Microchemical Journal 225 (2026) 118191

Contents lists available at ScienceDirect

Microchemical Journal

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

Direct determination of soil carbon stocks by NIRS: an advanced approach☆

Vitor Silveira Freitas [a,*], Lucas Raimundo Bento [a], Vitoria Napolitano [a,b], Patrícia Perondi Anchão Oliveira [c], José Ricardo Macedo Pezzopane [c], Alberto Carlos de Campos Bernardi [c], Débora Marcondes Bastos Pereira Milori [a], Ladislau Martin-Neto [a]

[a] Embrapa Instrumentação, Rua XV de Novembro 1452, 13560-970 São Carlos, SP, Brazil
[b] Federal University of São Carlos, 13565-905 São Carlos, SP, Brazil
[c] Embrapa Pecuária Sudeste, km 234 Washington Luiz Highway, 'Fazenda Canchim', São Carlos 13560-970, São Paulo, Brazil

ARTICLE INFO

Keywords:
NIRS
Soil bulk density
Soil carbon stock
Tropical pastures

ABSTRACT

Soil carbon stocks are obtained as the product of carbon content, soil bulk density, and specific layer thickness. The determination of soil bulk density requires field sampling procedures that can be labor-intensive and time-consuming, particularly when sampling deep soil layers or large areas. In this study, near-infrared spectroscopy (NIRS) was used to directly estimate volumetric soil carbon (Ps, kg C m^{-3}), allowing easier estimation of soil carbon stocks. This approach aims to reduce the difficulties and high cost of traditional soil sampling, which involves opening trenches and collecting undisturbed soil samples using a volumetric ring for soil bulk density determination, as well as using conventional techniques as CHN elemental analyzer. A dataset of 576 soil samples (480 for calibration and 96 for validation) was measured by NIRS, collected from a long-term field experiment encompassing different tropical pasture management systems and a reference native vegetation area: recovered pasture (RP), intercropped pasture (CON), extensive pasture (EX), and native forest (FO). The best model, employing the spectral first derivative as a preprocessing step and Partial Least Squares (PLS) regression as the multivariate algorithm, achieved performance of $R^2 = 0.79$, RMSEC = 4.5 kg C m^{-3}, RPD = 2.16 and RPIQ = 2.85 in calibration, and $R^2 = 0.85$, RMSEP = 4.1 kg C m^{-3}, RPD = 2.53 and RPIQ = 3.63 in validation. Carbon stocks calculated from NIRS-predicted values in the validation dataset showed no significant differences from reference measurements. These findings demonstrate that NIRS can provide reliable, rapid, and cost-effective estimates of soil carbon stocks, highlighting its potential application in carbon credit assessments.

1. Introduction

Accurate quantification of soil carbon stocks is essential for assessing the potential of agricultural systems to sequester soil organic carbon [1,2]. Soil carbon sequestration has been promoted as a strategy to help offset agricultural greenhouse gas emissions through conservative management practices that increase biomass production, enhance soil organic matter (SOM) content, improve soil fertility, and provide several additional co-benefits [3]. Soil carbon sequestration refers to the transference of carbon from atmospheric carbon dioxide (CO_2) to plants through photosynthesis. This carbon is subsequently incorporated into the soil through litter decomposition and root exudates, and it is typically quantified by comparing soil carbon stocks at different points in time to evaluate the impact of management practices [4].

Reliable measurements enable the identification of management practices that enhance soil carbon storage and promote ecosystem sustainability. In addition, robust and scalable quantification methods are crucial for implementing carbon accounting frameworks, such as those required to map carbon stocks across large agricultural areas for carbon credit-based projects that reward farmers for soil carbon sequestration [5].

Soil carbon stock is calculated as the product of carbon concentration (C; g kg^{-1}), soil bulk density (BD; g cm^{-3}), and layer thickness (cm). Soil carbon stocks are commonly measured to a depth of 30 cm, according to the Intergovernmental Panel on Climate Change (IPCC). However, deeper sampling (up to 100 cm), particularly in tropical areas, is

☆ This article is part of a Special issue entitled: 'Green food enviro' published in Microchemical Journal.

* Corresponding author.
E-mail address: vitor.freitas@colaborador.embrapa.br (V.S. Freitas).

https://doi.org/10.1016/j.microc.2026.118191
Received 28 November 2025; Received in revised form 20 March 2026; Accepted 24 April 2026
Available online 28 April 2026

recommended whenever possible. Such sampling generally requires specialized machinery, making the process more time-consuming and costly [6]. In a review of tropical Brazilian soils (*Cerrado* or Brazilian savanna), Carvalho et al. [7] reported that the replacement of native *Cerrado* vegetation by cultivated species reduced the original soil carbon stocks (0–30 cm) by 73%. However, when considering the 0–100 cm layer, 52.4% of the carbon stock data evaluated were higher under cultivated areas than under native *Cerrado* soils. According to Sá et al. [8] and Teixeira et al. [9], soil sampling to a depth of 100 cm is recommended for assessing soil carbon stocks in Brazilian tropical and subtropical soils.

Conventional methods for determining both soil carbon and BD are time-consuming, labor-intensive, and costly. The dry combustion using a CHN Analyzer is one of the most widely employed techniques for determining soil carbon due to its high accuracy and low measurement error (~ 0.30% wt C) [10]. This method requires costly consumables, including high-purity gases, combustion capsules, and catalysts or reagents used for gas conversion and detection. These requirements make the technique expensive and can limit its application for large soil sample sets, for example, in farmlands under periodic soil monitoring [11]. Furthermore, the traditional method for determining bulk density requires digging soil trenches (up to 1 m deep), involves significant field disturbance due to the use of heavy machinery (e.g., excavators), and demands trained personnel to operate the equipment. In addition, intact soil cores must be collected at specific depths using volumetric cylinders by trained personnel to minimize sampling errors.

The core method for bulk density sampling is susceptible to several sources of error, mainly in large farmland areas, including: (i) operator experience, (ii) sampling depth, (iii) the size of the core cylinder, (i*v*) soil moisture content, and (v) performing precise sampling [12]. The associated error of BD measurements is typically around 7–8%. However, when considering all potential sources of error, from field sampling to laboratory analyses, error propagation can occur, potentially resulting in overall uncertainties ranging from 10 to 40% [13].

These limitations stress the need for rapid, accurate, cost-effective, and environmentally friendly methods to analyze key soil parameters, such as soil carbon and bulk density. The use of infrared spectroscopy techniques, such as near-infrared (NIR, 700–2500 nm), visible–near infrared (Vis-NIR, 380–2500 nm), and mid-infrared (MIR, 2500–25,000 nm), stands out as a simplified alternative to traditional analytical methods [14–16].

Estimations of soil organic carbon using Vis-NIR under laboratory conditions are widely reported in the literature [17–20] and have proven to be a practical, cost-effective, and efficient alternative to the commonly used conventional elemental analyzer. Additionally, NIR and Vis-NIR spectroscopies were accredited by the Verified Carbon Standard (VCS) of VERRA (2023), as alternative methods for soil carbon quantification, enabling their application in measurement, reporting, and verification (MRV) projects to generate carbon credits [21].

Direct estimations of BD using infrared spectroscopy have been reported with low predictive accuracy [22–24]. This limitation arises because BD is affected by multiple factors, including texture, depth, soil organic matter content, management practices, and land use. In addition, BD is strongly influenced by structural pore-space characteristics, which are not detectable by NIRS [25], making it challenging to develop robust predictive models.

Recent studies have attempted to predict volumetric soil organic carbon (SOCv, g C cm^{-3}), defined as the product of soil total carbon and bulk density, using Vis–NIR spectroscopy, thereby enabling subsequent estimation of soil carbon stocks by multiplying SOCv by the corresponding layer thickness. With relatively small datasets, Allory et al. [26], using 137 soil samples, reported a coefficient of determination (R^2val) of 0.89 and an associated root mean square error of prediction (RMSEP) of 5.1. Allo et al. [27] achieved an R^2 of 0.87 and an RMSEP of 0.61 when testing 95 soil samples; however, these results were obtained by cross-validation without partitioning the data into separate calibration and validation sets. With a larger dataset of 584 soil samples, Cambou et al. [28] reported an R^2val of 0.60 and an RMSEP of 1.9, while Dharumarajan et al. [25], using 361 samples, obtained an R^2val of 0.52 and an RMSEP of 4.94.

