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

Sensing technologies for topsoil mapping and soil functional assessment

Nélida Elizabet Quiñonez Silvero

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

Piracicaba
2022

Nélida Elizabet Quiñonez Silvero
Agronomist

Sensing technologies for topsoil mapping and soil functional assessment
versão revisada de acordo com a resolução CoPGr 6018 de 2011

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

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

Piracicaba
2022

To my parents,

Patricio and Yolanda,

for their unconditional **support during this academic expedition.**

ACKNOWLEDGMENTS

CONTENTS

8

RESUMO

Tecnologias de sensoriamento para mapeamento do solo superficial e avaliação das funções do solo

Esta tese de doutorado trata do uso de tecnologias de sensoriamento remoto e proximal para o mapeamento de propriedades do solo e das possibilidades de uso destes produtos no estudo das funções que ele é capaz de oferecer. O mapeamento digital tem tomado força desde os anos 90's quando foram propostos os primeiros conceitos de pedometria e mapeamento preditivo do solo. Atualmente, existe uma demanda crescente por mapas de solos e as tecnologias de detecção estão desempenhando um papel fundamental para que isto seja possível. Considerando este cenário, o objetivo desta tese é introduzir exemplos de mapeamento de atributos do solo utilizando imagens de solo exposto obtidas a partir de séries temporais de satélites e de como usar estes dados em exemplos de mapeamento de outros atributos do solo relacionadas às suas funções. No Capítulo 1 se expõe uma introdução geral do trabalho e se abordam as principais problemáticas que se pretendem resolver com esta tese. No Capítulo 2 se exploram as propriedades de imagens de satélite em diferentes escalas e sua influência na obtenção de mapas de propriedades de solo no nível de fazenda. Foi observado que, de fato, os mapas produzidos com imagens de tamanho de pixel diferente proporcionam mapas diferentes. O impacto destes mapas em diferentes escalas foi avaliado tanto para fins de classificação como para o manejo do solo e foi observado que o delineamento de unidades de mapeamento pode ser afetado pela qualidade dos mapas. A mesma situação foi descrita para o caso do manejo do solo, onde pode haver uma inconsistência na distribuição espacial de propriedades do solo específicas o que pode levar a uma estratégia de manejo diferente dependendo de qual tipo de mapa e em qual escala é utilizado. O capítulo 3 propõe a junção de dados de dois satélites para a obtenção de maiores áreas de solo exposto que possam possibilitar o melhor estudo da variação do solo, similar ao capítulo 2, porém no nível regional abordando a variação do solo desde um ponto de vista macro. Neste capítulo, foi observado que os avanços em termos de maior disponibilidade de imagens de satélite ao longo do tempo proporcionam um melhor entendimento das variações do solo. Já no capítulo 4 se apresentam estratégias para o mapeamento de classes de drenagem em regiões tropicais. Este capítulo foi desenvolvido com o intuito de avaliar a capacidade dos mapas de propriedades do solo para estudar características que são mais complexas e difíceis de estudar. Finalmente, o capítulo 5 proporciona uma revisão das possibilidades dos sensores remotos e proximais para a medição e estudo das funções do solo. Aqui se descrevem desde sensores de laboratório até satélites que cobrem toda a região do espectro eletromagnético e de como estes podem ser utilizados para estudar os solos. No tópico final se apresentam três exemplos de aplicação no Brasil e se conclui esta tese discutindo as principais vantagens, limitações e o que ainda precisa ser feito para avançar no estudo do recurso solo fazendo uso das tecnologias disponíveis.

Palavras-chave: Mapeamento digital de solos, Sensoriamento remoto, Espectroscopia, Sensoriamento proximal, Aprendizado de máquina, Resolução espacial, Imagens multi-temporais, Funções do solo, Escala

ABSTRACT

Sensing technologies for topsoil mapping and soil functional assessment

This doctoral dissertation deals with the use of proximal and remote sensing technologies for mapping soil properties and the possibilities of using these products in the study of the functions that the soil is capable to offer. Digital soil mapping has gained strength since the 90's, when the first concepts of pedometrics and predictive soil mapping were proposed. Currently, there is a growing demand for soil maps and sensing technologies are playing a key role in making this possible. Considering this scenario, the objective of this dissertation is to introduce examples of soil attribute mapping using bare soil images obtained from satellite time series and how to use these data in mapping examples of other characteristics related to soil functions. In Chapter 1, a general introduction to the work is presented and addresses the main problems that are intended to be resolved with this thesis are addressed. Chapter 2 explores the properties of satellite imagery at different scales and their influence on obtaining maps of soil properties at the farm level. It was observed that, in fact, maps produced with images of different pixel sizes were different. The impact of these maps at different scales was evaluated both for soil classification purposes and for soil management and it was observed that the delineation of mapping units can be affected by the quality of soil property maps. The same situation was described for the case of soil management, where there may be an inconsistency in the spatial distribution of specific soil properties which can lead to a different management strategy depending on which type of map and at which scale it is used. Chapter 3 proposes the joining of data from two satellites to obtain larger areas of exposed soil that can enable a better study of soil variations, similar to chapter 2, but at the regional level, approaching soil variation from a different point of view. In this chapter, it was observed that advances in terms of greater availability of satellite images over time provide a better understanding of soil variations. Chapter 4 presents strategies for mapping drainage classes in tropical regions. This chapter was developed to assess the ability of soil property maps to study features that are more complex and difficult to measure. Finally, Chapter 5 provides a review of the possibilities of remote and proximal sensors for measuring and studying soil functions. Here we described sensors from the laboratory to satellites that cover the entire electromagnetic spectrum and how they can be used to study soils. In the final topic, three examples of application in Brazil are presented and this thesis is concluded by discussing the main advantages, limitations and what still needs to be done to advance in the study of the soil resource using the available technologies.

Keywords: Digital soil mapping, Remote sensing, Spectroscopy, Proximal sensing, Machine learning, Spatial resolution, Multi-temporal images, Soil functions, Scale

1. GENERAL INTRODUCTION

Soil is an extraordinarily complex entity (Scull et al., 2003). Its importance lies not only in agriculture as a means of plant growth and food production, but also in the performance of six other functions that are related to the good functioning of ecosystems and human well-being (Blum, 2005). Among these functions, soil plays an important role in water cleaning, filtering and buffering, in addition to being the habitat of an immense biodiversity of organisms (Keesstra et al., 2012; Makó et al., 2017; Bölter and Blume, 2002). Soils are also considered an archaeological and geological heritage reserve, which has helped to reveal important aspects of our recent history (Cummins et al., 2018). The most recent recognized role of soils is in mitigating climate change, by reducing CO_2 emissions to the atmosphere and sequestering carbon as a long-term component (Minasny et al., 2017).

Understanding and measuring these functions are essential to improve land use and management, assist in decision-making and in the construction of public policies related to it. But for that, it is first necessary to know the characteristics and variability of soils, both in space and possible changes that can happen over time. One of the most common ways of communicating this is through soil maps or data on soil's physical, chemical and mineralogical properties. The need for this kind of soil information is increasing considerably all over the world (Poggio et al., 2021) and even more in developing countries, where maps at suitable scales for land planning are scarce. This gap is being filled thanks to the widespread use of frameworks such as digital soil mapping (Scull et al., 2003; McBratney et al., 2003) and the widely availability of powerful computational tools, geographic information systems (GIS), remote and proximal sensors, cloud computing and data storage capacities. With these advances, the management of big soil data and complex statistics to model soil variability becomes possible.

Soil data in the form of maps or point information are useful for land planning but also as inputs in ecological and hydrological models, which are used to understand crucial processes happening in the soil system (Veerecken et al., 2016). Accurate soil maps are needed to best represent the soil variability and for this, huge amounts of data are needed, which are in part fulfilled by field descriptions available in legacy soil data or, more recently, by data obtained from technologies such as soil spectroscopy, remote and proximal sensing. Sensing technologies, which includes remote and proximal sensing, are playing a crucial role in providing fast and accurate information to assist in digital soil mapping. The advantage of using satellites is their capability of imaging large areas and provide a spatial overview of possible changes in landscapes at short or longer distances. They also provide continuous data at different temporal resolution, which is allowing the monitoring of possible short-term changes that can affect the spatial variability of soils (Demattê et al., 2020).

In this dissertation, we are concerned in addressing three main issues: a) digital soil mapping of soil properties at the farm and b) regional levels, c) search for strategies to map soil drainage classes using available soil data, expert knowledge and machine learning, and d) explore the possibilities of sensing technologies for studying and measuring soil functions. The first is addressed in Chapter 2, in which soil property maps obtained at multiple scales from different satellites are presented. In this study we evaluated the influence of spatial, spectral and temporal resolutions of three satellites on mapping soil properties (clay content, soil color and iron and organic matter content) in a small agricultural area in southeastern Brazil. We further evaluated the influence of the results on soil management and classification. In Chapter 3, a similar approach was followed but, in this case, the soil variability was studied using two satellites, Sentinel-2 and Landsat-8. We aimed to assess the potential of multi-temporal bare soil images (Synthetic Soil Image, SYSI) obtained from Landsat-8 and Sentinel-2 satellites (singly and combined) to predict topsoil properties (clay, sand, silt, organic matter, and color) using the Cubist algorithm and to compare their performances with those obtained from single-date images. Our hypothesis was that SYSI provides more continuous

soil reflectance values than single-date images and that the combined use of Landsat-8 and Sentinel-2 satellites can offer even more bare soil pixels, which can improve soil mapping in tropical regions.

Chapter 4 is a mixture of digital soil mapping and applicability of available soil maps. Here we wanted to study two strategies for mapping soil drainage classes using expert knowledge and machine learning. The best strategy was selected considering a set of criteria that represented the most accurate, simplest, transparent and transferable map of soil drainage class. Finally, in Chapter 5, we aimed at shedding light on the use of sensing technologies to assess and monitor the variability of soil properties and their use in the evaluation of soil functions by reviewing the most important aspects. In this chapter we provided a brief description of soil spectroscopy and proximal and remote sensing and their applicability in obtaining information concerning soil properties. Besides, a description of soil functions and their properties of major importance in the generation of indicators was given. The review ended with a discussion of the strengths, weaknesses and possible future scenarios of sensing technologies associated with obtaining soil information that contributes to a better understanding of soil functions.

This dissertation ends with a general discussion and a conclusion of the main findings.

References

Bölter, M., Blume, H.-P., 2002. Soils as Habitats for Microorganisms. Ecol. Stud. 154, 285–302. https://doi.org/10.1007/978-3-642-56318-8_16

Blum, W.E.H., 2005. Functions of soil for society and the environment. Rev. Environ. Sci. Biotechnol. 4, 75–79. https://doi.org/10.1007/s11157-005-2236-x

Cummins, T., Lewis, H., Ní Lionáin, C., Davis, S., 2018. Soils and Archaeology 267–280. https://doi.org/10.1007/978-3-319-71189-8_19

Demattê, J.A.M., Safanelli, J.L., Poppiel, R.R., Rizzo, R., Silvero, N.E.Q., de Mendes, W.S., Bonfatti, B.R., Dotto, A.C., Salazar, D.F.U., de Mello, F.A.O., da Paiva, A.F.S., Souza, A.B., dos Santos, N.V., Maria Nascimento, C., de Mello, D.C., Bellinaso, H., Gonzaga Neto, L., Amorim, M.T.A., de Resende, M.E.B., da Vieira, J.S., de Queiroz, L.G., Gallo, B.C., Sayão, V.M., da Lisboa, C.J.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.

Keesstra, S.D., Geissen, V., Mosse, K., Piiranen, S., Scudiero, E., Leistra, M., van Schaik, L., 2012. Soil as a filter for groundwater quality. Curr. Opin. Environ. Sustain. 4, 507–516. https://doi.org/10.1016/j.cosust.2012.10.007

Makó, A., Kocsis, M., Barna, G.Y., Tóth, G., 2017. Mapping the storing and filtering capacity of European soils - Publications Office of the EU. Luxembourg.

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

Minasny, B., Malone, B.P., McBratney, A.B., Angers, D.A., Arrouays, D., Chambers, A., Chaplot, V., Chen, Z.S., Cheng, K., Das, B.S., Field, D.J., Gimona, A., Hedley, C.B., Hong, S.Y., Mandal, B., Marchant, B.P., Martin, M., McConkey, B.G., Mulder, V.L., O'Rourke, S., Richer-de-Forges, A.C., Odeh, I., Padarian, J., Paustian, K., Pan, G., Poggio, L., Savin, I., Stolbovoy, V., Stockmann, U., Sulaeman, Y., Tsui, C.C., Vågen, T.G., van Wesemael, B., Winowiecki, L., 2017. Soil carbon 4 per mille. Geoderma 292, 59–86. https://doi.org/10.1016/J.GEODERMA.2017.01.002

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

Scull, P., Franklin, J., Chadwick, O.A., Mcarthur, D., 2003. Predictive soil mapping: a review. https://doi.org/10.1191/0309133303pp366ra

Vereecken, H., Schnepf, A., Hopmans, J.W., Javaux, M., Or, D., Roose, T., Vanderborght, J., Young, M.H., Amelung, W., Aitkenhead, M., Allison, S.D., Assouline, S., Baveye, P., Berli, M., Brüggemann, N., Finke, P., Flury, M., Gaiser, T., Govers, G., Ghezzehei, T., Hallett, P., Hendricks Franssen, H.J., Heppell, J., Horn, R., Huisman, J.A., Jacques, D., Jonard, F., Kollet, S., Lafolie, F., Lamorski, K., Leitner, D., McBratney, A., Minasny, B., Montzka, C., Nowak, W., Pachepsky, Y., Padarian, J., Romano, N., Roth, K., Rothfuss, Y., Rowe, E.C., Schwen, A., Šimůnek, J., Tiktak, A., Van Dam, J., van der Zee, S.E.A.T.M., Vogel, H.J., Vrugt, J.A., Wöhling, T., Young, I.M., 2016. Modeling Soil Processes: Review, Key Challenges, and New Perspectives. Vadose Zo. J. 15, vzj2015.09.0131. https://doi.org/10.2136/vzj2015.09.0131

2. SOIL PROPERTY MAPS WITH SATELLITE IMAGES AT MULTIPLE SCALES AND THEIR IMPACT ON SOIL MANAGEMENT AND CLASSIFICATION

Abstract

Soil maps at appropriate scales can aid decision-making in agriculture and the environment. In this context, remote sensing products have shown their power to investigate soil properties, but the assessment of its spatial information can be hampered by the presence of other objects than soil. To overcome this issue, soil scientists have been studying spatial patterns from multi-temporal satellite images aiming at improving soil property maps. In this work, we applied the cubist algorithm to predict topsoil properties (clay, sand, organic matter and iron contents, and soil color components) in south-eastern Brazil using multi-temporal (Landsat8-OLI, and Sentinel2-MSI) and single-date images (PlanetScope, Landsat8-OLI, and Sentinel2-MSI). We aimed to evaluate the influence of satellite's spatial, spectral and temporal resolutions on soil mapping. Predictive models were constructed with 120 soil samples and using four (vis-NIR) and six (vis-NIR-SWIR) spectral bands as predictors in a 10-fold cross-validation procedure. The multi-temporal image obtained from the Sentinel2-MSI satellite (with 10 m pixel size and six spectral bands), showed the best model performances in cross-validation (R^2 between 0.48 and 0.78). The PlanetScope image, which has only four bands in the vis-NIR region and spatial resolution of 3 m did not improve model performances, although the R^2 values were higher for soil color components ($R^2 > 0.5$). Satellite images in different spatial, spectral and temporal resolutions provide slightly different soil property maps which may promote different strategies regarding soil classification and management. However, satellite images should be used with caution, as they provide only surficial information about the soil variability and confirmation with field surveys is required.

2.1. Introduction

The growing demand for spatial information has encouraged the application of Digital Soil Mapping (DSM) strategies to obtain reliable soil maps (McBratney et al., 2003; Smith et al., 2010; Malone et al., 2011; Gholizadeh et al., 2018; Chabrillat et al., 2019). Maps of soil properties, such as for clay, sand and organic matter content, are of great value for guiding agricultural and environmental practices, such as the management zones, water storage, and drainage classes (Reyes et al., 2019; Hong et al., 2013; Zhao et al., 2008). In this sense, satellite images provide a synoptic view of the area under study, which have been of significant importance for obtaining accurate maps of soil properties.

Soil properties vary with scale (Nussbaum et al., 2011) and soil maps produced with satellite images with different spatial, spectral and temporal resolutions would be different as well. The spatial resolution is associated with the pixel size, characterizing high- and coarse-resolution images, with small and large pixels, respectively. Coarse-resolution images are predominantly used to represent the phenomena from a general point of view (Hengl, 2006) whereas high-resolution images can provide better details about terrestrial objects and are more useful in farm-scale studies (Gomez et al., 2015). Satellite images can also present different spectral resolutions. This is defined as the number of ranges (or bands) detected in the electromagnetic spectrum and is related with the spectral signature and soil compounds (Gholizadeh et al., 2018; Marais Sicre et al., 2020). Besides these, we also have the temporal resolution related with the frequency of revisit over a site (Ose et al., 2016). This characteristic plays an important role in soil mapping. The shorter the revisit time, the greater is the number of images that can be obtained and,

therefore, the soil will be better monitored over time (Croft et al., 2012). Thus, the discussion about these three components of an image has gained great importance to understand soil variability.

The influence of satellite resolutions on soil mapping has been reported in many studies. Some of them confirmed that images with high spectral resolution provided the best performances in soil mapping (Gomez et al., 2018; Gholizadeh et al., 2018) but others (Behrens et al., 2018; Forkuor et al., 2017; Sahwan et al., 2018) highlighted the importance of spatial resolution, as soils can behave differently at multiple scales. Samuel-Rosa et al. (2015) stated that the improvement in model performances using fine-scale covariates was modest, not recommending spending resources in obtaining covariates or predictors with high spatial resolution that poorly explain soil variability.

Satellites' resolution is not the only issue to be considered in soil mapping. Predictions of soil properties and further interpretations for soil management can also be hampered by the presence of vegetation, crop residues and clouds (Roberts et al., 2019; Castaldi et al., 2019). Demattˆe et al. (2018), for example, found that depending on the study area, a single satellite image can provide only 0.5% of valid bare soil pixels, which certainly would not bring suficient information for soil mapping. Castaldi et al. (2019) used the Normalized Burn Ratio 2 (NBR2) to exclude moist and vegetated pixels in a single Sentinel-2 image from August 2017 and obtained about 25% of valid bare soil pixels for soil organic carbon (SOC) mapping. In another research, Bartholomeus et al. (2011) found that SOC predictions drop dramatically (~58%) when the soil surface is partially vegetated. These researches exemplify the difficulties on understanding soils by images (influence of clouds, moisture and vegetated pixels).

To overcome the indicated issues, recent studies have been developing multi-temporal techniques to obtain more continuous bare soil pixels from satellite images to improve digital soil mapping. These methodologies were developed using Landsat missions (Diek et al., 2017; Demattˆe et al., 2018, 2020; Rogge et al., 2018) and Sentinel-2 images (Loiseau et al., 2019; Castaldi et al., 2019; Gholizadeh et al., 2018; Tziolas et al., 2020; Silvero et al., 2021). Successful results have been reported with these approaches at local and regional scales (Fongaro et al., 2018; Safanelli et al., 2020; Poppiel et al., 2020), which consolidates the use of multi-temporal images in several DSM applications.

The strategy to use images (which image, which satellite, which characteristics) on soil mapping will have direct impacts on decisions making related to soil management and, consequently, in agriculture and the environment (Casa et al., 2012; Demattˆe et al., 2016; Buttafuoco et al., 2017). Besides that, satellite images are also important tools to enhance soil class maps and classification (Ramcharan et al., 2018). The spatial coverage and spectral patterns can bring insights about soil variability, which is later linked with its classification by field survey. Despite the indicated importance of remote sensing for agriculture, we found only one research that explored the implications of different satellite resolutions for soil management. Breunig et al. (2020) used UAV cameras (0.20 m), PlanetScope (3 m) and Landsat-8 images (30 m) to delineate management zones for smallholder farming in Brazil. These authors found that PlanetScope image was better than UAV cameras in delineating management zones and Landsat-8 data was unable to provide suitable results, as the area under study was very small. This highlight (once more) the importance of scale and satellite's resolution to assist decision-making in soil management.

Multi-temporal images showed their usefulness for digital soil mapping but the applications for other soil approaches are still lacking. Recently, Silvero et al. (2021) compared single-date and multi-temporal images from two satellites (Landsat-8 and Sentinel-2) in predicting topsoil properties for regional soil mapping, but the authors did not evaluate the implications for soil management and classification. Besides that, at the farm level the above-mentioned satellites can provide very different results. Thus, in this study we evaluated the influence of spatial, spectral and temporal resolutions of satellites (PlanetScope, Sentinel2-MSI and Landsat8-OLI) on mapping soil properties (clay

content, soil color and iron and organic matter content) using single-date and multi-temporal images in a small agricultural area in southeastern Brazil. We further evaluated the influence of the results on soil management and classification. We constructed three hypotheses as follows: i) multi-temporal images are better than single-date ones to predict and represent the variability of soil properties; ii) spatial, spectral and temporal resolutions of satellite images affect the representation of soil variability and iii) maps of soil properties obtained from different satellites can affect soil classification and management.

2.2. Material and Methods

2.2.1. Study area

The study area has 182 ha and is located in Rafard city, São Paulo State, south-eastern Brazil in a Palaeozoic depression (Fig. 1a). This area was chosen because it has a considerable variation in soil types and geology, especially at short distances, which makes it a suitable area for studying the influence of satellite images at different resolutions. The most common soil classes are Acrisols, Lixisols, Chernozems, Nitisols, Arenosols and Cambisols (Fig. 1b) (IUSS Working Group WRB, 2015) developed from a diversified geology (diabase, siltite, metamorphosed siltite and alluvial deposits) (Fig. 1c). Sugarcane is the cultivated crop. The climate is classified as subtropical mesothermal (Cwa), with an average annual precipitation of 1200 mm and annual average temperature of 24 ℃ (Alvares et al., 2013). A flowchart of the procedures followed in this work is presented in Fig. 2.

2.2.2. Soil sampling and laboratory analysis

A regular grid of 100×100 m was used, where 162 soil samples were collected at 0.20 m depth (Fig. 1a). Soil samples were air dried, ground and sieved to 2 mm for physical and chemical determinations in the laboratory (Teixeira et al., 2017). The clay and sand contents were determined following the pipette method, the iron content of the clay fraction (Fe_2O_3) was obtained by the sulphuric acid digestion and the organic matter content (OM) was determined by the Walkley-Black oxidation method (Walkley and Black, 1934). Additionally, we acquired soil reflectance within the 350–2500 nm spectral range from the 2 mm fraction using a FieldSpec Pro spectroradiometer (Analytical Spectral Devices, Boulder, Colo.) and the Long Light geometry, as described by Romero et al. (2018). The components of the Munsell soil color system (hue, value and chroma) were calculated using the 350–780 nm spectral range (visible portion of soil spectra) and color space transformations, as described by Viscarra Rossel et al. (2006) and Silvero et al. (2021). The Munsell notation was chosen due to its popularity in the soil science community.
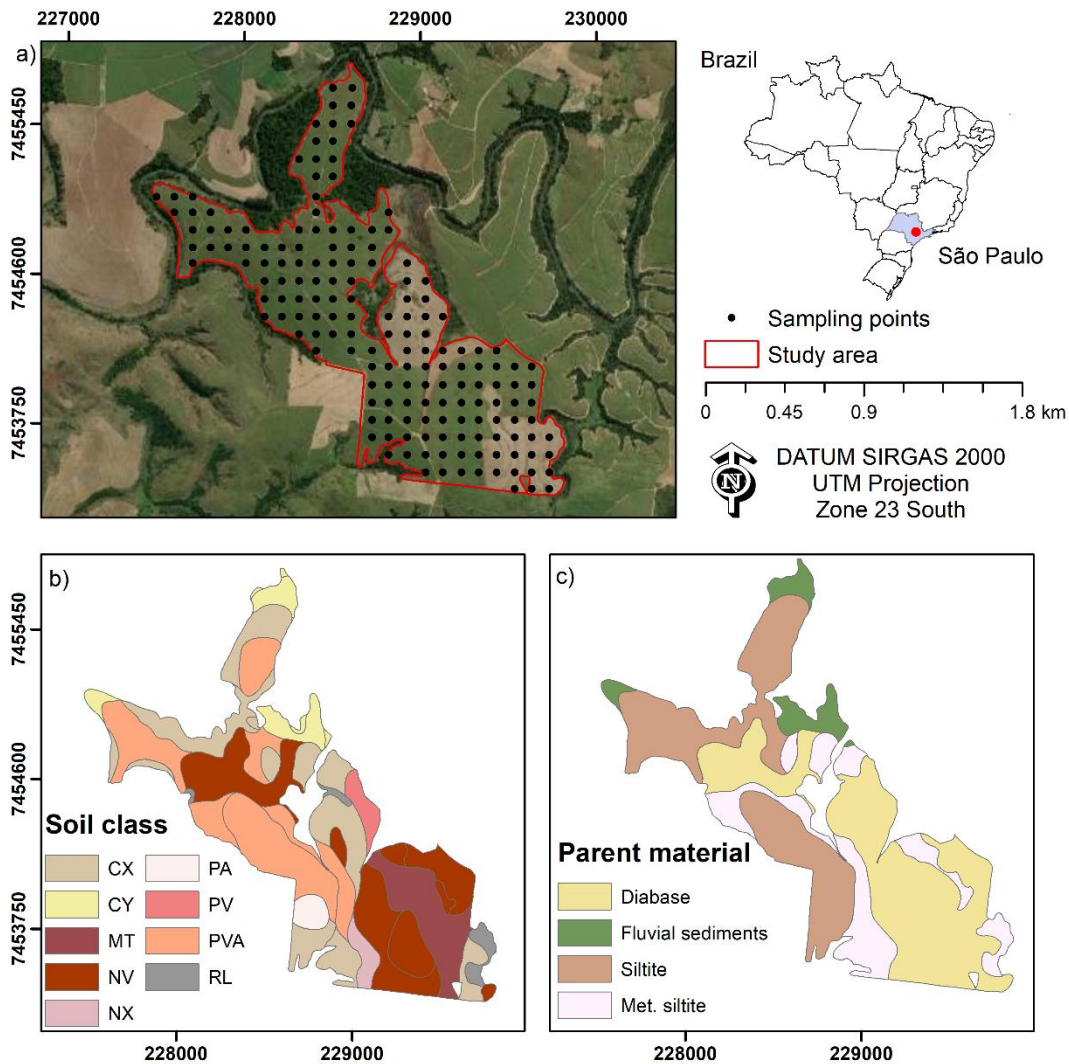
Figure 1. Study site and sampling points (a), soil class (b) and parent material (c) maps. CX and CY – Cambisol (Cambissolo Haplico and Cambissolo Flúvico), MT – Chernozems (Chernossolo Argilúvico), NV and NX – Nitisol (Nitossolo Vermelho and Nitossolo Haplico), PA, PV and PVA - Acrisol/Lixisol (Argissolo Amarelo, Argissolo Vermelho and Argissolo Vermelho Amarelo), RL – Arenosol (Neossolo Litolico).

2.2.3. Satellite images acquisition and processing

For this study, three satellites were used: Landsat8-OLI, Sentinel-2 MSI and PlanetScope. A single date image was obtained for each satellite and multitemporal images for Landsat8-OLI and Sentinel2-MSI, totaling five images that were used as predictors of soil properties in our study area. A brief description of each satellite is given in the next sections.
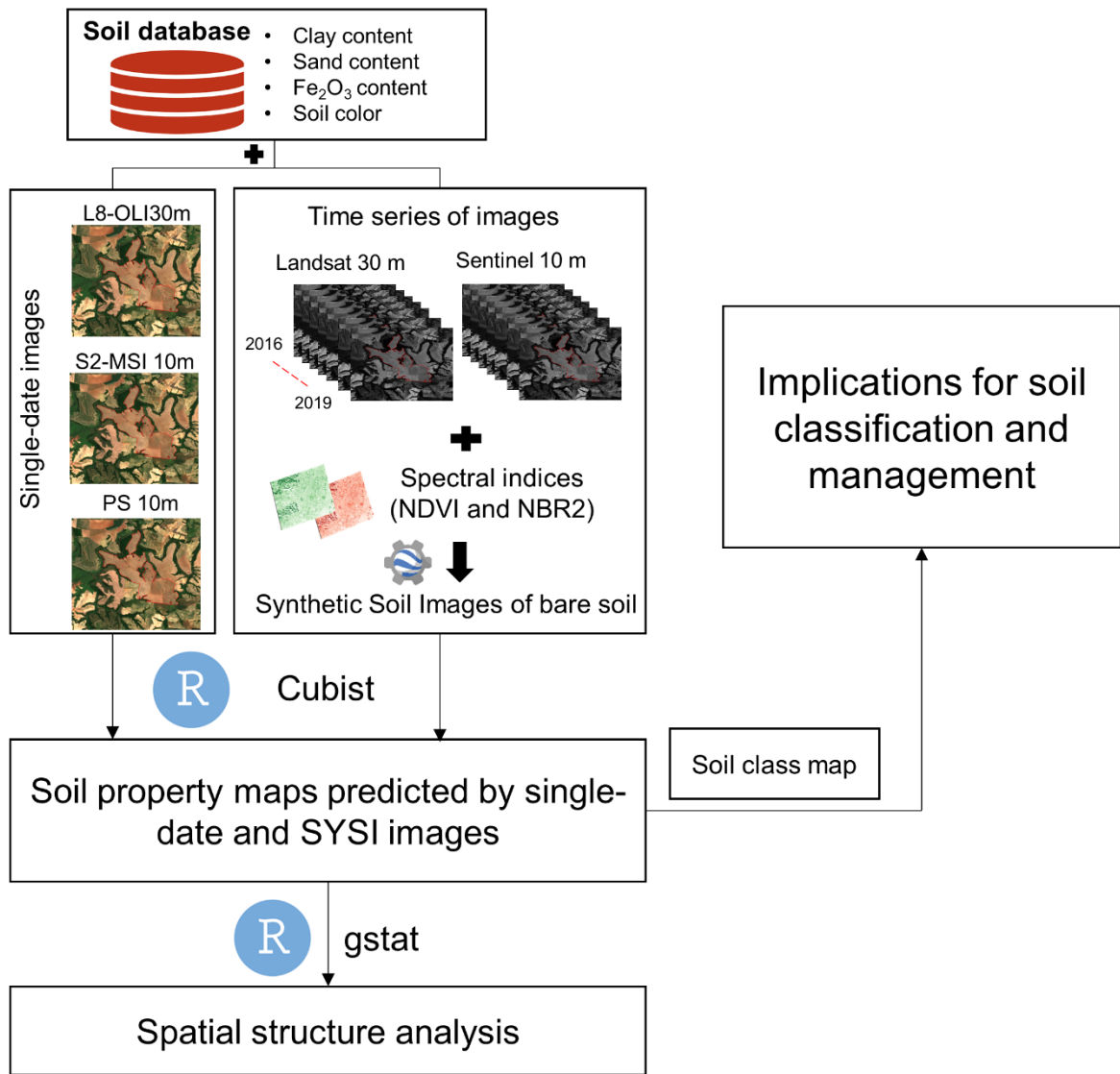
Figure 2. Flowchart of the methodology followed in this work.

2.2.3.1. Single-date satellite image

The single-date Landsat-8 TIER 1 image (L8-OLI), from December 2017, was acquired from the USGS Earth Explorer Platform. The image is from orbit 220 and path 76 and has no clouds or shadows. It was available in surface reflectance, processed by the Landsat Ecosystem Disturbance Adaptive Processing System – LEDAPS (Schmidt et al., 2013) and, therefore, no preprocessing was performed. The L8-OLI image has six spectral bands in the vis-NIR-SWIR region and spatial resolution of 30 m (Table 1).

The single-date Sentinel-2 (S2-MSI) top of atmosphere reflectance (Level-1C) image from November 2017 was downloaded from the Copernicus Open Access Hub web platform. It was atmospherically corrected to Level-2A Bottom-Of-Atmosphere surface reflectance, using the Sen2Cor plugin (Main-Knorn et al., 2017) adapted to R software (R Core Team, 2019). The image was acquired over tile 22KHV and during relative orbit 038. This image has nine spectral bands but we used six of them from the vis-NIR-SWIR region (the red-edge were excluded). The spatial resolutions are 10 and 20 m. The spectral bands that were in 20 m were resampled to 10 m using the

resample function and bilinear method in R software, in order to ensure that all bands were at the same spatial extent.

The single-date Level 3A PlanetScope surface reflectance image with four spectral bands (vis-NIR) and 3 m spatial resolution, also from November 2017, was downloaded from the Planet Labs platform. The Surface Reflectance (SR) Product is derived from the standard Planet Analytic Product (Radiance) and is processed to top of atmosphere reflectance and then atmospherically corrected to surface reflectance. The 6S radiative transfer model with ancillary data from MODIS is used to account for atmospheric effects (Planet Labs, 2020). More details about the characteristics of each satellite can be found in Table 1.

Table 1. Spectral bands, range and central wavelengths of the satellite images used.

Spectral bands	Landsat8-OLI	Sentinel2-MSI	PlanetScope	Landsat8-OLI	Sentinel2-MSI	PlanetScope
	Spectral range (nm)			Central wavelength (nm)		
Blue	450.0 – 510.0	457.5 – 522.5	464.0 – 517.0	482.0	492.4	490.5
Green	530.0 – 590.0	542.5 – 577.5	547.0 – 585.0	561.4	559.8	566.0
Red	640.0 – 670.0	650.0 – 680.0	650.0 – 682.0	654.5	664.6	666.0
*RedEdge1		697.5 – 712.5			704.1	
*RedEdge2		732.5 – 747.5			740.5	
*RedEdge3		733.0 – 793.0			782.8	
NIR	850.0 – 880.0	855.0 – 875.0	846.0 – 888.0	864.6	864.7	867.0
SWIR1	1570.0 – 1650.0	1565.0 – 1655.0		1608.8	1613.7	
SWIR2	2110.0 – 2290.0	2100.0 – 2280.0		2200.7	2202.4	

*The RedEdge bands of the Sentinel-2 MSI satellite were not used in this work. NIR: Near-infrared, SWIR1 and SWIR2: Shortwave infrared.

2.2.3.2. Multi-temporal images

Two multi-temporal bare soil images were obtained for Landsat and Sentinel satellites. For the first, forty-five images were acquired from 2016 to 2019 whereas for the second sixty-three images were used from the same period. The images were selected considering the cloud coverage (up to 20% clouds). A data mining technique called GeoSpatial Soil Sensing System (GEOS3) was used to obtain the multi-temporal images, called Synthetic Soil Image – SYSI (Demattê et al., 2018). The method consists of classification rules based on spectral indices coupled with scene quality products to mask water, burned areas, straw, vegetation, clouds, and shadows. Briefly, the process consisted of selecting images with <20% cloud and shadows. Afterwards, the NDVI (Normalized Vegetation Index) and NBR2 (Normalized Burn Ratio) indices were calculated for each image. Thresholds to separate soils from vegetation, burn areas and straw were used to classify the images as follows: NDVI < 0.25 = 1, NDVI > 0.25 = NA, NBR2 < -0.10 = 1 and NBR2 > 0.10 = NA, where 1 corresponded to bare soils and NA to non-bare soil. The quality masks provided with the images were also classified as 1 for clear sky conditions and NA for areas covered by clouds and shadows. After NDVI, NDBR2 and quality masks were classified, the images were masked, ordered by date and the median reflectance was calculated for each pixel.

