

Division - Soil in Space and Time | Commission - Pedometrics

Reflectance spectroscopy in the prediction of soil organic carbon associated with humic substances

Sharon Gomes Ribeiro^{(1)*} , Marcio Regys Rabelo de Oliveira⁽²⁾ , Letícia Machado Lopes⁽¹⁾ , Mirian Cristina Gomes Costa⁽³⁾ , Raul Shiso Toma⁽³⁾ , Isabel Cristina da Silva Araújo⁽⁴⁾ , Luis Clenio Jario Moreira⁽⁵⁾  and Adunias dos Santos Teixeira⁽⁴⁾ 

⁽¹⁾ Universidade Federal do Ceará, Programa de Pós-Graduação em Ciência do Solo, Fortaleza, Ceará, Brasil.

⁽²⁾ Universidade Federal do Ceará, Programa de Pós-Graduação em Engenharia Agrícola, Fortaleza, Ceará, Brasil.

⁽³⁾ Universidade Federal do Ceará, Departamento de Ciência do Solo, Fortaleza, Ceará, Brasil.

⁽⁴⁾ Universidade Federal do Ceará, Departamento de Engenharia Agrícola, Fortaleza, Ceará, Brasil.

⁽⁵⁾ Instituto Federal de Educação, Ciência e Tecnologia do Ceará, Departamento de Agronomia, Limoeiro do Norte, Ceará, Brasil.

ABSTRACT: Understanding organic carbon and predominant humic fractions in the soil allows contributes to soil quality management. Conventional fractionation techniques require time, excessive sampling, and high maintenance costs. In this study, predictive models for organic carbon in humic substances (HS) were evaluated using hyperspectral data as an alternative to chemical fractionation and quantification by wet digestion. Twenty-nine samples of *Neossolos Flúvicos* (Fluvents) - A1, and 36 samples of *Cambissolos* (Inceptisols) - A2 were used. The samples were also analyzed jointly, creating a third sample group - A1&A2. Untransformed spectral reflectance factors were obtained using the FieldSpec Pro FR 3 hyperspectral sensor (350–2500 nm). Pre-processing techniques were employed, including Savitzky–Golay smoothing and first- and second-order derivative analysis. After selecting variables using the Backward method, which removes spectral variables that are not statistically significant for the regression. Estimation models were built by Principal Components Regression (PCR) and Partial Least Squares Regression (PLSR). The spectral data were evaluated individually for soil classes A1 and A2, and jointly for A1&A2. The PLSR was more efficient than PCR, especially for the estimation models that used the first derivative of reflectance employing the three sample groups. For samples of A1, the best estimate was seen for humic acid (RPD = 6.09) and humin (RPD = 2.38); for A2, the best models estimated the OC in fulvic acid (RPD = 2.35) and humin (RPD = 2.51); and for the joint spectral data (A1&A2), the prediction was robust for humin only (RPD = 2.01). The most representative wavelengths were observed using the first derivative with PLSR and PCR, centred on the region between 1600 and 1800 nm. The first-derivative of reflectance calculated more-robust predictive models using PLSR than PCR. The best predictions occurred for organic carbon associated with humic acid in *Neossolos Flúvicos*, with fulvic acid in *Cambissolos*, and with humin in both soil classes.

Keywords: spectroradiometry, pedometrics, organic matter.

* Corresponding author:

E-mail: sharonribeiro@alu.ufc.br

Received: October 22, 2022

Approved: March 09, 2023

How to cite: Ribeiro SG, Oliveira MRR, Lopes LM, Costa MCG, Toma RS, Araújo ICS, Moreira LCJ, Teixeira AS. Reflectance spectroscopy in the prediction of soil organic carbon associated with humic substances. Rev Bras Cienc Solo. 2023;47:e0220143
<https://doi.org/10.36783/18069657rbc20220143>

Editors: José Miguel Reichert  and Leônidas Azevedo Carrijo Melo .

Copyright: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.



INTRODUCTION

Stabilization soil organic matter (SOM) and physical protection between solid soil particles are important factors that greatly influence the permanence of organic carbon (OC) in the soil matrix (Jiménez-González et al., 2019). The quality and stability of SOM can be assessed by quantifying the OC associated with its most stable fractions: the humic substances (HS), whose structures include around 70 % of the element (Primo et al., 2011).

Carbon associated with HS can be quantified using chemical fractionation of the SOM (Schnitzer, 1978; Kumada, 1987), verifying the solubility of each fraction in acid or alkali media, followed by digestion and quantification of the OC, carried out via dry combustion or wet digestion. These fractions are characterized as: i) fulvic acid (FA) - soluble, regardless of the pH of the environment; ii) humic acid (HA) - insoluble in acids; iii) humin (HUM) - insoluble in both acid and alkaline solution (Ebeling et al., 2013).

Although accurate, conventional techniques for quantifying OC are generally costly and routinely require equipment maintenance, giving slow results and producing excessive waste (Xiaoju et al., 2021). Among alternative methods to chemical analysis, remote sensing techniques have been evaluated to estimate humic substances and OC in the soil (Terra et al., 2013; Cambou et al., 2021; Raiesi, 2021; Xie et al., 2021).

When evaluating soil properties, reflectance spectroscopy in the region of 350-2500 nm is usually able to generate analytical results quickly and non-invasively (Tomazoni and Guimarães, 2015). Molecular vibrations characteristic of humic substances and humification can be inferred from specific wavelengths between 1000 and 2500 nm, and provide relevant information concerning their structure and reactivity (Canellas and Rumjanek, 2005; Granlund et al., 2021).

Spectral data sets present a significant number of variables for use. To reduce the dimensionality of the set and reveal the best spectral variables for predictive modelling, multivariate regression methods – such as Principal Components Regression (PCR) and Partial Least Squares Regression (PLSR) – have been successfully applied by several recent authors to building predictive models based on spectral variables (Pudelko and Chodak, 2020; Liu et al., 2020; Zhang et al., 2021; Ribeiro et al., 2021) to predict OC, and other soil properties using spectroscopy between 350 and 2500 nm. The approach has shown promise in estimating the carbon content of each humic substance (C-HS) in soil samples or isolated extracts (Madhavan et al., 2017; Gomez et al., 2020). However, research focused on predicting humic substances in semi-arid soils is still rare, and few studies are currently found in the literature.

Brazil's semi-arid region is considered heterogeneous regarding environmental and pedological conditions, and a variety of soil classes is found throughout the region, including from poorly weathered soils to the most. It is known that soils at different stages of development have properties that significantly affect the dynamics and degree of organic matter stabilization (Fontana et al., 2008; Cunha et al., 2010).

We hypothesized that reflectance spectroscopy (350 - 2500 nm) allied to PLSR and PCR is an efficient alternative to fractionation and chemical digestion for estimating the organic carbon present in the humic fractions of soil samples from the semi-arid region. This study aimed to: i) evaluate the correlation between variations in the levels of humic substances and the reflectance factors of the soil samples; ii) evaluate the efficiency of spectral transformations and the PLSR and PCR methodologies in predicting the organic carbon content in each humic substance; and iii) build linear regression models to estimate the carbon in humic acid, fulvic acid and humin.

MATERIALS AND METHODS

Soil properties and sample preparation

Samples were collected from the surface horizons (0.00-0.10 m) of the soils in two distinct areas of irrigated perimeters in the state of Ceará in the northeast of Brazil (Figure 1). The first area is located in the Irrigated District of Morada Nova (A1), in the Banabuiú hydrographic basin, with a predominance of *Neossolos Flúvicos* (Fluvents) whose textural classes fall between sandy loam and silty-clay loam. The second collection area (A2) is a part of the Jaguaribe-Apodi Irrigated District, located in the district of Limoeiro do Norte, in the basin of the Lower Jaguaribe, where *Cambissolos* (Inceptisols) of a sandy-loam to clayey texture predominate, with the significant presence of iron oxides (Jacomine et al., 1973; Ribeiro et al., 2021). Twenty-nine soil samples from A1, and 36 samples from A2 were used. Each collected soil sample was ground up, air-dried and sifted using a 2-mm mesh.

Chemical fractionation and quantification of organic carbon in the humic substances

Chemical fractionation was based on the characterization of humic substances as proposed by Swift (1996), described in figure 2. The procedures were carried out in the Pedology Laboratory of the Department of Soil Sciences at the Centre for Agricultural Sciences of the Federal University of Ceará (CCA-UFC).

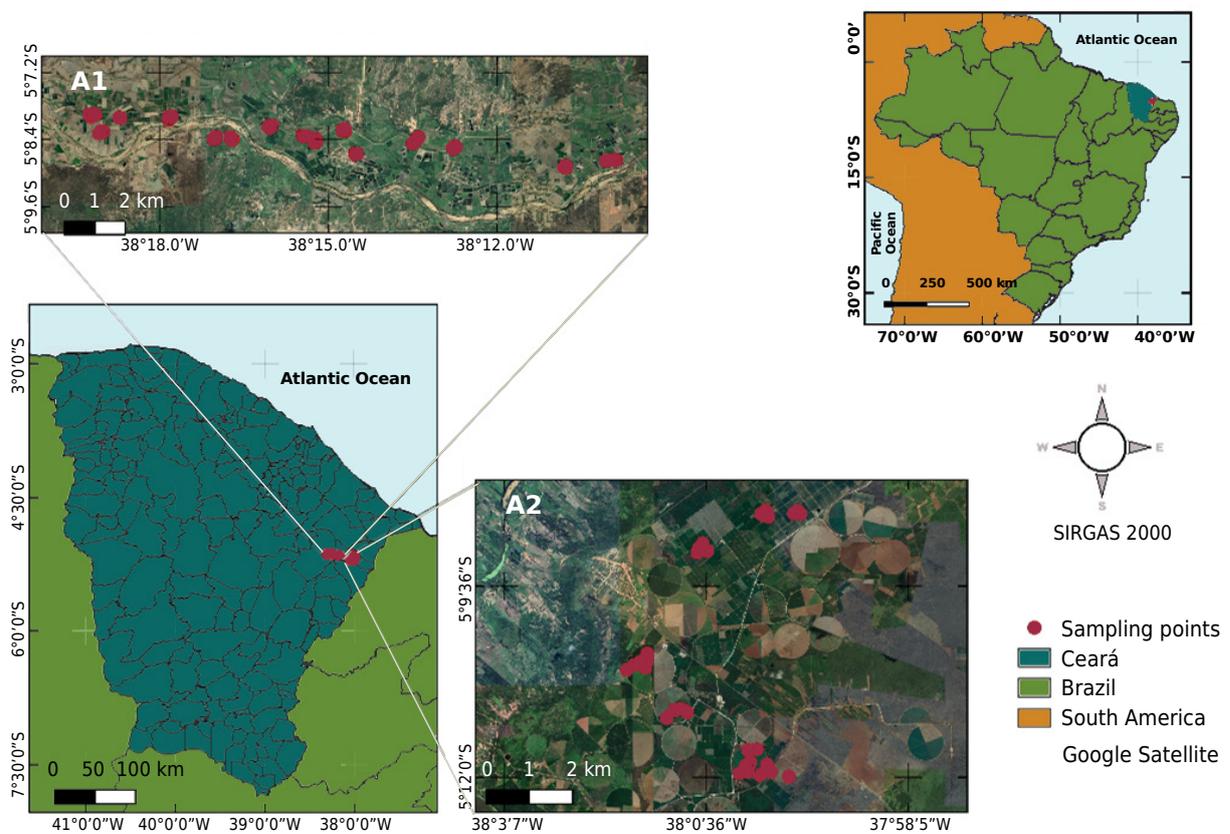


Figure 1. Distribution map of the collection points of the soil samples used in the study.

