Spatial Modeling of a Soil Fertility Index using Visible–Near-Infrared Spectra and Terrain Attributes

R. A. Viscarra Rossel*
CSIRO Land and Water
Bruce E. Butler Lab.
GPO Box 1666
Canberra ACT 2601, Australia

R. Rizzo
CSIRO Land and Water
Bruce E. Butler Lab.
GPO Box 1666
Canberra ACT 2601, Australia

and
Soil Science Dep.
Univ. de São Paulo
Piracicaba, S.P., Brazil

J.A.M. Demattê
Soil Science Dep.
Univ. de São Paulo
Piracicaba, S.P., Brazil

T. Behrens
Institute of Geography
Physical Geography
Univ. of Tübingen
Rümelinstraße 19-23
D-72074, Tübingen, Germany

Our objective was to develop a methodology to predict soil fertility using visible–near-infrared (vis–NIR) diffuse reflectance spectra and terrain attributes derived from a digital elevation model (DEM). Specifically, our aims were to: (i) assemble a minimum data set to develop a soil fertility index for sugarcane (*Saccharum officinarum* L.) (SFI-SC) for biofuel production in tropical soils; (ii) construct a model to predict the SFI-SC using soil vis–NIR spectra and terrain attributes; and (iii) produce a soil fertility map for our study area and assess it by comparing it with a green vegetation index (GVI). The study area was 185 ha located in São Paulo State, Brazil. In total, 184 soil samples were collected and analyzed for a range of soil chemical and physical properties. Their vis–NIR spectra were collected from 400 to 2500 nm. The Shuttle Radar Topographic Mission 3-arcsec (90-m resolution) DEM of the area was used to derive 17 terrain attributes. A minimum data set of soil properties was selected to develop the SFI-SC. The SFI-SC consisted of three classes: Class 1, the highly fertile soils; Class 2, the fertile soils; and Class 3, the least fertile soils. It was derived heuristically with conditionals and using expert knowledge. The index was modeled with the spectra and terrain data using cross-validated decision trees. The cross-validation of the model correctly predicted Class 1 in 75% of cases, Class 2 in 61%, and Class 3 in 65%. A fertility map was derived for the study area and compared with a map of the GVI. Our approach offers a methodology that incorporates expert knowledge to derive the SFI-SC and uses a versatile spectro-spatial methodology that may be implemented for rapid and accurate determination of soil fertility and better exploration of areas suitable for production.

Abbreviations: Al_{sat}, aluminum saturation; B_{sat}, base saturation; CEC, cation exchange capacity; DEM, digital elevation model; GVI, green vegetation index; HOC, horizontal curvature; MAC, maximum curvature; MDS, minimum data set; OM, organic matter; PC, principal component; PCA, principal component analysis; ROC, regression and operating characteristic; SFI-SC, soil fertility index for sugarcane; vis-NIR, visible–near infrared.

Global demand for food, fuel, and energy is rapidly increasing, together with population growth. Consequently, there is an urgent need to expand areas where both food and energy can be produced. The production of biofuels as a renewable source of energy is rapidly growing because oil is nonrenewable, polluting, and contributes to the accumulation of CO_2 in the atmosphere and thus the greenhouse gas effect. Many countries (e.g., Brazil, Australia, and the United States) are considering the use of sugarcane for the production of biofuels. The attraction is the high capacity of sugarcane to produce ethanol as well as the cogeneration of electricity during processing and the retention of CO_2 by these plants. For this reason, some countries are looking at the rational expansion of sugarcane cropping, as their aim is to explore areas that are fertile and suitable for its production. Several factors affect the development of sugarcane, but soil–landscape interactions are particularly important (Cesar et al., 1987). The soil is the main supplier for nutrients, water, and physical support for the plant (Maule et al., 2001). Therefore, understanding the factors that affect crop production and their interactions across the landscape is important (Lespch, 1987).

Many techniques have been developed to efficiently quantify and qualify the potential of a soil to support plant growth (Letey et al., 2003; Rezaei et al., 2006).

Soil Sci. Soc. Am. J. 74:1293–1300
Published online 2 June 2010
doi:10.2136/sssaj2009.0130
Received 31 Mar. 2009.
*Corresponding author (raphael.viscarra-rossel@csiro.au).

Mostly, they involve the use of a minimum data set (MDS) of soil properties and derivation of a soil fertility (or quality) index (e.g., Ortega and Santibañez, 2007). A MDS is generally assembled with soil properties that reflect the condition and function of the soils and their variation in space and time in relation to a specific crop. The selected properties should be easy to measure, verifiable, and well related to their intended use. Once the MDS is prepared, the relevant index can be developed and used to provide the farmer or land manager with information important for decision making (Rezaei et al., 2006).

Currently, the development of such indices relies on the selection and measurement of the soil properties in the MDS using conventional soil analyses, which is laborious, time consuming, and costly. The challenge then is to develop a methodology that may be used to derive these indices from information that accurately relates to the soil and landscape and can be more easily and inexpensively acquired.

Proximal soil sensing using vis–NIR diffuse reflectance spectroscopy (400–2500 nm) can be used to acquire soil information (Viscarra Rossel and McBratney, 1998; Chang et al., 2001; Dematt̂e et al., 2004) because it is rapid, nondestructive, reproducible, and cost effective (Viscarra Rossel et al., 2006b). Furthermore, measurements can be made in situ (e.g., Viscarra Rossel et al., 2009). Visible–NIR reflectance from soil is affected by the chemistry and biology of its constituents as well as its physical structure. Absorption of vis–NIR light occurs due to overtones and combinations of fundamental molecular absorptions in the mid-infrared region. Most of these are associated with the soil mineralogy and organic materials (Brown et al., 2006; Viscarra Rossel et al., 2006a).

