

Division - Soil In Space and Time | Commission - Pedometrics

Fine-scale soil mapping with Earth Observation data: a multiple geographic level comparison

José Lucas Safanelli⁽¹⁾ , José Alexandre Melo Demattê^{(1)*} , Natasha Valadares dos Santos⁽¹⁾ , Jorge Tadeu Fim Rosas⁽¹⁾ , Nélida Elizabet Quiñonez Silvero⁽¹⁾ , Benito Roberto Bonfatti⁽²⁾  and Wanderson de Sousa Mendes⁽¹⁾ 

⁽¹⁾ Universidade de São Paulo, Escolar Superior de Agricultura Luiz de Queiroz, Departamento de Ciência do Solo, Piracicaba, São Paulo, Brasil.

⁽²⁾ Universidade do Estado de Minas Gerais, Unidade Passos, Passos, Minas Gerais, Brasil.

ABSTRACT: Multitemporal collections of satellite images and their products have recently been explored in digital soil mapping. This study aimed to produce a bare soil image (BSI) for the São Paulo State (Brazil) to perform a pedometric analysis for different geographical levels. First, we assessed the potential of the BSI for predicting the surface (0.00-0.20 m) and subsurface (0.80-1.00 m) clay, iron oxides (Fe_2O_3), aluminum (m%) and bases saturation (V%) contents at the state level, which are important properties for soil classification. In this task, legacy soil samples, the BSI and terrain attributes were employed in machine learning. In a second moment, we evaluated the capacity of the BSI for clustering the landscape at the regional level, comparing the predicted patterns with a legacy semi-detailed soil map from a smaller reference site. In the final stage, the predicted soil maps from the state level were investigated at the farm level considering several sites distributed across the São Paulo state. Our results demonstrated that clay and Fe_2O_3 reached the best prediction performance for both depths at the state level, reaching a RMSE of less than 10 %, RPIQ higher than 1.6 and R^2 of at least 0.41. Additionally, the predicted landscape clusters had a significant association with the main pedological classes, subsurface color, soil mineralogy and texture from the legacy semi-detailed soil map. Illustrative examples at the farm level indicated great capacity of BSI in detecting the variations of soils, which were linked to several soil properties, such as texture, iron content, drainage network, among others. Therefore, this study demonstrates that BSI is valuable information derived from optical Earth Observation data that can contribute to the future of soil survey and mapping in Brazil (PronaSolos).

Keywords: remote sensing, soil cartography, soil mapping, pedometrics, PronaSolos.

* **Corresponding author:**
E-mail: jamdemat@usp.br

Received: July 07, 2021

Approved: September 24, 2021

How to cite: Safanelli JL, Demattê JAM, Santos NV, Rosas JTF, Silvero NEQ, Bonfatti BR, Mendes WS. Fine-scale soil mapping with Earth Observation data: a multiple geographic level comparison. Rev Bras Cienc Solo. 2021;45:e0210080.

<https://doi.org/10.36783/18069657rbc20210080>

Editors: José Miguel Reichert  and Marcos Gervasio Pereira 

Copyright: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.



INTRODUCTION

Digital soil mapping (DSM) has been explored in Brazil to produce updated and higher resolution maps of soil properties and soil taxonomic classes (Chagas et al., 2010; ten Caten et al., 2012; Samuel-Rosa et al., 2013; Giasson et al., 2015; Arruda et al., 2016; Vasques et al., 2016; Moura-Bueno et al., 2019). Different methods were also adapted for predicting soil taxonomical classes (Nolasco de Carvalho et al., 2015; Coelho et al., 2021), such as the probability mapping (Giasson et al., 2006), categorical mapping (Bazaglia Filho et al., 2013; Wolski et al., 2017; Moura-Bueno et al., 2019), Boolean logic (Rizzo et al., 2020), spatial disaggregation (Sarmiento et al., 2017), among others. The DSM relies on the relationship of soil observations with environmental data employed in a model, which delivers a predicted value for a specific location usually coupled with an uncertainty estimate (McBratney et al., 2003; Lagacherie et al., 2006; Minasny and McBratney, 2016).

In this regard, Earth Observation (EO) data are explored as model predictors due to their widespread availability and ability to cover large geographical areas. However, when employing DSM over huge areas with high or medium resolution EO data, the project may be limited by the high computing processing requirements and storage capacity. Especially in large countries, DSM has important obstacles to overcome. We can also mention the practical and economical limitations of sampling strategies that seek to improve the representation of soil diversity, as well as the lack or poor relationship of some environmental information with soils when calibrating prediction models.

Generally, EO products have been used as covariates for inferring soil spatial distribution, despite the continuous vegetation cover in both arable lands and natural forests (Chabrillat et al., 2019). For example, the *Manual Técnico de Pedologia (Technical Manual of Pedology)*, which describes some guidelines for soil survey and mapping in Brazil, indicates that remote sensing products are useful for soil mapping if coupled with the intensification of field surveys (IBGE, 2015). However, different from this view, multitemporal collections of satellite images increased its value in recent years, as these products provide useful insights related to soil properties (Chabrillat et al., 2019). This can support both the quantitative and qualitative interpretation of soils at the landscape level. However, multitemporal EO data in the optical domain can only assist the soil surveying and mapping if coupled with appropriate transformation or preprocessing techniques.

Many automated processing algorithms have recently been developed to produce soil reflectance maps from optical EO data (Diek et al., 2017; Demattê et al., 2018; Rogge et al., 2018). In Brazil, the Geospatial Soil Sensing System (GEOS3) has been used to map soil properties (Fongaro et al., 2018; Gallo et al., 2018; Poppiel et al., 2019a) and infer the spatial distribution of soil and lithological classes (Mendes et al., 2019; Poppiel et al., 2019b; Rizzo et al., 2020). The algorithm is based on the retrieval of historical soil exposures to the satellite measurements, which are aggregated into products namely Synthetic Soil Image (SYSI) and Soil Frequency (SF). These products reveal the direct multispectral response of soils and the cumulative alterations of soil surface for a given location, which can be correlated with soil properties and land intensification (Demattê et al., 2020), respectively. However, despite the importance and developments in the area, there is a lack of knowledge regarding the impact of such analysis in multiple geographic levels. Our hypothesis is that a bare soil image derived from multitemporal collections of satellite images can be employed in pedometric analysis at different geographical levels, despite its standard spatial resolution of 30 m.

This paper aimed to produce a direct representation of soil reflectance for the São Paulo State in Brazil using the historical Landsat image collections. With the spectral features derived from EO data, we aimed at assessing the relationship of the bare soil reflectance with soil properties commonly used for soil taxonomic classification at the state level, i.e., clay, iron oxides (Fe_2O_3), aluminum saturation (m%) and bases saturation (V%),

by calibrating machine learning models for spatial prediction. Further, we also aimed to study soil reflectance images' potential for clustering landscape units over a reference site, comparing its results to a legacy semi-detailed soil map. Finally, the medium-resolution maps were assessed on different farms from São Paulo state. Thus, this study explores the potential of medium-resolution products derived from EO data for assisting the soil survey and mapping in Brazil (PronaSolos).

MATERIALS AND METHODS

We performed the study across the state of São Paulo, Brazil, mainly in agricultural areas due to the availability of bare soils from cropped lands. We produced a bare soil composite for the state level using the historical collections of Landsat images. We used this product to assess the major surface spectral patterns across the state and evaluate the quantitative relationship with some soil properties used for soil classification. First, we explored and discussed the method and its impact on the state level. In a second moment, we explored the bare soil composite over a smaller reference area near the Piracicaba municipality. In this step, we compared the soil units of a semi-detailed legacy map with landscape units generated from a clustering method. Then, we explored the results at the farm level to complete our assessment. In the next sections, the geographical areas and statistical analysis are described.

Soil properties mapping at the state level

The major region explored in this study is the Brazilian state of São Paulo, which represents an area of 248,222 km² (Figure 1). São Paulo has a relatively high relief, with 85 % of its landscape between 300 and 900 m above sea level. In the region, seven distinct climate types appear, being predominant the tropical climate of altitude (CWA) and savanna climate (AW). The CWA is characterized by rainy summer and dry winter, while the Aw is warm and has a minimal rainfall (Alvares et al., 2013).

About 30 % of São Paulo's territory is composed of a crystalline basement, dominated by igneous (granite) and metamorphic rocks. The remaining 70 % consist of deposits of the Paraná and Bauru basins, which resulted in sedimentary rocks from the glacial-interglacial cycle that occurred in the southern portion of the Gondwana paleocontinent. The Paraná basin was the scenario of intense tectonic-magmatic processes during the Mesozoic to Cenozoic period, resulting in basaltic flows, dike swarms and several alkaline complexes.

