Division – Soil Use and Management | Commission – Soil Fertility and Plant Nutrition

Soil spectral library of Piauí State using machine learning for laboratory analysis in Northeastern Brazil

Wanderson de Sousa Mendes[(1)]* (iD), **Cácio Luiz Boechat**[(2)] (iD), **Adriano Venicius Santana Gualberto**[(2)] (iD), **Ronny Sobreira Barbosa**[(2)] (iD), **Yuri Jacques Agra Bezerra da Silva**[(2)] (iD), **Paloma Cunha Saraiva**[(2)] (iD), **Antonny Francisco Sampaio de Sena**[(3)] (iD) **and Lizandra de Sousa Luz Duarte**[(3)] (iD)

[(1)] OpenGeoHub Foundation, Wageningen, the Netherlands.
[(2)] Universidade Federal do Piauí, *Campus* Professora Cinobelina Elvas, Bom Jesus, Piauí, Brasil.
[(3)] Universidade Federal do Rio Grande do Sul, Departamento de Ciência do Solo, Programa de Pós-Graduação em Ciência do Solo, Porto Alegre, Rio Grande do Sul, Brasil.

ABSTRACT: Soil chemical and physical analyses are the major sources of data for agriculture. However, traditional soil analyses are time-consuming, not cost-efficient, and not environmentally friendly. An alternative to traditional soil analyses is soil spectroscopy. This technique is a low-cost and quick analytical method, which can be implemented in a laboratory and/or in-situ. Nevertheless, some spectrometers are expensive and do not contemplate the entire spectrum. Despite this limitation, the main objective of the study was to create a soil spectral library of the Piauí State using only the 1000–2500 nm range. In this sense, it was evaluated and standardized the soil spectral library by accessing the combination of smoothing, standard normal variate, continuum removal, and Savitzky-Golay derivative spectral preprocessing procedures with partial least squares, random forest, and cubist machine learning algorithms. It was collected 262 geo-referenced soil samples at the layer of 0.00–0.20 m across the entire Piauí State, representing most of its soil variability. The soil properties evaluated were pH(H_2O), sand, clay, and soil organic carbon (SOC) contents. This study demonstrated that the Standard Normal Variate was one of the most promising preprocessing procedures to improve model predictions for pH(H_2O), sand, and clay. For SOC and pH, the best overall results were without preprocessing the soil spectra. Moreover, the cubist model was the most accurate in predicting soil properties. Finally, our study showed evidence of the potential and feasibility of using this soil spectral library to estimate soil properties such as pH(H_2O), sand, clay, and SOC.

Keywords: soil spectroscopy, soil predictions, shortwave infrared, NIR.

INTRODUCTION

Soil chemical and physical analyses are the principal sources of data for agriculture. This information allows to better manage the carbon cycle (Guevara et al., 2020), plants growth (Soong et al., 2020), soil acidity and fertility (Cargnelutti Filho et al., 1996), and is part of the soil security framework (Amundson et al., 2015; Bennett et al., 2019). However, those analyses are performed in traditional soil laboratories that produce potential toxic residuals for the environment. For instance, one unique soil sample produces ~1.4 g of $Cr_2O_7^{2-}$ and $Fe(NH_4)_2(SO_4)_2.6H_2O$, and ~5 mL of H_2SO_4 for the analysis of soil organic carbon (SOC) only. In 2019, the Official Network of Soil and Plant Tissue Analysis Laboratories in the Brazilian states of Rio Grande do Sul and Santa Catarina performed 286,659 traditional soil analyses (ROLAS, 2019). A rough calculation shows that those analyses for SOC produced approximately 401.3 kg and 1,433.3 L of toxic residuals per year, which were dumped in the sewerage network (Afonso et al., 2003). Other issues of traditional soil analyses are time-consuming and not cost-efficient (Fang et al., 2018; Jaconi et al., 2019; Borges et al., 2020).

An alternative to traditional soil analyses is soil spectroscopy, which consists of using the emission and absorption of light and other radiation (electromagnetic radiation) by soils (matter). This technique is a low-cost and quick analytical method, which can be implemented in a laboratory and/or in-situ. Nevertheless, some spectrometers are expensive and do not contemplate the entire spectrum. Despite this limitation, it is important to evaluate the capacity of using specific ranges of the spectrum to quantify soil properties. In the last few decades, the diffuse reflectance has arisen and matured as a notable identifier of physical, chemical, and mineralogical components of soils (Stoner and Baumgardner, 1981; Ben-Dor and Banin, 1995; Nocita et al., 2015; Rossel et al., 2016; Demattê et al., 2019). The soil spectroscopy mainly encompasses the visible (Vis, 350–700 nm) and near-infrared (NIR, 700–2500 nm) ranges of the electromagnetic spectrum. In remote sensing, the NIR is usually split into two ranges because of the satellite sensors (Ogen et al., 2017; Crucil et al., 2019; Mendes et al., 2019) as NIR (700–1000 nm) and shortwave infrared (SWIR, 1000–2500 nm) (López-Maestresalas et al., 2016; Wilczyński et al., 2016). The composition presented in this study fixes in the SWIR region because our sensor only captures this range of the electromagnetic spectrum.

Even though there is the Brazilian Soil Spectral Library (Demattê et al., 2019), which was a national effort to build predictive models at distinct geographic levels, acquiring precise spectroscopic models continues a demanding assignment (Grunwald et al., 2015; Moura-Bueno et al., 2020), mainly at downscales. For instance, the number of samples from Piauí State in that soil spectral library was 67 units, which is a low sampling density to calibrate and validate a state quantitative model such as the area of Piauí State (~251,000 km^2). Those authors pointed out that some predictive models presented worse results because of the small number of samples at the state level. Thus, there is a gap to be fulfilled at regional and local scales developing specific soil spectral libraries that better describe quantitatively and qualitatively the physical, chemical, and mineralogical soil properties. Therefore, the main objective was to develop a specific spectral library for Piauí State and; in this sense, this study aimed to (i) evaluate the potentiality of using only the SWIR range to build predictive models and (ii) access which is the best combination of the spectral preprocessing procedure (smoothing, standard normal variate, continuum removal, and Savitzky-Golay derivatives) with three machine learning regression methods (partial least squares, random forest, and cubist) for pH(H_2O), sand, clay, and SOC at the regional scale.

MATERIALS AND METHODS

Study area, soil sampling, and analyses

The study area covers approximately 251,000 km^2 and is the entire extent of Piauí State in the Northeastern Brazil. The two ecotones of the region are Caatinga and Cerrado biomes

(Figure 1). The first one is characterized by large depressions, plateaus, mainly xerophytic flora, and is not as ecologically diverse as other Brazilian ecotones (Oliveira et al., 2019). The Cerrado biome is typified by highlands, savannah flora, and biodiversity. In both biomes, the main soil types, according to the Brazilian Soil Classification System (Santos et al., 2018), are *Latossolos, Argissolos, Cambissolos, Neossolos Regolíticos,* and *Neossolos Quartzarênicos,* which correspond to Ferralsols, Acrisols, Cambisols, Regosols, and Arenosols (IUSS Working Group WRB, 2015), respectively. Approximately 85 % of the Piauí's territory are constituted by phanerozoic sedimentary rocks, and the remaining 15 % are occupied by metamorphic and igneous rocks. Consequently, the soil types are predominately *Latossolos* (Ferralsols), *Neossolos* (Regosols and Arenosols), and *Argissolos* (Acrisols). The region is classified as tropical wet with dry winter (Aw) in the Cerrado biome, as hot semi-arid with dry winter (BShw) in the Caatinga biome, and as hot and humid tropical with dry winter on the Koppen classification system displaying the average annual temperature of 26.7 °C (Alvares et al., 2013).

