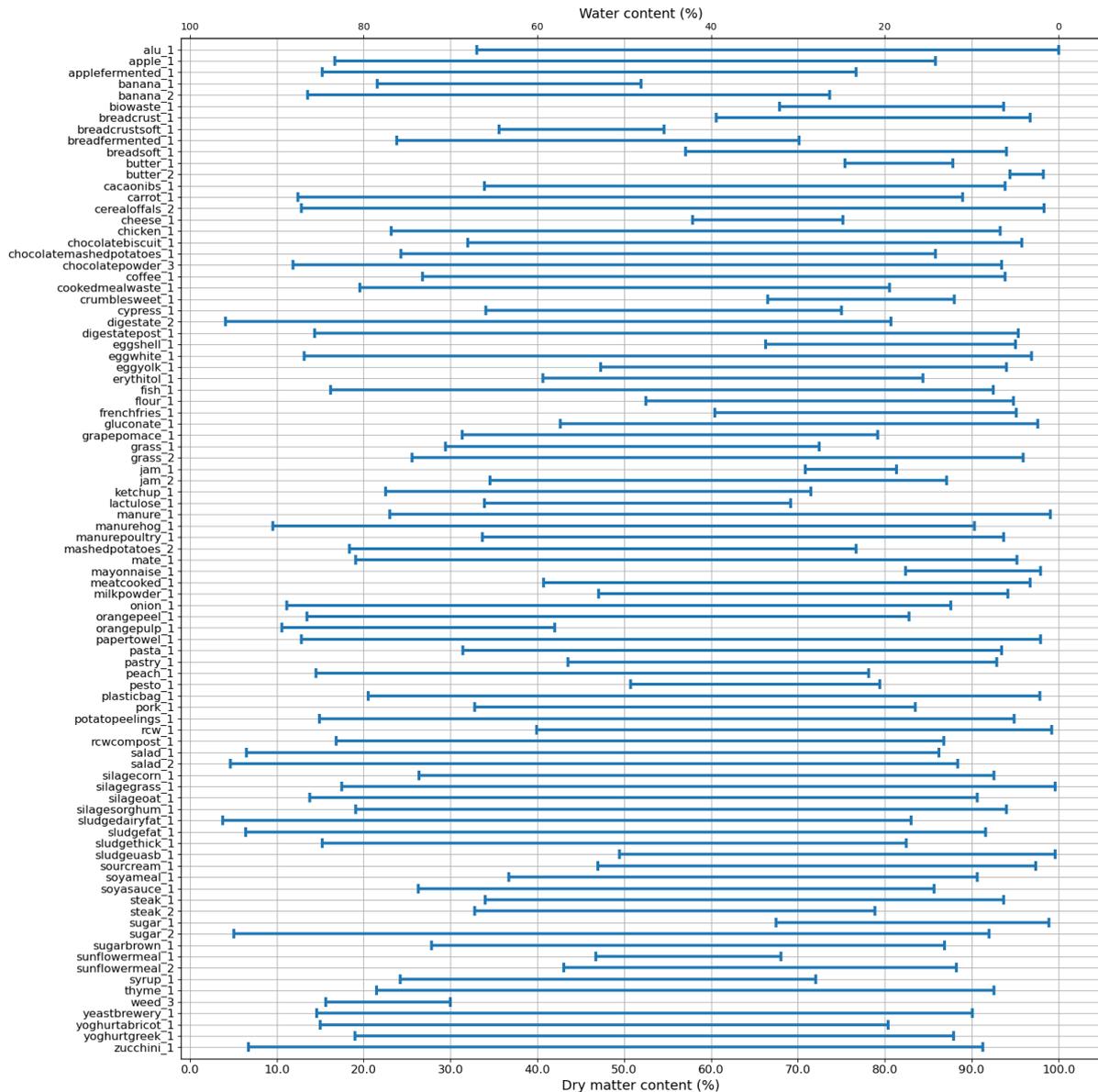


Figure 8. The range of moisture content covered for all substrates.

The temperature variations during drying and its effects

For each drying experiment, the temperature was monitored both inside the measurement cell and outside in the room. The mean temperature measured for all substrates is 28.3 °C, with a standard deviation of 1.8 °C. Such variations in temperature between each substrate drying experiment can be explained by the daily temperature differences from one experiment to another. Though the measurements were taken in a temperature controlled room, temperature differences were still observed. Moreover, the standard deviation of within-substrate temperature differences observed during drying is

1.15°C. Such sample temperature variations during drying can be explained by two factors: heating resulting from the spectrometer's lamp, and heating resulting from the absorption of water by the desiccant.

Unfortunately, variations of temperature may have a strong impact on the acquired spectra and may lead to the alteration of quantitative calibration models as many authors have shown^{125,261,263,297,302–304}. Indeed, as temperature rises, proportions of molecular vibrations within each molecular vibrational energy levels change, which has a direct impact on the absorption of photons (*i.e.*, the spectra). Visually, a horizontal shift of the broad absorbance bands can be observed in the spectra²²⁴, but in fact this relates to vertical absorption changes from the originating sub-bands. To have an idea of the magnitude of such changes, in the case of pure water at 22°C, it has been measured that at 1410 nm (free OH water peak), a +1°C temperature change increased the intensity of the absorbing peak by +0.8% (*i.e.*, temperature coefficient of 0.8% °C⁻¹)^{305,306}. However, as these authors highlighted, because scattering has little if no temperature dependence, the temperature coefficient applies exclusively on the absorption coefficient and not on the scattering coefficient. Though these changes could indeed alter the exact assignment of bands, they are very limited compared to the spectral variations induced by moisture content changes.

Water effects on NIR spectra: a combination of physical and chemical effects

Figure 9 shows some examples of near infrared spectral evolutions during drying, representative of the main types of evolution observed (spectral evolutions for all other substrates are provided in Appendix C of Paper I). These effects are also showed after a second derivation using Savitzky-Golay which deconvolutes the broad absorbance peaks (Figure 10). Different effects can be observed.

Firstly, water variation modifies strongly the global pseudo-absorbance level of the spectra: these baseline shifts probably relate to scattering modifications, as already reported^{91,92}. Interestingly, for suspensions, the pseudo-absorbance level increases with water content increase, while for the emulsions (cream, butter, oil), it decreases.

Secondly, for all substrates with intermediate and high moisture content levels (spectra in dark blue), well-known broad absorbance features due to OH vibrations are observed in the NIR spectra around 1210 nm, 1450 nm and 1940 nm. These are attributed respectively to the combination of the first overtone of the O-H stretching and O-H bending band, the first overtone of the O-H stretching band and the combination of the O-H stretching band and O-H bending band of water^{209,212}.

During the drying process (from blue to red), new absorbance features in relation with chemical composition progressively appear (related to OH vibrations of sugars or fatty acids, NH vibrations of

Chapter II

proteins, CH vibrations of alkanes, the C=C vibration of alkenes, and C=O vibrations of ketones/aldehydes).

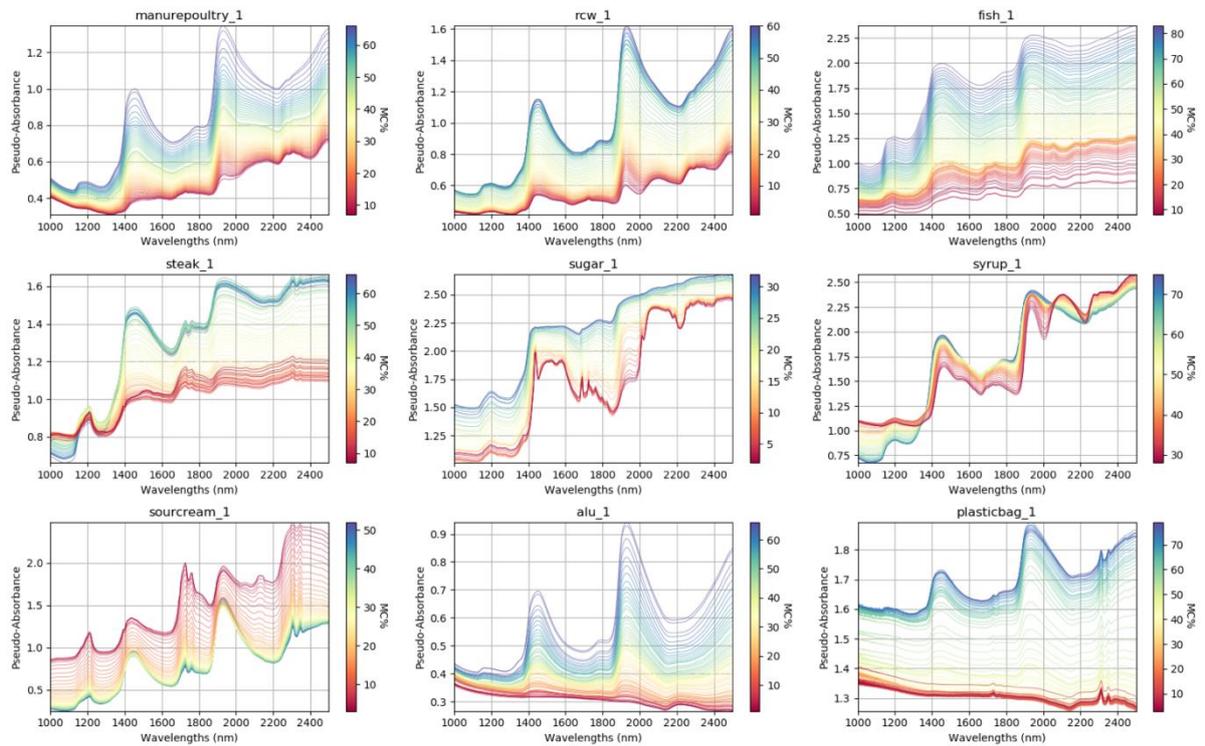


Figure 9. The raw pseudo-absorbance spectra colored by moisture content (blue for high MC, red for low MC). The spectral variations are presented for nine substrates representative of the diversity of biochemical compositions and physical properties (poultry manure, ramial chipped wood / rcw, fish, cooked steak, sugar, syrup, sour cream, aluminum and plastic bag).

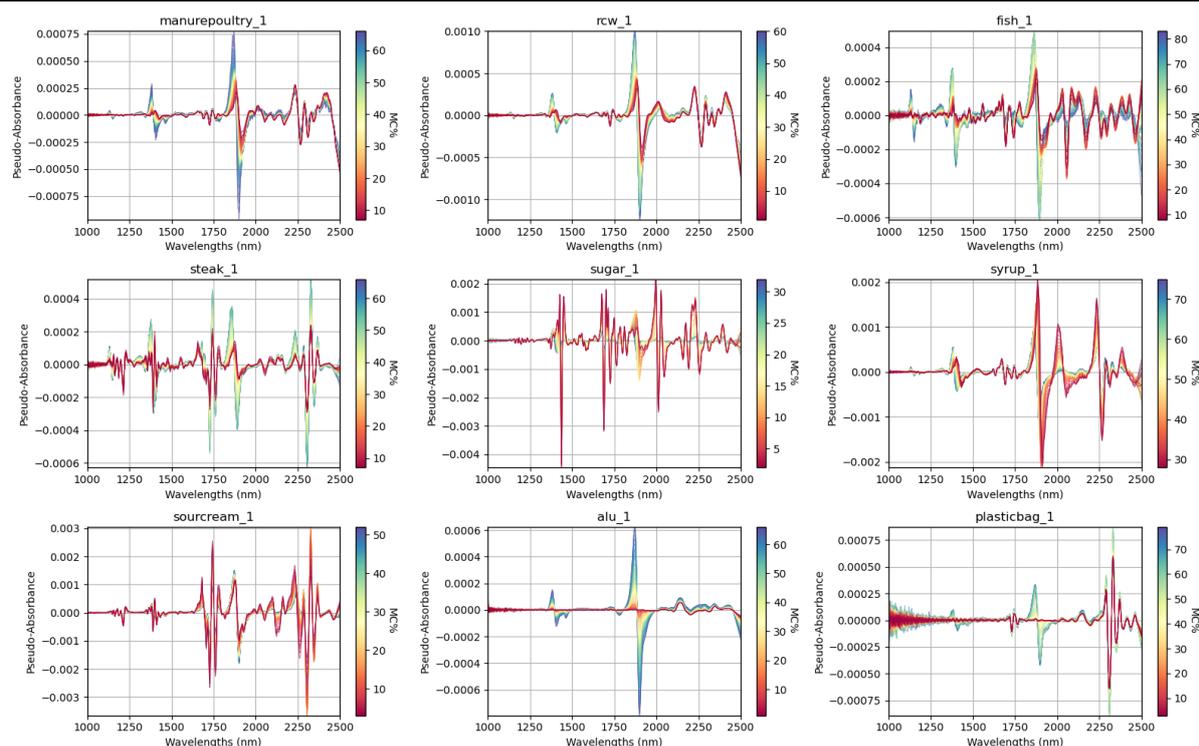


Figure 10. The preprocessed spectra (Savitzky-Golay second derivation with polynomial order of three and window length of 21) colored by moisture content (blue for high MC, red for low MC). The spectral variations are presented for nine substrates representative of the diversity of biochemical compositions and physical properties (poultry manure, ramial chipped wood / rcw, fish, cooked steak, sugar, syrup, sour cream, aluminum and plastic bag).

3. Opportunities

The opportunities provided by this experimental design and the resulting dataset are numerous:

- **Drawing a generic picture of water effects on NIRS.** Indeed, the wide range in moisture content and in biochemical types allows to provide a generic picture of water effects on NIRS. A variety of exploratory data analyses can be applied to describe the effects (*e.g.*, PCA, ICA, MCR-ALS, 2D-COS).
- **Building a model for dry matter content prediction.** This experimental design can be used simply for calibrating models for the prediction of dry matter content on organic waste. One of the advantage of the system is that the temperature remains stable and at ambient conditions (contrarily to other drying systems such as oven-drying or freeze-drying), which means that the built model can then be applied directly on measurements made in the laboratory or online.
- **Evaluating model correction methods.** Using this experimental design the spectral variations related to moisture content variations are captured. This allows to build models robust to variations of moisture

content. Various robustification methods can be developed and tested (orthogonal projections, shrinkage, regularization, transfer) as reviewed in Section 2.1.3.

- **A new way of measuring substrates: adding the moisture content dimension.** While the measurement in NIRS usually consists in a repeated measurement of a sample, the possibility of measuring the sample during its isothermic drying can be envisioned. Indeed, adding the moisture content dimension should provide richer information enabling the separation of substrates. In analogy, temperature-dependent spectra have been proposed where a substrate is scanned at various temperatures; which enables a better characterization of aqueous systems³⁰⁷⁻³⁰⁹.

- **The use of NIRS for better understanding the drying behaviours of food.** The drying behaviours of food samples as well as pharmaceuticals is of particular interest for the design and control of drying processes. The use of NIRS can be used to identify the drying curves phases (both physical and chemical changes), as well as to determine the end of the drying process

Chapter III. Limitations of global correction approaches for moisture content correction

This chapter covers results that are not yet published, but a paper (Paper I – Limitations of global correction approaches for moisture content correction in the context of organic waste characterization) is in preparation and will be submitted shortly.

1. Introduction

As reviewed in Section Chapter I.2.1.4., there is a wide range of global correction strategies that could be applied for accounting moisture content effects. However, these correction have never been assessed in the context of multiple organic waste (which implies diverse physical and biochemical characteristics) and on a wide range of moisture content. This chapter presents a study that addresses this issue by applying ten different global correction approaches in the aim of correcting an existing biochemical methane potential (BMP) prediction model trained on dry samples (MC%<5%). The following methods were employed:

- Two simple approaches which do not require any external data: a spectral range selection approach where water bands are cut (**the “cut water bands” approach**)^{196,260}; and an approach where the dataset is column scaled (**the scaling approach**)⁸⁰.
- The **model update approach** (also called “spiking”, “exhaustive model”, “data augmentation”)^{229,261–263}, which consists in adding the nuisance data to the calibration dataset and building a model classically using PLSR.
- The **External Parameter Orthogonalization (EPO) approach**^{125,128,204,268–273} which consists in projecting the calibration spectra orthogonally to the subspace spanned by the nuisance data.
- The **OSC-EPO approach**^{274,310} which is a variant of EPO where the calibration spectra are also projected orthogonally to the subspace spanned by the nuisance data. However, the nuisance subspace is found by priorily removing the information that is related to the predicted characteristic of interest, using orthogonal signal correction (OSC).
- The **skewing approach using generalized least squares weighing (GLSW)**^{124,289,290} which consists in down-weighting spectral regions of the calibration spectra depending on where and how much the nuisance spectra vary.

-
- The **repeatability file (REPPFILE)**^{260,285–287} **approach** which is a data augmentation strategy that consists in adding the nuisance spectra to the calibration spectra and setting the reference characteristic values to zero. Each of the two blocks (calibration and nuisance) are weighted according to its respective number of samples.
 - The **transfer approach using piecewise direct standardization (PDS)**^{277,278} which consists in building a linear model that transforms the spectra from the nuisance data into the calibration spectra. This methodology has been used successfully for instrument standardization.
 - Since each of these strategies may have their specific advantages and drawbacks, **an ensemble correction approach** is here proposed where the predictions made by all correction methods are stacked together using a PLS model.

2. Materials & Methods

All the methods were coded in Python, and were made available on <https://framagit.org/AlexM/nirsmodelexecution>.

In this study, the drying experiment dataset (presented in Chapter II Section 3) was used. Two datasets were defined:

- The “calibration data” which corresponded to the spectra X_{ref} of freeze-dried and ground substrates, and their corresponding biochemical characteristics Y_{ref} . This dataset allowed to model the relationship between the spectral variations and the biochemical characteristic of interest (here, the BMP).
- The “nuisance data” which corresponds to all the spectra $X_{nuisance}$ acquired during the drying experiment on the 89 substrates, and their corresponding biochemical characteristics $Y_{nuisance}$. This dataset represents the spectral variations from which the model needs to be insensitive to.

The methodology used to train and compare the correction strategies is presented in Figure 12. The train and test sets were defined for the calibration data using the Duplex algorithm³¹¹ applied on Y_{ref} . This allowed to ensure a good representativity in terms of BMP (Figure 11). The substrates’ spectra from the nuisance data were then attributed accordingly to train and test sets. This procedure ensured that no substrates were found both in train and test sets.

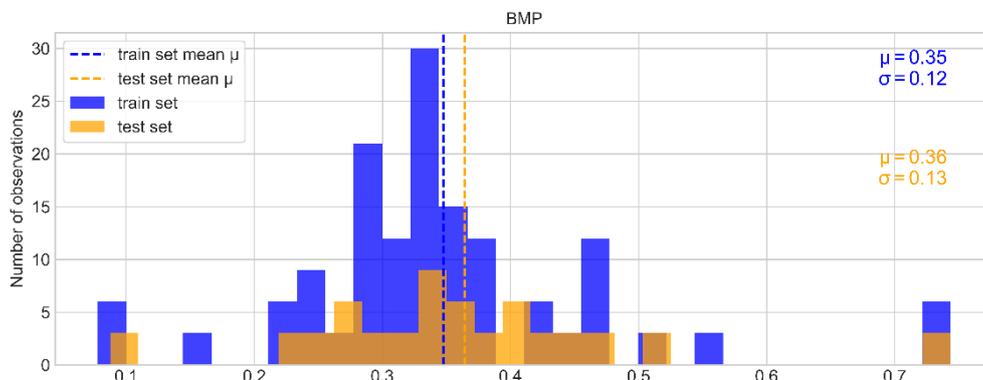


Figure 11. Histograms of BMP values for the train and test set (respectively in blue and orange).

The model correction comparison was run on raw spectra and preprocessed spectra using Savitsky-Golay second derivation (SG2), or VSN. However, because similar results were obtained, only the results on SG2 spectra are provided here.

With the train calibration and nuisance datasets, the model hyperparameters were tuned using a 5-fold cross-validation repeated 30 times. In the cross-validation process, each substrate's spectra were kept in a unique fold to increase the requirement for robust models. The parameter or combination of parameters which minimized the RMSE-CV was selected. Nevertheless, the evolutions of other criteria such as the mean absolute error (MAE), the coefficient of determination (R^2), the variance of b-coefficients and Durbin-Watson (DW) criteria applied to b-coefficients were also evaluated to adjust the final choice of the model parameters. Once the model parameters were selected, the model was trained on the calibration and nuisance train dataset, and predictions were made on the calibration and nuisance test datasets.

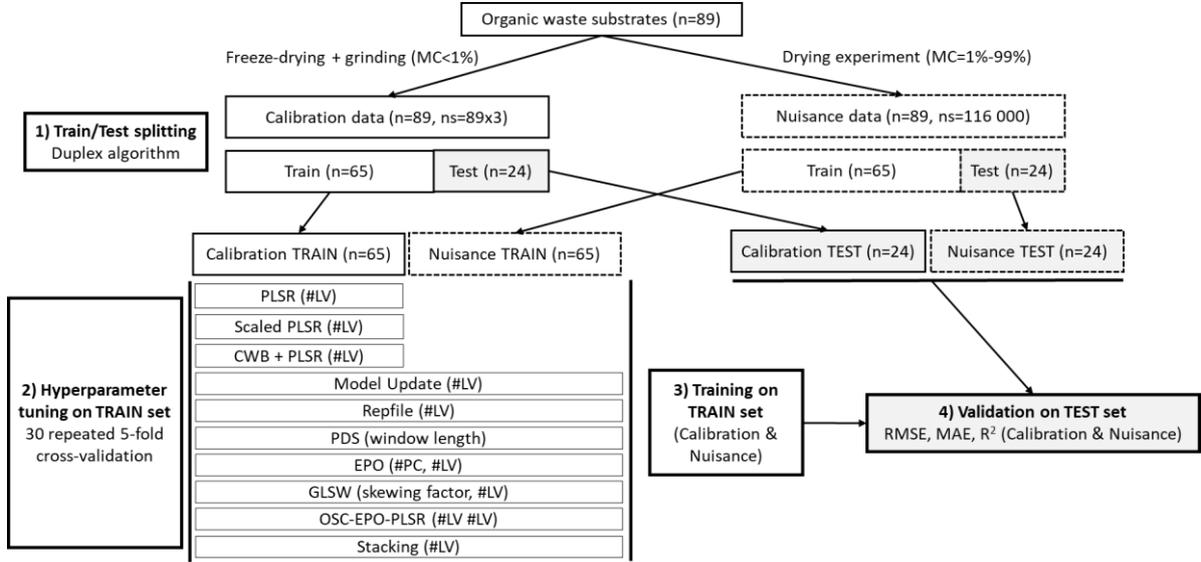


Figure 12. Flowchart for the comparison of moisture content correction strategies.

The formulas of the different statistics used (root mean squared error (RMSE), mean absolute error (MAE), coefficient of determination (R^2), squared Pearson's correlation coefficient (r^2), and bias (b)) are detailed below. Let \mathbf{y}_i be the observed reference variable, and \mathbf{f}_i the predicted reference variable for a given spectrum \mathbf{i} . Let $\bar{\mathbf{y}}$ and $\bar{\mathbf{f}}$ their respective mean values over the whole dataset (train or test set).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\mathbf{y}_i - \mathbf{f}_i)^2}{n}} \quad (\text{Eq. 5})$$

$$MAE = \frac{\sum_{i=1}^n |\mathbf{y}_i - \mathbf{f}_i|}{n} \quad (\text{Eq. 6})$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (\mathbf{y}_i - \mathbf{f}_i)^2}{\sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})^2} \quad (\text{Eq. 7})$$

$$r^2 = \left(\frac{\sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{f}_i - \bar{\mathbf{f}})}{\sqrt{\sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})^2 \sum_{i=1}^n (\mathbf{f}_i - \bar{\mathbf{f}})^2}} \right)^2 \quad (\text{Eq. 8})$$

$$b = \frac{\sum_{i=1}^n (\mathbf{f}_i - \mathbf{y}_i)}{n} \quad (\text{Eq. 9})$$

3. Results & Discussion

All the statistics (RMSE, MAE, R^2) obtained for the cross-validation step (CV), the calibration step (train) and the validation step (test) are represented in Table 3.

Looking at the performances of the control model (*i.e.*, when only PLSR is used on dry spectra), it appears that an RMSE of 65 mL(CH₄).gTS⁻¹ is obtained on the independent validation set, which is slightly higher than the cross-validation RMSE of 52 mL(CH₄).gTS⁻¹. This level of error is coherent with the errors of the models found in the literature (as reviewed in Section Chapter I.1.2.2.1) which vary between 40 and 80 mL(CH₄).gTS⁻¹. As expected, the moisture content considerably affects the prediction accuracy. For this type of model, the error obtained on the nuisance data (*i.e.*, when moisture content varies up to 95%) is more than three times more important with an RMSE of 198 mL(CH₄).gTS⁻¹.

All methods except the scaled pls approach (“SCALING-PLSR”) allowed to enhance the RMSE obtained on the nuisance data (wet form samples) compared to the control model (“PLSR”). Amongst the methods with the lowest errors, the simple cut water bands approach (“CWB-PLSR”) reached an error of 102 mL(CH₄).gTS⁻¹, and the model update approach (“UPDATED-PLSR”) reached an error of 93 mL(CH₄).gTS⁻¹. Compared to the error made on the dry form, there is a 30% error difference. Of course, while these errors are important, these models can be applied on a very wide range of biochemical types, on a very wide moisture content range, and most importantly, with no sample preparation. Unfortunately, the obtained coefficients of determination R^2 all show very low (and negative) values (<0.19). When R^2 is negative, this means that the model is highly biased and is not even better than a model which would always predict the mean of the calibration dataset. This shows that the models are not satisfactory. To better investigate this, the prediction values are plotted against the observed values (in Figure 13). It appears that while most correction methods allow to reduce the range of predictions for substrates with different moisture content levels (*i.e.*, the moisture content effect is reduced), some substrates still have a very wide range of predicted values for different moisture content levels.

Chapter III

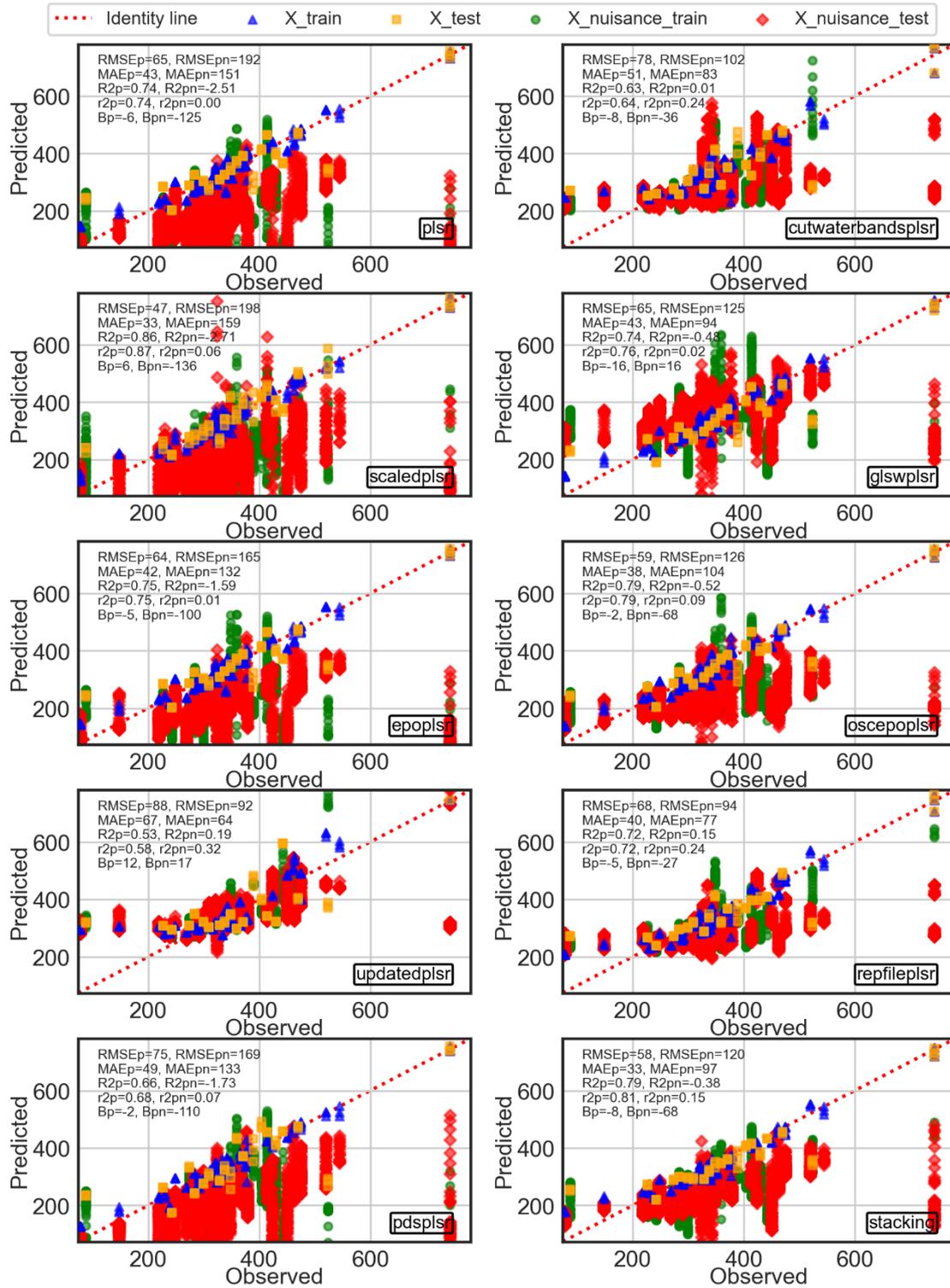


Figure 13. Predicted vs. observed values for the ten correction methods. The dry calibration train set is plot in blue, the dry calibration test set is plot in orange, the nuisance train set is plot in green, the nuisance test set is plot in red. The statistics (RMSE, MAE, R^2 , r^2 , Bias) are provided for the dry calibration test set (p) and the wet nuisance test set (pn).

Chapter III

Table 3. Quality of models corrected from moisture content. The RMSE, MAE and R² are provided for cross-validation (CV), calibration (C), prediction (P) (calibration data), prediction (PN) (nuisance data).

Model type	Parameters		All train data						Calibration data (dry form)						Nuisance data (wet form)					
	K	#LV	Cross-validation (CV)			Calibration (train)			Validation (test)			Calibration (train)			Validation (test)					
			RMSE	MAE	R ²	RMSE	MAE	R ²	RMSE	MAE	R ²	RMSE	MAE	R ²	RMSE	MAE	R ²			
PLSR		7	52	37	0.62	29	23	0.94	65	43	0.74	198	155	-2.6	193	152	-2.51			
CWB-PLSR		4	61	44	0.51	53	36	0.82	78	52	0.63	112	80	-0.14	102	84	0.01			
SCALING-PLSR		6	48	36	0.69	27	21	0.95	48	33	0.86	215	168	-3.21	198	159	-2.71			
GLSW-PLSR	0.05	7	66	53	0.46	28	22	0.95	66	43	0.74	110	76	-0.1	125	94	-0.48			
EPO-PLSR	1	7	51	37	0.64	29	22	0.94	64	42	0.75	173	131	-1.72	166	133	-1.59			
OSC-EPO-PLSR	3	6	50	37	0.67	26	21	0.95	59	39	0.79	133	95	-0.61	127	105	-0.52			
UPDATED-PLSR		3	65	48	0.54	65	45	0.72	89	67	0.53	82	49	0.39	93	64	0.19			
REPPFILE-PLSR		6	52	37	0.62	35	24	0.92	66	39	0.74	108	75	-0.05	101	86	0.04			
PDS-PLSR	21	9	46	35	0.70	24	18	0.96	75	50	0.66	169	128	-1.6	170	133	-1.73			
STACKING		1	47	34	0.73	31	22	0.94	58	33	0.8	130	97	-0.55	121	98	-0.38			

4. Concluding remarks

- To our knowledge, this was the first time that a global correction method was applied to NIRS calibration models built on solid organic waste.
- The application of global correction methods appears to be rather limited for building robust calibration models on diverse organic waste. While these were successful for reducing the influence of moisture content, the errors in prediction for one substrate at various moisture content levels remain very high.
- In addition, the range of predictions for some substrates at equal ranges of moisture content appear to be very different. This may mean two things: (i) that the sensitivity to moisture content is different from one substrate to another, and/or (ii) that the moisture content effects are spectrally different from one substrate to another.
- A deeper analysis of the water effects, and its dependency on the substrate type (biochemical composition and physical properties) remains to be assessed.

Chapter IV. Unveiling the non-linearity of moisture content effects in organic waste using principal components analysis (PCA)

This chapter covers results presented in Paper II – Unveiling non-linear water effects in near infrared spectroscopy: A study on organic wastes during drying using chemometrics (published in Waste Management journal). See section “Included Papers” for the full reference and published article.

1. Introduction

In Chapter III Section 4, it appeared that better understanding water effects and how they relate to the substrate properties could be key for the development of robust calibrations models on wet substrates. Indeed, groups could then be used for building local models, an approach which has been shown to be successful for BMP prediction on plant biomasses¹⁶⁶.

As mentioned in Chapter I Section 1.3.1, the effects of moisture content on near infrared spectra has been described for a wide variety of different matter types including soil^{190–196}, crops^{197–201}, food²⁰², plants²⁰³, wood²⁰⁴, pharmaceuticals²⁰⁵, object models^{206–208}, and water-dominant systems²⁰⁹. In addition, though not focused on the analysis of moisture content effects in NIRS, some studies use NIRS to monitor drying or hydration processes where moisture content varies^{210,211}.

However, no study has yet analyzed and compared moisture content effects in one comprehensive experiment with a wide variety of biochemical and physical types. In light of this, this chapter presents the analysis of the drying experiment dataset using PCA. The reader is invited to refer to Paper I for further details.

2. Materials & Methods

The choice of principal components analysis (PCA)

In order to disentangle the water effects on NIR spectra, different exploratory data analysis methods could be used (as reviewed in Section 1.2.1.2). However, PCA appeared as the most suitable method for

the initial analysis of such data with no *a priori* knowledge. Indeed, PCA simply finds orthogonal components that explain most of the variance, and has the benefit of being a nested method, meaning that contrarily to ICA or MCR-ALS, they will provide the same components when adding new components to the model. However, other methods are more suitable for obtaining independent or pure components, with a flexible framework proposed in MCR-ALS where various constraints can be added (as further investigated in Chapter VI).

Pretreatments

Because NIR spectra contain both physical and chemical information, a wide variety of pre-processing techniques⁷⁹ are often used to maximize the chemical information of interest in the spectra. However, the pre-processing steps may bring important artefacts and deport chemical information, which can make the assignment of bands more complicated¹³⁴. Because the database is highly heterogeneous, the baselines due to scattering are expected to be very different; so it appeared best to apply PCA directly on the raw spectra. In particular, this has the benefit of assessing the absolute amplitudes of the various sources of variations.

3. Results & Discussion

The high dimensionality of this dataset

While only the first eight components were studied in depth, the information dimensionality of the dataset is much higher. Different criteria allow to assess this, including the cumulative explained variance, and the Durbin-Watson (DW) criterion³¹². This latter criterion is classically used to detect the presence of autocorrelation in prediction residuals from regression models and ranges between 0 and 4. However, it is used here to appreciate the structure of the loadings for each component: are the loadings noisy (with a high DW value) or are the loadings still showing NIR-like spectral information (with a low DW value)? The evolutions of these criteria according to the principal component number are presented in Figure 14. While the explained variance would suggest about 15 components as the suitable dimension, the DW criterion would suggest more about 30 components. Indeed, when plotting several loadings up to the 100th (left subplot of Figure 14), the loadings show NIR-like features up to at least the 25th component. This illustrates very well the high complexity of the acquired dataset presented in Chapter II.

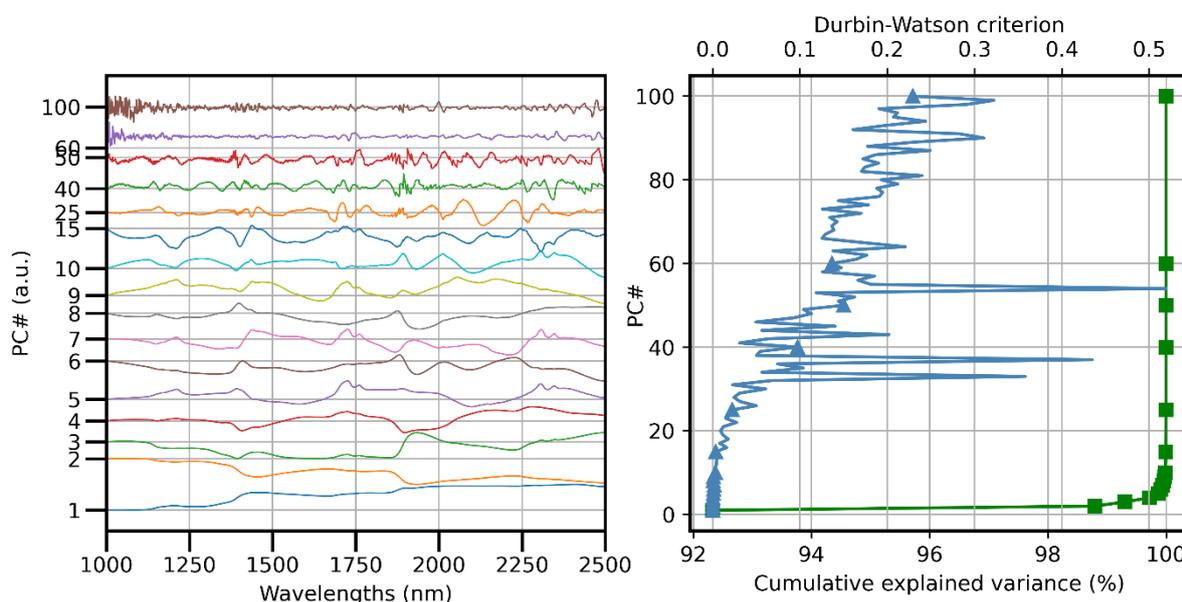


Figure 14. The evolution of loadings (left subplot) and the evolutions of cumulative explained variance and the Durbin-Watson (DW) criterion, according to the principal component number (PC#). The blue curve is the DW criterion and the green curve is the cumulative explained variance.

Analysis of scores' kinetics and loadings

In this study, the interpretation of the spectral regions contributing to each components (loadings) (Figure 16) and the analysis for each component of how each substrates' scores evolve along drying (Figure 17) allowed to reveal the complexity of moisture content effects. The detailed analysis can be found in Paper II.

Spectra reveal substrates with a wide range of physical properties (*i.e.*, different scattering levels)

The first component which accounts for more than 92 % of the total variance was shown to correspond to global pseudo-absorbance level differences which are unrelated to specific spectral regions (*i.e.*, specific chemical compound). Two groups were identified: substrates with high scores (Figure 17) such as sugar or raw fish exhibiting high pseudo-absorbance levels (~ 2.3 - 2.5); and substrates with low scores such as aluminum pellets or poultry manure exhibiting low pseudo-absorbance levels (~ 1.5 - 1.8). These two groups have very different physical properties which impact how light scatters within matter, and the resulting amount of reflecting light collected by the detector. Figure 15 presents the visual aspect of these samples to illustrate the physical differences: liquid samples (wet sucrose, syrup or dry sourcream) often show higher pseudo-absorbance levels than solid samples (dry poultry manure, ramial chipped wood). Indeed, a liquid state is often associated with a lower number of particle interfaces where light could be refracted, which implies that light travels deeper in the matter.



Figure 15. Visual aspect of substrates with low and high levels of pseudo-absorbance. Pictures were taken from above the NIRS measurement cell.

Moisture content affects physical properties of substrates (which affects light scattering)

What is interesting is how moisture content affects these physical properties. From looking at the scores of the first component (Figure 17), most samples show a global decrease of pseudo-absorbance as moisture content levels decrease. This means light is further reflected as moisture content decreases. The main reason are the changes along drying of the refractive index differences between particles ($n_{\text{water}} \approx 1.33$, $n_{\text{air}} \approx 1.0$, $n_{\text{organicmatter}} \approx 1.4$)³¹³. Indeed, for most substrates, as drying occurs, water is replaced by air, which leads to increased number and intensity of refractive index differences. This leads to an increased scattering with higher reflectance levels (*i.e.*, lower pseudo-absorbance levels).

However, this was shown to be different for samples such as sour cream (but also butter, yoghurt, pork meat), which show increasing scores (Figure 17) as moisture content decreases. When looking at the visual aspect (Figure 15), it appears that the dried sour cream has turned biphasic, with a high-fat content liquid in which light is further transmitted. In these systems, as drying occurs, water is replaced by fat and not air, which leads to lower refractive index differences ($n_{\text{vegetableoil}} \approx 1.47$).

Another important aspect found through this study concerns the non-linearity of the scores' evolutions (first principal component in Figure 17). Most substrates do not show a linear relationship with moisture content. This suggests that linear models such as PLS will require several latent variables to approximate this relationship. This is further analyzed in Chapter V.

Moisture content affects chemical composition of substrates (which affects light absorption)

By definition, moisture content affects the quantity of water in the system. This means that the level of light absorption is directly affected. In the second component, the loadings (Figure 16) show the three broad absorption peaks that can be found in pure water transmission spectra (at 1209 nm, 1456 nm, and 1933 nm). Most substrates show scores (Figure 17) that decrease along drying which shows that moisture content directly affects the quantity of light absorbed by water. Similarly, the levels of organic matter (PC4), fat (PC5), or carbohydrates (PC6) all vary along the moisture content variations.

Moisture content variations not only impacts the quantity of water in the system, but the state of water molecules. This means that the absorption pattern of water molecules, and more precisely of the OH bond vibrations, is itself modified along drying. Two different spectral patterns related to water's chemical interaction (*i.e.*, water state) were identified in the third and eighth components. Indeed, it was shown in the third component that small differences between the first overtone absorbance band at 1430 nm and the second combination absorbance band at 1940 nm is associated with the presence of simple carbohydrates or proteins, both of these molecules forming important interactions with water through hydrogen bonding. In addition, in the eighth component, as drying occurs, a shift of the OH absorbance bands from high energy vibrations to lower energy vibrations was highlighted for most substrates. While these chemical interactions and changes of water state certainly occur, they seem much more limited in terms of variance in comparison with the effects of moisture content on scattering. When analyzing datasets of complex materials with high differences in physical properties all together, these changes are difficult to assess. However, when analyzing each material separately, these effects become clearer.

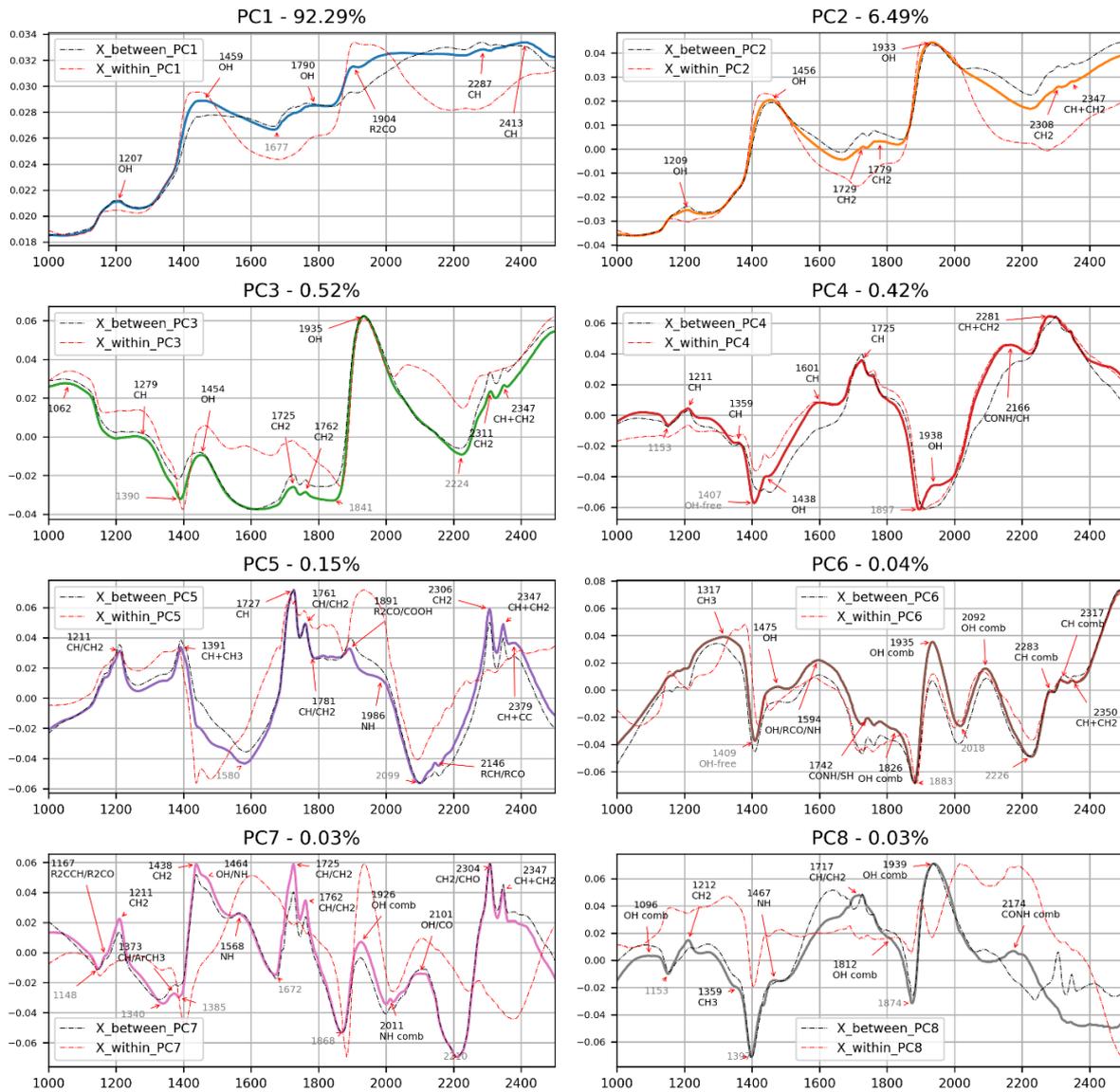


Figure 16. Loadings of the first eight components. Explained variance percentage of each principal component is given in the title. For each component, the corresponding eigenvector of the between-substrate variance-covariance matrix (see Eq. 8 in Paper I) is plot (in dashed black line), as well as the corresponding eigenvector of the within-substrate variance-covariance matrix (Eq. 9 in Paper I) (in dashed red line).

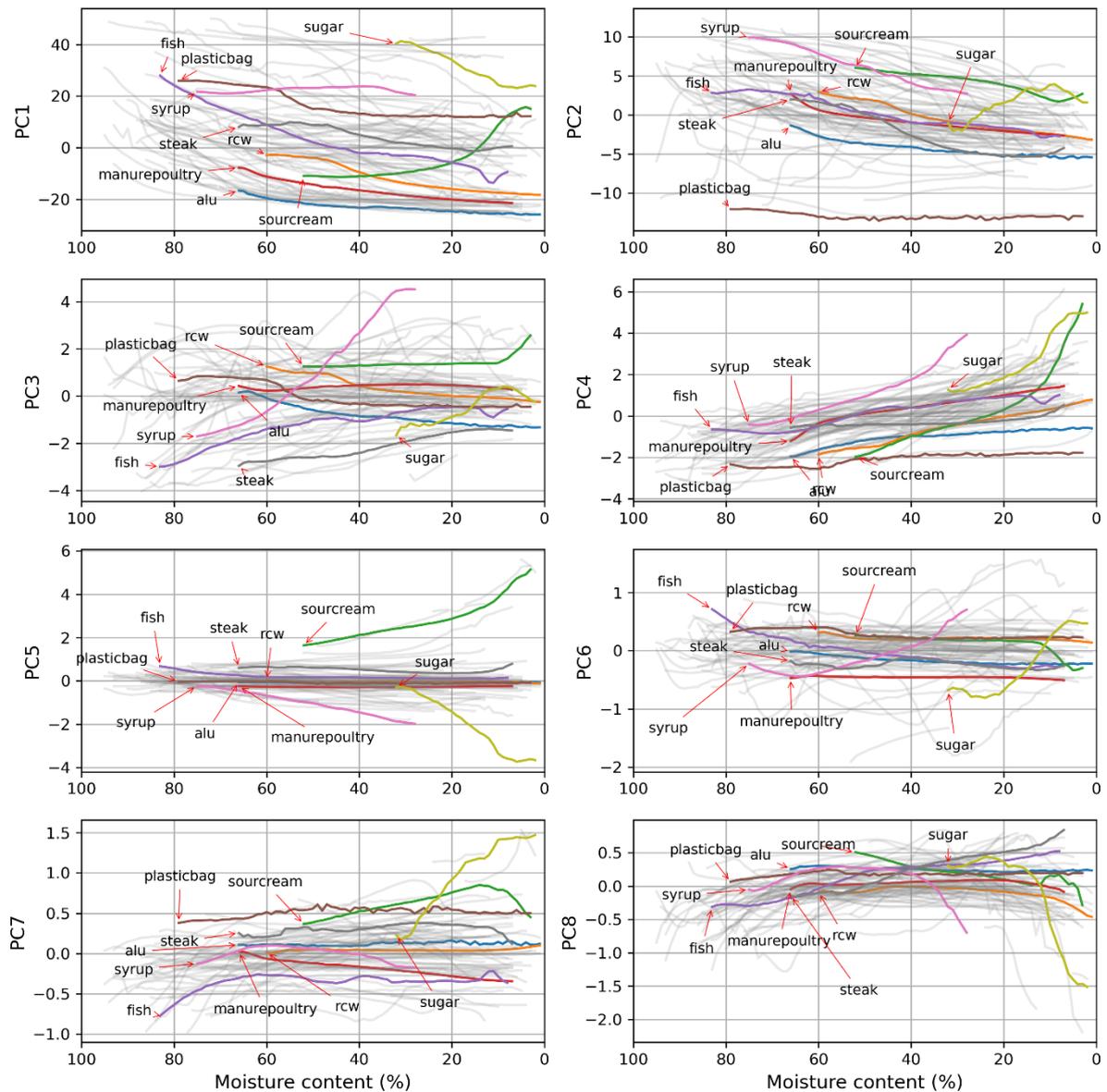


Figure 17. Evolutions of the first eight components' scores along drying for nine representative substrates. Other substrates are colored in grey.

