



HAL
open science

A new optical method coupling light polarization and Vis-NIR spectroscopy to improve the measurement of soil carbon content

Alexia Gobrecht, R. Bendoula, J.M. Roger, Véronique Bellon Maurel

► **To cite this version:**

Alexia Gobrecht, R. Bendoula, J.M. Roger, Véronique Bellon Maurel. A new optical method coupling light polarization and Vis-NIR spectroscopy to improve the measurement of soil carbon content. *Soil and Tillage Research*, 2016, 155, pp.461-470. <10.1016/j.still.2015.06.003>. <hal-01254732>

HAL Id: hal-01254732

<https://hal.science/hal-01254732v1>

Submitted on 12 Jan 2016

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.



HAL Authorization

1 A new optical method coupling light polarization and
2 Vis-NIR spectroscopy to improve the measurement of
3 soil carbon content

4 Alexia Gobrecht, Ryad Bendoula, Jean-Michel Roger, Véronique
5 Bellon-Maurel

6 *Irstea, UMR ITAP, 361 Rue Jean-François Breton, 34000 Montpellier, France*
7 corresponding author: alexia.gobrecht@irstea.fr ; + 33 4 67 16 64 60

8 **Abstract**

9 Visible and Near infrared (Vis-NIR) spectroscopy is now a common analyt-
10 ical method to measure different physical and chemical properties of soils,
11 including carbon content. However, prediction model accuracy is insufficient
12 for Vis-NIR spectroscopy to replace routine laboratory analysis. One of
13 the main issues this technique is facing up to is light scattering due to soil
14 particles. It causes departure in the assumed linear relationship between the
15 absorbance spectrum and the concentration of the chemical components as
16 stated by Beer-Lambert's law, which underpins the linear calibration models.

17 Therefore it is essential to improve the quality of the measured signal
18 in order to optimize the calibration results. Optics can help to mitigate
19 scattering effect on the signal.

20 The aims of this study were to test the feasibility of a new optical setup,
21 names PoLiS, coupling linearly polarized light with a Vis-VNIR (350 - 800
22 nm) spectrometer to free the measured spectra from multiscattering effect.
23 The measured signals were used to model the chemical absorbance of the soil
24 samples using Dahm's Equation in the frame of the Representative Layer
25 Theory (RLT).

26 The study was conducted using a set of 52 soil samples collected in France
27 (in the French calcareous Prealps) to predict soil Total Organic Carbon
28 (TOC) content. The PoLiS absorbance signal tended to be more linearly
29 related to the concentration of organic carbon, which is an important pre-
30 requisite to perform linear multivariate modeling. In a second step, the PLS
31 models achieved for TOC performed appreciably better than models based

32 on classical reflectance spectra. The standard errors of cross validation de-
33 creased from 20.8 $g.kg^{-1}$ to 17.6 $g.kg^{-1}$ and the coefficient of determination
34 R^2 improved from 0.82 to 0.87 on ground samples. To compare the added-
35 value of the PoLiS method we benchmarked the PoLiS prediction models
36 against models built from empirically preprocessed spectra. Again, the Po-
37 LiS method showed better performances.

38 This work confirmed that by optical means, it is possible to significantly
39 improve the quality of a spectroscopic signal and as a consequence, improve
40 also the quality of the linear model.

41 *Keywords:* Visible and near infrared spectroscopy, Light Polarization, Soil,
42 Total Organic Carbon, Representative Layer Theory

1. Introduction

With the goal of reducing the amount of greenhouse gases in the atmosphere, policy makers encourage practices intended to sequester carbon in soils (reforestation, changes in farming practices). New methods are required to rapidly and accurately measure soil Carbon at field- and landscape-scales. Although Visible and Near Infrared Spectroscopy (Vis-NIRS) is an analytical technology adapted to these specifications and becoming a very popular analytical technology in soil science, it is still steps away from being used as a routine analytical tool, both in field and laboratory. One of the reasons is that calibration models lack of robustness as soon as influence factors, which are numerous in soils, interfere. One of the main issues is that soils are highly scattering materials. As a direct consequence, the measurement conditions are far from the ideal conditions stated by Beer-Lambert's law where the absorbance should be linearly related to the chemical concentration (Gobrecht *et al.*, 2014). Light scattering depends on the physical structure of the soil samples and directly contributes to the shape of the measured spectrum by hiding (or overlapping) the chemically related information. The absorbance at wavelength λ is not linear with concentration and there is a real contradiction in building calibration models based on linear multivariate methods such as the commonly-used Partial Least Square Regression (PLS). Overcoming this signal quality issue is of great interest because the accuracy of the prediction is directly related to the quality of the measured signal (MacDougall & Crummett, 1980).

The most common strategy to reduce scattering effects is spectral pre-treatment. This preprocessing step is specifically designed to reduce multiplicative and additive effects caused by variations of physical properties (Rinnan *et al.*, 2009; Martens, 1991). Among them, standard normal variate (SNV) often associated to detrend (Barnes *et al.*, 1989), multiplicative signal correction (MSC) (Geladi *et al.*, 1985), Extended MSC (EMSC) (Martens, 1991), normalization or Optical Path Length Estimation and Correction (OPLEC) (Chen *et al.*, 2006; Jin *et al.*, 2012). However, these approaches remain questionable: they consider that scattering is nearly constant allover the wavelengths, which is not the case (Shi & Anderson, 2010); they may eliminate chemical-related information, which is very small with regard to scattering effects (Martens *et al.*, 2003); they are inappropriate when light scattering varies greatly from sample to sample (Steponavicius & Thennadil, 2011). As a consequence, the model may sometimes fail when applied on a

new set of samples.

Another option is to acquire the spectrum in a way that separates the part related to chemical absorption from the part related to scattering. Specific experimental techniques, based on the application of the light propagation theory or resolution of the Radiative Transfer Equation (Shi & Anderson, 2010) have been proposed: adding-doubling set-ups (Steponavicius & Thennadil, 2011; Prah, 1995; Steponavicius & Thennadil, 2009), spatially-resolved spectroscopy (Farrell *et al.*, 1992), time-resolved spectroscopy (Chauchard *et al.*, 2005; Abrahamsson *et al.*, 2005) and frequency-resolved spectroscopy (Martens, 1991). Although powerful, these methods have their limitations, particularly when applied on highly scattering samples. First, they may require complex and sometimes expensive optical implementations, which may not be compatible with conventional spectrometers or with highly scattering samples (for which transmission measurement is not possible). Secondly, as they rely on the estimation of absorption and scattering coefficients achieved by model inversion, parameters describing the studied medium (sample thickness, refractive index, particle size and shape...) must be known or approximated, which may be a troublesome task as they are often unknown in complex media (Steponavicius & Thennadil, 2011; Swartling *et al.*, 2003).

Bendoula *et al.* (2015) proposed to combine light polarization and VIS-NIR reflectance spectra acquisitions. The Polarized Light Spectroscopy (PoLiS) method is an original technique to reduce directly the effects of multiple scattering on the measured signal by using the wave theory of light (Lu *et al.*, 2006; Backman *et al.*, 1999). When linearly polarized light interacts with a scattering material, the backscattered light progressively loses its initial polarization and oscillates randomly in all the planes. Using the principle of polarization subtraction, Bendoula *et al.* (2015) measured a reflectance spectra being less impacted by multiple scattering. In Gobrecht *et al.* (2015), the signals measured with the PoLiS method were processed in the frame of Dahm's Representative Layer Theory (Dahm & Dahm, 2007) to propose a model of the absorbing power. The method was successfully tested on model particulate samples (sand + dye) showing that the newly computed absorbance signal is more linearly related to the concentration of dye in the sample.

The aim of this study is to test the PoLiS method on real soil samples to predict Total Organic Carbon (TOC) content in order to:

- validate that PoLiS absorbance measured on soil samples is more lin-

early related to TOC ;

- evaluate the benefit of using the PoLiS absorbance in TOC calibration models ;
- compare this “optical” preprocessing method to commonly used mathematical preprocessing methods.

