

Soil carbon prediction in Marajó island wetlands

David Lukas de Arruda^{(1)*} , João Carlos Ker⁽¹⁾ , Gustavo Vieira Veloso⁽¹⁾ , Renata Jordan Henriques⁽²⁾ , Elpídio Inácio Fernandes-Filho⁽¹⁾ , Danilo de Lima Camêlo⁽³⁾ , Lucas de Carvalho Gomes⁽⁴⁾  and Carlos Ernesto Gonçalves Renaud Schaefer⁽¹⁾ 

⁽¹⁾ Universidade Federal de Viçosa, Departamento de Solos, Viçosa, Minas Gerais, Brasil.

⁽²⁾ Universidade Federal de Minas Gerais, Departamento de Geografia, Belo Horizonte, Minas Gerais, Brasil.

⁽³⁾ Universidade Federal do Espírito Santo, Departamento de Agronomia, Alegre, Espírito Santo, Brasil.

⁽⁴⁾ Aarhus University Department of Agroecology, Aarhus, Denmark.

ABSTRACT: Soil is an important carbon repository in terrestrial ecosystems, serving a fundamental role in the intricate cycling of this elemental component. Wetlands are crucial components of the global carbon cycle, playing a significant role in carbon sequestration due to their remarkable productivity and unique sedimentary structures. Our study focuses on the wetlands east of Marajó island, recognized as the largest fluvial-marine plain in South America. In this study, we applied a methodological framework to optimize SOC content prediction in the wetlands of Marajó island using readily available environmental covariates. We collected and analyzed 81 soil samples from the most representative geoenvironments on the island at a layer of 0.00-0.20 m. Our database included vegetation indices, morphometric maps, and covariates based on distance from water bodies and archaeological sites. We tested five machine learning algorithms - Cubist, Linear Model, Random Forest, K Nearest Neighbor, and Support Vector Machine - to obtain the best prediction performance. Cubist model demonstrated the highest performance for training ($R^2 = 0.483$) and testing ($R^2 = 0.505$) datasets, making it the optimal choice for SOC prediction in the topsoil. The most important covariates selected by Cubist using recursive feature elimination were digital elevation model, topographic heterogeneity index, vertical distance between the summit and base of the slope, and Euclidean distance from water bodies. Geoenvironments characterized by dense alluvial rainforest with palms on Plinthosols and Gleysols, mangroves with Gleysols, and coastal muddy plains exhibited the highest SOC content in the topsoil.

Keywords: soil organic carbon, machine learning, spatial prediction, climate change.

*** Corresponding author:**

E-mail: david.solos@hotmail.com

Received: December 15, 2023

Approved: April 01, 2024

How to cite: Arruda DL, Ker JC, Veloso GV, Henriques RJ, Fernandes-Filho EI, Camêlo DL, Gomes LC, Schaefer CEGR. Soil carbon prediction in Marajó island wetlands. Rev Bras Cienc Solo. 2024;48:e0230162.

<https://doi.org/10.36783/18069657rbc20230162>

Editors: José Miguel Reichert  and Ricardo Simão Diniz Dalmolin .

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



INTRODUCTION

Soil is an important carbon repository in terrestrial ecosystems, playing a key role in the intricate cycling of this elemental component (Doetterl et al., 2015). On a global scale, soil organic carbon (SOC) storage is responsive to shifts in environmental conditions and climate (Nottingham et al., 2020), potentially accelerating climate change by releasing CO₂ and CH₄ into the atmosphere.

Wetlands play a crucial role in the global carbon cycle within natural systems, serving as key contributors to this complex process (Sjögersten et al., 2021; Zhang et al., 2023; Adame et al., 2024). Their remarkable productivity and sedimentary structure contribute to the highest carbon sequestration rate per unit area among all-natural systems (Rogers et al., 2019). Ecosystems in coastal wetlands, including tidal marshes, mangroves, and seagrasses, arouse wide attention and are given high hopes for carbon sinks in the current phase for their particularly efficient carbon sequestration (Duarte et al., 2013; Macreadie et al., 2021). Consequently, accurately assessing the spatial distribution of SOC content in wetlands is imperative for comprehending and monitoring the cycling of this elemental component (Zhang et al., 2023).

The carbon in soils primarily exists in organic form and is regulated by the equilibrium between input (deposition) and output (decomposition) of organic matter (Wiesmeier et al., 2019). In addition to inherent soil characteristics such as mineralogy and texture, SOC distribution is significantly influenced by local factors, including topography, slope, vegetation, and hydrology (Chen et al., 2020; Hein et al., 2020; Zhao et al., 2021).

Determining SOC content often involves manual soil sampling procedures and laboratory analysis, making the process time-consuming, labor-intensive, and challenging to implement in flood-prone and remote areas (Cui et al., 2021). This limitation results in a scarcity of data, particularly in areas of difficult to access (Chi et al., 2021).

The development of digital soil mapping (DSM) and its applications (McBratney et al., 2003) has been employed as a tool in the spatialization of SOC at wetlands and mangroves in various scales (Chi et al., 2021; Hidayatullah et al., 2023; Maxwell et al., 2023; Muñoz et al., 2024). Terrain attributes derived from digital elevation models (DEMs) are among the most commonly used predictor variables in the wetland DSM (Leonard et al., 2012; Kloiber et al., 2015). The topography is particularly important in wetlands because it directly influences the distribution and flow of water, affecting the hydrology of these environments and, consequently, their ecological dynamics, including the formation and maintenance of habitats for wildlife and carbon sequestration (Mitsch et al., 2000). However, predicting SOC remains challenging due to the high variability of factors influencing soil organic carbon (Scharlemann et al., 2014).

Nonlinear models, such as machine learning, are increasingly being used in DSM. However, it is necessary to incorporate more pedological knowledge for reasonable modeling from both scientific and technical perspectives (Chen et al., 2022). Machine learning-based models, including Random Forest (Muñoz et al., 2024), Cubist (Rudiyanto et al., 2018), and Support Vector Machine-SVM (Sun et al., 2023), have demonstrated successful applications in this domain.

The focus of our study includes the wetlands east of Marajó island, recognized as the largest fluvial-marine plain in South America (Rossetti and Valeriano, 2007; Francisquini et al., 2014). Climate projections suggest the eastern portion of the island could face permanent flooding in the coming decades (Barros and Albernaz, 2014). This transformation poses a threat to the livelihoods of the island's population, which is strongly connected to local practices such as hunting, fishing, buffalo breeding, rice cultivation, extractivism, and agriculture (Murrieta et al., 1999; Cohen et al., 2009; Lisboa, 2012).

It is crucial to ascertain the spatial distribution of surface SOC levels using readily accessible environmental covariates. This endeavor holds significant importance for informed decision-making, formulation of public policies concerning environmental

recovery and conservation, and the identification of alternative scenarios for SOC sequestration (Fernandes et al., 2021; Lembaid et al., 2022; Liu et al., 2022). Moreover, these assessments may furnish valuable information to underpin the initiation of new research endeavors, particularly pertinent to Marajó island, characterized by one of the lowest Human Development Index (HDI) values in Brazil (IBGE, 2010).

This study aimed to model the spatial distribution of SOC in the topsoil, considering various environments and utilizing easily obtainable environmental covariates. Our research addresses existing knowledge gaps through innovative approaches, such as utilizing readily accessible environmental covariates, diverse machine learning algorithms, advanced techniques for external validation, and an ecologically sustainable methodology. These contributions can potentially enhance the monitoring of wetland areas, particularly those vulnerable to global climate change. By contributing to improving and maintaining terrestrial ecosystem quality, our investigation aligns with the broader goals of environmental conservation within the soil science domain.

MATERIALS AND METHODS

Study area

Marajó island covers an area of 40,100.00 km² and is divided into two main parts: the highlands with better drainage, ranging from 10 to 40 m above sea level, located in the south and west and covered by tropical forest; and the wetlands to the east, ranging from 0 to 9 m above sea level and experiencing annual floods, with savanna vegetation, sandy ridges, paleochannels, savannas, and mangroves (Lisboa, 2012) (Figure 1a). The focus area of this study is the eastern part of Marajó island, which corresponds to the wetlands and covers an area of 14,464.455 km².

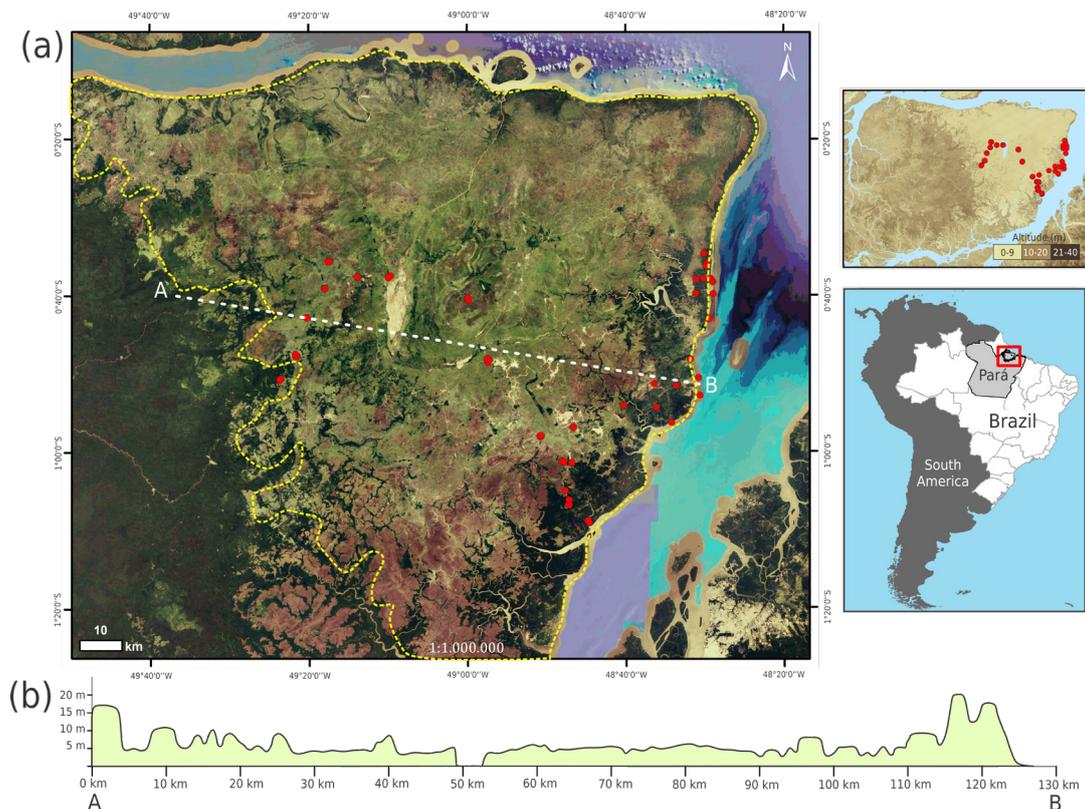


Figure 1. Location of the sampling points in different environments in the wetlands of Marajó island, Pará State, Northern Brazil (a); and the topographic section of wetlands of the island (b).

In the wetlands with fluvial or fluvio-marine influence, *Gleissolos* (Entisols) predominate (Correa et al., 1974), in areas corresponding to ancient mangroves in the central-east of the island (França et al., 2012; Smith et al., 2012), as well as *Vertissolos* (Vertisols) and *Cambissolos* (Inceptisols) distributed sporadically (Correa et al., 1974; Cerri and Volkoff, 1988; Murrieta et al., 1999; Lisboa, 2012). Sandy ridges also stand out in the wetlands, sediments influenced by fluvial or aeolian processes, distributed as *Neossolos Quartzarênicos* (Psamments) in elevated areas and *Espodossolos* (Spodosols) (Tatumi et al., 2008), as well as *Cambissolos Flúvicos* (Fluvents), consistent with paleochannels (Correa et al., 1974; Rossetti and Valeriano, 2007).

Numerous archaeological sites are distributed in these areas, dating to at least 3,000 years before the present (Meggers and Roosevelt, 1992). Many of these sites are concentrated in alluvial paleodykes, also called Tesos, which may reach several hundred meters in length with up to 12 m in height and may occur scattered across the wetland areas.

According to Köppen classification system (Alvares et al., 2013), the climate is monsoon and savannah, transitioning between Am and Aw with an average annual temperature ranging from 25 to 29 °C and average precipitation between 2,000 and 4,000 mm (Lima et al., 2005).