The process to obtain the SYSI from Landsat-8 images (SYSI L8-OLI) was carried out in the Google Earth Engine (GEE) cloud-based platform (Gorelick et al., 2017), where surface reflectance images are available and ready for use. The SYSI from Sentinel-2 MSI satellite (SYSI S2-MSI) was obtained in R software, after the images were downloaded and processed for atmospheric correction. Unfortunately, it was impossible to obtain a SYSI for PlanetScope, due to financial constraints (the images are paid) and therefore, only the single-date image was used.

2.2.4. Soil mapping procedure

2.2.4.1. Dataset

The final dataset used for this research had 120 soil observations, after samples without bare soil information were excluded. Descriptive statistics of soil properties were summarized with violin plots and the images' reflectance values for each spectral band were compared by the Tukey test (p-value < 0.05).

2.2.4.2. Model calibration

Prediction models between reflectance values of satellite images and soil properties were built using the Cubist algorithm, which is a rule-based regression tree model that represents linear models of the predictors (Kuhn and Johnson, 2013; Quinlan, 1992). The Cubist algorithm uses a boosting-like procedure named *committees*, which consist of a sequence of rule-based models, and the *k* most similar *neighbors* close enough to the prediction sample, to improve performances (Kuhn and Johnson, 2013). The caret package (Kuhn et al., 2018) with the default number of committees (1, 10 and 20) and neighbors (0, 5 and 9) were used to build the prediction models by 10-fold cross-validation. The standard procedure of data partition in training/validation sets was not performed because our dataset was not large enough to obtain an appropiate validation.

Two approaches were used to evaluate the influence of the spectral resolution (i.e., number of spectral bands): in the first, four bands from the vis-NIR region (RED, GREEN, BLUE, NIR) of all images (single-date and SYSI images) were used as predictors. In the second approach, six bands from the vis-NIR-SWIR region (RED, GREEN, BLUE, NIR, SWIR1, SWIR2) of four images (single-date and SYSIs, excluding the PS image), were used as predictors. The PS image was excluded from this second analysis because it has no bands in the SWIR region, which hampered the comparison with the other satellites.

The influence of the pixel size (spatial resolution) was evaluated by comparing the predictions obtained for each satellite while the influence of temporal resolution on predictions was restricted to the comparison between SYSI images. The Root Mean Square error (RMSE), the coefficient of determination (R^2), and the Mean Absolute Error (MAE) were used to select the best models. The following equations were used:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \hat{y})^2}$$

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

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|$$

where n is the number of samples; y_i is the measured soil property; $\hat{y}_i$ is the predicted soil property and $\hat{y}$ is the mean value of the measured soil property. The best models were used to predict seven soil properties (clay, sand, iron and OM contents and soil color components) at unseen locations using the five satellite images.

2.2.4.3. Spatial structure analysis of predicted soil property maps

In order to evaluate the differences in spatial structure among the soil property maps predicted by different satellites, variograms were obtained and their parameters (sill, range and nugget) compared for each soil property. The nugget to sill ratio [(nugget/sill) *100], defined as "the proportion of short-range variability that cannot be described by a geostatistical model in the studied field" (Kravchenko, 2003), was obtained in order to evaluate the spatial dependence. The number of samples for each predicted map and satellite was different, as the spatial resolution differed (3, 10 and 30 m). The following formula was used to obtain the empirical variograms:

$$\gamma(h) = \frac{1}{2n} \sum_{i=1}^{n} \{z(x_i) - z(x_i + h)\}^2$$

where $\gamma(h)$ is the average semivariance; n is the number of pairs of points; $z(x_i)$ is the value of z in the position (x_i); and $z(x_i + h)$ is the value of z at the position $(x_i + h)$. Exponential, gaussian and spherical theoretical models were tested to fit the experimental variograms. The geostatistical analysis was carried out in R software, using the gstat package.

2.2.5. Applications of predicted soil maps for management and classification

2.2.5.1. Soil legacy map

The soil property maps predicted by different satellites were related to a conventional soil class map developed by Bazaglia Filho et al. (2013). This soil class map is on a 1:10,000 scale. It was obtained using stereopairs of coloured aerial photos, topographic maps with contour lines at 5 m-intervals, semi-detailed soil maps and geological information. Pits (30) and soil observations (160) were also used for delineating the soil mapping units (SMU) and named according to the fourth level (suborder) of the Brazilian Soil Classification System (SiBCS, from the Portuguese name). More details can be found in Bazaglia Filho et al. (2013).

2.2.5.1. Relationship of soil class map and predicted soil properties

We selected a small area that presented a considerable contrast in soil texture, color, iron and organic matter contents to be compare with our predicted maps. The predicted maps of clay content were classified into five textural classes, as follows (g kg^{-1}): <150 = sandy, 150 – 250 = loam sandy, 250 – 350 = loam clay, 350 – 600 = clay, >600 = very clay. The soil color components (hue, value and chroma) were presented as Munsell's notation, and iron and organic matter contents in their original value. The soil color and iron content maps were used to compare their patterns to those from the legacy soil map. The mean and standard deviation of each soil property for each soil class in the small selected area was also calculated, in order to show if the values significantly varied within soil classes and how this can impact or improve other soil studies. Regarding the use of soil property maps for soil

classification, we used the Nitisol class and the clay content value (=>350 g kg^{-1}), described in the Brazilian Soil Classification System as the main criteria for the Nitisol class (SiBCS) (Santos et al., 2013). We used only the Nitisol class since it is considered as more homogeneous in depth, and the information from the 0–20 cm depth can be used to represent this soil class.

2.3.   Results

2.3.1.   Spectral patterns of satellite images and variability of soil properties

Single-date and SYSI images show very similar spectral patterns (Fig. 3), excluding the areas without bare soil information (black pixels in Fig. 3a, d–e). Single-date S2-MSI and PS images were acquired on the same date (11/17/2017) and it was possible to obtain bare soil pixels for the entire study area. This was not possible for the single-date L8-OLI image, acquired one month later (December 2017), when the sugarcane was already growing in some parts of the area (black pixels in Fig. 3a). The areas without bare soil pixels were excluded from further analysis (red lines in Fig. 3).
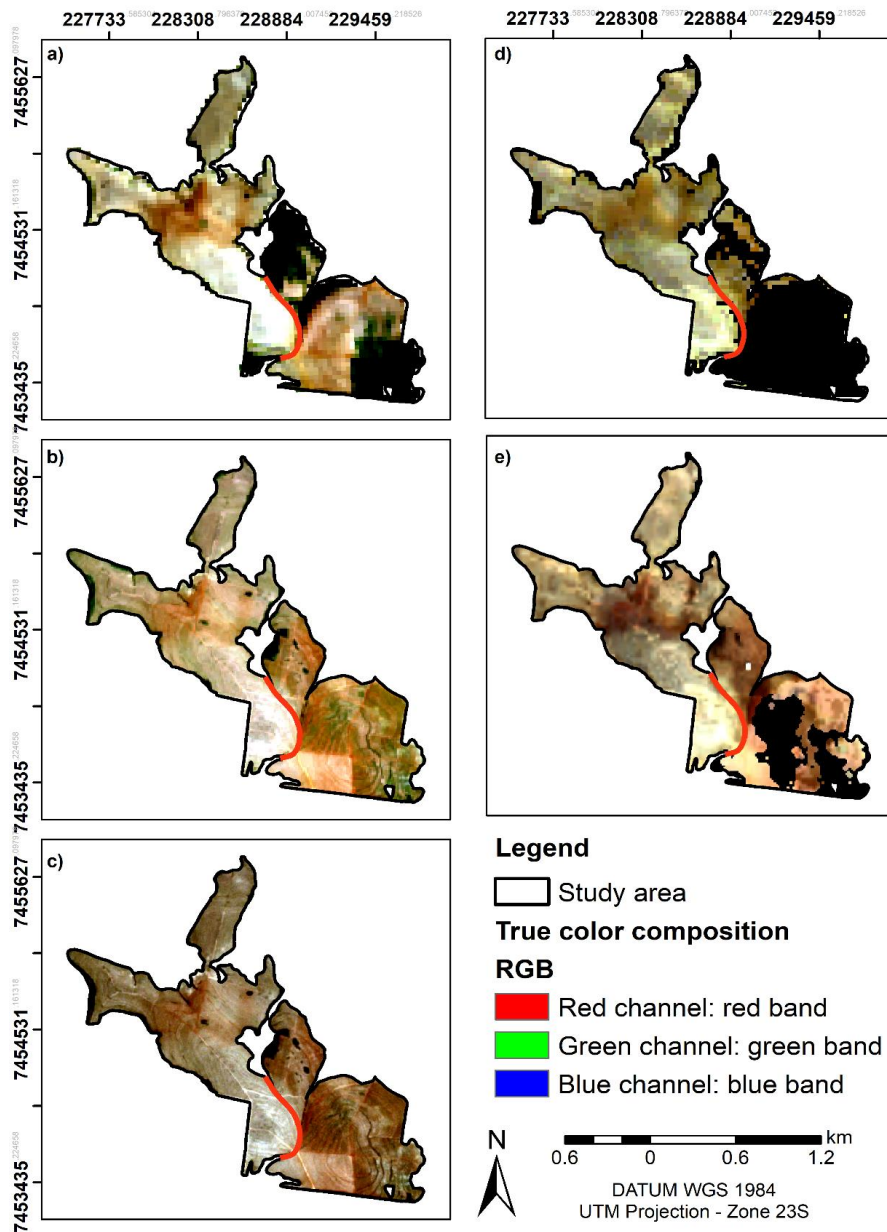
Figure 3. Single-date and multi-temporal images used in this work. a) Single-date Landsat8 – OLI image from December 2017, b) Single-date Sentinel2 – MSI image from November 16, 2017, c) Single-date PlanetScope image from November 16, 2017; d) Multi-temporal bare soil images from Landsat L8-OLI (SYSI L8-OLI) and e) Sentinel2-MSI (SYSI S2-MSI), respectively. Black colors in single-date and SYSI images are areas without bare soil pixels.

The reflectance values for each band of single-date and SYSI images varied significantly across all spectral regions (Fig. 4) but SYSI images show a more constant increase in their reflectance values from RED to SWIR1 bands than single-date ones. The single-date S2-MSI image has the highest mean reflectance in BLUE (8.57 10^{-2}), GREEN (12.43 10^{-2}) and RED (16.88 10^{-2}) bands while the L8-OLI bands have the highest standard deviations (Table 2). Among the SYSI images, the SYSI S2-MSI has the highest mean and standard deviation for all spectral bands (Table 2).

Clay, sand and iron contents (Fe_2O_3) are highly variable across the study area (Fig. 5), with mean and standard deviations of 325.26 ± 178.70, 481.84 ± 203.43 and 88.32 ± 79.37, respectively. The organic matter content (OM) was lower than 10% and soil colors varied from red to brown yellow (hues between 8 and 18).
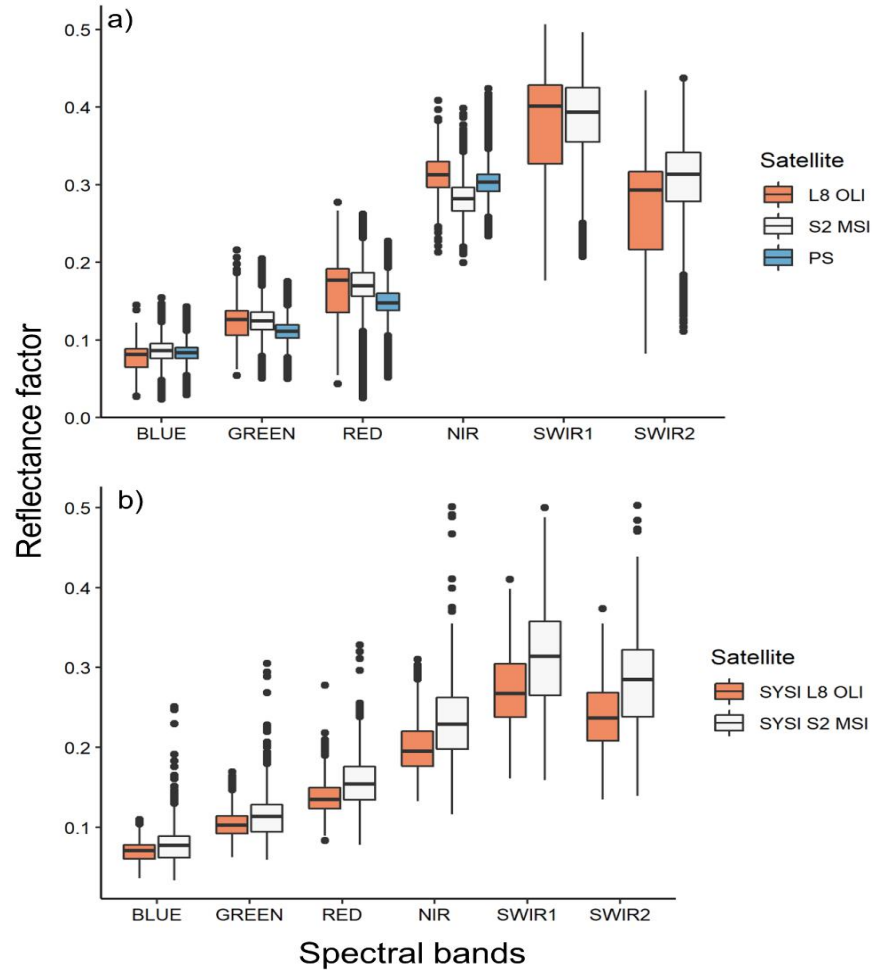
Figure 4. Distribution of reflectance values of satellites' spectral bands from: a) single-date and b) multi-temporal images (SYSI).

Table 2. Spectral bands, range and central wavelengths of the satellite images used.

	Single-date images						SYSI images			
	L8-OLI*		S2-MSI*		PS*		L8-OLI*		S2-MSI*	
B1	7.58±2.03	Dc	8.57±1.50	Aa	8.30±1.24	Bb	7.05±1.33	Eb	7.73±1.85	Ca
B2	12.19±2.44	Bb	12.43±1.82	Aa	11.09±1.42	Dc	10.47±1.80	Eb	11.29±2.37	Ca
B3	16.11±4.36	Bb	16.88±2.87	Aa	14.76±2.02	Dc	13.87±2.17	Eb	15.63±2.90	Ca
B4	31.30±2.46	Aa	28.10±2.17	Cc	30.20±1.72	Bb	20.13±3.27	Eb	23.05±4.41	Da
B5	37.26±7.94	Bb	38.99±4.55	Aa			27.14±4.68	Db	31.40±6.15	Ca
B6	26.38±7.79	Cb	31.11±4.41	Aa			23.94±4.05	Db	28.23±5.30	Ba

*Scaled by 10^{-2}. B1: Blue band, B2: Green band, B3: Red band, B4: Near Infrared band, B5: Shortwave Infrared 1 (SWIR1), B6: Shortwave Infrared 2 (SWIR2). Uppercase green letters correspond to contrast between four spectral bands of single-date and SYSI images. Uppercase orange letters correspond to contrast between B5 and B6 spectral bands of single-date and SYSI images. Lowercase red letters correspond to contrast within four spectral bands of single-date images. Lowercase blue letters correspond to contrast between four spectral bands of SYSI images. Lowercase black letters correspond to comparison between B5 and B6 of single images (except PS).
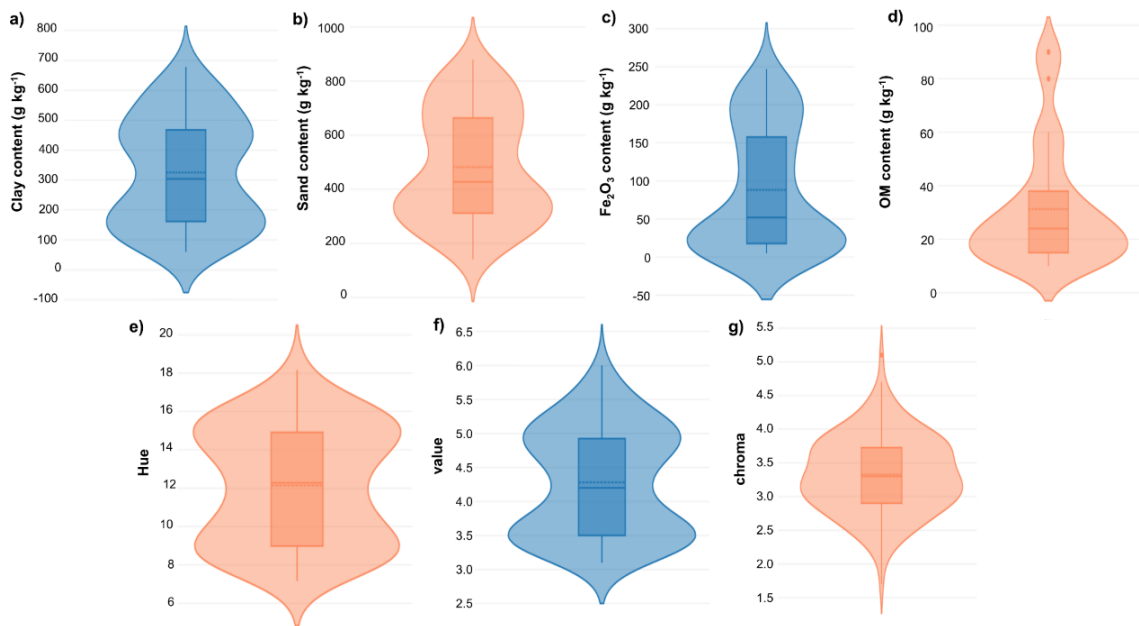
Figure 5. Distribution of reflectance values of satellites' spectral bands from: a) single-date and b) multi-temporal images (SYSI).

2.3.2. Model performances

The model performances of soil properties are shown in Table 3 (blue and red colors correspond to the best models obtained using four vis-NIR and six vis-NIR-SWIR spectral bands as predictors, respectively). We found that the SYSI S2-MSI presented the best performances in cross-validation for almost all soil properties either using four or six bands, except OM content and chroma, which were best predicted by single-date S2-MSI and L8-OLI, respectively. The difference in model performances using either four or six spectral bands was not very

significant but the performance was better using six spectral bands as predictors for almost all soil properties, except chroma and hue, which are components of soil color.

Table 3. Spectral bands, range and central wavelengths of the satellite images used.

Soil properties	NB	C	N	RMSE	R²	MAE	C	N	RMSE	R²	MAE
				Clay content (g kg⁻¹)					Hue		
PS	Four	10	9	124.75	0.54	97.87	1	9	1.95	0.61	1.58
S2-MSI	Four	10	9	116.65	0.59	94.28	20	9	1.71	0.69	1.36
L8-OLI	Four	20	5	120.4	0.62	94.19	1	0	1.75	0.67	1.42
SYSI S2-MSI	**Four**	**1**	**0**	**110.6**	**0.65**	**90.51**	**10**	**0**	**1.41**	**0.77**	**1.16**
SYSI L8-OLI	Four	20	0	130.78	0.52	107.28	20	9	2.17	0.45	1.77
S2-MSI	Six	20	9	106.81	0.68	83.72	10	9	1.75	0.67	1.35
L8-OLI	Six	10	5	110.21	0.66	85.71	1	0	1.78	0.65	1.44
SYSI S2-MSI	**Six**	**20**	**0**	**86.69**	**0.75**	**68.75**	**1**	**0**	**1.43**	**0.76**	**1.15**
SYSI L8-OLI	Six	10	0	111.25	0.64	90.64	1	9	2.06	0.53	1.66
				Sand content (g kg⁻¹)					Value		
PS	Four	10	0	168.62	0.48	136.92	20	0	0.48	0.60	0.39
S2-MSI	Four	20	0	156.75	0.47	126.86	20	9	0.41	0.71	0.32
L8-OLI	Four	10	5	159.14	0.47	121.03	10	0	0.45	0.68	0.36
SYSI S2-MSI	**Four**	**20**	**0**	**157.7**	**0.48**	**122.61**	**10**	**0**	**0.36**	**0.77**	**0.28**
SYSI L8-OLI	Four	10	0	156.42	0.46	126.83	10	0	0.51	0.55	0.41
S2-MSI	Six	10	0	145.09	0.55	116.22	10	0	0.42	0.70	0.34
L8-OLI	Six	10	0	151.86	0.52	123.86	10	0	0.42	0.73	0.34
SYSI S2-MSI	**Six**	**1**	**0**	**132.2**	**0.60**	**101.32**	**20**	**0**	**0.35**	**0.79**	**0.28**
SYSI L8-OLI	Six	10	0	138.68	0.58	113.35	1	9	0.47	0.52	0.38
				Fe₂O₃ content (g kg⁻¹)					chroma		
PS	Four	1	0	60.26	0.48	44.22	10	0	0.43	0.47	0.34
S2-MSI	Four	10	0	52.54	0.63	40.58	**10**	**0**	**0.43**	**0.51**	**0.34**
L8-OLI	Four	1	5	51.19	0.61	38.77	10	0	0.46	0.41	0.36
SYSI S2-MSI	**Four**	**10**	**0**	**48.73**	**0.64**	**37.44**	1	9	0.45	0.41	0.35
SYSI L8-OLI	Four	10	9	55.86	0.54	42.47	1	0	0.5	0.27	0.4
S2-MSI	Six	1	9	49.14	0.65	36.52	**20**	**0**	**0.43**	**0.45**	**0.34**
L8-OLI	Six	10	9	45.84	0.68	34.12	1	0	0.44	0.45	0.34
SYSI S2-MSI	**Six**	**20**	**0**	**38.57**	**0.78**	**29.92**	1	0	0.44	0.44	0.35
SYSI L8-OLI	Six	20	9	49.63	0.64	36.18	10	0	0.5	0.25	0.4
				OM content (g kg⁻¹)							
PS	Four	1	0	9.38	0.42	7.42					
S2-MSI	Four	1	5	8.21	0.46	6.27					
L8-OLI	**Four**	**1**	**9**	**8.1**	**0.54**	**6.38**					
SYSI S2-MSI	Four	20	0	8.37	0.47	6.33					
SYSI L8-OLI	Four	1	9	8.42	0.47	6.42					
S2-MSI	Six	10	0	8.24	0.50	6.26					
L8-OLI	**Six**	**10**	**9**	**8.03**	**0.56**	**6.18**					
SYSI S2-MSI	Six	10	0	7.89	0.53	6.06					
SYSI L8-OLI	Six	10	9	8.25	0.51	6.41					

NB: number of spectral bands used in predictions, C: committees or number of model trees or boosting iterations used, N: neighbors used for each iteration, R²: coefficient of determination, RMSE: Root Mean Square Error (g kg⁻¹), MAE: Mean absolute error (g kg⁻¹). PS, S2-MSI and L8-OLI correspond to single-date images from PlanetScope, Sentinel-2 and Landsat-8 satellites respectively. SYSI: multi-temporal bare soil images.

The predicted maps, except sand content, are shown in Fig. 6. Soil color maps are presented in Munsell's notation (hue value/chroma) and it can be noted that their spatial patterns appear very similar, with dark colors in the center of the study area and lighter in the other regions. Dark colors are associated with higher values of clay and Fe_2O_3 content (Fig. 6) but the values of these soil properties varied among satellites. The predicted maps of Fe_2O_3 showed values that differ among satellites, although their spatial patterns are similar. The single-date PlanetScope is prone to underestimate Fe_2O_3 content (200 g kg^{-1}). The same can be said for clay content but not for OM content, whose maps show a different spatial pattern among satellites, with different range of values. Although our study area has values of OM content of up to 80 g kg^{-1}, the satellites were unable to predict values above 50 g kg^{-1}.

2.3.3. Spatial structure analysis of predicted soil property maps

The variogram parameters helped us to explain the differences in spatial structure of the predicted maps (Table 4 and Fig. S1 in supplementary material). Exponential and spherical models were fitted to the experimental variograms and contrasting values of range, sill and nugget were found among the soil property maps predicted by different satellites. The range parameter varies from 137.05 to 1156.75 m and the highest values were observed for almost all soil properties predicted by the PS image, except for the hue component predicted by SYSI L8-OLI (1156.75 m). We observed that most of the soil properties predicted by SYSI L8-OLI have the highest values of sill (semivariance) and the lowest values of the range parameter, excluding the hue component mentioned above.

The nugget to sill ratio (N/S), which is a measure of the degree of spatial dependence was also evaluated (Table 4). Overall, the hue component and Fe_2O_3 content presented a strong spatial dependence, no matter which satellite was used to predict them. For both cases, the range parameter reached values between 400 and 500 m, except those predicted by SYSI L8-OLI, which were adjusted to exponential models and had the lowest (137.05 m) and highest (1156.75 m) values of the range parameter, for hue component and Fe_2O_3 content, respectively. For clay and sand contents, the spatial dependence was strong for single-date L8-OLI and SYSIs and weak and moderate for single-date PS and S2-MSI, respectively. For both soil properties, the lowest values of the range parameter and the highest values of the sill parameter were observed when they were predicted by the SYSI L8-OLI. The nugget parameter for sand content was the highest when it was predicted by the single-date S2-MSI image, which caused the weak spatial dependence (Table 4). OM content also had moderate dependence when predicted by single-date PS and S2-MSI images. The value component presented strong spatial dependence for almost all satellites, except when we used SYSI L8-OLI. Chroma, predicted by single-date S2-MSI and SYSI L8-OLI, showed moderate spatial dependence and had the lowest values of the range parameter when compared to the other satellites.
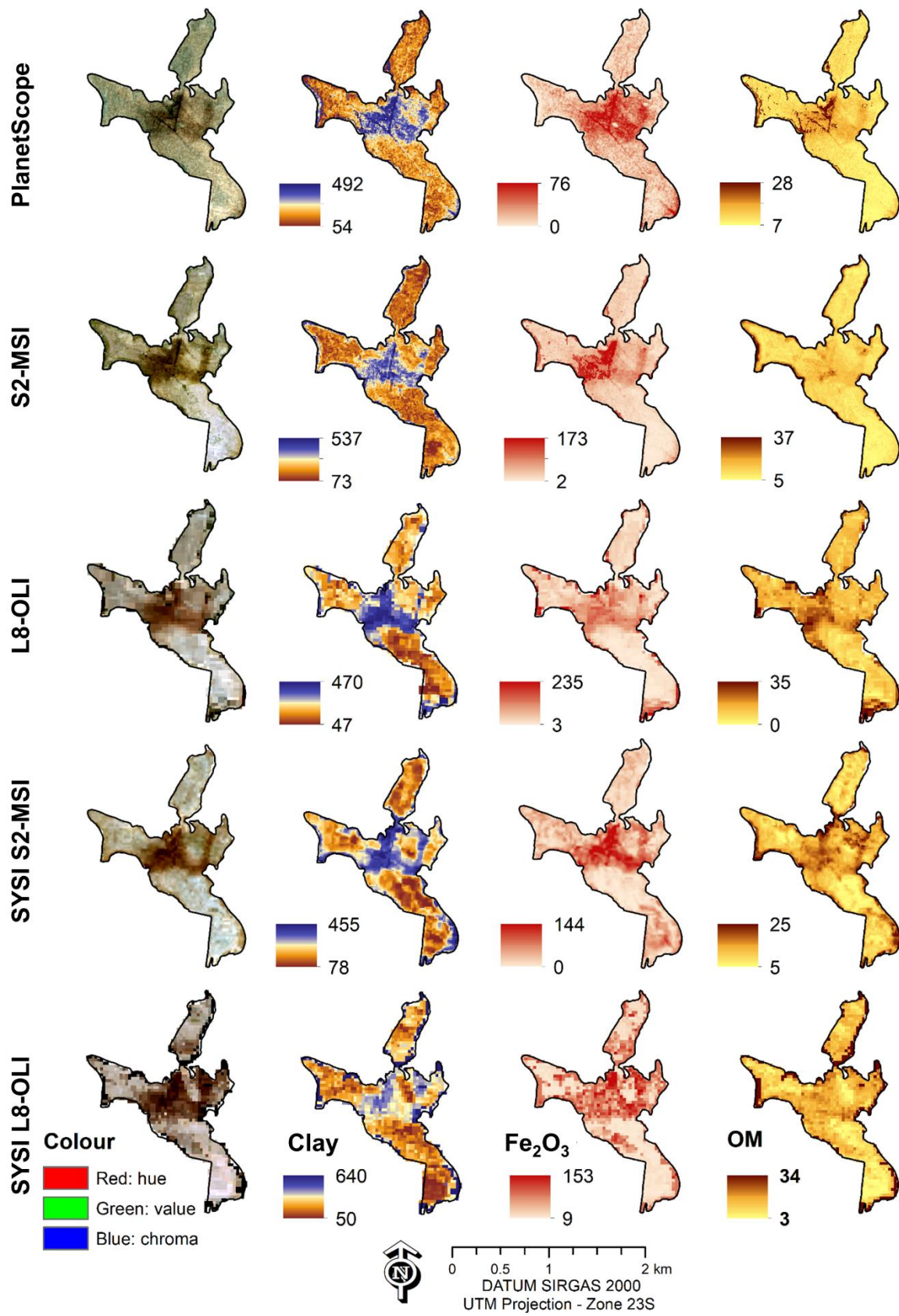
Figure 6. Predicted maps of soil properties. The soil color components (hue, value and chroma) are presented as RGB composition. Clay, Fe_2O_3 and OM contents are in g kg^{-1}.

Table 4. Semivariogram parameters of predicted soil properties

Soil properties	Model	nugget	range	sill	N/S (%)		Model	nugget	range	sill	N/S (%)	
Clay content										Hue		
PS	Sph	4578.93	**555.82**	14411.03	**31.77**	M	Sph	0.59	530.00	3.82	15.45	S
S2-MSI	Sph	4474.02	404.83	14993.10	29.84	M	Sph	0.97	429.96	7.32	13.25	S
L8-OLI	Sph	1523.04	414.82	14883.85	10.23	S	Sph	0.54	473.64	5.66	9.54	S
SYSI S2-MSI	Sph	1212.22	347.50	11036.14	10.98	S	Sph	0.00	511.24	4.44	0.00	S
SYSI L8-OLI	Sph	**6309.85**	252.09	**26050.60**	24.22	S	Exp	**2.54**	**1156.75**	**11.20**	**22.76**	S
Sand content										value		
PS	Sph	6597.23	**544.52**	17297.15	38.14	M	Sph	0.03	**528.06**	0.23	13.04	S
S2-MSI	Sph	**8395.48**	308.27	10662.60	**78.74**	W	Sph	0.056	466.70	0.39	14.51	S
L8-OLI	Sph	1327.68	380.45	18354.21	7.23	S	Sph	0.05	446.54	0.41	12.20	S
SYSI S2-MSI	Sph	1548.86	323.35	12259.58	12.63	S	Sph	0.004	511.51	0.31	1.27	S
SYSI L8-OLI	Sph	4854.45	242.63	**23867.24**	20.34	S	Sph	0.18	321.80	0.60	30.00	M
Fe_2O_3 content										chroma		
PS	Sph	81.04	**533.43**	569.49	14.23	S	Sph	0.026	**581.40**	0.13	20.63	S
S2-MSI	Sph	428.19	456.92	2322.84	18.43	S	Exp	0.009	406.26	0.02	**37.50**	M
L8-OLI	Sph	**860.94**	404.68	**4038.30**	21.32	S	Sph	0.004	516.43	0.16	2.44	S
SYSI S2-MSI	Sph	88.10	402.21	1946.11	4.53	S	Exp	0.000	180.49	0.10	0.00	S
SYSI L8-OLI	Exp	531.77	137.05	1941.28	**27.39**	S	Exp	**0.078**	175.63	**0.24**	32.77	M
OM content												
PS	Sph	9.50	**468.94**	27.10	**35.06**	M						
S2-MSI	Exp	15.83	141.01	50.86	31.12	M						
L8-OLI	Sph	6.48	329.16	42.41	15.28	S						
SYSI S2-MSI	Exp	3.37	155.09	30.22	11.15	S						
SYSI L8-OLI	Sph	**17.57**	232.51	**69.67**	25.22	S						

N/S: nugget to sill ratio expresses the degree of spatial dependence: S: strong, M: moderate, W: weak. Exp: Exponential model, Sph: Spherical model, PS, S2-MSI and L8- OLI correspond to single-date images from PlanetScope, Sentinel-2 and Landsat-8 satellites, respectively. SYSI: multi-temporal bare soil images. OM: organic matter, Fe2O3: iron content

2.3.4. Relationship of predicted soil property maps with legacy soil data

The predicted maps of soil color, clay (as textural classes) and iron contents were related to the soil class map (red and black lines in Figs. 7–9). The mean and standard deviation within each soil class and between satellites were also compared (Table 5).

All soil color maps distinguished the main variability patterns of the study area (Fig. 6) but the coarse spatial resolution of Landsat image, for example, affected some details (Fig. 7). A suggested delineation of similar colors was proposed in Fig. 7. However, their different spatial resolution led to draw somewhat different lines (dash orange lines in Fig. 7) which did not perfectly agree with the soil classes drawn in the legacy information (red lines in Fig. 7). The mean and standard deviation of soil color varied significantly within each soil class and among satellites, but visually the colors are similar, with few exceptions (Table 5). The soil color for CXd5 (Cambisol) and PVAdab (Acrisol/ Lixisol) predicted by SYSI L8-OLI and SYSI S2-MSI shows the most contrasting values compared to the other satellites. For CXd5, the values of hue varied from 2.7 to 3.5 YR (brown yellow) but those predicted by SYSI L8-OLI, the color appears redder (0.9YR 4.0/3.6). The same was observed for PVAdab predicted by SYSI S2-MSI, which shows a browner color (3.4YR 4.8/3.4).

The predicted maps of clay content (as textural classes) for each satellite are shown in Fig. 8. The high spatial resolution of PS image (3 m) provides a more fragmented texture map than those obtained by Sentinel-2 and Landsat-8 satellites (both single-date and SYSIs). Although similar patterns, the spatial resolution impacted in some details. For example, in the portion of the area corresponding to the PVAdeab4 soil class (number 4 in Fig. 8), the predominant textural classes in maps predicted by single-date L8-OLI and both SYSI S2-MSI and SYSI L8-OLI are

sandy and loam sandy. In contrast, the same information predicted by single-date PlanetScope and S2-MSI images have the loam sandy as predominant textural class. The same can be observed for the NVLef2 (Nitisol) soil class. This soil class is characterized by a clay texture, but a mixture of textures inside the polygon was found for all predicted maps (Fig. 8).

