

# Estimating texture and organic carbon of an Oxisol by near infrared spectroscopy<sup>1</sup>

## Estimación de textura y carbono orgánico de un Oxisol mediante espectroscopia de infrarrojo cercano

Felipe Fernández-Martínez<sup>2\*</sup>, Jesús Hernán Camacho-Tamayo<sup>3</sup>, Yolanda Rubiano-Sanabria<sup>4</sup>

**ABSTRACT** - Laboratory analyses are a fundamental basis for monitoring soil behavior. These analyses are usually tedious and expensive depending on the methodology used, which may limit data acquisition. The aim of this research was to evaluate the potential of Near Infrared (NIR) diffuse reflectance spectroscopy for the estimation of texture and Soil Organic Carbon (SOC) of an Oxisol. A total of 313 samples were collected at fixed depths of 0.0-0.10, 0.10-0.20, 0.20-0.30, 0.30-0.40 and 0.40-0.50 m in 70 points distributed in 248 ha, from which SOC and the fractions of sand, silt and clay were determined. The spectral signatures were obtained from a NIRFlex sensor, and the modeling was done applying partial least squares regression. A highly representative model was obtained for the SOC estimation, with a coefficient of determination ( $R^2$ ) of 0.97, Root Mean Square Error (RMSE) of 1.10 g kg<sup>-1</sup> and Residual Prediction Deviation (RPD) of 5.63. For the textural fractions, estimation models of lesser performance were obtained, with  $R^2$  values of 0.62; 0.44 and 0.62, RMSE values of 1.10%, 2.92% and 3.08%, and RPD values of 1.82, 1.61 and 1.81 for sand, silt and clay, respectively. By means of geostatistical interpolation surfaces, the behavior of the measured and spectrally estimated variables was compared. NIR spectroscopy proved to be a viable alternative for the precise estimation of SOC, while for the textural fractions it is convenient to explore the improvement of the estimates.

**Key words:** NIR spectroscopy. Pedometrics. Spectral modelling. Geostatistics. Chemometrics.

**RESUMEN** - Los análisis de laboratorio de suelos son una base fundamental para monitorear su comportamiento. Estos análisis suelen ser tediosos y de costo elevado según la metodología empleada, lo que puede limitar la obtención de esta información. El objetivo de esta investigación fue evaluar el potencial de la espectroscopía de reflectancia difusa de Infrarrojo Cercano (NIR) en la estimación de la textura y el Carbono Orgánico del Suelo (COS) de un Oxisol. Se recolectaron en total 313 muestras a profundidades fijas de 0.0-0.10, 0.10-0.20, 0.20-0.30, 0.30-0.40 y 0.40-0.50 m en 70 puntos distribuidos en 248 ha, y les fueron determinadas el COS, la fracción de arena, limo y arcilla. Con un sensor NIRFlex se obtuvieron las firmas espectrales, y la modelación se efectuó aplicando regresión por mínimos cuadrados parciales. Se obtuvo un modelo de alta representatividad para la estimación del SOC, con coeficiente de determinación ( $R^2$ ) de 0.97, Raíz del Error Cuadrático Medio (RMSE) de 1.10 g kg<sup>-1</sup> y Desviación Residual de la Predicción (DRP) de 5.63. Para las fracciones texturales se obtuvieron modelos de estimación de menor desempeño, con  $R^2$  de 0.62, 0.44 y 0.62, RMSE de 1.10%, 2.92% y 3.08%, y DRP de 1.82, 1.61 y 1.81 para arena, limo y arcilla respectivamente. Por medio de superficies de interpolación geoestadística se comparó el comportamiento de las variables medidas y estimadas espectralmente. La espectroscopía NIR mostró ser una alternativa viable para la estimación precisa del COS, mientras que para las fracciones texturales es conveniente explorar el mejoramiento de las estimaciones.

**Palabras clave:** Espectroscopía NIR. Pedometría. Modelo espectral. Geoestadística. Quimiometría.

DOI: 10.5935/1806-6690.20220020

Editor-in-Article: Prof. Adriel Ferreira da Fonseca - adrielff@gmail.com

\*Author for correspondence

Received for publication on 19/08/2021; approved on 23/09/2021

<sup>1</sup>Research project from the Bogotá headquarters Research Division of Universidad Nacional de Colombia

<sup>2</sup>Graduate Research Assistant, Civil and Environmental Engineering Department & PhD student at Mississippi State University, United States, ffernandezm@unal.edu.co (ORCID ID 0000-0003-4251-8820)

<sup>3</sup>Associate Professor at the School of Engineering of Universidad Nacional de Colombia, Bogotá Headquarters, Colombia, jhcamachot@unal.edu.co (ORCID ID 0000-0002-7066-369X)

<sup>4</sup>Associate Professor at the School of Agricultural Sciences of Universidad Nacional de Colombia, Bogotá Headquarters, Colombia, yrubianos@unal.edu.co (ORCID ID 0000-0002-5558-4504)

## INTRODUCTION

Soil is a non-renewable resource, whose knowledge and management are key in agricultural production. For this, laboratory analyses are required, which are usually demanding in time and resources (VENDRAME *et al.*, 2012; NOCITA *et al.*, 2014). Therefore, it is necessary to apply techniques that reduce processing times and costs.

Soil texture influences soil physical and chemical properties (JOVIĆ *et al.*, 2019), and hydrodynamic processes such as drainage and water retention (POGGIO; GIMONA, 2017). Its determination, in addition to serving as a factor for the classification of soils, can serve for decision-making for production and conservation processes.

Soils are the most important reserve of organic carbon in terrestrial ecosystems, small changes in this can significantly alter the concentration of CO<sub>2</sub> in the atmosphere and, therefore, contribute to the phenomenon of global warming (TEKIN; TUMSAVAS; MOUAZEN, 2012; ANGELOPOULOU *et al.*, 2019). Soil Organic Carbon (SOC) is a measurable component of soil organic matter, which is a key attribute to assess soil quality, since it has a great influence on its physical, chemical and biological properties, and is also related with macronutrients, such as nitrogen and phosphorus, which are determinants in terms of fertility (LAAMRANI *et al.*, 2019).

NIR spectroscopy is a technique used for prediction of the physical and chemical properties of various materials, and has shown great potential for application in soil science and the agricultural sector, given the quality of the estimates obtained for soil properties and the lower amount of resources used compared to other conventional laboratory analyses (SUMMERS *et al.*, 2011; VISCARRA ROSSEL *et al.*, 2016; NAWAR; MOUAZEN, 2019). Due to their sensitivity to wavelengths in the NIR range of the electromagnetic spectrum, certain soil organic and inorganic components can be evaluated qualitatively and quantitatively using spectral analysis (HOBLEY; PRATER, 2019). For texture and SOC, several spectral modeling studies have been reported for the estimation of properties (VISCARRA ROSSEL *et al.*, 2016; CAMACHO-TAMAYO *et al.*, 2017; LASHYA *et al.*, 2018; POPPIEL *et al.*, 2018; LIU *et al.*, 2019), in which factors such as the sensor that acquires the spectra (in the field or laboratory), the pretreatments applied to the spectral signatures, the multivariate analysis techniques used with the reference data and the robustness of the soil property databases are used to calibrate the models.

However, for equatorial oxisols, few studies have been carried out on the application of NIR spectroscopy for soil analysis (CAMACHO-TAMAYO; RUBIANO; HURTADO, 2014; CAMACHO-TAMAYO *et al.*, 2017). Given the agricultural expansion that these soils are

experiencing in Colombia, together with the need to adopt technologies that allow the monitoring of soil properties in a rapid way in the context of climate change (ANGELOPOULOU *et al.*, 2019), this research was proposed with the objective of evaluating prediction models made from NIR spectroscopy in estimating the texture (sand, silt and clay fractions) and the SOC of an equatorial Oxisol. The standard laboratory reference methods applied were: 1) the pipette for the texture and 2) the elemental analyzer for the SOC.

