

Special Issue: CONBEA 2024

Scientific Paper

Doi: <http://dx.doi.org/10.1590/1809-4430-Eng.Agric.v45nespe120240173/2025>

DEVELOPING PREDICTION MODELS FOR SOIL ORGANIC MATTER CONTENT USING HYPERSENSITIVE DATASETS IN VARIOUS CROP ROTATION SYSTEMS

**Amanda S. Reis^{1*}, Marcos R. Nanni¹, Marlon Rodrigues², Glaucio L. A. A. dos Santos¹,
Weslei A. Mendonça¹, Caio A. de Oliveira¹, Karym M. de Oliveira¹**

^{1*}Corresponding author. Universidade Estadual de Maringá/Maringá - PR, Brasil.
E-mail: asreis@uem.br | ORCID ID: <https://orcid.org/0000-0001-5006-4887>

KEYWORDS

spectral signature,
multivariate statistics,
soil management,
remote sensing.

ABSTRACT

Soil organic matter (SOM) varies significantly along soil profiles, directly influencing soil fertility and structure. The aim of this study was to estimate SOM levels at various soil depths using VisNIR-SWIR spectroscopy combined with partial least squares regression (PLSR). The experiment was conducted on a Dystroferic Red Latosol under different crop rotation systems maintained since 1985. Soil samples were collected in March 2019 and stratified into eight layers (0–40 cm), totaling 384 samples. Spectral readings were obtained using a FieldSpec 3 Jr spectroradiometer, and SOM content was determined using a colorimetric method. The PLSR models demonstrated strong predictive capacity, especially for the 10-cm layer ($R^2 = 0.96$, RMSE = 0.74 g dm³) and for the full dataset ($R^2 = 0.82$, RMSE = 3.20 g dm³), with RPD values above 2, indicating excellent performance. The most relevant spectral bands were found in the ranges 580–590 nm, 870–930 nm, and 2400 nm. It was concluded that VisNIR-SWIR spectroscopy combined with PLSR is a promising and sustainable tool for estimating SOM, and is applicable across various soil layers and agricultural management contexts.

INTRODUCTION

Soil organic matter (SOM) content is one of the main indicators of tropical soil quality and plays a fundamental role in nutrient dynamics, physical structuring, water retention, and resilience to climate variability (Francos et al., 2021). Besides promoting soil fertility, SOM is directly involved in mitigating climate change through atmospheric carbon sequestration. However, the accurate and efficient quantification of SOM remains a challenge, especially on a large scale and at various soil depths.

Conventional laboratory methods, such as colorimetric and dry combustion methods, are time-consuming, require specialized labor, and rely on chemical reagents, making them unsuitable for large-scale, high-frequency applications. Therefore, diffuse reflectance

spectroscopy, in the visible, near-infrared, and shortwave infrared regions (VisNIR-SWIR), has emerged as a rapid, nondestructive, sustainable, and increasingly accurate technique for SOM content prediction, especially when it is combined with multivariate modeling algorithms (Aiwa et al., 2023; Wu et al., 2024).

The use of hyperspectral sensors, both imaging and non-imaging, combined with robust statistical techniques, such as partial least squares regression (PLSR), allows the identification of specific spectral bands associated with soil organic compounds (Shen et al., 2020; Zhang et al., 2017; Moura-Bueno et al., 2019). These bands, often located in the ranges of 580–590 nm, 870–930 nm, and around 2400 nm, reflect the presence of C-H, O-H, and N-H functional groups, characteristic of organic matter.

¹ Universidade Estadual de Maringá/Maringá - PR, Brasil.

² Universidade Tecnológica Federal do Paraná/Medianeira - PR, Brasil.

Area Editor: Wellington Gonzaga do Vale

Received in: 9-25-2024

Accepted in: 5-23-2025

Despite technological advances, the spatial and vertical variability of SOM content, especially in tropical soils under various management systems, still limits the development of generalizable models. This calls for studies in specific environments with well-documented land use history and detailed sampling stratification throughout the soil profile.

Thus, the objective of this study was to estimate SOM content at various depths using VisNIR-SWIR spectroscopy combined with PLSR regression in a long-term experiment under various crop rotation systems in southern Brazil. This study aimed to contribute toward the advancement of rapid and high-resolution soil quality monitoring methods in tropical agricultural environments.

MATERIAL AND METHODS

The experiment was conducted at the experimental farm of COAMO Agroindustrial Cooperative, located in Campo Mourão, Paraná State, Brazil (24°05'41.6"S and 52°21'31.5"W). According to the Köppen classification, the climate is humid subtropical (Cfa), with an average annual temperature of 20°C, annual rainfall of 1601 mm, and relative humidity of 78% (data obtained from the local meteorological station). The soil in the experimental area is classified as clayey-textured dystroferic red latosol according to the Brazilian Soil Classification System (EMBRAPA, 2018).

The plots used in this study are part of a long-term experiment initiated in 1985 through a partnership between the Brazilian Agricultural Research Corporation (EMBRAPA) and COAMO. The treatments included crop rotation under no-tillage (NT) and conventional systems, as well as crop succession under NT. The winter–summer crop sequences listed below, were repeated every four years. Only treatment T5 was managed using conventional tillage with soil disturbance before summer planting. The others were under NT.

