Mapping Brazilian soil mineralogy using proximal and remote sensing data

Nícolas Augusto Rosin, José A.M. Demattê, Raul Roberto Poppiel, Nélida E.Q. Silvero, Heidy Rodriguez-Albarracin, Jorge Tadeu Fim Rosas, Lucas Tadeu Greschuk, Henrique Bellinaso, Budiman Minasny, Cecile Gomez, et al.

Mapping Brazilian soil mineralogy using proximal and remote sensing data

Nícolas Augusto Rosin [a], José A.M. Demattê [a,*], Raul Roberto Poppiel [a], Nélida E.Q. Silvero [a], Heidy S. Rodriguez-Albarracin [a], Jorge Tadeu Fim Rosas [a], Lucas Tadeu Greschuk [a], Henrique Bellinaso [a,b], Budiman Minasny [c], Cecile Gomez [d], José Marques Júnior [e], Kathleen Fernandes [e]

[a] Department of Soil Science, Luiz de Queiroz College of Agriculture (ESALQ), University of São Paulo (USP), Piracicaba, São Paulo zip code: 13418-900, Brazil
[b] Coordination of Integrate Technical Assistance of Secretariat of Agriculture and Supply - CATI/SAA, Piracicaba Regional, Campos Salles Street, 507 - Piracicaba, São Paulo State, zip code: 13400-200, Brazil
[c] School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW 2006, Australia
[d] LISAH, Univ. Montpellier, IRD, INRAE, Institut Agro, Montpellier, France
[e] Department of Agricultural Sciences at School of Agricultural and Veterinarian Sciences of the São Paulo State University (Unesp) in Jaboticabal, SP, Brazil

ARTICLE INFO

ABSTRACT

Minerals control many soil functions and play a crucial role in addressing global existential issues. Measuring the abundance of soil minerals is a laborious, costly, and time-consuming task; however, soil spectroscopy can be a useful tool to overcome this issue. This work aimed to map the abundance of major mineralogical components of soils in Brazil from surface to 1 m deep and at a spatial resolution of 30 m. Spectral data of the Brazilian Soil Spectral Library with Vis-NIR-SWIR was used to estimate the abundance of haematite, goethite, kaolinite, and gibbsite. These minerals were spatialized using digital soil mapping techniques. We also developed a novel framework to obtain bare soil reflectance for areas without natural or anthropic soil exposure (continuous image) and used it as covariate. Soil minerals and their abundances were successfully estimated by Vis-NIR-SWIR reflectance. Haematite predictions presented the most accurate results with Random Forest models, followed by gibbsite, kaolinite, and goethite. The spatial validation with reference mineralogical data found R^2 of 0.64 (haematite), 0.40 (goethite), 0.20 (kaolinite/Kt), 0.29 (gibbsite/Gbs), and 0.40 (Kt/Kt + Gbs). The resulting maps of soil minerals were in accordance with the geology, pedology, climate, and relief of Brazil and revealed the spatial distribution of mineral abundances at a finer resolution than existing geological and pedological maps, reaching a farm level detail.

1. Introduction

Soil minerals are the "gear" of many soil processes, such as carbon storage, nutrient and water fluxes, playing a fundamental role for humans' welfare by providing food and fuel, regulating nutrients and water cycling as well as contributing to climate change mitigation. Soil mineralogy controls the soil cation exchange capacity, soil aggregation, and structure, organic matter dynamics, water retention, adsorption of phosphorus, among others (de Oliveira et al., 2020; Gilkes and Prakongkep, 2016; Hassink, 1997; Heuvelink et al., 2021; Santos et al., 2017; Zhao et al., 2017). Soil minerals and soil organic carbon (SOC) levels have a close relationship with the potential of the soil to sequester

C by forming organic-mineral complexes (Oades, 1988, Torn et al., 1997, Baldock and Skjemstad, 2000, Singh et al., 2018). Iron oxides and clay minerals, such as haematite, goethite, kaolinite, and gibbsite control C stability and storage in tropical soils (Kirsten et al., 2021). Soil mineralogy also influences emission and fixation of nitrous oxides (Barrón et al. 2020). Thus, the understanding of soil mineralogy is key to soil genesis and geochemical processes (Macías and Camps-Arbestain, 2020).

Soil is the result of the interaction of forming factors: climate, organisms, relief, parent material, and time (Jenny, 1941). Several factors determine soil formation processes and soil mineralogy composition (Heimsath et al., 2012). The extensive Brazilian territory has diverse
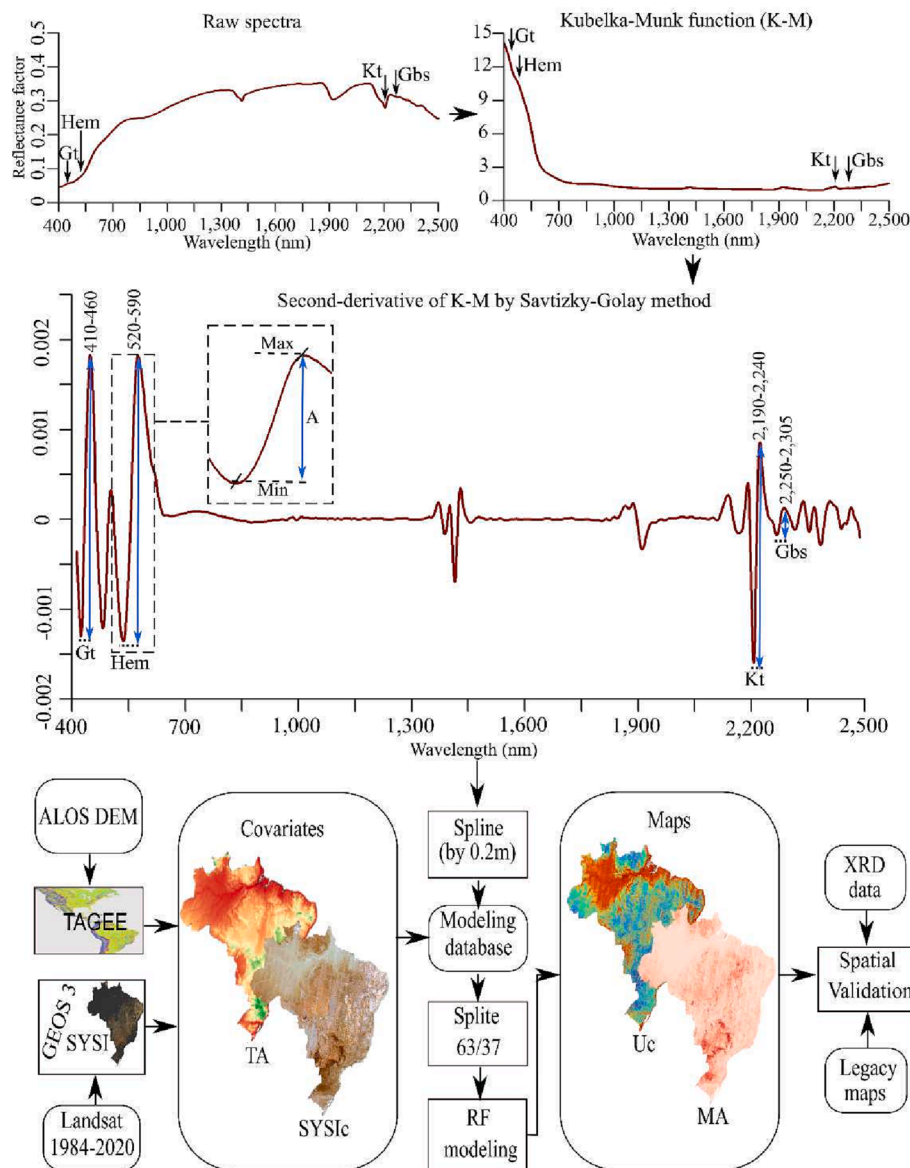
Fig. 1. Flowchart of mineral amplitude (MA) and spatialization procedures. Gt = goethite; Hem = haematite; Kt = kaolinite; Gbs = gibbsite; Max = maximum; Min = minimum; DEM = digital elevation model; TAGGE = Terrain Analysis in Google Earth Engine; GEOS3 = Geospatial Soil Sensing System; SySI = synthetic soil image; SySIc = synthetic soil image full coverage; Uc = uncertainty; RF = Random Forest; XRD = X-ray diffraction.

geology (Gómez et al., 2019), varied climate (Alvares et al., 2013), topography and biodiversity (Gomes et al., 2019), resulting in a diversity of soil types (Santos et al., 2018; Instituto Brasileiro de Geografia e Estatística (IBGE), 2021). However, information on soil mineralogy for the entire Brazilian territory does not exist. Maps of soil classes available for the Brazilian territory are mostly at a coarse scale (Lepsch, 2013; Nolasco de Carvalho et al., 2015), while soil maps at the semi-detailed scale are almost non-existent, for example, there are no detailed soil maps in the Northern region, where the Amazon rainforest is located (Mendonça-Santos and dos Santos, 2006).

Traditional methods for mineralogical characterization using X-ray diffraction (XRD) are laborious, expensive, time-consuming, and only provide qualitative or semi-quantitative information (Bahia et al., 2015; Chipera and Bish, 2001; Fang et al., 2018; McManus, 1991; Viscarra Rossel, 2011; Whittig, 1965; Kunze and Dixon, 1986). Consequently, mineralogical evaluation is restricted to scientific research and not available to direct users, such as farmers (Rosin et al., 2022). Additionally, the mineralogical analysis is not commercially available and obtaining mineralogical data for the entire Brazilian territory is even more challenging.

Therefore, a cost-effective method for soil mineralogical assessment becomes essential (Fang et al., 2018; Viscarra Rossel, 2011). Remote and proximal sensing techniques, specifically diffuse reflectance spectroscopy in the visible, near-infrared, and shortwave infrared (Vis-NIR-SWIR) ranges provide a fast, non-destructive, and environmentally friendly method for soil characterization (Nocita et al., 2015; Soriano-Disla et al., 2014). The 350–2500 nm spectral range is commonly used to estimate several soil attributes (Ng et al., 2019; Soriano-Disla et al., 2014; Viscarra Rossel et al., 2006), such as soil particle size distribution (Coblinski et al., 2020; Silva et al., 2019), SOC (Barthès et al., 2019; Moura-Bueno et al., 2020) and soil chemical properties (Vaudour et al., 2018; Wadoux et al., 2019; Zhao et al., 2021). According to Bahia et al. (2015), while conventional methods require more than 55 h per sample to obtain the iron oxide minerals contents, while Vis-NIR-SWIR spectroscopy require only about 20 min (time estimated for scanning the sample and all data processing procedures).

The Vis-NIR-SWIR range has specific wavelengths related to overtone and vibrational transition caused by the interaction of energy with
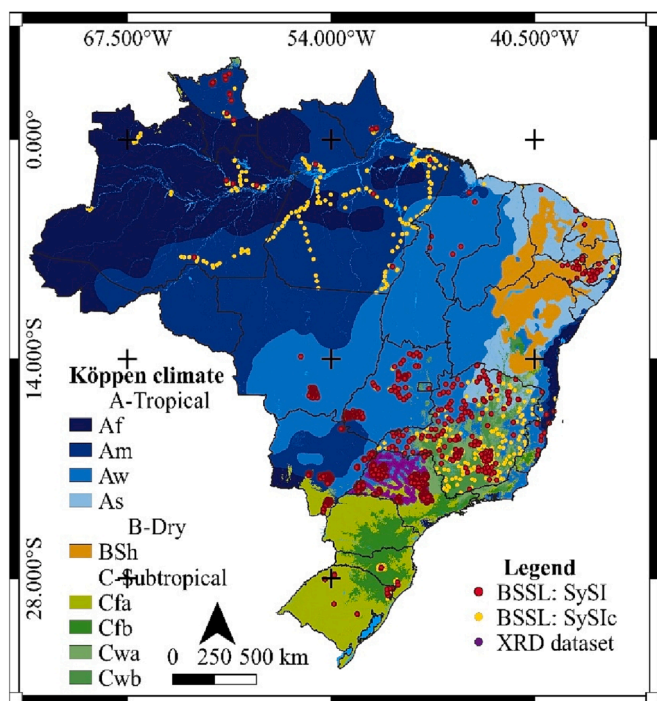
Fig. 2. Study site and soil observation points for 0–0.2 m depth of The Brazilian Soil Spectral Library (BSSL) and X-ray diffraction (XRD) dataset. BSSL: SySI = BSSL points with Synthetic soil image; BSSL: SySIc = BSSL points with predicted reflectance bands; A = Tropical zone, without dry season (Af), monsoon (Am), dry winter (Aw) or dry summer (As); B = Dry zone with semi-arid climate and low latitude and altitude (BSh) C. Subtropical zone with oceanic climate, without dry season, with hot summer (Cfa) or with temperate summer (Cfb) or subtropical zone with dry winter and hot summer (Cwa) or temperate summer (Cwb).

clay minerals and iron and aluminum oxides in the soils (Fang et al., 2018). The use of the spectral preprocessing method, such as the Kubelka-Munk function combined with the second derivative, can enhance spectral features related to soil minerals (Barrón and Torrent, 1986; Scheinost et al., 1998). In general, intensity of the band amplitude calculated from derivative values between minima and maxima at specific absorption features is proportional to the mineral amount in the sample (Kosmas et al., 1984; Mendes et al., 2021).

Several studies have investigated the fundamentals of the influence of soil mineralogy on the Vis-NIR-SWIR spectra (Barrón and Torrent, 1986; Clark et al., 1990; Madeira-Neto et al., 1995; Scheinost et al., 1998; Fernandes et al., 2004; Heller Pearlshtien and Ben-Dor, 2020; Sahwan et al., 2021). Some researchers developed equations and models to quantify the mineral contents by linking amplitudes or intensities of specific wavelengths and XRD data (Bahia et al., 2015; Canton et al., 2021; Fernandes et al., 2004; Madeira-Neto et al., 1995; Silva et al., 2020; Fernandes et al., 2020; Mendes et al., 2021). Mineral indices and relative mineral abundances derived from infrared spectral bands can also be mapped across the region (Fernandes et al., 2020; Mendes et al., 2021; Poppiel et al., 2020; Ramos et al., 2020; Viscarra Rossel et al., 2010; Viscarra Rossel, 2011) as well as mineral contents (Silva et al., 2020). Thus, spectroscopy becomes an important tool to obtain quantitative and spatial information on soil mineralogy.

Using the Vis-NIR-SWIR spectra and geostatistical techniques, the abundance of iron oxides in the soil were accessed and mapped across Australia (Viscarra Rossel et al., 2010). Environmental covariates representing the soil formation factors were used to digitally map kaolinite, illite, and smectite in Australian soils (Viscarra Rossel, 2011). For the western region of São Paulo State in Brazil, Silva et al. (2020) mapped the contents of haematite and goethite, while Fernandes et al. (2020)

obtained the kaolinite-gibbsite ratio for the same region. Ramos et al. (2020) mapped the haematite-goethite ratio for Rio Grande do Sul State. Mulder et al. (2013) mapped mica, kaolinite, smectite, and calcite for an area of 15,000 km² in northern Morocco, combining XRD and Vis-NIR-SWIR with ASTER data.

Remote sensing using infrared spectra has been used to identify direct mineral on planet Mars (Bibring et al., 2006; Poulet et al., 2005). However, the use of optical remote sensing of Vis-NIR-SWIR bands on the earth's soil surface is challenged by atmospheric conditions, clouds, and vegetation cover. Demattê et al. (2018) created a synthetic soil image (SySI) to obtain bare soil pixels from a time series of Landsat satellites from 1985 to present. The SySI bands can be used as predictors of soil minerals. For example, Poppiel et al. (2020) mapped the abundances of haematite, goethite, kaolinite, gibbsite, and 2:1 minerals for Goiás State in Brazil. Moreover, Mendes et al. (2021) mapped several key soil minerals at a regional scale in Brazil. Therefore, bare soil reflectance data (Demattê et al. 2018) can be considered an innovative and promising covariate, due to its strong correlation with soil attributes (Bellinaso et al., 2021; Demattê et al., 2018; Fongaro et al., 2018; Silvero et al., 2021). Nevertheless, SySI image can only be used in locations with naturally exposed soil or exposed through anthropic activities (cropping lands), which is a major limitation of its use.

This study aimed to map the abundance of soil minerals, such as haematite, goethite, kaolinite, and gibbsite for the entire Brazilian territory. The hypothesis is that soil mineral abundances estimated from the Vis-NIR-SWIR spectra can be spatialized to large areas using bare soil reflectance plus other covariates. We developed a novel framework to obtain a synthetic soil image with full coverage (SySIc) for bare soil reflectance as a covariate for the entire land area of Brazil. This study mapped the abundances of major soil mineralogical components at soil surface and subsurface with a spatial resolution of 30 m, which were validated with ground truth observations.

