XXXIII Congresso Brasileiro de Ciência do Solo
Solos nos biomas brasileiros: sustentabilidade e mudanças climáticas
31 de julho à 05 de agosto - Center Convention - Uberlândia/Minas Gerais

SOIL CLASSIFICATION FROM VISIBLE/NEAR-INFRARED DIFFUSE REFLECTANCE SPECTRA AT MULTIPLE DEPTHS

Gustavo de Mattos Vasques[(1)]**; José Alexandre Melo Demattê**[(2)]**; Leonardo Ramírez López**[(3)]**; Fabrício da Silva Terra**[(2)]

[(1)] Researcher; Embrapa Solos, Rua Jardim Botânico, 1024, Rio de Janeiro, RJ, 22460-000, Brazil; gustavo@cnps.embrapa.br; [(2)] Professor; Student; Departamento de Ciência do Solo, Universidade de São Paulo, Av. Pádua Dias 11, Caixa Postal 9, Piracicaba, SP, 13418-900, Brazil; [(3)] Student; Geographisches Institut, Universität Tübingen, Rümelinstraße 19-23, 72070, Tübingen, Germany

Abstract – Visible/near-infrared diffuse reflectance spectroscopy (VNIRS) offers an alternative to conventional analytical methods to estimate various soil attributes. However, the use of VNIRS in soil survey and taxonomic classification is still underexplored. We investigated the potential use of VNIRS to classify soils in a region with variable soils, geology, and topography in southeastern Brazil. We combined principal component (PC) analysis, and multinomial logistic regression to classify 291 soils at the levels of suborder (second highest), and suborder with textural classification (STC), described in the field according to the Brazilian Soil Classification System. Soil visible/near-infrared (400-2500 nm) spectra were collected from three depth intervals (0-20, 40-60, and 80-100 cm), and combined in sequence to compose a pseudo multi-depth spectral curve, which was used to derive the classification models. The percent of correctly classified soils at the suborder level was 79% using 20 PCs, and 96% using 30 PCs. At the STC level, soils were correctly classified in 100%, and 78% of the cases using 20, and 30 PCs, respectively. Given the inherent complexity and variability within soil taxonomic groups, and in contrast the similarity among different groups, combining spectral data from different depths in multivariate classification offered a simple and inexpensive solution to adequately distinguish soils. This novel approach could improve soil classification and survey in a cost-efficient manner, supporting sustainable use, and management of tropical soils.

Keywords: multivariate classification, pedometrics, spectroscopy

INTRODUCTION

Soil survey has been traditionally done by combining the surveyor's interpretation of soil-landscape relationships and field expertise with supporting maps, aerial and/or satellite images, and soil data. Albeit this strategy has been used across countries to map soils at a range of geographical scales, currently it still does not fully incorporate newly available forms of data collection and interpretation. This is the case, for example, of proximally sensed data, including soil electrical conductivity, and diffuse reflectance.

Visible/near-infrared diffuse reflectance spectroscopy (VNIRS) has been applied to estimate many soil attributes used in soil survey, including organic matter, carbon, pH, macro- and micronutrients, water content, and others (Chang et al., 2001; Viscarra Rossel et al., 2006; Vasques et al., 2008; Du & Zhou, 2009; Stenberg et al., 2010). Because VNIRS uses little sample preparation and chemicals, and can be used to simultaneously estimate various soil attributes, it can reduce time and cost of analyses. In this case, gain obtained from VNIRS applies to data collection and analysis, but only indirectly to final soil classification and survey.

Therefore, to improve efficiency of soil classification, we propose a direct application of VNIRS to derive soil classes according to the Brazilian Soil Classification System (SiBCS; EMBRAPA, 2006). We hypothesized that the diffuse reflectance spectra of soils from three depth intervals could be used to classify soils at the suborder level with acceptable accuracy (> 80% agreement rate), as the SiBCS uses color for distinguishing soils at this taxonomic level.

