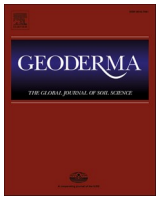
Geoderma 457 (2025) 117288

Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma



Integrating satellite radar vegetation indices and environmental descriptors with visible-infrared soil spectroscopy improved organic carbon prediction in soils of semi-arid Brazil

Erli Pinto dos Santos [a,*], Michel Castro Moreira [a], Elpídio Inácio Fernandes-Filho [b], José A.M. Demattê [c], Uemeson José dos Santos [d], Jean Michel Moura-Bueno [f], Renata Ranielly Pedroza Cruz [e], Demetrius David da Silva [a], Everardo Valadares de Sá Barreto Sampaio [g]

[a] Department of Agricultural Engineering, Federal University of Viçosa, Peter Henry Rolfs Avenue, Viçosa 36570-900 Minas Gerais, Brazil
[b] Department of Soils, Federal University of Viçosa, Peter Henry Rolfs Avenue, Viçosa 36570-900 Minas Gerais, Brazil
[c] Department of Soil Science, Luiz de Queiroz College of Agriculture, Universidade de São Paulo, Pádua Dias Avenue, Piracicaba 13418-900, Brazil
[d] Federal Institute of Education, Science, and Technology of Pará, Campus Óbidos, Rodovia PA 437, km 02, Óbidos 68250-000 Pará, Brazil
[e] Department of Agronomy, Federal University of Viçosa, Peter Henry Rolfs Avenue, Viçosa 36570-900 Minas Gerais, Brazil
[f] Soil Science Department, Federal University of Santa Maria, Roraima Avenue, 1000, Santa Maria 97105-900 Rio Grande do Sul, Brazil
[g] Department of Nuclear Energy, Federal University of Pernambuco, Professor Luís Freire Avenue, 1000 Recife, Pernambuco, Brazil

ARTICLE INFO

Handling Editor: Dr Budiman Minasny

Keywords:
Soil health
Soil spectral library
Agriculture
Forest
Carbon science
Carbon credits

ABSTRACT

Soil Organic Carbon (SOC) is a paramount soil attribute for climate regulation, soil fertility, and agricultural productivity. The global demand for SOC testing came in response to expanding soil management practices aimed at ensuring soil health. This study explores enhanced accuracy in predicting SOC using soil spectroscopy (proximal sensing). A Soil Spectral Library (SSL), made from 127 soil profiles in Northeast Brazil, mainly by using soils from a semi-arid region, was used. Four modeling scenarios were employed, incorporating distinct covariable sets: 1) diffuse reflectance from laboratory spectroscopy (SSL); 2) diffuse reflectance and radar vegetation indices from all-weather and globally available Sentinel-1 satellite data; 3) diffuse reflectance and environmental factors; 4) all covariables. Integration of radar vegetation indices and environmental factors significantly improved SOC estimates by soil spectroscopy. Predicting SOC solely from SSL reflectance data yielded an average RMSE of 4.54 g kg^{-1} and R^2 of 0.62. However, by using all covariables significantly reduced RMSE by approximately 13 % (to 3.94 g kg^{-1}) and increased R^2 by 14 % (to 0.71). This comprehensive approach, combining SSL, satellite radar vegetation indices, and environmental variables, substantially advances SOC spectroscopic prediction accuracy, offering valuable insights for applications in agriculture and environmental monitoring. These findings contribute to the reliability of proximal and remote sensing methodologies in soil testing.

1. Introduction

The soil has ecosystem functions important for the entire biosphere. Among the main ecological functions of soils (besides technical and cultural functions), we have: biomass production, water filtration and storage, nutrient storage and recycling, habitat for biological activity, and carbon storage (Wiesmeier et al., 2019). Soil organic carbon (SOC), which is the main fraction of soil organic matter, is a key soil attribute due to soil carbon storage being important not only for climate regulation but also for affecting all of the soil functions mentioned above (Wiesmeier et al., 2019) including controlling soil fertility and agricultural production (Jobbágy and Jackson, 2000).

The growing awareness of SOC's importance in soil conservation and its critical role in maintaining ecosystem functions has spurred the development and adoption of sustainable practices as key strategies for mitigating greenhouse gas emissions (FAO, 2020; Lal et al., 2018;

* Corresponding author.
E-mail addresses: erlipinto@gmail.com, erli.santos@ufv.br (E.P. dos Santos).

https://doi.org/10.1016/j.geoderma.2025.117288
Received 21 October 2024; Received in revised form 26 February 2025; Accepted 5 April 2025
Available online 9 April 2025

Paustian et al., 2019; Smith et al., 2020). Examples of farm-scale conservation practices include no-tillage, cover cropping, and agroforestry systems. These practices, and their increasing adoption, have heightened the demand for soil monitoring due to their effects on soil attributes such as SOC, soil structure, and salinity, ranging from farm to watershed scales. Spatial and temporal soil monitoring is essential for assessing soil's role in food and fiber production, greenhouse gas mitigation, and water and biodiversity protection (FAO, 2020; FAO, 2024). Furthermore, monitoring, reporting, and verifying SOC levels are critical for implementing and adapting site-specific soil management practices effectively (Angelopoulou et al., 2020).

However, the high global demand for soil analysis consumes a lot of chemical reagents used in the analytical determination of SOC, such as dichromate, ferrous ammonium sulfate, and sulfuric acid (Demattê et al., 2019a). Also, traditional chemical soil analysis methods are time-consuming and expensive. As a result, soil spectroscopy in the visible and near-infrared regions of the spectra is an alternative and complementary method to traditional chemical determinations of SOC. Spectroscopy is a relatively fast, low-cost, non-destructive, proximal soil sensing method (Bellon-Maurel and McBratney 2011; Demattê et al. 2019b; FAO 2020; Viscarra Rossel et al. 2016).

Soil spectroscopy consists of measuring the diffuse reflectance of soil samples, commonly manipulated in the laboratory, and building soil spectral libraries (SSL) (Demattê et al. 2019b; Viscarra Rossel et al., 2010). The method consists of quantifying the reflectance of samples at wavelengths within the visible (Vis: from 350 to 700 nm) and near-infrared (NIR: from 700 to 2500 nm). Although mid-infrared (MIR: 2500 to 25000 nm) can also be measured, it is less common than other spectral regions (de Mendes et al., 2022). SSL contains the reflectance of the samples in the visible and infrared spectral bands, and the chemical and/or physical attributes of the soil samples. With an SSL, the attributes of interest can be modeled with the spectral readings through chemometrics models (Pudełko and Chodak, 2020; Soriano-Disla et al., 2014). Typically, Machine Learning (ML) methodologies are used to build these models, by using regression methods such as Partial Least Squares, Support Vector Machine, Random Forest, and others (Ben-Dor et al., 2009; de Mendes et al., 2022; Moura-Bueno et al., 2019; dos Santos et al., 2022; Soriano-Disla et al., 2014).

Spectroscopy is said to be a promising methodology to increase efficiency and help monitor soil attributes (Nocita et al., 2015) and thus is in constant evolution to improve the accuracy and reliability of estimates. Among the advances in soil spectroscopy, we highlight the use of modeling strategies with independent samples from the same SSL to improve the generalization capacity of the models (Brown et al., 2005; dos Santos et al., 2023a,b; McBride, 2022; Viscarra Rossel et al., 2022). Ways to improve the understanding of soil attribute prediction models with SSL are also necessary and have been studied (McBride, 2022; Viscarra Rossel et al., 2022; Wadoux, 2023). Furthermore, recently, the use of other environmental variables (added to SSL) stood out as a way of increasing the accuracy and precision of models, especially for SOC prediction models (Adi et al., 2019; Moura-Bueno et al., 2021; Sabetizade et al., 2021; Wang et al., 2022).

In this context, Moura-Bueno et al. (2020, 2019) observed that stratifying an SSL into subsets based on homogeneity criteria defined by other environmental variables produced more accurate estimates of SOC. However, the downside of stratifying was the reduced number of samples to calibrate ML models (Moura-Bueno et al., 2020). To address the challenges posed by reduced sample sizes in stratified SSLs, the Cubist regression method, previously applied to global SSL modeling (Viscarra Rossel et al., 2016), has since been adopted in various studies (Adi et al., 2019; Moura-Bueno et al., 2021; Sabetizade et al., 2021; Wang et al., 2022).

The advantage of Cubist lies in its ability to automatically partition samples into subsets based on decision rules derived from continuous and/or categorical variables. This feature allows for more targeted modeling by capturing the variability in soil properties or SOC content across different environmental conditions. Furthermore, continuous variables are used in internal linear regression models to make predictions (Kuhn and Johnson, 2013; Kuhn and Quinlan, 2023; Quinlan, 1992). Notably, the environmental variables added to SSL must be related to soil formation factors and the variability of SOC content in the landscape.

Although there are still no criteria for choosing environmental variables specifically for SSL, to achieve SOC values, such variables must indicate soil formation factors (Moura-Bueno et al., 2021). The possible and available environmental variables must be SCORPAN factors, which describe soil variations in the landscape such as the soil itself, climate, organisms, relief, parental material, age, and spatial position (McBratney et al., 2003; Minasny and McBratney, 2016). Therefore, variables such as soil texture, climate classification, vegetation indices by orbital remote sensing (ORS), classifications of land use and cover, mineralogy, elevation, and others have been used (Adi et al., 2019; Meng et al., 2022; Moura-Bueno et al., 2021, 2020, 2019; Sabetizade et al., 2021; Wang et al., 2022).

Specifically, the vegetation indices by ORS indicated by Sabetizade et al. (2021) for modeling SOC with a national SSL have the advantage of representing both the presence and amount of vegetation (Zeng et al., 2022), which is the main input of organic matter into the soil after deposition and decomposition (Wiesmeier et al., 2019). These variables are continuous and can be used both to stratify an SSL and to predict SOC levels. Unfortunately, the problem with optical satellite products is the uncertainty regarding the availability of scenes due to cloud cover, mainly in tropical regions (Asner, 2001; dos Santos et al., 2022).

Hence, radar ORS is important because it is much less influenced by cloud cover than optical sensors (Flores-Anderson et al., 2019; Woodhouse, 2006). In this context, the vegetation indices for synthetic aperture radar (SAR) sensors emerged to put the high operability of radar to good use as an alternative to spectral vegetation indices through optical sensing. SAR vegetation indices aim to represent the amount of aboveground plant biomass or phenology (Bhogapurapu et al., 2022; dos Santos et al., 2021; Frison et al., 2018; Mandal et al., 2020b, 2020a; Periasamy, 2018). Therefore, they can be alternatives to optical ORS products in modeling SOC through soil spectroscopy, requiring only tests to prove their feasibility.

The work hypothesizes that the accuracy and precision of SOC prediction models using diffuse reflectance spectroscopy in the Vis-NIR spectrum tend to improve when environmental variables are added to the modeling in Cubist models. Specifically, we propose the use of SAR-derived vegetation indices (obtained via orbital radar sensors) alongside other environmental covariates to enhance SOC estimates. To test this approach, we utilized a regional soil spectral library (R-SSL) from Northeastern Brazil and evaluated SOC prediction models based on Vis-NIR. We assessed whether the inclusion of SAR vegetation indices and other environmental variables improves model performance and contributes to a more robust estimation of SOC levels.

2. Methodology

2.1. Study area description

The soil data was surveyed in the State of Pernambuco, Northeast Brazil. The studied soil profiles are between meridians 34° 48′ and 38° 48′ W and between latitudes 7° 18′ and 9° 18′ S (Fig. 1). The surveys were carried out in campaigns from 2011 to 2013.

The studied soil profiles are in three different climatic regions, according to the Köppen climate typology for Brazil (Alvares et al., 2013). From east to west, the first region is the coastal zone, which has an Am climate (tropical zone with monsoons), whose climate is hot and humid, with average annual temperatures varying between 24 and 26 °C and total annual rainfall between 1600 and 2500 mm year^{-1}; the second zone is subhumid and classified as As climate (tropical zone with dry summer), it is a hot and dry region (26–28 °C and 600–1000 mm

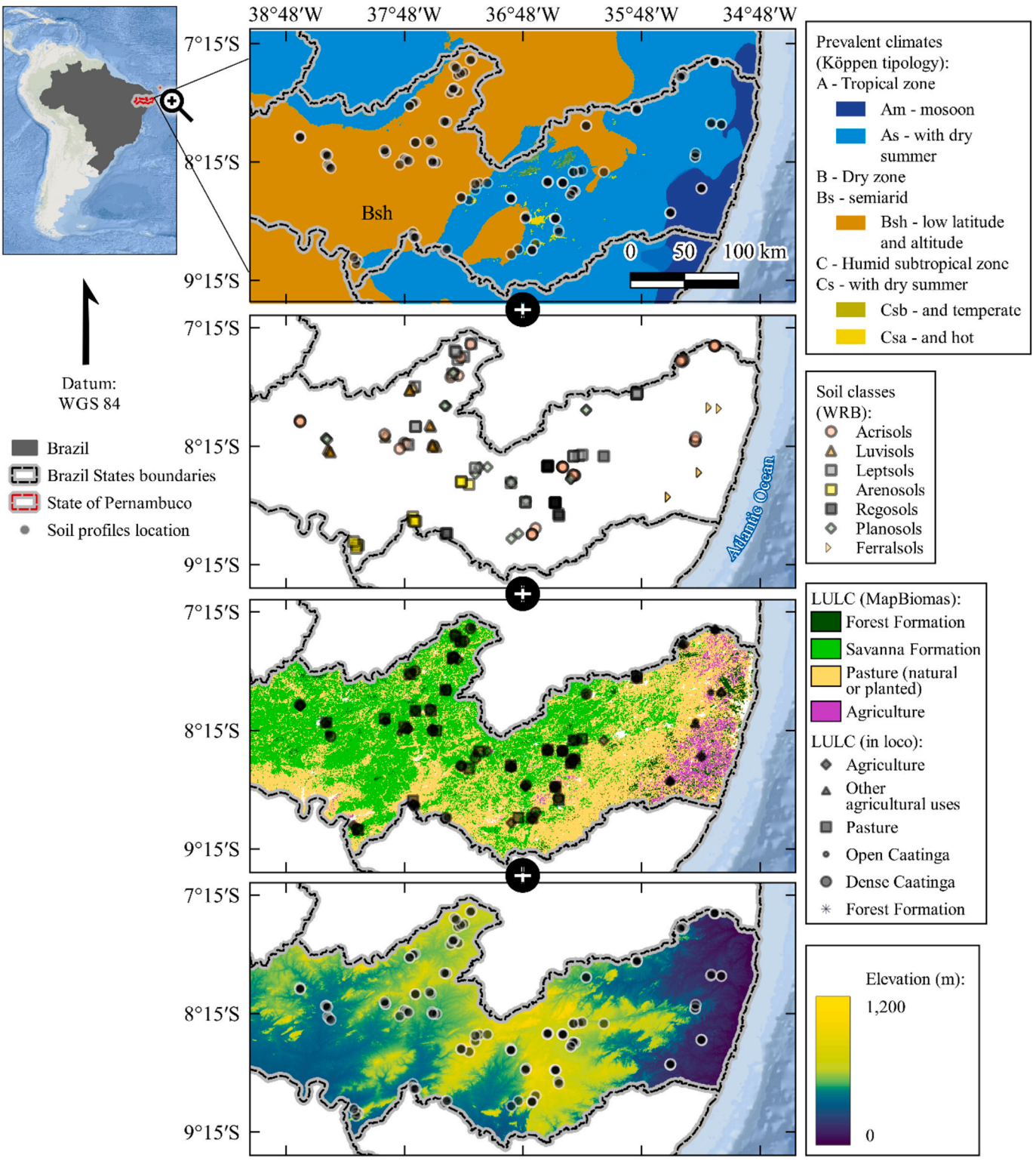


Fig. 1. Location of soil profiles sampled in the study area and their distribution according to the Köppen climate classification (Alvares et al., 2013), their respective soil classes, and land use and land cover classification (LULC) surveyed by MapBiomas (Coleção 6.0 do MapBiomas (Souza et al., 2020)), as well as the terrain elevation (NASA JPL, 2020).

year^{-1}); further west is the semi-arid region, which has a Bsh climate (dry zone, semiarid with low latitude and altitude), with a hot climate (27–29 °C) with low (400–800 mm year^{-1}) and irregular rainfall (dos Santos et al., 2020). Only one soil profile falls under the climate class Csa (humid subtropical zone with hot and dry summer). This class is only found in this region due to its elevation close to 1,000 m.

