**Brazil - 2021**

# II Pedometrics Brazil

# ANNALS

### ON-LINE 24th to 27th November 2021
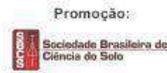
The event **II Pedometrics Brazil** was organized by Embrapa Soils and Federal Rural University of Rio de Janeiro (UFRRJ) and promoted by the Brazilian Soil Science Society (SBCS). The conference was held online from November 24th to 27th, in the year 2021, and was sponsored by Coordination for the Improvement of Higher Education Personnel (CAPES) and Carlos Chagas Filho Foundation for Research Support of the State of Rio de Janeiro (FAPERJ).

## **Organizing Committee**:

Waldir de Carvalho Junior – Embrapa Solos (Coordinator)

Helena Saraiva Koenow Pinheiro – UFRRJ (Sub-Coordinator )

Ricardo Simão Diniz Dalmolin – UFSM (Sub-Coordinator )

Marcos Bacis Ceddia – UFRRJ

Silvio Barge Bhering – Embrapa Solos

Alexandre ten Caten – UFSC

José Marques Júnior – Unesp/FCAV

Elpídio Inácio Fernandes Filho – UFV

Eduardo Guimarães Couto – UFMT

Gustavo Souza Valladares – UFPI

Fabrício da Silva Terra – UFVJM

Marcio Rocha Francelino – UFV

The II Pedometrics Brazil was an event promoted by the Pedometrics Commission of Division I – Soil in Space and Time, of the Brazilian Soil Science Society.

Pedometrics is one of the disciplines in soil science with the greatest scientific and technological development in the recent years, mainly due to the advances in computer systems and data availability. The development of technologies related to the use of computing and algorithms allows a better understanding of soil, as a phenomenon that varies at different scales in space and time.

Research in Pedometrics is highly relevant to society, since it allows for the soil information to be more accessible and useful for farmers and public administrators. The II Pedometrics Brazil took place at a significant moment for the country, with the beginning of the implementation of the National Soil Survey Program (PronaSolos). The main objective of PronaSolos is to map soils of the Brazilian territory, at scales compatible with soil governance and in much more detail than the existing soil surveys.

The II Pedometrics Brazil was organized by Embrapa Soils and UFRRJ and held on-line from 24th to 27th of November, 2021. The central theme was "Pedometrics: Innovations in the Tropics" and it featured four sessions:

01: Pedometrics: Innovation in Tropics

02: Legacy Data: How turn it useful?

03: Advances in Soil Sensing

04: Pedometric guidelines to systematic soil surveys

The goals of the II Pedometrics Brazil were to discuss pedometric methodologies and processes that may be implemented in the PronaSolos program, and to promote the exchange of experiences and knowledge between Brazilians and foreign researchers, and graduate students, working in Agronomy and related Soil Science fields.

# Modelling and mapping superficial soil texture through machine learning and limited legacy data

PEREIRA, Luís Flávio[1]; SANTOS, Cássio Marques Moquedace[2]; ROSOLEM, Gabriel Phelipe Nascimento[3]; DE SOUZA, Maria da Conceição[4]; FRANCELINO, Márcio Rocha[5]; FERNANDES-FILHO, Elpídio Inácio[6]

[1]Department of Soil, Federal University of Viçosa, luis.flavio@ufv.br; [2]Department of Soil, Federal University of Viçosa, cassio.moquedace@ufv.br; [3]Department of Rural Engineering, Federal University of Santa Catarina, gabriel.rosolem@posgrad.ufsc.br; [4]Department of Biology, Federal University of Ceará, conceicaosousa150410@gmail.com; [5]Department of Soil, Federal University of Viçosa, marcio.francelino@ufv.br; [6]Department of Soil, Federal University of Viçosa, elpidio@ufv.br

## Thematic Session: Pedometrics: Innovations in Tropics

**Abstract**

Soil texture influences soil's physical, chemical, and biological processes, being crucial for understanding soil functions and improving decision making. This paper aimed to map the surface soil texture in the *Minas Gerais* state, Brazil, through machine learning and legacy data. We used 668 georeferenced samples containing coarse sand, fine sand, silt, and clay contents at 0-20 cm depth and 109 covariates based on the SCORPAN model. Predictor's selection and predictive modeling of each particle size for five different models were performed 100 times, using 75% of samples for training and 25% for testing. The Random Forest was the best model for all-grain size distribution. The most important covariates were related to the parent material, soil properties, and climate. We conclude that machine learning techniques can produce fair consistent maps for superficial texture and the associated models' uncertainties, even using limited legacy data.

Keywords: soil physics; clay; silt; sand.

## Introduction

The texture or particle size distribution of soils has influence on many physical, chemical and biological processes among several compartments of the Earth's critical zone (Palm et al., 2007). Thus, surface texture maps can provide fundamental data for understanding soil functions and their ecosystem services. They also provide important inputs for environmental models, land use planning and management, and decision-making. However, spatial information in detailed scales and known uncertainties on contents of each soil's particle size still lack for many regions in the world. Legacy data might be useful for producing and improving low-cost maps of soil attributes, but standardized and large datasets are scarce. In this sense, the objective was to map the surface soil texture in the *Minas Gerais* state through machine learning techniques using limited legacy data.

## Methodology

The *Minas Gerais* state (586.528 km$^2$) is located in southeastern Brazil and present high geodiversity and climate variability (UFV 2010). To model and map surface texture for the whole area, we used 668 georeferenced samples containing standardized measures of coarse sand, fine sand, silt, and clay contents at 0-20 cm

depth (Souza et al. 2015), and 109 covariates based on the SCORPAN model. Covariates included climate data from WorldClim (Hijmans et al., 2005), parent material (CPRM 2004), volumetric water content in soil (Copernicus Climate Change Service, 2019), soil classes (UFV 2010), and morphometric data from digital elevation model processing (NASA JPL, 2020). All covariates were harmonized to 500 m of spatial resolution (O'Brien, 2020).

We applied sequential predictor selection, eliminating covariates from pairs with more than 95% spearman correlation and, subsequentially, the Recursive Feature Elimination was applied (Kuhn and Johnson, 2013). The predictive modeling for each particle size (coarse sand, fine sand, silt and clay) was done using five different models (Random Forest - RF, Cubist, Multivariate Adaptive Regression Spline – MARS, Support Vector Machine – SVM Radial, and Stochastic Gradient Boosting – GBM). These steps were performed 100 times, using 75% of samples for training and cross-validation and 25% for testing. The metrics adopted for validation were: coefficient of determination ($R^2$), mean absolute error (MAE) and root-mean-square error (RMSE). To assess models' consistency and uncertainties, we used the coefficient of variation (CV) of contents in the 100 predictions, the sum of modeled contents, and the calculated silt by difference of fractions (100 – (clay + fine sand + coarse sand)). All modellings and processing were carried out using the R software (R Core Team, 2021).

## Results and discussion

Considering the limited legacy data approach, our results showed a fair predictive performance of machine learning models, without overfitting (Figure 1).
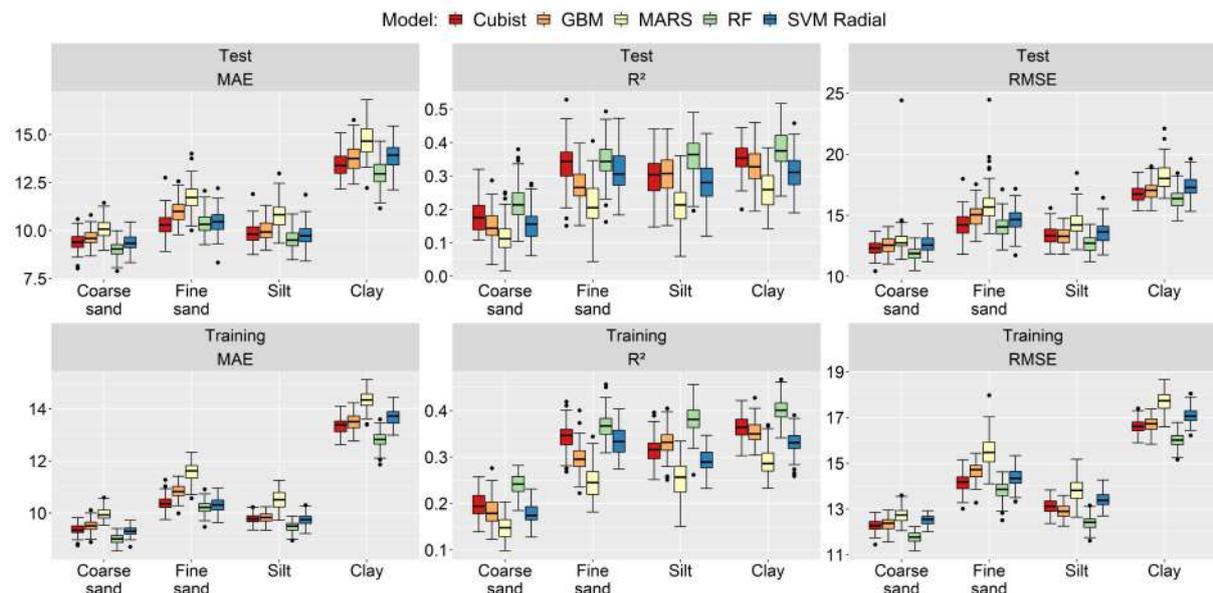


Figure 1. Models' performance of the 100 runs for each particle size in the test and training steps.

The Random Forest had the best performance for all soil particle sizes and was used for predicting the maps. A range of 10-15 predictors in the selection process was needed to stabilize the performance of the models. The most important predictors and its ranking variate among particle sizes, but were related to climate, parent material, relief, and soil properties, such as: air temperature, precipitation, rock texture, soil moisture and soil class. Relief properties (convexity, ruggedness and others) were especially important for silt contents prediction. Sand fractions presented the worst $R^2$ values, which indicates that merging coarse and fine sand in a single sand class might improve models' prediction. Both RMSE and MAE presented relatively low values for all particle size, which means that the machine learning models were accurate to predict clay, sand and silt contents (Figure 1).

These results are also confirmed by a visual-spatial assessment (Figure 2). In general, all particle sizes presented low CV values and low spatial and quantitative inconsistences of the values of sum-of-contents and silt-by-difference metrics. Field knowledge and cartographic comparisons also confirm a good spatial and quantitative consistency of maps, which could be improved by collecting more soil texture samples on low sample density areas, such as *Triângulo Mineiro* and *Central Mineira* mesoregions. Soils are mostly clayey at the surface, and fine sand particles are dominant in low-clay areas, which is mainly related to parent materials and weathering conditions.
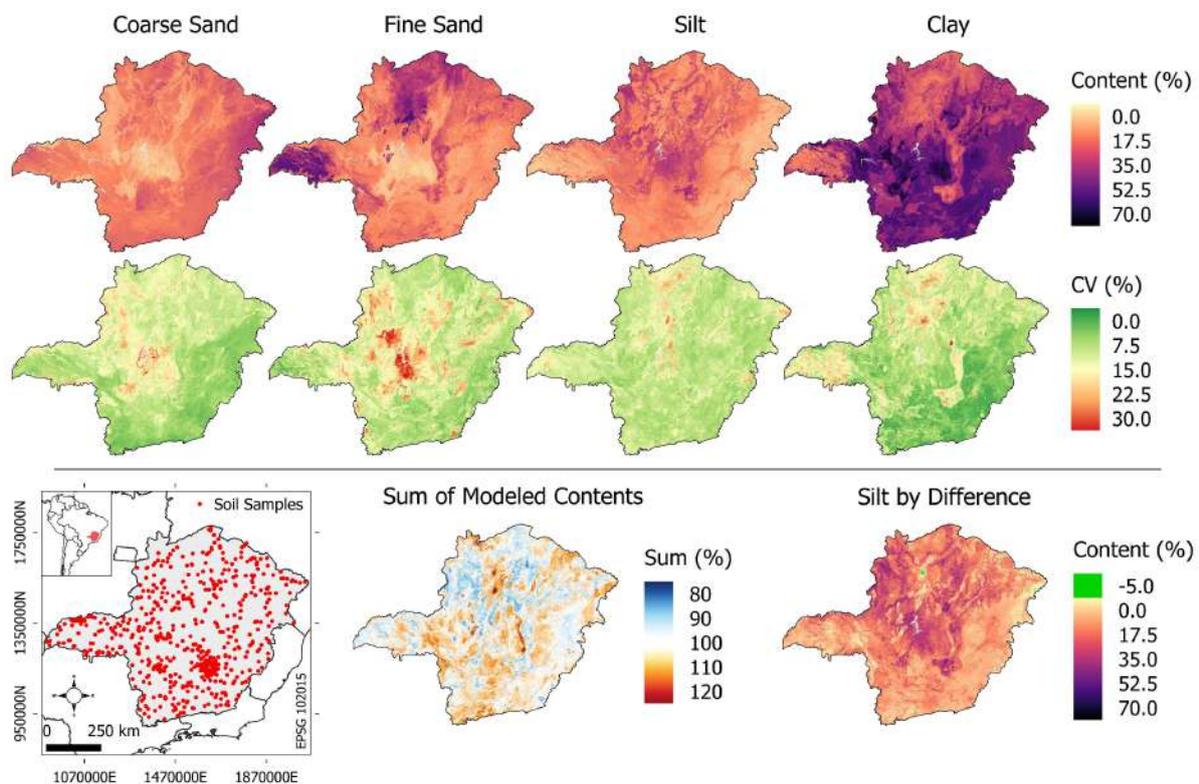


Figure 2. Modeled maps of mean content from the 100 runs and spatial uncertainties for each particle size using the Random Forest model.

## Conclusions

We conclude that machine learning techniques can produce fair consistent maps for superficial texture and the associated models' uncertainties, even using limited legacy data. Random Forest was the best model and the most important predictors were distributed across the SCORPAN factors.

## Acknowledgements

## References

COPERNICUS CLIMATE CHANGE SERVICE. **ERA5-Land hourly data from 2001 to present**. 2019.

CPRM. **Carta Geológica do Brasil ao Milionésimo: sistema de informações geográficas–SIG** [Geological Map of Brazil 1: 1.000. 000 scale: Geographic Information System–GIS]. 2004.

DE SOUZA, J.J.L.L. et al. Geochemistry and spatial variability of metal(loid) concentrations in soils of the state of Minas Gerais, Brazil. **Science of the Total Environment** 505, 338–349, 2015.

Hijmans, R.J. et al. Very high resolution interpolated climate surfaces for global land areas. **International Journal of Climatology** 25, 1965–1978. 2005.

KUHN, M.; JOHNSON, K., 2013. **Applied Predictive Modeling**. Springer New York, New York, NY.

NASA JPL. **NASADEM Merged DEM Global 1 arc second V001**. 2020.

O'BRIEN, J. **gdalUtilities: Wrappers for "GDAL" Utilities Executables**. 2020

PALM, C. et al. Soils: A Contemporary Perspective. **Annu. Rev. Environ. Resour**. 32, 99–129, 2007.

R CORE TEAM, 2021. **R: A Language and Environment for Statistical Computing**. R Foundation for Statistical Computing.

UFV et al. **Mapa de solos do Estado de Minas Gerais**. Universidade Federal de Viçosa; Fundação Centro Tecnológico de Minas Gerais; Universidade Federal de Lavras; Fundação Estadual do Meio Ambiente, Belo Horizonte, 2010.

# MULTIFRACTAL ANALYSIS OF BIOLOGICAL, PHYSICAL, AND CHEMICAL ATTRIBUTES OF SOIL

Silva, Raimunda Alves[1]; Siqueira, Glécio Machado[1]

[1] Universidade Federal do Maranhão, PPG BIONORTE, ray-234@hotmail.com; glecio.siqueira@ufma.br

**Thematic Session: Pedometrics: Innovations in Tropics**

**Abstract**
The objective of this study was to assess the multifractality of biological, physical, and chemical attributes of soil in the 0.0-0.05 m, 0.05-0.15 m and 0.15-0.3 m layers, in Parque Estadual do Mirador (PEM), Maranhão, Brazil. The sampled data comprised soil fauna, physical (sand, clay, silt, density, total porosity, macroporosity and microporosity), and chemical attributes of soil (organic carbon, pH, phosphorus, potassium, calcium, magnesium, sum of bases and capacity of cation exchange) in two savanna formations. The data were subjected to multifractal analysis. The predators group presented the highest degree of multifractality ($\Delta_{T1} = 0.64$ and $\Delta_{T2} = 0.33$). Silt and Ca were the attributes with the highest multifractality in the 0.0-0.05 m and 0.05-0.15 m layer. OC and Ca expressed the highest multifractality in the 0.15-0.3 m layer. The singularity spectra described that the data represent multifractal systems influenced by the parent material, soil type, and relief.
Keywords: invertebrates; soil fauna; physical properties; singularity spectra, multifractality.

## Introduction

The soil attributes tend expressing high natural variability in the landscape, resulting from soil formation factors. Thus, it is necessary to implement analysis tools that consider variability at different scales. Variability scales are characterized at different times in multifractal analysis (Halsey et al., 1986), allowing to describe the heterogeneity of systems (Vidal Vázquez et al., 2013). In line ith this, according to Banerjee et al. (2011), variables are highly heterogeneous and dependent on the scale magnitude. Given the above, this study has the following hypothesis: a) the biological, physical, and chemical attributes of soil have different scaling patterns and heterogeneity structure in landscape. Thus, the objective of this study was to assess the multifractality of biological, physical, and chemical attributes of soil in the 0.0-0.05 m, 0.05-0.15 m, and 0.15-0.3 m layers, in Parque Estadual do Mirador (PEM), Maranhão, Brazil.

## Methodology

The study was carried out at Parque Estadual do Mirador (PEM), in the State of Maranhão, Brazil (6° 42' 9.803" S and 44° 42' 54.936" W - Figure 1). In the experimental plots (Figure 1). Biological, physical, and chemical attributes of the soil were sampled from the experimental plots (Figure 1), in August 2017. T1 plot comprised a sparse Cerrado vegetation, with an average altitude of 308 m, and predominance of Oxisols, belonging to the geological formation of Sambaíba. T2 comprised a typical Cerrado vegetation with an average altitude of 432 m, and

II Pedometrics
Nov, 24th to 27th, 2021
Brazil
Apoio: FAPERJ  CAPES
Promoção: Sociedade Brasileira de Ciência do Solo
Organização: UFRRJ  Embrapa Solos

predominance of Entisols originated from the geological unit known as Corda (Figure 1).
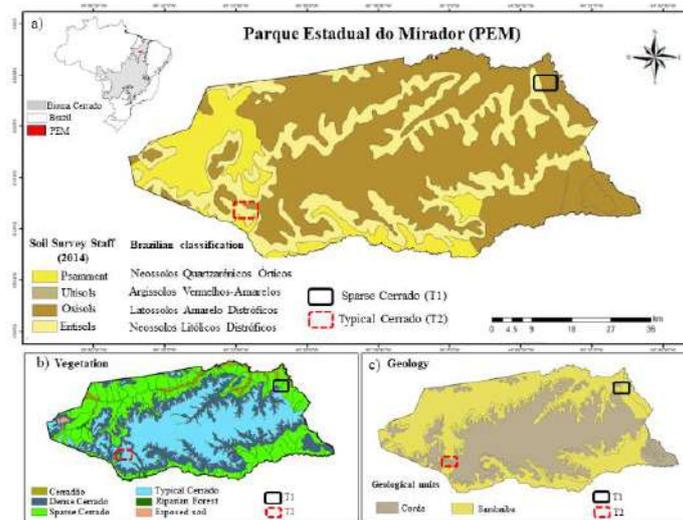


Figure 1. Map showing the location of Parque Estadual do Mirador (PEM), Maranhão (Brazil).

A transect with 128 sampling points was installed in each plot. The sampling points were installed with a distance of 3 m between the points, resulting in a total length of 381 m. The invertebrate fauna was sampled from pitfall traps. The physical and chemical attributes of the soil were collected from the 0.0-0.05 m, 0.05-0.15 m, and 0.15-0.3 m layers, from undisturbed soil samples. Soil physical attributes comprised: sand (g kg$^{-1}$), clay (g kg$^{-1}$), silt (g kg$^{-1}$), density bulk (BD - Mg m$^{-3}$), macroporosity (m$^3$ m$^{-3}$), microporosity (m$^3$ m$^{-3}$), total porosity (TP - m$^3$ m$^{-3}$), according to Camargo et al. (2009). Chemical attributes comprised: organic carbon (OC, g dm$^{-3}$), pH (in CaCl$_2$ solution), phosphorus (P, mg dm$^{-3}$), potassium (K, mmolc dm$^{-3}$), calcium (Ca, mmolc dm$^{-3}$), magnesium (Mg, mmolc dm$^{-3}$), sum of bases (SB), and cation exchange capacity (CEC, mmolc dm$^{-3}$), according to Raij et al. (2001). Multifractal analysis was carried ou based on the moment method (Halsey et al., 1986), and on the direct method (Chhabra and Jensen, 1989), considering the total length of the transect (δ = 381 m) divided into successive $2^k$ segments (k = 1, 2, 3...). The analysis involved the obtainment of the partition function (Vidal Vázquez et al., 2013), generalized dimension values (D - Hentschel and Procaccia, 1983), and singularity spectrum of the function of $f(α)$ versus $α$, according to Chhabra and Jensen (1989).

**Results and discussion**

The singularity spectra [$f(α)$ versus $α$] corresponding to the biological variables expressed multifractal behavior, with asymmetry for the spectrum branches (left and right) for the different variables in T1 and T2 (Figure 2a and 2b). The singularity spectra [$f(α)$ versus $α$] for physical and chemical soil attributes were asymmetric, with different degrees of multifractality (Figure 3). In general, the physical attributes in T1 and T2 expressed high measurement values, where silt showed the greatest differences in multifractality in the layers and between the plots (Figure 3b, 3e and 3f). The singularity spectrum [$f(α)$ versus $α$] for the chemical attributes showed

different degrees of multifractality in the layers and between the plots, with asymmetry of the branches to the left, indicating that there where high measurement values in the variable scales (Figure 3). CO content varied in the variability scales for all layers and plots. Ca (Figure 3m), and Mg (Figure 3l and 3m) showed differences in the asymmetry of branches at T2 in the 0.05-0.15 m and 0.15-0.30 m layers, with asymmetry of the branches to the right, indicating occurrence of low measurement values.

Soil biological indicators in T1          Soil biological indicators in T2
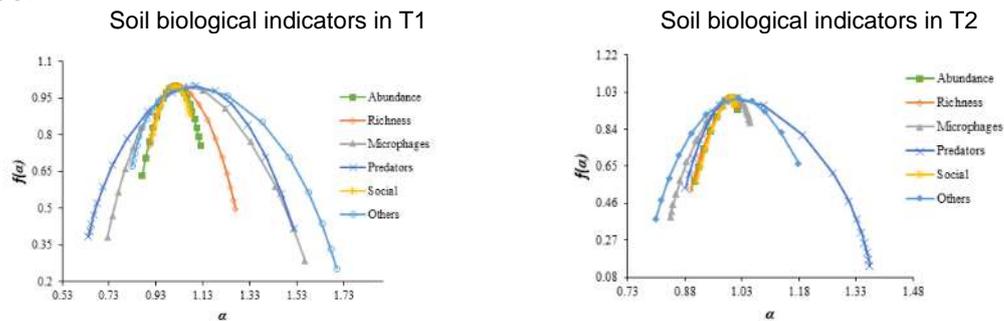


Figure 2. Spectrum of singularity for biological indicators of the soil in T1 (Sparse Cerrado) and in T2 (typical Cerrado) at Parque Estadual do Mirador, Maranhão (Brazil).

Soil physical attributes

a) T1 (0.0-0.05 m)          b) T1 (0.05-0.15 m)          c) T1 (0.15-0.3 m)



d) T2 (0.0-0.05 m)          e) T2 (0.05-0.15 m)          f) T2 (0.15-0.3 m)



Soil chemical attributes

g) T1 (0.0-0.05 m)          h) T1 (0.05-0.15 m)          i) T1 (0.15-0.3 m)



j) T2 (0.0-0.05 m)          l) T2 (0.05-0.15 m)          m) T2 (0.15-0.3 m)
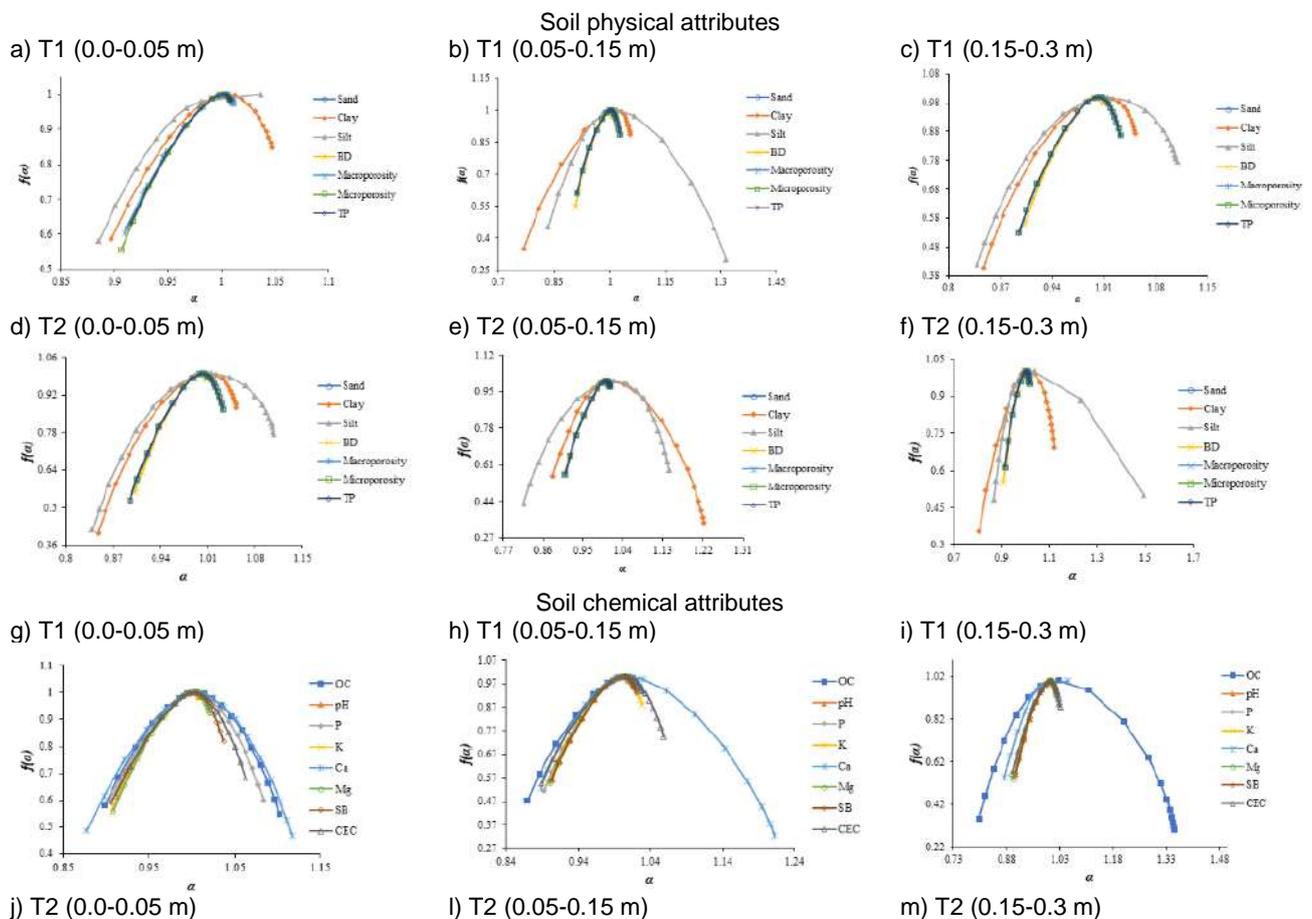
Figure 3. Spectrum of singularity for physical and chemical attributes of the soil. Physical attributes in T1: a) 0.0-0.05 m; b) 0.05-0.15 m; c) 0.15-0.3 m; physical attributes in T2: d) 0-0.05 m; e) 0.05-0.15 m; f) 0.15-0.3 m; chemical attributes at T1: g) 0.0-0.05 m; h) 0.05-0.15 m; i) 0.15-0.3 m; chemical attributes in T2: j) 0.0-0.05 m; l) 0.05-0.15 m; m) 0.15-0.3 m.

## Conclusions

The soil showed different degrees of multifractality. The multifractality of soil fauna was influenced by the vegetation gradient, and the multifractality of physical and chemical soil attributes was associated to the soil parent material, soil type, and to landscape relief. The multifractal analysis allowed to elucidate the dynamics of ecological relationships for soil fauna organisms, reflecting the tropic chain structures.
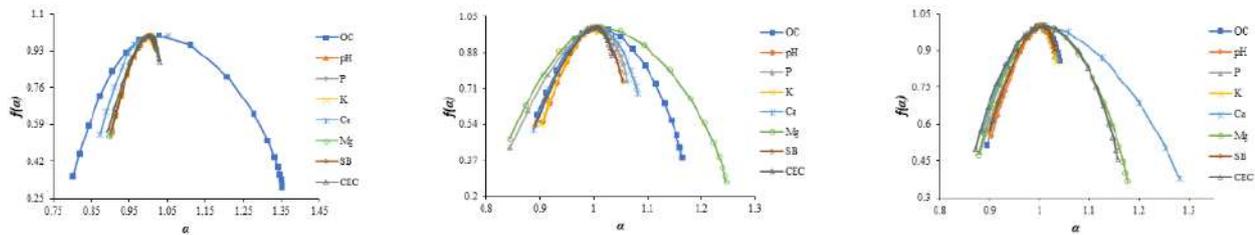
## References

Banerjee, S. et al. Spatial relationships between leaf area index and topographic factors in a semiarid grassland Joint multifractal analysis. **Australian Journal of Crop Science,** v.6, p.756-763, 2011.

Camargo, A. O. et al. **Métodos de análise química e física de solos do Instituto Agronômico de Campinas**. Campinas: IAC. 2009.

Chhabra, A.; Jensen, R. V. Direct determination of the $f(\alpha)$ singularity spectrum. **Physical Review Letters**, v.62, p.1327-1330, 1989.

Halsey, T. C. et al. Fractal measures and their singularities: the characterization of strange sets. **Physical Review A**, v.33, p.1141-115, 1986.

Hentschel, J. G. E.; Procaccia, I. The infinite number of generalized dimensions on fractals and strange attractors. **Physica D**, v.8, p.435-444, 1983.

Raij, B. et al. **Análise química para avaliação da fertilidade de solos tropicais. Campinas: Instituto Agronômico**. 2001. 285p.

Vidal Vázquez, E. et al. Multifractal analysis of soil properties along two perpendicular transects. **Vadose Zone Journal,** v.12, p.1-13, 2013.

# Spatial prediction of topsoil texture fractions using bare surfaces reflectance and Random Forest in Google Earth Engine

PELEGRINO, Marcelo Henrique Procópio[1]; MENEZES, Michele Duarte de[1]; GUILHERME, Luis Roberto Guimarães[1]; LIMA, Geraldo Jânio Eugênio de Oliveira[2]; POPPIEL, Raul Roberto[3]; DEMATTÊ, José Alexandre Melo[3].

[1] Federal University of Lavras, marcelo.pelegrino@estudante.ufla.br; michele.menezes@ufla.br; guilherm@ufla.br; [2]Agricultural and Environmental Technology Center (CAMPO), geraldo.lima@campoanalises.com.br; [3]University of São Paulo, raulpoppiel@usp.br; jamdemat@usp.br;

## Thematic Session: PEDOMETRICS: INNOVATION IN TROPICS

**Abstract**

Soil texture is management-key for agricultural in Brazil. Reflectance images of exposed soils can be related to properties, being useful as an environmental covariate in predictive models. This work aimed to create a new soil environmental covariate by recovering the bare earth surface image in order to get a continuous surface for a 3,295,860 km$^2$ area in Cerrado Biome, and assess their predictive power on soil texture modeling for site-specific management scale. For that, it was selected related covariates to generate a predictive map, thus, overlay the predicted map to fill in the gaps, and evaluate the performance of the recovered image versus an averaged image in predictive models using RMSE, R$^2$ and RPIQ. The predicted images had satisfactory correlation (Pearson) with the originals, which presented average values of 0.53, and, clay, silt and sand were predicted with proper accuracy using only the recovered image.

**Keywords:** spatial big data; google earth engine; random forest; pedometry; remote sensing.

## Introduction

The Cerrado Biome comprises the main and most challenging agricultural area in Brazil, responsible for 70% of national agricultural production (Wickramasinghe et al., 2012) with high potential to assure food security (Cerri et al., 2018). Soil texture fractions are key attributes for agricultural recommendation and soil management, presenting little or no change over time. Currently, there are no digital maps of soil fractions which may assist site-specific management of precision agriculture in the Cerrado Biome, so an effort to gather a legacy soil data and produce remote sensing-based covariates can be useful to fill this gap of continuous spatial information (Poppiel et at., 2019). However, a computational task for big data analysis to generate information proper to site-specific management in tropical conditions is a challenge. Site-specific management, also called precision agriculture, consist of the management of agricultural crops at a spatial scale smaller than that of the whole field (Plant, 2001).

Demattê et al. (2020) proposed a remote sensing data mining method to retrieve bare pixels over time and obtain a final image called Synthetic Soil Image (SySI). This image contains median values of the reflectance of bare surfaces of the Earth, including bare soil and rock outcrop. The spectral patterns retrieved from agricultural bare soils in the SySI are related to the simultaneous interaction of several soil

attributes from soil surfaces exposed during land conversion and soil tillage practices. Soil spectral patterns can be directly related to soil minerology (Viscarra Rossel et al., 2016), clay (Lagacherie et al., 2013) and soil classes (Novais et al., 2021), among others.

Thus, we aimed to predict clay, silt and sand contents at 0-20 cm soil depth using remote sensing data and Random Forest algorithms in R environment and Google Earth Engine cloud-based platform.

**Methodology**

The soil dataset comprises 32,239 samples with information of clay, sand and silt content from the topsoil layer (0-20 cm) covering 3,295,860 $Km^2$ of study area. This dataset was obtained from: a) the collaboration of several Brazilian soil laboratory analyses being taken from farms with higher technological level (precision agriculture as a management tool), comprising 78% of the database; b) the Brazilian Soil Spectral Library (BSSL) (Demattê et al, 2019); c) the Free Brazilian Repository for Open Soil Data (FEBR) (Samuel-Rosa et al., 2017); d) Platform WebGIS IDE-SISEMA. Once  the soil datasets were merged, the duplicates and typos were checked, outliers were removed according to Tukey's rule, and thus, dataset was finally divided into training (70%) and testing (30%) sets.

Covariates obtention: A) A 30 m spatial resolution Synthetic Soil Image (SySI) was generated using the method described in Demattê et al. (2020). B) A 30 m spatial resolution elevation data from the Advanced Land Observing Satellite (ALOS) was used to calculate thirteen terrain attributes using the package Terrain Analysis in GEE (Safanelli et al., 2020). C) Nineteen bioclimate data were obtained from the WorldClim V1 Bioclim database at 1 km resolution (Hijmans et al., 2005). D) Three soil-plant variables were from Penman-Monteith-Leuning Evapotranspiration V2 (PML_V2), transpiration from vegetation, direct evaporation from the soil and vaporization of intercepted rainfall from vegetation (Zhang et al., 2016). E) To represent a condition with the least disturbance of the landscape, where natural vegetation predominated, a Landsat mosaic obtained from 1984 to 1986 was used. F) Coarser spatial resolutions (climate and soil-plant data) were downscaled to 30 m pixel size.

Since SySI had gaps over areas with continuous vegetation cover within the period evaluated, which were filled (predicted) using covariates and random forest algorithm, as follow: A) Covariates were randomly sampled using two observations per $km^2$ in a 4 × 4 moving window. B) For each band, we selected the covariates by a 10-fold cross-validation Recursive Feature Elimination (RFE) algorithm, implemented on the Caret Package (Kuhn, 2017). In order to avoid wavy patterns, it was limited in two bioclimatic covariates for each prediction according to RFE criteria. C) Thereafter, a grid search to tune optimal Random Forest (RF) parameters *mTry*, *nTree, min node size (minNS), sample fraction* and *replacement* can be ran as folowing: (5%, 15%, 25%, 33% and 40% of the total features number)*(floor=number

of features/3) to *mTry*; 30, 50, 80 and 100 trees; 1, 3, 5 and 10 to *minNS*; *sample fraction* were 50, 63 and 80 %; with and without replacement, resulting in 120 combinations, using the ranger package version 0.11.1 (Wright and Ziegler, 2017) in the R software (R Core Team, 2019). Unlike usual, optimal models with the fewest number of trees possible was seek, aiming at lower loss of estimation stability (Probst et al, 2019), and decrease computational time, as this increases linearly with the number of trees (Boehmke and Greenwell, 2020).

Finally, the original SySI was overlaid on top of predicted bands and merged to obtain a continuous image, preserving the original values and incorporating the predicted ones on gaps. Thus, the spatially continuous image named here as filled-SySI, and an average image (original + predicted) named average-SySI were used as covariates soil texture predictive models.

The same process of tuning was applied for soil texture calibration considering two set of features: a) filled-SySI + soil-plant covariates and the absence of both; and b) averaged-SySI following the a) scheme. To assess the spatial prediction performance the metrics of RMSE, $R^2$ and ratio of the performance to interquartile distance (RPIQ) were calculated.

## Results and discussion

The predicted bare soil images (SySI) had satisfactory correlation (Pearson) with the originals, Blue 0.50; Green 0.52; LST 0.32; NIR 0.61; Red 0.58; SWIR1 0.57; SWIR2 0.58. Table 1 shows the performance metrics of the soil texture consisting of training and test dataset using filled-SySI and average-SySI with soil-plant covariates and the absence of them.

Table1. Performance metrics from training and testing sets of the default RF models and loss percentage of the optimal models for soil texture predictions

| | Training set | | | | | | Testing set | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | clay | silt | sand | clay | silt | sand | clay | silt | sand | clay | silt | sand |
| | RMSE (g kg⁻¹) | | | $R^2$ | | | RMSE (g kg⁻¹) | | | $R^2$ | | |
| **A** | 129.76 | 80.22 | 171.45 | 0.63 | 0.39 | 0.64 | 130.81 | 79.74 | 171.32 | 0.63 | 0.40 | 0.64 |
| **F** | 125.66 | 78.85 | 165.25 | 0.66 | 0.41 | 0.66 | 126.92 | 78.64 | 165.80 | 0.65 | 0.42 | 0.66 |
| **A*** | 122.45 | 77.34 | 157.88 | 0.67 | 0.41 | 0.64 | 123.2 | 77.8 | 163.33 | 0.67 | 0.43 | 0.67 |
| **F*** | 119.92 | 76.65 | 157.88 | 0.68 | 0.44 | 0.69 | 120.93 | 76.44 | 158.74 | 0.68 | 0.45 | 0.69 |
| | RPIQ | | | | | | Percentage of loss of the tuned models | | | | | |
| **A** | 2.60 | 2.25 | 2.73 | | | | **A** | -0.95 | -0.88 | -0.83 | | |
| **F** | 2.67 | 2.22 | 2.82 | | | | **F** | -0.91 | -1.09 | -0.72 | | |
| **A*** | 2.75 | 2.25 | 2.86 | | | | **A*** | -0.76 | -0.80 | -0.86 | | |
| **F*** | 2.81 | 2.29 | 2.95 | | | | **F*** | -0.78 | -0.77 | -0.88 | | |

A: average-SySI set; F: filled-SYSI; A*: average-SySI + soil-plant covariates; F*: filled-SySI + soil-plant covariates; RMSE: Root Mean Square Error; RPIQ: ratio of the performance to interquartile distance; $R^2$: Coeffcient of determination.

With the exception of silt, all optimal models (100 trees) showed less than 1% of loss compared to the default model (500 trees). In a correlation analysis, the filled-SySI showed satisfactory negative correlations with clay (NIR -0.61; SWIR2 -0.59; SWIR1 -0.56 and Red -0.55) and silt (NIR -0.46; Red -0.43; SWIR -0.39 and SWIR1 -0.37) and positive with sand (NIR 0.63; SWIR2 0.58; Red 0.57; SWIR1 0.56). Such results corroborates with the statement regarding the product ability in soil modeling (Poppiel et al., 2021; Novais et al., 2021).

## Conclusions

The gaps filled bare soil reflectance images (SySI) showed high performance for topsoil clay, silt and sand content spatial predictions. The filled image was slightly superior. And there was little or no difference in the presence or absence of auxiliary variables. The RFE criteria and model adjustments allowed eliminating redundant or noisy variables, enabling the reduction of the number of trees without significant losses in the predictions.

## Acknowledgements

## References

Cerri, C. R. P. et al. Reducing Amazon Deforestation through Agricultural Intensification in the Cerrado for Advancing Food Security and Mitigating Climate Change. **Sustainability**, v.10, n. 4, p. 989, 2018.

Demattê, J. A. M. et al. Bare earth's Surface Spectra as a proxy for Soil Resource Monitoring. **Sci. Rep.** v.10, n.1, 2020.

Novais, J. J. Digital Soil Mapping Using Multispectral Modeling with Landsat Time Series Cloud Computing Based. **Remote Sens**., n.13, p1181, 2021.

Plant, R. E. Site-specific management: the application of information technology to crop production. **Computers and Electronics in Agriculture**, n. 30, p9–29, 2001.

Poppiel, R. R. et al. High resolution middle eastern soil attributes mapping via open data and cloud computing. **Geoderma**. 385, 114890, 2021.

Poppiel, R. R. et al. Mapping at 30 m Resolution of Soil Attributes at Multiple Depths in Midwest Brazil**. Remote Sens***. v.*11*, p. 2905, 2019.

Safanelli, J.L. et al. **Terrain Analysis in Google Earth Engine**: A Method Adapted for High Performance Global-Scale Analysis. ISPRS Int. J. Geo-Information, 2020.

# Application of electrical conductivity profiling for the characterization and textural discretization of a Technosol

BARBOSA, Alexandre Muselli[1]; GUIMARÃES, Camila Camolesi[2]

[1] Institute for Technological Research (IPT), muselli@ipt.br; [2] Institute for Technological Research (IPT), camilacg@ipt.br.

## Thematic Session: Pedometrics: Innovations in Tropics

## Abstract

Obtaining large amounts of field data in a representative way is a great challenge, especially in the soil profile. This study aimed to evaluate the application of electrical conductivity profiling as a field method for the vertical characterization of soils. A field test was conducted in a Technosol profile, using a MH6534 electrical conductivity profiling probe, soil sampling, and granulometric characterization and electrical conductivity laboratory tests. The results indicated that this field method was able to identify and discretize the soil layers, having high correlation with clay contents obtained in the laboratory. However, it was not able to identify the organic layer. The method proved to be a good alternative for the vertical characterization of the tested soil, but soil samplings and analysis are necessary for data interpretation and models calibration.

**Keywords:** penetrometer; contaminated soil; non-invasive method; field data; real time.

## Introduction

The soil matrix is usually significantly heterogeneous and anisotropic is critically affected by the vertical and horizontal variability of the media (KARDANPOUR; JACOBSEN; ESBENSEN, 2014). The acquisition of large amounts of high-density data, in a representative way and with low cost is a challenge, especially when considering the vertical characterization of soils. An ideal penetrating data acquisition method must present characteristics that favor its field application, especially in contaminated sites, with simple operation, acquisition velocity and low level of soil disturbance. The latter is the most critical, as soil characteristics should not suffer great alterations, in order to avoid the migration of potential contaminations to deeper positions in the soil and/or groundwater. This study aims to evaluate the applicability of electrical conductivity profiling as an indirect, non-invasive characterization method of stratigraphic variation.

## Methodology

For the electrical conductivity profiling test, the Geoprobe® MH6534 probe was used, with electrical conductivity sensor in a horizontally aligned dipole-type arrangement, data acquisition in every 1.5 cm, and nails introduced by percussive system to a depth of 9 m. For the laboratory tests, the soil was collected by the direct push method (ASTM, 2014), using a pneumatic geohammer with 1.4 m jacketed rods, up

to a depth of 7.5 m. The samples were sent to the laboratory of the Investigations, Risks and Environmental Management Section of the Institute for Technological Research. The soil profile description was performed according to the Manual of Description and Collection of Soil in the Field (SANTOS et al., 2013), and its classification was based on the World Reference Base for Soil Resources (IUSS, 2015). The samples selection occurred according to the textural variation, resulting in 17 samples for granulometric analysis and electrical conductivity tests (TEIXEIRA et al., 2017).

## Results and discussion

The soil in the study area was classified as an Epileptic Humic Technosol, being under great influence of anthropic activities and addition of materials in the first layers. The pedogenetic processes observed along the soil profile, especially under 1.5 m, are controlled by fluvial deposition materials from the former floodplain of the Pinheiros river, in the Sedimentar Basin of São Paulo, as well as by the depth and groundwater level. The sediments are covered by technogenic deposits superimposed on the original floodplain (CARVALHO, 2006), which corroborates the soil profile identified in the area, representing a typical profile of the Pinheiros river fluvial material influence zones.

The electrical conductivity values obtained in the field were extracted as a function of the depths of the granulometric characterization tests, and the relation between the clay content and the electrical conductivity results for the discrete samples was analyzed (Figure 1). The coefficient of determination obtained was $R^2 = 0,9164$, a satisfactory result, considering that the sensor determination is an indirect process, which is susceptible to medium heterogeneity as well as to potential contact problems and/or mechanical and electrical failures. The same result was observed along the soil profile: the electrical responses only corresponded to the texture variations where clay was identified, not presenting any relation with other soil fractions.

The dipole-type array used in this system was able to identify clay lenses (Layer 07 – 5.87 m to 6.0 m and Layer 10 – 7.00 m to 7.05 m), with thickness ranging from 0.05 m to 0.13 m, identified during the visual-tactile soil description process. This result shows the sensitivity of the electrical conductivity sensors and the array system used for stratigraphic data acquisition of thin layers. The discretization of these layers corroborates the findings of Christy et al. (1994), who, when testing the responsiveness, repeatability and sensitivity of arrays, observed that dipole-type arrays are more sensitive to small variations in the medium. The method is also able to identify these variations, even though it is an array model more susceptible to lack of contact with the sides of the perforations, as it has only two electrodes horizontally arranged.
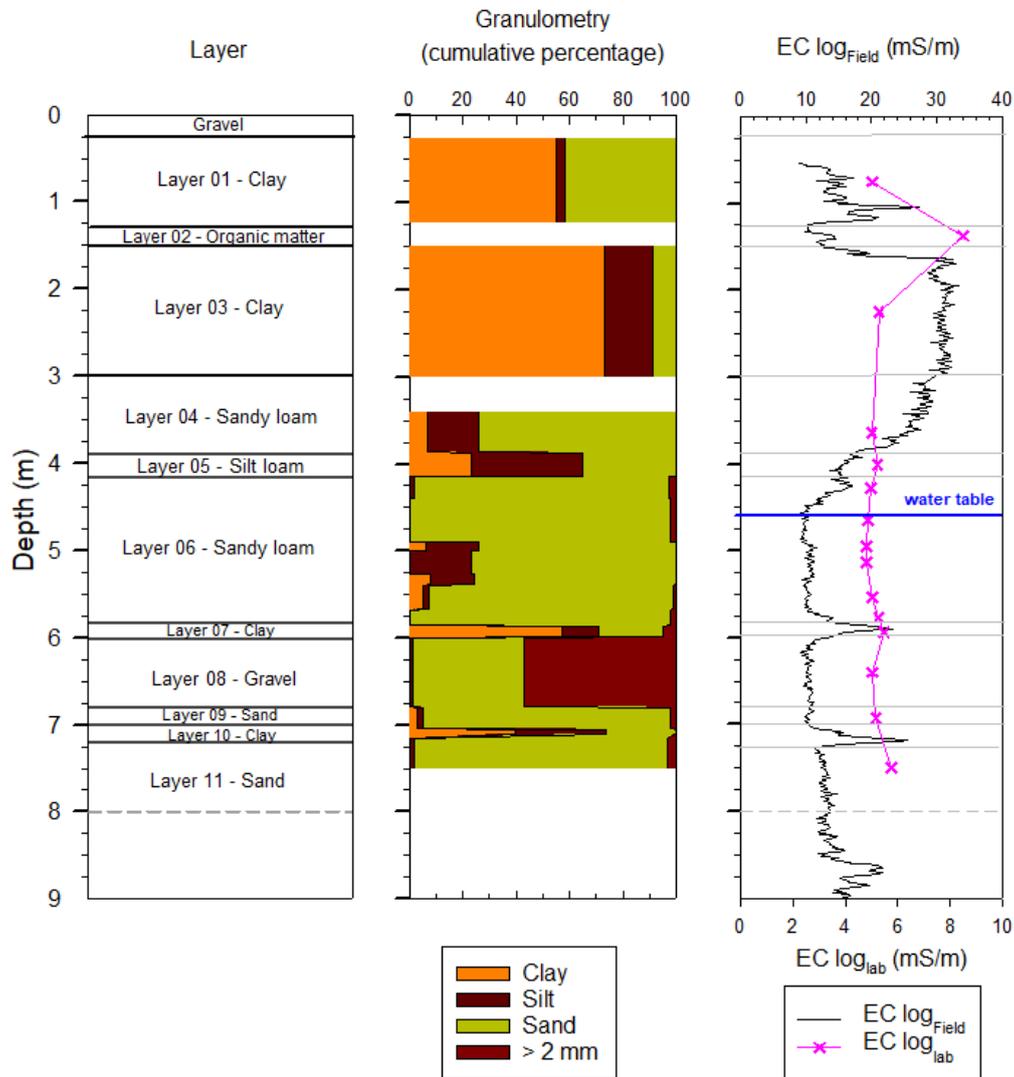
**Figure 1**. Tactile-visual textural division of soil layers and the results of granulometric and electrical conductivity characterization obtained in the field and in the laboratory

The results of electrical conductivity obtained in the laboratory showed little variation along the soil profile, with a value of 8.5 mS/m in the organic layer (Layer 02), an inverse behavior to that observed in the field test. This is due to the differences between the methods: in the laboratory, the electrical conductivity in the solid-liquid mixture is evaluated via exchange of cations associated with clay minerals and organic matter, whereas in the field method, the electrical conductivity is evaluated by a sensor in a solid medium, with direct contact with soil particles and its moisture. Thus, in the laboratory, the organic layers present a more conductive characteristic, and in the field test, the organic layers have a resistive characteristic, similar to sandy layers. In an indirect system for the determination of textural features, this behavior might lead to interpretation errors, thence soil samplings are necessary for the elaboration of calibration models.

## Conclusions

The application of the electrical conductivity profiling test with horizontal dipole-type sensor array for stratigraphic characterization is feasible, enabling the discretization of thin soil layers, even in a heterogeneous profile. The field electrical conductivity values are directly related to clay layers, and inversely related to organic layers, requiring the collection of samples for the calibration and interpretation of indirect data from the site. Further studies are necessary to evaluate the results of the electrical conductivity profiling method in other tropical soils.

## References

ASTM, American Society for Testing and Materials. **Standard Guide for Direct Push Soil Sampling for Environmental Site**. v. D6282/D628, p. 20, 2014. DOI: 10.1520/D6282_D6282M-14.

CARVALHO, D. L. R. **Indicadores geomorfológicos de mudanças ambientais no sistema fluvial do Alto Tietê (município de São Paulo): pesquisa documental.** 2006. 2v. Dissertação (Mestrado em Geografia Física) - Departamento de Geografia, FFLCH, USP, São Paulo.

CHRISTY, C. D.; CHRISTY, T. M.; WITTIG, V. **A Percussion Probing Tool for the Direct Sensing of Soil Conductivity**. Technical paper N 94-100, Direct Image. 1994. 15 p.

IUSS, International Union of Soil Sciences. **World reference base for soil resources 2014, update 2015**. Prepared by Schad P, van Huyssteen C, Micheli E. 192 pp. World Soil Resources Reports No. 106, FAO, Rome, 2015. 192 p.

KARDANPOUR, Z; JACOBSEN, O. S.; ESBENSEN, K. H. **Soil heterogeneity characterization using PCA (Xvariogram) - Multivariate analysis of spatial signatures for optimal sampling purposes**. Chemometrics and Intelligent Laboratory Systems, v. 136, 2014. p. 24–35

SANTOS, R.D. dos et al. **Manual de descrição e coleta de solo no campo**. 6. ed. Viçosa: Sociedade Brasileira de Ciência do Solo, 2013. 100 p.

TEIXEIRA, C. P. et al. **Manual de Métodos de Análises de Solo**. 3 ed. Brasilia: EMBRAPA, 2017. 577 p.

# Spatial distribution of soil organic carbon stocks in southwestern Amazônia predicted with machine learning

Santos, Cássio Marques Moquedace[1,2]; Gomes, Lucas Carvalho[1,3]; Lana, Marcos Guedes[1,4]; Veloso, Gustavo Vieira[1,5]; Fernandes-Filho, Elpídio Inácio[1,6]; Cardoso, Irene Maria[1,7]

[1]Federal University of Viçosa; [2]cassiomoquedace@gmail.com; [3]lucascarvalhogomes15@hotmail.com; [4]marcos.lana@ufv.br; [5]gustavo.v.veloso@gmail.com; [6]elpidio@ufv.br; [7]irene@ufv.br

**Thematic Session: (1) - Pedometrics: Innovation in tropics**

## Abstract

The soil high amounts of carbon and can act as a sink for greenhouse gases. Therefore, the aim of the study was to model and map SOC stocks in the state of Rondônia at 1 m depth. We used 2,914 soil profiles for modeling. We select the most important variables by correlation and by importance using recursive features elimination (RFE). We tested seven different machine learning algorithms. The random forest obtained the highest $R^2$ and the lowest MAE and RMSE and was selected for the prediction of SOC stocks. Soil classes and climate were the most important variables for SOC stocks. Rondônia stores 2,530.91 TgC. The greatest amounts of SOC are at low altitudes, where seasonal flooding occurs, and the genesis of the soil favors the accumulation of SOC. Lower SOC stocks is associated with Ferralsols under intense agricultural expansion and deforestation for livestock.

Keywords: random forest; Rondônia; digital soil mapping; climate changes; recursive feature elimination.

## Introduction

Soil accumulates two-thirds of all terrestrial carbon, and depending on of land use and management, it can be a source or sink of greenhouse gases (RUMPEL et al., 2020), playing an essential role on climate change mitigation (LAL, 2004). However, accurate spatial quantification of soil carbon (SOC) stocks is still scarce, especially in developing countries with a low density of soil samples as Brazil. Despite a lack of sampling points for the Amazon biome, the state of Rondônia has a soil database with more than 2,900 profiles, resulted from a detailed soil survey in the 1990s.

The spatial distribution of SOC stocks can be performed using digital soil mapping techniques, using remote sensing associated machine learning (MCBRATNEY et al., 2003). Then, mapping SOC stocks at a regional level has the potential to improve the quantification and distribution of total SOC stocks and better guide public policy makers. Therefore, the objective of the study was to model and map the SOC stocks in the state of Rondônia at 1 m depth.

## Methodology

Rondônia is located in the Amazon biome and the climate is Am (tropical monsoon), with altitude ranging from 33 to 1,100 m. We used information from 2,914 soil profiles from the soil survey used for the second approximation of the Socioeconomic Ecological Zoning of Rondônia to model the distribution of SOC stocks at standard depths (0-5, 5-15, 15-30, 30-60 and 60-100 cm). We used as predictors the variables available related to the SCORPAN function, which describe quantitatively the soil formation factors, including soil classes, lithology (COCHRANE; COCHRANE, 2006), vegetation indices, relief (NASA JPL, 2020) and climatic variables (FICK; HIJMANS, 2017). To select the most important variables, we applied sequential predictor

selection methods, first by correlation > |0.95|, and after by the interactive recursive feature elimination (RFE) algorithm.

We used seven different algorithms in the modeling process: cubist; multivariate adaptive regression spline - earth; stochastic gradient boosting - gbm; elastic-net regularized generalized linear models - glmnet; k-nearest neighbors - k-knn; random forest - rf; and support vector machines - svmRadial. The process was executed 100 times for each algorithm and depth, varying the subset for training (75%) and testing (25%). The performance indexes mean absolute error (MAE), root mean squared error (RMSE) and coefficient of determination ($R^2$) metrics were used to selected. Finally, we created, an average map from the 100 runs with the that presented the higher performance. The uncertainty of prediction was evaluated through the coefficient of variation of the 100 maps. The entire process was performed on R software (R CORE TEAM, 2021).

## Results and discussion

The algorithms presented different performances in the 100 runs for all soil depths (Fig. 1). Random forest had the highest $R^2$ and lowest MAE and RMSE at all depths and was selected for the prediction of SOC stocks. The variation in performance metrics demonstrates the importance of repeating the adjustment process in machine learning under different subsets. This approach may avoid results that are not plausible in the reality of the modeled phenomenon.
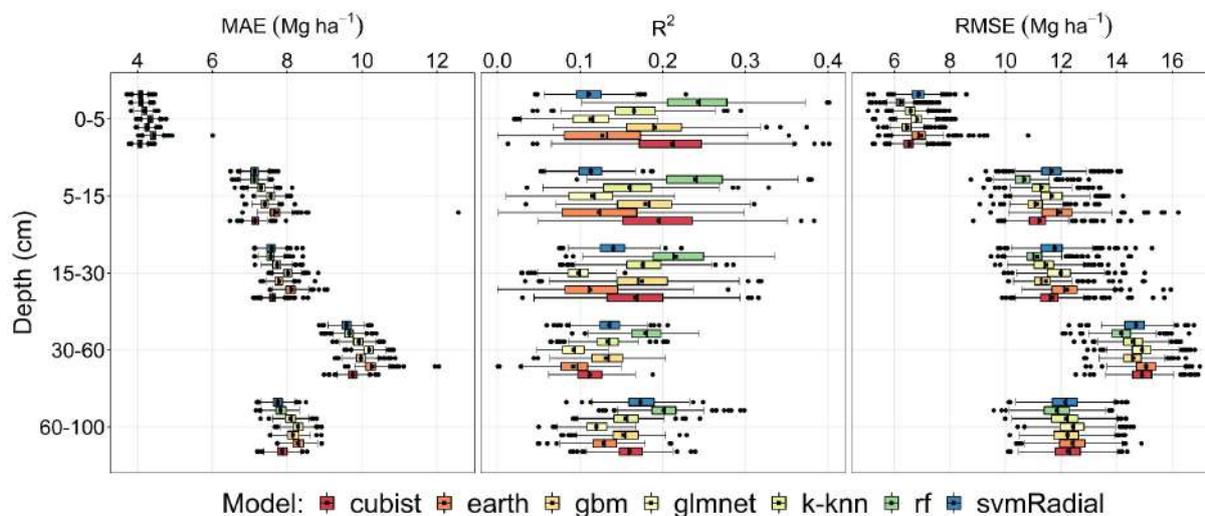


**Fig. 1.** Performance of different algorithms in the test set of soil organic carbon stocks at different depths, in Rondônia State, Brazil.

Random forest used between 12 and 20 predictors for performance stabilization. Soil classes and bioclimatic variables (temperature and precipitation) were the ones that stood out the most important variables for the prediction of SOC stocks. The soil class mainly describes texture and structure, which are determinant attributes in SOC stocks balance (BASILE-DOELSCH et al., 2020). The climate moderates the biological processes there affect carbon storage in soil (TAN et al., 2020), which explains the importance of these predictors in the in our modelling.

Rondônia stores 2,530.91 TgC and more than 50% of this carbon are located superficial layers (0-30 cm). The largest SOC stocks are found in the southern and

southeastern regions of the state (Fig. 2), where there is a predominance of Gleysols, and the smallest SOC stocks are in the central region of the state under Ferralsols and where the deforestation process is intense. The higher SOC stocks are associated the low altitudes, where soil genesis and seasonal flooding dynamics favor the accumulation of SOC (MAGHA et al., 2021). Regions with lower SOC stocks are associated to Ferralsols, which high weathering process reduced the amounts of SOC. In addition, the central region of Rondônia witnessed an intense agricultural expansion and deforestation for cattle farming, what can significantly reduce SOC stocks (MAIA et al., 2009). The coefficient of variation (CV) showed uncertainties bellow 25%, with the highest CV values in areas with high SOC stocks, evidencing greater difficulty of the model in making predictions in these places.
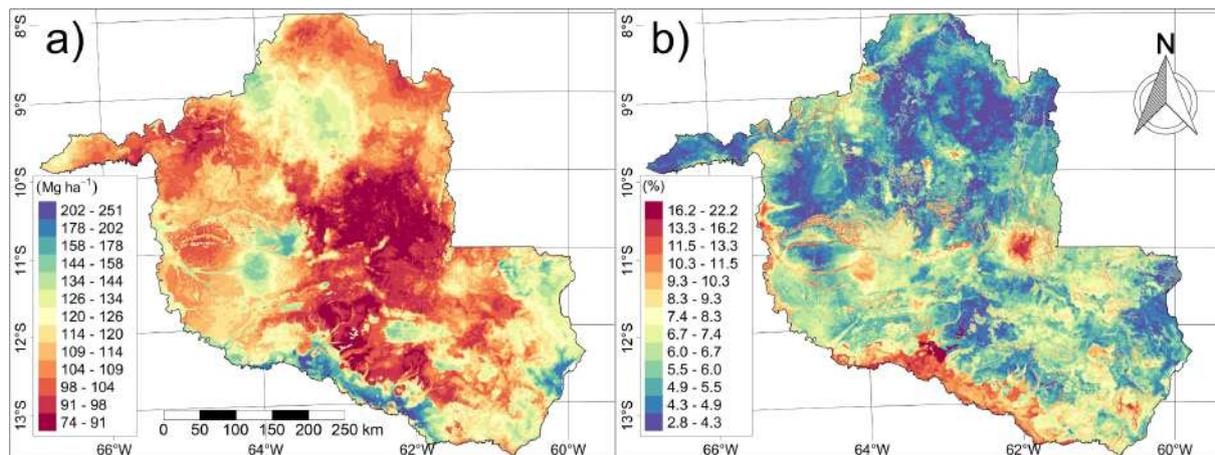


**Fig. 2.** Mean of 100 predicted soil organic carbon stock maps (a) and coefficient of variation (b) at 1 meter depth, Rondônia state, Brazil.

## Conclusions

Random forest had the best performance on modeling SOC stocks in Rondônia, and the most important variables were soil classes and climate. The SOC stock in the state of Rondônia at 1 meter depth is approximately 2,530.91 Tg. Wetlands hold the largest SOC stocks, while areas of Ferralsols under heavy anthropization contain the smallest values SOC in the state.

## Acknowledgements

## References

BASILE-DOELSCH, I. et al. Reviews and syntheses: The mechanisms underlying carbon storage in soil. **Biogeosciences**, v. 17, n. 21, p. 5223–5242, 2020. DOI: 10.5194/bg-17-5223-2020.

COCHRANE, T. T.; COCHRANE, T. A. Diversity of the land resources in the Amazonian state of Rondônia, Brazil. **Acta Amazonica**, v. 36, n. 1, p. 91–102, 2006. DOI: 10.1590/s0044-59672006000100011.

FICK, S. E.; HIJMANS, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. **International Journal of Climatology**, v. 37, n. 12, p. 4302–4315, 2017. DOI: 10.1002/joc.5086.

LAL, R. Soil carbon sequestration impacts on global climate change and food security. **Science**, v. 304, n. 5677, p. 1623–1627, 2004. DOI: 10.1126/science.1097396.

MAGHA, A. M. et al. Soil Water Characteristics of Gleysols in the Bamenda (Cameroon) Wetlands and Implications for Agricultural Management Strategies. **Applied and Environmental Soil Science**, v. 2021, p. 1–15, 2021. DOI: 10.1155/2021/6643208.

MAIA, S. M. F. et al. Effect of grassland management on soil carbon sequestration in Rondônia and Mato Grosso states, Brazil. **Geoderma**, v. 149, n. 1–2, p. 84–91, 2009. DOI: 10.1016/j.geoderma.2008.11.023.

MCBRATNEY, A. B. et al. On digital soil mapping. **Geoderma**, v. 117, n. 1–2, p. 3–52, 2003. DOI: 10.1016/S0016-7061(03)00223-4.

NASA JPL. **NASADEM Merged DEM Global 1 arc second V001NASA EOSDIS Land Processes DAAC.** NASA EOSDIS Land Processes DAAC, 2020. DOI: https://doi.org/10.5067/MEaSUREs/NASADEM/NASADEM_HGT.001.

R CORE TEAM. **R: A Language and Environment for Statistical ComputingR Foundation for Statistical Computing.** Vienna, Austria, 2021.

RUMPEL, C. et al. The 4p1000 initiative: Opportunities, limitations and challenges for implementing soil organic carbon sequestration as a sustainable development strategy. **Ambio**, v. 49, n. 1, p. 350–360, 2020. DOI: 10.1007/s13280-019-01165-2.

TAN, Q. et al. Clarifying the response of soil organic carbon storage to increasing temperature through minimizing the precipitation effect. **Geoderma**, v. 374, p. 114398, 2020. DOI: 10.1016/j.geoderma.2020.114398.

Nov, 24th to 27th, 2021
Brazil

# Modeling soils physical-hydric attributes through algorithms for quantitative pedology in Guapi-Macacu watershed, RJ

SANTOS, Priscilla A.[1]; PINHEIRO, Helena S. K.[2]; JÚNIOR, Waldir C.[3];
PEREIRA, Nilson R.[3]; BHERING, Silvio B.[3]; SILVA, Igor L.[4]

[1]UFRRJ Master in Geoscience - Petrology and Geotectonics Department, priscillaas@ufrrj.br;
[2]UFRRJ Soil Department Professor, koenow@ufrrj.br; [3]Embrapa Solos CNPS Researchers,
{waldir.carvalho@embrapa.br, nilson.pereira@embrapa.br, silvio.bhering@embrapa.br};[4]UFRRJ
Specialization in Applied Statistics – Mathematics Department, igorleite-ils@hotmail.com.

**Thematic Session: "Pedometrics: Innovation in Tropics"**

**Abstract**
The research goal is to analyze soil's properties and associate them with the behavior and vertical variability of soil basic infiltration speed (bir) and saturated hydraulic conductivity (ksat) in soils from Guapi-Macacu watershed using the Algorithm for Quantitative Pedology (AQP) package, in order to support predictive vertical modeling of soil attributes. To achieve the goals, 36 soil profiles were subjected to statistical analysis and then applied the AQP depth functions: standardization, slicing and aggregation methods. Thus, having the harmonized data set, the results were quantitatively and qualitatively evaluated, which pointed to high soil granulometric and physicochemical properties variability, maintaining a moderate to strong correlation with the physical-hydric attributes. It is concluded that the high soil properties variability can affect the vertical modeling in terms of prediction, as it tends to reduce the assertive degree in the training/validation of the models.
Keywords: AQP; Geoprocessing; Hydropedology; Digital Soil Mapping; Predictive Modeling.

## Introduction

Knowledge about soil physical-hydric attributes, such as is important to understand the water dynamics in watersheds (GARCÍA-SINOVAS et al., 2001). The water content stored and available affects the environmental functions of soils, the biodiversity and sustainability of this natural resource (FAO, 2017). Thus, the present work aims to understand soil`s physical-hydric attributes vertical variability, specifically the soil basic infiltration ratio speed (bir) and saturated hydraulic conductivity (ksat) from soils in Guapi-Macacu`s watershed and its relationship with other soils properties, such as particle size composition (sand, silt and clay), soil and particle density and porosity to apply pedotransfer functions (PTFs).
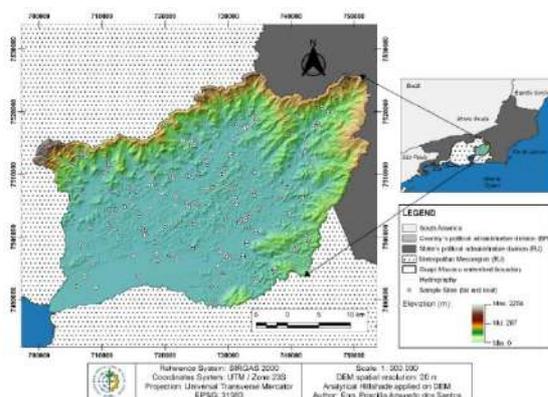
## Methodology

The studied area is composed by Guapi-Macacu whatershed (Figure 1), located in the Guanabara Bay Hydrographic Region (RH-V), in metropolitan region of Rio de Janeiro. Its domain is delimited by the political-administrative limits of the Itaboraí, Guapimirim and Cachoeiras de Macacu municipalities; and also by Guapiaçu town, reaching dimensions of 1250.78 km$^2$ of water catchment area and 199.2 km in perimeter extension (Projeto Macacu, 2010).

The soil profiles, as well as their physical-chemical analysis, come from the pedological

survey carried out in 2011 by Embrapa Solos company in partnership with the Research and Development Support Foundation - FAPED (CHAGAS et al., 2011). Data from 36 hydropedological sampling profiles were obtained by conducting a hydropedological survey with Guelph Permeameter.

The variables measured (bir and ksat) in two different layers (0-20 cm and 0-40 cm) were subjected to an exploratory statistical analysis. Then, soil depth functions were applied in final database, through the spline method implemented in a routine developed in RStudio v. 1.3.959 environment with R v.4.0.1 (R CORE TEAM, 2020), using Algorithms for Quantitative Pedology (AQP) package (BEAUDETTE et al., 2013). The soil profiles were sliced at a 1 cm interval (slice-wise method), and the data set was aggregated according to six predefined intervals of 20 cm in depth (slab-wise method). The results obtained through the AQP modeling were characterized in profile collections and analyzed according to each soil property. The methodology is shown at Figure 2.



**Figure 1.** Study area: Guapi-Macacu whatershed, Rio de Janeiro, Brazil.
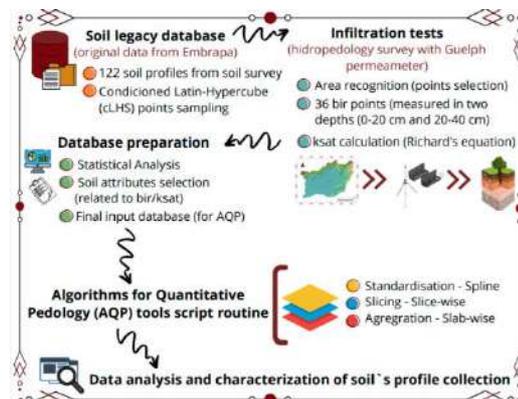


**Figure 2.** Proposed methodology flowchar.

## Results and discussion

The soil profiles collection presented characteristics such as: a total of 36 soil profiles, ranging from 0.30-1.15 cm in depth. The largest fraction of clay (400 to 700 g.kg$^{-1}$) (Figure 3) is found in the subsurface B horizon and increases with depth; the fraction of clay dispersed in water (350 to 450 g.kg$^{-1}$) (Figure 4) behaves inversely, being greater in the A horizon and decreasing in depth.

Sand (Figure 5) is well distributed along the profiles in the surface and sub-surface layers, ranging from 400-800 g.kg$^{-1}$. Soils with different particle sizes presented distinct behaviors in terms of easy water movement and particle translocation, greater in clay and silt, less in san, due to weight and size. Thus, indicating that bir and ksat is greater in superficial and subsurface horizons where the thin sand particle size portion (Figure 6) quantity is greater compared to coarse sand (Figure 7), and soil porosity (Figure 8) also increases in these saturated layers.

The ksat measured mainly increased in the limit between surface and subsurface layers (~20-40 cm), where clay (Figure 3) and silt (Figure 9) fraction decreases due to the fine sand proportion increase, affecting water flow in pores. Rain saturated soils shown a decrease in ksat value influencing measurements (low values in superficial layers and null values in subsurface portion).

Soil density (Figure 10) generally increases with soil depth of the Basin. Organic soils have low soil density values around 0-0.6 g.cm⁻³. Mineral soils (presence of horizon AB, BA and C) density decreased, while other soils density increased in depth. The behavior is quite different for particle density (Figure 11), which affects aggregation and water flow, associated with particle size (soil granulometry). The particle density values are higher than soil density, reaching a maximum of 2.65 g.cm⁻³.
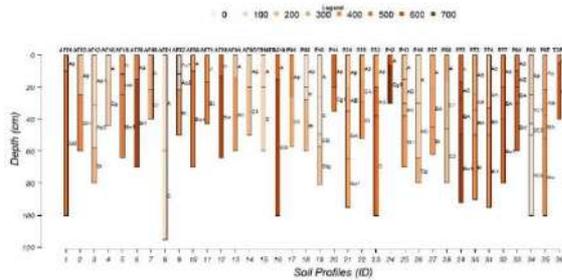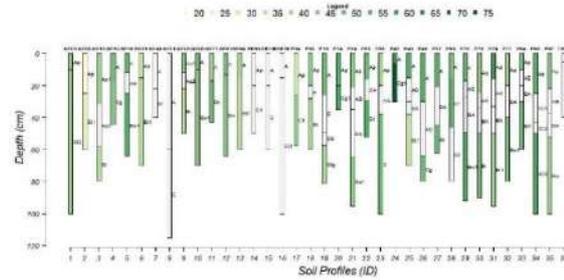


**Figure 3.** Clay attribute variability in Guapi-Macacu soils profile collection, in g.kg⁻¹.
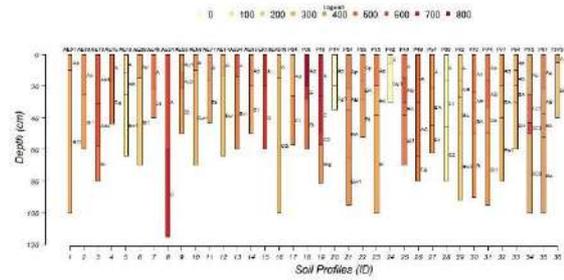


**Figure 7.** Coarse sand attribute variability in Guapi-Macacu soils profile collection, in g.kg⁻¹.
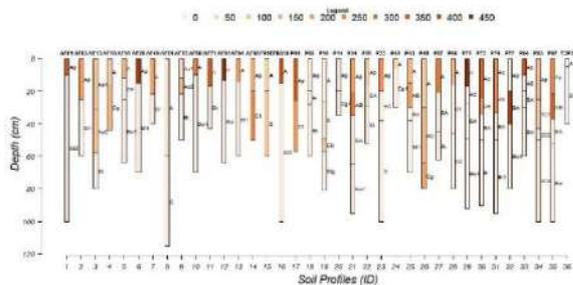


**Figure 4.** Dispersed clay attribute variability in Guapi-Macacu soils profile collection, in g.kg⁻¹.



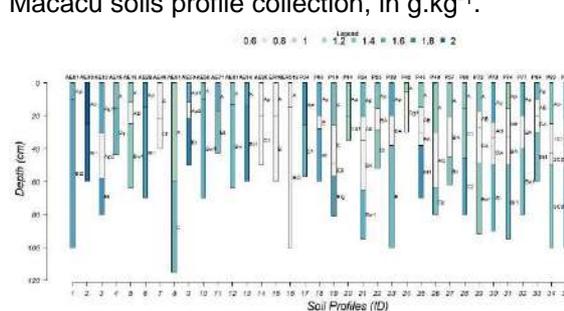**Figure 8**. Porosity attribute variability in Guapi-Macacu soils profile collection, in percentage (%).
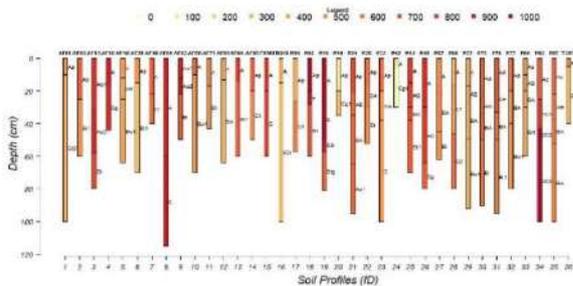


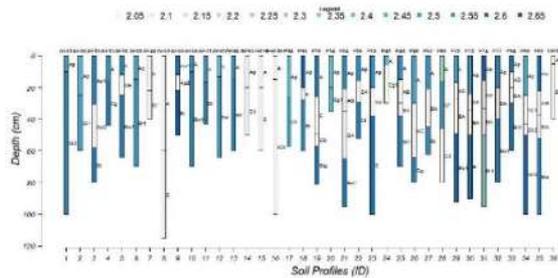**Figure 5.** Sand attribute variability in Guapi-Macacu soils profile collection, in g.kg⁻¹.
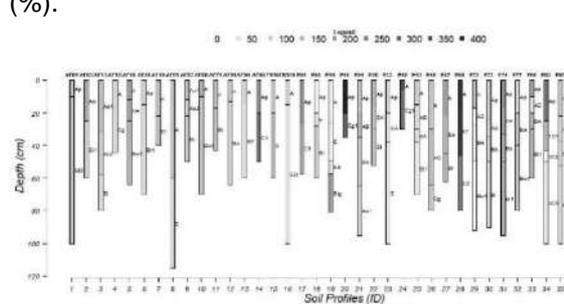


**Figure 9.** Silt attribute variability in Guapi-Macacu soils profile collection, in g.kg⁻¹.
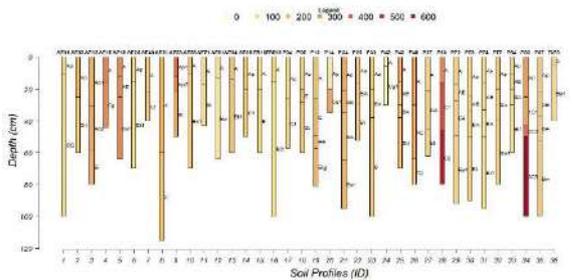


**Figure 6.** Thin sand attribute variability in Guapi-Macacu soils profile collection, in g.kg⁻¹.



**Figure 10**. Soil density attribute variability in Guapi-Macacu soils profile collection, in g.cm⁻³.

**Figure 11.** Particle density attribute variability in Guapi-Macacu soils profile collection, in g.cm$^{-3}$.

## Conclusions

The AQP use contributed to the texture variability analysis and soil physical-hydric attributes in depth, allowing the correlation between soil characteristics and its textural classification properties. Thus, this tool acts as a support for analysts in decision making at choosing input variables in predictive models' development (digital soil mapping), enabling the input data harmonization. Associated with machine learning methods and models, the AQP is a potential tool for preliminary studies in hydropedology, such as the implementation of pedotransfer functions, being recommended its use in research aimed at maintenance and conservation of soil water functions as directed by FAO (2017) (storage, availability and human supply, among others).