T1: forage radish maize/wheat–soybean/wheat–soybean / wheat–soybean

T2: forage radish maizemaize/wheat–soybean/second-crop maize–soybean / wheat–soybean

T3: oat–maize/second-crop corn–soybean/second-crop corn–soybean/wheat–soybean

T4: lupine–maize, oat–soybean, wheat–soybean, and wheat–soybean

T5: second-crop maize–maize+brachiaria/wheat–soybean / oat–soybean / wheat–soybean

T6: vetch–maize/wheat–soybean/wheat–soybean / wheat–soybean

T7: second-crop maize–maize + brachiaria/wheat–soybean/oat–soybean / wheat–soybean

T8: forage pea–maize/lupine–maize/second-crop maize–soybean/wheat–soybean

T9: oat+radish–soybean/second-crop maize–soybean / wheat–soybean / second-crop maize–soybean

T10: wheat–maize / second-crop maize–soybean / second-crop maize–wheat soybean/wheat–soybean

T11: wheat–soybean/wheat–wheat-soybean/wheat–wheat-soybean/wheat–soybean

T12: second-crop maize–maize/second-crop maize–soybean/ second-crop maize–soybean/second-crop maize–soybean

Soil samples were collected in March 2019 at the beginning of the winter crop season. Samples were taken at eight different depth layers, covering 0–40 cm in 5 cm increments, giving a total of 384 samples (12 treatments × 4 replicates × 8 layers). After collection, samples were oven-dried, sieved, and analyzed for soil organic carbon using the colorimetric method (Raij et al., 2001), with a conversion factor of 1.724 to estimate SOM content.

The 384 soil samples were ground and sieved through a 2-mm mesh to homogenize moisture and surface roughness effects (Epiphânio, 1992), before they were placed in Petri dishes with a diameter of 9 cm and a height of 1.5 cm. Spectral readings were obtained using a FieldSpec 3 Jr spectroradiometer, which covers the spectral range 350–2500 nm. The device was configured to take 30 readings for each soil sample to generate an average spectral curve for each soil sample. Samples were individually positioned 8 cm from the fiber optic sensor, which has a reading area of approximately 2 cm². A 650 W lamp, used as the light source, was positioned 35 cm above the platform at a 30° angle to the horizontal plane (Cezar et al., 2012). The equipment was calibrated using a white reference panel with 100% reflectance (Spectralon) (Figure 1). Each sample was subjected to three sequential readings with 120° rotations between them to accurately capture material variation.

Statistical analyses were based on average spectral readings, according to the methodology proposed by Nanni & Demattê (2006). After spectral data collection, the ViewSpecPro software was used to apply the “splice correction” procedure to eliminate discontinuities (offsets) at the junctions of the spectroradiometer’s operating ranges: VisNIR (350–1100 nm) and SWIR (1100–2500 nm).

Due to noise, the curve edges were removed, resulting in a working spectral range of 380–2450 nm. Prior to statistical analysis of the reflectance data, the spectral curves were smoothed to reduce noninformative inflections and improve quality of the analysis. The Savitzky-Golay algorithm was applied to all curves as a pre-processing step, to remove the scattering effects caused by diffuse reflectance and to reduce baseline shifts, peak overlap, and other factors detrimental to the signal-to-noise ratio (Tong et al., 2015).