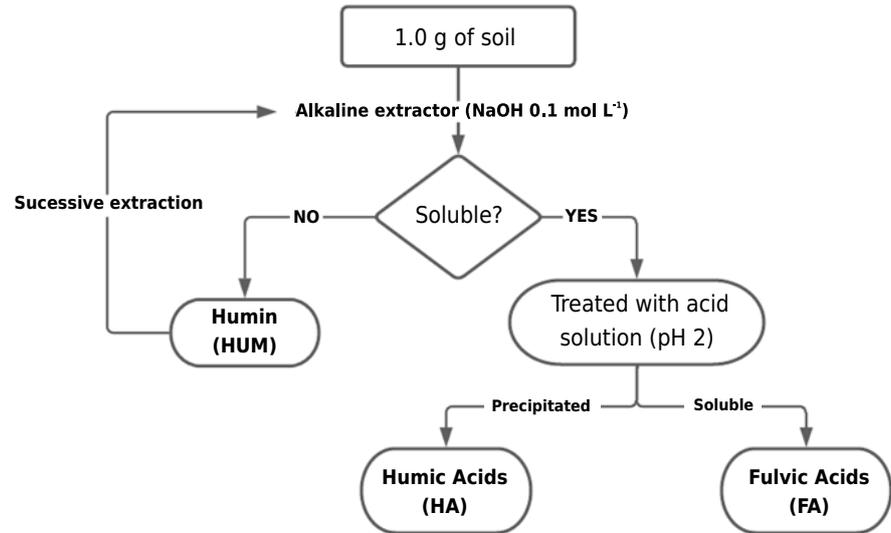


Figure 2. Flowchart of the chemical fractionation of the humic substances in the soil samples from this study.

Organic carbon was quantified using the Yeomans and Bremner method (1988), digesting in potassium dichromate and sulphuric acid with external heating, from when it was possible to evaluate the OC concentration in each humic fraction: humic acid (C-HA), fulvic acid (C-FA) and humin (C-HUM).

Acquisition and handling of the hyperspectral data

Hyperspectral data were acquired in the dark-room of the Geoprocessing Laboratory at the Centre for Agricultural Sciences of the Federal University of Ceará (CCA-UFC), as shown in figure 3.

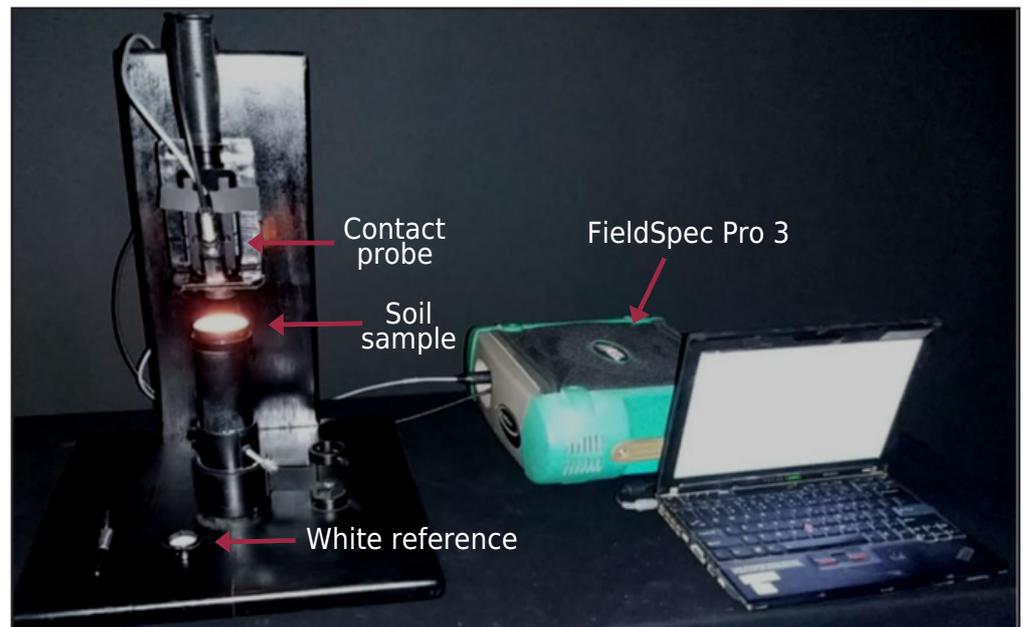


Figure 3. Geometry of the dark-room acquisition of hyperspectral data.

Oven-dried soil samples (45 °C for 24 h) were placed in a black polypropylene cylinder, measuring 0.05 m in diameter and 15 mm in height. Spectral readings were taken of the duly identified samples using the FieldSpec Pro FR 3 spectroradiometer (Figure 3), with the Hi-Bright Contact probe attached to the support to avoid instability and noise when taking readings in the VNIR-SWIR region (350-2500 nm).

The spectroradiometer was calibrated for maximum reflectance (white reference) using a spectral plate. The reflectance factors (RF) were subjected to first-order derivative transformation, as per Rudorff et al. (2007) (Equation 1), and from this, the second derivative was determined.

$$\frac{d\rho_{\lambda}}{dx} \simeq \frac{\rho_{i+1} - \rho_{i-1}}{2\Delta x} \quad \text{Eq. 1}$$

in which: Δx corresponds to the distance between two successive bands ($\Delta x = x_{i+1} - x_{i-1}$), allowing that $x_{i+1} > x_{i-1}$; ρ_{i+1} refers to the reflectance factor of the point following i ; ρ_{i-1} corresponds to the reflectance factor of the point preceding i .

The reflectance of the soil samples was also subjected to the smoothing of Savitzky and Golay (1964), which seeks to reduce random noise and avoids introducing distortions into the spectral data, preserving the shape of the spectrum, as per equation 2:

$$y_j^* = \frac{1}{N} \sum_{h=-k}^k C_h y_{j+h} \quad \text{Eq. 2}$$

in which: y_j^* is the new smoothed value; C_h represents the coefficients of the smoothing filter; N is the size of the smoothing window; k is the number of neighbours to the left and right of j .

Chemometric analysis was carried out using the collected data in two different ways: i) individual observations in A1 (*Neossolos Flúvicos* - Fluvents) and A2 (*Cambissolos* - Inceptisols); and ii) observations grouped into an A1&A2 dataset (*Neossolos Flúvicos* and *Cambissolos* - Fluvents and Inceptisols) to evaluate the effectiveness of the estimation models without considering the chemical or spectral heterogeneity of the two soil classes.

Descriptive statistics

Initially, an analysis was made of the frequency distribution of the C-HS data in the A1, A2 and A1&A2 sample sets, considering the Kolmogorov-Smirnov normality test at 5 %; the median distribution of the OC content in each fraction of the humic substances was also evaluated. The linear correlation between C-FA, C-HA and C-HUM and the non-transformed reflectance factor of the samples was analyzed for each of the wavelengths under study, as per the Pearson Correlation Equation (Equation 3).

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left[\sum_{i=1}^n (x_i - \bar{x})^2 \right] \left[\sum_{i=1}^n (y_i - \bar{y})^2 \right]}} \quad \text{Eq. 3}$$

in which: r represents the Pearson correlation coefficient; x_i and y_i are each of the measured variables (independent and dependent, respectively) for the i -th individual; and \bar{x} and \bar{y} represent the arithmetic mean of variables X and Y .

Estimating organic carbon in the humic substances

The contents of C-FA, C-HA and C-HUM were submitted to min-max normalisation (Equation 4) to reduce the effect of scale and magnitude between the parameters of the estimation

models. The individual data sets (A1 and A2) and the joint dataset (A1&A2) were normalized as per equation (4).

$$N_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad \text{Eq. 4}$$

in which: N_i corresponds to the normalized value of the i -th observation; x_i is the real value of variable x in the i -th observation; and $\min(x)$ and $\max(x)$ are the minimum and maximum values of x .

Wavelengths between 350 and 2500 nm that most influenced the variation in the content of each humic fraction were selected by the Backward method. This method removed redundant spectral variables and selected only those with the best statistical correlations for the variation in the content of C-HS for the regression (Shiferaw and Hergarten, 2014). With the spectral variables selected, mathematical models were built using Partial Least Squares Regression (PLSR) and Principal Components Regression (PCR) to predict the OC in the humic fractions of each data set (A1, A2 and A1&A2).

Of each set under evaluation, 30 % of the data were used as unpublished data and destined for the external validation process of the predictive models (Figure 4). Model validation was carried out using the following statistical metrics: coefficient of determination (R^2) (Equation 5); adjusted coefficient of determination ($R^2_{adj.}$) (Equation 6); Root Mean Square Error (RMSE) (Equation 7); and the Ratio of Prediction to Deviation (RPD) (Equation 8).

$$R^2 = 1 - \frac{\sum_{i=1}^N (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^N (Y_i - \bar{Y})^2} \quad \text{Eq. 5}$$

$$R^2_{adj.} = 1 - \frac{(N-1)(1-R^2)}{N-(k+1)} \quad \text{Eq. 6}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (\hat{Y}_i - Y_i)^2}{N}} \quad \text{Eq. 7}$$

$$RPD = \frac{\sigma_{Y_o}}{RMSE} \quad \text{Eq. 8}$$

in which: \hat{Y}_i represents the values calculated by the models for the i -th observation; Y_i are the values measured in the laboratory for the i -th observation; \bar{Y} represents the mean of the observed values; N is the number of observations; k is the total of independent variables; and σ_{Y_o} is the standard deviation of the measured or observed values.

