



Soil surface spectral data from Landsat imagery for soil class discrimination

Marcos Rafael Nanni^{1*}, José Alexandre Melo Demattê², Marcelo Luiz Chicati¹, Peterson Ricardo Fiorio³, Everson César¹ and Roney Berti de Oliveira¹

¹Departamento de Agronomia, Universidade Estadual de Maringá, Av. Colombo, 5790, 87020-900, Maringá, Paraná, Brazil. ²Departamento de Solos e Nutrição de Plantas, Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba, São Paulo, Brazil. ³Departamento de Engenharia Rural, Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba, São Paulo, Brazil. *Author for correspondence. E-mail: mnranni@uem.br

ABSTRACT. The aim of this study was to develop and test a method to determine and discriminate soil classes in the state of São Paulo, Brazil, based on spectral data obtained via Landsat satellite imagery. Satellite reflectance images were extracted from 185 spectral reading points, and discriminant equations were obtained to establish each soil class within the studied area. Sixteen soil classes were analyzed, and discriminant equations that comprised TM5/Landsat sensor bands 1, 2, 3, 4, 5, and 7 were established. The results showed that this methodology could effectively identify individual soil classes using discriminant analyses of the spectral data obtained from the surface. Success rates of > 40% were achieved for 14 of the 16 evaluated soil classes when applying the satellite image data. When the 10 soil classes containing the largest number of minimum cartographic areas were used, the hit rate increased to > 50%, for seven soil classes with a global hit rate of 52%. When the soil classes were grouped based on their parent materials, the hit rate increased to 70%. Thus, we concluded that the spectral method for soil classification was efficient.

Keywords: discriminant analysis, TM-Landsat, Brazilian soil classes, spectral response.

Informações espectrais de imagens Landsat da superfície do solo como indicativo na discriminação de classes de solos

RESUMO. O objetivo deste trabalho foi desenvolver e testar um método para a determinação da classe de solo e sua separabilidade na paisagem dos solos presentes em uma área de estudo localizada no estado de São Paulo. Um conjunto de equações discriminantes foi obtido utilizando-se o sistema SAS que permitiu estabelecer a classe de solo na área de estudo. Foram analisadas 16 classes de solos as quais foram estabelecidas equações discriminantes compostas pelas bandas 1, 2, 3, 4, 5, e 7 do sensor TM5/ Landsat. As leituras espectrais foram realizadas em 185 pontos da área de estudo, donde se extraiu a reflectância da imagem. Classes de solos podem ser individualizadas por meio de análise discriminante utilizando-se informações sobre seu comportamento espectral obtida pela metodologia apresentada. A análise discriminante apresentou índices de acerto acima de 40% dentro da classe de solo avaliada, para 14 das 16 classes de solos. Utilizando-se as dez classes com maior número de áreas mínimas cartografadas o acerto, dentro a classe, foi maior que 50% para sete classes de solos, com acerto global estabelecido em 52%. Quando se agrupou as classes de solos em função do seu material de origem, o acerto passou para 70%.

Palavras-chave: análise discriminante, TM-Landsat, solos do Brasil, resposta espectral.

Introduction

In recent decades, Brazil has become a prominent worldwide leader in agricultural exports due to the expansion into new agricultural frontiers, implementation of technological improvements, and exceptional efforts to address adversities among growers. To continue this growth, Brazil must adopt suitable practices for using and managing previously uncultivated land areas.

Soil surveys that include appropriate and detailed cartographic representations allow us to obtain a wealth of information that, when properly managed, can ena-

ble technicians and growers to employ methodologies and establish strategies that extend and even increase the productive capacity of their lands. However, there are an insufficient number of professionals available to map such a large country, and there are still millions of hectares of land left for the rational and sustainable production of food (DALMOLIN, 1999).

Demattê (2001) stated that the study and implementation all available technologies regarding soil spectral analyses is of critical significance. Spectral behavior analyses of soils using remote sensors might provide answers to the problem of

class discrimination. Such studies are based on the fact that each soil presents a spectral signature according to the absorption of specific wavelengths across the entire electromagnetic spectrum (BENDOR et al., 1999).

The fundamental theory of soil spectral analyses indicates that soils can be characterized according to their B horizon. This horizon is not detected using satellite data unless erosion has occurred. Therefore, how could spectral responses help in discriminating soil units? To answer this question, we begin with the assumption that many soil classes exhibit surface characteristics that differ from other classes and therefore can be used as a taxonomic identifier. This assumption has been combined with the use of aerial photographs for soil discrimination. This method does not detect B horizon values; however, it does infer information regarding the 'probable' soil class. A small number of published studies of soils and tropical soils have suggested that the use of satellite images can assist in the discrimination of surface information.

Because a spectral response is also an individual trait, the use of this technique allows for the separation of soil classes and may therefore aid in pedological surveys. Because soil properties are related to soil classes, and these soil properties are related to spectra, we can correlate spectra with soils to facilitate the detection of soil types. Therefore, the aim of this study was to develop and test a

method for determining and distinguishing different soil classes in the southwestern region of São Paulo State, Brazil, based on spectral data obtained from satellite images.

Material and methods

The studied area is located in the southwestern region of the state of São Paulo and is delimited by the geographical coordinates 23°0'31.37" - 22°58'53.97" south latitude and 53°39'47.81" - 53°37'25.65" west longitude in a region known as the Paleozoic depression (IPT, 1981). As part of the hydrographic basin of the Tietê River (OLIVEIRA et al., 1992), the area is bordered by the Capivari river and spans an area of approximately 198 ha with a perimeter of 11,045.80 m (Figure 1).

Geologically, the studied area is situated in the Itararé Formation and is part of the Tubarão Group. These units were formed between the Upper Carboniferous Period and Middle Permian Epoch and occur in the state a complex association of several lithofacies, almost all of which are detritic that formed in rapid vertical and horizontal successions (IPT, 1981). The predominant lithologies consist of mineralogically immature sandstones with heterogeneous granulation ranging from feldspathic sandstones to arkoses; these units range from thin layers to banks that are tens of meters thick (VIDAL-TORRADO; LEPSCH, 1999).