2. Material and Methods

2.1. Soil samples

The 52 studied soil samples, provided by Irstea EMGR research unit are a subset of a soil sample collection used in a previous research work published in Saenger *et al.* (2013). The samples were collected in the Vercors High Plateau Natural Reserve (VHPNR) a protected mountainous area in the French calcareous Prealps (44°97N - 5°42E). Soils of the VHPNR developed on Urgonien limestones and are generally neutral or basic. They comprise humiferous and very shallow Cambisols, Leptosols, Umbirsols and Anthrosols (FAO/IUSS/ISRIC 200). Detailed information on vegetation and soil types of the study area are provided in Saenger *et al.* (2013). The samples were collected from the Topsoil (0-10 cm) from the A horizon (Organo-mineral layer). The litter layer, when present, was removed prior to sampling.

After collection, soil samples were air dried and stored at 4°C until chemical and spectral analysis. Total Organic Carbon was measured by dry combustion after decarbonation according to NF ISO 10694, using a N/C-Analyzer (Thermo Scientific, FLASH 2000 NC Analyzer, France) (AFNOR, 1995) (Table 1).

Each sample was prepared to get different particle sizes, namely:

- The *Coarse* form obtained by hand crushing the air-dried soil to get aggregates smaller than 5 mm. This preparation resulted in a large variety of particle and aggregate sizes within and between samples, depending on the type of soil;
- The *Sieved* form at 2 mm, which is the classical soil preparation prior to spectral acquisition;
- The finely *Ground* form at 0.25 mm.

Each sample was carefully transferred to an adapted 5-cm diameter petri dish and moved in circles to get an even and horizontal surface before spectral analysis.

Table 1 HERE

2.2. Instrumentation

The PoLiS optical setup, schematized in figure 1, was composed of a halogen light source (150 W, Leica Cls) coupled with a 940 μm core diameter optical fiber (Numerical Aperture $NA = 0.25$, Sedi & ATI). The light delivered by the fiber was collimated by an aspheric lens (F220SMA-B - Thorlabs). The incident beam was a 1.5 cm diameter circular spot with 1° divergence. The incident and reflected beam were polarized through two broad-band (400 nm–800 nm) polarizers (NT52-557, Edmunds Optics). Incident light was linearly polarized and reflected light was collected in a narrow cone (1°). The output from the analyzer was coupled to an optical fiber ($NA = 0.25$, Sedi & ATI) by an aspheric lens (F220SMA-B - Thorlabs). This fiber was connected to a spectrometer (MMS1, Zeiss). Spectral data were collected in the 400 – 800 nm wavelength range at 3 nm intervals, resulting in measurements at 121 discrete wavelengths per spectrum. A constant angle of 70° was maintained between the incident and reflecting arms. This angle was chosen to optimize intensity of the reflected beam.

FIGURE 1 HERE

2.3. PoLiS spectral acquisition

Each sample was illuminated with linearly polarized light and the remitted light intensity was measured with the *PoLiS* setup with the analyzer set respectively parallel, $I_{\parallel}(\lambda)$, and perpendicular, $I_{\perp}(\lambda)$, with respect to the polarization of the illumination light (Figure 2). Dark current, $I_b(\lambda)$, i.e. signal without light, was systematically recorded for all measured spectra with the same optical configuration and subtracted to each measurement. All measurements were made in a dark room to limit ambient light effects.

FIGURE 2 HERE

A diffuse reflectance gray standard (Spectralon®SRS60, Labsphere) was used to collect a reference spectrum, $I_0(\lambda)$, to standardize spectra from non-uniformities of all components of the instrumentation (light source, fibers, lens, polarizer and spectrometer). The reference measurement was collected every 10 to 15 samples.

From these measurements, the backscattered reflectance, $R_{BS}(\lambda)$, and the weakly scattered reflectance, $R_{SS}(\lambda)$, were computed for each sample according to Bendoula *et al.* (2015):

$$R_{BS}(\lambda) = \frac{[I_{\parallel}(\lambda) - I_{b\parallel}(\lambda)] + [I_{\perp}(\lambda) - I_{b\perp}(\lambda)]}{[I_0(\lambda) - I_{b0}(\lambda)]} \quad (1)$$

$$R_{SS}(\lambda) = \frac{[I_{\parallel}(\lambda) - I_{b\parallel}(\lambda)] - [I_{\perp}(\lambda) - I_{b\perp}(\lambda)]}{[I_0(\lambda) - I_{b0}(\lambda)]} \quad (2)$$

2.4. PoLiS absorbance Abs_{PO}

As proposed in Gobrecht *et al.* (2015), the PoLiS absorbance $Abs_{PO}(\lambda)$ was computed from the backscattered reflectance $R_{BS}(\lambda)$ and low scattered reflectance $R_{SS}(\lambda)$ as :

$$Abs_{PO}(\lambda) = -\log \left(R_{SS}(\lambda) + \sqrt{(1 - R_{SS}(\lambda))^2 - \frac{R_{SS}(\lambda)}{R_{BS}(\lambda)} (1 - R_{BS}(\lambda))^2} \right) \quad (3)$$

For comparison, the backscattered absorbance $Abs_{BS}(\lambda)$ was also computed from the total backscattered reflectance signal $R_{BS}(\lambda)$ measured with PoLiS.

$$Abs_{BS}(\lambda) = -\log R_{BS}(\lambda) \quad (4)$$

2.5. Multivariate Analysis

2.5.1. Principal Component Analysis

An exploratory analysis of the backscattered absorbance spectra $Abs_{BS}(\lambda)$ and the PoLiS absorbance spectra $Abs_{PO}(\lambda)$, was carried out using Principal Component Analysis (PCA). In order to evaluate the impact of the soil preparation on the spectra, the spectral data were centered in two different ways:

- Mean centered, meaning that the mean spectrum of the entire data set (global mean) is removed from all samples (Coarse, sieved and ground) to analyze the global variance of the dataset;

- Centered according to the location: the mean of the three spectra (one for each particle size preparation) measured for each sample collected at one location was subtracted. This centering allowed us to examine the variance within samples having the same TOC content but presenting different physical structure.

The Wilk's lambda criterion (Λ_w) was applied on the scores of the PCA. Wilk's Lambda is the ratio of the within class variance to the total variance (Roger *et al.*, 2005). Λ_w ranges from 0 to 1. For a low value, the classes are well separated and a value close to one indicates that the classes are confounded.

2.5.2. Calibration with Partial Least Square Regression

Calibration models were built using PLS (Wold *et al.*, 2001), considered as the *benchmark* chemometric technique used for quantitative analysis of diffuse reflectance spectra. The different types of signals computed, $R_{BS}(\lambda)$, $Abs_{BS}(\lambda)$ and $Abs_{PO}(\lambda)$, were compared on the basis of the performances of leave-one-out cross-validation models built on the each particle size sample set to predict soil Total Organic Content (TOC).

Preprocessing methods such as Standard Normal Variate (SNV), Multiplicative Scatter Correction (MSC) and modified Optical Pathlength Estimation and Correction (OPLECM) were applied to the different spectra.

Finally, the robustness of the best models obtained in cross-validation for each particle size class was tested by predicting the other particle size sets.

The performances of the cross-validations, and validations using other particle size sets as validation samples, were assessed through the number of latent variables used in the models, the coefficient of determination R^2 and the Standard Error of Cross-Validation (SECV) and Standard error of prediction (SEP) corrected from the bias (Bellon-Maurel *et al.*, 2010).

All the computations were performed with Matlab software (Matlab R2012b, Mathworks).

3. Results and discussion

3.1. Spectral analysis

The different mean-per-quartile spectra measured for samples having different particle sizes are plotted in figure 3. In the studied wavelength range (400 nm - 800 nm), spectral signatures are of good quality, with no noise,

even for $Abs_{PO}(\lambda)$, which have low signal intensities (about 10% of the total absorbance signal). In this wavelength range, soil spectra do not show characteristic spectral features and appear “flat”. Hence, the differences in the intensity level are related to the brightness of the samples. The reflectance intensity is consistent with the total organic carbon content (TOC) of the studied samples, showing that the darker the sample, the higher the TOC content. This observation concerns all particle sizes classes.