Soil sampling, SOC quantification, and obtaining environmental covariates

A total of 81 composite samples were collected, covering the main geoenvironments of the eastern sector of Marajó island: fluvio-lacustrine plains with grassy vegetation on Gleysols, old salt-plain terrace with fluvio-lacustrine sediments, transition zone between savanna and rainforest, fluvio-lacustrine plains with grassy vegetation on Entisols (Paleolake Arari), terraces with hydromorphic Plinthosols on fluvio-lacustrine sediments, sandy flats with Spodosols under savanna, dense alluvial rainforest with palms on Plinthosols and Gleysols, plateau with Oxisols and secondary vegetation, apicum with Gleysols and halophytic plants, salt-plain with paleodunes, and mangrove with Gleysols (Figure 2). Subsamples were collected in the center and in each cardinal direction (N, S, E, and W) at about 3 to 5 m from the central point and at a depth of 0.00-0.20 m.

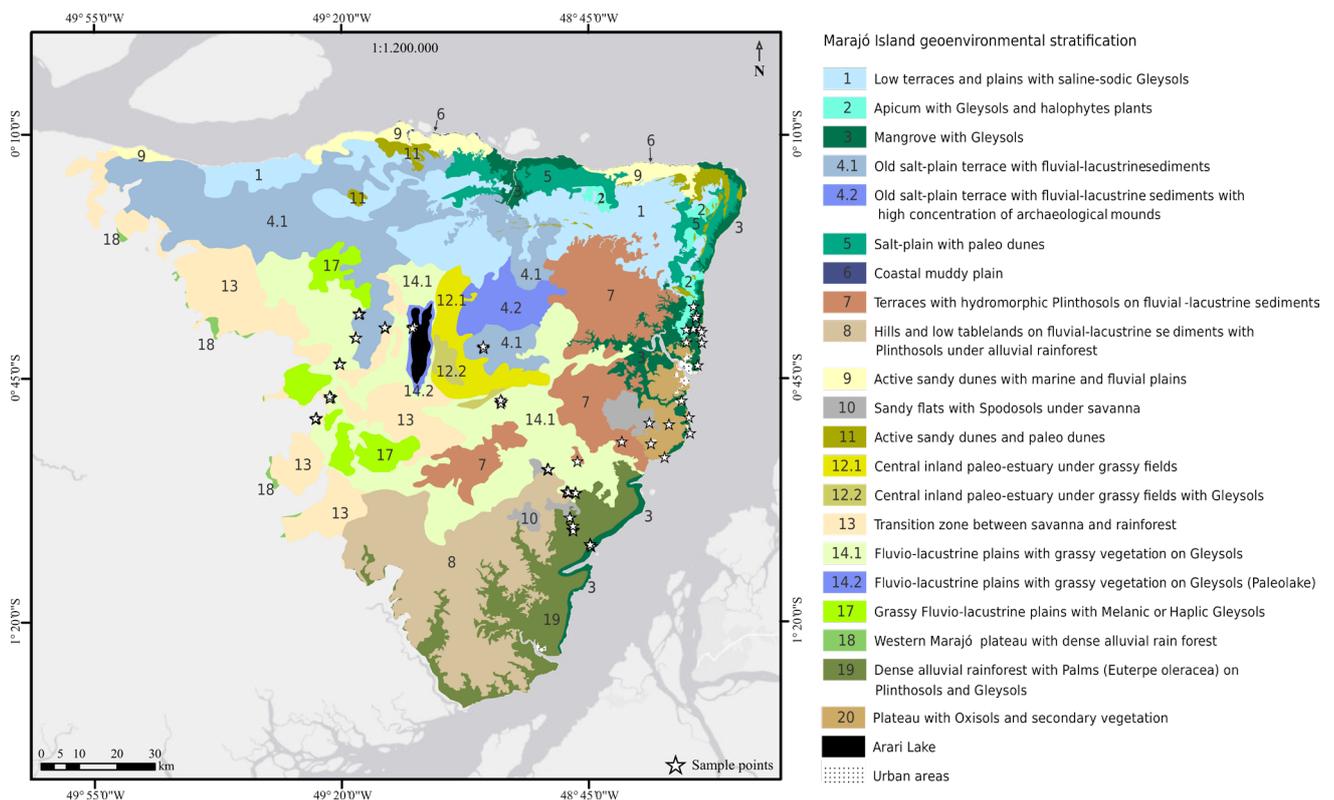


Figure 2. Distribution of sampling points in the wetlands of Marajó island over different geoenvironments.

Due to the difficulty of accessing much of the island's geoenvironments, many of which remain flooded even during the dry season, we focused the sample collection on areas near the roads. However, we did not overlook the flooded environments during this process. Under these circumstances, the samples were obtained near the roads to ensure a comprehensive representation of the various soil types found on the island, including those subject to permanent flooding. Collections were made in November 2019, at the peak of the dry period. Samples were air-dried, ground, and sieved through a 2 mm mesh to obtain the air-dried fine earth fraction (ADFE) for SOC determination. Soil organic carbon was determined using the Walkley-Black method (Teixeira et al., 2017).

Covariates employed in this study are easily accessible and pertain to factors that regulate the SOC dynamics in wetland areas, such as topography, slope, vegetation, and hydrology (Chen et al., 2020; Hein et al., 2020; Zhao et al., 2021). Soil samples geographic coordinates (X and Y) were used as covariates for spatializing SOC contents. The SIRGAS 2000/UTM zone 22S projected coordinate system (EPSG: 31983) was used as the standard during the SOC spatialization process.

Topographic covariables were derived from the NASADEM SRTM-Shuttle Radar Topography Mission digital elevation model (DEM) (LP DAAC, 2003), with a spatial resolution of approximately 30 m (1 arc-second). To obtain the covariables, R *software* version 4.10 (Team, 2021) was used, using the "RSAGA" (Brenning, 2008), "raster" (Robert, 2019), and "rgrass7" (Bivand et al., 2019) packages. The attributes generated based on DEM are described in table 1 (Sena et al., 2021; Paes et al., 2022). Covariate Y was also created and represents the latitude of the data collection points, measuring the distance in degrees from the Equator along the meridian.

Due to the complex drainage network (Rossetti and Valeriano, 2007), and the common presence of archaeological sites with "*Terra Preta de Índio*" - Amazonian Dark Earth (Anthrosol) (Meggers and Roosevelt, 1992), two covariables were created based on the Euclidean distance between the drainage channels and the sampling point, and the Euclidean distance between the archaeological sites and the sampling point (Figure 3).

Landsat 8 images, captured in the same period as the soil samples (November 2019), were employed to calculate the Normalized Difference Vegetation Index (NDVI) using the formula $(\text{Band 5} - \text{Band 4}) / (\text{Band 5} + \text{Band 4})$, as introduced by Rouse Jr et al. (1973). This index normalizes spectral data from the near-infrared and red bands. Despite being one of the earliest indices developed, its widespread use and proven efficacy make it one of the most frequently employed indices in the analysis and prediction of SOC (Gomes et al., 2019; Padilha et al., 2020; Zhang et al., 2023).

Variable selection and comparison of predictive models

The process of covariate removal/selection is designed to construct a model that requires lower computational cost, adhering to the principle of parsimony (Gomes et al., 2019; Reis et al., 2021; Paes et al., 2022). The elimination of covariates involved a three-stage process: firstly, considering variance; secondly, analyzing correlations; and thirdly, evaluating prediction importance for each algorithm (Figure 4).

Elimination based on variance involves removing covariates with minimal variance, as they do not significantly contribute to the predictive model. Additionally, these covariates have the potential to impact certain model performances and increase computational costs (Kern et al., 2017; Hujoel et al., 2018). The "nearZeroVar" function available in the R software Caret package (Kuhn, 2008) facilitated the removal of covariates with very low variance. The eliminated covariables were Hill (HI), Hill index (HINDEX), Valleu Index (VA) and Valley (VA).

Covariate removal based on correlation targets highly correlated variables, as they can impact final model performance and the efficiency of subsequent covariate elimination.

Table 1. Terrain variables generated from a digital elevation model

Terrain attribute	Abbreviation	Brief description
Convergence index	CI	Convergence index/divergence in relation to flow
Cross sectional curvature	CSC	Measure curvature perpendicular to slope direction
Digital elevation model	DEM	Represents the elevation in each cell of the model
Diurnal anisotropic heating	DAH	Continuous exposure-dependent energy measurement
Flow line curvature	FLC	Represents the projection of a gradient line on a horizontal plane
General curvature	GC	The combination of plane and profile bends
Geomorphons	GEO	Associated geometry using machine vision approach
Hill	HI	Analytical choline shading
Hill index	HIINDEX	Analytical index choline shading
Longitudinal curvature	LC	Scares curvature in the direction of the slope
Mass balance index	MBI	Balance index between erosion and deposition
Maximal curvature	MAXC	Maximum curvature in the local normal section
Morphometric Protection Index	MPI	Exposure/protection measure of a surrounding relief point
Mid-slope position	MSP	Represents the distance from top to valley, ranging from 0 to 1
Minimal curvature	MINC	Minimum curvature for local normal section
Multiresolution index of ridge top flatness	MRRTF	Indicates flat positions in high-altitude areas
Multiresolution index of valley bottom flatness	MRVBF	Indicates flat surfaces at the bottom of the valley
Normalized height	NH	Vertical distance between the base and the ridge of the normalized slope
Plan curvature	PLANC	Described as the curvature of the hypothetical contour line that passes through a specific cell
Profile curvature	PROC	Describes the curvature of the surface in the direction of the steepest slope
Real surface area	RSA	Actual calculation of cell area
Slope	S	Represents local angular slope
Slope height	SH	Vertical distance between the base and the ridge of the slope
Slope Index	SI	Represents a local angular slope index
Solrad Diffuse1	SolDiffuse1	Diffuse heatstroke for January
Solrad Diffuse2	SolDiffuse2	Diffuse heatstroke for July
Solrad dur 1	SolDur1	Duration of heat stroke for January
Solrad dur 2	SolDur2	Duration of Heat stroke for July
Solrad Direct1	SolDiret1	Direct heat stroke in January
Solrad Direct2	SolDiret2	Direct heat stroke in July
Solrad Ration1	SolRation1	The ratio between Direct Heat stroke and Diffuse Heatstroke in January
Solrad Ration2	SolRation2	The ratio between Direct Heat stroke and Diffuse Insolation in July
Solrad Sunrise1	SolSunrise1	Average sunrise time for January
Solrad Sunrise2	SolSunrise2	Average sunrise time of July
Solrad Sunset1	SolSunset1	Average sunset time of January
Solrad Sunset2	SolSunset2	Average sunset time in July
Solrad total1	SolTotal1	Total insolation for the month of January
Solrad total2	SolTotal2	Total insolation for the month of July
Standardized height	STANH	Vertical distance between the base and the standardized slope index
Surface specific points	SSP	Indicates differences between specific surface offset points
Tangencial curvature	TANC	Measured in the normal plane in a direction perpendicular to the gradient

Continue

Continuation

Terrain attribute	Abbreviation	Brief description
Terrain ruggedness index	TRI	Quantitative index of topographic heterogeneity
Terrain surface convexity	V	The ratio between the number of cells that have positive curvature and the number of all valid cells within a specified search radius
Terrain surface texture	TST	Divide surface texture into 8, 12, or 16 classes
Total curvature	TC	General measurement of surface curvature
Topographic position index	TPI	Difference between the elevation of the point A and the surrounding elevation
Valley depth	VD	Calculation of vertical distance at the level of the drainage base
Valley	VA	Nebulous valley of calculation using the Cartola approach
Valley Index	VA	Calculation of Diffuse valley index using the Cartola approach
Vector ruggedness measure	VRM	Roughness of the terrain
Topographic wetness index	TWI	Describes the tendency of each cell to accumulate water as a function of relief

Spearman correlation coefficients were initially calculated for all covariates to eliminate highly correlated covariates. Subsequently, pairs of covariates exhibiting a correlation of 95 % or higher were examined, and the covariate with the highest mean absolute correlation value with other variables was removed from the analysis (Paes et al., 2022). The “findcorrelation” function, available in the Caret package of the R software (Kuhn et al., 2020), was employed to remove highly correlated covariates. Eliminated covariables were: Cross sectional curvature (CSC), General curvature (GC), Longitudinal curvature (LC), Total curvature (TC), and Terrain ruggedness index (TRI).