2. Materials and methods

The first step of this work consisted of estimating the relative abundances of haematite, goethite, kaolinite, and gibbsite using point observations of Vis-NIR-SWIR spectra from the Brazilian Soil Spectral Library (BSSL) (Demattê et al. 2019). The second step comprised the SySIc creation. For the third step, the random forest model was used to link estimates of observation points of mineral abundance with SySIc and terrain attributes (TA) via the soil spatial prediction function. The spatial function was applied to the entire Brazilian territory to obtain maps of soil minerals and their uncertainties. Finally, the maps were validated with reference mineralogical data and compared with legacy soil, geology maps, and climate and terrain conditions at the national, regional, and farm levels. The flowchart of mineral amplitude (MA) calculation and spatialization procedures are given in Fig. 1.

2.1. Study site

The study site covered the entire Brazilian territory, with approximately 8.8 million km² (Fig. 2), comprising a high diversity of climate, relief, geology, and soil types. The climate is divided into tropical (81.4%), dry (4.9%), and subtropical (13.7%) zones, according to the Köppen classification (Alvares et al., 2013) (Fig. 2). The relief is composed mainly of slightly sloping land and altitudes between 200 and 400 m by low tectonic and volcanic activity and erosion processes (Ross, 2013). The Brazilian landscapes are composed main by plateaus, plains and depressions (Ross, 2013). The geology is composed of several types of metamorphic, igneous, and sedimentary rocks (Gómez et al., 2019). The main soil classes of the country are Latossolos (Ferralsols) and Argissolos (Acrisols/Lixisols/Alisols), which comprise more than 60% of the territory (Instituto Brasileiro de Geografia e Estatística (IBGE), 2021). Maps of the soil, geology, altitude, and slope are available in Supplementary Material 1.

Table 1
Selected bands for calculation of mineral amplitude (MA).

Soil mineral	Minimum band in literature (nm)	Maximum band in literature (nm)	Selected range (nm)	Reference
Haematite	~535	~580	520–590	Scheinost et al. (1998)
Goethite	~415	~455	410–460	Scheinost et al. (1998)
Kaolinite	2205	2225	2,190–2,240	Clark et al. (1990) and Madeira-Neto et al. (1995)
Gibbsite	2265	2295	2,250–2,305	Clark et al. (1990) and Poppiel et al. (2020)

Brazil also has great biodiversity with six distinct biomes (Amazonia, Cerrado, Mata Atlântica, Caantinga, Pantanal, and Pampa), with Amazonia (49.29%) and Cerrado (22%) covering most of the territory (Gomes et al., 2019). The Amazonia biome is a humid tropical forest with high biodiversity, mean annual rainfall higher than 3,100 mm, and mean annual temperature between 25.9 and 27.7 °C. The Cerrado biome is a semi-humid savanna with mean annual precipitation between 1,200 and 1,800 mm and mean annual temperature between 22 and 23 °C. The Mata Atlântica biome comprises the highest diversity of environments under tropical and subtropical climates, with mean annual precipitation between 700 and 1,500 mm and mean annual temperature between 11 and 26 °C. The Caatinga biome is a semi-arid savanna, with annual average rainfall of about 500 mm and mean annual temperature between 20 and 29 °C. The Pantanal has long flooding periods, with mean annual precipitation between 1,000 and 1,600 mm and mean annual temperature between 22 and 24 °C. The Pampa biome has subtropical climate and is covered by temperate grasslands, with mean annual precipitation between 1,300 and 2,500 mm and mean annual temperature between 14 and 20 °C (Gomes et al., 2019).

2.2. Soil observations

A legacy database with spectral data in the Vis-NIR-SWIR range (from 350 to 2500 nm), with spectral resolution of 1 nm, from the BSSL (Demattê et al., 2019), accessible at https://besbbr.com.br, was used in this study, generating 30,334 soil samples with geographical coordinates (Fig. 2) (Supplementary Material 2). Most BSSL observations were from 0 to 0.2 m (n = 10,306), 0.4–0.6 m (n = 7,676), and 0.8–1 m (n = 7,697) depth intervals and the remaining (n = 7,264) from various depths of soil horizons.

A second legacy database with values of haematite, goethite, kaolinite, and Kt/(Kt + Gbs) obtained by X-ray diffraction (XRD) of the clay faction was used to validate the digital mineral maps. This database comprises 221 samples from 0 to 0.2 m layer with georeferenced points in São Paulo State (Fig. 2) obtained by Silva et al. (2020) and Fernandes et al. (2020).

2.3. Calculation of mineral abundances and data harmonization

The raw reflectance spectra of all 30,334 samples (Supplementary Material 2) were transformed to absorbance using the Kubelka-Munk function (KM) (Barrón and Torrent, 1986) (Fig. 1). Since the KM spectra had overlapping bands, we calculated their 2nd derivative using the Savitzky-Golay method (SG) (Savitzky and Golay, 1964) to resolve and enhance the spectral features of interest. The parameters used in SG were: differentiation order (m) = 2; polynomial order (p) = 3; window size (w) = 25 and the output data ranged from 412 to 2478 nm. The intervals related to haematite, goethite, kaolinite, and gibbsite were defined in the SG spectra based on the literature (Table 1). Thus, the MA for each mineral was calculated as: MA = Maxλ - Minλ, where Maxλ and Minλ are the maximum (positive) and minimum (negative) values of the specific ranges.

The MA data from various depths of observations were harmonized by spline interpolation, due its accuracy (Bishop et al., 1999; Malone et al., 2009), using the GSIF R package (Hengl and MacMillan, 2019). The output data were divided into five layers by 0.2 m from surface up to 1 m depth (0–0.2, 0.2–0.4, 0.4–0.6, 0.6–0.8 and 0.8–1 m). After the spline interpolation procedure, we had 12,000 observation points for the 0–0.2 m layer (about 1.41×10^{-3} points per km^2) to 8,000 for the 0.8–1 m layer (about 9.41×10^{-4} points per km^2) (Supplementary Material 2).

2.4. Environmental covariates

2.4.1. SySI and TA

The environmental covariates used were SySI and terrain attributes (TA), at 30 m of spatial resolution (Fig. 1). SySI represents the following SCORPAN factors (McBratney et al., 2003): soil (directly), parent material (indirectly), and age (indirectly), while TA represents the relief (directly) (Poppiel et al., 2020). It is possible to infer soil attributes in the subsurface using soil reflectance of surface soils (Mendes et al., 2019). Mendes et al. (2021) demonstrated the use of SySI as a predictor for MA mapping at soil depths up to 1 m. Thus, we used SySI as a covariate also for subsurface layers.

SySI was obtained using the collection of Landsat images from 1984 to 2020 (Landsat 5, 6, 7 and 8) through the GEOS3 method (Demattê et al., 2018, 2020) in Google Earth Engine (GEE) (Gorelick et al., 2017). SySI had the same number of satellite bands used as input data for GEOS3, in this case (Landsat), six bands: Blue (450–520 nm), Green (520–600 nm), Red (630–690 nm), NIR (760–900 nm), SWIR1 (1550–1750 nm), and SWIR2 (2080–2350 nm). The GEOS3 uses the soil spectral trend and the Normalized Difference Vegetation Index (NDVI) and Normalized Burn Ratio 2 (NBR2) to create the soil mask for all satellite images (Demattê et al., 2018). Thus, bare soil pixels were selected when there was an increase in reflectance from Blue to SWIR1, and in NDVI between −0.15 and 0.25 and in NBR2 between −0.15 and 0.15. SySI pixels are the median of all bare soil pixels detected across the time series images.

A total of 13 terrain attributes (TA) were obtained from the digital elevation model (DEM) of Advanced Land Observing Satellite (ALOS) (Japan Aerospace Exploration Agency, 2021) available in GEE (Golerick et al., 2017), using the package terrain analysis in GEE (TAGEE) (Safanelli et al., 2020a).

The SySI and TA values were extracted for the observation points. The spatial predictors were selected based on the Spearman correlation with MA. The Spearman correlation >0.10 or <-0.10 was the criterion to select covariates for modeling. Thus, the covariates selected were all bands from SySI: Blue, Red, Green, NIR, SWIR1, and SWIR2 and four of TA: elevation, slope, maximal curvature, and hillshade. Further information on TA covariates can be found in Safanelli et al. (2020a). The results of the Spearman correlation analyses can be found in Table A1.

2.4.2. SySIc

SySI is directly related to spectral data obtained at the laboratory level and proved to be a powerful tool for spatial prediction of soil attributes (Bellinaso et al., 2021; Demattê et al., 2018; Fongaro et al., 2018; Poppiel et al., 2021, 2020; Safanelli et al., 2021a; 2021b; 2020b; Silvero et al., 2021). However, the GEOS3 method is applicable only in locations with naturally exposed soil or exposed to anthropic activities (*i.e.*, soil tillage). For the Brazilian territory, which has vast areas with native vegetation (*i.e.*, Amazon rainforest), SySI covers only roughly 30% of the entire territory. Therefore, to obtain continuous maps of soil mineralogy for the entire Brazilian territory, SySIc was created using environmental covariates and machine learning algorithms (Fig. 2). The covariates include TA, as mentioned in section 2.4.1 (elevation, slope,
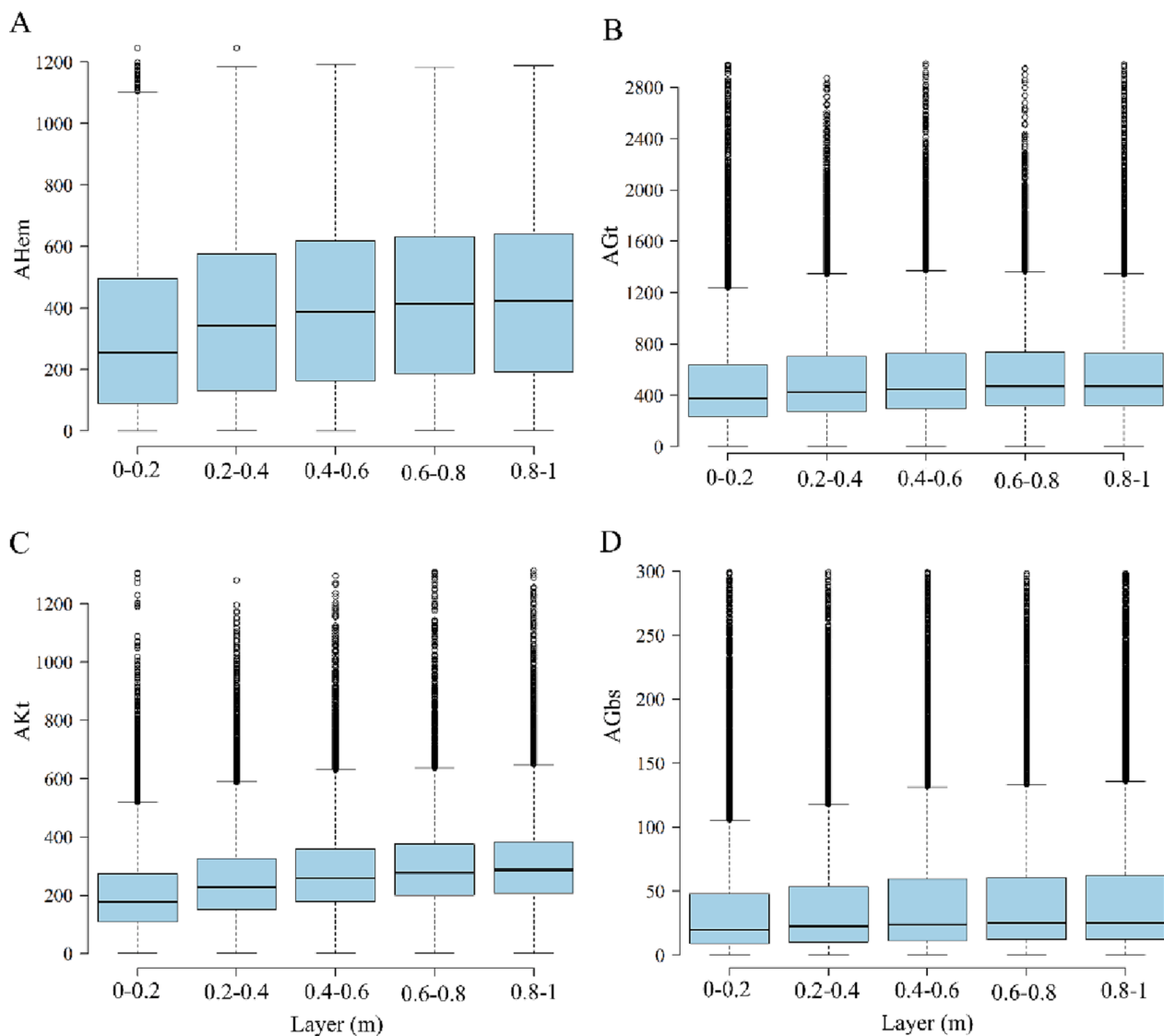
Fig. 3. Boxplot for haematite (AHem) (A), goethite (AGt) (B), kaolinite (AKt) (C), and gibbsite (AGbs) (D) amplitudes after spline procedure.

maximal curvature, and hillshade) and spectral bands (Blue, Green, Red, NIR, SWIR1, and SWIR2) of a median cloud-free mosaic, obtained from a time series of Landsat 5 satellite images from 1984, when natural vegetation cover was larger in the study site.

Random sampling was carried out in the extent of SySI intersected with TA and Landsat mosaic layers, generating 85,740 points. The correlation between covariates (Landsat mosaic and TA) and SySI was obtained as described in section 2.4.1 also available in Table A.1. The modeling was performed on each SySI band as a dependent variable and Landsat mosaic bands and TA as independent variables, resulting in six models. The modeling procedure, model evaluation (Table A.2), and spatial bootstrap predictions were performed, as described in section 2.5. Finally, SySI original bands were overlapped with the six bare soil predicted bands and then merged in a unique raster file, the SySIc. Thus, pixels with bare soils remained, while the "gaps" were filled with the predicted bands. Here, we only used observations with the actual bare soil reflectance pixels (SySI image) to build the MA models and the SySIc image to predict MA for the entire Brazilian territory.

2.5. Spatial prediction of soil minerals

The MA models were fitted using only points with SySI, with about 9,600 observation points for 0–0.2 m layer (about 1.09×10^{-3} points

per km^2) and with roughly 7,300 points for 0.8–1 m layer (about 8.29×10^{-4} points per km^2) (Supplementary Material 2). The selected environmental covariates (SySI and TA) were used as independent variables to obtain spatial models for amplitudes of haematite (AHem), goethite (AGt), kaolinite (AKt), and gibbsite (AGbs) for 0–0.2, 0.2–0.4, 0.4–0.6, 0.6–0.8, and 0.8–1 m depths. We used the Random Forest (RF) machine learning algorithm (Breiman, 2001), a popular and accurate method for soil attribute mapping (Padarian et al., 2020). We used the scikit-learn package in Python (Pedregosa et al., 2011) to implement a bootstrapping routine (Efron and Tibshirani, 1993) for training and testing the RF model, as described by Safanelli et al. (2021a; b). After this step, the spatial prediction was performed in GEE (Gorelick et al., 2017).

