SCIENTIA AGRICOLA

Soils and Plant Nutrition

Diffuse reflectance spectroscopy for estimating soil organic carbon and make nitrogen recommendations

Nicolas Augusto Rosin[iD], Ricardo Simão Diniz Dalmolin*[iD]; Taciara Zborowski Horst-Heinen[iD], Jean Michel Moura-Bueno[iD], Daniely Vaz da Silva-Sangoi[iD], Leandro Souza da Silva[iD]

Universidade Federal de Santa Maria/CCR, Depto. de Solos, Av. Roraima, 1000 – 97105-900 – Santa Maria, RS – Brasil.
*Corresponding author <dalmolin@ufsm.br>

ABSTRACT: Diffuse reflectance spectroscopy (DRS) has the potential to predict soil organic carbon (SOC). However, it is still little used as a matter of routine in soil laboratories in Brazil. The objective of this study was to make evaluations as to whether SOC predicted by spectral techniques can replace measurement by routine chemical methods with no loss in quality and be applied in the recommendation of nitrogen fertilizer as well as identifying the best prediction strategies to use. A data set containing 2,471 samples from six soil spectral libraries (SSL) was used to develop spectroscopic models for SOC content prediction, including consideration of sample stratification and preprocessing techniques. The SOC was quantified through the analytical-chemical methods of wet combustion with determination by titration, designated as the reference method (REM), and colorimeter, designated as the routine method (ROM in an independent data set). SOC contents predicted by the spectral analysis method (SAM) were compared to the REM and ROM results, converted to soil organic matter (SOM) and used for N recommendations. The best estimate for SOM content using the SAM was achieved through stratification of the SSL and application of the standard normal variate (SNV) preprocessing. The SOC predicted by spectral techniques proved capable of replacing the SOC measured by routine chemical methods with no loss of quality and supported by an appropriate nitrogen fertilizer recommendation, provided the models met the conditions and possessed the characteristics of the samples to be analyzed.

Keywords: soil attributes prediction, soil fertility, proximal soil sensing, chemometric, green chemistry

Introduction

Soil organic carbon (SOC) is fundamental to a profitable and sustainable agricultural system as an indicator of soil quality (Global Soil Partnership, 2017; Bailey et al., 2017). The SOC content, expressed in terms of soil organic matter (SOM) content, is also used by several fertilizer recommendation systems, in different states and zones such as Rio Grande do Sul and Santa Catarina (SBCS, 2016), Cerrado (Embrapa, 2004), Paraná (SBCS, 2019), São Paulo (IAC, 1996), and Minas Gerais (SBCS, 1999), to evaluate soil fertility and as a criterion for nitrogen (N) fertilizer recommendation for crops designated by these systems.

Reliable estimates of SOM are important in ensuring that N recommendations are adjusted for soil nutrient availability and crop demand, notably with the intention of avoiding underestimating recommendations that can result in yields below the crop potential, environmental contamination by N excess, and financial waste. Among different chemical analytical methods employed by the soil analysis laboratories to determinate SOC/SOM, it is very common to use those based on soil carbon oxidation by a sulfochromic solution, followed by spectrophotometric colorimeter determination (Tedesco et al., 1995). This is based on the original procedure of Walkley and Black (1934), using hazardous chemical reagents, such as the chromium solution, and

are relatively laborious and time-consuming to execute. Therefore, we were looking for an alternative method that can be used without these disadvantages, but with accuracy in SOC/SOM estimation. Soil spectral analysis by the diffuse reflectance spectroscopy (DRS) technique associated with mathematical models is an alternative for estimating SOC/SOM content with faster execution analysis, easier sample preparation, and with no chemical extractors (Viscarra Rossel et al., 2016; Wijewardane et al., 2016; Dotto et al., 2018). Studies have also addressed the DRS technique when estimating the nitrogen content (Li et al., 2015; Padmanabhi and Deshmukh, 2017; Conforti et al., 2018).

Even with the known potential of spectroscopic models to quantify the SOC, there is a lack of information about the applicability of models and the comparison between SOC obtained by these models and laboratory routine chemical methods in Brazil. The key question is if the SOM estimated by DRS is equivalent to SOM quantified by routine chemical analysis. This information is necessary to an evaluation of the accuracy of SOC/SOM content estimated by DRS to meet the demands for N recommendation reports. The objective of this work was to assess if SOC predicted by spectral techniques can replace SOC measured by routine chemical methods with no loss of quality with an appropriate nitrogen fertilizer recommendation as well as identify the best prediction strategies to accomplish this.

Materials and Methods

### Soil database and data stratification

Database information

A data set containing 2,471 legacy samples from six soil spectral libraries (SSL) from Rio Grande do Sul (RS) and Santa Catarina (SC), Brazil (Figure 1), was used in this study. The samples were taken according to different proposals, in layers or horizons by a random, intentional, or convenience strategy (Table 1). The data set is heterogeneous and it contains samples that represent the predominant soil class in the south of Brazil: Argissolos (Acrisols and Alisols), Cambissolos (Cambisols), Gleissolos (Gleysols), Latossolos (Ferralsols), Neossolos (Regosols), Nitossolos (Nitisols), Organossolos (Histosols) and Planossolos (Planosols) by classification of Brazilian System of Soil Classification - SiBCS (Santos et al., 2018) and World Reference Base (IUSS/WRB, 2015). They also include areas from different land use/land cover including forest, native grasslands, and field crops using tillage and no-till systems as well as the climate from these states ranging from humid subtropical without dry season with hot summer (Cfa) to humid subtropical without dry seasons with temperate summer (Cfb), according to the Köppen classification (Alvares et al., 2013).

The main difference between the various SSL is the parent material that they came from. While extrusive igneous rocks are predominant in four of the six SSLs, the sedimentary rocks are predominant in two of them. The SSL that came from extrusive rocks as parent material can be acidic or basic, and has a clayey texture or finer, and the mineralogy comprises a clay fraction composed of kaolinite, hematite and goethite. On the other hand, the SSL with sedimentary rocks as parent material has a sandy to very clayey soil texture and the mineralogy comprises a clay fraction composed of different contents of kaolinite, 2:1 minerals with hydroxy-Al between layers, hematite and goethite. This variation in mineralogical composition between the SSLs has a significant effect on their spectral behavior (Araújo et al., 2014).

In order to evaluate and compare the performance of these methods in predicting SOC/SOM for the nitrogen recommendation, an independent validation set (IVS) was used as an external validation. The IVS was composed of 69 samples distributed throughout the city of São Miguel das Missões, RS (Figure 1). This data set was selected because it is sourced from a field crop, which is the main recipient of nitrogen fertilizer application. In addition, it was the only set with routine laboratory determined SOC/SOM data readily available.

The SOC content in these SSL was determined from air-dried soil samples, ground in a mortar, by wet combustion with sulfochromic solution ($K_2Cr_2O_7$ + H_2SO_4) in the presence of external heating, and determined by titration with $Fe(NH_4)_2(SO_4)_2.6H_2O$ (Yeomans and Bremner, 1988), herein designated the reference method (REM). In the IVS, SOC content was quantified by REM and also by the method used by the Official Network of Soil Analysis Laboratories in Rio Grande do Sul and Santa Catarina (SBCS, 2016), according to the methodology described in Tedesco
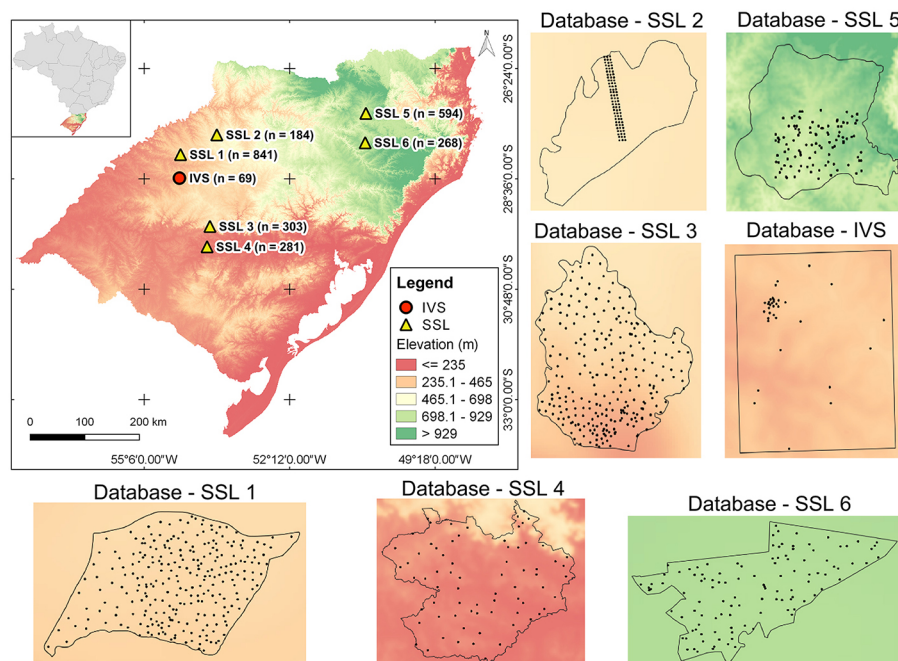
Figure 1 – Spatial distribution of samples from the six soil spectral libraries (SSL), independent validation set (IVS) and their respective geographic location in the States of Rio Grande do Sul and Santa Catarina.