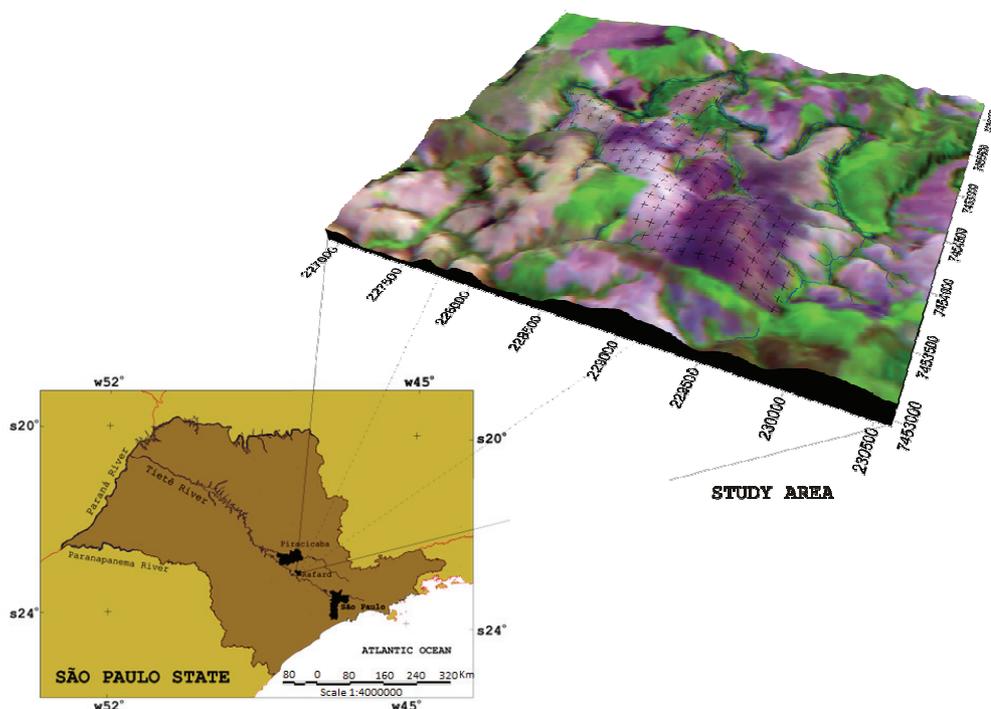


Figure 1. Study location and soil sampling grid representation (1 sample ha⁻¹).

The geographic information system “Sistema de Processamento de Informações Georreferenciadas (SPRING)” (INPE, 1999) was used to implement the cartographic study. This system has multiple functions and algorithms for processing georeferenced databases. A database was established using this system to incorporate the geospatialized information that was obtained from planimetric and altimetric charts and data that was collected in the field using GPS receivers and orbital images.

In the entire area, 185 augured points were selected (Figure 1). These points were marked in a grid fashion with one point per hectare (WOLKOWSKI; WOLLENHAUPT, 1994) following procedures used in soil surveys. Specifically, a synthetic approach was used (EMBRAPA, 1996) and the database was georeferenced with a differential global positioning system (DGPS) technique, which is a method used with GPS receivers to achieve higher precision with these data receptors in absolute positioning modes.

The collected samples were first dried in a forced ventilation oven at a constant temperature of 50°C for 48h and then sifted through a 2-mm sieve (air-dried soil). The textural soil groupings were conducted according to Embrapa (1999). The densitometric method was used to determine the total sand, silt and clay contents (CAMARGO et al., 1986). Organic matter (OM), active and residual acidity, pH and cation exchange capacity (CEC) determinations were performed according to Embrapa (1997). The sum of the cations (i.e., calcium, magnesium and potassium) (SC), base saturation ($V\% = [SC/CEC] \cdot 100$), and aluminum saturation ($m\% = [Al^{3+}/SC + Al^{3+}] \cdot 100$) were determined according to Van Raij and Quaggio (1989). Total iron (Fe), silicon (Si), and titanium (Ti) contents were determined by sulfuric acid digestion according to a methodology recommended by Embrapa (1997). The soil color was identified using the Munsell color chart (GRETAGMACBETH, 2000).

The physiographic unit boundaries were established after generating orbital image visual interpretations according to the recommendations of Donzeli et al. (1983) and Nanni and Rocha (1997). These boundaries were then used in the SPRING system after generating successive combinations of algorithms and filters to improve the visual quality as reported by Nanni and Rocha (1997). The soil classes were defined after analyzing the laboratory samples and open trenches at representative physiographic unit locations. The fieldwork procedures and the description and collection of the materials were performed according to the criteria established by

Lemos and Santos (1996). The soil classes were defined according to the Brazilian classification system (EMBRAPA, 1999) and correlated to the soil taxonomy (COSTA; NANNI, 2004).

Spectral data acquisition and processing

The information extracted from the pixels was used to extract the gray value levels for each TM-Landsat-5 image point that corresponded to a sampling location in the field. The gray value levels for each band were then converted into reflectance values according to the procedures established by Markham and Barker (1986) and Thome et al. (1997). For Rayleigh scattering and ozone absorption corrections, the 5S irradiative transfer code simulation (TANRÉ et al., 1992; VERMOTE et al., 1997) was used and properly corrected for atmospheric effects (KAHLE et al., 1980). After the conversion and correction procedures, the zero gray level images corresponded to 0% reflectance, whereas a gray level of 255 corresponded to 100% soil reflectance (DEMATTE; NANNI, 2003).

To verify whether the points in the image that corresponded to the field-collected points characterized the exposed soils (i.e., without plant cover), the normalized difference vegetation index (NDVI) methodology was applied (DEMATTE et al., 2000; JACKSON, 1983; KAUTH; THOMAS, 1976).

Statistical analyses

The Statistical Analysis System (SAS, 1992) software package version 6 was used to manage the obtained data and perform the statistical analyses. To organize the database for statistical analyses and the evaluation of hypotheses, the reflectance values for each soil were fitted to a data matrix used for this analysis, which consisted of the six TM-Landsat-5 bands for each sampled terrain.

Completely randomized variance analyses were performed on this matrix, and the reflectance means of each soil per selected wavelength range were analyzed using the Tukey test at the $p \leq 0.01$ and $p \leq 0.05$ levels. The general linear models (GLM) method of the SAS program was used to test each selected wavelength range for the hypothesis where H_0 : soil1=soil2=soil3=soil4=soil5. The rejection of this null hypothesis implies the acceptance of an alternative hypothesis where H_1 : at least two soils are statistically different. The purpose of this study was to determine which variables would have a higher or lower potential to yield models that best explain the problem. Therefore, the STEPDISC method (SAS, 1992) was initially employed to select

suitable bands. Discriminant analyses were then conducted to differentiate and characterize the soil with the goal of developing and testing the method for soil class determination based on either its spectral data or analytical characteristics. Thus, the DISCRIM method (SAS, 1992) was used. The parametric method was used to develop the discriminant linear functions.