## MATERIAL AND METHODS

### Study area

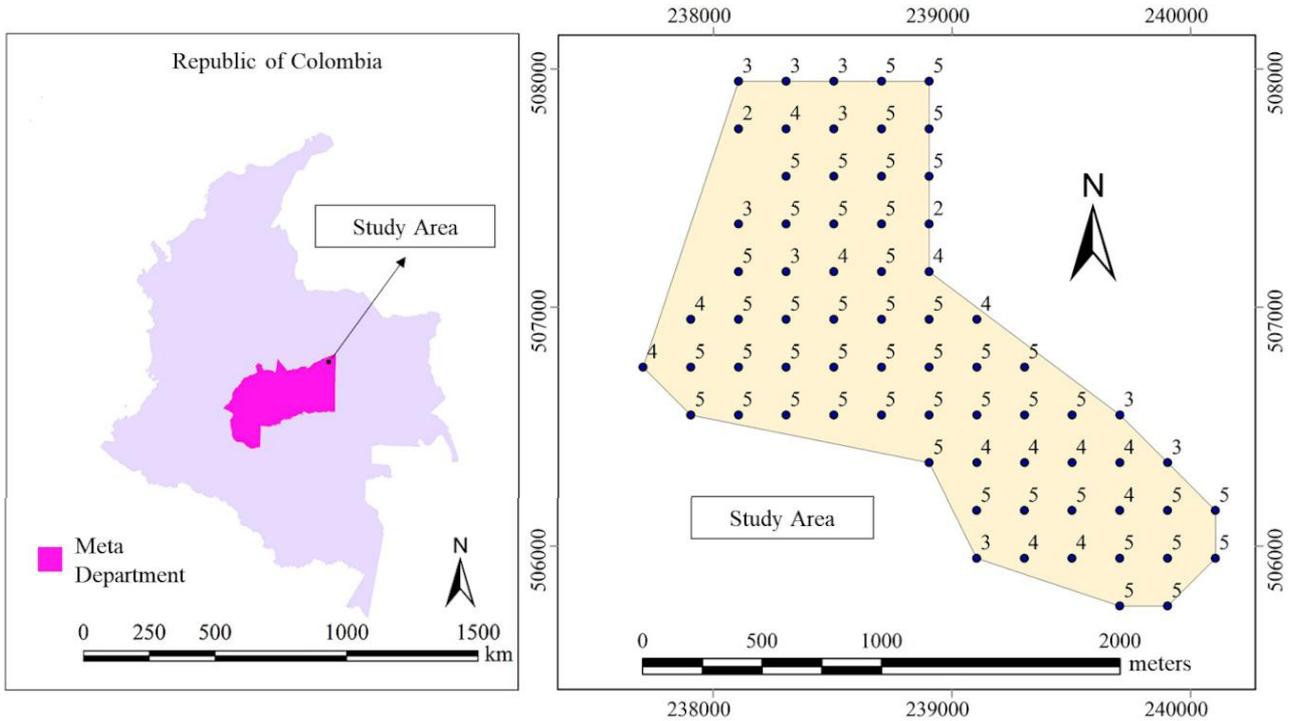
The study area is located in the Carimagua Research Center of the Colombian Agricultural Research Corporation - AGROSAVIA, located in the municipality of Puerto Gaitán, department of Meta, Colombia. It is located at the geographic coordinates of 4° 34'01.6"N 71° 19'58.0"W and has an area of 248 ha, an average altitude of 175 meters above sea level, average temperature of 28 °C and average annual rainfall of 2339 mm. The relief of the area is flat to slightly undulating, with slopes ranging from 0 to 7%. The soils in the study area are classified as Typic Hapludox, characterized by their low pH values and organic matter (CAMACHO-TAMAYO; RUBIANO; HURTADO, 2014).

### Sampling and laboratory analysis

The sampling was carried out by means of a rigid network separated perpendicularly every 200 m in the month of July 2018. It consisted of 70 cells, in which disturbed soil samples were extracted at depth ranges of 0.0-0.10, 0.10-0.20, 0.20-0.30, 0.30-0.40 and 0.40-0.50 m, that is, 5 samples for each trunk. In total, 313 soil samples were collected, of which 70 correspond to the first depth, 70 to the second, 68 to the third, 59 to the fourth and 46 to the last depth. The difference in samples by depth is due to the occurrence of the water table in some positions. Figure 1 shows the detail of the sampling, indicating the number of samples per point.

The soil samples were oven-dried at a temperature of 35 °C for 48 hours, and sieved through a 2 mm mesh. Soil texture was determined by sieving using the pipette method according to the procedure proposed by DAY (1965), applying sodium hexametaphosphate as a dispersing agent. In this procedure, the sand and clay fractions are calculated by sieving, and the silt fraction by the difference with the total mass of soil. Since the soils of the study area did not show the presence of carbonates, the determination of SOC was carried out using an Elemental Analyzer (TruSpec CN Carbon Nitrogen Determinator,

**Figure 1** - Location of the study area with the number of samples collected per sampling point



LECO co., St. Joseph, Mi, USA). The spectral signatures were obtained from a NIRFlex N-500 sensor (BÜCHI Labortechnik AG), which has a resolution of 8 cm<sup>-1</sup> and collects 1500 spectral response readings in the entire near infrared range (from 1000 to 2500 nm). Between 60 and 70 g of sieved soil sample was placed in Petri dishes 10 cm in diameter by 2 cm high. The scanned area of the Petri dish corresponded to a diameter of 9 cm, of which the sensor recorded a mean value of 32 reflectance scans for each wavelength value.

**Spectral analysis**

Prior to the application of Partial Least Squares Regression (PLSR) for the generation of models, the total of the samples was divided into two groups, a calibration group and a validation group. These groups corresponded to 70% and 30% of the total samples, that is, 219 for the calibration group and 94 samples for the external validation group. The division of these groups was carried out using the Kennard-Stone algorithm (KENNARD; STONE, 1969), which allowed the selection of the samples for each group in a uniform way within the predictor space, in such a way that they represented the total variability of the data set. The procedure followed by this algorithm is that for a calibration group of n samples ( $X_{CAL} = \{x_{CALj}\}_{j=1}^n$ ) to be selected from the total group of samples N ( $X = \{x_i\}_{i=1}^N$ , where  $n < N$ ):

The sample that best represents the mean in the predictor space is found in X, removing it from X and placing it in  $X_{CAL}$ . This sample will be named  $x_{CAL1}$ .

The sample of X that is most dissimilar to  $x_{CAL1}$  is found, it is removed from X and placed in  $X_{CAL}$ . This sample will be named  $x_{CAL2}$ .

The sample of X is found that is most dissimilar to those that are already part of  $X_{CAL}$ , it is removed from X and located in  $X_{CAL}$ .

Step 3 will be repeated n-4 times to complete the selection of  $X_{CAL}$  ( $x_{CAL4}, \dots, x_{CALn}$ ).

It is important to note that the dissimilarity between  $X_{CAL}$  and  $x_i$  is defined by the smallest distance of any sample assigned in  $X_{CAL}$  for each  $x_i$ . After the definition of groups, the spectral signatures obtained were subjected to pretreatments that consisted of the combination of transformations from reflectance to absorbance, first derivative of Savitzky-Golay and Standardized Normal Variate (SNV) in order to correct the possible noise of the spectra and optimize the prediction of the models to be generated. The PLSR was carried out after performing each combination of spectral pretreatments.

For evaluation of the models, the R<sup>2</sup> (Equation 1) was taken into account between the data measured and estimated by each model, the RMSE (Equation 2) of the

data derived from the spectral modeling, and the RPD corresponding to the factor by which the precision of the estimate has been increased was compared to the use of the mean of the reference data (Equation 3).

$$R^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{N}} \quad (2)$$

$$DRP = SD_{MED} / RMSE_{VAL} \quad (3)$$

Where  $y_i$  and  $\hat{y}_i$  correspond to the measured and estimated data respectively,  $\bar{y}_i$  corresponds to the mean of the measured data and  $N$  is the total number of samples. For the RPD,  $SD_{MED}$  corresponds to the standard deviation of the total data measured in the laboratory, while the  $RMSE_{VAL}$  corresponds to the root of the mean square error of the data from the external validation group.

The maximum number of components in modeling by partial least squares was defined at 10. By means of a 20-fold cross validation process, the optimal number of latent variables that were included as estimators in the calibration models was determined, which were based on the lowest calibration RMSE (PINHEIRO *et al.*, 2017; HOBLEY; PRATER, 2019; LAAMRANI *et al.*, 2019). The distribution of calibration and validation groups, the construction and subsequent validation of the models was carried out with the software R (R Core TEAM) and the “prospectr” and “pls” libraries.

The performance of the estimation models of each land property was based on the RPD values, applying the classification proposed by Viscarra Rossel, Mcglynn and McBratney (2006), a quantitative model that obtains a RPD less than 1.0 is defined as very poor; a RPD between 1.0 and 1.4 indicates a poor model; RPD values between 1.4 and 1.8 will indicate a regular model; a RPD between 1.8 and 2.0 will be a good model and between 2.0 and 2.5 will indicate a very good model; and a RPD greater than 2.5 will be considered an excellent model. Regarding the  $R^2$  values, according to McBratney and Viscarra Rossel (2008), the estimates of soil properties will be classified as: very good ( $R^2 > 0.80$ ), good ( $0.80 > R^2 > 0.60$ ), regular ( $0.60 > R^2 > 0.40$ ) and poor ( $R^2 < 0.40$ ).