Digital elevation models and derived terrain attributes have been used as predictors of soil attributes in digital soil mapping (e.g., Moore et al., 1991; Behrens et al., 2009). Their use in soil mapping is based on the fact that the landscape controls the distribution of water, sediments, and dissolved material, resulting in different soil characteristics and affecting the availability of nutrients for plant growth (Jenny, 1941; Pachepsky et al., 2001; Girgin and Frazier, 2005; McBratney et al., 2003).

We have developed a methodology to determine soil fertility using vis–NIR diffuse reflectance spectra and terrain attributes derived from a DEM, using sugarcane as an example. There is some literature on the use of soil vis–NIR for determining soil fertility. For example, Vagen et al. (2005) tested the potential of vis–NIR soil spectral libraries for predicting and mapping soil properties in the highlands of Madagascar. They derived a spectral soil fertility index using 10 common agronomic indicators of soil fertility. They suggested that the method was able to separate the soils of the study area into ordinal soil condition classes. Awiti et al. (2007) evaluated the use of near-infrared spectroscopy as a tool for diagnosing the soil condition for agriculture and environmental management. They derived three soil conditions: good, average, and poor. Their study indicated that vis–NIR spectroscopy was a good tool to indicate areas for different management. Neither study, however, considered any spatial analysis. We propose that combining spectroscopy with spatial (terrain) analysis provides useful information necessary to determine soil fertility for crop production. The aims of our study were to: (i) assemble a MDS based on current agronomic practices and develop a SFI-SC for biofuel production in tropical soils, (ii) construct a model that may be used to predict the SFI-SC using soil vis–NIR spectra and terrain attributes, and (iii) produce a SFI-SC map for our study area and assess it by comparing it to a vegetation index map that is correlated to sugarcane yield.

MATERIALS AND METHODS

Site Description

The study was conducted in a 185-ha area in Rafard County, in the southwest of São Paulo State, Brazil (23°0′31.37′ S, 53°39′47.81′ W). The climate is mesothermic with dry winters, in which the average temperature of the coldest month (July) is 18°C and the hottest month (February) is 24°C. The geology and soils in this region are heterogeneous and complex, with basaltic, arenic, and shale parent materials. Of the 12 orders in Soil Taxonomy (Soil Survey Staff, 1999), five were found in the study area, including Oxisols, Entisols, Alfisols, Ultisols, and Inceptisols.

Soil Sampling and Analyses

The soil sampling scheme was based on a 100-m grid. We note that systematic sampling on a 100-m grid may not account for short-range soil variability. This is a limitation of our data that we intend to improve in future work. In total, 184 samples were collected from a depth of 0 to 20 cm using an auger. The soils were dried in an oven at 50°C for 48 h and later passed through a 2-mm sieve. The following properties were analyzed: clay content was measured by the densitometer method (de Camargo et al., 1986); soil pH was measured in a 1:5 soil/water ratio (pH_{H_2O}); the OM content was measured using the Walkley–Black technique; and Ca^{2+}, Mg^{2+}, and Al^{3+} were measured using an exchangeable cation resin (van Raij and Quaggio, 1989). These results were used to calculate the B_{sat}, CEC, and aluminum saturation (Al_{sat}).

Visible–Near-Infrared Spectroscopy

For the spectroscopic analysis, the dried <2-mm soil samples were placed in 9-cm-diameter petri dishes and scanned in the laboratory using an Infrared intelligent spectroradiometer. The sensor has a range between 400 and 2500 nm and a spectral resolution of 2 nm from 400 to 1000 nm and 4 nm from 1000 to 2500 nm. The scanning geometry was set in a perpendicular position in relation to the sample, with 27 cm between the sensor and the sample. The halogen lamp used for illumination was positioned 61 cm away from the sample, creating a 20° angle with the zenith. The absolute white reference used for calibration for the instrument was a Spectralon white reference plate.

Partial least squares regression with leave-one-out cross-validation using the software ParLeS v3.0 (Viscarra Rossel, 2008) was used to show that the spectra can be used to predict the soil properties in the MDS and that there is a basis for the use of spectra to predict the SFI-SC.

Terrain Analysis

The DEM used was the 3-arcsec (90-m resolution) Shuttle Radar Topographic Mission (SRTM) preprocessed by Jarvis et al. (2008). We used the SRTM data to provide a cost-effective approach based on freely available data. The terrain analyses were realized according to Behrens et al. (2009). In total, 17 terrain attributes were derived comprising local,

regional, and combined attributes. These were elevation, difference from aspect angle (DB 0, DB 45, DB 90, and DB 135), horizontal curvature (HOC), local elevation, waxing and waning slopes, relative profile curvature, topographic roughness, relative hillslope position, relative horizontal curvature, profile curvature, project distance from stream, mean curvature, minimal curvature, and maximum curvature (MAC). Details on each of these are given in Behrens et al. (2009).

Minimum Data Set and Soil Fertility Index for Sugarcane

The idea of a MDS is to provide a small yet sufficient set of soil properties that can be used for the assessment of soil quality, fertility, and function (Rezaei et al., 2006). In this study, we selected soil properties commonly used in that region of Brazil (e.g., Prado, 2005) and assembled the MDS for determining soil fertility for sugarcane production.

According to common practice in the study region, three soil fertility classes were defined: Class 1, representing soils that are highly fertile and most suitable for sugarcane production; Class 2, the fertile soils that may be improved to Class 1 with few additional inputs and management; and Class 3, representing soils that are the least fertile in the study area and are unsuitable for sugarcane or may require a great amount of input and management. Based on the soil sample data and the knowledge of local sugarcane growers, the SFI-SC was derived using conditional functions that used a sequence of (if, and) arguments with (then, else) clauses. These describe the interaction between the soil properties in the MDS and were paired to the three SFI-SC classes. The functions were adapted from current tables used for soil fertility classification for sugarcane production (Prado et al., 1998; Prado, 2005). In this way, knowledge from local experts was introduced in the conditionals and the development of the SFI-SC. Table 1 shows the data used to define these conditionals for each fertility class.