A grid search procedure was performed for each mineral and depth to select the best hyperparameters combination, aiming to reduce overfitting possibilities. We trialed a range of values for number of trees (FS) (30, 60, 100, 200, and 500 trees), number of predictors in tree splits (nRP) (3, 5, 8, and 10) and minimum number of observations at leaves (minSL) (10, 20, 30, 40, 50, 100, 200, and 500). The optimal combination (FS, nPR, and minSL) with the least root mean square error (RMSE) for the validation set was selected for each mineral. However, when the RMSE difference between the hyperparameters combination was less than 1×10^{-5}, the combination with lower FS value was chosen to make the spatial prediction faster.
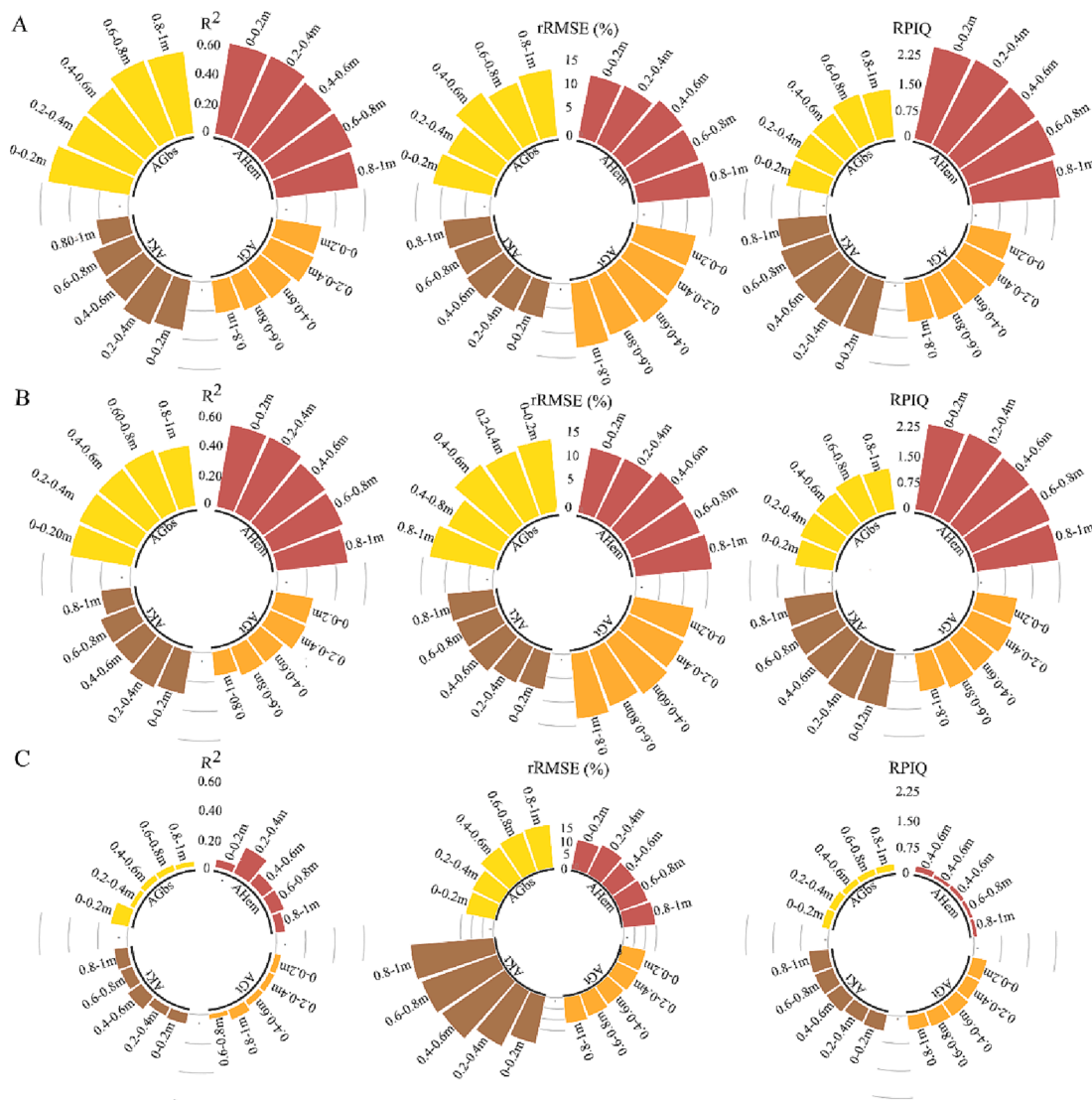
Fig. 4. Performance of Random Forest (RF) models for training set (A), validation set (B) and points with only predicted reflectance (synthetic soil image full coverage/SySIc) (C). R^2 = coefficient of determination; rRMSE = relative root mean square error; RPIQ = ratio of the performance to interquartile distance; AHem = haematite amplitude; AGt = goethite amplitude; AKt = kaolinite amplitude; AGbs = gibbsite amplitude.

The calibration set was composed of bootstrapped samples randomly sampled with replacement (63% of samples) for each regression tree, while observations that were not sampled (37% of samples) were used for validation. The bootstrapping procedure was performed from 30 to 500 times in each hyperparameter combination tested, similar to each FS tested, because one bootstrapping procedure was performed for each regression tree.

The accuracy parameters evaluated were coefficient of determination (R^2), RMSE, relative RMSE (rRMSE), and ratio of performance to interquartile distance (RPIQ) of calibration and validation datasets. The R^2 represents the variance explained, RMSE represents the accuracy, while RPIQ evaluates the error over the interquartile range of the data. The rRMSE was calculated as follow: rRMSE (%) = RMSE/(MaxMA - MinMA)*100, where MaxMA and MinMA are the maximum and minimum MA values for each mineral in the dataset after spline interpolation.

The spatial prediction was performed using the bootstrapped models in GEE (Gorelick et al., 2017) to obtain the MA maps and their uncertainties. The prediction was carried out using the SySIc image and TA to obtain continuous maps for the entire Brazilian territory. The final MA maps were obtained by the mean of 50 bootstrap predictions at 30 m

spatial resolution. The uncertainty maps were obtained from the 90% prediction interval (PI90 = 0.95Q – 0.5Q), where Q is the prediction quartile. The uncertainty maps were obtained with less detailed spatial resolution (250 m), to reduce the computational time.

The MA maps were also validated in locations without SySI, using part of the BSSL dataset with points located outside the areas of soil exposure (Supplementary Material 2). The statistics parameters also included R^2, RMSE, rRMSE, and RPIQ.

2.6. Data interpretation

First, the MA maps were stratified in three groups based on their percentiles to describe the distribution of MA values in the Brazilian territory:

1) less occurrence = samples with MA value in the lower percentile;
2) moderate occurrence = samples with MA value in the intermediary percentile;
3) high occurrence = sample with MA value in the upper percentile.

Second, the Hem/(Hem + Gt) and Kt/(Kt + Gbs) ratios were calculated from the MA maps. These ratios were frequently used in soil mineralogy studies to infer soil genesis (Schaefer et al., 2008). Further
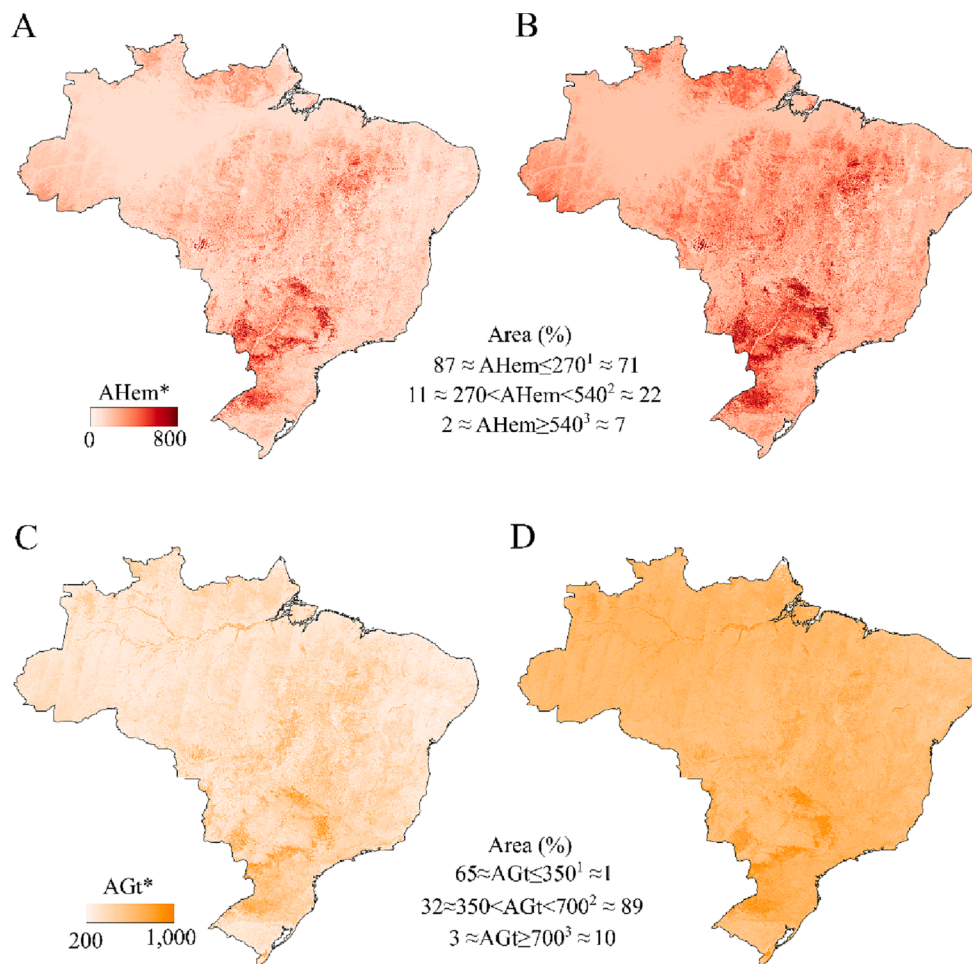
Fig. 5. Maps of haematite amplitude (AHem) for 0–0.2 m (A) and 0.8–1 m (B) and goethite amplitude (AGt) for 0–0.2 m (C) and 0.8–1 m (D) depths. *10^{-6} scale; [1]low occurrence; [2]moderate occurrence; [3]high occurrence.

information can be found in Supplementary Material 3.

Third, a spatial ternary plot was made to evaluate the distribution of the minerals, which showed in the Red band = iron oxides (Hem + Gt), Green = kaolinite, and Blue = gibbsite. Further information also is in Supplementary Material 3.

2.7. Spatial validation and study cases

The spatial validation was performed using an external georeferenced soil observation dataset with haematite, goethite, kaolinite, gibbsite, and the ratio Kt/(Kt + Gbs) data obtained by conventional methods. The values of MA and Kt/(Kt + Gbs) for 0.0–0.2 m were extracted for this dataset and the R^2, RMSE, and rRMSE were evaluated.

To assess spatial patterns of the digital maps, we compared MA, ratios, and ternary plot maps visually with 1:5 000 000 geology (Gómez et al., 2019) and 1:1 000 000 soil (Instituto Brasileiro de Geografia e Estatística (IBGE), 2021) maps and with climate (Alvares et al., 2013), elevation, and slope (Japan Aerospace Exploration Agency, 2021) patterns. Several case studies followed by visual evaluation were carried out and detailed information can be found in Supplementary Material 3. The legacy geological and pedological maps are available only at small scales for the entire Brazilian territory and were used to verify the mineralogy patterns at national and regional levels. Soil maps and other maps with detailed resolution were used to evaluate the MA maps at the finest scale, which is compatible with the degree of mineralogy information obtained (local and farm level).

3. Results

3.1. Descriptive statistics of MA data

The MA, which reflects the abundance of minerals in soil samples, increased with soil depth (Fig. 3) (Supplementary Material 2). The mean for AHem and AGt ranged from 318 to 433 ($\times 10^{-6}$) and from 506 to 597 ($\times 10^{-6}$), respectively, at surface and at 1 m depth. AKt and AGbs showed mean values from 211 to 317 ($\times 10^{-6}$) and from 51 to 50 ($\times 10^{-6}$) at surface and at 1 m depth, respectively. AGbs values showed stability above 0.40 m, while a more pronounced increase in deeper layers was verified for AKt, followed by AGt. The variation with depth can be related to soil types, because soils in Brazil have uniform profiles, such as Ferralsols, Plinthosols, and Nitisols, and with textural gradients, such as Acrisols, Lixisols, and Alisols (Supplementary Material 1).

A large variation in MA was observed (Fig. 3) (Supplementary Material 2). AGbs and AHem presented high variations, which can be related to specific conditions of relief, climate, geology, and pedology (Fig. 2) (Supplementary Material 1), with AGt and AKt more stable. AGbs has the highest coefficient of variation (CV) with values ranging from 116 to 127 %. Iron oxides showed intermediate values of CV ranging from 62 to 81% for AHem and from 69 to 82% for AGt. AKt showed the smallest values of CV, ranging from 55 to 69 %. Regarding soil depth, the highest CV values were verified for 0.0–0.2 m depth, with MA values showing lower variations in deep layers. The full summary of the statistics can be found in Supplementary Material 2, which also shows the raw MA dataset (before spline), points with SySI (modeling
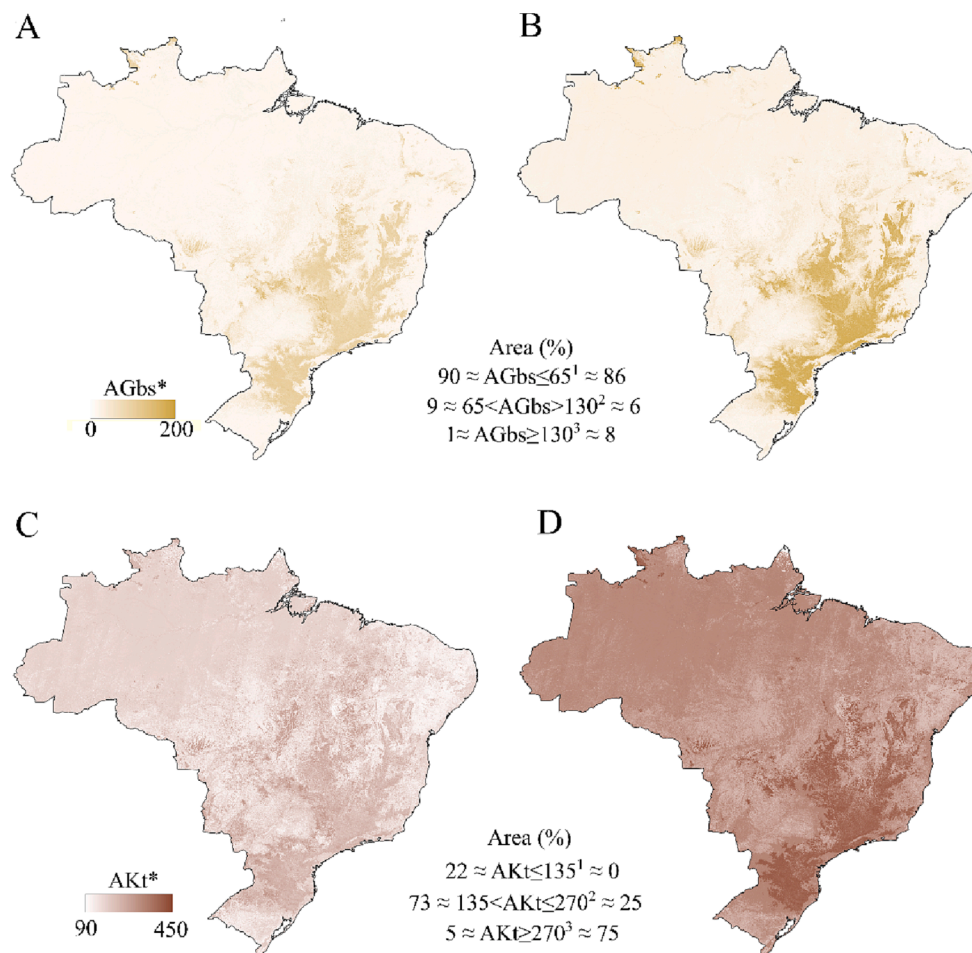
Fig. 6. Maps of gibbsite amplitude (AGbs) for 0–0.2 m (A) and 0.8–1 m (B) and kaolinite amplitude (AKt) (AGt) for 0–0.2 m (C), and 0.8–1 m (D) depths. *10^{-6} scale; [1]low occurrence; [2]moderate occurrence; [3]high occurrence.

dataset), and points without soil exposure (only SySIc image).

3.2. Performance of RF models for MA prediction and SySIc validation

Higher R^2 and smaller error in the validation set were found in the upper layers, decreasing accuracy with increasing depth (Fig. 4ab) (Supplementary Material 4). AHem reached R^2 for the validation set, ranging between 0.48 (deeper layer) and 0.56 (upper layer). RMSE ranged from 171 to 189 (x10^{-6}), rRMSE from 13.73 to 15.93 %, and RPIQ from 2.23 to 2.34. AGbs had R^2 values ranging between 0.41 (deeper layer) and 0.44 (layer 0.6 – 0.8 m), with RMSE ranging from 41 to 46 (x10^{-6}), rRMSE from 12.50 to 13.52%, and RPIQ from 1.03 to 1.18 in the validation set. Results were less accurate for AKt and AGt with R^2 values from 0.20 and 0.29 and from 0.16 to 0.25 in the validation data set, respectively. For AKt, RMSE ranged from 100 to 125 (x10^{-6}), rRMSE from 7.67 to 9.52 %, and RPIQ from 1.32 to 1.50, while for AGt, RMSE ranged from 372 to 402 (x10^{-6}), rRMSE from 12.07 to 15.33 %, and RPIQ from 1.03 to 1.19. RMSE values can be found in Supplementary Material 4.

SySIc was predicted with R^2 ranging from 0.25 to 0.36 in the validation set (Table A2). When MA in non-bare soils were predicted with SySIc, accuracy decreased (Fig. 4c) (Supplementary Material 4). Therefore up, R^2, RMSE, rRMSE, and RPIQ ranged from 0.03 to 0.19, from 40 to 499 (x10^{-6}), from 9.87 to 38%, and from 0.08 to 0.53, respectively. The decrease of accuracy in locations with predicted bands was observed for all minerals and depths.