MATERIALS AND METHODS

Soil sampling and field classification

The study was conducted near the city of Piracicaba, in the central-eastern part of the state of São Paulo, Brazil (Figure 1), in a region that has been primarily used for sugarcane production in the last 30 years. Mean annual precipitation, and temperature in the region are 1328 mm, and 21.6 ºC, respectively, while elevations vary from about 489 to 709 m, and slopes from 0 to 32%. Soils in the region are in most part derived from sandstone, siltstone, and shale, and less prominently from limestone, basalt, and colluvial deposits (Mezzalira, 1966).

A total of 291 soil profiles (11 pits, and 278 boreholes) were visited and classified in the field at the suborder level (second highest) according to the SiBCS (Table 1). Exceptionally, for Argissolo Vermelho soils, the *Latossólico* designation (as in PVL) was added to indicate the presence of an oxic horizon. Soil samples were taken at 0-20, 40-60, and 80-100 cm, and analyzed chemically and granulometrically according to Camargo et al. (1986). The clay content was used to refine soil suborder classification into five textural groups (Table 2) based on EMBRAPA (2006), thus constituting soil suborder with textural classification (STC) groups.

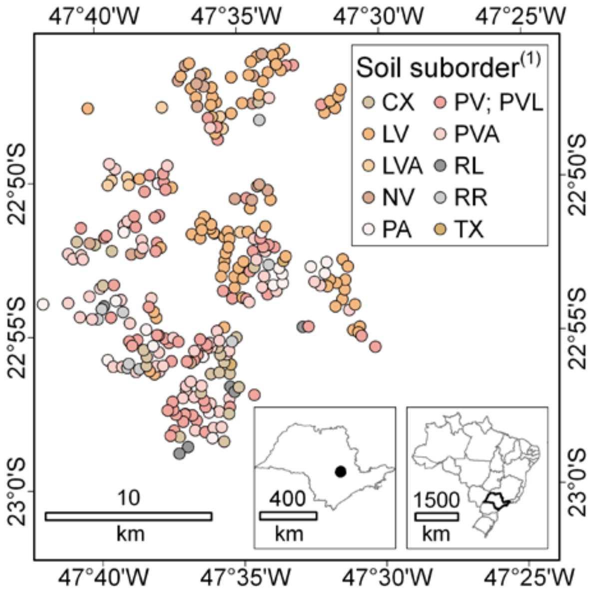

Figure 1. Soil sampling site locations within the state of São Paulo (left inset map), and Brazil (right inset map)
[(1)] CX, Cambissolo Háplico; LV, Latossolo Vermelho; LVA, Latossolo Vermelho-Amarelo; NV, Nitossolo Vermelho; PA, Argissolo Amarelo; PV, Argissolo Vermelho; PVA, Argissolo Vermelho-Amarelo; PVL, Argissolo Vermelho Latossólico; RL, Neossolo Litólico; RR, Neossolo Regolítico; TX, Luvissolo Háplico

Soil spectroscopy and multivariate classification

Visible/near-infrared (VNIR; 400-2500 nm) soil reflectance spectra were collected at the three sampled depth intervals from air-dried and sieved (2 mm) samples using a FieldSpec Pro sensor (Analytical Spectral Devices Inc., Boulder, CO), with 100 scans per sample, and resolution of 1 nm at 400-1100 nm, and 2 nm at 1100-2500 nm. Spectralon (Labsphere, North Sutton, NH) was used as white reference, and scanned at every 20 minutes (~15 samples).

The spectral curves containing 2100 bands were smoothed (Savitzky & Golay, 1964) using a third-order polynomial across five bands, and then reduced by simple averaging across ten bands, resulting in spectral curves containing 210 bands. For each soil profile, the pre-treated spectra from the three depth intervals were joined in sequence to create a pseudo *multi-depth spectral curve*. Thus, these multi-depth spectra contained 630 reflectance bands covering the three depths (210 bands each) seamlessly.

Soil classification at two taxonomic levels (suborder, and STC, respectively) was performed using multinomial logistic regression (MLR; Agresti, 2002). This method estimates the probability of a sample occurring in (or belonging to) each class of a finite group of classes, with final sample classification usually considering the class where the sample had the highest probability of occurrence.