We collected 262 geo-referenced soil samples at 0.00–0.20 m layer based on geology, relief, and under native vegetation or non-cultivated areas, which comprise areas
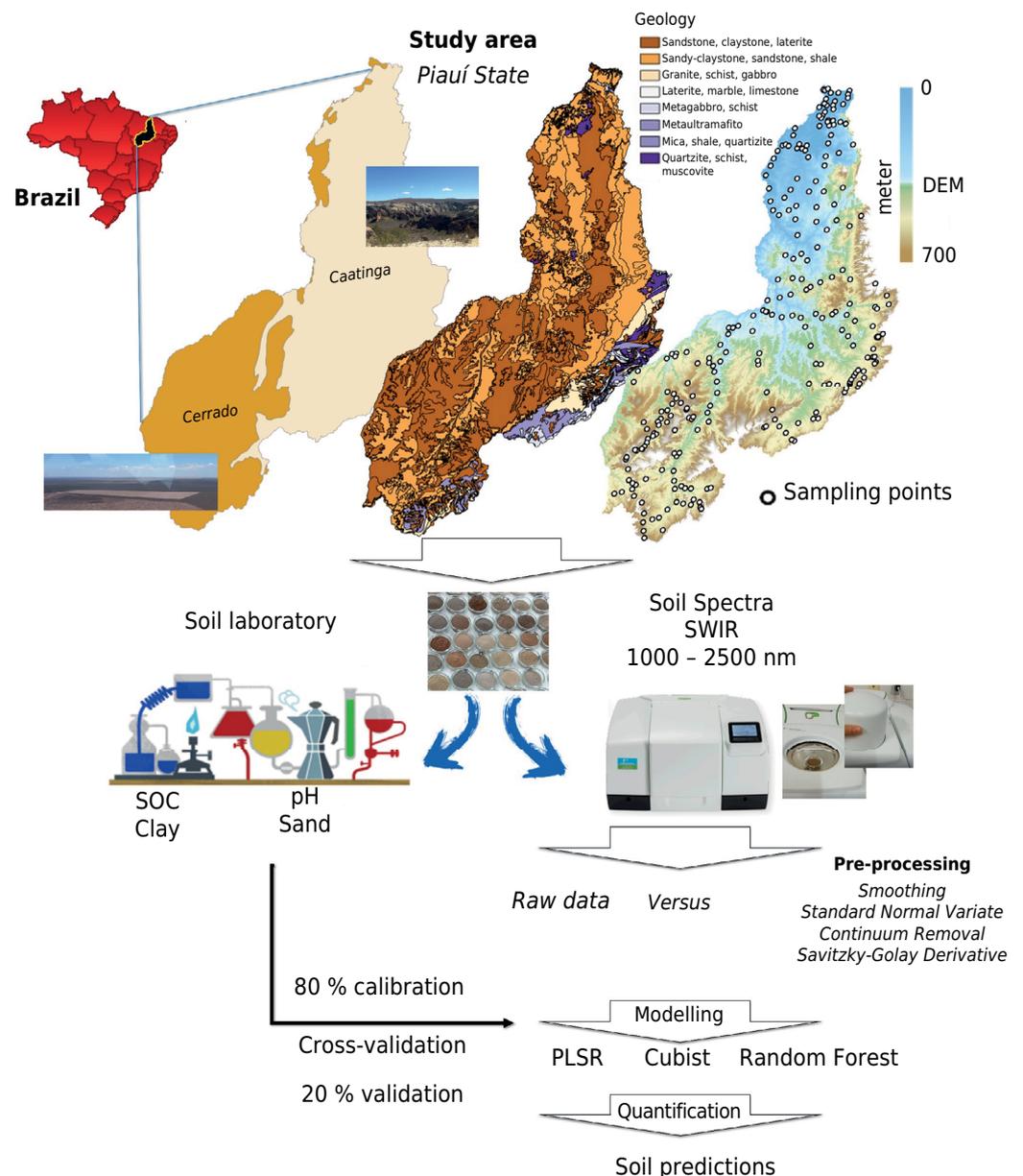

Figure 1. Methodological sequence of the research.

of low impact or landscapes with low or non-anthropic interference. Those samples covered most of the variability of the soil parent materials and types. To ensure a good representation, each soil sample was composed of five sub-samples collected 4 m from the central geo-referenced point using an auger, sieve, and bucket of stainless steel. The soil chemical and physical analyses were carried out in soil samples on oven-dried (48 h at 50 °C), ground, and sieved (<2.0 mm mesh). The soil pH was measured in an extract at 1:2.5 soil/water ratio, which was stirred and kept at rest for 1 hour at room temperature. The soil particle size, clay and sand, was determined by the Bouyoucos densimeter method (Beretta et al., 2014; Day, 2015). The soil organic carbon was determined by the Walkley-Black method by oxidation of the carbon in 10 mL of $K_2Cr_2O_7$ 0.0667 mol L^{-1} via wet and subsequent titration with $(NH_4)_2(SO_4)_2.6H_2O$ 0.102 mol L^{-1} (Walkley and Black, 1934). Following the recent recommendation stated by Minasny et al. (2020), drawing briefly the history of the van Bemmelen's factor (e.g., 1.724) to convert organic carbon into organic matter, we modeled the soil organic carbon instead of soil organic matter. Those authors pointed out that it would be further meticulous in reporting soil organic carbon.

Soil spectral acquisition and preprocessing

The soil spectral data were acquired using a multi-range Fourier Transform-Infrared/Near-Infrared spectrometer (FT-IR/NIR, PerkinElmer Inc., USA) in a range of 1000 to 2500 nm with a spectral resolution of 0.5 nm. The ground and sieved (<2.0 mm mesh) soil samples were placed in a petri dishes for homogeneous arrangement allowing a flat surface for performing the spectral reading. The detector type of the sensor is DTGS, which the signal is generated by the change in temperature altered by absorption of the infrared radiation operated at room temperature (Figure 1). The sensor covers an area of approximately 2 cm^3 in the center of each sample. A white plate called *Spectralon*, which represents a ~100 % reflectance pattern, was used to calibrate the sensor. This procedure was performed at the beginning of each reading to standardize and equalize the spectral measurements.

Beyond the raw spectra, four spectral preprocessing approaches were applied to mathematically correct measurements and noisy. Moreover, these four techniques were selected because of their ability to provide tuning models with high accuracy. The first one was the smoothing, which is a straightforward moving average of spectral data using a convolution function. We used an 11-nm search window to smooth our soil spectra data (Figure 2a). The second preprocessing was the Standard Normal Variate (SNV), which removes scatter in the spectral data. It normalizes each spectrum considering its mean value and the standard deviation (Figure 2b) (Barnes et al., 1989). The third was continuum removal (CR), which eliminates the continuum aspects of spectra and frequently isolates specific absorption characteristics (Figure 2c). The CR finds points placing on the convex hull of 100 % of reflection, joins the points by linear interpolation and standardizes the spectrum by splitting the input data by the interpolated line (Clark and Roush, 1984). Last but not least, the Savitzky-Golay (SG) smoothing minimizes the set of variation and enhances resolutions of spectral peak characteristics (Savitzky and Golay, 1964). We used SG with the first derivative with a first-order polynomial and window size of 5-nm (Figure 2d). The soil spectra preprocessing was performed using the Alradspectra graphical user interface in R programming (Dotto et al., 2019). The setting parameters for SNV and SG were selected after we performed the initial tests.

