Mapping clay fraction oxides in Brazil using Earth observation strategy

Jorge Tadeu Fim Rosas [a], José A.M. Demattê [a,d,*] , Nícolas Augusto Rosin [a] ,
Raul Roberto Poppiel [a,d], Nélida E.Q. Silvero [a], Merilyn Taynara Accorsi Amorim [a],
Heidy S. Rodríguez-Albarracín [a], Letícia Guadagnin Vogel [a] , Bruno dos Anjos Bartsch [a],
José João Lelis Leal de Souza [b], Lucas de Carvalho Gomes [c], Danilo César de Mello [b]

[a] *Department of Soil Science, Luiz de Queiroz College of Agriculture (ESALQ), University of São Paulo (USP), Piracicaba, São Paulo 13418-900, Brazil*
[b] *Department of Soil Science, Federal University of Viçosa, Viçosa, Minas Gerais 36570-900, Brazil*
[c] *Department of Agroecology, Aarhus University, Tjele 8830, Denmark*
[d] *Center for Carbon Research in Tropical Agriculture (CCARBON) - University of São Paulo, Piracicaba, São Paulo, Brazil*

ARTICLE INFO

ABSTRACT

The major oxides in the clay fraction of tropical soils are iron (Fe2O3), aluminum (Al2O3), and silicon (SiO2) oxides, which are responsible for the soil's capacity to provide multiple ecosystem services. Therefore, they are used to classify the soils into different pedological classes. Despite their importance of these oxides, quantifying them on a large scale presents significant challenges. The most common method is laboratory sulfuric acid digestion, which is expensive, complex, and environmentally detrimental. To overcome these issues and provide faster results, we developed a satellite technique associated with machine learning (ML) to map Fe_2O_3, Al_2O_3, and SiO_2 in Brazil in all agricultural areas at 30 m resolution. Additionally, we tested the feasibility of the generated maps to infer soil weathering, and assist in the construction of pedological maps. A dataset, comprising 5,330 sites (0–20 cm and 80–100 cm) across all 27 states was employed in prediction. Six spectral variables obtained from the historical Landsat series (bare soil) and seven terrain attributes derived from a digital elevation model were employed to generate the Fe_2O_3, Al_2O_3, and SiO_2 maps, using the Random Forest algorithm. The predicted maps of oxides covered nearly 3.48 million km^2 (~40 % of the national territory). The best predictions were observed for Fe_2O_3 in the 0–20 cm layer (RMSE = 49.8 $g.kg^{-1}$, RPIQ = 1.82, and R^2 = 0.62), while the worst predictions were for SiO_2 in the 80–100 cm layer (RMSE = 65.3 $g.kg^{-1}$, RPIQ = 1.50 and R^2 = 0.22). It was possible to infer soil weathering using the Ki index. Despite the models not showing such high R^2 values, the results are aligned with legacy maps, highly weathered soils were observed in the plateaus of the Cerrado biome, while younger soils were observed in the arid Caatinga biome and waterlogged soils in the Pantanal biome. The generated maps also demonstrated a high potential for grouping pedological soil classes. They also revealed a relationship between oxide contents and the NDVI of sugarcane crops, indicating potential applications in crop management. Moreover, this satellite-based technique, supported by ML, presents a plausible approach to predict oxide fraction at high spatial resolution for large areas.

1. Introduction

Soil is involved in several ecosystem services (ES) that are of utmost importance for the maintenance of life on the planet (McBratney et al., 2014), such as provisioning (food, fiber, and timber production), regulating (climate, flood, and water regulation), cultural and supporting (nutrient cycling, soil formation) (Silvero et al., 2023). This ability is determined by various intrinsic soil properties, such as texture, organic carbon content, depth, bulk density, available water, and cation exchange capacity (Adhikari and Hartemink, 2016; Hewitt et al., 2015). These properties have different spatial distributions and can vary considerably across the Earth's crust, influencing the soil's capacity to supply ecosystem services (Adhikari and Hartemink, 2016). In tropical soils, characterized by a high degree of weathering, the clay fraction oxides are distinguished in provisioning ES (Kirsten et al., 2021; Silva et al., 2021). Soil oxides play several roles in the soil matrix, influencing the maintenance of soil structure, potential for organic carbon and water retention, the pH level, and availability of nutrients for plants, among

---

others (Kirsten et al., 2021; Ukabiala et al., 2021). Thus, quantifying these oxides can help in understanding the ability of soil to provide ES.

The clay fraction (particles <0.002 mm) major oxides of tropical soils are iron oxides (Fe_2O_3), aluminum oxides (Al_2O_3), and silicon oxides (SiO_2) (Schaefer et al., 2008), and they are used in the Brazilian soil classification system (SiBCS). The Fe_2O_3 content is used in SiBCS for classification at the third level of the classification hierarchy, while Al_2O_3 and SiO_2 are used to measure the degree of soil weathering (Embrapa, 2018). Despite the influence of these oxides on the soil, there is a lack of effort to map them in the tropical regions of the world. Principally because, laboratory methodologies are expensive, time-consuming, and have a high potential for environmental pollution (its quantification is performed by using acid digestion with strong acids such as sulfuric acid), making it difficult to obtain information on these oxides for large areas, with high spatial resolutions (Silva et al., 2020). However, despite the challenge of mapping the contents of these oxides, innovative techniques for digital soil mapping (DSM) have the potential to create high-resolution maps detailing oxide contents, even with low sample densities and challenging conditions of complexity inherent in soil systems (de Mendes et al., 2022; Safanelli et al., 2021b).

The DSM approach involves deploying robust mathematical models capable of predicting a given soil property based on environmental covariates (Hengl et al., 2017; Vaysse and Lagacherie, 2015). Due to advances in computational techniques, it is currently possible to implement robust machine learning (ML) algorithms that aim to generate predictions with enhanced accuracy and precision (Wadoux et al., 2020). The Random Forest (RF) algorithm has been the most widely used in DSM studies (Padarian et al., 2020). RF creates robust models composed of multiple decision trees with low correlation among each other. Due to these characteristics, RF is able topredict with high accuracy and a minimal overfitting rate in the models (Sheykhmousa et al., 2020; Wadoux et al., 2020).

The final quality of DSM outputs depends on the environmental covariates used for the prediction of specific soil properties. The environmental covariates express soil formation factors and represent parts of the physical, chemical and organisms processes that govern the spatial variation of soil (McBratney et al., 2003; Wadoux et al., 2020). Among the environmental covariates, the most commonly used ones are digital elevation models, vegetation indices, climate covariates, geological maps, and surface reflectance obtained by satellites (Lamichhane et al., 2019; Ma et al., 2019). In addition to these covariates, recent studies propose using environmental covariates that represent bare soil reflectance (Roberts et al., 2019; Rosin et al., 2023; Safanelli et al., 2021b). Despite the great potential of bare soil reflectance in DSM, obtaining this reflectance is a major challenge primarily due to the persistent presence of vegetation covering soil surfaces throughout the year. To solve this problem, Demattê et al. (2018) developed an algorithm (Geospatial Soil Sensing System – GEOS3) capable of retrieving bare soil reflectance from a historical series of Landsat images by aggregating the pixels that were exposed at least once throughout the historical series into a synthetic soil image (SYSI).

Several studies worldwide have demonstrated the potential of exposed soil reflectance for soil attribute mapping (Gasmi et al., 2021; Rizzo et al., 2020; Roberts et al., 2019; Silvero et al., 2021a; Yang et al., 2020). The prediction of soil attributes based on reflectance has been consolidated through the use of laboratory spectroscopy (Barra et al., 2021). Soil attributes concentrations alters their spectral signature, at specific wavelengths or across the entire electromagnetic spectrum, such as the iron oxide content altering reflectance in the red region, or clay content altering reflectance intensity across the spectrum (Ackerson et al., 2015; Demattê et al., 2007). This characteristic makes the reflectance of exposed soil when applied in DSM, one of the most important environmental covariate in predicting soil attributes (Poppiel et al., 2019; Rosin et al., 2023).

We expect that the reflectance of topsoil measured by satellite will be related to the contents of Fe_2O_3, Al_2O_3, and SiO_2 oxides present in the clay fraction, as these oxides influence electronic transitions and molecular vibrations in the 350–2500 nm spectral range. Thus, we aimed to: (i) quantify and map the contents of Fe_2O_3, Al_2O_3, and SiO_2 using remote sensing data integrated with ML techniques at both surface and subsurface soil depths; and (ii) evaluate whether the resulting oxide maps can be used to infer the degree of weathering in Brazilian agricultural soils, support soil classification, and assist in agricultural crop management. The high-resolution maps (30 m of spatial resolution) generated for all agricultural areas of Brazil are expected to support researchers, farmers, and consultants in better understanding soil dynamics and guiding more informed land use decisions.

2. Materials and methods

2.1. Study area

The study area is represented by the entire agricultural territory of Brazil. Brazil is a country with large dimensions (approximately 8.5 million km^2) and high soil variability. This soil variability is due to the formation factors, mainly, the climate, the time of formation, the parent material, and the topography (Schaefer et al., 2023). Brazil's climate varies across the territory between tropical, subtropical and semiarid (Alvares et al., 2013). The time formation of parent material that gave rise to Brazilian soils primarily occurred during the Cenozoic, Mesozoic, Neoproterozoic, and Paleoproterozoic eras (Schaefer et al., 2023). Regarding the parent material, the majority of rocks are sedimentary, formed by sediments with diverse origins, such as weathering of the Andes (Amazon basin), ocean floor (São Francisco craton region), and desert sediment as in the Bauru and Botucatu formations (Schaefer et al., 2023). In addition to sedimentary rocks, a large part of Brazil's geology is composed of volcanic and plutonic rocks (de Alkmim, 2015). In the region encompassing the states of Paraná, São Paulo, Rio Grande do Sul, Minas Gerais, Goiás, and Mato Grosso do Sul, there is a large mass of basaltic rocks formed from basaltic eruptions (de Alkmim, 2015). These rocks are extrusive and contain higher levels of iron and easily weatherable minerals (Schaefer et al., 2008). Along the Brazilian coastline, many plutonic rocks were formed underground (i.e., intrusives) during the continental drift (de Alkmim, 2015). These rocks possess highly crystallized minerals like quartz. Another factor of formation with high variability in Brazil is the topography. Many mountains are present in coastal regions, while the interior of the country has large plateaus that favor soil weathering (Schaefer et al., 2023). There are also extensive floodplain basins, such as the Pantanal (Couto et al., 2023). All of these variations in soil formation factors contribute to the existence of soils with different levels of weathering and, consequently, different levels of oxides that were mapped in this study.

2.2. Soil observations

The soil observations were derived from two sources: i) publicly available national soil profile databases with nearly 9119 sites (Samuel-Rosa et al., 2019), and ii) the Brazilian Soil Spectral Library with nearly 45,000 sites (Demattê et al., 2019). The final database containing the soil observations used in this study was consolidated and modified in four different steps. The first step was to eliminate the sites with inaccurate coordinates (i.e., accuracy less than 100 m). In the second step, all the sites with information on at least one of the Fe_2O_3, SiO_2, and Al_2O_3 oxides were filtered. The pseudo-total concentrations of these oxides were obtained by sulfuric acid digestion, as proposed by Vettori (1969), with modifications proposed by Raij and Valadares (1974). This extraction method can access the oxides in the secondary minerals. Therefore, the extracted oxides mostly reflect the clay fraction of the soil.

Once the second step was completed, a total of 5330 sites were obtained, covering the entire Brazilian territory. These sites were submitted to a third step, which consisted of detecting those located in the areas

covered by SYSI. In the fourth step, the sites with irregular sampling (i. e., soil profiles), were subjected to interpolation along the profile to standardize the depths, making them suitable for our study, while the sites with 0–20 cm and 80–100 cm depths were maintained without interpolation. The interpolation method was the quadratic spline function implemented in the GSIF R package (Hengl and Macmillan, 2019). The use of the quadratic spline function for the interpolation of soil attributes along the profile was proposed by Bishop et al. (1999), who showed that, in tests comparing it with other similar methods, it outperformed the alternatives; since then, it has been widely used in soil science studies s (Ellili-Bargaoui et al., 2020; Sulaeman et al., 2013; Tayebi et al., 2021). After interpolation, we obtained layers of 20 cm, starting at the soil surface to 100 cm depth. The 0–20 and 80–100 cm depths were used in this study because: 1) There are more observation points in these layers (because some points are sampled at these depths; 2) We wanted to show the differences between the surface (A horizon) and the subsurface layers (B horizon), being the intermediary horizons less important for pedology studies in this case, where the SiBCS is considered and 3) The digital soil mapping for Brazilian territory is highly time-consuming. After all the steps were taken, the database was ready to be used. In the final database, each oxide had a different number of sites. In addition, the sampling density was also different for each depth used (Table 1).

2.3. Environmental covariates

The RF algorithm was used to fit predictive models for Fe_2O_3, Al_2O_3, and SiO_2 oxides using the synthetic soil image (SYSI) and terrain attributes as covariates. A flowchart of the prediction process is shown in Fig. 1.

The concentrations of the oxides are directly associated with soil formation factors, which was one of the criteria for choosing the environmental covariates. Despite the great importance of all soil formation factors, only the relief factor was considered. It was impossible to use factors such as parent material and time due to low spatial resolution or lack of information. The climatic factor was not included because the covariates available in WordClim have a spatial resolution of 1 km. Additionally, these variables exhibit artifacts across the Brazilian territory, such as unrealistic geometric patterns and longitudinal lines. Furthermore, for the organism factor, vegetation indices calculated from spectral information of the surface are commonly employed. According to Lamichhane et al. (2019), covariates of this nature exhibit an indirect relationship with soil properties. Consequently, we opted not to incorporate them into our analysis. To represent the relief factor, we used environmental covariates that were previously used to predict soil attributes with satisfactory results in Safanelli et al (2021a). These terrain attributes were as follows: slope, altitude, north and east slope, horizontal curvature, vertical curvature, and a relief shape index. These variables were based on a digital elevation model (DEM) from the Shuttle Radar Topography Mission (SRTM) with 30 m spatial resolution (Farr and Kobrick, 2000). To calculate the environmental variables, the Terrain Analysis in Google Earth Engine (TAGEE) algorithm proposed by

Safanelli et al. (2020b) was used. The individual information on the environmental covariates derived from the DEM is shown in Table 2.