We used principal component analysis (PCA; Harman, 1976) to reduce the number of independent variables, thus extracting the spectral information contained in the 630 reflectance bands into 30 principal components (PCs). Either the whole 30 PCs, or the first 20 PCs, were used as independent variables in the MLR models. In summary, this approach constituted a soil classification framework using MLR on the PCs of combined soil VNIR spectra from three depths.

RESULTS AND DISCUSSION

Multivariate classification at the suborder level

At the suborder level, the MLR models correctly classified 79% of the observations using 20 PCs, and 96% using the whole 30 PCs. These high agreement rates reflected the correlation between soil VNIR spectra and taxonomic classes. In effect, classification at the level of suborder takes into consideration not only the main pedogenetic features of soils, which are used to group soils at the order (highest) level, but also, and primarily among other factors, soil color to specifically separate soils within the same order into suborders. Among the soils used in this study (Table 1), only Cambissolo Háplico, Neossolo Litólico, and Neossolo Regolítico do not take soil color explicitly into consideration for suborder separation.

Considering the model with 30 PCs, the great majority of suborders (8 out of 11) were classified with 100% accuracy (Table 3). Combining soil spectra from multiple depths into a seamless spectral curve before PCA assured the inclusion of spectral information in the PCs to distinguish certain soil classes that otherwise could not be distinguishable based on the spectra of a single depth. Even so, some suborders were still incorrectly classified, including Latossolo Vermelho (LV), Nitossolo Vermelho (NV), and Argissolo Vermelho (PV), which were confounded for one another (Table 3).

These misclassifications were not surprising, given the great similarity in soil VNIR spectra (Figure 2), color, and other attributes among these soil types in the region. For example, the differences between NV and PV were subtle, requiring careful determination of the clay content of the A, and B horizons (including the B/A clay content ratio) to distinguish between them. Clay content is a property that only indirectly influences the VNIR spectra of soils, and thus small differences in clay content between these suborders were probably not captured by the VNIRS-MLR model. As another example, LV differs from NV and PV mostly in morphological, and physical attributes, whereas their VNIR spectra (Figure 2), and color are very similar, resulting in mixed-up classification of LV as NV or PV.

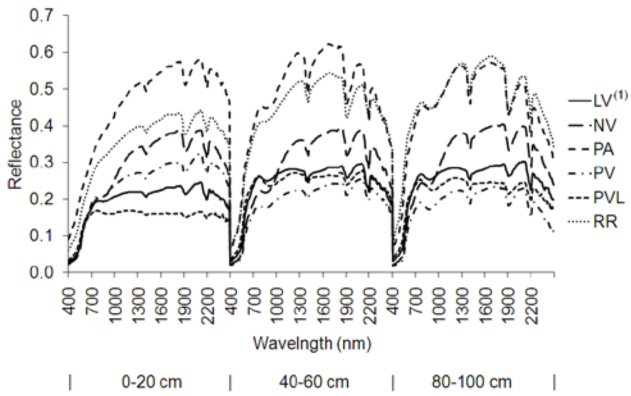

Figure 2. Pseudo multi-depth visible/near-infrared diffuse reflectance spectra of selected soils, after smoothing (Savitzky & Golay, 1964) using a third-order

polynomial across five bands, and averaging across ten bands
[(1)] LV, Latossolo Vermelho; NV, Nitossolo Vermelho; PA, Argissolo Amarelo; PV, Argissolo Vermelho; PVL, Argissolo Vermelho Latossólico; RR, Neossolo Regolítico

Similar agreement rates for validation (~95%) were obtained for Israeli soils using mid-infrared (2500-25,000 nm) attenuated total reflectance, and photoacoustic spectroscopy, respectively (Linker, 2008). The author used wavelet decomposition, and artificial neural networks to classify about 200 soil samples into five groups, and also observed great similarity of soil spectra among misclassified groups. In a similar study, Du et al. (2008) correctly classified 96% of the validation samples into the same five Israeli soil taxonomic groups using mid-infrared photoacoustic spectroscopy analyzed by PCA, and probabilistic neural networks. Even though the multivariate classification methods differed, in principle, high agreement rates were obtained in both cases (and also in this study) because soil infrared spectra contained information related to soil taxonomic evaluation.

