Data Article

Soil class map of the Rio Jardim watershed in Central Brazil at 30 meter spatial resolution based on proximal and remote sensed data and MESMA method

Raúl R. Poppiel a,*, Marilusa P.C. Lacerda a, José A.M. Demattè b, Manuel P. Oliveira Jr. a, Bruna C. Gallo b, José L. Safaneli b

a Faculty of Agronomy and Veterinary Medicine, University of Brasília; ICC Sul, Darcy Ribeiro University Campus, Asa Norte, Postal Box 4508, Brasília 70910-960, Brazil
b Department of Soil Science, College of Agriculture Luiz de Queiroz, University of São Paulo; Pâdua Dias Av., 11, Piracicaba, Postal Box 09, São Paulo 13416-900, Brazil

A R T I C L E   I N F O

Article history:
Received 2 May 2019
Received in revised form 17 May 2019
Accepted 20 May 2019
Available online 30 May 2019

Keywords:
Digital soil mapping
Soil management
Agricultural planning
Soil classification system
Landsat

A B S T R A C T

Geospatial soil information is critical for agricultural policy formulation and decision making, land-use suitability analysis, sustainable soil management, environmental assessment, and other research topics that are of vital importance to agriculture and economy. Proximal and Remote sensing technologies enables us to collect, process, and analyze spectral data and to retrieve, synthesize, visualize valuable geospatial information for multidisciplinary uses. We obtained the soil class map provided in this article by processing and analyzing proximal and remote sensed data from soil samples collected in toposequences based on pedomorphogeological relationships. The soils were classified up to the second categorical level (suborder) of the Brazilian Soil Classification System (SiBCS), as well as in the World Reference Base (WRB) and United States Soil Taxonomy (ST) systems. The raster map has 30 m resolution and its accuracy is 73% (Kappa

DOI of original article: https://doi.org/10.1016/j.geoderma.2019.04.028.
* Corresponding author.
E-mail address: raulpoppiel@gmail.com (R.R. Poppiel).

https://doi.org/10.1016/j.dib.2019.104070
2352-3409/© 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
coefficient of 0.73). The soil legend represents a soil class followed by its topsoil color.

© 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

### 1. Data

The dataset contains spatial variability of soil classes at detailed scale (30 m spatial resolution), classified according to the Brazilian Soil Classification System — SiBCS [1], the World Reference Base — WRB [2] and the United States Soil Taxonomy — ST [3].

The sampling points were distributed over the Rio Jardim Watershed, in Central Brazil (Fig. 1), where we collected soil samples of Ferralsols, Plinthosols, Regosols and Cambisols (Fig. 2). After analyzing and processing all the data, we mapped the soils and quantified the results (Fig. 3). The soil map was saved in a KMZ file and provided in a 3D movie overlapped on the digital elevation model (Supplementary materials).
2. Experimental design, materials and methods

2.1. Soil sampling

The soil sampling design based on the soil-geoform-geology (pedomorphogeological) relationships allowed to identify representative areas where we defined six toposequences and 34 sites to visit in the field. In each site, we performed morphological description according to Santos et al. [4], and we collected soil samples from surface A horizon and diagnostic subsurface horizons (B or C), as well as at 0–0.2, 0.2–0.4 and 0.8–1.0 m layers. The soils were classified up to the second categorical level (suborder), according to the Brazilian Soil Classification System – SIBCS [1] (Fig. 2). For the validation step, we used an external and independent dataset of 231 sites distributed on a regular grid of 1400 × 1400 m.

2.2. Soil attributes analysis

The soil samples were air-dried, ground and sieved (2 mm mesh) and analyzed for physical and chemical determination according to Embrapa [5]. The spectroscopic analysis was conducted using the

![Fig. 1. Dataset location, soil sampling points and aerial photography with natural color composite.](image1)

![Fig. 2. Soil observations according to the Brazilian Soil Classification System – SIBCS [1] and corresponding World Reference Base – WRB [2] and Soil Taxonomy – ST [3] classifications.](image2)
FieldSpec Pro sensor (Analytical Spectral Devices Inc., Boulder, CO, USA), which has 1 nm of spectral resolution from 350 nm to 2500 nm (VIS-NIR-SWIR).

2.3. Satellite images acquisition

We acquire a Landsat 5 TM Level-1 time series (VIS-NIR-SWIR range) with 10 images from the United States Geological Survey (USGS), corresponding to orbit 221 and point 71 between May and September (dry season) from 1984 to 2009, with up to 10% maximum cloud cover.

2.4. Data processing

Laboratory topsoil reflectance spectra were processed by removing the continuous spectrum (CR) [6] and by applying the second derivative (SD) of the Kubelka–Munk (K-M) function [7] from 350 to 2500 nm for mineralogical assessment. We clustered soils with similar attributes based on topsoil reflectance to obtain spectral patterns, which were convolved using a Gaussian function of Landsat 5 TM spectral bands to obtain endmembers. Then, we added to each cluster (a soil class) the main topsoil attribute for creating soil legend for the map. Poppiel et al. [8] described further information about the method.