The regression coefficients obtained from the PLSR provide qualitative information on the correlations between the data measured in the laboratory and the spectral signatures (SUMMERS *et al.*, 2011; TEKIN; TUMSAVAS; MOUAZEN, 2012). From this information, the most representative spectral bands were identified to estimate the SOC and textural fractions of the soil.

### Geostatistical analysis

Once the models were calibrated and validated, the experimental semivariograms were calculated for

the measured and spectrally estimated data. These semivariograms were adjusted to theoretical models, whose best adjustment performance was based on obtaining lower values of the sum of squares of the residuals, and the higher values of  $R^2$  and the Cross-Validation Coefficient (CVC). Once the semivariogram of each property was adjusted, the Degree of Spatial Dependence (DSD) was verified by the relationship between the nugget effect ( $C_0$ ) and the sill ( $C_0 + C_1$ ). DSD is classified as weak when it is less than 25%, moderate when it reaches values between 25% and 75%, and strong at values greater than 75% (CAMBARDELLA *et al.*, 1994). From the adjusted semivariograms, ordinary kriging was applied to perform spatial interpolation. Ordinary kriging is a widely used method for spatial prediction analysis, as it allows obtaining unbiased linear predictions with minimal variance (DIGGLE; RIBEIRO, 2007). The geostatistical processing of the data was carried out with the GS + <sup>TM</sup> v.9 software (Gamma Design Software, LLC, Plainwell, MI) in combination with ArcGIS® v. 10.8 (ESRI).

## RESULTS AND DISCUSSION

### Descriptive analysis of the measured variables

The predominant texture classes in the study area were silty loam for the first three depths (0.0-0.30 m) and silty clay loam for the remaining two depths (0.30-0.50 m). The finer texture present in the deeper thicknesses, compared to the surface thicknesses, is characteristic of Oxisols with an intense degree of physical weathering (CAMACHO-TAMAYO *et al.*, 2017). As for the SOC contents, a downward behavior is evident in the soil profile, that is, lower SOC contents were found with increasing profile depth. The descriptive analysis of the measured variables is shown in Table 1. The behavior of increase and decrease of SOC and textural fractions, similar to that of the present research, has also been reported by other studies carried out in oxisols in South America (RAMIREZ-LOPEZ; REINA-SÁNCHEZ; CAMACHO-TAMAYO, 2008; CAMACHO-TAMAYO; RUBIANO; HURTADO, 2014).

### Soil spectral signatures

Figure 2A shows all the spectral signatures obtained in the study area. The reflectance behavior of these soils proved to be similar to that of other studies previously reported in Oxisols, which were developed in Brazil (PINHEIRO *et al.*, 2017; POPPIEL *et al.*, 2018). To understand how a spectral response can explain a given response variable, it is useful to do a qualitative analysis of these spectra. In the spectral signatures obtained in this research, a variation in the reflectance of the samples is evidenced, which tends to increase as the depth within

**Table 1** - Descriptive analysis of the texture and SOC for each depth sampled

Property	Unit	Depth [m]	Mean	Minimum	Maximum	SD	CV
SOC	g kg <sup>-1</sup>	0.0-0.10	26.15	19.76	32.41	0.26	10.08
		0.10-0.20	17.49	13.30	21.57	0.18	10.51
		0.20-0.30	13.87	10.76	17.53	0.16	11.33
		0.30-0.40	11.24	8.46	15.04	0.17	14.78
		0.40-0.50	9.53	5.77	13.11	0.16	16.50
Sand	%	0.0-0.10	6.45	1.85	9.65	2.04	31.82
		0.10-0.20	4.47	1.53	7.53	1.44	32.26
		0.20-0.30	3.16	1.00	5.29	1.03	32.23
		0.30-0.40	2.72	1.03	5.00	0.94	33.68
		0.40-0.50	2.39	1.18	3.83	0.78	33.43
Silt	%	0.0-0.10	74.63	66.22	82.37	3.44	4.64
		0.10-0.20	73.89	63.44	83.89	4.28	5.81
		0.20-0.30	71.85	62.13	82.34	4.37	6.08
		0.30-0.40	69.40	60.47	81.60	4.69	6.73
		0.40-0.50	67.80	60.53	75.10	3.75	5.55
Clay	%	0.0-0.10	18.92	11.21	26.46	3.42	17.87
		0.10-0.20	21.64	11.77	32.23	4.39	20.06
		0.20-0.30	24.99	13.69	35.62	4.70	18.99
		0.30-0.40	27.88	16.96	36.71	4.26	15.28
		0.40-0.50	29.81	22.90	36.91	3.74	12.55

SD = standard deviation. CV = coefficient of variation

the soil profile increases (Figure 2B). This distinction is due to the characteristic contents of organic matter (or organic carbon) and iron oxides characteristic of Oxisols (CARNIELETTO *et al.*, 2018).

The effects of iron oxides and their relationship with color variation in soils can also be described from spectral data. Yellowish soils show a spectral behavior influenced by goethite, presenting higher spectral reflectance (POPPIEL *et al.*, 2018), while reddish soils, characteristic for their hematite content, tend to present lower reflectance (CARNIELETTO *et al.*, 2018). The highest proportion of clay in the samples corresponding to the depths of 0.30-0.50 m, in relation to the first three depths, is exhibited by the greater decrease in the reflectance of these deeper thicknesses from the wavelength 2200 nm, which has been reported in other investigations (SUMMERS *et al.*, 2011; CURCIO *et al.*, 2013). The highest SOC contents were found in the most superficial depths (Table 1), and the spectral responses corroborated this, since the reflectance decreased with increasing SOC content. This is due to the fact that the organic material present in the soil samples absorbs energy, which triggers a low intensity

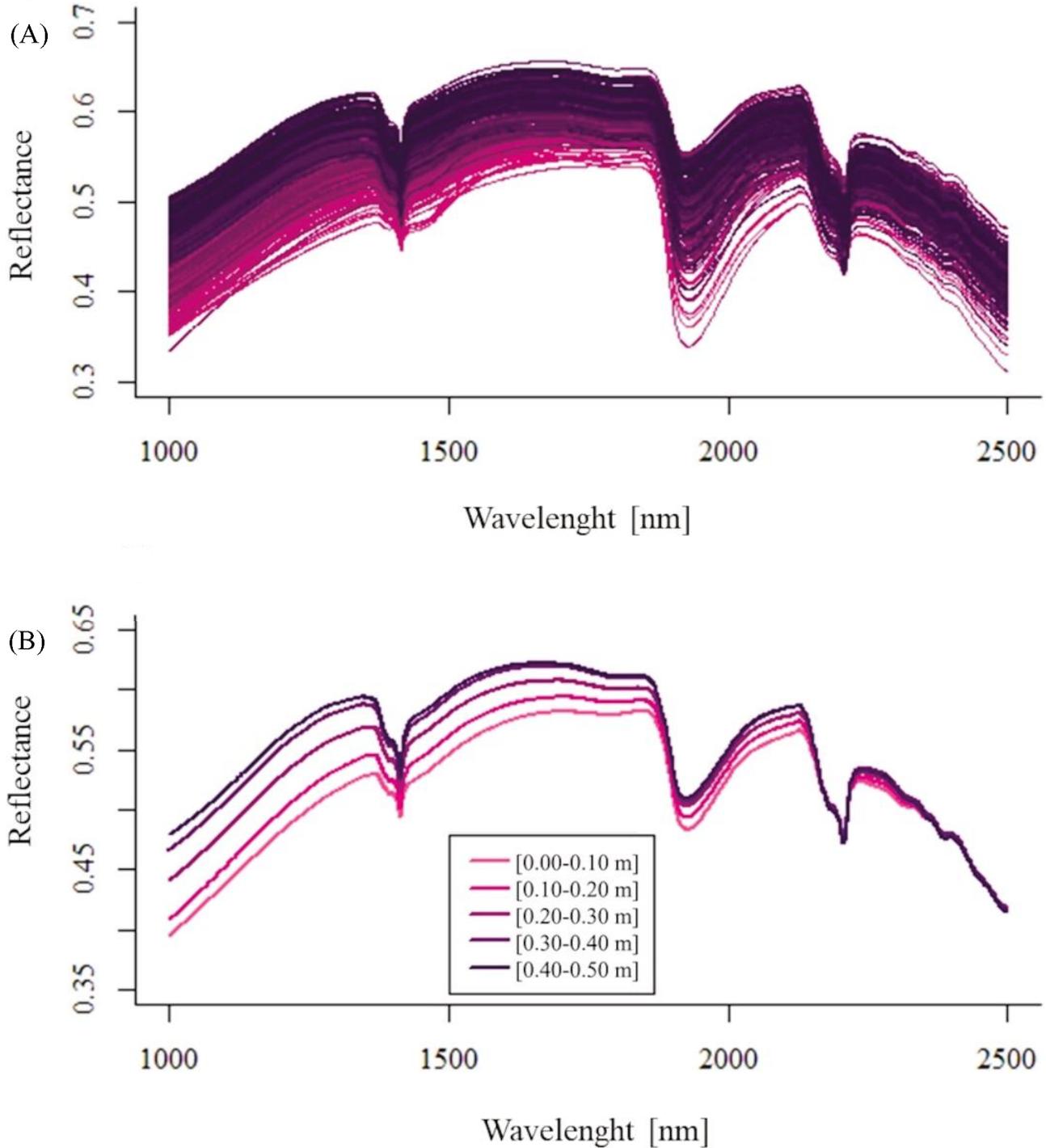
of reflectance throughout the spectrum, which tends to decrease at higher wavelengths (CAMACHO-TAMAYO; RUBIANO; HURTADO, 2014).

