Discover Environment

Research

Analysis of organic and mineral nitrogen, total organic carbon and humic fractions in Ferralsols: an approach using Vis-NIR-SWIR, MIR and X-ray fluorescence spectroscopy

Bruna Coelho de Lima[1] · Carlos H. dos Santos[1] · Carlos S. Tiritan[1] · José A. M. Demattê[2] · Andres M. R. Gomez[2] · Heidy S. R. Albarracín[2] · Bruno A. Bartsch[2]

Received: 4 March 2024 / Accepted: 31 May 2024
Published online: 16 July 2024

Abstract
This work aimed to develop suitable predictive models for ammonium, nitrate, total nitrogen, total organic carbon and soil humic fractions, for Ferralsols, using Vis-NIR-SWIR, MIR and X-ray fluorescence spectroscopic techniques in conjunction with machine learning algorithms, Cubist, PLSR, Random Forest and Support Vector Machine. Chemical analyzes were carried out to determine nitrate, total nitrogen, total organic carbon and chemical fractionation of soil organic matter, as well as spectral analyzes using Vis-NIR-SWIR spectroscopy, MIR and X-ray fluorescence. The spectroscopy results were processed using RStudio v. 4.1.3, applying Cusbist, PLSR, Random Forest and Support Vector Machine machine learning algorithms to create predictive models and describe spectral curves and Pearson correlation. Of the prediction models developed for nitrogen, total organic carbon and humic fractions, the PLSR and Support Vector Machine algorithms presented the best predictive performances. The descriptive analysis of the spectra identified the main absorption bands and the location of the bands sensitive to the attributes of interest. The correlation analysis proposed that the use of Vis-NIR-SWIR, MIR and XRF spectroscopic techniques were effective in predicting the contents of nitrogen, total organic carbon and humic fractions in soil with a medium sandy texture. However, it is important to highlight that each technique has its characteristic mechanism of action, Vis-NIR-SWIR and MIR detect the element based on overtones and fundamental tones, while XRF is based on the atomic number of the elements or elemental association.

Keywords Supervised machine learning · Spectroscopic techniques · Accurate fertilization Sandy soil · Ammonium and nitrate quantification · Organic matter quantification · Soil health · Machine Learning

1 Introduction

Sustainable soil use is an indispensable practice for the longevity of tropical agricultural soils, due to the difficulty in maintaining the stock and quality of organic matter in production systems [1]. The Oxisols, as they are called in the USA by soil taxonomy [2], are considered by the Brazilian soil classification system as Ferralsols, are weathered, low-fertility soils that occupy 7.5% of the global soil, 23% of the tropics [3], and more than 60% of Brazilian soils [4]. This class of soil is made up of 1:1 clay minerals such as Kaolinite, and Al and Fe oxides such as Gibbsite, Hematite, and Goethite [5]. Due to their particularities Ferralsols present challenges for achieving sustainable management, requiring an understanding

✉ Bruna Coelho de Lima, brunacoelho94@outlook.com.br | [1]University of Oeste Paulista, Presidente Prudente, São Paulo, Brazil. [2]Luiz de Queiroz" College of Agriculture, University of São Paulo, Piracicaba, São Paulo, Brazil.

of their intrinsic properties to establish agricultural practices that preserve the environment [6]. The implementation of conservation practices such as no-till farming, and crop rotation, among others [7], results in the concentration of plant residues on the soil surface, providing benefits not only to the soil but also to water, contributing to improvements in the physical, chemical and biological properties of the system [8], especially the carbon (C) and nitrogen (N) fractions present in the soil [9]. Soil organic matter (SOM) from crop residues is composed of humified fractions, called humic substances, which characterize 85% of the total organic carbon (TOC) present in SOM [1]. TOC is the fraction of carbon stored in organic matter, which represents 58% of MOS. The fractions of MOS are relevant to environmental quality, due to their representativeness in the global carbon stock [10]. In addition to contributing to carbon storage in the soil, MOS also adds other soil nutrients such as nitrogen [11]. The N is one of the essential nutrients most required by plants and is considered extremely important in the processes of synthesizing proteins, nucleotides, nucleic acids, alkaloids, as well as chlorophyll, in addition to being a key factor in the photosynthesis cycle [12]. Furthermore the organic N, there are also inorganic mineral forms, composed of ammonium (NH_4^+), nitrogen dioxide (NO_2^-), and nitrate (NO_3^-) [13]. In 2020, Brazil consumed approximately 5.3 million tons of N [14]. This underscores the importance of nitrogen fertilization for increasing agricultural productivity. However, there is still a great deal of doubt regarding the definition of the doses to be applied, as the indiscriminate use of nitrogen fertilizer in the forms of ammonium and urea causes soil acidification [15], as well as losses through the processes of volatilization, leaching, and denitrification [16], causing losses for both the farmer and the farming system. Usually, C and N availability in the soil can be assessed using chemical methods. However, there are restrictions on the chemical procedure, as it does not consider the mineralization and immobilization of the element, which occur during the process of decomposition of organic matter [17], in addition to increasing the time it takes to release the result, in the case of using traditional methods of determination in commercial routine laboratories. Therefore, the use of remote sensing techniques geotechnologies has provided alternatives for evaluating soil attributes, optimizing time, reducing costs, and increasing the number of samples evaluated [18]. One of these is supervised machine learning. The machine learning algorithms often used to predict soil attributes are PLSR, Cubist, Random Forest, and Support Vector Machine [19]. In conjunction with machine learning, spectroscopic techniques such as Vis-NIR-SWIR, MIR, and X-ray fluorescence have been increasingly used to add to soil analysis carried out in laboratories, with applications focused on various aspects of soil Science [20, 21, 22]. The importance of associating spectroscopic techniques with the prediction of nitrogen and total organic carbon in soil classes such as Ferralsols becomes fundamental, given the costly chemical methods for predicting these components in the soil, as well as the undeniable need to assess their contents to provide adequate management and reduce nitrogen losses in the soil-atmosphere system [23]. The hypothesis of the work is that the use of spectroscopic techniques in association with supervised machine learning, allows the construction of a predictive model for the components ammonium, nitrate, total nitrogen, total organic carbon and humic fractions, for the soil class Ferralsols .The present study was conducted to develop a predictive model for ammonium, nitrate, total nitrogen, total organic carbon, and the chemical fractions of organic matter, using Vis-NIR-SWIR, MIR, and X-ray fluorescence spectroscopic techniques in association with the Cubist, PLSR, Random Forest and Support Vector Machine learning algorithms.

2 Materials and methods

2.1 Description of the site and soil collection procedures

The study area is located on the experimental farm of the University of Oeste Paulista—UNOESTE, in Presidente Bernardes—SP/Brazil, according to the geographical coordinates, south latitude 22°17′13″ and west longitude 51°40′34″ (Fig. 1). The site is home to an existing experiment, set up in 2014, consisting of a consortium system made up of the forage plant *Panicum maximum* cv. Mombasa and the legumes *Cajanus cajan* (Guandú bean) and *Macrotyloma axillare* cv. Java.

The soil is classified as Ferralsols [24], and the regional climate, according to the [25] classification, is of the Cwa type. The soil was collected to begin developing the predictive models in this area but in a fallow, unmanaged plot. Composite soil samples were taken from the 0-20 and 80-100 cm layers at two points on the plot. The samples were analyzed for chemical parameters according to the methodology of [26], and the granulometric characterization according to the methodology of the [27] (Table 1).

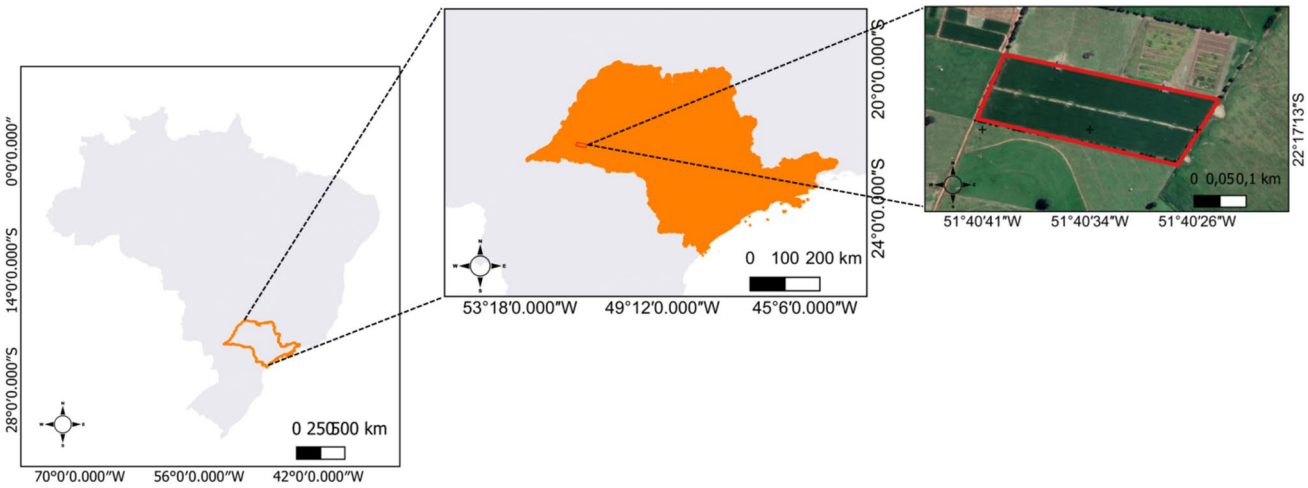

Fig. 1 Sketch of the experimental área

Table 1 Chemical and granulometric analysis of the soil in the experimental area

Depth (cm)	pH ($CaCl_2$)	O.M (g dm^{-3})	P	S	Al^{3+}	H+Al	K	Ca	Mg	SB	CEC	V (%)
			– (mg dm^{-3}) –		($mmol_c$ dm^{-3})							
0–20	5.0	17.6	4.2	2.8	0	21.1	1.8	15.8	9.4	27	48.2	56
80–100	4.1	11.7	2.0	22.7	2	21.6	0.8	3.3	4.6	8.7	30.3	28

Depth (cm)	Granulometry (g kg^{-1})			Textural class
	Sandy	Silt	Clay	
0–20	770.9	53.9	175.2	Sandy medium
80–100	706.2	63.3	230.5	Sandy medium

OM Organic matter, *SB* sum of bases, *CEC* cátion exchange capacity, *V* Base saturation

2.2 Experimental design and determination of treatments

The experimental design was entirely randomized, with two sampling depths, 0-20 and 80-100 cm of soil, and four increasing doses of 0, 100, 200, and 400 kg.ha^{1} of the mineral chemical compound ammonium nitrate, consisting of 34% N (50% N-nitric and 50% N-ammoniacal). Each depth had four doses and ten repetitions, totaling 80 samples. The respective doses of the ammonium nitrate compound were applied. The duly identified samples were then moistened to their field capacity and stored in an air-conditioned greenhouse for 30 days [28]. The soil content intended for N and CO determination was stored in a freezer at -15°C as soon as it was removed from the cups. In this way, the inorganic N content was preserved until the laboratory analysis began [29]. The plot for spectral analysis was dried in a forced ventilation oven at 45°C for 24 hours [30]. For the Vis-NIR-SWIR spectroradiometer and the XRF spectrometer, the soil samples were sieved through a 2.00 mm (9 mesh) [30, 31]. In MIR the samples were sieved through a 150µm (100 mesh) [32].