3.3. Maps of mineralogy and uncertainty

The maps of iron oxide revealed areas with higher occurrence of haematite and goethite at 0–0.2 m depth, in the Central-Western, Southwestern, and Southern regions in Brazil (Fig. 5). At this depth, 87, 11, and 2.1% of the area have low, moderate, and high AHem occurrence, respectively. Likewise, 65, 32, and 3% of the area had low, moderate, and high AGt occurrence, respectively.

For the 0.8–1 m depth, there were more areas with greater occurrence of haematite and goethite when compared with the 0–0.2 m layer and this increase was more pronounced for goethite (Fig. 5). In 0.8–1 m layer, 71, 22, and 7% of area for AHem and 1, 89, and 10% of area for AGt showed values of low, moderate, and high ranges, respectively. Thus, nearly all soils in Brazil have considerable levels of goethite in deeper layers.

Only a small part of the area had high gibbsite occurrence up to 1 m depth (Fig. 6). The distribution at 0.8–1 m depth showed that Brazilian soils had 86, 6, and 8% of the area with low, moderate, and high AGbs occurrence, respectively. A significant abundance of kaolinite was observed in Brazilian soils for all depths (Fig. 6). The map showed that 0, 24, and 75% of the area have low, moderate, and high AKt occurrence, respectively. Gibbsite and kaolinite abundances in the 0.8–1 m layer were higher than in the 0–0.2 m layer. For 0–0-2 m depth, 90, 10, and 0% of area have AGbs and 23, 72, 5% of the area have AKt of low, moderate, and high ranges, respectively.

Geology greatly influenced the distribution of iron oxide minerals, with higher values of AGt and AHem related to mafic volcanic rocks, such as basalt (Fig. 5) (Supplementary Material 1). AHem high values
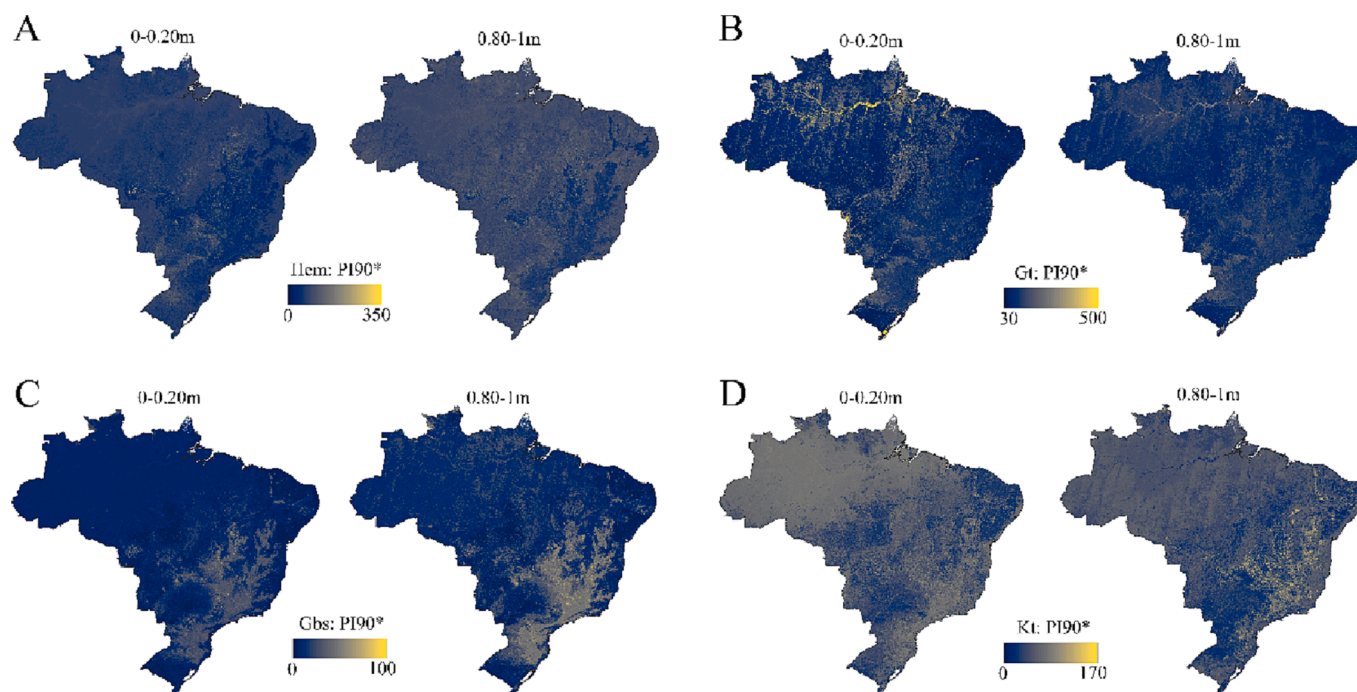
Fig. 7. Uncertainty of spatial prediction maps for haematite (Hem) (A), goethite (Gt) (B), gibbsite (Gbs) (C), and kaolinite (Kt) (D). *10^{-6} scale; PI90 = 90% prediction interval.

were also associated to Ferralsols and Nitosols. AHem high values occurred specifically in some locations under tropical and subtropical zones, with less abundance in dry environments (Fig. 2), while AGt had considerable occurrence in areas with limited drainage conditions (Supplementary Material 1).

High values of AGbs were related to high altitude (greater than 500 m) and granite and gneisses rocks (Fig. 6) (Supplementary material 1). Moderate abundance of gibbsite was observed, associated to iron oxides, in soils derived from volcanic rocks. Gibbsite occurrence was associated to several types of soils, such as Ferralsols, Acrisols, Lixisols, and Cambisols. The mineral map of kaolinite showed considerable abundance in all soil types from several parent materials. The complete summary statistics of predicted maps can be found in Supplementary Material 2.

In general, areas with high abundances of minerals showed higher uncertainty (90PI) in the 50 bootstrap predictions (Figs. 5, 6, and 7). AKt presented higher uncertainty, followed by AGb, AHem, and AGt. The Northern region in Brazil showed higher 90PI values for AKt and AGt. High altitude regions showed greater uncertainty for AGbs (Supplementary material 1), while AHem 90PI values were the most homogeneous in the entire territory (Fig. 7).

3.4. Ratios and spatial ternary plot

The Hem/(Hem + Gt) ratio map is shown in Fig. 8. It shows high ratio values in locations with varied parent material and climate conditions, except in dry zones (Northeast) and subtropical zones with temperate summer (South) (Fig. 8ab and Fig. 2) (Supplementary Material 1). The map shows dominance of goethite in Brazilian soils with similar patterns in the 0–0.2 m and 0.8–1 m layers. However, there were some exceptions, such as higher values for locations in the Western region of São Paulo State and lower values in the Amazon basin for 0.8–1 m compared to 0–0.2 m depths. The Kt/(Kt + Gbs) ratio shows a predominance of kaolinite in relation to gibbsite, with gibbsite occurrence concentrated in the South and Southwest, associated to soils from mafic volcanic rocks (Fig. 8cd) (Supplementary material 1). The ternary plot followed the patterns of each mineral map and showed predominance of kaolinite in Brazilian soils under several conditions (Fig. 8ef)

(Supplementary Material 1). The occurrence of iron oxides is located mainly in some places in the South, Southwest, and Central-West associated to the parent material, with predominance in soils from volcanic mafic rocks, while gibbsite occurred mainly in the South and Southwest, regions at higher altitudes.

3.5. Spatial validation with reference mineralogy data

The MA and Kt/(Kt + Gbs) maps were in accordance with reference values of mineralogy obtained by XRD (Table 2). Results were more accurate for haematite, with R^2 of 0.64, RMSE of 8.72 g kg^{-1} and rRMSE of 13.7 %, followed by goethite with R^2 of 0.40, RMSE of 4.45 g kg^{-1}, and rRMSE of 12.7 % and Kt/(Kt + Gbs) with R^2 of 0.40, RMSE of 0.14, and rRMSE of 17.9%. The worst results were for gibbsite and kaolinite, with R^2 of 0.29, RMSE of 14.41 %, and rRMSE of 19.1% and R^2 of 0.20, RMSE of 14.81 %, and rRMSE of 21.4%, respectively.

3.6. Case studies

The areas with higher iron oxides (haematite and goethite) abundance were related to basalt rocks and more weathered soil classes, such as red (hue equal to 2.5YR or more reddish) Ferralsols and Nitosols, covering the northern region of Rio Grande do Sul State, western region of Paraná State, and southern region of Mato Grosso do Sul State (Fig. 9a). The western region of Paraná State shows higher abundance of haematite and goethite for Ferralsols and Nitosols than for Acrisols, Lixisols and Alisols. The western region of Pará State (North Brazil) presented moderated haematite abundance and high Hem/(Hem + Gt) ratio for red-yellow (hue equals to 5YR) Acrisols, Lixisols and Ferralsols and low haematite and Hem/(Hem + Gt) for yellow (hue equals to 7.5YR or more yellowish) Acrisols, Lixisols, Ferralsols and Gleysol (Fig. 9b). The southern region of Goiás State (Central Brazil) presented moderate haematite abundance in red and clayey Ferralsols (clay content equal to or more than 350 g kg^{-1}), high in red, clayey and ferric Ferralsols (Fe$_2$O$_3$ equals to or more than 18 g kg^{-1}) and low content in Arenosols and other soil classes. Gibbsite abundance in Goiás State was also high in red and clayey Ferralsols, but restricted to flat and elevated
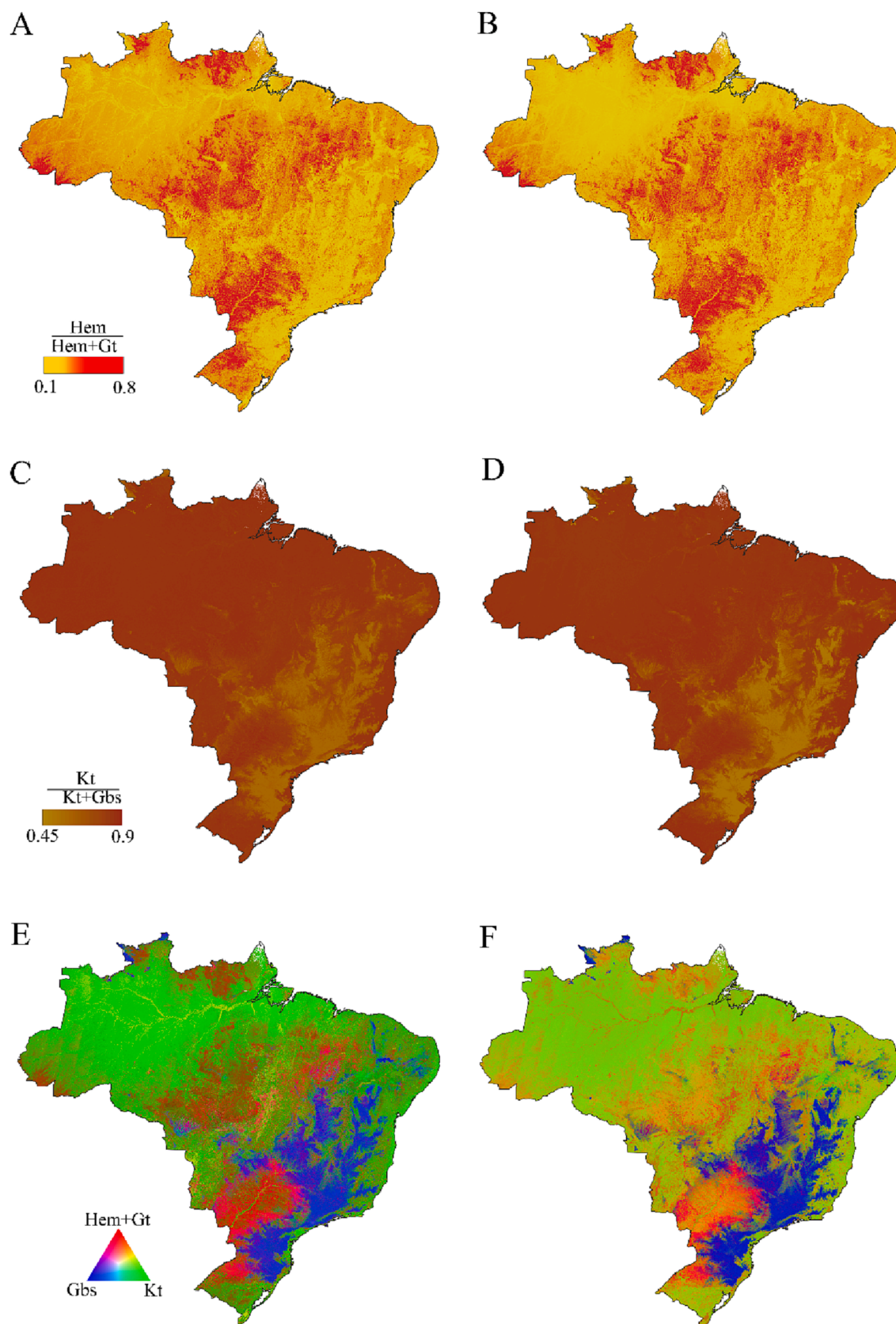
Fig. 8. Hem/(Hem + Gt) ratio for 0–0.2 m (A) and 0.8–1 m (B), Kt/(Kt + Gt) ratio for 0–0.2 m (C) and 0.8–1 m (D), and ternary plot (D) for 0–0.2 m (E) and 0.8–1 m (F) layers. Hem = haematite; Gt = goethite; Kt = kaolinite; Gbs = gibbsite.

Table 2
Validation of maps with mineralogy data obtained by traditional methods.

Soil mineral/index	Layer (m)	Descriptive statistic						Validation		
		n	Min	Max	Mean	SD	CV (%)	R^2	RMSE	rRMSE (%)
Haematite (g kg^{-1})	0.0–0.2	221	3.6	67.3	21.8	14.5	66	0.64	8.72	13.7
Goethite (g kg^{-1})	0.0–0.2	221	1.5	36.5	9.7	5.8	59	0.40	4.45	12.7
Kaolinite (%)	0.0–0.2	221	30.7	100	89.8	17.2	18	0.20	14.81	21.4
Gibbsite (%)	0.0–0.2	221	0.0	75.5	10.4	17.2	165	0.29	14.41	19.1
Kt/Kt+Gbs	0.0–0.2	221	0.22	1.0	0.89	0.18	20	0.40	0.14	17.9

areas (Supplementary Material 1).

The areas of Ferralsols in the West of São Paulo State showed a small increase of haematite and kaolinite from 0 to 0.2 m to 0.8–1 m, while the areas of Acrisols, Lixisols and Alisols presented a high increase (Fig. 10a). At the farm level, it is also possible to visualize the increase of haematite and kaolinite at depth in an area of Alisols. The areas of yellow Ferralsols in the southwest of Amazonas State had a significant increase in AGt abundance from 0 to 0.2 m to 0.8–1 m, while for red and ferric Ferralsols, the haematite abundance was similar in the topsoil and 1 m of depth (Fig. 10b).

In the case study in the municipality of Piracicaba, São Paulo Sate, more weathered and clayey soil classes, such as Ferralsols, showed high values of haematite and moderate values of gibbsite, while less weathered and sandy (clay content less than 350 g kg^{-1}) soil classes showed low values for both minerals (Fig. 11a). An area (farm) showed high haematite and moderate gibbsite abundance for red and ferric Ferralsols and Nitisols derived from basalt rocks, while the other soil classes from sandstone and siltstone, such as Gleysols, Plinthosols, Cambisols and Leptosols, red Acrisols and red-yellow Acrisols, showed low values for both minerals (Fig. 11b). Another area (farm) showed high haematite abundance for red and ferric Ferralsols derived from basalt and low haematite abundance for red Acrisols, red-yellow Acrisols, Cambisols, Leptosols and Gleysols from sandstones and other rocks (Fig. 11c).

4. Discussion

4.1. Distribution of minerals

The Brazilian territory shows the dominance of Ferralsols, Acrisols and Lixisols, which altogether cover almost 60% of the entire territory. These soils are characterized as highly weathered with intense base leaching and predominance of kaolinite, goethite, haematite, and gibbsite (in that order) (Macías and Camps-Arbestain, 2020; Schaefer et al., 2008). In addition, mineral amounts tend to increase with clay, silicon, iron, and aluminum contents, which increase at depth (horizon B) due to pedogenetic processes, such as illuviation (Blume and Schwertmann, 1969; Buol et al., 2011; Carroll, 1953; Macías and Camps-Arbestain, 2020). This knowledge confirmed the increase in AGt, AHem, AKt, and AGbs values at depth (Mendes et al., 2021; Poppiel et al., 2020).