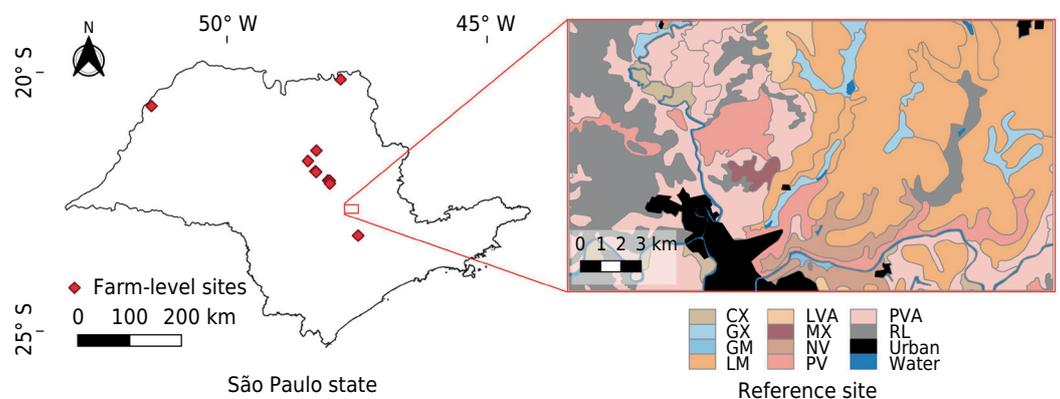


Figure 1. Study area representing the São Paulo State and several farm-level sites (left). The reference site near the Piracicaba municipality with a semi-detailed pedological map (right, scale 1:100,000) is also highlighted. CX: Cambisol (*Cambissolo*), GX: Gleysol (*Gleissolo*), GM: Histic Gleysol (*Gleissolo Hístico*), LV: Ferralsol (Reddish) (*Latossolo Vermelho*), LVA: Ferralsol (Yellowish-red) (*Latossolo Vermelho Amarelo*), MX: Haplic Phaeozem (*Chernossolo Háplico*), NV: Eutric Nitisol (*Nitossolo Eutrófico*), PV: Acrisol (*Argissolo Vermelho*), PVA: Lixisol (*Argissolo Vermelho Amarelo*), RL: Regosol (*Neossolo Regolítico*).

Conversely, the Bauru basin is represented by continental sandstones and constitutes about 45 % of the territory, corresponding to a major part of the substrate of the Western region of the state. Intracontinental and coastal Cenozoic basins, as well as Quaternary sedimentary deposits, also occur in minor proportions (Garcia et al., 2018).

The pedodiversity is also noticeable. Based on the World Reference Base (WRB) soil classification system (IUSS Working Group WRB, 2015), Lixisols and/or Acrisols predominate in more undulating topography, mostly in the Western part. Ferralsols, Arenosols and Eutric Nitisols are mostly found in gentle slopes in relatively geomorphological stable positions, similar to what is found in some portions of the Western and Central regions of the state. Conversely, Cambisols, Leptsols and Regosols are typical of unstable positions of the landscape, like hilly slopes and floodplains. Gleysols, Podzols and Histosols are inherent in coastal plains and wetlands, where the water table is persistent (Oliveira et al., 1999; Rossi, 2017).

Bare soil reflectance and terrain attributes

The algorithm Geospatial Soil Sensing System - GEOS3 (Demattê et al., 2018) - developed inside the Google Earth Engine, was used to generate the multispectral bare soil composite from 1982 to 2018 using the Landsat collections of 30-m-resolution surface reflectance images (USGS, 2018a,b). The GEOS3 is a data-mining algorithm that extracts soil features from the historical collection of satellite images and aggregates the spatially sparse bare soil fragments into a denominated as Synthetic Soil Image (SYSI). The SYSI is the reflectance image, composed of six spectral bands: blue (450–520 nm), green (520–600 nm), red (630–690 nm), NIR (760–900 nm), SWIR1 (1550–1750 nm) and SWIR2 (2080–2350 nm). A set of rules was used to identify bare soil pixels in satellite images, based on spectral indices coupled with quality assessment bands to remove cloud, cloud shadow, inland water, snow, photosynthetic vegetation and non-photosynthetic vegetation (crop residues). Bare soils were flagged when the pixel value had the Normalized Difference Vegetation Index (NDVI) values falling between the range of -0.05 and 0.30 (masking out green vegetation), and Normalized Burn Ratio 2 index (NBR2) values between the range of -0.15 and 0.15 (masking outcrop residues). The pixels identified as bare soil were marked as valid for aggregating the multitemporal collection by their median value. Thus, GEOS3 produces an almost-continuous representation of surface reflectance, increasing the mapped area of the soil surface by averaging the multitemporal measurements. Detailed information about GEOS3 and SYSI, as well as spectral indices and sensitivity analysis, can be found in previous studies (Demattê et al., 2018, 2020; Fongaro et al., 2018; Gallo et al., 2018). In addition, the Digital Elevation Model (DEM) derived from the Shuttle Radar Topography Mission (SRTM) with a resolution of 30 m was used to calculate the slope, northness, eastness, horizontal curvature, vertical curvature and shape index of terrain (Safanelli et al., 2020a). This terrain information was used as additional covariates in prediction models.

Legacy soil observations

The legacy soil observations gathered across the São Paulo State came from previous surveys developed by the Geotechnologies in Soil Science Group (<https://esalqgeocis.wixsite.com/geocis>) and others (Nanni et al., 2004; Terra et al., 2015; Rizzo et al., 2016; Fongaro et al., 2018; Gallo et al., 2018; Demattê et al., 2019; Mendes et al., 2019). Surface (0.00-0.20 m) and subsurface (0.80-1.00 m) samples ($n = 22,311$) were collected with augers in georeferenced locations, mostly following the sampling strategy of toposequences or grid schemes.

In this study, a few key soil properties explored by the Brazilian Soil Classification System - SiBCS (Santos et al., 2018) were selected for further analysis: clay content, total iron content (Fe_2O_3), base (V%) and aluminum saturation (m%). Clay content is used, for example, to evaluate the textural gradient between the surface and subsurface horizon,

which is a diagnostic attribute of Lixisols. The iron oxides content (total iron) is a diagnostic attribute for classifying soils at the third categorical level, which is related to color and mineralogical features. The saturation of aluminum or bases (Ca^{2+} , Mg^{2+} , K^+ , Na^+) in the sorption complex of soils are used to distinguish the soil chemical fertility at the third categorical level of SiBCS.

Soil samples had their chemical and particle size analysis following Brazilian standards. The densimeter method was used to obtain the soil particle size distribution (clay, silt and sand), whereas the chemical properties (V% and m%) and Fe_2O_3 were determined by ion-exchange resin method and sulfuric acid extraction, respectively, according to Raij et al. (2001). Although the soil data is composed of several legacy surveys performed in the last years, which can affect the performance of prediction models, they still provide a baseline of the soil spatial distribution across the São Paulo State.

Soil properties prediction models

Prediction models of clay, Fe_2O_3 , V% and m% for the 0.00-0.20 and 0.80-1.00 m layers were applied in the study area. The Random Forests (RF) algorithm, which is an ensemble of many regression trees built with bootstrapped samples and randomized predictors (Breiman, 2001), was chosen to calibrate the soil prediction models. For this, we performed a grid search for defining the best hyperparameters for each soil layer and soil property. The number of trees of the forest (FS), the number of random predictors tested at splits of each tree (nRP), and also the minimum number of samples at the tree leaves (minSL) were tested.

The training set was determined by collecting bootstrapped samples of each regression tree, whereas the testing set was defined by the remaining samples (out of the bag) that were not selected at each bootstrapping. Bootstrapping is a method that samples observations with replacement to the same size of the dataset, covering, in general, around 67 % of the entire dataset. We applied the bootstrapping method up to 500 times, which is equivalent to the maximum number of trees tested for the random forest. The performance metrics were determined by the mean value calculated across the collection of bootstrapped trees for both the training and the test sets. The minimum Root Mean Squared Error (RMSE) of the training set was used for defining the best hyperparameters (FS, nRP, and minSL). The range of values tested for FS was 30, 60, 100, 200 and 500 trees. The values of 3, 5, 8, 11 and 13 predictors were investigated for defining the nRP: SYSI blue, SYSI green, SYSI red, SYSI NIR, SYSI SWIR1, SYSI SWIR2, elevation, slope, northness, eastness, horizontal curvature, vertical curvature, and terrain shape index (Table 1). Further, the values of 10, 20, 30, 40, 50, 100, 200 and 500 observations were tested for minSL.

The final model performance was evaluated by applying the optimized models to the test set. The prediction accuracy was evaluated using the RMSE. The Coefficient of Determination (R^2) was calculated to assess the explained variance of the prediction models, and the Ratio of Performance to Interquartile range (RPIQ) was calculated to assess the consistency between the predicted values with the variability of soil observations (Safanelli et al., 2020a). The evaluation performance was defined by averaging the metrics from the out-of-the-bootstrapped estimates.

Landscape units at the regional level

At the second stage of multilevel analysis, a smaller reference site was selected near Piracicaba municipality due to the availability of a semi-detailed soil map and expressive soil variability. The Piracicaba region is characterized by a complex geology, diverse topography, and the occurrence of different soil types (Figure 1). The region is classified as humid subtropical by the Koppen's classification system, characterized by dry winter and hot rainy summers. The average annual rainfall is approximately 1273 mm yr^{-1} and

Table 1. Predictors used for model calibration

| Predictors | Source (spatial resolution) |
|---------------------------|-----------------------------|
| SYSI blue (450–520 nm) | Landsat collection (30 m) |
| SYSI green (520–600 nm) | Landsat collection (30 m) |
| SYSI red (630–690 nm) | Landsat collection (30 m) |
| SYSI NIR (760–900 nm) | Landsat collection (30 m) |
| SYSI SWIR1 (1550–1750 nm) | Landsat collection (30 m) |
| SYSI SWIR2 (2080–2350 nm) | Landsat collection (30 m) |
| Elevation | SRTM (30 m) |
| Slope | TAGEE/SRTM (30 m) |
| Northness | TAGEE/SRTM (30 m) |
| Eastness | TAGEE/SRTM (30 m) |
| Horizontal curvature | TAGEE/SRTM (30 m) |
| Vertical curvature | TAGEE/SRTM (30 m) |
| Terrain shape index | TAGEE/SRTM (30 m) |

the average annual temperature is 24 °C. The landscape is gently undulating and rolling uplands, with slopes between 0 and 42 % (Gallo et al., 2018).