RESULTS

Descriptive statistics

Figure 5a shows the median value for carbon in the HA and FA fractions, with the data separated by collection area. The samples from A1 showed a predominance of C-HA, with a median of 0.67 g kg⁻¹ OC, while C-FA showed a median of 0.42 g kg⁻¹ OC; whereas in the samples from A2, C-FA predominates, with a median of 0.70 g kg⁻¹ and C-HA showing lower values, with a median of around 0.56 g kg⁻¹.

When evaluating the A1&A2 joint samples, no significant difference was seen between C-FA and C-HA. The C-HUM content stood out regarding data distribution and carbon concentration, as shown in figure 5b. The mean value of the humin fraction was higher than that of the other humic substances in each sample set. The samples from A2 showed the highest contents and the smallest variations in carbon content.

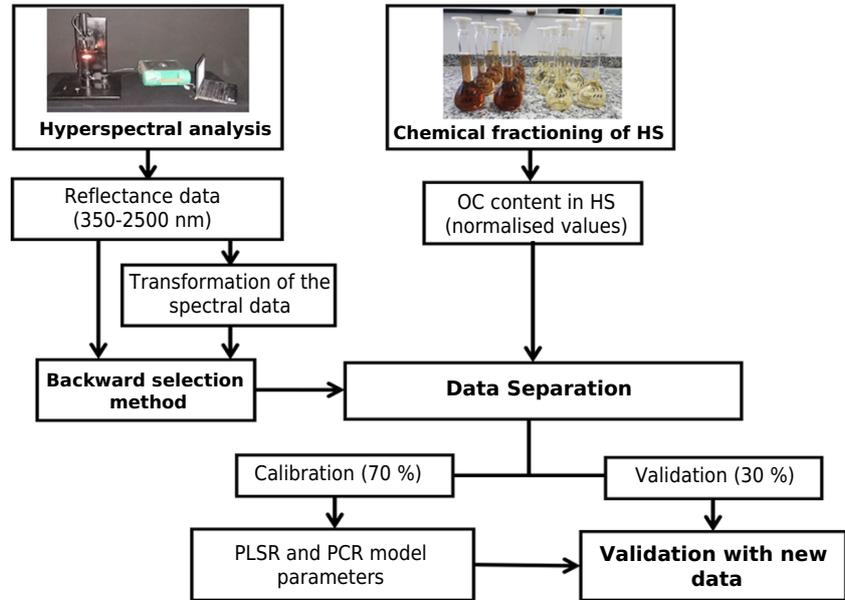


Figure 4. Methodological flowchart for the calibration and validation of the PLSR and PCR models.

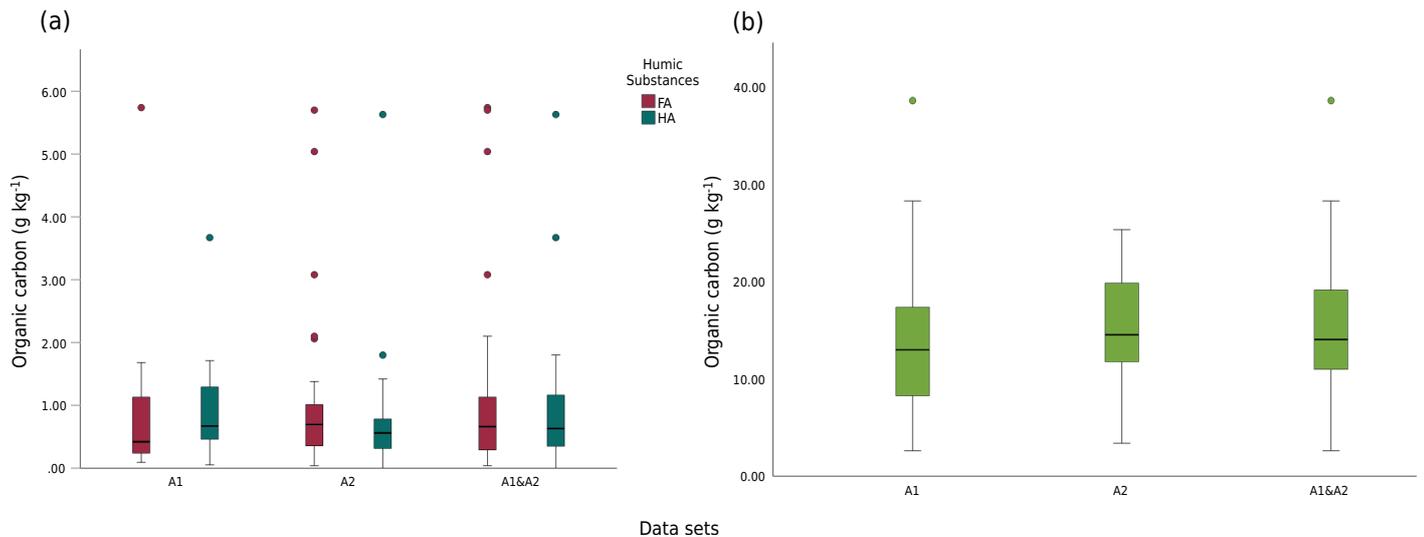


Figure 5. Boxplot of the organic carbon content (g kg^{-1}) obtained by chemical fractionation in the fulvic acid (FA) and humic acid (HA) fractions (a) and in the humin (HUM) fraction (b), separated by area (A1 and A2), and for the joint data set (A1&A2).

Pearson correlation (r) between the C-HS content and wavelength

Figures 6a, 6b and 6c show that for each set of samples, the correlation between C-FA, C-HA and wavelength with the reflectance of the samples shows $|r| < 0.3$. On the other hand, the variation in C-HUM showed negative correlations with the entire reflectance spectrum for the three sample sets, albeit never exceeding $|r| = 0.4$.

When using the spectral and chemical fractionation data from the A1&A2 set (Figure 6c), there is weak correlation between the variation in C-HS content and the reflectance of the soil samples, with behavior similar to that seen for the individual samples from A1 (Figure 6a) and A2 (Figure 6b).

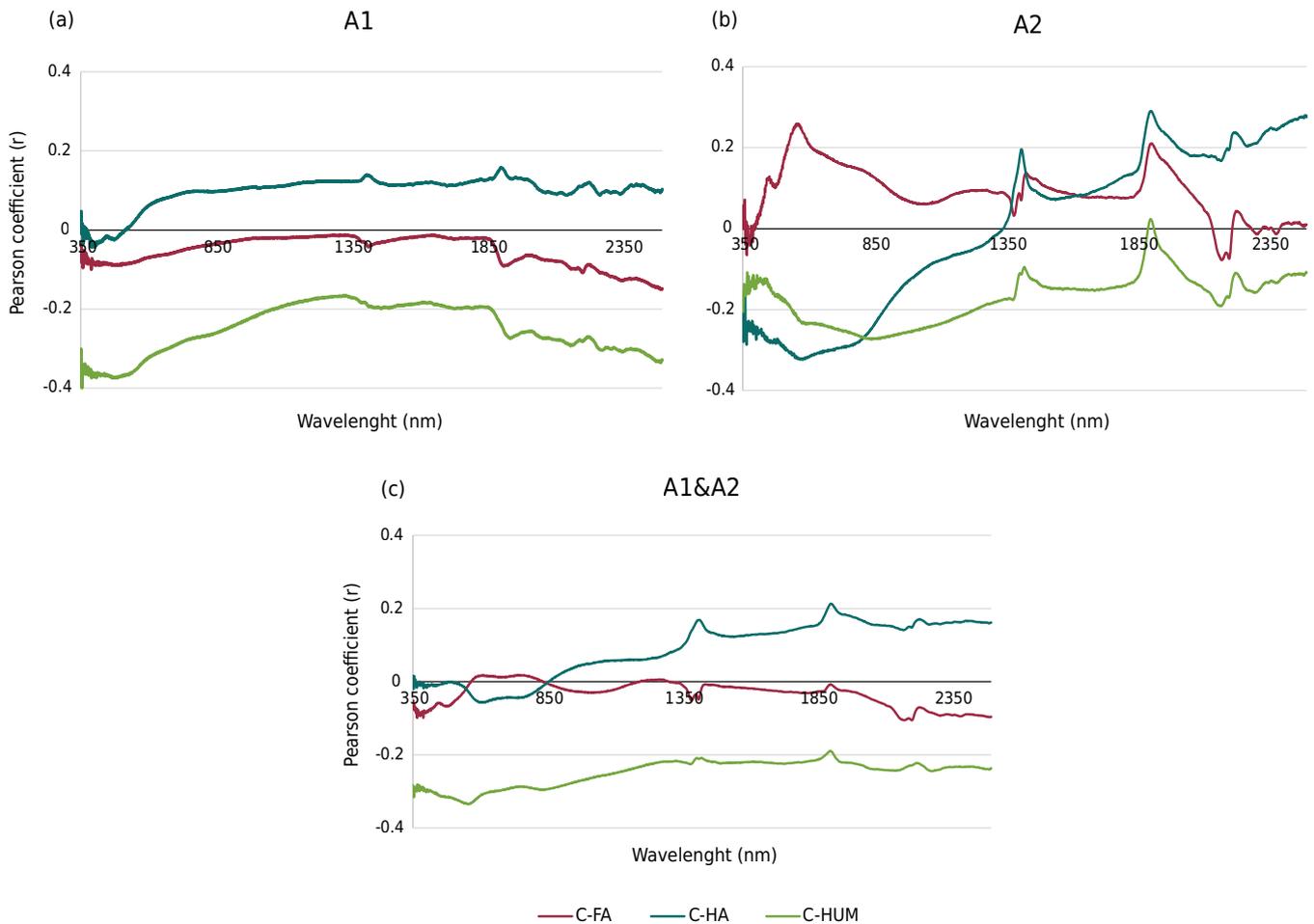


Figure 6. Pearson correlation between reflectance and the organic carbon content in humic substances obtained by chemical fractionation, individually in A1 (a) and A2 (b), and jointly in the A1&A2 data set (c).

Estimating organic carbon in the humic substances

Principal Components Regression (PCR)

The PCR estimation models showed better performance using RF transformed into the first derivative ($RPD > 2.0$ and $R^2_{adj} > 0.8$) for each sample set. The C-HA and C-HUM content is highlighted when using the spectral response of the soil samples from A1 (Figures 7a and 7b), and the C-FA and C-HUM content when using the samples from A2 (Figures 7c and 7d).

