Universidade de São Paulo
Escola Superior de Agricultura "Luiz de Queiroz"

Fine scale mapping of phosphorus stocks in brazilian soils by geotechnologies: implications for a sustainable agriculture

Jorge Tadeu Fim Rosas

Thesis presented to obtain the degree of Doctor Science. Area: Soil and Plant Nutrition

Piracicaba
2023

Jorge Tadeu Fim Rosas
Bachelor's in Agronomy

Fine scale mapping of phosphorus stocks in brazilian soils by geotechnologies: implications for a sustainable agriculture

Advisor:
Prof. Dr. **JOSÉ ALEXANDRE MELO DEMATTÊ**

Thesis presented to obtain the degree of Doctor Science. Area: Soil and Plant Nutrition

Piracicaba
2023

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA – DIBD/ESALQ/USP

Rosas, Jorge Tadeu Fim

Fine scale mapping of phosphorus stocks in brazilian soils by geotechnologies: implications for a sustainable agriculture/Jorge Tadeu Fim Rosas. - - Piracicaba, 2023.

89 p.

Tese (Doutorado) - - USP / Escola Superior de Agricultura “Luiz de Queiroz”.

1. Aprendizagem de máquina 2. Mapeamento digital de solos 3. Sensoriamento remoto 4. Solos tropicais I. Título

ACKNOWLEDGMENTS

Firstly, I would like to express my profound gratitude to the University of São Paulo (USP), a renowned institution that provided me with the opportunity to expand my academic horizons. The Luiz de Queiroz College of Agriculture (ESALQ) and the Graduate Program in Soils and Plant Nutrition were fundamental, providing the structure and academic environment necessary for the completion of this work.

I am grateful to CAPES - Coordination for the Improvement of Higher Education Personnel for the scholarship granted during my doctorate through process number: 88887.481452/2020-00 and to FAPESP (São Paulo Research Foundation) which, through projects 2014-22262-0, 2021-05129, allowed the financing of my research.

A special thanks to my advisor, Professor Dr. José A.M. Demattê. His guidance, patience, and knowledge were invaluable. His insight and wisdom were a constant inspiration and his encouragement in moments of doubt helped alleviate my anxieties.

To my parents, José Tadeu Gonçalves Rosas and Maria Aparecida Fim Rosas, I owe eternal gratitude. You gave me life, educated me, and always provided me with the necessary support. This work is also yours.

To my girlfriend Luciana, for the affection, support, friendship, concern, and encouragement. Thank you for understanding me in moments of stress and calming me down with the phrase “Everything will be alright”. Thank you for sharing this moment of great importance in my life with me.

To my siblings, Guilherme, Luana, and Pedrinho, I thank you for always believing in me and for being a constant source of friendship and support.

I thank my colleagues at GEOCIS, especially Bruno, Nicolas, Fernando, Gabriel, Soledad, Merilyn, Leticia, Jean, and Borges. You made the work environment a place of friendship and learning. Each of you contributed to my personal and academic experience.

To my dear colleagues from “Bela Vista 619” - Caio, Rafael, Felipe, and Adijailton - with whom I had the privilege of sharing the same home, my sincere gratitude. Our harmonious coexistence, full of laughter and celebrations, became an inexhaustible source of joy and comfort. Sunday lunches, always accompanied by good food and excellent company, became precious moments of relaxation and camaraderie. Without a doubt, your presence and the mutual support we shared made the entire academic process significantly lighter and more enjoyable.

Finally, I thank all the other friends who, directly or indirectly, were part of my daily life and contributed to my personal and academic formation. Each of you has a special place in this journey.

SUMMARY

RESUMO

Mapeamento de alta resolução dos estoques de fósforo nos solos brasileiros por meio de geotecnologias: implicações para uma agricultura sustentável

O fósforo, um nutriente essencial para a vida e crucial para a agricultura, é extraído de reservas minerais não renováveis, levantando preocupações sobre a sustentabilidade. Portanto, o mapeamento dos estoques de fósforo no solo é vital para a gestão eficiente desse recurso e para a sustentabilidade do planeta. Entretanto, mapear esses estoques em grandes áreas e alto nível de detalhes não é uma tarefa fácil. Felizmente, novas metodologias de mapeamento digital de solo (MDS) podem contribuir para a obtenção de mapas cada vez mais precisos. A hipótese central desta tese é que o uso de geotecnologias pode contribuir para o mapeamento dos estoques de P no brasil, com elevada precisão. Esta tese foi dividida em dois artigos. O primeiro teve como objetivo mapear os principais óxidos da fração argila do solo, que estão intimamente relacionados aos estoques de P. Esses mapas foram utilizados como covariáveis preditoras dos estoques de P no segundo capítulo, cujo objetivo principal foi mapear os estoques de fósforo total (PT) e fósforo disponível (PD) no Brasil. Para mapear os principais óxidos da fração argila, usamos um conjunto de dados de modelagem com 5.330 observações. Seis variáveis espectrais obtidas da série histórica Landsat e sete atributos do terreno derivados de um modelo digital de elevação foram utilizados para determinar Fe2O3, Al2O3 e SiO2 usando o algoritmo Random Forest. As melhores previsões foram observadas para Fe2O3 na camada superficial (RMSE = 47,0, RPIQ = 1,85 e R2 = 0,65), enquanto as menores previsões foram para SiO2 na camada subterrânea (RMSE = 66,7, RPIQ = 1,55 e R2 = 0,19). Os mapas dos óxidos na camada de 0-20 cm foram usados na predição dos estoques de P. Além desses óxidos, incluímos covariáveis ambientais relacionadas aos processos de formação do solo, como relevo, clima e organismos, e outros atributos, como por exemplo, carbono orgânico do solo e argila. Dividimos o Brasil em duas sub-regiões, representando áreas com cobertura nativa e áreas com cobertura antrópica. A partir disso, construímos modelos preditivos independentes para cada sub-região. Ao todo, 28.572 amostras para PD e 3.154 para PT foram usadas na modelagem. Nossos resultados mostraram que o Brasil possui um estoque de TP de 531 Mt e um estoque de AP de 17,4 Mt. As maiores médias de estoques de TP estão no bioma Mata Atlântica (73,8 g/m²), o que pode estar ligado aos maiores estoques de carbono orgânico no solo deste bioma. Os maiores estoques médios de AP estão no bioma Caatinga (2,51 g/m²) por apresentar solos mais jovens e com baixa capacidade de adsorção de fósforo. Descobrimos também que o uso de fertilizantes aumentou significativamente os estoques de AP, onde as áreas agrícolas sempre tiveram estoques de AP mais elevados do que as áreas nativas. A abordagem proposta foi capaz de quantificar os estoques de P do Brasil com distribuição espacial alinhada ao entendimento dos solos brasileiros. Além disso, foi possível mapear de forma inédita todo o território brasileiro com uma escala de 30m.

Palavras-chave: Aprendizagem de máquina, Mapeamento digital de solos, Sensoriamento remoto, Solos tropicais

ABSTRACT

Fine scale mapping of phosphorus stocks in brazilian soils by geotechnologies: implications for a sustainable agriculture

Phosphorus, an essential nutrient for life and crucial for agriculture, is extracted from non-renewable mineral reserves, raising concerns about sustainability. Therefore, mapping soil phosphorus stocks is vital for the efficient management of this resource and for the sustainability of the planet. However, mapping these stocks over large areas and at a high level of detail is not an easy task. Fortunately, new digital soil mapping (DSM) methodologies can contribute to obtaining increasingly accurate maps. The central hypothesis of this thesis is that the use of geotechnologies can contribute to the mapping of P stocks in Brazil, with high accuracy. This thesis was divided into two articles. The first aimed to map the main oxides of the soil clay fraction, which are closely related to P stocks. These maps were used as predictive covariates of P stocks in the second chapter, whose main objective was to map the stocks of total phosphorus (TP) and available phosphorus (AP) in Brazil. To map the main oxides of the clay fraction, we used a modeling dataset with 5,330 observations. Six spectral variables obtained from the Landsat historical series and seven terrain attributes derived from a digital elevation model were used to determine Fe2O3, Al2O3, and SiO2 using the Random Forest algorithm. The best predictions were observed for Fe2O3 in the superficial layer (RMSE = 47.0, RPIQ = 1.85, and R2 = 0.65), while the lowest predictions were for SiO2 in the underground layer (RMSE = 66.7, RPIQ = 1.55, and R2 = 0.19). The maps of the oxides in the 0-20 cm layer were used in predicting P stocks. In addition to these oxides, we included environmental covariates related to soil formation processes, such as relief, climate, and organisms, and other attributes, such as, for example, soil organic carbon and clay. We divided Brazil into two sub-regions, representing areas with native coverage and areas with anthropic coverage. From this, we built independent predictive models for each sub-region. In total, 28,572 samples for AP and 3,154 for TP were used in modeling. Our results showed that Brazil has a TP stock of 531 Mt and an AP stock of 17.4 Mt. The highest averages of TP stocks are in the Atlantic Forest biome (73.8 g/m²), which may be linked to the higher stocks of soil organic carbon in this biome. The highest average AP stocks are in the Caatinga biome (2.51 g/m²) for presenting younger soils and with low phosphorus adsorption capacity. We also found that the use of fertilizers significantly increased AP stocks, where agricultural areas always had higher AP stocks than native areas. The proposed approach was able to quantify Brazil's P stocks with spatial distribution aligned with the understanding of Brazilian soils. In addition, it was possible to map the entire Brazilian territory for the first time with a scale of 30m.

Keywords: Machine Learning, Digital Soil Mapping, Remote Sensing, Tropical Soils

1 GENERAL INTRODUCTION

Brazil is considered a breadbasket for the world due to the large quantities of food produced in its territory (Tollefson, 2010). One of the factors that contribute most to the large quantities of food produced is the country's large territorial extension as well as its climatic characteristics and wealth of water resources (Dias et al., 2016). At the end of the 20th century, what most marked the increase in Brazil's agricultural production was the expansion of agricultural frontiers, mainly in the Cerrado biome (Gibbs et al., 2010; Leite et al., 2012). At the beginning of the 21st century, with the implementation of public policies and concern on the part of society for the environment, the expansion of agricultural frontiers in Brazil had its pace slowed down (Graesser et al., 2015). However, agricultural production continues to grow. Year after year, grain production breaks records (Dias et al., 2016).

The explanation for these increases in Brazilian agricultural production is associated with the increased adoption of technologies by farmers, mainly the use of agricultural fertilizers (Barretto et al., 2013; Lopes and Guimarães, 2016). Among the fertilizers, phosphates have gained prominence in their use, as their use has greatly increased in recent Years (Withers et al., 2018). Studies report that the quantities of P fertilizers currently used by agriculture are about 30 times greater than in the 1960s, and this may be leading to the accumulation of P in Brazilian agricultural soils (Martinelli et al., 2010; Roy et al., 2016a).

The use of P in large quantities and agriculture's heavy dependence on this nutrient has sparked debates on this subject. The main concern regarding the use of P in agriculture is that the P used is primarily extracted from non-renewable mineral reserves, and their depletion could lead to food insecurity on the planet (Elser, 2012; Rosemarin and Ekane, 2016). The pursuit of sustainable P use has been a central theme of many studies. However, in recent decades, phosphorus has been used in agriculture without much control and often inefficiently (Roy et al., 2016b; Withers et al., 2015). Incorrect application of P can have serious consequences. Its excessive use can cause losses to aquatic ecosystems by causing primary overproduction, leading to eutrophication and significant ecosystem imbalance (Schipanski and Bennett, 2021; Steffen et al., 2015). Furthermore, in the soil, excessive doses can be adsorbed by soil clays and become unavailable to plants in a short period (Schipanski and Bennett, 2021).

The pursuit of efficient and sustainable use of phosphorus (P) involves strategies at the local level (on the farm), such as soil pH correction, application in varied doses, and increasing organic matter in the soil to neutralize charges capable of adsorbing P (Schröder et al., 2011; Simpson et al., 2011). In addition, it encompasses strategies at the global level, mainly highlighting the mapping of P stocks and the factors that influence the availability of this element to plants (Barbieri et al., 2021; Roy et al., 2016). Mapping soil P stocks can shed light on the spatial distribution of P in the soil, allowing for the assessment of the productivity of natural and agricultural ecosystems. Furthermore, knowledge of these stocks can guide P applications or the adoption of practices to preserve P in soils (Viscarra Rossel and Bui, 2016).

Mapping large areas ends up being a laborious process, which depends on a dense number of field samples and spatial modeling of the variable that one wishes to map. A very used modeling method for mapping soil attributes was kriging, which is based on the assumption that there is spatial dependence in the soil attributes,

and these can be interpolated as a function of distance (Adhikari et al., 2013; Mondal et al., 2017). Despite being widely used, a large number of samples is necessary to adjust models that are capable of explaining the spatial behavior of a certain attribute (Adhikari et al., 2013). However, new digital mapping techniques have been surpassing classical methods. These techniques are based on machine learning or deep learning algorithms that minimize errors (Padarian et al., 2019; Poppiel et al., 2019; Tziolas et al., 2020). These techniques make use of covariates that strongly correlate with the study variable, and these covariates are usually easy to acquire, and can be sampled in large quantities (Padarian et al., 2019).

Among the steps of digital soil mapping (DSM), the choice of covariates is the most critical. The SCORPAN model proposed by McBratney et al. (2003) suggests that environmental covariates should mainly represent the soil formation processes. In addition, these covariates can also be represented by soil properties with some degree of relation to the soil attribute to be mapped (McBratney et al., 2003). Among the covariates used in the mapping of soil attributes, spectral variables in the spectrum region varying from 0.4-2.5 μm obtained by optical sensors have shown to be very promising for mapping in large areas where soil observations tend to be scarce (Forkuor et al., 2017; Zeraatpisheh et al., 2019; Zhou et al., 2021). The sensors on board orbital platforms like Landsat have gained the most prominence in recent years. Generally, the products of orbital platforms have been used as indirect information to infer the spatial distribution of the soil due to the continuous coverage of the earth's surface (Chabrillat et al., 2019). In addition, the multi-temporal collections of images allow the use of data mining algorithms to generate an image that represents the exposed soil surface, which has a strong relationship with numerous soil attributes.

Recently, many automated processing algorithms have been developed to recover soil reflectance from multispectral images (Demattê et al., 2018; Diek et al., 2017; Rogge et al., 2018). In Brazil, the Geospatial Soil Sensing System (GEOS3) has been used to map soil atributes (Fongaro et al., 2018; Gallo et al., 2018; Poppiel et al., 2019) and infer spatial distribution of soils and lithological classes (Mendes et al., 2019; Poppiel et al., 2019; Rizzo et al., 2020). The algorithm is based on the recovery of historical soil exposures to satellite measurements, which are aggregated into products namely: Synthetic Soil Image (SYSI) and Soil Frequency (SF) (Demattê et al., 2018). These products reveal the direct multispectral response of the soils and the historical disturbance of the surface for a given location, which can be correlated with soil properties and morphological attributes and with soil intensification respectively (Demattê et al., 2020).

Other covariates that have been used in DSM are those related to the soil formation factors relief, climate, and organism. For example, multispectral images, obtained by satellites, can describe the spectral behavior of vegetation, representing the soil formation factor organism (Chen et al., 2022). Digital elevation models are used to represent the relief (Safanelli et al., 2020), while historical data, obtained by weather stations, such as rainfall and temperature, can be spatialized and used to represent the climate (Shary, 2023). In addition, soil properties can also be used to compose the covariates. In the case of phosphorus, its stocks can be greatly affected by the degree of weathering of the soils (Antoniadis et al., 2016). The degree of weathering influences the composition of the soils as well as the presence of aluminum and iron oxides that are large fixers of P in tropical soils (Pavinato et al., 2020).

In light of the above, the central hypothesis of this thesis proposes that the application of advanced DSM tools, such as the use of machine learning and environmental variables as representations of the underlying factors to the spatial variations in phosphorus stocks, will allow the construction of a comprehensive and detailed map these stocks throughout the Brazilian territory. This approach will offer an enhanced level of precision and understanding of the geospatial distributions of phosphorus, significantly contributing to the sustainable management of this crucial resource in the agricultural and environmental context. To test this hypothesis, this thesis was structured into two articles. The primary objective of the first article was to map the oxides of the soil clay fraction and the degree of chemical weathering of the soils in order to associate these attributes with the P stocks in Brazilian soils and use them as environmental covariates to predict these stocks. The main objective of the second article was to map the total and plant-available phosphorus stocks of Brazilian soils at a 30 m scale using advanced DSM techniques.

REFERENCES

Adhikari, K., Kheir, R.B., Greve, M.B., Greve, M.H., 2013. Comparing kriging and regression approaches for mapping soil clay content in a diverse danish landscape. Soil Science 178, 505–517. https://doi.org/10.1097/SS.0000000000000013

Antoniadis, V., Koliniati, R., Efstratiou, E., Golia, E., Petropoulos, S., 2016. Effect of soils with varying degree of weathering and pH values on phosphorus sorption. Catena (Amst) 139, 214–219. https://doi.org/10.1016/J.CATENA.2016.01.008

Barbieri, P., MacDonald, G.K., Bernard de Raymond, A., Nesme, T., 2021. Food system resilience to phosphorus shortages on a telecoupled planet. Nature Sustainability 5, 114–122. https://doi.org/10.1038/s41893-021-00816-1

Barretto, A.G.O.P., Berndes, G., Sparovek, G., Wirsenius, S., 2013. Agricultural intensification in Brazil and its effects on land-use patterns: An analysis of the 1975-2006 period. Global Change Biology 19, 1804–1815. https://doi.org/10.1111/gcb.12174

Chabrillat, S., Ben-Dor, E., Cierniewski, J., Gomez, C., Schmid, T., van Wesemael, B., 2019. Imaging Spectroscopy for Soil Mapping and Monitoring, Surveys in Geophysics. Springer Netherlands. https://doi.org/10.1007/s10712-019-09524-0

Chen, S., Arrouays, D., Leatitia Mulder, V., Poggio, L., Minasny, B., Roudier, P., Libohova, Z., Lagacherie, P., Shi, Z., Hannam, J., Meersmans, J., Richer-de-Forges, A.C., Walter, C., 2022. Digital mapping of GlobalSoilMap soil properties at a broad scale: A review. Geoderma 409, 115567. https://doi.org/10.1016/J.GEODERMA.2021.115567

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

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

Dias, L.C.P., Pimenta, F.M., Santos, A.B., Costa, M.H., Ladle, R.J., 2016. Patterns of land use, extensification, and intensification of Brazilian agriculture. Global change biology 22, 2887–2903. https://doi.org/10.1111/gcb.13314

Diek, S., Fornallaz, F., Schaepman, M.E., de Jong, R., 2017. Barest Pixel Composite for agricultural areas using landsat time series. Remote Sensing 9. https://doi.org/10.3390/rs9121245

Elser, J.J., 2012. Phosphorus: a limiting nutrient for humanity? Curr Opin Biotechnol 23, 833–838. https://doi.org/10.1016/J.COPBIO.2012.03.001

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

Forkuor, G., Hounkpatin, O.K.L., Welp, G., Thiel, M., 2017. High resolution mapping of soil properties using Remote Sensing variables in south-western Burkina Faso: A comparison of machine learning and multiple linear regression models. PLoS ONE 12, 1–21. https://doi.org/10.1371/journal.pone.0170478

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

Gibbs, H.K., Ruesch, A.S., Achard, F., Clayton, M.K., Holmgren, P., Ramankutty, N., Foley, J.A., 2010. Tropical forests were the primary sources of new agricultural land in the 1980s and 1990s. Proceedings of the National Academy of Sciences of the United States of America 107, 16732–16737. https://doi.org/10.1073/pnas.0910275107

Graesser, J., Aide, T.M., Grau, H.R., Ramankutty, N., 2015. Cropland/pastureland dynamics and the slowdown of deforestation in Latin America. Environmental Research Letters 10, 34017. https://doi.org/10.1088/1748-9326/10/3/034017

Leite, C.C., Costa, M.H., Soares-Filho, B.S., De Barros Viana Hissa, L., 2012. Historical land use change and associated carbon emissions in Brazil from 1940 to 1995. Global Biogeochemical Cycles 26, 1–13. https://doi.org/10.1029/2011GB004133

Lopes, A.S., Guimarães, G.L.R., 2016. A career perspective on soil management in the Cerrado region of Brazil, Advances in Agronomy. Elsevier Inc. https://doi.org/10.1016/bs.agron.2015.12.004

Martinelli, L.A., Naylor, R., Vitousek, P.M., Moutinho, P., 2010. Agriculture in Brazil: Impacts, costs, and opportunities for a sustainable future. Current Opinion in Environmental Sustainability 2, 431–438. https://doi.org/10.1016/j.cosust.2010.09.008

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

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. https://doi.org/10.1016/j.geoderma.2019.01.025

Mondal, A., Khare, D., Kundu, S., Mondal, S., Mukherjee, S., Mukhopadhyay, A., 2017. Spatial soil organic carbon (SOC) prediction by regression kriging using remote sensing data. Egyptian Journal of Remote Sensing and Space Science 20, 61–70. https://doi.org/10.1016/j.ejrs.2016.06.004

Padarian, J., Minasny, B., McBratney, A.B., 2019. Using deep learning to predict soil properties from regional spectral data. Geoderma Regional 16, e00198. https://doi.org/10.1016/j.geodrs.2018.e00198

Pavinato, P.Sergio., Rocha, G.Cassoni., Cherubin, M.Roberto., Harris, Ian., Jones, D.Leonard., Withers, P.J.Anthony., 2020. Map of total phosphorus content in native soils of Brazil. Sci Agric 78, e20200077. https://doi.org/10.1590/1678-992X-2020-0077

Poppiel, R.R., Lacerda, M.P.C., Safanelli, J.L., Rizzo, R., Oliveira, M.P., Novais, J.J., Demattê, J.A.M., 2019. Mapping at 30 m resolution of soil attributes at multiple depths in midwest Brazil. Remote Sensing 11. https://doi.org/10.3390/rs11242905

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

Rogge, D., Bauer, A., Zeidler, J., Mueller, A., Esch, T., Heiden, U., 2018. Building an exposed soil composite processor (SCMaP) for mapping spatial and temporal characteristics of soils with Landsat imagery (1984–2014). Remote Sensing of Environment 205, 1–17. https://doi.org/10.1016/j.rse.2017.11.004

Rosemarin, A., Ekane, N., 2016. The governance gap surrounding phosphorus. Nutr Cycl Agroecosyst 104, 265–279. https://doi.org/10.1007/S10705-015-9747-9/TABLES/2

Roy, E.D., Richards, P.D., Martinelli, L.A., Coletta, L. Della, Lins, S.R.M., Vazquez, F.F., Willig, E., Spera, S.A., VanWey, L.K., Porder, S., 2016a. The phosphorus cost of agricultural intensification in the tropics. Nature Plants 2, 2–7. https://doi.org/10.1038/NPLANTS.2016.43

Safanelli, J.L., Poppiel, R.R., Chimelo Ruiz, L.F., Bonfatti, B.R., de Oliveira Mello, F.A., Rizzo, R., Demattê, J.A.M., 2020. Terrain Analysis in Google Earth Engine: A Method Adapted for High-Performance Global-Scale Analysis. ISPRS International Journal of Geo-Information 9, 400. https://doi.org/10.3390/IJGI9060400

Schipanski, M.E., Bennett, E.M., 2021. The Phosphorus Cycle, in: Fundamentals of Ecosystem Science, Second Edition. Academic Press, pp. 189–213. https://doi.org/10.1016/B978-0-12-812762-9.00009-5

Schröder, J.J., Smit, A.L., Cordell, D., Rosemarin, A., 2011. Improved phosphorus use efficiency in agriculture: A key requirement for its sustainable use. Chemosphere 84, 822–831. https://doi.org/10.1016/J.CHEMOSPHERE.2011.01.065

Shary, P.A., 2023. Environmental Variables in Predictive Soil Mapping: A Review. Eurasian Soil Science 56, 247–259. https://doi.org/10.1134/S1064229322602384/FIGURES/6

Simpson, R.J., Oberson, Astrid, Culvenor, Richard A, Ryan, Megan H, Veneklaas, Erik J, Lambers, Hans, Lynch, Jonathan P, Ryan, Peter R, Delhaize, Emmanuel, Andrew Smith, F., Smith, Sally E, Harvey, Paul R, Richardson, Alan E, Wissuwa J Simpson, M.R., Culvenor, R A, Richardson, A E, Oberson, A, Ryan, M H, Veneklaas, E J, Lambers, H, Lynch, J P, Ryan, P R, Delhaize, E, Smith, F.A., Smith, S E, Harvey, P R, 2011. Strategies and agronomic interventions to improve the phosphorus-use efficiency of farming systems. Plant and Soil 349, 89–120. https://doi.org/10.1007/S11104-011-0880-1

Steffen, W., Richardson, K., Rockström, J., Cornell, S.E., Fetzer, I., Bennett, E.M., Biggs, R., Carpenter, S.R., De Vries, W., De Wit, C.A., Folke, C., Gerten, D., Heinke, J., Mace, G.M., Persson, L.M., Ramanathan, V., Reyers, B., Sörlin, S., 2015. Planetary boundaries: Guiding human development on a changing planet. Science (1979) 347. https://doi.org/10.1126/SCIENCE.1259855/SUPPL_FILE/STEFFEN-SM.PDF

Tollefson, J., 2010. Food: The global farm. Nature 466, 554–556. https://doi.org/10.1038/466554a

Tziolas, N., Tsakiridis, N., Ben-Dor, E., Theocharis, J., Zalidis, G., 2020. Employing a multi-input deep convolutional neural network to derive soil clay content from a synergy of multi-temporal optical and radar imagery data. Remote Sensing 12. https://doi.org/10.3390/RS12091389

Viscarra Rossel, R.A., Bui, E.N., 2016. A new detailed map of total phosphorus stocks in Australian soil. Science of the Total Environment 542, 1040–1049. https://doi.org/10.1016/J.SCITOTENV.2015.09.119

Withers, P.J.A., Rodrigues, M., Soltangheisi, A., De Carvalho, T.S., Guilherme, L.R.G., Benites, V.D.M., Gatiboni, L.C., De Sousa, D.M.G., Nunes, R.D.S., Rosolem, C.A., Andreote, F.D., Oliveira, A. De, Coutinho, E.L.M., Pavinato, P.S., 2018. Transitions to sustainable management of phosphorus in Brazilian agriculture. Scientific Reports 8, 1–13. https://doi.org/10.1038/s41598-018-20887-z

Withers, P.J.A., van Dijk, K.C., Neset, T.S.S., Nesme, T., Oenema, O., Rubæk, G.H., Schoumans, O.F., Smit, B., Pellerin, S., 2015. Stewardship to tackle global phosphorus inefficiency: The case of Europe. Ambio 44, 193–206. https://doi.org/10.1007/S13280-014-0614-8/FIGURES/5

Zeraatpisheh, M., Ayoubi, S., Jafari, A., Tajik, S., Finke, P., 2019. Digital mapping of soil properties using multiple machine learning in a semi-arid region, central Iran. Geoderma 338, 445–452. https://doi.org/10.1016/j.geoderma.2018.09.006

Zhou, T., Geng, Y., Ji, C., Xu, X., Wang, H., Pan, J., Bumberger, J., Haase, D., Lausch, A., 2021. Prediction of soil organic carbon and the C:N ratio on a national scale using machine learning and satellite data: A comparison between Sentinel-2, Sentinel-3 and Landsat-8 images. Science of the Total Environment 755, 142661. https://doi.org/10.1016/j.scitotenv.2020.142661

2 THE BRAZILIAN SOILS CLAY FRACTION MAJOR OXIDES SPATIALIZED BY EARTH OBSERVATION STRATEGY

ABSTRACT

The soils clay fraction major oxides of tropical soils are iron oxides (Fe2O3), aluminum oxides (Al2O3), and silicon oxides (SiO2). In tropical soils, these oxides are directly or indirectly responsible for the soil's capacity to provide ecosystem services. Additionally, they are used to classify soils into different pedological classes. Despite the importance of these oxides, quantifying them on a large scale is not an easy task. Moreover, the most common method for quantifying these oxides is laboratory sulfuric acid digestion, which is expensive, complex, and environmentally harmful. To overcome this issue and provide faster information, we developed a satellite technique associated with machine learning to map all agricultural areas in Brazil at a 30 m resolution. Additionally, we tested if the generated maps can be used to infer soil weathering, assist in the construction of pedological maps, and support agricultural crop management. We used a modeling dataset with 5,330 sites (0-20 cm and 80-100 cm) distributed across all 27 states. Six spectral variables obtained from the historical Landsat series (bare soil) and seven terrain attributes derived from a digital elevation model were used to determine Fe2O3, Al2O3, and SiO2 using the Random Forest algorithm. The predicted maps of oxides covered nearly 3.48 million km² (~40% of the national territory). The best predictions were observed for Fe2O3 in the surface layer (RMSE = 47.0, RPIQ = 1.85, and R2 = 0.65), while the lowest predictions were for SiO2 in the subsurface layer (RMSE = 66.7, RPIQ = 1.55, and R2 = 0.19). It was possible to infer soil weathering using the Ki index. Our results were consistent with legacy maps, where highly weathered soils were observed in the plateaus of the cerrado biome, while younger soils were observed in the arid Caatinga biome and waterlogged soils in the Pantanal biome. Our maps also demonstrated a high potential for grouping pedological soil classes. Furthermore, we observed a relationship between oxide contents and the vigor of sugarcane crops, indicating that our findings can assist in crop management. Moreover, our satellite-based technique supported by machine learning was capable of predicting this information for large areas with high spatial resolution.