The wavelength range of the PoLiS setup is limited to 400 nm - 800 nm by the range of the polarizer used. This range is not the optimal Vis-NIR region for soil carbon calibration but Viscarra Rossel *et al.* (2008), for example, suggest that the visible portion of the spectrum contains more information on the absorbance characteristics of soil organic carbon than the shortwave NIR (700 – 1100 nm) content. In regard of the objectives of this study, this range is sufficient.

FIGURE 3 HERE

Sample preparation, i.e. particle size, has an impact on the intensity level of the backscattered reflectance $R_{BS}(\lambda)$. As commonly seen in NIR diffuse reflection (Pasikatan *et al.*, 2001), the smaller the particles, the higher the reflectance. As a consequence, the absorbance computed as $Abs_{BS}(\lambda) = -\log R_{BS}(\lambda)$ shows a lower level for ground samples. Therefore, the differences in the intensity levels are due to the combined effect of the physical structure and the brightness of the soil samples.

For the PoLiS absorbance $Abs_{PO}(\lambda)$, the intensity is about ten times smaller than for backscattered absorbance $Abs_{BS}(\lambda)$. This is partly due to the fact that the PoLiS optical set up selects only a small part of the signal (the single scattered one). The shape is also slightly different, with a small shoulder at 600 nm.

For coarse samples, the absorbance spectra $Abs_{BS}(\lambda)$ of the highly concentrated samples (quartiles Q3 and Q4) are not clearly separated, meaning that the variance due to particle size differences, and therefore scattering, dominates the chemically related information in the spectra. On the contrary, the PoLiS absorbance spectra $Abs_{PO}(\lambda)$ for quartiles Q3 and Q4 are clearly separated. This indicates that part of the spectral information due to the physical structure has been removed. Chemically related information, characterized by the brightness, becomes more visible.

Figure 4 shows the score plots of the PCA performed on differently centered spectral datasets ($Abs_{BS}(\lambda)$ and $Abs_{PO}(\lambda)$) according to section 2.5.

FIGURE 4 HERE

On the mean-centered data and for both $Abs_{BS}(\lambda)$ and $Abs_{PO}(\lambda)$, PC1 explains more than 98 % of the variability. Because of a multiplicative effect, the spectra appear to be organized in a conic pattern represented by the triangle in figure 4. The score plot converges both to a minima close to zero and spreads on the opposite side. This conic pattern represented on figure 4, is characteristic of a multiplicative effect caused by variables of influence which are a combination of soil brightness (related to TOC) and soil physical structure, i.e. the particle size.

For Abs_{BS} , finely ground samples are clearly separated from the two other particle size classes. For $Abs_{PO}(\lambda)$, this separation is less obvious. The summit of the cone contains the darker samples of different particle size classes. The multiplicative effect is due to TOC content as scattering is supposedly lessen.

The score plot of the data centered per sample location confirms the previous observation: for $Abs_{BS}(\lambda)$, the ground soils are clearly separated from the two other classes (sieved and coarse) as for $Abs_{PO}(\lambda)$, the classes appear more confounded. The values of the Wilk's lambda, computed on the scores of the PCA confirm these statements. When particle size classes are separated, the Wilk's lambda is lower.

The PoLiS method corrects, to a certain extent, the effect of scattering on the signal, leading to an absorbance less sensitive to the physical structure of the samples.

3.2. Linearity between Absorbance and TOC Concentration

The assumption that, by correcting the signal from part of the multiscattering effect, the PoLiS absorbance $Abs_{PO}(\lambda)$ is more linearly related to TOC content can be assessed through the Pearson's correlation coefficient between the absorbance and the TOC content. The correlograms presented in figure 5 show the correlation between the two absorbance signals ($Abs_{BS}(\lambda)$ and $Abs_{PO}(\lambda)$) and TOC as a function of the wavelength and for each sample preparation.

FIGURE 5 HERE

For coarse and sieved samples, the Pearson's coefficient R between the absorbance and the TOC concentration is always higher for $Abs_{PO}(\lambda)$ than for $Abs_{BS}(\lambda)$, over all the wavelength range. For ground samples, the two correlogram are similar, although slightly better for $Abs_{PO}(\lambda)$ between 400 and 600 nm. It is coherent with the general acceptance that preparing the samples (sieving or grinding) has a direct impact on the signal quality and

consequently on the quality of the calibration models (Morgan *et al.*, 2009; Bellon-Maurel *et al.*, 2010). Here, PoLiS method leads to an additional improvement of the correlation between the Absorbance signal and TOC.

Another way to visualize this observation is to plot TOC versus the absorbance value at the optimal wavelength of $Abs_{BS}(\lambda)$, respectively 450 nm for the coarse samples, 600 nm for the sieved samples and 570 nm for the ground samples (figure 6).

FIGURE 6 HERE

The degree of linearity between $Abs_{PO}(\lambda)$ and TOC is improved for coarse and sieved samples, but this effect is less for ground samples, for which the linear correlation coefficient for $Abs_{BS}(\lambda)$ and $Abs_{PO}(\lambda)$ are very similar and high (>0.87).

To conclude, this analysis shows that $Abs_{PO}(\lambda)$ is more linearly related to the TOC concentration (Figure 6) and additionally that the particle size has less impact on its spectral signature (Figure 4). Therefore, calibration conditions are more appropriate for $Abs_{PO}(\lambda)$ than for $Abs_{BS}(\lambda)$ to use linear methods like PLS in order to predict TOC in soils.

3.3. Model analysis

3.3.1. Quality of the calibration models

Figure 7 shows the quality of the models calibrated on the spectra obtained with the different methods : the backscattered reflectance spectra ($R_{BS}(\lambda)$), the backscattered absorbance spectra ($Abs_{BS}(\lambda)$) and the PoLiS absorbance spectra ($Abs_{PO}(\lambda)$), with no preprocessing, for each category of particle size.

FIGURE 7 HERE

First, the prediction models built with the backscattered reflectance $R_{BS}(\lambda)$ are not satisfying. They show a characteristic “banana” shaped regression curve, typical of non-linearity. However, ground and sieved samples produce better predictions than coarse samples. The latter present a high structural variability which affects the spectra. The scattering effect dominates in the spectral information but in a different manner for all the samples. This confirms the discussion of the previous section: sieving or grinding soils improves the PLS models.

The log–transformation of the backscattered reflectance $R_{BS}(\lambda)$ into backscattered absorbance, $\{Abs_{BS}(\lambda) = -\log R_{BS}(\lambda)\}$, improves the quality of the models. Theoretically, the linear relation is between absorbance and concentration and not between reflectance and the concentration. In our case, the

log also plays the role of a mathematical preprocessing method as it transforms multiplicative effects (due to scattering) into additive effects (Hadoux *et al.*, 2014). The PLS algorithm is capable to discard this additive effect in the regression process. R^2 and SECV are improved but need a high number of latent variables to build the models (10 for the ground samples and 8 for the sieved samples). According to the principle of parsimony, there is a risk that models will lack in robustness (Bellon-Maurel & McBratney, 2011; Seasholtz & Kowalski, 1993).

The models built with $Abs_{PO}(\lambda)$ outperform all the other models built with $R_{BS}(\lambda)$ and $Abs_{BS}(\lambda)$, whatever the particle size. R^2 and SECV are improved and, in addition, the number of latent variables decreases. However, soil sample preparation still impacts the results. PoLiS method also takes benefit from sample preparation (ground or sieved). For coarse samples, predictions are not so good, although improved compared to the predictions of the models built with the backscattered absorbance Abs_{BS} .