## References

BEAUDETTE, D. E., ROUDIER, P., O'GEEN, A. T. Algorithms for quantitative pedology: a toolkit for soil scientists, **Computers & Geosciences**, vol. 52, pp. 258-268, 2013.

CHAGAS, C. S.; JÚNIOR, W.C., PEREIRA, N. R.; BHERING, S. B.; FONSECA, O. O. M.; PINHEIRO, H. S. K.; MUSELLI, A.; JEUNE, W. **Descrição e resultados das análises dos perfis de solo coletados nas bacias hidrográficas dos rios guapi-macacu e caceribu - contrato nº 6000.00419115.08.2**. Relatório Técnico, Embrapa Solos, Rio de Janeiro, 2011. Acesso: 29 Abril 2021, <https://ainfo.cnptia.embrapa.br>.

FAO. **Voluntary Guidelines for Sustainable Soil Management**. Food and Agriculture Organization of the United Nations guide, Rome, Italy, 26p., 2017.

GARCÍA-SINOVAS, D.; REGALADO, C.; MUÑOZ-CARPENA, R.; ÁLVAREZ-BENEDÍ, J. Comparación de los permeámetros de Guelph y Philip-Dunne para la estimación de la conductividad hidráulica saturada del suelo, **Actas de las V Jornadas sobre Investigación en la Zona no Saturada**, pp. 31-36, 2001.

HORA, A. F.; HWA, C. S.; HORA, M. A. G. M. **Projeto Macacu: Planejamento Estratégico da Região Hidrográfica dos Rios Guapi-Macacu e Caceribu-Macacu**. Fundação Euclides da Cunha, Universidade Federal Fluminense, Niterói, Rio de Janeiro, 2010.

R CORE TEAM (2020). R: **A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>.

# Multiscalar Geomorphometric Generalization: a flexible approach to improve modeling of soil-landscape relationships.

ARAÚJO, Cauan[1]; OLIVEIRA JUNIOR, Raimundo[2]; BELDINI, Troy[3]

[1] Universidade Federal do Oeste do Pará, cauan.araujo@ufopa.edu.br; [2] Embrapa Amazônia Oriental, raimundo.oliveira-junior@embrapa.br; [3] Universidade Federal do Oeste do Pará, tpbeldini@yahoo.com.

**Thematic Session: Pedometria: Inovação nos Trópicos.**

## Abstract

The soil-landscape relationship occurs through complex interactions, that could be represented by generalized covariates and multiscalar geomorphometric analysis. So, the objective of this study is to predict soil surface attributes on scales compatible with mappings of small farms, by applying multiscale geomorphometric generalization. MGG is a upscalling operation, based on a cartographic concept of generalization applied at vector and raster representations, with unique multiscalar reference. The study was conducted ain watersheds located on Alter-do-Chão geological formation, Eastern Amazon. The results has show improvements with multiscalar generalized covariates, on soil surface attributes mappings with Random Forest. Future research is needed on: scale transformations methods; vector-raster scale correspondences; scale-related pedogenesis models on pedoenvironments.

Keywords: Digital Soil Mapping; Upscaling; Dimension analysis; Random Forest.

## Introduction

The soil-landscape relationship is related to the concept of the catena, coined by Milne (1935). Subsequently, the analyses of soil-landscape relations proposed by Hugget (1975), contemplated three-dimensional models of the slopes. In the context of digital soil mapping, in the scorpan model paradigm (MCBRATNEY et al., 2003), soil-landscape process modeling can be described as an interdisciplinary object of the interface between pedometry-geomorphometry.

The scale issues of soil-landscape relationships are related to the complex interactions. The multiscalar influences of topography on particular soil distribution have two general aspects: overlay of pedological processes that occurred at different times (SCHMIDT & ANDREW, 2005; TARGULIAN & KRASILNIKOV, 2007); and driving forces in the present time, determined by the sum of forces better correlated with one, several, or many geomorphologic scales (HU et al., 2020).

Some studies tested the geomorphometric covariables scales selection, show highest DEM resolutions do not necessarily produce the highest accuracy for predictive soil mapping (CAVAZZI et al., 2012; SAMUEL-ROSA et al., 2015).

For a friendly interpretation of soil-landscape scale specific relationships, this study proposed a cartographic-based criterion to formalize the scale correspondence to pixel size for geomorphometric covariables. The present study tested the hypothesis whether multiscale geomorphic representation, obtained from cartographic generalization of a digital elevation model, can improve pedometric modeling.
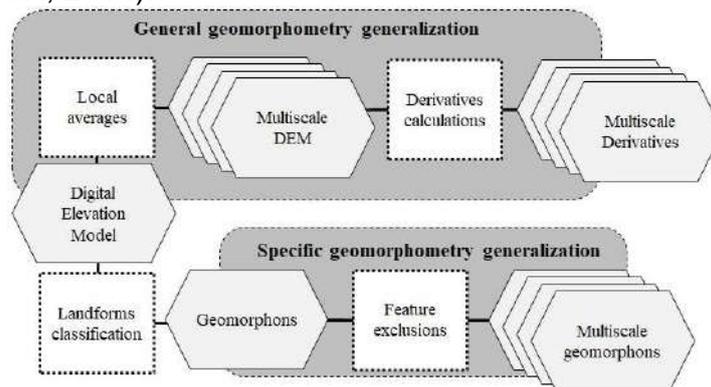
**Methodology**

The study area is in phanerozoic sedimentary basins, in the Alter do Chão geological formation, located in the Trombetas basin in Oriximiná-Pará in the Eastern Amazon. The geomorphic units of these watersheds are classified as a homogeneous dissection with coarse drainage density and weak incision depth.

In this study, the concept of minimum mappable area for soil surveys (IBGE, 2015) was considered to define pixel sizes in relation to cartographic scale. The detailed descriptions for each of the four scales used are in Table 1.

**Table 1.** Correspondence between scale and pixel size for Multiscale Geomorphometric Generalization (MGG), using the concept of minimum mappable area.

| Scale | Minimal mappable area (m²) | Pixel size (m) | Pixel area[a] (m²) | pa/mma[b] (%) |
|---|---|---|---|---|
| 1:25000 | 25000 | 30 | 22500 | 90 |
| 1:50000 | 100000 | 60 | 90000 | 90 |
| 1:75000 | 225000 | 90 | 202500 | 90 |
| 1:100000 | 400000 | 120 | 360000 | 90 |

a. For a 5x5 window. b. Ratio pixel area (pa) by minimal mappable area (mma), in percentage.

The geomorphic variables, at different scales, were obtained from SRTM DEM from two upscaling methods, as illustrated in Figure 2. Using local averages on covariable elevation, in 2x2, 3x3, 4x4 windows, for resolutions 60m, 90m and 120m, respectively, and subsequent derivatives covariable calculation. Classification of geomorphons was followed by the exclusion of polygons smaller than the minimum mappable area for each scale. Such methods correspond to cartographic generalization applied to general geomorphometry and specific geomorphometry, respectively (ZINCK, 2016).



**Figure 1.** Methodology flowchart of MGG for the topography covariables.

The modeling of the soil-landscape for prediction of soil superficial layer texture was done using Random Forest (BREIMAN, 2001). The sampling has done with 9 pilot areas, consisting of small farms, distributed in the upper and lower courses in both basins and adjacent basin boundaries, totaling 697 ha, approximately 1% of the extent of the watersheds. Each pilot area was sampled at 10 points, a sufficient density for semi-detailed soil surveys, compatible with the 1:25,000 scale soil maps (IBGE, 2015).

## Results and discussion

The prediction of sand and silt content, at original and multiscale generalized geo-morphometrics, is illustrated in Figure 2. In both variable groups, the MGG has pro-duced maps with less noise and more recognizable patterns related to geomorphic fea-tures. These results corroborate the hypothesis that the topography has an influence, in a larger spatial context, and has prevalence on prediction of soil particle size contents in the tested basin. In contrast, a case study with Random Forest with 30m and 90m DEM did not achieve significant differences in prediction (BHERING et al., 2016). Despite some similarity with co-variables importance, like Elev and RSP, the modeling is done on single scale datasets. In this sense, we can argue the importance of observing soil-landscape phenomena from a multiscale perspective.

The MGG was able to increase the accuracy of superficial layer soil texture classifications, as shown in Table 2. The user´s accuracy has a considerably higher result, so the MGG increased the reliability of each mapped class. In the same way, the Kappa index also has higher values for MGG geomorphometric variables.
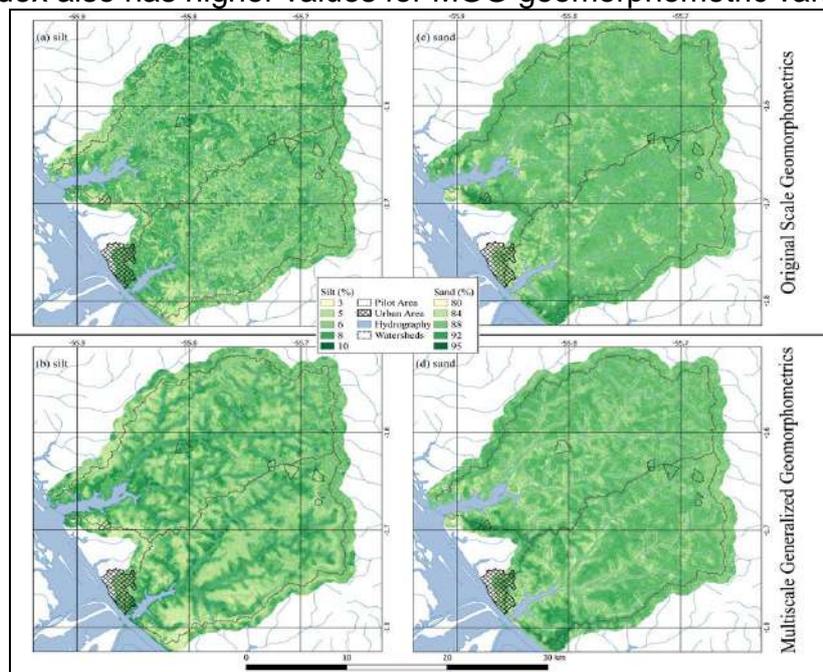


**Figure 2.** Predictive maps of silt (a,b) and sand (c,d) at original scale geomorphometrics and multiscale generalized geomorphometrics, respectively.

**Table 2.** Accuracy evaluation for soil superficial layer texture classification.

| Geomorphometric variables | User's / Producer's accuracy | | | Kappa Index |
|---|---|---|---|---|
| | MAr | ArMe | MeAr | All Classes |
| Original Scale | 75% / 84% | 72% / 58% | 20% / 20% | 43% |
| MGG | 81% / 88% | 76% / 71% | 100% / 67% | 62% |

## Conclusions

The MGG improved Random Forest model adjustment for silt and sand particles and also improved the accuracy of metrics of the soil texture classification of the surface

layer, especially for the most unusual classes, with the Kappa Index going from 43% to 62%. Topography influences on a coarser spatial scale and has prevalence on prediction of soil particle size contents in the studied watershed. Future research is needed on: scale transformations methods; vector-raster scale correspondences; scale-related pedogenesis models on pedoenvironments.

**References**

BHERING, S.B. CHAGAS, C.S. JUNIOR, W.C. PEREIRA, N.R. FILHO, B.C. and PINHEIRO, H.S.K. (2016) Mapeamento digital de areia, argila e carbono orgânico por modelos Random Forest sob diferentes resoluções espaciais**, Pesqui. Agropecu. Bras.**, 51, 1359–1370.

BREIMAN, L. (2001) Random forests. **Mach. Learn.**, 45, 5–32.

CAVAZZI, S. CORSTANJE, R. MAYR, T. HANNAM, J. and FEALY, R. (2012) Are fine resolution digital elevation models always the best choice in digital soil mapping?. **Geoderma**, 195, 111–121.

HU, G. R. LI, X. Y. and YANG, X. F. (2020) The impact of micro-topography on the interplay of critical zone architecture and hydrological processes at the hillslope scale: Integrated geo-physical and hydrological experiments on the Qinghai-Tibet Plateau. **J. Hydrol.,** 583, 12-46.

HUGGET, R.J. (1975) Soil landscape systems: A model of soil Genesis. **Geoderma**, 13, 1–22.

IBGE. (2015) Manual Técnico de Pedologia, 3a edition. Rio de Janeiro: IBGE.

MCBRATNEY A.B., MENDONÇA SANTOS, M.L., and MINASNY B. (2003) On digital soil mapping. **Geoderma**, 117, 3–52.

SAMUEL-ROSA, A. HEUVELINK, G.B.M. VASQUES, G.M. and ANJOS, L.H.C. (2015) Do more detailed environmental covariates deliver more accurate soil maps?. **Geoderma**, 243, 214–227.

SCHMIDT, J. and ANDREW, R. (2005) Multi-scale landform characterization. **Area**, 37, 341–350.

TARGULIAN, V.O. and KRASILNIKOV, P.V. (2007) Soil system and pedogenic processes: Self-organization, time scales, and environmental significance. **Catena**, 71, 373–381.

ZINCK, J.A. (2016) The Geomorphic Landscape: Criteria for Classifying Geoforms in Geo-pedology, Cham: Springer International Publishing, pp. 77–99.

# INFLUENCE OF SOIL COVER ON PORE DISTRIBUTION IN A FERRALSOL EVALUATED BY 3D COMPUTERIZED MICROCT

LEMES, Marcelo[1]; MACHADO, Alessandra[2] VASQUES, Gustavo[3]; RODRIGUES, Hugo[4]; LOPES Ricardo Tadeu[5]; ROSAS, Reiner Olíbano[6]

[1] Estacio de Sá University, marcelo.lemes@estacio,br; [2] Federal University of Rio de Janeiro; alessandramachado192@gmail.com; [3] Embrapa, gustavo.vasques@embrapa.br; [4] Federal Rural University of Rio de Janeiro, hugomr@id.uff.br; [5] Federal University of Rio de Janeiro, ricardo@lin.ufrj.br; [6] Federal Fluminense University, reiner_rosas@id.uff.br

**Thematic Session: Pedometrics: Innovations in Tropics**

**Abstract**
In the search for new techniques aimed at complementing and adding new data on Ferralsols, high resolution computer microtomography (microCT) appears as a non-destructive and fast analytical technique. MicroCT has been outstanding in the international scenarios and it is more and more present in soil analyses. Analyzing a soil sample by microCT allows obtaining knowledge, in microscale, on shape, size, distribution, volume, area and pore connectivity and having a 3D visualization of the soil sample and its structure. The objective of this work is to use microCT to compare the porosity distribution and pore connectivity density between a soil without vegetation cover and a soil with grass cover, both Ferralsols. Four undisturbed samples of each soil were collected. The results show that the MicroCT technique is an efficient and non-destructive tool for the analysis and characterization of the pore structure of soils protected from and degraded by erosion, underlining clear differences between them, as expected.

## Introduction

The pores of the soil are represented by cavities with different sizes and shapes, determined by the arrangement of solid particles which constitute a volumetric fraction of the soil filled with air, water and nutrients solution (Hillel, 1972). The soil porosity influences in aeration, water conduction and retention, resistance to penetration and branching of the roots in the soil and, consequently use of available water and nutrients (Tavares Filho, J. & Tessie, Stubff et al, 2014). Several techniques can be used to obtain soil porosity index. In the search for new techniques aimed at complementing and adding new data on Ferralsols, high resolution computer microtomography (microCT) appears as a non-destructive and fast analytical technique.

The microCT provides high resolution images with a set of volume data of an inspected sample that does not need to be modified and no preparation method has been submitted. Its physical principle is based on the attenuation of the X-rays when they interact with the object and are modulated according to the physical characteristics. To obtain microCT images, it is necessary to acquire many projections at constant angular steps and the reconstruction is performed with an appropriate algorithm based on the filtered overhead.

The objective of this work is to use microCT to compare the porosity distribution and pore connectivity density between a soil without vegetation cover and a soil with grass cover, both Ferralsols. Four undisturbed samples of each soil were collected.

## Methodology

The work developed on a slope within the limits of the Stream of Thorn Microbasin, which has a total area of $9.14km^2$ being located in the Farming Paiol, where agricultural activities related to the genetic improvement of cattle and dairy cattle are developed. The area is located in the municipality of Silva Jardim, State of Rio de Janeiro, and access is given by BR 101.

The undisturbed samples for the tests were collected on the half slope of a dissected hill within the limits established by erosion plots, totaling eight replicates packed in plastic film for transportation and handling without loss. The samples were distributed in the following way: four in soil with grass cover and four in soil without vegetation cover, using acrylic tubes measuring 50mm in height and 32mm in diameter.

The samples were scanned in a high energy system - Skyscan / Bruker, model1173. The system operates with voltage and current of 130 kV and 61 μA, respectively. A flat panel detector (2240 x 2240 pixels) was used to register the transmission of the X-ray beam. After acquisition, the image is captured and reconstructed using the FDK reconstruction algorithm, (Feldkamp; Davis; Kress, 1984).

In the present study, we chose to use a adaptive method segmentation. In this method for each voxel, the threshold is calculated as the mean of all pixel/voxel grayscales within a selected radius. In this way, the binary image is obtained, with the objects (soil matrix) in white and the background (pore) in black. Thus, it was possible to quantify the total porosity and density of connectivity. The schematic of this segmentation.

## Results and discussion

To acquire the data, the values of the parameters were adjusted to acquire the information within a pattern that responded in the most accurate way, searching for the openings with the smallest possible size, in order to identify more clearly the class of Micropores. In this sense the soil pore diameters analyzed were classified according to Brewer (1964) that define micropores that have a diameter smaller than 0.03mm and macropores larger than this value.

Covered soil (CS) presented the following percentages of total porosity: 21.1; 24.6; 27.3 and 36.7%; As the soil without cover (SWC) had porosities of 13.6; 20.5; 21.8 and 30.8%. The pore densities in the covered soil were 46.7; 76.7; 155.4 and 508.63 mm3, while in the uncovered terrain were 19.9; 45.8; 76.7 and 511.3 mm3. The values of soil porosity presented higher values in the soil covered with grass in relation to the same pairs of soil samples without vegetation cover. This trend was also observed in the values of density of connections between the pores. (Table 1)

The values corresponding to macroporosity were 7.37; 13.78; 9.01 and 12.54% on CS soil and 7.66; 7.48; 13.18; 12.52 not only SWC. Both coverages obtained an average very close to this index, varying only 0.46%. Already the values of microporosity for the soil CS was 13.75; 10.91; 18.37 and 24.17%, in SWC soil the indices were in the houses of 5.94; 13.08; 8.71 and 18.36. In this cyst it is noticed that there is evidence for CS soil, showing an average 5% higher than SWC soil. This can be explained because the permanence of the vegetal cover increases organic

matter in the soil, it keeps the root system active and assists in the stability of its aggregates generating an increase of the microporisdade according to Viana *et al* (2011).

In order to obtain a standard of comparison, the soil water retention curve test was carried out in soil samples from the same aforementioned environments (CS and SWC), which presents several practical, technical and scientific applications, such as: The determination of the soil field capacity, the permanent wilting point and the total availability of water in the soil, indispensable variables for an adequate irrigation management, soil water balance and macro and micro porosity (Table 2).

One of the ways to determine the water retention curve in the soil is to use the Richards pressure chamber, which simulates a determined tension in the soil sample and later, by weight difference (wet soil after being subjected to pressure - Soil dried in an oven at 105 ºC for 48 hours), the water content related to the applied voltage is determined. For all eight samples submitted to the assay an eight-strain sequence was applied on the following increasing scale: 0.01; 0.033; 0.06; 0.33; 1.00; 5.00 and 15.00 bar.

What is evident when carrying out a relationship with both methods is that the Retention Curve is larger than the Computed Microtomography because this method can identify a greater percentage of microporosity reaching the house of 23.17% more in the soil CS and 28.05% in soil SWC. This is due to the fact that the maximum resolution of the MCT reaches the equivalent of 0.03mm, however there are a series of pores with smaller sizes that are not identified by the tomographic sensors. Regarding macroporosity data, the Retention Curve showed an average of 12.11% for CS soil and 5.40% for SWC soil. In this same index for MCT, 10.21% for CC soils and 11.52 for SWC were presented. This demonstrates a greater capacity of the microtomography in the identification of pores larger than 0.03mm, mainly to environments that have already been worn in compacts for a period of five years.

| Samples | | Retention Curve | | | |
|---|---|---|---|---|---|
| | | Total Porosity (%) | Macropores (%) | Micropores (%) | Soil density (g/cm³) |
| Soil with grass Cover | 1 | 49.45 | 8.95 | 40.90 | 2.52 |
| | 2 | 52.86 | 15.16 | 42.10 | 2.63 |
| | 3 | 50.78 | 9.98 | 35.80 | 2.53 |
| | 4 | 53.66 | 14.36 | 41.10 | 2.59 |
| Soil without Vegetation | 5 | 46.15 | 5.25 | 40.50 | 2.59 |
| | 6 | 45.10 | 3.00 | 37.70 | 2.57 |
| | 7 | 47.03 | 11.23 | 40.80 | 2.53 |
| | 8 | 43.24 | 2.14 | 39.30 | 2.57 |

Table 1 - Distribution of porosity and connectivity density in different systens of use.

| Samples | | X-ray computed microtomography | | | |
|---|---|---|---|---|---|
| | | Total Porosity (%) | Macropores (%) | Micropores (%) | Connectivity Density (mm³) |
| Soil with grass — Cover | 1 | 30.88 | 12.52 | 18.36 | 511.13 |
| | 2 | 24.69 | 13.78 | 10.91 | 76.76 |
| | 3 | 27.38 | 9.01 | 18.37 | 155.42 |
| | 4 | 36.71 | 12.54 | 24.17 | 508.63 |
| Soil without Vegetation | 5 | 13.60 | 7.66 | 5.94 | 19.91 |
| | 6 | 20.56 | 7.48 | 13.08 | 45.84 |
| | 7 | 21.89 | 13.18 | 8.71 | 76.79 |
| | 8 | 21.12 | 7.37 | 13.75 | 46.03 |

Table 2 - Distribution of the values processed by Retention Curve.

## Conclusions

The results show that the presence of the vegetal cover is a relevant factor in the increase of the porosity of the superficial layers of the soil, because the root system develops ducts that are connected with the progress of its development. In this sense, identified a concentration of 18.60 g/dm3 of organic carbon in the A horizon of this soil, which may aid in the stability of aggregates and condition the occurrence of pores. On the other hand, the lack of vegetation cover considerably reduces the pore indices of the superficial layers of the soil, which can be explained by the fact that the sealing process of the exposed soil occurs when the material disaggregated by erosion caused by the impact of the raindrops (Splash) obliterates the pores, corroborating with this hypothesis identified, for the same area, very high apparent density values for horizon A, at the house of 1.45 g/dm3. Thus, the MicroCT technique demonstrated an efficient and non-destructive tool for the analysis and characterization of the pore structure of soils protected from and degraded by erosion, underlining clear differences between them, as expected.

## References

BREWER, R., 1964. Fabric and mineral analysis of soils. Huntington, N.Y.: R.E. Krieger. Crestana, S., 1992. Calibração e Uso de um Tomógrafo Computadorizado em Ciência do Solo. Revista Brasileira de Ciência do Solo.

FELDKAMP, L.A., DAVIS, L.C., KRESS, J.C., 1984. Pratical cone beam algorithm. Journal of the Optical Society of America. 1(6), 612-619.

HILLEL, D., 1972. Soil and water: physical principles and processes. 3rd. Ed. New York: Academic, 288.

TAVARES FILHO, J., TESSIER, D., 2009. Characterization of soil structure and porosity under long-term conventional tillage and no-tillage systems. Revista Brasileira de Ciência do Solo. 33, 1837-1844.

VIANA, E.T.; BATISTA, M. A.; TORMENA, C. A.; COSTA, A. C. S. & INOUE, T. T. Atributos físicos e carbono orgânico em latossolo vermelho sob diferentes sistemas de uso e manejo. Revista Brasileira de Ciência do Solo, 35:2105-2114, 2011.

# Spatial dependence of organic carbon in a soil profile of Lagoa Grande das Queimadas in Várzea Branca, PI.

MARQUES, Rabech Grasiely Gomes[1]; LIMA NETO, Miguel Alvores[2]; VALLADARES, Gustavo Souza[3]; FREITAS, Aline Gonçalves de[4]

[1] UFPI, rabechgrasiely1998@hotmail.com; [2] UFPI, miguel3limaalvores@gmail.com; [3] UFPI, valladares@ufpi.edu.br; [4] UFPI, tuttyfreitas@gmail.com.

## Thematic Session: Pedometrics: Innovations in Tropics

**Abstract**

The objective of this work was to evaluate the spatial dependence of the total organic carbon (TOC) of the archeological site Lagoa Grande das Queimadas. For this purpose, soil samples were collected up to 200 cm deep, totaling 38 samples, and the TOC was analyzed in the laboratory. The data obtained were analyzed using descriptive statistics and geostatistical. The results showed that there is spatial dependence. The kriging showed that the TOC up to 80 cm decays in depth and from 80 cm to 100 cm there is an increase and decaying again after 100 cm. The use of geostatistical techniques proved to be efficient to understand the spatial dependence of the data. The results indicate lithological discontinuity.

Keywords: Geostatistics; Semivariogram; Kriging; Archaeological record.

## Introduction

The applicability and use of geostatistics as a methodology for analyzing data in space or time is widespread in several branches of science and it is interested in determining the spatial dependence of observations on a variable. In soil science, geostatistics provides information about the spatial variability of its properties in a sampled area, and this knowledge allows for proper management planning. Many works use the technique to assess horizontal variability, few works on vertical variability in the soil profile.

Given the above, the objective of this work was to evaluate the spatial dependence in depth of the levels of TOC of an open trench in Lagoa Grande das Queimadas located in the southwest of the State of Piauí, using descriptive and geostatistical statistics.

## Methodology

The study area comprises Lagoa Grande das Queimadas located in Várzea Branca in the southwest of the state of Piauí, between two conservation units: the Serra da Capivara National Park and the Serra das Confusões National Park. The region in which the lagoon is located presents minimum temperatures of 18 ⁰C and maximum of 36 ⁰C, with a semi-arid, hot and dry climate. The vegetation is shrub-arboreal caatinga (CPRM, 2004).

The regional geology is composed of granites and schists from the Pre-Cambrian Sobradinho-Remanso Complex and tertiary-quaternary dendrite-laterite deposits. The

depressed shaped site has Holocene sediments. The municipality of Várzea Branca is located in the Canindé-Piauí Hydrographic Sub-Basin, whose main watercourse is the Piauí River. In the study area, a Haplic Gleysol occurs.

The research presents the TOC analysis of 38 samples collected up to 200 cm deep inside the Lagoa das Queimadas. Samples were collected every 10 cm, for the first 3 depths and the others every 5 cm.

The TOC were quantified following the Embrapa Soil Analysis Methods Manual (TEIXEIRA, et al., 2017). To measure the TOC, the soil sample is placed in an acidic medium and the TOC is determined by the oxidation of potassium dichromate, in the end the remaining dichromate is titrated with ferrous ion.

After this procedure, the data were tabulated and descriptive statistical analysis performed. To understand its spatial dependence, data were analyzed using semivariograms and kriging (VIEIRA et al., 2020). The depth of each layer was used as a geographic coordinate, starting from the surface where the value zero was assigned.

## Results and discussion

Data analysis through descriptive statistics showed that the soils have an average of 1.50 g kg-1 TOC, standard deviation and variance of 1, minimum value 0.18 and maximum value 3.00, showing that the data show moderate variability. In addition, asymmetry presented a value of -0.1 and kurtosis of -1.6, indicating that the values present asymmetry to the left and platykurtic distribution.

The spatial dependence of TOC was verified by adjusting the semivariogram (Figure 1). A good fit of the semivariogram can be observed, indicating a moderate spatial dependence.

The adjusted model had a nugget effect (C0 = 0.15), a level of 0.39 indicating spatial dependence and the range showed that these samples are correlated up to a depth of 11.58 cm (Figure 01).

The data have a structural variance (C1) of 0.24, with a coefficient of determination ($R^2$) of 0.71 and a degree of spatial dependence of 62%. Furthermore, the gaussian model was the one that best suited the observed data.

The figure 02 shows the depth distribution of TOC observed in the samples and estimated by Kriging. Up to a depth of approximately 85 cm there is a decrease in TOC, with observed and estimated values being close, that is, with small deviations. Between 85 and 110 cm, the TOC distribution is erratic, with greater deviations between measured and estimated values. From 110cm onwards, TOC contents are significantly smaller than up to 85cm and estimates return to few deviations. The

results indicate that there is a lithological discontinuity or fluvic character in the profile from 85 cm in depth.



Model : 0.14888*Nugget+0.24372*Gaussian(11,583,2)
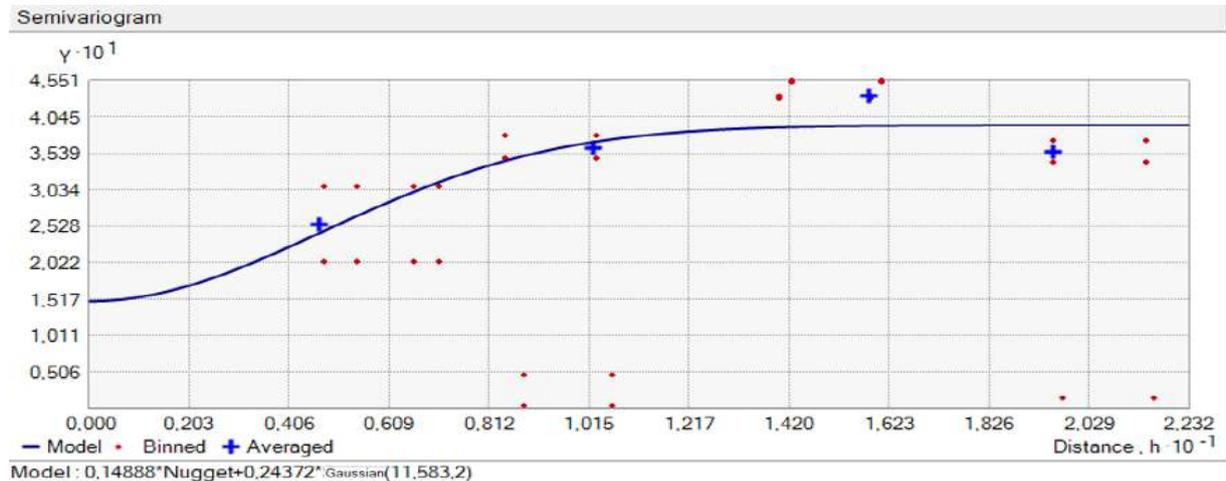
**Figure 01-** TOC semivariogram of Lagoa Grande das Queimadas archeological site.
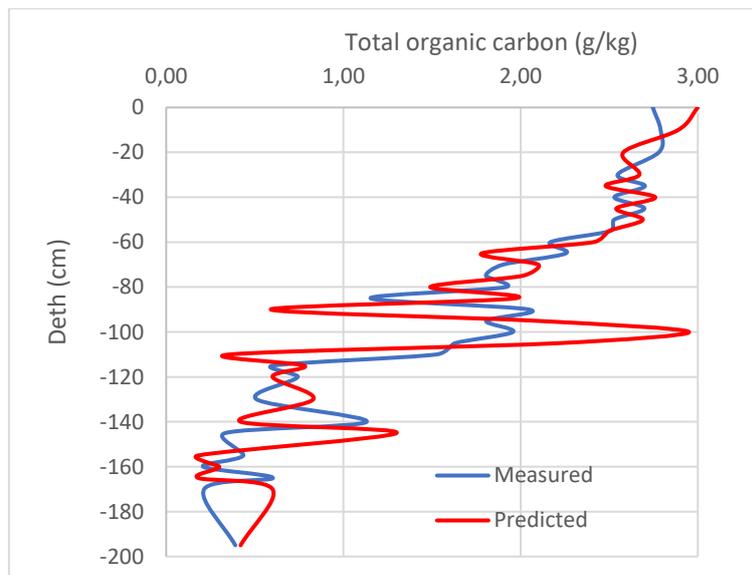


**Figure 02-** Depth distribution of observed and estimated organic carbon contents by kriging.

## Conclusions

The TOC contents in the profile studied had a good adjustment to the semivariogram and a good estimate can be affirmed by the kriging, being able to employ the method for interpolation. In view of the observed results, it was possible to make inferences about the origin of the parent material of the studied soil.

## References

CPRM. **Projeto cadastro de fontes de abastecimento por água subterrânea, estado do Piauí: diagnóstico do município de Várzea Branca**. - Fortaleza, 2004. Disponível em: http://rigeo.cprm.gov.br/xmlui/bitstream/handle/doc/16459/Rel_VarzeaBranca.pdf?sequence=1 Acesso em: 27. maio. 2020.

TEIXEIRA, P. C.; DONAGEMA, G. K.; FONTANA, A.; TEIXEIRA, W. G. (Ed. Técnicos). **Manual de métodos de análise de solo**, 3. ed. rev. e ampl. Brasília: Embrapa, 2017. 575p.

VIEIRA, S. R. Geoestatística em estudos de variabilidade espacial do solo. In: NOVAIS, R. F.; ALVARES, V. H.; SCHAEFER, C. E. G. R. (Ed.). **Tópicos em ciência do solo**. Viçosa, MG: Sociedade Brasileira de Ciência do Solo, 2000.

# Digital Soil Mapping: some issues of pedological concern

CARVALHO FILHO, Amaury de[1]; MOTTA, Paulo Emilio Ferreira da[2]; CARVALHO, Ana Letícia Lima de[3]

[1] Embrapa Solos, amaury.carvalho@embrapa.br; [2] Embrapa Solos, motta.pauloemilio@gmail.com; [3] Universidade Federal de Juiz de Fora, carvalho.ana@engenharia.ufjf.br

## PEDOMETRICS: INNOVATION IN TROPICS

**Abstract**

Aiming to contribute to the development and improvement of modern soil mapping techniques, an analysis of the basic concepts and practices adopted by the Digital Soil Mapping was performed, under the focus of pedological knowledge. The main deficiencies and inconsistencies concerning the Scorpan model and the conceptual approach of the method are identified and discussed. Some problems related to the soil maps produced are also pointed out. The importance of pedological knowledge for the development of soil mapping techniques and soil interpretation is emphasized, which in the authors' opinion should receive special attention from the scientific community dedicated to the topic.

Keywords: soil survey; soil genesis; pedological knowledge; scorpan model; soil science.

## Introduction

Since its introduction (McBratney et al., 2003), the Digital Soil Mapping (DSM) has been considered a revolutionary approach, and its advantages and future perspectives frequently exalted (McBratney et al., 2019). Despite the large number of studies carried out in nearly two decades, there remains a difficulty in reconciling standardized procedure for a comprehensive application. Efforts in this direction have focused on the adequacy of statistical and computational techniques and selection of covariates; issues concerning pedological foundations have received little or no attention. In this sense, this work aims to evaluate some aspects of the DSM from a pedological point of view, as contribution to modern techniques of soil mapping.

## Methodology

From a comprehensive evaluation of several studies applying DSM techniques, and having as reference review and synthesis papers and basic documents that present the fundamental principles of the method (McBratney et al., 2003; Dobos et al., 2006; SSDS, 2017; Rossiter, 2018, McBratney et al., 2019), an analysis of the main concepts, practices and results produced by the DSM methodology was performed, which are discussed below from the point of view of pedological knowledge.

## Results and discussion

As proposed by McBratney et al. (2003), the basic principle of Digital Soil Mapping (DSM) is the use of environmental variables related to soil formation factors for prediction of soil classes or soil properties, based on the so called Scorpan model, which is used as a foundation to express quantitative evaluations in a spatial context.

This approach originates from a supposed similarity with traditional procedures of making soil surveys, which would involve a predefined model of soil formation applied to soil properties data and other environmental variables that have significant impact on soil formation (Dobos et al., 2006).

The Scorpan model is based on the fundamental equation of soil-forming factors, proposed by Jenny (1941). It is formulated as an empirical quantitative function $S = f(s, c, o, r, p, a, n)$, where S (soil classes or soil attributes at a point in space and time) is estimated by seven environmental covariates — s: soil (other properties of the soil); c: climate; o: organisms; r: relief; p: parent material; a: age (the time factor); n: spatial position.

The genetic connotation implicit in the model is evident. In this sense, some questions arise. Given the great variability of environmental conditions related to soil genesis, and the strong interdependence among them, what are the objective conditions of establishing the formula factors in order to ensure reliable quantitative estimates? And how to take into account the interactions between the factors?

Besides the logical inconsistency of considering soil characteristics (factor s) as an independent variable in the equation, which implies a relationship of self-dependence in the genesis of a natural element (as something whose origin depends on itself), establishing the other factors on a quantitative basis seems a very complex problem. Thus, climate (climatic properties of the environment) and relief (landscape attributes) can be decomposed into a series of attributes, but how to quantify, or even to estimate, the participation of each one? On the other hand, organisms (vegetation or fauna or human activity) can be represented by vegetation or digital land cover data (SSDA, 2017); but how to take into account the influence of soil fauna constituents, microorganisms among them?

Even more difficult is to estimate the parent material, referred as lithology (McBratney et al., 2003), which can be derived from a geology map according to Soil Survey Manual (SSDS, 2017). At this point there is a contradiction between bedrock (lithology), the central theme of geological mappings, and parent material, as evidenced by the concept presented by the Soil Survey Manual itself (SSDS, 2017): "Soil parent material is not always residuum weathered directly from underlying bedrock. The material that developed into the modern soil may not be related to the underlying bedrock at all. In fact, most soils did not form in place but were subject to transport and deposition by wind, water, gravity, or human activities."

However, the most complex and perhaps most important factor for an adequate adjustment of the equation is time (age), which has influence on all the others. From the variations in environmental conditions over the time of soil formation results the recognized soil polygenetic character, whose understanding requires specific and detailed studies of soil genesis. As reminded by Arnold (1999): "The exact combination of physicochemical and biological reactions that have actually transformed materials over time into soil horizons of a specific soil can never be known with certainty."

In summary, the Scorpan model presents a set of basic inconsistencies, both from the point of view of soil genesis knowledge and the specification of the equation parameters. Another problem of DSM conceptual approach refers to prediction based on soil genesis inferences, which presents a strong speculative aspect and incorporates a hypothetical character to the results obtained. Here, a word is necessary on the supposed similarity between DSM and traditional soil survey. It is important to highlight that the soil survey is based in relationships established between soils and other recognizable elements of the environment, evaluated directly in the field. This procedure does not exclude the use of soil genesis knowledge (and consequently of formation factors) to establish relationships between soils and other environmental variables — but its application should be restricted as an element of inference, not of determination. Although mental models are applied throughout the mapping work, they are progressively tested and refined in a continuous process of adjustment. For this reason, the detail level of the soil survey, more than the map scale itself, has an essential significance as an index of reliability and as a guide for interpretation of results. Even though subtle, this aspect implies a strong distinction between the traditional method of soil survey and the DSM approach. At the same time, it reveals a distorted understanding of the soil mapping process, which is reflected both in the digital methodological procedures and in its products.

In practice, for DSM predictions, the Scorpan model is just partially applied regarding the formula factors. While some factors (relief, for example) are decomposed into numerous attributes (environmental covariates), others are related restrictedly to one characteristic, or may not even be considered, which occurs frequently with the time factor (age). Each environmental covariate data (including soil) can be directly obtained by field determinations (soil profiles, for instance) or derive from remote or proximal sensing images, or even from thematic maps. Therefore, there are a large number of possible variations, regarding the nature of covariates selected, the information sources, the predictive models and the data processing techniques used. As a result, subjective conditions are incorporated into the process, whereas the results depend, widely, on the model adopted (subject of the action) as well as on the attributes, many of which are not directly related to the soil (object of the evaluation).

The high degree of subjectivity is expressed in the prediction differences observed in numerous digital mapping studies. This is one of the reasons for the difficulty of establishing a standard protocol for applying DSM. Such difficulty has been attributed to sampling deficiency, the need to adjust prediction models and covariate selection techniques, or even to hidden factors (Rossiter, 2018). Its causes, however, have much deeper roots. They result from inconsistencies of the DSM method discussed above, which in essence are linked to the disconnection with the pedological knowledge, the basis for understanding soils (Arnold, 1999).

In fact, some propositions adopted by the DSM evidence insufficient soil knowledge. Among them, the attribution of spatial resolution to soil maps (instead of scale) and the mapping of soil classes — both devoid of meaning. In the first case, a characteristic of sensor images (that record information captured from a real-world

element) is applied to a thematic map, which corresponds to a representation of the geographic distribution of a natural element (not the phenomenon itself). In the second case, the fact that a soil class is just a concept, and as such impossible to map, is ignored. There is here a confusion, common to non-specialists, between taxonomic class and soil mapping unit as a landscape segment, which reinforces the importance of pedological knowledge for soil mapping and interpretation (Arnold, 1999).

The lack of perception of the inherent utilitarian character of soil maps is also explicit. This is exemplified by numerous DSM studies that present as product maps of soil classes at the order or suborder levels, which in themselves have no practical use. An eloquent example is the recent efforts to estimate the probability of occurrence of soil classes or properties. There is no technology that can solve the impasse concerning the practical use of maps indicating, for the same area, distinct possibilities of spatial distribution of soils (most probable and second-most-probable), as presented by McBratney et al. (2019; see Fig. 3).

## Conclusions

The DSM approach presents a series of deficiencies associated to non-observation of knowledge developed by Pedology, which compromises its suitability as a methodology for soil mapping, which should be considered by scientific community related to the subject.

## References

ARNOLD, R. W. **The soil survey: past, present, and future** (1999). Disponível em: https://www.nrcs.usda.gov/wps/PA_NRCSConsumption. Acesso em: 15 set. 2021.

DOBOS, E. et al. **Digital soil mapping as a support to production of functional maps**. 68 p. Official Publications of the European Communities, Luxemburg. 2006.

JENNY, H. **Factors of soil formation: a system of quantitative pedology**. McGrawHill, New York, 1941. 281 p.

McBRATNEY, A. B.; MENDONÇA SANTOS, M. L.; MINASNY, B. On digital soil mapping. **Geoderma**, v. 117, p. 3-52, 2003.

McBRATNEY, A.; GRUIJTER, J. de; BRYCE, A. Pedometrics timeline. **Geoderma**, v. 338, p. 568-575, 2019.

ROSSITER, D. G. Past, present and future of information technology in pedometrics. **Geoderma**, v. 324, p. 131-137, 2018.

SSDS (Soil Science Division Staff). U. S. Department of Agriculture. **Soil survey manual**. Washington, D.C., 2017. 603 p. (USDA. Agriculture Handbook, 18).

# Scale variability of soil invertebrate fauna under use and management systems.

SIQUEIRA, Glécio Machado[1]; PEREIRA, Daniel Rocha[1]; GUEDES FILHO, Osvaldo[2]; SILVA, Raimunda Alves[1]

[1] Universidade Federal do Maranhão, Programa de Pós-Graduação em Biodiversidade e Biotecnologia da Amazônia, glecio.siqueira@ufma.br; daniel.rocha.drp@gmail.com; ray-234@hotmail.com;

[2] Universidade Federal do Paraná, Campus de Jandaia do Sul, osvaldoguedes@ufpr.br

## Thematic Session: Pedometrics: Innovations in Tropics

**Abstract**

Soil invertebrate fauna quickly responds to environmental changes and has been used as environmental quality indicator. The objective of this work was to evaluate the scale variability of soil invertebrate fauna under use and management systems. Invertebrate fauna was sampled in a transect with 128 points under the following systems: Maize, Soybean, Eucalyptus stage I, Eucalyptus stage II, preserved Cerrado and disturbed Cerrado. Data were evaluated using multifractal analysis to verify the complexity and heterogeneity. Maize, preserved Cerrado and disturbed Cerrado presented higher richness of taxonomic groups. Invertebrate fauna showed multifractal behavior with the complexity and heterogeneity influenced by management. Soybean presented monofractal behavior, which reflects the low system diversity.

Keywords: scale heterogeneity; singularity spectrum; soil fauna diversity; spatial variability

## Introduction

The soil is habitat of many invertebrate taxonomic groups that play diverse ecological functions. They are sensitive to minimal environmental variations and has been used as environmental quality indicator. Invertebrate fauna has high complexity and scale variabilities in natural ecosystems (Silva et al., 2020). However, in anthropic systems the complexity decreases. According to Vidal-Vázquez et al. (2013) the heterogeneity description provides important information's about spatial and scale variabilities. The objective of this work was to evaluate the scale variability of soil invertebrate fauna under use and management systems.

## Methodology

The experiment was carried out at Fazenda Unha de Gato (3º70'80.88'' S and 43º18'71.27'' W – Figure 1), in areas under use and management systems: Maize (MI – 103 ha); Soybean (SO - 113 ha), Eucalyptus stage I (E1 - 3,79 ha), Eucalyptus stage II (E2 - 3,79 ha), preserved Cerrado (CP - 33,08 ha) and disturbed Cerrado (CA - 20,44 ha). Pitfall traps were installed in 05/01/2016 in transect with 128 sampling points, with three meters distant each other.

Data were analyzed by multifractal technique following the presumptions of moment method (Halsey et al., 1986; Evertsz and Mandelbrot 1992) and direct method (Chhabra e Jensen, 1989), generating successive partitions to k (k = 1, 2, 3…), where

at each scale δ, a number of segments, N(δ) = 2k with characteristic size length, δ = L × 2$^{-k}$, were obtained, covering the entire extent of the support, L. That allows the construction of partition function graphs (Evertsz and Mandelbrot 1992; Vidal Vázquez et al., 2013), and singularity spectrum (Chhabra and Jensen (1989).
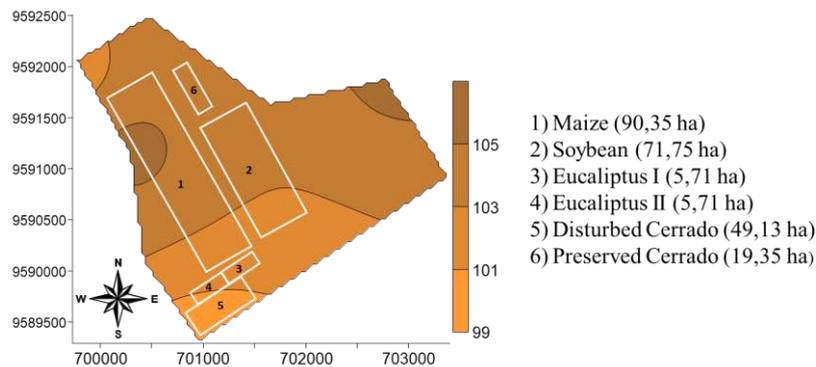
Figure 1. Sampling areas at Fazenda Unha de Gato, Maranhão, Brazil.

## Results and discussion

A total of 40,524 individuals were collected comprising 36 taxonomic groups (Figure 2a). The area under Maize presented higher abundance (15,502 – 38.25%), followed by Eucalyptus stage II (8,590 – 21.19%), Eucalyptus stage I (5,630 – 13.89%), preserved Cerrado (5,146 – 12.69%), disturbed Cerrado (3,882 – 9.57%) and Soybean (1,837 – 4.53%).

Partition functions were built to successive segments of 2k in k = 0 and k= 8 in order intervals −10 < q < 10 and had multifractality with adjustment greater than 0.97. In figure 2b is showed the partition function of richness of taxonomic groups for preserved Cerrado.
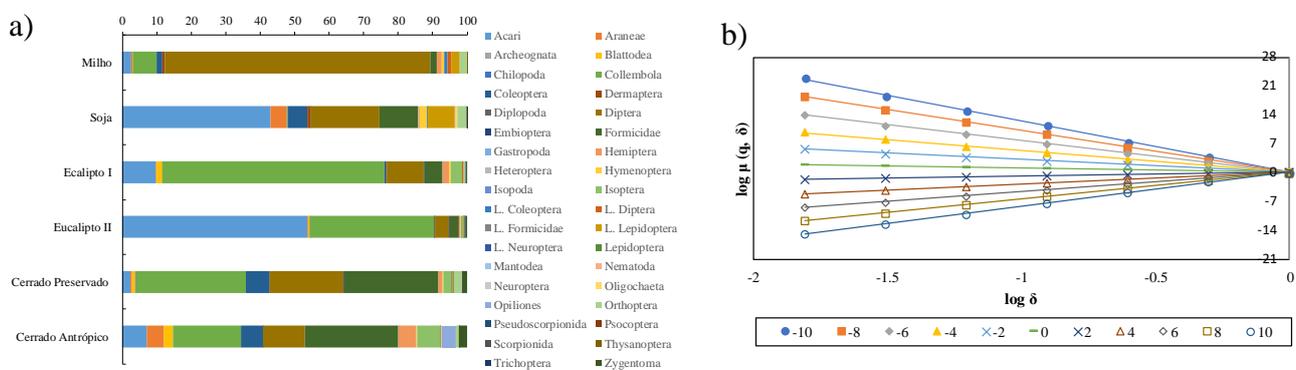
Figure 2. Identified taxonomic groups in experimental plots (2a) and partition functions (2b) of richness in preserved Cerrado.

The spectrum of the generalized dimension (Dq) describes a typical sigma-shaped curve and provides indicator parameters of properties of multifractal dimension (Vidal-

Vázquez et al., 2013). The left sides of the spectra represent the q negative moments, which correspond to higher measured concentrations, while the right side represents the q-positive moments, which correspond to the lower measured concentrations. Therefore, the differences in these three moments of the generalized dimension (Dq) were used to evaluate the heterogeneity of scale properties. When D0 = D1 = D2, the distribution of the data series is characterized as monofractal; however, if D0 > D1 > D2, the measure distribution is considered to be multifractal.

The singularity spectrum graph to organism abundance expressed in individual per trap per day (Figure 3a) described that organism abundance cover multifractal systems. maize, Eucalyptus stage I and II showed asymmetric branches to the left, indicating that along the transect there is domain of high values of measurements. Soybean, preserved and disturbed Cerrado had asymmetric branches to the right, demonstrating a domain of low values of scale measurements.
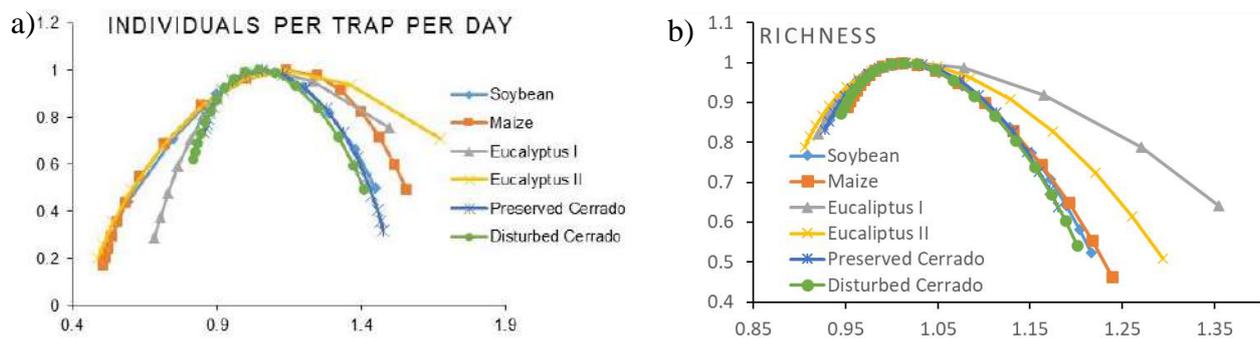


Figure 3. Spectrum of singularity to the number of individuals per trap per day (a) and richness of taxonomic groups (b) in the experimental plots.

The spectrum of singularity to richness of taxonomic groups presented asymmetric branches to the right, indicating a domain of low values of scale measurements. Eucalyptus stage I and II had spectrums with higher amplitude. We highlight that soil invertebrate fauna in Soybean and Maize cultivations had scale variability like Cerrado areas.

## Conclusions

Maize area had higher abundance and lower diversity of taxonomic groups. The variations found in scales of heterogeneity to abundance and richness indicates that each these parameters reflect a variability in distinct manners. Multifractal analysis pointed the differences in the soil use and management indicating a domain of low and high values of measurement and describing the heterogeneity of the evaluated systems.

## References

CHHABRA, A., Jensen, R.V., 1989. **Direct determination of the f (∝) singularity spectrum**. Physical Review Letters 62: 1327-1330.

EVERTSZ, C.J.G., MANDELBROT, B.B., 1992. **Multifractal measures**. In H. Peitgen, H. Jürgens, and D. Saupe (Eds.), Chaos and Fractals. Berlin: Springer. p. 921-953.

HALSEY, T.C., JENSEN, M.H., KANADOFF, L.P., PROCACCIA, I., SHARIMAN, B.I., 1986. **Fractal measures and their singularities: the characterization of strange sets**. Phys Rev A 33: 1141-1151.

VIDAL VÁZQUEZ, E., CAMARGO, O.A., VIEIRA, S.R., MIRANDA, J.G.V., MENK, J.R.F., SIQUEIRA, G.M., MIRÁS AVALOS, J.M., PAZ GONZÁLEZ, A., 2013. **Multifractal analysis of soil properties along two per¬pendicular transects**. Vadose Zone J 12: 1-13.
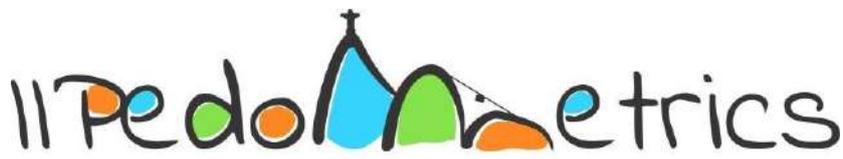
# Soil erodibility estimate in areas under natural and anthropped environments in the South of the Amazon

BRITO FILHO, Elilson[1]; CAMPOS, Milton César[1], HASSANE, Abdul[2]; SANTOS, Luís Antônio[3]; ASSIS, Juliana[2]; CUNHA, José Maurício[2]

[1] Universidade Federal da Paraíba – CCA/UFPB, bfsambiente@gmail.com, mcesarsolos@gmail.com; [2] Universidade Federal do Amazonas – IEAA/UFAM, assaneluis@gmail.com, juhmalta99@gmail.com, maujmc@gmail.com; [3] Universidade do Estado do Amazonas, santoslac@gmail.com

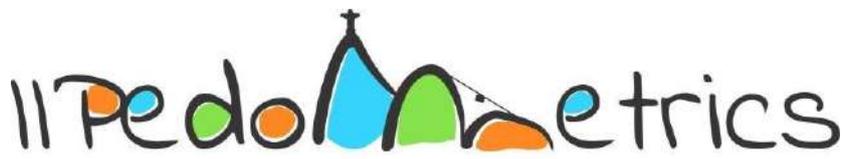**Thematic Session:** Pedometrics: Innovations in Tropics

## Abstract

The work aimed to estimate soil erodibility in natural and anthropogenic environments in the southern region of Amazonas. Areas of approximately one hectare were selected in each study environment, namely: native forest 1, native forest 2, cerrado, cerradão, pastagem, reforestation with teca, reforestation with jenipapo, misto reforestation (teca, jenipapo, andiroba, samaúma, mogno). In each environment soil samples were randomly collected at a depth of 0-20 cm, 32 per area, 256 in total. Texture analysis and soil organic carbon were performed, and then soil erodibility was estimated using indirect prediction methods. Data were subjected to descriptive and univariate analysis, as well as principal component analysis. According to the results obtained, it is observed that the evaluated areas of native forest 1 and 2, cerradão and pastagem show high susceptibility, reflecting the level of erodibility, this in relation to the cerrado, reforestation with jenipapo, teca and misto reforestation.

## Introduction

Soils in the southern region of Amazonas have been undergoing changes due to the replacement of forest areas by the most diverse use systems, without due knowledge and compliance with technical criteria, and this has been one of the main problems in the region (Frozzi et al., 2020 ), which has led to the acceleration of erosive processes, causing damage to the ecosystem.

In this context, in order to estimate the erodibility, some studies used indirect methods, as, among other characteristics, it has a low cost and provides important information for diagnosing the use and management activities. Of these, a study carried out by Brito et al. (2020), who studied the estimation of erodibility in areas of Terras Preta de Índio under the use of cocoa, pasture and coffee in the region of Apuí, AM.

Knowledge of changes in soil attributes caused by various anthropogenic uses provides assistance for the adoption of management practices that allow increasing the yield of the production process with the conservation of environments (Souza et al., 2020). For Oliveira et al. (2020), although several scientific works and the efforts of a large number of researchers have contributed to the advancement of knowledge of soils in the southern region of Amazonas, there is a need to expand this knowledge for a better compression of Amazonian ecosystems. Given the above, the objective of this

work was to estimate the soil erodibility in natural and anthropogenic environments in the southern region of Amazonas.

## Methodology

The study was carried out between January and March 2020 in areas of the municipality of Humaitá, south of the state of Amazonas (Brazil). The areas of native forest 1 (FN1) cerrado, cerradão are located at km 20, of the BR-319, towards Humaitá – Manaus, in an area belonging to the 54th Infantry Battalion of the Brazilian Army. Pastagem and native forest areas 2 (FN2) are located on a private property at km 45 of BR-319 towards Manaus. Reforestation areas with teca, jenipapo reforestation, misto reforestation (teca, jenipapo, andiroba, samaúma and mogno) is located on a private property located at km 18 of BR-319 towards Manaus, AM.

In each studied area (native forest 1 and 2, cerrado, pastagem area, teca, jenipapo and misto reforestation area), 32 collection points were selected in the central part of an area equivalent to one hectare and then samples were collected of soil randomly using a Dutch auger at a depth of 0-20 cm, making a total of 256 samples in the studied areas.

The collected soil samples were dried in the shade and later shrunk manually, passing them through a sieve with a diameter of 2.00 mm, obtaining the air-dried fine earth, then the total organic carbon content was determined (TOC) and estimated soil organic matter (SOM), as well as particle size analysis (texture) (Teixeira et al., 2017). To fractionate the sand, it was sieved, aiming to estimate the factors of soil erodibility, using a sieve shaker with 2mm meshes; 1mm; 0.5mm; 0.250mm; 0.125mm and 0.053mm. From the fractions, the soil erodibility variables were determined: K factor (global erodibility - Denardin (1990)), Ki factor, Kr and Tc (between-groove, furrow and critical shear erodibility - (Flanagan and Livingston, 1995)).

## Results and discussion

In relation to soil organic matter (SOM), this presented mean values ranging from 33.07 to 102.55 g kg$^{-1}$, in the following ascending order: teca> misto> pastagem > cerrado > cerradão>jenipapo> FN1 > FN2. The areas of the natural environment had higher content compared to the cultivated areas. This result is due to the management systems adopted in agricultural crops that have a great influence on the MOS stock, which may decrease, maintain or increase in relation to the area's native vegetation (Caetano et al., 2013), as well as the diversity of organisms present in natural environments which, in turn, a certain percentage is lost in the conversion of environments to anthropized places, these organisms are responsible for the decomposition of organic material and incorporation of carbon and nutrients into the soil (Gonçalvez and Santana, 2019).

The mean K factor ranged from (0.04 to 0.06 t.ha$^{-1}$.MJ$^{-1}$.mm$^{-1}$.ha.h), starting in the jenipapo > teca > misto > FN1> cerrado > cerradão > pastagem> FN2. Within the classification proposed by Castro et al. (2011), for the K factor, the jenipapo, teca and misto reforestation areas were classified as areas of high erodibility, whereas the FN1

and FN2, cerrado, cerradão and pastagem areas, these have very high erodibility, in both cases it is necessary a conservation management plan at these sites in order to mitigate and prevent evolution to higher levels of erosion, gullies in this case (Hernani et al., 1999).

The average Ki wepp factor ranging from ($3.10E^{+06}$ to $4.85E^{+06}$ kg.s.m$^{-4}$), starting in the area FN2 > cerrado > jenipapo > FN1 > misto > teca > pastagem > cerradão. It is observed that the natural (FN1 and FN2 and cerrado) and crop areas (Jenipapo, Teca and Misto) have lower averages compared to pastagem ($4.17E+06$ kg.sm-4) and cerradão ($4. .85E^{+06}$ kg.sm$^{-4}$), which corroborates the values presented by Brito et al (2020). This shows that pastagem and cerradão are more susceptible to erosion between ridges.

The results of the principal component analysis are shown in Figure 1. It was possible to verify that the cerrado area has a greater relationship with sand associated with global erodibility (K) and inter-groove erodibility (Ki), showing the easy detachment of particles in this environment of natural form (Corrêa et al., 2015).
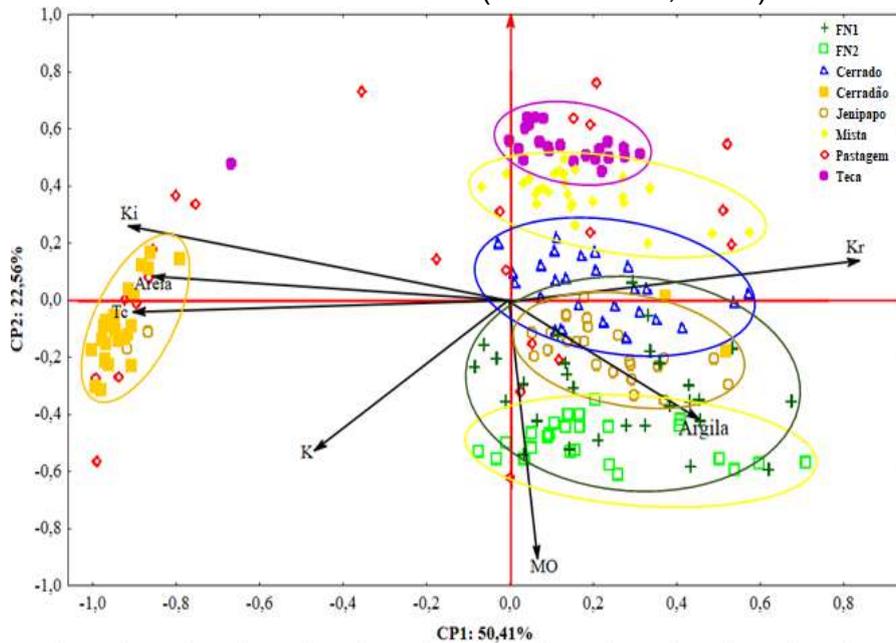


**Figura 1.** Principal component analysis of the soil attributes studied at a depth of 0–20 cm for the areas FN1, FN2, cerrado, cerradão, pastagem, jenipapo, teca and misto, in southern Amazonas. Areia: sand, Argila: clay, MO: soil organic matter.

The areas of FN1e FN2 and jenipapo presented a high content of soil organic matter (MOS), combined with a higher clay content, thus causing low K factor rates, and even with a smaller amount of vegetation cover, the cerrado also presented low evidence of lighter, shallower erosion (Kr). According to Costa (2013), the high content of organic matter in the soil influences its structure and stability. Thus, the importance of clay in the soil is justified as essential in the aggregation and stabilization of soil aggregates, as they will contribute to greater resistance against erosive processes (Brito et al., 2020).
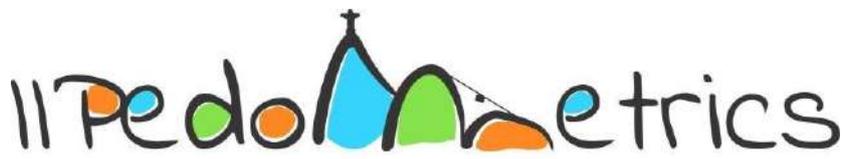
## Conclusions

Soil organic matter has mechanisms that are fundamental for reducing the risk of soil erosion in natural environments, being crucial as a form of mitigation in cultivated environments.

Of the areas evaluated, the cerradão and pasture show the highest erosion levels in relation to the other areas, a fact that is closely associated with higher levels of sand and low clay, even with high values of critical shear stress.

## References

BRITO, W.B.M. et al. Dynamics and spatial aspects of erodibility in Indian Black Earth in the Amazon, Brazil. **Catena**, v. 185, p. 104281, 2020.

CAETANO, J. O. et al. Dinâmica da matéria orgânica de um Neossolo Quartzarênico de Cerrado convertido para cultivo em sucessão de soja e milheto. **Revista Brasileira de Ciência do Solo**, v. 37, n. 5, p. 1245-1255, 2013.

CASTRO, W.J. et al. Erodibilidade de solos do cerrado goiano. **Revista em Agronegócios e Meio Ambiente**, v. 4, n. 2, p. 305-320, 2011.

CORRÊA, E. A. et al. Espacialização temporal das perdas de solo em uma microbacia hidrográfica com predomínio de solos arenosos. **Geografia**, v. 40, n. 1, p. 101-118, 2015

COSTA, E.M.; SILVA, H.F.; REBEIROS, P. R.D. Matéria orgânica do solo e o seu papel na manutenção produtividade dos sistemas agrícolas. **Revista Enciclopédia biosfera, Centro Científico Conhecer**- Goiânia. v.9,n.17;p. 1843, 2013.

DENARDIN, J.E. **Erodibilidade de solo estimada por meio de parâmetros físicos e químicos**. 1990. 81p. Tese (Doutorado) - Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, 1990.

FLANAGAN D. C.; LIVINGSTON S. J. **Water erosion prediction project: WEEP user summary**. West Lafayette: National Soil Research Laboratory & USDA – Agricultural Research Service, Report 11, p 25-26, 1995.

FROZZI, J.C. et al. Physical attributes and organic carbon in soils under natural and anthropogenic environments in the South Amazon region. **Environmental Earth Sciences,** v. 79, p. 251-266, 2020.

GONÇALVES, T. S.; SANTANA, C. S. Os organismos do solo e a manutenção da matéria orgânica. **Revista GeoAmazônia**, v. 7, n. 14, p. 139-159, 2019.

HERNANI, L. C. et al. Sistemas de manejo de solo e perdas de nutrientes e matéria orgânica por erosão. **Revista Brasileira de Ciência do Solo**, v. 23, n. 1, p. 145-154, 1999.

OLIVEIRA, I.A. et al. Pedoenvironmental indicators of soil in Western Amazonia, Brazil. **Environmental Monitoring and Assessment**, v. 192, p. 768-774, 2020.

SOUZA, F.G. et al. Aggregate stability and carbon stocks in Forest conversion to different cropping systems in Southern Amazonas, Brazil**. Carbon Management**, v. 11, p. 81-96, 2020.

TEIXEIRA, P.C. et al. **Manual de métodos de análise de solo**. 3ª. Ed. Embrapa Solos. Brasília. p.573, 2017.

# Soil data curation with the help of an expert system for soil classification

VAZ, Glauber J.[1]; SILVA JÚNIOR, Adalberto F.[2]; SILVA NETO, Luís de França[3]

[1] Embrapa Informática Agropecuária, glauber.vaz@embrapa.br; [2] Universidade Federal Rural de Pernambuco, adalbertofrancisco75@gmail.com; [3] Embrapa Solos, luis.franca@embrapa.br

**Thematic Session: Pedometrics: innovation in tropics**

**Abstract**

We used an expert system for soil classification to curate soil data and improve its quality. The system is the first to accurately classify soil profiles to the fourth level of the current version of the Brazilian Soil Classification System (SiBCS). We analyzed 94 soil profiles using the expert system, which guided the necessary changes on soil data to make it consistent with the corresponding classifications. About 45% of soil profiles did not require data treatment, and most changes were related to horizon symbols. Even after data treatment, changes in classification were necessary for almost 40% of the profiles on at least one categorical level. Therefore, using an expert system for soil classification can help identify inconsistencies in data and classifications of soil profiles, in addition to guiding the necessary changes. It can also help improve the SiBCS.

Keywords: data quality; soil profiles; digital tool.

## Introduction

Increasing the quantity and quality of soil data and information is essential for improving soil resource governance. It is one pillar of action for the Global Soil Partnership (GSP) (FAO, 2021), which aims to improve soil governance to guarantee healthy and productive soils as well as supporting the provision of essential ecosystem services. The National Soil Program of Brazil (Pronasolos) (POLIDORO et al., 2016) was proposed to provide richer information on Brazilian soils for decision making. The long-term program has five main lines of action, one of which deals with database and soil information.

Soil classification is an essential component of soil science. The Brazilian Soil Classification System—in Portuguese, *Sistema Brasileiro de Classificação de Solos* (SiBCS)—is the official taxonomic system for soil classification in the country. It is structured in the form of a taxonomic key up to the fourth categorical level. It also contains recommendations of qualifiers for the fifth level and suggested properties for the sixth categorical level (DOS SANTOS et al., 2018).

The correct classification relies on the consistency and completeness of soil data, which involves dozens of soil attributes. Vaz et al. (2019) developed an expert system for automatic soil classification and analyzed data from a widely used soil database, comparing the results of the system with the classifications registered in the database. They showed the need for greater data curation of available databases under the supervision of soil scientists and presented the system as a powerful tool to assist with this activity. However, their analysis was limited to the first level of SiBCS, and did not include data curation. In the present study, we used the same

expert system as Vaz et al. (2019) to examine soil data to the first four levels of SiBCS, as well as took steps to curate the data and improve its quality.

## Methodology

The expert system we used to analyze the soil profiles is based on the rules of SiBCS for its first four categorical levels. The classification provided by the system only considers the current version of SiBCS (DOS SANTOS et al., 2018).

We formed a team of soil and computer scientists to analyze the data with the help of the expert system. Once completely validated, the software should provide correct classifications in all cases since the data provided is correct and complete. In some situations, the software can generate wrong classifications due to the considerable complexity and the number of possible classes in the system. When this occurs, the software is corrected and starts to produce the expected result. Therefore, all the automatic classifications generated by the system in the present study were correct and verified by soil scientists.

We analyzed 94 soil profiles from the states of Pernambuco and Rio Grande do Norte in Brazil. These samples were collected during the GeoTab Project, which aimed to organize soil data from the Brazilian coastal tablelands and update the classifications of the profiles. These are available in '.doc' files, meaning that the data needed to be processed in order to be generated in the format required by the expert system. We did it using an app called SmartSolos.

After obtaining the automatic classification for a given soil profile, we compared it with the recorded one. When the classes were different, we analyzed the data to check its consistency and the rules of the software in order to verify its correctness. The source of such differences could be errors in software, soil attribute data, or classification. For each case, we made the necessary changes.

## Results and discussion

Table 1 shows the number of profiles analyzed for each first categorical level (order) of SiBCS. The 'Classification' columns provide the number of soil profiles from each order that were classified by the expert system according to the records made previously by soil scientists. The 'Data' columns indicate whether data treatment was required in order to obtain a correct classification.

The 'Ln' columns give the number of profiles whose records were correctly classified to the $n^{th}$ level. For example, the classification of nine out of 25 *argissolos* was consistent with verified records to the fourth level, while 15 *argissolos* had correct classifications to the third level, but not the fourth. Finally, one profile that was actually an *argissolo* had been labeled with entirely different classes.

Table 1: The classifications and the consistency of data for the analyzed soil profiles.

| Order | # Profiles | Classification | | | | | Data | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | L0 | L1 | L2 | L3 | L4 | OK | Horizon | Addition | Update |
| *Argissolos* | 25 | 1 | 0 | 0 | 15 | 9 | 12 | 9 | 1 | 3 |
| *Cambissolos* | 11 | 0 | 0 | 1 | 2 | 8 | 4 | 5 | 2 | 1 |
| *Chernossolos* | 2 | 0 | 0 | 0 | 0 | 2 | 1 | 1 | 1 | 0 |
| *Espodossolos* | 3 | 0 | 0 | 0 | 0 | 3 | 3 | 0 | 0 | 0 |
| *Gleissolos* | 8 | 0 | 0 | 3 | 0 | 5 | 6 | 0 | 2 | 0 |
| *Latossolos* | 11 | 0 | 0 | 1 | 2 | 8 | 8 | 2 | 1 | 0 |
| *Luvissolos* | 5 | 2 | 0 | 0 | 1 | 2 | 2 | 3 | 0 | 0 |
| *Neossolos* | 13 | 1 | 0 | 3 | 0 | 9 | 4 | 9 | 7 | 0 |
| *Nitossolos* | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| *Organossolos* | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| *Planossolos* | 7 | 2 | 0 | 0 | 1 | 4 | 1 | 6 | 0 | 0 |
| *Plintossolos* | 4 | 0 | 0 | 0 | 0 | 4 | 1 | 3 | 0 | 0 |
| *Vertissolos* | 2 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 |
| Unknown | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Total | 94 | 7 | 0 | 9 | 21 | 57 | 43 | 41 | 14 | 5 |

The 'Data' column group indicates the changes, if any, required for each classification:
• OK: data were consistent; therefore, no change was made.
• Horizon: changes in the horizon symbols.
• Addition: additional data were needed.
• Update: updates in some attributes.

In order to arrive at a consistent classification, profiles occasionally required changes to horizon symbols, attribute updates, or additional data. The sum of numbers in the 'Data' columns is not necessarily equal to the number of profiles examined from the corresponding order, as is the case of the 'Classification' columns. This might occur, for example, when a single profile requires changes in both horizon symbol and attribute update.

After data treatment, the system classified 60.6% (57/94) of all profiles in a manner consistent with the records at all four levels. Meanwhile, 22.3% (21/94) of profiles were consistent with the third level, with errors only arising in the fourth. In most of these cases, the registered class at the fourth level is no longer valid. As such, these errors were largely caused by incompatibilities across SiBCS versions, and the records had not yet been updated. In 9.6% (9/94) of profiles, only the first and second levels were correct. In 7.5% (7/94) of profiles, the classification was completely different from the original. Therefore, some change in classification was necessary for almost 40% (37/94) of the profiles.

To obtain a correct classification, data must be correct and complete. No data treatment was required for 45.7% (43/94) of profiles. Of the profiles that did require changes, most needed only the adjustment of symbol horizons, which can be quickly done by a specialist. Updating obsolete symbols to the current standard and adding a missing suffix were the most common changes. In 14.9% (14/94) of the profiles, it was necessary to add data that a specialist would be able to distinguish but were not explicitly registered. In some cases, it was necessary to replicate the dry color in other horizons or to add an attribute indicating, for example, cohesive qualifier, fluvic qualifier, or alterable primary materials. Data not related to horizon symbols only had to be updated in 5.3% (5/94) of cases, generally for a single attribute. Thus, incorrect attribute values were corrected after analysis by a domain specialist who identified the inconsistencies in the data. In many cases, they were only recognized because the classification obtained by the system was not equal to the one recorded—furthermore, the results from the expert system provided indications of the necessary changes.

It is important to note that one profile was classified by the system as "unknown" for the first level. The current version of SiBCS considers the predominance (>50%) of activity clay in the B horizon to classify *luvissolos* and *argissolos*. However, in the profile classified as "unknown", 50% of the B horizon had low-activity clay and 50% high-activity clay. Therefore, it is not classified either as a *luvissolo* or as an *argissolo*. This demonstrates another benefit of the expert system, namely its ability to validate SiBCS rules using software.

## Conclusions

Analyzing soil profiles with an automatic soil classification tool makes it easier to identify errors in data or classification of soil profiles and allows more reliable data curation. Additionally, the system can identify areas for improvement in the SiBCS.

## References

DOS SANTOS, H. G. et al. **Sistema brasileiro de classificação de solos**. Brasília, DF: Embrapa, 2018.