Multivariate classification at the STC level

At the STC level, contrary to the suborder models, the MLR model derived from 20 PCs had a higher agreement rate (100%) than the one derived from 30 PCs (78%). The 20 PCs carried enough spectral information to correctly classify the soils, but also were general (i.e. representative) enough to avoid overfitting, which was probably the case with 30 PCs. Compared to suborders, STC groups were more specific and contained a smaller number of samples per group. This minimized the within-group variation, and facilitated distinguishing among groups with a smaller number of PCs (20 as opposed to 30). Nonetheless, soil spectra were still similar among STC groups (not shown).

We expected that the large number of STC groups (34), and the fact that they were defined based on the clay content, would hinder formulation of a good VNIRS-based classification model. On the contrary, interestingly VNIRS allowed producing a robust model to separate soil types, even those with very similar characteristics, flawlessly.

Our results corroborate those obtained by Nanni et al. (2004; 185 observations, 18 classes, 91% agreement rate), and Fiorio et al. (2010; 473 observations, 23 classes, 81% agreement rate), who also used VNIR spectra to classify soils at a taxonomic level equivalent to STC. Instead of PCA, they used stepwise discriminant analysis to select predictors among 22 bands (or band intervals), and 13 so-called height differences previously selected based on expert knowledge from two sampled depth intervals.

CONCLUSIONS

1. Visible/near-infrared diffuse reflectance spectroscopy and multinomial logistic regression offer a framework to rapidly classify soils at the suborder level with acceptable accuracy;

2. Combining soil spectral curves from three depth intervals allows to better distinguish soils, even those with similar characteristics.

ACKNOWLEDGEMENTS

We thank the Soil Science Department at the University of São Paulo, and the Coordination for the Improvement of Higher Education Personnel (CAPES) for the support.

REFERENCES

AGRESTI, A. Categorical data analysis. 2.ed. New York, John Wiley & Sons, 2002. 734p.

CAMARGO, O.A.; MONIZ, A.C.; JORGE, J.A. & VALADARES, J.M.A.S. Métodos de análise química, mineralógica e física de solos do Instituto Agronômico de Campinas. Campinas, Instituto Agronômico, 1986. 94p.

CHANG, C.-W.; LAIRD, D.A.; MAUSBACH, M.J. & HURBURGH JR, C.R. Near-infrared reflectance spectroscopy–principal components regression analyses of soil properties. Soil Sci. Soc. Am. J., 65:480-490, 2001.

DU, C. & ZHOU, J. Evaluation of soil fertility using infrared spectroscopy: a review. Environ. Chem. Lett., 7:97-113, 2009.

DU, C.; LINKER, R. & SHAVIV, A. Identification of agricultural Mediterranean soils using mid-infrared photoacoustic spectroscopy. Geoderma, 143:85-90, 2008.

EMBRAPA (EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA). Sistema Brasileiro de Classificação de Solos. 2.ed. Brasília, EMBRAPA, 2006. 306p.

FIORIO, P.R.; DEMATTÊ, J.A.M.; NANNI, M.R. & FORMAGGIO, A.R. Diferenciação espectral de solos utilizando dados obtidos em laboratório e por sensor orbital. Bragantia, 69:453-466, 2010.

HARMAN, H.H. Modern factor analysis. 3.ed. Chicago, University of Chicago Press, 1976. 508p.

LINKER, R. Soil classification via mid-infrared spectroscopy. In: LI, D., ed. Computer and computing technologies in agriculture, volume 2. Boston, Springer, 2008. p.1137-1146.

MEZZALIRA, S. Folha geológica de Piracicaba. Folha SF-23-M-300. São Paulo, Instituto Geográfico e Geológico do Estado de São Paulo, 1966.