Keywords: Digital Soil mapping, ecosystem services, Soil weathering, machine learning, bare soil reflectance.

2.1 Introduction

The soil plays several ecosystem services (ES) that are of utmost importance for the maintenance of life on the planet (McBratney et al., 2014). The soil's ability to provide ecosystem services is determined by various intrinsic soil properties, such as texture, organic carbon content, depth, density, available water, and cation exchange capacity (Adhikari and Hartemink, 2016; Hewitt et al., 2015). These soil properties have different spatial distributions and can vary considerably across the Earth's crust, influencing the soil's capacity to deliver ecosystem services (Adhikari and Hartemink, 2016). In tropical soils, where the degree of weathering is high, one of the most important properties in the production of ES are clay fraction oxides (Kirsten et al., 2021; Silva et al., 2021). Soil oxides play several roles in the soil matrix. Among these functions, there is the maintenance of soil structure, potential for organic carbon retention, potential water retention, influences the pH magnitude, availability of nutrients for plants, among others (Kirsten et al., 2021; Ukabiala et al., 2021). Against this, the task of quantifying these oxides can help in understanding the ability of a given soil to produce ES.

The clay fraction major oxides of tropical soils are iron oxides (Fe2O3), aluminum oxides (Al2O3) and silicon oxides (SiO2) (Schaefer et al., 2008). In addition to the importance of these oxides in the production of ES, in Brazil, they are used in the Brazilian soil classification system (SiBCS). The Fe2O3 content is used in SiBCS for classification in the third categorical level, while Al2O3 and SiO2 are used to measure the degree of soil weathering (Embrapa Solos, 2018).

Despite the great importance of these oxides in the soil, few efforts are made to map them in the tropical regions of the world. This is due to the fact that laboratory methodologies are expensive, time consuming and have a high potential for environmental pollution (its quantification is performed by use of acid digestion with strong acids such as sulfuric acid), making difficult to obtain pedogenic oxides information for large areas, with high spatial resolutions (Silva et al., 2019). However, despite the great challenge of mapping the contents of these oxides, innovative techniques for digital soil mapping (DSM) can enable the construction of maps containing their contents with high spatial resolution and low levels of uncertainty, even with low sample densities (Mendes et al., 2022; Safanelli et al., 2021a).

In DSM, robust mathematical models capable of predicting a given soil property based on environmental covariates are used (Hengl et al., 2017; Vaysse and Lagacherie, 2015). Due to the advancements in computational techniques, it is currently possible to implement robust machine learning (ML) algorithms that exhibit high reliability (Wadoux et al., 2020a). Numerous algorithms are available for use in several scientific fields. However, in DSM studies, the Random Forest (RF) algorithm has been the most widely used (Padarian et al., 2020). The RF has great acceptance in the user community due to being a robust model composed of multiple decision trees that are not correlated with each other. Due to these characteristics, the RF has high accuracy in prediction processes with minimal chance of overfitting in the models (Sheykhmousa et al., 2020; Wadoux et al., 2020a). In addition, due to the structure of RF being formed by decision trees, it allows for working with linear regression and categorical data, encompassing various soil properties (Sheykhmousa et al., 2020).

The final quality of DSM depends greatly on the environmental covariates used for modeling a specific soil attribute. Typically, environmental covariates are used that represent part of the physical and chemical

processes that govern the spatial variation of soil (McBratney et al., 2003; Wadoux et al., 2020). Among the environmental covariates, the most commonly used ones are digital elevation models, vegetation indices, climate covariates, geological maps, and surface reflectance obtained by satellites (Lamichhane et al., 2019). In addition to these covariates, recent studies propose using environmental covariates that represent bare soil reflectance (Rosin et al., 2023; Safanelli et al., 2021b). Despite the great potential of bare soil reflectance in DSM, obtaining this reflectance is a major challenge due to the vegetation present on soils most of the time. In order to solve this problem, Demattê et al. (2018) developed an algorithm (Geospatial Soil Sensing System - GEOS3) capable of retrieving bare soil reflectance from historical series of Landsat images by aggregating the pixels that were exposed at least once throughout the historical series into a synthetic soil image (SYSI).

Several studies worldwide have demonstrated the potential of exposed soil reflectance for soil attribute mapping (Gasmi et al., 2021; Rizzo et al., 2020; Silvero et al., 2021; Yang et al., 2020). The prediction of soil attributes based on reflectance has been consolidated with the use of spectroscopy in laboratories (Barra et al., 2021). This consolidation is due to the fact that the concentrations of soil attributes alter their spectral signature. The spectral signature can be altered either at specific wavelengths or across the entire electromagnetic spectrum, such as the iron oxide content altering reflectance in the red region or clay content altering reflectance intensity across the spectrum (Ackerson et al., 2015; Demattê et al., 2007). This characteristic makes the reflectance of exposed soil when applied in DSM the most important environmental covariates in predicting soil attributes (Poppiel et al., 2019; Rosin et al., 2023).

We expect that topsoil reflectance measured by satellite will be related to the clay fraction major oxides Fe2O3, Al2O3, SiO2 by the electronic transition and molecular vibrations in the 350-2500 nm spectral range. With this, we expect to quantify and spacialize these properties by remote sensing associated with machine learning tehcniques at surface and undersurface. In addition, we expect that these oxides maps can be utilized to infer on the weathering of Brazilian agricultural soils, support soil class mapping, and aid in agricultural crop management. The information, for all agriculture areas of Brazil, in a high special definition (30 m) may bring light to researchers, farmers and consultors, on understanding their soil dynamics.

2.2 Materials and methods

2.2.1 Study area

The study area is represented by the entire agriculture territory of Brazil. Brazil is a country with a large dimension (with about 8.5 million km2) and its variability of soils. This soil variability is due to the great formation factors, mainly, the climate, the time of formation, the parental material and the topography. Brazil's climate varies across the territory between tropical, subtropical until semiarid being represented by 12 classes (Figure A1). The time of formation of the parental materials that gave rise to Brazilian soils are most formed in the Cenozoic, Mesozoic, Neoproterozoic and Paleoproterozoic eras (Figure A1). Regarding the parental material, the majority of rocks are sedimentary, formed by sediments with diverse origins, such as weathering of the Andes (Amazon basin), ocean floor (São Francisco craton region), desert sediment as in the Bauru and Botucatu

formations. In addition to sedimentary rocks, a large part of Brazil's geology is composed of volcanic and plutonic rocks. In the region encompassing the states of Paraná, São Paulo, Rio Grande do Sul, Minas Gerais, Goiás, and Mato Grosso do Sul, there is a large mass of basaltic rocks formed from basaltic eruptions. These rocks are extrusive and contain higher levels of iron and easily weatherable minerals. Along the Brazilian coastline, there are many plutonic rocks that were formed underground (intrusives) during continental drift. These rocks possess highly crystallized minerals like quartz.

Another factor of formation with high variability in Brazil is the topography. Many mountains are present in coastal regions, while the interior of the country has large plateaus that favor soil weathering. There are also extensive floodplain basins, such as the Pantanal. All of these variations in soil formation factors contribute to the existence of soils with different levels of weathering and, consequently, different levels of oxides that were mapped in this study.

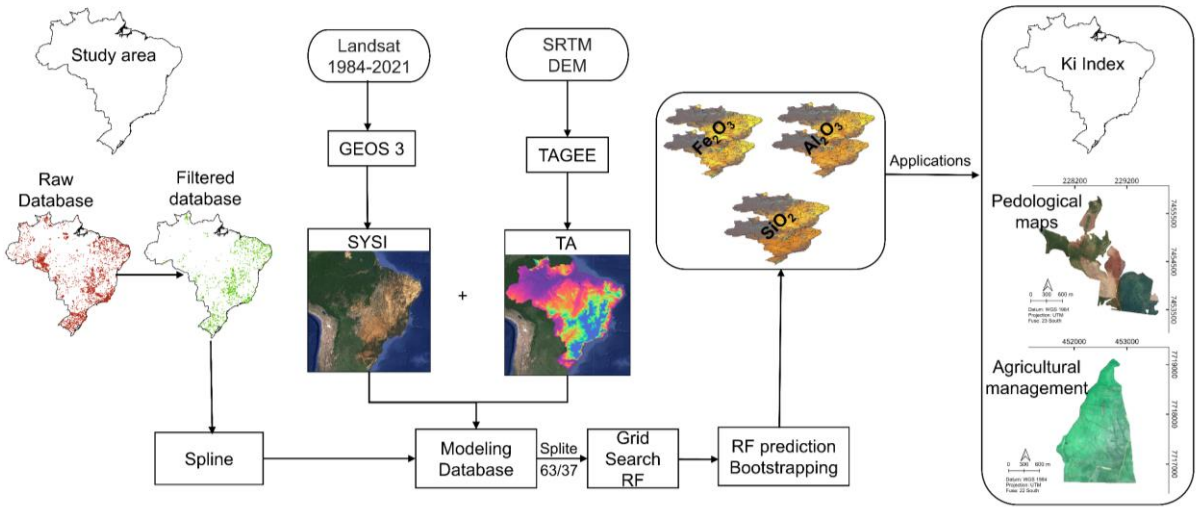

Figure 1. Flowchart of the soils clay fraction major oxides prediction and its possible applications tested. GEOS3: Geospatial Soil Sensing System; SYSI: Synthetic Soil Image; DEM: Digital Elevation Model; STRM: Shuttle Radar Topography Mission; TAGEE: Terrain Analysis in Google Earth Engine; TA: Terrain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

2.2.2 Soil observations

The soil observations were based several databases, which came from two different sources: i) Publicly available national soil profile databases with nearly 9.119 sites (Samuel-Rosa et al., 2019), ii) Brazilian Soil Spectral Library with nearly 45,000 sites (Demattê et al., 2019). The final database containing the soil observations used in this study were consolidated and modified in four different steps. The first step was to eliminate the sites with not accurate coordinates (accuracy less than 100 m). In the second step, all the sites that had information on at least one of the Fe_2O_3, SiO_2, Al_2O_3 oxides were filtered. The pseudo-total concentrations of these oxides were obtained by the sulfuric acid digestion proposed by Vettori (1969), with modifications proposed by Raij and Valadares (1974). This extraction method can access the oxides in the secondary minerals. In this way, the extracted oxides mostly make up from the clay fraction of the soil.

Once the second step was completed, a total of 5,330 sites were obtained covering the entire Brazilian territory. These sites were submitted to a third step, which consisted of detecting those located in the areas covered by SYSI. In the fourth step, each site remaining from the third step, containing information of oxides at more than one depth, was subjected to interpolation along the profile to standardize the depths making them suitable for our study. The interpolation method was the quadratic spline function implemented in the GSIF R package (Hengl and Macmillan, 2019). The use of the quadratic spline function for interpolation of soil attributes along the profile was proposed by Bishop et al. (1999), where, in tests comparing with other similar methods, it excelled, since then it has been widely used in soil science studies (Ellili-Bargaoui et al., 2020; Sulaeman et al., 2013; Tayebi et al., 2021). Once interpolation, we obtained layers of 20 cm, starting at the soil surface to 100 cm depth. Two depths were used in this study 0-20 and 80-100 cm. After all the steps taken, the database was ready to be used. In the final database, each oxide had a different number of sites. In addition, the sampling density was also different for each depth used (Table 1).

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

Oxide	**Layer**	**Nº observations**	
		Brazil	SISY Coverage
Fe_2O_3	0-20 cm	5211	2715
	80-100 cm	3671	1874
Al_2O_3	0-20 cm	4551	2253
	80-100 cm	3262	1608
SiO_2	0-20 cm	4750	2380
	80-100 cm	3393	1697

SYSI: Synthetic Soil Image

2.2.3 Environmental covariates

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

The concentrations of the oxides are directly associated with soil formation factors which was one of the criteria for choosing the environmental covariates. Despite the great importance of all formation factors, the relief factor was considered. It was not possible to use the factors as climate, parental material, time and organisms, due to low spatial resolution or lack of information. To represent the relief factor, we consider environmental covariates that were previously used to predict soil attributes with satisfactory results in Safanelli et al (2021b). These terrain attributes were as follows: slope, altitude, north and east slope, horizontal curvature, vertical curvature, and a relief shape index. These variables were based on a digital elevation model (DEM) from the Shuttle Radar Topography Mission (SRTM) with 30 m (Farr and Kobrick, 2000). To calculate the

environmental variables, the Terrain Analysis in Google Earth Engine (TAGEE) algorithm proposed by (Safanelli et al., 2020b) was used. The individual information on the environmental covariates derived from the DEM are shown in table 2.

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

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

Table 2. Environmental covariates used for modeling major oxides.

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

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

2.2.4 Soils clay fraction major oxides modeling by Random Forest

In this study, we employed the Random Forest (RF) machine learning algorithm to map the content of soil**s** clay fraction major oxides for Brazilian agriculture areas. The RF, proposed by Breiman (2001), is an ML algorithm that operates by constructing a set of independent decision trees. Each tree is trained on a random sample of the training data. During prediction, each tree generates an estimate, and the final prediction is determined by a combination of the estimates from all the trees (Breiman, 2001). This ensemble approach allows Random Forest to capture complex relationships between variables and produce more robust and accurate results (Cutler et al., 2012).

RF can be used for both linear regression and classification tasks. This characteristic makes the algorithm highly attractive for use in DSM, as it can handle a wide range of soil attributes effectively (Heung et

al., 2016). Due to this, RF is the most widely used ML algorithm in the construction of digital soil maps (Wadoux et al., 2020). In recent years, several studies have demonstrated the potential of RF to construct digital soil maps with a high level of detail and accuracy for various soil attributes. RF was used to map carbon stock (Tayebi et al., 2021), soil mineralogy (Rosin et al., 2023), soil texture (Dharumarajan and Hegde, 2022) , parent material of soils (Heung et al., 2014).

2.2.4.1 Model tuning, performance and validation

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

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

A bootstrapped sampling was used to determine the training dataset for the model while the test dataset was defined by the remaining samples (out of the bag) that were not selected in each bootstrapping. The bootstrapping sampling was performed individually for each decision tree in the random forest, and in the end, they were aggregated by mean. Bootstrapping is a resampling method that randomly samples observations with replacement from the original dataset to create a new dataset of the same size, covering, in general, around 67% of the entire dataset. We applied the bootstrap method up to 500 times, which is equivalent to the maximum number of trees tested for the random forest.

The optimal combination (FS, nPR and minSL) with the smallest Root Mean Squared Error (RMSE) for the test set was selected for each soil attribute. The accuracy of the developed models was evaluated by the coefficient of determination (R^2), RMSE, Ratio of Performance to InterQuartile distance (RPIQ) of the calibration and validation data sets. The calculation of these parameters was performed according to the equations 1, 2, 3, 4, 5.

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

$$R^2 = 1 - \frac{SS_{residuals}}{SS_{total}} \quad (2)$$

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

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

$$RPIQ = \frac{IQR_y}{RMSE_{\hat{y}}} \quad (5)$$

where y_i is the vector of measured values, $\hat{y}_i$ is the vector of predicted values, $\underline{y}$ is the mean of vector y, and IQR is the interquartile range.

2.2.4.2 Soils clay fraction major oxides and uncertainty mapping

After selecting the best FS, nPR and minSL optimal combination, it was used to predict the major soil oxides for the study area. The raster files corresponding to the environmental covariate were used as input data in Fe_2O_3, Al_2O_3 and SiO_2 models, and the output was the respective maps. The final estimated value for each pixel in the final maps was formed by the aggregated mean of the estimations obtained by each regression tree (Equation (6)). Additionally, the standard deviation of the estimations of the predictions was also calculated (Equation (7)). The standard deviation of the ensemble was used to calculate the spatial distribution of uncertainty, which was defined by the confidence interval (95% probability level) standardized for the average estimate at the pixel level (Equation (8)).

$$\hat{f}_p = \frac{1}{B}\sum_{b=1}^{B} f_b\left(X_p\right) \quad (6)$$

$$S_p = \sqrt{\frac{\sum_{b=1}^{B}\left(f_b(X_p)-\hat{f}_p\right)^2}{B-1}} \quad (7)$$

$$CR_{std} = \frac{\left(\hat{f}_p+1.96\frac{S_p}{\sqrt{B}}\right)-\left(\hat{f}_p-1.96\frac{S_p}{\sqrt{B}}\right)}{\hat{f}_p} \quad (8)$$

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

2.2.5 Oxides maps interpretation

The interpretation of the oxide maps obtained in the modeling was conducted using three distinct approaches. The first approach involved reclassifying the maps based on their oxide content into three distinct classes. Subsequently, the percentage of areas encompassing each class was quantified.

The Fe_2O_3 maps were divided according to the SiBCS classification system, where hypoferric soils had oxide contents below 80 g.kg-1, mesoferric soils ranged from 80 to <180 g.kg-1, and ferric soils ranged from 180 to 360 g.kg-1. However, for the Al_2O_3 and SiO_2 maps, this classification system was not applicable. Therefore, we proposed intervals based on the range of oxide contents obtained in the maps. We defined soils with oxide contents less than 80 g.kg-1 as low, soils ranging from 80 to <150 g.kg-1 as medium, and soils with oxide contents above 150 g.kg-1 as high, for both oxides.

The second interpretation focused on the spectral signature of the aforementioned classes for all oxides. We compared the average spectral signature for each class obtained from the sampled points with the average spectral signature of the classes in the predicted map.

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

2.2.6 Soils clay fraction major oxides maps applications

We tested three applications of the soil oxide maps: Infer to the degree of chemical weathering of soils; potential to contribute to the development of high resolution soil class maps; and finally, the prospect of utilizing these maps to guide management practices in agricultural fields.

2.2.6.1 Chemical weathering index calculation

To infer on the chemical weathering of soils, we calculated the Ki index (equation 9). The Ki calculation was performed on a pixel-by-pixel basis using the maps of SiO_2 and Al_2O_3. The Ki weathering index expresses an idea of the removal of Si by the soil weathering process. The higher the value of the index, the lower the degree of weathering of the soil.

$$Ki = \frac{SiO2(g.kg^{-1})}{Al2O3(g.kg^{-1})} x\ 1.70 \qquad (9)$$

2.2.6.2 Application in pedological maps

In the Brazilian Soil Classification System (SiBCS), the content of pedogenic oxides and the degree of weathering are used to soil classification. To test the potential contribution of our maps to the construction of pedological maps, we selected a 182-hectare farm located in the municipality of Rafard, São Paulo (Figure 1). The chosen area exhibits significant variations in lithology, with predominance of immature psammites with heterogeneous grain size, transitioning to feldspathic psammites and even arkosic sandstones. Concurrently with this lithology, eruptive elements of dikes from the Serra Geral formation occur, composed of intrusive bodies of tholeiitic basalt (Nanni and Demattê, 2006). This complex lithology has contributed to the formation of different soil classes. A conventional soil class map was constructed by Bazaglia Filho et al. (2013) for the area. In this map, five soil classes were identified at the first categorical level of SiBCS.

Using our oxide maps, we performed an unsupervised classification using the K-means algorithm. The aim of this classification was to generate a map containing groups based on the content of pedogenic oxides. All six oxide maps (Fe_2O_3, Al_2O_3, and SiO_2 in the 0-20 cm and 80-100 cm layers) were used as input variables in the classification. Additionally, we included the Ki maps from both layers as input variables. The number of groups was set to 5, representing the number of soil from conventional soil class map. Finally, we compared the resulting map from the unsupervised classification with the conventional soil class map.

2.2.6.3 Application in agricultural management

Oxides have a direct relationship with crop vigor (Bronick and Lal, 2005). The presence of these pedogenic oxides improves soil structure, increases soil organic matter content, and consequently enhances the soil's water-holding capacity (Barthès et al., 2008; Kirsten et al., 2021). In light of this, we tested whether the content of oxides found in our maps are related to the vegetative vigor of plants.

For it, we selected a sugarcane farm located in Itapura, São Paulo, with an area of 229 hectares (Figure 1). This area was chosen because it had two lithologies: sandstone and basalt, with an abrupt transition between them. The vegetative vigor of sugarcane plants was assessed using the Normalized Difference Vegetation Index (NDVI). NDVI is a widely used index for inferring vegetation vigor, as it utilizes the difference between near-infrared reflectance and red reflectance (Huang et al., 2021). The calculation of NDVI is performed using Equation 10, and its values range from -1 to +1. In vegetation, values close to 0 indicate low vegetative vigor, while values close to 1 represent high vegetative vigor.

$$NDVI = \frac{NIR-Red}{NIR+Red} \quad (10)$$

Where NIR represents the reflectance in the near-infrared band, and Red represents the reflectance in the red band.

The NDVI map was calculated using images obtained from the MSI sensor aboard the Sentinel-2 satellite. Since our objective was to map at the farm level, we chose Sentinel-2 over Landsat due to its higher spatial resolution (10m), which ensures a greater level of detail. We used a single image obtained by calculating the pixel-wise median from all images (with less than 20% cloud cover) acquired in May 2020. The image acquisition period preceded the harvest, representing the final stages of sugarcane development. The NDVI map was subjected to an unsupervised classification using the K-means algorithm. Two classes were defined in the final map, with one class representing plants with higher vegetative vigor and the other class representing lower vegetative vigor. The map containing the vegetative vigor classes was compared with the soil oxide maps.

2.3 Results and discussion

2.3.1 Descriptive statistics

The statistical analysis of the sample data containing the contents of the main soil oxides in the clay fraction shows high variability (Figure 2). The Fe_2O_3 contents ranged from 0 to 385 g kg^{-1}, SiO_2 varied from 0 to 390 g kg^{-1}, and Al_2O_3 ranged from 0.34 to 396 g kg^{-1}. The contents of these oxides in the soil clay fraction are related to two factors: the type of parent material and its degree of weathering (Schaefer et al., 2008). The high variability observed in oxide contents is explained by the extensive geodiversity throughout the Brazilian territory (de Paula Silva et al., 2021). Differences were also observed in the contents between the evaluated layers. All studied oxides have higher concentrations in the 80-100 cm layers (Figure 2). Minor differences were observed for Fe_2O_3, where the median content in the 0-20 cm layer was 41 gkg^{-1} and in the 80-100 cm layer it was 57 g kg^{-}

[1]. On the other hand, for SiO_2 and Al_2O_3 oxides, the differences were more pronounced, with median contents in the 0-20 cm layer of 118 and 100 g kg^{-1}, and in the 80-100 cm layer of 163 and 154 g kg^{-1}, respectively. The surface layers are in greater contact with the weathering agents and causes a more intense loss of elements than in the deeper layer, especially in tropical regions, resulting in lower quantities of oxides in the surface layers (Guimarães et al., 2021).

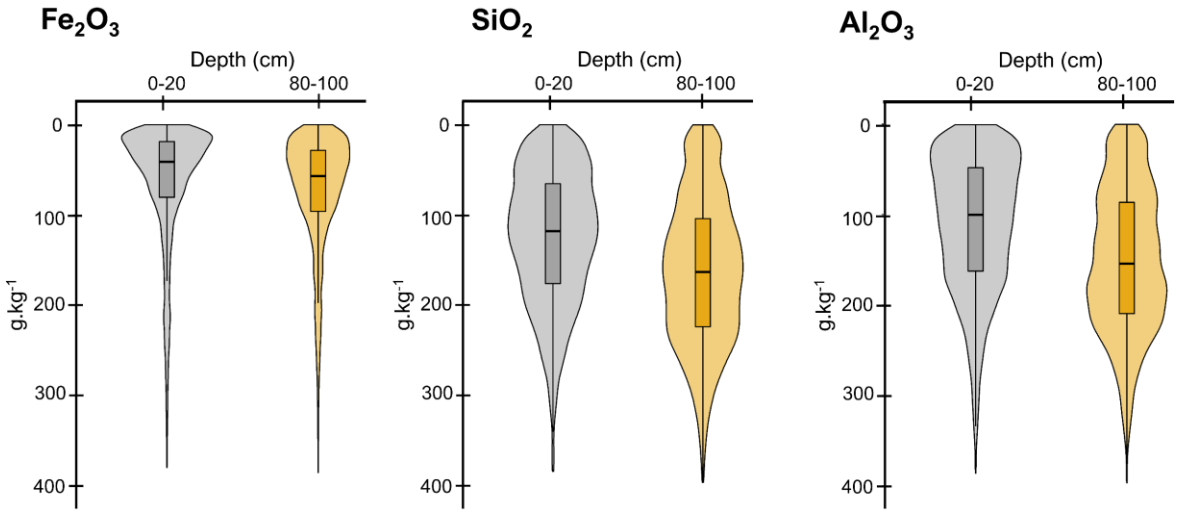

Figure 2. Descriptive statistics of major soil oxides data observed in the study area.

2.3.2 Spearman correlation between soils clay fraction major oxides and environmental covariates

The Spearman correlation analysis between the environmental covariates and the oxides revealed that, for all the oxides, the bands of SYSI showed the highest correlations (Figure 3). Among the bands of the SYSI, SWIR2 showed the highest correlation values, ranging from -0.46 (for SiO_2 in the 80-100 cm layer) to -0.63 (for Fe_2O_3 in the 0-20 cm layer). Pedogenic oxides are components of the soil clay fraction, and an increase in their content tends to reduce the reflectance intensity throughout the spectrum, resulting in negative correlation in the regions of the spectrum covered by the SYSI bands, particularly in the SWIR region (Terra et al., 2018). Different correlation values of oxides with SYSI were observed among the studied soil layers, with the 0-20 cm layer consistently showing higher correlations when compared to the 80-100 cm layer. This is due to the reflectance collected by satellites only comprising the surface layer of the soil, thus ensuring a direct relationship with oxide content in the 0-20 cm layer and an indirect relationship in the 80-100 cm layer (Mendes et al., 2019; Rosin et al., 2023b).