FAO. **Global soil partnership.** Disponível em: <www.fao.org/global-soil-partnership>. Acesso em: 23 set. 2021.

POLIDORO, J. C. et al. **Programa Nacional de Solos do Brasil (PronaSolos)**. Rio de Janeiro: Embrapa Solos, 2016.

VAZ, G. J. et al. Uma API para a classificação de solos do Brasil. In: CONGRESSO BRASILEIRO DE AGROINFORMÁTICA, 12., 2019, Indaiatuba. **Anais...** Ponta Grossa: SBIAGRO, 2019. p. 63-72.

# CONVERSION IN CULTIVATION AREAS UNDER ENVIRONMENTAL CHANGES IN CARBON AGGREGATE IN SOILS OF THE AMAZON ENVIRONMENT

**Silva, Joálison de Brito[1]; Souza, Fernando Gomes de[2]; Brito Filho, Elilson Gomes de [3]; Pinheiro, Elyenayra Nogueira[4]; Cunha, José Maurício da[5] & Campos, Milton César Costa[6]**

[1]Graduating in Agronomy from the Federal University of Paraíba (UFPB). E-mail: joalisonbrito2018@gmail.com
[2]Professor of Agronomy at the Federal University of Roraima (UFRR). E-mail: fernando.souza@ufrr.br
[3]Graduating in Agronomy from the Federal University of Amazonas (UFAM). E-mail: bfsambiente@gmail.com
[4]Master's Student in Environmental Sciences at the University of Amazonas (UFAM). E-mail: elyeng.7@gmail.com
[5]Professor at the Federal University of Amazonas (UFAM). E-mail: maujmc@gmail.com
[6]Professor of the Department of Soils and Rural Engineering (UFPB). E-mail: mcesarsolos@gmail.com

**Abstract**
Over the years, the process of converting natural ecosystems into cultivated areas, as well as land use and management systems, has provided changes in aggregate stability and in soil carbon stock levels. The present work aimed to evaluate the impact of native forest conversion into cropland and its impacts on carbon aggregation and storage in southern western Amazonia. Meshes with dimensions of 90 x 70 m, 90 x 56 m and 54 x 42 m were established, where 80 points were demarcated with regular spacing of 10 x 10 m, 10 x 8 m and 6 x 6 m, for the forest and guarana, urucum, cupuaçu, respectively. At each sampling point, clod and ring samples were collected to determine physical and chemical attributes, at three depths (0.00-0.05; 0.05-0.10; and 0.10-0.20 m), totaling 240 samples/area. The conversion influenced the soil aggregation state, evaluated by the increase of clay fraction dispersion and decrease of DMP, DMG values and aggregate classes > 2.00 mm; the attributes were at the limit of the degree of spatial dependence, ranging from moderate to strong; in the multivariate analysis, the forest and cupuaçu areas showed similar behavior, with values of CO, EC and aggregate classes 2 – 1 mm, < 1 mm and IEA, above the average.

**Keywords:** Amazonian soils, land use and management, environmental impacts.

**Introduction**

Currently, the occupation and replacement of previously forested areas by agricultural areas without proper knowledge and non-compliance with technical criteria, has been one of the main problems in the Amazon region. In this sense, a conversion of natural environments into agricultural systems, especially monoculture systems, has caused changes in the soil (FREITAS et al., 2015). Among the soil properties that change the most due to use and management, the structure stands out, associated with the formation of compacted layers with a decrease in macropores, aggregate size, water infiltration rate and increased resistance to penetration of the root system and density (COSTA et al., 2015). Studies have also shown that soil carbon stocks have been significantly affected by land use and management systems (CUNHA et al., 2017). So, intensive cultivation, combined with a high turnover rate, is responsible for reducing the content of organic matter in the soil, which is one of the main agents for the formation and stabilization of aggregates (CASTRO FILHO et al., 1998). Quantifying changes in the stability of soil aggregates can provide results that support agricultural production on a more sustainable basis. In order to optimize crop productivity, attention and maintenance of good aggregation, stability and, consequently, good structure are necessary (OLIVEIRA et al., 2013). Given the above, the stability of aggregates, as a physical

attribute, becomes fundamental in the assessment of structural quality, as it is sensitive to variations in land use, and can identify possible changes promoted by the conversion of the forest into agricultural areas. Thus, this study aimed to evaluate soil carbon aggregation and stock in areas undergoing forest conversion to different cropping systems, using traditional univariate, multivariate and geostatistical statistical techniques.

## Metodology

The study was carried out in two rural properties that are part of the São Francisco Settlement located in the municipality of Canutama, Amazonas, Brazil. Four areas were selected, being three areas under different crops: Urucum (Bixa orellana L.), Cupuaçu (Theobroma grandiflorum (Willd. Ex. Spreng) Schum), Guaraná (Paullinia cupana (Mart.) Ducke) and more forest area (Figure 1).
Knits were made according to the dimensions of the crop. In the areas of guaraná and forest, 90 x 70 m meshes were evaluated with regular spacing between the sampling points of 10 x 10 m, in the annatto area the chosen mesh was 90 x 56 m with spacing between the sampling points of 10 x 8 m , for the cupuaçu area, the mesh presented dimensions of 54 x 42 m, with regular spacing between the sampling points of 6 x 6 m. The were collected at the crossing points of the meshes, at depths of 0.00-0.05; 0.05-0.10; and 0.10-0.20 m, with 80 sampling points in each area, and totaling 240 per area. The points were georeferenced with a Garmin Etrex model GPS equipment (Datum South American´69).
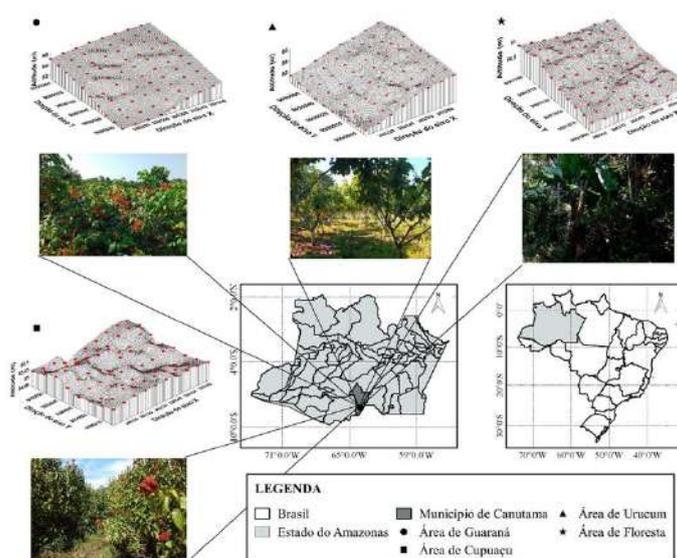


**Figure 1.** Location and digital elevation model of areas with guaraná, cupuaçu, urucum and forest, in the municipality of Canutama, southern Amazonas - AM.

From the soil collected and taken to the laboratory provided by aggregate stability by the method of Kemper and Chepil (1965) and then determined the gravimetric mean diameter (DMG), weighted mean diameter (DMP) and the aggregate stability index (IEA) . The degree of flocculation (GF), degree of dispersion (GD) and wet organic carbon (CO) was performed by the method of Teixeira et al. (2017), after the carbon analysis, the carbon stock (EC) was corrected. After obtaining data, univariate analysis of variance, geostatistics and multivariate analysis were included.

## Results and discussion

The analysis of variance (ANOVA) for the attributes evaluated in areas cultivated with guarana, annatto and cupuaçu in comparison with the forest area are presented in table 1 for the respective layers 0.00-0.05 m, 0.05-0 .10 m and 0.10-0.20 m.

**Table 1**. Mean test of soil attributes in the layers of 0.00-0.05 m, 0.05-0.10 m and 0.10-0.0.20 m for areas with different uses in southern Amazonas - AM.

| Descriptive statistics | CO | EC | DMG | DMP | Classes % | | | IEA | GF | GD |
|---|---|---|---|---|---|---|---|---|---|---|
| | $g\ kg^{-1}$ | $t\ ha^{-1}$ | mm | | >2,00 | 2,0-1,0 | <1,00 | | % | |
| **Layer (0,00-0,05 m)** | | | | | | | | | | |
| **Guaraná** | | | | | | | | | | |
| Average | 14,04d | 7,99b | 2,79a | 3,19a | 94,38a | 0,82bc | 4,80b | 93,92b | 78,93a | 21,07a |
| **Urucum** | | | | | | | | | | |
| Average | 16,52c | 8,79b | 2,76a | 3,18a | 94,77a | 0,70c | 4,53b | 92,87b | 76,60ab | 23,40a |
| **Cupuaçu** | | | | | | | | | | |
| Average | 23,42a | 11,18a | 2,53b | 3,08b | 91,17b | 1,15a | 7,68a | 91,12c | 68,55c | 31,45a |
| **Forest** | | | | | | | | | | |
| Average | 20,22b | 8,66b | 2,82a | 3,20a | 94,52a | 0,98ab | 4,51b | 95,64a | 72,91a | 27,09a |
| **Layer (0,05-0,10 m)** | | | | | | | | | | |
| **Guaraná** | | | | | | | | | | |
| Average | 12,11d | 6,63c | 2,44c | 2,96bc | 86,17b | 2,77a | 11,06a | 91,12b | 69,42a | 30,58a |
| **Urucum** | | | | | | | | | | |
| Average | 16,16b | 8,62b | 2,81 | 3,16a | 93,63a | 1,39b | 4,98c | 94,64a | 59,27a | 40,73a |
| **Cupuaçu** | | | | | | | | | | |
| Average | 23,30a | 13,38a | 2,59b | 3,04b | 89,17b | 3,49a | 7,35b | 93,74a | 57,78a | 42,22a |
| **Forest** | | | | | | | | | | |
| Average | 13,66c | 6,53c | 2,52bc | 2,94c | **87,61b** | 3,58a | 8,81b | 93,67a | 66,14a | 33,86a |
| **Layer (0,10-0,20 m)** | | | | | | | | | | |
| **Guaraná** | | | | | | | | | | |
| Average | 10,35b | 11,13b | 2,09b | 2,72bc | 77,54b | 5,33b | 17,13ab | 89,47c | 54,27ab | 45,73ab |
| **Urucum** | | | | | | | | | | |
| Average | 7,87c | 8,08d | 2,62a | 3,03a | 89,42a | 2,76c | 7,82c | 92,83a | 51,99ab | 48,01ab |
| **Cupuaçu** | | | | | | | | | | |
| Average | 8,05c | 9,40c | 2,07b | 2,62c | 73,95c | 7,89a | 18,16a | 90,92bc | 41,41b | 58,58a |
| **Forest** | | | | | | | | | | |
| Average | 12,70a | 13,32a | 2,15b | 2,75b | 80,45b | 5,47b | 14,09b | 91,86b | 61,44a | 38,66b |

CO: organic carbon; EC: carbon stock; DMG: geometric mean diameter; DMP: weighted average diameter; IEA: Aggregate Stability Index; GF: Degree of Flocculation; GD: Degree of Dispersion. Means followed by the same lowercase letter in the column do not differ by Tukey's test ($p < 0.05$).

By analyzing the results of CO and EC, it was possible to observe that there were differences between all areas analyzed, with higher values found in the area under cupuaçu cultivation in the 0.00-0.05 m and 0.05-0.10 m layers, with the exception of the 0.10-0.20 m layer, which the forest area presented the highest value.

The high values of CO and EC in the cupuaçu area, respectively in the 0.00-0.05 m and 0.05-0.10 m layers must be associated with the high biomass production, provided by the input of

residues from the culture itself, a fact that can also be attributed to the time of 7 years of cultivation with the crop without undergoing intensive practices in the soil. The respective results found for CO and EC corroborate those found by Cunha et al. (2017), which they attributed to the highs attributed to the vegetation cover verified in loco under the use of the Guandu bean in comparison with an area of native forest and pasture.

It is possible to observe an increase in aggregate values with an increase in organic carbon. This situation was proven in the studies of CAMPOS et al., (2016), where they highlighted a positive correlation between the distribution of aggregates with organic carbon, observing a percentage increase. Regarding the degree of dispersion (GD), it was possible to observe that the areas studied were not differentiated from each other by the Tukey test at all depths analyzed. The only exception was observed for the forest area at a depth of 0.10-0.20 m, which presented a lower degree of dispersion in relation to the areas of guaraná, annatto and cupuaçu.

Based on these results, studies highlight that normally soils with high flocculation (GF) and low dispersion (GD) are related to better soil physical conditions (VASCONCELOS et al., 2013), in addition to mentioning that both attributes are inversely proportional (VICENTE et al., 2012). Under analysis of the degree of spatial dependence (GDE) from geostatistical data, expressed by the ratio between the nugget effect and the plateau, following the classification by Cambardella et al. (1994), it was observed that the attributes differ in the limits of the degree of spatial dependence (GDE), ranging from moderate to strong dependence. Corroborating those found by Alho et al. (2014), they evaluated the spatial variability of aggregate stability and carbon stock in Cambisol and Argisol in Amazonas. In line with this, it was possible to observe higher GDE for GDM in the annatto area with 58% and 63.1% forest, respectively, in the 0.00-0.05 layer. In the 0.05-0.10 m layer, the highest DEG was observed for the attributes silt and IEA, with 53.8% and 54.5%, respectively, in the area of guaraná and annatto. At depth 0.10-0.20 m, the highest values occurred for the DMP attributes and aggregate classes> 2 mm, both in the guaraná area.

Figure 2 illustrates the distribution of scores in the different areas studied and the disposition of factor loadings of the soil attributes formed by PC1 and PC2
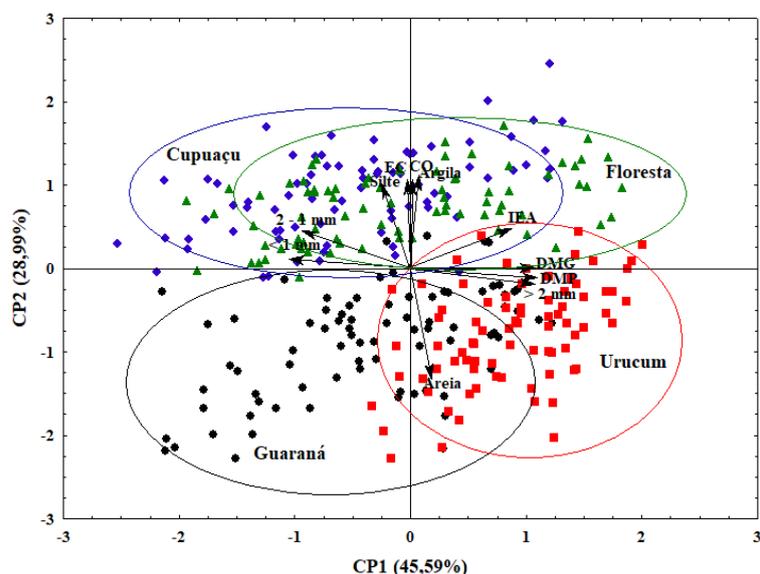


**Figure 2.** Analysis of the main components of the soil attributes studied at depth from 0.00 to 0.20 m, in an area of guaraná, urucum, cupuaçu and forest in the municipality of Canutama, AM.

It is possible to observe greater densification of the forest and cupuaçu scores in the first and second quadrants, which shows that both areas obtained values for the attributes CO, EC, silt, clay and aggregate classes 2 - 1 mm, <1mm and IEA, above of the media. On the other hand, an area cultivated with annatto was more distributed in the fourth quadrant, with attributes more focused on characteristic characteristics, the same values above the average, such structuring

condition may be related to the time of cultivation in the processing area. The opposite occurred in the area cultivated with guaraná, where values below the average were observed for the attributes that are related to soil conditions, such as DMG, DMP and aggregate classes > 2 mm.

## Conclusions

Areas cultivated with cupuaçu, after the conversion process, may present the same content of CO and EC or even surpass as areas of native forests. The conversion process influences the soil aggregation state, evaluated by the increase of clay fraction dispersion and decrease of DMP, DMG values and aggregate classes> 2 mm. The attributes defined in the studied areas define the limits of the degree of dependence, using the variation between moderate and strong. In the multivariate analysis, as forest and cupuaçu areas complement similarity, with values of CO, EC, silt, clay and aggregate classes 2 - 1 mm, < 1 mm and IEA, above the average.

## References

CAMBARDELLA, C.A. et al. Field-scale variability of soil properties in Central Iowa. **Soil Science Society of America Journal**, vol. 58, no. 5, p. 1501-1.1994.

CAMPOS, M.C.C. et al. Soil carbon stocks and Cambisol aggregates under different managements in southern Amazonas. **Environment and Water Magazine**, v. 11, no. 2, p. 339-348, 2016.

CASTRO FILHO, C. et al. Aggregate stability and its relationship with organic carbon content in a Dystrophic Purple Latosol, as a function of planting systems, crop rotations and methods of preparation. **Brazilian Journal of Soil Science**, vol. 22, no. 3, p. 527-538, 1998.

COSTA, N.R. et al. Soil attributes and carbon accumulation in crop-livestock integration in a No-Till System. **Brazilian Journal of Soil Science**, vol. 39, no. 3, p. 852-863, 2015.

CUNHA, J.M. et al. Physical attributes and soil carbon stocks in the Archaeological Preta Preta areas of the Amazon. **Environment & Water Magazine**, v. 12, no. 2, p. 263-281, 2017.

FREITAS, L. et al. Multivariate techniques for evaluating the attributes of a Red Oxisol under different managements. **Brazilian Journal of Agrarian Sciences / Brazilian Journal of Agrarian Sciences**, vol. 10, no. 1, p. 17-26. 2015

OLIVEIRA, P.R. et al. Structural quality of an Oxisol subjected to compaction. **Brazilian Journal of Soil Science** 37 (3): 604-612, 2013.

TEIXEIRA, P.C. et al. **Soil analysis methods manual**. 3rd ed. revised and updated. EMBRAPA, Brasilia. 573p, 2017.

VASCONCELOS, R.R.A. et al. Physical characteristics of saline-sodic soils in the semiarid region of Pernambuco as a function of different levels of gypsum. **Brazilian Journal of Agricultural and Environmental Engineering**, v. 17, no. 12, p. 1318–1325, 2013.

VICENTE, T.F.S. et al. Relationships of soil attributes and aggregate stability in sugarcane fields with and without vinasse. **Brazilian Journal of Agricultural and Environmental Engineering**, v. 16 n. 11, p. 1215–1222, 2012.

# Comparison between kriging and multiple linear regression maps of soil penetration resistance

SILVA JÚNIOR, João Fernandes da[1]; COSTA, Silmara Pereira[1]; MORAES, Irwing Jordan Almeida de[1]; MAGALHÃES, Deila da Silva[1]; PINHEIRO, Daniel Pereira[1]; GALVÃO, Jessivaldo Rodrigues[1]

[1] Federal Rural University of Amazon, email: joao.fernandes@ufra.edu.com.br

**Thematic Session: Pedometria: Inovação nos Trópicos**

**Abstract**
The mapping of soil attributes such as resistance to penetration is essential to assess soil compaction. The objective of the present study was to assess the similarity between the multiple linear regressions (MLR) and ordinary kriging (OK) to estimate the variability of soil penetration resistance (PR). The study was carried out at Fazenda Freitas, São Francisco of Pará city, in the Para State. The samples were georeferenced, totaling 71 points at 0.00–0.10m of depth, and we measured the PR in these places. We assessed the spatial pattern by geostatistical analysis, and after was interpolated by OK. We used a dataset with 15 relief covariates for MLR using the dependent variable: PR. We measured the correlation between the estimative from OK and MLR. We selected eleven relief covariates for the regression model with R2=0.70. The PR map estimated by MRL obtained 70% similarity with the map obtained by OK. MLR is an alternative to estimate PR when it is impossible to use geostatistical modeling for similar environments.

Keywords: Pedometrics; Geostatistics; Soil mapping; Pedotransfer.

## Introduction

The mapping of soil class and attributes on a detailed scale is still a major challenge for sustainable management. In digital mapping of soil attributes, it is common to use geostatistical techniques to create iso-values maps. However, this requires a large number of soil samples to be able to find spatial dependence structure and make interpolations. This makes this methodology costly, sometimes unfeasible, especially in small areas. An alternative is the use of remote sensing data, machine learning, relief attributes that are important covariates to estimate part of these soil attributes (MacBratney et al., 2003).

Another economical and cheap alternative is the use of digital soil mapping using relief factors (Menezes et al., 2014). These relief covariates are mostly correlated with most of the soil attributes (Silva Júnior et al., 2012). Based on the hypothesis that can be used as attribute predictors of soil penetration resistance using multiple linear regression analyses. The objective proposed in this work is to evaluate the similarity between the multiple linear regressions with the results of the penetration resistance map estimated by ordinary kriging.

## Methodology

The study was carried out at Fazenda Freitas, in the municipality of São Francisco do Pará, Northeast Pará. The soil is medium-textured Yellow Latosol

(Oxisol) (Santos et al., 2018). The study area was georeferenced with GPS geodesic to survey the coordinates of the limits of the sampling grid, totaling 71 sampling points in a regular grid with a spacing of 22m. Was measured the soil resistance to penetration (RP) using an impact penetrometer on the surface of the layer of 0.00-0.10m. Initially, we analyzed the data in an exploratory way to verify the presence of outliers and the need for transformation for geostatistical analysis.

We measured the spatial dependence based on the assumption of the intrinsic hypothesis, which was performed from the experimental semivariograms.

$$\hat{y}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2 \tag{1}$$

Where: $N(h)$ represents the number of experimental pairs;
$(h)$ is the regular interval that separates $Z(x_i)$ and $Z(x_i+h)$;
$Z(x_i)$ is the value of a variable $Z$ in position $x_i$; $Z(x_i+h)$ is the value of a variable $Z$ separated by a distance $h$ of position $x_i$.

The adjustment of the mathematical model is performed for a graphical representation of $\hat{y}(h)$ and h (Matheron, 1963).

The ordinary kriging was used for making spatial distribution maps for soil resistance to penetration, using software Surfer – 14 (Golden Software inc., 1999),

In our work, we used the method Stepwise Regression. The stepwise regression method is a combination of forwarding selection and backward elimination. We used the following variables in input: 15 independent variables (Digital Elevation Model (DEM), Analytical Hillshading (AH), Aspect (ASP), Cross-Sectional Curvature (CSC), Longitudinal Curvature (LC), Convergence Index (CI), Closed Depressions (CD), Flow Accumulation (FA), Topographic Wetness Index(TWI), Slope, LS Factor (LSF), Channel Network Base Level (CNBL), Vertical Distance to Channel Network (VDCN), Valley Depth (VD), Relative Slope Position (RSP), to estimate the penetration resistance of soil from ordinary kriging (PR_OK). We measured the coefficient of determination between the raster data from OK and MLR to assess the similarity.

**Results and discussion**

The semivariogram analysis showed the existence of spatial dependence for PR in the layer 0,00-0,10m. The mathematical model adjusted to the experimental semivariogram was spherical, with range: 203.37 m, sill:0.02, nugget:0.01, and degree of spatial dependence of moderate (Table I).

Table I. Parameters for fitting the theoretical model to the experimental semivariogram of soil penetration resistance.

| Soil property | Model | Target | | | Nugget/sill[1] | spatialstructure[2] |
| | | Range (m) | Partial sill | Nugget | | |
| --- | --- | --- | --- | --- | --- | --- |
| PR | Spherical | 203.37 | 0.02 | 0.01 | 0,50 | Medium |

[1]Calculated from the target data set; [2]Values < 0.25 being strong, 0.25-0.75 being medium, and > 0.75 being weak (Cambardella et al., 1994)

Table II presents the result of stepwise regression using SAGA GIS 2.3.2. In this study. were selected as controlled variables eleven relief variables. Other covariates tested in this study no selected indicates the smallest p-value than Alpha-

to-Enter (p= 0.000< 0.05). Therefore, the soil penetration resistance variable that enters into the model.

Table II. Summary of soil penetration resistance multiple linear regression model parameters

| Stepwise regression with variables | | | | |
|---|---|---|---|---|
| R multiple | 0.70 | | | |
| Coefficient of determination ($R^2$) | 0.70 | | | |
| $R^2$ adjusted | 0.70 | | | |
| Standard error of estimate | 0.03 | | | |
| Analysis of variance | | | | |
| | SS | df | MS | F | Sig. |
| Regression | 7.044 | 11 | 0.640 | 545.739 | 0.000 |
| Residue | 2.982 | 2.982 | 0.001 | | |
| Total | 10.026 | 2.993 | | | |

Covariates selected: Vertical Distance to Channel Network (VDCN), Slope (S), Channel Network Base Level (CNBL), Aspect (ASP), Flow Accumulation (FA), Convergence Index (CI), Longitudinal Curvature (LC), Valley Depth (VD), LS Factor (LSF), Topographic Wetness Index (TWI) and Relative Slope Position (RSP).

Based on the MLR coefficients, it was possible to elaborate the following equation to estimate the soil resistance to penetration, equation 2:

$$\hat{PR} = 1.804394 + 0.059068VDCN - 8.513399SLOPE - 0.060039CNBL - 0.008424ASP - 0.000023FA - 0.001869CI - 10.196843LC - 0.048566VD + 0.698430LSF - 0.022674TWI + 0.028259RSP$$

(2)

Forkuor et al. (2017) also managed to estimate soil attributes using MRL and remote sensing variables and concluded that for sand and clay MLR has offered a better predictive ability. The RP values of the KO and MRL maps ranged between 0.56 and 0.84 Mpa. These are within the acceptable range for the plants, established in the depths of 0.00–0.10 m (Figure I). According to Tormena et al. (1998), it was determined that 2.00 Mpa is the limiting level for root growth, and depending on the species, it can reach up to 4.00 Mpa (IMHOFF et al., 2000).
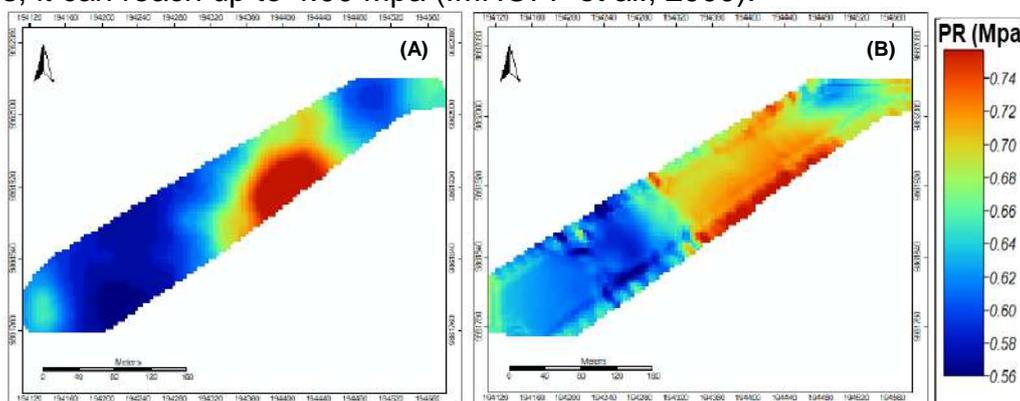


Figure I, Maps of the spatial distribution of soil resistance to penetration by KO (A) and by MRL (B).

Using this equation (2) and specializing the estimated values, we calculated the correlation between this map estimated by MLR and kriging and found an R2 de 0.70, as shown in figure II.

These results showed a 70% similarity between the maps. Our finding showed that as is possible to use the MRL as a quick assessment tool and non-expensive in

II Pedometrics

Apoio: FAPERJ CAPES  Promoção: Sociedade Brasileira de Ciência do Solo  Organização: UFRRJ Embrapa Solos  Nov, 24th to 27th, 2021 Brazil

caparison with kriging. Due to applying geostatistics, to use data samples a lot to find the spatial dependence structure.
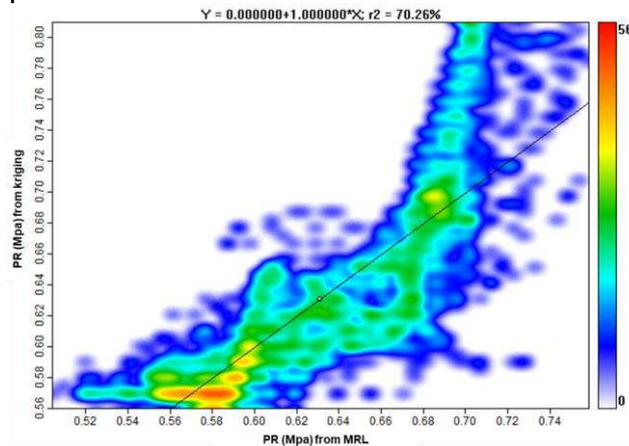


Figure II, Scatter plot between results from OK and MLR and coefficient of determination ($R^2$).

## Conclusions

The results from the final model of the multiple linear regressions by the stepwise method have 70% similarity of the values estimated by ordinary kriging. This shows you can have an alternative. The use of MLR is an alternative to estimate PR when it is impossible to use geostatistical modeling for similar environments. MLR is a quick assessment tool and non-expensive in a relationship with ordinary kriging.

## References

CAMBARDELLA, C. A. et al. Field-scale variability of soil properties in central Iowa soils. **Soil Science Society of America Journal**, v.58, p.1501-1511, 1994.

FORKUOR. G. et al. High-Resolution Mapping of Soil Properties Using Remote Sensing Variables in SouthWestern Burkina Faso: A Comparison of Machine Learning and Multiple Linear Regression Models, **PLOS ONE**, v.23 2017

GOLDEN SOFTWARE. **Surfer for Windows**: Surface mapping system. Versão 6.01. Golden, 1995.

GS+. GS+ **Geostatistical for environmental science**. Versão 10.0, Michigan: Gamma Design Software, 2014.

IMHOFF, S. et al.: Aplicações da curva de resistência no controle da qualidade física de um solo sob pastagem. **Pesquisa Agropecuária Brasileira**, Brasília, v.35, n. 7, p.1493-1500, 2000.

MATHERON, G. Principles of geostatistics. **Economic Geology**, v.58, p. 1246-1266, 1963.

MENEZES, M. D. et al. Solum depth spatial prediction comparing conventional with knowledge-based digital. Scientia Agricola, v. 71, n. 4, p. 316-323, 2014.

MCBRATNEY, A.B. et al. On digital soil mapping. **Geoderma**. 117, 3–52, 2003.

# Particle size distribution in a semiarid hydromorphic soil profile using geostatistics

LIMA NETO, Miguel Alvores[1]; MARQUES, Rabech Grasiely Gomes[2;] VALLADARES, Gustavo Sousa[3]; FREITAS, Aline Gonçalves de[4].

[1] UFPI, miguel3limaalvores@gmail.com; [2] UFPI, rabechgrasiely1998@hotmail.com; [3] UFPI, valladares@ufpi.edu.br; [4] UFPI, tuttyfreitas@gmail.com

**Thematic Session: Pedometrics: Innovations in Tropics**

## Abstract

The present work aimed to evaluate the distribution of granulometric fractions in depth in a hydromorphic soil profile by semivariograms interpretation. The study area is in the municipality of Várzea Branca-PI, more specifically in the archaeological site Lagoa Grande das Queimadas Piauí State, Northeastern Brazil. For the particle size analysis, the pipette and sieving method was used. Subsequently, the data were processed through kriging, which proved to be great tools for understanding the spatial distribution in depth.

Keywords: Geostatistics; Pedology; Granulometry; Archaeological record

## Introduction

The knowledge of the physical properties of soils related to particle size distribution are fundamental for understanding the active processes, potentials and weaknesses of any soil profile. So that there is knowledge of the soil horizons, collections in soil profiles and subsequent particle size analysis are fundamental.

In this sense, it is important to perform the spatialization of the results found, aiming at the spatial understanding of the distributions and acting processes. There are several ways to carry out this step, however, due to the difficulty of performing numerous collections, there is a need for methods that estimate values for non-sampled points, and the semivariograms presented here are a way of checking the quality of the estimation, given the mentioned, the present work aimed to evaluate the distribution of particle size fractions in depth by ordinary kriging.

## Methodology

The study area comprises Lagoa Grande das Queimadas located in Várzea Branca in the southwest of Piauí State. The region in which the lagoon is located presents minimum temperatures of 18 ºC and maximum of 36 ºC, with a semi-arid, hot and dry climate. The vegetation is shrub-arboreal caatinga (CPRM, 2004).

The regional geology is composed of granites and schists from the Pre-Cambrian Sobradinho-Remanso Complex and tertiary-quaternary dendrite-laterite deposits. The depressed shaped site has Holocene sediments. The municipality of Várzea

Branca is located in the Canindé-Piauí Hydrographic Sub-Basin, whose main watercourse is the Piauí River. In the study area, a Haplic Gleysol occurs.

The analyzed samples were collected at a depth of (0-160 cm, in 31 layers) in Lagoa das Queimadas, with the first 3 layers having a thickness of 10 cm and the other 5 cm. For the particle size analysis, the methodology present in the Embrapa Soil Methods and Analysis Manual (TEIXEIRA et al., 2017) was used, using the pipette and sieving method, and the clay, silt and sand contents were determined.

After this procedure, the data were tabulated and descriptive statistical analysis performed. To understand its spatial dependence, data were analyzed using semivariograms and kriging (VIEIRA et al, 2020). The depth of each layer was used as a geographic coordinate, starting from the surface where the value zero was assigned.

**Results and discussion**

The analysis of the data referring to the granulometry of the soils through descriptive statistics showed average values for sand of 149.07, for silt 501.85 and for clay 348.52 g kg[-1]. With variance 14948.54; 7519.79; 30427.88, respectively for sand, silt and clay, the standard deviation was smaller for the silt variable 86.72, indicating less variability in the silt values. The results indicate a predominance of silt and clay in the studied soil, and low degree of weathering.

The coefficients of variation presented values of 19.8% for the silt variable, 33.0% for clay and 47.4% for sand, representing moderate to high variability in the studied profile.

In addition to descriptive statistics, the data were analyzed using geostatistics by ordinary kriging, with the adjustment of semivariograms (Figure 1).

The semivariogram for the sand fraction showed the nugget effect 0 (C0 = 0), the plateau value (C0+C1), coincided with the structural variance (C1), both being 5242.6, the range was 220 cm, the best model to represent the data of this variable was the circular.

As for the silt variable, the semivariogram showed a nugget effect in the value of (C0=1308.2), structural variance was (C1= 13964), plateau (C0+C1= 15272) range 88.62 cm. The model that best fit the data was the Gaussian. For clay, as well as the sand fraction, the nugget effect was zero (C0= 0), therefore structural and level variance had identical values being (C0+C1=45219), (C1=45219). The range was 180cm, the semivariogram model that best represented the data was spherical.
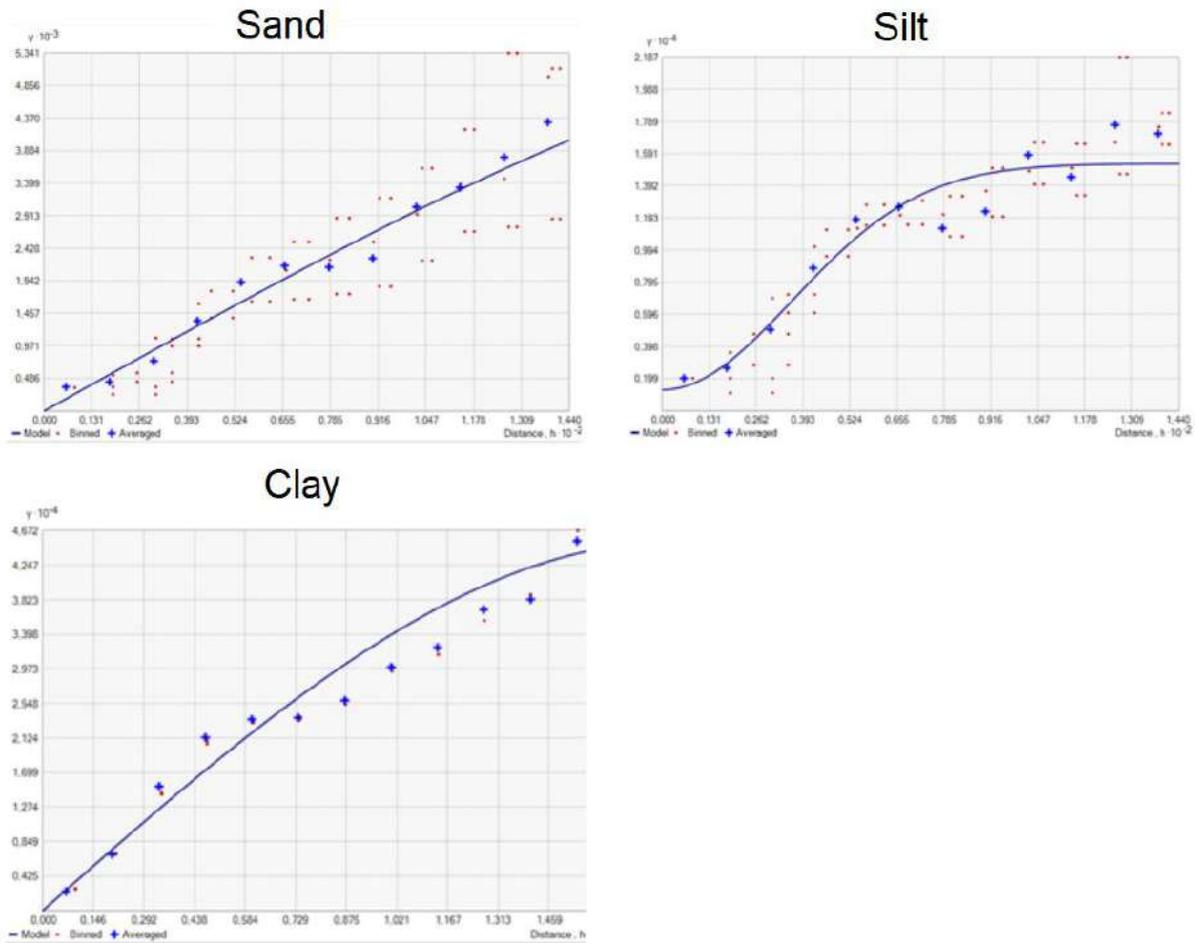
**Figure 1-** Semivariograms of soil particle size fractions.

Assessing the granulometry of the soils, it is possible to see lithological discontinuity occurring around 110 cm in depth. Since the sand and silt contents increase and the clay contents decrease significantly (Figure 2).
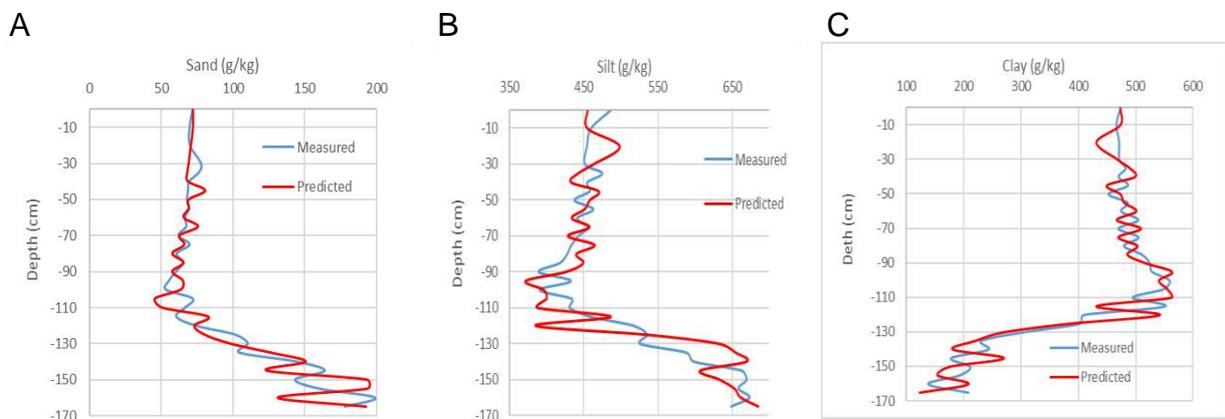


**Figure 2-** Depth distribution of sand (A), silt (B) and clay (C) contents observed and estimated by kriging.

## Conclusions

The different semivariogram models are great tools for understanding the spatial distribution of data, emphasizing the need for adjustment and choice of appropriate models according to the specificity of the data. The results indicated lithological discontinuity at about 110cm deep in the soil profile.

## References

CPRM. **Projeto cadastro de fontes de abastecimento por água subterrânea, estado do Piauí: diagnóstico do município de Várzea Branca**. - Fortaleza, 2004. Disponível em: http://rigeo.cprm.gov.br/xmlui/bitstream/handle/doc/16459/Rel_VarzeaBranca.pdf?sequence=1 Acesso em : 27. maio. 2020.

TEIXEIRA, P. C.; DONAGEMA, G. K.; FONTANA, A.; TEIXEIRA, W. G. (Ed. Técnicos**). Manual de métodos de análise de solo**, 3. ed. rev. e ampl. Brasília: Embrapa, 575p., 2017.

VIEIRA, S. R. Geoestatística em estudos de variabilidade espacial do solo. In: NOVAIS, R. F.; ALVARES, V. H.; SCHAEFER, C. E. G. R. (Ed.). **Tópicos em ciência do solo**. Viçosa, MG: Sociedade Brasileira de Ciência do Solo, 2000.

# New alternatives for collecting soil data as input of local DSM approaches

*Philippe Lagacherie,  INRAE, UMR LISAH, France*

In spite of their fine resolution, the Digital Soil Mapping (DSM) products that are now available at global and, in some countries, at national scale do not provide accurate representations of local soil patterns as required by the end-users acting at local level, e.g municipalities or watersheds, …(Rossiter et al, 2021). Several recent experiments (Somarathna et al, 2017 ; Lagacherie et al, 2020, Loiseau et al, 2021) revealed that better performances of DSM could be obtained by significantly increasing the spatial density of the sites with measurements of soil properties that are used for calibrating the (machine learning) DSM Models. However, such an increase often represents an inaccessible investment in soil information, especially for a local user. Alternatives to the collection of exact soil measurements exist and should be considered as surrogate data for enriching at acceptable costs the soil inputs used for calibrating the DSM models.

The sources of surrogate soil data are multiple: legacy data obtained by automated entry procedures, proximal and/or remote sensing estimations of soil properties (Vis-NIR spectrometry, electromagnetic induction, remote sensing, gamma-ray spectroscopy, ….), plant-based soil functional indicators (Thermal Infrared estimations of ETR, delta $C_{13}$) or qualitative soil observations provided either by experienced soil surveyors, farmers or even citizens. Some recent examples of collection of such surrogate soil data will be presented in the talk. We will also present the specific DSM approaches that have been developed and tested for using such surrogate data in synergy with the "classical" soil measurements to improve the DSM products (Zare et al, 2021; Styc et al, 2021).

## References

Lagacherie, P., Arrouays, D., Bourennane, H., Gomez, C., Nkuba-kasanda, L., 2020.  Analyzing the impact of soil spatial sampling on the performances of Digital Soil Mapping models and their evaluation : A numerical experiment on Quantile Random Forest using clay contents obtained from Vis-NIR-SWIR hyperspectral imagery. Geoderma 375.

Loiseau, T., Arrouays, D., Richer-de-forges, A.C., Lagacherie, P., Ducommun, C., Minasny, B., 2021. Density of soil observations in digital soil mapping : A study in the Mayenne region , France. Geoderma Reg. 24, e00358.

Somarathna, P.D.S.N., Minasny, B., Malone, B.P., 2017. More Data or a Better Model? Figuring what Matters Most for the Spatial Prediction of Soil Carbon. Soil Sci. Soc. Am. J. 81, 1413–1426.

Rossiter, David G; Beaudette, Dylan ; Libohova,Zamir; Poggio, Laura, 2021. How well does Predictive Soil Mapping represent soil geography? An investigation from the USA. Case Studies. Report 2021/03, ISRIC – World Soil Information, Wageningen (doi:10.17027/isricwdc-202103)

Styc, Q., Gontard, F., Lagacherie, P., 2021. Harvesting spatially dense legacy soil datasets for digital soil mapping of available water capacity in Southern France. Geoderma Reg. 24, e00353.

Zare, S., Abtahi, A., Rashid, S., Shamsi, F., Lagacherie, P., 2021. Combining laboratory measurements and proximal soil sensing data in digital soil mapping approaches. Catena 207, 105702.

# Use of variables derived from the Digital Elevation Model for the sustainable planning of soils in Bom Jardim – RJ

AGUIAR, Francisco Ivo dos Santos[1]; PINHEIRO, Helena Saraiva Koenow[1]; CARVALHO JUNIOR, Waldir de[2]

[1] Federal Rural University of Rio de Janeiro - UFRRJ, ivoaguiar222@hotmail.com; lenask@gmail.com; [2] Empresa Brasileira de Pesquisa Agropecuária, waldir.carvalho@embrapa.br

## Thematic Session: Legacy Data: How Turn It Useful?

## Abstract

The research goal is to analyze terrain attributes obtained from a digital elevation model (DEM) to represent soil forming factors used in digital soil mapping at Bom Jardim county, Rio de Janeiro state, and also to delimit arable areas. The DEM was generated to represent the altimetry of the study area, and subsequently, the basic attributes of the terrain were extracted: slope, slope height, aspect, valley depth index, topographic moisture index and factor - LS. Soil properties were correlated with terrain attributes through a Spearman correlation method. It is concluded that, the terrain covariates derived from the digital elevation model can represent the soil forming factors and be used at the digital soil mapping of the area, as well as be used to establish agricultural suitability areas in mountainous areas, such as Bom Jardim – RJ.

Keywords: digital soil mapping, topographic parameters, hillslope area, soil conservation and security.

## Introduction

Advances in remote sensing have occurred quickly, especially regarding spatial and spectral resolutions, both of great interest for digital soil mapping (DSM) (TEN CATEN, et al., 2011). Thus, factors and processes of soil formation can be modeled from variables measured by remote sensing and geoprocessing techniques. Therefore, it is essential to create a database with environmental variables that most influence the variability of soils in order to further develop digital soil mapping of soil attributes and classes in a interest area (CALDERANO FILHO et al., 2009).

Orbital remote sensing data and terrain attributes derived from a digital elevation models (DEM), have been used to understand the spatial and temporal relationships between soil classes, properties and different environmental variables (SANCHEZ et al., 2009). These attributes are commonly used as auxiliary variables in the spatial prediction of soil-landscape patterns and contribute to the improvement of the mapping of soil classes and properties, such as horizon thickness, elements in the soil solution, texture, color, moisture, among others (Gessler et al., 2000).

The objective of this work is to analyze terrain attributes derived from a digital elevation model (DEM), in Bom Jardim county, in order to select covariates that represent soil forming factors and also to delimit arable areas based.

**Methodology**

The study area corresponds to Bom Jardim county, Rio de Janeiro state, between the coordinates 22º 06' and 22º 18' S and 42º 12' and 42º 30' W. The total area have 385.04 km$^2$, with strong-wavy relief, altitudes ranging between 405 to 1,630m, and average slope of 38%. According to Calderano Filho et al. (2009), three main soil classes (Oxisols, Epodossolo and Argisol) occurs in the area. To compose the soil database, 209 soil profiles were used, collected and described according to SiBCS (SANTOS et al., 2005), in a total of 603 soil samples. From the entire dataset the following soil properties were selected to use in the present analysis: coarse sand, fine sand, clay, clay dispersed in water and organic carbon.

A digital elevation model (DEM) was generated to represent the altimetry of the study area, from the official digital cartographic base of the state of Rio de Janeiro - scale 1:25,000. The topographic database in vector format, contour lines with 10 meters of equidistance, quoted points and a hydrography were interpolated using the "TopotoRaster" tool to obtain a 20m resolution DEM. At the end of this procedure, the spurious depressions were corrected by using the "FillSink" tool, in order to make it hydrologically consistent.

Later, the DEM was imported into the System for Automated Geoscientific Analyzes (SAGA-GIS, 2020) program, the Terrain Analysis function was used to extract basic terrain attributes: slope, aspect, valley depth index, topographic wetness index, and factor - LS. the terrain attributes were associated with soil samples database, and the free software QGIS 3.10.4 was used to edit the maps and produce the legends.

**Results and discussion**

Analyzing the image (Figure 1A) it can be seen that the areas of the municipality present considerable variation in relief. Bom Jardim is a municipality of mountainous areas and fragile ecosystems.

Slope is one of the most important terrain attributes associated with pedogenetic processes, as it directly affects surface and subsurface water flow velocity and, consequently, soil water content, erosion/deposition potential and many other important exogenic processes. According to the map presented in Figure 1B the average declivity varies around 20.51 to 41.02º, characteristic of a heavily undulating to mountainous relief.

The Figure 1C present the LS factor where with higher LS Factor represent a region where there is loss of kinetic energy and the deposition of eroded material from higher

parts of the terrain takes place. The topographic moisture index (Figure 1D) describes a region's tendency to water storage. According to the parameters of Lin et al. (2006) the soils of Bom Jardim are well drained, but only in the highest parts of the landscape. These moisture conditions expressed in the map can also be associated with soil thickness, structure, density and permeability.

In Figure 1E we have the aspect map, this attribute has an influence on insolation, evapotranspiration, and the distribution and abundance of flora and fauna. On the map Figure 1F is represented the valley depth index, which describes how flat the bottom of a valley is. For mountainous regions these areas have great importance, since this particular region has a high susceptibility to erosion (CARVALHO JÚNIOR, et al., 2014), and a relevant role to agricutural production. This areas can be indicated both for silvopastoral activities in drier areas, and agroforestry activities at the humid areas.

The correlation matrix between soil properties and terrain attributes is presented in Figure 2. Coarse sand content and water-dispersed clay did not correlate with any terrain attribute, while fine sand content was correlated with topographic moisture index and DEM. The clay content was corrected with the valley depth index. In relation to organic carbon, only and correlated with DEM.
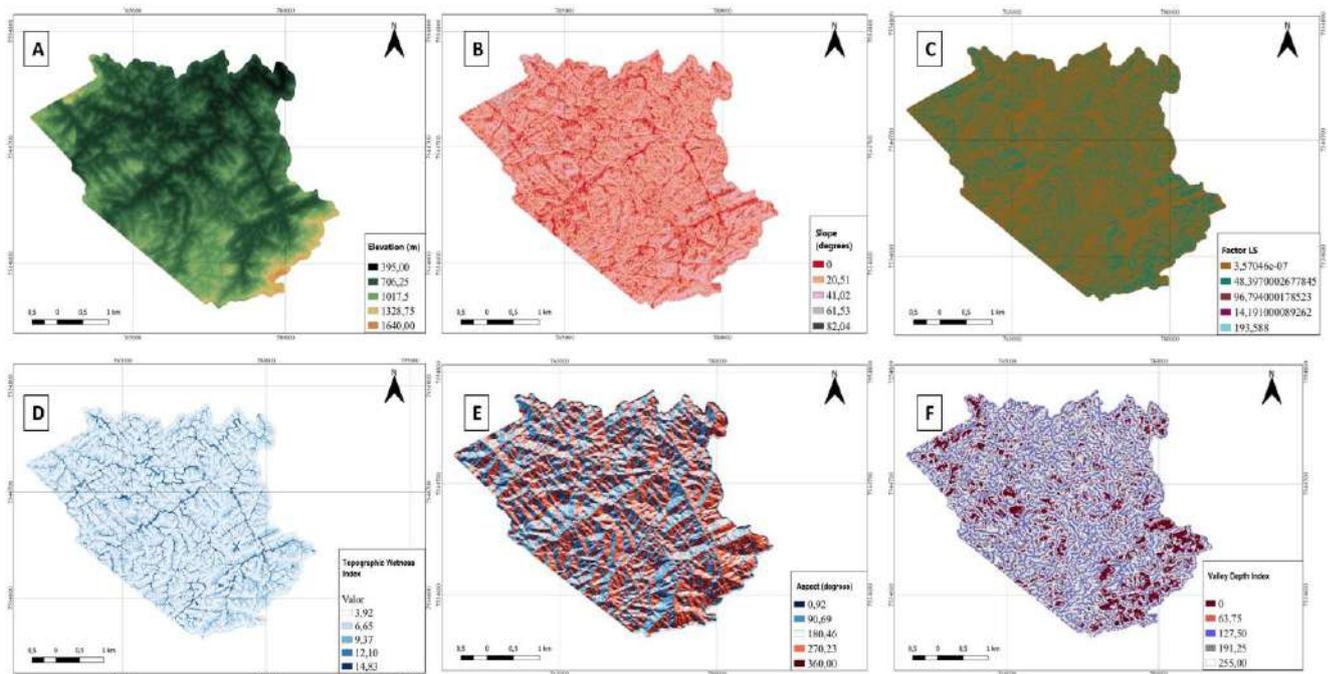


**Figure 1**. (A) Digital elevation model - DEM and digital model of environmental covariates, with a spatial resolution of 20 m, for the municipality of Bom Jardim – RJ. B (Slope); C (LS Factor); D (typographic moisture index); E (Aspect) and F (Valley depth index)
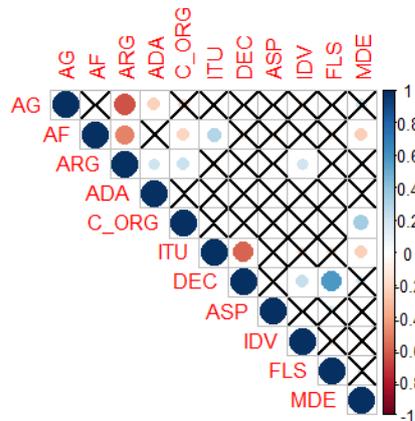
**Figure 2**. Correlation matrix between the properties of the soil surface layer with the environmental co-variables and terrain attributes. (AG = coarse sand; AF = fine sand; AGR = clay; ADA = clay dispersed in water; C_ORG = organic carbon; ITU = topographic wetness index; DEC = slope; ASP = aspect; IDV = valley depth index; MDE = digital elevation model.

## Conclusions

The environmental covariates derived from the digital elevation model can be used to represent soil forming factors, such as slope, LS factor and topographic wetness index. As well as the covariates aspect, topographic wetness index and valley depth can be used to establish agricultural suitability areas in mountainous areas, such as Bom Jardim county, in Rio de Janeiro state.

## References

CALDERANO FILHO, B. et al. Caracterização dos solos do município de Bom Jardim-RJ. **Anais**...Proceedings of XXXII Congresso Brasileiro de Ciências do Solo (Fortaleza SBCS CE). 2009.

CARVALHO JUNIOR, W. el al. A regional-scale assessment of digital mapping of soil attributes in a tropical hillslope environment. **Geoderma**, v. 232-234, p. 479–486, 2014

COMISSÃO DE FERTILIDADE DO SOLO DO ESTADO DE MINAS GERAIS. **Recomendações para o uso de corretivos e fertilizantes em Minas Gerais** – *5ª Aproximação.* Viçosa, MG, 1999. 359 p.

GESSLER, C. et al. Comparação de métodos para mapeamento digital de solos com utilização de sistema de informação geográfica. **Ciência Rural**, v. 40; p.2099-2106, 2001.

LIN,S. et al. Atmospheric correction of Landsat ETM+ land surface imagery. I. Methods. **IEEE Transactions on geoscience and remote sensing**, v. 39, n. 11, p. 2490-2498, 2006.

SANCHEZ, P.A. et al. Digital soil map of the world. **Science**, v.325, p.680-681, 2009.

TEM CATEN, A. et al. Componentes principais como preditores no mapeamento digital de classes de solos. **Ciência Rural**, v.41, n.7, p.1170-1176, 2011.

# Soil Organic Carbon Stock Estimation Using Legacy Data: A case study of North Fluminense Region-BR

NASCIMENTO, Mariana Melo[1]; COSTA, Elias Mendes[2] ; CEDDIA, Marcos Bacis [3]

[1] Undergraduate student in Agronomy at Federal University Rural of Rio de Janeiro (UFRRJ), Laboratory of Water and Soils in Agroecossistems (LASA). mariana.m.n2015@outlook.com;

[2] Post-Doctoral Student of the Postgraduate Program in Agronomy Soil Science at UFRRJ, Laboratory of Water and Soils in Agroecossistems (LASA). eliasmccosta@gmail.com;

[3] Professor at the Department of Agrotechnologies and Sustainability, Institute of Agronomy at UFRRJ, Laboratory of Water and Soils in Agroecossistems (LASA). marcosceddia@gmail.com.

## Abstract

The purpose of this work was to study and map the distribution of soils organic carbon stock (SOCS) up to 1m depth using legacy data in the north region of Rio de Janeiro State (Brazil). The study compares the performance of geoestatistical (ordinary kriging - Ok) and 3 Random Forest (RF) algorithms. The data belongs to PROJIR dataset, which was generated in 1983 and digitalized recently by the Laboratory of Water and Soils in Agroecosystems (LASA). The study found 161 soil profiles containing both soil organic carbon and soil bulk density, which allow the calculus of SOCS up to 1.0m soil depth.  Besides, to apply RF algorithm, 19 covariates were evaluated as predictor (13 derived from DEM and 6 from Landsat8 images). The map of SOCS generated by OK presented the best performance using the conventional metrics ($r^2$, RMSE and MAE) and the criteria of *Plausibility*, *Interpretability* and *Explainability*. The majority territorial extension of the study site has SOCS between 0-20 Kg.m$^{-2}$.

Keywords: Spatial prediction; Random Forest; Ordinary Kriging.

## Introduction

Knowledge in soil legacy data is at risk of being lost due to the complexity of maintaining files as paper, considering the high costs involved in exploratory soil surveys (ARROUAYS et.al, 2017). Thus, the relevance of retrieving this information is applicability in predictive modeling of soil attributes, scientific studies, and decision-making in public policies. Soil organic carbon (SOC) management is key to climate change mitigation and adaptation by achieving neutrality of land degradation, food security, and a more sustainable ecosystem. SOC in soils has a positive influence on multiple soil functions, where high values of SOCS make soils more resilient to climate change through increased water holding capacity, erosion, and flood control. Using tools such as digital mapping it is possible to estimate the soil organic carbon stock at a regional scale using the equation SCORPAN (GOMES et al, 2019). This modeling is important for calculating the carbon inventory and understanding the biophysical processes that can affect the soil carbon balance. Therefore, specific regionalized surveys are important to ensure an appropriate scale of the study, in addition to being a basis for intervening in soil restoration and a basis for the management of sustainable practices. So, this work sought to model the spatial variability of soil organic carbon stock up to 1m depth using legacy data for a region in the north of the state of Rio de Janeiro using Random Forest and ordinary kriging as an estimator.

## Methodology

In this study, data was collected from an area of approximately 250,000 hectares, representing about 30% of the northern region of Rio de Janeiro. One of the main economic activities of the region was the sugar and alcohol industry, which led, between 1981 and 1983, to carry out an edaphoclimatic survey coordinated by IAA, through National Sugarcane Improvement Program and which was called "Northern Fluminense Irrigation and Drainage Project", PROJIR in Portuguese (FARIAS, 2008). Nowadays, the PROJIR data legacy belongs to Federal Rural University of Rio de Janeiro. The database has several soil observations (soil profiles and soil boreholes) and, among these, 161 profiles have analyzes of organic carbon and soil density, which allowed the direct calculation of the carbon stock up to 1 meter deep using the following equation (GOMES et al, 2019): SOCS = [SOC x BD x T], where: SOCS - Soil Organic Carbon Stocks ($Kg.m^{-2}$); SOC - soil organic carbon content ($g.kg^{-1}$); BD - soil bulk density ($Kg.dm^{-3}$); T - soil layer thickness (m);To predict the SOCS stock 3 versions of the algorithm Random Forest (RF) were trained and 1 model using geostatistical ordinary kriging (OK). The OK is a univariate method that uses the primary variable (SOC) measured at sampled locations to predict the same primary variable at unsampled locations (CEDDIA et al, 2015). The RF algorithms used 19 environmental covariates, being 13 from the relief (DEM, slop, aspect, northernness, plan_cur, prof_curv, convergence, cat_area, twi, ls_factor, rsp, chnd, chnb) and 6 from Landsat8 images (NDVI, EVI, CLAY, SAVI, GSI, IRON). The RF model 1 (RF1) used all covariates (19), the RF model 2 (RF2) used only relief covariates (13), while RF model 3 (RF3) were used only Landsat 8 images. The model's efficiency were performed using $r^2$, RMSE and MAE metrics. The process of preparation of covariates, selection, calibration, evaluation of models, prediction and generation of maps were implemented in R software (R Core Team, 2019).

## Results and discussion

The results of the study is summarized in figure 1 and table 1(the metrics of the four algorithms Rf1, RF 2, RF 3 and OK is presented), respectively. In figure 1a, it's possible to note that the SOCS present regions with higher values (towards to the southeast direction) and lower values (center of the map). Besides, the histogram (figure 1b) shows that the SOCS data do not follows a normal distribution function (positive skewness).

| Table 1. Metrics of the algorithms to map SOCS up to 1m depth. | | | |
|---|---|---|---|
| Model | $R^2$ | RMSE | MAE |
| RF1 | 0.21 | 16.12 | 11.39 |
| RF2 | 0.18 | 17.20 | 11.94 |
| RF3 | 0.30 | 15.52 | 10.55 |
| OK | 0.33 | 15.74 | 8.78 |

The data was transformed using $\log_{10}$ function. The experimental semivariogram with its Gaussian model fitted is presented in figure 1c. A Gaussian model with a pure nugget effect ($C_o$) of 0.058, a contribution ($C_1$) of 0.025 and a range (a) of 4500 meters. The parabolic behavior of the semivariance at short distance shows the strong spatial continuity between points located at up to 3000 meters.

The maps generated by OK (figure 1 d and e, kriging and kriging variance) and the RF1, RF2 and RF3 (figures 1f, 1g and 1 h, respectively) is presented. The map generated by OK shows a more continuous isolines values along the study site, which is a consequence of the spatial continuity captured by the gaussian model. The maps generated by RF models reflect the importance of the covariates used. For example, the maps generated by RF1 and RF 2 shows that the relief covariates has a strong effect on the polygons of values of SOCS (figure 1f and 1g). On the other hands, in the map generated by RF3 (only Landsat8 image covariates) the Paraiba do Sul river is clearly highlighted as having higher SOCS. These results are important to show the importance of the modelling process evaluation, which can be done not only through the conventional metrics ($r^2$, RMSE and MAE) but also analyzing its *Plausibility* (validity considering the current knowledge and scientist theories), *Interpretability* (the translation of an abstract model or model output into terms understandable by humans) and *Explainability* (models must predict and explain the phenomenon) (WADOUX et al, 2020). The map generated by OK also presented a better performance (higher $R^2$ and lower MAE). The model RF3 was the second best (with lower RMSE), however, this map does not follow the criteria of *Plausibility*, *Interpretability* and *Explainability*.
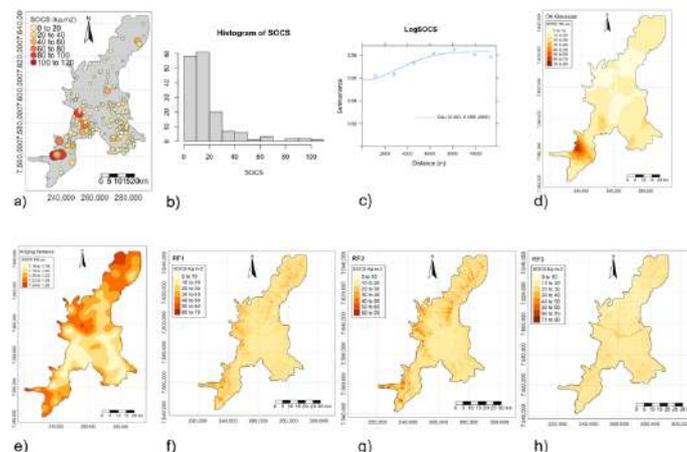


Figure1: a) SOCS distribution b) Histogram of SOCS; c) Semivariogram of log SOCS; d) OK-Gaussian Map; e) Kriging Variance Map; f) RF1 Map; g) RF2 Map; h) RF3 Map.

The models RF1 and RF3 doesn't represent the SOCS. This could be due to the database being collected in the 83's. At that time the Landsat 8 images does not exist and consequently it is not reasonable to explain the carbon dynamics using images from a time different from that when the soil was collected and analyzed. Therefore, it can be inferred that the result is due to a numerical coincidence. The predictive mapping of SOCS with terrain covariates using RF2 (figure 2g) presented the lowest $r^2$ values (0.17) and the worst RMSE result (17.20). The predicted values of SOCS using OK algorithm shows that the majority territorial extension of the study site has SOCS between 0- 20 Kg.m$^{-2}$ (figure 1d). Besides, the lower SOC (0-10 Kg.m$^{-2}$) is found

in the region from the central to northeast direction (figure 1d). The higher values of kriging variance (figure 1e) are observed along the regions with lower observations density, which, in general is commonly associated to the boundaries limits of PROJIR area (figure 1e). The map of kriging variance can be useful for helping the soil scientist to information.

## Conclusions

The map of SOCS generated by OK presented not only the best performance using the conventional metrics ($r^2$, RMSE and MAE) but also the criteria of *Plausibility*, *Interpretability* and *Explainability*. The majority territorial extension of the study site has SOCS between 0 - 20 Kg.m$^{-2}$. The product of this work can serve as input to various models of soil function assessment, for conservation and management purposes as well as soil security.

## References

ARROUAYS, D. et al. **Soil legacy data rescue via GlobalSoilMap and other international and national initiatives** (2017). Available in: https://www.sciencedirect.com/science/article/pii/S2214242816300699?via%3Dihub Accessed: 26 August 2021.

CORE TEAM, R., R: **a Language and Environment for Statistical Computing. R Foundation for Statistical Computing**, Vienna, Austria. ISBN 3-900051-07-0. 2019.

CEDDIA, M. B et all. **Spatial variability of soil carbono stoks in the Urucu river basin, Central Amazon-Brazil** 2015. Available in: http://dx.doi.org/10.1016/j.scitotenv.2015.03.121 Accessed: 07 October 2021.

FARIAS, L. N. **Spatial variability of physical hydric attributes of soils of the North Fluminense Region**. 2008. 65p. Dissertation (Master Science in Agronomy, Soil Science) Instituto de Agronomia, Departamento de Solos, Universidade Federal Rural do Rio de Janeiro, Seropédica, RJ, 2008.

GOMES, L. C. et al. **Modelling and mapping soil organic carbon stocks in Brazil** (2019).Available in: https://www.sciencedirect.com/science/article/pii/S0016706117320669?via%3Dihub Accessed: 26 August 2021.

OLIVEIRA, S. E.; REATTO, A. **Estoques de carbono do solo segundo os componentes da paisagem.** Cadernos de Ciência & Tecnologia, Brasília, v. 32, n.1/2, p. 71-93, jan./ago. 2015

WADOUX, A. M. J-C.; MINASNY, B.; MCBRATNEY A.B. **Machine learning for digital soil mapping: applications, challenges and suggested solutions** 2020. Available in: https://www.researchgate.net/publication/339068148_Machine_learning_for_digital_soil_mapping_applications_challenges_and_suggested_solutions. Accessed: 10 October 2021.

# Disaggregating maps with machine learning to increase the detail in the soil map of Municipality of São Sepé – RS.

SENA, Antonny F. S.[1*]; GIASSON, Elvio[1]; COSTA, José J. F. [1]; COELHO, Fabrício F. [1]; SILVA, Ryshardson G. P. O. [1]; SILVA, Saulo G[1].

[1] Programa de Pós-Graduação em Ciência do Solo - Universidade Federal do Rio Grande do Sul, *agro.antonnysampaio@gmail.com;

## Thematic Session: Pedometrics Guidelines To Systematic Soil Survey: Tropical Study Cases

**Abstract**
The availability of reliable soil data is scarce in many countries, including Brazil. A Digital Soil Mapping technique of the map disaggregation has proven to be an efficient and cost-effective alternative to produce reliable soil data. This research proposed to test an update map disaggregation methodology. The study was conducted using soil legacy data from the Soil Survey of the Municipality of São Sepé - RS, Brazil. The legacy data were digitized in a GIS environment and seven covariates were generated from the DEM. This dataset was processed by the GIS software's own random tree training and classification tools. The results were smoothed and evaluated by stratified cross-validation, error matrix and visual analysis. The method proposed in this work proved to be efficient in individualizing polygons of soil classes allocated in complex areas. The most important variables for the random tree model were altitude and geology.
Keywords: random decision tree; digital soil mapping; legacy soil map.

## Introduction

The demand for more accurate soil information has increased every year, especially for agricultural purposes, where technological advances allow an increasingly efficient use of this information (CHAVES *et al.*, 2021; ZERAATPISHEH *et al.*, 2020). The existing soil maps produced through field surveys provide basic information, but they can be very inaccurate, even more so when they are maps with less detailed scales, which makes it difficult to take decision in agriculture, especially in farms that operate on a smaller scale (FLORES et al., 2013). Digital Soil Mapping (DSM) has presented itself as an alternative for generating more information about soils, either by mapping new areas or as a means of adding greater reliability to existing maps (HÄRING *et al.*, 2012; VINCENT *et al.*, 2018). In this context, the map disaggregation technique, widely used in DSM, stands out for its ability to individualize simple mapping units, often allocated in more complex units, containing two or more soil classes. Among the tools available for map disaggregation in Geographic Information Systems (GIS) environment, decision trees have stood out, presenting good results in the disaggregation of legacy soil maps (MØLLER *et al.*, 2019; SARMENTO *et al.*, 2017). The objective of this work was to perform the disaggregation of a legacy map using the tools of image segmentation and classification by Random Trees available in a GIS environment in order to improve the quality of the legacy map and facilitate the disaggregation process.

## Methodology

The study area is in the municipality of São Sepé, in the central region of the state of Rio Grande do Sul, Brazil. It covers an area of approximately 2,203 km² between latitudes 29°53'S and 30°32'S and longitudes 53°08'W and 53°52'W; the altitude range is between 29 and 461 meters above sea level. The climate of the region according to the Köppen international climate classification system was determined according to (ALVARES *et al.*, 2013), which classifies the region as "Cfa", presenting a humid climate, with rainfall during all months of the year and in the hottest month of the year it presents temperatures above 22°C. The region is drained mainly by the Vacacaí River and its tributaries.

The first step was to digitize the soil map of the São Sepé municipality, originally available in 1:100,000 scale and noncolor paper printed (LEMOS *et al.*, 1972), where 26 soil classes were described. The second step was the extraction of the geomorphometric variables (aspect, slope, curvature, topographic wetness index and Euclidean distance from main drainage) derived from the Digital Elevation Model (DEM), which were obtained using the *Spatial Analyst* extension of the ArcGIS software. The DEM used was obtained from the remote data of the SRTM mission, with 30m resolution. To provide more information to the model, also were used layers with geology data (CPRM, 2010) and a landform map obtained using LandMapR (MACMILLAN, 2003), that subdivides the landscape into 15 landforms.

After creating the layer of sample points, the *Sample* tool was used to extract the data from each layer into a table. Once the spreadsheet was generated, variables selection was done using WEKA 3.8.5 software (FRANK; HALL; WITTEN, 2016), which also allowed to select the best model for classification. Back in ArcGIS, the selected variables were grouped using the *Composite Bands* tool, which groups the various *rasters* into a file with multiple bands and the *Segment Mean Shift* tool was used to group pixels with similar characteristics. These layers were processed by the *Train Random Trees Classifier* tool to generate the classification rules file (.ecd), compatible with the ArcGIS classifier. At this point the point file with the model training data is also added. After having classification rules, were used to classify the raster generated by the *Segment Mean Shift* tool, generating the final map.

The accuracy of the predicted map was verified by using cross-validation, error matrix, and visual observation of the new polygon's limits.

## Results and discussion

The overall map accuracy (agreement with legacy map) was 62.04% and the error was 37.96%. Machado et al. (2018) obtained an overall accuracy of 80.5% in the disaggregation of a map, however, the authors pointed out that maps disaggregated from legacy data with more complex components tend to increase the rate of correct predictions, since this complexity allows the combination of more than one soil class.

The error matrix also showed Planosols as having the highest agreement rate (2.102), which was expected, since this is the class that has the largest area extent (53,584 ha) in the original map and this gives it a predictive agreement rate for the model. The model had Mean Absolute Error (MAE) and Root Mean Squared Error

(RMSE) values of 0.038 and 0.138 respectively. The most important covariates for the construction of the decision tree rules were altitude and geology, which explains the delineation of the polygons obtained in the disaggregated map, as can be seen in Figure 1.
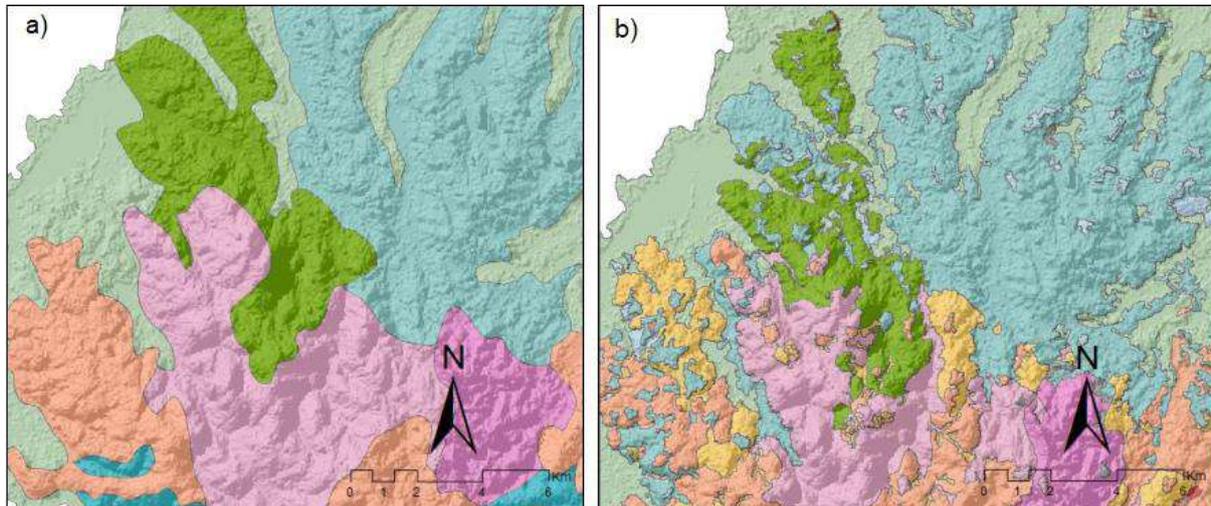


Figure 1. Detail of part of the study area showing combined map units of the legacy soil map (a), and the allocation of individual soils obtained by disaggregation (b).

The delineation of the disaggregated soil classes was richer in details and showed greater conformity with the relief, both in the lower areas, following the design of the slopes, and in the higher areas, such as the appearance of patches in the top areas (Fig. 1b). According to Minai, Libohova e Schulze (2020), legacy soil map units that extend to natural boundaries are often the consequence of delineations made on low-resolution imagery. The disaggregated map represents the distribution of MUs through pixels, enabling more gradual and continuous transitions between the soil classes. Geology, in turn, can explain part of the emergence or disappearance of classes in certain sections of the map.

**Conclusions**

The disaggregation increased the level of detail of the soil map of the São Sepé municipality. The methodology developed in this work was efficient in identifying the occurrence of simple soil classes in complex areas. The methodological changes implemented in this study simplified the disaggregation process.

**References**

ALVARES, C. A. et al. Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, [s. l.], v. 22, n. 6, p. 711–728, 2013.

CHAVES, D. A. et al. Soil mapping and characterization of the Mapou basin, Haiti. **Geoderma Regional**, [s. l.], v. 27, n. May 2020, p. e00432, 2021.

CPRM, S. G. do B.-. **Mapa Geológico do Estado do Piauí**, Geologia e recursos minerais do Estado do Piauí, Teresina: Serviço Geológico do Brasil, 2010.

ESRI. **ArcGIS Desktop: Release 10.8. Redlands, CA: Environmental Systems Research Institute.**, 2019.

FLORES, C. A. et al. Increase of detail in soil surveys and its contribution for managing wine quality in Vale dos Vinhedos , Brazil. **Ciência e Técnica Vitivinícola**, [s. l.], n. November 2014, p. 6, 2013.

HÄRING, T. et al. Geoderma Spatial disaggregation of complex soil map units : A decision-tree based approach in Bavarian forest soils. **Geoderma**, [s. l.], v. 185–186, p. 37–47, 2012.

LEMOS, R. C. De et al. **Levantamento dos solos do município de São Sepé**, Unuiversidade Federal do Rio Grande do Sul, 1972.

MACHADO, I. R. et al. Spatial Disaggregation of Multi-Component Soil Map Units Using Legacy Data and a Tree-Based Algorithm in Southern Brazil. [s. l.], p. 1–14, 2018.

MACMILLAN, R. A. **LandMapR Software Toolkit- C ++ Version: Users Manual.**, 2003.

