University of São Paulo
"Luiz de Queiroz" College of Agriculture

Remote sensing technologies for digital soil mapping: applications for agriculture and environmental planning

Fellipe Alcantara de Oliveira Mello

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

Piracicaba
2022

Fellipe Alcantara de Oliveira Mello
Geographer

Remote sensing technologies for digital soil mapping: applications for agriculture and environmental planning

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

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

Piracicaba
2022

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

Mello, Fellipe Alcantara de Oliveira

Remote sensing technologies for digital soil mapping: applications for agriculture and environmental planning / Fellipe Alcantara de Oliveira Mello. - - Piracicaba, 2022.

105 p.

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

1. Mapeamento digital de solos 2. Sensoriamento remoto 3. Redes de drenagem 4. Solos hidromórficos I. Título

DEDICATÓRIA

À Ana, Nilton, Ivanilza e Luccas

Por todo suporte, incentivo e apoio incondicional

Dedico.

AGRADECIMENTOS

Primeiramente agradeço a Deus pelas incontáveis bençãos que recebo em minha vida todos os dias. A fé depositada em Jesus Cristo me trouxe para o desafio da pós-graduação e me sustentou durante estes cinco anos e meio. Sem esta certeza não seria possível superar os obstáculos e por isso continuo dedicando minha fé naquele que fortalece.

Todos os dias deste doutorado foram compartilhados com minha querida esposa Ana. Os sorrisos, lágrimas e ansiedades eram primeiro recebidos por ela. Sem seu suporte e apoio essa conquista não seria possível e, por isso, sou profundamente grato.

Por meu pai Nilton que abriu mão de finais de semana para investir em meus estudos e sempre acreditou no meu potencial. A minha mãe Ivanilza que chorou quando saí de casa, mas que se orgulha muito dessa conquista. Sem vocês eu não estaria aqui. Ao meu irmão Luccas, agradeço por entender essa mudança, deixar de jogar basquete nos finais de semana e me apoiar nas lutas e dificuldades. Não seria o mesmo sem minha família.

Agradeço a meu sogro e sogra, Nivaldo e Hilda, por me receberem como um filho desde 2017 e pela amizade e confiança. Também a Pedro, Sara e Taís por toda amizade e parceria durante uma fase tão importante em nossas vidas (casamentos, mudanças, sonhos...).

Aos professores da ESALQ/USP de quem recebi ensinamentos e conhecimento incalculáveis. Ao Departamento de Ciência do Solo e do PPG de Solos e Nutrição de Plantas por toda estrutura e apoio necessários.

Em especial ao professor e orientador José Alexandre Melo Demattê, que me recebeu para o mestrado e confiou em meu trabalho para o doutorado. A oportunidade e o crescimento pessoal e profissional adquiridos neste período são muito maiores do que poderia imaginar. Obrigado por todo suporte e parceria ao longo destes anos.

Aos professores Tiago Osório Ferreira, Benito Roberto Bonfatti e Eliângela Benedet da Silva por todas as sugestões e correções na qualificação. Ao professor Pablo Vidal-Torrado e sua disciplina de Gênese e Classificação de Solos, que foi o diferencial para meu trabalho como pesquisador em pedologia. Ao professor Igo Fernando Lepsch, com o qual tive o prazer e honra de conviver e trabalhar no levantamento de solos.

Aos queridos colegas de trabalho do Geocis, com os quais aprendi e convivi durante este período: Henrique, Rodnei, Diego, Danilo, Clécia, Raul, Nélida, André, José Lucas, Karina, Wanderson, Ariane, Benito, Ruiz, Arnaldo, Natasha, Jorge, Nicolas, Salman, Mah, Soledad, Andrés, Lucas Greschuk, Cláudia, Gabriel, Maria Eduarda, Merylin, Fábio e aos demais que convivi, obrigado pelo trabalho em equipe e dedicação que torna o grupo Geocis forte e relevante no cenário internacional.

Finalmente agradeço a FAPESP pelo financiamento do grupo de pesquisa (Número do processo 2014/22262-0). O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Programa de Excelência Acadêmica (PROEX) - Brasil (CAPES) – Código de Financiamento 001.

Meu Muito Obrigado!

CONTENTS

RESUMO

Sensoriamento remoto e mapeamento digital de solos: aplicações para agricultura e planejamento ambiental

A presente tese de doutorado desenvolveu pesquisa sobre o mapeamento digital de solos (MDS) e seu potencial para o aprimoramento dos levantamentos de solo a partir de informações digitais. No primeiro capítulo há uma introdução geral sobre a pesquisa desenvolvida, uma contextualização geral sobre a importância do tema e suas aplicações para o avanço da ciência do solo. O segundo capítulo apresenta a inserção de novas covariáveis para o MDS relacionadas a redes de drenagem (RD). O trabalho combinou informações de satélite, relevo e RD para mapear os teores de argila, areia e carbono orgânico do solo para o município de Piracicaba no estado de São Paulo, Brasil. O terceiro capítulo aplica a metodologia do MDS para realizar o refinamento (aumento de escala) de um mapa geológico para a região de Pereira Barreto no estado de São Paulo, Brasil. O trabalho mostrou que mesmo com a reflectância espectral da superfície do solo é possível fazer relações e análises sobre as características da subsuperície. No quarto capítulo, foi aplicada a metodologia de detecção de solo exposto (imagem sintética de solo exposto – SySI) para a identificação e classificação de solos com ocorrência de hidromorfismo. A reflectância nas faixas espectrais do visível, infra-vermelho próximo e infra-vermelho de ondas curtas (Vis-NIR-SWIR) do solo exposto foi capaz de identificar feições relacionadas aos solos hidromórficos, as quais foram utilizadas para classificar estes solos para uma região de 863,577.9 km² entre o sudeste e centro-oeste do Brasil. Os resultados indicaram alta ocorrencia de hidromorfismo em áreas agrícolas, indicando a necessidade de maior monitoramento. O ultimo capítulo apresenta uma conclusão geral do trabalho com os principais resultados e suas aplicações na ciência do solo.

Palavras-chave: Mapeamento digital de solos, Modelagem espacial; Sensoriamento remoto, Redes de drenagem, Material de origem do solo, Solos hidromórficos

ABSTRACT

Remote sensing technologies for digital soil mapping: applications for agriculture and environmental planning

This doctorate thesis developed advanced research regarding digital soil mapping (DSM) and its potential for improving soil surveys using multiple and complex digital information. For the first chapter, it is presented a general introduction to the developed research, a general contextualization of the topic's importance and its applications for the advancement of soil science. The second chapter presents the insertion of new covariates for the DSM related to drainage networks (DN). The work combined satellite, relief, and DN information to map clay, sand, and soil organic carbon contents for the municipality of Piracicaba in the state of São Paulo, Brazil. The third chapter applies the DSM methodology to refine (enhance the scale) a geological map for the Pereira Barreto region in the state of São Paulo, Brazil. The work showed that even with the soil surface spectra, it is possible to analyze the characteristics of the subsurface. In the fourth chapter, the bare soil detection methodology (Synthetic Soil Image – SySI) was applied to identify and classify soils with hydromorphism. The reflectance in the visible, near-infrared and short-wave infrared (Vis-NIR-SWIR) spectral bands of the exposed soil was able to identify features related to hydromorphic soils, which were used to predict hydromorphic soils for a region of 863,577.9 km² between the Southeast and Midwest of Brazil. The results showed a high occurrence of hydromorphism in agricultural areas, indicating the need for greater monitoring. The last chapter presents a general conclusion of the work with the main results and their applications for soil science.

Keywords: Digital soil mapping, Spatial modelling, Remote sensing, Drainage network, Soil parent material, Hydromorphic soils

1. GENERAL INTRODUCTION

Agriculture has been a major target in order to provide food security to the growing world's population (Godfray et al., 2010; Wollenberg et al., 2016). The modernization of agro-industrial activities and the application of technology for management improved food production, as well as the ability to ensure a sustainable development (Wu and Clark, 2016). The advances in agriculture are at the core of the 17 Sustainable Development Goals presented by the United Nations in 2015, supporting initiatives that increase productivity while minimizing the impact on the environment (Ericksen et al., 2009; Lal and Stewart, 2010). Among the multiple efforts worldwide to ensure food security, soil conservation is held as a key factor, since soil is the source for plant growth and development and it provides the required information to achieve a productive agriculture with sustainable practices (McBratney et al., 2014). However, inadequate soil use result in soil degradation, erosion, contamination, low productivity, and other environmental issues (Brown and Brown, 2018; Oldeman, 1997).

Although soil is fundamental to achieve food security, soil data is often scarce and with low scale for an adequate planning (Miller, 2012). National and state inventories provide general soil databases and legacy soil maps in many countries, which allowed the development and expansion of agricultural activities. Until the end of the 1990s, most countries did not have a map of soil classes at an appropriate scale for agricultural activities. Brazil, for instance, basically relied on data derived from the RADAM Brazil project at a scale of 1:1,000,000 (Radambrasil, 1973). Other economically developed countries such as Australia and The Netherlands also did not have detailed maps of even their agricultural regions (McBratney et al., 2003).

However, the cost of soil survey is a limitation for the acquisition of new and detailed information about soil characteristics (Ma et al., 2019). Field incursions, laboratory analysis, and the work of a soil scientist require time and financial support, which increases proportionally with the scale of the survey (Sanchez et al., 2009). One of the solutions to assess more soil information at a lower cost was to combine the legacy soil information with geographic information systems (GIS), allowing the spatialization of multiple soil information for regions, watersheds, and whole countries (Minasny and McBratney, 2016). The advance of computational processing and the availability of geospatial information (i.e. satellite data) helped advancing soil studies in a more quantitative form, finding the relationship between soil information and the environment in a technique recognized as digital soil mapping (DSM) (McBratney et al., 2003).

DSM is defined as the creation and maintenance of spatial soil information systems through numerical models, in order to infer spatiotemporal variations of soil types and properties, using soil observation and knowledge of related environmental variables (Philippe Lagacherie et al., 2007). This was possible thanks to the advancement of technology, especially in the 1990s, when computers more suitable for personal use and with greater data processing capability emerged. In addition to computers, other technologies such as the global positioning system (GPS), satellite imagery and terrain information derived from digital elevation models have become more available and accessible to researchers (Sanchez et al., 2009). To understand and use large soil databases new statistical tools such as data mining and machine learning algorithms were applied to soil science. This made the use of spatial information easier and enabled new applications for spatial ground information.

The basis for the DSM is defined according to the soil forming factors presented by (Jenny, 1941), where Soil = f(c,o,r,p,t), (c) climate; (o) organisms; (r) relief; (p) source material; (t) time. The soil function is a mechanistic model to guide the understanding of factors that may have contributed to the formation of a given soil (Searle et al., 2021). The formation factors would act as environmental variables (independent), together with the spatial position, to predict a certain soil variable (dependent), through a mathematical model in which uncertainty would also be

quantified (Ma et al., 2019). The environmental variables used in the prediction models are diverse, but must be related to the soil formation process. If not, the variable should not be used to predict soil properties. Such variables must be used in raster format at an adequate spatial resolution (pixel size). In a GIS, the values of the environmental variables will be assigned to the ground points, so that a mathematical model can be calibrated. The created model will be applied on the environmental variables, generating a continuous digital soil map in the landscape. After generating the digital map, a second set of real soil information will be used to validate the map. Therefore, DSM studies emerged combining conventional soil survey and theories (i.e. fieldwork, laboratory analysis, pedological maps), digital information (i.e. satellite reflectance, topographic attributes, climate data), and statistical modelling (i.e. linear regression, decision trees, neural networks) (Arruda et al., 2016). The S variable (soil) can be used in two ways, being continuous (soil attributes) and categorical (soil classes). Continuous variables are more associated with soil attributes, such as texture, organic matter, and soil depth. On the other hand, categorical variables are associated with soil classes. Therefore, when mapping attributes, linear and non-linear regressions will be used, while classification models or supervised classification will be used for mapping classes. Thus, the DSM is consolidated as a low-cost, fast, and powerful tool to access soil information in areas with scarcity of data, which can help in agricultural activities and land use planning.

Therefore, this thesis was guided by the need to explore geospatial tools to improve soil mapping. The DSM methods were applied for different purposes with the help of multiple information related to the soil forming factors. The second chapter used a set of complex drainage network variables to perform digital maps of clay, sand, and soil organic carbon (SOC), for a 1378 km2 site in the São Paulo state, Brazil. The relationship between the drainage density (DD), drainage frequency (DF), channel sinuosity, and confluence angle with soil classes and attributes were analyzed. The third chapter explored a multi temporal Landsat image composition with bare soil reflectance to extract soil properties and distinguish discrepant lithological classes at the western plateau, São Paulo State, Brazil. The area is 247,737 ha large, where 981 soil samples were collected at 0–20 cm depth. We performed a DSM procedure to generate maps of attributes related to parent material. Finally, the fourth chapter evaluated and analyzed multitemporal bare soil image reflectance at various locations to verify the occurrence of hydromorphism. We used this information to predict hydromorphic soils for a large area and analyzed their distribution by federal states and land use/cover types. The main results and findings are compiled in the last chapter of this thesis.

References

Arruda, G.P. de, Demattê, J.A.M., Chagas, C. da S., Fiorio, P.R., Souza, A.B. e, Fongaro, C.T., 2016. Digital soil mapping using reference area and artificial neural networks. Sci. Agric. 73, 266–273. https://doi.org/10.1590/0103-9016-2015-0131

Brown, L.R., Brown, L.R., 2018. World Population Growth , Soil Erosion , and Food Security. Science (80-.). 214, 995–1002.

Ericksen, P.J., Ingram, J.S.I., Liverman, D.M., 2009. Food security and global environmental change: emerging challenges. Environ. Sci. Policy 12, 373–377. https://doi.org/10.1016/j.envsci.2009.04.007

Godfray, H.C.J., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., Muir, J.F., Pretty, J., Robinson, S., Thomas, S.M., Toulmin, C., 2010. Food Security: The Challenge of Feeding 9 Billion People. Science (80-.). 327, 812 LP – 818. https://doi.org/10.1126/science.1185383

Jenny, H., 1994. Factors of soil formation: a system of quantitative pedology. Courier Corporation.

Lal, R., Stewart, B.A., 2010. Food Security and Soil Quality, 1st ed. CRC Press, Boca Raton, FL. https://doi.org/https://doi.org/10.1201/EBK1439800577

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

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., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping, Geoderma. https://doi.org/10.1016/S0016-7061(03)00223-4

Miller, B.A., 2012. The Need to Continue Improving Soil Survey Maps. Soil Horizons 53, 11. https://doi.org/10.2136/sh12-02-0005

Minasny, B., McBratney, A.B., 2016. Digital soil mapping: A brief history and some lessons. Geoderma 264, 301–311. https://doi.org/10.1016/j.geoderma.2015.07.017

Oldeman, L.., 1997. Soil Degradation: A Threat to Food Security? Paper presented at the International Conference on Time Ecology: Time for Soil Culture – Temporal Perspectives on Sustainable Use of Soil.

Philippe Lagacherie, M.E., A.B., M., Voltz, 2007. Digital Soil Map An Introductory Perspective.

Radambrasil, P., 1973. Levantamento de recursos naturais. Ministério das Minas e Energia, Departamento Nacional da Produção Mineral ….

Sanchez, P.A., Ahamed, S., Carré, F., Hartemink, A.E., Hempel, J., Huising, J., Lagacherie, P., McBratney, A.B., McKenzie, N.J., Mendonça-Santos, M. de L., Minasny, B., Montanarella, L., Okoth, P., Palm, C.A., Sachs, J.D., Shepherd, K.D., Vågen, T.-G., Vanlauwe, B., Walsh, M.G., Winowiecki, L.A., Zhang, G.-L., 2009. Digital Soil Map of the World. Science (80-.). 325, 680 LP – 681. https://doi.org/10.1126/science.1175084

Searle, R., McBratney, A., Grundy, M., Kidd, D., Malone, B., Arrouays, D., Stockman, U., Zund, P., Wilson, P., Wilford, J., Van Gool, D., Triantafilis, J., Thomas, M., Stower, L., Slater, B., Robinson, N., Ringrose-Voase, A., Padarian, J., Payne, J., Orton, T., Odgers, N., O'Brien, L., Minasny, B., Bennett, J.M., Liddicoat, C., Jones, E., Holmes, K., Harms, B., Gray, J., Bui, E., Andrews, K., 2021. Digital soil mapping and assessment for Australia and beyond: A propitious future. Geoderma Reg. 24, e00359. https://doi.org/https://doi.org/10.1016/j.geodrs.2021.e00359

Wollenberg, E., Vermeulen, S.J., Girvetz, E., Loboguerrero, A.M., Ramirez-Villegas, J., 2016. Reducing risks to food security from climate change. Glob. Food Sec. 11, 34–43. https://doi.org/10.1016/j.gfs.2016.06.002

Wu, H., Clark, H., 2016. The sustainable development goals: 17 goals to transform our world, in: Furthering the Work of the United Nations. United Nations, pp. 36–54. https://doi.org/10.18356/69725e5a-en

2. COMPLEX HYDROLOGICAL KNOWLEDGE TO SUPPORT DIGITAL SOIL MAPPING

Abstract

Drainage network (DN) is the representation of all the stream channels developed over the landscape. The morphometry of DN describes the relationship between channel characteristics and basin geometry, which is regulated by a series of processes, such as weathering, geomorphology, sediment erosion/deposition. The interaction between these factors impacts soil formation, resulting in a relationship between DN morphometry and soil characteristics. Unstable surfaces produced shallow soils, enhancing surface runoff and high drainage density, while stable landscapes favor vertical infiltration through old and weathered soils. Digital soil mapping has benefited from multiple environmental variables, such as relief and satellite data, but DN information can offer great contributions for the prediction of soil attributes. In this work, we applied a set of complex DN variables to perform digital maps of clay, sand, and soil organic carbon (SOC), for a 1,378 km² site in the São Paulo state, Brazil. We analyzed the relationship between the drainage density (DD), drainage frequency (DF), channel sinuosity, and confluence angle with soil classes and attributes. The Cubist and Random Forest algorithms were tested to predict the soil information, and to evaluate the impact of the new hydrological variables. The results showed that landscapes predominated by clayey soils favor surface runoff and increase channel formation, with higher channel sinuosity and acute confluence angles. These new drainage variables contributed 35 to 55% for SOC prediction. DD and DF were the most important drainage variables on the models, ranging from 65 to 70%. The external validation reached R^2 of 0.72 and 0.56 for the prediction of clay and sand, respectively. The impact of DN information on the model performance suggests that more work is needed to better explore and understand the relationship between DN and soil information.

Keywords: Hydrological variables, Digital soil mapping, Soil-landscape relationship, Confluence angle, Channel sinuosity

Published as: Mello, F.A.O., Demattê, J.A.M., Rizzo, R., Mello, D.C. de, Poppiel, R.R., Silvero, N.E.Q., Safanelli, J.L., Bellinaso, H., Bonfatti, B.R., Gomez, A.M.R., Sousa, G.P.B., 2022. Complex hydrological knowledge to support digital soil mapping. Geoderma 409, 115638. https://doi.org/https://doi.org/10.1016/j.geoderma.2021.115638

2.1. Introduction

Drainage network (DN) is the representation of all the stream channels developed over the landscape (Charlton, 2008). The channels start close to the topographical drainage division, flowing down slope by the lowest part of a valley (thalweg) inside the watershed, forming junctions with tributary channels (Thorp et al., 2010). The morphometry of DN describes the relationship between channel characteristics and basin geometry (Gray, 1961; Moussa, 2003), which is regulated by a series of processes, such as weathering, geomorphology, sediment erosion/deposition. These forming factors determine the confluence angle, channel width, sinuosity, and length of the DN (Clubb et al., 2016; Strahler and Strahler, 1989). In addition, watershed morphometry studies are important to assess hydrological potential (Barroso et al., 2014), flood risks (Romshoo et al., 2012), landscape interpretation (Ray, 1960; Way, 1973), and environmental conservation (Patel et al., 2013).

The interaction between natural factors, such as geomorphology, lithology, soil, precipitation rate, and temperature, determines the local channel morphometry (Fowler et al., 2007; Strahler, 1957, 1952). Crave and Gascuel-Odoux (1997) pointed to the role of pedogeomorphology on the spatial variability of soil surface water dynamics. The water infiltrates more at gentle slopes combined with porous soil, while steep slopes stimulate the surface runoff and channel formation. Way (1973) explained that sites with impermeable rocks and soils with low infiltration rates result in a denser DN. Young and shallow soils tend to occur at steep slopes, where the surface runoff predominates, slowing the weathering activity and pedogenesis (Bonifacio et al., 1997; Schaetzl and Anderson, 2005). When derived from mafic and ultramafic parent materials, these soils present high clay content, which decreases water infiltration and promotes surface channel formation (Ker et al., 2015; Breemen and Buurman, 2002).

Channel meandering has been attributed to different natural factors such as Earth's rotation, gentle slope, flow energy, changes in river stage, bank erosion, sediment overload, and deposition (Chatley, 1938; Chitale, 1973; Eakin, 1910; Friedkin, 1945; Lacey, 1923; Lane, 1957; Quraishy, 1943; Russell, 1936). Langbein and Leopold (1966) related channel sinuosity to the stability of the landscape, where tortuous rivers are allocated at stable valleys formed after long periods of erosion. Yang (1971) concluded that a smooth sinuous meandering course is the only course that a natural unbraided stable channel can take. The braided shape begins as a straight channel with shoals in alternate sides of the stream, producing bank erosion, which is a result of the variations of water discharge, sediment concentration, channel slope and geometry, valley slope, and geological constraints (Ackers and Charlton, 1970; Yang, 1971). These classical works started the findings on channel morphometry, allowing new discoveries about hydrology and watershed management.

The channels of a drainage basin are connected through junctions, where two segments meet and form another one. This principle is the basis of hierarchical river ordering established by Strahler (1957). These river channel confluences are of interest to geomorphologists, sedimentologists, and engineers due to its connection with water flow and sedimentation (Best, 1986). In the past, studies about river channel confluences brought attention to alteration in hydraulic geometry (Richards, 1980), riverbed morphology and depositional sites (Alam et al., 1985; Best, 1988), and also the effect on flow mixing and pollutant dispersal (Bonakdari et al., 2011; Gaudet and Roy, 1995; Biron et al., 2004).

Despite DN being useful to understand soil formation, studies relating drainage morphometry and soils did not advance and were restricted to old descriptive analysis. On the other hand, terrain modeling was heavily used in digital soil mapping (DSM) (McBratney et al., 2003). In particular, most studies explored new terrain variables and the effect of scale on improving DSM accuracy (Hartemink et al., 2008; Miller and Schaetzl, 2016). Afterwards, some works explored the connection between DN morphometry and soils, with the results pointing to the differences between DN over soils of different texture and mineralogy, but there was no continuity on using drainage information on DSM (Jung et al., 2015, 2011; Vasques et al., 2015). Mello et al. (2021) used drainage information to extrapolate a soil class map for a larger area, which had good results but only used regular DN variables.

Therefore, our hypothesis is that there is a spatial correlation between soil attributes, soil classes, and DN, which can improve DSM. The main objective of this research was to use complex DN morphometric information coupled with other covariates (i.e., satellite and relief data) using machine learning algorithms for topsoil attributes mapping. We believe this approach can improve DSM and provide insights for other similar research.

2.2. Material and Methods

2.2.1. Study area

This study was conducted in the Piracicaba municipality in the São Paulo State, southeast Brazil (Fig. 1). The area has 1,378 km² and is located in a humid subtropical climate with humid summer and dry winter (Cfa) (Alvares et al., 2013). The annual precipitation rate varies from 1100 to 1450 mm. The region is part of a transitional zone between savanna and tropical forest vegetation, which was extensively changed for sugarcane and other crop productions (Barreto et al., 2006).

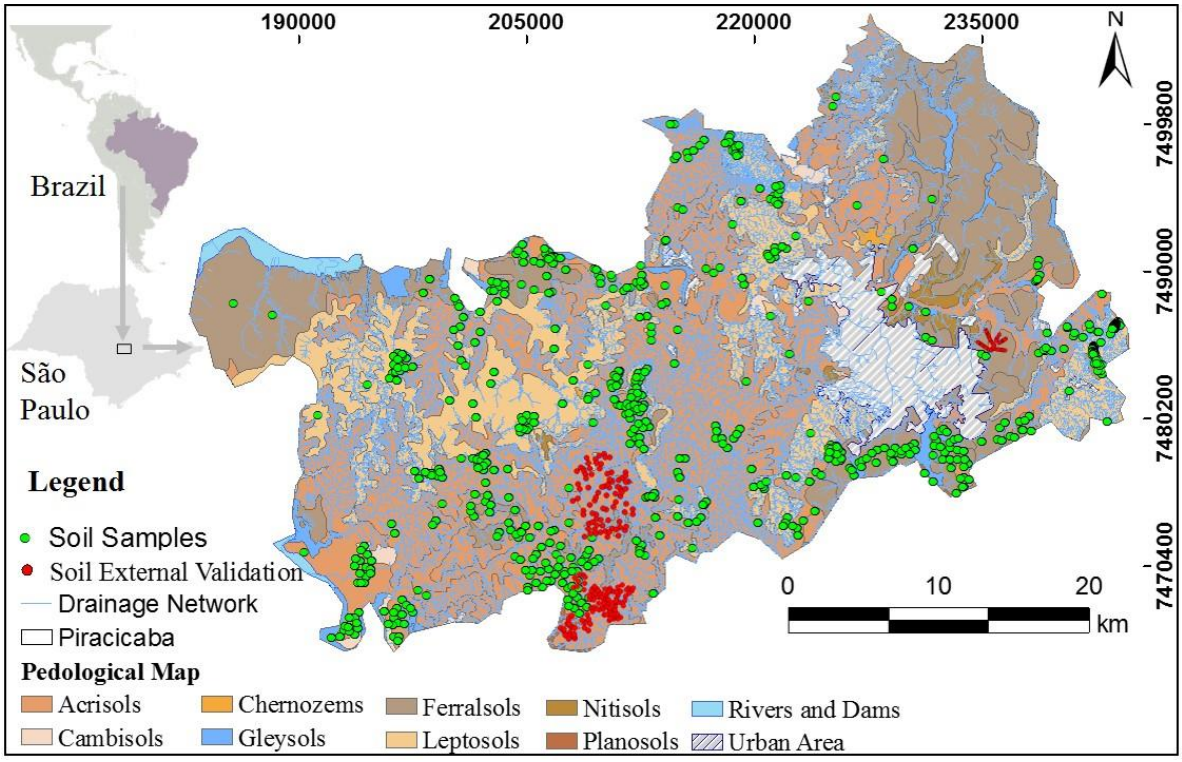

Figure 1. Maps of the study area with focus on the drainage network. a. Drainage network database derived from aerial photographs, soil class map developed from the agronomical institute of Campinas – SP (IAC) and adapted by Rossi (2017), and the 1748 soil samples collected at the area.

The study area is located in the Paulista Peripheral Depression (PPD), a complex geological region with igneous, metamorphic, and sedimentary rocks, in a weathered environment that formed a smooth relief (Vidal-Torrado and Lepsch, 1999). The Piracicaba river is the main channel bed of the PPD, with an average flow rate of 144 m³/s, which is supplied from a complex group of tributary channels and sub-basins (Barreto et al., 2006). The landscape is covered by different soil types, varying from young (e.g. Cambisols, Regosols) to old (e.g. Ferralsols, Nitisols) soils (IUSS Working Group, 2015). This complexity allowed the development of this work, which combined detailed environmental information to evaluate the soil impact on DN morphometry.

2.2.2. Soil data

The soil information used in this work was provided by the Geotechnologies in Soil Science Group (GeoCiS), which has carried out a series of soil studies in the Piracicaba region (Demattê et al., 2019). We used 1717 soil locations distributed around the study area that were sampled at 0-20 cm depth with an auger and prepared for physical and chemical laboratory analysis (Fig. 1). We also used an additional soil dataset with 318 soil samples from the GeoCiS, in order to perform an external validation on the predicted maps. The external soil dataset was distributed in agricultural areas mainly at the southern part of the study area, which was enough to represent soil distribution and validate the maps.

We selected sand, clay, and soil organic carbon (SOC) contents to use as dependent variables for spatial predictions. The textural information was used due to its relationship with the relief (Schaetzl and Anderson, 2005; Thomas et al., 1999), soil reflectance (Poppiel et al., 2019a; Viscarra Rossel et al., 2006), and DN (Demattê and Demétrio, 1998; França and Demattê, 1990). The SOC information was selected based on its relationship with soil

water content (Haynes and Swift, 1990; Hoogsteen et al., 2015), which is intrinsic to DN patterns (Pallard et al., 2009; Ritchie, 1981).

The samples were oven-dried for 48h at 50 °C, ground and sieved through a 2 mm mesh (EMBRAPA, 2011). The densimeter method was performed to analyze the soil particle size distribution using sodium hydroxide (0.1 mol L−1) and sodium hexametaphosphate (0.1 mol L−1) as dispersing agents (Camargo et al., 1986). The SOC content was determined using the Walkley–Black method (Teixeira et al., 2017).

2.2.3. Environmental information

2.2.3.1. Hydrological data

The DN information used in this work was obtained manually from 1 m orthorectified aerial images from the Paulista Company of Metropolitan Planning (EMPLASA). The images composed of 8-bit RGB photographs were captured using digital cameras attached to an aircraft with 30% of lateral overlapping, which allowed the stereoscopic effect for the photogrammetric analysis.

Table 1. Environmental variables used as predictors for DSM of soil attributes. The data is separated in drainage, relief and remote sensing, regarding the source of each variable. These variables were calculated in 30m of spatial resolution.

Class	Attribute	Description	Unit	Reference
Drainage	ANG	Confluence Angle	Degree	(Best, 1986)
	SIN	Flow Sinuosity	non-dimensional	(Langbein and Leopold, 1966)
	DD	Drainage Density	km/km²	(Strahler, 1952)
	DF	Drainage Frequency	non-dimensional	(Strahler, 1952)
	CNBL	Channel Network Base Level	m	(Bock and Köthe, 2008)
	VDCN	Vertical Distance to Channel Network	m	(Rennó et al., 2008)
Terrain	ASP	Aspect	Degree	(Florinsky, 2012)
	CTA	Catchment Area	m	(Tarboton et al., 1992)
	GCV	General Curvature	Degree	(Florinsky, 2012)
	LSF	LS Factor	non-dimensional	(Conrad et al., 2015)
	PCV	Profile Curvature	m^{-1}	(Florinsky, 2012)
	RSP	Relative Slope Position	non-dimensional	(McConkey et al., 1997)
	SLP	Slope	Degree	(Florinsky, 2012)
	TRI	Topographic Ruggedness Index	non-dimensional	(Riley et al., 1999)
	TWI	Topographic Wetness Index	non-dimensional	(Pei et al., 2010)
	VDP	Valley Depth	m	(Florinsky, 2012)
	VTR	Vector Terrain Ruggedness	non-dimensional	(Riley et al., 1999)
Remote Sensing	SySI Band 1	Blue	Reflectance factor	(Demattê et al., 2018)
	SySI Band 2	Green	Reflectance factor	(Demattê et al., 2018)
	SySI Band 3	Red	Reflectance factor	(Demattê et al., 2018)
	SySI Band 4	Near Infrared	Reflectance factor	(Demattê et al., 2018)
	SySI Band 5	Short Wave Infrared - 1	Reflectance factor	(Demattê et al., 2018)
	SySI Band 6	Short Wave Infrared - 2	Reflectance factor	(Demattê et al., 2018)

For the photogrammetric analysis, we used the 1 m stereoscopic images to create a three-dimensional effect using the PHOTOMOD Lite 6.3 software, a proper screen, and glasses. Through 3D visualization of the landscape, we manually mapped intermittent and perennial river channels from the top to lower relief positions, following the lowest part of a valley (thalweg), and exported it as a vector file format for further editing (Fig. 2a). We corrected

topological errors in ESRI ArcGIS 10.4 and transformed the data into raster format with a pixel size of 30 m for further analysis.

The DN shapefile was used to calculate a set of hydrological attributes with 30m of spatial resolution (Table 1). We separated the variables obtained from DN information in order to evaluate its performance on predicting soil attributes. Most hydrological attributes were calculated in the SAGA GIS Channels Library (Conrad et al., 2015) (Table 1). However, the drainage density (DD) and drainage frequency (DF) were calculated in ESRI ArcGIS 10.4, using the line density tool. The DD and DF were calculated according to Strahler (1952), respectively as follows:

$$DD = \frac{\mathrm{L}}{\mathrm{A}}$$

where, DD: drainage density, L: total channel length (km), and A: area surrounding the channels (km^2)

$$DF = \frac{\mathrm{n}}{\mathrm{A}}$$

where, DF: drainage frequency. n: number of channels, A: area surrounding the channels (km^2).

We also calculated the flow sinuosity (SIN) of the channels in order to evaluate its relationship with landscape attributes, previously stated in other studies (Langbein and Leopold, 1966; Parvis, 1950; Ray, 1960). The SIN was calculated for each river segment despite its length according to Charlton (2008), as follows:

$$SIN = \frac{\mathrm{CL}}{\mathrm{VL}}$$

where, SIN: channel sinuosity, CL: channel length, and VL: valley length.

To calculate the SIN in ESRI ArcGIS 10.4 we performed the following steps. We established the fluvial order for the channels according to Strahler (1957). We considered as a river segment any continuous channel that remained in the same fluvial order. When the order changed, a new river segment started (Fig. 2b). Then, we calculated the UTM coordinates at the start and end of each segment and computed the horizontal distances in the X and Y axis (Fig. 2b). The X and Y distances represented the opposite and adjacent sides of a triangle, where the hypotenuse, which is the valley length between the channel source and mouth, was calculated by the Pythagoras theorem, as follows:

$$VL = \sqrt{(\mathrm{Xs} - \mathrm{Xm})^2 + (\mathrm{Ys} - \mathrm{Ym})^2}$$

where, Xs: Channel source X coordinate, Xm: Channel mouth X coordinate, Ys: Channel source Y coordinate, and Ym: Channel mouth Y coordinate

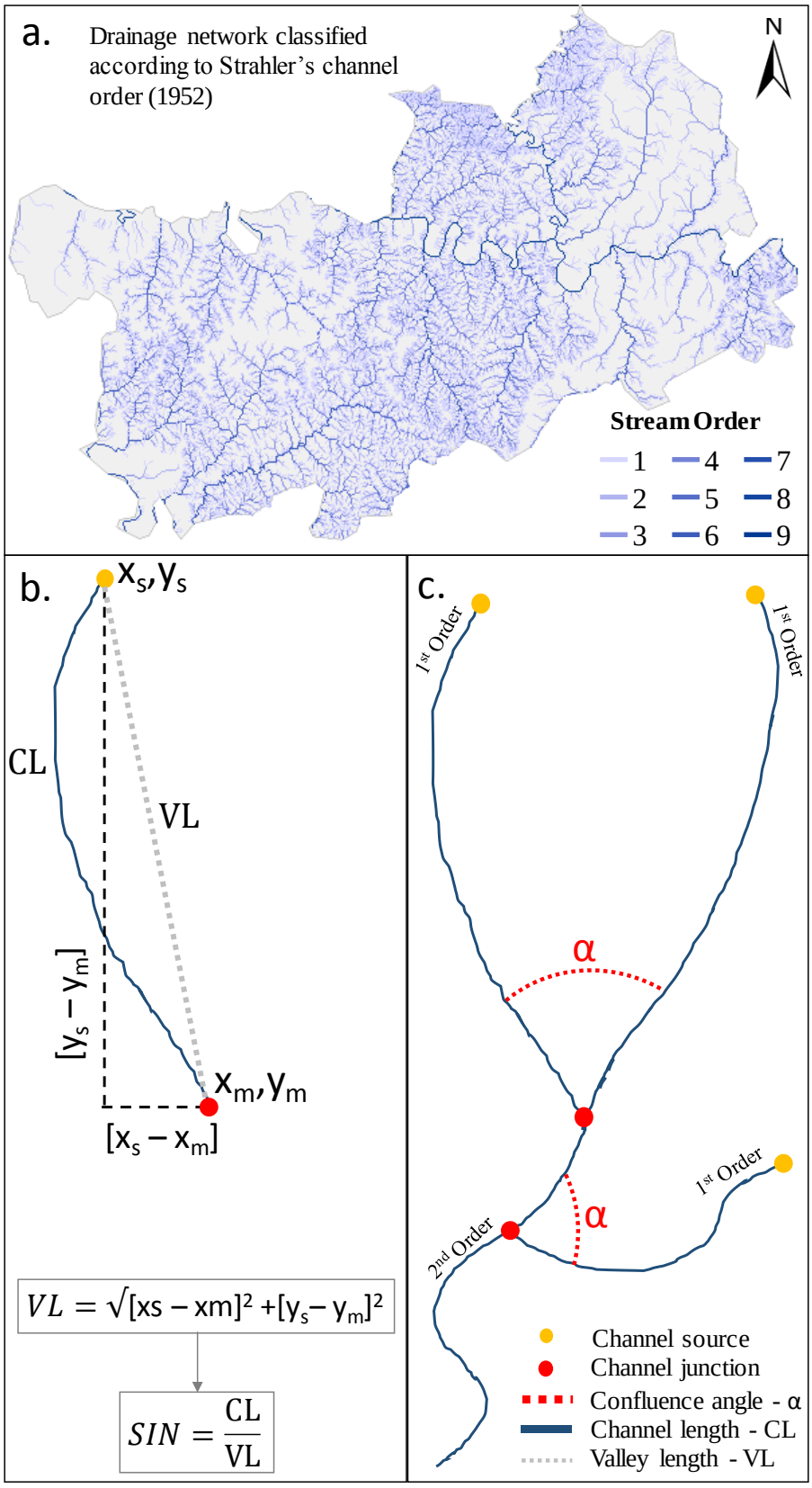

Figure 2. Graphical explanation of the calculation of a. Stream ordering following Strahler (1952), b. channel sinuosity and c. confluence angle. Xs: Channel source X coordinate; Xm: Channel mouth X coordinate; Ys: Channel source Y coordinate; Ym: Channel mouth Y coordinate; CL: Channel length; VL: Valley length; SIN: Channel sinuosity.

Based on the channel's segmentation, we also calculated the confluence angle at every junction between two segments (Fig. 2c). The confluence angle is defined as the internal angle between two channels, according to the flow direction of the channels (Best, 1986; Biron et al., 2004). We used the arcpy tool in ESRI ArcGIS 10.4, to calculate the angle at the channel junction, where the value of the angle was stored (Fig. 2c).

In order to have a continuous surface with the confluence angle and the channel sinuosity information for predicting soil attributes in the study area, we interpolated them by geostatistics. We used the values of confluence stored at each location as input data, while for the sinuosity we obtained the value to interpolate from a point generated in the middle of each channel. We employed the inverse distance weighted (IDW) method due to the assumption that things that are close to one another are more alike than those that are farther apart. This function is available with the geostatistical analyst package in ESRI ArcGIS 10.4. We defined the power function as 2, which is the rate at which the weights decrease as distance increases. For the search neighborhood parameter, we used the default value of 12.