The soil properties used in the MDS were kriged to show their spatial distribution across the study area. We used ordinary kriging with either exponential or spherical semivariograms (Webster and Oliver, 2001).

Modeling the Soil Fertility Index for Sugarcane

Once the SFI-SC was established, it was modeled independently using boosted decision trees (Quinlan, 1993). The classification tree models were built with 10 boosting trials using the software See5 version 2.05 (Rulequest Research, Sydney, Australia). The following predictors were used: the first five principal component analysis (PCA) scores from the vis–NIR spectra and the 17 terrain attributes. The PCA was performed to compress the 625 vis–NIR wavelengths into fewer principal components (PCs) that accounted for a large proportion of the variance in the spectra. Prediction accuracy to derive the SFI-SC model was determined using 10-fold cross-validation. Therefore, in future analyses, this model could be used to predict soil fertility using only soil vis–NIR spectra and terrain data.

From the model output, we recorded the attribute usage and the importance of the variables used in the models. The former can be used to assess how the individual predictors contribute to the classifier by reporting the percentage of the training cases where the attribute is used in predicting a class. The latter is only a rough estimate used in the classifications (Quinlan, 1993).

Table 1. Soil properties of clay content, cation exchange capacity (CEC), base saturation (B_{sat}), and organic matter content (OM) used to define the conditionals in the soil fertility index for sugarcane (SFI-SC).

SFI-SC class	Clay	CEC	B_{sat}	OM
	g kg^{-1}	$mmol_c$ kg^{-1}	%	g kg^{-1}
1	>350	>150	>70	>25
2	150–350	50–150	50–70	15–25
3	<150	<50	<50	<15

Mapping the Soil Fertility Index for Sugarcane across the Study Area

To derive a map of soil fertility across our study area, we interpolated the five PCs and elevation onto a 30-m grid using ordinary point kriging (Webster and Oliver, 2001). This was necessary to provide reasonable spatial detail for the predictions of the SFI-SC across the study area. The semivariograms of the five PCs were modeled using exponential or spherical functions, while the semivariogram of elevation was modeled using a Gaussian function. For the PCs, the nugget (C_0), sill ($C_0 + C$) and range (a) parameters were recorded and the proportion of nugget to sill variance ($C_0/C_0 + C$) calculated for interpretation of their spatial distribution (Webster and Oliver, 2001). A small nugget compared with relatively large sill values indicates that the data are spatially structured, while a large nugget compared with relatively small sill values indicates the opposite.

Kriging of the predictor variables generated a total of 2177 points in the study area. The elevation on this grid was used to derive the 17 terrain attributes. The decision tree model was then generalized on the 30-m grid using the PCs and the terrain attributes to produce the SFI-SC map for the study area.

Assessment of the Digital Soil Fertility Index for Sugarcane Map across the Study Area

As an independent pseudo-validation of our results, the predicted SFI-SC across the study area was compared with a map of the GVI derived from Landsat images using a logistic regression and operating characteristic (ROC) curves (Hastie et al., 2001). We used the GVI instead of the normalized difference vegetation index, because the GVI suffer less influence from soil brightness and provides a better measure of vegetation cover and plant vigor (Malet, 1996). The logistic regression was used to assess the probability of having one of the SFI-SC classes as a smooth function of the GVI. The ROC curves were used to assess how well the SFI-SC compared with the GVI. They show a plot of the sensitivity by (1 − specificity), for each SFI-SC class. The sensitivity is the probability that the given GVI values correctly predict a class of the SFI-SC; the specificity is the probability that the given GVI values correctly predict a misclassification. Then, for a given GVI, the probability of incorrectly predicting a misclassification is 1 − specificity. The area under the ROC curve is a common index used to summarize the information contained in the curve, thus the larger the area, the better the fit. We also calculated Spearman's ρ correlation coefficient between the SFI-SC and the GVI across the entire study area.

Table 2. Distribution of the soil properties of clay content, cation exchange capacity (CEC), base saturation (B_{sat}), pH in water (pH_{H_2O}), organic matter content (OM), and Al saturation (Al_{sat})

Statistic	Clay	CEC	B_{sat}	pH_{H_2O}	OM	Al_{sat}
	g kg^{-1}	mmol$_c$ kg^{-1}	%		g kg^{-1}	%
Mean	322	112.3	69.2	5.6	17	6.9
SD	178	72.6	18.5	0.5	10	13.4
Range	60–670	14–471	13- 96	4.2–6.7	2–52	0.5–62.8
CV	55	65	27	10	60	195

RESULTS AND DISCUSSION

Minimum Data Set

The statistical distribution of the soil properties considered for the MDS are shown in Table 2 and their correlations are shown in Table 3.

Clay, CEC, and OM were positively correlated; B_{sat} and pH_{H_2O} were positively correlated, as is commonly observed in tropical soils (van Raij, 1991), and Al_{sat} and pH_{H_2O} were negatively correlated. These relationships are in good agreement with other studies of soil fertility characteristics for sugarcane production (e.g., Maule et al., 2001; Ciotta et al., 2003).

We did not use pH_{H_2O} or Al_{sat} in the MDS because these are not commonly used to assess soil fertility for sugarcane production in our study region. Furthermore, they are correlated to B_{sat} and CEC (Table 3). These findings are consistent with other work on Brazilian soils (e.g., Catani and Gallo, 1955). When developing the MDS, we did not use properties such as Ca and Mg, which were used, for example, by Vagen et al. (2005) and Awiti et al. (2007). The reason was that these attributes were well correlated to B_{sat} (r = 0.75 and 0.72, respectively). The final MDS used to develop the SFI-SC thus consisted of clay content, CEC, B_{sat}, and OM. The spatial distribution of the soil properties in the MDS across the study area is shown in Fig. 1.

The southern and southeastern portions of the study area exhibited good fertility characteristics, with high clay content, CEC, OM, and B_{sat}, while the northern and northwestern areas showed lower values for the measured soil properties.