Elimination of covariates with low importance, for each assessed model, aims to achieve a more concise model by identifying the optimal set of covariates. This process is defined by the importance of covariates compared to others in model fitting, through the Recursive Feature Elimination (RFE) tool of the Caret package (Kuhn and Johnson, 2013). The RFE is an interactive method for covariate selection/removal, following the Backward type approach. It determines the optimal set of covariates by assessing various subsets with different quantities of covariates (Kuhn and Johnson, 2013). These subsets are defined by the operator and the complete set, always undergoing evaluation from the largest set to the smallest set, as outlined in Ferreira et al. (2021). Due to their unique characteristics,

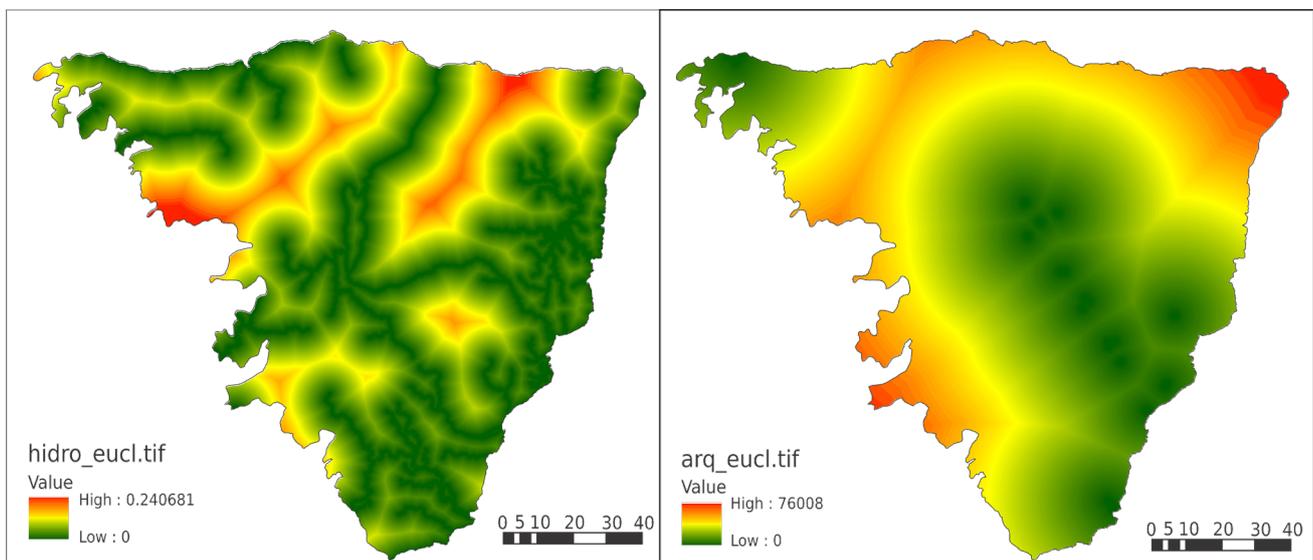


Figure 3. Euclidean distance between the sampling point and the hydrography and the sampling point and archaeological sites.

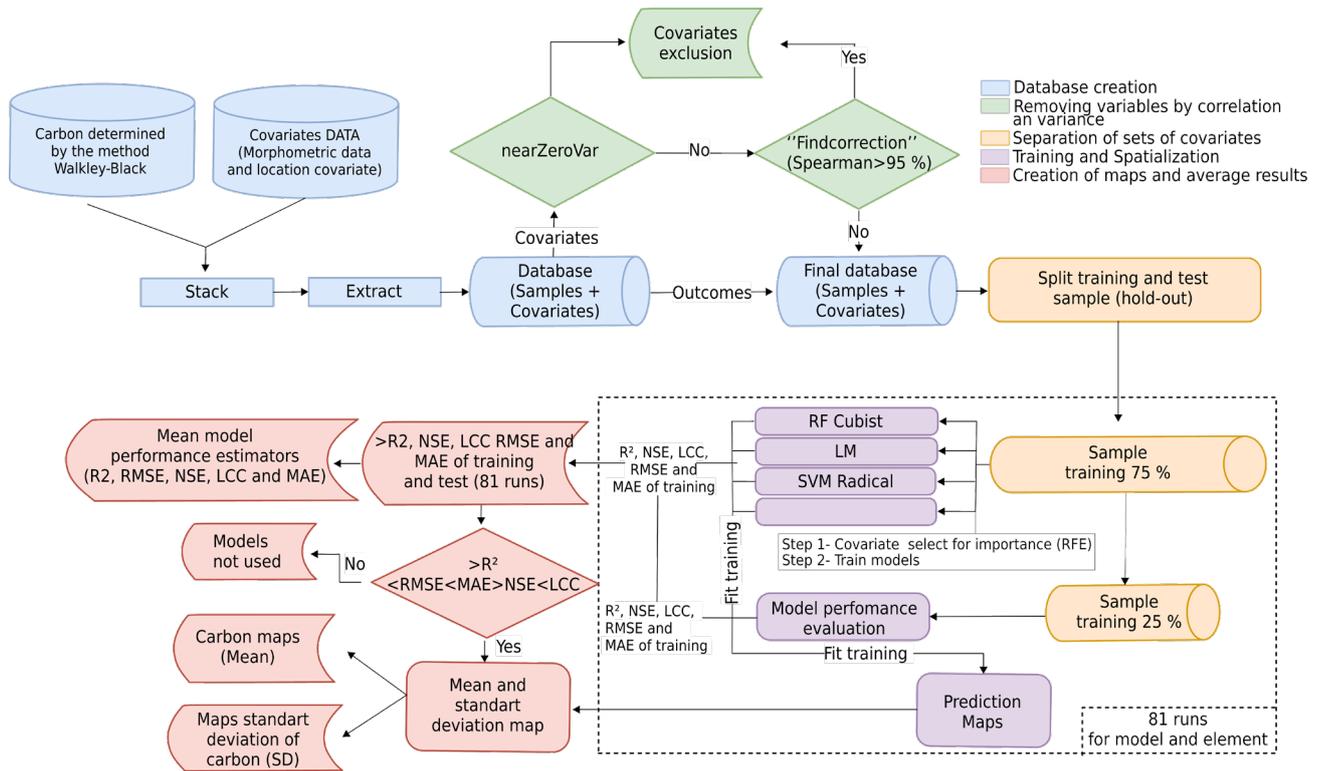


Figure 4. Flowchart showing the sequence of methodologies applied for spatial prediction of soil carbon.

RFE was applied separately to each algorithm under consideration. The defined subsets comprised various combinations of covariates, including the complete set of 38 covariates that passed the elimination steps for very low variance and correlation. The tested models included Cubist (CUB), Random Forest (RF), Support Vector Machine (SVM) with Radial Kernel, Linear Model (Lm) and K Nearest Neighbor (KKNN). Each algorithm has an important calculation methodology, and the functions are listed and considered “support functions” of the RFE (Kuhn and Johnson, 2013).

The dataset was divided into training and testing sets after eliminating covariates with minimal variance and highly correlated ones. Given the limited sample size (81), the Nested leave-one-out cross-validation (“LOOCV”) technique was employed for DATA partitioning (Ferreira et al., 2021; Paes et al., 2022). This method involves two loops: an inner loop and an outer loop. In the inner loop, one sample is excluded (resulting in $n - 2$ samples), while the remaining sample forms the outer loop. This process is repeated n times, where n is the total sample size. All samples traverse the outer loop, enabling the calculation of predicted values. These predictions are then matched with the observed values for the samples, facilitating the evaluation of the model’s performance in the final prediction encompassing the entire set of samples from the outer loop (Figure 4).

Model training was conducted using the optimal set of covariates identified for each selected algorithm through RFE. The final model was fine-tuned using the leave-one-out cross-validation (LOOCV) method, testing five values of internal hyperparameters for each algorithm under evaluation. Mean Absolute Error (MAE) served as the performance parameter for selecting the optimal hyperparameter values. The hyperparameters for each algorithm were sourced from the Caret package manual, chapter 6 (“Models described”), accessible at <https://topepo.github.io/caret/train-models-by-tag.html>. The study evaluated various performance metrics, including MAE (Equation 1), Root Mean Squared Error (RMSE) (Equation 2), Coefficient of Determination (R^2) (Equation 3), Lin’s Concordance Correlation Coefficient (LCCC) (Equation 4), and Nash-Sutcliffe Model Efficiency Coefficient (NSE) (Equation 5).

$$MAE = \frac{\sum_{i=1}^N |P_i - O_i|}{n} \quad \text{Eq. 1}$$

$$RMSE = \left[\frac{1}{N} \sum_{i=1}^N (P_i - O_i)^2 \right]^{\frac{1}{2}} \quad \text{Eq. 2}$$

$$R^2 = \frac{\sum_{i=1}^n (P_i - \overline{om})^2}{\sum_{i=1}^n (Om_i - \overline{om})^2} \quad \text{Eq. 3}$$

$$LCCC = \frac{2\rho\sigma_p\sigma_o}{\sigma_o^2 + \sigma_p^2 + (\overline{O} + \overline{P})} \quad \text{Eq. 4}$$

$$NSE = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \overline{O})^2} \quad \text{Eq. 5}$$

in which: p_i is the predicted sample values; o_i is the observed sample values; \bar{o} is the mean of observed values; σ_p is the variance of predicted samples; σ_o is the variance of observed values; ρ is the correlation coefficient between predicted and observed samples; and n is the number of samples.

We also calculated null model values (NULL_RMSE and NULL_MAE). The null model involves using the average value determined by the collected samples (Equations 6 and 7). It serves as the simplest model possible in the absence of specific training data, providing a single mean for numeric outcomes. The null model's reference point is the percentage of samples with the most prevalent class when class probabilities are requested. Models comparable to or worse than the null model should be disregarded. The best models exhibited lower RMSE and MAE results than those observed for NULL_MAE and NULL_RMSE. This indicates the final model surpasses the use of mean values, highlighting improved model quality.

If the NULL_RMSE and NULL_MAE values are lower than those seen during the algorithm's validation phase, it indicates that using the means of the property samples aligns with the machine learning model created by the algorithms. The calculations for NULL_RMSE and NULL_MAE were performed using the "null mode" function in the caret package (Kuhn et al., 2020, Mello et al., 2021).

$$NULL_RMSE = \left[\frac{1}{N} \sum_{i=1}^N (Om_t - O_i)^2 \right]^{\frac{1}{2}} \quad \text{Eq. 6}$$

$$NULL_MAE = \frac{\sum_{i=1}^N |Om_t - O_i|}{n} \quad \text{Eq. 7}$$

in which: Om_t is the mean of training samples; O_i is the test sample; n is the number of samples (loop).

Final maps of SOC and coefficient of variation

Of the five evaluated algorithms used to create spatial prediction maps of SOC, we selected the best-performing model based on the mean MAE value from the 81 runs (Gomes et al., 2019; Reis et al., 2021; Paes et al., 2022). Based on the 81 predicted SOC maps, the mean and standard deviation of each pixel's prediction were calculated within the study area to generate mean and standard deviation maps, respectively. Each pixel had a resolution of 30 m, and the maps were used for the discussion of the results. The

importance of covariates was also calculated from the results of the 81 runs, for which the mean covariate importance for each model was computed (Gomes et al., 2019).

The algorithm that performed the best (CUB) was used to generate maps predicting the SOC content in the study area. From the final map of SOC of the best model, the mean SOC values were determined for each geoenvironment, calculated from the mean value of all cells inserted in each geo-environment using the *zonal statistics* plugin of the QGIS package (<https://qgis.org/en/site/>).

RESULTS

SOC, selected covariates and performance of predictive models

The SOC contents in the topsoil (0.00-0.20 m) of the studied area ranged from 0.2 to 12 dag kg⁻¹, with a coefficient of variation (CV) of 77.6 % and skewness of 1.55 (Table 2). The evaluated models could predict the SOC content determined by the Walkley-Black method. The analysis revealed, on average, about 12 covariates produced performance comparable to using 30 covariates (Figure 5). On average, around 12 covariates yielded comparable performance to utilizing 30 covariates (Figure 5).