The geology includes fine to medium, reddish, silty-clay sandstone of the Botucatu and Pirambóia Formation; dike elements of basic intrusive diabase rocks of the Serra Geral Formation; sandstones from the Irati Formation (Passa Dois Group), where the relief is gently undulating to flat; and siltstones, that are the predominant lithology of the Itararé and Tatuí Formation (Tubarão group) (Bonfatti et al., 2020). Thus, the soil variability is mainly due to the complexity and diversity of the parent material. Soil derived from diabase rocks, such as Ferralsols, are situated in the highest position of the relief, while Nitisols occur on smooth to gentle undulating relief. Soils derived from sedimentary rocks include Leptsols, Arenosols, Lixisols, which may be associated or not. On the flat relief near to the drainage system, Planosols and Gleysols are found (Gallo et al., 2018). In Piracicaba region (Figure 1), a smaller reference site (area of ~350 km²) was selected to study the relationship of bare soil composite with soil mapping units of a semi-detailed pedological chart of the São Paulo state that corresponds to Piracicaba region. This map is a result of several surveys carried out in the area, providing the boundaries of soil units as a vector layer with a 1:100,000 scale (Rossi, 2017).

Principal components and k-means clustering analysis

The SYSI of the reference site was transformed by a principal component analysis (PCA) to reduce the spectral dimension and produce uncorrelated components. The first three components that accounted for 99 % of the spectral variability were employed to k-means clustering analysis to produce spatial units defined by their multispectral reflectance. The optimal number of clusters was defined when a decrease of the Akaike Information Criteria (AIC) was equal to or lower than 10 % from the range of 1 and 20 clusters. The AIC criteria was selected to define a parsimonious model for clustering the reference site, i.e., it was used to define the simplest landscape delineation with the least assumptions and number of clusters, but with the greatest explanatory capacity (Dotto et al., 2020). The mean reflectance of each cluster was calculated for spectral characterization.

The landscape units defined by k-means clustering were compared to the legacy soil mapping units using the correspondence analysis, which evaluates the association of two categorical groups. The Chi-squared test was applied to assess the association strength between clusters and soil mapping units, and the relationship between the levels of both groups was evaluated from a contingency table and the respective diagram of

principal coordinates. Thus, the relationships can be statistically and visually assessed (Demattê et al., 2019).

For this analysis, the soil mapping units were interpreted according to their predominant soil class since a few units were composed of soil associations with ranked abundance. The dominant taxonomic class was decomposed in interpreted information representing the soil type, color, fertility, iron oxides content, and subsurface texture. The decomposition followed the first, second, and third levels of SiBCS, with the subsurface texture being used as an additional soil property. The levels from the SiBCS classification are based on the presence or absence of soil characteristics, diagnostic horizons or properties that can be identified at the field, which are resulted from a set of dominant processes that acted during soil formation. In table 2, we present the original and translated categorical levels decomposed from the dominant soil taxonomic class of soil units.

Farm-level assessment

Farm-level assessment was performed by linking the aforementioned products (SYSI, terrain attributes, and soil properties maps) with field observations. We went to eight sites distributed across the São Paulo State to qualitatively assess the bare soil image and its derived products. First, we wanted to evaluate the performance of the traditional toposequence sampling strategy for depicting the soil variation across the

Table 2. Conversion of classification terms from the Brazilian Soil Classification System - SiBCS (Santos et al., 2018) to World Reference Base - WRB/FAO (IUSS Working Group WRB, 2015)

| Original (SiBCS) | Adapted to WRB/FAO |
|---------------------------------|--|
| <i>First categorical level</i> | |
| Argissolo | Acrisols, Lixisols*, Alisols |
| Cambissolo | Cambisols |
| Chernossolo | Phaeozems*, Kastanozems, Chernozems |
| Gleissolo | Gleysols*, Stagnosols |
| Latossolo | Ferralsols |
| Neossolo | Leptosols, Arenosols, Regosols*, Fluvisols |
| Nitossolo | Nitisols*, Lixisols or Alisols |
| <i>Second categorical level</i> | |
| Litólico | Lithic (continuous rock contact) |
| Háplico | Haplic (qualifier for general properties) |
| Vermelho | Reddish |
| Vermelho-Amarelo | Yellowish-red |
| Melânico | Histic horizon |
| <i>Third categorical level</i> | |
| Distrófico | Dystric (low fertility) qualifier |
| Distroférico | Dystric and Ferric (high Fe ₂ O ₃ content) |
| Eutrófico | Eutric (high fertility) qualifier |
| Eutroférico | Eutric and Ferric |
| <i>Subsurface texture</i> | |
| Média | Loam |
| Média/Argilosa | Clay loam |
| Argilosa | Clay |
| Argilosa/Muito Argilosa | Clay/Heavy clay |
| Muito Argilosa | Heavy clay |

* Selected when multiple options were available.

landscape. Without using the products derived from EO data, we distributed and made field observations with an auger in three different depths (0.00-0.20, 0.40-0.60 and 0.80-1.00 m) with subsequent soil analysis. This evaluation allowed the comparison of the toposequence sampling strategy with the SYSI and its derived maps.

In another stage, a different approach was tested. In this case, SYSI and its predicted maps were used to allocate observation points based on their spatial patterns. This would make it possible to detect features related to the drainage network detected by SYSI, which could be linked to redoximorphic features of Gleysols. Other sites with high variation of clay content at the farm level were also observed in situ to assess the discrimination potential of SYSI for smart sample allocation and soil survey. In the end, two detailed maps were delineated using all information available. This final analysis was decisive to complete the multilevel analysis of products derived from EO data.

RESULTS

Spectral features from EO data

The SYSI with 30 m resolution produced over the São Paulo State reveals different patterns that are related to soil and rock types (Figure 2). In the true color composition (Figure 2a), SYSI is presented with brown shades that resemble the true color of soils. The false color of SYSI (Figure 2b), composed in part by infrared bands (R: SWIR1 G: NIR, and B: red), allows better visualization of the soil variability at the state level (Figure 2c). It is also worthy to note the Southeastern region of the state is barely mapped with bare soils, possibly related to permanent land uses, such as natural forests, and places where metamorphic rocks prevail (Figure 2d).

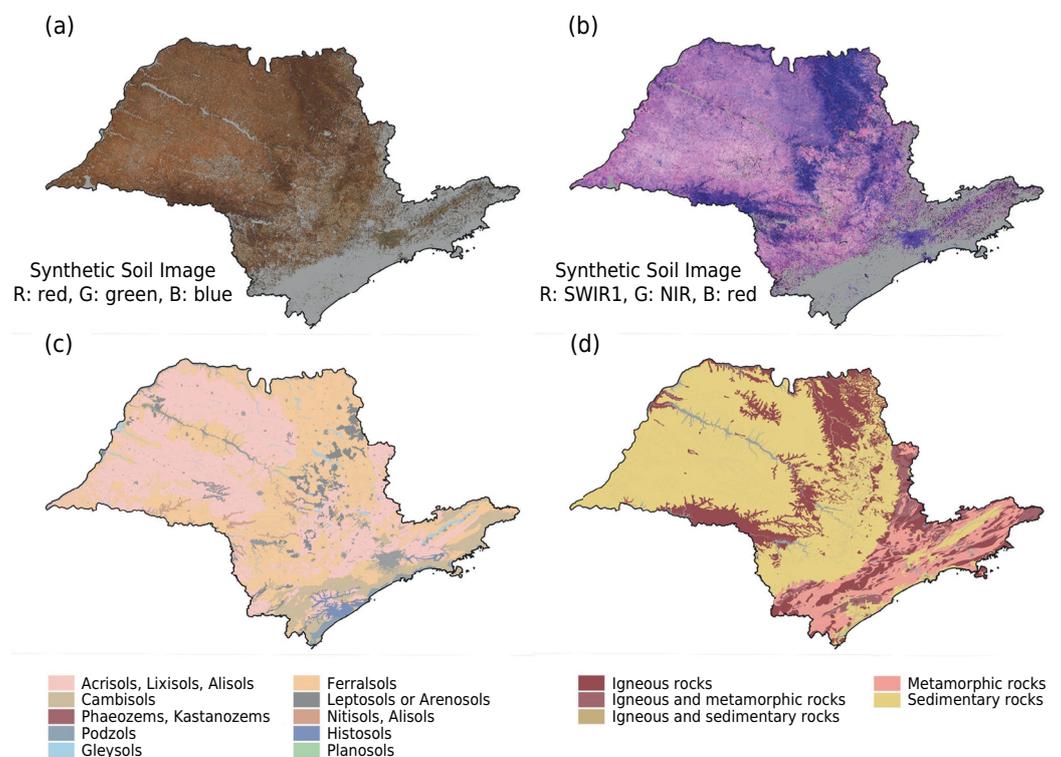


Figure 2. a) Synthetic Soil Image (SYSI) in true-color composition; b) SYSI in false-color composition; c) Main soil taxonomic classes (scale 1:500.000 from Rossi (2017)); d) Rock types (IPT - Instituto de Pesquisas Tecnológicas, 1981). Gray color indicates unmapped soil reflectance in SYSI images and the absence of information in soil and lithological maps. The readers are referred to figure 1 for spatial reference.

Soil properties prediction at the state level

When SYSI was employed with terrain attributes for calibrating prediction models of soil properties, the visual aspects confirmed its association to clay and Fe_2O_3 content (Table 3). The aluminum and base saturation of the sorption complex of soils presented unsatisfactory performance (either $R^2 < 0.4$ or $\text{RPIQ} < 1.4$, and RMSE around 20 %) for 0.00-0.20 and 0.80-1.00 m layers. For clay and Fe_2O_3 , the surface layer had the best performance metrics: the lowest RMSE and highest R^2 values between the layers, reaching an explained variance (R^2) between 0.41 and 0.69, and RMSE of less than 10 %.