For the A1&A2 joint set (Figure 8), only the C-HUM estimation model with the first derivative showed satisfactory performance ($RPD > 1.4$). The models that used untransformed reflectance data, the second derivative and smoothed reflectance, obtained an $RPD < 1.4$. The most efficient regression models ($RPD > 2.0$) are shown in table 1, with the respective coefficients generated by PCR for each significant wavelength. These models achieved a desirable performance only when using soil samples separated by region.

Partial Least Squares Regression (PLSR)

As with PCR, the estimation models showed better performance with PLSR when using the first-order derivative of the reflectance data, with an $RPD > 2.0$ and $R^2_{adj} > 0.8$ when validating (Figures 9a to 9d).

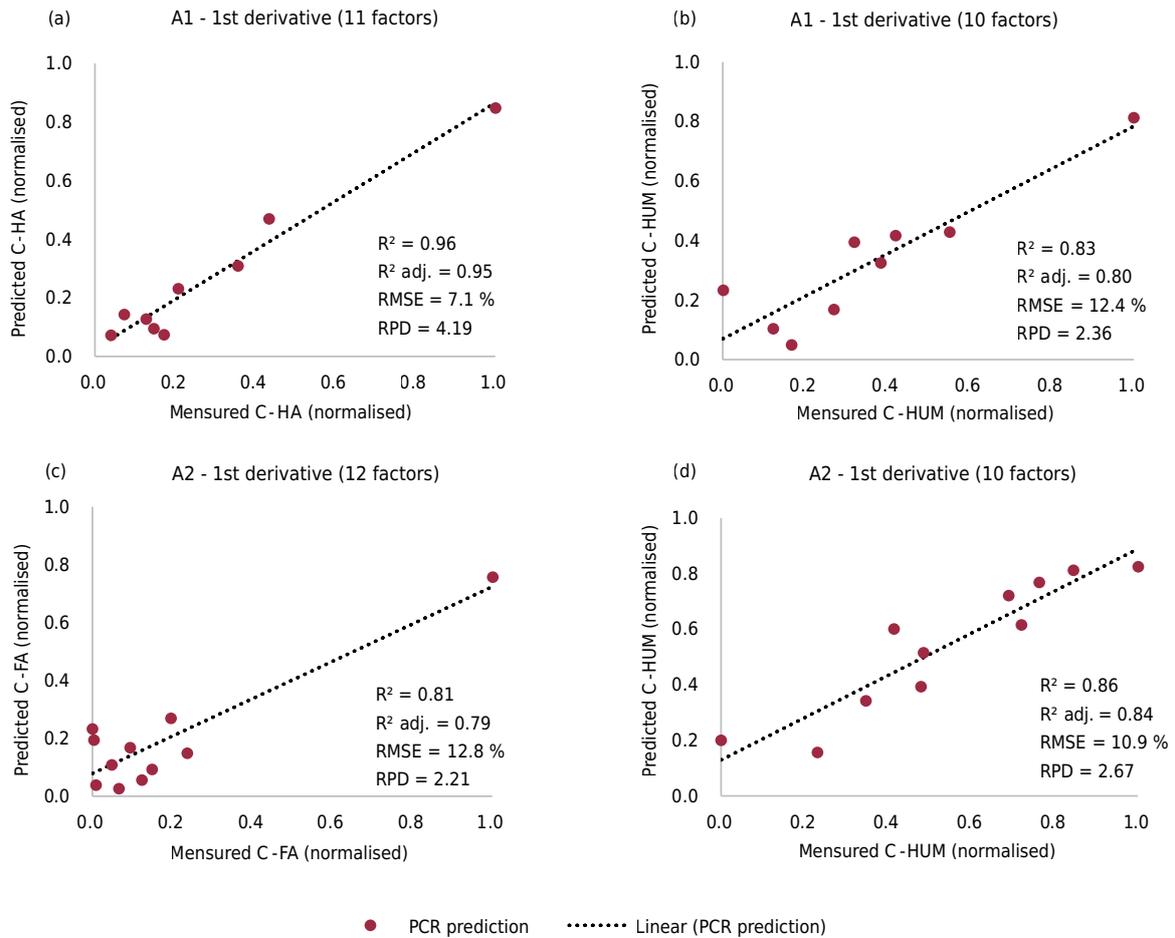


Figure 7. Validation of the best-fitting PCR models for predicting C-HA in A1 (a), C-HUM in A1 (b), C-FA in A2 (c), and C-HUM in A2 (d).

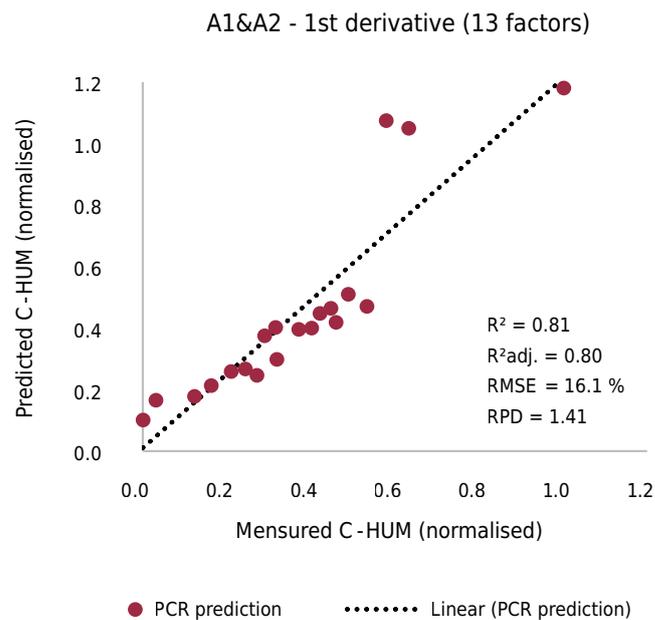


Figure 8. Validation of the best-fitting PCR model for predicting C-HUM in the A1&A2 joint data set.

Table 1. Equations of the best PCR models (RPD >2.0) for predicting organic carbon in humic substances, with the respective adjusted R²

Sample	Humic substances	Spectral data	Best PCR prediction models for C-HS (normalised)	R ² _{adj.}
A1	HA	First Derivative (ρ')	$0.475 - 2106.56 (\rho'_{832nm}) + 92.94 (\rho'_{2317nm}) - 230.83 (\rho'_{371nm}) + 2243.5 (\rho'_{912nm}) - 1789.59 (\rho'_{901nm}) + 2900.01 (\rho'_{843nm}) - 1461 (\rho'_{1223nm}) + 323.5 (\rho'_{1606nm}) - 987.23 (\rho'_{1661nm}) - 197.34 (\rho'_{2331nm}) + 360.52 (\rho'_{1821nm})$	0.95
	UM	First Derivative (ρ')	$-0.786 - 2493.96 (\rho'_{986nm}) - 1676.08 (\rho'_{1669nm}) - 996.82 (\rho'_{1851nm}) + 3227.71 (\rho'_{1229nm}) + 1501.33 (\rho'_{842nm}) + 455.81 (\rho'_{1610nm}) + 73.62 (\rho'_{1665nm}) + 1371.01 (\rho'_{1132nm}) + 96.89 (\rho'_{1231nm}) - 340.79 (\rho'_{1491nm})$	0.80
A2	FA	First Derivative (ρ')	$0.430 - 3950.04 (\rho'_{995nm}) - 1463.14 (\rho'_{1721nm}) - 31.01 (\rho'_{368nm}) + 2226.65 (\rho'_{987nm}) + 118.42 (\rho'_{357nm}) - 647.41 (\rho'_{1700nm}) - 196.15 (\rho'_{382nm}) + 1136.86 (\rho'_{1017nm}) - 646.25 (\rho'_{1726nm}) - 486.81 (\rho'_{1606nm}) + 59.99 (\rho'_{1737nm}) - 311.95 (\rho'_{431nm})$	0.79
	HUM	First Derivative (ρ')	$0.329 - 1384.06 (\rho'_{1719nm}) + 1384.47 (\rho'_{1023nm}) + 1050.97 (\rho'_{1813nm}) + 1311.44 (\rho'_{1263nm}) + 162.64 (\rho'_{1606nm}) - 648.93 (\rho'_{628nm}) - 801.13 (\rho'_{1044nm}) + 834.88 (\rho'_{2312nm}) - 211.00 (\rho'_{371nm}) + 40.66 (\rho'_{361nm}) + 533.06 (\rho'_{987nm})$	0.84

Figures 9a and 9b show that, once again, the estimation models for C-HA and C-HUM stood out when using the soil samples from A1; while models for the carbon content of the humic substances from A2 had the best results when validating for C-FA and C-HUM (Figures 9c and 9d). Using the joint A1&A2 spectral data, it was only possible to obtain an excellent estimation model for the C-HUM content (Figure 10), this being the only model to present an RPD >2.0 and R²_{adj.} >0.8 when validating.

In turn, the predictive models where RPD was greater than 2.0 are shown in table 2, with the coefficients of regression and the respective wavelengths that most stood out for the variation in OC in each fraction of the humic substances.