Physico-chemical changes occurring during drying

During drying, a wide variety of physico-chemical changes may occur such as sticking, caking, collapse, crystallization, agglomeration, loss of volatiles, browning and oxidation.³¹⁴ All these changes are likely to modify the light scattering and these transitions certainly make the measured NIR signal prone to non-linearities.

4. Concluding remarks

Through the use of PCA, the complexity of moisture content effects on NIRS were revealed. Moisture content effects were shown to affect both the biochemical composition and the physical properties of substrates. Such modifications directly impact the NIR signal that is measured: changes in physical properties affect the amount of light that is scattered and transmitted in the material, and changes in biochemical composition affect the amount and frequency of light that is absorbed by the material.

In addition, it was shown that the main spectral variance (represented by the first and second component) showed a non-linear evolution with moisture content. To investigate this more theoretically, the effects on scattering and absorption need to be analyzed independently.

Chapter V. Modeling the influence of water content on scattering: towards a new Bouguer-Beer-Lambert law for wet scattering samples

This chapter covers results presented in Paper III – Relating Near-Infrared Light Path-Length Modifications to the Water Content of Scattering Media in Near-Infrared Spectroscopy: Toward a New Bouguer – Beer – Lambert Law (published in Analytical Chemistry journal). See section “Included Papers” for the full reference and published article.

1. Introduction

In Chapter IV, moisture content variations of substrates were shown to strongly modify the light scattering. This study aims to further understand these effects. In order to isolate these effects, a model system composed of aluminum paper pellets mixed with liquid water was analyzed using the drying system with NIR acquisition (presented in Chapter II). The advantage of this sample was that the absorption of NIR light by aluminum is insignificant, which removed any potential absorption interaction term.

2. Results

A new formulation of the Bouguer-Beer-Lambert law that takes into account path-length modifications

For homogeneous isotropic media without particles, transmission measurements respect the Bouguer-Beer-Lambert law⁵⁰:

$$A_{\lambda,c} = -\log(T_{\lambda,c}) = \varepsilon_{\lambda} \cdot L \cdot c \quad (\text{Eq. 10})$$

With $A_{\lambda,c}$ the absorbance, $T_{\lambda,c}$ the transmittance, ε_{λ} the extinction coefficient, L the light path-length and c the concentration of the absorbing species.

Chapter V

In scattering media, two phenomena must be taken into account³¹⁵: a modification of path-length $L_{\lambda,c}$, and a loss of photons $f_{\lambda,c}$:

$$A_{\lambda,c} = \varepsilon_{\lambda} \cdot L_{\lambda,c} \cdot c + f_{\lambda,c} \quad (\text{Eq. 11})$$

In most studies interested in the multiplicative effects on NIR, $L_{\lambda,c}$ is modeled with a simple multiplicative constant k not directly related to the analyte concentration⁹⁶:

$$L_{\lambda,c} = k l_0 \quad (\text{Eq. 12})$$

with l_0 a constant path-length.

However, there are many cases where the absorptive species of interest is in fact directly responsible for the multiplicative effects. As discussed in Chapter IV water content variations during drying induce scattering modifications because of the increased number or intensity of refractive index differences between the gas phase, liquid phase and solid particles³¹⁶. In such cases, it seems reasonable to relate the light path-length modifications (multiplicative effect) directly to the concentration of the absorptive species (water content). Because such relationship is expected to involve a geometrical relationship, it is here proposed that the path-length function is set as to be a simple power function of the water content concentration c :

$$L_{\lambda,c} = l_0 \cdot c^{a_{\lambda}} \quad (\text{Eq. 13})$$

with a_{λ} a power coefficient dependent of the wavelength λ . Indeed, when looking at the absorption law as a probabilistic law, the more light is scattered through the media, the higher the chances for photons to be absorbed by the matter. It seems reasonable to suggest this direct relationship of path-length modifications to water content.

Combining (Eq. 11) and (Eq. 13) yields a new formulation of the absorption law for scattering media:

$$A_{\lambda,c} = \varepsilon_{\lambda} \cdot l_0 \cdot c^{a_{\lambda}+1} + f_{\lambda,c} \quad (\text{Eq. 14})$$

This additive baseline term $f_{\lambda,c}$ can be removed by a wide number of pretreatments, but in order to preserve spectroscopic relative amplitudes and positions, the extended multiplicative scatter correction (EMSC) framework was used. A pure water transmission spectrum acquired using a JASCO V560, was used as the reference spectrum (after transforming to absorbance using a logarithm). A constant, first-

Chapter V

order and second-order polynomial was included as in the original EMSC⁹⁶. An interferent pure spectrum was added as the measured spectrum of the fully dried aluminum. Finally, in order to evaluate the additive and multiplicative terms only in regions where water absorbance bands are known to interfere less, a weighted least-squares (WLS) was applied with weights set as the inverse of the pure water transmission spectrum. This resulted in the following EMSC model:

$$A_{observed,i}(\lambda) = a_i \cdot A_{pure\ water}(\lambda) + b_i + c_i \lambda + d_i \lambda^2 + e_i A_{pure\ aluminum}(\lambda) \quad (\text{Eq. 15})$$

Instead of correcting for both the additive and multiplicative effects, only additive effects were corrected for:

$$A_{corrected,i}(\lambda) = A_{observed,i}(\lambda) - (b_i + c_i \lambda + d_i \lambda^2 + e_i A_{pure\ aluminum}(\lambda)). \quad (\text{Eq. 16})$$

This yielded spectra corrected only for additive effects:

$$A_{corrected,i}(\lambda) = A_{\lambda,c} - f_{\lambda,c} = \epsilon_{\lambda} \cdot l_0 \cdot c^{a_{\lambda}+1} \quad (\text{Eq. 17})$$

Applying the log-transform then results in the following formula:

$$\log(A_{\lambda,c} - f_{\lambda,c}) = \log(\epsilon_{\lambda} \cdot l_0) + (a_{\lambda} + 1) \cdot \log(c). \quad (\text{Eq. 18})$$

For each wavelength λ , an ordinary least squares (OLS) regression was applied between $\log(A_{\lambda,c} - f_{\lambda,c})$ and $\log(c)$ to evaluate $a_{\lambda} + 1$ (the slope) and $\log(\epsilon_{\lambda} \cdot l_0)$ (the intercept). To evaluate goodness of fit, the coefficient of determination (R^2) was calculated.

In Figure 18, the results from this analysis are presented for a wavelength of 1450 nm where water OH bonds are known to absorb strongly. In Figure 18 (A) the raw absorbance at 1450 nm show non-linear evolutions with water content. Moreover, differences between absorbance values can be observed despite identical water content. These differences were well removed by the EMSC correction (as shown in Figure 18 (B)).

Figure 18 (C) shows a very good fit ($R^2 = 0.995$) between the log-transformed corrected absorbance values and the log-transformed water contents. This suggests that the path-length modifications induced

by water content at 1450 nm can be well modeled using a simple power law function of water content. The obtained slope of the regression line (1.524) is an estimation of $(a_\lambda + 1)$ in eq.(5) and eq.(9).

After running the same analysis to all wavelengths, it was shown that this law is valid for all wavelengths greater than 1150 nm ($R^2 > 0.9$).

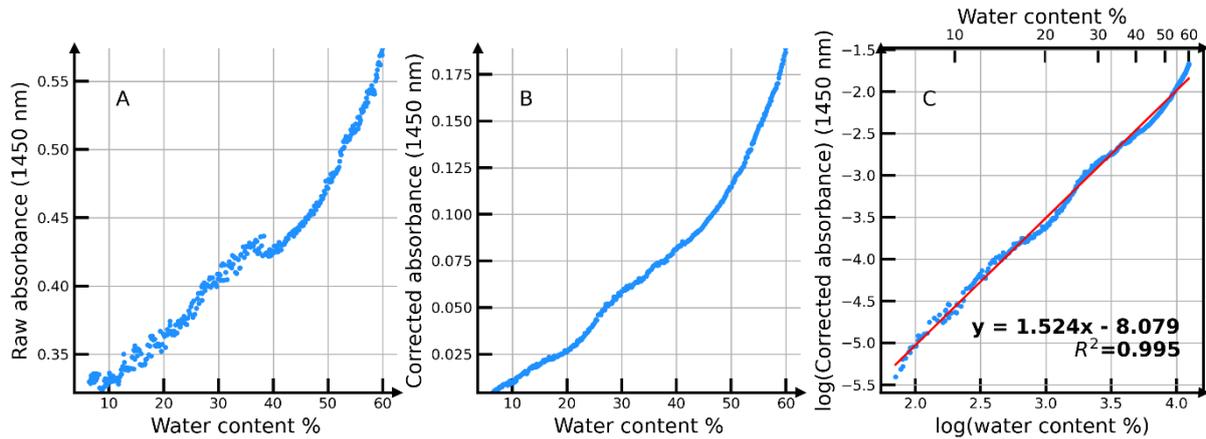


Figure 18. Evolutions with water content % of (A) raw absorbance values, (B) corrected absorbance values, and (C) log-transformed corrected absorbance values are provided. The latter log-transformed corrected absorbance values are plot with log-transformed water content %, In the (C) subplot, the OLS regression line is plot in red, with the slope, intercept and coefficient of determination (R^2).

Implications for fundamental water structure studies

By properly retrieving the additive and multiplicative effects, the real extinction coefficients (multiplied by a given factor l_0) were obtained ($\epsilon_\lambda \cdot l_0$) (exponential of the green curve in Figure 19). This signal can be further analyzed to better understand the water structure, and in particular, to better understand what constitutes the broad first overtone OH band at 1450 nm (green curve in Figure 19). Indeed, by applying Savitzky-Golay second derivative (an effective deconvolution technique amongst others³¹⁷), two clear negative sub peaks were identified at 1405 nm and 1469 nm (orange curve in Figure 19). The peak positions corresponded perfectly to the identified peaks in other studies on pure water measurements^{216,224}. Other peaks at 1335 nm and 1537 nm may also be identified, and relate well to different water molecular conformations as described by aquaphotomics²²⁷.

To summarize the point here, in highly diluted systems, the measured transmission spectra relate well with the chemical absorption, which allows to identify correctly the specific absorption bands by water components and structures. However, in scattering media, and reflectance measurements, these effects need to be taken into account, otherwise wrong assignments will be made.

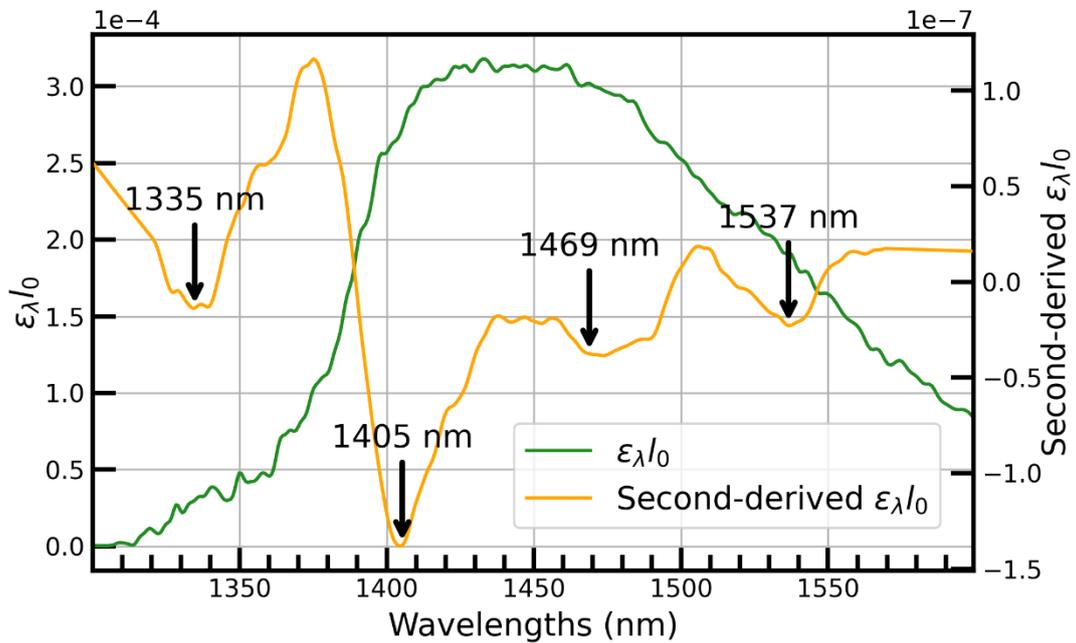


Figure 19. The exponential of the fitted intercept values (ϵ_{λ}/l_0) in eq.(9) (in green) and the corresponding Savitzky-Golay second derivative (in orange).

Implications for quantitative calibrations

It appeared that model calibration would benefit from using the observed power law relationship between the absorbance and water content. For example, a log transformation of both spectra X and reference values y , would make the relationship between absorbance and water become linear again, therefore enabling better water content prediction models. To showcase this, a comparison of four models was provided in Figure 20, and showed that with one PLS latent variable, the lowest RMSEC was obtained by a model where X was first corrected from additive effects using EMSC and then log-transformation was applied on both X and y to linearize the relationship as in (Eq. 18) (red curves in Figure 20). This showed the strong effect of linearizing the relationship between absorbance and water content: the model was simpler and thus more robust. Indeed, in order to reach the same prediction error value (0.8% of water content) of the one latent variable model obtained with $\log(\text{EMSC}_{\text{additive}}(X)) \sim \log(y)$, the other models $X \sim y$ (blue curve), $\text{EMSC}_{\text{additive}}(X) \sim y$ (orange curve) and $\text{EMSC}_{\text{complete}}(X) \sim y$ (green curve) needed respectively four, three and six more latent variables.

Of course, with sufficient number of latent variables (greater than 15), the models without the log transform appeared to follow the same error of prediction as the log-transformed model. This illustrated

well how the PLS algorithm is capable of taking into account non-linear relationships, though it is probable that these models were overfitted. Interestingly, applying a complete EMSC with removal of both additive and multiplicative effects appeared very bad in this case for water content prediction (green curves in Figure 20). Indeed, by removing multiplicative effects, the strong relationship between water content and these effects was lost. Of course, it should be noted that, as discussed in other studies¹¹³, applying a logarithm to y is not without consequences: the original distribution of y value is deformed which may cause problems for regression. Moreover, one necessary condition is that the additive effects including the dry spectrum should be removed by pre-processing before applying the log transformation. However, it appeared clear from these results that taking into account the power law relationship between water content and light path-length allowed a significant improvement of predictive models. Moreover, it was shown that removing these scattering effects (the objective of $EMSC_{complete}(X) \sim y$) to predict dry matter content may not be most effective, as it removes an information which is directly dependent of the moisture content.

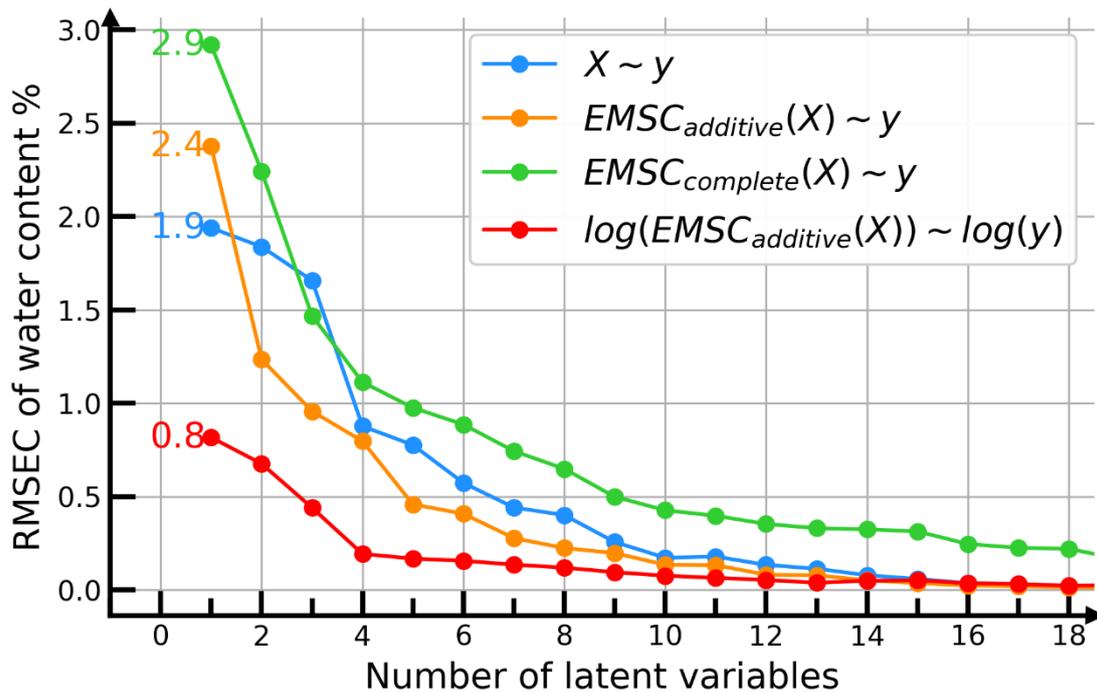


Figure 20. RMSEC curves obtained for four different models: $X \sim y$ (in blue), $EMSC_{additive}(X) \sim y$ (in orange), $EMSC_{complete}(X) \sim y$ (in green), and $\log(EMSC_{additive}(X)) \sim \log(y)$ (in red).

3. Concluding remarks

The introduction of path-length modifications directly related to a power law of moisture content into the BBL law allowed us to fully explain the variations observed in aluminum mixed with water. In this scattering model system, the dry matter component (aluminum) did not absorb any light, which enabled us to identify exclusively the scattering modifications induced by moisture content differences.

The implications of this new theoretical framework were discussed. Firstly, it should benefit fundamental water structure studies where the pure spectral features are of importance: the extinction coefficients were directly determined. Secondly, it was shown to benefit quantitative calibrations built on varying moisture content systems: models for dry matter content estimation are made simpler and therefore more robust.

After having provided a clear qualification of scattering effects induced by moisture content, it appeared necessary to investigate this theory in more complex systems with biochemical absorptions. While these may lead to complex interactions, it can be hypothesized that other absorbing contributions in fact also follow a power law relationship with moisture content. The higher the moisture content level, the higher the photons penetrate in the matter, leading to a higher probability of photons to be absorbed, should it be by water or by the biochemical components. However, as showed in Chapter IV, this behavior differs amongst substrates. The use of this law could allow a better identification of groups of spectra with common moisture content effects, as investigated in Chapter VI.

Chapter VI. A knowledge-based local approach to account for water effects in NIRS

1. Introduction

It appeared in Chapter III that global correction methods were not able to provide satisfactory models when applied to the whole range of biochemical types found in organic waste as well as with a wide range of moisture content. This observation was further explained by the non-linearity of water effects (Chapter IV). Indeed, with the use of PCA, the water effects were shown to be dependent of both physical and biochemical properties, as well as to the range of moisture content. From these results, one way of dealing with water effects that has been proposed relies on building local models (*i.e.*, models based on subsets of the whole dataset for which the water effects can be linearly modeled). In this chapter, a practical methodology to define these groups is proposed and discussed.

In Chapter 5, it was shown that water effects on scattering (in particular, path-length modifications) could be well modeled by a power law in simple systems such as aluminum. It seems that this new theoretical framework could be leveraged for answering two objectives: defining groups of samples with similar water effects, but also better identifying the spectral contributions of dry matter constituents within spectra. This chapter discusses these two possibilities.

2. Building predictive models on a local subset of substrates with homogeneous moisture content effects

2.1 Defining the local subset

In Chapter IV, the evolutions of scores for the first component according to moisture content (Figure 17) allow us to identify a first group of substrates with a common behavior towards moisture contents. Indeed, these substrates' spectra all seem to follow a power relationship with moisture content levels, like in the aluminum sample. A new PCA was run on these selected samples ($n=37$) and the first component scores are presented in Figure 21. A non-linear least squares fitting procedure (Levenberg-Marquardt algorithm)³¹⁸ was used to fit the scores T_1 of the first component with the following power relationship with water content c :

$$T_1(c) = k_1 \cdot c^{k_2} + k_3$$

(Eq. 19)

The fitted values are presented in Figure 21 as dotted red lines, along with a boxplot of the R^2 obtained for each of the substrates. The fitted values show a good adequacy with the observed values as confirmed by the very good R^2 values (>0.95). This further validates the power law framework presented in Chapter 3, by showing that moisture content effects on organic waste can be well-modeled by a power law relationship, and this even when the dry matter constituents absorb. However, it seems the parameters of this power law relationship (both the intercept, and the power constant) differ from one substrate to another.

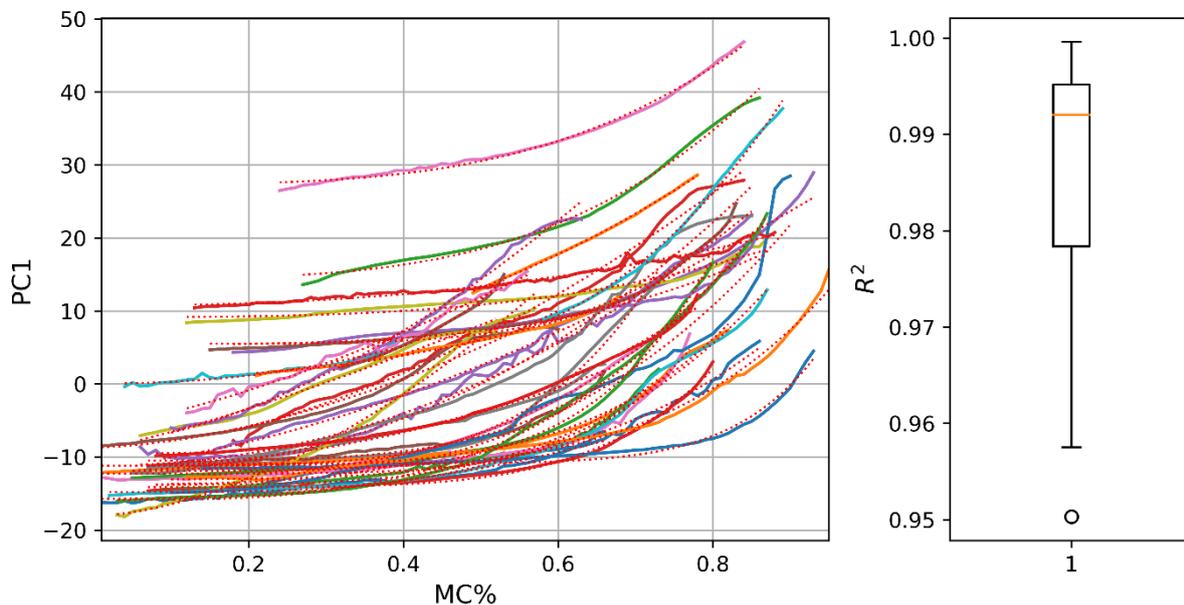


Figure 21. Scores kinetics for the first principal component of the identified subset with the fitted power laws in red dotted lines (left subplot). Boxplot of the R^2 obtained for each fitted power law.

Qualitatively, what relates all these substrates together appears to be that they all present stable porous solid structures: when drying occurs, the water is simply replaced by air. Within substrates that do not belong to this group, there are substrates which were shown to turn biphasic during drying (with opposite baseline variations): butter, cheese, sour cream, yoghurt, pork, pesto, french fries. In addition, there are also substrates which are highly soluble in water and turn from liquid transparent solutions to crystalline solids during drying: sucrose, soya sauce, jam. In these samples, the signal at high moisture content level appears flat. For these substrates, one possibility is that instead of following a power relationship, they follow a power relationship followed by a plateau as modeled by logistic or Gompertz functions for example³¹⁹.

2.2 Fitting the local model

While the moisture content effects appear more homogeneous within this group, there is still the need to know how well models can cope with these effects. To answer this, different models were built to predict biochemical methane potential as done in Chapter III. While all the strategies that were used could be employed, only the simple PLSR approach and the model update approach were compared. Indeed, in Chapter III, these two methods provided the worst and best results so they are well suited for benchmarking the different model qualities. In addition, models were built on more restricted moisture content ranges (moisture content under 60%, under 40% and under 20%) to better evaluate whether the poor performances observed in Chapter III come from the differences of moisture content effects between substrates, or due to the magnitude of moisture content effects themselves. All the models were built following the same model calibration procedure presented in Chapter III.

The predicted and observed values are presented for each of the obtained models in Figure 22. As expected, when the model is trained only on dry samples (left column), the prediction errors on the wet samples (nuisance data) are bad, with an RMSEPN of $196 \text{ mL}(\text{CH}_4).\text{gTS}^{-1}$ when using the full moisture content range (first line), and $124 \text{ mL}(\text{CH}_4).\text{gTS}^{-1}$ when using samples with moisture content under 20% (last line). Fortunately, correction methods such as the updated PLS-R help to reduce this error to $96 \text{ mL}(\text{CH}_4).\text{gTS}^{-1}$ when using the full range (right column). However, this corresponds to the same RMSE level as the one which was obtained for models built with all the substrates ($\text{RMSEPN} = 92 \text{ mL}(\text{CH}_4).\text{gTS}^{-1}$ in Figure 13). This means that the moisture content effects, while being more homogeneous, still remain high. However, when building the same models on more limited moisture content ranges (below 60% as presented in the second line of Figure 22, below 40% as presented in the third line of Figure 22 and below 20% as presented in the fourth line of Figure 22), the quality of models appear much better with RMSEP of respectively $87 \text{ mL}(\text{CH}_4).\text{gTS}^{-1}$, $74 \text{ mL}(\text{CH}_4).\text{gTS}^{-1}$ and $42 \text{ mL}(\text{CH}_4).\text{gTS}^{-1}$. This further confirms the potential of building local models based on both biochemical and physical properties, and on moisture content ranges.

Chapter VI

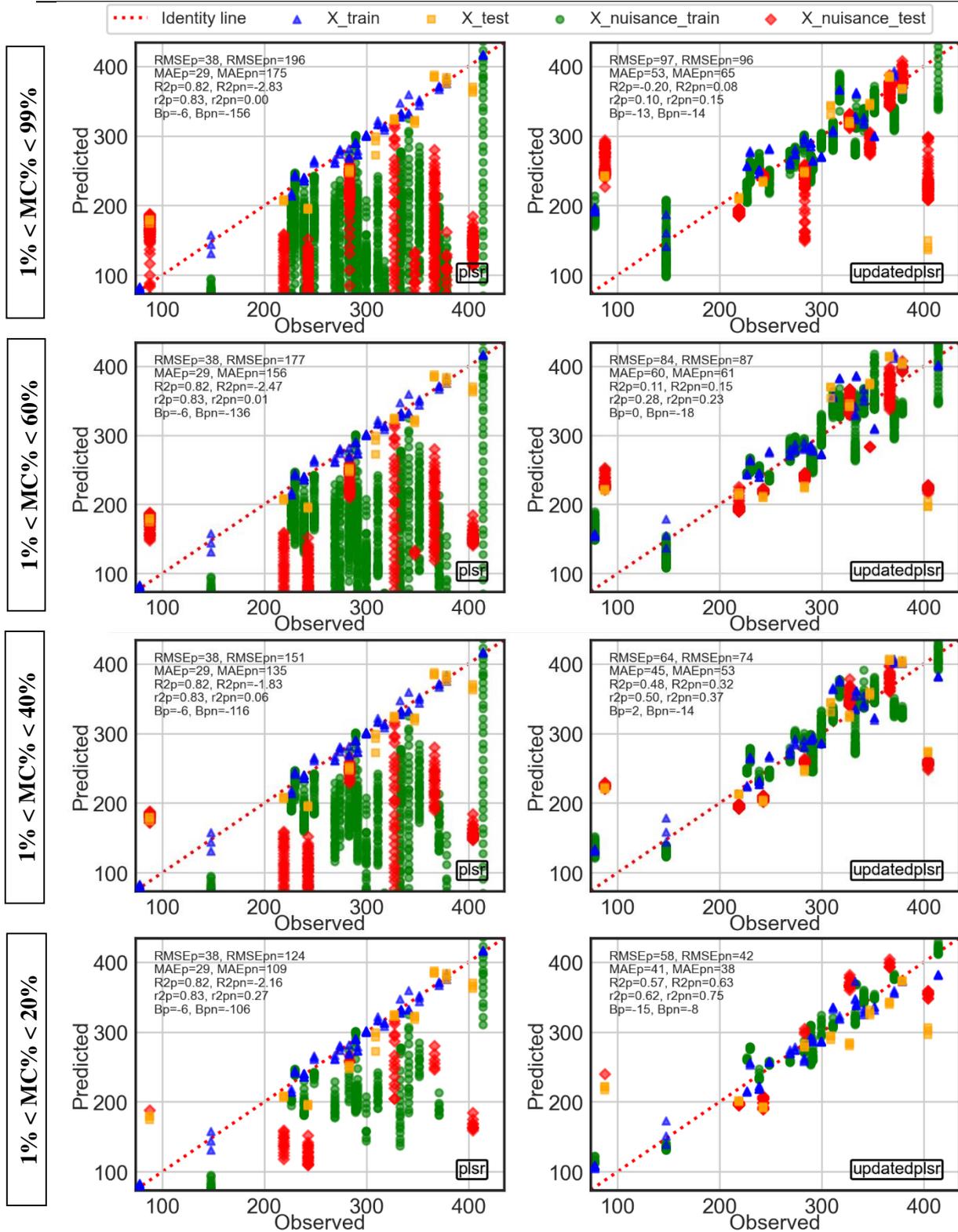


Figure 22. Predictions vs observed values for models (PLS-R or updated PLS-R) built on the group with homogeneous moisture content effects, with four moisture content ranges considered: MC < 99% (first line), MC < 60% (second line), MC < 40% (third line), MC < 20% (fourth line). The dry calibration train set is plot in blue, the dry calibration test set is plot in orange, the nuisance train set is plot in green, the nuisance test set is plot in red.

Chapter VI

plot in red. The statistics (RMSE, MAE, R2, r2, Bias) are provided for the dry calibration test set (p) and the wet nuisance test set (pn).

Different conclusions can be drawn from these results.

- A first reason that explains the limited performance of the models when built on the full range of moisture content may be that **the power relationship between the NIR spectra and moisture content cannot be well modeled by the PLS**, even by adding multiple components. In such a case, it seems that new calibration strategies could be developed by leveraging the fact that the moisture content effects are well modeled by a power relationship. For example, log transformation of both spectra and reference characteristics could be done to linearize the relationship (as proposed in Chapter V). However, in order to do this, the differences in global pseudo-absorbance levels between each substrates would first need to be removed. While a Savitzky-Golay second derivation could well remove these additive baselines, it transforms too much the spectra, with negative parts (which does not go well with the log transformation). Other strategies could involve the use of EMSC (as done in Chapter 5), but it requires to know the dry spectra in order to integrate it in the model.
- A second reason may be that **the subset of substrates still does not have a homogeneous behavior regarding water**. Indeed, while the evolutions of the first principal component scores are similar (power-like evolutions), score evolutions in other dimensions are still too different and cannot be taken into account by linear modeling approaches. This would require to further refine the local group. However, by reducing the number of substrates within the local group, the variability of the reference characteristic within this local group may end up being too low for the model to be built.
- Another reason may be that **the footprint of biochemical components is not sufficiently present in the spectra when moisture content levels are too high**. The spectral variations related to moisture content variations are simply too important compared to spectral variations related to biochemical methane potential differences. In such case, results presented in Figure 22 suggest that limiting to a range of moisture content below 40% can already provide satisfactory results.

3. Leveraging the model of moisture content effects on scattering to define classes of substrates with homogeneous effects

3.1 Introduction

As presented in the previous section, models built on a local subset of substrates which appeared to be affected in the same way by water content variations (*e.g.*, global variations are related to moisture content variations by a power function) still show limited performances. One of the reason evoked may be that the evolutions of the spectral contribution of dry matter constituents still differs too much between substrates. In other words, the local subset chosen should be further refined. To better qualify these differences, it is proposed to use multivariate curve resolution alternating least squares (MCR-ALS)³²⁰ to identify the spectroscopically-meaningful dry matter constituent contributions. In particular, the knowledge on the power-type relationship moisture content effect that has been proposed in Chapter V can be integrated in the analysis thanks to a hard-modeling constraint on the concentration profiles. This should enable a better resolution of the mixtures, and therefore a better knowledge of substrate behaviours towards moisture content.

3.2 Materials & Methods

Only some aspects of the MCR-ALS methodology are provided here for brevity, please refer to previous references for more details³²⁰. The method is a soft-modeling technique based on an alternating least squares (ALS) algorithm which iteratively calculates the concentrations \mathbf{C} and the pure spectra \mathbf{S}^T matrices by solving the following equation (Eq. 20):

$$\mathbf{X} = \mathbf{C}\mathbf{S}^T + \mathbf{E} \quad (\text{Eq. 20})$$

The procedure consists of the following steps:

1. Definition of the number of components.
2. Initialization of pure spectra \mathbf{S}^T (or concentration profiles \mathbf{C}).
3. Application of a constrained least squares regression to determine \mathbf{C} knowing \mathbf{S}^T and \mathbf{X} .
4. Application of a constrained least squares regression to determine \mathbf{S}^T knowing \mathbf{C} and \mathbf{X} .

5. Computation of \mathbf{CS}^T , and calculation of a reconstruction error. Repetition of steps 3, 4 and 5 until convergence of the reconstruction error.

MCR-ALS is a highly flexible procedure to solve mixture systems thanks to the constraints that can be added in steps 3 and 4. This allows to leverage the knowledge we have on the system to better resolve the observed mixture data. Constraints can be applied both to concentration profiles and pure spectra (and to all or part of them). Such constraints include non-negativity, smoothness³²¹, symmetry, unimodality^{322,323}, closure (mass balance condition), sparseness³²⁴, relationships with metavariables and/or between components (correlation constraints³²⁵, kinetic models^{326–329}), local rank/selectivity³³⁰ and multiset³³¹.

In this study, MCR-ALS is applied to the drying experiments of 2 substrates that are part of the group which was evaluated in section 2: a ligno-cellulosic substrate (wheat chaff [chaff is the envelope of cereal grains] mixed with water), a ligno-cellulosic substrate with soluble sugars (apple mixed with water).

While a variety of methods can be used to determine the number of components, here, we reasoned this choice by analyzing PCA scores and based on the interpretability of solutions. Initial estimates of the spectra (\mathbf{S}^T) were found thanks to the SIMPLISMA algorithm which finds most distinct spectra within \mathbf{X} as first approximates of pure spectra. Regarding the constraints used in this study, a non-negativity constraint was applied on both concentrations and spectra, a closure constraint (sum-to-one) was applied on concentrations, a unimodality constraint was applied on concentration profiles, and most importantly, a hard-modeling constraint was applied on the first component to fit a power model type as in (Eq. 19). This hard-modeling constraint was applied using a Levenberg-Marquardt fitting procedure³¹⁸.

Different metrics were calculated to evaluate the reconstruction error of the final solution (and to assess the convergence of the procedure), based on x_{ij} the elements of the observed data \mathbf{X} and x_{ij}^* the elements of the reconstructed data \mathbf{CS}^T :

$$\mathbf{Mean\ squared\ error\ (MSE)} = \frac{\sum_{ij} (x_{ij} - x_{ij}^*)^2}{ij} \quad (\text{Eq. 21})$$

$$\mathbf{Frobenius\ norm\ ratio} = \frac{\sum_{ij} x_{ij}^{*2}}{\sum_{ij} x_{ij}^2} \quad (\text{Eq. 22})$$

$$\text{Lack of fit (LOF)\%} = 100 \times \sqrt{\frac{\sum_{i,j} (x_{i,j} - x_{i,j}^*)^2}{\sum_{i,j} x_{i,j}^2}} \quad (\text{Eq. 23})$$

$$\text{Explained variance \%} = 100 \times \left(1 - \frac{\sum_{i,j} (x_{i,j} - x_{i,j}^*)^2}{\sum_{i,j} x_{i,j}^2} \right) \quad (\text{Eq. 24})$$

3.3 Results & Discussion

A ligno-cellulosic substrate (wheat chaff)

Figure 23 presents the results of the MCR-ALS decomposition applied to the drying of a ligno-cellulosic substrate (wheat chaff mixed with water). An optimal solution was found with three components and with an explained variance of 99.78% and low lack of fit of 1.89%. This adds further evidence to the validity of the proposed law. By applying the power function hard-modeling on C_1 (represented in blue in the left middle subplot), it appears that the procedure allowed to correctly identify a pure water spectrum S_1 (blue curve in right middle subplot). The second component (in orange) represents the absorption related to the dry matter constituents (as shown by the different peaks of S_2 at 1750 nm, 2090 nm, and 2330 nm). The concentration profile shows a non-linear behaviour, especially with a sharp drop when moisture content level exceeds about 65%. Indeed, the third component (in green) accounts for these different variations that can be found at moisture content levels above 65%. The pure spectrum S_3 appears to present only two peaks in the OH absorption regions. One possible interpretation can be found in the measurement mode: above 65% of moisture content, the wheat chaff floats above the liquid, which means most of the measured volume consists of pure water. This would lead to an increased transmission (*i.e.*, higher loss of photons). From these observations, it appears clear that the spectra from these two moisture content ranges (below and above 65%) cannot be modeled well by a PLS if put in the same subset.

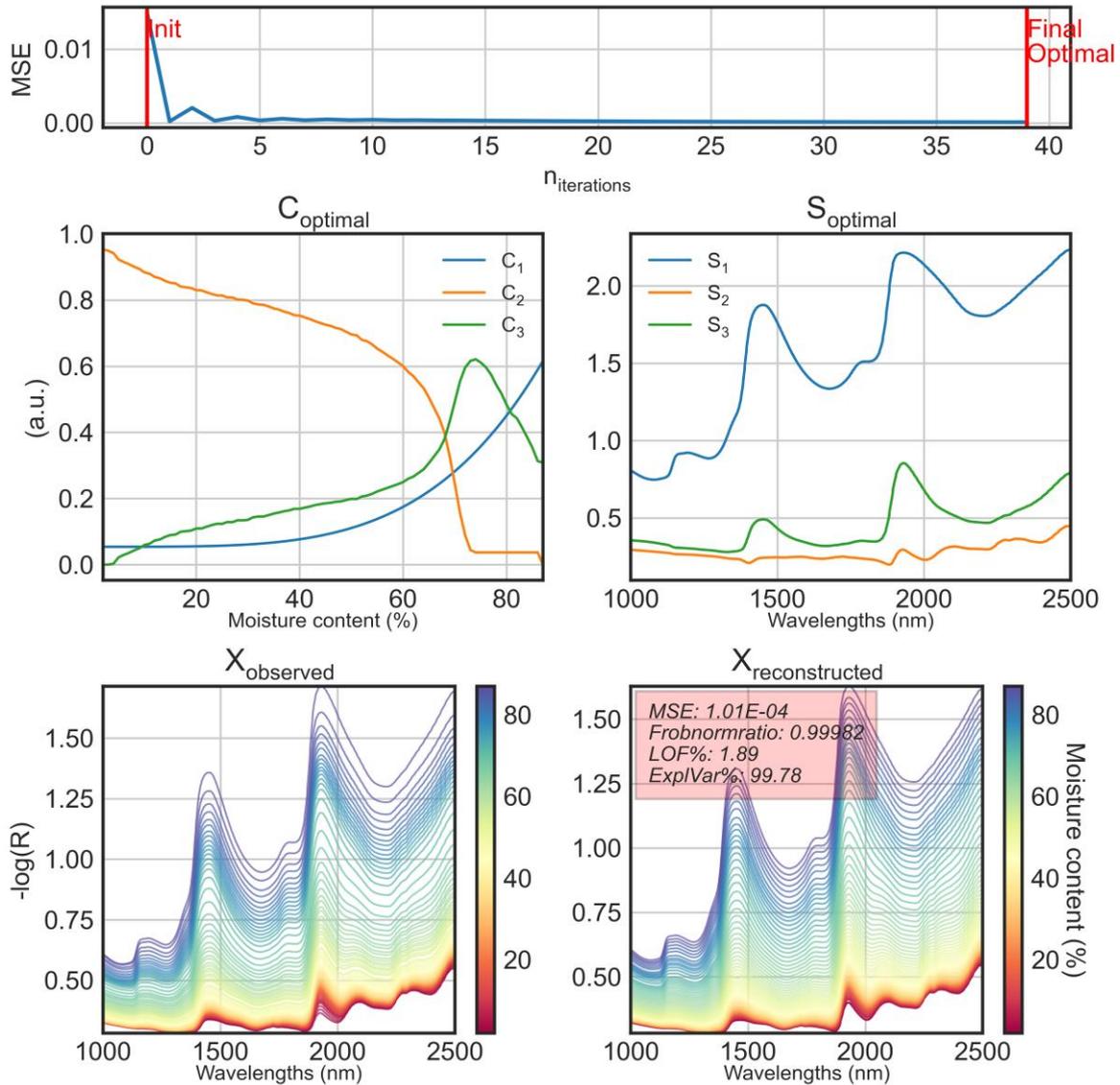


Figure 23. MCR-ALS results for the drying of ligno-cellulosic substrate (wheat chaff). The top subplot corresponds to the evolution of the reconstruction error (MSE) over the 40 iterations of MCR-ALS. The middle subplots corresponds to the concentration profiles (C_{optimal}) and the pure spectra (S_{optimal}) of the optimal solution found by MCR-ALS. The lower subplots corresponds to the observed raw spectra (X_{observed}) and the reconstructed spectra from MCR-ALS decomposition ($X_{\text{reconstructed}}$). The red box presents the different errors of reconstruction (MSE, Frobenius Norm Ratio, Lack of Fit % (LOF%), and Explained Variance %).

A ligno-cellulosic substrate with soluble sugars (apple)

Figure 24 presents the results of the MCR-ALS decomposition applied to the drying of a ligno-cellulosic substrate (apple mixed with water). An optimal solution was found with three components and with an explained variance of 99.33% and low lack of fit of 0.69%. The power law constraint could not be applied directly, due to a non-linearity for moisture content levels above 70%. Indeed, the third component S_3 (in green) resembles the most to the pure water spectrum, and while the concentration profile could be well modeled by a power law until 70% of moisture content, a distinct spectral variation

Chapter VI

represented by the first component S_1 (in blue) is found above 70%. A flat region above 1880 nm (with a negative peak at the OH absorption band maximum at 1930 nm), which can be interpreted as a saturation of the signal. This type of flattening can be seen in various other substrates which are highly transmitting (ketchup, lactulose, sucrose, oil).

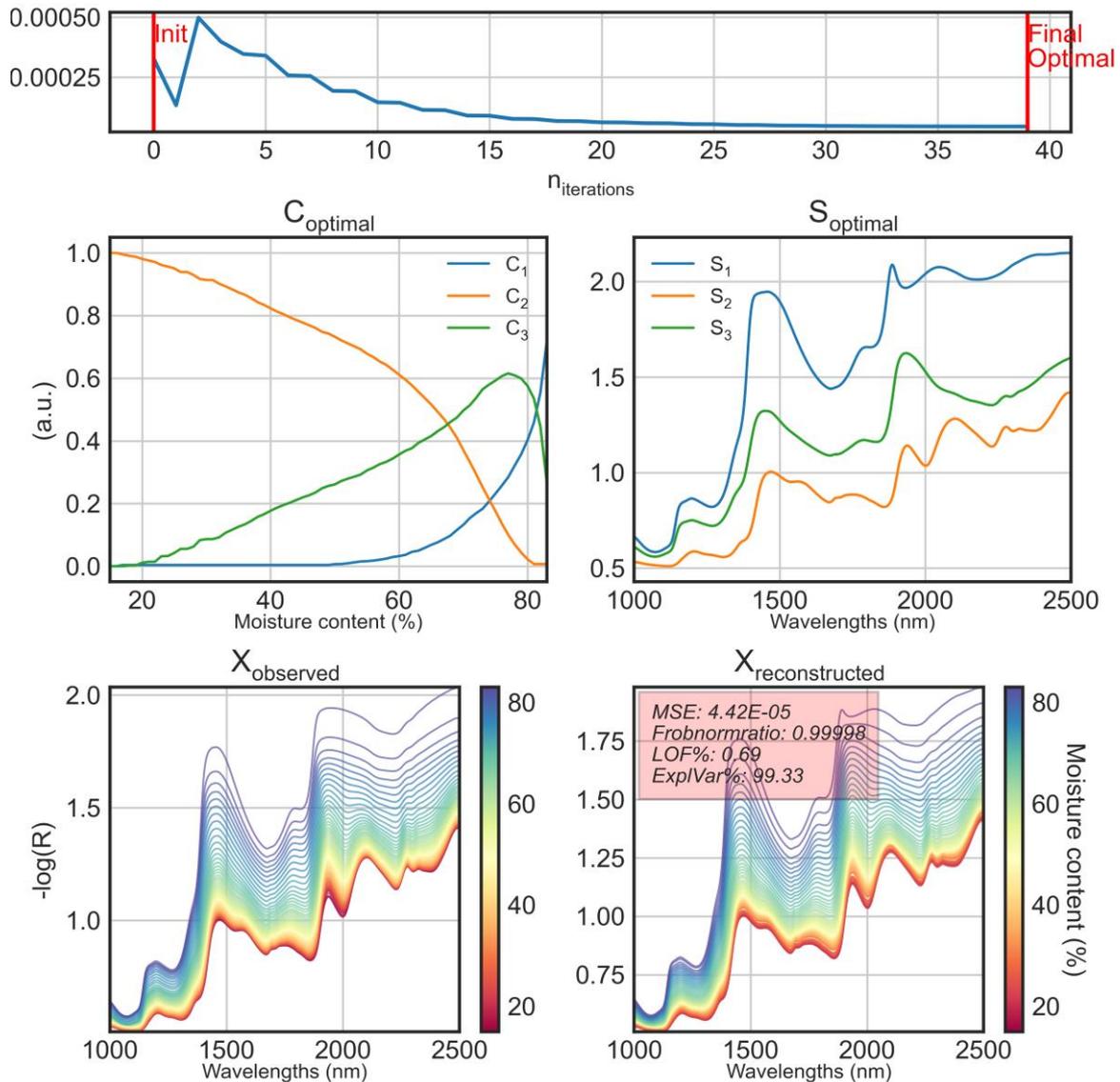


Figure 24. MCR-ALS results for the drying of substrate with soluble carbohydrates (apple). The top subplot corresponds to the evolution of the reconstruction error (MSE) over the 40 iterations of MCR-ALS. The middle subplots corresponds to the concentration profiles (C_{optimal}) and the pure spectra (S_{optimal}) of the optimal solution found by MCR-ALS. The lower subplots corresponds to the observed raw spectra (X_{observed}) and the reconstructed spectra from MCR-ALS decomposition ($X_{\text{reconstructed}}$). The red box presents the different errors of reconstruction (MSE, Frobenius Norm Ratio, Lack of Fit % (LOF%), and Explained Variance %).

In these two examples, we have shown how MCR-ALS could be used to refine the moisture content regions in which moisture content effects are homogeneous. Further steps could include the use of a

multiset structure, where all the substrates are added together, but constraints are applied per set (per substrate), and most importantly, the absence or presence of each of the components can be defined for each substrate (local rank/selectivity constraints). This should make the curve resolution more robust (reduction of the rotational ambiguity)³³¹.

4. Concluding remarks

In the first part of this chapter, it was shown that the new absorption law framework has allowed to better identify a group of substrates with common water influence. However, the local model built on this group does not show enhanced prediction performances. Only the reduction of moisture content range led to the obtention of significantly better prediction performances (up to a low error of 42 mL(CH₄),gTS⁻¹). Three different hypothesis were proposed regarding this: firstly that the non-linear effects were not well modeled by the PLS method, secondly that the local subset did not correspond to sufficiently homogeneous behaviours to moisture content variations; and thirdly, that moisture content effects were too strong when taking the full range of moisture content and masked any spectral differences between substrates of different biochemical types.

In the second part of this chapter, it was shown how the use of the new absorption law framework presented in Chapter V could provide a better resolution of the observed mixtures . Using MCR-ALS and a power law constraint, it was shown that the spectral components related to dry matter (which are of interest for quantification) could be better identified. This brings promises in better identifying the substrates behaviours to moisture content effects, and better defining subsets on which to build predictive models.

Chapter VII. General concluding remarks

1. Summary of results

Today, a great diversity of organic waste are valorized thanks to different bioprocesses such as anaerobic digestion. These organic waste are often co-digested with improved performances (both in yield and in stability). However, the wide range of biochemical and physical properties of these substrates brings important challenges for the optimization and control of the bioprocess. A continuous characterization of the feeding substrates appears as a necessary component of the monitoring system. In this context, the use of near infrared spectroscopy has been established as the most mature and cost-effective technology for characterizing these organic waste. However, for this technology to be more largely adopted (both by waste suppliers and energy producers), new developments need to be made regarding robustness issues. In particular, ways to overcome complex water effects on NIRS need to be found so as to apply the technology on raw materials directly on-site or online.

In this thesis, different developments were made to answer the two following scientific objectives: **(i) developing a better understanding of the moisture content effects on NIRS applied to a wide range of organic materials**, and **(ii) finding new ways of building models that are robust to moisture content effects**.

To answer the first objective, a unique and comprehensive experimental set-up was built with dynamic acquisition of near infrared spectra during the drying process of various substrates (Chapter II). This set-up allowed to capture the NIR spectral variations related to moisture content variations. Using chemometrics, the moisture content effects were shown to be highly complex, with both physical and chemical effects (Chapter IV). Chemically-speaking, the OH bonds of water are shown to strongly absorb in the NIR region masking the signal of other OH/CH/NH bonds present in organic matter. But depending on the number and type of hydrogen bonds made with other water molecules or biomolecules (proteins), this absorption pattern can also be different. Physically-speaking, water changes the physical state of the material (liquid/solid, crystallinity/amorphousity) which modifies the scattering behavior of light. For all these reasons, water effects were shown to be dependent of the biochemical type, the physical properties of the substrate and the moisture content range. To further understand how moisture content affects light scattering, a simple model system (composed of aluminum pellets and water) was studied using the drying system (Chapter V). Thanks to an insignificant absorption of light by aluminum,

the effects of water on scattering could be studied independently from the absorption of dry matter constituents. A modification of the Bouguer-Beer-Lambert law, also referred as the “ideal absorption law”, was introduced, where the light path-length is directly related to moisture content with a simple power function. This new modeling framework for wet scattering media holds great promises for spectroscopic studies on water and was shown to allow building more robust and accurate models for the prediction of dry matter content.

