

Comissão 1.3 - Pedometria

USING NUMERICAL CLASSIFICATION OF PROFILES BASED ON VIS-NIR SPECTRA TO DISTINGUISH SOILS FROM THE PIRACICABA REGION, BRAZIL⁽¹⁾

Rodnei Rizzo⁽²⁾, José A. M. Demattê⁽³⁾ & Fabrício da Silva Terra⁽⁴⁾

SUMMARY

Considering that information from soil reflectance spectra is underutilized in soil classification, this paper aimed to evaluate the relationship of soil physical, chemical properties and their spectra, to identify spectral patterns for soil classes, evaluate the use of numerical classification of profiles combined with spectral data for soil classification. We studied 20 soil profiles from the municipality of Piracicaba, State of São Paulo, Brazil, which were morphologically described and classified up to the 3rd category level of the Brazilian Soil Classification System (SiBCS). Subsequently, soil samples were collected from pedogenetic horizons and subjected to soil particle size and chemical analyses. Their Vis-NIR spectra were measured, followed by principal component analysis. Pearson's linear correlation coefficients were determined among the four principal components and the following soil properties: pH, organic matter, P, K, Ca, Mg, Al, CEC, base saturation, and Al saturation. We also carried out interpretation of the first three principal components and their relationships with soil classes defined by SiBCS. In addition, numerical classification of the profiles based on the OSACA algorithm was performed using spectral data as a basis. We determined the Normalized Mutual Information (NMI) and Uncertainty Coefficient (U). These coefficients represent the similarity between the numerical classification and the soil classes from SiBCS. Pearson's correlation coefficients were significant for the principal components when compared to sand, clay, Al content and soil color. Visual analysis of the principal component scores showed differences in the spectral behavior of the soil classes, mainly among Argissolos and the others soils. The NMI and U similarity coefficients showed values of 0.74 and 0.64, respectively, suggesting good similarity between the numerical and SiBCS classes. For example, numerical

⁽¹⁾ Part of the master dissertation of the first author submitted at Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo - USP. Received for publication on January 31, 2013 and approved on November 5, 2013.

⁽²⁾ M.Sc., Escola Superior de Agricultura "Luiz de Queiroz" - ESALQ/USP. Av. Pádua Dias, 11. Caixa Postal 9. CEP 13418-900 Piracicaba (SP), Brazil. E-mail: rodnei.rizzo@gmail.com

⁽³⁾ Professor, Soil Scientist, Department of Soil Science Luiz de Queiroz College of Agriculture, University of São Paulo. Padua Dias Avenue, 11. CEP 13416-900 Piracicaba (SP), Brazil. E-mail: jamdemat@usp.br

⁽⁴⁾ Professor, Institute of Agricultural Science (ICA), Campi Unai, Federal University of Jequitinhonha and Mucuri Valleys (UFVJM). Rua Vereador João Narciso, 1380, Bairro Cachoeira. CEP 38610-000 Unai (MG). E-mail: fabricio.terra@ufvjm.edu.br

classification correctly distinguished Argissolos from Latossolos and Nitossolos. However, this mathematical technique was not able to distinguish Latossolos from Nitossolos Vermelho férricos, but the Cambissolos were well differentiated from other soil classes. The numerical technique proved to be effective and applicable to the soil classification process.

Index terms: soil reflectance, soil classification, principal component analysis, OSACA algorithm.

RESUMO: CLASSIFICAÇÃO NUMÉRICA DE PERFIS COM BASE EM ESPECTROS VIS-NIR PARA DISTINÇÃO DE SOLOS DA REGIÃO DE PIRACICABA, SP

Atualmente, o espectro de reflectância do solo é uma informação subutilizada em processos de classificação. Sendo assim, os objetivos deste trabalho foram avaliar as relações entre os espectros de solos e seus atributos físicos e químicos, identificar padrões espectrais para cada classe de solo e por fim analisar o uso da classificação numérica de perfis, conciliada a dados espectrais, na distinção de solos. Para tanto, foram estudados 20 perfis da região de Piracicaba, SP, sendo descritos morfológicamente e classificados até o 3º nível categórico do SiBCS. Amostras foram coletadas dos horizontes pedogenéticos e foram submetidas às análises granulométrica e química. Posteriormente, o comportamento espectral de cada solo foi obtido e submetido à análise por componentes principais. Os escores dos componentes principais foram utilizados em análise de correlação linear com pH, matéria orgânica, P, K, Ca, Mg, Al, CTC, V% e m%. Além da interpretação dos três primeiros componentes principais, foram estabelecidas relações desses com as classes de solo definidas pelo SiBCS. Ainda, a classificação numérica dos perfis foi procedida a partir do algoritmo OSACA, utilizando os dados espectrais como base. A relação entre a classificação do SiBCS e a classificação numérica, foi determinada pelo índice de Informação Mútua Normalizada (IMN) e o Coeficiente de Incerteza (U). Foram encontradas correlações significativas entre os escores dos componentes principais e a areia (0,78), a argila (-0,74), a cor do solo e o teor de Al (0,73). A representação gráfica dos componentes principais e sua interpretação visual indicou diferenças no comportamento espectral das classes de solo, principalmente entre os Argissolos e as demais classes. Já a classificação numérica dos perfis, com base nos espectros dos solos, obteve valores para os índices de informação mútua normalizada e coeficiente de incerteza de 0,74 e 0,64, respectivamente. Esses valores indicam que a classificação numérica possui boa relação com a estabelecida pelo SiBCS, sendo capaz, por exemplo, de distinguir Argissolos de classes como Latossolos e Nitossolos. Além disso, observou-se que tal técnica não é capaz de diferenciar Latossolos e Nitossolos Vermelho férricos; no entanto, Cambissolos foram agrupados corretamente. A técnica evidenciou-se eficiente, demonstrando sua aplicabilidade em processos de classificação de solo.

Termos de indexação: comportamento espectral do solo, classificação de solo, análise por componentes principais, algoritmo OSACA.

INTRODUCTION

Agriculture is considered one of the foundations of the Brazilian economy, providing employment and increasing foreign exchange reserves. However, in recent decades, authorities and even increasing community awareness have required production increases based on sustainable practices. Therefore, implementation of proper management procedures and prior knowledge of the agricultural environment become essential.

Soil is one of the most important constituents of the environment. It is not only a support for plants but also supplies water and nutrients. Thus, knowing the soil properties and their spatial

variability is essential for implementation of any management technique (Bhatti et al., 1991). Soil maps are one of the most used sources of information for evaluating soil spatial variability and, when available at an appropriate scale, they allow the user to identify physical, chemical and morphological properties. Therefore, these maps aid in the decision-making process for agricultural planning, e.g., indicating sites with periodic flooding or variations in soil depth.

According to Mendonça-Santos & Santos (2006), approximately 35 % of Brazilian territory, which represents 17 of 26 States, has soil maps in intermediate scales (1:100,000 - 1:600,000). However, soil maps covering the whole of national territory are found only in the exploratory and schematic scales.

These are called organizational maps and were mainly developed by the Brazilian Agricultural Research Organization (Embrapa) and the Agronomic Institute of Campinas (IAC). They are usually offered free of charge and made available in digital databases. Currently, to respond to the demand for more detailed maps, several private agricultural enterprises have hired professionals to develop these maps. However, this information is not available in the public domain. In addition, there is a lack of interest from governmental institutions in producing detailed maps, due to the complexity, cost, and, especially, time requirements (Ben-Dor et al., 2008).

According to Brown et al. (2006), the cost for conventional soil characterization, calculated by the U.S. National Soil Survey Center, is about US\$2,500 per pedon and requires 6-12 months to be accomplished. Giasson et al. (2006) estimated a cost of approximately US\$ 2.21 ha⁻¹ (scale of 1:50,000) and US\$ 0.817 ha⁻¹ (scale of 1:100,000) for mapping the soil in two municipalities in the state of Rio Grande do Sul, Brazil. To improve the mapping process, researchers have developed equipment to provide soil information in real time at lower costs (Viscarra Rossel et al., 2009). Since the 1980s, numerous initiatives have evaluated the potential of spectra in obtaining quantitative soil data and, in many cases, the results were promising (Rivero et al., 2007; Ben-Dor et al., 2008; Viscarra Rossel et al., 2009). Several researchers (Moran et al., 1997; Ben-Dor et al., 1999) have recommended the development of methods using optical sensors, which may lend support to both laboratory processes and routine fieldwork. Some studies attempted to use the spectrum through visual interpretation to describe the soil class (Ben-Dor et al., 2008; Bellinaso et al., 2010). However, this process requires knowledge of spectral patterns for each soil class, as well as good references for comparisons (complete spectral libraries), a condition that is not always possible.