#### Performance of spectral models

The models that achieved a better performance according to the RPD and R<sup>2</sup> parameters for the evaluated properties can be seen in Table 2. The graphs of the linear regression for the measured and estimated data in the validation and calibration groups are shown in Figure 3. It is evident that for the validation of the models, according to McBratney and Viscarra Rossel (2008) for R<sup>2</sup>, and Viscarra Rossel, McGlynn and McBratney (2006) for RPD, an excellent estimate of SOC was achieved, followed by good estimates for sand and clay fractions, and finally, a regular performance for silt.

For SOC, the model showed an excellent performance with very good estimates according to its RPD (5.63) and its R<sup>2</sup> (0.97) respectively, and the RMSE in validation reached a value of 1.10 g kg<sup>-1</sup> (Table 2). This performance obtained with spectral models is similar to studies carried out on a large scale, such as the one

**Figure 2** - A) Spectral signatures obtained for all the samples. B) Spectral signatures averaged according to sampling depth



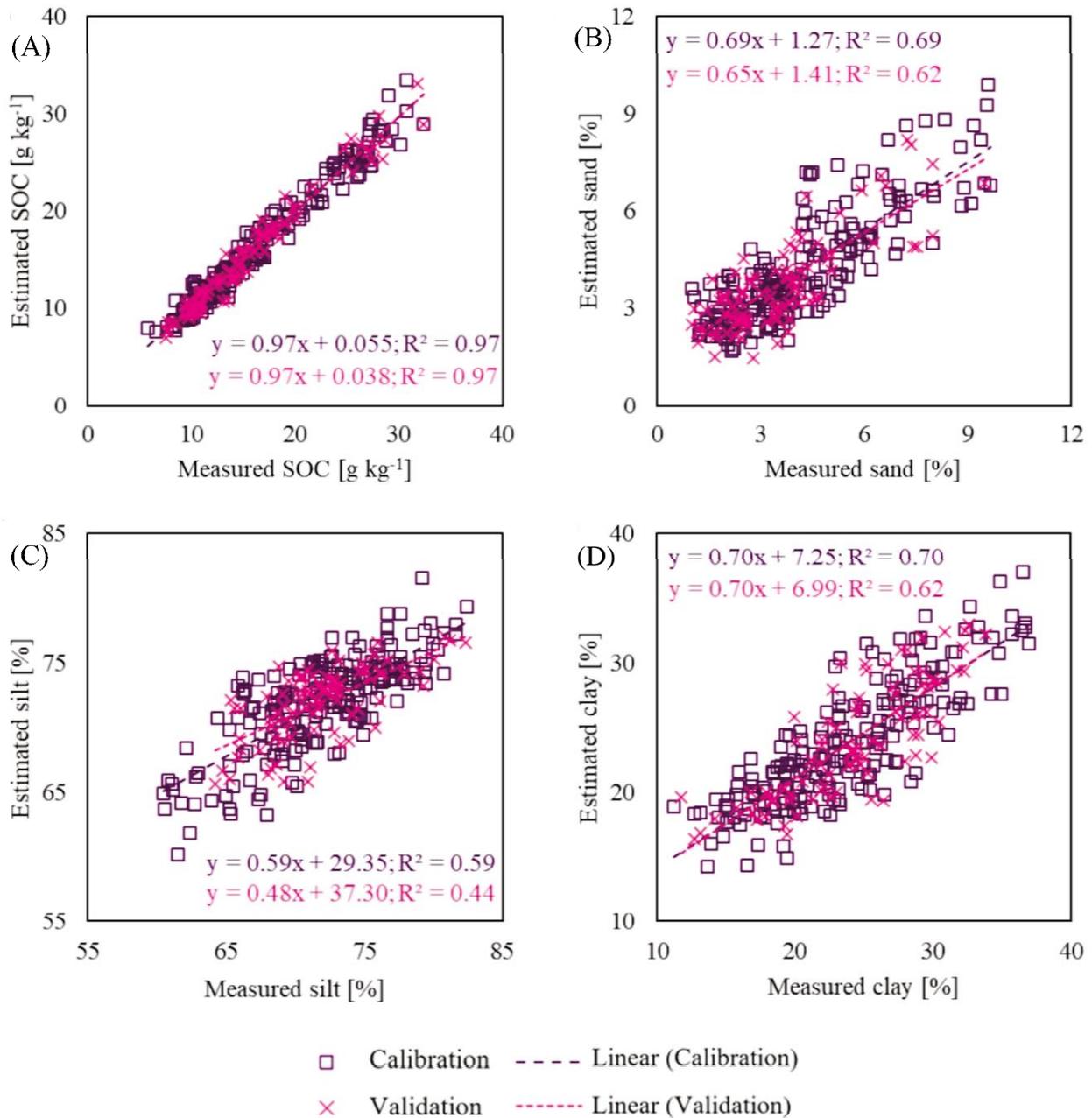
**Table 2** - Parameters for evaluating the performance of spectral models

Property	Unit	Calibration				Validation			
		Range	RMSE	R <sup>2</sup>	RPD	Range	RMSE	R <sup>2</sup>	RPD
SOC	g kg <sup>-1</sup>	5.77 - 32.41	1.14	0.97	5.46	7.48 - 32.40	1.10	0.97	5.63
Sand	%	1.00 - 9.65	1.13	0.69	1.76	1.03 - 9.50	1.10	0.62	1.82
Silt	%	60.47 - 82.37	3.03	0.59	1.56	64.24 - 82.34	2.92	0.44	1.61
Clay	%	11.77 - 36.91	3.12	0.70	1.79	11.21 - 33.80	3.08	0.62	1.81

developed by Liu *et al.* (2019), in which they obtained in the validation a performance with excellent RPD,  $R^2$  equal to 0.96 and an RMSE of  $2.90 \text{ g kg}^{-1}$  for a calibrated model with a total of 11213 soil samples. Jaconi, Don and Freibauer (2017) reported in their review of studies in Europe, on a regional and continental scale, with stratified samples, very good calibrations according to  $R^2$  and excellent for their RPD. For many applications, this degree of estimation is permissible, which makes NIR spectroscopy analysis a cost-effective option for the determination of

SOC, and can be applied to a wide range of soil types, when a representative dataset is available, robust enough to build the models (JACONI; DON; FREIBAUER, 2017). The sampling carried out at the different depths in this investigation allowed to generate a spectral model robust enough to estimate, in a way closer to reality, the variation of the SOC in the soil profile compared to other studies where, in general, one or two depths are sampled, either horizon A or B, or both, as is the case of the research of Camacho-Tamayo, Rubiano and Hurtado (2014).