The mean and standard deviation of clay content varied significantly inside each soil class in function of the satellite used (Table 5). For example, the mean value of clay content within the CXe3 (Cambisol), predicted by PlanetScope (270.2 g kg^{-1}), is two times higher than the one predicted by SYSI L8-OLI (136.7 g kg^{-1}) while the difference with the single-date L8-OLI is approximately 60 g kg^{-1}. The mean values of clay content for CXe3 (Cambisol) and PVAeadb4 (Acrisol/Lixisol) classes decreased from PS to SYSI L8-OLI while for CXd5 (Cambisol), the values increased as the pixel size increased (Table 5). For the other soil classes, very similar mean and standard deviations were observed but, for all of them, the map predicted by the SYSI L8-OLI shows the highest standard deviation, even >100 g kg^{-1}.

Iron content maps (Fe$_2$O$_3$) have spatial patterns similar to those observed in clay content and soil color maps (Fig. 9) and varied considerably inside each soil class and between satellites (Table 5). For almost all soil classes, except PVAdeab4 (Acrisol/Lixisol), the predicted mean values tend to increase as the pixel size increases from 3 to 30 m. Again, the highest values of mean and standard deviation are for the maps predicted by the SYSI L8-OLI and the single-date L8-OLI images. The most contrasting values were found within the PVAdab soil class (Acrisol/Lixisol), with mean and standard deviation ranging from 15.6 ± 17.4 to 72.1 ± 66.2, values predicted by the SYSI S2-MSI and the single L8-OLI, respectively.

The percentage of the Nitisol class area (NVLef2) that was correctly delineated in the legacy soil map was calculated considering the variability of the clay content values within it is shown in Table 6 for each satellite. We performed this task only for Nitisol, as this is the only soil class within our study area that has a criterion based on clay content (>350 g kg^{-1}) in the Brazilian Soil Classification system (SiBCS). We found that none of the predicted maps meet 100% of the criterion (clay content above > 350 g kg^{-1}, but the clay content maps predicted by the SYSI L8-OLI and the SYSI S2-MSI show the highest (71.58%) and lowest (47.92%) percentage of the area correctly delineated, respectively.
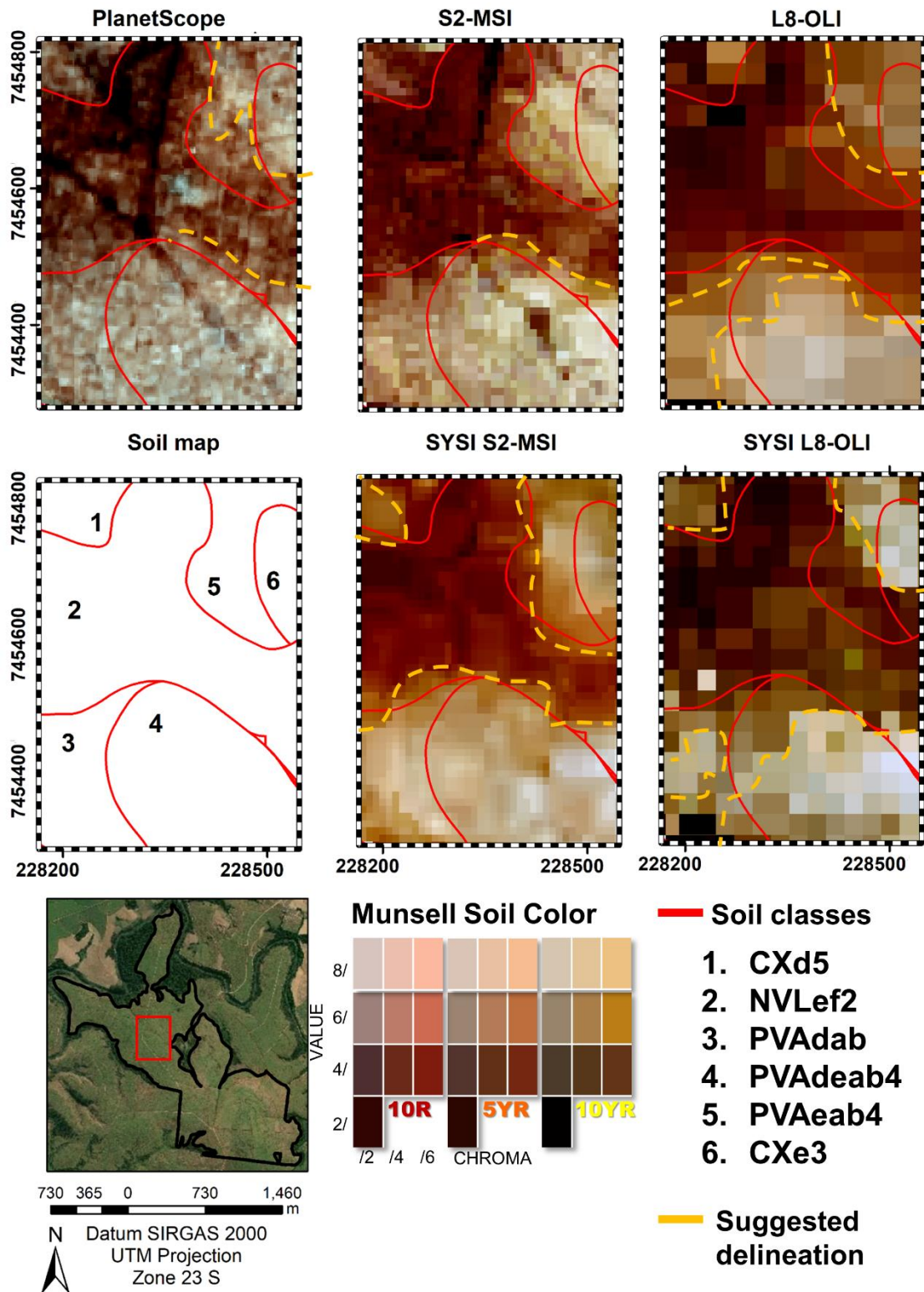
Figure 7. Predicted maps of soil color for each satellite. The soil classes were described as follows: 1) CXd5: Cambisol (Cambissolo Háplico distrófico muito arenoso), 2) NVLef2: Nitisol (Nitossolo Vermelho Latossólico eutroférrico argiloso), 3) PVAdab: Acrisol/Lixisol (Argissolo Vermelho Amarelo distrófico abrúptico), 4) PVAdeab4: Acrisol/Lixisol (Argissolo Vermelho Amarelo distrófico-eutrófico abrúptico textura arenosa), 5) PVAeab4: Acrisol/Lixisol (Argissolo Vermelho Amarelo eutrófico abrúptico textura arenosa), 6) CXe3: Cambisol (Cambissolo Háplico eutrófico textura média). The orange traced line corresponds to the suggested delineation of soil mapping units according to the color patterns of each map.

2.4.   DISCUSSIONS

2.4.1.   Spectral patterns of single-date and SYSI images

The study area is characterized by a wide range of parent materials and the studied soil properties followed this same pattern of variation. The five satellite images used in this work had similar spectral patterns, but for the SYSI images and the single-date L8-OLI, we were unable to retrieve bare soil pixels for the whole study area. The high incidence of clouds, the presence of vegetation, and the sugarcane management explained the absence of bare soil pixels in both SYSI and single-date L8- OLI images (Roberts et al., 2019; Castaldi et al., 2019).

Among the SYSI images, the SYSI S2-MSI had a higher percentage of pixels with bare soil than SYSI L8-OLI. This difference was a result of the different number of images used to obtain each SYSI (45 and 63 images, respectively). The temporal resolution played an important role in this regard, as the number of images relied on the revisit period. The Sentinel-2 satellite has a revisit period of approximately five days, which is more than three times shorter than that of Landsat-8 (16 days), which allowed a more constant monitoring of the soil's behaviour over time.

Single-date images had the whole study area with bare soil pixels (except L8-OLI), but the analysis of the reflectance values revealed that the SYSI images had spectral patterns more related to soils, in which the values generally increase from visible to SWIR1 and then decreases in the SWIR2 (Demattê et al., 2014; Poppiel et al., 2019). Regarding the single-date PS image, its spectral signature was difficult to analyze, as it lacks spectral bands in the SWIR region, which may limit the study of soil characteristics related to this region, such as texture and phyllosilicate minerals (Fang et al., 2018; Gomez et al., 2018). We found that the reflectance values of the SYSI images were lower than single-date ones and this may be related to the way in which SYSIs were calculated (Demattê et al., 2018), i.e., the median reflectance of a period of four years.

2.4.2.   Variability of soil properties explained by satellite images

For almost all soil properties, except OM content and chroma component, the best model performances were obtained using six spectral bands and the SYSI S2-MSI with 10-m pixel size. The difference between single-date and SYSI images and their spatial, spectral and temporal resolutions was not significant, but we observed that the satellite's resolutions are important factors for soil mapping. For example, the PlanetScope image has the highest spatial and the lowest spectral resolutions, but it did not improve the performance of the models. Probably this is due to the lack of spectral bands in the SWIR region, which were described as important predictors of some soil properties, such as texture (Gomez et al., 2015).

We expected to obtain better results with the single-date PlanetScope image, as it had the highest spatial resolution and would be able to represent the short-scale variability. In this work, we observed that the SYSI S2-MSI with a 10 m pixel size was the most adequate to explain the variability of soil properties in an area with <200 ha. In larger areas, at the regional level, the results may be different and satellites with larger pixel sizes may be more suitable to explain the variability. This was observed in a recent study conducted by Silvero et al. (2021), in the tropical region of Piracicaba, in Brazil. The authors reported that the SYSI obtained from the Landsat-8 satellite, with a 30 m pixel size, was better than the SYSI obtained from the Sentinel-2 satellite. Other authors have also reported that high spatial resolution images do not always offer improvements in soil mapping (Gomez et al., 2015; Guo et al., 2019).
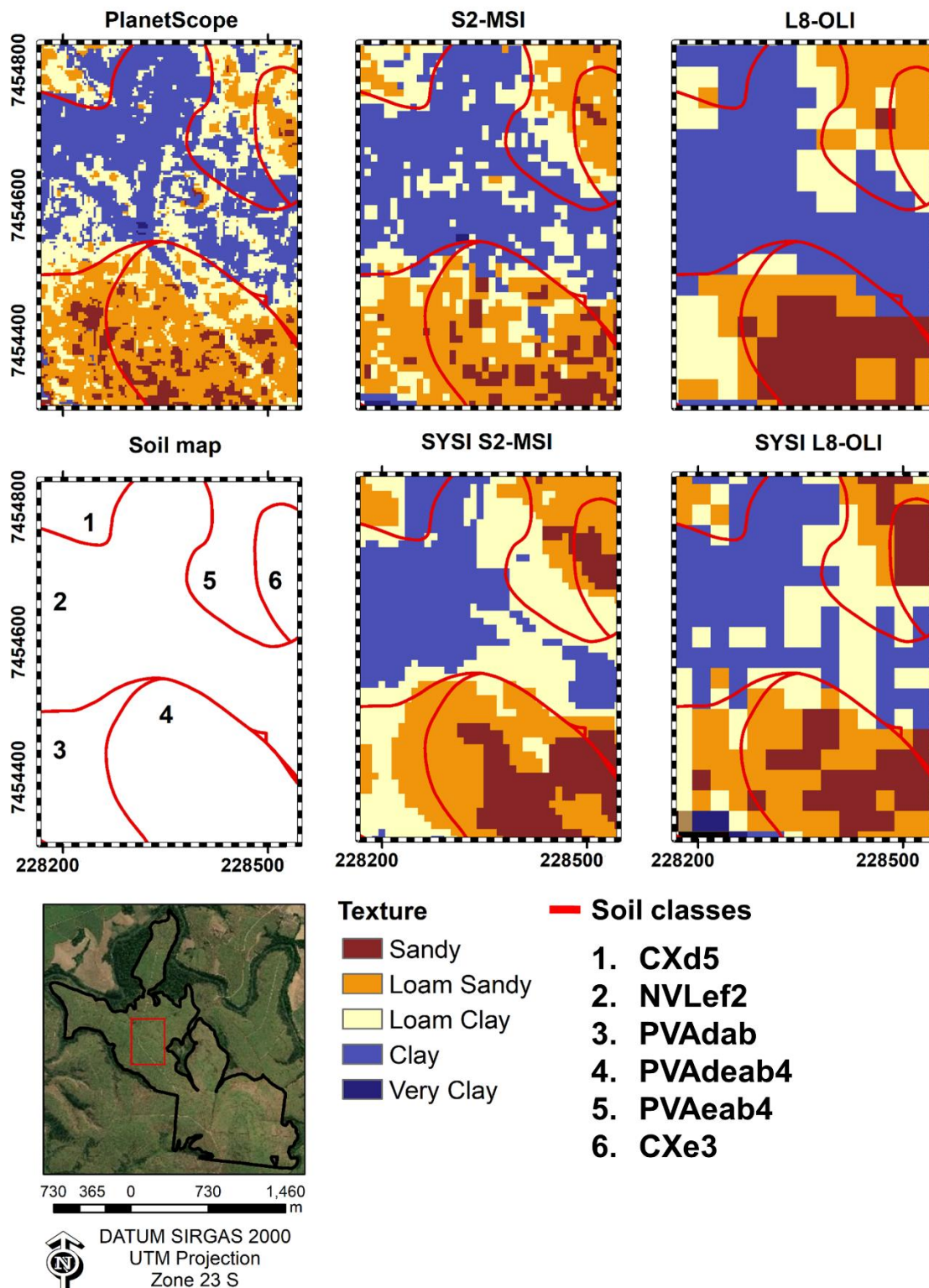
Figure 8. Predicted maps of soil texture for each satellite. The soil classes were described as follows: 1) CXd5: Cambisol (Cambissolo Háplico distrófico muito arenoso), 2) NVLef2: Nitisol (Nitossolo Vermelho Latossólico eutroférrico argiloso), 3) PVAdab: Acrisol/Lixisol (Argissolo Vermelho Amarelo distrófico abrúptico), 4) PVAdeab4: Acrisol/Lixisol (Argissolo Vermelho Amarelo distrofico-eutrófico abrúptico textura arenosa), 5) PVAeab4: Acrisol/ Lixisol (Argissolo Vermelho Amarelo eutrófico abrúptico textura arenosa), 6) CXe3: Cambisol (Cambissolo Háplico eutrófico textura média).

The temporal resolution, on the other hand, is important for obtaining continuous bare soil information of the area under study and to reduce the influence of vegetation, moisture and crop residues (Roberts et al., 2019;

Castaldi et al., 2019). In this study, the SYSI images showed spectral patterns more related to soils and their model performances were also better than single-date ones. Although suitable results were already reported with multi-temporal images (Diek et al., 2017; Fongaro et al., 2018; Loiseau et al., 2019; Tziolas et al., 2020; Demattê et al., 2020), they may not be suitable to study soil properties such as organic matter content, which can vary in short periods due to soil management (Macias and Chesworth, 1992). This is confirmed by our results, where the best performances for soil organic matter were observed using the single-date L8-OLI.

Single-date images from Landsat satellites are the most commonly used for soil mapping and similar results to those found here were reported (Demattê et al., 2007; Ducart et al., 2016; Forkuor et al., 2017; Zízala et al., 2019). Sentinel-2 and PlanetScope images, on the other hand, are still poorly used but the results reported in other studies were similar to ours (van der Werff and van der Meer, 2015; Poggio and Gimona, 2017; Gholizadeh et al., 2018; Gomez et al., 2018; Castaldi et al., 2019; Vaudour et al., 2019; Zízala et al., 2019). Multi-temporal images have been recently used to predict clay content in semiarid lands (Shabou et al., 2015), soil textural classes in tropical areas (Demattê et al., 2017), soil texture and organic carbon in Switzerland (Diek et al., 2017) and several soil properties over Europe (Safanelli et al., 2020), with similar results to those reported here.

The difference among soil property maps predicted by different satellites can be confirmed by their different spatial structure in the geostatistical analysis. For six, out of the seven soil properties studied, the use of the SYSI from Landsat-8 OLI satellite showed the highest semivariance values. Despite this, most of them (in five out of seven soil properties) presented the lowest values of the range parameter indicating high spatial variability and a strong spatial dependence. On the other hand, the range parameter values observed for the single-date PS image indicate the lowest spatial variability, although the spatial dependence for some soil properties was moderate. For all soil properties, the SYSI S2-MSI, which showed the best model performances, also presented a strong spatial dependence and one of the lowest values of semivariance.
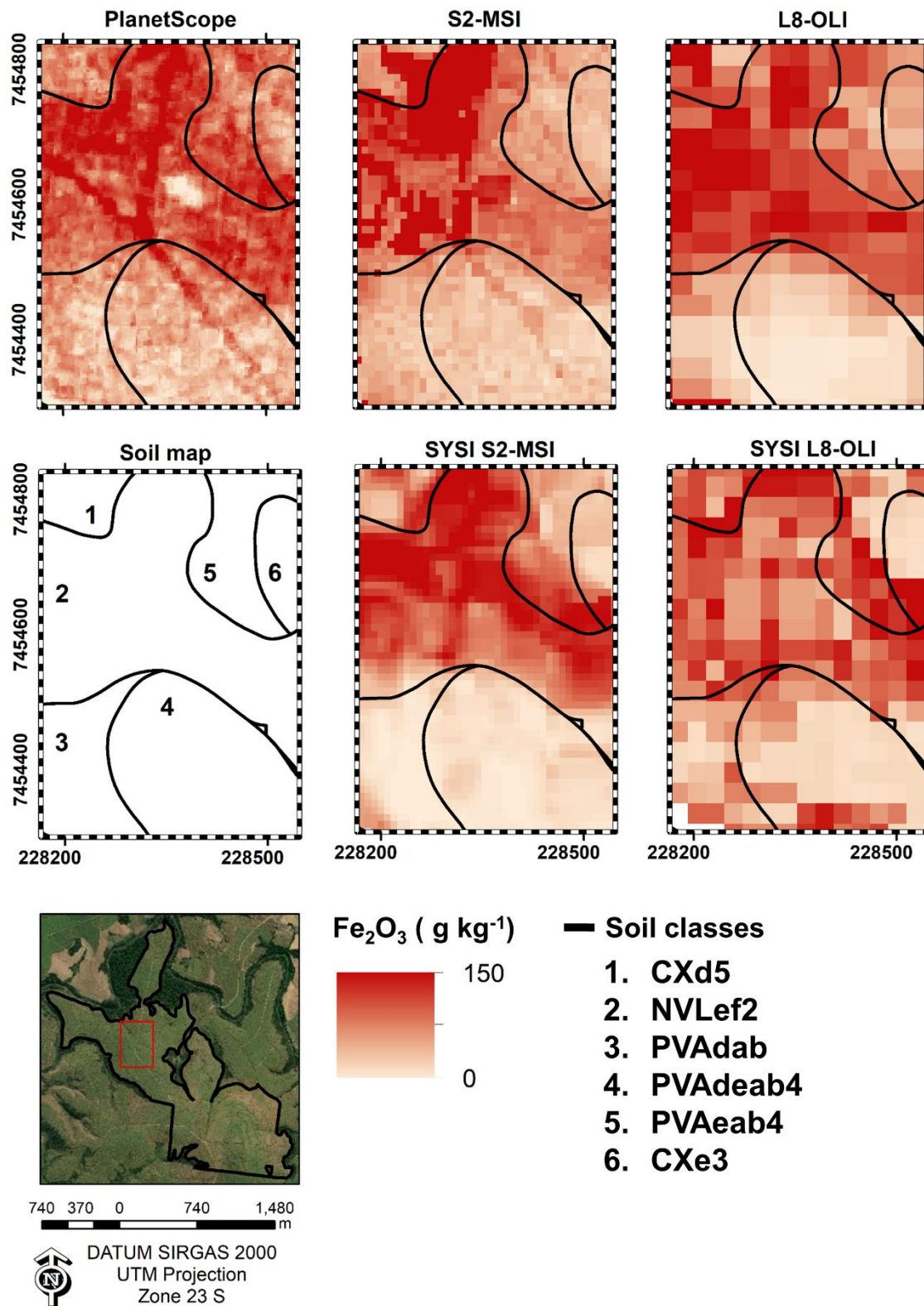
Figure 9. Predicted maps of iron content (Fe2O3) for each satellite. The soil classes were described as follows: 1) CXd5: Cambisol (Cambissolo Háplico distrófico muito arenoso), 2) NVLef2: Nitisol (Nitossolo Vermelho Latossólico eutroférrico argiloso), 3) PVAdab: Acrisol/Lixisol (Argissolo Vermelho Amarelo distrófico abrúptico), 4) PVAdeab4: Acrisol/Lixisol (Argissolo Vermelho Amarelo distrófico-eutrófico abrúptico textura arenosa), 5) PVAeab4: Acrisol/ Lixisol (Argissolo Vermelho Amarelo eutrófico abrúptico textura arenosa), 6) CXe3: Cambisol (Cambissolo Háplico eutrófico textura média).

2.4.3. Soil property maps as auxiliary tools to assist pedology and soil management

Through the analysis of predicted maps of soil properties (color, iron and clay content, Figs. 6–8) and their relationship with the legacy soil map, it was possible to note how satellite images and their different spatial, spectral and temporal resolutions can be useful to support and improve studies related to pedology and management.

Soil color, for example, is the main criterion in the Brazilian Soil Classification System to reach suborders of some soil classes (Santos et al., 2013). We found that the values of color components (hue, value and chroma) appeared slightly different within each soil class and this information can help to improve the delineation of soil groups and to reduce the associated uncertainty of soil color determination by pedologists.

Table 5. Mean of soil properties for each soil class and satellite

Soil class	CXe3	CXd5	NVLef2	PVAdab	PVAdeab4	PVAeadb4
Soil Colour (hue value/chroma)						
PS	3.5YR 4.7/3.5	3.5YR 4.6/3.6	0.8YR 3.9/3.6	3.5YR 4.6/3.2	4.7YR 4.9/3.2	2.5YR 4.5/3.6
S2-MSI	5.3YR 5.0/3.3	3.4YR 4.5/3.2	9.6R 3.7/3.7	3.1YR 4.5/3.4	5.4YR 5.1/3.0	3.9YR 4.6/3.4
L8-OLI	3.9YR 4.8/3.3	2.7YR 4.5/3.2	0.2YR 3.8/3.7	2.3YR 4.4/3.4	5.2YR 5.2/3.0	2.8YR 4.6/3.5
SYSI S2-MSI	3.6YR 4.7/3.1	3.1YR 4.6/3.5	10R 3.8/3.7	3.4YR 4.8/3.4	4.6YR 5.0/3.0	2.5YR 4.4/3.3
SYSI L8-OLI	3.9YR 4.8/3.2	0.9YR 4.0/3.6	9.4R 3.7/3.7	2.2YR 4.4/3.4	4.2YR 4.9/3.1	1.6YR 4.2/3.5
Clay content (g kg^{-1})						
PS	270.2±63.7	228.3±107.8	371.1±82.5	214.6±78.7	189.7±60.3	316.5±61.0
S2-MSI	224.4±57.0	259.4±110.1	376.4±78.8	264.8±103.0	188.8±80.5	261.7±85.5
L8-OLI	210.4±62.9	289.4±115.4	363.7±81.5	292.9±86.7	133.8±64.9	241.7±58.6
SYSI S2-MSI	173.9±67.3	306.9±103.5	346.3±53.7	245.7±53.0	169.2±72.8	251.1±51.6
SYSI L8-OLI	136.7±96.6	365.8±192.3	397.7±86.2	292.7±119.0	189.6±103.0	245.6±101.0
Fe$_2$O$_3$ (g kg^{-1})						
PS	37.9±14.1	23.1±17.2	56.1±18.3	21.0±13.4	18.9±12.4	48.4±13.7
S2-MSI	36.8±9.4	52.5±43.3	109.4±46.0	49.8±28.4	20.6±19.5	52.4±16.9
L8-OLI	53.2±14.5	74.2±67.3	112.4±36.5	72.1±66.2	16.7±24.5	66.1±24.8
SYSI S2-MSI	38.1±31.3	46.2±34.4	97.7±37.6	15.6±17.4	16.2±18.4	63.9±37.0
SYSI L8-OLI	85.1±56.9	73.9±43.3	88.9±32.8	51.0±30.0	33.1±33.1	92.6±48.9

The red colors highlighting the mean and standard deviation correspond to the higher values observed for each soil property inside each soil class. CXd5: Cambisol (Cambissolo Haplico distrófico muito arenoso), NVLef2: Nitisol (Nitossolo Vermelho Latossólico eutroférrico argiloso), PVAdab: Acrisol/Lixisol (Argissolo Vermelho Amarelo distrófico abrúptico), PVAdeab4: Acrisol/Lixisol (Argissolo Vermelho Amarelo distrófico-eutrófico abrúptico textura arenosa), PVAeab4: Acrisol/Lixisol (Argissolo Vermelho Amarelo eutrófico abrúptico textura arenosa), CXe3: Cambisol (Cambissolo Haplico eutrofico textura média). PS, S2-MSI and L8-OLI correspond to single-date images from PlanetScope, Sentinel-2 and Landsat-8 satellites respectively. SYSI: multi-temporal bare soil images.

Soil texture was also variable within each soil mapping unit, depending on the satellite used. For example, the CXe3 soil class showed high values of clay content when the single-date PlanetScope image, compared to the ones observed for the other satellites. The information provided by satellite images in different resolutions could be of great importance to improve the delineation of soil mapping units, as texture and color are important diagnostic attributes for several soil classes (Santos et al., 2013). In this sense, we also evaluated the performance of the satellite based on the criteria of clay content (>350 g kg^{-1}) for the Nitisol class as an example. The results were different, depending on which satellite was used. Surprisingly, the clay content map predicted by the SYSI L8-OLI, which did not necessarily have the best performance, had 70% of the area that meets the criteria for Nitisol soil class. If we

consider the clay content map predicted by the single-date PlanetScope image, only 58% of the area delineated in the legacy soil map meet the criteria. The SYSI S2-MSI, which showed the best performance, in this case had the lowest percentage of the area that met the criteria of clay content for Nitisol. These examples indicate that the traditional delineation would be incorrect and could be improved using the clay content maps predicted here and confirmed with field surveys.

Table 7. Percentage of the area of Nitisol class that meet the criterion of clay content values.

NVLef2	PS	S2-MSI	L8-OLI	SYSI S2-MSI	SYSI L8-OLI
Clay content			% of the area		
=>350 g kg^{-1}	58.14	64.85	62.43	47.92	71.58
<350 g kg^{-1}	41.86	35.14	37.57	52.07	28.42

NVLef2: Nitisol (Nitossolo Vermelho Latossolico eutroferrico in Brazilian Soil Classification system). PS, S2-MSI and L8-OLI correspond to single-date images from PlanetScope, Sentinel-2 and Landsat-8 satellites respectively. SYSI: multitemporal images

The iron content is generally higher in tropical areas than in other regions and has important implications for soil classification and mapping (Stucki et al., 1985; Santos et al., 2013). In the SiBCS, for example, the iron content is used to define some classes at third categorical levels (Santos et al., 2013). In our study area, the spatial distribution of iron content was variable and important differences between satellites and soil classes were observed. As the spatial resolution decreases, the iron content tends to increase, which can cause differences in the definition of soil classes depending on the iron content and the satellite used. For example, the Nitisol class (NVLef2) was classified as eutroferric in the legacy soil map, which means that the iron content must be between 80 and 180 g kg^{-1}. Among the Fe_2O_3 maps, only that predicted by the single-date PlanetScope image, with mean and standard deviation of 56.1 ± 18.3, does not meet this criterion in the surficial layer.

Soil color, texture and iron content have also been reported as important soil properties to relate to soil drainage (Evans and Franzmeier, 1988; Malone et al., 2018) and mineralogical studies (Viscarra Rossel et al., 2010; Poppiel et al., 2020; Ramos et al., 2020), which are also important for soil classification systems. By analyzing soil color, texture and iron content maps, we confirmed that it is important to consider the influence of different satellites' resolution on soil classification and mapping, coupled with field surveys to obtain reliable soil spatial information.

In terms of soil management, information such as texture and iron content are important. Areas with high clay content are highly susceptible to compaction and areas with different values will impact soil management strategies. Also, the variability of iron content can influence the management of phosphorus fertilization and promotes different strategies to help and reduce the specific adsorption of this nutrient (Camargo et al., 2013). Both soil texture and iron content maps can also assist in the delineation of management zones for precision agriculture of liming and fertilization rate (Safari et al., 2013; Meyer et al., 2019).

2.5. Conclusions

Our study shows the importance of satellite images in different spatial, spectral and temporal resolution for digital soil mapping and applications on classification and management. For our specific study area under tropical conditions, the SYSI S2-MSI showed to be the best predictor in explaining the variability of the majority of the soil

properties. In addition, the analysis of the spatial structure of predicted maps revealed that satellites' resolutions have strong influence on the representation of soil property variability.

Traditional pedologists usually use relief inferences to delineate soil mapping units. Our results showed that the soil properties predicted by multi-temporal images with bare soil, indicated areas where the delineation of soil mapping units on legacy maps can be improved. Thus, each pixel (with a soil property content) can bring light to the gaps brought from inferences using only relief.

Soil property maps obtained from satellites with different spatial, spectral and temporal resolutions can provide information that can promote different strategies in terms of soil classification and management. In this sense, the selection of a satellite with specific resolution would depend on the objectives of soil mapping and the decision makers' strategies. Results indicate that it is expected that the best combination should be an image with high-spatial and spectral resolution, as the multitemporal technique will depend on the user's objective.

Acknowledgments

Appendix A. Supplementary data

Supplementary data to this article can be found online at: https://doi. org/10.1016/j.geoderma.2021.115089.

References

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

Bartholomeus, H., Kooistra, L., Stevens, A., van Leeuwen, M., van Wesemael, B., Ben-Dor, E., Tychon, B., 2011. Soil Organic Carbon mapping of partially vegetated agricultural fields with imaging spectroscopy. Int. J. Appl. Earth Obs. Geoinf. 13, 81–88. https://doi.org/10.1016/j.jag.2010.06.009.

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

Behrens, T., Schmidt, K., MacMillan, R.A., Viscarra Rossel, R.A. 2018. Multi-scale digital soil mapping with deep learning. Sci. Rep. 8, 15244

Breunig, F.M., Galvão, L.S., Dalagnol, R., Santi, A.L., Della Flora, D.P., Chen, S. 2020. Assessing the effect of spatial resolution on the delineation of management zones for smallholder farming in southern Brazil. Remote Sens. Appl. Soc. Environ. 19. https://doi.org/10.1016/j.rsase.2020.100325.

Buttafuoco, G.., Castrignano, A.., Gucci, G.., Lacolla, G.. and Luca, F. 2017. Geostatistical modelling of within-field soil and yield variability for management zones delineation: a case study in a durum wheat field. Precis. Agric. 18, 37–58. https://doi.org/10.1007/s11119-016-9462-9

Camargo, L.A., Marques, J., Pereira, G.T., Alleoni, L.R.F. 2013. Spatial correlation between the composition of the clay fraction and contents of available phosphorus of an Oxisol at hillslope scale. Catena 100, 100–106. http://dx.doi.org/10.1016/j.catena.2012.07.016

Casa, R., Castaldi, F., Pascucci, S., Pignatti, S. 2012. Potential of hyperspectral remote sensing for field scale soil mapping and precision agriculture applications. Ital. J. Agron. **7**, 331–336. https://doi.org/10.4081/ija.2012.e43

Castaldi, F., Chabrillat, S., Don, A. and van Wesemael, B. 2019. Soil organic carbon mapping using LUCAS topsoil database and Sentinel-2 data: An approach to reduce soil moisture and crop residue effects. Remote Sens. 11, 2121. https://doi.org/10.3390/rs11182121

Chabrillat, S., Ben-Dor, E., Cierniewski, J., Gomez, C., Schmid, T., Van Wesemael, B. 2019. Imaging spectroscopy for soil mapping and monitoring. Surv. Geophys. 40, 361–399. https://doi.org/10.1007/s10712-019-09524-0.

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

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

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., Sayão, V.M., Rizzo, R., Fongaro, C.T. 2017. Soil class and attribute dynamics and their relationship with natural vegetation based on satellite remote sensing. Geoderma, 302, 39–51. https://doi.org/10.1016/j.geoderma.2017.04.019

Demattê, J.A.M., Alves, M.R., Terra, F. da S., Bosquilia, R.W.D., Fongaro, C.T. and Barros, P.P. da S. (2016). Is it possible to classify topsoil texture using a sensor located 800 km away from the surface? Rev. Bras. Ciência do Solo 40, e0150335. https://doi.org/10.1590/18069657rbcs20150335.

Demattê, J.A.M., Bellinaso, H., Romero, D.J., Fongaro, C.T. 2014b. Morphological Interpretation of Reflectance Spectrum (MIRS) using libraries looking towards soil classification. Sci. Agric. 71, 509–520. 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, J., Zullo Jr, J. 2007. Quantification of tropical soil attributes from ETM +/LANDSAT-7 data. Int. J. Remote Sens. 28, 3813–3829. https://doi.org/10.1080/01431160601121469

Diek, S., Fornallaz, F., Schaepman, M.E., Jong, R. De, Diek, S., Fornallaz, F., Schaepman, M.E., De Jong, R. 2017. Barest pixel composite for agricultural areas using Landsat time series. Remote Sens. 9, 1245. https://doi.org/10.3390/rs9121245

Ducart, D.F., Moreira Silva, A., Labouré, C., Toledo, B., Mozer De Assis, L. 2016. Mapping iron oxides with Landsat-8/OLI and EO-1/Hyperion imagery from the Serra Norte iron deposits in the Carajás Mineral Province, Brazil. Brazilian J. Geol. 46, 331–349. https://doi.org/10.1590/2317-4889201620160023

Evans, C. V., Franzmeier, D.P. 1988. Color index values to represent wetness and aeration in some Indiana soils. Geoderma 41, 353–368. https://doi.org/10.1016/0016-7061(88)90070-5

Fang, Q., Hong, H., Zhao, L., Kukolich, S., Yin, K., Wang, C. 2018. Visible and Near-Infrared reflectance spectroscopy for investigating soil mineralogy: A review. J. Spectrosc. 1-14. https://doi.org/10.1155/2018/3168974

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

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

Gholizadeh, A., Žižala, D., Saberioon, M., Borůvka, L. 2018. Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging. Remote Sens. Environ. 218, 89–103. https://doi.org/10.1016/j.rse.2018.09.015.

Gomez, C., Adeline, K., Bacha, S., Driessen, B., Gorretta, N., Lagacherie, P., Roger, J.M., Briottet, X. 2018. Sensitivity of clay content prediction to spectral configuration of VNIR/SWIR imaging data, from multispectral to hyperspectral scenarios. Remote Sens. Environ. 204, 18–30. https://doi.org/10.1016/j.rse.2017.10.047.

Gomez, C., Oltra-Carrió, R., Bacha, S., Lagacherie, P., Briottet, X. 2015. Evaluating the sensitivity of clay content prediction to atmospheric effects and degradation of image spatial resolution using Hyperspectral VNIR/SWIR imagery. Remote Sens. Environ. 164, 1–15. https://doi.org/10.1016/j.rse.2015.02.019.