Table 2. Descriptive statistics of soil organic carbon (SOC) in the study area (n = 81)

Min	Max	Mean	SD	Median	CV	Skewness
dag kg ⁻¹					%	
0.2	12.0	3.1	2.4	2.5	77.6	1.55

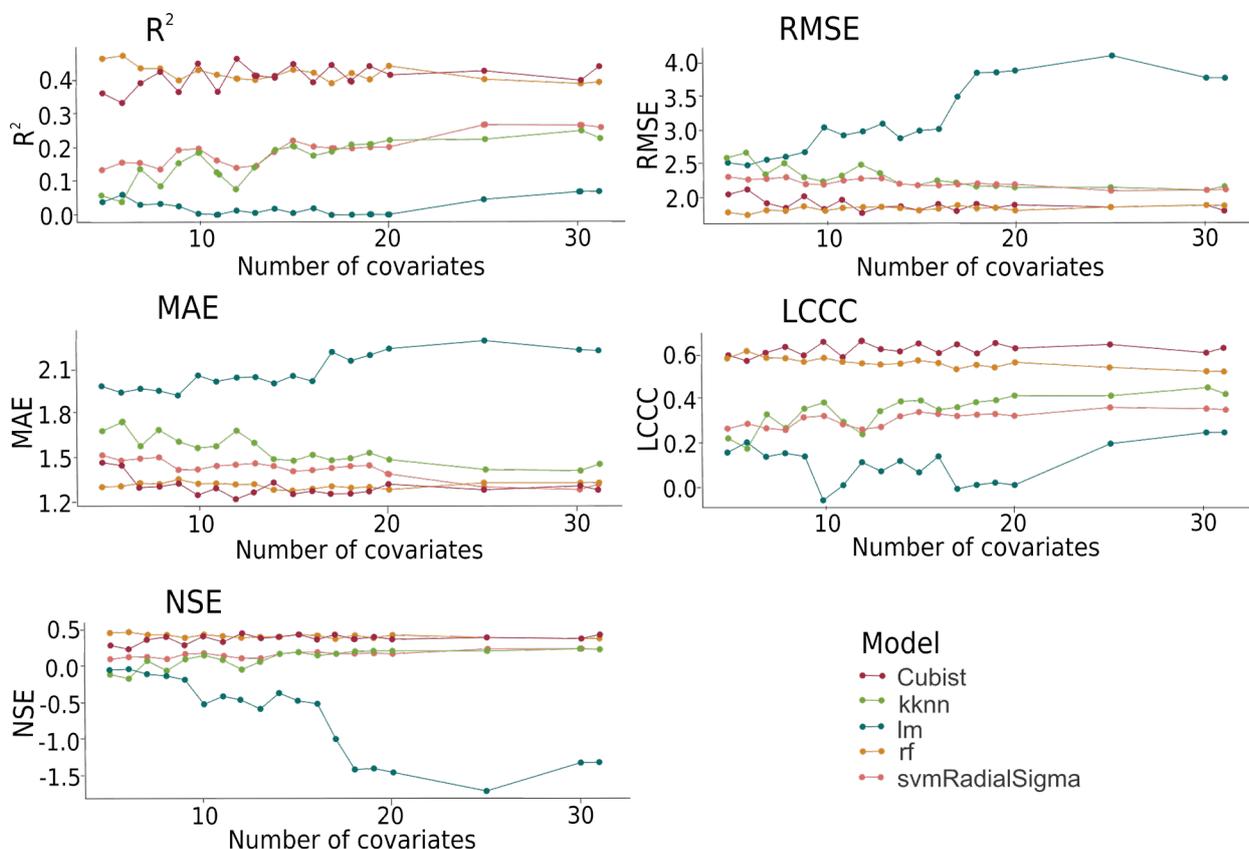


Figure 5. Performance of models in the process of selecting covariates for soil organic carbon (SOC), using Recursive Feature Elimination (RFE) methodology.

The CUB model demonstrated superior performance, with MAE and RMSE values lower than the null model (Null_Model) (Table 3). This suggests the CUB model surpasses the use of mean values. Additionally, the CUB algorithm achieved an R^2 of 0.483 for training and 0.505 for testing. The second-best performing model was Random Forest (RF), which also exhibited low MAE and RMSE values, though slightly higher than CUB. RF achieved an R^2 of 0.49 for training and 0.375 for testing (Table 3 and Figure 6). When comparing the validation R^2 of RF to CUB, there was only a 25.7 % decrease in performance. In contrast, the remaining algorithms—Lm, KKNN, and SVM—demonstrated moderate to poor performance (Table 3).

The machine learning algorithms selected covariates to predict SOC. Notably, the most crucial covariates for prediction were identified as elevation (MDE) and slope_height, both intricately linked to topography, each with an importance rating of 100 %. Additionally, Terrain surface convexity (TSC) played a significant role, with an importance of 90 %, followed by hydro_eucl (50 %), MRVBF (45 %), and Y (40 %), albeit with relatively lesser importance. In contrast, other covariables, such as MSP, VD, NH, MRRTF, and TWI had weaker impacts, while VRM proved negligible in predicting soil SOC (Figure 7).

Table 3. Performance of Cubist (CUB), Linear Model (Lm), Random Forests (RF), K Nearest Neighbor (KKNN), and Support Vector Machines (SVM) models in predicting carbon content, evaluated by mean absolute error (MAE), Lin's Concordance Correlation (LCCC), Nash-Sutcliffe Efficiency (NSE), root mean squared error (RMSE), and coefficient of determination (R^2) of the training and testing datasets

Performance metrics		Carbon	
		Training	Test
		dag kg ⁻¹	
MAE	CUB	1.217	1.218
	Lm	1.735	1.986
	KKNN	1.455	1.556
	SVM	1.292	1.387
	RF	1.245	1.362
	Null_Model	-	1.828
RMSE	CUB	1.803	1.745
	Lm	2.724	3.018
	KKNN	2.183	2.33
	SVM	2.124	2.184
	RF	1.752	1.932
	Null_Model	-	1.932
R2	CUB	0.483	0.505
	Lm	0.103	0.035
	KKNN	0.226	0.169
	SVM	0.278	0.208
	RF	0.49	0.375
	Null_Model	-	0.000
NSE	CUB	0.679	0.499
	Lm	0.24	-0.524
	KKNN	0.427	0.091
	SVM	0.376	0.193
	RF	0.633	0.376
	Null_Model	-	-
LCCC	CUB	0.679	0.691
	Lm	0.24	0.187
	KKNN	0.427	0.382
	SVM	0.376	0.343
	RF	0.633	0.547
	Null_Model	-	-

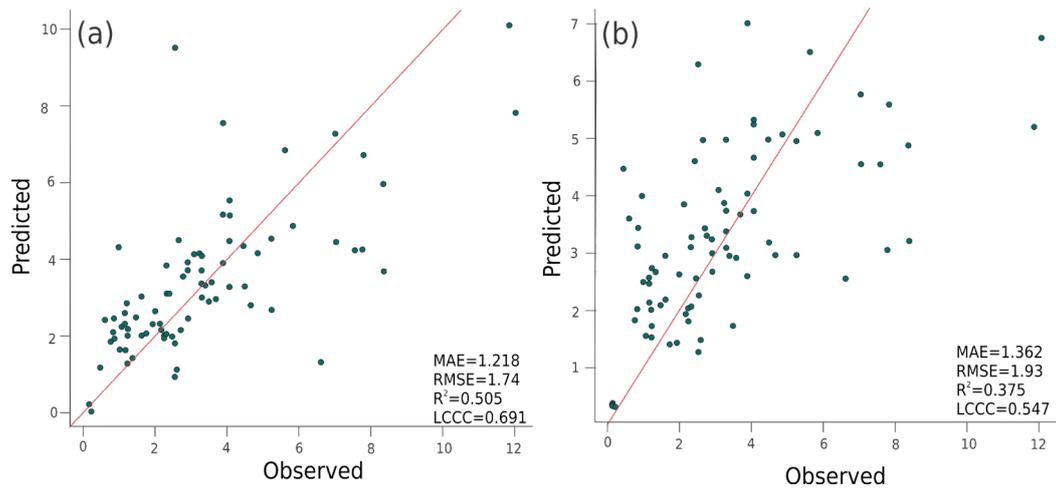


Figure 6. Predicted values *versus* laboratory-determined values for soil organic carbon contents. Predicted value *versus* observed values of soil organic carbon using Cubist (a) and Random Forest (b) algorithms.

Prediction of final maps and uncertainty of SOC

The highest SOC contents were found in the geoenvironments of dense alluvial rainforest with palms on Plinthosols and Gleysols, followed by mangrove with Gleysols, coastal muddy plain, western Marajó plateau with dense alluvial rainforest, and old salt-plain terrace with fluvial-lacustrine sediments with high concentration of archaeological mounds (Figures 8 and 9). In these geoenvironments, the average surface SOC contents varied from 3.76 to 5.44 dag kg⁻¹. The geoenvironments with intermediate SOC contents include active sandy dunes with marine and fluvial plains, plateau with Oxisols and secondary vegetation, grassy fluvio-lacustrine plains with Melanic or Haplic Gleysols, hills, and low tablelands on fluvial-lacustrine sediments with Plinthosols under alluvial rainforest, terraces with hydromorphic Plinthosols on fluvial-lacustrine sediments, and central inland paleoestuary under grassy fields with Gleysols. In these areas, SOC contents varied from 3.09 to 3.66 dag kg⁻¹ (Figures 8 and 9). The lowest SOC contents corresponded to the geoenvironments of sandy flats with Spodosols under savanna, low terraces and plains with saline-sodic Gleysols, and salt-plain with paleodunes. SOC contents in these environments ranged from 2.49 to 2.55 dag kg⁻¹.

The mean CV of the SOC predictions for the geoenvironments ranged from 19.51 to 30.50 % (Figure 8). The greatest uncertainties, as reflected by the highest CV, corresponded to the geoenvironments of hills and low tablelands on fluvial-lacustrine sediments with Plinthosols under alluvial rainforest and sandy flats with Spodosols under savanna.

DISCUSSION

Model performance and uncertainty

The methodological framework employed in this study optimized the prediction of SOC contents through digital soil mapping methodologies for covariate and model selection, coupled with an assessment of prediction uncertainty. Notably, the application of Recursive Feature Elimination (RFE) did not enhance model performance, aligning with findings from previous studies (Stevens et al., 2013; Gomes et al., 2019).

This framework facilitated the identification of the best-performing model based on a subset of covariates with higher confidence. These covariates were initially selected by their respective models using RFE. Furthermore, the framework provided a measure of prediction uncertainty, a valuable feature in DSM for modeling SOC contents. As a result, this methodological approach enables the creation of maps with known performance and is well-suited for mapping SOC contents in areas characterized by high environmental diversity.

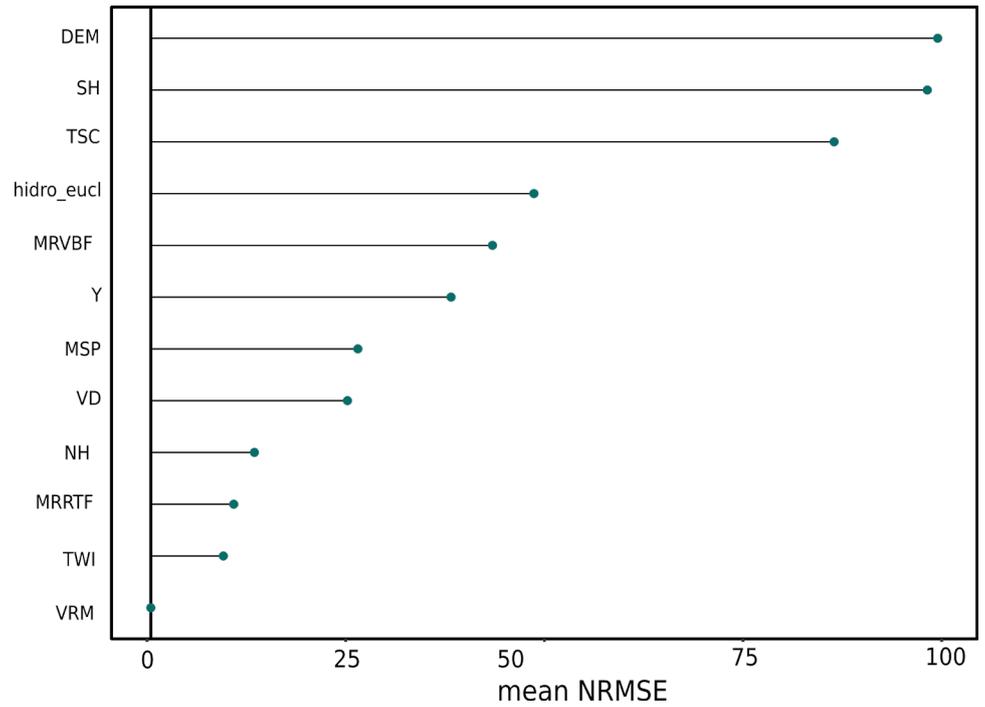


Figure 7. Importance of covariates on predicting soil organic carbon (SOC) contents using the best-performing algorithm.