To answer the second objective, an evaluation of global correction methods was proposed. A variety of methods including skewing, orthogonal projection, feature selection, or data augmentation, were evaluated to correct existing models working on dry samples to be insensitive to moisture content effects (Chapter III). Results showed that several techniques allowed to successfully robustify the models (*i.e.*, moisture content effects could be reduced up to ten-fold). However, in the context of biochemical methane prediction on diverse organic waste, the final prediction accuracy was not satisfactory (30% less compared to models on dried materials). These limited performances can be well explained by the complexity of water effects that was revealed with the drying experiment. For these reasons, it appears that adopting a local approach can enable a reduction of this complexity, and allow to build linear models within classes of same substrate types and moisture content ranges. This approach is demonstrated on a subset of samples which have a similar behavior with moisture content variations (Chapter VI). Indeed, the effects of moisture content on all these substrates appear to be well modeled by a power relationship as observed on the mixture of aluminum and water. Successful models could be built on this subset when limiting the moisture content range below 40%. However, it appeared that this subset could be further refined. In particular, the way dry matter constituents’ absorption signatures evolve along moisture content levels differs between substrates. To better understand how the dry matter constituents’ signatures evolve when moisture content varies, multivariate curve resolution (MCR-ALS) was proposed as a tool to model moisture content effects. Within this framework, different constraints based on the knowledge of moisture content effects (such as the power-type relationship) could be included.

2. Topics for future research

A variety of research perspectives can be outlined from this work:

- The proposed **adaptation of the Bouguer-Beer-Lambert for wet scattering media should be further investigated in more complex systems**. In particular, it can be anticipated that this model does not only apply to water, but applies successfully to any other solvents. More

importantly, how does this model apply to systems where the dry matter constituents absorb? How to express this absorption? As already mentioned for example, while the global variations (related to scattering) due to moisture content variations have been modeled by a power function, it appears that for highly soluble substrates like sucrose, the signal flattens and saturates; which could be modeled by logistic or Gompertz functions.

- **Investigate furthermore the knowledge-based local approach strategy.** Through the help of multivariate curve resolution, the moisture content effects can be further detailed for all substrates, and groups on which to build models can be defined accordingly.
- **Evaluate the potential of non-linear modeling strategies such as local PLS, SVM, regression trees, random forests or CNN.** Of course, the fact that non-linear moisture content behaviours have been outlined suggests that non-linear modeling strategies could be more suitable. These have great potential for complex datasets composed of diverse organic materials and a wide range of moisture content. In particular, the calibration process can be more easily automatized. However, as already introduced, these methods have different drawbacks such as the necessary high number of samples, the high risks of overfitting, and the lack of interpretability of the obtained models.
- **Study the possibility of using the NIRS measurements of a substrate during drying as one single observation on which to predict.** Since the spectral evolutions during drying have been shown to be specific to the substrate type (physical and biochemical characteristics), it is expected that by using all the spectra during drying as predictors, more accurate models could be built. In terms of modeling, this implies evaluating three-way methods such as N-PLS, PARAFAC or multi-way MCR-ALS. Of course, getting the spectra during the full drying could be difficult in operational conditions due to the drying time (1-3 days), but even the beginning of drying could be used as a predictor instead of one individual spectrum.
- While this thesis focused on the modeling approach to tackle moisture content effects, investigating both **sample preparation methods** and the **measurement configuration** are certainly complementary pathways that could be leveraged to develop effective online and at-site NIRS applications on raw organic waste.
- Further developments should be done on **the use of low-cost and handheld spectrometers applied to wet substrates.** First investigations were done in this thesis (Paper IV) on dry samples, but they should be pursued on wet samples.

Included papers

This section presents all the papers (as published) which were discussed in this thesis.

Paper I – Limitations of global correction approaches for moisture content correction in the context of organic waste characterization

Full reference not yet available (paper to be submitted).

Mallet, A.; Zennaro, B.; Charnier, C.; Latrille, É.; Bendoula, R.; Steyer, J.-P.; Roger, J.-M. Limitations of global correction approaches for moisture content correction in the context of organic waste characterization.

Paper II – Unveiling non-linear water effects in near infrared spectroscopy: A study on organic wastes during drying using chemometrics

Full reference:

Mallet, A.; Charnier, C.; Latrille, É.; Bendoula, R.; Steyer, J.-P.; Roger, J.-M. Unveiling Non-Linear Water Effects in near Infrared Spectroscopy: A Study on Organic Wastes during Drying Using Chemometrics. *Waste Manag.* **2021**, *122*, 36–48. <https://doi.org/10.1016/j.wasman.2020.12.019>.



Unveiling non-linear water effects in near infrared spectroscopy: A study on organic wastes during drying using chemometrics



Alexandre Mallet^{a,b,c,d,*}, Cyrille Charnier^c, Éric Latrille^{a,d}, Ryad Bendoula^b, Jean-Philippe Steyer^a, Jean-Michel Roger^{b,d}

^aINRAE, Univ Montpellier, LBE, 102 Av des Etangs, Narbonne F-11100, France

^bINRAE, UMR ITAP, Montpellier University, Montpellier, France

^cBIOENTECH Company, F-11100 Narbonne, France

^dChemHouse Research Group, Montpellier, France

ARTICLE INFO

Article history:

Received 5 October 2020

Revised 24 November 2020

Accepted 12 December 2020

Keywords:

Near infrared spectroscopy

Chemometrics

Robustness

Water effects

Drying

Organic wastes

ABSTRACT

In the context of organic waste management, near infrared spectroscopy (NIRS) is being used to offer a fast, non-destructive, and cost-effective characterization system. However, cumbersome freeze-drying steps of the samples are required to avoid water's interference on near infrared spectra. In order to better understand these effects, spectral variations induced by dry matter content variations were obtained for a wide variety of organic substrates. This was made possible by the development of a customized near infrared acquisition system with dynamic highly-resolved simultaneous scanning of near infrared spectra and estimation of dry matter content during a drying process at ambient temperature. Using principal components analysis, the complex water effects on near infrared spectra are detailed. Water effects are shown to be a combination of both physical and chemical effects, and depend on both the characteristics of the samples (biochemical type and physical structure) and the moisture content level. This results in a non-linear relationship between the measured signal and the analytical characteristic of interest. A typology of substrates with respect to these water effects is provided and could further be efficiently used as a basis for the development of local quantitative calibration models and correction methods accounting for these water effects.

© 2020 Elsevier Ltd. All rights reserved.

1. Introduction

A growing number of solid organic waste treatment processes such as anaerobic digestion, composting or pyrogaseification are currently being developed and industrialized. Usually, organic wastes cover a wide range of physical characteristics and biochemical compositions, making substrate characterization a key issue in optimizing any of these processes. Recently, near infrared spectroscopy (NIRS) has been used to offer a fast, non-destructive, and cost-effective waste characterization system in the anaerobic digestion context (Charnier et al., 2016; Fitamo et al., 2017; Godin et al., 2015; Lesteur et al., 2011; Mayer et al., 2013; Mortreuil et al., 2018) and composting context (Albrecht et al., 2008; Galvez-Sola et al., 2010; Vergnoux et al., 2009). However, a freeze-drying step is always required, due to strong interferences in the near infrared region related to the presence of water in the substrates (Lobell and Asner, 2002; Williams, 2009). Not only is this

drying step cumbersome and impedes any online application, but the volatilization process that takes place during drying makes some characteristics (volatile fatty acids) impossible to predict directly. Though some applications have been developed for the characterization of liquid samples with the presence of water, these are usually restricted to a limited moisture content range, as well as one substrate type (Jacobi et al., 2009; Stockl and Lichti, 2018). In fact, near infrared spectroscopy is sensitive to numerous factors including the spectrometer lamp temperature (Sánchez et al., 2003), sample presentation (Sørensen et al., 2014), light penetration depth (Padalkar and Pleshko, 2015), sample particle size distribution (Igne et al., 2014), sample temperature (Sánchez et al., 2003), and moisture content (Lobell and Asner, 2002). All these interfering factors need to be accounted for in order to build robust quantitative calibrations (Acharya et al., 2014; Zeaiter et al., 2004). Furthermore, these factors may interact together, leading to more complexity for their correction. Indeed, for example, a close relationship between moisture effects and temperature has been outlined (Renati et al., 2019; Wenz, 2018), leading to account for both factors in conjunction (Hans et al., 2019).

* Corresponding author.

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

The effect of moisture content on near infrared spectra has been described for a wide variety of different matter types including soil (Bogrekci and Lee, 2006; Bowers and Hanks, 1965; Chang et al., 2005; Knadel et al., 2014; Lobell and Asner, 2002; Sudduth and Hummel, 1993; Wu et al., 2009), crops (Gaines and Windham, 1998; Gergely and Salgó, 2003; Peiris et al., 2016; Popineau et al., 2005; Williams, 2009), food (Büning-Pfaue, 2003), plants (Carter, 1991), wood (Giordanengo et al., 2008), pharmaceuticals (Igne et al., 2014), object models (Reeves, 1995, 1994; Wenz, 2018), and water-dominated systems (Muncan and Tsenkova, 2019). In addition, though not focused on the analysis of moisture content effects in NIRS, some studies use NIRS to monitor drying or hydration processes where moisture content varies (Caponigro et al., 2018; Raponi et al., 2017). However, no study has yet analyzed and compared moisture content effects in one comprehensive experiment with a wide variety of biochemical and physical types. Better understanding water effects and how they relate to the substrate properties appears as key for the development of robust calibrations models on wet substrates. Indeed, groups could then be used for building local models, an approach which has been shown to be successful for biochemical methane potential (BMP) prediction on plant biomasses (Godin et al., 2015).

The main effect of moisture content variations on NIR spectra usually put forward in studies relates to the apparition of three broad OH absorbance bands (detailed further on); but one major effect of water relates to physical effects (ie. changes in scattering). This is why, when speaking about water effects, an important aspect to have in mind concerns the measurement mode. For transparent liquid samples such as pure water or clear suspensions, transmission or transflexion mode is usually preferred (Pasquini, 2003), while for solid samples like powders, diffuse reflection appears most suitable. When studying large moisture content variations, one substrate may cover various states from a clear suspension, to a sludge-type material, to a powder when fully dried.

Because near infrared spectra contain both physical information (such as granulometry) and chemical information (compound concentration of interest), a pre-processing step is commonly used to maximize the chemical information in the spectra. This is done by getting rid of baseline effects due to scattering (referred to additive and multiplicative effects), as well as using spectral derivation to deconvolve the peaks. A wide variety of pre-processing techniques are used (Rinnan et al., 2009; Zeaiter and Rutledge, 2009), sometimes even in combination (Roger et al., 2020). However, these pre-processing steps may bring important artefacts (Rabatel et al., 2019) in the spectra when applied inappropriately. As well, some pre-processing steps such as derivation may deport the chemical information on shifted peak positions which can make the assignment of bands more complicated (Oliveri et al., 2019). Nevertheless, such pre-processing steps will most likely remain necessary when building quantitative models.

In the context of highly diverse matter types, water effects are expected to vary at least according to the biochemical characteristics. Exploring such differences in effects is the aim of this article. A customized air-drying system was built, allowing the simultaneous monitoring of samples' moisture content and acquisition of near infrared spectra during drying. Using this system, spectral variations related to moisture content variations were obtained for a large variety of substrates. A principal component analysis was used to explore the various effects. The aim of this global PCA was to identify major groups of substrates in regards to water effects. This was done by analyzing the scores' kinetics of each substrate during drying in relation with the interpretation of each component loadings using band assignments (Williams and Antoniszyn, 2019; Workman and Weyer, 2012). Because the aim of the study was to explore the water effects, including baseline

modifications related to scattering effects, data analysis was done on the raw spectra, without any prior pre-processing steps.

2. Materials and methods

2.1. Sample preparation

The study was conducted on $c = 89$ substrates chosen to represent a wide range of organic wastes with different chemical compositions: fruits (banana, apple), vegetables (carrots, onions, salads, potato), farm wastes (manure, silage, soya meal, grass), dairy products (cream, yoghurt, butter), meat products (beef, grilled/fresh meat, fish), as well as food industry materials (sugar, sauces, fried potatoes, wheat flour). In order to provide control samples with simplified water effects due to limited water chemical interactions, a selection of packaging materials were also measured (wood, paper, aluminum, plastic). Because these packaging materials were found dry at their original state, samples were wetted artificially by adding water at the start of the experiment.

For each substrate, 50 g of fresh matter (initial mass before drying, M_0) were sampled and manually ground (to obtain a mixture with elements below 1 cm) for further drying and NIR analysis. To determine dry matter content before and after drying (respectively DM_0 and DM_f), two replicate samples of 5–10 g were weighed before and after 48 h of drying in a heat chamber at 105 °C.

2.2. Drying system

The drying system used (Fig. 1) was a customized system consisting in a closed tube loop, with an internal circulation of air generated by a peristaltic pump (Masterflex N°77521-47 6-600 RPM, with a head #7018-52) set at 500 RPM corresponding to a generation of a flow speed of 2000 ml min⁻¹. A strong desiccant (sodium hydroxide) was used to enable drying of the gas phase and therefore the substrate: indeed, sodium hydroxide allows to bring the relative humidity at about 8% at 25 °C (Greenspan, 1976). The drying circuit was connected to a hermetic spectrometer sampling cup in which the waste sample was placed. The sampling cup was placed over the spectrophotometer for continuous automatic near infrared acquisitions; and the desiccant was weighed continuously using a precision balance (Ohaus Traveler TA502), to enable the measurement of loss of water during drying. In addition, two temperature probes were installed on the system to monitor both the temperature inside the sample cup chamber and the room temperature, for investigation of temperature-induced spectral variations. Before closing the system and launching the acquisition, the circuit was flushed with nitrogen gas to limit oxidative reactions on the substrates. Using this drying system, substrates were dried during time periods varying from 12 h to 72 h.

2.3. Near infrared spectroscopic acquisition system

During the drying process, a spectrum of the sample was acquired from below every 90 s in reflectance mode over 10000–4000 cm⁻¹ (1000–2500 nm) range with a resolution of 8 cm⁻¹ (0.8–5 nm) by a BUCHI NIR-Flex N-500 solids spectrophotometer with a rotating add-on petri dish and high-performance sample cup (Buchi, Flawil, Switzerland). Each measurement consisted of an average of 96 scans acquired while rotating the sample at 360° to enhance sampling representativeness. In order to compute reflectance spectra from these measurements, an internal Spectralon® reference was scanned every 10 min. All spectra

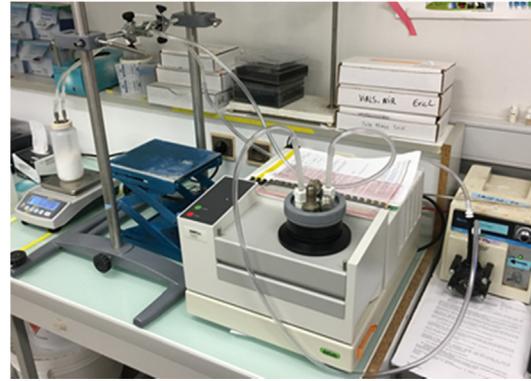
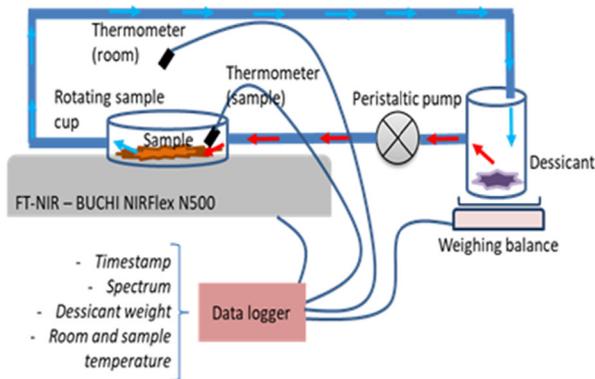


Fig. 1. Experimental set-up with: a NIRS acquisition under a quartz rotating sampling cup; a tube circuit with gas circulation, a desiccant weighed by a precision balance; the whole system is automatized and controlled by RS-232 serial connection.

were transformed into pseudo-absorbance units using log transformation:

$$\text{PseudoAbsorbance} = -\log_{10}(\text{Reflectance}) \quad (1)$$

2.4. Dry matter content estimation during the drying process

At a given time t during drying, the sample's water loss on drying $LOD_s(t)$ in g was measured by monitoring the weight of the desiccant $M_d(t)$. Using the dry matter content measured before drying DM_0 , the dry matter content of the measured sample during drying $DM_s(t)$ was estimated from :

$$LOD_s(t) = M_d(t) - M_d(t=0), \quad (2)$$

$$DM_s(t) = M_0 \times DM_0 / (M_0 - LOD_s(t = t_{final})). \quad (3)$$

As mentioned, after drying, dry matter content was measured classically (using 48 h oven-drying at 105 °C) to confirm the final obtained dry matter content given by the system.

2.5. Biochemical characterization of substrates

All the substrates were freeze-dried (using a Cosmos 20 k freeze-dryer (Cryotec, Saint-Gély-du-Fesc, France)) and ground to 1 mm (using an MF 10 basic Microfine grinder drive (IKA Works, Staufen, Germany)), to be scanned in vials by the same near infrared spectrometer. A previously calibrated model (Charnier et al., 2016) was applied to obtain carbohydrates content, lipid content, nitrogen content, chemical oxygen demand with respective obtained standard errors of prediction (RMSEP) of 53 mgO₂.gTS⁻¹, 3.2 * 10⁻² g.gTS⁻¹, 8.6 * 10⁻³ g.gTS⁻¹, 83 mgO₂.gTS⁻¹.

2.6. Chemometrics

2.6.1. Data preparation

The dataset consists of 116 000 spectra of 89 substrates covering different dry matter content ranges. To facilitate interpretation, spectra were then linearly interpolated on a common dry matter content range from 1% to 95% with a 1% step; but of course left to NaN values outside the measured dry matter content ranges. Indeed, this allowed to compare spectra of different substrates at strictly identical dry matter contents. This resulted in a matrix $X(n,p)$ with $n = 5011$ the number of spectra, and $p = 1501$ the number of wavelengths.

2.6.2. Data processing

All the data analysis was performed using Python 3.6.5: data wrangling with Pandas 0.25.1, NumPy 1.17.3, SciPy 1.3.1, principal component analysis with Scikit-learn 0.21.3, and plotting with Matplotlib 2.2.2 (Hunter, 2007; McKinney, 2010; Oliphant, 2010; Pedregosa et al., 2015; van Rossum and Drake, 2009; Virtanen et al., 2019).

A global principal component analysis (PCA using the singular value decomposition algorithm) was run with $k = 8$ components on the raw centered matrix

$$X_C = X - \frac{1}{n} J_n X, \quad (4)$$

with J_n the all-ones square matrix of size n .

This provided $T(n,k)$ matrix of scores and $P(p,k)$ matrix of loadings so that

$$X_C = TP^T + E \quad (5)$$

with E matrix of residuals.

In some cases, for a given principal component q to be analyzed, the raw spectra matrix deflated by the previous principal components was computed to further support the interpretation of loadings and scores.

$$X_{C_deflated[q]} = X_C - T_{q-1} P_{q-1}^T. \quad (6)$$

In addition, the first eigenvectors of the within-substrate and between-substrate variance-covariance matrices were computed (Roger et al., 2005). For this, a matrix C of size (n,c) was defined, containing the substrate's membership disjunctive encoding of the individuals, i.e. $y_{ij} = 1$ if the individual i belongs to the substrate j and 0 if not. Let

$$T = \frac{1}{n-1} X_C^T X_C, \quad (7)$$

be the full variance-covariance matrix,

$$B = \frac{1}{n-1} X_C^T C (C^T C)^{-1} C^T X_C, \quad (8)$$

be the between-substrates variance-covariance matrix,

$$W = T - B, \quad (9)$$

be the within-substrate variance-covariance matrix.

To evaluate autocorrelation (i.e. information content) in the signals (spectra or loadings), the Durbin-Watson statistic was used, defined as:

$$DW = \sum_{i=2}^n (r_i - r_{i-1})^2 / \sum_{i=1}^n r_i^2, \quad (10)$$

with r_i and r_{i-1} the successive values in a vector.

Let D be a matrix of size $(n,1)$ with all the estimated dry matter content (Eq. (3)) of each spectra from X ; and D_C its centered matrix version (Eq. (4)). To evaluate the zones in the spectra that are most correlated to dry matter content %, a correlation spectra was calculated, which corresponds to Pearson correlation coefficient calculated between each wavelength column of X and the dry matter content levels in D .

$$\begin{aligned} \text{CorrelationSpectra} &= \left[\frac{\text{cov}(X_1, D)}{\sigma_{X_1} \sigma_D}, \frac{\text{cov}(X_2, D)}{\sigma_{X_2} \sigma_D}, \dots, \frac{\text{cov}(X_p, D)}{\sigma_{X_p} \sigma_D} \right] \\ &= \frac{D_C^T X_C}{\text{diag}(X_C^T X_C)^{1/2} \text{diag}(D_C^T D_C)^{1/2} I_p}. \end{aligned} \tag{11}$$

3. Results & discussion

3.1. Data overview

3.1.1. Biochemical characteristics

Fig. 2 presents the predicted characteristics obtained using the near infrared spectroscopy calibrated model for freeze-dried and ground samples. Samples (detailed in Section 2.1) cover a very wide variety of biochemical types which is representative of the variety of inputs possibly used in the anaerobic digestion process, in particular in co-digestion plants. All biochemical characteristics show non-Gaussian distributions, which will impact the structure of the data. Some extreme samples will impact the variance in the spectra related to biochemical characteristics. Indeed, for example, the fat content histogram (Fig. 2) clearly highlights two populations: one population with no or very low fat content levels ($<0.2 \text{ g.gTS}^{-1}$) and another population with very high fat content levels ($>0.7 \text{ g.gTS}^{-1}$). Unfortunately, such structuring is difficult to avoid, as intermediate compositions with 0.5 g.gTS^{-1} of fat content level results in biphasic systems.

3.1.2. Dry matter content ranges

Fig. 3 presents for each substrate the range of dry matter content over which spectra were obtained. Contrarily to many studies that focused on limited dry matter content ranges (70–95%), a very wide range of dry matter content was obtained here (5–95%). However, substrates were not all measured along the same dry matter content range. Several reasons explain this including differences in the initial dry matter content (very low dry matter contents like *salad_1* or *digestate_1*, and very high dry matter contents like *butter_2*, *mayonnaise_1*), drying inefficiency related to highly bound water or intra-cellular water (*syrop_1*, *ketchup_1*, *banana_2*, *orange pulp_1*) as well as simple experimental drying interruptions due mostly to electric failures (*banana_1*, *crustbread_1*, *sunflower-meal_1*, *grass_1*, *weeds_3*). These latter samples were still kept in the dataset because they still represented useful spectral variance related to moisture content variations. Two families of substrates can already be defined from these drying behaviors: hydrophobic substrates for which low dry matter content levels are difficult to obtain but are easily dried (like *butter*, *sour cream*, *mayonnaise*), and hydrophilic substrates in which water is more difficult to extract (like *syrop*, *ketchup*, *banana*, *orange pulp*). Within hydrophilic substrates, the final moisture content to which the substrate was dried relates to numerous factors and their complex interaction such as the presence of gelling agents like pectin, or water soluble molecules like saccharides, as well as the interaction of proteins and starch controlling viscosity and swelling characteristics (Dehnad et al., 2016).

3.1.3. Experimental conditions

3.1.3.1. Dry matter content estimation validity. Validity of the dry matter content monitoring system was evaluated as illustrated in Fig. 4. Let

$$d_{\text{finalDM}} = DM_s(t = \text{final}) - DM_f, \tag{12}$$

the final dry matter content error, corresponding to the difference between the final dry matter content obtained in the experiment, and the one measured classically (using oven-drying). Fig. 4.1 and Fig. 4.3 both reveal four apparent outliers: dairy fat sludge, orange pulp, brewery yeast, and sunflower meal with respective dry matter

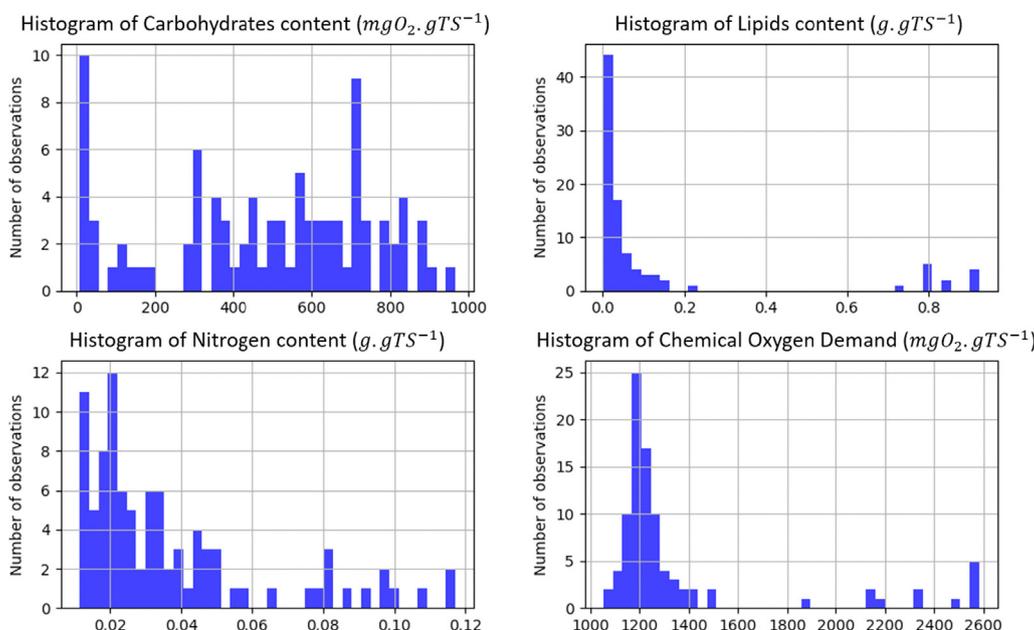


Fig. 2. Sample characteristics - Histograms of predicted characteristics: carbohydrate content ($\text{mgO}_2.\text{gTS}^{-1}$), fat content (g.gTS^{-1}), nitrogen content (g.gTS^{-1}), chemical oxygen demand ($\text{mgO}_2.\text{gTS}^{-1}$).

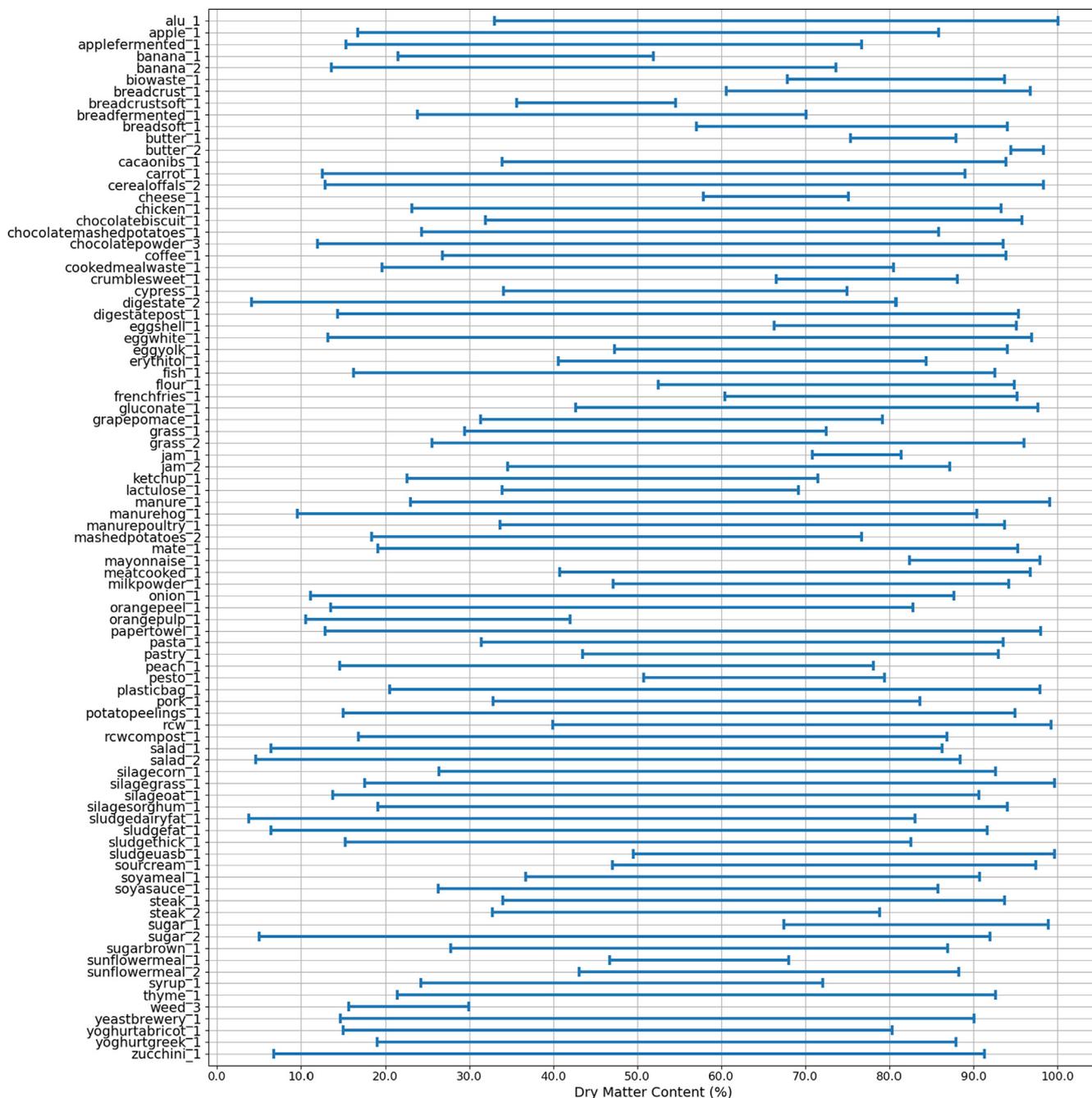


Fig. 3. Drying data: list of all samples with initial and final dry matter contents obtained in the experiment. Spectra were obtained within each of these ranges.

estimation error values of -5.81 g.g^{-1} , -6.82 g.g^{-1} , -6.98 g.g^{-1} and -7.63 g.g^{-1} . These substrates were consequently withdrawn from the dataset in the further analyses. Fig. 4.2 shows a good degree of agreement between the measured dry matter and the estimated dry matter using the system. Fig. 4.3 shows no obvious relationship between the differences and the mean which confirms homoscedasticity of the residuals. From the boxplot, it seems that the system slightly underestimates the measured dry matter content (-0.21%) with a standard deviation of $\pm 2.30\%$ (Fig. 4.1). This appears marginal compared to the large range of dry matter content studied here. However, this does imply that drawing conclusions on spectral effects due to water below 2–3% of dry matter content differences should be done carefully.

3.1.3.2. Temperature variations during drying. Similarly to how it was done on spectra (Eqs. (7)–(9)), temperature variations can be separated into two components: the temperature differences observed between each substrate drying experiment (between-substrates temperature differences), and the temperature variations occurring during drying for each substrate (within-substrates temperature differences). As shown in Fig. 5.1, the mean temperature measured for all substrates is $28.3 \text{ }^\circ\text{C}$, with a standard deviation of $1.8 \text{ }^\circ\text{C}$. Such variations in temperature between each substrate drying experiment can be explained by the daily temperature differences from one experiment to another. Though the measurements were taken in a temperature controlled room, temperature differences were still observed.

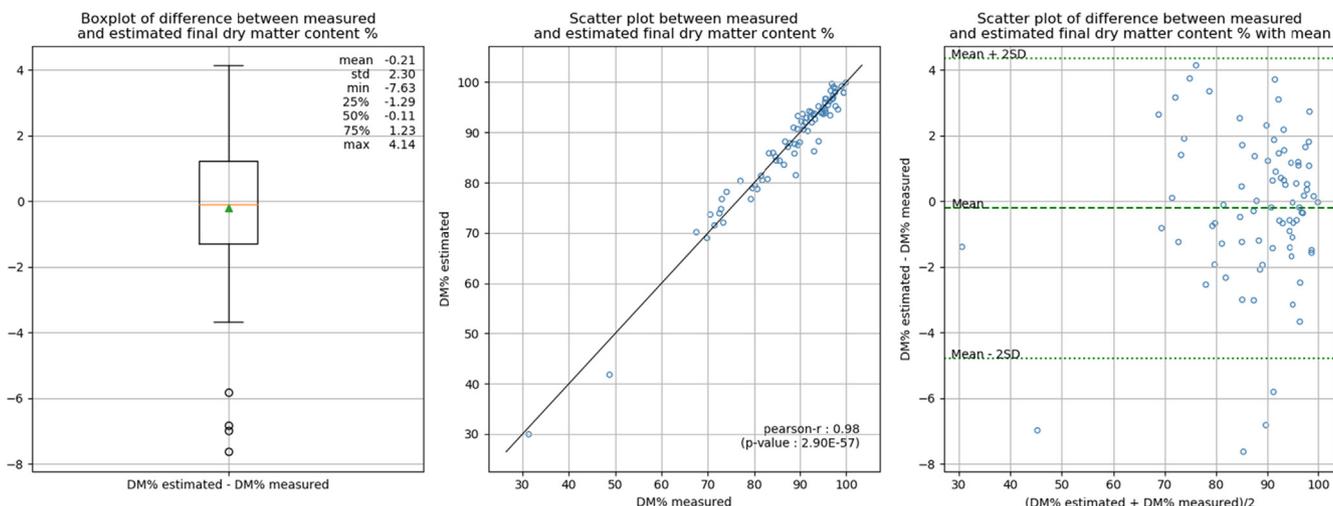


Fig. 4. Experimental conditions: (1) boxplot of final dry matter content errors; (2) measured vs. estimated final dry matter content, and (3) difference against mean.

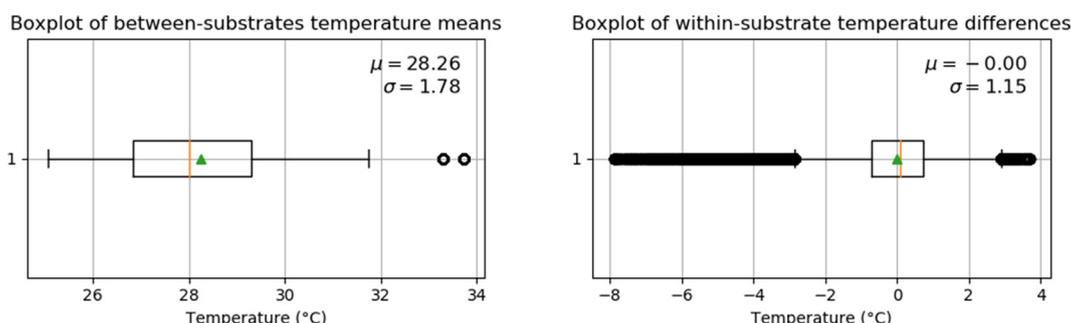


Fig. 5. Experimental conditions: boxplot of between-substrates temperature differences (1) and within-substrates temperature differences (2) during drying.

Moreover, as shown in Fig. 5.2, the standard deviation of within-substrate temperature differences observed during drying is 1.15 °C. Such sample temperature variations during drying can be explained by two factors: heating resulting from the spectrometer’s lamp, and heating resulting from the absorption of water by the desiccant.

Unfortunately, variations of temperature may have a strong impact on the acquired spectra and may lead to the alteration of quantitative calibration models as many authors have shown (Campos et al., 2018; Cozzolino et al., 2007; Dvořák et al., 2017; Golic and Walsh, 2006; Roger et al., 2003; Sun et al., 2020; Wülfert et al., 1998). Indeed, as temperature rises, proportions of molecular vibrations within each molecular vibrational energy levels change, which has a direct impact on the absorption of photons (ie. the spectra). Visually, a horizontal shift of the broad absorbance bands can be observed in the spectra (Renati et al., 2019), but in fact this relates to vertical absorption changes from the originating sub-bands. To have an idea of the magnitude of such changes, in the case of pure water at 22 °C, it has been measured that at 1410 nm (free OH water peak), a +1 °C temperature change increased the intensity of the absorbing peak by +0.8% (i.e. temperature coefficient of 0.8% °C⁻¹) (Cumming, 2013; Kou et al., 1993). However, as these authors highlighted, because scattering has little if no temperature dependence, the temperature coefficient applies exclusively on the absorption coefficient and not on the scattering coefficient. Though these changes could indeed alter the exact assignment of bands, these changes are very limited compared to the spectral variations induced by dry matter content changes.

3.1.4. Raw spectra analysis

Fig. 6 shows some examples of near infrared spectral evolutions during drying, representative of the main types of evolution observed (spectral evolutions for all other substrates are provided in Appendix C). Different effects can be observed.

Firstly, water variation modifies strongly the global pseudo-absorbance level of the spectra: these baseline shifts probably relate to scattering modifications, as reported in (Ilari et al., 1988; Isaksson and Naes, 1988). Interestingly, for suspensions, the pseudo-absorbance level increases with water content increase, while for the emulsions (cream, butter, oil) it decreases. As explained in Section 3.2.1, this can be linked to different refraction modifications according to which component replaces water along drying.

Secondly, for all substrates with intermediate and high moisture content levels (spectra in dark blue in Fig. 6 below), well-known broad absorbance features due to OH vibrations are observed in the NIR spectra around 1210 nm, 1450 nm and 1940 nm. These are attributed respectively to the combination of the first overtone of the O–H stretching and O–H bending band, the first overtone of the O–H stretching band and the combination of the O–H stretching band and O–H bending band of water (Luck, 1974; Muncan and Tsenkova, 2019).

During the drying process (spectra colored from blue to red in Fig. 6), new absorbance features in relation with chemical composition progressively appear (related to OH vibrations of sugars or fatty acids, NH vibrations of proteins, CH vibrations of alkanes, the C=C vibration of alkenes, and C=O vibrations of ketones/aldehydes) and they will be further discussed in Section 3.2. Surpris-

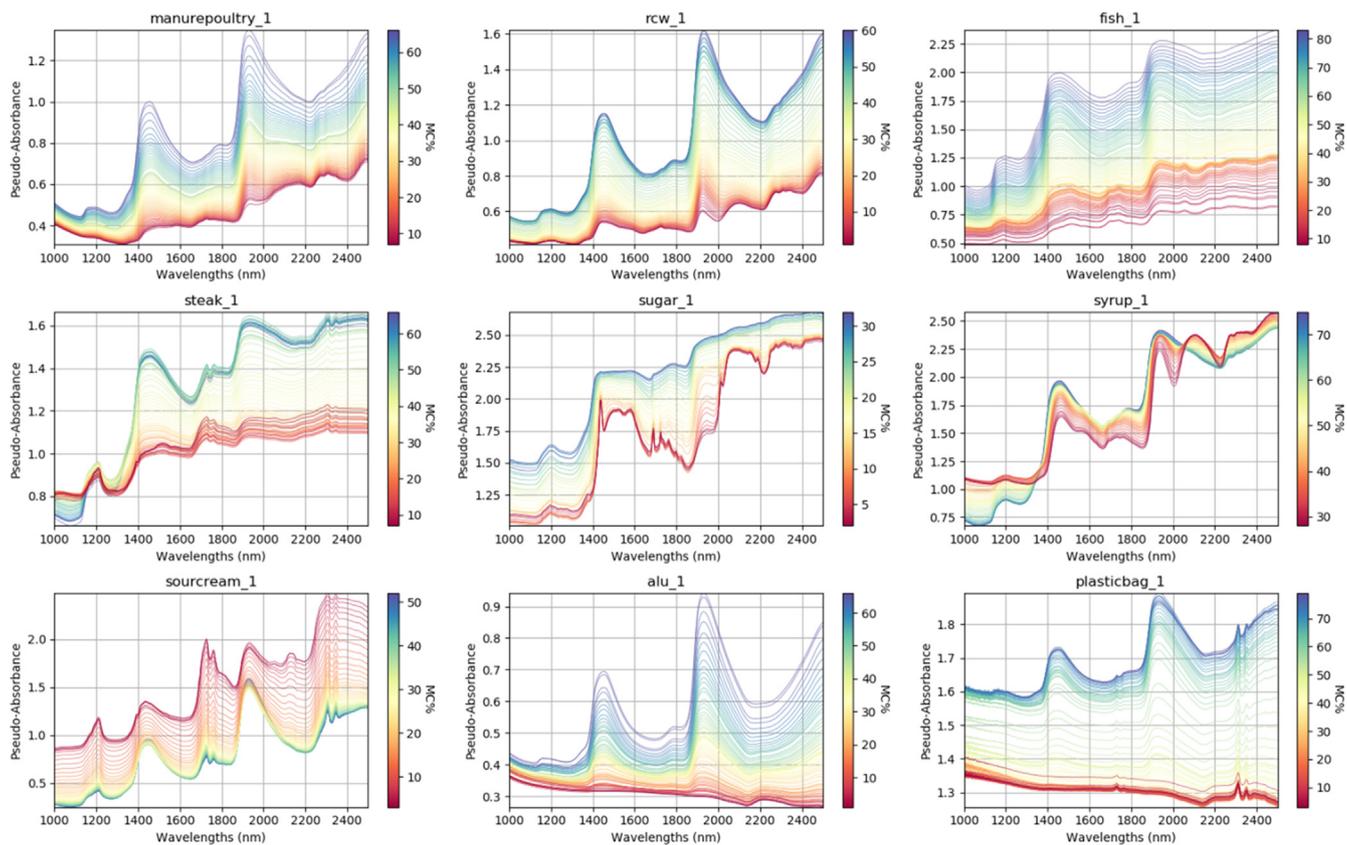


Fig. 6. Raw pseudo-absorbance spectra colored by moisture content (%) for nine substrates representative of the diversity of biochemical compositions and physical properties (poultry manure, ramial chipped wood / rcw, fish, cooked steak, sugar, syrup, sour cream, aluminum and plastic bag).

ingly, the plastic bag's dry state spectra appear flattened, but this corresponds to a scale issue: when water is present, pseudo-absorbance levels are very high (1.6–2.2), making low moisture content spectra peaks more flat, but when plotting the plastic bag spectra alone, typical peaks related to the polymeric structure of heteroatomic bonds present in plastics are well present.

3.2. Principal component analysis

The cumulative total explained variance percentage (not shown here, see Appendix A) reaches a plateau at the eighth component. Therefore, the analysis of loadings and scores (Fig. 7 and Fig. 8) was focused on these first eight components.

3.2.1. Analysis of the first component:

The first component's loadings of the PCA are fully positive, with a clear slope and no main absorbance peak can be identified (Fig. 7). This suggests that the first component corresponds to global additive variations of pseudo-absorbance level unrelated to specific spectral regions (i.e. specific chemical compound). Such observation is very common in near infrared spectroscopic data. Indeed, the first component's loadings usually resemble the mean spectrum even when data is mean-centered, and relate to light scattering differences between samples mostly due to physical differences such as granulometry (Ilari et al., 1988; Isaksson and Naes, 1988). However here, the first component's loadings do not look like the mean spectrum (that shows broad peaks at 1450 and 1940 nm) but resemble the first eigenvector of the between-substrates variance-covariance matrix (plot as a black dotted line in Fig. 7). This suggests that the first component relates to global light scattering differences observed between substrates.

In the first component's score plot (Fig. 8), substrates with high scores include *sugar*, *syrup* or *plastic bag*; and substrates with low scores include *aluminum*, *poultry manure* or *ramial chipped wood*. Indeed, the former substrates exhibit high general pseudo-absorbance levels (~2.3–2.5); while the latter substrates exhibit low general pseudo-absorbance levels (~1.5–1.8). These differences in pseudo-absorbance levels can be explained by different physical properties of the substrates: *sugar* in solution is transparent and reflects less light, *aluminum* reflects most of light. But the different chemical compositions also play a role: substrates like *sugar*, *syrup* with high contents in simple carbohydrates absorb more light than *manure* or *wood* which contain mostly complex carbohydrates like cellulose or lignin. Indeed, chemical composition and physical properties are intrinsically linked, as for example simple carbohydrates are more soluble and susceptible to form liquid transparent systems; while cellulose and lignin allow better formation of porous materials with multiple refractive interfaces. Therefore, the first component relates to global differences in pseudo-absorbance levels between substrates, both due to physical and chemical differences between substrates.

In addition, substrates show different scores' kinetics (Fig. 8). Most of the substrates (like *sugar*, *fish*, *manure* or *aluminum*) show decreasing scores along drying which means that the general pseudo-absorbance level decreases along drying. However, some substrates such as *sour cream* (but also *butter*, *mayonnaise* or *greek yoghurt* not shown here) show opposite scores' kinetics, with increasing scores along drying. These variations in the global pseudo-absorbance level along drying are due to changes in the refractive index differences between particles along drying ($n_{\text{water}} \approx 1.33$, $n_{\text{air}} \approx 1.0$, $n_{\text{organiccompounds}} \geq 1.4$) (Polyanskiy, 2008). For most substrates, as drying occurs, water is replaced by air, which leads to increased refractive index differences. As index dif-

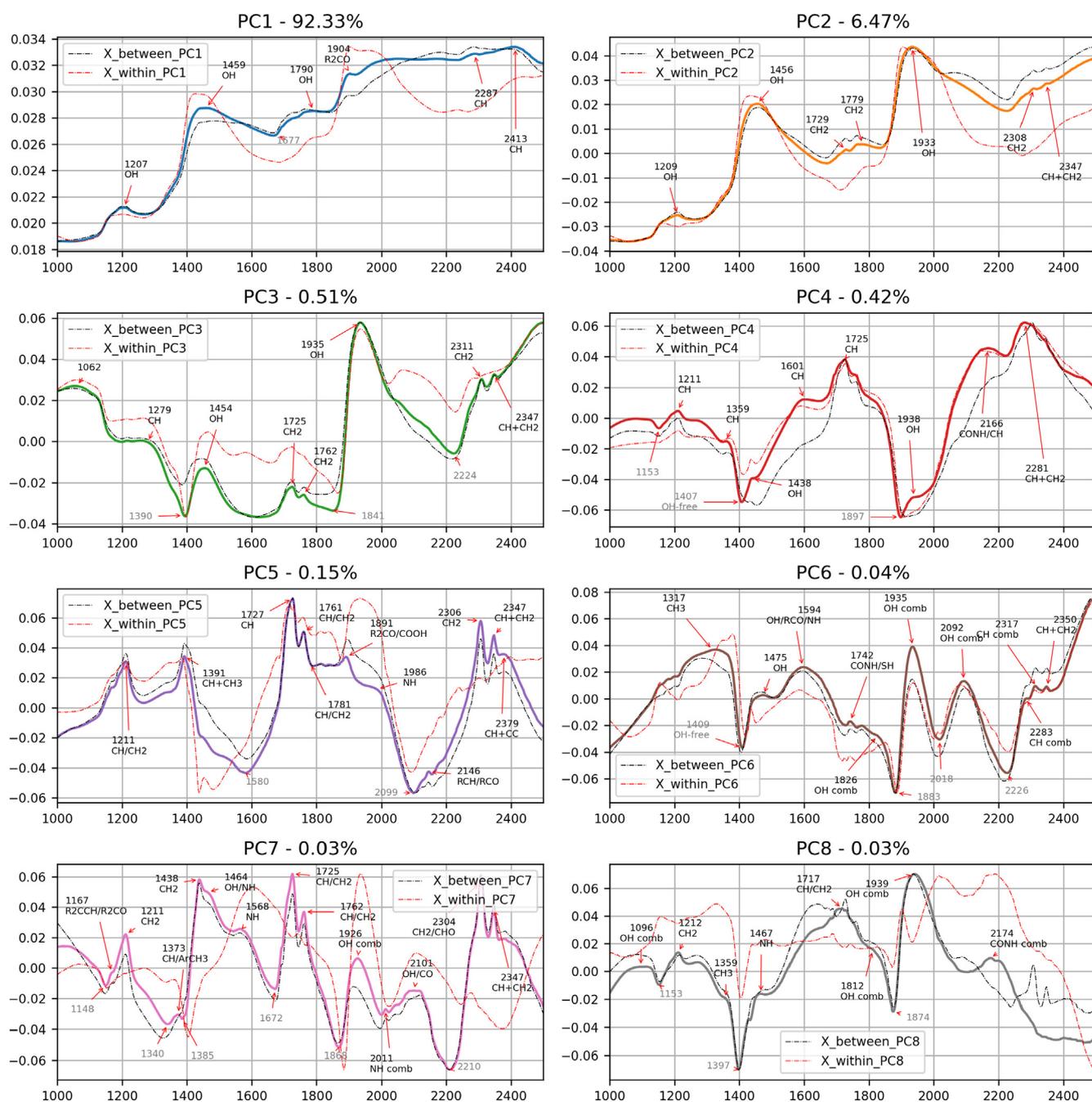


Fig. 7. Loadings from PCA of Xc (Eqs. (4) and (5)) with peak detection and chemical attributions (positive peaks annotated in black, and negative peaks in grey). Abscissa axis correspond to wavelengths (in nm). Explained variance percentage of each principal component is given in the title. For each component, the corresponding eigenvector of the between-substrate variance–covariance matrix (Eq. (8)) is plot (in dashed black line), as well as the corresponding eigenvector of the within-substrate variance–covariance matrix (Eq. (9)) (in dashed red line). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

ferences increase in number or intensity, scattering increases, leading to higher reflectance levels (ie. lower pseudo-absorbance levels). On the contrary, for substrates containing high levels of fat, water is replaced by fat and not air; leading to lower refractive index differences ($n_{vegetableoil} \approx 1.47$) (Polyanskiy, 2008), and therefore, an increasing global pseudo-absorbance level. In datasets which include these two groups of substrates, with opposite baseline evolutions in relation with moisture content, scatter correction pretreatments appear necessary.