Soil Fertility Index for Sugarcane

The statistics for the three fertility classes derived by the conditional functions are given in Table 4.

Clay content was significantly different in all three classes with the high-fertility class having an average clay content of 48%, the fertile class 27%, and the low-fertility class only 11% clay (Table 4). Clay content varied less in the low-fertility class. Similarly, CEC was greatest in Class 1 and decreased to relatively low values in Class 3. In all three fertility classes, the average B_{sat} was >50%, but only marginally in Class 3 with a B_{sat} of 56% (Table 4). In this class, there were instances where B_{sat} values were considerably <50%, which can restrict sugarcane production (Prado et al., 1998). The average OM was significantly different for all three classes, decreasing from Class 1 to Class 3 (Table 4). In weathered tropical soils with predominantly variable charge, OM provides a large portion of the CEC (Mendonça and Rowell, 1996). Organic matter also supports soil aggregation and water retention, which also improve soil fertility.

Table 3. Correlation coefficients for the soil properties of clay content, cation exchange capacity (CEC), base saturation (B_{sat}), pH in water (pH_{H_2O}), organic matter content (OM), and aluminum saturation (Al_{sat}). All correlations were significant at p = 0.05.

Property	Clay	CEC	B_{sat}	pH_{H_2O}	OM	Al_{sat}
	g kg^{-1}	mmol$_c$ kg^{-1}	%		g kg^{-1}	%
Clay	1					
CEC	0.69	1				
B_{sat}	0.40	0.52	1			
pH_{H_2O}	0.14	0.37	0.80	1		
OM	0.66	0.72	0.35	0.24	1	
Al_{sat}	–0.34	–0.35	–0.77	–0.69	–0.35	1

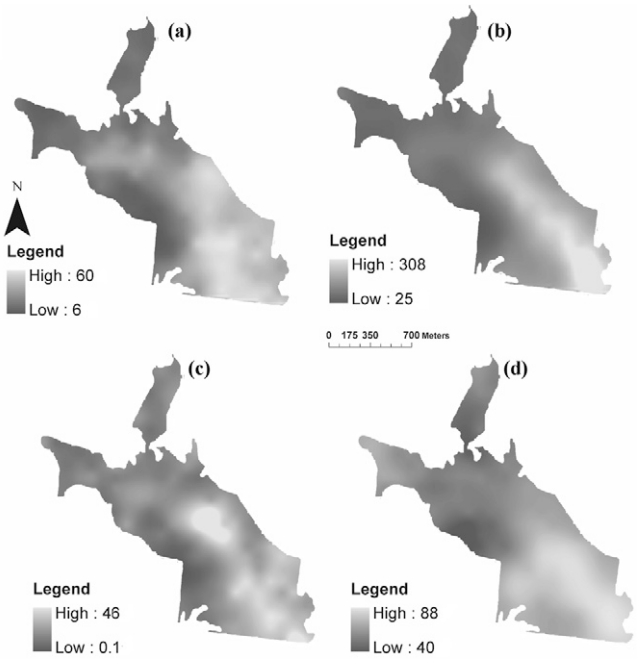

Fig. 1. Kriged soil attribute maps for the study area: (a) clay content (g kg^{-1}); (b) cation exchange capacity (CEC, mmol$_c$ kg^{-1}); (c) organic matter (OM, g kg^{-1}); (d) base saturation (B_{sat}, %).

Modeling the Soil Fertility Index for Sugarcane

The cross-validation results for the decision tree model of the SFI-SC derived from the conditionals are shown in Table 5.

The model was able to correctly classify the highly fertile Class 1 in 75% of the cases, the fertile Class 2 in 61%, and the least fertile Class 3 in 65% of all cases. The largest misclassifications occurred between Classes 2 and 3 (Table 5).

Table 4. Means ± standard deviations of the soil properties of clay content, cation exchange capacity (CEC), base saturation (B_{sat}), and organic matter content (OM) for the three conditional classes of the soil fertility index for sugarcane (SFI-SC).

SFI-SC class	Clay	CEC	B_{sat}	OM
	g kg^{-1}	mmol$_c$ kg^{-1}	%	g kg^{-1}
1	48.1 ± 11.6 a†	178.4 ± 66.8 a	80.5 ± 10.4 a	2.6 ± 0.9 a
2	26.5 ± 11.3 b	76.7 ± 16.7 b	64 ± 13.6 b	1.3 ± 0.5 b
3	11.2 ± 3.9 c	44.9 ± 10.6 c	56.1 ± 24.7 c	0.7 ± 0.3 c

† Means followed by different letters within a column indicate significant differences between the classes at α = 0.05.

Table 5. Contingency table of the cross-validation results for the decision tree model of the soil fertility index for sugarcane (SFI-SC) derived using the conditionals.

SFI-SC	Cross-validation prediction (%)		
	Class 1	Class 2	Class 3
Class 1	75	14	11
Class 2	20	65	15
Class 3	10	25	65

This model can be used to estimate soil fertility for sugarcane production in areas with similar growing conditions using only soil vis-NIR spectra and terrain data. For example, if a farmer wants to know whether he can grow sugarcane for biofuel production, then this model could be used to assess the fertility of the soil to help with the decision. One of the advantages of using conditionals to derive the SFI-SC is that they incorporate expert knowledge from the sugarcane producers into the methodology.

Importance of the Predictor Variables

The decision tree model used four of the five PCs and only three out of the 17 terrain predictor variables. The decision tree used PC1 and PC3 in 100% of the training cases, PC4 in 94%, and PC2 in 88%, while it used elevation in 93% of the training cases and MAC and HOC in 91 and 82%, respectively (Table 6).