2.3 Chemical soil analysis

2.3.1 Determination of organic matter and total organic C in the soil and chemical fractionation of organic matter and quantification of carbon in the fractions

To determine the organic matter content of the soil, 2g of soil from each sample was used and sieved through a 0.250 mm (60 mesh) sieve. The procedure was carried out using the [33], adapted from [34]. To extract humic fractions, the methodology described by [33], adapted from [34]. 1g of dry soil was used in 50ml centrifuge tubes and 10ml of NaOH solution (0.1 mol ℓ^{-1}) was pipetted into each tube. The tubes were then shaken in a vertical shaker for 1 hour at a speed

of 12 rpm and left motionless for 24 hours. After 24 h, the tubes were centrifuged at 3000 revolutions for 20 minutes. The supernatant resulting from the procedure was transferred to 100 ml beakers. 10 ml of NaOH (0.1 mol ℓ^{-1}) were pipetted again into each tube, shaken and left still for 1 h. The centrifuge procedures at 3000 revolutions and supernatant removal were repeated once again. The alkaline extract contained in the bekers, resulting from the removal of the supernatant, contains the fraction of Humic and Fulvic acids, and had the pH adjusted to 2 with the H_2SO_4 solution (20%). The residue remaining in the tubes after removing the supernatant contains the Humine fraction, which is taken to the oven at a temperature of 45°C for 72 hours. The extract containing Humic and Fulvic acid, after pH adjustment, was transferred to other centrifuge tubes, remaining motionless for 18 hours for total precipitation of the humic fraction. After the time had elapsed, they were centrifuged at 3000 turns for 5 minutes. The supernatant resulting from this procedure contains the Fulvic fraction and was transferred to 50 ml volumetric flasks with its volume measured with deionized water. The precipitates at the bottom of the centrifuge tubes, Humic fraction, were added to 30 ml of NaOH (0.1 mol ℓ^{-1}) for dilution and homogenization, being transferred to 50 ml volumetric flasks, completing the volume with the solution of NaOH (0.1 mol ℓ^{-1}). I) Quantification of organic carbon in the Humic and Fulvic fractions: 5 ml of the extract from the Humic or Fulvic fraction were pipetted into 100 ml tubes for digestion. 10 ml of $K_2Cr_2O_7$ solution (0.033 mol ℓ_{-1}) was added with a volumetric pipette, then 10 ml of concentrated H2SO4 was added. The tubes were placed in the digester block at 170°C and maintained at that temperature for 30 minutes. After cooling at room temperature, 5 drops of the ferroin indicator solution were added, being titrated with the ammonium ferrous sulfate solution (0.03 mol ℓ^{-1}). The titration turning point is clear, changing from green to violet, and possibly reaching red. Under the same conditions, we made 6 blank controls, where 3 blanks were taken to the digester block along with the samples, and the other 3 remained unheated, at room temperature. Unheated control blanks are important for calculating the total dichromate lost on heating in the absence of the sample. II) Quantification of organic carbon in the Humine fraction: 0.5 g of the humine fraction was weighed and transferred to the digestion tubes. 5 ml of $K_2Cr_2O_7$ (0.167 mol ℓ^{-1}) were added with the aid of a volumetric pipette, and then 15 ml of concentrated H_2SO_4 were added. The samples were digested in a digester block at 170°C and kept at that temperature for 30 minutes. After cooling at room temperature, the contents of each tube were transferred to a 250 ml Erlenmeyer flask and deionized water was added to bring the final volume to 80 ml. Subsequently, 0.3 ml of indicator solution was added and titration occurred with ammoniacal ferrous sulfate solution (0.1 mol ℓ^{-1}). The same procedures were carried out for the white controls.

2.3.2 Determination of inorganic nitrogen

To determine the inorganic forms of N (NH4+ and NO3-) in the soil, 5 cm^3 of soil was collected from each stored sample, which was determined according to the methodology of [35].

2.3.3 Determination of total nitrogen content

To determine the total N in the soil, 1g of soil was collected from each sample and the procedure was carried out according to the methodology described by [35].

2.4 Spectroscopic analysis

2.4.1 Visible, near infrared and shortwave infrared (Vis-NIR-SWIR)

The analysis was carried out according to the methodology described by [30], in which the soil samples were dried in a forced ventilation oven at a temperature of 45°C for 24 hours. They were then ground and sieved through a 2 mm mesh sieve (9 mesh), and deposited in 5g portions in Petri dishes, leveling the surface of the sample to reduce relief. Spectral data was obtained using the FieldSpec3 spectroradiometer (Analytical Spectral Devices, Boulder, CO, USA), using wavelengths of 350-2.500 nm.

2.4.2 Middle infrared (MIR)

The analysis was carried out according to the methodology described by [32], in which the soil samples were dried in a forced ventilation oven at 45°C for 24 hours and then sieved through a 150µm mesh (100 mesh) to obtain smaller soil particles. The reflectance spectra were obtained using the Alpha Sample Compartment RT-DLaTGS ZnSe spectrometer

(Bruker Optik GmbH), equipped with an accessory for the acquisition of diffuse reflectance (DRIFT). The spectra are obtained in the range between 4000 and 400 cm^{-1} (2.500-25.000 nm) with a spectral resolution of 2 cm^{-1} and 32 scans per minute per spectrum. For calibration, a gold reference plate was used as a standard every four measurements. A 1g portion of soil was used for the analysis.

2.4.3 X-Ray fluorescence (XRF)

The analysis was carried out according to the methodology described by [31], in which the soil samples were dried in a forced ventilation oven at 45°C for 24 hours, and then sieved through 2 mm mesh (9 mesh). The analysis was carried out using the Olympus Delta Professional portable X-ray fluorescence spectrometer (Olympus, USA), which has two excitation modes. The equipment is supplied with a 50 keV silver X-ray anode and a silicon drift detector, with 2,048 channels. It comes with factory-installed calibration methods called "Soil" and "Geochem", which act independently and read various elements. The soil samples were analyzed in "Soil" mode.

2.5 Data analysis

The data was extracted and tabulated in spreadsheets, together with the respective data from the chemical analyses for determining N and COT. The machine learning algorithms used were Cubist, Partial Least Squares Regression (PLSR), Randon Forest (RF), and Support Vector Machine (SVM). These algorithms were selected due to their recurrent use in studies to develop predictive models based on MIR and NIR soil spectral libraries [36, 37]. Of the samples, 70% (equivalent to 28 samples) were randomly selected for training, and 30% (equivalent to 12 samples) for model validation. This process was carried out for the datasets referring to soil depths of 0-20 and 80-100 cm. The algorithms were run using the Caret package in the RStudio® v. 4.1.3 software. The relationship between the observed and predicted values was assessed by the coefficient of determination (R^2), the root mean square error (RMSE), and the performance ratio for the interquartile range (RPIQ). The models were classified based on R2 values, according to the methodology [38]. Models with $R^2 \geq 0.75$: well-adjusted models to accurately predict soil properties; R2 0.75 - 0.50: satisfactory models that can be improved; and R2 ≤ 0.50: non-significant models with no predictive capacity.

3 Results and discussion

3.1 Descriptive statistics of the attributes under study

Table 2 Statistical parameters for the contents of ammonium, nitrate, total nitrogen, total organic carbon and humic fractions, at a depth of 0-20 cm in the soil (attached). Table 3 Statistical parameters for the contents of ammonium, nitrate, total nitrogen, total organic carbon and humic fractions, at a depth of 80-100 cm in the soil (attached).

3.2 Evaluation of calibration and validation models for Vis–NIR-SWIR, MIR, and XRF

The calibration and validation models for the set of samples collected from the 0–20 cm depth are listed in (Table 2). The data was used to assess the prediction performance of the models, as well as the wavelengths studied. The SVM

Table 2 Statistical parameters for the contents of ammonium, nitrate, total nitrogen, total organic carbon and humic fractions, at a depth of 0-20 cm in the soil

	Ammonium	Nitrate	Total. N.	TOC	Fulvic	Humic	Humine
Max.	4.15	1.77	2.46	1.09	1.70	0.32	0.47
Min.	0.04	0.14	2.10	0.93	0.008	0.00	0.11
Mean	0.75	1.21	2.30	1.00	0.39	0.12	0.33
S.E. Mean	0.10	0.08	0.01	0.01	0.08	0.01	0.01
Std. Deviation	0.65	0.55	0.09	0.04	0.43	0.08	0.07
Variance	0.42	0.30	0.01	0	0.18	0.01	0.01
Variation coefficient	0.86	0.46	0.04	0.04	1.08	0.63	0.22

^{R2} Coefficient of determination. ^{RMSE} Root mean square error, ^{RPIQ} Performance ratio for interquartile range

Table 3 Statistical parameters for the contents of ammonium, nitrate, total nitrogen, total organic carbon and humic fractions, at a depth of 80-100 cm in the soil

	Ammonium	Nitrate	Total. N.	TOC	Fulvic	Humic	Humine
Max.	1.52	1.78	2.14	0.67	0.46	0.51	0.98
Min.	0.02	0.72	1.75	0.49	0.01	0.02	0.60
Mean	0.8	1.37	1.97	0.57	0.14	0.16	0.80
S.E. Mean	0.08	0.06	0.02	0.01	0.02	0.02	0.01
Std. Deviation	0.48	0.35	0.10	0.05	0.11	0.10	0.09
Variance	0.23	0.12	0.01	0.00	0.01	0.01	0.01
Variation coefficient	0.60	0.30	0.05	0.09	0.80	0.60	0.11

^{R2} Coefficient of determination. ^{RMSE} Root mean square error, ^{RPIQ} Performance ratio for interquartile range

regression model expressed satisfactory and adequate coefficients of determination (R^2) within the classification scale [37]. As for the wavelengths studied, Vis-NIR-SWIR showed satisfactory predictions for ammonium (R^2 0,53), nitrate (R^2 0.82) and XRF for nitrate (R2 0.53) (Table 4).