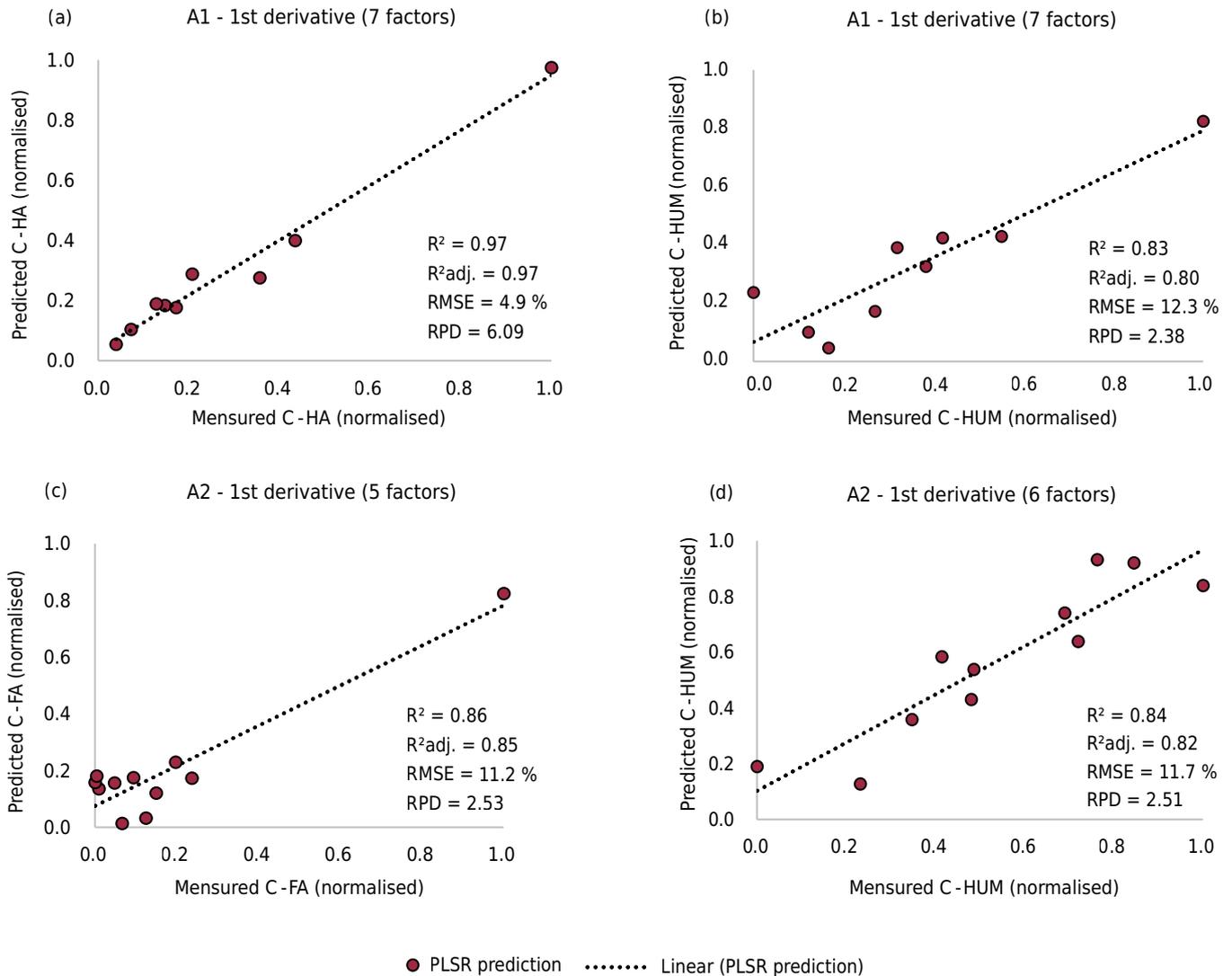


Figure 9. Validation of the best-fitting PLSR models for predicting C-HA in A1 (a), C-HUM in A1 (b), C-FA in A2 (c), and C-HUM in A2 (d).

DISCUSSION

Descriptive statistics

The predominance of C-HA in A1 suggests the selective loss of FA on the surface, as the high porosity of soils favors its mobility with coarser particles (Ebeling et al., 2011; Clemente et al., 2018). The opposite scenario is demonstrated by the samples from A2, indicating less potential for carbon loss due to the mobility of FA (Benites et al., 2003; Clemente et al., 2018).

In general, sandy soils are more likely to have a predominance of C-HA than of C-FA, while the latter is more often concentrated in soils with a high clay content (Ebeling et al., 2011). In view of the above, it is possible to corroborate the results found in the present study, in which the *Neossolos Flúvicos* from A1 present a sandy-loam to silty-clay loam texture, while, according to Ribeiro et al. (2021), the *Cambissolos* from A2 are characterized as sandy-loam to clayey soils.

In figure 5a, the descriptive statistics show a predominance in the mean value of OC in fulvic acid for samples from A2, while for samples from A1, there is a slight increase in

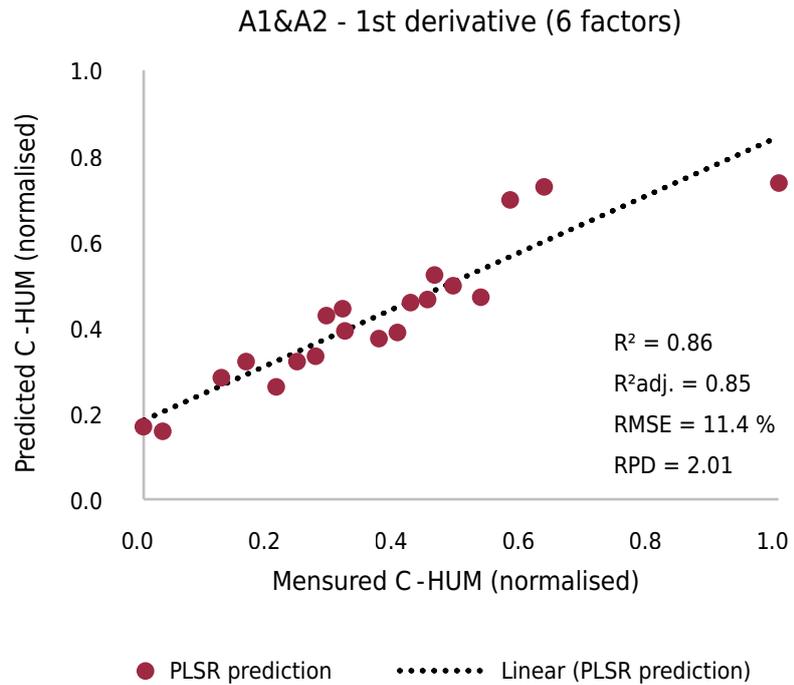


Figure 10. Validation of the best-fitting PLSR model for predicting C-HUM in the A1&A2 joint data set.

the mean value of the C-HA fraction. However, when evaluating the joint chemical results from A1&A2, no significant difference was seen between the content of either fraction, since neither predominate when the samples are evaluated without differentiating by area. This result was expected, since this dataset was built from a combination of the two previous datasets.

The results are similar to those reported by Fontana et al. (2005), who found a predominance of C-FA in *Cambissolos*, and of C-HA in *Neossolos* evaluated in Uruguay. The above authors consider that the degree of soil development can influence the dynamics of the most soluble fractions in the surface horizons. For the authors, soils with a low degree of evolution, such as *Neossolos*, have coarser material and less depth compared to *Cambissolos*, facilitating the loss of soluble organic particles on the soil surface.

The humin fraction (Figure 5b) is closely related to soil colloids, and its predominance in the samples has to do with its high molecular weight and strong interaction with clay minerals, which provides protection and stability to the OC in the fraction (Pham et al., 2021; Di Iorio et al., 2022). The humin fraction, therefore, stood out the most in terms of the distribution of the carbon data, as shown in figure 5b.

The C-HUM content showed the best mean values and smallest variations compared to C-FA and C-HA, this can be explained by 50 to 70 % of the C present in the humic substances being associated with the humin fraction (Pham et al., 2021). For this fraction, the highest median value (14.48 g kg^{-1}) and lowest variation are found in the samples from A2; this is explained by the strong interaction of the organic fraction with the iron oxides that can be found in the *Cambissolos* of the region (Mota et al., 2007; Moreira, 2013).

According to the Kolmogorov-Smirnov test at a level of 5 %, only the humin fraction showed low distortion tending towards normality, with a $p\text{-value} > 0.2$, demonstrating the heterogeneity and variation in the values of OC in the FAF and HAF for each of the sample sets under evaluation.

Table 2. Equations of the best PLSR models (RPD >2.0) for predicting organic carbon in humic substances, with the respective adjusted R²

Sample	Humic substances.	Spectral data	Best PLSR prediction models for C-HS (normalised)	R ² _{adj.}
A1	HA	First Derivative (ρ')	$0.264 - 1553.99 (\rho'_{832nm}) + 285.72 (\rho'_{2317nm}) - 298.36 (\rho'_{371nm}) + 2906.19 (\rho'_{912nm}) - 1555.22 (\rho'_{901nm}) + 1897.48 (\rho'_{843nm}) - 1763.80 (\rho'_{1223nm}) + 864.44 (\rho'_{1606nm}) - 387.49 (\rho'_{1661nm}) - 252.65 (\rho'_{2331nm}) - 1.46 (\rho'_{1821nm})$	0.97
	HUM	First Derivative (ρ')	$- 0.789 - 2460.43 (\rho'_{986nm}) - 1678.28 (\rho'_{1669nm}) - 966.92 (\rho'_{1851nm}) + 2986.89 (\rho'_{1229nm}) + 1535.40 (\rho'_{842nm}) + 532.88 (\rho'_{1610nm}) + 60.59 (\rho'_{1665nm}) + 1383.32 (\rho'_{1132nm}) + 161.16 (\rho'_{1231nm}) - 281.35 (\rho'_{1491nm})$	0.80
A2	FA	First Derivative (ρ')	$0.413 - 3502.33 (\rho'_{995nm}) - 1551.79 (\rho'_{1721nm}) - 570.99 (\rho'_{368nm}) + 1913.60 (\rho'_{987nm}) + 180.04 (\rho'_{357nm}) - 563.85 (\rho'_{1700nm}) - 232.96 (\rho'_{382nm}) + 798.81 (\rho'_{1017nm}) - 852.76 (\rho'_{1726nm}) - 111.49 (\rho'_{1606nm}) + 113.29 (\rho'_{1737nm}) - 198.84 (\rho'_{431nm})$	0.78
	HUM	First Derivative (ρ')	$0.369 - 1361.95 (\rho'_{1719nm}) + 801.05 (\rho'_{1023nm}) + 855.16 (\rho'_{1813nm}) + 1453.15 (\rho'_{1263nm}) + 389.22 (\rho'_{1606nm}) - 1002.90 (\rho'_{628nm}) - 453.46 (\rho'_{1044nm}) + 869.54 (\rho'_{2312nm}) - 253.17 (\rho'_{371nm}) - 39.20 (\rho'_{361nm}) + 940.23 (\rho'_{987nm})$	0.82
A1&A2	HUM	First Derivative (ρ')	$0.562 + 35.91 (\rho'_{1051nm}) + 591.03 (\rho'_{1608nm}) - 282.31 (\rho'_{981nm}) - 393.34 (\rho'_{966nm}) + 665.06 (\rho'_{1834nm}) + 257.55 (\rho'_{1828nm}) + 54.19 (\rho'_{1742nm}) + 287.71 (\rho'_{1853nm}) - 486.48 (\rho'_{1399nm}) - 32.39 (\rho'_{1462nm}) + 90.44 (\rho'_{388nm}) - 296.36 (\rho'_{1703nm}) - 589.75 (\rho'_{1976nm})$	0.85

Pearson correlation (r) between the OC content of humic substances and reflectance

Assuming that the closer $|r| = 1.0$, the more intense is the linear correlation between the variables, and when $|r| = 0.0$ there is no correlation (Schober et al., 2018), it was found that the Pearson correlations between the C-HS content and the reflectance of the soil samples (Figures 6a, 6b and 6c) were not good for the three sample groups under evaluation (A1, A2 and A1&A2), with each assuming an independent character in the analysis.