As was expected, the most important variables were the PCs of the vis-NIR spectra (Table 6) as these are directly related to the mineral and organic composition of the soil and their properties. The terrain attributes had a lower contribution to the models because, in our models, they are used to provide additional information on the spatial environmental processes affecting the soil. Further work with other data and using different algorithms is needed to understand these relationships. There is evidence in the literature, however, suggesting that terrain attributes, and particularly curvatures, affect soil processes and thus the development of A horizons, the presence or absence of E horizons (e.g., Gessler et al., 1995), as well as soil properties such as texture and soil water content (e.g., Pachepsky et al., 2001).

Visible–Near-Infrared Spectra as Predictors

The partial least squares regression cross-validation results are shown in Table 7. They show the potential of the vis–NIR spectra to predict the soil attributes in the MDS. These results agree with other studies that have suggested that vis–NIR can be used to predict these soils properties (e.g., Chang et al., 2001; Vagen et al., 2005; Awiti et al., 2007; Viscarra Rossel et al., 2006b).

Table 6. Attribute usage and estimated importance of the predictor variables identified by the decision tree modeling.

Variable†	Attribute usage	Estimated importance
	%	
PC1	100	24
PC2	88	13
PC3	100	8
PC4	99	13
Elevation	93	1
MAC	91	3
HOC	82	1

† PC, principal component; MAC, maximum curvature; HOC, horizontal curvature.

Table 7. Assessment statistics for the partial least square regression cross-validation predictions of the soil properties of clay content, cation exchange capacity (CEC), base saturation (B_{sat}), and organic matter content (OM) in the minimum data.

Statistic	Clay	CEC	B_{sat}	OM
	g kg^{-1}	$mmol_c$ kg^{-1}	%	g kg^{-1}
No. of factors	11	15	8	7
R^2	0.89	0.74	0.38	0.70
RMSE	58.6	37.13	14.67	5.8

The semivariograms of the first four PCs, the resultant maps, and the PCA loadings of each are shown in Fig. 2. The semivariograms of the four PCs in Fig. 2 show that the range of spatial variation (a) and the nugget to sill ratio ($C_0/C_0 + C$) increases from the first to the fourth PC. That is, as the proportion of the variance in the spectra that is explained by consecutive PCs decreases, so does their spatial structure.

The first PC accounted for 71% of the variability in the spectra. The semivariogram of the first PC shows that the spatial variability of this PC across the study area is well structured, with a relatively small nugget variance and a nugget/sill ratio of 18% (Fig. 2a, top row). We fitted an exponential model to the data. The PCA loadings of the first PC suggest that this first component is related to soil color and related soil properties, such as OM, as well as Fe oxides with absorptions in the visible range and near 900 nm, respectively.

Principal Component 2 accounted for 16% of the total variance in the spectra and that which was not accounted for by PC 1. Its semivariogram was fitted with a spherical model, although a small trend was evident in the data at ranges beyond 700 m (Fig. 2b, top row). The nugget/sill ratio of PC 2 was 44%. The PCA loadings of PC 2 suggest that it is related to Fe oxides with absorptions near 900 nm and clay minerals with absorptions near 1950 and 2200 nm, which may account for clay minerals with water and Al–OH in their structure, respectively.

The PCA loadings of the remaining PCs are harder to interpret because these components account for relatively smaller, residual variances in the spectra. The PC maps (Fig. 2) provide good representations of the soil spatial variability in the study area (e.g., Odlare et al., 2005). Table 8 shows the correlation between the PCs and the soil attributes in the MDS: PC 1 was well correlated to clay content, OM, and CEC; PC 2 to CEC, clay content, and B_{sat}; PC 3 to OM, CEC, clay content, and B_{sat}; while PC 4 was

Table 8. Correlations between the soil properties of clay content, cation exchange capacity (CEC), base saturation (B_{sat}), and organic matter content (OM) of the minimum data set and the principal components (PCs).

Property	PC 1 (71%)	PC 2 (16%)	PC 3 (7%)	PC 4 (2%)
Clay	−0.78**	−0.21*	−0.30**	−0.05
CEC	−0.32**	−0.40**	−0.36**	0.12
B_{sat}, %	−0.08	−0.19**	−0.18*	−0.10
OM, %	−0.36**	−0.13	−0.49**	−0.006

* Significant at $P = 0.05$.