Results and discussion

The area studied included a variety of parent materials consisting of reworked sandstone-diabase, sandstone-shale, diabase-shale materials, and a combination of all three. As a result of these geological conditions, 18 soil classes were established (Table 1).

The following soil subgroups and land areas dominated the study area: Paleudults and Paleudalfs, 58.12 ha (31.12%); Oxisols, 39.60 ha (21.21%); Dystrochrepts and Eutrochrepts, 36.92 ha (19.76%); Rhodic Paleudalfs, 19.08 ha (10.22%); Arguidolls, 12.76 ha (6.84%); Udorthent, 9.76 ha (10.22%); and Udifluvents, 10.52 ha (5.63%). The declivity conditions pre-established the possible presence of thin soils with depths ranging between 1.50 and 2.00 meters. Deeper soils were found in plains or gently rolling areas and reached up to tens of meters in depth. Because the cartographic land areas included in the strongly rolling and mountainous classes (with soils less than 0.80 cm in depth) were small, lower proportions of shallow soil classes were

observed. Rocky outcrops were also included in the shallow soil class.

As expected, the soil classes that contained a small number of samples showed smaller amplitudes or variability of the attributes allowed by the class. The B1, B2, B3, B4, B5, and B7 values corresponded to the bands from the TM sensor (Table 1). This data bank was structured according to the following guidelines: (1) the soil spectral patterns from a given area are used to develop statistics and equations; (2) when a soil sample cannot be classified, its spectral information must be acquired; (3) the spectral information must be applied to all previous equations (Table 1); and (4) the highest obtained data on equations results will be the desired soil. However, such reasoning is valid only for the soils and conditions that were observed in this study.

Furthermore, this reasoning created the potential to characterize different regions in the country, thus allowing for faster preliminary identifications, as stated by Coleman and Montgomery (1990), Demattê and Garcia (1999), and Nanni et al. (2004).

According Gerbermann and Neher (1979), if a dataset is obtained via an automated technique, soil maps will be produced faster than with conventional methods. Table 2 summarizes the hit percentages for the classification of the soils that were examined using the discriminant equations compared to the conventional classification methodology.

Table 1. Discriminant equations obtained using the TM-Landsat evaluation bands for all soil classes.

Class	n ¹	Discriminant equations ²
RUd ³	7	-34.9927 - 3.4605*B ₁ - 0.14081*B ₂ + 0.89447*B ₃ + 2.48853*B ₄ + 0.73691*B ₅ - 0.43438*B ₇
MTfr	14	-33.0937 - 6.4961*B ₁ + 0.76258*B ₂ + 0.67527*B ₃ + 2.74335*B ₄ + 1.43487*B ₅ - 1.8539*B ₇
CE1	4	-39.2013 - 6.70611*B ₁ + 1.12848*B ₂ + 0.9865*B ₃ + 2.52815*B ₄ + 1.3809*B ₅ - 1.60193*B ₇
CD1	6	-44.642 - 4.82595*B ₁ + 0.63345*B ₂ - 0.15483*B ₃ + 3.4661*B ₄ + 0.90695*B ₅ - 0.53662*B ₇
CE3	3	-31.4928 - 5.79768*B ₁ + 0.2626*B ₂ + 1.16703*B ₃ + 2.52334*B ₄ + 1.11786*B ₅ - 1.36596*B ₇
CE2	6	-31.2862 - 5.74454*B ₁ + 1.0568*B ₂ + 0.2449*B ₃ + 2.95226*B ₄ + 0.94872*B ₅ - 1.15765*B ₇
CD2	12	-47.5261 - 4.8144*B ₁ + 0.61794*B ₂ + 0.63277*B ₃ + 2.83464*B ₄ + 1.26189*B ₅ - 1.06166*B ₇
LVe	20	-30.4057 - 7.34009*B ₁ + 0.58157*B ₂ + 1.89452*B ₃ + 2.13164*B ₄ + 0.6073*B ₅ - 0.69839*B ₇
LVAe	9	-30.6845 - 6.71217*B ₁ + 0.19287*B ₂ + 1.98053*B ₃ + 1.7553*B ₄ + 0.73748*B ₅ - 0.45603*B ₇
LVAAd	5	-38.5065 - 4.51743*B ₁ + 0.48783*B ₂ - 0.05451*B ₃ + 3.19036*B ₄ + 0.93884*B ₅ - 0.6548*B ₇
PVAe1	27	-50.0303 - 4.77088*B ₁ + 0.45385*B ₂ + 0.90233*B ₃ + 2.78573*B ₄ + 0.72144*B ₅ - 0.08911*B ₇
PVAAd	2	-35.4309 - 6.44089*B ₁ + 1.06144*B ₂ + 0.76785*B ₃ + 2.79228*B ₄ + 0.83768*B ₅ - 0.87981*B ₇
PVAe2	15	-38.6123 - 4.66779*B ₁ + 0.84416*B ₂ + 0.49152*B ₃ + 2.62668*B ₄ + 0.79755*B ₅ - 0.43845*B ₇
PVAe3	4	-45.236 - 7.15908*B ₁ + 0.79169*B ₂ + 1.59665*B ₃ + 2.48002*B ₄ + 1.06265*B ₅ - 0.92726*B ₇
Rle2	2	-55.1678 - 8.65796*B ₁ + 4.80298*B ₂ - 1.77679*B ₃ + 3.79686*B ₄ + 1.42308*B ₅ - 2.04336*B ₇
Rle1	9	-30.0657 - 6.06055*B ₁ + 0.15969*B ₂ + 1.25555*B ₃ + 2.17189*B ₄ + 1.1738*B ₅ - 1.15031*B ₇
NVef1	14	-28.3508 - 7.2187*B ₁ + 0.27142*B ₂ + 1.96497*B ₃ + 1.93981*B ₄ + 1.06931*B ₅ - 1.41957*B ₇
NVef2	4	-22.2789 - 6.45176*B ₁ + 0.35227*B ₂ + 2.14137*B ₃ + 1.49931*B ₄ + 0.35021*B ₅ - 0.36348*B ₇