127 soil profiles were surveyed in different classes of land use and land cover. The profiles were collected in Acrisols (35 profiles), Luvisols (12), Leptsols (23), Arenosols (11), Regosols (12), Planosols (23), and Ferralsols (11).

Regarding the different classes of land use and land cover (LULC) in the Caatinga and Atlantic Forest biomes: Agriculture (31 profiles),

Pasture (27), Other agricultural uses (5), Open Caatinga (31), Dense Caatinga (27) and Forest (forest formation, 6 profiles). The Agriculture class corresponds to rainfed agricultural cultivation, mainly subsistence agriculture with the cultivation (mostly) of corn (*Zea mays* L.), beans (*Phaseolus vulgaris* L.), cowpea beans (*Vigna unquiculata* L. Walp.), and cassava (*Manihot esculenta* Crantz.). The Pasture class represents areas grazed and covered by native or introduced African grasses and other herbaceous plants. The dense Caatinga class represents areas of native Caatinga forest that had substrate and tree shrub covering between 60 and 80 % of the soil surface. The open Caatinga class has less soil covered by trees and shrubs, between 40 and 60 %, due to natural causes or human interference.

2.2. Collection and analysis of soil samples: analytical and spectroscopic determination

Soil samples were collected in trenches of 0.7 × 0.7 m in area in the soil profiles mentioned previously. After removing the superficial litter, the trenches were dug. The samples were collected in seven layers at standard depths (cm): 0 to 10, 10–20, 20–30, 30–40, 40–60, 60–80 and 80–100. Since not all profiles were deeper than or equal to 1 m, a total of 701 samples were collected (n = 701).

The samples underwent an analytical determination of their chemical and physical properties in the laboratory. The samples were dried and sieved through a 2 mm mesh. Soil organic carbon contents (SOC, g kg^{-1}) were determined with subsamples of approximately 10 mg via dry determination, using the TruSpec CHN-analyzer (LECO® 2006, St. Joseph, EUA).

To create the regional spectral library (R-SSL), soil samples were prepared in the laboratory for spectral readings in visible, near-infrared, and short-wave wavelengths (Vis-NIR). Hence, the FieldSpec 3 spectroradiometer (Analytical Spectral Devices Inc., Boulder, CO, USA) was used. The Fieldspec 3 measures radiance from 350 to 2500 nm and has spectral resolutions of 1 nm (from 350 to 700 nm), 3 nm (700 to 1400 nm), and 10 nm (1400 to 2500 nm). The sampling intervals of the output data are 1 nm with 2151 spectral bands. 20 g of each soil sample was placed in Petri dishes and distributed homogeneously on a flat surface. The source of electromagnetic radiation was two halogen lamps (50 W) (both non-collimated and with a zenith angle of 30°) positioned 35 cm from the sample with an angle of 90° between them. A fiber-optic cable, located 8 cm from the center of the sample surface, captured the reflected energy from an area of approximately 2 cm^2. For each sample, the average reflectance was calculated from three repetitions of readings in different positions, decreasing the shading effect. Each repetition consisted of 100 sensor readings, to maximize the signal-to-noise ratio. The instrument was calibrated before sample readings and every 20 min thereafter, using a white *Spectralon* plate.

2.3. Spectral library pre-processing

The original reflectance, without any transformation, of all R-SSL samples was subjected to two pre-processing steps. The first step consisted of smoothing the spectral curves using a moving average filter, for this purpose a convolution function with a band of 4 nm was used. The second step consisted of calculating the normalized reflectance from the smoothed curves by using the continuum removal algorithm of Clark and Roush (1984).

The normalized reflectance spectral curves in Vis-NIR were used to model and predict SOC contents. This was defined because the continuum removal technique highlights light absorption features by organic compounds (Viscarra Rossel et al., 2016) and performed best in modeling SOC contents with the same R-SSL (dos Santos et al., 2023a,b; dos Santos et al., 2020).

2.4. Model's environmental variables

In addition to the normalized reflectance in Vis-NIR, other environmental variables were used to predict SOC levels. The selected environmental variables are related to the formation and variation factors of SOC in the landscape, as recommended by (Moura-Bueno et al., 2021).

The set of environmental covariates can be divided between categorical and continuous. The categorical covariates used were the LULC classes and soil types (obtained from the field survey, Fig. 1), and the climate class of each soil profile. The climate classification was obtained from the Köppen classification mapping for Brazil (Alvares et al., 2013). The continuous covariates are elevation, which was obtained from the NASADEM digital elevation model (NASA JPL, 2020), and SAR vegetation indices. Table 1 summarizes all environmental variables used in the modeling process, and further details regarding the SAR vegetation indices are presented in the subsequent section.

2.5. Obtaining SAR vegetation indices from the Sentinel-1 mission

In this study, orbital remote sensing products were also used to predict SOC contents. The satellite data used in this study came from images of the SAR (Synthetic Aperture Radar) sensor on board the orbital platforms of the European Space Agency (ESA) Sentinel-1 mission. The Sentinel-1 images were used to obtain vegetation indices proposed in the literature for this sensor.

The Sentinel-1 mission satellites operate with SAR-type imaging radar sensors. These are active sensors, which in the case of Sentinel-1 satellites, operate in the C band (with a wave frequency of ≅ 5.4*GHz*) (ESA, 2012). Dual polarization images from the *Interferometric Wide Swath* (IW) imaging mode were used, which are pre-processed images with only the observed wave amplitude information, called GRD (Ground Range Detected) products.

Table 1
Summary of all environmental variables used in the study and their source.

Environmental variable	Variable data type	Dummies covariables for modeling	Spatial resolution	Source of the data
Land use and land cover (LULC)	Discrete/ Categorical (classes used as dummy covariables)	Pasture, Forest Formation, Farming, Agriculture, Open Caatinga, and Dense Caatinga	N/A	Field surveying
Climate	Discrete/ Categorical (classes used as dummy covariables)	Am, As, Csa, and BSh	100 m	Alvarez et al. (2013)
Soil classes	Discrete/ Categorical (classes used as dummy covariables)	Ferralsols, Regosols, Arenosols, Leptsols, Luvisols, and Acrisols	N/A	Soil surveying
Elevation	Continuous	–	30 m	NASADEM: reprocessing of Shuttle Radar Topography Mission (NASA JPL, 2020)
SAR vegetation indices	Continuous	–	10 m*	Sentinel-1 ESA mission (ESA, 2012)

Note: * pixel size of 10 m and variable spatial resolution; N/A: not available, since the data comes from field surveying, stead of any geospatial dataset.

Since the Sentinel-1 orbital mission started in 2014 (ESA, 2022), the Sentinel-1 IW GRD images obtained for the study were multi-temporal to the soil surveys. Although the survey was carried out from 2011 to 2013, the images that covered the entire study area (first selection criteria and scenes) were from 2017. Therefore, 28 products were used, from the relative orbits of numbers 9 and 82. The images were selected from dates (month) compatible with the dates of the survey (second criterion). This choice was based on studies that used products from remote sensing platforms to model SOC (mainly Sentinel satellites) and used observations from different dates than the soil surveys for SOC analysis (Kunkel et al., 2022; Shafizadeh-Moghadam et al., 2022; Sothe et al., 2022; Zhou et al., 2020a, 2020b). This methodology is common and valid when there are no significant changes in the factors that control SOC levels in the soil, such as changes in land use and cover (Wiesmeier et al., 2019), which was not observed for the study area.

To ensure no significant LULC changes occurred in the studied soil locations, an evolution analysis was conducted. It is available in the Supplementary Material of the article, summarized in Fig. S1. Details of the Sentinel-1 IW GRD images can also be found in the Supplementary Material (Table S1).

Sentinel-1 IW GRD images have two wave polarizations with greater global coverage. These polarizations are the VH, whose sensor emits a pulse of radiation in vertical polarization and measures the reflectivity detected in horizontal polarization, and VV, whose emission and detection are in vertical polarization. Dual polarization Sentinel-1 IW GRD images are required to determine Sentinel-1 SAR vegetation indices.

Sentinel-1 IW GRD images are formed after the sensor scans the Earth's surface over a wide swath of 250 km. This broad band is made up of three subbands (IW1, IW2, and IW3) obtained by the TOPSAR method – Terrain Observation with Progressive Scans SAR (De Zan and Guarnieri, 2006). Subbands IW1, IW2, and IW3 are scanned with incidence angles of 32.9°, 28.3°, and 43.1°, respectively. Furthermore, the spatial resolution of the images (in the range × azimuth directions) of the IW1, IW2, and IW3 subranges are 20.4 m × 22.5 m, 20.3 × 22.6, 20.5 × 22.6, respectively (ESA, 2022). The scenes used are from the Alaska Satellite Facility (ASF, 2022) portal due to their accessibility.

After the acquisition, the Sentinel-1 IW GRD products were processed using the following algorithms:

1) *Apply Orbit File*: obtains accurate satellite orbit and velocity vectors and generates accurate georeferenced images;
2) *Thermal Noise Removal*: removes antenna thermal noise that affects images;
3) *Border Noise Removal*: removes noise at the images' edges;
4) *Radiometric Calibration*: normalizes the amplitude of each polarization for a Radar Cross Section (RCS) and obtains the backscatter coefficient (reflectivity per unit area) in β^0 (RCS required to perform terrain corrections);
5) *Despeckling*: applies the Speckle noise filter using the Lee Sigma filter with a window of 11 × 11 pixels (*sigma* = 0.9);
6) *Radiometric Terrain Flattening (RTF)*: mitigates distortions in backscatter that are likely to occur due to the relief (slopes, hills, etc.) and the operating geometry of SAR sensors (of the side-looking type) (Small, 2011). At this stage, the digital elevation model used to represent the relief was the Copernicus 30 m Global. After the RTF algorithm, the symbology of the backscattering coefficient (reflectance for radar) is transformed from β^0 to γ^0 (more details can be found in the methodology of Small (2011)).
7) *Range-Doppler Terrain Correction*: orthorectification of the images from the Copernicus 30 m Global.

The described steps of the digital processing of Sentinel-1 images are necessary to transform the wave amplitude detected into a backscatter coefficient. Although the IW images have different spatial resolutions in terms of range and azimuth, after orthorectification the output image has a pixel size of 10 × 10 m. More details of the Sentinel-1 IW GRD image processing steps can be obtained in the texts Filipponi (2019) and dos Santos et al. (2021). Once we had the images with VH and VV polarizations calibrated to the backscatter coefficient at γ^0, the next step was to calculate the SAR vegetation indices.

Following the methodology of Santos et al. (2023), five products from Sentinel-1 images were used to predict SOC contents, such as the VH polarization, which is itself a polarimetric indicator of the presence of vegetation and the amount of aboveground biomass (da Bispo et al., 2020; Joshi et al., 2017; Saatchi, 2019; Santoro et al., 2021; Woodhouse, 2006), and four other vegetation indices. The vegetation indices calculated were: the Cross-ratio (CR), the Dual-polarization SAR Vegetation Index (DPSVI), the modified DPSVI (DPSVIm), and the Dual-polarization Radar Vegetation Index for GRD products (DpRVIc). Table 2 provides details about each of the indexes mentioned.

Python programming language resources, using SNAP (**Sentinel Application Platform**) software algorithms were used to download and process images and calculate vegetation indices. The algorithms used to process Sentinel-1 images were from the **SNAP Sentinel-1 toolbox** module, made available by ESA. **R** programming language resources (R Core Team, 2023) were used to sample the images using the geographic coordinates of the soil profiles. SAR images were sampled using precise geographical coordinates obtained in soil surveying (using a Global Navigation Satellite System receiver), and the pixel that the soil localization falls into was selected to add SAR backscattering and vegetation indices to the database. The codes used to process and sample the Sentinel-1 IW GRD images can be checked in the repository: <https://github.com/eupassarinho/sentinel-1-SAR-vegetation-indices.git> (dos Santos et al., 2023a,b).

2.6. Soil organic carbon content modeling approaches

Four scenarios were defined for modeling: *model set 1, model set 2, model set 3,* and *model set 4* (Fig. 2). In *model set 1,* only the Vis-NIR normalized reflectance spectral bands (obtained by proximal sensing) were used to predict SOC contents. In *model set 2*, in addition to the Vis-NIR spectral bands, the SAR satellite vegetation indices (obtained by remote sensing) were used (polarization VH, CR, DPSVI, DPSVIm, and DpRVIc). In *model set 3*, the Vis-NIR spectral bands and the covariates LULC, soil type, climate, and elevation were used. Finally, the *model set 4* used all covariates: Vis-NIR, vegetation indices SAR, LULC, soil type, climate, and elevation. The four scenarios followed the modeling method described in the following topics:

2.6.1. Regression methods and covariate selection

Cubist was the regression method used in all four scenarios to predict SOC levels. Cubist is a regression method based on regression trees. It divides the training data into homogeneous partitions for the covariates used. A series of rules using "if-then conditions" define the partitions. When a partition is created, at the end of the trees (final leaves) a linear regression model (ordinary least squares) is used to predict the soil attribute. Continuous or categorical variables can be used to define conditions, but only numerical variables are used in the regression equations. Details about how Cubist works can be found in Quinlan (1992) and Kuhn and Johnson (2013). Cubist was chosen for its ability to handle spectral data and heterogeneous datasets such as R-SSL (Dematte et al. 2019b; Moura-Bueno et al. 2021; Viscarra Rossel et al. 2016).

The **R** language implementation of **Cubist** has two hyperparameters that can be used to optimize Cubist models: the number of *committees* and *neighbors*. When setting up a combination of committees, Cubist adopts a boosting-like scheme that creates iterative model trees in sequence, which means that the argument *committees* controls the number of model trees. The argument *neighbors* controls the number of similar samples (from the training data with defined rules) that are used to adjust to predict a new sample (Kuhn and Johnson, 2013; Kuhn and

Table 2
Description of vegetation indices and polarimetric descriptors calculated from Sentinel-1 IW GRD images.