MINAI, J.; LIBOHOVA, Z.; SCHULZE, D. G. Disaggregation of the 1:100,000 Reconnaissance soil map of the Busia Area, Kenya using a soil landscape rule-based approach. **Catena**, [s. l.], v. 195, n. January, p. 104806, 2020.

MØLLER, A. B. et al. Improved disaggregation of conventional soil maps. **Geoderma**, [s. l.], v. 341, n. January, p. 148–160, 2019.

SARMENTO, C. E. et al. Geoderma Regional Disaggregating conventional soil maps with limited descriptive data : A knowledge-based approach in Serra Gaúcha , Brazil. **Geoderma Regional**, [s. l.], v. 8, p. 12–23, 2017.

VINCENT, S. et al. Geoderma Spatial disaggregation of complex Soil Map Units at the regional scale based on soil-landscape relationships. **Geoderma**, [s. l.], v. 311, p. 130–142, 2018.

ZERAATPISHEH, M. et al. Conventional and digital soil mapping in Iran: Past, present, and future. **Catena**, [s. l.], v. 188, n. December 2019, p. 104424, 2020.

# Geophysical data to modeling soil properties in tropical hillslope areas

BASTOS, Blenda[1]; PINHEIRO, Helena[1]; CARVALHO JUNIOR, Waldir[2]

[1]Universidade Federal Rural do Rio de Janeiro (UFRRJ), blenda.bastos@hotmail.com, lenask@gmail.com, [2]Embrapa solos/RJ, waldir.carvalho@embrapa.br

**Thematic Session: Legacy data: How turn it useful?**

## Abstract

Aero geophysical data is becoming an important source of environmental covariate in digital mapping. Airborne gamma-ray spectrometry is more common in digital soil mapping, because of the penetration potential of approximately 30-40 cm. However, the airborne magnetic method can be tested to add and improve the prediction of soil properties. Therefore, the objective of this work was to implement a preliminary study to model the spatial distribution of soil properties using pedological legacy data, aero geophysical data, and terrain covariates to discuss their importance to the digital soil mapping in Bom Jardim county, Rio de Janeiro, Brazil.

## Introduction

Considering the applicability of the airborne gamma-ray spectrometry to represent different sources of parental material, its use for digital soil mapping has increasing. The penetration potential of approximately 30-40 cm and the correlation with weathering and pedogenesis processes (Wilford et al., 1997) where discussed by Reinhardt et al. (2019), Bonfatti et al. (2020) and Loiseau et al. (2020). The airborne magnetic method, on the other hand, despite being less frequent, showed potential for soil studies as McCafferty et al. (2009), Siemon et al. (2020) and Iza et al. (2018). The research goal is to predict soil properties using legacy soil data, terrain covariates derived from the Digital Elevation Model (DEM) and aero geophysical data (AGD) through Random Forest model, to evaluate the potential of these covariates in digital soil mapping.

## Methodology

The soil dataset gathers 208 superficial soil samples collected in Bom Jardim county, between 2009 and 2011 (Figure 1), and from those samples some of the soil properties were addressed this study: Bases saturation (VV), Soil density (DEN), Clay and Sand contents. The procedures used to collect, describe and analyze the soil samples are detailed by Calderano Filho (2012). The DEM was obtained by interpolation of vectorial data from the official cartographic database of Rio de Janeiro state, at 1:25,000 scale, with a 20m of spatial resolution. From the DEM, 17 terrain covariates were derived in the SAGA-GIS open-source software.
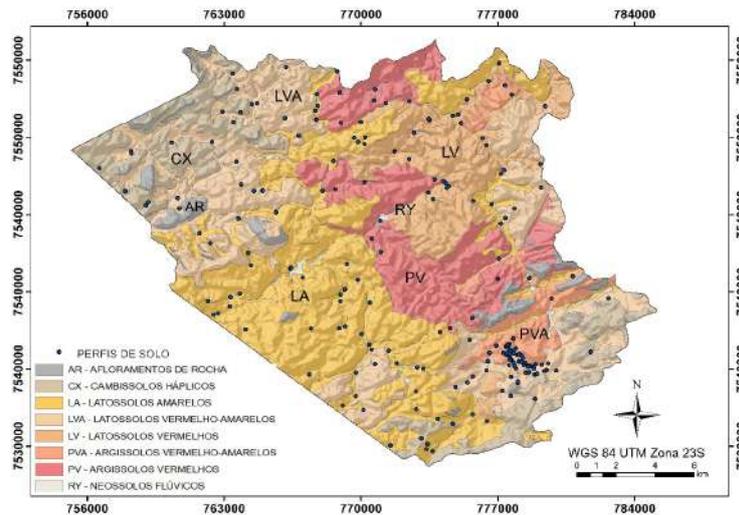
Figure 1: Location of soil profiles, Bom Jardim - RJ, modified from Calderano Filho, B. (2012).

The AGD was obtained from CPRM (2012), and the interpolation was performed using the methods minimum curvature for gamma-ray data (Briggs, 1974) and bidirectional for magnetic data (Geosoft, 2010), totaling 19 covariates with a resolution of 100 m, as suggested by Vasconcellos et al. (1994). After processing, all the products were resampled in the RStudio software to 20 m resolution to adapt them to the terrain covariates resolution. After processing the covariates, Spearman's correlation was applied with a critical value of 95% to exclude covariates that are not correlated with dependent variables. The Random Forest (Breiman, 2001) model was applied with the parameters: ntree=350 and mtry= 10, to modeling soil properties in RStudio software. The accuracy was evaluated through the coefficients $R^2$ and RMSE obtained by the cross-validation method.

## Results and discussion

Four airborne magnetic covariates were excluded from Spearman's correlation, including Total magnetic anomaly, Tilt angle and your absolute value (Miller e Singh, 1994), and Horizontal tilt angle (Cooper e Cowan, 2006). After that, the model RF was applied. Table 1 demonstrates the cross-validation $R^2$ and RMSE values.

Tabela 1: $R^2$ and RMSE for each dependent variable.

|  | DEN | Sand | Clay | VV |
|---|---|---|---|---|
| $R^2$ | 0.37 | 0.18 | 0.15 | 0.19 |
| RMSE | 0.46 g/cm3 | 45.63 g/kg | 87.62 g/kg | 19.07 % |

Carvalho Junior et al. (2014), from the same database, considering 0-5 cm soil depth and the method ordinary kriging, obtained $R^2$ values 0.19 and 0.17 to predict soil clay and sand contents, respectively. The study considering 18 environmental covariates derived from DEM and satellite imagery. Comparing with the $R^2$ values of this study for clay and sand, the AGD seems to be important to improve the prediction performance. Figure 2 shows the top five variables that most contributed to regression model performance.
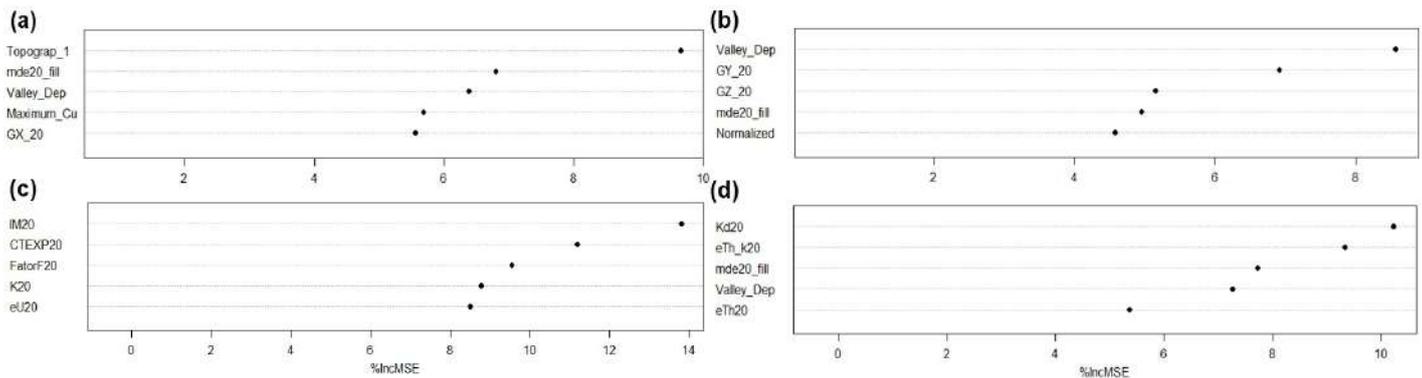
Figure 2: Top 5 most important variables for RF model performance: (a) Sand, (b) Clay, (c) Soil density (DEN), (d) Bases saturation (VV).

The AGD is present in the top 5 covariates to all properties studied (Figure 2). The magnetic data GX, GY and GZ as we can see in Barbosa et al. (2013) are related to the presence of magnetic bodies, in other words, the reflecting parental material characteristics. The Mafic Index (IM) was calculated by combing magnetic and gamma data and allows the removal of the influence of iron-rich soils (Barbosa et al., 2013) and according to Figure 2, has importance in the prediction of DEN. The importance of gamma-ray data was remarkable to predict DEN and VV. CTEXP is mostly related to the source material (high K, eU and eTh values). High values of eTh can be related to the parental material or clay related with intense weathering process (Wilford et al., 1997). Kd is the value of K (%) without the eTh contribution highlighting these element anomalies (Pires, 1995) and FatorF is calculated by the formula F=K*(eU/eTh), where high values show K (%) enrichment (Ribeiro et al., 2014), that can explain your contribution to the VV prediction.

## Conclusions

From the results observed, it was possible to conclude that the aero geophysical data have significant importance. AGD can be used in predictive modeling procedures to map soil properties as support with terrain covariates to understand the origin of soil property's spatial variability.

## References

BARBOSA, I. O. et al. Geology, airborne geophysics, geomorphology and soils in the individualization of the Niquelândia mafic-ultramafic complex, Goiás state, Brazil. Brazilian Journal of Geophysics, v. 31, n. 3, p. 463-481, 2013.

BONFATTI, B. R. et al. Digital mapping of soil parent material in a heterogeneous tropical area. Geomorphology, 367, p. 107305, 2020.

BREIMAN, L. Random forests. Machine learning, 45(1), p. 5-32, 2001.

BRIGGS, IAN C. Machine contouring using minimum curvature. Geophysics, v. 39, n.1, p. 39-48, 1974.

CALDERANO FILHO, B. Análise geoambiental de Paisagens Rurais Montanhosas da Serra do Mar utilizando Redes Neurais Artificiais. Subsídios a sustentabilidade ambiental de ecossistemas frágeis e fragmentados sob interferência antrópica. Embrapa Solos-Tese/dissertação, 2012.

CPRM. Relatório final Projeto Aerogeofísico Rio de Janeiro (Projeto 1.117), v. 1, 219 p., 2012.

DE CARVALHO JUNIOR, W. et al. A regional-scale assessment of digital mapping of soil attributes in a tropical hillslope environment. Geoderma, 232, p. 479-486, 2014.

GEOSOFT INC. Filtragem montaj MAGMAP. Processamento de dados de campos potenciais no domínio da frequência. Extensão para o Oasis Montaj, v. 7.1. Tutorial. e guia do usuário. Toronto, ON Canadá, 77p., 2010.

IZA, E. R. H. D. F. et al. Integration of geochemical and geophysical data to characterize and map lateritic regolith: an example in the Brazilian Amazon. Geochemistry, Geophysics, Geosystems, v. 19, n. 9, p. 3254-3271, 2018.

LOISEAU, T. et al. Could airborne gamma-spectrometric data replace lithological maps as co-variates for digital soil mapping of topsoil particle-size distribution? A case study in Western France. Geoderma Regional, v. 22, 2020.

MCCAFFERTY, A. E., VAN GOSEN, B. S. Airborne gamma-ray and magnetic anomaly signatures of serpentinite in relation to soil geochemistry, northern California. Applied Geochemistry, v. 24, n. 8, p. 1524-1537, 2009.

PIRES, A. C. B. Identificação geofísica de áreas de alteração hidrotermal, Crixás-Guarinos, Goiás. Revista Brasileira de Geociências, São Paulo, v. 25, n. 1, p. 61- 68, 1995.

REINHARDT, N., HERRMANN, L. Gamma-ray spectrometry as versatile tool in soil science: A critical review. Journal of Plant Nutrition and Soil Science, v. 182, n.1, p. 9-27, 2019.

RIBEIRO, V. B., MANTOVANI, M. S. M., LOURO, V. H. A. Aerogamaespectrometria e suas aplicações no mapeamento geológico. Terrae Didática, v.10, n.1, p. 29-51, 2014.

SIEMON, B. et al. Airborne electromagnetic, magnetic, and radiometric surveys at the German North Sea coast applied to groundwater and soil investigations. Remote Sensing, v. 12, n. 10, p. 1629, 2020.

WILFORD, J. R., BIERWIRTH, P. E., CRAIG, M. A. Application of airborne gamma-ray spectrometry in soil/regolith mapping and applied geomorphology. AGSO Journal of Australian Geology and Geophysics, v. 17, n.2, p. 201-216, 1997.

# How to use legacy soil data to plan new data collection?

COSTA, Elias Mendes[1]; CEDDIA, Marcos Bacis[1]
[1] Universidade Federal Rural do Rio de Janeiro, eliasmccosta@email@.com; marcosceddia@gmail.com[1]

## Thematic Session: Legacy Data: How turn it usefull?

**Abstract**
This study aimed to analyze and discuss how the use of legacy soil data and landscape similarity analysis can help in planning the allocation of preferred areas for detailed soil surveys in the scope of the PronaSolos-RJ project. Covariates were raised that represent the soil formation factors were raised for the entire state. These covariates were then used in the Gower dissimilarity index in order to evaluate the similarity from different reference areas (legacy soil survey). The territory of Rio Janeiro is not satisfactorily covered by pedagogical maps at scales and levels of detail to meet current and future demands. The reference area approach to representing a region through the Gower index similarity analysis can save time, money, and personnel resources.

## Introduction

There is an increasing demand for spatial information on soil types and their properties by scientists and decision-makers to better understand the effect of a growing population and an increasing demand for food in a climate-changing environment. In many places in the world, soil information is difficult to obtain and can be non-existent and in Brazil is not different. When no detailed map or soil observation is available in a region of interest, we can use a reference area (RA) with similar characteristics (Mallavan et al., 2010) to extrapolate the soil-landscape relationship. Considering that funding to obtain new soil surveys is very scarce in Brazil we have to optimize it by organizing the available existing data (legacy data) and planning the soil survey in representative areas where we can use the idea of transferability model to map other regions with similar characteristics (Grunwald et al., 2018).

A large quantity of soil data has already been produced in Brazil as part of soil surveys and research projects on the various aspects of soil science (Samule-Rosa et al., 2020), however, there is no standardization of this collection, not a single base where you can consult and have a quick answer about the level of survey, scale, format, title, authorship and where if you find such information. Realizing this need, the National Soil Program of Brazil (PronaSolos) was created, an ambitious project to investigate the Brazilian soils through soil surveys throughout the entire territory that will consolidate data integration and availability, decision-makers needs and collaborate with the advancement of knowledge of the soils in Brazil.

There will be several initiatives across the country with partnerships between the union, states and municipalities. One of the states that have already started the activities of planning and executing PronaSolos is the state of Rio de Janeiro (RJ), the state where the program was born. Seeking to optimize resources of time, money and people this study aimed to analyze and discuss how the use of legacy soil data

and landscape similarity analysis can help in planning the allocation of preferred areas for detailed soil surveys in the scope of the PronaSolos-RJ project.

## Methodology

The study area is the State of Rio de Janeiro (RJ) located between the geographical coordinates 41º and 45º W and 20º30 and 23º30 S and is about 44,000 km² in the Southeast of Brazil (Figure 1). The area is divided into six geopolitical mesoregions known as Baixadas, Centro Fluminense, Metropolitana do Rio de Janeiro, Noroeste Fluminense, Norte Fluminense e Sul Fluminense. The state also is characterized by eight large landscape types known as Serra da Bocaina, Coastal Plains, Mountainous Area, North-Northwest Fluminense, Paraíba do Sul River (Middle Valley), Serra Mantiqueira, Serra dos Órgãos, and Upper Itabapoana River (Plateau), described in Mendonça-Santos et al. (2008).
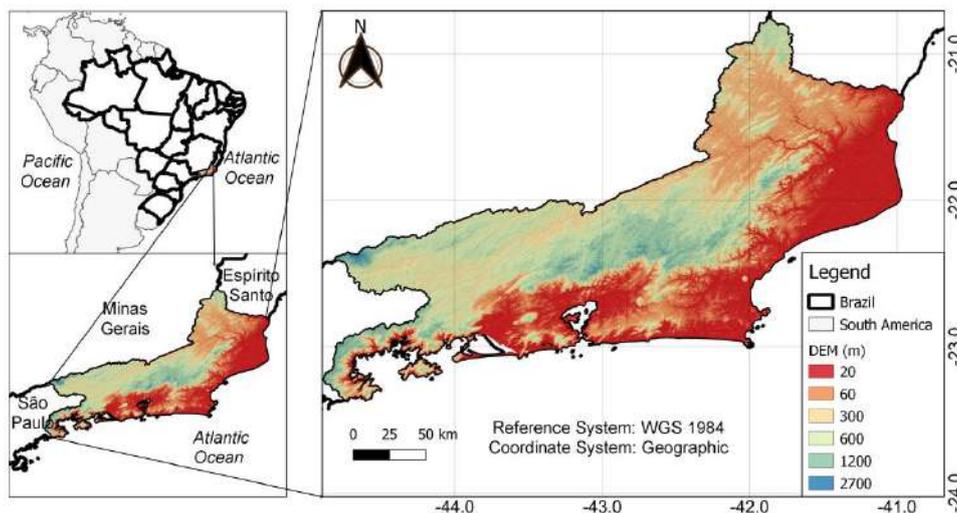


**Figure 1**. The study area location and elevation map, extracted from the SRTM DEM.

Covariates that represent the factors of soil formation were raised for the entire state, they were: Relief (Elevation, slope, topographic wetness index); Parent material (geology); Climate (precipitation and average temperature); Soil (legacy map 1:250,000 scale); Organism (Sentinel2 bands b2, b3, b4, b8 and SAVI index).

The Gower similarity index (GI) proposed by (Gower, 1971) as outlined by (Mallavan et al., 2010), was employed to measure the similarity among fields (legacy data-reference area).

$$Sij = \frac{1}{p}\sum_{k=1}^{p}(1 - \frac{|xik - xjk|}{range\ k})$$

where Sij is the GI between sites i and j; k represents each covariate; p is the number of covariates; range k is the value range of covariate k in the whole study area. Thus, Sij ranges between 0 and 1; a value of 1 means that the two individuals differ in no character whereas 0 means they differ maximally in all their characters. In our case, the interpretation is the opposite of the one presented above, that is, values of 1-Sij equal to 0 means that the two individuals differ in no character whereas 1 means they differ maximally in all their characters. All legacy soil survey from RJ state were

downloaded from the Geoinfo platform of the Brazilian Agricultural Research Corporation (Geoinfo-Embrapa). They were used as reference areas (RA) for computing the GI, one each time. The final value with the dissimilarity classes is an average of all maps that were classified following the GI criterion, greater than 0.12, "Dissimilar" and less than or equal 0.12 "Similar"

**Results and discussion**

As can be seen, none of the areas is capable of representing very well the entire state of Rio de Janeiro (Figure 2) and, as expected, represents its surroundings where the environmental characteristics are similar. For example, the RA "Região Serrana" represents very well the entire Serrana region of Rio, with GI less than 0.1, but does not represent very well the coastal regions or the north and northwest Fluminese regions

For example, the Projir region is representative of the northern region and some parts of the coast, with a GI less than 0.1, but it represents very well neither the Serrana region nor the mountainous region of the southern Fluminense in the "Serra do Mar" and "Serra da Mantiqueira" regions.
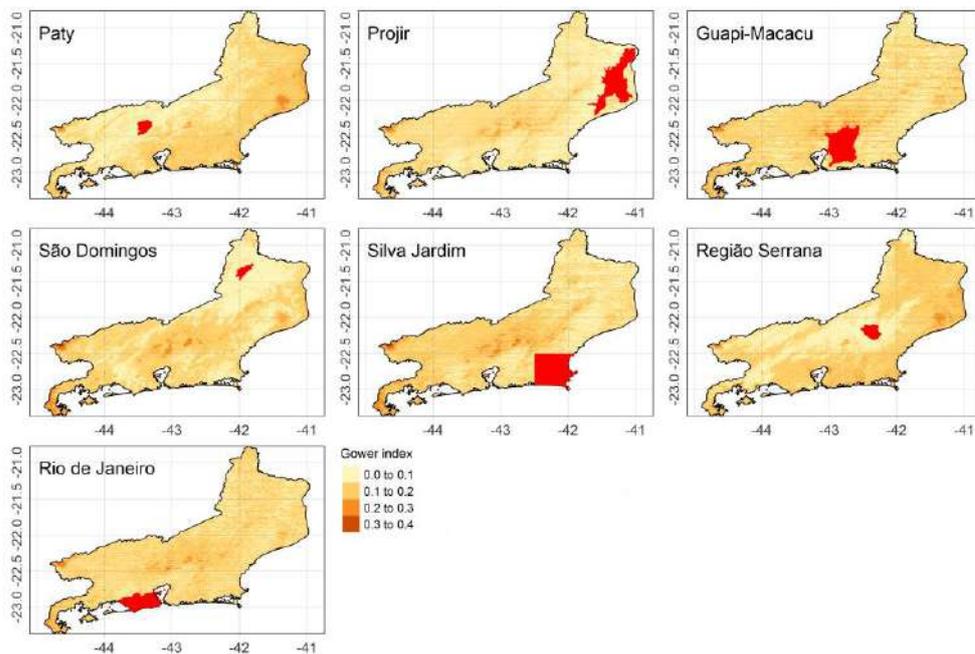


**Figure 2**. Gower index maps using as reference different regions of Rio de Janeiro where soil surveys were carried out.

When combining all GI maps, it is possible to see that the areas considered dissimilar are mainly the areas with rugged terrain in the Serrana region, the Sul Fluminense region and the Costa Verde (Figure 3).

Also, part of the north coast of Rio de Janeiro was classified as dissimilar when combined the indices of all Ras (Figure 3).

Following the idea of using the areas where soil surveys have already been carried out as representative RA, it is possible to define model extrapolation limits (transferability) for a given region. Combining this information, it is possible to define

areas that have little similarity to any existent RA, that is, priority areas for soil survey following the demands of PronaSolos. This procedure can and should optimize resources of time, money and people
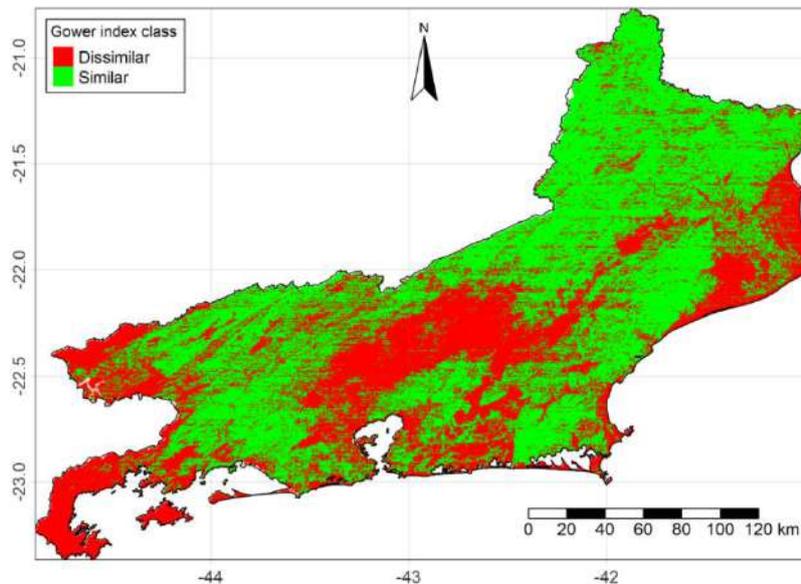


**Figure 3.** Dissimilarity class map based on average Gower index

## Conclusions

The territory of Rio Janeiro is not satisfactorily covered by pedological maps at scales and levels of detail to meet current and future demands.
The reference area approach to representing a region through the Gower index similarity analysis can save time, money, and personnel resources**.**

## References

Gower J. C. A General coefficient of similarity and some of its properties. **Biometrics**. v.27, n. 4. pp. 857-871. 1971

Grunwald, S.; Yu, C.; Xiong, X. Transferability and scalability of soil total carbon prediction models in Florida, USA. **Pedosphere**. 28(6), pp.856–872, 2018.

Mallavan, B.P.; Minasny, B.; McBratney A. B. Homosoil, a methodology for quantitative extrapolation of soil information across the globe. **Chapter 12. Digital Soil Mapping**. pp137-149. 2010

Mendonça-Santos, M.L.; Santos, H.G.; Dart, R.O.; Pares, J.G. Digital mapping of soil classes in Rio de Janeiro state, Brazil: Data, modelling and prediction. **Chapter 34. Digital Soil Mapping**. pp. 381-396. 2010

Samuel-Rosa, A.; Dalmolin, R. S. D.; Moura-Bueno, J. M.; Teixeira, W. G.; Alba, J. M. F. Open legacy soil survey data in Brazil: geospatial data quality and how to improve it. **Scientia Agricola.** v.77, n.1, pp.1-11. 2020

# Legacy data: exploratory analysis to digital soil class mapping.

Carvalho Junior, Waldir[1]; Bhering, Silvio Barge[1]; Pereira, Nilson Rendeiro[1]; Chagas, Cesar da Silva[1]; Lopes, Carlos Henrique Lemos[2].

[1]: Embrapa Solos, {waldir.carvalho; silvio.bhering; nilson.pereira; cesar.chagas}@embrapa.br; [2]: Secretaria de Estado de Meio Ambiente, Desenvolvimento Econômico, Produção e Agricultura Familiar – SEMAGRO, chlopes@semagro.ms.gov.br

## Thematic Session: Legacy Data – how to turn it usefull

**Abstract**

The use of legacy data improve the value of these data and promote the input of new data on the database. Also, reduces the primary data collection and analytical procedures. In MS State, the legacy data was 1325 soil profiles. We are looking for correlations with previus soil map, and litology and vegetation maps to reveal the profiles distribution. The DEM and derivatives was used also to compare with the profiles distribution. The spatial distribution of soil profiles follow the categorical maps units occurrence. The statistical values of DEM and slope are not significantly different. The spatial distribution of soil profiles can represent the covariates of the all area and are ready to be used to produce digital soil maps.
Keywords: soil database; Mato Grosso do Sul; Tropical.

## Introduction

The use of legacy data for soil surveys is important to reducing primary data collection, and enhance available data that could otherwise be neglected.

Looking for a methodology to verify the use of these legacy data against some predictor covariates, an exploratory spatial analysis of legacy data over the Paraguay river basin, in the state of Mato Grosso do Sul, with the exception of the wetland, was carried out (Figure 01).

The objective of this work is to verify the adequacy of the spatial distribution of the legacy data as a function of covariates such as altimetry, slope, lithology, biomes, vegetation and soils maps in the study area.

## Methodology

The dataset has 1325 soil profiles collected in previous works, without the use of statistical sampling techniques, important step in digital soil mapping. These data belong to the ZAE MS project of Embrapa Solos and will soon be available in the institution's databases. Numerical altimetry and slope covariates were obtained from NASA JPL (2020). Thematic covariates on Biomes, Lithology, Soils and vegetation were obtained from BDiA – IBGE (2021).

The study area is included in the cerrado biome, and represents approximately 27% of the state's area, with 96,960 km2. Figure 01 shows the study area and thematic covariates used in this study, namely, lithology, soils and vegetation.

## Results and discussion

The DEM and slope covariates has the follow characteristics (Table 01). The DEM values are close between study area and soil samples. Apparently the slope

has different distribution, but areas with slope above 45 % are less than 01% of the study area, and we can consider that the relation are maintained.

Table 02 shows 1325 soil samples grouped by class according to FAO (2021). Note an unbalanced distribution of classes, with predominancy of Ferralsols, Arenosols and Acrisols. The covariates are showed in Figure 01.

Table 01. Statistics values of DEM (meters) and slope (%) of study area and soil samples.

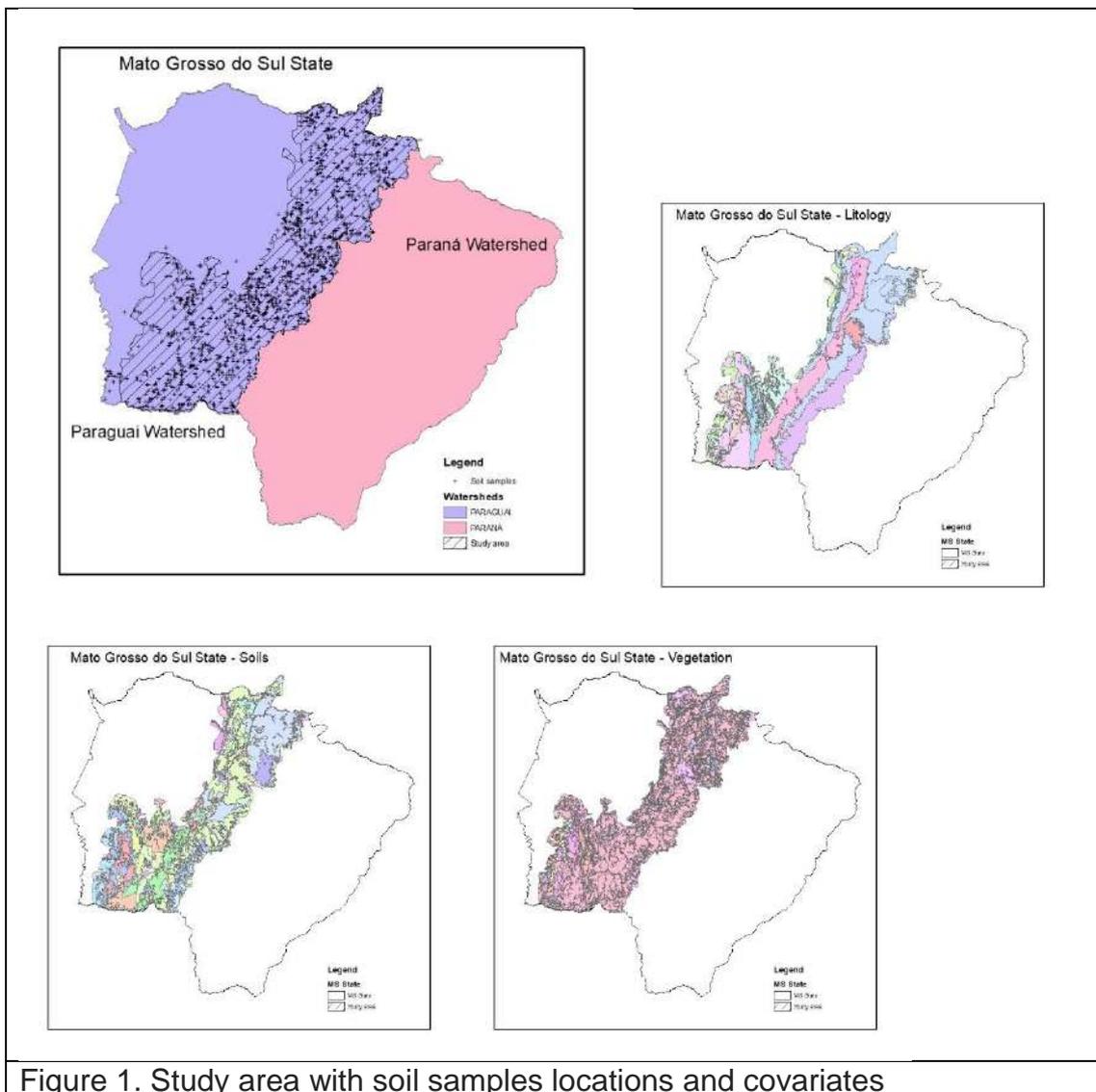|                    | min | max | mean | SD  |
|--------------------|-----|-----|------|-----|
| DEM study area     | 73  | 942 | 334  | 146 |
| Slope study area   | 0   | 370 | 7.5  | 7.9 |
| DEM soil profiles  | 76  | 890 | 344  | 146 |
| Slope soil profiles| 0   | 52  | 5.5  | 4.8 |



Figure 1. Study area with soil samples locations and covariates

Table 02. Samples Soil Class distribution according FAO soil taxonomy (FAO, 2021)

| FAO Class of soil samples | count | % |
|---|---|---|
| Acrisols | 145 | 10.9 |
| Arenosols | 250 | 18.9 |
| Cambisols | 45 | 3.4 |
| Chernozems | 42 | 3.2 |
| Ferralsols | 540 | 40.8 |
| Fluvisols | 1 | 0.1 |
| Gleysols | 46 | 3.5 |
| Leptosols | 69 | 5.2 |
| Luvisols | 4 | 0.3 |
| Nitisols | 34 | 2.6 |
| Planosols | 50 | 3.8 |
| Plinthosols | 46 | 3.5 |
| Regosols | 35 | 2.6 |
| Vertisols | 18 | 1.4 |

The result of join the soil samples and soil map, shows that the soil samples distribution has a quite relation with soils polygons distributions. The Table 3 shows this relation. We can note that only one soil unit map (Histosols) doesn´t have soil samples and that the greater soil units have the greater account of soil samples. The idea is to analysis which soil samples are within each soil unit, to best understand the soil sample distribution in relation to soil map, but the quantitative aspects reveals that the soil samples distribution follow the same distribution of soil unit map.

Table 3. Soil units map, percent and km2 distribution and soil samples within each soil unit.

| Soil units | % | km2 | soil samples |
|---|---|---|---|
| Rock outputs | 0.6 | 595 | 3 |
| urban | 0.0 | 20 | 0 |
| water | 0.1 | 89 | 0 |
| Plinthosols | 1.8 | 1,785 | 21 |
| Gleysols | 2.0 | 1,976 | 20 |
| Ferralsols | 25.9 | 25,120 | 482 |
| Chernozems | 3.7 | 3,619 | 27 |
| Nitisols | 3.5 | 3,352 | 66 |
| Histosols | 0.0 | 8 | 0 |
| Acrisols | 12.4 | 12,056 | 205 |
| Leptosols | 10.9 | 10,571 | 73 |
| Arenosols | 20.2 | 19,625 | 272 |
| Regosols | 8.9 | 8,659 | 72 |
| Planosols | 7.6 | 7,343 | 64 |
| Vertisols | 2.2 | 2,143 | 20 |

The vegetation covariate also has a relation between map units and soil samples spatial distribution. The same correlation occurs between soil samples and lithology map units. These correlations can be noted in Table 4.

Table 4. Correlation between soil samples spatial distribution and map units of vegetation and lithology. Note that only the lithology map units with soil samples are showed.

| Lithology map units | % | km² | soil samples | Vegetation map units | % | Km² | soil samples |
|---|---|---|---|---|---|---|---|
| Sand deposits | 1.1 | 1,049 | 15 | Contact | 32.2 | 31,260 | 370 |
| amphibolite | 0.1 | 57 | 1 | Water | 0.1 | 86 | |
| Arches, Conglomerate | 0.3 | 323 | 6 | Deciduous Seasonal Forest | 2.9 | 2,777 | 18 |
| Sandstone | 37.0 | 37,300 | 524 | Seasonal Semideciduous Forest | 1.0 | 928 | 7 |
| Biotite | 9.0 | 8,750 | 81 | Savannah | 62.6 | 60,704 | 920 |
| Limestone | 4.1 | 3,999 | 57 | Savannah-estepe | 1.2 | 1,195 | 10 |
| dacitus | 12.2 | 11,806 | 205 | Savannah-estepe | 1.2 | 1,195 | 10 |
| clay deposit | 6.5 | 6,260 | 62 | | | | |
| Diamictite, Shale | 15.2 | 14,705 | 217 | | | | |
| Philito | 4.8 | 4,607 | 59 | | | | |
| shale | 2.0 | 1,943 | 44 | | | | |
| Marble | 1.1 | 1,043 | 20 | | | | |
| quartzite | 1.4 | 1,342 | 22 | | | | |
| Schist | 3.4 | 3,312 | 12 | | | | |

**Conclusions**

The spatial location of soil profiles follows the spatial distribution of the covariates, which denotes that the spatial distribution of soil profiles can represent the covariates of the all area and can be used to produce digital soil maps.

**References.**

BDia – IBGE (2021). Banco de Informações Ambientais. from https://bdiaweb.ibge.gov.br/#/home. Accessed August 18, 2021

FAO (2021). World Reference Base For soil Resources 2014. http://www.fao.org/soils-portal/data-hub/soil-classification/world-reference-base. Accessed August 19, 2021.

NASA JPL (2020). *NASADEM Merged DEM Global 1 arc second nc V001*. NASA EOSDIS Land Processes DAAC. Accessed 2020-08-28 from https://doi.org/10.5067/MEaSUREs/NASADEM/NASADEM_NC.001. Accessed August 28, 2020

# Accuracy of Pedotransfer Functions to Estimate Soil Bulk Density in Brazil

HUF DOS REIS, Aline Mari[1]; TEIXEIRA, Wenceslau Geraldes[2]; FONTANA, Ademir[2];
BARROS, Alexandre Hugo Cezar[2]; VICTORIA, Daniel de Castro[3]
[1] ZARC - Embrapa Solos, aline.huf@colaborador.embrapa.br; [2] Embrapa Solos,
wenceslau.teixeiral@embrapa.br; [2] ademir.fontana@embrapa.br; [2] alexandre.barros@embrapa.br ;
[3] Embrapa Informática Agropecuária, daniel.victoria@embrapa.br

## Thematic Session: Legacy data: How turn it useful?

**Abstract**

Soil bulk density is an important soil physical property, used as a quality indicator. Its variation influences soil water content and carbon stock estimates. This study aims to evaluate the accuracy of pedotransfer functions that predict bulk density in Brazil. The predictive capacity of 14 pedotransfer functions were evaluated using the Pearson correlation ($r$), the mean standard error (MSE), and the root mean square error (RMSE). The best results were obtained by Benites *et al.* (2006) – B, Botula (2013), and Souza *et al.* (2016). However, the inaccuracy is not acceptable for some applications and new functions considering a hierarchical system of soil data considering soil class and depth (surface or subsurface), land use (agriculture, pasture, and forest) and management (no-tillage, conventional or livestock farming forest integration) will be developed and tested.

## Introduction

Soil bulk density (BD) is not determined in routine soil laboratories (De Vos *et al.*, 2005), mainly due to its specific applications and limited use in fertilizer recommendations. However, BD is a means to evaluate sustainable soil management practices (Botula, 2013), especially for soil structure quality, as it reflects compaction (Assouline, 2006). Additionally, Agricultural Zoning of Climatic Risk (ZARC) provides information on planting dates and the probability of unfavorable weather events for the entire territory. ZARC uses a time series of climate data, phenological information and available soil water (ASW). ASW data came from pedotransfer functions estimatives that in nowadays are estimated by a particle size distribution (Teixeira *et al.*, 2021). BD data may improve the ASW estimatives. Moreover, crop models as Decision Support System for Agrotechnology Transfer - DSSAT (Hoogenboom *et al.*, 2019) need the BD values of soil horizons to run the analyses and predictions. Another demand or BD data is as to estimate soil stock of carbon for the inventory Measurement, Reporting and Verification of greenhouse gas (GHG) mitigation, especially from agriculture (Bernoux *et al.*, 1998; De Vos *et al.*, 2005; FAO, 2020). Considering its importance, the lack of BD data and the possibility of using pedotransfer functions (PTFs) to estimate missing information, the objective of this study was to evaluate the accuracy of available PTFs to predict soil BD for Brazil.

## Methodology

The BD data associated with soil chemical and physical properties were obtained in the Hydrophysical Database for Brazilian Soils - HYBRAS (Ottoni *et al.*, 2018) and

other data sets totaling 2,635 soil samples data. The soil samples, parameters ranged as follows: 0.4, 0 and 0% for sand, silt, and clay minimum contents respectively; 98.8, 83, and 96% respectively for maximum values. Organic carbon (OC) values ranged from 0 to 62.1%; pH (water) ranged from 2.4 to 9.6, and the sum of bases from 0.3 to 502 mmol$_c$ kg$^{-1}$. 14 PTFs were analyzed. These PFTs were created for a large variety of soils from different locations. They use as estimator parameters: sand, silt and clay content, organic carbon (OC), pH, partial sum of cations, and sum of bases. For each BD-PTF estimative, the range limits were respected. The Pearson correlation coefficient (*r*) represents the precision of results. *i.g.*, the higher *r* values are the best, the mean standard error (MSE), and the root mean square error (RMSE), were used to analyze the accuracy of BD estimate values. Coefficient of variation (CV) was also used. The lower MSE and RMSE indicate the higher accuracy in the prediction.

**Results and discussion**

The 14 BD-PTF equations used in this study, the location where the most data are from, the predictive soil parameters and the precision and accuracy indices are shown in Table 1. The best indices (*r*, RMSE and MSE) were observed for the Benites – B, Botula and Souza PTFs (Table 1). Due to the dispersion of measured values in relation to deviations, the Benites – B function showed greater variation, especially for low BD values (Figure 1). The proposed function considered few samples with low BD, as well as high BD samples. Thus, low BD values are underestimated, and high BD values are overestimated. Similar trend was observed with the PTFs proposed by Botula (2013) and Souza *et al.* (2016). Our dataset has many sandy soil samples, which differs from the datasets used to generate these three PTFs, presenting a BD maximum value of 2,1 g cm$^{-3}$.

Boschi *et al.* (2016) obtained good performance for the PTF proposed by Benites *et al.* (2006) – B (RMSE=0.19) and other, after evaluating 25 PTFs for a set of 222 soil profiles from all Brazilian biomes.

Table 1 – Statistical metrics of precision and accuracy to describe the performance of 14 PTFs to estimate Bulk Density.

| PTF | Location | Parameters | n | *r* | MSE | RMSE | CV % |
|---|---|---|---|---|---|---|---|
| Alexander (1980) | California – USA | OC | 2627 | 0.37 | 0.00 | 0.22 | 11 |
| Manrique & Jones (1991) – A | USA | OC | 2630 | 0.39 | 0.00 | 0.22 | 12 |
| Manrique & Jones (1991) – B | USA | OC | 2626 | 0.42 | 0.00 | 0.20 | 9 |
| Bernoux *et al.* (1998) – A | Amazon - Brazil | Clay | 2538 | 0.45 | 0.00 | 0.28 | 7 |
| Bernoux *et al.* (1998) – B | Amazon - Brazil | Clay, OC | 2282 | 0.61 | 0.00 | 0.25 | 9 |
| Bernoux *et al.* (1998) – C | Amazon - Brazil | Clay, OC, pH | 1103 | 0.44 | 0.00 | 0.25 | 8 |

| Reference | Location | Predictors | n | r | | | |
|---|---|---|---|---|---|---|---|
| Bernoux *et al*. (1998) – D | Amazon - Brazil | Clay, sand, OC, pH | 1054 | 0.44 | 0.00 | 0.24 | 8 |
| Tomasella & Hodnett (1998) | Amazon - Brazil | Clay, silt, OC | 1769 | 0.50 | 0.00 | 0.23 | 10 |
| Kaur *et al*. (2002) | Almora - India | Clay, silt, OC | 2413 | 0.58 | 0.01 | 0.38 | 30 |
| Benites *et al*. (2006) – A | Brazil | Clay | 2635 | 0.49 | 0.00 | 0.20 | 7 |
| Benites *et al*. (2006) – B | Brazil | Clay, OC | 2385 | <u>0.63</u> | <u>0.00</u> | <u>0.19</u> | 11 |
| Benites *et al*. (2006) – C | Brazil | Clay, OC, SB* | 963 | 0.40 | 0.00 | 0.18 | 9 |
| Botula (2013) | Lower Congo | Clay, sand, OC | 2047 | <u>0.46</u> | <u>0.00</u> | <u>0.19</u> | 3 |
| Souza *et al*. (2016) | Rio Doce Basin - Brazil | Clay, OC, pH, SB | 688 | <u>0.35</u> | <u>0.01</u> | <u>0.18</u> | 10 |

SB*: partial sum of cations (Ca. Mg and K); n: number of samples; *r*: Pearson correlation coefficient; RMSE: root mean square error; MSE: mean standard error; CV: coefficient of variation.
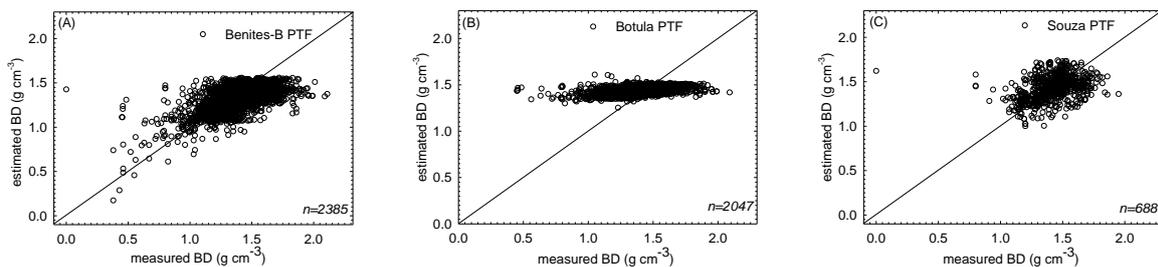


Figure 1 – Measured BD values *vs* deviation of estimates of three PTFs: Benites *et al*. (2006) – B (A). Botula (2013) (B), Souza *et al*. (2016) (C).

## Conclusions

BD predictions from all PTFs tested, show relatively low accuracy. It may be unacceptable in some BD data applications. New PTFs using hierarchical approaches to estimate BD are already under analysis.

## Acknowledgements

## References

ALEXANDER, E. B. Bulk densities of California soils in relation to other soil properties. **Soil Science Society of America Journal,** v. 44, p. 689-692, 1980.

ASSOULINE, S. Modeling the relationship between soil bulk density and the water retention curve. **Soil Science Society of America Journal,** v. 5, p. 554-563, 2006.

BENITES, V. M. et al. Funções de pedotransferência para estimativa da densidade dos solos brasileiros. Dados eletrônicos - Rio de Janeiro: Embrapa Solos, 2006.

BERNOUX, M. et al. Bulk densities of Brazilian soils related to other soil properties. **Soil Science Society of America Journal,** v. 62, p. 743-749, 1998.

BOSCHI, R. S. et al. How accurate are pedotransfer functions for bulk density for Brazilian soils? **Scientia Agricola**, v. 75, n. 1, p. 70-78, 2016.

BOTULA, Y.-D. 2013. **Indirect methods to predict hydrophysical properties of soils of Lower Congo**. PhD thesis, Ghent University.

DE VOS, B. et al. Predictive quality of pedotransfer functions for estimating bulk density of forest soils. **Soil Science Society of America Journal,** v. 69, p. 500-510, 2005.

HOOGENBOOM, G.; et al. The DSSAT crop modeling ecosystem. In: p.173-216 [K.J. Boote, editor] **Advances in Crop Modeling for a Sustainable Agriculture**. Burleigh Dodds Science Publishing, Cambridge, United Kingdom, 2019.

KAUR, R.; SANJEEV, K.; GURUNG, H. P. A pedo-transfer function (PTF) for estimating soil bulk density from basic data and its comparison with existing PTFs. **Australian Journal of Soil Research**, v. 40, p. 847-857, 2002.

MANRIQUE, L. A.; JONES, C. A. Bulk density of soils in relation to soil physical and chemical properties. **Soil Science Society of America Journal**, v. 55, p. 476-481, 1991.

OTTONI, M.V. et al. Hydrophysical database for Brazilian soils (HYBRAS) and pedotransfer functions for water retention. **Vadose Zone Journal**, v. 17, p. 1–17, 2018.

SOUZA, E. et al. Pedotransfer functions to estimate bulk density from soil properties and environmental covariates: Rio Doce basin. **Scientia Agricola**, v. 73, n. 6, p. 525-534, 2016.

TEIXEIRA, W. G. et al. Predição da água disponível no solo em função da granulometria para uso nas análises de risco no Zoneamento Agrícola de Risco Climático. **Boletim de pesquisa e desenvolvimento,** Rio de Janeiro - Embrapa Solos, p. 1-102, 2021.

TOMASELLA, J.; HODNETT, M. G. Estimating soil water retention characteristics from limited data in Brazilian Amazonia. **Soil Science**, v. 163, p. 190-202.

# Soil evaluation and socioeconomic factors influence on agricultural efficiency: a pilot study in Rio de Janeiro

GOMES, Eliane[1]; FIDALGO, Elaine Cristina[2]; OLIVEIRA, Elton[3]; MARTINS, Alba Leonor[4]; OLIVEIRA, Aline[5]; COELHO, Maurício[6]

[1] Embrapa / SIRE, eliane.gomes@embrapa.br; [2] Embrapa Solos, elaine.fidalgo@embrapa.br; [3] Universidade Federal Fluminense, oliveruff2@gmail.com; [4] Embrapa Solos, alba.leonor@embrapa.br; [5] Embrapa Solos, aline.oliveira@embrapa.br; [6] Embrapa Solos, mauricio.coelho@embrapa.br

**Thematic Session: 02: Legacy data: How turn it useful?**

**Abstract**

In this paper we evaluate the agricultural efficiency of the municipalities of the state Rio de Janeiro and assess the influence of contextual variables on the performance. The scores are computed by means of Data Envelopment Analysis (DEA) models, where inputs are land, labor and capital (or technology), and the value of crops and of livestock productions are the outputs. Covariates are related to soil evaluation (susceptibility to erosion and land use suitability) and to socioeconomic factors. The results show that high levels of susceptibility to erosion influence negatively and significantly the efficiency scores. The land suitability to agriculture and the land suitability to livestock are positively associated with performance. The presence of family-based farmers favors the agricultural performance of the assessed municipalities.

Keywords: Agricultural performance; Soil erosion susceptibility; Land use suitability; Data envelopment analysis; Fractional regression.

## Introduction

The farming sector in the state of Rio de Janeiro, Brazil, comprises mainly vegetables, fruits and grains productions. Dairy and beef cattle farming are also present in almost all the municipalities. These production chains strengthen the economy in the countryside (employment and income), boost rural communities and play an important role in food and nutritional security for the population of the State (EMATER-RIO, 2017, 2019). They are predominantly performed by family-based farmers (IBGE, 2019).

Given the importance of this activity, here we evaluate the agricultural efficiency of these municipalities by means of Data Envelopment Analysis (DEA) models. Input dimensions are proxies for land, labor and capital. Outputs are defined by the value of agricultural productions. These data were obtained from the 2017 Brazilian agricultural census (IBGE, 2019). In addition, we seek to identify exogenous variables that potentially affect performance. These covariates are variables from the PronaSolos databases, referring to soil evaluation (classes of soil susceptibility to erosion and to the different land use suitability), and socioeconomic factors from the 2017 census.

## Methodology

DEA (COOPER et al., 2007) is a mathematical programming approach that computes efficiency scores for a group of observations (so-called DMU). These measures are based on the level of resources used (inputs) and the results obtained (outputs) in a

production process. Each individual observation is optimized to estimate a piecewise linear efficient frontier, composed of the best practices from the sample (benchmarks).

Here we assume the variable returns to scale hypothesis (VRS) and output orientation, in accordance with other similar studies, as Souza et al. (2020). The envelope formulation of this DEA model is presented in (1), where $h_0$ is the efficiency score of DMU 0 under evaluation; $x_{ik}$ is the input $i$, $i=1...r$, of DMU $k$, $k=1...n$; $y_{jk}$ represents the output $j$, $j=1...s$, of DMU $k$; $x_{io}$ and $y_{jo}$ are the inputs $i$ and the outputs $j$ of the DMU 0; $\lambda_k$ is the contribution of DMU $k$ to the target of DMU 0 (benchmarks have non-zero $\lambda_k$).

$$Max\ h_0$$
$$subject\ to$$
$$x_{i0} - \sum_k x_{ik}\lambda_k \geq 0, \forall i; \quad -h_0 y_{j0} + \sum_k y_{jk}\lambda_k 0, \forall j; \quad \sum_k \lambda_k = 1; \quad \lambda_k \geq 0, \forall k$$

(1)

We considered three input dimensions: land, labor and capital. Land was defined as the sum of crops, forestry and livestock areas (hectares). Labor was represented by the total expenses on salaries (thousand R$). Capital, or technology, included expenses on different inputs, as services, fertilizers, seeds, pesticides, medicines for animals, salt, feed, transportation, electricity, machinery, fuels, among others (thousand R$). The outputs are the value of crops production (thousand R$) and the value of animal production (thousand R$). This approach allows municipalities with specialized production or with a good combined production arrangement to be efficient, as in Gomes et al. (2009). There were 89 municipalities with valid production data.

Due to the nature of DEA type responses (scores between 0 and 1), Ramalho et al. (2010) proposed fractional regression models to identify covariates that affect DEA scores. Let $z_j$ be the vector of covariates for municipality $j$. A fractional regression assumes $E\left(\tilde{h}_n\left(x_j, y_j \mid z_j\right)\right) = G(\delta' z_j)$, where $G(.)$ is a non-linear function with values in $(0,1]$ and $\delta$ is a vector of parameters. The model can be estimated by non-linear least squares or quasi-maximal likelihood.

The covariates used in the regression fit were: percentage of areas with susceptibility to erosion ($ps1 =$ very low; $ps2 =$ low; $ps3 =$ moderate; $ps4 =$ high; $ps5 =$ very high) (FERRAZ et al., 2021); percentage of areas with regular or restricted suitability for crops ($paplav$); percentage of areas with regular or restricted suitability for pastures and crops ($pappast$); percentage of areas with restricted suitability for forestry or unsuitable ($papoutra$) (CARVALHO FILHO et al., 2003); percentage of family-based farmers ($paf$); percentage of farmers that received technical assistance ($pat$); percentage of farmers that received credit/financing ($pfin$) (IBGE, 2019).

**Results and discussion**

Figure 1 shows the distribution of the efficiency scores. The average efficiency was 55.1%. This non-homogeneity in performance agrees with Souza (2019), regarding rural development in the State.
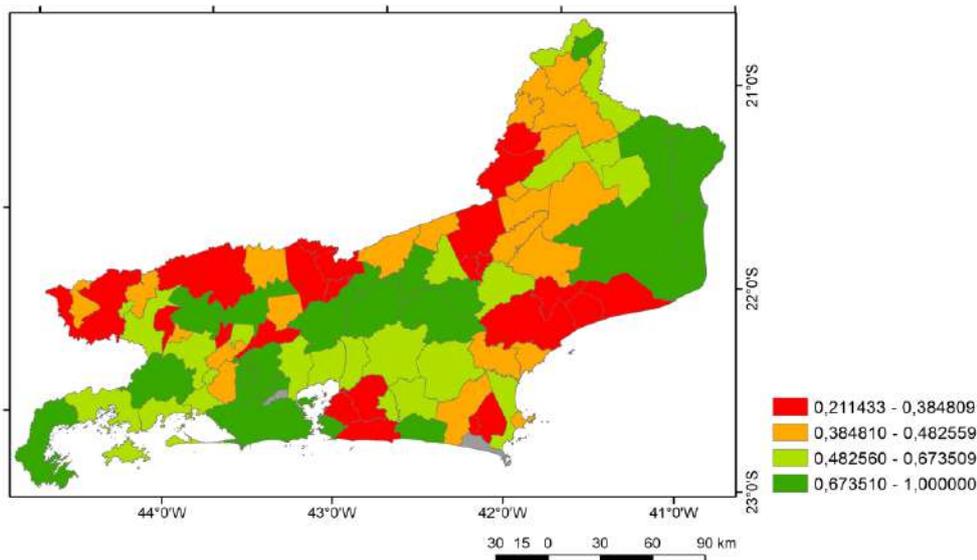
Figure 1: Geographical distribution of the DEA efficiency scores in Rio de Janeiro.

In Table 1 we can see the fractional regression fit. *pat* and *pfin* were not significant and were dropped from the model. Correlation between observed and predicted values is 72.1%. Municipalities with higher proportion of family-based farms are more efficient. The higher are the percentages of areas with moderate, high and very high susceptibility to erosion, the lower are the efficiency scores. Suitability to crops affects positively the efficiency. Suitability to livestock and crops has marginal positive effect.

Table 1: Fractional regression fit. Covariates were measured in log scale.

|  | Coefficient | Standard error | z | P>|z| | 95% Confidence interval | |
|---|---|---|---|---|---|---|
| *paf* | 1.235 | 0.342 | 3.61 | 0.000 | 0.565 | 1.904 |
| *ps1* | 0.020 | 0.100 | 0.20 | 0.843 | -0.175 | 0.215 |
| *ps2* | -0.247 | 0.204 | -1.21 | 0.225 | -0.647 | 0.152 |
| *ps3* | -0.797 | 0.280 | -2.85 | 0.004 | -1.346 | -0.248 |
| *ps4* | -0.343 | 0.096 | -3.58 | 0.000 | -0.530 | -0.155 |
| *ps5* | -0.433 | 0.186 | -2.33 | 0.020 | -0.797 | -0.068 |
| *paplav* | 0.167 | 0.063 | 2.65 | 0.008 | 0.044 | 0.291 |
| *pappast* | 0.364 | 0.224 | 1.62 | 0.105 | -0.076 | 0.804 |
| *papoutra* | -0.045 | 0.095 | -0.48 | 0.633 | -0.231 | 0.140 |
| constant | -1.552 | 1.690 | -0.92 | 0.359 | -4.864 | 1.761 |

**Conclusions**

The agricultural economic activity in the state of Rio de Janeiro has a medium overall performance. Soil evaluation factors are influential: moderate and higher levels of susceptibility to erosion affect significantly and negatively the performance; suitability to crops and to livestock productions have a positive association with performance. The presence of family-based production is also positive. These results may support (i) public policies related to soil governance (e.g., PronaSolos and the National Policy

for Soil and Water Conservation in Rural Environments), and (ii) the achievement of international agendas commitments at a national level, such as the Sustainable Development Goals (e.g., SDG 2) and the Global Soil Partnership.

## References

CARVALHO FILHO, A.; LUMBRERAS, J.F.; AMARAL, F.C.S.; NAIME, U.J.; SANTOS, R.D. dos; CALDERANO FILHO, B.; LEMOS, A.L.; OLIVEIRA, R.P.; AGLIO, M.L.D. **Avaliação da aptidão agrícola das terras do Estado do Rio de Janeiro.** Rio de Janeiro: Embrapa Solos, 2003. 32 p. (Boletim de Pesquisa e Desenvolvimento, 30).

COOPER, W.; SEIFORD, L.M.; TONE, K. **Data Envelopment Analysis: a comprehensive text with models, applications, references and DEA-Solver software**. New York: Springer, 2007. 489p.

FERRAZ, RPD; SIMÕES, M.; LUMBRERAS, J.F.; COELHO, M.R.; BACA, J.M.; FREITAS, P.L.;LIMA, E.P.; KUCHLER, P.C.; ALMEIDA, M.B.F. **Mapa de suscetibilidade dos solos à erosão hídrica do Brasil (Primeira aproximação)**. Disponível em: http://geoinfo.cnps.embrapa.br/documents/2916. Acesso em: 02 ago. 2021.

EMATER-RIO. **ASPA – Acompanhamento Sistemático da Produção Agrícola** (2019). Disponível em: http://www.emater.rj.gov.br/areaTecnica/cult2018.pdf. Acesso em: 02 ago. 2021.

EMATER-RIO. **Bovinocultura – Pecuária de Leite/Corte** (2017). Disponível em: http://www.emater.rj.gov.br/areaTecnica/RelBovi2017.pdf. Acesso em: 02 ago. 2021.

GOMES, E.G.; GREGO, C.R.; SOARES DE MELLO, J.C.C.B.; VALLADARES, G.S.; MANGABEIRA, J.A.C.; MIRANDA, E.E. Dependência espacial da eficiência do uso da terra em assentamento rural na Amazônia. **Produção**, v. 19, p. 417–432, 2009.

IBGE. Instituto Brasileiro de Geografia e Estatística. **Censo Agropecuário 2017** (2019). Disponível em: https://sidra.ibge.gov.br/pesquisa/censo-agropecuario/censo-agropecuario-2017. Acesso em: 19 jul. 2021.

RAMALHO, E.A.; RAMALHO, J.J.S.; HENRIQUES, P.D. Fractional regression models for second stage DEA efficiency analyses. **Journal of Productivity Analysis**, v. 34, p. 239–255, 2010.

SOUZA, G.S.; GOMES, E.G.; ALVES, E.R.A. Two-part. fractional regression model with conditional FDH responses: an application to Brazilian agriculture. **Annals of Operations Research**, 2020. DOI: https://link.springer.com/article/10.1007/s10479-020-03752-z.

SOUZA, R.P. O desenvolvimento rural no Estado do Rio de Janeiro a partir de uma análise multidimensional. **Revista de Economia e Sociologia Rural**, v. 57, n. 1, p. 109–126, 2019.

# Soil and rock depth mapping with aid of geomorphologic covariates in central Brazil

VIANA, J.H.M.[1], OTTONI, M.V.[2], CHARGEL, L.T.[3]

[1] EMBRAPA, joao.herbert@embrapa.br; [2] CPRM, marta.ottoni@cprm.gov.br, [3] CPRM, leotristaochargel@gmail.com

**Thematic Session: Legacy data: How turn it useful?**

## Abstract

Some hydrogeological models demand many spatially distributed data, which are often scarce or absent. Among them, the thickness of soil and saprolite and the depth to the rock basement are necessary for the understanding of the vadose zone processes. This work aimed to evaluate some interpolation procedures to map the soil and rock depth, based on the brazilian well database, and some geomorphometric covariates from the digital elevation model. The results indicate that the ranger procedure was the best ranked, but the soil depth results were much better than those for rock, measured by explained variance and mean relative error (RME).

**Keywords:** soil thickness, interpolation methods, SIAGAS, Carinhanha river, Verde Grande river

## Introduction

The use of models to aid the water resources' management is often precluded by the lack of data and the poor scale of the current environmental maps. This is specially true in central areas of Brazil, where the low density of geological and pedological information limits the reliability of the results. Nevertheless, the growing water scarcity is pushing the government agencies to adopt rational policies for water management, and they must to be based on sound data. This work is part of a project to supply such basic data, and is focused on some of the core properties demanded by the hydrogeological models that are being developed for critical areas in central Brazil.

## Methodology

The study area is located in central Brazil, in the states of Minas Gerais and Bahia. It is drained by two tributaries of the São Francisco river, the Verde Grande and Carinhanha rivers (Figure 1). The regional climate is a transition from dry tropical to semiarid. The main soil types are Oxisols and Ultisols, according to the Brazilian Soil Map
(http://geoinfo.cnps.embrapa.br/layers/geonode%3Abrasil_solos_5m_20201104).
The main hydrogeological domains are associated with the cretaceous sandstones of the Urucuia and Areado formations and with the karstic rocks of the Bambuí Group. The karstic features are widespread in the Verde Grande watershed, and are a challenge for the modelling of the aquifers. A set of 6238 wells were selected from the Brazilian Underground Water Information System (SIAGAS database http://siagasweb.cprm.gov.br/layout/), and contains the information of location and lithological description of the well's profiles. The well profile layers were classified as

soil, saprolite and rock, considering some key words in the lithological description. A set of rules was created for the determination of the soil lower limit and the rock depth, applied by a visual basic script. After the classification was done, the result was checked for all well profiles of the database by a soil expert. The classification was also done manually for those profile layers that could not be classified by the script. We are aware that this classification is strongly influenced by the quality of the data of the well profiles, which contain imprecisions, due to the lack of standards for description. A subset of the original SIAGAS's dataset was chosen in a polygon around the study area, and 2031 wells were used for the analyses. Additionally, a set of rock outcrop data was added to the data (+650 datapoints). For the soil layer data, a confidence interval based on all the soil data was used and the points with soil layers thicker than the 95% interval were discarded (soil depth bigger than 24 m). The interpolation was performed in a regular grid of 500 x 500 m, where the covariates were measured. The covariates were a set of ALOS_PALSAR's Digital Elevation Model geomorphometric derivatives, generated in QGIS 3.20.3-Odense (https://www.qgis.org). The best correlated covariates were chosen, based on the R2 and the p value (Table 1), and with minimum correlation in between. The interpolation was performed in R environment, using the packages "sp", "sf", "stars", "ggplot2", "GSIF", "randomForest", "ranger", "quantregForest" and "raster" and the best model was chosen based on the mean relative error (RME) and variance results. The outputs were the soil and rock depth maps and their variance counterparts.

## Results and discussion

The results of the correlation analyses indicate that the covariates have low values of $R^2$, though they are statistically significant with very low p values (Table 1), for both soil and rock depths. The soil depth map and its variance map are plotted in the Figure 2. The result of the model validation shows that the "ranger" model performed much better than the other models, explaining 70% of the variance and with a lower RME (Table 2). The rock depth map and its variance map (Figure 2) also were best fit by the "ranger" model, but with just a small difference among the models and a relatively poor fit (35%) and higher RME (Table 2). We speculate that the high number of rock outcrops, specially in the karstic area of the Verde Grande basin, may have an impact on the overall model performance, precluding better results and producing a result that can be seen only as a general trend.
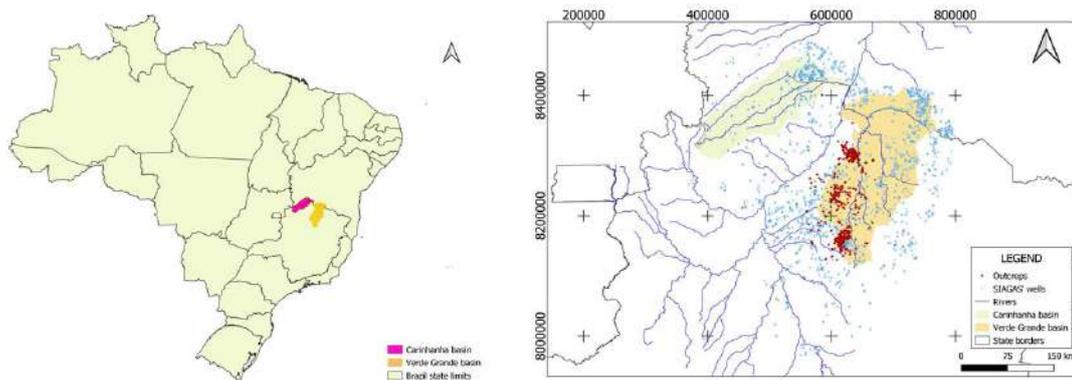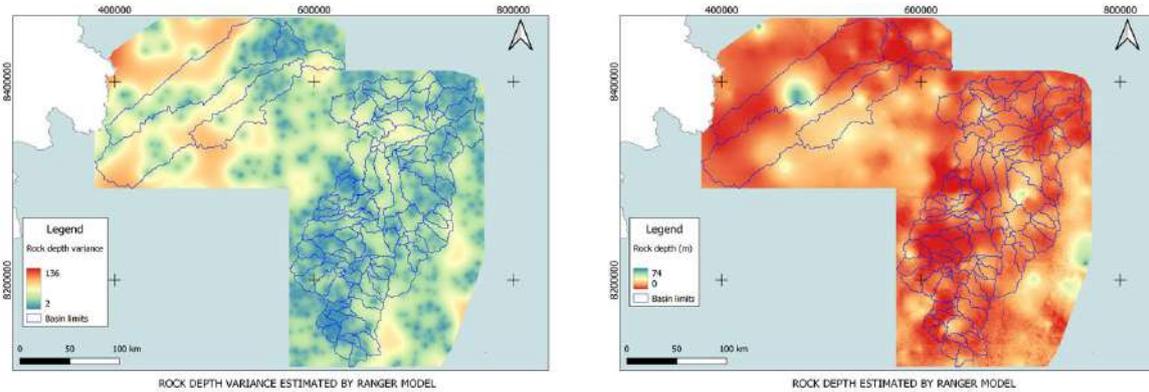
Figure 1 – Location of the watersheds in Brazil and the distribution of the wells and



ROCK DEPTH VARIANCE ESTIMATED BY RANGER MODEL

ROCK DEPTH ESTIMATED BY RANGER MODEL

outcrops in the study area. DATUM SIRGAS 2000, UTM 23S.

Figure 2 – Soil depth map and soil depth variance map. DATUM SIRGAS 2000, UTM 23S.



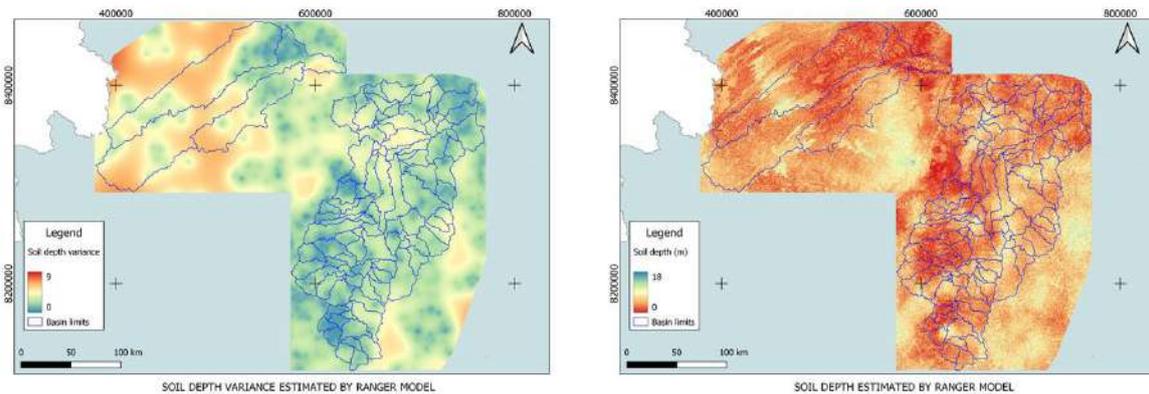SOIL DEPTH VARIANCE ESTIMATED BY RANGER MODEL

SOIL DEPTH ESTIMATED BY RANGER MODEL

Figure 3 – Rock depth map and rock depth variance map. DATUM SIRGAS 2000, UTM 23S.

Table 1 – Results for the best correlated attributes, used in the model fitting.

Rock depth

| Attribute | $R^2$ | p value |
|---|---|---|
| Terrain Ruggedness Index | -0.1700 | 0.0000 |
| Latitude | -0.1000 | 0.0000 |
| Valley Index | -0.0680 | 0.0021 |

Soil thickness

| Attribute | $R^2$ | p value |
|---|---|---|
| Altitude | 0.1200 | 0.0000 |
| Multiresolution Index of Ridge Top Flatness (MRRTF) | 0.0830 | 0.0004 |
| Minimal Curvature | 0.0690 | 0.0006 |
| Slope Index | 0.0680 | 0.0009 |

Table 2 – Results of the evaluation of the models.

Rock depth

| Method | RME, meter (validation) | Var explained |
|---|---|---|
| "ranger" | 11.15 | 35.1% |
| "random Forest" | 11.23 | 34.2% |
| "quantregForest' | 11.54 | 30.4% |
| "rpart" | 11.71 | 28.5% |
| "GLM" | 11.82 | 27.1% |

Soil thickness

| Method | RME, meter (validation) | Var explained |
|---|---|---|
| "ranger" | 3.001 | 70.4% |
| "random Forest" | 4.444 | 35.0% |
| "quantregForest' | 4.612 | 30.0% |
| "rpart" | 4.694 | 27.8% |
| "GLM" | 4.724 | 26.6% |

## Conclusions

The use of geomorphologic covariates for the interpolation of soil and rock depth produced better results for soil than for rock. The "ranger" algorithm was the best ranked, but the results for rock depth were poor, which could be the effect of a large number of rock outcrops in the karstic region of the river Verde Grande.

## Acknowledgements (optional)

## References

HENGL, T. 2019. **GSIF: Global Soil Information Facilities**. R package version 0.5-5/r245. https://R-Forge.R-project.org/projects/gsif/

LIAW, A. and WIENER, M. (2002). Classification and Regression by randomForest. **R News** 2(3), 18–22.

MEINSHAUSEN, NICOLAI (2017). **quantregForest: Quantile Regression Forests**. R package version 1.3-7. https://CRAN.R-project.org/package=quantregForest

R Core Team (2021). **R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria**. URL https://www.R-project.org/.

WRIGHT, MARVIN N., ZIEGLER, ANDREAS (2017). ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. **Journal of Statistical Software**, 77(1), 1-17. doi:10.18637/jss.v077.i01

# Rescue of saturated hydraulic conductivities and steady Infiltration rates data in the State of Rio de Janeiro - Brazil

PIMENTEL, Leticia Guimarães[1]; OTTONI, Marta Vasconcelos[2]; CAMPOS, Pablo Nieto[3]; HUF DOS REIS, Aline Mari[4] ; MARTINS, Alba Leonor[5]; TEIXEIRA, Wenceslau Geraldes[5]

[1]ZARC - Embrapa Solos - RJ, Brazil, leticia.pimentel@colaborador.embrapa.br; [2]CPRM - RJ, Brazil, marta.ottoni@cprm.gov.br; [3]CPRM - RJ, Brazil, pabloncampos12@gmail.com; [4]ZARC - Embrapa Solos - RJ, Brazil, aline.huf@@colaborador.embrapa.br; [5]Embrapa Solos, RJ, Brazil alba.leonor@embrapa.br; [5]Embrapa Solos - RJ, Brazil, wenceslau.teixeira@embrapa.br

## Abstract

Saturated hydraulic conductivity and steady infiltration rate are essential parameters in irrigation, and drainage projects, as well as in hydrological and climatological modelling. These data are hardly found in widely used soil databases in Brazil, and they are still reported in dispersed publications. This study aims to rescue SIR and $K_s$ data for the state of Rio de Janeiro and make them available to users. In total 934 measurements of $K_s$ and 24 of SIR were compiled. The soil types most frequent measured were Gleissolos and Neossolos, as well agriculture and pastures for the land-use system. The most popular methods to evaluate the $K_s$ were constant head in the lab and the Guelph permeameter in field evaluations, and for the SIR, the double ring method. The data ranged from 0.40 to 1073 mm h$^{-1}$. The lack of information in some regions of the Rio de Janeiro States indicates priority areas for increasing $K_s$ and SIR determination.

Keywords: constant head, Guelph permeameter, double ring

## Introduction

Evaluating soil water dynamics involves the determination of soil hydraulic parameters such as saturated hydraulic conductivity ($K_s$) and steady infiltration rate (SIR). These parameters have been largely used in hydrological models, irrigation, and drainage projects, as well as in studies related to the fate of nutrients and pesticides and water erosion. The saturated hydraulic conductivity ($K_s$) characterizes the capacity of the soil to transmit water in saturated conditions. Its measurement can be done in the laboratory using methods developed under transient conditions or under steady-state conditions, or even under field conditions, with the popular use of the Guelph permeameter. However, the determination of soil hydraulic properties is costly and time-consuming, and very difficult to be evaluated in large areas due to the high spatial variability of these parameters. SIR and $K_s$ data for the state of Rio de Janeiro have been retrieved and compiled. This study is part of a national effort of the Brazilian Soil Science Society (SIBCs) to rescue and make available the data of SIR and $K_s$ to Brazil.

## Methodology

The study was performed by obtaining data on basic SIR and/or $K_s$ obtained from lab or field evaluations in soils located in the state of Rio de Janeiro - Brazil. These studies were found by a bibliographic survey of articles, theses and dissertations, conference announcements, technical publications, and reports. The SIR and $K_s$ data were stored in a soil database (SDB) structured including other information related: sampling location (site description, geographic coordinates, land use), physical, chemical properties, and water retention at different matric potential. The measuring methods used to evaluate SIR and $K_s$ are also included in the SDB. Some parameters were mandatory, such as the county, soil classification, land use, granulometry, saturated hydraulic conductivity ($K_s$ – mm/h) (obtained in laboratory or in the field) or steady infiltration rate (SIR – mm/h) and methods for determination of $K_s$ or SIR. The land-use systems and soil classification were harmonized using the rules of classification used in the Project MapBiomas (MAPBIOMAS, 2020) and soil classification using the first order of classification of the last version of the Brazilian Soil Classification System (Santos, 2018). The data were classified according to land use and soil type. Descriptive statistics of mean, maximum, and minimum values for soil types in combinations with land use were calculated.