For choosing the environmental covariates, we also adopted the bare soil spectral signature. Several studies have demonstrated that these oxides alter the reflectance intensity of soils (Demattê et al., 2009; Terra et al., 2018). Thus, it is possible to estimate their concentrations from spectral information. To obtain the spectral variables, the GEOS3 algorithm (Demattê et al., 2020, Demattê et al., 2018), developed within the Google Earth Engine, was used to generate the multispectral composite of bare soil and the frequency of soil exposure from 1982 to 2020, using the Landsat Surface Reflectance Collections at 30 m resolution (USGS, 2021a; 2021b). GEOS3 is a data mining algorithm that extracts soil features from the historical collection of satellite images and aggregates the spatially sparse discovered soil fragments into a SYSI. The SYSI is the reflectance image, composed of six spectral bands (blue – band 1, green – band 2, red – band 3, NIR – band 4, SWIR1 – band 5, and SWIR2 – band 6).

A set of rules was used to identify bare ground pixels in satellite images, based on spectral indices coupled with quality assessment bands to remove clouds, cloud shadows, inland water, photosynthetic vegetation, and non-photosynthetic vegetation (e.g., harvest residues).. A pixel was considered bare soil when it had Normalized Difference Vegetation Index (NDVI) values between −0.05 and 0.30 (masking green vegetation) and Normalized Burn Ratio 2 (NBR2) values between −0.15 and 0.15 (masking crop residues). Pixels detected as bare soil were selected to compose the multitemporal collection. Each pixel that makes up the SYSI refers to the median value calculated among all pixels in the multitemporal collection for a given position. Thus, GEOS3 produced an almost continuous representation of the topsoil reflectance, increasing the mapped area of the soil surface by combining and averaging the estimates frommultitemporal measurements. Detailed information on GEOS3, spectral indices, and sensitivity analysis can be found in previous studies (Demattê et al., 2020; Fongaro et al., 2018; Gallo et al., 2018; Safanelli et al., 2020a). All six SYSI bands were used as environmental variables to predict oxides; they are shown in Table 2.
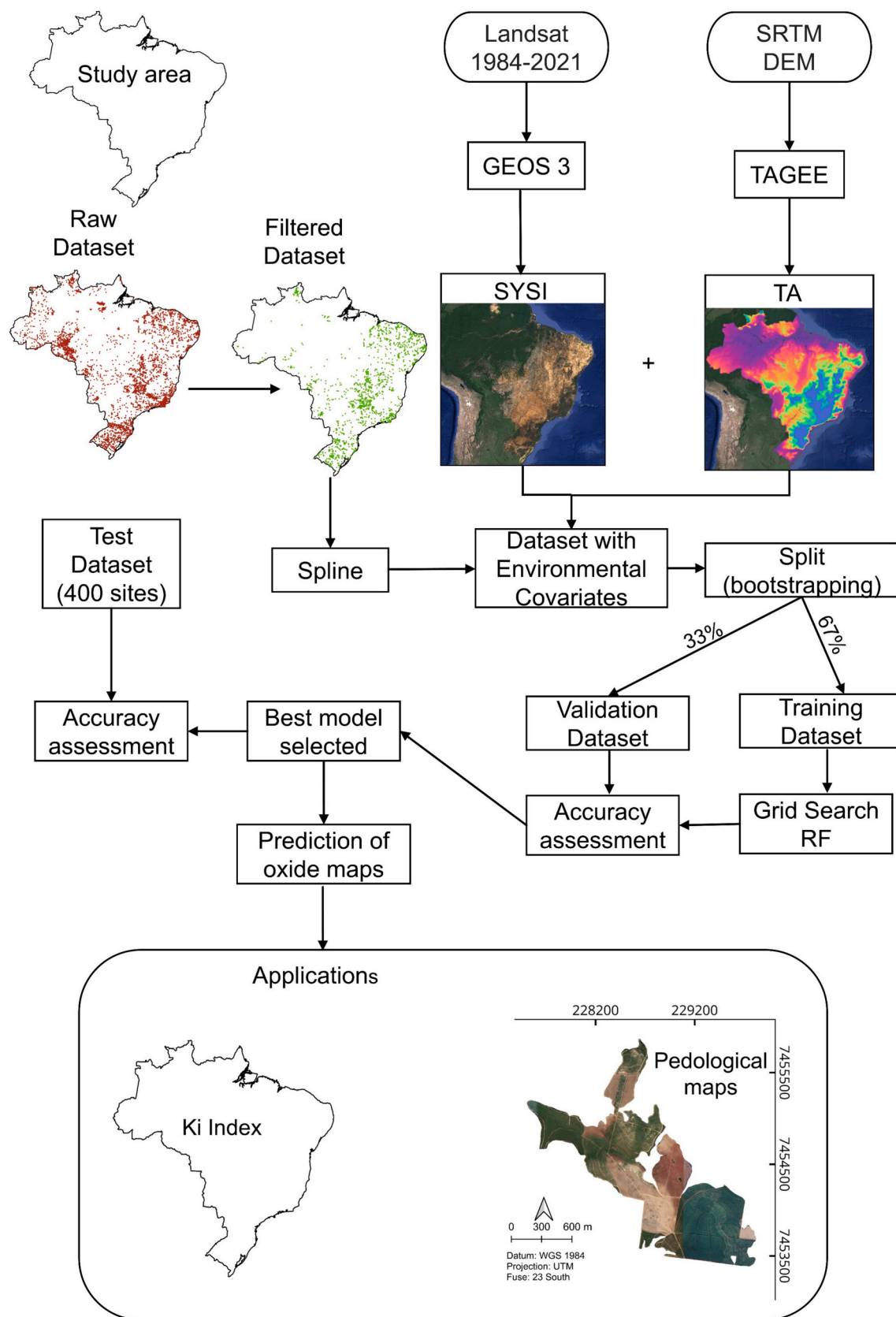
2.4. Soil clay fraction major oxides prediction by random forest

In this study, a quantile regression forest (QRF) ML algorithm, an extension of the RF algorithm, was implemented to map the content of major oxides in the soil clay fraction across Brazilian agricultural areas. The RF, proposed by Breiman (2001), is a ML algorithm that operates by constructing a set of independent decision trees. Each tree is trained on a random sample of the training data. During prediction, each tree generates an estimate, and the final prediction is determined by combining the estimates from all the trees (Breiman, 2001). The QRF proposed by Meinshausen (2006) provides a set of prediction quantiles using an empirical probability distribution rather than just a median value. These quantiles can be used to determine the prediction interval (PI), allowing inference about prediction uncertainties.

2.4.1. Model tuning, performance, and validation

The selected environmental covariates were used as independent variables to obtain soil oxide prediction models. At each soil sampling point, the values of the environmental covariates were collected using the bilinear sampling method, which considers some neighboring pixels within a radius of 100 m around the coordinate point. The bilinear sampling aims to reduce the effects of noisy pixels and the effect of coordinates with lower accuracy.

To optimize the best combination of hyperparameters for RF, a grid search procedure was performed to improve prediction accuracy. The following values were tested for each hyperparameter: number of trees in the forest (FS): 30, 60, 100, 200, and 500, number of random predictors tested at splits of each tree (nRP): 3, 5, 9, 11, and 13, and minimum number of samples at the tree (minSL): 10, 20, 30, 40, 50, 100, 200, and 500.

Table 1
Number of soil observations for each oxide throughout Brazil and in areas covered by SYSI.

Oxide	Layer	N° observations	
		Brazil	SISY Coverage
Fe_2O_3	0–20 cm	5211	2715
	80–100 cm	3671	1874
Al_2O_3	0–20 cm	4551	2253
	80–100 cm	3262	1608
SiO_2	0–20 cm	4750	2380
	80–100 cm	3393	1697

SYSI: Synthetic Soil Image.

Fig. 1. Flowchart of the soils clay fraction major oxides prediction and its possible applications tested. GEOS3: Geospatial Soil Sensing System; SYSI: Synthetic Soil Image; DEM: Digital Elevation Model; STRM: Shuttle Radar Topography Mission; TAGEE: Terrain Analysis in Google Earth Engine; TA: Terrain Attributes; RF: Random Forest; Ki: Weathering index.

Table 2
Environmental covariates used for prediction of major oxides.

Environmental covariate	Source (spatial resolution)	References
SYSI blue (450–520 nm)	Landsat collection (30 m)	(Demattê et al., 2018)
SYSI green (520–600 nm)	Landsat collection (30 m)	(Demattê et al., 2018)
SYSI red (630–690 nm)	Landsat collection (30 m)	(Demattê et al., 2018)
SYSI NIR (760–900 nm)	Landsat collection (30 m)	(Demattê et al., 2018)
SYSI SWIR1 (1550–1750 nm)	Landsat collection (30 m)	(Demattê et al., 2018)
SYSI SWIR2 (2080–2350 nm)	Landsat collection (30 m)	(Demattê et al., 2018)
Elevation	SRTM (30 m)	(Farr and Kobrick, 2000)
Slope	TAGEE/SRTM (30 m)	(Safanelli et al., 2020b)
Northness	TAGEE/SRTM (30 m)	(Safanelli et al., 2020b)
Eastness	TAGEE/SRTM (30 m)	(Safanelli et al., 2020b)
Horizontal curvature	TAGEE/SRTM (30 m)	(Safanelli et al., 2020b)
Vertical curvature	TAGEE/SRTM (30 m)	(Safanelli et al., 2020b)
Terrain shape index	TAGEE/SRTM (30 m)	(Safanelli et al., 2020b)

SYSI: Synthetic Soil Image; TAGEE: Terrain Analysis in Google Earth Engine; SRTM: Shuttle Radar Topography Mission.

Bootstrapped sampling was used to determine the training dataset for the model, while the validation dataset was defined by the remaining samples (out-of-bag) that were not selected in each bootstrapping. Bootstrapping is a method that randomly samples observations from the original dataset to create a new dataset for training prediction models. This training dataset is generally composed of approximately 67 % of the original dataset. We applied the bootstrap method up to 500 times, which corresponds to the maximum number of trees tested for the random forest.

The optimal combination (FS, nPR, and minSL) with the smallest Root Mean Squared Error (RMSE) for the validation set was selected for each soil attribute. The accuracy of the developed models was evaluated by the coefficient of determination (R^2), RMSE, and Ratio of Performance to InterQuartile distance (RPIQ) of the calibration, validation, and test data sets. The calculation of these parameters was performed according to the equations (1) to (5).

$$RMSE = \sqrt{\left(\frac{\sum_{i=1}^{n}(\widehat{y}_i - y_i)^2}{n}\right)} \tag{1}$$

$$R^2 = 1 - \frac{SS_{residuals}}{SS_{total}} \tag{2}$$

$$SS_{residuals} = \sum_{i=1}^{n}(\widehat{y}_i - y_i)^2 \tag{3}$$

$$SS_{total} = \sum_{i=1}^{n}(y_i - \overline{y})^2 \tag{4}$$

$$RPIQ = \frac{IQR_y}{RMSE_{\overline{y}}} \tag{5}$$

where y is the vector of measured values, $\widehat{y}$ is the vector of predicted values, $\overline{y}$ is the mean of vector y, n is the number of soil observations and, IQR is the interquartile range.

2.4.2. Evaluation of predictions

The evaluation of predictions was conducted following the selection of predictive models. For this evaluation, 400 geo-referenced sites covering the entire Brazilian territory were provided by private sources, exclusively for map validation purposes, with all site data de-identified. Each site consisted of two soil samples, one for each mapped layer. The sulfuric acid digestion method described in Section 2.2 was used to determine Fe_2O_3, Al_2O_3, and SiO_2. These samples comprised a new database called the test database, to which environmental covariates were also added, as described previously. The test database was not utilized during any of the training and validation steps. The selected predictive models were applied to this new database to predict the oxide contents. After the predictions, accuracy was assessed using the R^2, RMSE, and RPIQ metrics.

2.4.3. Soil clay fraction major oxides and uncertainty mapping

The best combination of FS, nPR, and minSL, was used to predict the major soil oxides for the study area. The raster files corresponding to the environmental covariates were used as input data in Fe_2O_3, Al_2O_3, and SiO_2 models, and the output consisted of the respective maps. The final estimated value for each pixel in the final maps was formed by the aggregated mean of the estimations obtained by each regression tree (Equation (6)).

$$\widehat{f}_p = \frac{1}{B}\sum_{b=1}^{B}f_b(X_p) \tag{6}$$

where X p is the set of environmental features for pixel; B is the forest size determined by the number of bootstrapping samples b; f_b is the regression tree fitted to each bootstrapped sample b; $\widehat{f}_p$ is the mean of the pixel.

The uncertainties of the predictions were inferred by calculating the ratio of the interquartile range over the median (prediction interval ratio, PIR) proposed by Poggio et al. (2021). The main reason for choosing PIR was its ability to compare models for predicting different oxides at different depths. The 90th prediction interval (PI90) was adopted in this study. The PIR maps were derived (Equation (7) from the maps corresponding to the quantiles of 0.05 (q0.05), 0.5 (q0.5), and 0.95 (q0.95), obtained through QRF. The PIR value ranges from 0 to 1, where a higher value indicates greater uncertainty in the prediction.

$$PIR = \frac{q0.95 - q0.05}{q0.50} \tag{7}$$

where q0.95 is the 95 % quantile of the modeled predictions; q0.05 is the 5 % quantile of the modeled predictions and, q0.5 is the 50 % quantile of the modeled predictions.

2.5. Oxide maps interpretation

Brazilian soils are typically highly weathered and exhibit advanced stages of pedogenesis, primarily due to the prevailing tropical and subtropical climates that have shaped the landscape over geological timescales. Much of the parent material is also rich in iron-bearing minerals. As a result of intense weathering and pedogenetic processes, particularly ferralitization, latosolization, and clay illuviation, adaptations to international soil classification systems, such as the WRB and USDA Soil Taxonomy, have been necessary, especially with regard to iron oxide (Fe_2O_3) content. To support these classifications, Fe_2O_3 content determined through sulfuric acid digestion has been categorized into four groups: (i) hypoferric (<80 g kg^{-1}), (ii) mesoferric (80 to <180 g kg^{-1}), (iii) ferric (180 to <360 g kg^{-1}), and (iv) hyperferric (≥360 g kg^{-1}). These thresholds are critical for distinguishing soil types and understanding the genesis, evolution, and functional properties of tropical and subtropical soils, as well as guiding land use and management strategies. This Fe_2O_3 based classification is especially relevant for Latosols, Nitosols, and Argisols, dominant soil types in Brazil which correspond to Ferralsols, Oxisols, Acrisols, Lixisols, and Alisols in international systems.

Fe_2O_3 content is also used to assess the degree of ferralitization, a pedogenetic process marked by intense chemical weathering and residual iron enrichment, primarily in the form of iron oxides (Schaefer et al., 2008; van Breemen & Buurman, 1998). These oxides contribute to the formation of oxic horizons (Soil Survey Staff, 2017) or ferralic B horizons (WRB, 2014), composed of highly weathered mineral constituents with variable grain sizes. The formation of such horizons generally requires iron-rich parent material, a climate conducive to deep

weathering, and prolonged exposure to weathering processes. The degree of ferralitization can be quantitatively assessed using the Ferralitization Index (SiO$_2$/Fe$_2$O$_3$), based on sulfuric acid digestion. Accordingly, the concentrations of Fe$_2$O$_3$ and SiO$_2$ measured as part of the Brazilian Soil Classification System (SiBCS) are essential for classifying soils and interpreting their formation.