¹Number of individuals; ²Discriminant equations obtained using the TM-Landsat analyses with the following bands: B₁ (450-520 nm); B₂ (520-600 nm); B₃ (630-690 nm); B₄ (760-900 nm); B₅ (1550-1750 nm); and B₇ (2080-2350 nm); Soil classes of studied area: Argissolo Vermelho-Amarelo distrófico, PVAAd (Typic Paleudult); Argissolo Vermelho-Amarelo eutrófico textura arenosa/média, PVAe1 (Typic Paleudult); Argissolo Vermelho-Amarelo eutrófico abrupto, PVAe3 (Arenic Abruptic Paleudalf); Argissolo Vermelho-Amarelo eutrófico textura média/argilosa, PVAe2 (Typic Paleudult); Cambissolos Háplicos Ta eutróficos lépticos substrato folhelhos da Formação Itararé, CE2 (Typic Eutrochrept); Cambissolos Háplicos Tb distrófico, CD2 (Typic Dystrochrept); Cambissolos Háplicos Tb distrófico e aluminosos substrato retrabalhamento de arenito e saprolito de folhelho da Formação Itararé, CD1 (Typic Dystrochrept); Cambissolos Háplicos Tb eutrófico típicos substrato diabásio, CE3 (Typic Eutrochrept); Cambissolos Háplicos Tb eutrófico típicos substrato regolito do retrabalhamento de arenito e diabásio, CE1 (Typic Eutrochrept); Chernossolo Argilúvico férrico saprolítico, MTfr (Typic Arguidoll); Latossolo Vermelho eutrófico típico - LVe (Typic Haplorthox); Latossolo Vermelho-Amarelo epieutrófico típico, LVAAd (Typic Haplorthox); Latossolo Vermelho-Amarelo eutrófico típico, LVAe (Typic Haplorthox); Neossolo Litólicos eutrófico chernossólicos e típicos, Rle2 (Typic Udorthent); Neossolos Flúvicos Tb distrófico típicos, RUd (Typic Udifluent); Nitossolo Vermelho eutrófico, NVef1 (Rhodic Paleudalf); and Nitossolo Vermelho eutrófico latossólico, NVef2 (Rhodic Paleudalf).

Table 2. Discriminant analyses, the number of samples, and the percent of soil classifications for all soil classes in the studied area.

Class	Total samples	Correct		Incorrect ¹	
		Samples	%	Samples	%
RUd	7	4	57.14	3	42.86
MTfr	14	9	64.29	5	35.71
CE1	4	3	75.00	1	25.00
CE2	6	3	50.00	3	50.00
CE3	3	1	33.33	2	66.66
CD1	6	2	33.33	4	66.66
CD2	12	0	0.00	12	100.00
LVe	20	9	45.00	11	55.00
LVAe	9	4	44.44	5	55.56
LVAAd	5	2	40.00	3	60.00
PVAe1	27	16	59.26	11	40.74
PVAe2	15	0	0.00	15	100.00
PVAe3	4	1	25.00	3	75.00
PVAd	2	0	0.00	2	100.00
Rle1	9	2	22.22	7	77.78
Rle2	2	1	50.00	1	50.00
NVef1	14	10	71.43	4	28.57
NVef2	4	3	75.00	1	25.00

¹Global error median=61.36%; ²Number of incorrect classifications.

Of the 18 analyzed classes (Table 3), only five showed results hit values of > 60% within the class. Among these classes, the best results were achieved for CE1 and NVef2 with a 75% hit rate within each class. However, most of the classes showed inadequate results. Total failure was observed (i.e., no observations were auto classified using the established discriminant equations) for the soil classes CD2, PVAd and PVAe2 with six, two, and five individual sites, respectively.

In contrast to the total failure observed, the following aspects should be considered: (1) a misidentification occurred for those classes with similar physical surface characteristics, such as organic matter content, iron, clay, sand, and color, which would be difficult to distinguish even in the field; (2) classes with higher iron contents due to their parent material (diabase) showed infrequent or no misidentifications for confusion with the classes derived from sandstone; (3) the Dystrochrepts soil, which is a transitional class, showed the highest number of classification errors; and (4) in some classes with low numbers of sites, an error in a single point could result in a high rate of error for the class, which would contribute to an increase in the global error rate. For example, the Rle2 class (i.e., litholic soil with a shale substrate) with only two sites was problematic. Because one of the sites was classified as a Eutrochrepts soil with a shale substrate (CE2), the error rate was 50%. Considering that these soils were similar at the surface and in satellite images, the reported error (50%) was high and affected the global error for the entire analysis. This class showed the highest global

median error value among all analyses performed for this study (61.36%).

In an attempt to eliminate this problem, only the 10 soil classes with the highest number of sites were evaluated. The classes that showed more than seven individual sites were selected following the procedure reported by Oliveira et al. (1982). To ensure that the statistical analyses could not exhibit a marked influence on the standard error of the average, a minimum of 7 samples per soil class were used.

Of the 18 identified classes in the study area, only 10 contained more than 7 observations, which were classified as the following: RUd, MTfr, CD1, CE2, CD2, LVe, LVAe, PVAe1, PVAe2, and NVef1. The obtained equations for the most abundant classes are shown in Table 3.

Table 4 summarizes the percentages of hits for the classification of the soils that were examined using the discriminant equations.

By limiting the number of soil classes, the global error was reduced to 48.25%. This reduction occurred for two primary reasons. The first reason was the elimination of classes with a small number of sites, which contributed to the increased classification error and the low hit percentages within each class. The second reason for the error reduction was the inclusion of those classes that were previously classified into one of the excluded classes. For example, the CD2 class had a hit rate within its own class of 0% based on an analysis that involved all classes. The hit rate increased to 33.33% when the analysis involved only the 10 classes with the highest number of sites.

Furthermore, the classes with similar surface characteristics caused misidentifications during the classification. Therefore, because the orbital data only referred to the surface soil layers, further discriminant analyses were conducted to group the soils according to their parent materials. The classes that were related to the parent material were designated as sand, shale, shale/diabase, and diabase.