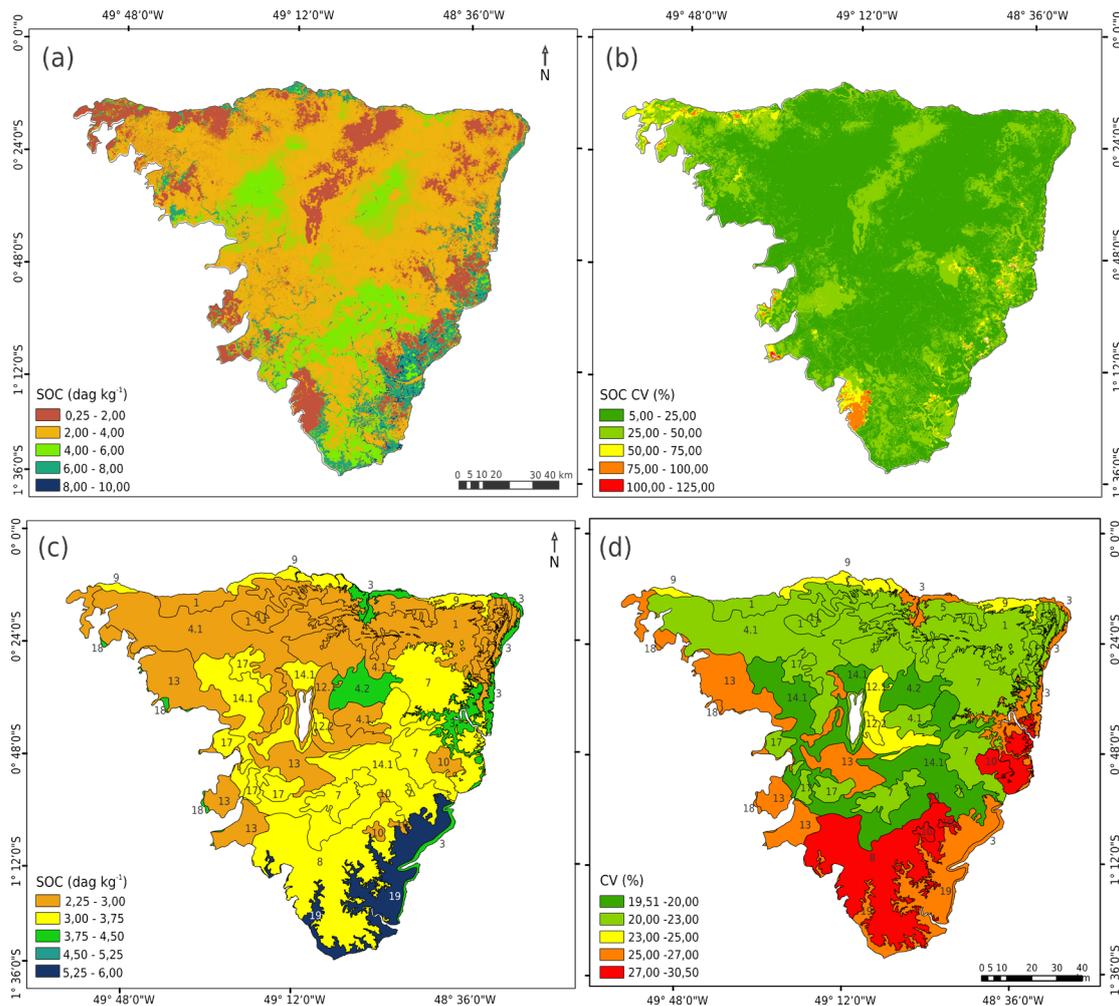


Figure 8. Soil carbon contents obtained from 81 runs using the best-performing algorithm (a); coefficient of variation - CV (%) of the 81 predictions with the Cubist model (b); mean organic carbon contents of soils per geoenvironment (c); and mean coefficient of variation - CV (%) per geoenvironment (d).

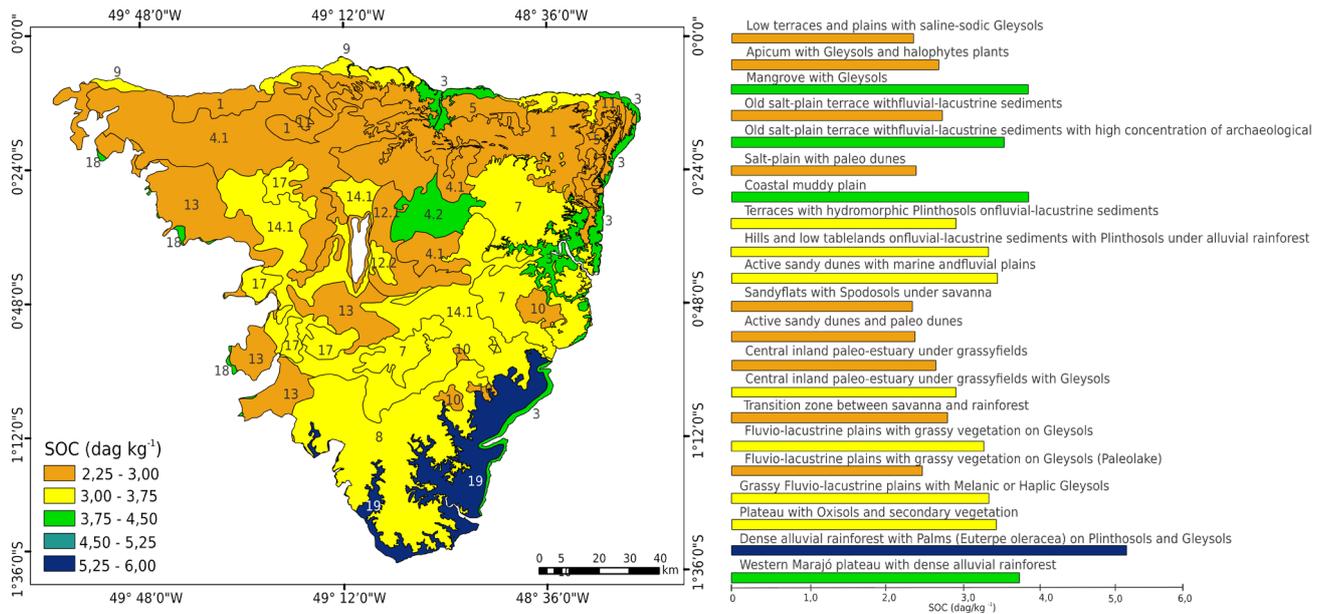


Figure 9. Mean soil carbon contents per geoenvironment, obtained by averaging the cells within each geoenvironment.

Our findings validate the efficacy of the Cubist model as a robust machine-learning algorithm for predicting SOC, aligning with observations by Goydaragh et al. (2021) and Zeraatpisheh et al. (2022). The prediction statistics indicate overfitting is not a concern, as evidenced by the similarity in R^2 , NSE, and LCCC values between the training and test sets (Table 3). This conclusion is further supported by the marginal differences observed in RMSE and MAE between the training and validation sets. Notably, the R^2 variation was less than 5% when predicting SOC contents in the training set. The slightly greater variation in the test sets suggests an increase in prediction uncertainty, possibly attributed to the smaller number of samples utilized for cross-validation compared to training or to the internal data selection process of the Cubist model.

The scarcity of soil data emerges as a notable source of error in DSM (Malone et al., 2018). In this context, we attribute the limited representativeness of soil samples as a plausible reason for the elevated prediction uncertainty observed in specific geoenvironments. Furthermore, the natural variability of the study area, influenced by climate changes at the end of the Pleistocene and the beginning of the Holocene, neotectonic events (Rossetti and Valeriano, 2007; Rossetti et al., 2012), and relative sea level variations (Cohen et al., 2005), has transformed the region into a complex pedogeomorphological landscape. These factors may have contributed to uncertainties in the mapping process.

Distribution of SOC concentration in the geoenvironments of the eastern Marajó island

In the geoenvironments of dense alluvial tropical forests, primary productivity is high, favoring SOC accumulation through litter deposition, which tends to be cycled by microorganisms even during the wetter periods of higher decomposition rates, remaining in the soil (Sanches et al., 2008; Ritter et al., 2018; Sakai and Kitajima, 2019). Even with the partial suppression of vegetation by the local population for livestock, subsistence agriculture, and wood extraction (Lisboa, 2012), these soils maintain SOC concentrations for some time, as pre-existing litter decomposes (Barros and Fearnside, 2016). These areas are predominantly composed of loamy, aluminous, kaolinitic soils, poor in iron (Rêgo, 1986; Horbe and Costa, 1997). Additionally, dense alluvial rainforests with palms on Plinthosols and Gleysols are commonly affected by seasonal flooding (Japiassú and Filho, 1974; Rossetti and Valeriano, 2007). In the Amazon biome, seasonally flooded forests have higher SOC concentrations on average when compared to non-flooded forests due to slower decomposition rates (Barbosa et al., 2017).

Located in floodplains, the old salt-plain terrace with fluvial-lacustrine sediments with high concentration of archaeological sites has a marked presence of paleochannels (Rossetti et al., 2009; Francisquini et al., 2014), with arboreal-shrub vegetation aligned with alluvial tropical forests (Japiassú and Filho, 1974). With naturally high SOC content, many of these areas were chosen as human settlements by pre-Columbian communities, both for their better drainage and proximity to water resources (Lisboa, 2012; Meggers and Roosevelt, 1992). These occupations increased the SOC through the deposition of bones and feces (Meggers and Evans, 1957).

Mangroves and coastal muddy soils are widely distributed in tropical and subtropical regions of the world, with high productivity and rapid deposition rates (Cui et al., 2021). In these geoenvironments, low oxygen concentrations result in a slow rate of SOC mineralization, favoring its long-term accumulation and stability (Sahrawat, 2003). In mangroves in the southeast and south areas, with greater fluvial influence, non-saline Gleysols predominate, while in the northeast and north areas, with greater marine influence, saline and sodic Gleysols predominate (Henriques et al., 2022).

The sandy plains with Spodosols under savanna (Cerrado) naturally have a low topsoil SOC content, either due to the natural occurrence of fire (Gomes et al., 2018), sandy texture low cationic adsorption capacity (Yost and Hartemink., 2019) or due to greater evapotranspiration and sparser vegetation in these areas (Henriques et al., 2022). The saline geoenvironments, Low Terraces and plains with saline-sodic Gleysols, Apicum with Gleysols and halophytes plants, and Salt-plain with paleo dunes also had low SOC content. Such environments have higher evapotranspiration and lower precipitation (Lisboa, 2012), resulting in soils affected by salts, such as Salic Gleysols (Henriques et al., 2022). Salinity and sodicity reduce primary productivity (Katerji et al., 2003), increase the potential for SOC decomposition (Setia et al., 2012), and the propensity for erosion (Wong et al., 2010).

Climate vulnerability of SOC in the geoenvironments of Marajó island

Impacts of climate change on SOC have been observed in various studies in the wetlands (Fitzgerald et al., 2008; Valiela et al., 2018; Alaniz et al., 2022; Hidalgo-Corrotea et al., 2023), revealing intensification of coastal erosion and habitat loss, culminating in reduced primary productivity and consequent decrease in soil carbon absorption.

Soils beneath dense alluvial tropical forest geoenvironments, which have the highest levels of SOC in the study area, may also be subject to these changes. Even though increased CO₂ may enhance primary productivity (Ainsworth and Long, 2005), intensifying dry and wet seasons would lead to increased tree mortality (Allen et al., 2010; Adams et al., 2017). Moreover, the global temperature increase of 1.5 °C predicted by IPCC (Stocker et al., 2014) would lead to a net increase in SOC release into the atmosphere via decomposition (Trumbore et al., 1996; Davidson and Janssens, 2006), as microbial activity tends to increase with temperature (He et al., 2010).