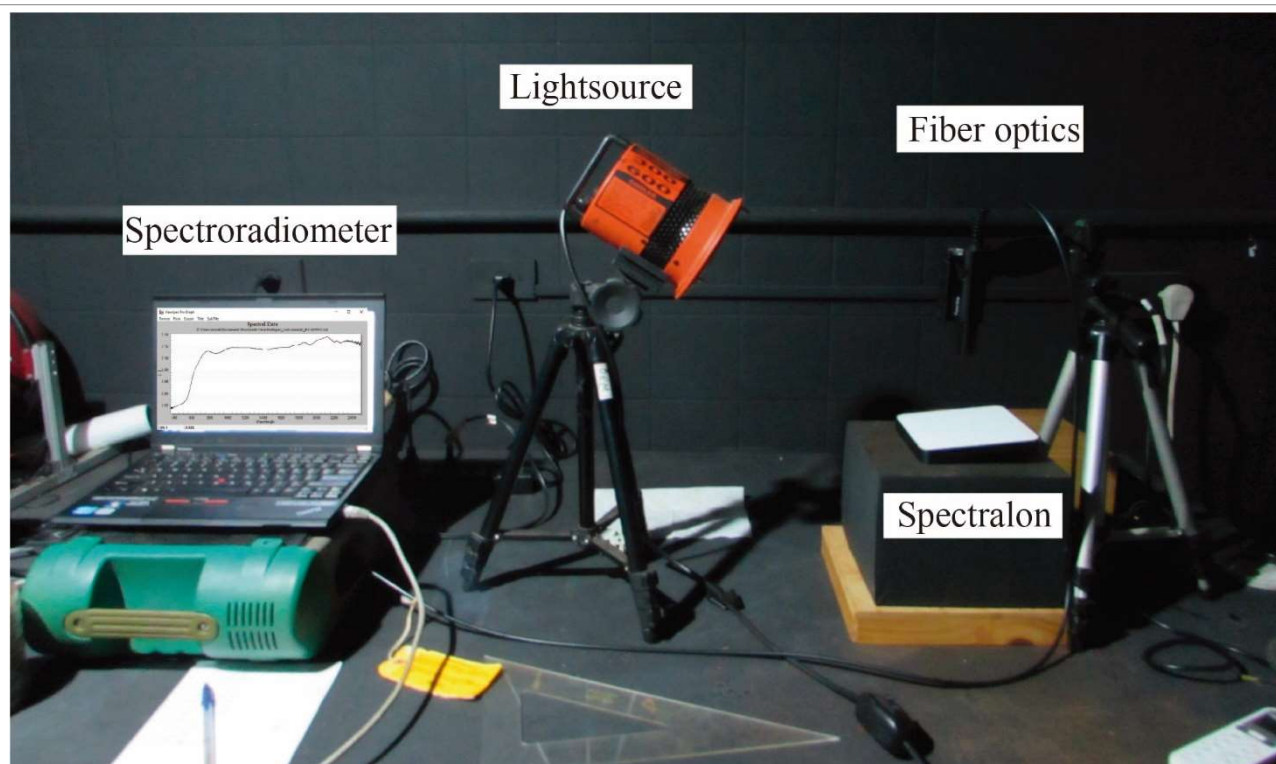


FIGURE 1. FieldSpec 3 Jr in the laboratory for acquiring spectral data from soil samples.

MULTIVARIATE STATISTICAL ANALYSES

Principal Component Analysis (PCA)

PCA was performed using the software *Unscrambler X 10.3* ($p \leq 0.05$) to identify sample groups with homogeneous spectra and enable the clustering of similar individuals. This technique transforms a set of original variables into another set of the same dimension, which is known as principal components (PCs), with minimal loss of information. Each PC was independent and was estimated to retain the maximum amount of variance from the original dataset.

The first principal component (PC1) corresponds to the largest eigenvalue and describes the greatest amount of variance among the samples in the X-space. When this set of samples can no longer be explained by PC1, a second component orthogonal to the first is used, and so on (Almeida, 2009). Determining the correct number of PCs to use is fundamental because the predictive values of these models depend directly on the number of principal components selected (Liebmann et al, 2010).

Partial Least Squares Regression

PLSR was also conducted using *Unscrambler X 10.3* to develop a prediction model for SOM content based on reflectance. This method extracts information from the spectral data matrix (wavelengths) and correlates it with reference data (SOM content). It assumes that the dependent variable can be estimated by a linear combination of explanatory variables (Wang et al., 2018).

Therefore, the calibration model must be optimized through cross-validation to obtain the smallest number of PLS factors that yield models with higher coefficients of determination (R^2) and RPD (Ratio of Performance to Deviation), and lower root mean square error (RMSE) values (Viscarra Rossel et al., 2006).

To construct the model, the 384 samples were randomly divided into two sets: calibration (70%–269 samples), which were used to develop the PLSR model, and external validation (30%–115 samples), which were used to test it. Data distribution was verified using Hotelling's T^2 and leverage models to identify outliers.

The predictive capacity of the model was evaluated based on R^2 (equation 1), RMSE (equation 2), and bias (equation 3) from the calibration and cross-validation phases, as well as the RPD from the prediction phase (equation 4), as shown in the equations below:

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

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

$$BIAS = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i) \quad (3)$$

$$RPD = \frac{DP}{SEP} \quad (4)$$

RESULTS AND DISCUSSION

Descriptive statistics for SOM content are presented in Table 1. SOM values ranged from 13.21 g dm⁻³ to 52.32 g dm⁻³, with a SD of 7.13 g dm⁻³ for the entire dataset. As the depth increased, the SD tended to decrease, reflecting a lower variability in SOM content in the deeper layers.

The coefficient of variation (CV) indicated moderate variability across the datasets, according to the classification by Qi-Yong et al. (2014). Among the 384 soil samples, 85.68% had SOM values above 20 g dm⁻³, significantly influencing the soil's spectral behavior. The 30-cm layer did not meet the statistical assumptions and was therefore excluded from the table.

TABLE 1. Minimum, maximum, mean, standard deviation (SD), and coefficient of variation (CV) of soil organic matter determined by the conventional method.

Layer (cm)	No. of observations	Min (g dm ⁻³)	Max (g dm ⁻³)	Mean (g dm ⁻³)	SD (g dm ⁻³)	CV (%)
0–40	384	13.21	52.32	26.32	7.13	27
5	48	19.92	52.33	39.56	6.81	17
10	48	22.63	42.91	31.03	4.31	14
15	48	21.68	44.66	27.82	3.73	13
20	48	19.60	30.78	25.03	2.45	10
25	48	17.05	28.54	22.79	2.77	12
35	48	14.50	35.25	22.52	3.77	16
40	48	13.22	26.78	21.13	2.68	13

SD: standard deviation; CV: coefficient of variation

VISUAL ANALYSIS OF THE SOIL REFLECTANCE SPECTRA

Figure 2 shows the average reflectance spectra for each treatment aggregated from the 0–40 cm layers. Although the differences were subtle, they became more evident in the near-infrared (NIR) region, particularly for treatments T7 and T8, which had the lowest SOM contents

(24.85 g dm⁻³ and 24.15 g dm⁻³, respectively) and higher reflectance values. This behavior is expected, as a lower SOM content tends to increase the spectral reflectance (Francos et al., 2021). Although SOM does not have specific absorption bands, it affects the albedo and shape of the reflectance spectrum. Several studies describe SOM's contribution to the spectral behavior of soils (Dotto et al., 2014; Jia et al., 2017; Moura-Bueno et al., 2019).