The premise to interpolate these hydrological parameters is that both attributes describe the surrounding landscape configuration. According to classical studies, the confluence angle and the channel sinuosity are dependent on the landscape, soil texture and lithology (Best, 1988; Demattê and Demétrio, 1998; Jung et al., 2011; Way, 1973). Therefore, we aimed to test these variables along with a set of other environmental variables to model soil attributes.

2.2.3.2. Relief attributes

We used contour lines with 5 m equidistance, acquired from planialtimetric maps at 1:10,000 scale, obtained from the Geographical and Cartographic Institute of the São Paulo State (IGC). These lines were used to interpolate a 30 x 30m digital elevation model (DEM) using the Topo to Raster function in ESRI ArcGIS 10.4. The resulting DEM was used to calculate a set of relief attributes in the SAGA GIS Terrain Analysis Library (Conrad et al., 2015) (Table 1).

2.2.3.3. Bare soil image composite

The soil surface patterns are related to subsurface variations and processes that occur within the soil profile, specific to each soil class (Dotto et al., 2020; Minasny and McBratney, 2008). Thus, we used the Geospatial Soil Sensing System (GEOS3) (Demattê et al., 2018) to a time-series of Landsat images using the Google Earth Engine (GEE) platform (Gorelick et al., 2017). We used the Landsat 4 Thematic Mapper (TM) (1982–1993), Landsat 5 TM (1984–2012), Landsat 7 Enhanced Thematic Mapper Plus (ETM +) (1999–2018) and the Landsat 8 Operational Land Manager (OLI) (2013–2018). These methods used only Tier 1 or surface reflectance data that were processed by the LEDAPS (Landsat 4, 5, and 7) (U.S.G.S., 2019a) and LASRC (Landsat 8) algorithms (U.S.G.S., 2019b).

The GEOS3 algorithm extracts soil information from a time series of Landsat images and aggregates the bare soil pixels into a single synthetic soil image (SySI), which is the reflectance image of the bare soil composite (Safanelli et al., 2020). The bare soil pixels were identified on single satellite images through a set of classification rules. The rules were based on spectral indices coupled with quality assessment bands, which removed cloud, cloud shadow, inland water, photosynthetic vegetation, and non-photosynthetic vegetation (crop residues) (Safanelli et al., 2020). The pixels were classified as soil based on the Normalized Difference Vegetation Index (NDVI), with a threshold between – 0.15 and 0.20 to mask out green vegetation, Normalized Burning Ratio (NBR2), with a – 0.15

and 0.15 to mask out crop residues, difference between bands 1 and 2 (B2 – B1) and between bands 2 and 3 (B3 – B2) (Demattê et al., 2020, 2018).

Afterwards, the bare soil pixels were used to calculate, pixel-by-pixel, the median values of topsoil reflectance for single bands and obtain the final reflectance value (Demattê et al., 2020, 2018). The SySI had six spectral bands from blue to short-wave infrared regions at 30 m resolution (Table 1).

2.2.4. Statistical analysis

Before using the database to predict the soil attributes, we performed a sequence of exploratory analysis in order to evaluate the relationship of the environmental variables with the soil information. We also intended to evaluate the results regarding the new hydrological attributes.

We explored the minimum, maximum, mean, median, standard deviation, and skewness values of the soil data in the R software (R Core Team, 2013). We calculated the texture class of the soil according to the USDA soil texture triangle (Soil Science Division Staff, 2017) using the "soiltexture" package in R (Moeys, 2018). Afterwards, we calculated the Spearman's correlation between the soil data and environmental variables including soil, relief, and DN using the "corrplot" package (Wei et al., 2017) in R software (R Core Team, 2013).

We analyzed the variables' variance through a principal component analysis (PCA) using the "factoextra" package (Kassambara, 2017) in the R software (R Core Team, 2013), and displayed the first and second components separated by the textural soil class to explore the variation present in the data set. The PCA is a multivariate analysis that reduces the dimensionality of the dataset by projecting each data point onto only the first few principal components to obtain lower-dimensional data while preserving as much of the data's variation as possible (Bryant and Yarnold, 1995). The PCA is often used for predictive models, and it was used for an exploratory analysis in this work.

Finally, we analyzed the SIN and confluence angle distribution for each soil type. All the information about the channel sinuosity and the confluence angle were separated according to the soil type and were explored through a boxplot graphic. The graphic was performed using the "ggplot" package (Wickham, 2011) in R (R Core Team, 2013). We excluded the soil types that represented less than 1% coverage of the study area, such as Chernozems and Planosols. The data were also submitted to a multi comparison test, in order to confirm if there were statistical differences between the drainage information and the soil types. To reduce the natural spatial dependency, we fitted a linear model using Generalized Least Squares (GLS) (Dormann et al., 2007). Afterwards, we applied an analysis of variance (ANOVA) ($p < 0.01$), measuring the mean value of the hydrological variables for each soil class.

2.2.5. Machine learning techniques

We used random forest (RF) and Cubist machine learning algorithms to predict soil attributes in the study area. Both techniques are common in DSM works, producing good results on the prediction of soil texture (Chagas et al., 2016; Fongaro et al., 2018; Lagacherie et al., 2019; Poppiel et al., 2019b), SOC (John et al., 2021; Li et al., 2021; Moura-Bueno et al., 2021), soil classes (Flynn et al., 2019; Lamichhane et al., 2021; Odgers et al., 2014; Vincent et al., 2018; Zeraatpisheh et al., 2019), and other environmental factors related to soil (Hengl et al., 2018; Styc et al., 2021; Zhang et al., 2021).

2.2.5.1. Soil data splitting (cross-validation)

We applied the k-fold cross-validation (CV) method to the soil data to construct the prediction models using the R package "caret" (Kuhn, 2008). CV is a resampling method originally developed to fix optimistic results of the predictive effectiveness of regression equations when applied to future observations (Mosier, 1951). The procedure consists of randomly dividing the data in k groups, using k -1 groups to fit a model, and one for validation (Browne, 2000; Efron and Tibshirani, 1995). This procedure is repeated k times, always leaving one group out of the calibration dataset (Hawkins et al., 2003). Finally, the result is summarized with the mean of the model skill scores. In this work, we applied the 10-fold cross-validation, which has been used in DSM works and produced good results (Brus et al., 2011; Horst-Heinen et al., 2021; Loiseau et al., 2021).

2.2.5.2. Random Forest and Cubist algorithms

The RF algorithm estimates a user-specified number of decision trees by randomly sampling an existing dataset (Breiman, 2001). However, at each node construction, a random sample of the dependent variables is used. The resulting decision tree is used to estimate the error rate by predicting the value of the remaining unsampled data and comparing with the known results (Gambill et al., 2016).

Cubist is a rule-based algorithm that creates a tree structure from a larger dataset of variables (Quinlan, 1992). The trees are produced by the algorithm through rules that use boosting training (Quinlan, 1993). The boosting training consists in giving more weight to stronger learners, so the weaker learners are also converted into strong learners (Khaledian and Miller, 2020). The final model is regulated by a set of nodes along the tree and two hyperparameters to reduce the prediction error. We used root mean square error (RMSE), mean absolute error (MAE), and r squared (R^2) to evaluate the RF and Cubist models. For Cubist, the variables' importance was also analyzed through the number of times the variables were used in the model.

2.2.5.3. Random Forest and Cubist algorithms

In order to evaluate the influence of the DN variables on soil predictions, we tested three dataset combinations to construct the models. The first one test combined the SySI bands with the terrain variables, the second used DN and terrain variables, and the third combined all the variables.

The prediction performance of the models was accessed by the CV using the whole soil data. We decided not to split the data into validation and calibration sets, instead we used the available data to fit the model using CV and used separated field samples to perform an external validation. We used the number of randomly selected predictors (mtry) to select the optimal RF model. The mtry regulates the number of variables that can be randomly sampled in each split of the trees, which resulted in 2, 14, and 27. We used 500 trees for stable variable estimates. We used two hyperparameters (committees and neighbors), to reduce the prediction error of Cubist models.

2.2.5.4. Spatial Predictions and External validation

After testing the models and evaluating their performance, we predicted the clay, sand, and SOC contents for the study area. We decided to present all the maps predicted by RF and Cubist models. However, we only analyzed the ones that had better performance.

Despite having the results from the k-fold cross-validation, we performed an external validation to the predicted maps using a soil dataset of 1015 samples to provide more confidence to the results. External validations have proved to be useful to evaluate the performance of a predicted map (Grinand et al., 2008) and have contributed for the evaluation of digital maps produced in the same region of this work (Bonfatti et al., 2020; Mello et al., 2021). This procedure is important on the evaluation of the soil attribute values regarding the DN morphometry, which is the focus of the work. We provide an analysis of the results based on the soil-landscape relationship, interpreting the predicted values according to the relief and DN patterns.

Finally, we presented an analysis of the DN variables across the landscape, and its impact on soil attribute prediction based on the rules of soil-landscape relationship (Milne, 1935; Ruhe, 1960). We explored the variability of some hydrological attributes in a toposequence inside the Marins watershed and related it with the predicted maps.

2.2 Results

2.3.1. Soil attribute statistics

The spatial distribution of the 1717 soil samples covered the major soil types of the study area (Fig. 1), and provided a representative soil data set. Overall, soils had higher sand contents at the 0-20 cm layer, a typical characteristic of the region (Table 2). Minimum and maximum values of sand ranged from 1 to 975 g kg^{-1}, being considerably higher than clay contents, which ranged from 0 to791 g kg^{-1} (Table 2). The average values of both attributes also showed discrepancy, with clay content being 245 g kg^{-1} and sand was 577 g kg^{-1} (Table 2).

Table 2. Descriptive statistics of clay, sand, and soil organic carbon (SOC) contents of the soil samples used on the digital soil mapping.

Attribute	n	Min.	Max.	Mean	Median	SD	Skewness
				g kg^{-1}			
Clay	1717	7	791	245	198	168.0	0.98
Sand		1	975	577	646	264.0	-0.32
SOC		0	28.4	5.04	4.06	4.18	1.34

The SOC ranged from 0 to 28.4 g kg^{-1}, with an average of 5.04 g kg^{-1} and standard deviation of 4.18 g kg^{-1} (Table 2). The maximum SOC value is considered high for both sandy and clayey soils, but the mean value was low, which is expected from sandy soils. Therefore, the descriptive statistics suggest a region with predominance of sand at the 0-20cm soil horizons, with low SOC contents.

2.3.2. Soil and environmental variables relationship

The soil samples were also used to plot a textural triangle, which corroborated with the descriptive statistics (Fig. 3a). The triangle indicates a predominance of sandy samples, having strong presence at the range of 70% to 100%, showing the predominance of sandy soils at the study area (Fig. 3a). There is also a significant presence of clayey and medium texture samples, where the levels of clay reached 80%, representing the clayey soils. This soil textural variation is determinant in a DSM study, since it allows the evaluation of patterns between environmental variables and soil attributes. We considered silt contents for the textual triangle analysis, but decided not to use it as a

main attribute in this work, since silt content in the Piracicaba region is notoriously low (Vidal-Torrado and Lepsch, 1999).

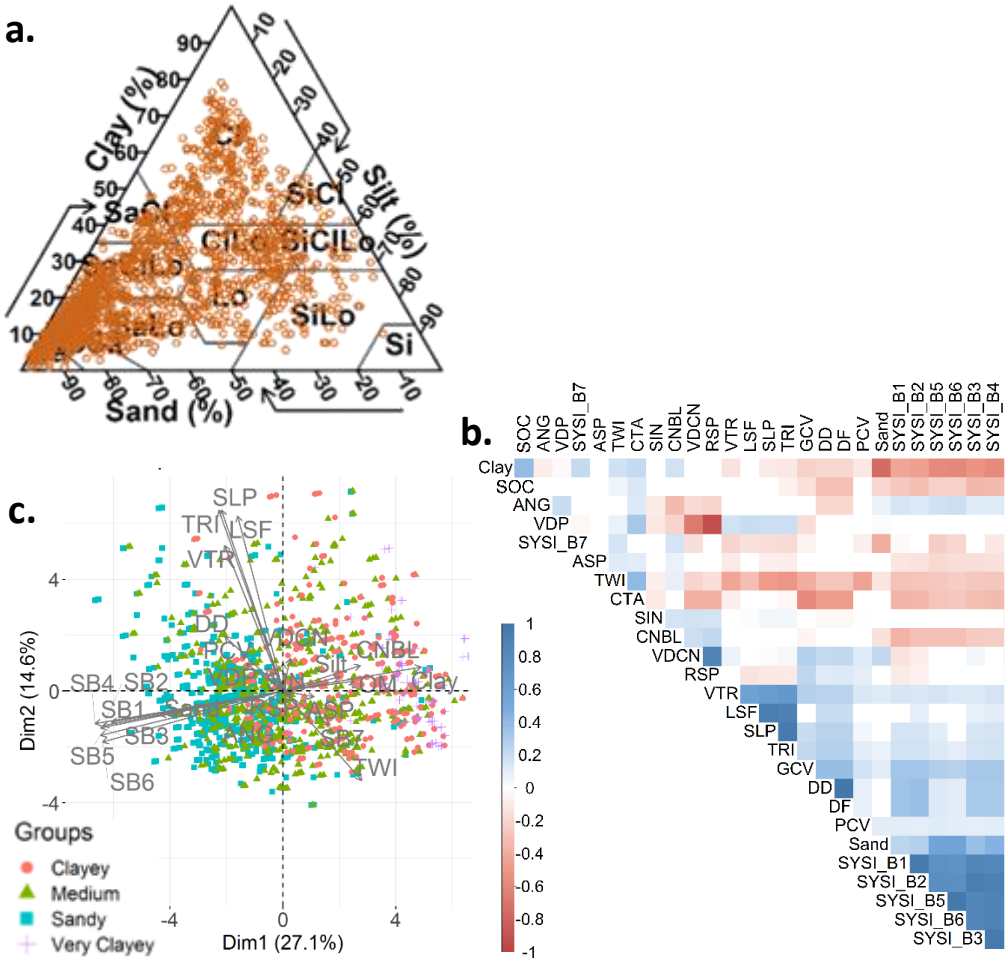

Figure 3. Three exploratory analysis of the database. a. textural triangle showing the predominance of sandy texture. The textural classes are according to the USDA parameters (Soil Science Division Staff, 2017). b. correlation matrix based on the Spearman's coefficient. c. principal component analysis. Biplot of first and second PC's: vectors representing the variables and the points are classified according to its textural class.

In Figure 3b we present a correlation matrix with all the variables used in this work. The drainage related variables had similar correlation coefficients when combined with the soil attributes (Fig. 3b). Overall, the correlation was not significant in most cases, but it was possible to gain insights regarding the new drainage variables. The channel sinuosity had a positive correlation with the three soil attributes, while the confluence angle presented a negative correlation with clay and OM. The DD and DF had a similar correlation (-0.31) with SOC and were the best correlated drainage variables. The traditional DN variables, CNBL and VDCN, had opposite results regarding the soil attributes and the highest result was of VDCN with sand (0.23).

The correlation between terrain and soil attributes was similar to the DN variables (Fig. 3b). Overall, they had a negative correlation with clay and OM, which is expected due to the natural accumulation of SOC in clayey soils (Hassink et al., 1993), and also the positive correlation between these soil attributes (Fig. 3b). Some terrain variables also had good correlation with the DN (Fig. 3b). The DD had a positive correlation with SLP, TRI, and

VTR, indicating the relationship between the relief forms and the channel development. However, the negative correlation between DD and TWI supports the principle that when there are more channels on the surface, the surface runoff is prioritized rather than the vertical infiltration. The confluence angle also had a negative correlation (-0.21) with the RSP, indicating that the slope position is affecting the junction of streams.

The biplot shows the opposite direction of clay and OM, compared to the SySI bands, caused by the negative correlation between these variables (Fig. 3c). The SySI vectors show a strong association with sand, as the vectors point to the same direction and most sandy samples are close to their vectors. In a lower intensity, the confluence angle was also associated with medium to sandy texture. The DD was more associated with the relief variables (SLP, TRI, LSF, VTR), which had low association with SySI, clay, and OM (Fig. 3c).

The soil samples were distributed according to their textural classes, placing the clayey and sandy samples on opposite ends of the plot (Fig. 3c). Although the new DN variables were not representative of the data variance, they were more associated with the soil texture and SOC content than some traditional relief variables (Fig. 3c).

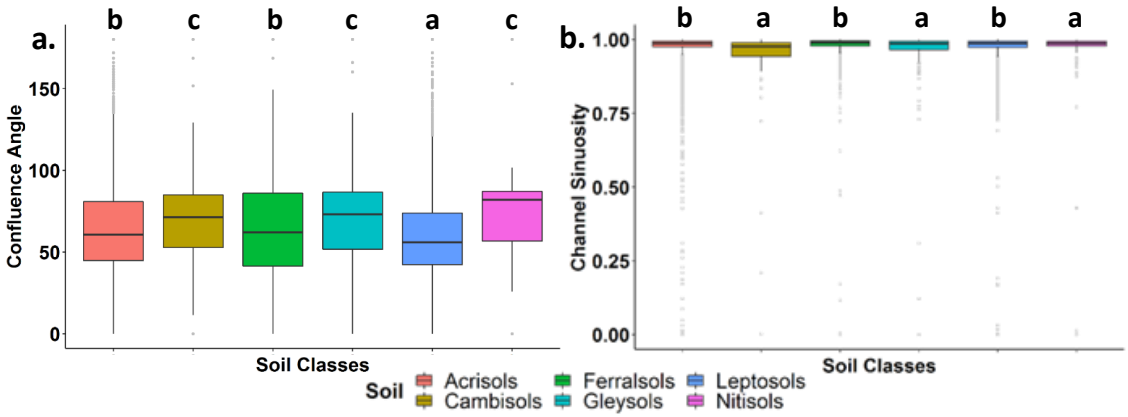

Figure 4. Boxplot and multi comparison test ($p < 0.01$) significance level for soil classes and drainage variables. a. channel confluence angles distribution for each soil class; b. channel sinuosity distribution for each soil class.

Despite not showing high correlation values or a strong association with the soil attributes, the DN variables showed a different pattern for the soil classes. We analyzed the distribution of the confluence angle and the channel sinuosity for each major soil class at the study area. The results showed that soils with similar physical characteristics tend to have similar confluence angle and channel sinuosity (Fig. 4).

For the analysis regarding the confluence angle, we highlight the results of Acrisols, Ferralsols, and Leptosols, which comprise around 86% of the study area. These three soil classes normally have similar textural and physical characteristics in the Piracicaba region and were separated from the other soil classes by the variance analysis (Fig. 4a). These classes showed high presence of observations on the 74th percentile and significant presence of potential outliers at the higher end of the boxplot (Fig. 4a). However, the interquartile range remained lower than the other classes, indicating that more observations had lower angles. The median bar also stayed around 60º, indicating a pattern of acute angles. The Cambisols, Gleysols and Nitisols were grouped into another class (Fig. 4a). These classes had less observations at the 75th percentile and few potential outliers. The median was consistently higher than class "a" and "b", indicating a pattern of higher angles in these soils.

The variance analysis of channel sinuosity had similar results on the soil classes grouping (Fig. 4b). Soils that have low infiltration rates were grouped together in class "a", being Cambisols, Gleysols, and Nitisols (Fig. 4b). The remaining classes were grouped in class "b" and showed more potential outliers at the lower end of the distribution (Fig. 4b). The overall display of the graphic suggests that there is not a great variation of the sinuosity values over the soil classes. This is due to the disproportionate number of first order river channels, which are mostly short and straight, pushing the distribution to higher values (Fig. 4b). However, the variance analysis was capable of identifying differences of sinuosity patterns for different soils.

2.3.3. Model assessments

Table 3 shows the performance of RF and Cubist regression models using 10-fold cross-validation for calibrating sand, clay, and SOC prediction models. The model performance for each dataset combination was similar for both RF and Cubist algorithms (Table 3). The models including SySI had better results than the ones with DN and terrain, for the prediction of sand and clay. However, for all the soil attributes and for both algorithms, the combination of DN and SySI reached the highest prediction performance, with 0.71 for sand and clay, and 0.41 for SOC (Table 3). Cubist prediction models had slightly higher R^2 values than RF for sand and OM, while both algorithms reached equal R^2 values for the clay prediction (Table 3). The RMSE values for Cubist prediction models were lower than RF models for the three soil attributes, with differences ranging from 0.03 to 1.85 g kg^{-1} (Table 3). The MAE values for Cubist models were also slightly lower than RF models with differences ranging from 0.09 to 1.0 g kg^{-1} (Table 3).

Table 3. Performance metrics of the prediction models for soil clay, sand, and soil organic carbon (SOC) contents.

Soil Att.	Dataset[1]	Random Forest				Cubist				
		mtry	RMSE	R^2	MAE	Commit[2]	Neigh[3]	RMSE	R^2	MAE
Clay	RS+Terrain	10	91.9	0.7	68.9	10	5	93.1	0.7	69.4
	DN+Terrain	17	93.73	0.69	70.6	20	5	93.5	0.69	70.8
	All	2	91.62	0.71	69.2	20	9	91.4	0.71	68.2
Sand	RS+Terrain	10	149.8	0.70	99.0	10	5	150.1	0.69	99.4
	DN+Terrain	9	154.9	0.67	104.7	20	5	154	0.68	102.2
	All	24	148.05	0.71	97.1	20	9	146.2	0.72	96.9
SOC	RS+Terrain	2	5.67	0.38	4.0	20	9	5.73	0.36	4.1
	DN+Terrain	9	5.64	0.4	3.9	20	9	5.63	0.39	3.9
	All	2	5.63	0.40	3.9	20	9	5.6	0.41	3.9

[1]Type of data included in the model, where RS: remote sensing, Terrain, DN: drainage network, and All: all variables included; [2]Number of committees used to select the best model; [3] Number of neighbors.

The model performance reached satisfactory levels for the three soil attributes, especially for clay and sand (Table 3). The addition of the hydrological attributes showed good potential for the prediction of soil physical attributes. The prediction of SOC had the lowest values, nevertheless the amount of variation explained by the models were reasonable (Table 3).

2.3.4. Variables' importance

In Figure 5 we presented the performance of each environmental variable on the prediction of soil attributes using RF and Cubist models with cross-validation. Overall, the bands 4 and 6 from SySI and DD had the highest contribution on the models (Fig. 5). Moreover, the SySI and drainage related variables were the most influential on both RF and Cubist models, followed by the relief information.

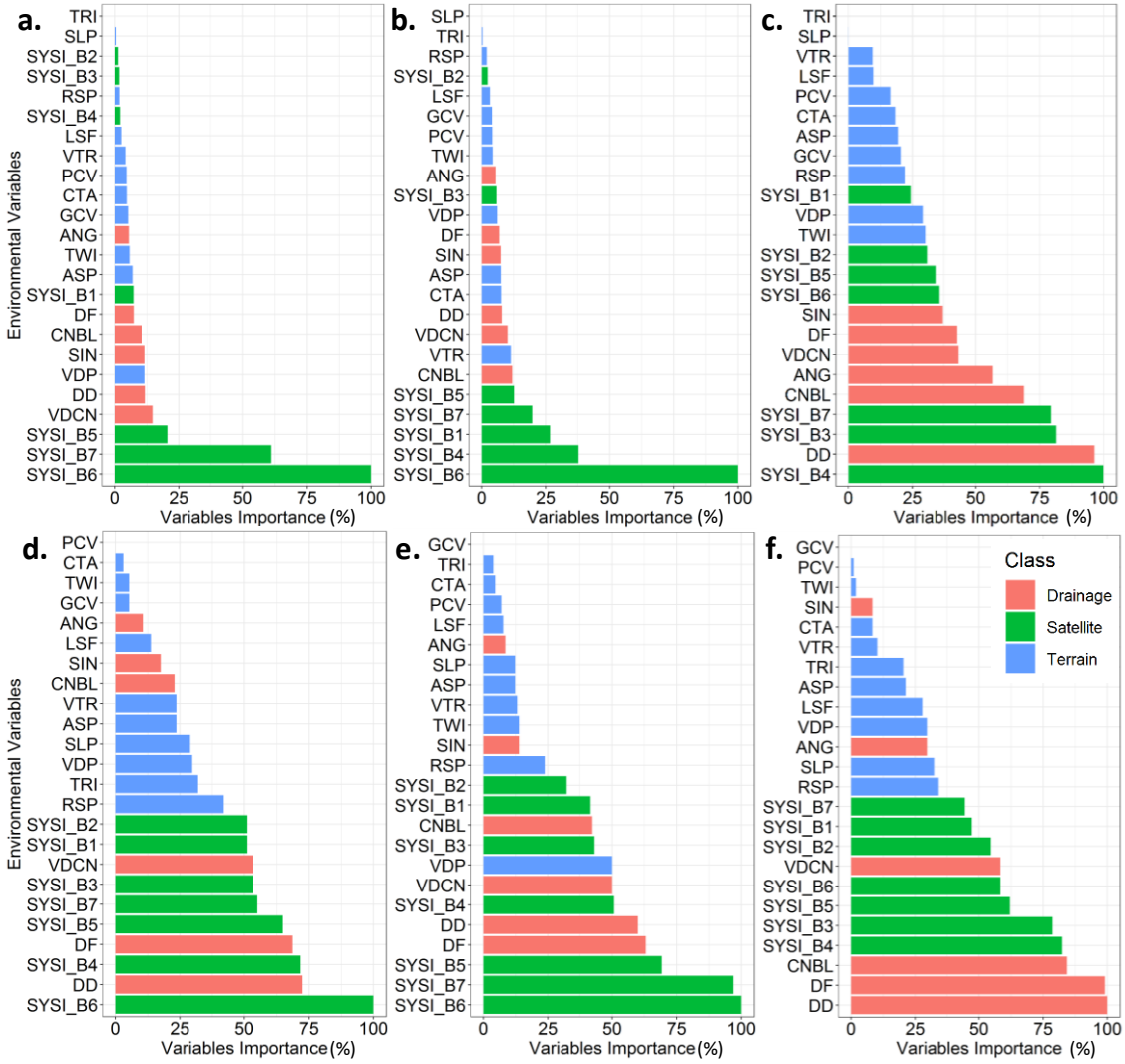

Figure 5. Graphics of Variables' importance for the prediction of sand (a,d), clay (b,e) and soil organic carbon (SOC) (c,f). The first three graphics (a-c) represent the results of random forest (RF) models, and the last three (d-f) represent the results of the cubist models.

For the prediction of sand, we found a different scenario between RF and Cubist models (Fig. 5a,d). For RF, the bands 2, 3, and 4 from SySI had low influence on the model, while for Cubist the same variables had more importance (Fig. 5a,d). Considering the OM predictions, the DN variables, such as DD, DF, and CNBL showed more weight on the model predictions (Fig. 5c,f). The relief attributes, commonly a major predictor in DSM studies, had lower importance on RF and Cubist models of the three attributes (Fig. 5).

Overall, the contribution of new hydrological attributes ranged between 20 and 100% for clay, sand, and OM prediction. The channel sinuosity ranged between 0 and 25% for the prediction of clay and sand (Fig. 5 a,b,d,e), while reaching 35% of importance on the prediction of SOC using RF (Fig. 5c). The confluence angle ranged between 10 and 20% for the prediction of clay and sand, and had a strong influence on SOC predictions using RF, reaching 55% (Fig. 5c). More conventional drainage attributes had higher importance on the models, with DD and DF ranging between 20 and 100%, being the most important attributes for the prediction of SOC using Cubist (Fig 5f). DD and DF had more influence on the Cubist algorithm, ranging from 65 to 70%, while on RF both hydrological attributes ranged from 15 to 10% for the prediction of clay and sand (Fig. 5a, b).

2.3.5. Digital soil maps

Figure 6 displays the predicted maps of sand, clay, and SOC for the study area. The digital maps showed that the western part of the study area presented sandy soils (Fig. 6a,b) with average sand content of 694 g kg^{-1} and clay of 186 g kg^{-1}, while the eastern region showed soils with higher clay content (Fig. 6c,d) with average sand and clay contents of 478 g kg^{-1} and 305 g kg^{-1}, respectively. The maps suggest an inverse distribution of the soil texture at the study area, which is caused by the negative correlation between sand and clay contents (Fig. 3b).

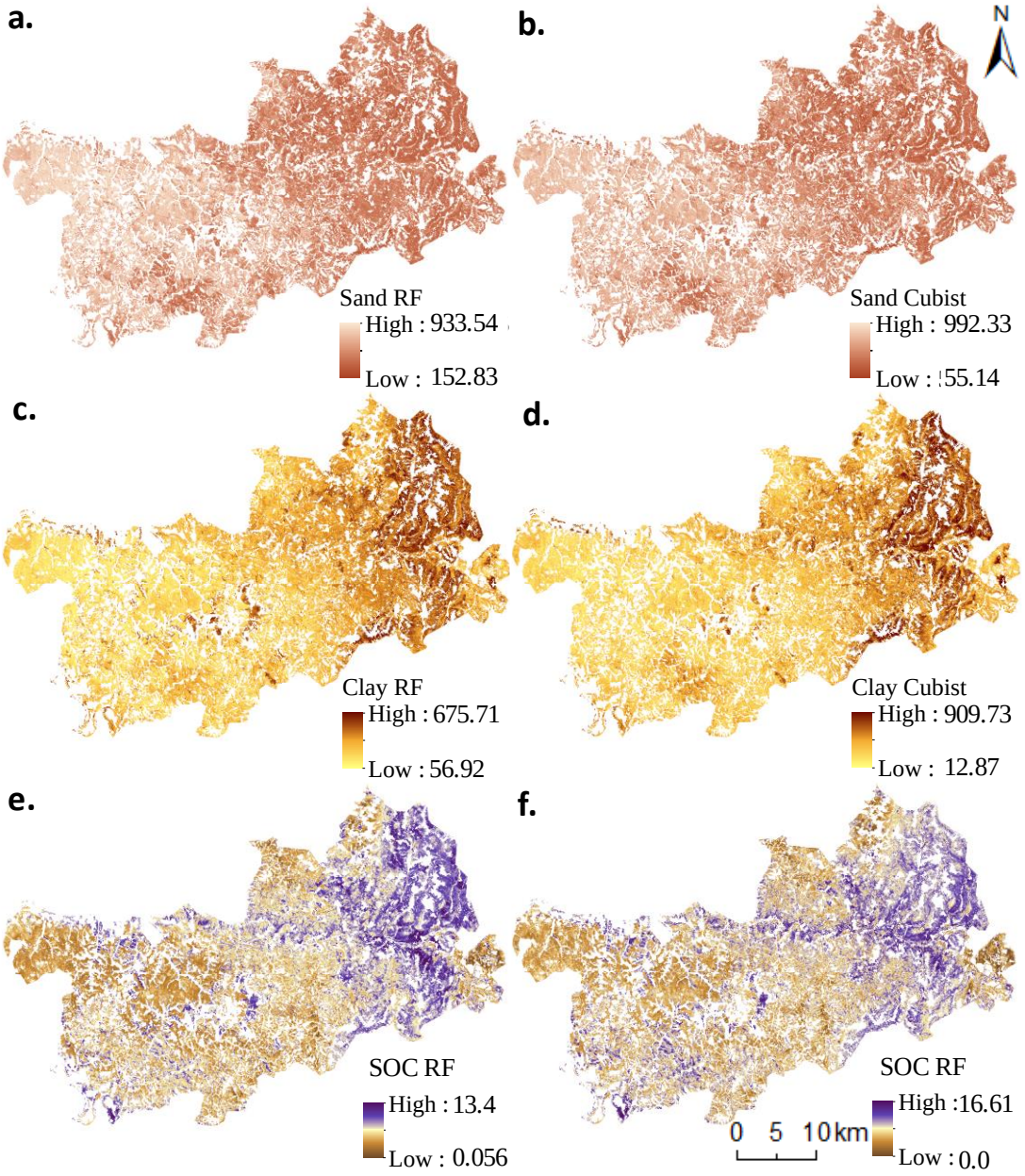

Fig. 6. Resulting maps of the random forest (RF) and cubist modelling techniques. a. sand content predicted through RF; b. sand content predicted through cubist; c. clay content predicted through RF; d. clay content predicted through cubist; e. soil organic carbon (SOC) content predicted through RF; f. SOC content predicted through cubist.

The Cubist algorithm overestimated the maximum values for sand (1244.98 g kg^{-1}) and clay (909.73 g kg^{-1}), while the maximum values found on the soil analysis were 975 g kg^{-1} and 791 g kg^{-1}, respectively (Table 2). The RF model was able to produce maps within the boundaries of the original minimum and maximum values for sand and clay.

The spatial distribution for the SOC content (Fig. 6e,f) was similar to the clay patterns (Fig. 6a,b) due to the positive correlation between these two attributes (Fig. 3b). Both algorithms underestimated the presence of high SOC contents, since the maximum predicted values were 23.05 g kg^{-1} for RF and 28.57 g kg^{-1} for Cubist, while the maximum observed value was 49 g kg^{-1}.

2.3.6. External validation

The external validation confirmed the robustness of the model (Table 4). The R^2 for clay was 0.72 and 0.71, for RF and Cubist respectively, which was similar to the model itself (Table 4). Sand had lower coefficients compared to the model, however, the Cubist map had R^2 of 0.56 and low error rates. This result was determinant to choose Cubist as the most appropriate model for this work. SOC had R^2 of 0.08 and high error rates, indicating that this external validation was not satisfactory to validate SOC prediction (Table 4).

Table 4. External validation performed by the combination of the digital soil maps and 318 extra soil samples.

		Random Forest				Cubist			
	n	R^2	RMSE	RPIQ	bias	R^2	RMSE	RPIQ	bias
Clay	318	0.74	159.7	2.75	-39.2	0.75	152.6	2.98	-45.4
Sand		0.52	217.6	2.16	-66.6	0.64	182.9	2.56	-34.2
SOC		0.2	16.1	1.27	-6.6	0.1	16.4	1.24	-6.6

2.4. Discussion

2.4.1. Effects of environmental variables on soil attribute predictions

The 1717 soil samples were representative, considering the geographical extent of the study area (1,378 km²) (Fig. 1). The required soil sampling to produce accurate digital soil maps has been the object of discussion in many works (Kidd et al., 2015; McBratney et al., 2003; Vincent et al., 2018). Besides there were attempts to produce digital soil maps with limited soil information (Stumpf et al., 2016; Zhang et al., 2016), the large and well distributed soil database available in this work allowed the execution of an adequate soil map (Fig. 1). In addition, using adequate covariates was necessary to accurately predict soil attributes in the study area, which is located in a complex geological and geomorphological region (Barreto et al., 2006). The consideration of DN variables also proved to be worthwhile, since it improved the model in multiple cases (Fig. 6).

The bare soil reflectance (SySI) was determinant to yield a R^2 of 0.71 and 0.72 for clay and sand predictions, respectively (Table 3). The use of synthetic soil images, retrieved from multitemporal satellite collections has been extensively explored recently to understand their potential for soil and environmental assessment (Mzid et al., 2021; Silvero et al., 2021; Vaudour et al., 2021). According to Demattê et al. (2020), bare soil information from SySI offers a proxy for lithological, pedological, carbon pools, and biome changes. However, SySI alone may not be able to identify all soil transitions and complex variations across the landscape, since it does not comprise the landscape altitude variation and other natural features. In this case, it is also necessary to use other information related to soil forming factors, such as relief and DN.

Relief variables had a minor contribution to the model's performance (Fig. 5). Digital terrain models and other landscape information have been a major source for DSM due to their widespread availability for soil mapping. But in this case, DN was more important on the prediction models (Fig. 5). Mello et al. (2021) also found a better performance of drainage related variables compared to relief, for soil class prediction. However, the relief is a soil forming factor which regulates water movement, erosional and depositional processes, and helps the identification of soil transitions, reminiscing the traditional techniques of soil mapping (Bazaglia Filho et al., 2013; Teramoto et al.,

2001). Therefore, the combined use of traditional relief information with new drainage variables can add detail and quality to soil mapping.

2.4.2. Drainage information for digital soil mapping

After employing the confluence angle and the sinuosity values through a statistical analysis (Fig. 4), and observing the weight of DN variables on soil modelling (Fig. 5), it was possible to state that DN information had higher importance than the relief variables on the model prediction. However, DSM works historically prioritized relief data as covariates and did not explore the full potential of DN information (McBratney et al., 2003; Minasny and McBratney, 2016).

During the 1990's, few works were conducted in order to understand the relationship between DN and soil (Demattê and Demétrio, 1995; França and Demattê, 1990). Demattê and Demétrio (1998) found DD of 8.4 and 7.65 km/km² in areas with Nitisols and Cambisols, and DD of 0.86 and 1.14 km/km² in Ferralsols areas, indicating the role of clay content on channel formation. In Figure 7c we analyzed a similar situation at a watershed within the study area. The DD varies across the landscape with the highest values over Leptosols (9.4 km/km²) at the backslope, and the lowest values over the summit and alluvial toeslope (Fig 7c).

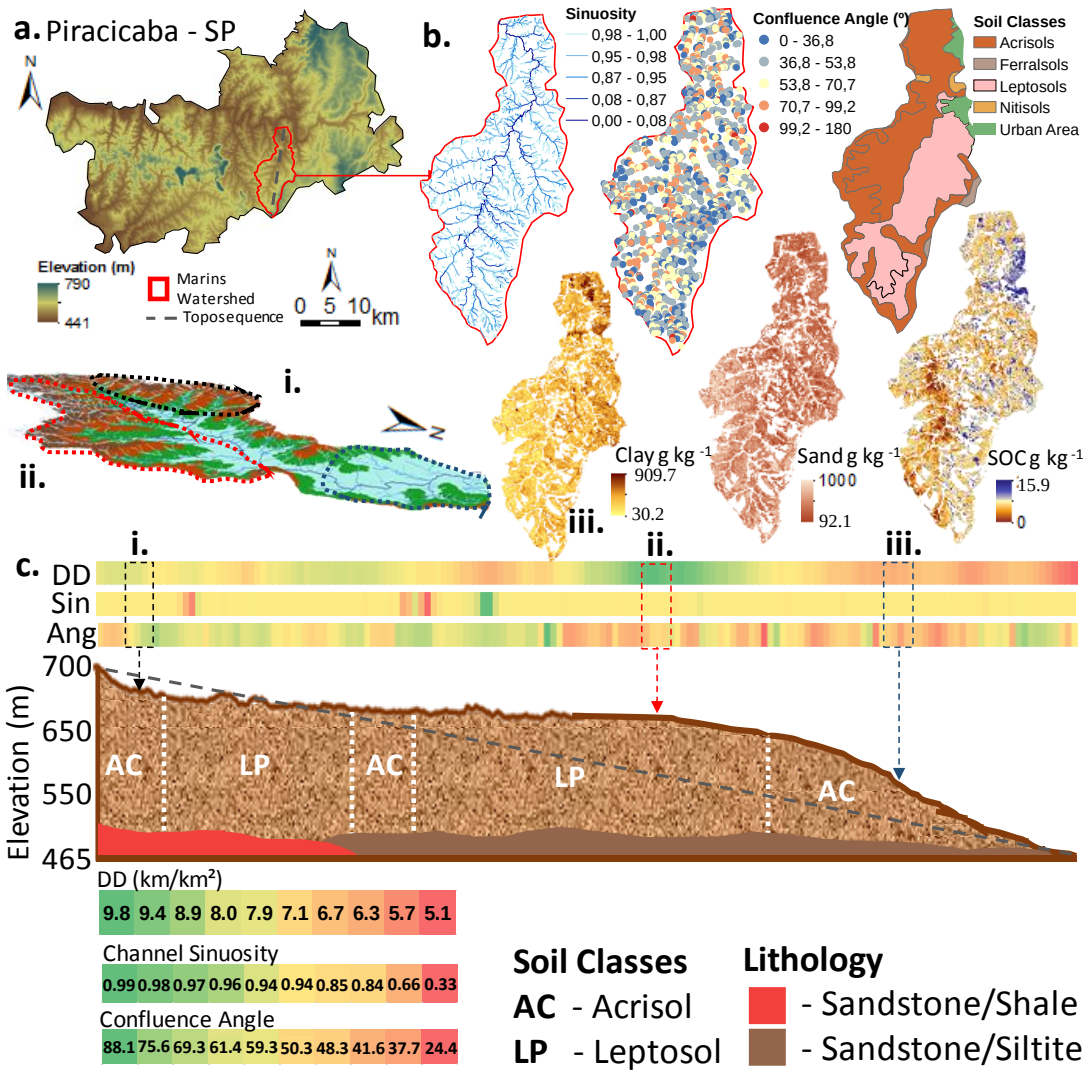

Figure 7. Analysis of drainage related variables and digital soil maps on the Marins watershed, at Piracicaba-SP. a. Piracicaba municipality, elevation, Marins watershed, and the toposequence transect; b. channel sinuosity, confluence angle, soil classes, and digital soil maps from cubist model; c. variations of hydrological variables, soil, and lithological classes across a toposequence within the Marins watershed.