Additionally, due to the advanced weathering typical of Brazilian soils, Al$_2$O$_3$ and SiO$_2$ contents (also obtained through sulfuric acid digestion) are used to estimate the degree of chemical weathering through the weathering index Ki (Equation (8). This index provides further insight into soil development and mineral transformation in tropical environments. The interpretation of the oxide maps obtained in the prediction was conducted using two distinct approaches. The first approach involved reclassifying the maps based on their oxide contents into three distinct classes. Subsequently, the percentage of the area encompassed by each class was quantified.

The Fe$_2$O$_3$ maps were divided according to the SiBCS classification system, where hypoferric, mesoferric, and ferric. However, for the Al$_2$O$_3$ and SiO$_2$ maps, this classification system was not applicable. Therefore, we proposed patterned intervals based on the range of oxide contents obtained in the maps. We defined soils with oxide contents less than 80 g.kg^{-1} as low, soils ranging from 80 to < 150 g.kg^{-1} as medium, and soils with oxide contents above 150 g.kg^{-1} as high, for both oxides.

The second interpretation was conducted based on the parent materials. We utilized a geological map constructed by Gómez et al. (2018) and calculated the average oxide contents for each parent material.

2.6. Soil clay fraction major oxides maps applications

We tested two applications of the soil oxide maps: Infer the degree of chemical weathering of soils and potential to contribute to the development of high-resolution soil class maps.

2.6.1. Chemical weathering index calculation

To infer the chemical weathering of soils, we calculated the Ki index (Equation (8) proposed by Embrapa Solos (1997). The Ki calculation was performed on a pixel-by-pixel basis using the maps of SiO$_2$ and Al$_2$O$_3$. The Ki weathering index expresses clay mineral destruction and Si leaching by the soil weathering process. The higher the index value, the lower the degree of weathering of the soil. Ki is one of the most widely used indices in Brazil, as it indicates soil mineralogy and weathering. The Ki index is used in the SiBCS to characterize and identify horizons "B latossolicos", an equivalent to ferralic horizons, which should have values below 2.2, indicating a high degree of weathering (Embrapa Solos, 2018). Low Ki values indicate processes of monosiallitization, ferralitization and alitization, leading to the formation of 1:1 clay minerals, iron and aluminum oxides, respectively (Lima et al., 2022). On the other hand, higher Ki values (greater than 2.2) indicate the occurrence of bissiallitization processes with a higher presence of 2:1 clay minerals, indicating a lower degree of weathering (Guimarães et al., 2021).

$$Ki = \frac{SiO2\left(g.kg^{-1}\right)}{Al2O3\left(g.kg^{-1}\right)} X1.70 \tag{8}$$

2.6.2. Application in pedological maps

Considering that, in the Brazilian Soil Classification System (SiBCS), the content of pedogenic oxides and the degree of weathering are used as criteria for soil classification, the generated maps were evaluated for their potential applicability. To this end, a 182 ha farm located in the municipality of Rafard, São Paulo state (Fig. 1), was selected to test the contribution of these maps to the development of pedological maps. The chosen area exhibits significant variations in lithology, with a predominance of immature psammites with heterogeneous grain sizes, transitioning to feldspathic psammites and even arkosic sandstones. Concurrently with this lithology, eruptive elements of dikes from the

Serra Geral formation occur, composed of intrusive bodies of tholeiitic basalt (Nanni et al., 2019). This complex lithology has contributed to the formation of different soil classes. A conventional soil class map was constructed by Bazaglia Filho et al. (2013) for the area. In this map, five soil classes were identified at the first categorical level of SiBCS.

Using our oxide maps, we performed an unsupervised classification using the K-means algorithm. This classification generated a map containing groups based on the content of oxides. All six oxide maps (Fe$_2$O$_3$, Al$_2$O$_3$, and SiO$_2$ in the 0–20 cm and 80–100 cm layers) were used as input variables in the classification. Additionally, we included the Ki maps from both layers as input variables. The number of groups was set to five, representing the number of soils from the conventional soil class map. Finally, we compared the resulting map from the unsupervised classification with the conventional soil class map.

3. Results and Discussion

3.1. Descriptive statistics

The main soil oxide contents are highly variable among samples and depths (Fig. 2). The Fe$_2$O$_3$ contents ranged from 0 to 385 g kg^{-1}, SiO$_2$ varied from 0 to 390 g kg^{-1}, and Al$_2$O$_3$ ranged from 0.34 to 396 g kg^{-1}. The contents of these oxides in the soil clay fraction are related to two factors: the type of parent material and its degree of weathering (Schaefer et al., 2008). The high variability observed in oxide contents is explained by the extensive geodiversity throughout the Brazilian territory (de Paula Silva et al., 2021). Differences were also observed in the contents between the evaluated layers. All studied oxides have higher concentrations in the 80–100 cm layers (Fig. 2). Minor differences were observed for Fe$_2$O$_3$, where the median content in the 0–20 cm layer was 41 g kg^{-1}, and in the 80–100 cm layer it was 57 g kg^{-1}. On the other hand, for SiO$_2$ and Al$_2$O$_3$ oxides, the differences were more pronounced, with median contents in the 0–20 cm layer of 118 and 100 g kg^{-1}, and in the 80–100 cm layer of 163 and 154 g kg^{-1}, respectively. The surface layers have greater interaction with the weathering agents causing a more intense loss of elements than in the deeper layer, especially in tropical regions, resulting in lower quantities of oxides in the surface layers (Guimarães et al., 2021).

3.2. Spearman correlation between soil clay fraction major oxides and environmental covariates

The Spearman correlation analysis between the environmental covariates and the oxides revealed that, for all the oxides, the bands of SYSI showed the highest correlations (Fig. 3a). Among the bands of the SYSI, SWIR2 showed the highest correlation values, ranging from −0.46 (for SiO$_2$ in the 80–100 cm layer) to −0.63 (for Fe$_2$O$_3$ in the 0–20 cm layer). Pedogenic oxides are components of the soil clay fraction, and an increase in their content tends to reduce the reflectance intensity throughout the spectrum, resulting in a negative correlation in the regions of the spectrum covered by the SYSI bands, principally in the SWIR region (Terra et al., 2018). Different correlation values of oxides with SYSI were observed among the studied soil layers, with the 0–20 cm layer consistently showing higher correlations when compared to the 80–100 cm layer. This is due to the reflectance collected by satellites only comprises the surface layer of the soil, thus ensuring a direct relationship with oxide content in the 0–20 cm layer and an indirect relationship in the 80–100 cm layer (de Mendes et al., 2019; Rosin et al., 2023).

Regarding the remaining environmental covariates (terrain attributes), a positive correlation with oxide concentrations was observed. According to Marques et al. (2004), most high-altitude areas in Brazil are characterized by older parent materials that have undergone more intensive weathering, resulting in higher concentrations of pedogenic oxides in the clay fraction of the soil. Among the terrain attributes, elevation and slope had the highest correlation values with oxides. Only
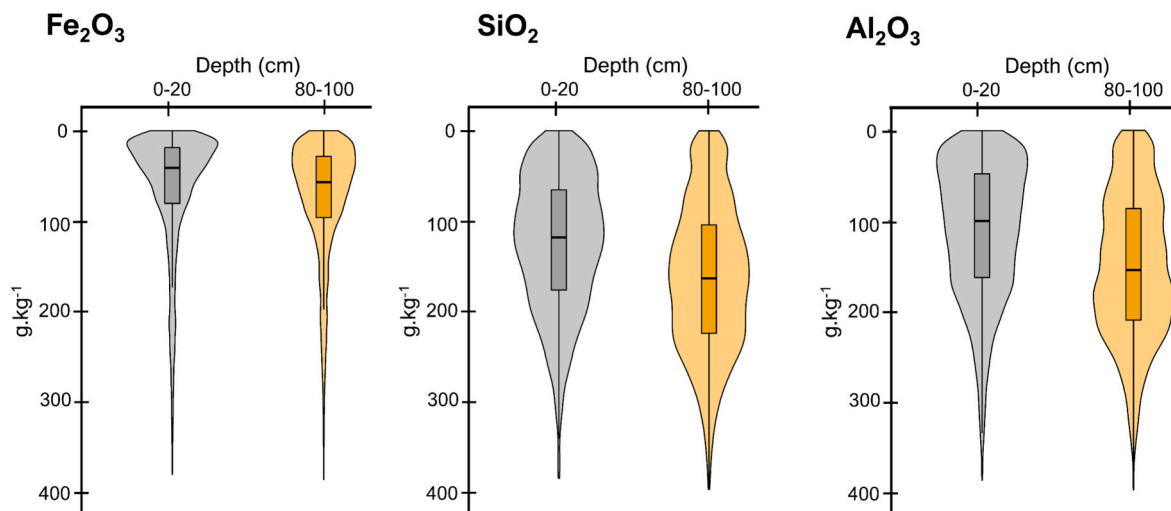
Fig. 2. Descriptive statistics of major soil oxides data observed in the study area.

SiO_2 in the 80–100 cm layer showed no correlation with elevation. The correlations with elevation ranged from 0.21 for SiO_2 in the 0–20 cm layer to 0.40 for Fe_2O_3 in the 0–20 cm layer. In addition, the correlations with slope ranged from 0.15 for SiO_2 in the 80–100 cm layer to 0.27 for Fe_2O_3 in the 0–20 cm layer. The higher correlation values observed between elevation, slope, and Fe_2O_3 content can be attributed to the presence of a basaltic lithological substrate (basalt from the Serra Geral Formation, which is rich in iron) and the occurrence of flattened summits. These geomorphological and geological conditions favor intense chemical weathering, soil homogenization, and ferralitization processes, leading to the development of deeper soils with elevated Fe_2O_3 contents (Mello et al., 2022, 2023). In contrast, soils at lower elevations are typically formed over siltstone substrates and fluvial sediments, which contain lower iron concentrations. In these areas, ferralitization is less advanced, and clay illuviation predominates, resulting in lower Fe_2O_3 contents. Additionally, drainage is often impeded by the presence of argillic horizons, and elutriation processes in the surface horizons further reduce iron accumulation by leaching fine particles, thereby increasing the relative concentration of SiO_2 (Mello et al., 2022, 2023).

3.3. Performance of RF models for soil clay fraction major oxides prediction

The best model performances were observed in the surface layers, for all oxides, especially for Fe_2O_3 in the 0–20 cm layer (RMSE = 49.8 $g.kg^{-1}$, RPIQ = 1.82 and R^2 = 0.62) (Fig. 3b). Conversely, Si_2O_3 presented the lower values, mainly for the 80–100 cm layer (RMSE = 65.3 $g.kg^{-1}$, RPIQ = 1.50 and R^2 = 0.22). This superior performance in the surface layer can be attributed to the SYSI bands that represent the soil surface reflectance, and as demonstrated in section 3.2, showed the highest correlations with the oxides among all environmental covariates.
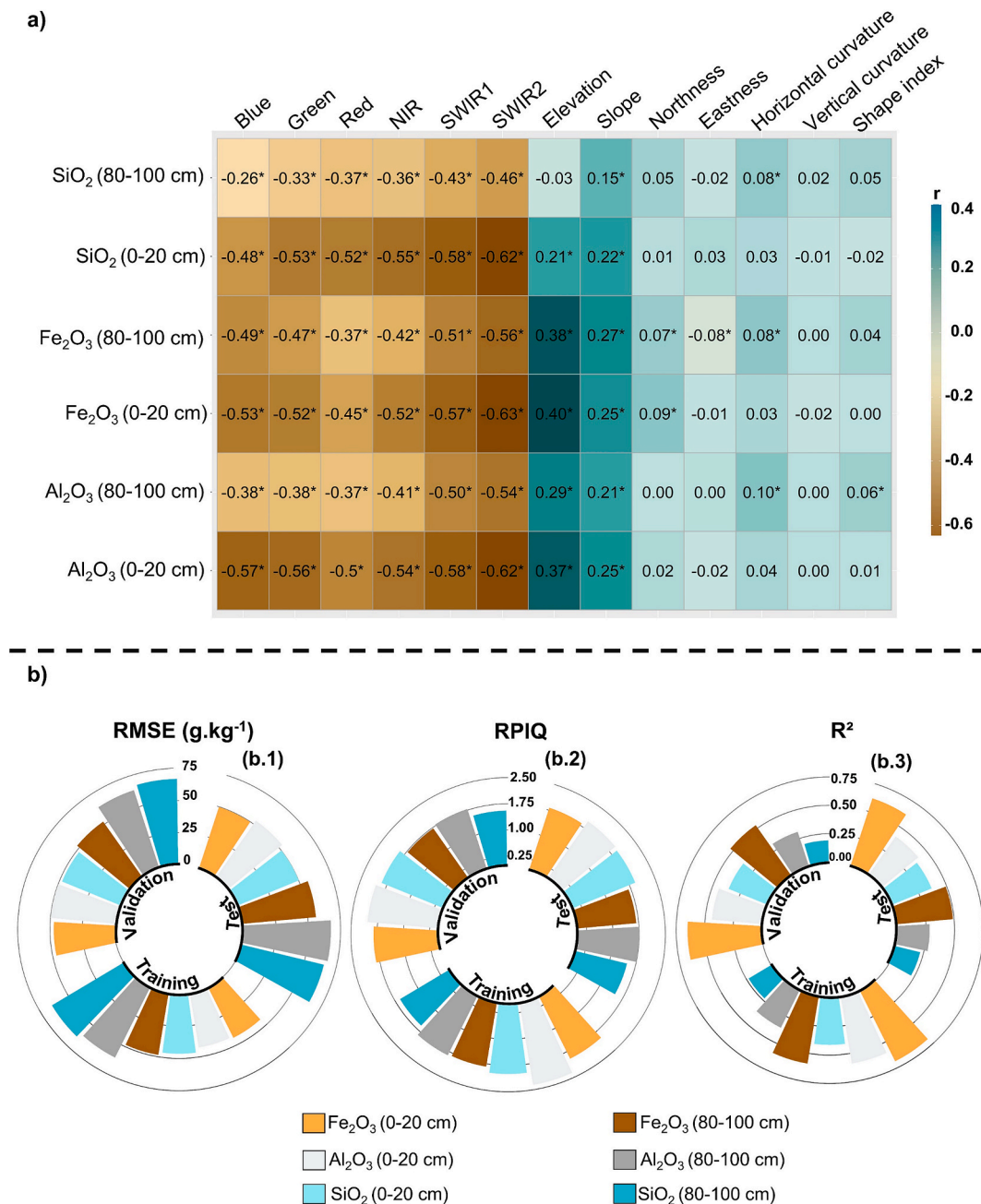
The Fe_2O_3 oxide showed more accurate models than SiO_2 and Al_2O_3 (Fig. 3b). It occurred due to the less generalized occurrence of Fe_2O_3 in soil, being present in higher concentrations under specific conditions of relief, climate, and parent material (Bigham et al., 2002; Wiriyakitnateekul et al., 2007). On the other hand, SiO_2 and Al_2O_3 are present in several soil minerals, such as kaolinite, illite, feldspar, and others, with generalized occurrence in Brazilian territory (Schaefer et al., 2008). Besides, the Fe_2O_3 showed higher Spearman correlations with environmental covariates, than SiO_2 and Al_2O_3, mainly with SYSI bands and elevation, as explained above, which is key to explaining the accuracy of RF prediction models.

