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

Alexandre Mallet, Cyrille Charnier, Eric Latrille, Ryad Bendoula,
Jean-Philippe Steyer, Jean-Michel Roger

▶ To cite this version:

Alexandre Mallet, Cyrille Charnier, Eric Latrille, Ryad Bendoula, Jean-Philippe Steyer, et al.. Unveiling non-linear water effects in near infrared spectroscopy: A study on organic wastes during drying using chemometrics. Waste Management, 2021, 122, pp.36-48. 10.1016/j.wasman.2020.12.019 . hal-03122828

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

Authors: 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]

a) INRAE, Univ Montellier, LBE, 102 Av des Etangs, Narbonne F-11100, France
b) INRAE, UMR ITAP, Montpellier University, Montpellier, France
c) BIOENTECH Company, F-11100 Narbonne, France
d) ChemHouse Research Group, Montpellier, France

Corresponding author : Alexandre MALLET, a.mallet10@gmail.com, +33 6 26 42 03 45

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.

Keywords

Near infrared spectroscopy; Chemometrics; Robustness; Water effects; Drying; Organic wastes;

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 bio-chemical 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).

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-dominant 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 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 48h of drying in a heat chamber at 105 °C.

2.2. Drying system

The drying system used (Figure 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 hours to 72 hours.

2.3. Near infrared spectroscopic acquisition system

During the drying process, a spectrum of the sample was acquired from below every 90 seconds 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 reference was scanned every 10 minutes. All spectra were transformed into pseudo-absorbance units using log transformation:

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

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 :

$$LOD_s(t) = M_d(t) - M_d(t = 0), \quad \text{(Eq. 2)}$$

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

As mentioned, after drying, dry matter content was measured classically (using 48h 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 20k 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 $mgO_2.gTS^{-1}$, $3.2*10^{-2}$ $g.gTS^{-1}$, $8.6*10^{-3}$ $g.gTS^{-1}$, 83 $mgO_2.gTS^{-1}$.

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

$$X_C = X - \frac{1}{n} J_n X, \qquad \text{(Eq. 4)}$$

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

$$X_C = TP^T + E, \qquad \text{(Eq. 5)}$$

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.

$$X_{C\_deflated[q]} = X_C - T_{q-1} {P_{q-1}}^T. \qquad \text{(Eq. 6)}$$

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

$$T = \frac{1}{n-1} {X_C}^T X_C, \qquad \text{(Eq. 7)}$$

be the full variance-covariance matrix,

$$B = \frac{1}{n-1} {X_C}^T C(C^T C)^{-1} C^T X_C, \qquad \text{(Eq. 8)}$$

be the between-substrates variance-covariance matrix,

$$W = T - B, \qquad \text{(Eq. 9)}$$

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:

$$\mathrm{DW} = \sum_{i=2}^{n} (r_i - r_{i-1})^2 \,/ \sum_{i=1}^{n} {r_i}^2, \qquad \text{(Eq. 10)}$$

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.

$$\mathrm{CorrelationSpectra} = \left[\frac{cov(X_1,\mathrm{D})}{\sigma_{X_1}\sigma_{\mathrm{D}}}, \frac{cov(X_2,\mathrm{D})}{\sigma_{X_2}\sigma_{\mathrm{D}}}, \dots, \frac{cov(X_p,\mathrm{D})}{\sigma_{X_p}\sigma_{\mathrm{D}}}\right] = \frac{{D_C}^T X_C}{diag({X_C}^T X_C)^{1/2} diag({D_C}^T D_C)^{1/2} I_p}. \qquad \text{(Eq. 11)}$$

3. Results & Discussion

3.1. Data overview

3.1.1. Biochemical characteristics

Figure 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 (Figure 2) clearly highlights two populations: one population with no or very low fat content levels (<0.2 g.gTS^{-1}) and another population with very high fat content levels (>0.7 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

Figure 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 (*syrup_1*, *ketchup_1*, *banana_2*, *orangepulp_1*) as well as simple experimental drying interruptions due mostly to electric failures (*banana_1*, *crustbread_1*, *sunflowermeal_1*, *grass_1*, *weeds_3*). Theses 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 *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 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 Figure 4. Let

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

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). Figure 4.1 and Figure 4.3 both reveal four apparent outliers: dairy fat sludge, orange pulp, brewery yeast, and sunflower meal with respective dry matter 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. Figure 4.2 shows a good degree of agreement between the measured dry matter and the estimated dry matter using the system. Figure 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 ± 2.30% (Figure 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 (Eq. 7)(Eq. 8)(Eq. 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 Figure 5.1, 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, as shown in Figure 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

Figure 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 Figure 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 on Figure 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. Surprisingly, 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 (Figure 7 and Figure 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 (Figure 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 Figure 7). This suggests that the first component relates to global light scattering differences observed between substrates.