NANNI, M.R.; DEMATTÊ, J.A.M. & FIORIO, P.R. Análise discriminante dos solos por meio da resposta espectral no nível terrestre. Pesq. Agrop. Bras., 39:995-1006, 2004.

SAVITZKY, A. & GOLAY, M.J.E. Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem., 36:1627-1639, 1964.

SOIL SURVEY STAFF. Keys to Soil Taxonomy. 11.ed. Washington, United States Department of Agriculture, 2010. 338p.

STENBERG, B.; VISCARRA ROSSEL, R.A.; MOUAZEN, A.M. & WETTERLIND, J. Visible and near infrared spectroscopy in soil science. In: SPARKS, D.L., ed. Advances in agronomy, volume 107. Amsterdam, Academic Press, 2010. p.163-215.

VASQUES, G.M.; GRUNWALD, S. & SICKMAN, J.O. Comparison of multivariate methods for inferential modeling of soil carbon using visible-near-infrared spectra. Geoderma, 146:14-25, 2008.

VISCARRA ROSSEL, R.A.; WALVOORT, D.J.J.; MCBRATNEY, A.B.; JANIK, L.J. & SKJEMSTAD, J.O. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma, 131:59-75, 2006.

Table 1. Soil suborders as classified in the field according to the Brazilian Soil Classification System (SiBCS; EMBRAPA, 2006), and corresponding classes in Soil Taxonomy (Soil Survey Staff, 2010)

Abbreviation	SiBCS suborder	Soil Taxonomy class	Observations
CX	Cambissolo Háplico	Udepts	21
LV	Latossolo Vermelho	Udox	82
LVA	Latossolo Vermelho-Amarelo	Udox	8
NV	Nitossolo Vermelho	Udalfs, Udults	18
PA	Argissolo Amarelo	Udalfs, Udults	18
PV	Argissolo Vermelho	Udalfs, Udults	55
PVA	Argissolo Vermelho-Amarelo	Udalfs, Udults	59
PVL	Argissolo Vermelho Latossólico	Udalfs, Udults	13
RL	Neossolo Litólico	Lithic Udorthents, Lithic Udipsamments	6
RR	Neossolo Regolítico	Udorthents, Udipsamments	9
TX	Luvissolo Háplico	Udalfs	2
Total			291

Table 2. Soil textural classes used to refine suborder classes

Code	Textural class	Original name	Clay content (%)
1	Very clayey	Muito argilosa	> 60
2	Clayey	Argilosa	35-60
3	Medium-clayey	Médio-argilosa	25-35
4	Medium-sandy	Médio-arenosa	15-25
5	Sandy	Arenosa	< 15

Table 3. Soil suborder multivariate classification results

Observed[1]	Predicted[1]											Agreement rate (%)
	CX	LV	LVA	NV	PA	PV	PVA	PVL	RL	RR	TX	
CX	21	0	0	0	0	0	0	0	0	0	0	100
LV	0	78	0	1	0	3	0	0	0	0	0	95
LVA	0	0	8	0	0	0	0	0	0	0	0	100
NV	0	4	0	14	0	0	0	0	0	0	0	78
PA	0	0	0	0	18	0	0	0	0	0	0	100
PV	0	4	0	1	0	50	0	0	0	0	0	91
PVA	0	0	0	0	0	0	59	0	0	0	0	100
PVL	0	0	0	0	0	0	0	13	0	0	0	100
RL	0	0	0	0	0	0	0	0	6	0	0	100
RR	0	0	0	0	0	0	0	0	0	9	0	100
TX	0	0	0	0	0	0	0	0	0	0	2	100
Overall agreement rate (%)												96

[1] CX, Cambissolo Háplico; LV, Latossolo Vermelho; LVA, Latossolo Vermelho-Amarelo; NV, Nitossolo Vermelho; PA, Argissolo Amarelo; PV, Argissolo Vermelho; PVA, Argissolo Vermelho-Amarelo; PVL, Argissolo Vermelho Latossólico; RL, Neossolo Litólico; RR, Neossolo Regolítico; TX, Luvissolo Háplico