**Figure 3** - Results of the validation of the spectral models for (A) SOC, (B) sand, (C) silt and (D) clay



For the texture fractions, contrasting results were obtained in the validation of the models compared to other investigations. This may be due to the fact that together, the sand, silt and clay fractions add up to 100%, and the uncertainty in their calculation may increase when obtaining a fraction indirectly, while in the case of SOC, it is obtained directly with laboratory methodology (AHMADI *et al.*, 2021). For sand, a good  $R^2$  (0.62) was obtained, a good quantitative prediction (RPD = 1.82), which agrees with what was found by Pinheiro *et al.* (2017) in soils with an average sand content of 39% in which they achieved an  $R^2$  of 0.62 and a RPD 1.61. In soils with average sand contents of 9.3% (varying between 1.9% and 82.9%), applying PLSR and various calibration models, Hobley and Prater (2019) obtained excellent RPD (from 3.90 to 4.49) together with an  $R^2$  of 0.95. For the clay fraction, the model reached a  $R^2$  of 0.62, which is considered good, and a RPD of 1.81. For soils with an average clay percentage of 32%, a validation RMSE of 8.49% and a very good RPD (2.35) are reported in the study developed by Viscarra Rossel and Webster (2012) in Australia. These good estimates for the calculation of clay have also been reported by Curcio *et al.* (2013) showing  $R^2$  greater than 0.80. Regarding the fraction of silt, the obtained models showed less satisfactory results ( $R^2 = 0.44$  and RPD = 1.61), which agrees with that reported by Tumsavas *et al.* (2018) and Pinheiro *et al.* (2017), the latter obtaining an  $R^2$  of 0.36 and a RPD of 1.25. The poor performance in predicting the fraction of silt may be due to errors associated with the laboratory methodology used to measure the texture. When the silt content is calculated as the remaining percentage of 100% after adding the clay and sand contents, this would then generate a greater degree of analytical error, which would be reflected in the silt fraction, as is the case with what happened in the study by Pinheiro *et al.* (2017). In the study carried out by Camacho-Tamayo *et al.* (2017) in the eastern plains of Colombia at the Research Center of Carimagua, estimation of the models of the textural

fractions was made using the reference data obtained from the Bouyoucos methodology. The pipette, compared to Bouyoucos, can disaggregate the soil particles in a greater proportion due to the longer stirring time in the procedure, which was evidenced in the increase in the silt fraction and the decrease in the sand fraction in comparison with the study carried out by Camacho-Tamayo *et al.* (2017), which was also reflected in the performance of the models calculated in these two investigations.

The application of multivariate statistical methods is necessary to find correlations of spectral signatures with soil properties, with PLSR being the most applied methods (ANGELOPOULOU *et al.*, 2019; NAWAR; MOUAZEN, 2019), which bases its operation on the description of linear relationships of the variables of interest. However, these relationships are not always linear, and it is necessary to apply techniques such as neural networks or machine learning, to optimize the estimates, especially with more robust databases. Although, for this research it was possible to obtain an excellent model for the estimation of SOC, for the texture, on the other hand, not so satisfactory results were obtained, for which it could be explored in the application of these other techniques to achieve better quality estimates, or make the reference database more robust. Although it is possible to obtain good results with samples external to the study area, it would be convenient to include these external samples in a recalibration of the model, in order to strengthen its predictive capacity and ensure that these estimated values have a greater representativeness, especially with studies at a local or regional scale (VISCARRA ROSSEL *et al.*, 2016).

### Representative wavelengths in spectral modeling

The beta regression coefficients ( $\beta$ ) as a function of wavelength, derived from the PLSR, are shown in Figure 4 for each of the properties addressed.

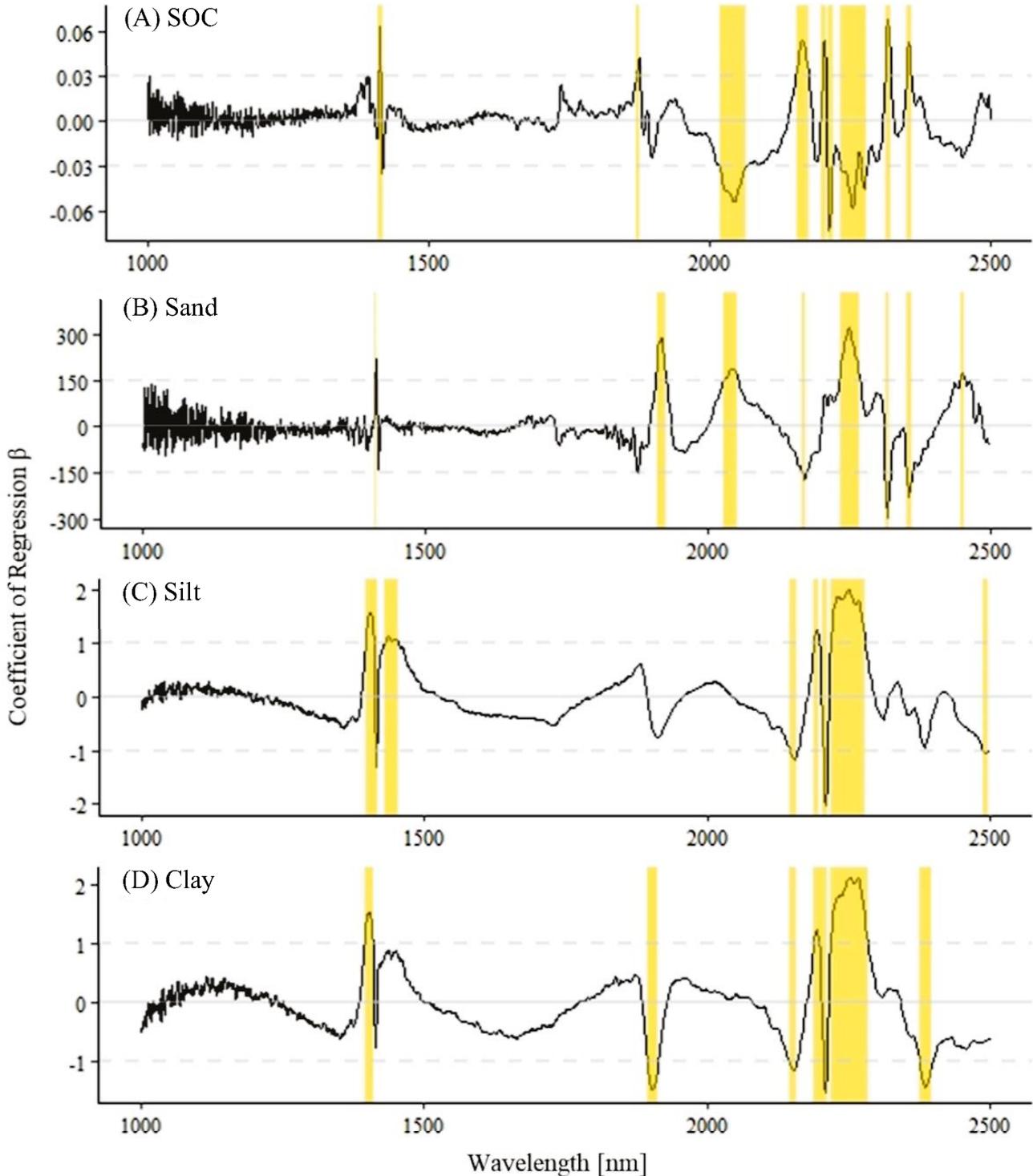
**Table 3** - Relative bands of greater weight in the construction of spectral models according to the estimated property

Model/ Property	Unit	$\beta$ Coefficient considered	Spectral bands with the greatest influence on the model	Number of spectral bands
SOC	$\text{g kg}^{-1}$	$\beta \geq 0.03; -0.03 \geq \beta$	1410-1418, 1869-1876, 2017-2065, 2155-2176, 2200-2207, 2210-2219, 2234-2279, 2312-2322, 2351-2359	163
Sand	%	$\beta \geq 150; -150 \geq \beta$	1410-1413, 1875-1876, 1909-1926, 2029-2053, 2168-2174, 2234-2269, 2314-2322, 2351-2361, 2448-2454	110
Silt	%	$\beta \geq 1.0; -1.0 \geq \beta$	1395-1416, 1430-1453, 2145-2159, 2189-2197, 2204-2211, 2218-2277, 2487-2495	140
Clay	%	$\beta \geq 1.0; -1.0 \geq \beta$	1395-1410, 1895-1913, 2145-2159, 2189-2211, 2218-2284, 2376-2397	156

These regression coefficients are obtained from the linear combination of the predictive values in the PLSR, and can be used to highlight wavelengths with a greater degree of influence on the property of interest (TEKIN; TUMSAVAS; MOUAZEN, 2012; TUMSAVAS

*et al.*, 2018). Table 3 shows the relative spectral bands of greater weight in the construction of the spectral model of each property, in which the greater number of relevant wavelengths is evidenced in the estimation of SOC in comparison with sand, silt and clay.