3.4. Soil clay fraction major oxides maps

The predicted maps of oxides (Fig. 4) comprised nearly 3.48 million km^2 (~40 % of the national territory), which matches the area covered by the SYSI with bare soils between 1982 and 2021. These maps mainly span the agricultural areas of the country, where land cover tends to be removed due to agricultural practices. The maps for the 80–100 cm layer showed higher concentrations for all oxides, mainly SiO_2, following the patterns observed in the soil dataset (Fig. 2). On the other hand, the maps of the 0–20 cm layer exhibited greater spatial variability. The increased spatial variability observed in the surface layer is a attributed of more intense weathering processes than in the subsurface layer (Poppiel et al., 2019; Rossel, 2011).

Fe_2O_3 exhibited the highest variation: 17.5–232 $g.kg^{-1}$ (0–20 cm layer) and 29–211 $g.kg^{-1}$ (80–100 cm layer). Subsequently, Al_2O_3 ranged from 26-207 $g.kg^{-1}$ (0–20 cm layer) to 88–192 $g.kg^{-1}$ (80–100 cm layer), while SiO_2 ranged from 50.5-170 $g.kg$-1 (0–20 cm layer) to 116–200 $g.kg^{-1}$ (80–100 cm layer). Despite having slightly higher contents in the 80–100 cm layer, the spatial distribution of Fe_2O_3 contents is similar in both mapped layers (Fig. 4a and 4b). Approximately 61 % of the mapped area in the 0–20 cm layer and 57 % in the 80–100 cm layer consist of soils with low iron content (hypoferric soils). Additionally, 9 % and 18 % correspond to soils with moderate iron content (mesoferric soils), while 30 % and 25 % represent soils with high iron content (ferric soils), respectively. Moreover, the clay translocation that occurs in some locations can increase the variability in the surface layer and decrease it in the subsurface (Schaefer et al., 2008).

Oxides Al and Si exhibit distinct spatial behaviors between the evaluated layers (Fig. 4d, 4e, 4g, and 4h). The contents of these oxides are significantly higher in the subsurface layer throughout the mapped area. In the 0–20 cm layer, 49 % and 38 % of the mapped area consist of soils with low contents (<80 $g.kg^{-1}$) of Al_3O_2 and SiO_2, respectively (Fig. 4f and 4i). In contrast, in the 80–100 cm layer, there are no soils containing low contents. Instead, approximately 64 % of the mapped areas contain soils with Al_2O_3 contents ranging from 80 to 150 $g.kg^{-1}$, and 62 % contain SiO_2 contents above 150 $g.kg^{-1}$. In tropical soils, the majority of aluminum (Al) and silicon (Si) are found within the structures of clay minerals, primarily in kaolinite, resulting from the process of monosialization. Due to the process of clay illuviation, these elements are transported to subsurface horizons (Quénard et al., 2011; Schaefer et al., 2008). On the other hand, iron (Fe) is predominantly found in pedogenic oxides such as hematite and goethite, which are products of intense weathering (Schwertmann and Taylor, 1989). The main pedogenetic process in soils with high iron content and intense weathering is

Fig. 3. Performance evaluation of soil clay fraction major oxides prediction using out-of-the-bag validation. (a): Spearman's correlation analysis of environmental covariates with soil oxides, (b) metrics evaluating the models used to predict soil oxide maps, (b.1): Root Mean Squared Error (RMSE), (b.2): Ratio of Performance to InterQuartile distance (RPIQ), (b.3): coefficient of determination (R2), * Significant at 5% probability level.
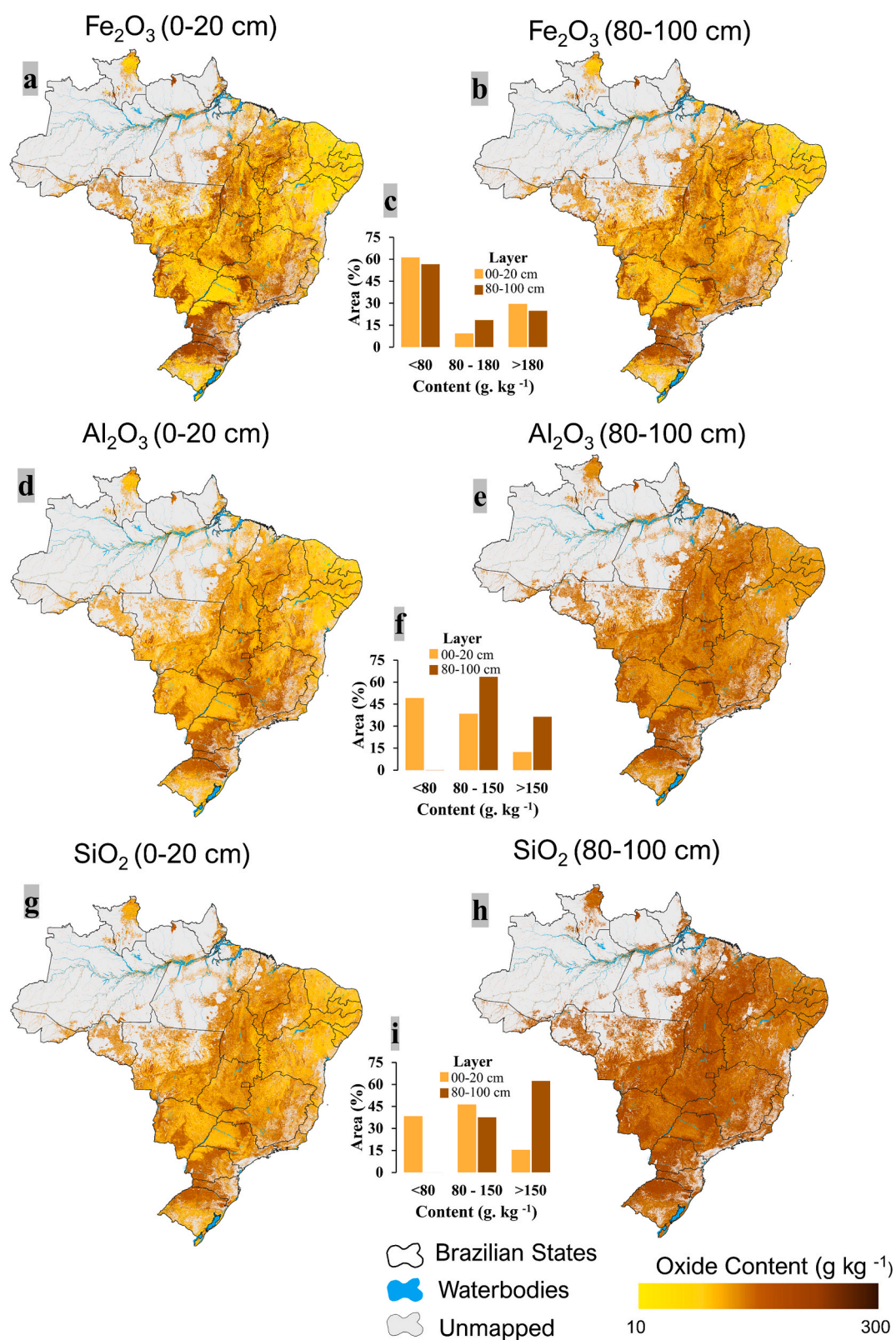
ferralitization (Breemen and Buurman, 1998). Iron oxides form stable complexes with organic matter and other soil particles, limiting their movement within the soil profile (Schwertmann and Taylor, 1989; Van Wambeke et al., 1983).

The predicted oxide contents are strongly correlated with parent materials. The highest contents were observed in regions composed of the basaltic flows of the Serra Geral formation, while the lowest contents were found in soils formed on sandstones (Fig. 5a). A more detailed view of the oxide contents across the main parent materials of Brazilian soils is presented in Fig. 5b. The highest average oxide contents were observed in soils developed on volcanic basaltic rocks (Vb) (Fe_2O_3: 172 and 162; Al_2O_3: 143 and 168; SiO_2: 140 and 188 g.kg^{-1} in the 0–20 and 80–100 cm layers, respectively), followed by soils derived from volcanic rhyolitic rocks (Vr) (Fe_2O_3: 133 and 133; Al_2O_3: 121 and 151; SiO_2: 125
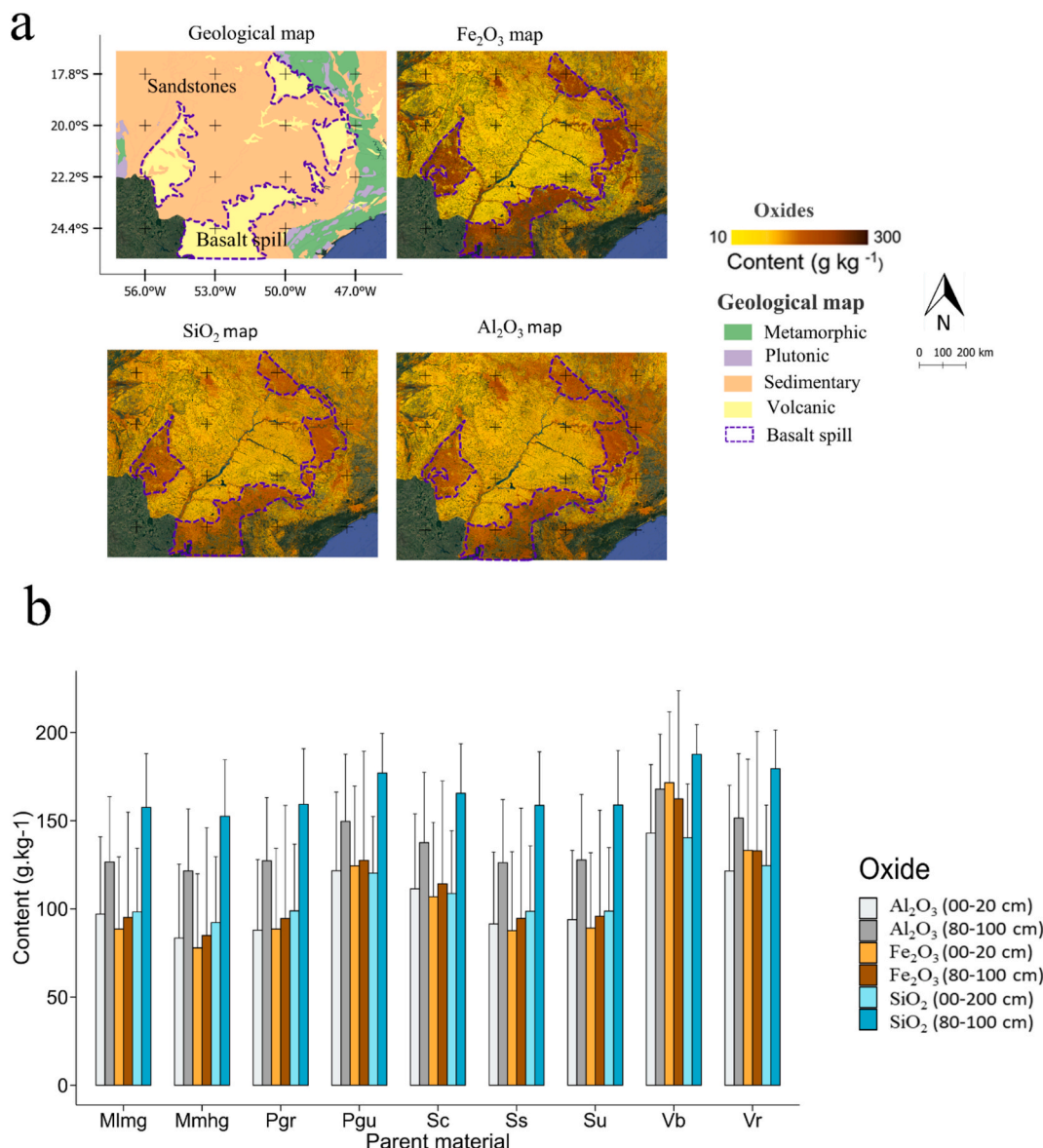
and 180 g.kg^{-1} in the 0–20 and 80–100 cm layers, respectively). The lowest contents were observed in soils derived from metamorphic medium to high-grade rocks (Mmhg) (Fe_2O_3: 78 and 85; Al_2O_3: 83 and 121; SiO_2: 92 and 153 g.kg-1 in the 0–20 and 80–100 cm layers, respectively) and sedimentary siliciclastic rocks (Ss) (Fe_2O_3: 88 and 195; Al_2O_3: 91 and 126; SiO_2: 99 and 159 g.kg^{-1} in the 0–20 and 80–100 cm layers, respectively).

Volcanic rocks are rich in minerals such as amphiboles, pyroxene, and olivine, which are less crystallized and more susceptible to chemical and physical weathering (Schenato et al., 2003). Additionally, these minerals contain significant amounts of iron, giving rise to ferric soils (Araujo et al., 2014; Schaefer et al., 2008). Due to the advanced weathering of these soils, they also tend to have higher clay contents, resulting in elevated concentrations of silicon and aluminum, which are

Fig. 4. Soils clay fraction major oxides predicted maps. (a) Fe$_2$O$_3$ map in the 0–20 cm layer; (b) Fe$_2$O$_3$ map in the 80–100 cm layer; (c) percentage of the mapped area as a function of Fe$_2$O$_3$ content; (d) Al$_2$O$_3$ map in the 0–20 cm layer; (e) Al$_2$O$_3$ map in the 80–100 cm layer; (f) percentage of mapped area as a function of Al$_2$O$_3$ content; (g) SiO$_2$ map in the 0–20 cm layer; (h) SiO$_2$ map in the 80–100 cm layer; (i) percentage of mapped area as a function of SiO$_2$ content.