Table 1 – Main characteristics of the Spectral Soil Library sites.

Database	State/ Geographic coordinates	Climate	Parent material	Soil type SiBCS (WRB)	Land use/ Land cover	n. total	Sampling type/ Dept (cm)	MS
SSL 1 (Moura-Bueno et al., 2019)	RS/ 28°6'59" S 54°22'35" W	Cfa	Extrusive igneous (basic)	Latossolo (Ferralsol) Gleissolo (Gleysol)	Field crops, Forest, Native grasslands	841	IM/0-5, 5-15, 15-30, 30-60, 60-100	C1, C2, C3
SLL 2 (Ramos et al., 2017)	RS/ 27°43'22" S 53°39'10" W	Cfa	Extrusive igneous (basic)	Latossolo (Ferralsol)	Field crops	184	C/horiz	C1, C2, C3
SSL 3 (Samuel-Rosa et al., 2013)	RS/ 29°32'55" S 53°47'26" W	Cfa	Sedimentary rocks	Argissolo (Acrisol/Alisol)	Field crops, Forest, Native grasslands	303	IM/0-20, 20-40	C1
SSL 4 (Flores et al., 2015)	RS/ 29°57'13" S 53°50'49" W	Cfa	Sedimentary rocks	Argissolo (Acrisol/Alisol), Planossolo (Planosol)	Field crops, Forest, Native grasslands	281	C/horiz	C1
SSL 5 (Dotto et al., 2018)	SC/ 27°17'42" S 50°40'53" W	Cfb	Extrusive igneous (acidic)	Cambissolo (Cambisol), Latossolo (Ferralsol), Neossolo (Regosol)	Forest, Native grasslands	594	cLHS/0-5, 5-15, 15-30, 30-60, 60-100	C1, C2
SSL 6 (Horst et al., 2018)	SC/ 27°52'48" S 50°41'42" W	Cfb	Extrusive igneous (acidic)	Cambissolo (Cambisol), Latossolo (Ferralsol), Neossolo (Regosol) Nitossolo (Nitisol)	Forest	268	cLHS/0-20, 20-40, 40-60, 60-100	C1, C2
IVS (Wolski et al., 2017)	RS/ 28°35'11" S 54°23'53" W	Cfa	Extrusive igneous (basic)	Latossolo (Ferralsol) Nitossolo (Nitisol)	Field crops	69	C/horiz	IVS

SSL = Soil Spectral Library, IVS = independent validation set; UTM = central site coordinates as Universal Transverse Mercator; Cfa = Köppen classification as humid, oceanic climate, with hot summer; Cfb = Köppen classification as subtropical without dry season, with temperate summer; SiBCS = Brazilian System of Soil Classification 2018; WRB = World Reference Base 2015 – FAO; MS = Modeling set; C1 = calibration set one composed by 2,471 samples; C2 = calibration set two composed by 1,887 samples; C3 = calibration set three composed by 1,025 samples; IM = irregular mesh sampled, C = convenient sampled, cLHS = conditioned Latin Hypercube Sampled, horiz = pedogenic horizon.

et al. (1995), herein designated the routine method (ROM). The spectral measurement was taken in the laboratory by a spectroradiometer, with a spectral range from 350 to 2500 nm, spectral resolution from 3 nm (up to 700 nm) to 10 nm and sampling interval of 1 nm. The spectroradiometer was calibrated every 20 sample readings using a white spectralon (barium sulfate) plate as a 100 % reflectance reference standard.

Data stratification

Looking for the best training set to predict SOM for nitrogen recommendation the set of six SSL, henceforth designated as calibration set one (C1, n = 2,471) was stratified into two groups on the basis of their similarity in relation to the IVS. This was done because SSL composed of heterogeneous samples usually decrease the predictive performance of models (Araújo et al., 2014; Viscarra Rossel et al., 2016; Wijewardane et al., 2016; Moura-Bueno et al., 2019). Furthermore, according to Grunwald et al. (2015) models must include local samples representative of the pedological characteristics of interest.

Taking this into account, the first stratification criterion was parent material. In this case, the SSL which came from regions with extrusive volcanic rocks as parent material (SSL1, SSL2, SSL5, and SSL6) were included in calibration set two (C2, n = 1,887 samples). The second stratification criterion took into account both the parent material of the rock acidity and the climate. In this case, only the SSLs which came from regions with extrusive volcanic rocks (basic) as parent material and climate Cfa (SSL1 and SSL2), were included in calibration set three (C3, n = 1,025 samples). Thereby the similarity between the calibration sets and the IVS increases at the same time that the sample number decreases. To explore the spectral variance of sets C1, C2, C3, and similarity with the IVS, principal component analysis (PCA) was conducted.

Spectral preprocessing, model fitting and validation

For each calibration set, three spectral preprocessing techniques were tested: i) smoothing using a 9-nm search window (SMO); ii) Savitzky-Golay 1st derivative, using a 1st order polynomial and 9-nm search window (SGD) (Savitzky and Golay, 1964) and iii) standard normal variate (SNV), making a total of nine models. These preprocessing techniques were used in order to reduce noise and signals without significance in the spectra, in addition to isolating specific absorption features (Rinnan et al., 2009). Preprocessing was performed using the prospectr package (Stevens and Ramírez-López, 2013) implemented by the R software program (R Core Team, 2019).

The method of soil spectral analysis (SAM) was used to obtain SOC predicted values through spectroscopic models. Models were fitted to

three calibration data sets (C1, C2 and C3), by the multivariate method of partial least squares regression (PLSR). The PLSR model is widely used in spectroscopic modeling (Dotto et al., 2018) and has exceptional ability in estimating SOC from subtropical soil spectra (Moura-Bueno et al., 2019). The number of latent variables used was 25 which was selected based on the best parameter of latent variables to get a robust estimate so that the cross-validation error (RMSE) is minimized. This multivariate model has been successfully used in spectroscopic modeling to predict SOC (Dotto et al., 2018, Moura-Bueno et al., 2019). The modeling was performed using an R software program (R Core Team, 2019), with the pls package (Mevik et al., 2016).

The models were evaluated through leave-one-out cross-validation and were also independently validated, comparing them to the IVS reference values. The statistical parameters of accuracy evaluated in cross-validation and independent validation were the coefficient of determination (R^2) Eq. (1), root mean squared error (RMSE) Eq. (2), Bias Eq. (3), and ratio of performance to interquartile distance (RPIQ) (Bellon-Maurel et al., 2010).

$$R^2 = \frac{\sum_{i=1}^{N}(\hat{y}_i - \bar{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y}_i)^2} \tag{1}$$

$$RMSE = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2 \tag{2}$$

$$Bias = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - \bar{y}_i) \tag{3}$$

$$RPIQ = \frac{(Q3 - Q1)}{RMSE} \tag{4}$$

where: $\hat{y}$ = predicted value; $\bar{y}$ = mean observed value; y = observed values; N = number of samples with i = 1, 2,..., n; Q1 = 25 % of samples; Q3 = 75 % of samples; Q3 - Q1 = interquartile distance, which represents the range containing 50 % of the population around the median (Bellon-Maurel et al., 2010).

Comparison of Methods for Quantification of SOM and Nitrogen Recommendation

The SOC values from IVS obtained by REM, ROM, and SAM were used for soil fertility assessment and nitrogen recommendation. The Manual for Liming and Fertilization for the states of Rio Grande do Sul and Santa Catarina (SBCS, 2016) were used in a field crop under simulated conditions. The first step was to convert SOC content from the samples of IVS obtained by each method (REM, ROM, and SAM) into SOM by multiplying the SOC value by 1.724, an adjustment known as the Van Bemmelen factor.

Given the SOM, the second step was the assignment of the values of the IVS samples into the classes of SOM as defined by the Manual for Liming and Fertilization for the states of Rio Grande do Sul and Santa Catarina (SBCS, 2016), according to its SOM content: ≤ 2.5 % (low), 2.6 to 5 % (medium) and > 5 % (high). The errors in the assignment of classes were quantified. The SOM values obtained by SAM were also compared to the reference values obtained by REM and ROM in order to compare the methods together with RMSE Eq. (2).