In this regard, numerical classification of profiles could be useful as a technique to support soil classification based on spectral data. According to Campbell et al. (1970), numerical classification of profiles allows application of a numerical procedure to determinate similarity between soil profiles and subsequently identify the most representative groups for the dataset. Various methods are described in the literature, suggesting the relevance of this technique and its applicability (Rayner, 1966; King & Girard, 1988; Carré & McBratney, 2005). More recently, Carré & Jacobson (2009) proposed a method of numerical classification of soil profiles called the *Outil Statistique d'Aide à la Cartogénèse Automatique* (OSACA). To date, all studies carried out have considered soil quantitative information obtained from laboratory analyses or qualitative properties obtained from field observations. However, no study has been carried out taking into account soil spectral behaviors as input data for numerical classification.

Thus, this study tested numerical classification using only Vis-NIR spectra as input data to distinguish soil profiles, assessing the potential and limitations of this information to classify soils. In addition, we aimed to assess the correlations among principal component scores of soil spectra and soil properties, and conduct visual evaluation of soil profile spectral patterns and their relationship to soil classes in the Brazilian Soil Classification System (SiBCS).

MATERIAL AND METHODS

Study area and data sampling

The area of study is located between latitudes 22° 42' 49" and 23° 0' 15" S, and longitudes 47° 57' 3" and 47° 30' 15" W, corresponding to the municipality of Piracicaba, São Paulo, Brazil. Soil samples from 20 profiles, previously collected and studied by Bellinaso (2009), were used in the analyses. The soil profiles were described and sampled according to Santos et al. (2005). The samples were analyzed for soil particle size using a densimeter with hexametaphosphate as a dispersing agent (Camargo et al., 1986). The resin method was used to determine pH (in H₂O and KCl), organic matter content (OM), and P exchangeable, K, Ca, Mg, and Al content. Cation exchange capacity (CEC), bases and Al saturation were calculated (Raij et al., 2001). The soil profiles were classified up to the 3rd category level according to Embrapa (2013) based on the chemical and soil particle size data, as well as soil morphological description and field observations. The profiles were also classified according to World Reference Base for soil Resources (WRB) (IUSS Working Group WRB, 2007) and are presented along with the Brazilian classification. In this work, the profiles classified as Latossolos (LV and LVA) (SiBCS) corresponded to Ferrasols (WRB), the Argissolos (PA, PVA, PV) are related to the Acrisols, Nitossolos (NV) with the Nitisols and Cambissolos (CX) are related to Cambisols and Alisols.

Soil color was determined using a Minolta CR-300 colorimeter adjusted for the Munsell color system (Campos et al., 2003). The values were then converted to the RGB color system by the COLOSOL software to use the color as a numerical variable in statistical analysis (Viscarra Rossel et al., 2006a).

After that, the spectra of soil samples were obtained using the FieldSpec Pro spectrometer (Analytical Spectral Devices, Boulder, Colo.), which has spectral resolution of 1 nm and performs readings in wavelengths in the spectral range from 350 to 2500 nm. Geometric acquisition of the spectral data was based on perpendicular positioning of the sensor to the sample separated by 27 cm. The lighting was positioned at 61 cm from the sample, forming a zenith angle of 20°. The absolute reference standard was a white spectralon plate. To cover the soil surface

analyzed, the samples were measured in triplicate, and then an average of these three readings was obtained.

Principal Component Analysis of spectra

Due to multicollinearity in high spectral resolution data (Chang et al., 2001) and the long time requirement for computational processing, transformation of spectra has been necessary. We used Principal Component Analysis (PCA) (Wold, 1982) carried out with the Parles 3.1 software (Viscarra-Rossel, 2008) to summarize the soil spectral data. This software determines the principal components (PCs), which are given by the linear combination of the variables X_1, X_2, \dots, X_j (in this case, the spectral wavelengths):

$$PC_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ij}X_j$$

in which a_{ij} corresponds to the load of the variable j in the calculation of the principal component i .

The Principal Component (PC) is calculated to represent the greatest possible variance in the dataset respecting the restriction applied to the load calculations:

$$a_{i1}^2 + a_{i2}^2 + \dots + a_{ij}^2 = 1$$

The process is repeated until the total of the principal components calculated is equal to the original number of variables or reaches a pre-established number of components. At the end of the analysis, the PCA provides the eigenvectors, eigenvalues and the principal component scores. The elements of an eigenvector are the loads a_{ij} described above, also known as loadings (Viscarra-Rossel, 2008). In this case, a set of loadings is associated with each principal component and each loading is associated with an original variable.

The loading value indicates the contribution of each variable, in this case, the reflectance of each spectral wavelength for each principal component. This information is interpreted by observing the proximity of the loading values in relation to value 0, i.e., the closer the loading value is to 0, the lower the contribution. Similarly, the more positive or more negative the loading value is, the greater the contribution of the variable to the principal component.

The eigenvalues represent the variance explained by each principal component, which decrease from the first to the last principal component, generally expressed in percentage. Finally, the scores are one of the most important pieces of data, allowing the construction of ordination diagrams of samples (scatter plots). Visual analysis of these scatter plots allows the similarity between samples to be identified or even outliers to be located.

Numerical classification of profiles

To perform numerical classification of the soil profiles, the first four PCs were considered as variables, explaining more than 98 % of the data variability.

The OSACA algorithm (Carré & Jacobson, 2009) was used for numerical classification. This algorithm aims to classify the soil profiles using k-means clustering (Diday, 1971) by comparing soil characteristics, taking the position of their respective horizons into account. First, the comparison of profiles is performed two by two, i.e., it compares profile “A” and “B”, then “A” and “C”, and so forth. In these comparisons, the 1st horizon in profile “A” is compared to the 1st horizon in profile “B”, the 2nd horizon in “A” with the 2nd horizon in “B”, and so forth (Figure 1). The values of these comparisons are given by distance metrics, in this case, the Euclidean distance (Gauch, 1982). The average Euclidean distance between each two profiles, called pedological distance (Dped), is added to the k-means cluster algorithm to classify the profiles. The pedological distance is described as follows:

$$Dped(S_a, S_b) = \frac{\sum_{j=1}^{M_a} D_h(h_{a,j}, h_{b,j}) + \sum_{j=M_a+1}^{M_b} D_h(h_{a,M_a}, h_{b,j})}{M_b}$$

in which S_a and S_b are profiles A and B; $h_{i,j}$ is horizon j of profile i ; $D_h(h_{a,j}, h_{b,j})$ is the Euclidian distance between the j horizons of profiles A and B; M_a and M_b is the number of horizons of profile A and B, considering that profile B has a higher number of horizons.

Comparison between numerical and SiBCS classifications

The Brazilian Soil Classification System is used to cluster soil profiles with similar properties and provides important information for soil management and conservation procedures. Therefore, we checked for agreement between the OSACA and SiBCS classifications. Considering that SiBCS takes the diagnostic horizons into account, while the numerical method uses the pedogenetic horizons, we do not aim to validate the numerical method, but rather observe, from an established system, whether the groups generated by OSACA are consistent. In this case, soil classes up to the 3rd categorical level, as well as information on soil texture and chemistry, were used

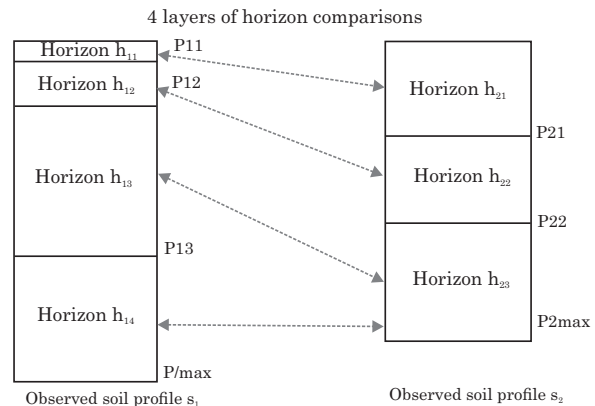


Figure 1. Illustration indicating the profile comparison method. Source: Carré & Jacobson (2009).

as a basis for comparison. Although the SiBCS classification does not take directly this information into account, it is important in soil management and conservation procedures. In addition, it is generally indicated next to the soil classes in maps.