**Figure 4 -** Regression coefficients  $\beta$  as a function of wavelength for the spectral models of (A) SOC, (B) sand, (C) silt and (D) clay



Considering regression coefficients as a function of the wavelength generated from the PLSR, spectral bands of greater representativeness are identified for the estimation models, which manifest themselves in the form of peaks, which are generated through the wavelength. These peaks are considered as indicators of the correlation between the infrared frequencies and the constituents of interest in the soil, in this case, the SOC and the texture fractions. Figure 4 shows approximate lengths of 1400, 1900 and 2200 nm that had the greatest influence on the generation of models for the four properties. These three characteristic bands are related to the bending and stretching of the O – H bonds in networks of minerals and water molecules, associated with clay minerals and other components of the soil matrix (SUMMERS *et al.*, 2011; TEKIN; TUMSAVAS; MOUAZEN, 2012).

Because it is an Oxisol, the soil in this study is characterized by the presence of clay minerals such as hematite, kaolinite, gibbsite and goethite, which have an influence on the spectral response in the NIR (VISCARRA ROSSEL; MCGLYNN; MCBRATNEY, 2006), as in wavelengths of 1400 and 2200 nm for kaolinite, gibbsite at approximately 2250 nm due to its OH-AL characteristics and goethite with weak distinctions at wavelengths close to 1700 nm (VENDRAME *et al.*, 2012). Figure 4 shows the behavior of peaks in the spectral bands adjacent to those reported for the characteristic presence of these minerals. For the soils in the study area, the predominant mineral in the clay fraction is kaolinite, which is manifested in the peaks of 1900 and 2200 nm (CAMACHO-TAMAYO *et al.*, 2017). It should be clarified that the estimation of clay and its associated minerals has achieved correlations with spectral bands in the visible spectrum (CARNIELETTO *et al.*, 2018; TUMSAVAS *et al.*, 2018; HOBLEY; PRATER, 2019), which is not part of this research.

The results of this research showed that the prediction performance of sand and clay were very similar (Figures 3D and 3D), while for silt a lower performance was achieved (Figure 3C). The similarity in the quality of the models is verified with the greater influence of similar or close wavelengths in the generation of the models (Figures 4B, 4C and 4D). This may be because clay, silt, and sand can be associated with similar spectral regions of the NIR range (LASHYA *et al.*, 2018). For the SOC, it was possible to obtain an accurate estimation model (Figure 3A), and in turn in the spectrum, it was possible to evidence a greater number of wavelengths of relevance for the generation of the model (Figure 4A) compared to those obtained for the estimation of the texture (Tabla 3). The wavelengths of greater weight in spectral models for SOC found in this investigation were mostly similar to those reported by other carbon studies with NIR spectroscopy (SUMMERS *et al.*, 2011; ANGELOPOULOU *et al.*, 2019; LAAMRANI *et al.*, 2019), such as, for example, the 2200 nm region, which has been shown to be correlated with lignin and humic acids, important in the prediction of SOC (SUMMERS *et al.*, 2011). The spectral bands highlighted in Table 3 may be useful for other investigations, with which it could be corroborated if they are analogous in more types of soil, or if they could only be applicable for the study of Oxisols.

#### Geostatistical interpolation surfaces

The semivariograms adjusted for the analyzed data are shown in Tables 4 and 5. It is important to mention that, for each of the properties at each depth, with measured and estimated data, no weak DSDs were presented; therefore, no nugget-type models were evidenced. The spatial structure of the adjusted theoretical semivariogram model was conserved in most cases between the measured and estimated data for each property. For the CVC and R<sup>2</sup> parameters, values greater than 0.70 were obtained in almost all of the semivariograms generated.

**Table 4** - Parameters of the theoretical semivariograms for SOC and sand obtained from measured and spectrally estimated data

Property	Depth [m]	Semivariogram model	C <sub>0</sub>	C <sub>0</sub> + C <sub>1</sub>	Range [m]	CVC	R <sup>2</sup>	DSD [%]
Measured SOC [g kg <sup>-1</sup> ]	0.00-0.10	Exponential	2.1E <sup>-2</sup>	5.9E <sup>-2</sup>	753	0.73	0.95	65
	0.10-0.20	Spherical	1.5E <sup>-2</sup>	3.3E <sup>-2</sup>	1046	0.72	0.96	56
	0.20-0.30	Spherical	8.1E <sup>-3</sup>	2.0E <sup>-2</sup>	695	0.85	0.83	59
	0.30-0.40	Exponential	1.1E <sup>-2</sup>	2.7E <sup>-2</sup>	711	0.77	0.90	59
	0.40-0.50	Spherical	1.0E <sup>-2</sup>	2.4E <sup>-2</sup>	538	0.76	0.91	58
Estimated SOC [g kg <sup>-1</sup> ]	0.00-0.10	Spherical	1.6E <sup>-1</sup>	1.2E <sup>+0</sup>	582	0.81	0.94	87
	0.10-0.20	Exponential	1.5E <sup>-1</sup>	3.8E <sup>-1</sup>	828	0.79	0.97	60
	0.20-0.30	Exponential	7.3E <sup>-2</sup>	1.8E <sup>-1</sup>	760	0.71	0.99	58
	0.30-0.40	Spherical	4.6 E <sup>-2</sup>	1.7E <sup>-1</sup>	641	0.87	0.88	72
	0.40-0.50	Spherical	2.6E <sup>-2</sup>	1.0E <sup>-1</sup>	490	0.94	0.85	75

Continuation Table 4

Measured sand [%]	0.00-0.10	Exponential	1.20	4.12	700	0.71	0.96	71
	0.10-0.20	Exponential	0.65	1.84	600	0.71	0.95	65
	0.20-0.30	Spherical	0.47	0.98	391	0.89	0.74	52
	0.30-0.40	Exponential	0.29	0.73	950	0.90	0.85	60
	0.40-0.50	Spherical	0.16	0.49	427	0.78	0.87	67
Estimated sand [%]	0.00-0.10	Spherical	0.16	1.24	582	0.81	0.94	87
	0.10-0.20	Exponential	0.15	0.38	828	0.79	0.97	60
	0.20-0.30	Exponential	0.07	0.18	760	0.71	0.99	58
	0.30-0.40	Spherical	0.05	0.17	641	0.87	0.88	72
	0.40-0.50	Spherical	0.03	0.10	490	0.94	0.85	75

**Table 5** - Parameters of the theoretical semivariograms for silt and clay obtained from measured and spectrally estimated data

Property	Depth [m]	Semivariogram model	$C_0$	$C_0+C_1$	Range [m]	CVC	$R^2$	DSD [%]
Measured silt [%]	0.00-0.10	Spherical	3.50	11.50	635	1.00	0.92	70
	0.10-0.20	Spherical	8.67	21.29	1029	0.85	0.95	59
	0.20-0.30	Spherical	8.15	19.12	763	0.79	0.96	57
	0.30-0.40	Spherical	6.29	21.94	575	0.88	0.92	71
	0.40-0.50	Spherical	4.17	16.93	891	0.91	0.97	75
Estimated silt [%]	0.00-0.10	Spherical	0.63	2.08	672	0.87	0.98	70
	0.10-0.20	Spherical	0.33	4.26	685	0.82	0.94	92
	0.20-0.30	Spherical	2.93	8.27	793	0.70	0.92	65
	0.30-0.40	Exponential	4.56	12.60	706	0.78	0.81	64
	0.40-0.50	Spherical	2.93	8.81	739	0.73	0.88	67
Measured clay [%]	0.00-0.10	Spherical	5.73	12.30	1230	0.93	0.97	53
	0.10-0.20	Spherical	9.06	20.60	1177	0.79	0.91	56
	0.20-0.30	Spherical	10.96	25.13	912	0.80	0.94	56
	0.30-0.40	Spherical	1.35	19.68	448	0.86	0.92	93
	0.40-0.50	Spherical	7.39	14.79	1171	0.99	0.91	50
Estimated clay [%]	0.00-0.10	Spherical	0.87	3.37	677	0.89	0.88	74
	0.10-0.20	Spherical	0.24	5.22	650	0.85	0.85	95
	0.20-0.30	Spherical	3.15	11.50	900	0.82	0.97	73
	0.30-0.40	Spherical	4.14	11.37	513	0.79	0.99	64
	0.40-0.50	Spherical	2.91	7.35	748	0.69	0.93	60