**Results and discussion**

The results totalizing 934 data of $K_s$ and 24 data of SIR (Bernardes, 2005; Bhering 2007; Duarte, 2004; Fabian, 1997; Instituto do Açúcar e do Álcool, 2011; Nacinovic, 2013; Silva, 2011). The county of Campos dos Goytacazes does have probably the largest Ks and SIR data base in Brazil. The most frequent soil types evaluated were Gleissolos (26%) followed by Neossolos (26 %) and Cambissolos (22%). Agriculture and pastures were the land uses systems more evaluated for SIR and $K_s$ in Rio de Janeiro (Figure 1)
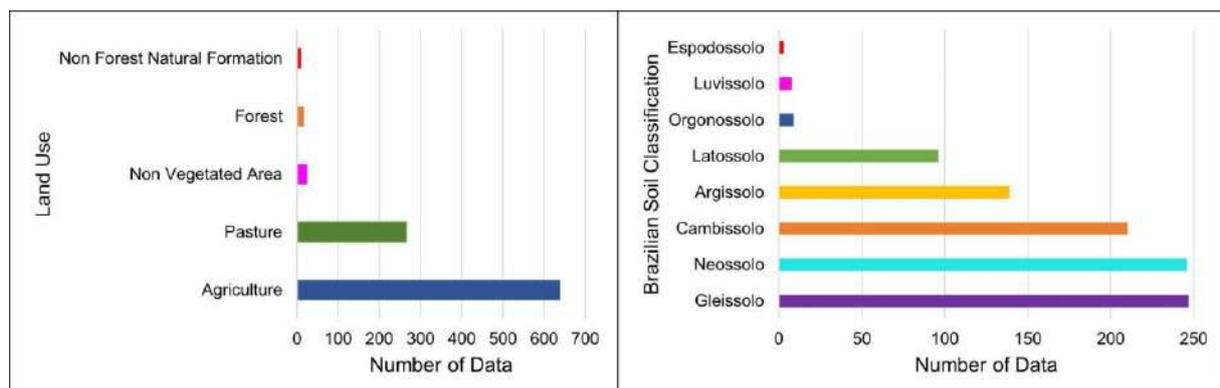


Figure 1 - Number of available data of Saturated Hydraulic Conductivity ($K_s$) or Steady Infiltration Rate (SIR) for different soil type and land uses in the State of Rio de Janeiro - Brazil.

The most frequent method to evaluate de $K_s$ was the constant head soil core for lab and well permeameters (i.e Guelph permeameter) for field evaluations. To evaluate the SIR, in a total of 24 data, the most frequently method was the double ring. The

highest $K_s$ value was 1073 mm h$^{-1}$ measured in a Neossolo and lowest values (0,4 mm h$^{-1}$) in a Neossolos and Latossolo covered with pastures. Espodossolos show the minimum value of $K_s$ 249 mm h$^{-1}$. The $K_s$ values > 500 mm h$^{-1}$ repeated in soil types were a consequence of the maximum rate of the method used. The SIR evaluations were also concentrated in Gleissolos and Argissolos. The lowest values 1 mm h$^{-1}$ G were measured in Gleissolos covered by pastures. However, Gleissolos also showed a SIR of 114 mm/h with an average of 140 mm/h (Table 1).

**Conclusions**

The State of Rio de Janeiro has more than 900 measurements available for $K_s$ or SIR. However, the most part of the data are concentrated in Gleissolos and Neossolos in the north part of the state near the cost. The soil covered by original Atlantic Forest is poorly measured. The most popular methods for $K_s$ measurements were constant head soil core and well permeameters in the field. The $K_s$ and SIR are valuable data for many applications. This rescue of dispersed data may be useful for many projects. The lack of information about hydraulic information in some regions or soil may indicate priority areas for evaluations. This soil database will be available together with the data from other states of Brazil in a public data bank.

**References**

BERNARDES, R. S. **Condutividade Hidráulica de Três Solos da Região Norte Fluminense**. 2005. 80 f. Mestrado -UENF, Campos dos Goytacazes - RJ. 2005.

BHERING, S. B. **Influência do Manejo do Solo e da Dinâmica da Água no Sistema de Produção do Tomate de Mesa:** subsídios à sustentabilidade agrícola do Noroeste Fluminense. 2007. 232 f. Tese- UFRJ, RJ. 2007.

DUARTE, A. P. L. **Avaliação de Propriedades Termo-Hidráulicas de Solos Requeridas na Aplicação da Técnica de Dessorção Térmica**. 2004. 290 f. Tese – PUC- RJ. 2004.

FABIAN, A. J.; OTTONI FILHO, T. B.; Determinação de Curvas de Infiltração Usando Uma Câmara de Fluxo. **R. Bras. Ci. Solo**, Campinas, v. 21, p. 325-333, 1997.

INSTITUTO DO AÇÚCAR E DO ÁLCOOL. **Projeto de irrigação e drenagem da cana-de-açúcar na região Norte Fluminense (Projir)**. 1983. Rio de Janeiro: Sondotécnica, 1983. 9 v.

LEAL, I. F. **Classificação e Mapeamento Físico-Hídricos de Solos do Assentamento Agrícola Sebastião Lan II, Silva Jardim - RJ**. 2011. 136 f. Dissertação – UFRJ - RJ, Rio de Janeiro. 2011.

MAPBIOMAS. **Código das Classes da Legenda e Paleta de Cores Utilizadas na Coleção 5 do MapBiomas.** Disponível em: https://mapbiomas-br-

site.s3.amazonaws.com/downloads/C%C3%B3digos_das_classes_da_legenda_e_p aleta_de_cores.pdf. Acesso em: 05 out. 2021.

NACINOVIC, M. G. G. **Avaliação de Erosão Hídrica Superficial em Parcelas Experimentais**. 2013. 155 f. Tese – UFRJ – RJ. 2013.

SILVA, A. P. **Influência da Forma e Posição da Encosta nas Características do Solo e na Regeneração Natural de Espécies Florestais em Áreas de Pastagens Abandonadas**. 2011. 79 f. Dissertação - - UFRRJ – Seropédica. 2011.

SANTOS, H. G. et al. **Sistema Brasileiro de Classificação de Solos.** 2018. Brasília, DF: Embrapa, 2018.

Table 1 - Descriptive statistics of saturated hydraulic conductivity ($K_s$) and steady infiltration rate (SIR) for different soil types and land used in the state of Rio de Janeiro.

| Soil Type | Land Use | | $K_s$ [mm h$^{-1}$] | | SIR [mm h$^{-1}$] | | |
|---|---|---|---|---|---|---|---|
| SiBCs | Mapbiomas | | average | maximum | minimum | average | maximum |
| Gleissolo | Agriculture | 1 | 29 | >500 | 7 | 140 | 338 |
| | Pasture | 1 | 38 | >500 | 1 | 57 | 114 |
| | Non forest | 2 | 39 | 161 | - | - | - |
| Neossolo | Agriculture | 1 | 156 | 1073 | - | - | - |
| | Pasture | 0,4 | 403 | >500 | - | - | - |
| | [1]Non forest | 1 | 9 | 31 | - | - | - |
| Cambissolo | Agriculture | 1 | 42 | 748 | - | - | - |
| | Pasture | 1 | 23 | >500 | - | - | - |
| Argissolo | Agriculture | 3 | 131 | >500 | 17 | 48 | 83 |
| | Pasture | 2 | 117 | 819 | - | - | - |
| | Non forest | 1 | 8 | 30 | - | - | - |
| Latossolo | Agriculture | 38 | 113 | 226 | - | - | - |
| | Pasture | 0,4 | 27 | 177 | - | - | - |
| | Forest | 17 | 316 | 998 | - | - | - |
| Organossolo | Agriculture | 12 | 67 | 152 | - | - | - |
| | Pasture | 44 | 66 | 88 | 98 | 98 | 98 |
| Luvissolo | Pasture | - | - | - | 12 | 74 | 180 |
| Espodossolo | Agriculture | 249 | 416 | >500 | - | - | - |

1 - Non forest natural vegetation (Mapbiomas classification system. SiBCS – Sistema Brasileiro de Classificação de Solos.

# REDUCING COMPLEXITY OF INPUT DATASET TO MAP $Fe_2O_3$, Nb AND $TiO_2$ CONTENTS IN MORRO DOS SEIS LAGOS-AM, BRAZIL

RODRIGUES, Niriele Bruno[1]; SILVA, Júlio Cesar Lopes da [2]; SILVA, Renan Pereira Marinatti da [3]; PINHEIRO, Helena Saraiva Koenow [4]

[1] UFRRJ, nirielebr@ufrrj.br; [2] UFRJ, jlopes@geologia.ufrj.br; [3] UFRRJ, marinatti.renan@gmail.com; [4] UFRRJ,lenask@gmail.com.

**Thematic Session: Legacy data: How turn it useful?**

**Abstract**

The study goal is to assess the performance of recursive feature elimination (RFE) to reduce the covariates from input dataset to predict $Fe_2O_3$, Nb and $TiO_2$ contents through Gradient Boosted Machine (GBM) and Random Forest (RF), in Morro dos Seis Lagos-AM, Brazil. The input dataset gathers 344 sample points (soil, sediments and rock materials) with topographic covariates and remote sensing data, from Sentinel-2A. The best performance to modeling the elements was achieved with RFE and RF model (Nb $R^2$=0.19, $TiO_2$ $R^2$=0.19 and $Fe_2O_3$ $R^2$=0.24).

**Keywords:** Pedometrics; Soil chemical attributes; Data mining.

**Introduction**

In the last decades, machine learning models has been widely used in mineral prospection. These techniques can support mineral prospection of Fe, Nb, Ti to attend the increasing demand of this elements for industrial and energetical purposes (MITCHELL, 2015). Nowadays, the availability of multiple geographic data is useful for geological mapping and mineral prospection, and contributes for high-quality and low-cost predictions. Recent studies by Cracknell et al. (2014), Costa *et a*l. (2019), Pinheiro (2021) and Wang *et al.* (2021) have used machine-learning techniques in geological analysis. When multiscale hyperspectral sensing techniques were integrated it tend to upgrade conventional surveys techniques.

Morro dos Seis Lagos complex contain a thick laterite crust (>200 m), where weathering processes of siderite carbonatite produced a goethite/hematite crust (CPRM, 2019). Carbonatites and laterites host trace minerals with elements that have economic potential (as Nb and $TiO_2$, for example). This occurrence is located in Amazon biome under dense forest on environment protection areas, with difficult access. The goal of the study was to assess the performance of REF and Stepwise methods to select covariates to predict $Fe_2O_3$, Nb and $TiO_2$ contents through of data mining methods (GBM and RF) at Morro dos Seis Lagos, São Gabriel da Cachoeira, Amazonas, Brazil.

**Methodology**

This research used 344 samples from the geochemical dataset from the Geological Survey of Brazil (CPRM), available on the GeoSBG platform (CPRM, 2019). A 20m spatial resolution hydrologically consistent digital elevation model (DEM) was obtained by interpolation of vectorial data (contour lines, quoted points and hydrography), from the cartographic database of Brazilian Institute of Geography and Statistics (IBGE)- with scale 1:25.000. The terrain covariates derived from the DEM were generated in SAGA-GIS v.2 .1.2 software (CONRAD, 2007), and they are:Catchment Area; Surface Area; Aspect; Flow Accumulation; Curvature; Profile Curvature; Plan Curvature; Slope; Geomorphons; LS Factor; Topographic Wetness Index (TWI); Convergence Index; Relative Slope (RSP); Terrain Ruggedness (TRI); Valley Index. The remote sensing data from the MSI Sentinel-2A sensor were used, as well derived indexes from spectral bands: Ferrous Silicates, Laterite, Iron Oxide and Gossan (VAN DER MEER et al., 2014).

As a first criteria to select input covariates a Spearman correlation analysis was performed to eliminate highly correlated covariates, with a threshold of 0.95 (KUHN et al., 2013). Excluding Topographic Wetness Index, Mass Balance Index, Multiresolution Lower Valley Flatness Index (MRVBF), B7, B8, B5, Terrain Ruggedness Index (TRI), Slope degre, Longitudinal Curvature, Ferrous Silicate, once they present self-correlation (0.95 threshold). Subsequently, Recursive Feature Elimination -RFE (JEONG et al. 2017) method was tested to select an optimized covariate dataset, aiming to reduce the dimensionality of the input covariates for each mapped element. To assess the performance of the predictive models, leave-one-out cross-validation (LOOCV) method was applied, via the Caret package (KUHN et al., 2017).

Gradient Boosted Machine (GBM) and Random Forest (RF) were applied in the spatial modelling procedures through gbm and randomForest packages (LIAW and WIENER, 2018; GREENWELL, et al. 2019) in R environment. The RF hyperparameters were 500 mtry (default) and 1/3 of the set of covariates for each element; and GBM hyperparameters were adjusted as Gaussian parameter (square error), number of trees equal 300 (n.trees), interaction depth equal to 5, shrinkage equal 0.3, and for bag.fraction 0.1 was used. A synthesis of methodological procedures is presented in Figure 1.
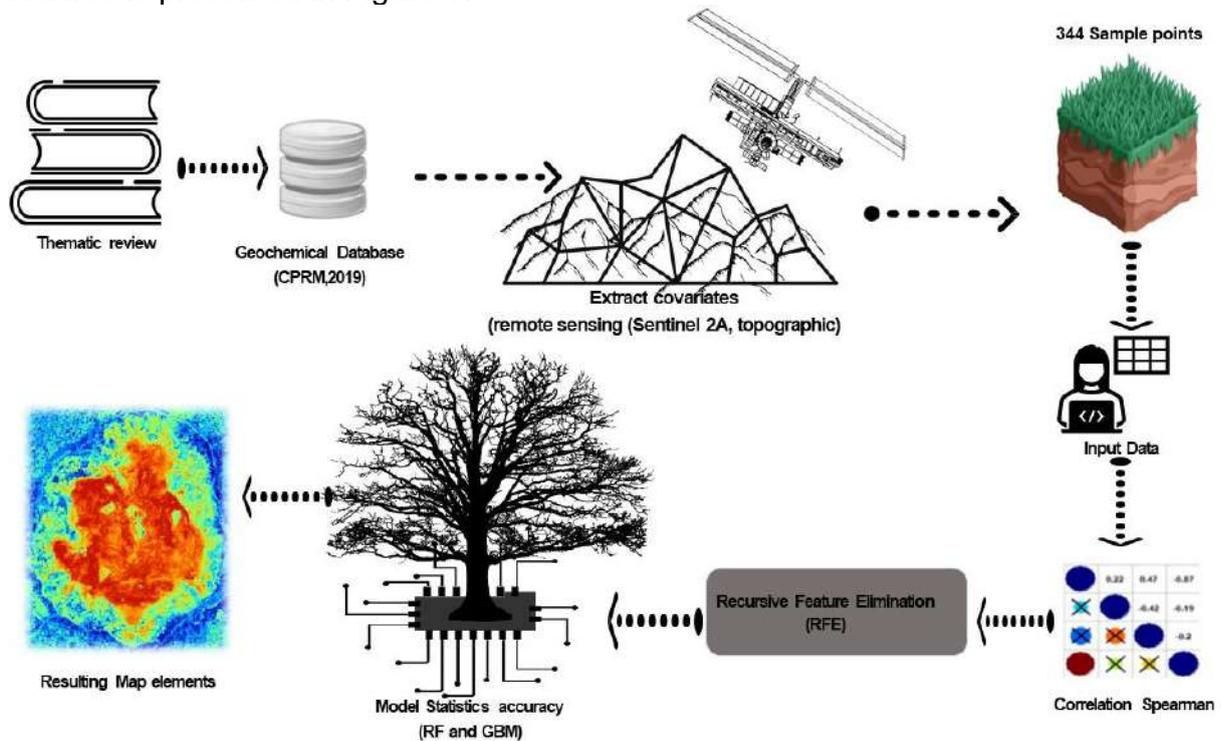


**Figure 1.** Flowchart of the methodological procedures.

### Results and discussion

In general, the predictor covariates for both models presented heterogeneity in the composition of their sets (Table 1). According to the degree of importance to predict the elements, it was notice that the terrain covariates presented greater influence when compared with indexes from Sentinel -2A sensor.

The models presented similar behavior, for the importance of covariates through RFE. It is probably due the similar structure from tree-based models, adopting similar predictive variables based on the importance ranking (LIAW and WIENER, 2012; GREENWELL, et al. 2019). In this study, both models showed similar covariates and validation values to accuracy indexes to predict $Fe_2O_3$, Nb AND $TiO_2$. Wang et al. (2020) in a study for detection of geochemical anomalies in the Jingdezhen region, found good performance by using this hybrid

approach (RFE-RF) to select covariates to map the anomalies.

In the GBM model, the covariates Elevation, SAGA Wetness Index and Gossan showed more importance. The importance of covariates ranked by Random Forest showed that Elevation, Terrain surface and Gossan, were more important to predict the elements.

**Table 1.** Performance of the RFE and Spearman with Random Forest (RF) and Gradient Boosted Machine (GBM) models to predict Nb, $TiO_2$, and $Fe_2O_3$.

| Elements (% Wt) | Methods | Selected Covariates | Model | RMSE | MSE | $R^2$ |
|---|---|---|---|---|---|---|
| **Nb** | RFE | Elevation +Standardiz height + Aspect + Terrain Surface + Mid- Slope Positon + B11 | GBM | 0.81 | 0.65 | 0.17 |
| | | | RF | 0.75 | 0.57 | 0.19 |
| | Only Spearman | Terrain Surface + SAGA Weteness Index + MRRTF + B4+B3 | GBM | 1.05 | 1.10 | 0.003 |
| | | | RF | 0.83 | 0.70 | 0.05 |
| **$TiO^2$** | RFE | SAGA Weteness Index+ Elevation + Real Surface Area+ Terrain Surface | GBM | 4.08 | 16.65 | 0.15 |
| | | | RF | 3.68 | 13.71 | 0.19 |
| | Only Spearman | Vector Terrain Ruggedness (VRM) + Standardiz height + SAGA Weteness Index + Normalized+ MRRTF+ Ferric Oxide + Elevation+ Flow line Curvature | GBM | 4.17 | 17.40 | 0.12 |
| | | | RF | 3.76 | 14.19 | 0.15 |
| **$Fe_2O_3$** | RFE | Elevation + SAGA Weteness Index + Gossan + Vector Terrain Ruggedness (VRM) | GBM | 22.75 | 517.65 | 0.17 |
| | | | RF | 19.90 | 398.10 | 0.24 |
| | Only Spearman | Aspect+ Vector Terrain Ruggedness (VRM) + Terrain Surface + Terrain1 + Standardiz height +Real Surface+ Normalized+ Mid- Slope Positon + Elevation+ Curvature total+ Tangential curvature +Curvature profile+ Curvature plan+ Curvature maximum+ General curvature + +Convergence Index + Cross-Sectional Curvature +B4+B3+B2 | GBM | 20.72 | 429.47 | 0.24 |
| | | | RF | 19.41 | 376.94 | 0.28 |

RF: Random Forest; BRT: Boost regression tree; RMSE: Root mean square error; MSE: Mean square error; $R^2$: Rsquared.

In general, models related to RFE, showed satisfactory performance according the accuracy indexes obtained from cross-validation. The Random Forest model, presented better performance through RFE selection (Table 1). Regarding the use of RFE, both methods to reported similar variability to fit the observed values according to the metrics (RMSE and MSE).

The results showed that RFE optimized the prediction of the elements to both machine learning models. The method works eliminating covariates recursively by interaction between the covariates and the predicted variable, considering collinearity among the covariates (multicollinearity) (SVETNIK *et al.*, 2004).

## Conclusions

The Random Forest model showed better performance to predict the elements, indicating as important covariates: Elevation, Gossan and SAGA Wetness Index, which were pre-selected through recursive feature elimination (RFE) method. In this sense, future research should to address other spectral indices from remote sensing to improve the resulting models, as well other predictive algorithms, as artificial neural networks for example.

# References

CONRAD, O. SAGA – **System für Automatisierte Geowissenschaftliche Analysen**. Presentation held at FOSSGIS 2007, Berlin, Germany, 14. March 2007.

COSTA, I.S.L. et al. Predictive lithologic mapping using machine learning methods: a case study in the Grey Lineament, Carajas Province, Brazil. **Journal of the Geological Survey of Brazil**, [S. l.], v. 2, n. 1, p. 26-36, 2019.

CPRM. **Project Assessment of the Rare Earth Potential in Brazil: Morro dos Seis Lagos area, Northwest Amazonas** / Organized by Lucy Takehara. 4. Ed. Brasília: CPRM, 2019.112 p.

CRACKNELL, M. J. et al. Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. **Computers & Geosciences**, v. 63, p. 22-33, 2014.

GREENWELL, B. et al. **Package 'gbm'**. R package version, v. 2, n. 5, 2019. Disponível em: https://CRAN.R-project.org/package=gbm

JEONG, G. et al. Spatial soil nutrients prediction using three supervised learning methods for assessment of land potentials in complex terrain. **Catena**, v. 154, p. 73–84, 2017.

KUHN, M. et al. **Caret**: classification and regression training, 2017. Disponível em: https://cran.r-project.org/package=caret

KUHN, M. et al. **Applied predictive modeling**. New York: Springer, 2013.

LIAW A, WIENER M. **Classification and regression with random forest**. R package version 4.6-14; 2018. Disponível em: https://CRAN.R-project.org/

MITCHELL, ROGER H. Primary and secondary niobium mineral deposits associated with carbonatites. **Ore Geology Reviews**, v. 64, p. 626-641, 2015.

PINHEIRO, Helena Saraiva Koenow. DIGITAL MAPPING OF P-LI-BE CONTENTS: A PROPOSAL TO RECOGNIZE LCT TYPE PEGMATITES. In: TOLEDO, Catarina; MA, Julia B. Curto; BARBOSA, Paola (ed.). CONTRIBUIÇÕES: à ciência e técnica dos (as) jovens geólogos (as) brasileiros (as). Brasília: Anais 50° Congresso Brasileiro de Geologia, 2021. Cap. 18. p. 03-378.

SVETNIK, V. et al. Application of Breiman's random forest to modeling structure-activity relationships of pharmaceutical molecules. **In: International Workshop on Multiple Classifier Systems**. Springer, Berlin, Heidelberg, p.334-343, 2004.

WANG, C. et al. Indicator element selection and geochemical anomaly mapping using recursive feature elimination and random forest methods in the Jingdezhen region of Jiangxi Province, South China. **Applied Geochemistry**, v. 122, p. 104760, 2020.

VAN DER MEER, F. D. et al. Potential of ESA's Sentinel-2 for geological applications. **Remote sensing of environment**, v. 148, p. 124-133, 2014.

# Spatial prediction of soil pH in a grape growing region of Southern Brazil

SILVA, Ryshardson G. P. O.[1]; GIASSON, Elvio[2]; MENEZES, Michele D.[3]; SENA, Antonny F. S.[2]; SILVA, Saulo G.[2]

[1] Department of Soil Science, Universidade Federal do Rio Grande do Sul, rgeovane@outlook.com.br; [2] Department of Soil Science, Universidade Federal do Rio Grande do Sul; [3] Department of Soil Science, Universidade Federal de Lavras

**Thematic Session: Legacy Data: How turn it useful?**

## Abstract

The present study aimed to create spatial predictive models for soil pH $H_2O$ in an important grape growing region from the south of Brazil, assessing the relationships between this soil attribute and continuous environmental covariates. The study was carried out in the Vale dos Vinhedos, an 81,180 km² wine-producing region located in the Rio Grande do Sul State (Brazil). The soil dataset was obtained from "Os Solos do Vale dos Vinhedos" project. The results of soil pH analysis of 163 soil profiles were used in this study, in three fixed depth intervals: 0–5, 5–15, 15–30 cm. A set of 25 morphometric maps derived from a DEM, 2 vegetation indexes derived from a Sentinel-2 image, soil classification and 5 climate variables from the WorldClim Data Portal were used as covariates. Linear regression (LM) and random forest (RF) were performed, and the accuracy of each predictive model were evaluated. The best performance was achieved with the random forest models to predicting soil pH between 0 and 30 cm in this study, with the best values of RMSE and MAE.
Keywords: Regression Kriging; Linear Model; Random Forest; Vineyard.

## Introduction

The Rio Grande do Sul State is one of the most relevant producers of grapes and wines of Brazil (MELLO; MACHADO, 2020). The Vale dos Vinhedos (Vineyard Valley), located in the Rio Grande do Sul State, has national recognition as one of the most traditional wine producing centers in Brazil (TONIETTO et al., 2012, 2013). Some soil attributes, such as pH, have a direct influence on the composition and quality of wines (COIPEL et al., 2006). Knowledge of its spatial distribution is necessary both to improve the management of vineyard sites, as well as to understand the functioning of the soil in the ecosystem and indicate how it can be better managed. (ODEH et al., 2007).

For pedometrics, the soil variability can be investigated in two ways: through non-geostatistical techniques (multiple linear regression, generalized additive model, etc.) or using geostatistical techniques (ordinary kriging, simple kriging, etc.) (McBRATNEY et al., 2000; ODEHA et al., 1994). The combination of these methods (hybrid methods) often results in more accurate local predictions of soil attributes (GOOVAERTS, 1999), as they use secondary information available as auxiliary variables in making the predictions (YIGINI; PANAGOS, 2016; ZHU; LIN, 2010).

The present study aimed to create spatial predictive models for soil pH H2O in an important grape growing region of Brazil, assessing the relationships between this soil attribute and continuous environmental covariates. Additionally, it was also sought to apply appropriate DSM functions to map this attribute.

**Methodology**

The study was carried out in the Vale dos Vinhedos, an 81,180 km² wine-producing region (29.08–29.14° S latitude, 51.29–51.37° W longitude) located in the Rio Grande do Sul State (Brazil). The climate is classified as Cfb, subtropical with a mild summer (ALVARES et al., 2014). The soil dataset were obtained from the project: "Os Solos do Vale dos Vinhedos" (FLORES et al., 2012). Sampling was done with 163 soil profiles described and analyzed following Brazilian standard methods (EMBRAPA, 1997, 2006). The soil pH $H_2O$ results were used in three fixed depth intervals (0–5, 5–15, 15–30 cm), following GlobalSoilMap.Net specifications (ARROUAYS et al., 2014). For this, an interpolation with an equal-area spline function was performed (MALONE et al., 2009).

In total, 88 covariates were selected to model the spatial distribution. The covariates were: morphometric maps derived from a 5 x 5 m grid resolution DEM (FLORES et al., 2012), to be listed: analytical hillshading (AH), aspect (AS), convexity (CV), direct insolation (DI), elevation (EL), flow direction (FD), LS factor (LS), curvature (maximal (Cmax), minimal (Cmin), total (CT), plan (CPL) and profile (CPF)), mass balance index (MBI), mid slope position (MSP), multi-resolution valley bottom flatness (MRVBF), normalized height (NH), overland flow distance to channel network (OFD), slope (SL), slope height (SH), standardized height (STH), topographic position index (TPI), topographic wetness index (TWI), valley depth (VD), vertical distance to channel network (VDCN) and vector ruggedness measure (VRM); a soil class map converted to Soil Taxonomy classification (12th ed, 2014); NDVI and SAVI obtained from a Sentinel-2 image (2016-09-10); maps with the annual temperature (mean (Tme), maximal (Tmax) and minimal (Tmin) in °C), annual precipitation (mm) (Prec) and solar radiation (kJ $m^{-2}$ $day^{-1}$) (SR) from the WorldClim Data Portal (FICK; HIJMANS, 2017). The covariate maps were interpolated onto a common grid of 15 m resolution.

A classical test for association between paired samples using Pearson's product–moment correlation coefficient was applied and p-value less than 0.05 was considered to represent a significant correlation. Then a Mantel test was used to determine the spatial structure for soil pH, being considered significant a p-value less than 0.20. Linear regression (*LM*) and random forest (*RF*) were performed with only correlated landscape covariates. We evaluated the accuracy of each predictive model by calculating the Root Mean Squared Error (*RMSE*) and Mean Absolute Error (*MAE*) using 10–fold cross-validation. The most accurate models were selected to fit residual variograms and apply the regression kriging (*RK*). The final RK predictions were produced using a sum of the regression and simple kriging parts.

**Results and discussion**

The results of Pearson's correlation analysis showed that, at all depths, there was a significant correlation for the covariates: EL, LS, SL, SH, STH, VD, Prec02-12, SR01-03, SR07, SR10, Tmax01-12, Tmed01-12 e Tmin01-12. The AH and MBI covariates showed a significant correlation only for the most superficial layer. MRVBF was significantly correlated at depths 0-5 and 5-15 cm, while TPI showed a significant correlation only at depths 0-5 and 15-30 cm. The SR05 and SR06

covariates showed a significant correlation in depths 5-15 and 15-30 cm, respectively (Table 1). The Mantel test showed significant values at all depths evaluated, with p-values remaining lower than specified (Table 1).

Table 1. Number of environmental covariables correlated, Mantel p-values and accuracy metrics of predictive models.

| Variable | Total of correlated covariates | Mantel p-value | LM RMSE | RF RMSE | LM MAE | RF MAE | [1]RF ΔRMSE% |
|---|---|---|---|---|---|---|---|
| pH.1 | 62 | 0.14 | 0.62 | 0.45 | 0.49 | 0.37 | +37.8 |
| pH.2 | 60 | 0.11 | 0.65 | 0.47 | 0.53 | 0.39 | +38.3 |
| pH.3 | 60 | 0.14 | 0.68 | 0.49 | 0.56 | 0.40 | +38.8 |

[1] ΔRMSE% indicates improvement in *RMSE* in percentages compared to the *LM* model.

At all 3 depths that the soil pH attribute was evaluated, the best values of RMSE and MAE were achieved with the RF models. The RMSE ranged between 0.45 – 0.49 with RF models and between 0.62 – 0.68 with GLM models, while the MAE ranged between 0.37 – 0.40 with RF models and between 0.49 – 0.56 with GLM models (Table 1). The column ΔRMSE% in Table 1 shows the relative improvement in accuracy of the RF models compared to the LM models. The relative improvement between 37.8–38.8% of random forests models show that its use improves the evaluated soil pH mapping accuracy.

Random forests can fit complex non-linear relationships and has no requirements considering the probability distribution of the target variable (HENGL et al., 2015), unlike linear regression. These two factors may be the answer to the better performance of RF models to predicting soil pH in this study.
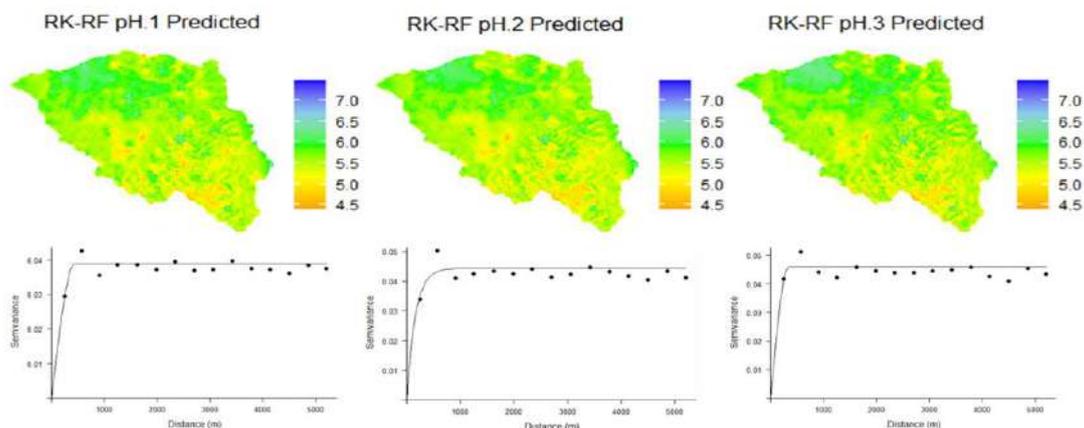


Fig. 1. Spatial prediction of soil pH at depths of 0-5, 5-15 and 15-30 cm based on the RK and respective variograms of residuals.

## Conclusions

The correlation analysis indicates a significant correlation between soil pH and 62 covariates at depth 1 and 60 covariates at depths 2 and 3. The Mantel test was significant at all depths evaluated. The best performance was achieved with the random forest models to predicting soil pH in this study.

# References

ALVARES, C. A. et al. Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, v. 22, n. 6, p. 711–728, 2014.

ARROUAYS, D. et al. GlobalSoilMap: Toward a Fine-Resolution Global Grid of Soil Properties. **Advances in Agronomy**, v. 125, p. 93–134, 1 jan. 2014.

COIPEL, J. et al. «Terroir» effect, as a result of environmental stress, depends more on soil depth than on soil type (Vitis vinifera L. cv. Grenache noir, Côtes du Rhône, France, 2000). **Journal International des Sciences de la Vigne et du Vin**, v. 40, n. 4, p. 177–185, 2006.

EMBRAPA - EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. **Manual de métodos de análise de solo** (M. E. C. Claessen et al., Eds.). Rio de Janeiro - RJ: EMBRAPA, Centro Nacional de Pesquisa de Solos, 1997. Disponível em: <https://www.agencia.cnptia.embrapa.br/Repositorio/Manual+de+Metodos_000fzvhotqk02wx5ok0q43a0ram31 wtr.pdf>.

EMBRAPA - EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. **Sistema brasileiro de classificação de solos**. 2. ed. Rio de Janeiro, RJ: Centro Nacional de Pesquisa de Solos, 2006.

FICK, S. E.; HIJMANS, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. **International Journal of Climatology**, v. 37, n. 12, p. 4302–4315, 1 out. 2017.

FLORES, C. A. et al. **Os Solos do Vale dos Vinhedos**. 1ª ed. Brasília - DF: Embrapa Clima Temperado, 2012.

GOOVAERTS, P. Geostatistics in soil science: state-of-the-art and perspectives. **Geoderma**, v. 89, n. 1–2, p. 1–45, 1 abr. 1999.

HENGL, T. et al. Mapping Soil Properties of Africa at 250 m Resolution: Random Forests Significantly Improve Current Predictions. **PLoS ONE**, v. 10, n. 6, p. e0125814, 2015.

MALONE, B. P. et al. Mapping continuous depth functions of soil carbon storage and available water capacity. **Geoderma**, v. 154, p. 138–152, 2009.

MCBRATNEY, A. B. et al. An overview of pedometric techniques for use in soil survey. **Geoderma**, v. 97, n. 3–4, p. 293–327, 1 set. 2000.

MELLO, L. M. R. DE; MACHADO, C. A. E. **Vitivinicultura brasileira: panorama 2019**. Bento Gonçalves, RS. Embrapa Uva e Vinho, 2020. Disponível em: <https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/1124189/1/COMUNICADO-TECNICO-214-Publica-602-versao-2020-08-14.pdf>. Acesso em: 19 dez. 2020

ODEH, I. O. A.; CRAWFORD, M.; McBRATNEY, A. B. Digital Mapping of Soil Attributes for Regional and Catchment Modelling, using Ancillary Covariates, Statistical and Geostatistical Techniques. In: LAGACHERIE, P.; MCBRATNEY, A. B.; VOLTZ, M. (Eds.). **Developments in Soil Science**. Elsevier, 2007. v. 31p. 437–622.

ODEHA, I. O. A.; MCBRATNEY, A. B.; CHITTLEBOROUGH, D. J. Spatial prediction of soil properties from landform attributes derived from a digital elevation model. **Geoderma**, v. 63, n. 3–4, p. 197–214, 1 nov. 1994.

TONIETTO, J. et al. O Clima Vitícola das Regiões Produtoras de Uvas Para Vinhos Finos do Brasil. In: TONIETTO, J.; RUIZ, V. S.; GÓMEZ-MIGUEL, V. D. (Eds.). **Clima, Zonificación y Tipicidad del Vino en Regiones Vitivinícolas Iberoamericanas**. 1. ed. Madrid: CYTED - Ciencia y Tecnología para el Desarrollo, 2012. p. 111–146.

TONIETTO, J. et al. **O regulamento de uso da Denominação de Origem Vale dos Vinhedos: Vinhos Finos Tranquilos e Espumantes**. Bento Gonçalves, RS: Embrapa Uva e Vinho, 2013. Disponível em: <http://www.cnpuv.embrapa.br>. Acesso em: 12 out. 2020.

YIGINI, Y.; PANAGOS, P. Assessment of soil organic carbon stocks under future climate and land cover changes in Europe. **Science of The Total Environment**, v. 557–558, p. 838–850, 2016.

ZHU, Q.; LIN, H. S. Comparing ordinary kriging and regression kriging for soil properties in contrasting landscapes. **Pedosphere**, v. 20, n. 5, p. 594–606, 2010.

# Advances in soil observation and knowledge via proximal sensing

BAHIA, Angélica Santos Rabelo de Souza[1]; MARQUES JR., José[2]; REIS, Ricardo Andrade[3]

[1]State Univ. of São Paulo (UNESP), angelica.bahia@unesp.br; [2]State Univ. of São Paulo (UNESP), jose.marques-junior@unesp.br; [3]State Univ. of São Paulo (UNESP), ricardo.reis@unesp.br

**Thematic Session: Advances in Soil Sensing**

**Abstract**

With the advancement of Agriculture 4.0, the use of proximal sensors is one of the most innovative solutions in the search for optimizing time and inputs, maximize profitability and reducing environmental impact. In this context, the aims of this work were to evaluate the prediction capacity of soil C and N stocks by diffuse reflectance spectroscopy (DRS) and magnetic susceptibility (MS), and evaluate the spatial variability of this attributes. Soil samples were collected at 372 sites. Were used partial least squares regression (PLSR) for DRS data and linear regression for MS. Good prediction accuracy parameters were obtained between the C stock with DRS (r = 0.87; RMSE = 1.24) and MS (r = 0.88; RMSE = 1.98), as well as the N stock with DRS (r = 0.82; RMSE = 2.1) and MS (r = 0.78; RMSE = 2.4), revealing that these are good predictors of important soils properties. Therefore, these close sensing techniques have the potential to meet the demand of modern agriculture.

Keywords: pedometrics; chemometrics; soil informaton systems; spectral signature; magnetic signature.

## Introduction

Soil total C and N perform essential roles in soil health and ecosystem dynamics. Soil total C is the driving force of biological activity, serving as a primary source of energy and nutrients for many soil organisms and an important factor affecting N mineralization and immobilization in soils. While total N, an essential macronutrient for plant growth, is also one of the main determinants and indicators of soil fertility and quality. Thus, the prediction of these stocks is necessary for a range of agricultural and environmental applications.

For decades, classic laboratory methodologies have been used to obtain soil attributes. However, these methods are time-consuming, expensive, destroy the soil sample during analysis and generate chemical residues. Thus, the development of fast, accurate and low-cost methods to quantify soil attributes is of paramount importance. In this sense, proximal sensors such as diffuse reflectance spectroscopy (DRS) and magnetic susceptibility (MS) meet this need, as they are two tools used to effectively estimate soil attributes (MCBRATNEY et al., 2006; BAHIA et al. al., 2017). Therefore, the objectives of this work were to investigate whether DRS and MS can be applied to predict soil C and N stocks in a sandstone-basalt transition region and characterize the spatial distribution of these attributes.

## Methodology

The 900 ha study area is located in the São Paulo State in Guatapará Town, Brazil (21º28'45''S and 48º01'01''W). This area was chosen due to the presence of great variation in soil classes, landscape forms and a geological transition between sandstone-basalt in the geomorphological unit of the Western Paulista Plateau, close to the limit of the Basaltic Cuestas. This type of geological transition represents about 44,000 ha in the state of São Paulo (IPT, 1981). So the results can be extrapolated to other regions. In the study area, a regular sampling grid containing 372 points was installed, with spacing between points varying from 142 to 174 m. Soil samples were collected at a depth of 0.0-0.2 m.

The total contents of C and N in the soil were determined by dry combustion, in the LECO CN-2000 equipment. The stocks of C and N (in t ha$^{-1}$) were calculated based on the equivalent soil mass. Reflectance values were recorded on a 950 UV/Vis/NIR spectrophotometer, equipped with an integrating sphere 150 mm in diameter, at 0.5 nm intervals along the visible (Vis) and near infrared (NIR) range (380 to 2300 nm). MS was analyzed in soil samples by the MS2 Bartington meter.

Data were submitted to descriptive statistics using the SAS software. In order to verify the statistical differences between the mean values of the attributes studied considering the compartments, the Tukey test at 5% probability was applied to the data. To develop pedotransfer functions (PTFs) based on soil spectra and laboratory data, partial least squares regression (PLSR) using Parles was used (VISCARRA ROSSEL, 2008). The PTFs based on MS were calibrated by linear regression between magnetic and laboratory data. The evaluation of the precision of the PTFs was made through the analysis of the coefficients of determination ($R^2$) and correlation (r), RMSE (root mean square error), RPD (residual prediction deviation) and the Willmott agreement index (d). The characterization of the spatial variability was performed through geostatistical analysis with modeling of experimental variograms and subsequent interpolation by ordinary kriging.

**Results and discussion**

Soil C and N stocks ranged, respectively, from 19.5 to 52.6 t ha$^{-1}$ (mean 36.6 t ha$^{-1}$) and from 2.9 to 4.4 t ha$^{-1}$ (mean of 3.7 t ha$^{-1}$). Positive correlations were found for C stock with ERD ($R^2$ = 0.75; r = 0.87; RMSE = 1.24; RPD = 2.11; d = 0.96) and MS ($R^2$ = 0.78; r = 0.88; RMSE = 1.98; RPD = 2.621; d = 0.92), as well as for N stock with ERD ($R^2$ = 0.68; r = 0.82; RMSE = 2.10; RPD = 2.76; d = 0.73) and MS ($R^2$ = 0.61; r = 0.78; RMSE = 2.40; RPD = 2.89; d = 0.71). The PTFs presented good precision and accuracy parameters (higher values of $R^2$ and RPD > 2, and lower values of RMSE), mainly for the prediction of C. The good performance of FPTs by DRS is due to the fact of the NIR region of the spectrum present information related to these elements, due to various chemical bonds (CC, C=C, CH, CN, NH). On the other hand, MS is directly related to C and N because these attributes are related to the dynamics of microbial activity in the soil.

All attributes analyzed presented spatial dependence, expressed through adjustments of variograms. The spherical model was adjusted for all attributes (Table 1). This model adjusts to attributes that present abrupt variations across the landscape (CAMBARDELLA et al., 1994). These variations may be related to the types of source material (geology), relief and soils found in the study area, showing the relationship between these factors and the detailed characterization of spatial variability and the definition of mapping units. The C stock showed a strong degree of spatial dependence (SDD ≤ 25%), while the N stock showed moderate DGE (25% < SDD ≤ 75%) (CAMBARDELLA et al., 1994). All had high ranges, and the values were similar comparing the observed and predicted data. Thus, it can be inferred that both DRS and MS are indicated to predict these attributes in a more similar way to real field data.

Table 1. Spherical variogram parameters of best fit to soil C and N stocks in the prediction dataset (observed laboratory data and predictions made by DRS and MS).

| Data set | | $C_0$ | $C_0+C_1$ | SDD (%) | A (m) | $R^2$ | SSR | Cross validation | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | a | b |
| | Observed | 3.02 | 16.50 | 21 | 1562 | 0.90 | 2.65 | -0.01 | 1.00 |
| C stock | DRS | 3.05 | 16.11 | 22 | 1548 | 0.90 | 2.15 | 0.01 | 1.00 |
| | MS | 2.70 | 15.28 | 13 | 1721 | 0.90 | 2.95 | 0.01 | 1.03 |
| | Observed | 0.01 | 0.047 | 30 | 1640 | 0.99 | 1.7E-05 | 0.00 | 1.00 |
| N stock | DRS | 0.01 | 0.045 | 29 | 1718 | 0.99 | 2.0E-05 | 0.00 | 1.00 |
| | MS | 0.02 | 0.052 | 33 | 1855 | 0.99 | 3.2E-06 | 0.00 | 1.04 |

N=372. $C_0$- nugget effect; $C_0+C_1$- step; SDD- degree of spatial dependence $[C_0/(C_0+C_1)]*100$; A- range; $R^2$- coefficient of determination of the adjusted model; SSR- sum of squares of the residuals; a- linear regression coefficient; b- angular coefficient of regression.
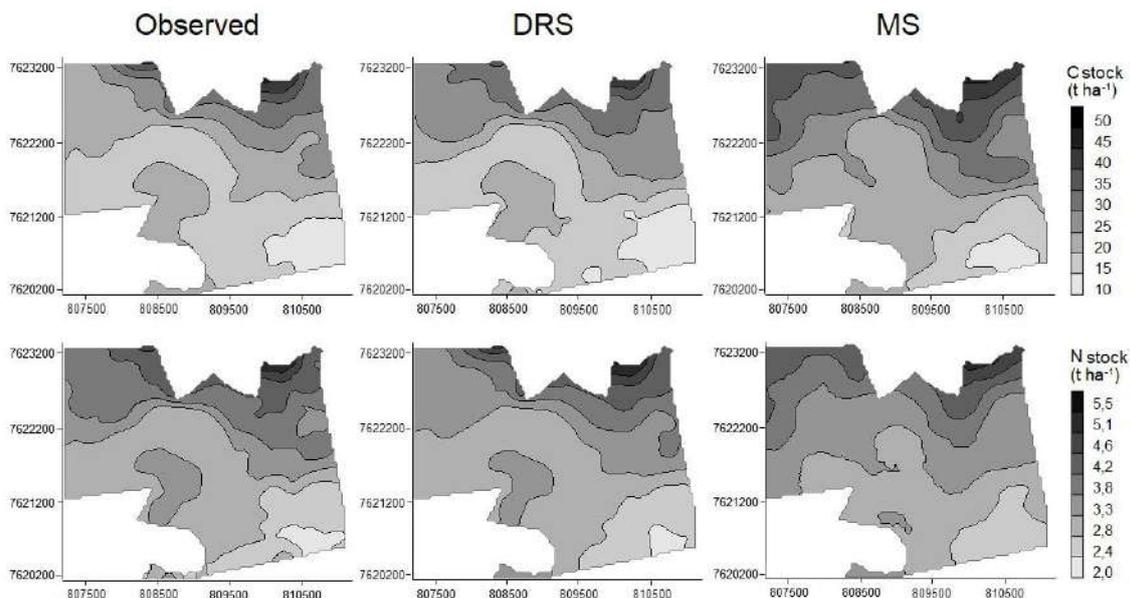


Figure 1. Spatial distribution maps of soil C and N stock in the prediction dataset (observed laboratory data and predicted by DRS and MS).

The spatial pattern of the mapped attributes was similar, as shown in Figure 1. It is observed that the map predicted by the DRS, mainly, was the one that most resembled the map obtained with the observed data, which is confirmed by the accuracy parameters in the models obtained and cross-validation (Table 1). It is known that the determination of soil attributes for mapping purposes requires a large amount of sample and this becomes onerous, making it difficult to carry out detailed mapping of large areas (BAHIA et al., 2017). For this reason, the calibration of mathematical models using DRS and MS is promising in order to predict these attributes in unmeasured samples, showing advances in soil sensing.

## Conclusions

The DRS and MS correlate positively with the values of C and N stocks in the soil, and can be used in prediction models for these attributes. The interpolated maps based on the prediction of attributes by DRS and MS show a pattern of variability similar to maps based on observed data, which is confirmed by the accuracy and cross-validation parameters. Therefore, both DRS and MS have the potential to predict C and N stocks in unknown samples, which can be used for fast and efficient evaluation of these attributes. This demonstrates the importance of these tools in mapping large areas and with a detailed scale, showing advances in soil sensing.

## Acknowledgements (optional)

## References

BAHIA, A. S. R. S.; MARQUES JR., J.; LA SCALA, N.; CERRI, C. E. P.; CAMARGO, L. A. Prediction and mapping of soil attributes using diffuse reflectance spectroscopy and magnetic susceptibility. **Soil Science Society American Journal**, v. 81, p. 1450-1462, 2017.

CAMBARDELLA, C. A.; MOORMAN, T. B.; NOVAK, J. M.; PARKIN, T. B.; KARLEN, D. L.; TURCO, R. F.; KONOPKA, A. E. Field-scale variability of soil properties in central Iowa soils. **Soil Science Society of America Journal**, Madison, v. 58, p. 1501-1511. 1994.

INSTITUTO DE PESQUISAS TECNOLOGICAS DO ESTADO DE SÃO PAULO (IPT). **Mapa geomorfológico do Estado de São Paulo**, 94 p., 1981.

MCBRATNEY, A. B.; MINASNY, B.; VISCARRA ROSSEL, R. A. Spectral soil analysis and inference systems: A powerful combination for solving the soil data crisis. **Geoderma**, Amsterdam, v. 136, P. 272–278, 2006.

VISCARRA ROSSEL, R. A. ParLeS: software for chemometric analysis of spectroscopic data. **Chemometrics and Intelligent Laboratory Systems,** v. 90, p. 72-83, 2008.

# Ground penetrating radar non-invasively positions an underground dam and estimates its water reservoir shape and volume

VASQUES, Gustavo M.[1]; RODRIGUES, Hugo M.[1]; HUBER, Emanuel[1]; TAVARES, Sílvio R. L.[1]; MARQUES, Flávio A.[1]; SILVA, Maria Sônia L.[1]

[1] Embrapa Solos, gustavo.vasques@embrapa.br, rodrigues.machado.hugo@gmail.com, silvio.tavares@embrapa.br, flavio.marques@embrapa.br, sonia.lopes@embrapa.br; [2] GEOTEST Ltd, emanuel.huber@protonmail.com

**Thematic Session: Advances in soil sensing**

**Abstract**
The objectives are to use ground penetrating radar (GPR) to: (a) find a suitable length and location for a future underground dam; and (b) model the shape and estimate the volume of its water reservoir. Radargrams were obtained in a 0.85-ha area in northeastern Brazil along five survey lines parallel to, and eight lines transverse to the slope, starting from the proposed location of the dam. The 13 radargrams were pre-processed and migrated. The terrain surface and the top of the regolith delineated in the radargrams were interpolated, plotted in 3D showing the shape of the water reservoir, and used to estimate its volume. From the radargrams, the future dam should be 45 m longer and centered 22 m south of the proposed location, increasing the water reservoir by 50% to a total of 6 million L. By assessing the terrain surface and regolith, the GPR allows to adjust the length and position future underground dams and assess their water reservoir shape and volume non-invasively.
Keywords: GPR; Geophysics; Brazilian semiarid region; Soil depth; Regolith

## Introduction

In the Brazilian semiarid region, underground dams have been used to bar and store rainwater in the subsurface providing water for reuse (SILVA et al., 1998). Despite its importance, the volume of water stored by the underground dam is typically not quantified due to the high cost of the identification and mapping of the impermeable layer that limits the vertical flow of water at the bottom of the reservoir, which requires drilling or opening soil trenches.

The ground penetrating radar (GPR) has been used to map soil restrictive layers (NOVÁKOVÁ et al., 2013; SCHALLER et al., 2020) and to find the ideal location for building an underground dam (LIMA et al., 2018). However, the estimation of the water reservoir volume in the underground dam accumulation area from GPR data remains an open research task.

Thus, the objectives are to use a GPR to: (a) find a suitable length and location for a future underground dam; and (b) model the shape and estimate the volume of its water reservoir.

## Methodology

The area is located in Santana do Ipanema, Alagoas state, in the semiarid region of northeastern Brazil, at coordinates 9°23'47.5" S and 37°13'39.4" W. The area has a 3% slope gradient and lies around an intermittent stream, where an underground dam will be built. Soils include *Neossolos Flúvicos* (Fluvisols; Fluvents), *Neossolos Regolíticos* (Regosols; Psamments), and *Planossolos Háplicos* (Planosols; Aqualfs).

A survey was done using a MALÅ GroundExplorer GPR (Guideline Geo AB, Sundbyberg, Sweden), carrying a 450-MHz shielded antenna, consisting of five lines parallel to, and eight lines transverse to the slope (Figure 1a), with line 1 representing the proposed length and location of the future dam (Figure 1a, hatched feature).

Soil trenches were opened at nine sites close to the GPR survey lines (Figure 1a, b) to classify the soil and mark the top of the regolith, which constitutes the bottom of the water reservoir. The GPR radargrams were pre-processed (zero-time correction, dewow, gain, eigenvalue filter, constant offset correction, time-to-depth conversion) and migrated (Kirchhoff), and the top of the regolith was delineated in all radargrams (HUBER; HANS, 2019; R CORE TEAM, 2020). The terrain surface and the top of the regolith elevations were interpolated with 2-m resolution across the area by multilevel B-splines.

The soil pore volume in each pixel was calculated by multiplying the soil depth (terrain surface minus top of regolith) by the pixel area (4 m$^2$) by a soil porosity of 0.38 m$^3$ m$^{-3}$ estimated from similar soils of the region (JACOMINE et al., 1975). The water reservoir volume was estimated by summing up the pore volume of all pixels.

For this exercise, it is assumed that: the water reservoir is limited at the top by the terrain surface, at the bottom by the top of the regolith, and laterally by the boundaries of the GPR survey lines; a single soil porosity value represents the whole reservoir both horizontally and vertically; and all pores are available to store water.

**Results and discussion**
Along the slope, the soil depth increases from the upper (northwest) to the lower part of the area close to the dam (southeast) (Figure 1b, red lines). Across the slope, it increases from both sides towards the intermittent stream thalweg, which is closest to line 10 and trenches P1, P4 and P7 (Figure 1a, b). The soil depth varies between 45 (P8) and 160 cm (P5 and P7) as observed in the soil trenches.

The radargram of survey line 1 corresponds to the proposed length and location of the future underground dam (Figure 1c). The shape of the terrain surface and top of the regolith at line 1 shows that the proposed dam is not centered at the intermittent stream thalweg and its length of 65 m is too short to bar the underground water flow in the southern side of the area. Thus, the underground dam should be extended about 45 m, and centered about 22 m to the left (southwards), assuming that the cross-sectional shape of the slope is symmetrical.

The interpolated terrain surface, and top of the regolith elevations ranged between 213.5 and 215 m, and 210 to 213.5 m, from the dam to the top, respectively (Figure 2). In this exercise, their map extent represents the underground dam accumulation area of about 0.85 ha, and the region between them corresponds to the shape of the water reservoir. The water reservoir is thickest close to the dam along the slope, and to the thalweg across the slope (Figure 2).

The estimated water reservoir volume in this 0.85-ha area is 4013 m$^3$. Extending the underground dam 45 m southwards to a total of 110 m in length and centering it at the stream thalweg would increase the accumulation area to about 1.44 ha, assuming a proportional increase keeping its rectangular shape. The water reservoir

volume would increase to about 6020 m$^3$, assuming that the area is symmetrical across the stream. Also, this would stop the underground water from flowing outwards laterally at the southern border. Additionally, the top of the dam should be built above the terrain surface to store water aboveground for irrigation, animal watering, and other uses, and capture surface and runoff water during heavy rains.

## Conclusions

The results obtained from this exercise show the potential of the GPR to non-invasively identify, visualize and map the terrain surface and the top of the regolith, supporting adjusting the length and location of a future underground dam and estimating its water reservoir shape and volume, with minimal need for soil trenches. The individual radargrams and derived terrain surface and top of the regolith maps can be plotted in 3D to show their variations and the shape of the water reservoir vertically and horizontally, along and across the slope.

The analytical methods, functions, and tools provided by the free and open-source RGPR package in R covered the whole framework of GPR data processing. Thus, using the RGPR package is recommended for its completeness and to reduce the cost of data analysis by replacing proprietary software.

## References

HUBER, E.; HANS, G. RGPR: **Ground-penetrating radar (GPR) data visualisation, processing and delineation**. R package version 0.0.7. 2019.

JACOMINE, P. K. T. et al. **Levantamento exploratório - Reconhecimento de solos do estado de Alagoas**. Recife: Embrapa, 1975. (Boletim Técnico, v. 35)

LIMA, A. O. et al. GPR 3D profile of the adequateness of underground dams in a sub-watershed of the Brazilian Semiarid. **Revista Caatinga**, v. 31, p. 523-531, 2018.

NOVÁKOVÁ, E. et al. Evaluation of ground penetrating radar and vertical electrical sounding methods to determine soil horizons and bedrock at the locality Dehtáře. **Soil and Water Research**, v. 8, p. 105-112, 2013.

R CORE TEAM. **R: A language and environment for statistical computing**. Vienna, R Foundation for Statistical Computing, 2020.

SCHALLER, M. et al. Comparison of regolith physical and chemical characteristics with geophysical data along a climate and ecological gradient, Chilean Coastal Cordillera (26 to 38° S). **SOIL**, v. 6, p. 629-647, 2020.

SILVA, M. S. L. et al.Exploração agrícola em barragens subterrâneas. **Pesquisa Agropecuária Brasileira,** v. 33, p. 975-980, 1998.
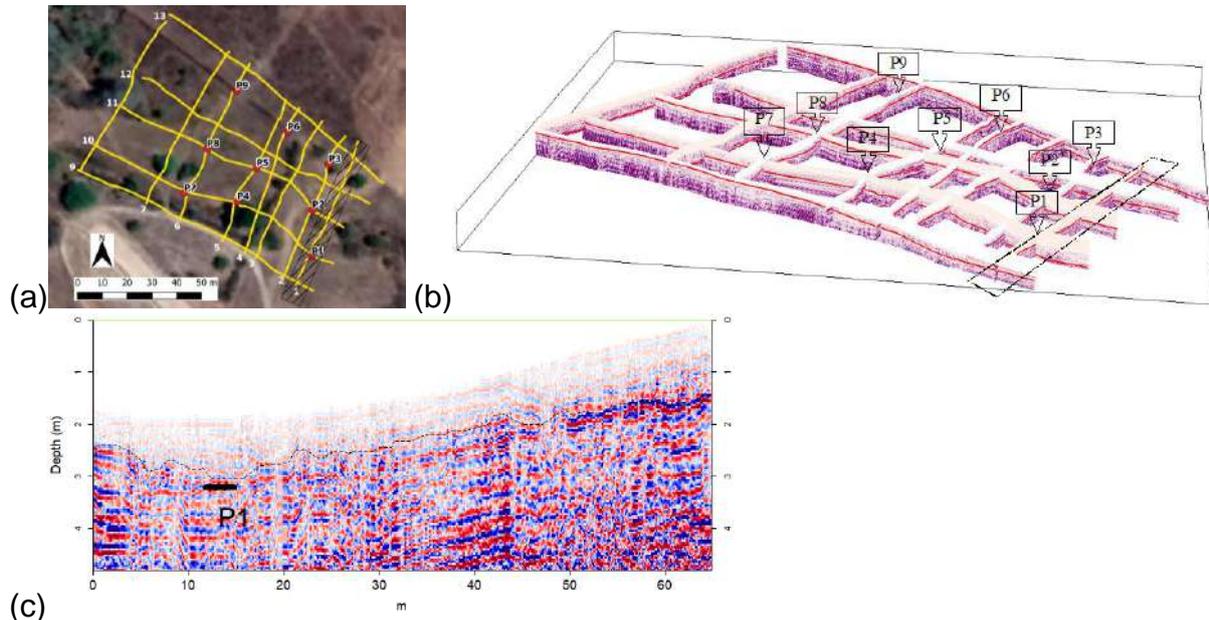
Figure 1. (a) GPR survey lines (1 to 13) and soil trenches (P1 to P9); (b) Migrated radargrams showing the delineated top of the regolith as red lines; and (c) Radargram of line 1 corresponding to the proposed length and location of the future underground dam, showing the delineated top of the regolith (thin dashed line), and the position of the top of the regolith recorded in P1 (thick dash).
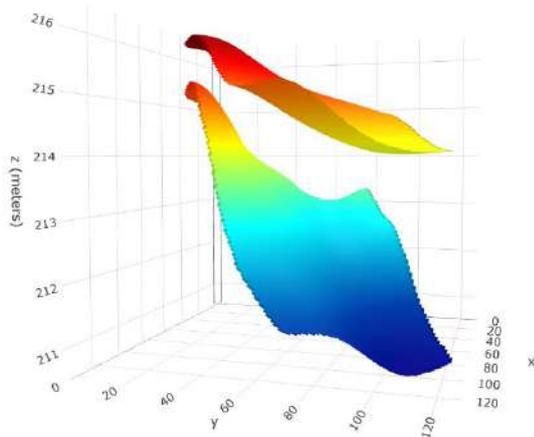


Figure 2. Terrain surface and top of the regolith elevations interpolated from the 13 GPR survey lines. The region between them corresponds to the shape of the water reservoir. Elevations (z) are exaggerated relative to x and y coordinates.

# Spatial variability attributes of a Ultisols hydrosequence with aid of portable X-ray fluorescence spectrometry (pXRF)

Bócoli, Fernanda[1]; Silva, Sérgio[2]; Santos, Walbert[3]; Inda, Alberto[4]; Curi, Nilton[5]

[1] UFLA, bocolifernanda@gmail.com; [2] UFLA, sergio.silva@dcs.ufla.br; [3]IFSULDEMINAS, walbert.santos@ifsuldeminas.edu.br; [4] UFRGS, alberto.inda@ufrgs.br; UFLA, niltcuri@ufla.br

## Thematic Session: Advances in soil sensing

### Abstract

Proximal sensors are drawing attention in many pedological studies. Pedological investigations using proximal sensors are increasing in tropical regions, but more in-depth studies on spatial variability of properties at short distances are still scarce. This study evaluated the spatial variability of three Ultisols profiles in a hydrosequence by digital morphometrics, combining sensors data. Soil samples were collected for elemental distribution mapping on a regular grid design (0.2 x 0.2 m) and analyzed via portable X-ray fluorescence (pXRF) spectrometry. The spatial evaluation showed great variations in elementary content distribution. Being possible to identify changes on parent material along the hydrosequence and into the soil profiles.

Keywords: Ultisols with different drainage; proximal sensors; digital morphometrics.

## Introduction

Pedological investigations using proximal sensors for in-depth evaluations of soil properties variability at short distances are still scarce in tropical regions, although proximal sensors have become very helpful for soil-related studies (Stockmann et al., 2016). The spatial variability of attributes in a profile, as the elemental contents, has been explored in few studies (Gauer-Gray, Hartemink, 2018; Silva et al., 2018; Sun et al., 2020, Mancini et al., 2021). However, understanding the spatial dynamic of soil profile attributes can aid in numerous studies related to pedology or agricultural management. In this regard, Hartemink and Minasny (2014) proposed the soil digital morphometrics approach, which is a technique that measures and quantifies the soil profile attributes and its spatial variability with the aid of proximal sensors.

This study evaluated the spatial variability of three Ultisols profiles in a hydrosequence via portable X-ray fluorescence (pXRF) spectrometry. This study hypothesizes that the profile position on the landscape rules the attributes variations. This study aims to investigate the causes of the soil profiles differences along the hydrosequence, characterize and evaluate the attributes that varied more in the Ultisols, with proximal sensors associated with the morphological description. Therefore, the portable X-ray fluorescence spectrometer (pXRF) and magnetic susceptibilimeter (MS) data were analyzed individually and combined to understand the soil profile genesis, elemental, and magnetic minerals interactions.

## Methodology

We analyzed three Ultisols profiles located at Lavras county, Minas Gerais state, Brazil (Figure 1). All the profiles were located under the native vegetation (Atlantic Forest), and the parent material of the Ultisols is the gneiss.
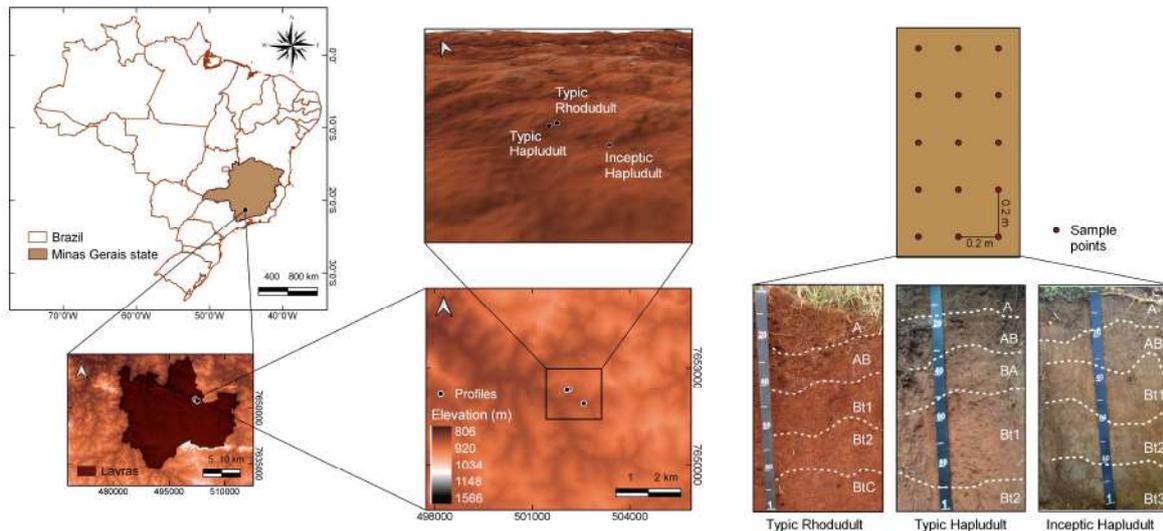


Figure 1. Location, topography, altitude, and sample points of the three Ultisol profiles: Typic Rhodudult, Typic Hapludult, and Inceptic Hapludult, in Lavras, Minas Gerais state, Brazil.

Samples were collected in a regular grid design (0,2 x 0.2 m) throughout the profiles and were air-dried and sieved to 2 mm. Approximately 10 g were analyzed via a Bruker® spectrometer model S1 Titan LE, on trace mode, in triplicate, for 60 seconds (Weindorf; Chakraborty, 2016). The spatial analyses of the elemental contents obtained by pXRF were conducted via interpolation through the Multilevel B-splines (Lee et al., 1997) using the software QGIS 3.16 (QGIS Development Team, 2021).

## Results and discussion

All profiles showed high MS values (Table 1) on the surface and lower in-depth. The Typic Hapludult and Inceptic Hapludult profiles probably received sediments deposition from soils derived from mafic rocks with magnetic minerals on their constitution, as maghemite and magnetite (Dearing, 1999) present in upper areas of the study area (Curi et al., 2017). The pXRF analyses showed that total Fe contents (Figure 2) tended to decrease in-depth. The Typic Rhodudult had the highest MS value on the horizons A and AB. Since this profile is located at the upper third of the hillslope, its better drainage probably favored more magnetic minerals formation than the other profiles increasing its magnetism. Moreover, a gneiss lithologic discontinuity from the Typic Rhodudult to the other profiles is possible, modifying its elemental composition.

Table 1. Results from magnetic susceptibility of the analyzed profiles.

| Soil | Typic Rhodudult | | | | | Typic Hapludult | | | | | Inceptic Hapludult | | | | |
|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| Horizon | A | AB | Bt1 | Bt2 | BTC | A | AB | BA | Bt1 | Bt2 | A | AB | Bt1 | Bt2 | Bt3 |
| MS* | 7.7 | 9.9 | 7.5 | 5.9 | 4.6 | 5.3 | 4.6 | 3.9 | 3.3 | 3.9 | 8.5 | 5.5 | 3.8 | 2.9 | 2.3 |

*MS= magnetic susceptibility ($10^{-7}$ m$^3$ kg$^{-1}$).

In Figure 2, the total elemental contents are higher in the Inceptic Hapludult than the Typic Hapludult. Only the K content is higher on the latter profile. In the Typic Rhodudult, Al, Si, and K, total contents are higher than the contents of the other profiles. More weathered soils generally have high contents of non-mobile elements as Al, Ti, and Fe because the mobile elements have already been leached out from these soils (Stockmann et al., 2016). Still, the elemental contents depend on the dominant soil-forming process and slight variations in the soil parent material. Stockmann et al. (2016) and Mancini et al. (2021) used the Ti and Zr total contents as parent material tracers, indicating, for example, lithologic discontinuities in some profiles. These residual elements here show a uniform distribution in Typic Rhodudult and Typic Hapludult, which did not occur with the Inceptic Hapludult, which can indicate a parent material variation. Mancini et al. (2021) found a wide variation in the gneiss composition when studying a 4.5m Oxisol profile, verifying laminations richer in biotite, others in muscovite, in an area near this Ultisols hydrosequence study.
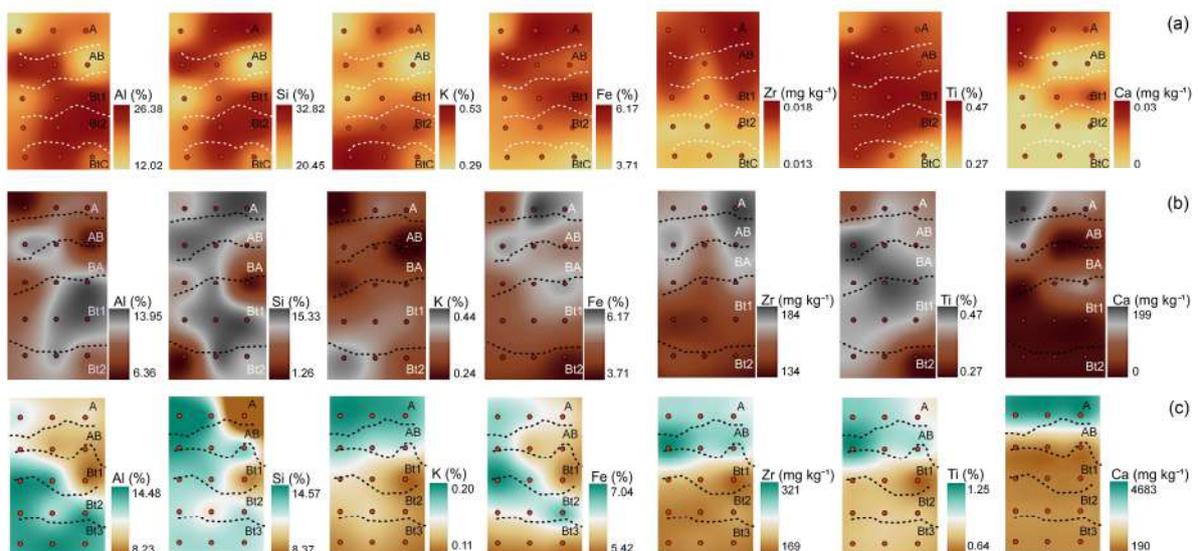


Figure 2. Maps of total elemental distribution obtained with the pXRF aid for the profiles: Typic Rhodudult (a), Typic Hapludult (b), and Inceptic Hapludult (c), in Lavras municipality, Minas Gerais state, Brazil.

## Conclusions

The profiles differed by their color and elemental distribution. Although the soils have as parent material the gneiss, the area seems to have a lithologic discontinuity from the Typic Rhodudult to the other profiles distinguishing their elemental distribution. The Inceptic Hapludult may have a deposition of sediments from upper areas of the landscape, which is shown by its elemental contents and magnetic susceptibility variation comparing to other horizons. The spatial analyses, mainly with pXRF data, demonstrated that even morphologically homogenous horizons of the same profile, there might be a wide variation in the distribution of the elements as Fe, Si, and Al.

## Acknowledgements

## References

CURI, N. et al. **Mapeamento de solos e magnetismo no campus da UFLA como traçadores ambientais.** Lavras: UFLA, 2017. 147 p.

DEARING, J. Environmental Magnetic Susceptibility. **Using the Bartington MS2 system**, 43 p., 1999.

GRAUER-GRAY, J.; HARTEMINK, A.E. Raster sampling of soil profiles. **Geoderma**, v. 318, p. 99-108, maio 2018. Elsevier BV. http://dx.doi.org/10.1016/j.geoderma.2017.12.029.

HARTEMINK, A.E.; MINASNY, B. Towards digital soil morphometrics. **Geoderma**, v. 230-231, p. 305-317, out. 2014. Elsevier BV. http://dx.doi.org/10.1016/j.geoderma.2014.03.008.

LEE, S. et al. **Scattered Data Interpolation with Multilevel B-Splines**. IEEE Transactions On Visualisation And Computer Graphics, v.3, n.3., p.228-244, 1997.

MANCINI, M. et al. Formation and variation of a 4.5 m deep Oxisol in southeastern Brazil. **Catena**, v. 206, p. 105492, nov. 2021. Elsevier BV. http://dx.doi.org/10.1016/j.catena.2021.105492.

QGIS Development Team, 2021. **QGIS Geographic Information System**. Open Source Geospatial Foundation. https://qgis.org/en/site/

SILVA, S.H.G. et al. Soil weathering analysis using a portable X-ray fluorescence (PXRF) spectrometer in an Inceptisol from the Brazilian Cerrado. **Applied Clay Science**, v. 162, p. 27-37, set. 2018. Elsevier BV. http://dx.doi.org/10.1016/j.clay.2018.05.028.

STOCKMANN, U. et al. Utilizing portable X-ray fluorescence spectrometry for in-field investigation of pedogenesis. **Catena**, v. 139, p. 220-231, abr. 2016. Elsevier BV. http://dx.doi.org/10.1016/j.catena.2016.01.007.

SUN, F. et al. Enhanced soil profile visualization using portable X-ray fluorescence (PXRF) spectrometry. **Geoderma**, v. 358, p. 1-11, jan. 2020. Elsevier BV. http://dx.doi.org/10.1016/j.geoderma.2019.113997.

WEINDORF, D.C.; CHAKRABORTY, S. Portable X-ray Fluorescence Spectrometry Analysis of Soils. **Methods of Soil Analysis**, v. 1, n. 1, p. 1–8, 2016.

# Pattern recognition of soil profiles based only their mid-IR reflected energy.

TERRA, Fabrício da Silva[1]; DEMATTÊ, José Alexandre Melo[2]; HORÁK-TERRA, Ingrid[1]

[1] Federal University of Jequitinhonha and Mucuri Valleys, fabricio.terra@ufvjm.edu.br, ingrid.horak@ufvjm.edu.br; [2] University of São Paulo, jamdemat@usp.br

**Thematic Session: ADVANCES IN SOIL SENSING**

**Abstract**

Soil reflectance spectra have a great potential to be directly used as auxiliary tool for pedogenesis studies and for soil surveys, classification, and mapping, but not simply as input for quantitative models. However, there is a lack of studies about spectral pedology using mid-IR data. So, we aimed to classify 20 soil profiles by integrating reflectance and pedological metrics and evaluate the influence of soil properties in this classification. We used 20 soil classes (profiles), mid-IR spectra (Nicolet 6700 FTIR), and taxonomic distance implemented in the OSACA program. A coherent classification of profiles was possible by combining spectra and pedological metrics, since 20 profiles were reorganized into 11 clusters and 30 % of them were individually clustered as follows: Nitisol, 2 Gleysols, Humic Ferralsol, Rhodic Ferralsol, Acrisol. Combining mid-IR spectra and pedological metric is very effective for pattern recognition of representative profiles of some Brazilian soil classes.

## Introduction

How can Brazil reach an efficient production combined with conservation of natural resources without knowing soils? Suitable knowledge of soil spatial and temporal variability inevitably implies in a large amount of fieldwork and sampling necessarily resulting in large amounts of analyses, waste generation, and high financial costs. Proximal soil sensing by reflectance spectroscopy is based on obtaining information by reflected energy after macroscopic and microscopic interactions between electromagnetic radiation and soil components (organic, mineral, and water) (Baungardner et al., 1985). So, this technique shows great utility in pedological assessments where soil spectra can be applied on preliminary evaluation of large areas with higher sampling density, quickly, cost-effectively and without impacting the environment. Soil reflectance spectra have a great potential to be directly used as auxiliary tool for helping and optimizing pedogenesis studies and soil surveys, classification, and mapping, but not only and simply as input for predictive models of soil properties. Soil spectra from visible to shortwave infrared (400 to 2500 nm) have been tested in pedological studies (Demattê e Terra, 2014; Terra et al., 2018). However, there is still a lack of researches about spectral pedology using mid-infrared spectra (mid-IR: 4000 to 400 $cm^{-1}$), which can be efficiently used to distinguish and cluster soil profiles. By only using mid-IR, our aims were to a) classify 20 soil profiles by integrating reflectance data and pedological metrics and b) evaluate the influence of weathering and pedogenesis in this resulting classification.

## Methodology

A soil database with physic-chemical and mineralogical analyses was used for it, where the sampling sites are scattered over four Brazilian States: São Paulo, Minas Gerais, Mato Grosso do Sul, and Goiás. Soil mid-IR reflectance spectra were obtained by the Nicolet 6700 Fourier-Transform Infrared equipped with Smart Diffuse Reflectance (1.2 nm spectral resolution and 64 readings per spectrum), and they were log-transformed into absorbance values, baseline corrected, and transformed by the mean-centered Principal Component Analysis (PCA). For clustering soil profiles taking into account differences in spectral behavior in depth comparing horizon by horizon, we used 20 soil classes (profiles) from the database with similarities and differences among them mainly regarding to clay content and mineralogy. The taxonomic distance (pedological metric) implemented in the OSACA (*Outil Statistique d'Aide à la Cartogénèse Automatique*) program (Carré and Jacobson, 2009) was applied as an automatic system for clustering soil profiles by comparing their properties in depth (horizon by horizon). Similarities between two soil profiles are calculated as a mean distance between horizons of each profile (Equation 1). This metric takes in account sequence of horizons in a soil profile. Having a profile more horizons than other one, the deepest horizon of the shallowest profile is used once and again in calculations.

$$D_{ped}\left(S_a,S_b\right)=\frac{\sum_{j=1}^{M_c} D_h\left(h_{a,j},h_{b,j}\right)+\sum_{j=M_c+1}^{M_b} D_h\left(h_{a,M_c},h_{b,j}\right)}{Mb}$$

(1)

Where: $D_{ped}$ is the taxonomic distance (pedological metric) between two soil profiles A ($S_a$) and B ($S_b$), M is the number of horizons of each profile, $D_h$ is the Euclidean distance (Equation 2) between horizons (h) of each profile.

$$D_h\left(h_a,h_b\right)=\left(\sum_{f=1}^{F}\left(V_{a,f}-V_{b,f}\right)^2\right)^{0.5}$$

(2)

Where: V is a vector of F variables (soil properties) describing each horizon.