The visual aspect of the predicted maps of clay and Fe_2O_3 (Figure 3) followed the same patterns of SYSI and rock types (Figure 2). The dark shades of both SYSIs had higher estimates for both clay and Fe_2O_3 contents in both predicted soil layers. The places that predominate igneous rocks are also associated with regions of high clay and Fe_2O_3 contents (Figure 2). The distinct predicted layers also reveal that subsurface had higher

Table 3. Performance evaluation of soil attributes prediction using out-of-the-bag validation

| Property ⁽¹⁾ | Layer ⁽²⁾ | n ⁽³⁾ | RMSE ⁽⁴⁾ | RPIQ ⁽⁵⁾ | R ²⁽⁶⁾ | FS ⁽⁷⁾ | nRP ⁽⁸⁾ | minSL ⁽⁹⁾ |
|-------------------------|----------------------|------------------|---------------------|---------------------|-------------------|-------------------|--------------------|----------------------|
| | m | | | | | | | |
| Clay | 0.00-0.20 | 22311 | 8.56 | 1.67 | 0.69 | 60 | 13 | 200 |
| Clay | 0.80-1.00 | 20557 | 9.31 | 1.70 | 0.65 | 200 | 13 | 200 |
| Fe_2O_3 | 0.00-0.20 | 476 | 4.17 | 1.72 | 0.57 | 60 | 6 | 10 |
| Fe_2O_3 | 0.80-1.00 | 181 | 8.06 | 2.09 | 0.41 | 100 | 3 | 20 |
| m | 0.00-0.20 | 21933 | 16.91 | 0.69 | 0.07 | 200 | 13 | 500 |
| m | 0.80-1.00 | 20509 | 25.51 | 2.04 | 0.20 | 30 | 13 | 200 |
| V | 0.00-0.20 | 21982 | 19.56 | 1.46 | 0.06 | 30 | 13 | 500 |
| V | 0.80-1.00 | 20555 | 20.57 | 1.80 | 0.15 | 30 | 11 | 200 |

⁽¹⁾ Fe_2O_3 (%): total iron oxides content; m (%): aluminum saturation; V (%): bases saturation. ⁽²⁾ Layer: soil depth (m). ⁽³⁾ n: number of available samples per soil attributes and layer. ⁽⁴⁾ RMSE: root mean squared error. ⁽⁵⁾ RPIQ: ratio of performance to the interquartile range. ⁽⁶⁾ R^2 : coefficient of determination. ⁽⁷⁾ FS: hyperparameter forest size. ⁽⁸⁾ nRP: hyperparameter number of random predictors. ⁽⁹⁾ minSL: hyperparameter minimum number of samples at leaves.

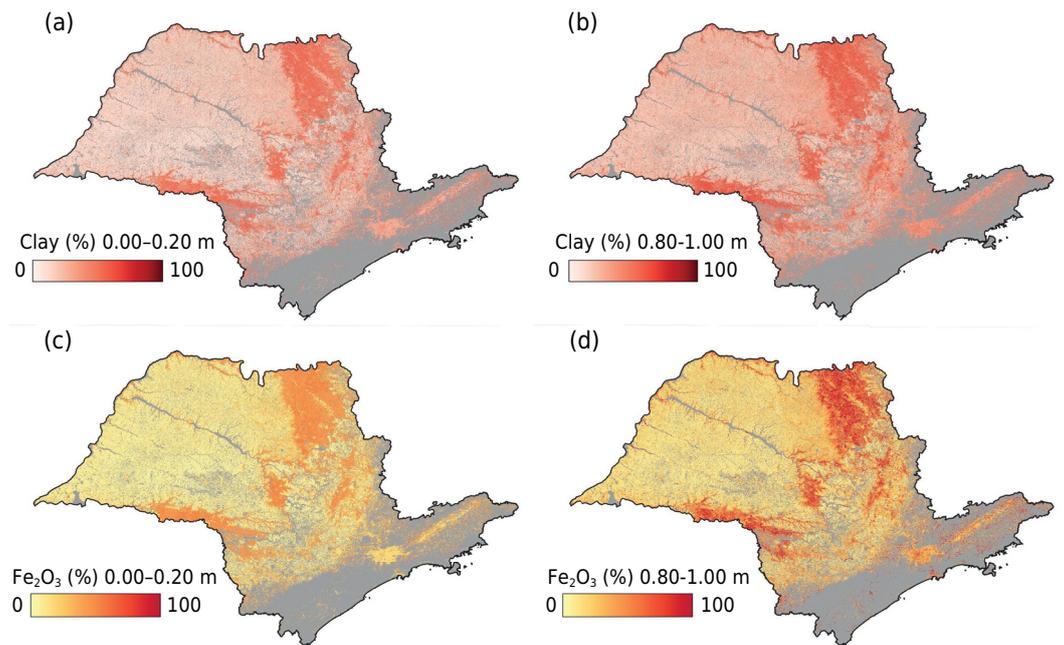


Figure 3. a) Predicted clay content for the 0.00-0.20 m layer; b) predicted clay content for the 0.80-1.00 m layer; c) Predicted Fe_2O_3 content for the 0.00-0.20 m layer; d) Predicted Fe_2O_3 content for the 0.80-1.00 m layer. The readers are referred to figure 1 for spatial reference.

estimates for both soil attributes (Figures 3b and 3d), as demonstrated by the slightly darker shades of subsurface maps (0.80-1.00 m).

Landscape discrimination at the regional level

From the visualization of the state level, we went further in detail to understand the spatial patterns at regional level. However, it is worth mentioning that the spatial resolution of the bare soil image and its products is 30 m, changing only the visualization level. The SYSI of the reference site was firstly processed by PCA to reduce the spectral dimension and produce uncorrelated features. The first three components accounted for almost 99 % of the spectral variability, which were influenced by the overall intensity, the effect of the infrared region, and the contrasting effects of correlated bands.

In the clustering analysis, the first three components promoted the identification of seven spectral groups across the reference site. The spatial distribution of cluster revealed a relationship with some soil taxonomic classes found in the reference site (Figure 4). The regions dominated by Ferralsols (LV) and Nitisol (NV) coincided with the ones with darker shades in the true color composite of SYSI (Figures 4a and 4b) and clusters with low reflectance intensity (Figure 4c). Conversely, the high predominance of Regosols (RL) and Lixisols (PVA) in the opposite side of Ferralsols coincided with brighter colors of SYSI and the remaining clusters. The soil types of high reflectance intensity are situated in the lower parts of the reference site, while the Ferralsols and dark-colored soils prevail on the higher portions of the landscape (Figure 4a).

The spatial association of the spectral clusters with the pedological map is confirmed ($p < 0.01$) by the correspondence analysis (Figure 5). For the first classification level, the soil units composed of Nitisol, Ferralsol and Gleysol had a significant association to clusters number 2 and 3 (Figure 5a), which presented a lower reflectance intensity and are situated in the darker shaded area of SYSI (Figure 4b). Soil units composed of Cambisol, Regosol, Lixisol and Phaeozem are associated with clusters 4, 6 and 7. A similar association was found for the second categorical level, where the soil color and other general properties are defined (Figure 5b). Reddish soils are linked to cluster 2 and 3, while yellowish-red soils are linked to groups 5 and 7. The lithic and general qualifiers (Haplic) were grouped in the middle of the diagram close to clusters 1, 4 and 6 (Figure 5b). For the qualifiers of soil fertility and iron content (Figure 5c), the third categorical level, it is possible to find an association of soils with high Fe_2O_3 content to clusters of numbers 2 and 3, which have a lower reflectance. Finally, the subsurface texture also reveals the grouping of high clay content soils linked to the spectral groups 1, 2 and 3 (Figure 5d).

Farm-level assessment

The illustrations and observations at the farm level revealed that SYSI and its related products provided great assistance to soil survey and mapping, despite its limited resolution (Figure 6). Figure 6a demonstrates that regular satellite images or aerial photographs present vegetation patterns, restricting the visualization of the soil surface. In a first example, after SYSI generation (Figure 6b), a drainage pattern was observed using the true color composition, since a pale color was detected within a reddish soil zone. The field observation performed at that site confirmed the presence of redoximorphic features (Figure 6c).

Figure 6d represents an apparently homogeneous area by the regular satellite image. However, the SYSI false-color composite (Figure 6e) depicted the spatial variation across the entire area and the clay content could be quantified by a prediction model, which followed the same patterns of SYSI (Figure 6f). As ratified by field observations, the area with light magenta color were classified with low clay content, while the darker ones were estimated with higher clay content. It is important to note that there