Haematite and gibbsite variations (Fig. 5ab and Fig. 6ab) are related to specific conditions such as relief, climate, geology, and pedology (Schaefer et al., 2008). Kaolinite and gibbsite showed less variability (Figs. 6 and 8) due to stability in the environment (Fink et al., 2016; Schaefer et al., 2008). The greater variability found in the 0–0.2 m layer can be explained by the soil types of Brazil. Soil classes with uniform profiles have similar amounts of minerals on the surface and subsurface, while soil classes with contrasting texture have an accumulation of minerals in the B horizon (Lelong et al., 1976; Santos et al., 2018).

4.2. Prediction models and accuracy of maps

4.2.1. Predictive MA models

The use of the SySI as covariates in the digital soil mapping (DSM) framework explains the more accurate results for the surface layer than

for deep layers (Fig. 4) (Supplementary Material 4). SySI is a direct measurement of topsoil reflectance, shown by temporal satellite images (Demattê et al., 2018), indirectly related to soil sub surface layers (Mendes et al., 2019).

Accurate results of haematite modeling were in accordance with the literature (Fig. 4) (Supplementary Material 4) (Bahia et al., 2015; Canton et al., 2021; Poppiel et al., 2020; Sellitto et al., 2009; Silva et al., 2020; Mendes et al., 2021). On the other hand, goethite predictions showed the worst results. In the literature, the results for goethite are more variable than haematite and almost always reported with low accuracy (Canton et al., 2021; Mendes et al., 2021; Poppiel et al., 2020; Sellitto et al., 2009; Silva et al., 2020). The electromagnetic transitions of haematite in the 535–580 nm and of goethite in the 415–455 nm range are well separated, allowing to differ iron forms (Scheinost et al., 1998). Haematite has a more pronounced and well-defined absorption feature than goethite (Demattê and Garcia, 1999; Kosmas et al., 1984). Substitution of iron for aluminum in the goethite structure can reach up to 40% in Brazilian Ferralsols and can cause less stability in the absorption feature (Jiang et al., 2014, 2022; Kosmas et al., 1984; Scheinost et al., 1998). Moreover, the range of goethite detection is near the 350–400 nm range, which is normally associated to high noise to signal ratio in the ASD FieldSpec.

The second-best result for mineral modeling was gibbsite followed by kaolinite (Fig. 4) (Supplementary material 4). The estimation of these minerals by spectroscopy has been less explored than iron oxides with variable results and gibbsite is more accurately predicted than kaolinite (Fernandes et al., 2020; Madeira-Neto et al., 1995; Mendes et al., 2021; Poppiel et al., 2020; Viscarra Rossel, 2011). The OH and metal-OH groups of these minerals show vibrational activity in the 2,200 to 2,300 nm range; however, this vibrational activity is also related to halloysite and 2:1 minerals, such as montmorillonites, muscovite and illite (Clark et al., 1990; Dufréchou et al., 2015). As a result, this spectral region cannot accurately detect Al-OH groups in all cases (Clark et al., 1990). The halloysite and montmorillonite (near 2,200 nm) have similar absorption features (Clark et al., 1990; Goetz et al., 2009) and may explain the low accuracy for the kaolinite models.

Few studies use environmental covariates and DSM procedures for MA mapping at fine resolution. Poppiel et al. (2020) mapped an area of 851,000 km^2 with spatial resolution of 30 m and found R^2 of 0.71, 0.72, and 0.72 for AHem, 0.45, 0.45, and 0.24 for AGt, 0.47, 0.55, and 0.59 for AKt, and 0.55, 0.64, and 0.65 for AGbs, for 0–0.2, 0.2–0.6, and 0.6–1 m layers, using TA, a synthetic vegetation image, SySI, and climate covariates. Poppiel et al. (2020) also used the RF algorithm and the models were validated by 10-fold cross-validation, which can overestimate the accuracy parameters compared to other data split methods (Volkan Bilgili et al., 2010), such as bootstrapping, used in this study. Viscarra Rossel (2011) used the Cubist method and Landsat bands, TA, climate, geological and gamma-ray data to predict AKt in Australia and reached R^2 ranging from 0.50 to 0.53 for 0–0.2 m layer and from 0.45 to 0.48 for 0.6–0.8 m layers. In a regional study area (2,274 km^2) in Brazil, Mendes et al. (2021) obtained mineral maps with 30 m of resolution using the SySI and RF algorithms. The authors reported R^2 of 0.54, 0.17, and 0.62 for AHem, 0.16, 0.10, and 0.24 for AGt, 0.32, 0.00, and 0.38 for AKt, and 0.17, 0.09, and 0.62 for AGbs, for 0–0.2, 0.4–0.6, and 0.8–1 m layers, respectively.
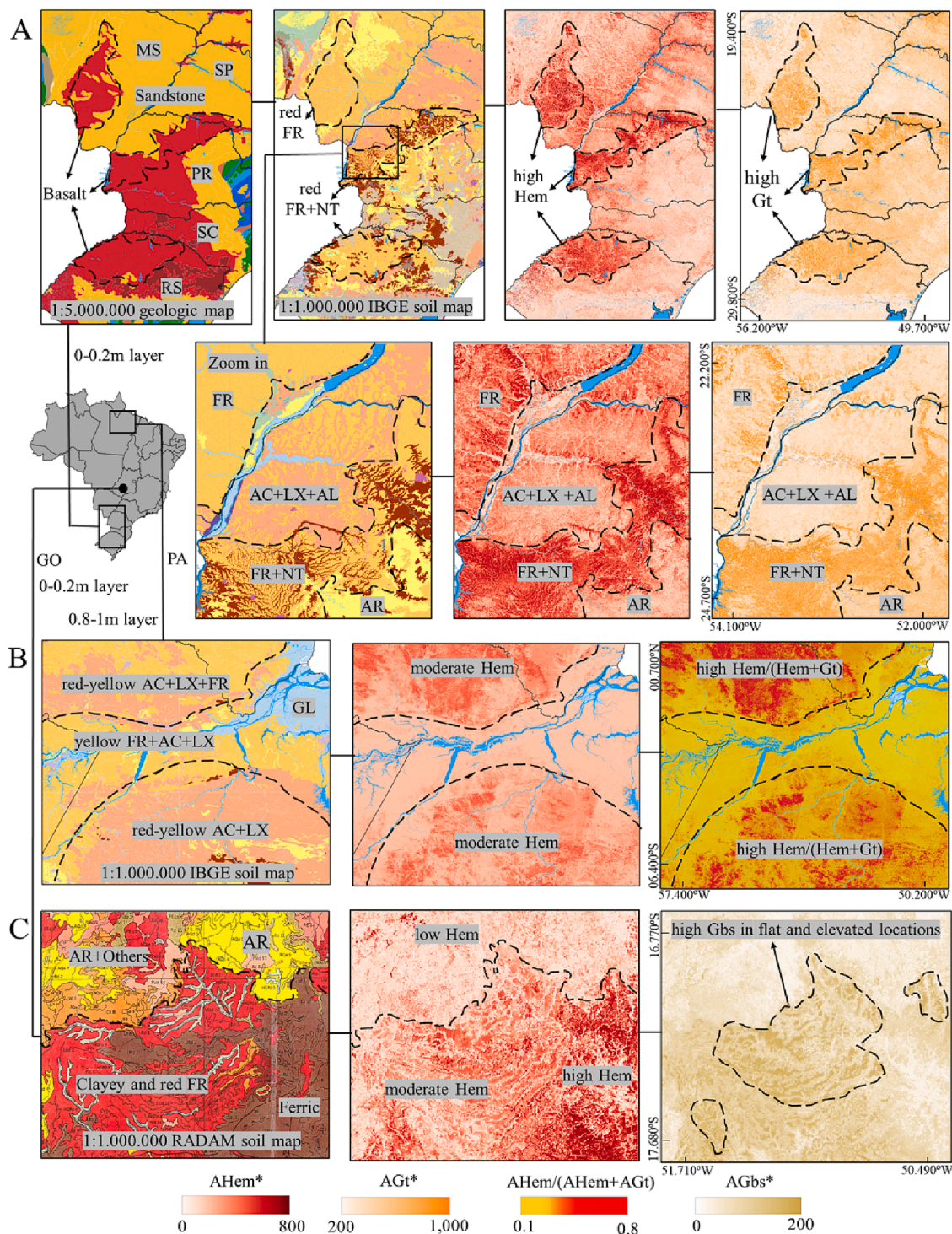
Fig. 9. Case studies for amplitude of haematite (AHem) and goethite (AGt) amplitudes, covering the states of Rio Grande do Sul (RS), Santa Catarina (SC), Paraná (PR), São Paulo (SP) and Mato Grosso do Sul (MS), with zoom in western Paraná (PR) State (A), for AHem and AHem/(AHem + AGt) ratio covering northeastern Pará (PA) State (B) and for AHem and Gbs amplitude (AGbs) covering southern Goiás (GO) State (C). FR = Ferralsol; NT = Nitisol; AC = Acrisol; LX = Lixisol; AL = Alisol. The maps of geology (adapted), the first soil, and the second soil were obtained from Gómez et al. (2019), Instituto Brasileiro de Geografia e Estatística (IBGE) (2021), and RADAMBRASIL (1983), respectively. Red soils have a hue equals to 2.5YR or more reddish and red-yellow soils have a hue equals to 5YR. Ferric soils had Fe$_2$O$_3$ equals to or more than 18 g kg^{-1} and clayey soil had clay equals or more than 350 g kg^{-1}. *10^{-6} scale.

In our study, it was not possible to avoid autocorrelation between soil observations in the modeling database, which could influence the accuracy parameters (Brus et al., 2011; Wadoux et al., 2020); however, there was not a strong autocorrelation in the database used. Autocorrelation can occur when legacy databases, without a sampling design, are used to predict soil attributes (Wadoux et al., 2020), as in this study.

To obtain more soil observations or remove some observations close to the others to avoid autocorrelation, as proposed by Wadoux et al. (2019), were not viable for this study. Besides, the use of spatial based calibration–validation sets split promote a pessimistic map accuracy assessment (Wadoux et al. 2021).
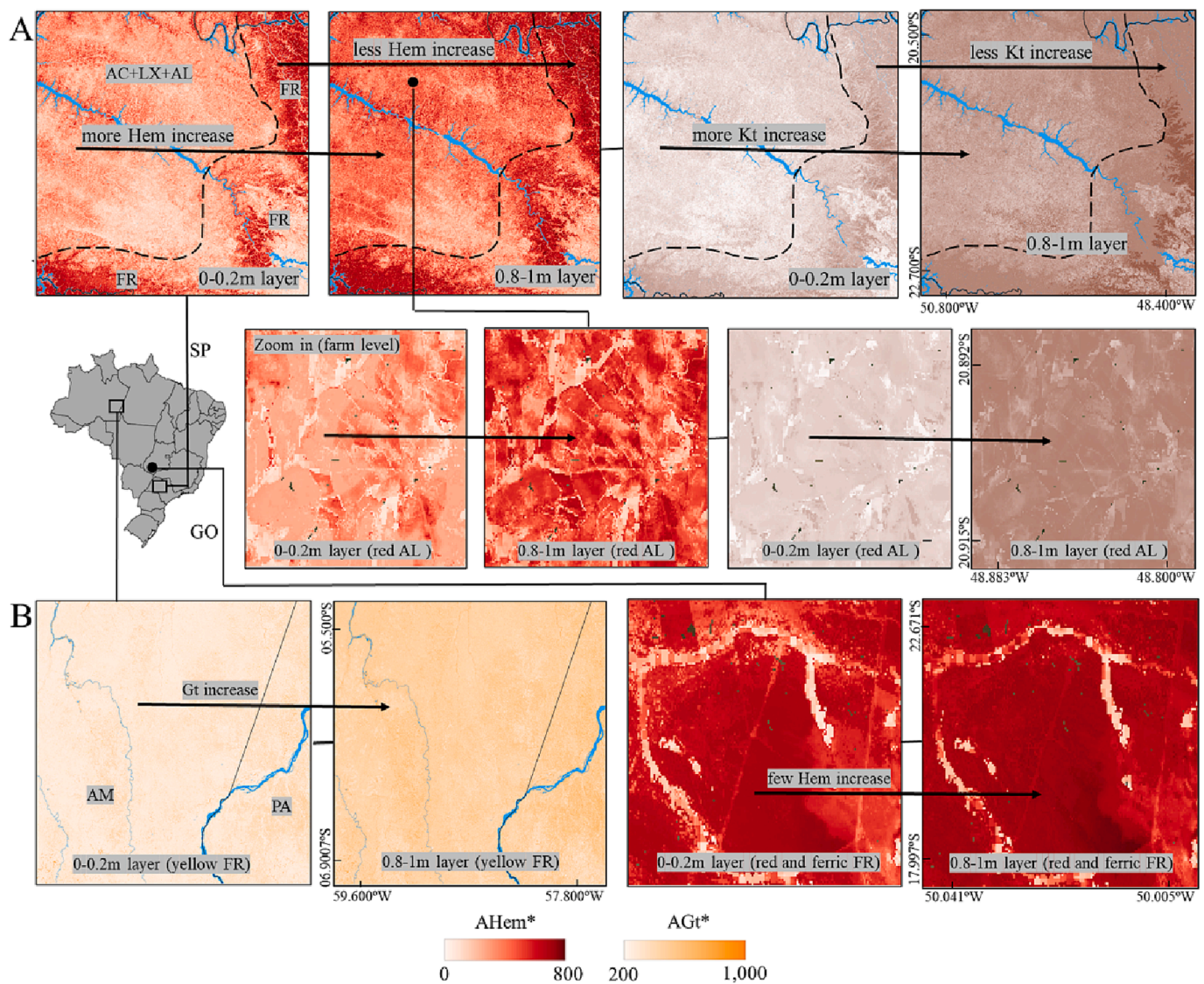
Fig. 10. Examples in 0–0.2 m and 0.8–1 m layers for haematite (AHem) and kaolinite (AKt) amplitudes covering the West São Paulo (SP) State, with a zoom at farm level (A), for goethite amplitude (AGt) in the Southwest of Amazonas (AM) and AHem in the South of Goiás (GO) States (B). FR = Ferralsol; AC = Acrisol; LX = Lixisol; AL = Alisol. Red soils have a hue equals to 2.5YR or more reddish, red-yellow soils have a hue equals to 5YR and yellow soils have hue equals to 7.5YR or more yellowish. Ferric soils had Fe_2O_3 equals to or more than 18 g kg^{-1}. *10^{-6} scale.

4.2.2. Validation of SySIc

It was possible to predict bare soil reflectance for locations without bare soil exposure (Table A2) and to use these estimates to map the soil mineral patterns. Bare soil reflectance showed a higher correlation with vegetation reflectance than terrain attributes (Table A1), meaning that vegetation spectral data can be used as a proxy. Studies have used vegetation reflectance to map soil information, such as Takata et al. (2007) and Mahmoudabadi et al. (2017). The decrease in accuracy for points without soil exposure (Fig. 4c) can be explained by propagation of uncertainty from predicted value of reflectance to mineral maps.

4.2.3. Validation with reference mineralogical data

The results of spatial validation with mineralogical data obtained by XRD showed a coherence of MA values obtained by Vis-NIR-SWIR spectra (Table 2). Results were more accurate for haematite and goethite predictions than for kaolinite and gibbsite. Higher accuracy for haematite than for goethite can be explained by the characteristic of spectral features, replacements of Fe for Al in the goethite structure and the noise in the spectral ranges (Kosmas et al. 1984; Scheinost et al.,

1998; Demattê and Garcia, 1999; Jiang et al., 2022, 2014; Sun, 2021), as discussed in section 4.2.1. For kaolinite and gibbsite, the OH groups in the infrared range used in this study are often related to other soil constituents (Clark et al., 1990; Dufréchou et al., 2015).

