Chapter

Pedometric Tools Applied to Zoning Management of Areas in Brazilian Semiarid Region

Helena Saraiva Koenow Pinheiro,
Pedro Armentano Mudado Xavier,
Lúcia Helena Cunha dos Anjos, Cesar da Silva Chagas
and Waldir de Carvalho Júnior

Abstract

Brazilian semi-arid region is a recent frontier in the country for agribusiness. The objective of this study is to apply pedometric tools to zoning areas with distinct potential and limitations to agricultural purposes. The research was set in three main steps: (i) to compile a database with all complete profile data collection; (ii) to analyze the vertical variability of soil properties and select a set of soil key properties useful to define the land potential and limitations; and (iii) to classify the area according to potential for agriculture, considering a medium technological level of the farmers. The quantitative methods applied are supported by geographic information systems (GIS) and spatial statistics. The soil data compilation was based on legacy data, with corresponding topographical data and information from remote sensing images of the area. Tree-based and geostatistical algorithms were applied to predict the spatial variability of the soil key properties. The definition of management zones was based on Iso Cluster and Maximum Likelihood Classification tools. The results pointed three different management zones according to risks of salinization and requirements for irrigation control. The approach showed to be a simple and useful way to select and recommend primarily potential areas for agriculture based on soil properties.

Keywords: AQP package, pedometrics, soil-depth functions, soil key properties, zoning management areas

1. Introduction

Soils are essential to human survival, and they provide a wide diversity of ecosystem services. Among them, soils are at the base of food production, carbon storage and cycles to atmosphere, water availability, turning biomass into valuable nutrients and degrading toxic elements, and supporting biodiversity [1]. Soil surveys are a relevant source of soil information and environmental data for many uses in agriculture and, otherwise, are fundamental for evaluating land capability and preventing degradation [2]. Ramalho Filho and Pereira [3] highlight the importance

of assessing regional conditions and soil properties to recommend adequate land-use management.

However, in Brazil, even though this is well-known, the soil class information is not commonly taken into account in planning or land-use zoning. One of the problems is the small scale of available maps from systematic soil surveys (1:250,000–1:1,000,000) in the country; this small scale is not suitable for recommendations regarding soil management at the farm level [4]. This limits the implementation of land-use zoning and soil sustainable management practices and policies.

The northeast region of Brazil presents complex and heterogeneous natural features in relation to rainfall, soil, and vegetation. It represents a huge challenge for the use and management of soil and water in sustainable agricultural systems. The soil degradation observed in some areas of this region is due to irregular rainfall, soil fertility conditions, and population pressures in a typically fragile environment [5].

Despite of the environmental variability, the northeast region of Brazil has a unique biome with a particular vegetation type (named "caatinga"). The majority of soils present shallow superficial horizons and profiles and very low organic carbon contents [6–8]. The rainfall deficit together with an inappropriate soil management for agricultural use can lead to desertification processes and to increase soil salinity in many areas, further degrading the land. Also common is the presence of rocks on the soil mass and/or surface layers and soils with high clay activity, both implying strong limitations for tillage. Thus, knowing the distribution of soil classes and properties is a requirement for land-use zoning.

The manipulation and analysis of a large soil profile collection, in the existing datasets, can be difficult due to the wide variability of horizon depths and thickness, as well as the variability among and between soil classes and orders. Addressing this issue, studies about the variability of soil properties along with depth were developed by [9–15]. Beaudette et al. [16] introduced the algorithm for quantitative pedology (AQP) package, which gathers tools to produce standard profile sketches highlighting differences in soil properties between horizons. It allows to standardize profile sketches, according to Munsell color chart, and dataset harmonization by applying the slice-wise algorithm, among other useful tools to express variations inside a given soil profile collection.

Quantitative studies have been developed in soil science through numerical modeling relationships between environmental variables and soil, which are applied to a geographic dataset to create a preliminary or predictive map [17, 18]. All these techniques are known as pedometrics [17], which represents an interface between statics and soil science, supporting predictive analysis related with soil classes and attributes [19, 20]. Based in pedometric concepts, the digital soil mapping (DSM) allows to study the relationships between soil and forming factors represented by spectral bands from remote sensing data, surface numerical models, and thematic maps [17]. Remote sensing data deserves an important place in digital soil mapping, particularly in flat landscapes, where morphometric covariates have lesser influence in soil formation [21].

Landsat spectral bands, particularly in the short-wave infrared range (SWIR), are commonly used to represent the environmental covariates of parent material and/or soil. Different mineral assemblages will have different spectral reflectances, which may be separable by analyzing bands 1–5 and 7 [22, 23]. The normalized difference vegetation index (NDVI) values range between −1 and +1. High positive values usually indicate the occurrence of dense green vegetation, pointing to an appropriate supply of water and nutrients. Low values express limited photosynthetic activity, and negative ones correspond to sparse lacking ground coverage [24]. The soil enhancement ratios of Landsat spectral band ratios 3:2, 3:7, and 5:7

have been interpreted to accentuate carbonate radicals, ferrous iron, and hydroxyl radicals, respectively, in exposed soil and geologic materials [25].

The identification of homogeneous zones of management consists in grouping areas whose environmental characteristics, as relief and soil, are similar. The design of homogeneous zones must be done according to the soil management, considering as well the risk of degradation, and it contributes to the sustainable land-use planning. Sánchez and Silva [26] pointed out that management areas present similar response when applied similar agronomic practices and they are subject to the same risks and limitations of agricultural use.

The cluster analysis deals with segmentation of a set of N objects into clusters (groups) in a fashion that the same type of datasets falls in a cluster that is different from those with dissimilar datasets [27]. The results of cluster analysis reveal internal data structure and improve understanding of data. There are many cluster algorithms that are available for data partitioning into. This analysis was used to delineate management zone methods [28], and the authors prefer to apply fuzzy cluster algorithms, and they are commonly used to define management zones for precision agriculture. To develop management zones [29], the cluster analysis was conducted by Management Zone Analyst 1.0.1 (Agricultural Research Services, University of Missouri, Columbia), with a combination of variables (elevation-Electrical Conductivity-EC, ECah-%Na; elevation-ECah-%Na, elevation-pH 1:1-%Na).

The hypothesis of this study is that the use of soil legacy data allied with modern tools that are able to analyze large datasets can assist the definition of soil potential and limitations to agricultural practices and thus assist the land-use planning and zoning. The work was developed from a legacy dataset of soil surveys made by the *Companhia de Desenvolvimento do Vale do São Francisco* (CODEVASF) [30] in the Brazilian semiarid region.

The specific goals of this study are (i) to identify soil key properties for definition of different management zones based on area databased and literature; (ii) to harmonize the dataset according predefined properties; (iii) to produce soil key property maps through digital soil mapping; and (iv) to create a map zoning the area according the potential to agricultural uses. This approach was used to establish the main trends from the profile collection by means of soil-depth functions and to select soil key properties adequate for the area (flatland landscape and semiarid climate), in this way, separating the areas according to their agriculture potentials.

2. Materials and methods

2.1 Study area and dataset

The study was conducted in an area located between 9°53′0″ and 9°36′30″ S and 40°34′30″ and 40°23′30″W (with 34,437.82 hectares), in the municipality of Juazeiro in Bahia State, northeast region of Brazil (**Figure 1**).