In the first component's score plot (Figure 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 (Figure 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 ($n_{water} \simeq 1.33$, $n_{air} \simeq 1.0$, $n_{organic\ compounds} \geq 1.4$) (Polyanskiy, 2008). For most substrates, as drying occurs, water is replaced by air, which leads to increased refractive index differences. As index differences 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 ($n_{vegetable\ oil} \simeq 1.47$) (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 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 (Figure 7) match well with the first eigenvector of the within-substrate variance-covariance matrix (plotted as a red dotted line in Figure 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 (Figure 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 (Figure 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 (Figure 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 Figure 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 (Figure 7), all of them being related to bands present in organic matter (CH/CH_2/NH). For this component, all the substrates exhibit increasing scores along drying (in particular *sugar*, *steak* or *sour cream*) (Figure 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 (Figure 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 (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 (Figure 7) which relate to CH, CH_2 and CH_3 combination bands. Moreover, the fifth component's scores separate very clearly the substrates rich in lipids from the ones rich in simple carbohydrates (Figure 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/CH_2/CH_3 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) (Figure 7). Scores exhibit two groups (Figure 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 (Figure 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 (Figure 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 (Figure 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 (Figure 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 (Figure 8), particularly in the high moisture content range (60 to 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.

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 Sousbie for their technical support during the setting-up of the experiment; as well as the ChemHouse group for limitless discussions on chemometrics.

References

Acharya, U.K., Walsh, K.B., Subedi, P.P., 2014. Robustness of partial least-squares modelsto 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., Terrom, 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

Bogrekci, 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

Büning-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

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. 170, 244–255. https://doi.org/10.1097/01.ss.0000162289.40879.7b

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 Manag. 59, 140–148. https://doi.org/10.1016/j.wasman.2016.10.029

Cozzolino, D., Liu, L., Cynkar, W.U., Dambergs, 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/CCHEM.1998.75.3.386

Galvez-Sola, L., Moral, R., Perez-Murcia, M.D., Perez-Espinosa, A., Bustamante, M.A., Martinez-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/Artn 803\rDoi 10.1051/Forest:2008065

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., Alinejadian, 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. a, 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

Meer, F. van der, 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

Oliphant, 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 and Analytical Applications 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 Deoxynivalenol Calibration. Cereal Chem. J. 93, 631–637. https://doi.org/10.1094/CCHEM-04-16-0120-R

Polyanskiy, M., 2008. Refractive index database. [WWW Document]. RefractiveIndex.INFO. URL 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, 111449. https://doi.org/10.1016/j.molliq.2019.111449

Rinnan, Å., Berg, F. van den, Engelsen, S.B., 2009. Review of the most common pre-processing 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, 103975. 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

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, 111117. 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, Sercond 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

Zeaiter, M., Rutledge, D., 2009. Preprocessing Methods. Compr. Chemom. 3, 121–231. https://doi.org/10.1016/B978-044452701-1.00074-0