** Significant at $P = 0.01$.

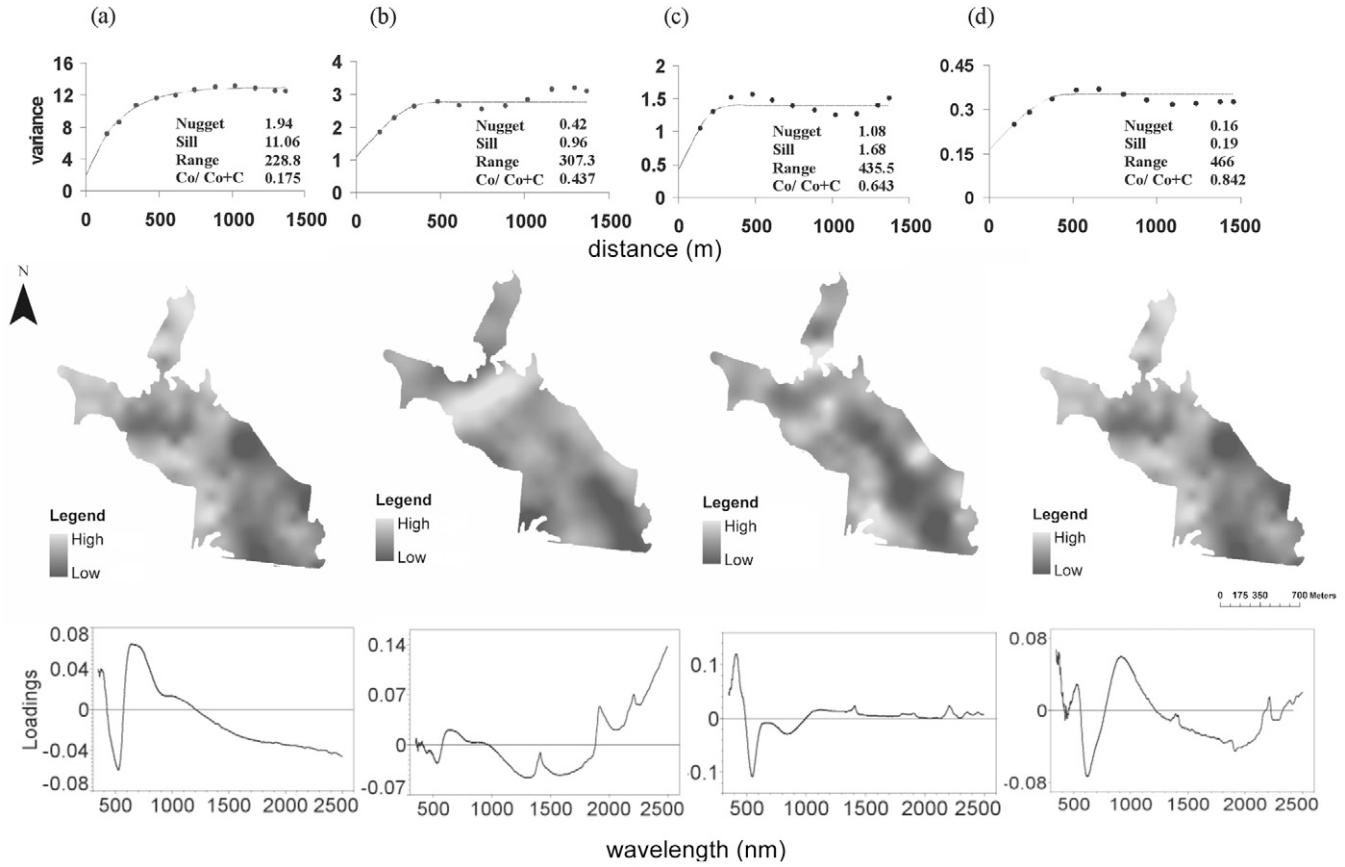

Fig. 2. Semivariograms of principal component (PC) scores, their kriged maps, and their loadings for (a) PC 1, (b) PC 2, (c) PC 3, and (d) PC 4.

less well correlated to these soil properties as this PC represents noise (Table 8).

Terrain Attributes as Predictors

Terrain attributes were less important in the models (Table 6). Nevertheless, elevation, MAC, and HOC were important in the predictions (Fig. 3).

Elevation can be related to various soil properties, including water movement and OM content, where higher elevations could represent smaller water and OM contents (e.g., Sumfleth and Duttmann, 2008). Terrain curvatures can affect A-horizon depth and the presence or absence of E horizons (Gessler et al., 1995). Pachepsky et al. (2001) suggested that profile curvature is a good predictor of soil texture and variations in soil water content.

Table 9 shows the correlation between the terrain attributes used in the models and the soil attributes in the MDS. Elevation was correlated to all soil properties in the MDS; MAC was correlated to clay content, OM, and CEC; and HOC was correlated to clay content and OM (Table 9).

Mapping of Soil Fertility across the Study Field

The SFI-SC map is shown in Fig. 4a. The GVI across the study area (Fig. 4b) provided us with a pseudo-independent measure of fertility.

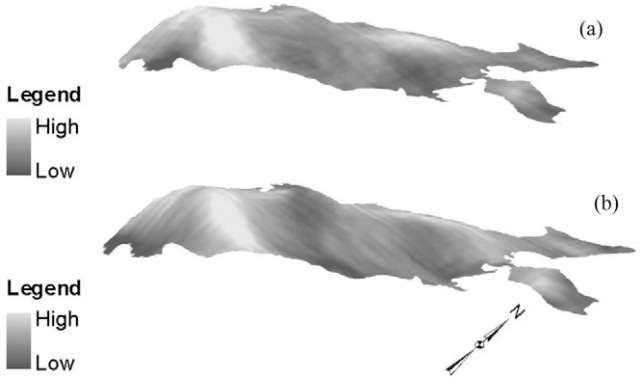

Fig. 3. Terrain attributes (a) maximum curvature (MAC) and (b) horizontal curvature (HOC) used in the model to predict soil fertility, draped over elevation.

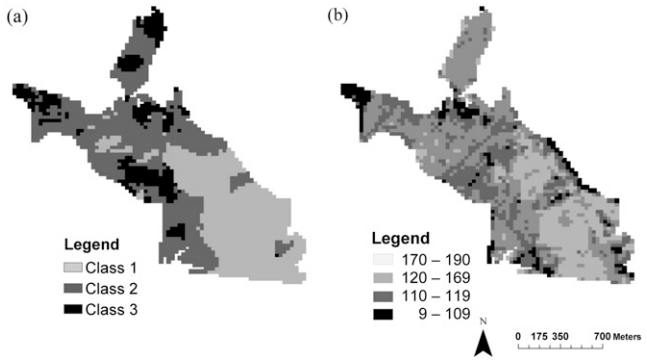

Fig. 4. Thematic maps for the study area: (a) the soil fertility map, and (b) the green vegetation index (GVI) map, which is highly correlated to sugarcane yield.

The logistic probability plot (Fig. 5a) shows that the probability attributed to Class 1 of the SFI-SC (the highly fertile soil) increases with increasing GVI values >150, which indicate the most vigorous plant growth. The distance between the lower and upper curves represents the probability attributed to Class 2 of the SFI-SC (the fertile soil), and it shows that it is larger for mid-range GVI values of around 110, which indicate good plant growth. The distance from the upper curve to the top of the graph represents the probability attributed to Class 3 of the SFI-SC (the least fertile soil), and it shows that it is larger for smaller GVI values, which represent poor plant growth (Fig. 5a). The ROC curves (Fig. 5b) complement these results, showing that Class 1 of the SFI-SC showed the best classification, corresponding to the higher GVI values, while Classes 2 and 3 were less well classified. The overall Spearman's ρ correlation coefficient between the GVI and the SFI-SC classes was 0.45.