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

Guo, L., Shi, T., Linderman, M., Chen, Y., Zhang, H., Fu, P. 2019. Exploring the influence of spatial resolution on the digital mapping of soil organic carbon by airborne hyperspectral VNIR imaging. Remote Sens. 11, 1032. https://doi.org/10.3390/rs11091032.

Hengl, T. 2006. Finding the right pixel size. Comput. Geosci. 32, 1283–1298. https://doi.org/10.1016/j.cageo.2005.11.008.

Hong, S.Y., Minasny, B., Han, K.H., Kim, Y., Lee, K., 2013. Predicting and mapping soil available water capacity in Korea. PeerJ e71. https://doi.org/10.7717/peerj.71.

IUSS Working Group WRB. 2015. World reference base for soil resources. http://www.fao.org.

Kravchenko, A.N. 2003. Influence of spatial structure on accuracy of interpolation methods. Soil Sci. Soc. Am. J. 67, 1564–1571. https://doi.org/10.2136/sssaj2003.1564.

Kuhn, M., Weston, S., Keefer, C., Coulter, N., Quinlan, R. 2018. Package "Cubist." 13. https://github.com/topepo/Cubist/issues

Kuhn, M., Johnson, K., 2013. Applied Predictive Modeling, Springer. Springer, New York. https://doi.org/10.1007/978-1-4614-6849-3

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/10.1016/j.jag.2019.101905.

Macias, F., Chesworth, W. 1992. Weathering in humid regions, with emphasis on igneous rocks and their metamorphic equivalents. Dev. Earth Surf. Process. 2, 283–306. https://doi.org/10.1016/B978-0-444-89198-3.50017-9

Malone, B.P., McBratney, A.B., Minasny, B. 2011. Empirical estimates of uncertainty for mapping continuous depth functions of soil attributes. Geoderma, 160, 614–626. https://doi.org/10.1016/j.geoderma.2010.11.013.

Malone, B.P., Mcbratney, A.B., Minasny, B. 2018. Description and spatial inference of soil drainage using matrix soil colours in the Lower Hunter Valley, New South Wales, Australia. PeerJ 6, e4659. https://doi.org/10.7717/peerj.4659

Main-Knorn, M, B, P., Louis, J., Debaecker, V., Muller-Wilm, U., Gascon, F., 2017. Sen2Cor for Sentinel-2, in: Image and Signal Processing for Remote Sensing. p. 12. https://doi.org/10.1117/12.2278218.

Marais Sicre, C., Fieuzal, R., Baup, F. 2020. Contribution of multispectral (optical and radar) satellite images to the classification of agricultural surfaces. Int. J. Appl. Earth Obs. Geoinf. 84, 101972.

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

Meyer, S., Kling, C., Vogel, S., Schröter, I., Nagel, A., Kramer, E., Gebbers, R., Philipp, G., Lück, K., Gerlach, F., Scheibe, D., Ruehlmann, J. 2019. Creating soil texture maps for precision liming using electrical resistivity and gamma ray mapping. In: Precision Agriculture 2019 - Papers Presented at the 12th European Conference on Precision Agriculture, ECPA 2019, pp. 539–546. Wageningen Academic Publishers.

Nussbaum, M., Ettlin, L., Çöltekin, A., Suter, B., Egli, M., 2011. The relevance of scale in soil maps. Bull. BGS 63–70.

Ose, K., Corpetti, T., Demagistri, L., 2016. Multispectral Satellite Image Processing, in: Baghdadi, N., Zribi, M. (Eds.), Optical Remote Sensing of Land Surface: Techniques and Methods. Elsevier Inc., Oxford, pp. 58–124. https://doi.org/10.1016/B978-1-78548-102-4.50002-8

Planet Labs. 2020. Planet Imagery Specifications.

Poggio, L., Gimona, A. 2017. Assimilation of optical and radar remote sensing data in 3D mapping of soil properties over large areas. Sci. Total Environ. 579, 1094–1110. https://doi.org/10.1016/j.scitotenv.2016.11.078

Poppiel, R.R., Lacerda, M.P.C., Rizzo, R., Safanelli, J.L., Bonfatti, B.R., Silvero, N.E.Q., Demattê, J.A.M. 2020. Soil color and mineralogy mapping using proximal and remote sensing in Midwest Brazil. Remote Sens. 12, 1197. https://doi.org/10.3390/rs12071197

Poppiel, R.R., Lacerda, M.P.C., Demattê, J.A.M., Oliveira, M.P., Gallo, B.C., Safanelli, J.L. 2019. 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

Quinlan, J.R. (1992). Learning wth continuous classes, in: Proceedings AI'92, 5th Australian Conference on Artificial Intelligence.World Scientific, pp. 343–348.

R Core Team, 2019. R: A Language and Environment for Statistical Computing.

Ramcharan, A., Hengl, T., Nauman, T., Brungard, C., Waltman, S., Wills, S., Thompson, J., 2018. Soil Property and Class Maps of the Conterminous United States at 100-Meter Spatial Resolution. Soil Sci. Soc. Am. J. 82, 186–201. https://doi.org/10.2136/sssaj2017.04.0122

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

Reyes, J., Wendroth, O., Matocha, C., Zhu, J., 2019. Delineating Site-Specific Management Zones and Evaluating Soil Water Temporal Dynamics in a Farmer's Field in Kentucky. Vadose Zo. J. 18, 1–19. https://doi.org/10.2136/vzj2018.07.014.

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

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

Romero, D.J., Ben-Dor, E., Demattê, J.A.M., Souza, A.B., Vicente, L.E., Tavares, T.R., Martello, M., Strabeli, T.F., Barros, P.P.S., Fiorio, P.R., Gallo, B.C., Sato, M.V., Eitelwein, M.T. 2018. Internal soil standard method for the Brazilian soil spectral library: Performance and proximate analysis. Geoderma 312, 95–103. https://doi.org/10.1016/j.geoderma.2017.09.014.

Safanelli, J.L., Chabrillat, S., Ben-Dor, E., Demattê, J.A.M. 2020. Multispectral models from bare soil composites for mapping topsoil properties over Europe. Remote Sens. 12, 1369. https://doi.org/10.3390/rs12091369.

Safari, Y., Esfandiarpour Boroujeni, I., Kamali, A., Salehi, M.H., Bagheri Bodaghabadi, M. 2013. Mapping of the soil texture using geostatistical methods (a case study of the Shahrekord plain, central Iran). Arab. J. Geosci. 6, 3331–3339. https://doi.org/10.1007/s12517-012-0559-9.

Sahwan, W., Lucke, B., Kappas, M., Bäumler, R. 2018. Assessing the spatial variability of soil surface colors in northern Jordan using satellite data from Landsat-8 and Sentinel-2. European Journal of Remote Sens. 51, 850–862. https://doi.org/10.1080/22797254.2018.1502624

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

Santos, H.G., Jacomine, P.K.T., Anjos, L.H.C., Oliveira, V.A., Oliveira, J.B., Coelho, J.F. and Cunha, T.J.F. (2013). Sistema brasileiro de classificação de solos.

Schmidt, G.L., Jenkerson, C.B., Masek, J., Vermote, E., and Gao, F., 2013, Landsat ecosystem disturbance adaptive processing system (LEDAPS) algorithm description: U.S. Geological Survey Open-File Report 2013–1057, 17 p.

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

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

Smith, S., Bulmer, C., Flager, E., Frank, G., Filatow, D. 2010. Digital soil mapping at multiple scales in British Columbia, Canada. Program and Abstracts, 4th Global Workshop on Digital Soil Mapping, 17.

Stucki, J.W., Goodman, B.A., Schwertmann, U. 1985. Iron in Soils and Clay Minerals, Soil Science. https://doi.org/10.1097/00010694-198904000-00013.

Teixeira, P.C., Donagema, G.K., Fontana, A., Teixeira, W.G. 2017. Manual de Métodos de Análise de Solo. https://www.embrapa.br.

Tziolas, N., Tsakiridis, N., Ben-Dor, E., Theocharis, J., Zalidis, G. 2020. Employing a multi-input deep convolutional neural network to derive soil clay content from a synergy of multi-temporal optical and radar imagery data. Remote Sens. 12, 1389. https://doi.org/10.3390/rs12091389

Vaudour, E., Gomez, C., Fouad, Y., Lagacherie, P. 2019. Sentinel-2 image capacities to predict common topsoil properties of temperate and mediterranean agroecosystems. Remote Sens. Environ. 223, 21–33. https://doi.org/10.1016/j.rse.2019.01.006.

Viscarra Rossel, R.A., Bui, E.N., De Caritat, P., Mckenzie, N.J. 2010. Mapping iron oxides and the color of Australian soil using visible–near-infrared reflectance spectra. J. Geophys. Res. 115, 4031. https://doi.org/10.1029/2009JF001645

Viscarra Rossel, R.A., Minasny, B., Roudier, P., McBratney, A.B. 2006. Colour space models for soil science. Geoderma 133, 320–337. https://doi.org/10.1016/j.geoderma.2005.07.017

Walkley, A., Black, I.A. 1934. An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. Soil Science 37, 29–38.

van der Werff, H., van der Meer, F. 2015. Sentinel-2 for mapping iron absorption feature parameters. Remote Sens. 7, 12635–12653. https://doi.org/10.3390/rs71012635.

Zhao, Z., Chow, T.L., Yang, Q., Rees, H.W., Benoy, B., Xing, Z., Meng, F.R., 2008. Model prediction of soil drainage classes based on digital elevation model parameters and soil attributes from coarse resolution soil maps. Can. J. Soil. Sci 88, 787–799

Žížala, D., Minařík, R., Zádorová, T. 2019. Soil organic carbon mapping using multispectral remote sensing data: Prediction ability of data with different spatial and spectral resolutions. Remote Sens. 11, 2947. https://doi.org/10.3390/rs11242947.

3. SOIL VARIABILITY AND QUANTIFICATION BASED ON LANDSAT AND SENTINEL MULTITEMPORAL IMAGES: A COMPARISON

Abstract

There is a worldwide need for detailed spatial information to support soil mapping, mainly in the tropics, where main agricultural areas are concentrated. In this line, satellite images are useful tools that can assist in obtaining soil information from a synoptic point of view. This study aimed at evaluating how satellite images at different resolutions (spatial, spectral and temporal) can influence the representation of soil variability over time, the percentage of bare soil areas and spatial predictions of soil properties in southeastern Brazil. We used single-date and multi-temporal images (SYSI, Synthetic Soil Images) of bare soil pixels from the Sentinel2-MultiSpectral Instrument (S2-MSI) and the Landsat-8 Operational Land Imager (L8-OLI) to conduct this research. Two SYSIs were obtained from images acquired in four years (2016–2019) for each satellite (SYSI S2-MSI and SYSI L8-OLI) and a third SYSI, named SYSI Combined, was obtained by combining the images from both satellites. The single-date images for each satellite were acquired in September, when the influence of clouds was low and bare soil pixels was predominant. Single-date images and SYSIs were compared by means of their spectral patterns and ability to predict topsoil properties (clay, sand, silt, and organic matter contents and soil color) using the Cubist algorithm. We found that the SYSIs outperformed single-date images and that the SYSI Combined and SYSI L8- OLI provided the best prediction performances. The SYSIs also had the highest percentage of areas with bare soil pixels (~30–50%) when compared to the single-date images (~20%). Our results suggest that bare soil images obtained by combining Landsat-8 and Sentinel-2 images are more important for soil mapping than spatial or spectral resolutions.

3.1. Introduction

Maps of soil properties are essential to understand soil's productive potential and, consequently, are valuable tools for many agricultural and environmental issues. In this sense, soil maps have been obtained by digital mapping techniques, in which mathematical and statistical tools are used to explain soils' behavior by relating its properties to environmental variables. As traditional soil mapping techniques have historical limitations related to financial and time constraints, digital soil mapping (DSM) emerged as an alternative to meet the demand for soil maps around the world (Lagacherie et al., 2007; McBratney et al., 2003; Minasny and McBratney, 2016; McBratney et al., 2019; Hengl et al., 2017; Hengl et al., 2014; Viscarra Rossel et al., 2016; EMBRAPA, 2016).

Among the tools available for mapping soil properties, satellite images have been of significant importance, as they provide a synoptic view of the area under study (Grunwald et al., 2011; Mulder et al., 2011; de Padilha et al., 2020) and allow monitoring soil behavior over time (Fathololoumi et al., 2020). Landsat missions are the most commonly used satellites to study soil properties. The first works used multiple linear regression and geostatistical analysis to relate soil properties such as clay content, soil color and organic matter to Landsat spectral bands (Escadafal et al., 1989; Mathieu et al., 1998; Mattikalli, 1997; Ishida and Ando, 1999; Demattê et al., 2007; Liao et al., 2013). More recently, studies have been using the latest Landsat8-OLI and sophisticated nonlinear algorithms to improve predictions of soil properties (de Arruda et al., 2016; Forkuor et al., 2017).

Although Landsat missions are still widely used, recently launched satellites such as the Sentinel-2 MultiSpectral Instrument (S2-MSI), which has improved characteristics in terms of spatial, spectral and temporal

resolutions (Helder et al., 2018), are offering new opportunities to enhance soil studies but their use is still restricted to the European region. The works of van der Werff and van der Meer (2015) and Castaldi et al. (2016) were the first studies that reported the use of the spectral responses of Sentinel-2 to estimate iron features, soil texture and organic carbon. The authors used, respectively, hyperspectral images and spectral libraries to simulate Sentinel-2 bands and found encouraging results ($R^2 > 0.5$ from Castaldi et al., 2016). Poggio and Gimona (2017) also used Sentinel-2 images coupled with MODIS, Landsat and radar data to estimate soil properties and their associated uncertainty over Scotland while Gomez et al. (2018) studied the sensitivity of clay content prediction using spectral configurations from multispectral to hyperspectral sensors, including Sentinel-2. Later, Castaldi et al. (2019a) evaluated the capability of Sentinel-2 to predict soil organic carbon (SOC) and found that this satellite was adequate to describe the variability of SOC at field and regional scales.

The aforementioned studies of Landsat and Sentinel-2 satellites used data simulated from spectral libraries or single-date images for soil studies. To use single-date images, clear-sky conditions and areas without vegetation or straw are required to achieve suitable results (Blasch et al., 2015; Roberts et al., 2019). In this sense, recent studies have been developing methodologies to obtain bare soil mosaics or synthetic images with bare soil pixels for use in soil mapping. Examples that used Landsat images are the Barest Pixel Composite (Diek et al., 2017), the GeoSpatial Soil Sensing System (GEOS3) (Demattê et al., 2020; Demattê et al., 2018) and the Soil Composite Processor (SCMaP) (Rogge et al., 2018), which were applied to several studies and proved to be promising tools to improve mapping of topsoil properties (Diek et al., 2017; Fongaro et al., 2018; Gallo et al., 2018; de Mendes et al., 2019; Žížala et al., 2019; Poppiel et al., 2020; Rizzo et al., 2020; Safanelli et al., 2020) and correlated areas such as geology (Bonfatti et al., 2020).

For Sentinel-2 satellites, no consolidated methodology was reported to obtain bare soil information from time series, but some works applied methods similar to those described above for Landsat. For example, Castaldi et al. (2019b) tested several thresholds of the NBR2 (Normalized Burn Ratio2) spectral index to apply to Sentinel-2 images and predict SOC over North-eastern Germany. Loiseau et al. (2019) followed the same principle to map clay content over mainland France but using NDVI (Normalized Difference Vegetation Index) thresholds to obtain bare soil mosaics and Žížala et al. (2019) used the SCMap to obtain bare soil composites to study SOC variability in Czech Republic. The most recent research is from Tziolas et al. (2020) who also used NDVI thresholds to obtain bare soil pixels from Sentinel-2 images of a three-year period and combined it with radar information to predict clay content, with promising results ($R^2 = 0.60$).

Such researches certainly enrich the literature related to the use of satellite images for soil studies, but a comprehensive analysis that highlights the main advantages of using bare soil images over single-date ones is still missing. Besides that, a possible combination of images from both satellites in order to obtain more bare soil pixels can be useful to improve soil studies and needs to be explored. In this research, we aim to assess the potential of multi-temporal bare soil images (Synthetic Soil Image, SYSI) obtained from Landsat-8 and Sentinel-2 satellites (singly and combined) to predict topsoil properties (clay, sand, silt, organic matter, and color) using the Cubist algorithm and to compare their performances with those obtained from single-date images. Our hypothesis is that SYSI provides more continuous soil reflectance values than single-date images and that the combined use of Landsat-8 and Sentinel-2 satellites can offer even more bare soil pixels, which can improve soil mapping in tropical regions.

3.2. Material and Methods

3.2.1. Study area and soil dataset

The study area comprises 4814.62 km² within the State of São Paulo, southeastern Brazil (Fig. 1) and is characterized by a humid subtropical climate, with hot summers and dry winters (Cwa), according to the Koppen climatic classification (Alvares et al., 2013). With altitudes varying from 441 to 661 m above sea level, the region is characterized by a wide variability in parent materials such as siltstones, shales, argillites, and sandstones formed during the Paleozoic era. The geological complexity developed a wide range of soils such as Ferralsols, Lixisols, Nitisols, Cambisols, Arenosols, and Gleysols (IUSS Working Group WRB, 2015). Most of the area is used for sugarcane production and pasture.

We used 1500 soil observations obtained at 0–0.20 m depth from soil surveys performed between 2010 and 2019. These soil samples were oven-dried at 45 °C for 24 h, grounded and sieved to 2 mm, and then analyzed in the laboratory for chemical and physical determination (Teixeira et al., 2017). The organic matter content was determined by the Walkley Black oxidation method (Walkley and Black, 1934) and the sand, silt and clay contents were obtained by the pipette method. The *soiltexture* package (Moeys, 2018) was used to classify the soil samples into nine textural classes, according to the Soil Taxonomy, in R software (R Core Team, 2019).

The laboratory soil spectra were obtained from Bellinaso et al. (2010), who used the FieldSpec Pro spectroradiometer (Analytical Spectral Devices, Boulder, Colo.) to acquire the reflectance spectra from the visible, near and shortwave-infrared regions (vis-NIR-SWIR, 350–2500 nm). The spectra from the visible region (350–780 nm) were used to calculate the Munsell soil color components (hue, value and chroma) by applying color matching functions and transformation among color spaces, as described by Viscarra Rossel et al. (2006). The formula and steps for soil color calculations are described in the supplementary material.
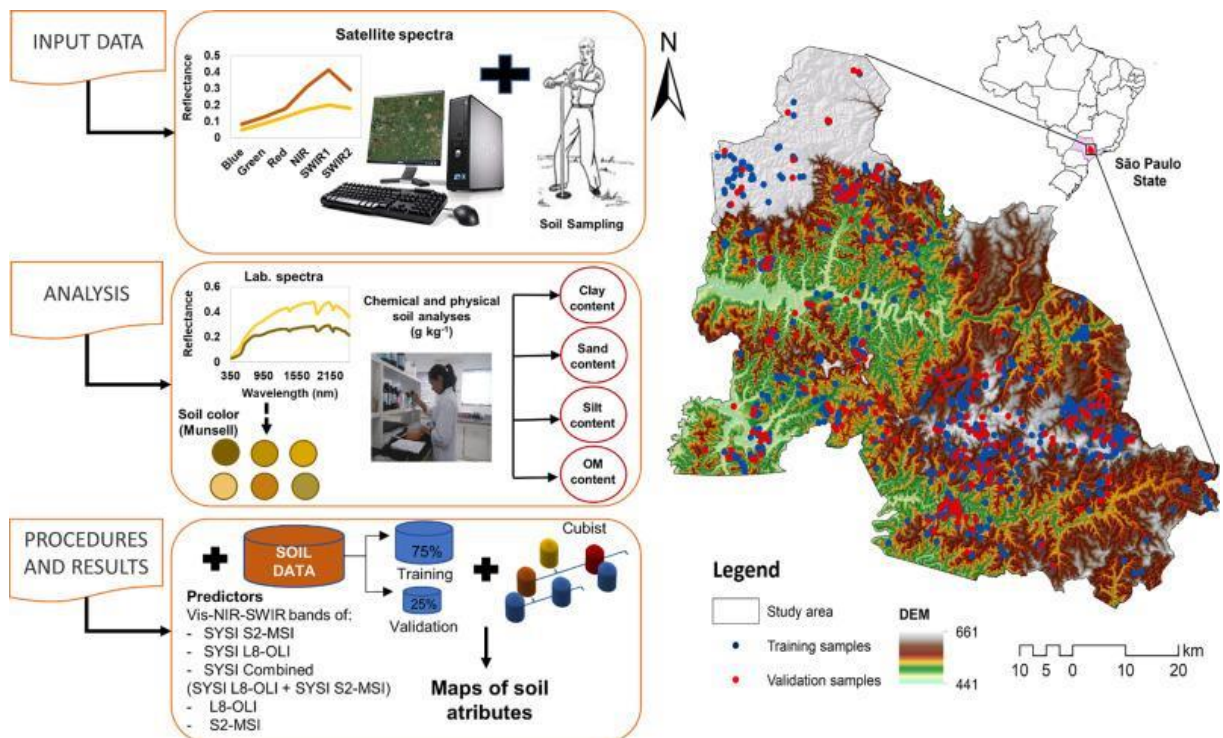
Figure 1. Location of the study area, soil observations and methodology flowchart.

3.2.2. Satellite data and multitemporal bare soil technique

3.2.2.1. Satellite data

Landsat8-OLI and Sentinel2-MSI satellite data, as surface reflectance products, were used for this work. The Landsat-8 Operational Land Imager (L8-OLI) was launched by the National Aeronautics and Space Administration (NASA) in 2013 (Markham et al., 2015) and is the last satellite in a series of missions that began in 1972, commonly used for Earth observation and study. Its main characteristics are the spatial resolution of 30 m and the revisit time of the same point every 16 days. It has nine spectral bands that cover the visible (coastal, blue, green and red), near-infrared (NIR), shortwave infrared (SWIR1 and SWIR2) and thermal infrared regions (see Table S1 in the supplementary material for details of the spectral ranges).

The Sentinel-2 MultiSpectral Instrument (S2-MSI) mission was launched by the European Space Agency (ESA) in 2015. Two satellites (2A and 2B) were placed in orbit, which combined provide approximately five-days revisit time. Differently from Landsat8- OLI, Sentinel2-MSI has spatial resolutions ranging from 10 to 60 m and thirteen spectral bands that cover the visible (coastal, blue, green and red), near-infrared (B8 and NIR), red-edge (RedEdge1, RedEdge2, and RedEdge3), shortwave-infrared (SWIR), and water and cirrus regions (see Table S1 in the supplementary material for more details), which made it very popular for the study of the land.

3.2.2.2. Single-date and multi-temporal bare soil images for each satellite

Multi-temporal bare soil images, named SYSI (Synthetic Soil Image) were obtained by applying the Geospatial Soil Sensing System method (GEOS3) (Demattê et al. (2018) to Landsat8-OLI and Sentinel2-MSI images. We used 45 L8-OLI images from 2016 to 2019, which have at least 80% of the area without clouds and shadows (see Fig. S1 and Table S2 in the supplementary material for details). The L8-OLI images were processed by the LasRC algorithm (Vermote et al., 2016) and are available as surface reflectance products (TIER 1) on the Google Earth Engine platform (GEE) (Gorelick et al., 2017).
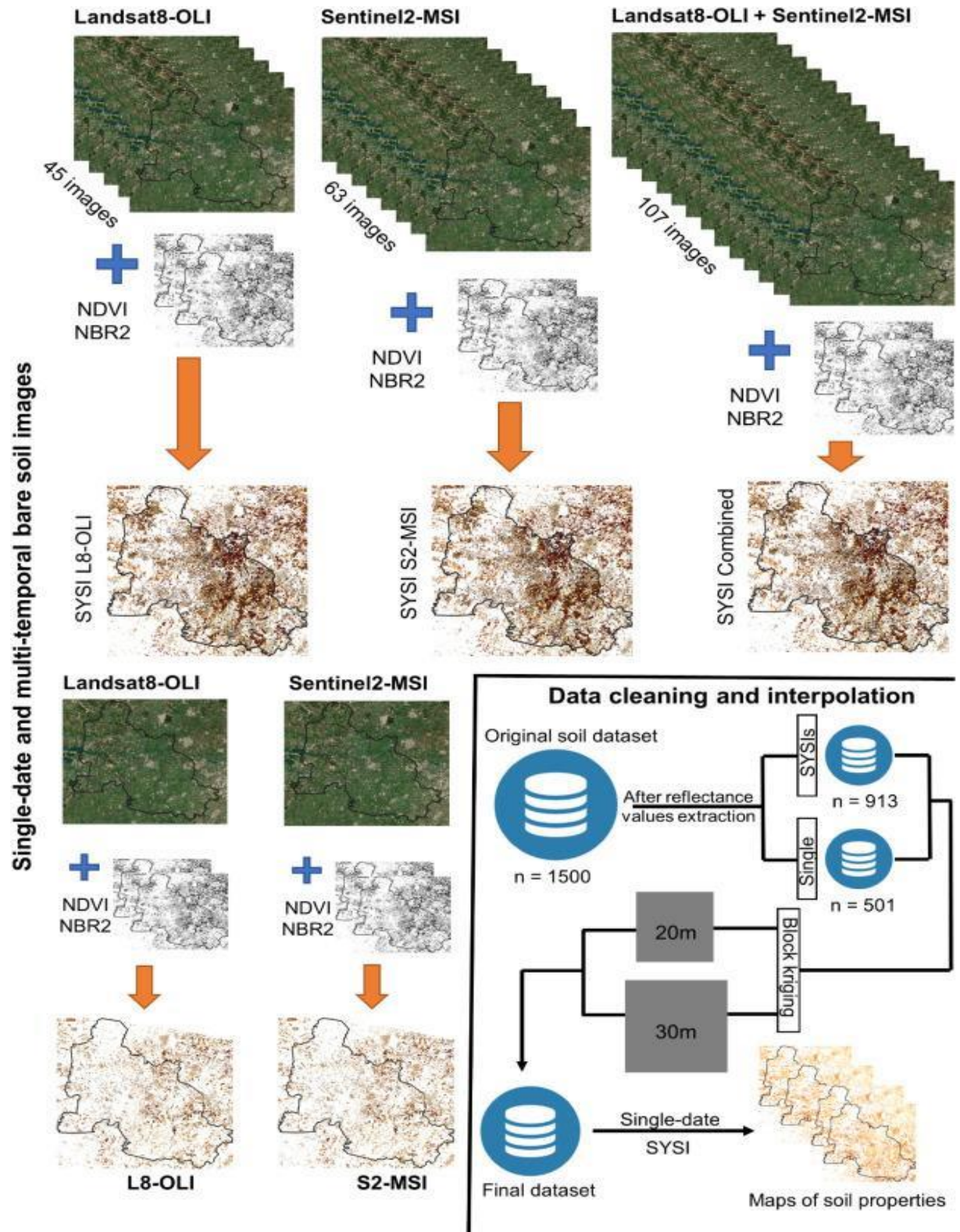
Figure 2. Detailed flowchart of the procedures for obtaining bare soil areas and data cleaning and interpolation.

The S2-MSI images were not available as surface reflectance products (Level 2a) in the GEE platform at the time this work was carried out (in GEE, the surface reflectance images for Brazil have been available since December 2018) and therefore, 63 Level-1C (Top of Atmosphere reflectance, TOA) images from 2016 to 2019 were downloaded from the Copernicus website with the *getSpatialData* package

(https://github.com/16EAGLE/getSpatialData) in R software. The images were chosen using as criterion the cloud coverage (< 20%) and were atmospherically corrected and converted to surface reflectance using the Sen2Cor plugin (Main-Knorn et al., 2017) adapted to R software. The spatial resolution of S2-MSI images ranges from 10 to 60 m but we used spectral bands with 10 and 20 m pixel sizes. The NIR bands (band 8A) were available in 10 m pixel size and were resampled to 20 m using the resample function and the bilinear method in R software, in order to ensure the same spatial extent of all spectral bands. The SYSI from L8-OLI images (SYSI L8-OLI) was obtained in the GEE platform, since the images were available and ready for use.

The SYSI from S2-MSI images (SYSI S2-MSI) was obtained in R software, due to the unavailability of surface reflectance for our study area, mentioned before. The GEOS3 method consists of (1) obtaining the images, (2) masking pixels that were non-bare soil, and (3) calculating the median reflectance of the masked images. After obtaining the images, spectral indices such as NDVI (Normalized Difference Vegetation Index) and NBR2 (Normalized Burn Ratio 2) were calculated and used to mask the images (Fig. 2). Following the GEOS3 method, the range between 0 and 0.25 for NDVI were considered as bare soil and the values above this threshold were flagged as NA. NBR2 values > 0.075 were also flagged as NA, and corresponded to sites covered with straw or burned. The quality masks provided with each image were used to exclude areas with clouds and shadows. Finally, the masked images were ordered by date and the median reflectance was calculated for each satellite. Two SYSIs, with nine and six spectral bands were obtained, named SYSI S2- MSI and SYSI L8-OLI, respectively (Fig. 2), which were later used as predictors of soil properties.

Single-date Landsat8-OLI and Sentinel-2 images acquired in 09/09/2019 were also downloaded from the GEE platform, as surface reflectance products. Both images did not present either cloud nor shadows and the pixels with vegetation, straw and burned were excluded using the NDVI and NBR2 thresholds, as described earlier for SYSI images (Fig. 2). They were used later to carry our predictions of soil attributes.

3.2.2.3. Multi-temporal bare soil from Landsat and sentinel data combined

A third SYSI was obtained by combining L8-OLI and S2-MSI images, in order to assess the gains (or losses) of bare soil pixels and its potential for predicting topsoil properties (Fig. 2). We denoted it as SYSI Combined. We used respectively 45 and 62 images from L8-OLI and S2-MSI satellites (for details about the dates and images used, the reader is referred to Table S2 in the supplementary material), totaling 107 images. We excluded one S2-MSI image because it was from the same date as Landsat-8 (October 2019) and we decided to maintain that from Landsat since we had a lower number of images from this satellite. The S2-MSI images have spatial resolution of 20 m and, to group with L8-OLI, they were resampled to 30 m, using the bilinear method and the resample function in R software. The SYSI Combined was also obtained by applying the GEOS3 method, but in R software, due to the unavailability of S2-MSI images in surface reflectance for the period studied and code adaptation problems in the GEE platform. The result was a SYSI Combined with six spectral bands covering the visible (red, green and blue), near-infrared (NIR), and the shortwave-infrared (SWIR 1 and SWIR2) regions. The RedEdge bands from S2-MSI images were excluded from the SYSI Combined as the L8-OLI images do not have these spectral bands and the bare soil areas of the RedEdge would represent only those from S2-MSI images.

3.2.2.4. Number of images required for bare soil pixels

It is interesting to define an optimal number of images to obtain at least 90% of the area with bare soil pixels, similar as performed by Diek et al. (2017) on the Swiss Plateau, to find out how many years of satellite images would be needed to obtain continuous soil reflectance values. However, as the Sentinel-2 satellite has only been available since 2015 and the influence of clouds in tropical regions, we agree that it would be difficult to fulfill this condition. To find out how the percentage of bare soil areas would be affected by clouds and the period of the year, we obtained four SYSIs. The first and second SYSIs were obtained using images with 20% cloud and from the dry (May to September) and moist season (October to April), respectively. The third and fourth SYSIs were also obtained separately for dry and moist seasons but without considering the cloud coverage as a filter. We calculated the percentage of bare soil pixels in each to understand how image quality and period of the year can affect bare soil pixel retrieval from tropical regions instead of defining an optimal number of images. These SYSIs were not used in predictions and were used only to explain the limitations of bare soil images in tropical regions.

3.2.3. Statistical analysis

3.2.3.1. Assessment of reflectance spectra, bare soil percentage, and soil properties

We matched the lab soil spectra with those from satellites by convolving the first to L8-OLI and S2-MSI spectral regions, using their relative spectral response functions as described in Zhang et al. (2018). The spectral patterns of simulated spectra and those from the three SYSIs and single-date images were qualitatively analyzed and compared. The quality of the bare soil of each SYSI and single-date images was evaluated by the soil line analysis (Baret et al., 1993) to show if the GEOS3 method, originally developed for Landsat data, could effectively provide bare soil information from Sentinel-2 data and from the images combined. For that, the reflectance values of NIR and Red bands were plotted on a XY plot and regression analyses were carried out among them. The closer the scatter points to the 1:1 line, the greater the coefficient of determination (R^2) and, therefore, the method was able to accurately detect bare soil areas. The percentage of bare soil in each SYSI and single-date images was also measured by calculating the number of pixels with bare soil and dividing it by the total number of pixels in the area.

To confirm the spectral similarities among S2-MSI and L8-OLI spectral bands (see Table S1 in the supplementary material), we performed a canonical correlation analysis only between the three SYSIs. The spectral bands of each SYSI were reduced to separate canonical variables through a multivariate linear transformation (Webster, 1977). The greater the canonical correlations, the more similar were the spectral signatures among the SYSIs. The procedure for obtaining the canonical correlations was carried out in R software with the CCA package (González et al., 2008).

The soil properties were analyzed, first by checking the normality distribution and the presence of outliers and then, by studying their relationships to SYSIs and single-date spectral bands by Pearson's correlation analysis.

3.2.3.2. Soil dataset cleaning and interpolation

The dataset used contained 1500 soil observations and due to the different number of pixels with bare soil in each image, the dataset was reduced to 1415, 1258, 913, 525 and 501 soil observations for SYSI Combined, SYSI S2-MSI, SYSI L8-OLI and single-dates S2-MSI and L8-OLI images, respectively. For a better comparison among the predictions' ability of SYSI and single-date images, we used the lowest number of soil samples to perform predictions (i.e., 913 and 501 soil observations, respectively).

To account for the within-pixel variability, we decided to use block kriging to estimate pixel average values of soil properties from point measurements. Block kriging consists in the estimation of a linear average for an attribute within supports, which is defined as the shape, size and orientation of the volume associated with any observation (Olea, 1999). In our case, we considered two supports or block sizes: 30 and 20 m, which correspond to the pixel sizes of Landsat8-OLI and Sentinel2-MSI, respectively (Fig. 2). To conduct block kriging on soil samples, we first calculated the semivariance, which is based on the intrinsic hypothesis and principle of regionalized variables (Isaaks and Srivastava, 1989), following the formula:

$$\gamma(h) = \frac{1}{2n(h)} \sum_{i=1}^{N(h)} [z(x_i) - z(x_i + h)]^2$$

where, $\gamma(h)$ is the semivariance; $n(h)$ is the number of pairs of points separated by the distance h; $z(x_i)$ is the value of z in the position, and $z(x_i + h)$ is the value of z in the position $x_i + h$. Second, the kriging estimate was performed following the formula:

$$\hat{z}(x_0) = \sum_{i=1}^{N} \lambda_i\, z(x_i)$$

where $\hat{z}(x_0)$ is the kriging estimator at the x0 point; $z(x_i)$ represents the measured soil property value at the xi point; and λ_i is the kriging weight attributed to closest $z(x_i)$ values to estimate $\hat{z}(x_0)$. The estimated values then represented the average values of soil properties at pixel sizes of 20 and 30 m.