The agreement between the SiBCS and the numerical classification was determined by the Normalized Mutual Information (NMI) and the Uncertainty Coefficient (U) indices. The NMI is used to describe the degree of dependence between original classes and numerical groups (Santos, 2009). In other words, this index is a measure of the disagreement between the groups obtained and the classes indicated (Liu & Navath, 2006). Therefore, random groups would have mutual information worth zero, while high values of this measurement would indicate high similarity. One of the possible formulations for NMI is defined as (Shi & Ghosh, 2003):

$$\tau_{\text{IMN}} = \frac{\sum_{i=1}^g \sum_{j=1}^l T_{ij} \times \log_2 \left(\frac{n \times T_{ij}}{|P_i| \times |L_j|} \right)}{\left[\sum_{i=1}^g |P_i| \times \log_2 \left(\frac{|P_i|}{n} \right) \right] \times \left[\sum_{j=1}^l |L_j| \times \log_2 \left(\frac{|L_j|}{n} \right) \right]}$$

in which P_i represents one of the g groups; L_j is one of the l classes; T_{ij} is the element of the contingency matrix that represents the number of profiles attributed to the P_i group, which are also members of L_j class; and n is the total number of profiles. In this case, the NMI value ranges from 0 to 1, and the higher this value, the greater the relationship between the original classes and the groups.

The U indicates if the variation of reference classification can be explained by the groups generated. Among the positive aspects of this coefficient, it does not penalize the algorithm if it generates groups that subdivide a class. Thus, if profiles belonging to a specific SiBCS class are subdivided into two groups, the U considers the division correct, without penalizing U. The index is defined as (Shannon & Weaver, 1963):

$$U(\bar{I} | J) = \frac{H(\bar{I}) - H(\bar{I} | J)}{H(\bar{I})}$$

in which $H(\bar{I})$ corresponds to the entropy of distributions and reference classification; and $H(\bar{I} | J)$ refers to the value of the conditional entropy. By applying the index, we obtain a value between 0 and 1, where higher values suggest more associated classes and groups.

RESULTS AND DISCUSSION

Soil profiles

The profiles were placed in different landscape positions, with elevation ranging from 438 to 781 m

(Figure 2). Those located on the east side of the study area in the highest places are derived from basaltic rocks. These soils showed high clay and iron oxide contents and, in some cases, high base saturation. An example was the profile P52 (Table 1) which corresponded to a Latossolo Vermelho (LV) férrico with a heavy clay texture. There were also soil profiles classified as Nitossolo Vermelho (NV), with clay or heavy clay texture and, in some cases, high iron oxide content (18 % < ferric < 36 %). The profiles located in the hillside areas with intermediate elevation (Figure 2) correspond to Argissolo Vermelho and Argissolo Vermelho-Amarelo, with texture ranging from clay to loam. These soils were derived from claystones and siltstones of the Corumbataí and Iratí formations (Mezzalira, 1966) and are represented by profiles P26 and P54 (Table 1).

The profiles in the central and western areas corresponded to the loam textured soils, with color ranging from Yellow-Red to Yellow. Profiles placed in flat areas were related mostly to Latossolos, while those allocated to gentle sloping areas corresponded to Argissolos. Soils in steeper slopes at lower altitudes (Figure 2) were classified as Cambissolo Háplico. The parental materials of these soils are claystones and sandstones from the same geological formations described above (Mezzalira, 1966). The representative profiles of these soils were identified as P32, P37, P40 and P49, which are shown in table 1.

Relationship between soil spectral data and physico-chemical properties

The relationship between soil properties and spectral data were evaluated based on Pearson correlation analysis (Table 2). In this analysis, soil properties determined in the laboratory were compared to the PCA scores of the spectral data. In addition, the PCA loadings were plotted and observed to indicate which spectral range affected each PC score (Figure 3). By evaluating the PC1 loadings (Figure 3a), we observed that the scores were affected by the entire spectral reflectance intensity (albedo) along the wavelengths after 600 nm. Thus, variability in the PC1 was related to the spectral albedo, corroborating results found by Galvão et al. (2001) and Bellinaso et al. (2010). The correlation between clay content and PC1 scores showed a value of -0.74 (Table 2). Considering that PC1 was related to spectral intensity, this result is in accordance with those obtained by Bowers & Hanks (1965), Stoner (1979) and Demattê et al. (2004), which indicated that variations in soil particle sizes caused changes in the reflectance of the soil spectral behavior. Sand contents are also related to PC1, showing a correlation of 0.78 (Table 2). In this case, reduction in the clay content of the soil samples leads to a proportional increase in the percentage of other fractions, e.g., sand.

The loadings showed that PC2 scores were primarily related to 350-600 nm (Figure 3b).

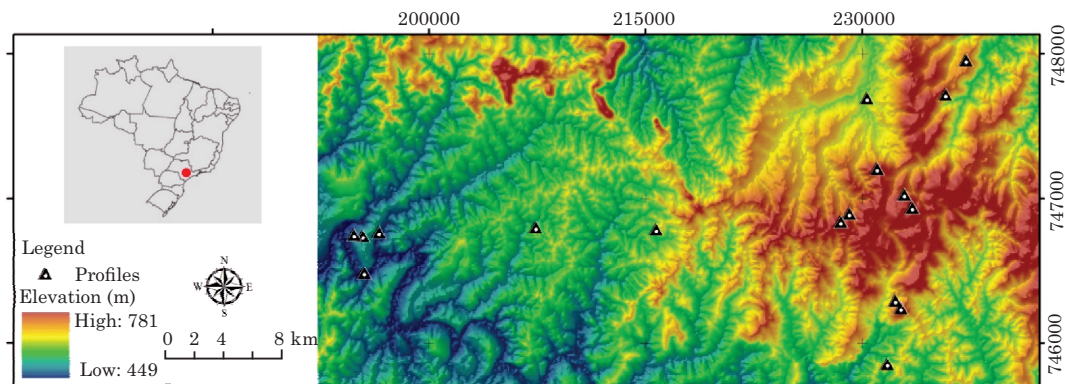


Figure 2. Profile location and landscape elevation in the survey area.

Reflectance variations in the visible region (400-700 nm) results from the presence of Fe oxides (Demattê & Garcia, 1999) which are directly related to variations in soil color (Fernandes et al., 2004; Chicatti, 2011). This is in agreement with the results observed in table 2, where the color represented by the values of R, G and B showed significant correlations with PC2 of 0.78 (G) and 0.71 (B). In addition to soil color, we also observed a significant correlation between exchangeable Al and PC2 (r value of 0.73). According to Stenberg et al. (2010), cations in the soil do not have a direct relationship to the spectrum in the visible and near infrared range. According to them, there is no specific spectral feature in the Vis-NIR for elements retained in CEC. Moreover, the authors attribute the good results in some studies (Chang et al., 2001; Groenigen et al., 2003; Pereira et al., 2004; Nanni & Demattê, 2006) to the existence of local covariation between spectrally active soil properties and the cations evaluated. In our study, soil color is considered a spectrally active property, which showed significant correlation between the Al^{3+} and R, G and B values of 0.74, 0.82 and 0.80, respectively (Table 2). Therefore, there was an indirect relationship between oxides and Al contents.

The most important wavelengths related to the PC3 (Figure 3c) were around 1,400; 1,900 and 2,200 nm. The features in 1,400 and 1,900 nm are related to molecular vibration of the hydroxyl group (OH) in hygroscopic water and 2:1 clay mineral structures (Lindberg & Snyder, 1972). In contrast, variations in 2200 nm are attributed to kaolinite (Goetz et al., 2009). Other ranges that had less influence on PC3 scores were from 350 to 600 nm and 850 to 1,100 nm, both related to the presence of hematite and goethite in soil samples (Leone & Sommer, 2000). The significant correlation between PC3 and R and G values, 0.79 and 0.77, respectively (Table 2), supported the statement that variations in PC3 were related to Fe oxides and hydroxides and, consequently, soil color. According to Viscarra Rossel et al. (2006b), this spectral range is also affected by OM, although in our study there was no significant correlation between the scores and that property. The lack of

correlation between OM contents and reflectance data (Table 2) is in disagreement with several studies. Viscarra Rossel et al. (2010) evaluated the correlation between spectral PCs and OM in the municipality of Rafard, São Paulo, Brazil, and found significant values for PC3. Karmanov (1968) suggested that the interactions between OM and soil spectral properties are due to the accumulation of humic substances, which produced reduction of soil spectral reflectance.