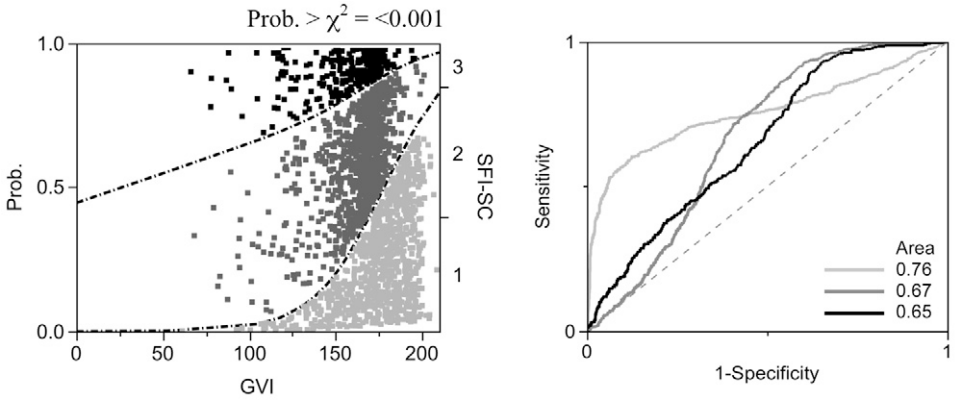

Fig. 5. Comparisons between the soil fertility index for sugarcane (SFI-SC) classes and the green vegetation index (GVI): (a) cumulative logistic probability plot showing the observed significance probability for the model χ^2 test, and (b) regression and operating characteristic (ROC) curves for the three SFI-SC classes, showing the areas under each of the curves.

Although the above comparisons are not strict validations of the technique presented here, they show that there is a fair relationship between the developed SFI-SC and the GVI. Appropriate validation would require ground measurements.

CONCLUSIONS

The soil attributes selected for the minimum data set—clay content, cation exchange capacity, bases saturation, and organic matter—were useful attributes for the assessment of soil fertility for sugarcane production in the region where our study site is located.

Visible–near-infrared spectra, a digital elevation model, and derived terrain attributes can be used to derive a soil fertility classification because (i) the spectra account for the soil mineral and organic composition and (ii) the terrain data account for soil and landscape interactions (e.g., water movement and nutrient availability), which are important for crop growth.

The cross-validation of the decision tree models showed greater misclassifications for Class 2, the fertile soils, and Class 3, the least fertile soils, than for Class 1, the highly fertile soils, which were more accurately classified.

A comparison between the derived SFI-SC and the GVI provided an independent pseudo-validation of our methodology. The results from the logistic comparison between the SFI-SC classes and the GVI were in good agreement with those from the cross-validation. Overall, the SFI-SC map was in fair agreement with the map of the GVI.

We presented a versatile spectro-spatial methodology that can be used for rapid and accurate determination of soil fertility for crop production.

Table 9. Correlations between the soil properties of clay content, cation exchange capacity (CEC), base saturation (B_{sat}), and organic matter content (OM) of the minimum data set and the terrain attributes of elevation, maximum curvature (MAC), and horizontal curvature (HOC).

Property	Elevation	MAC	HOC
Clay	0.47**	0.30**	0.22**
CEC	0.32**	0.19*	0.10
B_{sat}, %	0.30**	0.08	−0.02
OM, %	0.24**	0.22**	0.16*

* Significant at $P = 0.05$.

** Significant at $P = 0.01$.

REFERENCES

Awiti, A.O., M.G. Walsh, K.D. Shepherd, and J. Kinyamario. 2007. Soil condition classification using infrared spectroscopy: A proposition for assessment of soil condition along a tropical forest–cropland chronosequence. Geoderma 143:73–84.

Behrens, T., K. Schmidt, A.-X. Zhu, and T. Scholten. 2009. Multi-scale digital terrain analysis and feature selection for digital soil mapping. Geoderma 55:175–185..

Brown, D.J., K.D. Shepherd, M.G. Walsh, M.D. Mays, and T.G. Reinsch. 2006. Global soil characterization with VNIR diffuse reflectance spectroscopy. Geoderma 132:273–290.

Catani, R.A., and F.P. Gallo. 1955. Avaliação da exigência em calcário dos solos do Estado de São Paulo mediante correlação entre pH e a porcentagem de saturação em bases. Rev. Agric. Piracicaba 30:49–60.

Cesar, M.A.A., A.A. Thin, A.P. de Camargo, B.M.A. Bissolino, and F.C. Whistles. 1987. Capacidade de fosfatos naturais e artificiais em elevar o teor de fósforo no caldo de cana-de-açucar visando o processo industrial. STAB: Acúcar, Álcool Subptodutos 6:32–38.

Chang, C.W., D.A. Laird, M.J. Mausbach, and C.R. Hurburgh, Jr. 2001. Near-infrared reflectance spectroscopy: Principal component regression analyses of soil properties. Soil Sci. Soc. Am. J. 65:480–490.

Ciotta, M.N., C. Bayer, S.M.V. Fontoura, P.R. Ernani, and J.A. Albuquerque. 2003. Soil organic matter and cation exchange capacity increases in a low activity clay soil under no-tillage system. (In Portuguese, with English abstract.) Cienc. Rural 33:1161–1164.

de Camargo, A.O., A.C. Moniz, J.A. Jorge, and J.M. Valadares. 1986. Chemical, physical and mineralogical analyses methods of the IAC soils. (In Portuguese.) Bol. Téc. 106. Inst. Agron., Campinas, S.P., Brazil.

Demattê, J.A.M., R.C. Campos, M.C. Alves, P.R. Fiorio, and M.R. Nanni. 2004. Visible–NIR reflectance: A new approach on soil evaluation. Geoderma 121:95–112.