Regarding the remaining environmental covariates (terrain attributes), a positive correlation with oxide concentrations was observed. According to Marques et al. (2004), most high-altitude areas in Brazil are characterized by older parent materials that have undergone greater weathering, resulting in higher concentrations of pedogenic oxides in the clay fraction of the soil. Among the terrain attributes, elevation and slope show the

highest correlation values with oxides. Only SiO_2 in the 80-100 cm layer showed no correlation with elevation. The correlations with elevation ranged from 0.21 for SiO_2 in the 0-20 cm layer to 0.40 for Fe_2O_3 in the 0-20 cm layer. In addition, the correlations with slope ranged from 0.15 for SiO_2 in the 80-100 cm layer to 0.27 for Fe_2O_3 in the 0-20 cm layer. The higher correlation values between these variables and Fe_2O_3 content can be attributed to the fact that areas with higher altitudes are predominantly located within the region formed by the Serra Geral Formation, resulting from basaltic spill that led to soils with higher iron content (Famelli et al., 2021; Silva et al., 2020). On the other hand, soils located at lower elevations often consist of poor drainage condition soils, favoring Gleization processes that can translocate or remove iron (Couto et al., 2023; de Almeida, 2023; Ferreira et al., 2023). The environmental covariates representing terrain curvature had virtually no correlation with the oxides. We believe that this lack of correlation is due to curvature capturing spatial variability at short distances. However, our dataset has a low sample density, which does not adequately cover these shorter distances.

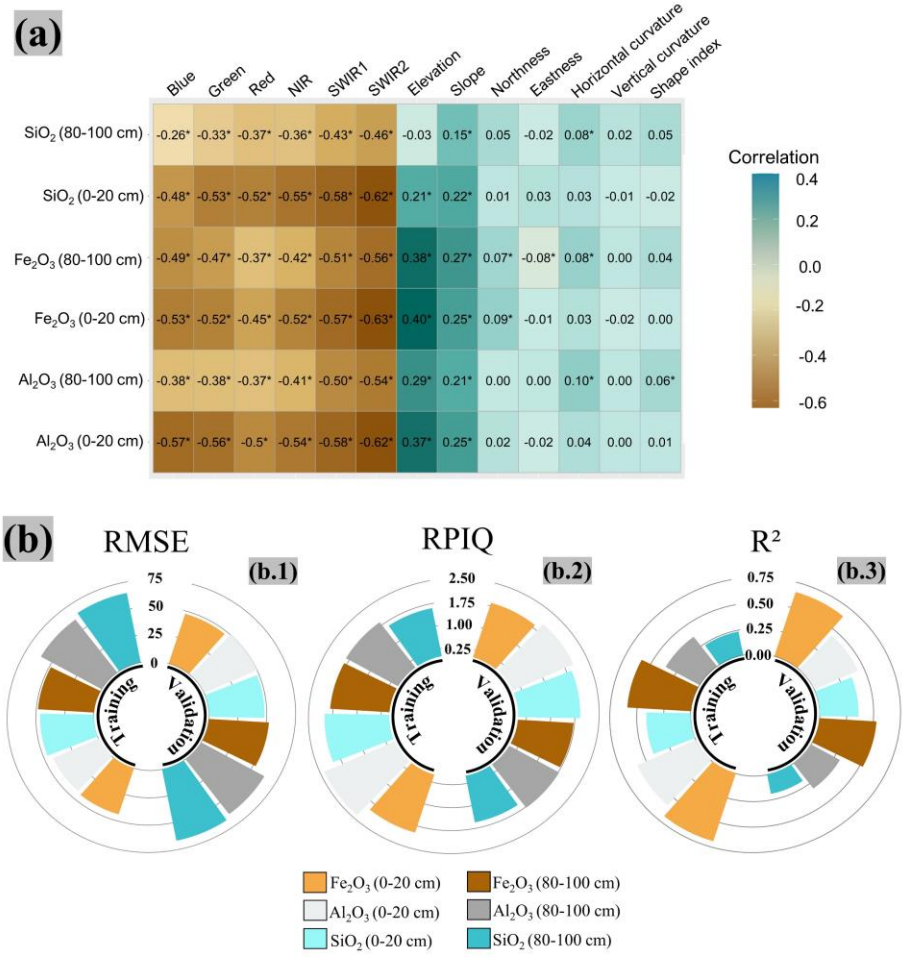

Figure 3. Performance evaluation of soils clay fraction major oxides prediction using out-of-the-bag validation. (a): Speraman's correlation analysis of environmental covariates with soil oxides, (b) parameters of models used to predict soil oxide maps, (b.1): Root Mean Squared Error (RMSE), (b.2): Ratio of Performance to InterQuartile distance (RPIQ), (b.3): coefficient of determination (R^2), * Signifcant at 5% probability level.

2.3.3 Performance of RF models for soils clay fraction major oxides prediction

The best performances were observed in the surface layers, especially for Fe_2O_3 in the 0-20 cm layer (RMSE = 47, RPIQ = 1.85 and R^2 = 0.65) (Fig. 3b). Conversely, Si_2O_3 presented the lower values, mainly for the 80-100 cm layer (RMSE = 66.7, RPIQ = 1.55 and R^2 = 0.19). Evaluating the performance of the models among the studied layers, the 0-20 cm layer performed better for all oxides. This superior performance in the surface layer can be attributed to the SYSI bands that represent the soil surface reflectance, and as demonstrated in section 3.2, showed the highest correlations with the oxides among all environmental covariates.

The Fe_2O_3 oxide showed more accurate models than SiO_2 and Al_2O_3 (Figure 3b). It occurred due to the less generalized occurrence of Fe_2O_3 in soil, being present in high concentration under specific conditions of relief, climate and parent material (Bigham et al., 2002; Wiriyakitnateekul et al., 2007). On the other hand, SiO_2 and Al_2O_3 are present in several soil minerals, such as kaolinite, with generalized occurrence in Brazilian territory (Schaefer et al., 2008). Besides, the Fe_2O_3 showed high Spearman correlations with environmental covariates, than SiO_2 and Al_2O_3, manly with SYSI bands and elevation, as explained above. which is key to explain the accuracy of RF prediction models.

2.3.4 Soils clay fraction major oxides maps

The predicted maps of oxides (Figures 4) comprised nearly 3.48 million km² (~40% of the national territory) that matches the area covered by the SYSI with bare soils between 1982 and 2021. These maps mainly span the agricultural areas of the country, where land cover (such as sugar cane, soy, grains, pasture, reforestation and others) tends to be removed due to agricultural practices. The maps for the 80-100 cm layer presented higher concentrations for all oxides, mainly SiO_2, following the patterns observed in the soil dataset (Figure 2). On the other hand, the maps of the 0-20 cm layer exhibit greater spatial variability. The increased spatial variability observed in the surface layer is a result of more intense weathering processes than in the subsurface layer (Poppiel et al., 2019; Rossel, 2011).

FeO2 exhibited the highest variation: 17.5-232 g.kg-1 (0-20 cm layer) and 29-211 g.kg-1 (80-100 cm layer). Subsequently, Al_2O_3 ranged from 26-207 g.kg-1 (0-20 cm layer) to 88-192 g.kg-1 (80-100 cm layer), while SiO_2 ranged from 50.5-170 g.kg-1 (0-20 cm layer) to 116-200 g.kg-1 (80-100 cm layer). Despite having slightly higher contents in the 80-100 cm layer, the spatial distribution of Fe_2O_3 contents is similar in both mapped layers (figures 4a and 4b). Approximately 61% of the mapped area in the 0-20 cm layer and 57% in the 80-100 cm layer consist of soils with low iron content (hypoferric soils). Additionally, 9% and 18% correspond to soils with moderate iron content (mesoferric soils), while 30% and 25% represent soils with high iron content (ferric soils), respectively.

Oxides Al3O2 and SiO_2 exhibit distinct spatial behaviors among the evaluated layers (Figures 4d, 4e, 4g, and 4h). The contents of these oxides are significantly higher in the subsurface layer throughout the mapped area. In the 0-20 cm layer, 49% and 38% of the mapped area consist of soils with low contents (>80 g.kg-1) of Al3O2 and SiO_2, respectively (Figure 4f and 4i). In contrast, in the 80-100 cm layer, there are no soils containing

low contents. Instead, approximately 64% of the mapped areas contain soils with Al_2O_3 contents ranging from 80 to 150 g.kg-1, and 62% contain SiO_2contents above 150 g.kg-1. In tropical soils, the majority of aluminum (Al) and silicon (Si) are found within the structures of clay minerals, primarily in kaolinite resulting from the process of monosialization. Due to the process of clay illuviation, these elements are transported to subsurface horizons (Quénard et al., 2011; Schaefer et al., 2008). On the other hand, iron (Fe) is predominantly found in pedogenic oxides such as hematite and goethite, which are products of intense weathering (Schwertmann and Taylor, 1989). The main pedogenetic process in soils with high iron content and intense weathering is ferralitization (Breemen and Buurman, 1998). Iron oxides form stable complexes with organic matter and other soil particles, limiting their movement within the soil profile (Schwertmann and Taylor, 1989; Van Wambeke et al., 1983).

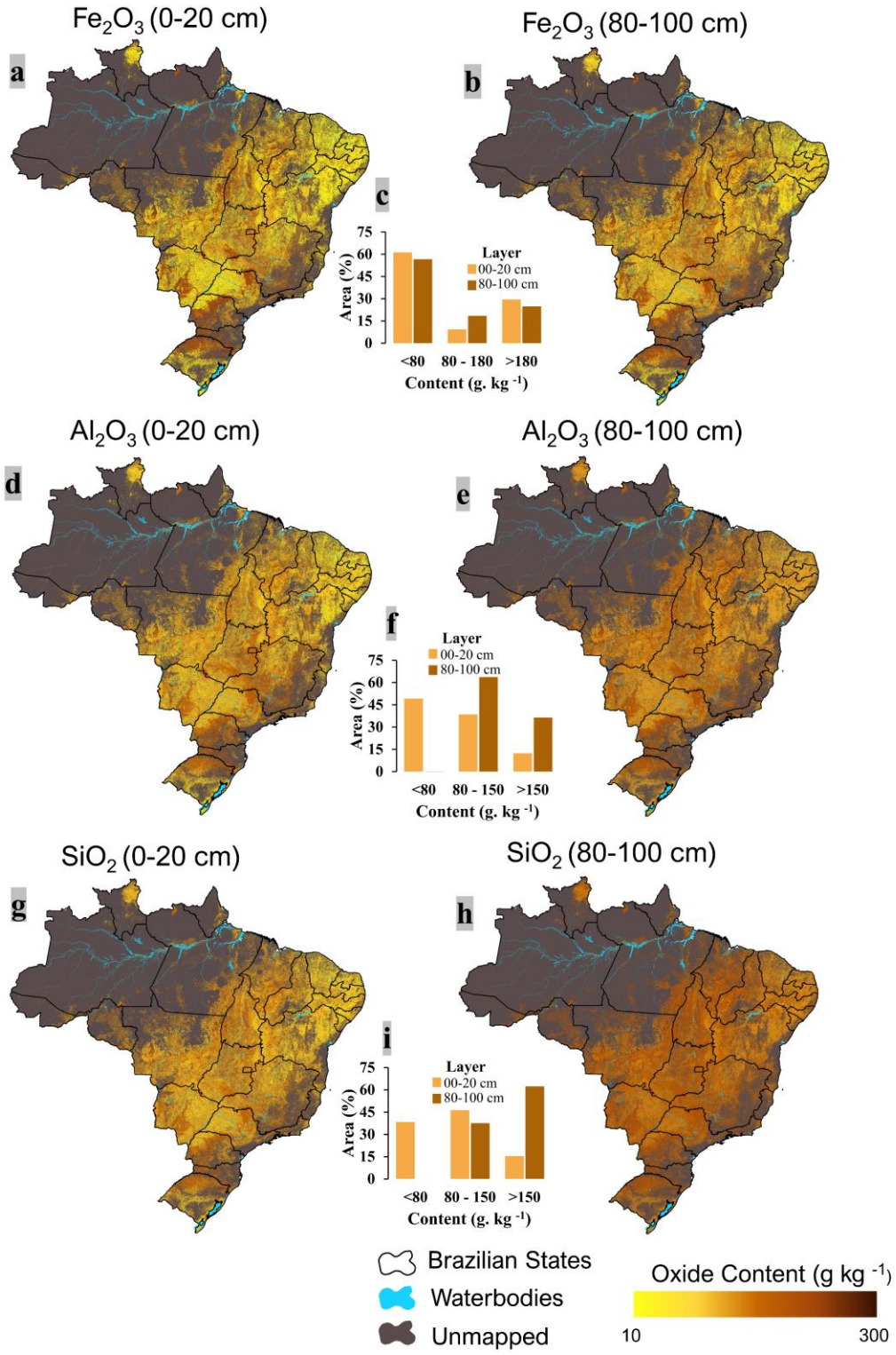

Figura 4. Soils clay fraction major oxides predicted maps. (a) Fe_2O_3 map in the 0-20 cm layer; (b) Fe_2O_3 map in the 80-100 cm layer; (c) percentage of mapped area as a function of Fe_2O_3 content; (d) Al_2O_3 map in the 0-20 cm layer; (e) Al_2O_3 map in the 80-100 cm layer; (f) percentage of mapped area as a function of Al_2O_3 content; (g) SiO_2 map in the 0-20 cm layer; (h) SiO_2 map in the 80-100 cm layer; (i) percentage of mapped area as a function of SiO_2 content.

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

Volcanic rocks are rich in minerals such as amphiboles, pyroxene, and olivine, which are less crystallized and more susceptible to chemical and physical weathering (Schenato et al., 2003). Additionally, these minerals contain significant amounts of iron, giving rise to ferric soils (Araujo et al., 2014; Schaefer et al., 2008). In addition, due to the advanced weathering of these soils, they have higher clay contents in their texture, resulting in higher levels of silicon and aluminum, constituents of kaolinites and gibbsite, being found (Schaefer et al., 2008). The lower contents observed in soils derived from Mmhg rocks are due to their chemical composition. These rocks are usually gneisses, which undergo significant transformations in their mineralogy due to the intensity of metamorphism (White et al., 2017). Despite the mineralogical and structural variations, gneisses predominantly exhibit minerals that are resistant to weathering, such as K-feldspars, muscovite, and quartz, along with accessory minerals like zircon (Buol and Weed, 1991; Câmara et al., 2021). Similar to Mmhg rocks, Sedimentary Siliciclastic rocks also predominantly consist of minerals that are highly resistant to chemical weathering, particularly quartz (Garzanti, 2019). This results in soils with lower clay contents and consequently lower oxide contents (Martins et al., 2005; L. S. Silva et al., 2019).

In soils developed from plutonic rocks, the oxide contents were higher in Plutonic Gabbroic and Ultramafic (Pgu) rocks compared to Plutonic Granitic (Pg) rocks. Pgu rocks have a mineralogical composition very similar to basalt, with a notable presence of ferromagnesian minerals that are more easily weathered (Guimarães et al., 2017; Silva et al., 2022). On the other hand, Pg rocks predominantly consist of highly crystallized minerals that are highly resistant to chemical weathering (Heuze, 1983). Median contents (above 80 $g.kg^{-1}$) for the predicted oxides were observed in soils developed on Sedimentary Carbonatic (Sc) rocks. These values differ from the literature, especially for Fe_2O_3, where soils developed on Sc rocks usually have lower contents than 80 $g.kg^{-1}$ (de Souza Oliveira et al., 2021; Pinheiro Junior et al., 2021; Silva et al., 2017). This inconsistency can be explained by the low mapping scale of lithology (1:5,000,000), where other rocks with high iron contents may be included in the map regions interpreted as Sc rocks.

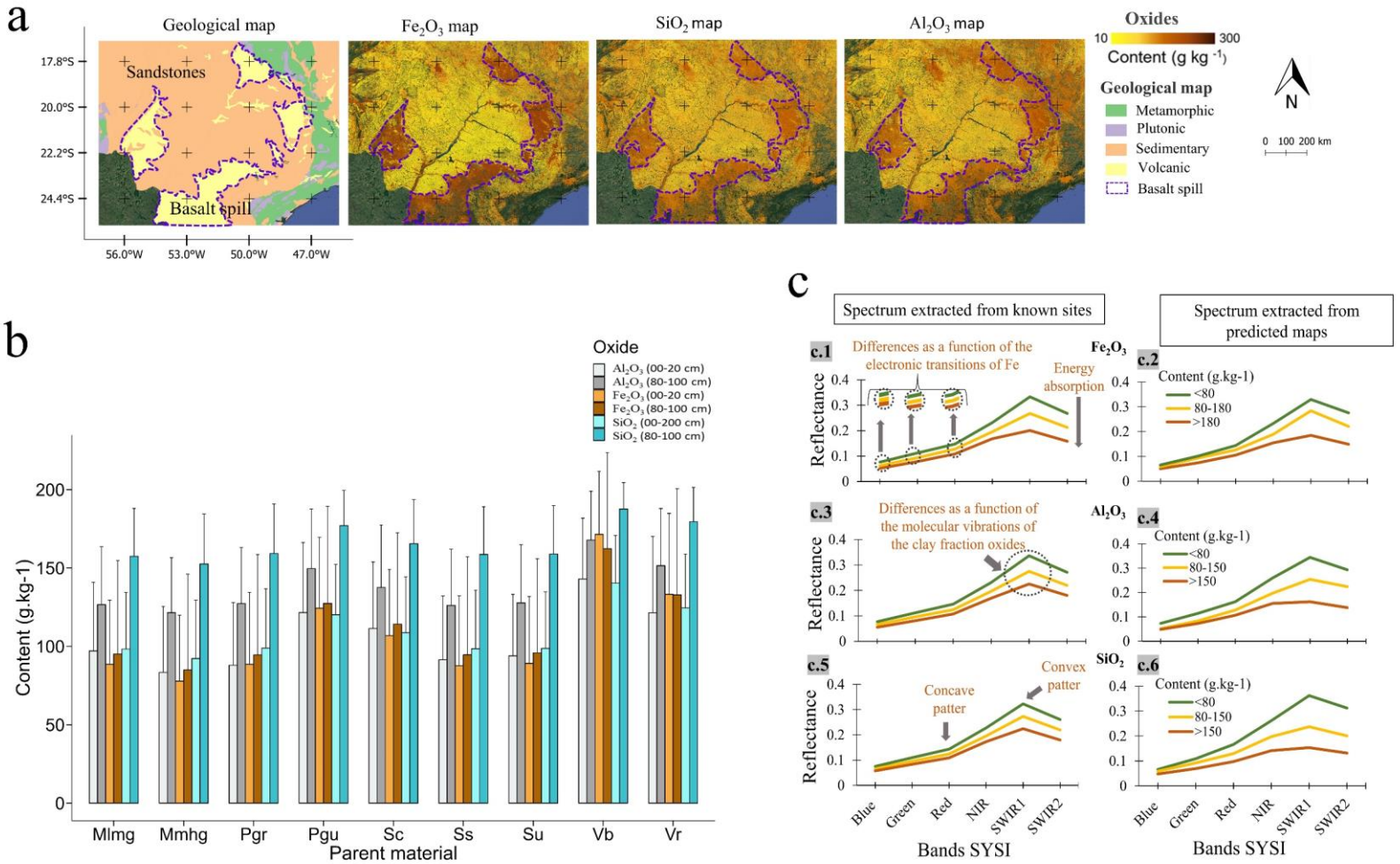

Figure 5. Interpretations of soils clay fraction major oxides predicted maps. (a) Comparison between the predicted oxide maps in the 0-20 cm layer with the geological map; (b) Oxide content in soils as a function of parent material; (c) Comparison between the average spectral signature of soils with different oxide contents obtained punctually in laboratory data with the average spectral signature obtained in mapped regions with different oxide contents, (c.1) average spectrum for laboratory data as a function of Fe_2O_3, (c.2) average spectrum for predicted maps as a function of Fe_2O_3, (c.3) average spectrum for laboratory data as a function of Al_2O_3, (c.4) average spectrum for predicted maps as a function of Al_2O_3, (c.5) average spectrum for laboratory data as a function of SiO_2, (c.6) average spectrum for predicted maps as a function of SiO_2. Vr: Volcanic Rhyolitic; Vb: Volcanic Basaltic; Su: Sedimentary Undifferentiated; Ss: Sedimentary Siliciclastic; Sc: Sedimentary Carbonatic; Pgu: Plutonic Gabbroic and Ultramafic; Pgr: Plutonic Granitic; Mmhg: Metamorphic Medium to high grade; Mlmg: Metamorphic Low to medium grade.

2.3.4.1 Spectral behavior as a function of oxide content

Spectral signature (from image pixels compared with punctual controlling pixels with known wet laboratory information) of types of oxides indicated important variancy (Figure 5c) which gives support to the statistical approach. As we may see, the morphology trend is similar for all oxides of know positions to unknown ones (Figure 5c.1, c.2, c.3, c.4, c.5 and c.6) which bring an information of spectra indicating that the pixels were acquired with quality. We observe that the hypoferric soils and ferric soils have different trends due to their constitution and interaction with energy. The same happens with SiO_2. important to mention that this silica is not from quartz, since sulphuric acid digestion extracts Si from clay fraction oxides. This, it is coherent to see high intensity with low SiO_2. This SiO_2 are in the structure of minerals in clay fraction (i.e., hematite, goethite, gibbsite, kaolinite and montmorillonite).

2.3.4.2 Soils clay fraction major oxides predictions uncertainty

In general, the spatial uncertainty (Figure 6) ranged from 0 to 20%, suggesting a low variation around the mean predicted value. The uncertainty represents the variation of the set of predicted values for each pixel in relation to the mean. Thus, the low uncertainty values validate the oxides prediction process adopted in this study. Low uncertainty values were also observed by Safanelli et al. (2021a) with uncertainty values were between 0-25% in the prediction of clay and carbon stock in Brazil. When analyzing the uncertainty maps in a spatial way, it is possible to see that for all minerals, the regions where the lowest concentrations were observed, the highest values of uncertainty were also observed. This reveals that there are higher estimates in areas with low mineral content.

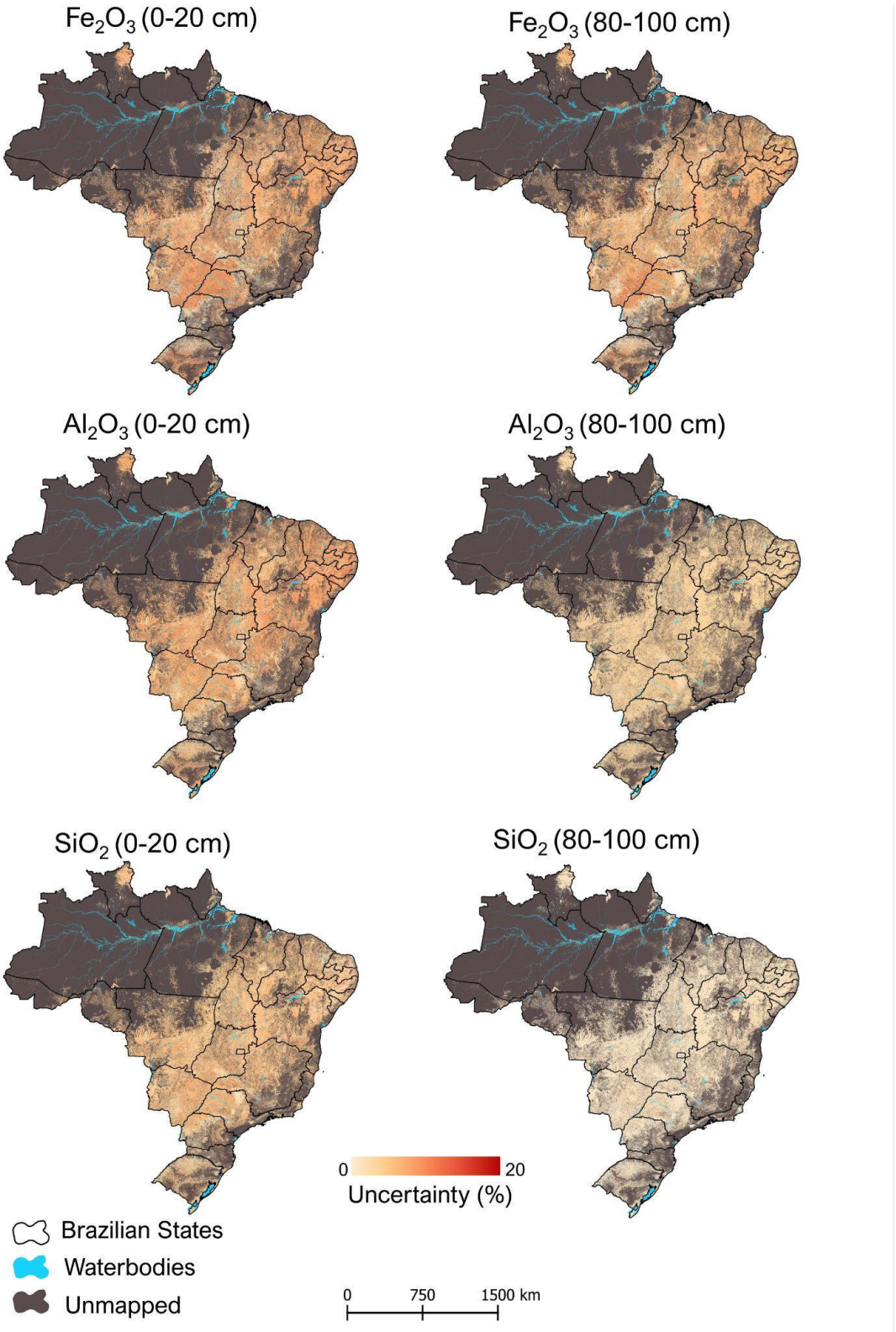

Figure 6. Uncertainty Predicted major soil oxides.

2.3.4.3 A discussion related with spatial resolution and geographical extension

The technique maintains the spatial resolution with 30 m (Figure 7). As we can observe, when we look at the São Paulo state (Figure 7a), it is clearly detected the spill of basalt from south to north. This spill cross the state with the basalt origin and generates a specific spectra at 3r2g1b and 5r4g3b compositions (Landsat 5), in agreement with Fongaro et al. (2018). The comparison of the spectra with the iron quantification are well correlated. Sometimes, these maps are well visualized because are with a low dimension. On the other hand, when we make a zoom in a regional scale, in an area with 500.000 ha (Figura 7b), we still see variations and with greater differences. Going towards to a farm scale, such as the one with 160 ha (Figure 7c), the differences are still there. These images indicate that the mechanism of SYSI is a powerful technique, that results in multiple geographical extensions from a country to a farm and maintain the same information.