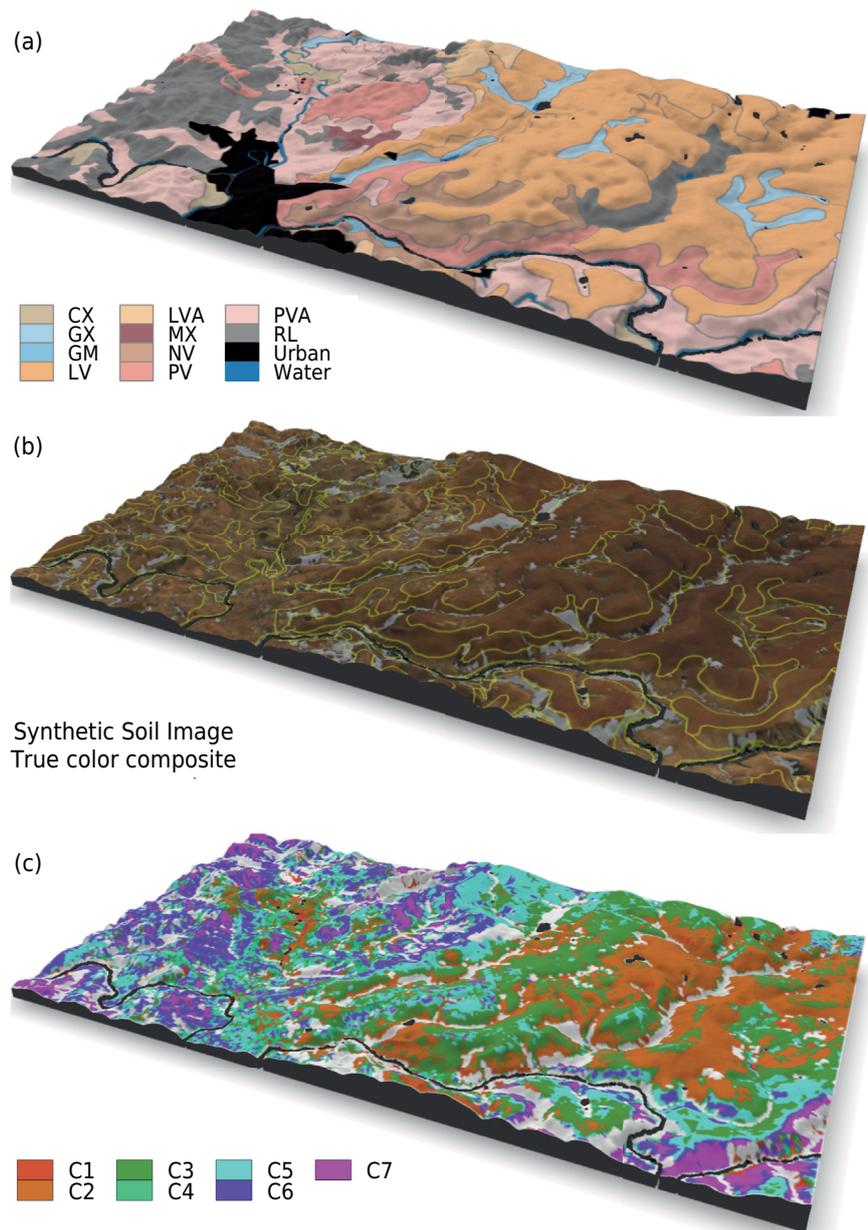


Figure 4. a) 3D visualization of the soil taxonomic map (scale 1:100,000); b) Synthetic Soil Image (SYSI) in the true-color composite with map units in yellow lines; c) Spatial distribution of spectral clusters (prefix C) across the reference site. The readers are referred to figure 1 for spatial reference and soil classes names.

was no continuous trend of these patterns along the terrain, and thus, the traditional inference of conventional soil mapping could fail in detecting those soil transitions. Indeed, as observed from figures 6d and 6e, the classical delineation following the terrain patterns is limited, while the information derived from EO data allows qualitative and quantitative analysis.

A more impacting change about the terrain variation is indicated in figures 6g and 6h. As it represents a smaller area, the clayey soil does not follow the terrain trend, as the continuum is broken by a sandy soil. If a pedologist demarked the toposequences 1 and 2, he/she could surely detect the soil variation, but could not detect the same transition in position 3. Another good example is provided from figures 6i to 6l.

A strong relief characterizes the site represented in figures 6i to 6l with variable lithology (shale, sandstone, and basalt). With a regular aerial photograph or satellite image (Figure 6i), it is impossible to detect most of the landscape variations due to vegetation

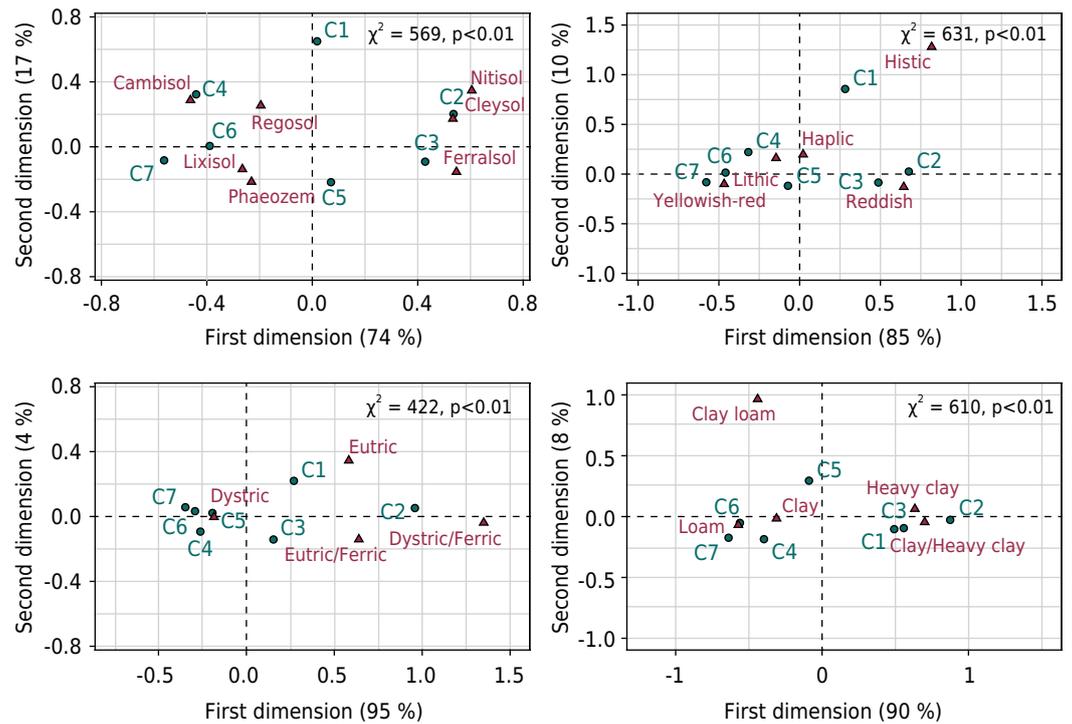


Figure 5. Correspondence analysis of landscape clusters with information derived from the dominant class from the semi-detailed soil map: first categorical level (a), second categorical level (b), third categorical level, and subsurface soil texture (d).

and residues cover. When the elevation gradient is used (Figure 6j), a contrasting relief is revealed. However, only with the bare soil composite that a synoptic view of the soil landscape can be generated (Figure 6k), which brings an additional comprehension about the soil types. Integrating terrain information with SYSI patterns make possible the delineation of soil types with higher precision, as expressed in figure 6l with the resulting detailed map.

Figures 6m, 6n and 6o illustrate the differentiation of soils in a relatively homogenous area with flatter terrain. In this situation, the soil surface is depicted only by the bare soil reflectance, making it difficult for a pedologist to work with only elevation gradients. In fact, the locations with question marks reveal the places with patterns that seem to be related to the same terrain positions.

DISCUSSION

Predictions at the state level

The darker shades from both the true color and false composites of SYSI (Figure 2) illustrated the soils with a lower reflectance intensity in the visible channels (blue, green and red), which are related to clay, iron oxides and organic carbon (Dematté et al., 2019). In this sense, the spatial variation revealed by the pedological map (Figure 2c) presented a certain degree of association with the true and false-color compositions of SYSI, where the darker shades were linked to Ferralsols and the brighter ones to Acrisols, Lixisols, or Arenosols. Besides that, it is evident the correlation of SYSI darker shades to igneous rocks, in agreement with study of Bonfatti et al. (2020).

When SYSI was employed in machine learning, the performance metrics (Table 3) confirmed its association to clay and Fe_2O_3 for both soil layers. These satisfactory results for clay and Fe_2O_3 are in accordance with previous studies that used bare soil images and other environmental information for DSM (Fongaro et al., 2018; Poppiel et al., 2019;