Figure 7b indicates that there are no great differences of the confluence angle within the Marins watershed (Fig. 7b). However, the confluence angles on the eastern side of the area tend to be acute, while on the western side there are more right angles (Fig. 7b). According to Way (1973), these changes on the drainage system morphometry are related to variations on the soil texture and rock porosity. The right angles occurred more over Acrisols, which commonly have a sandy surface horizon, usually linked to an eluvial horizon (dos Santos et al., 2018; IUSS Working Group, 2015). In contrast, the acute angles were located over Leptosols, which in this case, is a young soil with shallow surface horizon and parent material close to the surface.

This scenario is similar to the one described by Parvis (1950), who found a dendritic drainage pattern over sandstone/shale with symmetry of its drainage lines. Parvis (1950) also pointed out that soils in these areas contain plastic clay and erode in V-shaped gullies, resulting in acute angles. Way (1973) linked the dendritic pattern (acute angles) with a homogeneous distribution of soils and rocks, which is the case at the Marins watershed (Fig. 7). Although the confluence angle is an innovative variable, its relationship with the soil information proved to be an

efficient covariable for DSM. In this tropical scenario, with a developed DN, and a relatively stable geomorphology (Cunha et al., 2005), it was possible to investigate the variations of the confluence angles over different soils.

The channel sinuosity information was more difficult to relate with soil attributes. The sinuosity is more pronounced at longer channels, which occur at more stable positions in the landscape. Fig. 7c indicates that the sinuosity was homogeneous across the landscape, with more sinuous channels at flat positions, where meandering is possible (Chitale, 1973; Fowler et al., 2007). However, the differences on channel sinuosity over soil classes indicated a potential variable to distinguish soil information (Fig. 4). The sinuosity changes across the landscape from the summit (stable), backslope (unstable), and floodplain (stable) and for different channel orders, being higher in advanced orders. Moreover, the channel sinuosity had an important contribution to the model and should be considered in other DSM works, especially in areas with different climatic conditions and water regime, since this study investigated this relationship in a tropical region with a high annual precipitation rate.

Soil attribute variation on the landscape can be related to the drainage information, since DN is an indicator of multiple processes that take place over the landscape. According to the soil landscape relationship rules, soils at the summit tend to be homogeneous due to the surface's stability which results in a well drained, weathered, and deep profile (Milne, 1935; Ruhe, 1960) with few channels (Dunne, 1980). As the relief changes to backslope and footslope positions, the water dynamics change due to shallow soils and steep slopes, affecting the channels' pattern across the landscape (Jung et al., 2011; Parvis, 1950). Finally, Vasques et al. (2015) concluded that the drainage patterns around the soil is directly linked to processes of soil formation and water movement, which is used in the Brazilian Soil Classification System (Santos et al., 2018).

Numerous theories and hypotheses have been proposed through the years concerning the formation and development of channel sinuosity and confluence angles (Best, 1988; Fowler et al., 2007; Lane, 1957; Yang, 1971). Although there are many works relating the channel morphometry with geology, terrain, and other natural factors (Jung et al., 2015; Lin et al., 2006; Miller, 1958), this work addressed how some channel characteristics are affected by the soil and how DSM can benefit from this relationship.

The slight improvement in model performance achieved by the addition of DN covariates calls for a more specific analysis regarding the role of soil on channel morphometry (Table 3). Drainage variables can be analyzed by other means, e.g., at each hierarchical channel order, including other DN information, testing hydrography in different scales, and predicting other soil attributes. Besides that, DN contributions to model performance may be different according to the study site, such as Mello et al. (2021), which found expressive improvements in soil class predictions in a nearby site. In this work, we confirmed the proposition of more research regarding DN as a DSM covariate, seeking to achieve higher correlation with soil spatial distribution.

2.5. Conclusions

The bare soil image was the most important variable for the prediction of the soil attributes. Nonetheless, the hydrological attributes presented great importance on soil predictions, validating our hypothesis.

SySI was the most important environmental variable in the prediction model, followed by DN and terrain attributes. SySI is an established data for DSM, but the addition of DN information improved the model performance, confirming the association between soil variability and drainage morphometry. It was also the one with most variation on the toposequence, indicating its significance to understand the soil-landscape relationship.

Despite the importance of traditionally used terrain variables (i.e., slope, aspect, and others), DN presented greater significance on the attribute's quantification. Indeed, the confluence angle presented notable variability over

different soil classes. The representative channel sinuosity values are only manifested from third order channels. Therefore, we conclude that it is worth exploring the relationship between channel sinuosity and soils, focusing on more representative channels.

For this tropical region, with complex lithological and pedological characteristics, RF and Cubist models performed similarly, producing good results for soil attribute mapping.

Acknowledgments

We are grateful to Geotechnologies in Soil Science Group (GeoCiS, ESALQ - USP; <http://esalqgeocis.wixsite.com/english>) for team support. We also thank the anonymous reviewers. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. Funding was provided by São Paulo Research Foundation (FAPESP, grant n. 2014/ 22262-0;).

References

Ackers, P., Charlton, F.G., 1970. The Geometry of Small Meandering Streams. Proc. Inst. Civ. Eng. 47, 80.

Alam, M.M., Crook, K.A.W., Taylor, G., 1985. Fluvial herring-bone cross-stratification in a modern tributary mouth bar, Coonamble, New South Wales, Australia. Sedimentology 32, 235–244. https://doi.org/10.1111/j.1365-3091.1985.tb00506.x

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

Barreto, A.G.O.P., Sparovek, G., Giannotti, M.P.P.-P., 2006. Atlas rural de Piracicaba.

Barroso, G.F., Gonçalves, M.A., Garcia, F. da C., 2014. The Morphometry of Lake Palmas, a Deep Natural Lake in Brazil. PLoS One 9, e111469.

Bazaglia Filho, O., Rizzo, R., Lepsch, I.F., Prado, H. do, 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. Ciência do Solo 37, 1136–1148.

Best, J.L., 1988. Sediment transport and bed morphology at river channel confluences. Sedimentology 35, 481–498. https://doi.org/10.1111/j.1365-3091.1988.tb00999.x

Best, J.L., 1986. The morphology of river channel confluences. Prog. Phys. Geogr. Earth Environ. 10, 157–174. https://doi.org/10.1177/030913338601000201

Biron, P.M., S., R.A., Sangsoo, H., 2004. Three-Dimensional Numerical Modeling of Mixing at River Confluences. J. Hydraul. Eng. 130, 243–253. https://doi.org/10.1061/(ASCE)0733-9429(2004)130:3(243)

Bonakdari, H., Lipeme-Kouyi, G., Wang, X., 2011. Experimental Validation of CFD Modeling of Multiphase Flow through Open Channel Confluence. World Environ. Water Resour. Congr. 2011, Proceedings. https://doi.org/doi:10.1061/41173(414)227

Bonfatti, B.R., Demattê, J.A.M., Marques, K.P.P., Poppiel, R.R., Rizzo, R., Mendes, W. de S., Silvero, N.E.Q., Safanelli, J.L., 2020. Digital mapping of soil parent material in a heterogeneous tropical area. Geomorphology 367, 107305. https://doi.org/https://doi.org/10.1016/j.geomorph.2020.107305

Bonifacio, E., Zanini, E., Boero, V., Franchini-Angela, M., 1997. Pedogenesis in a soil catena on serpentinite in north-western Italy. Geoderma 75, 33–51.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1007/9781441993267_5

Browne, M.W., 2000. Cross-Validation Methods. J. Math. Psychol. 44, 108–132. https://doi.org/https://doi.org/10.1006/jmps.1999.1279

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

Bryant, F.B., Yarnold, P.R., 1995. Principal-components analysis and exploratory and confirmatory factor analysis.

Camargo, O.A., Moniz, A.C., Jorge, J.A., Valadares, J., 1986. Métodos de análise química, mineralógica e física de solos do Instituto Agronômico de Campinas.

Chagas, C. da S., de Carvalho Junior, W., Bhering, S.B., Calderano Filho, B., 2016. Spatial prediction of soil surface texture in a semiarid region using random forest and multiple linear regressions. CATENA 139, 232–240. https://doi.org/https://doi.org/10.1016/j.catena.2016.01.001

Charlton, R., 2008. Fundamentals of Fluvial Geomorphology, 1st ed. Routledge, London.

Chatley, H., 1938. Hydraulics of large rivers. J. Jr. Inst. Eng 8, 16.

Chitale, S. V, 1973. Theories and relationships of river channel patterns. J. Hydrol. 19, 285–308. https://doi.org/https://doi.org/10.1016/0022-1694(73)90104-2

Clubb, F.J., Mudd, S.M., Attal, M., Milodowski, D.T., Grieve, S.W.D., 2016. The relationship between drainage density, erosion rate, and hilltop curvature: Implications for sediment transport processes. J. Geophys. Res. Earth Surf. 121, 1724–1745. https://doi.org/10.1002/2015JF003747

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J., 2015. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. Geosci. Model Dev. 8. https://doi.org/10.5194/gmd-8-1991-2015

Crave, A., Gascuel-Odoux, C., 1997. The influence of topography on time and space distribution of soil surface water content. Hydrol. Process. 11, 203–210.

Cunha, P., Marques Júnior, J., Curi, N., Pereira, G.T., Lepsch, I.F., 2005. Superfícies geomórficas e atributos de Latossolos em uma seqüência Arenítico-Basáltica da região de Jaboticabal (SP) . Rev. Bras. Ciência do Solo .

Demattê, J.A.M., Demétrio, V.A., 1998. Caracterização de solos por padrões de drenagem e sua relação com índices de intemperismo. Pesqui. Agropecu. Bras. 33, 87–95.

Demattê, J.A.M., Demètrio, V.A., 1995. Fotointerpretação de padrões de drenagem de bacias hidrográficas na caracterização de solos desenvolvidos de rochas eruptivas básicas no estado do Paraná. Sci. Agric. 52, 569–577. https://doi.org/10.1590/s0103-90161995000300026

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/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, 4461. https://doi.org/10.1038/s41598-020-61408-1

Donagema, G.K., Campos, D.V.B. de, Calderano, S.B., Teixeira, W.G., Viana, J.H.M., Donagemma, G.K., Campos, D.V.B. de, Calderano, S.B., Teixeira, W.G., Viana, J.H.M., 2011. Manual de métodos de análise de solo, 2 rev. ed, Embrapa Solos. https://doi.org/1517-2627

dos Santos, H.G., Jacomine, P.K.T., Dos Anjos, L.H.C., De Oliveira, V.A., Lumbreras, J.F., Coelho, M.R., de Almeida, J.A., de Araujo Filho, J.C., de Oliveira, J.B., Cunha, T.J.F., 2018. Sistema brasileiro de classificação de solos. Brasília, DF: Embrapa, 2018.

Dotto, A.C., Demattê, J.A.M., Viscarra Rossel, R.A., Rizzo, R., 2020. Soil environment grouping system based on spectral, climate, and terrain data: a quantitative branch of soil series. SOIL 6, 163–177. https://doi.org/10.5194/soil-6-163-2020

Dunne, T., 1980. Formation and controls of channel networks. Prog. Phys. Geogr. 211. https://doi.org/10.1177/030913338000400204

Eakin, H.M., 1910. The Influence of the Earth's Rotation upon the Lateral Erosion of Streams. J. Geol. 18, 435–447.

Efron, B., Tibshirani, R.J., 1995. Cross-validation and the bootstrap: Estimating the error rate of a prediction rule. Division of Biostatistics, Stanford University.

F. Dormann, C., M. McPherson, J., B. Araújo, M., Bivand, R., Bolliger, J., Carl, G., G. Davies, R., Hirzel, A., Jetz, W., Daniel Kissling, W., Kühn, I., Ohlemüller, R., R. Peres-Neto, P., Reineking, B., Schröder, B., M. Schurr, F., Wilson, R., 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography (Cop.). 30, 609–628. https://doi.org/https://doi.org/10.1111/j.2007.0906-7590.05171.x

Flynn, T., Rozanov, A., de Clercq, W., Warr, B., Clarke, C., 2019. Semi-automatic disaggregation of a national resource inventory into a farm-scale soil depth class map. Geoderma 337, 1136–1145. https://doi.org/10.1016/j.geoderma.2018.11.003

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. 10. https://doi.org/10.3390/rs10101555

Fowler, A.C., Kopteva, N., Oakley, C., 2007. The Formation of River Channels. SIAM J. Appl. Math. 67, 1016–1040. https://doi.org/10.1137/050629264

França, G.V., Demattê, J.A.M., 1990. Parâmetros da rede de drenagem de solos da região de Iracemápolis (SP). An. da Esc. Super. Agric. Luiz Queiroz 47, 541–555. https://doi.org/10.1590/s0071-12761990000200015

Friedkin, J.F., 1945. A laboratory study of the meandering of alluvial rivers. United States Waterways Experiment Station.

Gambill, D.R., Wall, W.A., Fulton, A.J., Howard, H.R., 2016. Predicting USCS soil classification from soil property variables using Random Forest. J. Terramechanics 65, 85–92. https://doi.org/10.1016/j.jterra.2016.03.006

Gaudet, J.M., Roy, A.G., 1995. Effect of bed morphology on flow mixing length at river confluences. Nature 373, 138–139. https://doi.org/10.1038/373138a0

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

Gray, D.M., 1961. Interrelationships of watershed characteristics. J. Geophys. Res. 66, 1215–1223. https://doi.org/10.1029/JZ066i004p01215

Grinand, C., Arrouays, D., Laroche, B., Martin, M.P., 2008. Extrapolating regional soil landscapes from an existing soil map: Sampling intensity, validation procedures, and integration of spatial context. Geoderma 143, 180–190. https://doi.org/https://doi.org/10.1016/j.geoderma.2007.11.004

Hartemink, A.E., McBratney, A., Mendonça-Santos, M. de L., 2008. Digital Soil Mapping with Limited Data, Springer. https://doi.org/10.1007/978-1-4020-8592-5

Hassink, J., Bouwman, L.A., Zwart, K.B., Bloem, J., Brussaard, L., 1993. Relationships between soil texture, physical protection of organic matter, soil biota, and C and N mineralization in grassland soils, in: Brussaard, L., Kooistra, M.J.B.T.-S.S.B.I. (Eds.), . Elsevier, Amsterdam, pp. 105–128. https://doi.org/https://doi.org/10.1016/B978-0-444-81490-6.50059-5

Hawkins, D.M., Basak, S.C., Mills, D., 2003. Assessing Model Fit by Cross-Validation. J. Chem. Inf. Comput. Sci. 43, 579–586. https://doi.org/10.1021/ci025626i

Haynes, R.J., Swift, R.S., 1990. Stability of soil aggregates in relation to organic constituents and soil water content. J. Soil Sci. 41, 73–83. https://doi.org/10.1111/j.1365-2389.1990.tb00046.x

Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B.M., Gräler, B., 2018. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ 6, e5518. https://doi.org/10.7717/peerj.5518

Hoogsteen, M.J.J., Lantinga, E.A., Bakker, E.J., Groot, J.C.J., Tittonell, P.A., 2015. Estimating soil organic carbon through loss on ignition: effects of ignition conditions and structural water loss. Eur. J. Soil Sci. 66, 320–328. https://doi.org/10.1111/ejss.12224

Horst-Heinen, T.Z., Dalmolin, R.S.D., ten Caten, A., Moura-Bueno, J.M., Grunwald, S., Pedron, F. de A., Rodrigues, M.F., Rosin, N.A., da Silva-Sangoi, D.V., 2021. Soil depth prediction by digital soil mapping and its impact in pine forestry productivity in South Brazil. For. Ecol. Manage. 488, 118983. https://doi.org/https://doi.org/10.1016/j.foreco.2021.118983

IUSS Working Group, W.R.B., 2015. World reference base for soil resources. World Soil Resour. Rep. 103.

John, K., Kebonye, N.M., Agyeman, P.C., Ahado, S.K., 2021. Comparison of Cubist models for soil organic carbon prediction via portable XRF measured data. Environ. Monit. Assess. 193, 197. https://doi.org/10.1007/s10661-021-08946-x

Jung, K., Marpu, P.R., Ouarda, T.B.M.J., 2015. Improved classification of drainage networks using junction angles and secondary tributary lengths. Geomorphology 239, 41–47. https://doi.org/https://doi.org/10.1016/j.geomorph.2015.03.004

Jung, K., Niemann, J.D., Huang, X., 2011. Under what conditions do parallel river networks occur? Geomorphology 132, 260–271. https://doi.org/10.1016/j.geomorph.2011.05.014

Kassambara, A., 2017. Practical guide to principal component methods in R: PCA, M (CA), FAMD, MFA, HCPC, factoextra. Sthda.

Ker, J.C., Curi, N., Schaefer, C.E.G.R., Vidal-Torrado, P., 2015. Pedologia: fundamentos.

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

Kidd, D., Malone, B., McBratney, A., Minasny, B., Webb, M., 2015. Operational sampling challenges to digital soil mapping in Tasmania, Australia. Geoderma Reg. 4, 1–10. https://doi.org/https://doi.org/10.1016/j.geodrs.2014.11.002

Kuhn, M., 2008. Building Predictive Models in R Using the caret Package. J. Stat. Software; Vol 1, Issue 5 . https://doi.org/10.18637/jss.v028.i05

Lacey, J.M., 1923. Some Problems connected with the Rivers and the Canals in Southern India., in: Minutes of the Proceedings of the Institution of Civil Engineers. Thomas Telford-ICE Virtual Library, pp. 150–160.

Lagacherie, P., Arrouays, D., Bourennane, H., Gomez, C., Martin, M., Saby, N.P.A., 2019. How far can the uncertainty on a Digital Soil Map be known?: A numerical experiment using pseudo values of clay content obtained from Vis-SWIR hyperspectral imagery. Geoderma 337, 1320–1328. https://doi.org/https://doi.org/10.1016/j.geoderma.2018.08.024

Lamichhane, S., Kumar, L., Adhikari, K., 2021. Updating the national soil map of Nepal through digital soil mapping. Geoderma 394, 115041. https://doi.org/https://doi.org/10.1016/j.geoderma.2021.115041

Lane, E.W., 1957. A study of the shape of channels formed by natural streams flowing in erodible material.

Langbein, W.B., Leopold, L.B., 1966. River meanders - Theory of minimum variance, Professional Paper. Washington, D.C. https://doi.org/10.3133/pp422H

Li, X., Ding, J., Liu, J., Ge, X., Zhang, J., 2021. Digital Mapping of Soil Organic Carbon Using Sentinel Series Data: A Case Study of the Ebinur Lake Watershed in Xinjiang. Remote Sens. . https://doi.org/10.3390/rs13040769

Lin, H., Bouma, J., Pachepsky, Y., Western, A., Thompson, J., van Genuchten, R., Vogel, H.-J., Lilly, A., 2006. Hydropedology: Synergistic integration of pedology and hydrology. Water Resour. Res. 42. https://doi.org/10.1029/2005WR004085

Loiseau, T., Arrouays, D., Richer-de-Forges, A.C., Lagacherie, P., Ducommun, C., Minasny, B., 2021. Density of soil observations in digital soil mapping: A study in the Mayenne region, France. Geoderma Reg. 24, e00358. https://doi.org/https://doi.org/10.1016/j.geodrs.2021.e00358

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

Mello, F.A.O., Demattê, J.A.M., Rizzo, R., Dotto, A.C., Poppiel, R.R., Mendes, W.S., Guimarães, C.C.B., 2021. Expert-based maps and highly detailed surface drainage models to support digital soil mapping. Geoderma 384, 114779. https://doi.org/https://doi.org/10.1016/j.geoderma.2020.114779

Miller, B.A., Schaetzl, R.J., 2016. History of soil geography in the context of scale. Geoderma 264, 284–300. https://doi.org/10.1016/j.geoderma.2015.08.041

Miller, J.P., 1958. High mountain streams: effects of geology on channel characteristics and bed material. State Bureau of Mines and Mineral Resources, New Mexico Institute of Mining ….

Milne, G., 1935. Composite units for the mapping of complex soil associations. Trans. 3rd Int. Congr. Soil Sci 1, 345–347.

Minasny, B., McBratney, A.B., 2016. Digital soil mapping: A brief history and some lessons. Geoderma 264, 301–311. https://doi.org/10.1016/j.geoderma.2015.07.017

Minasny, B., McBratney, A.B., 2008. Regression rules as a tool for predicting soil properties from infrared reflectance spectroscopy. Chemom. Intell. Lab. Syst. 94, 72–79. https://doi.org/10.1016/j.chemolab.2008.06.003

Moeys, J., 2018. The soil texture wizard: R functions for plotting, classifying, transforming and exploring soil texture data. CRAN. R-Project.

Mosier, C.I., 1951. I. Problems and Designs of Cross-Validation 1. Educ. Psychol. Meas. 11, 5–11. https://doi.org/10.1177/001316445101100101

Moura-Bueno, J.M., Dalmolin, R.S.D., Horst-Heinen, T.Z., Grunwald, S., ten Caten, A., 2021. Environmental covariates improve the spectral predictions of organic carbon in subtropical soils in southern Brazil. Geoderma 393, 114981. https://doi.org/https://doi.org/10.1016/j.geoderma.2021.114981

Moussa, R., 2003. On morphometric properties of basins, scale effects and hydrological response. Hydrol. Process. 17, 33–58. https://doi.org/10.1002/hyp.1114

Mzid, N., Pignatti, S., Huang, W., Casa, R., 2021. An Analysis of Bare Soil Occurrence in Arable Croplands for Remote Sensing Topsoil Applications. Remote Sens. . https://doi.org/10.3390/rs13030474

Odgers, N.P., Sun, W., McBratney, A.B., Minasny, B., Clifford, D., 2014. Disaggregating and harmonising soil map units through resampled classification trees. Geoderma 214–215, 91–100. https://doi.org/10.1016/j.geoderma.2013.09.024

Pallard, B., Castellarin, A., Montanari, A., 2009. A look at the links between drainage density and flood statistics. Hydrol. Earth Syst. Sci. 13, 1019.

Parvis, M., 1950. Drainage pattern significance in airphoto identification of soils and bedrocks. Photogramm. Eng. 16, 375–409.

Patel, D.P., Gajjar, C.A., Srivastava, P.K., 2013. Prioritization of Malesari mini-watersheds through morphometric analysis: a remote sensing and GIS perspective. Environ. earth Sci. 69, 2643–2656.

Poppiel, R.R., Lacerda, M.P.C., Demattê, J.A.M., Oliveira, M.P., Gallo, B.C., Safanelli, J.L., 2019a. Pedology and soil class mapping from proximal and remote sensed data. Geoderma 348, 189–206. https://doi.org/10.1016/j.geoderma.2019.04.028

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

Quinlan, J.R., 1992. Learning with continuous classes. Mach. Learn. 92, 343–348. https://doi.org/10.1.1.34.885

Quinlan, J.R.B.T.-M.L.P. 1993, 1993. Combining Instance-Based and Model-Based Learning. Morgan Kaufmann, San Francisco (CA), pp. 236–243. https://doi.org/https://doi.org/10.1016/B978-1-55860-307-3.50037-X

Quraishy, M.S., 1943. River meandering and the earth's rotation. Curr. Sci. 12, 278.

R Core Team, 2013. R: A language and environment for statistical computing.

Ray, R.G., 1960. Aerial photographs in geologic interpretation and mapping, Professional Paper. https://doi.org/10.3133/pp373

Richards, K.S., 1980. A note on changes in channel geometry at tributary junctions. Water Resour. Res. 16, 241–244. https://doi.org/10.1029/WR016i001p00241

Ritchie, J.T., 1981. SOIL WATER AVAILABILITY. Plant Soil 58, 327–338.

Romshoo, S.A., Bhat, S.A., Rashid, I., 2012. Geoinformatics for assessing the morphometric control on hydrological response at watershed scale in the Upper Indus Basin. J. Earth Syst. Sci. 121, 659–686.

Ruhe, R. V, 1960. Elements of the soil landscape. Trans. 7th int. Congr. Soil Sci. 4, 165–170.

Russell, R.J., 1936. Physiography of the lower Mississippi delta: Louisiana Geol. Survey, Geol. Bull.

Safanelli, L.J., Chabrillat, S., Ben-Dor, E., Demattê, A.M.J., 2020. Multispectral Models from Bare Soil Composites for Mapping Topsoil Properties over Europe. Remote Sens. . https://doi.org/10.3390/rs12091369

Schaetzl, R., Anderson, S., 2005. Soils. Genesis and Geomorphology. Cambridge University Press, New York.

Silvero, N.E.Q., Demattê, J.A.M., Vieira, J. de S., Mello, F.A. de O., Amorim, M.T.A., Poppiel, R.R., Mendes, W. de 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/https://doi.org/10.1016/j.geoderma.2021.115089

Soil Science Division Staff, 2017. Soil survey manual. Government Printing Office Washington, DC.

Strahler, A.N., 1957. Quantitative analysis of watershed geomorphology. Eos, Trans. Am. Geophys. Union 38, 913–920.

Strahler, A.N., 1952. Dynamics basis of geomorphology. Bull. Geol. Soc. Am. 63, 923–938.

Strahler, A.N., Strahler, A.H., 1989. Elements of Physical Geograpgy, 4th ed. John Wiley & Sons, Inc.

Stumpf, F., Schmidt, K., Behrens, T., Schönbrodt-Stitt, S., Buzzo, G., Dumperth, C., Wadoux, A., Xiang, W., Scholten, T., 2016. Incorporating limited field operability and legacy soil samples in a hypercube sampling design for digital soil mapping. J. Plant Nutr. Soil Sci. 179, 499–509. https://doi.org/https://doi.org/10.1002/jpln.201500313

Styc, Q., Gontard, F., Lagacherie, P., 2021. Harvesting spatially dense legacy soil datasets for digital soil mapping of available water capacity in Southern France. Geoderma Reg. 24, e00353. https://doi.org/https://doi.org/10.1016/j.geodrs.2020.e00353

Teixeira, P.C., Donagemma, G.K., Fontana, A., Teixeira, W.G., 2017. Manual de métodos de análise de solo. Rio Janeiro 4.

Teramoto, E.R., Lepsch, I.F., Vidal-Torrado, P., 2001. Relações solo, superfície geomórfica e substrato geológico na microbacia do ribeirão Marins (Piracicaba - SP) . Sci. Agric. .

Thomas, A.L., King, D., Dambrine, E., Couturier, A., Roque, J., 1999. Predicting soil classes with parameters derived from relief and geologic materials in a sandstone region of the Vosges mountains (Northeastern France). Geoderma 90, 291–305. https://doi.org/https://doi.org/10.1016/S0016-7061(98)00135-9

Thorp, J., Thoms, M.C., Delong, M.D., 2010. The riverine ecosystem synthesis: toward conceptual cohesiveness in river science. Elsevier.

U.S.G.S., 2019a. Landsat 4-7 Surface Reflectance (Ledaps) Product Guide 32. https://doi.org/10.1016/0042-207X(74)93024-3

U.S.G.S., 2019b. Landsat 8 Surface Reflectance Code (LASRC) Poduct Guide. (No. LSDS-1368 Version 2.0). 40.

van Breemen, N., Buurman, P., 2002. Soil Formation - Second Edition.

Vasques, G.M., Demattê, J.A.M.M., Viscarra Rossel, R.A., Ramírez López, L., Terra, F.S., Rizzo, R., De Souza Filho, C.R., 2015. Integrating geospatial and multi-depth laboratory spectral data for mapping soil classes in a geologically complex area in southeastern Brazil. Eur. J. Soil Sci. 66, 767–779. https://doi.org/10.1111/ejss.12255

Vaudour, E., Gomez, C., Lagacherie, P., Loiseau, T., Baghdadi, N., Urbina-Salazar, D., Loubet, B., Arrouays, D., 2021. Temporal mosaicking approaches of Sentinel-2 images for extending topsoil organic carbon content mapping in croplands. Int. J. Appl. Earth Obs. Geoinf. 96, 102277. https://doi.org/https://doi.org/10.1016/j.jag.2020.102277

Vidal-Torrado, P., Lepsch, I.F., 1999. Relações material de origem / solo e pedogênese em uma seqüência de solos predominantemente argilosos e Latossólicos sobre psamitos na depressão periférica Paulista: Paulo State Peripheral Depression, southeastern Brazil . Rev. Bras. Ciência do Solo .

Vincent, S., Lemercier, B., Berthier, L., Walter, C., 2018. Spatial disaggregation of complex Soil Map Units at the regional scale based on soil-landscape relationships. Geoderma 311, 130–142. https://doi.org/10.1016/j.geoderma.2016.06.006

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

Way, D.S., 1973. Terrain analysis: a guide to site selection using aerial photographic interpretation. Stroudsburg, PA: Dowden, Hutchinson and Ross.

Wei, T., Simko, V., Levy, M., Xie, Y., Jin, Y., Zemla, J., 2017. Package 'corrplot.' Statistician 56, e24.

Wickham, H., 2011. ggplot2. WIREs Comput. Stat. 3, 180–185. https://doi.org/https://doi.org/10.1002/wics.147

Yang, C.T., 1971. On river meanders. J. Hydrol. 13, 231–253. https://doi.org/https://doi.org/10.1016/0022-1694(71)90226-5

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/https://doi.org/10.1016/j.geoderma.2018.09.006

Zhang, L., Yang, L., Ma, T., Shen, F., Cai, Y., Zhou, C., 2021. A self-training semi-supervised machine learning method for predictive mapping of soil classes with limited sample data. Geoderma 384, 114809. https://doi.org/https://doi.org/10.1016/j.geoderma.2020.114809

Zhang, S.-J., Zhu, A.-X., Liu, J., Yang, L., Qin, C.-Z., An, Y.-M., 2016. An heuristic uncertainty directed field sampling design for digital soil mapping. Geoderma 267, 123–136. https://doi.org/https://doi.org/10.1016/j.geoderma.2015.12.009

3. SOIL PARENT MATERIAL PREDICTION THROUGH SATELLITE MULTISPECTRAL ANALYSIS ON A REGIONAL SCALE AT THE WESTERN PAULISTA PLATEAU, BRAZIL.

Abstract

Parent material is the main source for soil textural, mineralogical, and other physical attributes. The knowledge over this factor is explored generally in low scale geology maps, insufficient for most users. Remote sensing can offer assistance in this regard, since it allows the evaluation of soil properties, as largely indicated in literature, being a potential tool to delineate parent material. Thus, we explored a multi temporal Landsat image composition with bare soil reflectance to extract soil properties and distinguish discrepant lithological classes at the western plateau, São Paulo State, Brazil. The area is 247,737 ha large, where 981 soil samples were collected at 0 – 20 cm depth. We acquired the synthetic soil image and linked the pixel's spectra with soil attributes. We performed a digital soil mapping procedure to generate maps of attributes related to parent material. The soil maps offered a great input on identifying the transitions between sandstone and basalt as soils from these formations have significant differences in clay, sand, Fe_2O_3 and TiO_2 contents. Therefore, the use of remote sensing coupled with digital soil mapping is a strong alternative to conventional methods to improve low scale PM maps to enhance detail on regional and local scales.

Keywords: Remote sensing, Digital Soil Mapping, Parent material, Satellite data, Synthetic soil image, Lithological transitions, Multiple soil classes delineation

Published as: Mello, F.A.O., Bellinaso, H., Mello, D.C., Safanelli, J.L., Mendes, W.D.S., Amorim, M.T.A., Gomez, A.M.R., Poppiel, R.R., Silvero, N.E.Q., Gholizadeh, A., Silva, S.H.G., Curi, N., Demattê, J.A.M., 2021. Soil parent material prediction through satellite multispectral analysis on a regional scale at the Western Paulista Plateau, Brazil. Geoderma Reg. e00412. https://doi.org/https://doi.org/10.1016/j.geodrs.2021.e00412

3.1. Introduction

Parent material (PM) governs most soil physical properties and characteristics (Schaetzl, J Randall and Anderson, 2005; Wilson, 2019). The mineral composition of the rock influences soil color, mineralogy, texture, water retention capability, magnetic susceptibility and other properties (Ker et al., 2015; Ma et al., 2019; Mokma and Sprecher, 1994; Richardson and Daniels, 1993; Schwertmann, 1993), which are important for soil management in agricultural areas. Therefore, multiple strategies have been used to map PM through information contained in soil maps (Florea et al., 2015; Miller et al., 2008; Prokopovich, 1984). These efforts are important since PM maps are necessary for human activities such as mining, agriculture, water, fuel, and others.

Despite the role of soil and geological information for economical and sustainable activities (Maltman, 2012; Prokopovich, 1984), most available databases are limited and with coarse spatial resolution, hampering its application for local and regional planning (Adhikari and Hartemink, 2016; Dobos et al., 2013; McBratney et al., 2014; Nolasco de Carvalho et al., 2015).

Digital soil mapping (DSM) and remote sensing (RS), emerged as important tools to fill this gap and produce more detailed maps of natural resources (Dharumarajan et al., 2021; Searle et al., 2021; Styc et al., 2021; Wadoux et al., 2019) and to improve legacy databases (Lamichhane et al., 2021; Nauman and Thompson, 2014; Odgers et al., 2014; Pelegrino et al., 2016; Vincent et al., 2018).

The use of RS coupled with geographical information systems (SIG) has resulted in new representative environmental data, which assisted on the prediction of soil attributes, landscape processes, and geological information (Grimm et al., 2008; Minasny and Hartemink, 2011; Poppiel et al., 2019b). Recently, Demattê et al. (2018a), developed the Synthetic Soil Image (SySI) which is a multi-temporal composition of bare soil retrieved from Landsat time series. SySI has proved to aid on the prediction of soil chemical and physical attributes, such as clay,

sand, organic matter, and Fe_2O_3 (Fongaro et al., 2018; Mendes et al., 2019; Poppiel et al., 2019a), soil classes (Mello et al., 2021; Rizzo et al., 2020), and PM classes (Bonfatti et al., 2020; Gallo et al., 2018).

The combination of RS and soil information has more to contribute for the mapping of geological features, based on the relationship between soil physical attributes and PM (Blatnik et al., 2020). The challenge stands on satellites capturing electromagnetic energy reflected from the Earth's surface, which can be affected by the physical and chemical attributes of the ground (Baumgardner et al., 1986). Moreover, the reflected energy responds to the upper few centimeters of the soil surface, presenting no direct effects of PM reflectance. However, soil attributes such as clay, sand, Fe_2O_3, and TiO_2 are directly affected by the PM and are also manifested on the spectral reflectance (Dematte et al., 2009; Izawa et al., 2019; Silva et al., 2020).

Therefore, our hypotheses are that soil attributes from 0-20 cm layer and bare soil reflectance are related to the soil PM, and that it is possible to upscale a geological map from national to regional extent using SySI and DSM techniques. The objectives in this study were: i) evaluate the relationship between SySI and soil attributes mainly affected by parent material in the Western Paulista Plateau; ii) predict PM classes using SySI, environmental variables, and random forest modeling.

3.2. Material and methods

3.2.1. Study area

The study area is located at the Pereira Barreto municipality, São Paulo State, at the Western Paulista Plateau (Fig.1). The area consists of a 247.7 ha agricultural site that is used for sugarcane production (Caldarelli and Gilio, 2018). The area has a savanna climate (Aw) with wet summers and dry winters, according to the Köppen climate classification (Alvares et al., 2013). The temperature ranges between 21.5 and 27.0 ºC with an annual rainfall of 1,100 millimeters (mm).

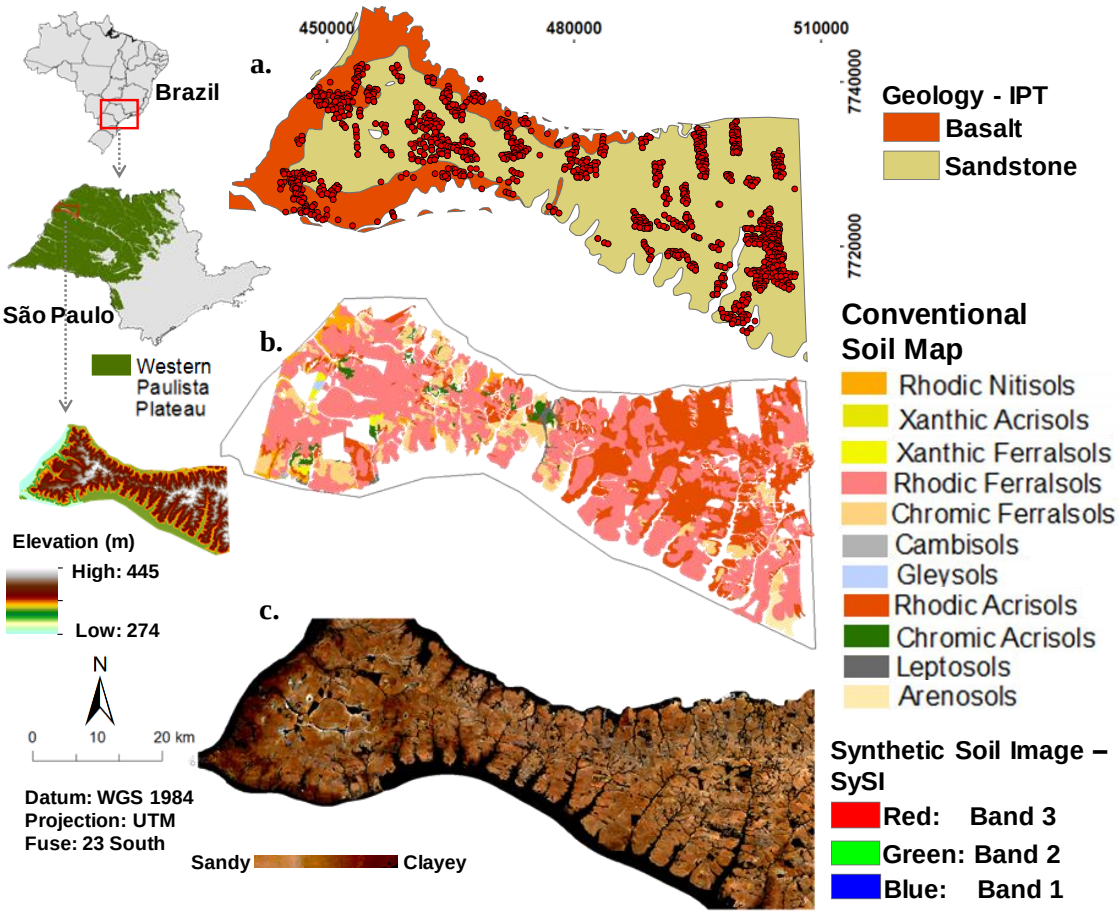

Figure 1: Study area located at the São Paulo State, Brazil. a. geological map from Paulista Institute of Technology (IPT, 1981) covered by the soil collection points, separated into calibration and validation datasets. b. Detailed soil map designed by a soil mapping specialist with the soil classification according to WRB (IUSS Working Group, 2014). c. Synthetic Soil Image (SYSI) developed by the GEOS3 method (Demattê et al., 2018).