3.2.3.3. Prediction of soil properties

The soil datasets were randomly partitioned into training (75%) and validation (25%) using the *createDataPartition* function from the caret package in R software. The predictions of soil properties using SYSI images had 684 and 229 soil samples for training and validation, respectively; while for single-date images the number of samples for training and validation was 376 and 125. The predictions of soil properties using single-date and SYSI images from Sentinel-2 satellite were performed following two approaches: a) using nine (vis-RedEdge-NIR-SWIR) and b) six spectral bands (vis-NIR-SWIR) as predictors. To predict topsoil properties using SYSI L8-OLI and SYSI Combined, we used their six spectral bands (vis-NIR-SWIR) as predictors. In addition, for soil color components (hue, value, and chroma), a third approach was considered, using the three visible bands (blue, green, and red), since soil color is more related to this portion of the electromagnetic spectrum (Odeh and McBratney, 2005).

The Cubist method was used to calibrate regression tree models using the train function of the caret package. Cubist is a rule-based algorithm whose main characteristics came from Quinlan's M5 model (Kuhn et al., 2018; Quinlan, 1992). Rule-based models are characterized by being easy to interpret (Kuhn and Johnson, 2013) and, unlike other decision trees algorithms, such as Random Forest, Cubist uses linear regression models at each node instead of the average, to avoid overfitting (Kuhn and Johnson, 2013). In addition, the algorithm uses a boosting-like method called committees, which controls the number of model's trees that are sequentially created and the most k common neighbors or instances to build the trees (Khaledian and Miller, 2020). The Cubist algorithm does not use all predictors to construct the rules and build the trees (Khaledian and Miller, 2020) and provide only the most important for predictions. We used the default number of committees (1, 10 and 20) and neighbors (0, 5, and 9) from the train function, which together gave us nine models. The root mean square error (RMSE) was used to select the best models.

3.2.3.4. Assessment of model performances and spatial structure analysis

The coefficient of determination (R2train), the root mean square error (RMSEtrain) and the mean absolute error (MAE), which were provided with the Cubist model results, were used to evaluate the performance of the training set's prediction models. The performance of the validation sets was evaluated by R^2val, RMSEval, bias, concordance correlation coefficient (CCC) and mean square error (MSE), which were calculated using the goof function of the ithir package in R software. The maps of soil properties were obtained using the predict function and SYSI and single-date images in R software.

After predictions and performance assessments were carried out, we selected, presented, and discussed those that had the best performances in training and validation sets. Afterwards, we assessed the spatial structure of predicted values by analyzing their semivariogram parameters, which measured the spatial dependence of soil properties (Gomez et al., 2012). We used the semivariograms obtained in section 2.3.2, when the interpolation of the observed values was performed. Then, we calculated the semivariograms of values predicted by SYSI and single-date images, using the whole dataset (913 and 501 soil samples).

3.3. Results

3.3.1. Bare soil images and spectral pattern descriptions

Single-date and SYSI bare soil images are shown in Fig. 3. Among them, the SYSI Combined had the highest percentage of pixels with bare soil (45.57%), followed by the SYSI S2-MSI (43.93%) and the SYSI L8- OLI (31.89%). The single-date images had the lowest percentages, but the S2-MSI image had almost 2% more areas with bare soil than the L8- OLI image. We found that the best in representing bare soil patterns in the study area was the SYSI L8-OLI, which has the highest coefficient of determination (R2: 0.94) in the soil line analysis, while the single-date L8-OLI images has the lowest (R2: 0.70) (Fig. 3). The canonical correlation between SYSIs showed that the SYSI Combined and SYSI S2-MSI has the highest coefficient (0.94) (Fig. 4).

We were unable to define an optimal number of images to obtain at least 90% of bare soil pixels in our study area, due to the short period studied (four years) and other factors such as the influence of clouds and shadows in the images. In Fig. 5 we are showing why it would be difficult to define an optimal number of images to obtain the highest number of pixels with bare soil in tropical regions. With the images acquired from the dry season only, we found that the percentage of bare soil pixels was similar for Landsat data using either 34 or 19 images, mainly due to the influence of clouds and shadows. The same can be observed for Sentinel-2, from which we obtained 48 (20% cloud) and 81 (without cloud percentage criterion) images in the dry season and the percentage of bare soil pixels only differed in 1%. From the moist season we obtained more images than from the dry season, but the percentage of bare soil pixels from both satellites were low (< 15%).
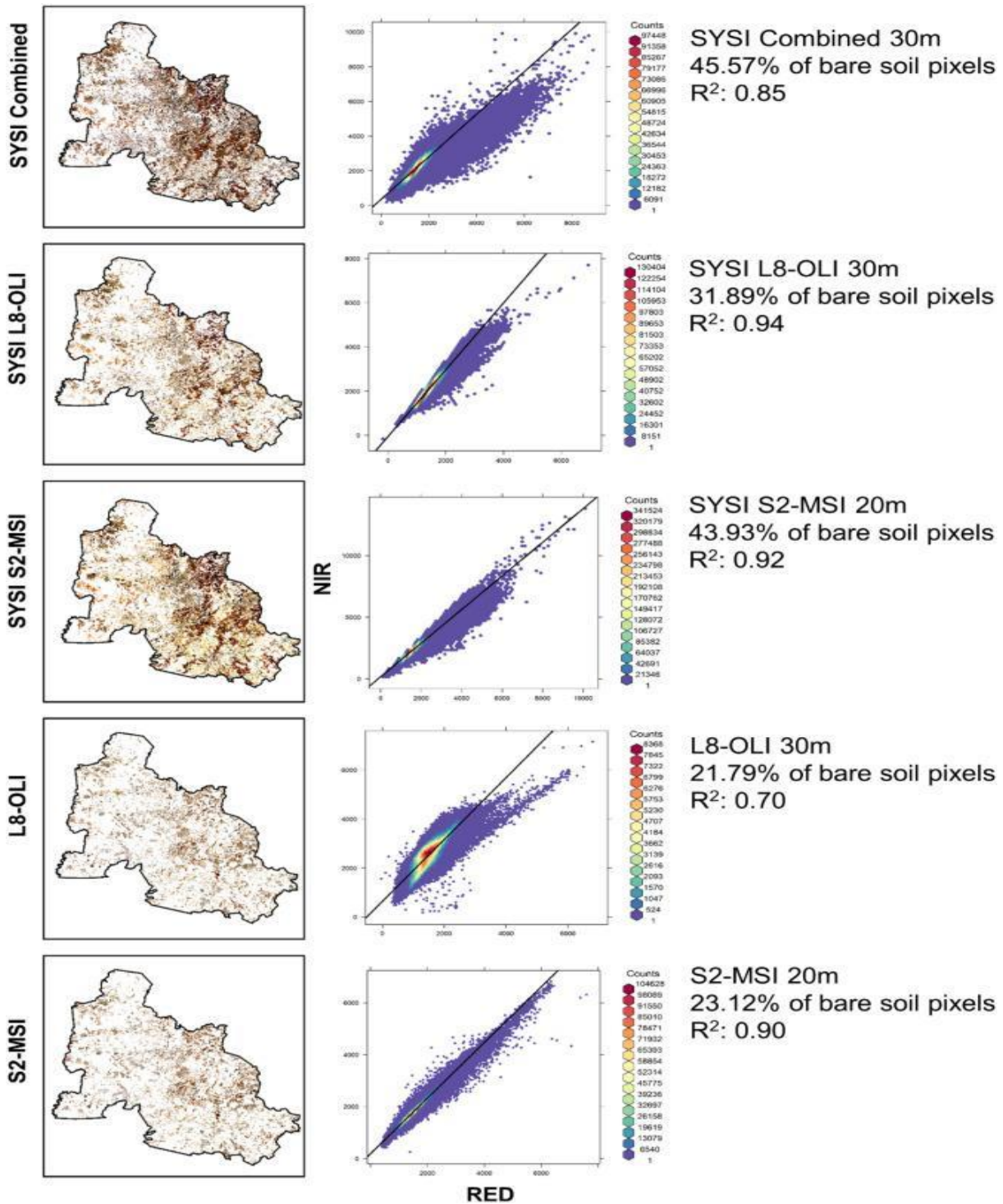
Figure 3. True color composition (RGB: 432) of SYSI and single-date images. The SYSI L8-OLI and SYSI S2-MSI correspond to the bare soil images obtained by using images of four years from Landsat8-OLI and Sentinel2-MSI satellites, respectively. The SYSI Combined was obtained by combining the images from both satellites, after resampling Sentinel-2 MSI images to 30 m. Both single-date L8-OLI and S2-MSI images were acquired on 09/09/2019. The percentage values correspond to areas with bare soil obtained by applying the GEOS3 method. Relationships between NIR and Red bands are presented to show the soil line for each image. The value of the coefficient of determination (R^2) was used as an indication of bare soil retrieval

The spectral patterns of satellite images and lab spectra (this last simulated the spectral responses of Sentinel and Landsat) are presented (Fig. 6). We found that, for simulated S2-MSI and L8-OLI spectra, the reflectance values increase as the clay content decreases (Fig. 6a-b) and that their spectral patterns appear very similar

(Fig. 6c), excluding the Red-Edge bands, which are present only in Sentinel-2 satellites. Two contrasting textural classes (clay and sandy), among the nine presented, were used to compare the simulated spectra with those extracted from satellite images (Fig. 6d-f). The sandy textures have reflectance values higher than clay, but the differences between single-date and SYSIs are more prominent in the clay texture and in the NIR-SWIR region (Fig. 6d). Simulated and satellite spectra are also shown, separately for each textural class (Fig. 6e-f). In almost all spectral regions, simulated lab spectra of sandy texture have higher reflectance values than satellite images, especially in the NIR-SWIR region and those from clay texture are more similar to SYSI than to single-date images.
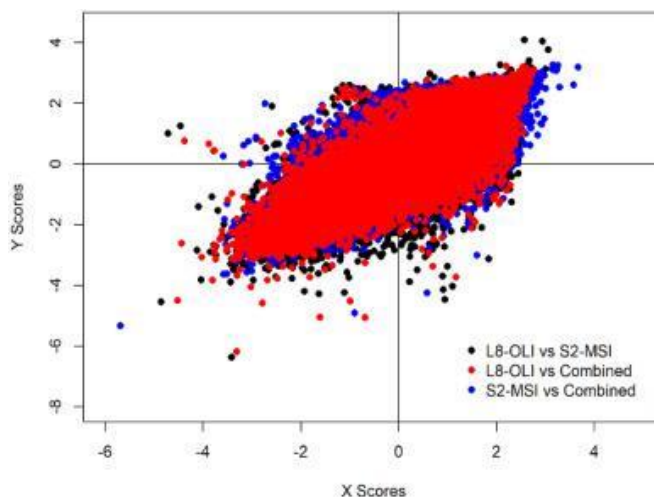
Figure 4. Canonical correlation analysis between the three SYSIs.

3.3.2. Descriptive statistics of soil properties and correlation with bare soil images

The distribution and descriptive statistics of the soil datasets used for single-date and SYSI images are shown in Fig. 7. For the first dataset (n = 913), the mean values of soil color components are 5.85YR 4.14/ 3.98 (hue = 15.85, value = 4.14 and chroma = 3.98), which were very similar to those observed in the second dataset (n = 501) and that denotes the predominance of yellow-brown soil color in the study area. Clay, sand, and silt contents were highly variable in the first, and the sand content was dominant with mean and standard deviation of 503.78 $\pm$ 236.40 g kg^{-1}. Some values of clay content > 600 g kg^{-1} were also found. In the second dataset, high mean values of sand content were observed (633.23 g kg^{-1}) while clay and silt contents were lower than the values observed in the first dataset. In both datasets, the organic matter content (OM) was low, with average <2% and maximum up to 5% (Fig. 7).

All soil properties have higher correlation coefficients with SYSI than with single-date images and the SYSI Combined shows the highest correlations (Table 1). Soil color components, hue and value, and sand content have positive correlation values with satellite's bands while chroma, clay, silt and OM content show negative correlations. All soil properties, except OM content and chroma component, show suitable correlation coefficients with the RedEdge bands of the SYSI S2-MSI but the correlations are low with those from the single-date S2-MSI.
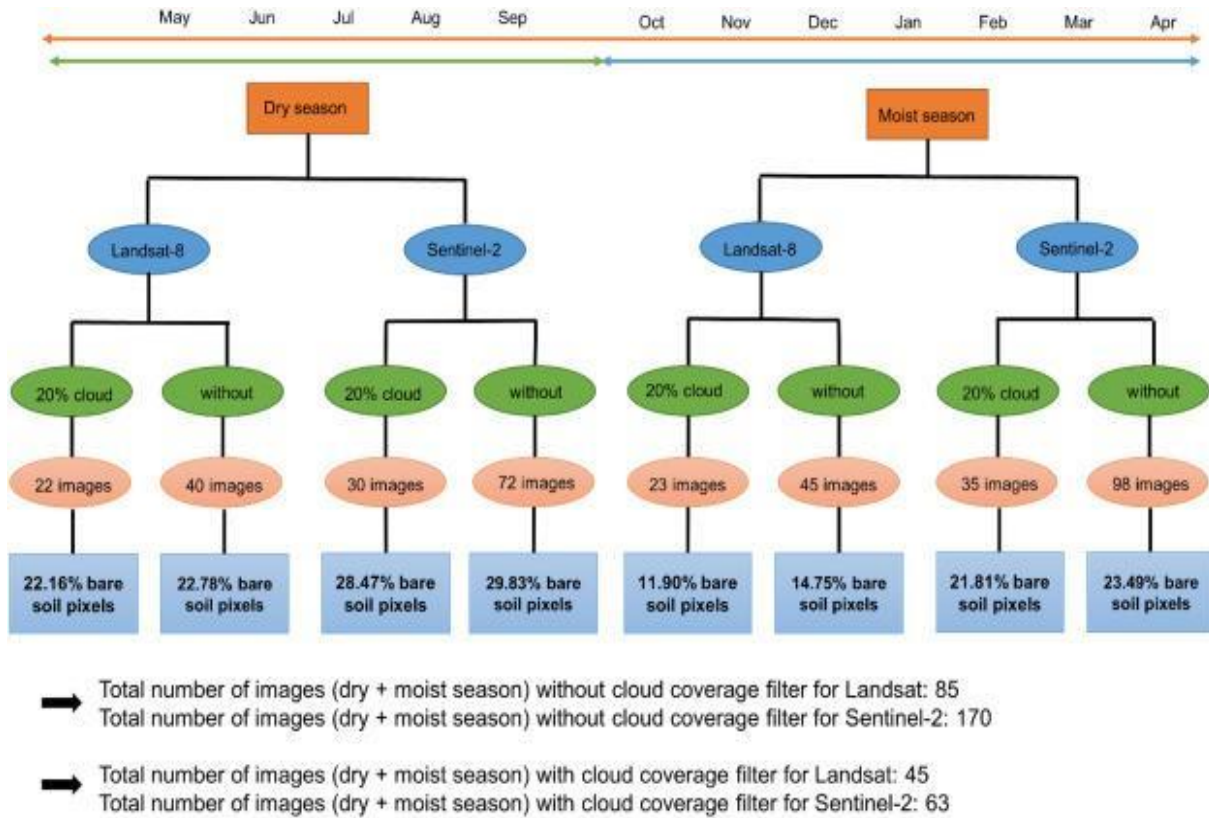
Figure 5. Number of images and percentage of bare soil pixels for dry and moist seasons for each satellite. 20% and without corresponding to the criterion applied to select the images. 20%: only images with 20% cloud coverage, without: no criterion of cloud was considered.

3.3.3. Prediction performances

In Table 2 we present the best performances for each image and soil property. Other performances are described in Tables S3 and S4 in the supplementary material. We found that six spectral bands from the vis-NIR-SWIR region are sufficient to describe soil properties variability and that the RedEdge bands of Sentinel-2 satellite did not improve the performances. The performances are better with SYSIs than with single-date images, and the differences are considerable. The prediction's performances for training and validation sets revealed that the SYSI Combined and SYSI L8-OLI are the best predictors for almost all soil properties, except silt content (Table 2). Chroma, silt and OM content have the lowest performances ($R^2 < 0.5$).

The SYSI Combined explained 72% of the clay content variability in the training set (R^2train:0.72). There is no significant difference between performances of clay content predicted by SYSI S2-MSI and SYSI L8-OLI but for the single-date images, the performances were low (Table 2). In the validation set, the SYSI Combined also has the highest performance (R^2val: 0.74). The sand content was better predicted by the SYSI Combined and the silt content has one of the lowest performances (R^2train: 0.34). Unlike the clay and sand content, the performance of the validation set is low for the silt content. The performance of the OM content is also low (R^2train: 0.31), but is slightly higher in the validation set (R^2val > 0.36).

Unlike the other soil properties, to predict soil color components we used only the observed values because it was impossible to fit theoretical models to their semivariograms and the interpolation by block kriging was impossible to perform. The SYSI L8-OLI is the best predictor of soil color components (R^2train: 0.57, 0.59, 0.35 for hue, value and chroma, respectively), but the performances in the validation set were low. We found that the six

spectral bands in the vis-NIR-SWIR region are more suitable to describe soil color variability than only those from the visible region (see Tables S3 and S4 in the supplementary material).
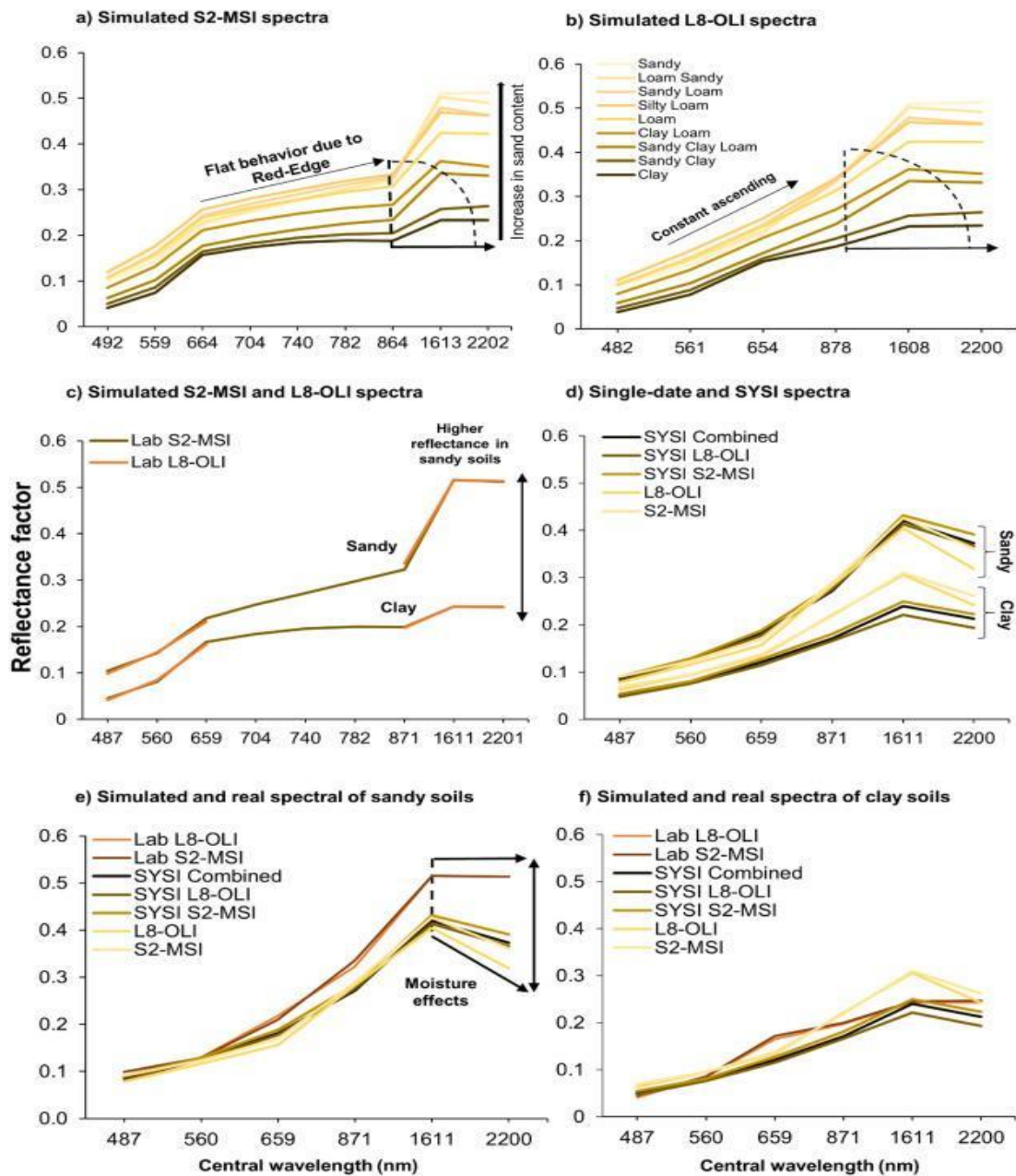
Figure 6. Spectral curves of lab spectra simulated to a) Sentinel2-MSI and b) Landsat8-OLI with nine and six bands, respectively. c) Comparison between simulated S2-MSI and L8-OLI spectral bands. Note that L8-OLI has six bands, d) spectral curves of SYSIs and single-date images, e) and f) simulated and real spectral curves for sandy and clay soils, respectively. The x-axis for c-f were averaged using the central wavelengths of both satellites.

3.3.4. Predicted soil maps and spatial structure analysis

The spatial distribution maps of clay, sand and OM contents and soil color components predicted by SYSI and single-date images are illustrated in Figs. 8 and 9. Silt content maps are not shown as they had the lowest performances. The predicted maps show very similar spatial patterns regardless of the differences in bare soil areas and spatial resolution. Both SYSI and single-date images were able to represent the variability of soil properties, but the mean and standard deviation are slightly different between them. Clay content map predicted by SYSI

Combined, which had the best performance, shows high values of mean and standard deviation in contrast to the other predicted maps. The OM content map shows the same behavior but the sand content is lower than the other SYSIs and single-date images (Fig. 8). Predicted maps of hue and value based on the SYSI Combined (Fig. 9), show slightly lower values of mean and standard deviation than the other images while for chroma, the mean values are very similar among the images.
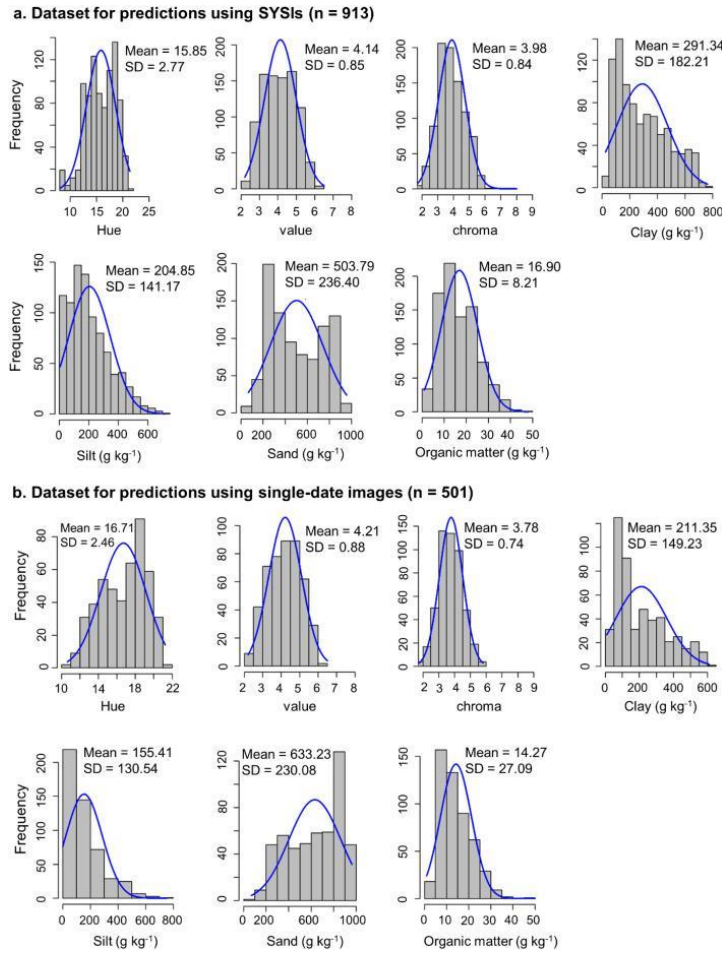
Figure 7. Histograms of soil datasets used. Hue, value and chroma are dimensionless. Hue was normalized to numeric values as follow: <10 = 10R, 10–12.5 = 2.5YR, 12.5–15 = 5YR, 15–17.5 = 7.5YR, >20 = 10YR.

In Fig. 10, we show a small area to demonstrate the effect of bare soil areas on the predicted maps of hue and clay and sand contents. It can be noted that the differences in spatial coverage are considerable among single-date and SYSI images. As single-date L8-OLI and S2-MSI were acquired on the same date, the differences between them are not significant. Between maps based on SYSI images, we can describe some differences, especially for the hue component predicted by SYSI L8-OLI, which shows high values, indicating a yellower color, where the other SYSIs show low values, indicating a redder color.

The spatial distributions of predicted soil properties were similar, which can be confirmed by the similarity of their semivariogram parameters (Fig. 11). Most of the predicted soil properties had their experimental semivariograms described by exponential models, and in some cases, gaussian and spherical models were fitted (Table S5 in the supplementary material). We do not present the semivariograms for silt content, as they had the lowest performances.

The semivariograms of measured soil properties (black lines in Fig. 11) have the highest semivariance values, although the range parameters, which define the spatial variability, are very similar between the measured and the predicted soil properties (Table S5 in the supplementary material). The difference in the semivariance between observed and predicted values is even higher for OM content (Fig. 11c). Unfortunately, it was impossible to fit theoretical models for value component predicted by SYSI images and exponential models described the spatial patterns of value maps based on single-date L8-OLI and S2- MSI (Fig. 11e). For the chroma component, none of the tested models were fitted and reached inaccurate results, with very high range values (> 20,000 m) that are inconsistent with the result obtained for other soil properties.

Table 1 Pearson's correlation analysis between SYSIs spectral bands and soil attributes

SYSIs	Bands	Hue	Value	Chroma	Clay	Sand	Silt	OM
					--------------------g kg^{-1}---------------------			
SYSI Combined	BLUE	0.68	0.67	-0.37	-0.70	0.50	-0.09	-0.52
	GREEN	0.67	0.69	-0.32	-0.68	0.46	-0.06	-0.51
	RED	0.59	0.62	-0.21	-0.66	0.49	-0.11	-0.50
	NIR	0.63	0.64	-0.26	-0.73	0.56	-0.17	-0.53
	SWIR1	0.62	0.57	-0.29	-0.80	0.67	-0.28	-0.57
	SWIR2	0.61	0.58	-0.32	-0.81	0.69	-0.29	-0.59
SYSI L8-OLI	BLUE	0.65	0.67	-0.37	-0.67	0.48	-0.10	-0.51
	GREEN	0.65	0.67	-0.29	-0.64	0.45	-0.07	-0.49
	RED	0.58	0.60	-0.19	-0.60	0.45	-0.12	-0.47
	NIR	0.61	0.61	-0.23	-0.69	0.55	-0.20	-0.52
	SWIR1	0.62	0.56	-0.26	-0.76	0.65	-0.29	-0.56
	SWIR2	0.63	0.57	-0.28	-0.77	0.67	-0.30	-0.57
L8-OLI	BLUE	0.54	0.42	-0.34	-0.43	0.33	-0.10	-0.32
	GREEN	0.52	0.43	-0.27	-0.39	0.29	-0.07	-0.30
	RED	0.41	0.36	-0.21	-0.31	0.24	-0.08	-0.25
	NIR	0.38	0.31	-0.22	-0.46	0.41	-0.20	-0.35
	SWIR1	0.37	0.28	-0.26	-0.52	0.49	-0.27	-0.38
	SWIR2	0.48	0.39	-0.35	-0.55	0.50	-0.25	-0.40
SYSI S2-MSI	BLUE	0.54	0.60	-0.19	-0.62	0.47	-0.13	-0.46
	GREEN	0.63	0.66	-0.36	-0.66	0.48	-0.11	-0.48
	RED	0.62	0.67	-0.30	-0.65	0.45	-0.08	-0.47
	RedEdge1	0.53	0.60	-0.19	-0.63	0.48	-0.14	-0.46
	RedEdge2	0.53	0.61	-0.19	-0.64	0.49	-0.15	-0.47
	RedEdge3	0.54	0.62	-0.20	-0.65	0.51	-0.16	-0.48
	NIR	0.56	0.55	-0.28	-0.76	0.65	-0.29	-0.54
	SWIR1	0.56	0.57	-0.31	-0.78	0.67	-0.30	-0.55
	SWIR2	0.57	0.62	-0.24	-0.70	0.56	-0.20	-0.50
S2-MSI	BLUE	0.33	0.21	-0.20	-0.41	0.29	-0.05	-0.32
	GREEN	0.28	0.17	-0.13	-0.38	0.25	-0.02	-0.30
	RED	0.14	0.06	-0.04	-0.28	0.19	-0.02	-0.24
	RedEdge1	0.12	0.05	-0.04	-0.32	0.23	-0.04	-0.26
	RedEdge2	0.10	0.02	-0.03	-0.39	0.30	-0.09	-0.33
	RedEdge3	0.09	0.02	-0.04	-0.40	0.31	-0.10	-0.33
	NIR	0.09	0.00	-0.06	-0.42	0.34	-0.12	-0.34
	SWIR1	0.08	-0.04	-0.07	-0.48	0.42	-0.20	-0.37
	SWIR2	0.20	0.08	-0.15	-0.53	0.45	-0.19	-0.39

OM: organic matter content

Table 2 Model performances of soil attributes for calibration and validation datasets

Soil attributes	NB	C	N	RMSE$_{train}$	R²$_{train}$	MAE	R²$_{val}$	RMSE$_{val}$	CCC	MSE	Bias
					Clay content						
SYSI Combined	**Six**	**20**	**0**	**56.99**	**0.72**	**44.12**	**0.74**	**55.68**	**0.85**	**3101.17**	**1.19**
SYSI L8-OLI	Six	10	0	62.80	0.66	47.50	0.68	61.44	0.81	3775.28	-3.58
L8-OLI	Six	20	0	111.97	0.42	85.97	0.58	108.39	0.68	11750.16	-21.80
SYSI S2-MSI	Six	20	0	62.20	0.67	48.69	0.70	59.41	0.83	3529.42	-1.98
S2-MSI	Six	20	0	116.36	0.38	87.53	0.42	121.09	0.59	14663.98	-25.06
					Sand content						
SYSI Combined	**Six**	**20**	**0**	**153.23**	**0.57**	**115.66**	**0.58**	**153.37**	**0.75**	**23523.08**	**18.29**
SYSI L8-OLI	Six	10	9	121.82	0.50	125.33	0.54	158.92	0.71	25254.84	9.63
L8-OLI	Six	20	0	188.87	0.34	147.76	0.48	180.63	0.60	32627.16	42.48
SYSI S2-MSI	Six	20	0	160.13	0.53	121.95	0.52	165.38	0.71	27351.72	24.48
S2-MSI	Six	20	0	193.66	0.34	150.15	0.29	200.88	0.51	40354.57	27.15
					Silt content						
SYSI Combined	Six	1	9	117.23	0.25	86.27	0.18	131.19	0.36	17210.46	-1.28
SYSI L8-OLI	Six	10	9	119.52	0.22	89.39	0.21	128.47	0.37	16505.35	-2.77
L8-OLI	Six	20	0	123.11	0.17	86.77	0.20	122.03	0.32	14892.65	-36.97
SYSI S2-MSI	**Nine**	**20**	**9**	**79.52**	**0.34**	**60.70**	**0.20**	**128.53**	**0.34**	**16520.52**	**2.65**
S2-MSI	Six	20	0	128.18	0.15	90.81	0.09	123.32	0.24	15207.40	-27.74
					OM content						
SYSI Combined	**Six**	**10**	**0**	**5.58**	**0.35**	**4.18**	**0.38**	**6.11**	**0.53**	**37.37**	**-1.08**
SYSI L8-OLI	Six	10	0	5.69	0.32	4.29	0.36	6.20	0.50	38.40	-1.20
L8-OLI	Six	20	0	6.39	0.21	4.74	0.27	5.92	0.42	35.13	-1.84
SYSI S2-MSI	Nine	10	0	5.75	0.31	4.39	0.38	6.12	0.51	37.47	-1.23
S2-MSI	Nine	20	0	6.56	0.19	4.82	0.22	6.10	0.36	37.26	-1.20
					Hue						
SYSI Combined	Six	20	0	1.66	0.55	1.29	0.56	1.66	0.71	2.75	-0.03
SYSI L8-OLI	**Six**	**20**	**9**	**1.64**	**0.57**	**1.28**	**0.54**	**1.72**	**0.71**	**2.94**	**-0.10**
L8-OLI	Six	20	0	1.71	0.50	1.38	0.67	1.43	0.78	2.05	0.11
SYSI S2-MSI	Nine	10	9	1.75	0.50	1.35	0.55	1.68	0.72	2.83	-0.17
S2-MSI	Nine	20	0	1.82	0.49	1.40	0.60	0.49	0.74	2.22	-0.21
					value						
SYSI Combined	Six	20	9	0.59	0.56	0.47	0.60	0.56	0.76	0.31	-0.04
SYSI L8-OLI	**Six**	**10**	**9**	**0.57**	**0.59**	**0.46**	**0.55**	**0.59**	**0.72**	**0.35**	**-0.01**
L8-OLI	Six	20	0	0.70	0.31	0.56	0.41	0.68	0.60	0.47	0.01
SYSI S2-MSI	Nine	10	9	0.58	0.56	0.46	0.51	0.62	0.70	0.39	-0.01
S2-MSI	Nine	20	0	0.72	0.37	0.58	0.42	0.67	0.59	0.45	-0.07
					chroma						
SYSI Combined	Six	10	0	0.67	0.32	0.49	0.32	0.66	0.48	0.43	-0.03
SYSI L8-OLI	**Six**	**10**	**0**	**0.65**	**0.35**	**0.49**	**0.27**	**0.68**	**0.45**	**0.46**	**-0.04**
L8-OLI	Six	20	0	0.66	0.21	0.51	0.17	0.68	0.33	0.47	0.04
SYSI S2-MSI	Nine	10	0	0.65	0.35	0.48	0.34	0.65	0.51	0.42	-0.06
S2-MSI	Six	20	0	0.66	0.21	0.51	0.17	0.68	0.33	0.47	0.04

SYSI L8-OLI and SYSI S2-MSI correspond to bare soil images obtained from images of four years of Landsat-8 and Sentinel-2 satellites. SYSI Combined is the bare soil image obtained by combining the images from both satellites and L8-OLI and S2-MSI correspond to single-date Landsat and Sentinel-2 images acquired in 09/09/2019. NB: number of spectral bands used in predictions, C: committees or number of model trees or boosting iterations used, N: neighbors used for each iteration, RMSE$_{train}$: Root Mean Square Error of training set (g kg^{-1}), R²$_{train}$: coefficient of determination of training set, MAE: Mean Absolute error, RMSE$_{val}$: Root Mean Square Error of validation set (g kg^{-1}), R²$_{val}$: coefficient of determination of validation set, CCC: concordance correlation coefficient, MSE: Mean Square Error.