The raw digital numbers in the images were converted to radiance and then to reflectance-at-surface values [9] by applying the module Fast Line-of-sight Atmospheric Analysis of Hyperspectral cubes (FLAASH) using the ENVI software. We obtained bare soil images by applying the Normalized
Difference Vegetation Index — NDVI [10] to suppress vegetated areas (dense, moderate and sparse), and the Middle Infrared Index-MIDII [11] to mask areas with straw or burned. Then, the bare soil reflectance was retrieved in a bare soil composite (single image), denominated Synthetic Soil Image (SYSI) by selecting the lowest value of the MIDII which is correlated with low moisture content using as inputs data the SYSI and the endmembers to select the mixture model (a soil class) that best fits each pixel. To improve spatial coherence, we smoothed the resulting image by applying a median filter, using a $7 \times 7$ kernel.

For assessment of the map accuracy, we calculated the Kappa coefficient [14] based on a confusion matrix, and its values were classified as proposed by [15]. Poppiel et al. [8] presented a flowchart to illustrate the complete methodology. We showed summary quantifications of the mapped soil classes in Fig. 3.

Acknowledgements

The authors gratefully acknowledge the University of Brasilia — UnB, the Brazilian Federal Agency for Support and Evaluation of Graduate Education — CAPES for the first author master scholarship (finance code 001), the São Paulo Research Foundation — FAPESP (grant numbers: 2014-22262-0 and 2016/01597-9) for providing the equipment used in the VIS-NIR-SWIR reflectance analysis, the Geoprocessing and Pedomorphology Laboratory — Geoped at the FAV/UnB for provided data, the Geo-technologies in Soil Science group — GeoSS (http://esalqgeocis.wixsite.com/geocis/equipe) for the support.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.dib.2019.104070.

References

[1] Embrapa - Brazilian Agricultural Research Corporation - National Soils Research Center, Brazilian System of Soil Classification, fourth ed., Embrapa-Cnps, Brasilia, Brazil, 2014.
[2] IUSS Working Group WRB, World Reference Base for Soil Resources 2014: International Soil Classification System for Naming Soils and Creating Legends for Soil Maps, Food and Agriculture Organization, Rome, 2015. http://www.fao.org/3/a-i5794en/i5794en.pdf.
[3] Soil Survey Staff, Keys to Soil Taxonomy, United States Department of Agriculture, Washington, 2014. http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/FS.getDocumenti3794en/I3794EN.pdf.
[4] R.D. Santos, R.C. Lemos, H.G. Santos, J.C. Ker, L.H.C. Anjos, Manual of Field Soil Description and Collection, sixth ed., Sbc/S Embrapa-Cnps, Vçosa, Brazil, 2013.
[5] Embrapa - Brazilian Agricultural Research Corporation - National Soils Research Center, Manual of Soil Analysis Methods, third ed., Embrapa Solos, Brasilia, DF, 2017. http://ainfo.cnptia.embrapa.br/digital/bitstream/item/171907/1/Manual-de-Metodos-de-Analise-de-Solos-2017.pdf.
[6] R.N. Clark, T.L. Roush, Reflectance spectroscopy: quantitative analysis techniques for remote sensing applications, J. Geophys. Res. 89 (1984) 6329–6340, https://doi.org/10.1029/JB089iB07p06329.
[7] A.C. Scheinost, A. Chavernas, V. Barrón, J. Torrent, 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, Clay Clay Miner. 46 (1998) 528–536, https://doi.org/10.1346/CCMN.1998.0460506.
[8] R.R. Poppiel, M.P.C. Lacerda, J.A.M. Dematté, M.P. Oliveira Jr., B.C. Gallo, J.L. Safanelli, Pedology and soil class mapping from proximal and remote sensed data, Geoderma 348 (2019) 189–206, https://doi.org/10.1016/j.geoderma.2019.04.028.
[9] T. Cooley, G.P. Anderson, G.W. Felle, M.L. Hoke, A.J. Rakowski, J.H. Chetwynd, J.A. Gardner, S.M. Adler-Golden, M.W. Matthew, A. Berk, L.S. Bernstein, P.K. Acharya, D. Miller, P. Lewis, FLAASH, a MODTRAN4-based atmospheric correction algorithm, its application and validation, Proc. Int. Geosci. Remote Sens. Symp. - IGARSS. 3 (2002) 1414–1418, https://doi.org/10.1109/IGARSS.2002.1026134.
[10] J.W. Rouse, R.H. Haas, J.A. Schell, D.W. Deering, Monitoring Vegetation Systems in the Great Plains with ERTS, Proceeding ERTS Symp, 1973, pp. 309–317, citeulike-article-id:12009708.
[11] J. Madeira Netto, Etude quantitative des relations constitutants minéralogiques - réflectance diffuse des latosols brésiliens: Application à l’utilisation pédologique des données satellitaires TM (region de Brasilia), Doctorat thesis (sciences des sols), Université Paris-VI, 1991.
[12] R Core Team, R: A Language and Environment for Statistical Computing, 2018. https://www.r-project.org/.
[13] D.A. Roberts, K. Halligan, P. Dennison, VIPER Tools User Manual, University of California, Santa Barbara, 2007. http://www.vipertools.org/.

[14] J. Cohen, A coefficient of agreement of nominal scales, Educ. Psychol. Meas. 20 (1960) 37–46, https://doi.org/10.1177/001316446002000104.

[15] J.R. Landis, G.G. Koch, The measurement of observer agreement for categorical data, Biometrics 33 (1977) 159, https://doi.org/10.2307/2529310.