Table 4 Prediction of ammonium, nitrate, and organic matter fractions in soil samples in the 0–20 cm layer using the Cubist, PLSR, Random Forest, and Support Vector Machine learning algorithms

Cubist			
Vis–NIR-SWIR		MIR	XRF
R^2/RMSE/RPIQ			
Ammonium	0.52/3.90/0.23	0.04/5.21/0.18	0.12/6.30/0.35
Nitrate	0.38/18.16/0.86	0.23/19.51/0.87	0.11/21.85/0.52
Total nitrogen	0.19/0.04/0.58	0.16/58.59/0.89	0.09/66.71/0.61
TOC	0.09/1.15/0.5	0.09/0.87/0.72	0.10/0.99/0.61
Fulvic acid	0.16/0.54/0.58	0.20/0.45/0.51	0.09/0.64/0.41
Humic acid	0.08/0.48/0.56	0.08/0.39/0.42	0.08/0.38/0.57
Humine	0.07/0.71/0.30	0.10/0.67/0.31	0.07/0.67/0.46
PLSR			
Ammonium	0.39/4.94/0.64	0.23/5.76/0.56	0.09/4.21/0.25
Nitrate	0.76/11.94/1.12	0.06/26.52/0.47	0.15/20.19/0.85
Total nitrogen	0.10/62.74/0.49	0.05/72.18/0.46	0.29/53.79/0.47
TOC	0.14/0.89/0.52	0.04/1.08/0.58	0.17/0.84/0.50
Fulvic acid	0.30/0.54/0.54	0.13/0.54/0.60	0.10/0.43/0.68
Humic acid	0.11/0.30/0.64	0.02/0.49/0.31	0.04/0.36/0.74
Humine	0.13/0.54/0.30	0.08/0.65/0.32	0.19/0.58/0.43
Random forest			
Ammonium	0.42/4.05/0.26	0.11/4.97/0.27	0.34/5.63/0.40
Nitrate	0.32/17.69/0.90	0.14/23.96/0.79	0.26/19.82/0.84
Total nitrogen	0.22/44.30/0.79	0.08/54.30/0.47	0.03/56.76/0.49
TOC	0.07/0.76/0.61	0.13/0.81/0.75	0.02/0.58/0.62
Fulvic acid	0.09/0.59/0.33	0.10/0.43/0.47	0.17/0.38/0.62
Humic acid	0.11/0.30/0.64	0.17/0.41/0.61	0.06/0.43/0.63
Humine	0.13/0.52/0.51	0.03/0.45/0.56	0.08/0.63/0.52
Support vector machine			
Ammonium	0.53/3.82/0.70	0.03/6.60/0.16	0.25/5.53/0.19
Nitrate	0.82/9.66/1.49	0.28/7.82/0.63	0.53/15.23/0.82
Total nitrogen	0.35/9.95/0.95	0.25/5.52/0.69	0.07/0.59/0.82
TOC	0.17/0.88/0.59	0.14/1.01/0.33	0.21/0.67/0.73
Fulvic acid	0.34/0.44/0.71	0.05/0.54/0.50	0.08/0.53/0.51
Humic acid	0.06/0.54/0.40	0.03/0.44/0.56	0.16/0.34/0.58
Humine	0.14/0.88/0.29	0.08/0.63/0.51	0.06/0.50/0.46

^{R2} Coefficient of determination. ^{RMSE} Root mean square error, ^{RPIQ} Performance ratio for interquartile range

Evaluating the models for the 80–100 cm depth (Table 5), the PLSR and SVM models were only slightly better at predicting nitrate and ammonium. Among the wavelengths, only Vis-NIR-SWIR showed satisfactory values for the elements nitrate (R^2 0.68) in PLSR, and ammonium (R^2 0.50) in SVM.

In Table 6, the data obtained at depths of 0-20 and 80-100 cm were statistically evaluated together, with the aim of making the predictive models robust and verifying the values of each attribute based on this premise.

In the 0-20 cm soil layer (Table 4), the SVM regression model showed consistent validations compared to the other machine learning models. The Vis-NIR-SWIR wavelength predicted ammonium ($R^2$0.53) and nitrate (R^2 0.82), and XRF ($R^2$0.53) . The same behavior was seen for [39], integrating hyperspectral images and machine learning techniques for mapping nitrogen fractions in the soil. In the study, the authors observed that the SVM model expressed satisfactory results for total nitrogen, ammonium, and nitrate (R^2 0.94, 0.70, and 0.82) respectively, compared to the PLSR and RF methods [39]. In the 80-100cm layer (Table 5), the predictive performance of the models was significant for the PLSR and SVM algorithms. The Vis-NIR-SWIR spectrum predicted nitrate (R^2 0.68) in PLSR and ammonium (R^2 0.50) in SVM. TOC and the chemical fractions of organic matter had low coefficients of determination in both soil layers (0-20 and 80-100 cm) (Tables 4 and 5). In the literature, this may be linked to factors such as temperature, rainfall, anthropogenic activity, and

Table 5 Prediction of ammonium, nitrate, and organic matter fractions in soil samples in the 80–100 cm layer, using the Cubist, PLSR, Random Forest, and Support Vector Machine learning algorithms

Cubist			
Vis–NIR-SWIR		MIR	XRF
R^2/RMSE/RPIQ			
Ammonium	0.32/10.79/0.20	0.29/12.09/0.19	0.07/13.85/0.12
Nitrate	047/1455/159	0.14/19.15/0.60	0.10/21.32/1.02
Total nitrogen	0.13/25.31/0.50	0.17/26.14/0.79	0.07/32.16/0.62
TOC	0.10/0.51/0.62	0.10/0.57/0.48	0.10/0.56/0.54
Fulvic acid	0.13/1.00/0.37	0.07/103/0.28	0.06/1.07/0.35
Humic acid	0.18/0.78/0.40	0.10/0.76/0.38	0.07/0.93/0.22
Humine	0.10/2.80/0.44	0.06/2.26/0.64	0.06/2.92/0.49
PLSR			
Ammonium	0.49/8.93/0.37	0.27/10.87/0.23	0.16/18.46/0.17
Nitrate	0.68/12.83/1.67	0.09/23.12/0.50	0.08/35.14/0.44
Total nitrogen	0.23/23.32/0.63	0.08/35.14/0.44	0.07/92.06/0.20
TOC	0.13/0.52/0.57	0.08/0.63/0.33	0.03/0.57/0.52
Fulvic acid	0.34/1.03/0.57	0.04/1.02/0.28	0.18/0.96/0.44
Humic acid	0.04/1.43/0.20	0.09/0.83/0.40	0.06/0.70/0.39
Humine	0.11/3.07/0.41	0.13/2.54/0.43	0.10/1.93/0.61
Random forest			
Ammonium	0.37/10.43/0.30	0.18/10.35/0.26	0.15/11.46/0.21
Nitrate	0.42/15.29/1.42	0.10/18.16/1.07	0.09/18.70/0.72
Total nitrogen	0.33/23.71/0.69	0.10/24.46/0.80	0.07/25.70/0.52
TOC	0.05/0.53/0.46	0.16/0.40/0.99	0.09/0.36/0.67
Fulvic acid	0.04/0.72/0.54	0.08/0.81/0.61	0.12/1.09/0.33
Humic acid	0.05/0.60/0.47	0.04/0.88/0.24	0.13/0.62/0.34
Humine	0.05/3.28/0.32	0.03/2.42/0.45	0.07/1.47/0.62
Support vector machine			
Ammonium	0.50/9.21/0.40	0.18/10.85/0.22	0.05/16.63/0.50
Nitrate	0.44/12.92/0.71	0.26/18.21/1.07	0.11/21.13/0.44
Total nitrogen	0.06/28.94/0.45	0.06/25.04/0.78	0.10/34.40/0.06
TOC	0.22/0.71/0.33	0.02/0.58/0.40	0.06/34.38/0.22
Fulvic acid	0.05/0.87/0.40	0.03/0.93/0.43	0.14/0.88/0.05
Humic acid	0.21/0.83/0.44	0.06/1.05/0.38	0.01/0.92/0.21
Humine	0.04/3.24/0.35	0.06/2.36/0.51	0.11/3.22/0.04

[R2] Coefficient of determination. [RMSE] Root mean square error, [RPIQ] Performance ratio for interquartile range

Table 6 Prediction of ammonium, nitrate, and organic matter fractions in soil samples in the 0-20 and 80-100cm layer using the Cubist, PLSR, Random Forest, and Support Vector Machine learning algorithms

Cubist			
Vis–NIR-SWIR		MIR	XRF
R^2/RMSE/RPIQ			
Ammonium	0.05/9.71/0.21	0.24/9.82/0.17	0.03/10.84/0,18
Nitrate	0.36/16.26/1.13	0.10/20.57/0.76	0.03/21.90/0.94
Total nitrogen	0.56/48.36/0.74	0.62/46.50/0.75	0.05/74.27/0.40
TOC	0.86/1.2/2.63	0.93/0.98/3.1	0.07/3.59/0.81
Fulvic acid	0.45/0.67/1.1	0.47/0.69/1.48	0.07/0.95/0.89
Humic acid	0.05/1.04/0.22	0.04/0.49/0.45	0.04/1.09/0.22
Humine	0.51/2.12/0.93	0.54/1.92/0.77	0.04/3.22/0.43
PLSR			
Ammonium	0.31/0.78/0.11	0.32/7.94/0.25	0.07/14.28/0.12
Nitrate	0.72/12.56/1.80	0.41/16.56/1.18	0.04/19.35/0.94
Total nitrogen	0.45/54.30/0.71	0.65/45.79/0.84	0.07/65.86/0.40
TOC	0.94/0.83/3.37	0.94/0.86/3.50	0.20/3.08/1.07
Fulvic acid	0.33/0.89/1.05	0.44/0.78/1.04	0.08/1.12/0.81
Humic acid	0.06/0.67/0.24	0.10/0.63/0.43	0.04/0.56/0.51
Humine	0.56/1.92/0.57	0.67/1.82/1.81	0.10/3.39/0.33
Random forest			
Ammonium	0.10/9.82/0.16	0.01/10.33/0.24	0.02/7.36/0.20
Nitrate	0.11/20.04/1.10	0.17/17.87/0.93	0.15/18.10/1.19
Total nitrogen	0.61/44.99/0.80	0,28/63.56/0.78	0.36/52.86/0.61
TOC	0.96/0.68/4.90	0.86/1.22/2.45	0.82/1.40/2.40
Fulvic acid	0.26/0.75/1.26	0.35/0.71/1.29	0.35/0.77/0.95
Humic acid	0.05/0.54/0.41	0.04/0.79/0.32	0.06/0.50/0.52
Humine	0.64/1.61/0.66	0.60/1.79/0.57	0.45/2.23/0.50
Support vector machine			
Ammonium	0.38/8.17/0.20	0,26/9.37/0.26	0.32/7.29/0.21
Nitrate	0.60/11.75/1.42	0.33/17.34/0.97	0.24/20.97/1.03
Total nitrogen	0.52/55.29/0.60	0.68/39.03/1.02	0.21/62.24/0.58
TOC	0.97/0.59/5.16	0.95/0.80/4.15	0.76/1.90/1.58
Fulvic acid	0.49/0.70/1.38	0.28/0.90/1.10	0.22/0.82/1.15
Humic acid	0.04/0.63/0.36	0.00/1.16/0.19	0.05/0.53/0.44
Humine	0.49/2.04/1.15	0.70/1.60/0.70	0.15/3.14/0.77