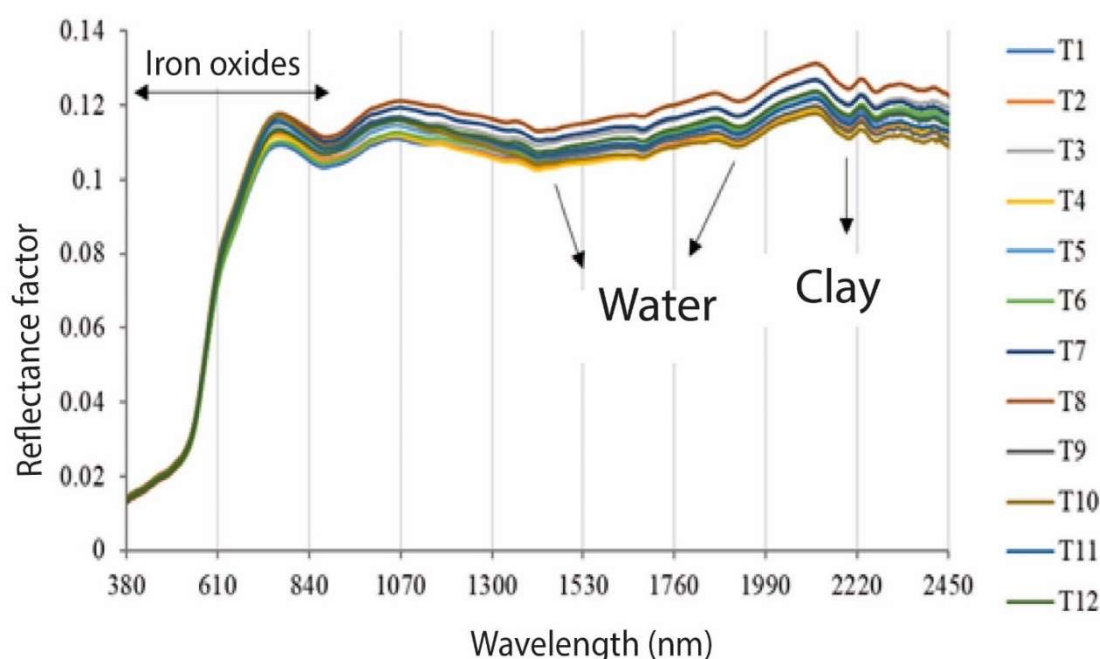


FIGURE 2. Averages of the soil spectral curves of each treatment, from 0–40 cm layer.

Generally, the spectra exhibited similar shapes and absorption features (Figure 2). In the visible region (380–700 nm), absorption is mainly associated with iron-containing minerals, which indicate the presence of goethite and hematite (Drăguț & Dornik, 2016), which are common in Latosols. In the NIR–SWIR region (700–2450 nm), weak tones and combination bands arise due to the stretching and bending of functional group bonds (Viscarra Rossel et al., 2006).

Well-defined absorption features associated with O–H functional groups were observed around 1400 and 1900 nm, and those associated with kaolinite presence were observed near 2200 nm (Demattê & Garcia, 1999). An inflection near 2300 nm, linked to molecular vibrations of C–

H, C–O, C=O, C–N, and N–H groups, associated with SOM (Viscarra Rossel & Behrens, 2010), was also observed.

PRINCIPAL COMPONENT ANALYSIS (PCA)

Following the visual comparison of the soil spectral curves, PCA was conducted using the full dataset ($n = 384$) corresponding to the 0–40 cm layer and using the individual layers ($n = 48$) (Figure 3). This analysis was performed to reduce the number of variables (wavelengths) into a set of new, derived variables (principal components or factors) that summarize the original SOM content values.

Figure 3 shows the percentage of variance explained by each principal component (PC). For the full dataset, PC1, PC2, and PC3 accounted for 89%, 8%, and 3% of the

variance, respectively. Similar proportions of variance were observed in the other sets, indicating the potential to discriminate the SOM within each spectral dataset.

Based on surface (0–10 cm) and subsurface (10–30 cm) soil samples, Jiang et al. (2017) identified differences

sample spectra using PCA, with almost 98% of the variance being explained by models tailored to each depth. They observed that surface spectra largely overlapped with subsurface spectra, indicating spectral similarity between the two groups.