3.3.2. Comparison of optical and mathematical spectral preprocessing

The PoLiS method can be considered as an “optical preprocessing” method: prior to the calibration step, the different components of the total spectra are selected in order to compute an absorbance spectrum. The main objective of this optical preprocessing step is to enhance the quality of the signal by reducing the effect of multiscattering. We compared the calibration results using the PoLiS method with three mathematical preprocessing methods (SNV, MSC and modified OPLEC) usually applied on spectra to reduce the multiplicative and additive effects due to scattering.

Figure 8 present the R^2 and the SECV values for each models built.

FIGURE 8 HERE

The TOC prediction models built with the PoLiS absorbance spectra $Abs_{PO}(\lambda)$ always show better figures of merit than for the models built with $R_{BS}(\lambda)$ and $Abs_{BS}(\lambda)$, even when they are preprocessed.

The backscattered reflectance spectra $R_{BS}(\lambda)$ are highly impacted by light scattering. Hence, the preprocessing methods improve the performances of the prediction models, in particular for the sieved and ground samples. SNV and MSC have almost the same behavior on these spectral data, which is often stressed out by authors (Fearn *et al.*, 2009). Modified OPLEC gives good results and seems to be a promising preprocessing method as it specifically removes the multiplicative effect. For coarse samples however, none of the preprocessing methods applied did significantly increase the quality

parameters. These samples present a high sample-to-sample heterogeneity and as a consequence, different levels of light – matter interactions, which are more difficult to capture and correct by the different preprocessing method. Preprocessing the backscattered absorbance spectra Abs_{BS} does not significantly changes the quality of the models, although the number of latent variables decreases from 10 to 7.

For $Abs_{PO}(\lambda)$, none of the preprocessing methods have a positive impact on the figures of merit compared to the raw absorbance spectra. On the contrary, preprocessing the PoLiS absorbance $Abs_{PO}(\lambda)$ highly degrades the quality of the models. It is known that mathematical preprocessing methods suppresses part of the spectral information, sometimes not exclusively due to physical influence but which can also be related to chemical information.

As a conclusion, the PoLiS method produces an optimal absorbance signal, which does not need to be preprocessed prior calibration as the models built from $Abs_{PO}(\lambda)$ always outperform the other models, for all the particle sizes.

3.3.3. Behaviour of the PoLiS method regarding particle size

The main assumption made for the PoliS method is that it reduces the multiscattering effect on the absorbance spectra. Yet, multiscattering is dependent of the particle size of the sample. In section 3.1, the PCA analysis on the data concluded that $Abs_{PO}(\lambda)$ is less impacted by the preparation of the samples than $Abs_{BS}(\lambda)$, although, the ground samples still behave differently. Table 2 show the quality parameter (R^2 , bias and Standard Error of Prediction corrected from the bias (SEP_c) and slope) of the models built on one particle size class and applied to another particle size class.

TABLE 2 HERE

First, each time finely ground samples (< 0.25 mm) are involved, either in the calibration set or in the test set, PoLiS method do not produce better predictions. R^2 is lower with $Abs_{PO}(\lambda)$ than with $Abs_{BS}(\lambda)$ and the SEP_c , the bias and the slope are worse. We previously observed that for ground samples, $Abs_{BS}(\lambda)$ and $Abs_{PO}(\lambda)$ show a very similar correlogram, meaning that both absorbance signals show a relative linearity with TOC. Here, the PoLiS method seems to reach its limits when the particle size of the particulate samples are very small. Grinding finely the samples affects the way light travels in the samples and probably also the depolarization process. As a consequence, the backscattered reflectance $R_{BS}(\lambda)$ and the low scattered reflectance $R_{SS}(\lambda)$ used to compute the PoliS absorbance $Abs_{PO}(\lambda)$ (equation

3) are not completely reliable.

When particle sizes are larger than 2 mm, i.e. sieved or coarse, the models built with $Abs_{PO}(\lambda)$ always produce better results than with $Abs_{BS}(\lambda)$, as shown in figure 9.

FIGURE 9 HERE

Although the PoLiS calibration model built on coarse samples was the less performant in cross-validation (see figure 7), the prediction are not degraded when it is applied on the sieved samples. Moreover, the bias, which is a good indicator of robustness, remains small. On the other way, when the model built on sieved samples is applied on coarse samples, the figures of merit are not as good as in cross validation, but still, the results are much better with $Abs_{PO}(\lambda)$ than with $Abs_{BS}(\lambda)$. And again, the bias is very small for $Abs_{PO}(\lambda)$ compared to the high bias value for $Abs_{BS}(\lambda)$.

These results show that PoLiS is a promising measurement technique in the perspective of reducing the sample preparation as it is less sensitive to changes of the physical structure of the samples and well adapted to low processed samples.

4. Conclusions

For the first time, the issue of light scattering in Vis-NIR spectroscopy applied to soils has been studied from an optical point of view. In this study, PoLiS, an original optical setup based on light polarization spectroscopy, was used to select backscattered light being less impacted by multiscattering effects due to particles composing soil samples. The absorbance signal computed from the PoLiS measurements was compared to the absorbance traditionally computed by taking the log of the backscattered reflectance.

The aim of this study was to verify the assumptions underpinning the PoLiS method. We can make following statements and concluding remarks :

- On soil samples, the method produced spectral signatures of good quality, with no noise, despite the low intensity in the PoLiS wavelength range;
- Removing part of the multiscattering improved the degree of linearity between the PoLiS absorbance and the TOC, over all the wavelength range (400 - 800 nm) for coarse and sieved samples.

- TOC prediction models build with the PoLiS absorbance always outperformed the models built with the backscattered absorbance, even when mathematically preprocessed. This is an important result confirming that a signal of better quality improves the quality of the prediction models.
- The PoLiS absorbance is less impacted by a change of particle size of the samples but an effect is still visible, particularly for ground samples. As a consequence, the predictive potential of the PoLiS absorbance when only the physical structure of the sample changes is higher than the backscattered absorbance, when the particle size is > 2 mm. For finely ground samples, PoLiS seems to reach its limits.

This study confirms the high potential of the PoLiS method for the spectral analysis of soil properties. Solving the technical limits which would make the PoLiS method work beyond 800 nm, would allow to take an important step in the metrological quality of the soil carbon content measurement by NIRS.

5. Acknowledgments

The authors acknowledge the financial support of ADEME (French Environment and Energy Management Agency) within the INCA Project (In-field Spectroscopy for Carbon Accounting) of the GESSOL program.

Abrahamsson, C., Johansson, J., Andersson-Engels, S., Svanberg, S., & Folestad, S. 2005. Time-Resolved NIR Spectroscopy for Quantitative Analysis of Intact Pharmaceutical Tablets. *Analytical Chemistry*, **77**(4), 1055–1059.

AFNOR. 1995. *NF ISO 10694, Soil quality, Determination of organic and total carbon after dry combustion elementary analysis*.

Backman, V., Gurjar, R., Badizadegan, K., Itzkan, I., Dasari, R. R., Perelman, L. T., & Feld, M. 1999. Polarized light scattering spectroscopy for quantitative measurement of epithelial cellular structures in situ. *Selected Topics in Quantum Electronics, IEEE Journal of*, **5**(4), 1019–1026.

Barnes, R. J., Dhanoa, M. S., & Lister, Susan J. 1989. Standard Normal Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra. *Applied Spectroscopy*, **43**(5), 772–777.