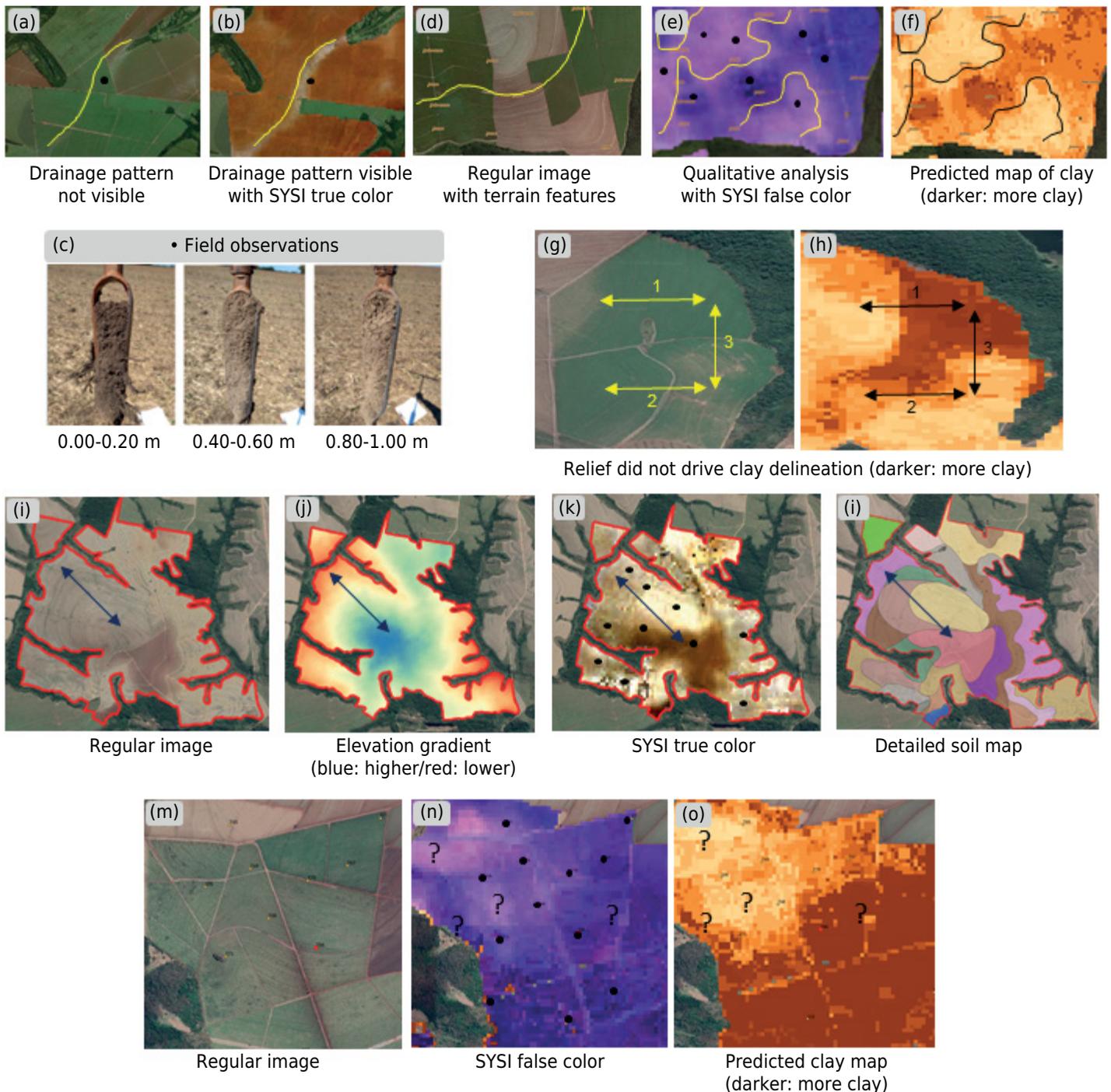


Figure 6. Farm-level illustrations: a) aerial photograph of site 1 (lat: -48.28047, lon: -21.50934), b) SYSI true color of site 1, c) field observation at site 1, d) aerial photograph of site 2 (lat: -48.05636, lon: -22.08765), e) SYSI false color of site 2, f) surface clay content of site 2, g) aerial photograph of site 3 (lat: -48.02316, lon: -22.09264), h) surface clay content of site 3, i) aerial photograph of site 4 (lat: -47.47652, lon: -23.15061), j) digital elevation model of site 4, k) SYSI true color of site 4, l) detailed soil map of site 4, m) aerial photograph of site 5 (lat: -48.29843, lon: -21.90618), n) SYSI false color of site 5, o) surface clay content of site 5.

Safanelli et al., 2020a). In addition, the visual correspondence of the predicted maps of clay and Fe_2O_3 with rock types (Figure 2) confirmed the association of soil reflectance with soil minerals and its granulometric composition (Dalmolin et al., 2005), which are important characteristics used for soil taxonomic classification. The poor prediction performance of chemical properties (m% and V%) found in this study is also present in other studies (Poppiel et al., 2019a; Rizzo et al., 2020; Safanelli et al., 2021). This can be linked to several factors, such as an insufficient calibration dataset, the spatial-temporal variation of these soil attributes, and the weak relationship of the environmental predictors tested in this study with the chemical soil properties.

Landscape patterns at the regional level

Principal component analysis was used before the clustering analysis to compact information and produce uncorrelated features, similar to what other studies have explored (ten Caten et al., 2011; Viscarra Rossel et al., 2011). Overall, the results of PCA followed the expected effects of iron oxides at the NIR and red bands, while soils with higher sand content were similarly grouped in high reflectance clusters (Dalmolin et al., 2005). In addition, the optimal number of spectral clusters was close to previous studies that similarly explored spectral segmentation methods (Viscarra Rossel et al., 2016; Demattê et al., 2019; Dotto et al., 2020). The clustering analysis applied to the regional level confirmed that reflectance patterns is a valuable proxy for soil discrimination and quantitative analysis (Nanni et al., 2004; Bellinaso et al., 2010).

Soil landscape patterns at the farm level

Field observations confirmed that SYSI can express the irregular soil variations across the landscape (Figure 6). It is worth noting that there is a gain of information from the regular satellite images or aerial photographs, which is increased by the terrain attributes and completed by a bare soil composite. However, the bare soil images do not give the last word, as the final representations of soil types are better achieved when merging all the available information. The detailed soil maps represented in figure 6 were delineated by integrating field observations, terrain attributes, and SYSI.

If the bare soil technique were not used, as demonstrated in figure 6, many places could probably not be detected by the pedologist in the field. This happens because the pedologist uses his/her tacit knowledge with terrain inference, while the bare soil image can produce a synoptic view about soil surface reflectance. In turn, the main limitation is that SYSI only represents the surface variation, and for some specific soil types, the subsurface can be contrasting. Thus, soils with textural gradient, such as Lixisols or Acrisols, could not be detected only with surface data.

Another point that can be discussed is that the bare soil images can provide information about soil variability before going to the field, assisting in this case, to the allocation of more assertive observations. This can save time, resources and improve the representation of soil variability of the final maps. Some soil spots revealed in figure 6 demonstrated that the surface variations are irregular and not detected by terrain patterns. Thus, the illustrations of figure 6 and figure 3 of SD confirm that smarter sampling strategies using bare soil images or derived products could improve the soil survey and soil mapping.

Some considerations regarding the PronaSolos

Brazil, as a continental country, requires detailed soil information for its sustainable development. In this aspect, PronaSolos has gained attention and can become one of the most outstanding projects of soil survey and mapping, which will impact the future of Brazilian agriculture and other environmental sectors (Polidoro et al., 2016). Detailed and updated information on soils brings a myriad of benefits to our society. Our results reveal the potential of an effective method to depict the soil spatial variability using computational systems and EO data by producing a synthetic soil image (YSI). In this sense, capacitation courses (e.g., Universidade PronaSolos), free and open access datasets and products (e.g., febr) and software (e.g., Python, R), collaborative work and others, will be necessary to increase the assimilation of these technologies in future works (Dalmolin et al., 2020; Samuel-Rosa et al., 2020; Costa et al., 2021).

Another point that must be addressed is the integration of new technologies and products with the legacy methods of soil survey and mapping. Since the soil reflectance comes from EO satellite sensors, the observation of the soil surface across the landscape becomes synoptic. In the 1960s, EO was performed by aerial photographs that indicated the differences between soils without looking at the subsurface but making inferences

from the relief. In the 1970s, pedologists from Radam project used radar sensor on board of airplanes to relate the image results to soil variability (IBGE, 2015). Thus, the present technique based on multispectral reflectance is an evolution of these legacy methods and can be assimilated or integrated. The multitemporal reflectance of SYSI, which is closely related to several soil properties and attributes, is an effective proxy that brings direct information of clay, iron and other relatively stable soil properties.

In turn, one also shall take into consideration the subsurface layers, for sure the most challenging target for EO satellite sensors. However, it is absolutely clear that EO products associated with field information can improve soil survey and mapping. In a recent study, Rossiter et al. (2021) performed a study to evaluate the predictive soil mapping (PSM) with geographic distribution of soil properties, where some limitations were identified. Although there still exist some discrepancies between PSM and field soil survey, the authors clearly addressed that PSM is still a valuable tool. They claimed PSM is more reproducible and objective, integrating field data with environmental covariates and learning algorithms, which can at least provide a useful pre-map for planning sampling and field survey. In this sense, the pedologist can go directly to key locations previously identified by SYSI or other landscape information, gaining time and reducing potential costs. At the end, the final decision is still made by the pedologists or soil surveyors by examining the soil at the landscape.

CONCLUSIONS

A direct representation of bare soil reflectance of São Paulo State was produced, representing most of the historical and recent cropped area. The relationship of this information was confirmed with clay and Fe_2O_3 spatial distribution (surface and subsurface layer) by a satisfactory prediction performance when employed to machine learning with auxiliary spatial information (terrain). We also demonstrated the potential of bare soil image for clustering landscape units linked to soil taxonomic classes over a reference site in São Paulo. The predicted spatial clusters had a significant association with soil classes and their qualifiers, such as the main pedological class, subsurface color, soil mineralogy and soil texture. The potential of SYSI for soil survey and soil mapping was illustrated with several examples at the farm level. Therefore, this study demonstrates that bare soil images are valuable information derived from optical Earth Observation data, which can contribute to the future of soil survey and mapping in Brazil (i.e., PronaSolos).

ACKNOWLEDGMENTS

This study had financial support from the São Paulo Research Foundation (FAPESP), grants numbers 2018/21356-1, 2016/01597-9, and 2014/22262-0. The authors are also grateful to Geotechnology in Soil Science Group (<https://esalqgeocis.wixsite.com/english>).

SUPPLEMENTARY DATA

Supplementary data to this paper can be found online at https://www.rbcjournal.org/wp-content/uploads/articles_xml/1806-9657-rbcs-45-e0210080/1806-9657-rbcs-45-e0210080-suppl01.pdf.

AUTHOR CONTRIBUTIONS

Conceptualization:  José Alexandre Melo Demattê (equal) and  José Lucas Safanelli (equal).

Data curation:  José Lucas Safanelli (equal).

Formal analysis:  José Lucas Safanelli (equal).

Investigation:  Benito Roberto Bonfatti (equal) and  José Lucas Safanelli (equal).

Methodology:  Benito Roberto Bonfatti (equal),  Jorge Tadeu Fim Rosas (equal),  José Lucas Safanelli (equal),  Natasha Valadares dos Santos (equal),  Nélida Elizabet Quiñonez Silvero (equal) and  Wanderson de Sousa Mendes (equal).

Project administration:  José Alexandre Melo Demattê (lead).

Resources:  José Alexandre Melo Demattê (lead).

Software:  José Lucas Safanelli (equal),  Natasha Valadares dos Santos (equal) and  Wanderson de Sousa Mendes (equal).