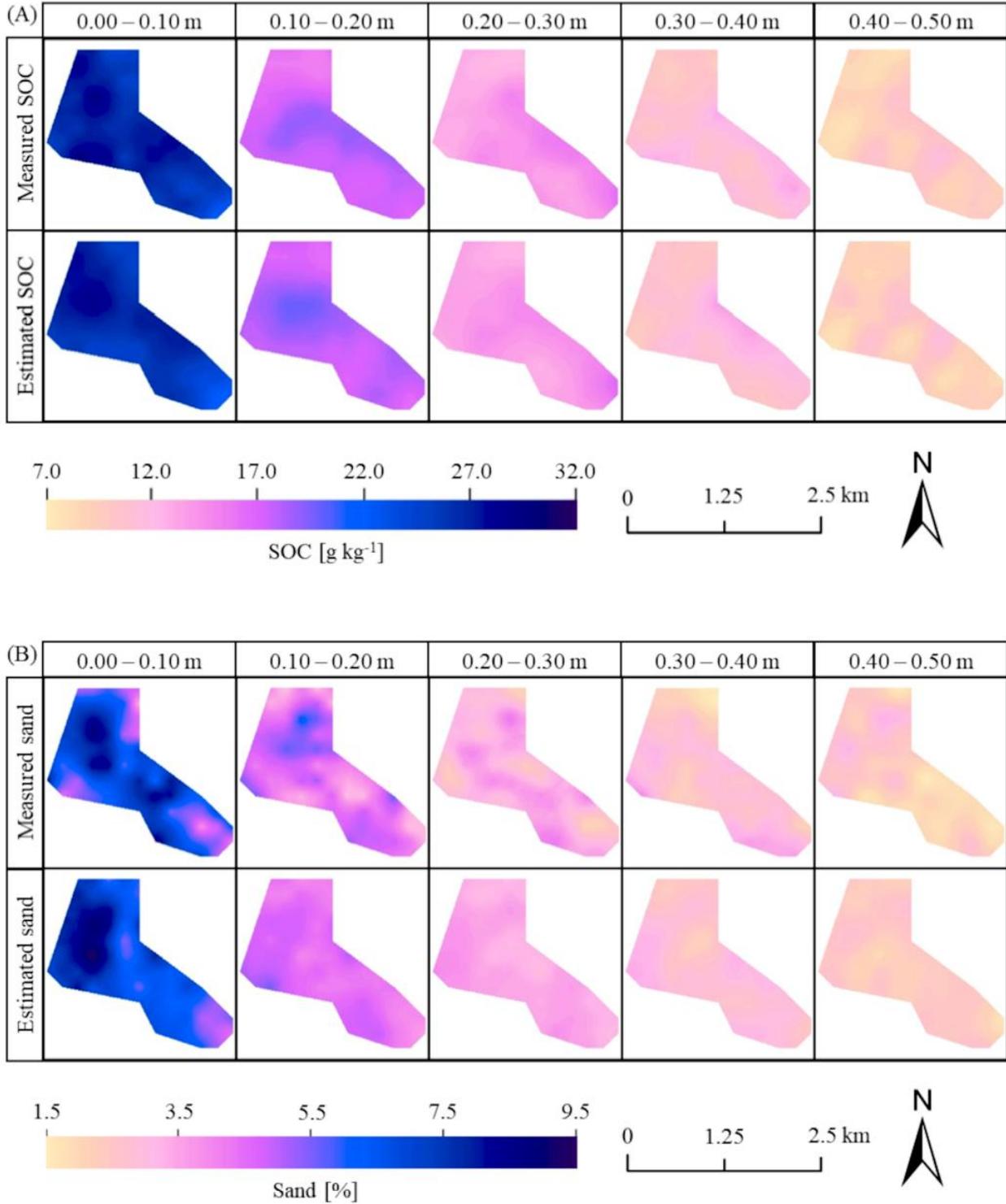
For the geostatistical interpolation surfaces (Figures 5 and 6) it was observed that the spatial behavior of the variables remained between the measured and estimated data as reported by other studies (SUMMERS *et al.*, 2011; CAMACHO-TAMAYO; RUBIANO; HURTADO, 2014; TUMSAVAS *et al.*, 2018). However, it is important to clarify that the level of correspondence

that interpolation surfaces may have between measured and estimated data is due to the performance of the spectral model. For the SOC (Figure 5A), there were no differences between maximum and minimum values per interpolation surface of measured and estimated data at the level that did occur for the textural fractions (Figure 5B, 6A and 6B), which is attributed to lower representativeness

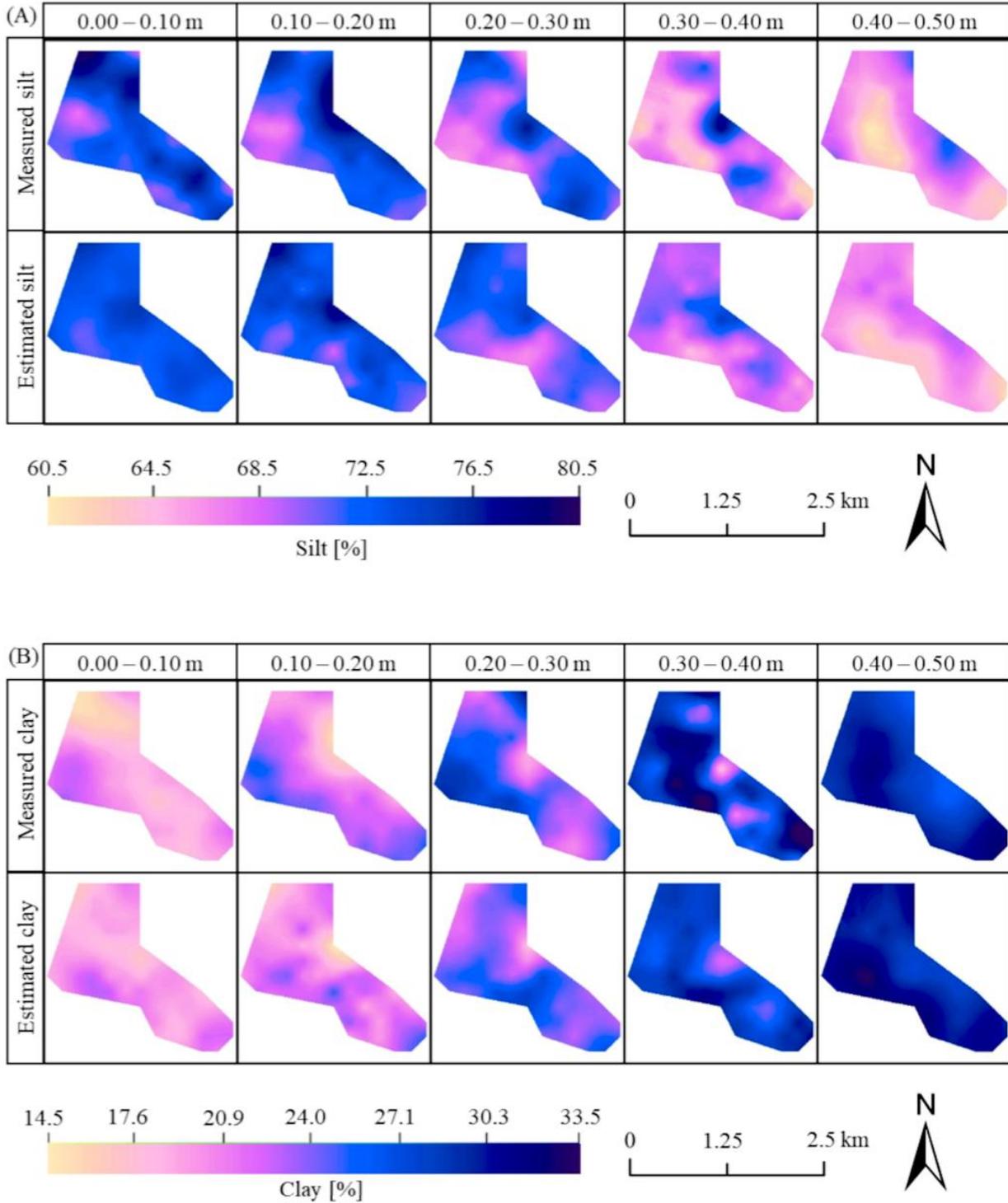
of the models of the textural fractions in comparison with the spectral model of SOC (Table 2). The foregoing is also evidenced when examining the behavior of each property according to the sampling depth, where although for

measured and estimated data, the trend in the behavior of SOC, sand and silt, is reduced to greater sampling depths, and that of clay when increasing to greater depths, the correspondence obeys the representativeness of the model.