Finally, the first component accounts for up to 93% of the total spectral variance. As determined above, the first component relates to global variations of pseudo-absorbance due to light scattering

differences between substrates, both related to their physical properties and chemical compositions. Moreover, these light scattering differences are shown to vary along drying. One of the outcomes from this result is that the main effect of moisture content variations on near infrared spectra is a physical one : global variations of pseudo-absorbance. More generally, this illustrates how much very little specific chemical-related information is present in raw near infrared spectra compared to physical-related information (Martens et al., 2003). Some authors have leveraged this observation by focusing on the baseline variations for online prediction of dry matter content instead of attempting scatter correction pretreatments as commonly done (Bogomolov et al., 2018). Though

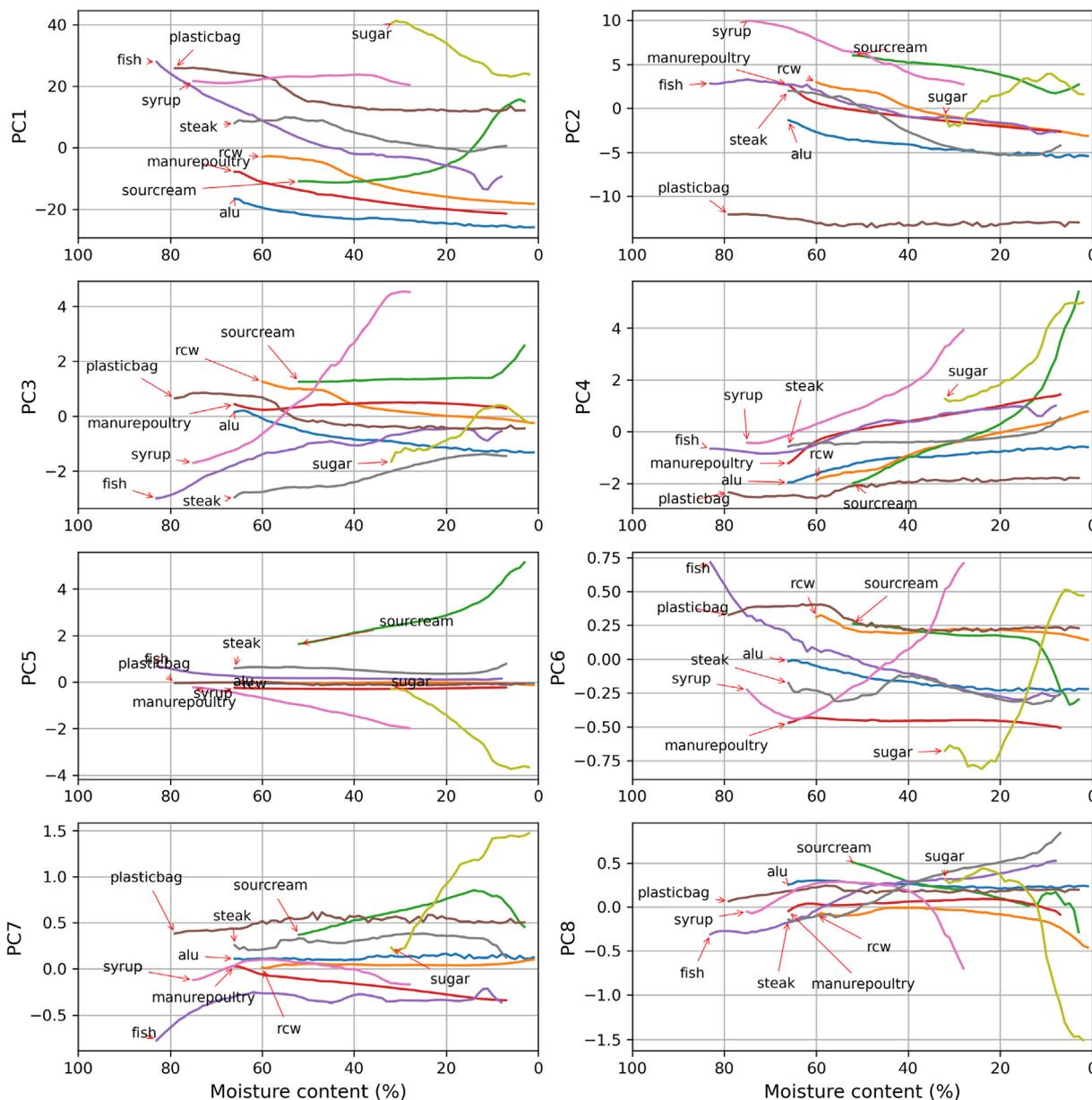


Fig. 8. Scores from PCA of Xc - All abscissa axes correspond to moisture content (%). Representative substrates were selected and plotted. See all other substrates scores in Appendix D.

this achieved promising results, it was shown here that these global levels of pseudo-absorbance are highly dependent of chemical properties; and that applying such a methodology on samples with different biochemical compositions would not be sufficient.

3.2.2. Analysis of second component:

The second component’s loadings (Fig. 7) match well with the first eigenvector of the within-substrate variance–covariance matrix (plotted as a red dotted line in Fig. 7) which suggests it relates to the main spectral variations that occur during each substrate’s drying. Three broad peaks can be found at 1209 nm, 1456 nm and 1933 nm (Fig. 7) which are attributed to pure water OH bonds’ broad absorption bands. This means that the second component relates to the varying expression of pure water spectrum during drying. In accordance, most scores (Fig. 8) show a decrease along drying, with an overall linear relationship with dry matter content. Unlike in the first component, high fat content substrates like *sourcream* also show decreasing scores along drying.

However, some samples such as *sugar* (but other samples not shown here like *lactulose*, *soya sauce* or *eggwhite*) show bell-curve-like-shape (increasing then decreasing) scores along drying. This can be explained by an excessive level of forward scattering for these substrates over certain levels of moisture content. Indeed, forward scattering is at such a high level, that the measured reflectance is similarly low for all wavelengths, and therefore, water’s OH absorbance peaks appear low. One outcome of this observation is that though a linear relationship of the pure water spectrum component with dry matter content seems valid for many substrates, this remains true only within a certain range of dry matter content and depends on the substrate’s scattering properties (ie. how ‘transparent’ the sample is in the near infrared region).

Another important characteristic of these loadings is the positive slope. During drying, not only the water’s OH absorbance bands height vary according to the moisture content, but also the general slope of the spectra is modified. As explained, most scores show a decrease along drying, which means that at high moisture

content ranges, the spectra have higher absorbance levels at high wavelengths (1700 nm and above) than at low wavelengths (below 1700 nm); and as moisture content decreases these differences are diminished. These slope modifications are again, due to changes in the physical structure of the substrates as moisture content varies.

3.2.3. Analysis of third component:

The third component's loadings show no slope and contain the same two broad peaks (as in the second component's loadings) situated at 1454 nm and 1935 nm that can be attributed to water OH absorbance bands (Fig. 7). However, their relative importance is very different: the peak at 1935 nm is much higher than the peak at 1454 nm (absolute value of 0.06 compared to 0.01). In Fig. 8, two groups of substrates can be distinguished based on the third component scores: substrates showing positive decreasing scores along drying (*aluminum, ramial chipped wood, or poultry manure*); and substrates showing negative increasing scores along drying (*syrup, sugar, steak, fish*). Referring to the upper interpretation of the loadings, this means that for the former group, as drying occurs, the peak at 1935 nm decreases relatively more than the peak at 1454 nm; while for the latter group of substrates, the peak at 1935 nm decreases relatively less than the peak at 1454 nm. Such differences in the relative expression of the two OH broad absorbance bands is related to chemical water interactions as some authors have suggested (Gorretta et al., 2019). Indeed, the latter group gathers substrates with high content levels in carbohydrates or proteins which are known to interact with water through non-covalent H-bonding (Laage et al., 2017).

3.2.4. Analysis of fourth component:

The fourth component's loadings exhibit two sharp negative peaks at 1407 and 1897 nm related to water OH absorbance bands; as well as several sharp positive peaks at 1211, 1359, 1725, 2166 and 2281 nm (Fig. 7), all of them being related to bands present in organic matter (CH/CH₂/NH). For this component, all the substrates exhibit increasing scores along drying (in particular *sugar, steak or sour cream*) (Fig. 8), with a clear linear relationship with moisture content (ie. dry matter content). To confirm this, the Pearson correlation spectra with dry matter content (Eq. (11)) was plot (see Appendix B), and the exact same shape is obtained. This implies that the bands associated to free water molecules may be formally identified here as the negative peaks in these loadings at 1407 nm and 1897 nm.

Though the majority of the fourth component's scores show an increase throughout the drying process (Fig. 8), some samples such as *plasticbag, aluminum (or digested sludge not shown here)* show almost flat score evolutions along drying. In these substrates, organic matter levels are very low, if not inexistent (for aluminum). As near infrared photons are absorbed for the most part by organic molecular bonds, it is expected that the dry matter fingerprint (near infrared spectrum related to dry matter) for these substrates is nearly inexistent. Though some information may still indeed be present due to interactions between minerals and OH as some authors in mineral chemistry have outlined (van der Meer, 2018), the fingerprint should be very limited. As a consequence, the fourth component is related to the organic matter content (per fresh mass) rather than the dry matter content.

Furthermore, though all the rest of the substrates show an increase along drying, the rates of increase vary along the substrate types. Some substrates such as *sugar, syrup, or sour cream* show much larger and steeper variations than others. What gathers these substrates is their liquid structure. In these substrates, light penetrates more in the matter, which means that the measured volume is higher, and therefore the absorbance differences due to the moisture content differences are more marked.

3.2.5. Analysis of fifth component:

The fifth component's loadings exhibit positive very sharp peaks situated at 1211, 1391, 1727, 1761, 1891, 2306, and 2347 nm (Fig. 7) which relate to CH, CH₂ and CH₃ combination bands. Moreover, the fifth component's scores separate very clearly the substrates rich in lipids from the ones rich in simple carbohydrates (Fig. 8). Indeed, a first group constituted by *sour cream (and butter, pesto, mayonnaise, or egg yolk not shown here)* exhibits highly increasing scores, while a second group constituted by *sugar, syrup (and ketchup, fermented apple not shown here)* exhibits highly decreasing scores. Between these two groups, a third intermediate group exhibits close-to-zero fluctuations in the scores: *fish, rcw (salad, grass or soya meal not shown here)*. This suggests that the fifth component relates to high fat content substrates, and in particular, to the CH/CH₂/CH₃ bonds that are highly concentrated in fatty acids and triglycerides, and where combination bands are therefore expected to be active. Furthermore, these two groups of substrates can be easily distinguished even at very high levels of moisture content (at least for moisture contents up to 60%). This is a promising outcome in regards to the feasibility of building fat content predictive models on fresh wastes, as there is still information allowing to distinguish substrates based on their fat content.

3.2.6. Analysis of sixth component:

The sixth component's loadings consist of various peaks related to combination bands such as OH combinations (1594, 1935 or 2092 nm), and CH combinations (2283 and 2317) (Fig. 7). Scores exhibit two groups (Fig. 8): increasing scores for samples such as *sugar, syrup, (and chocolate powder, or apricot yoghurt not shown here)*, and decreasing scores for samples such as *fish (chicken not shown here)*.

What distinguishes these groups chemically is the presence or absence of carbohydrates, may it be simple carbohydrates (glucose, sucrose) or complex carbohydrates (starch, cellulose). This suggests that the sixth component is specific to the expression of carbohydrates. Indeed, the band at 2092 has been specifically assigned to combinations of OH vibrations in substrates with high content in starch and cellulose. However, it seems here that such OH combination bands are also expressed in simpler sugars such as glucose and sucrose (*sugar, syrup*). One of the outcomes from this is that the sixth component is a good indicator of the total level of carbohydrates in a substrate.

3.2.7. Analysis of seventh component:

The seventh component's loadings show the same sharp peaks related to CH₂ at 1725, 1762 and 2304, and 2347 nm (Fig. 7) that were already found in the fifth component loadings. Therefore, as expected, substrates with high fat content levels like *sour cream (and butter, mayonnaise not shown here)* all exhibit high scores (Fig. 8). However, the *sugar* substrate also exhibits very high scores compared to the rest which implies that the bands at 1725, 1762, 2304 and 2347 nm are also expressed in *sugar* spectra for low moisture content levels (<10%). This suggests that the bands that allowed a clear separation in the fifth component between *sugar* and the substrates rich in fat, are the other bands at 1391 nm and 1891 nm.

Compared with the fifth component's loadings (Fig. 7), two new peaks are identified: a very sharp peak at 1438 nm, as well as the same OH combination band (2101 nm) that was assigned to the presence of carbohydrates in the sixth component. As pointed out by some authors (Williams, 2009), peaks in the 1430 nm region may not always relate to water's OH bonds. Indeed, OH is present in many different molecules such as within hydroxyl groups in alcohols and carbohydrates or carboxylic groups in fatty acids. However, other authors have assigned the 1438 nm band to be

specifically related to water molecules forming one hydrogen bond (Muncan and Tsenkova, 2019), which is the case of water molecules surrounding sucrose for example. Further investigations would be required to be able to conclude on the specific assignment.

3.2.8. Analysis of eighth component:

In the eighth component's loadings (Fig. 7), negative and positive peaks are positioned on each side of the two main water OH absorbance bands' maximums: a negative peak at 1397 nm and a positive peak at 1467 nm; together with a negative peak at 1874 nm and a positive peak at 1939 nm. In addition, all substrates show increasing scores along drying (Fig. 8), particularly in the high moisture content range (60–100%). These negative and positive peaks represent shifts of the OH-bond absorbance bands that occur from lower wavelengths to higher wavelengths along drying. This shift of the OH-bond absorbance bands has been explained by some authors by the change of the water population types: from free to bound water (from free water molecules to water molecules forming dimers, trimers, quadrimers as well as hydration shells) (Kuroki et al., 2019; Maeda et al., 1995). Of course, at low moisture content ranges (<20%), most substrates show decreasing scores, which suggests that such bound water absorbance bands are disappearing as drying occurs.

3.2.9. Summary of principal components' meanings in regards to water effects

It was shown that one of the main effects of water on near infrared spectra concerns global changes in scattering due to water's crucial role in biomolecules' structure and the resulting physical properties of the substrates. Indeed, the first two components accounting for almost 99% of the total variance relate to the appearance of global additive baselines, as well as a multiplicative effect shown by the modification of spectra slope. As seen, these scattering modifications due to modifications of physical properties vary according to the chemical composition of substrates. For example, the presence of fat may form emulsions leading to decreased scattering levels during drying, while suspensions or porous media formed by solid ligno-cellulosic component show increased scattering levels during drying. As well, the presence of soluble components such as sucrose may lead to transparent solutions with important forward scattering levels. A complex interaction between chemical composition and physical scattering properties has therefore been outlined.

Secondly, a strong overlap of water OH absorbance bands has been highlighted and shown in the second and third components, masking other more minor OH absorbance bands present in carbohydrates, fatty acids, or alcohols.

Thirdly, two different spectral patterns related to water's chemical interaction (ie. water state) were identified in the third and eighth components. Indeed, it was shown in the third component that small differences between the first overtone absorbance band at 1430 nm and the second combination absorbance band at 1940 nm is associated with the presence of simple carbohydrates or proteins, both of these molecules forming important interactions with water. In addition, in the eighth component, as drying occurs, a shift of the OH absorbance bands from high energy vibrations to lower energy vibrations was highlighted for most substrates.

Fourthly, the fourth component was found linearly dependent of dry matter content in most substrates. However, it was shown that the rate of this dependence differed over substrates depending on its physical properties.

Finally, different components related to the substrates' chemical composition were found. Indeed, the fifth, sixth and seventh components differentiated substrates based on carbohydrates

levels, as well as fat content levels. This is promising in regards to the possibility of developing calibrations on high moisture content substrates as there is still information related to the chemical composition: wet substrates spectra are not "just water spectra".

4. Conclusion

The present study investigated the complexity of water effects in near infrared spectroscopy and highlighted the close dependency with the biochemical and physical characteristics of samples.

A customized acquisition system allowed to obtain a unique dataset comprising NIR spectral variations related to water content modifications in standard conditions (ambient temperature/humidity) with no heating nor chemical altering (oxidation, Maillard reactions). Such water spectral variations were obtained on a very wide variety of biochemical types (including carbohydrate substrates, protein substrates, fat substrates as well as packaging materials), allowing a comprehensive analysis of the water effects in near infrared spectroscopy.

A detailed analysis of the dataset using principal component analysis revealed water's complex effects, combining both physical and chemical effects. The fact that water effects depend both on the dry matter content range and the nature of the substrates (both biochemical composition, and physical structure) leads to important challenges for its correction in the context of organic waste characterization. These results encourage future research on the correction of water effects to focus on the development of local and clustered approaches, to correct water effects within groups of substrates with common physical properties and dry matter content range.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the French Agency of National Research and Technology (ANRT) [grant number 2018/0461].

Authors would like to thank Guillaume Guizard and Philippe Soubie for their technical support during the setting-up of the experiment; as well as the ChemHouse group for limitless discussions on chemometrics.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.wasman.2020.12.019>.

References

- Acharya, U.K., Walsh, K.B., Subedi, P.P., 2014. Robustness of partial least-squares models to change in sample temperature: II. Application to fruit attributes. *J. Near Infrared Spectrosc.* 22, 287–295. <https://doi.org/10.1255/jnirs.1119>.
- Albrecht, R., Joffre, R., Gros, R., Le Petit, J., Terron, G., Périssol, C., 2008. Efficiency of near-infrared reflectance spectroscopy to assess and predict the stage of transformation of organic matter in the composting process. *Bioresour. Technol.* 99, 448–455. <https://doi.org/10.1016/j.biortech.2006.12.019>.
- Bogomolov, A., Mannhardt, J., Heinzerling, O., 2018. Accuracy improvement of in-line near-infrared spectroscopic moisture monitoring in a fluidized bed drying process. *Front. Chem.* 6. <https://doi.org/10.3389/fchem.2018.00388>.
- Bogrekcı, I., Lee, W.S., 2006. Effects of soil moisture content on absorbance spectra of sandy soils in sensing phosphorus concentrations using UV-VIS-NIR spectroscopy. *Trans. ASABE* 49, 1175–1180.
- Bowers, S.A., Hanks, R.J., 1965. Reflection of radiant energy from soils. *Soil Sci.* 100, 130–138. <https://doi.org/10.1097/00010694-196508000-00009>.

- Bünning-Pfaue, H., 2003. Analysis of water in food by near infrared spectroscopy. *Food Chem.* 82, 107–115. [https://doi.org/10.1016/S0308-8146\(02\)00583-6](https://doi.org/10.1016/S0308-8146(02)00583-6).
- Campos, M.I., Antolin, G., Debán, L., Pardo, R., 2018. Assessing the influence of temperature on NIRS prediction models for the determination of sodium content in dry-cured ham slices. *Food Chem.* 257, 237–242. <https://doi.org/10.1016/j.foodchem.2018.02.131>.
- Caponigro, V., Marini, F., Gowen, A., 2018. Hydration of hydrogels studied by near-infrared hyperspectral imaging. *J. Chemom.* 32, 1–19. <https://doi.org/10.1002/cem.2972>.
- Carter, G.A., 1991. Primary and secondary effects of water content on the spectral reflectance of leaves. *Am. J. Bot.* 78, 916–924. <https://doi.org/10.1002/j.1537-2197.1991.tb14495.x>.
- Chang, C.-W., Laird, D.A., Hurburgh, C.R., 2005. Influence of soil moisture on near-infrared reflectance spectroscopic measurement of soil properties. *Soil Sci. Soc. Am. J.* 170, 244–255. <https://doi.org/10.1097/01.s0000162289.40879.7b>.
- Charnier, C., Latrille, E., Jimenez, J., Lemoine, M., Boulet, J.C., Miroux, J., Steyer, J.P., 2016. Fast characterization of solid organic waste content with near infrared spectroscopy in anaerobic digestion. *Waste Manage.* 59, 140–148. <https://doi.org/10.1016/j.wasman.2016.10.029>.
- Cozzolino, D., Liu, L., Cynkar, W.U., Damberg, R.G., Janik, L., Colby, C.B., Gishen, M., 2007. Effect of temperature variation on the visible and near infrared spectra of wine and the consequences on the partial least square calibrations developed to measure chemical composition. *Anal. Chim. Acta* 588, 224–230. <https://doi.org/10.1016/j.aca.2007.01.079>.
- Cumming, J.B., 2013. Temperature dependence of light absorption by water. *Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers Detect. Assoc. Equip.* 713, 1–4. <https://doi.org/10.1016/j.nima.2013.02.024>.
- Dehnad, D., Jafari, S.M., Afrasiabi, M., 2016. Influence of drying on functional properties of food biopolymers: From traditional to novel dehydration techniques. *Trends Food Sci. Technol.* 57, 116–131. <https://doi.org/10.1016/j.tifs.2016.09.002>.
- Dvořák, L., Fajman, M., Sustova, K., 2017. Influence of sample temperature for measurement accuracy with FT-NIR spectroscopy. *J. AOAC Int.* 100, 499–502. <https://doi.org/10.5740/jaoacint.16-0264>.
- Fitamo, T., Triolo, J.M., Boldrin, A., Scheutz, C., 2017. Rapid biochemical methane potential prediction of urban organic waste with near-infrared reflectance spectroscopy. *Water Res.* 119, 242–251. <https://doi.org/10.1016/j.watres.2017.04.051>.
- Gaines, C.S., Windham, W.R., 1998. Effect of wheat moisture content on meal apparent particle size and hardness scores determined by near-infrared reflectance spectroscopy. *Cereal Chem.* 75, 386–391. <https://doi.org/10.1094/CHEM.1998.75.3.386>.
- Galvez-Sola, L., Moral, R., Perez-Murcia, M.D., Perez-Espinosa, A., Bustamante, M.A., Martínez-Sabater, E., Paredes, C., 2010. The potential of near infrared reflectance spectroscopy (NIRS) for the estimation of agroindustrial compost quality. *Sci. Total Environ.* 408, 1414–1421. <https://doi.org/10.1016/j.scitotenv.2009.11.043>.
- Gergely, S., Salgó, A., 2003. Changes in moisture content during wheat maturation - What is measured by near infrared spectroscopy? *J. Near Infrared Spectrosc.* 11, 17–26. <https://doi.org/10.1255/jnirs.350>.
- Giordanengo, T., Charpentier, J.P., Roger, J.M., Roussel, S., Brancheriau, L., Chaix, G., Bailleres, H., 2008. Correction of moisture effects on near infrared calibration for the analysis of phenol content in eucalyptus wood extracts. *Ann. For. Sci.* 65. <https://doi.org/10.1007/s10355-008-0065-0>.
- Godin, B., Mayer, F., Agneessens, R., Gerin, P., Dardenne, P., Delfosse, P., Delcarte, J., 2015. Biochemical methane potential prediction of plant biomasses: Comparing chemical composition versus near infrared methods and linear versus non-linear models. *Bioresour. Technol.* 175, 382–390. <https://doi.org/10.1016/j.biortech.2014.10.115>.
- Golic, M., Walsh, K.B., 2006. Robustness of calibration models based on near infrared spectroscopy for the in-line grading of stonefruit for total soluble solids content. *Anal. Chim. Acta* 555, 286–291. <https://doi.org/10.1016/j.aca.2005.09.014>.
- Gorretta, N., Nouri, M., Herrero, A., Gowen, A., Roger, J.M., 2019. Early detection of the fungal disease “apple scab” using SWIR hyperspectral imaging, in: *Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote Sensing*. <https://doi.org/10.1109/WHISPERS.2019.8921066>.
- Greenspan, L., 1976. Humidity Fixed Points of Binary Saturated Aqueous Solutions. *J. Res. Natl. Bur. Stand. Phys. Chem.* 81A, 89–96. <https://doi.org/10.6028/jres.081A.011>.
- Hans, G., Allison, B., Bruce, A., 2019. Temperature and Moisture Insensitive Prediction of Biomass Calorific Value from Near-Infrared Spectra Using External Parameter Orthogonalization. *J. Near Infrared Spectrosc.* 3–23. <https://doi.org/10.1177/0967033519840742>.
- Hunter, J.D., 2007. *Matplotlib: A 2D graphics environment*. *Comput. Sci. Eng.* 9, 90–95.
- Igne, B., Hossain, M.N., Drennen, J.K., Anderson, C.A., 2014. Robustness considerations and effects of moisture variations on near infrared method performance for solid dosage form assay. *J. Near Infrared Spectrosc.* 22, 179–188. <https://doi.org/10.1255/jnirs.1097>.
- Ilari, J.L., Martens, H., Isaksson, T., 1988. Determination of Particle Size in Power By Scatter Correction in Diffuse Near-Infrared Reflectance. *Appl. Spectrosc.* 42, 722–728. <https://doi.org/10.1366/0003702884429058>.
- Isaksson, T., Naes, T., 1988. Effect of multiplicative scatter correction (MSC) and linearity improvement in NIR spectroscopy. *Appl. Spectrosc.* 42, 1273–1284. <https://doi.org/10.1366/0003702884429869>.
- Jacobi, H.F., Moschner, C.R., Hartung, E., 2009. Use of near infrared spectroscopy in monitoring of volatile fatty acids in anaerobic digestion. *Water Sci. Technol.* 60, 339–346. <https://doi.org/10.2166/wst.2009.345>.
- Knadel, M., Deng, F., Alinejad, A., Wollesen de Jonge, L., Moldrup, P., Greve, M.H., 2014. The Effects of Moisture Conditions—From Wet to Hyper dry—On Visible Near-Infrared Spectra of Danish Reference Soils. *Soil Sci. Soc. Am. J.* 78, 422. <https://doi.org/10.2136/sssaj2012.0401>.
- Kou, L., Labrie, D., Chylek, P., 1993. Refractive indices of water and ice in the 065– to 25- μm spectral range. *Appl. Opt.* 32, 3531. <https://doi.org/10.1364/ao.32.003531>.
- Kuroki, S., Tsenkova, R., Moyankova, D., Muncan, J., Morita, H., Atanassova, S., Djilianov, D., 2019. Water molecular structure underpins extreme desiccation tolerance of the resurrection plant *Haberlea rhodopensis*. *Sci. Rep.* 9, 1–12. <https://doi.org/10.1038/s41598-019-39443-4>.
- Laage, D., Elsaesser, T., Hynes, J.T., 2017. Water Dynamics in the Hydration Shells of Biomolecules. *Chem. Rev.* 117, 10694–10725. <https://doi.org/10.1021/acs.chemrev.6b00765>.
- Lesteur, M., Latrille, E., Maurel, V.B., Roger, J.M., Gonzalez, C., Junqua, G., Steyer, J.P., 2011. First step towards a fast analytical method for the determination of Biochemical Methane Potential of solid wastes by near infrared spectroscopy. *Bioresour. Technol.* 102, 2280–2288. <https://doi.org/10.1016/j.biortech.2010.10.044>.
- Lobell, D.B., Asner, G.P., 2002. Moisture Effects on Soil Reflectance. *Soil Sci. Soc. Am. J.* 66, 722. <https://doi.org/10.2136/sssaj2002.7220>.
- Luck, W.A., 1974. *Structure of water and aqueous solutions*. Verlag Chemie.
- Maeda, H., Ozaki, Y., Tanaka, M., Hayashi, N., Kojima, T., 1995. Near Infrared Spectroscopy and Chemometrics Studies of Temperature-Dependent Spectral Variations of Water: Relationship between Spectral Changes and Hydrogen Bonds. *J. Near Infrared Spectrosc.* 3, 191–201. <https://doi.org/10.1255/jnirs.69>.
- Martens, H., Nielsen, J.P., Engelsen, S.B., 2003. Light scattering and light absorbance separated by extended multiplicative signal correction. Application to near-infrared transmission analysis of powder mixtures. *Anal. Chem.* 75, 394–404. <https://doi.org/10.1021/ac020194w>.
- Mayer, F., Noo, A., Sinnaeve, G., Dardenne, P., Gerin, P., 2013. Prediction of the biochemical methane potential (BMP) of maize silages reduced to a powder using NIR spectra from wet and dried samples, in: *NIR2013 Proceedings: Picking Up Good Vibrations*. pp. 458–463.
- 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/majora-92bf1922-00a>.
- van der Meer, F., 2018. Near-infrared laboratory spectroscopy of mineral chemistry: A review. *Int. J. Appl. Earth Obs. Geoinf.* 65, 71–78. <https://doi.org/10.1016/j.jag.2017.10.004>.
- Mortreuil, P., Lagnet, C., Schraauwers, B., Algae, F.M., 2018. Fast prediction of organic wastes methane potentials by Near Infrared Reflectance Spectroscopy (NIRS): a successful tool for agricultural biogas plant. *Uest.Ntua.Gr*. <https://doi.org/10.1177/0734242X18778773>.
- Muncan, J., Tsenkova, R., 2019. Aquaphotomics—From Innovative Knowledge to Integrative Platform in Science and Technology. *Molecules* 24, 2742. <https://doi.org/10.3390/molecules24152742>.
- Olyphant, T.E., 2010. *Guide to NumPy*, Methods. Trelgol Publishing USA.
- Oliveri, P., Malegori, C., Simonetti, R., Casale, M., 2019. The impact of signal pre-processing on the final interpretation of analytical outcomes – A tutorial. *Anal. Chim. Acta* 1058, 9–17. <https://doi.org/10.1016/j.aca.2018.10.055>.
- Padalkar, M.V., Pleshko, N., 2015. Wavelength-dependent penetration depth of near infrared radiation into cartilage. *Analyst* 140, 2093–2100. <https://doi.org/10.1039/c4an01987c>.
- Pasquini, C., 2003. *Review Near Infrared Spectroscopy : Fundamentals*. *Practical Aspects Anal. Appl.* 14, 198–219.
- Pedregosa, F., Varoquaux, G., Buitinck, L., Louppe, G., Grisel, O., Mueller, A., 2015. *Scikit-learn: Machine Learning in Python*. *J. Mach. Learn. Res.* 19, 29–33.
- Peiris, K.H.S., Dong, Y., Bockus, W.W., Dowell, F.E., 2016. Moisture Effects on the Prediction Performance of a Single-Kernel Near-Infrared Deoxyvalenol Calibration. *Cereal Chem. J.* 93, 631–637. <https://doi.org/10.1094/CHEM-04-16-0120-R>.
- Polyanskiy, M., 2008. *Refractive index database*. [WWW Document]. <http://refractiveindex.info/> (accessed 1.27.20).
- Popineau, S., Rondeau-Mouro, C., Sulpice-Gaillet, C., Shanahan, M.E.R., 2005. Free/bound water absorption in an epoxy adhesive. *Polymer (Guildf.)* 46, 10733–10740. <https://doi.org/10.1016/j.polymer.2005.09.008>.
- Rabatel, G., Marini, F., Walczak, B., Roger, J., 2019. VSN: Variable sorting for normalization. *J. Chemom.* 1–16. <https://doi.org/10.1002/cem.3164>.
- Raponi, F., Ferri, S., Monarca, D., Moscetti, R., Colantoni, A., Massantini, R., 2017. Real-time monitoring of organic apple (var. Gala) during hot-air drying using near-infrared spectroscopy. *J. Food Eng.* 222, 139–150. <https://doi.org/10.1016/j.jfoodeng.2017.11.023>.
- Reeves, J.B., 1995. Efforts to Quantify Changes in Near-Infrared Spectra Caused by the Influence of Water, pH, Ionic Strength, and Differences in Physical State. *Appl. Spectrosc.* 49, 181–187. <https://doi.org/10.1366/0003702953963788>.
- Reeves, J.B., 1994. Effects of water on the spectra of model compounds in the short-wavelength near infrared spectral region (14,000–9091 cm^{-1} or 714–1100 nm) 212, 199–212.
- Renati, P., Kovacs, Z., De Ninno, A., Tsenkova, R., 2019. Temperature dependence analysis of the NIR spectra of liquid water confirms the existence of two phases, one of which is in a coherent state. *J. Mol. Liq.* 292. <https://doi.org/10.1016/j.molliq.2019.111449>.

- Rinnan, Å., van den Berg, F., Engelsen, S.B., 2009. Review of the most common preprocessing techniques for near-infrared spectra. *TrAC - Trends Anal. Chem.* 28, 1201–1222. <https://doi.org/10.1016/j.trac.2009.07.007>.
- Roger, J., Biancolillo, A., Marini, F., 2020. Sequential preprocessing through ORTHogonalization (SPORT) and its application to near infrared spectroscopy. *Chemom. Intell. Lab. Syst.* 199, <https://doi.org/10.1016/j.chemolab.2020.103975>.
- Roger, J.M., Chauchard, F., Bellon-Maurel, V., 2003. EPO-PLS external parameter orthogonalisation of PLS application to temperature-independent measurement of sugar content of intact fruits. *Chemom. Intell. Lab. Syst.* 66, 191–204. [https://doi.org/10.1016/S0169-7439\(03\)00051-0](https://doi.org/10.1016/S0169-7439(03)00051-0).
- Roger, J.M., Palagos, B., Guillaume, S., Bellon-Maurel, V., 2005. Discriminating from highly multivariate data by Focal Eigen Function discriminant analysis; application to NIR spectra. *Chemom. Intell. Lab. Syst.* 79, 31–41. <https://doi.org/10.1016/j.chemolab.2005.03.006>.
- Sánchez, N.H., Lurol, S., Roger, J.M., Bellon-Maurel, V., 2003. Robustness of models based on NIR spectra for sugar content prediction in apples. *J. Near Infrared Spectrosc.* 11, 97–107. <https://doi.org/10.1255/jnirs.358>.
- Sørensen, M., Larsen, A., Esbensen, K.H., 2014. Visualisation of Sampling Error Effects in near Infrared Analysis—Comparison between Petri Dish, Roll Bottle and Spiral Sampler. *NIR news* 25, 11–15. <https://doi.org/10.1255/nirn.1414>.
- Stockl, A., Lichti, F., 2018. Near-infrared spectroscopy (NIRS) for a real time monitoring of the biogas process. *Bioresour. Technol.* 247, 1249–1252. <https://doi.org/10.1016/j.biortech.2017.09.173>.
- Sudduth, K.A., Hummel, J.W., 1993. Soil organic matter, CEC, and moisture sensing with a portable NIR spectrophotometer. *Trans. - Am. Soc. Agric. Eng.* 36, 1571–1582. <https://doi.org/10.13031/2013.28498>.
- Sun, X., Subedi, P., Walsh, K.B., 2020. Achieving robustness to temperature change of a NIRS-PLSR model for intact mango fruit dry matter content. *Postharvest Biol. Technol.* 162, <https://doi.org/10.1016/j.postharvbio.2019.111117>.
- van Rossum, G., Drake, F.L., 2009. *Python 3 Reference Manual*, Scotts Valley, CA. CreateSpace, Scotts Valley, CA.
- Vergnoux, A., Guiliano, M., Le Dréau, Y., Kister, J., Dupuy, N., Doumenq, P., 2009. Monitoring of the evolution of an industrial compost and prediction of some compost properties by NIR spectroscopy. *Sci. Total Environ.* 407, 2390–2403. <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., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, I., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., 2019. *SciPy 1.0—Fundamental algorithms for scientific computing in python*. arXiv arXiv:1907.10121.
- Wenz, J.J., 2018. Examining water in model membranes by near infrared spectroscopy and multivariate analysis. *Biochim. Biophys. Acta - Biomembr.* 1860, 673–682. <https://doi.org/10.1016/j.bbamem.2017.12.007>.
- Williams, P., 2009. Influence of water on prediction of composition and quality factors: The Aquaphotomics of low moisture agricultural materials. *J. Near Infrared Spectrosc.* 17, 315–328. <https://doi.org/10.1255/jnirs.862>.
- Williams, P., Antoniszyn, J., 2019. Near-infrared Technology: Getting the best out of light, Near-infrared Technology: Getting the best out of light. *AFRICAN SUN MEDIA*. <https://doi.org/10.18820/9781928480310>
- Workman, L., Weyer, J., 2012. *Practical Guide and Spectral Atlas for Interpretive Near-Infrared Spectroscopy, Second Edition, Practical Guide and Spectral Atlas for Interpretive Near-Infrared Spectroscopy, Second Edition Spectroscopy*, Second Edition. CRC press. <https://doi.org/10.1201/b11894>
- Wu, C.Y., Jacobson, A.R., Laba, M., Baveye, P.C., 2009. Alleviating moisture content effects on the visible near-infrared diffuse-reflectance sensing of soils. *Soil Sci.* 174, 456–465. <https://doi.org/10.1097/SS.0b013e3181b21491>.
- Wülfert, F., Kok, W.T., Smilde, A.K., 1998. Influence of Temperature on Vibrational Spectra and Consequences for the Predictive Ability of Multivariate Models. *Anal. Chem.* 70, 1761–1767. <https://doi.org/10.1021/ac9709920>.
- Zeaiter, M., Roger, J.M., Bellon-Maurel, V., Rutledge, D.N., 2004. Robustness of models developed by multivariate calibration. Part I: The assessment of robustness. *TrAC - Trends Anal. Chem.* 23, 157–170. [https://doi.org/10.1016/S0165-9936\(04\)00307-3](https://doi.org/10.1016/S0165-9936(04)00307-3).
- Zeaiter, M., Rutledge, D., 2009. Preprocessing Methods. *Compr. Chemom.* 3, 121–231. <https://doi.org/10.1016/B978-044452701-1.00074-0>.

Paper III – Relating Near-Infrared Light Path-Length Modifications to the Water Content of Scattering Media in Near-Infrared Spectroscopy: Toward a New Bouguer – Beer – Lambert Law

Full reference:

Mallet, A.; Tsenkova, R.; Muncan, J.; Charnier, C.; Latrille, E.; Bendoula, R.; Steyer, J. P.; Roger, J. M. Relating Near-Infrared Light Path-Length Modifications to the Water Content of Scattering Media in Near-Infrared Spectroscopy: Toward a New Bouguer-Beer-Lambert Law. *Anal. Chem.* **2021**, *93* (17), 6817–6823. <https://doi.org/10.1021/acs.analchem.1c00811>.

Relating Near-Infrared Light Path-Length Modifications to the Water Content of Scattering Media in Near-Infrared Spectroscopy: Toward a New Bouguer–Beer–Lambert Law

Alexandre Mallet,* Roumiana Tsenkova, Jelena Muncan, Cyrille Charnier, Éric Latrille, Ryad Bendoula, Jean-Philippe Steyer, and Jean-Michel Roger



Cite This: <https://doi.org/10.1021/acs.analchem.1c00811>



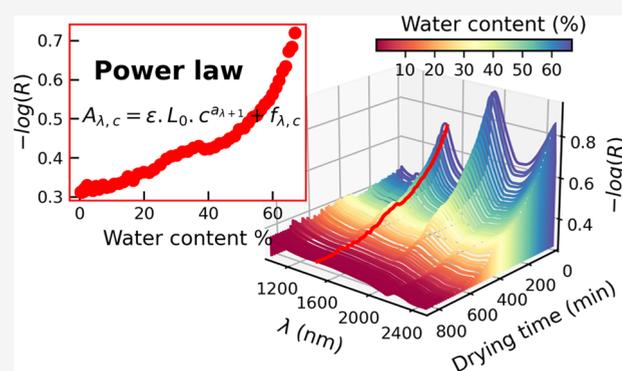
Read Online

ACCESS |

Metrics & More

Article Recommendations

ABSTRACT: In near-infrared spectroscopy (NIRS), the linear relationship between absorbance and an absorbing compound concentration has been strictly defined by the Bouguer–Beer–Lambert law only for the case of transmission measurements of nonscattering media. However, various quantitative calibrations have been successfully built both on reflectance measurements and for scattering media. Although the lack of linearity for scattering media has been observed experimentally, the sound multivariate statistics and signal processing involved in chemometrics have allowed us to overcome this problem in most cases. However, in the case of samples with varying water content, important modifications of scattering levels still make calibrations difficult to build due to nonlinearities. Moreover, even when calibration procedures are successfully developed, many preprocessing methods used do not guarantee correct spectroscopic assignments (in the sense of a pure chemical absorbance). In particular, this may prevent correct modeling and interpretation of the structure of water. In this study, dynamic near-infrared spectra acquired during a drying process allow the study of the physical effects of water content variations, with a focus on the first overtone OH absorbance region. A model sample consisting of aluminum pellets mixed with water allowed us to study this specifically, without any other absorbing interaction terms related to the dry mass-absorbing constituents. A new formulation of the Bouguer–Beer–Lambert law is proposed, by expressing path length as a power function of water content. Through this new formulation, it is shown that a better and simpler prediction model of water content may be developed, with more precise and accurate identification of water absorbance bands.



INTRODUCTION

Although apparently a simple molecule (H₂O), water shows complex behaviors and presents many physical anomalies compared to other liquids.^{1,2} This is mostly explained by its dominating intermolecular hydrogen bonding. As surprising as it may seem, its structure model is far from having reached a broad consensus.³ One of the analytical techniques used to reveal details of the water structure and its functionalities in aqueous solutions is near-infrared (NIR) spectroscopy. However, very different models and interpretations of what constitutes the first overtone OH absorbance region in NIR have been proposed. Using second derivative, up to six underlying water species were detected and said to correspond to different water species: protonated water (Sr), and water with no, one, two, three, and four hydrogen bonds (S0–S4).⁴ Using nonlinear fitting procedures, these authors fit six Gaussian peaks to reconstruct the observed spectra. This was later confirmed by other authors studying water-glucose solutions.⁵ Though good reconstruction errors were obtained,

these nonlinear fitting procedures may suffer from user-guided hypothesis and initialization biases. In another study, using multivariate curve-resolution-asymmetric least squares (MCR-ALS), three components were found best to explain water spectra at different temperatures.⁶ Furthermore, it was shown that though water with salt showed the same trends, the position of these three components depended on the ionic strength of salts. The fact that the components were not forced to be Gaussian in this study could explain why a lower number of components were obtained. A two-state model (water species with weaker and stronger hydrogen bonds) has also

Received: February 22, 2021

Accepted: April 12, 2021

been outlined using second derivatives, two-dimensional correlation spectroscopy (2D-COS), and principal component analysis (PCA) applied on temperature-dependent NIR spectroscopy (NIRS).⁷ Recently, this model was supported by new temperature-dependent data that suggest the presence of a coherent state.⁸ In this experiment, an isosbestic point has been observed around 1438 nm, which suggests the existence of an equilibrium between two populations/states of water, referred to as the “mixture-model”.⁹ Using second derivation, the authors have found two negative subpeaks with stable positions under temperature variations (positioned at 1412 and 1462 nm). However, they pointed out that the isosbestic point was not perfectly constant and could be dependent on temperature, as previously also observed by Gowen et al.,⁶ which seems to be still an unresolved question today: is there a third population of vibration?⁷ An interaction term? Baseline artifacts? A continuous distribution of hydrogen bond geometries?¹⁰

A concept that has been developed in the recent years, called aquaphotomics,¹¹ deals with the water–light interaction over the whole electromagnetic spectrum. It has been found that the first overtone OH region consists of 12–14 different water absorbance bands, each related to a given water structure.¹² Various water structures have been experimentally identified in various systems containing water and later found to be in agreement with theoretical calculations, such as water solvation shells at 1364–1384 nm, water molecules confined by ions with no hydrogen bonding at 1396–1403 nm,¹³ free OH water with no hydrogen bonding like in water vapor at 1403–1418 nm (S0), protein hydration shells at 1418–1430 nm, water molecules with one, two, three, or four hydrogen bonds (S1–S4), respectively, at 1432–1454, 1458–1468, 1472–1482, and 1482–1495 nm, and strongly bound water at 1506–1516 nm. Shao et al.,¹⁴ on the other hand, in the same region found 10 spectral components (Gaussian peaks) corresponding to nine different water molecular structures, using knowledge-based genetic algorithm. However, the correct assignment of these bands to water states implies that the peaks truly correspond to absorbing species and are not due to scattering artifacts. While this seems fully respected in pure water or diluted solution measurements, the Bouguer–Beer–Lambert law remains unobeyed for scattering media (such as suspensions and powders).¹⁵ This makes it difficult to consolidate water-state-specific assignments in scattering media.

To solve this, a variety of preprocessing algorithms have been proposed, based on the hypothesis that scattering provokes additive and multiplicative effects. These effects come from factors such as particle size¹⁶ or water content.¹⁷ To remove additive effects, detrend¹⁸ is a simple preprocessing algorithm allowing the removal of polynomial baselines of different orders. To remove multiplicative effects, one strategy is to apply logarithm to transform the multiplicative effect into an additive one. Other algorithms allow us to correct additive and multiplicative effects together: standard normal variate (SNV),¹⁸ multiplicative scatter correction (MSC),^{19–21} extended multiplicative scatter correction (EMSC),^{22–24} or optical path-length estimation and correction (OPLEC).^{25,26}

One concern regarding the application of these methods is that the additive and multiplicative effects may be strictly identified only in some regions (where no or very little chemical absorbance is present). This led to further developments and extensions of these methods to be applied on specific wavelengths regions only. For example, the normalized

spectral ratio (NSR),²⁷ which can be seen as a more restricted SNV, assumes that in a spectrum, there is a wavelength subjected to only an additive effect and another wavelength subjected to only a multiplicative effect. The robust normal variate (RNV)²⁸ extends the SNV as well, by calculating the mean and standard deviation only on a percentile of the spectrum. Similarly, probabilistic quotient normalization (PQN)²⁹ has been proposed in the field of ¹H NMR analysis to avoid spectral regions where single endogenous or drug metabolites absorb and negatively influence the classical integral normalization. In the EMSC framework, a weighted least-squares (WLS) estimation of the multiplicative and additive coefficients has been proposed.²³ Authors have proposed different methodologies to determine these weights, including the simple inverse of the mean spectra, or in a more sophisticated way using a robust regression method such as random sample consensus (RANSAC)³⁰ as proposed in the variable sorting for normalization (VSN) algorithm.³¹

Another noteworthy approach to deal, more generally, with the influence of an interference is prewhitening using generalized least-squares weighing (GLSW)³² or orthogonal projections like in external parameter orthogonalization (EPO).³³ These have been successfully applied on soils to deal with moisture effects³⁴ or sucrose solutions to deal with temperature effects.³⁵

Finally, some other algorithms coming from the signal processing community have been used, such as Savitsky–Golay derivation (usually first or second),³⁶ continuum removal (CR),³⁷ asymmetric least squares (ALS), or discrete and continuous wavelet transforms (DWT and CWT, respectively).^{14,38} It can be noted that when the preprocessing is part of a calibration modeling pipeline, these may be combined.^{39–41} Though these latter algorithms undoubtedly allow successful calibration models to be developed, a drawback is that spectroscopic interpretation of the resulting signal is more difficult than former physics-based models such as EMSC.⁴² In particular, this comes as an important aspect when looking for specific assignments of water structure in NIR spectra: the chemical extinction coefficients’ relative intensity and positions are of importance. Furthermore, this task is made more difficult by the fact that absorptivity of overtones in the near-infrared spectral region (10 000–4000 cm⁻¹) is about 10 times less intense than the absorptivity of fundamentals in the infrared spectral region (4000–1000 cm⁻¹).

The present paper proposes a new phenomenological equation relating light path-length modifications to water content for scattering media containing water. Such equation is illustrated in a model medium composed of aluminum paper pellets and water, and demonstrated through the use of the extended multiplicative scatter (EMSC) correction framework. This equation could lead to further developments of preprocessing algorithms that take into account this relationship specifically. Better understanding of the water content effects in the samples containing water could also lead to a more precise determination of which regions in the NIR spectra are specific to water state information.

■ MATERIALS AND METHODS

Experimental Setup. Aluminum paper pellets of 5 mm diameter and with a density of 0.21 g cm⁻³ were mixed with water, and 50 g was sampled. The initial dry matter content was measured classically by taking two replicate samples of 10

g, which were weighed before and after 48 h of drying in a heat chamber at 105 °C. The samples were then placed individually into a custom-built air-drying system with simultaneous online acquisition of NIR spectra and determination of moisture content. For full details on the materials and methods, see Mallet et al. (2021).⁴³ The drying system consisted of a closed-tube loop with a 2000 mL min⁻¹ internal flow of air generated by a peristaltic pump. A strong desiccant (sodium hydroxide) was used to maintain the gas-phase relative humidity at about 8%, and the drying circuit was connected to a hermetic spectrometer sampling cup in which the aluminum sample was placed. The thickness of the sample in the spectroscopic measurement chamber was 3 cm. The sampling cup was placed over a Fourier transform–NIR (FT-NIR) spectrophotometer (Buchi NIR-Flex N-500) with a tungsten halogen lamp source and an extended-range InGaAs detector for continuous automatic NIR reflectance acquisitions. No level of zero filling was used before applying the Fourier transform. In addition to the spectral measurements, the desiccant was weighed continuously using a precision balance (Ohaus Traveler TA502) to enable the measurement of loss of water during drying. The temperature of the sample was monitored using a temperature-logging module (DTM5080 LKMelectronic), and the mean temperature was 25.9 °C with a standard deviation of 0.4 °C. The temperature dependence of light absorption by water was shown to be maximal at 1410 nm, with a temperature coefficient of 0.8% °C⁻¹.⁴⁴ This means that the maximum absorbance variation due to temperature that could be expected in this experimental setup is of only 0.34% of the absorbance, which is negligible.

The dataset consisted of 573 reflectance spectra from 4000 to 10 000 cm⁻¹ (4 cm⁻¹ resolution) covering a range of water content from 0.02 to 67%. This resulted in a matrix $X(n, p)$ with number of spectra $n = 573$ and number of wavelengths $p = 1501$.

All of the data analyses were performed using Python 3.6.5: data wrangling with NumPy 1.17.3, signal processing with SciPy 1.3.1, and plotting with Matplotlib 2.2.2.^{45–48}

Theory and Data Analysis. For homogeneous isotropic media without particles, transmission measurements respect the Bouguer–Beer–Lambert law¹⁵

$$A_{\lambda,c} = -\log(T_{\lambda,c}) = \varepsilon_{\lambda} \cdot L \cdot c \quad (1)$$

with $A_{\lambda,c}$ the absorbance, $T_{\lambda,c}$ the transmittance, ε_{λ} the extinction coefficient, L the light path length, and c the concentration of the absorbing species.

In scattering media, two phenomena must be taken into account:⁴⁹ a modification of path length $L_{\lambda,c}$ and a loss of photons $f_{\lambda,c}$

$$A_{\lambda,c} = \varepsilon_{\lambda} \cdot L_{\lambda,c} \cdot c + f_{\lambda,c} \quad (2)$$

In most studies interested in the multiplicative effects on NIR, $L_{\lambda,c}$ is modeled with a simple multiplicative constant k not directly related to the analyte concentration²⁴

$$L_{\lambda,c} = k l_0 \quad (3)$$

with l_0 a constant path length.