Figures

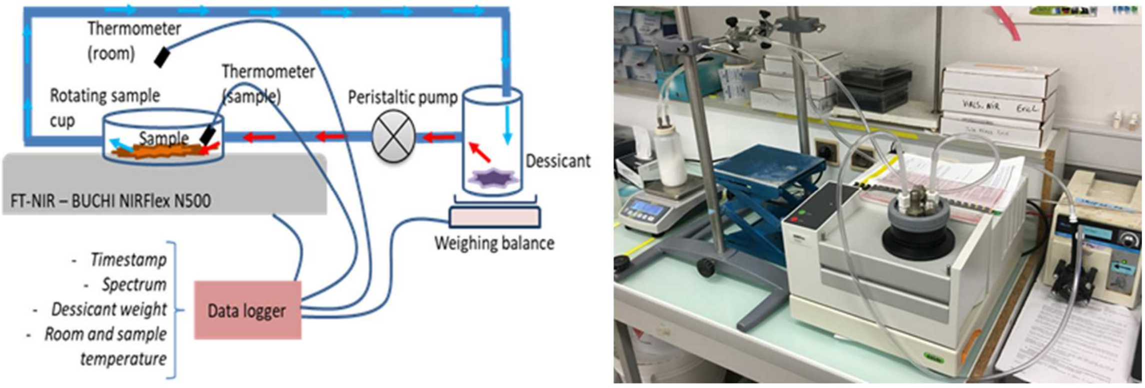


Figure 1 - Experimental set-up with: a NIRS acquisition under a quartz rotating sampling cup; a tube circuit with gas circulation, a desiccant weighed by a precision balance; the whole system is automatized and controlled by RS232 serial connection.

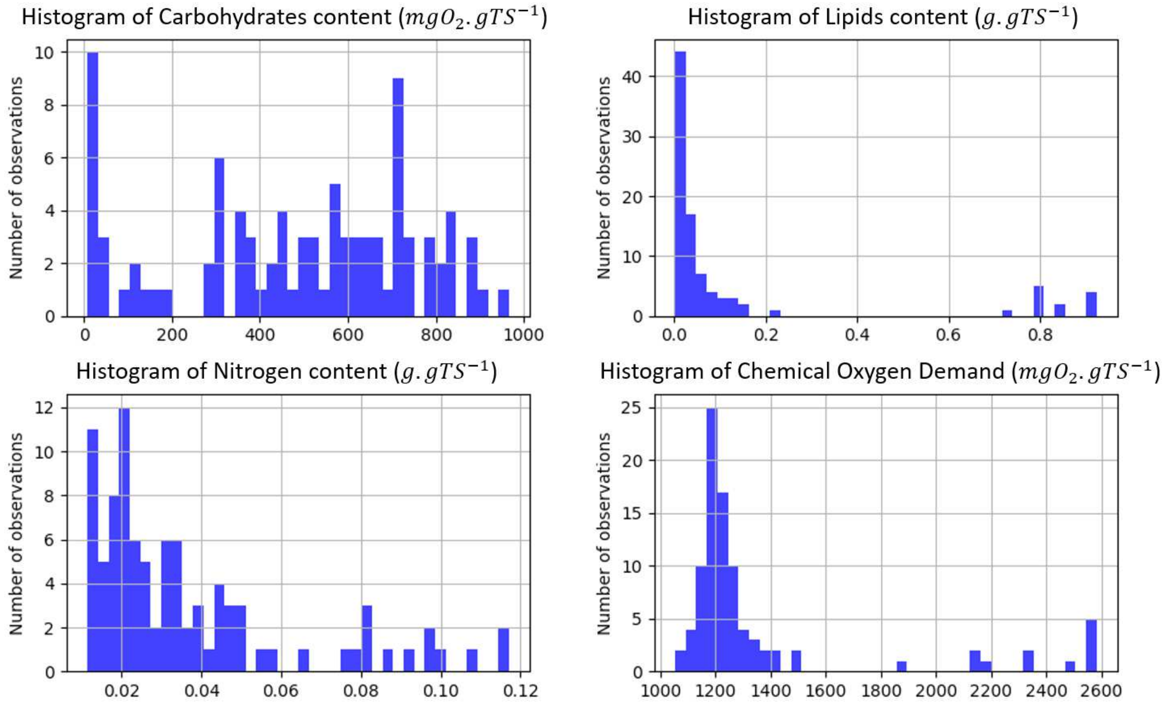


Figure 2 - Sample characteristics - Histograms of predicted characteristics: carbohydrate content ($mgO_2.gTS^{-1}$), fat content ($g.gTS^{-1}$), nitrogen content ($g.gTS^{-1}$), chemical oxygen demand ($mgO_2.gTS^{-1}$)