3.4. Discussions

3.4.1. Bare soil images retrieval

We have explored the use of single-date and multi-temporal bare soil images from two satellites in different spatial, spectral and temporal resolutions to predict topsoil properties in tropical environments. We have found that the SYSIs outperformed single-date images in bare soil pixel retrieval and predictions of soil properties.
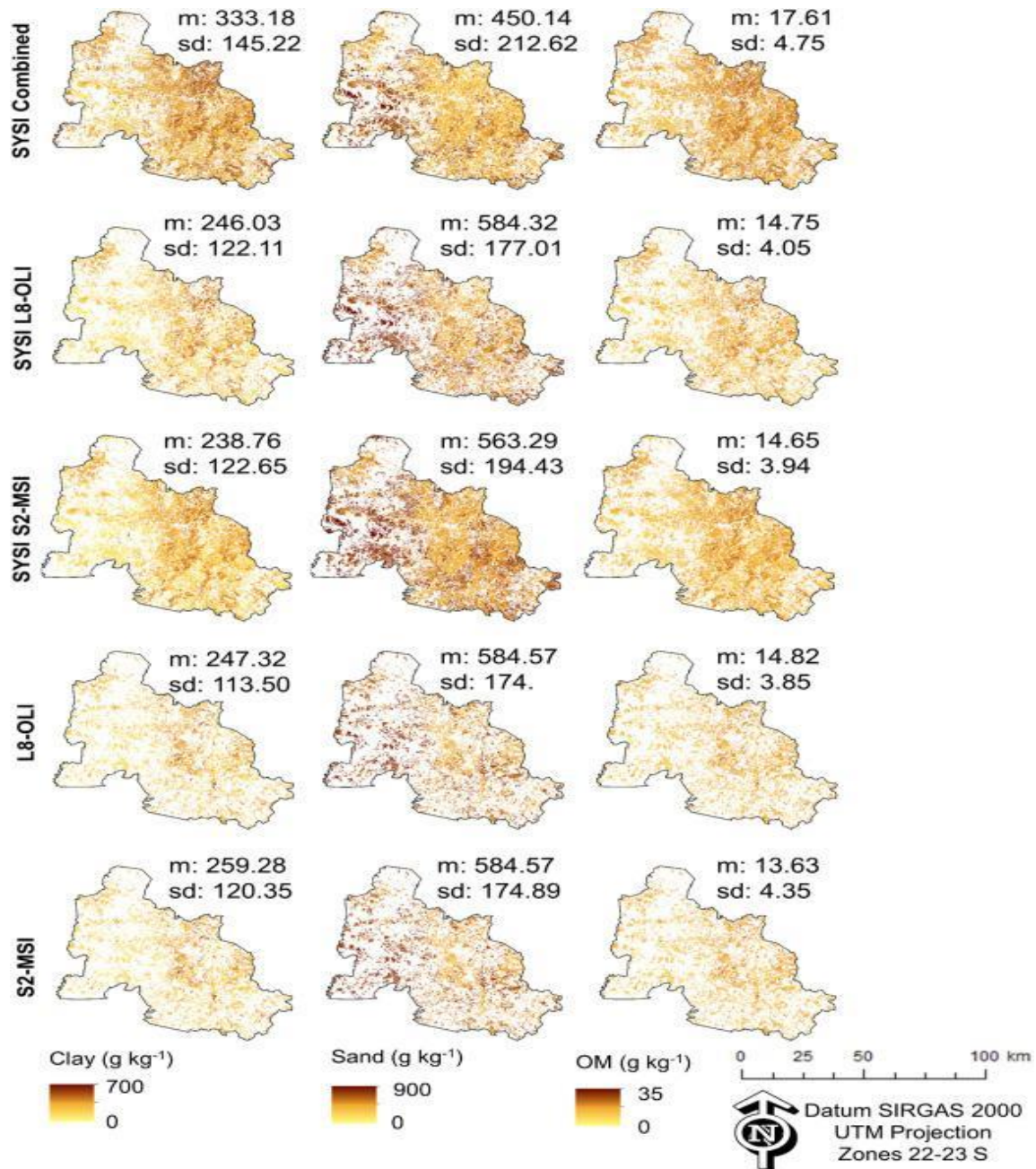
Figure 8. Predicted maps and mean (m) and standard deviation (sd) of clay, sand and organic matter (OM) contents.

Although the GEOS3 method was originally conceived for Landsat data (Demattê et al., 2018), it also has potential to retrieve bare soil pixels from Sentinel-2 images. The GEOS3 method efficiently retrieved bare soil pixels for both SYSIs and single-date images, confirmed by the closeness of scatter points to the 1:1 line (Baret et al., 1993). Considering the percentage of the area with bare soil pixels, the SYSI Combined had 3 and 13% more than SYSI S2-

MSI and SYSI L8-OLI and almost twice as much as single-date images. By combining the images from both satellites, we have more than two images per month and the revisit time has to do with this. The short temporal resolution provided by both satellites combined helped us to detect changes in land use (Shao et al., 2019) and increased the number of pixels with bare soil when compared to the SYSIs for each satellite separately. SYSI S2-MSI also had more bare soil pixels than SYSI L8-OLI, and the shorter revisit time of Sentinel-2 satellites (~ five days) is an advantage, as more images can be obtained and, consequently, more bare soil pixels. Some issues need to be considered regarding the combination of the images from both satellites, though. The SYSI Combined had lost the best spatial and spectral resolution provided by Sentinel-2 images, since those were resampled to 30 m and the RedEdge bands were removed to allow the combination with Landsat images. Fortunately, these issues did not affect the prediction accuracies, as revealed by our results.
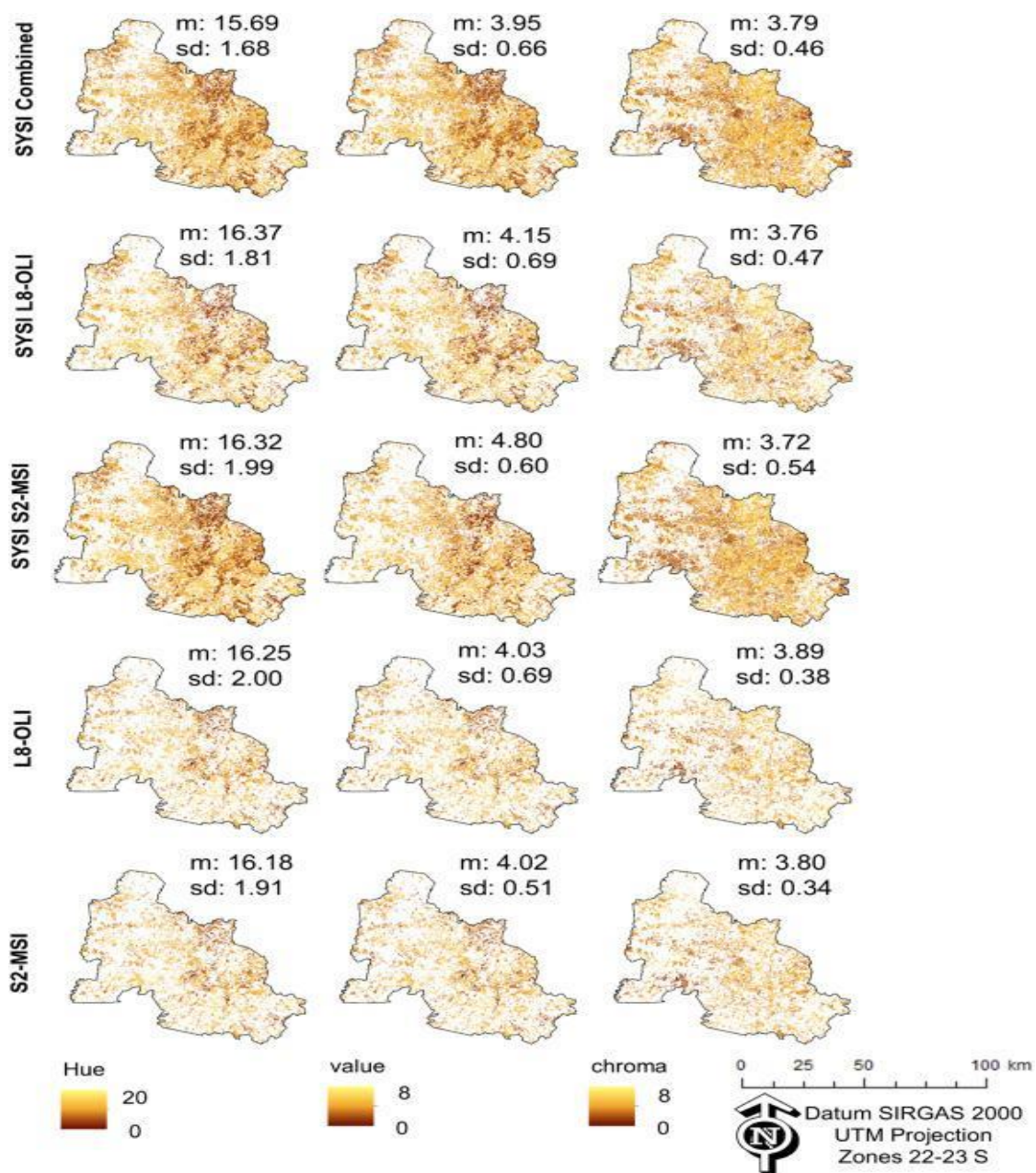
Figure 9. Predicted maps and mean (m) and standard deviation (sd) of hue, value and chroma components (soil color).

In this research, we have not defined an optimal number of images to obtain at least 90% bare soil pixels, as performed by Diek et al. (2017) in Swiss Plateau, because there are several factors that need to be considered. First, according to our knowledge and experience working with bare soil images in tropical regions, the number of images is important but more critical is its quality, which would define the amount of bare soil pixels in a single image. That is because the influence of clouds and shadows are substantial in tropical regions. For example, Demattê et al. (2018) were able to retrieve 68% and 92% of bare soil pixels in a single scene, but using ~40 years of Landsat images. As we used only images with 20% cloud coverage from 2016 to 2019, we acquired 45 and 63 images from 85 and 170 images available for Landsat and Sentinel-2, respectively (Fig. 5), which provided <50% bare soil pixels. Besides that, Sentinel-2 images are available only since 2015, which hampers further studies of the optimal number of images from this satellite. Another issue is related to the period when the images were taken. Twenty images from the dry period would provide more bare soil pixels than 20 images from the moist season, due to cloud and management influences.
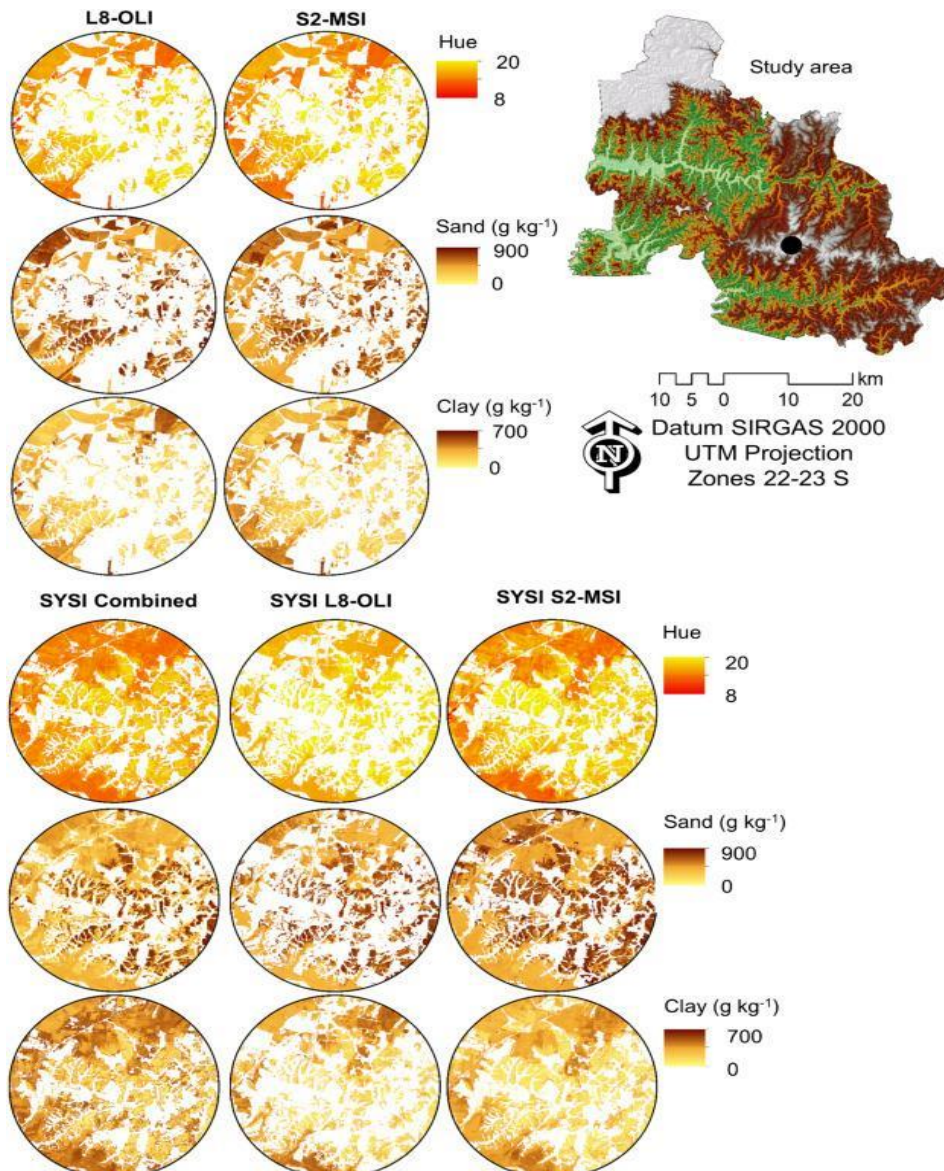
Figure 10. Details of spatial variability of soil properties predicted by SYSI and single-date images.

The spectral signatures of lab spectra, single-date and SYSI images were very similar and we confirmed that they follow the typical spectral patterns described for soils (Demattê et al., 2014). But we found that satellite's spectra were slightly lower than lab spectra and this can be related to several factors. First, lab spectra are acquired under controlled conditions, without influence of the environment while reflectance spectra captured by satellites are more susceptible to atmospheric conditions and geometric distortions (Richter and Schläpfer, 2002). Besides that, soil surface conditions such as moisture level and roughness can also influence the spectra (Chang et al., 2005; Janik et al., 2016; Lobell and Asner, 2002; Silvero et al., 2020). We found that the SYSIs have spectral patterns very similar to those from single-date images and this can confirm that the median statistic performed to obtain the SISY's reflectance has low influence of moisture, as the single-date images were acquired in the dry season. This issue was raised by Žížala et al. (2019) who was concerned about the usefulness of multi-temporal images for studying soil organic carbon. By the soil line analysis and the comparison of lab spectra with those from satellites, we confirmed that SYSI and single-date images represented bare soil reflectance patterns from our study area.
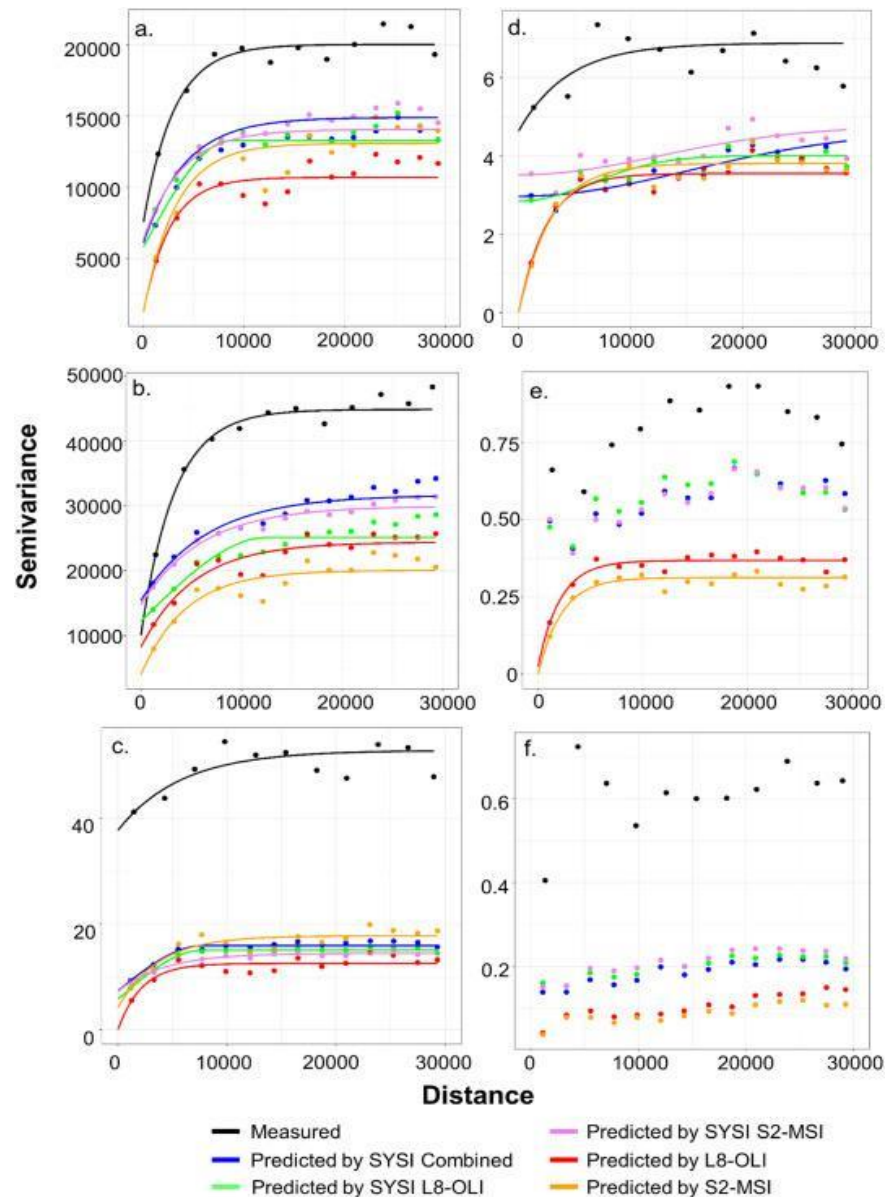
Figure 11. Semivariogram of predicted and observed values of soil properties. a. clay content, b. sand content, c. organic matter content, d. hue, e. value and f. chroma. Silt content was excluded due to its low performances.

3.4.2. Relationship of soil properties with bare soil images

Our study area is highly variable in parent materials and soil types (Bazaglia Filho et al., 2013; Fongaro et al., 2018; Gallo et al., 2018; Rizzo et al., 2020), which influenced the spatial variability of soil properties. We found very low to very high contents of clay, sand and silt and the OM content was low (< 5%), which is related to climate influences and the historical soil management by sugarcane farmers (de Bordonal et al., 2018; Cherubin et al., 2016; Franco et al., 2015). We observed that soil properties were highly correlated with SISY's bands, except chroma component and silt content; however, with single-date images, correlation values were low. The majority of the works that used single-date images found suitable correlation with soil properties ($r > 0.5$) (Ben-Dor et al., 2003; Ben-Dor et al., 1997; da Chagas et al., 2016; Forkuor et al., 2017), slightly higher than those observed here.

Interestingly, all spectral bands were highly correlated with soil color components, except chroma and those from single-date S2-MSI, which show the lowest correlations. Soil color determination is related (physically) to the visible portion of the electromagnetic spectrum (Fernandez and Schulze, 1987; Odeh and McBratney, 2005; Sovdat et al., 2019) and we expected the highest values of correlation between soil color and these spectral bands. The high correlation values between satellite spectral bands and soil color was already described by Escadafal et al. (1989) and Mattikalli (1997), who used six simulated Landsat spectral bands (vis-NIR-SWIR) to relate to Munsell soil color. Color is a very important soil property that have been used to support studies in pedology (Bigham et al., 1993; Siqueira et al., 2015; Stockmann et al., 2018;) drainage, wetlands, soil moisture mapping (Barrett, 2002; Bigham et al., 1993; Pretorius et al., 2017; Schmidt and Ahn, 2019), and mineralogical studies (Barron and Torrent, 1986; Nagano and Nakashima, 1989; Nagano et al., 1992; Richardson and Daniels (1993); Campos et al., 2003; Viscarra Rossel et al., 2010). Although an important property, studies that relate soil color with satellite spectral bands are scarce and only in recent years have been more research about it (Poppiel et al., 2020; Ramos et al., 2020).

3.4.3. Influence of spectral, spatial and temporal resolutions on prediction performances

Spatial, spectral and temporal resolutions of satellite images influenced the prediction performances. Single-date images had the lowest performances and multi-temporal images were more efficient in predicting soil properties, probably due to their ability to capture soil variability over time (Fathololoumi et al., 2020). The SYSI Combined and the SYSI L8-OLI, with spatial resolution of 30 m and six spectral bands, were the best predictors. Images from the Sentinel-2 satellite (single and SYSI), which had the best spatial and spectral resolutions (20 m and nine spectral bands), did not necessarily provide the best results, although the SYSI was better than the single-date image from this satellite. The RedEdge bands, present in Sentinel-2 and absent in Landsat-8, are more related to vegetation than to soil compounds (Delegido et al., 2011; Pinheiro et al., 2019; Sun et al., 2020) and although they showed significant correlations with soil properties, they were not substantial to explain the variability of soil properties.

In general, Landsat-8 data were better than Sentinel-2 to predict topsoil properties. Žížala et al. (2019) tested several multispectral sensors, including Landsat and Sentinel-2 to assess the influence of spatial and spectral resolution on soil organic carbon (SOC) mapping. In contrast to our results, they found that Sentinel-2 was better than Landsat-8 to predict SOC. Similarly, Rosero-Vlasova et al. (2018) also reported that Sentinel-2 provided better results than Landsat-8 in predicting SOC and texture in Spain, but they used simulated spectra acquired in laboratory.

As far as we know, there are no studies that merged images from Landsat-8 and Sentinel-2 to obtain bare soil pixels and predict topsoil properties, but there are countless that used both single-date and, more recently, multi-temporal bare soil images and reported suitable results. Demattê et al. (2007) and Liao et al. (2013) used Landsat

ETM data to predict soil texture by multiple linear regression while Xiao et al. (2006) used both Landsat ETM and Landsat TM to develop a topsoil grain size index based on correlation analysis between soil texture and satellites´ spectral bands. Ahmed and Iqbal (2014) mapped soil texture and OM content using Landsat data and multiple linear regression. The authors stated that the results for OM were satisfactory despite the low content measured in the study area (<1%). On the other hand, Zhai et al. (2006) and Forkuor et al. (2017) used machine learning algorithms to predict several topsoil properties using remote sensing products, including Landsat. Both studies found reasonable results and were similar to those found here for single-date images.

As Sentinel-2 satellites have been available only since 2015, there are still few studies about its potential for soil mapping, mostly in the European region. Castaldi et al. (2016) used simulated spectral data of Sentinel-2, Landsat and forthcoming hyperspectral sensors to estimate soil texture and SOC values. The authors studied the influence of spectral resolution (number of spectral bands) and they found that hyperspectral sensors would be more feasible to study soil properties. Differently from our results, they found that Sentinel-2 outperformed Landsat data for clay and sand content estimates (R2 > 0.5). Although testing hyperspectral sensors is beyond the scope of this research, we can highlight the contrasting results obtained. While we found that the Landsat data performed better than Sentinel-2, this seems to be more suitable for soil properties studies in temperate regions. Zhou et al. (2020) tested different models using DEM derivatives, radar and optical data from Sentinel to predict SOC and nitrogen and although they concluded that this satellite proved useful for soil mapping, they found better results by combining optical and radar data.

We found that the use of multi-temporal bare soil images was better than single-date images to predict topsoil properties in tropical areas. The use of bare soil pixels, either from single-date or from time series have been reported for both Landsat and Sentinel data. Diek et al. (2017) obtained the barest soil composite over Swiss Plateau and found reasonable results for predictions of soil texture and SOC, although their results were lower than those presented here (R2 < 0.5). Other researchers used the GEOS3 method and reported similar to our results for tropical areas (Fongaro et al., 2018; Gallo et al., 2018; de Mendes et al., 2019; Poppiel et al., 2020; Poppiel et al., 2019a; Poppiel et al., 2019b; Rizzo et al., 2020; Bonfatti et al., 2020) and over Europe (Safanelli et al., 2020). The Soil Composite Processor (SCMaP) developed by Rogge et al. (2018) was applied by Žížala et al. (2019) to obtain bare soil pixels from Sentinel-2 images and reasonable results for SOC were reported, even better than ours.

For Sentinel-2 data, some works reported the use of NDVI and NBR2 thresholds to obtain bare soil pixels, either from single-date images or from time series. Castaldi et al. (2019a) used simulated Sentinel-2 data to obtain models for SOC predictions. The authors tested several NBR2 thresholds to obtain bare soil pixels from Sentinel-2 images to which they applied the models obtained with simulated data. They accurately map SOC over a large area. Vaudour et al. (2019) predicted eight soil properties in two French areas, using Sentinel-2 images, with 8 and 39% of the area with bare soil pixels. Their results for soil texture were fairly lower (R2 < 0.5) than those observed here. In mainland France, Loiseau et al. (2019) used bare soil mosaics obtained from two years of Sentinel-2 images coupled with other covariates and satellites to map clay content using quantile forest regression. They found that satellite data integration, in which Sentinel-2 is included, very slightly improved the prediction of the soil clay content but showed potential for soil mapping following the directions of the GlobalSoilMap project. Gomez et al. (2019) also used Sentinel-2 time-series to classify soil texture, but only six images from three months (February–April) were used, which resulted in moderate performances.

To our knowledge, no such work was reported using multi-temporal images from Sentinel-2 in tropical regions, which hampered our results with those from similar climatic conditions.

3.4.4. Predicted map and spatial structure analysis

Spatial maps from single-date and SYSI images coming from different satellites were very similar. We found that their spatial variability and structure was also similar, as revealed by their similar range values. The measured data showed higher values of semivariance than those observed for predicted values. The components of soil color, Value and chroma, had none of the theoretical models fitted to their experimental semivariogram, which can be related to the number of samples used, being insufficient to explain their spatial structure. Similar results were also found by Vaudour et al. (2019) who reported that observed values had semivariance values higher than those from predicted values for clay and iron content. Gomez et al. (2018) also reported that the semivariogram of predicted values by different satellite/sensor were very similar.

3.4.5. Advantages, limitations and future research

We show here and other researches already highlighted the high potential of satellite images for soil properties mapping and study. Nowadays, with the free open access of satellite images in cloud-based platforms such as the GEE, obtaining bare soil images is becoming very efficient, mainly from Landsat missions, which have a plenty of data of almost four decades. Although GEE still has limited capacity for Sentinel-2, due to its short time in orbit, its potential cannot be underestimated.

There are limitations, though. We observed that it is still difficult to obtain full bare soil pixels from tropical areas due to the influence of clouds, shadows and soil management. Future researchers should pursue better strategies to obtain bare soil mosaics for Sentinel-2 data, to fully exploit its potential for soil mapping. Besides that, further studies are needed on strategies of how Landsat and Sentinel data can be fully operational together. Although there are efforts in developing harmonized Landsat-Sentinel data and data fusion (Claverie et al., 2018; Shao et al., 2019) further research on how to obtain bare soil pixels from the integration of two satellites is needed.

3.5. Conclusions and final remarks

Multi-temporal images with bare soil pixels, obtained from time series of Landsat-8 and Sentinel-2 satellites, were suitable to obtain soil property maps from tropical regions. Although the Sentinel-2 images provided the best spatial, spectral and temporal resolution, the combination of images from both satellites was more suitable for soil mapping purposes. Single-date images from both satellites, although useful, have still limited capacity to provide bare soil information for soil mapping purposes for large areas.

The better temporal resolution, expressed by the higher revisit time provided by the combination of the images from Landsat8-OLI and Sentinel2-MSI satellites, was the most important factor in our study. The SYSI Combined showed the highest bare soil coverage and the best model's performances with six spectral bands and 30 m-pixel size. The RedEdge extra bands of the Sentinel-2 satellite did not add suitable information to the models, probably due that it is more related to vegetation characteristics than soil compounds.

Soil maps produced using the SYSI Combined can be used to meet the demands of detailed soil maps, especially in tropical regions. With them, sustainable land use and management practices can be achieved.

Acknowledgments

Appendix A. Supplementary data

Supplementary data to this article can be found online at https:// doi.org/10.1016/j.rse.2020.112117.

References

Ahmed, Z., Iqbal, J., 2014. Evaluation of Landsat TM5 multispectral data for automated mapping of surface soil texture and organic matter in GIS. Eur. J. Remote Sens. 47, 557–573. https://doi.org/10.5721/EuJRS20144731.

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

Baret, F., Jacquemoud, S., Hanocq, J.F., 1993. About the soil line concept in remote sensing. Adv. Sp. Res. 13, 281–284. https://doi.org/10.1016/0273-1177(93)90560-X.

Barrett, L.R., 2002. Spectrophotometric color measurement in situ in well drained sandy soils. Geoderma 108, 49–77. https://doi.org/10.1016/S0016-7061(02)00121-0.

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

Bazaglia Filho, O., Rizzo, R., Lepsch, I.F., Prado, H., Gomes, F.H., Mazza, J.A., Demattê, J.A.M., 2013. Comparison between detailed digital and conventional soil maps of an area with a complex geology. Rev. Bras. Ciência do Solo do Solo 37, 1136–1148.

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.

Ben-Dor, E., Inbar, Y., Chen, Y., 1997. The reflectance spectra of organic matter in the visible near-infrared and short-wave infrared region (400 - 2500 nm) during a con-troled decomposition process. Remote Sens. Environ. 61, 1–15.

Ben-Dor, E., Goldlshleger, N., Benyamini, Y., Agassi, M., Blumberg, D.G., 2003. The spectral reflectance properties of soil structural crusts in the 1.2- to 2.5-µm spectral region. Soil Sci. Soc. Am. J. 67, 289–299. https://doi.org/10.2136/sssaj2003.2890.

Bigham, J.M., Ciolkosz, E.J., Post, D.F., Bryant, R.B., Batchily, A.K., Huete, A.R., Levine, S.J., Mays, M.D., 1993. Correlations between field and laboratory measurements of soil color, in: Soil color. Springer Nat. 35–49. https://doi.org/10.2136/ sssaspecpub31.c3. SSSA Special Publication No.31.

Blasch, G., Spengler, D., Itzerott, S., Wessolek, G., 2015. Organic matter modeling at the lanEK-SCape scale based on multitemporal soil pattern analysis using RapidEye data. Remote Sens. 7, 11125–11150. https://doi.org/10.3390/rs70911125.

Bonfatti, B.R., Demattê, J.A.M., Marques, K.P.P., Poppiel, R.R., Rizzo, R., de Mendes, W.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/10.1016/j.geomorph.2020.107305.

Campos, R.C., Demattê, J.A.M., Quartaroli, C.F., 2003. Determinação do teor de hematita no solo a partir de dados de colorimetria e radiometria. Pesqui. Agropecuária Bras. 38, 521–528.

Castaldi, F., Palombo, A., Santini, F., Pascucci, S., Pignatti, S., Casa, R., 2016. Evaluation of the potential of the current and forthcoming multispectral and hyperspectral im-agers to estimate soil texture and organic carbon. Remote Sens. Environ. 179, 54–65. https://doi.org/10.1016/J.RSE.2016.03.025.

Castaldi, F., Chabrillat, S., Don, A., van Wesemael, B., 2019a. Soil organic carbon map-ping using LUCAS topsoil database and Sentinel-2 data: an approach to reduce soil moisture and crop residue effects. Remote Sens. 11, 2121. https://doi.org/10.3390/rs11182121.

Castaldi, F., Hueni, A., Chabrillat, S., Ward, K., Buttafuoco, G., Bomans, B., Vreys, K., Brell, M., van Wesemael, B., 2019b. Evaluating the capability of the sentinel 2 data for soil organic carbon prediction in croplands. ISPRS J. Photogramm. Remote Sens. 147, 267–282. https://doi.org/10.1016/j.isprsjprs.2018.11.026.

Chang, C.-W., Laird, D.A., Hurburgh, C.R., 2005. Influence of soil moisture on near-in-frared reflectance spectroscopic measurement of soil properties. Soil Sci. 170, 244–255. https://doi.org/10.1097/01.ss.0000162289.40879.7b.

Cherubin, M.R., Karlen, D.L., Cerri, C.E.P., Franco, A.L.C., Tormena, C.A., Davies, C.A., Cerri, C.C., 2016. Soil quality indexing strategies for evaluating sugarcane expansion in Brazil. PLoS One 11, e0150860. https://doi.org/10.1371/journal.pone.0150860.

Claverie, M., Ju, J., Masek, J.G., Dungan, J.L., Vermote, E.F., Roger, J.C., Skakun, S.V., Justice, C., 2018. The harmonized Landsat and Sentinel-2 surface reflectance data set. Remote Sens. Environ. 219, 145–161. https://doi.org/10.1016/j.rse.2018.09.002.

da Chagas, C.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/10.1016/j.catena. 2016.01.001.

de Arruda, G.P., Demattê, J.A.M., da Chagas, C.S., Fiorio, P.R., Souza, A.B., 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.

de Bordonal, R.O., Carvalho, J.L.N., Lal, R., Figueiredo, E.B., Oliveira, B.G., La Scala Jr., N., 2018. Sustainability of sugarcane production in Brazil. A review. Agron. Sustain. Dev. 38, 1–23. https://doi.org/10.1007/s13593-018-0490-x.

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

de Padilha, M.C.C., Vicente, L.E., Demattê, J.A.M., dos Santos Wendriner Loebmann, D.G., Vicente, A.K., Salazar, D.F.U., Guimarães, C.C.B., 2020. Using Landsat and soil clay content to map soil organic carbon of oxisols and Ultisols near São Paulo, Brazil. Geoderma Reg. 21, e00253. https://doi.org/10.1016/j.geodrs.2020.e00253.

Delegido, J., Verrelst, J., Alonso, L., Moreno, J., 2011. Evaluation of sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content. Sensors 11, 7063–7081. https://doi.org/10.3390/s110707063.

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. https://doi.org/10.1080/01431160601121469.

Demattê, J.A.M., Bellinaso, H., Romero, D.J., Fongaro, C.T., 2014. Morphological interpretation of reflectance Spectrum (MIRS) using libraries looking towards soil classi-fication. Sci. Agric. 71, 509–520. https://doi.org/10.1590/0103-9016-2013-0365.