The PC4 was affected mainly by the 550-700 nm and 850-1,400 nm bands (Figure 3d), which were related to the presence of Fe oxides and hydroxides in the soil. Given that PC4 showed significant correlation to sand ($r = 0.66$) and clay ($r = -0.59$) contents (Table 2), the iron oxides in the clay fraction were probably responsible for the correlation between this property and PC4.

Spectral analysis of soil classes based on PCA

The determination of dissimilarity of soil classes, based on spectral reflectance, should be performed considering the qualitative analysis of both spectra from the surface horizon (HA) and the diagnostic subsurface horizons (HB). Despite good results (Demattê et al., 2004; Ben-Dor et al., 2008), this method requires knowledge of soil spectra and their characteristic properties for each soil class, which has not been established yet. Therefore, we evaluated the differences in spectral behavior of soil classes based on the scatter plots of the PCA scores (Figure 4), which was more easily interpreted. In this case, we considered the surface and subsurface diagnostic horizons.

Knowing the relationship between soil properties and the reflectance spectrum, it is possible to draw inferences about clay content, color and mineralogy from visual interpretation analysis. In addition, assessment of soil profile spectra and surface and subsurface horizons, allows identification of the degree of weathering and environmental conditions to which soils have been subjected (Demattê et al., 2004).

In this context, the scatter plots of the scores from Argissolo profiles (Figure 4 a,b,c) showed distinct

Table 1. Chemical and particle size analysis of soil profile samples

Id.	SIBCS. ⁽¹⁾	Hor.	Sand	Silt	Clay	OM	P	K	Ca	Mg	Al	H	SB	CEC	Bsat	Alsat	pH(H ₂ O)	pH(KCl)	Grad.	Fe ₂ O ₃	Ki	Silt/clay	Hue	v ⁽²⁾	c ⁽³⁾
P26A	PA1-3	Ap	600	180	220	11	4	4	25	11	1	11	40	51	78	2	5.1	3.6	2.3	6.10	2.86	0.40	9.9YR	4.1	1.8
P26C		Bt	300	200	500	7	1	1.3	8	9	35	51	18	69	26	66	4.3	3.6	2.3	6.10	2.86	0.40	8.7YR	4.2	2.5
P32A	CX4-3	Ap	840	60	100	10	3	0.7	14	6	5	16	21	37	56	19	5						8.9YR	3.8	1.4
P32C		Bt	780	60	160	7	1	0.4	8	3	35	43	11	54	21	75	4.6	3.9	1.6	1.15	1.83	0.38	7.6YR	4.6	3.3
P37A	PVA3-3	A	680	100	220	17	6	1.6	28	13	0	17	43	60	71	0	6.1						8.3YR	3.7	1.7
P37C		Bt	530	100	370	13	1	0.4	15	6	12	34	21	55	39	36	4.8	3.9	1.7	3.36	1.84	0.27	6.5YR	3.8	2.6
P40A	LVA4-3	Ap	680	100	220	10	2	0.6	9	4	9	34	14	48	29	40	4.9						5.9YR	3.6	2
P40C		Bw ₁	620	120	260	9	1	0.2	6	3	6	20	9.2	29	32	39	4.8	4	1.2	2.20	1.48	0.46	4.8YR	4.1	3
P49A	LVA4-1	Ap	720	120	160	15	342	1.8	42	6	2	15	50	65	77	4	6.2						7.5YR	3.6	1.7
P49C		Bt ₁	700	100	200	7	1	1.1	19	4	0	6	24	30	80	0	6.3	5.5	1.2	1.81	1.24	0.5	4.5YR	3.8	2.7
P52A	LVf1-1	Ap	280	80	640	22	6	2.4	14	4	5	47	20	67	30	20	4.7						2.0YR	3.4	2.3
P52C		Bw2	200	90	710	14	2	0.8	27	9	1	22	37	59	63	3	5.8	4.9	1.1	23.04	1.99	0.12	1.1YR	3.3	2.2
P54A	PVA2-1	Ap	580	140	280	15	3	0.4	27	5	5	90	32	122	26	13	5.4						5.9YR	3.8	2.1
P54C		Bt ₁	400	120	480	7	2	0.4	44	8	0	9	52	61	85	0	5.5	5	1.7	7.44	1.69	0.25	5.6YR	3.8	3.5

(1) Nomenclature according to Embrapa (2013). The first number following the soil class represents the textural class: (1) heavy clay: > 35 % and ≤ 60 %; (3) clay loam: > 25 % and ≤ 35 %; (4) sandy loam: > 15 % and ≤ 25 %; and (5) sand: ≤ 15 %. The second number represents the diagnostic chemical properties: (1) B_{sat} ≥ 50 %; (2) B_{sat} < 50 %; (3) Al_{sat} ≥ 50 %, (2) Value, (3) Chroma.

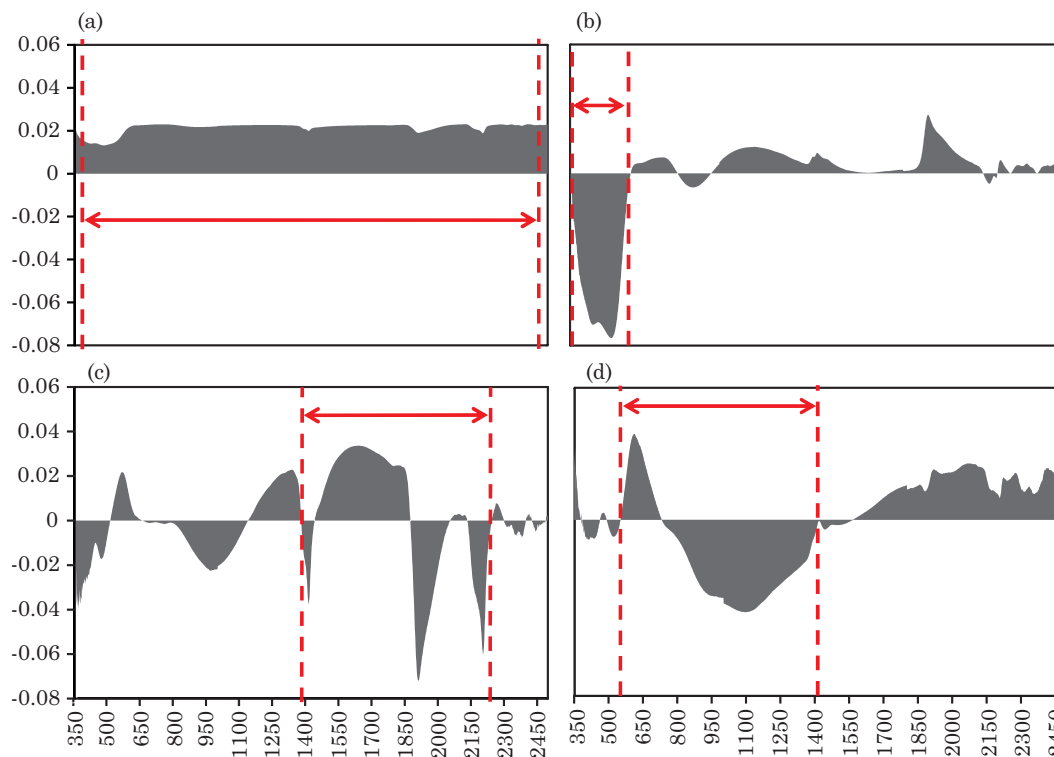


Figure 3. PCA loadings of the (a) first principal component - PC1, (b) second principal component - PC2, (c) third principal component - PC3, and (d) fourth principal component - PC4.

values for HA in relation to HB. By analyzing these graphs, we observed that the values are greater in samples from horizon A for PC1 and PC2. This difference was more pronounced in Figure 4c, where the plot represented the values of PC2 and PC3. In this case, samples from HA of Argissolos were at the top left of the figure, while the samples from HB were found at the bottom right (Figure 4c). The PCA loadings (Figure 3) showed that the second component was related to the spectral range from 350 to 600 nm and, therefore, Fe oxides (Demattê & Garcia, 1999). Thus, the sample distribution on that scatter plot indicated variation in Fe oxide contents between the subsurface horizons and their respective surface horizons. This is in agreement with Lepsch et al. (1977), who evaluated soils in a toposequence in the western plateau of São Paulo, Brazil, and found that Argissolos had a significant variation of Fe_2O_3 at depth. According to the authors, this characteristic is related to iron accumulation in the subsurface horizons caused by clay degradation in the HA of the profiles. The PC3 scores (Figure 4c) were higher in samples from HB than HA, which were influenced by the features in 1,400, 1,900 and 2200 nm, as evidenced in the loadings, indicating the presence of 1:1 and 2:1 clay minerals (Lindberg & Snyder, 1972), and therefore variation in the clay content at increasing depth in the profile.