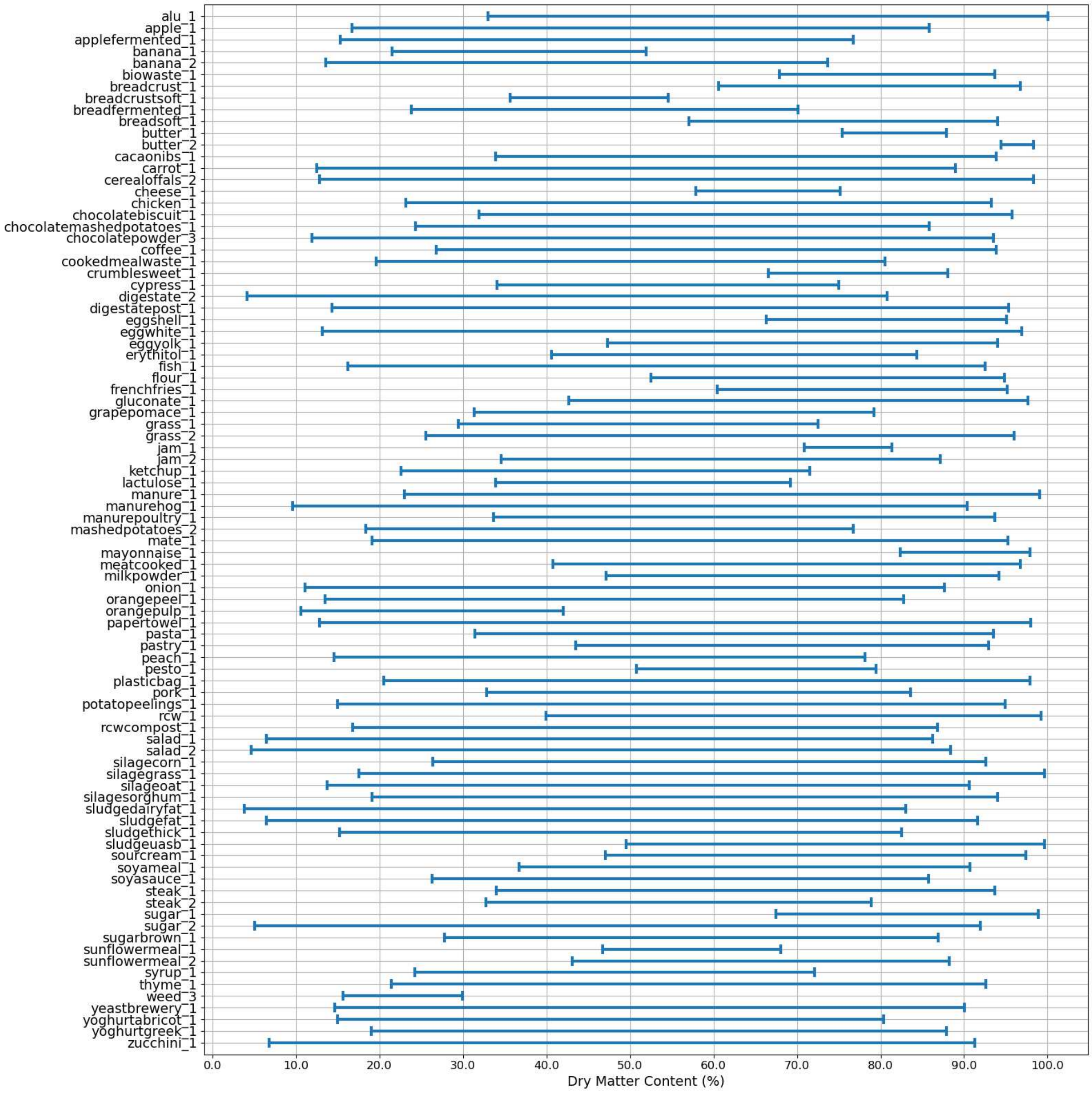


Figure 3 - Drying data: list of all samples with initial and final dry matter contents obtained in the experiment. Spectra were obtained within each of these ranges.