- Bellon-Maurel, V., & McBratney, A.B. 2011. Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils - Critical review and research perspectives. *Soil Biology and Biochemistry*, **43**(7), 1398–1410.
- Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J.M., & McBratney, A.B. 2010. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. *TrAC Trends in Analytical Chemistry*, **29**(9), 1073–1081.
- 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. *Applied Spectroscopy*, **69**(1).
- Chauchard, F., Roger, J.M., Bellon-Maurel, V., Abrahamsson, C., Andersson-Engels, S., & Svanberg, S. 2005. MADSTRESS: A linear approach for evaluating scattering and absorption coefficients of samples measured using time-resolved spectroscopy in reflection. *Applied Spectroscopy*, **59**(10), 1229–1235.
- Chen, Z.-P., Morris, J., & Martin, E. 2006. Extracting Chemical Information from Spectral Data with Multiplicative Light Scattering Effects by Optical Path-Length Estimation and Correction. *Analytical Chemistry*, **78**(22), 7674–7681.
- Dahm, D.J., & Dahm, K.D. 2007. *Interpreting Diffuse Reflectance and Transmittance: A Theoretical Introduction to Absorption Spectroscopy of Scattering Materials*. 1 edn. NIR Publications.
- Farrell, T.J., Patterson, M.S., & Wilson, B. 1992. A diffusion theory model of spatially resolved, steady-state diffuse reflectance for the noninvasive determination of tissue optical properties in vivo. *Medical Physics*, **19**(4), 879–888.
- Fearn, T., Riccioli, C., Garrido-Varo, A., & Guerrero-Ginel, J. E. 2009. On the geometry of SNV and MSC. *Chemometrics and Intelligent Laboratory Systems*, **96**(1), 22–26.

- Geladi, P., MacDougall, D., & Martens, H. 1985. Linearization and Scatter-Correction for Near-Infrared Reflectance Spectra of Meat. *Applied Spectroscopy*, **39**(3), 491–500.
- Gobrecht, A., Roger, J.J., & Bellon-Maurel, V. 2014. Chapter Four - Major Issues of Diffuse Reflectance NIR Spectroscopy in the Specific Context of Soil Carbon Content Estimation: A Review. *Pages 145–175 of: Sparks, Donald L. (ed), Advances in Agronomy 123*, vol. 123. Academic Press.
- 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. *Analytica Chimica Acta*, **853**(0), 486 – 494.
- Hadoux, X., Gorretta, N., Roger, J.M., Bendoula, R., & Rabatel, G. 2014. Comparison of the efficacy of spectral pre-treatments for wheat and weed discrimination in outdoor conditions. *Computers and Electronics in Agriculture*, **108**, 242–249.
- Jin, J.W., Chen, Z.P., Li, L.M, Steponavicius, R., Thennadil, S. N., Yang, J., & Yu, R.Q. 2012. Quantitative Spectroscopic Analysis of Heterogeneous Mixtures: The Correction of Multiplicative Effects Caused by Variations in Physical Properties of Samples. *Analytical Chemistry*, **84**(1), 320–326.
- Lu, B., Morgan, S. P., Crowe, J. A., & Stockford, I. M. 2006. Comparison of Methods for Reducing the Effects of Scattering in Spectrophotometry. *Applied Spectroscopy*, **60**(10), 1157–1166.
- MacDougall, D., & Crummett, W. B. 1980. Guidelines for data acquisition and data quality evaluation in environmental chemistry. *Analytical Chemistry*, **52**(14), 2242–2249.
- Martens, H. and Stark, E. 1991. Extended multiplicative signal correction and spectral interference subtraction: New preprocessing methods for near infrared spectroscopy. *Journal of Pharmaceutical and Biomedical Analysis*, **9**(8), 625–635.
- 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. *Analytical Chemistry*, **75**(3), 394–404.

- Morgan, C.L. S., Waiser, T. H., Brown, D. J., & Hallmark, C. T. 2009. Simulated in situ characterization of soil organic and inorganic carbon with visible near-infrared diffuse reflectance spectroscopy. *Geoderma*, **151**(3-4), 249–256.
- Pasikatan, M. C., Steele, J. L., Spillman, C. K., & Haque, E. 2001. Near infrared reflectance spectroscopy for online particle size analysis of powders and ground materials. *Journal of Near Infrared Spectroscopy*, **9**(3), 153–164.
- Prahl, S. A. 1995. The adding-doubling method. *Pages 101–129 of: Optical-thermal response of laser-irradiated tissue*. Springer.
- Rinnan, A., van den Berg, F., & Engelsen, S. B. 2009. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry*, **28**(10), 1201–1222.
- 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. *Chemometrics and Intelligent Laboratory Systems*, **79**(1), 31–41.
- Saenger, A., Cécillon, L., Sebag, D., & Brun, J.J. 2013. Soil organic carbon quantity, chemistry and thermal stability in a mountainous landscape: A Rock-Eval pyrolysis survey. *Organic Geochemistry*, **54**(0), 101 – 114.
- Seasholtz, M. B., & Kowalski, B.R. 1993. The parsimony principle applied to multivariate calibration. *Analytica Chimica Acta*, **277**(2), 165–177.
- Shi, Z., & Anderson, C. A. 2010. Pharmaceutical applications of separation of absorption and scattering in near-infrared spectroscopy (NIRS). *Journal of Pharmaceutical Sciences*, **99**(12), 4766–4783.
- Steponavicius, R., & Thennadil, S. N. 2009. Extraction of Chemical Information of Suspensions Using Radiative Transfer Theory to Remove Multiple Scattering Effects: Application to a Model Two-Component System. *Analytical Chemistry*, **81**(18), 7713–7723.
- Steponavicius, R., & Thennadil, S. N. 2011. Extraction of Chemical Information of Suspensions Using Radiative Transfer Theory To Remove Multiple

- Scattering Effects: Application to a Model Multicomponent System. *Analytical Chemistry*, **83**(6), 1931–1937.
- Swartling, J., Dam, J.S., & Andersson-Engels, S. 2003. Comparison of spatially and temporally resolved diffuse-reflectance measurement systems for determination of biomedical optical properties. *Applied Optics*, **42**(22), 4612–4620.
- Viscarra Rossel, R.A., Fouad, Y., & Walter, C. 2008. Using a digital camera to measure soil organic carbon and iron contents. *Biosystems Engineering*, **100**(2), 149–159.
- Wold, S., Sjöström, M., & Eriksson, L. 2001. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, **58**(2), 109–130.

Table legend:

Table 1 : Total Organic Carbon ($g.kg^{-1}$) descriptive statistics for the whole dataset. Q1, Q2 and Q3 correspond respectively to the first quartile, the median and the upper quartile. SD: standard deviation.

Table 2 : Performance of the models built with $Abs_{BS}(\lambda)$ and $Abs_{PO}(\lambda)$ on one particle size sample set and tested on another particle size sample set. L.V. is the number of latent variables used for the calibration model, R^2 is the coefficient of determination, SEP_c is standard error of prediction corrected form the bias in $g.kg^{-1}$.

Table 1

	n	Mean	SD	Min	Q1	Q2	Q3	Max	Skewness
TOC	52	88.6	48.04	11.4	50.20	88.75	115.0	248.0	0.86

Table 1: Descriptive statistics for the whole dataset. Q1, Q2 and Q3 correspond respectively to the first quartile, the median and the upper quartile. SD: standard deviation for TOC: Total Organic Carbon ($g.kg^{-1}$)

Table 2

Particle size of the Calibration set	Particle size of the Test set	Signal	L.V.	R^2	SEP_c	Bias	Slope
Coarse	Sieved	$Abs_{BS}(\lambda)$	5	0.64	29	-6.5	0.74
		$Abs_{PO}(\lambda)$	5	0.76	24	-5.7	0.86
	Ground	$Abs_{BS}(\lambda)$	5	0.67	28	-44	0.70
		$Abs_{PO}(\lambda)$	5	0.62	31	-33	0.50
Sieved	Coarse	$Abs_{BS}(\lambda)$	8	0.53	37	24.5	0.78
		$Abs_{PO}(\lambda)$	5	0.67	28	6.0	0.72
	Ground	$Abs_{BS}(\lambda)$	8	0.75	24	-20	0.72
		$Abs_{PO}(\lambda)$	5	0.70	28	-34	0.54
Ground	Coarse	$Abs_{BS}(\lambda)$	10	0.45	44	12	0.8
		$Abs_{PO}(\lambda)$	4	0.50	52	23	1.1
	Sieved	$Abs_{BS}(\lambda)$	10	0.70	27	11	0.84
		$Abs_{PO}(\lambda)$	4	0.69	43	31	1.28

Table 2: Performance of the models built with $Abs_{BS}(\lambda)$ and $Abs_{PO}(\lambda)$ on one particle size sample set and tested on another particle size sample set. L.V. is the number of latent variables used for the calibration model, R^2 is the coefficient of determination, SEP_c is standard error of prediction corrected form the bias in $g.kg^{-1}$.