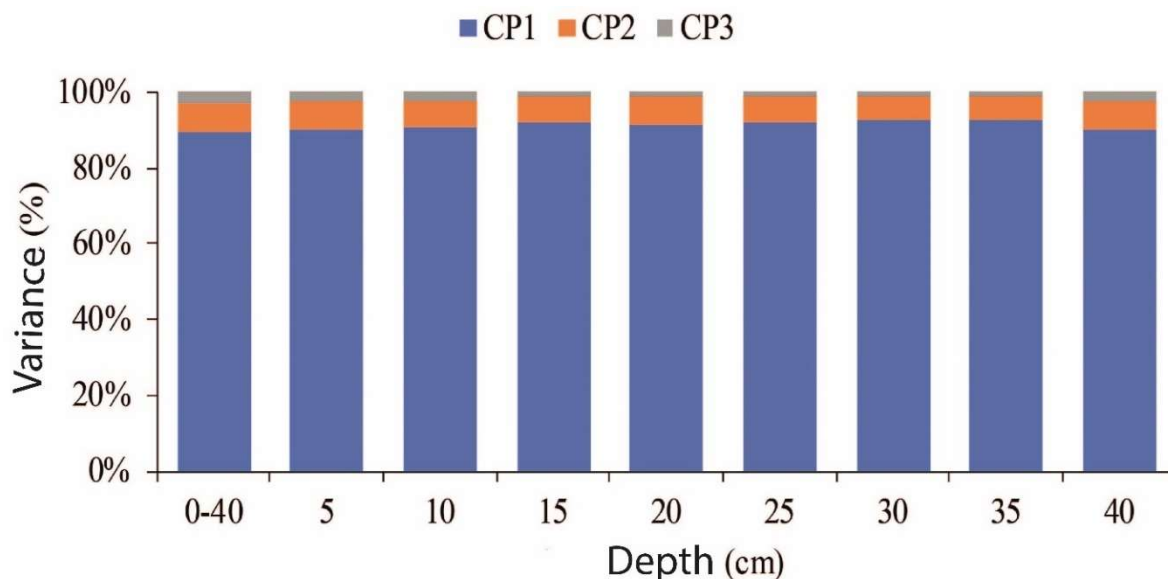


FIGURE 3. Principal component (CP) analysis of the 0–40 layer ($n=384$) and in the stratified layers ($n=48$).

PREDICTION MODELS FOR SOIL ORGANIC MATTER CONTENT

Based on the dataset composed of reflectance spectral curves and SOM values, a predictive model was developed using PLSR. Models were constructed for each individual layer ($n = 48$) and for the total dataset ($n = 384$), as shown in Table 2. The 15- and 35-cm layers were excluded because the number of PLS components for these layers was insufficient to explain the data variance, preventing adequate model fitting. This limitation is commonly related to cross-validation methods that select fewer components to prevent overfitting, which may reduce the model's ability to fully capture the data structure (Deng et al., 2015). Further studies should consider more robust machine learning algorithms for such cases.

In all datasets, the calibration phase produced higher correlation (r) and coefficients of determination (R^2) than the cross-validation phase, resulting in higher RMSE values. The 10-cm layer exhibited the best calibration results: $r = 0.98$, $R^2 = 0.96$ and the lowest RMSE (0.74 g dm^{-3}). This suggests that surface layers provide better precision for estimating SOM content. These findings can support further research using imaging sensors to map areas with different soil management practices and contribute to soil quality maintenance.

RMSE values varied across datasets, with the lowest being recorded for the 10-cm layer (0.74 g dm^{-3}) and the

highest being recorded for the full dataset (3.20 g dm^{-3}). These differences may be attributed to the sample size, specific characteristics of each data set, and model parameters (Lazzaretti et al., 2020). In all cases, bias was close to zero, indicating that the model did not present systematic errors (Cezar et al., 2019).

These results are consistent with those of Jaconí (2011), who used spectroscopy to analyze carbon in sugarcane soils and developed a model with the following parameters: $R^2 = 0.95$ and $\text{RMSE} = 1.53 \text{ g dm}^{-3}$. Similarly, Shi et al. (2014) developed a China-based spectral library for SOM content prediction and obtained a model with an R^2 of 0.90 and RMSE of 3.66 g dm^{-3} . Comparable results were also reported by Cezar et al. (2019) and Zhang et al. (2017) who constructed models with R^2 values of 0.86 and 0.90, respectively.

When assessing model quality by RPD, the 10-cm layer and full dataset (0–40 cm) showed the best performance, with values of 2.03 and 2.07, respectively. RPD values above 2 indicate excellent regression models for soil analysis (Xu et al., 2021), suggesting that the models developed in this study are excellent.

Finally, the developed model was applied to an external dataset comprising 115 samples (30% of the total), revealing a positive correlation between the estimated and observed SOM content values. Regression analysis yielded an R^2 of 0.83 and an RMSE of 2.69 g dm^{-3} , lower than values obtained during calibration and cross-validation.

TABLE 2. Statistical parameters of PLSR for predicting soil organic matter content.

Depth (cm)	Spectral samples	r_c	r_{cv}	R^2_c	R^2_{cv}	RMSE _c (g dm ⁻³)	RMSE _{vc} (g dm ⁻³)	BIAS _{cv}	RPD
0-40	384	0,90	0,89	0,82	0,79	3,20	3,44	-0,008	2,07
5	48	0,91	0,81	0,83	0,65	2,53	3,65	-0,01	1,86
10	48	0,98	0,85	0,96	0,72	0,74	2,12	0,077	2,03
20	48	0,94	0,66	0,89	0,44	0,78	1,79	-0,33	1,37
25	48	0,64	0,53	0,41	0,28	2,20	2,46	0	1,13
30	48	0,55	0,35	0,30	0,12	2,48	2,87	-0,112	1,31
40	48	0,86	0,62	0,74	0,38	1,48	2,37	-0,001	1,13

c: calibration; cv: cross validation.

Based on the results in Table 2, showing that the best-performing models were those based on the 10-cm layer and the full dataset, the latter was selected for developing a model that could predict the SOM content across various depths. When applied to the external dataset (115 samples), the model showed a strong positive

correlation between the estimated and observed SOM content values (Figure 4). Regression analysis yielded an R^2 of 0.83 and an RMSE of 2.69 g dm⁻³, which were lower than those in the calibration and cross-validation phases. Similar results were obtained by Nanni et al. (2021) and Reis et al. (2021), with R^2 values of 0.80 and 0.70, respectively.