Fig. 5. Interpretations of soils clay fraction major oxides predicted maps. (a) Comparison between the predicted oxide maps in the 0–20 cm layer with the geological map; (b) Oxide content in soils as a function of parent material. Vr: Volcanic Rhyolitic; Vb: Volcanic Basaltic; Su: Sedimentary Undifferentiated; Ss: Sedimentary Siliciclastic; Sc: Sedimentary Carbonatic; Pgu: Plutonic Gabbroic and Ultramafic; Pgr: Plutonic Granitic; Mmhg: Metamorphic Medium to high grade; Mlmg: Metamorphic Low to medium grade.

constituents of kaolinite and gibbsite (Schaefer et al., 2008). The lower contents observed in soils derived from Mmhg rocks are due to their chemical composition. These rocks are typically gneisses, which have undergone significant transformations in their mineralogy due to intense metamorphism (White et al., 2017). Despite the mineralogical and structural variations, gneisses predominantly exhibit minerals that are resistant to weathering, such as K-feldspars, muscovite, and quartz, along with accessory minerals like zircon (Buol and Weed, 1991; Câmara et al., 2021). Similar to Mmhg rocks, Sedimentary Siliciclastic rocks also predominantly consist of minerals that are highly resistant to chemical weathering, particularly quartz (Garzanti, 2019). This results in soils with lower clay contents and consequently lower oxide concentrations (Martins et al., 2005; L. S. Silva et al., 2019).

In soils developed from plutonic rocks, the oxide contents were higher in Plutonic Gabbroic and Ultramafic (Pgu) rocks compared to Plutonic Granitic (Pg) rocks. Pgu rocks have a mineralogical composition similar to basalt, with a notable presence of ferromagnesian minerals that are more easily weathered (Guimarães et al., 2017; Silva et al.,

2022). On the other hand, Pg rocks are predominantly composed of highly crystallized minerals that are strongly resistant to chemical weathering (Heuze, 1983). Median oxide contents exceeding 80 g·kg^{-1} were observed in soils developed on Sedimentary Carbonatic (Sc) rocks. These values differ from those commonly reported in the literature, especially for Fe_2O_3, as soils formed on Sc rocks usually exhibit contents below 80 g·kg^{-1} (de Souza Oliveira et al., 2021; Pinheiro Junior et al., 2021; Silva et al., 2017). This inconsistency may be attributed to the coarse lithological mapping scale (1:5,000,000), in which regions interpreted as Sc rocks may actually include other lithologies with higher iron contents.

3.4.1. Soil clay fraction major oxides predictions uncertainty

The uncertainties of the predictions were represented by the PIR (Fig. 6). In general, PIR values ranged from 0.1 to 0.5, with higher values observed in the subsurface layers. The lowest uncertainties were observed in the Fe_2O_3 maps, which also exhibited the highest prediction accuracy metrics. Across all maps, higher PIRs were observed in regions
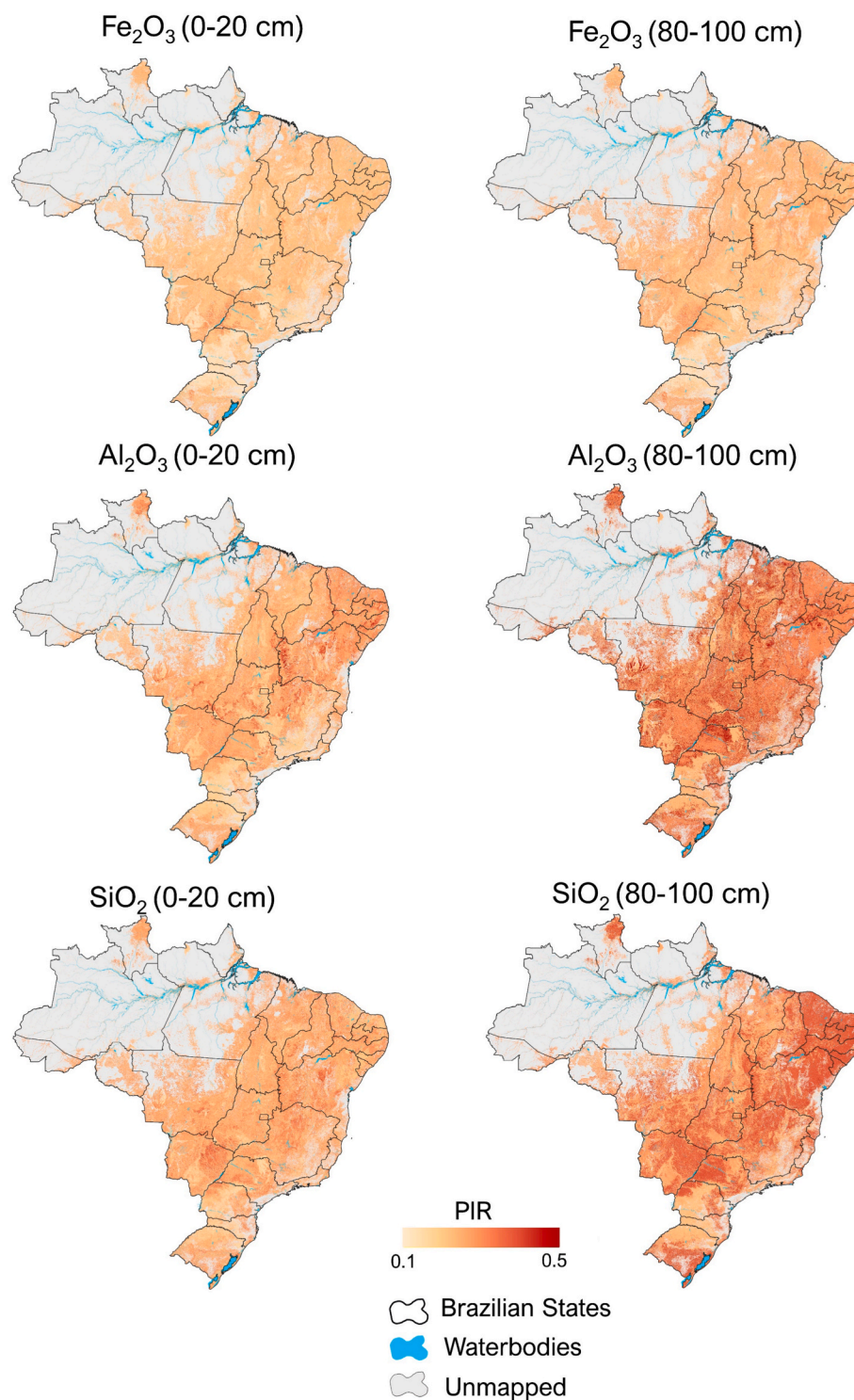
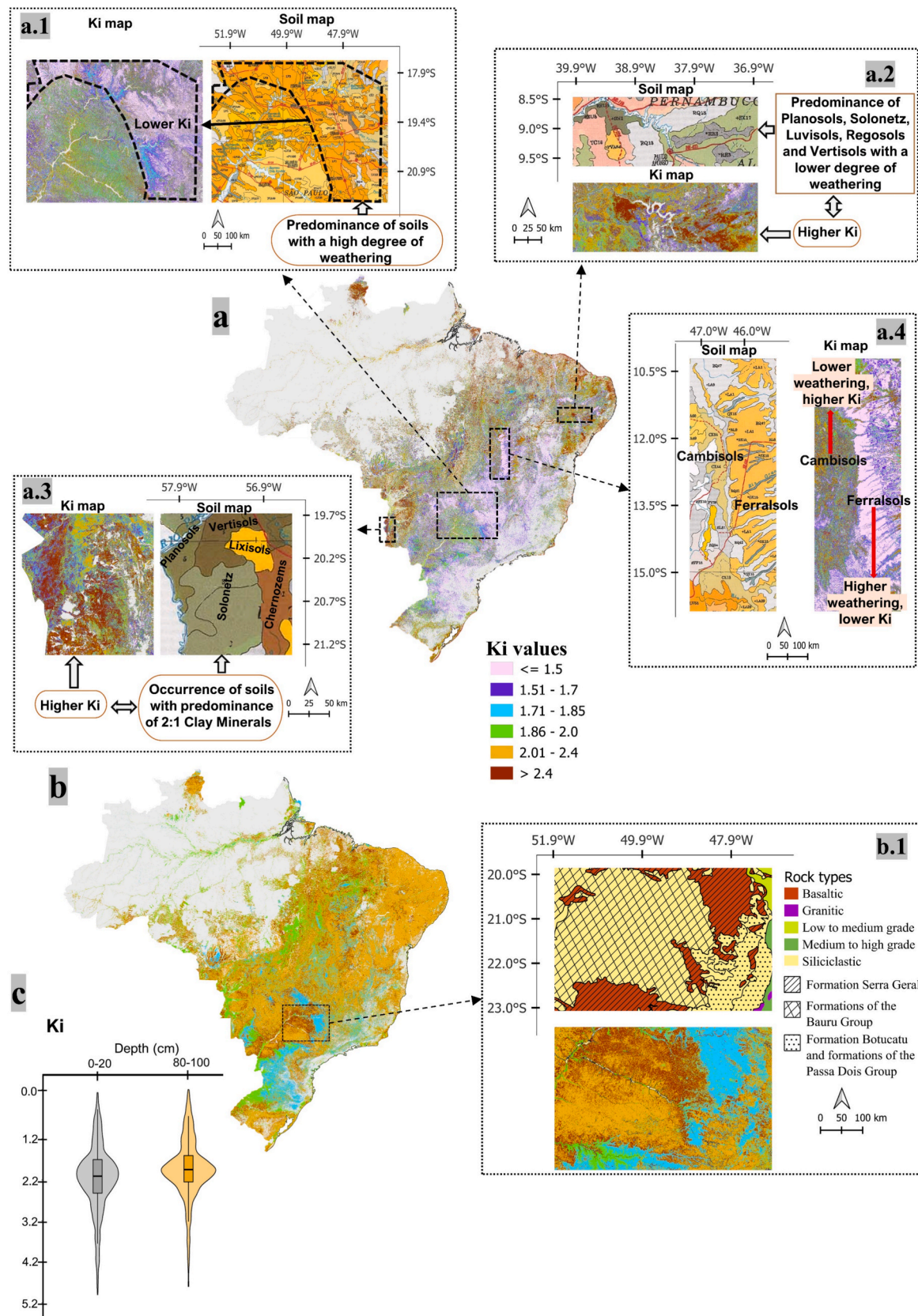**Fig. 6.** Uncertainty Predicted major soil oxides.

with lower sampling densities. As suggested by Wadoux et al. (2020), the estimation of uncertainties can be used to identify under-sampled locations and guide future sampling efforts. The widest PIR ranges were observed for Al_2O_3, and SiO_2 in the 80–100 cm layer, indicating that these maps are the least reliable among those generated in this study.

3.5. Soil clay fraction major oxides maps applications

3.5.1. Maps of weathering index (Ki)

The Ki index maps calculated based on SiO2 and Al_2O_3 contents are presented in Fig. 7. We generated Ki maps for two soil layers (0–20 cm and 80–100 cm). The map of the surface layer (Fig. 7a) revealed greater spatial variability in Ki values, ranging from 0.67 to 4.2, with values below 2 predominating across most of the mapped area. In contrast, the subsurface layer map showed lower spatial variability, with values

Fig. 7. Maps of Ki index. (a) Ki map of the 0–20 cm layer; (a.1) Region representing the Cerrado biome with highly weathered soils; (a.2) Region representing the Caatinga biome with less weathered soils; (a.3) Region representing the Pantanal biome with less weathered hydromorphic soils; (a.4) Region exhibiting a contrast between highly weathered and less weathered soils; (b) Ki map of the 80–100 cm layer; (b.1) Region with different geological formations; (c) Descriptive statistics of the Ki index calculated from laboratory analyses in the two studied layers. The pedological map used for comparisons was the legacy soil map of Brazil published by IBGE, (2021); The geological map used for comparisons was published by Gómez et al. (2018).

between 1.5 and 2.9 (Fig. 7b), and a significant prevalence of areas containing Ki values above 2. These maps are consistent with the Ki values calculated from laboratory-analyzed soil samples, as shown in Fig. 7c.

Ki values predominantly equal to or below 1.5 (in the 0–20 cm layer) are observed in the region represented in Fig. 7a.1. This region corresponds to a portion of the Cerrado biome, where the predominant soils are Ferralsols. The Ferralsols of the Cerrado exhibit varying origins, ranging from sedimentary rocks such as the sandstones of the Paraná Basin to basaltic rocks of the Serra Geral Formation and Cenozoic Lateritic Cover (Oliveira et al., 2023). The occurrence of plateaus associated with the tropical climate has contributed to an intense degree of weathering of these soils, making them the most weathered in Brazil (Curi and Franzmeier, 1984; Oliveira et al., 2023). In general, their mineralogy exhibits a lower presence of primary weatherable minerals and a higher presence of kaolinite and oxides, with the soils tend to be more gibbsitic in the higher areas of the plateaus (Schaefer et al., 2008).

The Ki maps represent the spatial variations in the degree of weathering in soils formed under different circumstances, as can be observed in Fig. 7a.4. The delimited area in the figure encompasses a region between the states of Bahia and Tocantins, as well as the northern part of Goiás. This region exhibits a strong contrast between highly weathered soil classes (Ferralsols) on one side, which are found in the extensive plateau of the Urucuia geological formation, and Cambisols and Regosols formed on the Bambuí group on the other side. Values of Ki ranging from 1.91 to 3.10 were observed by Maranhão et al, (2016) when studying profiles formed on the Bambuí group in this region, which is consistent with our results.

Higher values of Ki in the surface layer (predominantly $\geq$ 2) are observed in the regions encompassing the Pantanal and Caatinga biomes, where younger soils such as Leptosols, Regosols, Luvissols, Planosols, Solonetz, and Vertisols are predominant. Fig. 7a.2 and 7a.3 provide a zoom-in on these regions, offering a more detailed visualization. The Pantanal is a vast floodplain, and as such, hydromorphism is the main pedogenetic process in its soils, reducing weathering rates (de Coringa et al., 2012). Additionally, the parent materials in the Pantanal region often consist of carbonate sediments or exhibit a significant presence of sodium (Andrade et al., 2020; de Oliveira et al., 2021). The chemical elements released during the weathering of these sediments that remain in the soil solution increase the pH, favoring the neoformation of 2:1 clay minerals (Couto et al., 2023). Contrary to the Pantanal, the Caatinga biome is one of the largest tropical dry areas in the world (Araujo et al., 2022). Furthermore, the Caatinga is the oldest semiarid landscape in South America, and its semiarid climate strongly influences the formation of its soils (Araújo Filho et al., 2023). The majority of soils in this biome contain significant amounts of primary minerals and 2:1 clay minerals due to the low rates of weathering (Oliveira et al., 2019).