^{R2} Coefficient of determination. ^{RMSE} Root mean square error, ^{RPIQ} Performance ratio for interquartile range

soil texture, which favor changes in the soil's organic carbon content [40]. The different predictive performances of the models may be related to the inherent peculiarities of the elements and the machine learning algorithm. Each statistical method has its advantages and disadvantages, which must be considered according to the purpose of the research and the data [39]. It is therefore necessary for all research to test the different modeling methods to determine which model is the most appropriate [41]. The machine learning algorithms applied to data from depths 0-20 and 80-100cm together (Table 6) showed that all validation models presented good predictive performance for the attributes nitrate, total nitrogen, TOC and humine, indicating that the combination of data from both depths resulted in satisfactory validations for these variables. In a study [42], state that increasing the number of samples to more than 50, to develop validation models for COT and NT, promotes not only greater precision but also lower prediction errors than in models with fewer of samples. The Vis-NIR-SWIR spectrum showed prediction for nitrate, total nitrogen and TOC, while MIR predicted total nitrogen, TOC and the humine fraction (Table 6). Some studies report that the MIR spectrum achieved better results for pH, SOM and sand content than the Vis-NIR spectrum [43, 44, 45]. On the other hand, for attributes such as organic carbon, CEC and total nitrogen, MIR does not stand out over Vis-NIR, as some characteristic vibrations of these elements can be masked due to the strong absorption of minerals from the soil [82]. For fluorescence X-ray, was obtained (R^2 0.76) for COT (Table 6). In a rapid analysis of organic matter with EDXRF and multivariate analysis, [46] achieved R^2 of 0.60 for TOC

validation in the PLSR model. This proposes that SVM machine learning also has the capability for prediction of soil total organic carbon with XRF spectrum. The non-linear SVM model allows it to surpass linear models such as PLSR, due to its capacity and flexibility in handling non-linear information, specifically in predictions of high variability soil attributes [47]. The Vis-NIR-SWIR graphs for the 0-20 and 80-100 cm layers (Fig. 2 a and b) showed the regions normally related to soil spectroscopy diagnostics [48].

3.3 Descriptive analysis of Vis–NIR-SWIR, MIR and XRF spectra

3.3.1 Vis-NIR-SWIR

The 0-20 and 80-100 cm layers of soil from a Ferralsols showed high reflectance in the Vis-NIR-SWIR wavelength spectral regions (Fig. 2 a) and b). It is possible to observe humic and fulvic acid absorption features at 400 nm, COS reflectance peaks at 410 and 550 nm and humine absorption features at 460 and 570 nm. [49, 50]. Fulvic acid and nitrate absorption features were identified at 700 and 760 nm. In the 1400 nm region, there is an absorption feature for water, and at 1442 nm, a reflectance peak for nitrate [51]. In the spectral region of 1800 and 1870 nm, absorption features for humic acid (1850 nm) and COS (1870 nm) are observed. In the 1900 nm absorption feature, water is determined. The inorganic nitrogen nitrate and ammonium can be seen in the spectral region of 2088, 2115, and 2120 nm respectively [52]. Further on, there is an absorption feature belonging to humic acid at 2150 nm and a reflectance peak of the kaolinite mineral located at 2200 nm [53]. Absorption features referring to total nitrogen were found at 2250, 2330, and 2430 nm [54]. In

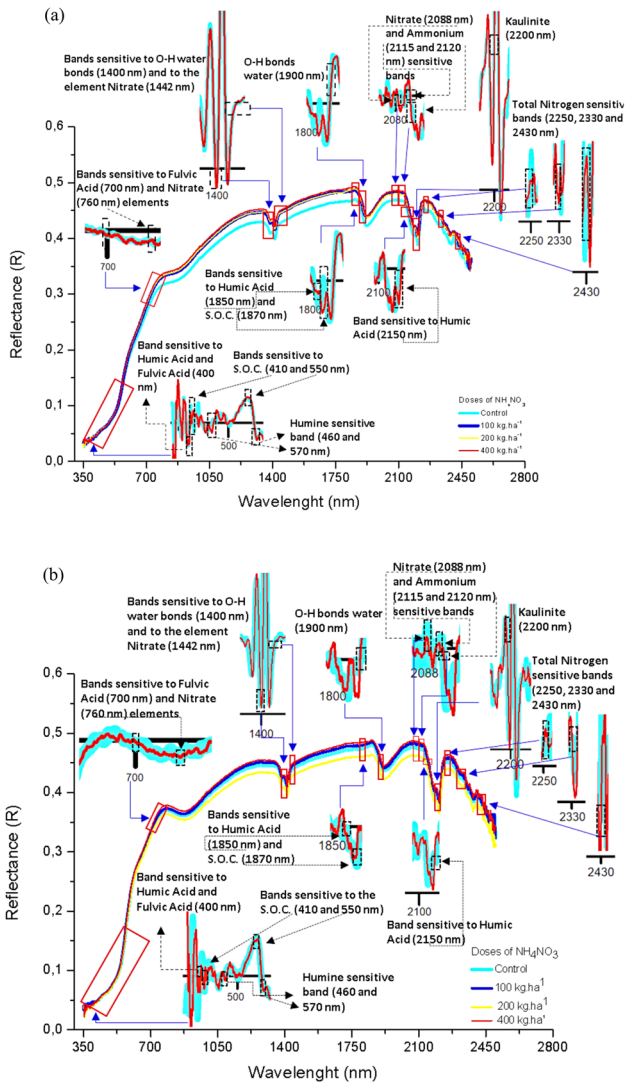

Fig. 2 Vis–NIR-SWIR spectrum: **a** 0–20 cm layer; **b** 80–100 cm soil layer

the visible region (400-700 nm), we observed bands associated with humic and fulvic acids, COS, and humine (Fig. 2 a and b), according to [55], these bands are related to MO and iron oxides, which are highly relevant for the determination of humic compounds and carbon. From 700 to 1000 nm, there are bands related to the elongation of C-H bonds, involving relevant information about humic substances as well as nitrogen fractions [55]. At wavelengths of 1400 nm, 1900 nm, and 2200 nm, important absorption features occur (Fig. 2 a and b), indicating the presence of water and 1:1 (Kaolinite) and 2:1 clay minerals (which can be different types of clay minerals, for example, illite, montmorillonite, and vermiculite) [56]. Still in the 1400 to 1900 nm range, we obtained a reflectance peak referring to nitrate, according to the author [57], there is a strong occurrence of amine groups (-R-NH_2), primary amides (-RC=H-O-NH_2), and secondary amides (-R-C=H-O-NH-R) present in aliphatic chains in the soil, which characterizes the possible presence of nitrate in this spectrum band. The spectrum shows absorption features for humic acid at 1850 and for CO at 1870nm. From 1400 to 1870nm, vibrations occur in the molecular bonds between C-H, N-H, O-H, and S-H [58] (Fig. 2 a and b). In this spectral range, there are differentiations between each humic fraction [59]. The presence of ammonium was determined in the 2115 and 2120nm bands, according to the authors [60], ammonium can be found in 2:1 minerals such as illite and montmorillonite. The total nitrogen absorption features seen in bands 2250, 2330, and 2430 are bands associated with soil water content (O-H bonds) and total carbon [61] (Fig. 2 a and b). Thus, due to the inherent link between total carbon and total nitrogen [62], the total nitrogen content in the soil can be indirectly predicted and is highly related to these wavelengths [54].

3.3.2 MIR

It can be seen that at wavenumbers of 3690 and 3620 cm^{-1}, there are peaks of reflectance of the clay mineral Kaolinite, total nitrogen at 3560 and 3151 cm^{-1}, and a water absorption feature at 3400 cm^{-1} [63, 64, 65] (Fig. 3 a and b). At 3152 cm^{-1} there is an ammonium reflectance peak and in the spectral range from 2960 to 2870 cm^{-1}, there are MOS absorption

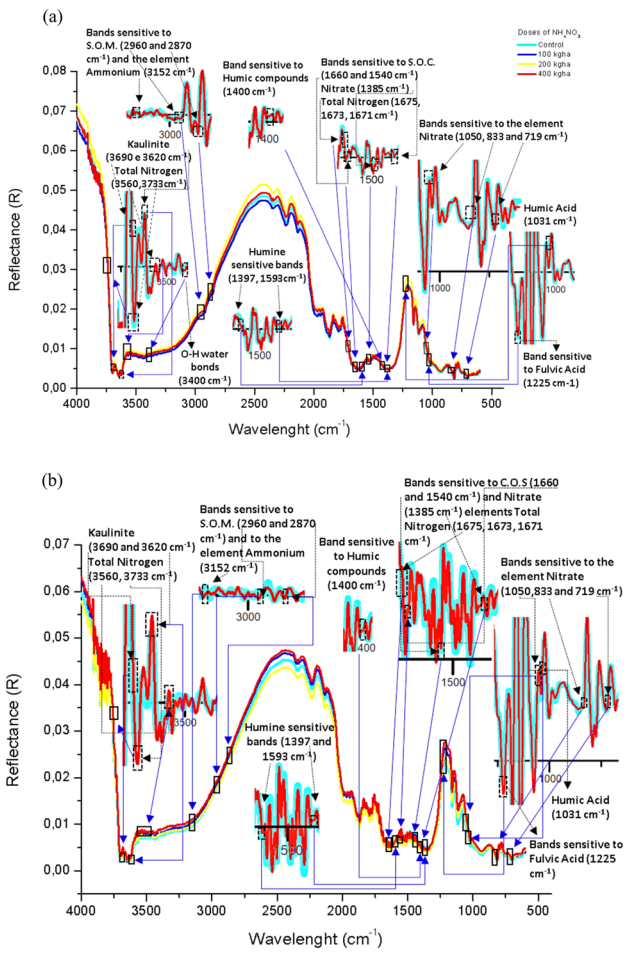

Fig. 3 MIR spectrum: **a** 0–20 cm layer; **b** 80–100 cm soil layer

features [66, 67]. Bands related to humic compounds were determined at 1400 and 1031 $cm^{-1}$with peaks in reflectance [68, 69]. The COT element showed absorption features at 1660 and 1540 cm^{-1}, next to the nitrate element at 1385 cm-1, and total nitrogen at (1675, 1673, and 1671 cm^{-1}) [66, 70]. The 1050, 833, and 719 cm^{-1} spectral bands determined the nitrate element through reflectance peaks [66]. Fulvic acid showed absorption features at 1225 cm^{-1} and humine showed reflectance peaks at 1397, 1593 cm^{-1} [69].In MIR (Fig. 3 a and b), reflectance peaks of the kaolinite mineral were identified at 3690 and 3620 cm^{-1}, according to the authors [71], these peaks are caused by O-H bond stretching vibrations in kaolinite and 2:1 aluminosilicates such as illite and smectite. In the 2960 to 3152 cm^{-1}range, bands sensitive to MOS and nitrate were found, possibly due to the C-O, C=O, C-C, C-H, O-H, N-H, N=C, and S-H bonds in this spectral region, making up the soil's organic matter and manipulating its interactions with other soil constituents such as minerals, bacteria, and water [67]. The reflectance peaks of the humic compounds at 1400 and 1030 $cm^{-1}$are attributed to the CH_2 and CH_3 bonds and the symmetrical elongation of the COO band [72]. Organic carbon showed features at 1660 and 1540 cm^{-1}, attributed to C=O stretching of amide groups and stretching of C=N bonds [58]. The spectral region between 1740 and 1600 cm^{-1} can be characterized by the determination of hydrophilic organic compounds, sensitive to soil organic carbon and total nitrogen [73] , explaining the bands 1675, 1673, and 1671 cm^{-1}. A feature was observed at 1385 $cm^{-1}$for NO_3^-. According to the literature, the spectral regions from 1385 to 1050 cm^{-1} are defined as C-N molecular vibrations in amine molecules, which are highly favorable for determining nitrogen in the soil [74]. The same occurs in the spectrum between 1050, 800, and 500 cm^{-1}, a range known for multiple mineral vibrations, containing particularities of aromatic amines C-H, and N-H [75] which contributed to the prediction of nitrate in the 833 and 719 cm^{-1} bands.