Unlike previous studies, the objective of this study was to evaluate the potential of near-infrared spectroscopy, excluding the visible range, for assessing soil carbon stocks. Specifically, the study aimed to: (i) directly predict volumetric soil carbon (Ƥs kg C m^{-3}) using NIRS; (ii) estimate soil carbon stocks up to 1 m depth from these predictions and compare them with values obtained using conventional methods; and (iii) evaluate the performance of two modeling approaches, partial least squares regression (PLS), a well-established multivariate method widely used in soil spectroscopy, and random forest (RF), a machine learning algorithm capable of capturing nonlinear relationships between spectral information and soil properties. We hypothesize that, since soil carbon can be accurately predicted by NIRS, and BD is correlated with soil properties such as texture and soil organic matter, which are also predictable using NIRS, it is possible to develop robust models for estimating Ƥs. This approach has the potential to simplify and optimize the quantification of soil carbon stocks, representing a noteworthy contribution to sustainable agriculture, climate change mitigation, and the carbon credit market.

2. Material and methods

2.1. Study site

The study was conducted at Embrapa Livestock Southeast Research Center (*Embrapa Pecuária Sudeste*) in São Carlos, Brazil (21°57′ S, 47°50′ W; 860 m above sea level), within a long-term field experiment established in 2011. The soil at the experimental site is classified as a Typic Haplustox (US Soil Taxonomy) and has a sandy-loam texture. The regional climate is classified as tropical highland (Köppen classification: Cwa), marked by distinct winter and summer seasons [29].

The experiment comprises three pasture-based beef cattle production systems, each replicated in three plots, and a native forest used as a reference (Fig. 1). i) Recovered pasture (RP) of *Urochloa decumbens* (syn. *Brachiaria decumbens)* Stapf cv. Basilisk, with soil amendment and mineral fertilization using macro and micronutrients, under intensive management with a moderate stocking rate; ii) Intercropped pasture (CON) of *Urochloa* (syn. *Brachiaria*) *decumbens* Stapf cv. Basilisk, with soil amendment and mineral fertilization using macro and micronutrients (except nitrogen), intercropped with pigeon pea (*Cajanus cajan*), under intensive management with a moderate stocking rate; iii) Extensive pasture (EX) consisting of a mixture of *Urochloa brizantha* Stapf cv. Marandu and *Urochloa decumbens* Stapf cv. Basilisk, without the annual addition of mineral fertilizer. iv) Forest (FO) corresponds to a Semideciduous seasonal forest of the Atlantic Foret biome near the experimental area, with soil texture and classification identical to the pasture areas.

2.2. Experimental design and soil sampling

Each management treatment (RP, CON, and EX) comprised three experimental replicates. In each replicate, two soil trenches (1 m deep) were excavated, totaling six trenches per treatment, in addition to six trenches in the forest area (FO). Soil sampling followed the protocol established by Oliveira [30]. Undisturbed samples were collected from three faces, using volumetric rings, of each trench for bulk density determination, while disturbed samples were obtained for soil carbon analysis at depths of 0–5, 5–10, 10–20, 20–30, 30–40, 40–60, and 80–100 cm. Around each trench, four transects were established from the corners, along which three disturbed samples were taken at 5 m intervals at the same depths as those sampled from the trench faces (Fig. 2).

Overall, 576 soil samples were collected. Samples were air-dried and

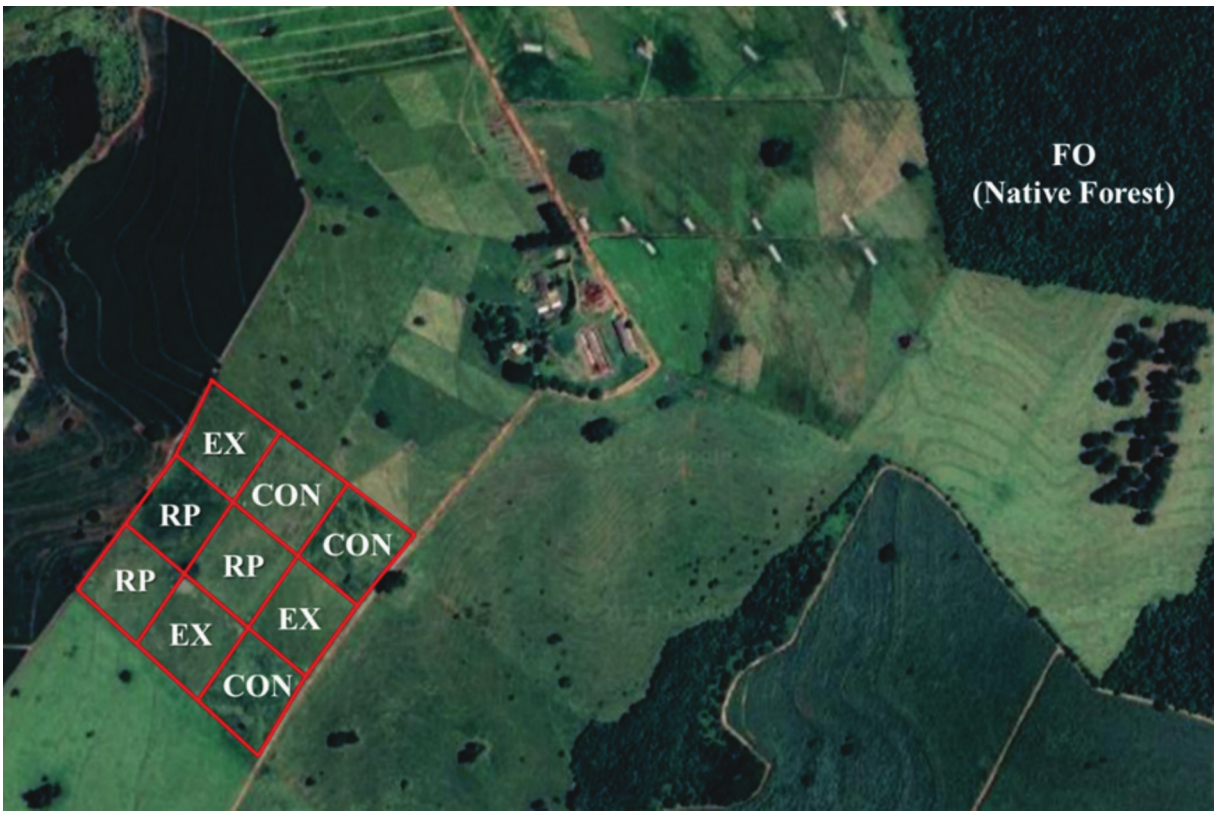

Fig. 1. Experimental site of tropical pastureland showing three replicates for each management treatment: Recovered Pasture (RP), Extensive Pasture (EX), Intercropped Pasture (CON), and Native Forest sampling site (FO).

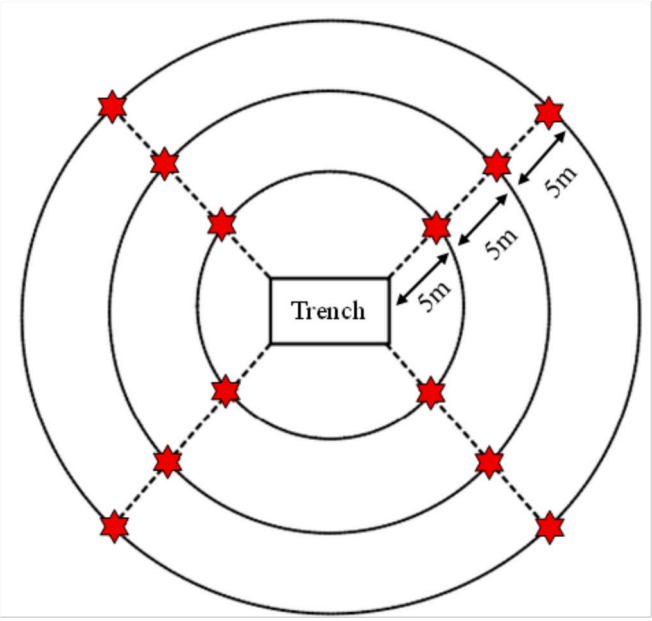

Fig. 2. Soil sampling around the trenches to improve detection of soil carbon variability (Adapted from Oliveira et al., 2014).

sieved through a 2 mm mesh. The chemical composition of the 0–20 cm soil layer was characterized by the following parameters: pH (H_2O) = 5.9, pH ($CaCl_2$) = 5.2, organic matter = 23.7 g dm^{-3}, P = 4.3 mg dm^{-3}, K^+ = 1.1 mmolc dm^{-3}, Ca^{2+} = 24.4 mmolc dm^{-3}, Mg^{2+} = 12.2 mmolc dm^{-3}, Al^{3+} = 0.6 mmolc dm^{-3}, H + Al = 27.9 mmolc dm^{-3}, SB = 37.8 mmolc dm^{-3}, CEC = 65.6 mmolc dm^{-3}, V = 57.4%, and m = 1.9% [25].