It is necessary to consider that the reference values of iron oxides were obtained with a semi-quantitative method, where standard samples of haematite and goethite were mixed in different proportions to access the Hem/(Hem + Gt) ratio. Thus, haematite and goethite concentrations were calculated based on the difference of free and low crystalline iron forms, determined by wet chemistry and the Hem/(Hem + Gt) ratio (Kämpf and Schwertmann, 1983). On the other hand, kaolinite and gibbsite were determined by the Rietveld method (Rietveld, 1969), which is an adjustment of the XRD diffractograms, where Kt + Gbs = 100%, disregarding the other minerals in the clay faction. Finally, it is necessary consider that, regardless of the method, the reference mineralogical information obtained by XRD also contains uncertainty (Kahle et al., 2002; Zhou et al., 2018) and the results can be considered a rough estimate of reality (Kahle et al., 2002).
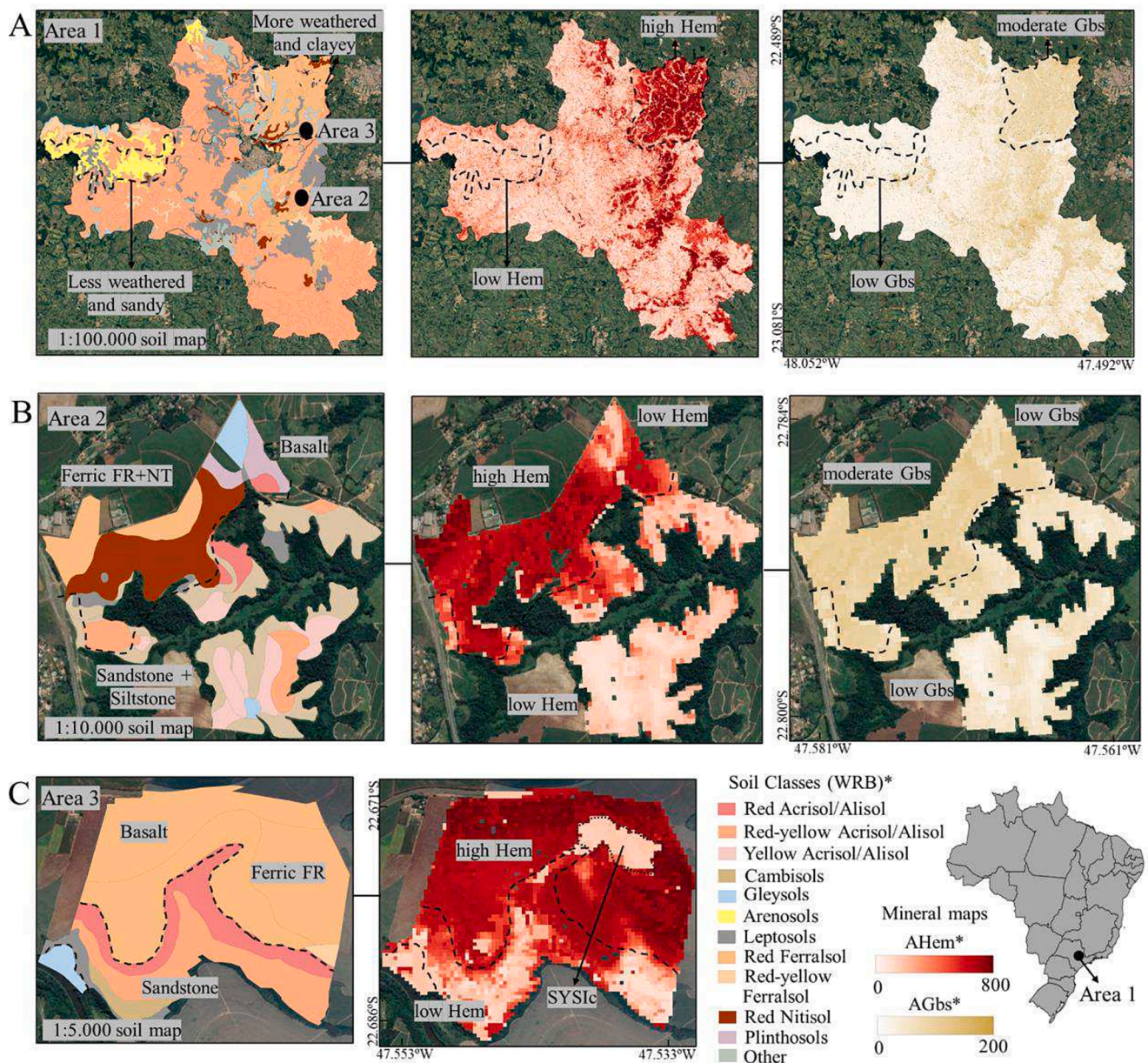
Fig. 11. Case studies in the 0.8–1 m layer for hematite (AHem) and gibbsite (AGt) amplitudes in regional (municipality of Piracicaba, São Paulo Sate) (A) and farm (B) (C) levels. The soil maps of 1:100.000 (adapted), 1:10.000, and 1:5.000 were from Oliveira and Prado (1989), Souza (2020), and Demattê et al. (2004), respectively. Red soils have a hue equals to 2.5YR or more reddish, red-yellow soils have a hue equals to 5YR and yellow soils have a hue equals to 7.5YR or more yellowish. Ferric soils had Fe$_2$O$_3$ equals to or more than 18 g kg^{-1}. *10^{-6} scale. SySIc = Soil synthetic image with full coverage.

4.2.4. Spatial uncertainty and quality of maps

Generally, the MA maps obtained were coherent with the soil class and the geological legacy maps at national, regional, and farm levels (Figs. 5, 6, 8, 9, 10 and 11) (Supplementary Material 1). However, some errors can be observed in some locations, such as in the Northern region of Brazil (Amazon rainforest) (Figs. 5 and 6). These errors may be associated to: 1) low density of soil observations, and 2) use of SySIc as a predictor (locations without any bare soil pixels, as explained in section 2.4.2). The uncertainty represented by 90PI were higher for areas with high predicted MA values (Figs. 5, 6, and 7). Kaolinite, which has higher contents, presented expressive uncertainty in almost all the Brazilian territory. This can be an indicator of an underestimation of the predicted mean values in these areas. Safanelli et al. (2021a) observed over-estimation of predicted values associated to high uncertainty in areas with low content for clay mapping. There was a tendency of higher PI90 values in locations with SySIc and few observation points. Poggio et al.

(2021) related high PI90 values in areas with low sampling density for several soil attributes.

The entire land area of Brazil had a low density of soil observations (Supplementary Material 2) (ranging from 8.41 × 10^{-4} to 1.41 × 10^{-3} points per km^2) concentrated in the southwestern and central-western regions, with few samples in the southern, northwestern, and northern regions (Fig. 2). This occurs because the legacy data from BSSL had samples located mainly in agricultural areas (Demattê et al., 2019).

As previously mentioned, SySI is a direct measurement of topsoil reflectance, while SySIc is predicted and thus with greater uncertainty. The "gaps" of SySI were predicted using a Landsat mosaic of vegetation and TA (Table A2). There were also artifacts of longitudinal lines in SySIc (Supplementary Material 1) related to clouds that interfered in Landsat images used in this mosaic. These artifact lines can also be seen in MA maps (Figs. 5 and 6). Fongaro et al. (2018) also reported artifacts in clay maps caused by covariates. Some areas with only predicted

Table A1

Spearman correlation between the Amplitude of minerals (AM), terrain attributes (TA), Synthetic soil image (SySI) bands and Landsat mosaic of 1984.

Covariate	Amplitude of minerals				SYSI bands					
	AHem	AGt	AKt	AGbs	blue	green	red	NIR	SWIR1	SWIR2
Elevation	−0.19	0.07	0.02	**0.43**	−0.23	−0.20	−0.10	−0.10	−0.07	−0.09
Slope	−0.12	−0.02	0.01	0.05	−0.18	−0.14	−0.20	−0.07	−0.08	−0.15
Aspect	−0.02	−0.01	−0.02	−0.06	−0.03	0.00	−0.08	0.00	0.00	−0.03
Hillshade	**0.12**	0.00	0.02	0.00	**0.19**	**0.20**	**0.18**	**0.14**	**0.16**	**0.19**
Northness	−0.03	−0.06	−0.04	0.00	0.03	0.05	0.02	0.07	0.07	0.05
Eastness	0.02	0.02	0.01	0.07	0.03	0.01	0.09	0.01	0.00	0.03
Horizontal curvature	−0.07	−0.05	−0.07	−0.09	−0.04	−0.05	−0.03	−0.03	−0.02	−0.02
Vertical curvature	−0.01	−0.02	0.01	−0.02	−0.06	−0.06	−0.02	−0.02	−0.01	0.00
Mean curvature	−0.05	−0.05	−0.04	−0.07	−0.06	−0.06	−0.02	−0.02	−0.01	0.00
Minimal curvature	0.07	0.01	0.02	0.00	0.02	0.00	0.07	0.01	0.04	0.09
Maximal curvature	−0.14	−0.08	−0.08	−0.12	−0.06	−0.06	−0.02	−0.02	−0.01	0.00
Gaussian curvature	0.00	−0.01	0.00	−0.03	0.00	0.00	0.01	0.00	0.00	0.01
Shape index	0.04	0.00	0.01	−0.01	−0.08	−0.07	−0.02	−0.02	−0.01	0.00
SySI blue	−0.63	−0.40	−0.36	−0.38						
SySI green	−0.60	−0.40	−0.36	−0.40						
SySI red	−0.47	−0.37	−0.33	−0.42						
SySI NIR	−0.48	−0.39	−0.36	−0.44						
SySI SWIR1	−0.51	−0.43	−0.36	−0.42						
SySI SWIR2	−0.49	−0.42	−0.36	−0.43						
Mosaic blue					0.33	0.29	0.28	0.18	0.18	0.20
Mosaic green					0.26	0.24	0.26	0.16	0.14	0.17
Mosaic red					0.18	0.16	0.21	0.11	0.11	0.12
Mosaic NIR					0.05	0.11	0.18	0.23	0.19	0.22
Mosaic SWIR1					0.26	0.24	0.30	0.29	0.34	0.32
Mosaic SWIR2					0.20	0.17	0.24	0.20	0.24	0.25

Where: Min = minimum value; Max = maximum value; SD = standard deviation; CV = coefficient of variation; R^2 = coefficient of determination; RMSE = root mean square error; rRMSE = relative root mean square error.

Table A2

Accuracy parameters for Synthetic soil image with full coverage (SYSIc) modelling.

SySI band	Calibration set			Validation set		
	R^2	RMSE	RPIQ	R^2	RMSE	RPIQ
blue	0.35	174.2	1.54	0.28	181.6	1.48
green	0.36	202.6	1.55	0.29	211.3	1.49
red	0.32	229.1	1.52	0.25	239.2	1.49
NIR	0.32	333.6	1.62	0.25	348.7	1.56
SWIR1	0.46	511.0	1.72	0.36	557.3	1.58
SWIR2	0.38	429.3	1.49	0.32	450.1	1.42

Where: R^2 = coefficient of determination; RMSE = root mean square error; RPIQ = ratio of the performance to interquartile distance.

reflectance values showed uncertain MA values, such as in the mountainous regions along the coast of the Atlantic Ocean and the Amazon rainforest (Figs. 5, 6, and 8).

4.3. Spatial and depth variation of minerals

4.3.1. Iron oxides

Geology has a crucial role in abundance of iron oxide minerals (Figs. 5, 8, and 9a) (Supplementary Material 1). Mafic volcanic rocks, such as basalt, are rich in iron and magnesium and produce soils with great amounts of iron oxide minerals (Schaefer et al., 2008; Cornell and Schwertmann, 2003; Long et al., 2011). Among the soil forming factors, parent material has a direct influence on soil mineralogy (Buol et al., 2011). The case studies in the states of Rio Grande do Sul, Santa Catarina, Paraná, São Paulo, and Mato Grosso do Sul showed higher influence of geology in the Central-West (Mato Grosso do Sul State) and geologic plus climate in the South of Brazil (Rio Grande do Sul and Paraná States) in iron oxides abundance, mainly haematite (Fig. 2) (Schaefer et al., 2008). Haematite occurrence is related mainly to more weathered and red soils, such as Ferralsols and Nitisols (Supplementary Material 1) characterized as clayey with high iron content (Fig. 9c, Fig. 10b, and Fig. 11).

High haematite abundance occurred in some locations under tropical and subtropical zones, with lower abundance in dry environments. (Figs. 2 and 5) (Supplementary Material 1). Occurrences of iron oxides are characteristic of warm and humid climates (Long et al., 2011; Macías and Camps-Arbestain, 2020; Cornell and Schwertmann, 2003; Schaefer et al., 2008). Haematite is commonly found in well-drained soils and less common in flat areas with poor drainage, such as the Pantanal biome in the western region of Mato Grosso do Sul State and in the Amazon basin (northern Brazil). On the other hand, goethite occurs mostly under limited drainage conditions (Figs. 5 and 10b) (Supplementary Material 1). Goethite is more stable than haematite and present in more variable environments (Macedo and Bryant, 1989; Resende et al., 1986; Cornell and Schwertmann, 2003; Schaefer et al., 2008).

Soils with homogeneous profiles have similar mineral distribution at depth, as the case of Ferralsols, originated from basalt rocks (Figs. 5 and 10a) (Supplementary Material 1). Conversely, soil types with textural gradients have higher iron oxides concentration at deeper layers. The increase in haematite at depth can be associated to the occurrence of red Acrisols, Lixisols, and Alisols. Higher predominance of goethite than haematite in the 0.8–1 m layer can be related to intermediate weathered soil types, such as red-yellow and yellow Acrisols, Lixisols, and Alisols.

Goethite abundance was more pronounced in the Amazon region (northern Brazil), which has a predominance of soils with 5YR or more yellowish (Fig. 5). However, areas with yellow Ferralsols also showed increased goethite abundance at depth, while it did not occur for haematite in red Ferralsols (Fig. 10b). This phenomenon is possibly explained by the masking of iron oxides by SOC in the 0–0.2 m layer (Croft et al., 2012; Heller Pearlshtien and Ben-Dor, 2020). This effect is more pronounced for goethite than for haematite, as haematite pigmentation directly affects the spectral response of the visible region. Additionally, goethite has a higher affinity to organic matter than haematite (Kaiser and Guggenberger, 2000).

The variation of Hem/(Hem + Gt) ratio verified in Fig. 8ab is directly related to soil color (Kosmas et al., 1984; Scheinost et al., 1998). A mixture of one part of haematite for four parts of goethite is enough to confer to the soil a hue of 5YR (Kosmas et al., 1984). Therefore, haematite is almost absent in yellow soils, while a small amount of

haematite is enough to mask the presence of goethite (Schaefer et al., 2008; Kosmas et al., 1984). In southern Brazil, Hem/Gt + Hem is driven mainly by climatic conditions, while in the central region, the parent material, climatic conditions, and drainage are the main driving factors (Schaefer et al., 2008).

A higher Hem/(Gt + Hem) ratio was observed in locations with middle elevation and slope (Fig. 8ab) (Supplementary Material 1). The areas with flat relief promoted the occurrence of goethite due to poor drainage, while areas with high slope had larger haematite occurrence (Breemen and Buurman, 2002; Macías and Camps-Arbestain, 2020). Generally, there was no variation in the Hem/(Gt + Hem) ratio at depth, except for the Amazon basin (North region) due to the increase of goethite abundance related to the occurrence of soils with hue 5YR or more yellow in the subsurface layer or because of the masking of goethite in the surface layer, as mentioned above.

4.3.2. Gibbsite

Gibbsite occurrence in the Brazilian territory was related to the high altitude (>500 m) and felsic plutonic and metamorphic rocks (Figs. 6 and 8) (Supplementary Material 1). Gibbsite can occur in different parent materials; nevertheless, there is propensity of gibbsite occurrence in metamorphic rocks (Buol et al., 2011). Flat and elevated locations in the southern region of Goiás State is an example of a high gibbsite occurrence (Fig. 9c). Poppiel et al. (2020) also verified this condition using a combined proximal and remote sensing approach in the central-western region in Brazil. Several studies also related gibbsite occurrence with altitude and more weathered soils in Brazil (Gomes et al., 2004; Poppiel et al., 2020; Reatto et al., 2008; Schaefer et al., 2008).

Gibbsite occurrence was associated to several soil types, such as highly weathered Ferralsols, intermediate weathered Acrisols and Lixisols, and less weathered Cambisols (Figs. 6 and 8) (Supplementary Material 1). Moderate gibbsite abundance occurred in soils derived from mafic rocks, associated to iron oxides (Figs. 5 and 6). This partially contrasts with the literature that attributes gibbsite abundance in highly weathered soils from mafic rocks, due to the small amount of silicon that favors its formation (Schaefer et al., 2008). However, gibbsite occurrence is not related exclusively to high-weathered soils formed under humid tropical climate (Macías Vazquez, 1981). Two processes that occur under free drainage conditions, low silica activity and few base contents can explain gibbsite occurrence in the soil, in the neo formation in the initial phases of Al-silicates weathering, and in the intense weathering process with dissolution of kaolinite (Macías Vazquez, 1981; Schaefer et al., 2008; Tardy et al., 1973).

According to the literature, gibbsite can increase, decrease, or stay the same at depth (Buol et al., 2011; Macedo and Bryant, 1989) and gibbsite showed the smallest increases at depth when compared to other minerals (Fig. 6).