Supported by lateritic crusts, already in the process of dissolution under the humid climate of the island (Horbe and Costa, 1997), these forested areas will be subject to sea-level rise (Barros and Albernaz, 2014), which tends to undercut the base of lateritic slopes, causing the dissolution of Fe and Al oxyhydroxides and resulting in the dismantling of the relief (Henriques et al., 2022). The mortality of economically valuable tree species for local populations (Açaí, cupuaçu, murici), combined with the loss of suitable areas for the regeneration of these species, may lead to the disappearance of these plants, representing a burden for several extractive families that depend on these natural resources for survival (Lisboa, 2012; Evangelista-Vale et al., 2021).

Climate projections for the island suggest, in the coming decades, a significant portion of Marajó's wetlands may face permanent flooding (Barros and Albernaz, 2014). This anticipated inundation, particularly in geoenvironments with fluvial-lacustrine influence, currently sustained by freshwater (Santos et al., 2008; Rosário et al., 2009), could adversely impact soil microorganisms. The entry of saline water into these areas might directly harm soil microorganisms, leading to increased osmotic stresses and specific ionic toxicity (Wong et al., 2010). Additionally, it could indirectly affect microbial activity, diminishing SOC content and plant residue input (Singh, 2016). Moreover, the permanent flooding of these regions would render buffalo farming economically unviable and reduce the available areas for rice cultivation, which are crucial economic activities on Marajó island (Lisboa, 2012).

Considering the projected rise in sea levels, Marajó island may revert to conditions reminiscent of the mid-Holocene, when sea level rise facilitated the establishment of mangroves in the central regions of the floodplains (França et al., 2012; Smith et al., 2012). In such conditions, and as biological communities adapt (Singh, 2016), the long-term accumulation of SOC may be favored (Rogers et al., 2019; Watanabe et al., 2019). This is because the area of carbon deposition under anoxic conditions would likely increase (Sahrawat, 2003).

CONCLUSIONS

Machine learning algorithms have demonstrated their effectiveness in mapping SOC content and its associated uncertainty in topsoil, utilizing readily available covariates. The methodological framework outlined in this study optimizes the prediction process without compromising performance. It achieves this by selecting only the most crucial covariates and the best-fitting model, enabling the prediction of SOC content alongside its corresponding uncertainty. This approach opens avenues for conducting long-term monitoring studies of SOC, particularly in remote and challenging-to-access regions characterized by significant geoenvironmental diversity. Furthermore, the spatial visualization of mapping uncertainties holds potential for future research and mapping endeavors. It can be employed to identify priority areas for sample collection, enhancing the efficiency of soil data gathering and contributing to more accurate and reliable SOC mapping efforts.

The most influential covariates that shape the distribution of SOC stocks in the wetlands of Marajó island include the elevation of each cell, the distance between the base and the summit of the slope, the convexity of the terrain surface, and the proximity to water bodies. The Cubist model outperformed other machine learning methods in predicting surface SOC content in the wetlands of Marajó island, achieving higher accuracy.

Global climate change may directly impact SOC in the geoenvironments of the wetland areas of Marajó island and its local population, which traditionally relies on land use-related activities. However, surface SOC content alone does not fully reflect the soil's actual capacity to store carbon. Modeling studies of carbon stocks in the diverse geoenvironments of the wetland areas of Marajó island are deemed necessary for a comprehensive understanding.

DATA AVAILABILITY

The data will be provided upon request.

ACKNOWLEDGMENTS

We would like to thank the field logistics support provided by the Emilio Goeldi Museum (MPEG), the Federal University of Pará.

FUNDING

The CNPq (MCTIC/CNPq - 21/2017) supported this study, and CAPES granted the research fellowship.

AUTHOR CONTRIBUTIONS

Conceptualization:  Carlos Ernesto Schaefer (equal),  Danilo de Lima Camêlo (equal),  David Lukas de Arruda (equal),  Gustavo Vieira Veloso (equal) and  Lucas de Carvalho Gomes (equal).

Data curation:  David Lukas de Arruda (equal),  Gustavo Vieira Veloso (equal),  Lucas de Carvalho Gomes (equal) and  Renata Jordan Henriques (equal).

Formal analysis:  David Lukas de Arruda (equal) and  Gustavo Vieira Veloso (equal).

Funding acquisition:  Carlos Ernesto Schaefer (lead).

Investigation:  Carlos Ernesto Schaefer (equal),  David Lukas de Arruda (equal) and  Gustavo Vieira Veloso (equal).

Methodology:  Gustavo Vieira Veloso (equal),  Lucas de Carvalho Gomes (equal) and  Renata Jordan Henriques (equal).

Project administration:  Carlos Ernesto Schaefer (equal) and  Lucas de Carvalho Gomes (equal).

Software:  Gustavo Vieira Veloso (lead).

Supervision:  Carlos Ernesto Schaefer (equal),  Elpídio Inácio Fernandes-Filho (equal),  João Carlos Ker (equal),  Gustavo Vieira Veloso (equal) and  Lucas de Carvalho Gomes (equal).

Validation:  Elpídio Inácio Fernandes-Filho (equal),  Gustavo Vieira Veloso (equal) and  Lucas de Carvalho Gomes (equal).

Visualization:  Danilo de Lima Camêlo (equal) and  Lucas de Carvalho Gomes (equal).

Writing - original draft:  Carlos Ernesto Schaefer (equal),  Danilo de Lima Camêlo (equal),  David Lukas de Arruda (equal),  Elpídio Inácio Fernandes-Filho (equal),  Gustavo Vieira Veloso (equal),  João Carlos Ker (equal),  Lucas de Carvalho Gomes (equal) and  Renata Jordan Henriques (equal).

Writing - review & editing:  Carlos Ernesto Schaefer (equal),  Danilo de Lima Camêlo (equal),  David Lukas de Arruda (equal),  Elpídio Inácio Fernandes-Filho (equal),  Gustavo Vieira Veloso (equal),  João Carlos Ker (equal),  Lucas de Carvalho Gomes (equal) and  Renata Jordan Henriques (equal).

REFERENCES

Adame MF, Hasan S, Buelow CA. Tropical intertidal wetlands are hotspots of carbon storage and nutrient transformations. In: Baird D, Elliott M, editors. *Treatise on estuarine and coastal science*. Amsterdam: Academic Press; 2024. v. 3. p. 508-18. <https://doi.org/10.1016/B978-0-323-90798-9.00036-6>

Adams HD, Zeppel MJB, Anderegg WRL, Hartmann H, Landhäusser SM, Tissue DT, Huxman TE, Hudson PJ, Franz TE, Allen CD, Anderegg LDL, Barron-Gafford GA, Beerling DJ, Breshears DD, Brodrick TJ, Bugmann H, Cobb RC, Collins AD, Dickman LT, Duan H, Ewers BE, Galiano L, Galvez DA, Garcia-Forner N, Gaylord ML, Germino MJ, Gessler A, Hacke UG, Hakamada R, Hector A,

- Jenkins MW, Kane JM, Kolb TE, Law DJ, Lewis JD, Limousin JM, Love DM, Macalady AK, Martínez-Vilalta J, Mencuccini M, Mitchell PJ, Muss JD, O'Brien MJ, O'Grady AP, Pangle RE, Pinkard EA, Piper FI, Plaut JA, Pockman WT, Quirk J, Reinhardt K, Ripullone F, Ryan MG, Sala A, Sevanto S, Sperry JS, Vargas R, Vennetier M, Way DA, Xu C, Yopez EA, McDowell NG. A multi-species synthesis of physiological mechanisms in drought-induced tree mortality. *Nat Ecol Evol.* 2017;1:1285-91. <https://doi.org/10.1038/S41559-017-0248-X>
- Ainsworth EA, Long SP. What have we learned from 15 years of free-air CO₂ enrichment (FACE)? A meta-analytic review of the responses of photosynthesis, canopy properties and plant production to rising CO₂. *New Phytologist.* 2005;165:351-72. <https://doi.org/10.1111/J.1469-8137.2004.01224.X>
- Alaniz AJ, Carvajal MA, Marquet PA, Vergara PM, Meneses L, Moreira-Arce D. Analyzing the spatiotemporal patterns of forests carbon sink and sources between 2000 and 2019. 2022;10:e2021EF002560. <https://doi.org/10.1029/2021EF002560>
- Allen CD, Macalady AK, Chenchouni H, Bachelet D, McDowell N, Vennetier M, Kitzberger T, Rigling A, Breshears DD, Hogg EH (Ted), Gonzalez P, Fensham R, Zhang Z, Castro J, Demidova N, Lim J-H, Allard G, Running SW, Semerci A, Cobb N. A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. *For Ecol Manage.* 2010;259:660-84. <https://doi.org/10.1016/j.FORECO.2009.09.001>
- Alvares CA, Stape JL, Sentelhas PC, Gonçalves JLM, Sparovek G. Köppen's climate classification map for Brazil. *Meteorol Z.* 2013;22:711-28. <https://doi.org/10.1127/0941-2948/2013/0507>
- Barbosa RI, Castilho CV, Perdiz RO, Damasco G, Rodrigues R, Fearnside PM. Decomposition rates of coarse woody debris in undisturbed Amazonian seasonally flooded and unflooded forests in the Rio Negro-Rio Branco Basin in Roraima, Brazil. *For Ecol Manage.* 2017;397:1-9. <https://doi.org/10.1016/j.FORECO.2017.04.026>
- Barros D, Albernaz A. Possible impacts of climate change on wetlands and its biota in the Brazilian Amazon. *Braz J Biol.* 2014;74:810-20. <https://doi.org/10.1590/1519-6984.04013>
- Barros HS, Fearnside PM. Soil carbon stock changes due to edge effects in central Amazon forest fragments. *For Ecol Manage.* 2016;379:30-6. <https://doi.org/10.1016/j.FORECO.2016.08.002>
- Bivand R, Krug R, Neteler M, Jeworutzki S. Rgrass7: Interface between GRASS 7 geographical information system and R. R package version 0.2-1. *GrassGis*; 2019. Available from: <https://r-forge.r-project.org/projects/spgrass/>.
- Brenning A. Statistical geocomputing combining R and SAGA: The example of landslide susceptibility analysis with generalized additive models. In: Böhner J, Blaschke T, Montanarella L, editors. *SAGA--Seconds Out (Hamburger Beiträge Zur Physischen Geographie Und Landschaftsökologie)*. 2008. v. 19. p. 23-32.
- Cerri CC, Volkoff B. Matéria orgânica de três solos dos campos inundáveis da Ilha de Marajó (PA). *Rev Bras Cienc Solo.* 1988;12:93-100.
- Chen S, Arrouays D, Mulder VL, Poggio L, Minasny B, Roudier P, Libohova Z, Lagacherie P, Shi Z, Hannam J, Meersmans J, Richer-de-Forges AC, Walter C. Digital mapping of GlobalSoilMap soil properties at a broad scale: A review. *Geoderma.* 2022;409:115567. <https://doi.org/10.1016/j.GEODERMA.2021.115567>
- Chen S, Zou J, Hu Z, Lu Y. Temporal and spatial variations in the mean residence time of soil organic carbon and their relationship with climatic, soil and vegetation drivers. *Global Planet Change.* 2020;195:103359. <https://doi.org/10.1016/j.GLOPLACHA.2020.103359>
- Chi Y, Liu D, Xie Z. Zonal simulations for soil organic carbon mapping in coastal wetlands. *Ecol Indic.* 2021;132:108291. <https://doi.org/10.1016/j.ECOLIND.2021.108291>
- Cohen MCL, Behling H, Lara RJ, Smith CB, Matos HRS, Vedel V. Impact of sea-level and climatic changes on the Amazon coastal wetlands during the late Holocene. *Veg Hist Archaeobot.* 2009;18:425-39. <https://doi.org/10.1007/s00334-008-0208-0>
- Cohen MCL, Souza Filho PWM, Lara RJ, Behling H, Angulo RJ. A model of Holocene mangrove development and relative sea-level changes on the Bragança Peninsula (northern Brazil). *Wetl Ecol Manag.* 2005;13:433-43. <https://doi.org/10.1007/s11273-004-0413-2>