The three profiles of Cambissolos showed spectra with distinct patterns with high values for PC1, PC2 and PC3 (Figure 4d,e,f). For that reason, these profiles

had a distinct position in the PC graph when compared to the soils of other classes (Figure 4). These soils presented a lower degree of weathering and tended to have higher contents of 1:1 and 2:1 clay minerals, which were identified in the spectra and produced higher PC3 values. The loam texture soils with predominantly yellowish color (Table 1) generated greater values of PC1 and PC2. This was confirmed by the significant correlation between clay content and PC1, as well as between soil color and PC2 (Table 2). Moreover, by Figure 4e and 4f, we observed that surface and subsurface horizons may be distinguished from the more intense degree of weathering in the soil surface layer, which led to variations in soil properties, such as clay type and content, as well as the presence of Fe oxides (Lepsch et al., 1977). In spite of a visible spectral distinction between HA and HB, it was not as pronounced as in Argissolos, thus excluding the possibility of a mistake between these soil classes when assessed only for their spectral reflectance.

In regard to the Latossolos and Nitossolos, the PC scores of both soil classes showed similar spectral patterns, in which both PC1 and PC3 showed medium values, while the PC2 values were generally lower (Figure 4g,h,i,j,k,l). In addition, the samples from HA and HB did not show significant differences in regard to the scores. That similarity indicated homogeneity in the characteristics of samples at different depths, which is characteristic for these two soil classes. Through the scatter plots, we observed close

Table 2. Pearson correlation coefficients indicating the relationship between spectra PCA scores and physico-chemical attributes of soil

	Sand	Silt	Clay	OM	P	K	Ca	Mg	Al	H	SB	CEC	Bsat%	Alsats%	pH(H ₂ O)	pH(KCl)	R	G	B	PC1	PC2	PC3	PC4
Sand	1	-0.34	-0.98*	-0.47	-0.44	0.18	-0.42	-0.55	0.39	0.21	-0.47	-0.22	-0.29	0.4	-0.24	-0.3	0.68	0.59	0.31	0.78*	0.2	0.59	0.66*
Silt		1	0.16	-0.27	-0.08	-0.09	-0.05	0.13	0.24	0.22	0	0.21	-0.13	0.26	-0.26	-0.25	0.18	0.3	0.3	-0.36	0.48*	0.29	-0.51
Clay			1	0.55	0.48	-0.17	0.45	0.55	-0.46	-0.27	0.5	0.19	0.33	-0.47	0.3	0.37	-0.75	-0.69*	-0.38	-0.74*	-0.31*	-0.68	-0.59*
OM				1	0.28	-0.24	0.29*	0.17*	-0.41	-0.18	0.25*	0.05*	0.2	-0.4	0.4	0.33	-0.68*	-0.65*	-0.44	-0.15	-0.48	-0.45	-0.01
P					1	-0.14	0.52	0.18	-0.3	-0.26	0.43	0.14	0.47	-0.46	0.37	0.4	-0.36	-0.43	-0.32	-0.38	-0.39	-0.24	-0.18
K						1	0.05	0.12	-0.01	-0.09	0.12*	0.02	0.19	-0.08	0	0	0.03	0.04	-0.08	-0.02	0.11	0.01	-0.12
Ca							1	0.76*	-0.51*	-0.52	0.97*	0.38*	0.91*	-0.77*	0.79*	0.82*	-0.43*	-0.45	-0.52	-0.3	-0.11	-0.46	-0.26
Mg								1	-0.32*	-0.29*	0.88*	0.53*	0.64*	-0.55*	0.52*	0.55	-0.43*	-0.33	-0.28	-0.6	0.18	-0.63	-0.47
Al									1	0.93*	-0.47	0.5	-0.67*	0.85*	-0.64	-0.59	0.74*	0.82*	0.8*	0.18	0.73*	0.6	0.16
H										1	-0.48	0.56*	-0.72*	0.79*	-0.64*	-0.6	0.52	0.62*	0.67*	0.09	0.6	0.44	0.14
SB											1	0.45*	0.88*	-0.75*	0.75*	0.78*	-0.45*	-0.43	-0.48	-0.4	-0.02	-0.54	-0.34
CEC												1	0.09	0.1	0.05	0.11	0.1	0.22	0.23	-0.29	0.59	-0.06	-0.18
Bsat%													1	-0.91*	0.89*	0.88*	-0.5*	-0.53	-0.64	-0.23	-0.28	-0.47	-0.21
Alsats%														1	-0.86*	-0.82*	0.74*	0.79*	0.81*	0.26	0.56	0.64	0.21
pH(H ₂ O)															1	0.96*	-0.56	-0.6	-0.59	-0.08	-0.39	-0.53	-0.03
pH(KCl)																1	-0.52	-0.57	-0.5	-0.13	-0.33	-0.57	-0.07
R																	1	0.95*	0.72	0.5	0.63	0.79*	0.33
G																		1	0.79*	0.35	0.78*	0.77*	0.23
B																			1	0.17	0.71*	0.47	0.19
PC1																				1	-0.01	0.52	0.8*
PC2																					1	0.44	-0.12
PC3																						1	0.26
PC4																							1

* significant at p<0.0001.