4.3.3. Kaolinite

The map of kaolinite shows considerable abundance in all soil types from different parent materials and generally increased with depth (Fig. 6). The literature does not mention a widespread occurrence of kaolinite in Brazilian soils, occurring mainly in Ferralsols, Nitisols, Acrisols, Lixisols, and Plinthosols (Schwertmann and Kämpf, 1985). Kaolinite can occur in several conditions, mainly in humid and warmer climates, with free drainage conditions, low pH, and non-excessive Si leaching (Schaefer et al., 2008). The regions with moderate values of kaolinite have sedimentary parent material or dry climate conditions, such as some regions in Northeast Brazil (Figs. 2, 6, and 8) (Supplementary Material 1). In the western region of São Paulo State, there was an increase in kaolinite content from topsoil to 0.8–1 m due to illuviation in Acrisols, Lixisols, and Alisols and the low variation at depth in Ferralsols (Fig. 10a).

5. Conclusion

The relative abundances of major soil minerals goethite, haematite, kaolinite, and gibbsite were successfully identified and estimated by Vis-NIR-SWIR reflectance using samples from the Brazilian Soil Spectral Library (BSSL). These soil minerals presented a significant correlation with spectral data of synthetic soil image (SySI) with bare soil pixels obtained from Landsat time series images. Haematite presented more accurate results in spatial prediction with R^2 ranging from 0.48 to 0.56, followed by gibbsite (0.42 to 0.44), kaolinite (0.20 to 0.31), and goethite (0.16 to 0.26).

The spatial distribution of minerals was predicted for the entire Brazilian territory. For that purpose, a novel modeling procedure was implemented to obtain a bare soil reflectance with full coverage (SySIc), even for areas without soil exposure for the time period between 1984 and 2020. However, this estimated bare soil reflectance had lower accuracy in the prediction of soil minerals. Therefore, future studies should be carried out to improve SySIc accuracy.

The mineral maps were validated with reference mineralogical data and showed R^2 of 0.64 (haematite), 0.40 (goethite), 0.20 (kaolinite), 0.29 (gibbsite), and 0.40 (Kt/Kt + Gbs). The maps were in accordance with legacy maps of geology and pedology as well as with climate and terrain conditions. Moreover, the maps revealed the spatial distribution of mineral abundance at a finer scale than geological and pedology maps available in Brazil, reaching the farm level.

The proposed approach reveals the distribution of mineral abundances in the Brazilian territory and provides an efficient method to obtain information on soil mineralogy for large areas.

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.

Data availability

The authors do not have permission to share data.

Appendix A

.

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.geoderma.2023.116413.

References

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. Zeitschrift 22, 711–728. https://doi.org/10.1127/0941-2948/2013/0507.

Bahia, A.S.R.S., Marques Júnior, J., Siqueira, D.S., 2015. Procedures using diffuse reflectance spectroscopy for estimating hematite and Goethite in Oxisols of São Paulo, Brazil. Geoderma Reg. https://doi.org/10.1016/j.geodrs.2015.04.006.

Baldock, J., Skjemstad, J., 2000. Role of the soil matrix and minerals in protecting natural organic materials against biological attack. Org. Geochem. 31, 697–710. https://doi.org/10.1016/S0146-6380(00)00049-8.

Barrón, V., Torrent, J., 1986. Use of the Kubelka-Munk theory to study the influence of iron oxides on soil colour. J. Soil Sci. 37, 499–510. https://doi.org/10.1111/j.1365-2389.1986.tb00382.x.

Barrón, V., Méndez, J.M., Balbuena, J., Cruz-Yusta, M., Sánchez, L., Giménez, C., Sacristán, D., González-Guzmán, A., Sánchez-Rodríguez, A.R., Skiba, U.M., Inda, A. V., Marques, J., Recio, J.M., Delgado, A., del Campillo, M.C., Torrent, J., 2020. Photochemical emission and fixation of NOX gases in soils. Sci. Total Environ. 702, 134982 https://doi.org/10.1016/j.scitotenv.2019.134982.

Barthès, B.G., Kouakoua, E., Clairotte, M., Lallemand, J., Chapuis-Lardy, L., Rabenarivo, M., Roussel, S., 2019. Performance comparison between a miniaturized and a conventional near infrared reflectance (NIR) spectrometer for characterizing soil carbon and nitrogen. Geoderma 338, 422–429. https://doi.org/10.1016/j.geoderma.2018.12.031.

Bellinaso, H., Silvero, N.E.Q., Ruiz, L.F.C., Accorsi Amorim, M.T., Rosin, N.A., Mendes, W. de S., Sousa, G.P.B. de, Sepulveda, L.M.A., Queiroz, L.G. de, Nanni, M.R., Demattê, J.A.M., 2021. Clay content prediction using spectra data collected from the ground to space platforms in a smallholder tropical area. Geoderma 399, 115116. 10.1016/j.geoderma.2021.115116.

Bibring, J.-P., Langevin, Y., Mustard, J.F., Poulet, F., Arvidson, R., Gendrin, A., Gondet, B., Mangold, N., Pinet, P., Forget, F., Berthé, M., Bibring, J.-P., Gendrin, A., Gomez, C., Gondet, B., Jouglet, D., Poulet, F., Soufflot, A., Vincendon, M., Combes, M., Drossart, P., Encrenaz, T., Fouchet, T., Merchiorri, R., Belluci, G., Altieri, F., Formisano, V., Capaccioni, F., Cerroni, P., Coradini, A., Fonti, S., Korablev, O., Kottsov, V., Ignatiev, N., Moroz, V., Titov, D., Zasova, L., Loiseau, D., Mangold, N., Pinet, P., Douté, S., Schmitt, B., Sotin, C., Hauber, E., Hoffmann, H., Jaumann, R., Keller, U., Arvidson, R., Mustard, J.F., Duxbury, T., Forget, F., Neukum, G., 2006. Global Mineralogical and Aqueous Mars History Derived from OMEGA/Mars Express Data. Science (80-.). 312, 400–404. https://doi.org/10.1126/science.1122659.

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.

Blume, H.P., Schwertmann, U., 1969. Genetic Evaluation of Profile Distribution of Aluminum, Iron, and Manganese Oxides. Soil Sci. Soc. Am. J. 33, 438.

Breemen, N.V., Buurman, P., 2002. Soil Formation, 2nd ed. Springer, Amsterdam.

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

Brus, D.J., Kempen, B., Heuvelink, G.B.M., 2011. Sampling for validation of digital soil maps. Eur. J. Soil Sci. 62, 394–407. https://doi.org/10.1111/j.1365-2389.2011.01364.x.

Buol, S.W., Southard, R.J., Graham, R.C., McDaniel, P.A., 2011. Soil Genesis and Classification, 6a. ed. John Wiley & Sons Ltd, Chichester, UK.

Canton, L.C., de Souza Júnior, I.G., Silva, L.S., Marques Júnior, J., da Costa, A.C.S., 2021. Identification and quantification of iron oxides by diffuse reflectance spectroscopy with Praying Mantis accessory and integration sphere. Catena. https://doi.org/10.1016/j.catena.2020.104899.

Carroll, D., 1953. Clay Minerals in a Limestone Soil Profile1. Clays Clay Miner. 2, 171–182. https://doi.org/10.1346/CCMN.1953.0020115.

Chipera, S.J., Bish, D.L., 2001. Baseline Studies of the Clay Minerals Society Source Clays Powder X-ray Diffraction Analyses. Clays Clay Miner. 49, 398–409. https://doi.org/10.1346/CCMN.2001.0490507.

Clark, R.N., King, T.V.V., Klejwa, M., Swayze, G.A., Vergo, N., 1990. High spectral resolution reflectance spectroscopy of minerals. J. Geophys. Res. 95 https://doi.org/10.1029/jb095ib08p12653.

Coblinski, J.A., Giasson, É., Demattê, J.A.M., Dotto, A.C., Costa, J.J.F., Vašát, R., 2020. Prediction of soil texture classes through different wavelength regions of reflectance spectroscopy at various soil depths. Catena 189. https://doi.org/10.1016/j.catena.2020.104485.

Cornell, R.M., Schwertmann, U., 2003. The Iron Oxides: Structure, Properties, Reactions, Occurences and Uses, Second. ed. Wiley-VCH, Weinheim.

Croft, H., Kuhn, N.J., Anderson, K., 2012. On the use of remote sensing techniques for monitoring spatio-temporal soil organic carbon dynamics in agricultural systems. CATENA 94, 64–74. https://doi.org/10.1016/j.catena.2012.01.001.

de Oliveira, J.S., Inda, A.V., Barrón, V., Torrent, J., Tiecher, T., de Oliveira Camargo, F. A., 2020. Soil properties governing phosphorus adsorption in soils of Southern Brazil. Geoderma Reg. https://doi.org/10.1016/j.geodrs.2020.e00318.

Demattê, J.A., Campos, R.C., Alves, M.C., Fiorio, P.R., Nanni, M.R., 2004. Visible–NIR reflectance: a new approach on soil evaluation. Geoderma 121, 95–112. https://doi.org/10.1016/j.geoderma.2003.09.012.

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., Barreto, M.C. d. 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. 10.1016/j.geoderma.2019.05.043.

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. 10.1038/s41598-020-61408-1.

Demattê, J.A.M., Garcia, G.J., 1999. Alteration of Soil Properties through a Weathering Sequence as Evaluated by Spectral Reflectance. Soil Sci. Soc. Am. J. 63, 327–342. https://doi.org/10.2136/sssaj1999.03615995006300020010x.

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.

dos Santos, P.G., de Almeida, J.A., Sequinatto, L., 2017. Mineralogy of the Clay Fraction and Chemical Properties of Soils Developed from Sedimentary Lithologies of Pirambóia, Sanga-the-Cabral and Guará Geological Formations in Southern Brazil. Rev. Bras. Ciência do Solo 41. https://doi.org/10.18069/18069657rbcs20160344.

Dufréchou, G., Grandjean, G., Bourguignon, A., 2015. Geometrical analysis of laboratory soil spectra in the short-wave infrared domain: Clay composition and estimation of the swelling potential. Geoderma 243–244, 92–107. https://doi.org/10.1016/j.geoderma.2014.12.014.

Efron, B., Tibshirani, R.J., 1993. An Introduction to the Bootstrap. Chapman & Hall, Londres.

Fang, Q., Hong, H., Zhao, L., Kukolich, S., Yin, K., Wang, C., 2018. Visible and Near-Infrared Reflectance Spectroscopy for Investigating Soil Mineralogy: A Review. J. Spectrosc. 2018, 1–14. https://doi.org/10.1155/2018/3168974.

Fernandes, R.B.A., Barrón, V., Torrent, J., Fontes, M.P.F., 2004. Quantificação de óxidos de ferro de Latossolos brasileiros por espectroscopia de refletância difusa. Rev. Bras. Ciência do Solo 28, 245–257. https://doi.org/10.1590/S0100-06832004000200003.

Fernandes, K., Marques Júnior, J., Bahia, A.S.R. de S., Demattê, J.A.M., Ribon, A.A., 2020. Landscape-scale spatial variability of kaolinite-gibbsite ratio in tropical soils detected by diffuse reflectance spectroscopy. Catena. 10.1016/j.catena.2020.104795.

Fink, J.R., Inda, A.V., Tiecher, T., Barrón, V., 2016. Iron oxides and organic matter on soil phosphorus availability. Ciência e Agrotecnologia 40, 369–379. https://doi.org/10.1590/1413-70542016404023016.

Fongaro, C., Demattê, J., Rizzo, R., Lucas Safanelli, J., Mendes, W., Dotto, A., Vicente, L., Franceschini, M., Ustin, S., 2018. Improvement of Clay and Sand Quantification Based on a Novel Approach with a Focus on Multispectral Satellite Images. Remote Sens. 10, 1555. https://doi.org/10.3390/rs10101555.

Gilkes, R.J., Prakongkep, N., 2016. How the unique properties of soil kaolin affect the fertility of tropical soils. Appl. Clay Sci. 131, 100–106. https://doi.org/10.1016/j.clay.2016.01.007.

Goetz, A.F.H., Curtiss, B., Shiley, D.A., 2009. Rapid gangue mineral concentration measurement over conveyors by NIR reflectance spectroscopy. Miner. Eng. 22, 490–499. https://doi.org/10.1016/j.mineng.2008.12.013.

Gomes, J.B.V., Curi, N., Motta, P.E.F., Ker, J.C., Marques, J.J.G.S.M., Schulze, D.G., 2004. Análise de componentes principais de atributos físicos, químicos e mineralógicos de solos do bioma cerrado. Rev. Bras. Ciência do Solo 28, 137–153. https://doi.org/10.1590/S0100-06832004000100014.

Gomes, L.C., Faria, R.M., de Souza, E., Veloso, G.V., Schaefer, C.E.G.R., Filho, E.I.F., 2019. Modelling and mapping soil organic carbon stocks in Brazil. Geoderma 340, 337–350. https://doi.org/10.1016/j.geoderma.2019.01.007.

Gómez, J., Schobbenhaus, C., Montes, N.E., 2019. Geological Map of South America 2019. Scale 1:5 000 000. 10.32685/10.143.2019.929.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27. 10.1016/j.rse.2017.06.031.

Hassink, J., 1997. The capacity of soils to preserve organic C and N by their association with clay and silt particles. Plant Soil 191, 77–87. https://doi.org/10.1023/A:1004213929699.

Heimsath, A.M., DiBiase, R.A., Whipple, K.X., 2012. Soil production limits and the transition to bedrock-dominated landscapes. Nat. Geosci. 5, 210–214. https://doi.org/10.1038/ngeo1380.

Heller Pearlshtien, D., Ben-Dor, E., 2020. Effect of Organic Matter Content on the Spectral Signature of Iron Oxides across the VIS–NIR Spectral Region in Artificial Mixtures: An Example from a Red Soil from Israel. Remote Sens. 12, 1960. https://doi.org/10.3390/rs12121960.

Hengl, T., MacMillan, R.A., 2019. Predictive Soil Mapping with R [WWW Document]. URL https://soilmapper.org/ (accessed 7.7.21).

Heuvelink, G.B.M., Angelini, M.E., Poggio, L., Bai, Z., Batjes, N.H., Bosch, R., Bossio, D., Estella, S., Lehmann, J., Olmedo, G.F., Sanderman, J., 2021. Machine learning in space and time for modelling soil organic carbon change. Eur. J. Soil Sci. 72, 1607–1623. https://doi.org/10.1111/ejss.12998.

Instituto Brasileiro de Geografia e Estatística (IBGE), 2021. Mapa de Solos do Brasil [WWW Document]. URL https://www.ibge.gov.br/geociencias/downloads-geociencias.html (accessed 9.10.21).

Japan Aerospace Exploration Agency (JAEA): ALOS Research and Application Project, 2021. Advanced Land Observing Project [WWW Document]. accessed 7.15.21. https://www.eorc.jaxa.jp/ALOS/en/index_e.htm.

Jenny, H., 1941. Factors of Soil Formation: A System of Quantitative Pedology. Dover Publications, Nova York.

Jiang, Z., Liu, Q., Colombo, C., Barrón, V., Torrent, J., Hu, P., 2014. Quantification of Al-goethite from diffuse reflectance spectroscopy and magnetic methods. Geophys. J. Int. 196, 131–144. https://doi.org/10.1093/gji/ggt377.

Jiang, Z., Liu, Q., Roberts, A.P., Dekkers, M.J., Barrón, V., Torrent, J., Li, S., 2022. The Magnetic and Color Reflectance Properties of Hematite: From Earth to Mars. Rev. Geophys. 60 https://doi.org/10.1029/2020RG000698.

Kahle, M., Kleber, M., Jahn, R., 2002. Review of XRD-based quantitative analyses of clay minerals in soils: The suitability of mineral intensity factors. Geoderma 109, 191–205. https://doi.org/10.1016/S0016-7061(02)00175-1.

Kaiser, K., Guggenberger, G., 2000. The role of DOM sorption to mineral surfaces in the preservation of organic matter in soils. Org. Geochem. 31, 711–725. https://doi.org/10.1016/S0146-6380(00)00046-2.

Kämpf, N., Schwertmann, U., 1983. Goethite and hematite in a climosequence in southern Brazil and their application in classification of kaolinitic soils. Geoderma 29, 27–39. https://doi.org/10.1016/0016-7061(83)90028-9.

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. Sci. Rep. 11, 5076. https://doi.org/10.1038/s41598-021-84777-7.

Kosmas, C.S., Curi, N., Bryant, R.B., Franzmeier, D.P., 1984. Characterization of Iron Oxide Minerals by Second-Derivative Visible Spectroscopy. Soil Sci. Soc. Am. J. 48, 401–405. https://doi.org/10.2136/sssaj1984.03615995004800020036x.

Kunze, G.W., Dixon, J.B., 1986. Pretreatment for Mineralogical Analysis, in: Methods of Soil Analysis: Part 1 Physical and Mineralogical Methods, Second Edition. pp. 91–100.