Spectral modeling and model assessment

The spectral modeling was performed employing the partial least squares regression (PLSR), random forest, and cubist machine learning methods in the R programming (R Development Core Team, 2020). The dataset was randomly split into calibration and validation. The models were calibrated using 80 % of the dataset and implemented in R
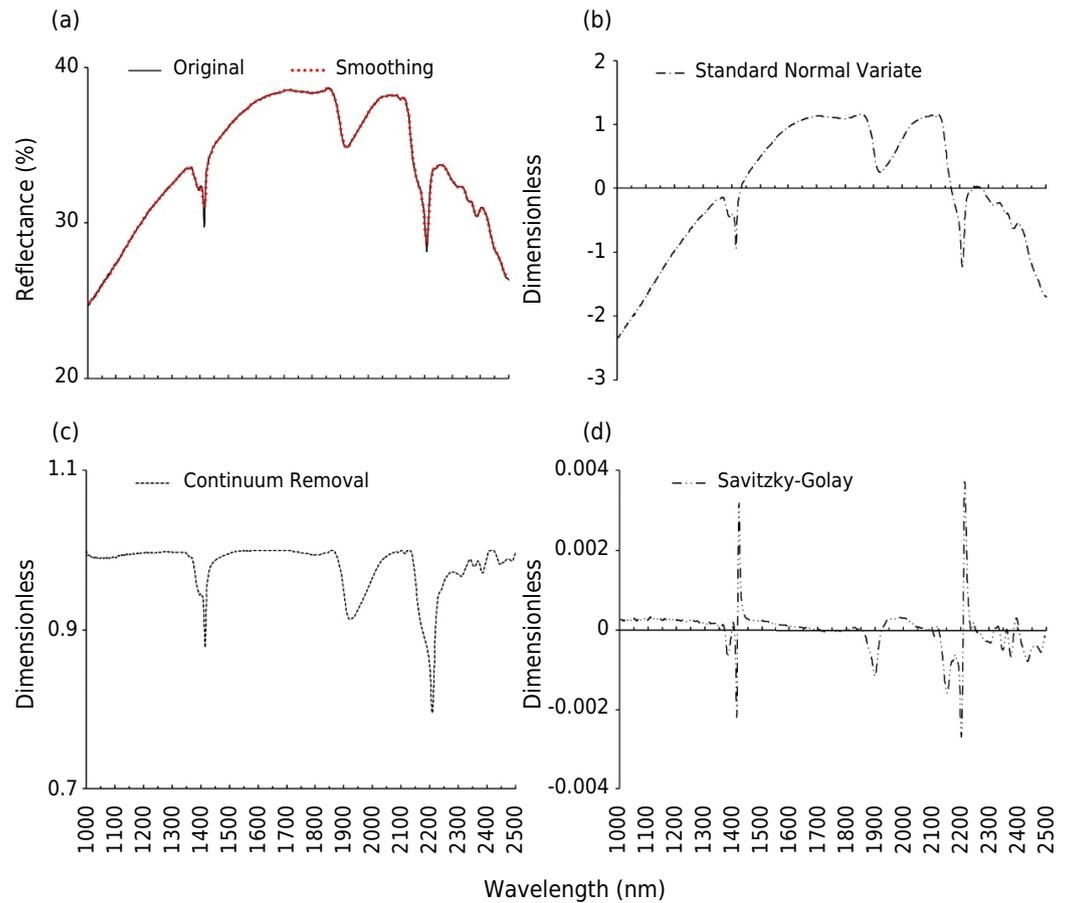
Figure 2. Median of spectral curves preprocessed using smoothing (a), standard normal variate (b), continuum removal (c), and Savitzky-Golay derivative (d) approaches.

package "caret" (Kuhn, 2008). The tunning parameters for each algorithm are described in table 1. The optimization was performed using a 10-fold repeated cross-validation method, executed ten times for each model. The best models were selected based on the lowest RMSE and highest R^2 of calibration and validation dataset.

The PLSR joins features from multiple regression, and principal components analysis diminishing complex spectral matrix solving single- and multi-label learning issues and is widely applied in soil spectroscopic modeling (Ergon, 2013). The adjustable parameter of PLSR is the selection of principal components that explain 95 % of the variance of the latent variables (number of principal components). The cubist model is based on the construction of regression trees, wherein the set of response variables are split into subsets by similarity. The partitions are defined by if-then conditions that generate a sequence of rules. The regulating arguments of the cubist model are interactive model trees created in sequence (committee) and the nearest neighbours, which determine the average of a training set points. Detailed information is described in Quinlan (1993). Last but not least, the random forest computes a mean decrease impurity estimating the feature importance. The adjustable parameters are the number of trees and the number of variables randomly sampled as candidates at each split (Breiman, 2001).

The model evaluation was done using 20 % of the observations left for external validation and the performance of the prediction models assessed employing the root mean square error (RMSE; Equation 1), coefficient of determination (R^2; Equation 2), the ratio of performance to deviation (RPD; Equation 3), and the ratio of performance to interquartile distance (RPIQ; Equation 4).

Table 1. Internal model parameterization of the three predictive methods analyzed for pH(H₂O), sand, clay, and organic carbon using the R package "caret"

Target	Treatment	PLSR	Random Forest	Cubist	
		Ncomp	M_{try}	Committee	Neighbors
pH(H₂O)	Original	3	1501	20	0
	Smoothing	3	1491	10	0
	SNV	2	1501	20	0
	CR	3	54	20	5
	SGD	3	1491	20	9
Sand (g kg⁻¹)	Original	3	1501	20	0
	Smoothing	3	54	20	0
	SNV	2	1501	20	0
	CR	1	54	10	9
	SGD	3	54	20	5
Clay (g kg⁻¹)	Original	2	1501	20	0
	Smoothing	3	1491	20	0
	SNV	3	1501	10	0
	CR	1	54	20	0
	SGD	3	54	20	0
SOC (dag kg⁻¹)	Original	3	1501	10	0
	Smoothing	3	1499	10	0
	SNV	2	54	20	9
	CR	1	54	20	0
	SGD	3	54	20	5

PLSR: Partial Least Squares Regression; SNV: Standard Normal Variate; CR: Continuum Removal; SGD: Savitzky-Golay Derivatives; SOC: soil organic carbon; Ncomp: number of principal components.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}} \qquad \text{Eq. 1}$$

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

$$RPD = \frac{SD}{RMSE} \qquad \text{Eq. 3}$$

$$RPIQ = \frac{IQ}{RMSE} \qquad \text{Eq. 4}$$

In which $\hat{y}$ is the predicted value; $\bar{y}$ is the mean observed value; y is the observed values; n is the number of samples with i = 1,2,3, ..., n; SD is the standard deviation of the observed values; and IQ is the interquartile range of data distribution calculated by the difference between the 3rd and 1st quartile. The RPD values were used just for comparison to the literature. Bellon-Maurel et al. (2010) pointed out that this index's thresholds are arbitrary and without statistical or utilitarian basis.

RESULTS

Soil descriptive and spectral qualitative analyses

Table 2 presents the distribution of pH(H₂O), sand, clay, and soil organic carbon in the study area. The pH(H₂O) values varied from extremely acid (3.5–4.5) to strongly alkaline

(>8). However, the more predominant pH(H$_2$O) values are around 4.52. This pattern is an intrinsic characteristic of the region. The clay content presented values ranging from 107.00 to 508.00 g kg^{-1}, with a median of 163.40 g kg^{-1}. This shows the predominance of sandy soils. The soil organic carbon (SOC) content was up to 2.5 dag kg^{-1}. The skewness and kurtosis indices for all properties demonstrated a satisfactory normal distributed dataset. These explained patterns help to understand the importance of regional/local datasets to build a trustable and representative spectral library.

The soil spectra have both information qualitative and quantitative. The qualitative aspects of the soil spectra are on the absorption and reflectance intensities called valleys and peaks, respectively. Figure 3 shows the median spectral curves of the clay, pH(H$_2$O), and SOC groups. The clay contents were grouped in clay (350–600 g kg^{-1}), loam sand (150–350 g kg^{-1}), sand (100–150 g kg^{-1}), and very sand (<100 g kg^{-1}). Qualitatively, there is a slight difference between sandy and very sandy soil spectral signature, which could be grouped into one class in our specific case. Clayey soils absorb more electromagnetic magnetic energy than sandy soils (Figure 3a). The SOC were divided into three classes: low (<1 dag kg^{-1}), medium (1–2 dag kg^{-1}), and high (>2.5 dag kg^{-1}). We suggested these classes to better distinguish SOC contents in our dataset because our spectrometer has only the SWIR window (Figure 3b). The arrangement for the pH(H$_2$O) values was: very low (3–4.5), low (4.5–5.5), good (5.5–6), high (6–7), and very high (>7). Grouping the pH(H$_2$O) values into these classes, we could qualitatively discriminate the pH(H$_2$O) content at 1900 nm (Figure 3c).