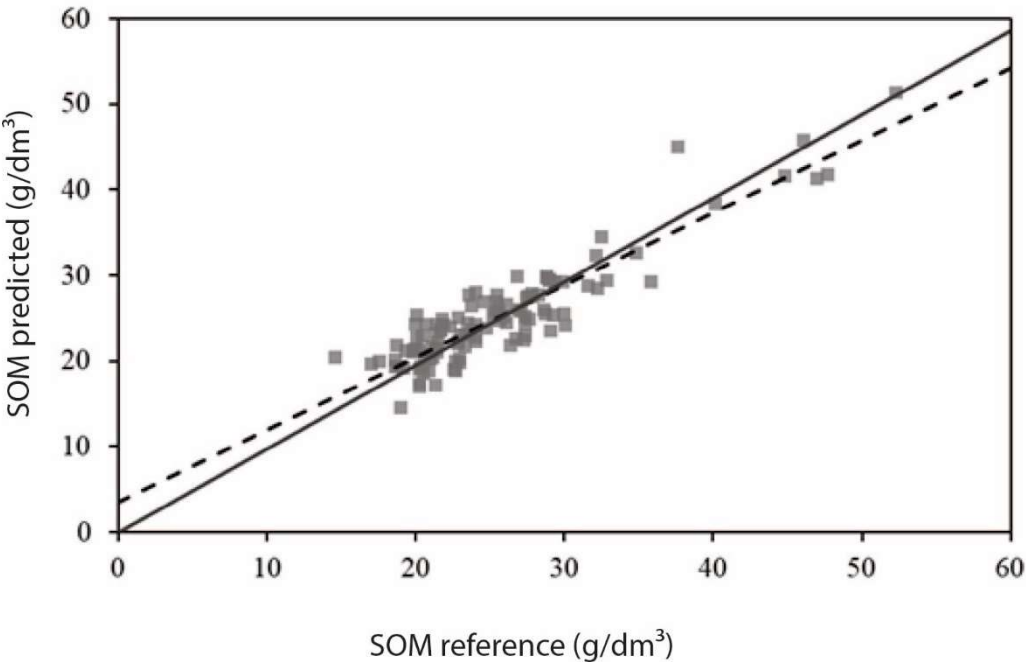


FIGURE 4. Scatter plot of soil organic matter (SOM) obtained during the external validation phase through PLSR. 1:1 line (solid line); Regression line (dashed line).

PLSR Regression Coefficients and Relevant Wavelengths

PLSR theory states that all wavelengths contribute to model construction to varying degrees; thus a two-dimensional graph was created to visualize the regression coefficients, as shown in Figure 5.

In the VisNIR region, the PLSR regression coefficients showed prominent peaks with positive correlations around 580–590 nm and 870–930 nm. These ranges were also highlighted by Shen et al. (2020) when

estimating SOM content using PLSR. The peaks in the 500–850-nm region correspond to the O–H and C–H combination bands typical of organic matter.

In the SWIR region, coefficients were more distributed, with a positive peak near 2400 nm, corresponding to C–H combination stretching zones (Stenberg, 2010). This region was also strongly linked to SOM by Wu et al. (2024) and Vibhute et al. (2018). Therefore, the 580–590, 870–930, and 2400 nm bands were highly relevant for SOM content prediction in this study.

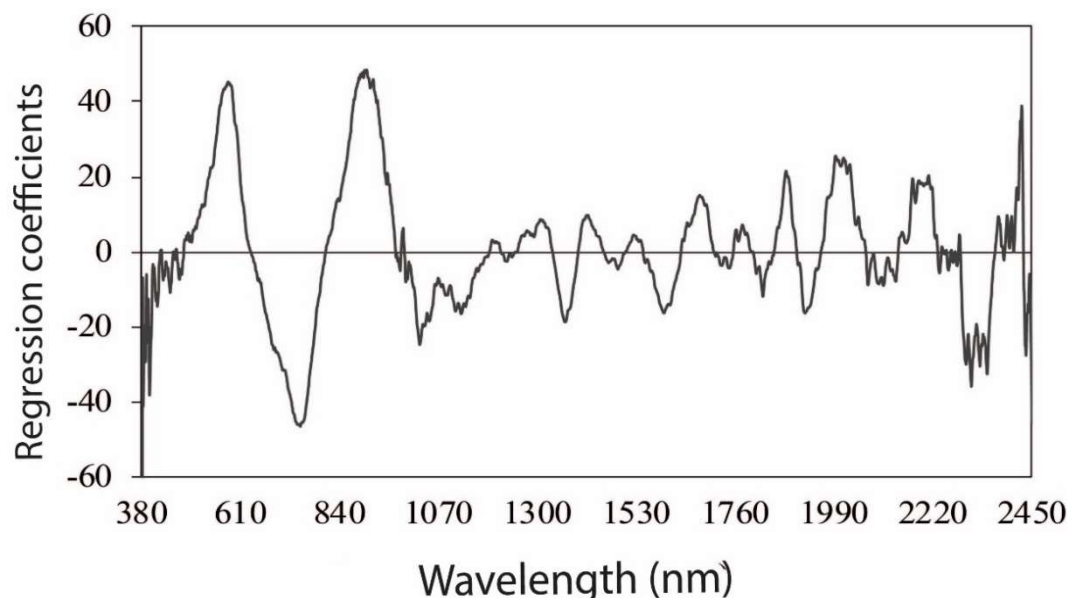


FIGURE 5. Coefficients of the PLSR regression model for predicting soil organic matter.

CONCLUSIONS

This study developed prediction models of SOM content under various soil management and crop rotation systems using soil reflectance. In all datasets, principal component analysis explained 97% of the spectral variance, with the NIR and SWIR wavelength ranges providing the most significant contributions.