The map of the 80–100 cm layer showed that subsurface soils are less weathered and more similar to the parent material than the surface soils. It is possible to observe that lower Ki values are found in regions composed of rocks that undergo weathering more easily, such as basalt. In Fig. 7b.1, we selected a region where soils have developed on different parent materials. Soils derived from quartzose rocks (Bauru Group Formation, Botucatu Formation, and Passa Dois Formation) exhibit Ki values above 2 in the subsurface layer. On the other hand, soils formed on the Serra Geral Formation (ultramafic basaltic rocks) display Ki values lower than 2, indicating greater weathering in the subsurface.

As observed by Wilford (2012), the intensity of weathering plays a crucial role in the transformation of primary minerals into secondary components, such as clays and oxides, which in turn influence the hydrological, geochemical, and geophysical properties of soils. Although our research focused on spectral and terrain covariates, other methodologies—such as the approach proposed by Wilford (2012)—provide valuable insights into how the integration of topographic and radiometric data can enhance our understanding of regolith properties.

Weathering intensity, which integrates airborne gamma-ray spectrometry data with elevation information from the Shuttle Radar Topography Mission (SRTM), can serve as a promising tool to improve soil property modeling, particularly in areas where soil characteristics exert strong influence. Although we did not incorporate radiometric data in our study due to the lack of high-resolution information, the weathering intensity methodology using radiometric inputs may be employed in future research, should such data become available—thus refining the prediction of soil attributes in our study area.

Moreover, Wilford's (2012) methodology can be adapted to assess weathering intensity across different landscapes and regions, as demonstrated by Mello et al. (2023) in tropical landscape soils, providing complementary insights to our terrain-based analysis.

3.5.2. Application in pedological maps

The study area comprises five soil classes: Cambisols, Chernozems, Nitisols, Lixisols, and Leptosols (Fig. 8a), with a predominance of Lixisols. Unsupervised classification, performed using the oxide maps, created groups that represented the observed soil classes on the pedological map (Fig. 8b). The best representations were found in the regions corresponding to Nitisols and Lixisols, represented by groups (1) and (2), respectively. The Nitisols in these regions developed over iron-rich basaltic rocks and easily weatherable primary minerals, thus exhibiting high oxide content (Silvero et al., 2021b), due to intense weathering and ferralitization processes, in contrast to the other soils in the area that have developed on parent materials moderately rich in quartz (Nanni et al., 2019). This characteristic allowed for the distinction of this soil class from the others due to its high oxide content and pronounced degree of weathering. In contrast, in this region, Lixisols are soils with a moderate degree of weathering when compared to Nitisols, where the pedogenetic process of clay illuviation is predominant (Embrapa Solos, 2018). This process favors the formation of sandy surface layers and clayey subsurface layers. As a result, there are significant differences in oxide content between the layers (Quénard et al., 2011). This characteristic allowed for the discrimination of these soils into a single group.

In the regions encompassing the soil classes Cambisols, Chernozems, and Leptosols, there was no clear discrimination between the groups. These soils share common characteristics such as a low degree of weathering, significant amounts of primary minerals and 2:1 clay minerals, and shallow depth (Embrapa Solos, 2018). We believe that due to these similarities, it was challenging to individualize these soils into distinct groups.

3.6. Limitations and challenges

Mapping large territories, such as Brazil, is a laborious task that poses several challenges, especially when soil attributes are difficult to determine in the laboratory. In this study, clay fraction oxides were mapped, which are difficult to determine in the laboratory due to the high costs of analysis and the polluting potential of the reagents involved. However, these oxides are of great importance for the understanding of soils.

One of the major limitations of this study was the low sampling density, with many areas still under-sampled. The uncertainty maps indicate that many locations require further efforts for future collections and analyses. Another limitation was the absence of environmental covariates that represented the parent material in greater detail since soil oxides have a direct relationship with this formation factor. Despite this, the soil oxides obtained were consistent with the existing geological maps at coarse scales.

The soil maps showed great agreement with the legacy pedological maps and the understanding of Brazilian soils. However, the accuracy metrics, especially R^2, were not high, ranging from 0.22 to 0.62, and the RMSE was not so low, ranging from 49.8 to 65.3 g.kg^{-1}, suggesting that the maps may contain significant errors in some regions, as shown in the
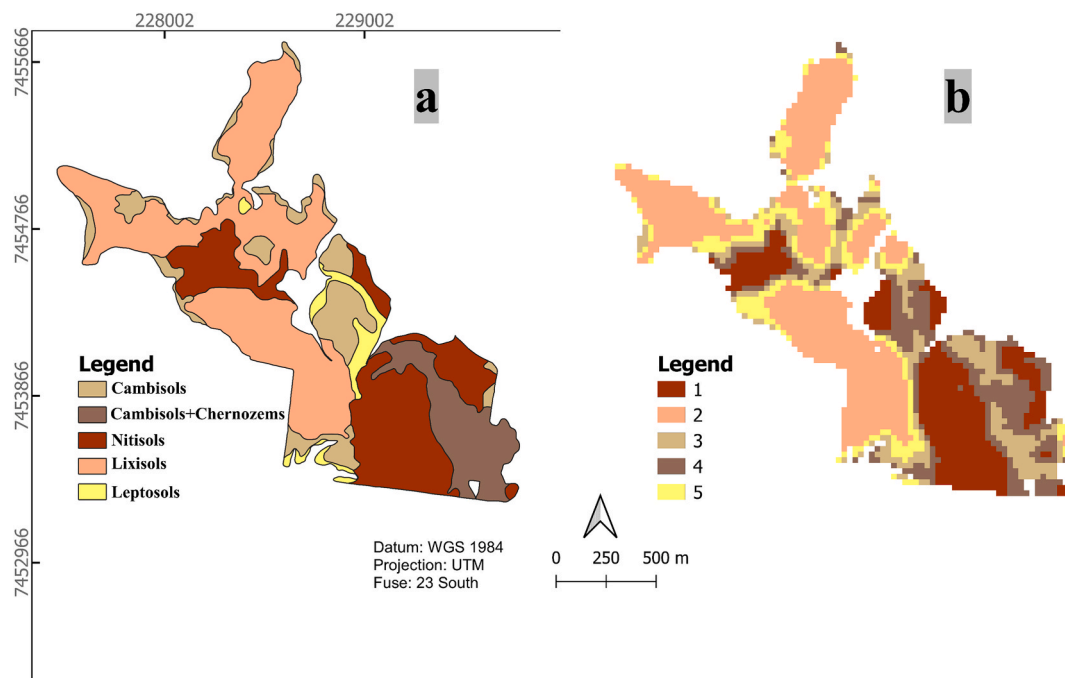
Fig. 8. Comparisons between (a) a detailed pedological map, (b) Class map obtained by unsupervised classification of soil oxide, and Ki maps.

uncertainty maps, especially in the subsurface maps. Despite this, the maps follow the expected spatial distribution. Another major limitation is the mapping in areas where native vegetation remains intact. In our study, SYSI was the most important covariate, but it is not possible to obtain this covariate for the entire Brazilian territory, limiting the map only to agricultural areas.

In future soil mapping, some major challenges remain, including: (I) improving the database to contain data from under-sampled areas; (II) developing covariates that represent parent materials with greater detail; and (III) developing covariates that can map these oxides in areas where vegetation has never been removed and where SYSI cannot be obtained.

4. Conclusions

It was possible to predict soil clay fraction major oxides (Fe_2O_3, SiO_2, Al_2O_3) with high spatial resolution (30 m) using satellite-derived data and ML. Furthermore, the prediction of our maps was considered satisfactory, although the R^2 values are not high, ranging from 0.22 to 0.62, and the RMSE was not as low, ranging from 49.8 to 65.3 $g.kg^{-1}$. The predicted oxide maps covered approximately 3.48 million km^2 (~40 % of the national territory), encompassing all areas that were used for agriculture during the period from 1982 to 2022.

The oxide contents were strongly related to the pedogenetic processes of soil formation, such as hydromorphism, clay illuviation, monosialitization, and ferralitization. Additionally, a significant relationship between soil-forming factors, such as parent material and climate, was observed.

The maps of soil clay fraction major oxides showed potential for inferring the degree of weathering of soils using the weathering index Ki. The Ki index demonstrated consistency with legacy pedological maps, where higher weathering was observed in regions predominantly composed of Ferralsols, while less weathered soils were found in regions dominated by Neosols, Luvissols, Planosols, Solonetz, and Vertisols.

Oxide maps exhibited potential to assist in the construction of pedological maps, capable of distinguishing regions belonging to the same soil class. Furthermore, the findings provide valuable information for all agricultural regions in Brazil, with high spatial resolution. This information can aid researchers, farmers, and consultants in understanding the dynamics of their soils and making informed decisions. For future mappings, major challenges include improving the database by incorporating data from under-sampled areas and developing detailed covariates to represent parent materials. Additionally, there is a need to develop covariates capable of mapping oxides in areas where natural vegetation remains and SYSI data cannot be obtained.

CRediT authorship contribution statement

Jorge Tadeu Fim Rosas: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **José A.M. Demattê:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Nícolas Augusto Rosin:** Writing – review & editing, Writing – original draft, Data curation, Conceptualization. **Raul Roberto Poppiel:** Writing – review & editing, Writing – original draft, Visualization, Conceptualization. **Nélida E.Q. Silvero:** Writing – review & editing, Writing – original draft, Visualization. **Merilyn Taynara Accorsi Amorim:** Writing – review & editing, Writing – original draft, Visualization. **Heidy S. Rodríguez-Albarracín:** Writing – review & editing, Writing – original draft, Visualization. **Letícia Guadagnin Vogel:** Writing – review & editing, Data curation. **Bruno dos Anjos Bartsch:** Writing – review & editing, Data curation. **José João Lelis Leal de Souza:** Writing – review & editing, Data curation. **Lucas de Carvalho Gomes:** Writing – review & editing. **Danilo César de Mello:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

17876-8, CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) process number: 88887.481452/2020-00, CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) process number: 307190-2021-8, CCARBON (Center for Carbon Research in Tropical Agriculture) at the University of São Paulo, sponsored by the FAPESP under grant 2021/10573-4, the Geotehcnologies on Soil Science (https://esalqgeocis.wixsite.com/geocis) and FEALQ - Fundação de Estudos Agrários Luiz de Queiroz (https://fealq.org.br/).

Data availability

The dataset generated during the current study is available in Google's Earth Engine App provided in the link: https://geocis.users.earthengine.app/view/soil-oxidesrosas-etal. These data have a spatial resolution of 90 m and can be viewed and downloaded by any user.

References

Ackerson, J.P., Demattê, J.A.M., Morgan, C.L.S., 2015. Predicting clay content on field-moist intact tropical soils using a dried, ground VisNIR library with external parameter orthogonalization. Geoderma 259–260, 196–204. https://doi.org/10.1016/J.GEODERMA.2015.06.002.

Adhikari, K., Hartemink, A.E., 2016. Linking soils to ecosystem services — a global review. Geoderma 262, 101–111. https://doi.org/10.1016/J.GEODERMA.2015.08.009.

Alvares, C.A., Stape, J.L., Sentelhas, P.C., De Moraes Gonçalves, J.L., Sparovek, G., 2013. Köppen's climate classification map for Brazil. Meteorol. Z. 22, 711–728. https://doi.org/10.1127/0941-2948/2013/0507.

Andrade, G.R.P., Furquim, S.A.C., Nascimento, T.T.V., do Brito, A.C., Camargo, G.R., Souza, G.C. de, 2020. Transformation of clay minerals in salt-affected soils, Pantanal wetland. Brazil. Geoderma 371, 114380. https://doi.org/10.1016/J.GEODERMA.2020.114380.

Araújo Filho, J.C., Correa, M.M., Paiva, A.Q., Costa, O.V., Valladares, G.S., Ribeiro (in memoriam), M.R., Schaefer, Carlos E. G. R., 2023. Semi-arid Soils of the Caatinga Biome of Northeastern Brazil, in: Schaefer, C. E. G. R. (Ed.), The Soils of Brazil. Springer, Cham, pp. 175–193. https://doi.org/10.1007/978-3-031-19949-3_6.

Araujo, H.F.P., Garda, A.A., Girão e Silva, W.A. de, Nascimento, N.F.F., do Mariano, E. de F., Silva, J.M.C. da, 2022. The Caatinga region is a system and not an aggregate. J. Arid Environ. 203, 104778. https://doi.org/10.1016/J.JARIDENV.2022.104778.

Araujo, M.A., Pedroso, A.V., Amaral, D.C., Zinn, Y.L., 2014. Paragênese mineral de solos desenvolvidos de diferentes litologias na região sul de Minas Gerais. Rev. Bras. Cienc. Solo 38, 11–25. https://doi.org/10.1590/S0100-06832014000100002.

Barra, I., Haefele, S.M., Sakrabani, R., Kebede, F., 2021. Soil spectroscopy with the use of chemometrics, machine learning and pre-processing techniques in soil diagnosis: recent advances–a review. TrAC Trends Anal. Chem. 135, 116166. https://doi.org/10.1016/J.TRAC.2020.116166.

Bazaglia Filho, O., Rizzo, R., Lepsch, I.F., do Prado, H., Gomes, F.H., Mazza, J.A., Demattê, J.A.M., 2013. Comparison between detailed digital and conventional soil maps of an area with complex geology. Rev. Bras. Cienc. Solo 37, 1136–1148. https://doi.org/10.1590/S0100-06832013000500003.

Bigham, J.M., Fitzpatrick, R.W., Schulze, D.G., 2002. Iron Oxides, in: Joe B. Dixon, Darrel G. Schulze (Eds.), Soil Mineralogy with Environmental Applications. John Wiley & Sons, Ltd, pp. 323–366. https://doi.org/10.2136/SSSABOOKSER7.C10.

Bishop, T.F.A., McBratney, A.B., Laslett, G.M., 1999. Modelling soil attribute depth functions with equal-area quadratic smoothing splines. Geoderma 91, 27–45. https://doi.org/10.1016/S0016-7061(99)00003-8.

Breemen, N., Buurman, P., 1998. Ferralitization, in: Soil Formation. Springer, Dordrecht, pp. 291–312. https://doi.org/10.1007/978-0-585-31788-5_13.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324/METRICS.

Buol, S.W., Weed, S.B., 1991. Saprolite-soil transformations in the Piedmont and Mountains of North Carolina. Geoderma 51, 15–28. https://doi.org/10.1016/0016-7061(91)90064-Z.