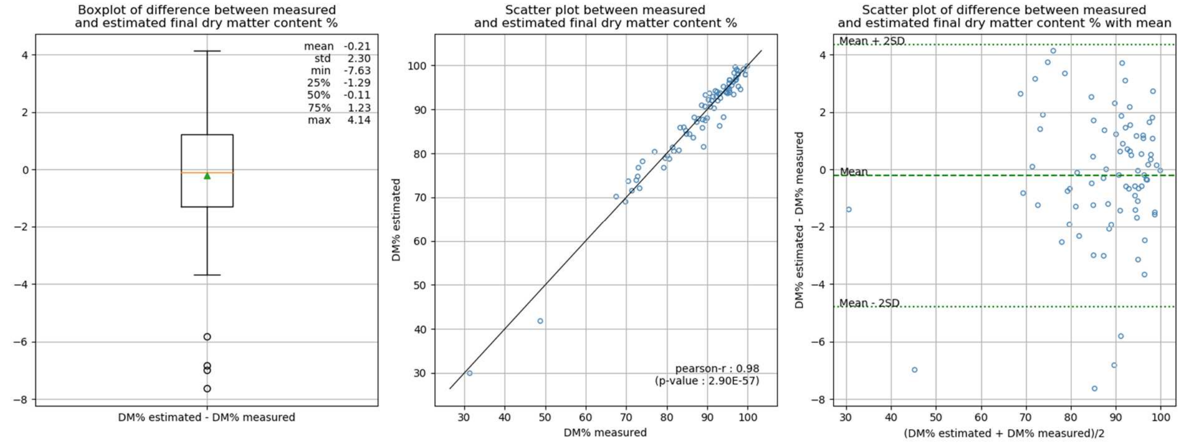


Figure 4 - Experimental conditions: (1) boxplot of final dry matter content errors; (2) measured vs. estimated final dry matter content, and (3) difference against mean

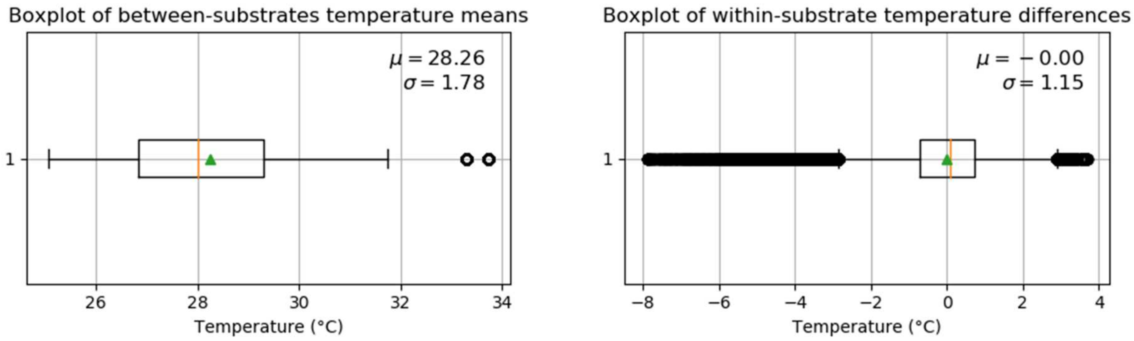


Figure 5 - Experimental conditions: boxplot of between-substrates temperature differences (1) and within-substrates temperature differences (2) during drying