In OSACA, fitting a profile into a cluster is defined by the shortest taxonomic distance between a profile and central profile of a cluster, and the optimization of this process is achieved by changing a profile from a cluster to another in order to reduce the squared error. We performed two clustering analyses, where firstly the following properties were used as variables: sand; clay; organic carbon (OC); exchangeable phosphorus (P), potassium (K), calcium (Ca), magnesium (Mg), and aluminum ($Al^{3+}$); potential acidity ($H^+ + Al^{3+}$); sum of bases (SB); cation exchange capacity (CEC), clay activity (CA); base (V%) and aluminum (m%) saturations; and pH in water, important for soil management, survey, and classification. Secondly, the PCA's scores from mid-IR data were tested. Contingency matrices were used to assess the efficiency of the spectra in performing pattern recognition of soil profiles by comparing the number of resulting clusters after each clustering.

**Results and discussion**

According the IUSS Working Group WRB (2015), the soil classes (acronym of the Brazilian classes between parentheses) selected were: Arenosol (profile 1), Acrisol (profile 2), Lixisol (profile 5), Nitisol (profile 14), Cambisol (profile 15), Gleysol

(profiles 17 and 20), Planosol (profile 18), Leptsol (profile 19), Haplic Ferralsol (profiles 3, 6, and 7), Rhodic Ferralsol (profiles 4, 9, 10, 11, 12, and 13), Xanthic Ferralsol (profile 8), Humic Ferralsol (profile 16). For each soil property, the mean and standard deviation values were: 462 and 305 g kg$^{-1}$ for sand, 438 and 270 g kg$^{-1}$ for clay, 8 and 5 g kg$^{-1}$ for OC, 5 and 13 mg kg$^{-1}$ for P, 1.3 and 2 mmol$_c$ kg$^{-1}$ for K, 9 and 9 mmol$_c$ kg$^{-1}$ for Ca, 4 and 4 mmol$_c$ kg$^{-1}$ for Mg, 9 and 11 mmol$_c$ kg$^{-1}$ for Al$^{3+}$, 35 and 20 mmol$_c$ kg$^{-1}$ for H$^+$+Al$^{3+}$, 14 and 14 mmol$_c$ kg$^{-1}$ for SB, 50 and 26 mmol$_c$ kg$^{-1}$ for CEC, 152 and 106 mmol$_c$ kg$^{-1}$ for CA, 27 and 19% for V%, 43 and 32% for m%, and 5 and 0.6 for pH in water.

Using soil properties, the twenty profiles were divided into 7 clusters (figure 1a). The profiles 4, 13, 16, and 18 were grouped in the cluster 1 based on higher values of OC, H$^+$+Al$^{3+}$, and CEC (distances from 5% to 9%). Ferralsols 7 and 12 (cluster 2 - distance of 14%) were grouped because similarity in number of profiles, higher contents of P, K, Ca, Mg, and SB. Cambisol (15) was separately clustered due to its clay particle distribution and lower values of CA. Ferralsols 3, 9, and 11 were grouped in the cluster 4 (distances from 5% to 7%) because intermediate values of all properties. Profiles 17 and 19 were grouped in the cluster 5 (distance of 15%) due to reduced number of profiles (2 profiles). Profiles 1, 2, 5, 6, 8, and 20 were grouped in the cluster 6 (distances from 7% to 21%) based on higher contents of sand and CA and lower values of OC. The cluster 7 joined the profiles10 and 14 with distance of 10% due to higher values of clay, OC, and H$^+$+Al$^{3+}$.
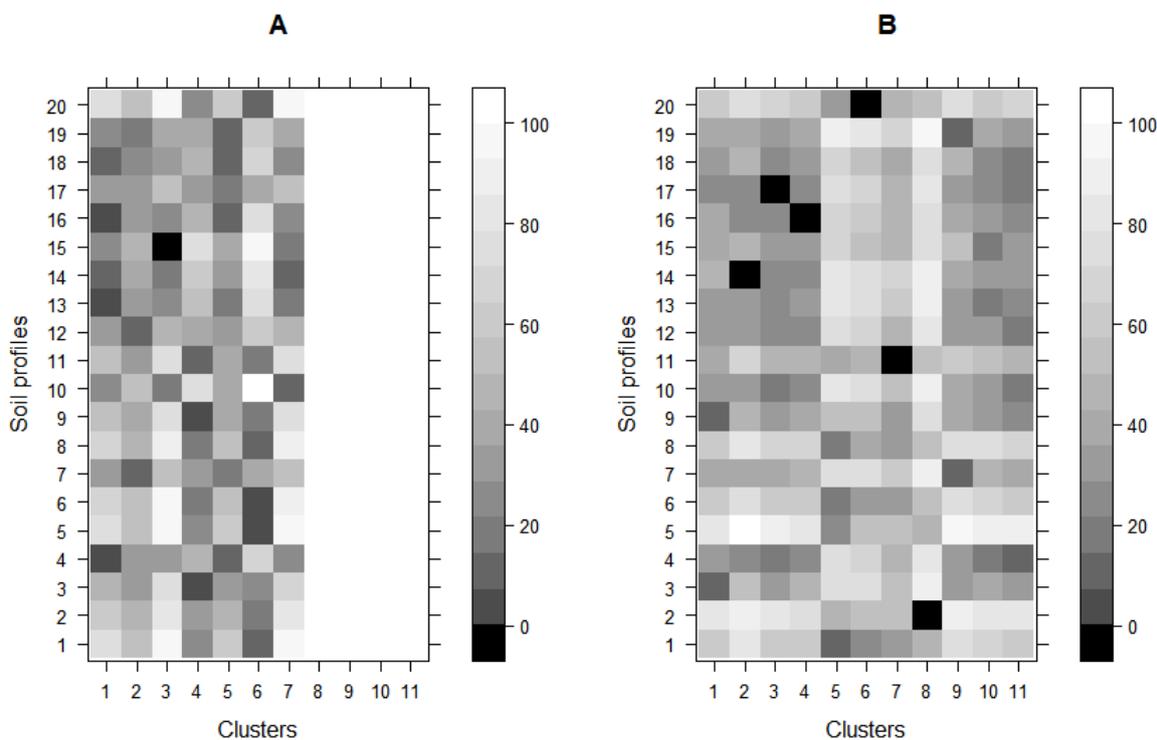


Figure 1. Matrix representation of the relative taxonomic distances (in %) between soil profiles and center of clusters using soil properties (A) and mid-IR spectra (B).

In Based on mid-IR spectra, 11 clusters were obtained basically based on variations of reflectance intensity and absorption features of these soil profiles represented by their PCA's scores. A better discrimination among soil profiles was possible due to contributions of mineralogy, weathering and pedogenetic processes brought by the profiles' spectral behaviors. The following soil profiles were individually clustered: 2, 11, 14, 16, 17, and 20. Ferralsols 3 and 9 were grouped in the cluster 1 (distance of 13%). Profiles with higher sand content by horizons (1, 5, 6, and 8) were grouped in the cluster 5 (distance up to 24%). The cluster 9 jointed profiles with the fewest number of horizons (distance of 14%). Profiles 13 and 15 were clustered together (cluster 10 - distance of 17%) due to similar clay content and lower values of SB and V% influenced by mineralogy. The cluster 11 grouped Rhodic Ferralsols (4, 10, and 12) and Planosol (18) with ferric characteristics and considerable contents of $Al^{3+}$, $H^+$ $+Al^{3+}$, and m% (distances from 10% to 21%).

## Conclusions

Combining mid-IR spectra and pedological metric is very effective for pattern recognition of representative profiles of some Brazilian soil classes, since soil spectra bring information about particle size distribution, mineralogy, and some chemical properties. Ferralsols, Nitisol, Acrisol, Lixisol, Arenosol, Gleysol, Cambisol, and Leptsol can be correctly discriminated. Based on our results, mid-IR spectral behavior of soil profiles can be reliably integrated with conventional properties and morphological characteristics for a better automated discrimination and clustering of soil classes. Mid-IR is presented here as a useful auxiliary tool for direct applications in pedological assessments, particularly, in soil survey and classification and optimizations of these activities.

## Acknowledgements (optional)

## References

BAUMGARDNER, M. F. et al. Reflectance properties of soils. **Advances in Agronomy**, Amsterdam, v. 38, p. 1-43, 1985.

DEMATTÊ, J. A. M.; TERRA, F. S. Spectral pedology: A new perspective on evaluation of soils along pedogenetic alterations. **Geoderma**, Amsterdam, v. 217-218, p. 190-200, 2014.

TERRA, F. S.; DEMATTÊ, J. A. M.; VISCARRA ROSSEL, R. A. Proximal spectral sensing in pedological assessments: vis-NIR spectra for soil classification based on weathering and pedogenesis. **Geoderma**, Amsterdam, v. 318, p. 123-136, 2018.

IUSS WORKING GROUP WRB. 2014. **World Reference Base for Soil Resources 2014**. International soil classification system for naming soils and creating legends for soil maps. World Soil Resources Reports Nº. 106. FAO, Rome.

# Mapping soil salinity: a case study from Marajó Island, Brazilian Amazonia

HENRIQUES, Renata Jordan[1]; de OLIVEIRA, Fábio Soares[2]; SCHAEFER, Carlos Ernesto Gonçalves Reynaud[3]; FRANCELINO, Márcio Rocha[4]

[1] PhD Student, Universidade Federal de Minas Gerais, renatajhques@gmail.com
[2] Professor, Universidade Federal de Minas Gerais, fabiosolos@gmail.com
[3] Professor, Universidade Federal de Viçosa, reyschaefer@yahoo.com.br
[4] Professor, Universidade Federal de Viçosa, marcio.francelino@gmail.com

**Thematic Session: ADVANCES IN SOIL SENSING**

## Abstract

Remote sensing is a key way to identifying surface salinity and its particularities. The scenario of this study is Marajó Island, located in the Amazon River mouth. This fluvial-marine ecosystem has one of the largest areas of salinity in Brazil. The main goal is to present the applicability of spectral indices in the recognition of surface salinity using Sentinel-2. The work was developed in three main stages: (i) fieldwork for soil sampling; (ii) laboratory procedures for obtaining electrical conductivity; and (iii) digital processing of Sentinel-2 for obtained the salt plains data. As result, salinity indices are satisfactory for recognize salt-affected soils and its particularities. The eastern Marajó Island showed a great geodiversity, with potential to increase Holocene landscape evolution and climate change projections by its salinity environments.

## Introduction

Salts accumulation on the earth's surface results from physic-chemical and human processes, representing considerable environmental hazard (Gorji et al., 2017). Remote sensing is a key way to identifying surface salinity and its particularities, be analysing spectral signatures and algebra between sensor bands (Wu et al., 2018). Marajó Island is the largest fluvial-marine archipelago on the planet, with one of the most extensive areas of salt-affected soils in Brazil. Saline soils at coastline are result of seawater intrusion, intensive evapotranspiration, or improper soil and water management (Paz et al., 2020; Wu et al., 2018). The excess of soluble salts (saline soils), the dominance of exchangeable sodium in the soil exchange complex (sodic soils), and both combination (saline-sodic) can affect several chemic-physical landscape processes (Gorji et al., 2017). In this context, the main goal of this work is to present the applicability of spectral indices in the recognition of surface salinity in Marajó Island, using Sentinel-2 images. The study is justified by necessity of

increase knowledge of soil salinity in Brazil, and its particularities in a fluvial-marine ecosystem.

**Methodology**

The study was developed in three main stages: (i) field work for soil sampling and in situ observations; (ii) laboratory procedures for obtained the electrical conductivity of soil samples; and (iii) digital processing of Sentinel-2A image. In the first step, five pedological horizons were obtained according the landscape diversity of non-saline areas, apicums and salt plains. The second stage consisted in obtained the electrical conductivity of saturated soil past extract ($E_c$) (USDA, 2014). In the last step, the Sentinel-2 image was obtained from 2020 January; corrected the atmospheric effects by Sen2Cor plugin (Main-Knor, 2017); applied the Normalized Difference Salinity Index (NDSI re1) (Wang et al., 2019), Intensity Index (Int1) (Fourati et al., 2015), Salinity Index 1 (SI1) (Khan et al., 2001), Salinity Index red-edge 1 (SI1 re1) (Wang et al., 2019), and Normalized Difference Vegetation Index (NDVI) (Tab 1), and obtained the spectral signature of each horizon sampled.

Tab.1: Spectral indices of salt-affected soils

| Index | Equation | Sentinel-2 equation | Reference |
|---|---|---|---|
| NDSI re1 | (red-edge 1 – NIR)/ (red-edge1 + NIR) | (B5 – B8A)/(B5+B8A) | Wang et al. (2019) |
| Int1 | (G + R)/2 | (B3+B4)/2 | Fourati et al. (2015) |
| SI1 | (G X R) 0.5 | (B3 x B4) 0.5 | Khan et al. (2001) |
| SI1 re1 | (B + red-edge1) 0.5 | (B3 x B5) 0.5 | Wang et al. (2019) |
| NDVI | (NIR – R) / (NIR + R) | (B4 – B8A)/(B4+B8A) | |

**Results and discussion**

The spectral indices showed high concentrations of salts in almost of 59.1 km² of the study area, with potential to be applicably in the entire open plains of Marajó Island. Were identified considerable diversity of salt land distribution patterns. The northern portion reveals tendency to accumulate superficial rainfall with high evapotranspiration processes, that causes high salt concentrations at the surface and the most tendence to form saline crusts during dry periods. In the context of climate change, the sea-level rising has potential to raising water base level of Marajó Bay, enhancing the saline accumulation (Fig. 1).

The southern portion reveals predominance of salt dispersed within the soils, and less subjected to waterlogging. The main indicator of salinity in the landscape are wide saline plains associated with low-lying halophyte plants, such as *Sesuvium portulacastrum* and palm trees as *Copernicia prunifera* (Miller) H.E. Moore (*Arecaceae*) (carnaúba). This portion has extensive apicums associated to Marajó's mangrove ecosystem, that made a very fragile landscape in front of climate change and sea-level rise (Fig. 1).

The spectral signatures and the saline indices attained better results where salt is concentrated near the surface. The NDVI is an indirect indication of the saline concentration (Adj. R² 0.669), by inference of vegetation stress. Higher concentrations of salts imply in an environment with greater setbacks for most of the plants, revealed by the lower NDVI values (Fig. 2).
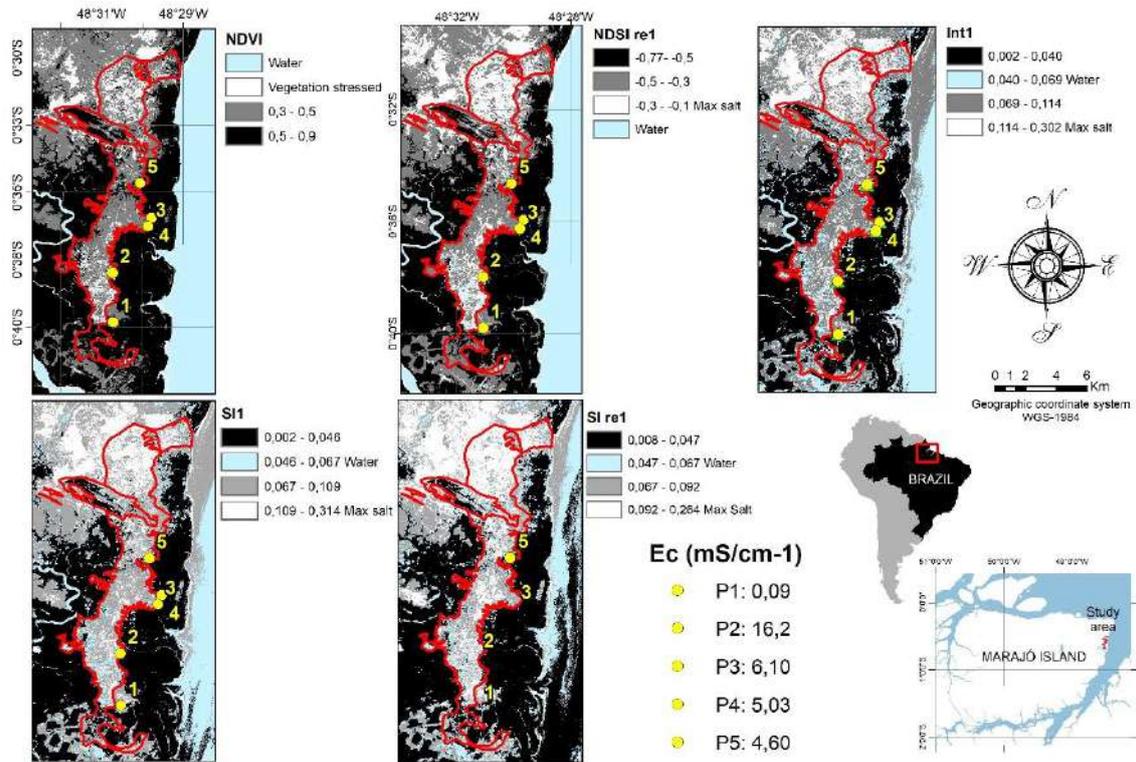
Fig. 1: Salt-affected soils in the eastern Marajó Island, Brazil
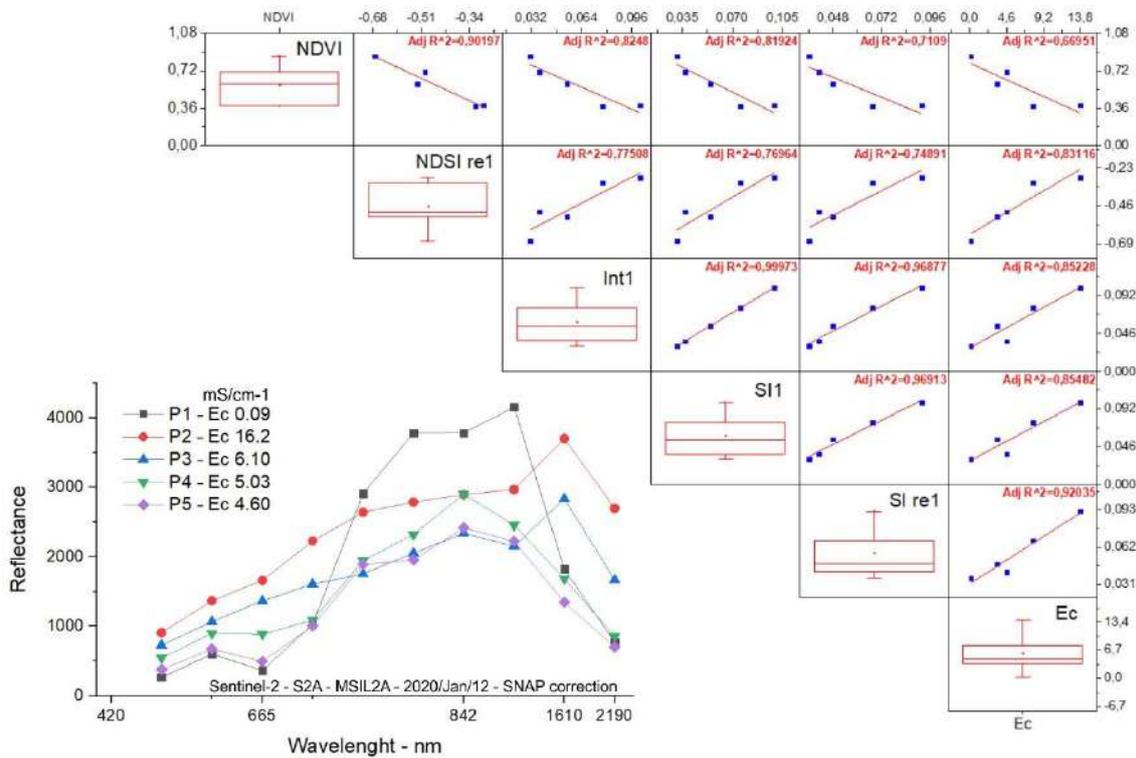


Fig. 2: Indices and spectral signature of salt-affected soils

The salinity indices (NDSI re1, Int1, SI1, and SI1 re1) showed good association with the electrical conductivity varying. The better index was SI re1, with Adj. $R^2$ 0.920, followed by SI1 (Adj. $R^2$ 0.854), Int1 (Adj. $R^2$ 0.852), and NDSI re1 (0.831). The spectral signatures showed a reflectance peak at 1610 nm (Sentinel-2, B11)

for most salt-affected soils, such as P2 ($E_c$ 16.2 mS/cm$^{-1}$) and P3 ($E_c$ 6.10 mS/cm$^{-1}$), suggesting a specific signature for salinity in Marajó Island (Fig. 2). The P1 is a non-saline soil ($E_c$ 0.09 mS/cm$^{-1}$), which spectral signature indicate the most development vegetation, with reflectance peak between 600 and 900 nm.

## Conclusions

The saline indices are satisfactory for the spatial distribution analysis of salt-affected soils in Marajó Island. The electrical conductivity ($E_c$) was best explained by the saline indices SI re1 (R²adj 0.92), SI1 (R²adj 0.85), Int1 (R²adj 0.85), and NDSI re1 (R²adj 0.83). NDVI showed R²adj 0.66 with $E_c$, represented an indirect indication of salinity according inference of vegetation stress. The salt-affect soils of eastern Marajó Island indicated a great geodiversity, with signs of marine incursion, tidal effects, rainfall accumulation, intensive evapotranspiration processes, and sea-level rising tendences. The study evidenced the great potential of remote sensing in recognize saline areas in Brazil, especially in a fluvial-marine ecosystem such as Marajó archipelago.

## References

GORJI, T. et al. Monitoring soil salinity via remote sensing technology under data scarce conditions: A case study from Turkey. **Ecol. Indic**. 74, 384–391, 2017.

KHAN M. N. et al. Mapping Salt-Affected Soils Using Remote Sensing Indicators - A Simple Approach with the Use of G1S IDRISI. **Anais…** 22nd Asian Conference on Remote Sensing, 1, 25-29, 2001.

MAIN-KNOR, M. et al. Sen2Cor for Sentinel-2. **Conference paper**, 1-12, 2017.

PAZ, A.M. et al. Prediction of soil salinity and sodicity using electromagnetic conductivity imaging. **Geoderma** 361, 2020.

USDA, 2014. **Soil Survey Field and Laboratory Methods Manual**. United States Department of Agriculture, Natural Resources Conservation Service.

FOURATI, H.T. et al. Modeling of soil salinity within a semi-arid region using spectral analysis. **Arabian Journal of Geosciences**, v.8, p.1-8, 2015.

WANG, J. et al. Capability of Sentinel-2 MSI data for monitoring and mapping of soil salinity in dry and wet seasons in the Ebinur Lake region, Xinjiang, China. **Geoderma**, 353, 172–187, 2019

WU, S. et al. Mapping the Salt Content in Soil Profiles using Vis-NIR Hyperspectral Imaging. **Soil Sci. Soc. Am. J**. 82, 1259–1269, 2018.

# Management zones from proximal soil sensors capture within-field soil property and terrain variations

VASQUES, Gustavo M.[1]; RODRIGUES, Hugo M.[1]; OLIVEIRA, Ronaldo P.[1]; HERNANI, Luís Carlos[1]; TAVARES, Sílvio R. L.[1]

[1] Embrapa Solos, gustavo.vasques@embrapa.br, rodrigues.machado.hugo@gmail.com, ronaldo.oliveira@embrapa.br, luis.hernani@embrapa.br, silvio.tavares@embrapa.br

**Thematic Session: Advances in soil sensing**

**Abstract**
The objectives are to delineate soil management zones from soil proximal sensor data, and compare soil property values among zones in a 72-ha crop field in southeastern Brazil. Apparent electrical conductivity (aEC) and magnetic susceptibility (aMS), and equivalent Th (eTh) and U (eU) were measured across the field by a Geonics EM38-MK2 and a Medusa MS1200 sensors, respectively. These properties were kriged and used as input for delineating three management zones by fuzzy k-means clustering. Soil properties were measured at 0-10 cm at 72 sites, and their means were compared among the zones. Soil clay, organic C and exchangeable Ca and Mg vary significantly among the zones, according to Brown-Forsythe and Games-Howell tests (p=0.05), while pH, available P and exchangeable K do not. Zone delineation from proximal sensor data constitutes an efficient data-driven approach to separate the field into meaningful parts for soil, irrigation and crop management based on soil variation.
Keywords: Geophysics; Electrical conductivity; Gamma radiometrics; Geostatistics; Precision agriculture

## Introduction

Site-specific crop management has been proposed as an alternative to conventional cropping that accounts for the within-field variation of soils, relief, crops and other factors aiming to increase profitability by increasing productivity, optimizing inputs and minimizing negative environmental impacts. This can be accomplished by splitting the field into homogeneous zones, so-called management zones, based on the variation of soils and/or other factors across the field. Then, each zone is managed differently by varying the rates of sowing, fertilizers, amendments, pesticides, irrigation, and other inputs, according to the characteristics of the zone.

However, assessing the soil variation can be costly if soil samples are taken on a grid with, say, one sample per hectare. Proximal and remote sensors can efficiently provide input soil data for management zone delineation, expediting sampling and reducing costs by measuring (usually electromagnetic) soil properties at many (hundreds to thousands) sites covering the field in a single survey. These sensors have been used in different regions and soil types to delineate management zones (BENEDETTO et al., 2013; HAGHVERDI et al., 2015; SCUDIERO et al., 2018; ORTUANI et al., 2019; VALLENTIN et al., 2020).

Thus, the objectives are to: (a) delineate soil management zones from soil proximal sensor data; and (b) compare soil property values among zones.

**Methodology**

The study was conducted in a 72-ha crop field under no-till crop rotation system and central pivot irrigation located in Itaí, São Paulo, southeastern Brazil, with central coordinates 23.5854° S and 48.9395° W. Soils in the area are *Latossolos* (Oxisols, Ferralsos).

To delineate soil management zones, a EM38-MK2 sensor (Geonics, Mississauga, Canada) (1-m coil spacing, vertical orientation) dragged on a rubber mat behind a pickup truck, and a MS1200 gamma radiometer (Medusa, Groningen, Netherlands) mounted on the bull bar of the truck, were used to take 4306 apparent electrical conductivity (aEC) and magnetic susceptibility (aMS), and 4896 equivalent thorium (eTh) and uranium (eU) measurements, respectively, along 25 parallel lines about 40 m apart across the field (Figure 1a). The four proximal sensor variables (aEC, aMS, eTh and eU) were kriged across the area with 5-m spatial resolution, and the kriged maps were used as input to delineate three soil management zones by fuzzy k-means clustering.

To compare soil property values among the delineated zones, a regular grid comprising one site per hectare (Figure 1a) was derived across the study area, and soil samples were taken at 0-10 cm at the 72 sites and analyzed for clay, organic C (OC), pH, available P, and exchangeable bases, according to Teixeira et al. (2017). Soil property means were compared among soil management zones by Brown-Forsythe tests (p=0.05), followed by Games-Howell post hoc tests (p=0.05).

**Results and discussion**

The spatial variations of aMS and eTh are very similar, and differ from those of aEC and especially eU (Figure 1b-e). The aEC, aMS and eTh variograms were best fitted by spherical models, while a Gaussian model was used for eU, explaining the smoother eU spatial patterns. Variogram ranges were 500, 495, 668 and 443 m for aEC, aMS, eTh and eU, respectively.

The distinct spatial features in the southwest portion of the area, observed in the aEC and aMS maps (Figure 1b, c), are due to the presence of a catchment area of a spring at the extreme southwest. This constitutes one of the delineated management zones (Figure 1f, "Southwest" zone). In comparison, the "North" zone has distinct eTh and eU values from the other zones, and the "Southeast" zone differs in aEC, aMS, eTh and eU from the other two zones (Figure 1b-e).

Soil clay, OC and exchangeable Ca and Mg vary significantly among the zones, according to Brown-Forsythe tests, with significant differences found between the "North" and "Southeast" zones for all of them, and between the "North" and "Southwest" zones for clay, according to Games-Howell tests (Table 1). Mean soil pH, available P and exchangeable K are not statistically different among zones, but P and K have noticeably larger means in the "Southeast" zone, especially against the "Southwest" zone, though they vary too much for the differences to be statistically significant (Table 1).

The "North" and "Southwest" zones could be merged based on the similarity of aMS and eTh (Figure 1c, d), and of soil properties at 0-10 cm (Table 1). However, the

"Southwest" zone has steeper slopes and higher clay content as it encompasses a catchment, while the "North" zone has plain terrain and smaller clay content. Also, the "Southwest" zone has wetter soils, due to its relief position, which is evident from the aEC map (Figure 1b). Thus, keeping the "Southwest" zone apart from the other zones for soil and irrigation management is recommended.

On the other hand, the northernmost portion of the area has similar aEC, aMS and eTh values to the "Southeast" zone (Figure 1b, c, d, f) in contrast to the "North" zone where it belongs. It also has higher clay, OC, Ca, Mg and K compared to the rest of the "North" zone (not shown). Thus, it could be split from the "North" zone, and whether managing it separately is worth the extra effort could be evaluated.

## Conclusions

Proximal soil sensors capture soil variation patterns across the field, providing a large amount of data that can be used to delineate soil management zones efficiently. The properties the sensors measure (aEC, aMS, eTH and eU) are affected by the soil constituents and by relief and water dynamics, and thus, they indirectly carry information on soil formation factors and processes, which is encapsulated in the delineated zones, reducing the need for extra data.

In turn, the proposed delineated zones need to be judged from: a soil perspective with the aid of field soil samples besides proximal sensor data; a terrain perspective, if the area has variable, irregular terrain, which is the case in the "Southwest" zone; and from an agronomic and logistic perspective, pondering soil, irrigation, and crop management. As such, the results presented in the study are open for discussion, field testing and decision making, for which the farmer needs to be involved.

## Acknowledgements

## References

BENEDETTO, D. et al. Field partition by proximal and remote sensing data fusion. **Biosystems Engineering**, v. 114, p. 372-383, 2013.

HAGHVERDI, A. et al. Perspectives on delineating management zones for variable rate irrigation. **Computers and Electronics in Agriculture**, v. 117, p. 154-167, 2015.

ORTUANI, B. et al. Integrating geophysical and multispectral data to delineate homogeneous management zones within a vineyard in Northern Italy. **Sensors**, v. 19, n. 3974, p. 1-22, 2019.

SCUDIERO, E. et al. Workflow to establish time-specific zones in precision agriculture by spatiotemporal integration of plant and soil sensing data. **Agronomy**, v. 8, n. 253, p. 1-21, 2018.

TEIXEIRA, P. C. et al. **Manual de métodos de análise de solo**. 3. ed. rev e ampl. Brasília: Embrapa, 2017.

VALLENTIN, C. et al. Delineation of management zones with spatial data fusion and belief theory. **Precision Agriculture**, v. 21, p. 802-830, 2020.

Table 1. Variation of soil properties among management zones. Equal letters indicate equal means among zones, according to Games-Howell tests at p=0.05.

| Property | N | Mean | Stdev | N | Mean | Stdev | N | Mean | Stdev |
|---|---|---|---|---|---|---|---|---|---|
| | | North | | | Southeast | | | Southwest | |
| Clay (g kg$^{-1}$) | 33 | 392[b] | 49 | 27 | 430[a] | 38 | 12 | 433[a] | 21 |
| OC (g kg$^{-1}$) | 33 | 14[b] | 1 | 27 | 16[a] | 1 | 12 | 15[ab] | 2 |
| pH | 33 | 6.6[a] | 0.3 | 27 | 6.6[a] | 0.3 | 12 | 6.4[a] | 0.4 |
| P (mg dm$^{-3}$) | 33 | 141[a] | 83 | 27 | 151[a] | 63 | 12 | 127[a] | 55 |
| Ca (cmol$_c$ dm$^{-3}$) | 33 | 6.0[b] | 0.8 | 27 | 6.7[a] | 0.8 | 12 | 6.4[ab] | 1.0 |
| Mg (cmol$_c$ dm$^{-3}$) | 33 | 1.8[b] | 0.2 | 27 | 2.1[a] | 0.3 | 12 | 1.9[ab] | 0.3 |
| K (cmol$_c$ dm$^{-3}$) | 33 | 451[a] | 880 | 27 | 583[a] | 1094 | 12 | 197[a] | 52 |

N, number of observations; Stdev, standard deviation.
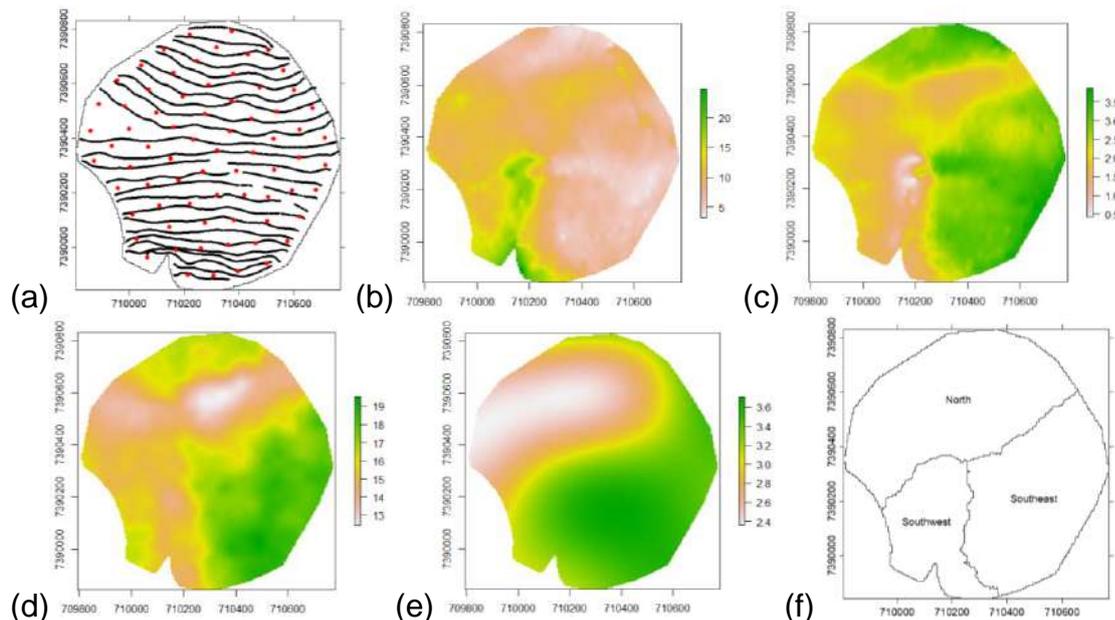


Figure 1. (a) Proximal sensor survey lines (black lines) and soil sampling sites (red dots); (b-e) Kriged maps of aEC (mS m$^{-1}$), aMS (ppt), eTh (ppm) and eU (ppm), respectively; and (f) Delineated soil management zones. Coordinates are in UTM zone 22S.

# The granulometry causes change on determination of soil color via Nix Pro 2 sensor

CONDÉ, Viviane Flaviana[1]; BUENO, Thays Vieira[2]; OLIVEIRA, Jéssica Ribeiro[3]; IBRAIMO, Anifo Soares Mamudo[4]; FRANCELINO, Márcio Rocha[5]; FERNANDES-FILHO, Elpídio Inácio[6]

[1] Universidade Federal de Viçosa (UFV), vivianefconde@gmail.com; [2] UFV, thaysvierabueno@gmail.com; [3] UFV, jessica.r.ribeiro@ufv.br; [4] UFV, anifo.ibraimo@ufv.br; [5] UFV, marcio.francelino@ufv.br; [6] UFV, elpidio@ufv.br

**Thematic Session: Advances in Soil Sensing**

**Abstract**

Usually, Soil Color is visually estimated in the field with the Munsell Soil Color Charts. Visual definitions are subject to uncontrolled effects that can change results. Our work aimed to evaluate the influence of granulometry in the determination of soil color using the sensor Nix Pro 2. Eighteen soil samples in different granulometries (2 mm, 0.297 mm, 0.149 mm and 0.074 mm) were used and analyzed with Nix Pro 2. The RGB values obtained in the analysis with the sensor were converted to Munsell standard through an R software script. The particle size changes the spectral behavior of soil samples for frequencies in the visible region and the Nix Pro 2 color sensor can optimize soil color determination.

Keywords: Soil sensing, Tropical soil, Munsell Soil Color Charts, RGB

## Introduction

Soil Color is used to soil identification and classification, for estimate various soil properties such as the presence of iron, organic matter, drainage conditions and texture. Spectral measuring instruments make it possible to determine colors accurately and there are many color sensors. The Nix Pro 2 is a low-cost, proximal, active color sensor that emits a light beam with a spectrum in the visible region ("Nix Sensor", 2021). It's necessary to improve techniques using sensors for color determination in tropical soils and sample preparation. The objective of this work was to evaluate the influence of sample particle size in the determination of soil color using the Nix Pro 2 sensor.

## Methodology

Eighteen samples of soil profiles collected in the states of Goiás and Tocantins were air-dried, ground in an agate mortar and sieved in a stainless steel sieve with an opening of 2 mm, 0.297 mm, 0.149 mm and 0.074 mm. 9 g of each sample were put in a bottle and the surface was covered with plastic PVC food film. The sensor settings were made so that all samples were analyzed under the same conditions of light (D50), opening angle (10º) and humidity with 15 repetitions of readings.

The RGB values obtained by Nix Pro 2 were used to calculate the hue, value and chroma. And from these three variables were possible to calculate the correspondent variable in Munsell color system. This analyses were made through the R software (R CORE TEAM, 2020) and the packages: aqp (BEAUDETTE; ROUDIER; O'GEEN, 2013) , dplyr (WICKHAM et al., 2021), modest (PALLMANN, 2017), patchwork (PEDERSEN, 2020), readr (WICKHAM; HESTER; FRANCOIS, 2020), WriteXLS (SCHWARTZ, 2021).

**Results and discussion**

The influence of soil granulometry on color is related to the specific surface of the soil components and the spectral response. The uniformity of the analyzed material influences the performance of the sensors. Sifting the soil is an alternative to homogenize. However, in the breaking process, quartz grains from sand particles and other soil minerals will have their granulometry reduced, which may change the perception of colors. Studies indicate that in sandy-textured soils, small variations in organic matter can lead to large variations in color (DEMATTÊ et al., 2011; MORITSUKA et al., 2014).

When converting to the Munsell system, there were small variations in value and chroma for practically all soil samples in the four particle sizes used, but 11 of the 18 samples showed equal results in the 2 mm and 0.297 mm particle sizes. Five soil samples (1, 5, 7, 16 and 17) had a variation in hue, which represents an important change in color nomenclature (Table 1). In general, the 2 mm and 0.297 mm particle sizes presented a performance similar to the readings taken in the field. In the field classification, the samples were grouped into ten classes up to the second categorical level, as follows: ARGISSOLO AMARELO (17,18); ARGISSOLO VERMELHO (8); CAMBISSOLO HÁPLICO (4,9,14,15); LATOSSOLO AMARELO (6,7); LATOSSOLO VERMELHO (11); NEOSSOLO QUARTZARÊNICO (5); NITOSSOLO VERMELHO (10); ORGANOSSOLO HÁPLICO (12, 13); PLINTOSSOLO HÁPLICO (1); PLINTOSSOLO PÉTRICO (2,3,16) (SANTOS et al., 2018).

Red soils have 10R and 2.5YR hues, red-yellow soils have 5YR hues and yellow soils have 7.5YR and 10YR hues (SANTOS et al., 2018). Of the seven soil samples classified with color in the second categorical level, only two soil samples did not match with the determination made in the field with Munsell Soil Color Charts (MSCC) when analyzed using a Nix Pro 2 color sensor. Sample 8 (ARGISSOLO VERMELHO) was classified as red-yellow, 5YR hue regardless of particle size. Sample 7 (LATOSSOLO AMARELO) was classified as yellow-red (hue 7.5YR) in the particle size 0.074 mm and red-yellow (hue 5YR) for the other particle sizes.

Table 1: Colors of soil samples in Munsell Soil Color Charts (MSCC) in different particle sizes

| Soil | mm | MSCC* | Color name | Soil | mm | MSCC* | Color name |
|---|---|---|---|---|---|---|---|
| 1 | 2 | 2.5YR 6/4 | Light yellowish brown | 10 | 2 | 2.5YR 3/4 | Dark olive brown |
| | 0,297 | 2.5YR 6/4 | Light yellowish brown | | 0,297 | 2.5YR 4/5 | Olive brown |
| | 0,149 | 5YR 6/5 | Light reddish brown | | 0,149 | 2.5YR 4/6 | Olive brown |
| | 0,074 | 5YR 6/5 | Light reddish brown | | 0,074 | 2.5YR 4/6 | Olive brown |
| 2 | 2 | 2.5YR 5/6 | Light olive brown | 11 | 2 | 2.5YR 4/5 | Olive brown |
| | 0,297 | 2.5YR 5/6 | Light olive brown | | 0,297 | 2.5YR 4/5 | Olive brown |
| | 0,149 | 2.5YR 6/6 | Olive yellow | | 0,149 | 2.5YR 4/5 | Olive brown |
| | 0,074 | 2.5YR 5/7 | Light olive brown | | 0,074 | 2.5YR 4/5 | Olive brown |
| 3 | 2 | 5YR 5/6 | Yellowish red | 12 | 2 | 5YR 3/1 | Very dark gray |
| | 0,297 | 5YR 5/5 | Reddish brown | | 0,297 | 5YR 3/1 | Very dark gray |
| | 0,149 | 5YR 5/5 | Reddish brown | | 0,149 | 5YR 4/2 | Dark reddish gray |
| | 0,074 | 5YR 6/5 | Light reddish brown | | 0,074 | 5YR 4/2 | Dark reddish gray |
| 4 | 2 | 7.5YR 6/4 | Light brown | 13 | 2 | 5YR 3/1 | Very dark gray |
| | 0,297 | 7.5YR 6/4 | Light brown | | 0,297 | 5YR 3/2 | Dark reddish brown |
| | 0,149 | 7.5YR 6/4 | Light brown | | 0,149 | 5YR 4/2 | Dark reddish gray |
| | 0,074 | 7.5YR 7/4 | Pink | | 0,074 | 5YR 5/2 | Reddish gray |
| 5 | 2 | 5YR 5/4 | Reddish brown | 14 | 2 | 5YR 4/5 | Reddish brown |
| | 0,297 | 7.5YR 6/4 | Light brown | | 0,297 | 5YR 4/5 | Reddish brown |
| | 0,149 | 7.5YR 7/4 | Pink | | 0,149 | 5YR 5/5 | Reddish brown |
| | 0,074 | 7.5YR 7/3 | Pink | | 0,074 | 5YR 5/5 | Reddish brown |
| 6 | 2 | 7.5YR 5/3 | Brown | 15 | 2 | 5YR 4/5 | Reddish brown |
| | 0,297 | 7.5YR 5/3 | Brown | | 0,297 | 5YR 4/5 | Reddish brown |
| | 0,149 | 7.5YR 6/4 | Light brown | | 0,149 | 5YR 5/5 | Reddish brown |
| | 0,074 | 7.5YR 6/3 | Light brown | | 0,074 | 5YR 5/5 | Reddish brown |
| 7 | 2 | 5YR 6/5 | Light reddish brown | 16 | 2 | 7.5YR 6/4 | Light brown |
| | 0,297 | 5YR 5/5 | Reddish brown | | 0,297 | 5YR 6/4 | Light reddish brown |
| | 0,149 | 5YR 6/5 | Light reddish brown | | 0,149 | 5YR 6/4 | Light reddish brown |
| | 0,074 | 7.5YR 6/5 | Light brown | | 0,074 | 5YR 6/4 | Light reddish brown |
| 8 | 2 | 5YR 4/4 | Reddish brown | 17 | 2 | 7.5YR 5/4 | Brown |
| | 0,297 | 5YR 5/5 | Reddish brown | | 0,297 | 7.5YR 5/5 | Brown |
| | 0,149 | 5YR 5/5 | Reddish brown | | 0,149 | 10YR 6/5 | Light yellowish brown |
| | 0,074 | 5YR 5/5 | Reddish brown | | 0,074 | 10YR 6/4 | Light yellowish brown |
| 9 | 2 | 7.5YR 7/3 | Pink | 18 | 2 | 7.5YR 5/4 | Brown |
| | 0,297 | 7.5YR 7/3 | Pink | | 0,297 | 7.5YR 5/4 | Brown |
| | 0,149 | 7.5YR 7/3 | Pink | | 0,149 | 7.5YR 6/4 | Light brown |
| | 0,074 | 7.5YR 7/3 | Pink | | 0,074 | 7.5YR 6/4 | Light brown |

## Conclusions

The particle size changes the spectral behavior of soil samples for frequencies in the visible region. The 2mm and 0.297mm grain sizes performed similarly to field readings. The Nix Pro 2 color sensor can optimize soil color determination, however

analytical procedures require a careful methodology to ensure reproducibility and repeatability.

**References**

BEAUDETTE, D. E.; ROUDIER, P.; O'GEEN, A. T. Algorithms for quantitative pedology: A toolkit for soil scientists. **Computers & Geosciences**, v. 52, p. 258–268, mar. 2013.

DEMATTÊ, J. A. M. et al. Quantificação de matéria orgânica do solo através de modelos matemáticos utilizando colorimetria no sistema Munsell de cores. **Bragantia**, v. 70, n. 3, p. 590–597, 30 set. 2011.

MORITSUKA, N. et al. Soil color analysis for statistically estimating total carbon, total nitrogen and active iron contents in Japanese agricultural soils. **Soil Science and Plant Nutrition**, v. 60, n. 4, p. 475–485, 4 jul. 2014.

**Nix Sensor**. Disponível em: <https://www.nixsensor.com/>. Acesso em: 2 abr. 2021.

PALLMANN, P. **Package: modest (Model-Based Dose-Escalation Trials)**. . [S.l: s.n.]. , 2017

PEDERSEN, T. L. **Package ' patchwork ': The Composer of Plots**. . [S.l: s.n.]. , 2020

R CORE TEAM. **R: A Language and Environment for Statistical Computing**. . Vienna, Austria: [s.n.]. , 2020

SANTOS, H. G. dos et al. **Sistema Brasileiro de Classificação de Solos**. 5. ed. rev ed. Brasília, DF: Embrapa, 2018., 2018.

SCHWARTZ, M. **WriteXLS: Cross-Platform Perl Based R Function to Create Excel 2003 (XLS) and Excel 2007 (XLSX) Files**. . [S.l: s.n.]. , 2021

WICKHAM, H. et al. Package ' dplyr '. 2021.

WICKHAM, H.; HESTER, J.; FRANCOIS, R. **Package ' readr ': Read Rectangular Text Data**. . [S.l: s.n.]. , 2020

# Soil Organic Carbon mapping using different algorithms as soil sensing technique applied to extreme environment.

MEIER, Martin[1]; SCHMITZ, Daniela; FRANCELINO, Márcio R.
[1] Federal University of Viçosa, martin.meier@ufv.br; [2] Federal University of Viçosa, daniela.schmitz@ufv.br; [3] Federal University of Viçosa, marcio.francelino@ufv.br

**Thematic Session: Soil Sensing**

**Abstract**

The importance of evaluating the level of carbon in soils affects ecological and agricultural decisions. In addition, soil sensing techniques are very useful and are allowing a revolution in agriculture and environmental management, providing to decision makers important information. To demonstrate the usefulness of soil sensing technique applied to a mapping of nutrients in a determinate area, the objective of this study was to determine carbon content and distribution, comparing two different interpolation algorithms to perform geoprocessing mapping of an ice-free area in the Antarctica. We tested two different algorithms mapping processes. Both algorithms (IDW and Kernel) did confirm that carbon is concentrated at the northern part of the ice-free area, but also can be founded at the eastern part. This levels of carbon concentration are evidences of nutrient contribution that bird colonies perform to soil nutrient, mainly to the surface soil layer as presented in the results.

Keywords: climate change; environmental management; geoprocessing.

## Introduction

The importance of carbon level in soils has its bulwarks on the ecological and agricultural point of view (Bhattacharyya et al., 2015). In addition, soil sensing techniques are very useful to agricultural and environmental management, providing important information about the projects sites peculiarities to decision makers.

Using soil sensing as a technique applied in pedometric and pedology science, is useful and cheaper in several situations. Nowadays it can be used in a well-developed and perfectly controlled agronomic projects in the tropics, e.g. agriculture 4.0 or precision agriculture. By the other hand, can be applied to studies or research' missions on extreme environments where time and weather are not predictable and may change in seconds, like both Arctic and Antarctic polar environments.

The latter environment is well researched by the Terrantar (scientific group part of the National Institute of Science and Technology – INCT/Federal University of Viçosa - UFV) regarding to pedology and soil classification. To demonstrate the usefulness of soil sensing technique applied to a mapping of nutrients in a determinate area, the objective of this study was to determine carbon content and distribution, comparing two different interpolation algorithms to perform geoprocessing carbon contend mapping in an ice-free area in the Antarctica.

## Methodology

### Study area
The Stansbury peninsula (62°15 S 58°59 W) is located in the northern part of Nelson Island, part of the South Shetland Islands, with a total area of 165 km², and only 5% of ice-free areas (Rodrigues et al. 2020). The sub-Antarctic maritime climate classification, according to Köppen, includes strong winds, high weather variability, and relatively mild temperatures varying from average annual air temperature of -2.2 ° C and the average summer air temperature slightly greater than 0 ° C (Ferron et al. 2004). Precipitation varies from 350 and 500 mm per year with rainfalls during summer.

### Sample Physics and Chemic variables
We used 33 sampling points across the ice-free area collecting soil samples in the superficial soil layer (0 to 5 cm depth). Soil texture was determined by dispersion in distilled water, sieved and weighted to separate the coarse and fine sand fractions, followed by sedimentation to determine the silt and clay fractions (Gee and Bauder, 1986). The chemical analysis was performed by routine process to determine the main soil variables (pH2O, pKCl, Ca, Mg, K, Na, P, Al, Al+H, SB, t, T, V, m, Isna, C and Prem), in the Soil Fertility Laboratory of the University of Viçosa – UFV. The specific attribute that was object of this study, total organic carbon (TOC) content, was estimated after the determination of organic carbon using Walkley Black adapted method, without heating (Bhattacharyya et al., 2015).

### Exploratory Statistical and Geospacial analysis

Using the results obtained by the physical and chemical analysis performed to the soil samples, we performed a descriptive and experimental statistical analysis and a statistically correlation through Pearson's Matrix among the attributes. The attributes that show correlations with carbon have been discussed. We used all values in a dataset converted to points georeferenced to perform an interpolation test for TOC using Geostatistical Analyst Tools of the ArcGIS 10.3 software. We tested two different algorithms mapping processes: Inverse distance weighted (IDW) interpolation; and Kernel Interpolation. Both these maps will be used to discuss concentration of TOC in the soil superficial layer across the ice-free area.

## Results and discussion

At the surface layer, soils presented textural class mostly Sandy loam, with a maximum clay content of 25%, and an average contend of 16,33%. The chemical characteristics suggest an incipient process of soil formation, a general eutrophic character and high bases saturation ($V_{mean}$ = 85,24%).

The soil attributes studied of the ice-free area, show that TOC in the according to Pearson's correlation matrix (Figure 1) had positive correlation (in blue circles) with coordinate y (metric - UTM), phosphorous ($P_{mean}$ = 144,10 mg.dm$^{-3}$ / $P_{median}$ = 78,30

mg.dm⁻³) and modest correlation with potential acidity ($H+Al_{mean}$ = 4,91 $cmol_c.dm^{-3}$). In the other hand, TOC presented negative correlation (in red circles) with pH2O and base saturation (V), and also a discrete negative correlation with Calcium (Ca).
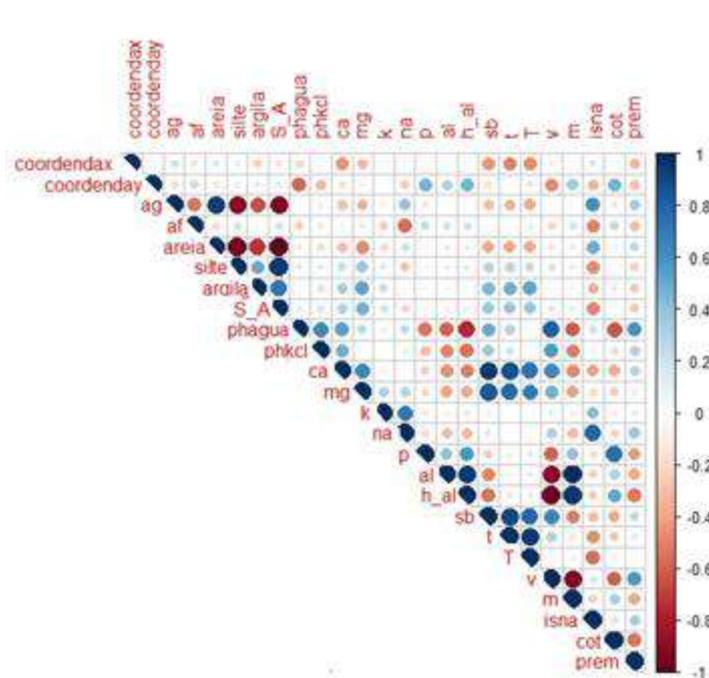


**Figure 1.** Person's correlation matrix of all variables of the 33 sampling points. Red indicates negative correlation, blue positive correlation and white no correlation. Circle radius determine magnitude of correlation (From 1 to -1).
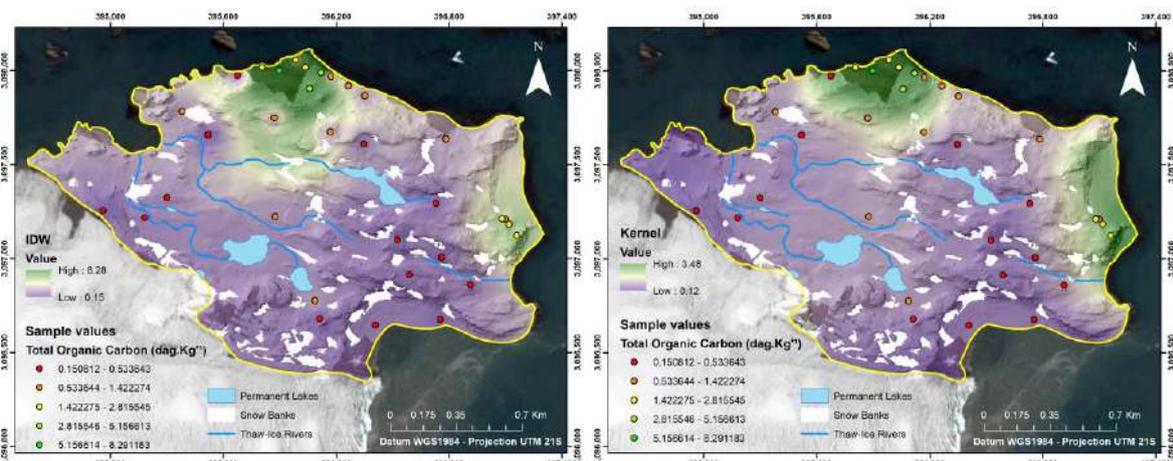


**Figure 2.** Comparison of maps produced by the two different interpolation algorithms IDW and Kernel, of the ice-free area in Nelson Island, Antarctica.

The positive correlation with the metric attribute "coordinate Y" (projection used was Universal Transversa of Mercator – UTM 21S) reveals the gradient that latitude is producing for TOC contend in the soil surface layers around the ice-free area. The

most at north of the peninsula Stansbury you may go, more soil carbon there you will find.

The detached correlation indicated by Persons Matrix over the position in the environment discussed in this brief abstract can be confirmed when performing the mapping process with interpolation algorithms as presented in Figure 2. Both algorithms (IDW and Kernel) did confirm that carbon is concentrated at the northern part of the ice-free area, but also can be founded at the eastern part. This levels of carbon concentration are evidences of nutrient contribution that bird colonies perform to soil nutrient, mainly to the surface soil layer as presented.

## Conclusions

There is an expressive concentration of soil organic carbon at two different regions of the ice-free area, with predominance to the northwest region followed by the east side, both next to the sea. These areas can be distinguished by the rest of the ice-free area, once the latter lacks on wild life presence. Both areas where TOC was higher demonstrate in numbers the pronounced difference in the maps produced by both algorithms. In addition, comparing which differences were observed among them, IDW had higher sensibility and detached more pronouncedly the northwest region to the eastern region, in comparison to Kernel.

## Acknowledgements

## References

Bhattacharyya, T., et al. (2015). Walkley-Black Recovery Factor to Reassess Soil Organic Matter: Indo-Gangetic Plains and Black Soil Region of India Case Studies. Communications in Soil Science and Plant Analysis, 46(20), 2628–2648. https://doi.org/10.1080/00103624.2015.1089265.

Gee, GW, Bauder, JW, 1986. Particle-size analysis. In: Klute, A. (Ed.), Methods of Soil Analysis. Part 1: Physical and Mineralogical Methods. Soil Science Society of America, Madison, pp. 383–412.

Rodrigues, et al. (2021). Phosphatization under birds' activity: Ornithogenesis at different scales on Antarctic Soilscapes. Geoderma, 391(February). https://doi.org/10.1016/j.geoderma.2021.114950).

# Proximal and remote sensor data fusion for in-depth salinization mapping in the Brazilian semiarid via machine learning

TAVARES, Sílvio R. L.[1]; VASQUES, Gustavo M.[1]; OLIVEIRA, Ronaldo P.[1]; DANTAS, Marlon M.[2]; RODRIGUES, Hugo M.[3]

[1] Embrapa Solos, silvio.tavares@embrapa.br, gustavo.vasques@embrapa.br, ronaldo.oliveira@embrapa.br, luis.hernani@embrapa.br; [2] Instituto Federal do Rio Grande do Norte, dantasmm1@gmail.com; [3] Universidade Federal Rural do Rio de Janeiro, rodrigues.machado.hugo@gmail.com

**Thematic Session: Advances in soil sensing**

## Abstract

Mapping the salinization in irrigated cropland is a challenging practice. As an alternative, data from proximal and remote sensors have been implemented together via datafusion and machine learning algorithms. The present work was carried out on a farm with 11 ha and used data from the proximal sensor EM38-MK2 associated with radar C-band data obtained by the Sentinel1 satellite. The salinization classes were created from electrical conductivity data measured at 35 points using a 50 x 50 m sampling grid and at three depths: 0 – 10, 10 – 30, and 30 – 50 cm using conventional laboratory approach. The accuracy values of the class prediction models presented values between 0.66 and 0.74 and Kappa values between 0.43 and 0.59 using Random Forest. The salinization decreased in layers 0 - 10 and 10 - 30 cm due to implementing a surface drainage system but the depth 30 - 50 cm had the highest occurrence of Salic classes, with a potentially harmful effect on the roots.

## Introduction

The salinized regions in high-temperature areas occur via water evaporation from the soil, transpiration for vegetables and the carry-over of salts that settle on the surface. Some agricultural techniques have been proposed to mitigate these effects, such as using water with lower electrical conductivity values and applying drainage systems to remove salts. However, methods to monitor the occurrence of salinization after the application of mitigating activities are challenging, as identifying the occurrence of soil salinization requires a high number of soil samples for laboratory analysis through the analysis of electrical conductivity in saturated paste this being a financially costly method that does not allow covering large areas.

Proximal sensors have been proposed as an alternative to monitoring areas where salinization occurs. Apparent electrical conductivity and apparent magnetic susceptibility data have been reported as potential attributes that allow to identify and map salinization advance or retreat since these sensor attributes are closely related to clay content, moisture, cation exchange capacity, and pH (LOPES; MONTENEGRO, 2019). As an aid to mapping the occurrence of salinization in irrigated plantation areas, it is also possible to combine data from proximal sensors with other data sources, such as radar data obtained by satellites (HUANG; PROCHAZKA; TRIANTAFILIS, 2016) and which are freely available for use.

Therefore, the objective of the work was to spatialize the occurrence of soil salinization from the predictive mapping of the salinization classes reflected by the electrical conductivity measured in the laboratory as a function of proximal and remote sensor data via machine learning algorithm Random Forest.

**Methodology**

The study was conducted on a family farm at Baixo Açu irrigated perimeter and has approximately 11 ha. It is in northeastern Brazil in the region of Alto do Rodrigues – RN (Figure 2). The region's climate is Aw via Köppen-Geiger and has a rainfall regime with an average annual occurrence of 400 mm and an average annual maximum temperature of 34°C. Due to its semiarid conditions, the area is subject to natural soil salinization processes that have been increased by irrigated crop production.

The EM38-MK2 (Geonics Limited, Mississauga, Canada) was used for continuous aEC and aMS readings (N=5,168 points) on zig-zag footstep tracks in "1 m" (aEC and aMS 1 m) and "0. 5 m" (aEC and aMS 0.5 m) coil separation mode on vertical orientation. These data were spatially characterized using semivariogram adjustments and then interpolated by ordinary kriging with 10 m resolution. One thousand five hundred fifty-one points or 30% of the original dataset were intended to validate the four maps produced (Figure 2).

The two vertical-vertical and vertical-horizontal polarizations present in the C-band of the Sentinel-1 satellite were selected (Figure 1. h. and i.).

For soil salinity from laboratory analysis, soil core samples were collected in a 35 points uniform grid (50 x 50 m; Figure 1), at 0 – 10, 10 – 30, and 30 – 50 cm depth, and the samples were analyzed using the method of the electrical conductivity measured in the saturated paste ($EC_{lab}$) as described in Embrapa's methods manual (TEIXEIRA et al., 2017). In addition, the pH data at the 35 points were also measured, and then these were spatialized using the inverse square distance method.

The salinity data for the three depths were classified according to their degree of salinization using the limits defined for the characteristics "Not saline" ($EC_{lab} < 4$ dS m$^{-1}$), "Saline" ($4$ dS m$^{-1} < EC_{lab} < 7$ dS m$^{-1}$), and "Salic" ($7$ dS m$^{-1} < EC_{lab}$) using the limits defined in the Brazilian Soil Classification System (SANTOS et al., 2018).

The set of covariates comprises four maps of proximal sensors stacked with the two radar polarized band maps and the three pH maps, totaling nine predictor covariates. $EC_{lab}$ data were modeled and mapped from the model's fit for each salinity class by depth using the Random Forest classifier present in the caret package in the R software and will be evaluated for accuracy using the kappa index and leaving one out cross-validation.

**Results and discussion**

The aEC and aMS maps showed higher values in the central locations with a tendency to grow to the east, agreeing with the direction of drainage (Figure 1; a., b.). The pH maps show high values (> 7) for the entire study area, reinforcing the presence of salts in the soil (Figure 1; e., f., g.). The radar C-band also showed higher values in the west and northwest regions of the study area, agreeing with the drainage orientation (Figure 1; h., i.).

The aEC and aMS maps showed external validation errors less than 80 mS/m and 0.2 ppt, respectively. The pH covariates for all depths were the most important in all salinization Random Forest models (Table 1). While the 1 m aMS was the second most important for the 10 – 30 and 30 – 50 cm salinization models considering the data from

proximal sensors. The 0 – 10 cm salinization model did not show high significance between remote and proximal sensor data. Accuracy values for all models were around 0.7, while kappa values were close to 0.5 (Table 1).

The depth salinity map (Figure 3) shows a higher concentration of the "Not saline" class in the center and northeast of the maps, in agreement with the behavior of the aEC maps shown in Figure 2. The occurrence of salinization shows a decrease as it approaches the surface due to the existence of drainage channels built to drain the water used, represented in figure 3 in blue lines. The 30 - 50 cm map has a higher occurrence of the "Salic" class, demonstrating that it is a layer with toxic effects for some crops where the drainage activity may not have been enough to cannot attenuate the effects of salinization.

## Conclusions

The data from proximal electromagnetic sensors combined with radar data obtained by remote sensors allowed to spatialize the phenomenon of salinization in an irrigated crop area with good accuracy via adjustments to prediction models using the Random Forest algorithm.

The pH data measured in the laboratory was fundamental for the construction of predictive models of salinization.

Data from remote and proximal sensors proved to be essential tools for monitoring and mapping the effects of salinization on the soil.

## References

HUANG, J.; PROCHAZKA, M. J.; TRIANTAFILIS, J. Irrigation salinity hazard assessment and risk mapping in the lower Macintyre Valley, Australia. **Science of the Total Environment**, v. 551–552, p. 460–473, 2016.

LOPES, I.; MONTENEGRO, A. A. D. A. Spatialization of electrical conductivity and physical hydraulic parameters of soils under different uses in an alluvial valley. **Revista Caatinga**, v. 32, n. 1, p. 222–233, 2019.

SANTOS, H. G. DOS; JACOMINE, P. K. T.; ANJOS, L. H. C. DOS; OLIVEIRA, V. Á. DE; LUMBRERAS, J. F.; COELHO, M. R.; ALMEIDA, J. A. DE; FILHO, J. C. DE A.; OLIVEIRA, J. B. DE; CUNHA, T. J. F. **Sistema brasileiro de classificação de solos**. 5. ed. Brasília, Brazil: Embrapa, 2018.

TEIXEIRA, P. C.; DONAGEMMA, G. K.; FONTANA, A.; TEIXEIRA, W. G. **Manual de métodos de análise de solo**. 3. ed. rev ed. Brasília, DF: Embrapa, 2017.
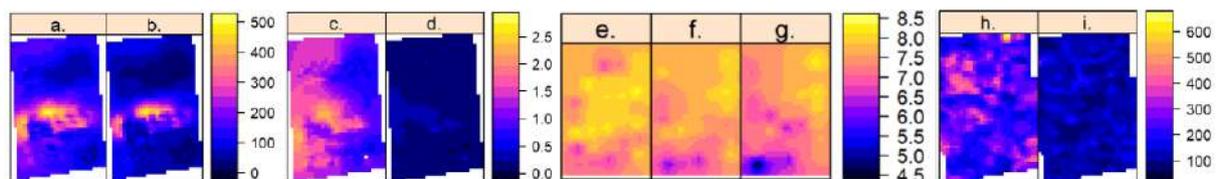
Figure 1: a. aEC 1 m; b. aEC 0.5 m; c. aMS 1 m; d. aMS 0.5 m; e. pH 0-10 cm; f. pH 10-30 cm; g. pH 30-50 cm; h. C-band vertical-vertical Sentinel-1; C-band vertical-horizontal Sentinel-1.
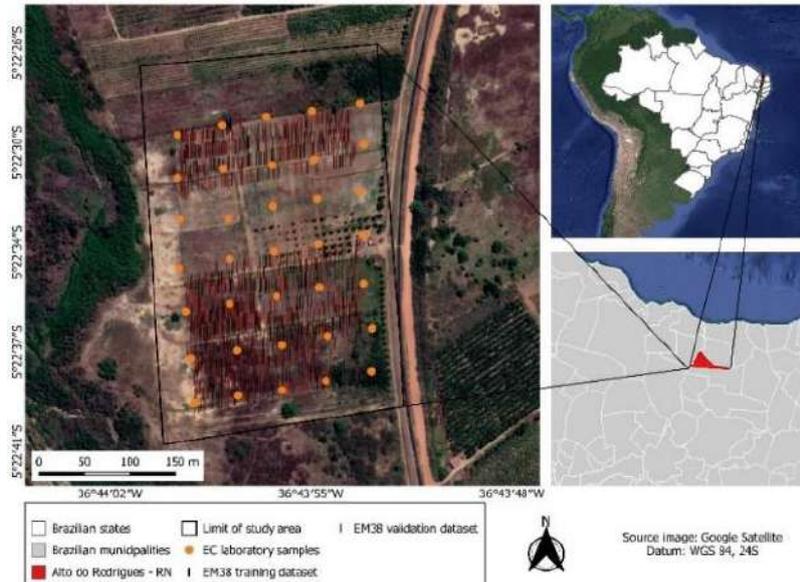
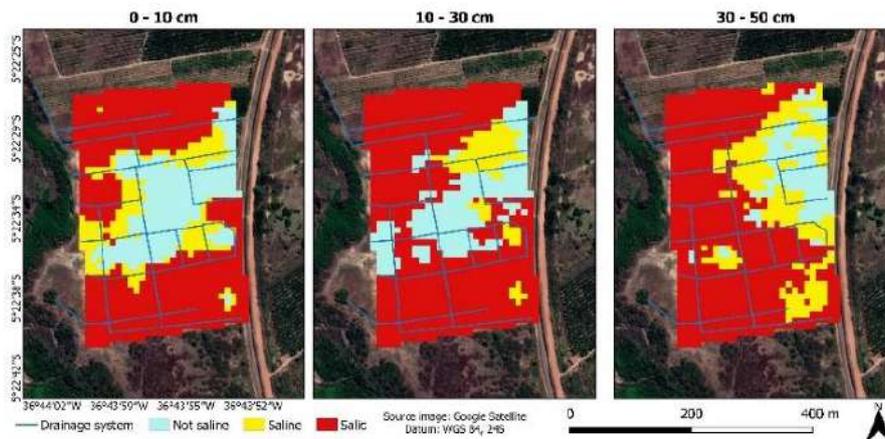Figure 2: Location map of the study area and the sampling design.



Figure 3: Spatialized occurrence of soil salinization in depth. A) 0 – 10 cm in depth; B) 10 – 30 cm in depth; C) 30 – 50 cm deep.

Table 1: Ranking of the importance of covariates in the Random Forest model and the accuracy and kappa values for each layer.

| Salinity 0 - 10 cm | importance | Salinity 10 - 30 cm | importance | Salinity 30 - 50 cm | importance |
|---|---|---|---|---|---|
| pH 0 – 10 cm | 100 | pH 0 – 10 cm | 100 | pH 30 – 50 cm | 100 |
| B2 | 10.943 | pH 10 – 30 cm | 62.58 | pH 0 – 10 cm | 57.38 |
| pH 10 – 30 cm | 9.17 | aMS 1 m | 48.3 | aMS 1 m | 52.05 |
| aMS 0.5 m | 5.27 | aEC 0.5 m | 44.64 | B1 | 27.35 |
| aMS 1 m | 4.998 | pH 30 – 50 cm | 43.51 | aEC 0.5 m | 25.41 |
| aEC 0.5 m | 3.835 | aEC 1 m | 32.86 | pH 10 – 30 cm | 19.91 |
| pH 30 – 50 cm | 3.485 | B2 | 30.38 | aEC 1 m | 13.13 |
| aEC 1 m | 3.028 | aMS 0.5 m | 20.16 | aMS 0.5 m | 12.02 |
| B1 | 0 | B1 | 0 | B2 | 0 |

| | Salinity 0 - 10 cm | Salinity 10 - 30 cm | Salinity 30 - 50 cm |
|---|---|---|---|
| Accuracy | 0.74 | 0.71 | 0.66 |
| Kappa | 0.59 | 0.52 | 0.43 |

# Near-infrared spectroscopy: PLS calibration models for prediction of potentially toxic elements in soil and sediments from Ipojuca river watershed, Pernambuco

MAIA, Angelo Jamil[1,a]; SILVA, Yuri Jacques Agra Bezerra[2]; NASCIMENTO, Rennan Cabral[1,b]; VERAS, Germano[3]; NASCIMENTO, Clístenes Williams Araújo[1,c]; SILVA, Ygor Jacques Agra Bezerra[1,d];

[1] Universidade Federal Rural de Pernambuco, Departamento de Agronomia,
[a]request.angelo@gmail.com, [b]rennancabral2@gmail.com, [c]clistenes.nascimento@ufrpe.br,
[d]ygor.silva@ufrpe.br; [2] Universidade Federal do Piauí, Departamento de Agronomia,
yuriufrpe@yahoo.com.br; [3] Universidade Estadual da Paraíba, Laboratório de Química Analítica e Quimiometria, germano@servidor.uepb.edu.br.

## Thematic Session: Advances in soil sensing

**Abstract**

The Ipojuca river watershed carries several environmental impact activities that can contribute to soil and water contamination by potentially toxic elements (PTEs). Monitoring PTEs concentrations is important for controlling transfer processes to water bodies, but traditional analytical methods present several limitations. Thus, the aim of this work was to create calibration models based on near-infrared (NIR) spectroscopy for PTEs prediction in soil and sediments from Ipojuca river watershed. It were collected 101 soil and sediments samples, which had its spectra measured at the range of 1000 - 2500 nm. Optical emission spectrometry was carried out for measuring 24 PTEs concentrations. Three sample sets were subdivided and models were generated for each one: I. All samples; II. Channel banks III. Bedload sediments. Reasonable results were obtained, strengthening the idea of infrared spectroscopy as a viable tool for predicting the concentrations of several potentially toxic elements.

Keywords: Spectrometry; pollution; proximal sensing; watershed management.

## Introduction

In Ipojuca river watershed there are several economic activities presenting high potential environmental impact (Barros et al., 2013), such as intensive agriculture, lack of basic sanitation in urban areas (Sobral et al., 2016), and also Suape port industrial complex (Muniz et al., 2005). The presence of such activities raises concerns regarding the environmental quality within this watershed, which can be affected due to the input of pollutants.

Potentially toxic elements (PTEs) are those which present a relative impact in human health and environmental quality. Common examples of PTEs are heavy metals and rare earth elements (Liu et al., 2021). The contamination of these elements in soil pose threat to human health (Shaheen et al., 2020), and the association of PTEs with sediments is known to occur in processes of transference from soil to water environments (Usman et al., 2021), leading to environmental, social and economic issues (Samiee et al., 2019). PTEs concentrations in bed sediments can be used as indicators of contamination (Silva et al., 2017), and can also provide geochemical

and pedogenetic information when combined with soil data (Fernandes et al., 2018; Bantan et al., 2020).

The analysis of PTEs concentrations is a basic step on monitoring the presence of these elements in the environment. However, the traditional analytical methods for determining PTEs concentrations are mainly expensive, destructive, time-consuming, and use hazardous reagents (Cozzolino et al., 2016). Near-infrared (NIR) spectroscopy is an alternative method for predicting several elements concentrations using chemometric calibration models (Camargo et al., 2018; Zhang et al., 2019). Thus, the aim of this work was to test near-infrared spectroscopy as an alternative method for prediction of several PTEs in soil and bed sediments samples from Ipojuca river watershed.

## Methodology

The study area comprises the Ipojuca river watershed (3400km²) which is completely inserted in Pernambuco state, with its main water course extending for 320 km from upstream semiarid to downstream Atlantic forest and coastal outlet. It was collected 101 soil and bed sediments samples distributed in different land uses, in order to represent the geological and pedological diversity within the area: *Caatinga* drylands savannah (n=12), sugarcane crops (n=15), channel banks (n=31), unpaved roads (n=13), bed sediments (n=25), corn crops (n=2), and pastures (n=3).

All samples were air-dried and sieved in 2 mm mesh sieves for NIR spectroscopy analysis, and a fraction of each sample was separated and sieved in 38 µm mesh sieves for total digestion (10 ml HF, 10 ml $HNO_3$, 6 ml $HClO_4$ and 5 ml HCl) and posterior analysis in inductively coupled plasma optical emission spectrometry, for obtaining Al, Ba, Be, Co, Cr, Fe, Mn, Mo, Ni, Pb, Sn, Sr, Ti, V, Zn, Ce, Gd, La, Nd, Pr, Sc, Sm, Y, and Th concentrations. Certified sample SRM 2710a Montana I soil (NIST, 2002) was used to ensure the analytical quality.

The spectra of the samples were measured in 1000 – 2500 nm range using a FT-IR/NIR Spectrometer (Frontier/PerkinElmer) coupled with a Reflectance Accessory, with 2 nm resolution at a 0.5 nm window size and 32 accumulations for each sample.

PLS models for concentration prediction were built for each element using raw spectral data and also Savitzky-Golay and Standard Normal Variate preprocessings. The models were built using three different sample sets: all samples, channel banks and bed sediments. Calibration and cross-validation steps were evaluated through determination coefficient (R²), root mean square error (RMSE) and bias. Also, the elliptical joint confidence region (EJCR) test was used in calibration to assess the occurrence of systematic errors in the models.

## Results and discussion

The elements average concentrations decreased in the following order (mg kg$^{-1}$): Al (91,159.5), Fe (30,627.77), Ti (4,662.93), Ba (924.39), Mn (505.64), Sr (196.83), Ce (131.8), Zn (62.53), La (58.4), V (51.21), Pb (48.5), Cr (33.78), Th (30.49), Nd (30.25), Ni (14.81), Y (14.77), Pr (11.44), Sm (7.62), Co (6.83), Sn (6.48), Sc (5.94), Gd (5.6), Mo (2.15), Be (2.02).

The models which obtained the best results (higher R², lower RMSE and bias) and were also evaluated in EJCR test (has no systematic errors in calibration step) were marked in bold at Table 1.