The reflectance of the soil samples, without spectral transformation, showed little correlation with the variation in C-FA content in the sample groups under evaluation. Despite the peaks observed in the region around 564 nm in the soil samples from A2 (Figure 6b), the correlation between the chemical and spectral variables can be considered fragile ($|r| = 0.25$). The results are consistent with studies carried out by Henderson et al. (1992), who reported that fulvic acid has no significant influence on soil reflectance.

The correlation of reflectance with the C-HA content stood out for the A2 soils (Figure 6b), showing a weak inverse relationship in the visible and near-infrared region (350-1200 nm), with $|r| = 0.32$ at 570 nm - opposite behavior to the correlation with the C-FA content of the same sample group. The results show an inverse correlation between C-FA and C-HA around 570 nm for both variables. Despite the weak correlation, the variation in OC in fulvic acid shows a direct relationship with the untransformed reflectance, while in humic acid, this relationship is reversed.

The region between 550 and 880 nm is useful for suggesting the presence of hematite in the form of free iron oxide in the soil samples (Pearlshtien and Ben-Dor, 2020). In general, features typical of hematite and goethite can be seen from absorption troughs in the visible and near-infrared spectrum (450-950 nm), with hematite most affecting the absorption around 550 nm (Demattê et al., 2015; Lin et al., 2021).

Cambissolos from the area of Limoeiro do Norte, Ceará, have a significant iron-oxide content which, due to the intense affinity with the humic substances in the soil, is capable of forming organometallic complexes (Fontana et al., 2008; Moreira, 2013). The strong bonding of these organometallic complexes may therefore be an important factor in the behavior of the correlation between C-HAF and C-FAF (Figure 6b). The reflectance factors in the samples from A2 around 570 nm corroborate the study by Ribeiro et al. (2021), which showed the influence of iron oxides on this spectrum region.

The variations in the reflectance of the soil samples from A1 and the joint data from A1&A2 showed that there was no linear relationship with the variations in the C-FA and C-HA content, as shown in figures 6a to 6c, in which the correlations between the chemical and spectral variables always remained very close to zero.

For each of the sample groups, the untransformed reflectance was inversely correlated with the C-HUM content over the entire spectrum from 350-2500 nm, despite the weak correlation; this is possibly due to the darker color of this humic fraction masking the reflectance factors of the soil samples (Ribeiro, 2021).

Predicting organic carbon in the humic substances by PCR

The PCR regression models with first-derivative spectral data performed better when predicting the OC in the predominant humic substances in each of the areas of interest. In other words, the best predictions were for C-HA and C-HUM in the samples from A1 (Figures 7a and 7b) and for C-FA and C-HUM in the samples from A2 (Figures 7c and 7d). The inefficiency of the reflectance factors as components of the predictive models for OC was expected due to the weak correlation between the reflectance factors and the variation in C-HS. The same result can be expected when using smoothed reflectance

since Savitzky-Golay smoothing results in no significant changes to the original spectrum (Ribeiro, 2021).

Using the second derivative of reflectance also did not efficiently calibrate the predictive models. This method has the benefit of ease of execution by the computer and of removing the noise that still exists after applying the first derivative (Ennes, 2008); however, it was unable to adjust the reflectance factors for use as predictive components the C-HS content.

No transformation was efficient in generating an RPD ≥ 1.4 when predicting the C-FA content in samples from A1 or the C-HA content in samples from A2, possibly due to insufficient C-HS content to give a reliable prediction. As a result, according to the Chang classification (Chang et al., 2001), the predictive models with an RPD < 1.4 could not be considered reliable for predicting C-HS.

The predictive performance for C-HA in A1 stood out for its high RPD value (4.19) and $R^2_{adj.}$ of 0.95 (Figure 7a), reducing the dimensionality of the data to 11 factors or latent variables. The predictive model for C-HUM (Figure 7b), with an RPD = 2.36, $R^2_{adj.}$ = 0.80 and RMSE = 12.4 %, used 10 latent variables. The high RPD that resulted in the reliability of the C-HA prediction, was due to the lowest RMSE seen between all of the predictive models, i.e., the prediction variation was only 7.1 % with unused data. Reliable predictive models, therefore, tend to have a higher RPD and reduced RMSE compared to models of low (RPD < 1.4) or intermediate ($1.4 < \text{RPD} < 2.0$) reliability.

The results for RPD and $R^2_{adj.}$ seen in this study are consistent with those found by Xie et al. (2021) for predicting the levels of particulate organic matter in soil samples of fluvial origin using PCR. The authors found that the first derivative of reflectance performed extremely well in building the predictive models for SOM a with a particle size of 0.15 mm exhibiting an RPD = 2.06, and $R^2 = 0.79$.

The results, therefore, show the adjustment efficiency of PCR models in predicting small organic particles in the soil using spectral data transformed into the first derivative. The transformation technique can be considered very efficient in improving the resolution of the reflectance factors, removing noise and overlapping spectra (O'Haver, 1979). It can be seen that the best models performed well only when the samples were separated by area, showing that predicting the OC content was efficient only for the most predominant humic substances.

The coefficients and respective wavelengths using the first derivative of reflectance are shown in table 1. The most significant wavelengths for predicting C-HS in A1 and A2 are located at specific points on the spectrum. The Backward method, using the spectral samples of the *Neossolos Flúvicos* from A1, selected only the 371 nm band in the visible region. This result agrees with Viscarra-Rossel et al. (2006), who found a positive influence from humic acid in the region around 400 nm, with prominent absorption peaks in the visible spectrum.

The wavelengths at 2317 and 2330 nm, considered significant for predicting C-HA using the samples from A1, and at 2312 nm for C-HUM in A2 (Table 1), are associated with the aromatic humic-acid groups (Ben-Dor et al., 1997) or the aliphatic carboxylic bonds (Workman and Weyer, 2008) related to the structure of the organic matter itself (Meissl et al., 2007). As aromatic rings are the principal component of humic acid and humin molecules, absorption peaks in the infrared spectrum are expected in regions of aromatic vibration, such as around 1600-1660 nm (Terra et al., 2013), in the same way, that significant wavelengths were found for predicting C-HA and C-HUM.

Wavelengths of 1726, 1760 and 1761 nm are influenced by the aliphatic bonds found in the structure of humic substances (Ben-Dor et al., 1997; Fidêncio et al., 2002). The Backward method selected regions close to the above wavelengths as significant for predicting C-FA. Carboxylic bonds present in the carbon biomass have been shown to influence absorption peaks in the region of 1650-1800 nm (Vaidyanathan et al., 1999), and recent authors (Cambou et al., 2021; Ribeiro et al., 2021) also reported the influence of wavelengths around 1600-1880 nm on predicting the organic carbon content of the soil (g kg^{-1}). These ranges are within those found in this study as significant for the variation in OC content in each of the humic substances under evaluation.

Therefore, the absorption peaks in the infrared region can be considered a method of characterization that provides important information about the nature of humic substances, as well as investigating the predominant functional groups in the organic extract of the soil (Tomazoni and Guimarães, 2015).

Predicting organic carbon in the humic substances by PLSR

As with PCR, the PLSR models showed better performance in predicting the C-HA and C-HUM content in the samples from A1 (Figures 9a and 9b), and the C-FA and C-HUM content in the samples from A2 (Figures 9c and 9d), using reflectance factors transformed into the first derivative.

The models built to predict the C-FA content using the spectral data of the samples from A1, showed an RPD < 1.4 for all of the spectral treatments under analysis, characterizing them as inefficient for predicting the carbon in the fraction. Among these, in the best model, albeit inefficient as a predictor, the spectrum transformed into the first derivative of reflectance was also used, suggesting the efficiency of this transformation in improving the prediction, with a lower mean square error (RMSE = 8.2 %). Transforming the reflectance factors between 350 and 2500 nm leads to a better illustration of the variations in absorbance relative to the spectral bands, which demonstrates the potential relationship between the chemical variables of the soil and the first-order derivative of reflectance (Bou-Orm et al., 2020).

For predicting C-HA and C-HUM using the spectral samples from A1 shown in figures 9a and 9b, the best performances had an $R^2_{\text{adj}} = 0.97$ and 0.80 , and RPD = 6.09 and 2.38, respectively, classifying both models as excellent predictors (Chang et al., 2001). The PLSR was able to reduce the dimensionality of the spectral data from A1 to only seven latent variables for each predictive model.

The predictive model for C-FA in the soil samples from A2, had an $R^2_{\text{adj}} = 0.85$, RPD = 2.53 and RMSE = 11.2 %, and reduced the dimensionality of the data to six latent variables (Figure 9c). The results show the efficiency of using spectral data to predict the level of humic substances, as do those of Vergnoux et al. (2009), who obtained predictions for fulvic acid with an $R^2 = 0.98$ and error of 7.9 % when validating PLSR models for the 1000-2500 nm region of the spectrum.

Using samples of the *Cambissolos* from A2, the prediction was also efficient for the C-HUM content (Figure 9d), generating a predictive model of five latent variables and an $R^2_{\text{adj}} = 0.82$, RMSE = 11.7 % and RPD = 2.51. The prediction of OC in the humin fraction was the only one that resulted from reliable models (RPD > 2.0) using PLSR for each of the sample groups evaluated in this study. With PLSR, it was also possible to predict the C-HUM content efficiently, with an RPD = 2.01 when using the first derivative on the samples from a set of A1&A2 data (Figure 10), which ignores the spectral and chemical differences between them.

Unlike PCR, it was possible to train PLSR predictive models for the humin fraction in the three sample groups, demonstrating the efficiency of the technique in calibrating models of moderate reliability. When using PCR to calibrate the predictive model for C-HUM using the first derivative of reflectance with A1&A2, the RPD was 1.41 (Figure 8), which reflects the need for alternative ways of adjusting and improving the prediction efficiency (Chang et al., 2001). The PLSR, in turn, generated an RPD = 2.01 for the same conditions (Figure 10); it can therefore be inferred that the improvement in prediction efficiency is a result of the change in the multivariate technique for calibrating the models.

The PCR is considered the simplest and most easily interpreted technique; however, the results are still less robust than those obtained with PLSR. In the present study, PCR showed prediction efficiency for the same humic substances as did PLSR, since the models presented similar responses, differing only in the values of the coefficients of regression. Similar results when comparing the two statistical methods are also seen in the literature, in studies by Mouazen et al. (2010), Shiferaw and Hergarten (2014) and Xie et al. (2021).