- Correa PRS, Peres RN, Vieira LS. Levantamento exploratorio de solos da Folha SA.22 Belém. In: Brasil. Projeto Radam Folha SA.22 Belém; geologia, geomorfologia, solos, vegetação e uso potencial da terra. Rio de Janeiro: Departamento Nacional de Produção Mineral; 1974. p. 132-290. Available from: <https://biblioteca.ibge.gov.br/biblioteca-catalogo?id=224022&view=detalhes>.
- Cui L, Sun H, Du X, Feng W, Wang Y, Zhang J, Jiang J. Dynamics of labile soil organic carbon during the development of mangrove and salt marsh ecosystems. *Ecol Indic.* 2021;129:107875. <https://doi.org/10.1016/j.ECOLIND.2021.107875>
- Davidson EA, Janssens IA. Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. *Nature.* 2006;440:165-73. <https://doi.org/10.1038/nature04514>
- Doetterl S, Stevens A, Six J, Merckx R, van Oost K, Pinto MC, Casanova-Katny A, Muñoz C, Boudin M, Venegas EZ, Boeckx P. Soil carbon storage controlled by interactions between geochemistry and climate. *Nat Geosci.* 2015;8:780-3. <https://doi.org/10.1038/ngeo2516>
- Duarte CM, Losada IJ, Hendriks IE, Mazarrasa I, Marbà N. The role of coastal plant communities for climate change mitigation and adaptation. *Nat Clim Change.* 2013;3:961-8. <https://doi.org/10.1038/nclimate1970>
- Evangelista-Vale JC, Weihs M, José-Silva L, Arruda R, Sander NL, Gomides SC, Machado TM, Pires-Oliveira JC, Barros-Rosa L, Castuera-Oliveira L, Matias RAM, Martins-Oliveira AT, Bernardo CSS, Silva-Pereira I, Carnicer C, Carpanedo RS, Eisenlohr P V. Climate change may affect the future of extractivism in the Brazilian Amazon. *Biol Conserv.* 2021;257:109093. <https://doi.org/10.1016/j.BIOCON.2021.109093>
- Fernandes MM, Fernandes MRM, Garcia JR, Matricardi EAT, Lima AHS, Filho RNA, Filho RRG, Piscoya VC, Piscoya TOF, Cunha Filho M. Land use and land cover changes and carbon stock valuation in the São Francisco river basin, Brazil. *Environ Chall.* 2021;5:100247. <https://doi.org/10.1016/j.ENVC.2021.100247>
- Ferreira RG, Silva DD, Elesbon AAA, Santos GR, Veloso GV, Fraga MS, Fernandes Filho EI. Geostatistical modeling and traditional approaches for streamflow regionalization in a Brazilian Southeast watershed. *J South Am Earth Sci.* 2021;108:103355. <https://doi.org/10.1016/j.JSAMES.2021.103355>
- FitzGerald DM, Fenster MS, Argow BA, Buynevich I V. Coastal impacts due to sea-level rise. *Annu Rev Earth Planet Sci.* 2008;36:601-47. <https://doi.org/10.1146/annurev.earth.35.031306.140139>
- França MC, Francisquini MI, Cohen MCL, Pessenda LCR, Rossetti DF, Guimarães JTF, Smith CB. The last mangroves of Marajó Island - Eastern Amazon: Impact of climate and/or relative sea-level changes. *Rev Palaeobot Palynol.* 2012;187:50-65. <https://doi.org/10.1016/j.revpalbo.2012.08.007>
- Francisquini MI, Lima CM, Pessenda LCR, Rossetti DF, França MC, Cohen MCL. Relation between carbon isotopes of plants and soils on Marajó Island, a large tropical island: Implications for interpretation of modern and past vegetation dynamics in the Amazon region. *Palaeogeogr Palaeoclimatol Palaeoecol.* 2014;415:91-104. <https://doi.org/10.1016/j.PALAEO.2014.03.032>
- Gomes L, Miranda HS, Bustamante MMC. How can we advance the knowledge on the behavior and effects of fire in the Cerrado biome? *For Ecol Manage.* 2018;417:281-90. <https://doi.org/10.1016/j.FORECO.2018.02.032>
- Gomes LC, Faria RM, Souza E, Veloso GV, Schaefer CEGR, Fernandes Filho EI. Modelling and mapping soil organic carbon stocks in Brazil. *Geoderma.* 2019;340:337-50. <https://doi.org/10.1016/j.GEODERMA.2019.01.007>
- Goydaragh MG, Taghizadeh-Mehrjardi R, Jafarzadeh AA, Triantafyllis J, Lado M. Using environmental variables and Fourier Transform Infrared Spectroscopy to predict soil organic carbon. *Catena.* 2021;202:105280. <https://doi.org/10.1016/j.CATENA.2021.105280>
- He X, Lin Y, Han G, Guo P, Tian X. The effect of temperature on decomposition of leaf litter from two tropical forests by a microcosm experiment. *Eur J Soil Biol.* 2010;46:200-7. <https://doi.org/10.1016/j.EJSOBI.2010.02.001>
- Hein CJ, Usman M, Eglinton TI, Haghypour N, Galy VV. Millennial-scale hydroclimate control of tropical soil carbon storage. *Nature.* 2020;581:63-6. <https://doi.org/10.1038/s41586-020-2233-9>

- Henriques RJ, Oliveira FS, Schaefer CEGR, Francelino MR, Lopes PRC, Senra EO, Lourenço VR. Soils and landscapes of Marajó island, Brazilian Amazonia: Holocene evolution, geoarchaeology and climatic vulnerability. *Environ Earth Sci.* 2022;81:254. <https://doi.org/10.1007/S12665-022-10310-2/TABLES/4>
- Hidalgo-Corrotea C, Alaniz AJ, Vergara PM, Moreira-Arce D, Carvajal MA, Pacheco-Cancino P, Espinosa A. High vulnerability of coastal wetlands in Chile at multiple scales derived from climate change, urbanization, and exotic forest plantations. *Sci Total Environ.* 2023;903:166130. <https://doi.org/10.1016/J.SCITOTENV.2023.166130>
- Hidayatullah MF, Kamal M, Wicaksono P. Species-based aboveground mangrove carbon stock estimation using WorldView-2 image data. *Remote Sens Appl.* 2023;30:100959. <https://doi.org/10.1016/J.RSASE.2023.100959>
- Horbe AMC, Costa ML. Solos gerados a partir do intemperismo de crostas lateríticas sílico-ferruginosas. *Acta Amazon.* 1997;27:241-56. <https://doi.org/10.1590/1809-43921997274256>
- Hujoel IA, Murphree DH, van Dyke CT, Choung RS, Sharma A, Murray JA, Rubio-Tapia A. Machine learning in detection of undiagnosed celiac disease. *Clin Gastroenterol H.* 2018;16:1354-55. <https://doi.org/10.1016/J.CGH.2017.12.022>
- Instituto Brasileiro de Geografia e Estatística - IBGE. Censo demográfico 2010. Brasília, DF: IBGE; 2010 [cited 2022 Feb 4]. Available from: <https://www.ibge.gov.br/estatisticas/multidominio/cultura-recreacao-e-esporte/9662-censo-demografico-2010.html>.
- Japiassú A, Goés Filho L. As regiões Fitoecológicas, sua natureza e seus recursos econômicos. In: Brasil. Projeto Radam Folha SA.22 Belém; geologia, geomorfologia, solos, vegetação e uso potencial da terra. Rio de Janeiro: Departamento Nacional de Produção Mineral; 1974. p. 291-390 Available from: <https://biblioteca.ibge.gov.br/biblioteca-catalogo.html?view=detalhes&id=281391>.
- Katerji N, van Hoorn JW, Hamdy A, Mastrorilli M. Salinity effect on crop development and yield, analysis of salt tolerance according to several classification methods. *Agric Water Manag.* 2003;62:37-66. [https://doi.org/10.1016/S0378-3774\(03\)00005-2](https://doi.org/10.1016/S0378-3774(03)00005-2)
- Kern AN, Addison P, Oommen T, Salazar SE, Coffman RA. Machine learning based predictive modeling of debris flow probability following wildfire in the intermountain Western United States. *Math Geosci.* 2017;49:717-35. <https://doi.org/10.1007/S11004-017-9681-2>
- Kloiber SM, Macleod RD, Smith AJ, Knight JF, Huberty BJ. A semi-automated, multi-source data fusion update of a wetland inventory for east-central Minnesota, USA. 2015;35:335-48. <https://doi.org/10.1007/s13157-014-0621-3>
- Kuhn M. Building predictive models in R using the caret package. *J Stat Softw.* 2008;28:1-26. <https://doi.org/10.18637/JSS.V028.I05>
- Kuhn M, Johnson K. Applied predictive modeling. New York: Springer; 2013. <https://doi.org/10.1007/978-1-4614-6849-3>
- Kuhn M, Wing J, Weston S, Williams A, Keefer C, Engelhardt A, Kenkel B. classification and regression training. R package version 6.0-86. Cambridge: Astrophysics Source Code Library; 2020.
- Lembaid I, Moussadek R, Mrabet R, Bouhaouss A. Modeling soil organic carbon changes under alternative climatic scenarios and soil properties using DNDC model at a semi-arid mediterranean environment. *Climate.* 2022;10:23. <https://doi.org/10.3390/CLI10020023>
- Leonard PB, Baldwin RF, Homyack JA, Wigley TB. Remote detection of small wetlands in the Atlantic coastal plain of North America: Local relief models, ground validation, and high-throughput computing. *For Ecol Manage.* 2012;284:107-15. <https://doi.org/10.1016/j.foreco.2012.07.034>
- Lima AMM, Oliveira LL, Fontinhas RL, Lima RJS. Ilha do Marajó: Revisão histórica, hidroclimatologia, bacias hidrográficas e propostas de gestão. *Holos Environ.* 2005;5:65. <https://doi.org/10.14295/holos.v5i1.331>
- Lisboa PLB. A terra dos Aruã: Uma história ecológica do arquipélago do Marajó. Brasília, DF: Museu Paraense Emílio Goeldi; 2012.