The third step was the simulation of the nitrogen recommendation. The class assignment of the best model that provided the smallest error and RMSE from step two was chosen to represent the recommendation by SAM. For each class of REM, ROM, and SAM there was a specific N rate recommended, according to crop, yield expectations, and previous land use. Considering corn production with yield expectation of 7 t ha⁻¹ (average yield in the state of RS) (CONAB, 2019a) and fallow as previous land use the N and urea (45 % N) recommendations were calculated. The resulting SOC from each method was estimated, considering the price for urea in the southern Brazil market (approximately 350.00 USD t⁻¹, at an exchange rate of 1.00 USD = 4.00 BRL) (CONAB, 2019b). The flowchart of the methodology is presented in Figure 2.

Results and Discussion

Data description

SOC content in C1 varied between 0.20 % and 15.92 %, with a mean of 2.13 % and standard deviation (SD) of 1.65 % (Figure 3A). Most of the samples have between 1 and 4 % SOC (Figure 3A), values commonly found in subtropical soils in the south of Brazil (Boddey et al., 2010) while SOC values above 8 %, characteristic of organic soils, are uncommon. C2, containing only soils whose parent material are acidic and basic extrusive igneous rocks, showed the same SOC amplitude as C1, with a mean of 2.30 % and SD of 1.52 % (Figure 3B). Both data sets contain soil samples from multiple land uses that provide different inputs of plant material to the soil, which, added to climatic differences, justify the high amplitude in SOC values. This amplitude of SOC values in C3 was from 0.24 % to 5.64 % with low data dispersion (SD = 0.95 %) (Figure 3C) as expected, since the samples were from land use predominantly for field crops and under Latossolos (Ferralsols) with a mineralogy composed basically of kaolinite and iron oxides (hematite).

The IVS presented minimum value of 0.26 % and maximum value of 3.85 %, with a mean of 1.03 % and SD of 0.69 % (Figure 3D). The mean SOC values of the data sets are similar to each other, but higher than the IVS. However, the amplitude of SOC data in IVS is within the SOC interval of all calibration sets, showing measures of central tendency and dispersion closest to set C3.

The spectral variation of sets C1, C2, C3, and IVS can be observed by the PC1 and PC2 scores shown
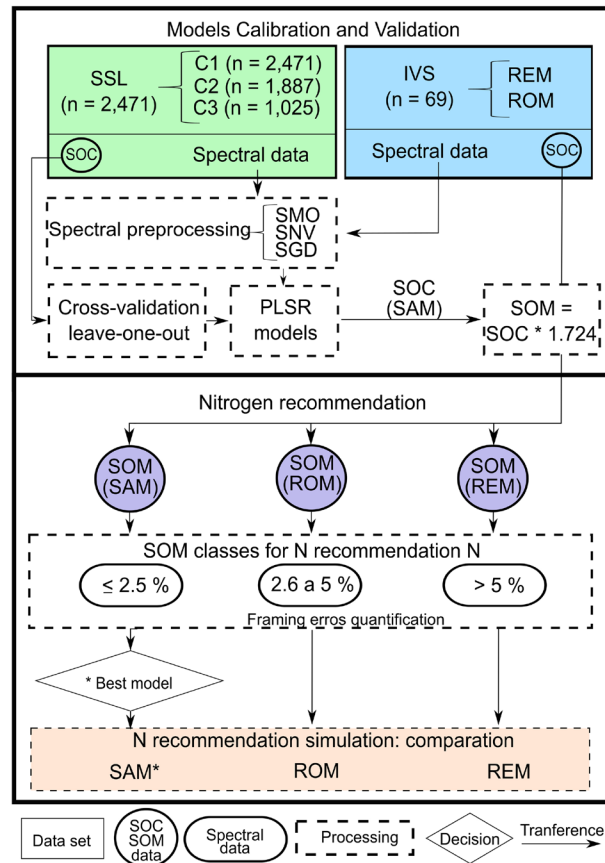
Figure 2 – Flowchart of the methodology. In which: SSL = soil spectral library; C1 = calibration set 1; C2 = calibration set 2; C3 = calibration set 3; IVS = independent validation set; SOC = soil organic carbon; SOM = soil organic matter; REM = reference method; ROM = routine method; SAM = spectral analysis method; SMO = smoothing; SGD = Savitzky-Golay derivative; SNV = standard normal variate; PLSR = partial least squares regression; N = nitrogen.

in Figure 4. The largest dispersion of spectral data is observed in the samples of set C1, with high data dispersion in the upper and lower right quadrants (Figure 4A). The spectral variation in the samples in the C2 set (Figure 4B) is smaller compared to the C1 set; however, it is greater compared to C3. The smallest spectral variation was observed in the C3 sets (Figure 4C), with most samples concentrated near the central axis. The samples from the IVS set showed lower spectral variation compared to the C3 sets, followed by C1 and C2. This showed that the IVS set had samples with more homogeneous chemical, physical and mineralogical characteristics. In addition, it is noted that the spectral variation of the IVS set is close to that observed in most samples from sets C1, C2, and C3, indicating that the IVS set can be used to validate calibrated models with these three sets.

Model Calibration and Validation

### Calibration and Leave-one-out Cross Validation Models

Models fitted to the C2 data set had the best adjustments, with R^2 ranging from 0.79 to 0.83 and RMSE values from 0.63 to 0.70 %. The C2-SGD model had better performance for calibration, where R^2 = 0.83 and RMSE = 0.63 % (Figure 5). The smallest errors were observed in models fitted to set C3, which had RMSE ranging from 0.46 to 0.52 %, with emphasis on the C3-SMO model (R^2 = 0.75 and RMSE = 0.46 %) (Figure 5). Corroborating our results, Moura-Bueno et al. (2019), using a local SSL in southern Brazil, found higher predictive accuracy of SOC associated with the PLSR model with SGD preprocessing (R^2 = 0.74 and RMSE = 0.52 %). Similarly, Knox et al. (2015), using soil libraries (n = 1,014) on a regional scale in Florida, USA, obtained an R^2 of 0.87 and RMSE of 0.40 (log g kg^{-1}).

The models adjusted to C1 had the worst adjustments, with R^2 ranging from 0.67 to 0.72 and RMSE from 0.87 to 0.95 %. These R^2 values are close to those of Pinheiro et al. (2017), in a study with 434 samples in the Central Amazon region, Brazil, in which they found R^2 = 0.71. However, the RMSE achieved by the authors was 0.57 %, which is lower than that observed in the present study. Although the prediction amplitude presented by the authors (SOC range between 0.01 and 10.56 %) is close to that of the present study (Figure 3), the observed difference in accuracy is attributed to the pedological heterogeneity. Even considering a lower predictive interval in C3, the RMSE ranged from 0.46 to 0.52 %, presenting lower accuracy than that observed by Terra et al. (2015), who found RMSE values = 0.16 % using a spectral library with 1,259 samples of tropical soils from Brazil, with a prediction interval between 0.2 and 7.0 % of SOC.