2.3. Soil analysis by reference methods

Soil carbon concentration was determined by an elemental analyzer (Gerhardt DUMATHERM C/N). Bulk density was obtained from undisturbed soil cores collected with 100 cm^3 volumetric rings inserted into the trench walls at the specified depths. Bulk density was calculated gravimetrically, based on the oven-dry mass of each core after drying at 105 °C [9]. Soil texture was analyzed using the pipette method [31].

2.4. Soil carbon stock calculations

For the direct estimation of soil carbon stocks using NIRS, we calculated the product of soil carbon concentration × bulk density (volumetric soil carbon), referred to as Ƥs (Eq. 1). Subsequently, per-layer carbon stock was calculated following Eq. 2, and the values were summed to obtain the total stock to a depth of 1 m.

$$Ƥs\ (kg\ C\ m^{-3}) = C \times BD \quad (1)$$

$$Cstock\ (Mg\ C\ ha^{-1}) = Ƥs \times f \times 0.1 \quad (2)$$

where Cstock is the carbon stock in each layer in Mg C ha^{-1}; C is the soil total carbon in g kg^{-1}; BD is the soil bulk density in g cm^{-3}; f is the layer thickness in centimeters. The coarse fragment fraction (> 2 mm) was not considered in the calculation of soil carbon stocks, as the study areas are predominantly devoid of rock fragments.

2.5. NIRS data acquisition

Grounded dry samples (2 mm) were transferred to flat-bottom glass tubes for spectral acquisition. Measurements were carried out in the range of 10.000 to 4.000 cm^{-1} (1000–2500 nm) in reflectance mode using an Antaris II FT-NIR Analyzer (ThermoFisher). Spectra were

acquired at a resolution of 8 cm^{-1}, with each spectrum representing the average of 32 scans. Three independent spectra were collected for each sample, and the final spectrum used in the analyses corresponded to the average of these three measurements. The instrument automatically converted the spectra from reflectance to absorbance.

2.6. Spectral pre-processing and calibration models for volumetric soil carbon prediction

Two spectral preprocessing strategies were evaluated: (i) baseline correction (BL) and (ii) Savitzky–Golay first derivative filtering (SG). Baseline correction is a common step in NIRS soil analyses, and the Savitzky–Golay algorithm reduces noise by fitting low-order polynomials within local windows, preserving key spectral features such as peaks and inflection points. In addition, applying the first derivative of the spectra enhances subtle spectral features. These preprocessing methods were applied to the reflectance spectra to minimize physical variability caused by light scattering and to enhance relevant spectral characteristics [32,33]. Spectral preprocessing and modeling were performed using the R software.

The dataset comprising 576 samples was divided into a calibration set (480 samples) and a validation set (96 samples). The validation set was constructed by randomly selecting one soil trench for each treatment (RP, CON, EX, and FO). Table 1 presents the summary statistics of the soil properties for the entire dataset, as well as for the calibration and validation sets.

Calibration models for the quantification of Ps were evaluated using two modeling approaches: Partial Least Squares Regression (PLS-BL, PLS-SG) and Random Forest Regression (RF-BL, RF-SG), which were applied with both preprocessing strategies.

PLS is a widely used linear multivariate method in infrared spectroscopic modeling. Spectral information is condensed into a smaller set of orthogonal components, known as latent variables (LVs), which capture the variance most relevant to the property of interest [34]. The optimal number of LVs for each PLS model was determined using a 10-fold cross-validation procedure, ensuring robust prediction of the soil property. In the calibration set, PLS-BL used 6 LVs, while PLS-SG used 7 LVs.

RF is a machine learning method that combines bootstrap aggregation with random feature selection to improve prediction performance. At each node of a decision tree, a random subset of predictors is considered, and the best split among them is chosen. Trees are grown until reaching a specified number of nodes [35]. For both calibration models (RF-BL and RF-SG), the parameters were set as follows: 200 trees, 20 predictor variables per split (mtry = 20), and 10 bootstrap resamples using the function optimism_boot.

Table 1
Summary of soil properties for the full dataset, calibration, and validation set.

	n		Sand (%)	Clay (%)	Silt (%)	BD (g dm^{-3})	C (g kg^{-1})
Full set	576	Min	41	11	1	0.82	3.8
		Max	84	50	27	1.67	56.8
		Mean	61	31	8	1.31	13.4
		Skewness	0,18	-0,16	1,01	−0.20	1.66
Calibration set	480	Min	41	11	1	0.82	3.8
		Max	84	50	27	1.67	56.8
		Mean	61	31	8	1.31	13.6
		Skewness	0,20	−0,15	1,03	−0.20	1.72
Validation set	96	Min	48	18	1	0.96	4.6
		Max	77	42	16	1.57	37.6
		Mean	62	30	7	1.32	12.4
		Skewness	0.21	−0.34	0.58	0.01	1.38

Number of samples (n); bulk density (BD); total carbon (C); Minimum values (Min); Maximum values (max).

2.7. Model metrics evaluation

Model performance was evaluated using the coefficient of determination (R^2), root mean square error (RMSE), ratio of performance to deviation (RPD), ratio of performance to interquartile range (RPIQ), and *p*-value. R^2 reflects the proportion of variance explained by the model, with values closer to 1 indicating a better fit. RMSE measures prediction precision, with lower values representing higher accuracy. RPD expresses the ratio between the standard deviation of observed values and the RMSE, with higher values indicating better predictive performance and model reliability. RPIQ relates the interquartile range of observed data to model error, with higher values indicating superior predictive performance [36]. The *p*-value tests whether predicted and observed values differ significantly, with *p*-values >0.05 indicating no significant difference and suggesting that the model predictions are consistent with the observed data.

3. Results

3.1. Soil properties

After splitting the dataset into calibration ($n = 480$) and validation ($n = 96$) subsets, we compared their distributions with the full dataset ($n = 576$) to ensure a representative split. The calibration set encompassed a broad range of soil textures, with sand varying from 41 to 84%, clay from 11 to 50%, and silt from 1 to 27% (Table 1). Bulk density (BD) varied between 0.82 and 1.67 g cm^{-3}, while soil carbon content ranged from 3.8 to 56.8 g kg^{-1}. Together, these ranges captured the variability observed in the full dataset, supporting the suitability of the calibration set for model calibration.

The use of a large dataset (576 soil samples in this work) enhances the representativeness of the calibration set by incorporating a broader range of variability in soil properties and their corresponding spectral responses. A larger and more diverse dataset optimizes model robustness and predictive performance, as it allows the algorithm to better capture the relationship between spectral information and the soil property being modeled.

Including the complete variability of soil texture, bulk density, and carbon concentration in the calibration dataset is crucial to minimizing extrapolation errors and improving model robustness. If the calibration set does not adequately cover the entire range of variability, the algorithms may be forced to extrapolate during validation, which may result in biased or unreliable predictions. The validation subset exhibited comparable ranges, with sand at 48–77%, clay at 18–42%, silt at 1–16%, BD 0.96–1.57 g cm^{-3}, and carbon concentration at 4.6–37.6 g kg^{-1}. These similarities confirm a balanced dataset division suitable for model evaluation.

3.2. Spectral characteristics and preprocessing

Representative near-infrared (NIR) spectra are shown in Fig. 3, illustrating the effect of preprocessing techniques. The Fig. 3a displays the original spectrum. Baseline correction (Fig. 3b) removed scattering effects, while the Savitzky–Golay first derivative (Fig. 3c) enhanced subtle spectral features associated with organic and mineral components. These spectra correspond to a soil sample with 24 g kg^{-1} carbon, 67.6% sand, 28.2% clay, 4.2% silt, and a bulk density (BD) of 1.3 g cm^{-3}.