**Table 1.** R² values for cross-validation step ($R^2_{cv}$)

| Elements | All samples (n=101) | | | Channel banks (n=31) | | | Bed sediments (n=25) | | |
|---|---|---|---|---|---|---|---|---|---|
| | RD | SG | SNV | RD | SG | SNV | RD | SG | SNV |
| Al | 0.87 | 0.89 | 0.87 | **0.89** | **0.89** | **0.88** | **0.81** | **0.85** | **0.90** |
| Ba | 0.58 | 0.57 | 0.51 | 0.51 | 0.56 | **0.58** | 0.43 | 0.28 | 0.07 |
| Be | 0.59 | 0.52 | 0.57 | 0.35 | 0.51 | 0.52 | 0.62 | 0.55 | 0.66 |
| Ce | -0.04 | 0.18 | 0.02 | 0.11 | 0.30 | 0.20 | -0.38 | -0.02 | 0.02 |
| Co | 0.51 | 0.56 | 0.42 | 0.21 | 0.28 | **0.55** | 0.19 | -0.19 | -0.04 |
| Cr | 0.41 | 0.52 | 0.46 | 0.07 | 0.23 | 0.18 | 0.21 | 0.21 | -0.24 |
| Fe | 0.84 | 0.85 | 0.83 | **0.75** | 0.68 | 0.72 | **0.83** | **0.81** | **0.83** |
| Gd | 0.37 | 0.43 | 0.49 | 0.53 | 0.58 | **0.74** | 0.28 | 0.43 | 0.50 |
| La | 0.27 | 0.25 | 0.19 | 0.16 | 0.35 | **0.51** | -0.46 | 0.03 | -0.11 |
| Mn | 0.05 | 0.22 | 0.16 | 0.22 | -0.06 | 0.38 | 0.13 | 0.05 | -0.11 |
| Mo | 0.77 | 0.80 | 0.79 | 0.74 | 0.79 | 0.74 | **0.80** | 0.61 | 0.68 |
| Nd | 0.17 | 0.15 | -0.01 | 0.26 | 0.32 | **0.53** | 0.15 | 0.28 | 0.29 |
| Ni | 0.52 | 0.58 | 0.55 | -0.03 | 0.41 | 0.37 | 0.07 | -0.04 | -0.10 |
| Pb | -0.11 | -0.08 | -0.18 | -0.19 | -0.08 | -0.47 | 0.27 | 0.12 | 0.32 |
| Pr | 0.20 | 0.16 | 0.08 | 0.20 | 0.38 | 0.30 | -0.68 | -0.19 | -0.11 |
| Sc | 0.77 | 0.78 | 0.73 | **0.68** | 0.61 | **0.70** | 0.56 | 0.65 | 0.64 |
| Sm | 0.16 | 0.23 | 0.22 | 0.21 | 0.26 | 0.30 | -0.48 | 0.28 | 0.30 |
| Sn | 0.31 | 0.50 | 0.40 | 0.06 | 0.18 | -0.43 | 0.04 | -0.03 | -0.05 |
| Sr | 0.30 | **0.53** | 0.26 | 0.42 | 0.34 | 0.18 | 0.21 | 0.31 | 0.37 |
| Th | 0.13 | 0.04 | -0.26 | -0.04 | 0.05 | -0.03 | 0.04 | 0.10 | -0.42 |
| Ti | 0.88 | **0.90** | 0.91 | **0.88** | **0.87** | 0.80 | **0.86** | **0.89** | **0.89** |
| V | 0.79 | **0.82** | 0.80 | **0.72** | 0.78 | 0.81 | 0.61 | 0.70 | 0.71 |
| Y | 0.56 | 0.61 | 0.59 | 0.59 | 0.63 | **0.79** | 0.44 | 0.54 | 0.58 |
| Zn | 0.36 | 0.16 | -0.06 | 0.24 | 0.32 | 0.20 | 0.06 | -0.26 | -0.67 |

*RD Raw data, SG Savitzky-Golay, SNV Standard Normal Variate*

Considering the models that were validated in calibration step and also kept a R²>0.80 in cross-validation step, it is possible to highlight: for all samples, Ti and V using SG; for channel banks, Al using raw data, SG and SNV and Ti using raw data and SG; for bed sediments, Al, Fe and Ti using raw data, SG and SNV and Mo using raw data. Al, Fe and Ti are the elements with higher concentrations in the samples, and have also presented good or reasonable result in the models for all three sample sets. It is possible that high concentrations of the elements and the performance of the models are correlated. The good result of Mo in bed sediments set using raw spectral data is possibly explained by the adsorption correlation of Fe-minerals and chemical species of Mo in anoxic sediments (Xu et al., 2006), since the average concentration of the element is low (2.15 mg kg$^{-1}$).

## Conclusions

It is possible to build viable calibration models for prediction of potentially toxic elements concentrations in soil and bed sediments from Ipojuca river watershed. Overall, the results support the idea of NIR spectroscopy as a viable alternative method for prediction of chemical elements in soil and bed sediments in general.

# References

BANTAN, R. A.; AL-DUBAI, T. A.; AL-ZUBIERI, A. G. **Geo-environmental assessment of heavy metals in the bottom sediments of the Southern Corniche of Jeddah, Saudi Arabia**. Marine Pollution Bulletin, v. 161, p. 111721, 2020.

BARROS, A.M.L.; SOBRAL, M.C.; GUNKEL, G. **Modelling of point and diffuse pollution: application of the Moneris model in the Ipojuca river basin, Pernambuco State, Brazil**. Water science and technology, v. 68, n. 2, p. 357-365, 2013.

CAMARGO, L.A. et al. **Predicting potentially toxic elements in tropical soils from iron oxides, magnetic susceptibility and diffuse reflectance spectra**. Catena, v. 165, p. 503-515, 2018.

COZZOLINO, D. **Near infrared spectroscopy as a tool to monitor contaminants in soil, sediments and water—State of the art, advantages and pitfalls**. Trends in Environmental Analytical Chemistry, v. 9, p. 1-7, 2016.

FERNANDES, A. R. et al. **Quality reference values and background concentrations of potentially toxic elements in soils from the Eastern Amazon, Brazil.** Journal of Geochemical Exploration, v. 190, p. 453-463, 2018.

LIU, J. et al. **Quantifying and predicting ecological and human health risks for binary heavy metal pollution accidents at the watershed scale using Bayesian Networks.** Environmental Pollution, v. 269, p. 116125, 2021.

MUNIZ, K. et al. **Hydrological impact of the port complex of Suape on the Ipojuca River (Pernambuco-Brazil).** Journal of Coastal Research, v. 21, n. 5, p. 909-914, 2005.

NIST-National Institute of Standards and Technology. **Standard reference materials -SRM 2709, 2710 and 2711**. Addendum Issue Date: January 18 2002.

SAMIEE, F. et al. **Exposure to heavy metals released to the environment through breastfeeding: A probabilistic risk estimation**. Science of The Total Environment, v. 650, p. 3075-3083, 2019.

SHAHEEN, S. M. et al. **Soil contamination by potentially toxic elements and the associated human health risk in geo-and anthropogenic contaminated soils: A case study from the temperate region (Germany) and the arid region (Egypt).** Environmental Pollution, v. 262, p. 114312, 2020.

SILVA, Y. J. A. B. et al. **Bedload as an indicator of heavy metal contamination in a Brazilian anthropized watershed**. Catena, v. 153, p. 106-113, 2017.

SOBRAL, M. et al. **Evolution of the Monitoring Water Quality System in Ipojuca River Basin, Brazil**. Watershed and River Basin Management, p. 94-100, 2016.

USMAN, Q.A. et al. **Spatial distribution and provenance of heavy metal contamination in the sediments of the Indus River and its tributaries, North Pakistan: Evaluation of pollution and potential risks**. Environmental technology & innovation, v. 21, p. 101184, 2021.

XU, N.; CHRISTODOULATOS, C.; BRAIDA, W. **Adsorption of molybdate and tetrathiomolybdate onto pyrite and goethite: effect of pH and competitive anions.** Chemosphere, v. 62, n. 10, p. 1726-1735, 2006.

ZHANG, X. et al. **Predicting cadmium concentration in soils using laboratory and field reflectance spectroscopy**. Science of the Total Environment, v. 650, p. 321-334, 2019.

## Soil Moisture Monitoring in the Federal District, using Geotechnologies

OLIVEIRA, Carlos Eduardo Meireles.
[1] Brasília University, carlos.meireles.oliveira@gmail.com;

## Thematic Session: ADVANCES IN SOIL SENSING

### Abstract

The present work aimed at monitoring soil moisture through geotechnologies, mainly through active remote sensing, by obtaining undeformed soil samples in different locations in the Federal District in order to analyze and understand the relationships between soil moisture and radar (Sentinel-1) images from each location.
Keywords: Pedology, Remote Sensing, Sentinel-1, Water.

### Introduction

Soil moisture content is considered an indicator of climate change. The water crisis that occurred in the Federal District (DF) in 2017 led to the rationing of water resources that impacted social and agro-environmental activities. Thus, the need to monitor soil moisture arises since, in the water balance, soils are the means by which rainwater supplies groundwater and water bodies, including water reservoirs.

### Methodology

By consulting cartographic material, references on soil properties and geotechnologies, especially remote sensing, it was possible to subdivide the region into 4 sectors and thus determine 40 sampling points. Subsequently, field trips were made to collect deformed samples, which were submitted to laboratory tests to determine the physical and chemical attributes, as well as the pedological classification according to the Brazilian Soil Classification System - SiBCS. As for the undeformed samples, necessary for determining soil moisture, they were collected synchronously with the dates of the Sentinel-1 satellite passages. The measured soil moisture data

were modeled by means of simple linear regression with the radar data as a covariate for moisture recovery.

**Results and discussion**

The pedomorphogeological relationships of the soils and the land use and occupation helped for a characterization and classification of the soils. The data generated enabled the preparation of correlation graphs necessary to observe the relationships between variables such as soil attributes and the Sentinel-1 data backscatter values. The Sentinel-1 images allowed for a reasonable mapping of the soil moisture indices of the Federal District, representing the distribution of soil moisture in each of the 4 predetermined sectors.
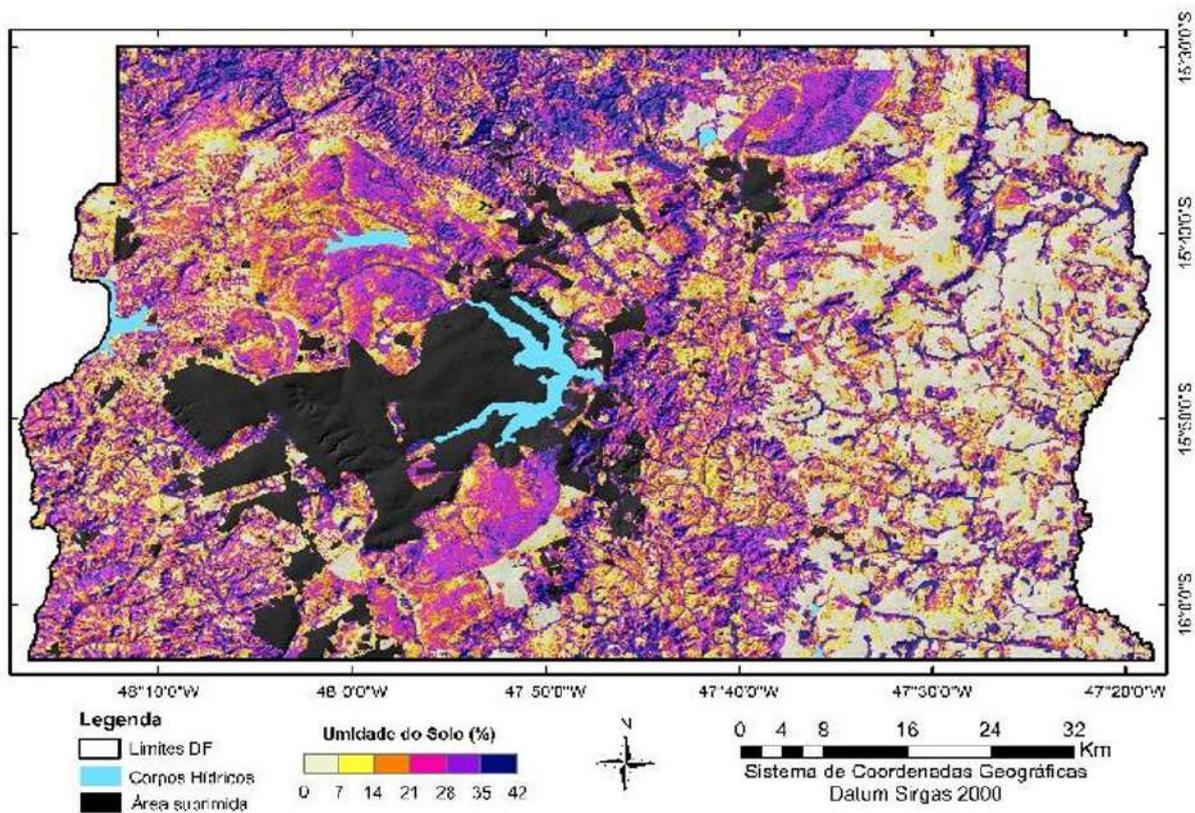


Figure 1 - Moisture rate map of the DF soils produced from the treatment of the obtained data.

## Conclusions

Variations in soil moisture produce significant changes in the energy balance of the earth's surface. Thus, this soil attribute configures an effective parameter in the modeling of hydrological processes and soil formation and consequently shows the importance of its determination by remote sensing data, since the spatial analysis of large areas makes the work unfeasible. The use of remote sensing, geoprocessing of images and other geotechnological tools, makes spatial analysis feasible because classic survey techniques present difficulties such as lack of multi-temporal data, the need for many field measurements, etc. However, environmental factors that affect the soil moisture recovered by radar data such as vegetation cover and soil roughness must be considered.

## Acknowledgements

## References

SANTOS, R. D. dos; LEMOS, R. C. de; SANTOS, H. G. dos; KER, J. C.; ANJOS, L. H. C. dos; SHIMIZU, S. H. Manual de descrição e coleta de solo no campo. 7. ed. rev. e ampl. Viçosa, MG: Sociedade Brasileira de Ciência do Solo, 2015. 102 p.

Sistema Brasileiro de Classificação de Solos / Humberto Gonçalves dos Santos … [et al.]. – 5. ed., rev. e ampl. − Brasília, DF: Embrapa, 2018. 356 p.: il. color.; 16 cm x 23 cm.

# Iron rods in support of the use of ground penetrating radar with 450 MHz monostatic antenna for imaging argillic horizon

NASCIMENTO, Carlos[1]; VASQUES, Gustavo M.[2]; RODRIGUES, Hugo[3]; CEDDIA, Marcos[4]

[1] Universidade Federal Rural do Rio de Janeiro, carloswagner.geologia@yahoo.com; [2] Embrapa Solos, gustavo.vasques@embrapa.br; [3] Universidade Federal Rural do Rio de Janeiro, rodrigues.machado.hugo@gmail.com; [4] Universidade Federal Rural do Rio de Janeiro, marcosceddia@gmail.com

**Thematic Session: Advances in soil sensing**

## Abstract

This work aimed to use a GPR equipped with a 450 MHz antenna to identify the depth of the clayey horizon marked with iron rods in Planosols. This study was carried out in an agroecological farm (Fazendinha Agroecológica do km 47), located in Seropédica municipality, Rio de Janeiro state, Brazil. First, three Planosols were described, and iron rods with dimensions of 80 cm in length and 0.8 cm in diameter were inserted in the transitions among soil horizons. Next, radargrams from the soil profiles and from one transect with those profiles were acquired using a GPR porting a shielded monostatic antenna of 450 MHz. Then, a depth model was adjusted to the transect radargram using the arithmetic average from the pulse velocities measured among the individual radargrams from the soil profiles. As the result, the average velocity created a depth model for the transect radargram that produced similar depths to those transitions viewed in the field for the argillic horizon.
Keywords: soil survey; shallow geophysics; Planosols.

## Introduction

The ground penetrating radar (GPR) has been shown potential to identify and map soil features (Ucha et al., 2002). One of those features can be the bottom of the E horizon in Planosols, that is, the depth to the argillic horizon. Its estimation is essential in order to study the soil water storage volume, the depth of the sandy horizons for soil tillage purposes, and other uses. In addition, this feature could also be used to locate the boundaries among soil types in the field, possibly expediting soil survey (Ucha et al., 2002). Then, this work aimed to analyze the feasibility of iron rods in support of a GPR porting a monostatic antenna of 450 MHz to determine the depth of the argillic horizon in an area with Planosols.

## Methodology

This work was carried out in an agroecological farm (Fazendinha Agroecológica do km 47), located in Seropédica municipality, Rio de Janeiro state, southeastern Brazil. Three trenches were opened, where soil profiles were described according to Santos et al. (2015) and classified according to the World Reference Base (IUSS Working Group WRB, 2015) as Planosols, named of P2 (with six horizons), P5 (with seven horizons) and P6 (with six horizons). Disturbed soil samples were collected at each horizon and analyzed in the laboratory to measure the particle size fractions (PSF) and the gravimetric water content, according to Teixeira et al. (2017).
Iron rods with dimensions of 80 cm in length and 0.8 cm in diameter were inserted in the transitions among soil horizons. Then, the three soil profiles (2-m long by 1.5-m wide by 1.20 to 1.72-m deep) and one transect (220-m long) with those profiles were imaged using the GPR MALÅ GroundExplorer (Guideline Geo AB, Sundbyberg,

Sweden), creating radargrams using a monostatic shielded antenna of 450 MHz. The radargrams obtained in the field were processed using the ReflexW software (Sandmeier, 2009), and two pre-processing procedures were done in sequence: static correction and dewow. After pre-processing, the hyperbolas from the iron rods were identified in the radargrams, and the pulse velocity estimations were obtained in ReflexW, followed by the conversion of the Y-axis of the radargram from time (ns) to depth units (m).

**Results and discussion**

The main results of the physical attributes from the horizons are shown in Table 1, which identify the abrupt transitions (the top of the argillic horizon). Each soil profile has a higher clay content in the B horizon concerning the E horizon, leading to a higher K (dieletric constant) contrast between these horizons (De Benedetto et al., 2010), which may help see the argillic horizon in the 450 MHz radargram.

Table 1. Physical attributes of the P2, P5, and P6 soil profiles.

| Profile | Horizon | Horizon number | Depth (cm) | PSF (g kg$^{-1}$) Sand | Clay | Gravimetric water (%) | Horizon transition |
|---------|---------|----------------|------------|------|------|----------------------|--------------------|
| P2 | E | 3 | 22-69 | 831 | 92 | 2.1 | 3 |
|    | Bt1 | 4 | 69-92 | 564 | 388 | 5.9 | |
| P5 | E5 | 6 | 145-158 | 717 | 20 | 6.1 | 6 |
|    | Bt | 7 | 158-172$^{+}$ | 621 | 267 | 4.2 | |
| P6 | E | 3 | 32-44 | 878 | 45 | 4.1 | 3 |
|    | Btg1 | 4 | 44-66 | 703 | 216 | 12.3 | |

Then, the GPR was used on the soil profiles and the pulse velocities were measured in the radargrams using the ReflexW (Table 2). From those velocities was calculated an average to apply to the depth model in the transect radargram (Figure 1).

Table 2. Pulse velocities acquired at the horizon transitions in the radargrams.

| Profile | Horizon transition | Pulse velocities |
|---------|--------------------|------------------|
| P2 | 1, 2, 3 | 0.11 m ns$^{-1}$ |
| P5 | 1, 2, 3 | 0.11 m ns$^{-1}$ |
| P6 | 1, 3, 4 | 0.10 m ns$^{-1}$ |

The iron rod at the abrupt change in P5 (Table 1) was not possible to observe (Figure 1b). It may be caused by its higher depth (at 1.58 m) in relation to P2 (at 69 cm) and P6 (at 44 cm) (Porsani, 1999), even with the upper horizons with low water content and composed by sandy texture (Doolittle and Collins, 1995).
The arithmetic average from those velocities is 0.107 m ns$^{-1}$ for that specific day of soil survey. It means that according to the environmental conditions of the day, those velocities may change, for instance, due to the water content in the soil (Doolittle and Collins, 1995). Then, considering that average to the transect radargram, the depths of the argillic horizons were very similar to those viewed in the field description (in P2 and P6) (Table 1) and for the last rod seen in the P5 (at 38 cm, third transition). This means that the iron rods can be feasible to map the argillic horizon in large areas where Planosols appear, using the frequency of 450 MHz for shallower argillic horizons. However, since antennas with lower frequencies can map deeper regions in the soil (Porsani, 1999), iron rods can also help map greater depths.
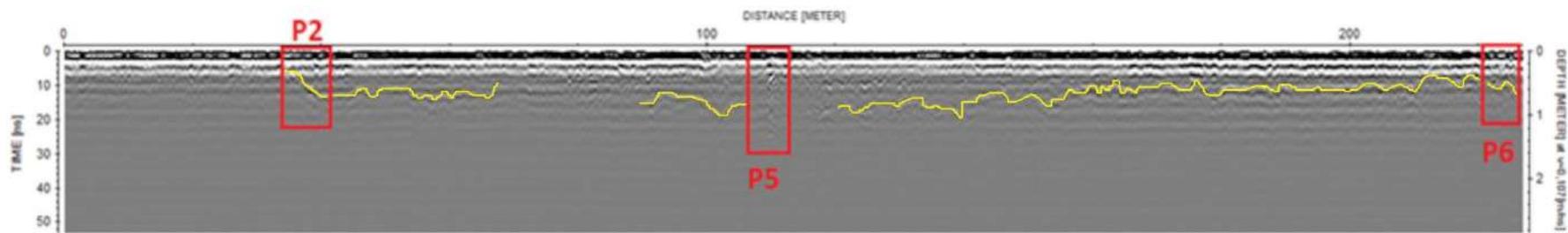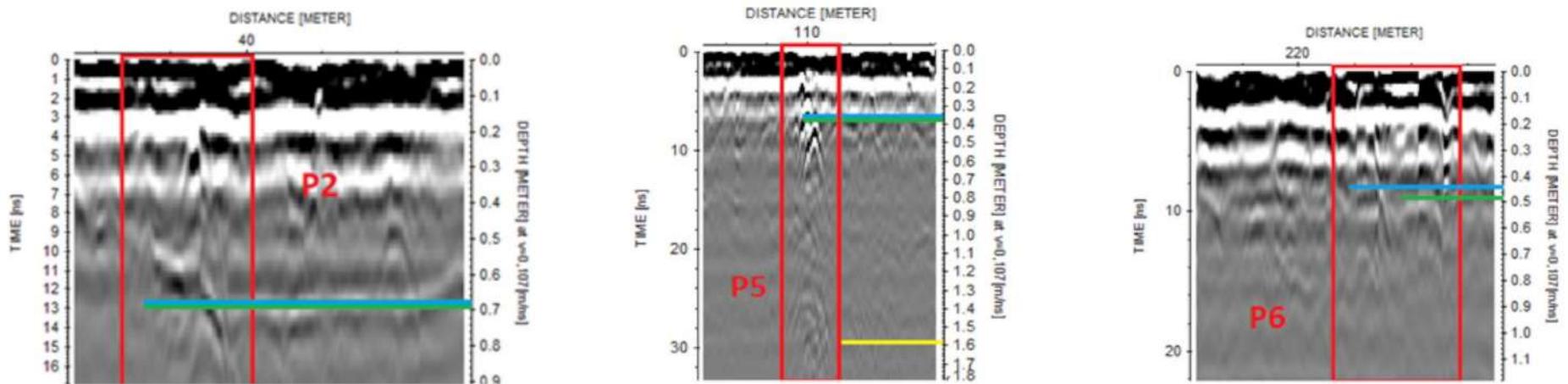
Figure 1. In (a), the transect radargram with the P2, P5, and P6 profiles (in red) and the top of the argillic horizon marked in yellow, when it is possible to be seen in the radargram. In (b) is represented the detail of each profile from the transect radargram. The blue bar in each radargram demarcates the hyperbola of the deepest iron rod in each profile, while the green bar represents the real depth seen in the field. In P2 and P6, the green bars are close to the actual depth of the base of the E horizon (69 and 44 cm depth, simultaneously). In P5, the yellow bar represents the base depth of the E5-to-Bt transition (at 1.58 m).

## Conclusions

The iron rods supporting the 450 MHz antenna showed a good potential to identify argillic horizon where Planosols dominate in significant areas, especially when the E-to-B transitions are placed in shallower depths.

## Acknowledgements

## References

DE BENEDETTO D. et al. Spatial relationship between clay content and geophysical data. **Clay Minerals Journal**, Cambridge, v. 45, p. 197-207, 2010.

DOOLITTLE, J. A.; COLLINS, M. E. Use of soil information to determine application of ground penetrating radar. **Journal of Applied Geophysics**, Amsterdam, v. 33, p. 101-108, 1995.

IUSS Working Group WRB. **World Reference Base for soil resources 2014, update 2015** – International soil classification system for naming soils and creating legends for soil maps. Rome, Italy, 2015. 192 p.

PORSANI, J. L. **Ground penetrating radar (GPR)**: proposta metodológica de emprego em estudos geológico-geotécnicos nas regiões de Rio Claro e Descalvado-SP. 1999. 145 f. Tese (Doutorado em Geofísica) – Universidade Estadual Paulista, Rio Claro. 1999.

SANDMEIER, K. J. **ReflexW Version 8.5** – Windows™ XP/7/8/10-program for the processing of seismic, acoustic or electromagnetic reflection, refraction and transmission data. Karlsruhe, Germany, 2009. 634 p.

SANTOS, R. D. et al. **Manual de descrição e coletas de solos no campo**. Viçosa, MG, 2015. 102 p.

TEIXEIRA, P. C. et al. **Manual de métodos de análise de solo**. Brasília, DF, 2017. 573 p.

UCHA, J. M. et al. Uso do radar penetrante no solo (GPR) na investigação dos solos dos tabuleiros costeiros no litoral norte do estado da Bahia. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 26, p. 373-380, 2002.

# An optimized sample for assessing soil property variations across the field and within management zones efficiently

VASQUES, Gustavo M.[1]; RODRIGUES, Hugo M.[1]; TAVARES, Sílvio R. L.[1]; HERNANI, Luís Carlos[1]; OLIVEIRA, Ronaldo P.[1]

[1] Embrapa Solos, gustavo.vasques@embrapa.br, rodrigues.machado.hugo@gmail.com, ronaldo.oliveira@embrapa.br, luis.hernani@embrapa.br, silvio.tavares@embrapa.br

**Thematic Session: Advances in soil sensing**

## Abstract

An optimized sampling design to assess soil property variation across the field and within management zones is proposed and validated in a 72-ha crop field in southeastern Brazil. An optimized sample (18 sites) was derived by spatial simulated annealing from proximal sensor covariates. Soil properties were measured at 0-10 cm and validated against those measured at 72 sites on a regular grid. The optimized and regular grid samples had equal global spatial trend models and means for soil clay, pH and exchangeable Ca, Mg and K, and different ones for organic C and available P. Within zones, equal means between sampling designs were found for all soil properties in the "North" zone, and for most properties in the other two zones. Soil property correlations against proximal sensor variables were honored by the optimized samples in most cases, both globally and within zones. The optimized soil sample reduces costs while keeping most soil information for guiding management decisions.

Keywords: Proximal soil sensing; Spatial simulated annealing; Spatial trends; Precision agriculture

## Introduction

Site-specific soil management requires knowing the spatial distribution of soil properties that guide management recommendations. Producing this information using uniform soil sampling on a regular grid across the field may be expensive due to soil sampling and analysis costs. Alternatively, on-the-go field sensors measure soil properties at many sites covering the field efficiently (ADAMCHUK et al., 2004) and can provide data to delineate management zones (VASQUES et al., 2021) and optimize soil sampling (DOMENECH et al., 2017).

For optimizing soil sampling, it is desirable that the number of sites is reduced while keeping enough soil information to support management decisions. For that, an optimized sampling design can be proposed, considering management zones and soil variation measured by proximal sensors, and validated to confirm that it represents soil property variation across the field and within zones.

Thus, the objectives are to: (a) produce an optimized sampling design to assess soil property variation; (b) compare global spatial trend models from optimized *versus* regular grid samples; and (c) compare soil property means and correlations against proximal sensor variables from optimized *versus* regular grid samples, globally and within management zones.

## Methodology

Three management zones were delineated on a 72-ha no-till irrigated crop field in Itaí, São Paulo, southeastern Brazil, by k-means clustering based on kriged maps of

proximal sensor variables, including apparent electrical conductivity (aEC) and magnetic susceptibility (aMS) measured by a EM38-MK2 sensor (Geonics, Mississauga, Canada), and equivalent thorium (eTh) and uranium (eU) contents measured by a MS1200 gamma radiometer (Medusa, Groningen, Netherlands) (VASQUES et al., 2021). Soils in the field are *Latossolos* (Oxisols, Ferralsos).

To assess soil property variation across the field, a regular grid sampling design comprising 72 sites was derived (Figure 1a, black dots). An optimized sampling design comprising 18 sites (Figure 1a, red dots) was derived by selecting six sites in each zone by spatial simulated annealing (SAMUEL-ROSA, 2019) reproducing the marginal distributions and correlations among aEC, aMS, eTh and eU. Soil samples were taken at 0-10 cm at the 90 sites (72+18) and analyzed for clay, organic C (OC), pH, available P, and exchangeable bases, according to Teixeira et al. (2017) (Figure 1b-h). Sensor variable values from their kriged maps were extracted to the 90 sites.

To check whether the optimized samples capture the global spatial trends of soil properties, analyses of variance and F tests (p=0.05) were used comparing first-degree spatial trend models – soil property=f(x*y) – against full models including the sampling design and interaction terms – soil property=f(x*y*sampling design). In addition, spatial trend models were derived from optimized and regular grid samples, respectively, and compared by Chow's test (p=0.05).

Welch's analysis of variance was used to compare soil property means from the optimized *versus* regular grid samples globally, using all observations from both sets, and locally at each zone, respectively. Soil property correlations against proximal sensor variables from the optimized *versus* regular grid samples were compared at p=0.05 using Fisher r-to-z transformation of correlation coefficients, both globally and at each zone, respectively.

**Results and discussion**
The global spatial trend models did not differ significantly between optimized and regular samples for all soil properties except OC and available P, according to both F and Chow's tests. Soil OC and available P models differed significantly between sampling designs in the regression intercepts, but not in the slopes of either the x or y variable, that is, the geographic coordinates. This shows that all soil property trends described by the regular grid samples in both the E-W and N-S directions were captured by the optimized samples.

Globally, the Welch's tests showed that only OC and available P differed significantly between optimized and regular grid samples, though OC means were similar (Table 1). Locally, all soil properties had equal means between sampling designs in the "North" zone, while significant differences were found for pH, available P and exchangeable Mg in the "Southeast" zone, and for clay, OC and available P in the "Southwest", though their means were similar between designs, except P. Mean soil exchangeable K varies between designs, but their high within-group variances hinder statistically significant differences.

Globally, correlations among soil properties and proximal sensor variables from the regular grid samples were honored by the optimized samples for all paired variables except pH x aEC, and P x eU. The same behavior was observed within the

management zones, where most soil property-proximal sensor correlations were respected by the optimized samples. Significant differences in correlations between sampling designs were observed for: pH x aEC, Ca x aEC, Mg x aEC, and Mg x aMS in the "North"; Ca x eTh in the "Southeast"; and Mg x eTh in the "Southwest".

Overall, the optimized samples captured the global spatial trends of most properties and honored their mean values both globally and locally within management zones, as mean property values were very close between designs (except for available P and exchangeable K) despite significant differences in some cases (Table 1). They also captured the correlations among soil properties and proximal sensor variables both globally and within zones. This represents a reduction of 75% (from 72 to 18 sites) in soil sampling and analysis costs, while keeping most soil information.

## Conclusions

Soil sampling and analytical costs can be reduced considerably by reducing the sample size while keeping most soil information across the field and within management zones. For that, a combination of proximal sensor surveys that catch soil variations efficiently across the field and a sample optimization algorithm like spatial simulated annealing can be used with positive results, as shown in this paper.

In principle, management decisions based on soil data obtained at the optimized sampling sites would be mostly correct. Along these lines, whether investing in more samples, say one sample per hectare, provides more accurate management decisions that are worth the extra cost is open for debate and further research.

## References

ADAMCHUK, V. I. et al. On-the-go soil sensors for precision agriculture. **Computers and Electronics in Agriculture**, v. 44, p. 71-91, 2004.

DOMENECH, M. B. et al. Sampling scheme optimization to map soil depth to petrocalcic horizon at field scale. Geoderma, v. 290, p. 75-82, 2017.

SAMUEL-ROSA, A. **spsann: Optimization of sample configurations using spatial simulated annealing**. R package version 2.2.0. 2019

TEIXEIRA, P. C. et al. **Manual de métodos de análise de solo**. 3. ed. rev e ampl. Brasília: Embrapa, 2017.

VASQUES, G. M. et al. Management zones from proximal soil sensors capture within-field soil property and terrain variations. In: ENCONTRO BRASILEIRO DE PEDOMETRIA, 2, 2021, Rio de Janeiro. **Anais…** Rio de Janeiro: Embrapa Solos, 2021.

Table 1. Soil property means from regular grid and optimized samples. Equal letters indicate equal means between sampling designs globally, and within management zones, respectively, according to Welch's tests at p=0.05.

| Property | N | Mean | N | Mean | N | Mean | N | Mean | N | Mean | N | Mean | N | Mean | N | Mean |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Global | | | | North | | | | Southeast | | | | Southwest | | | |
| | Grid | | Optimized | | Grid | | Optimized | | Grid | | Optimized | | Grid | | Optimized | |
| Clay (g kg$^{-1}$) | 72 | 413[a] | 18 | 424[a] | 33 | 392[a] | 6 | 367[a] | 27 | 430[a] | 6 | 463[a] | 12 | 433[b] | 6 | 443[a] |
| OC (g kg$^{-1}$) | 72 | 15[a] | 18 | 14[b] | 33 | 14[a] | 6 | 13[a] | 27 | 16[a] | 6 | 15[a] | 12 | 15[a] | 6 | 13[b] |
| pH | 72 | 6.6[a] | 18 | 6.5[a] | 33 | 6.6[a] | 6 | 6.6[a] | 27 | 6.6[b] | 6 | 6.7[a] | 12 | 6.4[a] | 6 | 6.2[a] |
| P (mg dm$^{-3}$) | 72 | 143[a] | 18 | 99[b] | 33 | 141[a] | 6 | 137[a] | 27 | 151[a] | 6 | 79[b] | 12 | 127[a] | 6 | 81[b] |
| Ca (cmol$_c$ dm$^{-3}$) | 72 | 6.3[a] | 18 | 5.9[a] | 33 | 6.0[a] | 6 | 5.7[a] | 27 | 6.7[a] | 6 | 6.2[a] | 12 | 6.4[a] | 6 | 5.9[a] |
| Mg (cmol$_c$ dm$^{-3}$) | 72 | 1.9[a] | 18 | 2.0[a] | 33 | 1.8[a] | 6 | 1.9[a] | 27 | 2.1[b] | 6 | 2.2[a] | 12 | 1.9[a] | 6 | 1.9[a] |
| K (cmol$_c$ dm$^{-3}$) | 72 | 458[a] | 18 | 501[a] | 33 | 451[a] | 6 | 173[a] | 27 | 583[a] | 6 | 1110[a] | 12 | 197[a] | 6 | 220[a] |

N, number of observations; Stdev, standard deviation.
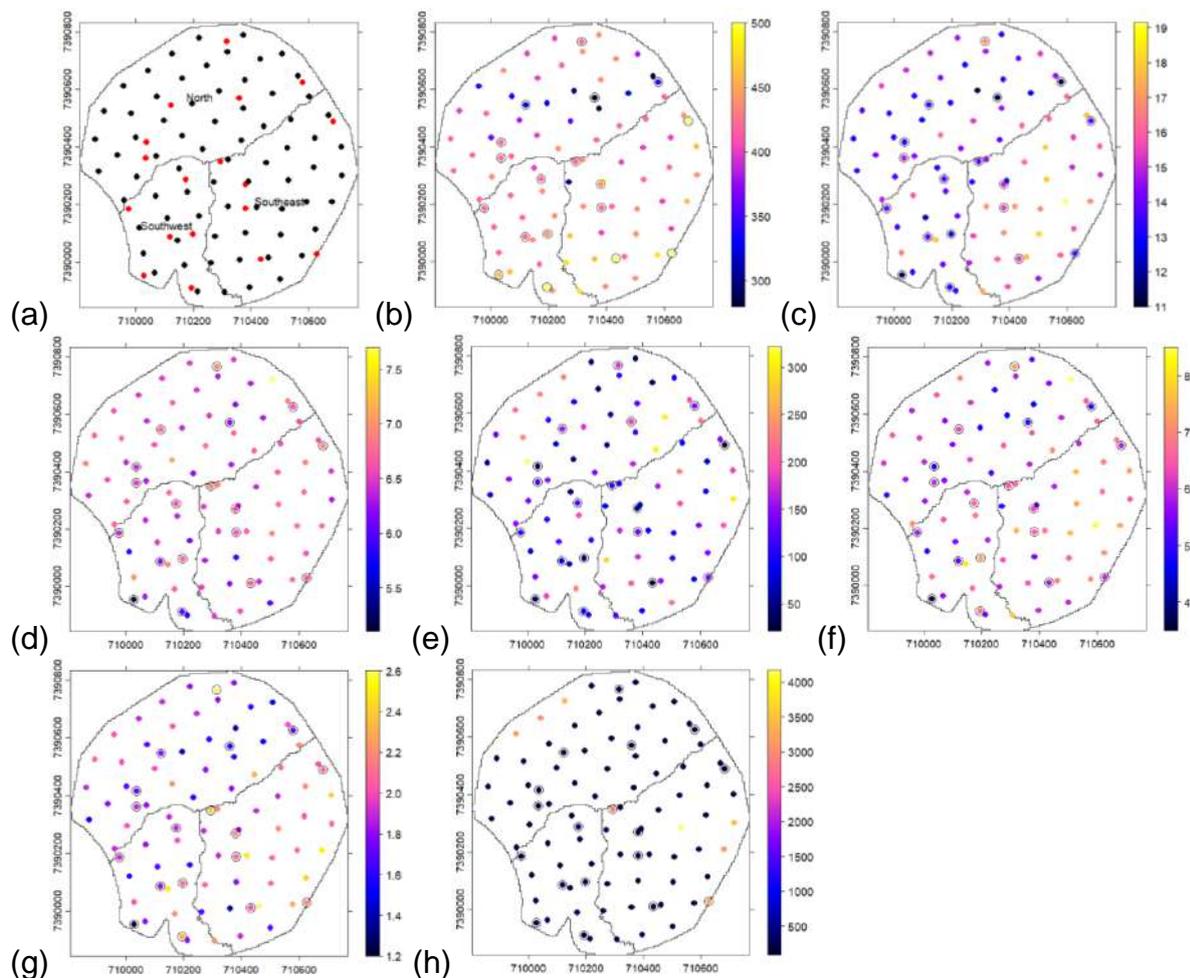
Figure 1. (a) Soil management zones, regular grid samples (black dots) and optimized samples (red dots); (b-h) Soil clay (g kg$^{-1}$), organic C (g kg$^{-1}$), pH, available P (mg dm$^{-3}$), and exchangeable Ca, Mg and K (cmol$_c$ dm$^{-3}$), respectively. Optimized samples are circled in the soil property maps. Coordinates are in UTM zone 22S.

# Identificação espectral da mineralogia e granulometria de Planossolo e Luvissolo no Semiárido Baiano.

FERREIRA, Higo Batista[1]; SOUZA, Deorgia Tayane Mendes de[2]
[1] Universidade Estadual de Feira de Santana, higoferreira8@gmail.com; [2] Universidade Estadual de Feira de Santana, deorgiasouza.geo@gmail.com

**Sessão temática: 03 - Avanços em Sensoriamento do solo**

**Resumo**
O presente trabalho teve como objetivo analisar o comportamento espectral de duas classes de solos, afim de identificar suas assinaturas espectrais e verificar a relação entre as propriedades mineralógicas e granulométricas. As amostras foram extraídas de um perfil de Planossolo e de um Luvissolo, oriundas da região semiárida da Bahia. Os procedimentos metodológicos aplicados foram, análise de granulometria e espectroscopia de reflectância VIS-NIR-SWIR, em seguida os dados obtidos foram analisados e interpretados a partir dos valores granulométricos e comportamento espectral das amostras de solo. Os resultados encontrados na granulometria sinalizam as mudanças texturais abruptas entre horizontes A e B de ambos perfis. Os resultados obtidos por meio da espectroscopia de reflectância proporcionou a verificação da mudança textural por meio do padrão de absorção no horizonte B textural, com destaque para a concentração de argilominerais 2:1. Desse modo, é possível afirmar que associação entre mineralogia e granulometria contribuíram para a diferenciação entre horizontes e perfis, representando uma diversidade pedológica numa mesma região.

**Palavras-chave:** Espectrorradiometria; Sensoriamento Remoto; Solos.

## Introdução

Os solos oriundos do clima semiárido, apresentam diversidade de processos pedogenéticos, isso se dá aos fatores de formação relevo e material de origem que potencializam os diversos processos de formação dos solos, consequentemente, promove uma diversidade pedológica. São solos com peculiaridades, onde se destaca a predominância do intemperismo físico em detrimento do químico, a conservação dos materiais de origem e também a presença de minerais primários intemperizáveis e argilominerais 2:1 (SOUZA, 2020). Assim, o estudo e a investigação da mineralogia e da granulometria das diversas frações dos solos, atrelado a sua caracterização morfológica, física e química é de extrema relevância para compreensão dos atributos pedológicos para suas aplicações em melhoramento na escala cartográfica dos mapas, como também suporte para manejo adequado dos solos (OLIVEIRA, 2007).

A grande demanda acerca de informações pedológicas para maiores intervenções, seja na exploração agrícola como na sua compreensão com a relação dos sistemas naturais (BAPTISTA et al., 2019), necessita cada vez mais de rapidez nos resultados, como também baixos custos. A espectrorradiometria surge como uma alternativa, sendo uma técnica rápida e de baixo custo, e que contribui na diminuição

do tempo e na quantidade de resíduos laboratoriais não impactando negativamente o meio ambiente, além de ser eficiente na predição de diversos atributos do solo e análises envolvendo grandes quantidades de dados (BELLINASO, 2009).

Deste modo, a presente pesquisa tem como objetivo analisar o comportamento espectral de duas classes de solos, afim de identificar suas assinaturas espectrais e verificar a relação entre as propriedades granulométricas e mineralógicas.

## Metodologia

As amostras analisadas correspondem aos horizontes dos perfis de Planossolo e Luvissolo que foram coletadas nos municípios de Araci e Juazeiro (Figura 1). Ambos estão localizados no clima semiárido baiano. A escolha de áreas distantes referente a escala é devido a necessidade de observar de que modo outros fatores como, material de origem interfere na resposta espectral como também influencia na granulometria.
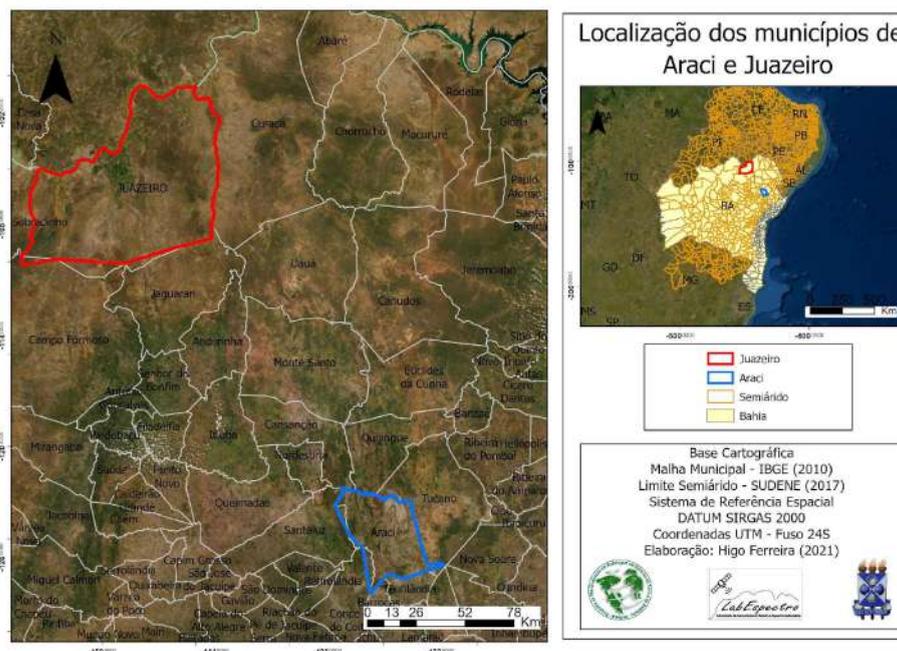


Figura 1 – Mapa de localização

**Os procedimentos metodológicos seguiram a sequência:**

**a) Seleção das amostras, tratamento, medição e análise granulométrica:** Para a obtenção das medidas espectrais, as amostras foram tratadas de acordo com o protocolo de Bendor et. al, (2015). Em seguida, foi feita a coleta das medidas espectrais, onde se utilizou o espectrorradiômetro ASD FieldSpec® 4 Hi-Res, que abrange um intervalo de 350 a 2500 nm.

**b) Análise granulométrica:** a execução para análise granulométrica foi por meio da Embrapa (1997) - Análise Granulométrica (Dispersão total).

**c) Processamento das curvas, análise e interpretação dos resultados:** O Software utilizado foi o ViewSpec Pro 6.0 ASDInc, com o objetivo de reduzir os ruídos, seguido da análise e interpretação que se deram a partir das curvas espectrais, que foram caracterizadas com o auxílio do Software ENVI 5.3 (Harris Geospatial Solutions).

**Resultados e Discussão**

Com os dados obtidos do fracionamento das partículas, foi possível analisar a mudança textural abrupta entre horizontes em ambos perfis (Tabela 1) em seus horizontes diagnósticos – Bt, onde há uma maior concentração de partículas em tamanho argila. Esta característica também influencia em suas classes texturais, onde no Perfil 1, Planossolo háplico há uma mudança significativa de g/kg de argila entre os horizontes A e Bt. O mesmo acontece no perfil 2, Luvissolo crômico, que embora indique uma classe textural franca, apresenta o valor de 2,08 de gradiente, atributo fundamental na definição de sua ordem (SANTOS et al. 2018).

Tabela 1 – dados granulométricos

| Horizonte | Profundidade (cm) | Areia Total (g/kg) | Silte (g/Kg) | Argila (g/kg) | Gradiente textural - B/A | Classe textural |
|---|---|---|---|---|---|---|
| Perfil 1 – Planossolo háplico eutrófico típico | | | | | | |
| A | 0 –12 | 822 | 28 | 150 | - | Are – fr |
| AE | 12 – 33 | 792 | 32 | 176 | - | Fr – are |
| E | 33 – 51 | 793 | 32 | 175 | - | Fr – are |
| Bt | 51 – 67 | 576 | 22 | 402 | - | Arg – are |
| Cr | 67 – 79+ | - | - | - | - | - |
| Perfil 2 – Luvissolo crômico órtico vertissólico hipocarbonático | | | | | | |
| A | 0 – 7 | 495,9 | 382,5 | 121,6 | - | Fr |
| Bt | 7 – 70 | 459,0 | 290,5 | 250,6 | 2.08 | Fr |
| C | 70 – 106 | 276,1 | 518,4 | 205,6 | | Fr. siltosa |

Are – Arenosa; Fr – Franca; Arg – Argilosa.

As assinaturas espectrais geradas por reflectância com contínuo removido (Figura 2) mostraram padrões de absorções, sinalizando a presença de minerais do tipo 2:1 montmorillonite, que agrega aos solos concentrações de argila, principalmente no horizonte B de ambos os perfis. E além disso, é possível diagnosticar no perfil 2 a concentração de calcita, o que classifica o mesmo em hipocarbonático (altos teores de $CaCO_3$). É importante destacar também que, os altos valores de areia são resultados da presença de minerais primários nesta fração.
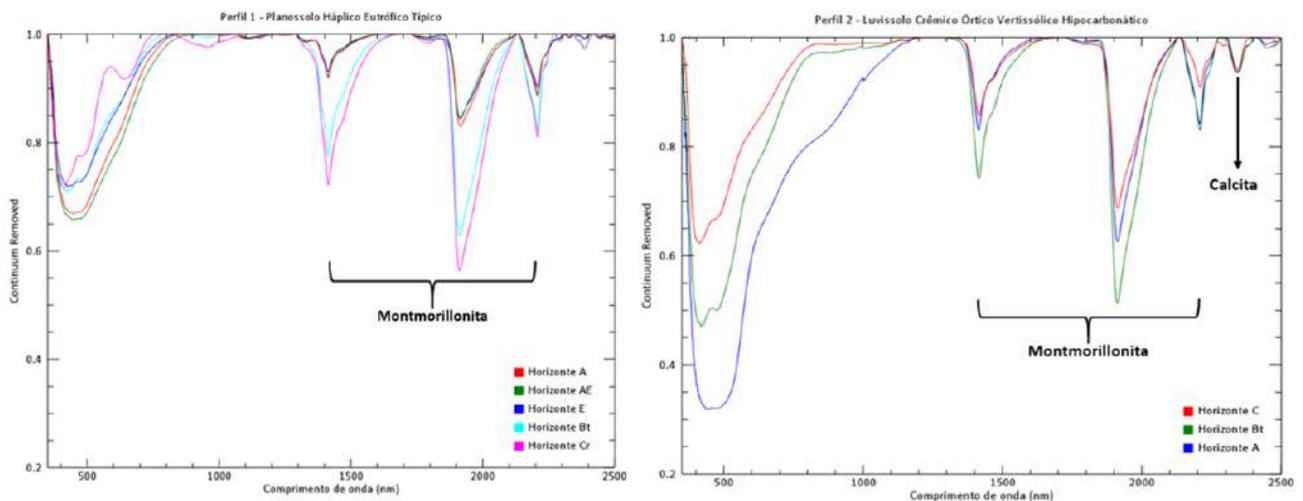
Figura 2 – Assinaturas espectrais dos perfis

## Conclusões

Verificou-se, a partir da identificação espectral e a associação entre mineralogia e granulometria, que a diversidade pedológica do semiárido deve ser cada vez mais investigada e estudada. A presença de minerais primários facilmente intemperizáveis na fração areia, atribui percentagem alta na granulometria, entretanto, por meio da espectroscopia é possível verificar a concentração de argilominerais e a natureza dos mesmos.

## Referências

BAPTISTA, G. M. de M.; MADEIRA NETTO, J. da S.; SOUZA, D. T. M. de. Reflectância dos Solos. *In*: MENESES, P. R.; ALMEIDA, T. de; BAPTISTA, G. M. de M**. Reflectância dos materiais terrestres: análises e interpretações**. Oficina de Textos: São Paulo, 2019, p. 163-187.

BEN-DOR, E.; ONG, C.; LAU, I. C. Reflectance measurements of soils in the laboratory: Standards and protocols. **Geoderma**. Perth, Austrália, 2015, p. 112-124.

BELLINASO, H. **Biblioteca espectral de solos e sua aplicação na quantificação de atributos e classificação**. Piracicaba, 2009, 265 f Dissertação (Mestrado em Agronomia) - Universidade de São Paulo. Piracicaba, 2009.

EMBRAPA. Centro Nacional de Pesquisa de Solos (Rio de Janeiro, RJ). **Manual de Métodos de Análise de Solo**, Centro Nacional de Pesquisa de Solos. 2 Ed. Rio de Janeiro, 1997, p. 212.

OLIVEIRA, L. B**. Mineralogia, micromorfologia, gênese e classificação de Luvissolos e Planossolos desenvolvidos de Rochas no semiárido do Nordeste Brasileiro.** Tese (Doutorado em Solos e Nutrição de Plantas) - Universidade Federal de Viçosa, Viçosa, 2007. 189 pp.

SANTOS, H. G.et al. **Sistema Brasileiro de Classificação de Solos**. 5. Ed. Embrapa: Brasília, 2018. 353 p.

SOUZA, D. T. M. **Espectroscopia de reflectância aplicada a estudos pedogenéticos em ambientes semiáridos**. 153 f Tese (Doutorado em Geociências Aplicadas e Geodinâmica) - Universidade de Brasília, Brasília, 2020.

# Soil spectral behavior in a tropical toposequence on metasedimentary rocks

NOVAIS, Jean J. [1]

[1] Faculty of Agronomy and Veterinary Medicine – University of Brasília, Campus Darcy Ribeiro, Brasília, Federal District Brazil, jjnagron@gmail.com

**Thematic Session: ADVANCES IN SOIL SENSING**

## Abstract

Investigations about the pedomorphogeological dynamics can answer to several characteristics of soil classes. Such conditions alter the soil attributes, in turn, must they influence the spectral response. In this sense, this paper aimed to relate soil attributes with their spectral behavior. For this purpose, five soil profiles in a tropical toposequence on metasedimentary rocks of the Paranoá group were classified and characterized according to physic-chemical attributes. Afterward, these data were related to the spectral curve for each soil class. The morphological interpretation of reflectance spectra supported the qualitative assessments. We observed the main features which characterize the soil class, namely, goethite (0.48 and 0.95 μm) and hematite (0.53 and 0.85 μm), kaolinite (1.4 and 2,205 μm), gibbsite (2.265 μm), 2:1 clay mineral (1.405 and 1.9 μm). Iron oxide features are highlighted in Ferralsols and organic matter obliterating Haplic Gleysol and Hemic Histosol curves.

## Introduction

Factors such as climate, organisms, relief, rocky matrix, age, and soil arrangement in space are responsible for pedogenesis. These factors provide characteristics and attributes that allow the distinction between soils, for instance: texture, natural fertility, structure, water-holding capacity (LACERDA et al., 2016).

Such characteristics can be determined by morphological in the field, physical and chemical laboratory analysis, including remote sensing techniques such as reflectance spectroscopy (from 0.35 μm to 2.5 μm) (DEMATTÊ et al., 2014). The spectral behavior reveals several soil compounds and characteristics, such as mineralogy, texture, moisture and organic matter (NOVAIS et al. 2021).

Thereby the objective was to classify the soils and evaluate them regarding the influence of pedological attributes on the spectral behavior of soils in a tropical toposequence formed on metasedimentary rocks, Paranoá Group.

## Methodology

This work was developed in a microwatershed in the Água Limpa experimental Farm (FAL), belonging to the University of Brasília (UnB), in Federal District, Brazil and it was performed between 2019 and 2020. It was performed at the Laboratory of Geoprocessing and Pedomorphogeology – GeoPed, which is linked to the Faculty of Agronomy and Veterinary Medicine – FAV of UnB.

Fieldworks followed Soil Survey Staff (2014) recommendations in a toposequence formed by five soil profiles (Figure 1). Thus, the samples were collected at depth 0-20 cm, for the surface, and 10-100 cm for the subsurface layer. These samples were air-dried, mashed, and sieved (particles lower than 2 mm). Then, they were submitted to

physical (texture) and chemical (assortment complex) analyses in the laboratory, according to Teixeira et al. (2017). With results, the profiles were classified according to World Reference Bases for soil resources (IUSS WORKING GROUP WRB, 2014).
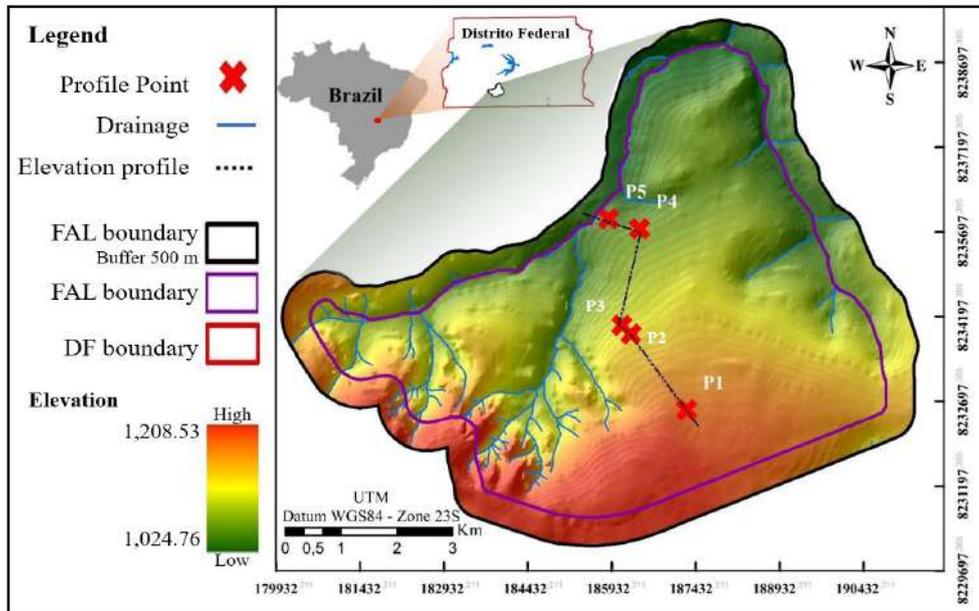


**Figure 1.** Localization map of the study area showing a) digital elevation model, b) geological groups, c) geomorphological surface, and d) slope. All with the sampling spots studied. MNPpa: slates, MNPpr3: clayey metarrythmits; GS-1 High Plateaus and GS-2 Region of Intermediate Dissection.

Surface soil samples also were analyzed by reflectance spectroscopy from visible and shortwave infrared (SWIR) range, using the Fieldspec Pro hyperspectral sensor (MALVERN PANALYTICAL Inc, 2021), which were morphologically interpreted according to main absorption features, identifying some attributes that characterize the soils, such as mineralogy, organic matter, and texture (DEMATTÊ et al., 2014). Finally, the soils were interrelated according to position in the landscape, tracing a profile of soil evolution based on the silt/clay ratio, which the pedomorphogeological relationships conditions were considered as factors that alter soil spectral behavior.

**Results and discussion**

Typically, the soil presented texture ranging from very clayey to loam-sandy. According to IUSS Working Group WRB (2014), base saturation below 50% characterize Dystric soils. Tropical soils tend to present low natural fertility, acidic character, and high iron oxides and aluminum content (NOVAIS et al., 2021).
The soil classes were determined as P1 – clayic, dystric, Rhodic Ferralsol; P2 – clayic, dystric, Rhodic Ferralsol; P3 – clayic, petroplinthic, dystric, Haplic Ferralsol; P4 – clayic, dystric Haplic Gleysol; P5 – loam-sandy hemic Haplic Histosol.
As long as the altitude decreases, the weathering action also decreases in a toposequence because of the insufficient local drainage, which favors the formation

of hydromorphic soils (LACERDA et al., 2016). Profiles in the upper of toposequence exhibited a high degree of evolution than the bottom.

It was possible to note the occurrence of specific processes of soil formation such as Ferratilization in P1, P2, and P3 and plinthization and oxidation in the P4 and P5 profiles, caused by reductions in the internal water flow, such changes gradually influence the general behavior of the spectra. However, more detailed studies in the domain of time and space are necessary to explain the pedogenetic relationship that confers the distinctive properties of soils within spectral reflectance (Figure 2).
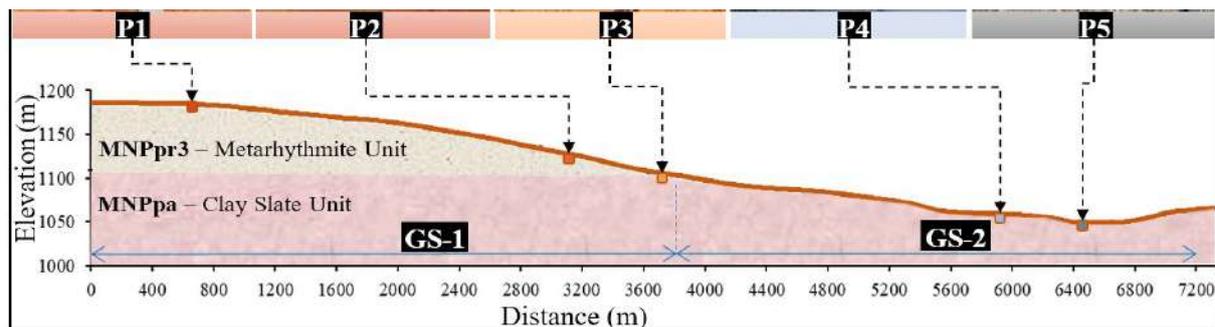


**Figure 2.** Topographic profile of toposequence and relative location of soil profiles (P) on geomorphological surfaces (SG).

The Morphological interpretation of reflectance spectra showed the typical features of primary and secondary minerals, as well as oxides influence and obliterating action of organic matter throughout the spectra. Different features were observed in the toposequence profiles. Basically, they presented features of iron oxides such as goethite (0.48 and 0.95 μm) and hematite (0.53 and 0.85 μm), kaolinite (1.4 and 2,205 μm), gibbsite (2.265 μm), 2:1 clay mineral, and water adsorbed on the particle surface (1.405 and 1.9 μm), as described by Demattê et al. (2014) and Novais et al. (2021). These features are highlighted in Figure 3, at the end of this paper.

**Conclusions**

Features of primary and secondary minerals, iron oxides, and organic matter were observed in the spectral curves in the toposequence. They were related to chemical-physical analyses, reinforcing their utility for soil classification. A quantitative approach can support the potential of spectral information for soil mapping.

**References**

DEMATTÊ, J. A. M. et al. Morphological Interpretation of Reflectance Spectrum (MIRS) using libraries looking towards soil classification. **Scientia Agricola**, v. 71, n. 6, p. 509-520, 2014.

IUSS WORKING GROUP WRB, **World reference base for soil resources 2014** International soil classification system, No. 106. FAO, 2015. 206 p.

LACERDA, M. P. C. et al. Tropical texture determination by proximal sensing using a regional spectral library and its relationship with soil classification. **Remote Sensing**. v. 8, n. 701, 2016.

MALVERN PANALYTICAL Inc. **ASD Fieldspec® 4 User manual**. 6th ed. Falls Church, VA, USA, 2019. 102 p.

NOVAIS, J. J. et al. Digital soil mapping using multispectral modeling with Landsat time series cloud computing based. **Remote Sensing**, v. 13, p. 1181, 2021.

SOIL SURVEY STAFF. **Soil Survey Field and Laboratory Methods Manual**. United States Department of Agriculture, NRCS. 2014. 1031 p.
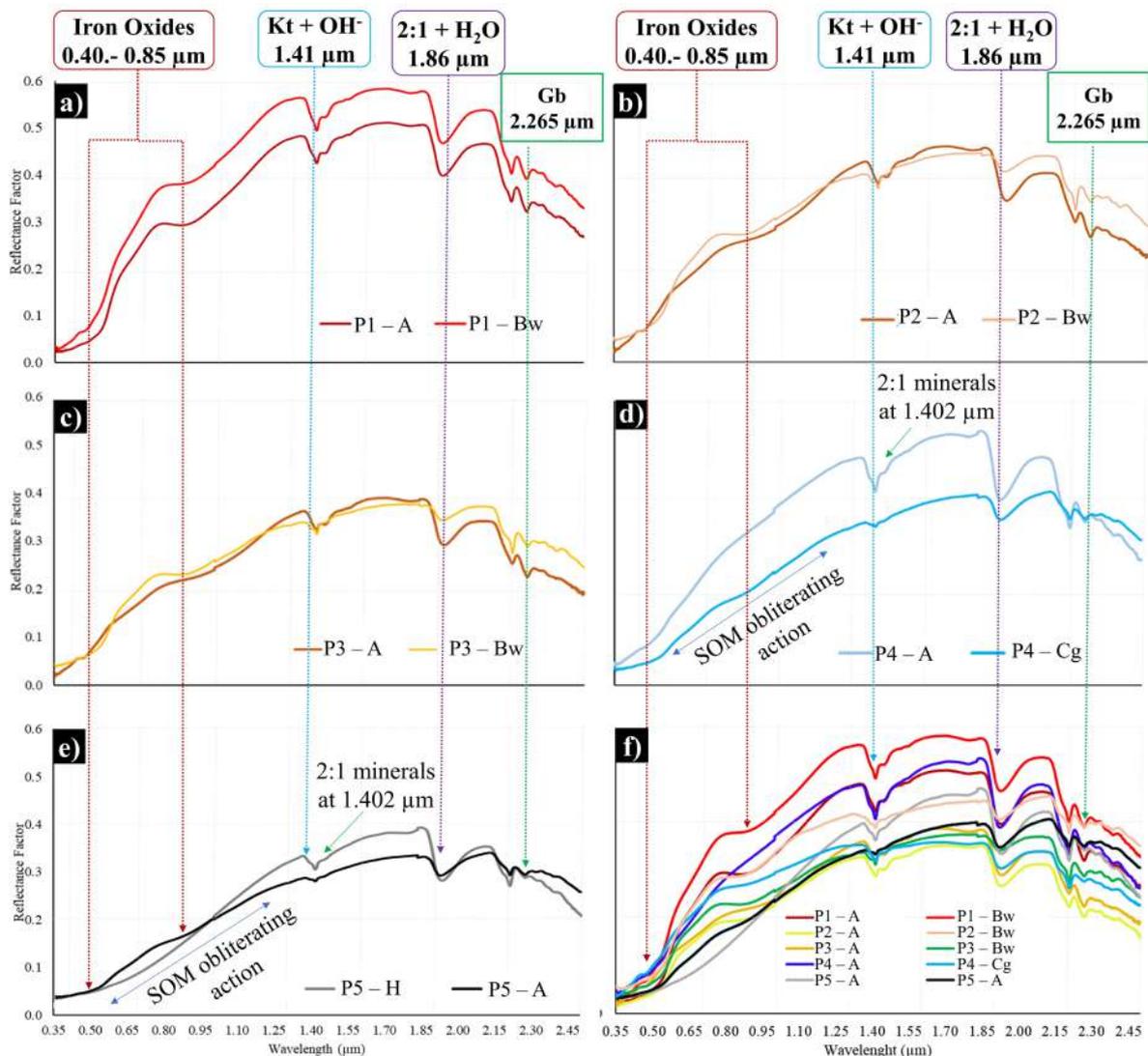
**Figure 4.** Morphological interpretation of soil horizons spectra with features highlighted. Where: a) and b) Rhodic Ferralsol; c) Haplic Ferralsol, d) Haplic Gleysol; e) Hemic Histosol and f) Spectral library of studied soils; (Kt) kaolinite, (Gb) gibbsite, (Ht) hematite, (Gt) goethite, and (SOM) Soil Organic Matter.

# DIGITAL MAPPING OF SOIL CLASSES COMPARING MACHINE LEARNING ALGORITHMS IN INTERIOR TABLELANDS

SAMPAIO, Felipe[1]; BARRETO, Ícaro[2]; POELKING, Everton[3]; SOUZA, Luciano da Silva[4]; COSTA, Oldair[5]

[1] UFRB, ftsampaio.ufrb@gmail.com; [2] UFRB, icaro@ufrb.edu.br; [3] UFRB, everton@ufrb.edu.br; [4] UFRB, lsouza@ufrb.edu.br; [5] UFRB, oldair@ufrb.edu.br.

**Thematic Session: 04 Pedometric protocols for systematic soil surveys**

## Abstract

Land use management requires knowledge about the properties and attributes that characterize. The objective of this work was to test the use of machine learning in soil prediction, aiming to compare the respective performances. The selected area is located in the region of Tabuleiros Interioranos of Recôncavo da Bahia, Brasil. The adopted method used for training the observation points and legacy data models, the variables were generated through maps of source material, land use, geomorphometric data and satellite images. In the R software, prediction calculations were performed. The tested algorithms were: Random Forest (RF), Decision Trees (C5), K-Nearest Neighbors (KNN), Gradient Boosting Machines (GBM) and Support-vector machine (SVM). The accuracy of digital mapping was measured using the index Kappa and general accuracy. The C5 model showed the highest kappa and accuracy index.

Keywords: soil class prediction; algorithm test; pedometrics; land use planning.

## Introduction

The earth's natural resources are not inexhaustible, and the growing demand imposed exceeds the natural capacity to replenish those resources. Land use management requires knowledge and information about the properties and attributes that characterize it (CARVALHO JUNIOR et al., 2014).

The MDS is a current, quantitative technique, which through mathematical models integrate environmental covariates that influence the soil formation factors to perform the classification (BAGATINI; GIASSON; TESKE, 2015). According to ten Caten et al.,2012 one of the main classifiers used in digital soil mapping in Brazil until 2011 are decision trees.

The objective of this paper was to investigate the performance of machine learning algorithms on soil mapping in interior tablelands, area that encompasses the campus of the Universidade Federal do Recôncavo da Bahia (UFRB) in the city of Cruz das Almas – Bahia.

## Methodology

The area is located in UFRB, with an area of 1.367 hectares. There are three distinct geological units: Pre-Cambrian metamorphic rocks from the Granulitic Complex; tertiary sediment (detritic-latetritic deposits) and quaternary alluvial sediments. The

geomorphology is mainly presented as a board, in the dissection phase, classified in the geomorphological unit of the Inland Tablelands by RADAMBRASIL (1981), belonging to the morpho-structural domain of the Platores Inumados. The classification by Köppen (Köppen classification system), indicates that the climate is "Af" typy, the average annual rainfall is 1.200 mm and temperature of 24.2 °C.

Using legacy data from previous traditional surveys, 43 complete soil profiles were described up to the fourth taxonomic level identified according to the Brazilian Soil Classification System - SiBCS (Santos et al., 2018). Additionally, 100 extra samples were collected (mini-pits and auger samples), characterized, and classified by a senior pedologist, for a total of 143 sites of soil sampling. The soils were correlated with the geomorphic surfaces in which they occur prospecting (along topo sequences) by this method, it is possible to establish correlations between classes of soil, texture, drainage, depth, slope, length and shape of slopes and position and exposure of soils in relation to slopes according to IBGE (2015).

Thirty existing soil classes were identified in the study area, these classes were divided into 10 mapping units, namely: 1 – LATOSSOLO AMARELO Distrocoeso típico (LAdx1), 2 - CAMBISSOLO HÁPLICO Tb Distrófico petroplíntico + ARGISSOLO AMARELO Distrocoeso abrúptico (PAdx+C1), 3 - CAMBISSOLO HÁPLICO Ta Eutrófico vertissólico (C2), 4 – CHERNOSSOLO EBÂNICO Órtico vertissólico (ME), 5 - VERTISSOLO HÁPLICO Órtico gleissólico (V), 6 - PLANOSSOLO HÁPLICO Eutrófico solódico (P), 7 - ARGISSOLO VERMELHO-AMARELO Eutrófico típico (PVAe), 8 - ARGISSOLO ACINZENTADO Eutróficos típico + PLANOSSOLOS NÁTRICOS Órticos gleissólicos + GLEISSOLO HÁPLICO Tb Eutrófico típico (PSG), 9 - GLEISSOLO MELÂNICO Ta Eutróficos vertissólico (G1), 10 - LATOSSOLO AMARELO Distrocoeso argilsólico (LAdx2).

Using the R 4.1.1  (R Development Core Team, 2016)  with the RSaga package (Brenning et al., 2018) to access the SAGA GIS 6.2 software tools, variables were generated (flow accumulation, aspect, capitation area, watercourse, curvature, profile curvature, planar curvature, slope, flow direction, shading, Topographic wetness index, stream power index) through the digital elevation model - DEM generated from Advanced Land Observing Satellite (ALOS), 12.5 m of spatial resolution. Spectral imagery of the CBERS 04A satellite, acquired on November 23, 2019: band 1 blue (0.45 – 0.52 μm), band 2 green (0.52 – 0.59 μm), band 3 red (0.63 – 0.69 μm), band 4 infra-red (0.77 – 0.89 μm), were obtained through the Instituto Nacional de Pesquisas Espaciais (INPE), from the CERBS 4A satellite – WPM sensor with a resolution of 8 meters. The Land Use Map was produced with ESRI's ArcGis Pro software using CERBS 4A images. The material map was adapted from the map published by the SRTM de Geologia do Brasil at a scale of 1:250,000.

For this work the training used 10-fold cross-validation repeated 3 times. This method is suggested as the fairest method in evaluation of performance to comparing machine learning algorithms. Were tested the control parameters that resulted in better prediction performance of the mathematical models Random Forest (RF), Decision Trees (C5), K-Nearest Neighbors (KNN), Gradient Boosting Machines

(GBM) and Support-vector machine (SVM) with help of Caret package (Kuhn, 2013). To evaluate the algorithms, was used the confusion matrix (Congalton and Green, 2009) to derive the Kappa indexes (Landis e Koch, 1977) and the overall accuracy. Statistical results such as mean, median and variance were analyzed to compare the predictors.

**Results and discussion**

The results show that classifiers C5.0 and Random Forest have great potential for predicting soil classes for the data set used. The machine learning algorithms exhibited Kappa index ranging from 0.16 to 0.39 and means of overall accuracy ranging from 0.25 to 0.45, this variance shows that not all models were efficient for class prediction in this study (Figure 1).
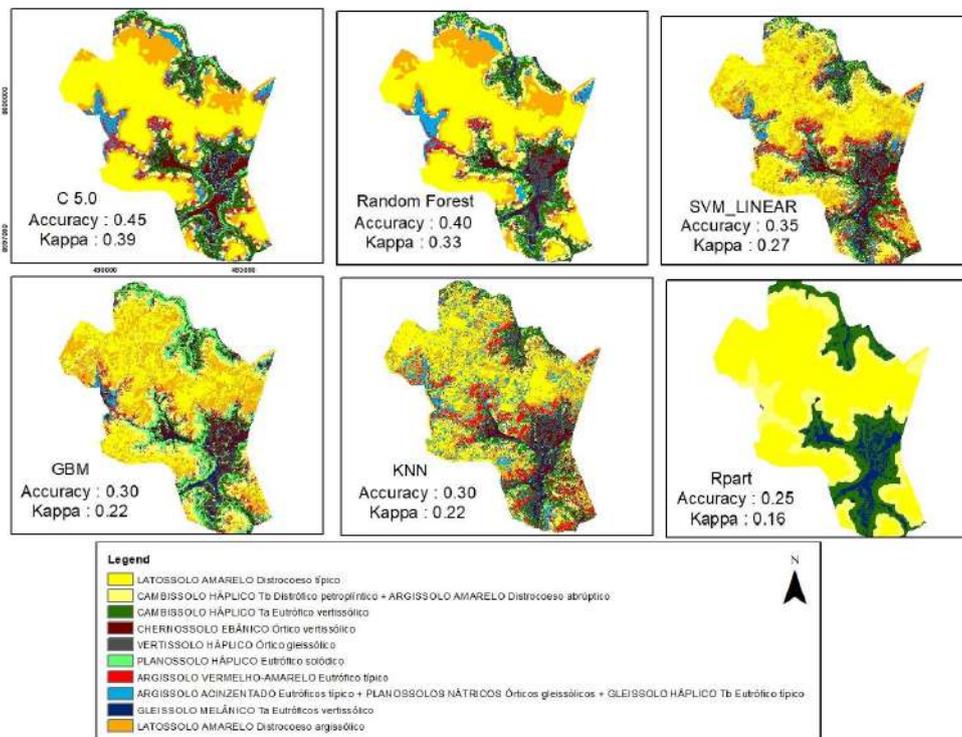


**Figure 1: Prediction maps of Soil Classes of a tropical Interior Tablelands for different algorithms.**

Observing the generated maps, it is possible to see that, except for the algorithms C5.0, RF and Rpart, there is difficulty in separating the mapping units (UM), analyzing the UMs LAdx1 and LAdx2 it is possible to recognize areas of LAdx2 located in the center of the tablelands, which is not a field reality. The Rpart algorithm was not able to identify all mapping units, showing high BIAS with a tendency to underfitting compared to the other tested algorithms. RF and C5.0 are decision family models, both have a low bias, fit well to the dataset and tend to overfit, in the present study, these characteristics enabled a good separation of the mapping units, satisfactorily approaching the field reality.

## Conclusions

Comparing the results found, was observed that the C5.0 classifier obtain the best performance in digital soil mapping classes of a tropical interior tablelands with the data set used. The variance of the results indicates that using more than one algorithm for prediction is a good practice in digital soil mapping. The distribution of cartographic units is consistent with the field reality and with previous surveys. Predominance of Latossolos in the flat tops and upper third of the landscape, followed by Argissolo on the slope, Cambissolo in the lower third and in the lowlands a complex composed of Vertssolo, Chernosoil and Planossolo and finally, along the drainage line, the predominance of Gleissolo.

## References

BAGATINI, T.; GIASSON, E.; TESKE, R. Selection of sampling density based on data from already mapped areas for training decision tree models in digital soil mapping. Brazilian Journal of Soil Science, vol. 39, no. 4, p. 960–967, 1 Jul. 2015.