As spectral models usually assume patterns of spectral variation associated with changes in the SOC content of samples, heterogeneous samples in terms of mineralogical composition imply greater spectral variation that is not associated with SOC (Araújo et al., 2014; Moura-Bueno et al., 2019; Viscarra Rossel and Behrens, 2010). As regards the preprocessing techniques tested, in general, the accuracy of the model increased as the techniques became more complex. SGD showed the best performance, with R^2 ranging between 0.72 and 0.83 and RMSE between 0.48 and 0.87 %. This preprocessing removes unimportant reference signals from samples through the derivative of the measured responses relative to the wavelength number and promotes better pattern recognition. SNV preprocessing had R^2 ranging from 0.70 to 0.79 and RMSE from 0.52 to 0.91 % (Figure 5). The SMO preprocessing showed the worst adjustments, with R^2 ranging from 0.67 to 0.77 and RMSE from 0.46 to 0.95 % (Figure 5). However, the C2-SMO model performed better than C2-SNV. The C3-SMO model, on the other hand, performed better than C3-SNV and C3-SGD models. Although the literature suggests the use of spectral
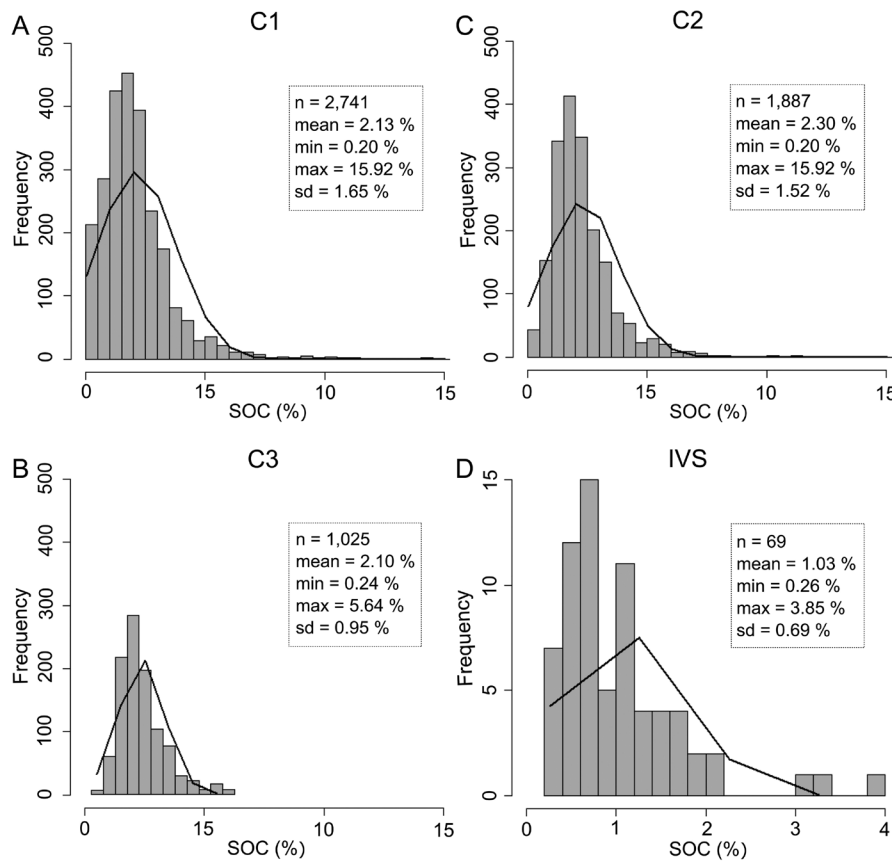
Figure 3 – Histograms and descriptive statistics of SOC data. In which: C1 = calibration set 1; C2 = calibration set 2; C3 = calibration set 3; IVS = independent validation set; SOC = soil organic carbon; n = number of samples; min = minimum value; max = maximum value; sd = standard deviation.
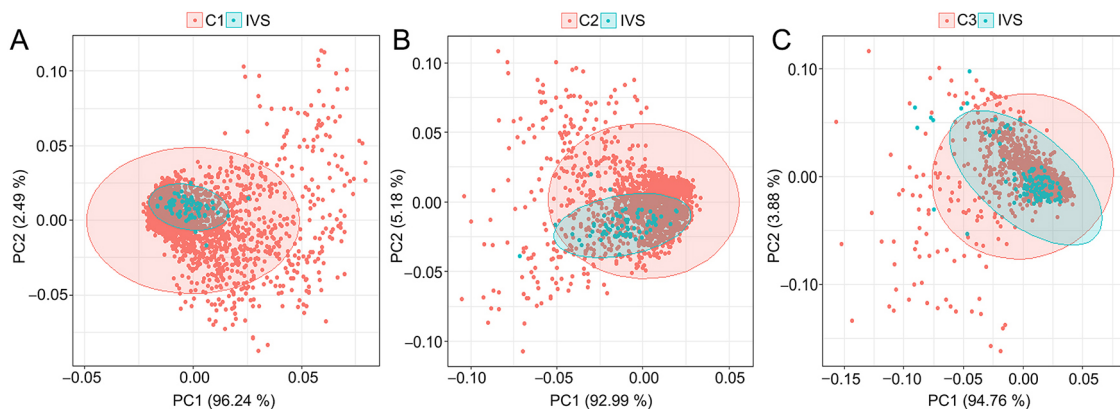

Figure 4 – PC1 and PC2 scores derived from PCA of spectral data identifying data set C1, C2, C3 and IVS. In which: C1 = calibration set 1; C2 = calibration set 2; C3 = calibration set 3; IVS = independent validation set; PC1 = principal components 1; PC2 = principal components 2.

preprocessing techniques (Vasques et al., 2008; Dotto et al., 2018; Moura-Bueno et al., 2019), which, however, do not always improve the predictions (Kooistra et al., 2001). Because of this uncertainty of outcomes, there is no spectral technique that can be considered as universally better. Preliminary tests are therefore

essential to matching preprocessing techniques to each data set of characteristics, especially for clayey to very clayey soils with high iron oxide content (Moura-Bueno et al., 2019).

Despite the significant reduction in the number of samples between sets, from 2,471 in C1 to 1,025 in
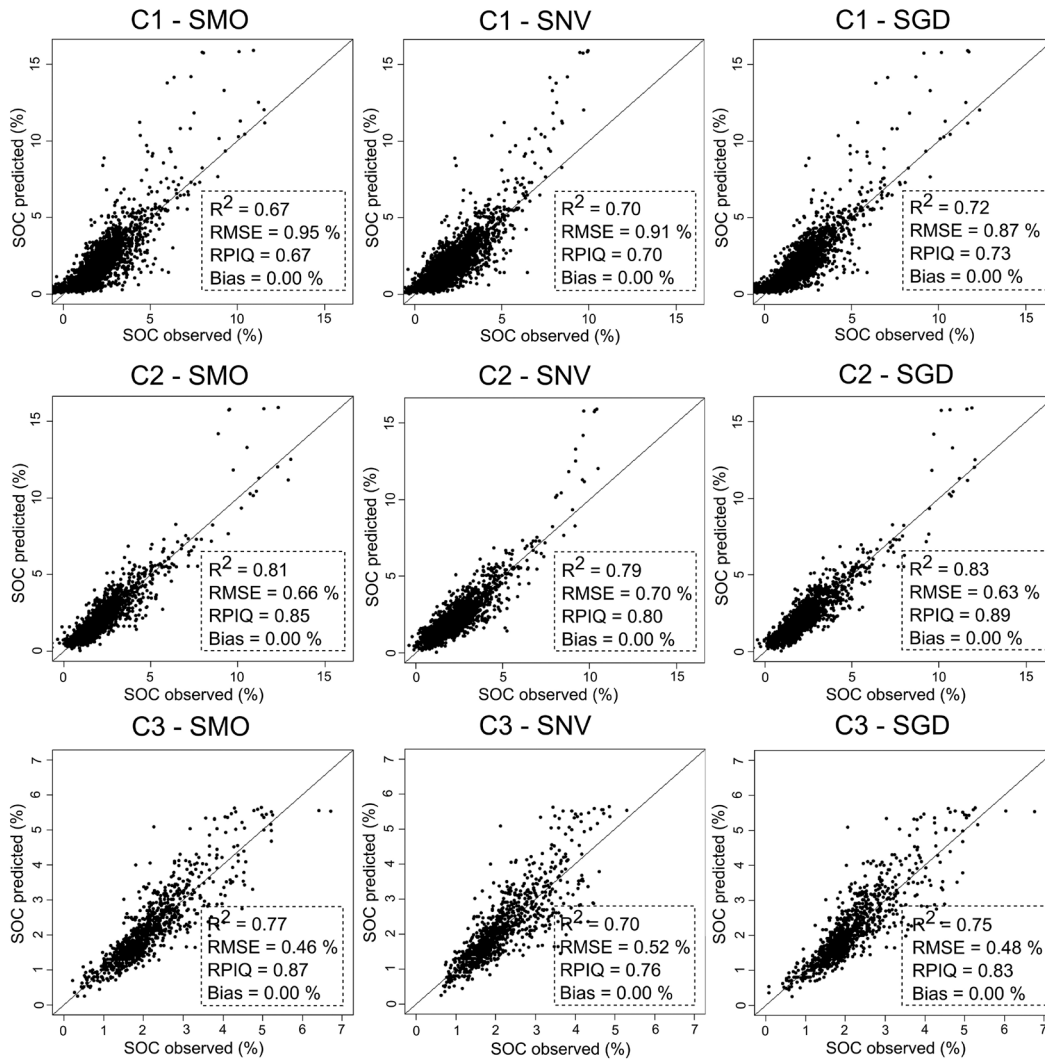
Figure 5 – Adjustment of the leave-one-out cross-validation for models. In which: C1 = calibration set 1; C2 = calibration set 2; C3 = calibration set 3; SOC = soil organic carbon; SMO = smoothing; SGD = Savitzky-Golay derivative; SNV = standard normal variate; R² = coefficient of determination; RMSE = root mean squared error; RPIQ = ratio of performance to interquartile distance.

C3, the predictive performance of the models did not depend only on the number of samples in the spectral library. More homogeneous sets in relation to sample characteristics may present more explicit numerical relationships and thus calibrate more accurate spectral models (Araújo et al., 2014; Viscarra Rossel et al., 2016; Wijewardane et al., 2016; Moura-Bueno et al., 2019).