The NIRS spectra of the samples exhibited the main absorption features in the regions of 4000–4700 cm^{-1}, 4900–5300 cm^{-1}, and 6750–7500 cm^{-1}, along with two weak absorptions at 6400 and 6600 cm^{-1}. The first region (4000–4700 cm^{-1}) corresponds to absorptions from organic molecules, including aliphatic carbon, carbohydrates, and polysaccharides [37]. Additionally, absorptions in this region may be attributed to aluminum and iron oxides present in the clay fraction of the soil, such as smectite, illite, and kaolinite [38].

A strong absorption in the 4900–5300 cm^{-1} region corresponds to

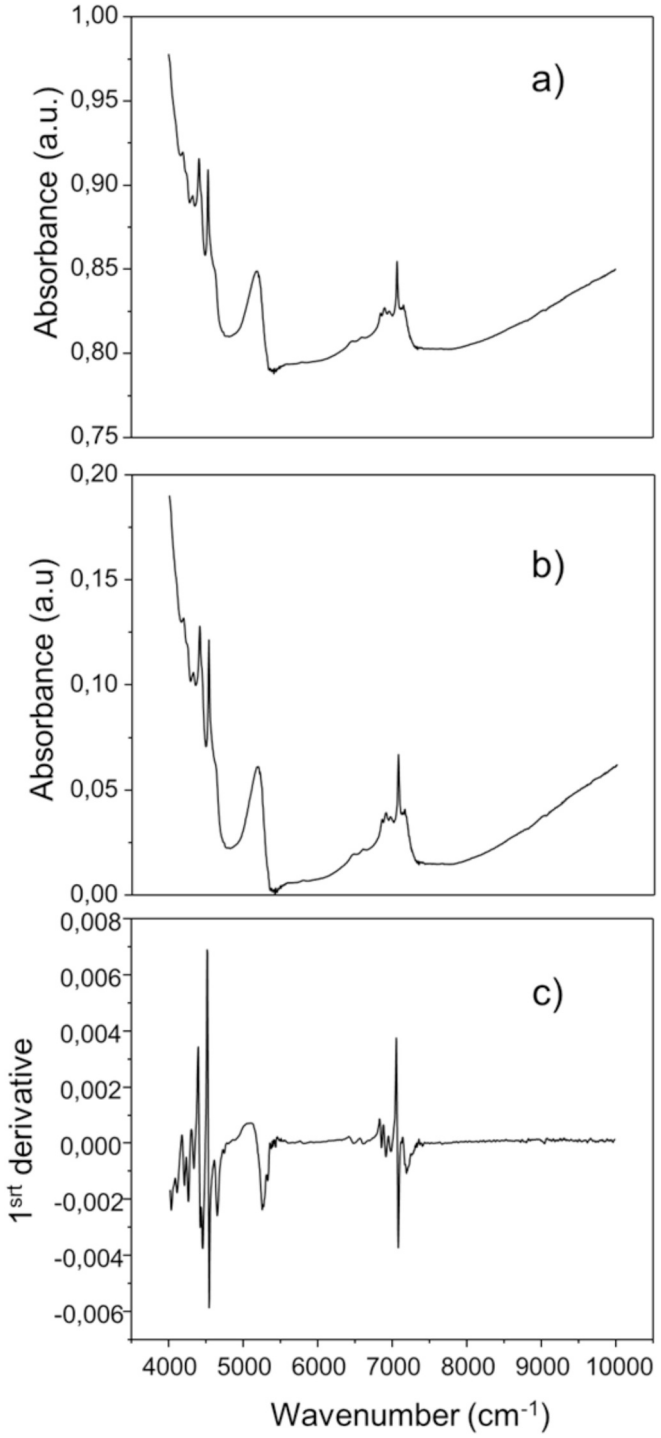

Fig. 3. a) NIRS spectrum; b) baseline corrected spectrum; c) spectrum after Savitzky–Golay first derivative filtering.

the combination of O—H stretching and H—O—H bending vibrations of water molecules structurally bound within the crystal lattice [39]. Weak absorptions in 6400 and 6600 cm^{-1} are associated with the presence of amine and amide functional groups in soil organic matter (SOM). The absorption features in the 6750–7500 cm^{-1} region are attributed to water bound within the interlayer spaces of clay minerals, as well as carboxylic acids and hydroxyl groups from SOM [34].

These overlapping features jointly reflect the soil organic–mineral matrix that governs volumetric carbon variability (Ƥs). Table 2 summarizes the main absorption regions for minerals and soil organic matter components in the NIR range.

Table 2
Spectral regions of fundamental mid-IR absorptions and corresponding overtone and combination bands of soil constituents in the NIR range (adapted from Viscarra Rossel and Behrens, 2010).

Constituent	Fundamental (cm^{-1})	NIR (cm^{-1})	NIR mode
Water	ν1 O—H (3278)	5221,93	ν2 + ν3
	ν2 H—O—H (1645)	6872,85	2ν2 + ν3
	ν3 O—H (3484)	7246,38; 8810,57	ν1 + ν3, ν1 + ν2 + ν3
Hydroxyl	ν1 O—H (3575)	7142,86	2ν1
Kaolin doublet (Clay)	ν1a O—H (3695)	7168,46	2ν1a
	ν1b O—H (3620)	7067,14	2ν1b
Kaolin (Clay)	δ Al—OH (915)	4629,63; 4528,99	ν1a + δ, ν1b + δ
Carbonate	ν3 CO_3^{-2} (1415)	4280,82	3ν3
Aromatics	ν1 C—H (3030)	6060,61; 9090,91	2ν1, 3ν1
Amine	δ N—H (1610)	4854,37	ν1 + δ
	ν1 N—H (3330)	6666,67; 100,000	2ν1, 3ν1
Alkyl	ν3 C—H (2930)	5861,66	2ν3
	ν1 C—H (2850)	5701,25; 8787,35; 8547,01	2ν1, 3ν3,3ν1
Carboxylic acids	ν1 C═O (1725)	5181,35; 6901,31	3ν1, 4ν1
Amides	ν1 C═O (1640)	4918,84; 6561,68	3ν1, 4ν1
Aliphatics	ν1 C—H (1465)	4395,60; 5861,66	3ν1, 4ν1
Methyl	ν1 C—H (1445–1350)	4050,22–4334,63	3ν1
Phenolics	ν1 C—OH (1275)	6020,47	4ν1
Polysaccharides	ν1 C—O (1170)	4679,46	4ν1
Carbohydrates	ν1 C—O (1050)	4199,92	4ν1

ν1: Symmetric stretching; ν2: Bending; ν3: Asymmetric stretching; δ: Bending.

3.3. Model calibration and validation performance

A multivariate algorithm (PLS), previously applied to predict volumetric soil carbon using Vis–NIR spectroscopy [25–28], and a machine learning algorithm (RF), used for soil organic carbon prediction with infrared spectroscopy and for bulk density estimation via pedotransfer functions [19,40,41], were evaluated in combination with two spectral preprocessing strategies (baseline correction and first derivative) using the full NIRS spectrum to predict volumetric soil carbon (Ƥs). This approach aims to simplify the metrics required for soil carbon stock estimation and eliminate the need for separate measurements of soil carbon and bulk density.

During calibration, RF-BL and RF-SG models achieved better performance, with R^2 = 0.96, RPIQ values of 5.95 and 6.17, and lower errors (RMSEC = 2.2 and 2.1 kg C m^{-3}), respectively. However, their accuracy decreased during the validation phase (R^2 = 0.77–0.80; RMSEP ~4.9 kg C m^{-3}), indicating mild overfitting. PLS models using first-derivative preprocessing performed better during validation. The PLS-SG model yielded the most robust results in the validation, with R^2 = 0.85, RMSEP = 4.1 kg C m^{-3}, RPD = 2.53, and RPIQ = 3.63. The performance of the models is summarized in Table 3. The close alignment of measured and predicted Ƥs values for the PLS-SG model along the 1:1 line (Fig. 4) indicates that the model provides accurate predictions.

3.4. Prediction of specific layer and summed soil carbon stocks

Predicted and reference carbon stocks per layer and cumulative to 1 m depth are presented in Table 4 and Fig. 5 for the best model (PLS-SG). Across management systems, a *t*-test at the 95% confidence level showed no significant differences between NIRS predictions and reference measurements, indicating good agreement between methods. In addition, the propagated uncertainty (σ_{total}) associated with the total carbon stock estimates provides an integrated measure of variability,

Table 3
Model performance of volumetric soil carbon prediction.