- Liu Q, Yang D, Cao L, Anderson B. Assessment and prediction of carbon storage based on land use/land cover dynamics in the tropics: A case study of Hainan Island, China. *Land*. 2022;11:244. <https://doi.org/10.3390/LAND11020244>
- Land Processes Distributed Active Archive Center – LP DAAC. NASADEM_HGT v001 - NASADEM Merged DEM Global 1 arc second. Sioux Falls: LP DAAC; 2003 [cited 2024 Mar 22]. Available from: https://lpdaac.usgs.gov/products/nasadem_hgtv001/.
- Macreadie PI, Costa MDP, Atwood TB, Friess DA, Kelleway JJ, Kennedy H, Lovelock CE, Serrano O, Duarte CM. Blue carbon as a natural climate solution. *Nat Rev Earth Environ*. 2021;2:826-39. <https://doi.org/10.1038/s43017-021-00224-1>
- Malone BP, Odgers NP, Stockmann U, Minasny B, McBratney Alex B. Digital mapping of soil classes and continuous soil properties. In: McBratney A, Minasny B, Stockmann U, editors. *Pedometrics: Progress in soil science*. Cham: Springer; 2018. p. 373-413. https://doi.org/10.1007/978-3-319-63439-5_12
- Maxwell TL, Hengl T, Parente LL, Minarik R, Worthington TA, Bunting P, Smart LS, Spalding MD, Landis E. Global mangrove soil organic carbon stocks dataset at 30 m resolution for the year 2020 based on spatiotemporal predictive machine learning. *Data Brief*. 2023;50:109621. <https://doi.org/10.1016/j.DIB.2023.109621>
- McBratney AB, Santos MLM, Minasny B. On digital soil mapping. *Geoderma*. 2003;117:3-52. [https://doi.org/10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4)
- Meggers BJ, Evans C. Archeological investigations at the mouth of the Amazon. *Bur Am Ethnol Bull*. 1957;167:1-664. Available from: <https://repository.si.edu/handle/10088/15461>
- Meggers BJ, Roosevelt AC. Moundbuilders of the Amazon: Geophysical Archaeology on Marajo Island, Brazil. *J Field Archaeol*. 1992;19:399. <https://doi.org/10.2307/529927>
- Mello DC, Veloso GV, Lana MG, Mello FAO, Poppiel RR, Cabrero DRO, Di Raimo LADL, Schaefer CEGR, Fernandes Filho EI, Leite EP, Demattê JAM. A new methodological framework for geophysical sensor combinations associated with machine learning algorithms to understand soil attributes. *Geosci Model Dev*. 2022;15:1219-46. <https://doi.org/10.5194/gmd-15-1219-2022>
- Mitsch WJ, Gosselink JG. The value of wetlands: importance of scale and landscape setting. *Ecol Econ*. 2000;35:25-33. [https://doi.org/10.1016/S0921-8009\(00\)00165-8](https://doi.org/10.1016/S0921-8009(00)00165-8)
- Muñoz ASM, Alvis AIG, Martínez IFB. A random forest model to predict soil organic carbon storage in mangroves from Southern Colombian Pacific coast. *Estuar Coast Shelf Sci*. 2024;299:108674. <https://doi.org/10.1016/j.ECSS.2024.108674>
- Murrieta RSS, Dufour DL, Siqueira AD. Food consumption and subsistence in three Caboclo populations on Marajo Island, Amazonia, Brazil. *Hum Ecol*. 1999;27:455-75. <https://doi.org/10.1023/A:1018779624490>
- Nottingham AT, Meir P, Velasquez E, Turner BL. Soil carbon loss by experimental warming in a tropical forest. *Nature*. 2020;584:234-7. <https://doi.org/10.1038/S41586-020-2566-4>
- Padilha MCC, Vicente LE, Demattê JAM, Loebmann DGSW, Vicente AK, Salazar DFU, Guimarães CCB. Using Landsat and soil clay content to map soil organic carbon of Oxisols and Ultisols near São Paulo, Brazil. *Geoderma Reg*. 2020;21:e00253. <https://doi.org/10.1016/j.GEODRS.2020.E00253>
- Paes EC, Veloso GV, Fonseca AA, Fernandes Filho EI, Fontes MPF, Soares EMB. Predictive modeling of contents of potentially toxic elements using morphometric data, proximal sensing, and chemical and physical properties of soils under mining influence. *Sci Total Environ*. 2022;817:152972. <https://doi.org/10.1016/j.SCITOTENV.2022.152972>
- Rêgo RS. Caracterização e gênese de solos com plintita na Ilha de Marajó [thesis]. Seropédica: Universidade Federal Rural do Rio de Janeiro; 1986.
- Reis GB, Silva DD, Fernandes Filho EI, Moreira MC, Veloso GV, Fraga MS, Pinheiro SAR. Effect of environmental covariable selection in the hydrological modeling using machine learning models to predict daily streamflow. *J Environ Manage*. 2021;290:112625. <https://doi.org/10.1016/j.JENVMAN.2021.112625>

- Ritter CD, Zizka A, Roger F, Tuomisto H, Barnes C, Nilsson RH, Antonelli A. High-throughput metabarcoding reveals the effect of physicochemical soil properties on soil and litter biodiversity and community turnover across Amazonia. *PeerJ*. 2018;6:e5661. <https://doi.org/10.7717/PEERJ.5661/SUPP-10>
- Robert JH. Raster: geographic data analysis and modeling. R package version 2.8-19. CRAN – Package raster; 2019 [cited 2022 Feb 8]. Available from: <https://cran.r-project.org/web/packages/raster/index.html>
- Rogers K, Kelleway JJ, Saintilan N, Megonigal JP, Adams JB, Holmquist JR, Lu M, Schile-Beers L, Zawadzki A, Mazumder D, Woodroffe CD. Wetland carbon storage controlled by millennial-scale variation in relative sea-level rise. *Nature*. 2019;567:91-5. <https://doi.org/10.1038/s41586-019-0951-7>
- Rosário RP, Bezerra MO, Vinzón SB. Dynamics of the saline front in the northern channel of the Amazon river - influence of fluvial flow and tidal range (Brazil). *J Coast Res*. 2009;56:1414-8.
- Rossetti DF, Góes AM, Toledo PM. Archaeological mounds in Marajó Island in northern Brazil: A geological perspective integrating remote sensing and sedimentology. *Geoarchaeology*. 2009;24:22-41. <https://doi.org/10.1002/GEA.20250>
- Rossetti DF, Souza LSB, Prado R, Elis VR. Neotectonics in the northern equatorial Brazilian margin. *J South Am Earth Sci*. 2012;37:175-90. <https://doi.org/10.1016/j.jsames.2012.03.004>
- Rossetti DF, Valeriano MM. Evolution of the lowest Amazon basin modeled from the integration of geological and SRTM topographic data. *Catena*. 2007;70:253-65. <https://doi.org/10.1016/j.catena.2006.08.009>
- Rouse Jr JW, Haas RH, Deering DW, Schell JA, Harlan JC. Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation. Texas: Texas A&M University; 1973. Available from: <https://ntrs.nasa.gov/api/citations/19740022555/downloads/19740022555.pdf>
- Rudiyanto, Minasny B, Setiawan BI, Saptomo SK, McBratney AB. Open digital mapping as a cost-effective method for mapping peat thickness and assessing the carbon stock of tropical peatlands. *Geoderma*. 2018;313:25-40. <https://doi.org/10.1016/J.GEODERMA.2017.10.018>
- Sahrawat KL. Organic matter accumulation in submerged soils. *Adv Agron*. 2003;81:169-201. [https://doi.org/10.1016/S0065-2113\(03\)81004-0](https://doi.org/10.1016/S0065-2113(03)81004-0)
- Sakai S, Kitajima K. Tropical phenology: Recent advances and perspectives. *Ecol Res*. 2019;34:50-4. <https://doi.org/10.1111/1440-1703.1131>
- Sanches L, Valentini CMA, Pinto Júnior OB, Nogueira JS, Vourlitis GL, Biudes MS, Silva CJ, Paulino B, Lobo FA, Valentini MA, Pinto Júnior OB, Nogueira JS, Vourlitis GL, Biudes MS, Silva CJ, Bambi P, Lobo FA. Seasonal and interannual litter dynamics of a tropical semideciduous forest of the southern Amazon Basin, Brazil. *J Geophys Res Biogeo*. 2008;113:4007. <https://doi.org/10.1029/2007JG000593>
- Santos MLS, Medeiros C, Muniz K, Feitosa FAN, Schwamborn R, Macêdo SJ. Influence of the Amazon and Pará Rivers on water composition and phytoplankton biomass on the adjacent shelf. *J Coast Res*. 2008;24:585-93. <https://doi.org/10.2112/05-0538.1>
- Scharlemann JPW, Tanner EVJ, Hiederer R, Kapos V. Global soil carbon: Understanding and managing the largest terrestrial carbon pool. *Carbon Manag*. 2014;5:81-91. <https://doi.org/10.4155/CMT.13.77>
- Sena NC, Veloso GV, Lopes AO, Francelino MR, Fernandes Filho EI, Senra EO, Silva Filho LA, Condé VF, Silva DLA, Araújo RW. Soil sampling strategy in areas of difficult access using the cLHS method. *Geoderma Reg*. 2021;24:e00354. <https://doi.org/10.1016/J.GEODRS.2020.E00354>
- Setia R, Smith P, Marschner P, Gottschalk P, Baldock J, Verma V, Setia D, Smith J. Simulation of salinity effects on past, present, and future soil organic carbon stocks. *Environ Sci Technol*. 2012;46:1624-31. <https://doi.org/10.1021/ES2027345>
- Singh K. Microbial and enzyme activities of saline and sodic soils. *Land Degrad Dev*. 2016;27:706-18. <https://doi.org/10.1002/LDR.2385>
- Sjögersten S, de la Barrera-Bautista B, Brown C, Boyd D, Lopez-Rosas H, Hernández E, Monroy R, Rincón M, Vane C, Moss-Hayes V, Gallardo-Cruz JA, Infante-Mata D, Hoyos-Santillan J, Vidal

- Solórzano J, Peralta-Carreta C, Moreno-Casasola P. Coastal wetland ecosystems deliver large carbon stocks in tropical Mexico. *Geoderma*. 2021;403:115173. <https://doi.org/10.1016/j.geoderma.2021.115173>
- Smith CB, Cohen MCL, Pessenda LCR, França MC, Guimarães JTF. Holocene proxies of sedimentary organic matter and the evolution of Lake Arari-Amazon Region. *Catena*. 2012;90:26-38. <https://doi.org/10.1016/j.catena.2011.10.002>
- Stevens A, Nocita M, Tóth G, Montanarella L, van Wesemael B. Prediction of soil organic carbon at the European scale by visible and near infrared reflectance spectroscopy. *PLoS One*. 2013;8:e66409. <https://doi.org/10.1371/JOURNAL.PONE.0066409>
- Stocker TF, Qin D, Plattner GK, Tignor MMB, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of IPCC the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press; 2014. <https://doi.org/10.1017/CBO9781107415324>
- Sun Y, Ma J, Zhao W, Qu Y, Gou Z, Chen H, Tian Y, Wu F. Digital mapping of soil organic carbon density in China using an ensemble model. *Environ Res*. 2023;231:116131. <https://doi.org/10.1016/j.envres.2023.116131>
- Tatumi SH, Silva LP, Pires EL, Rossetti DF, Góes AM, Munita CS. Datação de sedimentos pós-barreiras no norte do Brasil: Implicações paleogeográficas. *Rev Bras Geocienc*. 2008;38:514-24. <https://doi.org/10.25249/0375-7536.2008383514524>
- Team RC. The R project for statistical computing. Team RC; 2021 [cited 2022 Mar 25]. Available from: <https://ci.nii.ac.jp/naid/10027310073/>.
- Teixeira PC, Donagemma GK, Fontana A, Teixeira WG. *Manual de métodos de análise de solo*. 3. ed. rev e ampl. Brasília, DF: Embrapa; 2017.
- Trumbore SE, Chadwick OA, Amundson R. Rapid exchange between soil carbon and atmospheric carbon dioxide driven by temperature change. *Science*. 1996;272:393-6. <https://doi.org/10.1126/SCIENCE.272.5260.393>
- Valiela I, Lloret J, Bowyer T, Miner S, Remsen D, Elmstrom E, Cogswell C, Robert Thieler E. Transient coastal landscapes: Rising sea level threatens salt marshes. *Sci Total Environ*. 2018;640-641:1148-56. <https://doi.org/10.1016/j.scitotenv.2018.05.235>
- Watanabe K, Seike K, Kajihara R, Montani S, Kuwae T. Relative sea-level change regulates organic carbon accumulation in coastal habitats. *Glob Chang Biol*. 2019;25:1063-77. <https://doi.org/10.1111/GCB.14558>
- Wiesmeier M, Urbanski L, Hobbey E, Lang B, von Lützw M, Marin-Spiotta E, van Wesemael B, Rabot E, Ließ M, Garcia-Franco N, Wollschläger U, Vogel HJ, Kögel-Knabner I. Soil organic carbon storage as a key function of soils - A review of drivers and indicators at various scales. *Geoderma*. 2019;333:149-62. <https://doi.org/10.1016/j.geoderma.2018.07.026>
- Wong VNL, Greene RSB, Dalal RC, Murphy BW. Soil carbon dynamics in saline and sodic soils: A review. *Soil Use Manag*. 2010;26:2-11. <https://doi.org/10.1111/j.1475-2743.2009.00251.x>
- Yost JL, Hartemink AE. Soil organic carbon in sandy soils: A review. *Adv Agron*. 2019;158:217-310. <https://doi.org/10.1016/bs.agron.2019.07.004>
- Zeraatpisheh M, Garosi Y, Reza Owliaie H, Ayoubi S, Taghizadeh-Mehrjardi R, Scholten T, Xu M. Improving the spatial prediction of soil organic carbon using environmental covariates selection: A comparison of a group of environmental covariates. *Catena*. 2022;208:105723. <https://doi.org/10.1016/j.catena.2021.105723>
- Zhang S, Tian J, Lu X, Tian Q. Temporal and spatial dynamics distribution of organic carbon content of surface soil in coastal wetlands of Yancheng, China from 2000 to 2022 based on Landsat images. *Catena*. 2023;223:106961. <https://doi.org/10.1016/j.catena.2023.106961>
- Zhao W, Zhao X, Zhou T, Wu D, Tang B, Wei H. Climatic factors driving vegetation declines in the 2005 and 2010 Amazon droughts. *PLoS One*. 2017;12:e0175379. <https://doi.org/10.1371/JOURNAL.PONE.0175379>