Independent Validation Models

The independent validation was performed taking into consideration the prediction of SOC content for IVS samples (n = 69) and the accuracy values were calculated based on the predicted value for each spectroscopic model fitted (Figure 6).

The most complex preprocessing does not necessarily lead to increased accuracy in the independent validation. The best results as regards accuracy were verified in SGD preprocessing with R² ranging from 0.21 to 0.71 and RMSE from 0.39 to 0.77 %, followed by SMO, with R² ranging from 0.24 to 0.65 and RMSE from 0.59 to 0.80 % and SNV, with R² ranging from 0.14 to 0.78 and RMSE from 0.38 to 0.90 %.

C1 models presented the lowest degrees of accuracy in independent validation, with R² ranging from 0.14 to 0.21 and RMSE ranging from 0.77 to 0.90 % (Figure 6). Set C2 presented intermediate degrees of accuracy, with R² ranging from 0.41 to 0.62 and RMSE ranging from 0.50 to 0.89 %. The best results were found in set C3, which had R² ranging from 0.65 to 0.78, and the lowest RMSE, ranging from 0.38 to 0.39 %, highlighting the C3-SNV model (R² = 0.78, RMSE = 0.38 %) obtained the highest degree of prediction accuracy. Lucà et al. (2017)

found that, in homogeneous samples of Italian soils, the predictive capacity of PLSR models increases with increases in the sample set.

However, under the conditions of the present study, the increase in the number of samples (from C3 to C2, for example) resulted in prediction losses due to the inclusion of samples with pedological differences compared to the IVS in the calibration of models (Figure 3), which resulted in an increase in spectral variance as presented in Figure 4. This behavior was expected since C3 models were calibrated exclusively with samples with origin and characteristics similar to the IVS. Lucà et al. (2017) also found that, in homogeneous samples of Italian soils in which the predictive capacity of PLSR models increases with increases in the sample set. This confirms the importance of representativeness of the samples in the calibration database and suggests that spectroscopic models should consider local samples for

calibration in order to achieve greater accuracy in SOC prediction (Grunwald et al., 2015).

The best adjustments and, consequently, the smallest prediction errors in the validation, occurred as a result of the smaller amplitude of the calibration set SOC data and pedological similarities between the calibration set samples and the IVS. It is fair to state that the main factor related to the accuracy of spectroscopic models is the amplitude of SOC content in the dataset, associated with pedological variation, thereby agreeing with Viscarra Rossel and Behrens (2010), Zeng et al. (2016) and Moura-Bueno et al. (2019).

Comparison of methods for quantification of SOM and nitrogen recommendation

Comparison of methods for quantification of SOM

The IVS SOM content quantified by REM had a
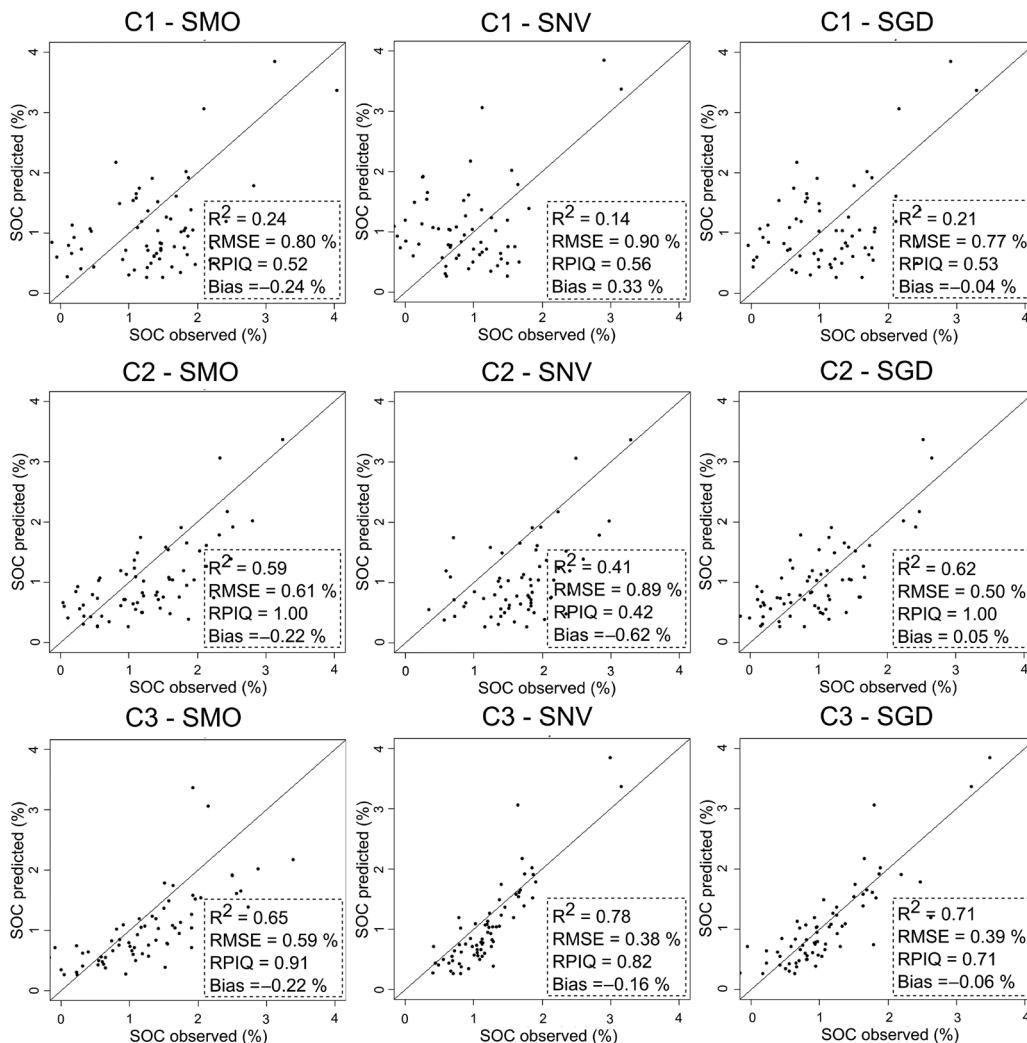
Figure 6 – Adjustment of the independent validation for models. In which: C1 = calibration set 1; C2 = calibration set 2; C3 = calibration set 3; SOC = soil organic carbon; SMO = smoothing; SGD = Savitzky-Golay derivative; SNV = standard normal variate; R^2 = coefficient of determination; RMSE = root mean squared error; RPIQ = ratio of performance to interquartile distance.

minimum value of 0.39 %, a maximum value of 5.83 % and an average of 1.57 %, while the SOM quantified by ROM presented a minimum value of 0.37 %, a maximum value of 4.90 % and average of 1.35 % (Figure 7). Overall, the ROM underestimated SOM values (Bias = – 0.22) (Table 2) and showed the largest deviations from the REM in the samples with the highest SOM content. Errors (samples: 35, 37 and 69) were found (Figure 8), which could not be checked because of the absence of repetitions in the routine laboratory. The RMSE verified between REM and ROM was 0.54 % and there were seven different (errors) recommendation classes for N (Table 2).

It is common to get errors in the determination of SOC by wet combustion, especially underestimation of high values since carbon oxidation is not complete and average correction factors become necessary (Nelson and Sommers, 1975). Due to a linear relationship with the dry combustion method, the Walkley-Black method can be reliably used within the range of 0.42 % to 8 % SOC (De Vos et al., 2007). Although the routine method is related to the Walkley-Black method, the sensitivity of the colorimetric method is very dependent on the amplitude of the samples in the calibration between the sulfochromic solution absorbance and the actual soil SOM values (Tedesco et al., 1995), which may increase the inaccuracy in quantifying SOM. The differences observed in the deviations of SOM estimates between ROM and SAM methods (Figure 8) showed that the spectroscopic models present greater accuracy. This result was also observed in a study by Souza et al. (2016), who subjected the wet combustion method employed in routine Brazilian laboratories to an accuracy and precision test and found lower accuracy than that obtained by spectroscopic models.