The biome in the region studied was the caatinga [31], an ecoregion characterized by xeromorphic vegetation with broadleaf thorny shrubs and trees that shed their leaves seasonally, cactus plants, and sparse arid-adapted grasses [32]. The climate is characterized as semiarid with dry winter and rainy summer, and the coldest month mean temperature remains above 18°C (BSwh' in the Köppen classification). The average annual rainfall is 400 mm, with the rainy season extending from November to April (highest precipitation in March); and average annual temperature is around 26°C. The xerothermic indexes are between 200 and 150, comprising of 7 to 8 dry months.
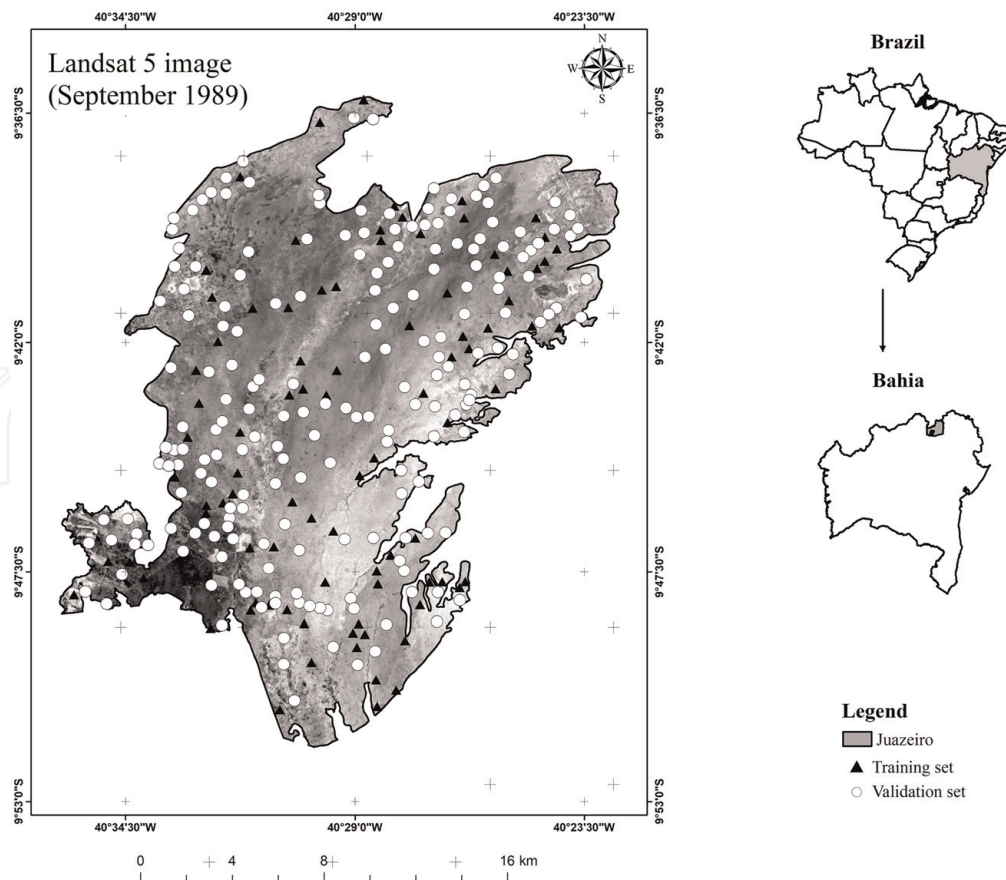
Figure 1.
Study area and location of soil profiles, over Landsat image (band 3).

At the study site, a portion of the original vegetation has been removed, and signs of land degradation such as bare soils are common. The landscape is mainly compounded by flat surfaces, with maximum slopes of 8%, and plateaus are common in the region. Geology comprises of limestone rocks from the caatinga formation, gneiss-granite rocks of the Caraiba-Paramirim complex, and recent colluvium/alluvium sediments [33].

The soils types in the northeast region of Brazil show large variation, according to parent material and relief, from shallow and with high content of basic cations (Ca, Mg, K, and Na) to deep and leached profiles [34]. Dominant soil classes according to the World Reference Base for Soil Resources [35] are Vertisols, Cambisols, and Planosols, with smaller extension of Regosols, Acrisols, and Luvisols.

2.2 Soil input data and covariates

The soil properties used in this study were sand, clay, and cation exchange capacity (CEC), determined according to [36]. The selection of soil properties was based on two main conditions: (i) presenting variability according to depth and ii) having importance to agricultural management (e.g., clay or sodium (Na) content).

From the original dataset, with 523 profiles, the ones with complete morphological description and analytical data were selected, performed by *Companhia de Desenvolvimento do Vale do São Francisco* (CODEVASF) in 1989 [29]. So, the input dataset comprised of the topsoil (0–20 cm) and subsoil (20–60 cm) layers for 290 soil profiles from a soil survey (legacy data). The sampling was performed according to the requirements of national soil survey service, correspondent to the detailed soil survey level [37, 38].

As covariates, Landsat 5 TM spectral bands were used in this study (**Table 1**), from September 1989, available in http://www.dgi.inpe.br/CDSR/. The NDVI values range between −1 and + 1. In the study area, most of the NDVI values were below 0.1, indicating little vegetation cover in the study area [39].

2.3 Modeling procedures

Quantitative soil properties, such as clay and sand content, can be estimated along the depth of soil profile through slice-wise algorithm [16] implemented in R software through algorithm for quantitative pedology (AQP) package [46]. The functions calculate an estimative of the main trend of values, represented by median and the variation interval by quartiles (25th and 75th of data distribution). The values shown in a column corresponds to percentage of soil profiles contributing with the function at each depth.

Initially, Pearson's linear correlation analysis was used to measure the linear association between variables, to determine among the soil properties defined by AQP, which correlated with environmental covariates (**Table 1**). This analysis was implemented in R [46], through the function cor.test, according to [42, 44, 47]. In Pearson's correlation the p-value defines whether or not two variables are statistically correlated, and for this study, it was defined that values lower than 0.02 indicate that the correlation is significant.

The modeling procedure to execute the random forest (RF) prediction was performed in R software, through randomForest (RF). RF is a nonparametric technique developed by Breiman [48] as an extension of classification and regression tree (CART) systems, to improve the performance of the predictors. To implement the RF models, three parameters are necessary: number of trees in the forest (ntree); minimum amount of data in each terminal node (nodesize); and number of covariates used in each tree (mtry) [49]. The ntree value was set to system default (500), although more stable results can be achieved with a larger number [50]. The

Covariates	Spatial resolution (m)	Spectral range (μm)*	References
Band 1	30	0.450–0.515	[23, 39–42]
Band 2	30	0.525–0.605	
Band 3	30	0.630–0.690	
Band 4	30	0.755–0.900	
Band 5	30	1.550–1.750	
Band 7	30	2.090–2.350	
NDVI (Band 4 - Band 3)/(Band 4 + Band 3)	30	—	[11, 23, 41–44]
Band 3/Band 2	30	—	
Band 3/Band 7	30	—	
Band 5/Band 7	30	—	
GSI (Band 3 - Band 1)/(Band 3 + Band 2 + Band 1)	30	—	[45]

Landsat 5 satellite

Table 1.
Covariates used in spatial prediction of soil properties.

nodesize value was set to 5 for each terminal node, as usually selected in regression studies. The mtry value chosen in this study was according to Liaw and Wiener [49], which proposes an amount corresponding to one third of the total number of predictor variables for regression problems.

Although Na showed a significant correlation at the two depths, according to the Pearson correlation analysis, the preliminary results using the random forest (RF) model were very unsatisfactory. Thus, exceptionally for this property, the RF model has been replaced by ordinary kriging (OK). Semivariograms were used to analyze the spatial structure of the Na, and to generate predictive maps, in both depths. The OK was performed in R software, through krige function [46]. OK model is the most familiar type of kriging and provides an accurate estimate for an area around a measure sample [51].

The model's performance was evaluated based on independent validation set, which was not used in the training procedure. Thereby, the 290 soil samples were randomly divided into 2 independent datasets in the R software; one of these was used in the training process (200 soil samples) and another for the validation process (90 soil samples). The analysis of the model's performance was based on the correlation between the measured values (validation samples) and estimated values, calculated by the coefficient of determination (R^2), the root mean square error (RMSE), and mean error (ME), presented as Eqs. (1) and (2):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} d_i^2} \tag{1}$$

$$ME = \frac{1}{n}\sum_{i=1}^{n} d_i \tag{2}$$

where "d" is the difference between the observed and estimated values and "n" is the number of samples used in the validation process.

The RMSE is a measure of the overall error of the estimation and commonly is used to estimate the error or uncertainty in places where the error was not measured directly; thereby, the higher the values of RMSE, the greater the differences between the datasets [52]. The ME gives the bias and allows evaluation of overestimation (positive values) or underestimation (negative values); values close to zero are preferable.

2.4 Definition of management zones

Management zones were defined in this study according to potential for agriculture, considering variability of soil key properties along profile depth, importance of soil properties for the land management, and the performance of the models to predict the spatial variation of the properties. Based on the maps for the selected soil key properties, an unsupervised classification was performed by using a series of input raster bands (Na, CEC, clay, and sand) using the Iso Cluster and Maximum Likelihood Classification tools from ArcGIS Desktop 10.3.

The Iso Cluster tool uses a modified iterative optimization clustering procedure, also known as the migrating means technique. The algorithm separates all cells into the user-specified number of distinct unimodal groups in the multidimensional space of the input bands; the iso prefix of the isodata clustering algorithm is an abbreviation for the iterative self-organizing way of performing clustering. In the clustering process, during each iteration, all samples are assigned to existing cluster centers, and new means are recalculated for every class. The optimal number of classes to specify is usually unknown. Therefore, it is advised to enter a

conservatively high number, analyze the resulting clusters, and rerun the function with a reduced number of classes [53].

3. Results and discussion

3.1 Algorithm for quantitative pedology (AQP)

The AQP package allows to gather a set of functions to work and to analyze large soil profile collections. Depth functions of soil key properties used to distinguish the profiles are presented in **Figure 2** and **Figure 3**. The percentage values plotted along the profile shows the relative quantity of profiles used in the soil-depth function to calculate the statistics (median and quartiles) at each depth.

The study area presents shallow soils, sometimes with high base saturation; but the main limitation for agricultural use is the soil texture and high clay activity, which will influence the soil moisture to adequately manage the soils due to their high plasticity.