Câmara, E.R.G., dos Santos, J.C.B., de Araújo Filho, J.C., Schulze, S.M.B.B., Corrêa, M.M., Ferreira, T.O., de Sousa, J.E.S., de Souza Júnior, V.S., 2021. Parent rock–pedogenesis relationship: how the weathering of metamorphic rocks influences the genesis of Planosols and Luvisols under a semiarid climate in NE Brazil. Geoderma 385, 114878. https://doi.org/10.1016/J.GEODERMA.2020.114878.

de Coringa, E., Couto, E.G., Perez, X.L.O., Torrado, P.V., 2012. Atributos de solos hidromórficos no pantanal norte matogrossense. Acta Amazon. 42, 19–28. https://doi.org/10.1590/S0044-59672012000100003.

Couto, E.G., Corrêa, G.R., Oliveira, V.A., do Nascimento, A.F., Vidal-Torrado, P., Beirigo, R., Schaefer, C.E.G.R., 2023. Soils of Pantanal: The Largest Continental Wetland, in: The Soils of Brazil. Springer, Cham, pp. 239–267. https://doi.org/10.1007/978-3-031-19949-3_9.

Curi, N., Franzmeier, D.P., 1984. Toposequence of oxisols from the central plateau of Brazil. Soil Sci. Soc. Am. J. 48, 341–346. https://doi.org/10.2136/sssaj1984.03615995004800020024X.

de Alkmim, F.F., 2015. Geological Background: A Tectonic Panorama of Brazil, in: Landscapes and Landforms of Brazil. Springer, pp. 9–17. https://doi.org/10.1007/978-94-017-8023-0_2/COVER.

de Paula Silva, J., Alves, G.B., Ross, J.L.S., de Oliveira, F.S., do Nascimento, M.A.L., Felini, M.G., Manosso, F.C., Pereira, D.Í., 2021. The geodiversity of Brazil: quantification, distribution, and implications for conservation areas. Geoheritage 13, 1–21. https://doi.org/10.1007/S12371-021-00598-0/FIGURES/9.

de Souza Oliveira, N., Schiavo, J.A., Cirilo de Souza, A., Laranjeira, L.T., Viana de Moraes, E.M., Pereira, M.G., 2021. Mineralogy and genesis in an alkaline soil system in the southern Pantanal wetland, Brazil. J. South Am. Earth Sci. 111, 103456. https://doi.org/10.1016/J.JSAMES.2021.103456.

Demattê, J.A.M., Dotto, A.C., Paiva, A.F.S., Sato, M.V., Dalmolin, R.S.D., de Araújo, M. do S.B., da Silva, E.B., Nanni, M.R., ten Caten, A., Noronha, N.C., Lacerda, M.P.C., de Araújo Filho, J.C., Rizzo, R., Bellinaso, H., Francelino, M.R., Schaefer, C.E.G.R., Vicente, L.E., dos Santos, U.J., de Sá Barretto Sampaio, E.V., Menezes, R.S.C., de Souza, J.J.L.L., Abrahão, W.A.P., Coelho, R.M., Grego, C.R., Lani, J.L., Fernandes, A.R., Gonçalves, D.A.M., Silva, S.H.G., de Menezes, M.D., Curi, N., Couto, E.G., dos Anjos, L.H.C., Ceddia, M.B., Pinheiro, É.F.M., Grunwald, S., Vasques, G.M., Marques Júnior, J., da Silva, A.J., de Barreto, M.C.V., Nóbrega, G.N., da Silva, M.Z., de Souza, S.F., Valladares, G.S., Viana, J.H.M., da Silva Terra, F., Horák-Terra, I., Fiorio, P.R., da Silva, R.C., Frade Júnior, E.F., Lima, R.H.C., Alba, J.M.F., de Souza Junior, V.S., Brefin, M.D.L.M.S., Ruivo, M.D.L.P., Ferreira, T.O., Brait, M.A., Caetano, N.R., Bringhenti, I., de Sousa Mendes, W., Safanelli, J.L., Guimarães, C.C.B., Poppiel, R.R., e Souza, A.B., Quesada, C.A., do Couto, H.T.Z., 2019. The Brazilian Soil Spectral Library (BSSL): a general view, application and challenges. Geoderma 354, 113793. https://doi.org/10.1016/J.GEODERMA.2019.05.043.

Demattê, J.A.M., Fioriob, P.R., Ben-Dorc, E., 2009. Estimation of soil properties by orbital and laboratory reflectance means and its relation with soil classification. The Open Remote Sensing J. 2, 12–23. https://doi.org/10.2174/187541390100201012.

Demattê, J.A.M., Fongaro, C.T., Rizzo, R., Safanelli, J.L., 2018. Geospatial Soil Sensing System (GEOS3): a powerful data mining procedure to retrieve soil spectral reflectance from satellite images. Remote Sens. Environ. 212, 161–175. https://doi.org/10.1016/j.rse.2018.04.047.

Demattê, J.A.M., Nanni, M.R., Formaggio, A.R., Epiphanio, J.C.N., Dematte, J.A.M., Dematte, D., Nanni, M.R., 2007. Spectral reflectance for the mineralogical evaluation of Brazilian low clay activity soils. 28, 4537–4559. https://doi.org/10.1080/01431160701250408.

Demattê, J.A.M., Safanelli, J.L., Poppiel, R.R., Rizzo, R., Silvero, N.E.Q., Mendes, W. de S., Bonfatti, B.R., Dotto, A.C., Salazar, D.F.U., Mello, F.A. de O., Paiva, A.F. da S., Souza, A.B., Santos, N.V. dos, Maria Nascimento, C., Mello, D.C. de, Bellinaso, H., Gonzaga Neto, L., Amorim, M.T.A., Resende, M.E.B. de, Vieira, J. da S., Queiroz, L.G. de, Gallo, B.C., Sayão, V.M., Lisboa, C.J. da S., 2020. Bare Earth's Surface Spectra as a Proxy for Soil Resource Monitoring. Sci Rep 10, 1–11. https://doi.org/10.1038/s41598-020-61408-1.

Ellili-Bargaoui, Y., Walter, C., Michot, D., Lemercier, B., 2020. Mapping soil properties at multiple depths from disaggregated legacy soil maps in the Brittany region, France. Geoderma Regional 23, e00342. https://doi.org/10.1016/J.GEODRS.2020.E00342.

Solos, E., 1997. Manual of Methods of Soil Analysis. Ministry of Agriculture, Rio de Janeiro, Brazil.

Solos, E., 2018. Sistema Brasileiro de Classificação de Solos. Embrapa Solos, Rio de Janeiro.

Farr, T.G., Kobrick, M., 2000. Shuttle radar topography mission produces a wealth of data. Eos Trans. AGU 81, 583–585. https://doi.org/10.1029/EO081I048P00583.

Fongaro, C.T., Demattê, J.A.M., Rizzo, R., Safanelli, J.L., de Mendes, W.S., Dotto, A.C., Vicente, L.E., Franceschini, M.H.D., Ustin, S.L., 2018. Improvement of clay and sand quantification based on a novel approach with a focus on multispectral satellite images. Remote Sens. (Basel) 10. https://doi.org/10.3390/rs10101555.

Gallo, B.C., Demattê, J.A.M., Rizzo, R., Safanelli, J.L., de Mendes, W.S., Lepsch, I.F., Sato, M.V., Romero, D.J., Lacerda, M.P.C., 2018. Multi-temporal satellite images on topsoil attribute quantification and the relationship with soil classes and geology. Remote Sens. (Basel) 10. https://doi.org/10.3390/rs10101571.

Garzanti, E., 2019. Petrographic classification of sand and sandstone. Earth Sci. Rev. 192, 545–563. https://doi.org/10.1016/J.EARSCIREV.2018.12.014.

Gasmi, A., Gomez, C., Lagacherie, P., Zouari, H., Laamrani, A., Chehbouni, A., 2021. Mean spectral reflectance from bare soil pixels along a Landsat-TM time series to increase both the prediction accuracy of soil clay content and mapping coverage. Geoderma 388, 114864. https://doi.org/10.1016/J.GEODERMA.2020.114864.

Gómez, J., Schobbenhaus, C., Montes, N.E., 2018. Geological Map of South America At a Scale of 1: 5M. Commisslon for the Geological Map of the World (CGMW), Colombian Geological Survey and Geological Survey of Brazil.

Guimarães, C.C.B., Demattê, J.A.M., de Azevedo, A.C., da Silva, R.C., Salazar, D.F.U., 2021. Weathering of soils originated from diabase in the Paraná Basin, São Paulo State, Brazil. Rev. Ciência Agronômica 52, e20196677. https://doi.org/10.5935/1806-6690.20210063.

Guimarães, I.P., Silva Filho, A.F., Armstrong, R., 2017. Origin and age of coeval gabbros and leucogranites in the northern subprovince of the borborema province, NE Brazil. J. South Am. Earth Sci. 76, 71–93. https://doi.org/10.1016/J.JSAMES.2017.02.014.

Hengl, T., De Jesus, J.M., Heuvelink, G.B.M., Gonzalez, M.R., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S., Kempen, B., 2017. SoilGrids250m: Global gridded soil information based on machine learning. PLoS One 12, e0169748. https://doi.org/10.1371/JOURNAL.PONE.0169748.

Hengl, T., Macmillan, R.A., 2019. Predictive Soil Mapping with R, 1st ed, Predictive soil mapping with R. OpenGeoHub foundation, Wageningen, the Netherlands.

Heuze, F.E., 1983. High-temperature mechanical, physical and thermal properties of granitic rocks— a review. Int. J. Rock Mech. Mining Sci. Geomech. Abstracts 20, 3–10. https://doi.org/10.1016/0148-9062(83)91609-1.

Hewitt, A., Dominati, E., Webb, T., Cuthill, T., 2015. Soil natural capital quantification by the stock adequacy method. Geoderma 241–242, 107–114. https://doi.org/10.1016/J.GEODERMA.2014.11.014.

IBGE, I.B. de G. e E., 2021. Mapa de Solos do Brasil [WWW Document]. https://www.ibge.gov.br/geociencias/downloads-geociencias.html.

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.

Kirsten, M., Mikutta, R., Vogel, C., Thompson, A., Mueller, C.W., Kimaro, D.N., Bergsma, H.L.T., Feger, K.H., Kalbitz, K., 2021. Iron oxides and aluminous clays selectively control soil carbon storage and stability in the humid tropics. Scientific Reports 2021 11:1 11, 1–12. https://doi.org/10.1038/s41598-021-84777-7.

Lamichhane, S., Kumar, L., Wilson, B., 2019. Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: a review. Geoderma 352, 395–413. https://doi.org/10.1016/J.GEODERMA.2019.05.031.

Lima, A.P.B., Inda, A.V., Zinn, Y.L., da Silva, E.R., do Nascimento, P.C., 2022. Soil formation and properties along a sedimentary lithosequence in the ecotonal Cerrados of Mato Grosso, Brazil. Catena (amst) 219, 106599. https://doi.org/10.1016/J.CATENA.2022.106599.

Ma, Y., Minasny, B., Malone, B.P., Mcbratney, A.B., 2019. Pedology and digital soil mapping (DSM). Eur. J. Soil Sci. 70, 216–235. https://doi.org/10.1111/EJSS.12790.

Maranhão, D.D.C., Pereira, M.G., Collier, L.S., dos Anjos, L.H.C., Azevedo, A.C., de Cavassani, R., 2016. Genesis and classification of soils containing carbonates in a toposequence of the Bambuí Group. Rev Bras Cienc Solo 40, e0150295. https://doi.org/10.1590/18069657RBCS20150295.

Marques, J.J., Schulze, D.G., Curi, N., Mertzman, S.A., 2004. Major element geochemistry and geomorphic relationships in Brazilian Cerrado soils. Geoderma 119, 179–195. https://doi.org/10.1016/S0016-7061(03)00260-X.

Martins, F.B., Ferreira, P.M.V., Flores, J.A.A., Bressani, L.A., Bica, A.V.D., 2005. Interaction between geological and geotechnical investigations of a sandstone residual soil. Eng. Geol. 78, 1–9. https://doi.org/10.1016/J.ENGGEO.2004.10.003.

McBratney, A., Field, D.J., Koch, A., 2014. The dimensions of soil security. Geoderma 213, 203–213. https://doi.org/10.1016/J.GEODERMA.2013.08.013.

McBratney, A.B., Santos, M.L.M., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52. https://doi.org/10.1016/S0016-7061(03)00223-4.

Meinshausen, N., ridgeway, G. Quantile regression forests. Journal of machine learning research, 7, n. 6, 2006.

Mello, D.C., Ferreira, T.O., Veloso, G.V., Lana, M.G., Mello, F.A.O., Di Raimo, L.A.D.L., Schaefer, C.E.G.R., Francelino, M.R., Fernandes-Filho, E.I., Demattê, J.A.M., 2022. Pedogenetic processes operating at different intensities inferred by geophysical sensors and machine learning algorithms. Catena 216, 106370. https://doi.org/10.1016/j.catena.2022.106370.

Mello, D.C., Ferreira, T.O., Veloso, G.V., Lana, M.G., Mello, F.A.O., Di Raimo, L.A.D.L., Cabrero, D.R.O., Souza, J.J.L.L., Fernandes-Filho, E.I., Francelino, M.R., Demattê, J.A.M., 2023. Digital mapping of soil weathering using field geophysical sensor data coupled with covariates and machine learning. J. S. Am. Earth Sci. 128, 104449. https://doi.org/10.1016/j.jsames.2023.104449.

de Mendes, W.S., Demattê, J.A.M., Minasny, B., Silvero, N.E.Q., Bonfatti, B.R., Safanelli, J.L., Rizzo, R., da Costa, A.C.S., 2022. Free iron oxide content in tropical soils predicted by integrative digital mapping. Soil Tillage Res. 219, 105346. https://doi.org/10.1016/J.STILL.2022.105346.

de Mendes, W.S., Medeiros Neto, L.G., Demattê, J.A.M., Gallo, B.C., Rizzo, R., Safanelli, J.L., Fongaro, C.T., 2019. Is it possible to map subsurface soil attributes by satellite spectral transfer models? Geoderma 343, 269–279. https://doi.org/10.1016/j.geoderma.2019.01.025.

Nanni, M.R., Demattê, J.A.M., 2006. Spectral reflectance methodology in comparison to traditional soil analysis. Soil Sci. Soc. Am. J. 70, 393–407. https://doi.org/10.2136/SSSAJ2003.0285.