However, there are many cases where the absorptive specie of interest is in fact directly responsible for the multiplicative effects. It has been shown, for example, that water content variations during drying induce scattering modifications because of the increased number or intensity of refractive

index differences between the gas phase, liquid phase, and solid particles.⁴³ In such cases, it seems reasonable to relate the light path-length modifications (multiplicative effect) directly to the concentration of the absorptive species (water content). Because such relationship is expected to involve a geometrical relationship, it is here proposed that the path-length function is set as to be a simple power function of the water content concentration c

$$L_{\lambda,c} = l_0 \cdot c^{a_{\lambda}} \quad (4)$$

with a_{λ} a power coefficient dependent on the wavelength λ .

Combining eqs 2 and 4 yields a new formulation of the absorption law for scattering media

$$A_{\lambda,c} = \varepsilon_{\lambda} \cdot l_0 \cdot c^{a_{\lambda}+1} + f_{\lambda,c} \quad (5)$$

As already reviewed, the additive baseline term $f_{\lambda,c}$ can be removed by a wide number of pretreatments. To preserve spectroscopic relative amplitudes and positions, the extended multiplicative scatter correction (EMSC) framework was used. A pure water transmission spectrum acquired using a JASCO V560 was used as the reference spectrum (after transforming to absorbance using a logarithm). A constant, first-order and second-order polynomial was included as in the original EMSC.²⁴ An interferent pure spectrum was added as the measured spectrum of the fully dried aluminum. Finally, to evaluate the additive and multiplicative terms only in regions where water absorbance bands are known to interfere less, a weighted least-squares (WLS) was applied with weights set as the inverse of the pure water transmission spectrum. This resulted in the following EMSC model

$$A_{\text{observed},i}(\lambda) = a_i \cdot A_{\text{pure water}}(\lambda) + b_i + c_i \lambda + d_i \lambda^2 + e_i A_{\text{pure aluminum}}(\lambda) \quad (6)$$

Instead of correcting for both the additive and multiplicative effects, only additive effects were corrected for

$$A_{\text{corrected},i}(\lambda) = A_{\text{observed},i}(\lambda) - (b_i + c_i \lambda + d_i \lambda^2 + e_i A_{\text{pure aluminum}}(\lambda)) \quad (7)$$

This yielded spectra corrected only for additive effects. Thus, eq 5 can be rewritten as such

$$A_{\lambda,c} - f_{\lambda,c} = \varepsilon_{\lambda} \cdot l_0 \cdot c^{a_{\lambda}+1} \quad (8)$$

Applying the log transform then results in the following formula

$$\log(A_{\lambda,c} - f_{\lambda,c}) = \log(\varepsilon_{\lambda} \cdot l_0) + (a_{\lambda} + 1) \cdot \log(c) \quad (9)$$

For each wavelength λ , an ordinary least-squares (OLS) regression was applied between $\log(A_{\lambda,c} - f_{\lambda,c})$ and $\log(c)$ to evaluate $a_{\lambda} + 1$ (the slope) and $\log(\varepsilon_{\lambda} \cdot l_0)$ (the intercept). To evaluate goodness of fit, the coefficient of determination (R^2) was calculated.

To investigate the impact of such a new relationship on the development of calibration models, a comparison of different models is proposed, with different preprocessing of spectra (X) and water content values (y). A first model was built on raw X and raw y . A second model was built on additive-only EMSC-corrected X (as in eq 7) and raw y . A third model was built on a complete EMSC-corrected X (with both additive and multiplicative terms corrected for) and raw y . Finally, a fourth

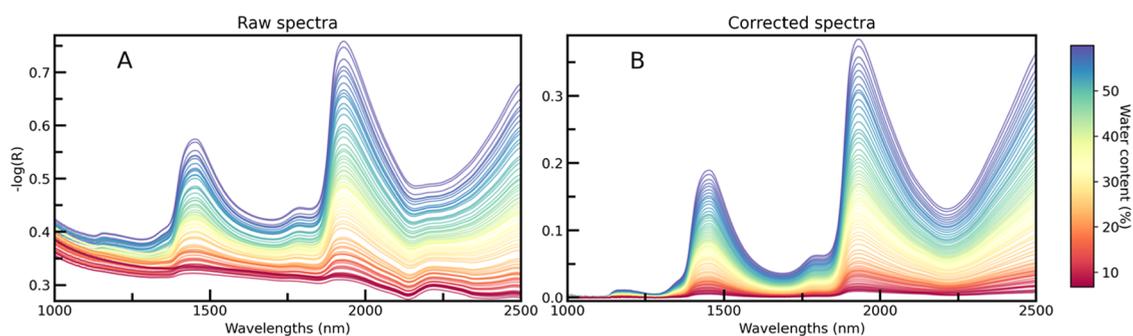


Figure 1. (A) Raw absorbance spectra of aluminum pellets during drying and (B) absorbance spectra corrected for additive effects using the EMSC framework as detailed in eq 6. Spectra are colored by water content (%) from red (low water content) to blue (high water content). For better representation, a subset of spectra is displayed (72 spectra uniformly sampled).

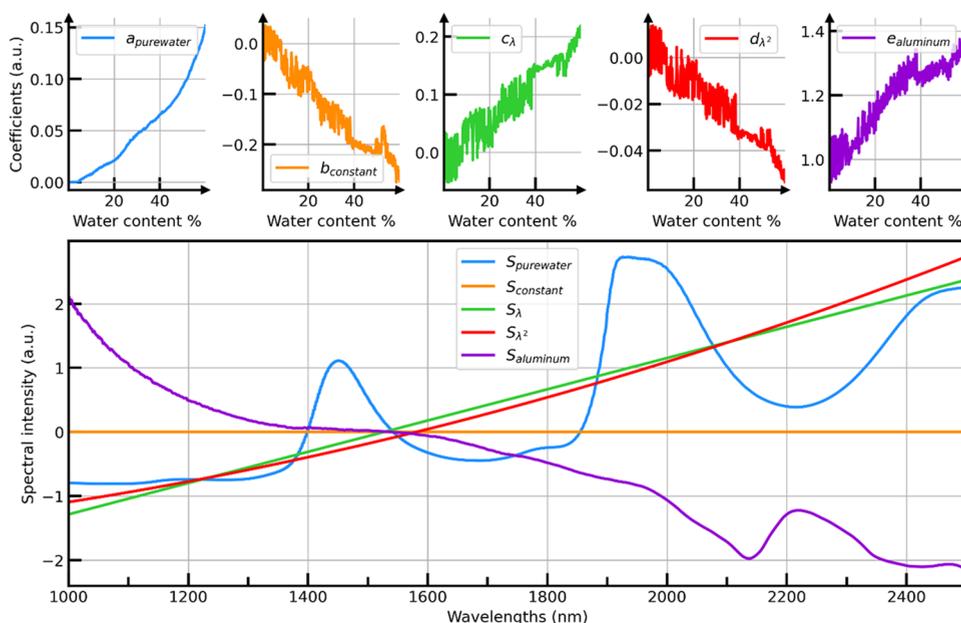


Figure 2. Extended multiplicative scatter correction (EMSC) model components as detailed in eq 6: coefficients (upper plots) and pure spectra (lower plot). The pure spectra were mean-centered and reduced as in a standard normal variate for better visual comparison of spectra shapes.

model was built on log-transformed additive-only EMSC-corrected X and log-transformed y . All models were calibrated using a partial least-squares regression (PLS1-R) with NIPALS algorithm^{50,51} with up to 10 latent variables. Models were fit on a training set representing 30% of the dataset, obtained by the Duplex algorithm⁵² on y or $\log(y)$, to obtain a uniform distribution. To evaluate the performance of these models, the root-mean-square error of calibration (RMSEC) was calculated.

RESULTS AND DISCUSSION

Raw Absorbance Spectra and Correction for Additive Effects. The near-infrared spectral evolutions observed during drying of aluminum are shown in Figure 1. Raw spectra (A) present strong variations of the global absorbance level with water content variations. Such global variations relate to scattering modifications due to changes in intensity or number of the differences in refraction indices between the different phases, as previously reported.^{19,20,43}

Moreover, well-known broad absorbance features due to OH vibrations around 1190, 1450, and 1940 nm are clearly observed in the NIR spectra when water is present (water

content superior to 5%). These are attributed, respectively, to the combination of the first overtone of the O–H stretching and O–H bending band, the first overtone of the O–H stretching band and the combination of the O–H stretching band and O–H bending band of water.^{12,53} A shoulder peak can as well be noticed at 1790 nm as observed by other authors.⁵⁴ In the corrected spectra (Figure 1B), the variations at 1000 nm are absent, just like they are absent in the transmission spectra of pure water.

Figure 2 presents the extended multiplicative scatter correction (EMSC) model components defined in eq 6. Though aluminum is a high reflective material, its dry spectrum (as presented in the lower plot of Figure 2) is not fully flat. For example, a small absorption band at 2220 nm can be observed. As some authors have highlighted, this band could arise from an association between OH groups and the silicon of the quartz sample cell window, as also seen in soil spectra rich in silicate.⁵⁴ In addition, the dry aluminum spectrum shows a clear negative slope, especially between 1000 and 1300 nm. Indeed, this region is particularly affected by scattering events, as from a theoretical standpoint, scattering occurs when the particle size is greater than the wavelength of the radiation. It can be concluded that the pure aluminum

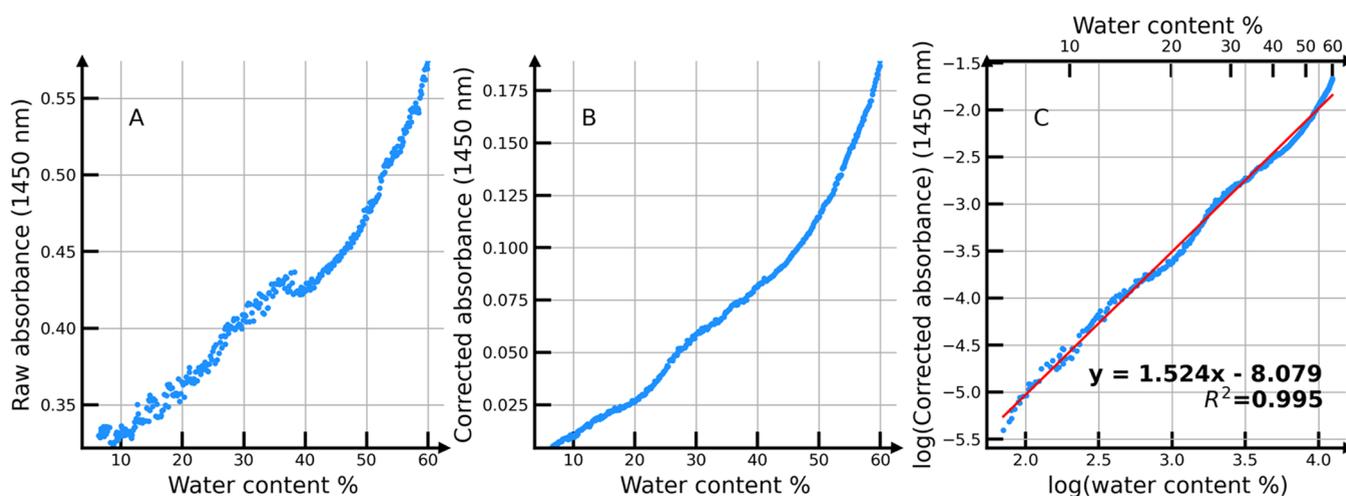


Figure 3. Evolutions with water content % of (A) raw absorbance values, (B) corrected absorbance values, and (C) log-transformed corrected absorbance values. The latter log-transformed corrected absorbance values are plotted with log-transformed water content %. In the (C) subplot, the OLS regression line is plotted in red, with the slope, intercept, and coefficient of determination (R^2).

spectrum is in fact representative of the full instrumental setup condition: a reflectance measurement of scattering aluminum paper pellets through a quartz cell window.

The coefficients of the proposed EMSC model (upper plots of Figure 2) show different evolutions with water content. While additive baselines (constant b_{constant} , first-order c_{λ} , and second-order d_{λ}^2 polynomials, and the dry aluminum signature $\epsilon_{\text{aluminum}}$) show linear evolutions with water content, the pure water coefficient ($a_{\text{pure water}}$) exhibits a nonlinear evolution investigated in the following section.

Relating Path-Length Modifications to Water Content. In Figure 3A, the evolutions of raw absorbance at 1450 nm show nonlinear evolutions with water content. Moreover, differences between absorbance values can be observed despite identical water content. These differences were well removed by the EMSC correction (Figure 3B).

Figure 3C shows a very good fit ($R^2 = 0.995$) between the log-transformed corrected absorbance values and the log-transformed water contents. This suggests that the path-length modifications induced by water content at 1450 nm can be well modeled using a simple power law function of water content. The obtained slope of the regression line (1.524) is an estimation of $(a_{\lambda} + 1)$ in eqs 5 and 9.

After running the same analysis to all wavelengths (Figure 4), it can be shown that this law is valid for all wavelengths greater than 1150 nm ($R^2 > 0.9$) (red curve in Figure 4A). The slope values for all regressions (blue curve in Figure 4B) appear to be constant for wavelengths greater than 1350 nm, which allows us to provide a global estimate of $(a_{\lambda} + 1)$ in eqs 4 and 9: $(a_{\lambda} + 1) \approx 1.6$.

Implications for Spectroscopic Assignments on Water Structure. Hence, by properly retrieving the additive and multiplicative effects, the real extinction coefficients (multiplied by a given factor l_0) are obtained ($\epsilon_{\lambda} \cdot l_0$) (exponential of the green curve in Figure 4). This signal can be further analyzed to better understand the water structure, and in particular, to better understand what constitutes the broad first overtone OH band at 1450 nm (green curve in Figure 5). Indeed, by applying the Savitzky–Golay second derivative (an effective deconvolution technique among

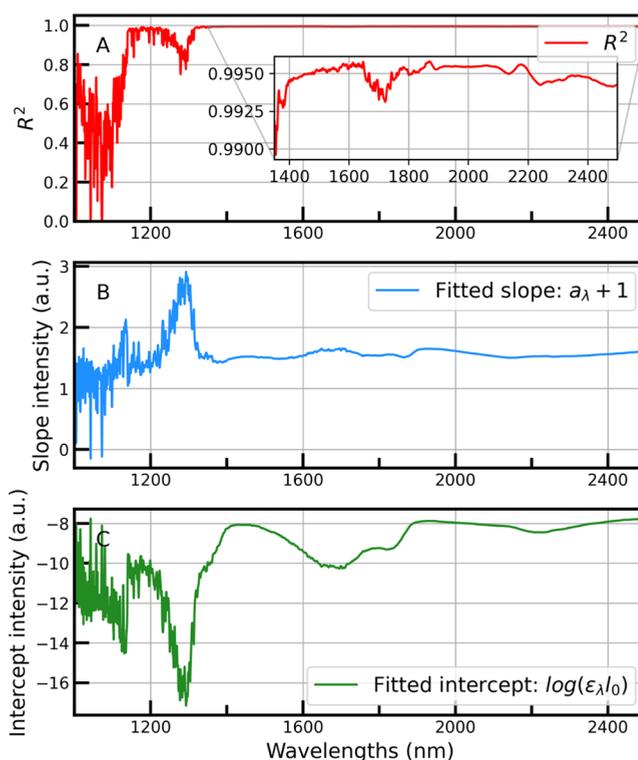


Figure 4. Log–log regression fit results for all wavelengths is presented with (A) the obtained coefficient of determination (R^2), (B) the fitted slope (corresponding to $(a_{\lambda} + 1)$ in eq 9), and (C) the fitted intercept (corresponding to $\log(\epsilon_{\lambda} \cdot l_0)$ in eq 9).

others⁵⁵), two clear negative subpeaks are identified at 1405 and 1469 nm (orange curve in Figure 5). The peak positions correspond perfectly to the identified peaks in other studies on pure water measurements.^{8,56} Other peaks at 1335 and 1537 nm may also be identified and relate well to different water molecular conformations as described by aquaphotomics.¹¹

Implications for Model Calibration with Varying Water Content Samples. It appears that model calibration would benefit from using the observed power law relationship between the absorbance and water content. For example, a log transformation of both spectra X and reference values y would

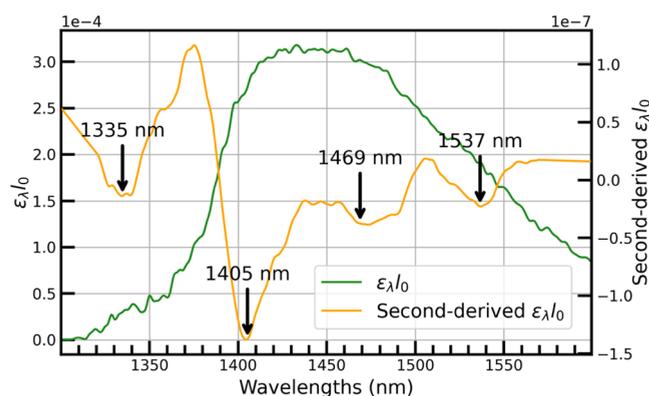


Figure 5. Exponential of the fitted intercept values ($\varepsilon_{\lambda}/l_0$) in eq 9 (in green) and the corresponding Savitzky–Golay second derivative (in orange).

make the relationship between absorbance and water become linear again, therefore enabling better water content prediction models. To showcase this, a comparison of four models is provided in Figure 6 and shows that with one PLS latent

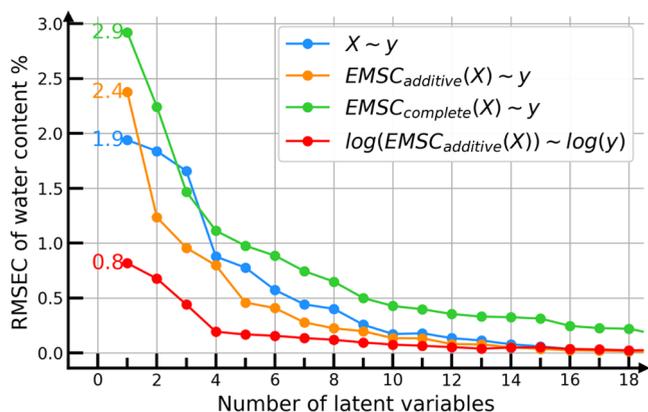


Figure 6. RMSEC curves obtained for four different models: $X \sim y$ (in blue), $EMSC_{\text{additive}}(X) \sim y$ (in orange), $EMSC_{\text{complete}}(X) \sim y$ (in green), and $\log(EMSC_{\text{additive}}(X)) \sim \log(y)$ (in red).

variable, the lowest RMSEC are obtained by a model where X is first corrected from additive effects using EMSC and then log transformation is applied on both X and y to linearize the relationship as in eq 9 (red curves in Figure 6). This shows the strong effect of linearizing the relationship between absorbance and water content: the model is simpler and thus more robust. Indeed, to reach the same prediction error value (0.8% of water content) of the one latent variable model obtained with $\log(EMSC_{\text{additive}}(X)) \sim \log(y)$, the other models $X \sim y$ (blue curve), $EMSC_{\text{additive}}(X) \sim y$ (orange curve), and $EMSC_{\text{complete}}(X) \sim y$ (green curve) need, respectively four, three, and six more latent variables. With sufficient number of latent variables (greater than 15), the models without the log transform appear to follow the same error of prediction as the log-transformed model. This illustrates well how the PLS algorithm is capable of taking into account nonlinear relationships, though it is probable that these models are overfitted. Interestingly, applying a complete EMSC with removal of both additive and multiplicative effects appears very bad in this case for water content prediction (green curves in Figure 6). Indeed, by removing multiplicative effects, the strong relationship between water content and these effects is

lost. Of course, it should be noted that, as discussed in other studies,³¹ applying a logarithm to y is not without consequences: the original distribution of y value is deformed, which may cause problems for regression. However, it appears clear from these results that taking into account the power law relationship between water content and light path length allows a significant improvement of predictive models.

CONCLUSIONS

In most drying experiments, the presence of the dry matter's chemical absorption signature makes it difficult to study independently the scattering modifications induced by water content. By studying the drying of a mixture of aluminum pellets and water, the physical effects induced by water content could be well isolated. A new modeling of path length is proposed in the Bouguer–Beer–Lambert law and consists of relating the path length directly to a power of water content. As shown through the simple use of logarithm to linearize the relationship, this new phenomenological equation could lead to significant enhancements of NIR calibrations on wet scattering media and to further developments of preprocessing algorithms that take specifically into account this relationship. Better consideration of the water content effects in samples containing water could as well lead to a more accurate determination of what regions in NIR spectra are specific to water state information.

AUTHOR INFORMATION

Corresponding Author

Alexandre Mallet – INRAE, Univ Montpellier, LBE, 11100 Narbonne, France; INRAE, UMR ITAP, Montpellier University, 34000 Montpellier, France; Bioentech, 11100 Narbonne, France; ChemHouse Research Group, 34000 Montpellier, France; orcid.org/0000-0003-2497-5550; Email: alexandre.mallet@bioentech.eu

Authors

Roumiana Tsenkova – Biomeasurement Technology Laboratory, Kobe University, 657-8501 Kobe, Japan

Jelena Muncan – Biomeasurement Technology Laboratory, Kobe University, 657-8501 Kobe, Japan

Cyrille Charnier – Bioentech, 11100 Narbonne, France

Eric Latrille – INRAE, Univ Montpellier, LBE, 11100 Narbonne, France; ChemHouse Research Group, 34000 Montpellier, France

Ryad Bendoula – INRAE, UMR ITAP, Montpellier University, 34000 Montpellier, France

Jean-Philippe Steyer – INRAE, Univ Montpellier, LBE, 11100 Narbonne, France

Jean-Michel Roger – INRAE, UMR ITAP, Montpellier University, 34000 Montpellier, France; ChemHouse Research Group, 34000 Montpellier, France; orcid.org/0000-0003-2123-5266

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.analchem.1c00811>

Author Contributions

The manuscript was written through contributions of all authors. CReDit authorship contribution is as follows: Alexandre Mallet: conceptualization, methodology, software, formal analysis, investigation, writing—original draft; Roumiana Tsenkova: conceptualization, methodology, investigation, writing—review & editing, supervision; Jelena Muncan:

conceptualization, methodology, investigation, writing—review & editing; Cyrille Charnier: conceptualization, methodology, writing—review & editing, supervision; Eric Latrille: conceptualization, methodology, writing—review & editing, supervision; Ryad Bendoula: conceptualization, methodology, writing—review & editing, supervision; Jean-Philippe Steyer: conceptualization, methodology, writing—review & editing, supervision; Jean-Michel Roger: conceptualization, methodology, writing—review & editing, supervision. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research work was partly completed in the context of the Agreenium EIR-A program at the Kobe University in Japan. Financial support from 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.

REFERENCES

- (1) Chaplin, M. F. Anomalous Properties of Water, 2020, http://www1.lsbu.ac.uk/water/water_anomalies.html. (accessed Mar 15, 2020).
- (2) Nostro, P.; Lo, Ninham, B. W. *Aqua Incognita: Why Ice Floats on Water and Galileo 400 Years On*; Connor Court Publishing, 2014.
- (3) Chaplin, M. F. *Encycl. Water* **2019**, 1–19.
- (4) Maeda, H.; Ozaki, Y.; Tanaka, M.; Hayashi, N.; Kojima, T. *J. Near Infrared Spectrosc.* **1995**, 3, 191–201.
- (5) Cui, X.; Cai, W.; Shao, X. *RSC Adv.* **2016**, 6, 105729–105736.
- (6) Gowen, A. A.; Amigo, J. M.; Tsenkova, R. *Anal. Chim. Acta* **2013**, 759, 8–20.
- (7) Segtnan, V. H.; Šašić, Š.; Isaksson, T.; Ozaki, Y. *Anal. Chem.* **2001**, 73, 3153–3161.
- (8) Renati, P.; Kovacs, Z.; De Ninno, A.; Tsenkova, R. *J. Mol. Liq.* **2019**, 292, No. 111449.
- (9) Wilse Robinson, G. W.; Cho, C. H.; Urquidi, J. *J. Chem. Phys.* **1999**, 111, 698–702.
- (10) Geissler, P. L. *J. Am. Chem. Soc.* **2005**, 127, 14930–14935.
- (11) Tsenkova, R. *J. Near Infrared Spectrosc.* **2009**, 17, 303–313.
- (12) Muncan, J.; Tsenkova, R. *Molecules* **2019**, 24, 2742.
- (13) Kojić, D.; Tsenkova, R.; Tomobe, K.; Yasuoka, K.; Yasui, M. *ChemPhysChem* **2014**, 15, 4077–4086.
- (14) Tan, J.; Sun, Y.; Ma, L.; Feng, H.; Guo, Y.; Cai, W.; Shao, X. *Chemom. Intell. Lab. Syst.* **2020**, 206, No. 104150.
- (15) Mayerhöfer, T. G.; Pahlow, S.; Popp, J. *ChemPhysChem* **2020**, DOI: 10.1002/cphc.202000464.
- (16) Tamburini, E.; Vincenzi, F.; Costa, S.; Mantovi, P.; Pedrini, P.; Castaldelli, G. *Sensors* **2017**, 17, No. 2366.
- (17) Lobell, D. B.; Asner, G. P. *Soil Sci. Soc. Am. J.* **2002**, 66, 722.
- (18) Barnes, R. J.; Dhanoa, M. S.; Lister, S. J. *Appl. Spectrosc.* **1989**, 43, 772–777.
- (19) Isaksson, T.; Naes, T. *Appl. Spectrosc.* **1988**, 42, 1273–1284.
- (20) Ilari, J. L.; Martens, H.; Isaksson, T. *Appl. Spectrosc.* **1988**, 42, 722–728.
- (21) Martens, H.; Jensen, S. A.; Geladi, P. In *Multivariate Linearity Transformation for Near-Infrared Reflectance Spectrometry*, Proceedings of the Nordic Symposium on Applied Statistics, 1983; pp 205–234.
- (22) Afseth, N. K.; Kohler, A. *Chemom. Intell. Lab. Syst.* **2012**, 117, 92–99.
- (23) Martens, H.; Stark, E. *J. Pharm. Biomed. Anal.* **1991**, 9, 625–635.
- (24) Martens, H.; Nielsen, J. P.; Engelsen, S. B. *Anal. Chem.* **2003**, 75, 394–404.
- (25) Chen, Z. P.; Morris, J.; Martin, E. *Anal. Chem.* **2006**, 78, 7674–7681.
- (26) Jin, J.; Chen, Z.; Li, L.; Steponavicius, R.; Thennadil, S. N.; Yang, J.; Yu, R. *Anal. Chem.* **2012**, 84, 320–326.
- (27) Li, L.; Peng, Y.; Yang, C.; Li, Y. *Postharvest Biol. Technol.* **2020**, 162, No. 111101.
- (28) Guo, Q.; Wu, W.; Massart, D. L. *Anal. Chim. Acta* **1999**, 382, 87–103.
- (29) Dieterle, F.; Ross, A.; Schlotterbeck, G.; Senn, H. *Anal. Chem.* **2006**, 78, 4281–4290.
- (30) Fischler, M. A.; Bolles, R. C. *Commun. ACM* **1981**, 24, 381–395.
- (31) Rabatel, G.; Marini, F.; Walczak, B.; Roger, J. M. *J. Chemom.* **2020**, 34, No. e3164.
- (32) Martens, H.; Høy, M.; Wise, B. M.; Bro, R.; Brockhoff, P. B. *J. Chemom.* **2003**, 17, 153–165.
- (33) Roger, J.-M.; Chauchard, F.; Bellon-Maurel, V. *Chemom. Intell. Lab. Syst.* **2003**, 66, 191–204.
- (34) Minasny, B.; McBratney, A. B.; Bellon-Maurel, V.; Roger, J. M.; Gobrecht, A.; Ferrand, L.; Joalland, S. *Geoderma* **2011**, 167–168, 118–124.
- (35) Acharya, U. K.; Walsh, K. B.; Subedi, P. P. *J. Near Infrared Spectrosc.* **2014**, 22, 279–286.
- (36) Savitzky, A.; Golay, M. J. E. *Anal. Chem.* **1964**, 36, 1627–1639.
- (37) Clark, R. N.; Roush, T. L. *J. Geophys. Res.: Solid Earth* **1984**, 89, 6329–6340.
- (38) Leung, A. K. M.; Chau, F. T.; Gao, J. Bin. *Anal. Chem.* **1998**, 70, 5222–5229.
- (39) Giguere, S.; Boucher, T.; Carey, C. J.; Mahadevan, S.; Dyar, M. D. *Appl. Spectrosc.* **2017**, 71, 1457–1470.
- (40) Roger, J.-M.; Biancolillo, A.; Marini, F. *Chemom. Intell. Lab. Syst.* **2020**, 199, No. 103975.
- (41) Mishra, P.; Roger, J. M.; Marini, F.; Biancolillo, A.; Rutledge, D. N. *Chemom. Intell. Lab. Syst.* **2020**, No. 104190.
- (42) Oliveri, P.; Malegori, C.; Simonetti, R.; Casale, M. *Anal. Chim. Acta* **2019**, 1058, 9–17.
- (43) Mallet, A.; Charnier, C.; Latrille, É.; Bendoula, R.; Steyer, J.-P.; Roger, J.-M. *Waste Manag.* **2021**, 122, 36–48.
- (44) Cumming, J. B. *Nucl. Instrum. Methods Phys. Res., Sect. A* **2013**, 713, 1–4.
- (45) van Rossum, G.; Drake, F. L. *Python 3 Reference Manual*; CreateSpace: ScottsValley, CA, 2009; <https://doi.org/10.5555/1593511>.
- (46) Oliphant, T. E. *Guide to NumPy*; Trelgol Publishing: USA, 2010; Vol. 1.
- (47) 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.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, I.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P. *Nat. Methods* **2020**, 17, 261–272.
- (48) Hunter, J. D. *Comput. Sci. Eng.* **2007**, 9, 90–95.
- (49) Gobrecht, A.; Roger, J. M.; Bellon-Maurel, V. *Adv. Agron.* **2014**, 145–175.
- (50) Næs, T.; Martens, H. *TrAC, Trends Anal. Chem.* **1984**, 3, 266–271.
- (51) Wold, H. *Nonlinear Iterative Partial Least Squares (NIPALS) Modelling: Some Current Developments*; Academic press, Inc., 1973; <https://doi.org/10.1016/b978-0-12-426653-7.50032-6>.
- (52) Snee, R. D. *Technometrics* **1977**, 19, 415–428.
- (53) Luck, W. A. *Structure of Water and Aqueous Solutions*; Verlag Chemie, 1974.
- (54) Williams, P. J. *J. Near Infrared Spectrosc.* **2009**, 17, 315–328.
- (55) Shao, X.; Cui, X.; Wang, M.; Cai, W. *Spectrochim. Acta, Part A* **2019**, 213, 83–89.
- (56) Segtnan, V. H.; Šašić, Š.; Isaksson, T.; Ozaki, Y. *Anal. Chem.* **2001**, 73, 3153–3161.

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

Full reference:

Mallet, A.; Pérémé, M.; Awhangbo, L.; Charnier, C.; Roger, J.-M.; Steyer, J.-P.; Latrille, É.; Bendoula, R. Fast At-Line Characterization of Solid Organic Waste: Comparing Analytical Performance of Different Compact near Infrared Spectroscopic Systems with Different Measurement Configurations. *Waste Manag.* **2021**, *126*, 664–673. <https://doi.org/10.1016/j.wasman.2021.03.045>.



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



Alexandre Mallet^{a,b,c,d,*}, Margaud Pérémé^{a,d}, Lorraine Awhangbo^{a,d}, Cyrille Charnier^c, Jean-Michel Roger^{b,d}, Jean-Philippe Steyer^a, Éric Latrille^{a,d}, Ryad Bendoula^b

^a INRAE, Univ Montpellier, LBE, 102 Av des Etangs, Narbonne F-11100, France

^b INRAE, UMR ITAP, Montpellier University, Montpellier, France

^c Bioentech, F-11100 Narbonne, France

^d ChemHouse Research Group, Montpellier, France

ARTICLE INFO

Article history:

Received 25 January 2021

Revised 23 March 2021

Accepted 25 March 2021

Available online xxx

Keywords:

Near infrared spectroscopy
Anaerobic digestion
Process monitoring
Biochemical methane potential
Compact systems
Measurement modes

ABSTRACT

Fast characterization of solid organic waste using near infrared spectroscopy 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 of applicability and high costs. Recent evolutions in the technology have given rise to both more compact and more modular low-cost near infrared systems which could allow a larger scale deployment. The current study investigates the relevance of these new systems by evaluating four different Fourier transform near-infrared spectroscopic systems with different compactness (laboratory, portable, micro spectrometer) but also different measurement configurations (polarized light, at distance, in contact). Though the conventional laboratory spectrometer showed the best performance on the various biochemical parameters tested (carbohydrates, lipids, nitrogen, chemical oxygen demand, biochemical methane potential), the compact systems provided very close results. Prediction of the biochemical methane potential was possible using a low-cost micro spectrometer with an independent validation set error of only 91 NmL(CH₄).gTS⁻¹ compared to 60 NmL(CH₄).gTS⁻¹ for a laboratory spectrometer. The differences in performance were shown to result mainly from poorer spectral sampling; and not from instrument characteristics such as spectral resolution. Regarding the measurement configurations, none of the evaluated systems allowed a significant gain in robustness. In particular, the polarized light system provided better results when using its multi-scattered signal which brings further evidence of the importance of physical light-scattering properties in the success of models built on solid organic waste.

© 2021 Elsevier Ltd. All rights reserved.

1. Introduction

In anaerobic digestion processes, different organic waste are often co-digested to enhance the production of both biogas and fertilizers (Hagos et al., 2017). A tremendous diversity of waste is concerned by these bioprocesses such as agricultural residues (animal manure, crop stems/stalks, silage), food industry waste (brewery, sugar refinery), urban solid waste, meat waste or catering waste. This implies that these waste cover a large range of biochemical composition and physical properties. Moreover, such

properties may fluctuate according to factors such as crop seasonality, transport or storage. This brings important challenges for ensuring the stability of the process and the efficiency of biogas production in digesters (Wu et al., 2019). To answer this, online monitoring of the feeding substrate quality could allow the direct adaptation of the feeding strategy to the feeding substrate 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.

* Corresponding author at: INRAE, Univ Montpellier, LBE, 102 Av des Etangs, Narbonne F-11100, France.

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

¹ Full postal address: INRAE-LBE, 102, Avenue des Etangs, F-11100 Narbonne.

Near infrared spectroscopy (NIRS), coupled with multivariate analysis techniques (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 process, NIRS was used to monitor the degradation phases of compost (Albrecht 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 was initially used for in-situ monitoring of dry solids (DS), volatile solids (VS), chemical oxygen demand (COD) and volatile fatty acids (VFA) in digesters (Jacobi et al., 2009; Lomborg et al., 2009; Stockl and Lichti, 2018; Wolf et al., 2011). NIRS was then proposed for the determination of biochemical methane potential (BMP) 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; Triolo et al., 2014; Yang et al., 2021). Today, NIRS appears the most suitable method for predicting BMP on various organic substrates (Rodriguez et al., 2019). More recently, NIRS was used to estimate complementary characteristics such as carbohydrates content, lipid content, nitrogen content, COD, and kinetic parameters (Charnier et al., 2017). In terms of process monitoring, these developments allow time-consuming reference measurements (which last typically one to two 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 and high costs (Wu et al., 2019). Indeed, freeze-drying and grinding steps are necessary to avoid water and particle size effects in NIRS (Mallet et al., 2021), which currently limits the online applicability of such system. Moreover, the high costs of 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 NIRS is applied directly on fresh waste or with a prior freeze-drying step, there is a need to develop cheap and reliable instruments which can be used on a wide range of substrate types in order to promote a greater adoption of NIRS in co-digestion plants. This could be addressed by an at-site use of low-cost and compact near infrared (NIR) systems.

In the past few years, the use of NIRS has developed out of laboratories, thanks to important progress in the miniaturization of instruments. In particular, handheld Fourier transform near infrared (FT-NIR) micro spectrometers have appeared in the market, and make use of a micro-electro-mechanical systems-based (MEMS) Michelson interferometer (Beć et al., 2021). While conventional Michelson interferometers are made of discrete elements (including the moving mirror actioned by a motor, the fixed mirror, and beam splitter), MEMS technology enables a monolithic integration of these elements on a single chip, with the particularity that the moving mirror is operated by an electrical signal. Amongst the spectrometers making use of this technology, the NeoSpectra instrument has shown good analytical performance results for soil organic and total carbon content characterization (Shariffar et al., 2019; Tang et al., 2020), or authenticity screening in food (McVey et al., 2021). These compact spectrometers allow the measurements to be performed on site, thanks to their compactness, robustness and cost. However, these compact portable spectrometers tend to have poorer instrument performances than laboratory spectrometer, with lower resolution, spectral range, and signal-to-noise ratio (Beć et al., 2020; Crocombe, 2018). Therefore, the suitability of such systems for the characterization of diverse organic waste still needs to be assessed.

Another aspect of these compact systems concerns their modularity and the possibility of testing different measurement configurations, in order to enhance the measured signal. Indeed, in complex matter such as solid organic waste, the Bouguer-Beer-Lambert law does not hold due to important light scattering (Dahm and Dahm, 2004). To answer this, spectral pre-processing

has been proposed to remove both additive and multiplicative effects (Rabatel et al., 2020; Rinnan et al., 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 promising optical pre-processing method, based on polarized light spectroscopy (Backman et al., 1999) has been proposed to improve the absorbance signal measurement on such scattering samples (Bendoula et al., 2015; Gobrecht et al., 2015; Xu et al., 2019). Such system has shown analytical performance improvements for soils (Gobrecht et al., 2016), and more recently, for digestate (Awhangbo et al., 2020), but has never been evaluated on solid organic waste. Still in the aim of enhancing the measured spectra, time-resolved spectroscopy shows a promising future : applied to pharmaceutical tablets, collected photons with a particular propagation time were shown to be most informative for quantification (Alayed and Deen, 2017; Johansson et al., 2002). However, the cost of this technology still remains prohibitive for the organic waste management sector. Finally, the measurement mode (at distance or in contact, in reflectance or in interactance) also plays an important role in the final accuracy for estimating biochemical properties (Hemrattrakun et al., 2021; Khodabakhshian et al., 2019; Schaare and Fraser, 2000). Though current used laboratory spectrometers make use of a distance reflectance measurement, a contact immersed probe measurement has been shown to be useful for prediction of parameters on digestates (Awhangbo et al., 2020). Authors observe higher reflectance levels with less noise in the collected spectra, as well as new chemical features which were not apparent in a remote probe configuration. In light of this, it appears that the use of different measurement configurations could enable the calibration of more accurate and more robust NIRS models on diverse solid organic waste.

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

2. Materials and methods

2.1. Sample preparation and reference analyses

Thirty-three substrates were selected amongst various waste types that have been collected in rural, territorial and industrial anaerobic digestion plants in France. These substrates cover a wide range of biochemical and physical properties: solid cellulosic waste (like silage, cereals and corn cobs), liquid cellulosic suspensions (such as manure), liquid fat suspensions (catering waste or bio-waste), sweet emulsions (such as lactoserum), or protein and fat solid pastes (such as egg waste or cacao butter). The visual aspect of some of these substrates in raw form is presented in Appendix A.

For spectral measurements, each substrate sample was freeze-dried and ground to 1 mm. The dataset is fully described in a data paper and available online [<https://doi.org/10.15454/SQQTUU>].

Biochemical characterization of substrates was obtained by using a NIRS calibrated model (Charnier et al., 2017), with errors on independent test sets of 53 mg(O₂).gTS⁻¹ for carbohydrates content, 3.2*10⁻² g.gTS⁻¹ for lipids content, 8.6 mg.gTS⁻¹ for nitrogen content, and 83 mg(O₂).gTS⁻¹ for COD. The histograms of obtained prediction values are presented in Fig. 1.

2.2. Spectroscopic systems

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

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⁻¹ to 10 000 cm⁻¹ (1000–2500 nm) and a resolution of 4 cm⁻¹. An external white reference (Spectralon®) signal $I_0(\lambda)$ is automatically taken every 10 min. For each sample, an intensity signal $I(\lambda)$ was collected, and the pseudo-absorbance signal $A_{lab}(\lambda)$ was computed:

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

2.2.2. Portable spectrometer with immersed contact probe

The immersed contact probe system consists of a FT-NIR Rocket spectrometer (Arcoptix, Neuchatel, Switzerland) scanning in reflectance mode with a spectral range of 3 800 cm⁻¹ to 11 000 cm⁻¹ (900–2500 nm) and a resolution of 4 cm⁻¹. The spectrometer was connected to two optical fibers (for illumination and signal collection) of 1000 μm core diameter and numerical aperture of 0.39 (BFY1000, Thorlabs). A tungsten-halogen source (Ocean Optics HL-200-FHSA) was used for illumination. For each sample, the intensity $I(\lambda)$ was collected. A white reference (SRS99, Spectralon®) was scanned every hour during the measurements resulting in $I_0(\lambda)$. Finally, a dark current signal $I_n(\lambda)$ corresponding to the instrumental noise was recorded and subtracted to all spectra. A pseudo-absorbance signal $A_{ip}(\lambda)$ was thus calculated:

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

2.2.3. Portable spectrometer with polarized light spectroscopy

The polarized light system consists of the same elements (spectrometer, light source, optical fibers) as the immersed contact probe system, however, measurements were made at a distance of 5 cm from the samples, and a polarized light component (Awhangbo et al., 2020) was connected to the spectrometer. This 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 reflected light in an s-polarized and p-polarized image, corresponding to parallel $I_{\parallel}(\lambda)$ 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 collected. Three signals were then calculated following Bendoula et al. (2015) formula (Bendoula et al., 2015): the single scattering reflectance $R_{ss}(\lambda)$, the multiple scattering reflectance $R_{ms}(\lambda)$, and the total backscattering reflectance $R_{bs}(\lambda)$:

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

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

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

2.2.4. Handheld micro spectrometer

The micro spectrometer system consists of a MEMS FT-NIR NeoSpectra spectrometer (Si-Ware, Cairo, Egypt) scanning in reflectance mode with a spectral range of 3 921 cm⁻¹ to 7 407 cm⁻¹ (1350–2550 nm) and a resolution of 66 cm⁻¹. 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:

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

2.3. Data analysis: Model calibration

All the data analysis was performed using Python 3.6.5: data wrangling with Pandas 0.25.1, NumPy 1.17.3, SciPy 1.3.1, Scikit-learn 0.21.3, and plotting with Matplotlib 2.2.2 (Harris et al., 2020; Hunter, 2007; McKinney, 2010; Pedregosa et al., 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 immersed 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_{μ} 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.

In order to evaluate the models built on each spectroscopic system, a validation test set was constituted. With the aim of producing a representative validation test set, the Duplex algorithm (Snee, 1977) was run for each reference characteristic (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. To assess the representativeness of the validation test set in terms of spectral variability, a principal components analysis (Cordella, 2012) was done, and obtained scores were plot in Fig. 2 and Appendix C.

Models were built using a partial least squares regression (PLS1-R) with NIPALS algorithm (Næs and Martens, 1984; WOLD, 1973). To determine the number of latent variables, a cross-validation was done using a repeated randomized 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 independence. For each cross-validation run, various metrics were then calculated: the root-mean-square error (RMSE), the mean absolute error (MAE) (Willmott and Matsuura, 2005), the coefficient of determination (R^2), and B-coefficients

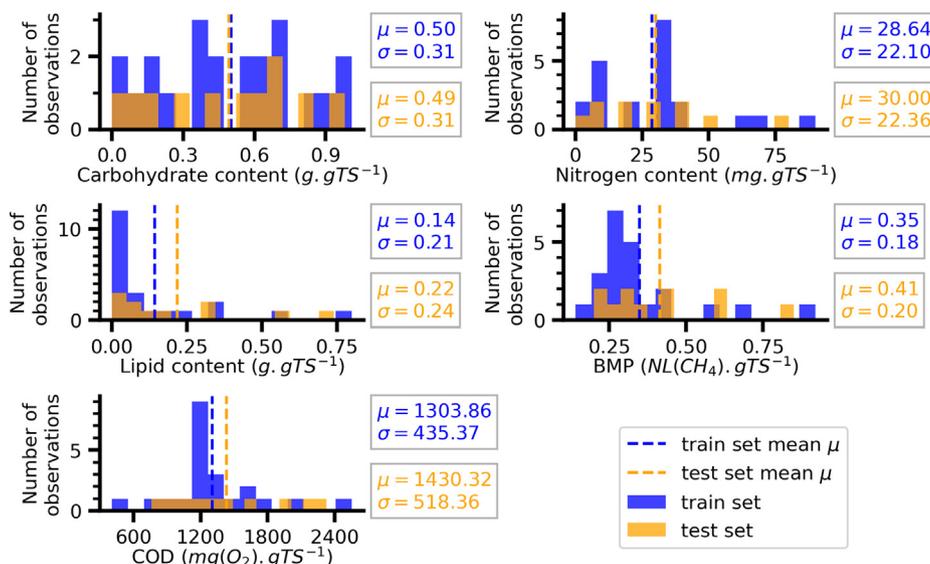


Fig. 1. Histograms of reference characteristics. Both train and test sets (respectively in blue and in orange) obtained by a Duplex split are presented. Respective mean (labeled as μ) and standard deviation (labeled as σ) are presented for train and test sets (respectively in blue, and in orange). Dotted lines represent respective mean values. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Spectral measurement protocol characteristics.

	Laboratory spectrometer	Immersed contact probe	Polarization system	Micro spectrometer
Measurement replicates	3	3	3	3
Number of scans per measurement	96 scans	10 scans	10 scans	28 s scan time
Measurement sampling method (measured area per scan)	360° rotation (~5 cm ²)	Fixed point (~0.05 cm ²)	Fixed point (~1 cm ²)	Fixed point (~25 cm ²)
Protocol between replicates	Mix the whole sample	Change the measured surface position	Change the measured surface position	Change the measured surface position

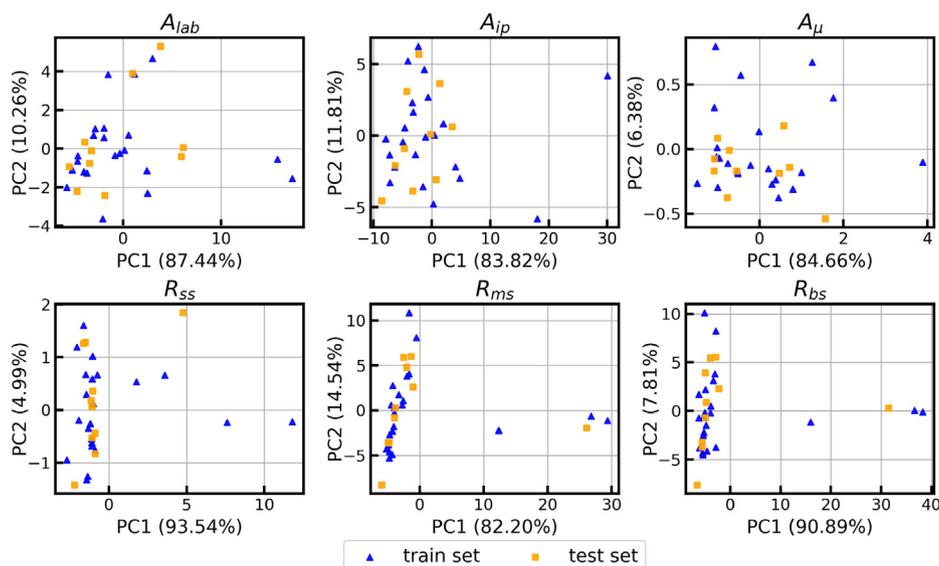


Fig. 2. Principal component analysis (PCA) score plots of train and test sets (respectively in blue triangles, and in orange squares) for each signal type (from the four spectroscopic configurations). Each subplot represents the scores of the first and second principal components (PC1 and PC2). The percentage 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. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

metrics which are the Durbin-Watson statistic (DW) and the variance (Rutledge and 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 R², and detecting rate increase of DW and variance of B-coefficients).

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.

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:

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

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

$$S_r = \sqrt{\sum_{i=1}^{n_{\text{test_samples}}} s_{r_i}^2} \quad (\text{Eq.8})$$

3. Results & discussion

3.1. Data overview

3.1.1. Training set and validation test set

For each reference characteristic, train and test set distributions (respectively in blue and orange) are presented as histograms in Fig. 1. Very similar distributions (same mean, same standard deviation) for all characteristics show that the Duplex algorithm succeeded in obtaining a representative test set in terms of biochemical 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 type in Fig. 2 (only the first and second component scores are displayed, but scores up to the tenth component were checked). As shown, the test set signals (in orange) cover most of the range covered by train set signals. However, in some cases, the variability of train set signals is not fully well represented in the test set. For example, in the upper left score plot representing the laboratory spectrometer configuration signal A_{lab} , no test set signal (orange square) is found in the far right plot where there are two train set signals (blue triangles). This is consistent with the fact that the Duplex algorithm was run on the reference values and not on the spectral values, so there is no guarantee for test set spectral representativeness. Although this is not optimal for evaluating calibration models alone, such methodology appeared to be the best to compare different spectroscopic systems on identical samples without bias.

3.1.2. Raw spectra analysis

Fig. 3 presents the raw reflectance signals obtained with each spectroscopic system. In all signals, the main peaks found in organic waste were apparent: the CH, CH₂ and CH₃ combination bands particularly present in fat (1731 nm, 1764 nm, 2310 nm, 2350 nm), the OH bands present in simple sugars (1436 nm, 1932 nm), the OH combination bands in starch or cellulose (2092 nm), and the NH combination bands present in proteins (2180 nm) (Williams and Antoniszyn, 2019; Workman and Weyer, 2012). However, the relative amplitude of these peaks seems to differ. For example, in the micro spectrometer signal (R_{μ}) the CH₂ combination bands at 2310 nm and 2350 nm seem much less sharp than in the laboratory spectrometer signal (R_{lab}). This can be well explained by the lower resolution of the micro-spectrometer (66 cm⁻¹) compared to the laboratory spectrometer (4 cm⁻¹). A consequence of this is that the models built on compact systems such as the micro spectrometer will be based on more simple features, which could lead them to be less accurate models but potentially also more robust.