Geology dates from higher Cretaceous (88 - 65 million years) and is generally composed of sedimentary rocks of Bauru and Caiuá groups (SR) (84%), most sandstones (57%), and basaltic spills of Serra Geral (SG) formation (15%) (Fernandes et al., 2007). The study area comprises two geological formations: Santo Anastácio (Caiuá group), with large presence of sedimentary rocks, mainly sandstones, and *SG*, with strong presence of igneous rocks like basalt (Fig. 1 a).

3.2.2. Soil data

First, we indicate a flow chart of activities applied in the work which will be described in detail on further topics.

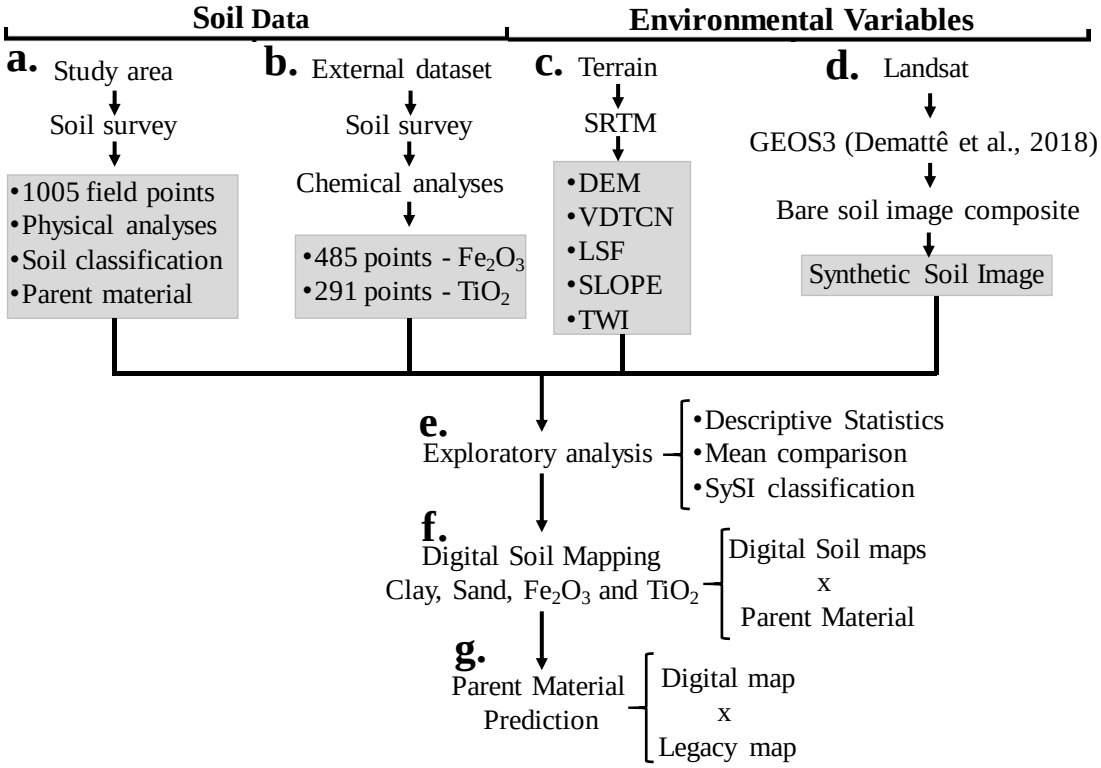

Figure 2: Flow chart of methodology activities. Indicating the construction of digital soil maps and the attempts to enhance detail on the geological map.

A conventional soil survey took place at the study area (Fig. 1b). The work was conducted by a team of soil scientists guided by a specialist. The survey provided information about the soil-landscape relationship, soil samples, and soil profile analysis according to the directions of the Brazilian Soil Classification system (dos Santos et al., 2018) and the Brazilian Soil Survey Manual (Santos et al., 1995). We collected 1005 soil samples at 0 – 20 cm depth (Table 1) using traditional auger methods and analyzed 44 soil profiles to produce a final soil map on 1:20,000 scale (Fig. 1b).

Table 1: Soil suborders according to the Brazilian System of Soil Classification(dos Santos et al., 2018) and correspondent soil classes on the World Reference Base (IUSS Working Group, 2015). Soil parent material was verified at the field.

Code	SiBCS	WRB	Parent Material	n
CX1	Cambissolo Háplico Muito Argiloso	Dystric Cambisol	Basalt	10
CX2	Cambissolo Háplico Argiloso	Dystric Cambisol	Basalt	2
LV2	Latossolo Vermelho Argiloso férrico	Rhodic Ferralsol	Basalt	13
LV3	Latossolo Vermelho Médio Argiloso	Rhodic Ferralsol	Sandstone	392
LV4	Latossolo Vermelho Médio Arenoso	Rhodic Ferralsol	Sandstone	220
NV1	Nitossolo Vermelho Muito Argiloso férrico	Rhodic Nitisol	Basalt	12
NV2	Nitossolo Vermelho Argiloso férrico	Rhodic Nitisol	Basalt	9
PV1	Argissolo Vermelho Muito Argiloso	Rhodic Acrisol	Sandstone	5
PV2	Argissolo Vermelho Argiloso	Rhodic Acrisol	Sandstone	37
PV3	Argissolo Vermelho Médio Argiloso	Rhodic Acrisol	Sandstone	191
PV4	Argissolo Vermelho Médio Arenoso	Rhodic Acrisol	Sandstone	92
RL2	Neossolo Litólico Argiloso	Leptosol	Basalt	12
RL3	Neossolo Litólico Médio Argiloso	Leptosol	Sandstone	4
RL4	Neossolo Litólico Médio Arenoso	Leptosol	Sandstone	2
RQ	Neossolo Quartzarênico	Arenosol	Sandstone	4

The soil samples were submitted to physico-chemical analysis (Fig. 2a). They were oven-dried for 48h at 50°C, ground and sieved through a 2mm mesh. The densimeter method was performed to analyze the soil particle size distribution using sodium hydroxide (0.1 mol L−1) and sodium hexametaphosphate (0.1 mol L−1) as dispersing agents (Camargo et al., 1986). Afterwards, the percentages of clay and sand were used to determine the soil texture class, follow the system of United States Department of Agriculture (USDA) (Soil Survey Staff, 2010).

The area has predominantly been covered with soils originated from sandstones, (e.g. Ferralsols, Acrisols and Arenosols) with high sand content (Fig. 1b). These soils are located at the higher landscape positions since the sandstone was formed through sediment deposition over the basalt layer, which is found at lower positions. The SG occurs close to drainage channels where the sandstone was weathered and eroded. The soil developed from basalt contains high clay and Fe_2O_3 contents (e.g. Ferralsols, Nitisols, Leptosols and Cambisols) (Fig. 2b) (Campos et al., 2012). Overall, the soil classes of the study site are medium texture Rhodic Ferralsols, on the flat hillslope surface under sandstone alteration, whereas on downslope the same soil class has a clayey texture, due to gradually changes to basalt alteration (Fig. 2b) (Garcia et al., 2018; Meireles et al., 2012).

3.2.2.1. External soil data

As the main goal in this research was to evaluate the identification of PM through soil analysis, we used an external soil database that contained the quantification of Iron oxide (Fe_2O_3) and Titanium oxide (TiO_2) (Fig. 2b). The soil data available for the study area did not contain this information, but we decided to use this external database due to the correlation between these oxides and the PM, and to investigate its presence in soils derived from basalt and sandstone (Lu et al., 2008; Schwertmann and Taylor, 1989).

Fe_2O_3 and TiO_2 contents were determined following the methodology proposed by Teixeira et al, (2017). We used 485 soil samples to determine Fe_2O_3 through sulfuric digestion, which consists in a determination by atomic absorption spectrometry in the sulfuric extract**.** We also used 291 soil samples to determine TiO_2 through sulfuric extract by atomic absorption.

This external soil data is part of a larger soil database which has been tested in multiple soil studies (Bellinaso et al., 2021; Mendes et al., 2021; Silvero et al., 2021). We used this information to fit a random forest model representative for the study area. Although the Fe_2O_3 and TiO_2 samples are located in a different area, we believe its use will provide benefits for our work. The modeling approach is fully covered in section 2.4.

3.2.3. Environmental variables

3.2.3.1. Terrain variables

To spatialize the soil information, we used a dataset of terrain variables (Table 2), which have shown their importance on improvement of soil attributes prediction in previous studies (Guo et al., 2019; Hengl et al., 2017; Poppiel et al., 2019a). We derived a digital elevation model (DEM) with 30m spatial resolution from the shuttle radar topography mission (SRTM) satellite image with of 1 arc-second of global resolution (Slater et al., 2006). We used the DEM to calculate more terrain attributes (Table 2) using the terrain analysis library in SAGA GIS (Conrad et al., 2015) (Fig. 2c). These variables are often applied for soil studies since their relationship with soil distribution (Florinsky, 2012; Mello et al., 2021; Ruhe, 1960).

Table 2: Environmental variables used as predictors for digital soil mapping of clay, sand, Fe_2O_3 and TiO_2.

Code	Attribute	Description	Unit	Reference
DEM	Digital Elevation Model	Elevation grid representing altitude	m	(Zhang and Montgomery, 1994)
VDTCN	Vertical Distance to Channel Network	Altitude above the channel network in the same units as the elevation data	m	(Rennó et al., 2008)
LSF	LS Factor	Slope length	non-dimensional	(Conrad et al., 2015)
SLOPE	Slope	Slope - a space curve defined for the set of nonspecial points on the surface	Degree	(Florinsky, 2012)
TWI	Topographic Wetness Index	Indicator of soil moisture distribution at different landscape positions	non-dimensional	(Pei et al., 2010)
B1	Landsat Band 1	Blue (450-520 nm)	Reflectance factor	(Demattê et al., 2018)
B2	Landsat Band 2	Green (520-600 nm)	Reflectance factor	(Demattê et al., 2018)
B3	Landsat Band 3	Red (630-690 nm)	Reflectance factor	(Demattê et al., 2018)
B4	Landsat Band 4	Near Infrared (NIR) (760-900 nm)	Reflectance factor	(Demattê et al., 2018)
B5	Landsat Band 5	Short Wave Infrared – 1 (1550-1750 nm)	Reflectance factor	(Demattê et al., 2018)
B7	Landsat Band 7	Short Wave Infrared – 2 (2080-2350 nm)	Reflectance factor	(Demattê et al., 2018)

3.2.3.2. Bare soil image composite

Multi-temporal Landsat series were used in this study to predict soil attributes and to evaluate the weight of the PM on bare soil reflectance (Fig. 2d). Soil surface reflectance provided by Landsat images offers a great tool to identify and quantify soil attributes (Rogge et al., 2018; Shabou et al., 2015). However, the soil must be exposed to express its characteristics, which hardly happens continuously at agricultural areas. To solve this issue, we used a novel approach named Geospatial Soil Sensing System (GEOS3) to produce a continuous bare soil image over the landscape (Demattê et al., 2018). We applied this method to a time-series of Landsat images using the Google Earth Engine (GEE) (Gorelick et al., 2017). We used the Landsat 4 Thematic Mapper (TM) 9 (1982–1993), Landsat 5 TM (1984–2012), Landsat 7 Enhanced Thematic Mapper Plus (ETM+) (1999–2018) and the Landsat 8 Operational Land Manager (OLI) (2013–2018). These methods use only Tier 1 or surface reflectance processed by the LEDAPS (Landsat 4, 5, and 7) and LASRC (Landsat 8) algorithms (U.S.G.S., 2019a, 2019b).

The GEOS3 algorithm was produced in GEE to extract soil reflectance from the collection of historical Landsat images and to aggregate the spatially bare soil fragments into a synthetic soil image (SySI), which is the reflectance image of the bare soil composite (Safanelli et al., 2020). To identify the bare soil pixels from single satellite images, a set of identification rules were used. They were based on spectral indices coupled with quality assessment bands, which removed cloud, cloud shadow, inland water, snow, photosynthetic vegetation, and non-photosynthetic vegetation (crop residues) (Safanelli et al., 2020). Each pixel was classified as soil based on the

Normalized Difference Vegetation Index (NDVI), with a threshold between –0.15 and 0.20 to mask out green vegetation, Normalized Burning Ratio (NBR2), with a −0.15 and 0.15 to mask out crop residues, difference between bands 1 and 2 (B2 – B1) and bands 2 and 3 (B3 – B2) (Demattê et al., 2020, 2018). Afterwards, bare soil pixels were used to calculate, pixel-by-pixel, the median values of topsoil reflectance for single bands and obtain the final (Demattê et al., 2020, 2018a). The SySI had six spectral bands from blue to short-wave infrared regions at 30 m resolution.

3.2.4. Soil attribute modeling

We used the soil samples in the study area to extract the values of the environmental variables (Table 2). This dataset was used to fit a random forest (RF) model, which had the sand and clay contents as dependent variables, while the environmental information was the independent variables. We chose these two soil attributes due to their relationship with the PM (sandstone and basalt) (Fig. 1a), which forms soils with different texture and mineral composition (van Breemen and Buurman, 2002).

3.2.4.1. Random forest and resampling methods

The RF modeling approach has been extensively used for DSM works (Khaledian and Miller, 2020; Lagacherie et al., 2019; Teng et al., 2018). We decided to use this algorithm because it has proven to be an efficient method to predict soil attributes in different ranges of data, scale, soil variability, and also to extract the relationships between soil and the landscape (Heung et al., 2014; Loiseau et al., 2019; Malone and Searle, 2021; Santra et al., 2017). We did not test other machine learning algorithms since our goal was to focus on how RS data can improve PM mapping.

Breiman (2001) explains that RF estimates a user-specified number of decision trees by randomly sampling an existing dataset. However, at each node construction, a random sample of the dependent variables is used. The resulting decision tree is used to estimate the out-of-bag error rate by predicting the value of the remaining unsampled data and comparing with the known results (Gambill et al., 2016).

We tested four resampling methods to fit our calibration model, being the out-of-bag, bootstrap, ten-fold cross-validation, and the repeated cross-validation. The model calibration process was performed using the RF algorithm in the R package caret (Kuhn, 2008). For the out-of-bag and bootstrap calibration, we randomly partitioned the data into calibration (70%) and validation (30%) subsets, following other DSM works (Fongaro et al., 2018; Poppiel et al., 2019a). We also performed a descriptive statistics analysis on the calibration and validation subsets. For cross-validation methods we used the complete soil dataset.

We compared the results of the model calibration and selected the best model based on the r-squared (R^2), root mean square error (RMSE), Ratio of Performance to InterQuartile distance (RPIQ), and bias (Hawkins et al., 2003). Afterwards, we chose the best model and applied it to predict the soil attributes for the study area.

3.2.4.2. Application and validation of internal and external models

The calibration step was applied on the three soil datasets, the internal (clay and sand) and the external (Fe_2O_3 and TiO_2). The internal model contained samples inside the study area (Fig. 1), which also contained the parent material information. After testing the results of the four resampling methods, we selected the one with best performance to predict sand and clay values in the study area.

The external model was fitted with data from the Piracicaba region, and was applied at the study area (Fig. 2b). The use of external soil datasets was explored by Wetterlind and Stenberg (2010), which compared the performance of prediction models using local and external soil data. The results pointed that the best scenario would be the combination of local and external datasets, which was also addressed in other DSM works (Bellinaso et al., 2021; Brown, 2007; Guy et al., 2015; Nawar and Mouazen, 2017; Sankey et al., 2008). We decided to test the external model for the prediction of Fe_2O_3 and TiO_2 at the study area, since we aimed to relate the PM with these two attributes.

The validation of the digital maps of sand and clay was performed for all the tested resampling methods. In the cases of out-of-bag and bootstrapping, the remaining 30% data was compared with the predicted maps. For the methods using cross validation the complete dataset was applied, resulting in a single validation value. For the validation of Fe_2O_3 and TiO_2, we applied the RF model on the 30% remaining dataset, predicting new values. This was considered as an external validation, since these datasets were external from the study area.

3.2.5. Exploratory statistical analysis

In this section we present the analyses regarding the distribution of soil texture, SySI, and PM classes. These analyses aimed to evaluate how the different is the soil texture and satellite reflectance over basalt and sandstone, according to the geology map provided by the Paulista Institute of Technology with scale of 1:500,000 (IPT, 1981) (Fig. 1a).

3.2.5.1. Soil texture, SySI, and geological classes

We used the internal soil database evaluate the distribution of sand and clay values over sandstone and basalt PM classes. The same process was applied to evaluate the distribution of SySI's six bands over the two PM classes. All the pixels inside the two PM classes were accounted in the analysis. These data were submitted to a statistical test to evaluate the mean differences between the values over basalt and sandstone (Fig. 2 f).

To reduce the spatial dependency effect, we fitted a linear model using Generalized Least Squares (GLS) (Dormann et al., 2007), and applied it to an analysis of variance (ANOVA) ($p < 0.01$), measuring the mean value according with the geological classes (Fig. 2 f). We used the multcompView package (Graves et al., 2015) to perform the mean comparison test in R software (R Core Team, 2013).

3.2.5.2. Bare soil spectral classification

Different studies stated the distinct spectral behavior of soils developed from sandstones and basaltic rocks (Demattê et al., 2012, Demattê et al., 2015). Based on this assumption, and in the fact of the reference legacy geology map used also presents only two classes (sandstones and basalt) for the study area, we performed a two-classes supervised classification on the area.

Using ESRI ArcGIS 10.4, we collected SySI reflectance signatures of both geologies at the study area that was considered pattern spectral signatures. Using the patterns information, we applied a two-classes maximum likelihood supervised classification to reclassify the image in new two classes.

We performed a mean test of sand and clay content within the two new classes and calculated the accuracy between the resulting map and the IPT geological map (IPT, 1981). Dobos et al. (2013) applied a similar approach to map PM using satellite images (MODIS and SRTM).

The analysis focused on the potential to indicate areas where the conventional map could be improved. Field observations also gave the in-situ information about geology, allowing the link between field, spectra, attribute analysis and the legacy map (Table 1). Finally, we were able to determine the inconsistencies on the generalized map and improve it.

3.2.6. Soil parent material prediction

After a complete evaluation of the relationship between SySI and soil attributes, we went further to predict PM classes using terrain and satellite information (Fig. 2f). The soil survey in the study area collected information about the PM (Table 1), which allowed its use on the prediction of PM classes.

We used the field points with PM information to extract the reflectance values from SySI's six bands and the five terrain attributes (Table 2). We used the RF algorithm and tested the same resampling methods used on the prediction of soil attributes. We did not apply the digital soil maps in this step for three reasons. The first is that they are a product from the same covariables (Table 2), which would result in autocorrelation. The second reason is that the digital soil maps contain error, which makes the use of the original data more adequate. The last reason is that we aimed to promote the use of free environmental data, such as SySI and terrain variables, to improve legacy maps and make quality information more accessible.

To validate the model, we analyzed the importance of each variable and calculated the accuracy and kappa coefficients (Congalton and Green, 2019). The final map was compared with the geology legacy map (1:500,000) and with the digital maps of soil attributes in order to identify transitional zones between the two classes.

3.3. Results and discussion

3.3.1. Exploratory analysis

3.3.1.1. Descriptive statistics of soil attributes

The analysis helped to understand the soil texture distribution over the study area at 0 – 20 cm depth (Table 3). In general, surface horizons were sandy loam. This is a result of an expressive presence of sandstone at the higher landscape positions (Fig. 1). Although there are different soil types in the area (Fig. 1b), clayey surface horizons are minority and are located at the lower altitudes where the basalt presence is stronger (Fig. 1c). Moreover, clay provided higher variability (higher CV), which shows its distribution is more heterogeneous than sand particles. Having sandy loam soils as dominant would prevent the study of geological transitions.

Table 3: Descriptive statistics of sand, clay, Fe_2O_3 and TiO_2 separated in calibration and validation datasets for digital soil mapping.

	n	Min.	Median	Max.	Mean	SD	CV (%)
Calibration							
Clay (g kg^{-1})	688	45	180	705	203.3	92.4	45
Sand (g kg^{-1})	688	111	760.5	945	729.8	122.0	17
Fe_2O_3 (g kg^{-1})	361	0.8	24.8	252.0	60.4	65.0	107.6
TiO_2 (g kg-1)	216	1.1	7.5	53.7	15.6	15.0	96.1
Validation							
Clay (g kg^{-1})	295	60	180	585	202.2	94.8	47
Sand (g kg^{-1})	295	231	762	918	730.9	125.6	17
Fe_2O_3 (g kg^{-1})	121	3.6	38.2	247.0	64.3	61.8	96.2
TiO_2 (g kg-1)	72	1.4	6.7	58.7	15.5	16.0	103.5

n: number of observations; min: minimum value; max: maximum value; SD: standard deviation; CV: coefficient of variation.

3.3.1.2. Sand, clay, and SySI distribution

The mean comparison test shows clay and sand contents as well as spectra over basalt and sandstone (Fig. 3a,c). In general, soil texture and SySI values were all statistically different for each geological class, which points to the PM influence on soil formation (Birkeland, 1984; Schaetzl and Anderson, 2005).

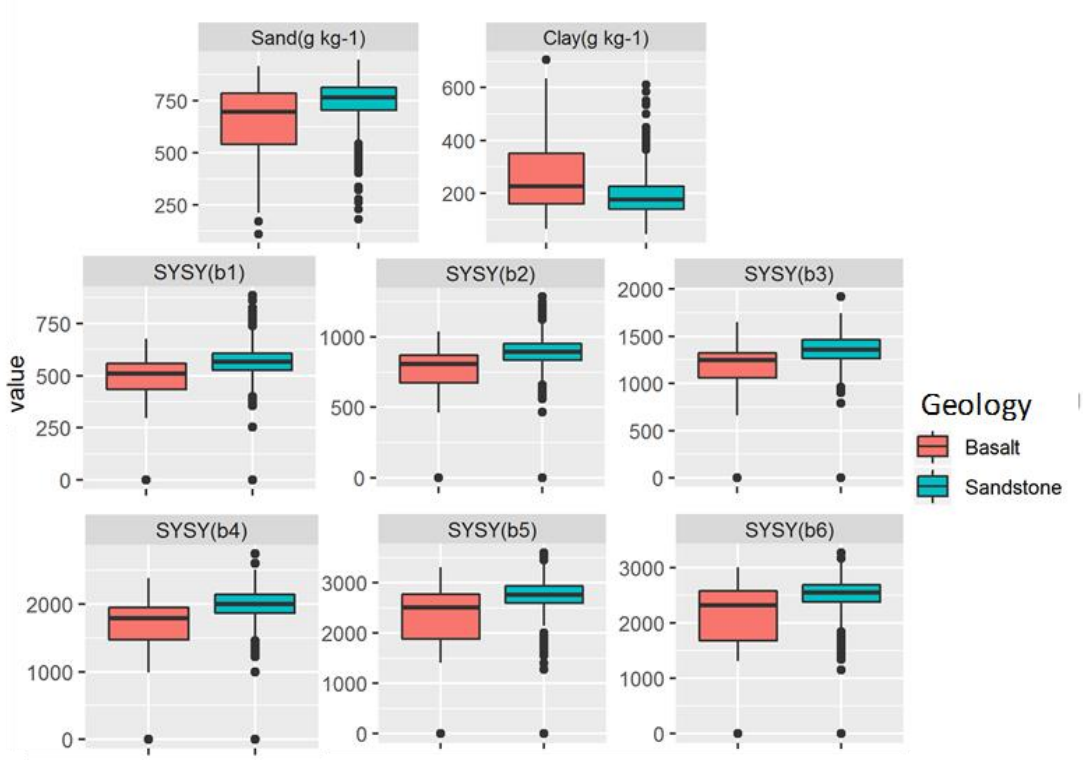

Figure 3: Multi comparison test performed to evaluate the mean difference between lithologies related to soil texture and SYSI bands. The Tukey test was performed with 0.05 of significance level. The test showed significant differences between geological formations for all the variables applied.

The mean clay values of points over basalt were low (268 g kg-1) (Fig. 3). This value over basalt (268 g kg^{-1}) was not expected, since mafic igneous rocks have low silica and high ferromagnesian minerals content (Meschede and Warr, 2019; Waroszewski et al., 2019). These minerals, after weathering and pedogenesis in tropical environments, result in the formation of oxidic clayey minerals (Schwertmann, 1988) and clay soil texture (Neel et al., 2007). These results are corroborated by (He et al., 2008), who found clayey to very clayey texture in basalt-derived soils attributing it to the mineralogy of the rock's constituent.

The low average value of clay over basalt (Fig. 3) could be explained by different process, such as landscape dynamics where erosion in the upper landscape positions (sandstone) and deposition in the lower (basalt). The interruption of pedogeochemistry activity by geomorphic processes, hampering mineral formation during basalt weathering (Waroszewski et al., 2019). Finally, clay accumulation processes that can result in textural differentiation in some soil classes (Breemen and Buurman, 2002). In addition, many of the collection points are located in transition areas between basalt and sandstone (Fig 1a). These observations, plus the low scale of the geological map, may cause a poor representation of reality.

The sandy texture of soils over the sandstone lithological compartment can be explained by the mineralogy of the *Santo Anastácio* formation, which predominates high levels of silica and mineral-poor cementing (Stradioto and Chang, 2020; Suguio et al., 1984). The predominance of sand over sandstone lithology indicates the influence of the PM for soil physical attributes. A similar result was found by Araujo et al. (2017), indicating that soil texture and mineralogy vary considerably over different PM. Cámara et al. (2017) explained that the soil texture heterogeneity is controlled by the lithology.

Soil reflectance extracted from satellite has been used to identify different minerals in soils (Demattê et al., 2007; Poppiel et al., 2019b). Opaque minerals, often derived from mafic rocks, tend to absorb energy and express a lower reflectance, while quartz minerals reflect most of the transmitted energy (Formaggio et al., 1996; Bellinaso et al., 2010).

The six SySI bands showed significant differences in terms of geology, even with the map's coarse resolution (Fig. 3). Bands one, two and three, inside the visible (VIS) spectral range are red, green and blue (RGB) and are often approached in the literature related to soil mineralogy and organic matter (Demattê et al., 2007; Madeira Netto, 1996). The result shows the mean reflectance on sandstone derived soils was higher for these three bands (Fig. 3). Band four represents the near infrared region (NIR) which is often related to vegetation characteristics (Demattê et al., 2007; Boettinger et al., 2008).

Bands five and six represent the short wave infrared (SWIR) and are related to soil texture due to the expression of phyllosilicates, quartz, and gibbsite (Boettinger et al., 2008; Janik et al., 2007). The variation observed on these two spectral ranges was more expressive than the others (Fig. 3).

3.3.1.3. Sand, clay, and SySI distribution

The darker color in Figure 4 is a result of more opaque minerals, derived from mafic rocks, which decreases the reflectance. The blue line shows the geological transition from the geology map (IPT, 1981) and the soil reflectance suggests an overestimation of the basalt area (Fig. 4).

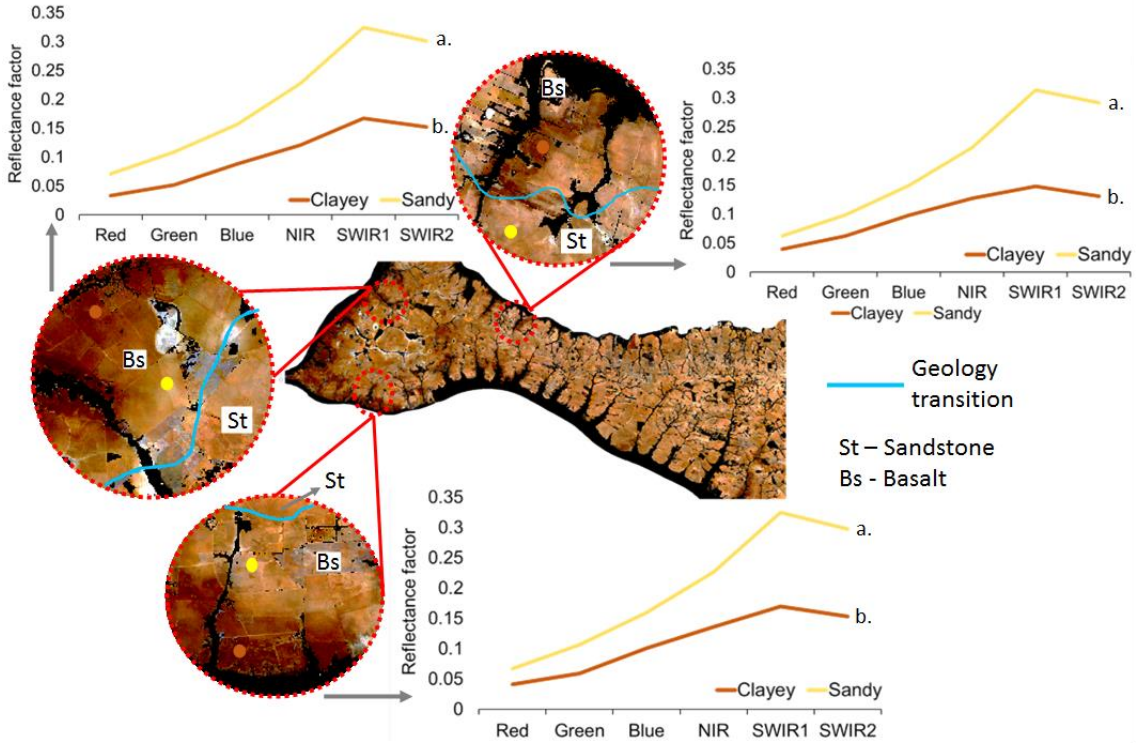

Figure 4: Three spectral patterns derived from SYSI. a. high clay content soils with low reflectance intensity and high sand content soils with high reflectance intensity.

Soils originated from basalt with expressive clay content (>35 %), often contain Fe oxides and are able to retain water (Silva et al., 2020). All these aspects influence the spectral reflectance decreasing its intensity in soils due to the presence of opaque mineral (Fig 4a). Sandy soils have low clay content (<15 %), primary minerals like feldspar and quartz, which contribute to a low water retention(Saxton et al., 1986).

We zoomed some areas to compare the geological transition and the reflectance values (Fig. 4). The RGB region showed lower intensity for basalt derived soils, indicating darker soils with high clay and oxides content (Fig. 4b) while soils with lower clay content and intense presence of quartz express higher reflectance (Fig. 4a). The spectral signatures in the study area is similar to the results found by Gallo et al. (2018) who studied diabase, siltstone and sandstone which the mean reflectance was different for all according to Tukey test ($p < 0.05$).

After performing the supervised classification on SySI, we observed that two distinct classes were formed (Fig. 5). The spectral signatures produced two classes (1 and 2), similar to the geology map (Fig. 5a). As expected, the sand content was higher in class 2, while the clay content was higher in class 1 (Fig. 5b). We also performed a comparison between the two new classes with the geology classes from the legacy map (IPT, 1981) reaching 81% of accuracy (Fig. 5).

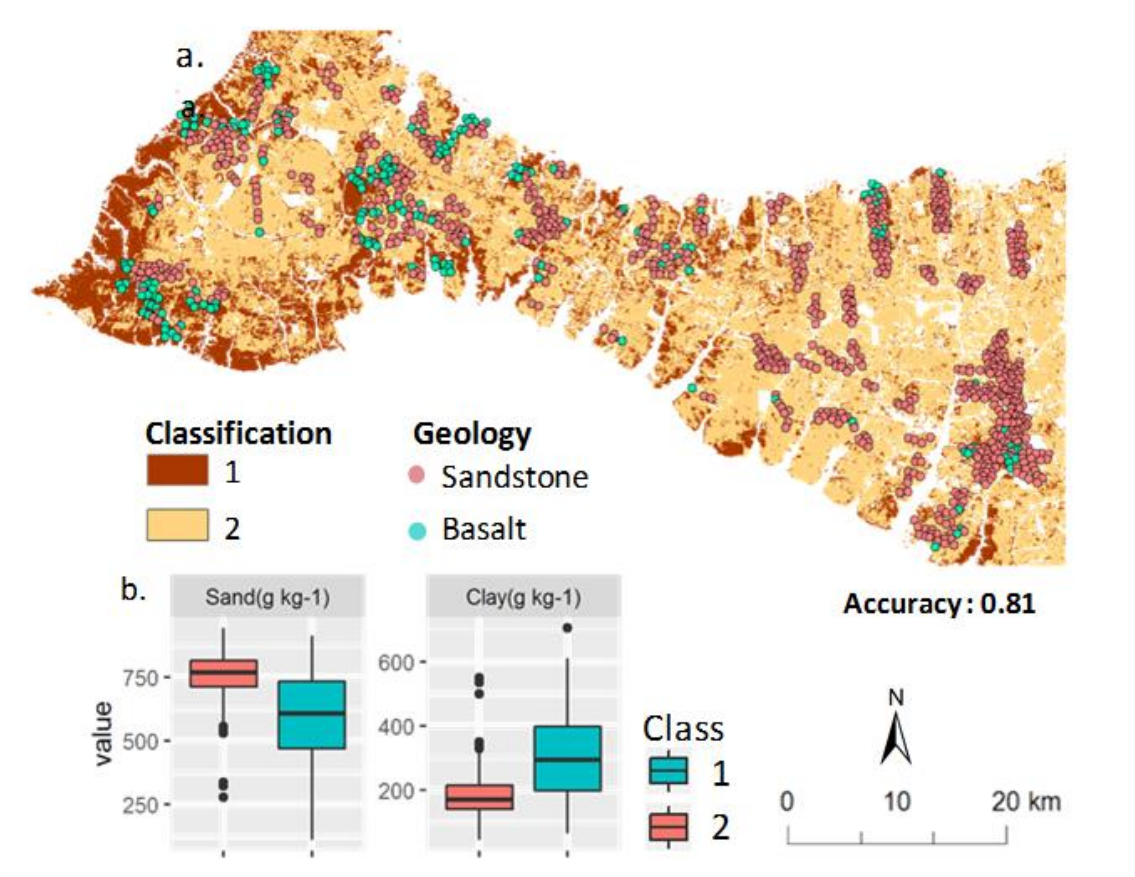

Figure 5: SYSI classification compared to geology and soil texture. a. SYSI's supervised classification using spectral signatures from basalt and sandstone areas. The points were classified based on the geological map. b. Multi comparison test between soil texture and the supervised classification.

This discrepancy was appointed by Gallo et al. (2018) and (Bonfatti et al., 2020) which compared the mean reflectance in soils developed from different PM and found a significant differences between igneous and sedimentary rocks. These results confirm the potential to use bare soil reflectance to detect the soil's parent material and show that surface processes, such as erosion, do not compromise the analysis. Poppiel et al. (2019b) and Gray et al. (2016) evaluated the variation of soil attributes over different parent materials and reached the conclusion that physical and chemical attributes are considerably different over discrepant geologies and can be used as an indicator of PM.

3.3.2. Model evaluation

3.3.2.1 Sand and clay prediction (internal model)

The use of multitemporal satellite information for environmental analysis is a powerful way to access and monitor natural resources (Charrua et al., 2021; Varin et al., 2021; Vivekananda et al., 2021). SySI has proved to aid significant improvement on DSM, especially regarding soil texture (Bellinaso et al., 2021; Demattê et al., 2018). However, its application on PM mapping is still challenging, since satellites only retrieve surface spectra (Vrieling, 2006). In this section we present digital maps of four soil attributes and compare them with a geology map, in order to demonstrate SySI's potential to map attributes associated with the PM.

After testing four resampling methods, we analyzed the results based on the coefficients of determination and errors (Table 4). We partitioned the dataset on 70% calibration and 30% validation for the use of out-of-bag and bootstrapping resampling methods. Although they had an overall low bias, their results were inferior to the cross-validation methods (Table 4). The ten-fold cross-validation was slightly better than the repeated cross-validation, with R^2 of 0.53 for clay, and 0.54 for sand prediction (Table 4). The RMSE and RPIQ were also similar in both cross-validation methods, which did not compromise the model performance.

Table 4: Evaluation of model predictions for sand and clay contents using the internal soil dataset. Tests of four resampling methods for calibrating and validating the model.

Attribute		Calibration				Validation			
		R^2	RMSE	RPIQ	bias	R^2	RMSE	RPIQ	bias
Clay ($g\ kg^{-1}$)	boot	0.52	67.5	1.18	0.1	0.44	66.3	0.53	-1.52
	oob	0.54	63.8	1.3	-0.25	0.44	65.87	0.53	-2.83
	10-cv	-	-	-	-	0.53	84.0	0.35	-1.5
	rep	-	-	-	-	0.49	66.6	0.35	-1.5
Sand ($g\ kg^{-1}$)	boot	0.53	86.2	2.15	0.49	0.44	86.3	0.88	2.21
	oob	0.56	82.7	2.17	0.03	0.44	87.23	0.87	3.03
	10-cv	-	-	-	-	0.54	83.9	0.19	3.07
	rep	-	-	-	-	0.53	87.1	0.18	2.12

R^2: r-squared; RMSE: rooth mean squared error; RPIQ: Ratio of Performance to InterQuartile distance; boot: bootstrapping; oob: out-of-bag; 10-cv: ten-fold cross-validation; rep: repeated cross-validation.

The results indicated that ten-fold cross-validation was the best model for the prediction of clay and sand. Thus, we applied this model in the study area and produced the digital map of these attributes (Fig. 6).