Oliveira, G.D.C., Francelino, M.R., Arruda, D.M., Fernandes-Filho, E.I., Schaefer, C.E.G.R., 2019. Climate and soils at the Brazilian semiarid and the forest-Caatinga problem: new insights and implications. Environ. Res. Lett. 14, 104007. https://doi.org/10.1088/1748-9326/AB3D7B.

de Oliveira, N.S., Schiavo, J.A., Cirilo de Souza, A., Laranjeira, L.T., Viana de Moraes, E.M., Pereira, M.G., 2021. Mineralogy and genesis in an alkaline soil system in the southern Pantanal wetland, Brazil. J. South Am. Earth Sci. 111, 103456. https://doi.org/10.1016/J.JSAMES.2021.103456.

Oliveira, V.A., Santos, G.G., Ker, J.C., Couto, E.G., Jacomine (in memoriam), P.K., Corrêa, G.R., Curi, N., Schaefer, Carlos E. G. R., 2023. Soils of Cerrados, the Brazilian Savannas, in: Schaefer, C. E. G. R. (Ed.), The Soils of Brazil. Springer, Cham, pp. 129–173. https://doi.org/10.1007/978-3-031-19949-3_5.

Padarian, J., Minasny, B., McBratney, A.B., 2020. Machine learning and soil sciences: a review aided by machine learning tools. Soil 6, 35–52. https://doi.org/10.5194/SOIL-6-35-2020.

Pinheiro Junior, C.R., Cipriano da Silva, R., da Silva Neto, E.C., de Azevedo, A.C., Nascimento, C.W.R., do Fontana, A., de Souza Júnior, V.S., dos Anjos, L.H.C., Pereira, M.G., 2021. Influence of carbonate rocks on soil properties in the humid tropical climate of atlantic forest, Rio de Janeiro –Brazil. J. South Am. Earth Sci. 112, 103582. https://doi.org/10.1016/J.JSAMES.2021.103582.

Poggio, L., De Sousa, L.M., Batjes, N.H., Heuvelink, G.B.M., Kempen, B., Ribeiro, E., Rossiter, D., 2021. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. Soil 7, 217–240. https://doi.org/10.5194/soil-7-217-2021.

Poppiel, R.R., Lacerda, M.P.C., Safanelli, J.L., Rizzo, R., Oliveira, M.P., Novais, J.J., Demattê, J.A.M., 2019. Mapping at 30 m resolution of soil attributes at multiple depths in midwest Brazil. Remote Sensing 2019, Vol. 11, Page 2905 11, 2905. https://doi.org/10.3390/RS11242905.

Quénard, L., Samouëlian, A., Laroche, B., Cornu, S., 2011. Lessivage as a major process of soil formation: a revisitation of existing data. Geoderma 167–168, 135–147. https://doi.org/10.1016/J.GEODERMA.2011.07.031.

Raij, B., Valadares, J.M.A.S., 1974. Análise dos elementos maiores de rochas, argilas e solos (IAC. Boletim Técnico, 16). Campinas .

Rizzo, R., Medeiros, L.G., de Mello, D.C., Marques, K.P.P., de Mendes, W.S., Quiñonez Silvero, N.E., Dotto, A.C., Bonfatti, B.R., Demattê, J.A.M., 2020. Multi-temporal bare surface image associated with transfer functions to support soil classification and mapping in southeastern Brazil. Geoderma 361, 114018. https://doi.org/10.1016/J.GEODERMA.2019.114018.

Roberts, D., Wilford, J., Ghattas, O., 2019. Exposed soil and mineral map of the Australian continent revealing the land at its barest. Nat. Commun. 10 (1), 5297. https://doi.org/10.1038/s41467-019-13276-1.

Rosin, N.A., Demattê, J.A.M., Poppiel, R.R., Silvero, N.E.Q., Rodriguez-Albarracin, H.S., Rosas, J.T.F., Greschuk, L.T., Bellinaso, H., Minasny, B., Gomez, C., Marques Júnior, J., Fernandes, K., 2023. Mapping Brazilian soil mineralogy using proximal and remote sensing data. Geoderma 432, 116413. https://doi.org/10.1016/J.GEODERMA.2023.116413.

Rossel, R.A.V., 2011. Fine-resolution multiscale mapping of clay minerals in Australian soils measured with near infrared spectra. J. Geophys. Res. Earth Surf. 116, 4023. https://doi.org/10.1029/2011JF001977.

Safanelli, J.L., Chabrillat, S., Ben-Dor, E., Demattê, J.A.M., 2020a. Multispectral Models from Bare Soil Composites for Mapping Topsoil Properties over Europe. Remote Sensing 2020, Vol. 12, Page 1369 12, 1369. https://doi.org/10.3390/RS12091369.

Safanelli, J.L., Demattê, J.A.M., Chabrillat, S., Poppiel, R.R., Rizzo, R., Dotto, A.C., Silvero, N.E.Q., de Mendes, W.S., Bonfatti, B.R., Ruiz, L.F.C., ten Caten, A., Dalmolin, R.S.D., 2021a. Leveraging the application of Earth observation data for mapping cropland soils in Brazil. Geoderma 396, 115042. https://doi.org/10.1016/J.GEODERMA.2021.115042.

Safanelli, J.L., Demattê, J.A.M., Dos Santos, N.V., Rosas, J.T.F., Silvero, N.E.Q., Bonfatti, B.R., de Mendes, W.S., 2021b. Fine-scale soil mapping with Earth Observation data: a multiple geographic level comparison. Rev Bras Cienc Solo 45, e0210080. https://doi.org/10.36783/18069657RBCS20210080.

Safanelli, J.L., Poppiel, R.R., Chimelo Ruiz, L.F., Bonfatti, B.R., de Oliveira Mello, F.A., Rizzo, R., Demattê, J.A.M., 2020b. Terrain analysis in Google Earth Engine: a method adapted for high-performance global-scale analysis. ISPRS Int. J. Geoinf. 9. https://doi.org/10.3390/ijgi9060400.

Samuel-Rosa, A., Dalmolin, R.S.D., Moura-Bueno, J.M., Teixeira, W.G., Alba, J.M.F., 2019. Open legacy soil survey data in Brazil: geospatial data quality and how to improve it. Sci. Agric. 77, e20170430. https://doi.org/10.1590/1678-992X-2017-0430.

Schaefer, C.E.G.R., Farbris, J.D., Ker, J.C., 2008. Minerals in the clay fraction of Brazilian Latosols (Oxisols): a review. Clay Miner. 43, 137–154. https://doi.org/10.1180/CLAYMIN.2008.043.1.11.

Schaefer, C.E.G.R., Corrêa, G.R., de Oliveira, F.S., Filho, E.I.F., Francelino, M.R., Gomes, L.C., 2023. The Making of Brazilian Soilscapes: A Geosystemic Vista on Neotropical Pedology, in: The Soils of Brazil. Springer, Cham, pp. 25–70. https://doi.org/10.1007/978-3-031-19949-3_2.

Schenato, F., Formoso, M.L.L., Dudoignon, P., Meunier, A., Proust, D., Mas, A., 2003. Alteration processes of a thick basaltic lava flow of the Paraná Basin (Brazil): petrographic and mineralogical studies. J. South Am. Earth Sci. 16, 423–444. https://doi.org/10.1016/S0895-9811(03)00098-1.

Schwertmann, U., Taylor, R.M., 1989. Iron Oxides, in: Dixon, J.B.W.S.B. (Ed.), Minerals in Soil Environments. John Wiley & Sons, Ltd, pp. 379–438. https://doi.org/10.2136/SSSABOOKSER1.2ED.C8.

Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., Homayouni, S., 2020. Support vector machine versus random forest for remote sensing image classification: a meta-analysis and systematic review. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13, 6308–6325. https://doi.org/10.1109/JSTARS.2020.3026724.

Silva, F.M., Silva, S.H.G., Teixeira, A.F.S., Inda, A.V., Fruett, T., Weindorf, D.C., Guilherme, L.R.G., Curi, N., 2022. Using proximal sensors to assess pedogenetic development of Inceptisols and Oxisols in Brazil. Geoderma Reg. 28, e00465. https://doi.org/10.1016/J.GEODRS.2021.E00465.

Silva, L.S., Marques Júnior, J., Barrón, V., Gomes, R.P., Teixeira, D.D.B., Siqueira, D.S., Vasconcelos, V., 2020. Spatial variability of iron oxides in soils from Brazilian sandstone and basalt. Catena (amst) 185, 104258. https://doi.org/10.1016/J.CATENA.2019.104258.

Silva, M.B., dos Anjos, L.H.C., Pereira, M.G., Schiavo, J.A., Cooper, M., de Cavassani, R.S., 2017. Soils in the karst landscape of Bodoquena plateau in cerrado region of Brazil. Catena (amst) 154, 107–117. https://doi.org/10.1016/J.CATENA.2017.02.022.

Silva, S.H.G., Ribeiro, B.T., Guerra, M.B.B., de 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, D.C., 2021. pXRF in tropical soils: methodology, applications, achievements and challenges. Adv. Agron. 167, 1–62. https://doi.org/10.1016/BS.AGRON.2020.12.001.

Silva, S.H.G., Silva, E.A., Poggere, G.C., Junior, A.L.P., Gonçalves, M.G.M., Guilherme, L.R.G., Curi, N., 2019. Modeling and prediction of sulfuric acid digestion analyses data from PXRF spectrometry in tropical soils. Sci. Agric. 77, e20180132. https://doi.org/10.1590/1678-992X-2018-0132.

Silvero, N.E.Q., Demattê, J.A.M., Amorim, M.T.A., dos Santos, N.V., Rizzo, R., Safanelli, J.L., Poppiel, R.R., de Mendes, W.S., Bonfatti, B.R., 2021a. Soil variability and quantification based on Sentinel-2 and Landsat-8 bare soil images: a comparison. Remote Sens. Environ. 252, 112117. https://doi.org/10.1016/J.RSE.2020.112117.

Silvero, N.E.Q., Demattê, J.A.M., Vieira, J.S., Mello, F.A.O., Amorim, M.T.A., Poppiel, R. R., Mendes, W.S., Bonfatti, B.R., 2021b. Soil property maps with satellite images at multiple scales and its impact on management and classification. Geoderma 397, 115089. https://doi.org/10.1016/J.GEODERMA.2021.115089.

Silvero, N.E.Q., Demattê., J.A.M., Minasny, B., Rosin, N.A., Nascimento, J.G., Rodríguez-Albarracín, H.S., Bellinaso, Gómez, A.M.R. 2023. Sensing technologies for characterizing and monitoring soil functions: A review. Chapter three. In: Advances in Agronomy, Volume 177. Pages 125-168. ISSN0065-2113. https://doi.org/10.1016/bs.agron.2022.08.002.

Soil Survey Staff, 2017. Soil Survey Manual. USDA Handbook No. 18. United States Department of Agriculture, Natural Resources Conservation Service, Washington, D. C.

Sulaeman, Y., Minasny, B., McBratney, A.B., Sarwani, M., Sutandi, A., 2013. Harmonizing legacy soil data for digital soil mapping in Indonesia. Geoderma 192, 77–85. https://doi.org/10.1016/J.GEODERMA.2012.08.005.

Tayebi, M., Rosas, J.T.F., Mendes, W. de S., Poppiel, R.R., Ostovari, Y., Ruiz, L.F.C., Dos Santos, N.V., Cerri, C.E.P., Silva, S.H.G., Curi, N., Silvero, N.E.Q., Demattê, J.A.M., 2021. Drivers of organic carbon stocks in different LULC history and along soil depth for a 30 years image time series. Remote Sens. 2021, Vol. 13, 2223. https://doi.org/10.3390/RS13112223.

Terra, F.S., Demattê, J.A.M., Viscarra Rossel, R.A., 2018. Proximal spectral sensing in pedological assessments: vis–NIR spectra for soil classification based on weathering and pedogenesis. Geoderma 318, 123–136. https://doi.org/10.1016/J.GEODERMA.2017.10.053.

Ukabiala, M.E., Kolo, J., Obalum, S.E., Amhakhian, S.O., Igwe, C.A., Hermensah, 2021. Physicochemical properties as related to mineralogical composition of floodplain soils in humid tropical environment and the pedological significance. Environ. Monit. Assess. 193, 1–15. https://doi.org/10.1007/S10661-021-09329-Y/TABLES/3.

USGS, 2021a. Landsat 4-7 Surface Reflectance Code LEDAPS Product Guid [WWW Document]. URL https:// www.usgs.gov/media/files/landsat-4-7-surface-reflectance-code-ledaps-product- guide (accessed 5.19.21).

USGS, 2021b. Landsat 8 Surface Reflectance Code LaSRC Product Guid [WWW Document]. URL : https:// www.usgs.gov/media/files/landsat-8-surface-reflectance-code-lasrc-product-guide (accessed 5.19.21).

Van Breemen, N., Buurman, P., 1998. Soil Formation. Springer, Dordrecht.

Van Wambeke, A., Eswaran, H., Herbillon, A.J., Comerma, J., 1983. Chapter 9 Oxisols, in: Developments in Soil Science. Elsevier, pp. 325–354. https://doi.org/10.1016/S0166-2481(08)70620-2.

Vaysse, K., Lagacherie, P., 2015. Evaluating digital soil mapping approaches for mapping globalsoilmap soil properties from legacy data in languedoc-roussillon (France). Geoderma Reg. 4, 20–30. https://doi.org/10.1016/J.GEODRS.2014.11.003.

Vettori, L., 1969. Métodos de análise de solo-Boletim técnico n° 7. Rio de Janeiro: Equipe de Pedologia e Fertilidade do Solo, 1969., Rio de Janeiro.

Wadoux, A.M.J.C., Minasny, B., McBratney, A.B., 2020. Machine learning for digital soil mapping: applications, challenges and suggested solutions. Earth Sci. Rev. 210, 103359. https://doi.org/10.1016/J.EARSCIREV.2020.103359.

White, R.W., Palin, R.M., Green, E.C.R., 2017. High-grade metamorphism and partial melting in Archean composite grey gneiss complexes. J. Metam. Geol. 35, 181–195. https://doi.org/10.1111/JMG.12227.

Wilford, J., 2012. A weathering intensity index for the Australian continent using airborne gamma-ray spectrometry and digital terrain analysis. Geoderma 183–184, 124–142. https://doi.org/10.1016/j.geoderma.2010.12.022.

Wiriyakitnateekul, W., Suddhiprakarn, A., Kheoruenromne, I., Smirk, M.N., Gilkes, R.J., 2007. Iron oxides in tropical soils on various parent materials. Clay Miner. 42, 437–451. https://doi.org/10.1180/CLAYMIN.2007.042.4.02.

Yang, H., Zhang, X., Xu, M., Shao, S., Wang, X., Liu, W., Wu, D., Ma, Y., Bao, Y., Zhang, X., Liu, H., 2020. Hyper-temporal remote sensing data in bare soil period and terrain attributes for digital soil mapping in the Black soil regions of China. Catena (amst) 184, 104259. https://doi.org/10.1016/J.CATENA.2019.104259.