Vegetation index	Formula	Theoretical bounds	Bibliographic reference
DPSVI	$DPSVI_{(i,j)} = \frac{VH_{(i,j)}\left[\begin{matrix}(VV_{max} \bullet VH_{(i,j)} - VV_{(i,j)} \bullet VH_{(i,j)} + VH_{(i,j)}{}^2)+ \\ (VV_{max} \bullet VV_{(i,j)} - VV_{(i,j)}{}^2 + VH_{(i,j)} \bullet VV_{(i,j)})\end{matrix}\right]}{\sqrt{2} \bullet VV_{(i,j)}}$	$0 \leq DPSVI$	(Periasamy, 2018)
DPSVIm	$DPSVIm_{(i,j)} = \frac{VV^2_{(i,j)} + VV_{(i,j)} \bullet VH_{(i,j)}}{\sqrt{2}}$	$0 \leq DPSVIm$	(dos Santos et al., 2021)
CR	$CR_{(i,j)} = \frac{VV_{(i,j)}}{VH_{(i,j)}}$	$1 \leq CR$	(Frison et al., 2018)
DpRVIc	$DpRVIc_{(i,j)} = \frac{q_{(i,j)} \bullet \left(q_{(i,j)} + 3\right)}{\left(1 + q_{(i,j)}\right)^2}$; in which $q_{(i,j)} = \frac{VH_{(i,j)}}{VV_{(i,j)}}$	$0 \leq DpRVIc \leq 1$	(Bhogapurapu et al., 2022)

Note: $VV_{(i,\,j)}$ and $VH_{(i,\,j)}$ correspond to the backscatter coefficient of polarizations VV e VH in pixel (i, j).

Quinlan, 2023). In all *model sets*, the committees and neighbors arguments were tuned via the *Search grid* declared in the *train* function of the **caret** library (Classification and Regression Training for **R**, (Kuhn, 2008). The number of declared committees was 50 and 100, and the number of neighbors was 5 and 9.

To optimize model training by Cubist, a Vis-NIR covariate selection step was added to the modeling. The Vis-NIR spectral covariates to predict SOC levels were selected using the LASSO (Least Absolute Shrinkage and Selection Operator) regression method, following the methodology of dos Santos et al. (2023a,b). LASSO is a regression and selection method based on the principle of parsimony, in which unimportant and/or highly correlated covariates are eliminated. LASSO fits a multiple linear regression model (using the ordinary least squares method) that has a covariate penalty parameter: when the slope coefficient of a covariate is equal to zero, then the covariate is eliminated from the model (James et al., 2013; Tibshirani, 1996). Covariate selection with LASSO was applied to Vis-NIR covariates in all modeling scenarios.

The LASSO implementation provided by the **glmnet** package in the R language was used (Friedman et al., 2010). LASSO's penalizing hyperparameter is λ, which was also tuned via the *Search grid* using a vector with numbers ranging from 0 to 2 in intervals of 0.05. In addition to the hyperparameter λ, the hyperparameter *alpha* was set up, which is the *elastic net mixing parameter*. To fit LASSO models, *alpha* was kept constant at one (1) (Friedman et al., 2010; Tay et al., 2023). Furthermore, as a requirement of the LASSO method, to correctly penalize the covariates, they must be on the same scale. Therefore, as pre-processing steps for training data in the **caret** *train* function, the following was set up: centering and scaling by the mean and standard deviation of the covariates, respectively.

The Recursive Feature Elimination (RFE) algorithm (Guyon et al., 2002; Kuhn and Johnson, 2013) was implemented right after LASSO's spectral covariables selection. RFE was implemented to apply covariable selection not only to normalized reflectance but also to SAR and environmental covariables. The *caretFuncs* from **caret** was used and a **Cubist** kernel was set to perform covariable selection. The *size* parameter, i.e., the number of predictors to be tested, was set as 5, 10, 20, and the total number of covariables, having other optimal intermediary numbers of predictors defined by the algorithm.

2.6.2. Dividing data to train and test (holdout) the models and cross-validation

The database was divided into a training set (for training Cubist models) and a holdout test using the proportion of 70 % and 30 % of the data, respectively (Brus et al., 2011). Although the division was done randomly, it was ensured that all samples from the same soil profile were in just one data set. This was done to ensure that the samples used to calibrate were independent of the ones used to test the models, following recommendations from (Brown et al., 2005; dos Santos et al., 2023a,b; Malmir et al., 2019; Poggio et al., 2017). Therefore, from the 127 soil profiles studied, 88 samples (n = 487, 70 % of the total) were used to train the models. The remaining 39 soil profiles (n = 214, 30 % of the total) were used to test the models.

The models were trained using 88 soil profiles (and their 487 samples). The cross-validation method used to optimize the models' hyperparameters was Leave-Soil-Profile-Out (LSPO CV). LSPO CV is an object-oriented k-Fold cross-validation that subsamples entire soil profiles for each of the calibration and validation partitions (dos Santos et al., 2023a,b). LSPO CV also aims to use independent soil samples to calibrate and validate the model (Brown et al., 2005), especially when the dependent variable is SOC, which may have a spatial dependence structure between nearby soil layers (dos Santos et al., 2023a,b). In the LSPO cross-validation, 10-Folds were configured. The *CreateSpacetimeFolds* method, from the **CAST** library (Meyer, 2021; Meyer et al., 2018), was used to randomly select which soil profiles were allocated to each of the *Folds*.

The division of the data into training and testing was done only once, but the division of the training data into different subsets of the LSPO cross-validation was done 100 times. Therefore, soil profiles were randomly drawn 100 times to select different samples to calibrate and validate the models. Hence, for each modeling scenario, 100 models were adjusted. The objective of repeating the modeling process is to evaluate the variability of model uncertainty when different data are used to train them (Gomes et al., 2019; Kuhn and Johnson, 2013; Mishra et al., 2022).

2.6.3. Model assessment

The models were adjusted to minimize the RMSE (Root Mean Squared Error) in training. However, the accuracy and correlation of the models' predictions on the test data were measured by the statistical metrics: RMSE, MAE (Mean Absolute Error), MSE (Mean Squared Error), coefficient of determination (R^2), Lin's concordance correlation coefficient (CCC) and Nash-Sutcliffe model efficiency (NSE).

The different *model sets* (e.g. *model set 1* versus *model set 3*) were compared using non-parametric statistical tests. The results of the correlation and error statistics, between the predicted and observed SOC values (in the test data), from each *model set* were statistically compared to verify whether the statistical values differed significantly, that is, whether they came from different distributions. For this, the Kruskal-Wallis test was applied, a non-parametric test for three or more groups of continuous variables. If a statistical difference was verified in this, the Dunn test was applied (test for paired groups, after the Kruskal-Wallis test) (McKight and Najab, 2010). In both cases, a 95 % confidence interval (P = 0.05) was adopted.

The relative importance of the covariates for model prediction was obtained by Variable Importance Plots – VIP (Greenwell et al., 2020; Greenwell and Boehmke, 2020; Kuhn and Johnson, 2013). This was done to identify and evaluate the relative importance of different sources of covariates (spectral covariates, SAR vegetation indices, and other environmental covariates) in the Cubist model. Regarding Cubist,

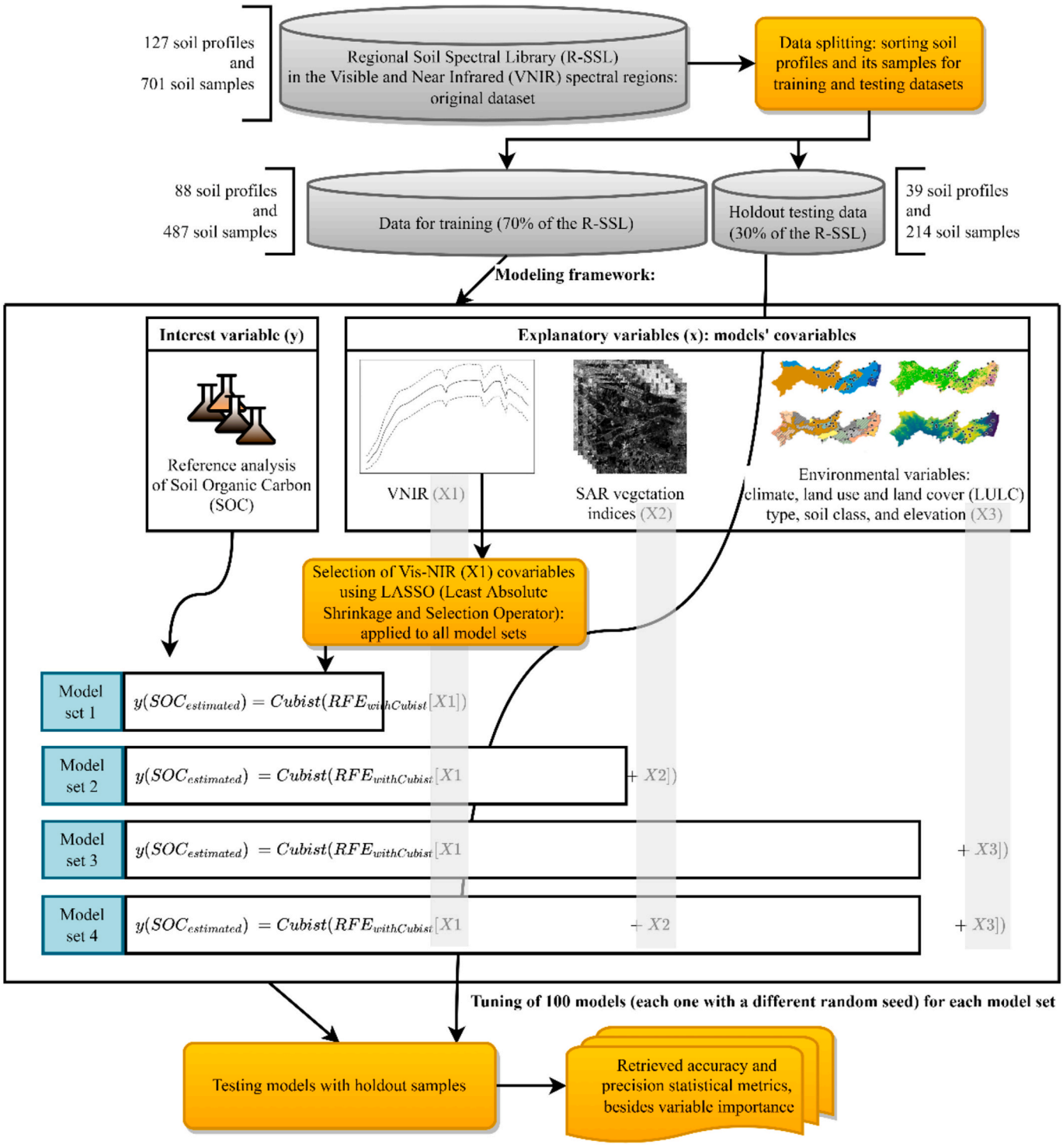


Fig. 2. Soil organic carbon (SOC) content modeling scheme using different predictor variables.

the relative importance of covariates consists of the average between the percentage of times that a given covariate was used: to create a rule (in decision trees) and in an internal linear regression. Fig. 2 displays a general scheme of the modeling process, mainly the modeling scenarios, for SOC contents.

3. Results

3.1. Environmental variables and SOC levels

The highest levels of SOC were found in forest areas and dense Caatinga whereas the lowest levels were in pasture and agriculture areas (Fig. 3a).

Regarding the SOC levels as a function of climate, in Fig. 3b, the lowest SOC levels were in the semi-arid climate (BSh). The highest SOC levels were found in the Csa climate, in a soil profile in the Dense

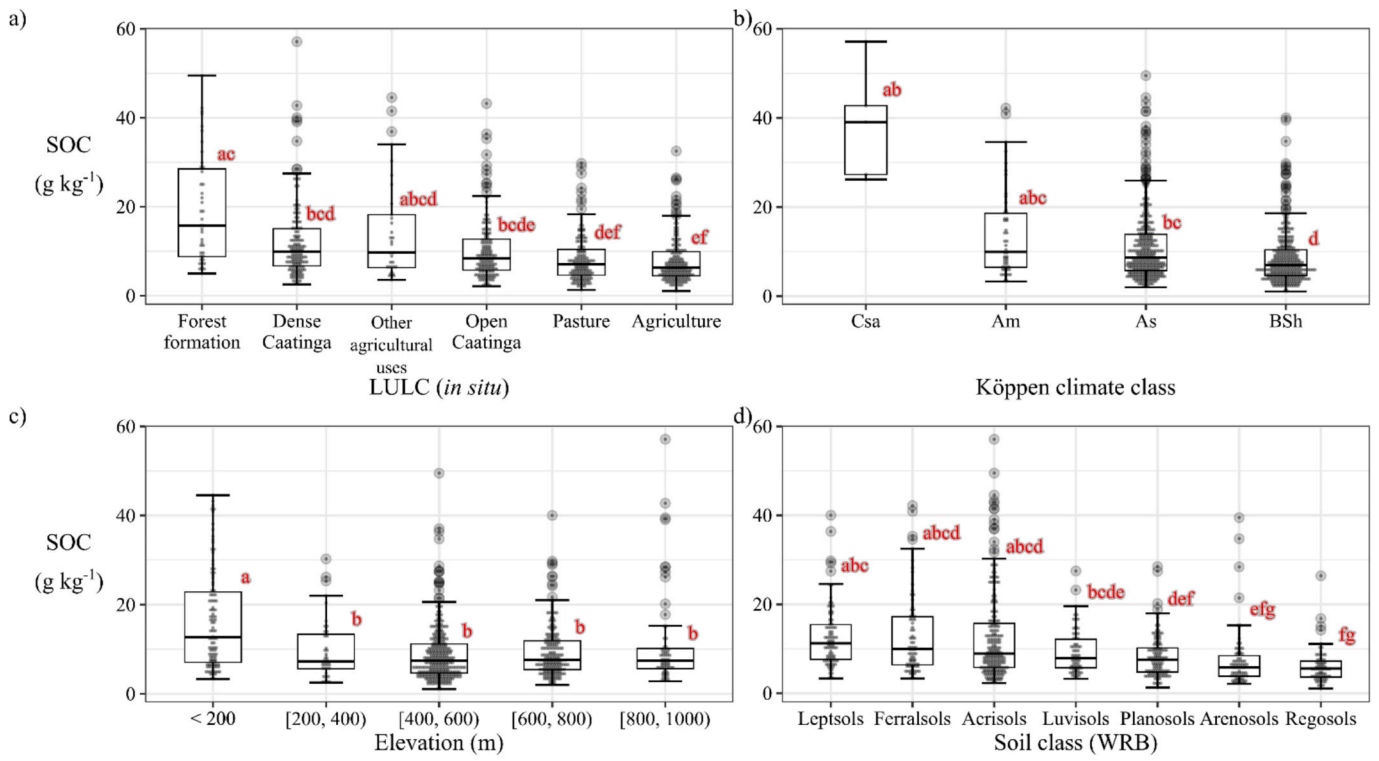


Fig. 3. Distribution of SOC (Soil Organic Carbon) levels in classes of landscape attributes. The following are displayed: a) SOC levels depending on land use and land cover (LULC); b) SOC levels in different climates; c) SOC levels at elevation ranges; d) the SOC levels in each soil class. Groups of distributions accompanied by different letters differ statistically from each other according to the non-parametric Kruskal-Wallis and Dunn tests (P = 0.05).

Caatinga area (with an elevation of around 970 m).

According to the local elevation (Fig. 3c), the highest SOC levels were observed at altitudes < 200 m, a region where the Am and As climates predominate. In the other elevation ranges, the SOC levels are similar, but with high SOC levels in areas >800 m.

Finally, regarding soil classes (Fig. 3d), a dummy variable, also adopted in the modeling strategies, the highest SOC levels were observed in Leptsols, Ferralsols, and Acrisols. From these classes onwards until Regosols, SOC levels decrease.

SAR vegetation indices indicate the amount of vegetation, which can be correlated with the amount of aboveground biomass, the leaf area index, or another biological parameter. Therefore, Fig. 4 shows the distributions of SOC levels and SAR vegetation index values in each LULC class.

The graphs from Fig. 4 show that the DPSVI and DPSVIm indices follow the same pattern of SOC: the highest levels are concentrated in forest areas, dense Caatinga, and other agricultural uses, decreasing values up to the pasture and agriculture classes.