	Calibration set (n = 480)					Validation set (n = 96)				
Model	R^2	RMSEC	RPIQ	RPD	p_value	R^2	RMSEP	RPIQ	RPD	p_value
PLS - BL	0.78	4.6	2.82	2.14	0.90	0.82	4.5	2.76	2.30	0.07
PLS - SG	0.79	4.5	2.85	2.16	0.92	0.85	4.1	3.63	2.53	0.55
RF - BL	0.96	2.2	5.95	4.50	0.96	0.77	5.3	2.79	1.94	0.30
RF - SG	0.96	2.1	6.17	4.67	0.97	0.80	4.9	3.06	2.13	0.61

Partial least squares regression with baseline correction (PLS-BL); partial least squares regression with Savitzky–Golay first derivative (PLS-SG); random forest regression with baseline correction (RF-BL); and random forest regression with Savitzky–Golay first derivative (RF-SG).

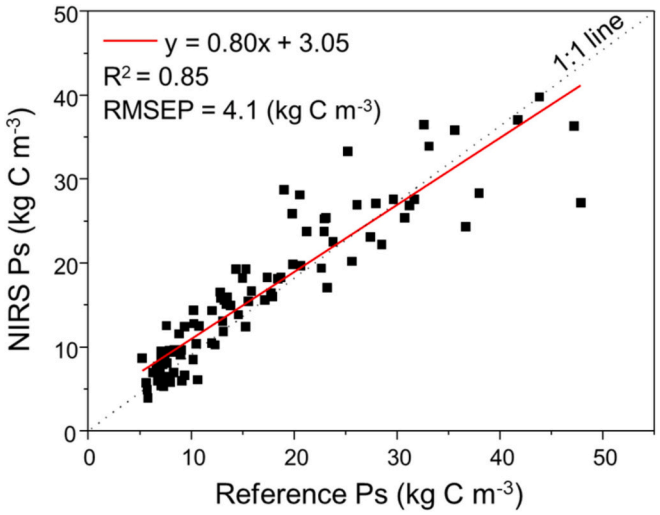

Fig. 4. Scatter plot comparing NIRS predicted versus reference measurements of C x BD (Ps), using PLS-SG strategy.

accounting for the combined uncertainties involved in the stock calculation.

Predicted and reference carbon stocks per layer and cumulative to 1 m depth are presented in Table 4 and Fig. 5 for the best model (PLS-SG). Across management systems, A t-test at the 95% confidence level showed no significant differences between NIRS predictions and reference measurements, demonstrating good model performance.

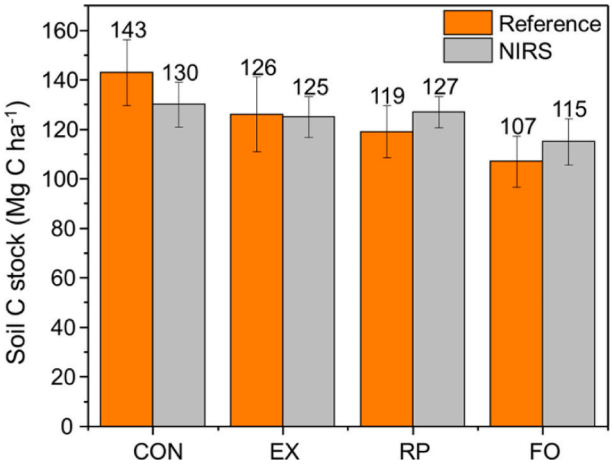

Fig. 5. Soil carbon stocks calculated from reference measurements, obtained by CHN analyzer to C content and volumetric ring to soil bulk density, and NIRS predicted Ps.

Table 4
Layer-specific soil carbon stocks (Mg C ha^{-1}), standard deviation (SD), and overall error (σ_{total}) for the reference method and predicted by NIRS.

	Treatment							
	CON				RP			
Depth (cm)	C stock (reference)	SD (reference)	C stock (NIRS)	SD (NIRS)	C stock (reference)	SD (reference)	C stock (NIRS)	SD (NIRS)
00–05	21.1	3.0	18.7	1.1	17.9	2.6	17.9	0.8
05–10	15.8	3.1	13.8	3.2	10.3	5.7	14.1	2.1
10–20	25.6	4.2	22.0	2.4	18.3	3.9	19.6	1.6
20–30	19.6	5.0	16.5	0.1	14.7	1.0	16.3	0.9
30–40	15.3	7.0	14.4	5.4	11.4	2.1	13.3	3.5
40–60	19.4	6.8	19.0	4.8	19.1	4.7	19.6	3.9
60–80	14.7	1.8	13.4	1.5	14.0	3.4	14.3	0.3
80–100	12.0	4.1	12.3	3.1	13.0	4.3	11.4	1.9
Total C stock	143.4		130.2		118.6		126.6	
σ_{total}		13.3		9.0		10.6		6.3
	EX				**FO**			
Depth (cm)	**C stock (reference)**	**SD (reference)**	**C stock (NIRS)**	**SD (NIRS)**	**C stock (reference)**	**SD (reference)**	**C stock (NIRS)**	**SD (NIRS)**
00–05	14.2	1.3	13.5	0.1	16.6	1.9	14.3	1.0
05–10	11.5	2.1	12.7	0.5	9.9	2.0	11.9	3.3
10–20	22.1	2.0	20.3	0.5	15.1	2.0	17.5	3.5
20–30	15.6	2.3	16.0	0.4	10.3	4.8	14.2	0.9
30–40	12.5	5.7	11.0	4.1	9.5	5.7	11.1	4.5
40–60	19.3	8.7	21.9	5.4	17.2	6.0	14.4	4.3
60–80	16.1	6.6	15.8	1.7	15.1	1.0	16.0	2.1
80–100	14.4	7.8	14.0	4.3	13.3	1.8	15.5	4.4
Total C stock	125.6		125.2		106.9		115.0	
σ_{total}		15.1		8.2		10.3		9.3

Intercropped pasture (CON); Recovered pasture (RP); Extensive pasture (EX); Forest (FO); Standard deviation (SD), carbon stock (C stock); overall error (σ_{total}).

Since the validation set consisted of one 1-m-deep trench per area (treatment), with three samples collected at each depth, we calculated soil C stock (Eq. 2) and the associated error using Ps values predicted by NIRS and compared them with the reference method (Fig. 5). In the validation set, soil carbon stocks ranged from 118.6 to 143.4 Mg C ha^{-1} in pastures and from 106.9 to 115.0 Mg C ha^{-1} in forest soils, considering both the reference measurements and NIRS predicted values. The overall error (σ_{total}) for NIRS predicted soil C stocks ranged from 6.3 to 9.3 Mg C ha^{-1} and was lower than for reference measurement (bulk density–elemental analyzer), which ranged from 10.6 to 15.1 Mg C ha^{-1} (Table 4). Moreover, NIRS predictions exhibited reduced within-layer variability, reflecting the method's higher analytical precision and lower sensitivity to soil heterogeneity.

4. Discussion

4.1. Preprocessing effects on models' performance

The direct estimation of soil carbon stocks using NIRS is based on the principle that soil carbon can be accurately predicted from the absorption features of soil organic matter functional groups, such as C═O, C—H, O—H, C═C, and others, which arise from bending and stretching vibrations in the NIR region [25], and bulk density (BD) is strongly correlated with soil carbon and texture, and influences key properties such as porosity and water retention [18,40]. An integrated approach was used to develop spectral models for estimating volumetric soil carbon using NIRS.

Evaluating the *p*-values of the proposed models, all of them indicated no significant difference between the predicted and observed values ($p > 0.05$). The Savitzky–Golay first derivative preprocessing strategy improved model performance for both PLS and RF algorithms, consistent with its ability to remove baseline drifts and amplify minor spectral features [35,42], which in turn enhances the performance of multivariate and machine learning algorithms.