The following soil classes were combined to compose the sand group: RUd, CE1, CD1, CD2, LVAd, PVAe1, PVAd, PVAe2, and PVAe3. The CE2 and Rle2 classes formed the shale group. The

shale/diabase group contained the MTfr class, and the diabase group consisted of the CE3, LVe, LVAe, Rle1, NVef1, and NVef2 soil classes. The discriminant equations for these groupings are shown in Table 5.

Among the six bands that were evaluated using STEPDISC from the SAS system (Table 5), only the TM_B2 band was not used to assemble the discriminant equations. All of the other bands showed statistical significance at $p \leq 0.01$. After the discriminant equations were established, the classification data were generated as shown in Table 6.

Table 3. Discriminant equations obtained using TM-Landsat-evaluated bands for the 10 soil classes with the highest number of observations in the studied area.

Class	n ¹	Discriminant equations ²
RUd	7	-34.2896 - 3.55919*B ₁ + 0.13371*B ₂ + 0.63909*B ₃ + 2.42146*B ₄ + 1.18612*B ₅ - 1.12938*B ₇
MTfr	14	-34.6111- 6.78654*B ₁ + 1.50043*B ₂ + 0.12797*B ₃ + 2.84146*B ₄ + 1.9366*B ₅ - 2.66193*B ₇
CD1	6	-44.6683 - 5.21448*B ₁ + 1.28492*B ₂ - 0.58129*B ₃ + 3.47132*B ₄ + 1.44051*B ₅ - 1.42153*B ₇
CE2	6	-32.2268 - 6.15409*B ₁ + 1.8236*B ₂ - 0.28028*B ₃ + 3.01722*B ₄ + 1.42041*B ₅ - 1.92273*B ₇
CD2	12	-47.9958 - 5.18662*B ₁ + 1.20571*B ₂ + 0.18731*B ₃ + 2.88435*B ₄ + 1.82645*B ₅ - 1.98184*B ₇
LVe	20	-29.8542 - 7.38688*B ₁ + 1.14133*B ₂ + 1.43477*B ₃ + 2.09416*B ₄ + 0.97403*B ₅ - 1.24361*B ₇
LVAe	9	-30.0024 - 6.66141*B ₁ + 0.59336*B ₂ + 1.62149*B ₃ + 1.70646*B ₄ + 1.09234*B ₅ - 1.00113*B ₇
PVAe1	27	-48.885 - 5.04352*B ₁ + 0.92914*B ₂ + 0.54099*B ₃ + 2.74134*B ₄ + 1.22831*B ₅ - 0.90683*B ₇
PVAe2	15	-38.464 - 5.05779*B ₁ + 1.43548*B ₂ + 0.07636*B ₃ + 2.6489*B ₄ + 1.27334*B ₅ - 1.22066*B ₇
NVef1	14	-28.3791 - 7.18731*B ₁ + 0.77905*B ₂ + 1.51134*B ₃ + 1.94458*B ₄ + 1.45759*B ₅ - 1.99804*B ₇

¹Number of individuals; ²Discriminant equations obtained using TM-Landsat analyses with the following bands: B₁ (450-520 nm); B₂ (520-600 nm); B₃ (630-690 nm); B₄ (760-900 nm); B₅ (1550-1750 nm); and B₇ (2080-2350 nm).

Table 4. Discriminant analyses, the number of samples, and soil classification percentages for the 10 soil classes with the highest number of observations.

Class	Total of samples	Correct		Error ¹	
		Samples	%	Samples	%
RUd	7	6	85.71	1	14.29
MTfr	14	10	71.43	4	28.57
CE2	6?	3	50.00	3	50.00
CD1	6?	3	50.00	3	50.00
CD2	12	4	33.33	8	66.67
LVe	20	9	45.00	11	55.00
LVAe	9	5	55.56	4	44.44
PVAe1	27	15	55.56	12	44.44
PVAe2	15	1	6.67	14	93.33
NVef1	14	9	64.29	5	35.71

¹Global error median = 48.25%; ²Number of error classifications.

Table 5. Discriminant equations obtained using TM-Landsat-evaluated bands for the parental material groups in the studied area.

Groups	n ¹	Discriminant equations ²
Sandstones	78	-36.45681 - 4.61984*B ₁ + 1.00592*B ₂ + 2.55491*B ₃ + 0.65769*B ₄ - 0.50672*B ₇
Shales	8	-28.66557 - 5.15408*B ₁ + 0.64817*B ₂ + 2.91957*B ₃ + 0.86388*B ₄ - 1.26487*B ₇
Shales/Diabase	14	-28.74755 - 6.00838*B ₁ + 1.09110*B ₂ + 2.52247*B ₃ + 1.18156*B ₄ - 1.68701*B ₇
Diabase	63	- 26.23708 + 1.87910*B ₁ + 1.92418*B ₂ + 0.71534*B ₃ - 0.93404*B ₇

¹Number of individuals; ²Discriminant equations obtained using TM-Landsat analyses with the following bands: B₁ (450-520 nm); B₂ (520-600 nm); B₃ (630-690 nm); B₄ (760-900 nm); B₅ (1550-1750 nm); and B₇ (2080-2350 nm).

Table 6. The number of observations and percent of soils classified using TM-Landsat bands for the parent material groups.

Groups ¹	Sandstone	Shales/Diabase	Shales	Diabase	Total
Sandstone	60.00 ²	4.00	10.00	4.00	78.00
Shales/Diabase	76.92%	5.13	12.82	5.13	100.00
	0.00	10.00	2.00	2.00	14.00
Shales	0.00	71.43%	14.29	14.29	100.00
	0.00	1.00	6.00	1.00	8.00
Diabase	0.00	12.50	75.00%	12.50	100.00
	1.00	10.00	5.00	47.00	63.00
Total	1.59	15.87	7.94	74.60%	100.00
	61.00	25.00	23.00	54.00	163.00
Percent	37.42	15.34	14.11	33.13	100.00

¹Identification of the soil class groupings with equal parental material; ²Total frequencies for all classes.

A significant improvement was observed when the soil classes were grouped according to their parental material (Table 7). The global error for the Chi-square test was 25.51%, which was well below the 48.25% error for the classification using the 10 most populous classes and far better than the 61% error when all classes were used separately. Differences between the parent materials were found using the spectral data that were obtained by the TM Landsat sensor in the soil line analyses (HUETE, 1989) as described by Nanni and Demattê (2006). Nanni and Demattê (2006) found higher variations in soil lines for the soils classes that were developed in areas with sandstone and low iron levels relative to those that contained high amounts of clay and iron.