The pH and the calcium (Ca) content presented a linear trend of median values along depth (**Figure 2**), even though a smaller number of soil profiles (less than 25%) were used in the estimative for deeper than 150 cm depth. However, different patterns were shown for potassium (K), sodium (Na), electrical conductivity (EC), and cation exchange capacity (CEC). The soil key properties Ca, K, and CEC showed more variability around the median value, thus presenting a better predictive potential to distinguishing soils with different parent materials and high clay activity, respectively.

It is pertinent to select which soil key properties could indicate differences in soil behavior along depth and are relevant for land management. From the analyses of **Figure 3**, it is possible to conclude that clay and sand contents have an opposite and large variability among the profiles of the collection; thus, they have great potential to aid in the definition of soil management zones. Since both properties are
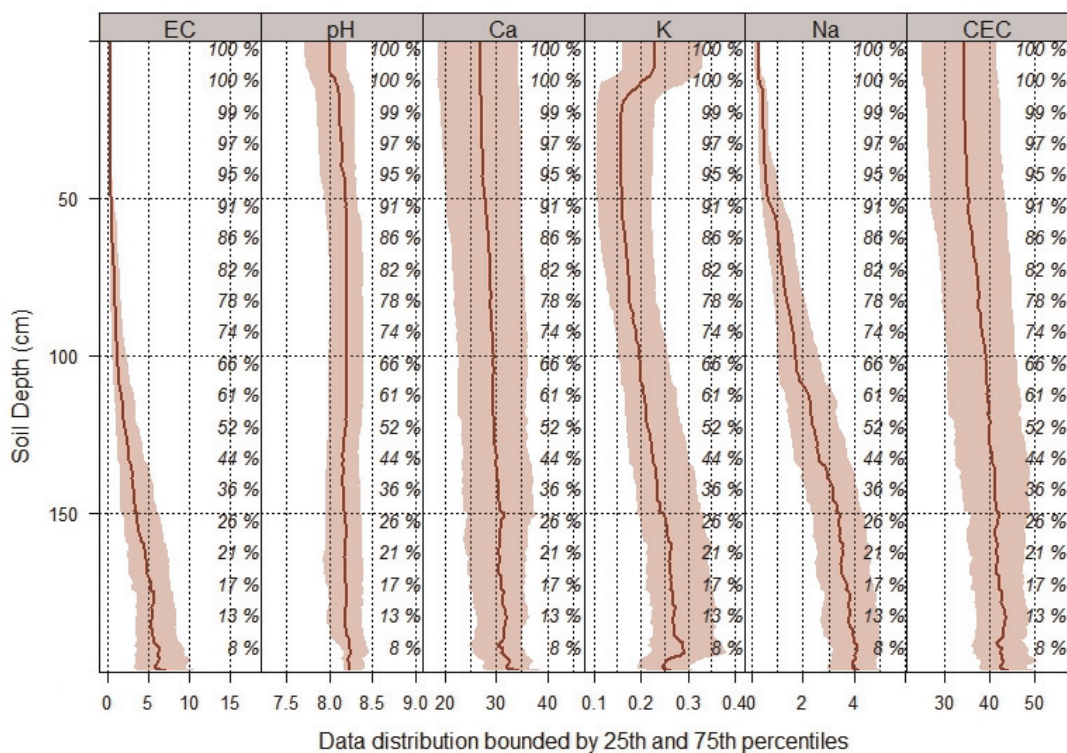

Figure 2.
Depth functions for soil key chemical properties. EC, electrical conductivity; CEC, cation exchange capacity.
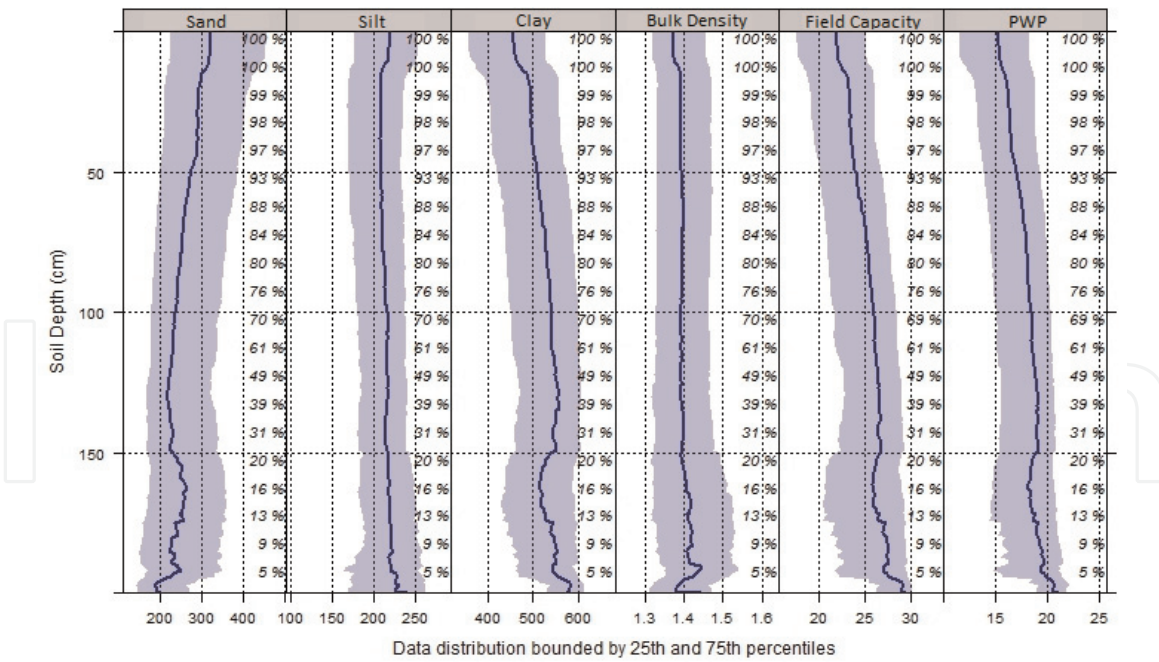
Figure 3.
Depth functions for soil key physical properties. PWP, permanent wilting point.

important to irrigation and mechanization practices, they were selected as criteria for agricultural zoning.

3.2 Predictive models

3.2.1 Pearson's correlation

The Pearson correlation coefficients (**Table 2**) showed that in general, the environmental covariates were significantly correlated with all soil properties analyzed ($p < 0.05$). The sand content was significantly correlated with covariates b5, b7,

Variables	Soil properties							
	Sand		Clay		CEC		Na	
Depth (cm)	0–20	20–60	0–20	20–60	0–20	20–60	0–20	20–60
b1	0.303	0.212	0.236	0.599	0.289	0.667	0.001	0.000
b2	0.198	0.321	0.431	0.177	0.848	0.723	0.001	0.000
b3	0.281	0.396	0.631	0.274	0.360	0.687	0.001	0.000
b4	0.037	0.040	0.009	0.072	0.002	0.016	0.019	0.000
b5	0.000	0.001	0.000	0.002	0.000	0.001	0.099	0.001
b7	0.000	0.002	0.000	0.003	0.000	0.000	0.152	0.001
NDVI	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
b3/b2	0.267	0.218	0.712	0.291	0.028	0.051	0.108	0.041
b3/b7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
b5/b7	0.702	0.990	0.274	0.718	0.000	0.001	0.617	0.082
GSI	0.000	0.000	0.000	0.000	0.649	0.493	0.010	0.000

Table 2.
Pearson's correlation (p-value) between soil properties (sand, clay, CEC, and Na) and environmental covariates (b1, b2, b3, b4, b5, b7, NDVI, b3/2, b3/b7, b5/b7, and GSI).

NDVI, b3/b7 ratio, and GSI (**Table 2**). These results do not agree with [44], who found no correlation between sand content and the Landsat 5 TM image data. Only the covariate b3/b7 was strongly correlated (r = −0.58 and −0.49, topsoil and subsoil, respectively) with the sand content [54], whereas the NDVI was moderately correlated (r = 0.39 and 0.36, topsoil and subsoil, respectively) and the other covariates weakly correlated, with r values below 0.27 (positive or negative) (**Table 3**).

Inverse relationships were observed between clay content and the environmental covariates, but magnitudes were the same as those observed for sand (**Table 2**), except for the covariate b4 (topsoil) which had no correlation with sand; similar results were reported by [42, 55]. The most relevant covariates were b3/b7 (r = −0.56 and −0.51, topsoil and subsoil, respectively) and NDVI (r = −0.36 and −0.37, topsoil and subsoil, respectively). These results vary among authors in the literature. Significant correlations were obtained between clay, the NDVI index, and b3/b2 and b5/b7 band ratios by [44], while there was no correlation between clay and the b3/b7 ratio band. On the other hand, Ahmed and Iqbal [56] found significant correlations between clay and bands 4 and 6 using Landsat 5 TM.