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., de Mendes, W.S., Bonfatti, B.R., Dotto, A.C., Salazar, D.F.U., de Mello, F.A.O., da Paiva, A.F.S., Souza, A.B., dos Santos, N.V., Maria Nascimento, C., de Mello, D.C., Bellinaso, H., Gonzaga Neto, L., Amorim, M.T.A., de Resende, M.E.B., da Vieira, J.S., de Queiroz, L.G., Gallo, B.C., Sayão, V.M., da Lisboa, C.J.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.

Diek, S., Fornallaz, F., Schaepman, M.E., De Jong, R., 2017. Barest pixel composite for agricultural areas using Landsat time series. Remote Sens. 9, 1245. https://doi.org/ 10.3390/rs9121245.

Escadafal, R., Girard, M.C., Courault, D., 1989. Munsell soil color and soil reflectance in the visible spectral bands of landsat MSS and TM data. Remote Sens. Environ. 27, 37–46. https://doi.org/10.1016/0034-4257(89)90035-7.

Fathololoumi, S., Vaezi, A.R., Alavipanah, S.K., Ghorbani, A., Saurette, D., Biswas, A., 2020. Improved digital soil mapping with multitemporal remotely sensed satellite data fusion: A case study in Iran. Sci. Total Environ. 721, 137703. https://doi.org/10.1016/j.scitotenv.2020.137703.

Fernandez, R.N., Schulze, D.G., 1987. Calculation of soil color from reflectance Spectra1. Soil Sci. Soc. Am. J. 51, 1277. https://doi.org/10.2136/sssaj1987. 03615995005100050033x.

Fongaro, C., Demattê, J., Rizzo, R., Lucas Safanelli, J., Mendes, W., Dotto, A., Vicente, L., Franceschini, M., Ustin, S., 2018. Improvement of clay and sand quantification based on a novel approach with a focus on multispectral satellite images. Remote Sens. 10, 1555. https://doi.org/10.3390/rs10101555.

Forkuor, G., Hounkpatin, O.K.L., Welp, G., Thiel, M., 2017. High resolution mapping of soil properties using remote sensing variables in South- Western Burkina Faso: A comparison of machine learning and multiple linear regression models. PLoS One 12, 1–21 (doi:10.1371/).

Franco, A.L.C., Cherubin, M.R., Pavinato, P.S., Cerri, C.E.P., Six, J., Davies, C.A., Cerri, C.C., 2015. Soil carbon, nitrogen and phosphorus changes under sugarcane expansion in Brazil. Sci. Total Environ. 515–516, 30–38. https://doi.org/10.1016/j.scitotenv.2015.02.025.

Gallo, B., Demattê, J., Rizzo, R., Safanelli, J., Mendes, W., Lepsch, I., Sato, M., Romero, D., Lacerda, M., 2018. Multi-temporal satellite images on topsoil attribute quantification and the relationship with soil classes and geology. Remote Sens. 10, 1–21. https://doi.org/10.3390/rs10101571.

Gomez, C., Lagacherie, P., Coulouma, G., 2012. Regional predictions of eight common soil properties and their spatial structures from hyperspectral Vis-NIR data. Geoderma 189–190, 176–185. https://doi.org/10.1016/j.geoderma.2012.05.023.

Gomez, C., Adeline, K., Bacha, S., Driessen, B., Gorretta, N., Lagacherie, P., Roger, J.M., Briottet, X., 2018. Sensitivity of clay content prediction to spectral configuration of VNIR/SWIR imaging data, from multispectral to hyperspectral scenarios. Remote Sens. Environ. 204, 18–30. https://doi.org/10.1016/j.rse.2017.10.047.

Gomez, C., Dharumarajan, S., Féret, J.-B., Lagacherie, P., Ruiz, L., Sekhar, M., 2019. Use of Sentinel-2 time-series images for classification and uncertainty analysis of inherent biophysical property: case of soil texture mapping. Remote Sens. 11, 565. https://doi.org/10.3390/rs11050565.

González, I., Déjean, S., Martin, P.G.P., Baccini, A., 2008. CCA: an R package to extend canonical correlation analysis. J. Stat. Softw. 23, 1–14. https://doi.org/10.18637/jss.v023.i12.

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.

Grunwald, S., Thompson, J.A., Boettinger, J.L., 2011. Digital soil mapping and modeling at continental scales: finding solutions for global issues. Soil Sci. Soc. Am. J. 75, 1201–1213. https://doi.org/10.2136/sssaj2011.0025.

Helder, D., Markham, B., Morfitt, R., Storey, J., Barsi, J., Gascon, F., Clerc, S., LaFrance, B., Masek, J., Roy, D.P., Lewis, A., Pahlevan, N., 2018. Observations and recommendations for the calibration of Landsat 8 OLI and sentinel 2 MSI for improved data interoperability. Remote Sens. 10, 1340. https://doi.org/10.3390/rs10091340.

Hengl, T., De Jesus, J.M., Macmillan, R.A., Batjes, N.H., Heuvelink, G.B.M., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Gonzalez, M.R., 2014. SoilGrids1km — global soil information based on automated mapping. PLoS One 9. https://doi.org/10.1371/journal.pone.0105992.

Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., Ruiperez Gonzalez, M., Kilibarda, M., Blagoti, 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 in-formation based on machine learning. PLoS One 12, e0169748. https://doi.org/10.1371/journal.pone.0169748.

Isaaks, E.H., Srivastava, R.M., 1989. An Introduction to Applied Geostatistics. Oxford University Press, Inc., New York, NY. Ishida, T., Ando, H., 1999. Use of disjunctive cokriging to estimate soil organic matter from Landsat thematic mapper image. Int. J. Remote Sens. 20, 1549–1565. https://doi.org/10.1080/014311699212605.

IUSS Working Group WRB, 2015. World Reference Base for Soil Resources.

Janik, L.J., Soriano-Disla, J.M., Forrester, S.T., McLaughlin, M.J., 2016. Moisture effects on diffuse reflection infrared spectra of contrasting minerals and soils: A mechanistic interpretation. Vib. Spectrosc. 86, 244–252. https://doi.org/10.1016/J.VIBSPEC.2016.07.005.

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

Kuhn, M., Johnson, K., 2013. Applied Predictive Modeling, Springer. Springer, New York. https://doi.org/10.1007/978-1-4614-6849-3.

Kuhn, M., Weston, S., Keefer, C., Coulter, N., Quinlan, R., 2018. Package "Cubist". Lagacherie, P., McBratney, A.B., Voltz, M., 2007. Digital Soil Mapping: An Introductory Perspective. Elsevier, Amsterdam.

Liao, K., Xu, S., Wu, J., Zhu, Q., 2013. Spatial estimation of surface soil texture using remote sensing data. Soil Sci. Plant Nutr. 59, 488–500. https://doi.org/10.1080/00380768.2013.802643.

Lobell, D.B., Asner, G.P., 2002. Moisture effects on soil reflectance. Soil Sci. Soc. Am. J. 66, 722. https://doi.org/10.2136/sssaj2002.7220.

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/10.1016/j.jag.2019.101905.

Main-Knorn, M.B.P., Louis, J., Debaecker, V., Muller-Wilm, U., Gascon, F., 2017. Sen2Cor for Sentinel-2. In: Image and Signal Processing for Remote Sensing, pp. 12.

Markham, B., Storey, J., Morfitt, R., 2015. Landsat-8 sensor characterization and cali-bration. Remote Sens. 7, 2279–2282. https://doi.org/10.3390/rs70302279.

Mathieu, R., Pouget, M., Cervelle, B., Escadafal, R., 1998. Relationships between satellite- based radiometric indices simulated using laboratory reflectance data and typic soil color of an arid environment. Remote Sens. Environ. 66, 17–28.

Mattikalli, N.M., 1997. Soil color modeling for the visible and near-infrared bands of Landsat sensors using laboratory spectral measurements. Remote Sens. Environ. 59, 14–28. https://doi.org/10.1016/S0034-4257(96)00075-2.

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

McBratney, A.B., Field, D., Morgan, C.L.S., Huang, J., 2019. On soil capability, capacity, and condition. Sustainability 11, 3350. https://doi.org/10.3390/su11123350.

Minasny, B., McBratney, A.B.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.

Moeys, J., 2018. The Soil Texture Wizard: R Functions for Plotting, Classifying, Transforming and Exploring Soil Texture Data.

Mulder, V.L., De Bruin, S., Schaepman, M.E., Mayr, T.R., 2011. The use of remote sensing in soil and terrain mapping — A review. Geoderma 162, 1–19. https://doi.org/10.1016/j.geoderma.2010.12.018.

Nagano, T., Nakashima, S., 1989. Study of colors and degrees of weathering of granitic rocks by visible diffuse reflectance spectroscopy. Geochem. J. 23, 75–83. https://doi.org/10.2343/geochemj.23.75.

Nagano, T., Nakashima, S., Nakayama, S., Osada, K., Senoo, A.M., 1992. Color variations associated with rapid formation of goethite from proto-ferrihydrite at pH 13 and 40 C. Clay Clay Miner. 40, 6013–6607.

Odeh, I.O.A., McBratney, A.B., 2005. Pedometrics. In: Hillel, D. (Ed.), Encyclopedia of Soils in the Environment. Elsevier, pp. 166–175. https://doi.org/10.1016/B0-12-348530-4/00020-5.

Olea, R.A., 1999. Block kriging. In: Olea, R.A. (Ed.), Geostatistics for Engineers and Earth Scientists. Springer US, pp. 187–208. https://doi.org/10.1007/978-1-4615-5001-3_12.

Pinheiro, H.S.K., Barbosa, T.P.R., Antunes, M.A.H., de Carvalho, D.C., Nummer, A.R., de Junior, W.C., da Chagas, C.S., Fernandes-Filho, E.I., Pereira, M.G., 2019. Assessment of phytoecological variability by red-edge spectral indices and soil-lanEK-SCape relationships. Remote Sens. 11, 2448. https://doi.org/10.3390/RS11202448.

Poggio, L., Gimona, A., 2017. Assimilation of optical and radar remote sensing data in 3D mapping of soil properties over large areas. Sci. Total Environ. 579, 1094–1110. https://doi.org/10.1016/j.scitotenv.2016.11.078.

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, 2905. https://doi.org/10.3390/rs11242905.

Poppiel, R.R., Lacerda, M.P.C., Rizzo, R., Safanelli, J.L., Bonfatti, B.R., Silvero, N.E.Q., Demattê, J.A.M., 2020. Soil color and mineralogy mapping using proximal and remote sensing in Midwest Brazil. Remote Sens. 12, 1197. https://doi.org/10.3390/rs12071197.

Pretorius, M.L., Van Huyssteen, C.W., Brown, L.R., 2017. Soil color indicates carbon and wetlands: developing a color-proxy for soil organic carbon and wetland boundaries on sandy coastal plains in South Africa. Environ. Monit. Assess. 189, 556. https://doi.org/10.1007/s10661-017-6249-z.

Quinlan, J.R., 1992. Learning with continuous classes. In: Proceedings AI'92, 5th Australian Conference on Artificial Intelligence.World Scientific, pp. 343–348.

R Core Team, 2019. R: A Language and Environment for Statistical Computing.

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

Richardson, J.L., Daniels, R.B., 1993. Stratigraphic and hydraulic influences on soil color development. In: Bigham, J.M., Ciolkosz, E.J. (Eds.), Soil Color, Special Publication No. 31. Soil Science Society of America, Madison, WI, pp. 109–126. https://doi.org/10.2136/sssaspecpub31.c7.

Richter, R., Schläpfer, D., 2002. Geo-atmospheric processing of airborne imaging spec-trometry data. Part 2: Atmospheric/topographic correction. Int. J. Remote Sens. https://doi.org/10.1080/01431160110115834.

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

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

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 char-acteristics of soils with Landsat imagery (1984–2014). Remote Sens. Environ. 205, 1–17. https://doi.org/10.1016/J.RSE.2017.11.004.

Rosero-Vlasova, O.A., Vlassova, L., Pérez-Cabello, F., Montorio, R., Nadal-Romero, E., 2018. Modeling soil organic matter and texture from satellite data in areas affected by wildfires and cropland abandonment in Aragón, Northern Spain. J. Appl. Remote. Sens. 12, 1. https://doi.org/10.1117/1.jrs.12.042803.

Safanelli, J.L., Chabrillat, S., Ben-Dor, E., Demattê, J.A.M., 2020. Multispectral models from bare soil composites for mapping topsoil properties over Europe. Remote Sens. 12, 1369. https://doi.org/10.3390/RS12091369.

Schmidt, S.A., Ahn, C., 2019. A comparative review of methods of using soil colors and their patterns for wetland ecology and management. Commun. Soil Sci. Plant Anal. 50, 1293–1309. https://doi.org/10.1080/00103624.2019.1604737.

Shao, Z., Cai, J., Fu, P., Hu, L., Liu, T., 2019. Deep learning-based fusion of Landsat-8 and Sentinel-2 images for a harmonized surface reflectance product. Remote Sens. Environ. 235. https://doi.org/10.1016/j.rse.2019.111425.

Silvero, N.E.Q., Di Raimo, L.A.D.L., Pereira, G.S., de Magalhães, L.P., da Terra, F.S., Dassan, M.A.A., Salazar, D.F.U., Demattê, J.A.M., 2020. Effects of water, organic matter, and iron forms in mid-IR spectra of soils: assessments from laboratory to satellite-simulated data. Geoderma 375, 114480. https://doi.org/10.1016/j.geoderma.2020.114480.

Siqueira, D.S., Marques, J., Pereira, G.T., Teixeira, D.B., Vasconcelos, V., Carvalho Júnior, O.A., Martins, E.S., 2015. Detailed mapping unit design based on soil–lanEK-SCape re- lation and spatial variability of magnetic susceptibility and soil color. Catena 135, 149–162. https://doi.org/10.1016/j.catena.2015.07.010.

Sovdat, B., Kadunc, M., Batič, M., Milčinski, G., 2019. Natural color representation of Sentinel-2 data. Remote Sens. Environ. 225, 392–402. https://doi.org/10.1016/j.rse.2019.01.036.

Stockmann, U., Jones, E.J., Odeh, I.O.A., McBratney, A.B., 2018. Pedometric treatment of soil attributes, in: pedometrics. Springer Nat. 115–153. https://doi.org/10.1007/978-3-319-63439-5_5.

Sun, Y., Qin, Q., Ren, H., Zhang, T., Chen, S., 2020. Red-edge band vegetation indices for leaf area index estimation from Sentinel-2/MSI imagery. IEEE Trans. Geosci. Remote Sens. 58, 826–840. https://doi.org/10.1109/TGRS.2019.2940826.

Teixeira, P.C., Donagema, G.K., Fontana, A., Teixeira, W.G., 2017. Manual de Métodos de Análise de Solo.

Tziolas, N., Tsakiridis, N., Ben-Dor, E., Theocharis, J., Zalidis, G., 2020. Employing a multi-input deep convolutional neural network to derive soil clay content from a synergy of multi-temporal optical and radar imagery data. Remote Sens. 12, 1389. https://doi.org/10.3390/RS12091389.

van der Werff, H., van der Meer, F., 2015. Sentinel-2 for mapping iron absorption feature parameters. Remote Sens. 7, 12635–12653. https://doi.org/10.3390/rs71012635.

Vaudour, E., Gomez, C., Fouad, Y., Lagacherie, P., 2019. Sentinel-2 image capacities to predict common topsoil properties of temperate and Mediterranean agroecosystems. Remote Sens. Environ. 223, 21–33. https://doi.org/10.1016/J.RSE.2019.01.006.

Vermote, E., Justice, C., Claverie, M., Franch, B., 2016. Preliminary analysis of the per- formance of the Landsat 8/OLI land surface reflectance product. Remote Sens. Environ. 185, 46–56. https://doi.org/10.1016/j.rse.2016.04.008.

Viscarra Rossel, R.A., Minasny, B., Roudier, P., McBratney, A.B., 2006. Colour space models for soil science. Geoderma 133, 320–337. https://doi.org/10.1016/J.GEODERMA.2005.07.017.

Viscarra Rossel, R.A., Bui, E.N., De Caritat, P., Mckenzie, N.J., 2010. Mapping iron oxides and the color of Australian soil using visible–near-infrared reflectance spectra. J. Geophys. Res. 115, 4031. https://doi.org/10.1029/2009JF001645.

Viscarra Rossel, R.A., Behrens, T., Ben-Dor, E., Brown, D.J., Demattê, J.A.M., Shepherd, K.D., Shi, Z., Stenberg, B., Stevens, A., Adamchuk, V., Aïchi, H., Barthès, B.G., Bartholomeus, H.M., Bayer, A.D., Bernoux, M., Böttcher, K., Brodský, L., Du, C.W., Chappell, A., Fouad, Y., Genot, V., Gomez, C., Grunwald, S., Gubler, A., Guerrero, C., Hedley, C.B., Knadel, M., Morrás, H.J.M., Nocita, M., Ramirez-Lopez, L., Roudier, P., Campos, E.M.R., Sanborn, P., Sellitto, V.M., Sudduth, K.A., Rawlins, B.G., Walter, C., Winowiecki, L.A., Hong, S.Y., Ji, W., 2016. A global spectral library to characterize the world's soil. Earth-Sci. Rev. 155, 198–230. https://doi.org/10.1016/J.EARSCIREV.2016.01.012.

Walkley, A., Black, I.A., 1934. An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. Soil Sci. 37, 29–38.

Webster, R., 1977. Canonical correlation in pedology: how useful? J. Soil Sci. 28. Xiao, J., Shen, Y., Tateishi, R., Bayaer, W., 2006. Development of topsoil grain size index for monitoring desertification in arid land using remote sensing. Int. J. Remote Sens. 27, 2411–2422. https://doi.org/10.1080/01431160600554363.

Zhai, Y., Thomasson, J.A., Boggess, J.E., Sui, R., 2006. Soil texture classification with artificial neural networks operating on remote sensing data. Comput. Electron. Agric. 54, 53–68. https://doi.org/10.1016/j.compag.2006.08.001.

Zhang, H.K., Roy, D.P., Yan, L., Li, Z., Huang, H., Vermote, E., Skakun, S., Roger, J.C., 2018. Characterization of sentinel-2A and Landsat-8 top of atmosphere, surface, and nadir BRDF adjusted reflectance and NDVI differences. Remote Sens. Environ. 215, 482–494. https://doi.org/10.1016/j.rse.2018.04.031.

Zhou, T., Geng, Y., Chen, J., Pan, J., Haase, D., Lausch, A., 2020. High-resolution digital mapping of soil organic carbon and soil total nitrogen using DEM derivatives, Sentinel-1 and Sentinel-2 data based on machine learning algorithms. Sci. Total Environ. 729. https://doi.org/10.1016/j.scitotenv.2020.138244.

Žížala, D., Minařík, R., Zádorová, T., 2019. Soil organic carbon mapping using multi- spectral remote sensing data: prediction ability of data with different spatial and spectral resolutions. Remote Sens. 11, 2947.

4. STRATEGIES FOR MAPPING SOIL DRAINAGE CLASSES: A COMPARISON OF EXPERT KNOWLEDGE AND MACHINE LEARNING APPROACHES

Abstract

Two potential strategies for mapping soil drainage classes in a GIS were evaluated for the state of São Paulo, Brazil. One strategy was to utilize the wealth of expert knowledge (EK) to develop a simple model based on a few covariate layers. Another strategy was to utilize a more objective process of machine learning (ML) and a larger stack of carefully selected covariates. The operational consideration of these two approaches is particularly important in areas like Brazil, where information for soil maps is scarce or incomplete. For considering the full, operational implementation of soil mapping, this study rated the two approaches in terms of accuracy, labor-efficiency, transferability, and accessibility. Usually, soil drainage classes are defined as a function of morphological characteristics that are recorded during soil profile descriptions. The first strategy based on EK used soil color and texture as base maps to define the drainage classes. The ML-based strategy, in turn, used a decision tree algorithm and a set of environmental covariates to represent the main processes that occur in the landscape. An existing soil drainage class map obtained from a digital map of soil classes was used as a reference. The maps obtained by EK and ML strategies had, respectively, an agreement of 50 and 53% with the reference map. Considering the results and the evaluation of the selection criteria for the best strategy, the soil drainage map obtained using EK received the highest score in accuracy. This strategy also was the most-labor efficient, since it is based on two soil properties that are easily obtained from other legacy data or soil databases. As this strategy was based on simple rules regarding soil color groups and textural classes that are easy to apply to other areas, their transferability is also higher when compared to the ML-based strategy. Finally, because ML-based models are still considered as 'black-boxes' and their interpretability is difficult, the EK-based strategy also received the higher score in the accessibility criteria. Therefore, the EK-based strategy was selected and can be considered as the most accurate, labor-efficient, transferable, and accessible way to map soil drainage classes in tropical environments.

4.1. Introduction

Demand for detailed soil information is consistently increasing over the years as a means of supporting not only land management and taxation related to agriculture, but also other uses related to the role of soils in ecosystems (Brevik et al., 2015). The gap between this demand and the state of legacy soil maps is becoming even greater in developing countries such as Brazil. Here, most of the available soil maps are less detailed, incomplete or no longer sufficient to support public policies and other applications involving soil resources (Mendonça Santos et al., 2015).

One of the important characteristics that is lacking in most Brazilian maps is the natural soil drainage class, which is defined as the frequency and duration of wet periods or degree and frequency the soil matrix is free of water saturation (Soil Science Division Staff, 2017). It is used as an indicator of the general conditions of water movement in the soil, which is primarily influenced by gravity and determined by soil texture, structure, and topography (Fausey, 2005; Gerardin and Duerue, 1990; Troeh, 1964). Information on soil drainage is important because it controls several processes in the soil such as the decomposition of soil organic matter, distribution of soil organic carbon (Raymond et al., 2012), microbial adaptations (Graça et al., 2021), cow stocking rates and dry matter production (Fitzgerald et al., 2008), and which plant species are more suitable for a specific area (Levine et. al, 1994).

The description of soil drainage is usually performed in terms of seven discrete units or classes varying from very well to poorly drained, depending on the proportion and rate of water infiltration. Rather than directly

observing the long-term hydrologic conditions of a site, drainage class is interpreted from soil morphological features such as soil color, texture class, and presence of redoximorphic features that are observed when soil profiles are described in the field. Soil color and redoximorphic features reflect processes that take place in the soil due to water movement, specifically oxidation and reduction of iron forms. Well aerated and drained soils usually present reddish and brownish colors, which are characteristic of the presence of iron in the oxidized form (Franzmeier and Kladivko, 2001). In these soils, water is removed rapidly, especially if coarse-textured. In poorly drained soils, however, gray and dark colors reflect the presence of organic matter and silicate minerals, unraveled after iron forms were reduced or removed from the system due to the presence of water during extended periods.

Depth to the water table is another characteristic that in some cases is more useful than morphological characteristics for describing soil drainage class (Mackintosh and van der Hulst, 1979). The concept of the depth to the water table is intimately related to the sequence of soils along a slope, commonly known as catena. This approach is profoundly rooted in soil-landscape relationship, in which hillslope positions (upperslope, summit, backslope, toeslope and foot slope) are interpreted as key players in governing water movement (Schaetzl, 2013). Soils with deep water tables usually are located in plain areas where well drained soils are presented, usually on uplands and upper backslopes. Poorly drained soils, in turn, have the highest water tables experiencing longer periods of water saturation, usually located at the bottom of the slope.

The evaluation of entire landscapes and allocation of land areas into drainage classes is a complex operation involving the spatial prediction of drainage classes based on variables that are more readily observed from the surface or above. Traditionally, the creation of soil drainage class maps involves two steps: describing it in the field and then extending the knowledge observed in soil profiles to areas where it is impractical to observe the soil below the surface. In this second process, each pedologist uses his/her own mental model and knowledge of soil-landscape relationships. Earlier works for describing soil drainage class of an area or a specific location were based on expert knowledge, morphological characteristics and pedologist's experience on soil-landscape relationship (Hudson et al., 1992). An example of the use of expert knowledge is the work of Gerardin and Duereu (1990), who proposed an approach for evaluating natural soil drainage in forest mineral soils to aid non-specialists. The authors proposed a system based on the identification and measurement of several variables related to the geomorphology, topography, soil and vegetation of the area under study. The results were promising but it was a very site-specific approach and not much utilized at that time it was proposed.

While these techniques have enabled the production of valuable soil maps for more than a century, the reliance on tacit knowledge can be prone to be inconsistent among experts (Kerebel and Holden, 2013; Bouma et al., 1973). Therefore, quantitative frameworks for mapping soil drainage classes emerged to reduce, at least in part, the inconsistencies associated with tacit knowledge. These quantitative approaches used mathematical calculations or statistical tools to describe the landscape and relate it to soil drainage classes, unlike expert knowledge, which relies on an expert's opinion and experience. An example of this is the work of Troeh (1964), who developed 3D equations to obtain landscape parameters such as slope, rate of change and curvatures. He related these parameters to soil drainage and found a strong correlation, concluding that the parameters obtained from 3D equations can be used to accurately predict soil drainage classes. Bell et al. (1992), proposed a "soil-landscape model" based on multivariate discriminant analysis with the aim of identifying the subset of landscape variables that could optimally separate soil drainage classes. In their work, the authors achieved an overall agreement of 74% when comparing their predictions of soil drainage classes with those described in field surveys.

Nowadays, digital soil mapping is the most used framework to spatially predict soil drainage classes. Optical and radar remote sensing images (Cialella et al., 1997; Peng et al., 2003; Niang et al., 2012), terrain derivatives (Zhao et al., 2008), electrical conductivity (Kravchenko et al., 2002), magnetic susceptibility (Grimley et al. 2008; Asgari et al. 2018) and soil color (Malone et al., 2018) were the most common environmental covariates used to spatially describe the variability of soil drainage classes across landscapes. However, description of the soil drainage class in the field is still needed before using digital soil mapping approaches. Most of the mentioned works were based on soil drainage classes described in the field or retrieved from legacy soil maps. The work of Malone et al. (2018) is an exception, where the authors tried to define soil drainage classes based on soil color groups. These color-based soil drainage classes were then predicted using terrain attributes and machine learning. This kind of approach is not common to find in literature, but it was a first attempt to quantitatively define soil drainage classes using information on soil color gathered from legacy data.

In this research, we attempted to compare two strategies for mapping soil drainage classes in Brazil. The first was based on expert knowledge (EK), in which maps of soil color and soil texture were used to define the soil drainage class. The second strategy was based on machine learning (ML), in which soil sampling points with the labeled soil drainage class were used in conjunction with environmental covariates to spatially predict the distribution of soil drainage classes in the study area. An existing soil drainage class map was taken as reference. Our main objective was to select the best strategy for mapping soil drainage classes, considering four criteria: accuracy, labor-efficiency, transferability, and accessibility.

4.2. Material and Methods

4.2.1. Study site description

The study area is in southeastern Brazil, in the state of São Paulo, covering an area of approximately 2,574 km2 (Figure 1). It is characterized by gentle slopes, undulating hills, and rolling uplands, with elevation varying between 450 and 950 m.a.s.l. In this region, 83% of the land is used for agriculture and pasture, 9% for forest formation, savanna, and forest plantation, and the remaining corresponds to urban areas and water bodies. More than 40% of the agricultural land is cultivated with sugarcane, almost 15% with pasture, 22% with a mixture of agriculture and pasture and 3% with soybean and other perennial and temporary crops (MAPBIOMAS, 2019). The climate of the region is classified as Cwa in Koppen's climatic classification, which is characterized by dry winter and hot summer with mean annual precipitation of 1200 mm and average annual temperature of 24 °C (Alvares et al., 2013).

The study area is part of the Rio Paraná basin, where a complex geological variability is found (Figure 1). The main stratigraphic units are the Itararé, Corumbataí, Pirambóia and Serra Geral Formations. The Itararé Formation is the oldest, which was formed in the Paleozoic Era during the Permian Period. This formation is composed of sandstones, siltstones, varvites and diamictites. The Corumbataí formation, also from the Permian Period, is above the Itararé Group and is characterized by siltstones with lenses of fine sandstones, mudstones, limestones, and dolomitic levels. Pirambóia and Serra Geral formations were formed in the Mesozoic Era, during Triassic and Cretaceous Periods, respectively. The Pirambóia Formation has sandstones with rounded and spherical grains, while basalt and diabase rocks are characteristic of the Serra Geral formation. For more details about the geology of the area, the reader is referred to the work of Bonfatti et al. (2020).
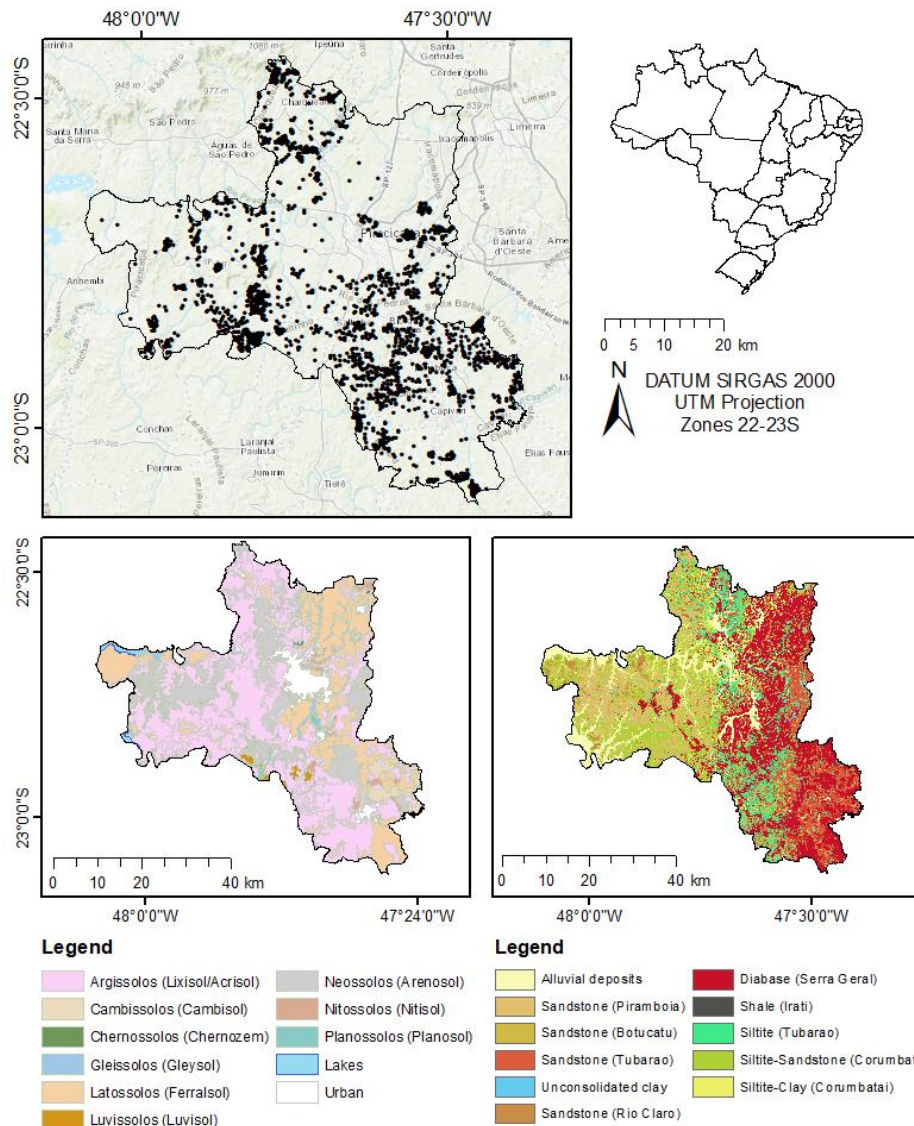
Figure 1. Study site with sampling points, soil class and geology maps. Adapted from Bonfatti et al. (2020) and Mendes (2020).

Predominating soils in the Brazilian Soil Classification System (with the corresponding WRB in parenthesis) are Argissolos (Acrisols/Lixisol), Latossolos (Ferralsol), Neossolos (Arenosols), Cambissolos (Cambisol), and Nitossolos (Nitisol) (Figure 1). Considering the soil genesis, both Argissolos and Latossolos are the oldest. The main characteristic of Argissolos is the presence of the Bt horizon, where accumulation of clay is predominant. In the study area, three Argissolos are found (PA, PVA and PV), whose main differences are related to soil color. PA (Argissolo Amarelo) has a hue of 7.5YR in the first 100 cm of the B horizon, PV (Argissolo Vermelho) has a hue of 2.5YR, and PVA (Argissolo Vermelho-Amarelo) has a hue varying from red to yellow, between the other two categories. Latossolos are the most weathered and well-developed soils. A homogeneous soil profile is characteristic of this soil order, with a strong presence of iron oxides and kaolinite. Four Latossolos can be found in our study area: LH (Latossolo humico), LA (Latossolo Amarelo), LV (Latossolo Vermelho), and LVA (Latossolo Vermelho-Amarelo), whose differences are also interpreted from soil color, except for LH, which has higher organic matter content than the others. Neossolos and Cambissolos are poorly evolved soils, with no B horizon (Neossolos) or an incipient one (Cambissolos). Two suborders of Neossolos are found in the study area: RQ (Neossolo

Quartzarenico), which is a sandy soil and RL (Neossolo litolico), which has a lithic contact in the first 50 cm. Finally, Nitossolos are similar to Latossolos but have a B horizon with waxiness or clay coatings with a granular structure and a clayey or very clayey texture. Nitosolos Vermelhos (NVf) with high presence of iron oxides are common in our study area. The second type of Nitossolo is the NX (Nitossolo Háplico), which is a soil class that does not fit in the other categories of this soil order.

4.2.2. Strategies for mapping soil drainage classes

In this research, our main objective was to select the best strategy for mapping soil drainage classes. The two strategies were based on expert knowledge (EK) and machine learning (ML). We used an existing soil drainage class map as reference, which was compared to the resulting maps of both strategies. The best map was selected based on a set of criteria related to the mapping procedure. A throughout detail of each step is described in the next sections.