3.3.3 XRF

In the 0-20 and 80-100 cm soil layers (Fig. 4 a and b), the XRF spectrum showed the intensity of the energy emitted (photon count per second), through the electronic transition of the elements present in the soil. Eleven elements were determined and the Compton and Thomson scattering peaks (Fig. 4 a and b). The lowest intensity peaks were found

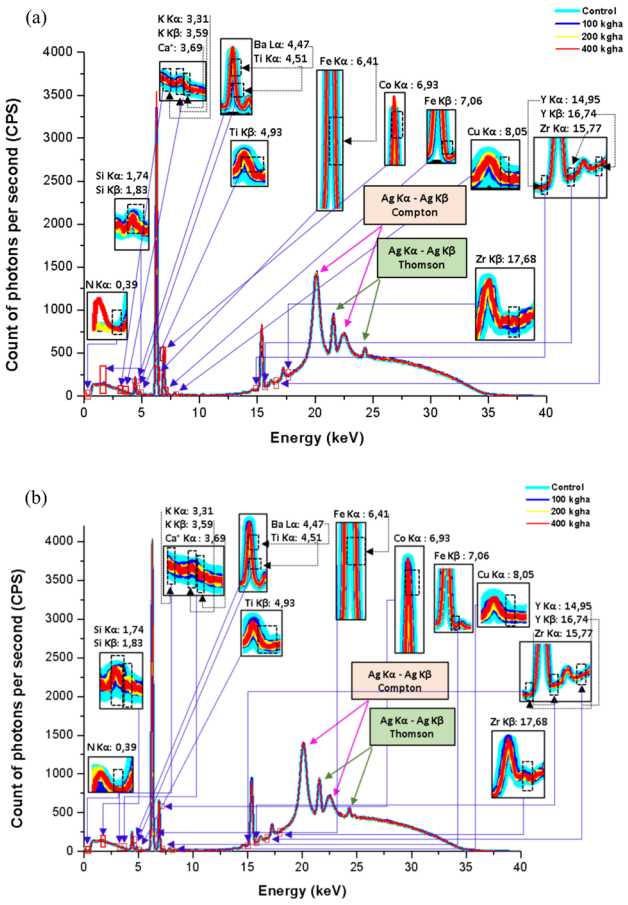

Fig. 4 XRF spectrum: **a** 0–20 cm layer; **b** 80–100 cm layer of soil

in the elements N, K, Ca+, Ti kα, Fe kβ, Cu kα, Y, and Zr [76, 77, 78, 79]. More intense peaks were seen for Si, Ba, Ti kβ, Fe kα e Co [77, 78, 79]. In XRF (Fig. 4 a and b), according to [80], significant levels of Al, Si, and Fe are highly related to the mineralogical composition of the soil. As Si can be determined in quartz (SiO_2) and muscovite ($KAl_2(AlSi_3O_{10})(OH)_2$), both predominant in the sand fraction, and also in kaolinite ($Al_4(Si_4O_{10})(OH)_8$), present in the clay fraction, Al can be seen in gibbsite ($Al(OH)_3$), as well as in muscovite and kaolinite [18]. The Fe element is usually associated with the presence of iron oxides such as hematite (Fe_2O_3), goethite (FeO(OH)), and magnetite (Fe_3O_4) [80]. The XRF is not able to determine light elements [96], such as Na, C, N, H, and Li), due to their stable electronic aspects and low fluorescent energy [81]. However, C and N can be measured indirectly due to elemental interactions. TOC interacts with oxides such as Fe, Al, Ti, Mn, and Pb, among others [82]. The total nitrogen interacts elementally with Al, Si, and Zn [83]. The XRF spectrometer has been used for successful predictions of soil physical and chemical properties such as sand, silt, clay content, CEC and base saturation, just through correlation with elements determined by the spectrum [84].

3.4 Pearson's correlation Vis-NIR-SWIR, MIR and XRF

3.4.1 Vis-NIR-SWIR

The correlation between the Vis-NIR-SWIR spectral data and the contents of ammonium, nitrate, total nitrogen, total organic carbon, and the humic fractions of organic matter was evaluated using Pearson's correlation coefficient for each soil layer (0–20 and 80–100 cm) (Fig. 5a and b). Coefficients above 0.7 indicate a very close relationship. A range of 0.4 to 0.7 indicates that the relationship is close. The range from 0.2 to 0.4 indicates that the relationship is normal [85]. It can be seen that for the 0–20 cm layer (Fig. 5 a), the ammonium content showed no correlations with the 2115 and 2120 nm wavelengths. Nitrate showed close positive correlations with the 760, 1442, and 2088 nm spectral bands. The total nitrogen content did not correlate with the 2250 and 2330 nm wavelengths, with only one positive correlation considered normal, at 2430 nm. The TOC content showed a close correlation with the 410 wavelength in the visible spectrum, and no correlation with the 550 and 1870 nm bands. The fulvic fraction showed no correlation with the 400 and 700 nm spectral regions. The same was seen for humic acid with the 400, 1850, and 2150 nm bands. In contrast, the humine fraction showed very close correlations with the 460 and 570 nm wavelengths. For the 80–100 cm layer (Fig. 5b), only the ammonium content showed correlations that were considered close, to the 2115 and 2120 nm bands of the Vis-NIR-SWIR spectrum. The other attributes such as nitrate, total nitrogen, total organic carbon, and humic fractions showed no correlations with their respective bands.

In (Fig. 5a and b), according to [86], soil is a complex mixture of numerous chemical constituents, which differ in physical state and also in terms of particle size, soil aggregation, and water content. Another important factor is the interaction of electromagnetic energy with the soil and its constituents. In organic compounds (organic matter), the main chemical bonds are OH, NH, CH, CC, CN, and others. In mineral compounds, the bonds are between Al–OH and Si–OH, present in clays. The absorbance in the NIR region is due to the harmonic tones, making the fundamental bands weaker, more overlapping, and less distinguishable from each other than, for example, in the MIR spectrum [87]. This argument may explain the low correlation of the elements under study with the Vis-NIR-SWIR spectrum in the 80–100 cm layer, given the depth and constituents of the soil.

3.4.2 MIR

The correlation between the Vis-NIR-SWIR spectral data and the contents of ammonium, nitrate, total nitrogen, total organic carbon, and the humic fractions of organic matter was evaluated using Pearson's correlation coefficient for each soil layer (0–20 and 80–100 cm) (Fig. 6a and b). Coefficients above 0.7 indicate a very close relationship. A range of 0.4 to 0.7 indicates that the relationship is close. The range from 0.2 to 0.4 indicates that the relationship is normal [85]. It can be seen that for the 0–20 cm layer (Fig. 6a), the ammonium content showed no correlations with the 2115 and 2120 nm wavelengths. Nitrate showed close positive correlations with the 760, 1442, and 2088 nm spectral bands. The total nitrogen content did not correlate with the 2250 and 2330 nm wavelengths, with only one positive correlation considered normal, at 2430 nm. The TOC content showed a close correlation with the 410 wavelength in the visible spectrum, and no correlation with the 550 and 1870 nm bands. The fulvic fraction showed no correlation with the 400 and 700 nm spectral regions. The same was seen for humic acid with the 400, 1850, and 2150 nm bands. In contrast, the humine fraction showed very close correlations with the 460 and 570 nm wavelengths. For the 80–100 cm layer (Fig. 6b), only the ammonium content showed correlations that were considered close, to the 2115 and 2120 nm bands of the

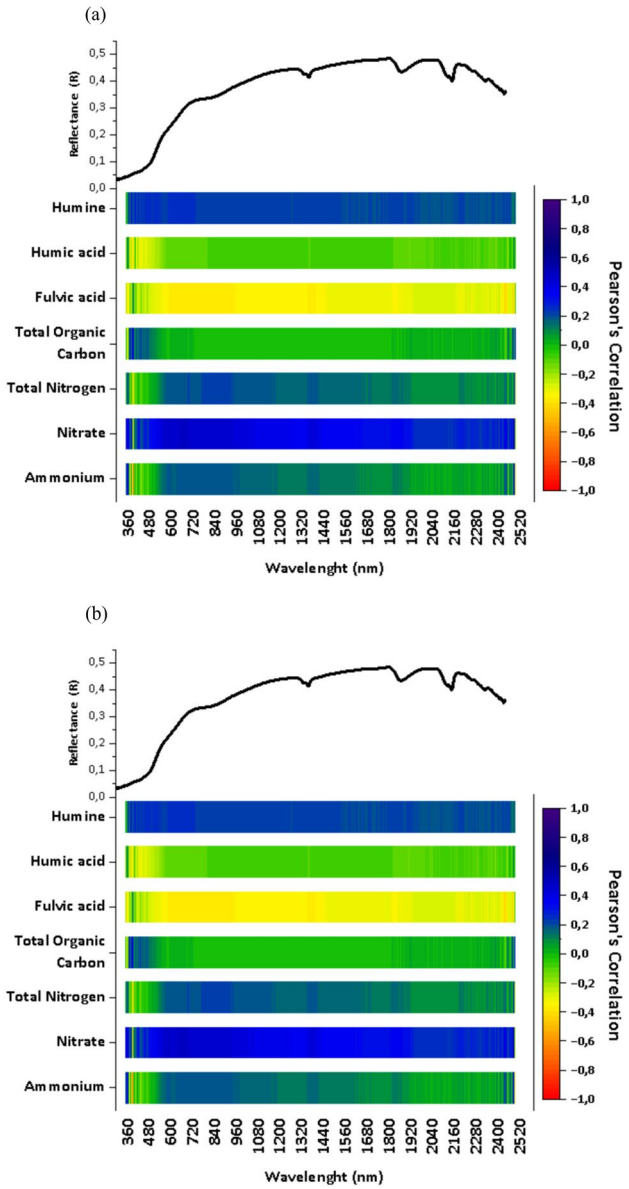

Fig. 5 Pearson's correlation coefficient between the Vis-NIR-Swir wavelength and the attributes Ammonium, Nitrate, Total Nitrogen, Total Organic Carbon, Fulvic and Humic Acids, and Humine for the layers **a** 0–20 cm; **b** 80–100 cm of soil. Significant values were. Significant values were displayed in a color gradient: red (negative correlations) and blue (positive correlations). Intense correlations (darker color)

Vis-NIR-SWIR spectrum. The other attributes such as nitrate, total nitrogen, total organic carbon, and humic fractions showed no correlations with their respective bands.