The CEC was significantly correlated with the covariates b4, b5, b7, NDVI, and b3/b7 and b5/b7 ratios (**Table 2**). Only the covariate b3/b7 was strongly correlated (r = −0.45 and −0.44, topsoil and subsoil, respectively) with the CEC [54], whereas the other covariates were weakly correlated, with r values below 0.27 (positive or negative) (**Table 3**). A strong correlation was observed by [55] between the CEC and covariates measured by ASTER spectral bands (1–8). The Na content was significantly correlated with most of the covariates, except with b7 (topsoil) and b3/b2 and b5/b7 ratios (**Table 2**). However, none of the covariates had a strong or moderate correlation with Na content (**Table 3**), which may explain the very low performance of the random forest model in the prediction of this soil property.

The study performed by Demattê et al. [57] highlighted that the correlation with a particular spectral band is directly related with soil characteristics in specific regions, explaining the differences between study cases.

Variables	Soil properties							
	Sand		Clay		CEC		Na	
Depth (cm)	0–20	20–60	0–20	20–60	0–20	20–60	0–20	20–60
b1	0.06	0.07	−0.07	−0.03	−0.06	−0.03	0.19	0.32
b2	−0.08	−0.06	0.05	0.08	−0.01	0.02	0.20	0.34
b3	−0.06	−0.05	0.03	0.06	−0.05	−0.02	0.19	0.32
b4	0.12	0.12	−0.15	−0.11	−0.18	−0.14	0.14	0.27
b5	0.23	0.20	−0.25	−0.19	−0.24	−0.20	0.10	0.20
b7	0.21	0.18	−0.24	−0.18	−0.27	−0.23	0.09	0.19
NDVI	0.39	0.36	−0.36	−0.37	−0.22	−0.23	−0.21	−0.27
B3/b2	0.07	0.07	−0.02	−0.06	0.13	0.11	−0.10	−0.12
b3/b7	−0.58	−0.49	0.56	0.51	0.45	0.44	0.24	0.29
b5/b7	−0.02	0.00	0.06	0.02	0.21	0.20	−0.03	−0.10
GSI	−0.27	−0.26	0.21	0.24	0.03	0.04	0.15	0.23

Table 3.
Pearson's correlation coefficient (r) between soil properties (sand, clay, CEC, and Na) and environmental covariates (b1, b2, b3, b4, b5, b7, NDVI, b3/2, b3/b7, b5/b7, and GSI).

3.2.2 Predictive models

In the training process, the RF model only maintains the covariates that had moderate or strong correlation with the soil properties (**Table 3**). The results obtained by the predictive models (RF and OK) using an independent dataset for validation (90 samples) are illustrated in **Figure 4**. In this study, the results obtained by RF models in the topsoil layer were higher than those obtained by
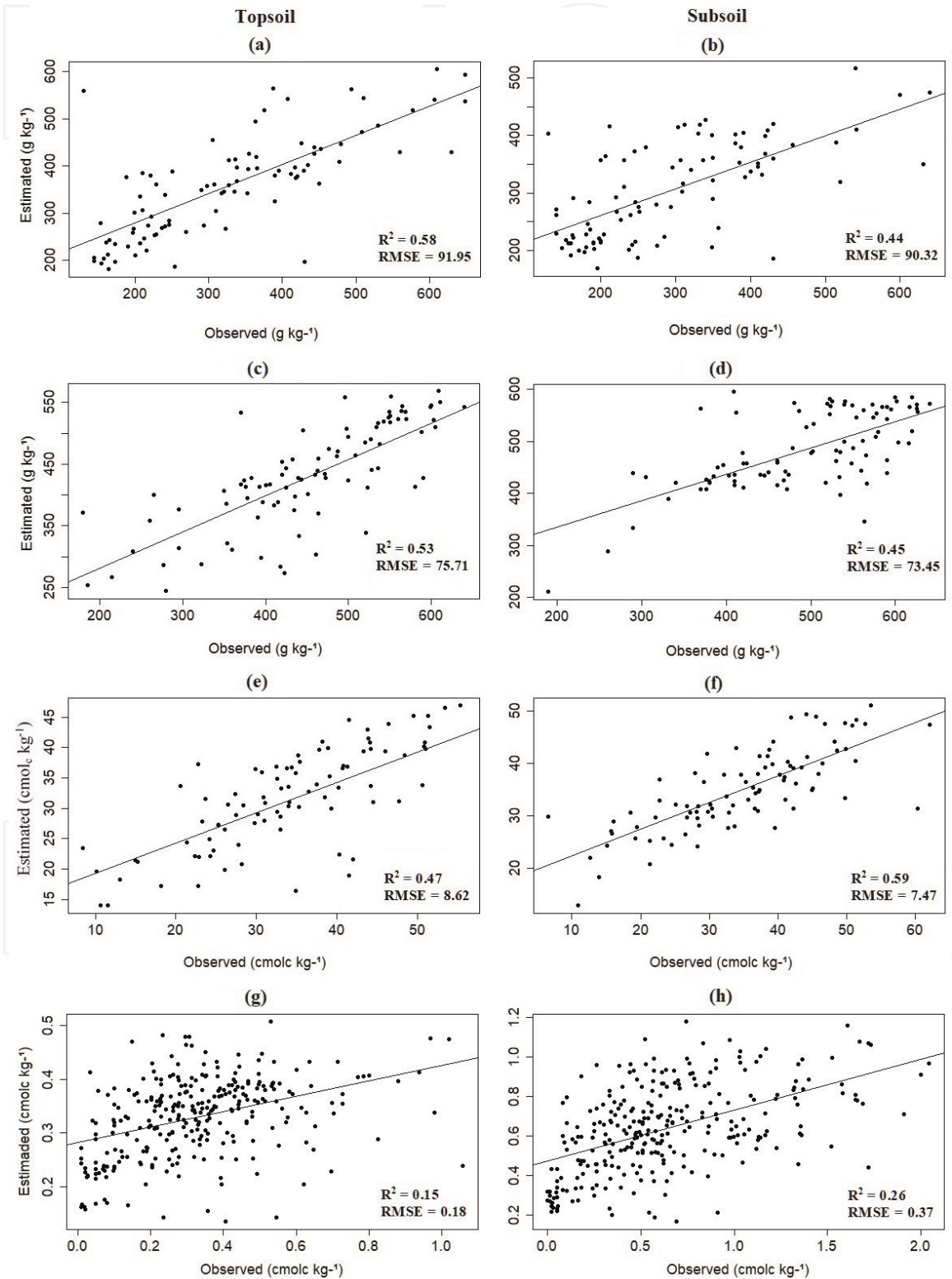

Figure 4.
Results obtained by prediction models using independent validation samples. (a) Topsoil sand—RF; (b) subsoil sand—RF; (c) topsoil clay—RF; (d) subsoil clay—RF; (e) topsoil CEC—RF; (f) subsoil CEC—RF; (g) topsoil Na—OK; (h) subsoil sand—OK.

subsoil for sand (R^2 = 0.58 and 0.44 and RMSE = 91.95 and 90.32 g kg^{-1}, topsoil and subsoil, respectively) and clay (R^2 = 0.53 and 0.45 and RMSE = 75.71 and 73.45 g kg^{-1}, topsoil and subsoil, respectively). The results were higher in subsoil than topsoil for CEC (R^2 = 0.44 and 0.59 and RMSE = 8.62 and 7.47 cmol$_c$ kg^{-1}, topsoil and subsoil, respectively) and Na content (R^2 = 0.15 and 0.26 and RMSE = 0.18 and 0.37 cmol$_c$ kg^{-1}, topsoil and subsoil, respectively) (**Figure 4**).

The results achieved by RF models were superior (topsoil) or similar (subsoil) to [58] that used terrain properties derived from a digital elevation model, for sand (variance explained = 30%) and clay (variance explained = 43%). In a study in Nigeria [59], percentages of variance explained (Var$_{ex}$) for RF models were of 48–49% for sand and 53–56% for clay in the top soil layer (0–15 cm). These values are inferior or similar to those obtained in this study for sand, similar for clay in the topsoil and superior in the subsoil. Therefore, the RMSE results for sand (19.26–19.67 g kg^{-1}) and clay (13.11–13.59 g kg^{-1}) were lower than those in the present study for the topsoil and subsoil layers (**Figure 4**). Lower Var$_{ex}$ values were reported by [60] for sand (33–35%) and clay (31–35%).