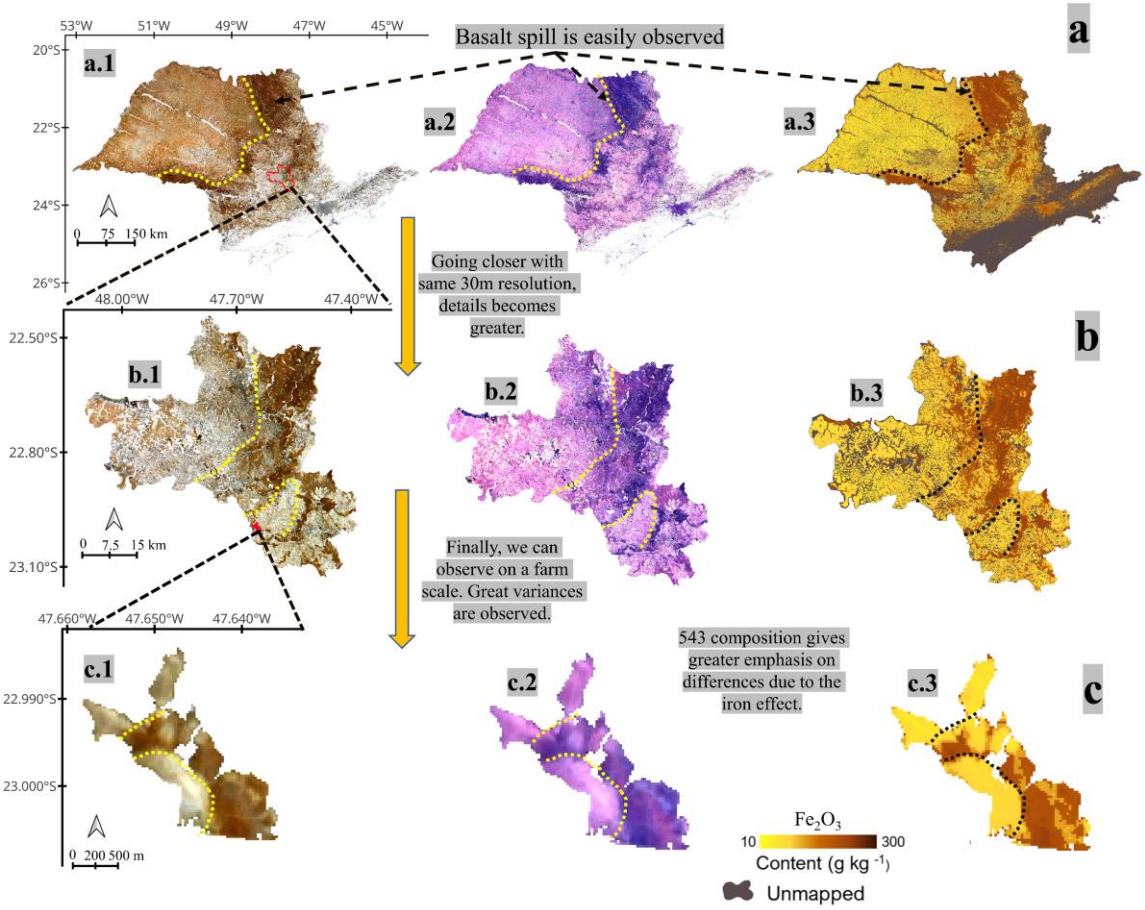

Figure 7: The figure represents the same region observed from different scales where (a) State scale, (b) Regional scale, and (c) farm scale. From left to right we go from a Landsat composition (SYSI) 321 and 543 RGB respectively, and finally a Fe_2O_3 map.

2.3.5 Soils clay fraction major oxides maps applications

2.3.5.1 Maps of Weathering Index (Ki)

The Ki maps calculated based on SiO2 and Al_2O_3 contents are presented in Figure 8. We obtained Ki maps for two soil layers (0-20 cm and 80-100 cm). The map of the surface layer (Figure 8a) revealed greater spatial variability in Ki values, ranging from 0.67 to 4.2, with values below 2 predominating throughout the map. In contrast, the subsurface layer map showed lower spatial variability, with values between 1.5 and 2.9 (Figure 6b), and a significant prevalence of areas containing Ki values above 2. These maps are consistent with the Ki values obtained through database calculations (laboratory analyses), as shown in Figure 8c. Ki is one of the most widely used indices in Brazil, as it indicates soil mineralogy and weathering. The Ki index is used in the Brazilian Soil Classification System to characterize and identify oxic horizons, which should have values below 2.2, indicating a high degree of weathering (Embrapa Solos, 2018). Low Ki values indicate processes of monosiallitization and alitization, leading to the formation of 1:1 clay minerals and aluminum oxides (Lima et al., 2022). On the other hand, higher Ki values (greater than 2.2) indicate the occurrence of bissiallitization processes with a higher presence of 2:1 clay minerals, indicating a lower degree of weathering (Guimarães et al., 2021).

Values of Ki predominantly ≤ 1.5 (in the 0-20 cm layer) are observed in the region represented in Figure 8a.1. This region comprises a portion of the Cerrado biome, where the predominant soils are Ferralsols. The Ferralsols of the Cerrado exhibit varying origins, ranging from sedimentary rocks such as the sandstones of the Paraná Basin to basaltic rocks of the Serra Geral Formation and Cenozoic Lateritic Cover (Oliveira et al., 2023). The occurrence of plateaus associated with the tropical climate has contributed to an intense degree of weathering of these soils, making them the most weathered in Brazil (Curi and Franzmeier, 1984; Oliveira et al., 2023). In general, their mineralogy exhibits a low presence of primary weatherable minerals and a higher presence of kaolinite and oxides, with the soils being more gibbsitic in the higher areas of the plateaus (Schaefer et al., 2008).

The Ki maps represent the spatial variations of weathering degree in soils formed under different circumstances, as can be observed in Figure 8a.4. The delimited area in the figure encompasses a region between the states of Bahia and Tocantins, as well as the northern part of Goiás. This region exhibits a strong contrast between highly weathered soil classes (Ferralsols) on one side, which are found in the extensive plateau of the Urucuia geological formation, and Cambisols and Regosols formed on the Bambuí group on the other side. Values of Ki ranging from 1.91 to 3.10 were observed by Maranhão et al, (2016) when studying profiles formed on the Bambuí group in this region, which is consistent with our results.

Higher values of Ki in the surface layer (predominantly ≥ 2) are observed in the regions encompassing the Pantanal and Caatinga biomes, where younger soils such as Neosols, Luvissols, Planosols, Solonetz, and Vertisols are predominant. A zoom in on these regions is presented in Figures 8a.2 and 8a.3 for a more detailed visualization. The Pantanal is a vast floodplain, and as such, hydromorphism is the main pedogenetic process in its soils, reducing weathering rates (Coringa et al., 2012). Additionally, the parent materials in the Pantanal region often consist of carbonate sediments or exhibit a significant presence of sodium (Andrade et al., 2020; Oliveira et al., 2021). The chemical

elements released during the weathering of these sediments that remain in the soil solution increase the pH, favoring the neoformation of 2:1 clay minerals (Couto et al., 2023). Contrary to the Pantanal, the Caatinga biome is one of the largest tropical dry areas in the world (Araujo et al., 2022). Furthermore, the Caatinga is the oldest semi-arid landscape in South America, and its semi-arid climate strongly influences the formation of its soils (Araújo Filho et al., 2023). The soils in this biome contain significant amounts of primary minerals and 2:1 clay minerals due to the low rates of weathering (Oliveira et al., 2019).

The map of the 80-100 cm layer showed that subsurface soils are less weathered and more similar to the parent material than the surface soils. It is possible to observe that lower Ki values are found in regions composed of rocks that undergo weathering more easily, such as basalt. In Figure 8b.1, we selected a region where soils have developed on different parent materials. Soils derived from quartzose rocks (Bauru Group Formation, Botucatu Formation, and Passa Dois Formation) exhibit Ki values above 2 in the subsurface layer. On the other hand, soils formed on the Serra Geral Formation (ultramafic basaltic rocks) display Ki values lower than 2, indicating greater weathering in the subsurface.

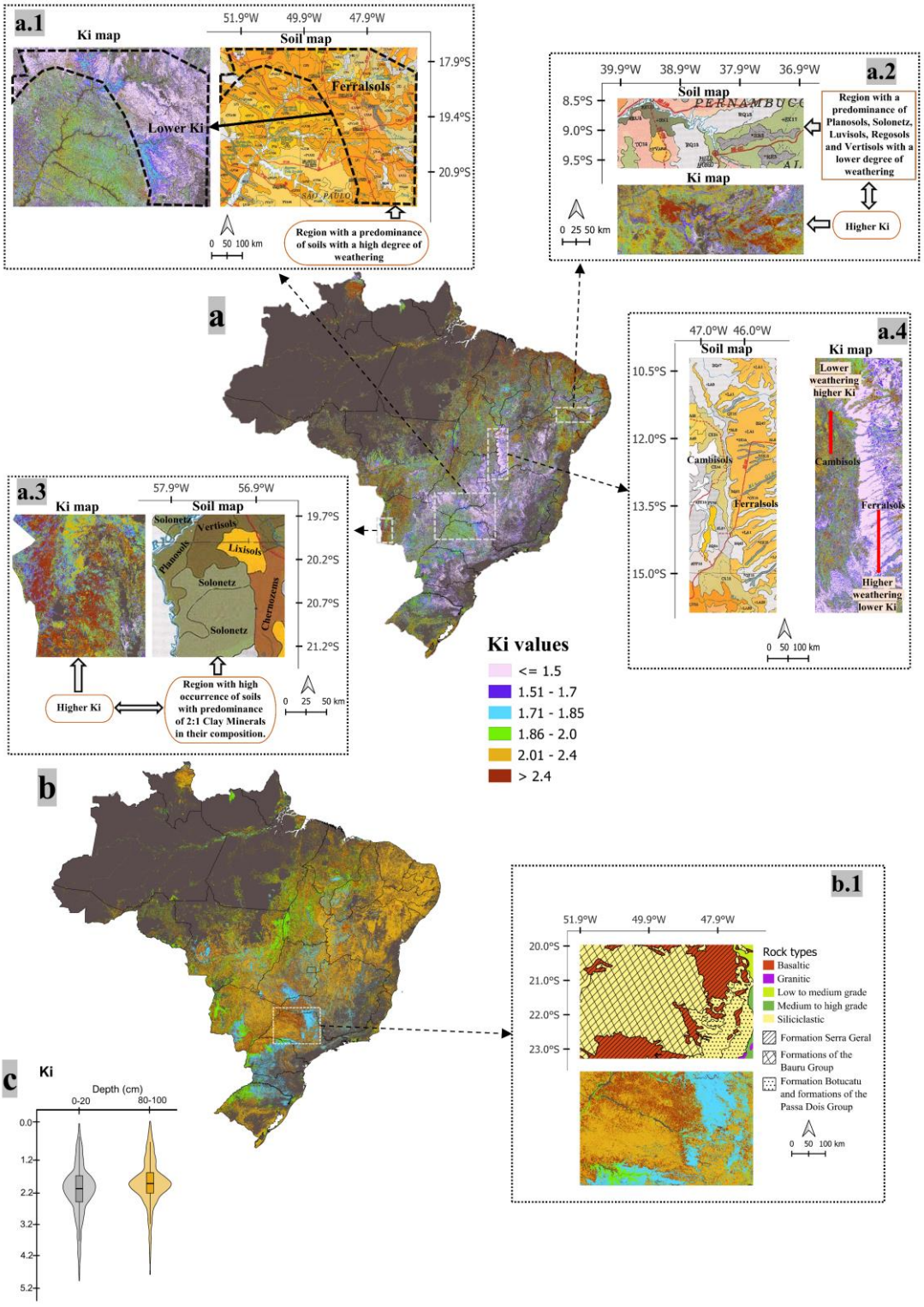

Figure 8. Maps of Ki index. (a) Ki map of the 0-20 cm layer; (a.1) Region representing the Cerrado biome with highly weathered soils; (a.2) Region representing the Caatinga biome with less weathered soils; (a.3) Region representing the Pantanal biome with less weathered hydromorphic soils; (a.4) Region exhibiting a contrast between highly weathered and less weathered soils; (b) Ki map of the 80-100 cm layer; (b.1) Region with different geological formations; (c) Descriptive statistics of the Ki index calculated from laboratory analyses in the two studied layers. The pedological map used for comparisons was the legacy soil map of Brazil published by IBGE, (2021); The geological map used for comparisons was published by Gómez et al. (2018).

2.3.5.2 Application in pedological maps

The study area comprises five soil classes: Cambisols, Chernozems, Nitisols, Lixisols, and Leptosols (Figure 9a), with a predominance of Lixisols. Unsupervised classification performed using the oxide maps created groups that represented the observed soil classes on the pedological map (Figure 9b). The best representations were found in the regions corresponding to Nitisols and Lixisols, represented by groups 1 and 2, respectively. The Nitisols in these regions develop on iron-rich ultrabasic basaltic rocks and easily weatherable primary minerals, thus exhibiting high oxide content (Silvero et al., 2021). In contrast to the other soils in the area that have developed on parent materials rich in quartz (Nanni and Demattê, 2006). This characteristic allowed for the distinction of this soil class from the others due to its high oxide content and pronounced degree of weathering. In contrast, Lixisols are soils with a moderate degree of weathering when compared to Nitisols, where the pedogenetic process of clay illuviation is predominant (Embrapa Solos, 2018). This process favors the formation of sandy surface layers and clayey subsurface layers. As a result, there are significant differences in oxide content between the layers (Quénard et al., 2011). This characteristic allowed for the discrimination of these soils into a single group.

In the regions encompassing the soil classes Cambisols, Chernozems, and Leptosols, there was not a clear discrimination between the groups. These soils share common characteristics such as low weathering degree, significant amounts of primary minerals and 2:1 clay minerals, and shallow depth (Embrapa Solos, 2018). We believe that due to these similarities, it was challenging to individualize these soils into distinct groups.

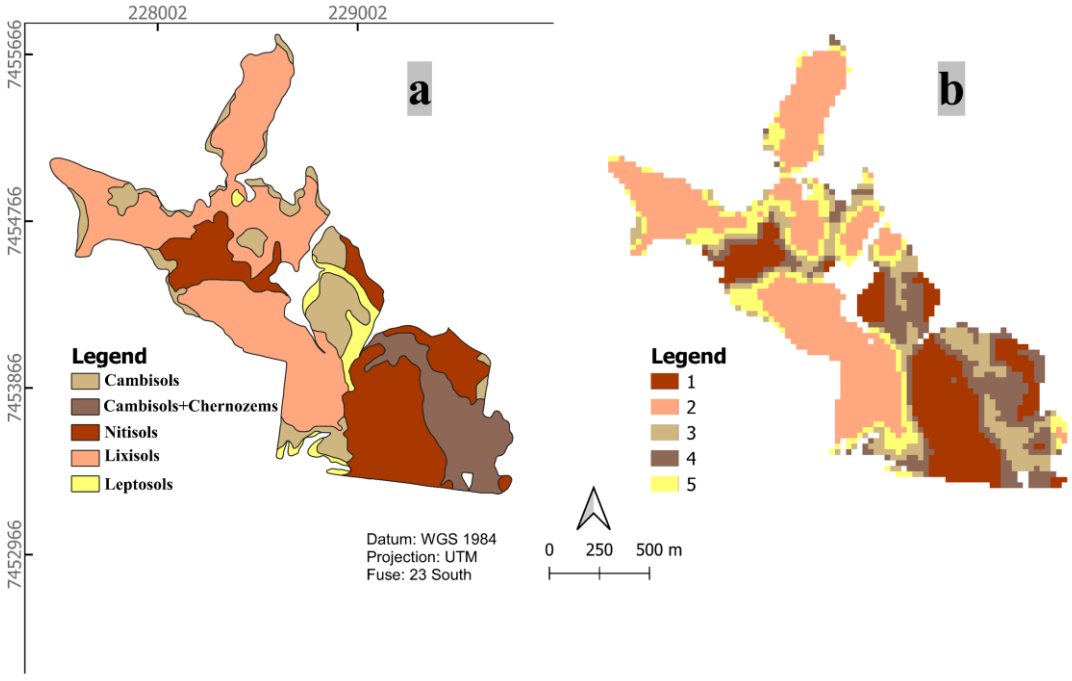

Figure 9. Comparisons between (a) a detailed pedological map, (b) Class map obtained by unsupervised classification of soil oxide and Ki maps

2.3.5.3 Application in agricultural management

The NDVI is widely used to infer, in general, the vegetation health (Lisboa et al., 2018; Piedallu et al., 2019). When we observe the NDVI of the sugarcane field, it is possible to detect two regions with opposite values (Figure 10a). The regions to the north and west presented higher NDVI values, indicating that the plants are healthier. While at south east the NDVI values are lower, indicating that plants may be suffering some type of stress. Unsupervised classification using the NDVI map (Figure 10b) showed that the area is divided into two distinct clusters. When we compare this map of groups with the maps of oxides (Figure 10c, d and e), it is possible to observe that cluster 1, where the healthiest plants are contained, has higher concentrations of oxides. According to Silva et al, (2022), soils containing higher levels of these oxides, mainly iron oxides, are generally soils originated from mafic or ultramafic rocks, which also provide higher levels of nutrients for plants due to their mineralogical composition. In addition, the oxides can form complexes with Soil organic matter (SOM) in the soil, increasing their concentrations as indicated by Hemingway et al, (Kleber et al., 2021). The increase in SOM content may favor the water storage capacity in the soil, increasing the availability of water for plants in dry periods (Wiesmeier et al., 2019).

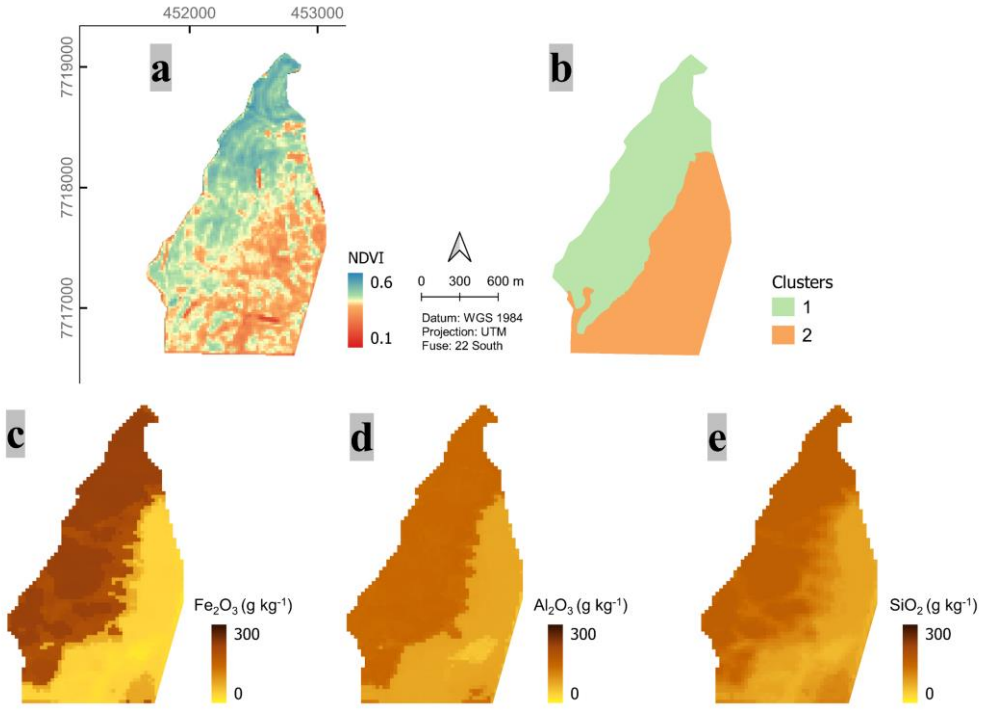

Figure 10. Relationship between the vigor of the sugarcane and the contents of oxides. (a) NDVI calculated with Sentinel-2 images at the final stage of sugarcane development; (b) unsupervised classification based on the NDVI, where cluster 1 represents plants with greater vigor and cluster 2 plants with lower vigor; (c) Fe_2O_3 map in the 0-20 cm layer; (d) Al_2O_3 map in the 0-20 cm layer; (e) SiO2 map in the 0-20 cm layer.

2.4 Conclusions

It was possible to predict soils clay fraction major oxides (Fe2O3, SiO2, Al2O3) with high spatial resolution (30m) using satellite-derived data and machine learning. Furthermore, the modeling of our maps was considered satisfactory, with R^2 and RMSE ranging from 0.19 to 0.65 and from 66.7 to 47 $g.kg^{-1}$, respectively. The predicted oxide maps covered approximately 3.48 million km² (~40% of the national territory), encompassing all areas that were used for agriculture during the period from 1982 to 2020.

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

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

Our maps showed great potential to assist in the construction of pedological maps, capable of distinguishing regions belonging to the same soil class. Additionally, we observed a relationship between the vigor of sugarcane crops and the oxide contents, where plants cultivated in soils with higher oxide contents exhibited greater vegetative vigor.

Furthermore, our findings provide valuable information for all agricultural regions in Brazil, with high spatial resolution. This information can aid researchers, farmers, and consultants in understanding the dynamics of their soils and making informed decisions.

Acknowledgments

We acknowledge FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) project 2014-22262-0, 2021-05129, CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) process number: 88887.481452/2020-00 and the Geotehcnologies on Soil Science (https://esalqgeocis.wixsite.com/geocis).

REFERENCES

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

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

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

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

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

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

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

Barthès, B.G., Kouakoua, E., Larré-Larrouy, M.C., Razafimbelo, T.M., de Luca, E.F., Azontonde, A., Neves, C.S.V.J., de Freitas, P.L., Feller, C.L., 2008. Texture and sesquioxide effects on water-stable aggregates and organic matter in some tropical soils. Geoderma 143, 14–25. https://doi.org/10.1016/J.GEODERMA.2007.10.003

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

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

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

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

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

Bronick, C.J., Lal, R., 2005. Soil structure and management: a review. Geoderma 124, 3–22. https://doi.org/10.1016/J.GEODERMA.2004.03.005

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

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

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

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

Curi, N., Franzmeier, D.P., 1984. Toposequence of Oxisols from the Central Plateau of Brazil. Soil Science Society of America Journal 48, 341–346. https://doi.org/10.2136/SSSAJ1984.03615995004800020024X

Cutler, A., Cutler, D.R., Stevens, J.R., 2012. Random Forests, in: Ensemble Machine Learning. Springer, New York, NY, New York, NY, pp. 157–175. https://doi.org/10.1007/978-1-4419-9326-7_5

da Silva, R.J.A.B., da Silva, Ygor Jacques Agra Bezerra, van Straaten, P., do Nascimento, C.W.A., Biondi, C.M., da Silva, Yuri Jacques Agra Bezerra, de Araújo Filho, J.C., 2022. Influence of parent material on soil chemical characteristics in a semi-arid tropical region of Northeast Brazil. Environ Monit Assess 194, 1–21. https://doi.org/10.1007/S10661-022-09914-9/TABLES/9

de Almeida, J.A., 2023. Soils of Pampa Gaúcho, the Mixed Prairies of Southern Brazil, in: The Soils of Brazil. Springer, Cham, pp. 299–342. https://doi.org/10.1007/978-3-031-19949-3_11

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

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

Demattê, J.A.M., Dotto, A.C., Paiva, A.F.S., Sato, M. V., Dalmolin, R.S.D., de Araújo, M. do S.B., da Silva, E.B., Nanni, M.R., ten Caten, A., Noronha, N.C., Lacerda, M.P.C., de Araújo Filho, J.C., Rizzo, R., Bellinaso, H., Francelino, M.R., Schaefer, C.E.G.R., Vicente, L.E., dos Santos, U.J., de Sá Barretto Sampaio, E. V., Menezes, R.S.C., de Souza, J.J.L.L., Abrahão, W.A.P., Coelho, R.M., Grego, C.R., Lani, J.L., Fernandes, A.R., Gonçalves, D.A.M., Silva, S.H.G., de Menezes, M.D., Curi, N., Couto, E.G., dos Anjos, L.H.C., Ceddia, M.B., Pinheiro, É.F.M., Grunwald, S., Vasques, G.M., Marques Júnior, J., da Silva, A.J., Barreto, M.C. de V., Nóbrega, G.N., da Silva, M.Z., de Souza,

S.F., Valladares, G.S., Viana, J.H.M., da Silva Terra, F., Horák-Terra, I., Fiorio, P.R., da Silva, R.C., Frade Júnior, E.F., Lima, R.H.C., Alba, J.M.F., de Souza Junior, V.S., Brefin, M.D.L.M.S., Ruivo, M.D.L.P., Ferreira, T.O., Brait, M.A., Caetano, N.R., Bringhenti, I., de Sousa Mendes, W., Safanelli, J.L., Guimarães, C.C.B., Poppiel, R.R., e Souza, A.B., Quesada, C.A., do Couto, H.T.Z., 2019. The Brazilian Soil Spectral Library (BSSL): A general view, application and challenges. Geoderma 354, 113793. https://doi.org/10.1016/J.GEODERMA.2019.05.043

Demattê, J.A.M., Fioriob, P.R., Ben-Dorc, E., 2009. Estimation of Soil Properties by Orbital and Laboratory Reflectance Means and its Relation with Soil Classification. The Open Remote Sensing Journal 2, 12–23. https://doi.org/10.2174/1875413901002010 12

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

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

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

Dharumarajan, S., Hegde, R., 2022. Digital mapping of soil texture classes using Random Forest classification algorithm. Soil Use Manag 38, 135–149. https://doi.org/10.1111/SUM.12668

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

Embrapa Solos, 2018. Sistema brasileiro de classificação de solos, Embrapa Solos. Rio de Janeiro.

Famelli, N., Lima, E.F., Carmo, I. de O., 2021. Lithostratigraphy of the Serra Geral Formation in the northern portion of the Paraná-Etendeka Igneous Province: A tool for tracking Early Cretaceous paleoenvironmental changes. Journal of Volcanology and Geothermal Research 410, 107152. https://doi.org/10.1016/J.JVOLGEORES.2020.107152

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

Ferreira, T.O., Otero, X.L., Nóbrega, G.N., Queiroz, H.M., Barcellos, D., Vidal-Torrado, P., 2023. Mangroves Along the Brazilian Coast, in: The Soils of Brazil. Springer, Cham, pp. 411–421. https://doi.org/10.1007/978-3-031-19949-3_15

Fongaro, C.T., Demattê, J.A.M., Rizzo, R., Safanelli, J.L., Mendes, W. de S., Dotto, A.C., Vicente, L.E., Franceschini, M.H.D., Ustin, S.L., 2018. Improvement of clay and sand quantification based on a

novel approach with a focus on multispectral satellite images. Remote Sens (Basel) 10. https://doi.org/10.3390/rs10101555

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

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

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

Gómez, J., Schobbenhaus, C., Montes, N.E., 2018. Geological Map Of South America At a Scale of 1: 5M. Commisslon for the Geological Map of the World (CGMW), Colombian Geological Survey and Geological Survey of Brazil. https://doi.org/10.32685/10.143.2019.929

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

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

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

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

Heung, B., Bulmer, C.E., Schmidt, M.G., 2014. Predictive soil parent material mapping at a regional-scale: A Random Forest approach. Geoderma 214–215, 141–154. https://doi.org/10.1016/J.GEODERMA.2013.09.016

Heung, B., Ho, H.C., Zhang, J., Knudby, A., Bulmer, C.E., Schmidt, M.G., 2016. An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. Geoderma 265, 62–77. https://doi.org/10.1016/J.GEODERMA.2015.11.014

Heuze, F.E., 1983. High-temperature mechanical, physical and Thermal properties of granitic rocks— A review. International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts 20, 3–10. https://doi.org/10.1016/0148-9062(83)91609-1

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

Huang, S., Tang, L., Hupy, J.P., Wang, Y., Shao, G., 2021. A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. J For Res (Harbin) 32, 1–6. https://doi.org/10.1007/S11676-020-01155-1/FIGURES/2

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

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

Kleber, M., Bourg, I.C., Coward, E.K., Hansel, C.M., Myneni, S.C.B., Nunan, N., 2021. Dynamic interactions at the mineral–organic matter interface. Nature Reviews Earth & Environment 2021 2:6 2, 402–421. https://doi.org/10.1038/s43017-021-00162-y

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

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

Lisboa, I.P., Damian, M., Cherubin, M.R., Barros, P.P.S., Fiorio, P.R., Cerri, C.C., Cerri, C.E.P., 2018. Prediction of Sugarcane Yield Based on NDVI and Concentration of Leaf-Tissue Nutrients in Fields Managed with Straw Removal. Agronomy 2018, Vol. 8, Page 196 8, 196. https://doi.org/10.3390/AGRONOMY8090196

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

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

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

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

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

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

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. https://doi.org/10.1016/j.geoderma.2019.01.025

Nanni, M.R., Demattê, J.A.M., 2006. Spectral Reflectance Methodology in Comparison to Traditional Soil Analysis. Soil Science Society of America Journal 70, 393–407. https://doi.org/10.2136/SSSAJ2003.0285

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

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

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

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

Piedallu, C., Chéret, V., Denux, J.P., Perez, V., Azcona, J.S., Seynave, I., Gégout, J.C., 2019. Soil and climate differently impact NDVI patterns according to the season and the stand type. Science of The Total Environment 651, 2874–2885. https://doi.org/10.1016/J.SCITOTENV.2018.10.052

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

Poppiel, R.R., Lacerda, M.P.C., Safanelli, J.L., Rizzo, R., Oliveira, M.P., Novais, J.J., Demattê, J.A.M., 2019. Mapping at 30 m Resolution of Soil Attributes at Multiple Depths in Midwest Brazil. Remote Sensing 2019, Vol. 11, Page 2905 11, 2905. https://doi.org/10.3390/RS11242905

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

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

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

Rosin, N.A., Demattê, J.A.M., Poppiel, R.R., Silvero, N.E.Q., Rodriguez-Albarracin, H.S., Rosas, J.T.F., Greschuk, L.T., Bellinaso, H., Minasny, B., Gomez, C., Marques Júnior, J., Fernandes, K., 2023a.