RF captured nonlinear spectral property relationships effectively but exhibited overfitting, evidenced by the performance drop from calibration ($R^2 = 0.96$; RMSEC = 2.1) to validation ($R^2 = 0.80$; RMSEP = 4.9). PLS, in contrast, achieved better generalization, improving from $R^2 = 0.79$ to 0.85 in validation. Because PLS forces a linear approximation of the data, the calibration fit is inherently weaker. However, it shows better generalization performance when applied to independent data, resulting in better validation performance. The performance of PLS-SG suggests that linear models can outperform nonlinear ones when preprocessing effectively enhances spectral relevance. Taken together, these results enable inferring that while RF is powerful in modeling complex and nonlinear spectral patterns, it is more susceptible to overfitting [41]. In contrast, PLS can deliver superior generalization on unseen data despite its more constrained calibration.

4.2. Models evaluation metrics

The ratio of performance to interquartile range (RPIQ) was used to assess model accuracy by evaluating the central distribution of the data, making it less sensitive to skewed values and outliers. Providing a low-cost and reliable tool for assessing model performance in NIRS-based estimation of soil carbon, particularly in the context of the carbon credits market [36]. Nawar and Mouazen [43] proposed an interpretation of RPIQ values, classifying models as excellent (RPIQ ≥2.5), very good (2.5 > RPIQ ≥2.0), good (2.0 > RPIQ ≥1.7), fair (1.7 > RPIQ ≥1.4), and poor (RPIQ <1.4). In our study, all prediction models achieved excellent performance, with PLS-SG performing best (RPIQ = 3.63).

However, for the Random Forest algorithm, the best model (RF-SG) showed a pronounced decrease in RPIQ, from 6.17 in calibration to 3.06 in validation, which strongly indicates overfitting. Conversely, the PLS-SG model improved its RPIQ from 2.85 in calibration to 3.63 in validation. This increase suggests that the validation dataset represented the central, less noisy portion of the calibration data, with fewer outliers and reduced noise, facilitating accurate prediction.

4.3. Predicted soil carbon stocks, influence of BD and layer thickness

The carbon stocks and their associated standard deviations were calculated for each soil layer (Table 4), as well as the overall error between the reference measurements and the NIRS-predicted carbon stocks, using the square root of the sum of the squared standard deviations of each layer, as described in [44]. Fig. 5 presents the carbon stocks for the 1 m soil profile, estimated from NIRS-predicted Ps, which did not differ significantly from the reference values at the 95% confidence level based on a *t*-test.

The absolute error of BD was higher in the superficial 0–30 cm layers, ranging from 0.36 to 0.88, compared to the deeper 30–100 cm layer, where the absolute errors ranged from 0.17 to 0.21 across the different treatments. Table S1 highlights this pattern, presenting the BD values and their associated layer-specific standard deviations as well as the absolute errors for both the 0–30 cm and 30–100 cm layers in the validation dataset.

This difference can be attributed to the higher carbon and soil organic matter contents, as well as greater microbial activity, in the surface layers, which lead to increased soil heterogeneity and spatial variability. In contrast, deeper layers are less influenced by root activity and generally receive lower organic matter inputs, resulting in lower variability and, consequently, lower BD measurement errors.

By using NIR spectroscopy, we were able to reduce the influence of errors associated with both BD and carbon measurements obtained through conventional methods. As shown in Table 4, for all treatments, the carbon stock estimated by NIRS exhibited lower overall error compared to the reference method.

Moreover, in agricultural systems, the standard deviations among the triplicates were consistently lower for the NIRS-predicted values than for the reference measurements. It is important to emphasize that the three samples per layer were not true analytical replicates, but rather spatial replicates collected in proximity within the same depth interval. Therefore, the lower variability observed for NIRS suggests higher internal precision and consistency of the spectroscopic predictions.

This suggests that NIRS provides more stable estimates of soil carbon stocks and is less affected by the small-scale spatial heterogeneity that typically influences conventional laboratory measurements. Together, these results demonstrate that NIRS reduces error propagation in soil carbon stock estimates, particularly when deeper layers (30–100 cm) are included in the calculations. In a recent study, dos Santos et al. [40] used pedotransfer functions to estimate BD ($R^2 = 0.74$; RMSE = 0.12), and when these predicted BD values were used to calculate soil carbon stocks, a strong correlation ($R^2 = 0.97$) was observed between predicted and reference stocks. Showing that uncertainties in BD estimation do not necessarily translate into large errors in soil carbon stock estimates.

The best model achieved an R^2 of 0.79 and RMSEC of 4.5 kg C m^{-3} in the calibration set, with a corresponding R^2 of 0.85 and RMSEP of 4.1 kg C m^{-3} in the independent validation set, and RPIQ values of 2.85 and 3.63, respectively. These results indicate a robust predictive performance for volumetric soil carbon at the layer level, with slightly better generalization on the validation set. Compared with previous studies using Vis-NIR spectroscopy, our model performs favorably. For instance, Allory et al. [26] reported R^2val = 0.89 and RMSEP = 5.1 with 137 samples, while Allo et al. [27] obtained $R^2 = 0.87$ and RMSEP = 0.61, though without separating calibration and validation sets. Studies with larger datasets, such as Cambou et al. [28] (R^2val = 0.60, RMSEP = 1.9, $n = 584$) and Dharumarajan et al. [25] (R^2val = 0.52, RMSEP = 4.94, $n = 361$), reported lower predictive performance, highlighting the challenge of scaling models to heterogeneous soils.

4.4. Final remarks on NIRS performance

An important distinction of this work is the use of the NIR range rather than Vis-NIR. Unlike Vis-NIR spectroscopy, which includes the visible and near-infrared regions, our approach relies exclusively on NIRS, excluding the visible range. The visible range contributes to models by capturing information related to soil color, including variations in pedogenic oxides and soil organic matter composition, which can help explain differences in soil properties across heterogeneous samples [45].

Most VIS–NIR spectrometers used for soil analysis employ monochromators or diffraction gratings to obtain reflectance signals directly as a function of wavelength, without applying the Fourier transformation [46]. Instruments operating exclusively in the NIR region, on the other hand, often use interferometers to collect the reflectance signal generated by differences in optical path length. The resulting interferogram is then processed through a Fourier transform to produce the spectrum as a function of wavenumber, providing higher spectral resolution that can improve model performance [47,48]. Overall, our findings demonstrate that, with proper data preprocessing and modeling, NIRS alone can provide reliable layer-specific predictions of volumetric soil carbon for a local and heterogeneous dataset.

NIRS predictions showed lower within-layer variability (Table 4), compared with conventional methods, suggesting higher analytical precision and reduced susceptibility to small-scale heterogeneity. Additionally, conventional soil carbon stock determination requires substantial sampling effort, including opening trenches and collecting BD samples, which may introduce uncertainties associated with BD determination using the volumetric ring method. Moreover, total carbon analysis using an elemental analyzer involves time-consuming sample preparation, such as grinding samples to particle sizes <150 μm and precise weighing on an analytical balance, in addition to the high cost of consumables. In this context, NIRS emerges as a practical and efficient alternative. With high analytical precision, low operational costs, and minimal sample preparation, NIRS stands out as a robust approach for the large-scale quantification of soil carbon stocks as well as for carbon accounting and monitoring. When coupled with robust calibration models, this approach could support scalable and consistent soil carbon measurements in tropical systems.

4.5. Limitations and implications for large-scale applications

Limitations of this method may arise from the dataset used to develop the model. The model was calibrated using a local dataset composed of samples with relatively similar physicochemical characteristics, including a specific soil type and comparable edaphoclimatic conditions. As a result, the predictive performance observed in this study may be influenced by the relative homogeneity of the sample's physicochemical properties, as they were collected from the same experimental area.

Considering the potential scale-up of this approach to regional or global applications, a much broader diversity of soils would need to be included in the calibration dataset. Agricultural soils vary widely in texture, bulk density, color, mineralogy, and soil organic matter content, which can significantly affect spectral responses and model transferability.

Although increasing the number of samples generally improves model robustness, datasets containing highly heterogeneous soils may require more advanced modeling strategies. In such cases, machine learning approaches, such as artificial neural networks, could be explored to better capture complex spectral patterns and improve predictive performance across diverse soil conditions.

Future research should therefore focus on expanding the dataset to include a wider range of soil types, management systems, and environmental conditions. Such efforts would contribute to the development of more generalizable models and support the application of spectroscopy-based approaches for large-scale soil monitoring. In this context, improving the scalability and transferability of these models is particularly relevant to soil carbon accounting initiatives and carbon credit frameworks that require the reliable quantification of soil carbon stocks across extensive agricultural areas.