Soil spectral quantification

Analyzing the Pearson's correlation index (Figure 3d), the clay and SOC contents had inverse correlation up to -0.5. The pH(H$_2$O) and sand values had a positive correlation up to 0.5 with the reflectance. The predictive metrics of the models was accessed by checking the external validation dataset left out, which simulate in-situ new sampling (Table 3). Assessing the models for pH(H$_2$O) prediction, the SNV displayed as the best preprocessing procedure among other models (Table 3). The cubist model (RMSE = 0.54; R^2 = 0.72; and RPIQ = 1.62) was the best fitted comparing to PLSR and RF. The spectral modeling of pH(H$_2$O) that showed higher RMSE (1.02), low R^2 (0.26), and RPIQ (0.92) was the RF using the smoothing preprocessing technique.

Table 2. Descriptive statistics of pH in water, sand, clay, and organic carbon contents in the Piauí State, Brazil

Parameters	pH(H$_2$O)	Sand	Clay	Soil Organic Carbon
		——————— g kg^{-1} ———————		dag kg^{-1}
Minimum	3.11	188.90	1.00	0.01
1st Quartile	3.79	645.30	107.00	0.37
Median	4.52	761.40	163.40	0.64
Mean	4.68	729.00	171.40	0.88
3rd Quartile	5.33	844.10	220.40	1.36
Maximum	8.57	990.10	508.00	2.50
CV (%)	22.78	20.87	54.88	78.82
SD (±)	1.06	152.10	94.00	0.69
Skewness	0.94	-0.67	0.25	0.79
Kurtosis	0.57	-0.29	-0.43	-0.42

SD: Standard deviation; CV: coefficient of variation; The soil pH was measured in an extract at 1:2.5 soil/water ratio, which was stirred and kept at rest for one hour at room temperature. The soil particle size, clay, and sand, was determined by the Bouyoucos densimeter method (Beretta et al., 2014; Day, 2015). The soil organic carbon was determined by the Walkley-Black method by oxidation of the carbon in 10 mL of K$_2$Cr$_2$O$_7$ 0.0667 mol L^{-1} via wet and subsequent titration with (NH$_4$)$_2$(SO$_4$)$_2$.6H$_2$O] 0.102 mol L^{-1} (Walkley and Black, 1934).
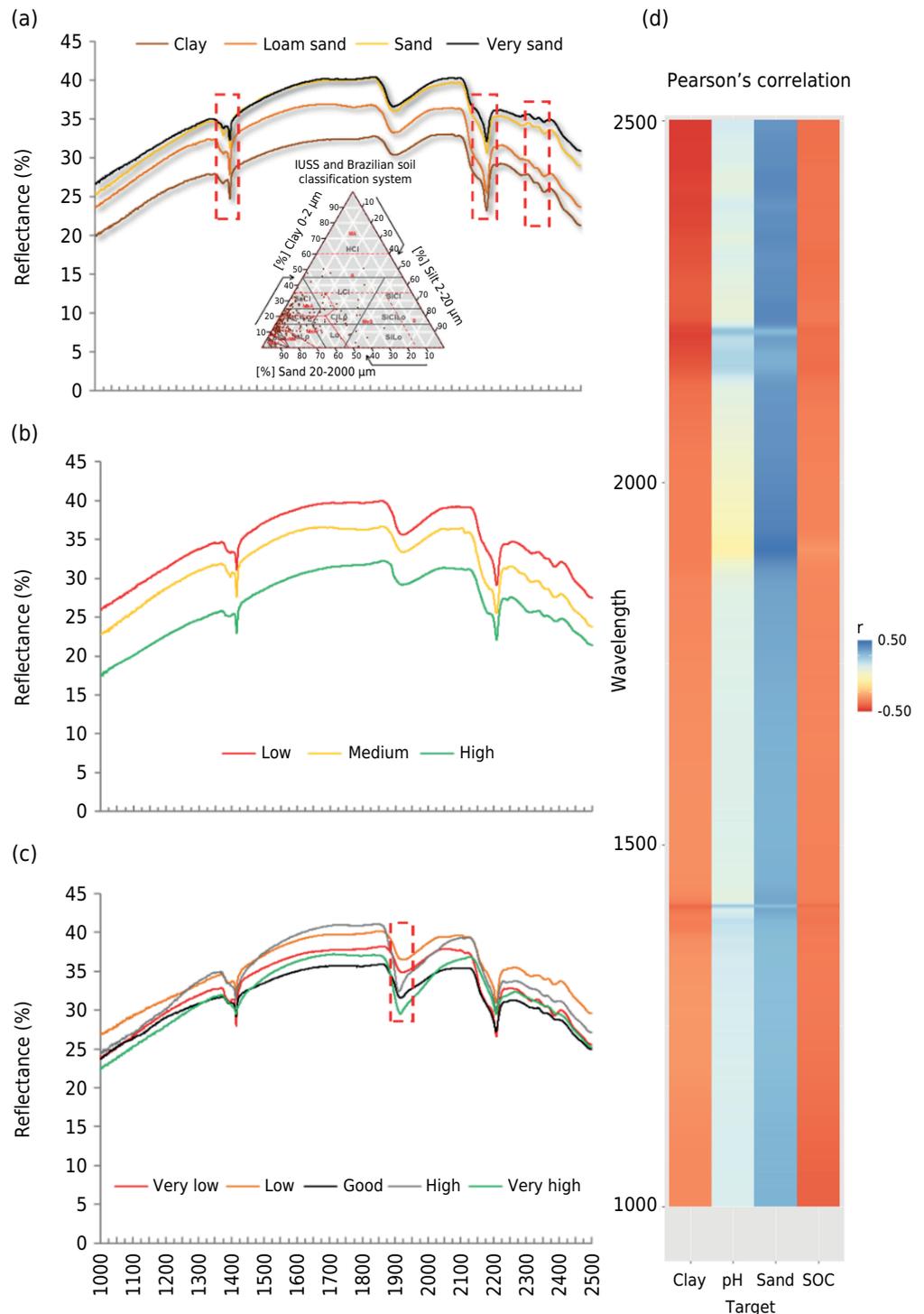
Figure 3. Median of spectral curves grouped by clay (a), soil organic carbon (b), and pH(H$_2$O) (c) contents. Pearson correlation index between the soil properties and spectral reflectance in the shortwave infrared (d).

Overall, the SNV data was better fitted in all models for sand predictions. Using the PLSR model, the RMSE, R^2, RPD, and RPIQ values were 128.10 g kg^{-1}, 0.19, 1.13, and 0.91, respectively (Table 3). The results using RF were 110.10 g kg^{-1}, 0.43, 1.31, and 1.06 for RMSE, R^2, RPD, and RPIQ. However, the best fitted model for sand predictions was using the cubist algorithm (RMSE = 98.7 g kg^{-1}, R^2 = 0.52, RPD = 1.46, and RPIQ = 1.18). Applying the PLSR, the SGD dataset better fitted the clay predictions than other datasets preprocessed or not. For the other two algorithms, the SNV dataset was still the most advised dataset to be used in the calibration of RF and cubist models. However, the

Table 3. Internal and external model evaluation of the three predictive methods analyzed for pH(H2O), sand, clay, and organic carbon