The best C-HS predictions resulted from the reflectance factors' spectral transformation into the first derivative, showing the potential of using the infrared spectrum for soil samples. It can therefore be suggested that regions of the infrared spectrum are effective for evaluating organic and inorganic properties of the soil, since they show the influence of functional groups and molecular vibrations that may be associated with humic substances (Janik et al., 1998).

In this study, the best performance of PLSR was the same as for PCR: for C-HA and C-HUM in the *Neossolos Flúvicos* of A1, and C-FA and C-HUM in the *Cambissolos* of A2. When using the samples without differentiating between the soil types, such as with A1&A2, the best performance was when predicting C-HUM.

Humin fraction is strongly associated with clay minerals (Pham et al., 2021); therefore, the wavelengths selected to predict C-HUM, around 1400 and 1900 nm, using A1&A2 (Table 2) may be related to water molecules in the clay minerals that are associated with the organic fraction in the samples, regardless of the class of soil.

Wavelengths at 966 and 981 nm were selected by the Backward method for estimating C-HUM using the joint samples from A1&A2; in the literature, this region of the spectrum is portrayed as typically influenced by iron oxides in the samples, mixed with the organic components (Vasques et al., 2009; Viscarra-Rossel and Behrens, 2010). Wavelengths between 1700 and 1880 nm were selected by all of the predictive models (Table 2), showing that this region of the spectrum is able to indicate vibrations related to the molecular structures that are associated with organic carbon (Fidêncio et al., 2002; Stenberg et al., 2010). Such results are expected, since the humic fractions are responsible for storing part of the soil OC in their structure (Santos et al., 2013).

As both models have the same predictions, PLSR generated predictive models with the same significant wavelengths as PCR. Table 2, therefore, also shows the spectral bands around 1600 and 1660 nm when estimating C-HA and C-HUM in the soils of A1 and predicting C-FA and C-HUM in the soils of A2, using the first derivative of reflectance. These spectral regions are characteristic of aromatic structures in the skeleton of the molecules that make up the humic substances. In turn, those in the 1700-1800 nm region, are mainly influenced by the aliphatic structures of the carboxylic groups (Table 2) (Ben-Dor et al., 1997; Ribeiro et al., 2021).

Observing spectral wavelengths in the VNIR-SWIR region to predict the levels of humic substances is not a technique widely found in the international literature but has proved to be efficient in predicting organic carbon using the techniques discussed in this study. In

this respect, it should be noted that the authors of this study intend to carry out further evaluations to consolidate their results.

CONCLUSION

For each set of soil samples, the untransformed reflectance showed a weak correlation with the variation in the organic carbon content of the humic substances. The PLSR, together with the spectral data transformed into the first derivative of reflectance, produced more robust predictive models than PCR for the organic carbon content associated with the humic acid in *Neossolos Flúvicos*, fulvic acid in *Cambissolos*, and humin in both of these classes. Regression models from reflectance spectroscopy showed the 1600-1800 nm region as significant for observing organic structures in soil samples.

ACKNOWLEDGMENTS

The authors would like to thank all the members of the Geoprocessing, Automation and Agricultural Management (GAMA) research group who helped in the methodological process in the laboratory and, especially, the Council for Scientific and Technological Development (CNPq), CAPES/PROAP and the National Institute of Science in Salinity Technology (INCTSal) for their support while conducting this study.

APPENDIX A. SUPPLEMENTARY DATA

Supplementary data to this article can be found online at https://www.rbcjournal.org/wp-content/uploads/articles_xml/1806-9657-rbcs-47-e0220143/1806-9657-rbcs-47-e0220143-suppl01.pdf.

AUTHOR CONTRIBUTIONS

Conceptualization:  Adunias dos Santos Teixeira (equal),  Mirian Cristina Gomes Costa (supporting) and  Sharon Gomes Ribeiro (lead).

Formal analysis:  Adunias dos Santos Teixeira (equal),  Isabel Cristina da Silva Araújo (equal),  Luis Clenio Jario Moreira (equal),  Marcio Regys Rabelo de Oliveira (equal),  Mirian Cristina Gomes Costa (equal) and  Raul Shiso Toma (equal).

Investigation:  Letícia Machado Lopes (equal),  Marcio Regys Rabelo de Oliveira (equal) and  Sharon Gomes Ribeiro (equal).

Methodology:  Isabel Cristina da Silva Araújo (supporting),  Letícia Machado Lopes (supporting),  Marcio Regys Rabelo de Oliveira (supporting) and  Sharon Gomes Ribeiro (lead).

Project administration:  Adunias dos Santos Teixeira (lead).

Resources:  Isabel Cristina da Silva Araújo (equal) and  Mirian Cristina Gomes Costa (equal).

Supervision:  Isabel Cristina da Silva Araújo (equal) and  Raul Shiso Toma (equal).

Visualization:  Mirian Cristina Gomes Costa (lead).

Writing - original draft:  Sharon Gomes Ribeiro (lead).

Writing - review & editing:  Adunias dos Santos Teixeira (equal),  Letícia Machado Lopes (equal),  Luis Clenio Jario Moreira (equal),  Marcio Regys Rabelo de Oliveira

(equal),  Mirian Cristina Gomes Costa (equal),  Raul Shiso Toma (equal) and  Sharon Gomes Ribeiro (equal).

REFERENCES

- Ben-Dor E, Inbar Y, Chen Y. The reflectance spectra of organic matter in the visible near-infrared and short-wave infrared region (400-2500 nm) during a controlled decomposition process. *Remote Sens Environ.* 1997;61:1-15. [https://doi.org/10.1016/S0034-4257\(96\)00120-4](https://doi.org/10.1016/S0034-4257(96)00120-4)
- Benites VM, Madari B, Machado PLOA. Extração e fracionamento quantitativo de substâncias húmicas do solo: um procedimento simplificado de baixo custo. Rio de Janeiro: Embrapa Solos; 2003.
- Bou-Orm N, Alromaiti AA, Elrmeithi M, Mohammad A, Nazzal Y, Howari FM, Aydaros FA. Advantages of first-derivative reflectance spectroscopy in the VNIR-SWIR for the quantification of olivine and hematite. *Planet Space Sci.* 2020;188:104957. <https://doi.org/10.1016/j.pss.2020.104957>
- Cambou A, Allory V, Cardinael R, Vieira LC, Barthes BG. Comparison of soil organic carbon stocks predicted using visible and near infrared reflectance (VNIR) spectra acquired *in situ* vs. on sieved dried samples: Synthesis of different studies. *Soil Security.* 2021;5:100024. <https://doi.org/10.1016/j.soisec.2021.100024>
- Canellas LP, Rumjanek VM. Espectroscopia na região do infravermelho. In: Canellas LP, Santos GA, editors. *Humosfera: Tratado preliminar sobre a química das substâncias húmicas*. Campos dos Goytacazes: Universidade Estadual do Norte Fluminense; 2005. p. 201-23.
- Chang CW, Laird DA, Mausbach MJ, Hurburgh JCR. Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties. *Soil Sci Soc Am J.* 2001;65:480-90. <https://doi.org/10.2136/sssaj2001.652480x>
- Clemente EDP, Oliveira FS, Machado MR, Schaefer CEGR. Fracionamento da matéria orgânica dos solos da Ilha da Trindade. *Rev Dep Geo.* 2018;36:48-62. <https://doi.org/10.11606/rdg.v36i0.147796>
- Cunha TJF, Petrere VG, Silva DJ, Mendes AMS, Melo RF; Oliveira Neto MB, Silva MSL, Alvarez IA. Principais solos do semiárido tropical brasileiro: caracterização, potencialidades, limitações, fertilidade e manejo. In: Sa IB, Silva PCG, editors. *Semiárido brasileiro: Pesquisa, desenvolvimento e inovação*. Petrolina: Embrapa Semiárido; 2010. p. 50-87.
- Demattê JAM, Araujo SR, Fiorio PR, Fongaro CT, Nanni MR. Espectroscopia VIS-NIR-SWIR na avaliação de solos ao longo de uma topossequência em Piracicaba/SP. *Rev Cienc Agron.* 2015;46:679-88. <https://doi.org/10.5935/1806-6690.20150054>
- Di Iorio E, Circelli L, Angelico R, Torrent J, Tan W, Colombo C. Environmental implications of interaction between humic substances and iron oxide nanoparticles: A review. *Chemosphere.* 2022;303:135172. <https://doi.org/10.1016/j.chemosphere.2022.135172>
- Ebeling AG, Anjos LHC, Pereira MG, Pinheiro EFM, Valladares GS. Substâncias húmicas e relação com atributos edáficos. *Bragantia.* 2011;70:157-65. <https://doi.org/10.1590/S0006-87052011000100022>
- Ebeling AG, Anjos LHCD, Pereira MG, Valladares GS, Pérez DV. Substâncias húmicas e suas relações com o grau de subsidência em Organossolos de diferentes ambientes de formação no Brasil. *Rev Cienc Agron.* 2013;44:225-33. <https://doi.org/10.1590/S1806-66902013000200003>
- Ennes R. Potencial das imagens hiperespectrais orbitais na detecção de componentes opticamente ativos no reservatório de Itupararanga [thesis]. Presidente Prudente: Universidade Estadual Paulista; 2008.
- Fidêncio PH, Poppi RJ, Andrade JC, Cantarella H. Determination of organic matter in soil using near-infrared spectroscopy and partial least squares regression. *Commun Soil Sci Plant Anal.* 2002;33:1607-15. <https://doi.org/10.1081/CSS-120004302>
- Fontana A, Anjos LHC, Sallés JM, Pereira MC, Rossiello ROP. Carbono orgânico e fracionamento químico da matéria orgânica em solos da Sierra de Ánimas - Uruguai. *Flor@m.* 2005;12:36-43.