3.2. *Performance of the models in predicting SOC levels with different covariates*

Fig. 5 displays the results of the accuracy and correlation statistics of the predictions of the four *model sets*. Model predictions were obtained using the test dataset, with samples independent from the model training. Furthermore, the results are expressed in boxplots since for each of the four *model sets*, 100 values of each metric were obtained, totaling 400 trained models.

In the RMSE, MAE, and MSE error metrics of each *model set*, it is observed that the more explanatory variables are added to the modeling, the more accurate predictions become (Fig. 5a). The average RMSE in *model set 1* was 4.54 g kg^{-1}. Adding the SAR vegetation indices, the average RMSE was 7.60 % lower, equal to 4.19 g kg^{-1}. Using spectral covariates plus environmental covariates, the average RMSE was 4.12 g kg^{-1}, 9.33 % lower than in *model set 1*. Finally, in *model set 4*, which used all available covariates to model SOC levels, the average RMSE was 3.94 g kg^{-1}, 13.12 % lower than *model set 1*. The same behavior was noted for the MAE and MSE metrics.

Fig. 5b) displays the results of the R^2, CCC, and NSE metrics for each *model set*. The result obtained with this set of correlation metrics between predictions and observations corroborates the pattern observed in Fig. 5a): adding environmental covariates improved the model performance, making them more efficient and accurate. The average R^2 obtained in *model set 1* was 0.62, but when radar and environmental covariates were added (*model set 4*) this value increased by 15 % to 0.72. The same pattern of improving predictions could be noticed in the other metrics of agreement (CCC) and model efficiency (NSE).

It is important to highlight that there was a significant statistical difference between the RMSE, MAE, MSE, R^2, CCC, and NSE values obtained in the four *model sets*. The results of the Kruskal-Wallis non-parametric test can be seen in Table 3. Furthermore, the results of the paired test (for the metric groups of each *model set*) and Dunn's non-parametric test (after the Kruskal-Wallis test) can be checked on the Supplementary Material (Table S2).

Numerically, we identified that by adding predictor covariates to the Cubist models the models' error dropped. Although SOC predictions using only Vis-NIR bands in Cubist models are already very good, by comparing the predicted and observed SOC values, as shown in Fig. 6, the predictions became even better and more accurate.

Fig. 6 shows the scatterplots between observed and predicted SOC values. The plotted points represent the average value of the predictions of the 100 models, while the vertical bars represent the range of the 100 predicted values for each sample, from the 100 training sessions of each *model set*. In Fig. 6a), which predicted SOC samples using only the Vis-NIR bands, the samples are more dispersed (further away from the 1:1 ratio) than in Fig. 6d). In Fig. 6d) SOC samples were predicted using the Vis-NIR bands plus all other predictor variables.

Furthermore, comparing Fig. 6a) with Fig. 6d), we noticed that the

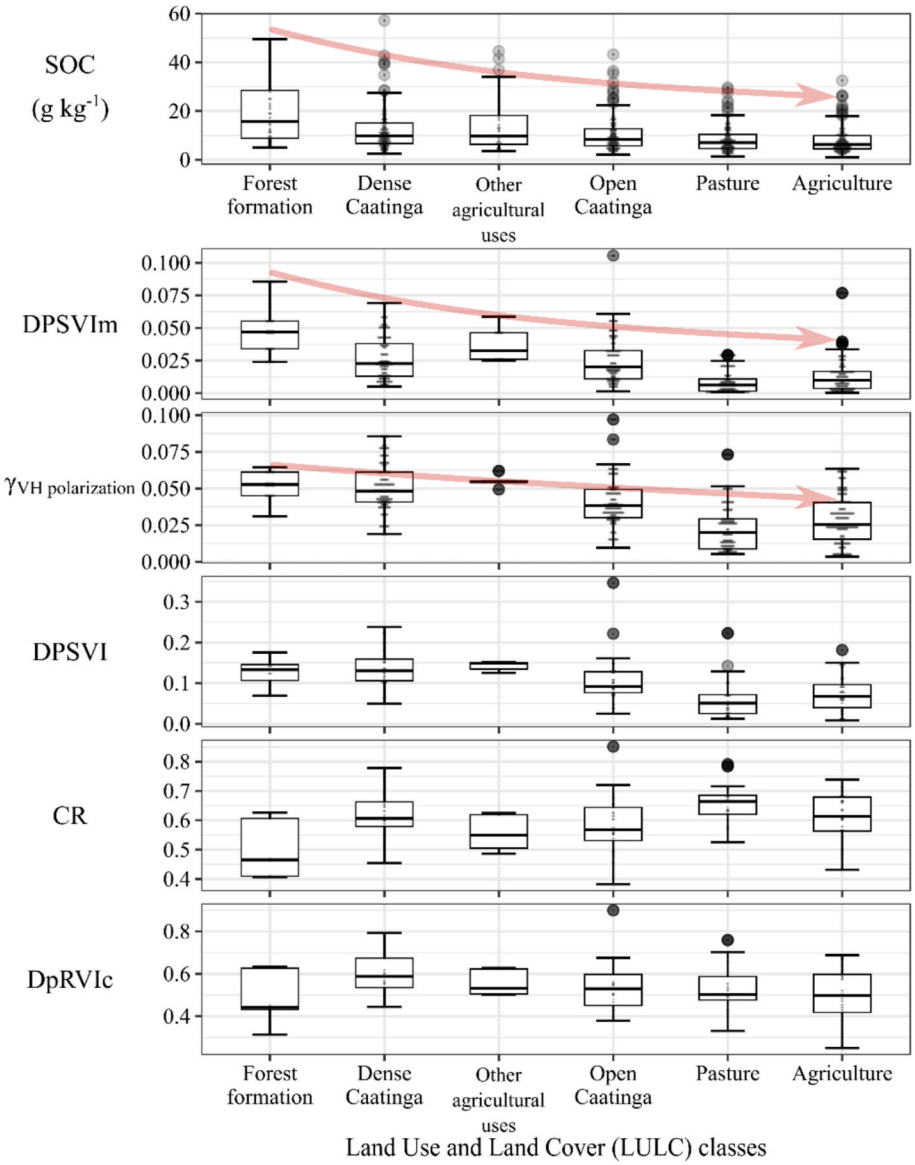


Fig. 4. Distribution of SOC (Soil Organic Carbon) levels and SAR vegetation indices as a function of land use and land cover (LULC). DPSVI, Dual-polarization SAR vegetation index; DPSVIm, modified DPSVI; CR, cross-ratio; DpRVIc, Dual-polarization Radar Vegetation index for Sentinel-1 GRD products.

amplitude of the 100 predictions for each SOC sample decreased. In other words, predictions became more accurate when more variables were used.

3.3. Contribution of different covariates in estimating SOC levels

A maximum of 76 Vis-NIR bands were selected by LASSO in *model set 1*. However, after the RFE-Cubist selection, not all were used in all repetitions. This can be observed in Fig. 7. While there is a predominance of using all or almost all bands (76 or 72 Vis-NIR bands, which occurred 42 and 50 times), in some repetitions smaller RMSE was obtained with least number of bands. The same is observed for all other *model sets*.

We observed, in the previous results, improvements in the accuracy and precision of SOC content estimates when covariates were added to the spectroscopy models. However, it is necessary to better understand the contribution of these environmental variables (vegetation types and indices, climate classes, soil in addition to elevation) used for the estimates. Thus, Fig. 8 displays the relative importance of these covariates.

Fig. 8a) shows the relative importance of individual spectral bands selected by the LASSO method for Cubist regression. In this graph, the plotted points represent the highest importance measured, while the vertical bars represent the amplitude of the measured importance for each band. In the graphs in Fig. 8b), the relative importance of the covariates was grouped into boxplots to aid the comparison between the variable groups. The sub-graphs in Fig. 8b) were divided according to the model's settings.

Different spectral bands were selected by LASSO for SOC modeling via Cubist. Fig. 8a) shows the importance and indicates specific wavelengths at which the measured reflectance is intrinsically related to soil organic particles. The spectral bands had the highest importance for SOC prediction, being used by Cubist in its internal decision trees, as well as in internal linear models.

After the spectral bands, the highest importance values were given to the covariates derived from the Sentinel-1 SAR: the importance of these variables ranged from 19 to 65 %. It is important to highlight that the importance of covariates for Cubist reflects the average number of times a variable was used for data division and prediction (in decision trees and linear models, respectively). We observed the least important covariate among the categorical variables were soil class, land use and

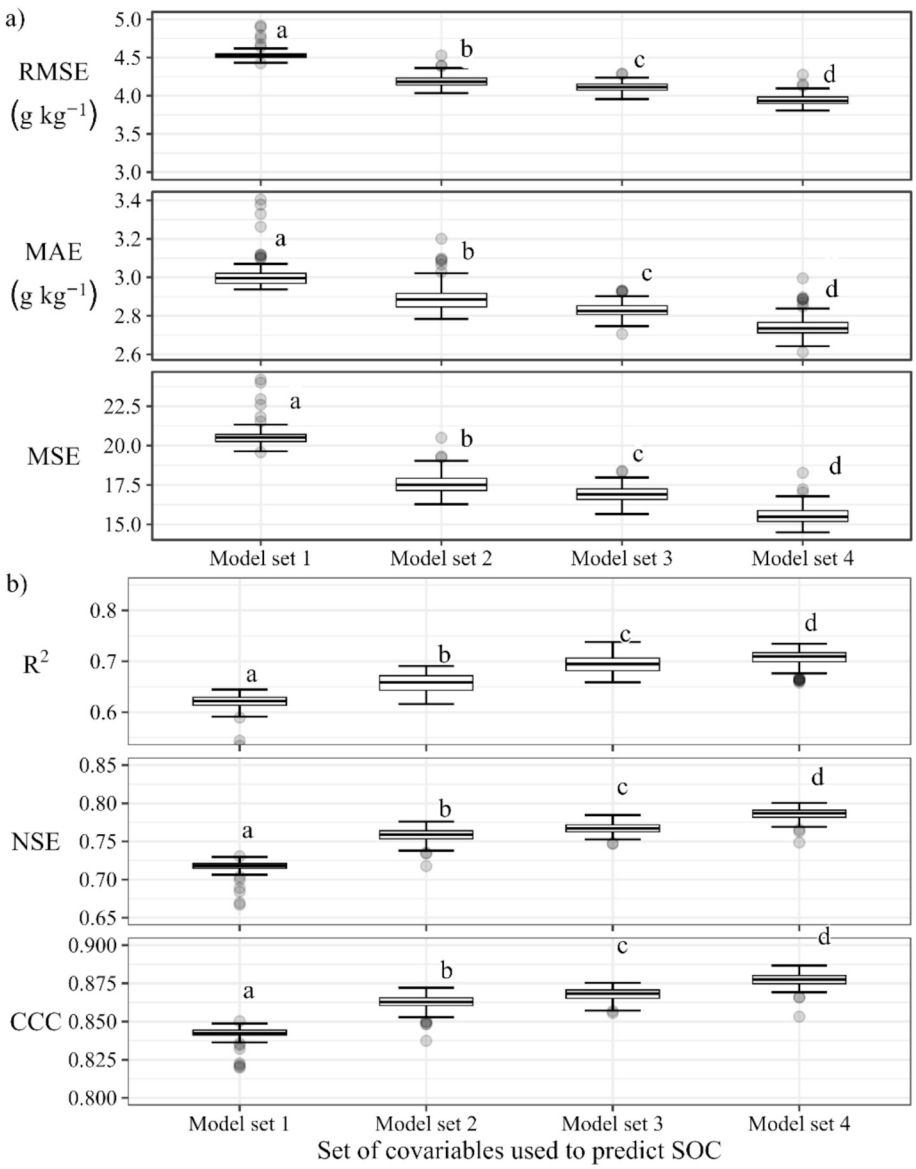


Fig. 5. Distribution of the 100 error and correlation values for each metric (Root Mean Squared Error [RMSE], Mean Absolute Error [MAE], Mean Squared Error [MSE], coefficient of determination [R^2], Lin's concordance correlation coefficient [CCC], and Nash- Sutcliffe model efficiency [NSE]) from the SOC (Soil Organic Carbon) predictions of the four *model sets*. Different lowercase letters indicate significant statistical differences between groups of the same statistical metric (P = 0.05).

Table 3
Result of the Kruskal-Wallis test for each statistical metric (Root Mean Squared Error [RMSE], Mean Absolute Error [MAE], Mean Squared Error [MSE], coefficient of determination [R^2], Lin's concordance correlation coefficient [CCC], and Nash- Sutcliffe model efficiency [NSE]): χ^2 is the chi-square statistic of the test, df is the degrees of freedom and (*) indicates a significant statistical difference (P = 0.05) between the *model sets*.

Tested statistical metric (groups of *model sets*):	χ^2	df	Adjusted *p*-value
RMSE	338.49	3	$4.64x10^{-73}$*
MAE	305.93	3	$5.17x10^{-66}$*
MSE	338.51	3	$4.59x10^{-73}$*
R^2	317.80	3	$1.40x10^{-68}$*
CCC	338.84	3	$3.89x10^{-73}$*
NSE	338.49	3	$4.64x10^{-73}$*

Note: the asterisk (*) next to the adjusted p-value indicates that the null hypothesis (which establishes that the compared groups are statistically equal) was rejected and the adjusted p-value is statistically significant at P = 0.05.

cover, and climate; as well as terrain elevation (continuous variable). To assess individual covariable relative importance for *model set 4*, it is possible to consult Fig. S2 in the Supplementary Material.

4. Discussion

4.1. Relation of environmental variables with SOC levels

There are variations in the SOC levels of the R-SSL samples that can be associated with different environmental (LULC, climate, and elevation) and pedological (soil classes) covariates, as shown in Fig. 3. Regarding the LULC's vegetation classes, SOC levels were higher in places where the soil receives a greater contribution of plant residues and is not managed (such as the disturbance of surface layers). This was observed in Fig. 3a) in the forest formation and Dense Caatinga areas. SOC levels were lower in places where the input of organic material of plant origin is lower, pasture, and agriculture, which receive agricultural and/or pastoral management.