The Var$_{ex}$ obtained for CEC using the validation samples was 47% (R2 = 0.47) for topsoil layer and 59% (R2 = 0.59) for subsoil layer. The goodness of fit estimated by RMSE was 8.62 and 7.47 cmol$_c$ kg^{-1} (topsoil and subsoil, respectively). Regarding the model's performance, the results can be considered satisfactory; besides that, few studies in literature have used RF to predict soil CEC, and no one has used only remote sensing data as main covariates. The present results showed worst performance from statistical indexes when compared with those obtained by Lagacherie et al. [61], which reached 79% for Var$_{ex}$ for the layer between 15 and 30 cm and 3.4 cmol$_c$ kg^{-1} for RMSE, using as input data terrain attributes and hyperspectral data in the visible and near infrared (AISA-Dual) with 5 m of spatial

Statistics	Dataset				Predictive model			
					Random forest			Ordinary kriging
	Sand	Clay	CEC	Na	Sand	Clay	CEC	Na
	g kg^{-1}		cmolc kg^{-1}		g kg^{-1}			cmolc kg^{-1}
	Topsoil							
Min	75	133	5.71	0.01	175	235	12.30	0.14
Max	802	640	61.41	3.59	626	594	54.33	0.57
Mean	334	449	33.29	0.37	338	449	34.27	0.34
SD	144	113	11.91	0.37	103	78	8.52	0.07
CV	43	25	36	100	30	17	25	206
	Subsoil							
Min	100	160	6.59	0.01	165	204	10.69	0.17
Max	672	673	62.15	7.99	576	623	51.28	1.26
Mean	308	484	33.99	0.71	308	483	35.24	0.65
SD	125	102	11.54	0.69	85	76	7.86	0.21
CV	41	21	34	97	28	16	22	32

Min, minimum; Max, maximum; SD, standard deviation; CV, coefficient of variation

Table 4.
Descriptive statistics from soil samples and predictive models for topsoil and subsoil.
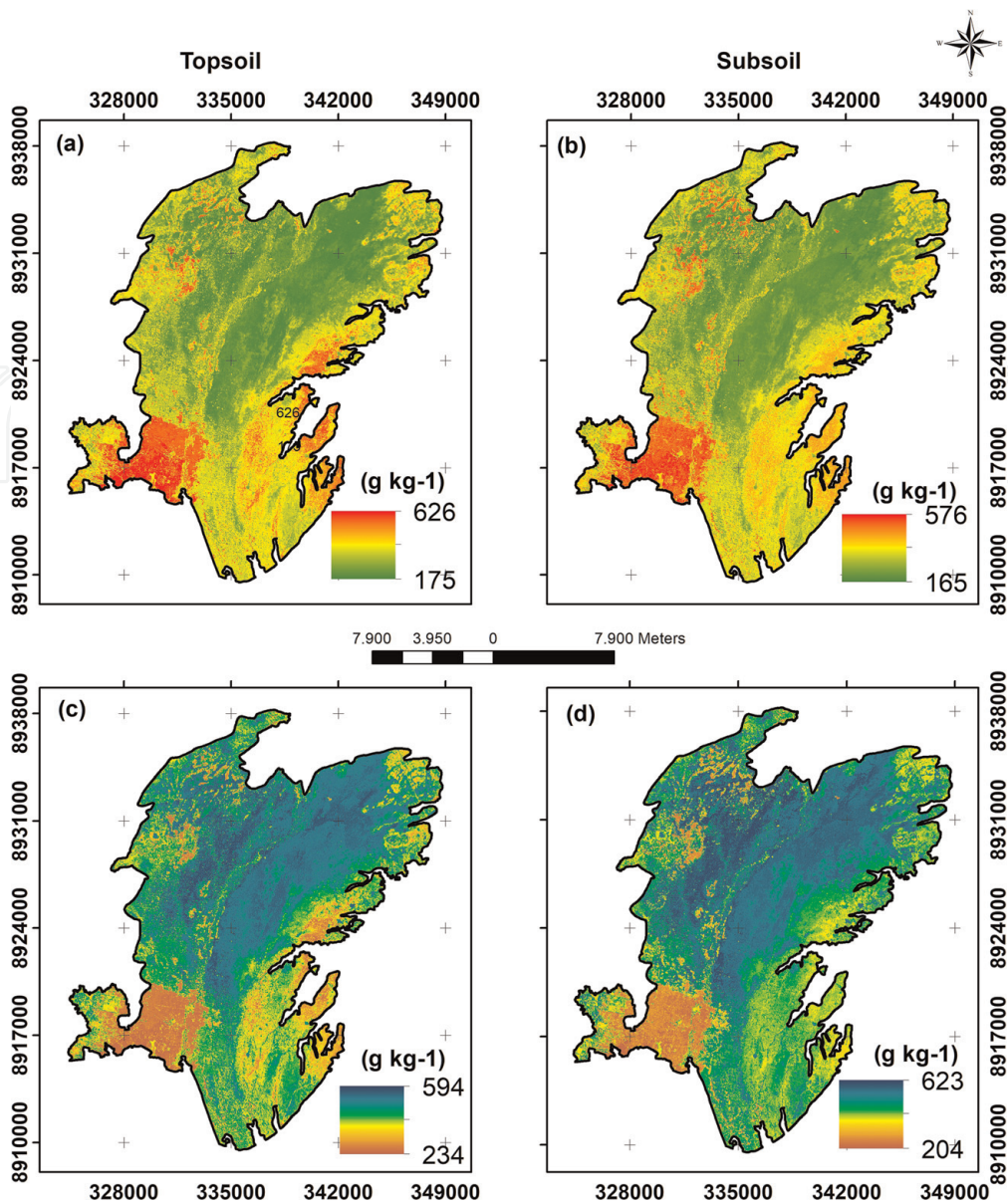
Figure 5.
Spatial distribution of soil physical properties estimated by the RK models. (a) Topsoil sand; (b) subsoil sand; (c) topsoil clay; (d) subsoil clay.

resolution. The difference from this study may be related to the coarser spatial resolution of images from Landsat 5 (30 m), in comparison with Lagacherie et al. [61] who used hyperspectral data (5 m). The influence of spatial resolution in the prediction of soil properties is reported by other studies [62, 63].

However, in this study the performance of metrics for CEC prediction was better when compared to the values (9.49 $cmol_c$ kg^{-1}) achieved by [60]. The poor performance was explained by the authors as due to the small-scale variation of parental material and erosion/deposition rates, which were not captured by the spatial resolution of the covariates (100 m). The authors also highlighted importance of input dataset to improve the models' performance.

The semivariogram obtained for Na content (**Figure 4**) provides a description of the spatial dependence and indicates processes related with the spatial distribution [65]. For both depths, the best semivariogram was the exponential model. The semivariograms for topsoil and for subsoil (**Figure 4**) used to estimate the Na content present a R2 of 0.26 and RMSE of 0.37 $cmol_c$ kg^{-1} and R2 of 0.15 and RMSE of 0.18 $cmol_c$ kg^{-1}, respectively. The descriptive statistics of soil properties prediction for the predictive models are presented in **Table 4**.
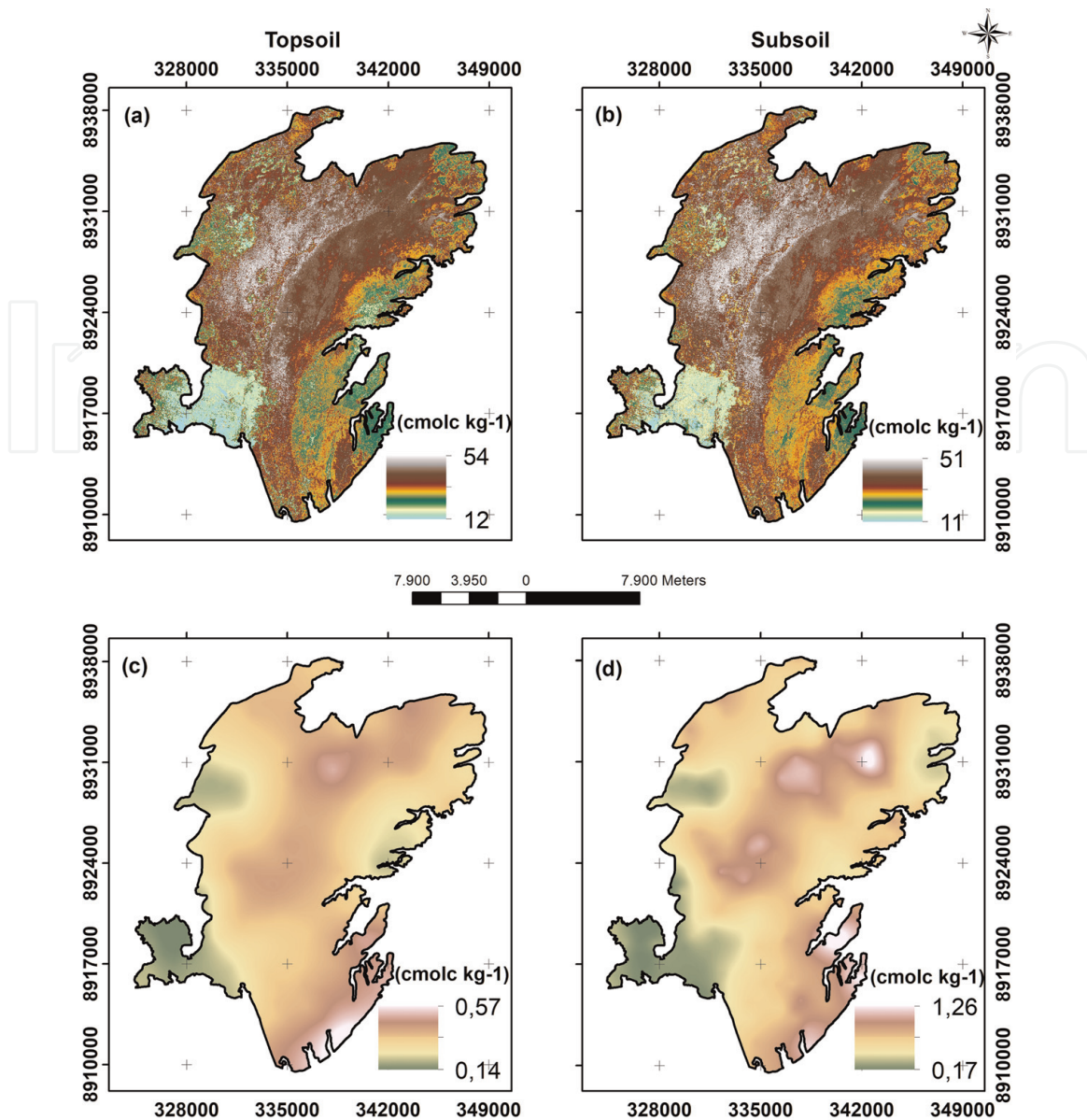
Figure 6.
Spatial distribution of soil chemical properties estimated by the RF and OK models. (a) Topsoil CEC; (b) subsoil CEC; (c) topsoil Na content; (d) subsoil Na content.