Figure legend

Figure 1 : Schematic diagram of polarized light spectroscopy system (PoLiS).

Figure 2 : Principle of the measurement of the two components $I_{\parallel}(\lambda)$ and $I_{\perp}(\lambda)$ of the totally backscattered light by means of linear light polarization

Figure 3 : Mean reflectance $R_{BS}(\lambda)$, backscattered absorbance $Abs_{BS}(\lambda)$, PoLiS absorbance $Abs_{PO}(\lambda)$ per quartile of TOC concentration for the three different particle sizes (a.) coarse < 5mm, (b.) sieved < 2 mm and (c.) ground < 0.25 mm

Figure 4 : Scores plots of the two principal components of the Principal Component Analysis performed on the absorbance spectra $Abs_{BS}(\lambda)$ (first line) and $Abs_{PO}(\lambda)$ (second line) for different data centering (mean centering and centering per sample location) methods.

Figure 5 : Correlogram between Absorbance and TOC for the wavelength range 400 - 800 nm. Vertical line indicates the wavelength at which the correlation coefficient for $Abs_{BS}(\lambda)$ is the highest.

Figure 6 : Plot of the backscattered absorbance $Abs_{BS}(\lambda)$ and the PoLiS absorbance $Abs_{PO}(\lambda)$ at wavelength λ vs the TOC concentration (in $g \cdot kg^{-1}$) for the three different particle sizes: coarse < 5 mm, sieved < 2 mm and ground < 0.25 mm) with linear fitting. R is the Pearson's coefficient.

Figure 7 : Predicted vs measured total organic carbon content from leave-one-out cross validation models calibrated with backscattered reflectance spectra (R_{BS}), backscattered absorbance ($Abs_{BS}(\lambda)$) and PoLiS Absorbance ($Abs_{PO}(\lambda)$) for the three different particle sizes: (a.) coarse < 5mm , (b.) sieved < 2 mm and (c.) finely ground < 0.25 mm) . R^2 : coefficient of determination; SECV: standard error of cross validation; LV: number of latent variables

Figure 8 : Comparison of the determination coefficient R^2 and the Standard Error of cross validation (SECV) of the prediction models built on the three types of samples. Dotted lines correspond to the performances of the models built with $Abs_{PO}(\lambda)$.

Figure 9 : Predicted vs measured total organic carbon content. Models were calibrated with the backscattered absorbance ($Abs_{BS}(\lambda)$) and the PoLiS Absorbance ($Abs_{PO}(\lambda)$) on one particle size class and tested on another particle size class. (upperline: coarse < 5 mm on sieved < 2 mm and lower line: sieved < 2 mm on coarse <5 mm) . R^2 : coefficient of determination, SEP_c : standard error of Prediction corrected from the bias in $g \cdot kg^{-1}$.

Figure 1

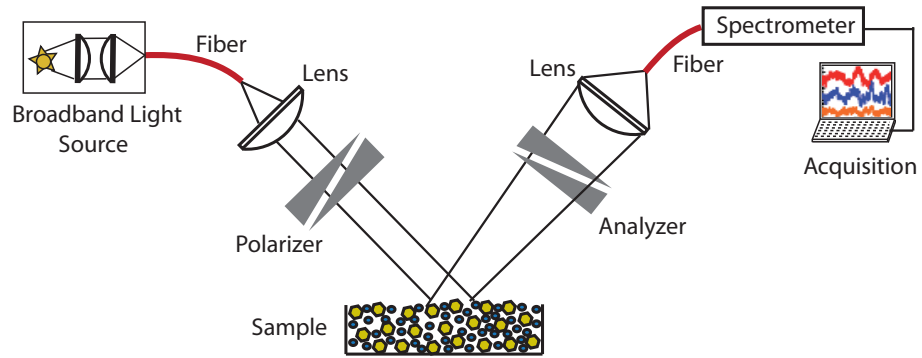


Figure 1: Schematic diagram of polarized light spectroscopy system (PoLiS).

Figure 2

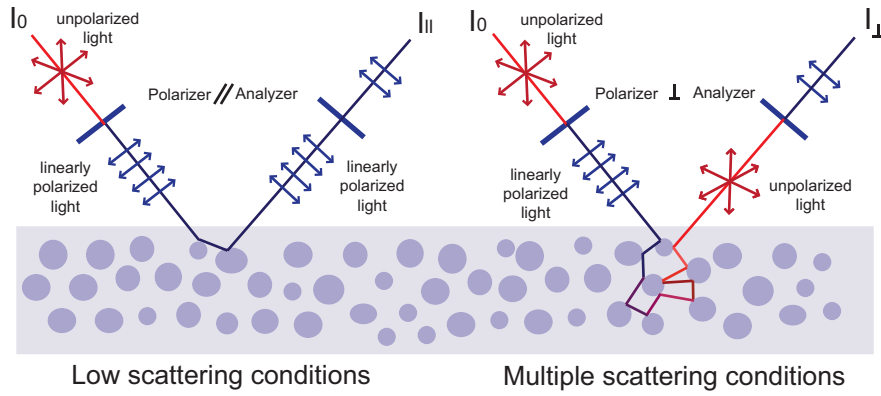


Figure 2: Principle of the measurement of the two components $I_{||}(\lambda)$ and $I_{\perp}(\lambda)$ of the totally backscattered light by means of linear light polarization

Figure 3

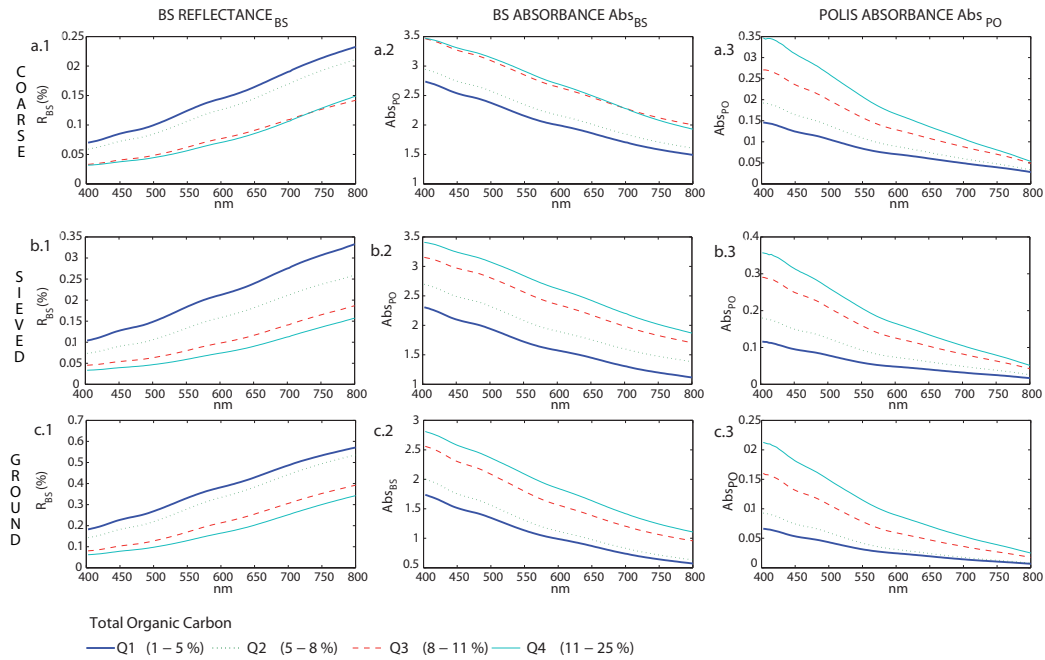


Figure 3: Mean reflectance $R_{BS}(\lambda)$, backscattered absorbance $Abs_{BS}(\lambda)$, PoLiS absorbance $Abs_{PO}(\lambda)$ per quartile of TOC concentration for the three different particle sizes (a.) coarse < 5mm, (b.) sieved < 2 mm and (c.) ground < 0.25 mm