Supervision:  José Alexandre Melo Demattê (lead).

Validation:  Jorge Tadeu Fim Rosas (equal),  José Lucas Safanelli (equal) and  Natasha Valadares dos Santos (equal).

Visualization:  José Alexandre Melo Demattê (equal) and  José Lucas Safanelli (equal).

Writing - original draft:  José Lucas Safanelli (lead).

Writing - review & editing:  Jorge Tadeu Fim Rosas (equal),  José Alexandre Melo Demattê (equal),  José Lucas Safanelli (equal),  Natasha Valadares dos Santos (equal),  Nélida Elizabet Quiñonez Silvero (equal) and  Wanderson de Sousa Mendes (equal).

REFERENCES

- Alvares CA, Stape JL, Sentelhas PC, Moraes Gonçalves JL, Sparovek G. Köppen's climate classification map for Brazil. *Meteorol Zeitschrift*. 2013;22:711-28. <https://doi.org/10.1127/0941-2948/2013/0507>
- Arruda GP, Demattê JAM, Chagas CS, Fiorio PR, Souza AB, Fongaro CT. Digital soil mapping using reference area and artificial neural networks. *Sci Agric*. 2016;73:266-73. <https://doi.org/10.1590/0103-9016-2015-0131>
- Bazaglia Filho O, Rizzo R, Lepsch IF, Prado H, Gomes FH, Mazza JA, Demattê JAM. Comparison between detailed digital and conventional soil maps of an area with complex geology. *Rev Bras Cienc Solo*. 2013;37:1136-48. <https://doi.org/10.1590/S0100-06832013000500003>
- Bellinaso H, Demattê JAM, Romeiro SA. Soil Spectral Library and Its Use in Soil Classification. *Rev Bras Cienc Solo*. 2010;34:861-70. <https://doi.org/10.1590/S0100-06832010000300027>
- Bonfatti BR, Demattê JAM, Marques KPP, Poppiel RR, Rizzo R, Mendes WS, Silvero NEQ, Safanelli JL. Digital mapping of soil parent material in a heterogeneous tropical area. *Geomorphology*. 2020;367:107305. <https://doi.org/10.1016/j.geomorph.2020.107305>
- Breiman L. Random forests. *Mach Learn*. 2001;45:5-32. <https://doi.org/10.1023/A:1010933404324>
- Chabrilat S, Ben-Dor E, Cierniewski J, Gomez C, Schmid T, van Wesemael B. Imaging spectroscopy for soil mapping and monitoring. *Surv Geophys*. 2019;40:361-99. <https://doi.org/10.1007/s10712-019-09524-0>
- Chagas CS, Fernandes Filho EI, Rocha MF, Carvalho Júnior W, Souza Neto NC. Avaliação de modelos digitais de elevação para aplicação em um mapeamento digital de solos. *Rev Bras Eng Agric Ambient*. 2010;14:218-26. <https://doi.org/10.1590/S1415-43662010000200014>
- Coelho FF, Giasson E, Campos AR, Tiecher T, Costa JFF, Coblinski JA. Digital soil class mapping in Brazil: a systematic review. *Sci Agric*. 2021;78:e20190227. <https://doi.org/10.1590/1678-992x-2019-0227>
- Costa EM, Ceddia MB, Santos FN, Silva LO, Rezende IPT, Fernandes DAC. Training pedologist for soil mapping: Contextualizing methods and its accuracy using the project pedagogy approach. *Rev Bras Cienc Solo*. 2021;45:e0200130. <https://doi.org/10.36783/18069657rbcs20200130>

- Dalmolin RSD, Gonçalves CN, Klamt E, Dick DP. Relationship between the soil constituents and its spectral behavior. *Cienc Rural*. 2005;35:481-9. <https://doi.org/10.1590/S0103-84782005000200042>
- Dalmolin RSD, Moura-Bueno JM, Samuel-Rosa A, Flores CA. How is the learning process of digital soil mapping in a diverse group of land use planners? *Rev Bras Cienc Solo*. 2020;44:e0190037. <https://doi.org/10.36783/18069657rbc20190037>
- Demattê JAM, Dotto AC, Paiva AFS, Sato MV, Dalmolin RSD, Araújo MSB, Silva EB, Nanni MR, ten Caten A, Noronha NC, Lacerda MPC, Araújo Filho JC, Rizzo R, Bellinaso H, Francelino MR, Schaefer CEGR, Vicente LE, Santos UJ, Sá EVBR, Menezes RSC, Souza JJLL, Abrahão WAP, Coelho RM, Grego CR, Lani JL, Fernandes AR, Gonçalves DAM, Silva SHG, Menezes MD, Curi N, Couto EG, Anjos LHC, Cedia MB, Pinheiro ÉFM, Grunwald S, Vasques GM, Marques Júnior J, Silva AJ, Barreto MCV, Nóbrega GN, Silva MZ, Souza SF, Valladares GS, Viana JHM, Terra FS, Horák-Terra I, Fiorio PR, Silva RC, Frade Júnior EF, Lima RHC, Alba JMF, Souza Junior VS, Brefin MDLMS, Ruivo MDLP, Ferreira TO, Brait MA, Caetano NR, Bringhenti I, Mendes WS, Safanelli JL, Guimarães CCB, Poppiel RR, Souza AB, Quesada CA, Couto HTZ. The Brazilian soil spectral library (BSSL): A general view, application and challenges. *Geoderma*. 2019;354:113793. <https://doi.org/10.1016/j.geoderma.2019.05.043>
- Demattê JAM, Fongaro CT, Rizzo R, Safanelli JL. Geospatial Soil Sensing System (GEOS3): A powerful data mining procedure to retrieve soil spectral reflectance from satellite images. *Remote Sens Environ*. 2018;212:161-75. <https://doi.org/10.1016/j.rse.2018.04.047>
- Demattê JAM, Safanelli JL, Poppiel RR, Rizzo R, Silvero NEQ, Mendes WS, Bonfatti BR, Dotto AC, Salazar DFU, Mello FAO, Paiva AFS, Souza AB, Santos NV, Nascimento CM, Mello DC, Bellinaso H, Gonzaga Neto L, Amorim MTA, Resende MEB, Vieira JS, Queiroz LG, Gallo BC, Sayão VM, Lisboa CJS. Bare earth's surface spectra as a proxy for soil resource monitoring. *Sci Rep*. 2020;10:4461. <https://doi.org/10.1038/s41598-020-61408-1>
- Diek S, Fornallaz F, Schaepman ME, Jong R. Barest pixel composite for agricultural areas using landsat time series. *Remote Sens*. 2017;9:1245. <https://doi.org/10.3390/rs9121245>
- Dotto AC, Demattê JAM, Rossel RAV, Rizzo R. Soil environment grouping system based on spectral, climate, and terrain data: a quantitative branch of soil series. *Soil*. 2020;6:163-77. <https://doi.org/10.5194/soil-6-163-2020>
- Fongaro C, Demattê J, Rizzo R, Safanelli JL, Mendes W, Dotto A, Vicente L, Franceschini M, Ustin S. Improvement of clay and sand quantification based on a novel approach with a focus on multispectral satellite images. *Remote Sens*. 2018;10:1555. <https://doi.org/10.3390/rs10101555>
- Gallo B, Demattê J, Rizzo R, Safanelli JL, Mendes W, Lepsch I, Sato M, Romero D, Lacerda M. Multi-temporal satellite images on topsoil attribute quantification and the relationship with soil classes and geology. *Remote Sens*. 2018;10:1571. <https://doi.org/10.3390/rs10101571>
- Garcia MGM, Brilha J, Lima FF, Vargas JC, Pérez-Aguilar A, Alves A, Campanha GAC, Duleba W, Faleiros FM, Fernandes LA, Fierz MSM, Garcia MJ, Janasi VA, Martins L, Raposo MIB, Ricardi-Branco F, Ross JLS, Filho WS, Souza CRG, Bernardes-de-Oliveira MEC, Brito Neves BB, Campos Neto MC, Christofolletti SR, Henrique-Pinto R, Lobo HAS, Machado R, Passarelli CR, Perinotto JAJ, Ribeiro RR, Shimada H. The inventory of geological heritage of the state of São Paulo, Brazil: Methodological basis, results and perspectives. *Geoheritage*. 2018;10:239-58. <https://doi.org/10.1007/s12371-016-0215-y>
- Giasson E, ten Caten A, Bagatini T, Bonfatti B. Instance selection in digital soil mapping: a study case in Rio Grande do Sul, Brazil. *Cienc Rural*. 2015;45:1592-8. <https://doi.org/10.1590/0103-8478cr20140694>
- Giasson E, Clarke RT, Inda Junior AV, Merten GH, Tornquist CG. Digital soil mapping using multiple logistic regression on terrain parameters in southern Brazil. *Sci Agric*. 2006;63:262-8. <https://doi.org/10.1590/S0103-90162006000300008>
- Instituto Brasileiro de Geografia e Estatística - IBGE. Manual técnico de pedologia. 3. ed. Rio de Janeiro: IBGE; 2015.