These maps contribute for the application of SySI to infer the PM, since the soil texture is mainly affected by the rock and the soil formation processes that occur (Breemen and Buurman, 2002). For instance, basalt and diabase from Serra Geral formation produce several ferromagnesian minerals (Mullins, 1977) that undergoes intense weathering in a tropical environment, forming silicate and oxidic clay minerals and soils with a clayey to very clayey texture and red/dark color (De Jong et al., 2000; Schwertmann, 1988). The sandstone of the Santo Anastácio formation, has high levels of silica and low levels of primary minerals susceptible to pedogenesis of clay after weathering (Suguio et al., 1984). Thus, the formation of sandy to medium sandy soils predominates over sandstones, as found by (Campos et al., 2007). It is clear that soil spectra are much affected by these opaque minerals (Fig. 4b) raising the argument of a correct PM classification studying the soil.

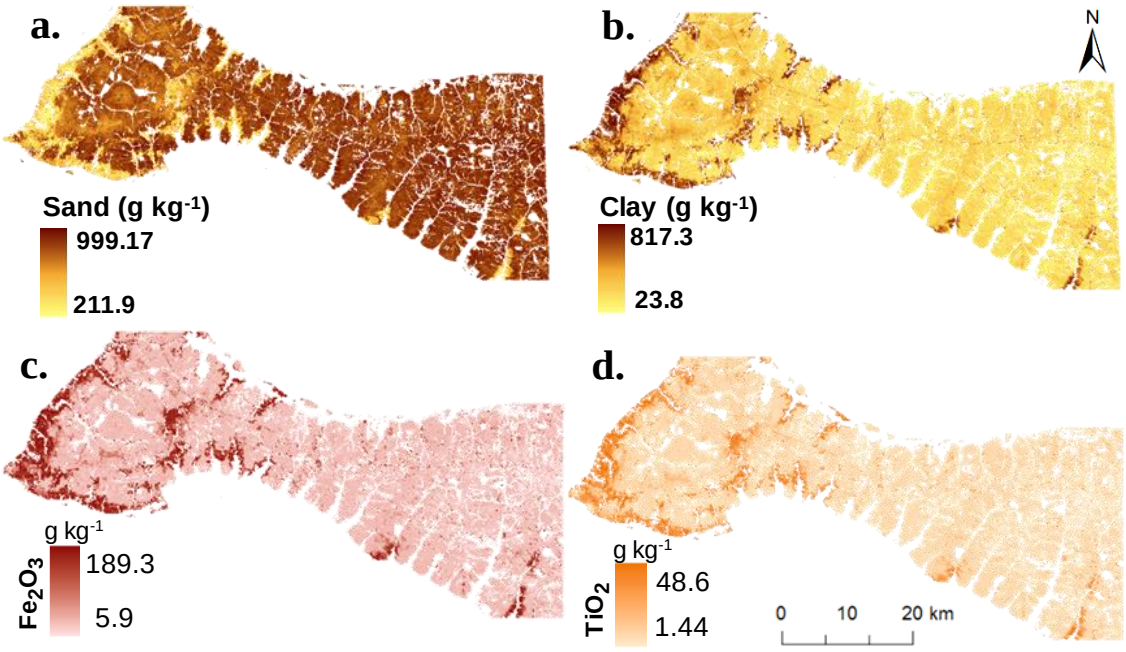

Figure 6: Digital soil maps of soil attributes using random forest. a. sand, b. clay, c. Fe_2O_3 and TiO_2.

However, sandstones from the Santo Anastácio Formation have high levels of silica and low levels of ferromagnesian minerals. As a result, sand formation is more favorable than clay. Therefore, these minerals tend to form soils with sandy texture and lighter color (Campos et al., 2007). These physical attributes will influence soil directly and will also affect the spectral reflectance increasing its values (Fig 4).

3.3.2.2 Fe2O3 and TiO2 prediction (external model)

The model performance for Fe_2O_3 and TiO_2 prediction was generally better than clay and sand (Table 5). The calibration coefficients for out-of-bag and bootstrapping were overly optimistic with R^2 of 0.93 (Fe_2O_3) and 0.92 (TiO_2), and a strong decrease on validation with R^2 0.64 (Fe_2O_3) and 0.69 (TiO_2) (Table 5). The cross-validation resampling methods proved to be more efficient on the prediction of these attributes, specifically the repeated cross-validation, with producing R^2 of 0.71 (Fe_2O_3) and 0.74 (TiO_2), and RMSE of 36.9 (Fe_2O_3) and 7.9 (TiO_2) (Table 5).

Table 5: Evaluation of model predictions for Fe2O3 and TiO2 contents using the external soil dataset. Tests of four resampling methods for calibrating and validating the model.

Attribute		Calibration				Validation			
		R^2	RMSE	RPIQ	bias	R^2	RMSE	RPIQ	bias
	boot	0.93	17.92	0.51	0.27	0.64	38.96	0.25	0.15
Fe_2O_3	oob	0.93	17.9	0.51	0.06	0.64	39.03	0.25	0.13
(g kg^{-1})	10-cv	-	-	-	-	0.67	35.9	0.27	1.2
	rep	-	-	-	-	0.71	36.9	0.61	0.71
	boot	0.92	4.4	0.54	0.03	0.69	8.12	0.27	0.13
TiO_2	oob	0.92	4.39	0.54	0.04	0.69	8.16	0.27	0.1
(g kg-1)	10-cv	-	-	-	-	0.63	9.16	0.25	1.51
	rep	-	-	-	-	0.74	7.9	0.62	-0.07

R^2: r-squared; RMSE: rooth mean squared error; RPIQ: Ratio of Performance to InterQuartile distance; boot: bootstrapping; oob: out-of-bag; 10-cv: ten-fold cross-validation; rep: repeated cross-validation.

The Fe_2O_3 and TiO_2 predictions used an external soil dataset, which prevented the validation process. Demattê et al. (2018) and Bellinaso et al. (2021) used this approach for clay quantification and demonstrated the potential of this method specially when applied in multitemporal bare soil images. Fe_2O_3 and TiO_2 maps are important for an exploratory analyses, because soils from São Paulo state, developed from a mafic rocks will have higher contents, if compared with soils derived from a sedimentary rock, such as sandstone (Schaetzl and Anderson, 2005; Breemen and Buurman, 2002).

The digital maps of clay, sand, Fe_2O_3 and TiO_2 content presented strong contrast between high and low landscape positions (Fig. 6). The higher positions have a stronger influence of sedimentary rocks from Santo Anastácio formation. As the altitude of landscape decreases, the drainage pathways are more evident and higher clay contents are more evident and higher Fe_2O_3 and TiO_2 contents are exposed (Fig. 6b). This process suggests geomorphological activity, which a steep landscape favors surface erosion, lowering the landscape and uncovering older rocks that had previously been covered by younger ones. In this case, the exposed rocks close to the drainage valleys were from Serra Geral formation, favoring the formation of iron oxides in soils (Ker et al., 2015).

Furthermore, there is presence of magnetite and ilmenite in the sand fraction, opaque minerals. On the other hand, sandstones from the Santo Anastácio Formation have high levels of silica and low of ferromagnesian minerals. The quartz presence results soils with lighter color (Campos et al., 2007) and increase reflectance values (Fig 4).

The Fe_2O_3 concentration in soils (Fig. 6c) is related to the PM, weathering and pedogenetic processes of accumulation and removal (Ker et al., 2015). In fact, PM stands out as the main supplier for Fe_2O_3 formation (Carvalho Filho et al., 2015; Schwertmann, 1993). Silva et al. (2020) worked with soils from Western Paulista Plateau and found higher Fe_2O_3 in soils developed from basalt rocks from Serra Geral formation and the minor values were found in sandy soils derived from sandstones.

Cunha et al. (2005) studied soils derived from the same geological formations and found a pattern similar to the encountered in this study, with soils developed from sandstones in higher landscape positions, while in lower positions the basalt is affecting the soil formation. Soils developed from sandstone contained low Fe_2O_3 contents and as the altitude decreased, clay content became higher and soils showed an eutrophic character (Cunha et al., 2005). At valley positions soils were ferric and clayey, indicating a strong influence of basalt. The clay (Fig. 6b) and

the Fe_2O_3 (Fig. 6c) maps highlight the same pattern in which there is a progressive increase of these attributes as the landscape is steepened.

TiO_2 is a mineral inherited from igneous and metamorphic rocks and appears normally in low concentration in soils with heavy minerals due to its resistance to weathering. However, some soils that are very weathered, leachate and derived from mafic rocks might have a higher accumulation of these oxides compared to temperate soils (Ker et al., 2015). Similar to our results, Nogueira et al. (2018) found a negative correlation between TiO_2 and sand contents of soils at the São Paulo State.

The mafic rocks from Serra Geral formation, tend to be rich in TiO_2 as result of their own natural formation (crystallization from liquids at 1000°C in the deep crust) (Green and Pearson, 1986). This rock is highly weathered, when exposed to natural conditions, remaining only TiO_2 rich minerals (e.g. Zircon, Titanomagnetite and Titanomaghemite), which are resistant to weathering. Schaefer et al. (2008) explained that for amphibolite (or dolerite), tholeiitic basalt and diabase, Ti or Al are the main isomorphic substituents in the Fe oxides.

3.3.3. Soil parent material prediction

The exploratory analysis suggested that the soil physical attributes are significantly different over basalt and sandstone (Fig. 3). The literature also points to a series of pedogenetic processes affecting the contents of sand, clay, Fe_2O_3, and TiO_2 (Cunha et al., 2005; Silva et al., 2020). In this regard, SySI has proved the potential to predict soil attributes associated to the PM. Therefore, we moved forward and applied SySI for the prediction of PM classes.

The discrepant geological formations and the only two PM classes in the study area positively influenced the prediction performance of the model (Table 6). We also tested the four resampling methods linked with RF, the same used for the DSM step. In this case the cross-validation methods had significant lower performance, favoring the use of the partitioned datasets. The best model used bootstrapping, resulting in 0.85 of accuracy and 0.49 of kappa coefficient for calibration, and 0.75 of accuracy and 0.40 of kappa coefficient for validation (Table 6).

Table 6: Evaluation of model performance and resampling method for the prediction of parent material classes.

Attribute		Calibration		Validation	
		Acc[2]	Kappa	Acc[2]	Kappa
PM Classes[1]	boot	0.85	0.49	0.75	0.40
	oob	0.91	0.63	0.57	0.23
	10-cv	-	-	0.26	0.03
	rep	-	-	0.27	0.04

[1] Parent material classes; [2] prediction accuracy; boot: bootstrapping; oob: out-of-bag; 10-cv: ten-fold cross-validation; rep: repeated cross-validation.

Although the low number of classes improved the model performance, other works presented strong results applying a similar strategy in more complex areas (Bonfatti et al., 2020; Gray et al., 2016; Kassai and Sisák, 2018). Mancini et al. (2020) used a proximal sensing to classify 12 PM classes using x-ray fluorescence information of soils. Based on these results and after exploring the differences in soil texture, mineralogy and satellite reflectance over basalt and sandstone, we confirmed that the environmental variables (Table 2) could be used to determine the PM through DSM techniques.

Although there was available soil analysis, we used SySI and a set of terrain variables to fit the RF model. SySI bands had the highest contribution followed by elevation data (Fig. 7). The variables importance graphic

indicates the strong input of bare soil reflectance, which proves that the bare soil reflectance is capable to identify soil variation and, consequently, parent material (Fig. 7). The digital elevation model was the most important terrain variable on the model (Fig. 7). The performance of the other terrain variables was similar to the results presented by Heung et al. (2014), which used terrain information to predict nine geological classes.

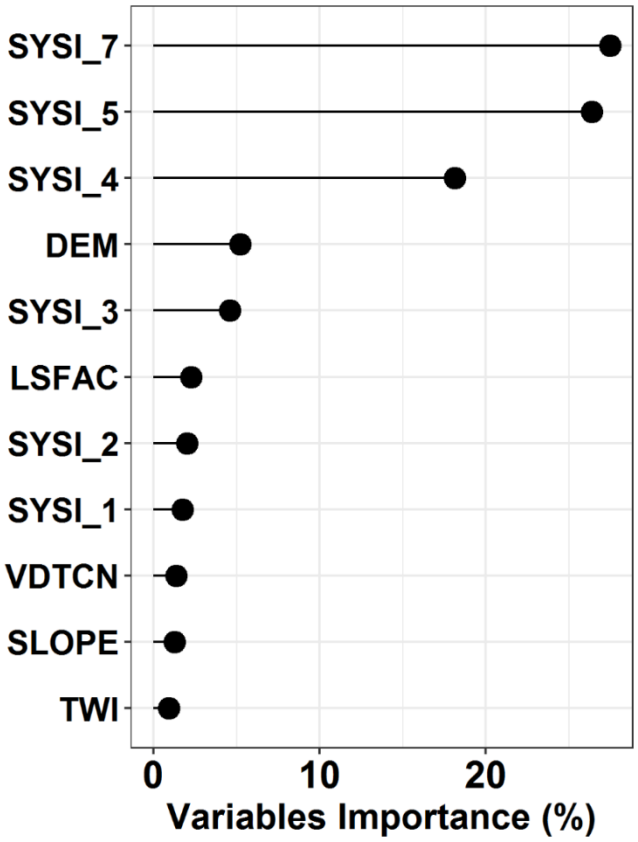

Figure 7: Variables importance on the prediction of the soil's parent material using random forest machine learning algorithm. SySI_1: SySI's band 1, SySI_2: SySI's band 2, SySI_3: SySI's band 3, SySI_4: SySI's band 4, SySI_5: SySI's band 5, SySI_7: SySI's band 7, DEM: digital elevation model, LSFAC: LS factor, SLOPE: terrain slope, VDTCN: vertical distance to channel network, TWI: topographic wetness index.

As stated by Fongaro et al. (2018), SySI's bands 4 (NIR), 5 (SWIR1), and 7 (SWIR2) respond to iron oxides, quartz, and moisture contents, explaining their good performance on the model prediction (Fig. 7). Hematite, kaolinite, and gibbsite contents are higher in clayey soils, decreasing and flattening the spectral curve, while higher contents of quartz increases the reflectance across the spectrum (Demattê et al., 2018).

The application of modern DSM techniques to predict PM classes is an update of traditional studies on landforms and bedrock distribution (Brevik and Miller, 2015). The early applications of soil knowledge on geological studies guided field surveys on the recognition of different geological formations and to fill gaps between existing geology maps (Rodgers, 1953; Thorp, 1949). Although soil maps were used to provide extra data, it was an important tool in periods where information was not easily accessible and available (Birkeland, 1999; Lindholm, 1994; Prokopovich, 1984).

Recent use of soil data brought more applications for geological studies (Dobos et al., 2013; Florea et al., 2015). Miller and Burras (2015) obtained good agreement of surficial geology maps derived from Soil Survey maps with surficial geology maps. Jang et al. (2021) used soil texture maps, topographic data, gamma-ray data, and

multitemporal bare soil image, classified into clusters to generate a PM map for a 1700 km² region in Australia. Richter et al. (2019) combined expert knowledge with quantitative digital terrain attributes, and digital soil mapping to predict PM classes, reaching an overall accuracy of 0.79 among pedisediment and residuum. Lacoste et al. (2011) also used machine learning to predict 20 PM classes, reaching 0.78 of kappa coefficient, and suggesting the method's application for other environmental studies.

The predicted map had similarities with SySI itself, clay, sand, Fe_2O_3, TiO_2, and the supervised classification maps (Fig. 8). This result is in agreement with our hypothesis, since there is a strong relationship between the contents of these attributes, spectral behavior, and PM (Dobos et al., 2013; Kassai and Sisák, 2018). The predicted map also highlights the overestimation of basalt areas in the legacy map with 1:500,000 scale (IPT, 1981) (Fig. 8b). Besides the clayey soils in the lower landscape positions, some areas maintain the sandy surface horizon from upper to lower parts (Fig. 8b). This high sand content is detected by SySI with higher reflectance values (Fig. 4) but it might add an error to the analysis. The sandy surface horizons can be formed from sediment deposition from upper positions and not represent correctly the PM.

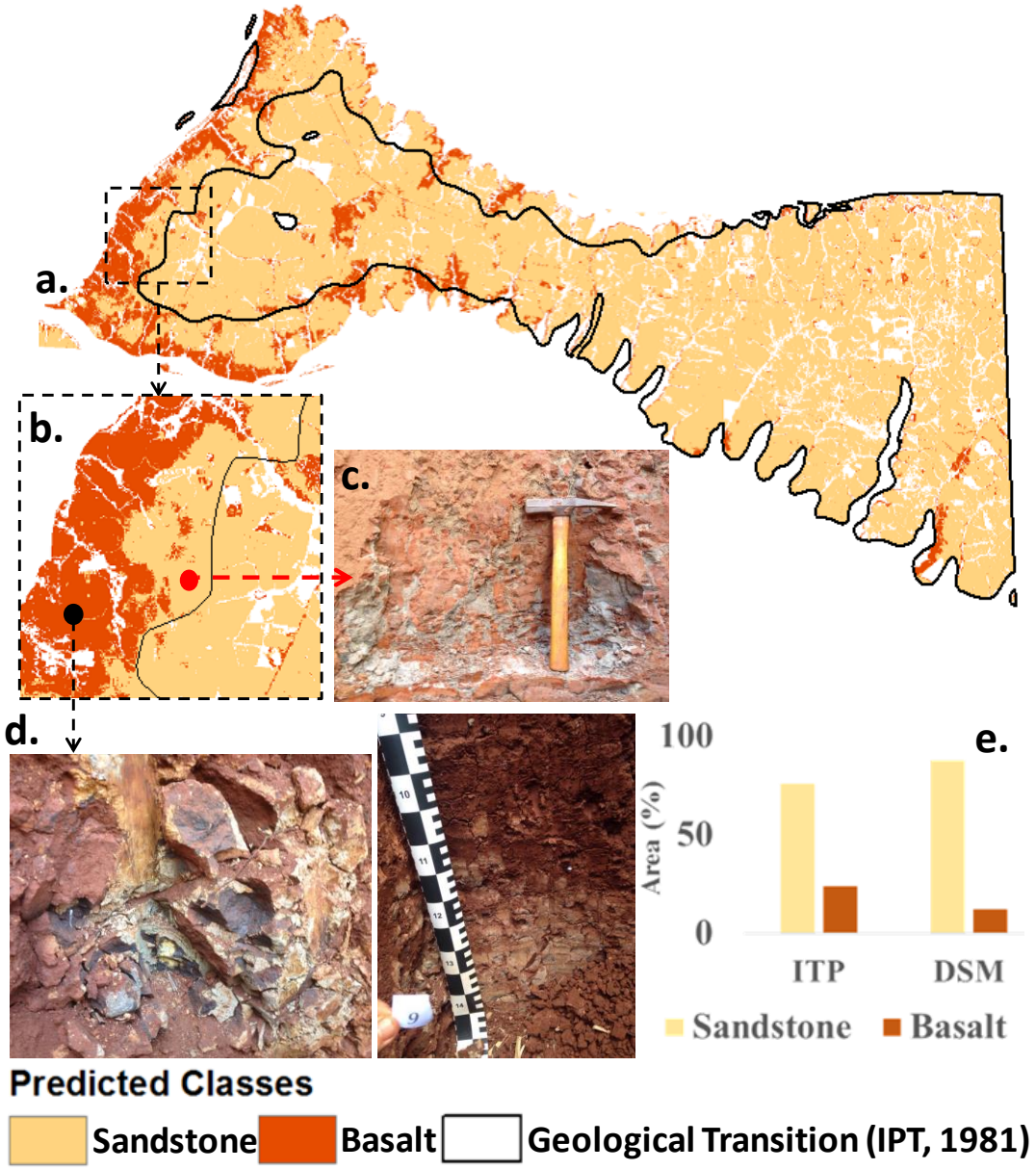

Figure 8: Soil parent material prediction using random forest. a. digital map of parent material classes with geological transition from a 1:500,000 geological map (IPT, 1981); b. highlight on the differences in parent material transition between the conventional and the digital map; c. sandstone rock observed in a soil profile; d. basalt rock observed in a Cambisol profile and e. area proportions between the conventional and digital maps.

Campos et al. (2012, 2007) in works carried out in the same region, described the pattern of increasing clay content from the surface from the highest to the lowest parts of the relief, following the transition from PM sandstone to basalt. The same authors highlighted the occurrence of a transition zone where the soils were influenced by the two materials in their formation.

The high accuracy and kappa coefficients indicate that soil information can be used to predict the PM, offering new applications for soil surveys and DSM techniques. The improvement in PM delineation can help on multiple agricultural practices and in land conservation. In this case, we had a decrease in basalt areas from 23 to 12% (Fig. 8e), an important information for land use planning. The sandstone areas increased from 76 to 87% (Fig. 8e), showing a stronger role on soil formation around the study area. Land use planning would be different in these two scenarios and a more adequate one could be developed from this technique.

3.4. Conclusions

The digital soil mapping techniques are advancing towards the assessment of geological information with focus on the PM. This fast way to classify SySI showed a potential to identify different geological formations through soil. Using a more detailed geological map would bring even more possibilities to identify better transitional areas and geological polygons.

The surface reflectance was indicative of soil PM due to their relationship with soil formation and attributes. The textural and mineralogical aspects of soils tended to differ from mafic and sedimentary rocks, affecting directly the reflectance. Specifically, the greater presence of quartz and greater sand content in soils developed from sandstones and the greater presence of opaque minerals (magnetite, ilmenite and titanomagnetite) and greater sand content in soils developed from basalt. Satellite images worked as proxies to infer the PM, since they retrieve surface reflectance.

The environmental variables worked satisfactorily to indicate lithological transitions as a result of the textural and mineral distribution over the landscape and the effects in surface reflectance. In areas where geological maps have low scale, the automated method using environmental variables offers a great advance in saving surveying time and costs, due to its capability to accurately identification of the transitions.

The prediction of PM classes by bare soil reflectance (SySI), terrain attributes and soil data proved to be efficient and accurate. This approach can improve a set of land use decisions if the procedures and variables applied are adequate and the use of the machine learning is correct.

Acknowledgments

We are grateful to Geotechnologies in Soil Science Group (GEOSS/ESALQ-USP; http://esalqgeocis.wixsite.com/english) for team support. We also thank the anonymous reviewers to improve this paper. The scholarship of the first author was provided by the Coordination of Superior Level Staff Improvement (CAPES). Funding was provided by São Paulo Research Foundation (FAPESP, grant numbers 2014/22262-0,2016/01597-9, 2016/26124-6, 2016/01597-9, 2018/12532-0).

References

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

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

Araujo, M.A., Zinn, Y.L., Lal, R., 2017. Soil parent material, texture and oxide contents have little effect on soil organic carbon retention in tropical highlands. Geoderma 300, 1–10.

Baumgardner, M.F., Silva, L.R.F., Biehl, L.L., Stoner, E.R., 1986. Reflectance properties of soils. Adv. Agron. 38, 1–44. https://doi.org/10.1016/S0065-2113(08)60672-0

Bellinaso, H., Demattê, J.A.M., Romeiro, S.A., 2010. Soil spectral library and its use in soil classification. Rev. Bras. Ciência do Solo 34, 861–870. https://doi.org/10.1590/s0100-06832010000300027

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

Birkeland, P.W., 1999. Soils and Geomorphology (Book Review). Geogr. Bull. 41, 121.

Birkeland, P.W., 1984. Soils and geomorphology. Oxford University Press.

Blatnik, M., Culver, D.C., Gabrovšek, F., Knez, M., Kogovšek, B., Kogovšek, J., Liu, H., Mayaud, C., Mihevc, A., Mulec, J., 2020. Lithomorphogenesis of Karst Surface, in: Karstology in the Classical Karst. Springer, pp. 33–98.

Boettinger, J.L., Ramsey, R.D., Bodily, J.M., Cole, N.J., Kienast-Brown, S., Nield, S.J., Saunders, A.M., Stum, A.K., 2008. Landsat spectral data for digital soil mapping, in: Digital Soil Mapping with Limited Data. Springer, pp. 193–202.

Bonfatti, B.R., Demattê, J.A.M., Marques, K.P.P., Poppiel, R.R., Rizzo, R., Mendes, W. de S., Silvero, N.E.Q., Safanelli, J.L., 2020. Digital mapping of soil parent material in a heterogeneous tropical area. Geomorphology 367, 107305. https://doi.org/https://doi.org/10.1016/j.geomorph.2020.107305

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1007/9781441993267_5

Brevik, E.C., Miller, B.A., 2015. The Use of Soil Surveys to Aid in Geologic Mapping with an Emphasis on the Eastern and Midwestern United States. Soil Horizons 56, sh15-01–0001. https://doi.org/10.2136/sh15-01-0001

Brown, D.J., 2007. Using a global VNIR soil-spectral library for local soil characterization and landscape modeling in a 2nd-order Uganda watershed. Geoderma 140, 444–453. https://doi.org/https://doi.org/10.1016/j.geoderma.2007.04.021

Caldarelli, C.E., Gilio, L., 2018. Expansion of the sugarcane industry and its effects on land use in São Paulo: Analysis from 2000 through 2015. Land use policy 76, 264–274.

Cámara, J., Gómez-Miguel, V., Martín, M.Á., 2017. Lithologic control on soil texture heterogeneity. Geoderma 287, 157–163.

Campos, M.C.C., Junior, J.M., de Souza, Z.M., Siqueira, D.S., Pereira, G.T., 2012. Análise multivariada de atributos do solo na discriminação de superfícies geomórficas em uma litossequência arenito-basalto. Rev. Ciência Agronômica 43, 429–438.

Campos, M.C.C., Marques, J., Pereira, G.T., Montanari, R., Camargo, L.A., 2007. Relações solo-paisagem em UMA litosseqüência arenito-basalto na Região de Pereira Barreto, SP. Rev. Bras. Cienc. do Solo 31, 519–529. https://doi.org/10.1590/s0100-06832007000300012

Carvalho Filho, A. de, Inda, A.V., Fink, J.R., Curi, N., 2015. Iron oxides in soils of different lithological origins in Ferriferous Quadrilateral (Minas Gerais, Brazil). Appl. Clay Sci. 118, 1–7. https://doi.org/https://doi.org/10.1016/j.clay.2015.08.037

Charrua, A.B., Padmanaban, R., Cabral, P., Bandeira, S., Romeiras, M.M., 2021. Impacts of the Tropical Cyclone Idai in Mozambique: A Multi-Temporal Landsat Satellite Imagery Analysis. Remote Sens. . https://doi.org/10.3390/rs13020201

Congalton, R.G., Green, K., 2019. Assessing the Accuracy of Remotely Sensed Data Principles and Practices, 3rd ed. Taylor & Francis Group, Boca Raton, FL. https://doi.org/10.1017/CBO9781107415324.004

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J., 2015. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. Geosci. Model Dev. 8. https://doi.org/10.5194/gmd-8-1991-2015

Cunha, P., Marques Júnior, J., Curi, N., Pereira, G.T., Lepsch, I.F., 2005. Superfícies geomórficas e atributos de Latossolos em uma seqüência Arenítico-Basáltica da região de Jaboticabal (SP) . Rev. Bras. Ciência do Solo .

De Jong, E., Pennock, D.J., Nestor, P.A., 2000. Magnetic susceptibility of soils in different slope positions in Saskatchewan, Canada. Catena 40, 291–305. https://doi.org/10.1016/S0341-8162(00)00080-1

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, Galdos, M. V, Guimarães, R. V, Genú, A.M., Nanni, M.R., Zullo Jr, J., 2007. Quantification of tropical soil attributes from ETM+/LANDSAT-7 data. Int. J. Remote Sens. 28, 3813–3829.

Dematte, J.A.M., Huete, A.R., Ferreira Jr., L.G., Nanni, M.R., Alves, M.C., Fiorio, P.R., 2009. Methodology for Bare Soil Detection and Discrimination by Landsat TM Image. Open Remote Sens. J. 2, 24–35. https://doi.org/10.2174/1875413900902010024

Demattê, J. A.M., Nanni, M.R., Formaggio, A.R., Epiphanio, J.C.N., 2007. Spectral reflectance for the mineralogical evaluation of Brazilian low clay activity soils. Int. J. Remote Sens. 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, 4461. https://doi.org/10.1038/s41598-020-61408-1

Dharumarajan, S., Kalaiselvi, B., Suputhra, A., Lalitha, M., Vasundhara, R., Kumar, K.S.A., Nair, K.M., Hegde, R., Singh, S.K., Lagacherie, P., 2021. Digital soil mapping of soil organic carbon stocks in Western Ghats, South India. Geoderma Reg. 25, e00387. https://doi.org/https://doi.org/10.1016/j.geodrs.2021.e00387

Dobos, E., Seres, A., Vadnai, P., Michéli, E., Fuchs, M., Láng, V., Bertóti, D., Kovács, K., 2013. Soil parent material delineation using MODIS and SRTM data. Hungarian Geogr. Bull. 62, 133–156.

dos Santos, H.G., Jacomine, P.K.T., Dos Anjos, L.H.C., De Oliveira, V.A., Lumbreras, J.F., Coelho, M.R., de Almeida, J.A., de Araujo Filho, J.C., de Oliveira, J.B., Cunha, T.J.F., 2018. Sistema brasileiro de classificação de solos. Brasília, DF: Embrapa, 2018.

Fernandes, L.A., de Castro, A.B., Basilici, G., 2007. Seismites in continental sand sea deposits of the Late Cretaceous Caiuá Desert, Bauru Basin, Brazil. Sediment. Geol. 199, 51–64. https://doi.org/10.1016/j.sedgeo.2005.12.030

Florea, N., Mocanu, V., Coteţ, V., Dumitru, S., 2015. Map of soil parent materials in Romania. Res. J. Agric. Sci. 47, 57–64.

Florinsky, I. V., 2012. Digital Terrain Analysis in Soil Science and Geology. Academic Press.

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. 10. https://doi.org/10.3390/rs10101555

Formaggio, A., Epiphanio, J., Valeriano, M., Oliveira, J., 1996. Comportamento espectral (450-2.450 nm) de solos Tropicals de Sao Paulo. Rev. Bras. ciência do solo 20, 467–474.

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. 10. https://doi.org/10.3390/rs10101571

Gambill, D.R., Wall, W.A., Fulton, A.J., Howard, H.R., 2016. Predicting USCS soil classification from soil property variables using Random Forest. J. Terramechanics 65, 85–92. https://doi.org/10.1016/j.jterra.2016.03.006

Garcia, M. da G.M., Brilha, J., de Lima, F.F., Vargas, J.C., Pérez-Aguilar, A., Alves, A., Campanha, G.A. da C., Duleba, W., Faleiros, F.M., Fernandes, L.A., Fierz, M. de S.M., Garcia, M.J., Janasi, V. de A., Martins, L., Raposo, M.I.B., Ricardi-Branco, F., Ross, J.L.S., Filho, W.S., Souza, C.R. de G., Bernardes-de-Oliveira, M.E.C., de Brito Neves, B.B., Campos Neto, M. da C., Christofoletti, S.R., Henrique-Pinto, R., Lobo, H.A.S., Machado, R., Passarelli, C.R., Perinotto, J.A. de J., Ribeiro, R.R., Shimada, H., 2018. The Inventory of Geological Heritage of the State of São Paulo, Brazil: Methodological Basis, Results and Perspectives. Geoheritage 10, 239–258. https://doi.org/10.1007/s12371-016-0215-y

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

Graves, S., Piepho, H.-P., Selzer, M.L., 2015. Package 'multcompView.' Vis. Paired Comp.

Gray, J.M., Bishop, T.F.A., Wilford, J.R., 2016. Lithology and soil relationships for soil modelling and mapping. Catena 147, 429–440. https://doi.org/10.1016/j.catena.2016.07.045

Green, T.H., Pearson, N.J., 1986. Ti-rich accessory phase saturation in hydrous mafic-felsic compositions at high P,T. Chem. Geol. 54, 185–201. https://doi.org/https://doi.org/10.1016/0009-2541(86)90136-1

Grimm, R., Behrens, T., Märker, M., Elsenbeer, H., 2008. Soil organic carbon concentrations and stocks on Barro Colorado Island — Digital soil mapping using Random Forests analysis. Geoderma 146, 102–113. https://doi.org/https://doi.org/10.1016/j.geoderma.2008.05.008

Guo, Z., Adhikari, K., Chellasamy, M., Greve, M.B., Owens, P.R., Greve, M.H., 2019. Selection of terrain attributes and its scale dependency on soil organic carbon prediction. Geoderma 340, 303–312. https://doi.org/10.1016/j.geoderma.2019.01.023

Guy, A.L., Siciliano, S.D., Lamb, E.G., 2015. Spiking regional vis-NIR calibration models with local samples to predict soil organic carbon in two High Arctic polar deserts using a vis-NIR probe. Can. J. Soil Sci. 95, 237–249. https://doi.org/10.4141/cjss-2015-004

Hawkins, D.M., Basak, S.C., Mills, D., 2003. Assessing Model Fit by Cross-Validation. J. Chem. Inf. Comput. Sci. 43, 579–586. https://doi.org/10.1021/ci025626i

He, Y., Li, D.C., Velde, B., Yang, Y.F., Huang, C.M., Gong, Z.T., Zhang, G.L., 2008. Clay minerals in a soil chronosequence derived from basalt on Hainan Island, China and its implication for pedogenesis. Geoderma 148, 206–212.

Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., Ruiperez Gonzalez, M., 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

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/https://doi.org/10.1016/j.geoderma.2013.09.016

Izawa, M.R.M., Cloutis, E.A., Rhind, T., Mertzman, S.A., Applin, D.M., Stromberg, J.M., Sherman, D.M., 2019. Spectral reflectance properties of magnetites: Implications for remote sensing. Icarus 319, 525–539. https://doi.org/https://doi.org/10.1016/j.icarus.2018.10.002

Jang, H.J., Dobarco, M.R., Minasny, B., McBratney, A., 2021. Creating a soil parent material map digitally using a combination of interpretation and statistical techniques. Soil Res.

Janik, L.J., Merry, R.H., Forrester, S.T., Lanyon, D.M., Rawson, A., 2007. Rapid prediction of soil water retention using mid infrared spectroscopy. Soil Sci. Soc. Am. J. 71, 507–514.

Kassai, P., Sisák, I., 2018. The role of geology in the spatial prediction of soil properties in the watershed of Lake Balaton , Hungary. Geol. Croat. 71, 29–39. https://doi.org/10.4154/gc.2018.04

Ker, J.C., Curi, N., Schaefer, C.E.G.R., Vidal-Torrado, P., 2015. Pedologia: fundamentos.

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

Kuhn, M., 2008. Building Predictive Models in R Using the caret Package. J. Stat. Software; Vol 1, Issue 5 . https://doi.org/10.18637/jss.v028.i05

Lacoste, M., Lemercier, B., Walter, C., 2011. Regional mapping of soil parent material by machine learning based on point data. Geomorphology 133, 90–99. https://doi.org/10.1016/j.geomorph.2011.06.026

Lagacherie, P., Arrouays, D., Bourennane, H., Gomez, C., Martin, M., Saby, N.P.A., 2019. How far can the uncertainty on a Digital Soil Map be known?: A numerical experiment using pseudo values of clay content obtained from Vis-SWIR hyperspectral imagery. Geoderma 337, 1320–1328. https://doi.org/https://doi.org/10.1016/j.geoderma.2018.08.024

Lamichhane, S., Kumar, L., Adhikari, K., 2021. Updating the national soil map of Nepal through digital soil mapping. Geoderma 394, 115041. https://doi.org/https://doi.org/10.1016/j.geoderma.2021.115041

Lindholm, R.C., 1994. Information derived from soil maps: Areal distribution of bedrock landslide distribution and slope steepness. Environ. Geol. 23, 271–275. https://doi.org/10.1007/BF00766742

Loiseau, T., Chen, S., Mulder, V.L., Román Dobarco, M., Richer-de-Forges, A.C., Lehmann, S., Bourennane, H., Saby, N.P.A., Martin, M.P., Vaudour, E., Gomez, C., Lagacherie, P., Arrouays, D., 2019. Satellite data integration for soil clay content modelling at a national scale. Int. J. Appl. Earth Obs. Geoinf. 82, 101905. https://doi.org/https://doi.org/10.1016/j.jag.2019.101905

Lu, S.-G., Xue, Q.-F., Zhu, L., Yu, J.-Y., 2008. Mineral magnetic properties of a weathering sequence of soils derived from basalt in Eastern China. CATENA 73, 23–33. https://doi.org/https://doi.org/10.1016/j.catena.2007.08.004

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

Madeira Netto, J.D.S., 1996. Spectral reflectance properties of soils. Photo Interpret. 34, 59–76.

Malone, B., Searle, R., 2021. Updating the Australian digital soil texture mapping (Part 2*): spatial modelling of merged field and lab measurements<a class="reftools" href="#FN1"/>. Soil Res.

Maltman, A., 2012. Geological maps: an introduction. Springer Science & Business Media.

Mancini, M., Silva, S.H.G., dos Santos Teixeira, A.F., Guilherme, L.R.G., Curi, N., 2020. Soil parent material prediction for Brazil via proximal soil sensing. Geoderma Reg. e00310. https://doi.org/https://doi.org/10.1016/j.geodrs.2020.e00310

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

Meireles, H.T., Marques Júnior, J., Campos, M.C.C., Pereira, G.T., 2012. Relações solo-paisagem em topossequência de origem basáltica. Pesqui. Agropecuária Trop. 42, 129–136.

Mello, F.A.O., Demattê, J.A.M., Rizzo, R., Dotto, A.C., Poppiel, R.R., Mendes, W.S., Guimarães, C.C.B., 2021. Expert-based maps and highly detailed surface drainage models to support digital soil mapping. Geoderma 384, 114779. https://doi.org/https://doi.org/10.1016/j.geoderma.2020.114779

Mendes, W. de S., Demattê, J.A.M., Silvero, N.E.Q., Rabelo Campos, L., 2021. Integration of multispectral and hyperspectral data to map magnetic susceptibility and soil attributes at depth: A novel framework. Geoderma 385, 114885. https://doi.org/https://doi.org/10.1016/j.geoderma.2020.114885

Mendes, W.D.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

Meschede, M., Warr, L.N., 2019. The Geology of Germany: A Process-Oriented Approach. Springer.

Miller, B.A., Lee Burras, C., 2015. Comparison of Surficial Geology Maps Based on Soil Survey and In Depth Geological Survey. Soil Horizons 56, sh14-05–0005. https://doi.org/https://doi.org/10.2136/sh14-05-0005

Minasny, B., Hartemink, A.E., 2011. Predicting soil properties in the tropics. Earth-Science Rev. 106, 52–62. https://doi.org/10.1016/j.earscirev.2011.01.005

Mokma, D.L., Sprecher, S.W., 1994. Water table depths and color patterns in soils developed from red parent materials in Michigan, USA. CATENA 22, 287–298. https://doi.org/https://doi.org/10.1016/0341-8162(94)90039-6

Mullins, C.E., 1977. Magnetic susceptibility of the soil and its significance in soil science–a review. J. soil Sci. 28, 223–246.