BRENNING A, BANGS D, BECKER B, SCHRATZ P, POLAKOWSKI F. RSAGA: SAGA Geoprocessing and terrain analysis. R package version 1.2.0; 2018. Available from: https://CRAN.R-project.org/ package=RSAGA

CARVALHO JUNIOR, Waldir de et al. Método do hipercubo latino condicionado para a amostragem de solos na presença de covariáveis ambientais visando o mapeamento digital de solos. Rev. Bras. Ciênc. Solo [online]. 2014.

CONGALTON RG, Green K. Assessing the accuracy of remotely sensed data: principles and practices. 2nd ed. Boca Raton: CRC Press; 2009.
Instituto Brasileiro de Geografia e Estatística - IBGE. Manual técnico de pedologia. 3. ed. Rio de Janeiro: IBGE; 2015.

KUHN M. Predictive modeling with R and the caret Package. Google Scholar; 2013.
Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33:159-74. https://doi.org/10.2307/2529310
R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria; 2016. Available from: http://www.R-project.org/.

RADAMBRASIL – Levantamento de recursos naturais. Vol. 24. Folha SD.24-Salvador: Geologia, Geomorfologia Pedologia, Vegetação, Uso potencial da Terra. Rio de Janeiro, Ministério das Minas e Energia. Projeto Radam Brasil. 1981.

SANTOS HG, JACOMINE PKT, ANJOS LHC, OLIVEIRA VA, OLIVEIRA JB, COELHO MR, LUMBRERAS JF, CUNHA TJF. Sistema brasileiro de classificação de solos. 3. ed. r ev. ampl. Rio de Janeiro: Embrapa Solos; 2013.

TEN CATEN, A. et al. Digital mapping of soil classes: characteristics of the Brazilian approach. Rural Science, v. 42, no. 11, p. 1989–1997, Nov. 2012.

# Do different Digital Elevation Models influence the performance on modelling soil classes?

SIQUEIRA, Rafael Gomes[1]; SANTOS, Cássio Marques Moquedace[2], Elpídio Inácio FERNANDES-FILHO[3]

[1] Federal University of Viçosa, rafael.g.siqueira@ufv.br; [2] Federal University of Viçosa, cassio.moquedace@ufv.br; [3] Federal University of Viçosa, elpidio@ufv.br

## Thematic Session: (1) - Pedometrics: Innovation in tropics

**Abstract**
Currently, there are many digital elevation models (DEM) with different spatial resolutions that can be used for digital soil mapping. Overall, it is assumed that the higher the DEMs spatial resolution, the better the digital soil maps would be. To test this assumption, we evaluated the suitability of several DEMs to map soil classes for a mountainous area in Brazil, using the machine learning model random forest. In general, the 90 m DEMs allowed more accurate soil predictions, highlighting the COPERNICUS GLO-90. Among the 30 m DEMs, the TOPODATA provided the best results, whereas the ALOS PALSAR of 12.5 m presented accuracy similar to the 30 m DEMs. The results did not show a clear pattern between the DEMs spatial resolution and the uncertainty of the prediction. We show that finer resolution DEMs are not always the best choice for predicting soil classes and that the relationship between the scale of the DEMs and the digital soil maps needs to be analyzed more closely.
Keywords: accuracy; digital soil mapping; machine learning; spatial resolution; uncertainty.

## Introduction

Digital Soil Mapping (DSM) was developed from the evolution of the computational, mathematical and statistical techniques, being an alternative to overcome the limitations of the highly subjective and poorly reproducible traditional soil mapping (GOMES et al., 2019). The DSM principle consists in intersecting soil observation points and layers of secondary environmental data, so that a model of some structure is fitted to describe the relationship between the soil and the environment (MINASNY et al., 2014).

Morphometric predictors extracted from Digital Elevation Models (DEMs) play a key role as secondary data for DSM. First, because they represent one of the main factors controlling the soil distribution on the terrain surface, and second due to the greater availability of DEMs with different spatial resolutions, obtained from spaceborne-sensor data (SENA et al., 2020).

Overall, it is assumed that the higher the DEMs spatial resolution, the better the digital soil maps would be, due to the increasing acquisition of information as the spatial resolution increases. Cavazzi et al. (2013) contest this assumption, stating that fine resolutions DEMs can mislead the soil prediction. Actually, we little know about the suitability of the distinct DEMs for digital soil maps with varied scales and degree of detail. In this work, we evaluated the suitability of 11 freely available DEMs to map soil classes at 2nd categoric level for a mountainous area in the Southeastern Brazil.

## Methodology

The study was carried out for the Belo Horizonte Southern Metropolitan Region Environmental Protection Area, located in the Minas Gerais State. The area of 1,627 km$^2$ has an altitude varying between 700 and 2080 m a.s.l.

We used 11 DEMs with different spatial resolutions (ALOS PALSAR at 12.5 m; ASTER GDEM, ALOS World 3D, COPERNICUS GLO-30, NASADEM, SRTM1 and TOPODATA at 30 m; COPERNICUS GLO-90, MERIT, SRTM3 and TanDEM-X at 90 m), whose predictive capabilities in classifying soils were evaluated using machine learning.

Initially, 42 morphometric variables were derived from each DEM. Then, a principal component analysis was used to select the best set of variables from each DEM to be used in obtaining the soil samples with the Latin Hypercube Sampling procedure (cLHS). The number of cLHS iterations was weighted by the DEMs spatial resolutions. Thus, we collected 2,000 points, using the labels of soil classes from the previous conventional soil map of the area (CPRM, 2005).

For the modeling process we used the random forest model. The model was fitted separately for each DEM following standardized methodology (GOMES et al., 2019). The samples were separated in 75% for training and 10-folds cross-validation and 25% for testing. We applied two sequential methods to select the best variables, one by correlation and the other by importance (recursive feature elimination). The model's performance was evaluated using the Accuracy and Kappa indices obtained from the cross-validation and test data. All the process was performed with 100 runs for each DEM, randomizing the training and test subsets.

The final predicted soil maps obtained from the 100 runs modal values were compared (overall Kappa) with the conventional soil map as external validation. We also evaluated the uncertainty of the prediction through the variety analysis (number of different classes that each pixel presented during the 100 runs).

## Results and discussion

Considering the Accuracy values, the best performance of training was reached with the SRTM3, followed by the COPERNICUS GLO-90, TOPODATA and MERIT. A similar pattern was found for the test, with the SRTM3-based model reaching an average Accuracy of 0.67 (Fig. 1).

As to the Kappa values, the Copernicus GLO-90 ensured the best results for training, followed by the SRTM3 and TOPODATA. With the test sets, the COPERNICUS GLO-90-based model kept the best performance (0.36). In turn, the worst performance was obtained with the 30 m DEMs NASADEM and ALOS World 3D (Fig. 1).

Lastly, considering the overall Kappa, we found that the COPERNICUS GLO-90-based prediction again had the best performance (0.32), followed by the TanDEM-X and SRTM3. The worst prediction happened with the ASTER GDEM and ALOS World 3D (Fig. 2).

Among the 90 m DEMS, the SRTM3 provided the highest certainty in the prediction, with 52% of the total pixels presenting only one soil class (the mode) in the final map. Among the 30 m DEMs, the SRTM1 and ASTER GDEM provided the best certainty,

with 58% and 57% of pixels presenting one class. In turn, the DEMs MERIT and ALOS World 3D provided the highest uncertainty of prediction (Fig. 2).
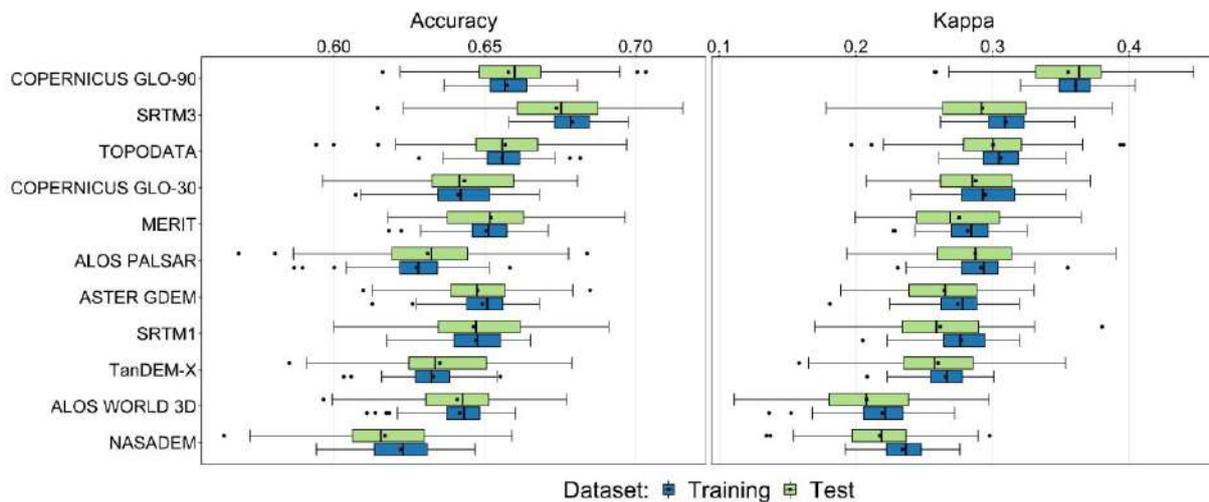


**Fig. 1.** Accuracy and Kappa for the 100 runs of the random forest algorithm with each DEM.

The 12.5 m ALOS PALSAR also presented high uncertainty, which, added to the ordinary prediction accuracy and extremely time-consuming processing (predictions based on it spent ± 3 times more than on 30 m DEMs and ± 30 times than on 90 m DEMs), meant a poor performance of this finer resolution DEM for our soil prediction.
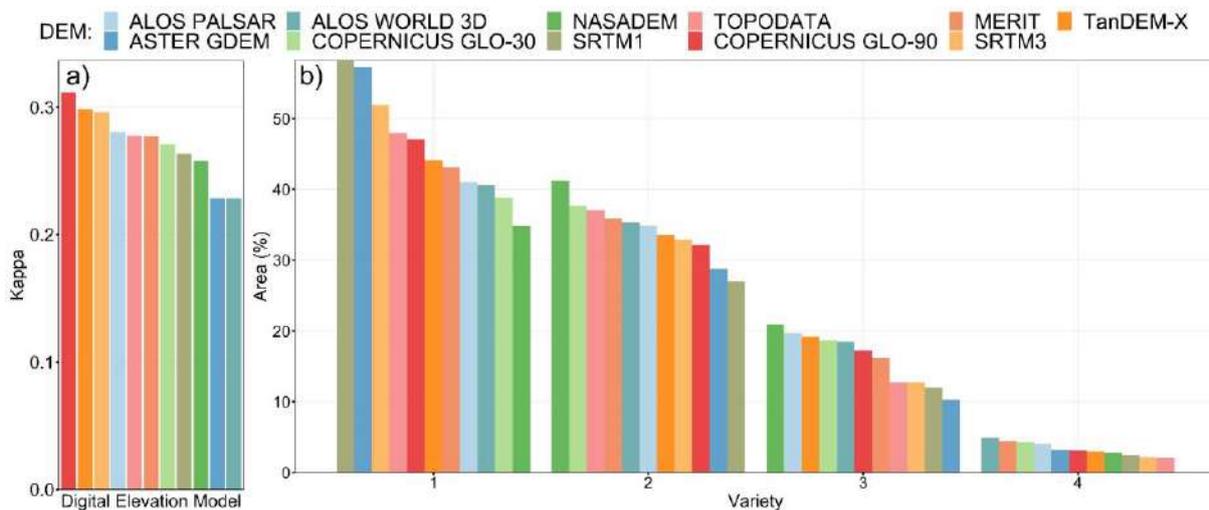


**Fig. 2.** Overall Kappa (a) and variety analysis indicating the uncertainty (b) for prediction with each DEM.

We assume that the best prediction's performance from the DEMs with lower spatial resolution (90 m) can be associated with the generalized character of the soil map predicted in this work (2nd categoric level). At the same time, the worse performance of 30 m and 12.5 DEMs may be associated with the greater amount of information which might carry out to a lot of noise that misleads the prediction.

So, we hypothesize that when predicting soil classes at higher categoric levels, and consequently producing soil maps at higher scales, the results obtained might be different, with finer resolution DEMs being favored by the greater need for more detailed information in the prediction.

Nevertheless, our results show that finer resolution DEMs are not always the best choice in DSM, which agrees with other works (CAVAZZI et al. 2013; SENA et al. 2020). Thus, the relationship between the DEMs' scale and the digital soil maps needs to be analyzed more closely for advances in DSM. Nowadays, this is an approach more needed than ever in the context of the Brazilian Soil Program (PronaSolos), which aims to soil map all the Brazilian territory at detailed scales using the DSM techniques.

**Conclusions**

In general, the 90 m DEMs allowed more accurate soil predictions, highlighting the COPERNICUS GLO-90. Among the 90 m DEMs, the MERIT provided the worst results. Among the 30 m DEMs, the TOPODATA provided the best results, whereas the ALOS PALSAR of 12.5 m presented accuracy similar to the 30 m DEMs.

The results did not show a clear pattern between the spatial resolution of DEMs and the uncertainty of prediction, with 90 m and 30 m DEMs among the DEMs providing the best certainty (SRTM1, SRTM3) and the worst uncertainty of prediction (MERIT, ALOS World 3D).

**References**

CAVAZZI, S. et al. Are fine resolution digital elevation models always the best choice in digital soil mapping? **Geoderma**, v. 195-196, p. 111-121, 2013.

CPRM. **Mapa de Solos da Área de Proteção Ambiental Sul da Região Metropolitana de Belo Horizonte**. Belo Horizonte: CPRM/EMBRAPA/SEMAD, 2005.

GOMES, L.C. et al. Modelling and mapping soil organic carbon stocks in Brazil. **Geoderma**, v. 340, p. 337–350, 2019.

MINASNY, B. et al. Pedometrics. **Reference Module in Earth Systems and Environmental Sciences**, p. 1-10, 2014.

SENA, N.C. et al. Analysis of terrain attributes in different spatial resolutions for digital soil mapping application in southeastern Brazil. **Geoderma Regional**, v. 21, p. 1–12, 2020.

# Applied Morphometry To Digital Mapping In Detailed Scale

VALLADARES, Gustavo S. .[1]

[1] Federal University of Piaui, valladares@ufpi.edu.br

**Thematic Session: PEDOMETRICS GUIDELINES TO SYSTEMATIC SOIL**

## Abstract

The objective of this study was to show a methodology to obtain a preliminary legend of a soil map, it can guide the pedologists in the field works to elaborate soil class maps in detailed scale. The study area is located in a mountainous relief in the Atlantic Plateau, in Jundiaí, São Paulo State.. Is a small area with 59 hectares. The original soil map of the area was made in 1:10,000 scale. Was generated a digital terrain model (DTM) with 4 m of spatial resolution. Based in the DTM are generated derivated maps like altitude, curvature, slope and distance from the streams. Soil orders were identified in 104 sampled points. At these points, morphometric information was used to classify the soils using the random forest method. The digital map produced was interpolated using the R software. The digital map generated had a Kappa accuracy index of 0.79, higher than that of the traditional map, which was 0.75. Overlapping the two maps overall 70% global equivalence, being satisfactory.
Keywords: Digital terrain model; digital soil mapping; machine learning; soil survey.

## Introduction

This work applied a pedological mapping methodology, in an experimental center of "APTA-Frutas" in Jundiaí, SP, using the morfometric parameters and machine learning to elaborate pedologic maps. The aim of the work is show a methodology to obtain a preliminary legend of a soil map (digital map), it can guide the pedologists in the field works to elaborate soil class maps. The objective is compare the preliminary map with other map made by tradicional pedological (traditional map) methodologies.

## Methodology

The study area is located in a mountainous relief in the Atlantic Plateau, in Jundiaí, São Paulo State. The land use and land cover in the area are predominantly apple, vineyard, peach, citrus and natural vegetation. Is a small area with 59 hectares. The original soil map of the area was made in 1:10,000 scale in the detailed level. Was generated a digital terrain model (DTM) with 4 m of spatial resolution based in a topographical map in 1:10,000 scale, the level curves are equidistant of 5m. Was used the TOPOGRID function with ArcInfo software. Based in the DTM are generated derivated maps like altitude, curvature, slope and distance from the streams.

Soil orders were identified in 104 sampled points. At these points, morphometric information was used to classify the soils using the random forest method (BREIMAN, 2001; LIAW & WIENER, 2002). The digital map produced was interpolated using the R software.

The validation of the digital map and verification of accuracy was performed by calculating the Kappa index (K). The digital map was also overlaid with the traditional one, and compared.

## Results and discussion

The altitude map vary from 690 m to 757 m. The slope map vary from 0 to 72 % (Figure 1). The soil identified in the study area was Latosol, Argisol, Cambisol and Gleisol.

Considering digital map, of the 104 sampling points used for training, 90% were correct in the classification, reaching a K of 0.79, which is considered a very good result. Considering traditional map, of the 104 sampling points used for training, 88% were correct in the classification, reaching a K of 0.75, result lower than that obtained for digital mapping. The class that had the worst classification in the traditional map was the Gelisols, as a result of the polygon in the traditional map being very narrow, that is, there was an underestimation of the area occupied by this soil class. Assessing each class individually, the traditional map was more successful for Cambisols and Latosols. And the digital map for Gleisols and Argisols.

The traditional pedological map was crossing with digital soil map (Figure 2), with global equivalence of 70%, and K equal 0.38, in the Cambisol the equivalence in area was 84%, and in the Gleisol the equivalence was 67%, this two soils types the equivalence are high. In the Argisol the equivalence was 46% and in the Latosol the equivalence was only 21% in area. The Latosol and Argisol had confuse with Cambisol.

## Conclusions

This methodology presents in the present work showed adequate to effect the preliminary mapping of some types of soils. The methodology cannot completely replace traditional mapping and is very useful in guiding sampling and preliminary mapping.

## Acknowledgements

## References

BREIMAN, L. Random forests. **Machine Learning**, v.45, n.1, p.5–32, 2001.

LIAW, A.; WIENER, M. Classification and regression by randomForest. **R News,** v.2, n.3, p.18–22, 2002.
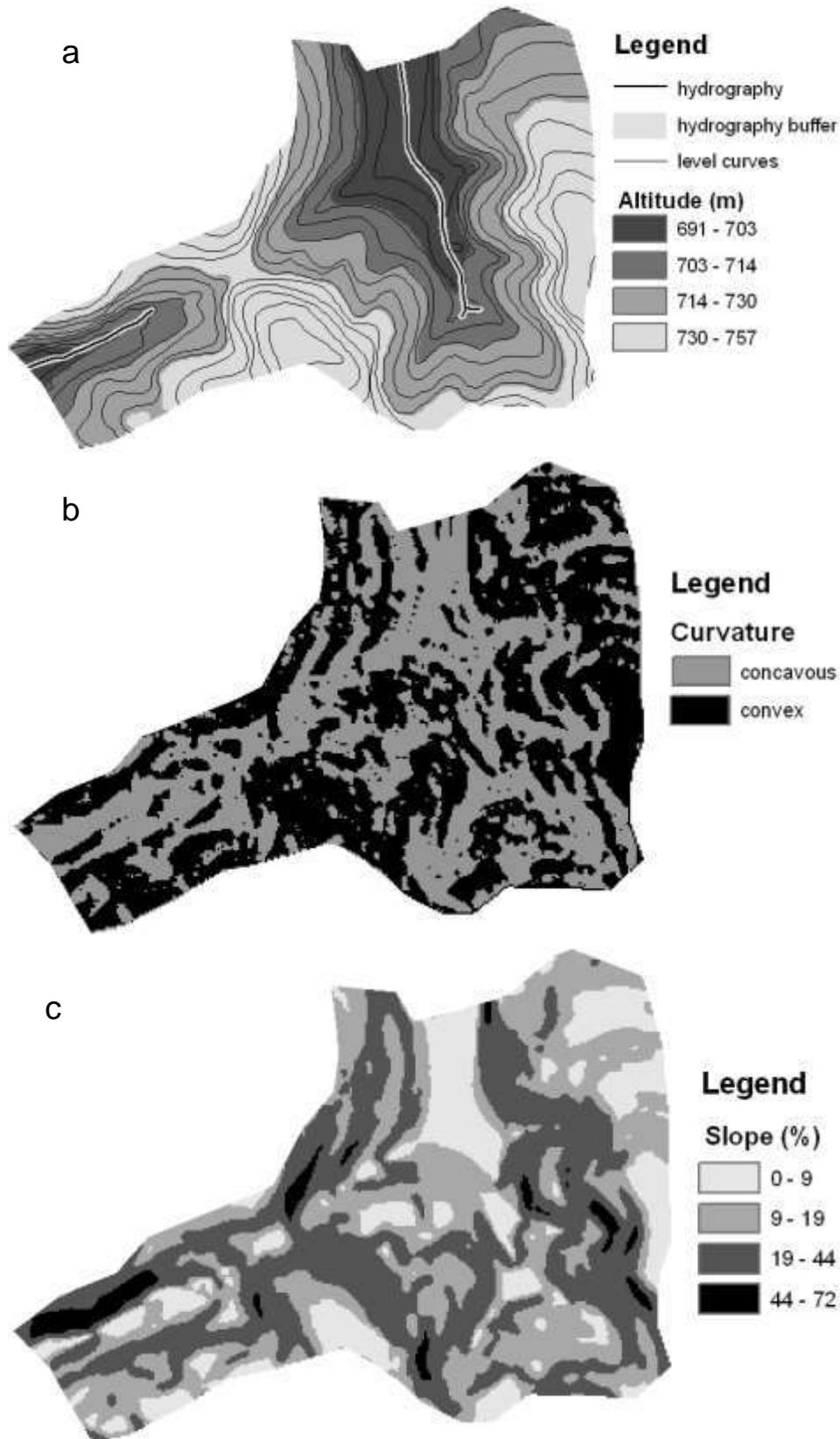
a



b



c



Figure 1. Maps derived from DEM of CAPTA-Frutas, Jundiaí, SP, Brazil. (a) altitude with level curves, hydrography and buffer; (b) curvature; (c) slope.

a

b

Legend
CX
GX
LA
PA

Figure 2. Digital soil map derived from DTM (a), and traditional soil mapping of CAPTA-Frutas, Jundiaí, SP, Brazil. CX: Haplic Cambisol; GX: Haplic Gleisol; LA: Yellow Latosol; PA: Yellow Argisol.

# Prediction Of Soil Carbon Stock In The Piaui State Coast By Remote Sensing

PORTELA, Mirya G. T.[1]; VALLADARES, Gustavo S.; PEREIRA, Marcos G.; CABRAL, Léya, J.R.S.; AMORIM, João V.A.; ESPINDOLA, Giovana M. de
[1] Federal University of Piaui, mgagronoma@yahoo.com.br

**Thematic Session: PEDOMETRICS GUIDELINES TO SYSTEMATIC SOIL**

**Abstract**
The objective of this study was to determine the organic carbon content (SOC) and soil carbon stocks under five different vegetations in the Parnaíba River Delta (PRD) in the Caatinga domain, located in the Brazilian state of Piauí, and to estimate them by using three predictive methods and the spectral bands and vegetation indexes derived from the Landsat 8 images. The Soil samples were collected for 40 points distributed in the area, where SOC and carbon stocks were determined, under vegetation classified as Psammophile pioneer vegetation (PPV), Dune subevergreen vegetation (DSV), Mangrove evergreen vegetation (MEV), Floodplain vegetation (FV), and vegetation associated with Carnaubals (VC). Afterward, the SOC of 0-10 cm and carbon stocks of 0-100 cm (CS100) were predicted using three predictive methods: multiple linear regression (MLR), ordinary kriging (OK) and regression kriging (RK). The results are that the soils under the mangrove evergreen vegetation showed the highest averages for carbon concentrations and carbon stocks for all stratus. In the SOC and CS100 predictions, it was observed that regression kriging (RK) was the best method, and this methodology can be used in soil carbon mapping in other areas.
Keywords: Parnaíba Delta; digital soil mapping; prediction methods; blue carbon.

## Introduction

The SOC is a dynamic property of this compartment and what represents the critical component of forest ecosystems, considered as potential carbon stores.

Coastal environments support biodiverse habitats of conservation interest and provide other essential benefits, such as carbon sequestration, due to the high rates of soil carbon accumulation. This carbon, called blue carbon, plays an essential role in climate change mitigation strategies, and presents variability depending on the region and factors such as soil and existing vegetation.

More research is needed in order to map SOC and CS in coastal regions, especially tropical rich mangroves. In these areas access is commonly difficult, as is soil sampling. Among the main difficulties in soil studies, the fact that SOC measurements require soil sampling, being expensive and time-consuming. Consequently, the number of samples available in a given area is generally scarce and does not reflect the actual level of variation that may be present in the study sites. Therefore, the precise interpolation of carbon concentrations in unsampled locations is necessary for better planning and management of theses areas.

Some techniques, from simple linear models to sophisticated techniques have been used to estimate the soil organic carbon. Multiple linear regression was more substantial compared to other methods in predicting organic carbon in soils in southern Brazil and in the Central Amazon.

This study aimed to determine the concentrations and carbon stocks in the soils of the Parnaíba River Delta (PRD), Piauí, employing digital soil mapping techniques.

**Methodology**

The study area is located in the state of Piauí, in the western portion of the northeast region of Brazil and comprises part of the *Parnaíba River Delta Environmental Protection Area* (APA), and portion of the *Parnaíba Delta Marine Extractive Reserve* (Resex). It occupies, more precisely, the region limited by the Igaraçu River on the southeast, Parnaíba River on the west, and the Atlantic Ocean, covering the municipality of Ilha Grande and part of Parnaíba, occupying an area of approximately 282 $km^2$, with approximately 8 $km^2$ of it on the *Parnaíba Delta Marine Extractive Reserve.* This APA is characterized by presenting a mosaic of ecosystems intersected by bays and estuaries, also being a very dynamic fluvial-marine region formed by the ecological tension between Caatinga, Cerrado, and marine systems.

The study was carried out from December 2016 to February 2017, in five areas with different vegetation types, classified as Psammophile pioneer vegetation (PPV), Dune subevergreen vegetation (DSV), Mangrove evergreen vegetation (MEV), Floodplain vegetation (FV), and vegetation associated with Carnaubals (VC). The points for carrying out the soil survey were previously defined, considering areas of high representativeness regarding the structure of the vegetation, based on the preliminary knowledge of the area, based on the photo interpretation also to obtain more significant variability, and on the Normalized Difference Vegetation Index (NDVI) and its variations. Through this index, it was possible to verify the density of photosynthetically active vegetation and, in this way, highlight the sample points in the study area.

Besides, considering that it is an area of native vegetation, the access routes are difficult. Consequently, collection points with greater accessibility for the work team were chosen, however, making sure to move at least 100 m away from the edges of roads and looking for the points preferably closer to those previously defined.
The soils were collected from 40 sample points of representative areas of the different types of vegetation, making a total of 242 samples of layers or horizons of soil profiles up to 100 cm deep. The soils collected represents 11 classes.

SOC contents were quantified using the wet oxidation method, which is based on the oxidation of organic carbon using dichromate ions in sulfuric medium. To determine the bulk density (Bd), the volumetric ring method was used. After determining the soil density, carbon stocks were calculated following Batjes (2000).

The spectral variables used to estimate the SOC and CS100, were obtained from images of the OLI sensor (Operational Land Imager), Landsat 8, orbit/point 219/062. The images obtained from the OLI instrument consist of nine multispectral bands, but in the study, only six bands were used (band 2 to band 7), the ones in the rang of the visible and infrared spectrums.

The image was collected in June 21st, 2016, obtained from the United States Geological Survey (USGS), with cloud coverage of 5.7%, solar elevation angle of 52.20, and azimuth angle of 44.36 degrees.

From the bands, eight indexes were generated: RVI, NDVI, SAVI, EVI, NDWI, GNDVI, MNDWI, CTVI. All remote sensing covariate rasters were assembled in a Geographic Information System (GIS) and their values extracted to the field soil data in ArcGis deriving the database used to build the forecasting models.

Three methods were used to predict the SOC and CS100, including multiple linear regression (MLR), ordinary kriging (OK) and regression kriging (RK). The MLR consists of determining the adjusted equations considering the soil variables as a dependent variable and all vegetation bands and indexes as the independent variables. The Stepwise backward algorithm was used to choose the most significant independent variables in the regression. The variable selection model (p <0.05) was calculated in XLSTAT, an extension of Microsoft Excel, for the CS100. The OK is a univariate method that uses the primary variable (SOC or CS) measured at sample points to predict them in non-sampled locations.  In the method, the mean is taken as a constant, but unknow, value and its satationary is assumed only within a local neighborhood centered on the location being forecast.   The RK is a hybrid geostatistical method, as it encompasses two approaches: first, it uses regression to predict a variable, and then it uses simple kriging to interpolate the residuals of the regression model (Hengl et al., 2004), and by difference improve the estimates. The RK used in this study was the type C, which involves an ordinary regression model followed by kriging the values of the regression residuals. The models of organic carbon concentration and carbon stocks were validated with 20% of the data, using three statistical parameters: RMSE, MAE, and $R^2$.

## Results and discussion

The SOC under vegetation ranged from 0.03 to 92.76 g $kg^{-1}$ of soil, with the highest average levels observed in soils under mangroove evergreen vegetation and the lowest average levels associated with soils under dune subevergreen vegetation Among the vegetations, the levels of organic carbon in MEV ranged from 18.02 to 92.76 g $kg^{-1}$, in FV ranged from 1.16 to 66.73 g $kg^{-1}$, in VC ranged from 0.25 to 46.59 g $kg^{-1}$, in DSV ranged from 0.03 to 3.67 g $kg^{-1}$, and in PPV ranged from 0.13 to 6.70 g $kg^{-1}$.

The CS100  values, regardless of vegetation, ranged from 5.83 to 466.63 Mg $ha^{-1}$. It was observed that the highest average value of carbon stock in the soil is associated with the mangrove evergreen vegetation, and the lowest averages associated with psammophile pioneer vegetation and dune subevergreen vegetation (p <0.01) (Figure 1), as well as observed for carbon content.
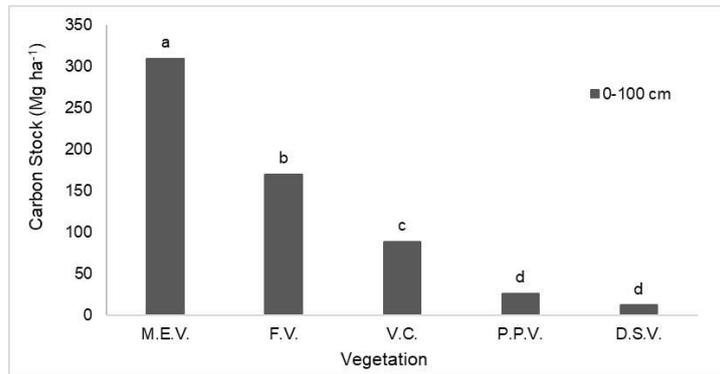
Figure 1. Average soil carbon stocks, in the different vegetation types in the Parnaíba River Delta, at 0-100 cm depths. MEV = Magrove evergreen vegetation; FV = Floodplain vegetation; VC = Carnaubal vegetation; PPV = Pioneer psamophilic vegetation; DSV=Dune subvergreen vegetation. Means of the same letter do not differ statistically. CV%0-100cm= 18,10%.

In the SOC and CS100 predictions, it was observed that RK had the lowest RMSE (5.54 $g.kg^{-1}$ and 38.35 $Mg.ha^{-1}$, respectively) and highest $R^2$ (0.97 and 0.95, respectively), being considered the best method for predicting these variables in the study area. The independent variables that showed the best responses for predicting CS100 were band 3, band 6 and band 7 .

## Conclusions

Soils under mangrove evergreen vegetation have higher SOC compared to other vegetations, as well as carbon stocks. In this environment in sandy soils, the lowest SOC and CS values were verified. In all soils, regardless of vegetation, carbon concentrations have decreased in depth. Among the prediction methods, RK was the most suitable for predicting CS, and the independent variables that presented the best responses were band 3, band 6, and band 7 (for CS100). The models used were satisfactory in the digital mapping of CS in the soils of the Parnaíba River Delta, and can be applied to others areas.

## Acknowledgements

## References

BATJES, N.H. Effects of mapped variation in soil conditions on estimates of soil carbon and nitrogen stocks for South America. **Geoderma**, v.97, n.1-2, p.135-144, 2000.
HENGL, T., HEUVELINK, G.B.M., STEIN, A. A generic framework for spatial prediction of soil variables based on regression-kriging. Geoderma, v.12, n.1-2, p.75-93, 2004.

# Combined proximal sensors for soil discrimination: A complementary tool for pedologists

GRESCHUK, Lucas T.[1]; DEMATTÊ, José A. M.[2]; RODRIGUEZ ALBARRACIN, H. Soledad [3]; BELLINASO, Henrique [4]; SILVERO, Nelida E. Q.[5]; ROSIN, Nícolas A.[6]

[1,2,3,4,5,6] Escola Superior de Agricultura Luiz de Queiroz (Esalq/USP), lucasgreschuk@usp.br

**Thematic Session: (4) Pedometrics guidelines to systematic soil survey: tropical study cases.**

## Abstract

Proximal sensing is a tool of relevance to pedology, as it provides suitable and quick information about soil attributes. The majority of the studies focus on the isolated use of proximal sensors in different wavelengths of the spectrum to getdata soil properties. Although sensors work on a specific range, the combination of them tends to increase the amount of information about soil. This study aimed to carry out a discriminant analysis of soil profiles, through the evaluation of spectral data from several bands along the electromagnetic spectrum together, including X-Ray fluorescence (XRF), visible (Vis), near (NIR) and short-wave infrared (SWIR), 350 - 2500 nm, and mid-infrared (Mid-IR, 2500 - 25000 nm). For that, this study had 5 steps. Fifteen soil profiles were morphologically described and collected in the state of Maranhão, Brazil, with subsequent laboratory analyses and taxonomic classification (Step 1), followed by the spectral acquisition of soil samples in the ranges X-Ray, Vis-NIR-SWIR e Middle Infrared (Step 2). After, was performed the Pearson correlation of the raw spectrum with soil attributes to choose of spectrum ranges further importants (Step 3). Thereafter, the cluster analysis of the soil profile considering color and spectral behavior. Last, the grouped soil was characterized and discriminated due to the spectral behavior of the soil profile, generating the description of the similarities between the soil profiles grouped together (Step 5). It is perceptible that the methodology used in this work was effective in complementing the soils discrimination, in terms of soil color, mineralogy and drainage conditions.

**Keywords:** digital mapping; proximal sensing; remote sensing; pedometrics; soil monitoring.

## Introduction

The determination of soil attributes using laboratory methods, such as physical, chemical, and X-ray diffraction (XRD) analyses, frequently is not a simple task because they are time-consuming, have considerable costs, and usually require wet-chemistry analyses (Zhang and Hartemink, 2019). As an alternative, diffuse reflectance spectroscopy can generate data that captures electromagnetic radiation interaction across the spectrum using proximal sensors to the soil sample, providing information at low cost, quickly and for a wide range of applications (Linker, 2007), in addition to not generating chemical effluents (Silva et al., 2021).

Proximal sensors can be an effective tool in the obtainment of soil information about the physical, chemical, and biological proprieties, performing your characterization and discrimination. When we use spectral sensors to analyze a soil profile, we are

making morphometrics, i.e., the application of tools that allow the measurement and quantification of the object (Hartemink and Minasny, 2014). Therefore, this study aimed to evaluate the joint spectral data obtained from various ranges along the electromagnetic spectrum, including X-ray Fluorescence (XRF), visible (Vis), near (NIR) and short-wave infrared (SWIR), 350 - 2500 nm, and middle infrared (Mid-IR, 2500 - 25000 nm) for complementing soil discrimination. We hope that the joint use of sensors is a complementary alternative to soil discrimination.

**Methodology**

This study had 5 steps (Figure 1). First, we carried out the collection of soil samples (Step 1), followed by the spectral acquisition of soil samples in the ranges X-Ray, Vis-NIR-SWIR e Middle Infrared (Step 2). After, was performed the Pearson correlation of the raw spectrum with soil attributes to choose of spectrum ranges further importants (Step 3). Thereafter, the cluster of the soil profile (Step 4.1) was done using only the visible range spectrum, due to the importance of soil color for its characterization and discrimination (Campos; Demattê, 2004). Then, a second grouping strategy (Step 4.2) was made up using the selected spectral bands by the Pearson correlation and considering the grouping of soil profiles by the result of their color. Last, the grouped soil was characterized and discriminated due to the spectral behavior of the soil profile, generating the description of the similarities between the soil profiles grouped together (Step 5).



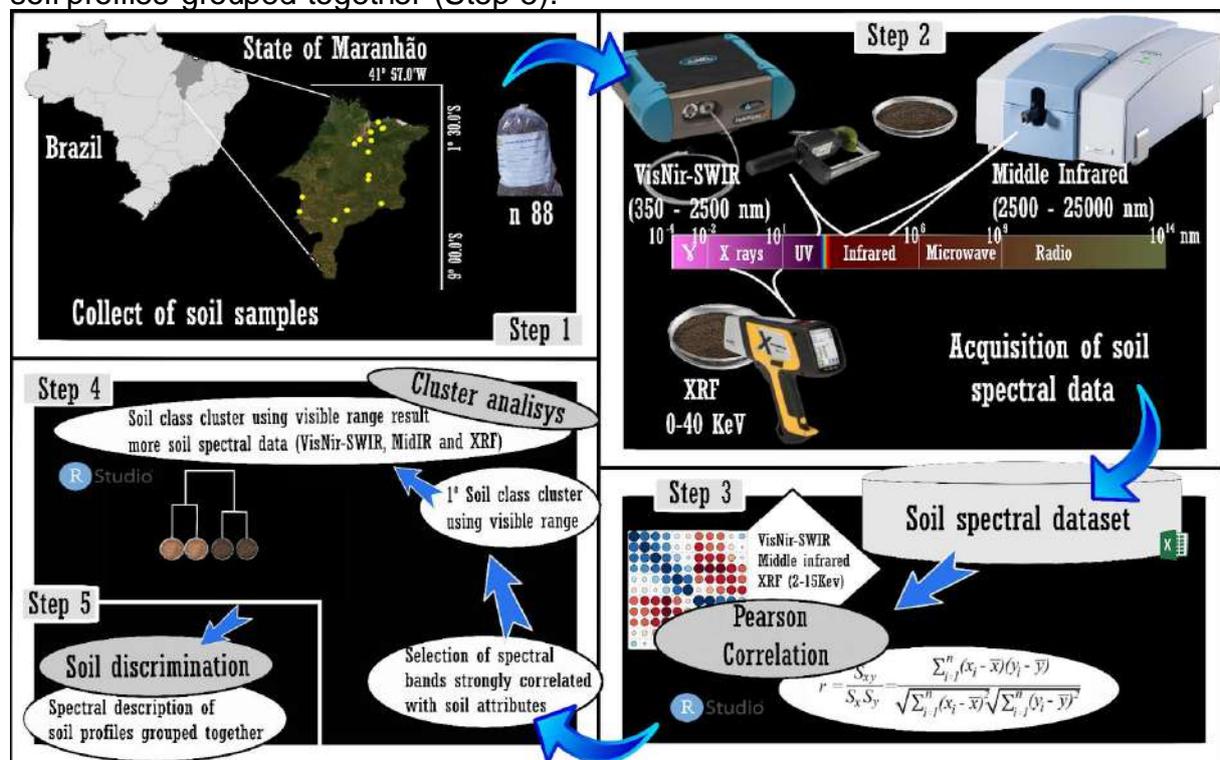Figure 1. Flowchart about the methodology of study.

**Results and discussion**

First, the color of the soil grouped the profiles by drainage condition (Figure 2a and b). Barbosa et al. (2019) mention that soils with gray and brown colors (Figure

2a) are strongly influenced by hydromorphic processes, generally associated with flat areas with seasonal accumulation of water and SOM presence. These colors come from the deferrification process that takes place in the soil (Barbosa et al., 2019). In agreement with Campos and Demattê., (2004), the yellow and red colors of soils are related to the presence of goethite and hematite, respectively (Figure 2b). The processes of oxidation and reduction of iron determine morphological characteristics of soils, giving them distinct colors (Barbosa et al., 2019). Grouping by spectrum presented groups of Vertissolos (MA-02 and MA-09), Plitossolo (MA-03) and Planossolo (MA-04), Plitossolo (MA-12) and Gleissolo (MA-13), and Espodossolo (MA-01 and MA-16) for poorly drained soils or with a higher concentration of SOM (Figure 2a). For soils with reddish colors, the result of grouping by spectrum was Latossolo Amarelo (MA-10) with Latossolo Vermelho Amarelo (MA-11), Latossolo Vermelho (MA-11) with Nitossolo Vermelho (MA-07), and Luvissolos (MA-06 and MA-08). This result comes from the spectral similarity between the soil profiles.
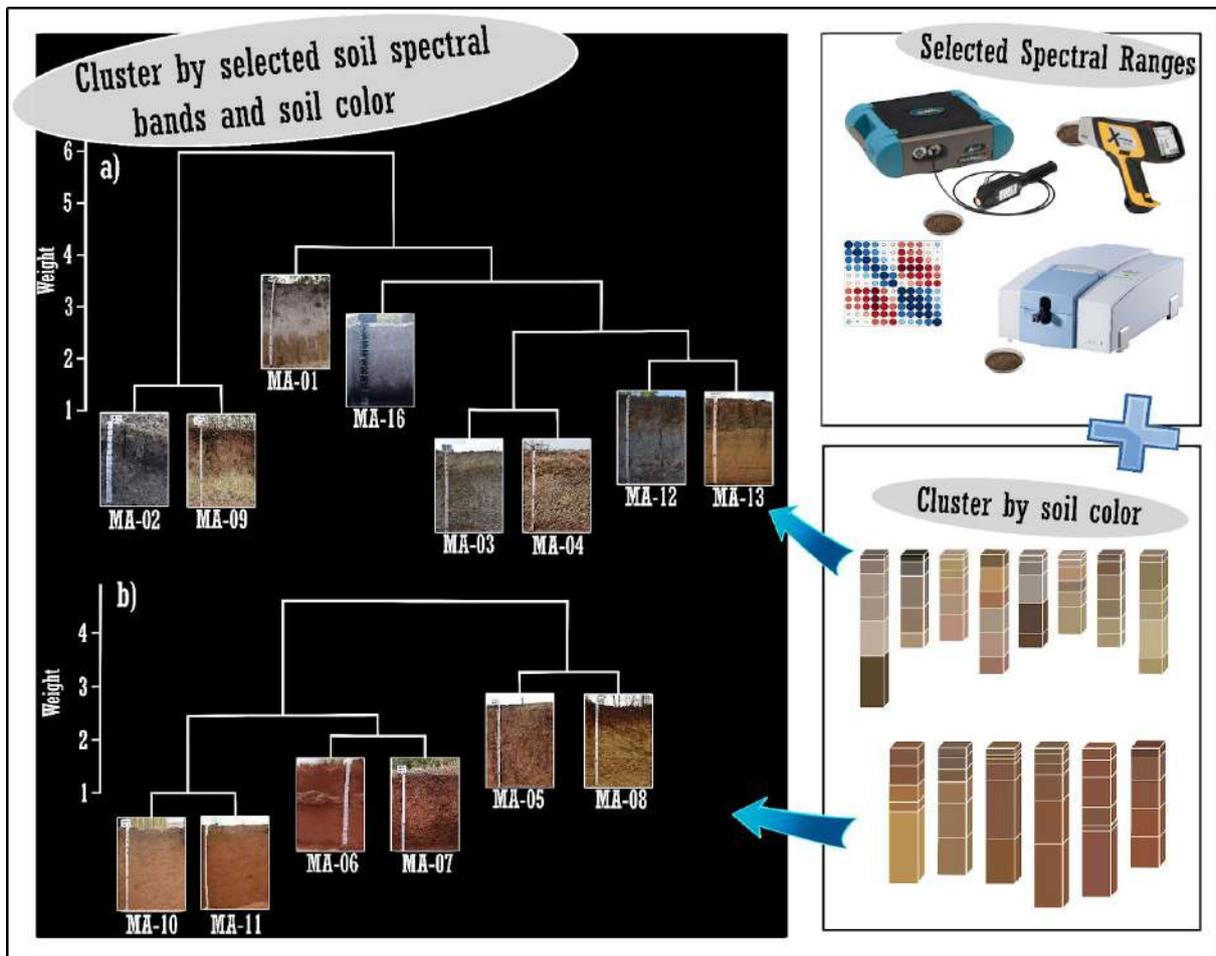


Figure 2. Cluster group a (soils with grayisher and brownisher colors) and cluster group b (soils with yellowisher and reddisher colors). The selected bands of the different spectra were: vis-NIR-SWIR (400 - 700, 1350 - 1450, 1900 - 2370 nm) and Mid-IR (3700 - 3300, 2760 - 2500, 2150 - 1875, 1200 - 900, 740 - 625 cm$^{-1}$).

## Conclusions

The methodological approach using several spectral ranges in a simultaneous evaluation demonstrated that complementary discrimination of soil profiles via sensors is feasible. The spectral analyses can assess several soil attributes important for soil classification, such as mineralogy, color, soil organic matter, drainage, and texture. With its specific spectral range, each equipment furnish different and additional information to assess the soil attributes and their discrimination. Finally, there is no better spectral range, but the information is complementary. Each one brings differential information related to the soil properties. The combined use of sensors can assist pedologists in soil survey, discrimination, and classification, bringing accurate, fast, low cost and environmentally friendly information about soil attributes.

## References

BARBOSA, G. S.; MALTONI, K. L.; PANOSSO, A. R. As cores do solo como parâmetro para delimitação do ambiente de uma vereda. Bol Goiano Geogr. 2019;39:1-25. https://doi.org/10.5216/bgg.v39i0.52021

CAMPOS, R.C.; DEMATTÊ, J.A.M. Soil color: approach to a conventional assessment method in comparison to an automatization process for soil classification. Revista Brasileira de Ciência do Solo. 2004;v.28;p.853-863.https://doi.org/10.1590/S0100-06832004000500008

HARTEMINK, A. E.; MINASNY B. Towards digital soil morphometrics. Geoderma. 2014. 230-231:305-17. https://doi.org/10.1016/j.geoderma.2014.03.008

LINKER, R. Soil Classification Via Mid-infrared Spectroscopy BT - Computer And Computing Technologies In Agriculture, Volume II. 2007; Boston, MA. Springer US; 2007. p. 1137-46.

ROSSEL R, JEON Y, ODEH I, MCBRATNEY A. Using a legacy soil sample to develop a mid-IR spectral library. Aust J Soil Res. 2008;46:1-16. https://doi.org/https://doi.org/10.1071/SR07099

SILVA, S. H. G.; RIBEIRO, B. T.; GUERRA, M. B. B.; CARVALHO, H. W. P.; LOPES G.; CARVALHO, G. S.; GUILHERME, L. R. G.; RESENDE, M.; MANCINI, M.; CURI N.; RAFAEL, R. B. A.; CARDELLI, V.; COCCO, S.; CORTI, G.; CHAKRABORTY, S.; LI, B. WEINDORF DC, 2021. pXRF in tropical soils: Methodology, applications, achievements and challenges. Advances in Agronomy. 167:1–62. https://doi.org/10.1016/bs.agron.2020.12.001

ZHANG, Y.; HARTEMINK, A.; E. Soil horizon delineation using vis-NIR and pXRF data. Catena. 2019; 180:298-308. https://doi.org/10.1016/j.catena.2019.05.001

# DETERMINATION OF SOIL ORGANIC MATTER AS A FUNCTION OF SOIL ATTRIBUTES IN TROPICAL SUGAR CANE AREAS

SILVA, Paulo Alexandre da Silva[1]; PANOSSO, Alan Rodrigo[2]; LA SCALA Jr, Newton[3]; PERUZZI, Nelson José[4]; VICENTINI, Maria Elisa[5]; ROLIM, Glauco De Souza[6]

[1] Universidade Estadual Paulista "Júlio de Mesquita Filho", paullo-alex@outlook.com; [2] Universidade Estadual Paulista "Júlio de Mesquita Filho", alan.panosso@unesp.br; [3] Universidade Estadual Paulista "Júlio de Mesquita Filho", lascala@fcav.unesp.br; [4] Universidade Estadual Paulista "Júlio de Mesquita Filho", peruzzi@fcav.unesp.br; [5] Universidade Estadual Paulista "Júlio de Mesquita Filho", mevicentini@gmail.com; [6] Universidade Estadual Paulista "Júlio de Mesquita Filho", rolim@fcav.unesp.br

## Abstract

Objective: use of the multiple linear regression technique to quantify soil organic matter, using soil attribute variables, in commercial sugarcane fields in Brazil. The experiment was carried out in a sugarcane field in the municipality of Motuca, Guariba, Pradópolis and Aparecida do Taboado. The technique used was multiple linear regression and the dependent variable was soil organic matter (SOM) and the independent variables were available phosphorus (P), cation exchange capacity (CEC), air temperature (Tair), air temperature. soil (Ts), particle density (PD) and soil moisture (SM). The multiple linear regression learning technique estimated the SOM as a function of the variables P, CEC, Tair, Ts, PD and SM, in areas of raw sugarcane, showing the existence of a high relationship between independent and dependent variables. Studies like this one, related to the determination of soil parameters are important for farmers, as the information generated can be used in the management and decision-making of the productive and financial process, being an alternative for the characterization not only of SOM, but also of other soil attributes.

Keywords: organic matter; multiple linear regression; soil physical and chemical attributes; Multivariate analysis.

## Introduction

The covering formed by the sugarcane straw provides a condition that favors the accumulation of carbon in the soil, contributing to the mitigation of greenhouse gases (GHGs) in the agricultural production of ethanol and sugar. According to data from the National Supply Company - CONAB (2020), in the 2018/2019 harvest the planted area was around 8.59 million hectares and production was 620.44 million tons. The maintenance of straw in the crop allows the increase of soil organic matter, which can generate an amount of residues of 10-20 t ha[-1] (TRIVELLIN et al., 2013). Organic residues that are formed by animal and vegetable residues, in different stages of decomposition (SILVA & MENDONÇA 2007), which is left on the soil after harvesting, resulting from the interactions that occur between the straw and the atmospheric, physical and chemical attributes of the soil.

The spatial variability of edaphoclimatic conditions in Brazilian regions directly interferes with organic matter, affecting the soil organic carbon dynamics. In this context, the objective of this work was to develop a mechanistic-empirical model, using the exploratory multivariate analysis of multiple linear regression data to predict the influence of each element of the soil-plant-atmosphere system in the soil organic matter quantification process, using the variables of the physical and chemical

attributes of the soil and climatic data, in commercial areas cultivated with the sugarcane culture in Brazil.

## Methodology

The study was carried out in a commercial area with sugarcane cultivation, under the raw cane management system, in Guariba - SP, Pradópolis - SP, Motuca - SP and Aparecida do Taboado - MS. In SP, the soils were classified as Latossolo Vermelho Eutroférrico and in MS as Latossolo Vermelho Distroférrico. Both had a clayey texture and the climate was Aw, with an average annual temperature of 23.7 ℃. Soil temperature was recorded by the LI-COR system (LI-8100). Soil moisture (SM) was measured by TDR equipment (Time Domain Reflectometry - Hydrosense TM, Campbell Scientific, Australia). After completion of the measurements, soil samples were collected at a depth of 0 to 0.10 m and later sieved in a 2 mm mesh. For chemical analysis, the following attributes were extracted: cation exchange capacity (CEC) and available phosphorus (P) content of the soil (RAIJ et al., 2001). From these same samples, the PD was also determined (EMBRAPA, 1997). the air temperature (Tair) was obtained by the NASA Power platform. All assessments were carried out at the beginning of the sugarcane crop development phase. Data were initially analyzed using descriptive statistics and the Shapiro-Wilk test at a 5% probability level. Then, the correlation matrix was performed using the Pearson method. The statistical technique used for the elaboration of the empirical-mechanistic model was the exploratory analysis of the multiple linear regression data. To evaluate the performance of the model, the following statistical indices were used: Coefficient of determination ($R^2$), Adjusted coefficient of determination ($R^2$ adj), Mean squared error (MSE), Mean squared error (RMSE), Akaike information criterion (AIC) and Schwarz Bayesian Criterion (BIC). For the realization of the model, 24 were used, with 18 points being used for the prediction of the model and 6 sampling points for validation.

## Results and discussion

The multiple linear regression equation that was estimated to explain the SOM as a function of the variables P, CEC, Tair, Ts, PD and SM was represented by Equation 1:

$$SOM = - 62.8932 - 0.5957\ P + 0.4946\ CEC - 3.2200\ e^{-15}\ Tair - 5.3290\ e^{-15}\ Ts + 22.4228\ PD - 2.2200\ e^{-15}\ SM$$

Where: SOM: soil organic matter (g $dm^{-3}$); P: available phosphorus (mg $dm^{-3}$); CEC: cation exchange capacity (mmolc $dm^{-3}$); Tair: air temperature up to 2 meters above the ground (℃); Ts: soil temperature (℃); SM = soil moisture (% of volume); PD: particle density (kg $dm^{-3}$); SM: Soil moisture (mm).

The value of the coefficient of determination of this model was $R^2$ = 0.999, and its respective adjusted value was, that is, 98.5% of the variations in the SOM estimate can be explained by variations in the values of the independent variables. The remaining 1.5% are sources of variations that are explained by other factors. Regarding the likelihood study of the model, the Akaike test (AIC) was performed, whose value was -1430 and the Schwarz Bayesian Criterion test (BIC), whose measured value was -1422. In both cases, it's observed that the independent

variables showed strong relationships with the dependent variable, showing that there is a strong relationship between the studied variables.

As a function of the predicted model for the SOM, the value of the linear coefficient was -62.8932 and presented statistical significance within the model (p-value = 0.000), showing that this coefficient strongly impacts the determination of the model. When the slopes were analyzed, the following results were obtained for the dependent variables: P = -0.5957, CEC = 0.4946, Tair = $-3.2200e^{-15}$, Ts = $-5.3290e^{-15}$, PD = 22.4228 and SM = $-2.2200e^{-15}$. Regarding the level of significance, p-values of 0.000, 0.000, 0.477, 0.433, 0.000 and 0.516, respectively, were observed. The variables that showed strong explanatory power in the model, through the level of significance, were P, CEC and PD. The SM, Tair and Ts variables did not show a significant level that could impact the model, but if this model is replicated for another data set, a greater relationship can be observed.

**Observed SOM versus Estimated SOM**

The linear regression formula that was estimated to validate the estimated SOM values as a function of the observed values was represented by Equation 2:

SOM_estimated = $-1.243e^{-14}$ + 1.000 SOM_observed

Where: SOM_estimated: soil organic matter (g dm$^{-3}$) estimated; SOM_observed: soil organic matter (g dm$^{-3}$) observed.

Similar to the study of the prediction of SOM, the value of the coefficient of determination of this model was $R^2 = 0.999$, and the value of the adjusted coefficient of determination was that is, 98.5% of the variations in the estimate of the SOM can be explained by variations in values of the independent variables. The remaining 1.5% are sources of variations that are explained by factors. In addition, the Akaike test (AIC) was also performed, whose value was -1468. The value found by the Schwarz Bayesian Criterion (BIC) to describe the interactions between the variables that maximize the probability of choosing the true model, whose measured value was -1466. For the two evaluative parameters, strong interactions of the independent variable with the dependent variable were observed, evidencing the strong linteraction of the observed SOM with the estimated SOM.
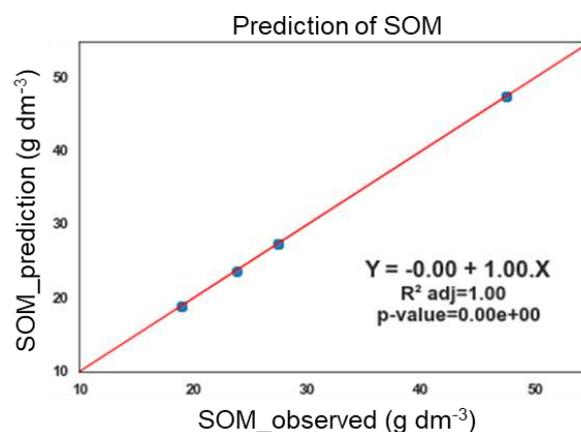


Figure 1. Plotting observed SOM data versus estimated SOM by the regression equation.

Figure 1 shows the graph of observed SOM versus estimated SOM. It is observed that the straight line angle, in relation to the linear fit was 45 degrees, characterizing the strong relationship between the observed SOM and the estimating SOM. In addition, also obtained the values of the RMSE (root mean squared error) which means the "root mean squared error" and also the MSE (mean squared error) which means the "mean squared error", values of 0 were found. stops the two parameters, showing that the observed and estimated values had high correlations.

## Conclusions

It was observed that the multiple linear regression learning technique was able to estimate the SOM as a function of the variables P, CEC, Tair, Ts, PD and SM, in areas of raw sugarcane, showing the existence of a high relationship between independent and dependent variables.

Studies like this one, related to the determination of soil parameters are important for farmers, as the information generated can be used in the management and decision-making of the productive and financial process, being an alternative for the characterization not only of SOM, but also of other soil attributes.

## Acknowledgements

Coordination for the Improvement of Higher Education Personnel (CAPES), Universidade Estadual Paulista (UNESP), Soil Characterization for Specific Management Purposes (CSME) and the Group of Agrometeorological Studies (GAS) for their support.

## References

CONAB, 2020. Acompanhamento da safra brasileira: cana–de–açúcar, segundo levantamento, maio/2020. In: Brasília. Conab. Companhia Nacional de Abastecimento Disponível: https://www.conab.gov.br/info-agro/safras/cana, Safra 2019/2020, (acess on out 13 2020).

EMBRAPA - Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA-CNPS). **Manual de métodos de análise de solo (In Portuguese), Centro Nacional De Pesquisa De Solos (2nd ed)**. Rio de Janeiro, Brazil, 1997.

RAIJ, B.V. **Análise química para avaliação da fertilidade de solos tropicais**. Campinas: Instituto Agronômico, 2001, 285 p.

SILVA IR & MENDONÇA ES. 2007. Matéria orgânica do solo. In: NOVAIS RF et al. **Fertilidade do solo**. Viçosa: SBCS. p. 275-374.

TRIVELLIN, P. C. O., FRANCO, H. C. J., OTTO, R., FERREIRA, D. A., VITTI, A. C., FORTES, C., FARON, C. E., OLIVEIRA, E. C. A., CANTARELLA, H. Impact of sugarcane trash on fertilizer requirements for São Paulo, Brazil. **Scientia Agricola**, Piracicaba, v. 70, n. 5, p. 345-352, 2013. Disponível em: http://dx.doi.org/10.1590/S0103-90162013000500009

# ANALYSIS OF SOIL CARBON STABILITY AS A FUNCTION OF SOIL ATTRIBUTES IN AREAS WITH RAW CANE MANAGEMENT

SILVA, Paulo Alexandre da Silva[1]; PANOSSO, Alan Rodrigo[2]; LA SCALA Jr, Newton[3]; PERUZZI, Nelson José[4]; VICENTINI, Maria Elisa[5]; ROLIM, Glauco De Souza[6]

[1] Universidade Estadual Paulista "Júlio de Mesquita Filho", paullo-alex@outlook.com; [2] Universidade Estadual Paulista "Júlio de Mesquita Filho", alan.panosso@unesp.br; [3] Universidade Estadual Paulista "Júlio de Mesquita Filho", lascala@fcav.unesp.br; [4] Universidade Estadual Paulista "Júlio de Mesquita Filho", peruzzi@fcav.unesp.br; [5] Universidade Estadual Paulista "Júlio de Mesquita Filho", mevicentini@gmail.com; [6] Universidade Estadual Paulista "Júlio de Mesquita Filho", rolim@fcav.unesp.br

## Abstract

Objective: To determine the structure of spatial variability of soil carbon losses, expressed by k, and its relationship with soil attributes, in sugarcane fields in the Brazilian Cerrado. The experiment was carried out in a sugarcane field in the municipality of Aparecida do Taboado (APT), in the state of Mato Grosso do Sul. The determination of $CO_2$ emissions from the soil was recorded by the LI-COR system (LI-8100). Principal components 1 and 2 (CP1 and CP2) explained 40.57% and 23.66% of the total variability of the data, respectively. The physical attributes of the soil correlates independently with CP1, linked to soil water. For CP2 the correlations were observed with chemical attributes, linked to soil fertility. The analysis of parameters corresponding to the spatial patterns of k with CP1 and CP2 previous and equal. The potential for carbon accumulation in the soil presents high spatial variability on a small scale, determined by changes in spatial patterns of k, proposing specific management regions.

Keywords: Spatial variability; principal component analysis; soil physical and chemical attributes, $FCO_2$; Multivariate analysis.

## Introduction

The covering formed by the sugarcane straw provides a condition that favors the accumulation of carbon in the soil, contributing to the mitigation of GHGs in the agricultural production of ethanol and sugar. According to data from the National Supply Company - CONAB (2019), in the 2018/2019 harvest the planted area was around 8.59 million hectares and production was 620.44 million tons. The State of São Paulo was the largest producer in the country, with 4.43 million hectares and production of 332.88 million tons, representing 53.65% of the processed sugarcane. The state of Mato Grosso do Sul is the fourth largest producer in the country, with 647.4 thousand hectares and production of 49.50 million tons, where it was responsible for 7.54% of Brazilian production. The proper use of soil management practices is of great importance for the process of mitigating GHG emissions in the soil. The practices of eliminating stumps by preparing the soil in sugarcane areas can lead to significant losses of soil carbon due to $CO_2$ emissions. The soil attributes that most influence the process of production and transfer of $CO_2$ into the soil are: soil density, soil texture, free water porosity, soil temperature and soil moisture (MOITINHO et al., 2015). The objective of this work was to determine the multivariate structure of the spatial variability of carbon losses, via $CO_2$ emissions, in agricultural

soils of the Brazilian Cerrado, under a commercial area cultivated with sugarcane, and to analyze its relationship with the attributes. physical and chemical soil.

## Methodology

The study was carried out in a commercial area under the cultivation of sugarcane, in the raw cane management system, in Motuca – SP and Aparecida do Taboado – MS. The soils were classified as Eutroferric Red Latosol and Dystroferric Red Latosol respectively. Both had a clayey texture and the climate was Aw, with an average annual temperature of 23.7 ºC. Soil $CO_2$ emission (Fm) and soil temperature were recorded by the LI-COR system (LI-8100). Soil moisture (Us) was measured by TDR equipment (Time Domain Reflectometry - Hydrosense TM, Campbell Scientific, Australia). The carbon stock (Estc) was calculated for a depth of 0.10 m (Estc = (CO.Ds.E).0.1; Estc = carbon stock (Mg ha$^{-1}$); CO = organic carbon content oxidizable (g kg$^{-1}$); Ds = soil density (kg dm$^{-3}$); E = thickness of the studied layer (0.10 m)). After completion of the measurements, soil samples were collected at a depth from 0 to 0.10 m and later sieved in a 2 mm mesh. For chemical analysis, the following attributes were extracted: cation exchange capacity (CTC) and available phosphorus (P) content of the soil (RAIJ et al., 2001). From these same samples, the Ds was also determined (EMBRAPA, 1997). All assessments were carried out at the beginning of the sugarcane crop development stage. Data were initially analyzed using descriptive statistics and the Shapiro-Wilk test at a 5% probability level. Then, the correlation matrix was performed using the Pearson method. The Factor Analysis multivariate statistical method was used to determine the processes between $CO_2$ flux and soil attributes. To determine the explanatory power of the main components, the data covariance matrix was used, from which the eigenvalues that originate the eigenvectors were extracted (Kaiser, 1958).

## Results and discussion

The first principal component, CP1, explained 59.20% of the total variance of the soil properties, while 14.36% was explained by CP2, totaling 73.56% of the variability of the soil properties. The formation of two groups was also observed: group I, with greater dispersion of points in the two-dimensional representation, located on the left side of the biplot plot of points, from the experimental area of Motuca, and group II, located to the right of CP1 and formed by Aparecida do Taboado. In the first main component and in order of importance, the attributes that presented the highest correlation coefficients were free water porosity - PLA (0.949), soil moisture - Um (-0.940), cation exchange capacity - CTC ( -0.939), carbon stock – EstC (-0.742), available phosphorus – P (-0.724) and Macroporosity - Macro (0.564). In the second component, the only variable that showed importance was the emission of $CO_2$ from the soil – Fm (0.890). There was a contrast with respect to soil attributes, where Motuca had higher values of P, CTC, Um and EstC, when compared to Aparecida do Taboado, which in turn had higher PLA and Macro. These relationships directly interfered with the stability of carbon in the soil, directly impacting the processes of decomposition of organic matter and the flow of carbon in the soil (Figure 1).
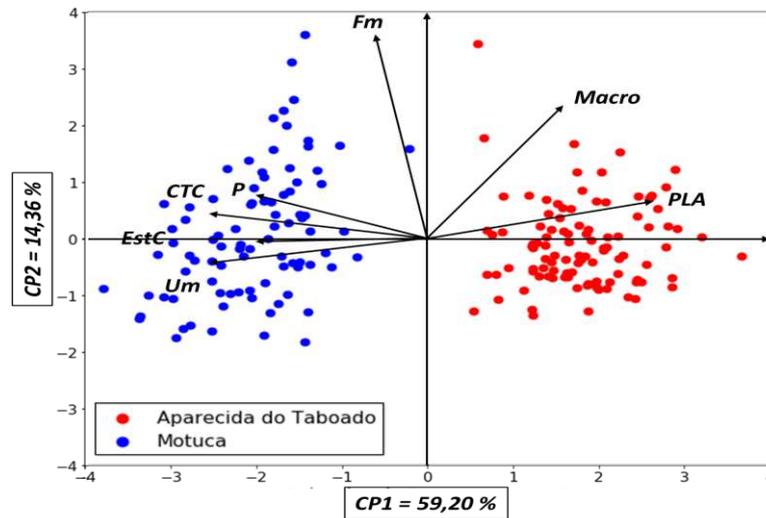
Figure 1. Biplot plot of CP1 and CP2 principal components of the principal component analysis with all sampling points and the following variables: soil $CO_2$ emission (Fm), soil carbon stock (Estc), free water porosity (PLA), available phosphorus (P), macroporosity (Macro), cation exchange capacity (CTC) and soil moisture (Um).

CP1 - main components 1; CP2 - main components 2; Fm = $CO_2$ emission factor ($\mu$mol $m^{-2}s^{-1}$); Um = soil moisture (%); Macro = macroporosity (%); P = available phosphorus (mg $dm^{-3}$); CTC = cation exchange capacity (mmolc $dm^{-3}$); Estc = soil carbon stock (Mg $ha^{-1}$); PLA = water free porosity (%).

In the cluster analysis, a contrast was observed between the points of high (A) and low (B) emissions in Motuca, which also happened in Aparecida do Taboado - MS, but the formation of an overlapping region, located more to the center, with points of high (A) and low (B) emissions, with a separation at the end, but with low intensity (Figure 2). The attributes of the soils that most influenced the Fm process were PLA, Um, CTC and P. When comparing this information with the cluster analysis, it is evident that the values of higher emissions represented by A are positively correlated with PLA, CTC, P and negatively with Um; whereas, for the points with higher emission, there is high PLA, CTC, P. For lower values of Fm, higher values of Um are observed.
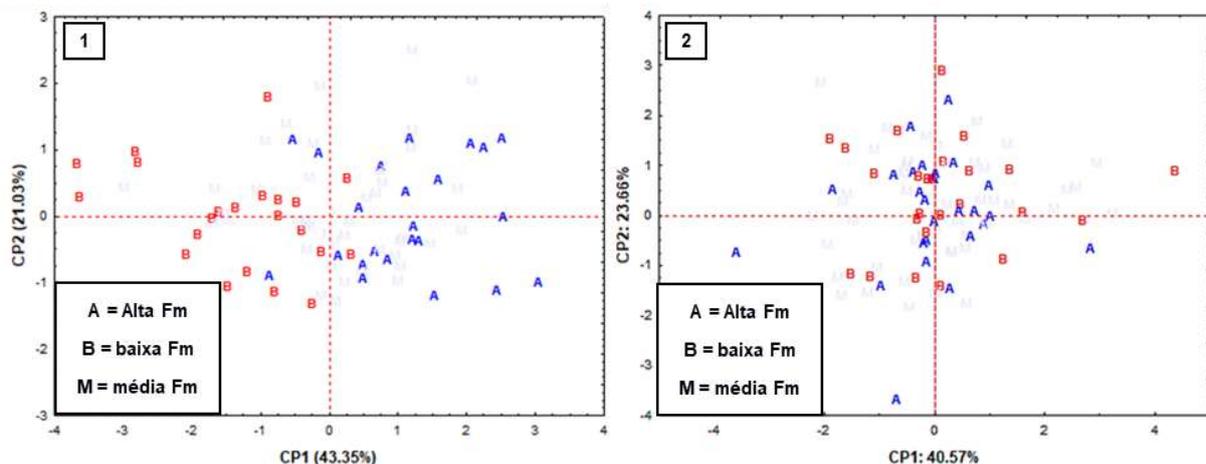
Figure 2: Variables in three classes of soil $CO_2$ emission groups for Motuca (1) and Aparecida do Taboado (2).

## Conclusions

Significant correlation coefficients were observed between the main components and the k factor in all study areas, where the physical attributes of the soil were the main responsible for $CO_2$ emissions, in areas of raw sugarcane, in the municipality of Motuca, in the state of Sao Paulo. In the municipality of Aparecida do Taboado - MS, the chemical properties of the soil were the main responsible for changing the spatial patterns of stability of carbon in the soil.

The results showed that within the same area, there were changes in the spatial patterns of k, leading to the occurrence of regions with potential accumulation or sequestration of carbon in the soil in sugarcane cultivation areas. used to improve the performance of agricultural practices, especially those related to soil preparation, use of agricultural inputs, accumulation of OM in the form of straw, water management, soil water content and mainly to promote the mitigation of greenhouse gases.

## Acknowledgements

## References

MOITINHO, M. R., PADOVAN, M. P., PANOSSO, A. R., TEIXEIRA, D. B., FERRAUDO, A. S., LA SCALA, N. On the spatial and temporal dependence of $CO_2$ emission on soil properties in sugarcane (Saccharum spp.) production, Amsterdam, **Soil and Tillage Research**, v. 148, p. 127-132, 2015.

EMBRAPA - Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA-CNPS). **Manual de métodos de análise de solo (In Portuguese), Centro Nacional De Pesquisa De Solos (2nd ed)**. Rio de Janeiro, Brazil, 1997.

KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. New York, **Psychometrika**, v. 23, p. 187-200, 1958.

RAIJ, B.V. **Análise química para avaliação da fertilidade de solos tropicais**. Campinas: Instituto Agronômico, 2001, 285 p.

# Field Navigation Mobile App – Agroecological zoning of the Mato Grosso do Sul state – ZAE/MS

**Bhering, Silvio Barge[1]; Dias, Hiran Silva[2]; Carvalho Junior, Waldir[1]; Chagas, Cesar da Silva[1]; Pereira, Nilson Rendeiro[1]; Oliveira, Aline Pacobahyba; Lopes, Carlos Henrique Lemos[3]**

[1]Embrapa Solos, {silvio.bhering;waldir.carvalho; nilson.pereira; cesar.chagas; aline.oliveira}@embrapa.br;

[2]Serviço Geológico do Brasil - CPRM, hiran.dias@cprm.gov.br

[3]Secretaria de Estado de Meio Ambiente, Desenvolvimento Econômico, Produção e Agricultura Familiar – SEMAGRO,chlopes@semagro.ms.gov.br

**Thematic Session: Pedometrics guidelines to systematics soil survey**

## Abstract

Obtaining primary soil data to support zoning studies is rare in Brazil. Navigation and adequate location of defined sampling points are limitations in the application of sampling design statistical techniques in large areas. The development of an mobile app, based on a platform of geographic information system for orientation and adequate location of sampling points, presented facilities compared to traditional navigation techniques in the field.

Keywords: soil survey; pedrometrics; sample design

## Introduction

Zoning studies are a traditional instrument of planning and ordering. Zoning studies may have various aspects, its land use and occupation are widely applied in urban and rural areas.

Due to the lack of data on natural resources in Brazil at compatible scales (=>1:100,000), these studies are traditionally based on secondary data. This zoning study is entirely based on the use of primary soil data and of pedometric techniques and it presented some new challenges.

Despite the advances made in the preparation of soil surveys, particularly with the adoption of digital soil mapping, either for attributes or for soil classes, the exact location and access by statistical techniques to predefined sampling points has been a great challenge.

This work presents the solution adopted in the Agroecological Zoning of the State of Mato Grosso do Sul (under development by Embrapa in partnership with the State Government) for navigation and access to description points and soil samples gathering.

## Methodology

The environmental covariates used to classify were selected from an initial set of 25 covariates. From this set, the variables were selected by eliminating the ones with high correlation (nonlinear correlation above 95%) to avoid the collinearity effect. In a second stage, among the categorical variables, those with high similarity were identified. Afterwards, the most important variables were ranked using the "Importance" function of the Randon Forest algorithm. The final set were composed of eight covariates. The morphometric variables related to the terrain were obtained through the R software with the R-Saga package, from the Digital Elevation Model (NASA JPL, 2020).

Categorical maps of the main geomorphological compartments of the state were additionally used, in addition to the lithology, based on the geological map (IBGE, 2021), vegetation (IBGE, 2021), soils (IBGE, 2021) and geodiversity (CPRM, 2009).

The sampling points were obtained through the geographically distributed stratified random sampling technique. Spatial restriction of a distance of less than 250m from the access roads available in the information plan of the basic cartographic material was used to favor access to the points. Latin Hypercube Conditioning method available in the R software (R Core team, 2020) was used in this procedure.

The first challenge in browser development was planning all the content (information layers and their features). The information layers, the form of presentation, display scales and symbologies used were defined in this stage. The entire set of basic information, including the basemap, consists of a satellite image and needs to be configured individually. The set is later encapsulated to enable its operation in offline mode since in areas with large geographic coverage and far from municipal headquarters internet connection is normally not available.

The municipal division with the location of the respective municipal headquarters, the access roads, individualized by type of pavement and jurisdiction, the hydrography and the mapping of categorical covariates already used in the definition of the sampling design were used as support information layers. This happened in this construction to facilitate access to predefined sampling points in the sampling design.

The application for field navigation was developed using ArcGis Explorer. All steps are performed in ArcGis Pro and the generated files are encapsulated for loading in the ArcGis Explorer software in this construction.

**Results and discussion**

Relevant aspect to be considered in the sampling design refers to the possible concentration of points in top elevation areas associated with water dividers, preferred location of most access roads. Thus, performing geographically distributed adjustments to the sampling plan with the adoption of toposequences was totally satisfactory for navigation and sampling (Figure 1).

Therefore, the indication of the spatial location of the soil samples corresponds to the spatial distribution of the covariates.

**Figure 1. Sampling design for Agroecological Zoning of Mato Grosso do Sul state – Phase 3.**

The standardization of maximum and minimum viewing scales of information layers in the file inserted in the application is associated with the need for visual details necessary for full navigation and with the size of the file generated for encapsulation. Though its size is limited by the mobile device hardware characteristics.

The standard area to be covered must be taken into account in the file development to be into the application used for field navigation. This area should be determined by maximum zoom scale desired. The scale must be suitable for navigating visual detail that can be identified by standard and alternative access to sampling points and the size of the file to be encapsulated.

Two example screens of the navigation application for mobile devices are presented below. It shows the navigation screen with the estimated location of points, the basic map used for navigation and in another a list of information layers enabled for support and information.

**Figure 2. Navigation mobile appscreens.**

## Conclusions

Field Navigator use enabled the adoption of the sampling statistics technique since the main challenge encountered was orientation difficulty and sampling points access.

This difficulty was consolidated by the use of traditional global positioning system devices associated with outdated altimetric bases and satellite images in printed media.

The application overcame the traditional techniques difficulties which incorporated a high time consumption,     navigation uncertainty and accessbility difficulty of alternative routes.

## References

BDia – IBGE (2021). Banco de Dados de Informações Ambientaisfromhttps://bdiaweb.ibge.gov.br/#/home. Accessed June18, 2021

NASA JPL (2020). *NASADEM Merged DEM Global 1 arc second nc V001*. NASA EOSDIS                                Land                            Processes DAAC. from https://doi.org/10.5067/MEaSUREs/NASADEM/NASADEM_NC.001.    Accessed August 28, 2020

CPRM (2021). Mapa de Geodiversidade do Estado do Mato Grosso do Sul from https://rigeo.cprm.gov.br/jspui/handle/doc/14703. Accessed July 21, 2021

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for StatisticalComputing, Vienna, Austria. URL https://www.R-project.org/