The RF model produced predicted values for sand, clay, and CEC within the range of the original values (**Table 4**), with a smaller standard deviation and coefficient of variation, as expected for this model [64]. The same was true for Na content in the OK model. However, the map produced presented a smaller range of values than the original dataset, which, according to [65], means that this model had low accuracy to describe the spatial variation of Na content, corroborating results found when predicting this property at both depths. The mean values for sand, clay, CEC, and Na content in both depths (**Table 4**) are closer to the values of the original dataset. The CV for both models was smaller than the CV from the original dataset, except for Na content in the topsoil.

These results could be explained by the moderate correlation of soil properties with the covariates and predominance of short-scale variations that could not be modeled from the set of profiles used. The results were considered satisfactory, except for Na content, and they can be ascribed to the physical interference of soil properties in the incident and reflected energy. However, the quantification of soil properties using an orbital sensor is not an easy task due to complexity of soil

dynamics and formation [66]. The spatial distribution of soil properties according to the predictive models is shown in **Figures 5** and **6**.

3.3 Definition of the management zones

Based on the maps created for each soil key property, the zoning procedure was applied by using Iso Cluster and Maximum Likelihood Classification tools. The descriptive statistics and the map produced are presented on **Table 5** and **Figure** 7, respectively.

Management zones	MIN	MAX	Range	Mean	STD
Na (0–20 cm)					
1	0.14	0.57	0.43	0.35	0.05
2	0.14	0.57	0.43	0.33	0.08
3	0.14	0.57	0.43	0.29	0.09
Na (20–60 cm)					
1	0.17	1.21	1.05	0.72	0.16
2	0.17	1.25	1.09	0.62	0.19
3	0.17	1.26	1.10	0.52	0.29
CEC (0–20 cm)					
1	26.90	54.33	27.43	41.88	3.87
2	15.17	47.03	31.86	31.55	4.40
3	12.30	35.81	23.50	20.85	3.63
CEC (20–60 cm)					
1	24.32	51.28	26.96	41.70	4.29
2	14.39	48.22	33.83	33.18	4.46
3	10.69	37.09	26.41	23.15	4.48
Clay content (0–20 cm)					
1	401.4	593.7	192.3	520.1	25.9
2	321.2	536.7	215.5	427.2	27.7
3	234.2	501.3	267.1	313.0	38.3
Clay content (20–60 cm)					
1	457.3	623.0	165.7	552.4	26.0
2	325.8	574.9	249.0	457.2	30.9
3	204.2	488.3	284.1	362.4	55.3
Sand content (0–20 cm)					
1	175.0	370.9	195.9	242.7	31.6
2	221.0	538.1	317.1	369.1	37.2
3	230.7	626.1	395.5	512.3	51.8
Sand content (20–60 cm)					
1	164.5	321.4	156.9	228.8	24.3
2	201.5	518.3	316.9	335.7	37.9
3	209.9	576.3	366.4	447.0	47.5

MIN, minimum; MAX, maximum; STD, standard deviation

Table 5.
Descriptive statistics of the management zones separated by layers 0–20 and 20–60 cm.

The conditions to define the agricultural zones have to be adjusted according to available data and heterogeneity of soil properties in the study area, in this case located in a semiarid region.

Zone 1 presents the greater amount of Na in both layers, and it is associated with greater values of CEC as indicated by the mean values, than Zones 2 and 3. The same condition is verified for clay content, where Zone 1 has the greatest mean value in both layers (0–20 and 20–60 cm). As expected, sand content shows the
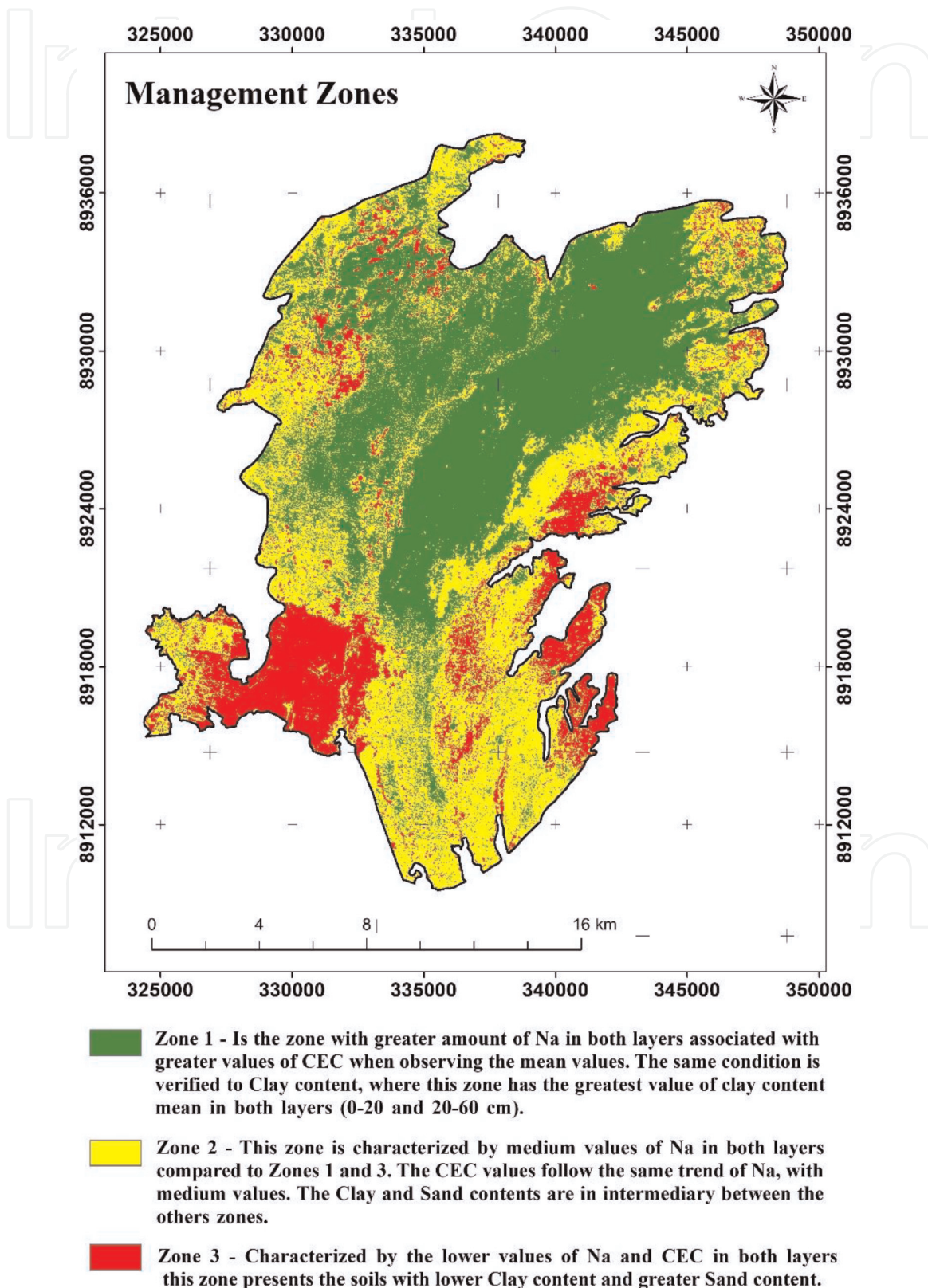

Zone 1 - Is the zone with greater amount of Na in both layers associated with greater values of CEC when observing the mean values. The same condition is verified to Clay content, where this zone has the greatest value of clay content mean in both layers (0-20 and 20-60 cm).