This fact observed by Nanni and Demattê (2006) opens new perspectives and possibilities for geological surveys because this method can set realistic boundaries between the different parent materials. This method confirms that the different parent materials can be discriminated according to their reflectance due to physical interactions with electromagnetic energy. To reinforce the discriminant analyses, a simulation was performed in which 80% of the sampled sites were used to generate a discriminant model that could be tested using the remaining 20% of the data. The sampled sites were randomly selected; i.e., the SAS system randomly selected the components that would take part in the discriminant analyses (80%) and those that would be used to test the obtained models (20%). The procedure was tested using

50 simulations with the system set to randomly choose 80% of the sampled sites, which generated the discrimination model that was used to test the remaining 20% of the sites. Only the 10 soil classes with the highest number of observations were used. Furthermore, classes with a reduced number of sites might not show the variability that exists within the class or might have characteristics that resemble another class. These issues could interfere with the safe and reliable analyses. In addition, as previously described, classes with a small number of sites greatly increased the global error.

After the simulation, the system yielded results that showed the data classification percentage within the model (Table 7).

The hit rate was low when 80% of the observations were applied to the model. Of the 5825 matches that were generated by the 50 simulations for the model-examined data, the model produced 3102 misclassifications with a global error of 53.3%, whereas the classifications matched in 2723 instances with an accuracy of 46.7%. Furthermore, 11% of the total data was lost (632 times). Hence, a difference between the numbers presented in Table 7 (5193) and the total number (5825) was observed. Despite this difference, the presented model was highly significant according to the Chi-square test ($p \leq 0.01$). The correlation between the classes that were observed and estimated by the model showed an r-square value of 83.8% as defined by the contingency coefficient.

Table 7. The number of observations and soil percentages classified using the 10 soil classes with the highest number of observations in the studied area and 80 % of the observations to develop the model.

Class ¹	RUd	MTfr	CD1	CE2	CD2	LVc	LVAc	PVAc1	PVAc2	NVefl	Total ²
RUd	213.00 ³ 77.45 ⁴	0.00	39.00	3.00	11.00	0.00	0.00	2.00	7.00	0.00	275
MTfr	0.00	376	14.18	1.09	4.00	0.00	0.00	0.73	2.55	0.00	554
	0.00	67.87	0.00	15.88	0.00	13.72	0.18	0.00	0.18	2.17	
CD1	2.00	0.00	110.00	2.00	47.00	0.00	0.00	41.00	36.00	0.00	238
	0.84	0.00	46.22	0.84	19.75	0.00	0.00	17.23	15.13	0.00	
CE2	0.00	49.00	26.00	123.00	0.00	25.00	7.00	0.00	0.00	1.00	231
	0.00	21.21	11.26	53.25	0.00	10.82	3.03	0.00	0.00	0.43	
CD2	42.00	18.00	73.00	16.00	192.00	0.00	0.00	63.00	74.00	0.00	478
	8.79	3.77	15.27	3.35	40.17	0.00	0.00	13.18	15.48	0.00	
LVc	0.00	57.00	0.00	4.00	0.00	396.00	124.00	0.00	0.00	229.00	810
	0.00	7.04	0.00	0.49	0.00	48.89	15.31	0.00	0.00	28.27	
LVAc	1.00	0.00	6.00	0.00	0.00	32.00	193.00	32.00	3.00	82.00	349
	0.29	0.00	1.72	0.00	0.00	9.17	55.3	9.17	0.86	23.5	
PVAc1	91.00	0.00	125.00	6.00	95.00	1.00	71.00	612.00	89.00	0.00	1090
	8.35	0.00	11.47	0.55	8.72	0.09	6.51	56.15	8.17	0.00	
PVAc2	99.00	48.00	44.00	52.00	104.00	0.00	25.00	115.00	114.00	0.00	601
	16.47	7.99	7.32	8.65	17.3	0.00	4.16	19.13	18.97	0.00	
NVefl	42.00	17.00	0.00	8.00	0.00	53.00	53.00	0.00	0.00	394.00	567
	7.41	3.00	0.00	1.41	0.00	9.35	9.35	0.00	0.00	69.49	
										Total ⁵	5193
											100

¹Classes with the highest number of observations; ²Total frequencies in the model; ³Frequency in the class; ⁴% correct within the class; ⁵Total frequencies for all classes.

Most of the errors occurred between classes that had very similar parent materials or surface layer properties. The only non-justifiable error was observed with the PVAe2 class, which was corrected classified only once in 554 trials (0.02%) by the model. Because the characteristics of these soils are quite distinct, this result indicates that the model requires adjustment.

Among all classes, RUd was the best classified with the smallest error and a 77.45% correct classifications. The worst performance was observed for the PVAe2 class, which produced a 18.97% frequency of correct identification. As previously discussed, this class originated from a sandstone deposit overlying shale or diabase. Therefore, the characteristics of PVAe2 soils result in an intermediate classification between several soil classes within the area. Even so, the highest rate of misidentification occurred among the classes with high sand fractions at the surface, such as RUd, CD1, CD2, PVAe1 and LVAe classes. Nevertheless, classification errors occurred with the MTfr (7.99%), CE2 (8.65%) and LVAe (4.16%) classes, indicating that the model was not reliable for these classes.

Table 8 summarizes the match percentages for the classification of soils that were examined using the discriminant equations when only 20% of the observations were tested in the model generated using the remaining 80% of the data.

When using 20% of the classes for the application of the model established by the remaining 80% of the data, the hit rate was low, producing a large number of misidentifications among the classes and a low frequency of observations within the classes. Based on 50 simulations, the model produced 962 misclassifications of 1475 points with a global error of 65.2%, whereas the model correctly classified 34.8% of

the data (513 points). Furthermore, 11% of the total data was lost (168 points). Hence, a difference between the frequencies shown in Table 8 (1307 test results) and the total frequency (1475 test results) was observed. In spite of this difference, the presented model was highly significant, as determined by the Chi-square test ($p \leq 0.01$). The correlation among the true and model-estimated classes showed an r-square value of 78.7%, as defined by the contingency coefficient.