In MIR (Fig. 6a and b), according to [88], the MIR spectrum is more sensitive to the organic and mineral constituents of the soil, due to the fundamental molecular vibrations that are absorbed at the specific wavenumber of the electromagnetic radiation, which makes it clearer and more precise than the visible and near-infrared region (Vis-NIR). Strong correlations between the different carbon and nitrogen fractions were also obtained by [87]. High correlations of the MIR wavenumber were observed with ammonium, nitrate and total nitrogen, as well as with total organic carbon and humic fractions of organic matter [89]. Found a good correlation between MIR, N and C. The authors add that generally, the good correlation between total nitrogen and total organic carbon are similar, due to the direct correlation of N with soil organic matter and thus, frequently associated to total soil organic carbon [90].

3.4.3 XRF

According to the correlation for the 0–20 cm soil layer (Fig. 7a, it can be seen that the ammonium element showed a close correlation with the energy levels of the elements N kα at 0.39 keV, Ba Lα (4.47), Ti kα (4.51), Fe kα (6.41), Fe kβ (7.06), Co kα (6.93) and Zr kα (15.77). The nitrate content of the soil correlated with N kα (0.39), Fe kα (6.41), Fe kβ (7.06),

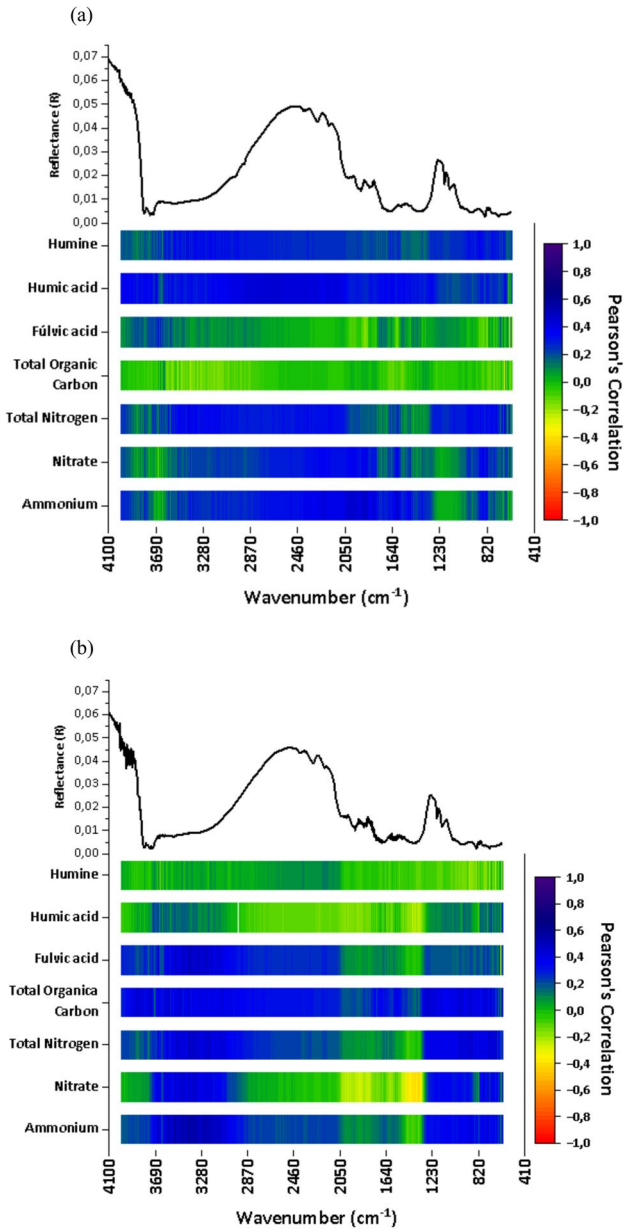

Fig. 6 Pearson's correlation coefficient between the MIR wavenumber and the attributes Ammonium, Nitrate, Total Nitrogen, Total Organic Carbon, Fulvic Acids, Humic Acids, and Humine for the layers **a** 0–20 cm; **b** 80–100 cm of soil. Significant values were displayed in a color gradient: red (negative correlations) and blue (positive correlations). Intense correlations (darker color)

and Co kα (6.93). The total nitrogen content correlated positively with N kα (0.39), Fe kα (6.41), and Co kα (6.93). The total organic carbon and the fractions fulvic acid and humine showed no correlation with any element determined in the XRF spectrum. The humic acid content showed a positive correlation with Ba Lα (4.47), Ti kα (4.51), Fe kα (6.41), Fe kβ (7.06), Co kα (6.93). For the 80–100 cm layer of soil (Fig. 7b), the elements ammonium and nitrate showed low correlations with the elements obtained in the XRF spectrum. However, the content of total nitrogen and total organic carbon showed a correlation only with N kα (0.39). Of the humic fraction, only humic acid showed a correlation close to the wavelength of the Fe kα element (6.41).

Observing (Fig. 7a and b), it can be seen that both nitrogen and carbon and humic fractions showed weak correlations with the XRF spectrum. According to [91], X-rays are strongly dependent on the atomic number of the elements. Light elements, that is, elements with atomic numbers below 12, are difficult to determine because they have low emission energies. The elements under study, nitrogen and carbon, have atomic numbers 7 and 6, respectively, which promotes low emission energy, compromising the determination in the spectrum. However, it is possible to use the elemental content quantified in XRF spectra to estimate soil properties, which are somehow associated with them [92].

In (Fig. 7a and b), it can be seen that most of the variables under study presented approximate correlations with the energy level equivalent to that of the element Fe kα (6.41), Fe kβ (7.06). In most soils, except in flooded areas, most organic C and N

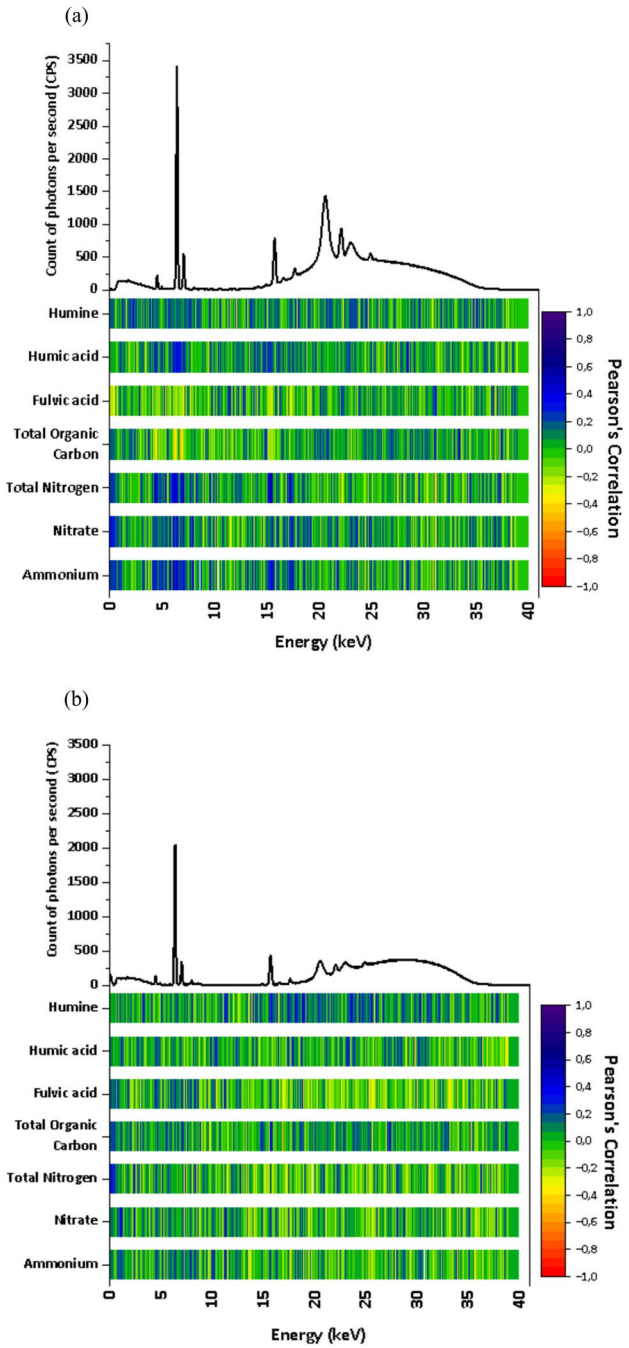

Fig. 7 Pearson's correlation coefficient between the XRF energy level and the attributes Ammonium, Nitrate, Total Nitrogen, Total Organic Carbon, Fulvic, Humic, and Humine Acids for layers **a** 0–20 cm; **b** 80–100 cm of soil. Significant values were displayed in a color gradient: red (negative correlations) and blue (positive correlations). Intense correlations (darker coloring)

fractions are strongly related to soil minerals [93]. TOC can interact with elemental oxides such as Fe and Al [82]. Nitrogen, when it percolates through the soil, can be retained in lower horizons where Fe and other oxides accumulate, associating with specific metallic phases. According to [94] iron oxides can camouflage other soil constituents [95]. Confirmed that soils with high levels (> 4%) of iron oxides can mask the spectral manifestation of soil organic matter, and consequently other elements associated with it, such as total organic carbon.

4 Conclusion

Of the prediction models developed for nitrogen, total organic carbon and soil humic fractions, the PLSR and Support Vector Machine algorithms presented the best predictive performances. The descriptive analysis of the spectra identified the main absorption bands and the location of the bands sensitive to the attributes of interest. The correlation analysis proposed that the use of Vis-NIR-SWIR, MIR and XRF spectroscopic techniques were effective in predicting the contents of nitrogen, total organic carbon and humic fractions in soil with medium sandy texture. However, it is important to highlight that each technique has its characteristic mechanism of action, Vis-NIR-SWIR and MIR detect the element based on harmonics and fundamental tones, while XRF is based on the atomic number of the elements or elemental association.

Author contributions Conceptualization: BCL, CHS, CST, JAMD. Methodology: JAMD. Fieldwork and sampling: BCL. Laboratory analysis: BCL, AMRG, HSRA, BAB. Writing—original draft preparation: BCL. Writing—review and editing: BCL, JAMD. Supervision: JAMD.

Funding The research was funded by the Coordination for the Improvement of Higher Education Personnel-Brazil (CAPES) Code-001. Support from the University of Oeste Paulista (UNOESTE), and the Geotechnologies in Soil Science Group (GeoCis).

Data availability Data and materials will be available on request.