Lelong, F., Tardy, Y., Grandin, G., Trescases, J.J., Boulange, B., 1976. Pedogenesis, chemical weathering and processes of formation of some supergene ore deposits. In: Wolf, K.H. (Ed.), Handbook of Strata-Bound and Stratiform Ore Deposits. Elsevier Scientific Publishing, Amsterdam, p. 81.

Lepsch, I.F., 2013. Status of Soil Surveys and Demand for Soil Series Descriptions in Brazil. Soil Horizons 54. https://doi.org/10.2136/sh2013-54-2-gc.

Long, X., Ji, J., Balsam, W., 2011. Rainfall-dependent transformations of iron oxides in a tropical saprolite transect of Hainan Island, South China: Spectral and magnetic measurements. J. Geophys. Res. 116, F03015. https://doi.org/10.1029/2010JF001712.

Macedo, J., Bryant, R.B., 1989. Preferential Microbial Reduction of Hematite Over Goethite in a Brazilian Oxisol. Soil Sci. Soc. Am. J. 53, 1114–1118. https://doi.org/10.2136/sssaj1989.03615995005300040022x.

Macías, F., Camps-Arbestain, M., 2020. A biogeochemical view of the world reference base soil classification system: Homage to Ward Chesworth, 1st ed, Advances in Agronomy. Elsevier Inc. 10.1016/bs.agron.2019.11.002.

Macías Vazquez, F., 1981. Formation of gibbsite in soils and saprolites of temperate-humid zones. Clay Miner. 16, 43–52. https://doi.org/10.1180/claymin.1981.016.1.03.

Madeira-Neto, J., Bedidi, A., Pouget, M., Cervelle, B., Flay, N., 1995. Spectral (MIR) determination of kaolinite and gibbsite contents in lateritic soils. Comptes Rendus - Acad. des Sci. Ser. II Sci. la Terre des Planetes 321, 119–128.

Mahmoudabadi, E., Karimi, A., Haghnia, G.H., Sepehr, A., 2017. Digital soil mapping using remote sensing indices, terrain attributes, and vegetation features in the rangelands of northeastern Iran. Environ. Monit. Assess. 189, 500. https://doi.org/10.1007/s10661-017-6197-7.

Malone, B.P., McBratney, A.B., Minasny, B., Laslett, G.M., 2009. Mapping continuous depth functions of soil carbon storage and available water capacity. Geoderma 154, 138–152. https://doi.org/10.1016/J.GEODERMA.2009.10.007.

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

McManus, D.A., 1991. Suggestions for authors whose manuscripts include quantitative clay mineral analysis by X-ray diffraction. Mar. Geol. 98, 1–5. https://doi.org/10.1016/0025-3227(91)90030-8.

Mendes, W. de 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. 10.1016/j.geoderma.2019.01.02.

Mendes, W. de S., Demattê, J.A.M., Bonfatti, B.R., Resende, M.E.B., Campos, L.R., Costa, A.C.S. da, 2021. A novel framework to estimate soil mineralogy using soil spectroscopy. Appl. Geochemistry 127. 10.1016/j.apgeochem.2021.104909.

Mendonça-Santos, M.L., dos Santos, H.G., 2006. Chapter 3 The State of the Art of Brazilian Soil Mapping and Prospects for Digital Soil Mapping. pp. 39–601. 10.1016/S0166-2481(06)31003-3.

Moura-Bueno, J.M., Dalmolin, R.S.D., Horst-Heinen, T.Z., ten Caten, A., Vasques, G.M., Dotto, A.C., Grunwald, S., 2020. When does stratification of a subtropical soil spectral library improve predictions of soil organic carbon content? Sci. Total Environ. 737, 139895. https://doi.org/10.1016/j.scitotenv.2020.139895.

Mulder, V.L., de Bruin, S., Weyermann, J., Kokaly, R.F., Schaepman, M.E., 2013. Characterizing regional soil mineral composition using spectroscopy and geostatistics. Remote Sens. Environ. 139, 415–429. https://doi.org/10.1016/j.rse.2013.08.018.

Ng, W., Minasny, B., Montazerolghaem, M., Padarian, J., Ferguson, R., Bailey, S., McBratney, A.B., 2019. Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra. Geoderma 352, 251–267.

Nocita, M., Stevens, A., van Wesemael, B., Aitkenhead, M., Bachmann, M., Barthès, B., Dor, E.B., Brown, D.J., Clairotte, M., Csorba, A., Dardenne, P., Demattê, J.A.M., Genot, V., Guerrero, C., Knadel, M., Montanarella, L., Noon, C., Ramirez-Lopez, L., Robertson, J., Sakai, H., Soriano-Disla, J.M., Shepherd, K.D., Stenberg, B., Towett, E. K., Vargas, R., Wetterlind, J., 2015. Soil Spectroscopy: An Alternative to Wet Chemistry for Soil Monitoring. Adv. Agron. 132, 139–159. https://doi.org/10.1016/bs.agron.2015.02.002.

Nolasco de Carvalho, C.C., Nunes, F.C., Homem Antunes, M.A., Nolasco, M.C., 2015. Soil Surveys in Brazil and Perspectives in Soil Digital Mapping. Soil Horizons 56. https://doi.org/10.2136/sh14-01-0002.

Oades, J.M., 1988. The retention of organic matter in soils. Biogeochemistry 5, 35–70. https://doi.org/10.1007/BF02180317.

Oliveira, J.B., Prado, H., 1989. Carta pedológica de Piracicaba. Escala 1, 100 .000.

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.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Bertrand, T., 2011. Scikit-learn: machine learning in python. J. Mach. Learn. Res. 12, 2825–2830.

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., Rizzo, R., Safanelli, J.L., Bonfatti, B.R., Silvero, N.E.Q., Demattê, J.A.M., 2020. Soil color and mineralogy mapping using proximal and remote sensing in Midwest Brazil. Remote Sens. 12, 1–30. https://doi.org/10.3390/rs12071197.

Poppiel, R.R., Demattê, J.A.M., Rosin, N.A., Campos, L.R., Tayebi, M., Bonfatti, B.R., Ayoubi, S., Tajik, S., Afshar, F.A., Jafari, A., Hamzehpour, N., Taghizadeh-Mehrjardi, R., Ostovari, Y., Asgari, N., Naimi, S., Nabiollahi, K., Fathizad, H., Zeraatpisheh, M., Javaheri, F., Doustaky, M., Naderi, M., Dehghani, S., Atash, S., Farshadirad, A., Mirzaee, S., Shahriari, A., Ghorbani, M., Rahmati, M., 2021. High resolution middle eastern soil attributes mapping via open data and cloud computing. Geoderma 385, 114890. https://doi.org/10.1016/j.geoderma.2020.114890.

Poulet, F., Bibring, J.P., Mustard, J.F., Gendrin, A., Mangold, N., Langevin, Y., Arvidson, R.E., Gondet, B., Gomez, C., 2005. Phyllosilicates on Mars and implications for early Martian climate. Nature 438 (7068), 623–627.

RADAMBRASIL, 1983. Mapa Exploratório de Solos: Goiânia Folha SE 22.

Ramos, P.V., Inda, A.V., Barrón, V., Siqueira, D.S., Marques Júnior, J., Teixeira, D.D.B., 2020. Color in subtropical brazilian soils as determined with a Munsell chart and by diffuse reflectance spectroscopy. Catena 193, 104609. https://doi.org/10.1016/j.catena.2020.104609.

Reatto, A., Bruand, A., de Souza Martins, E., Muller, F., da Silva, E.M., de Carvalho, O.A., Brossard, M., 2008. Variation of the kaolinite and gibbsite content at regional and local scale in Latosols of the Brazilian Central Plateau. Comptes Rendus Geosci. 340, 741–748. https://doi.org/10.1016/j.crte.2008.07.006.

Resende, M., Allan, J., Coey, J.M.D., 1986. The magnetic soils of Brazil. Earth Planet. Sci. Lett. 78, 322–326. https://doi.org/10.1016/0012-821X(86)90071-3.

Rietveld, H.M., 1969. A profile refinement method for nuclear and magnetic structures. J. Appl. Crystallogr. 2, 65–71. https://doi.org/10.1107/S0021889869006558.

Rosin, N.A., Demattê, J.A.M., Leite, M.C.A., de Carvalho, H.W.P., Costa, A.C., Greschuk, L.T., Curi, N., Silva, S.H.G., 2022. The fundamental of the effects of water, organic matter, and iron forms on the pXRF information in soil analyses. CATENA 210, 105868. https://doi.org/10.1016/j.catena.2021.105868.

Ross, J.L.S., 2013. Brasilian Relief: Structures and Forms. Rev. do Dep. Geogr. –USP 25, 20–36. https://doi.org/10.7154/RDG.2013.0025.0002.

Safanelli, J.L., Chabrillat, S., Ben-Dor, E., Demattê, J.A.M., 2020b. Multispectral Models from Bare Soil Composites for Mapping Topsoil Properties over Europe. Remote Sens. 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., Mendes, W. de 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. 10.1016/j.geoderma.2021.115042.

Safanelli, J. L., Demattê, J.A.M., Santos, N.V. dos, Rosas, J.T.F., Silvero, N.E.Q., Bonfatti, B.R., Mendes, W. de S., 2021b. Fine-scale soil mapping with Earth Observation data: a multiple geographic level comparison. Rev. Bras. Ciência do Solo 45. 10.36783/18069657rbcs20210080.

Safanelli, J.L., Poppiel, R., Ruiz, L., Bonfatti, B., Mello, F., Rizzo, R., Demattê, J., 2020a. Terrain Analysis in Google Earth Engine: A Method Adapted for High-Performance Global-Scale Analysis. ISPRS Int. J. Geo-Information 9, 400. https://doi.org/10.3390/ijgi9060400.

Sahwan, W., Lucke, B., Sprafke, T., Vanselow, K.A., Bäumler, R., 2021. Relationships between spectral features, iron oxides and colours of surface soils in northern Jordan. Eur. J. Soil Sci. 72, 80–97. https://doi.org/10.1111/ejss.12986.

Santos, H.G., Jacomine, P.K.T., Anjos, L.H.C., Oliveira, V.A., Lumbreras, J.F., Coelho, M. R., Almeida, J.A., Araújo Filho, J.C., Oliveira, J.B., Cunha, T.J.F., 2018. Brazilian Soil Classification System, 5 ed. EMBRAPA, Brasilia.

Savitzky, A., Golay, M.J.E., 1964. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Anal. Chem. 36, 1627–1639. https://doi.org/10.1021/ac60214a047.

Schaefer, C.E.G.R., Fabris, 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.

Scheinost, A.C., Chavernas, A., Barron, V., Torrent, D.J., 1998. Use and Limitations of Second-Derivative Diffuse Reflectance Spectroscopy in the Visible to Near-Infrared Range to Identify and Quantify Fe Oxide Minerals in Soils. Clays Clay Miner. 46, 528–536. https://doi.org/10.1346/CCMN.1998.0460506.

Schwertmann, U., Kämpf, N., 1985. Properties of goethite and hematite in kaolinitic soils of southern and central Brazil. Soil Sci. 139, 344–350.

Sellitto, V.M., Fernandes, R.B.A., Barrón, V., Colombo, C., 2009. Comparing two different spectroscopic techniques for the characterization of soil iron oxides: Diffuse versus bi-directional reflectance. Geoderma 149, 2–9. https://doi.org/10.1016/j.geoderma.2008.11.020.

Silva, E.B., Giasson, É., Dotto, A.C., ten Caten, A., Demattê, J.A.M., Bacic, I.L.Z., da Veiga, M., 2019. A Regional Legacy Soil Dataset for Prediction of Sand and Clay Content with Vis-Nir-Swir, in Southern Brazil. Rev. Bras. Ciência do Solo 43. https://doi.org/10.1590/18069657rbcs20180174.

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. https://doi.org/10.1016/j.catena.2019.104258.

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

Singh, M., Sarkar, B., Sarkar, S., Churchman, J., Bolan, N., Mandal, S., Menon, M., Purakayastha, T.J., Beerling, D.J., 2018. Stabilization of Soil Organic Carbon as Influenced by Clay Mineralogy. pp. 33–84. 10.1016/bs.agron.2017.11.001.

Soriano-Disla, J.M., Janik, L.J., Viscarra Rossel, R.A., MacDonald, L.M., McLaughlin, M. J., 2014. The performance of visible, near and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties. Appl. Spectrosc. Rev. 49, 139–186. https://doi.org/10.1080/05704928.2013.811081.

Souza, A.B., 2020. Proximal and remote sensing on the soil processes: from punctual to spatial approaches. Universaty of São Paulo.

Sun, X.-L., 2021. Errors induced by spectral measurement positions and instrument noise in soil organic carbon prediction using vis-NIR on intact soil. Geoderma 382, 114731. https://doi.org/10.1016/j.geoderma.2020.114731.

Takata, Y., Funakawa, S., Akshalov, K., Ishida, N., Kosaki, T., 2007. Spatial prediction of soil organic matter in northern Kazakhstan based on topographic and vegetation information. Soil Sci. Plant Nutr. 53, 289–299. https://doi.org/10.1111/j.1747-0765.2007.00142.x.

Tardy, Y., Bocquier, G., Paquet, H., Millot, G., 1973. Formation of clay from granite and its distribution in relation to climate and topography. Geoderma 10, 271–284. https://doi.org/10.1016/0016-7061(73)90002-5.

Torn, M.S., Trumbore, S.E., Chadwick, O.A., Vitousek, P.M., Hendricks, D.M., 1997. Mineral control of soil organic carbon storage and turnover. Nature 389, 170–173. https://doi.org/10.1038/38260.

Vaudour, E., Cerovic, Z., Ebengo, D., Latouche, G., 2018. Predicting Key Agronomic Soil Properties with UV-Vis Fluorescence Measurements Combined with Vis-NIR-SWIR Reflectance Spectroscopy: A Farm-Scale Study in a Mediterranean Viticultural Agroecosystem. Sensors 18, 1157. https://doi.org/10.3390/s18041157.

Viscarra Rossel, R.A., 2011. Fine-resolution multiscale mapping of clay minerals in Australian soils measured with near infrared spectra. J. Geophys. Res. 116, F04023. https://doi.org/10.1029/2011JF001977.

Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J., Skjemstad, J.O., 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131, 59–75. https://doi.org/10.1016/j.geoderma.2005.03.007.

Viscarra Rossel, R.A., Bui, E.N., De Caritat, P., McKenzie, N.J., 2010. Mapping iron oxides and the color of Australian soil using visible-near-infrared reflectance spectra. J. Geophys. Res. Earth Surf. 115, 1–13. https://doi.org/10.1029/2009JF001645.

Volkan Bilgili, A., van Es, H.M., Akbas, F., Durak, A., Hively, W.D., 2010. Visible-near infrared reflectance spectroscopy for assessment of soil properties in a semi-arid area of Turkey. J. Arid Environ. 74, 229–238. https://doi.org/10.1016/j.jaridenv.2009.08.011.

Wadoux, A.M.J.C., Heuvelink, G.B.M., de Bruin, S., Brus, D.J., 2021. Spatial cross-validation is not the right way to evaluate map accuracy. Ecol. Modell. 457, 109692 https://doi.org/10.1016/J.ECOLMODEL.2021.109692.

Wadoux, A.-M.-J.-C., Padarian, J., Minasny, B., 2019. Multi-source data integration for soil mapping using deep learning. SOIL 5, 107–119. https://doi.org/10.5194/soil-5-107-2019.

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

Whittig, L.D., 1965. X-Ray Diffraction Techniques for Mineral Identification and Mineralogical Composition. In: Black, C.A. (Ed.), Methods of Soil Analysis: Part 1 Physical and Mineralogical Properties, Including Statistics of Measurement and Sampling. American Society of Agronomy, Madison, pp. 671–698. https://doi.org/10.2134/agronmonogr9.1.c49.

Zhao, D., Arshad, M., Li, N., Triantafilis, J., 2021. Predicting soil physical and chemical properties using vis-NIR in Australian cotton areas. CATENA 196, 104938. https://doi.org/10.1016/j.catena.2020.104938.

Zhao, J., Chen, S., Hu, R., Li, Y., 2017. Aggregate stability and size distribution of red soils under different land uses integrally regulated by soil organic matter, and iron and aluminum oxides. Soil Tillage Res. 167, 73–79. https://doi.org/10.1016/j.still.2016.11.007.

Zhou, X., Liu, D., Bu, H., Deng, L., Liu, H., Yuan, P., Du, P., Song, H., 2018. XRD-based quantitative analysis of clay minerals using reference intensity ratios, mineral intensity factors, Rietveld, and full pattern summation methods: A critical review. Solid Earth Sci. 3, 16–29. https://doi.org/10.1016/j.sesci.2017.12.002.