In addition, the sensitivity with respect to the spectral range appears to differ from one spectrometer to another. For the micro spectrometer, the measured signal below 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_{ip}) or polarized signals (R_{ss} , R_{ms} , R_{bs}), 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 (Fig. 4). The reflectance levels of R_{lab} , R_{bs} , and R_{μ} are much higher (with 75% of the values that range respectively between 0.46 and 0.73, 0.43 and 0.66, and 0.60 and 0.90), than the reflectance levels of R_{ip} (with 75% of the values that range between 0.15 and 0.35). This is mostly related to the way the signal is acquired (i.e. the measurement configuration). Indeed, reflectance levels are the result of both the absorption level (dependent of chemical composition) and the scattering level (modifications of light optical path-length, and photon leakage (Gobrecht et al., 2014)). Therefore, the chosen measurement configuration might favour one or the other, leading to differences in the measured reflectance levels. Results show here that the distant mode systems (i.e. R_{lab} , R_{bs} , and R_{μ}) collect much more scattering photons than the contact mode system (i.e. R_{ip}). Regarding the advantage of one measurement mode over the other, this will mostly be dependent on the characteristic to be predicted, and its dependency on physical properties.

In the polarized light system, a clear difference of reflectance level can also be observed: 75% of the values of the multiple scattering signal R_{ms} range between 0.52 and 0.76, against 0.05 and 0.11 for the single scattering signal R_{ss} . This is consistent with the sole principle of polarized spectroscopy where R_{ss} corresponds to single scattering photons with low penetration in the media while R_{ms} 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.

3.2. Model performances

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

For prediction models built using the laboratory spectrometer system, the errors obtained on the test set (RMSEP) were of 0.108 g.gTS⁻¹ for carbohydrates content, 5.8 mg.gTS⁻¹ for nitrogen content, 0.034 g.gTS⁻¹ for lipids content, 0.060 NL(CH₄).gTS⁻¹ for BMP, and 136.4 mg(O₂).gTS⁻¹ for COD. These are all consistent to the performances of reference models (Charnier et al., 2017). The slightly lower performances obtained can be explained by the more limited number of samples on which these models were built (22 samples) compared to the original models (about 80 samples).

For all predicted characteristics, the laboratory system (A_{lab}) showed better analytical performance results than the compact systems (A_{ip} , A_{μ} , R_{ss} , R_{ms} , R_{bs}). However, in many cases, these latter systems showed similar performances to the laboratory system. For example, for BMP prediction, the prediction error (RMSEP) of the model obtained with the laboratory spectrometer signal A_{lab} was 60 mL(CH₄).gTS⁻¹. In comparison, for the polarized system signals R_{ss} , R_{ms} , R_{bs} , the model prediction errors were of respectively 115 mL(CH₄).gTS⁻¹, 111 mL(CH₄).gTS⁻¹ and 100 mL(CH₄).gTS⁻¹, while for the micro spectrometer signal A_{μ} the model's error was of only 91 mL(CH₄).gTS⁻¹. Similarly, for carbohydrates content

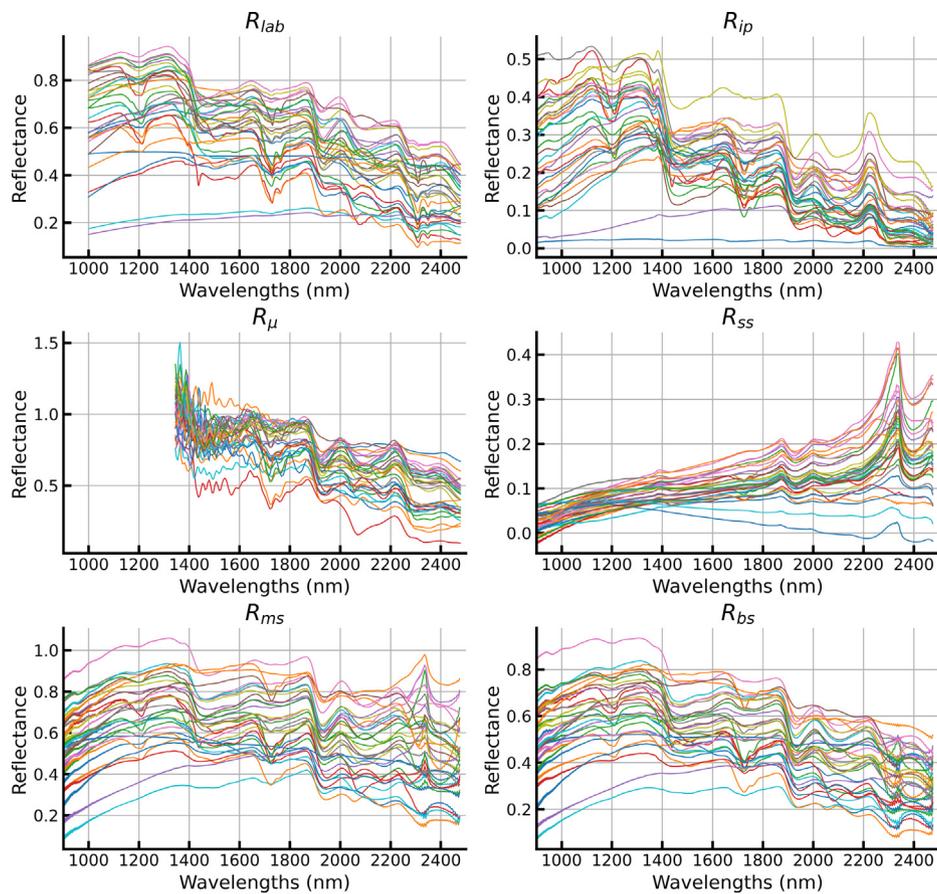


Fig. 3. Raw reflectance spectra of each spectroscopic system (R_{lab} : laboratory spectrometer, R_{ip} : immersed probe system, R_{μ} : micro-spectrometer, R_{ss} : single scattered signal of polarized system, R_{ms} : multiple scattered signal of polarized system, R_{bs} : total back-scattered signal of polarized system). Each spectrum corresponds to the mean of the triplicate measurements.

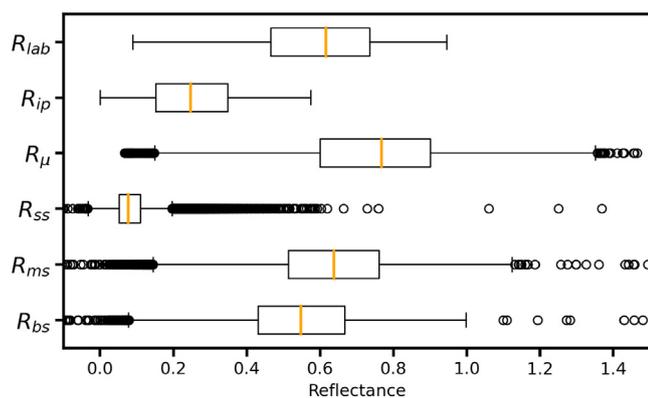


Fig. 4. Boxplots of raw reflectance spectra values for each signal type (R_{lab} : laboratory spectrometer, R_{ip} : immersed probe system, R_{μ} : micro-spectrometer, R_{ss} : single scattered signal of polarized system, R_{ms} : multiple scattered signal of polarized system, R_{bs} : total back-scattered signal of polarized system). Each boxplot was obtained on the flattened matrix (reflectance values for all samples and for all wavelengths). Median values are presented with orange lines. The box limits represent the first and third quartile values (respectively Q1 and Q3), and the lines that extend from the box show the lowest and largest data points excluding any outliers (respectively $Q1 - 1.5 \times (Q3 - Q1)$ and $Q1 + 1.5 \times (Q3 - Q1)$). Outliers are presented in empty black circles. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

prediction, while the prediction error for the laboratory spectrometer was 0.108 g.gTS^{-1} , the errors for the micro-spectrometer and the immersed probe system were of only 0.134 g.gTS^{-1} and 0.104 g.gTS^{-1} . Such observation can be made

for all the other characteristics studied. This is very promising because these models have acceptable errors compared to the laboratory spectrometer. This means that the models built on organic waste rely on sufficiently simple features so that the lower spectral resolution of compact spectrometers does not affect too much the performances. Such result is consistent with similar studies on low-cost compact NIR spectrometers with limited spectral range that are applied to herbaceous feedstock such as corn stover or sorghum (Wolfrum et al., 2020). Knowing that these spectrometers are low-cost (while the laboratory spectrometer costs about 50 000€, the immersed contact probe and polarized systems each cost about 20 000€ and the micro spectrometer only costs about 3 000 €), and can be used at-site, this holds great promises regarding an increased adoption of NIRS for robust solid organic waste characterization in anaerobic digestion plants.

Models built on the three signals R_{ss} , R_{ms} , R_{bs} obtained using the polarized spectroscopy system show different performances. While R_{bs} is a signal very similar to those obtained from the other set-ups because it includes information from all the backscattering light, R_{ss} and R_{ms} differ in terms of type of photons that are captured by the spectrometer (respectively single scattering photons, and multiple scattering photons). For all characteristics, models built on R_{bs} signal show better performances than models built on the R_{ss} and R_{ms} signals. For example, for COD prediction, while the prediction error (RMSEP) of the model using the total back-scattering signal (R_{bs}) is of $129.2 \text{ mg(O}_2\text{).gTS}^{-1}$, it is of $147.8 \text{ mg(O}_2\text{).gTS}^{-1}$ using R_{ms} and $273.9 \text{ mg(O}_2\text{).gTS}^{-1}$ using R_{ss} . This suggests that the use of polarized spectroscopy for predicting these characteristics on organic waste may not be particularly recommended.

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_{lab} , A_{ip} , A_{μ} , R_{ss} , R_{ms} , R_{bs}) 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^2_C , R^2_P).

Experiment Reference	Signal	Optimized parameters Spectral range (nm)	Pretreatment	#LV	Performance metrics				
					RMSEC	RMSECV	RMSEP	R^2_C	R^2_P
Carbohydrates $g.gTS^{-1}$	A_{lab}	1400–2240	SNV + SG1	2	0.143	0.192	0.108	0.78	0.83
	A_{ip}	1450–2240	SNV	2	0.188	0.169	0.104	0.82	0.84
	A_{μ}	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_{ms}	1200–2240	Raw	4	0.121	0.185	0.112	0.82	0.82
	R_{bs}	1300–2300	SNV + SG1	3	0.119	0.158	0.130	0.74	0.91
Nitrogen $mg.gTS^{-1}$	A_{lab}	1200–2240	SNV + SG1	10	3.6	15.0	5.8	0.98	0.89
	A_{ip}	1200–2240	SNV + SG1	5	9.9	18.4	7.5	0.87	0.83
	A_{μ}	1400–2500	SNV	5	10.1	20.7	13.1	0.85	0.51
	R_{ss}	1200–2240	SG2	3	17.2	23.1	21.4	0.62	−0.4
	R_{ms}	1200–2240	SNV + SG1	5	12.6	22.6	12.1	0.80	0.54
	R_{bs}	1300–2300	SG1 + SNV	4	10.7	17.6	11.7	0.83	0.61
Lipids $g.gTS^{-1}$	A_{lab}	1400–2240	SNV	6	0.025	0.056	0.034	0.99	0.98
	A_{ip}	1200–2240	VSN	9	0.057	0.066	0.066	0.99	0.93
	A_{μ}	1400–2500	SG1 + SNV	6	0.039	0.081	0.067	0.98	0.92
	R_{ss}	1200–2240	Raw	5	0.084	0.110	0.124	0.91	0.74
	R_{ms}	1200–2240	SG2	5	0.029	0.059	0.068	0.99	0.92
	R_{bs}	1200–2480	SG2	6	0.029	0.066	0.066	0.99	0.93
BMP $NL(CH_4).gTS^{-1}$	A_{lab}	1400–2240	SNV + SG1	9	0.020	0.078	0.060	0.99	0.90
	A_{ip}	1200–2240	SG1	2	0.088	0.115	0.110	0.77	0.73
	A_{μ}	1600–2500	SG2	2	0.102	0.109	0.091	0.74	0.82
	R_{ss}	1200–2240	SG1	4	0.085	0.115	0.115	0.79	0.71
	R_{ms}	1350–2240	SG1	3	0.087	0.115	0.111	0.78	0.73
	R_{bs}	1300–2300	SG1	6	0.046	0.120	0.100	0.94	0.78
COD $mg(O_2).gTS^{-1}$	A_{lab}	1400–2240	SNV	6	63.0	186.8	136.4	0.97	0.94
	A_{ip}	1450–2240	SNV + SG1	7	106.2	228.3	160.2	0.93	0.90
	A_{μ}	1400–2500	SG1 + SNV	3	180.4	303.1	196.9	0.81	0.85
	R_{ss}	1200–2240	SG2	4	192.0	247.4	273.9	0.78	0.72
	R_{ms}	1350–2240	SG1	3	112.5	151.4	147.8	0.92	0.92
	R_{bs}	1200–2480	SG1 + SNV	4	128.1	237.6	129.2	0.90	0.94

In fact, for lipids content, COD and nitrogen content, the models built with R_{ss} signals show much greater errors than models built with R_{ms} signals (respectively 0.124 $g.gTS^{-1}$ greater than 0.068 $g.gTS^{-1}$, 273.9 $mg(O_2).gTS^{-1}$ greater than 147.8 $mg(O_2).gTS^{-1}$, and 21.4 $mg.gTS^{-1}$ greater than 12.1 $mg.gTS^{-1}$). This came as a surprise, as the single scattering signal is theoretically supposed to be more related to absorbing constituents and less impacted by scattering effects (Gobrecht et al., 2015). This can be explained by the fact that for a dataset with such diverse solid organic substrates, the biochemical composition is closely related to the physical properties. For example, high lipid content substrates (which also correspond to substrates with high COD) tend to form liquid transparent solutions (like oil), which transmit light much more than low lipid content substrates which usually form highly scattering porous media. This relatively poor performance obtained with the single scattering signal is consistent with results obtained on digestates where physical structure appeared determinant (Awhangbo et al., 2020). One additional reason can be found in the measurement technique itself: the intensity captured for the single scattering signal is much more limited than a classical total backscattering signal, leading to higher signal-to-noise ratios. Further investigations could concentrate on the use of a multi-block approach combining these three complementary signals, as proposed on digestates (Awhangbo et al., 2020).

Apart from the single scattering signal, all signals allowed to build satisfactory models for the biochemical characterization of organic waste. However, no spectroscopic system allowed to surpass the analytical performance of the laboratory spectrometer system. Fig. 5 presents the observed and prediction plots for each signal for the prediction of biochemical potential. In A_{lab} (upper left subplot), the 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 spectroscopic systems, the

predictions for each of the three replicate spectra are very different. For example, the sample with a BMP of 0.63 $NL(CH_4).gTS^{-1}$ has predictions that vary for A_{lab} between 0.869 $NL(CH_4).gTS^{-1}$ and 0.870 $NL(CH_4).gTS^{-1}$, while the predictions for A_{ip} and A_{μ} vary respectively between 0.863 $NL(CH_4).gTS^{-1}$ and 0.940 $NL(CH_4).gTS^{-1}$ and between 0.709 $NL(CH_4).gTS^{-1}$ and 0.771 $NL(CH_4).gTS^{-1}$. This is observed for all the other characteristics as shown in Appendix D. It appears that in all the compact systems, the replicate spectra vary much more from each other than with the laboratory spectrometer. This could explain the greater errors obtained using the compact systems. The following result may be investigated more quantitatively by calculating the global repeatability standard deviations as presented in Fig. 6. Indeed, for all characteristics, the compact systems show much higher repeatability standard deviations than the laboratory spectrometer (in red). Such differences are due to the way the spectral measurements are acquired. As detailed in Table 1, the systems do not have the same number of scans and sampling surface. While for one measurement, the laboratory spectrometer collects scans during a full rotation of the sampling cup, the other systems only collect scans on a fixed point of the sample's surface. This means that in

compact systems the spectral measurement is much less representative of the total sample. It appears that the performance of compact systems could be enhanced by optimizing the way the spectra are taken: increasing the number of scans and the number of replicates to ensure a better stability of the measurements.

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

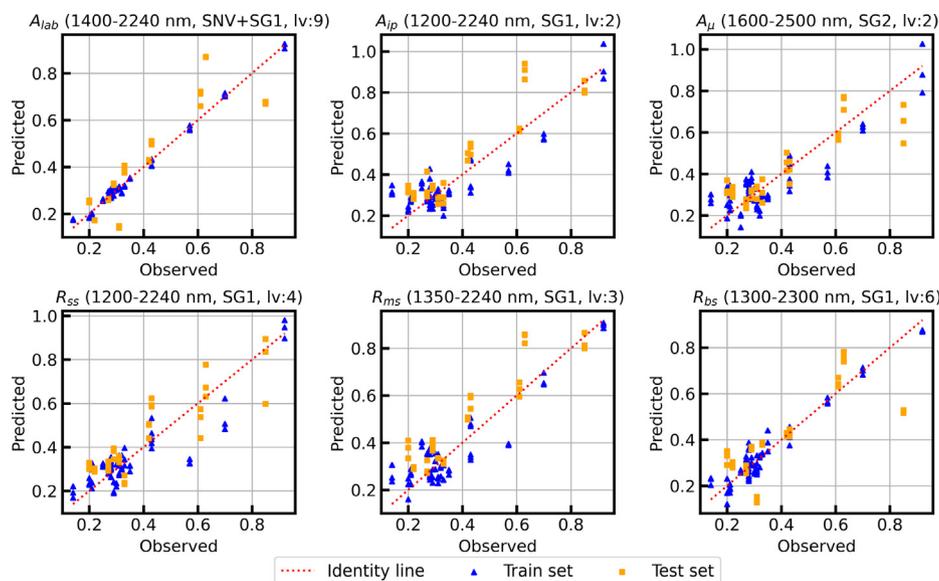


Fig. 5. Scatter plots of predicted and observed values for the selected prediction models of biochemical methane potential (BMP). Each subplot corresponds to the best model selected for a signal type (from the four spectroscopic configurations). Values for train and test sets are respectively presented in blue triangles and orange squares. The spectral range, pre-processing type and latent variable number (lv) of the models are provided in the titles. Results for other reference characteristics are provided in Appendix D. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

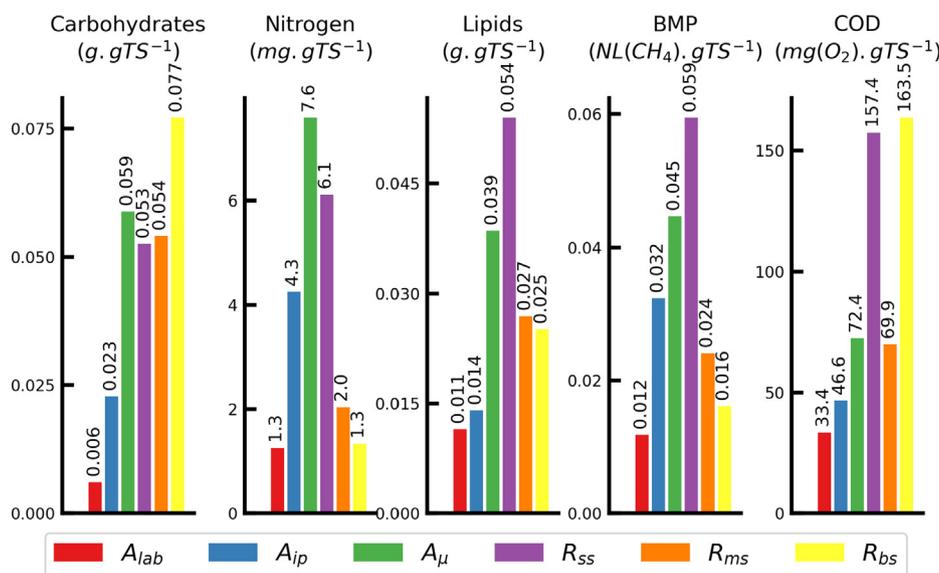


Fig. 6. Global prediction repeatability standard deviations as calculated with (Eq. (7)) and (Eq. (8)). For each reference characteristic (carbohydrates, nitrogen, lipids, BMP, COD), a bar plot colored by signal type is provided (A_{lab} in red, A_{ip} in blue, A_{μ} in green, R_{ss} in violet, R_{ms} in orange, and R_{bs} in yellow). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

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.

4. Conclusions

Results have shown that compact and low-cost systems including a hand-held micro spectrometer are suitable for online characterization of diverse solid organic waste. However, the use of new measurement configurations such as the polarized mode was not

shown to be an effective way to enhance the quality of predictive models. This suggests that the physical scattering properties of the substrates are 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 systems appear to lie in further optimization of the sampling protocol. These results set the path to a new era of low-cost and on-site NIRS analysis of the feeding substrates in co-digestion plants.

CRediT authorship contribution statement

Alexandre Mallet: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing - original draft, Supervision.

Margaud Pérémé: Methodology, Software, Formal analysis, Investigation, Writing - review & editing. **Lorraine Awhangbo:** Conceptualization, Methodology, Writing - review & editing, 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, Writing - review & editing, Supervision. **Ryad Bendoula:** Conceptualization, Methodology, Writing - review & editing, Supervision.

Declaration of Competing Interest

The authors declared that there is no conflict of interest.

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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.wasman.2021.03.045>.

References

- Alayed, M., Deen, M.J., 2017. Time-resolved diffuse optical spectroscopy and imaging using solid-state detectors: Characteristics, present status, and research challenges. *Sensors (Switzerland)* 17, 2115. <https://doi.org/10.3390/s17092115>.
- Albrecht, R., Joffre, R., Gros, R., Le Petit, J., Terron, G., Périsol, C., 2008. Efficiency of near-infrared reflectance spectroscopy to assess and predict the stage of transformation of organic matter in the composting process. *Bioresour. Technol.* 99 (2), 448–455. <https://doi.org/10.1016/j.biortech.2006.12.019>.
- Awhangbo, L., Bendoula, R., Roger, J.M., Béline, F., 2020. Multi-block SO-PLS approach based on infrared spectroscopy for anaerobic digestion process monitoring. *Chemom. Intell. Lab. Syst.* 196, 103905. <https://doi.org/10.1016/j.chemolab.2019.103905>.
- Backman, V., Gurjar, R., Badizadegan, K., Itzkan, I., Dasari, R.R., Perelman, L.T., Feld, M.S., 1999. Polarized light scattering spectroscopy for quantitative measurement of epithelial cellular structures in situ. *IEEE J. Sel. Top. Quantum Electron.* 5, 1019–1026. <https://doi.org/10.1109/2944.796325>.
- Barnes, R.J., Dhanoa, M.S., Lister, S.J., 1989. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Appl. Spectrosc.* 43 (5), 772–777. <https://doi.org/10.1366/000370289402201>.
- Beć, K.B., Grabska, J., Huck, C.W., 2021. Principles and Applications of Miniaturized Near-Infrared (NIR) Spectrometers. *Chem. - A Eur. J.* 27 (5), 1514–1532. <https://doi.org/10.1002/chem.v27.5.10.1002/chem.202002838>.
- Beć, K.B., Grabska, J., Siesler, H.W., Huck, C.W., 2020. Handheld near-infrared spectrometers: Where are we heading?. *NIR news* 31 (3–4), 28–35. <https://doi.org/10.1177/0960336020916815>.
- Bendoula, R., Gobrecht, A., Moulin, B., Roger, J.-M., Bellon-Maurel, V., 2015. Improvement of the chemical content prediction of a model powder system by reducing multiple scattering using polarized light spectroscopy. *Appl. Spectrosc.* 69 (1), 95–102. <https://doi.org/10.1366/14-07539>.
- Charnier, C., Latrille, E., Jimenez, J., Lemoine, M., Boulet, J.C., Miroux, J., Steyer, J.P., 2017. Fast characterization of solid organic waste content with near infrared spectroscopy in anaerobic digestion. *Waste Manag.* 59, 140–148. <https://doi.org/10.1016/j.wasman.2016.10.029>.
- Cordella, C.B.Y., 2012. PCA: The Basic Building Block of Chemometrics, in: Krull, I.S. (Ed.), *Analytical Chemistry*. IntechOpen, pp. 1–46. <https://doi.org/10.5772/51429>.
- Crocombe, R.A., 2018. Portable Spectroscopy. *Appl. Spectrosc.* 72 (12), 1701–1751. <https://doi.org/10.1177/0003702818809719>.
- Dahm, K.D., Dahm, D.J., 2004. Relation of representative layer theory to other theories of diffuse reflection. *J. Near Infrared Spectrosc.* 12 (3), 189–198. <https://doi.org/10.1255/jnirs.426>.
- Doublet, J., Boulanger, A., Ponthieux, A., Laroche, C., Poitrenaud, M., Cacho Rivero, J. A., 2013. Predicting the biochemical methane potential of wide range of organic substrates by near infrared spectroscopy. *Bioresour. Technol.* 128, 252–258. <https://doi.org/10.1016/j.biortech.2012.10.044>.
- Fitamo, T., Triolo, J.M., Boldrin, A., Scheut, C., 2017. Rapid biochemical methane potential prediction of urban organic waste with near-infrared reflectance spectroscopy. *Water Res.* 119, 242–251. <https://doi.org/10.1016/j.watres.2017.04.051>.
- Gobrecht, A., Bendoula, R., Roger, J.M., Bellon-Maurel, V., 2016. A new optical method coupling light polarization and Vis-NIR spectroscopy to improve the measurement of soil carbon content. *Soil Tillage Res.* 155, 461–470. <https://doi.org/10.1016/j.still.2015.06.003>.
- Gobrecht, A., Bendoula, R., Roger, J.M., Bellon-Maurel, V., 2015. Combining linear polarization spectroscopy and the Representative Layer Theory to measure the Beer-Lambert law absorbance of highly scattering materials. *Anal. Chim. Acta* 853, 486–494. <https://doi.org/10.1016/j.aca.2014.10.014>.
- Gobrecht, A., Roger, J.M., Bellon-Maurel, V., 2014. Major Issues of Diffuse Reflectance NIR Spectroscopy in the Specific Context of Soil Carbon Content Estimation. A Review., *Advances in Agronomy*. <https://doi.org/10.1016/B978-0-12-420225-2.00004-2>.
- Godin, B., Mayer, F., Agneessens, R., Gerin, P., Dardenne, P., Delfosse, P., Delcarte, J., 2015. Biochemical methane potential prediction of plant biomasses: Comparing chemical composition versus near infrared methods and linear versus non-linear models. *Bioresour. Technol.* 175, 382–390. <https://doi.org/10.1016/j.biortech.2014.10.115>.
- Hagos, K., Zong, J., Li, D., Liu, C., Lu, X., 2017. Anaerobic co-digestion process for biogas production: Progress, challenges and perspectives. *Renew. Sustain. Energy Rev.* 76, 1485–1496. <https://doi.org/10.1016/j.rser.2016.11.184>.
- Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M.H., Brett, M., Haldane, A., del Río, J.F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., Oliphant, T.E., 2020. Array programming with NumPy. *Nature* 585 (7825), 357–362. <https://doi.org/10.1038/s41586-020-2649-2>.
- Hemrattrakun, P., Nakano, K., Boonyakiat, D., Ohashi, S., Maniwa, P., Theanjumpol, P., Seehanam, P., 2021. Comparison of Reflectance and Interference Modes of Visible and Near-Infrared Spectroscopy for Predicting Persimmon Fruit Quality. *Food Anal. Methods* 14 (1), 117–126. <https://doi.org/10.1007/s12161-020-01853-w>.
- Hunter, J.D., 2007. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* 9 (3), 90–95. <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. Technol.* 102 (7), 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 monitoring of volatile fatty acids in anaerobic digestion. *Water Sci. Technol.* 60, 339–346. <https://doi.org/10.2166/wst.2009.345>.
- Jacobi, H.F., Ohl, S., Thiessen, E., Hartung, E., 2012. NIRS-aided monitoring and prediction of biogas yields from maize silage at a full-scale biogas plant applying lumped kinetics. *Bioresour. Technol.* 103 (1), 162–172. <https://doi.org/10.1016/j.biortech.2011.10.012>.
- Johansson, J., Folestad, S., Josefson, M., Sparén, A., Abrahamsson, C., Andersson-Engels, S., Svanberg, S., 2002. Time-resolved NIR/vis spectroscopy for analysis of solids: Pharmaceutical tablets. *Appl. Spectrosc.* 56 (6), 725–731. <https://doi.org/10.1366/000370202760077676>.
- Khodabakhshian, R., Emadi, B., Khojastehpour, M., Golzarian, M.R., 2019. A comparative study of reflectance and transmittance modes of Vis/NIR spectroscopy used in determining internal quality attributes in pomegranate fruits. *J. Food Meas. Charact.* 13 (4), 3130–3139. <https://doi.org/10.1007/s11694-019-00235-z>.
- Krapf, L.C., Nast, D., Gronauer, A., Schmidhalter, U., Heuwinkel, H., 2013. Transfer of a near infrared spectroscopy laboratory application to an online process analyser for in situ monitoring of anaerobic digestion. *Bioresour. Technol.* 129, 39–50. <https://doi.org/10.1016/j.biortech.2012.11.027>.
- Lesteur, M., Latrille, E., Maurel, V.B., Roger, J.M., Gonzalez, C., Junqua, G., Steyer, J.P., 2011. First step towards a fast analytical method for the determination of Biochemical Methane Potential of solid wastes by near infrared spectroscopy. *Bioresour. Technol.* 102 (3), 2280–2288. <https://doi.org/10.1016/j.biortech.2010.10.044>.
- Lomborg, C.J., Holm-Nielsen, J.B., Oleskowicz-Popiel, P., Esbensen, K.H., 2009. Near infrared and acoustic chemometrics monitoring of volatile fatty acids and dry matter during co-digestion of manure and maize silage. *Bioresour. Technol.* 100 (5), 1711–1719. <https://doi.org/10.1016/j.biortech.2008.09.043>.
- Mallet, A., Charnier, C., Latrille, É., Bendoula, R., Steyer, J.-P., Roger, J.-M., 2021. Unveiling non-linear water effects in near infrared spectroscopy: A study on organic wastes during drying using chemometrics. *Waste Manag.* 122, 36–48. <https://doi.org/10.1016/j.wasman.2020.12.019>.
- McKinney, W., 2010. Data Structures for Statistical Computing in Python, in: *Proceedings of the 9th Python in Science Conference*, pp. 56–61.
- McVey, C., McGrath, T.F., Haughey, S.A., Elliott, C.T., 2021. A rapid food chain approach for authenticity screening: The development, validation and transferability of a chemometric model using two handheld near infrared spectroscopy (NIRS) devices. *Talanta* 222, 121533. <https://doi.org/10.1016/j.talanta.2020.121533>.
- Næs, T., Martens, H., 1984. Multivariate calibration. II. Chemometric methods. *Trends Anal. Chem.* 3 (10), 266–271. [https://doi.org/10.1016/0165-9936\(84\)80044-8](https://doi.org/10.1016/0165-9936(84)80044-8).
- Varoquaux, G., Buitinck, L., Louppe, G., Grisel, O., Pedregosa, F., Mueller, A., 2015. *Scikit-learn: Machine Learning in Python*. *J. Mach. Learn. Res.* 19 (1), 29–33.
- Rabatel, G., Marini, F., Walczak, B., Roger, J.M., 2020. VSN: Variable sorting for normalization. *J. Chemom.* 34, 1–16. <https://doi.org/10.1002/cem.3164>.

- Rinnan, Å., Berg, F.V.D., Engelsen, S.B., 2009. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC - Trends Anal. Chem.* 28 (10), 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. Comparative analysis of methods and models for predicting biochemical methane potential of various organic substrates. *Sci. Total Environ.* 649, 1599–1608. <https://doi.org/10.1016/j.scitotenv.2018.08.270>.
- Rutledge, D.N., Barros, A.S., 2002. Durbin-Watson statistic as a morphological estimator of information content. *Anal. Chim. Acta* 454 (2), 277–295. [https://doi.org/10.1016/S0003-2670\(01\)01555-0](https://doi.org/10.1016/S0003-2670(01)01555-0).
- Savitzky, A., Golay, M.J.E., 1964. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* 36 (8), 1627–1639. <https://doi.org/10.1021/ac60214a047>.
- Schaare, P.N., Fraser, D.G., 2000. Comparison of reflectance, interactance and transmission modes of visible-near infrared spectroscopy for measuring internal properties of kiwifruit (*Actinidia chinensis*). *Postharvest Biol. Technol.* 20 (2), 175–184. [https://doi.org/10.1016/S0925-5214\(00\)00130-7](https://doi.org/10.1016/S0925-5214(00)00130-7).
- Sharififar, A., Singh, K., Jones, E., Ginting, F.I., Minasny, B., 2019. Evaluating a low-cost portable NIR spectrometer for the prediction of soil organic and total carbon using different calibration models. *Soil Use Manag.* 35, 607–616. <https://doi.org/10.1111/sum.12537>.
- Skvaril, J., Kyprianidis, K.G., Dahlquist, E., 2017. Applications of near-infrared spectroscopy (NIRS) in biomass energy conversion processes: A review. *Appl. Spectrosc. Rev.* 52 (8), 675–728. <https://doi.org/10.1080/05704928.2017.1289471>.
- Snee, R.D., 1977. Validation of Regression Models: Methods and Examples. *Technometrics* 19 (4), 415–428. <https://doi.org/10.1080/00401706.1977.10489581>.
- Stockl, A., Lichti, F., 2018. Near-infrared spectroscopy (NIRS) for a real time monitoring of the biogas process. *Bioresour. Technol.* 247, 1249–1252. <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. <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 Infrared Reflectance Spectroscopy (NIRS) for rapid determination of biochemical methane potential of plant biomass. *Appl. Energy* 116, 52–57. <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>
- Vergnoux, A., Guiliano, M., Le Dréau, Y., Kister, J., Dupuy, N., Doumenq, P., 2009. Monitoring of the evolution of an industrial compost and prediction of some compost properties by NIR spectroscopy. *Sci. Total Environ.* 407 (7), 2390–2403. <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., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A., H., Pedregosa, F., van Mulbregt, P., Vijaykumar, A., Bardelli, A.P., Rothberg, A., Hilboll, A., Kloeckner, A., Scopatz, A., Lee, A., Rokem, A., Woods, C.N., Fulton, C., Masson, C., Häggström, C., Fitzgerald, C., Nicholson, D.A., Hagen, D.R., 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., Schönberger, J.L., de Miranda Cardoso, J.V., Reimer, J., Harrington, J., Rodríguez, J.L.C., Nunez-Iglesias, J., Kuczynski, J., Tritz, K., Thoma, M., Newville, M., Kümmerer, M., Bolingbroke, M., Tarte, M., Pak, M., Smith, N.J., Nowaczyk, N., Shebanov, N., Pavlyk, O., Brodtkorb, P.A., Lee, P., McGibbon, R.T., Feldbauer, R., Lewis, S., Tygier, S., Sievert, S., Vigna, S., Peterson, S., More, S., Pudlik, T., 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 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* 17, 261–272. <https://doi.org/10.1038/s41592-019-0686-2>.
- Williams, P., Manley, M., Antoniszyn, J. (Eds.), 2019. Near-infrared Technology: Getting the best out of light. *AFRICAN SUN MeDIA*.
- 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.* 30, 79–82. <https://doi.org/10.3354/cr030079>.
- WOLD, H., 1973. Nonlinear Iterative Partial Least Squares (NIPALS) Modelling: Some Current Developments, Multivariate Analysis-III. Academic press, Inc. <https://doi.org/10.1016/b978-0-12-426653-7.50032-6>
- Wolf, Dieter, von Canstein, Harald, Schröder, Cathrin, 2011. Optimisation of biogas production by infrared spectroscopy-based process control. *J. Nat. Gas Sci. Eng.* 3 (5), 625–632. <https://doi.org/10.1016/j.jngse.2011.07.006>.
- Wolfrum, Edward J., Payne, Courtney, Schwartz, Alexa, Jacobs, Joshua, Kressin, Robert W., 2020. A Performance Comparison of Low-Cost Near-Infrared (NIR) Spectrometers to a Conventional Laboratory Spectrometer for Rapid Biomass Compositional Analysis. *Bioenergy Res.* 13 (4), 1121–1129. <https://doi.org/10.1007/s12155-020-10135-6>.
- Workman Jr., J., Weyer, L., 2012. Practical Guide and Spectral Atlas for Interpretive Near-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. <https://doi.org/10.1016/j.rser.2018.12.039>.
- Xu, Jun-Li, Gobrecht, Alexia, Gorretta, Nathalie, Héran, Daphné, Gowen, Aoife A., Bendoula, Ryad, 2019. Development of a polarized hyperspectral imaging system for investigation of absorption and scattering properties. *J. Near Infrared Spectrosc.* 27 (4), 314–329. <https://doi.org/10.1177/0967033519857732>.
- Yang, Gaixiu, Li, Ying, Zhen, Feng, Xu, Yonghua, Liu, Jinming, Li, Nan, Sun, Yong, Luo, Lina, Wang, Ming, Zhang, Lingling, 2021. Biochemical methane potential prediction for mixed feedstocks of straw and manure in anaerobic co-digestion. *Bioresour. Technol.* 326, 124745. <https://doi.org/10.1016/j.biortech.2021.124745>.
- Zeaiter, M., Roger, J.-M., Bellon-Maurel, V., 2005. Robustness of models developed by multivariate calibration. Part II: The influence of pre-processing methods. *TrAC - Trends Anal. Chem.* 24 (5), 437–445. <https://doi.org/10.1016/j.trac.2004.11.023>.

Paper V – 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

Full reference:

Péréomé, M.; Mallet, A.; Awhangbo, L.; Charnier, C.; Roger, J.; Steyer, J.; Latrille, É.; Bendoula, R. 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 Br.* **2021**, *36*, 107126. <https://doi.org/10.1016/j.dib.2021.107126>.



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^{a,d,e}, Alexandre Mallet^{a,b,c,e,*}, Lorraine Awhangbo^{a,e},
Cyrille Charnier^c, Jean-michel Roger^{b,e}, Jean-philippe Steyer^a,
Éric Latrille^{a,e}, Ryad Bendoula^b

^a INRAE, Univ Montpellier, LBE¹, 102 Av des Etangs, Narbonne F-11100, France

^b INRAE, UMR ITAP, Montpellier University, Montpellier F-34000, France

^c BIOENTECH Company, Narbonne F-11100, France

^d ENSCM, 240 Av du professeur Emile Jeanbrau, Montpellier F-34090, France

^e 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 (A. Mallet).

¹ @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"> - NIRFlex N-500 FT-NIR (BUCHI, France) - Immersed probe consisting of two fibers (one for illumination, the other for signal collection) plugged to a Rocket FTNIR-L1-025-2TE (Arcoptics, Switzerland) - Polarized light system (PoLiS) plugged to a Rocket FTNIR-L1-025-2TE (Arcoptics, Switzerland) - Neospectra-micro + Raspberry-Pi (Si-Ware, Egypt)
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: https://doi.org/10.15454/SQQTUU
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. https://doi.org/10.1016/j.wasman.2021.03.045 .

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*".

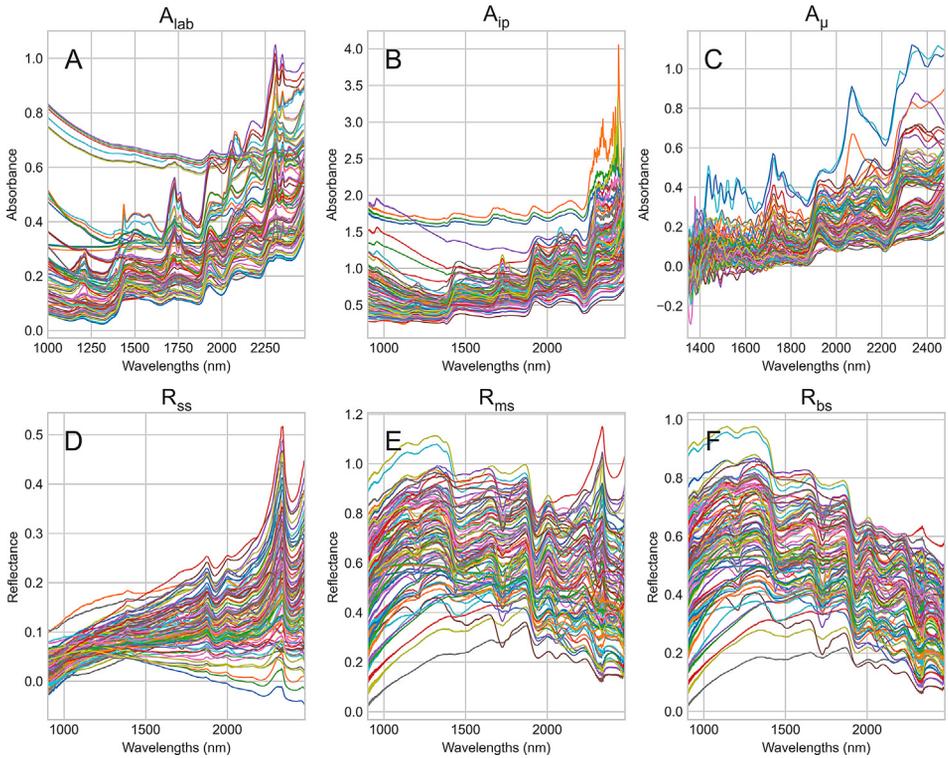


Fig. 1. Raw spectra obtained on the 33 organic waste samples with A) the laboratory spectrometer (absorbance, in nm, referred as A_{lab}), B) the compact spectrometer with an immersed probe (absorbance, in nm, referred as A_{ip}), C) the micro spectrometer (absorbance, in nm, referred as A_{μ}), and 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 cm^{-1} , using a BUCHI NIR-Flex N-500 solids spectrophotometer fitted with a vial accessory (Buchi, Flawil,

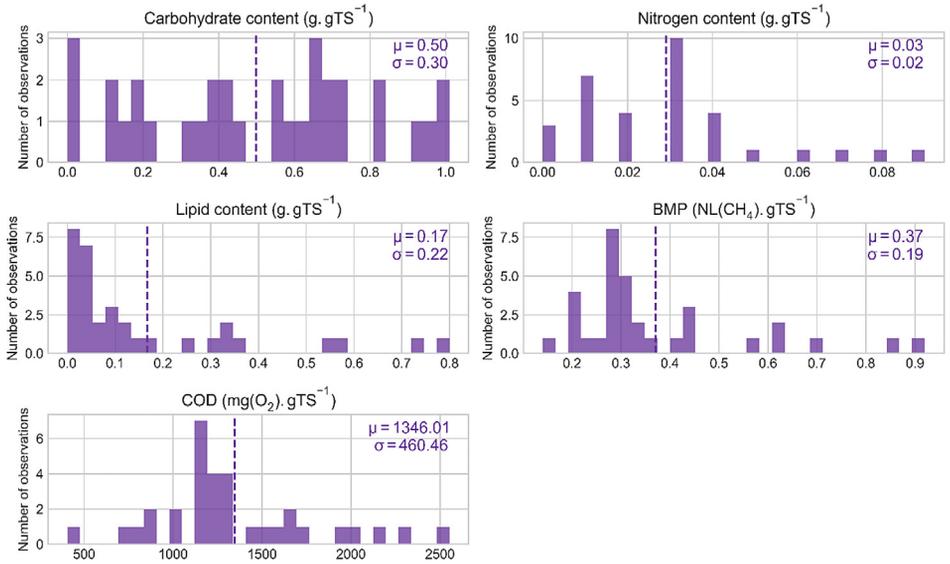


Fig. 2. Histograms of reference values obtained on the 33 organic waste samples using NIRS prediction model. For each subplot, both the mean value (μ) and standard deviation value (σ) are provided; and the dashed line represents the mean.

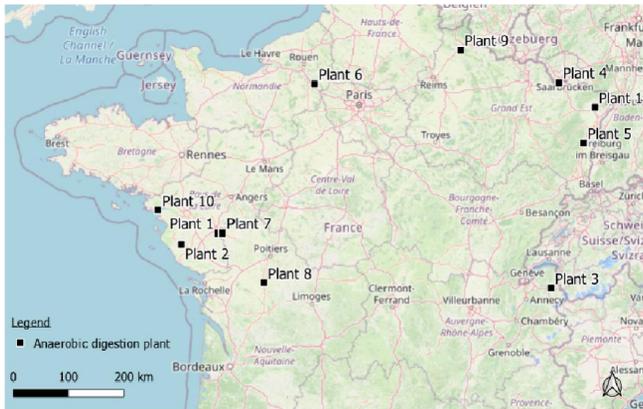


Fig. 3. Map (focused on France) showing the positions of the anaerobic digestion plants from where the substrates were collected.

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 cm^{-1} (900–2500 nm) with a resolution of 4 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 cm^{-1} to 7407 cm^{-1} (1350–2550 nm) and a resolution of 66 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).

All references

- (1) The Core Writing Team; Pachauri, R.; Meyer, L. *IPCC, 2014: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; 2014.
- (2) Ameli, H.; Qadrdan, M.; Strbac, G. Techno-Economic Assessment of Battery Storage and Power-to-Gas: A Whole-System Approach. *Energy Procedia* **2017**, *142*, 841–848. <https://doi.org/10.1016/j.egypro.2017.12.135>.
- (3) Lecker, B.; Illi, L.; Lemmer, A.; Oechsner, H. Biological Hydrogen Methanation – A Review. *Bioresour. Technol.* **2017**, *245* (August), 1220–1228. <https://doi.org/10.1016/j.biortech.2017.08.176>.
- (4) Brémond, U.; Bertrandias, A.; Steyer, J. P.; Bernet, N.; Carrere, H. A Vision of European Biogas Sector Development towards 2030: Trends and Challenges. *J. Clean. Prod.* **2021**, *287*. <https://doi.org/10.1016/j.jclepro.2020.125065>.
- (5) Moletta, R. *La Méthanisation*, 3rd editio.; Lavoisier, T. et D.-, Ed.; 2015.
- (6) Adekunle, K. F.; Okolie, J. A. A Review of Biochemical Process of Anaerobic Digestion. *Adv. Biosci. Biotechnol.* **2015**, *06* (03), 205–212. <https://doi.org/10.4236/abb.2015.63020>.
- (7) Baena-Moreno, F. M.; le Saché, E.; Pastor-Pérez, L.; Reina, T. R. Membrane-Based Technologies for Biogas Upgrading: A Review. *Environ. Chem. Lett.* **2020**, *18* (5), 1649–1658. <https://doi.org/10.1007/s10311-020-01036-3>.
- (8) Shinde, A. M.; Dikshit, A. K.; Odlare, M.; Thorin, E.; Schwede, S. Life Cycle Assessment of Bio-Methane and Biogas-Based Electricity Production from Organic Waste for Utilization as a Vehicle Fuel. *Clean Technol. Environ. Policy* **2021**, No. 0123456789. <https://doi.org/10.1007/s10098-021-02054-7>.
- (9) Rostkowski, K. H.; Criddle, C. S.; Lepech, M. D. Cradle-to-Gate Life Cycle Assessment for a Cradle-to-Cradle Cycle: Biogas-to-Bioplastic (and Back). *Environ. Sci. Technol.* **2012**, *46* (18), 9822–9829. <https://doi.org/10.1021/es204541w>.
- (10) Guilayn, F.; Rouez, M.; Crest, M.; Patureau, D.; Jimenez, J. *Valorization of Digestates from Urban or Centralized Biogas Plants: A Critical Review*; Springer Netherlands, 2020; Vol. 19. <https://doi.org/10.1007/s11157-020-09531-3>.
- (11) Narihiro, T.; Sekiguchi, Y. Microbial Communities in Anaerobic Digestion Processes for Waste and Wastewater Treatment: A Microbiological Update. *Curr. Opin. Biotechnol.* **2007**, *18* (3), 273–278. <https://doi.org/10.1016/j.copbio.2007.04.003>.
- (12) de Jonge, N.; Davidsson, Å.; la Cour Jansen, J.; Nielsen, J. L. Characterisation of Microbial Communities for Improved Management of Anaerobic Digestion of Food Waste. *Waste Manag.* **2020**, *117*, 124–135. <https://doi.org/10.1016/j.wasman.2020.07.047>.
- (13) Mir, M. A.; Hussain, A.; Verma, C. Design Considerations and Operational Performance of Anaerobic Digester: A Review. *Cogent Eng.* **2016**, *3* (1). <https://doi.org/10.1080/23311916.2016.1181696>.
- (14) Zamri, M. F. M. A.; Hasmady, S.; Akhilar, A.; Ideris, F.; Shamsuddin, A. H.; Mofijur, M.; Fattah, I. M. R.; Mahlia, T. M. I. A Comprehensive Review on Anaerobic Digestion of Organic Fraction of Municipal Solid Waste. *Renew. Sustain. Energy Rev.* **2021**, *137* (May 2020), 110637. <https://doi.org/10.1016/j.rser.2020.110637>.
- (15) Kumar, A.; Samadder, S. R. Performance Evaluation of Anaerobic Digestion Technology for Energy Recovery from Organic Fraction of Municipal Solid Waste: A Review. *Energy* **2020**, *197*, 117253. <https://doi.org/10.1016/j.energy.2020.117253>.
- (16) Diamantis, V.; Eftaxias, A.; Stamatelatos, K.; Noutsopoulos, C.; Vlachokostas, C.; Aivasidis, A. Bioenergy in the Era of Circular Economy: Anaerobic Digestion Technological Solutions to Produce Biogas from Lipid-Rich Wastes. *Renew. Energy* **2021**, *168*, 438–447. <https://doi.org/10.1016/j.renene.2020.12.034>.
- (17) European Biogas Association. *EBA Statistical Report 2020*; 2020.
- (18) Wouters, C.; Buseman, M.; van Tilburg, J.; Berg, T.; Cihlar, J.; Villar Lejarreta, A.; Jens, J.; Wang, A.; Peters, D.; van der Leun, K. *Market State and Trends in Renewable and Low-Carbon Gases in Europe*; 2020.