Declarations

Ethics approval and consent to participate No tests were performed on humans or animals when conducting this study. All authors have read, understood, and have complied as applicable with the statement on "Ethical responsibilities of Authors" as found in the Instructions for Authors.

Consent for publication The final version of the manuscript was reviewed and approved by all authors.

Competing interests The authors declare no competing interests.

References

1. Sales RP, Pegoraro RF, Portugal AF, et al. Organic matter fractions of an irrigated oxisol under no-till and conventional tillage in the Brazilian Semi-Arid Region. Rev Caatinga. 2017;30:303–12. https://doi.org/10.1590/1983-21252017v30n205rc.
2. Soil Survey Staff. Keys to soil taxonomy, 13th edition. USDA natural resources conservation service. 2022.
3. Michéli E, Schád P, Spaargaren O, et al. A framework for international classification, correlation and communication - Worldwide reference base for soil resources. 2nd ed. FAO: Roma; 2006.
4. Santos HG, Jacomine PKT, Anjos LHC, Oliveira VA, Lumbreras JF, Coelho MR, Almeida JA, Araujo Filho JC, de Oliveira JB, Cunha TJF. Brazilian soil classification system. 5th ed. Brasilia: EMBRAPA; 2019.
5. Schaefer CEGR, Fabris JD, Ker JC. Minerals in the clay fraction of Brazilian Latosols (Oxisols): a review. Clay Miner. 2008;43:137–54. https://doi.org/10.1180/claymin.2008.043.1.11.
6. Azevedo RP, de da Silva LCM, Pereira FAC, Peche PM, Pio LAS, Mancini M, Curi N, Silva BM. Interactions between intrinsic soil properties and deep tillage in the sustainable management of perennial crops. Sustainability. 2022;15:760. https://doi.org/10.3390/su15010760.
7. Derpsch R, Franzluebbers AJ, Duiker SW, Reicosky DC, Koeller K, Friedrich T, Sturny WG, Sá JCM, Weiss K. Why do we need to standardize no-tillage research? Soil Tillage Res. 2014;137:16–22. https://doi.org/10.1016/j.still.2013.10.002.
8. Raphael JPA, Calonego JC, Milori DMBP, Rosolem CA. Soil organic matter in crop rotations under no-till. Soil Tillage Res. 2016;155:45–53. https://doi.org/10.1016/j.still.2015.07.020.
9. Rigon JPG, Franzluebbers AJ, Calonego JC. Soil aggregation and potential carbon and nitrogen mineralization with cover crops under tropical no-till. J Soil Water Conserv. 2020;75:601–9. https://doi.org/10.2489/jswc.2020.00188.
10. Ramesh T, Bolan NS, Kirkham MB, et al. Soil organic carbon dynamics: Impact of land use changes and management practices: a review. Amsterdam: Elsevier; 2019. p. 1–107.
11. Bojko O, Kabala C. Organic carbon pools in mountain soils—sources of variability and predicted changes about climate and land use changes. CATENA. 2017;149:209–20. https://doi.org/10.1016/j.catena.2016.09.022.

12. Bloch SE, Ryu MH, Ozaydin B, Broglie R. Harnessing atmospheric nitrogen for cereal crop production. Curr Opin Biotechnol. 2020;62:181–8. https://doi.org/10.1016/j.copbio.2019.09.024.
13. Xu G, Fan X, Miller AJ. Plant nitrogen assimilation and use efficiency. Annu Rev Plant Biol. 2012;63:153–82. https://doi.org/10.1146/annurev-arplant-042811-105532.
14. PNF. Plano Nacional de Fertilizantes 2050—Uma Estratégia para os Fertilizantes no Brasil. Brasilia. p 195, 2021.
15. Caires EF, Haliski A, Bini AR, Scharr DA. Surface liming and nitrogen fertilization for crop grain production under no-till management in Brazil. Eur J Agron. 2015;66:41–53. https://doi.org/10.1016/j.eja.2015.02.008.
16. Fernandez JA, DeBruin J, Messina CD, Ciampitti IA. Late-season nitrogen fertilization on maize yield: a meta-analysis. F Crop Res. 2020;247:107586. https://doi.org/10.1016/j.fcr.2019.107586.
17. Jenkinson DS. Chemical tests for potentially available nitrogen in soil. J Sci Food Agric. 1968;19:160–8. https://doi.org/10.1002/jsfa.2740190310.
18. Benedet L, Faria WM, Silva SHG, Mancini M, Demattê JAM, Guilherme LRG, Curi N. Soil texture prediction using portable X-ray fluorescence spectrometry and visible near-infrared diffuse reflectance spectroscopy. Geoderma. 2020;376:114553. https://doi.org/10.1016/j.geoderma.2020.114553.
19. Mahesh B. Machine learning algorithms—a review. Int J Sci Res. 2020;9:381–6. https://doi.org/10.21275/ART20203995.
20. Ribeiro BT, Silva SHG, Silva EA, Guilherme LRG. Portable X-ray fluorescence (pXRF) applications in tropical Soil Science. Ciência e Agrotecnol. 2017;41:245–54. https://doi.org/10.1590/1413-70542017413000117.
21. Silva S, Poggere G, Menezes M, Carvalho G, Guilherme L, Curi N. Proximal sensing and digital terrain models applied to digital soil mapping and modeling of brazilian latosols (Oxisols). Remote Sens. 2016;8:614. https://doi.org/10.3390/rs8080614.
22. Skoog DA, Leary JJ. Principles of instrumental analysis. 4th ed. Philadelphia: Saunders; 1992.
23. Witheetrirong Y, Tripathi NK, Tipdecho T, Parkpian P. Estimation of the effect of soil texture on nitrate-nitrogen content in groundwater using optical remote sensing. Int J Environ Res Public Health. 2011;8:3416–36. https://doi.org/10.3390/ijerph8083416.
24. Embrapa—Empresa Brasileira de Pesquisa Agropecuária. Sistema brasileiro de classificação de solos, 2018.
25. Köppen W, Geiger R. Klimate der Erde. Gotha: Verlag Justus Perthes Wall-map 150cmx200cm 91–102, 1928.
26. Raij BV, Andrade JC, Cantarella H, Quaggio JÁ. Análise química para avaliação da fertilidade de solos tropicais. Potafós: Campinas; 2001.
27. Embrapa—Empresa Brasileira De Pesquisa AGROPECUÁRIA. Análise Granulométrica. Manual de métodos de análise de solo, p. 574, 2017.
28. Rocha FA, Martinez MA, Matos AT, et al. Modelo numérico do transporte de nitrogênio no solo. Parte II: reações biológicas durante a lixiviação. Rev Bras Eng Agrícola Ambient. 2008;12:54–61. https://doi.org/10.1590/S1415-43662008000100008.
29. Mattos Junior D, Cantarella H, Raij B. Manuseio e conservação de amostras de solo para preservação do nitrogênio inorgânico. Rev Bras Ciência do Solo. 1995;19:423–31.
30. Bellinaso H, Demattê JAM, Romeiro AS. Soil spectral library and its use in soil classification. Rev Bras Ciência do Solo. 2010;34:861–70. https://doi.org/10.1590/S0100-06832010000300027.
31. Rosin NA, Demattê JAM, Leite MCA, et al. The fundamental of the effects of water, organic matter, and iron forms on the pXRF information in soil analyses. CATENA. 2022;210:105868. https://doi.org/10.1016/j.catena.2021.105868.
32. Silvero NEQ, Di Raimo LADL, Pereira GS, et al. Effects of water, organic matter, and iron forms in mid-IR spectra of soils: assessments from laboratory to satellite-simulated data. Geoderma. 2020;375:114480. https://doi.org/10.1016/j.geoderma.2020.114480.
33. Matos ES, Mendonça ES, Morales MM et al. Carbono Total E Frações Químicas De Cabono Do Solo. In: Mendoça E S, Matos Es (eds). Matéria Orgânica do Solo—Métodos de Análises, 2 edn. p 221, 2017.
34. Yeomans JC, Bremner JMA. A rapid and precise method for routine determination of organic carbon in soil. Soil Sci Plant Anal. 1988;19:1467–76.
35. Cantarella H, Trivelin PCO. Determinação de nitrogênio inorgânico em solo pelo método da destilação a vapor. In: Análise química para avaliação da fertilidade de solos tropicais. Campinas: Instituto Agronômico; 2001. p. 285.
36. Dangal S, Sanderman J, Wills S, Ramirez-Lopez L. Accurate and precise prediction of soil properties from a large mid-infrared spectral library. Soil Syst. 2019;3:11. https://doi.org/10.3390/soilsystems3010011.
37. Deiss L, Margenot AJ, Culman SW, Demyan MS. Optimizing acquisition parameters in diffuse reflectance infrared Fourier transform spectroscopy of soils. Soil Sci Soc Am J. 2020;84:930–48. https://doi.org/10.1002/saj2.20028.
38. Terra FS, Demattê JAM, Viscarra Rossel RA. Spectral libraries for quantitative analyses of tropical Brazilian soils: Comparing vis–NIR and mid-IR reflectance data. Geoderma. 2015;255–256:81–93. https://doi.org/10.1016/j.geoderma.2015.04.017.
39. Xu S, Wang M, Shi X, et al. Integrating hyperspectral imaging with machine learning techniques for the high-resolution mapping of soil nitrogen fractions in soil profiles. Sci Total Environ. 2021;754:142135. https://doi.org/10.1016/j.scitotenv.2020.142135.
40. Qi Q, Zhang D, Zhang M, Tong S, et al. Spatial distribution of soil organic carbon and total nitrogen in disturbed Carex tussock wetland. Ecol Indic. 2021;120:106930. https://doi.org/10.1016/j.ecolind.2020.106930.
41. Nyarko F, Tack FMG, Mouazen AM. Potential of visible and near-infrared spectroscopy coupled with machine learning for predicting soil metal concentrations at the regional scale. Sci Total Environ. 2022;841:156582. https://doi.org/10.1016/j.scitotenv.2022.156582.
42. Kuang B, Mouazen AM. Influence of the number of samples on prediction error of isible and nearinfrared spectroscopy of selected soil properties at the farm scale. Eur J Soil Sci2012;63(3):421–429.
43. Li S, et al. In situ measurements of organic carbon in soil profiles using Vis-NIR spectroscopy onthe Qinghai–Tibet Plateau. Environ Sci Technol 2015;49(8):4980–4987.
44. Chen S, et al. Study on the characterization of VNIR-MIR spectra and prediction of soil organicmatter in paddy soil (in chinese). Spectrosc Spectr Anal 2016;36:1712–1716.
45. Hong Y, et al. Fusion of visible-to-near-infrared and mid-infrared spectroscopy to estimate soilorganic carbon. Soil Tillage Res 2022;217:105284. https://doi.org/10.1016/j.still.2021.105284.
46. Morona F, et al. Quick analysis of organic matter in soil by energy-dispersive X-ray fluorescenceand multivariate analysis. Appl Radiat Isot 2017;130:13–20.
47. Lucà F, et al. Effect of calibration set size on prediction at local scale of soil carbon by Vis-NIRspectroscopy. Geoderma 2017;288:175–183.
.