Mapping Brazilian soil mineralogy using proximal and remote sensing data. Geoderma 432, 116413. https://doi.org/10.1016/J.GEODERMA.2023.116413

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

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

Safanelli, J.L., Demattê, J.A.M., Chabrillat, S., Poppiel, R.R., Rizzo, R., Dotto, A.C., Silvero, N.E.Q., 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. https://doi.org/10.1016/J.GEODERMA.2021.115042

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

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

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

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

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

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

Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., Homayouni, S., 2020. Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review. IEEE J Sel Top Appl Earth Obs Remote Sens 13, 6308–6325. https://doi.org/10.1109/JSTARS.2020.3026724

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

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

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

Silva, S.H.G., Ribeiro, B.T., Guerra, M.B.B., de Carvalho, H.W.P., Lopes, G., Carvalho, G.S., Guilherme, L.R.G., Resende, M., Mancini, M., Curi, N., Rafael, R.B.A., Cardelli, V., Cocco, S., Corti, G., Chakraborty, S., Li, B., Weindorf, D.C., 2021. pXRF in tropical soils: Methodology, applications, achievements and challenges. Advances in Agronomy 167, 1–62. https://doi.org/10.1016/BS.AGRON.2020.12.001

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

Silvero, Nélida Elizabet Quiñonez, 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. https://doi.org/10.1016/J.RSE.2020.112117

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

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

Tayebi, M., Rosas, J.T.F., Mendes, W. de S., Poppiel, R.R., Ostovari, Y., Ruiz, L.F.C., Dos Santos, N.V., Cerri, C.E.P., Silva, S.H.G., Curi, N., Silvero, N.E.Q., Demattê, J.A.M., 2021. Drivers of Organic Carbon Stocks in Different LULC History and along Soil Depth for a 30 Years Image Time Series. Remote Sensing 2021, Vol. 13, Page 2223 13, 2223. https://doi.org/10.3390/RS13112223

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

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

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

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

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

Vaysse, K., Lagacherie, P., 2015. Evaluating Digital Soil Mapping approaches for mapping GlobalSoilMap soil properties from legacy data in Languedoc-Roussillon (France). Geoderma Regional 4, 20–30. https://doi.org/10.1016/J.GEODRS.2014.11.003

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

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

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

Wiesmeier, M., Urbanski, L., Hobley, E., Lang, B., von Lützow, M., Marin-Spiotta, E., van Wesemael, B., Rabot, E., Ließ, M., Garcia-Franco, N., Wollschläger, U., Vogel, H.J., Kögel-Knabner, I., 2019. Soil organic carbon storage as a key function of soils - A review of drivers and indicators at various scales. Geoderma 333, 149–162. https://doi.org/10.1016/J.GEODERMA.2018.07.026

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

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

APPENDIX

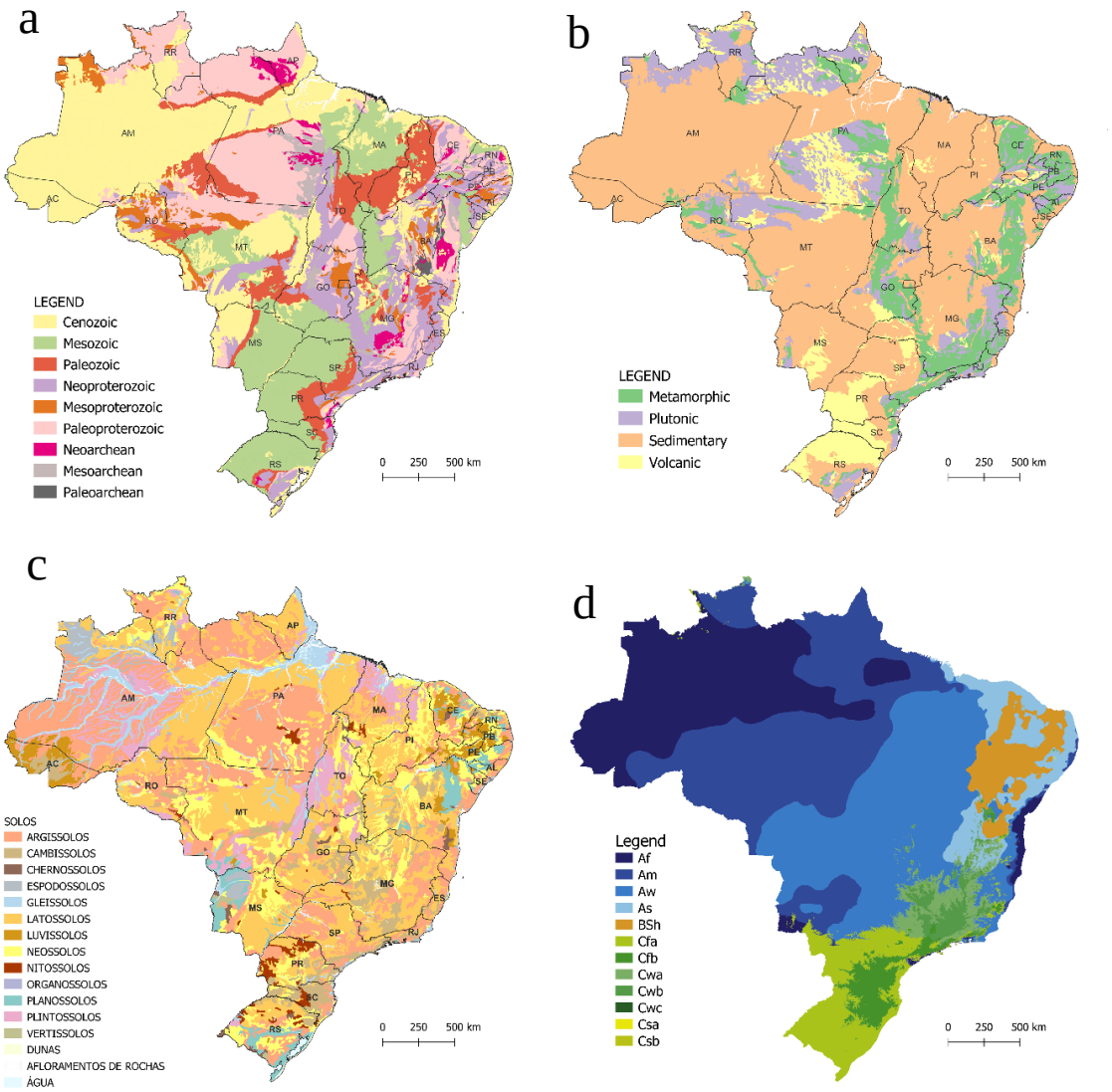

Figure A1. Geological map of Brazil (a) and map of parent materials of Brazil (b) (adapted from Gómes et al., 2019); pedological map of Brazil (Instituto Brasileiro de Geografia e estatística 2021) (c); map of Kopen's climatic classes of Brazil (Alvarez et al., 2013) (d).

3 GEOTECHNOLOGIES ON THE PHOSPHORUS STOCKS DETERMINATION IN TROPICAL SOILS: GENERAL IMPACTS ON SOCIETY

ABSTRACT

Phosphorus (P) is a critical nutrient for primary production in both terrestrial and aquatic ecosystems, and its use in fertilizers is essential for the success of global agriculture. However, given the non-renewability of P mineral reserves, the discussion about its sustainable use has been growing, with the aim of ensuring food security for future generations. In this study, our primary objective was to map the stocks of available P (AP) and total P (TP) in the soil throughout the territory of Brazil at a 30 m scale. Furthermore, we hope that the maps generated here can be used to achieve a sustainable use of P in agriculture and the environment systems. We used the Random Forest machine learning algorithm, which was fed with a legacy database containing 28,572 samples for AP and 3,154 for TP distributed throughout the Brazilian territory (0-20 cm). The P stocks were predicted from environmental covariates related to soil formation processes and soil properties correlated with P. We divided Brazil into two sub-regions, representing areas with native coverage and areas with anthropogenic coverage. From this, we built independent predictive models for each sub-region. Our results showed that Brazil has a TP stock of 531 Mt and an AP stock of 17.4 Mt. The highest averages of TP stocks are in the Atlantic Forest biome (73.8 $g.m^2$), which may be linked to the higher stocks of organic carbon in the soil of this biome. The highest averages AP stocks are in the Caatinga biome (2.51 $g.m^{-2}$) since this has younger soils with a low capacity to adsorb phosphorus. We also found that the use of fertilizers significantly increased AP stocks, where agricultural areas always had higher AP stocks than native areas. Furthermore, results indicated that Brazil's agricultural production is heavily influenced by AP stocks in the soil. We found a correlation ranging from 0.20 for coffee crops to 0.46 for soybean, between crop productivity and AP stocks in Brazilian soils.

Keywords: Digital Soil mapping, Sustainability, Remote Sensing, Machine Learning, soil health.

3.1 Introduction

Phosphorus (P), an emblematic chemical element, is one of the main nutrients that limit primary production in terrestrial and aquatic ecosystems (Schipanski and Bennett, 2021). Its use in agriculture, in the form of chemical fertilizers, has been the main driver of the significant increase in productivity in tropical soils (Withers et al., 2018). Although phosphorus is abundant in the soil, both in organic and inorganic forms, it is a limiting element for crop development and growth due to its low availability to plants. Plants absorb phosphorus through the root system in the form of orthophosphate, but only about 0.1% of the phosphorus is readily available to them (Schipanski and Bennett, 2021; Zhu et al., 2011).

The majority of P used in agriculture comes from non-renewable mineral reserves, which has sparked debates about its management, as the reserves are at risk of depletion, creating food insecurity for future generations (Elser, 2012; Rosemarin and Ekane, 2016). For decades, phosphorus has been used in agriculture without much control and often inefficiently (Roy et al., 2016; Withers et al., 2015). In the soil, phosphorus applied in the form of chemical fertilizers in excessive doses can become unavailable to plants due to its fixation or be lost to aquatic environments, causing eutrophication (Schipanski and Bennett, 2021). Therefore, the efficient use of phosphorus is indispensable for us to achieve sustainability on the planet (Steffen et al., 2015). The search for the efficient and sustainable use of P involves strategies at the local level (on the farm), such as the correction of soil pH, the application at a varied rate, and the increase of organic matter in the soil to neutralize charges capable of adsorbing P (Schröder et al., 2011; Simpson et al., 2011). In addition, it covers strategies at the global level, mainly highlighting the mapping of P stocks and the factors that influence the availability of this element to plants (Barbieri et al., 2021; Roy et al., 2016).

The P stocks in the soil are the result of a complex interaction of climatic, biotic, and landscape factors over time (Menge et al., 2023). Soil formation factors influence the P stocks in the soil as well as the proportions in which these stocks are available or not for plants (Hou et al., 2018; Smeck, 1985). Soils formed from ultramafic rocks have higher phosphorus stocks than those formed from chemically poor parent material (Daly et al., 2015; Mage and Porder, 2013). However, the degree of weathering affects the proportion of available phosphorus to total phosphorus. Highly weathered soils have most of their stocks adsorbed on metallic oxides, while younger soils have larger quantities of phosphorus in its available form (Antoniadis et al., 2016). Climate formation factor also plays a crucial role in the phosphorus stocks of soils around the planet. Mild climates tend to preserve carbon in the soil and with it the phosphorus contained in the soil organic matter (SOM) (Hou et al., 2018). In addition to formation factors, due to the use of phosphate fertilizers in agriculture, phosphorus stocks also tend to increase in agricultural areas, as residual phosphorus adds to the soil's natural stocks (Pavinato et al., 2020).

In Brazil, where oxidic soils with a high degree of weathering predominate (Schaefer et al., 2008), the understanding of the dynamics of the phosphorus element has always had a generalist tendency (Withers et al., 2018). Where there is practically total adsorption of phosphorus on oxyhydroxides from clay minerals. Such reasoning disregards all the geospatial variation of Brazilian soils, which are highly heterogeneous and have different capacities to adsorb phosphorus or not. In a mapping of the potential for soil phosphorus retention, Barbosa et al, (2022) demonstrated that about 70% of Brazilian soils have low

to medium potential to adsorb phosphorus. These characteristics lead us to review the understanding of the spatial behavior of the phosphorus element throughout Brazilian territory. Once the phosphorus made available in the weathering process, in the mineralization of soil organic matter (SOM), and in the chemical fertilization of crops may or may not be adsorbed in the soil, directly influencing the phosphorus stocks in the soil.

Phosphorus stocks in Brazilian soils have never been quantified at a national level with detailed scale. The pioneering study conducted by Pavinato et al., (2020) mapped the content of native total phosphorus in Brazilian soils. However, this map presents a low level of detail, following the resolution of the soil map of Brazil, which is 1:5000000, preventing farm-level planning. In addition, mapping only total phosphorus does not provide a clear idea about the phosphorus available to plants, as there is no direct relationship between total phosphorus and available phosphorus (Nwoke et al., 2003; Smeck, 1985).

In a country where its dimensions are continental like Brazil, mapping phosphorus stocks is not an easy task. Brazil has about 8.5 million square kilometers, and a large part of the Brazilian territory is under native vegetation (about 64%), mainly in the northern region in the Amazon biome (MapBiomas, 2023). This contributes to a scarcity of data, making it difficult to map with a high level of detail and low uncertainties. Despite this, thanks to recent advances in digital soil mapping (DSM), these problems are being overcome. Several studies demonstrate the potential of Digital Soil Mapping (DSM) to generate assertive maps of different soil properties, e.g soil mineralogy (Rosin et al., 2023), soil texture (Dharumarajan and Hegde, 2022; Žížala et al., 2022), soil organic carbon stock (Gomes et al., 2019; Safanelli et al., 2021; Tayebi et al., 2021) and even total phosphorus stocks (Viscarra Rossel and Bui, 2016).

The use of predictive models based on machine learning (ML) algorithms, along with environmental covariates with increasing levels of detail and quality, has significantly contributed to the development of more accurate maps (Khaledian and Miller, 2020; Samuel-Rosa et al., 2015; Wadoux et al., 2020). Among the ML algorithms, Random Forest (RF) proposed by Breiman, (2001) has stood out in soil science studies. According to Padarian et al. (2020), this algorithm is the most used in this theme, having become popular due to its robustness provided by its architecture composed of numerous uncorrelated trees. Furthermore, Random Forest is easy to implement, allowing for the construction of a grid search to determine the best hyperparameters to be used, which often results in superior performance compared to other machine learning algorithms (Sheykhmousa et al., 2020; Wadoux et al., 2020). Additionally, as soil properties can be of a quantitative or qualitative nature, the structure of Random Forest, based on decision trees, can handle both. This is because the algorithm works with both linear regression and classification (Sheykhmousa et al., 2020).

In DSM, the choice of environmental covariates is extremely important for obtaining assertive maps (Lamichhane et al., 2019). The SCORPAN model proposed by McBratney et al. (2003) is popularly used in DSM as a strategy in determining environmental covariates. Many covariates can be used to represent the factors described in the SCORPAN model. For example, multispectral images, obtained by satellites, can describe the spectral behavior of vegetation, representing the soil formation factor organism (Chen et al., 2022). Digital elevation models are used to represent the relief (Safanelli et al., 2020), while

historical data, obtained by weather stations, such as rainfall and temperature, can be spatialized and used to represent the climate (Shary, 2023). In addition, data mining algorithms, such as GEOS3 developed by Demattê et al., (2018), can be used to manipulate historical data series and extract point information highly correlated with soil properties. GEOS3 is fed with the entire historical series of Landsat images and outputs a synthetic soil image (SYSI) composed of pixels that represent the soil reflectance at a time when it was uncovered. The SYSI represents a direct measure of the soil and has a high correlation with various soil properties (Demattê et al., 2020), being used as a proxy for DSM in several studies (Gómez et al., 2023; Mello et al., 2023; Rosin et al., 2023; Safanelli et al., 2021). Furthermore, due to the correlation between various soil properties, they can also be used as environmental covariates (McBratney et al., 2003). Soil properties such as clay, iron oxides, aluminum oxides, and SOC are generally used to map soil phosphorus (Pavinato et al., 2020; Viscarra Rossel and Bui, 2016).

We expect that combination of advanced MDS techniques, including machine learning and environmental covariates with a high level of detail, will enable us to quantify and map available and total phosphorus stocks in the surface layer of soil throughout the Brazilian territory at a spatial resolution of 30 meters. Additionally, we expect that these maps can be utilized to support future strategies for the sustainable management of phosphorus in both agriculture and the environment.

3.2 Material and methods

3.2.1 Study area

The study covered all of Brazil, a country of continental dimensions with six main biomes. The Amazon and the Atlantic Forest are characterized by vast rainforests, while the Cerrado and the Caatinga feature extensive savannas. The Pantanal is one of the largest floodable areas in the world, and the Pampa has steppe-type vegetation similar to that of temperate climate regions. Brazilian agriculture is diverse, producing commodities such as soybeans, coffee, cotton, corn, and sugarcane, in addition to annual crops like beans and vegetables, perennial crops like tropical fruits, and planted forests. There are also large areas dedicated to livestock, about 19% of the entire national territory (MapBiomas, 2023).

Brazil exhibits a variety of climates according to the Köppen climate classifications, including equatorial (Af, Am, Aw, As), tropical (Aw, BSh), and subtropical (Cwa, Cwb, Cwc, Cfa, Cfb), influenced by its location predominantly in the Tropical Zone (92% above the Tropic of Capricorn) and the extensive coastline that contributes to the prevailing humidity (Alvares et al., 2013). The climates that stand out are the equatorial, characterized by high rainfall, and the subtropical, with moderate temperatures and well-distributed rains.

The geology of Brazil, the result of complex geological processes over billions of years, encompasses crystalline shields, sedimentary basins, and volcanic terrains, playing a fundamental role in the formation of soils and in determining their fertility(de Alkmim, 2015). The crystalline shields, representing 36% of the country, include the Guiana Shield, Central Brazil Shield, and Atlantic Shield, with enormous heterogeneity in the soils formed on plutonic rocks, which underwent metamorphism, physical damage, and were exposed to large climatic variations over the geological ages (Schaefer et al.,

2023).The sedimentary basins, covering about 60% of the territory, originated infertile soils due to the chemically poor sandstones that form these basins, such as those of the Paraná River basin and the Amazon basin, for example (Schaefer et al., 2023). The second largest volcanic spill in the world, which occurred at the beginning of the Cretaceous, originated magmatic rocks such as diabase and basalt in the Serra Geral geological formation (Alkmim, 2015). In these regions, soils with significant fertility originated due to the richness in nutrients of these rocks (Schaefer et al., 2023). The diversity of Brazilian soils, shaped by geological factors such as climate and relief, results in a wide variety of pedological characteristics throughout the country. In total, Brazilian soils are classified into 13 classes at the first categorical level according to the Brazilian soil classification system (Embrapa Solos, 2018).

All this diversity found throughout Brazil directly affects the P cycle. Thus, the P stocks in the soil are a consequence of this diversity, varying significantly in space.

3.2.2 Soil observations

In this study, we used a legacy database composed of data obtained from two different sources: i) Publicly available national soil profile databases (Samuel-Rosa et al., 2019), ii) Brazilian Soil Spectral Library (Demattê et al., 2019). These data were filtered according to two criteria: accuracy of geographic coordinates (error less than 100 meters) and the presence of available or total phosphorus (P) data. After the filtering process, we applied a vertical harmonization to the profile using the spline method proposed by Bishop et al. (1999), in order to standardize our samples to the 0-20 cm layer. We obtained 28,572 samples containing available phosphorus (AP) content and 3,154 samples containing total phosphorus (TP) content in the 0-20 cm layer. The P samples were distributed throughout the study area. AP was extracted using the resin method initially proposed by Amer et al. (1955) and modified for laboratory use in Brazil by Van Raij et al. (1986). The resin extraction method involves the use of both an anion-exchange and cation-exchange resin saturated with sodium bicarbonate, which adsorbs phosphorus from the soil in an aqueous suspension. The TP was determined through sulfuric acid digestion, as proposed by Vettori (1969), with modifications as proposed by Van Raij e Valadares (1974). This method accesses P present in the clay fraction of the soil, organic P, and is also capable of accessing P present in the primary mineral apatite.

The contents of AP and TP were transformed into P stocks using equation 1.

$$Pstocks = [P\ content\ X\ Ds\ X\ D]\ X\ 10 \quad (1)$$

Where, P stocks are reported in $g.m^{-2}$, P content ($g.kg^{-1}$), soil bulk density (Ds) ($g.cm^{-3}$), and soil thickness (D) (cm).

Soil bulk density (BD) was estimated using a pedotransfer function calibrated with the clay and soil organic carbon (SOC) contents for Brazilian soils and reached an R^2 value of 0.63 and standard error of 0.11 g cm^{-3} (Benites et al., 2007). This function is widely used in soil science studies in Brazil (Bonfatti et al., 2016; Gomes et al., 2019; Safanelli et al., 2021; Tayebi et al., 2021). All soil samples containing phosphorus content also had measurements of soil organic carbon (SOC) and clay content, which were used for bulk density (BD) calculations. The determinations of clay and SOC content followed the Brazilian soil analysis standards (Teixeira et al., 2017).

3.2.3 Strategies for mapping the entire national territory

Brazil is a continental country, where a significant portion of its vegetation cover remains native. This directly impacts data acquisition, rendering data scarcer in these areas compared to regions where vegetation cover has already been altered by human activities. In our study, we independently mapped these areas. The primary objective of this strategy was to generate a more accurate map where data availability was higher. To achieve our goal, we divided Brazil into two sub-regions based on vegetation cover, namely Anthropogenic Vegetation Cover (AVC) and Native Vegetation Cover (NVC). This segmentation was carried out using the entire historical series of Landsat images (from 1982 to 2022) through the GEOS3 algorithm (Dematt$\hat{e}$ et al., 2018). Regions classified as AVC were those where vegetation had been removed, leaving the soil exposed at least once during this period. Locations where this criterion was not met were considered as NVC regions in our study. As a result, the database containing soil observations was also segmented. The number of observations in each sub-region can be consulted in Table 1. In Figure 1, we present a workflow to illustrate the strategy adopted.

Table 1. Number of soil observations for vegetation cover

	AVC	**NVC**	**TOTAL**
TP	1,673	1,481	3,154
AP	22,025	6,547	28,572

TP: total phosphorus; AP: available phosphorus; AVC: Anthropogenic Vegetation Cover; NVC: Native Vegetation Cover.

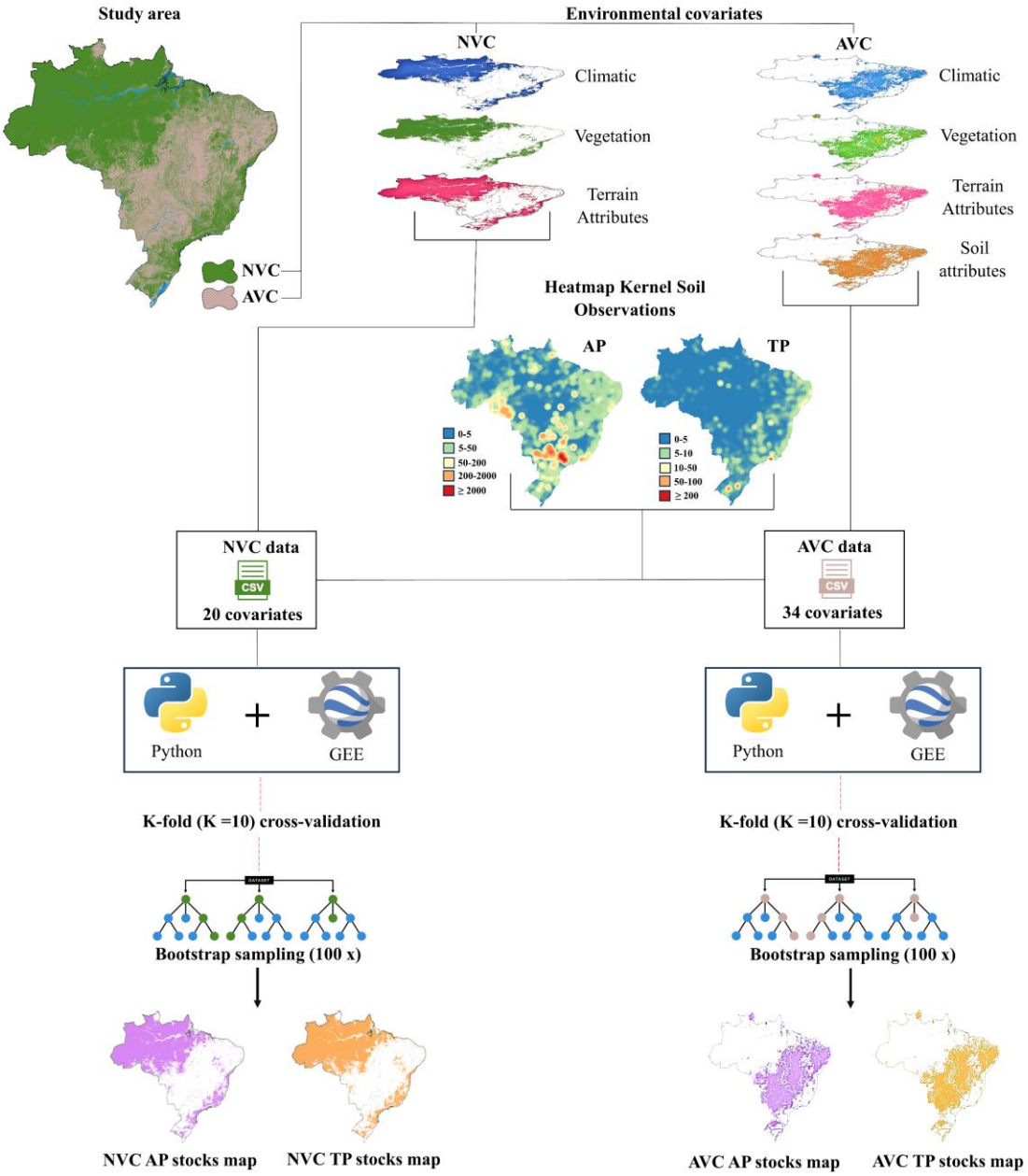

Figure 1. Workflow for mapping phosphorus stocks in Brazilian soils. Total phosphorus (TP); available phosphorus (AP); anthropogenic vegetation cover (AVC) and native vegetation cover (NVC)

3.2.4 Environmental covariates

Our primary concern in selecting environmental covariates was their ability to represent processes governing the phosphorus cycle in the Earth's crust. In general, we considered climatic variables, variables describing vegetation, and variables describing soil. In total, 34 covariates were used for modeling TP and AP. However, in areas of Low Data Availability (NVC), not all covariates were accessible. In Table 2, we specify which variables were employed for each mapped sub-region.