Nauman, T.W., Thompson, J.A., 2014. Semi-automated disaggregation of conventional soil maps using knowledge driven data mining and classification trees. Geoderma 213, 385–399. https://doi.org/https://doi.org/10.1016/j.geoderma.2013.08.024

Nawar, S., Mouazen, A.M., 2017. Predictive performance of mobile vis-near infrared spectroscopy for key soil properties at different geographical scales by using spiking and data mining techniques. CATENA 151, 118–129. https://doi.org/https://doi.org/10.1016/j.catena.2016.12.014

Neel, C., Soubrand-Colin, M., Piquet-Pissaloux, A., Bril, H., 2007. Mobility and bioavailability of Cr, Cu, Ni, Pb and Zn in a basaltic grassland: Comparison of selective extractions with quantitative approaches at different scales. Appl. Geochemistry 22, 724–735.

Nogueira, T.A.R., Abreu-Junior, C.H., Alleoni, L.R.F., He, Z., Soares, M.R., Santos Vieira, C. dos, Lessa, L.G.F., Capra, G.F., 2018. Background concentrations and quality reference values for some potentially toxic elements in soils of São Paulo State, Brazil. J. Environ. Manage. 221, 10–19. https://doi.org/10.1016/j.jenvman.2018.05.048

Nolasco de Carvalho, C.C., Nunes, F.C., Homem Antunes, M.A., Nolasco, M.C., 2015. Soil surveys in Brazil and perspectives in digital soil mapping. Soil Horizons 56. https://doi.org/10.2136/sh14-01-0002

Odgers, N.P., Sun, W., McBratney, A.B., Minasny, B., Clifford, D., 2014. Disaggregating and harmonising soil map units through resampled classification trees. Geoderma 214–215, 91–100. https://doi.org/10.1016/j.geoderma.2013.09.024

Pelegrino, M.H.P., Silva, S.H.G., Menezes, M.D. de, Silva, E. da, Owens, P.R., Curi, N., 2016. Mapping soils in two watersheds using legacy data and extrapolation for similar surrounding areas . Ciência e Agrotecnologia .

Poppiel, R.R., Lacerda, M.P.C., Demattê, J.A.M., Oliveira, M.P., Gallo, B.C., Safanelli, J.L., 2019a. Pedology and soil class mapping from proximal and remote sensed data. Geoderma 348, 189–206. https://doi.org/10.1016/j.geoderma.2019.04.028

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

Prokopovich, N.P., 1984. Use of Agricultural Soil Survey Maps for Engineering Geologic Mapping. Environ. Eng. Geosci. xxi, 437–447. https://doi.org/10.2113/gseegeosci.xxi.4.437

Richardson, J.L., Daniels, R.B., 1993. Stratigraphic and Hydraulic Influences on Soil Color Development. Soil Color, SSSA Special Publications. https://doi.org/https://doi.org/10.2136/sssaspecpub31.c7

Richter, J., Owens, P.R., Libohova, Z., Adhikari, K., Fuentes, B., 2019. Mapping parent material as part of a nested approach to soil mapping in the Arkansas River Valley. Catena 178, 100–108. https://doi.org/10.1016/j.catena.2019.02.031

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/https://doi.org/10.1016/j.geoderma.2019.114018

Rodgers, J., 1953. Geologic map of east Tennessee with explanatory text: Tennessee Department of Conservation. Div. Geol. Bull. 58, 168.

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 Sens. Environ. 205, 1–17. https://doi.org/10.1016/j.rse.2017.11.004

Ruhe, R. V, 1960. Elements of the soil landscape. Trans. 7th int. Congr. Soil Sci. 4, 165–170.

Safanelli, L.J., Chabrillat, S., Ben-Dor, E., Demattê, A.M.J., 2020. Multispectral Models from Bare Soil Composites for Mapping Topsoil Properties over Europe. Remote Sens. . https://doi.org/10.3390/rs12091369

Sankey, J.B., Brown, D.J., Bernard, M.L., Lawrence, R.L., 2008. Comparing local vs. global visible and near-infrared (VisNIR) diffuse reflectance spectroscopy (DRS) calibrations for the prediction of soil clay, organic C and inorganic C. Geoderma 148, 149–158. https://doi.org/https://doi.org/10.1016/j.geoderma.2008.09.019

Santos, H.G., Hochmüller, D.P., Cavalcanti, A.C., Rêgo, R.S., Ker, J.C., Panoso, L.A., Amaral, J.A.M. do., 1995. Procedimentos normativos de levantamentos pedológicos. Brasília, DF: EMBRAPA-SPI; Rio de Janeiro: EMBRAPA-CNPS, 1995.

Santra, P., Kumar, M., Panwar, N., 2017. Digital soil mapping of sand content in arid western India through geostatistical approaches. Geoderma Reg. 9, 56–72. https://doi.org/https://doi.org/10.1016/j.geodrs.2017.03.003

Saxton, K.E., Rawls, W., Romberger, J.S., Papendick, R.I., 1986. Estimating generalized soil-water characteristics from texture 1. Soil Sci. Soc. Am. J. 50, 1031–1036.

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/DOI: 10.1180/claymin.2008.043.1.11

Schaetzl, J Randall and Anderson, S., 2005. Soil Genesis and Geomorphology. Cambridge University Press, New York.

Schaetzl, R., Anderson, S., 2005. Soils. Genesis and Geomorphology. Cambridge University Press, New York.

Schwertmann, U., 1993. Relations Between Iron Oxides, Soil Color, and Soil Formation, in: Soil Color, SSSA Special Publication SV - 31. Soil Science Society of America, Madison, WI, pp. 51–69. https://doi.org/10.2136/sssaspecpub31.c4

Schwertmann, U., 1988. Occurrence and formation of iron oxides in various pedoenvironments, in: Iron in Soils and Clay Minerals. Springer, pp. 267–308.

Schwertmann, U., Taylor, R.M., 1989. Iron Oxides. Miner. Soil Environ., SSSA Book Series. https://doi.org/https://doi.org/10.2136/sssabookser1.2ed.c8

Searle, R., McBratney, A., Grundy, M., Kidd, D., Malone, B., Arrouays, D., Stockman, U., Zund, P., Wilson, P., Wilford, J., Van Gool, D., Triantafilis, J., Thomas, M., Stower, L., Slater, B., Robinson, N., Ringrose-Voase, A., Padarian, J., Payne, J., Orton, T., Odgers, N., O'Brien, L., Minasny, B., Bennett, J.M., Liddicoat, C., Jones, E., Holmes, K., Harms, B., Gray, J., Bui, E., Andrews, K., 2021. Digital soil mapping and assessment for Australia and beyond: A propitious future. Geoderma Reg. 24, e00359. https://doi.org/https://doi.org/10.1016/j.geodrs.2021.e00359

Shabou, M., Mougenot, B., Chabaane, Z., Walter, C., Boulet, G., Aissa, N., Zribi, M., 2015. Soil clay content mapping using a time series of Landsat TM data in semi-arid lands. Remote Sens. 7, 6059–6078. https://doi.org/10.3390/rs70506059

Silva, L.S., Júnior, J.M., 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 185, 104258.

Silvero, N.E.Q., Demattê, J.A.M., Vieira, J. de S., Mello, F.A. de O., Amorim, M.T.A., Poppiel, R.R., Mendes, W. de 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/https://doi.org/10.1016/j.geoderma.2021.115089

Slater, J.A., Garvey, G., Johnston, C., Haase, J., Heady, B., Kroenung, G., Little, J., 2006. The SRTM data "finishing" process and products. Photogramm. Eng. Remote Sensing 72, 237–247. https://doi.org/10.14358/PERS.72.3.237

Staff, S.S., 2010. Keys to soil taxonomy. United States Dep. Agric. Soil Conserv. Serv. Washington, DC.

Stradioto, M.R., Chang, H.K., 2020. Sandstone Diagenesis of the Bauru Group in the Sao Paulo State. Ciência e Natura; Vol 42 40 YEARS - Anniv. Ed. https://doi.org/10.5902/2179460X42694

Styc, Q., Gontard, F., Lagacherie, P., 2021. Harvesting spatially dense legacy soil datasets for digital soil mapping of available water capacity in Southern France. Geoderma Reg. 24, e00353. https://doi.org/https://doi.org/10.1016/j.geodrs.2020.e00353

Suguio, K., Barcelos, J.H., Guedes, M.G., Verdiani, A.C., 1984. Canal de Pereira Barreto: local de transição entre os arenitos Caiuá, Santo Anastácio e Adamantina. Rev. do Inst. Geológico 5, 25–37. https://doi.org/10.5935/0100-929x.19840004

Teng, H., Viscarra Rossel, R.A., Shi, Z., Behrens, T., 2018. Updating a national soil classification with spectroscopic predictions and digital soil mapping. CATENA 164, 125–134. https://doi.org/https://doi.org/10.1016/j.catena.2018.01.015

Thorp, J., 1949. Interrelations of Pleistocene Geology and Soil Science. GSA Bull. 60, 1517–1526. https://doi.org/10.1130/0016-7606(1949)60[1517:IOPGAS]2.0.CO;2

U.S.G.S., 2019a. Landsat 8 Surface Reflectance Code (LASRC) Poduct Guide. (No. LSDS-1368 Version 2.0). 40.

U.S.G.S., 2019b. Landsat 4-7 Surface Reflectance (Ledaps) Product Guide 32. https://doi.org/10.1016/0042-207X(74)93024-3

van Breemen, N., Buurman, P., 2002. Soil Formation - Second Edition.

Varin, M., Bournival, P., Fink, J., Chalghaf, B., 2021. Mapping Vernal Pools Using LiDAR Data and Multitemporal Satellite Imagery. Wetlands 41, 34. https://doi.org/10.1007/s13157-021-01422-9

Vincent, S., Lemercier, B., Berthier, L., Walter, C., 2018. Spatial disaggregation of complex Soil Map Units at the regional scale based on soil-landscape relationships. Geoderma 311, 130–142. https://doi.org/10.1016/j.geoderma.2016.06.006

Vivekananda, G.N., Swathi, R., Sujith, A., 2021. Multi-temporal image analysis for LULC classification and change detection. Eur. J. Remote Sens. 54, 189–199. https://doi.org/10.1080/22797254.2020.1771215

Vrieling, A., 2006. Satellite remote sensing for water erosion assessment: A review. CATENA 65, 2–18. https://doi.org/https://doi.org/10.1016/j.catena.2005.10.005

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

Waroszewski, J., Sprafke, T., Kabała, C., Kobierski, M., Kierczak, J., Musztyfaga, E., Loba, A., Mazurek, R., Łabaz, B., 2019. Tracking textural, mineralogical and geochemical signatures in soils developed from basalt-derived materials covered with loess sediments (SW Poland). Geoderma 337, 983–997.

Wetterlind, J., Stenberg, B., 2010. Near-infrared spectroscopy for within-field soil characterization: small local calibrations compared with national libraries spiked with local samples. Eur. J. Soil Sci. 61, 823–843. https://doi.org/https://doi.org/10.1111/j.1365-2389.2010.01283.x

Wilson, M.J., 2019. The importance of parent material in soil classification: A review in a historical context. CATENA 182, 104131. https://doi.org/https://doi.org/10.1016/j.catena.2019.104131

4. USING REMOTE SENSING TOOLS TO IDENTIFY HYDROMORPHIC SOILS HIDDEN UNDER AGRICULTURE AND IGNORED BY ENVIRONMENTAL REGULATION

Abstract

The need to secure food and water for the world's population increased over the past decades, having soil as a major natural resource to achieve this goal. However, the advance of soil degradation requires advanced monitoring techniques to access soil data. We evaluated multitemporal bare soil image reflectance at various locations to verify the occurrence of hydromorphism, predicted hydromorphic soils for a large area, and analyzed their distribution. The hydromorphic soils did not present the typical concavity features of Fe oxides in the region of the 900 nm band and the typical hematite amplitude located between 520 and 580 nm. Slope and SySI band 1 contributed the most for the prediction model with 98 and 69%, respectively. The optimal model was random forest with cross validation, reaching accuracy of 0.92 and Kappa of 0.77. Hydromorphic soils were generally predicted at concave and flat landforms, at the summit or footslope positions. Mato Grosso do Sul had 28% of the area analyzed in this study classified as hydromorphic, while São Paulo had 6%. Soybean and pasture areas had up to 14.9% of hydromorphic soils. Finally, SySI was able to map hydromorphic soils, a powerful tool to improve monitoring at agricultural areas.

Keywords: Remote sensing; Digital soil mapping; SySI; Hydromorphic soils; Machine learning

4.1. Introduction

Over the past decades multiple environmental challenges were addressed in international conventions and worldwide initiatives for sustainable development (CEC, 2006; Wu and Clark, 2016). The need to provide food and water for the world's population while minimizing the impact on climate raised awareness on the effort to achieve Food Security, Water Security, Energy Security, Climate Change Abatement, Biodiversity Protection and Ecosystem Service Delivery (Bouma and McBratney, 2013; Godfray et al., 2010; McBratney et al., 2014). In this regard, McBratney et al. (2014) addressed that soil has an important role on the achievement of such goals, but so far has been poorly applied in models to investigate these global challenges. The lack of soil knowledge and the advance of soil degradation caused by agriculture pose a threat as population is estimated to be 9 billion by the middle of the 21st century (Baker, 1994; Lal and Stewart, 2010).

Soil is a natural body comprised of solids (minerals and organic matter), liquid, and gases that is formed through the interaction of five natural factors (climate, organisms, relief, parent material, and time) (Dokuchaev, 1883; Jenny, 1994). Soil formation varies according to the activity of each forming factor, and each soil will have a contribution for the environment and ecosystem regulation (Banwart et al., 2011; Chorover et al., 2007). In this regard, hydromorphic soils stand out as a regulator of hydrological and biogeochemical cycles with fauna and flora (Mitsch et al., 2015). These soils are connected with the water table and represent a supplier for water recharge, nutrients, and sediments for riverine ecosystems (Lehrback et al., 2016; Santana and Barroso, 2014).

Hydromorphic soils are a result of prolonged water saturation and seasonal alternation between water logging and drainage, which promote anoxic conditions for at least part of the year (Buol and Rebertus, 1988; Duchaufour, 1982; van Breemen and Buurman, 1998). The permanent or periodic saturation of the soil is caused by the groundwater or from above (rain or irrigation) (Buol et al., 2011; van Breemen and Buurman, 1998). When the soil pores are filled with water, gas diffusion becomes very slow, decreasing the supply of O_2 (van Breemen and Buurman, 1998). This lack of oxygen favors the activity of anaerobic micro-organisms on the reduction of oxidized soil components, such as Fe^{3+}, Mn^{3+}, and Mn^{4+}, into Fe^{2+} and Mn^{2+} (Bedard-Haughn, 2011; Buol and Rebertus, 1988; Özcan et al., 2018). As a result, the zones where these oxides were reduced become a non-mottled and greyish

soil (chroma ≤ 2) due to the permanent water saturation, and are classified as gley soils (Buol et al., 2011; Pavlović et al., 2017; Schaetzl and Anderson, 2005). The water saturation occurs due to their landscape position at poorly drained floodplains and flat surfaces with constant influence of the water table, causing the gley phenomena (dos Santos et al., 2018; Ker et al., 2015).

Although hydromorphic soils are fragile and important for the environment, the agricultural expansion scenario represent an environmental pressure for these ecosystems (Gebresllassie et al., 2014; Tilman, 1999; Zou et al., 2018). Multiple studies presented ground water contamination through agricultural activities involving fertilizers, pesticides, and other chemical agents, implying the fragility of such ecosystems and the impact of unregulated anthropic activities (Bera et al., 2021; Goss et al., 1998; Miller, 1972; Nath et al., 2021; Ritter, 1990; Yang et al., 2021). According to the Brazilian Forestry Code, the margin areas of a river, lake, and water source must be preserved with the natural vegetation (Piedade et al., 2012; Sparovek et al., 2011). However, the areas to be preserved start counting from the riverbed at the drought season, which end up excluding potential wetland soils that will only be affected during the flood season. These soils will naturally be included in agricultural sites at risk of getting contaminated, since they are difficult to map and ignored by environmental regulations.

The definition of these seasonally saturated soils differs in soil classification systems, considering different depths for redoximorphic activity, duration of the saturated condition, color, and others (Table 1). Thus, the identification and mapping of such soils is fundamental for the conservation of water resources. However, there is a challenge on identifying and mapping theses soils over large areas. They often represent a small percentage among other soil types, which hampers their mapping (Mello et al., 2021b).

Table 1: Different terminologies and descriptions for soils with hydromorphic activity.

Terminology	Description	Source
Gleissolos	Mineral soils with Gley horizon within the first 50 cm from the surface, or between 50 and 150 cm deep. The Gley horizon must be under a shallow surface horizon insufficient to be classified as Organosols.	(dos Santos et al., 2018)
Gleysols	A wetland soil (hydric soil) that unless drained is saturated with groundwater for long enough to develop a characteristic gleyic color pattern. layer ≥ 25 cm thick, and starting ≤ 40 cm from the mineral soil surface, that has gleyic properties throughout and reducing conditions in some parts of every sublayer.	(IUSS Working Group, 2015)
Entisols, Inceptsols, Mollisols (Aqu-suborders)	Aquic conditions and sulfidic materials within 50 cm of the mineral soil surface; Permanent saturation with water and a reduced matrix in all horizons below 25 cm from the mineral soil surface; In a layer above a densic, lithic, or paralithic contact or in a layer at a depth between 40 and 50 cm below the mineral soil surface, whichever is shallower, aquic conditions for some time in normal years.	(Staff, 2010)
Wetlands	Areas where water covers the soil, or is present either at or near the surface of the soil all year or for varying periods of time during the year, including during the growing season.	(Kadlec and Wallace, 2008)

Therefore, we established the following goals to proceed with this study: (i) evaluate and analyze multitemporal bare soil image reflectance at various locations to verify the occurrence of hydromorphism; ii)

combine satellite and relief data with a machine learning algorithm to classify hydromorphic soils for a large area; iii) analyze the hydromorphic locations regarding the federal states and land use/cover types.

4.2. Material and methods

4.2.1. Study area

This study was conducted in an 863,577.9 km² area located across the southeast and mid-west regions of Brazil (Fig. 1). The region comprises tropical and subtropical climates classified as Aw (Savanna), Cwb (Subtropical highlands), Cfa, and Cwa (Humid subtropical) according to the Köppen climate classification (Alvares et al., 2013). The rainfall varies between 1000 to 2200 mm year^{-1} and the mean annual temperature varies between 18 to 24 °C (Alvares et al., 2013).

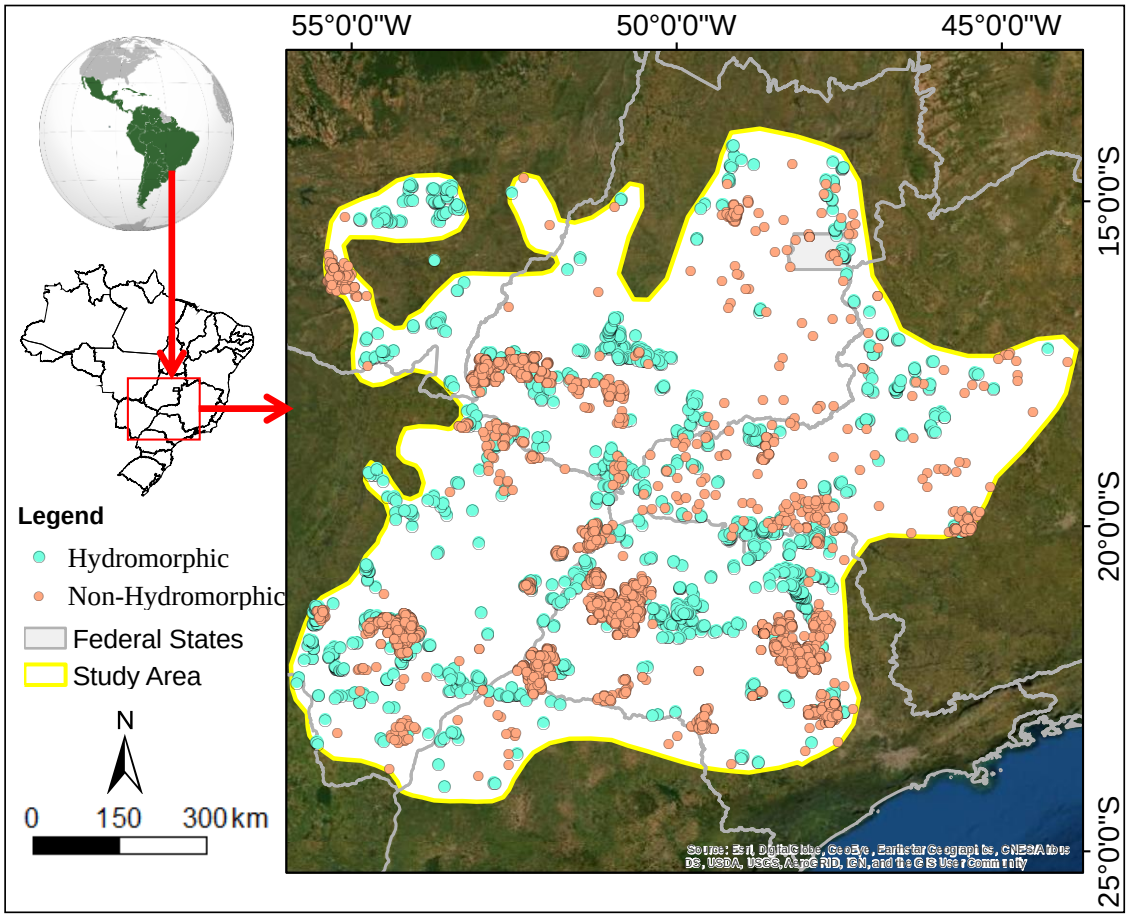

Figure 1. Study area located between the southeast and mid-west regions of Brazil. The map shows the point locations where soils were identified as hydromorphic and non-hydromorphic, depending on their conditions analyzed in the field and through remote sensing.

4.2.2. Soil dataset

In order to gather a large amount of data regarding hydromorphic soils, we combined three soil datasets from open and closed soil inventories in Brazil. The first soil dataset is private and was provided by the Geotechnologies in Soil Science Group (GeoSS), which has developed a series of soil studies in many regions of Brazil, especially in the São Paulo state (Demattê et al., 2019). The second soil dataset is an open source database

organized by Cooper et al. (2005), which obtained soil information from the Radambrasil project (Radambrasil, 1973) and other regional surveys. The last soil dataset used was provided by Samuel-Rosa et al. (2020), which assessed legacy soil observations from multiple private studies in Brazil, and made it available for download online. These two open source datasets were filtered and only the samples at 0 – 20 cm depth that contained a soil class and spatial coordinates were selected.

Table 2: Soil dataset orders according to the World Reference Base – WRB (IUSS Working Group, 2015) and the corresponding classes in the Soil Taxonomy (Soil Science Division Staff, 2017) and Brazilian Soil Classification System (dos Santos et al., 2018).

Code	WRB	Soil Taxonomy	SiBCS	Observations
Ac	Acrisol	Ultisols	Argissolo	4313
Cx	Cambisol	Inceptisols	Cambissolo	884
Ch	Chernozems	Molisols	Chernossolo	190
Pz	Podzols	Spodosols	Espodossolo	11
Gx	Gleysols	Entisols	Gleissolo	528
Fe	Ferralsol	Oxisols	Latossolo	9157
Lv	Luvisols	Aridisols	Luvissolo	6
Lp	Leptsols, Arenosols, Regosols	Entisols	Neossolo	2151
Ni	Nitisols	Ultisols	Nitossolo	232
Ps	Planosols	Alfisols	Planossolo	351
Total				17823

Most of these soil samples were acquired from traditional soil surveys, which consist of a soil specialist using conventional methods of soil surveying and mapping to select the sampling locations based on pedogeomorphological relationships (Santos et al., 1995). After combining the three datasets, we excluded the Gleysols and Planosols observations in order to gather only not hydromorphic soils, since their morphology is not affected by water saturation through the course of a year (Ker et al., 2015).

The hydromorphic soil observations were acquired from field works and remote sensing analysis. First, we identified soils in the field with water accumulation below 40 cm depth. These locations were often inside croplands and not classified as Gleysols or Planosols in legacy pedological maps (dos Santos et al., 2011). Afterwards, we combined the field observations with satellite image analysis to determine all these points, which were classified as hydromorphic. The final soil dataset had the spatial coordinates and the class determined as hydromorphic and not hydromorphic.

4.2.3. Environmental information

4.2.3.1. Synthetic soil image (SySI)

We implemented the Geospatial Soil Sensing System (GEOS3) (Demattê et al., 2018) to a time-series of Landsat images using the Google Earth Engine (GEE) platform (Gorelick et al., 2017) in order to generate a 30 meter spatial resolution Synthetic Soil Image (SySI). We used the Landsat 4 Thematic Mapper (TM) (1982–1993), Landsat 5 TM (1984–2012), Landsat 7 Enhanced Thematic Mapper Plus (ETM +) (1999–2018) and the Landsat 8 Operational Land Manager (OLI) (2013–2018). These methods used only Tier 1 or surface reflectance processed by the LEDAPS (Landsat 4, 5, and 7) (U.S.G.S., 2019a) and LASRC (Landsat 8) algorithms (U.S.G.S., 2019b).

Table 3: Environmental variables used as predictors for the digital soil mapping (DSM).

Class	Attribute	Description	Unit	Reference
Terrain	DEM	Elevation	meter	TAGEE (Safanelli et al., 2020)
	SLO	Slope	degree	
	NRT	Northernness	ND	
	EST	Easternness	ND	
	HCV	Horizontal Curvature	meter	
	VCV	Vertical Curvature	meter	
	SID	Shape Index	ND	
Remote Sensing	B1	Landsat Band 1 – Blue	Ref. factor	GEOS3 (Demattê et al., 2018)
	B2	Landsat Band 2 – Green	Ref. factor	
	B3	Landsat Band 3 – Red	Ref. factor	
	B4	Landsat Band 4 – Near Infrared	Ref. factor	
	B5	Landsat Band 5 – SWIR1	Ref. factor	
	B7	Landsat Band 7 – SWIR2	Ref. factor	

ND: dimensionless.

The GEOS3 algorithm extracts soil information from the collection of Landsat images and aggregates the spatially bare soil fragments into a synthetic soil image, which is the reflectance image of the bare soil composite (Fongaro et al., 2018). The bare soil pixels are identified on single satellite images through a set of identification rules. The rules were based on spectral indices coupled with quality assessment bands, which removed cloud, cloud shadow, inland water, photosynthetic vegetation, and non-photosynthetic vegetation (crop residues) (Safanelli et al., 2020). The pixels were classified as soil based on the Normalized Difference Vegetation Index (NDVI), with a threshold between – 0.15 and 0.20 to mask out green vegetation, Normalized Burning Ratio (NBR2), with a – 0.15 and 0.15 to mask out crop residues, difference between bands 1 and 2 (B2 – B1) and bands 2 and 3 (B3 – B2) (Demattê et al., 2020, 2018).

Afterwards, the bare soil pixels were applied to calculate, pixel-by-pixel, the median values of topsoil reflectance for single bands and obtain the final reflectance value (Demattê et al., 2018). The SySI had six spectral bands from blue to short-wave infrared regions at 30 m resolution (Table 3).

4.2.3.2. Terrain covariates

The terrain data was selected to compose the set of environmental variables due to its relevance as a soil forming factor (Dokuchaev, 1883; Jenny, 1994). Moreover, hydromorphic soils tend to occupy specific landscape positions (Bedard-Haughn and Pennock, 2002), indicating that terrain covariates are likely to contribute on the mapping of these soils.

We used the Terrain Analysis in Google Earth Engine (TAGEE) package in GEE to calculate a set of terrain spatial data (Safanelli et al., 2020b). The method uses the Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM) with a spatial resolution of 30 x 30 m to calculate multiple topographic variables (Table 3). With a 3x3 moving window rolling across the DEM, the first and second partial derivatives of terrain were calculated to estimate flux and form attributes (Safanelli et al., 2020b).

4.2.4. Statistical and spectral analysis

We combined the soil dataset with the environmental variables to perform a statistical analysis. The soil locations were separated by class (hydromorphic and not hydromorphic) and explored through a boxplot graphic to see the distribution of these soils according to the environmental information. The graphic was performed using the "ggplot" package (Wickham, 2011) in R ((R Core Team, 2013).

When SySI was displayed in true color composition (red, green, blue), it was possible to visually identify a change in hue and intensity of colors (Fig. 4a,b). This pattern is normally observed next to drainage channels and close to the water sources (Fig. 6b). Based on the geographical position and the differences in SySI's colors at these locations, soil moisture could be the cause for the abrupt change in the spectral response (Lobell and Asner, 2002; Weidong et al., 2002). However, we selected soil field samples with laboratory spectral analysis to quantitatively analyze the observed features.

With the objective of supporting and evaluating the applicability of the use of SySI in the identification of hydromorphic soils, an analysis of spectral signatures collected in the laboratory of hydromorphic soils was carried out, using the methodology proposed by Demattê et al. (2014). We used spectral information of soils with hydromorphic features, obtained by Demattê et al. (2017), Marques et al. (2019), and Demattê et al. (2019). Furthermore, we convolved the same dates for the SySI bands (Ben-Dor and Banin, 1995; Demattê et al., 2018), in order to evaluate the SySI spectral signature of these same soils.

Finally, for a better understanding of the spectral differences of hydromorphic and not hydromorphic class soils, we used the database of Demattê et al. (2019) and the areas classified as hydromorphic and not hydromorphic to select nearby points for application of the second derivative of the Kubelka-Munk function, as a way to verify changes in the amounts of iron oxides (Scheinost et al., 1998). The application of the second derivative aims to highlight smooth spectral changes not perceptible in the raw spectral signature (Poppiel et al., 2019). Soil samples in the same toposequence, classified as hydromorphic and not hydromorphic, were compared by the second derivative of the Kubelka Munk function of spectral laboratory data. The procedures for acquiring the hyperspectral data of the samples used, such as the type of sensor, its spectral resolution and geometric configurations used can be obtained in Demattê et al. (2019).

4.2.5. Random forest classification

The random forest (RF) algorithm was selected to perform the hydromorphic soils mapping, since its relevance in DSM (Khaledian and Miller, 2020; Taghizadeh-Mehrjardi et al., 2020; Teng et al., 2018; Zeraatpisheh et al., 2019). RF estimates a user-specified number of decision trees by randomly sampling an existing dataset (Breiman, 2001). However, at each node construction, a random sample of the dependent variables is used. The resulting decision tree is used to estimate the out-of-bag error rate by predicting the value of the remaining unsampled data and comparing with the known results (Gambill et al., 2016).

4.2.5.1 Model tuning and calibration

We used the soil observations to extract the environmental variables' values at each point using the "Extract multi values to point" function in ESRI ArcGIS 10.4. The dataset was used to calibrate a random forest model and to test different sets of hyperparameters using the caret package in R software (Kuhn, 2008). We performed a grid search to select the optimal hyperparameters, which were the maximum depth (150), maximum features (3), minimum samples leaf (1), minimum samples split (10), and number of trees (300). These parameters regulate the number of variables that can be randomly sampled in each split of the trees, the tree depth by setting the minimal number of samples for the terminal nodes, and the number of trees (Probst et al., 2019).

In order to calibrate the RF model, we tested three resampling methods coupled with the RF model using the caret package in R software (Kuhn, 2008). The first test used k-fold cross-validation (CV) method to fit the prediction models. CV is a resampling method used to fix optimistic results of the predictive effectiveness of regression equations (Mosier, 1951). The method randomly divided the data in k groups, using k - 1 groups to fit a model, and one for validation (Browne, 2000; Tibshirani and Efron, 1993). The procedure is repeated k times, always leaving one group out of the calibration dataset (Hawkins et al., 2003). Afterwards, the results are summarized with the mean of the model scores. We used the 10-fold CV to fit the RF model, compute the prediction performance, and apply it to the environmental variables to predict the classes for the study area.

The second method was the bootstrapping, which is a data resampling technique for estimating the statistical parameters of an unknown distribution and a robust method for optimal model selection (Efron, 1992; Shao, 1996). Bootstrapping technique randomly chooses n samples with replacement, so the same sample can be chosen multiple times. This process is repeated t times and the predictive performance of the validation set of those repeats are recorded and averaged as the final estimation of generalization performance of the model. We randomly partitioned the data as a training set (70%) and the validation set (30%). We also tested the out-of-bag resampling method which is a method of measuring the prediction error of random forests (Bischl et al., 2012).

4.2.5.2 Model performance and variables importance

The prediction performance of the data was accessed using the three default parameters of caret for classification models, being the number of randomly selected predictors (mtry), accuracy, and kappa coefficient. The mtry regulates the number of variables that can be randomly sampled in each split of the trees, which resulted in 2, 20, and 39. We used 300 trees for stable variable estimates.

As many environmental information were used as covariables to fit the RF model, we analyzed the variables' importance through the mean decrease Gini index. The analysis helped indicating how was the contribution of the terrain, climate, and remote sensing variables.

4.2.6. Spatial prediction and model validation

After testing the resampling methods and fitting the RF model, we selected the optimal model and used its parameters to predict the hydromorphic and not hydromorphic classes across the study area with the raster package in R software (Hijmans et al., 2013). The resulting binary map was exported and opened in ArcGIS 10.4. for further analysis.

We used the full dataset to perform the CV and randomly partitioned the dataset into calibration (70%) and validation (30%) datasets to perform the bootstrapping and out-of-bag resampling methods. We also calculated a confusion matrix for the optimal model and analyzed the errors of inclusion (commission errors), errors of exclusion (omission errors), user accuracy (UA), producer accuracy (PA), and global accuracy (GA) (Congalton and Green, 2019; Poppiel et al., 2019).

We also evaluated the areas mapped as hydromorphic according to their relief position (infiltration or surface runoff environment), proximity to the channel network or water source, and according to the pattern registered by SySI. We selected an area to analyze the distribution of hydromorphic soils across a toposequence, based on the soil-landscape relationship rules (Milne, 1935; Ruhe, 1960) and the channel network patterns (Shreve, 1967; Strahler, 1952).

4.2.7. Spatial prediction and model validation

With the digital map of hydromorphic soils, we were able to analyze the current land use situation of these soils. First, we masked the pixels classified as hydromorphic and exported them to a new raster in ESRI ArcGIS 10.4. The new raster was plotted over a land use and landcover map from the MapBiomas project. MapBiomas is a governmental initiative aimed to reconstruct annual land use and land cover information between 1985 and 2017 for Brazil based on random forest applied to Landsat archive using Google Earth Engine (Souza et al., 2020). The dataset is available for download at their repository in GEE and at their website (https://mapbiomas.org/) (Souza et al., 2020).

Finally, we quantified the areas of hydromorphic soils and identified the land uses at the areas. We also computed the total areas of hydromorphic soils for each federal state included in the study area (Fig. 1). The result was presented in a table with the total area of hydromorphic soils for each land use class and federal state.

4.3. Results

4.3.1. Statistical and spectral analysis

As the soil dataset has more observations of not hydromorphic soils, their values are noticeable across the whole TAGEE variables range (Fig. 2). The hydromorphic soils showed distinct patterns for the slope, horizontal curvature, shape index, and for the SySI bands (Fig. 2). The slope boxplot indicates that hydromorphic soils are commonly located at flat and more stable surfaces with more gentle slopes, while not hydromorphic soils are distributed across the landscape regardless of the slope (Fig. 2). The horizontal curvature and shape index presented negative values for hydromorphic soils, which indicate that these soils are often located at concave landforms favoring water accumulation (Florinsky, 2012) (Fig. 2). The vertical curvature also presented a tendency to negative values, a case of landforms with relative deceleration of gravity-driven substance flows, or a stable and flat landform (Florinsky, 2012) (Fig. 2).

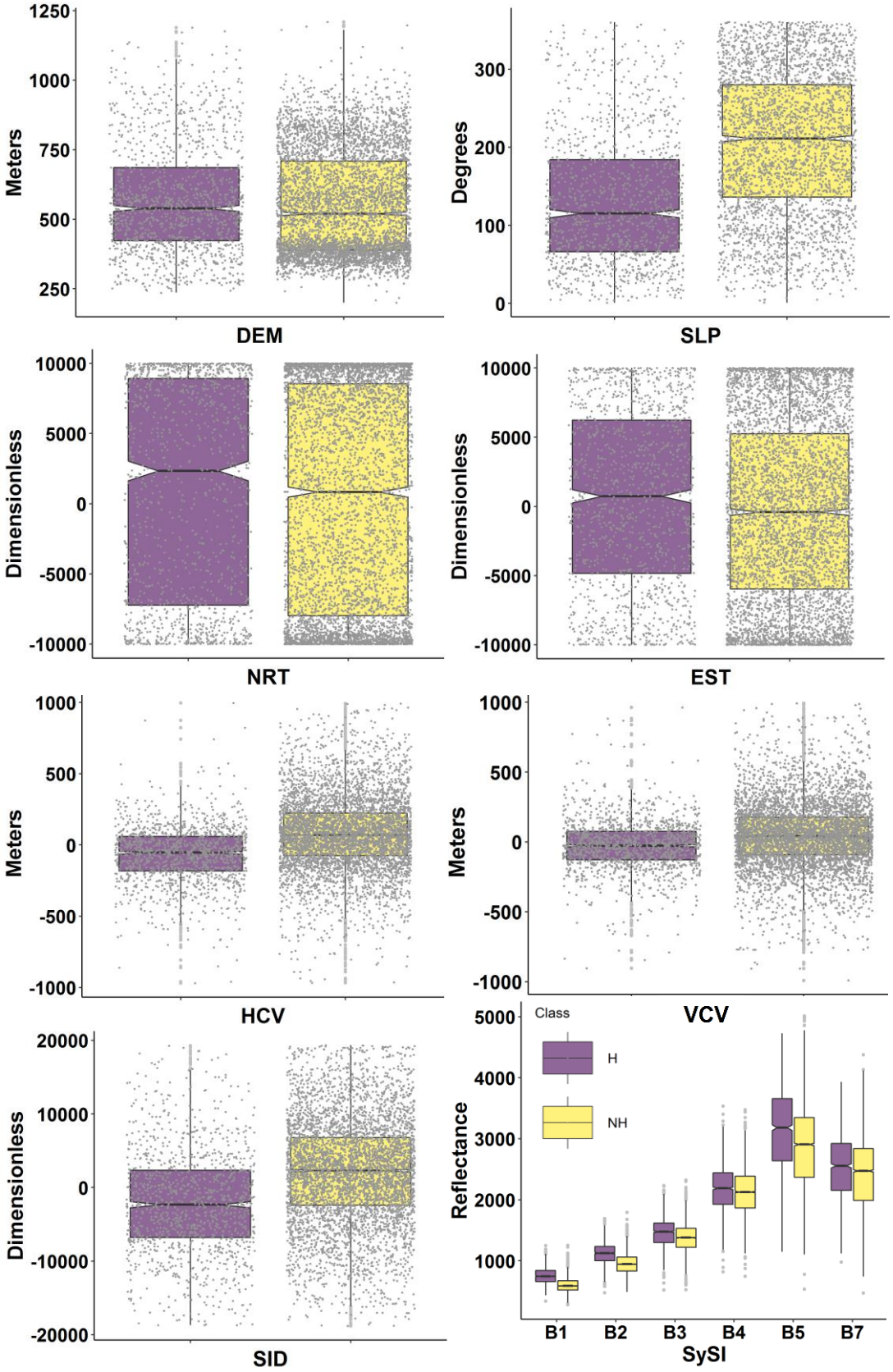

Figure 2: TAGEE and SySI values distribution according to the soil classification as H: hydromorphic and NH: not hydromorphic. DEM: digital elevation model; SLP: slope; NRT: northernness; EST: easternness; HCV: horizontal curvature; VCV: vertical curvature; SID: shape index; B1 to B7 refers to the Landsat bands used in SySI.