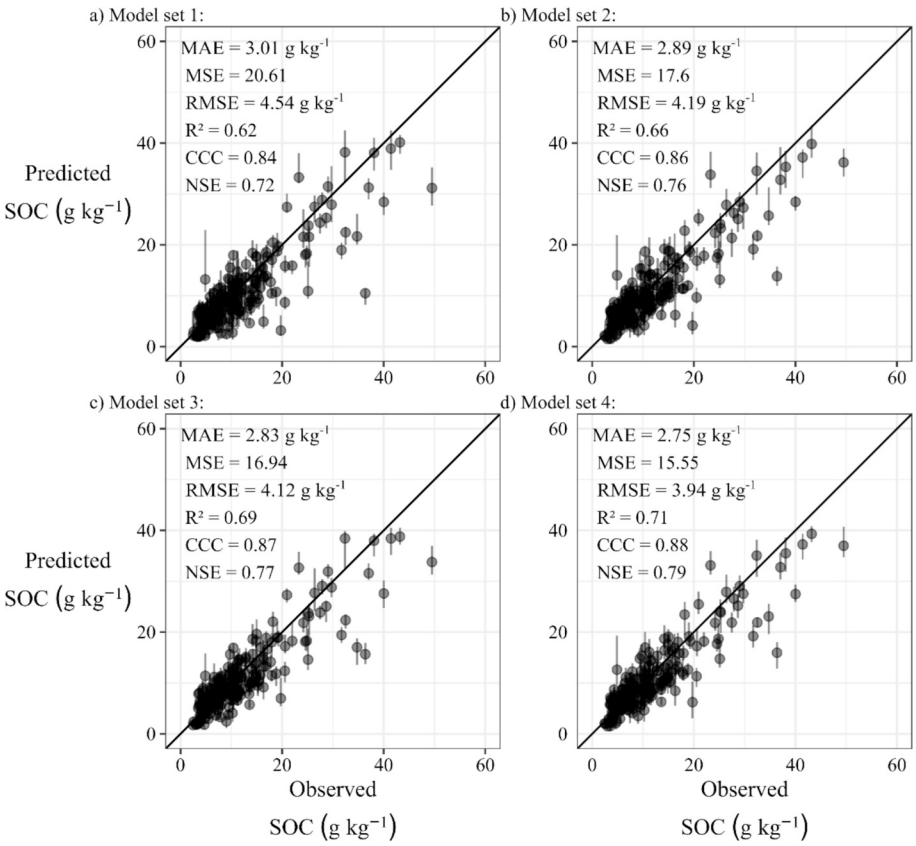


Fig. 6. Scatterplots between observed and predicted soil organic carbon (SOC) values in the four model sets: a) using only spectral variables (Vis-NIR); b) Vis-NIR bands plus SAR vegetation indices; c) Vis-NIR plus environmental covariates; d) all predictors. The plotted points represent the average of the predictions of the 100 models, while the vertical bars represent the range of the 100 predicted values for a given sample.

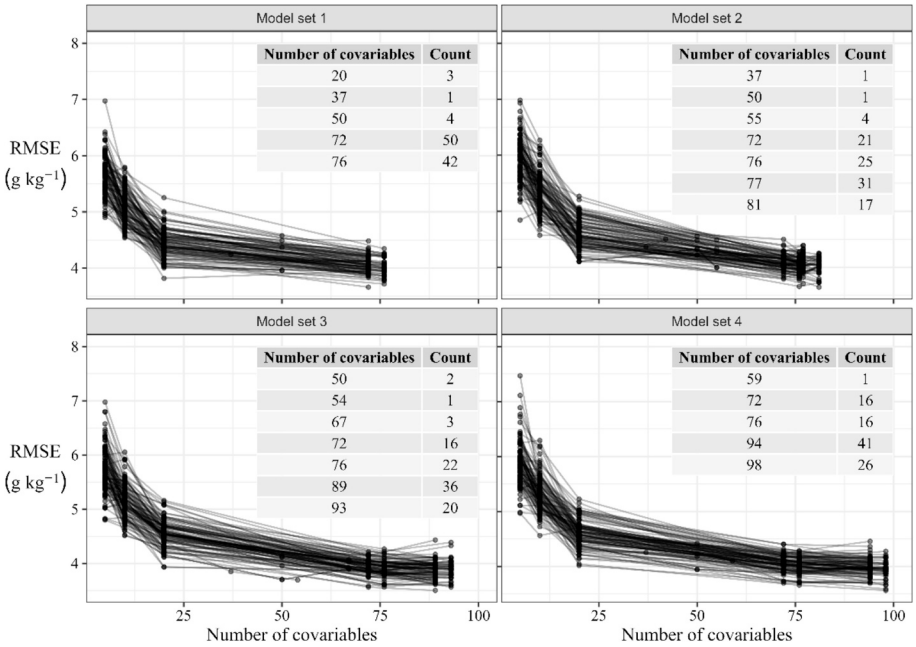


Fig. 7. Recursive Feature Elimination results: points represent RMSE values for each iteration and repetition. The number of covariables refers to the number of predictors tested and selected by RFE, and Count displays how often that set of covariables occurred along the repetitions.

The input of organic plant residues is the main source of soil organic matter, and the vegetation stratum is an indicator of SOC content in the soil (Jobbágy and Jackson, 2000; Wiesmeier et al., 2019). Although it is known that in areas occupied by pastures, SOC levels may exceed those from original forest conditions (Guo and Gifford, 2002), including in the Caatinga biome (Ferreira et al., 2016), from where the majority of R-SSL

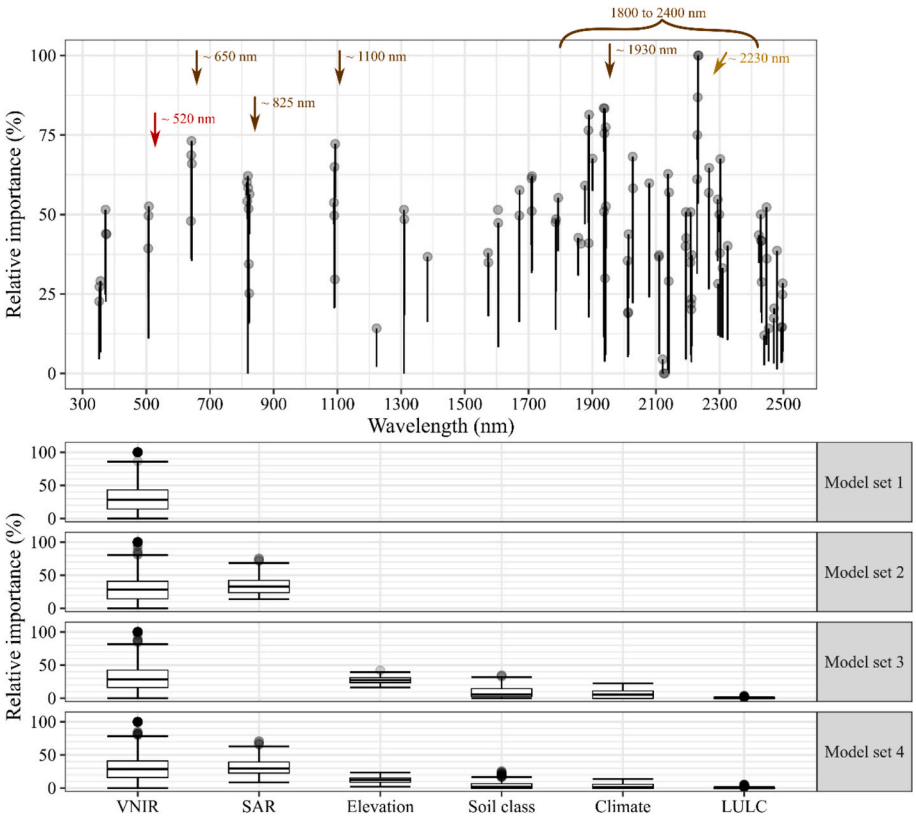


Fig. 8. Relative importance graphs of different predictor covariables for modeling soil organic carbon (SOC) content using the Cubist regression method. The importance of the Vis-NIR spectral bands for *model set 1* is detailed in subplot a), while graph b) displays the grouped relative importance of the other covariates for all *model sets*: boxplots summarize individual covariables importance for the grouped variables: Vis-NIR, SAR, Soil class, Elevation, Climate, LULC.

samples in this study come from, this also depends on the management of the pastures (de Medeiros et al., 2020), including the degree of degradation, if any. Hence, the differences in SOC levels between the classes with the highest amount of above-ground biomass (Forest Formation and Dense Caatinga) and those with the lowest level of plant biomass (Pasture and Agriculture) (Fig. 3a).

Climate is also a source of variation in SOC levels in R-SSL samples. Temperature and rainfall, both meteorological variables used in existing climate typologies, are the main factors in soil SOC storage at regional scales. Temperature and rainfall control the primary vegetation production rates (input of organic matter into the soil). Furthermore, temperature influences soil microbial activity, microorganisms responsible for soil respiration, and decomposition of organic matter (Guo and Gifford, 2002; Wiesmeier et al., 2019).

The lowest SOC levels can be found in Fig. 3b) in areas with a semi-arid climate, with an average SOC of 8.63 g kg^{-1}. In these areas, temperatures are higher, and annual rainfall is lower, which increases the rate of decomposition of organic matter and reduces the input of biomass, compared to areas under other climate typologies. The highest levels (average SOC of 38.48 g kg^{-1}) were observed in a profile under a humid subtropical climate with a dry and hot summer (Csa). These climate types occur in the semi-arid region of Brazil mainly due to the relief in plateau regions. In this case, it is an area with an elevation between 800 and 1,000 m in the Borborema Plateau, in the State of Pernambuco (Alvares et al., 2013; de Souza et al., 2022).

There was no significant difference between the SOC levels found in elevations higher than 200 m above sea level (Fig. 3c). There were, however, higher SOC levels in elevations lower than 200 m, as well as outliers in elevations greater than 800 m. We can say that in both cases, the SOC levels can be related to the climate. In the first case, in elevations lower than 200 m, there is a predominance of tropical climates with monsoons and dry summers, where precipitation is also higher (1,000 to 2,500 mm $year^{-1}$). The average SOC level with elevation $<$ 200 m was 16.45 g kg^{-1}. In the second case, in elevations greater than 800 m, although the average SOC level was 12.25 g kg^{-1}, very high SOC values ($>$ 30 g kg^{-1}) were found under Csa climate.

A possible explanation for this SOC behavior regarding elevation is that elevation affects temperature, and consequently, the climate. In other words, on a regional scale, the effect of elevation on SOC variations is related to the influence of elevation on environmental parameters that affect other soil formation factors, notably average annual temperature and rainfall (Hobley et al., 2015, 2016; Wiesmeier et al., 2019). Areas with a humid subtropical climate in the Brazilian semi-arid region, for example, have elevation as the main factor for their occurrence, as the aforementioned Borborema Plateau (Alvares et al., 2013; de Souza et al., 2022).

Soil type is also strongly associated with organic carbon storage. In many cases and classification systems, SOC content is part of the classification criteria (Mayes et al., 2014). However, the soil type only reflects soil properties that influence the supply and storage of organic carbon.

During pedogenesis, weathering reactions lead to changes in soil mineralogy, which strongly influence the surface area of mineral reactivity and carbon storage (Wiesmeier et al., 2019). Therefore, SOC levels may be higher in soils with a greater propensity for physical protection of particles, whether in soil aggregates or by clay and silt particles (Six et al., 2002; Stewart et al., 2008; Yu et al., 2019). Furthermore, metal sesquioxides (such as Fe and Al oxides) also contribute to the stabilization of SOC through the affinity between Fe and Al oxides and organic compounds (Baldock and Skjemstad, 2000; Wiesmeier et al., 2019; Yu et al., 2019). This is an indicator of what happens with the Ferralsols and Acrisols classes, which appear in Fig. 3d) with the highest SOC levels. On the other hand, the Arenosols (sandy soils) and Regosols appear among the classes with the lowest SOC levels.

4.2. Accuracy and precision improvement in modeling strategies

Specifically dealing with *model set 1*, whose SOC predictors were only the spectral bands in Vis-NIR, the accuracy and precision values obtained are compatible with those found in the literature. The average RMSE was about 4.5 g kg^{-1} while the accuracies reported in the literature range from 0.3 to 25 g kg^{-1} (Dematt$\hat{e}$ et al. 2019b; Soriano-Disla et al. 2014). This error amplitude found in the literature, however, depends on the size, location, representation scale of the spectral library used, and amplitude of the SOC values. Furthermore, a similar RMSE value was found by dos Santos et al. (2023a,b), around 4.2 g kg^{-1}, applying the same validation/test methodology on the same R-SSL.

Noteworthy, the spectral bands selected by LASSO for input into the Cubist models are intrinsically related to either the SOC levels or the presence of clay minerals and iron oxides. In the Vis-NIR spectra (400 to 1,000 nm) the spectral bands can be related to iron oxides, especially in the peaks around 520 and 650 nm (Viscarra Rossel and Behrens, 2010), as well as to the presence of organic compounds (650 and 825 nm peaks) (Ben-Dor, 1997; Nocita et al., 2015; Stevens et al., 2013; Viscarra Rossel and Behrens, 2010), as highlighted in Fig. 8a. The peaks highlighted in the NIR region (around 1100 nm) are associated with water and organic compounds (Viscarra Rossel and Behrens, 2010). In the infrared shortwave, specifically from 1800 to 2400 nm, the selected spectral bands can be associated with different organic compounds, such as amines, alkyls, carboxylic acids (especially at 1930 nm), amides, aliphatic compounds, methyls, phenolic compounds, polysaccharides and carbohydrates (Coates, 2006; Knadel et al., 2015; Vasques et al., 2010; Viscarra Rossel and Behrens, 2010).

Among the selected spectral bands, there are also clay mineral signatures. This association can be made with the relative importance of the bands around 2230 nm (also highlighted in Fig. 8a), which indicates the presence of aluminum hydroxides, Al–OH (Viscarra Rossel and Behrens, 2010; Zheng et al., 2016).

The same selected spectral bands (Fig. 8a) were used in all model configurations, from *model sets 1* to *4*. Therefore, the results of the statistical performance metrics of the *model sets* (Fig. 5 and Fig. 6) show that both accuracy and precision improved in the Vis-NIR spectroscopy predictions as environmental variables were added to the model. This corroborates the results found by Moura-Bueno et al. (2021), in subtropical soils from southern Brazil, and Wang et al. (2022), in soils in northern China.

When adding SAR vegetation indices to the Vis-NIR spectral bands, the average RMSE dropped about 7.2 %, from 4.54 to 4.19 g kg^{-1} (see Fig. 6a and Fig. 6b). Also, the average R^2 increased 6.18 %, from 0.62 to 0.66. Furthermore, we observed in Fig. 8b that the SAR vegetation indices have relative importance ranging from 20 to 65 %, both in *model set 2* and in *model set 4*.

SAR vegetation indices are good indicators of the amount of aboveground biomass (AGB). Although the backscatter from the vegetated surface for SAR sensors is not a direct measure of AGB (Woodhouse et al., 2012), the backscatter observed over these areas in cross-polarizations (VH or HV) is directly associated with the AGB content of that area (da Bispo et al., 2020; dos Santos et al., 2021; Joshi et al., 2017; Saatchi, 2019; Santoro et al., 2021). The plant organs that most interact with microwave radiation usually depend on the wavelength used. While radiation in the C band (with a wavelength of approximately 6 cm) tends to interact more with leaves and branches, in areas with denser vegetation, radiation in the L band ($\lambda \cong$ 23 cm) can interact with tree trunks (Flores-Anderson et al., 2019). In any case, in crossed polarizations, the surface elements that change the polarization state of the electromagnetic wave that reflects to the sensor appear brighter (Mitchard et al., 2011) in a wave reflection mechanism known as volumetric backscattering (Woodhouse, 2006).

The vegetation indices employed use the VV and VH polarizations of the Sentinel-1 images, to measure and represent the biophysical parameters of the detected vegetation (dos Santos et al., 2023a,b). DpRVIc is an index based on the polarization degree that vegetation causes in microwave radiation (Bhogapurapu et al., 2022; Mandal et al., 2020a). The polarization degree measures how much of the total energy backscattered by the targets has had its polarization modified. In the case of DpRVI and DpRVIc, both have been applied in the discretization of phenological stages of annual crops such as corn, canola, wheat, etc. (Bhogapurapu et al., 2022; Mandal et al., 2020a).

The DPSVI and DPSVIm indices are also based on the depolarization degree of the signal, however, they seek to distinguish surfaces of water bodies and bare soil from vegetated areas. The values of the indices are close to zero for water bodies and bare soil and increase with the amount of AGB (dos Santos et al., 2021; Periasamy, 2018). The difference between DPSVI and DPSVIm is that the latter is more sensitive to different levels of biomass in forest areas (dos Santos et al., 2021). For this reason, DPSVIm has incorporated the CR index (Frison et al., 2018) to ease the separation of different AGB levels in these areas.