**Figure 5.** Geostatistical interpolation surfaces obtained with measured and estimated data for (A) SOC and (B) the sand fraction



**Figure 6** - Geostatistical interpolation surfaces obtained with measured and estimated data for the silt fraction (A) and the clay fraction (B)



## CONCLUSIONS

1. This research confirms that the estimation of SOC performed by NIR spectroscopy in an Oxisol from the equatorial tropics, in combination with chemometric

methods, shows to be a useful tool that would save time and resources over the use of standard laboratory methodologies that, despite being quite accurate, may not be environmentally friendly like the Walkley Black case or expensive like the elemental analyzer. On the

other hand, the models for estimating the clay, silt and sand fractions showed a lower performance compared to that obtained for SOC, which should be explored in future research with different methodologies for obtaining reference data and multivariate analysis for this type of soil. With these results, the present study allows us to have a greater detail for the management of soils and their potential for carbon storage;

2. NIR spectroscopy applied in soils allows the analysis of larger volumes of samples in shorter periods of time, which would improve the uncertainty at the spatial scale of soil studies. A greater number of investigations in which different types of soil, their characteristics and their agronomic conditions are evaluated, will be imperative to expand the application of NIR spectroscopy in estimations of textural fractions, SOC and other properties of interest. This can be achieved by strengthening the soil spectral libraries, which would imply a high initial investment in obtaining reference data, which would be offset as the models built are implemented and with the decrease in the execution of conventional laboratory analyses.

## ACKNOWLEDGEMENT

This research was funded by the Research Division of the Universidad Nacional de Colombia (HERMES code No. 37685).

## REFERENCES

- AHMADI, A. *et al.* Soil properties prediction for precision agriculture using visible and near-infrared spectroscopy: A systematic review and meta-analysis. **Agronomy**, v. 11, n. 3, p. 1-14, 2021.
- ANGELOPOULOU, T. *et al.* Remote Sensing Techniques for Soil Organic Carbon Estimation: A Review. **Remote Sensing**, v. 11, n. 6, p. 1-18, 2019.
- CAMACHO-TAMAYO, J. H. *et al.* Evaluación de textura del suelo con espectroscopía de infrarrojo cercano en un oxisol de Colombia. **Colombia Forestal**, v. 20, n. 1, p. 5-18, 2017.
- CAMACHO-TAMAYO, J. H.; RUBIANO, Y. S.; HURTADO, M. DEL P. Near-infrared (NIR) diffuse reflectance spectroscopy for the prediction of carbon and nitrogen in an Oxisol. **Agronomía Colombiana**, v. 32, n. 1, p. 86-94, 2014.
- CAMBARDELLA, C. A. *et al.* Field-Scale Variability of Soil Properties in Central Iowa Soils. **Soil Science Society of America Journal**, v. 58, n. 5, p. 1501-1511, 1994.
- CARNIELETTO, A. *et al.* A systematic study on the application of scatter-corrective and spectral- derivative preprocessing for multivariate prediction of soil organic carbon by Vis-NIR spectra. **Geoderma**, v. 314, p. 262-274, 2018.
- CURCIO, D. *et al.* Prediction of soil texture distributions using VNIR-SWIR reflectance spectroscopy. **Procedia Environmental Sciences**, v. 19, p. 494-503, 2013.
- DAY, P. R. Particle fractionation and particle size analysis. In BLACK, C. A. **Method of soil analysis, Part I**. 1. ed. Soil Science Society of America, 1965. cap. 43, p. 545-567.
- DIGGLE, P.; RIBEIRO, P. **Model-based Geostatistics**. 1. ed. New York: Springer, 2007. 228 p.
- HOBLEY, E.; PRATER, I. Estimating soil texture from vis-NIR spectra. **European Journal of Soil Science**, v. 70, n. 1, p. 83-95, 2019.
- JACONI, A.; DON, A.; FREIBAUER, A. Prediction of soil organic carbon at the country scale: stratification strategies for near-infrared data. **European Journal of Soil Science**, v. 68, n. 6, p. 1-11, 2017.
- JOVIĆ, B. *et al.* Empirical equation for preliminary assessment of soil texture. **Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy**, v. 206, p. 506-511, 2019.
- KENNARD, R. W.; STONE, L. A. Computer Aided Design of Experiments. **Technometrics**, v. 11, n. 1, p. 137-148, 1969.
- LAAMRANI, A. *et al.* Ensemble Identification of Spectral Bands Related to Soil Organic Carbon Levels over an Agricultural Field in Southern Ontario, Canada. **Remote Sensing**, v. 11, n. 11, p. 1-15, 2019.
- LASHYA, P. *et al.* Visible-Near-Infrared Spectroscopy Prediction of Soil Characteristics as Affected by Soil-Water Content. **Soil Science Society of America Journal**, v. 82, n. 6, p. 1333-1346, 2018.
- LIU, S. *et al.* Estimating forest soil organic carbon content using vis-NIR spectroscopy: Implications for large-scale soil carbon spectroscopic assessment. **Geoderma**, v. 348, p. 37-44, 2019.
- MCBRATNEY, A. B.; VISCARRA ROSSEL, R. A. Diffuse Reflectance Spectroscopy as a Tool for Digital Soil Mapping. In HARTEMINK, A. E.; MCBRATNEY, A.; MENDONÇA SANTOS, M. L. **Digital Soil Mapping with Limited Data**. 1. ed. Springer, 2008. cap. 13, p. 165-172.
- NAWAR, S.; MOUAZEN, A. M. On-line vis-NIR spectroscopy prediction of soil organic carbon using machine learning. **Soil & Tillage Research**, v. 190, p. 120-127, 2019.
- NOCITA, M.; *et al.* Prediction of soil organic carbon content by diffuse reflectance spectroscopy using a local partial least square regression approach. **Soil Biology and Biochemistry**, v. 68, p. 337-347, 2014.
- PINHEIRO, É. F. M. *et al.* Prediction of Soil Physical and Chemical Properties by Visible and Near-Infrared Diffuse Reflectance Spectroscopy in the Central Amazon. **Remote Sensing**, v. 9, n. 4, p. 1-22, 2017.
- POGGIO, L.; GIMONA, A. 3D mapping of soil texture in Scotland. **Geoderma Regional**, v. 9, p. 5-16, 2017.
- POPPIEL, R. R. *et al.* Surface Spectroscopy of Oxisols, Entisols and Inceptisol and Relationships with Selected Soil Properties. **Revista Brasileira de Ciência Do Solo**, v. 42, p. 1-26, 2018.

RAMIREZ-LOPEZ, L.; REINA-SÁNCHEZ, A.; CAMACHO-TAMAYO, J. H. Variabilidad espacial de atributos físicos de un TypicHaplustox de los Llanos Orientales de Colombia. **Engenharia Agrícola**, v. 28, n. 1, p. 55–63, 2008.

SUMMERS, D. *et al.* Visible near-infrared reflectance spectroscopy as a predictive indicator of soil properties. **Ecological Indicators**, v. 11, p. 123–131, 2011.

TEKIN, Y.; TUMSAVAS, Z.; MOUAZEN, A. M. Effect of Moisture Content on Prediction of Organic Carbon and pH Using Visible and Near-Infrared Spectroscopy. **Soil Science Society of America Journal**, v. 76, n. 1, p. 188–199, 2012.

TUMSAVAS, Z. *et al.* Prediction and mapping of soil clay and sand contents using visible and near-infrared spectroscopy. **Biosystems Engineering**, v. 177, p. 90–100, 2018.

VENDRAME, P. R. S. *et al.* The potential of NIR spectroscopy to predict soil texture and mineralogy in Cerrado Latosols. **European Journal of Soil Science**, v. 63, n. 5, p. 743–753, 2012.

VISCARRA ROSSEL, R. A. *et al.* A global spectral library to characterize the world's soil. **Earth-Science Reviews**, v. 155, p. 198–230, 2016.

VISCARRAROSSEL, R. A.; MCGLYNN, R. N.; MCBRATNEY, A. B. Determining the composition of mineral-organic mixes using UV – vis – NIR diffuse reflectance spectroscopy. **Geoderma**, v. 137, p. 70–82, 2006.

VISCARRA ROSSEL, R. A.; WEBSTER, R. Predicting soil properties from the Australian soil visible – near infrared spectroscopic database. **European Journal of Soil Science**, v. 63, p. 848–860, 2012.



This is an open-access article distributed under the terms of the Creative Commons Attribution License