All references

- (19) Yang, X.; Liu, Y.; Thrän, D.; Bezama, A.; Wang, M. Effects of the German Renewable Energy Sources Act and Environmental, Social and Economic Factors on Biogas Plant Adoption and Agricultural Land Use Change. *Energy. Sustain. Soc.* **2021**, *11* (1), 1–22. <https://doi.org/10.1186/s13705-021-00282-9>.
- (20) Theuerl, S.; Herrmann, C.; Heiermann, M.; Grundmann, P.; Landwehr, N.; Kreidenweis, U.; Prochnow, A. The Future Agricultural Biogas Plant in Germany: A Vision. *Energies*. 2019, pp 1–32. <https://doi.org/10.3390/en12030396>.
- (21) Capson-Tojo, G.; Trably, E.; Rouez, M.; Crest, M.; Steyer, J. P.; Delgenès, J. P.; Escudíé, R. Dry Anaerobic Digestion of Food Waste and Cardboard at Different Substrate Loads, Solid Contents and Co-Digestion Proportions. *Bioresour. Technol.* **2017**, *233*, 166–175. <https://doi.org/10.1016/j.biortech.2017.02.126>.
- (22) Hagos, K.; Zong, J.; Li, D.; Liu, C.; Lu, X. Anaerobic Co-Digestion Process for Biogas Production: Progress, Challenges and Perspectives. *Renew. Sustain. Energy Rev.* **2017**, *76* (March 2016), 1485–1496. <https://doi.org/10.1016/j.rser.2016.11.184>.
- (23) Pellerá, F. M.; Gidarakos, E. Anaerobic Digestion of Solid Agroindustrial Waste in Semi-Continuous Mode: Evaluation of Mono-Digestion and Co-Digestion Systems. *Waste Manag.* **2017**, *68*, 103–119. <https://doi.org/10.1016/j.wasman.2017.06.026>.
- (24) Bong, C. P. C.; Lim, L. Y.; Lee, C. T.; Klemeš, J. J.; Ho, C. S.; Ho, W. S. The Characterisation and Treatment of Food Waste for Improvement of Biogas Production during Anaerobic Digestion – A Review. *J. Clean. Prod.* **2018**, *172*, 1545–1558. <https://doi.org/10.1016/j.jclepro.2017.10.199>.
- (25) Chow, W. L.; Chong, S.; Lim, J. W.; Chan, Y. J.; Chong, M. F. Anaerobic Co-Digestion of Wastewater Sludge : A Review of Potential Co-Substrates and Operating. *Processes* **2020**, *8*(1) (39), 1–21.
- (26) Rajagopal, R.; Massé, D. I.; Singh, G. A Critical Review on Inhibition of Anaerobic Digestion Process by Excess Ammonia. *Bioresour. Technol.* **2013**, *143*, 632–641. <https://doi.org/10.1016/j.biortech.2013.06.030>.
- (27) Zhang, C.; Xiao, G.; Peng, L.; Su, H.; Tan, T. The Anaerobic Co-Digestion of Food Waste and Cattle Manure. *Bioresour. Technol.* **2013**, *129*, 170–176. <https://doi.org/10.1016/j.biortech.2012.10.138>.
- (28) Dima, A. D.; Pârvulescu, O. C.; Mateescu, C.; Dobre, T. Optimization of Substrate Composition in Anaerobic Co-Digestion of Agricultural Waste Using Central Composite Design. *Biomass and Bioenergy* **2020**, *138* (May), 8–10. <https://doi.org/10.1016/j.biombioe.2020.105602>.
- (29) Wei, L.; Qin, K.; Ding, J.; Xue, M.; Yang, C.; Jiang, J.; Zhao, Q. Optimization of the Co-Digestion of Sewage Sludge, Maize Straw and Cow Manure: Microbial Responses and Effect of Fractional Organic Characteristics. *Sci. Rep.* **2019**, *9* (1), 1–10. <https://doi.org/10.1038/s41598-019-38829-8>.
- (30) Vijin Prabhu, A.; Antony Raja, S.; Avinash, A.; Pugazhendhi, A. Parametric Optimization of Biogas Potential in Anaerobic Co-Digestion of Biomass Wastes. *Fuel* **2021**, *288* (October 2020), 119574. <https://doi.org/10.1016/j.fuel.2020.119574>.
- (31) Almomani, F.; Bhosale, R. R. Enhancing the Production of Biogas through Anaerobic Co-Digestion of Agricultural Waste and Chemical Pre-Treatments. *Chemosphere* **2020**, *255*, 126805. <https://doi.org/10.1016/j.chemosphere.2020.126805>.
- (32) Du, M.; Liu, X.; Wang, D.; Yang, Q.; Duan, A.; Chen, H.; Liu, Y.; Wang, Q.; Ni, B. J. Understanding the Fate and Impact of Capsaicin in Anaerobic Co-Digestion of Food Waste and Waste Activated Sludge. *Water Res.* **2021**, *188*, 116539. <https://doi.org/10.1016/j.watres.2020.116539>.
- (33) Mata-Alvarez, J.; Dosta, J.; Romero-Güiza, M. S.; Fonoll, X.; Peces, M.; Astals, S. A Critical Review on Anaerobic Co-Digestion Achievements between 2010 and 2013. *Renewable and Sustainable Energy Reviews*. Elsevier 2014, pp 412–427. <https://doi.org/10.1016/j.rser.2014.04.039>.
- (34) Wang, Z.; Jiang, Y.; Wang, S.; Zhang, Y.; Hu, Y.; Hu, Z. hu; Wu, G.; Zhan, X. Impact of Total Solids Content on Anaerobic Co-Digestion of Pig Manure and Food Waste: Insights into Shifting of the Methanogenic Pathway. *Waste Manag.* **2020**, *114*, 96–106. <https://doi.org/10.1016/j.wasman.2020.06.048>.
- (35) Karki, R.; Chuenchart, W.; Surendra, K. C.; Shrestha, S.; Raskin, L.; Sung, S.; Hashimoto, A.; Kumar Khanal, S. Anaerobic Co-Digestion: Current Status and Perspectives. *Bioresource Technology*. Elsevier Ltd 2021, p 125001. <https://doi.org/10.1016/j.biortech.2021.125001>.
- (36) Ward, A. J.; Hobbs, P. J.; Holliman, P. J.; Jones, D. L. Optimisation of the Anaerobic Digestion of Agricultural Resources. *Bioresour. Technol.* **2008**, *99* (17), 7928–7940. <https://doi.org/10.1016/j.biortech.2008.02.044>.

All references

- (37) Batstone, D. J.; Keller, J.; Angelidaki, I.; Kalyuzhnyi, S. V.; Pavlostathis, S. G.; Rozzi, A.; Sanders, W. T.; Siegrist, H.; Vavilin, V. a. *Adm1. Iwa Publ.* **2002**, No. 1, 1–68.
- (38) De Clercq, D.; Wen, Z.; Fei, F.; Caicedo, L.; Yuan, K.; Shang, R. Interpretable Machine Learning for Predicting Biomethane Production in Industrial-Scale Anaerobic Co-Digestion. *Sci. Total Environ.* **2020**, *712*, 134574. <https://doi.org/10.1016/j.scitotenv.2019.134574>.
- (39) Wade, M. J. Not Just Numbers: Mathematical Modelling and Its Contribution to Anaerobic Digestion Processes. *Processes* **2020**, *8* (8). <https://doi.org/10.3390/PR8080888>.
- (40) Keucken, A.; Habagil, M.; Batstone, D.; Jeppsson, U.; Arnell, M. Anaerobic Co-Digestion of Sludge and Organic Food Waste-Performance, Inhibition, and Impact on the Microbial Community. *Energies* **2018**, *11* (9). <https://doi.org/10.3390/en11092325>.
- (41) Charnier, C.; Latrille, E.; Jimenez, J.; Torrijos, M.; Sousbie, P.; Miroux, J.; Steyer, J. P. Fast ADM1 Implementation for the Optimization of Feeding Strategy Using near Infrared Spectroscopy. *Water Res.* **2017**, *122*, 27–35. <https://doi.org/10.1016/j.watres.2017.05.051>.
- (42) Jimenez, J.; Aemig, Q.; Doussiet, N.; Steyer, J. P.; Houot, S.; Patureau, D. A New Organic Matter Fractionation Methodology for Organic Wastes: Bioaccessibility and Complexity Characterization for Treatment Optimization. *Bioresour. Technol.* **2015**, *194*, 344–353. <https://doi.org/10.1016/j.biortech.2015.07.037>.
- (43) Gaida, D.; Wolf, C.; Bongards, M. Feed Control of Anaerobic Digestion Processes for Renewable Energy Production: A Review. *Renew. Sustain. Energy Rev.* **2017**, *68*, 869–875. <https://doi.org/10.1016/j.rser.2016.06.096>.
- (44) Jimenez, J.; Latrille, E.; Harmand, J.; Robles, A.; Ferrer, J.; Gaida, D.; Wolf, C.; Mairet, F.; Bernard, O.; Alcaraz-Gonzalez, V.; Mendez-Acosta, H.; Zitomer, D.; Totzke, D.; Spanjers, H.; Jacobi, F.; Guwy, A.; Dinsdale, R.; Premier, G.; Mazhegrane, S.; Ruiz-Filippi, G.; Seco, A.; Ribeiro, T.; Pauss, A.; Steyer, J. P. Instrumentation and Control of Anaerobic Digestion Processes: A Review and Some Research Challenges. *Rev. Environ. Sci. Biotechnol.* **2015**, *14* (4), 615–648. <https://doi.org/10.1007/s11157-015-9382-6>.
- (45) Williams, P.; Antoniszyn, J. *Near-Infrared Technology: Getting the Best out of Light*; African Sun Media, 2019. <https://doi.org/10.18820/9781928480310>.
- (46) Pasquini, C. Near Infrared Spectroscopy: A Mature Analytical Technique with New Perspectives – A Review. *Anal. Chim. Acta* **2018**, *1026*, 8–36. <https://doi.org/10.1016/j.aca.2018.04.004>.
- (47) Guilhem, S. *Spectroscopies Vibratoires. Théorie, Aspects Pratiques et Applications*, Editions d.; 2020. <https://doi.org/10.17184/eac.9782813002556>.
- (48) Schaare, P. N.; Fraser, D. G. Comparison of Reflectance, Interactance and Transmission Modes of Visible-near Infrared Spectroscopy for Measuring Internal Properties of Kiwifruit (*Actinidia Chinensis*). *Postharvest Biol. Technol.* **2000**, *20* (2), 175–184. [https://doi.org/10.1016/S0925-5214\(00\)00130-7](https://doi.org/10.1016/S0925-5214(00)00130-7).
- (49) Pasquini, C. Near Infrared Spectroscopy: Fundamentals, Practical Aspects and Analytical Applications. *J. Braz. Chem. Soc.* **2003**, *14* (2), 198–219. <https://doi.org/10.1590/S0103-50532003000200006>.
- (50) Mayerhöfer, T. G.; Pahlow, S.; Popp, J. The Bouguer-Beer-Lambert Law: Shining Light on the Obscure. *ChemPhysChem* **2020**. <https://doi.org/10.1002/cphc.202000464>.
- (51) Wold, S. Chemometrics; What Do We Mean with It, and What Do We Want from It? *Chemom. Intell. Lab. Syst.* **1995**, *30* (1), 109–115. [https://doi.org/10.1016/0169-7439\(95\)00042-9](https://doi.org/10.1016/0169-7439(95)00042-9).
- (52) Lavine, B.; Workman, J. Chemometrics. *Anal. Chem.* **2008**, *80* (12), 4519–4531. <https://doi.org/10.1021/ac800728t>.
- (53) Kalivas, J. H.; Brown, S. D. *Calibration Methodologies*, Second Edi.; Elsevier, 2020; Vol. 3. <https://doi.org/10.1016/b978-0-12-409547-2.14666-9>.
- (54) Cordella, C. B. Y. PCA: The Basic Building Block of Chemometrics. In *Analytical Chemistry*; Krull, I. S., Ed.; IntechOpen, 2012; pp 1–46. <https://doi.org/10.5772/51429>.
- (55) Geladi, P.; Kowalski, B. R. Partial Least-Squares Regression: A Tutorial. *Anal. Chim. Acta* **1986**, *185* (185), 1–17. [https://doi.org/10.1016/0003-2670\(86\)80028-9](https://doi.org/10.1016/0003-2670(86)80028-9).
- (56) Malik, A.; De Juan, A.; Tauler, R. Multivariate Curve Resolution: A Different Way to Examine Chemical Data. *ACS Symp. Ser.* **2015**, *1199*, 95–128. <https://doi.org/10.1021/bk-2015-1199.ch005>.
- (57) Wang, G.; Ding, Q.; Hou, Z. Independent Component Analysis and Its Applications in Signal Processing for Analytical Chemistry. *TrAC - Trends Anal. Chem.* **2008**, *27* (4), 368–376. <https://doi.org/10.1016/j.trac.2008.01.009>.

All references

- (58) Preda, C.; Saporta, G. Clusterwise PLS Regression on a Stochastic Process. *Comput. Stat. Data Anal.* **2005**, *49* (1), 99–108. <https://doi.org/10.1016/j.csda.2004.05.002>.
- (59) Tøndel, K.; Indahl, U. G.; Gjuvslund, A. B.; Vik, J. O.; Hunter, P.; Omholt, S. W.; Martens, H. Hierarchical Cluster-Based Partial Least Squares Regression (HC-PLSR) Is an Efficient Tool for Metamodelling of Nonlinear Dynamic Models. *BMC Syst. Biol.* **2011**, *5*, 1–17. <https://doi.org/10.1186/1752-0509-5-90>.
- (60) Narayanan, H.; Sokolov, M.; Butté, A.; Morbidelli, M. Decision Tree-PLS (DT-PLS) Algorithm for the Development of Process: Specific Local Prediction Models. *Biotechnol. Prog.* **2019**, *35* (4), 1–11. <https://doi.org/10.1002/btpr.2818>.
- (61) Lesnoff, M.; Metz, M.; Roger, J. M. Comparison of Locally Weighted PLS Strategies for Regression and Discrimination on Agronomic NIR Data. *J. Chemom.* **2020**, No. November 2019, 10–12. <https://doi.org/10.1002/cem.3209>.
- (62) Lindgren, F.; Geladi, P.; Wold, S. The Kernel Algorithm for PLS. *J. Chemom.* **1993**, *7* (1), 45–59. <https://doi.org/10.1002/cem.1180070104>.
- (63) Shen, G.; Lesnoff, M.; Baeten, V.; Dardenne, P.; Davrieux, F.; Ceballos, H.; Belalcazar, J.; Dufour, D.; Yang, Z.; Han, L.; Fernández Pierna, J. A. Local Partial Least Squares Based on Global PLS Scores. *J. Chemom.* **2019**, *33* (5), 1–3. <https://doi.org/10.1002/cem.3117>.
- (64) Metz, M.; Lesnoff, M.; Abdelghafour, F.; Akbarinia, R.; Masegla, F.; Roger, J. M. A “Big-Data” Algorithm for KNN-PLS. *Chemom. Intell. Lab. Syst.* **2020**, *203* (March), 104076. <https://doi.org/10.1016/j.chemolab.2020.104076>.
- (65) Ramirez-Lopez, L.; Behrens, T.; Schmidt, K.; Rossel, R. A. V.; Demattê, J. A. M.; Scholten, T. Distance and Similarity-Search Metrics for Use with Soil Vis-NIR Spectra. *Geoderma* **2013**, *199*, 43–53. <https://doi.org/10.1016/j.geoderma.2012.08.035>.
- (66) Beyer, K.; Goldstein, J.; Ramakrishnan, R.; Shaft, U. When Is “Nearest Neighbor” Meaningful? *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **1998**, *1540*, 217–235. https://doi.org/10.1007/3-540-49257-7_15.
- (67) Naes, T.; Isaksson, T.; Kowalski, B. Locally Weighted Regression and Scatter Correction for Near-Infrared Reflectance Data. *Anal. Chem.* **1990**, *62* (7), 664–673. <https://doi.org/10.1021/ac00206a003>.
- (68) Holmes, G.; Hall, M.; Prank, E. Generating Rule Sets from Model Trees. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **1999**, *1747*, 1–12. https://doi.org/10.1007/3-540-46695-9_1.
- (69) Eriksson, L.; Trygg, J.; Wold, S. PLS-Trees®, a Top-down Clustering Approach. *J. Chemom.* **2009**, *23* (11), 569–580. <https://doi.org/10.1002/cem.1254>.
- (70) de Santana, F. B.; de Souza, A. M.; Poppi, R. J. Visible and near Infrared Spectroscopy Coupled to Random Forest to Quantify Some Soil Quality Parameters. *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.* **2018**, *191*, 454–462. <https://doi.org/10.1016/j.saa.2017.10.052>.
- (71) Nawar, S.; Mouazen, A. M. Comparison between Random Forests, Artificial Neural Networks and Gradient Boosted Machines Methods of on-Line Vis-NIR Spectroscopy Measurements of Soil Total Nitrogen and Total Carbon. *Sensors (Switzerland)* **2017**, *17* (10), 1–22. <https://doi.org/10.3390/s17102428>.
- (72) Cunha, C. L.; Torres, A. R.; Luna, A. S. Multivariate Regression Models Obtained from Near-Infrared Spectroscopy Data for Prediction of the Physical Properties of Biodiesel and Its Blends. *Fuel* **2020**, *261* (September 2019), 116344. <https://doi.org/10.1016/j.fuel.2019.116344>.
- (73) Borin, A.; Ferrão, M. F.; Mello, C.; Maretto, D. A.; Poppi, R. J. Least-Squares Support Vector Machines and near Infrared Spectroscopy for Quantification of Common Adulterants in Powdered Milk. *Anal. Chim. Acta* **2006**, *579* (1), 25–32. <https://doi.org/10.1016/j.aca.2006.07.008>.
- (74) Devos, O.; Ruckebusch, C.; Durand, A.; Duponchel, L.; Huvenne, J. P. Support Vector Machines (SVM) in near Infrared (NIR) Spectroscopy: Focus on Parameters Optimization and Model Interpretation. *Chemom. Intell. Lab. Syst.* **2009**, *96* (1), 27–33. <https://doi.org/10.1016/j.chemolab.2008.11.005>.
- (75) Marini, F.; Bucci, R.; Magrì, A. L.; Magrì, A. D. Artificial Neural Networks in Chemometrics: History, Examples and Perspectives. *Microchem. J.* **2008**, *88* (2), 178–185. <https://doi.org/10.1016/j.microc.2007.11.008>.
- (76) Acquarelli, J.; van Laarhoven, T.; Gerretzen, J.; Tran, T. N.; Buydens, L. M. C.; Marchiori, E. Convolutional Neural Networks for Vibrational Spectroscopic Data Analysis. *Anal. Chim. Acta* **2017**, *954*,

All references

- 22–31. <https://doi.org/10.1016/j.aca.2016.12.010>.
- (77) Cui, C.; Fearn, T. Modern Practical Convolutional Neural Networks for Multivariate Regression: Applications to NIR Calibration. *Chemom. Intell. Lab. Syst.* **2018**, *182* (February), 9–20. <https://doi.org/10.1016/j.chemolab.2018.07.008>.
- (78) Zeaiter, M.; Rutledge, D. Preprocessing Methods. *Compr. Chemom.* **2009**, *3*, 121–231. <https://doi.org/10.1016/B978-044452701-1.00074-0>.
- (79) Roger, J.; Boulet, J.-C.; Zeaiter, M.; Rutledge, D. N. Pre-Processing Methods. In *Comprehensive Chemometrics*; Elsevier, 2020; Vol. 3, pp 1–75. <https://doi.org/10.1016/b978-0-12-409547-2.14878-4>.
- (80) Bro, R.; Smilde, A. K. Centering and Scaling in Component Analysis. In *Journal of Chemometrics*; 2003; Vol. 17, pp 16–33. <https://doi.org/10.1002/cem.773>.
- (81) Barnes, R. J.; Dhanoa, M. S.; Lister, S. J. Standard Normal Variate Transformation and De-Trending of near-Infrared Diffuse Reflectance Spectra. *Appl. Spectrosc.* **1989**, *43* (5), 772–777. <https://doi.org/10.1366/0003702894202201>.
- (82) Clark, R. N.; Roush, T. L. Reflectance Spectroscopy: Quantitative Analysis Techniques for Remote Sensing Applications. *J. Geophys. Res. Solid Earth* **1984**, *89* (B7), 6329–6340. <https://doi.org/10.1029/JB089iB07p06329>.
- (83) Eilers, P. H. C.; Boelens, H. F. M. Baseline Correction with Asymmetric Least Squares Smoothing. *Life Sci.* **2005**, 1–26.
- (84) Peng, J.; Peng, S.; Jiang, A.; Wei, J.; Li, C.; Tan, J. *Asymmetric Least Squares for Multiple Spectra Baseline Correction*; 2010; Vol. 683. <https://doi.org/10.1016/j.aca.2010.08.033>.
- (85) Eilers, P. H. C. A Perfect Smoother. *Analytical Chemistry*. 2003, pp 3631–3636. <https://doi.org/10.1021/ac034173t>.
- (86) Savitzky, A.; Golay, M. J. E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* **1964**, *36* (8), 1627–1639. <https://doi.org/10.1021/ac60214a047>.
- (87) Leung, A. K. M.; Chau, F. T.; Gao, J. Bin. Wavelet Transform: A Method for Derivative Calculation in Analytical Chemistry. *Anal. Chem.* **1998**, *70* (24), 5222–5229. <https://doi.org/10.1021/ac9803737>.
- (88) Tan, J.; Sun, Y.; Ma, L.; Feng, H.; Guo, Y.; Cai, W.; Shao, X. Knowledge-Based Genetic Algorithm for Resolving the near-Infrared Spectrum and Understanding the Water Structures in Aqueous Solution. *Chemom. Intell. Lab. Syst.* **2020**, *206* (May). <https://doi.org/10.1016/j.chemolab.2020.104150>.
- (89) Chen, Z. P.; Morris, J.; Martin, E. Extracting Chemical Information from Spectral Data with Multiplicative Light Scattering Effects by Optical Path-Length Estimation and Correction. *Anal. Chem.* **2006**, *78* (22), 7674–7681. <https://doi.org/10.1021/ac0610255>.
- (90) Jin, J.; Chen, Z.; Li, L.; Steponavicius, R.; Thennadil, S. N.; Yang, J.; Yu, R. Quantitative Spectroscopic Analysis of Heterogeneous Mixtures : The Properties of Samples. *Anal. Chem.* **2012**, 320–326.
- (91) Isaksson, T.; Naes, T. Effect of Multiplicative Scatter Correction (MSC) and Linearity Improvement in NIR Spectroscopy. *Appl. Spectrosc.* **1988**, *42* (7), 1273–1284. <https://doi.org/10.1366/0003702884429869>.
- (92) Ilari, J. L.; Martens, H.; Isaksson, T. Determination of Particle Size in Power By Scatter Correction in Diffuse Near-Infrared Reflectance. *Appl. Spectrosc.* **1988**, *42* (5), 722–728. <https://doi.org/10.1366/0003702884429058>.
- (93) Martens, H.; Jensen, S. A.; Geladi, P. Multivariate Linearity Transformation for Near-Infrared Reflectance Spectrometry. In *Proceedings of the Nordic symposium on applied statistics*; 1983; pp 205–234.
- (94) Afseth, N. K.; Kohler, A. Extended Multiplicative Signal Correction in Vibrational Spectroscopy, a Tutorial. *Chemom. Intell. Lab. Syst.* **2012**, *117*, 92–99. <https://doi.org/10.1016/j.chemolab.2012.03.004>.
- (95) Martens, H.; Stark, E. Extended Multiplicative Signal Correction and Spectral Interference Subtraction: New Preprocessing Methods for near Infrared Spectroscopy. *J. Pharm. Biomed. Anal.* **1991**, *9* (8), 625–635. [https://doi.org/10.1016/0731-7085\(91\)80188-F](https://doi.org/10.1016/0731-7085(91)80188-F).
- (96) Martens, H.; Nielsen, J. P.; Engelsen, S. B. Light Scattering and Light Absorbance Separated by Extended Multiplicative Signal Correction. Application to near-Infrared Transmission Analysis of Powder Mixtures. *Anal. Chem.* **2003**, *75* (3), 394–404. <https://doi.org/10.1021/ac020194w>.
- (97) Martens, H.; Nielsen, J. P.; Engelsen, S. B. Light Scattering and Light Absorbance Separated by Extended Multiplicative Signal Correction. Application to near-Infrared Transmission Analysis of Powder Mixtures. *Anal. Chem.* **2003**, *75* (3), 394–404. <https://doi.org/10.1021/ac020194w>.

All references

- (98) Gallagher, N. B.; Blake, T. A.; Gassman, P. L. Application of Extended Inverse Scatter Correction to Mid-Infrared Reflectance Spectra of Soil. *J. Chemom.* **2005**, *19* (5–7), 271–281. <https://doi.org/10.1002/cem.929>.
- (99) Peiris, K. H. S.; Bean, S. R.; Jagadish, S. V. K. Extended Multiplicative Signal Correction to Improve Prediction Accuracy of Protein Content in Weathered Sorghum Grain Samples. *Cereal Chem.* **2020**, *97* (5), 1066–1074. <https://doi.org/10.1002/cche.10329>.
- (100) Afseth, N. K.; Segtnan, V. H.; Wold, J. P. Raman Spectra of Biological Samples: A Study of Preprocessing Methods. *Appl. Spectrosc.* **2006**, *60* (12), 1358–1367. <https://doi.org/10.1366/000370206779321454>.
- (101) Lieber, C. A.; Mahadevan-Jansen, A. Automated Method for Subtraction of Fluorescence from Biological Raman Spectra. *Appl. Spectrosc.* **2003**, *57* (11), 1363–1367. <https://doi.org/10.1366/00037020322554518>.
- (102) Thennadil, S. N.; Martin, E. B. Empirical Preprocessing Methods and Their Impact on NIR Calibrations: A Simulation Study. *J. Chemom.* **2005**, *19* (2), 77–89. <https://doi.org/10.1002/cem.912>.
- (103) Bertram, H. C.; Kohler, A.; Böcker, U.; Ofstad, R.; Andersen, H. J. Heat-Induced Changes in Myofibrillar Protein Structures and Myowater of Two Pork Qualities. A Combined FT-IR Spectroscopy and Low-Field NMR Relaxometry Study. *J. Agric. Food Chem.* **2006**, *54* (5), 1740–1746. <https://doi.org/10.1021/jf0514726>.
- (104) Kohler, A.; Böcker, U.; Warringer, J.; Blomberg, A.; Omholt, S. W.; Stark, E.; Martens, H. Reducing Inter-Replicate Variation in Fourier Transform Infrared Spectroscopy by Extended Multiplicative Signal Correction. *Appl. Spectrosc.* **2009**, *63* (3), 296–305. <https://doi.org/10.1366/000370209787598906>.
- (105) Li, Q.; Gao, Q.; Zhang, G. Improved Extended Multiplicative Scatter Correction Algorithm Applied in Blood Glucose Noninvasive Measurement with FT-IR Spectroscopy. *J. Spectrosc.* **2013**, *2013*, 1–5. <https://doi.org/10.1155/2013/916351>.
- (106) Sockalingum, G. D.; Tobin, M.; Bahrami, F.; Yang, Y.; Kohler, A.; Sule, J.; Pijanka, J.; Dumas, P.; Cotte, M.; Pittius, D. G. Van; Parkes, G.; Martens, H. Estimating and Correcting Mie Scattering in Synchrotron-Based Microscopic Fourier Transform Infrared Spectra by Extended Multiplicative Signal Correction. *Appl. Spectrosc.* **2008**, *62* (3), 259–266.
- (107) Rasskazov, I. L.; Singh, R.; Carney, P. S.; Bhargava, R. Extended Multiplicative Signal Correction for Infrared Microspectroscopy of Heterogeneous Samples with Cylindrical Domains. *Appl. Spectrosc.* **2019**, *73* (8), 859–869. <https://doi.org/10.1177/0003702819844528>.
- (108) Konevskikh, T.; Lukacs, R.; Blümel, R.; Ponomosov, A.; Kohler, A. Mie Scatter Corrections in Single Cell Infrared Microspectroscopy. *Faraday Discuss.* **2016**, *187*, 235–257. <https://doi.org/10.1039/c5fd00171d>.
- (109) Solheim, J. H.; Gunko, E.; Petersen, D.; Großertüschkamp, F.; Gerwert, K.; Kohler, A. An Open-Source Code for Mie Extinction Extended Multiplicative Signal Correction for Infrared Microscopy Spectra of Cells and Tissues. *J. Biophotonics* **2019**, *12* (8), 1–14. <https://doi.org/10.1002/jbio.201800415>.
- (110) Skogholt, J.; Liland, K. H.; Indahl, U. G. Preprocessing of Spectral Data in the Extended Multiplicative Signal Correction Framework Using Multiple Reference Spectra. *J. Raman Spectrosc.* **2019**, *50* (3), 407–417. <https://doi.org/10.1002/jrs.5520>.
- (111) Surowka, A. D.; Birarda, G.; Szczerbowska-Boruchowska, M.; Cestelli-Guidi, M.; Ziomber-Lisiak, A.; Vaccari, L. Model-Based Correction Algorithm for Fourier Transform Infrared Microscopy Measurements of Complex Tissue-Substrate Systems. *Anal. Chim. Acta* **2020**, *1103*, 143–155. <https://doi.org/10.1016/j.aca.2019.12.070>.
- (112) Mancini, M.; Toscano, G.; Rinnan, Å. Study of the Scattering Effects on NIR Data for the Prediction of Ash Content Using EMSC Correction Factors. *J. Chemom.* **2019**, *33* (4), 1–10. <https://doi.org/10.1002/cem.3111>.
- (113) Rabatel, G.; Marini, F.; Walczak, B.; Roger, J. M. VSN: Variable Sorting for Normalization. *J. Chemom.* **2020**, *34* (2), 1–16. <https://doi.org/10.1002/cem.3164>.
- (114) Isaksson, T.; Kowalski, B. Piece-Wise Multiplicative Scatter Correction Applied to near-Infrared Diffuse Transmittance Data from Meat Products. *Appl. Spectrosc.* **1993**, *47* (6), 702–709. <https://doi.org/10.1366/0003702934066839>.
- (115) Bi, Y.; Tang, L.; Shan, P.; Xie, Q.; Hu, Y.; Peng, S.; Tan, J.; Li, C. Interference Correction by Extracting the Information of Interference Dominant Regions: Application to near-Infrared Spectra. *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.* **2014**, *129*, 542–550. <https://doi.org/10.1016/j.saa.2014.03.080>.

All references

- (116) Liland, K. H.; Kohler, A.; Afseth, N. K. Model-Based Pre-Processing in Raman Spectroscopy of Biological Samples. *J. Raman Spectrosc.* **2016**, *47* (6), 643–650. <https://doi.org/10.1002/jrs.4886>.
- (117) Windig, W.; Shaver, J.; Bro, R. Loopy MSC: A Simple Way to Improve Multiplicative Scatter Correction. *Appl. Spectrosc.* **2008**, *62* (10), 1153–1159. <https://doi.org/10.1366/000370208786049097>.
- (118) Chen, H.; Lin, Z.; Mo, L.; Wu, H.; Wu, T.; Tan, C. Continuous Wavelet Transform-Based Feature Selection Applied to near-Infrared Spectral Diagnosis of Cancer. *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.* **2015**, *151* (8), 286–291. <https://doi.org/10.1016/j.saa.2015.06.109>.
- (119) Banskota, A.; Falkowski, M. J.; Smith, A. M. S.; Kane, E. S.; Meingast, K. M.; Bourgeau-Chavez, L. L.; Miller, M. E.; French, N. H. Continuous Wavelet Analysis for Spectroscopic Determination of Subsurface Moisture and Water-Table Height in Northern Peatland Ecosystems. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55* (3), 1526–1536. <https://doi.org/10.1109/TGRS.2016.2626460>.
- (120) Barros, A. S.; Rutledge, D. N. Segmented Principal Component Transform-Principal Component Analysis. *Chemom. Intell. Lab. Syst.* **2005**, *78* (1), 125–137. <https://doi.org/10.1016/j.chemolab.2005.01.003>.
- (121) Rännar, S.; MacGregor, J. F.; Wold, S. Adaptive Batch Monitoring Using Hierarchical PCA. In *Chemometrics and Intelligent Laboratory Systems*; 1998; Vol. 41, pp 73–81. [https://doi.org/10.1016/S0169-7439\(98\)00024-0](https://doi.org/10.1016/S0169-7439(98)00024-0).
- (122) Vigário, R. N. Extraction of Ocular Artefacts from EEG Using Independent Component Analysis. *Electroencephalogr. Clin. Neurophysiol.* **1997**, *103* (3), 395–404. [https://doi.org/10.1016/S0013-4694\(97\)00042-8](https://doi.org/10.1016/S0013-4694(97)00042-8).
- (123) Bouveresse, D. J. R.; Benabid, H.; Rutledge, D. N. Independent Component Analysis as a Pretreatment Method for Parallel Factor Analysis to Eliminate Artefacts from Multiway Data. *Anal. Chim. Acta* **2007**, *589* (2), 216–224. <https://doi.org/10.1016/j.aca.2007.02.061>.
- (124) Martens, H.; Høy, M.; Wise, B. M.; Bro, R.; Brockhoff, P. B. Pre-Whitening of Data by Covariance-Weighted Pre-Processing. *J. Chemom.* **2003**, *17* (3), 153–165. <https://doi.org/10.1002/cem.780>.
- (125) Roger, J. M.; Chauchard, F.; Bellon-Maurel, V. EPO-PLS External Parameter Orthogonalisation of PLS Application to Temperature-Independent Measurement of Sugar Content of Intact Fruits. *Chemom. Intell. Lab. Syst.* **2003**, *66* (2), 191–204. [https://doi.org/10.1016/S0169-7439\(03\)00051-0](https://doi.org/10.1016/S0169-7439(03)00051-0).
- (126) Andrew, A.; Fearn, T. Transfer by Orthogonal Projection: Making near-Infrared Calibrations Robust to between-Instrument Variation. *Chemom. Intell. Lab. Syst.* **2004**, *72* (1), 51–56. <https://doi.org/10.1016/j.chemolab.2004.02.004>.
- (127) Zeaiter, M.; Roger, J. M.; Bellon-Maurel, V. Dynamic Orthogonal Projection. A New Method to Maintain the on-Line Robustness of Multivariate Calibrations. Application to NIR-Based Monitoring of Wine Fermentations. *Chemom. Intell. Lab. Syst.* **2006**, *80* (2), 227–235. <https://doi.org/10.1016/j.chemolab.2005.06.011>.
- (128) Minasny, B.; McBratney, A. B.; Bellon-Maurel, V.; Roger, J. M.; Gobrecht, A.; Ferrand, L.; Joalland, S. Removing the Effect of Soil Moisture from NIR Diffuse Reflectance Spectra for the Prediction of Soil Organic Carbon. *Geoderma* **2011**, *167–168*, 118–124. <https://doi.org/10.1016/j.geoderma.2011.09.008>.
- (129) Acharya, U. K.; Walsh, K. B.; Subedi, P. P. Robustness of Partial Least-Squares Model to Change in Sample Temperature: I. A Comparison of Methods for Sucrose in Aqueous Solution. *J. Near Infrared Spectrosc.* **2014**, *22* (4), 279–286. <https://doi.org/10.1255/jnirs.1113>.
- (130) Giguere, S.; Boucher, T.; Carey, C. J.; Mahadevan, S.; Dyar, M. D. A Fully Customized Baseline Removal Framework for Spectroscopic Applications. *Appl. Spectrosc.* **2017**, *71* (7), 1457–1470. <https://doi.org/10.1177/0003702817695624>.
- (131) Roger, J.; Biancolillo, A.; Marini, F. Sequential Preprocessing through ORThogonalization (SPORT) and Its Application to near Infrared Spectroscopy. *Chemom. Intell. Lab. Syst.* **2020**, *199* (February), 103975. <https://doi.org/10.1016/j.chemolab.2020.103975>.
- (132) Mishra, P.; Roger, J. M.; Marini, F.; Biancolillo, A.; Rutledge, D. N. Parallel Pre-Processing through Orthogonalization (PORTO) and Its Application to near-Infrared Spectroscopy. *Chemom. Intell. Lab. Syst.* **2020**, No. July, 104190. <https://doi.org/10.1016/j.chemolab.2020.104190>.
- (133) Mishra, P.; Nordon, A.; Roger, J. M. Improved Prediction of Tablet Properties with Near-Infrared Spectroscopy by a Fusion of Scatter Correction Techniques. *J. Pharm. Biomed. Anal.* **2021**, *192*, 113684. <https://doi.org/10.1016/j.jpba.2020.113684>.
- (134) Oliveri, P.; Malegori, C.; Simonetti, R.; Casale, M. The Impact of Signal Pre-Processing on the Final

All references

- Interpretation of Analytical Outcomes – A Tutorial. *Anal. Chim. Acta* **2019**, *1058*, 9–17. <https://doi.org/10.1016/j.aca.2018.10.055>.
- (135) Mishra, P.; Rutledge, D. N.; Roger, J. M.; Wali, K.; Khan, H. A. Chemometric Pre-Processing Can Negatively Affect the Performance of near-Infrared Spectroscopy Models for Fruit Quality Prediction. *Talanta* **2021**, *229*, 122303. <https://doi.org/10.1016/j.talanta.2021.122303>.
- (136) Schoot, M.; Kapper, C.; van Kollenburg, G. H.; Postma, G. J.; van Kessel, G.; Buydens, L. M. C.; Jansen, J. J. Investigating the Need for Preprocessing of Near-Infrared Spectroscopic Data as a Function of Sample Size. *Chemom. Intell. Lab. Syst.* **2020**, *204* (April), 104105. <https://doi.org/10.1016/j.chemolab.2020.104105>.
- (137) Bogomolov, A.; Mannhardt, J.; Heinzerling, O. Accuracy Improvement of In-Line Near-Infrared Spectroscopic Moisture Monitoring in a Fluidized Bed Drying Process. *Front. Chem.* **2018**, *6* (October). <https://doi.org/10.3389/fchem.2018.00388>.
- (138) Lourenço, N. D.; Lopes, J. A.; Almeida, C. F.; Sarraguça, M. C.; Pinheiro, H. M. Bioreactor Monitoring with Spectroscopy and Chemometrics: A Review. *Anal. Bioanal. Chem.* **2012**, *404* (4), 1211–1237. <https://doi.org/10.1007/s00216-012-6073-9>.
- (139) Nordberg, Å.; Hansson, M.; Sundh, I.; Nordkvist, E.; Carlsson, H.; Mathisen, B. Monitoring of a Biogas Process Using Electronic Gas Sensors and Near- Infrared Spectroscopy (NIR). *Water Sci. Technol.* **2000**, *41* (3), 1–8. <https://doi.org/10.2166/wst.2000.0049>.
- (140) Skvaril, J.; Kyprianidis, K. G.; Dahlquist, E. Applications of Near-Infrared Spectroscopy (NIRS) in Biomass Energy Conversion Processes: A Review. *Appl. Spectrosc. Rev.* **2017**, *52* (8), 675–728. <https://doi.org/10.1080/05704928.2017.1289471>.
- (141) Jacobi, H. F.; Moschner, C. R.; Hartung, E. Use of near Infrared Spectroscopy in Monitoring of Volatile Fatty Acids in Anaerobic Digestion. *Water Sci. Technol.* **2009**, *60* (2), 339–346. <https://doi.org/10.2166/wst.2009.345>.
- (142) Holm-Nielsen, J. B.; Andrée, H.; Lindorfer, H.; Esbensen, K. H. Transflexive Embedded near Infrared Monitoring for Key Process Intermediates in Anaerobic Digestion/ Biogas Production. *J. Near Infrared Spectrosc.* **2007**, *15* (2), 123–135. <https://doi.org/10.1255/jnirs.719>.
- (143) Krapf, L. C.; Gronauer, A.; Schmidhalter, U.; Heuwinkel, H. Near Infrared Spectroscopy Calibrations for the Estimation of Process Parameters of Anaerobic Digestion of Energy Crops and Livestock Residues. *J. Near Infrared Spectrosc.* **2011**, *19* (6), 479–493. <https://doi.org/10.1255/jnirs.960>.
- (144) Krapf, L. C.; Gronauer, A.; Schmidhalter, U.; Heuwinkel, H. Evaluation of Agricultural Feedstock-Robust near Infrared Calibrations for the Estimation of Process Parameters in Anaerobic Digestion. *J. Near Infrared Spectrosc.* **2012**, *20* (4), 465–476. <https://doi.org/10.1255/jnirs.1013>.
- (145) Stockl, A.; Lichti, F. Near-Infrared Spectroscopy (NIRS) for a Real Time Monitoring of the Biogas Process. *Bioresour. Technol.* **2018**, *247* (September 2017), 1249–1252. <https://doi.org/10.1016/j.biortech.2017.09.173>.
- (146) Awhangbo, L.; Bendoula, R.; Roger, J. M.; Béline, F. Multi-Block SO-PLS Approach Based on Infrared Spectroscopy for Anaerobic Digestion Process Monitoring. *Chemom. Intell. Lab. Syst.* **2020**, *196* (December 2019), 103905. <https://doi.org/10.1016/j.chemolab.2019.103905>.
- (147) Reed, J. P.; Devlin, D.; Esteves, S. R. R.; Dinsdale, R.; Guwy, A. J. Performance Parameter Prediction for Sewage Sludge Digesters Using Reflectance FT-NIR Spectroscopy. *Water Res.* **2011**, *45* (8), 2463–2472. <https://doi.org/10.1016/j.watres.2011.01.027>.
- (148) Awhangbo, L.; Bendoula, R.; Roger, J. M.; Béline, F. Multi-Block Data Analysis for Online Monitoring of Anaerobic Co-Digestion Process. *Chemom. Intell. Lab. Syst.* **2020**, *205* (January). <https://doi.org/10.1016/j.chemolab.2020.104120>.
- (149) Stuth, J.; Jama, A.; Tolleson, D. Direct and Indirect Means of Predicting Forage Quality through near Infrared Reflectance Spectroscopy. *F. Crop. Res.* **2003**, *84* (1–2), 45–56. [https://doi.org/10.1016/S0378-4290\(03\)00140-0](https://doi.org/10.1016/S0378-4290(03)00140-0).
- (150) Raposo, F.; De La Rubia, M. A.; Fernández-Cegrí, V.; Borja, R. Anaerobic Digestion of Solid Organic Substrates in Batch Mode: An Overview Relating to Methane Yields and Experimental Procedures. *Renew. Sustain. Energy Rev.* **2012**, *16* (1), 861–877. <https://doi.org/10.1016/j.rser.2011.09.008>.
- (151) Hafner, S. D.; de Laclos, H. F.; Koch, K.; Holliger, C. Improving Inter-Laboratory Reproducibility in Measurement of Biochemical Methane Potential (BMP). *Water (Switzerland)* **2020**, *12* (6), 1752. <https://doi.org/10.3390/w12061752>.

All references

- (152) Lesteur, M.; Bellon-Maurel, V.; Gonzalez, C.; Latrille, E.; Roger, J. M.; Junqua, G.; Steyer, J. P. Alternative Methods for Determining Anaerobic Biodegradability: A Review. *Process Biochem.* **2010**, *45* (4), 431–440. <https://doi.org/10.1016/j.procbio.2009.11.018>.
- (153) Ward, A. J. Near-Infrared Spectroscopy for Determination of the Biochemical Methane Potential: State of the Art. *Chem. Eng. Technol.* **2016**, *39* (4), 611–619. <https://doi.org/10.1002/ceat.201500315>.
- (154) Charnier, C.; Latrille, E.; Jimenez, J.; Lemoine, M.; Boulet, J. C.; Miroux, J.; Steyer, J. P. Fast Characterization of Solid Organic Waste Content with near Infrared Spectroscopy in Anaerobic Digestion. *Waste Manag.* **2017**, *59* (1), 140–148. <https://doi.org/10.1016/j.wasman.2016.10.029>.
- (155) Grieder, C.; Mittweg, G.; Dhillon, B. S.; Montes, J. M.; Orsinia, E.; Melchinger, A. E. Determination of Methane Fermentation Yield and Its Kinetics by near Infrared Spectroscopy and Chemical Composition in Maize. *J. Near Infrared Spectrosc.* **2011**, *19* (6), 463–477. <https://doi.org/10.1255/jnirs.959>.
- (156) Kandel, T. P.; Gislum, R.; Jørgensen, U.; Lærke, P. E. Prediction of Biogas Yield and Its Kinetics in Reed Canary Grass Using near Infrared Reflectance Spectroscopy and Chemometrics. *Bioresour. Technol.* **2013**, *146*, 282–287. <https://doi.org/10.1016/j.biortech.2013.07.092>.
- (157) Buffiere, P.; Loisel, D.; Bernet, N.; Delgenes, J. P. Towards New Indicators for the Prediction of Solid Waste Anaerobic Digestion Properties. *Water Sci. Technol.* **2006**, *53* (8), 233–241. <https://doi.org/10.2166/wst.2006.254>.
- (158) Edwiges, T.; Frare, L.; Mayer, B.; Lins, L.; Mi Triolo, J.; Flotats, X.; de Mendonça Costa, M. S. S. Influence of Chemical Composition on Biochemical Methane Potential of Fruit and Vegetable Waste. *Waste Manag.* **2018**, *71*, 618–625. <https://doi.org/10.1016/j.wasman.2017.05.030>.
- (159) Lesteur, M.; Latrille, E.; Maurel, V. B.; Roger, J. M.; Gonzalez, C.; Junqua, G.; Steyer, J. P. First Step towards a Fast Analytical Method for the Determination of Biochemical Methane Potential of Solid Wastes by near Infrared Spectroscopy. *Bioresour. Technol.* **2011**, *102* (3), 2280–2288. <https://doi.org/10.1016/j.biortech.2010.10.044>.
- (160) Raju, C. S.; Ward, A. J.; Nielsen, L.; Møller, H. B. Comparison of near Infra-Red Spectroscopy, Neutral Detergent Fibre Assay and in-Vitro Organic Matter Digestibility Assay for Rapid Determination of the Biochemical Methane Potential of Meadow Grasses. *Bioresour. Technol.* **2011**, *102* (17), 7835–7839. <https://doi.org/10.1016/j.biortech.2011.05.049>.
- (161) Wolf, D.; von Canstein, H.; Schröder, C. Optimisation of Biogas Production by Infrared Spectroscopy-Based Process Control. *J. Nat. Gas Sci. Eng.* **2011**, *3* (5), 625–632. <https://doi.org/10.1016/j.jngse.2011.07.006>.
- (162) Jacobi, H. F.; Ohl, S.; Thiessen, E.; Hartung, E. NIRS-Aided Monitoring and Prediction of Biogas Yields from Maize Silage at a Full-Scale Biogas Plant Applying Lumped Kinetics. *Bioresour. Technol.* **2012**, *103* (1), 162–172. <https://doi.org/10.1016/j.biortech.2011.10.012>.
- (163) Doublet, J.; Boulanger, A.; Ponthieux, A.; Laroche, C.; Poitrenaud, M.; Cacho Rivero, J. A. Predicting the Biochemical Methane Potential of Wide Range of Organic Substrates by near Infrared Spectroscopy. *Bioresour. Technol.* **2013**, *128*, 252–258. <https://doi.org/10.1016/j.biortech.2012.10.044>.
- (164) Mayer, F.; Noo, A.; Sinnaeve, G.; Dardenne, P.; Gerin, P. a; Delfosse, P. Prediction of the Biochemical Methane Potential (BMP) of Maize Silages Reduced to a Powder Using NIR Spectra from Wet and Dried Samples. In *NIR 2013 France*; Citeseer, 2013; pp 458–463.
- (165) Triolo, J. M.; Ward, A. J.; Pedersen, L.; Løkke, M. M.; Qu, H.; Sommer, S. G. Near Infrared Reflectance Spectroscopy (NIRS) for Rapid Determination of Biochemical Methane Potential of Plant Biomass. *Appl. Energy* **2014**, *116*, 52–57. <https://doi.org/10.1016/j.apenergy.2013.11.006>.
- (166) Godin, B.; Mayer, F.; Agneessens, R.; Gerin, P.; Dardenne, P.; Delfosse, P.; Delcarte, J. Biochemical Methane Potential Prediction of Plant Biomasses: Comparing Chemical Composition versus near Infrared Methods and Linear versus Non-Linear Models. *Bioresour. Technol.* **2015**, *175*, 382–390. <https://doi.org/10.1016/j.biortech.2014.10.115>.
- (167) Wahid, R.; Ward, A. J.; Møller, H. B.; Sjøgaard, K.; Eriksen, J. Biogas Potential from Forbs and Grass-Clover Mixture with the Application of near Infrared Spectroscopy. *Bioresour. Technol.* **2015**, *198*, 124–132. <https://doi.org/10.1016/j.biortech.2015.08.154>.
- (168) Preys, S.; Lallemand, J.; Roussel, S. Flash BMP®: Calibration for the Biochemical Methane Potential (BMP) of Solid Organic Waste Using NIRS. In *NIR 2015 Brazil*; 2015; p 80006099.
- (169) Fitamo, T.; Triolo, J. M.; Boldrin, A.; Scheutz, C. Rapid Biochemical Methane Potential Prediction of Urban Organic Waste with Near-Infrared Reflectance Spectroscopy. *Water Res.* **2017**, *119*, 242–251.