The SAR indices with the greatest relative importance in SOC modeling were the VH polarization and the DPSVIm index, whose relative importance ranged from 45 to 65 %. Considering the characteristics of the SAR vegetation indices, we can conclude that they contribute to the prediction of SOC levels because they are capable of representing the plant residue intake in the soil, which is more correlated with the SOC contents found in the surface layers of the soil (Guo and Gifford, 2002; Hobley et al., 2015; Wiesmeier et al., 2019).

By adding soil formation factors (elevation, soil type, vegetation type, and climate) to R-SSL to predict SOC levels the model's performance also improved. So, from *model set 1* to *model set 3*, the average RMSE fell about 9.33 %, from 4.54 to 4.12 g kg^{-1}. The average R^2 obtained in *model set* 3 was 0.69, 12 % higher than in *model set 1*. This can be seen in Fig. 6a and Fig. 6c.

The relative importance of the covariates elevation, soil type, vegetation type, and climate for modeling with R-SSL ranged from 0 to 40 % (Fig. 8b). Except for elevation, a numerical variable that was used for both stages: decision and regression; all other covariates are categorical, and were used only in the decision trees of the Cubist model to separate the samples based on similar SOC levels.

Using this set of categorical variables, Cubist builds spectral sublibraries with samples grouped by similarity to reduce the SOC variance as a function of the classes of environmental variables. Moura-Bueno et al. (2020, 2019) studied the effect of SSL stratification on SOC prediction performance. The authors tested the hypothesis that reducing SOC variance, after stratifying SSL based on environmental, pedological, and spectral class criteria/variables, could improve the accuracy of SOC estimates. Moura-Bueno et al. (2020) concluded in their study with SSL of subtropical soils in southern Brazil that stratification may increase accuracy as long as there is a significant difference in SOC between classes of environmental variables. Hence, the stratification with the categorical LULC and Physiographic Region produced better predictions of SOC. However, the disadvantage found in manual stratification was the reduction in the number of samples available for model calibration (Moura-Bueno et al., 2020).

For *model set 3*, after the Vis-NIR normalized reflectance bands, the relative importance given to elevation ranged from 22 to 37 %. Although there is no statistically significant difference between most elevation classes (Fig. 3c), in the Northeast region of Brazil, elevation is closely related to the occurrence of different climates (Alvares et al., 2013). Where the elevation was less than 200 m, there were wetter and rainier areas and higher SOC levels (see Fig. 3c).

In turn, the climate, which is defined by the normal pattern of rainfall and temperature, will influence both the rate of primary production of vegetation (organic carbon input), decomposition, and emission through microbial respiration (carbon output). So much so that in *model set 3* the BSh climate class, with higher temperatures and lower rainfall, had the greatest importance (relative importance of around 17 %) compared to the other classes in the stratification of the SOC samples for Cubist.

Hobley et al. (2015, 2016) studied SOC variations in Western Australian soils as a function of different indicators and drivers (SOC input/output modulators). The authors identified that climate is an important driver of SOC, with higher rainfall associated with higher proportions of organic carbon in humus (Hobley et al., 2016). For this reason, when predicting SOC, numerical variables, mainly average annual rainfall, had greater importance for the results. Corroborating the relationship that rainfall has on primary production (Michaletz et al., 2014) and SOC storage mainly in surface layers (Jobbágy and Jackson, 2000; Wiesmeier et al., 2019).

In *model set 3*, the pedological variables, essentially the dummy variables for the Arenosols and Regosols classes, had greater relative importance (about 44 %), after the Vis-NIR covariates. The stratification of samples of these two soil types of soil, which are mainly high sandy soils, among the others, was important to improve the modeling accuracy.

This result corroborates the conclusions of the study by Jaconi et al. (2017). These authors studied SSL stratification strategies in the NIR spectrum, at a country scale (in Germany), created from soil samples under agricultural and pasture use. The best strategy for the accuracy of SOC estimates was to separate samples of sandy soils from samples of other textural classes, calibrating two models separately (Jaconi et al., 2017). The basis for this is the already discussed relationship of greater SOC storage capacity in soils with higher clay and silt contents, due to organic matter protection mechanisms.

Although the relationship between the LULC classes and SOC variations is known, as can be seen in Fig. 3b) discussed in the previous topic, the lowest relative importance in *model set 3* was given to these variables by Cubist. This does not necessarily mean that these variables are not important for predicting SOC. Variations in uses and coverage are considered important indicators and drivers of SOC at different scales, mainly in the surface layers of the soil (Guo and Gifford, 2002; Hobley et al., 2015, 2016; Moura-Bueno et al., 2021; Wiesmeier et al., 2019). However, for the Cubist models of *model set 3* with R-SSL, stratification of samples using climate, elevation, and pedology were more important.

The combination of all predictors, *model set 4*: Vis-NIR spectral bands, SAR vegetation indices, elevation, soil type, climate, and vegetation cover; produced the best SOC estimates. From *model set 1* to *model set 4*, the average RMSE dropped about 13 %, from 4.54 to 3.94 g kg^{-1}. Following, the average R^2 increased by 14 %, going from 0.62 (*model set 1*) to 0.71 (in *model set 4*). The other statistical metrics followed the same trends (see Fig. 6).

Moura-Bueno et al. (2021) obtained, by comparing a model with only Vis-NIR spectral signatures to another with spectral bands plus a set of auxiliary covariates, a reduction in RMSE of about 22 %, and an R^2 increasing from 0.76 to 0.85. In a similar approach, but to predict soil organic matter content, Wang et al. (2022) obtained an RMSE of 3.85 g kg^{-1} and R^2 of 0.85 with the addition of auxiliary environmental covariates and spectral classification to the Vis-NIR spectral signatures. When using only spectral signatures the RMSE and R^2 were 4.30 g kg^{-1} and 0.76, respectively.

Gains in accuracy with covariates added to SSL can be considered relevant or irrelevant for SOC prediction, depending on the desired application of these estimates (Moura-Bueno et al., 2021). Overall, accuracy gains while other assumptions are kept, such as the understanding and generalization capacity of the models (dos Santos et al., 2023a,b; Jaconi et al., 2017; McBride, 2022; Moura-Bueno et al., 2021; Viscarra Rossel et al., 2022), help improving Vis-NIR spectroscopy of soils as an alternative method for testing soil samples. Furthermore, the advantage of adding auxiliary covariates to SSL in regression methods such as Cubist is that gains in accuracy can be obtained without sample losses due to manual stratification.

For more precise applications, in which it is necessary to make as few mistakes as possible, more reliable estimates are required. This is the case with precision agriculture applications, whose goal sometimes is to map small variations in soil properties on a more detailed scale (Camargo et al., 2022). Other applications of SOC estimates by spectroscopy include fertilization recommendations (Barra et al., 2021; Rosin et al., 2020), and soil monitoring to implement good soil management practices (Angelopoulou et al., 2020). In these situations, reducing the error when using auxiliary variables is justified.

It is necessary to take into account that there is a cost with the adoption of auxiliary remote sensing products for SSLs, which is the need for more data processing. In this case, the processing of satellite images to generate the vegetation indices commonly used in digital soil mapping. However, the advantage is that high-resolution products are needed for estimates at smaller scales, from watershed (Kunkel et al., 2022) to field scale for prediction in agricultural areas (Nguyen et al., 2022).

An important issue related to SAR vegetation indices is that indices derived from optical orbital sensors are already widely used for digital mapping of soil organic carbon/organic matter. However, the use of radar remote sensing has three major advantages, inherent to microwave systems and the Sentinel-1 mission, which can provide operability for monitoring soil, agriculture, forestry, and environmental activities with SSLs. The first advantage is the low (or almost zero) cloud interference in imaging, which is a problem in optical remote sensing, mainly for tropical regions (Asner, 2001; Carrasco et al., 2019; dos Santos et al., 2022; Flores-Anderson et al., 2019; Roy and Yan, 2020). The second advantage is the high temporal and spatial resolution since the high operability of the Sentinel-1 mission with the constellation of two satellites (Sentinel-1A, −1B [inoperative since 2022] and −1C [planned for launch]) allows an almost weekly revisit rate (for some regions) (Mandal et al., 2020a; Periasamy, 2018). Furthermore, the GRD products, for which the vegetation indices used were designed, are distributed in global coverage and at high spatial resolution (ESA, 2012, ESA, 2022). Therefore, satellite products can be applied at different soil monitoring scales, and when obtained in time series, they can represent vegetation dynamics and/or land cover management (Kunkel et al., 2022).

In addition to high-resolution remote sensing, future work at local scales may benefit from other relief parameters that aid SSLs. This is the case with the topographic wetness index (TWI). The TWI has a good capacity to help explain SOC variations in the landscape on small scales, as it is an indicator of water movement and availability in the soil (Hobley et al., 2015; Sørensen et al., 2006; Wiesmeier et al., 2019).

5. Conclusions

Adding SAR vegetation indices (obtained by orbital remote sensing) to the R-SSL in Vis-NIR (obtained by proximal sensing) improved SOC spectral estimates. The combination of categorical and continuous variables describing the environment and soil formation factors (climate, soil type, land use and land cover class, and elevation) to the R-SSL in Vis-NIR resulted in a significant improvement in the estimates of SOC.

The spectroscopy-based models with the greatest accuracy and precision for predicting SOC levels were the models adjusted with all auxiliary covariates, with all covariates used being related to the SOC variation.

The average RMSE and R^2 obtained in predicting SOC with only Vis-NIR spectral data was 4.54 g kg^{-1} and 0.62, respectively. But when all environmental and radar remote sensing covariates were added, the RMSE reduced by approximately 13 % (to 3.94 g kg^{-1}) and the R^2 increased by 14 % (to 0.71).

The proposed methodology expands and validates the use of Sentinel-1 SAR mission data for land-related applications, such as soil carbon monitoring, by integrating optical and radar capabilities with landscape descriptors. Overall, by applying a more interpretable machine learning framework, we demonstrate that Vis-NIR reflectance spectroscopy can benefit from additional data sources. However, these data introduce constraints on the choice of machine learning methods, which must either inherently handle discrete covariates or include

appropriate preprocessing steps.

CRediT authorship contribution statement

Erli Pinto dos Santos: Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Michel Castro Moreira:** Writing – review & editing, Project administration, Investigation, Funding acquisition. **Elpídio Inácio Fernandes-Filho:** Writing – review & editing, Methodology, Investigation. **José A. M. Demattê:** Writing – review & editing, Methodology, Investigation. **Uemeson José dos Santos:** Writing – review & editing, Data curation. **Jean Michel Moura-Bueno:** Writing – review & editing, Methodology, Investigation. **Renata Ranielly Pedroza Cruz:** Writing – review & editing, Methodology, Formal analysis. **Demetrius David da Silva:** Writing – review & editing, Methodology, Investigation. **Everardo Valadares de Sá Barreto Sampaio:** Writing – review & editing, Data curation.

Ethics approval

This article does not contain any studies or data with human or animal subjects.

Funding

This research was funded by the Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), grant number APQ-01562-23; by the Coordenação de Aperfeiçoamento Pessoal de Nível Superior (CAPES), Finance code 001; and also by the CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.geoderma.2025.117288.

Data availability

Data will be made available on request.

References

Adi, S.H., Grunwald, S., Tafakresnanto, C., 2019. Fusing environmental variables into soil spectroscopy modeling using a novel two-step regression method. IOP Conf. Ser. Earth Environ. Sci. 393, 012100. https://doi.org/10.1088/1755-1315/393/1/012100.

Alvares, C.A., Stape, J.L., Sentelhas, P.C., de Moraes Gonçalves, J.L., Sparovek, G., 2013. Köppen's climate classification map for Brazil. Meteorol. Z. 22, 711–728. https://doi.org/10.1127/0941-2948/2013/0507.

Angelopoulou, T., Balafoutis, A., Zalidis, G., Bochtis, D., 2020. From laboratory to proximal sensing spectroscopy for soil organic carbon estimation—A review. Sustainability 12, 443. https://doi.org/10.3390/su12020443.

ASF, 2022. Copernicus Sentinel data 2017, 2018, and 2019. Retrieved from ASF DAAC, processed by ESA. [WWW Document]. URL https://asf.alaska.edu/ (accessed 11.17.22).

Asner, G.P., 2001. Cloud cover in Landsat observations of the Brazilian Amazon. Int. J. Remote Sens. 22, 3855–3862. https://doi.org/10.1080/01431160010006926.

Baldock, J.A., Skjemstad, J.O., 2000. Role of the soil matrix and minerals in protecting natural organic materials against biological attack. Org. Geochem. 31, 697–710. https://doi.org/10.1016/S0146-6380(00)00049-8.

Barra, I., Haefele, S.M., Sakrabani, R., Kebede, F., 2021. Soil spectroscopy with the use of chemometrics, machine learning and pre-processing techniques in soil diagnosis: Recent advances–A review. TrAC Trends Anal. Chem. 135, 116166. https://doi.org/10.1016/j.trac.2020.116166.

Bellon-Maurel, V., McBratney, A., 2011. Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils – Critical review and research perspectives. Soil Biol. Biochem. 43, 1398–1410. https://doi.org/10.1016/j.soilbio.2011.02.019.

Ben-Dor, E., 1997. The reflectance spectra of organic matter in the visible near-infrared and short wave infrared region (400–2500 nm) during a controlled decomposition process. Remote Sens. Environ. 61, 1–15. https://doi.org/10.1016/S0034-4257(96)00120-4.

Ben-Dor, E., Chabrillat, S., Demattê, J.A.M., Taylor, G.R., Hill, J., Whiting, M.L., Sommer, S., 2009. Using Imaging Spectroscopy to study soil properties. Remote Sens. Environ. 113, S38–S55. https://doi.org/10.1016/j.rse.2008.09.019.

Bhogapurapu, N., Dey, S., Mandal, D., Bhattacharya, A., Karthikeyan, L., McNairn, H., Rao, Y.S., 2022. Soil moisture retrieval over croplands using dual-pol L-band GRD SAR data. Remote Sens. Environ. 271. https://doi.org/10.1016/J.RSE.2022.112900.

da C. Bispo, P., Rodríguez-Veiga, P., Zimbres, B., do Couto de Miranda, S., Henrique Giusti Cezare, C., Fleming, S., Baldacchino, F., Louis, V., Rains, D., Garcia, M., Del Bon Espírito-Santo, F., Roitman, I., Pacheco-Pascagaza, A.M., Gou, Y., Roberts, J., Barrett, K., Ferreira, L.G., Shimbo, J.Z., Alencar, A., Bustamante, M., Woodhouse, I. H., Eyji Sano, E., Ometto, J.P., Tansey, K., Balzter, H., 2020. Woody aboveground biomass mapping of the brazilian savanna with a multi-sensor and machine learning approach. Remote Sens. 12, 2685. https://doi.org/10.3390/rs12172685.

Brown, D.J., Bricklemyer, R.S., Miller, P.R., 2005. Validation requirements for diffuse reflectance soil characterization models with a case study of VNIR soil C prediction in Montana. Geoderma 129, 251–267. https://doi.org/10.1016/j.geoderma.2005.01.001.

Brus, D.J., Kempen, B., Heuvelink, G.B.M., 2011. Sampling for validation of digital soil maps. Eur. J. Soil Sci. 62, 394–407. https://doi.org/10.1111/j.1365-2389.2011.01364.x.

Camargo, L.A., do Amaral, L.R., dos Reis, A.A., Brasco, T.L., Magalhães, P.S.G., 2022. Improving soil organic carbon mapping with a field-specific calibration approach through diffuse reflectance spectroscopy and machine learning algorithms. Soil Use Manag. 38, 292–303. https://doi.org/10.1111/sum.12775.