The elevation, northernness, and easternness have similar patterns for hydromorphic and not hydromorphic soils (Fig. 2). The boxplot indicates that the geographical position and altitude are not as relevant as the terrain landforms in order to promote the conditions for a hydromorphic soil formation.

The spectra were higher at the hydromorphic soils for all SySI bands, contrary to what is expected from soils with high levels of humidity for most part of the year (Staff, 2010) (Fig. 2). However, satellite spectra retrieves only a few centimeters below the surface, not necessarily being affected by the soil moisture (Chabrillat et al., 2019). Nonetheless, hydromorphism lowers the formation of Fe oxides and oxyhydroxides due to redoximorphic environment, promoting higher levels of sand and 2:1 clay minerals which increases the reflectance factor (Lobell and Asner, 2002).

4.3.1.1 Spectral Characterization of hydromorphic soils

After observing the features in SySI that suggested the occurrence of hydromorphism, we selected ten samples of hydromorphic soils (Gleysols and Planosols) previously classified in the studies of Demattê et al. (2017), Demattê et al. (2019), and Marques et al. (2019) for spectral analysis (Fig. 3). The main characteristic observed was the absence of a concave feature in the region of 900 nm, typical of the presence of iron oxyhydroxides (Rossel and Behrens, 2010; Sherman and Waite, 1985). There is also an attenuation of features caused by the presence of organic matter in the soil between 350 and 1300 nm, with some of the curves showing a concave-rectilinear pattern between these bands (Fig. 3a). In addition, some samples showed a small convex feature between 350 and 450 nm. Therefore, the main spectral signature characteristics of hydromorphic soils were observed in the range from 350 to 1350 nm (Fig. 3a).

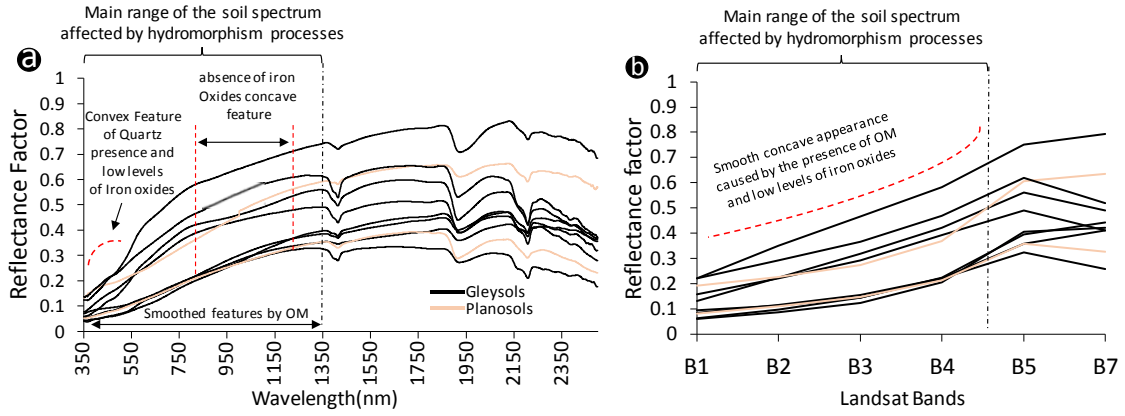

Figure 3: Laboratory topsoil spectra from A horizon of hydromorphic soils provided by (Demattê et al., 2017; Marques et al., 2019). a. Vis-NIR-SWIR spectra with low presence of iron oxides; b. the spectral curves were convolved into the Landsat/TM bands to analyze the satellite spectral behavior.

Regarding the convolved spectral signatures, a concave-rectilinear pattern was observed between SySI bands B1 to B4 (Fig. 3b), indicating that part of the characteristics presented in the spectral signatures of laboratory data are reflected in the SySI signatures, as, for instance, the attenuation caused by organic matter and absence of iron oxides.

4.3.1.2 Spectral Characterization of hydromorphic soils

Two sites were chosen to illustrate the spectral differentiation between hydromorphic and not hydromorphic soils (Fig. 4). The first site is a toposequence with clayey soils (Fig. 4a) not hydromorphic (highest position on the ground) and hydromorphic (lowest position on the ground). And a second site (Fig. 4b), with sandy soils following the same terrain arrangement described for clayey soils.

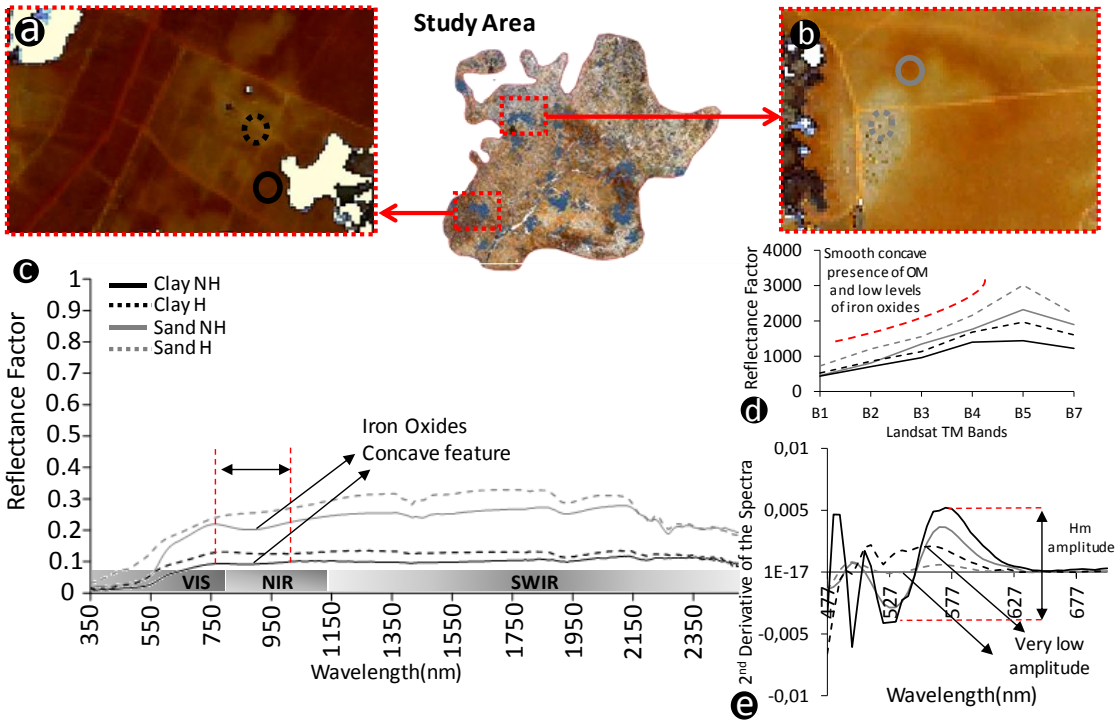

Figure 4: Laboratory and satellite topsoil spectra from locations with the hydromorphic feature found in SySI. a. soil field samples with laboratory spectra collected in a clayey area, one sample within the hydromorphic locations in SySI and one sample outside; b. soil field samples with laboratory spectra collected in a sandy area, one sample within the hydromorphic locations in SySI and one sample outside; c. laboratory spectra showing low iron oxides; d. SySI spectra collected at the same field locations; e. 2nd derivative of the spectra to highlight the lower features of iron oxides at the hydromorphic locations.

In both situations, the laboratory data (Fig 4c) show that: hydromorphic soils have greater reflectance than not hydromorphic ones, hydromorphic soils do not present the typical concavity feature of the presence of iron oxides in the region of the 900 nm band, as expected. The absence of iron oxides can also be confirmed by the absence of the typical hematite amplitude located between 520 and 580 nm in the second derivative of the Kubelka-Munk function (Fig. 4d). It is possible to verify that the hydromorphic soil samples present only a slight peak around 560 nm, but not an absorption at 525 nm. For sandy soils, a higher reflectance between 350 and 450 nm stands out.

For SySI spectral signatures at the same locations, it is possible to verify a higher reflectance for hydromorphic soils (Fig. 4d), and a slightly more concave shape between B1 and B4. Therefore, it is concluded that there is a basis for spectral behavior that allows the use of SySI as a predictor variable to differentiate areas with hydromorphic and not hydromorphic soils.

4.3.2. Model assessments

We tested three resampling methods to fit the RF model in order to predict the hydromorphic soil classes across the study area (Table 4). Overall, the combination of terrain (TAGEE) and SySI covariates had the highest accuracy for all resampling methods (0.92 and 0.91) (Table 4). Out-of-bag and bootstrapping methods had very similar results for calibration and validation, always varying 0.01 for each coefficient.

Table 4: Performance metrics of the prediction models for hydromorphic soils. Tests of different resampling methods available in R package caret (Kuhn, 2008).

Soil	Model[1]	Metrics[2]	Covariates[3]		
			Terrain	SySI	All
Hydromorphic Class	RF cv	mtry	4	2	7
		Acc	0.85	0.88	0.92
		Kappa	0.47	0.63	0.77
	RF oob		*Calibration*		
		mtry	2	2	13
		Acc	0.85	0.88	0.92
		Kappa	0.47	0.60	0.75
			Validation		
		Acc	0.85	0.88	0.91
		Kappa	0.49	0.61	0.72
		Sens	0.70	0.78	0.85
		Spec	0.87	0.89	0.92
	RF boot		*Calibration*		
		mtry	4	2	7
		Acc	0.84	0.87	0.91
		Kappa	0.46	0.58	0.73
			Validation		
		Acc	0.85	0.88	0.91
		Kappa	0.50	0.60	0.72
		Sens	0.71	0.78	0.84
		Spec	0.87	0.89	0.92

[1] Type of resampling methods tested, where RF: random forest model; cv: cross validation; oob: out-of-bag; and boot: bootstrapping. [2] Metrics used to evaluate the prediction performance, where mtry: hyperparameter that regulates the number of variables that can be randomly sampled in each split of the trees; Acc: Accuracy; Kappa coefficient; Sens: model sensitivity; and Spec: model specificity. [3] Type of data included in the model, where Terrain: TAGEE data; SySI: the six bands from SySI; and All: combination of TAGEE and SySI data.

The models containing only terrain covariates had good accuracies ranging from 0.84 to 0.85, but low Kappa coefficients ranging from 0.46 to 0.50 (Table 4). SySI performed better than the terrain covariates, with accuracies ranging from 0.87 to 0.88, and Kappa coefficients ranging from 0.58 to 0.63 (Table 4). Finally, the optimal model was achieved with terrain and SySI variables, using RF with CV as the resampling method (Table 4).

Table 5: Confusion matrix regarding the predicted classes of hydromorphic soils for the study area.

	Class	Reference		Total	UA	OE	CE
		H	NH				
Predicted	H	**1027**	299	1326	77%	23%	13%
	NH	148	**5059**	5207	97%	3%	6%
	Total	1175	5358	6533			
	PA	87%	94%		**6086**		

UA: user accuracy; OE omission error; CE: commission error; PA: producer accuracy. Bold represents the sum of the major diagonal (The total correctly classified soil classes).

The confusion matrix represents the RF model with CV, reaching a PA of 87% for the hydromorphic and 94% for the not hydromorphic class (Table 5). The model was able to correctly classify 6086 out of 6533 soil observations displaced across the study area. The prediction performance corroborates with the previous analyses which highlighted the terrain covariates as a major information to differentiate the location of hydromorphic soils.

4.3.3. Variables' importance

After fitting the RF model to predict hydromorphic soils, we evaluated the importance of each environmental variable (Fig. 5). The slope was the most important covariate for the model prediction, contributing with almost 100% (Fig. 5). This result is corroborated by the boxplot analysis, showing that hydromorphic soils are normally located in specific landforms (Fig. 2).

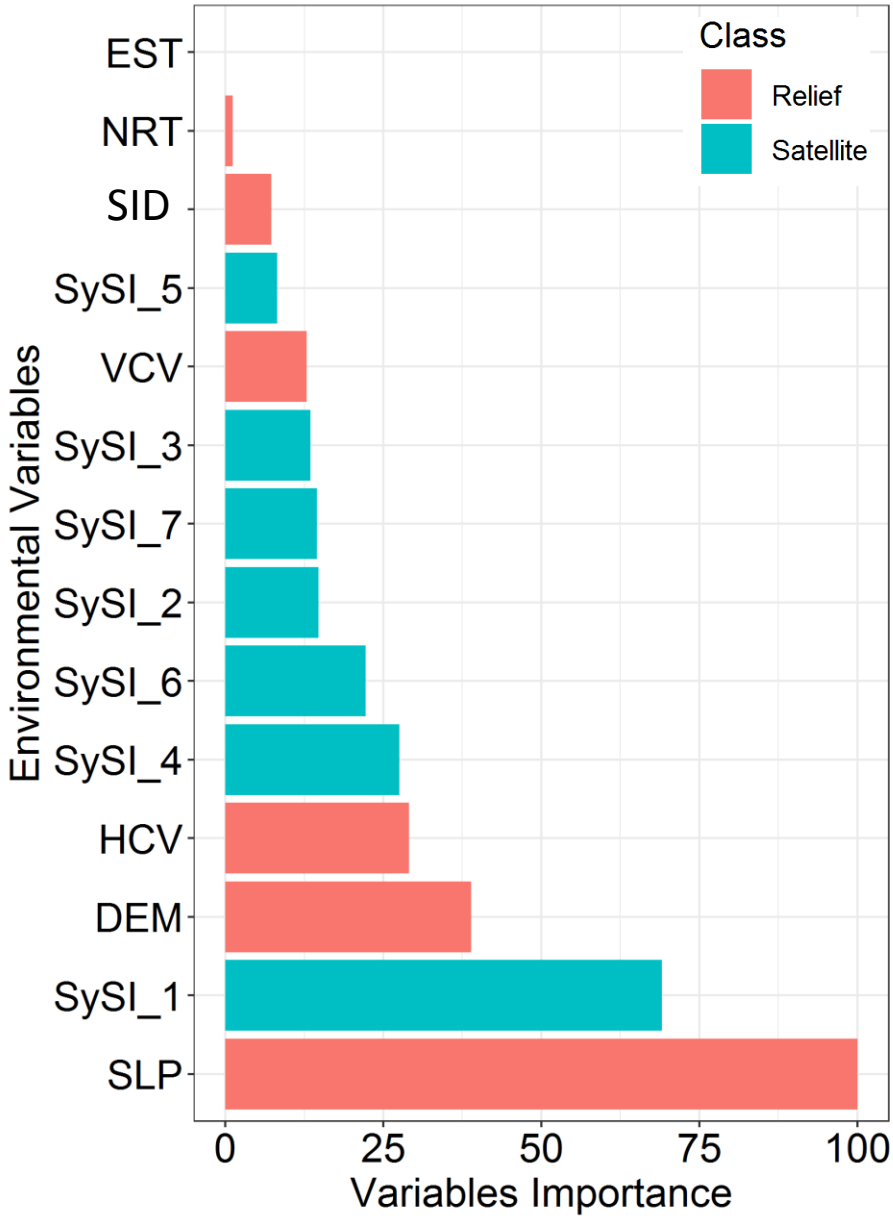

Figure 5: Graphics of Variables' importance for the prediction of hydromorphic soils with random forest (RF) model. The variables are separated by relief and satellite classes.

SySI had similar contribution for bands 2, 3, 4, 6, and 7, while band 1 had the second highest contribution for the model with 69% (Fig. 5). Although the soils affected by hydromorphism were identified by SySI, the terrain covariates had strong contribution for the prediction model (Fig. 5). This pattern reinforces the importance of relief for soil modelling even at large scales.

Contrary to the boxplot analysis, the elevation data had high contribution for the model prediction (Fig. 5). The elevation (altitude) was not significantly different between hydromorphic and not hydromorphic soils according to the boxplot analysis (Fig. 2). However, it had 38% of importance for classifying areas with hydromorphism, which highlights the relevance of relief for soil modelling.

The horizontal curvature, which indicates concave and convex landforms, had 29% of importance (Fig. 5). This covariate was more important than five SySI bands, indicating the specific terrain condition in which these soils

are formed. The remaining terrain covariates showed low contribution to the prediction model, confirming that hydromorphic soils are not related to geographical position as well as the local landform (Fig. 5).

4.3.4. Digital soil mapping (DSM)

After fitting the RF prediction model with cross validation, we applied it to the raster covariates and classified the pixels as hydromorphic and not hydromorphic soils. The result was a binary raster map in which one class represented soils with hydromorphic conditions and the other class represented regular soils (Fig. 6a). The hydromorphic class was not regularly distributed across the study area, indicating a possible restriction to the formation of these soils. The restriction factor can be related to topographic position, climate, annual mean precipitation (water availability), and others.

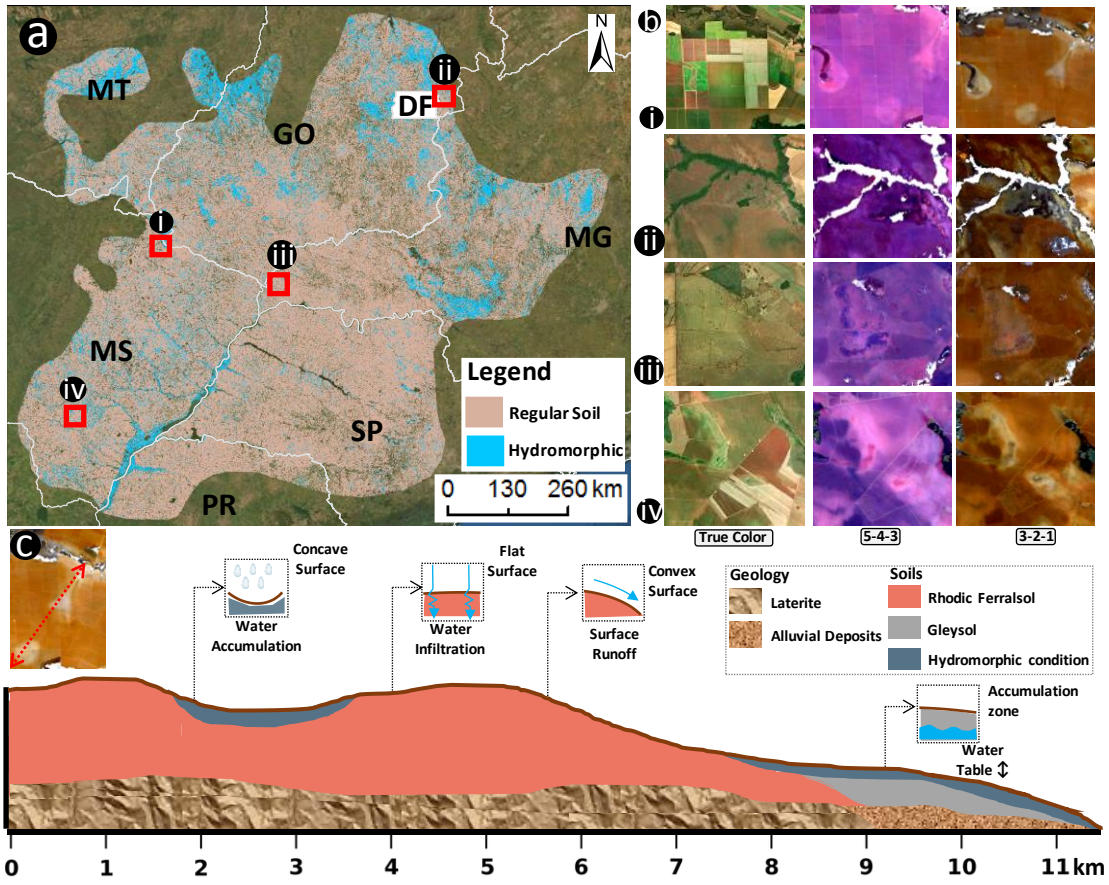

Figure 6: Results of the modelling and the relief pattern of areas classified as hydromorphic. a. Predicted map of hydromorphic soils for the study area; b. indication of SySI as a tool to identify hidden hydromorphic soils; c. toposequence extracted from area (i) indicating the relief positions of hydromorphic soils, the geology and soil classes (IBGE, 2019).

We selected four locations to observe and analyze the areas classified as hydromorphic soils (Fig. 6b). The analysis showed soils with hydromorphic conditions at different landscape positions, but it was not possible to identify these soils using only the Landsat images in true color composition (Fig. 6b). We also displayed the SySI

images in RGB 543 composite with visual enhancement for the same locations (Demattê et al., 2018). It was possible to identify some bright and darker features in the images, showing differences in the reflectance factor (Fig. 6b).

However, the SySI in RGB 321 composite was the best band composition to visually identify the features regarding hydromorphism (Fig. 6b). The differences in the reflectance factor usually indicates textural variation, Fe oxides presence, higher organic carbon content, soil moisture, and others (Mello et al., 2021; Mendes et al., 2022; Naimi et al., 2021; Silvero et al., 2021). In this case, the locations presented bright and dark pixels regarding hydromorphic soils, depending on the landscape position and landform (Fig. 6b). Figure 6bi presented three possible hydromorphic soils, one surrounding a channel's water source enhancing the possible area of Gleysols. In the same location there is an intermittent channel of second order (Strahler, 1957), which is an overland flow path. Finally, a closed depression with concave landform that functions as an accumulation zone (Fig. 6bi).

Figure 6bii represents a footslope, a flat landform next to a thalweg and normally influenced by groundwater level fluctuations (Schaetzl and Anderson, 2005). The Landsat 321 RGB composite shows an area with lower reflectance intensity at the footslope, indicating a possible Gleysol (Fig. 6bii). Figure 6biii and biv show a flat are at the summit and an overland flow path with the same features of lower reflectance intensity. These areas are displaced across the landscape, retaining water within the soil due to landforms or a soil characteristic that hampers water infiltration (Mello et al., 2021b).

We selected a location within the study area to explore the distribution of hydromorphic soils (Fig. 6bi). We plotted the relief profile in a toposequence to evaluate how the relief contributes for the formation of a hydromorphic soil (Fig. 6c). Although the predicted map proved how SySI was capable to classify the pixels as hydromorphic, the formation of these soils is also dependent on the landforms as demonstrated previously in the slope and horizontal curvature boxplots (Fig. 2).

The toposequence showed a hydromorphic soil located over a Ferralsol at the summit position (Fig. 6c). The hydromorphism is a clear product of the concave landform, which orientates the overland water flow to the same direction, forming a water accumulation zone. The other hydromorphic soil is located next to the flow channel at the footslope, over a Gleysol, and affected by groundwater level fluctuation (Fig. 6c). The soil map was able to identify the hydromorphic soil at the footslope but not the one at the summit, since the map is a legacy data with a coarse scale of 1:250.000 (Fig. 6c).

The hydromorphic soils occurred over two soil types, a Ferralsol and a Gleysol (Fig. 6c). The Ferralsols are weathered soils normally located at flat surfaces, which favor water infiltration and prevent the formation of drainage channels (Buol et al., 2011). The Ferralsol was also located at the hillslope with convex surface, favoring surface runoff and drainage channel formation (Fig. 6c). The Gleysols are normally located at lower relief positions that constantly receives and accumulates sediment and water, favoring redoximorphic activity in the soils (Schaetzl and Anderson, 2005).

4.4. Discussion

4.4.1. SySI reflectance of hydromorphic soils

The features observed in SySI were often near drainage network systems, which are zones normally affected by groundwater level fluctuations and that concentrate the drained water from upper positions of the watershed (Ahmadi and Sedghamiz, 2007; Jan et al., 2007). This observation suggested the influence of soil moisture in the reflectance response, causing differences in color intensity similar to the features found by Haubrock et al. (2008).

The soil moisture content affects the albedo, decreasing the reflectance factor throughout the spectral range (Chabrillat et al., 2019).

However, the spectral analysis of the soil locations with the SySI hydromorphic feature showed higher reflectance than the nearby locations without the SySI feature (Fig. 4). This was caused by a low presence of Fe oxides and higher contents of quartz (sand and silt fractions) in the soil surface, a result from permanent or periodic saturation of the soil by water (Duchaufour, 1982). The satellite reflectance only retrieves information about the surface, although it can be related to subsurface characteristics (Bonfatti et al., 2020; Gallo et al., 2018; Mello et al., 2021a; Mendes et al., 2019). The surface of Gleysols (Hydromorphic) are not necessarily wet, contrary to the subsurface where water saturation promotes anoxic conditions (Ker et al., 2015). Nonetheless, the surface mineral and textural properties can be affected by the water saturation from below (groundwater) or from above (rain or irrigation water) and removal or reduction of Fe3+ to Fe2+ (van Breemen and Buurman, 2002).

Thus, we were able to confirm the occurrence of hydromorphism at the locations with the SySI feature (Fig. 1). The Vis-NIR-SWIR laboratory analysis showed the absence of Fe oxides (oxides and oxyhydroxides) features in the region of 900 nm (Fig. 4). In both clayey and sandy soils, the Fe oxides features were present in the nearby soil locations, but not in the site with the hydromorphic feature observed in SySI (Fig. 4a,b).

The use of satellite spectra to predict hydromorphic soils proved to be an efficient and robust tool, with accuracy similar to Rapinel et al. (2019), that reached 82% of accuracy on the attempt to map multiple classes of wetlands through the combination of satellite, radar, and terrain information, and field locations classified as hydromorphic and non-hydromorphic soils. Whyte et al. (2018) also reached similar results for predicting different land use classes, 83.3% of accuracy, using satellite, radar, and terrain information. Although remote sensed data has been used to predict wetlands, SySI offers an approach without radar information that can classify areas with indication of hydromorphism, an advance over multiple studies regarding the mapping of hydromorphic soils (Chaplot et al., 2003, 2000; Pennock et al., 2014; Thompson et al., 1997).

4.4.2. Effects of environmental variables on hydromorphic soil prediction

SySI reflectance was determinant for the identification and classification of hydromorphic soils. SySi has been used as a covariate to predict multiple soil attributes (Bonfatti et al., 2020; Gallo et al., 2018; Mello et al., 2022; Naimi et al., 2021; L. J. Safanelli et al., 2020; Silvero et al., 2021). However, SySI was not used for the identification of soils affected by hydromorphism. After analyzing the laboratory Vis-NIR-SWIR spectra of gleysols and planosols, it was possible to affirm that the satellite spectra had features related to hydromorphic soils (Fig. 3 and 4). The modelling tests using SySI alone had results comparable to other works that mapped wetlands and hydromorphic soils (Lee et al., 2019; Rapinel et al., 2019; Whyte et al., 2018), proving the efficiency of the tool (Table 4).

The terrain covariates derived from TAGEE also had major importance for the prediction performance (Table 4). The statistical analysis also showed that the slope and horizontal curvature were significantly different for hydromorphic soils, even at a large scale framework (Fig. 2). The relief is a soil forming factor, determines the water flow direction, and the groundwater height (Schaetzl, J Randall and Anderson, 2005; van Breemen and Buurman, 2002). The formation of a hydromorphic soil relief on the types of landforms with long flat floodplains, with predominance of alluvial and colluvial deposits (Miller and Juilleret, 2020), closed depressions where water concentrates, and concave landforms across the landscape (Milne, 1935; Ruhe, 1960; Schaetzl and Anderson, 2005). Curmi et al. (1998) also pointed that topography determines the position of hydromorphic soils, where well drained

and hydromorphic soils are positioned at different landforms and separated by a transition zone whose extension depends on the slope.

The study area represented a large portion of southeast, mid-west, and southern regions of Brazil (Fig. 1). We used terrain information with 30x30 m of spatial resolution, a large scale information for a large study area. The model improvement and the relevance of relief on the formation of hydromorphic soils indicate that more terrain data should be considered for the classification and monitoring of hydromorphic soils.

4.4.3. Environmental pressures on hydromorphic soils

The hydromorphic soils identified by the suggested protocol (e.g. Gleysols, and Planosols) have multiple environmental functions, such as protecting living organisms, determine the flood amplitude, serves as a nutrient recycling and transfer environment, and works as a water recharge source for drainage channels (Junk, 2002; Magha et al., 2021; Maltby, 1991). The use of these soils for agricultural purposes promotes an environmental pressure and requires careful management practices, since it can produce morphological alterations, alterations in the landscape (Buol et al., 2011).

According to the Brazilian Forest Code (Ahrens, 2003), riverbeds up to 10 m of width require a permanent preservation area (PPA) of 30 m on both sides of the channel. The code also establishes a PPA of 50 m for the water source, in order to protect these natural resources from pollution, silting, and other risks (Sparovek et al., 2011). However, the technique explored in this work showed cases where intermittent rivers and water sources were not protected, or the protected area did not represent the actual resource that needed to be preserved (Fig. 7). The SySI RGB 321 composite with visual enhancement indicated many intermittent channels and water sources unpreserved and explore for agricultural purposes (Fig. 7a, b).

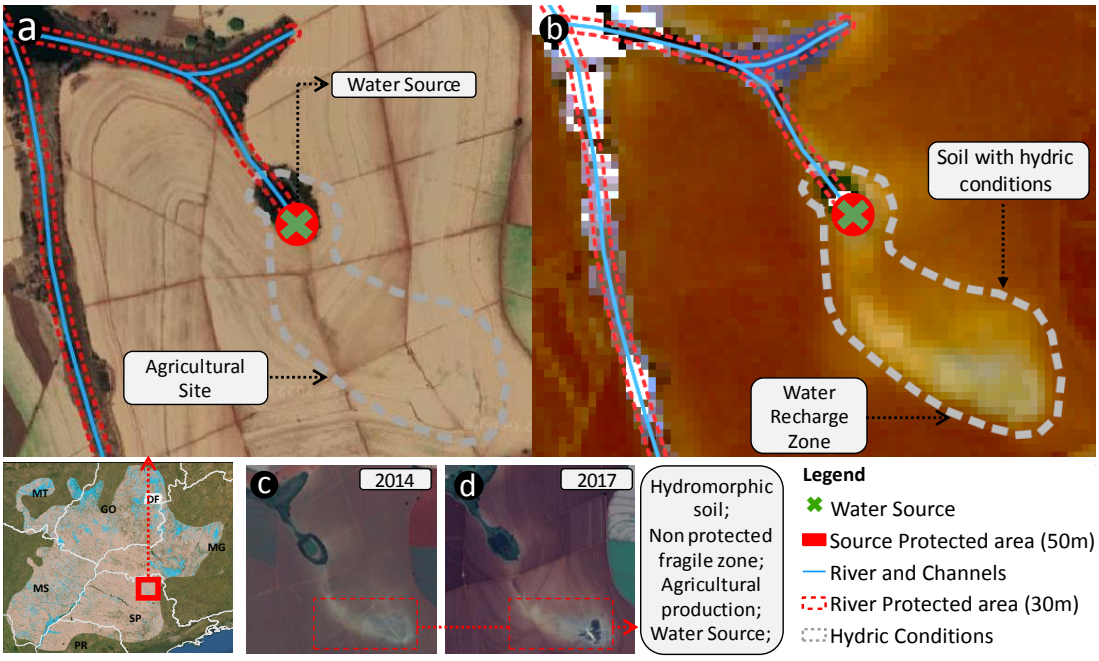

Figure 7: Example of hydromorphic soils within an agricultural site. a. the supposed water source is protected with vegetation; b. SySI identifies abrupt difference in reflectance indicating hydromorphic conditions; c and d. Landsat true color in 2014 and 2017 with exposed soil showing accumulation of water in the actual water source.

Figure 7a shows an agricultural site with a PPA surrounding the channels and the water source. Observing the Landsat RGB true color composite, there is hydromorphic soils and no apparent irregularities at the site. However, figure 7b show a prolonged area with soils with hydric conditions, suggesting that a larger area should be preserved. The hydromorphic soil is hidden under an agricultural area in contact with chemical fertilizers and pesticides, common for management practices. Vlasenko (2009) found many waterlogged lands in agricultural areas using aerial photographs. The author also related the development of hydromorphic sites due to changes in the physiographical features of the territory, the development of subsidence, the appearance of new depressions, and the expansion of the existing ones (Vlasenko, 2009).

Agricultural practices over hydromorphic soils can accelerate nutrient loss, affect particle aggregation, distribution and mineralogy of Fe oxides between particle-size fractions, and the interaction with organic matter stabilization (Giannetta et al., 2022; Shaheen et al., 2022). Hydromorphic soils also increase soil organic carbon reserves, which potentially improves water infiltration and minimizes runoff, which supplies the water table (da Silveira et al., 2022; Likhanova et al., 2022).

Two Landsat images from the same area in 2014 and 2017 indicate water accumulation in a different position of the PPA (Fig. 7c,d). The images indicate a different location for the water source and an intermittent channel connecting with the actual PPA at the site. The technique identified these soils and showed that multiple hydromorphic areas are located at agricultural sites, promoting degradation (Table 5).

Table 6: Predicted areas of hydromorphic soils for the study area analyzed by federal states and land use classifications according to Souza et al. (2020).

		Area (km²)		
Study Area	TA (km²)	*NH*	*H*	*H (%)*
	735953,8	629605,7	106348,2	14,5
States	PA (%)	*NH*	*H*	*H (%)*
Federal District	89.4	3074.2	2077.5	40.3
Goiás	57.9	162072.2	35116.5	17.8
Minas Gerais	22.8	112925.2	20646.0	15.5
Mato Grosso do Sul	44.8	137812.8	22293.0	13.9
Mato Grosso	5.8	37488.6	14617.5	28.1
Paraná	19.9	37065.8	2646.1	6.7
São Paulo	59.7	139166.9	8951.4	6.0
Land Use	TA (km²)	*NH*	*H*	*H (%)*
Sugar cane	86872.4	84209.1	2663.3	3.1
Soybean	113553.9	97037.2	16516.7	14.5
Forest Plantation	19771.9	17728.0	2043.8	10.3
Pasture	267460.2	227509.8	39950.3	14.9
Temporary Crop	11439.8	10015.0	1424.8	12.5
Perennial Corp	5345.4	4926.1	419.3	7.8

TA: total area of each land use class; PA: percentage area of each state included in the analysis; NH: not hydromorphic soil; H: hydromorphic soil.

Although the Brazilian Forest Code is a federal law applied for all states of the federation and any agricultural activity, there is a large difference between the percentages of hydromorphic soils by state and land use (Table 5). Furthermore, SySI information regards soils that were exposed at least one time through the Landsat TM time series (Demattê et al., 2018), which are predominantly agricultural soils.

The São Paulo state had only 6% of its area classified as hydromorphic, while Mato Grosso do Sul, Minas Gerais, and Goiás had up to 17.8% (Table 5). Mato Grosso had only 5.8% of its area included in this work, by had 28% classified as hydromorphic. These differences in hydromorphic areas require further investigations on the correct application Forest Code, since these states represent a large part of the Brazilian agriculture (IBGE, 2011).

The sugar cane areas had 3.1% of hydromorphic areas, an indication of adequate application of PPAs for a major agricultural activity in the country (Bordonal et al., 2018). Cherubin et al. (2016) explained that sugar cane expansion reintegrates degraded pasturelands into a more productive system, improving soil quality. On the other hand, pastures across the study area had 14.9% of hydromorphic areas, enhancing the risks of contamination since this land use is normally degraded due to due to low grass productivity and inadequate grazing management (Bordonal et al., 2018, 2017; Oliveira et al., 2016). Although sugar cane activity also impacts the environment, there is a smaller presence of hydromorphic soils, indicating a better management and application of PPA's law (Ahrens, 2003).

The soybean areas also had around 14% of areas classified as hydromorphic, followed by forest plantation and temporary crops with 10 and 12%, respectively (Table 5). Soybean is the predominant crop system in the states of Mato Grosso, Mato Grosso do Sul, and Goiás (Bonato and Dall'Agnol, 2022), which were the states with higher percentage of hydromorphic soils (Table 5). Sugar cane is the predominant agricultural system in São Paulo state (Bordonal et al., 2018), which explains the lower occurrence of hydromorphic soils. This result highlights the need to verify if the PPAs represent all the drainage network and further natural resources that need preservation.

The observed differences in hydromorphic soils regarding federal states and land use suggest a further investigation on the application of the Brazilian Forest Code. Multiple initiatives discussed the real benefits and limitations of the defined areas of preservation, discussing whether they should be larger in order to preserve natural resources (A. et al., 2017; da Silva et al., 2017; Stickler et al., 2013). However, this technique was able to identify multiple channel networks (intermittent and perennial) and water sources inside agricultural sites (Fig. 6 and 7). Finally, more remote sensing tools must be applied to monitor the control natural resources, which should be preserved and well managed in order to avoid future degradation (Goldshleger et al., 2010; Shoshany et al., 2013; Žížala et al., 2019).

4.5. Conclusions

In this study, we reached the following conclusions: i) The observed features in SySI were confirmed as hydromorphic soils through Vis-NIR-SWIR laboratory spectral analysis, allowing its use for land classification and mapping; ii) The relief covariates regarding curvature and slope were important to improve the prediction performance along with SySI to classify hydromorphic soils; iii) The catchment and accumulation areas were underestimated by land owners at multiple locations, resulting in unprotected channel network and water sources; iv) São Paulo state, known for its high sugar cane production, had the lowest area percentage for hydromorphic soils, while Mato Grosso, Goiás, Mato Grosso do Sul, e Minas Gerais had up to 28% of hydromorphic soil areas; v) Sugar cane showed the lowest percentage of hydromorphic soils, while soybean, pasture, and temporary crops had up to 14.9%, showing the imbalance on the monitoring and application of the Brazilian Forest Code.

Finally, SySI proved to be a robust tool on the identification and mapping of hydromorphic soils for agricultural areas, enhancing the possibilities to use remote sensing information to improve the monitoring of agricultural areas.

Acknowledgments

We are grateful to Geotechnologies in Soil Science Group (GEOSS/ESALQ-USP; http://esalqgeocis.wixsite.com/english) for team support. We also thank the anonymous reviewers to improve this paper. The scholarship of the first author was provided by the Coordination of Superior Level Staff Improvement (CAPES). Funding was provided by São Paulo Research Foundation (FAPESP, grant numbers 2014/22262-0,2016/01597-9, 2016/26124-6, 2016/01597-9, 2018/12532-0).

References

1. Wu, H.; Clark, H. The Sustainable Development Goals: 17 Goals to Transform Our World. In Furthering the Work of the United Nations; United Nations, 2016; pp. 36–54.

2. Godfray, H.C.J.; Beddington, J.R.; Crute, I.R.; Haddad, L.; Lawrence, D.; Muir, J.F.; Pretty, J.; Robinson, S.; Thomas, S.M.; Toulmin, C. Food Security: The Challenge of Feeding 9 Billion People. Science (80-.). 2010, 327, 812 LP – 818, doi:10.1126/science.1185383.

3. McBratney, A.; Field, D.J.; Koch, A. The Dimensions of Soil Security. Geoderma 2014, 213, 203–213, doi:10.1016/j.geoderma.2013.08.013.