Gessler, P.E., I.D. Moore, N.J. McKenzie, and P.J. Ryan. 1995. Soil landscape modeling and spatial prediction of soil attributes. Int. J. Geogr. Inf. Sci. 9:421–432.

Girgin, B.N., and B.E. Frazier. 2005. Landscape position and surface curvature effects on soils developed in the Palouse area, Washington. Proc. SPIE 2818:61–69.

Hastie, T., R. Tibshirani, and J. Friedman. 2001. The elements of statistical learning.

Data mining, inference, and prediction. Springer, New York.
Jarvis, A., H.I. Reuter, A. Nelson, and E. Guevara. 2008. SRTM 90m digital elevation data, version 4. Available at srtm.csi.cgiar.org (verified 25 Apr. 2010). CGIAR Consortium for Spatial Inf., Int. Water Manage. Inst., Battaramulla, Sri Lanka.
Jenny, H. 1941. Factors of soil formation: A system of quantitative pedology. McGraw-Hill, New York.
Lespch, I.F. 1987. Influence of soil factors in the production. (In Portuguese.) p. 83–98. *In* P.R.C. Castro et al. (ed.) Ecofisiologia da produção. Assoc. Brasileira para Pesquisa da Potassa e do Fosfato, Piracicaba, S.P., Brazil.
Letey, J., R.E. Sojka, D.R. Upchurch, D.K. Cassel, K.R. Olsen, W.A. Payne, S.E. Petrie, G.H. Price, R.J. Reginato, H.D. Scott, P.J. Smethurst, and G.B. Triplett. 2003. Deficiencies in the soil quality concept and its application. J. Soil Water Conserv. 58:180–182.
Malet, P. 1996. Classifying the geometry of canopies from time variations of red and near infrared reflectance. Remote Sens. Environ. 56:164–171.
Maule, R.F., J.A. Mazza, and G.B. Martha, Jr. 2001. Different productivity of sugar cane varieties in different soils and harvest periods. (In Portuguese with English abstract.) Sci. Agric. 58:295–301.
McBratney, A.B., M.L. Mendonça Santos, and B. Minasny. 2003. On digital soil mapping. Geoderma 117:3–52.
Mendonça, E.S., and D.L. Rowell. 1996. Mineral and organic fractions of two Oxisols and their influence on the effective cation-exchange capacity. Soil Sci. Soc. Am. J. 60:1888–1892.
Moore, I.D., R.B. Grayson, and A.R. Ladson. 1991. Digital terrain modelling: Review of hydrological, geomorphological, and biological applications. Hydrol. Processes 5:3–30.
Odlare, M., K. Svensson, and M. Pell. 2005. Near infrared reflectance spectroscopy for assessment of spatial soil variation in an agricultural field. Geoderma 126:193–202.
Ortega, R.A., and O.A. Santibáñez. 2007. Determination of management zones in corn (*Zea mays* L.) based on soil fertility. Comput. Electron. Agric. 58:49–59.
Pachepsky, Y.A., D.J. Timlin, and W.J. Rawls. 2001. Soil water retention as related to topographic variables. Soil Sci. Soc. Am. J. 65:1787–1795.
Prado, H. 2005. Brazilian soils: Genesis, morphology, classification, mapping and management. (In Portuguese). H. Prado, Piracicaba, Brazil.
Prado, H., M.G.A. Landell, R. Rossetto, M.P. Campana, L. Zimback, and M.A. Silva. 1998. Relation between chemical sub surface conditions of subsoils and sugarcane yield. p. 232. *In* World Congr. of Soil Sci., 16th, Montpellier, France. 20–26 Aug. 1998. ORSTOM, Montpellier.
Quinlan, J.R. 1993. C4.5: Programs for machine learning. Morgan Kaufman, San Francisco.
Rezaei, S.A., R.J. Gilkes, and S.S. Andrews. 2006. A minimum data set for assessing soil quality in rangelands. Geoderma 136:229–234.
Soil Survey Staff. 1999. Soil Taxonomy: A basic system of soil classification for making and interpreting soil surveys. 2nd ed. Agric. Handbk. 436. U.S. Gov. Print. Office, Washington, DC.
Sumfleth, K., and R. Duttmann. 2008. Prediction of soil property distribution in paddy soil landscapes using terrain data and satellite information as indicators. Ecol. Indic. 8:485–501.
Vagen, T., K.D. Shepherd, and M.G. Walsh. 2005. Sensing landscape level change in soil fertility following deforestation and conversion in the highlands of Madagascar using vis-NIR spectroscopy. Geoderma 133:281–294.
van Raij, B. 1991. Fertilidade do solo e adubação. Agronômica Ceres, São Paulo, Brazil.
van Raij, B., and J.A. Quaggio. 1983. Métodos de análise de solo para fins de fertilidade. Bol. Téc. 81. Inst. Agron., Campinas, S.P., Brazil.
Viscarra Rossel, R.A. 2008. ParLeS: Software for chemometric analysis of spectroscopic data. Chemom. Intell. Lab. Syst. 90:72–83.
Viscarra Rossel, R.A., S.R. Cattle, A. Ortega, and Y. Fouad. 2009. In situ measurements of soil colour, mineral composition and clay content by vis–NIR spectroscopy. Geoderma 150:253–266.
Viscarra Rossel, R.A., and A.B. McBratney. 1998. Laboratory evaluation of a proximal sensing technique for simultaneous measurements of soil clay and water content. Geoderma 85:19–39.
Viscarra Rossel, R.A., R.N. McGlynn, and A.B. McBratney. 2006a. Determining the composition of mineral–organic mixes using UV–vis–NIR diffuse reflectance spectroscopy. Geoderma 137:70–82.
Viscarra Rossel, R.A., D.J.J. Walvoort, A.B. McBratney, L.J. Janik, and J.O. Skjemstad. 2006b. Visible, near-infrared, mid-infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131:59–75.
Webster, R., and M.A. Oliver. 2001. Geostatistics for environmental scientists. John Wiley & Sons, Chichester, UK.