Carrasco, L., O'Neil, A., Morton, R., Rowland, C., 2019. Evaluating combinations of temporally aggregated Sentinel-1, Sentinel-2 and Landsat 8 for land cover mapping with Google Earth Engine. Remote Sens. 11, 288. https://doi.org/10.3390/rs11030288.

Clark, R.N., Roush, T.L., 1984. Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. J. Geophys. Res. Solid Earth 89, 6329–6340. https://doi.org/10.1029/JB089iB07p06329.

Coates, J., 2006. Interpretation of infrared spectra, a practical approach. Encyclopedia of Analytical Chemistry. John Wiley & Sons, Ltd, Chichester, UK https://doi.org/10.1002/9780470027318.a5606.

de Souza, J.J.L.L., Souza, B.I., Xavier, R.A., Cardoso, E.C.M., de Medeiros, J.R., da Fonseca, C.F., Schaefer, C.E.G.R., 2022. Organic carbon rich-soils in the brazilian semiarid region and paleoenvironmental implications. CATENA 212, 106101. https://doi.org/10.1016/j.catena.2022.106101.

De Zan, F., Guarnieri, A.M., 2006. TOPSAR: terrain observation by progressive scans. IEEE Trans. Geosci. Remote Sens. https://doi.org/10.1109/TGRS.2006.873853.

Demattê, J.A.M., Dotto, A.C., Bedin, L.G., Sayão, V.M., Souza, A.B.E., 2019a. Soil analytical quality control by traditional and spectroscopy techniques: Constructing the future of a hybrid laboratory for low environmental impact. Geoderma 337, 111–121. https://doi.org/10.1016/j.geoderma.2018.09.010.

Demattê, J.A.M., Dotto, A.C., Paiva, A.F.S., Sato, M.V., Dalmolin, R.S.D., de Araújo, M., do, S.B., da Silva, E.B., Nanni, M.R., ten Caten, A., Noronha, N.C., Lacerda, M.P.C., de Araújo Filho, J.C., Rizzo, R., Bellinaso, H., Francelino, M.R., Schaefer, C.E.G.R., Vicente, L.E., dos Santos, U.J., de Sá Barretto Sampaio, E.V., Menezes, R.S.C., de Souza, J.J.L.L., Abrahão, W.A.P., Coelho, R.M., Grego, C.R., Lani, J.L., Fernandes, A. R., Gonçalves, D.A.M., Silva, S.H.G., de Menezes, M.D., Curi, N., Couto, E.G., dos Anjos, L.H.C., Ceddia, M.B., Pinheiro, É.F.M., Grunwald, S., Vasques, G.M., Marques Júnior, J., da Silva, A.J., de Barreto, M.C.V., Nóbrega, G.N., da Silva, M.Z., de Souza, S.F., Valladares, G.S., Viana, J.H.M., da Silva Terra, F., Horák-Terra, I., Fiorio, P.R., da Silva, R.C., Frade Júnior, E.F., Lima, R.H.C., Alba, J.M.F., de Souza Junior, V.S., Brefin, M.D.L.M.S., Ruivo, M.D.L.P., Ferreira, T.O., Brait, M.A., Caetano, N.R., Bringhenti, I., de Sousa Mendes, W., Safanelli, J.L., Guimarães, C.C.B., Poppiel, R.R., e Souza, A.B., Quesada, C.A., do Couto, H.T.Z., 2019b. The Brazilian Soil Spectral Library (BSSL): A general view, application and challenges. Geoderma 354, 113793. https://doi.org/10.1016/j.geoderma.2019.05.043.

dos Santos, E.P., da Silva, D.D., do Amaral, C.H., 2021. Vegetation cover monitoring in tropical regions using SAR-C dual-polarization index: seasonal and spatial influences. Int. J. Remote Sens. 42, 7581–7609. https://doi.org/10.1080/01431161.2021.1959955.

dos Santos, E.P., da Silva, D.D., do Amaral, C.H., Fernandes-Filho, E.I., Dias, R.L.S., 2022. A Machine Learning approach to reconstruct cloudy affected vegetation indices imagery via data fusion from Sentinel-1 and Landsat 8. Comput. Electron. Agric. 194, 106753. https://doi.org/10.1016/j.compag.2022.106753.

dos Santos, E.P., Moreira, M.C., Fernandes-Filho, E.I., Demattê, J.A.M., dos Santos, U.J., da Silva, D.D., Cruz, R.R.P., Moura-Bueno, J.M., Santos, I.C., de Sampaio, E.V., S.b.,, 2023a. Improving the generalization error and transparency of regression models to estimate soil organic carbon using soil reflectance data. Ecol. Inform. 77, 102240. https://doi.org/10.1016/j.ecoinf.2023.102240.

ESA, 2012. Sentinel-1: ESA's Radar Observatory Mission for GMES Operational Services. European Space Agency.

ESA, E.S.A., 2022. Sentinel-1 SAR Technical Guide [WWW Document]. URL https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-1-sar (accessed 11.18.22).

Fao, 2020. A protocol for measurement, monitoring, reporting and verification of soil organic carbon in agricultural landscapes: GSOC-MRV Protocol. FAO, Rome, Italy. https://doi.org/10.4060/cb0509en.
FAO, 2024. Time to Address Global Soil Monitoring? 9. Food and Agriculture Organization of the United Nations, Rome, Italy.
Ferreira, A.C.C., Leite, L.F.C., de Araújo, A.S.F., Eisenhauer, N., 2016. Land-use type effects on soil organic carbon and microbial properties in a semi-arid region of northeast Brazil. Land Degrad. Dev. 27, 171–178. https://doi.org/10.1002/ldr.2282.
Filipponi, F., 2019. Sentinel-1 GRD Preprocessing Workflow. Proceedings 18, 11. https://doi.org/10.3390/ECRS-3-06201.
Flores-Anderson, A.I., Herndon, K.E., Thapa, R.B., Cherrington, E. (Eds.), 2019. The Synthetic Aperture Radar (SAR) Handbook: Comprehensive Methodologies for Forest Monitoring and Biomass Estimation. NASA, Huntsville https://doi.org/10.25966/nr2c-s697.
Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. 33, 1–22, 10.18637/jss.v033.i01.
Frison, P.-L., Fruneau, B., Kmiha, S., Soudani, K., Dufrêne, E., Toan, T.L., Koleck, T., Villard, L., Mougin, E., Rudant, J.-P., 2018. Potential of Sentinel-1 data for monitoring temperate mixed forest phenology. Remote Sens. 10, 2049. https://doi.org/10.3390/rs10122049.
Gomes, L.C., Faria, R.M., de Souza, E., Veloso, G.V., Schaefer, C.E.G.R., Filho, E.I.F., 2019. Modelling and mapping soil organic carbon stocks in Brazil. Geoderma 340, 337–350. https://doi.org/10.1016/j.geoderma.2019.01.007.
Greenwell, B., Boehmke, B., Gray, B., 2020. vip: Variable Importance Plots.
Greenwell, B.M., Boehmke, B.C., 2020. Variable Importance Plots—An Introduction to the vip Package. R J. 12, 343. https://doi.org/10.32614/RJ-2020-013.
Guo, L.B., Gifford, R.M., 2002. Soil carbon stocks and land use change: a meta analysis. Glob. Change Biol. 8, 345–360. https://doi.org/10.1046/J.1354-1013.2002.00486.X.
Guyon, I., Weston, J., Barnhill, S., Vapnik, V., 2002. Gene selection for cancer classification using support vector machines. Mach. Learn. 46, 389–422. https://doi.org/10.1023/A:1012487302797.
Hobley, E., Wilson, B., Wilkie, A., Gray, J., Koen, T., 2015. Drivers of soil organic carbon storage and vertical distribution in Eastern Australia. Plant Soil 390, 111–127. https://doi.org/10.1007/s11104-015-2380-1.
Hobley, E.U., Baldock, J., Wilson, B., 2016. Environmental and human influences on organic carbon fractions down the soil profile. Agric. Ecosyst. Environ. 223, 152–166. https://doi.org/10.1016/j.agee.2016.03.004.
Jaconi, A., Don, A., Freibauer, A., 2017. Prediction of soil organic carbon at the country scale: stratification strategies for near-infrared data. Eur. J. Soil Sci. 68, 919–929. https://doi.org/10.1111/ejss.12485.
James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning, Performance Evaluation, Springer Texts in Statistics. Springer, New York, New York, NY, 10.1007/978-1-4614-7138-7.
Jobbágy, E.G., Jackson, R.B., 2000. The vertical distribution of soil organic carbon and its relation to climate and vegetation. Ecol. Appl. 10, 423–436. https://doi.org/10.1890/1051-0761(2000)010[0423:TVDOSO]2.0.CO;2.
Joshi, N., Mitchard, E.T.A., Brolly, M., Schumacher, J., Fernández-Landa, A., Johannsen, V.K., Marchamalo, M., Fensholt, R., 2017. Understanding "saturation" of radar signals over forests. Sci. Rep. 7, 1–11. https://doi.org/10.1038/s41598-017-03469-3.
Knadel, M., Thomsen, A., Schelde, K., Greve, M.H., 2015. Soil organic carbon and particle sizes mapping using vis–NIR, EC and temperature mobile sensor platform. Comput. Electron. Agric. 114, 134–144. https://doi.org/10.1016/j.compag.2015.03.013.
Kuhn, M., 2008. Building predictive models in R using the caret package. J. Stat. Soft. 28. https://doi.org/10.18637/jss.v028.i05.
Kuhn, M., Johnson, K., 2013. Applied Predictive Modeling. Springer New York, New York, NY, 10.1007/978-1-4614-6849-3.
Kuhn, M., Quinlan, R., 2023. Cubist: Rule- and Instance-Based Regression Modeling.
Kunkel, V.R., Wells, T., Hancock, G.R., 2022. Modelling soil organic carbon using vegetation indices across large catchments in eastern Australia. Sci. Total Environ. 817, 152690. https://doi.org/10.1016/J.SCITOTENV.2021.152690.
Lal, R., Smith, P., Jungkunst, H.F., Mitsch, W.J., Lehmann, J., Nair, P.K.R., McBratney, A. B., de Sá, J.C.M., Schneider, J., Zinn, Y.L., Skorupa, A.L.A., Zhang, H.-L., Minasny, B., Srinivasrao, C., Ravindranath, N.H., 2018. The carbon sequestration potential of terrestrial ecosystems. J. Soil Water Conserv. 73, 145A–A152. https://doi.org/10.2489/jswc.73.6.145A.
Malmir, M., Tahmasbian, I., Xu, Z., Farrar, M.B., Bai, S.H., 2019. Prediction of soil macro- and micro-elements in sieved and ground air-dried soils using laboratory-based hyperspectral imaging technique. Geoderma 340, 70–80. https://doi.org/10.1016/j.geoderma.2018.12.049.
Mandal, D., Kumar, V., Ratha, D., Dey, S., Bhattacharya, A., Lopez-Sanchez, J.M., McNairn, H., Rao, Y.S., 2020a. Dual polarimetric radar vegetation index for crop growth monitoring using sentinel-1 SAR data. Remote Sens. Environ. 247, 111954. https://doi.org/10.1016/j.rse.2020.111954.
Mandal, D., Ratha, D., Bhattacharya, A., Kumar, V., McNairn, H., Rao, Y.S., Frery, A.C., 2020b. A radar vegetation index for crop monitoring using compact polarimetric SAR data. IEEE Trans. Geosci. Remote Sens. 58, 6321–6335. https://doi.org/10.1109/TGRS.2020.2976661.
Mayes, M., Marin-Spiotta, E., Szymanski, L., Akif Erdoğan, M., Ozdoğan, M., Clayton, M., 2014. Soil type mediates effects of land use on soil carbon and nitrogen in the Konya Basin, Turkey. Geoderma 232–234, 517–527. https://doi.org/10.1016/j.geoderma.2014.06.002.
McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52. https://doi.org/10.1016/S0016-7061(03)00223-4.
McBride, M.B., 2022. Estimating soil chemical properties by diffuse reflectance spectroscopy: Promise versus reality. Eur. J. Soil Sci. 73, e13192. https://doi.org/10.1111/ejss.13192.
McKight, P.E., Najab, J., 2010. Kruskal-Wallis Test. The Corsini Encyclopedia of Psychology. John Wiley & Sons, Ltd, pp. 1–1 https://doi.org/10.1002/9780470479216.corpsy0491.
Medeiros, A. de S., Maia, S.M.F., dos Santos, T.C., de A. Gomes, T.C., 2020. Losses and gains of soil organic carbon in grasslands in the Brazilian semi-arid region. Sci. Agric. 78, e20190076. https://doi.org/10.1590/1678-992X-2019-0076.
de S. Mendes, W., Demattê, J.A.M., Rosin, N.A., da S. Terra, F., Poppiel, R.R., Urbina-Salazar, D.F., Boechat, C.L., Silva, E.B., Curi, N., Silva, S.H.G., José dos Santos, U., Souza Valladares, G., 2022. The Brazilian soil mid-infrared spectral library: the power of the fundamental range. Geoderma 415, 115776. https://doi.org/10.1016/j.geoderma.2022.115776.
Meng, X., Bao, Y., Zhang, X., Wang, X., Liu, H., 2022. Prediction of soil organic matter using different soil classification hierarchical level stratification strategies and spectral characteristic parameters. Geoderma 411, 115696. https://doi.org/10.1016/j.geoderma.2022.115696.
Meyer, H., 2021. CAST: "caret" Applications for Spatial-Temporal Models.
Meyer, H., Reudenbach, C., Hengl, T., Katurji, M., Nauss, T., 2018. Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. Environ. Model. Softw. 101, 1–9. https://doi.org/10.1016/j.envsoft.2017.12.001.
Michaletz, S.T., Cheng, D., Kerkhoff, A.J., Enquist, B.J., 2014. Convergence of terrestrial plant production across global climate gradients. Nature 512, 39–43. https://doi.org/10.1038/nature13470.
Minasny, B., McBratney, A.B., 2016. Digital soil mapping: A brief history and some lessons. Geoderma 264, 301–311. https://doi.org/10.1016/j.geoderma.2015.07.017.
Mishra, U., Yeo, K., Adhikari, K., Riley, W.J., Hoffman, F.M., Hudson, C., Gautam, S., 2022. Empirical relationships between environmental factors and soil organic carbon produce comparable prediction accuracy to machine learning. Soil Sci. Soc. Am. J. https://doi.org/10.1002/saj2.20453.
Mitchard, E.T.A., Saatchi, S.S., Lewis, S.L., Feldpausch, T.R., Woodhouse, I.H., Sonké, B., Rowland, C., Meir, P., 2011. Measuring biomass changes due to woody encroachment and deforestation/degradation in a forest–savanna boundary region of central Africa using multi-temporal L-band radar backscatter. Remote Sens. Environ., DESDynI VEG-3D Special Issue 115, 2861–2873. https://doi.org/10.1016/j.rse.2010.02.022.
Moura-Bueno, J.M., Dalmolin, R.S.D., Horst-Heinen, T.Z., Grunwald, S., ten Caten, A., 2021. Environmental covariates improve the spectral predictions of organic carbon in subtropical soils in southern Brazil. Geoderma 393, 114981. https://doi.org/10.1016/j.geoderma.2021.114981.
Moura-Bueno, J.M., Dalmolin, R.S.D., Horst-Heinen, T.Z., ten Caten, A., Vasques, G.M., Dotto, A.C., Grunwald, S., 2020. When does stratification of a subtropical soil spectral library improve predictions of soil organic carbon content? Sci. Total Environ. 737, 139895. https://doi.org/10.1016/j.scitotenv.2020.139895.
Moura-Bueno, J.M., Dalmolin, R.S.D., ten Caten, A., Dotto, A.C., Demattê, J.A.M., 2019. Stratification of a local VIS-NIR-SWIR spectral library by homogeneity criteria yields more accurate soil organic carbon predictions. Geoderma 337, 565–581. https://doi.org/10.1016/j.geoderma.2018.10.015.
NASA JPL, 2020. NASADEM Merged DEM Global 1 arc second V001 . NASA EOSDIS Land Processes DAAC [WWW Document]. https://doi.org/10.5067/MEaSUREs/NASADEM/NASADEM_HGT.001.
Nguyen, T.T., Pham, T.D., Nguyen, C.T., Delfos, J., Archibald, R., Dang, K.B., Hoang, N. B., Guo, W., Ngo, H.H., 2022. A novel intelligence approach based active and ensemble learning for agricultural soil organic carbon prediction using multispectral and SAR data fusion. Sci. Total Environ. 804, 150187. https://doi.org/10.1016/j.scitotenv.2021.150187.
Nocita, M., Stevens, A., van Wesemael, B., Aitkenhead, M., Bachmann, M., Barthès, B., Ben Dor, E., Brown, D.J., Clairotte, M., Csorba, A., Dardenne, P., Demattê, J.A.M., Genot, V., Guerrero, C., Knadel, M., Montanarella, L., Noon, C., Ramirez-Lopez, L., Robertson, J., Sakai, H., Soriano-Disla, J.M., Shepherd, K.D., Stenberg, B., Towett, E. K., Vargas, R., Wetterlind, J., 2015. Soil spectroscopy: An alternative to wet chemistry for soil monitoring. In: Advances in Agronomy, pp. 139–159 https://doi.org/10.1016/bs.agron.2015.02.002.
Paustian, K., Collier, S., Baldock, J., Burgess, R., Creque, J., DeLonge, M., Dungait, J., Ellert, B., Frank, S., Goddard, T., Govaerts, B., Grundy, M., Henning, M., Izaurralde, R.C., Madaras, M., McConkey, B., Porzig, E., Rice, C., Searle, R., Seavy, N., Skalsky, R., Mulhern, W., Jahn, M., 2019. Quantifying carbon for agricultural soil management: from the current status toward a global soil information system. Carbon Manag. 10, 567–587. https://doi.org/10.1080/17583004.2019.1633231.
Periasamy, S., 2018. Significance of dual polarimetric synthetic aperture radar in biomass retrieval: An attempt on Sentinel-1. Remote Sens. Environ. 217, 537–549. https://doi.org/10.1016/j.rse.2018.09.003.
Poggio, M., Brown, D.J., Bricklemyer, R.S., 2017. Comparison of Vis–NIR on in situ, intact core and dried, sieved soil to estimate clay content at field to regional scales. Eur. J. Soil Sci. 68, 434–448. https://doi.org/10.1111/ejss.12434.
Pudełko, A., Chodak, M., 2020. Estimation of total nitrogen and organic carbon contents in mine soils with NIR reflectance spectroscopy and various chemometric methods. Geoderma 368, 114306. https://doi.org/10.1016/j.geoderma.2020.114306.
Quinlan, J., 1992. Learning with continuous classes. Proceedings AI'92. World Scientific, Singapore.
R Core Team, R., 2023. R: A Language and Environment for Statistical Computing.