The findings of this study showed that PLSR is a promising technique for estimating SOM content from reflectance spectra, especially for surface layers. The model demonstrated high accuracy when applied to the full dataset, with R^2 values of 0.82, and 0.79 in calibration and cross-validation, respectively, and an RPD of 2.07. The 10-cm layer, in particular, showed the best performance, with an R^2 of 0.96 and an RPD of 2.03, demonstrating the method's potential in areas with high organic matter contents.

These findings support the need for future research focusing on the application of orbital or suborbital sensors for monitoring areas under various soil management systems.

ACKNOWLEDGMENTS

COAMO Cooperativa Agroindustrial for providing the experimental area and for all assistance provided. To the National Council for Scientific and Technological Development – CNPq for granting the Junior Post-Doctorate scholarship.

REFERENCES

- Aiwa, M., Navqui, S. M. Z. A., Zhang, H., Li, L., Zhang, W., Awwad, F. A., Ismail, E. A. A., Khan, M. I., Reghavan, V., & Hu, J. (2023). AI and machine learning for soil analysis: An assessment of sustainable agricultural practices. *Bioresource and Bioprocessing*, 10, 90. <https://doi.org/10.1186/s40643-023-00710-y>
- Almeida, M. R. N. (2009). *Espectroscopia de Infravermelho Próximo com Transformada de Fourier (FTNIR) na Caracterização de Farinhas para Alimentação Pueril* [Master's dissertation, Universidade de Lisboa].
- Cezar, E., Nanni, M. R., Chicati, M. L., Junior, I. G. S., & Costa, A. C. S. (2012). Evaluation and quantification of silt, sand and clay fractions through their respective reflectances. *Revista Brasileira de Ciência do Solo*, 36(4), 1157–1165. <https://doi.org/10.1590/S0100-06832012000400010>
- Cezar, E., Nanni, M. R., Guerrero, C., Junior, C. A. S., Crusiol, L. G. T., Chicati, M. L., & Silva, G. F. C. (2019). Organic matter and sand estimates by spectroradiometry: Strategies for the development of models with applicability at a local scale. *Geoderma*, 340, 224 - 233. <https://doi.org/10.1016/j.geoderma.2019.01.021>
- Demattê, J. A. M., & Garcia, G. J. (1999). Evaluation of attributes of a Haplomox and a Tropohumult from Guarapuava, Paraná, by reflected energy. *Revista Brasileira de Ciência do Solo*, 23(2). <https://doi.org/10.1590/S0100-06831999000200018v>
- Deng, B., Yun, Y., Liang, Y., Cao, D., Xu, Q., Yi, L., & Huang, X. (2015). A new strategy to prevent overfitting in partial least squares models based on model population analysis. *Analytica Chimica Acta*, 880, 32–41. <https://doi.org/10.1016/j.aca.2015.04.045>
- Dotto, A. C., Dalmolin, R. S. D., Pedron, F. A., Caten, A., & Ruiz, L. F. C. (2014). Mapeamento digital de atributos: granulometria e matéria orgânica do solo utilizando espectroscopia de reflectância difusa. *Revista Brasileira de Ciência do Solo*, 38(6), 1663–1671. <https://doi.org/10.1590/S0100-06832014000600001>
- Drăguț, L., & Dornik, A. (2016). Land-surface segmentation as a method to create strata for spatial sampling and its potential for digital soil mapping. *International Journal of Geographical Information Science*, 30(7), 1359–1376. <https://doi.org/10.1080/13658816.2015.1131828>