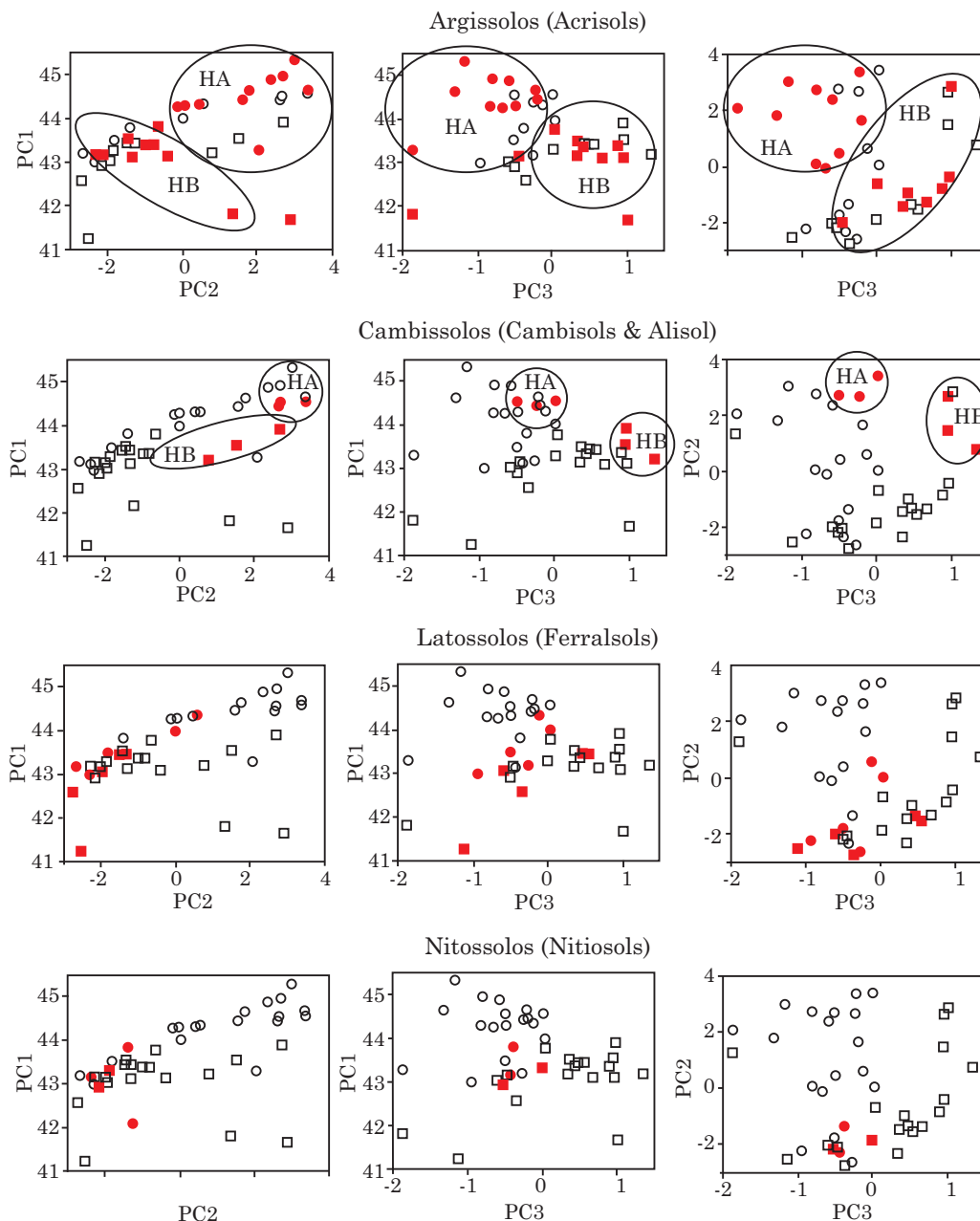


Figure 4. Graphs representing the principal components 1, 2 and 3 of the spectra from the profiles of Argissolos, Cambissolos, Latossolos and Nitossolos in the surface (●) and subsurface (■) layers.

positioning of these soil samples at the left (Figure 4 g, j) or closer to the origin of the graph (Figure 4h,i,k,l). Thus, these characteristics did not allow these soil classes to be distinguished from one another. However, that aspect could be differentiated from other soils evaluated in this study. Bellinaso et al. (2010) evaluated the spectra of 233 soil profiles from southeastern and Midwestern Brazil and found that Latossolos and Nitossolos had a similar spectral behavior. The author also indicated that the differentiation between these soils should be carried out through the presence of gibbsite in Latossolos, which provides a typical and more pronounced spectral

feature in 2,265 nm. The evidence of this mineral is due to a more advanced weathering process in LVs, providing a higher content of Fe and Al oxides and a greater loss of SiO₂ (Boul et al., 1997). Thus, we were able to confirm that the incapability of distinguishing LVs and NVs was, in part, related to the fact that the PCA was not able to represent spectral variations at 2265 nm as suggested by the loadings (Figure 3). In this case, it is necessary to observe the spectral behavior of these soil classes to distinguish them. However, even in studies with more detailed information, there is still great confusion between LVs and NVs (Cezar et al., 2013).

Numerical classification of soil profiles based on soil spectra

Based on numerical classification, seven groups were identified within the dataset evaluated (Table 3). The method is an unsupervised classification, i.e., no previous rules or conditions were established for the OSACA algorithm. The clustering of soil profiles was conducted based only on spectral behavior, taking into account their horizons. This method was used to investigate if the spectra are able to distinguish different soil clusters by themselves. In general, the comparison between the soil groups obtained from the numerical method and the soil classes from SiBCS presented similarities (Table 4). It proved that spectra were relevant to soil analyses and classification.

The SiBCS classification at the first categorical level generated four soil classes. Considering the second categorical level, this value was increased to seven classes. Adding information regarding texture to the suborder level, 13 different soil classes were obtained (Table 3). By evaluating the NMI index, we observed that the relationship between the cluster generated by OSACA and the classes of the Brazilian System of Soil Classification corresponded to 0.42 at the 1st level (Table 3). When these clusters were compared to the 2nd level + texture, this value increased to 0.74. The NMI index increased to 0.78 when the comparison was performed with the classification at the 2nd level plus texture and chemical properties (Table 3). The U index presented a greater value when considering the 2nd level + texture, with a value of 0.64 (Table 3). By soil classification at the 2nd level + texture + chemistry, the U index showed a slight reduction, with a value of 0.62 (Table 3). If we consider both indices in our analyses (NMI and U), we found better correlations between conventional classification and numerical clustering at the 2nd level + texture (Table 3).

By evaluating the differentiation carried out by numerical classification, we observed a trend toward grouping soil profiles with similar clay content (Figure 5a) and color (Figure 5b). In general, soils with higher clay content and red hue were allocated to group 1, while sandy soils with yellow hue were classified in group 7 (Figure 5). Groups 2, 3, 4 and 6, however, ranged from loam textured to clay soils, with hues from 5YR to 7.5YR (Figure 5b). In contrast, group 5

consisted of one profile with 10YR, which was distinct from the others (Figure 5b). Considering that the parameters used by OSACA were the PC scores, and PC1 is correlated with clay and the second and third components with soil color, we observed that this technique was able to cluster soil profiles with similarities in these properties.

The contingency table (Table 4) indicates the comparison between the classes obtained in the SiBCS and the OSACA groups. Overall, we observed that group 1 corresponded to Latossolos Vermelhos and Nitossolos Vermelhos. Therefore, the numerical technique was not able to distinguish soils belonging to these classes. In part, this was due to the fact that distinction between these two soil classes by spectral reflectance is carried out mainly by the absorption feature in 2265 nm. This feature is related to gibbsite, which is more pronounced in LVs than NVs (Bellinaso et al., 2010). In addition, given that these soils have high clay and iron oxide contents (Table 1) and, in some cases, the presence of magnetite, the spectral aspects able to distinguish these two classes might have been masked by these properties (Formaggio et al., 1996). Groups 2, 3 and 5 consisted of Argissolo profiles (Table 4), in which these soil profiles were divided into several groups, probably related to differences in clay contents and color, as well as the distinct variation patterns of these properties in depth. This may be exemplified by the comparison between groups 3 and 5 (Table 4). In this case, group 3 had one soil profile with loam texture, Yellow-Red Munsell color, classified as Argissolo; and soil profiles with clay texture, Yellow-Red color, classified as Argissolo. In contrast, group 5 consisted of one profile with clay texture, Yellow color, also classified as Argissolo (Table 4).

Group 6 consisted of profiles classified as Latossolo Vermelho Amarelo and Argissolo Vermelho Amarelo, both loam-textured (Table 4). These profiles were clustered in the same group because both were highly weathered, with a predominance of 1:1 clay minerals and Fe oxides. In this case, the main factor for distinguishing both classes is the variation in intensity of reflectance between surface and subsurface layers due to the texture gradient in Argissolos, as well as the feature in 2265 nm, due to the higher concentrations of gibbsite in Latossolos

Table 3. Normalized Mutual Information and Uncertainty coefficient for the Brazilian Soil Classification System (SiBCS)

Categorical level of classification SiBCS ⁽¹⁾	No. of classes SiBCS	No. of groups OSACA	NMI ⁽²⁾	U ⁽³⁾
Order	4	7	0.42	0.33
Suborder	7	7	0.61	0.57
Suborder + texture	13	7	0.74	0.64
Suborder + texture + chemical	18	7	0.78	0.62

⁽¹⁾ According to Embrapa (2013); ⁽²⁾ Normalized Mutual Information; ⁽³⁾ Uncertainty coefficients.

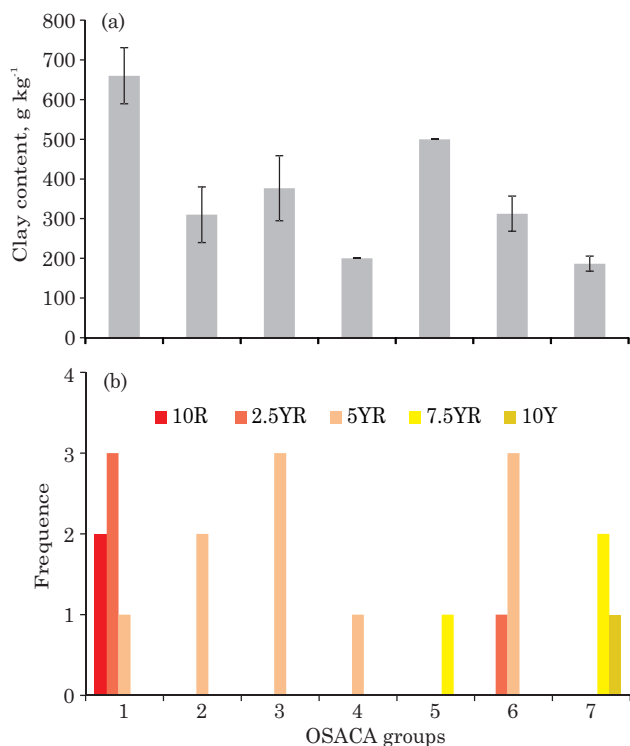


Figure 5. Average clay content and standard deviation for each OSACA group (a); and frequency of soil profile hues in the different OSACA groups (b).