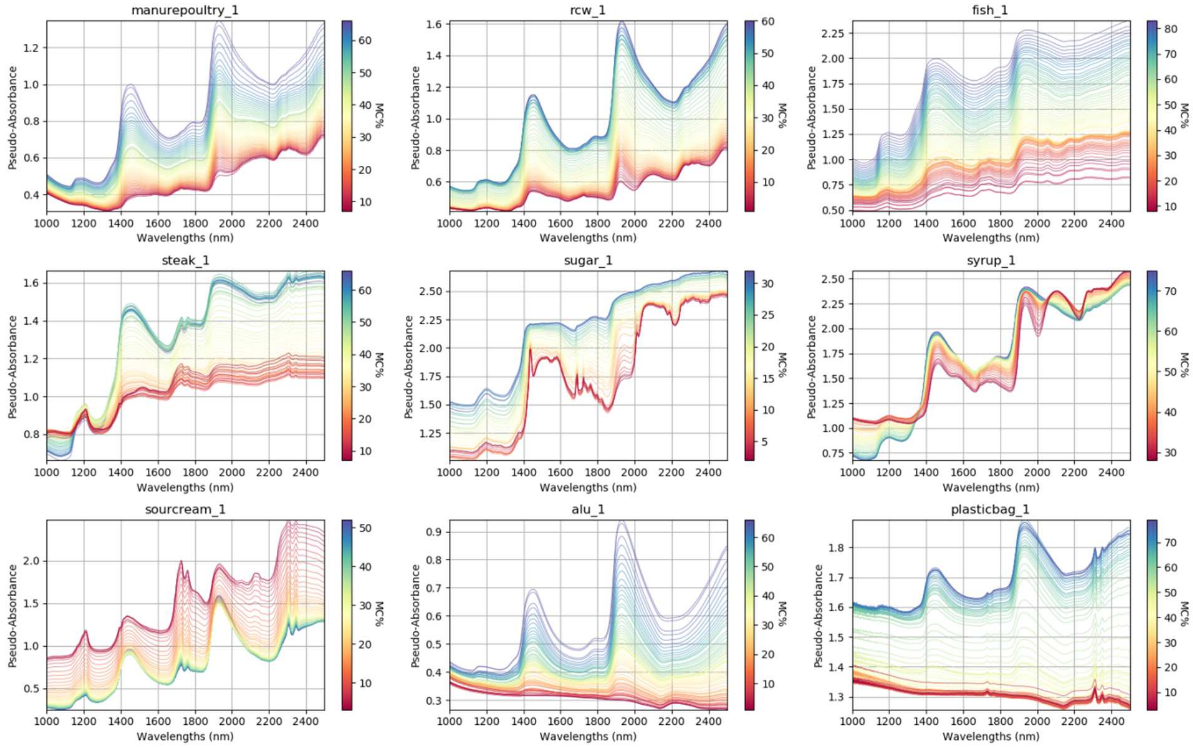


Figure 6 - Raw pseudo-absorbance spectra colored by moisture content (%) for nine substrates representative of the diversity of biochemical compositions and physical properties (poultry manure, ramial chipped wood / rcw, fish, cooked steak, sugar, syrup, sour cream, aluminum and plastic bag).