- Embrapa. (2018). *Sistema Brasileiro de Classificação de Solos* (5^a ed.). Embrapa. <http://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/1094003>
- Epiphânio, J. C. N. (1992). *Comportamento espectral dos solos do estado de São Paulo* [Master's dissertation, INPE].
- Francos, N., Ogen, Y., & Bem-Dor, E. (2021). Spectral assessment of organic matter with different composition using reflectance spectroscopy. *Remote Sensing*, 13(8), 1549. <https://doi.org/10.3390/rs13081549>
- Jaconi, A. (2011). *O uso da espectroscopia no infravermelho próximo na quantificação de carbono em solos sob o cultivo de cana-de-açúcar* [Master's dissertation, Universidade de São Paulo]. <https://doi.org/10.11606/D.64.2011.tde-22032012-141049>
- Jia, X., Chen, S., Yang, Y., Zhou, L., Yu, W., & Shi, Z. (2017). Organic carbon prediction in soil cores using VNIR and MIR techniques in an alpine landscape. *Scientific Reports*, 7(1), 1–9. <https://doi.org/10.1038/s41598-017-02061-z>
- Jiang, Q., Li, Q., Wang, X., Wu, Y., Yang, X., & Liu, F. (2017). Estimation of soil organic carbon and total nitrogen in different soil layers using VNIR spectroscopy: Effects of spiking on model applicability. *Geoderma*, 293, 54–63. <https://doi.org/10.1016/j.geoderma.2017.01.030>
- Lazzaretti, B. P., Silva, L. S., Drescher, G. L., Dotto, A. C., Darines, B., & Nörnberg, J. L. (2020). Prediction of soil organic matter and clay contents by near-infrared spectroscopy - NIRS. *Ciência Rural*, 50(1). <https://doi.org/10.1590/0103-8478cr20190506>
- Liebmann, B., Friedl, A., & Varmuza, K. (2010). Applicability of near-infrared spectroscopy for process monitoring in bioethanol production. *Biochemical Engineering Journal*, 52(2–3), 187–193. <https://doi.org/10.1016/j.bej.2010.08.006>
- Moura-Bueno, J. M., Dalmolin, R. S. D., Caten, A., Dotto, A. C., & Demattê, J. A. M. (2019). Stratification of a local VIS-NIR-SWIR spectral library by homogeneity criteria yields more accurate soil organic carbon predictions. *Geoderma*, 337, 565–581. <https://doi.org/10.1016/j.geoderma.2018.10.015>
- Nanni, M. R., & Demattê, J. A. M. (2006). Spectral reflectance methodology in comparison to traditional soil analysis. *Soil Science Society of America Journal*, 70(2), 393–407. <https://doi.org/10.2136/sssaj2003.0285>
- Nanni, M. R., Demattê, J. A. M., Rodrigues, M., Santos, G. L. A. A., Reis, A. S., Oliveira, K. M., Cezar, E., Furlanetto, R. H., Crusiol, L. G. T., & Sun, L. (2021). Mapping particle size and soil organic matter in tropical soil based on hyperspectral imaging and non-imaging sensors. *Remote Sensing*, 13(9), 1782. <https://doi.org/10.3390/rs13091782>
- Qi-Yong, Y., Zhon-cheng, J., Wen-jun, L., & Hui, L. (2014). Prediction of soil organic matter in peak-cluster depression region using kriging and terrain indices. *Soil and Tillage Research*, 144, 126–132. <https://doi.org/10.1016/j.still.2014.07.011>
- Raij, B. van, Andrade, J. C., Cantarella, H., & Quaggio, J. A. (2001). *Análise química para avaliação da fertilidade de solos tropicais*. Instituto Agrônômico.
- Reis, A. S., Rodrigues, M., Santos, G. L. A. A., Oliveira, K. M., Furlanetto, R. H., Crusiol, L. G. T., Cezar, E., & Nanni, M. R. (2021). Detection of soil organic matter using hyperspectral imaging sensor combined with multivariate regression modeling procedures. *Remote Sensing Applications: Society and Environment*, 22. <https://doi.org/10.1016/j.rsase.2021.100492>
- Shen, L., Gao, M., Yan, J., Li, Z., Leng, P., Yang, Q., & Duan, S. (2020). Hyperspectral estimation of soil organic matter content using different spectral preprocessing techniques and PLSR method. *Remote Sensing*, 12(7), 1206. <https://doi.org/10.3390/rs12071206>
- Shi, Z., Wan, Q., Peng, J., Ji, W., Liu, H., Li, X., & Viscarra Rossel, R. A. (2014). Development of a national VNIR soil-spectral library for soil classification and prediction of organic matter concentrations. *Science China Earth Sciences*, 57(7), 1671–1680. <https://doi.org/10.1007/s11430-013-4808-x>
- Stenberg, B. (2010). Effects of soil sample pretreatments and standardised rewetting as interacted with sand classes on Vis-NIR predictions of clay and soil organic carbon. *Geoderma*, 158(1–2), 15–22. <https://doi.org/10.1016/j.geoderma.2010.04.008>
- Tong, P., Du, Y., Zheng, K., Wu, T., & Wang, J. (2015). Improvement of NIR model by fractional order Savitzky-Golay derivation (FOSGD) coupled with wavelength selection. *Chemometrics and Intelligent Laboratory Systems*, 143, 40–48. <https://doi.org/10.1016/j.chemolab.2015.02.017>
- Vibhute, A. D., Dhumal, R., Nagne, A., Surase, R., Varpe, A., Gaikwad, S. V., Kale, K., & Mehrotra, S. (2018). Assessment of soil organic matter through hyperspectral remote sensing data (VNIR spectroscopy) using PLSR method. *2nd International Conference on Man and Machine Interfacing*, 1–6. <https://doi.org/10.1109/MAMI.2017.8307888>
- Viscarra Rossel, R. A., Walvoort, D. J. J., McBratney, A. B., Janik, L. J., & Skjemstad, J. O. (2006). Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma*, 131(1–2), 59–75. <https://doi.org/10.1016/j.geoderma.2005.03.007>
- Viscarra Rossel, R. A., & Behrens, T. (2010). Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma*, 158(1–2), 46–54. <https://doi.org/10.1016/j.geoderma.2009.12.025>

- Xu, M., Chu, X., Fu, Y., Wang, C., & Shaohua, W. (2021). Improving the accuracy of soil organic carbon content prediction based on visible and near-infrared spectroscopy and machine learning. *Environmental Earth Sciences*, 80, 326. <https://doi.org/10.1007/s12665-021-09582-x>
- Wang, F., Gao, J., & Zha, Y. (2018). Hyperspectral sensing of heavy metals in soil and vegetation: Feasibility and challenges. *ISPRS Journal of Photogrammetry and Remote Sensing*, 136, 73–84. <https://doi.org/10.1016/j.isprsjprs.2017.12.003>
- Wu, M., Huang, Y., Zhao, X., Jin, J., & Ruan, Y. (2024). Effects of different spectral processing methods on soil organic matter prediction based on VNIR-SWIR spectroscopy in karst areas, Southwest China. *Journal of Soils and Sediments*, 24, 914–927. <https://doi.org/10.1007/s11368-023-03691-9>
- Zhang, Y., Biswas, A., Ji, W., & Adamchuk, V. I. (2017). Depth-specific prediction of soil properties in situ using vis-NIR spectroscopy. *Soil Science Society of America Journal*, 81(5), 993–1004. <https://doi.org/10.2136/sssaj2016.08.0253>