(Bellinaso et al., 2010). However, the texture gradient was probably not pronounced enough to be detected by the spectra. Finally, group 7 was associated with soil profiles identified as Cambissolo Háplico with loam texture (Table 4). When compared to the highly weathered soils, such as Argissolos, Latossolos and Nitossolos, great differences between them were observed (Bellinaso et al., 2010). That was reinforced in this study, where group 7 was exclusively associated with Cambissolo profiles (Table 4).

CONCLUSIONS

1. The soil spectra do not have features directly related to soil chemical properties, consequently no significant correlations are expected between those two sets of information.

2. Descriptive analysis of soil spectral data, based on the diagnostic surface and subsurface horizons, allows different soil classes to be distinguished. It is observed that variations in PC scores between the surface and subsurface horizon of Argissolos enable a satisfactory level of distinction of this soil as compared to other classes. In this case, graphic representation of PC2 vs PC3 allows better discrimination of Argissolos.

Table 4. Contingency table of SiBCS classes by OSACA groups

SiBCS classif. ⁽¹⁾	WRB classif.	OSACA Classif.							Total	
		1	2	3	4	5	6	7		
CX4-5	Haplic Cambisol (Alumic, Dystric)	0	0	0	0	0	0	0	2	2
CX4-3	Haplic Alisol (Alumic, Profondic, Arenic)	0	0	0	0	0	0	0	1	1
LV2-1	Haplic Ferralsol (Eutric, Clayic, Rhodic)	1	0	0	0	0	0	0	0	1
LV2-2	Haplic Ferralsol (Dystric, Rhodic)	1	0	0	0	0	0	0	0	1
LVA4-3	Haplic Ferralsol (Dystric)	0	0	0	0	0	1	0	0	1
LVA3-2	Haplic Ferralsol (Dystric)	0	0	0	0	0	1	0	0	1
LVf1-1	Haplic Ferralsol (Eutric, Clayic, Rhodic)	1	0	0	0	0	0	0	0	1
NVf1-1	Haplic Nitisol (Dystric, Rhodic)	1	0	0	0	0	0	0	0	1
NV1-3	Haplic Nitisol (Alumic, Dystric, Rhodic)	1	0	0	0	0	0	0	0	1
PA1-3	Haplic Acrisol (Alumic, Epieutric, Clayic)	0	0	0	0	1	0	0	0	1
PVA3-2	Haplic Acrisol (Epieutric)	0	0	1	0	0	0	0	0	1
PV3-3	Haplic Acrisol (Abruptic, Alumic, Profondic, Arenic)	0	0	0	0	0	1	0	0	1
PVA1-1	Haplic Acrisol (Epieutric, Clayic)	1	0	0	0	0	0	0	0	1
PVA2-1	Haplic Acrisol (Abruptic, Arenic)	0	0	2	0	0	0	0	0	2
PVA3-3	Haplic Acrisol (Alumic)	0	1	0	0	0	0	0	0	1
PVA4-2	Haplic Acrisol (Abruptic, Epieutric, Profondic, Arenic, Chromic)	0	1	0	0	0	0	0	0	1
PVA3-3	Haplic Acrisol (Alumic)	0	0	0	0	0	1	0	0	1
LVA4-1	Haplic Ferralsol (Eutric, Arenic)	0	0	0	1	0	0	0	0	1
Total		6	2	3	1	1	4	3	20	

⁽¹⁾ Nomenclature according to Embrapa (2013). The first number following the soil class represents the textural class: (1) heavy clay: > 60 %; (2) clay: > 35 % and ≤ 60 %; (3) clay loam: > 25 % and ≤ 35 %; (4) sandy loam: >15% and ≤ 25 %; and (5) sand: ≤ 15 %. The second number represents the diagnostic chemical properties: (1) Bs_{at} ≥ 50 %; (2) Bs_{at} < 50 %; (3) Als_{at} ≥ 50 %.

3. Latossolos and Nitossolos have similar spectra and had PCs with great similarity in the surface and subsurface horizons. Consequently, distinction between those classes requires further information, such as morphological features.

4. Spectral reflectance data associated with the OSACA algorithm is an efficient methodology for soil classification. The results show a consistent clustering of soil profiles, as well as a good relationship to the SiBCS classes.

5. The similarity between numerical classification and the Brazilian system were higher when considering the 2nd categorical level + texture. When considering the 2nd categorical level + texture + chemical properties, the similarity decreases. This is indicative of spectral limitations in the classification process. Therefore, in situations where some chemical properties are required to distinguish soil classes correctly, soil spectra might not be the best option.

ACKNOWLEDGEMENTS

We thank the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for scholarships to the first and third authors, and to the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for a scholarship to the second author.

LITERATURE CITED

- BELLINASO, H.; DEMATTÊ, J.A.M. & ARAÚJO, S.R.A. Soil spectral library and its use in soil classification. *R. Bras. Ci. Solo*, 34:861-870, 2010.
- BELLINASO, H. Biblioteca espectral de solos e sua aplicação na quantificação de atributos e classificação. Piracicaba, Escola Superior de Agricultura "Luiz de Queiroz", 2009. 264p. (Dissertação de Mestrado)
- BEN-DOR, E.; HELLER, D. & CHUDNOVSKY, A. A novel method of classifying soil profiles in the field using optical means. *Soil Sci. Soc. Am. J.*, 72:1113-1123, 2008.
- BEN-DOR, E.; IRONS, J.R. & EPEMA, G.F. Soil reflectance. In: RENCEZ, A.N., ed. *Remote sensing for the earth science: Manual of remote sensing*. New York, John Wiley, 1999. p.111-188.
- BHATTI, A.U.; MULLA, D.J. & FRAZIER, B.E. Estimation of soil properties and wheat yields on complex eroded hills using geostatistics and thematic mapper images. *Rem. Sens. Environ.*, 37:181-191, 1991.
- BOWERS, S.A. & HANKS, R.J. Reflection of radiant energy from soils. *Soil Sci.*, 100:130-138, 1965.
- BROWN, D.J.; SHEPHERD, K.D.; WALSH, M.G.; DEWANE MAYS, M. & REINSCH, T.G. Global soil characterization with Vis-NIR diffuse reflectance spectroscopy. *Geoderma*, 132:273-290, 2006.
- BOUL, S.W.; HOLE, F.D.; McCRACKEN, R.J. & SOUTHARD, R.J. *Soil genesis and classification*. 4.ed. Ames, Iowa State University Press, 1997. 527p.
- 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 de Campinas, 1986. 94p. (Boletim Técnico, 106)
- CAMPBELL, N.A.; MULCAHY, M.J. & McARTHUR, W.M. Numerical classification of soil profiles on the basis of field morphological properties. *Aust. J. Soil Res.*, 8:43-58, 1970.
- CAMPOS, R.C.; DEMATTÊ, J.A.M. & QUARTAROLI, C.F. Determinação indireta do teor de hematita no solo a partir de dados de colorimetria e radiometria. *Pesq. Agropec. Bras.*, 38:521-528, 2003.
- CARRÉ F. & JACOBSON, V. Numerical classification of soil profile data using distance metrics. *Geoderma*, 148:336-345, 2009.
- CARRÉ, F. & McBRATNEY, A.B. Digital terrain mapping. *Geoderma*, 128:340-353, 2005.
- CEZAR, E.; NANNI, M.R.; CHICATI, M.L.; OLIVEIRA, R.B. & DEMATTÊ, J.A.M. Discriminação entre solos formados em região transicional por meio de resposta espectral. *Biosci. J.*, 29:644-654, 2013.
- CHANG, C.W.; LAIRD, D.A.; MAUSBACH, M.J. & HURBURGH, C.R. Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties. *Soil Sci. Soc. Am. J.*, 65:480-490, 2001.
- CHICATTI, M.L. Avaliação espectral de solos em áreas alagáveis do noroeste do Paraná. Maringá, Universidade Estadual de Maringá, 2011. 258p. (Tese de Doutorado)
- DEMATTÊ, J.A.M.; CAMPOS, R.C.; ALVES, M.C.; FIORIO, P.R. & NANNI, M.R. Visible-NIR reflectance: A new approach on soil evaluation. *Geoderma*, 121:95-112, 2004.
- DEMATTÊ, J.A.M. & GARCIA, G.J. Alteration of soil properties through a weathering sequence as evaluated by spectral reflectance. *Soil Sci. Soc. Am. J.*, 63:327-342, 1999.
- DIDAY, E. Une nouvelle méthode en classification automatique et reconnaissance des formes: la méthode des nuées dynamiques. *R. Stat. Appliq.*, 19:283-300, 1971.
- EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - EMBRAPA. Centro Nacional de Pesquisa de Solos. Sistema brasileiro de classificação de solos. 3.ed. Rio de Janeiro, 2013. 353p.
- FERNANDES, R.B.A.; BARRÓN, V.; TORRENT, J. & FONTES, M.P.F. Quantificação de óxidos de ferro de Latossolos brasileiros por espectroscopia de reflectância difusa. *R. Bras. Ci. Solo*, 28:245-257, 2004.
- FORMAGGIO, A.R.; EPIPHANIO, J.C.N. & VALERIANO, M.M. Comportamento espectral (450-2.450 nm) de solos tropicais de São Paulo. *R. Bras. Ci. Solo*, 20:467-474, 1996.
- GALVÃO, L.S.; PIZARRO, M.A. & EPIPHANIO, J.C.N. Variations in reflectance of tropical soils: Spectral-chemical compositions relationships from AVIRIS data. *Remote Sens. Environ.*, 75:245-255, 2001.