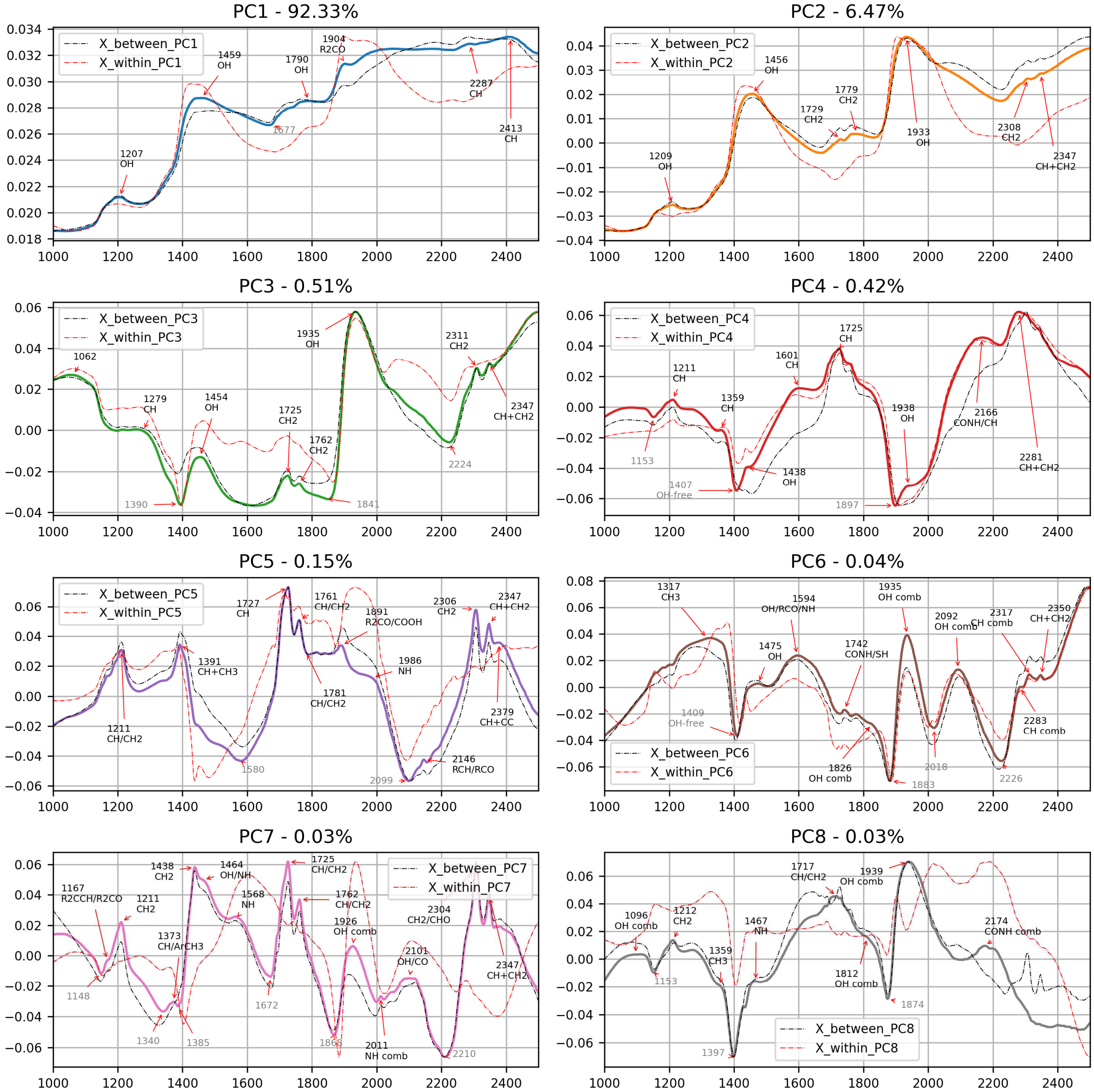


Figure 7 - Loadings from PCA of Xc (Eq. 4)(Eq. 5) with peak detection and chemical attributions (positive peaks annotated in black, and negative peaks in grey). Abscissa axis correspond to wavelengths (in nm). Explained variance percentage of each principal component is given in the title. For each component, 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).