Treatment	Partial Least Square Regression						Random Forest						Cubist					
	$RMSE_{cal}$	R^2_{cal}	$RMSE_{val}$	R^2_{val}	RPD	RPIQ	$RMSE_{cal}$	R^2_{cal}	$RMSE_{val}$	R^2_{val}	RPD	RPIQ	$RMSE_{cal}$	R^2_{cal}	$RMSE_{val}$	R^2_{val}	RPD	RPIQ
pH(H$_2$O)																		
Orig.	0.82	0.40	0.78	0.42	1.33	1.12	0.84	0.37	0.91	0.22	1.14	0.96	0.64	0.65	0.54	0.72	1.94	1.62
SM	0.77	0.46	0.99	0.31	1.19	0.95	0.81	0.38	1.02	0.26	1.15	0.92	0.59	0.66	0.79	0.55	1.49	1.19
SNV	0.84	0.46	0.61	0.57	1.55	1.24	0.76	0.54	0.60	0.61	1.60	1.27	0.62	0.68	0.52	0.71	1.80	1.45
CR	0.76	0.51	0.95	0.35	1.26	0.56	0.68	0.58	0.84	0.49	1.43	0.63	0.63	0.64	0.79	0.55	1.52	0.67
SGD	0.69	0.62	0.70	0.55	1.50	0.94	0.68	0.62	0.67	0.60	1.58	0.99	0.66	0.64	0.62	0.64	1.70	1.06
Sand (g kg^{-1})																		
Orig.	126.50	0.33	142.90	0.10	1.07	0.55	130.80	0.30	226.4	0.02	0.67	0.35	119.70	0.41	122.20	0.34	1.25	0.64
SM	129.90	0.29	139.80	0.11	1.04	0.96	134.50	0.27	137.8	0.16	1.06	0.97	120.10	0.40	140.10	0.17	1.04	0.96
SNV	135.40	0.27	128.10	0.19	1.13	0.91	135.70	0.27	110.1	0.43	1.31	1.06	130.90	0.30	98.70	0.52	1.46	1.18
CR	138.50	0.20	147.20	0.25	1.10	0.80	129.40	0.29	140.4	0.22	1.15	0.84	137.70	0.20	155.10	0.10	1.04	0.76
SGD	120.30	0.38	135.70	0.19	1.09	0.85	124.90	0.37	125.6	0.25	1.18	0.92	130.80	0.30	129.00	0.23	1.15	0.90
Clay (g kg^{-1})																		
Orig.	78.90	0.32	86.30	0.25	1.17	0.69	81.80	0.29	98.60	0.10	1.02	0.60	81.50	0.29	86.10	0.25	1.17	0.69
SM	76.50	0.35	98.10	0.04	0.98	0.57	81.20	0.31	96.90	0.08	0.99	0.58	76.20	0.33	94.90	0.07	1.01	0.59
SNV	86.20	0.23	86.90	0.11	1.05	0.67	87.10	0.20	81.10	0.20	1.12	0.72	79.20	0.30	80.20	0.25	1.14	0.73
CR	89.30	0.21	99.70	0.15	1.09	0.55	79.00	0.24	97.90	0.18	1.11	0.56	82.10	0.22	101.40	0.14	1.07	0.54
SGD	82.50	0.28	77.30	0.21	1.13	0.76	84.90	0.26	81.30	0.13	1.08	0.72	85.00	0.23	111.10	0.01	0.79	0.53
SOC (dag kg^{-1})																		
Orig.	0.61	0.25	0.60	0.14	1.08	0.34	0.64	0.19	0.55	0.29	1.19	0.37	0.60	0.28	0.49	0.42	1.33	0.41
SM	0.62	0.23	0.58	0.20	1.10	0.40	0.63	0.23	0.57	0.21	1.11	0.41	0.61	0.29	0.53	0.29	1.20	0.44
SNV	0.67	0.08	0.69	0.12	1.06	0.37	0.63	0.16	0.68	0.13	1.08	0.38	0.64	0.16	0.72	0.09	1.03	0.36
CR	0.67	0.09	0.70	0.00	0.98	0.48	0.60	0.27	0.64	0.14	1.09	0.53	0.60	0.27	0.64	0.18	1.08	0.52
SGD	0.65	0.17	0.56	0.36	1.28	0.65	0.63	0.14	0.65	0.16	1.10	0.56	0.61	0.21	0.62	0.24	1.17	0.59

Orig.: raw data; SM: Smoothing; SNV: Standard Normal Variate; CR: Continuum Removal; SGD: Savitzky-Golay Derivatives; RMSE: root mean square error; RPD: ratio of performance to deviation; RPIQ: ratio of performance to interquartile distance; SOC: soil organic carbon; cal: calibration dataset; val: external validation dataset. Models with the best efficiency are highlighted in grey for each treatment.

best models for predicting clay content were PLSR using SGD dataset (RMSE = 77.30 g kg^{-1}; R^2 = 0.21; and RPIQ = 0.76), and cubist using SNV dataset ((RMSE = 80.20 g kg^{-1}; R^2 = 0.25; and RPIQ = 0.73). The SGD dataset fitted more accurate SOC predictions than using the raw and other preprocessed datasets through PLSR models. However, the most overall accurate result was that using raw dataset with cubist models (RMSE = 0.49 dag kg^{-1}; R^2 = 0.42; and RPIQ = 0.41).

DISCUSSION

The more predominant pH(H$_2$O) values in the study area were around 4.52, which is considered a very strongly acid environment, following the classification proposed by van Raij et al. (2001). Leite et al. (2010) analyzed the chemical properties of different soil-uses in Piauí State's savannah, and presented a dataset with pH(H$_2$O) values between 4.17 and 5.04 for native Cerrado forest at 0.00–0.40 m layer. This pattern was also similar in our dataset, including the Caatinga Biome. The predominance of sandy soils influences the soil organic carbon pools, reflecting in the water retention and other soil properties (Syers et al., 1970; Rawls et al., 2003; Vinther et al., 2006). It is not surprising that soil organic carbon (SOC) values are usually low. The symmetric parameters (skewness and kurtosis) for the properties demonstrated a satisfactory normal distributed dataset, which is not normally easy to find in a natural environment (Vereecken et al., 2016).

Vis-NIR field measurements can be influenced by water at 1400 and 1900 nm (Rossel et al., 2017). However, after the samples are oven-dried and Vis-NIR laboratory readings conducted, these two absorption features reveal the presence of phyllosilicates such as 2:1, 1:1 or interstratified minerals (Rossel et al., 2016; Demattê et al., 2017; Di Iorio et al., 2019). Evaluating our datasets' spectral features qualitatively, they corroborated the fact that the soils in Piauí are rich in kaolinite and gibbsite (Figure 3a). These spectral features are presented by the characteristic doublet absorption valleys at 1400 and 2200 nm. Marques et al. (2019) evaluated the qualitative and quantitative spectral features of 16 soil profiles and pointed out that spectral features and a flat shape at 2200 and 2265 nm indicate the presence of kaolinite and gibbsite. This is caused due to the O-H bonds at 1400 nm and combined O-H stretch with Al-OH lattice at 1900 nm vibrations of water in the structure of phyllosilicates or absorbed on mineral surfaces (Clark et al., 1990). Kaolinitic and gibbsitic soils are the most dominant in tropical environments (Schwertmann and Herbillon, 2015), which corroborates our findings. Demattê et al. (2014) built a qualitative soil spectral library, and highlighted that the morphological features to identify organic matter are between 350 and 1100 nm, which comprises the visible up to the shortwave infrared ranges. This same trend was pointed out by Rossel et al. (2016), who created the global soil spectral library. The pH(H_2O) does not have active spectral features, but it is still measurable in the Vis-NIR because of the intrinsic relationship with the primary constituents of soils such as clay (Ge et al., 2020). As explained before, the absorption features at this wavelength present a H-O-H vibration of water that was in the mineral structure or absorbed on its surfaces (Rossel et al., 2016).

The inverse correlation between the SWIR spectral reflectance values and clay and SOC (Figure 3d) corroborated with the synthesis of high clay and SOC contents display high absorption spectral behavior (Clark et al., 1990; Demattê et al., 2019; Liu et al., 2020). The positive correlation of pH(H_2O) with the SWIR spectrum was because of the water molecular vibration, and the sand was caused due to the higher presence of quartz. Soil samples with high sand content present high reflectance (Demattê et al., 2018, 2019). Rossel et al. (2017), monitoring pH and SOC stocks in Northern New South Wales-Australia, obtained RMSE of 0.70 and 0.41 % and R^2 values of 0.71 and 0.81, respectively, using the Cubist machine learning algorithm with 317 samples to creating a local library for a 600 hectare. Demattê et al. (2019) produced the Brazilian Soil Spectral Library and evaluated some preprocessing and quantitative models for soil spectra at State, Region and National scales, found RMSE of 1.38 and R^2 of 0.29, and RPIQ of 1.68 for pH predictions using the cubist model and the CR preprocessing procedure in the Piauí State. Thus, our regional soil spectral library proves to be more accurate because of the sampling density and distribution influence by the greater representation of soils in the landscape. For the prediction models of sand, even though the R^2 and RPIQ values were lower than those presented by Demattê et al. (2019) (RMSE = 169.40 g kg^{-1}; R^2 = 0.80; and RPIQ = 3.83), the RMSE was lower in our study. The lower results in our findings for those parameters (e.g., R^2 and RPIQ) could be explained by the fact the equipment used in this study does not have the Vis-NIR window from 350–1000 nm. This spectral range strongly helps to identify the oxides, soil color and particles, and SOC (Clark et al., 1990; Ben-Dor and Banin, 1995; Ben-Dor et al., 2015; Nocita et al., 2015).