- GAUCH, H.G. Multivariate analysis in community ecology. Cambridge, Cambridge University Press, 1982. 298p.
- GIASSON, E.; INDA JUNIOR, A.V. & NASCIMENTO, P.C. Estimativa do benefício econômico potencial de dois levantamentos de solos no Estado do Rio Grande do Sul. *Ci. Rural*, 36:478-486, 2006.
- GOETZ, A.F.H.; CURTISS, B. & SHILEY, D.A. Rapid gangue mineral concentration measurement over conveyors by NIR reflectance spectroscopy. *Min. Eng.*, 22:490-499, 2009.
- GROENIGEN, J.W.; MUTTERS, C.S.; HORWATH, W.R. & KESSEL, C. NIR and DRIFT-MIR spectrometry of soils for predicting soil and crop parameters in a flooded field. *Plant Soil*, 250:155-165, 2003.
- IUSS Working Group WRB. 2007. World Reference Base for Soil Resources 2006, first update 2007. World Soil Resources Reports No. 103. FAO, Rome.
- KARMANOV, I.I. Genesis and geography of soils: Application of spectrophotometric coefficients to the study of soil-forming processes. *Sov. Soil Sci.*, 2:152-165, 1968.
- KING, D. & GIRARD, M.C. Reflexions sur la classification des profils de la couverture pedologique. Proposition dun algorithme: VLADIMIR. *Sci. Sol.*, 26:239-254, 1988.
- LEONE, A.P. & SOMMER, S. Multivariate analysis of laboratory spectra for the assessment of soil development and soil degradation in the southern Apennines. *Rem. Sens. Environ.*, 72:346-359, 2000.
- LEPSCH, I.F.; BUOL, S.W. & DANIELS, R.B. Soil-landscape relationships in the Occidental Plateau of São Paulo State, Brazil: II. Soil Genesis, Morphology and Classification. *Soil Sci. Soc. Am. J.*, 41:109-115, 1977.
- LINDBERG, J.D. & SNYDER, D.G. Diffuse reflectance spectra of several clay minerals. *Am. Mineral.*, 57:485-493, 1972.
- LIU, Y. & NAVATHE, S.B. Text analysis of MEDLINE for discovering functional relations among genes; evaluation of keyword extraction weighting schemes. *Intern. J. Data Min. Bioin.*, 1:88-110, 2006.
- MENDONÇA-SANTOS, M.L. & SANTOS, H.G. The state of the art of Brazilian soil mapping and prospects for digital soil mapping. In: LAGACHIERE, P.; McBRATNEY, A.B. & VOLTZ, M., eds. Developments in soil science. Amsterdam, Elsevier, 2006. p.39-54.
- MEZZALIRA, S. Folha Geologia de Piracicaba: SF 23-M 300. São Paulo, Instituto Geográfico e Geológico do Estado de São Paulo, 1966. Escala 1:100.000.
- MORAN, M.S.; IONOUE, Y. & BARNES, E.M. Opportunities and limitation for image-based sensing in the precision crop management. *Rem. Sens. Environ.*, 61:319-346, 1997.
- NANNI, M.R. & DEMATTÊ, J.A.M. Spectral reflectance methodology in comparison to traditional soil analysis. *Soil Sci. Soc. Am. J.*, 70:393-407, 2006.
- PEREIRA, A.G.; GÓMEZ, A.H. & HE, Y. Near-infrared spectroscopy potential to predict N, P, K and OM content in a loamy mixed soil and its combination with precision farming. In: CONFERÊNCIA INTERNACIONAL CIGR, Beijing, 2004. Proceedings... Beijing, 2004. p.11-14.
- RAIJ, B.van; ANDRADE, J.C.; CANTARELLA, H. & QUAGGIO, J.A. Análise química para avaliação da fertilidade de solos tropicais. Campinas, Instituto Agrônômico de Campinas, 2001. 285p.
- RAYNER, J.H. Classification of soils by numerical methods. *J. Soil Sci.*, 17:79-92, 1966.
- RIVERO, R.G.; GRUNWALD, S. & BRULAND, G.L. Incorporation of spectral data into multivariate geostatistical models to map soil phosphorus variability in a Florida wetland. *Geoderma*, 40:428-443, 2007.
- SANTOS, C.K. Análise de agrupamentos em coleção de documentos: uma abordagem baseada na detecção espectral de comunidades em redes complexas. Rio de Janeiro, Universidade Federal do Rio de Janeiro, 2009. 201p. (Dissertação de Mestrado)
- SANTOS, R.D.; LEMOS, R.C.; SANTOS, H.G.; KER, J.C. & ANJOS, L.H. Manual de descrição e coleta de solos no campo. 5.ed. Viçosa, MG, Sociedade Brasileira de Ciência do Solo, 2005. 100p.
- SHANNON, C.E. & WEAVER, W. The mathematical theory of communication. Champaign, University of Illinois, 1963. 117p.
- SHI, Z. & GHOSH, J. Model-based clustering with soft balancing. In: IEEE INTERNATIONAL CONFERENCE, 3., Melbourne, 2003. Anais... Melbourne, IEEE Computer Society, 2003. p.459-466.
- 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. Burlington, Academic Press, 2010. p.163-215.
- STONER, E.R. Physicochemical, site and bi-directional reflectance factor characteristics of uniformly moist soils. West Lafayette, Purdue University, 1979. 132p. (Tese de Doutorado)
- VISCARRA ROSSEL, R.A.; RIZZO, R.; DEMATTÊ, J.A.M. & BEHRENS, T. Spatial modelling of a soil fertility index using vis-NIR spectra and terrain attributes. *Soil Sci. Soc. Am. J.*, 74:1293-1300, 2010.
- VISCARRA ROSSEL, R.A.; CATTLE, S.R.; ORTEGA, A. & FOUAD, Y. In situ measurements of soil colour, mineral composition and clay content by vis-NIR spectroscopy. *Geoderma*, 150:253-266, 2009.
- VISCARRA ROSSEL, R.A. ParLeS: Software for chemometric analysis of spectroscopic data. *Chemom. Intelligent Lab. Syst.*, 90:72-83, 2008.
- VISCARRA ROSSEL, R.A.; MINASNY, B.; ROUDIER, P. & MCBRATNEY, A.B. Colour space models for soil science. *Geoderma*, 133:320-337, 2006a.
- VISCARRA ROSSEL, R.A.; WALWORT, D.J.J.; MCBRATNEY, A.B.; JANIK, L.J. & SKJESMSTAD, J.O. Visible near infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma*, 131:59-75, 2006b.
- WOLD, H. Systems under indirect observation. Amsterdam, Elsevier, 1982. 668p.