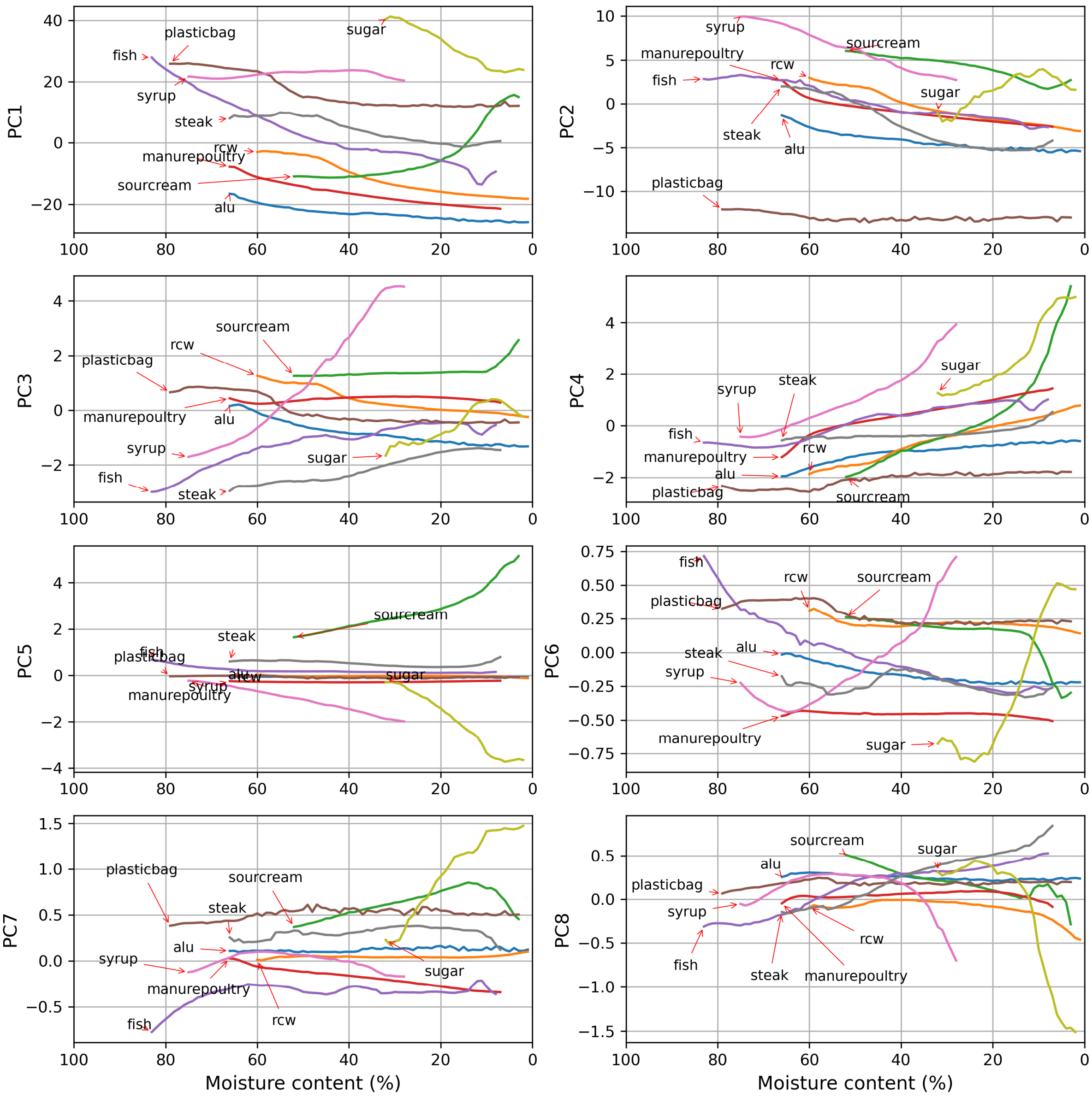
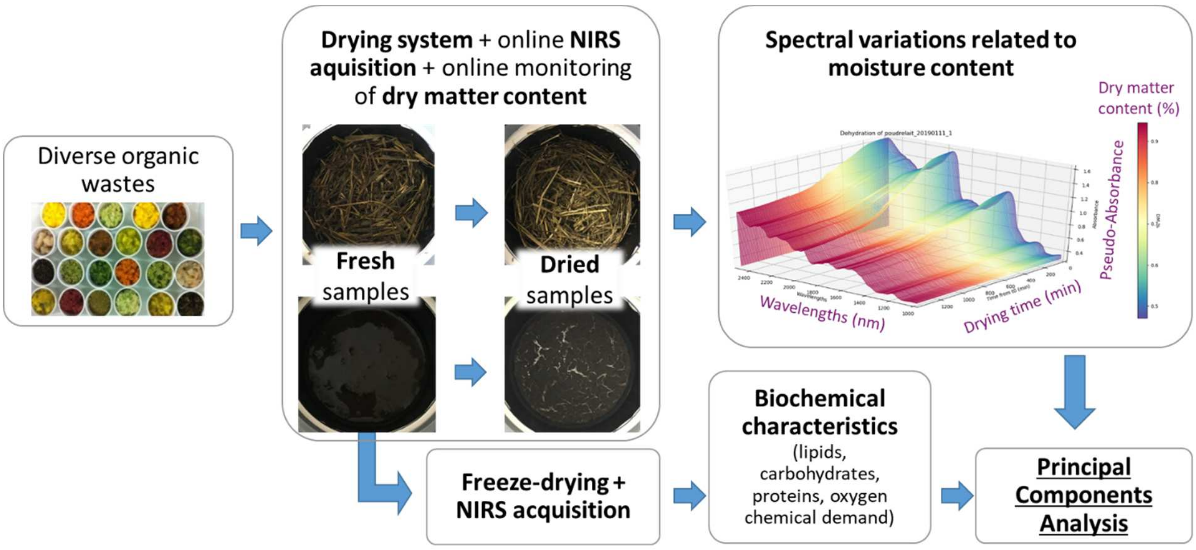


Figure 8 - Scores from PCA of Xc - All abscissa axes correspond to moisture content (%). Representative substrates were selected and plotted. See all other substrates scores in Appendix D.

Graphical abstract



Graphical Abstract - Summary of methodology: near infrared spectral variations related to moisture content variations are obtained for a variety of substrates, and application of principal components analysis is used to analyze the effects of water. The biochemical characteristics of substrates are obtained to investigate water effects' dependency to chemical types.