5. Conclusion

Accurate volumetric soil carbon predictions can be achieved using NIRS. Results demonstrated that applying a simple spectral preprocessing step (first derivative) combined with a widely used multivariate regression model (PLS) yielded robust performance, with an R^2_{val} of 0.85, an RMSEP of 4.09 kg C m^{-3}, RPD of 2.53, and an RPIQ of 3.63. Using the validation set, we estimated carbon stocks for each layer and summed them across the 1-m soil profile to assess differences between reference and NIRS measurements. The results highlight the strong potential of NIR spectroscopy in predicting soil carbon stock, as no significant statistical differences were observed between NIRS and reference data.

Moreover, the NIRS predicted soil carbon stocks exhibited lower variability, indicating greater consistency among replicates. These findings suggest that NIRS can provide reliable, rapid, and cost-effective estimates of soil carbon stocks, supporting its application as a precision agriculture tool and as a more practical method for soil carbon stocks quantification, with potential use in the carbon credits market.

CRediT authorship contribution statement

Vitor Silveira Freitas: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Lucas Raimundo Bento:** Writing – review & editing, Writing – original draft, Visualization, Validation. **Vitoria Napolitano:** Methodology, Investigation. **Patrícia Perondi Anchão Oliveira:** Writing – review & editing, Supervision, Resources, Funding acquisition. **José Ricardo Macedo Pezzopane:** Writing – review & editing, Resources, Funding acquisition. **Alberto Carlos de Campos Bernardi:** Writing – review & editing, Resources, Funding acquisition. **Débora Marcondes Bastos Pereira Milori:** Writing – review & editing, Funding acquisition. **Ladislau Martin-Neto:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Funding acquisition.

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.

Acknowledgements

We acknowledge the financial support provided by the São Paulo State Research Foundation (FAPESP) through project numbers: 2023/02444-5 and 2023/18440-9. We also thank the National Laboratory of Agri-Photonics (Sisfoton/CNPq: 440226/2021–0). *V.S*·F acknowledges the Post Doc fellowship from FAPESP (grant number: 2024/14266-7).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.microc.2026.118191.

Data availability

The original contributions presented in the study are included in the article. Further inquiries can be directed to the corresponding author.

References

[1] B. Minasny, B.P. Malone, A.B. McBratney, D.A. Angers, D. Arrouays, A. Chambers, V. Chaplot, Z.S. Chen, K. Cheng, B.S. Das, D.J. Field, A. Gimona, C.B. Hedley, S. Y. Hong, B. Mandal, B.P. Marchant, M. Martin, B.G. McConkey, V.L. Mulder, S. O'Rourke, A.C. Richer-de-Forges, I. Odeh, J. Padarian, K. Paustian, G. Pan, L. Poggio, I. Savin, V. Stolbovoy, U. Stockmann, Y. Sulaeman, C.C. Tsui, T. G. Vågen, B. van Wesemael, L. Winowiecki, Soil carbon 4 per mille, Geoderma 292 (2017) 59–86, https://doi.org/10.1016/j.geoderma.2017.01.002.

[2] Y. Garosi, S. Ayoubi, M. Nussbaum, M. Sheklabadi, Effects of different sources and spatial resolutions of environmental covariates on predicting soil organic carbon using machine learning in a semi-arid region of Iran, Geoderma Reg. 29 (2022), https://doi.org/10.1016/j.geodrs.2022.e00513.

[3] R. Lal, Soil carbon sequestration to mitigate climate change, Geoderma 123 (2004) 1–22, https://doi.org/10.1016/j.geoderma.2004.01.032.

[4] A. Don, F. Seidel, J. Leifeld, T. Kätterer, M. Martin, S. Pellerin, D. Emde, D. Seitz, C. Chenu, Carbon sequestration in soils and climate change mitigation—definitions and pitfalls, Glob. Chang. Biol. 30 (2024), https://doi.org/10.1111/gcb.16983.

[5] D.A. Bossio, S.C. Cook-Patton, P.W. Ellis, J. Fargione, J. Sanderman, P. Smith, S. Wood, R.J. Zomer, M. von Unger, I.M. Emmer, B.W. Griscom, The role of soil carbon in natural climate solutions, Nat. Sustainability 3 (2020) 391–398, https://doi.org/10.1038/s41893-020-0491-z.

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

[7] A.M. de Carvalho, D.R. de Jesus, T.R. de Sousa, M.L.G. Ramos, C.C. de Figueiredo, A.D. de Oliveira, R.L. Marchão, F.P. Ribeiro, R. de A. Dantas, L. de A.B. Borges, Soil carbon stocks and greenhouse gas mitigation of agriculture in the Brazilian Cerrado—a review, Plants 12 (2023), https://doi.org/10.3390/plants12132449.

[8] J.C. de Moraes Sá, R. Lal, K. Lorenz, Y. Bajgai, C. Gavilan, A. De Oliveira Ferreira, C. Briedis, T.M. Inagaki, D.R.P. Gonçalves, J.K. Bortoluzzi, Net zero and net negative emissions in brazilian biomes by no-till system, Sci. Total Environ. 1004 (2025), https://doi.org/10.1016/j.scitotenv.2025.180720.

[9] Paulo César Teixeira, Guilherme Kangussu Donagemma, Ademir Fontana, Wenceslau Geraldes Teixeira, Manual de métodos de análise de solo, 3ª edição, Embrapa, Brasília - DF, 2017. https://www.embrapa.br.

[10] D.V. Babos, W.N. Guedes, V.S. Freitas, F.P. Silva, M.L. de L. Tozo, P.R. Villas-Boas, L. Martin-Neto, D.M.B.P. Milori, Laser-induced breakdown spectroscopy as an analytical tool for total carbon quantification in tropical and subtropical soils: evaluation of calibration algorithms, Front. Soil Sci. 3 (2024), https://doi.org/10.3389/fsoil.2023.1242647.

[11] A. Chatterjee, R. Lal, L. Wielopolski, M.Z. Martin, M.H. Ebinger, Evaluation of different soil carbon determination methods, CRC, Crit. Rev. Plant Sci. 28 (2009) 164–178, https://doi.org/10.1080/07352680902776556.

[12] A.A.G. Al-Shammari, A.Z. Kouzani, A. Kaynak, S.Y. Khoo, M. Norton, W. Gates, Soil bulk density estimation methods: a review, Pedosphere 28 (2018) 581–596, https://doi.org/10.1016/s1002-0160(18)60034-7.

[13] A.F. Fowler, B. Basso, N. Millar, W.F. Brinton, A simple soil mass correction for a more accurate determination of soil carbon stock changes, Sci. Rep. 13 (2023) 2242, https://doi.org/10.1038/s41598-023-29289-2.

[14] J. Mészáros, Z. Kovács, P. László, S. Vass-Meyndt, S. Koós, B. Pirkó, N. Szűcs-Vásárhelyi, Z. Bakacsi, A. Laborczi, K. Balog, L. Pásztor, Vis-NIR soil spectral library of the hungarian soil degradation observation system, Sci. Data 12 (2025), https://doi.org/10.1038/s41597-025-04667-9.

[15] N. Asgari, S. Ayoubi, J.A.M. Demattê, A.C. Dotto, Carbonates and organic matter in soils characterized by reflected energy from 350–25000 nm wavelength, J. Mt. Sci. 17 (2020) 1636–1651, https://doi.org/10.1007/s11629-019-5789-9.

[16] S. Naimi, S. Ayoubi, L.A.D.L. Di Raimo, J.A.M. Demattе, Quantification of some intrinsic soil properties using proximal sensing in arid lands: application of Vis-NIR, MIR, and pXRF spectroscopy, Geoderma Reg. 28 (2022), https://doi.org/10.1016/j.geodrs.2022.e00484.

[17] S. Liu, H. Shen, S. Chen, X. Zhao, A. Biswas, X. Jia, Z. Shi, J. Fang, Estimating forest soil organic carbon content using Vis-NIR spectroscopy: implications for large-scale soil carbon spectroscopic assessment, Geoderma 348 (2019) 37–44, https://doi.org/10.1016/j.geoderma.2019.04.003.

[18] R.A. Viscarra Rossel, W.S. Hicks, Soil organic carbon and its fractions estimated by visible-near infrared transfer functions, Eur. J. Soil Sci. 66 (2015) 438–450, https://doi.org/10.1111/ejss.12237.