48. Clingensmith CM, Grunwald S. Predicting soil properties and interpreting Vis-NIR models from across the continental United States. Sensors. 2022;22:3187. https://doi.org/10.3390/s22093187.
49. Ladoni M, Bahrami HA, Alavipanah SK, Norouzi AA. Estimating soil organic carbon from soil reflectance: a review. Precis Agric. 2010;11:82–99. https://doi.org/10.1007/s11119-009-9123-3.
50. Viscarra Rossel RA, McGlynn RN, McBratney AB. Determining the composition of mineral-organic mixes using UV–vis–NIR diffuse reflectance spectroscopy. Geoderma. 2006;137:70–82. https://doi.org/10.1016/j.geoderma.2006.07.004.
51. Wei Y, Zhu X, Li C, et al. Monitoring soil nitrate nitrogen based on hyperspectral data in the apple orchards. Agric Sci. 2017;08:21–32. https://doi.org/10.4236/as.2017.81002.
52. Yang K, Browne PRL, Huntington JF, et al. Characterising the hydrothermal alteration of the Broadlands-Ohaaki geothermal system, New Zealand, using short-wave infrared spectroscopy. J Volcanol Geotherm Res. 2001;106:53–65. https://doi.org/10.1016/S0377-0273(00)00264-X.
53. Islam K, Singh B, McBratney A. Simultaneous estimation of several soil properties by ultra-violet, visible, and near-infrared reflectance spectroscopy. Soil Res. 2003;41:1101. https://doi.org/10.1071/SR02137.
54. Shi T, Cui L, Wang J, et al. Comparison of multivariate methods for estimating soil total nitrogen with visible/near-infrared spectroscopy. Plant Soil. 2013;366:363–75. https://doi.org/10.1007/s11104-012-1436-8.
55. Clark RN. Spectroscopy of rocks and minerals and principles of spectroscopy. In: Remote Sensing for the Earth Sciences (Manual of Remote Sensing). p 3–58, 1999.
56. Meneses Pr, Almeida T De, Baptista Gmdm. Reflectância dos materiais terrestres : análise e interpretação. São Paulo, 2019.
57. Vaz SJ. Técnicas analíticas em Química Analítica Aplicada a Poluentes Emergentes. Berlim, Alemanha, 2018
58. Steinberg CEW. Ecology of humic substances in freshwaters. Berlin Heidelberg, Berlin, Heidelberg: Springer; 2003.
59. Ukalska-Jaruga A, Bejger R, Debaene G, et al. Characterization of soil organic matter individual fractions (fulvic acids, humic acids, and humines) by spectroscopic and electrochemical techniques in agricultural soils. Agronomy. 2021;11:1067. https://doi.org/10.3390/agronomy11061067.
60. Yang K, Browne PRL, Huntington JF, et al. Characterising the hydrothermal alteration of theBroadlands–Ohaaki geothermal system, New Zealand, using short-wave infrared spectroscopy. JVolcanol Geotherm Res 2001;106:53–65. https://doi.org/10.1016/S0377-0273(00)00264-X.
61. Rossel RAV, Behrens T. Using data mining to model and interpret soil diffuse reflectance spectra. Geoderma. 2010;158:46–54. https://doi.org/10.1016/j.geoderma.2009.12.025.
62. Fystro GA. The prediction of C and N content and their potential mineralization in heterogeneous soil samples using Vis–NIR spectroscopy and comparative methods. Plant Soil. 2002. https://doi.org/10.1023/A:1020612319014.
63. Janik LJ, Forrester ST, Rawson A. The prediction of soil chemical and physical properties from mid-infrared spectroscopy and combined partial least-squares regression and neural networks (PLS-NN) analysis. Chemom Intell Lab Syst. 2009;97:179–88. https://doi.org/10.1016/j.chemolab.2009.04.005.
64. Li H, Wang J, Zhang J, et al. Combining variable selection and multiple linear regression for soil organic matter and total nitrogen estimation by DRIFT-MIR spectroscopy. Agronomy. 2022;12:638. https://doi.org/10.3390/agronomy12030638.
65. Rossel RAV, Jeon YS, Odeh IOA, et al. Using a legacy soil sample to develop a mid-IR spectral library. Soil Res. 2008;46:1. https://doi.org/10.1071/SR07099.
66. Boonmung S, Riley MR. Quantitative analysis of added ammonium and nitrate in silica sand and soil using diffuse reflectance infrared spectroscopy. Spectrosc Lett. 2003;36:251–74. https://doi.org/10.1081/SL-120024358.
67. Parikh SJ, Goyne KW, Margenot AJ. Soil chemical insights provided through vibrational spectroscopy. Amsterdam: Elsevier; 2014. p. 1–148.
68. Ding G, Novak JM, Amarasiriwardena D. Soil organic matter characteristics as affected by tillage management. Soil Sci Soc Am J. 2002;66:421–9. https://doi.org/10.2136/sssaj2002.4210.
69. Dick DP, Santos JHZ, Ferranti EM. Chemical characterization and infrared spectroscopy of soil organic matter from two southern Brazilian soils. Rev Bras Ciência do Solo. 2003;27:29–39. https://doi.org/10.1590/S0100-06832003000100004.
70. Sabetizade M, Gorji M, Roudier P, et al. Combination of MIR spectroscopy and environmental covariates to predict soil organic carbon in a semi-arid region. CATENA. 2021;196:104844. https://doi.org/10.1016/j.catena.2020.104844.
71. Madhavan DB, Kitching M, Mendham DS. Mid-infrared spectroscopy for rapid assessment of soil properties after land use change from pastures to Eucalyptus globulus plantations. J Environ Manage. 2016;175:67–75. https://doi.org/10.1016/j.jenvman.2016.03.032.
72. Celi L, Schnitzer M, Nègre M. Analysis of carboxyl groups in soil humic acids by a wet chemical method, fourier-transform infrared spectrophotometry, and solution-state carbon-13 nuclear magnetic resonance. A comparative. Study Soil Sci. 1997;162:189–97. https://doi.org/10.1097/00010694-199703000-00004.
73. Hutongs C, Ludwig B, Jung A, et al. Comparison of portable and bench-top spectrometers for mid-infrared diffuse reflectance measurements of soils. Sensors. 2018;18:993. https://doi.org/10.3390/s18040993.
74. Sisouane M, Cascant MM, Tahiri S, et al. Prediction of organic carbon and total nitrogen contents in organic wastes and their composts by Infrared spectroscopy and partial least square regression. Talanta. 2017;167:352–8. https://doi.org/10.1016/j.talanta.2017.02.034.
75. Cascant MM, Sisouane M, Tahiri S, et al. Determination of total phenolic compounds in compost by infrared spectroscopy. Talanta. 2016;153:360–5. https://doi.org/10.1016/j.talanta.2016.03.020.
76. Tavares TR, Molin JP, Nunes LC, et al. Multi-sensor approach for tropical soil fertility analysis: comparison of individual and combined performance of VNIR, XRF, and LIBS spectroscopies. Agronomy. 2021;11:1028. https://doi.org/10.3390/agronomy11061028.
77. Tavares TR, Mouazen AM, Alves EEN, et al. Assessing soil key fertility attributes using a portable x-ray fluorescence: a simple method to overcome matrix effect. Agronomy. 2020;10:787. https://doi.org/10.3390/agronomy10060787.
78. USEPA. Method 6200: field portable X-ray fluorescence spectrometry for the determination of elemental concentrations in soil and sediment. In: United States Environmental Protection Agency. Washington: USEPA; 2007.
79. Laperche V, Lemière B. Possible pitfalls in the analysis of minerals and loose materials by portable XRF, and how to overcome them. Minerals. 2020;11:33. https://doi.org/10.3390/min11010033.

80. Curi N, Silva SHG, Poggere GC, et al. Mapeamento de solos e magnetismo no campus da UFLA como traçadores ambientais. 1st ed. Lavras: UFLA; 2017.
81. Wang D, Chakraborty S, Weindorf DC, et al. Synthesized use of VisNIR DRS and PXRF for soil characterization: Total carbon and total nitrogen. Geoderma. 2015;243–244:157–67. https://doi.org/10.1016/j.geoderma.2014.12.011.
82. Gomes LC, Faria RM, Souza E, et al. Modelling and mapping soil organic carbon stocks in Brazil. Geoderma. 2019;340:337–50. https://doi.org/10.1016/j.geoderma.2019.01.007.
83. Zhang Y, Hartemink AE. Data fusion of vis-NIR and PXRF spectra to predict soil physical and chemical properties. Eur J Soil Sci. 2020;71:316–33. https://doi.org/10.1111/ejss.12875.
84. Rawal A, et al. Determination of base saturation percentage in agricultural soils via portable X-ray fluorescence spectrometer. Geoderma. 2019;338:375–82.
85. Song J, Gao J, Zhang Y et al. Estimation of soil organic carbon content in coastal wetlands withmeasured VIS-NIR spectroscopy using optimized support vector machines and random forests.Remote Sens 2022;14:4372. https://doi.org/10.3390/rs14174372.
86. FAO. A primer on soil analyses using visible and near-infrared (vis-NIR) and mid-infrared (MIR) spectroscopy. In: Soil spectroscopy. Roma: Food and Agriculture Organization of the United Nations; 2022.
87. Wartini N, Budiman M, Sang Ho J, et al. Mid-infrared spectroscopy for accurate measurement of an extensive set of soil properties for assessing soil functions. Soil Secur. 2022;6:100043. https://doi.org/10.1016/j.soisec.2022.100043.
88. Soriano-disla JM, et al. The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties. Appl Spectrosc Rev. 2014;49(2):139–86.
89. Ramírez PB, et al. Applying NIR and MIR spectroscopy for C and soil property prediction in northern cold-region ecosystems. Which approach works better? Geod Reg. 2023;32: e00617.
90. Schirrmann M, Gebbers R, Kramer E. Performance of automated near-infrared reflectance spectrometry for continuous in situ mapping of soil fertility at field scale. Vad Zone J. 2013;12(4):1–14.
91. Declercq Y, et al. A comprehensive study of three different portable XRF scanners to assess the soil geochemistry of an extensive sample dataset. Remote Sens. 2019;11(21):2490.
92. Greenberg I, et al. Evaluation of mid-infrared and X-ray fluorescence data fusion approaches for prediction of soil properties at the field scale. Sensors. 2023;23(2):662.
93. Sollins P, et al. Sequential density fractionation across soils of contrasting mineralogy: evidence for both microbial- and mineral-controlled soil organic matter stabilization. Biogeochemistry. 2009;96(1–3):209–31.
94. Dalmolin RSD, et al. Relação entre os constituintes do solo e seu comportamento espectral. Ciência Rural. 2005;35(2):481–9.
95. Stoner ER, Baumgardner MF. Characteristic variations in reflectance of surface soils. Soil Sci Soc Am J. 1981;45(6):1161–5.
96. OLYMPUS. Periodic table of detectable elements, 2013.