The highest degree of accuracy of the SAM was found in the C3-SNV model, with 0.58 % RMSE and six class placement errors (Figure 8). Overall, SAM models overestimated SOM content (Table 2). The smallest amount of errors in placing the SOM content in the recommendation classes were verified for set C3, which ranged from 6 to 15 errors. Followed by C2, ranging from 6 to 30 errors, and set C1, which ranged from 10 to 19 errors. The C3-SNV and C2-SGD models had the smallest number of errors, both with 6 errors (Figure 8). As regards the spectral preprocessing, the smallest errors were achieved by SGD, which ranges from 6 to 17 errors, followed by SNV, which ranges from 6 to 30 errors, and SMO, which ranges from 10 to 19 errors
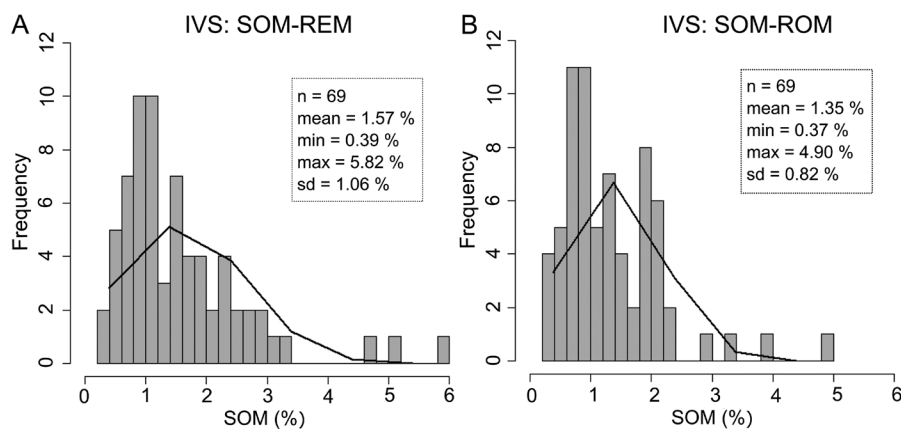

Figure 7 – Histograms and descriptive statistics of SOM data. In which: IVS = independent validation set; SOM = soil organic matter; n = number of samples; min = minimum value; max = maximum value; sd = standard deviation; REM = reference method; ROM = routine method.

Table 2 – Statistics of the applicability test for data from routine method (ROM) and predicted by spectral analysis method (SAM).

Preprocessing	SMO			SNV			SGD		
Data set	RMSE	Errors	Bias	RMSE	Errors	Bias	RMSE	Errors	Bias
	%		%	%		%	%		%
C1	0.89	19	0.41	1.36	10	–0.50	1.16	17	0.05
C2	0.92	13	0.34	1.34	30	0.93	0.75	6	–0.08
C3	0.65	15	0.38	0.58	6	0.24	0.59	7	0.09
Data set	RMSE	Errors	Bias						
	%		%						
Routine SOM	0.54	7	–0.22						

In which: C1 = calibration set 1; C2 = calibration set 2; C3 = calibration set 3; SOM = soil organic matter; SMO = smoothing; SGD = Savitzky-Golay derivative; SNV = standard normal variate; RMSE = root mean squared error.
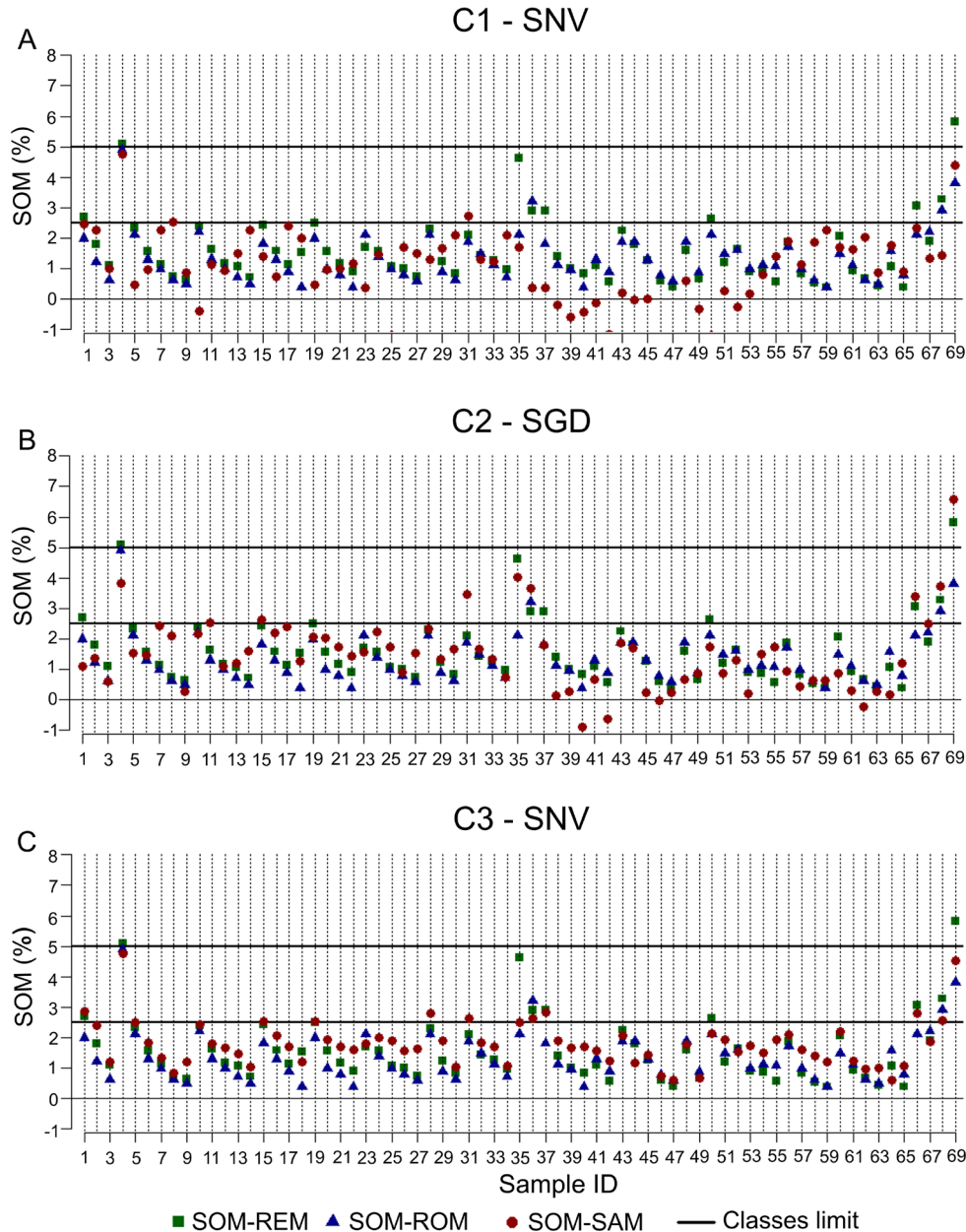
Figure 8 – Deviations verified between SOM analyzed by Walkley-Black method (reference), colorimetric method (in routine laboratory) and predicted by the best models of each sample set for each IVS sample. In which: C1 = calibration set 1; C2 = calibration set 2; C3 = calibration set 3; SOM = soil organic matter; SGD = Savitzky-Golay derivative; SNV = standard normal variate; REM = reference method; ROM = routine method; SAM = spectral analysis method.

(Figure 8). However, considering the RMSE of each preprocessing, SMO stood out, ranging from 0.65% to 0.89 %, followed by SGD, ranging from 0.59 % to 1.16 % and SNV, ranging from 0.58 % to 1.36 % (Table 2). The more complex preprocessing resulted in fewer placement errors; however, they increased the SOM RMSE. Thus, it can only be inferred that the best preprocessing is the one that best fits the data. It is thus recommended to perform preliminary tests for each data set.

Overall, SAM spectroscopic models overestimated the SOM content (Table 2). However; there are cases of underestimation, which, although not predominant, are not isolated and were found in practically all the models. This was more frequently observed in sets C1 and C2 and with a very low frequency in the best model with the best independent validation (C3-SNV). Prediction of negative values were found mainly in sets C1 and C2 and for samples with very low SOM content. This may be

an effect of the SOC range from the calibration sets C1 and C2 that have a higher mean than the IVS. However, a constant calibration and quality control of spectral models are necessary in order to ensure the reliability and applicability of SAM predicted SOC values.

Nitrogen Recommendation Simulation

Considering the IVS SOM data obtained by the REM, ROM and SAM methods, we obtained different amounts of recommended N. According to the SOM values obtained by the REM, the average recommendation was 102 kg N ha^{-1}. For ROM, the estimated average recommendation was 104 kg N ha^{-1} with RMSE of 6.34 kg N ha^{-1}. For SAM, using the most accurate spectroscopic model (C3-SNV), 102 kg N ha^{-1} with RMSE of 5.90 kg N ha^{-1} was recommended. The ROM provided a higher total N recommendation than the other methods due to systematic underestimation of SOM content, while the SAM presented a total N similar to the REM (Figure 9A).