3.2.4.1 Climatic environmental covariates

The variables we employed to represent the climate were Annual Mean Temperature (ANMT) and Annual Precipitation (ANP) along with Land Surface Temperature (LST). We retrieved annual mean temperature and precipitation data from the WorldClim BIO V1 dataset (Hijmans et al., 2005). These data correspond to a specialized global climate time series at a spatial resolution of 1 km. For our study, we resampled the covariates to a spatial resolution of 30 m. The Land Surface Temperature (LST) was calculated from the thermal band (band 10) of Landsat 8 imagery. In the Google Earth Engine (GEE) environment, we created a cloud mask and selected all images with less than 20% cloud cover. After image selection, we computed the temperature using the methods proposed by Ermida et al. (2020). We aggregated all LST images obtained through median calculation, resulting in an image representing the median LST for the period from 2013 to 2022 across the entire study area.

3.2.4.2 Vegetation environmental covariates

To represent vegetation, we created a synthetic image that depicted the mean Vegetation Reflectance (mVR). We utilized the complete Landsat 8 collection from 2013 to 2022. The process began by filtering out all images with less than 20% cloud cover. Subsequently, we generated a mask to select only vegetation and exclude water surfaces and exposed soil. To create the mask, we employed the Normalized Difference Vegetation Index (NDVI). We established a threshold of 0.5, and all pixels with values exceeding this threshold were aggregated using the median. At the conclusion of this process, we obtained a synthetic image covering the entire study area, containing spectral information about the vegetation. The mVR was represented by the six spectral bands of Landsat 8.

3.2.4.3 Soil environmental covariates

The selected covariates to represent the soil included topography, Synthetic Soil Image (SYSI), and soil attributes directly correlated with soil phosphorus content. Topography was represented by terrain attributes that were previously used to digital soil map with satisfactory results in Safanelli et al (2021). These terrain attributes were as follows: slope, elevation, north and east slope, horizontal curvature, vertical curvature, mean curvature, maximal curvature, aspect, hillshade and a relief shape index. These variables were based on a digital elevation model (DEM) from the Shuttle Radar Topography Mission (SRTM) with 30 m (Farr and Kobrick, 2000). To calculate the environmental variables, the Terrain Analysis in Google Earth Engine (TAGEE) algorithm proposed by (Safanelli et al., 2020) was used. The individual information on the environmental covariates derived from the DEM are shown in table 2.

We utilized the GEOS3 algorithm (Demattê et al., 2020, 2018) to generate the Synthetic Soil Image (SYSI). The SYSI was created using the time series of images from the entire Landsat collection (USGS, 2021a, 2021b) covering the period from 1982 to 2022. This image represents the mean reflectance of bare soil and is composed of six spectral bands (Blue - band 1, Green - band 2, Red - band 3, NIR - band 4, SWIR1 - band 5, and SWIR2 - band 6). We employed a set of rules to identify bare soil

pixels in the satellite images, based on spectral indices and quality assessment bands, to remove clouds, cloud shadows, freshwater, photosynthetic vegetation, and non-photosynthetic vegetation (crop residues). A pixel was considered as bare soil when it exhibited Normalized Difference Vegetation Index (NDVI) values between -0.05 and 0.30 (excluding green vegetation) and Normalized Burn Ratio 2 (NBR2) values between -0.15 and 0.15 (excluding crop residues). Pixels identified as bare soil were selected to compose the multi-temporal collection. Each pixel in the SYSI represents the median value calculated among all pixels in the multi-temporal collection for a given location.

In tropical soils, phosphorus (P) content depends significantly on its interaction with soil properties. We selected attributes directly related to soil P content as environmental covariates, which included clay, sand, silt, SiO2, Al2O3, Fe2O3, Soil Organic Carbon (SOC), and soil pH. For this study, we utilized 30-meter spatial resolution maps that were made available and specialized using the methods proposed by Safanelli et al., (2021), and Rosas et al., (2023). These attribute maps covered only the AVC areas due to the proxy used for their prediction, which was SYSI. All the maps were specialized using machine learning with the Random Forest algorithm, and the error metrics for each map are presented in the supplementary table A1. Further details on the spatialization of these attributes can b e found in (Safanelli et al., 2021, Rosas et al., 2023).

Table 2. Environmental covariates used for modeling phophorus stocks

Environmental covariate	Source (spatial resolution)	References	AVC	NVC
Annual Mean Temperature	WorldClim BIO Variables V1 (30 m)	(Hijmans et al., 2005)	yes	yes
Annual Precipitation	WorldClim BIO Variables V1 (30 m)	(Hijmans et al., 2005)	yes	yes
LST	Landsat collection (30 m)	(Ermida et al., 2020)	yes	yes
Elevation	SRTM (30 m)	(Farr and Kobrick, 2000)	yes	yes
Slope	TAGEE/SRTM (30 m)	(Safanelli et al., 2020)	yes	yes
Northness	TAGEE/SRTM (30 m)	(Safanelli et al., 2020)	yes	yes
Eastness	TAGEE/SRTM (30 m)	(Safanelli et al., 2020)	yes	yes
Horizontal curvature	TAGEE/SRTM (30 m)	(Safanelli et al., 2020)	yes	yes
Vertical curvature	TAGEE/SRTM (30 m)	(Safanelli et al., 2020)	yes	yes
Aspect	TAGEE/SRTM (30 m)	(Safanelli et al., 2020)	yes	yes
Hillshade	TAGEE/SRTM (30 m)	(Safanelli et al., 2020)	yes	yes
Terrain shape index	TAGEE/SRTM (30 m)	(Safanelli et al., 2020)	yes	yes
Mean curvature	TAGEE/SRTM (30 m)	(Safanelli et al., 2020)	yes	yes
Maximal Curvature	TAGEE/SRTM (30 m)	(Safanelli et al., 2020)	yes	yes
SYSI blue (450–520 nm)	Landsat collection (30 m)	(Demattê et al., 2018)	yes	no
SYSI green (520–600 nm)	Landsat collection (30 m)	(Demattê et al., 2018)	yes	no
SYSI red (630–690 nm)	Landsat collection (30 m)	(Demattê et al., 2018)	yes	no
SYSI NIR (760–900 nm)	Landsat collection (30 m)	(Demattê et al., 2018)	yes	no
SYSI SWIR1 (1550–1750 nm)	Landsat collection (30 m)	(Demattê et al., 2018)	yes	no
SYSI SWIR2 (2080–2350 nm)	Landsat collection (30 m)	(Demattê et al., 2018)	yes	no
mblue (450–520 nm)	Landsat collection (30 m)	In this study	yes	yes
mgreen (520–600 nm)	Landsat collection (30 m)	In this study	yes	yes
mred (630–690 nm)	Landsat collection (30 m)	In this study	yes	yes
mNIR (760–900 nm)	Landsat collection (30 m)	In this study	yes	yes
mSWIR1 (1550–1750 nm)	Landsat collection (30 m)	In this study	yes	yes
mSWIR2 (2080–2350 nm)	Landsat collection (30 m)	In this study	yes	yes
Clay	30 m	In this study	yes	no
Sand	30 m	In this study	yes	no

Silt	30 m	In this study	yes	no
SiO_2	30 m	(Rosas et al., 2023)	yes	no
Al_2O_3	30 m	(Rosas et al., 2023)	yes	no
Fe_2O_3	30 m	(Rosas et al., 2023)	yes	no
SOC	30 m	In this study	yes	no
pH_H_2O	30 m	In this study	yes	no

Synthetic Soil Image (SYSI); Terrain Analysis in Google Earth Engine (TAGEE); Shuttle Radar Topography Mission (SRTM). Land Surface Temperature (LST); Bands of mean Vegetation Reflectance (mBLUE, mGREEN, mRED, mNIR, mSWIR1, mSWIR2); Soil Organic Carbon (SOC); anthropogenic vegetation cover (AVC) and native vegetation cover (NVC).

3.2.5 Model tuning, performance and validation

The AP and TP maps were generated using Random Forest (RF) models trained independently for each defined sub-region (AVC and NVC). We employed a geospatial RF modeling pipeline developed by Van Den Hoogen et al. (2021). To determine the hyperparameters of the models, we conducted a grid search and tested various combinations to optimize the best ones. In the grid search, the following hyperparameters were explored: Variables per split (we tested the entire covariate set with an interval between tests of 4, for example, 4, 8, 12, ... n), minimum leaf population (we tested values of 3, 6, 9), and the maximum number of leaf nodes in each tree (we tested values of 500, 1000, 1500). The number of trees was fixed at 30, a value determined through prior testing to achieve the best possible performance with adequate computational cost.

To evaluate the model performance, we employed K-fold cross-validation. We set the number of folds (K) to 10, where sampling in each fold was random and stratified by biomes (Van Den Hoogen et al., 2021). We assessed the models using the following statistical metrics: coefficient of determination (R2), root mean square error (RMSE), and Ratio of Performance to Inter Quartile distance (RPIQ). The model selection was based on the lowest RMSE value, serving as the reference for mapping. The importance of predictor variables was evaluated using variable importance and ranked by the average importance across all models.

3.2.6 Phosphorus Stocks mapping

After selecting the best predictive models, they were used to predict the stocks of AP and TP in the study area. The raster files corresponding to the environmental covariates served as input data for the models, and the output consisted of AP and TP maps. Stratified bootstrapping sampling (by biome) was employed in the mapping process to create statistically valid pixel-wise confidence intervals and uncertainty estimates. In total, we conducted the bootstrapping procedure 100 times. The estimated value for each pixel in the final maps was formed by aggregating the means of the estimates obtained from each bootstrapping sample. Additionally, the standard deviation of the prediction estimates was also calculated. The standard deviation of the ensemble was used to compute the spatial distribution of uncertainty, defined by the 95% standardized confidence interval for the pixel-level mean estimate.

3.2.7 Statistical Analysis and Interpretation of Phosphorus Stock Maps

Our maps were interpreted in a logical sequence composed of three steps. First, we sought to understand the coherence of these maps with our understanding of Brazilian soils. In a second moment, we sought to understand how phosphorus stocks are distributed in Brazil. Lastly, we analyzed the importance of these stocks for the country's agricultural production.

3.2.7.1 Evaluating the coherence of the phosphorus stock maps of the soils

To verify the coherence of our maps with our understanding of Brazilian soils, we correlated our maps with soil observations contained in the Brazilian Soil Spectral Library database (Demattê et al., 2019), The chosen soil properties were pH, SOC, and the relative abundance of soil minerals such as hematite, goethite, gibbsite, and kaolinite. The choice of these properties was due to their governing the adsorption or availability of phosphorus in the soil. The pH used was measured in water, and the SOC was determined through wet combustion (Teixeira et al., 2017).

The relative abundance of minerals was determined based on soil reflectance spectra obtained in the Vis-NIR-SWIR region. The soil spectra were transformed to absorbance using the Kubelka-Munk function (KM) (Barron and Torrent, 1986). 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 features related to each mineral were defined in the SG spectra based on the literature. The range representing the feature of each mineral was used to calculate the mineral amplitude (MA), representing their relative abundance. The MA for each mineral was calculated as: $MA=Max\lambda-Min\lambda$, where $Max\lambda$ and $Min\lambda$ are the maximum (positive) and minimum (negative) values of the specific ranges. More details about the process steps can be found in Rosin et al, (2023).

3.2.7.2 Understanding the Distribution of Phosphorus Stocks in Brazil

To understand how phosphorus stocks are distributed throughout Brazilian soils, we calculated the average stocks of AP and TP according to the soil coverage within each Brazilian biome. In addition, we calculated the total stocks per biome. To determine soil coverage, we used the data collection from the MapBiomas project (MapBiomas, 2023), which provides annual soil coverage maps. In this study, we considered the soil coverage that occurred most in the last decade, for this we calculated the mode of the set of maps for the period from 2012 to 2022.

3.2.7.3 The Importance of Stocks for Brazil's Agricultural Production

The importance of phosphorus stocks for Brazilian agriculture was determined by correlating the average productivity of the country's main agricultural crops with the phosphorus stocks obtained in

our maps. We chose the five main agricultural crops produced: coffee, citrus, rice, sugarcane, and soy. The productivity of these crops was obtained from the database provided by the Brazilian Institute of Geography and Statistics (IBGE, 2021). We calculated the average of the harvests from 2017 to 2021, accounting for five harvests for each Brazilian municipality. The chosen period aimed to represent a current agricultural panorama with a five-year interval to smooth the influence of extreme and abnormal climatic events. Again, we used the data collection from the MapBiomas project on soil coverage to construct a mask and extract the average of phosphorus stocks only in pixels cultivated with each agricultural crop.

3.3 Results

Brazilian soils are naturally low in P, however, significant amounts of this element are fixed in the soil clays, making P unavailable to plants. Figure 2a presents the descriptive statistics of the TP stocks in the soil, calculated for the soil observations in the database used in the prediction. The median stocks of TP are approximately 40 $g.m^{-2}$, being slightly higher in the areas of NVC. The stocks of AP are much lower than the stocks of TP (Figure 2b), with median values of 0.64 and 1.7 $g.m^{-2}$ for areas of NVC and AVC respectively. The higher stocks in the AVC areas are due to the use of fertilizers in agriculture.

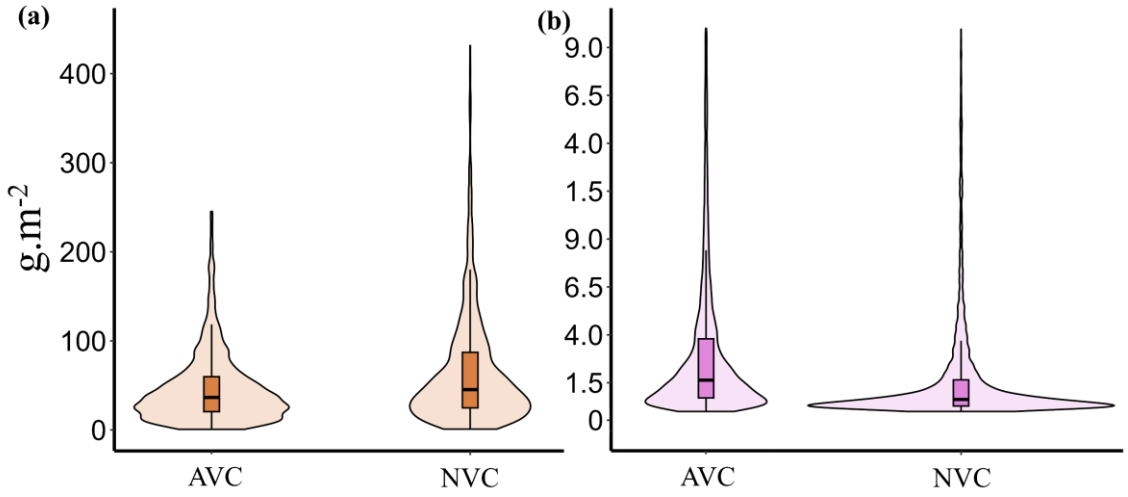

Figure 2. Descriptive statistics of total phosphorus stocks (a) and available phosphorus (b) in areas of anthropogenic vegetation cover (AVC) and in areas of native vegetation cover (NVC).

Four predictive models were adjusted to predict TP and AP stocks in the two defined sub-regions. The modeling results are presented in Figure 3. The predictive models adjusted for AVC areas showed better performance than in NVC areas. Performance was very similar among different forms of P (Figure 3). The highest R^2 was observed for TP stocks in AVC (0.35), which also exhibited the highest RPIQ (1.71); however, the smallest error, represented by RMSE, was observed in AP stocks in AVC areas (3.82%). On the other hand, the lowest R^2 and RPIQ were in the predictive model for AP in NVC areas, 0.27 and 1.30 respectively. The highest errors were observed in the TP stocks prediction model in NVC areas with an RMSE of 10.5%.

The environmental covariates that most influenced the prediction of TP stocks in AVC areas, in descending order, were Fe_2O_3, ANMT, ANP, LST, and the blue band of SYSI (Figure 3d). Notably, Fe_2O_3 had the highest importance order = 1 (maximum). As for TP stocks in NVC areas, where not all covariates were available, climatic covariates ANMT, ANP, and LST took on a prominent role. In addition to these, the SWIR1 band of mVR and altitude also held great importance. In the model adjusted to predict AP stocks in AVC areas, altitude (TA1) had the highest importance (0.98), followed by pH and the Nir band of mVR. Climatic variables ANMT and ANP also had high importance. In NVC areas, the most important variables were ANP with an importance of 1, LST with an importance of 0.77, and altitude with an importance of 0.67.

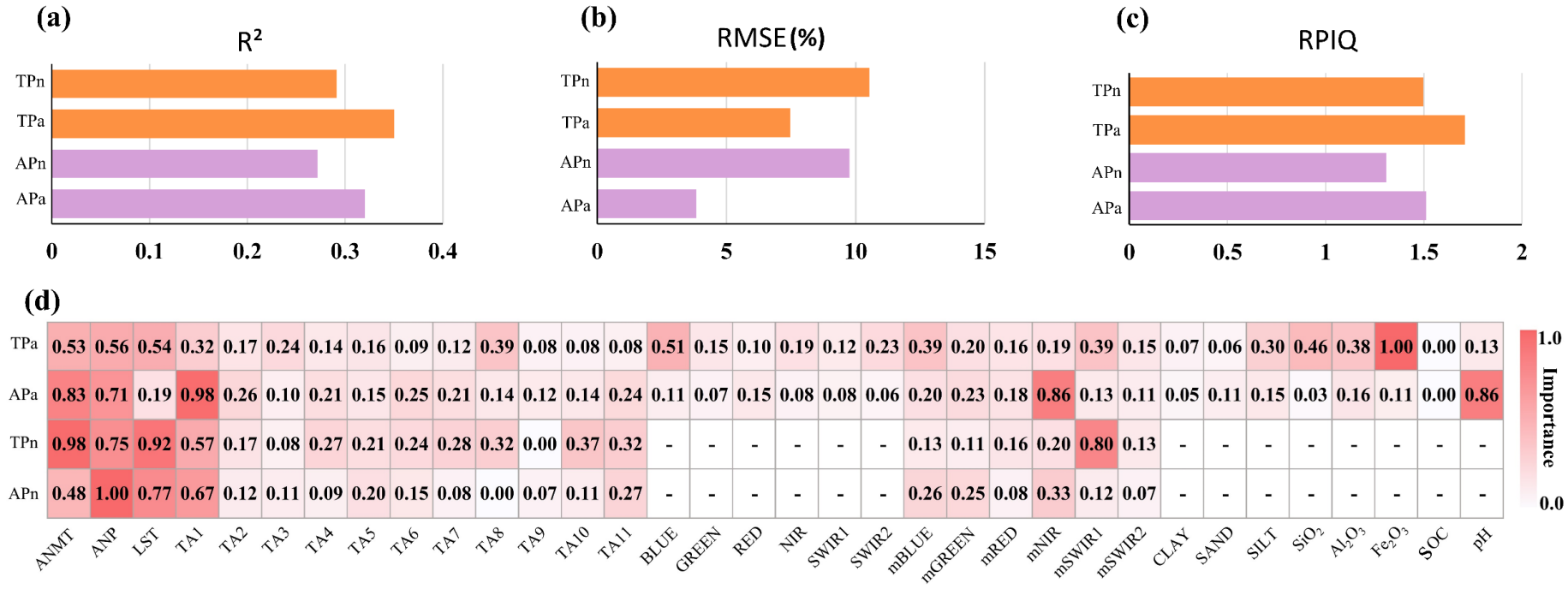

Figure 3. Performance evaluation of the prediction of soil phosphorus stocks using K-fold cross-validation with the Random Forest algorithm. Parameters of models used to predict soil phosphorus stocks (a, b and c); importance of environmental covariates in predicting soil phosphorus stocks (d). Root Mean Squared Error RMSE; Ratio of Performance to InterQuartile distance (RPIQ); coefficient of determination (R^2); total phosphorus stocks in the soil in areas of native vegetation cover (TPn); total phosphorus stocks in the soil in areas of anthropogenic vegetation cover (TPa); available phosphorus stocks in the soil in areas of native vegetation cover (APn); available phosphorus stocks in the soil in areas of anthropogenic vegetation cover (APa); Annual Mean Temperature (ANMT); Annual Precipitation (ANP); Land Surface Temperature (LST); Elevation (TA1); Slope (TA2); Eastness (TA3); Northness (TA4); Vertical Curvature (TA5); Horizontal Curvature (TA6); Aspect (TA7); Hillshade (TA8); Terrain shape index (TA9); Mean curvature (TA10); Maximal Curvature (TA11); synthetic soil image bands (BLUE, GREEN, RED, NIR, SWIR1, SWIR2); Bands of mean Vegetation Reflectance (mBLUE, mGREEN, mRED, mNIR, mSWIR1, mSWIR2; Soil Organic Carbon (SOC).

The maps of TP and AP stocks are presented in Figure 4. Our results showed that Brazil has a TP stock of 531 Mt and an AP stock of 17.4 Mt. The spatial distribution of P stocks aligns with the understanding of Brazilian soils. The highest TP stocks are found in soils originating from basaltic rocks (Serra Geral formation), in regions of higher altitude with milder climates, and in soils with higher carbon content. Soils with lower TP stocks are formed from sedimentary rocks, primarily in the drier regions of Brazil, such as the Caatinga biome. AP stocks are low across almost the entire Brazilian territory, with stocks below 5 $g.m^{-2}$ in approximately 96% of the area. Higher stocks can be observed in areas dedicated to agriculture, where phosphatic fertilizers are applied. Additionally, slightly higher AP stocks can be seen in the northeastern region of Brazil and the Pantanal biome, where less weathered soils predominate.

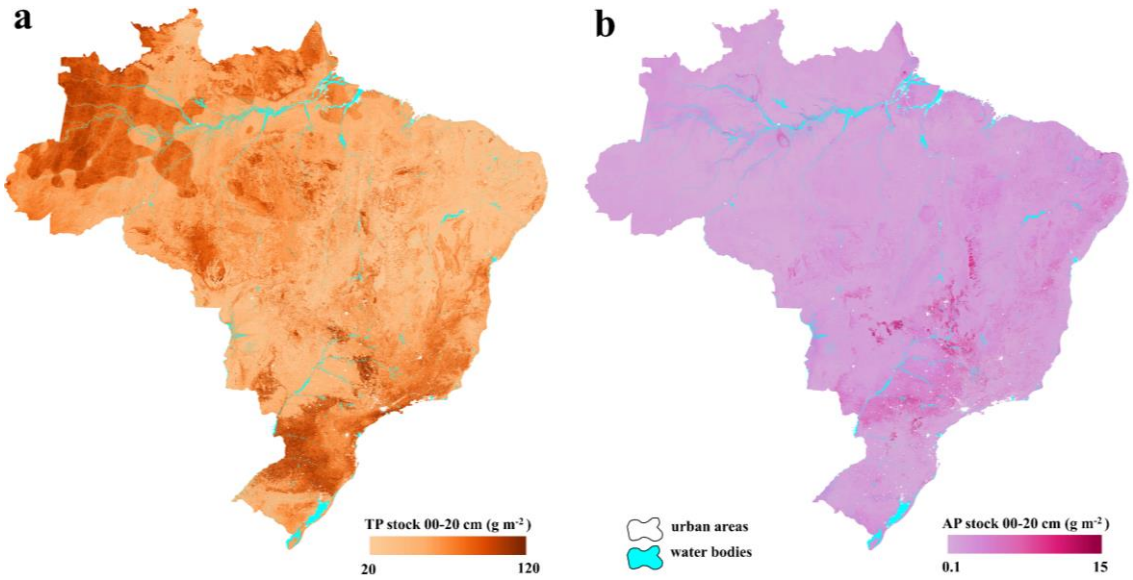

Figure 4. Maps of phosphorus stocks in Brazilian soils. Map of total phosphorus (TP) stocks in the soil layer from 0-20 cm (a); Map of available phosphorus (AP) stocks in the soil layer from 0-20 cm (b).

The calculation of the percentage of AP stocks in relation to TP stocks revealed significant variation across the Brazilian territory (Figure 5). This variation depends on agricultural fertilization and the soil's ability to adsorb P, making it unavailable to plants. The vast majority of Brazilian soils have less than 3% of the total P available to plants. Higher percentages can be observed in agricultural areas, particularly where soils are composed of sandstones, such as in the Paraná River basin (Bauru sequence) and in the Barreira region in the state of Bahia, which exhibits a combination of low clay content soils and increased agricultural activity in recent decades.

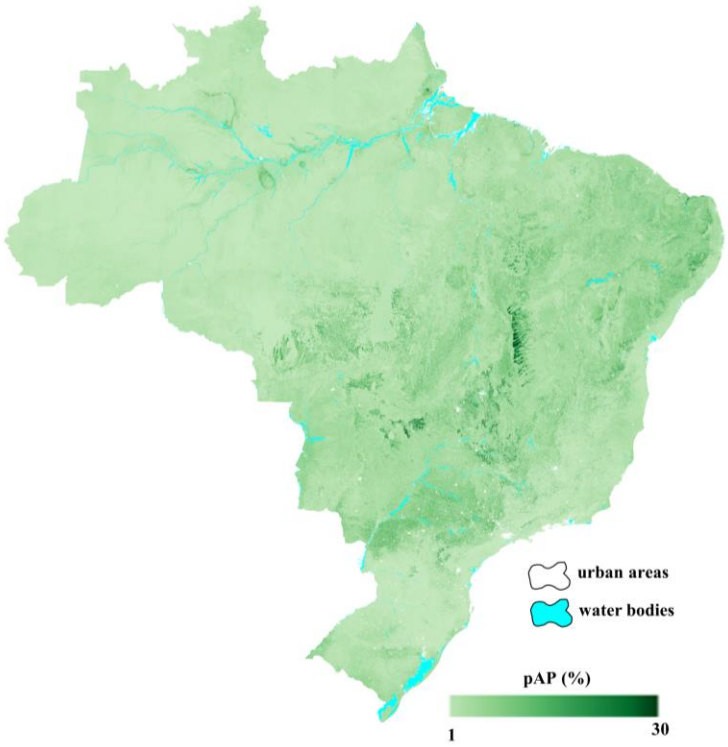

Figure 5. Map of the percentage of available phosphorus stocks in relation to total phosphorus stocks in Brazilian soils. Percentage of available phosphorus (pAP).

3.4 Discussion

We adopted a strategy that allowed us to map the P stocks throughout the Brazilian territory. The predictive models adjusted for the areas of AVC outperformed the models adjusted for the areas of NVC. In regions of the country where natural vegetation has been replaced by human intervention, there is a greater abundance of available data. This allows us to generate maps of attributes highly correlated with soil P contents with a high degree of accuracy, such as Fe_2O_3, clay, and SOC (Safanelli et al., 2021; Safanelli et al., 2021). Furthermore, in areas where natural vegetation has been removed, it is possible to recover the reflectance of bare soil, which has been used as a proxy for digital soil mapping (Demattê et al., 2020). Due to this greater abundance of data, the models performed better in areas of AVC.