- Instituto de Pesquisas Tecnológicas - IPT. Mapa geomorfológico do Estado de São Paulo, escala 1:1.000.000. São Paulo: IPT;1981.
- 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: Food and Agriculture Organization of the United Nations; 2015. (World Soil Resources Reports, 106).
- Lagacherie P, McBratney AB, Voltz M. Digital soil mapping: an introductory perspective. Amsterdam: Elsevier Science; 2006.
- McBratney A, Mendonça Santos M, Minasny B. On digital soil mapping. *Geoderma*. 2003;117:3-52. [https://doi.org/10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4)
- Mendes WS, Medeiros Neto LG, Demattê JAM, Gallo BC, Rizzo R, Safanelli JL, Fongaro CT. Is it possible to map subsurface soil attributes by satellite spectral transfer models? *Geoderma*. 2019;343:269-79. <https://doi.org/10.1016/j.geoderma.2019.01.025>
- Minasny B, McBratney AB. Digital soil mapping: A brief history and some lessons. *Geoderma*. 2016;264:301-11. <https://doi.org/10.1016/j.geoderma.2015.07.017>
- Moura-Bueno JM, Dalmolin RSD, Horst-Heinen TZ, Cancian LC, Schenato RB, Dotto AC, Flores CA. Prediction of soil classes in a complex landscape in Southern Brazil. *Pesq Agropec Bras*. 2019;54:e00420. <https://doi.org/10.1590/s1678-3921.pab2019.v54.00420>
- Nanni MR, Demattê JAM, Fiorio PR. Análise discriminante dos solos por meio da resposta espectral no nível terrestre. *Pesq Agropec Bras*. 2004;39:995-1006. <https://doi.org/10.1590/S0100-204X2004001000007>
- Carvalho CCN, Nunes FC, Antunes MAH, Nolasco MC. Soil surveys in Brazil and perspectives in digital soil mapping. *Soil Horizons*. 2015;56:1-9. <https://doi.org/10.2136/sh14-01-0002>
- Oliveira JB, Camargo MN, Rossi M, Calderano Filho B. Mapa pedológico do estado de São Paulo, escala 1:500.000. Campinas: Instituto Agrônomo de Campinas; 1999.
- Polidoro JC, Mendonça-Santos MDL, Lumberras JF, Coelho MR, Carvalho Filho A, Motta PEF, Carvalho Junior W, Araujo Filho JC, Curcio GR, Correia JR, Martins EDS, Spera ST, Oliveira SRDM, Bolfe EL, Manzatto CV, Tosto SG, Venturieri A, Sa IB, Oliveira VA, Shinzato E, Anjos LHC, Valladares GS, Ribeiro JL, Medeiros PSC, Moreira FMDS, Silva LSL, Sequinatto L, Aglio MLD, Dart RDO. Programa Nacional de Solos do Brasil (PronaSolos). Rio de Janeiro: Embrapa Solos; 2016.
- Poppiel RR, Lacerda MPC, Safanelli JL, Rizzo R, Oliveira MP, Novais JJ, Demattê JAM. Mapping at 30 m resolution of soil attributes at multiple depths in midwest Brazil. *Remote Sens*. 2019a;11:2905. <https://doi.org/10.3390/rs11242905>
- Poppiel RR, Lacerda MPC, Demattê JAM, Oliveira MP, Gallo BC, Safanelli JL. Pedology and soil class mapping from proximal and remote sensed data. *Geoderma*. 2019b;348:189-206. <https://doi.org/10.1016/j.geoderma.2019.04.028>
- Raij B van, Andrade JCC, Cantarella H, Quaggio JAA. Análise química para avaliação de solos tropicais. Campinas: Instituto Agrônomo de Campinas; 2001.
- Rizzo R, Demattê JAM, Lepsch IF, Gallo BC, Fongaro CT. Digital soil mapping at local scale using a multi-depth Vis-NIR spectral library and terrain attributes. *Geoderma*. 2016;274:18-27. <https://doi.org/10.1016/j.geoderma.2016.03.019>
- Rizzo R, Medeiros LG, Mello DC, Marques KPP, Mendes WS, Quiñonez Silvero NE, Dotto AC, Bonfatti BR, Demattê JAM. Multi-temporal bare surface image associated with transfer functions to support soil classification and mapping in southeastern Brazil. *Geoderma*. 2020;361:114018. <https://doi.org/10.1016/j.geoderma.2019.114018>
- Rogge D, Bauer A, Zeidler J, Mueller A, Esch T, Heiden U. Building an exposed soil composite processor (SCMaP) for mapping spatial and temporal characteristics of soils with Landsat imagery (1984-2014). *Remote Sens Environ*. 2018;205:1-17. <https://doi.org/10.1016/j.rse.2017.11.004>
- Rossi M. Mapa pedológico do Estado de São Paulo: revisado e ampliado. São Paulo: Instituto Florestal; 2017.

- Rossiter DG, Poggio L, Beaudette D, Libohova Z. How well does Predictive Soil Mapping represent soil geography? An investigation from the USA. *Soil Discuss* [Preprint]. 2021. <https://doi.org/10.5194/soil-2021-80>
- Safanelli JL, Chabrillat S, Ben-Dor E, Demattê JAM. Multispectral models from bare soil composites for mapping topsoil properties over Europe. *Remote Sens*. 2020a;12:1369. <https://doi.org/10.3390/rs12091369>
- Safanelli JL, Demattê JAM, Chabrillat S, Poppiel RR, Rizzo R, Dotto AC, Silvero NEQ, Mendes WS, Bonfatti BR, Ruiz LFC, ten Caten A, Dalmolin RSD. Leveraging the application of Earth observation data for mapping cropland soils in Brazil. *Geoderma*. 2021;396:115042. <https://doi.org/10.1016/j.geoderma.2021.115042>
- Safanelli JL, Poppiel RR, Ruiz LFC, Bonfatti BR, Mello FAO, Rizzo R, Demattê JAM. Terrain analysis in google earth engine: A method adapted for high-performance global-scale analysis. *ISPRS Int J Geo-Information*. 2020b;9:400. <https://doi.org/10.3390/ijgi9060400>
- Samuel-Rosa A, Dalmolin RSD, Miguel P. Building predictive models of soil particle-size distribution. *Rev Bras Cienc Solo*. 2013;37:422-30. <https://doi.org/10.1590/S0100-06832013000200013>
- Samuel-Rosa A, Dalmolin RSD, Moura-Bueno JM, Teixeira WG, Alba JMF. Open legacy soil survey data in Brazil: geospatial data quality and how to improve it. *Sci Agric*. 2020;77:e20170430. <https://doi.org/10.1590/1678-992x-2017-0430>
- Santos HG, Jacomine PKT, Anjos LHC, Oliveira VA, Lumberras JF, Coelho MR, Almeida JA, Araújo Filho JC, Oliveira JB, Cunha TJF. Sistema brasileiro de classificação de solos. 5. ed. rev. ampl. Brasília, DF: Embrapa; 2018.
- Sarmiento EC, Giasson E, Weber EJ, Flores CA, Hasenack H. Disaggregating conventional soil maps with limited descriptive data: A knowledge-based approach in Serra Gaúcha, Brazil. *Geoderma Reg*. 2017;8:12-23. <https://doi.org/10.1016/j.geodrs.2016.12.004>
- ten Caten A, Dalmolin RSD, Mendonça-Santos ML, Giasson E. Mapeamento digital de classes de solos: características da abordagem brasileira. *Cienc Rural*. 2012;42:1989-97. <https://doi.org/10.1590/S0103-84782012001100013>
- ten Caten A, Dalmolin RSD, Pedron FA, Santos MLM. Componentes principais como preditores no mapeamento digital de classes de solos. *Cienc Rural*. 2011;41:1170-6. <https://doi.org/10.1590/S0103-84782011000700011>
- Terra FS, Demattê JAM, Rossel RAV. Spectral libraries for quantitative analyses of tropical Brazilian soils: Comparing vis-NIR and mid-IR reflectance data. *Geoderma*. 2015;255-256:81-93. <https://doi.org/10.1016/j.geoderma.2015.04.017>
- United States Geological Survey - USGS. Landsat 4-7 Surface reflectance code LEDAPS product guide. Virginia: USGS; 2018a [cited 2019 Mar 11]. Available from: <https://www.usgs.gov/media/files/landsat-4-7-surface-reflectance-code-ledaps-product-guide>
- United States Geological Survey - USGS. Landsat 8 Surface reflectance code LaSRC product guide. Virginia: USGS; 2018b [cited 2019 Mar 11]. Available from: <https://www.usgs.gov/media/files/landsat-8-surface-reflectance-code-lasrc-product-guide>
- Vasques GM, Coelho MR, Dart RO, Oliveira RP, Teixeira WG. Mapping soil carbon, particle-size fractions, and water retention in tropical dry forest in Brazil. *Pesq Agropec Bras*. 2016;51:1371-85. <https://doi.org/10.1590/s0100-204x201600900036>
- Viscarra Rossel RA, Behrens T, Ben-Dor E, Brown DJ, Demattê JAM, Shepherd KD, Shi Z, Stenberg B, Stevens A, Adamchuk V, Aichi H, Barthès BG, Bartholomeus HM, Bayer AD, Bernoux M, Böttcher K, Brodský L, Du CW, Chappell A, Fouad Y, Genot V, Gomez C, Grunwald S, Gubler A, Guerrero C, Hedley CB, Knadel M, Morrás HJM, Nocita M, Ramirez-Lopez L, Roudier P, Campos EMR, Sanborn P, Sellitto VM, Sudduth KA, Rawlins BG, Walter C, Winowiecki LA, Hong SY, Ji W. A global spectral library to characterize the world's soil. *Earth-Sci Rev*. 2016;155:198-230. <https://doi.org/10.1016/j.earscirev.2016.01.012>
- Viscarra Rossel RA, Chappell A, De Caritat P, McKenzie NJ. On the soil information content of visible-near infrared reflectance spectra. *Eur J Soil Sci*. 2011;62:442-53. <https://doi.org/10.1111/j.1365-2389.2011.01372.x>

Wolski MS, Dalmolin RSD, Flores CA, Moura-Bueno JM, ten Caten A, Kaiser DR. Digital soil mapping and its implications in the extrapolation of soil-landscape relationships in detailed scale. *Pesq Agropec Bras.* 2017;52:633-42. <https://doi.org/10.1590/s0100-204x2017000800009>