As regards the urea (45 % N) recommendation, the amount was 227 kg ha^{-1} for the REM, 231 kg ha^{-1} for the ROM (RMSE of 14.15 kg ha^{-1}), and 227 kg ha^{-1} for SAM (RMSE of 13.11 kg ha^{-1}) (Figure 9B). It means a cost of 79.33 USD ha^{-1} for REM, 80.88 USD ha^{-1} (RMSE of 4.95 USD ha^{-1}) for ROM, and 79.33 USD ha^{-1} (RMSE of 4.58 USD ha^{-1}) for SAM, (Figure 9C). The SAM had satisfactory results for the N recommendation, with errors lower than the ROM, currently used to evaluate the SOM content for N recommendation for crops. The SAM, in addition to providing an improvement in the accuracy of the recommendation, compared to the ROM, would significantly reduce the SOC with chemical soil analysis (Viscarra Rossel and McBratney, 2008) as well as being 90 % more economical than the Walkley-Black method in the determination of SOC (O'Rourke and Holden, 2011). Therefore, there is room for its use in routine laboratories, considering it is a faster method, with lower final SOC and proven effectiveness, provided that calibrated models that contemplate the conditions and characteristics of the samples to be analyzed are used.

The SAM had satisfactory results for the N recommendation, with errors lower than the ROM, and is currently used to evaluate the SOM content for N fertilizer recommendation for crops. The SAM, in addition to providing an improvement in the accuracy of the recommendation, compared to the ROM, significantly reduced the SOC with chemical soil analysis (Viscarra Rossel and McBratney, 2008) as well as being 90 % more economical than the Walkley-Black method in the determination of SOC (O'Rourke and Holden, 2011). Therefore, there is room for its use in routine laboratories, considering it is a faster method, with lower final SOC and proven effectiveness, provided that calibrated models meet the conditions and possess the characteristics of the samples to be analyzed are used.

Conclusions

The SOC predicted by spectral techniques proved to be able to replace the measured by routine chemical methods without loss quality on the recommendation of nitrogen fertilizer, provide the models meet the conditions and possess the characteristics of the samples to be analyzed are used.

The sample set composed only by samples with pedological similarities with the independent validation set showed higher accuracy for calibration and validation. We recommend adjusting local models and calibrating them with samples similar to the ones that are going to be predicted.

Preprocessing techniques more complex resulted in better assignments into the classes for assessment of SOM content for N recommendation. However, we suggest performing preliminary tests and using spectral preprocessing techniques adapted to each data set.
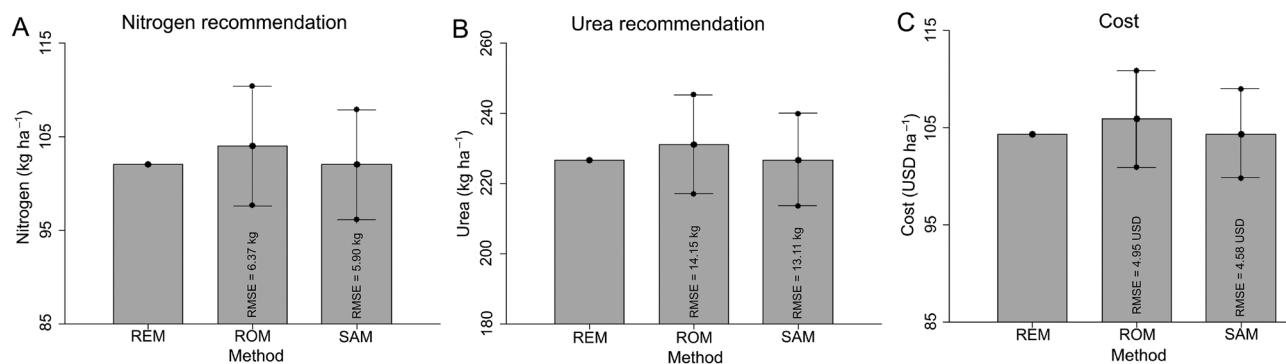
Figure 9 – N recommendation (A); urea recommendation (B) and respective cost in urea (C) in the independent validation set (IVS) simulating the representation of one (1) ha; SOM = soil organic matter; REM = soil organic matter content obtained from soil organic carbon determined by the reference method; ROM = soil organic matter content obtained by the routine analysis method; SAM = soil organic matter content obtained by the prediction model using diffuse reflectance spectroscopy.

Acknowledgements

Authors' Contributions

Conceptualization: Dalmolin, R.S.D.; Rosin, N.A.; Horst-Heinen, T.Z.; Moura-Bueno, J.M. Silva-Sangoi, D.V.; Silva, L.S. **Data acquisition:** Rosin, N.A.; Horst-Heinen, T.Z.; Moura-Bueno, J.M. **Data analysis:** Rosin, N.A.; Horst-Heinen, T.Z.; Moura-Bueno, J.M. **Design of methodology:** Dalmolin, R.S.D.; Rosin, N.A.; Horst-Heinen, T.Z.; Moura-Bueno, J.M. **Writing and editing:** Dalmolin, R.S.D.; Rosin, N.A.; Horst-Heinen, T.Z.; Moura-Bueno, J.M.; Silva-Sangoi, D.V.; Silva, L.S.

References

Alvares, C.A.; Stape, J.L.; Sentelhas, P.C.; Gonçalves, J.L.M., Sparovek, G. 2013. Köppen's climate classification map for Brazil. Meteorologische Zeitschrift 22: 711-728.

Araújo, S.R.; Wetterlind, J.; Demattê, J.A.M.; Stenberg, B. 2014. Improving the prediction performance of a large tropical vis-NIR spectroscopic soil library from Brazil by clustering into smaller subsets or use of data mining calibration techniques. European Journal of Soil Science 65: 718-729.

Bailey V.L.; Bond Lamberty B.; Deangelis K.; Grandy A.S.; Hawkes C.V.; Heckman K.; Lajtha K.; Phillips R.P.; Sulman B.N.; Todd–Brown K.E.O.; Wallenstein M.D. 2017. Soil carbon cycling proxies: Understanding their critical role in predicting climate change feedbacks. Global Change Biology 24: 895-905.

Boddey, R.M.; Jantalia, C.P.; Conceição, P.C.; Zanatta, J.A.; Bayer, C.; Mielniczuk, J.; Dieckowz, J.; Santos, H.P.; Denardin, J.E.; Aita, C.; Giacomini, S.; Alves, B.J.R. 2010. Carbon accumulation at depth in Ferralsols under zero-till subtropical agriculture. Global Change Biology 16: 784-795.

Companhia Nacional de Abastecimento [CONAB]. 2019a. Agricultural Information Portal. Agricultural Observatory - Grains: Historical Series = Portal de Informações Agropecuárias. Observatório Agrícola – Grãos: Série Histórica. CONAB, Brasília, DF, Brazil (in Portuguese).

Companhia Nacional de Abastecimento [CONAB]. 2019b. Agricultural Information Portal. Production Costs: Agricultural Inputs = Portal de Informações Agropecuárias. Custos de Produção: Insumos Agropecuários. CONAB, Brasília, DF, Brazil (in Portuguese).

Conforti, M.; Matteucci, G.; Buttafuoco, G. 2018. Using laboratory Vis-NIR spectroscopy for monitoring some forest soil properties. Journal of Soils and Sediments 18: 1009-1019.

De Vos, B.; Lettens, S.; Muys, B. Deckers, J.A. 2007. Walkley-Black analysis of forest soil organic carbon: recovery, limitations and uncertainty. Soil Use and Management 23: 221-229.

Dotto, A.C.; Dalmolin, R.S.D.; Ten Caten, A.; Grunwald, S. 2018. A systematic study on the application of scatter-corrective and spectral-derivative preprocessing for multivariate prediction of soil organic carbon by Vis-NIR spectra. Geoderma 314: 262-274.

Empresa Brasileira de Pesquisa Agropecuária [EMBRAPA]. 2004. Soil Correction and Fertilization = Correção do Solo e Adubação. 2ed. Embrapa Cerrado, Planaltina, DF, Brazil (in Portuguese).

Flores, C.A.; Alba, J.M.F.; Natchigall, S.D.; Cunha, H.N.; Moura-Bueno, J.M.; Deobald, G.A.; Pedron, F.A.; Flores, J.P.M.; Dalmolin, R.S.D.; Bernich, E.B.; Lange, M.C.; Nascimento, P.R.R.; Lopes, R.T. 2015. Semi-detailed soil survey, city of Santa Maria, state of Rio Grande do Sul = Levantamento semidetalhado de solos, município de Santa Maria, estado do Rio Grande do Sul. p. 11-79. In: Flores, C.A; Alba, J.M.F. Edaphic zoning of crops for the city of Santa Maria/RS, aiming at territorial planning = Zoneamento edáfico de culturas para o município de Santa Maria/RS, visando o ordenamento territorial. Embrapa Clima Temperado, Pelotas, RS, Brazil (in Portuguese).

Global Soil Partnership. 2017. Global Soil Organic Carbon Map – Leaflet. FAO, Rome, Italy.