P is present in soils in organic and inorganic forms. In tropical soils with a high degree of weathering, inorganic P is mostly adsorbed to iron oxides (mainly hematite and goethite) (Alovisi et al., 2020; de Oliveira et al., 2020). Therefore, there is a direct relationship between the Fe_2O_3 of the clay fraction with the TP stocks in the soil. This relationship ensures the greatest importance of this environmental variable in predicting TP stocks in areas of AVC. Organic P composes the soil organic matter (SOM). Despite this, SOC was not important in predicting P stocks. In the tropics, where microbial activity is intense, SOC levels are much lower compared to temperate climates (Wiesmeier et al., 2019). Furthermore, the low SOC levels of these soils are strongly complexed to soil oxides (Kleber et al., 2021). This association leads to a high degree of collinearity between Fe_2O_3 and SOC, resulting in the exclusion of this covariate to predict P stocks.

The predominant charges in the old soils of the tropics are positive, which attract phosphate ions adsorbing to soil colloids. Therefore, pH has a great influence on this adsorption. Soils with a pH lower than the point of zero charge (PZC) tend to adsorb more P, decreasing its availability to plants (Gérard, 2016; Stoop, 1980). This behavior explains the strong contribution of pH in predicting the stocks of AP in areas of AVC.

Soil formation factors are intrinsically associated with the P cycle in the Earth's crust and consequently with its stocks in the soil. The degree of soil formation changes the soil's capacity to adsorb P or make it available, allowing its migration to other ecosystems (Antoniadis et al., 2016). The inclusion in the modeling of variables that represented the soil formation factors such as relief and climate brought significant gains, mainly in the areas of NVC where not all covariates were available. In addition, these formation factors have a strong relationship with soil attributes such as pH, Fe_2O_3, and clay (Wadoux et al., 2020), which were of great importance in predicting P stocks in areas of AVC. Thus, these attributes were indirectly represented in the NVC areas.

The spectral characteristic of vegetation was important for the model adjusted for AP in areas of AVC, but it was not important for areas of NVC. The P content in the soil has a direct effect on plant development, which can alter its reflectance (Siedliska et al., 2021). However, the vegetation in NVC areas is extremely adapted to develop in soils with low P levels (Lambers et al., 2020). On the other hand, in AVC areas, agricultural crops that have undergone extensive genetic improvement processes throughout history tend to be extremely responsive to soil P levels (Ahmad et al., 2001). Therefore, the spatial variations of P levels in the soil influence the reflectance of the crop canopy much more than native vegetation.

Brazilian soils show a significant variation in TP and AP stocks. The variation in TP stocks aligns with soil mineralogy and soil organic carbon (SOC), as expected. We observed positive correlations between TP and SOC (0.56), and with the relative abundance of minerals such as Hematite (0.47), Gibbsite (0.40), and Goethite (0.35) (Figure 6). These correlations are due to the predominance of P adsorption on these minerals and the organic P present in the SOC (Alovisi et al., 2020; Dos Santos et al., 2008) .

The highest P stocks found in soils developed over basaltic rocks from the basaltic spill in the Serra Geral Formation are due to the chemical composition of these rocks. Ultramafic basaltic rocks are chemically richer, presenting higher P contents compared to other rocks present in Brazil, for example, sandstones (Daly et al., 2015; Mage and Porder, 2013). However, these rocks are also rich in metallic elements such as Fe and Al, which, during weathering processes, form minerals with high P adsorption potential, such as hematite, goethite, and gibbsite (Riffel et al., 2016).

The highest total TP stocks observed in the Amazon Forest and the Atlantic Forest are due to the greater accumulation of SOC in these biomes due to their dense biomass (Englund et al., 2017; Schaefer et al., 2008). In the southern region of Brazil, there are soils formed over basaltic rocks in regions of higher altitude, which contribute to the accumulation of SOC (Gomes et al., 2019). In this region are the country's largest TP stocks. The high TP stocks observed in the Rio Negro basin in the Amazon were not detected in the study on the native P contents in Brazilian soils conducted by Pavinatto et al (2020). There is a predominance of sandy and hydromorphic soils in these regions. Due to

hydromorphism and the large plant biomass produced by the forest, carbon is accumulated in the soils (Montes et al., 2011). We believe that for this reason, the vast majority of P present in these soils is in the organic form. However, there are very few soil observations in these regions, and more studies should be conducted for more concrete conclusions.

The low TP stocks observed in the northeastern region of Brazil, in the Paraná River basin, and in some plateaus such as the region of Barreiras in the state of Bahia are mainly due to the low P contents in the parent material. These regions are dominated by sandstones composed mainly of quartz and low P contents or by felsic rocks that also contain low P contents (da Silva et al., 2022; Schaefer et al., 2023). In addition, in the northeastern region, the arid climate negatively influences the inputs of organic P into the soil, reducing its stocks (Menezes et al., 2012). On the other hand, in these arid regions, the youngest soils are concentrated with a predominance of 2:1 clay minerals and a higher pH compared to the averages of Brazilian soils (de Araújo Filho et al., 2023), as a consequence, the stocks of AP are larger. In general, soils with a more alkaline pH showed higher stocks of AP, where the linear correlation analysis between the predicted AP stocks and the soil pH was 0.32 (Figure 6).

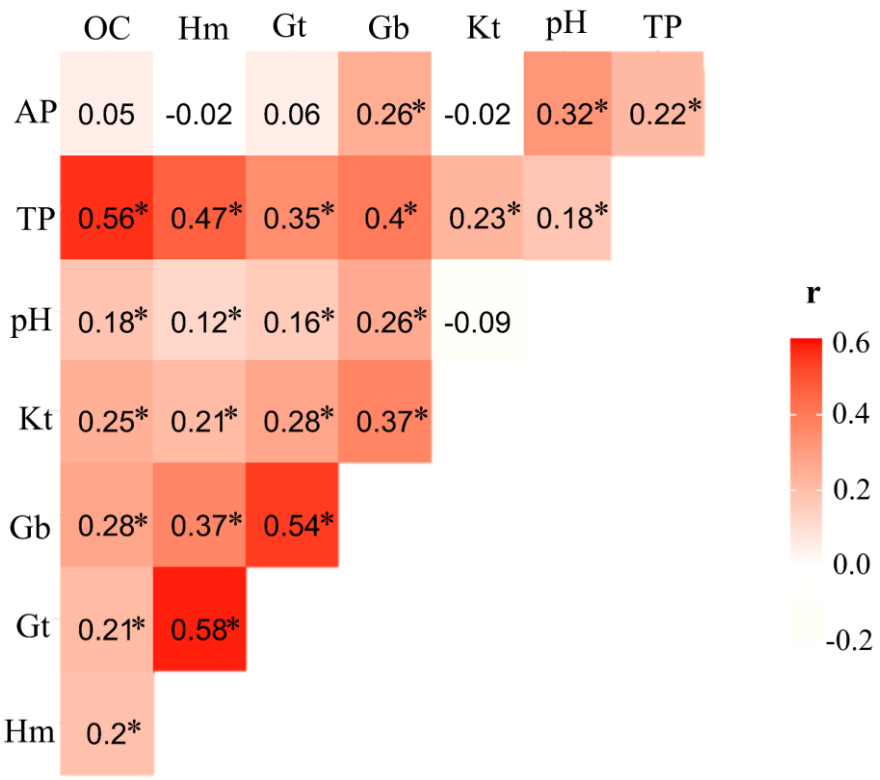

Figure 6. Pearson's linear correlation analysis between soil observations and the maps of phosphorus stocks in Brazilian soils. Soil organic carbon (OC); hematite (Hm); goethite (Gt); gibbsite (Gb); kaolinite (Kt), soil pH measured in water (pH); total phosphorus stock (TP); available phosphorus stock (AP); * Signifcant at 5% probability level.

The pedogenetic processes that occur in the soils of the Pantanal contribute to the formation of soils with lower degrees of weathering that have larger natural AP stocks than most Brazilian soils (Coringa et al., 2012; Couto et al., 2023). The larger AP stocks in the soils of the Pantanal are due to the low potential for P adsorption in these soils. The low P adsorption is a result of the predominance of 2:1

type clay minerals and more alkaline pH due to higher calcium contents in the parent material (De Souza Oliveira et al., 2021). In addition, the flooding of the plains creates a reducing environment, releasing the P from the clays and making it available (Hamilton et al., 1997). The sediments that are deposited in the floodplains also contribute to the increase in AP stocks in the soils (Owens, 2020). The interaction of all the above factors results in the spatial behavior of the AP stocks presented in Figure 7a, where the stocks of the low areas (floodable) are always higher than the stocks of the high areas.

Agricultural areas that receive constant applications of phosphate fertilizers have therefore increased their AP stocks compared to native areas. Not all P applied to the soil is exported by crops, generating a residual P that may or may not be adsorbed to the soil clays, this P is called legacy P (Rodrigues et al., 2016). In Figure 7b we can observe an example of the increases in AP stocks in areas cultivated with soybeans compared to native areas and pasture areas in soils with similar pedology. In addition, in regions where there is greater technological input, such as in circular areas (irrigated regions), AP stocks are even larger. According to Pavinato et al., (2020) the legacy P of the soil already amounted to about 33.4 Tg accumulated in agricultural soils between the period of 1967 to 2016.

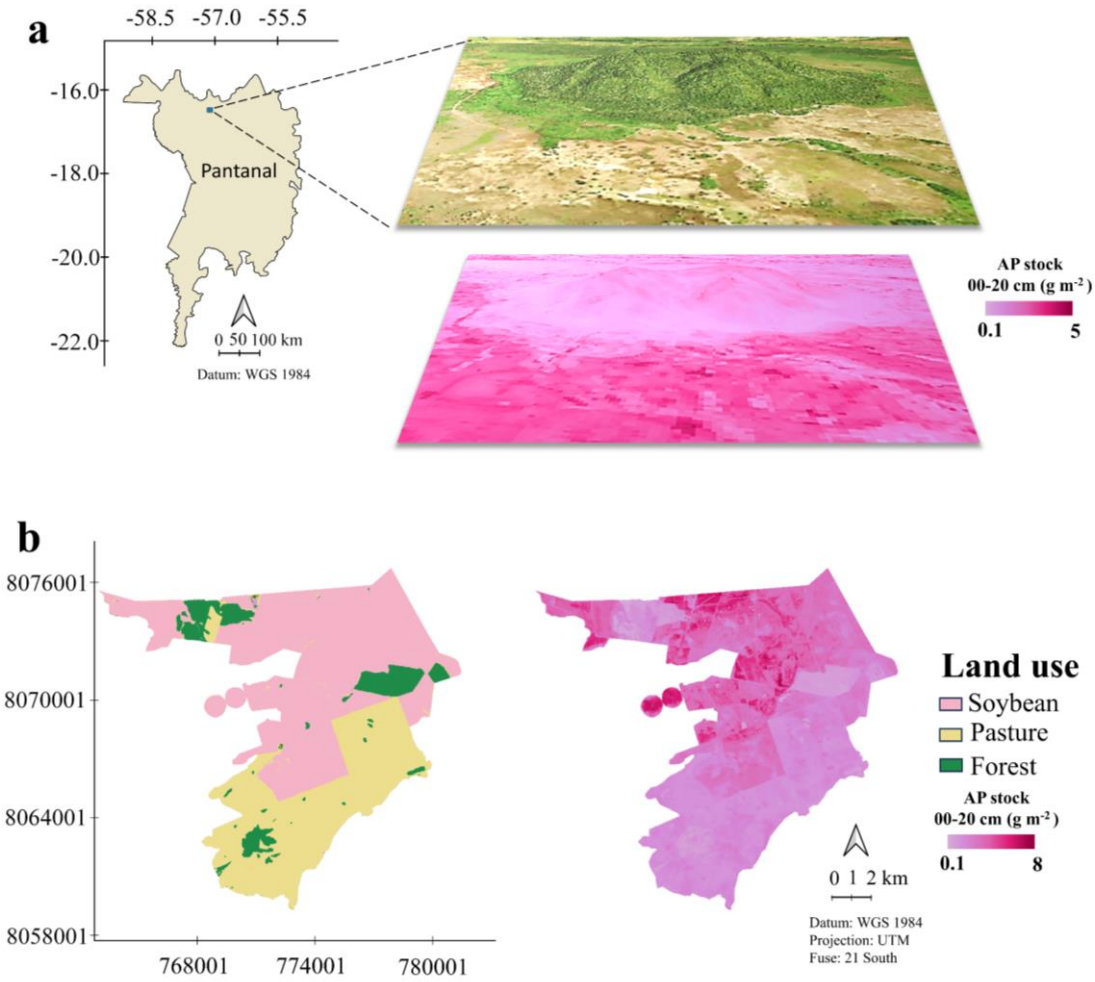

Figure 7. Spatial behavior of available phosphorus (AP) stocks in different landscapes. (a) typical landscapes of the Brazilian Pantanal, where the lower parts are flooded for long periods, (b) landscape dominated by similar pedology with a clear contrast in AP stocks between areas cultivated with soybean, pasture, and native forest.

3.4.1 Detailing phosphorus stocks at the level of soil coverage

Upon quantifying the P stocks by biome, we note that the largest stocks are found in the Amazon biome, containing 54% of the TP stocks and 39% of the AP stocks (Figure 8a and 8b). Despite the highest average values of the stocks not being found in this biome, its vast extent contributed to the results. The Amazon biome has the lowest average values of AP stocks, with 1.58 $g.m^{-2}$ (Figure 8c). According to Cunha et al., (2022) the soils of this biome have low phosphorus content, even limiting the productivity of the Amazon Forest. The highest average stocks of AP were observed in the Caatinga biome with 2.51 $g.m^{-2}$ and the Cerrado biome with 2.46 $g.m^{-2}$. The soils of the Caatinga biome mostly have a lower capacity to adsorb P, contributing to higher levels of this nutrient available for plants (de Araújo Filho et al., 2023). In addition, the arid climate contributes to minimal P exports to aquatic ecosystems (de Araújo Filho et al., 2023). On the other hand, the arid climate limits the production of biomass by vegetation, and consequently, the levels of organic P are lower, reducing the TP stocks (Menezes et al., 2012). In the Cerrado, the soils are naturally poor in P, however, the great agricultural advance in this biome contributed to the increase in P stocks due to the application of fertilizers (Rodrigues et al., 2016). In the Atlantic Forest is the highest average of TP stocks in Brazil, 73.8 $g.m^{-2}$, due to the organic P present in the high carbon stocks of this biome (Gomes et al., 2019).

Within the Brazilian biomes, large variations in P stocks were observed in different soil covers (Figure 8d). In general, forest formations and mangrove areas presented the highest TP stocks, ranging from approximately 50 $g.m^{-2}$ in the Pantanal to 80 $g.m^{-2}$ in the Atlantic Forest. These higher contents in the forest formation are associated with the P contained in the carbon stocks that tend to be larger in forest formations (Gomes et al., 2019). Mangroves only occurred in the Amazon, Atlantic Forest, and Caatinga biomes, and the TP stocks were approximately 80 $g.m^{-2}$ in the Amazon and Atlantic Forest biomes, while in the Caatinga these stocks were around 70 $g.m^{-2}$. Mangroves are areas of nutrient accumulation such as P, which is exported from terrestrial to aquatic ecosystems, hence these areas have larger P stocks (Pérez et al., 2021).

Higher TP stocks were also observed in areas occupied by soybean cultivation in the Atlantic Forest and in areas occupied by sugarcane cultivation in the Cerrado. These higher stocks are due to the cultivation of these crops (in these biomes) in highly P-fixing oxidic soils formed from basaltic rocks (Schaefer et al., 2008). In all biomes, the transformation of native areas such as forest formation and Savanna Formation into agricultural areas culminated in the reduction of total P stocks. However, the Pampa biome is the one that presents the smallest contrasts between areas with agricultural cover and with native cover. The transformation of native areas into agricultural areas reduces the carbon stocks of the soil in the Brazilian biomes on a progressive scale of 1 to 26% as demonstrated by Medeiros et al., (2022). This occurs due to disturbances in the soil caused by the management of agricultural crops leading to the oxidation of SOM (Zinn et al., 2005). The oxidation of SOM results in the mineralization of organic P, making it available to plants. This P ends up being exported by the crops, reducing their stocks in the soil (Spohn, 2020). In the Pampa, where the climate is milder than the average in Brazil, the decomposition of SOM is slower, preserving it even if native ecosystems are converted into agricultural ecosystems (Kaschuk et al., 2011).

As expected, the stocks of available phosphorus (AP) tended to be higher in agricultural areas, with emphasis on cotton, soybeans, and sugarcane crops with average values ranging from 4 to 7.5 $g.m^{-2}$. This confirms what we discussed earlier, AP stocks tend to increase in agricultural areas due to the legacy P originating from chemical fertilizations (Pavinato et al., 2020). In addition, the wetlands in the Pantanal also presented considerable stocks of AP (approximately 3.0 $g.m^{-2}$). This happens due to the low capacity to adsorb P from soils in hydromorphic environments of the Pantanal (Couto et al., 2023). In general, the forest formation was the soil cover that presented the lowest stocks of AP in all biomes.

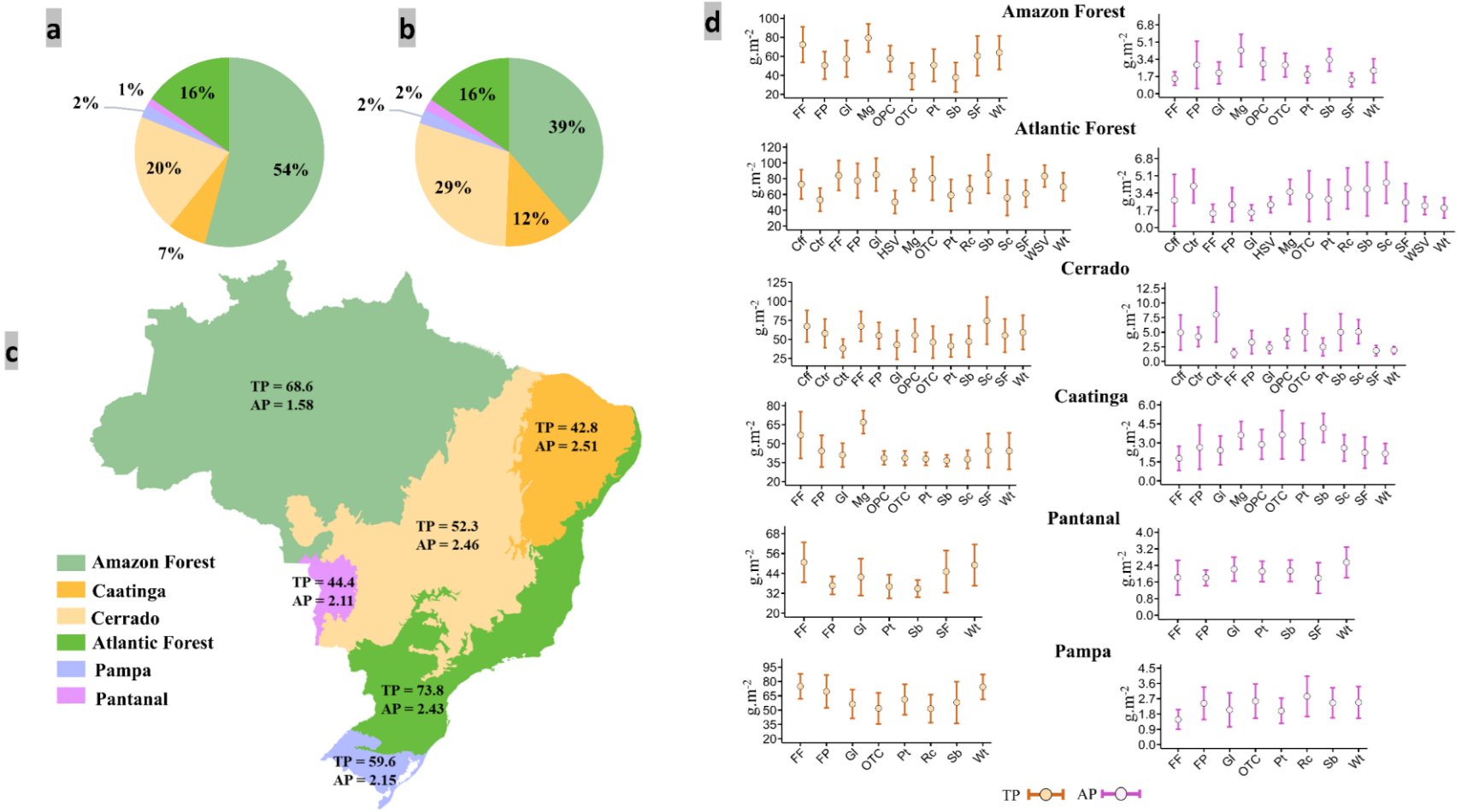

Figure 8: Spatial distribution of phosphorus stocks in Brazilian territory; proportion in percentage of total phosphorus stocks in Brazilian biomes (a); proportion in percentage of available phosphorus stocks in Brazilian biomes (b); average phosphorus stocks in the soils of different Brazilian biomes in $g.m^{-2}$ (c); Phosphorus stock as a function of soil cover (d). Total phosphorus stock (TP); available phosphorus stock (AP); FF: Forest Formation; SF: Savanna Formation; Mg: Mangrove; WSV: Wooded Sandbank Vegetation; Wt: Wetland; Gl: Grassland; HSV: Herbaceous Sandbank Vegetation; Pt: Pasture; Sb: Soybean; Sc: Sugar cane; Rc: Rice; Ctt: Cotton; OTC: Other Temporary Crops; Cff: Coffee; Ctr: Citrus; OPC: Other Perennial Crops; FP: Forest Plantation.

3.4.2 Impact of phosphorus stocks on Brazilian agriculture

Brazil is a global leader in food production and one of the largest exporters (Calil and Ribeira, 2019). The success of Brazilian agriculture is driven by the expansion of the agricultural frontier, mainly in the Cerrado biome (Soterroni et al., 2019). Brazil pioneered achieving high productivity in acidic and infertile soils, such as those of the Brazilian savannas (Nehring, 2022). This was made possible thanks to the application of fertilizers and agricultural correctives, especially phosphate fertilizers, due to the low availability of P in Brazilian soils (Rodrigues et al., 2016).

The combination of low P stocks in Brazilian soils and the success of Brazilian agriculture, driven by the incorporation of chemical fertilizers, makes Brazilian agriculture vulnerable to the depletion of mineral P reserves in the world. This is exacerbated by the high rates of P adsorption in the main agricultural areas (Roy et al., 2016). In addition, the main P reserves in the world are not in Brazil, making Brazilian agriculture dependent on international fertilizer trade (Barbieri et al., 2021).

In this study, we correlated the productivity of the main agricultural crops with the P stocks in Brazilian soils. We observed a correlation ranging from 0.2 for coffee and 0.46 for soybeans with AP stocks (Figure 9). These correlations are considered high, given that P is just one of many factors that can affect crop productivity. The correlations between crop productivity and TP stocks were not high for most crops, which was expected, as there is not a high correlation between available AP stocks and TP. However, the correlations show that the higher the percentage of AP relative to TP, the higher the productivity of agricultural crops. These findings reinforce the need for sustainable P management strategies, given the dependence of Brazilian agriculture on P and the imminent global scarcity of this mineral. The maps in this study provide a basis for the development of strategies for the sustainable use of P in Brazilian agriculture.

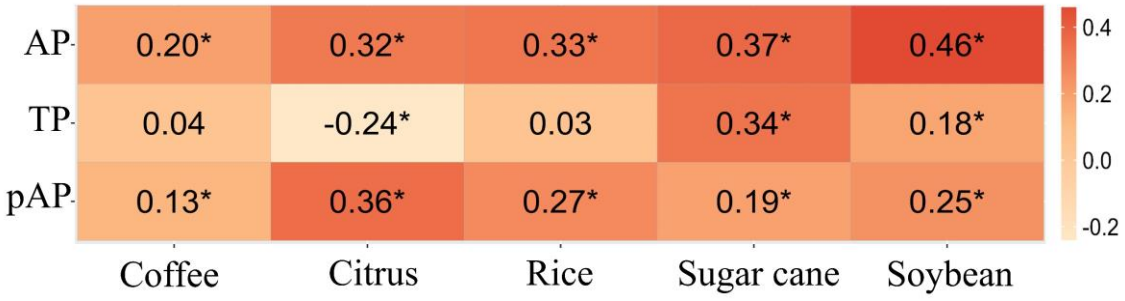

Figure 9. Pearson correlation analysis between the productivity of the main Brazilian agricultural crops and the phosphorus stocks in the soil. Total phosphorus stock (TP); available phosphorus stock (AP); percentage of available phosphorus (pAP); * Signifcant at 5% probability level.

3.5 Conclusions

We were able to predict the stocks of available phosphorus (AP) and total phosphorus (TP) with high spatial resolution (30m) using satellite-derived data, soil properties, and machine learning. Despite low determination coefficients, ranging from 0.27 for AP stocks in NVC areas to 0.35 for TP stocks in AVC areas, the obtained maps align with our understanding of Brazilian soils. Predictions generally performed better in AVC areas where more data and covariates were available.

The maps displayed expected spatial patterns. The highest TP stocks were observed in regions with soils formed from basaltic rocks and soils with higher SOC accumulation. The highest AP stocks were observed in soils with lower P retention, in arid regions of Brazil and the Pantanal biome, and in agricultural areas where P is added as fertilizer. In total, P stocks in Brazil were 531 Mt for TP and 17.4 Mt for AP. About 54% of TP stocks and 39% of AP stocks are in the Amazon biome, despite its soils not having the highest average values of these stocks. The Caatinga biome has the highest average AP stocks, with 2.51 $g.m^{-2}$, and the Atlantic Forest biome holds the highest average TP stocks with 73.8 $g.m^{-2}$.

Correlation was verified between TP stocks and soil mineralogy, mainly minerals responsible for P adsorption in soils such as hematite and gibbsite, which had correlations of 0.47 and 0.40, respectively. TP stocks also correlated with SOC at 0.56. AP stocks correlated positively with soil pH. In addition, AP stocks showed correlation with the productivity of the main Brazilian agricultural crops, varying from 0.2 for coffee to 0.46 for soybeans, indicating the existence of vulnerability if global P reserves were to be depleted.

In conclusion, the findings of this study provide a solid foundation for the sustainable management of phosphorus in Brazilian agriculture. The identification of areas with low phosphorus stocks and the correlation with crop productivity allow for the implementation of more efficient phosphorus management strategies. This could include the more targeted application of phosphate fertilizers, minimizing waste and maximizing productivity. Furthermore, understanding the dependence of Brazilian agriculture on phosphorus underscores the need to explore alternative sources of phosphorus and technologies to improve phosphorus use efficiency. These actions are crucial to ensure the sustainability of Brazilian agriculture in the face of the imminent global phosphorus scarcity.

Acknowledgments

We acknowledge FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) project 2014-22262-0, 2021-05129, CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) process number: 88887.481452/2020-00 and the Geotehcnologies on Soil Science (https://esalqgeocis.wixsite.com/geocis).