As it was for sand predictions, the prediction models of clay presented lower RMSE comparing to those findings reported by Demattê et al. (2019), which presented RMSE, R^2, and RPIQ values of 105.00 g kg^{-1}, 0.5, and 1.58, respectively. Our study presented better predictive metrics (e.g., RMSE) because the reduction in the number of samples undesirably interferes with the model performance for areas with high spectral and pedological variation (Moura-Bueno et al., 2020). Those authors had only 67 observations from Piauí State and thus, presented higher R^2 and RMSE. To improve the predictive power, it is probably necessary to add the Vis-NIR (350–1000 nm) or increase sampling density. In our specific case, better metric parameters of the clay and sand predictions

could be achieved by increasing sampling density. Most of the studies reported the soil organic matter or carbon predictions using Vis-NIR spectroscopy with transformed data by van Bemmelen's factor (Rossel et al., 2016; Demattê et al., 2019). Rossel et al. (2006) reported RMSE values of 0.36 dag kg^{-1} for organic carbon predictions in the Vis-NIR range (350–1050 nm). As mentioned before, the most suitable ranges to predict SOC is in the Vis-NIR (350–1050 nm) and potentially in the MIR range (2500–25000 nm), as reported by Salazar et al. (2020) and Silvero et al. (2020). However, we found RMSE of 0.49 dag kg^{-1} and R^2 of 0.42 using only the SWIR window and RMSE of 0.64 dag kg^{-1} using the same preprocessing and modeling procedure that those authors (e.g., Cubist and CR). Thus, in our study, we proved that if the sensor has a limited range of the Vis-NIR spectrum, it is still viable to use only the SWIR range to quantify some soil attributes. The raw data preprocessed via SNV improved spectral models of pH, sand, and clay predictions. Nevertheless, the best overall results were without preprocessing the soil spectra for SOC and pH. The most accurate model was the cubist, and that is the most applied model in the literature for soil predictions using Vis-NIR spectroscopy (Rossel et al., 2016; Demattê et al., 2019). As proved by Ng et al. (2019), soil spectral libraries with up to 2000 observations fit better the models using cubist and over 2000 observations it suggests to use a convolutional neural network. The performance overall results for pH(H$_2$O), sand, clay, and SOC were in according to those ranges reported by Soriano-Disla et al. (2014), Rossel et al. (2016), and Demattê et al. (2019).

CONCLUSION

This study demonstrated that the Standard Normal Variate was one of the most promising preprocessing procedures to improve model predictions for pH(H$_2$O) and sand, and SGD for clay. The best overall results were without preprocessing the soil spectra for SOC and pH. Moreover, the cubist model was the most accurate in predicting soil properties. Finally, our study showed evidence of the potential and feasibility of using only the SWIR spectral window to estimate soil properties such as pH(H$_2$O), sand, clay, and SOC. The regional soil spectral preprocessing recommendations for SWIR spectroscopy laboratory analysis to be applied in northeastern Brazil is to preprocess the raw/original data using SNV for pH, sand, and clay, and use the raw/original data for SOC. Furthermore, the main machine learning algorithm suggested is cubist. However, as the spectral library increases the number of observations, it is advised to perform a new evaluation using the convolutional neural network.

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS

Conceptualization: ⓘD Wanderson de Sousa Mendes (lead) and ⓘD Cácio Luiz Boechat (supporting).

Methodology: ⓘD Wanderson de Sousa Mendes (lead) and ⓘD Cácio Luiz Boechat (supporting).

Software: ⓘD Wanderson de Sousa Mendes (lead).

Validation: ⓘD Wanderson de Sousa Mendes (lead).

Formal analysis: ⓘD Wanderson de Sousa Mendes (lead).

Investigation: Antonny Francisco Sampaio de Sena (equal) and Lizandra de Sousa Luz Duarte (equal).

Resources: Cácio Luiz Boechat (lead).

Data curation: Wanderson de Sousa Mendes (lead).

Writing – original draft: Wanderson de Sousa Mendes (lead), Adriano Venicius Santana Gualberto (equal), and Cácio Luiz Boechat (equal).

Writing – review and editing: Wanderson de Sousa Mendes (equal), Cácio Luiz Boechat (equal), Adriano Venicius Santana Gualberto (equal), Ronny Sobreira Barbosa (equal), Yuri Jacques Agra Bezerra da Silva (equal), Paloma Cunha Saraiva (equal), Antonny Francisco Sampaio de Sena (equal), and Lizandra de Sousa Luz Duarte (equal).

Visualization: Wanderson de Sousa Mendes (equal), and Cácio Luiz Boechat (equal).

Supervision: Wanderson de Sousa Mendes (lead).

Project administration: Cácio Luiz Boechat (lead).

Funding acquisition: Cácio Luiz Boechat (lead).

REFERENCES

Afonso JC, Noronha LA, Felipe RP, Freidinger N. Gerenciamento de resíduos laboratoriais: recuperação de elementos e preparo para descarte final. Quim Nova. 2003;26:602-11. https://doi.org/10.1590/S0100-40422003000400027

Alvares CA, Stape JL, Sentelhas PC, Gonçalves JLM, Sparovek G. Köppen's climate classification map for Brazil. Meteorol Zeitschrift. 2013;22:711-28. https://doi.org/10.1127/0941-2948/2013/0507

Amundson R, Berhe AA, Hopmans JW, Olson C, Sztein AE, Sparks DL. Soil and human security in the 21st century. Science. 2015;348:1261071. https://doi.org/10.1126/science.1261071

Barnes RJ, Dhanoa MS, Lister SJ. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. Appl Spectrosc. 1989;43:772-7. https://doi.org/10.1366/0003702894202201

Bellon-Maurel V, Fernandez-Ahumada E, Palagos B, Roger J-MM, McBratney A. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. TrAC - Trends Anal Chem. 2010;29:1073-81. https://doi.org/10.1016/j.trac.2010.05.006

Ben-Dor E, Banin A. Quantitative analysis of convolved Thematic Mapper spectra of soils in the visible near-infrared and shortwave-infrared spectral regions (0·4–2·5 μm). Int J Remote Sens. 1995;16:3509-28. https://doi.org/10.1080/01431169508954643

Ben-Dor E, Ong C, Lau IC. Reflectance measurements of soils in the laboratory: Standards and protocols. Geoderma. 2015;245-246:112-24. https://doi.org/10.1016/j.geoderma.2015.01.002

Bennett JM, McBratney A, Field D, Kidd D, Stockmann U, Liddicoat C, Grover S. Soil security for Australia. Sustainability. 2019;11:3416. https://doi.org/10.3390/su11123416

Beretta AN, Silbermann AV, Paladino L, Torres D, Bassahun D, Musselli R, García-Lamohte A. Soil texture analyses using a hydrometer: modification of the Bouyoucos method. Cienc Investig Agrar. 2014;41:263-71. https://doi.org/10.4067/S0718-16202014000200013

Borges CS, Weindorf DC, Carvalho GS, Guilherme LRG, Takayama T, Curi N, Lima GJEO, Ribeiro BT. Foliar elemental analysis of Brazilian crops via portable X-ray fluorescence spectrometry. Sensors. 2020;20:2509-25. https://doi.org/10.3390/s20092509

Breiman L. Random forests. Mach Learn. 2001;45:5-32.