Figure 4

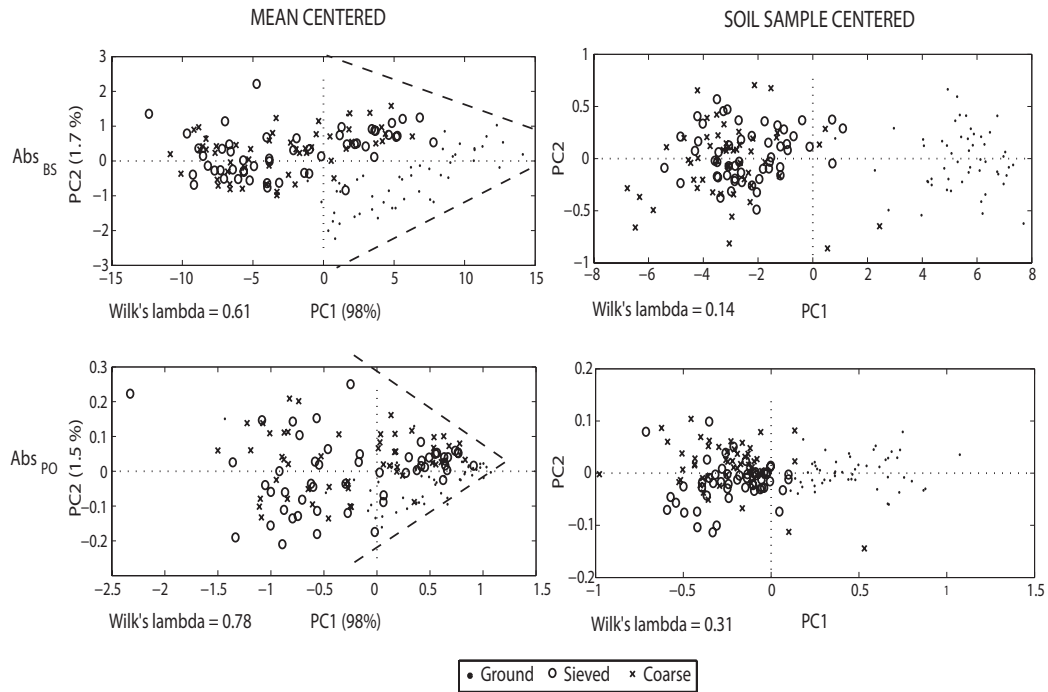


Figure 4: Scores plots of the two principal components of the Principal Component Analysis performed on the absorbance spectra $Abs_{BS}(\lambda)$ (first line) and $Abs_{PO}(\lambda)$ (second line) for different data centering (mean centering and centering per sample location) methods.

Figure 5

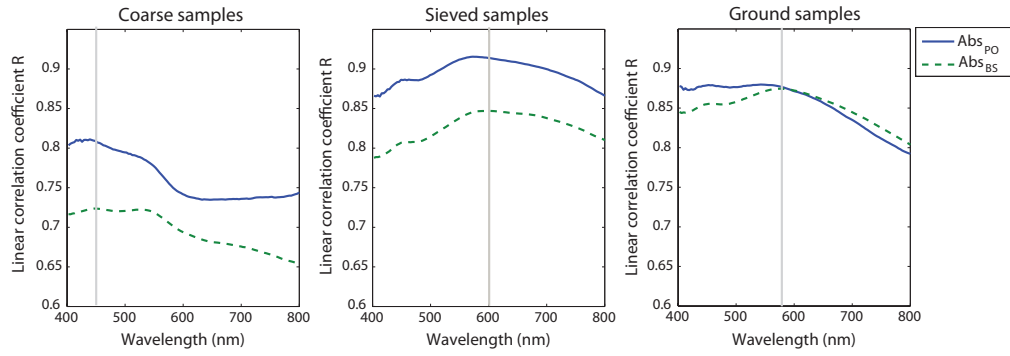


Figure 5: Correlogram between Absorbance and TOC for the wavelength range 400 - 800 nm. Vertical line indicates the wavelength at which the correlation coefficient for $Abs_{BS}(\lambda)$ is the highest.

Figure 6

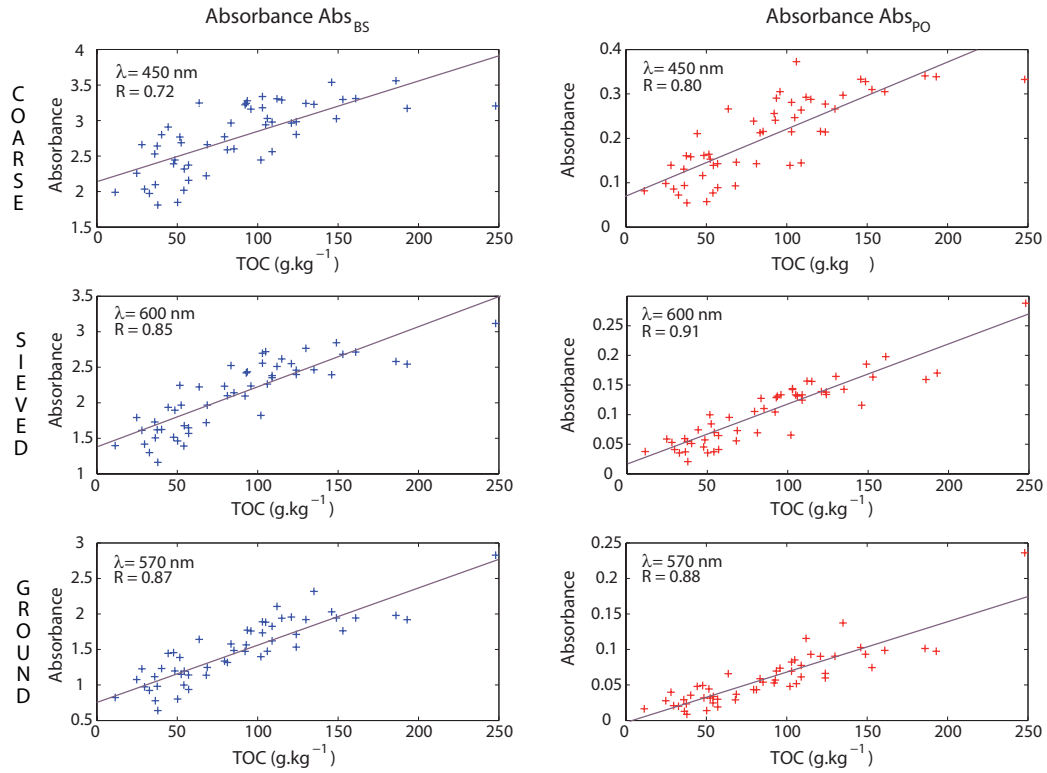


Figure 6: Plot of the backscattered absorbance $Abs_{BS}(\lambda)$ and the PoLiS absorbance $Abs_{PO}(\lambda)$ at wavelength λ vs the TOC concentration (in $g \cdot kg^{-1}$) for the three different particle sizes: coarse < 5 mm, sieved < 2 mm and ground < 0.25 mm) with linear fitting. R is the Pearson's coefficient.