Rosin, N.A., Dalmolin, R.S.D., Horst-Heinen, T.Z., Moura-Bueno, J.M., da Silva-Sangoi, D.V., da Silva, L.S., 2020. Diffuse reflectance spectroscopy for estimating soil organic carbon and make nitrogen recommendations. Sci. Agric. 78, e20190246. https://doi.org/10.1590/1678-992X-2019-0246.

Roy, D.P., Yan, L., 2020. Robust Landsat-based crop time series modelling. Remote Sens. Environ. 238, 110810. https://doi.org/10.1016/j.rse.2018.06.038.

Saatchi, S., 2019. SAR methods for mapping and monitoring forest biomass. In: Flores-Anderson, A.I., Herndon, K.E., Thapa, R.B., Cherrington, E. (Eds.), The Synthetic Aperture Radar (SAR) Handbook: Comprehensive Methodologies for Forest Monitoring and Biomass Estimation. NASA, Huntsville https://doi.org/10.25966/nr2c-s697.

Sabetizade, M., Gorji, M., Roudier, P., Zolfaghari, A.A., Keshavarzi, A., 2021. Combination of MIR spectroscopy and environmental covariates to predict soil organic carbon in a semi-arid region. CATENA 196, 104844. https://doi.org/10.1016/j.catena.2020.104844.

Santoro, M., Cartus, O., Carvalhais, N., Rozendaal, D.M.A., Avitabile, V., Araza, A., de Bruin, S., Herold, M., Quegan, S., Rodríguez-Veiga, P., Balzter, H., Carreiras, J., Schepaschenko, D., Korets, M., Shimada, M., Itoh, T., Moreno Martínez, Á., Cavlovic, J., Cazzolla Gatti, R., da Conceição Bispo, P., Dewnath, N., Labrière, N., Liang, J., Lindsell, J., Mitchard, E.T.A., Morel, A., Pacheco Pascagaza, A.M., Ryan, C. M., Slik, F., Vaglio Laurin, G., Verbeeck, H., Wijaya, A., Willcock, S., 2021. The global forest above-ground biomass pool for 2010 estimated from high-resolution satellite observations. Earth Syst. Sci. Data 13, 3927–3950. https://doi.org/10.5194/essd-13-3927-2021.

dos Santos, E.P., Moreira, M.C., Fernandes-Filho, E.I., Demattê, J.A.M., Dionizio, E.A., da Silva, D.D., Cruz, R.R.P., Moura-Bueno, J.M., dos Santos, U.J., Costa, M.H., 2023b. Sentinel-1 imagery used for estimation of soil organic carbon by dual-polarization SAR vegetation indices. Remote Sens. 15, 5464. https://doi.org/10.3390/rs15235464.

dos Santos, U.J., de M. Demattê, J.A., Menezes, R.S.C., Dotto, A.C., Guimarães, C.C.B., Alves, B.J.R., Primo, D.C., de S.B. Sampaio, E.V., 2020. Predicting carbon and nitrogen by visible near-infrared (Vis-NIR) and mid-infrared (MIR) spectroscopy in soils of Northeast Brazil. Geoderma Reg. 23, e00333. https://doi.org/10.1016/j.geodrs.2020.e00333.

Shafizadeh-Moghadam, H., Minaei, F., Talebi-khiyavi, H., Xu, T., Homaee, M., 2022. Synergetic use of multi-temporal Sentinel-1, Sentinel-2, NDVI, and topographic factors for estimating soil organic carbon. CATENA 212, 106077. https://doi.org/10.1016/J.CATENA.2022.106077.

Six, J., Conant, R.T., Paul, E.A., Paustian, K., 2002. Stabilization mechanisms of soil organic matter: Implications for C-saturation of soils. Plant Soil 241, 155–176. https://doi.org/10.1023/A:1016125726789.

Small, D., 2011. Flattening Gamma: Radiometric Terrain Correction for SAR Imagery. IEEE Trans. Geosci. Remote Sens. 49, 3081–3093. https://doi.org/10.1109/TGRS.2011.2120616.

Smith, P., Soussana, J., Angers, D., Schipper, L., Chenu, C., Rasse, D.P., Batjes, N.H., Egmond, F., McNeill, S., Kuhnert, M., Arias-Navarro, C., Olesen, J.E., Chirinda, N., Fornara, D., Wollenberg, E., Álvaro-Fuentes, J., Sanz-Cobena, A., Klumpp, K., 2020. How to measure, report and verify soil carbon change to realize the potential of soil carbon sequestration for atmospheric greenhouse gas removal. Glob. Change Biol. 26, 219–241. https://doi.org/10.1111/gcb.14815.

Sørensen, R., Zinko, U., Seibert, J., 2006. On the calculation of the topographic wetness index: evaluation of different methods based on field observations. Hydrol. Earth Syst. Sci. 10, 101–112. https://doi.org/10.5194/hess-10-101-2006.

Soriano-Disla, J.M., Janik, L.J., Viscarra Rossel, R.A., Macdonald, L.M., McLaughlin, M. J., 2014. The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties. Appl. Spectrosc. Rev. 49, 139–186. https://doi.org/10.1080/05704928.2013.811081.

Sothe, C., Gonsamo, A., Arabian, J., Snider, J., 2022. Large scale mapping of soil organic carbon concentration with 3D machine learning and satellite observations. Geoderma 405, 115402. https://doi.org/10.1016/j.geoderma.2021.115402.

Souza, C.M., Shimbo, Z.J., Rosa, M.R., Parente, L.L., Alencar, A.A., Rudorff, B.F.T., Hasenack, H., Matsumoto, M., Ferreira, G.L., Souza-Filho, P.W.M., de Oliveira, S.W., Rocha, W.F., Fonseca, A.V., Marques, C.B., Diniz, C.G., Costa, D., Monteiro, D., Rosa, E.R., Vélez-Martin, E., Weber, E.J., Lenti, F.E.B., Paternost, F.F., Pareyn, F.G. C., Siqueira, J.V., Viera, J.L., Neto, L.C.F., Saraiva, M.M., Sales, M.H., Salgado, M.P. G., Vasconcelos, R., Galano, S., Mesquita, V.V., Azevedo, T., 2020. Reconstructing three decades of land use and land cover changes in Brazilian biomes with landsat archive and earth engine. Remote Sens. 12, 2735. https://doi.org/10.3390/rs12172735.

Stevens, A., Nocita, M., Tóth, G., Montanarella, L., van Wesemael, B., 2013. Prediction of soil organic carbon at the european scale by visible and near infrared reflectance spectroscopy. PLoS ONE 8, e66409. https://doi.org/10.1371/journal.pone.0066409.

Stewart, C.E., Plante, A.F., Paustian, K., Conant, R.T., Six, J., 2008. Soil carbon saturation: linking concept and measurable carbon pools. Soil Sci. Soc. Am. J. 72, 379–392. https://doi.org/10.2136/sssaj2007.0104.

Tay, J.K., Narasimhan, B., Hastie, T., 2023. Elastic net regularization paths for all generalized linear models. J. Stat. Softw. 106, 1–31. https://doi.org/10.18637/jss.v106.i01.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B Methodol. 58, 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

Vasques, G.M., Grunwald, S., Harris, W.G., 2010. Spectroscopic models of soil organic carbon in Florida. USA. J. Environ. Qual. 39, 923–934. https://doi.org/10.2134/jeq2009.0314.

Viscarra Rossel, R.A., Behrens, T., 2010. Using data mining to model and interpret soil diffuse reflectance spectra. Geoderma 158, 46–54. https://doi.org/10.1016/j.geoderma.2009.12.025.

Viscarra Rossel, R.A., Behrens, T., Ben-Dor, E., Brown, D.J., Demattê, J.A.M., Shepherd, K.D., Shi, Z., Stenberg, B., Stevens, A., Adamchuk, V., Aïchi, H., Barthès, B.G., Bartholomeus, H.M., Bayer, A.D., Bernoux, M., Böttcher, K., Brodský, L., Du, C.W., Chappell, A., Fouad, Y., Genot, V., Gomez, C., Grunwald, S., Gubler, A., Guerrero, C., Hedley, C.B., Knadel, M., Morrás, H.J.M., Nocita, M., Ramirez-Lopez, L., Roudier, P., Campos, E.M.R., Sanborn, P., Sellitto, V.M., Sudduth, K.A., Rawlins, B.G., Walter, C., Winowiecki, L.A., Hong, S.Y., Ji, W., 2016. A global spectral library to characterize the world's soil. Earth-Sci. Rev. 155, 198–230. https://doi.org/10.1016/j.earscirev.2016.01.012.

Viscarra Rossel, R.A., Behrens, T., Ben-Dor, E., Chabrillat, S., Demattê, J.A.M., Ge, Y., Gomez, C., Guerrero, C., Peng, Y., Ramirez-Lopez, L., Shi, Z., Stenberg, B., Webster, R., Winowiecki, L., Shen, Z., 2022. Diffuse reflectance spectroscopy for estimating soil properties: A technology for the 21st century. Eur. J. Soil Sci. 73, e13271. https://doi.org/10.1111/ejss.13271.

Viscarra Rossel, R.A., McBratney, A.B., Minasny, B. (Eds.), 2010. Proximal Soil Sensing. Springer, Netherlands, Dordrecht. https://doi.org/10.1007/978-90-481-8859-8.

Wadoux, A.-M.-J.-C., 2023. Interpretable spectroscopic modelling of soil with machine learning. Eur. J. Soil Sci. 74, e13370. https://doi.org/10.1111/ejss.13370.

Wang, Z., Ding, J., Zhang, Z., 2022. Estimation of soil organic matter in arid zones with coupled environmental variables and spectral features. Sensors 22, 1194. https://doi.org/10.3390/s22031194.

Wiesmeier, M., Urbanski, L., Hobley, E., Lang, B., von Lützow, M., Marin-Spiotta, E., van Wesemael, B., Rabot, E., Ließ, M., Garcia-Franco, N., Wollschläger, U., Vogel, H.-J., Kögel-Knabner, I., 2019. Soil organic carbon storage as a key function of soils - A review of drivers and indicators at various scales. Geoderma 333, 149–162. https://doi.org/10.1016/j.geoderma.2018.07.026.

Woodhouse, I.H., 2006. Introduction to Microwave Remote Sensing. CRC Press, Boca Raton.

Woodhouse, I.H., Mitchard, E.T.A., Brolly, M., Maniatis, D., Ryan, C.M., 2012. Radar backscatter is not a "direct measure" of forest biomass. Nat. Clim. Change 2, 556–557. https://doi.org/10.1038/nclimate1601.

Yu, M., Wang, Y., Jiang, J., Wang, C., Zhou, G., Yan, J., 2019. Soil organic carbon stabilization in the three subtropical forests: importance of clay and metal oxides. J. Geophys. Res. Biogeosciences 124, 2976–2990. https://doi.org/10.1029/2018JG004995.

Zeng, Y., Hao, D., Huete, A., Dechant, B., Berry, J., Chen, J.M., Joiner, J., Frankenberg, C., Bond-Lamberty, B., Ryu, Y., Xiao, J., Asrar, G.R., Chen, M., 2022. Optical vegetation indices for monitoring terrestrial ecosystems globally. Nat. Rev. Earth Environ. https://doi.org/10.1038/s43017-022-00298-5.

Zheng, G., Jiao, C., Zhou, S., Shang, G., 2016. Analysis of soil chronosequence studies using reflectance spectroscopy. Int. J. Remote Sens. 37, 1881–1901. https://doi.org/10.1080/01431161.2016.1163751.

Zhou, T., Geng, Y., Chen, J., Liu, M., Haase, D., Lausch, A., 2020a. Mapping soil organic carbon content using multi-source remote sensing variables in the Heihe River Basin in China. Ecol. Indic. 114, 106288. https://doi.org/10.1016/J.ECOLIND.2020.106288.

Zhou, T., Geng, Y., Chen, J., Pan, J., Haase, D., Lausch, A., 2020b. High-resolution digital mapping of soil organic carbon and soil total nitrogen using DEM derivatives, Sentinel-1 and Sentinel-2 data based on machine learning algorithms. Sci. Total Environ. 729, 138244. https://doi.org/10.1016/J.SCITOTENV.2020.138244.