However, even when testing 20% of the data, the classes that showed the highest rates of misidentification originated from similar parent materials. Thus, we were able to establish that the use of soil spectral curves for the discrimination of parent material classes showed satisfactory to good results when applying orbital-level data. For the discrimination of soil classes, the models that were generated with the obtained reflectance data at the orbital level showed satisfactory results for some classes and were inconsistent for others, indicating the need for further studies to improve the quality of the results.

Misclassifications occurred between similar soil data. CE and CD are moderate shallow soils that occur in a rolling relief. These soils were misclassified with other soil types under the same conditions (Table 2). Other misclassifications can also be explored, such as NVe1 is being confounded with NVe2. The only difference between these soil classes is that NVe1 is a highly clayey soil and NVe2 is a clayey soil. PVA soils also occur in rolling relief and can be misclassified as CE or CD. Thus, there are many cases where misclassifications may occur due to overlapping soil characteristics.

Table 8. Number of observations and soil percentages classified using the 10 soil classes with the highest number of observations in the studied area and 20% of the observations as developed by the model.

Class ¹	RUd	MTfr	CD1	CE2	CD2	LVe	LVAe	PVAe1	PVAe2	NVef1	Total ²
RUd	26.00 ³	0.00	22.00	1.00	4.00	0.00	5.00	4.00	13.00	0.00	75
	34.67 ⁴	0.00	29.33	1.33	5.33	0.00	6.67	5.33	17.33	0.00	
MTfr	0.00	98.00	0.00	22.00	0.00	19.00	0.00	0.00	0.00	7.00	146
	0.00	67.12	0.00	15.07	0.00	13.01	0.00	0.00	0.00	4.79	
CD1	1.00	0.00	14.00	8.00	19.00	0.00	1.00	9.00	10.00	0.00	62
	1.61	0.00	22.58	12.9	30.65	0.00	1.61	14.52	16.13	0.00	
CE2	0.00	27.00	12.00	13.00	0.00	8.00	3.00	0.00	5.00	1.00	69
	0.00	39.13	17.39	18.84	0.00	11.59	4.35	0.00	7.25	1.45	
CD2	9.00	6.00	22.00	9.00	24.00	0.00	0.00	31.00	21.00	0.00	122
	7.38	4.92	18.03	7.38	19.67	0.00	0.00	25.41	17.21	0.00	
LVe	0.00	11.00	0.00	1.00	0.00	84.00	42.00	0.00	1.00	51.00	190
	0.00	5.79	0.00	0.53	0.00	44.21	22.11	0.00	0.53	26.84	
LVAe	1.00	0.00	1.00	0.00	0.00	17.00	39.00	5.00	1.00	37.00	101
	0.99	0.00	0.99	0.00	0.00	16.83	38.61	4.95	0.99	36.63	
PVAe1	23.00	0.00	37.00	3.00	31.00	0.00	18.00	135.00	13.00	0.00	260
	8.85	0.00	14.23	1.15	11.92	0.00	6.92	51.92	5.00	0.00	
PVAe2	30.00	11.00	16.00	24.00	23.00	1.00	12.00	28.00	4.00	0.00	149
	20.13	7.38	10.74	16.11	15.44	0.67	8.05	18.79	2.68	0.00	
NVef1	7.00	19.00	0.00	0.00	0.00	10.00	21.00	0.00	0.00	76.00	133
	5.26	14.29	0.00	0.00	0.00	7.52	15.79	0.00	0.00	57.14	
										Total ⁵	1307
											100

¹Classes with the highest number of observations; ²Total frequency in the model; ³Frequency in the class; ⁴% correct within the class; ⁵Total frequency for all classes.

Most of the results indicated that as a first analysis, this classification system could be important for assisting in the next step of soil classification. Discrepancies between the classifications and field results were due to several parameters, such as the relief and soil surface property, which are not related to the undersurface. In fact, the subsurface is more significant when classifying soils. This horizon contains less organic matter and influences soil management. In addition, the mineralogy is generally more accurate in the subsurface due to less weathering. For many reasons, the surface information provides important indications to assist the convergence of evidence and ultimately aid in soil classification. In the field, the surface of each soil type usually exhibits different appearances due to the dynamics of water, organic matter, relief, and humidity. All of these factors interfere with the spectral information and are responsible for the observed differences.

Therefore, we believe that the emergence of new sensors installed on orbital or sub-orbital platforms with a greater number of bands or spectral ranges will be very important for discriminating soils with a noticeable reduction of classification errors.

Conclusion

The surface information collected assisted in the discrimination of soils in the studied area using statistical analyses and the data obtained at the orbital level. Based on the discriminant analyses, the hit rate increased as the sample number decreased. The simulated statistical test was efficient for establishing errors and matches during the discriminant analyses.

Satellite-derived surface data can provide valuable and relevant information for soil class discrimination and assist in soil surveys. In addition, soil classes that were grouped according to their parent material were identified via discriminant analyses.

Acknowledgements

The authors thank the São Paulo Research Support Foundation (Fapesp) for the acquisition of the IRIS spectroradiometer (Process no. 95/6259 6), the Coordination for the Improvement of Higher Level Personnel (Capes) for providing funding for the first author's research, and the Brazilian National Research Council (CNPq) for providing funding for the second author's research (Process no. 98/01059 7).

References

BEN-DOR, E.; IRONS, J. R.; EPEMA, G. F. Soil reflectance. In: RENCZ, N. (Ed.). **Remote sensing for the earth sciences: manual of remote sensing**. New York: Wiley, 1999. p. 111-118.

CAMARGO, O. R.; MONIZ, A. C.; JORGE, J. A.; VALADARES, J. M. **Methods of chemical, mineralogical, and physical analysis of soils**. Campinas: IAC, 1986.

COLEMAN, T. L.; MONTGOMERY, O. L. Assessment of spectral characteristics for differentiating among soil categories in the southeastern United States. **Photogrammetric Engineering and Remote Sensing**, v. 52, n. 1, p. 1659-1663, 1990.

COSTA, A. C. S.; NANNI, M. R. **Brazilian classification: encyclopedia of soil science**. New York: Taylor and Francis, 2004.

DALMOLIN, R. S. D. Missing soil scientists in Brazil. **Boletim da Sociedade Brasileira de Ciência do Solo**, v. 24, n. 4, p. 13-15, 1999.

DEMATTÊ, J. A. M. The pedologist and precision agriculture. **Boletim da Sociedade Brasileira de Ciência do Solo**, v. 26, n. 1, p. 17-18, 2001.