All references

- <https://doi.org/10.1016/j.watres.2017.04.051>.
- (170) Mortreuil, P.; Baggio, S.; Lagnet, C.; Schraauwers, B.; Monlau, F. Fast Prediction of Organic Wastes Methane Potential by near Infrared Reflectance Spectroscopy: A Successful Tool for Farm-Scale Biogas Plant Monitoring. *Waste Manag. Res.* **2018**, *36* (9), 800–809. <https://doi.org/10.1177/0734242X18778773>.
- (171) Yang, G.; Li, Y.; Zhen, F.; Xu, Y.; Liu, J.; Li, N.; Sun, Y.; Luo, L.; Wang, M.; Zhang, L. Biochemical Methane Potential Prediction for Mixed Feedstocks of Straw and Manure in Anaerobic Co-Digestion. *Bioresour. Technol.* **2021**, *326* (December 2020), 124745. <https://doi.org/10.1016/j.biortech.2021.124745>.
- (172) Liu, J.; Jin, S.; Bao, C.; Sun, Y.; Li, W. Rapid Determination of Lignocellulose in Corn Stover Based on Near-Infrared Reflectance Spectroscopy and Chemometrics Methods. *Bioresour. Technol.* **2021**, *321* (November 2020), 124449. <https://doi.org/10.1016/j.biortech.2020.124449>.
- (173) Mallet, A.; Péréme, M.; Awhangbo, L.; Charnier, C.; Roger, J. M.; Steyer, J. P.; Latrille, É.; Bendoula, R. Fast At-Line Characterization of Solid Organic Waste: Comparing Analytical Performance of Different Compact near Infrared Spectroscopic Systems with Different Measurement Configurations. *Waste Manag.* **2021**, *126*, 664–673. <https://doi.org/10.1016/j.wasman.2021.03.045>.
- (174) Jacobi, H. F.; Moschner, C. R.; Hartung, E. Use of near Infrared Spectroscopy in Online-Monitoring of Feeding Substrate Quality in Anaerobic Digestion. *Bioresour. Technol.* **2011**, *102* (7), 4688–4696. <https://doi.org/10.1016/j.biortech.2011.01.035>.
- (175) Steyer, P.; Bouvier, J. C.; Conte, T.; Gras, P.; Harmand, J.; Delgenes, J. P. On-Line Measurements of COD, TOC, VFA, Total and Partial Alkalinity in Anaerobic Digestion Processes Using Infra-Red Ectrometry. *Water Sci. Technol.* **2002**, *45* (10), 133–138. <https://doi.org/10.2166/wst.2002.0310>.
- (176) Bekiaris, G.; Triolo, J. M.; Peltre, C.; Pedersen, L.; Jensen, L. S.; Bruun, S. Rapid Estimation of the Biochemical Methane Potential of Plant Biomasses Using Fourier Transform Mid-Infrared Photoacoustic Spectroscopy. *Bioresour. Technol.* **2015**, *197*, 475–481. <https://doi.org/10.1016/j.biortech.2015.08.050>.
- (177) Bekiaris, G.; Peltre, C.; Barsberg, S. T.; Bruun, S.; Sørensen, K. M.; Engelsen, S. B.; Magid, J.; Hansen, M.; Jensen, L. S. Three Different Fourier-Transform Mid-Infrared Sampling Techniques to Characterize Bio-Organic Samples. *J. Environ. Qual.* **2020**, *49* (5), 1310–1321. <https://doi.org/10.1002/jeq2.20106>.
- (178) Rodrigues, R. P.; Rodrigues, D. P.; Klepacz-Smolka, A.; Martins, R. C.; Quina, M. J. Comparative Analysis of Methods and Models for Predicting Biochemical Methane Potential of Various Organic Substrates. *Sci. Total Environ.* **2019**, *649*, 1599–1608. <https://doi.org/10.1016/j.scitotenv.2018.08.270>.
- (179) Wu, D.; Li, L.; Zhao, X.; Peng, Y.; Yang, P.; Peng, X. Anaerobic Digestion: A Review on Process Monitoring. *Renew. Sustain. Energy Rev.* **2019**, *103* (December 2018), 1–12. <https://doi.org/10.1016/j.rser.2018.12.039>.
- (180) Derikx, P. J. L.; Willers, H. C.; ten Have, P. J. W. Effect of PH on the Behaviour of Volatile Compounds in Organic Manures during Dry-Matter Determination. *Bioresour. Technol.* **1994**, *49* (1), 41–45. [https://doi.org/10.1016/0960-8524\(94\)90171-6](https://doi.org/10.1016/0960-8524(94)90171-6).
- (181) Buffiere, P.; Frederic, S.; Marty, B.; Delgenes, J. P. A Comprehensive Method for Organic Matter Characterization in Solid Wastes in View of Assessing Their Anaerobic Biodegradability. *Water Sci. Technol.* **2008**, *58* (9), 1783–1788. <https://doi.org/10.2166/wst.2008.517>.
- (182) Petersen, L.; Minkkinen, P.; Esbensen, K. H. Representative Sampling for Reliable Data Analysis: Theory of Sampling. *Chemom. Intell. Lab. Syst.* **2005**, *77* (1-2 SPEC. ISS.), 261–277. <https://doi.org/10.1016/j.chemolab.2004.09.013>.
- (183) Yan, H.; Siesler, H. W. Hand-Held near-Infrared Spectrometers: State-of-the-Art Instrumentation and Practical Applications. *NIR news* **2018**, *29* (7), 8–12. <https://doi.org/10.1177/0960336018796391>.
- (184) Beć, K. B.; Grabska, J.; Siesler, H. W.; Huck, C. W. Handheld Near-Infrared Spectrometers: Where Are We Heading? *NIR news* **2020**, *31* (3–4), 28–35. <https://doi.org/10.1177/0960336020916815>.
- (185) Crocombe, R. A. Portable Spectroscopy. *Appl. Spectrosc.* **2018**, *72* (12), 1701–1751. <https://doi.org/10.1177/0003702818809719>.
- (186) Beć, K. B.; Grabska, J.; Huck, C. W. Principles and Applications of Miniaturized Near-Infrared (NIR) Spectrometers. *Chem. - A Eur. J.* **2021**, *27* (5), 1514–1532. <https://doi.org/10.1002/chem.202002838>.
- (187) Tang, Y.; Jones, E.; Minasny, B. Evaluating Low-Cost Portable near Infrared Sensors for Rapid Analysis of Soils from South Eastern Australia. *Geoderma Reg.* **2020**, *20* (3), e00240. <https://doi.org/10.1016/j.geodrs.2019.e00240>.
- (188) Sharififar, A.; Singh, K.; Jones, E.; Ginting, F. I.; Minasny, B. Evaluating a Low-Cost Portable NIR

All references

- Spectrometer for the Prediction of Soil Organic and Total Carbon Using Different Calibration Models. *Soil Use Manag.* **2019**, *35* (4), 607–616. <https://doi.org/10.1111/sum.12537>.
- (189) McVey, C.; McGrath, T. F.; Haughey, S. A.; Elliott, C. T. A Rapid Food Chain Approach for Authenticity Screening: The Development, Validation and Transferability of a Chemometric Model Using Two Handheld near Infrared Spectroscopy (NIRS) Devices. *Talanta* **2021**, *222* (August 2020), 121533. <https://doi.org/10.1016/j.talanta.2020.121533>.
- (190) Chang, C.-W.; Laird, D. A.; Hurburgh, C. R. Influence of Soil Moisture on Near-Infrared Reflectance Spectroscopic Measurement of Soil Properties. *Soil Sci.* **2005**, *170* (4), 244–255. <https://doi.org/10.1097/01.ss.0000162289.40879.7b>.
- (191) Lobell, D. B.; Asner, G. P. Moisture Effects on Soil Reflectance. *Soil Sci. Soc. Am. J.* **2002**, *66* (3), 722. <https://doi.org/10.2136/sssaj2002.7220>.
- (192) Knadel, M.; Deng, F.; Alinejadian, A.; Wollesen de Jonge, L.; Moldrup, P.; Greve, M. H. The Effects of Moisture Conditions—From Wet to Hyper Dry—On Visible Near-Infrared Spectra of Danish Reference Soils. *Soil Sci. Soc. Am. J.* **2014**, *78* (2), 422. <https://doi.org/10.2136/sssaj2012.0401>.
- (193) Bogrekci, I.; Lee, W. S. Effects of Soil Moisture Content on Absorbance Spectra of Sandy Soils in Sensing Phosphorus Concentrations Using UV-VIS-NIR Spectroscopy. *Trans. ASABE* **2006**, *49* (4), 1175–1180.
- (194) Bowers, S. A.; Hanks, R. J. Reflection of Radiant Energy from Soils. *Soil Sci.* **1965**, *100* (2), 130–138. <https://doi.org/10.1097/00010694-196508000-00009>.
- (195) Sudduth, K. A.; Hummel, J. W. Soil Organic Matter, CEC, and Moisture Sensing with a Portable NIR Spectrophotometer. *Trans. - Am. Soc. Agric. Eng.* **1993**, *36* (6), 1571–1582. <https://doi.org/10.13031/2013.28498>.
- (196) Wu, C. Y.; Jacobson, A. R.; Laba, M.; Baveye, P. C. Alleviating Moisture Content Effects on the Visible Near-Infrared Diffuse-Reflectance Sensing of Soils. *Soil Sci.* **2009**, *174* (8), 456–465. <https://doi.org/10.1097/SS.0b013e3181b21491>.
- (197) Williams, P. Influence of Water on Prediction of Composition and Quality Factors: The Aquaphotomics of Low Moisture Agricultural Materials. *J. Near Infrared Spectrosc.* **2009**, *17* (6), 315–328. <https://doi.org/10.1255/jnirs.862>.
- (198) Gaines, C. S.; Windham, W. R. Effect of Wheat Moisture Content on Meal Apparent Particle Size and Hardness Scores Determined by Near-Infrared Reflectance Spectroscopy. *Cereal Chem.* **1998**, *75* (3), 386–391. <https://doi.org/10.1094/CCHEM.1998.75.3.386>.
- (199) Peiris, K. H. S.; Dong, Y.; Bockus, W. W.; Dowell, F. E. Moisture Effects on the Prediction Performance of a Single-Kernel Near-Infrared Deoxynivalenol Calibration. *Cereal Chem. J.* **2016**, *93* (6), 631–637. <https://doi.org/10.1094/CCHEM-04-16-0120-R>.
- (200) Gergely, S.; Salgó, A. Changes in Moisture Content during Wheat Maturation - What Is Measured by near Infrared Spectroscopy? *J. Near Infrared Spectrosc.* **2003**, *11* (1), 17–26. <https://doi.org/10.1255/jnirs.350>.
- (201) Popineau, S.; Rondeau-Mouro, C.; Sulpice-Gaillet, C.; Shanahan, M. E. R. Free/Bound Water Absorption in an Epoxy Adhesive. *Polymer (Guildf.)* **2005**, *46* (24), 10733–10740. <https://doi.org/10.1016/j.polymer.2005.09.008>.
- (202) Büning-Pfaue, H. Analysis of Water in Food by near Infrared Spectroscopy. *Food Chem.* **2003**, *82* (1), 107–115. [https://doi.org/10.1016/S0308-8146\(02\)00583-6](https://doi.org/10.1016/S0308-8146(02)00583-6).
- (203) Carter, G. A. Primary and Secondary Effects of Water Content on the Spectral Reflectance of Leaves. *Am. J. Bot.* **1991**, *78* (7), 916–924. <https://doi.org/10.1002/j.1537-2197.1991.tb14495.x>.
- (204) Giordanengo, T.; Charpentier, J. P.; Roger, J. M.; Roussel, S.; Brancheriau, L.; Chaix, G.; Bailleres, H. Correction of Moisture Effects on near Infrared Calibration for the Analysis of Phenol Content in Eucalyptus Wood Extracts. *Ann. For. Sci.* **2008**, *65* (8). <https://doi.org/Artn 803rDoi 10.1051/Forest:2008065>.
- (205) Igne, B.; Hossain, M. N.; Drennen, J. K.; Anderson, C. A. Robustness Considerations and Effects of Moisture Variations on near Infrared Method Performance for Solid Dosage Form Assay. *J. Near Infrared Spectrosc.* **2014**, *22* (3), 179–188. <https://doi.org/10.1255/jnirs.1097>.
- (206) Wenz, J. J. Examining Water in Model Membranes by near Infrared Spectroscopy and Multivariate Analysis. *Biochim. Biophys. Acta - Biomembr.* **2018**, *1860* (3), 673–682. <https://doi.org/10.1016/j.bbamem.2017.12.007>.
- (207) Reeves, J. B. Effects of Water on the Spectra of Model Compounds in the Short-Wavelength near Infrared

All references

- Spectral Region (14,000-9091 cm^{-1} or 714-1100 nm). **1994**, 212, 199–212.
- (208) Reeves, J. B. Efforts to Quantify Changes in Near-Infrared Spectra Caused by the Influence of Water, pH, Ionic Strength, and Differences in Physical State. *Appl. Spectrosc.* **1995**, 49 (2), 181–187. <https://doi.org/10.1366/0003702953963788>.
- (209) Muncan, J.; Tsenkova, R. Aquaphotomics—From Innovative Knowledge to Integrative Platform in Science and Technology. *Molecules* **2019**, 24 (15), 2742. <https://doi.org/10.3390/molecules24152742>.
- (210) Caponigro, V.; Marini, F.; Gowen, A. Hydration of Hydrogels Studied by Near-Infrared Hyperspectral Imaging. *J. Chemom.* **2018**, 32 (1), 1–19. <https://doi.org/10.1002/cem.2972>.
- (211) Raponi, F.; Ferri, S.; Monarca, D.; Moschetti, R.; Colantoni, A.; Massantini, R. Real-Time Monitoring of Organic Apple (Var. Gala) during Hot-Air Drying Using near-Infrared Spectroscopy. *J. Food Eng.* **2017**, 222, 139–150. <https://doi.org/10.1016/j.jfoodeng.2017.11.023>.
- (212) Luck, W. A. *Structure of Water and Aqueous Solutions*; Verlag Chemie, 1974.
- (213) Reeves III, J. B. Influence of Water on the Near Infrared Spectra of Model Compounds. *J. AOAC Int.* **1993**, 76 (4), 741–748. <https://doi.org/10.1093/jaoac/76.4.741>.
- (214) Giangiacomo, R. Study of Water-Sugar Interactions at Increasing Sugar Concentration by NIR Spectroscopy. *Food Chem.* **2006**, 96 (3), 371–379. <https://doi.org/10.1016/j.foodchem.2005.02.051>.
- (215) Tsenkova, R.; Pollner, B.; Kovacs, Z. Essentials of Aquaphotomics and Its Chemometrics Approaches. **2018**, 6 (August), 1–25. <https://doi.org/10.3389/fchem.2018.00363>.
- (216) Segtnan, V. H.; Šašić, Š.; Isaksson, T.; Ozaki, Y. Studies on the Structure of Water Using Spectroscopy and Principal Component Analysis. *Anal. Chem.* **2001**, 73 (13), 3153–3161. <https://doi.org/10.1021/ac010102n>.
- (217) Chaplin, M. F. Anomalous properties of water http://www1.lsbu.ac.uk/water/water_anomalies.html (accessed Mar 15, 2020).
- (218) Nostro, P. Lo; Ninham, B. W. *Aqua Incognita: Why Ice Floats on Water and Galileo 400 Years On*; Connor Court Publishing, 2014.
- (219) Chaplin, M. F. Structure and Properties of Water in Its Various States. *Encycl. Water* **2019**, 1–19. <https://doi.org/10.1002/9781119300762.wsts0002>.
- (220) Maeda, H.; Ozaki, Y.; Tanaka, M.; Hayashi, N.; Kojima, T. Near Infrared Spectroscopy and Chemometrics Studies of Temperature-Dependent Spectral Variations of Water: Relationship between Spectral Changes and Hydrogen Bonds. *J. Near Infrared Spectrosc.* **1995**, 3 (4), 191–201. <https://doi.org/10.1255/jnirs.69>.
- (221) Cui, X.; Cai, W.; Shao, X. Glucose Induced Variation of Water Structure from Temperature Dependent near Infrared Spectra. *RSC Adv.* **2016**, 6 (107), 105729–105736. <https://doi.org/10.1039/c6ra18912a>.
- (222) Gowen, A. A.; Amigo, J. M.; Tsenkova, R. Characterisation of Hydrogen Bond Perturbations in Aqueous Systems Using Aquaphotomics and Multivariate Curve Resolution-Alternating Least Squares. *Anal. Chim. Acta* **2013**, 759 (June 2012), 8–20. <https://doi.org/10.1016/j.aca.2012.10.007>.
- (223) Segtnan, V. H.; Šašić, Š.; Isaksson, T.; Ozaki, Y. Studies on the Structure of Water Using Two-Dimensional near-Infrared Correlation Spectroscopy and Principal Component Analysis. *Anal. Chem.* **2001**, 73 (13), 3153–3161. <https://doi.org/10.1021/ac010102n>.
- (224) Renati, P.; Kovacs, Z.; De Ninno, A.; Tsenkova, R. Temperature Dependence Analysis of the NIR Spectra of Liquid Water Confirms the Existence of Two Phases, One of Which Is in a Coherent State. *J. Mol. Liq.* **2019**, 292, 111449. <https://doi.org/10.1016/j.molliq.2019.111449>.
- (225) Robinson, G. W.; Cho, C. H.; Urquidi, J. Isosbestic Points in Liquid Water: Further Strong Evidence for the Two-State Mixture Model. *J. Chem. Phys.* **1999**, 111 (2), 698–702. <https://doi.org/10.1063/1.479349>.
- (226) Geissler, P. L. Temperature Dependence of Inhomogeneous Broadening: On the Meaning of Isosbestic Points. *J. Am. Chem. Soc.* **2005**, 127 (42), 14930–14935. <https://doi.org/10.1021/ja0545214>.
- (227) Tsenkova, R. Aquaphotomics: Dynamic Spectroscopy of Aqueous and Biological Systems Describes Peculiarities of Water. *J. Near Infrared Spectrosc.* **2009**, 17 (6), 303–313. <https://doi.org/10.1255/jnirs.869>.
- (228) Kojić, D.; Tsenkova, R.; Tomobe, K.; Yasuoka, K.; Yasui, M. Water Confined in the Local Field of Ions. *ChemPhysChem* **2014**, 15 (18), 4077–4086. <https://doi.org/10.1002/cphc.201402381>.
- (229) Peiris, K. H. S.; Bean, S. R.; Chiluwal, A.; Perumal, R.; Jagadish, S. V. K. Moisture Effects on Robustness of Sorghum Grain Protein Near-Infrared Spectroscopy Calibration. *Cereal Chem.* **2019**, 96 (4), 678–688. <https://doi.org/10.1002/cche.10164>.

All references

- (230) Park, H. S.; Lee, J. K.; Fike, J. H.; Kim, D. A.; Ko, M. S.; Ha, J. K. Effect of Sample Preparation on Prediction of Fermentation Quality of Maize Silages by near Infrared Reflectance Spectroscopy. *Asian-Australasian J. Anim. Sci.* **2005**, *18* (5), 643–648. <https://doi.org/10.5713/ajas.2005.643>.
- (231) Reeves, J. B.; Blosser, T. H.; Colenbrander, V. F. Near Infrared Reflectance Spectroscopy for Analyzing Undried Silage. *J. Dairy Sci.* **1989**, *72* (1), 79–88. [https://doi.org/10.3168/jds.S0022-0302\(89\)79082-2](https://doi.org/10.3168/jds.S0022-0302(89)79082-2).
- (232) Lovett, D. K.; Deaville, E. R.; Givens, D. I.; Finlay, M.; Owen, E. Near Infrared Reflectance Spectroscopy (NIRS) to Predict Biological Parameters of Maize Silage: Effects of Particle Comminution, Oven Drying Temperature and the Presence of Residual Moisture. *Anim. Feed Sci. Technol.* **2005**, *120* (3–4), 323–332. <https://doi.org/10.1016/j.anifeedsci.2005.02.001>.
- (233) Ikoyi, A. Y.; Younge, B. A. Influence of Forage Particle Size and Residual Moisture on near Infrared Reflectance Spectroscopy (NIRS) Calibration Accuracy for Macro-Mineral Determination. *Anim. Feed Sci. Technol.* **2020**, *270* (May), 114674. <https://doi.org/10.1016/j.anifeedsci.2020.114674>.
- (234) Finzi, A.; Oberti, R.; Negri, A. S.; Perazzolo, F.; Coccolo, G.; Tambone, F.; Cabassi, G.; Provolo, G. Effects of Measurement Technique and Sample Preparation on NIR Spectroscopy Analysis of Livestock Slurry and Digestates. *Biosyst. Eng.* **2015**, *134*, 42–54. <https://doi.org/10.1016/j.biosystemseng.2015.03.015>.
- (235) Thyholt, K.; Isaksson, T. Differentiation of Frozen and Unfrozen Beef Using Near-Infrared Spectroscopy. *J. Sci. Food Agric.* **1997**, *73* (4), 525–532. [https://doi.org/10.1002/\(sici\)1097-0010\(199704\)73:4<525::aid-jsfa767>3.3.co;2-3](https://doi.org/10.1002/(sici)1097-0010(199704)73:4<525::aid-jsfa767>3.3.co;2-3).
- (236) Stenberg, B. Effects of Soil Sample Pretreatments and Standardised Rewetting as Interacted with Sand Classes on Vis-NIR Predictions of Clay and Soil Organic Carbon. *Geoderma* **2010**, *158* (1–2), 15–22. <https://doi.org/10.1016/j.geoderma.2010.04.008>.
- (237) Roberts, J. J.; Cozzolino, D. Wet or Dry? The Effect of Sample Characteristics on the Determination of Soil Properties by near Infrared Spectroscopy. *TrAC - Trends Anal. Chem.* **2016**, *83*, 25–30. <https://doi.org/10.1016/j.trac.2016.08.002>.
- (238) Saleh, K.; Forny, L.; Guigon, P.; Pezron, I. Dry Water: From Physico-Chemical Aspects to Process-Related Parameters. *Chem. Eng. Res. Des.* **2011**, *89* (5), 537–544. <https://doi.org/10.1016/j.cherd.2010.06.005>.
- (239) Binks, B. P.; Murakami, R. Phase Inversion of Particle-Stabilized Materials from Foams to Dry Water. *Nat. Mater.* **2006**, *5* (11), 865–869. <https://doi.org/10.1038/nmat1757>.
- (240) Alfaro, G.; Meurens, M.; Birth, G. S. Liquid Analysis by Dry-Extract near-Infrared Reflectance on Fiberglass. *Appl. Spectrosc.* **1990**, *44* (6), 979–986. <https://doi.org/10.1366/0003702904086687>.
- (241) Kumar Acharya, U. *Extending the Use of NIRS in the Assessment of Fruit Maturation and Quality*; 2015.
- (242) Meurens, M. Fast Drying Systems for NIR Analysis. *NIR news* **1992**, *3* (5), 12–13. <https://doi.org/10.1255/nirn.149>.
- (243) IRSTEA. Dairy Products. In *NIR 2013 Proceedings*; La Grande-, 2013.
- (244) Thyholt, K.; Isaksson, T. Near Infrared Spectroscopy of Dry Extracts from High Moisture Food Products on Solid Support - A Review. *Journal of Near Infrared Spectroscopy*. SAGE Publications Sage UK: London, England October 2, 1997, pp 179–193. <https://doi.org/10.1255/jnirs.112>.
- (245) Núñez-Sánchez, N.; Martínez-Marín, A. L.; Polvillo, O.; Fernández-Cabanás, V. M.; Carrizosa, J.; Urrutia, B.; Serradilla, J. M. Near Infrared Spectroscopy (NIRS) for the Determination of the Milk Fat Fatty Acid Profile of Goats. *Food Chem.* **2016**, *190*, 244–252. <https://doi.org/10.1016/j.foodchem.2015.05.083>.
- (246) Segtnan, V. H.; Isaksson, T. Evaluating near Infrared Techniques for Quantitative Analysis of Carbohydrates in Fruit Juice Model Systems. *J. Near Infrared Spectrosc.* **2000**, *8* (2), 109–116. <https://doi.org/10.1255/jnirs.270>.
- (247) Acharya, U. K.; Subedi, P. P.; Walsh, K. B. Evaluation of a Dry Extract System Involving NIR Spectroscopy (DESIR) for Rapid Assessment of Pesticide Contamination of Fruit Surfaces. *Am. J. Anal. Chem.* **2012**, *03* (08), 524–533. <https://doi.org/10.4236/ajac.2012.38070>.
- (248) Beseos, J. M. E.; Yaptenco, K. F.; Esguerra, E. B.; Peralta, E. K. Evaluation of Dry Extract System Involving NIR Spectroscopy (DESIR) for Pesticide Residues Detection on Fresh Carabao Mango (*Mangifera Indica* L. Cv ‘Carabao’) Fruit. *J. Adv. Agric. Technol.* **2018**, *5* (1), 24–30. <https://doi.org/10.18178/joaat.5.1.24-30>.
- (249) Steidle Neto, A. J.; Lopes, D. C.; Silva, W. A. Feasibility of Vis/Nir Spectroscopy to Detect and Estimate Fungicide Residues on Intact Lettuces. *Adv. Hortic. Sci.* **2020**, *34* (2), 223–232.

All references

- <https://doi.org/10.13128/ahsc-8125>.
- (250) Backman, V.; Gurjar, R.; Badizadegan, K.; Itzkan, I.; Dasari, R. R.; Perelman, L. T.; Feld, M. S. Polarized Light Scattering Spectroscopy for Quantitative Measurement of Epithelial Cellular Structures in Situ. *IEEE J. Sel. Top. Quantum Electron.* **1999**, *5* (4), 1019–1026. <https://doi.org/10.1109/2944.796325>.
- (251) Gobrecht, A.; Bendoula, R.; Roger, J. M.; Bellon-Maurel, V. Combining Linear Polarization Spectroscopy and the Representative Layer Theory to Measure the Beer-Lambert Law Absorbance of Highly Scattering Materials. *Anal. Chim. Acta* **2015**, *853* (1), 486–494. <https://doi.org/10.1016/j.aca.2014.10.014>.
- (252) Bendoula, R.; Gobrecht, A.; Moulin, B.; Roger, J. M.; Bellon-Maurel, V. Improvement of the Chemical Content Prediction of a Model Powder System by Reducing Multiple Scattering Using Polarized Light Spectroscopy. *Appl. Spectrosc.* **2015**, *69* (1), 95–102. <https://doi.org/10.1366/14-07539>.
- (253) Xu, J. L.; Gobrecht, A.; Gorretta, N.; Héran, D.; Gowen, A. A.; Bendoula, R. Development of a Polarized Hyperspectral Imaging System for Investigation of Absorption and Scattering Properties. *J. Near Infrared Spectrosc.* **2019**, *27* (4), 314–329. <https://doi.org/10.1177/0967033519857732>.
- (254) Gobrecht, A.; Bendoula, R.; Roger, J. M.; Bellon-Maurel, V. A New Optical Method Coupling Light Polarization and Vis-NIR Spectroscopy to Improve the Measurement of Soil Carbon Content. *Soil Tillage Res.* **2016**, *155*, 461–470. <https://doi.org/10.1016/j.still.2015.06.003>.
- (255) Alayed, M.; Deen, M. J. Time-Resolved Diffuse Optical Spectroscopy and Imaging Using Solid-State Detectors: Characteristics, Present Status, and Research Challenges. *Sensors (Switzerland)* **2017**, *17* (9), 2115. <https://doi.org/10.3390/s17092115>.
- (256) Johansson, J.; Folestad, S.; Josefson, M.; Sparen, A.; Abrahamsson, C.; Andersson-Engels, S.; Svanberg, S. Time-Resolved NIR/Vis Spectroscopy for Analysis of Solids: Pharmaceutical Tablets. *Appl. Spectrosc.* **2002**, *56* (6), 725–731. <https://doi.org/10.1366/000370202760077676>.
- (257) Hemrattrakun, P.; Nakano, K.; Boonyakiat, D.; Ohashi, S.; Maniwaru, P.; Theanjumol, P.; Seehanam, P. Comparison of Reflectance and Interactance Modes of Visible and Near-Infrared Spectroscopy for Predicting Persimmon Fruit Quality. *Food Anal. Methods* **2021**, *14* (1), 117–126. <https://doi.org/10.1007/s12161-020-01853-w>.
- (258) Khodabakhshian, R.; Emadi, B.; Khojastehpour, M.; Golzarian, M. R. A Comparative Study of Reflectance and Transmittance Modes of Vis/NIR Spectroscopy Used in Determining Internal Quality Attributes in Pomegranate Fruits. *J. Food Meas. Charact.* **2019**, *13* (4), 3130–3139. <https://doi.org/10.1007/s11694-019-00235-z>.
- (259) Chauchard, F.; Roger, J. M.; Bellon-Maurel, V. Correction of the Temperature Effect on near Infrared Calibration - Application to Soluble Solid Content Prediction. *J. Near Infrared Spectrosc.* **2004**, *12* (3), 199–205. <https://doi.org/10.1255/jnirs.427>.
- (260) Tillmann, P.; Paul, C. The Repeatability File - A Tool for Reducing the Sensitivity of near Infrared Spectroscopy Calibrations to Moisture Variation. *J. Near Infrared Spectrosc.* **1998**, *6* (1–4), 61–68.
- (261) Golic, M.; Walsh, K. B. Robustness of Calibration Models Based on near Infrared Spectroscopy for the In-Line Grading of Stonefruit for Total Soluble Solids Content. *Anal. Chim. Acta* **2006**, *555* (2), 286–291. <https://doi.org/10.1016/j.aca.2005.09.014>.
- (262) Andries, E.; Kalivas, J. H.; Gurung, A. Sample and Feature Augmentation Strategies for Calibration Updating. *J. Chemom.* **2019**, *33* (1), 1–20. <https://doi.org/10.1002/cem.3080>.
- (263) Campos, M. I.; Antolin, G.; Debán, L.; Pardo, R. Assessing the Influence of Temperature on NIRS Prediction Models for the Determination of Sodium Content in Dry-Cured Ham Slices. *Food Chem.* **2018**, *257* (February), 237–242. <https://doi.org/10.1016/j.foodchem.2018.02.131>.
- (264) Wijewardane, N. K.; Ge, Y.; Morgan, C. L. S. Prediction of Soil Organic and Inorganic Carbon at Different Moisture Contents with Dry Ground VNIR: A Comparative Study of Different Approaches. *Eur. J. Soil Sci.* **2016**, *67* (5), 605–615. <https://doi.org/10.1111/ejss.12362>.
- (265) Nocita, M.; Stevens, A.; Noon, C.; Van Wesemael, B. Prediction of Soil Organic Carbon for Different Levels of Soil Moisture Using Vis-NIR Spectroscopy. *Geoderma* **2013**, *199*, 37–42. <https://doi.org/10.1016/j.geoderma.2012.07.020>.
- (266) An, X.; Li, M.; Zheng, L.; Sun, H. Eliminating the Interference of Soil Moisture and Particle Size on Predicting Soil Total Nitrogen Content Using a NIRS-Based Portable Detector. *Comput. Electron. Agric.* **2015**, *112*, 47–53. <https://doi.org/10.1016/j.compag.2014.11.003>.
- (267) Hong, Y.; Yu, L.; Chen, Y.; Liu, Y.; Liu, Y.; Liu, Y.; Cheng, H. Prediction of Soil Organic Matter by VIS-

All references

- NIR Spectroscopy Using Normalized Soil Moisture Index as a Proxy of Soil Moisture. *Remote Sens.* **2018**, *10* (1), 1–17. <https://doi.org/10.3390/rs10010028>.
- (268) Hans, G.; Allison, B.; Bruce, A. Temperature and Moisture Insensitive Prediction of Biomass Calorific Value from Near-Infrared Spectra Using External Parameter Orthogonalization. *J. Near Infrared Spectrosc.* **2019**, 3–23. <https://doi.org/10.1177/0967033519840742>.
- (269) de Santana, F. B.; de Giuseppe, L. O.; de Souza, A. M.; Poppi, R. J. Removing the Moisture Effect in Soil Organic Matter Determination Using NIR Spectroscopy and PLSR with External Parameter Orthogonalization. *Microchem. J.* **2019**, *145* (October 2018), 1094–1101. <https://doi.org/10.1016/j.microc.2018.12.027>.
- (270) Wijewardane, N. K.; Ge, Y.; Morgan, C. L. S. Moisture Insensitive Prediction of Soil Properties from VNIR Reflectance Spectra Based on External Parameter Orthogonalization. *Geoderma* **2016**, *267* (April), 92–101. <https://doi.org/10.1016/j.geoderma.2015.12.014>.
- (271) Ackerson, J. P.; Morgan, C. L. S.; Ge, Y. Penetrometer-Mounted VisNIR Spectroscopy: Application of EPO-PLS to in Situ VisNIR Spectra. *Geoderma* **2017**, *286*, 131–138. <https://doi.org/10.1016/j.geoderma.2016.10.018>.
- (272) Ackerson, J. P.; Demattê, J. A. M.; Morgan, C. L. S. Predicting Clay Content on Field-Moist Intact Tropical Soils Using a Dried, Ground VisNIR Library with External Parameter Orthogonalization. *Geoderma* **2015**, *259–260*, 196–204. <https://doi.org/10.1016/j.geoderma.2015.06.002>.
- (273) Roudier, P.; Hedley, C. B.; Lobsey, C. R.; Viscarra Rossel, R. A.; Leroux, C. Evaluation of Two Methods to Eliminate the Effect of Water from Soil Vis–NIR Spectra for Predictions of Organic Carbon. *Geoderma* **2017**, *296*, 98–107. <https://doi.org/10.1016/j.geoderma.2017.02.014>.
- (274) Preys, S.; Roger, J. M.; Boulet, J. C. Robust Calibration Using Orthogonal Projection and Experimental Design. Application to the Correction of the Light Scattering Effect on Turbid NIR Spectra. *Chemom. Intell. Lab. Syst.* **2008**, *91* (1), 28–33. <https://doi.org/10.1016/j.chemolab.2007.10.007>.
- (275) Shenk, J. S.; Westerhaus, M. O. Optical Instrument Calibration System - US Patent US4866644A, 1991.
- (276) Ji, W.; Viscarra Rossel, R. A.; Shi, Z. Accounting for the Effects of Water and the Environment on Proximally Sensed Vis-NIR Soil Spectra and Their Calibrations. *Eur. J. Soil Sci.* **2015**, *66* (3), 555–565. <https://doi.org/10.1111/ejss.12239>.
- (277) Ji, W.; Viscarra Rossel, R. A.; Shi, Z. Improved Estimates of Organic Carbon Using Proximally Sensed Vis-NIR Spectra Corrected by Piecewise Direct Standardization. *Eur. J. Soil Sci.* **2015**, *66* (4), 670–678. <https://doi.org/10.1111/ejss.12271>.
- (278) Wang, Y.; Veltkamp, D. J.; Kowalski, B. R. Multivariate Instrument Standardization. *Anal. Chem.* **1991**, *63* (23), 2750–2756. <https://doi.org/10.1021/ac00023a016>.
- (279) OGEN, Y.; FAIGENBAUM-GOLOVIN, S.; GRANOT, A.; SHKOLNISKY, Y.; GOLDSHLEGER, N.; BEN-DOR, E. Removing Moisture Effect on Soil Reflectance Properties: A Case Study of Clay Content Prediction. *Pedosphere* **2019**, *29* (4), 421–431. [https://doi.org/10.1016/S1002-0160\(19\)60811-8](https://doi.org/10.1016/S1002-0160(19)60811-8).
- (280) Andries, E.; Kalivas, J. H. Interrelationships between Generalized Tikhonov Regularization, Generalized Net Analyte Signal, and Generalized Least Squares for Desensitizing a Multivariate Calibration to Interferences. *J. Chemom.* **2013**, *27* (5), 126–140. <https://doi.org/10.1002/cem.2501>.
- (281) Hansen, P. W. Pre-Processing Method Minimizing the Need for Reference Analyses. *J. Chemom.* **2001**, *15* (2), 123–131. [https://doi.org/10.1002/1099-128X\(200102\)15:2<123::AID-CEM636>3.0.CO;2-8](https://doi.org/10.1002/1099-128X(200102)15:2<123::AID-CEM636>3.0.CO;2-8).
- (282) Andersson, C. A. Direct Orthogonalization. *Chemom. Intell. Lab. Syst.* **1999**, *47* (1), 51–63. [https://doi.org/10.1016/S0169-7439\(98\)00158-0](https://doi.org/10.1016/S0169-7439(98)00158-0).
- (283) Zhu, Y.; Fearn, T.; Samuel, D.; Dhar, A.; Hameed, O.; Bown, S.; Lovat, L. *Error Removal by Orthogonal Subtraction (EROS): A Customised Pre-Treatment for Spectroscopic Data*; 2008; Vol. 22. <https://doi.org/10.1002/cem.1117>.
- (284) Marbach, R. A New Method for Multivariate Calibration. *J. Near Infrared Spectrosc.* **2005**, *13* (5), 241–254. <https://doi.org/10.1255/jnirs.555>.
- (285) Westerhaus, M. O. Improving Repeatability of NIR Calibrations across Instruments. Agricultural Research Centre Publishing, Gembloux (Belgium) 1991.
- (286) Baker, C. W.; Givens, D. I.; Deaville, E. R. Prediction of Organic Matter Digestibility in Vivo of Grass Silage by near Infrared Reflectance Spectroscopy: Effect of Calibration Method, Residual Moisture and Particle Size. *Anim. Feed Sci. Technol.* **1994**, *50* (1–2), 17–26. <https://doi.org/10.1016/0377->

All references

- 8401(94)90006-X.
- (287) Pérez-Marín, D.; Garrido-Varo, A.; De Pedro, E.; Guerrero-Ginel, J. E. Chemometric Utilities to Achieve Robustness in Liquid NIRS Calibrations: Application to Pig Fat Analysis. *Chemom. Intell. Lab. Syst.* **2007**, *87* (2), 241–246. <https://doi.org/10.1016/j.chemolab.2007.02.004>.
- (288) DiFoggio, R. Desensitizing Models Using Covariance Matrix Transforms or Counter-Balanced Distortions. *J. Chemom.* **2005**, *19* (4), 203–215. <https://doi.org/10.1002/cem.925>.
- (289) Jiang, Q.; Chen, Y.; Guo, L.; Fei, T.; Qi, K. Estimating Soil Organic Carbon of Cropland Soil at Different Levels of Soil Moisture Using VIS-NIR Spectroscopy. *Remote Sens.* **2016**, *8* (9). <https://doi.org/10.3390/rs8090755>.
- (290) Igne, B.; Allen, R.; Gahkani, A.; Kojic, D.; Kovacs, Z.; McCormick, R.; Mohan, S. Summary of the 2018 International Diffuse Reflectance Conference (IDRC) Software Shoot-Out. *NIR news* **2019**, *30* (1), 6–11. <https://doi.org/10.1177/0960336018814094>.
- (291) Huang, G.; Chen, X.; Li, L.; Chen, X.; Yuan, L.; Shi, W. Domain Adaptive Partial Least Squares Regression. *Chemom. Intell. Lab. Syst.* **2020**, *201* (March), 103986. <https://doi.org/10.1016/j.chemolab.2020.103986>.
- (292) Mishra, P.; Nikzad-Langerodi, R. Partial Least Square Regression versus Domain Invariant Partial Least Square Regression with Application to Near-Infrared Spectroscopy of Fresh Fruit. *Infrared Phys. Technol.* **2020**, *111*, 103547. <https://doi.org/10.1016/j.infrared.2020.103547>.
- (293) Mishra, P.; Roger, J. M.; Rutledge, D. N.; Woltering, E. Two Standard-Free Approaches to Correct for External Influences on near-Infrared Spectra to Make Models Widely Applicable. *Postharvest Biol. Technol.* **2020**, *170*, 111326. <https://doi.org/10.1016/j.postharvbio.2020.111326>.
- (294) Franceschini, M. H. D.; Demattê, J. A. M.; Kooistra, L.; Bartholomeus, H.; Rizzo, R.; Fongaro, C. T.; Molin, J. P. Effects of External Factors on Soil Reflectance Measured On-the-Go and Assessment of Potential Spectral Correction through Orthogonalisation and Standardisation Procedures. *Soil Tillage Res.* **2018**, *177* (August 2017), 19–36. <https://doi.org/10.1016/j.still.2017.10.004>.
- (295) Ji, W.; Viscarra Rossel, R. A.; Shi, Z. Accounting for the Effects of Water and the Environment on Proximally Sensed Vis-NIR Soil Spectra and Their Calibrations. *Eur. J. Soil Sci.* **2015**, *66* (3), 555–565. <https://doi.org/10.1111/ejss.12239>.
- (296) Roudier, P.; Hedley, C. B.; Lobsey, C. R.; Rossel, R. A. V.; Leroux, C. Evaluation of Two Methods to Eliminate the Effect of Water from Soil Vis – NIR Spectra for Predictions of Organic Carbon. *Geoderma* **2017**, *296*, 98–107. <https://doi.org/10.1016/j.geoderma.2017.02.014>.
- (297) Sun, X.; Subedi, P.; Walsh, K. B. Achieving Robustness to Temperature Change of a NIRS-PLSR Model for Intact Mango Fruit Dry Matter Content. *Postharvest Biol. Technol.* **2020**, *162* (January), 111117. <https://doi.org/10.1016/j.postharvbio.2019.111117>.
- (298) Acharya, U. K.; Walsh, K. B.; Subedi, P. P. Robustness of Partial Least-Squares Modelsto Change in Sample Temperature: II. Application to Fruit Attributes. *J. Near Infrared Spectrosc.* **2014**, *22* (4), 287–295. <https://doi.org/10.1255/jnirs.1119>.
- (299) Knadel, M.; Deng, F.; Alinejadian, A.; Wollesen de Jonge, L.; Moldrup, P.; Greve, M. H. The Effects of Moisture Conditions—From Wet to Hyper Dry—On Visible Near-Infrared Spectra of Danish Reference Soils. *Soil Sci. Soc. Am. J.* **2014**, *78* (2), 422. <https://doi.org/10.2136/sssaj2012.0401>.
- (300) Rinnert, E.; Carteret, C.; Humbert, B.; Fragneto-Cusani, G.; Ramsay, J. D. F.; Delville, A.; Robert, J. L.; Bihannic, I.; Pelletier, M.; Michot, L. J. Hydration of a Synthetic Clay with Tetrahedral Charges: A Multidisciplinary Experimental and Numerical Study. *J. Phys. Chem. B* **2005**, *109* (49), 23745–23759. <https://doi.org/10.1021/jp050957u>.
- (301) Dehnad, D.; Jafari, S. M.; Afrasiabi, M. Influence of Drying on Functional Properties of Food Biopolymers: From Traditional to Novel Dehydration Techniques. *Trends Food Sci. Technol.* **2016**, *57* (September 2016), 116–131. <https://doi.org/10.1016/j.tifs.2016.09.002>.
- (302) Dvořák, L.; Fajman, M.; Sustova, K. Influence of Sample Temperature for Measurement Accuracy with FT-NIR Spectroscopy. *J. AOAC Int.* **2017**, *100* (2), 499–502. <https://doi.org/10.5740/jaoacint.16-0264>.
- (303) Cozzolino, D.; Liu, L.; Cynkar, W. U.; Damberg, R. G.; Janik, L.; Colby, C. B.; Gishen, M. Effect of Temperature Variation on the Visible and near Infrared Spectra of Wine and the Consequences on the Partial Least Square Calibrations Developed to Measure Chemical Composition. *Anal. Chim. Acta* **2007**, *588* (2), 224–230. <https://doi.org/10.1016/j.aca.2007.01.079>.

All references

- (304) Wülfert, F.; Kok, W. T.; Smilde, A. K. Influence of Temperature on Vibrational Spectra and Consequences for the Predictive Ability of Multivariate Models. *Anal. Chem.* **1998**, *70* (9), 1761–1767. <https://doi.org/10.1021/ac9709920>.
- (305) Cumming, J. B. Temperature Dependence of Light Absorption by Water. *Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip.* **2013**, *713*, 1–4. <https://doi.org/10.1016/j.nima.2013.02.024>.
- (306) Kou, L.; Labrie, D.; Chylek, P. Refractive Indices of Water and Ice in the 065- to 25-Mm Spectral Range. *Appl. Opt.* **1993**, *32* (19), 3531. <https://doi.org/10.1364/ao.32.003531>.
- (307) Cui, X.; Zhang, J.; Cai, W.; Shao, X. Selecting Temperature-Dependent Variables in near-Infrared Spectra for Aquaphotomics. *Chemom. Intell. Lab. Syst.* **2018**, *183* (October), 23–28. <https://doi.org/10.1016/j.chemolab.2018.10.006>.
- (308) Cui, X.; Liu, X.; Yu, X.; Cai, W.; Shao, X. Water Can Be a Probe for Sensing Glucose in Aqueous Solutions by Temperature Dependent near Infrared Spectra. *Anal. Chim. Acta* **2017**, *957*, 47–54. <https://doi.org/10.1016/j.aca.2017.01.004>.
- (309) Watanabe, A.; Morita, S.; Ozaki, Y. Study on Temperature-Dependent Changes in Hydrogen Bonds in Cellulose I β by Infrared Spectroscopy with Perturbation-Correlation Moving-Window Two-Dimensional Correlation Spectroscopy. *Biomacromolecules* **2006**, *7* (11), 3164–3170. <https://doi.org/10.1021/bm0603591>.
- (310) Al Makdessi, N.; Ecarnot, M.; Roumet, P.; Rabatel, G. A Spectral Correction Method for Multi-Scattering Effects in Close Range Hyperspectral Imagery of Vegetation Scenes: Application to Nitrogen Content Assessment in Wheat. *Precis. Agric.* **2019**, *20* (2), 237–259. <https://doi.org/10.1007/s11119-018-9613-2>.
- (311) Snee, R. D. Validation of Regression Models: Methods and Examples. *Technometrics* **1977**, *19* (4), 415–428. <https://doi.org/10.1080/00401706.1977.10489581>.
- (312) Rutledge, D. N.; Barros, A. S. Durbin-Watson Statistic as a Morphological Estimator of Information Content. *Anal. Chim. Acta* **2002**, *454* (2), 277–295. [https://doi.org/10.1016/S0003-2670\(01\)01555-0](https://doi.org/10.1016/S0003-2670(01)01555-0).
- (313) Polyanskiy, M. Refractive index database. <http://refractiveindex.info/> (accessed Jan 27, 2020).
- (314) Oyinloye, T. M.; Yoon, W. B. Effect of Freeze-Drying on Quality and Grinding Process of Food Produce: A Review. *Processes* **2020**, *8* (3), 1–23. <https://doi.org/10.3390/PR8030354>.
- (315) Gobrecht, A.; Roger, J. M.; Bellon-Maurel, V. *Major Issues of Diffuse Reflectance NIR Spectroscopy in the Specific Context of Soil Carbon Content Estimation. A Review.*; 2014; Vol. 123. <https://doi.org/10.1016/B978-0-12-420225-2.00004-2>.
- (316) Mallet, A.; Charnier, C.; Ladrille, É.; Bendoula, R.; Steyer, J. P.; Roger, J. M. Unveiling Non-Linear Water Effects in near Infrared Spectroscopy: A Study on Organic Wastes during Drying Using Chemometrics. *Waste Manag.* **2021**, *122*, 36–48. <https://doi.org/10.1016/j.wasman.2020.12.019>.
- (317) Shao, X.; Cui, X.; Wang, M.; Cai, W. High Order Derivative to Investigate the Complexity of the near Infrared Spectra of Aqueous Solutions. *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.* **2019**, *213*, 83–89. <https://doi.org/10.1016/j.saa.2019.01.059>.
- (318) Gavin, H. P. The Levenberg-Marquardt Algorithm For Nonlinear Least Squares Curve-Fitting Problems. In *Duke University*; 2019; pp 1–19.
- (319) Winsor, C. P. The Gompertz Curve as a Growth Curve. *Proc. Natl. Acad. Sci.* **1932**, *18* (1), 1–8. <https://doi.org/10.1073/pnas.18.1.1>.
- (320) de Juan, A.; Tauler, R. Multivariate Curve Resolution: 50 Years Addressing the Mixture Analysis Problem – A Review. *Anal. Chim. Acta* **2021**, *1145*, 59–78. <https://doi.org/10.1016/j.aca.2020.10.051>.
- (321) Hugelier, S.; Devos, O.; Ruckebusch, C. *A Smoothness Constraint in Multivariate Curve Resolution-Alternating Least Squares of Spectroscopy Data*, 1st ed.; Elsevier B.V., 2016; Vol. 30. <https://doi.org/10.1016/B978-0-444-63638-6.00014-0>.
- (322) de Juan, A.; Vander Heyden, Y.; Tauler, R.; Massart, D. L. Assessment of New Constraints Applied to the Alternating Least Squares Method. *Anal. Chim. Acta* **1997**, *346* (3), 307–318. [https://doi.org/10.1016/S0003-2670\(97\)90069-6](https://doi.org/10.1016/S0003-2670(97)90069-6).
- (323) Bro, R.; Sidiropoulos, N. D. Least Squares Algorithms under Unimodality and Non-Negativity Constraints. *J. Chemom.* **1998**, *12* (4), 223–247. [https://doi.org/10.1002/\(SICI\)1099-128X\(199807/08\)12:4<223::AID-CEM511>3.0.CO;2-2](https://doi.org/10.1002/(SICI)1099-128X(199807/08)12:4<223::AID-CEM511>3.0.CO;2-2).
- (324) Hugelier, S.; Piqueras, S.; Bedia, C.; de Juan, A.; Ruckebusch, C. Application of a Sparseness Constraint

All references

- in Multivariate Curve Resolution – Alternating Least Squares. *Anal. Chim. Acta* **2018**, *1000*, 100–108. <https://doi.org/10.1016/j.aca.2017.08.021>.
- (325) Azzouz, T.; Tauler, R. Application of Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) to the Quantitative Analysis of Pharmaceutical and Agricultural Samples. *Talanta* **2008**, *74* (5), 1201–1210. <https://doi.org/10.1016/j.talanta.2007.08.024>.
- (326) Amigo, J. M.; de Juan, A.; Coello, J.; Maspoch, S. A Mixed Hard- and Soft-Modelling Approach for the Quantitative Determination of Oxipurines and Uric Acid in Human Urine. *Anal. Chim. Acta* **2006**, *567* (2), 236–244. <https://doi.org/10.1016/j.aca.2006.03.041>.
- (327) Diewok, J.; De Juan, A.; Maeder, M.; Tauler, R.; Lendl, B. Application of a Combination of Hard and Soft Modeling for Equilibrium Systems to the Quantitative Analysis of PH-Modulated Mixture Samples. *Anal. Chem.* **2003**, *75* (3), 641–647. <https://doi.org/10.1021/ac026248j>.
- (328) Mas, S.; Bendoula, R.; Agoda-Tandjawa, G.; de Juan, A.; Roger, J. M. Study of Time-Dependent Structural Changes of Laponite Colloidal System by Means of near-Infrared Spectroscopy and Hybrid Hard- and Soft-Modelling Multivariate Curve Resolution-Alternating Least Squares. *Chemom. Intell. Lab. Syst.* **2015**, *142*, 285–292. <https://doi.org/10.1016/j.chemolab.2014.12.009>.
- (329) De Juan, A.; Maeder, M.; Martínez, M.; Tauler, R. Combining Hard- and Soft-Modelling to Solve Kinetic Problems. *Chemom. Intell. Lab. Syst.* **2000**, *54* (2), 123–141. [https://doi.org/10.1016/S0169-7439\(00\)00112-X](https://doi.org/10.1016/S0169-7439(00)00112-X).
- (330) Tauler, R.; Smilde, A.; Kowalski, B. Selectivity, Local Rank, Three-way Data Analysis and Ambiguity in Multivariate Curve Resolution. *J. Chemom.* **1995**, *9* (1), 31–58. <https://doi.org/10.1002/cem.1180090105>.
- (331) Tauler, R.; Maeder, M.; de Juan, A. Extended Multivariate Curve Resolution; 2009; pp 473–505.