Grunwald, S.; Vasques, G.M.; Rivero, R.G. 2015. Fusion of soil and remote sensing data to model soil properties. p. 1-109. In: Sparks, D.L., ed. Advances in Agronomy. Elsevier, Amsterdam, The Netherlands.

Horst, T.Z.; Dalmolin R.S.D.; Ten Caten A.; Moura-Bueno, J.M.; Cancian, L.C.; Pedron, F.A.; Schenato, R.B. 2018. Edaphic and topographic factors and their relationship with dendrometric variation of Pinus taeda L. in a high altitude subtropical climate. Revista Brasileira de Ciência do Solo 42: e0180023.

IUSS Working Group. 2015. World Reference Base for Soil Resources 2014: International Soil Classification System for Naming Soils and Creating Legends for Soil Maps. FAO, Rome, Italy. (World Soil Resources Reports, 106).

Knox, N.M.; Grunwald, S.; McDowell, M.L.; Bruland, G.L.; Myers, D.B.; Harris, W.G. 2015. Modelling soil carbon fractions with visible near-infrared (VNIR) and mid-infrared (MIR) spectroscopy. Geoderma 239-240: 229-239.

Kooistra, L.; Wehrens, R.; Leuven, R.S.E.W.; Buydens, L.M.C. 2001. Possibilities of visible-near-infrared spectroscopy for the assessment of soil contamination in river floodplains. Analytica Chimica Acta 446: 97-105.

Li, S.; Ji, W.; Chen, S.; Peng, J.; Zhou, Y.; Shi, Z. 2015. Potential of VIS-NIR-SWIR Spectroscopy from the Chinese soil spectral library for assessment of nitrogen fertilization rates in the paddy-rice region, China. Remote Sensing 7: 7029-7043.

Lucà, F.; Conforti, M.; Castrignanò, A.; Matteucci, G.; Buttafuoco G. 2017. Effect of calibration set size on prediction at local scale of soil carbon by Vis-NIR spectroscopy. Geoderma 288: 175-183.

Instituto Agronômico de Campinas [IAC]. 1996. Fertilization and Liming Recommendations for the State of São Paulo = Recomendações de Adubação e Calagem para o Estado de São Paulo. 2ed. IAC, Campinas, SP, Brazil. (Boletim Técnico, 100) (in Portuguese).

Moura-Bueno, J.M.; Dalmolin, R.S.D.; Ten Caten A.; Dotto A.C.; Demattê J.A. 2019. Stratification of a local VIS-NIR-SWIR spectral library by homogeneity criteria yields more accurate soil organic carbon predictions. Geoderma 337: 565-581.

Nelson, D.W.; Sommers, L.E. 1975. A rapid and accurate procedure for estimation of organic carbon in soil. Proceedings of Indiana Academy of Science 84: 456-462.

Padmanabhi, A.D.; Deshmukh, R.R. 2017. Hyperspectral analysis of soil total nitrogen using PLSR method: a review. International Journal of Theoretical and Applied Mechanics 12: 357-364.

Pinheiro, E.F.M.; Ceddia, M.B.; Clingensmith, C.M.; Grunwald, S.; Vasques, G.M. 2017. Prediction of soil physical and chemical properties by visible and near-infrared diffuse reflectance spectroscopy in the central Amazon. Remote Sensing 9: 293.

Ramos, P.V.; Dalmolin, R.S.D.; Marques Júnior, J.; Siqueira, D.S.; Almeida, J.A.; Moura-Bueno, J.M. 2017. Magnetic susceptibility of soil to differentiate soil environments in southern Brazil. Revista Brasileira de Ciência do Solo 41: 160-189.

Rinnan, A.; Berg, F.V.; Engelsen, S.B. 2009. Review of the most common pre-processing techniques for near-infrared spectra. Trends in Analytical Chemistry 28: 1201-1222.

Samuel-Rosa, A.; Dalmolin, R.S.D.; Miguel, P. 2013. Building predictive models of soil particle-size distribution. Revista Brasileira Ciência do Solo 37: 422-430.

Santos, H.G.; Jacomine, P.K.T.; Anjos, L.H.C.; Oliveira, V.A.; Lumbreras, J.F.; Coelho, M.R.; Almeida, J.A.; Araújo Filho, J.C.; Oliveira, J.B.; Cunha, T.J.F. 2018. Brazilian Soil Classification System. 5ed. Embrapa, Brasília, DF, Brazil.

Savitzky, A.; Golay, M.J.E. 1964. Smoothing and differentiation of data by simplified least-squares procedures. Analytical Chemistry 36: 1627-1639.

Sociedade Brasileira de Ciência do Solo [SBCS]. 1999. Approximation 5: Recommendations for the Use of Concealers and Fertilizers in Minas Gerais = 5ª Aproximação: Recomendações para o Uso de Corretivos e Fertilizantes em Minas Gerais. SBCS, Viçosa, MG, Brazil (in Portuguese).

Sociedade Brasileira de Ciência do Solo [SBCS]. 2016. Liming and Fertilizing Manual for the States of Rio Grande do Sul and Santa Catarina = Manual de Calagem e Adubação para os Estados do Rio Grande do Sul e de Santa Catarina SBCS, Viçosa, MG, Brazil (in Portuguese).

Sociedade Brasileira de Ciência do Solo [SBCS]. 2019. Fertilization and Liming Manual for the State of Paraná = Manual de Adubação e Calagem para o Estado do Paraná. 2ed. SBCS, Londrina, PR, Brazil (in Portuguese).

Souza, A.M.; Filgueiras, P.R.; Coelho, M.R.; Fontana, A.; Winkler, T.C.B.; Valderramac, P.; Poppi. R.J. 2016. Validation of the near infrared spectroscopy method for determining soil organic carbon by employing a proficiency essay for fertility laboratories. Journal of Near Infrared Spectroscopy 24: 293-303.

Tedesco, M.J.; Gianello, C.; Bissani, C.A.; Bohnen, H.; Volkweiss, S.J. 1995. Soil, Plant and other Materials Analysis = Análise de Solo, Plantas e outros Materiais. 2.ed. UFRGS, Porto Alegre, RS, Brazil (Soil Technical Bulletin, 5) (in Portuguese).

Terra, F.S.; Demattê, J.Á.M.; Viscarra-Rossel, R.A. 2015. Spectral libraries for quantitative analyses of tropical Brazilian soils: comparing vis–NIR and mid-IR reflectance data. Geoderma 255: 81-93.

Vasques, G.M.; Grunwald, S.; Sickman, J.O. 2008. Comparison of multivariate methods for inferential modeling of soil carbon using visible/near-infrared spectra. Geoderma 146: 14–25.

Viscarra Rossel, R.A.; McBratney, A.B. 2008. Diffuse reflectance spectroscopy as a tool for digital soil mapping. p. 165-172. In: Hartemink, A.E.; McBratney, A.B.; Mendonça-Santos, M.L. Digital soil mapping with limited data. Springer, Berlin, Germany.

Viscarra Rossel, R.A.; Behrens, T. 2010. Using data mining to model and interpret soil diffuse reflectance spectra. Geoderma 158: 46-54.

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

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

Wijewardane, N.K.Y.; Wills, S.; Loecke, T. 2016. Prediction of Soil Carbon In The Conterminous United States: Visible And Near Infrared Reflectance Spectroscopy Analysis Of The Rapid Carbon Assessment Project. Soil Science Society of America Journal 80: 973-982.

Wolski, M.S.; Dalmolin, R.S.D.; Flores, C.A.; Moura-Bueno, J.M.; Ten Caten, A.; Kaiser, D.R. 2017. Digital soil mapping and its implications in the extrapolation of soil-landscape relationships in detailed scale. Pesquisa Agropecuária Brasileira 52: 633-642.

Yeomans, J.C.; Bremner, J.M. 1988. A rapid and precise method routine determination of organic carbon in soil. Communications in Soil Science and Plant Analysis 19: 1467-1476.

Zeng, R.; Zhao, Y.-G.; Li, D.-C.; Wu, D.-W.; Wei, C.-L.; Zhang, G.-L. 2016. Selection of "local" models for prediction of soil organic matter using a regional Soil Vis-NIR spectral library. Soil Science 181: 13-19.Zhang, G.-L.; Liu, F.; Song, X.-D. 2017. Recent progress and future prospect of digital soil mapping: a review. Journal of Integrative Agriculture 16: 2871-2885. https://doi.org/10.1016/S2095-3119(17)61762-3

Zuur, A.F.; Ieno, E.N.; Elphick, C.S. 2010. A protocol for data exploration to avoid common statistical problems. Methods in Ecology and Evolution 1: 3-14. https://doi.org/10.1111/j.2041-210X.2009.00001.x