DEMATTÊ, J. A. M.; GARCIA, G. J. Alteration of soil properties through a weathering sequence as evaluated by spectral reflectance. **Soil Science Society of America Journal**, v. 63, n. 2, p. 327-342, 1999.

DEMATTÊ, J. A. M.; NANNI, M. R. Weathering sequence of soils developed from basalt as evaluated by laboratory (IRIS), airborne (AVIRIS) and orbital (TM) sensors. **International Journal of Remote Sensing**, v. 24, n. 3, p. 4715-4738, 2003.

DEMATTÊ, J. A. M.; HUETE, A.; FERREIRA JR., L.; ALVES, M. C.; NANNI, M. R.; CERRI, C. E. P. Evaluation of tropical soils through ground and orbital sensors. **International Conference of Geospatial Information in Agriculture and Forestry 2**, v. 2, n. 1, p. 34-41, 2000.

DONZELI, P. L.; VALÉRIO FILHO, M.; NOGUEIRA F. P.; PEREZ FILHO, A.; KOFFLER, N. F. Satellite images and radar on the physiographic setting of standards applied to soils. **Revista Brasileira de Ciência do Solo**, v. 7, n. 1, p. 89-94, 1983.

EMBRAPA-Empresa Brasileira de Pesquisa Agropecuária. **Criteria and standards for soil surveys**. Rio de Janeiro: Embrapa, 1996.

EMBRAPA-Empresa Brasileira de Pesquisa Agropecuária. **Manual of methods for soil analysis**. Rio de Janeiro: Embrapa, 1997.

EMBRAPA-Empresa Brasileira de Pesquisa Agropecuária. **Brazilian system of soil classification**. Brasília: Embrapa, 1999.

GERBERMANN, A. H.; NEHER, D. D. Reflectance of varying mixtures of a clay soil and sand. **Photogrammetric Engineering and Remote Sensing**, v. 45, n. 8, p. 1145-1151, 1979.

GRETAGMACBETH. **Munsell soil color charts – year 2000, revised**. New York: USA, 2000.

HUETE, A. R. Soil influences in remotely sensed vegetation-canopy spectra. In: ASRAR, G. (Ed.). **Theory and application of optical remote sensing**. New York: Wiley Interscience, 1989. p. 107-141.

INPE-Instituto Nacional de Pesquisas Espaciais. **Spring tutorial: basic Spring**. São José dos Campos: INPE, 1999.

- IPT-Instituto de Pesquisas Tecnológicas. **Mapa Geológico do Estado de São Paulo**. São Paulo: Brasil, 1981.
- JACKSON, R. D. Spectral indices in space. **Remote Sensing of Environmental**, v. 13, n. 1, p. 409-421, 1983.
- KAHLE, A. B.; MADURA, D. P.; SOHA, J. M. Middle infrared multispectral aircraft scanner data analysis for geological applications. **Applied Optics**, v. 19, n. 1, p. 2279-2290, 1980.
- KAUTH, R. J.; THOMAS, J. S. The tasseled cape graphic description of the spectral temporal development of agriculture crops as seen by Landsat. **Symposium on Machine Processing of Remotely Sensed Data**, v. 1, n. 1, p. 41-51, 1976.
- LEMONS, R. C.; SANTOS, R. D. **Manual of description and soil samples from the field**. Campinas: SBCS, 1996.
- MARKHAM, B. L.; BARKER, J. L. Landsat MSS and TM post-calibration dynamic ranges, exoatmospheric reflectances and at-satellite temperatures. **EOSAT Technical Notes**, n. 1, 1986.
- NANNI, M. R.; DEMATTÊ, J. A. M. Spectral reflectance methodology in comparison to traditional soil analysis. **Soil Science Society of America Journal**, v. 70, n. 2, p. 393-407, 2006.
- NANNI, M. R.; ROCHA, H. O. Integration of GIS technology, remote sensing and multivariate analysis in the delimitation of physiographic units for pedological mapping. **USP Scientific Series: IG report**, v. 28, n. 1, p. 129-143, 1997.
- NANNI, M. R.; DEMATTÊ, J. A. M.; FIORIO, P. R. Soil discrimination analysis by spectral response in the ground level. **Pesquisa Agropecuária Brasileira**, v. 39, n. 10, p. 995-1006, 2004.
- OLIVEIRA, J. B.; JACOMINE, P. K. T.; CAMARGO, M. N. **General soil classes from Brazil: a guide to assist recognition**. São Paulo: Funep/Unesp, 1992.
- OLIVEIRA, J. B.; MENCK, J. R. F.; BARBIERI, J. L. **Semidetailed pedological survey of São Paulo State: Araras quadrangle**. Campinas: IAC, 1982.
- SAS-Statistical Analysis System. **User's guide, version 6.0**. Cary: Statistical Analysis System Institute, 1992.
- TANRÉ, D.; HOLBEN, B. N.; KAUFMAN Y. J. Atmospheric correction algorithm for NOAA-AVHRR products: theory and application. **IEEE Transaction on Geoscience and Remote Sensing**, v. 30, n. 1, p. 231-248, 1992.
- THOME, K. B.; MARKHAM, B. L.; BARKER, J. L.; SLATER, P.; BIGGAR, S. Radiometric calibration of landsat. **Photogrammetric Engineering and Remote Sensing**, v. 63, n. 7, p. 853-858, 1997.
- VAN RAIJ, B.; QUAGGIO, J. A. **Methods of soil analysis for fertility**. Campinas: IAC, 1989.
- VERMOTE, E. F.; TANRÉ, D.; DEUZÉ, J. L.; HERMAN, L.; MORCRETTE, J. J. Second simulation of the satellite signal in the solar spectrum, 6S: an overview. **IEEE Transaction on Geoscience and Remote Sensing**, v. 35, n. 1, p. 675-686, 1997.
- VIDAL-TORRADO, P.; LEPSCH, I. F. Soil/parent material relations and pedogenesis on a slope dominated by clayey oxidic soils over sandstone at the S. Paulo State peripheral depression, southeastern Brazil. **Revista Brasileira de Ciência do Solo**, v. 23, n. 2, p. 357-369, 1999.
- WOLKOWSKI, R. P.; WOLLENHAUPT, N. C. Grid soil Sampling. **Better Crops**, v. 78, n. 4, p. 6-9, 1994.

Received on January 12, 2011.

Accepted on May 23, 2011.

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.