Cargnelutti Filho A, Storck L, Bartz HR. Estatísticas dos resultados das análises de laboratório de solo. Cienc Rural. 1996;26:401-6.https://doi.org/10.1590/S0103-84781996000300010

Clark RN, King TVV, Klejwa M, Swayze GA, Vergo N. High spectral resolution reflectance spectroscopy of minerals. J Geophys Res. 1990;95:12653-80. https://doi.org/10.1029/jb095ib08p12653

Clark RN, Roush TL. Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. J Geophys Res Solid Earth. 1984;89:6329-40. https://doi.org/10.1029/JB089iB07p06329

Crucil G, Castaldi F, Aldana-Jague E, van Wesemael B, Macdonald A, Van Oost K. Assessing the performance of UAS-Compatible multispectral and hyperspectral sensors for soil organic carbon prediction. Sustainability. 2019;11:1889-906. https://doi.org/10.3390/su11071889

Day PR. Particle fractionation and particle-size analysis. In: Black CA, editor. Methods of soil analysis: Physical and mineralogical properties, including statistics of measurement and sampling. Chichester, UK: John Wiley & Sons, Inc.; 2015. Part 1. p. 545-67.

Demattê JAM, Bellinaso H, Romero DJ, Fongaro CT. Morphological Interpretation of Reflectance Spectrum (MIRS) using libraries looking towards soil classification. Sci Agric. 2014;71:509-20. https://doi.org/10.1590/0103-9016-2013-0365

Demattê JAM, Dotto AC, Paiva AFS, Sato MV, Dalmolin RSD, Araújo MSB, Silva EB, Nanni MR, ten Caten A, Noronha NC, Lacerda MPC, Araújo Filho JC, Rizzo R, Bellinaso H, Francelino MR, Schaefer CEGR, Vicente LE, Santos UJ, Sá EBS, Menezes RSC, Souza JJLL, Abrahão WAP, Coelho RM, Grego CR, Lani JL, Fernandes AR, Gonçalves DAM, Silva SHG, Menezes MD, Curi N, Couto EG, Anjos LHC, Ceddia MB, Pinheiro ÉFM, Grunwald S, Vasques GM, Marques Júnior J, Silva AJ, Barreto MCV, Nóbrega GN, Silva MZ, Souza SF, Valladares GS, Viana JHM, Silva Terra F, Horák-Terra I, Fiorio PR, Silva RC, Frade Júnior EF, Lima RHC, Alba JMF, Souza Junior VS, Brefin MDLMS, Ruivo MDLP, Ferreira TO, Brait MA, Caetano NR, Bringhenti I, Sousa Mendes W, Safanelli JL, Guimarães CCB, Poppiel RR, Souza AB, Quesada CA, Couto HTZ. The brazilian soil spectral library (BSSL): A general view, application and challenges. Geoderma. 2019;354:113793. https://doi.org/10.1016/j.geoderma.2019.05.043

Demattê JAM, Fongaro CT, Rizzo R, Safanelli JL. Geospatial soil sensing system (GEOS3): A powerful data mining procedure to retrieve soil spectral reflectance from satellite images. Remote Sens Environ. 2018;212:161-75. https://doi.org/10.1016/j.rse.2018.04.047

Demattê JAM, Horák-Terra I, Beirigo RM, Terra FS, Marques KPP, Fongaro CT, Silva AC, Vidal-Torrado P. Genesis and properties of wetland soils by VIS-NIR-SWIR as a technique for environmental monitoring. J Environ Manage. 2017;197:50-62. https://doi.org/10.1016/J.JENVMAN.2017.03.014

Di Iorio E, Circelli L, Lorenzetti R, Costantini EAC, Egendorf SP, Colombo C. Estimation of andic properties from Vis-NIR diffuse reflectance spectroscopy for volcanic soil classification. Catena. 2019;182:104109-22. https://doi.org/10.1016/j.catena.2019.104109

Dotto AC, Dalmolin RSD, ten Caten A, Gris DJ, Ruiz LFC. AlradSpectra: a quantification tool for soil properties using spectroscopic data in R. Rev Bras Cienc Solo. 2019;43:e0180263. https://doi.org/10.1590/18069657rbcs20180263

Ergon R. Principal component regression (PCR) and partial least squares regression (PLSR). In: Granato D, Ares G, editors. Mathematical and statistical methods in food science and technology. Chichester, UK: John Wiley & Sons, Inc.; 2013. p. 121-42.

Fang Q, Hong H, Zhao L, Kukolich S, Yin K, Wang C. Visible and near-infrared reflectance spectroscopy for investigating soil mineralogy: A review. J Spectrosc. 2018;2018:3168974. https://doi.org/10.1155/2018/3168974

Ge Y, Morgan CLS, Wijewardane NK. Visible and near-infrared reflectance spectroscopy analysis of soils. Soil Sci Soc Am J. 2020;84:1495-502. https://doi.org/10.2136/msa2017.0040

Grunwald S, Vasques GM, Rivero RG. Fusion of soil and remote sensing data to model soil properties. Adv Agron. 2015;131:1-109. https://doi.org/10.1016/bs.agron.2014.12.004

Guevara M, Arroyo C, Brunsell N, Cruz CO, Domke G, Equihua J, Etchevers J, Hayes D, Hengl T, Ibelles A, Johnson K, Jong B, Libohova Z, Llamas R, Nave L, Ornelas JL, Paz F, Ressl R, Schwartz A, Victoria A, Wills S, Vargas R. Soil organic carbon across Mexico and the conterminous United States (1991–2010). Global Biogeochem Cy. 2020;34:e2019GB006219. https://doi.org/10.1029/2019GB006219

IUSS Working Group WRB. World reference base for soil resources 2014, update 2015: International soil classification system for naming soils and creating legends for soil maps. Rome: Food and Agriculture Organization of the United Nations; 2015. (World Soil Resources Reports, 106).

Jaconi A, Vos C, Don A. Near infrared spectroscopy as an easy and precise method to estimate soil texture. Geoderma. 2019;337:906-13. https://doi.org/10.1016/J.GEODERMA.2018.10.038

Kuhn M. Building predictive models in R using the caret Package. J Stat Softw. 2008;28:1-26. https://doi.org/10.18637/jss.v028.i05

Leite LFC, Galvão SRS, Holanda Neto MR, Araújo FS, Iwata BF. Atributos químicos e estoques de carbono em Latossolo sob plantio direto no cerrado do Piauí. Rev Bras Eng Agric Ambient. 2010;14:1273-80. https://doi.org/10.1590/S1415-43662010001200004

Liu J, Xie J, Han J, Wang H, Sun J, Li R, Li S. Visible and near-infrared spectroscopy with chemometrics are able to predict soil physical and chemical properties. J Soils Sediments. 2020;20:2749-60. https://doi.org/10.1007/s11368-020-02623-1

López-Maestresalas A, Keresztes JC, Goodarzi M, Arazuri S, Jarén C, Saeys W. Non-destructive detection of blackspot in potatoes by Vis-NIR and SWIR hyperspectral imaging. Food Control. 2016;70:229-41. https://doi.org/10.1016/j.foodcont.2016.06.001

Marques KPP, Rizzo R, Dotto AC, Souza AB, Mello FAO, Neto LGM, Anjos LHC, Demattê JAM. How qualitative spectral information can improve soil profile classification? J Near Infrared Spectrosc. 2019;27:156-74. https://doi.org/10.1177/0967033518821965

Mendes WS, Medeiros Neto LG, Demattê JAM, Gallo BC, Rizzo R, Safanelli JL, Fongaro CT. Is it possible to map subsurface soil attributes by satellite spectral transfer models? Geoderma. 2019;343:269-79. https://doi.org/10.1016/j.geoderma.2019.01.025

Minasny B, McBratney AB, Wadoux AMJ-C, Akoeb EN, Sabrina T. Precocious 19th century soil carbon science. Geoderma Reg. 2020;22:e00306. https://doi.org/10.1016/j.geodrs.2020.e00306