REFERENCE

Ahmad, Z., Gill, M.A., Qureshi, R.H., 2001. Genotypic variations of phosphorus utilization efficiency of crops. J Plant Nutr 24, 1149–1171. https://doi.org/10.1081/PLN-100106973

Alovisi, A.M.T., Cassol, C.J., Nascimento, J.S., Soares, N.B., da Silva Junior, I.R., da Silva, R.S., da Silva, J.A.M., 2020. Soil factors affecting phosphorus adsorption in soils of the Cerrado, Brazil. Geoderma Regional 22, e00298. https://doi.org/10.1016/J.GEODRS.2020.E00298

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

Amer, F., Bouldin, D.R., Black, C.A., Duke, F.R., 1955. Characterization of soil phosphorus by anion exchange resin adsorption and P32-equilibration. Plant Soil 6, 391–408. https://doi.org/10.1007/BF01343648/METRICS

Antoniadis, V., Koliniati, R., Efstratiou, E., Golia, E., Petropoulos, S., 2016. Effect of soils with varying degree of weathering and pH values on phosphorus sorption. Catena (Amst) 139, 214–219. https://doi.org/10.1016/J.CATENA.2016.01.008

Barbieri, P., MacDonald, G.K., Bernard de Raymond, A., Nesme, T., 2021. Food system resilience to phosphorus shortages on a telecoupled planet. Nature Sustainability 5, 114–122. https://doi.org/10.1038/s41893-021-00816-1

Barbosa, J.Z., Poggere, G., Mancini, M., Silva, S.H.G., Motta, A.C.V., Marques, J.J.G. de S. e. M., Curi, N., 2022. National-scale spatial variations of soil phosphorus retention capacity in Brazil. Physics and Chemistry of the Earth, Parts A/B/C 128, 103271. https://doi.org/10.1016/J.PCE.2022.103271

Barron, V., Torrent, J., 1986. Use of the Kubelka—Munk theory to study the influence of iron oxides on soil colour. Journal of Soil Science 37, 499–510. https://doi.org/10.1111/J.1365-2389.1986.TB00382.X

Benites, V.M., Machado, P.L.O.A., Fidalgo, E.C.C., Coelho, M.R., Madari, B.E., 2007. Pedotransfer functions for estimating soil bulk density from existing soil survey reports in Brazil. Geoderma 139, 90–97. https://doi.org/10.1016/J.GEODERMA.2007.01.005

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

Bonfatti, B.R., Hartemink, A.E., Giasson, E., Tornquist, C.G., Adhikari, K., 2016. Digital mapping of soil carbon in a viticultural region of Southern Brazil. Geoderma 261, 204–221. https://doi.org/10.1016/J.GEODERMA.2015.07.016

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

Calil, Y.C.D., Ribeira, L., 2019. Brazil's Agricultural Production and Its Potential as Global Food Supplier. Choices 1–12.

Chen, S., Arrouays, D., Leatitia Mulder, V., Poggio, L., Minasny, B., Roudier, P., Libohova, Z., Lagacherie, P., Shi, Z., Hannam, J., Meersmans, J., Richer-de-Forges, A.C., Walter, C., 2022. Digital mapping of

GlobalSoilMap soil properties at a broad scale: A review. Geoderma 409, 115567. https://doi.org/10.1016/J.GEODERMA.2021.115567

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

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

Cunha, H.F.V., Andersen, K.M., Lugli, L.F., Santana, F.D., Aleixo, I.F., Moraes, A.M., Garcia, S., Di Ponzio, R., Mendoza, E.O., Brum, B., Rosa, J.S., Cordeiro, A.L., Portela, B.T.T., Ribeiro, G., Coelho, S.D., de Souza, S.T., Silva, L.S., Antonieto, F., Pires, M., Salomão, A.C., Miron, A.C., de Assis, R.L., Domingues, T.F., Aragão, L.E.O.C., Meir, P., Camargo, J.L., Manzi, A.O., Nagy, L., Mercado, L.M., Hartley, I.P., Quesada, C.A., 2022. Direct evidence for phosphorus limitation on Amazon forest productivity. Nature 608, 558–562. https://doi.org/10.1038/s41586-022-05085-2

da Silva, R.J.A.B., da Silva, Ygor Jacques Agra Bezerra, van Straaten, P., do Nascimento, C.W.A., Biondi, C.M., da Silva, Yuri Jacques Agra Bezerra, de Araújo Filho, J.C., 2022. Influence of parent material on soil chemical characteristics in a semi-arid tropical region of Northeast Brazil. Environ Monit Assess 194, 1–21. https://doi.org/10.1007/S10661-022-09914-9/TABLES/8

Daly, K., Styles, D., Lalor, S., Wall, D.P., 2015. Phosphorus sorption, supply potential and availability in soils with contrasting parent material and soil chemical properties. Eur J Soil Sci 66, 792–801. https://doi.org/10.1111/EJSS.12260

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

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

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 Regional 22, e00318. https://doi.org/10.1016/J.GEODRS.2020.E00318

de Souza Medeiros, A., dos Santos, T.C., Maia, S.M.F., 2022. Effect of long-term and soil depth on soil organic carbon stocks after conversion from native vegetation to conventional tillage systems in Brazil. Soil Tillage Res 219, 105336. https://doi.org/10.1016/J.STILL.2022.105336

De Souza Oliveira, N., Schiavo, J.A., Laranjeira, L.T., De Moraes, E.M.V., Lima, M.F., Nunes, G.P., Pereira, M.G., 2021. Forms of inorganic phosphorus in carbonatic soils in the Pantanal of Mato Grosso do Sul, Brazil. Soil Research 59, 737–745. https://doi.org/10.1071/SR21007

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. de V., Nóbrega, G.N., da Silva, M.Z., de Souza, S.F., Valladares, G.S., Viana, J.H.M., da Silva Terra, F., Horák-Terra, I., Fiorio, P.R., da Silva, R.C., Frade Júnior, E.F., Lima, R.H.C., Alba, J.M.F., de Souza Junior, V.S., Brefin, M.D.L.M.S., Ruivo, M.D.L.P., Ferreira, T.O., Brait, M.A., Caetano, N.R., Bringhenti, I., de Sousa Mendes, W., Safanelli, J.L., Guimarães, C.C.B., Poppiel, R.R., e Souza, A.B., Quesada, C.A., do Couto, H.T.Z., 2019. The Brazilian Soil Spectral Library (BSSL): A general view, application and challenges. Geoderma 354, 113793. https://doi.org/10.1016/J.GEODERMA.2019.05.043

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

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

Dharumarajan, S., Hegde, R., 2022. Digital mapping of soil texture classes using Random Forest classification algorithm. Soil Use Manag 38, 135–149. https://doi.org/10.1111/SUM.12668

Dos Santos, D.R., Gatiboni, L.C., Kaminski, J., 2008. Fatores que afetam a disponibilidade do fósforo e o manejo da adubação fosfatada em solos sob sistema plantio direto. Ciência Rural 38, 576–586. https://doi.org/10.1590/S0103-84782008000200049

Elser, J.J., 2012. Phosphorus: a limiting nutrient for humanity? Curr Opin Biotechnol 23, 833–838. https://doi.org/10.1016/J.COPBIO.2012.03.001

Embrapa Solos, 2018. Sistema brasileiro de classificação de solos, Embrapa Solos. Rio de Janeiro.

Englund, O., Sparovek, G., Berndes, G., Freitas, F., Ometto, J.P., Oliveira, P.V.D.C.E., Costa, C., Lapola, D., 2017. A new high-resolution nationwide aboveground carbon map for Brazil. Geo 4, e00045. https://doi.org/10.1002/GEO2.45

Ermida, S.L., Soares, P., Mantas, V., Göttsche, F.M., Trigo, I.F., 2020. Google Earth Engine Open-Source Code for Land Surface Temperature Estimation from the Landsat Series. Remote Sensing 12, 1471. https://doi.org/10.3390/RS12091471

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

Gérard, F., 2016. Clay minerals, iron/aluminum oxides, and their contribution to phosphate sorption in soils — A myth revisited. Geoderma 262, 213–226. https://doi.org/10.1016/J.GEODERMA.2015.08.036

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, A.M.R., de Jong van Lier, Q., Silvero, N.E.Q., Inforsato, L., de Melo, M.L.A., Rodríguez-Albarracín, H.S., Rosin, N.A., Rosas, J.T.F., Rizzo, R., Demattê, J.A.M., 2023. Digital mapping of the soil available water capacity: tool for the resilience of agricultural systems to climate change. Science of The Total Environment 882, 163572. https://doi.org/10.1016/J.SCITOTENV.2023.163572

Hamilton, S.K., Sippel, S.J., Calheiros, D.F., Melack, J.M., 1997. An anoxic event and other biogeochemical effects of the Pantanal wetland on the Paraguay River. Limnol Oceanogr 42, 257–272. https://doi.org/10.4319/LO.1997.42.2.0257

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25, 1965–1978. https://doi.org/10.1002/JOC.1276

Hou, E., Chen, C., Luo, Y., Zhou, G., Kuang, Y., Zhang, Y., Heenan, M., Lu, X., Wen, D., 2018. Effects of climate on soil phosphorus cycle and availability in natural terrestrial ecosystems. Glob Chang Biol 24, 3344–3356. https://doi.org/10.1111/GCB.14093

IBGE, 2021. Sistema IBGE de Recuperação Automática - SIDRA [WWW Document]. URL https://sidra.ibge.gov.br/home/pimpfbr/brasil (accessed 12.1.23).

Kaschuk, G., Alberton, O., Hungria, M., 2011. Quantifying effects of different agricultural land uses on soil microbial biomass and activity in Brazilian biomes: Inferences to improve soil quality. Plant Soil 338, 467–481. https://doi.org/10.1007/S11104-010-0559-Z/TABLES/3

Khaledian, Y., Miller, B.A., 2020. Selecting appropriate machine learning methods for digital soil mapping. Appl Math Model 81, 401–418. https://doi.org/10.1016/J.APM.2019.12.016

Kleber, M., Bourg, I.C., Coward, E.K., Hansel, C.M., Myneni, S.C.B., Nunan, N., 2021. Dynamic interactions at the mineral–organic matter interface. Nature Reviews Earth & Environment 2, 402–421. https://doi.org/10.1038/s43017-021-00162-y

Lambers, H., de Britto Costa, P., Oliveira, R.S., Silveira, F.A.O., 2020. Towards more sustainable cropping systems: lessons from native Cerrado species. Theor Exp Plant Physiol 32, 175–194. https://doi.org/10.1007/S40626-020-00180-Z

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

Mage, S.M., Porder, S., 2013. Parent Material and Topography Determine Soil Phosphorus Status in the Luquillo Mountains of Puerto Rico. Ecosystems 16, 284–294. https://doi.org/10.1007/S10021-012-9612-5/TABLES/2

MapBiomas, Projeto., 2023. Coleção 7.1 da Série Anual de Mapas de Uso e Cobertura da Terra do Brasil [WWW Document]. URL https://plataforma.brasil.mapbiomas.org/cobertura?activeBaseMap=9&layersOpacity=100&activeModule=coverage&activeModuleContent=coverage%3Acoverage_main&activeYear=2022&mapPosition=-15.072124%2C-51.459961%2C4&timelineLimitsRange=1985%2C2022&baseParams[territoryType]=1&baseParams[territories]=1%3BBrasil%3B1%3BPa%C3%ADs%3B0%3B0%3B0%3B0&baseParams[activeClassTreeOptionValue]=default&baseParams[activeClassTreeNodeIds]=1%2C7%2C8%2C9%2C10%2C11%2C2%2

C12%2C13%2C14%2C15%2C16%2C17%2C3%2C18%2C19%2C28%2C30%2C31%2C32%2C33%2C34%2C29%2C35%2C36%2C37%2C38%2C20%2C21%2C4%2C22%2C23%2C24%2C25%2C5%2C26%2C27%2C6&baseParams[activeSubmodule]=coverage_main&baseParams[yearRange]=1985-2022 (accessed 11.30.23).

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

Mello, F.A.O., Demattê, J.A.M., Bellinaso, H., Poppiel, R.R., Rizzo, R., de Mello, D.C., Rosin, N.A., Rosas, J.T.F., Silvero, N.E.Q., Rodríguez-Albarracín, H.S., 2023. Remote sensing imagery detects hydromorphic soils hidden under agriculture system. Scientific Reports 13, 1–10. https://doi.org/10.1038/s41598-023-36219-9

Menezes, R.S.C., Sampaio, E.V.S.B., Giongo, V., Pérez-Marin, A.M., 2012. Biogeochemical cycling in terrestrial ecosystems of the Caatinga Biome. Brazilian Journal of Biology 72, 643–653. https://doi.org/10.1590/S1519-69842012000400004

Menge, D.N.L., Kou-Giesbrecht, S., Taylor, B.N., Akana, P.R., Butler, A., Pereira, K.A.C., Cooley, S.S., Lau, V.M., Lauterbach, E.L., 2023. Terrestrial Phosphorus Cycling: Responses to Climatic Change. Annu Rev Ecol Evol Syst. https://doi.org/10.1146/annurev-ecolsys-110421

Montes, C.R., Lucas, Y., Pereira, O.J.R., Achard, R., Grimaldi, M., Melfi, A.J., 2011. Deep plant-derived carbon storage in Amazonian podzols. Biogeosciences 8, 113–120. https://doi.org/10.5194/BG-8-113-2011

Nehring, R., 2022. The Brazilian Green Revolution. Polit Geogr 95, 102574. https://doi.org/10.1016/J.POLGEO.2021.102574

Nwoke, O.C., Vanlauwe, B., Diels, J., Sanginga, N., Osonubi, O., Merckx, R., 2003. Assessment of labile phosphorus fractions and adsorption characteristics in relation to soil properties of West African savanna soils. Agric Ecosyst Environ 100, 285–294. https://doi.org/10.1016/S0167-8809(03)00186-5

Owens, P.N., 2020. Soil erosion and sediment dynamics in the Anthropocene: a review of human impacts during a period of rapid global environmental change. J Soils Sediments 20, 4115–4143. https://doi.org/10.1007/S11368-020-02815-9/FIGURES/13

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

Pavinato, P.S., Cherubin, M.R., Soltangheisi, A., Rocha, G.C., Chadwick, D.R., Jones, D.L., 2020. Revealing soil legacy phosphorus to promote sustainable agriculture in Brazil. Sci Rep 10, 1–11. https://doi.org/10.1038/s41598-020-72302-1

Pavinato, P.S., Rocha, G.C., Cherubin, M.R., Harris, I., Jones, D.L., Withers, P.J.A., 2020. Map of total phosphorus content in native soils of Brazil. Sci Agric 78, e20200077. https://doi.org/10.1590/1678-992X-2020-0077

Pérez, A., Machado, W., Sanders, C.J., 2021. Anthropogenic and environmental influences on nutrient accumulation in mangrove sediments. Mar Pollut Bull 165, 112174. https://doi.org/10.1016/J.MARPOLBUL.2021.112174

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

Riffel, S.B., Vasconcelos, P.M., Carmo, I.O., Farley, K.A., 2016. Goethite (U–Th)/He geochronology and precipitation mechanisms during weathering of basalts. Chem Geol 446, 18–32. https://doi.org/10.1016/J.CHEMGEO.2016.03.033

Rodrigues, M., Pavinato, P.S., Withers, P.J.A., Teles, A.P.B., Herrera, W.F.B., 2016. Legacy phosphorus and no tillage agriculture in tropical oxisols of the Brazilian savanna. Science of The Total Environment 542, 1050–1061. https://doi.org/10.1016/J.SCITOTENV.2015.08.118

Rosemarin, A., Ekane, N., 2016. The governance gap surrounding phosphorus. Nutr Cycl Agroecosyst 104, 265–279. https://doi.org/10.1007/S10705-015-9747-9/TABLES/2

Rosas, J.T.F., Demattê, J.A.M., Rosin, N.A., Poppiel, R.R., Silvero, N.E.Q., Amorim, M.T.A., Rodriguez-Albarracin, H.S., 2023. The Brazilian soils clay fraction oxides and weathering index spatialized by Earth Observation Strategy. European Journal of Soil Science. In revision.

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

Roy, E.D., Richards, P.D., Martinelli, L.A., Coletta, L. Della, Lins, S.R.M., Vazquez, F.F., Willig, E., Spera, S.A., VanWey, L.K., Porder, S., 2016. The phosphorus cost of agricultural intensification in the tropics. Nature Plants 2, 1–6. https://doi.org/10.1038/nplants.2016.43

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., 2021. Leveraging the application of Earth observation data for mapping cropland soils in Brazil. Geoderma 396, 115042. https://doi.org/10.1016/J.GEODERMA.2021.115042

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

Safanelli, J.L., Poppiel, R.R., Chimelo Ruiz, L.F., Bonfatti, B.R., de Oliveira Mello, F.A., Rizzo, R., Demattê, J.A.M., 2020. Terrain Analysis in Google Earth Engine: A Method Adapted for High-Performance Global-Scale Analysis. ISPRS International Journal of Geo-Information 9, 400. https://doi.org/10.3390/IJGI9060400

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

Samuel-Rosa, A., Heuvelink, G.B.M., Vasques, G.M., Anjos, L.H.C., 2015. Do more detailed environmental covariates deliver more accurate soil maps? Geoderma 243–244, 214–227. https://doi.org/10.1016/J.GEODERMA.2014.12.017

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/ASSET/AC60214A047.FP.PNG_V03

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

Schaefer, C. E.G.R., do Amaral, E.F., Mendonça, B.A.F., Oliveira, H., Lani, J.L., Costa, L.M., Fernandes Filho, E.I., 2008. Soil and vegetation carbon stocks in Brazilian Western Amazonia: Relationships and ecological implications for natural landscapes. Environ Monit Assess 140, 279–289. https://doi.org/10.1007/S10661-007-9866-0/METRICS

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

Schipanski, M.E., Bennett, E.M., 2021. The Phosphorus Cycle, in: Fundamentals of Ecosystem Science, Second Edition. Academic Press, pp. 189–213. https://doi.org/10.1016/B978-0-12-812762-9.00009-5

Schröder, J.J., Smit, A.L., Cordell, D., Rosemarin, A., 2011. Improved phosphorus use efficiency in agriculture: A key requirement for its sustainable use. Chemosphere 84, 822–831. https://doi.org/10.1016/J.CHEMOSPHERE.2011.01.065

Shary, P.A., 2023. Environmental Variables in Predictive Soil Mapping: A Review. Eurasian Soil Science 56, 247–259. https://doi.org/10.1134/S1064229322602384/FIGURES/6

Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., Homayouni, S., 2020. Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review. IEEE J Sel Top Appl Earth Obs Remote Sens 13, 6308–6325. https://doi.org/10.1109/JSTARS.2020.3026724

Siedliska, A., Baranowski, P., Pastuszka-Woźniak, J., Zubik, M., Krzyszczak, J., 2021. Identification of plant leaf phosphorus content at different growth stages based on hyperspectral reflectance. BMC Plant Biol 21, 1–17. https://doi.org/10.1186/S12870-020-02807-4/TABLES/5

Simpson, R.J., Oberson, Astrid, Culvenor, Richard A, Ryan, Megan H, Veneklaas, Erik J, Lambers, Hans, Lynch, Jonathan P, Ryan, Peter R, Delhaize, Emmanuel, Andrew Smith, F., Smith, Sally E, Harvey, Paul R, Richardson, Alan E, Wissuwa J Simpson, M.R., Culvenor, R A, Richardson, A E, Oberson, A, Ryan, M H, Veneklaas, E J, Lambers, H, Lynch, J P, Ryan, P R, Delhaize, E, Smith, F.A., Smith, S E, Harvey, P R, 2011. Strategies and agronomic interventions to improve the phosphorus-use efficiency of farming systems. Plant and Soil 349, 89–120. https://doi.org/10.1007/S11104-011-0880-1

Smeck, N.E., 1985. Phosphorus dynamics in soils and landscapes. Geoderma 36, 185–199. https://doi.org/10.1016/0016-7061(85)90001-1

Soterroni, A.C., Ramos, F.M., Mosnier, A., Fargione, J., Andrade, P.R., Baumgarten, L., Pirker, J., Obersteiner, M., Kraxner, F., Câmara, G., Carvalho, A.X.Y., Polasky, S., 2019. Expanding the soy moratorium to Brazil's Cerrado. Sci Adv 5, 7336–7353. https://doi.org/10.1126/SCIADV.AAV7336/SUPPL_FILE/AAV7336_SM.PDF

Spohn, M., 2020. Increasing the organic carbon stocks in mineral soils sequesters large amounts of phosphorus. Glob Chang Biol 26, 4169–4177. https://doi.org/10.1111/GCB.15154

Steffen, W., Richardson, K., Rockström, J., Cornell, S.E., Fetzer, I., Bennett, E.M., Biggs, R., Carpenter, S.R., De Vries, W., De Wit, C.A., Folke, C., Gerten, D., Heinke, J., Mace, G.M., Persson, L.M.,

Ramanathan, V., Reyers, B., Sörlin, S., 2015. Planetary boundaries: Guiding human development on a changing planet. Science (1979) 347. https://doi.org/10.1126/SCIENCE.1259855/SUPPL_FILE/STEFFEN-SM.PDF

Stoop, W.A., 1980. Ion adsorption mechanisms in oxidic soils; implications for point of zero charge determinations. Geoderma 23, 303–314. https://doi.org/10.1016/0016-7061(80)90070-1

Tayebi, M., Rosas, J.T.F., Mendes, W. de S., Poppiel, R.R., Ostovari, Y., Ruiz, L.F.C., Dos Santos, N.V., Cerri, C.E.P., Silva, S.H.G., Curi, N., Silvero, N.E.Q., Demattê, J.A.M., 2021. Drivers of Organic Carbon Stocks in Different LULC History and along Soil Depth for a 30 Years Image Time Series. Remote Sensing 13, 2223. https://doi.org/10.3390/RS13112223

Teixeira, P.C., Donagemma, G.K., Fontana, A., Teixeira, W.G., 2017. Manual de métodos de análise de solo, 3rd ed. Embrapa Solos - Livro técnico (INFOTECA-E), Brazilia.

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

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

Van Den Hoogen, J., Robmann, N., Routh, D., Lauber, T., Van Tiel, N., Danylo, O., Crowther, T.W., 2021. A geospatial mapping pipeline for ecologists. https://doi.org/10.1101/2021.07.07.451145

van Raij, B., Quaggio, J.A., da Silva, N.M., 1986. Extraction of phosphorus, potassium, calcium, and magnesium from soils by an ion-exchange resin procedure. Commun Soil Sci Plant Anal 17, 547–566. https://doi.org/10.1080/00103628609367733

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

Viscarra Rossel, R.A., Bui, E.N., 2016. A new detailed map of total phosphorus stocks in Australian soil. Science of the Total Environment 542, 1040–1049. https://doi.org/10.1016/J.SCITOTENV.2015.09.119

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

Wiesmeier, M., Urbanski, L., Hobley, E., Lang, B., von Lützow, M., Marin-Spiotta, E., van Wesemael, B., Rabot, E., Ließ, M., Garcia-Franco, N., Wollschläger, U., Vogel, H.J., Kögel-Knabner, I., 2019. Soil organic carbon storage as a key function of soils - A review of drivers and indicators at various scales. Geoderma 333, 149–162. https://doi.org/10.1016/J.GEODERMA.2018.07.026

Withers, P.J.A., Rodrigues, M., Soltangheisi, A., De Carvalho, T.S., Guilherme, L.R.G., Benites, V.D.M., Gatiboni, L.C., De Sousa, D.M.G., Nunes, R.D.S., Rosolem, C.A., Andreote, F.D., Oliveira, A. De, Coutinho, E.L.M., Pavinato, P.S., 2018. Transitions to sustainable management of phosphorus in Brazilian agriculture. Scientific Reports 8, 1–13. https://doi.org/10.1038/s41598-018-20887-z

Withers, P.J.A., van Dijk, K.C., Neset, T.S.S., Nesme, T., Oenema, O., Rubæk, G.H., Schoumans, O.F., Smit, B., Pellerin, S., 2015. Stewardship to tackle global phosphorus inefficiency: The case of Europe. Ambio 44, 193–206. https://doi.org/10.1007/S13280-014-0614-8/FIGURES/5

Zhu, F., Qu, L., Hong, X., Sun, X., 2011. Isolation and Characterization of a Phosphate-Solubilizing Halophilic Bacterium Kushneria sp. YCWA18 from Daqiao Saltern on the Coast of Yellow Sea of China. Evid Based Complement Alternat Med 2011. https://doi.org/10.1155/2011/615032

Zinn, Y.L., Lal, R., Resck, D.V.S., 2005. Changes in soil organic carbon stocks under agriculture in Brazil. Soil Tillage Res 84, 28–40. https://doi.org/10.1016/J.STILL.2004.08.007

Žížala, D., Minařík, R., Skála, J., Beitlerová, H., Juřicová, A., Reyes Rojas, J., Penížek, V., Zádorová, T., 2022. High-resolution agriculture soil property maps from digital soil mapping methods, Czech Republic. Catena (Amst) 212, 106024. https://doi.org/10.1016/J.CATENA.2022.106024

APPENDIX

Table A1. Performance evaluation of the prediction of soil attributes used as covariates in the mapping of phosphorus stocks in Brazil.

Atributes	RMSE_train	RPIQ_train	R2_train	RMSE_test	RPIQ_test	R2_test
Fe2O3	43.88	1.97	0.69	47.07	1.85	0.65
Al2O3	42.79	2.36	0.59	49.85	2.04	0.44
SiO2	46.34	1.99	0.41	47.98	1.91	0.36
Clay	105.48	2.32	0.67	114.52	2.15	0.61
Sand	156.25	2.40	0.62	170.65	2.19	0.55
Silt	92.83	1.21	0.34	100.24	1.14	0.23
SOC	6.66	1.66	0.56	6.93	1.60	0.52
pH_H2O	0.51	1.36	0.37	0.53	1.30	0.32

4 FINAL REMARKS

The absence of detailed maps describing the spatial behavior of phosphorus stocks in Brazil hinders the implementation of strategies to manage the use of this element on farms and at the regional level. In this thesis, we used geotechnologies to map, at a spatial resolution of 30m, the total and available stocks of phosphorus for plants in the surface layer of the soil (0-20 cm) throughout the Brazilian territory.

In the first chapter, we carried out the mapping of soil oxides, such as Fe_2O_3, Al_2O_3, and SiO_2, which have a direct relationship with phosphorus stocks. These maps were incorporated into the set of environmental variables used to map phosphorus stocks in the second chapter.

The main conclusions of the study are:

1) The combination of advanced digital soil mapping techniques, including machine learning, remote sensing, geoprocessing, and the use of legacy databases, allowed the mapping of phosphorus stocks of Brazilian soils with high spatial resolution, aligned with the understanding of Brazilian soils.

2) The synthetic soil image (SYSI) contributed as a proxy to map soil attributes that have a strong relationship with P stocks, and these attributes could be used as environmental covariates to predict the stocks. Among the attributes, the best performance in modeling was observed for Fe_2O_3, with RMSE = 47 $g.kg^{-1}$, RPIQ = 1.85, and R^2 = 0.65.

3) The strategy of dividing Brazil into two sub-regions, Anthropogenic Vegetation Cover (AVC) and Native Vegetation Cover (NVC), allowed the use of covariates that were present only in AVC areas resulting in better performances in mapping these regions.

4) Despite low determination coefficients, ranging from 0.27 for AP stocks in NVC areas to 0.35 for TP stocks in AVC areas, the obtained maps align with our understanding of Brazilian soils.

5) The highest TP stocks were observed in regions with soils formed from basaltic rocks and soils with higher SOC accumulation. The highest AP stocks were observed in soils with lower P retention, in arid regions of Brazil and the Pantanal biome, and in agricultural areas where P is added as fertilizer.

6) In total, P stocks in Brazil were 531 Mt for TP and 17.4 Mt for AP. About 54% of TP stocks and 39% of AP stocks are in the Amazon biome, despite its soils not having the highest average values of these stocks.

7) The Caatinga biome has the highest average AP stocks, with 2.51 g.m-2, and the Atlantic Forest biome holds the highest average TP stocks with 73.8 g.m-2.

8) AP stocks showed correlation with the productivity of the main Brazilian agricultural crops, varying from 0.2 for coffee to 0.46 for soybeans, indicating the existence of vulnerability if global P reserves were to be depleted.

9) The unprecedented maps generated in this thesis have the potential to shed light on researchers, policymakers, and farmers in the quest for sustainable phosphorus management in Brazil.