- Fontana A, Benites VM, Pereira MG, Anjos LHC. Substâncias húmicas como suporte à classificação de solos brasileiros. *Rev Bras Cienc Solo*. 2008;32:2073-80. <https://doi.org/10.1590/S0100-06832008000500028>
- Gomez C, Chevallier T, Moulin P, Bouferra I, Hmaid K, Arrouays D, Jolivet C, Barthès BG. Prediction of soil organic and inorganic carbon concentrations in Tunisian samples by mid-infrared reflectance spectroscopy using a French national library. *Geoderma*. 2020;375:114469. <https://doi.org/10.1016/j.geoderma.2020.114469>
- Granlund L, Keinanen M, Tahvanainen T. Identification of peat type and humification by laboratory VNIR/SWIR hyperspectral imaging of peat profiles with focus on fen-bog transition in aapa mires. *Plant Soil*. 2021;460:667-86. <https://doi.org/10.1007/s11104-020-04775-y>
- Henderson TL, Baumgardner MF, Franzmeier DE, Stott DE, Coster DC. High dimensional reflectance analysis of soil organic matter. *Soil Sci Soc Am J*. 1992;56:865-72. <https://doi.org/10.2136/sssaj1992.03615995005600030031x>
- Jacomine PKT, Almeida JC, Medeiros LAR. Levantamento exploratório: Reconhecimento de solos do estado do Ceará. Recife: SUDENE-DRN; 1973 [cited 2021 Sep 15]. Available from: <https://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/331170>.
- Janik LJ, Merry RH, Skjemstad JO. Can mid infrared diffuse reflectance analysis replace soil extractions? *Aust J Exp Agric*. 1998;38:681-96. <https://doi.org/10.1071/EA97144>
- Jiménez-González MA, Álvares AM, Carral P, Almendros G. Chemometric assessment of soil organic matter storage and quality from humic acid infrared spectra. *Sci Total Environ*. 2019;685:1160-8. <https://doi.org/10.1016/j.scitotenv.2019.06.231>
- Kumada K. Chemistry of soil organic matter. Amsterdam: Elsevier; 1987.
- Liu J, Han J, Xie J, Wang H, Tong W, Ba Y. Assessing heavy metal concentrations in earth-cumulo-orthic anthrosols soils using Vis-NIR spectroscopy transform coupled with chemometrics. *Spectrochim Acta A Mol Biomol Spectrosc*. 2020;226:117639. <https://doi.org/10.1016/j.saa.2019.117639>
- Lin Z, Natoli JM, Picuri JC, Shaw SE, Bowyer WJ. Replication of the conversion of goethite to hematite to make pigments in both furnace and campfire. *J Archaeol Sci Rep*. 2021;39:103134. <https://doi.org/10.1016/j.jasrep.2021.103134>
- Madhavan DB, Baldock JA, Read ZJ, Murphy SC, Cunningham SC, Perring MP, Herrmann T, Lewis T, Cavagnaro TR, England JR, Paul KI, Weston CJ, Baker TG. Rapid prediction of particulate, humus and resistant fractions of soil organic carbon in reforested lands using infrared spectroscopy. *J Environ Manage*. 2017;193:290-9. <https://doi.org/10.1016/j.jenvman.2017.02.013>
- Meissl K, Smidt E, Schwanninger M. Prediction of humic acid content and respiration activity of biogenic waste by means of Fourier transform infrared (FTIR) spectra and partial least squares regression (PLS-R) models. *Talanta*. 2007;72:791-9. <https://doi.org/10.1016/j.talanta.2006.12.005>
- Moreira LJS. Caracterização de solos, concreções e nódulos ferruginosos em uma topossequência na Chapada do Apodi - CE [thesis]. Fortaleza: Universidade Federal do Ceará; 2013.
- Mota JCA, Assis Junior RN, Amaro Filho J, Romero RE, Mota FOB, Libardi PL. Atributos mineralógicos de três solos explorados com a cultura do melão na Chapada do Apodi - RN. *Rev Bras Cienc Solo*. 2007;31:445-54. <https://doi.org/10.1590/S0100-06832007000300004>
- Mouazen AM, Kuang B, Baerdemaeker J, Ramon H. Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy. *Geoderma*. 2010;158:23-31. <https://doi.org/10.1016/j.geoderma.2010.03>
- O'Haver TC. Derivative and wavelength modulation spectrometry. *Anal Chem*. 1979;51:91A-9A. <https://doi.org/10.1021/ac50037a008>

- Pearlshtien DH, Ben-Dor E. Effect of organic matter content on the spectral signature of iron oxides across the VIS-NIR spectral region in artificial mixtures: An example from a red soil from Israel. *Remote Sens.* 2020;12:1960. <https://doi.org/10.3390/rs12121960>
- Pham DM, Kasai T, Yamaura M, Katayama A. Humin: No longer inactive natural organic matter. *Chemosphere.* 2021;269:128697. <https://doi.org/10.1016/j.chemosphere.2020.128697>
- Primo DC, Menezes RSC, Silva TO. Substâncias húmicas da matéria orgânica do solo: Uma revisão de técnicas analíticas e estudos no nordeste brasileiro. *Sci Plena.* 2011;7:059901.
- Pudelko A, Chodak M. Estimation of total nitrogen and organic carbon contents in mine soils with NIR reflectance spectroscopy and various chemometric methods. *Geoderma.* 2020;368:114306. <https://doi.org/10.1016/j.geoderma.2020.114306>
- Raiesi F. The quantity and quality of soil organic matter and humic substances following dry-farming and subsequent restoration in an upland pasture. *Catena.* 2021;202:105249. <https://doi.org/10.1016/j.catena.2021.105249>
- Ribeiro SG. Espectroscopia de reflectância na avaliação do carbono orgânico em solos do semiárido [thesis]. Fortaleza: Universidade Federal do Ceará; 2021.
- Ribeiro SG, Teixeira AS, Oliveira MRR, Costa MCG, Araújo ICS, Moreira LCJ, Lopes FB. Soil organic carbon content prediction using soil-reflected spectra: A comparison of two regression methods. *Remote Sens.* 2021;13:4752. <https://doi.org/10.3390/rs13234752>
- Rudorff CM, Novo EMLM, Galvão LS, Pereira Filho W. Análise derivativa de dados hiperespectrais medidos em nível de campo e orbital para caracterizar a composição de águas opticamente complexas na Amazônia. *Acta Amaz.* 2007;37:269-80. <https://doi.org/10.1590/S0044-59672007000200014>
- Santos LL, Lacerda JJJ, Zinn YL. Partição de substâncias húmicas em solos brasileiros. *Rev Bras Cienc do Solo.* 2013;37:955-68. <https://doi.org/10.1590/S0100-06832013000400013>
- Savitzky A, Golay MJE. Smoothing and differentiation of data by simplified least squares procedures. *Anal Chem.* 1964;36:1627-39. <https://doi.org/10.1021/ac60214a047>
- Schnitzer M. Humic substances: Chemistry and reactions. In: Schnitzer M, Khan SU, editors. *Soil organic matter*. New York: Elsevier; 1978. p. 1-64.
- Schober P, Boer C, Schwarte LA. Correlation coefficients: Appropriate use and interpretation. *Anesth Analg.* 2018;126:1763-8. <https://doi.org/10.1213/ANE.0000000000002864>
- Shiferaw A, Hergarten C. Visible near infra-red (VisNIR) spectroscopy for predicting soil organic carbon in Ethiopia. *J Ecol Nat Environ.* 2014;6: 26-39. <https://doi.org/10.5897/JENE2013.0374>
- Stenberg B, Viscarra-Rossel RA, Mouazen A M, Wetterlind J. Visible and near-infrared spectroscopy in soil science. *Adv Agron.* 2010;107:163-215. [https://doi.org/10.1016/S0065-2113\(10\)07005-7](https://doi.org/10.1016/S0065-2113(10)07005-7)
- Swift RS. Organic matter characterization. In: Sparks DL, Page AL, Helmke PA, editors. *Methods of soil analysis: Part 3 - Chemical methods*. Madison: Soil Science Society of America; 1996. p. 1011-69.
- Terra FDS, Demattê JAM, Viscarra-Rossel R. Discriminação de solos baseada em espectroscopia de reflectância VIS-NIR. In: XVI Simpósio Brasileiro de Sensoriamento Remoto; 13-18 abril 2013; Iguazu, Brasil. Iguazu: Instituto Nacional de Pesquisas Espaciais; 2013. p. 9224-32.
- Tomazoni JC, Guimarães E. Características espectrais das frações húmica e ácido húmico da matéria orgânica total dos solos da bacia do rio Passo da Pedra. *Rev Bras Geogr Fis.* 2015;8:721-35. <https://doi.org/10.5935/1984-2295.20150027>
- Vaidyanathan S, Mcneil B, Macaloney G. Fundamental investigations on the near-infrared spectra of microbial biomass as applicable to bioprocess monitoring. *The Analyst.* 1999;124:157-62. <https://doi.org/10.1039/A806847J>
- Vasques GM, Grunwald S, Sickman JO. Modeling of soil organic carbon fractions using visible-near-infrared spectroscopy. *Soil Sci Soc Am J.* 2009;73:176-84. <https://doi.org/10.2136/sssaj2008.0015>

- Vergnoux A, Guiliano M, Le Dreau Y, Kister J, Dupuy N, Doumenq P. Monitoring of the evolution of an industrial compost and prediction of some compost properties by NIR spectroscopy. *Sci Total Environ.* 2009;407:2390-403. <https://doi.org/10.1016/j.scitotenv.2008.12.033>
- Viscarra-Rossel RA, Behrens T. Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma.* 2010;158:46-54. <https://doi.org/10.1016/j.geoderma.2009.12.025>
- Viscarra-Rossel RA, Walvoort DJJ, Mcbratney AB, Janik LJ, Skjemstad JO. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma.* 2006;131:59-75. <https://doi.org/10.1016/j.geoderma.2005.03.007>
- Workman JJ, Weyer L. Practical guide to interpretative nearinfrared spectroscopy. Boca Raton: CRC Press; 2008. <https://doi.org/10.1201/9781420018318>
- Xiaoju N, Tongqian Z, Yanyan S. Fossil fuel carbon contamination impacts soil organic carbon estimation in cropland. *Catena.* 2021;196:104889. <https://doi.org/10.1016/j.catena.2020.104889>
- Xie S, Li Y, Wang X, Liu Z, Ma K, Ding L. Research on estimation models of the spectral characteristics of soil organic matter based on the soil particle size. *Spectrochim Acta A Mol Biomol Spectrosc.* 2021;260:119963. <https://doi.org/10.1016/j.saa.2021.119963>
- Yeomans JC, Bremner JM. A rapid and precise method for routine determination of organic carbon in soil. *Commun Soil Sci Plant Anal.* 1988;19:1467-76. <https://doi.org/10.1080/00103628809368027>
- Zhang Z, Ding J, Zhu C, Wang J, Ma G, Ge X, Li Z, Han L. Strategies for the efficient estimation of soil organic matter in salt-affected soils through Vis-NIR spectroscopy: Optimal band combination algorithm and spectral degradation. *Geoderma.* 2021;382:114729. <https://doi.org/10.1016/j.geoderma.2020.114729>