Moura-Bueno JM, Dalmolin RSD, Horst-Heinen TZ, ten Caten A, Vasques GM, Dotto AC, Grunwald S. When does stratification of a subtropical soil spectral library improve predictions of soil organic carbon content? Sci Total Environ. 2020;737:139895. https://doi.org/10.1016/j.scitotenv.2020.139895

Ng W, Minasny B, Mendes W de S, Demattê JAM. Estimation of effective calibration sample size using visible near infrared spectroscopy: deep learning vs machine learning. Soil Discuss. 2019;48:1-21. https://doi.org/10.5194/soil-2019-48

Nocita M, Stevens A, van Wesemael B, Aitkenhead M, Bachmann M, Barthès B, Ben Dor E, Brown DJ, Clairotte M, Csorba A, Dardenne P, Demattê JAM, Genot V, Guerrero C, Knadel M, Montanarella L, Noon C, Ramirez-Lopez L, Robertson J, Sakai H, Soriano-Disla JM, Shepherd KD, Stenberg B, Towett EK, Vargas R, Wetterlind J. Soil Spectroscopy: An alternative to wet chemistry for soil Monitoring. Adv Agron. 2015;132:139-59. https://doi.org/10.1016/bs.agron.2015.02.002

Ogen Y, Goldshleger N, Ben-Dor E. 3D spectral analysis in the VNIR–SWIR spectral region as a tool for soil classification. Geoderma. 2017;302:100-10. https://doi.org/10.1016/J.GEODERMA.2017.04.020

Oliveira GC, Francelino MR, Arruda DM, Fernandes-Filho EI, Schaefer CEGR. Climate and soils at the Brazilian semiarid and the forest-Caatinga problem: new insights and implications for conservation. Environ Res Lett. 2019;14:104007. https://doi.org/10.1088/1748-9326/ab3d7b

Quinlan JR. C4.5: programs for machine learning. San Francisco, Calif.: Morgan Kaufmann Publishers; 1993.

R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria; 2020. Available from: http://www.R-project.org/.

Rawls WJ, Pachepsky YA, Ritchie JC, Sobecki TM, Bloodworth H. Effect of soil organic carbon on soil water retention. Geoderma. 2003;116:61-76. https://doi.org/10.1016/S0016-7061(03)00094-6

Rede oficial de laboratórios de análise de solo e de tecido vegetal dos estados do Rio Grande do Sul e de Santa Catarina - ROLAS. Passo Fundo: Laboratório de Solos da Embrapa Trigo / Sociedade Brasileira de Ciência do Solo, Núcleo Regional Sul; 2019 [cited 2020 Jun 20]. Available from: https://rolas.cnpt.embrapa.br/publico/pNumAmostrasAnalisadas.

Rossel RAV, Behrens T, Ben-Dor E, Brown DJ, Demattê JAM, Shepherd KD, Shi Z, Stenberg B, Stevens A, Adamchuk V, Aïchi H, Barthès BG, Bartholomeus HM, Bayer AD, Bernoux M, Böttcher K, Brodský L, Du CW, Chappell A, Fouad Y, Genot V, Gomez C, Grunwald S, Gubler A, Guerrero C, Hedley CB, Knadel M, Morrás HJM, Nocita M, Ramirez-Lopez L, Roudier P, Campos EMR, Sanborn P, Sellitto VM, Sudduth KA, Rawlins BG, Walter C, Winowiecki LA, Hong SY, Ji W. A global spectral library to characterize the world's soil. Earth-Science Rev. 2016;155:198-230. https://doi.org/10.1016/j.earscirev.2016.01.012

Rossel RAV, Lobsey CR, Sharman C, Flick P, Mclachlan G. Novel proximal sensing for monitoring soil organic C stocks and condition. Environ Sci Technol. 2017;51:5630-41. https://doi.org/10.1021/acs.est.7b00889

Rossel RAV, 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

Salazar DFU, Demattê JAM, Vicente LE, Guimarães CCB, Sayão VM, Cerri CEP, Padilha MC, Mendes WDS. Emissivity of agricultural soil attributes in southeastern Brazil via terrestrial and satellite sensors. Geoderma. 2020;361:114038. https://doi.org/10.1016/j.geoderma.2019.114038

Santos HG, Jacomine PKT, Anjos LHC, Oliveira VA, Lumbreras JF, Coelho MR, Almeida JA, Araújo Filho JC, Oliveira JB, Cunha TJF. Sistema brasileiro de classificação de solos. 5. ed. rev. ampl. Brasília, DF: Embrapa; 2018.

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

Schwertmann U, Herbillon AJ. Some aspects of fertility associated with the mineralogy of highly weathered tropical soils. In: Lal R, Sanchez PA, editors. Myths and science of soils of the tropics. Chichester, UK: John Wiley & Sons, Inc.; 2015. p. 47-59.

Silvero NEQ, Di Raimo LADL, Pereira GS, Magalhães LP, Terra FS, Dassan MAA, Salazar DFU, Demattê JAM. Effects of water, organic matter, and iron forms in mid-IR spectra of soils: Assessments from laboratory to satellite-simulated data. Geoderma. 2020;375:114480. https://doi.org/10.1016/j.geoderma.2020.114480

Soong JL, Janssens IA, Grau O, Margalef O, Stahl C, Van Langenhove L, Urbina I, Chave J, Dourdain A, Ferry B, Freycon V, Herault B, Sardans J, Peñuelas J, Verbruggen E. Soil properties explain tree growth and mortality, but not biomass, across phosphorus-depleted tropical forests. Sci Rep. 2020;10:2302. https://doi.org/10.1038/s41598-020-58913-8

Soriano-Disla JM, Janik LJ, Rossel RAV, Macdonald LM, McLaughlin MJ. The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties. Appl Spectrosc Rev. 2014;49:139-86. https://doi.org/10.1080/05704928.2013.811081

Stoner ER, Baumgardner MF. Characteristic variations in reflectance of surface soils. Soil Sci Soc Am J. 1981;45:1161-5. https://doi.org/10.2136/sssaj1981.03615995004500060031x

Syers JK, Campbell AS, Walker TW. Contribution of organic carbon and clay to cation exchange capacity in a chronosequence of sandy soils. Plant Soil. 1970;33:104-12. https://doi.org/10.1007/BF01378202

van Raij B, Andrade JCH, Cantarella Quaggio JA. Análise química para avaliação de solos tropicais. Campinas: IAC; 2001.

Vereecken H, Schnepf A, Hopmans JW, Javaux M, Or D, Roose T, Vanderborght J, Young MH, Amelung W, Aitkenhead M, Allison SD, Assouline S, Baveye P, Berli M, Brüggemann N, Finke P, Flury M, Gaiser T, Govers G, Ghezzehei T, Hallett P, Hendricks Franssen HJ, Heppell J, Horn R, Huisman JA, Jacques D, Jonard F, Kollet S, Lafolie F, Lamorski K, Leitner D, McBratney A, Minasny B, Montzka C, Nowak W, Pachepsky Y, Padarian J, Romano N, Roth K, Rothfuss Y, Rowe EC, Schwen A, Šimůnek J, Tiktak A, Van Dam J, van der Zee SEATM, Vogel HJ, Vrugt JA, Wöhling T, Young IM. Modeling soil processes: Review, key challenges, and new perspectives. Vadose Zone J. 2016;15:vzj2015.09.0131. https://doi.org/10.2136/vzj2015.09.0131

Vinther FP, Hansen EM, Eriksen J. Leaching of soil organic carbon and nitrogen in sandy soils after cultivating grass-clover swards. Biol Fertil Soils. 2006;43:12-9. https://doi.org/10.1007/s00374-005-0055-4

Walkley A, Black IA. An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. Soil Sci. 1934;37:29-38.

Wilczyński S, Koprowski R, Marmion M, Duda P, Błońska-Fajfrowska B. The use of hyperspectral imaging in the VNIR (400–1000 nm) and SWIR range (1000–2500 nm) for detecting counterfeit drugs with identical API composition. Talanta. 2016;160:1-8. https://doi.org/10.1016/j.talanta.2016.06.057