Zone 2 - This zone is characterized by medium values of Na in both layers compared to Zones 1 and 3. The CEC values follow the same trend of Na, with medium values. The Clay and Sand contents are in intermediary between the others zones.

Zone 3 - Characterized by the lower values of Na and CEC in both layers this zone presents the soils with lower Clay content and greater Sand content.

Figure 7.
Zoning of the study area according to selected soil key properties.

opposite pattern. Zone 1 occupies 42.6% of the area (1467 ha), and it has greater risk of salinization due to the higher mean values of Na. Even though Zone 1 has the greater mean values of CEC, it is associated to the high Na content and high content of clay, and possibly high clay activity (2:1). These properties limit the soil potential and irrigation management, further increasing salinization risks.

Zone 2 is characterized by medium values of Na in both layers, when compared to Zones 1 and 3. The CEC values follow the same trend of Na, with medium values, and clay and sand contents are intermediary. This zone occupies 41.7% of the area with 1436 ha. Zone 2 presents medium effort and limitation to soil management, considering risks of salinization and requirements for irrigation control.

Zone 3 is characterized by the lowest values of Na and CEC in both layers; it has the soils with lowest clay contents and proportionally greatest sand contents. This zone occupies 15.7% of the area with 539 ha. Zone 3 has the lowest mean values of CEC and Na, which decrease risks of salinization; plus, the lower clay content is less limiting to soil mechanization.

4. Conclusions

The approach based on the combination of orbital remote sensing data, with random forest (RF) and OK models to predict properties of soils in the Brazilian semiarid region, was considered satisfactory for sand, clay and CEC contents, and unsatisfactory for Na content.

The use of remote sensing data to map soil properties in flat relief areas, where common topographic covariates may not result in adequate maps by digital soil technics, is an advance. However, more research is needed to improve the quality of the covariates used as an input data.

The zoning procedures through multivariate analysis allowed to identify areas with higher salinization risks and where the management of irrigation is more complex. In the same way, the approach allowed to recognize zone of soils with lower sodium content and clay activity, which are preferred to grow annual crops.

Author details

Helena Saraiva Koenow Pinheiro[1*], Pedro Armentano Mudado Xavier[1],
Lúcia Helena Cunha dos Anjos[1], Cesar da Silva Chagas[2]
and Waldir de Carvalho Júnior[2]

1 Soil's Department, Agronomy Institute, Federal Rural University of Rio de Janeiro (UFRRJ), Seropédica, RJ, Brazil

2 Embrapa Solos, Rio de Janeiro, RJ, Brazil

*Address all correspondence to: lenask@gmail.com

IntechOpen

References

[1] FAO, ITPS. Status of the World's Soil Resources (SWSR)—Technical Summary. Rome, Italy: Food and Agriculture Organization of the United Nations and Intergovernmental Technical Panel on Soils; 2015

[2] Beek KJ. Land Evaluation for Agricultural Development: Some Explorations of Land-Use Systems Analysis with Particular Reference to Latin America. Wageningen: International Institute for Land Reclamation and Improvement; 1978. 333p. (ILRI Publication, 23)

[3] Ramalho Filho A, Pereira LC. Aptidão agrícola das terras do Brasil: potencial de terras e análise dos principais métodos de avaliação. Rio de Janeiro: Embrapa Solos; 1999. 36p. (Documentos, 1). (in Portuguese)

[4] Streck EV, Kämpf N, Dalmolin RSD, Klamt E, Nascimento PC, Schneider P, et al. Solos do Rio Grande do Sul. 2.ed ed. rev. e ampl. Porto Alegre: Emater/RS; 2008. 222p. (in Portuguese)

[5] Araújo Filho JÁ. Manejo Sustentável da Caatinga. Recife: Projeto Dom Helder Camara; 2013. 200 p. (in Portuguese)

[6] Lepsch IF. Formação e Conservação dos Solos. São Paulo: Oficina de Textos; 2002. 178p. (in Portuguese)

[7] Boletim informativo do Núcleo Regional Nordeste da Sociedade Brasileira de Ciência do Solo. Cruz das Almas, BA: NRNE/SBCS; 2017:1(1) (in Portuguese)

[8] de Castro SS, Hernani LC. Solos frágeis: caracterização, manejo e sustentabilidade. Brasília, DF: Embrapa; 2015. 367 p. (in Portuguese)

[9] Bishop TFA, McBratney AB, Laslett GM. Modelling soil attribute depth functions with equal-area quadratic smoothing splines. Geoderma. 1999;**91**:27-45. DOI: 10.1016/S0016-7061(99)00003-8

[10] Mishra U, Lal R, Slater B, Calhoun F, Liu D, Meirvenne MV. Predicting soil organic carbon stock using profile depth distribution functions and ordinary kriging. Soil Science Society of America Journal. 2007;**73**:614-621

[11] Malone BP, Minasny B, McBratney AB. Mapping continuous soil depth functions in the Edgeroi District, NSW, Australia, using terrain attributes and other environmental factors. In: Purves R, Gruber S, Straumann R, editors. Proceedings of Geomorphometry. Zurich: University of Zurich. 2009. pp. 90-97

[12] Malone BP, McBratney AB, Minasny B. Empirical estimates of uncertainty for mapping continuous depth functions of soil attributes. Geoderma. 2011;**160**:614-626

[13] Lacoste M, Minasny B, McBratney AB, Micho D, Viaud V, Walter C. High-resolution 3D mapping of soil organic carbon in a heterogeneous agricultural landscape. Geoderma. 2014;**213**:296-311

[14] Wiese L, Ros I, Rozanov A, Boshoff A, de Clercq W, Seifert T. An approach to soil carbon accounting and mapping using vertical distribution functions for known soil types. Geoderma. 2015;**263**:264-273

[15] Liu F, Rossiter DG, Song XD, Zhang GL, Yang RM, Zhao YG, et al. A similarity-based method for three-dimensional prediction of soil organic matter concentration. Geoderma. 2016;**263**:254-263

[16] Beaudette DE, Roudier P, O'Geen AT. Algorithms for quantitative

pedology: A toolkit for soil scientists. Computers and Geosciences. 2013;**52**: 258-268

[17] McBratney AB, Odeh IOA, Bishop TFA, Dunbar MS, Shatar TM. An overview of pedometric techniques for use in soil survey. Geoderma. 2000;**97**: 293-327. DOI: 10.1016/S0016-7061(00) 00043-4

[18] Ten Caten A. Mapeamento digital de solos: metodologias para atender a demanda por informação espacial em solos [thesis]. Universidade Federal de Santa Maria; 2011. 108p. (in Portuguese)

[19] Silveira CT, Oka-Fiori C, Santos LJC, Sirtoli AV, Silva CR. Pedometria apoiada em atributos topográficos com operações de tabulação cruzada por álgebra de mapas. Revista Brasileira de Geomorfologia. 2012;**13**(2):125-137. (in Portuguese)

[20] Coelho FF, Giasson E. Métodos para mapeamento digital de solos com utilização de sistema de informação geográfica. Revista Ciência Rural, Santa Maria. 2010;**40**(10):2099-2106. (in Portuguese)

[21] Chagas CS, Carvalho Júnior W, Pinheiro HSK, Xavier PAM, Bhering SB, Pereira NR, et al. Mapping soil cation exchange capacity in a semiarid region through predictive models and covariates from remote sensing data. Revista Brasileira de Ciência do Solo. 2018;**42**:1-12, e0170183. DOI: 10.1590/18069657rbcs20170183

[22] Sirtoli AE. Atributos do terreno e índices espectrais integrados por redes neurais artificiais [thesis]. Curitiba, PR: UFPR; 2008. 114p. (in Portuguese)

[23] Boettinger JL, Ramsey RD, Bodily JM, Cole NJ, Kienast-Brown S, Nield SJ, et al. Landsat spectral data for digital soil mapping. In: Hartemink AE, McBratney AB, Mendonça-Santos ML,

editors. Digital Soil Mapping with Limited Data. New York: Springer-Verlag; 2008. pp. 192-202

[24] Sumfleth K, Duttmann R. Prediction of soil property distribution in paddy soil landscapes using terrain data and satellite information as indicators. Ecological Indicators. 2008;**8**:485-501

[25] Amen A, Blaszczynski J. Integrated Landscape Analysis. Denver, CO: U.S. Department of the Interior, Bureau of Land Management, National Science and Technology Center; 2001. pp. 2-20

[26] Sánchez OR, SILVA TC. Zoneamento ambiental: uma estratégia de ordenamento da paisagem. Cad. Geociências, Rio de Janeiro. 1995;**14**: 47-53. (in Portuguese)