[19] S. Nawar, A.M. Mouazen, On-line Vis-NIR spectroscopy prediction of soil organic carbon using machine learning, Soil Tillage Res. 190 (2019) 120–127, https://doi.org/10.1016/j.still.2019.03.006.

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

[21] C. Black, C. Brummit, N. Campbell, M. DuBuisson, D. Harburg, L. Matosziuk, M. Motew, G. Pinjuv, E. Smith, Improved Agricultural Land Management, 2024.

[22] B. Minasny, A.B. McBratney, G. Tranter, B.W. Murphy, Using soil knowledge for the evaluation of mid-infrared diffuse reflectance spectroscopy for predicting soil physical and mechanical properties, Eur. J. Soil Sci. 59 (2008) 960–971, https://doi.org/10.1111/j.1365-2389.2008.01058.x.

[23] C.S. Moreira, D. Brunet, L. Verneyre, S.M.O. Sá, M.V. Galdos, C.C. Cerri, M. Bernoux, Near infrared spectroscopy for soil bulk density assessment, Eur. J. Soil Sci. 60 (2009) 785–791, https://doi.org/10.1111/j.1365-2389.2009.01170.x.

[24] S. Katuwal, M. Knadel, T. Norgaard, P. Moldrup, M.H. Greve, L.W. de Jonge, Predicting the dry bulk density of soils across Denmark: comparison of single-parameter, multi-parameter, and Vis–NIR based models, Geoderma 361 (2020), https://doi.org/10.1016/j.geoderma.2019.114080.

[25] S. Dharumarajan, C. Gomez, C.G. Kusuma, R. Vasundhara, B. Kalaiselvi, M. Lalitha, R. Hegde, Prediction of soil organic carbon stock along layers and profiles using Vis-NIR laboratory spectroscopy, Catena (Amst) 257 (2025), https://doi.org/10.1016/j.catena.2025.109150.

[26] V. Allory, A. Cambou, P. Moulin, C. Schwartz, P. Cannavo, L. Vidal-Beaudet, B. G. Barthès, Quantification of soil organic carbon stock in urban soils using visible and near infrared reflectance spectroscopy (VNIRS) in situ or in laboratory conditions, Sci. Total Environ. 686 (2019) 764–773, https://doi.org/10.1016/j.scitotenv.2019.05.192.

[27] M. Allo, P. Todoroff, M. Jameux, M. Stern, L. Paulin, A. Albrecht, Prediction of tropical volcanic soil organic carbon stocks by visible-near- and mid-infrared spectroscopy, Catena (Amst). 189 (2020). https://doi.org/https://doi.org/10.1016/j.catena.2020.104452.

[28] A. Cambou, V. Allory, R. Cardinael, L.C. Vieira, B.G. Barthès, Comparison of soil organic carbon stocks predicted using visible and near infrared reflectance (VNIR) spectra acquired in situ vs. on sieved dried samples: synthesis of different studies, Soil Secur. 5 (2021), https://doi.org/10.1016/j.soisec.2021.100024.

[29] J.R.M. Pezzopane, P.P.A. de Oliveira, A. de F. Pedroso, W.L. Bonani, C. Bosi, H. B. Brunetti, R.P. Neto, A.J. Furtado, P.H.M. Rodrigues, Intercropping of tropical grassland and pigeon pea: Impact on microclimate, soil water, and forage production, Rangel. Ecol. Manag. 95 (2024) 1–10, https://doi.org/10.1016/j.rama.2024.04.005.

[30] E. Pecuária Sudeste São Carlos, in: P. Perondi Anchão Oliveira Editor Técnico (Ed.), Protocolo para quantificação dos estoques de carbono do solo da rede de pesquisa Pecus Empresa Brasileira de Pesquisa Agropecuária Embrapa Pecuária Sudeste Ministério da Agricultura, Pecuária e Abastecimento, 2014. www.embrapa.br/pecuaria-sudeste.

[31] O.A. De Camargo, A.C. Moniz, J.A. Jorge, J.M.A.D.S. Valadares, Métodos de Análise Química, Mineralógica e Física de Solos do Instituto Agronômico de Campinas, 2009, https://doi.org/10.1210/jcem-65-1-105.

[32] Y. Wang, M. Li, R. Ji, M. Wang, L. Zheng, Comparison of soil total nitrogen content prediction models based on Vis-NIR spectroscopy, Sensors (Switzerland) 20 (2020) 1–20, https://doi.org/10.3390/s20247078.

[33] S.K. Shin, S.J. Lee, J.H. Park, Prediction of soil properties using Vis-NIR spectroscopy combined with machine learning: a review, Sensors 25 (2025), https://doi.org/10.3390/s25165045.

[34] S. Wold, M. Sjostrom, L. Eriksson, PLS-regression: a basic tool of chemometrics, Chemom. Intell. Lab. Syst. 58 (2001) 109–130. www.elsevier.comrlocaterchemometrics.

[35] R.J. Vestergaard, H.B. Vasava, D. Aspinall, S. Chen, A. Gillespie, V. Adamchuk, A. Biswas, Evaluation of optimized preprocessing and modeling algorithms for prediction of soil properties using Vis-NIR spectroscopy, Sensors 21 (2021), https://doi.org/10.3390/s21206745.

[36] V. Bellon-Maurel, E. Fernandez-Ahumada, B. Palagos, J.M. Roger, A. McBratney, Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy, TrAC Trends Anal. Chem. 29 (2010) 1073–1081, https://doi.org/10.1016/j.trac.2010.05.006.

[37] B. Yu, C. Yan, J. Yuan, N. Ding, Z. Chen, Prediction of soil properties based on characteristic wavelengths with optimal spectral resolution by using Vis-NIR spectroscopy, Spectrochim. Acta A Mol. Biomol. Spectrosc. 293 (2023), https://doi.org/10.1016/j.saa.2023.122452.

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

[39] R.N. Clark, T.V.V. King, M. Klejwa, G.A. Swayze, N. Vergo, High spectral resolution reflectance spectroscopy of minerals, J. Geophys. Res. 95 (1990), https://doi.org/10.1029/jb095ib08p12653.

[40] W.P. dos Santos, C.M.P. Vaz, L. Martin-Neto, A. Anselmi, J. Tomasella, F. de Souza Costa, J.A. Albuquerque, Q. de Jong van Lier, R. Galbieri, F.J. Perina, Predicting bulk density in Brazilian soils for carbon stocks calculation: a comparative study of multiple linear regression and random forest models using continuous and categorical variables, Discov. Soil 2 (2025), https://doi.org/10.1007/s44378-025-00035-6.

[41] F.B. de Santana, A.M. de Souza, R.J. Poppi, Visible and near infrared spectroscopy coupled to random forest to quantify some soil quality parameters, Spectrochim. Acta A Mol. Biomol. Spectrosc. 191 (2018) 454–462, https://doi.org/10.1016/j.saa.2017.10.052.

[42] K. Heil, U. Schmidhalter, An evaluation of different NIR-spectral pre-treatments to derive the soil parameters c and n of a humus-clay-rich soil, Sensors 21 (2021) 1–24, https://doi.org/10.3390/s21041423.

[43] S. Nawar, A.M. Mouazen, Predictive performance of mobile Vis-near infrared spectroscopy for key soil properties at different geographical scales by using spiking and data mining techniques, Catena (Amst). 151 (2017) 118–129, https://doi.org/10.1016/j.catena.2016.12.014.

[44] P.R. Bevington, D.K. Robinson, Data Reduction and Error Analysis for the Physical Sciences, 3rd ed, 2003.

[45] D.H. Pearlshtien, E. Ben-Dor, Effect of organic matter content on the spectral signature of iron oxides across the VIS-NIR spectral region in artificial mixtures: an

example from a red soil from Israel, Remote Sens. 12 (2020), https://doi.org/10.3390/rs12121960.
[46] B. Stenberg, R.A. Viscarra Rossel, A.M. Mouazen, J. Wetterlind, Visible and Near Infrared Spectroscopy in Soil Science, 2010, https://doi.org/10.1016/S0065-2113(10)07005-7.
[47] P.R. Griffiths, J.A. De Haseth, Fourier Transform Infrared Spectrometry, 2007.
[48] C. Pasquini, Near infrared spectroscopy: a mature analytical technique with new perspectives – a review, Anal. Chim. Acta 1026 (2018) 8–36, https://doi.org/10.1016/j.aca.2018.04.004.