Figure 7

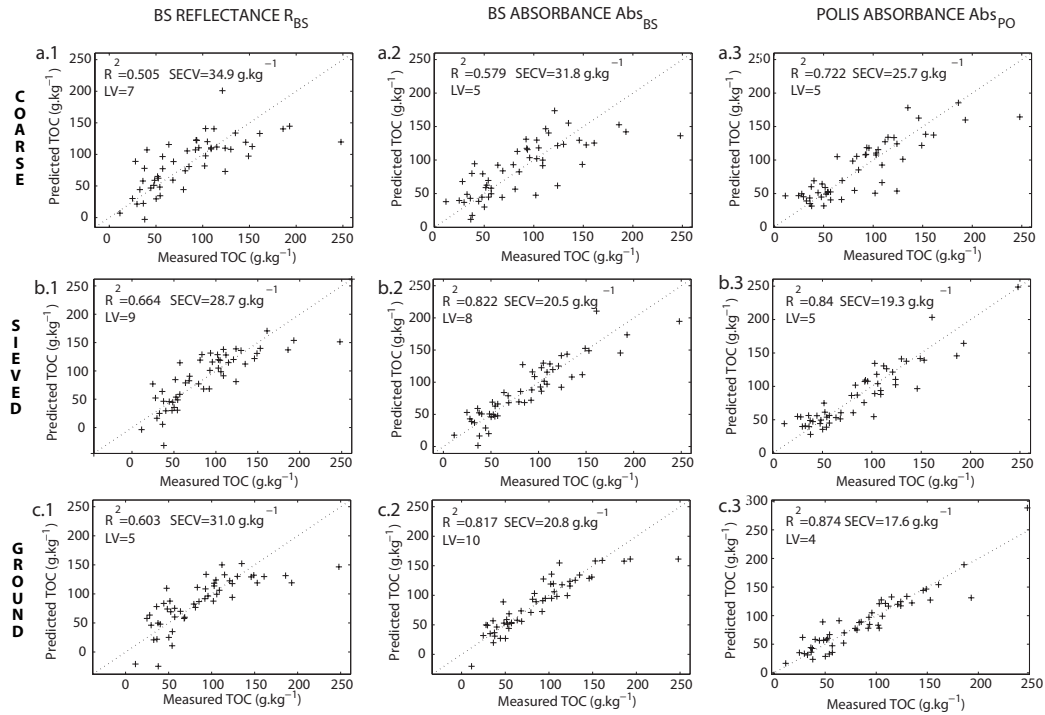


Figure 7: Predicted vs measured total organic carbon content from leave-one-out cross validation models calibrated with backscattered reflectance spectra (R_{BS}), backscattered absorbance ($Abs_{BS}(\lambda)$) and PoLiS Absorbance ($Abs_{PO}(\lambda)$) for the three different particle sizes: (a.) coarse < 5mm , (b.) sieved < 2 mm and (c.) finely ground < 0.25 mm) . R^2 : coefficient of determination; SECV: standard error of cross validation; LV: number of latent variables

Figure 8

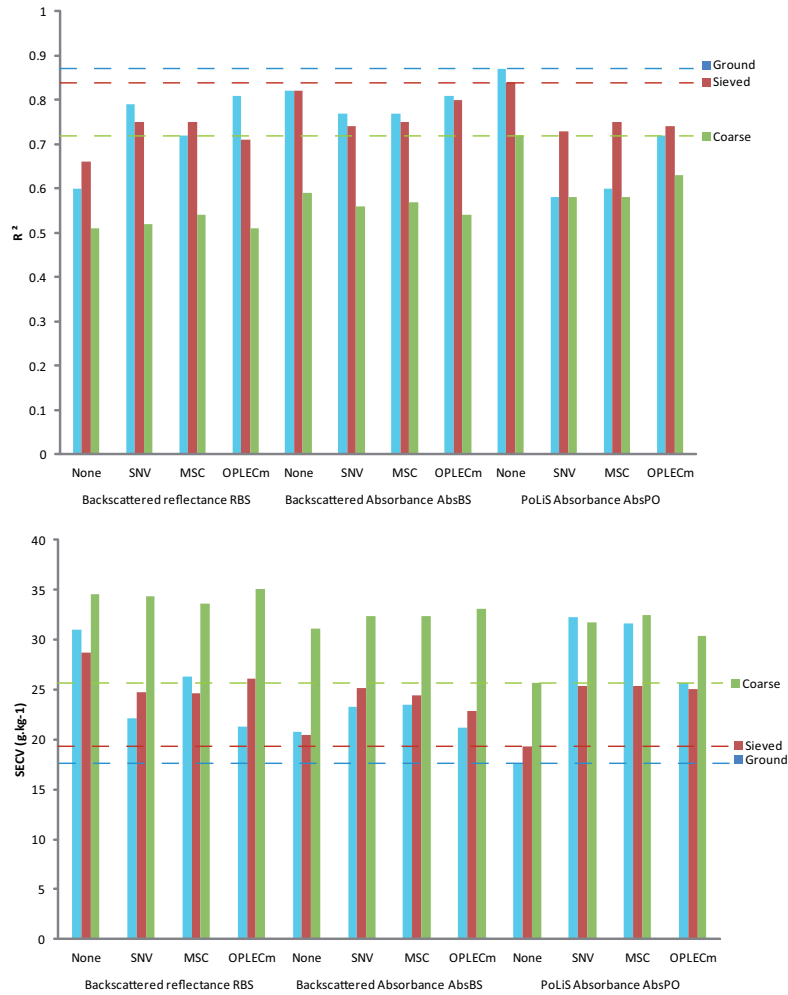


Figure 8: Comparison of the determination coefficient R^2 and the Standard Error of cross validation (SECV) of the prediction models built on the three types of samples. Dotted lines correspond to the performances of the models built with $Abs_{PO}(\lambda)$.

Figure 9

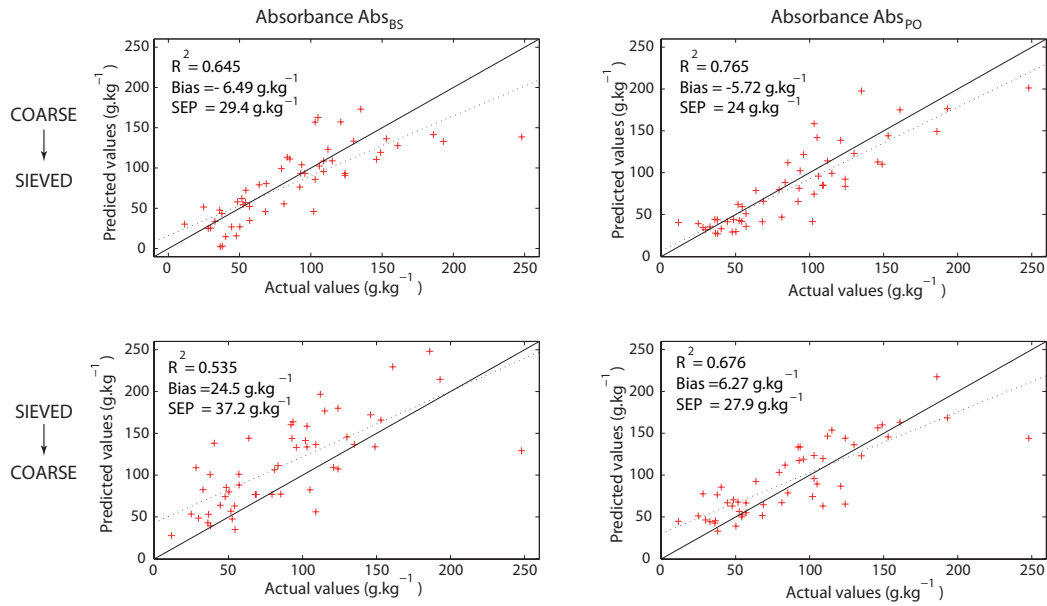


Figure 9: Predicted vs measured total organic carbon content. Models were calibrated with the backscattered absorbance ($Abs_{BS}(\lambda)$) and the PoLiS Absorbance ($Abs_{PO}(\lambda)$) on one particle size class and tested on another particle size class. (upperline: coarse < 5 mm on sieved < 2 mm and lower line: sieved < 2 mm on coarse < 5 mm). R^2 : coefficient of determination, SEP_c : standard error of Prediction corrected from the bias in g.kg⁻¹.