[27] Verma RR, Manjunath BL, Singh NP, Kumar A, Asolkar T, Chavan V, et al. Soil mapping and delineation of management zones in the Western Ghats of coastal India. Land Degradation and Development. 2018; **29**(12):4313-4322

[28] Reyes J, Wendroth O, Matocha C, Zhu JF. Delineating site-specific management zones and evaluating soil water temporal dynamics in a Farmer's field in Kentucky. Vadose Zone Journal. 2019;**18**(1):1-19

[29] He YB, DeSutter T, Norland J, Chatterjee A, Casey F, Clay D. The measurement, prediction, and development of soil management zones in low-relief sodic soils. Precision Agriculture. 2018;**19**(5):858-875

[30] Companhia de Desenvolvimento dos Vales do São Francisco - Codevasf. Projeto Salitre: levantamento detalhado de solo e classificação de terras para irrigação. Codevasf; 1989. (Projetos técnicos, 15). (in Portuguese)

[31] Veloso HP, Filho ALRR, Lima JCA. Classificação da vegetação brasileira,

adaptada a um sistema universal. Rio de Janeiro: IBGE; 1991. (in Portuguese)

[32] Xue Y, Sellers PJ, Kinter JL, Shukla J. A simplified biosphere model for global climate studies. Journal of Climate. 1991;**4**:345-364

[33] Souza DJ, Kosin M, Melo RC, Santos RA, Teixeira LR, Sampaio AR, et al. Mapa geológico do Estado da Bahia—escala 1:1.000.000. Salvador: CPRM; 2003. Versão 1.1. Programas Carta Geológica do Brasil a milionésimo e Levantamentos geológicos básicos do Brasil (in Portuguese)

[34] Vieira RMSP, Cunha APMA, Alvalá RCS, Carvalho VC, Ferraz Neto S, Sestini MF. Land use and land cover map of a semiarid region of Brazil for meteorological and climatic models. Revista Brasileira de Meteorologia. 2013; **28**:129-138

[35] IUSS Working Group WRB. World Reference Base for Soil Resources 2014. World Soil Resources Reports 106. Rome: Food and Agriculture Organization of the United Nations; 2014

[36] EMBRAPA. Centro Nacional de Pesquisa de Solos. Manual de métodos de análise de solo/Centro Nacional de Pesquisa de Solos. 2nd ed. rev. atual.—Rio de Janeiro; 1997. 212 p.: il. (EMBRAPA-CNPS. Documentos; 1) (in Portuguese)

[37] EMBRAPA. Centro Nacional de Pesquisa de Solos. Procedimentos normativos de levantamentos pedológicos. Brasília: EMBRAPA-SPI; 1995. 101p. (in Portuguese)

[38] EMBRAPA. Serviço Nacional de Levantamento e Conservação de Solos. Critérios para distinção de classes de solos e de fases de unidades de mapeamento: normas em uso pelo SNLCS. Rio de Janeiro: EMBRAPA-SNLCS; 1988. (in Portuguese)

[39] Liao K, Xu S, Wu J, Zhu Q. Spatial estimation of surface soil texture using remote sensing data. Soil Science and Plant Nutrition. 2013;**59**:488-500

[40] Bishop TFA, McBratney AB. A comparison of prediction methods for the creation of field-extent soil property maps. Geoderma. 2001;**103**:149-160

[41] Wu C, Wu J, Luo Y, Zhang L, De Gloria SD. Spatial prediction of soil organic matter content using cokriging with remotely sensed data. Soil Science Society of America Journal. 2009;**73**: 1202-1208

[42] Chagas CS, Carvalho Junior W, Bhering SB, Calderano Filho B. Spatial prediction of soil surface texture in a semiarid region using random forest and multiple linear regressions. Catena. 2016;**139**:232-240

[43] Saunders AM, Boettinger JL. Incorporating classification trees into a pedogenic understanding raster classification methodology, Green River Basin, Wyoming, USA. In: Lagacherie P, McBratney AB, Voltz M, editors. Digital Soil Mapping: An Introductory Perspective. Developments in Soil Science. Vol. 31. Amsterdam: Elsevier; 2007. pp. 389-399

[44] Carvalho Junior W, Lagacherie P, Chagas CS, Calderano Filho B, Bhering SB. A regional-scale assessment of digital mapping of soil attributes in a tropical hillslope environment. Geoderma. 2014;**232**:479-486

[45] Xiao J, Shen Y, Tateishi R, Bayaer W. Development of topsoil grain size index for monitoring desertification in arid land using remote sensing. International Journal of Remote Sensing. 2006;**12**:2411-2422

[46] R Development Core Team, R A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

2013. Available from: http://www.R-project.org/isbn 3-900051-07-0 [Accessed: May 08, 2015]

[47] Ciampalini R, Lagacherie P, Hamrouni H. Documenting Globa lSoilMap.net grid cells from legacy measured soil profile and global available covariates in Northern Tunisia. In: Minasny B, Malone BP, McBratney AB, editors. Digital Soil Assessments and Beyond. London: CRC Press/Balkema; 2012. pp. 439-444

[48] Breiman L. Random Forests. Technical Report for Version 3. 2001. Available from: http//oz.berkeley.edu/users/breiman/randomforest2001.pdf [Accessed: December 28, 2014]

[49] Liaw A, Wiener M. Classification and regression by random forest. R News. 2002;**2**(3):18-22

[50] Grimm R, Behrens T, Märker M, Elsenbeer H. Soil organic carbon concentrations and stocks on Barro Colorado Island—Digital soil mapping using Random forests analysis. Geoderma. 2008;**146**:102-113

[51] Pang S, Li TX, Zhang XF, Wang YD, Yu HY. Spatial variability of cropland leadand its influencing factors: A case study in Shuangliu county, Sichuan province,China. Geoderma. 2011;**162**: 223-230

[52] Holmes KW, Chadwick OA, Kyriakidis PC. Error in a USGS 30-meter digital elevation model and its impact on terrain modeling. Journal of Hydrology. 2000;**233**:154-173

[53] ArcGIS 10 Help. Available from: http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-analyst-toolbox/iso-cluster-unsupervised-classification.htm [Accessed: June 29, 2019]

[54] Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: Erlbaum; 1988

[55] Souza Junior JG, Demattê JA, Araújo SR. Modelos espectrais terrestres e orbitais na determinação de teores de atributos dos solos: potencial e custos. Bragantia. 2011;**70**:610-621. (in Portuguese)

[56] Ahmed Z, Iqbal J. Evaluation of Landsat TM5 multispectral data for automated mapping of surface soil texture and organic matter in GIS. European Journal of Remote Sensing. 2014;**47**:557-573

[57] Demattê JAM, Fiorio PR, Ben-Dor E. Estimation of soil properties by orbital and laboratory reflectance means and its relation with soil classification. Open Remote Sensing Journal. 2009;**2**:12-23

[58] Ließ M, Glaser B, Huwe B. Uncertainty in the spatial prediction of soil texture: Comparison of regression tree and random forest models. Geoderma. 2012;**170**:70-79

[59] Akpa SIC, Odeh IOA, Bishop TFA, Hartemink AE. Digital mapping of soil particle-size fractions for Nigeria. Soil Science Society of America Journal. 2014;**78**:1953-1966

[60] Vaysse K, Lagacherie P. Evaluating digital soil mapping approaches for mapping GlobalSoilMap soil properties from legacy data in Languedoc-Roussillon (France). Geoderma Regional. 2015;**4**:20-30

[61] Lagacherie P, Sneep AR, Gomez C, Bacha S, Coulouma G, Hamrouni MH, et al. Combining VisNIR hyperspectral imagery and legacy measured soil profiles to map subsurface soil properties in a Mediterranean area (cap-bon, Tunisia). Geoderma. 2013;**209-210**: 168-176

[62] Smith MP, Zhu AX, Burt JE, Stiles C. The effects of DEM resolution and neighborhood size on digital soil survey. Geoderma. 2006;**137**:58-69

[63] Ruiz-Navarro A, Barberá GG, García-Haro J, Albaladejo J. Effect of the spatial resolution on landscape control of soil fertility in a semiarid area. Journal of Soils and Sediments. 2012;**12**:471-485

[64] Hengl T, Heuvelink GBM, Kempen B, Leenaars JGB, Walsh MG, Shepherd KD. Mapping soil properties of Africa at 250 m resolution: Random forests significantly improve current predictions. PLoS One. 2015;**10**(6):1-26, e0125814. DOI: 10.1371/journal. pone.0125814

[65] Liao KH, Xu SH, Wu JC, Ji SH, Qing LIN. Cokriging of soil cation exchange capacity using the first principal component derived from soil physico-chemical properties. Agricultural Sciences in China. 2011;**10**: 1246-1253

[66] Demattê JAM, Galdos MV, Guimarães RV, Genú AM, Nanni MR, Zullo J Jr. Quantification of tropical soil attributes from ETM+/LANDSAT-7 data. International Journal of Remote Sensing. 2007;**28**:3813-3829