4. Bouma, J.; McBratney, A. Framing Soils as an Actor When Dealing with Wicked Environmental Problems. Geoderma 2013, 200–201, 130–139, doi:https://doi.org/10.1016/j.geoderma.2013.02.011.

5. Baker, R.M. Soil Resilience and Sustainable Land Use; Szabolos, D.J., I., G. and, Eds.; 2008/10/03.; Cambridge University Press: Wallingford, Oxfordshire, 1994; Vol. 31;.

6. Lal, R.; Stewart, B.A. Food Security and Soil Quality; Lal, R., Stewart, B.A., Eds.; 1st ed.; CRC Press: Boca Raton, FL, 2010; ISBN 0429130538.

7. Dokuchaev, V. V The Russian Chernozem Report to the Free Economic Society. Imp. Univ. St. Petersbg. St. Petersburg)[in Russ. 1883.

8. Jenny, H. Factors of Soil Formation: A System of Quantitative Pedology; Courier Corporation, 1994; ISBN 0486681289.

9. Banwart, S.; Bernasconi, S.M.; Bloem, J.; Blum, W.; Brandao, M.; Brantley, S.; Chabaux, F.; Duffy, C.; Kram, P.; Lair, G.; et al. Soil Processes and Functions in Critical Zone Observatories: Hypotheses and Experimental DesignAll Rights Reserved. No Part of This Periodical May Be Reproduced or Transmitted in Any Form or by Any Means, Electronic or Mechanical, Including Photocopying, . Vadose Zo. J. 2011, 10, 974–987, doi:10.2136/vzj2010.0136.

10. Chorover, J.; Kretzschmar, R.; Garcia-Pichel, F.; Sparks, D.L. Soil Biogeochemical Processes within the Critical Zone. Elements 2007, 3, 321–326, doi:10.2113/gselements.3.5.321.

11. Mitsch, W.J.; Bernal, B.; Hernandez, M.E. Ecosystem Services of Wetlands. Int. J. Biodivers. Sci. Ecosyst. Serv. Manag. 2015, 11, 1–4, doi:10.1080/21513732.2015.1006250.

12. Santana, S.E.; Barroso, G.F. Integrated Ecosystem Management of River Basins and the Coastal Zone in Brazil. Water Resour. Manag. 2014, 28, 4927–4942, doi:10.1007/s11269-014-0754-4.

13. Lehrback, B.D.; Neto, R.R.; Barroso, G.F.; Bernardes, M.C. Sources and Distribution of Sedimentary Organic Matter in the Northwestern Portion of Victoria Bay, ES. Brazilian J. Aquat. Sci. Technol. 2016, 20, 79–92.

14. Buol, S.W.; Rebertus, R.A. Soil Formation under Hydromorphic Conditions. In The Ecology and Management of Wetlands; Hook, D.D., McKee, W.H., Smith, H.K., Gregory, J., Burrell, V.G., DeVoe, M.R., Sojka, R.E., Gilbert, S., Banks, R., Stolzy, L.H., Brooks, C., Matthews, T.D., Shear, T.H., Eds.; Springer US: New York, NY, 1988; pp. 253–262 ISBN 978-1-4684-8378-9.

15. van Breemen, N.; Buurman, P. Hydromorphic Soils. In Soil Formation; Springer Netherlands: Dordrecht, 1998; pp. 151–181 ISBN 978-0-585-31788-5.

16. Duchaufour, P. Hydromorphic Soils BT - Pedology: Pedogenesis and Classification. In; Duchaufour, P., Ed.; Springer Netherlands: Dordrecht, 1982; pp. 335–372 ISBN 978-94-011-6003-2.

17. Buol, S.W.; Southard, R.J.; Graham, R.C.; McDaniel, P.A. Soil-Forming Processes. In Soil Genesis and Classification; Buol, S.W., Southard, R.J., Graham, R.C., McDaniel, P.A., Eds.; Wiley Online Books; 2011; pp. 163–179 ISBN 9780470960622.

18. Bedard-Haughn, A. Gleysolic Soils of Canada: Genesis, Distribution, and Classification. Can. J. Soil Sci. 2011, 91, 763–779, doi:10.4141/cjss10030.

19. Özcan, H.; Ekinci, H.; Sungur, A.; Everest, T. Gleysols. In The Soils of Turkey; Kapur, S., Akça, E., Günal, H., Eds.; Springer International Publishing: Cham, 2018; pp. 313–320 ISBN 978-3-319-64392-2.

20. Schaetzl, R.; Anderson, S. Soils. Genesis and Geomorphology; Cambridge University Press: New York, 2005; ISBN 9780521812016.

21. Pavlović, P.; Kostić, N.; Karadžić, B.; Mitrović, M. Order of Hydromorphic Soils. In The Soils of Serbia; Pavlović, P., Kostić, N., Karadžić, B., Mitrović, M., Eds.; Springer Netherlands: Dordrecht, 2017; pp. 157–173 ISBN 978-94-017-8660-7.

22. Ker, J.C.; Curi, N.; Schaefer, C.E.G.R.; Vidal-Torrado, P. Pedologia: Fundamentos. 2015.

23. dos Santos, H.G.; Jacomine, P.K.T.; Dos Anjos, L.H.C.; De Oliveira, V.A.; Lumbreras, J.F.; Coelho, M.R.; de Almeida, J.A.; de Araujo Filho, J.C.; de Oliveira, J.B.; Cunha, T.J.F. Sistema Brasileiro de Classificação de Solos.; Brasília, DF: Embrapa, 2018., 2018; ISBN 8570358172.

24. Zou, Y.; Wang, L.; Xue, Z.; E, M.; Jiang, M.; Lu, X.; Yang, S.; Shen, X.; Liu, Z.; Sun, G.; et al. Impacts of Agricultural and Reclamation Practices on Wetlands in the Amur River Basin, Northeastern China. Wetlands 2018, 38, 383–389, doi:10.1007/s13157-017-0975-4.

25. Gebresllassie, H.; Gashaw, T.; Mehari, A. Wetland Degradation in Ethiopia: Causes, Consequences and Remedies. J. Environ. Earth Sci. 2014, 4, 40–48.

26. Tilman, D. Global Environmental Impacts of Agricultural Expansion: The Need for Sustainable and Efficient Practices. Proc. Natl. Acad. Sci. 1999, 96, 5995 LP – 6000, doi:10.1073/pnas.96.11.5995.

27. Miller, J.C. Nitrate Contamination of the Water-Table Aquifer in Delaware; Newark, DE: Delaware Geological Survey, University of Delaware, 1972;

28. Bera, A.; Mukhopadhyay, B.P.; Chowdhury, P.; Ghosh, A.; Biswas, S. Groundwater Vulnerability Assessment Using GIS-Based DRASTIC Model in Nangasai River Basin, India with Special Emphasis on Agricultural Contamination. Ecotoxicol. Environ. Saf. 2021, 214, 112085, doi:https://doi.org/10.1016/j.ecoenv.2021.112085.

29. Yang, L.; Zheng, C.; Andrews, C.B.; Wang, C. Applying a Regional Transport Modeling Framework to Manage Nitrate Contamination of Groundwater. Groundwater 2021, 59, 292–307, doi:https://doi.org/10.1111/gwat.13047.

30. Nath, A.; Samanta, S.; Banerjee, S.; Danda, A.A.; Hazra, S. Threat of Arsenic Contamination, Salinity and Water Pollution in Agricultural Practices of Sundarban Delta, India, and Mitigation Strategies. SN Appl. Sci. 2021, 3, 560, doi:10.1007/s42452-021-04544-1.

31. Ritter, W.F. Pesticide Contamination of Ground Water in the United States - A Review. J. Environ. Sci. Heal. Part B 1990, 25, 1–29, doi:10.1080/03601239009372674.

32. Goss, M.J.; Barry, D.A.J.; Rudolph, D.L. Contamination in Ontario Farmstead Domestic Wells and Its Association with Agriculture:: 1. Results from Drinking Water Wells. J. Contam. Hydrol. 1998, 32, 267–293, doi:https://doi.org/10.1016/S0169-7722(98)00054-0.

33. Piedade, M.T.F.; Junk, W.J.; de Sousa Jr, P.T.; da Cunha, C.N.; Schöngart, J.; Wittmann, F.; Candotti, E.; Girard, P. As Áreas Úmidas No Âmbito Do Código Florestal Brasileiro. Código Florest. ea ciência o que nossos Legis. ainda precisam saber. Sumários Exec. Estud. científicos sobre impactos do Proj. Código Florest. 2012, 9–17.

34. Sparovek, G.; Barretto, A.; Klug, I.; Papp, L.; Lino, J. A Revisão Do Código Florestal Brasileiro. Novos Estud. CEBRAP 2011, 89, 111–135, doi:https://doi.org/10.1590/S0101-33002011000100007.

35. Mello, F.A.O.; Demattê, J.A.M.; Rizzo, R.; Dotto, A.C.; Poppiel, R.R.; Mendes, W.S.; Guimarães, C.C.B. Expert-Based Maps and Highly Detailed Surface Drainage Models to Support Digital Soil Mapping. Geoderma 2021, 384, 114779, doi:https://doi.org/10.1016/j.geoderma.2020.114779.

36. IUSS Working Group, W.R.B. World Reference Base for Soil Resources. World Soil Resour. Rep. 2015, 103.

37. Staff, S.S. Keys to Soil Taxonomy. United States Dep. Agric. Soil Conserv. Serv. Washington, DC 2010.

38. Kadlec, R.H.; Wallace, S. Treatment Wetlands; CRC press, 2008; ISBN 1420012517.

39. Alvares, C.A.; Stape, J.L.; Sentelhas, P.C.; De Moraes Gonçalves, J.L.; Sparovek, G. Köppen's Climate Classification Map for Brazil. Meteorol. Zeitschrift 2013, 22, 711–728, doi:10.1127/0941-2948/2013/0507.

40. 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.; et al. The Brazilian Soil Spectral Library (BSSL): A General View, Application and Challenges. Geoderma 2019, 354, 113793, doi:https://doi.org/10.1016/j.geoderma.2019.05.043.

41. Cooper, M.; Mendes, L.M.S.; Silva, W.L.C.; Sparovek, G. A National Soil Profile Database for Brazil Available to International Scientists. Soil Sci. Soc. Am. J. 2005, 69, 649–652, doi:https://doi.org/10.2136/sssaj2004.0140.

42. Radambrasil, P. Levantamento de Recursos Naturais; Ministério das Minas e Energia, Departamento Nacional da Produção Mineral …, 1973;

43. Samuel-Rosa, A.; Dalmolin, R.S.D.; Moura-Bueno, J.M.; Teixeira, W.G.; Alba, J.M.F. Open Legacy Soil Survey Data in Brazil: Geospatial Data Quality and How to Improve It. Sci. Agric. 2020, 77, 1–11, doi:10.1590/1678-992x-2017-0430.

44. Soil Science Division Staff Soil Survey Manual; Ditzler, C., Scheffe, K., Monger, H.C., Eds.; Government Printing Office Washington, DC, 2017;

45. Santos, H.G.; Hochmüller, D.P.; Cavalcanti, A.C.; Rêgo, R.S.; Ker, J.C.; Panoso, L.A.; Amaral, J.A.M. do. Procedimentos Normativos de Levantamentos Pedológicos; Brasília, DF: EMBRAPA-SPI; Rio de Janeiro: EMBRAPA-CNPS, 1995., 1995; ISBN 85-85007-58-3.

46. dos Santos, H.G.; CARVALHO JUNIOR, W. de; Dart, R. de O.; Áglio, M.L.D.; de Sousa, J.S.; Pares, J.G.; Fontana, A.; Martins, A.L. da S.; de OLIVEIRA, A.P. O Novo Mapa de Solos Do Brasil: Legenda Atualizada; Embrapa Solos: Rio de janeiro, 2011;

47. Demattê, J.A.M.; Fongaro, C.T.; Rizzo, R.; Safanelli, J.L. Geospatial Soil Sensing System (GEOS3): A Powerful Data Mining Procedure to Retrieve Soil Spectral Reflectance from Satellite Images. Remote Sens. Environ. 2018, 212, 161–175, doi:10.1016/j.rse.2018.04.047.

48. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone. Remote Sens. Environ. 2017, 202, 18–27, doi:10.1016/J.RSE.2017.06.031.

49. U.S.G.S. Landsat 4-7 Surface Reflectance (Ledaps) Product Guide. 2019, 32, doi:10.1016/0042-207X(74)93024-3.

50. U.S.G.S. Landsat 8 Surface Reflectance Code (LASRC) Poduct Guide. (No. LSDS-1368 Version 2.0). 2019, 40.

51. Safanelli, J.L.; Poppiel, R.R.; Ruiz, L.F.; Bonfatti, B.R.; Mello, F.A.; Rizzo, R.; Demattê, J.A.M. Terrain Analysis in Google Earth Engine: A Method Adapted for High-Performance Global-Scale Analysis. ISPRS Int. J. Geo-Information 2020, 9.

52. 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. Improvement of Clay and Sand Quantification Based on a Novel Approach with a Focus on Multispectral Satellite Images. Remote Sens. 2018, 10, doi:10.3390/rs10101555.

53. Safanelli, L.J.; Chabrillat, S.; Ben-Dor, E.; Demattê, A.M.J. Multispectral Models from Bare Soil Composites for Mapping Topsoil Properties over Europe. Remote Sens. 2020, 12.

54. 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.; et al. Bare Earth's Surface Spectra as a Proxy for Soil Resource Monitoring. Sci. Rep. 2020, 10, 4461, doi:10.1038/s41598-020-61408-1.

55. Bedard-Haughn, A.K.; Pennock, D.J. Terrain Controls on Depressional Soil Distribution in a Hummocky Morainal Landscape. Geoderma 2002, 110, 169–190, doi:https://doi.org/10.1016/S0016-7061(02)00229-X.

56. Wickham, H. Ggplot2. WIREs Comput. Stat. 2011, 3, 180–185, doi:https://doi.org/10.1002/wics.147.

57. R Core Team R: A Language and Environment for Statistical Computing. 2013.

58. Lobell, D.B.; Asner, G.P. Moisture Effects on Soil Reflectance. Soil Sci. Soc. Am. J. 2002, 66, 722–727, doi:https://doi.org/10.2136/sssaj2002.7220.

59. Weidong, L.; Baret, F.; Xingfa, G.; Qingxi, T.; Lanfen, Z.; Bing, Z. Relating Soil Surface Moisture to Reflectance. Remote Sens. Environ. 2002, 81, 238–246, doi:https://doi.org/10.1016/S0034-4257(01)00347-9.

60. Demattê, J.A.M.; Bellinaso, H.; Romero, D.J.; Fongaro, C.T. Morphological Interpretation of Reflectance Spectrum (MIRS) Using Libraries Looking towards Soil Classification. Sci. Agric. 2014, 71, 509–520.

61. Demattê, J.A.M.; Horák-Terra, I.; Beirigo, R.M.; Terra, F. da S.; Marques, K.P.P.; Fongaro, C.T.; Silva, A.C.; Vidal-Torrado, P. Genesis and Properties of Wetland Soils by VIS-NIR-SWIR as a Technique for Environmental Monitoring. J. Environ. Manage. 2017, 197, 50–62, doi:https://doi.org/10.1016/j.jenvman.2017.03.014.

62. Marques, K.P.; Rizzo, R.; Carnieletto Dotto, A.; Souza, A.B. e; Mello, F.A.; Neto, L.G.; Anjos, L.H.C. do.; Demattê, J.A. How Qualitative Spectral Information Can Improve Soil Profile Classification? J. Near Infrared Spectrosc. 2019, 27, 156–174, doi:10.1177/0967033518821965.

63. Ben-Dor, E.; Banin, A. Near-Infrared Analysis as a Rapid Method to Simultaneously Evaluate Several Soil Properties. Soil Sci. Soc. Am. J. 1995, 59, 364–372, doi:https://doi.org/10.2136/sssaj1995.03615995005900020014x.

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

65. Poppiel, R.R.; Lacerda, M.P.C.; Demattê, J.A.M.; Oliveira, M.P.; Gallo, B.C.; Safanelli, J.L. Pedology and Soil Class Mapping from Proximal and Remote Sensed Data. Geoderma 2019, 348, 189–206, doi:10.1016/j.geoderma.2019.04.028.

66. Teng, H.; Viscarra Rossel, R.A.; Shi, Z.; Behrens, T. Updating a National Soil Classification with Spectroscopic Predictions and Digital Soil Mapping. CATENA 2018, 164, 125–134, doi:https://doi.org/10.1016/j.catena.2018.01.015.

67. Zeraatpisheh, M.; Ayoubi, S.; Jafari, A.; Tajik, S.; Finke, P. Digital Mapping of Soil Properties Using Multiple Machine Learning in a Semi-Arid Region, Central Iran. Geoderma 2019, 338, 445–452, doi:https://doi.org/10.1016/j.geoderma.2018.09.006.

68. Khaledian, Y.; Miller, B.A. Selecting Appropriate Machine Learning Methods for Digital Soil Mapping. Appl. Math. Model. 2020, 81, 401–418, doi:https://doi.org/10.1016/j.apm.2019.12.016.

69. Taghizadeh-Mehrjardi, R.; Schmidt, K.; Eftekhari, K.; Behrens, T.; Jamshidi, M.; Davatgar, N.; Toomanian, N.; Scholten, T. Synthetic Resampling Strategies and Machine Learning for Digital Soil Mapping in Iran. Eur. J. Soil Sci. 2020, 71, 352–368, doi:10.1111/ejss.12893.

70. Breiman, L. Random Forests. Mach. Learn. 2001, 45, 5–32, doi:10.1007/9781441993267_5.

71. Gambill, D.R.; Wall, W.A.; Fulton, A.J.; Howard, H.R. Predicting USCS Soil Classification from Soil Property Variables Using Random Forest. J. Terramechanics 2016, 65, 85–92, doi:10.1016/j.jterra.2016.03.006.

72. Kuhn, M. Building Predictive Models in R Using the Caret Package. J. Stat. Softw. 2008, 28, 1–26, doi:10.18637/jss.v028.i05.

73. Probst, P.; Wright, M.N.; Boulesteix, A.-L. Hyperparameters and Tuning Strategies for Random Forest. WIREs Data Min. Knowl. Discov. 2019, 9, e1301, doi:https://doi.org/10.1002/widm.1301.

74. Mosier, C.I. I. Problems and Designs of Cross-Validation 1. Educ. Psychol. Meas. 1951, 11, 5–11, doi:10.1177/001316445101100101.

75. Browne, M.W. Cross-Validation Methods. J. Math. Psychol. 2000, 44, 108–132, doi:https://doi.org/10.1006/jmps.1999.1279.

76. Tibshirani, R.J.; Efron, B. An Introduction to the Bootstrap. Monogr. Stat. Appl. Probab. 1993, 57, 1–436.

77. Hawkins, D.M.; Basak, S.C.; Mills, D. Assessing Model Fit by Cross-Validation. J. Chem. Inf. Comput. Sci. 2003, 43, 579–586, doi:10.1021/ci025626i.

78. Efron, B. Bootstrap Methods: Another Look at the Jackknife. In Breakthroughs in statistics; Springer, 1992; pp. 569–593.

79. Shao, J. Bootstrap Model Selection. J. Am. Stat. Assoc. 1996, 91, 655–665, doi:10.1080/01621459.1996.10476934.

80. Bischl, B.; Mersmann, O.; Trautmann, H.; Weihs, C. Resampling Methods for Meta-Model Validation with Recommendations for Evolutionary Computation. Evol. Comput. 2012, 20, 249–275, doi:10.1162/EVCO_a_00069.

81. Hijmans, R.J.; van Etten, J.; Mattiuzzi, M.; Sumner, M.; Greenberg, J.A.; Lamigueiro, O.P.; Bevan, A.; Racine, E.B.; Shortridge, A. Raster Package in R 2013.

82. Congalton, R.G.; Green, K. Assessing the Accuracy of Remotely Sensed Data Principles and Practices; 3rd ed.; Taylor & Francis Group: Boca Raton, FL, 2019; ISBN 9788578110796.

83. Milne, G. Composite Units for the Mapping of Complex Soil Associations. Trans. 3rd Int. Congr. Soil Sci 1935, 1, 345–347.

84. Ruhe, R. V Elements of the Soil Landscape. Trans. 7th int. Congr. Soil Sci. 1960, 4, 165–170.

85. Shreve, R.L. Infinite Topologically Random Channel Networks. J. Geol. 1967, 75, 178–186.

86. Strahler, A.N. Dynamics Basis of Geomorphology. Bull. Geol. Soc. Am. 1952, 63, 923–938.

87. Souza, C.M.; Z. Shimbo, J.; Rosa, M.R.; Parente, L.L.; A. Alencar, A.; Rudorff, B.F.T.; Hasenack, H.; Matsumoto, M.; G. Ferreira, L.; Souza-Filho, P.W.M.; et al. Reconstructing Three Decades of Land Use and Land Cover Changes in Brazilian Biomes with Landsat Archive and Earth Engine. Remote Sens. 2020, 12.

88. Florinsky, I. V. Digital Terrain Analysis in Soil Science and Geology; Academic Press, 2012; ISBN 978-0-12-385036-2.

89. Chabrillat, S.; Ben-Dor, E.; Cierniewski, J.; Gomez, C.; Schmid, T.; van Wesemael, B. Imaging Spectroscopy for Soil Mapping and Monitoring. Surv. Geophys. 2019, 40, 361–399, doi:10.1007/s10712-019-09524-0.

90. Sherman, D.M.; Waite, T.D. Electronic Spectra of Fe3+ Oxides and Oxide Hydroxides in the near IR to near UV. Am. Mineral. 1985, 70, 1262–1269.

91. Rossel, R.A.V.; Behrens, T. Using Data Mining to Model and Interpret Soil Diffuse Reflectance Spectra. Geoderma 2010, 158, 46–54, doi:https://doi.org/10.1016/j.geoderma.2009.12.025.

92. IBGE, C. de R.N. e; Ambientais, E. Províncias Estruturais, Compartimentos de Relevo, Tipos de Solos, Regiões Fitoecológicas e Outras Áreas; IBGE.; IBGE: Rio de Janeiro, 2019; ISBN 978-85-240-4487-8.

93. Mello, F.A.O.; Bellinaso, H.; Mello, D.C.; Safanelli, J.L.; Mendes, W.D.S.; Amorim, M.T.A.; Gomez, A.M.R.; Poppiel, R.R.; Silvero, N.E.Q.; Gholizadeh, A.; et al. Soil Parent Material Prediction through Satellite Multispectral Analysis on a Regional Scale at the Western Paulista Plateau, Brazil. Geoderma Reg. 2021, e00412, doi:https://doi.org/10.1016/j.geodrs.2021.e00412.

94. Mendes, W. de S.; Demattê, J.A.M.; de Resende, M.E.B.; Chimelo Ruiz, L.F.; César de Mello, D.; Fim Rosas, J.T.; Quiñonez Silvero, N.E.; Ferracciú Alleoni, L.R.; Colzato, M.; Rosin, N.A.; et al. A Remote Sensing Framework to Map Potential Toxic Elements in Agricultural Soils in the Humid Tropics. Environ. Pollut. 2022, 292, 118397, doi:https://doi.org/10.1016/j.envpol.2021.118397.

95. Naimi, S.; Ayoubi, S.; Demattê, J.A.M.; Zeraatpisheh, M.; Amorim, M.T.A.; de Oliveira Mello, F.A. Spatial Prediction of Soil Surface Properties in an Arid Region Using Synthetic Soil Image and Machine Learning. Geocarto Int. 2021, 1–22, doi:10.1080/10106049.2021.1996639.

96. Silvero, N.E.Q.; Demattê, J.A.M.; Vieira, J. de S.; Mello, F.A. de O.; Amorim, M.T.A.; Poppiel, R.R.; Mendes, W. de S.; Bonfatti, B.R. Soil Property Maps with Satellite Images at Multiple Scales and Its Impact on Management and Classification. Geoderma 2021, 397, 115089, doi:https://doi.org/10.1016/j.geoderma.2021.115089.

97. Strahler, A.N. Quantitative Analysis of Watershed Geomorphology. Eos, Trans. Am. Geophys. Union 1957, 38, 913–920.

98. Ahmadi, S.H.; Sedghamiz, A. Geostatistical Analysis of Spatial and Temporal Variations of Groundwater Level. Environ. Monit. Assess. 2007, 129, 277–294, doi:10.1007/s10661-006-9361-z.

99. Jan, C.-D.; Chen, T.-H.; Lo, W.-C. Effect of Rainfall Intensity and Distribution on Groundwater Level Fluctuations. J. Hydrol. 2007, 332, 348–360, doi:https://doi.org/10.1016/j.jhydrol.2006.07.010.

100. Haubrock, S. -N.; Chabrillat, S.; Lemmnitz, C.; Kaufmann, H. Surface Soil Moisture Quantification Models from Reflectance Data under Field Conditions. Int. J. Remote Sens. 2008, 29, 3–29, doi:10.1080/01431160701294695.

101. Mendes, W.D.S.; Medeiros Neto, L.G.; Demattê, J.A.M.; Gallo, B.C.; Rizzo, R.; Safanelli, J.L.; Fongaro, C.T. Is It Possible to Map Subsurface Soil Attributes by Satellite Spectral Transfer Models? Geoderma 2019, 343, 269–279, doi:10.1016/j.geoderma.2019.01.025.

102. 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. Multi-Temporal Satellite Images on Topsoil Attribute Quantification and the Relationship with Soil Classes and Geology. Remote Sens. 2018, 10, doi:10.3390/rs10101571.

103. Bonfatti, B.R.; Demattê, J.A.M.; Marques, K.P.P.; Poppiel, R.R.; Rizzo, R.; Mendes, W. de S.; Silvero, N.E.Q.; Safanelli, J.L. Digital Mapping of Soil Parent Material in a Heterogeneous Tropical Area. Geomorphology 2020, 367, 107305, doi:https://doi.org/10.1016/j.geomorph.2020.107305.

104. van Breemen, N.; Buurman, P. Soil Formation - Second Edition; 2002; ISBN 9789401741408.

105. Rapinel, S.; Fabre, E.; Dufour, S.; Arvor, D.; Mony, C.; Hubert-Moy, L. Mapping Potential, Existing and Efficient Wetlands Using Free Remote Sensing Data. J. Environ. Manage. 2019, 247, 829–839, doi:https://doi.org/10.1016/j.jenvman.2019.06.098.

106. Whyte, A.; Ferentinos, K.P.; Petropoulos, G.P. A New Synergistic Approach for Monitoring Wetlands Using Sentinels -1 and 2 Data with Object-Based Machine Learning Algorithms. Environ. Model. Softw. 2018, 104, 40–54, doi:https://doi.org/10.1016/j.envsoft.2018.01.023.

107. Chaplot, V.; Walter, C.; Curmi, P. Improving Soil Hydromorphy Prediction According to DEM Resolution and Available Pedological Data. Geoderma 2000, 97, 405–422, doi:https://doi.org/10.1016/S0016-7061(00)00048-3.

108. Chaplot, V.; Walter, C.; Curmi, P. Testing Quantitative Soil-Landscape Models for Predicting the Soil Hydromorphic Index at a Regional Scale. Soil Sci. 2003, 168.

109. Pennock, D.; Bedard-Haughn, A.; Kiss, J.; van der Kamp, G. Application of Hydropedology to Predictive Mapping of Wetland Soils in the Canadian Prairie Pothole Region. Geoderma 2014, 235–236, 199–211, doi:https://doi.org/10.1016/j.geoderma.2014.07.008.

110. Thompson, J.A.; Bell, J.C.; Butler, C.A. Quantitative Soil-Landscape Modeling for Estimating the Areal Extent of Hydromorphic Soils. Soil Sci. Soc. Am. J. 1997, 61, 971–980, doi:https://doi.org/10.2136/sssaj1997.03615995006100030037x.

111. Mello, F.A.O.; Demattê, J.A.M.; Rizzo, R.; Mello, D.C. de; Poppiel, R.R.; Silvero, N.E.Q.; Safanelli, J.L.; Bellinaso, H.; Bonfatti, B.R.; Gomez, A.M.R.; et al. Complex Hydrological Knowledge to Support Digital Soil Mapping. Geoderma 2022, 409, 115638, doi:https://doi.org/10.1016/j.geoderma.2021.115638.

112. Lee, S.; Yeo, I.-Y.; Lang, M.W.; McCarty, G.W.; Sadeghi, A.M.; Sharifi, A.; Jin, H.; Liu, Y. Improving the Catchment Scale Wetland Modeling Using Remotely Sensed Data. Environ. Model. Softw. 2019, 122, 104069, doi:https://doi.org/10.1016/j.envsoft.2017.11.001.

113. Schaetzl, J Randall and Anderson, S. Soil Genesis and Geomorphology; Cambridge University Press: New York, 2005; ISBN 9780521812016.

114. Miller, B.A.; Juilleret, J. The Colluvium and Alluvium Problem: Historical Review and Current State of Definitions. Earth-Science Rev. 2020, 103316, doi:https://doi.org/10.1016/j.earscirev.2020.103316.

115. Curmi, P.; Durand, P.; Gascuel-Odoux, C.; Mérot, P.; Walter, C.; Taha, A. Hydromorphic Soils, Hydrology and Water Quality: Spatial Distribution and Functional Modelling at Different Scales BT - Soil and Water Quality at Different Scales: Proceedings of the Workshop "Soil and Water Quality at Different Scales" Held 7–9 August 1. In; Finke, P.A., Bouma, J., Hoosbeek, M.R., Eds.; Springer Netherlands: Dordrecht, 1998; pp. 127–142 ISBN 978-94-017-3021-1.

116. Junk, W.J. Long-Term Environmental Trends and the Future of Tropical Wetlands. Environ. Conserv. 2002, 29, 414–435, doi:DOI: 10.1017/S0376892902000310.

117. Maltby, E. Wetland Management Goals: Wise Use and Conservation. Landsc. Urban Plan. 1991, 20, 9–18, doi:https://doi.org/10.1016/0169-2046(91)90085-Z.

118. Magha, A.M.; Azinwi Tamfuh, P.; Mamdem, L.E.; Shey Yefon, M.C.; Kenzong, B.; Bitom, D. Soil Water Characteristics of Gleysols in the Bamenda (Cameroon) Wetlands and Implications for Agricultural Management Strategies. Appl. Environ. Soil Sci. 2021, 2021, 6643208, doi:10.1155/2021/6643208.

119. Ahrens, S. O Novo Código Florestal Brasileiro: Conceitos Jurídicos Fundamentais. In Proceedings of the Embrapa Florestas-Artigo em anais de congresso (ALICE); In: CONGRESSO FLORESTAL BRASILEIRO, 8., 2003, São Paulo. Benefícios …, 2003.

120. Vlasenko, V.P. Development of Hydromorphism in Soils of Agrolandscapes in the Depressions of Western Ciscaucasia. Eurasian Soil Sci. 2009, 42, 488–495, doi:10.1134/S1064229309050032.

121. Shaheen, S.M.; Wang, J.; Baumann, K.; Ahmed, A.A.; Hsu, L.-C.; Liu, Y.-T.; Wang, S.-L.; Kühn, O.; Leinweber, P.; Rinklebe, J. Stepwise Redox Changes Alter the Speciation and Mobilization of Phosphorus in Hydromorphic Soils. Chemosphere 2022, 288, 132652, doi:https://doi.org/10.1016/j.chemosphere.2021.132652.

122. Giannetta, B.; Oliveira de Souza, D.; Aquilanti, G.; Celi, L.; Said-Pullicino, D. Redox-Driven Changes in Organic C Stabilization and Fe Mineral Transformations in Temperate Hydromorphic Soils. Geoderma 2022, 406, 115532, doi:https://doi.org/10.1016/j.geoderma.2021.115532.

123. Likhanova, I.A.; Deneva, S. V; Kholopov, Y. V; Kuznetsova, E.G.; Shakhtarova, O. V; Lapteva, E.M. The Effect of Hydromorphism on Soils and Soil Organic Matter during the Primary Succession Processes of Forest Vegetation on Ancient Alluvial Sands of the European North-East of Russia. For. 2022, 13.

124. da Silveira, L.J.; Barbosa, R.A.; Spletozer, A.G.; Dias, H.C.T. Soil Characteristics in Veredas (Brazilian Palm Swamps) of Alto São Francisco, Minas Gerais, Brazil. In OPEN SCIENCE RESEARCH I; 2022; pp. 216–228 ISBN 978-65-5360-055-3.

125. IBGE, I.B. de G. e E. Atlas Do Espaço Rural Brasileiro; IBGE, 2011; ISBN 8524042281.

126. Bordonal, R. de O.; Carvalho, J.L.N.; Lal, R.; de Figueiredo, E.B.; de Oliveira, B.G.; La Scala, N. Sustainability of Sugarcane Production in Brazil. A Review. Agron. Sustain. Dev. 2018, 38, 13, doi:10.1007/s13593-018-0490-x.

127. Cherubin, M.R.; Karlen, D.L.; Franco, A.L.C.; Cerri, C.E.P.; Tormena, C.A.; Cerri, C.C. A Soil Management Assessment Framework (SMAF) Evaluation of Brazilian Sugarcane Expansion on Soil Quality. Soil Sci. Soc. Am. J. 2016, 80, 215–226, doi:https://doi.org/10.2136/sssaj2015.09.0328.

128. Oliveira, D.M. da S.; Paustian, K.; Davies, C.A.; Cherubin, M.R.; Franco, A.L.C.; Cerri, C.C.; Cerri, C.E.P. Soil Carbon Changes in Areas Undergoing Expansion of Sugarcane into Pastures in South-Central Brazil. Agric. Ecosyst. Environ. 2016, 228, 38–48, doi:https://doi.org/10.1016/j.agee.2016.05.005.

129. Bordonal, R. de O.; Lal, R.; Ronquim, C.C.; de Figueiredo, E.B.; Carvalho, J.L.N.; Maldonado, W.; Milori, D.M.B.P.; La Scala, N. Changes in Quantity and Quality of Soil Carbon Due to the Land-Use Conversion to Sugarcane (Saccharum Officinarum) Plantation in Southern Brazil. Agric. Ecosyst. Environ. 2017, 240, 54–65, doi:https://doi.org/10.1016/j.agee.2017.02.016.

130. Bonato, E.R.; Dall'Agnol, A. Soybean in Brazil-Production and Research. In Proceedings of the World Soybean Research Conference III: Proceedings; CRC Press, 2022; pp. 1248–1256.

131. A., A.A.; Raoni, R.; A., C.M.; C., S.M.C.; N., M.M.; P., dos R.T.N.; Ane, A.; S., S.-F.B.; Rayane, P. Limits of Brazil's Forest Code as a Means to End Illegal Deforestation. Proc. Natl. Acad. Sci. 2017, 114, 7653–7658, doi:10.1073/pnas.1604768114.

132. Stickler, C.M.; Nepstad, D.C.; Azevedo, A.A.; McGrath, D.G. Defending Public Interests in Private Lands: Compliance, Costs and Potential Environmental Consequences of the Brazilian Forest Code in Mato Grosso. Philos. Trans. R. Soc. B Biol. Sci. 2013, 368, 20120160, doi:10.1098/rstb.2012.0160.

133. da Silva, N.M.; Angeoletto, F.; Santos, jeater W.M.C.; Paranhos Filho, A.C.; Vacchiano, M.C.; Bohrer, J.F.C.; Candido, A.K.A.A. The Negative Influences of the New Brazilian Forest Code on the Conservation of Riparian Forests. Eur. J. Ecol. 2017, 3, 116–122, doi:10.1515/eje-2017-0019.

134. Goldshleger, N.; Ben-Dor, E.; Lugassi, R.; Eshel, G. Soil Degradation Monitoring by Remote Sensing: Examples with Three Degradation Processes. Soil Sci. Soc. Am. J. 2010, 74, 1433–1445, doi:https://doi.org/10.2136/sssaj2009.0351.

135. Žížala, D.; Juřicová, A.; Zádorová, T.; Zelenková, K.; Minařík, R. Mapping Soil Degradation Using Remote Sensing Data and Ancillary Data: South-East Moravia, Czech Republic. Eur. J. Remote Sens. 2019, 52, 108–122, doi:10.1080/22797254.2018.1482524.

136. Shoshany, M.; Goldshleger, N.; Chudnovsky, A. Monitoring of Agricultural Soil Degradation by Remote-Sensing Methods: A Review. Int. J. Remote Sens. 2013, 34, 6152–6181, doi:10.1080/01431161.2013.793872.

5. GENERAL DISCUSSIONS AND CONCLUSIONS

After exploring the multiple possibilities of DSM and applying its techniques for multiple purposes, a few general conclusions were reached regarding the three chapters presented in this thesis, as follows: a. DSM do not intend to eliminate traditional soil knowledge and methods of soil survey, but improve soil mapping by reducing cost and time, decrease the dependence of qualitative interpretations and use quantitative analysis, and by applying large scale geospatial data to make predictions of soil attributes or classes; b. although DSM deals with soil spatial variability using quantitative and nummerical analysis, the interpretation of the results should consider a conventional soil knowledge, following the soil-landscape relationship sules; c. DSM is a fundamental tool to improve the scale and quality of soil inventories, make more soil data available for the public, and slower soil degradation, which are determinant for food security.

Furthermore, the three chapters regarding this research produced other specific conclusions, listed as follows: for the second chapter, a. drainage network information had great importance on soil predictions; b. multitemporal satellite bare images, established in this work as a synthetic soil image (SySI) was a powerful tool for the prediction of clay, sand, and soil organic carbon contents; and c. despite the importance of traditionally used terrain variables (i.e., slope, aspect, and others), the confluence angle and channel sinuosity presented greater significance on the attribute's quantification. The final conclusion is that more studies regarding the relationship between channel sinuosity and soils should be performed. For a tropical region, with complex lithological and pedological characteristics, random forest and Cubist models performed similarly, producing good results for soil attribute mapping. The conclusions for the third chapter were as follows: a. the surface reflectance was indicative of soil parent material due to their relationship with soil formation and attributes; b. the textural and mineralogical aspects of soils tended to differ from mafic and sedimentary rocks, affecting directly the reflectance; c. satellite images worked as proxies to infer the parent material, since they retrieve surface reflectance. The final conclusions were that in areas where geological maps have low scale, the automated method using environmental variables offers a great advance in saving surveying time and costs, due to its capability to accurately identification of the transitions. The prediction of parent material classes by bare soil reflectance (SySI), terrain attributes and soil data proved to be efficient and accurate. The fourth chapter reached the following conclusions: a. the observed features in SySI were confirmed as hydromorphic soils through Vis-NIR-SWIR laboratory spectral analysis, allowing its use for land classification and mapping; b. the relief covariates regarding curvature and slope were important to improve the prediction performance along with SySI to classify hydromorphic soils; c. the catchment and accumulation areas were underestimated by land owners at multiple locations, resulting in unprotected channel network and water sources. Finally, SySI proved to be a robust tool on the identification and mapping of hydromorphic soils for agricultural areas, enhancing the possibilities to use remote sensing information to improve the monitoring of agricultural areas.