4.2.2.1. Reference soil drainage class map

An existing soil drainage class map was used as reference for comparing the two alternative estrategies (Figure 2). This reference map was produced by taking soil classes that have been previously mapped by Mendes (2020) and relating them with drainage classes. The map produced by Mendes was obtained following a digital soil mapping framework, obtaining an accuracy of 0.84. The relationship between drainage class and soil class can be found in Table S1 of the supplementary material. Although the Brazilian Soil Classification system includes eight drainage classes, the reference soil drainage map uses only five classes to represent drainage ranging from very well (Class 1) to poorly drained (Class 5). This map shows that the study area is predominantly (39%) Class 2 (well drained), followed by Class 3 (moderately drained) with 37% of the area. Drainage classes 1, 4, and 5 represent 13%, 9% and 2% of the area, respectively.
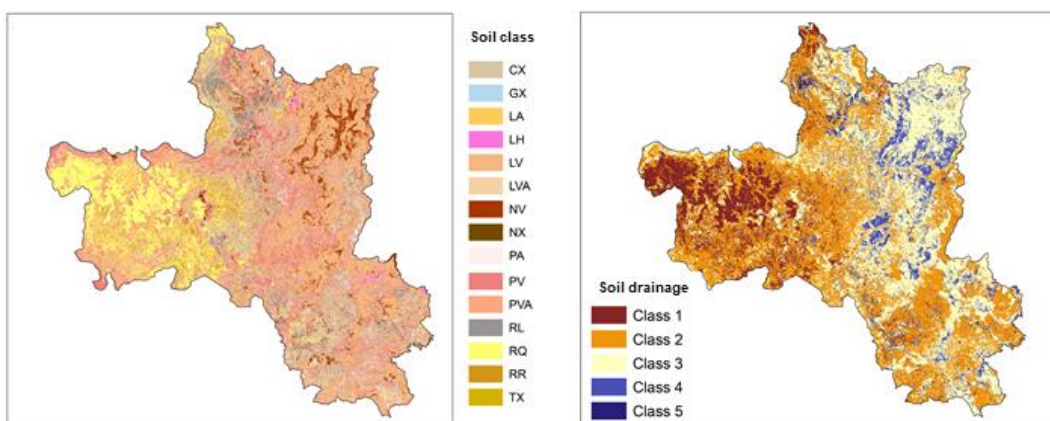
Figure 2. Soil class map and the resulting soil drainage class map used as reference (from Mendes (2020)). Cambisols (CX), Gleisols (GX), Ferralsols (LH, LA, LV, LVA), Nitisols (NV, NX), Acrisol/Lixisol (PA, PV, PVA), Arenosol/Regosol (RL, RQ, RR), Planosol (TX).

4.2.2.2. Soil drainage mapping by expert knowledge using color and texture (EK)

The strategy based on EK used soil color and textural classes as base maps. It is well known that soil color is a good indicator of drainage conditions. In general, red and yellow colors are associated with the iron oxidation process in soil environments where oxygen is abundant, indicating a very well to moderately drained soil. In other words, the soil is not permanently flooded, and infiltration conditions are optimal. On the other extreme, low chroma values or gleying colors are indicators of poor drainage. This coloration is caused by water staying longer in the soil, favoring the reduction of iron. Soil textural classes, in turn, were used because they play a key role in governing water movement through the soil. Water will infiltrate through sandy soil more quickly than through clayey soil.

Soil color (hue, value and chroma) and texture maps were obtained from Mendes et al. (2021), who predicted these properties for three layers (A = 0-20cm, B = 40-60cm and C = 80-100cm). We used the A layer map as the topsoil and the B and C layers were averaged to represent the subsoil. The soil color maps were visually classified into five groups (red, brown, yellow, gray, and black) to represent the sequence of five soil drainage classes, varying from well to poorly drained soils, similar to the approach presented in Malone et al. (2018). The textural class maps were obtained after classifying clay content raster maps in five classes: very sandy (<150 g kg^{-1}), sandy (150-250 g kg^{-1}), loam (250-350 g kg^{-1}), clayey (350-600 g kg^{-1}) and very clayey (>600 g kg^{-1}).

The classified soil color and textural maps were combined in ArcGIS Pro using the "combine" tool to join the maps and define a numerical code to represent the color and textural group. For example, a red soil with clayey texture had the code 1/5 while a gray soil with a clayey texture had the code 4/5. With this procedure, we ended up with a combination of possible 24 texture/color code for each layer (topsoil and subsoil, Table S1 of the supplementary material). After this procedure, a drainage class was assigned to each texture/color code by an expert pedologist for each layer. Both layers were then considered to define the final drainage class. An example is given below in Figure 3. More details can be found in Table S2 of the supplementary material.

Color	Code	Textural classes	Code	Drainage
Red	1	Very sandy (Clay content <150 g kg^{-1})	1	Class 1 (very well drained)
Brown	2	Sandy (Clay content <250 g kg^{-1})	2	Class 2 (well drained)
Yellow	3	Loam (250 > Clay content <350 g kg^{-1})	3	Class 3 (moderately drained)
Gray	4	Clayey (350 > Clay content <=600 g kg^{-1})	4	Class 4 (imperfectly to poorly drained)
Black	5	Very clayey (Clay content > 600 g kg^{-1})	5	Class 5 (very poorly drained)

C/T topsoil	C/T subsoil	Drainage Topsoil	Drainage Subsoil	Final drainage
1/5	1/5	Class 3	Class 3	Class 3

C/T: Color/texture code
Code 1/5: red color and very clayey soil
Class 3 (drainage): moderately drained

Red soil with very clayey texture (1/5) in both topsoil and subsoils
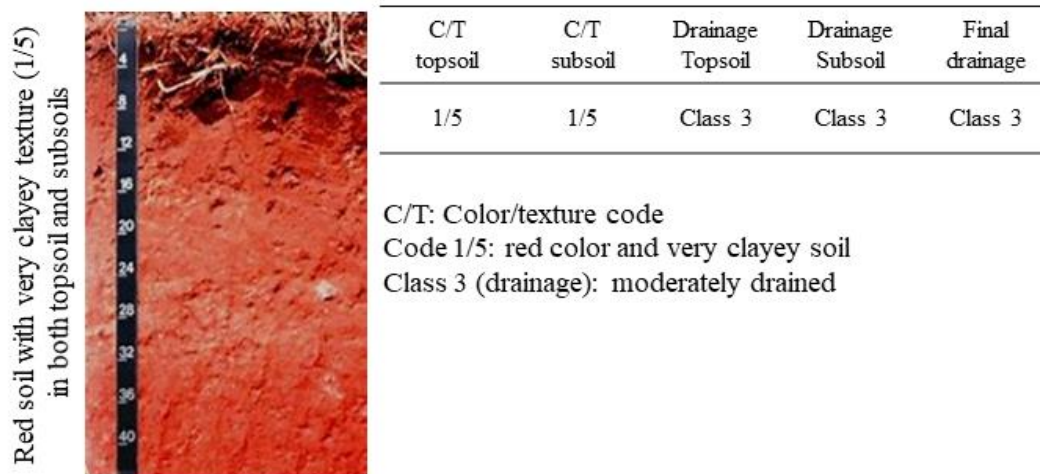
Figure 3. Example of allocation of drainage class based on soil color and texture of the topsoil and subsoil.

4.2.2.3.   Soil drainage class mapping by machine learning (ML)

The second strategy was based on mapping soil drainage class by ML. We used a soil dataset with 2,309 locations observed by the Geotechnologies in Soil Science Group between 2000 and 2020. Each sample point had its respective soil drainage class determined by extracting the value from the reference map for the respective locations. The strategy based on ML consisted of using environmental covariates (terrain, soil, and vegetation), the sampling points with the labeled soil drainage class and a decision tree algorithm (C5.0) to build a model for predicting soil drainage classes. Unlike the EK, the ML-based strategy is not relying on using only two soil data (color and texture) but instead a set of variables that seek to represent processes related to the distribution of soil drainage classes. An explanation of each step is given below.

4.2.2.3.1. Environmental covariates

A total of 603 environmental covariates were used in a digital soil mapping framework. All terrain derivatives were calculated in SAGA GIS, GRASS GIS and ArcGIS Pro. In this pool of covariates, the majority (552) were terrain derivatives at different analysis scales, obtained from a digital elevation model with a resolution of 5 m. Briefly, some terrain derivatives are scale dependent and calculated in a GIS with a moving window. The analysis scale consists of the resolution cell size, multiplied by the quantity of cells defining the width of the moving

window. The 3x3 cell moving window is the most used window size, but here we considered analysis scales with analysis scales ranging from 15 to 5,070 m. Eight terrain derivatives - including slope gradient, northness, eastness, relative elevation, as well as profile, plan, cross-sectional, and longitudinal curvatures - were obtained at different analysis scales, totaling 69 raster layers for each. More details about the use of covariates with multiple analysis scales can also be found in Miller et al. (2012).

In addition, other terrain derivatives that either do not use a moving window or could not have the analysis window size varied were included. These terrain derivatives were vertical distance to channel network (VDCN), topographic wetness index (TWI), mid-slope position (mslp), and multi-resolution valley bottom flatness (MRVBF). Remote sensing products from Sentinel-2A and Landsat were also included in the covariate stack. For the case of Sentinel-2A, images of the dry (April to October) and moist seasons (October to April) from 2015 to 2020 were obtained and averaged in the Google Earth Engine platform, after atmospheric correction was performed. The Normalized Difference Vegetation Index (NDVI) and Normalized Difference Red-Edge Index (NDRE) were calculated from the S2A images, also for both seasons. A Synthetic Soil Image (SYSI) was obtained in the Google Earth Engine platform following the methodology proposed by Demattê et al. (2018). This SYSI is a median reflectance image with six spectral bands, which represents the bare soil pixels in the study area.

Table 1. Environmental covariates used for modeling soil drainage classes

Environmental covariates	Abbreviations	Analysis scales	Selected
Slope gradient	slp	15-5070m	slp_075m
Profile curvature	prc	15-5070m	prc_115m
Plan curvature	plc	15-5070m	plc_1110m
Northness	nnes	15-5070m	nnes_3630m
Eastness	enes	15-5070m	enes_3630
Cross-sectional curvature	ccurv	15-5070m	ccurv_115m
Longitudinal curvature	lcurv	15-5070m	lcurv_115m
Relative elevation	rel	15-5070m	rel_3090m
Topographic wetness index	topidx	NA	YES
Mid-slope position	mslp	NA	YES
Vertical distance to channel network	VDCN	NA	YES
Multi-resolution valley bottom flatness	MRVBF	NA	YES
Soil color subsoil and topsoil (6 variables)	(Hue, value and chroma)	NA	YES
Sentinel-2A bands dry season (9 bands)	S2A dry	NA	S2A_dry_swir1
Sentinel-2A bands moist season (9 bands)	S2A moist	NA	S2A_moist_swir1
Sentinel-2A Normalized Difference Vegetation Index (dry and moist season)	S2A NDVI dry S2A NDVI moist	NA	S2A_NDVI_moist
Sentinel-2A Normalized Difference Red-Edge Index (dry and moist season)	S2A NDRE dry S2A NDRE moist	NA	S2A_NDRE_dry
Synthetic Soil Image (6 bands)	SYSI	NA	SYSI_swir2

Not all environmental covariates were used for spatial modeling, however. A Spearman rank correlation analysis was performed to select those that had the highest correlation with the soil drainage class. Considering the raster layers that were obtained at different analysis scales, only the one with the highest correlation was selected. For example, among the 69 raster layers of slope gradient at different analysis scales, the slope gradient at 115m (slp_115m, window size of 15x15 cells) was chosen. The same was performed for all terrain derivatives with multiple analysis scales, satellite images, and the SYSI. The final covariate stack consisted of 23 raster layers. The environmental covariates that were selected and used in the modeling process are shown in Table 1.

4.2.2.3.2. Building and validating the ML model

The soil dataset was partitioned into two sets. Seventy percent of the data were used for model building (1,616 samples), while the remaining soil samples (30%, 693 samples) were used for model validation. The C5.0

algorithm was used, which is a decision tree method based on the C4.5 and ID3 algorithms developed by Quinlan (1986). It is a supervised machine learning algorithm that develops a classification model using the training data with known labeled classes to determine the best rule that splits the data into pure classes. This algorithm has the advantage of being easy to implement and interpret the results. For validating the model, the test dataset was used, and the drainage classes were predicted for each of these sampling points. The accuracy of the model, which is the quantity of correct predictions divided by the total number of predictions, was calculated using the observed and predicted soil drainage classes to build a confusion matrix. Besides the accuracy, the kappa index, as well as the user's (UA) and producer's accuracy (PA) were also calculated. The PA represents the accuracy from the point of view of the map maker. In other words, it represents how often the drainage class observed in the field is correctly shown on the classified map. The UA, also known as reliability, represents the accuracy from the point of view of the map user. Finally, a map was obtained by predicting the soil drainage classes for the entire study area. Model building, validation and prediction were implemented in the R software using the C5.0, caret, and raster packages.

To evaluate the environmental covariates that were considered the most essential to predict drainage patterns in our study area, we used Shapley's value. This statistic is recommended for improving the interpretability of "black box" ML algorithms (Lundberg and Lee, 2017). The Shapley's value is based on game theory, in which members of a coalitional game should receive payments proportional to their marginal contributions. In our case, the game is the prediction of values, and the players are the environmental covariates used in the model building. Therefore, the Shapley's value represented the marginal contribution of each variable to the prediction of soil drainage classes.

4.2.3. Comparison between the reference soil drainage map and EK and ML strategies

The soil drainage maps obtained from the EK and ML strategies were compared with the reference soil drainage map in two ways. Firstly, the soil drainage class at each location of the sampling points was extracted from the three maps and two confusion matrices were built. After extracting the soil drainage class for each sampling point, it was observed that Class 5 was not present in the ML-based soil drainage map and therefore, this class was excluded from the comparisons.

Secondly, a comparison was performed on a cell-by-cell basis by subtracting the EK and ML-based maps to the reference map, to represent their agreement. The product was two difference maps, representing how much the EK and ML-based drainage maps differed from the reference data. The difference maps had negative and positive values. A negative value meant that the drainage class obtained by either EK or ML was higher (i.e., wetter) than the reference map. A positive value, in turn, represented a lower (drier) drainage class. For example, if at a specific cell the value of the subtracted map is -2, it means that the strategy for mapping soil drainage class predicted a Class 3 (moderately drained) soil drainage, where the reference had a Class 1 (very well drained). The reverse applies for positive values.

4.2.4. Selecting the best strategy for mapping soil drainage classes

Our objective with this research was to select the best strategy for mapping soil drainage classes in tropical environments. For this, each map was evaluated considering four important criteria for the overall soil map production process: accuracy, labor-efficiency, transferability, and accessibility. A value of 1 (good), 2 (moderate) or 3 (poor) was assigned to each criterion. These values were assigned considering our experience in obtaining base maps for strategies based on EK, or the use of ML to obtain the soil drainage map.

For the first criterion (accuracy), it was evaluated the model performances expressed as R^2 or kappa values. The EK strategy used as base maps soil color and texture, whose accuracies in R^2 were reported in Mendes (2020) and Mendes et al. (2021). The accuracy for the ML-based strategy was reported in this study. Accuracies higher than 0.6 received a value of 1 while accuracies between 0.4-0.6 and lower than 0.4 received the values of 2 and 3, respectively. The labor-efficiency was considered as the ease or difficulty in obtaining data for each strategy. In other words, if the base maps or environmental covariates are easier to calculate or to obtain from other sources. Because the ML strategy required a higher level of knowledge to process and calculate the environmental covariates from the DEM, it received a higher value when compared to the EK strategy. We argue that maps of soil color and texture are now easily available in platforms such as SoilGrids, for example. Although the resolution of the base maps might to be sufficient for a specific area, it is a straightforward way to have an idea of the variability of texture and color.

Transferability was evaluated in an analogous way as the labor-efficiency criterion. This means that if a strategy was considered as easy to be transferred to any other region like to the conditions of our study area, the value would be the lowest. In other words, the transferability would be higher under this assumption. For our case here, the EK strategy was considered as having the lowest value, because it is based on simple rules using only two soil properties, which can be easily applied to any other area. Conversely, the ML strategy is more complex, as it used a set of environmental covariates and a decision tree model to reveal patterns that would help explain the variability of soil drainage classes in our study area. Therefore, the value assigned to this criterion for the ML strategy was the highest, considering that the transferability of ML models is limited by the feature space (Meyer and Pebesma, 2021).

Finally, to evaluate accessibility, it was considered how easy it is to understand or use the strategies. For the ML-based strategy, the highest value was assigned because ML models are still considered black boxes. Although we used a simple decision tree to simplify the process, the selection of covariates that would have more importance in the model and how this process happens inside the model is still not well understood. The EK strategy is a set of rules based on soil color and texture groups, whose concept is simple to understand, as it is rooted in well-defined rules regarding the role that clay content and color play in soil drainage.

4.3. Results

4.3.1. Retrieving soil drainage classes by EK and ML

The strategy based on EK used topsoil and subsoil color and textural classes for building the numerical code and map drainage classes. These maps and numerical codes are presented in Figure S2 and Tables S2-S3 of the supplementary material. Grayer colors were more commonly found in the subsoil than in the topsoil. The most common color found in the topsoil (represented in the Munsell system) was the 10YR 6/4 (hue value/chroma), with more than 25% of the area. This color in the subsoil represented only 0.1%. Other common colors were 7.5YR 6/4 and 5YR 4/6, with 21 and 16% of the area, respectively. In depth, we observed that the 10R 6/2 Munsell color was the most common, which together with 10R 4/2 cover almost 30% of the area. The 5YR 4/6 color was also present in the subsoil, with more than 20% of coverage. In some areas (~17%), a yellower color was observed (7.5YR 6/6) in the subsoil, which is closely similar to the 7.5YR 6/4 found in the topsoil. The textural class maps for the topsoil and subsoil were highly variable across the study area. In the topsoil, mostly very sandy (36%) to sandy texture (37%) were found. In the subsoil, although the sandy texture showed a high percentage (37%), the loam and clay texture together represented more than 55%. Very clayey textures, in which the clay content is above 600 g kg^{-1} are uncommon in the study area, with just 0.1% and 0.3% in the topsoil and subsoil, respectively (Figure S2).

The combination of these colors and textural classes produced a soil drainage map in which 58% of the area had the Class 3 (moderately drained), which was mostly placed on areas with loam to clayey texture. The second most common drainage found was Class 2 (well drained), which was characterized as having brownish colors and sandy to loam textures. This soil drainage class covered approximately 29% of the area. Class 1 and Class 4 were in less than 10% of the area and the Class 5, which represented poorly drained soils, was even more poorly distributed, just in the areas where low chroma or glei colors were present, especially near the rivers.

The strategy based on ML provided information on the accuracy of the model and the contribution of the environmental covariates. These results are presented in Table S4 and Figure S3 of the supplementary material. In general, the ML model had a satisfactory kappa value (0.56). The PA and UA values for almost all soil drainage classes were above 50% (except for Class 5), which can be considered fair. Regarding the importance of environmental covariates, we found that the shortwave infrared 2 band of the bare soil image (SYSI_swir2) was the one that contributed the most to explain the distribution of soil drainage classes. The second and third variables that contributed the most were two soil color components: hue and value from the subsoil. The value component from the topsoil was the fifth variable that contributed the most to the model while the topographic wetness index occupied the fourth place. Variables such as the chroma component from the topsoil, the SWIR1 band from Sentinel2, planform curvature at 115m analysis scale and VDCN had an equal contribution to the model in general and for each class separately. The longitudinal curvature (lcurv_115m) and the profile curvature (prc_115m) had a negligible contribution to the model, similar to the chroma component form the subsoil and the slope gradient at 75m analysis scale (slp_075m).

The resulting ML-based soil drainage map had the Class 1 (very well drained) as the most common, with 33% of the area, followed by Class 2, Class 3 and Class 4, with 24%, 22% and 20% of the area, respectively. The Class 5 (poorly drained) was barely represented in just 0.11% of the area.

4.3.2. ML- and EK-based soil drainage maps and comparison with reference data

The difference maps represented the values obtained after subtracting the reference map from the predicted EK- and ML-based drainage maps (Figure 4). The higher the negative or positive values, the higher the difference between the maps. The negative values indicate areas where the estimated ML or EK-based map is predicting a higher value. In other words, the negative values represent areas where both strategies are predicted wetter drainage classes than those observed in the reference drainage map. For example, the -1 is indicating that the reference map could be Class 1 to 4 while the EK or ML-based drainage maps were between Class 2 and Class 5, being just one level higher. The same applied for positive values.
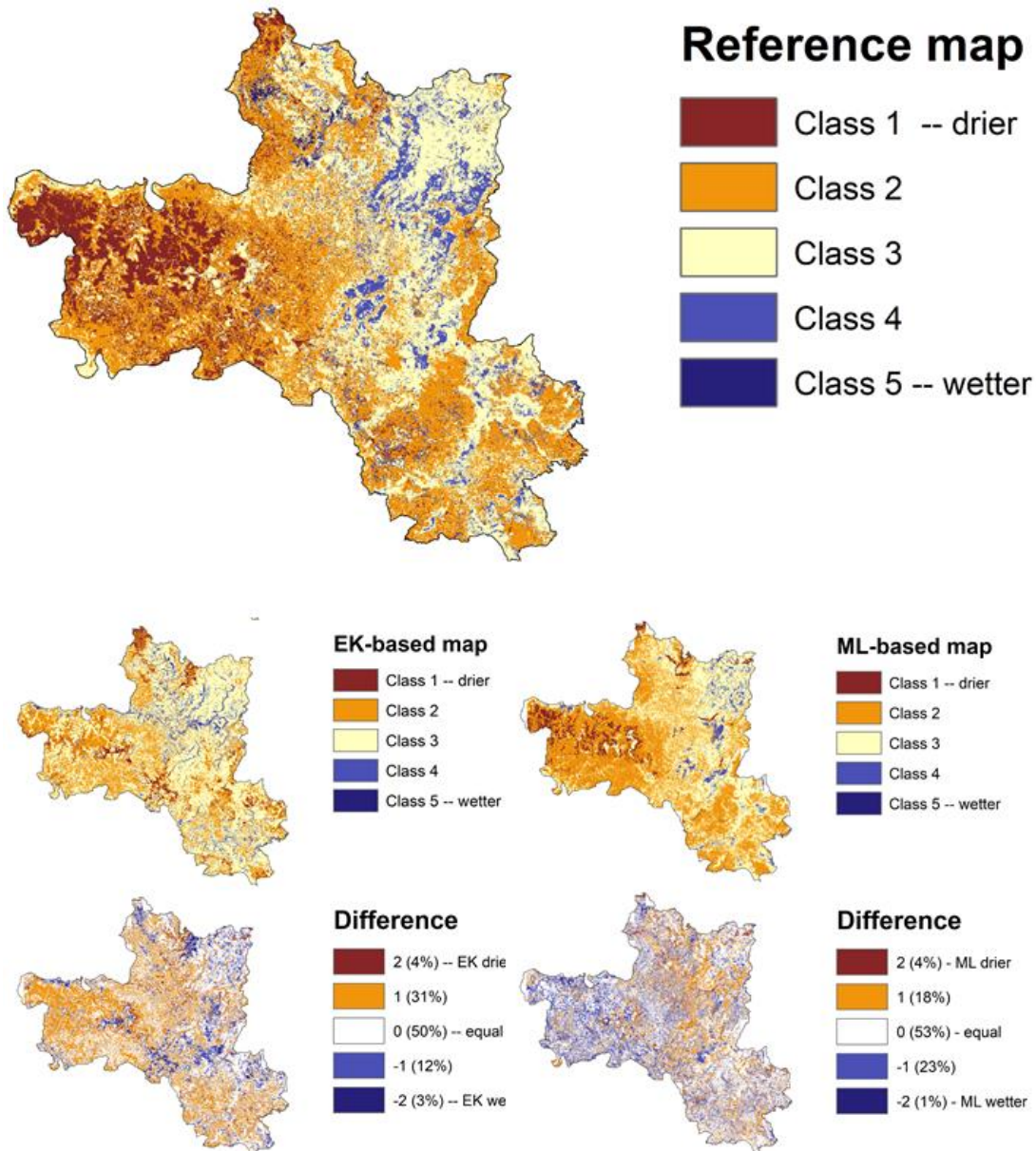
Figure 4. Example of allocation of drainage class based on soil color and texture of the topsoil and subsoil

We found that both strategies had an agreement of ~50% with the reference map. The drainage map obtained by ML had the highest agreement (53%) and ~41% was between ±1 Class (values 1 and -1 values in Figure 4). The difference map of the reference and the EK map revealed that the EK-based strategy is more prone to underestimate the drainage class, especially in sandy soils. Our results revealed that for the EK strategy, 43% of the predicted drainage classes were between ±1 class from the drainage class observed in the reference map (values 1 and -1 in Figure 4).

The confusion matrix built using the sampling points is presented in Table 2. Because Class 5 was not present in any of the sampling points, it was excluded from this confusion matrix, reducing the total number of samplings from 2309 to 2246. Between the reference sampling points and the ML-based strategy, we found an accuracy of 0.67 and kappa value of 0.5. For the EK-based strategy, the accuracy and kappa values were slightly lower (0.57 and 0.33, respectively).

Table 2. Confusion matrix between the reference drainage map and the resulting drainage maps of the two strategies.

		Reference soil drainage map					
		Class1	Class2	Class3	Class4	Total	UA
ML-based drainage	Class1	109	22	5	0	136	80%
	Class2	159	753	194	41	1147	66%
	Class3	23	154	498	99	774	64%
	Class4	0	1	41	147	189	78%
	Total	291	930	738	287	2246	
	PA	38%	81%	68%	51%		
	Accuracy	0.67					
	kappa	0.50					
		Class1	Class2	Class3	Class4	Total	UA
EK-based drainage	Class1	21	28	11	13	73	29%
	Class2	233	587	138	42	1000	59%
	Class3	37	312	578	146	1073	54%
	Class4	0	3	11	86	100	86%
	Total	291	930	738	287	2246	
	PA	7%	63%	78%	30%		
	Accuracy	0.57					
	kappa	0.33					

The users' (UA) and producers' accuracy (PA) were also calculated, in which high values represented high agreement between the maps. The PA is associated with the error of misclassification while the UA is dedicated to the error of over-classification. They help to describe how much the predicted soil drainage class is the same as the reference map. For example, the Class 3 of the ML-based strategy had PA and UA of 68% and 64%, respectively. This means that 68% of the reference "Class 3" was correctly identified as "Class 3", but 64% of the areas identified as "Class 3" in the classification were actually "Class 3". Considering the same example for the EK-based strategy, the PA and UA for "Class 3" were 78% and 54% respectively, meaning that the ML-based strategy was better in predicting this drainage class.

4.3.3. Selecting the best strategy for mapping soil drainage classes

Table 3 depicts the scoring for each of the four criteria used to select the best strategy for mapping soil drainage classes. The first criteria, accuracy, was easy to obtain from the base maps used in the EK-based strategy and from the ML model presented in this work. Soil color and texture maps used for the EK strategy had an R^2 of 0.65 while the kappa value of the ML strategy was 0.57. These accuracies received a value of 2, as explained in the methods section. Regarding transferability of these both strategies, the EK-based soil drainage map received a value of 2, considering the easiness or not to transfer or to repeat the strategy to develop these maps. A value of 3 was given to the ML-based soil drainage map since it requires more data and machine learning models sometimes are

more difficult to replicate. For the labor-efficiency it was considered the data needed to build each strategy and the effort to obtain this data. For both EK and ML a value of 2 was attributed because they require more data. Finally, the accessibility considered the way in which this data can be obtained. For example, for the ML map the highest value was assigned because it can be more difficult to understand because machine learning models are still considered black boxes. The rules for mapping soil drainage classes based on expert knowledge using soil color and texture maps had a lower value, because we understand that the base maps used to build these strategies are easier to obtain from soil databases or legacy data.

Table 3. Scoring values for each criterion considered for selecting the best strategy

Criteria	EK	ML
Accuracy	2	2
Transferability	2	3
Labor-efficiency	1	2
Accessibility	1	3

4.4. Discussion

The two strategies presented here for mapping soil drainage classes in Brazil reached a reasonable result, represented by approximately 50% agreement of the strategies with the reference map. The EK-based soil drainage map obtained from soil color and texture could appear subjective at some point. The advantage of this strategy is that it relies on two easily obtainable data layers. For this reason, an EK-based strategy was a suitable first attempt to represent spatial patterns of drainage class. Both properties are useful in the way that soil color is a result of the drainage, and the soil texture can determine how fast or slow the water moves through the soil profile.

Several researchers have demonstrated the importance of soil color for mapping soil drainage because it can reveal insights into the local hydrological regime (Schoonover and Crim, 2015). The work of Malone et al. (2018) is an example, where color descriptions found in legacy soil maps were used to describe and spatially predict a soil drainage index in New South Wales, Australia. They used fuzzy set theory to allocate soil colors in pre-defined five color groups (red, brown, yellow, gray and black), which represented drainage classes ranging from 1 to 5 (very well to poorly drained). The authors argued that their soil drainage index incorporated tacit knowledge and that it can be used in other areas since the soil color/soil drainage relationships were well defined by an expert.

The main disadvantage of using expert knowledge to define soil drainage classes is that in some cases it can be difficult to replicate. However, it has the advantage of being based on soil data that might be more accessible. Maps of soil properties such as soil texture can be easily obtained from soil databases such as the Soilgrids platform (Poggio et al., 2021). Soil color, in turn, can be easily described from legacy data or estimated from remote sensing images or soil spectral data (Poppiel et al., 2020; Silvero et al., 2021). However, the rules defined here may not completely apply to other areas and this limitation needs to be considered when aiming to use the strategy based on EK presented here. The description of soil drainage classes based on soil color can not be useful in all cases, however. This was observed in the work of Kidd et al. (2014), who tested the use of soil color, indirectly by

measuring the depth to mottling. They found that direct description of soil drainage classes in the field were better than using the depth to mottling as a surrogate or proxy of the soil drainage conditions.

An example of an earlier work using EK to define soil drainage classes was presented by Geradin and Duerue (1990), who aimed to build an objective approach to be easily accessible by non-specialists and easily applied to forest soils. They used topographical variables (hillslope position, terrain shape, effective upper slope position and declivity), abundance and mottling depth, solum depth and humus thickness to build rules and allocate soil drainage classes considering the variability of these parameters. Although particularly useful, this approach was probably not so much used because it appears overly site-specific and, at some point, difficult to replicate.

The result obtained with the ML-based strategy was as satisfactory as that obtained with the EK strategy but the agreement with the reference map was higher (53%). The accuracy with this approach was 0.66, which is considered satisfactory and can be improved if a better representativity of soil samples in poorly drained areas (Class 4 and Class 5) is used. In our case, the poorly drained areas were poorly represented with just 10% of the samples. The main advantage of using digital soil mapping approaches based on machine learning is that a set of environmental covariates are used as explanatory variables of the spatial distribution of drainage classes in the area. Although it was previously highlighted that the definition of soil drainage classes from soil color can still be subjective, we found that soil color variables such as hue and value were one of the variables that most contributed to the model based in ML, even more than those environmental variables that were considered to have a direct relationship with water movement, such as topographic wetness index (topidx) and multi-resolution valley flatness bottom (MRVFB).

This kind of approach using ML is the most common framework used for mapping soil drainage classes nowadays. Its main limitation is that before modeling, data on soil drainage class obtained from field surveys are needed. In our case this was a drawback because we could not find a proper description of the main drainage conditions and used the reference data to predict soil drainage classes using digital soil mapping.

Several types of environmental covariates were used ranging from soil-landscape relationships and information from proximal sensing such as magnetic susceptibility. One of the earlier works using prediction models was the work of Bell et al. (1994), who related some landscape attributes to obtain soil drainage class probability. They also found some difficulty in mapping poorly and moderately drained areas. In a similar research, Levine et al. (1993) studied the relationship of environmental covariates with soil drainage classes and found that they were positively correlated with values of NDVI, being the highest values associated with well drained soils. This is an indication of the conditions of the site for plant growth as poorly drained soils do not provide a suitable rhizosphere environment for most roots, but they are more fertile due to the accumulation of organic matter. Zhao et al. (2008) used topographic and hydrological variables and artificial neural networks to build a model for soil drainage classification. They used data from coarse resolution maps and aimed at improving the resolution of the soil drainage map by using machine learning and digital soil mapping frameworks. A 52% agreement was found with the reference data while 94% of the predicted drainage classes were within +-1 class. Among the variables used, they found a strong relationship between the drainage classes and the vertical slope position, which is similar to the mid-slope position (mslp) used in this work. Unlike these results, the contribution of the mslp on our model was low. Beucher et al. (2019) also used artificial neural networks (ANN) for predicting soil drainage classes in Denmark and compared its performance to that obtained by a decision tree classification (DTC) model. They used 31 covariates, which included topographical, hydrological, and soil variables. They found that the ANN outperformed the DTC, but the difference was just 2%, reporting an overall accuracy of 54 and 52%, respectively. Considering the contribution of

the environmental covariates, they found that the clay content in depth (100-200cm) was the one that contributed the most. Among the ten most important variables we also found slope to channel network, geology, vertical and horizontal distance to channel network, wetness index and depth to groundwater.

It is worth mentioning that the results obtained with the DTC model were similar to those reported here. It was our interest to use the simplest model, in which the rules are easy to follow and to comprehend. This was achieved with the C5.0 algorithm, also a decision tree model, which uses information ratio gain to build the rules. Cialella et al. (1997) also predicted soil drainage classes by using remote sensing data and topographical variables in a decision tree algorithm. They achieved satisfactory results, with an average accuracy of 78% and topography being the most correlated variable with the drainage classes followed by NDVI. In our case, the model revealed that the most important variable was the shortwave infrared band 2 (SWIR2) of the synthetic soil image (SYSI). This happened probably because the SWIR2 band is related to the variability of clay content, whose absorption feature due to minerals is centered at 2200 nm (Gomez et al., 2018)

Our final analysis is about selecting the best strategy for soil drainage class mapping in tropical environments. After considering four criteria, the best map was that obtained from expert knowledge using soil color and soil texture as inputs. This was selected because the strategy had the lowest value in the evaluation criteria, considering accuracy, transferability, labor-efficiency, and accessibility/transparency. This strategy was selected because the maps from which it was derived had a satisfactory accuracy (R2 >0.6), are easy to obtain from soil databases or legacy soil maps, the rules to define the drainage classes are easy to understand and replicate because they are based on just two properties, which makes this strategy more transparent and accessible. We are advocating for a simple strategy to map soil drainage classes in tropical environments based on expert knowledge and soil properties; however, the disadvantages presented here for each strategy and others not covered needs to be considered when choosing one of them.

4.5. Conclusions

In this research, EK and ML strategies were tested and compared for their ability to produce soil drainage class maps in a complex landscape with limited field observations. Both strategies achieved an agreement with the reference map of about 50%. The soil drainage class map based on ML presented a higher agreement with the reference map than the EK approach. However, considering the criteria defined for selecting the best strategy, the soil drainage map based on EK was more suitable. This strategy was simpler because it used only two variables that can be easily obtained from a soil database or estimated from remote and proximal sensors, in addition to having only a few rules that are easy to apply. Therefore, if no data on field survey or legacy descriptions in soil maps are available, soil drainage class can be accurately obtained from just two soil properties: soil color and soil texture.

The ML-based strategy appears suitable when a deeper understanding of the variables that affect the spatial variability of drainage classes are needed. The ML approach can take advantage of environmental covariates that have the potential to explain relationships with processes. Future works can focus their attention on revealing these key processes.

Acknowledgments

References

Asgari, N., Ayoubi, S., Demattê, J.A.M., 2018. Soil drainage assessment by magnetic susceptibility measures in western Iran. Geoderma Reg. 13, 35–42. https://doi.org/10.1016/j.geodrs.2018.03.003

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

Bell, J.C., Cunningham, R.L., Havens, M.W., 1992. Calibration and validation of a soil-landsape model for predicting soil drainage class. Soil Sci. Soc. Am. J. 56, 1860–1866. https://doi.org/10.2136/sssaj1992.03615995005600060035x

Bell, J.C., Cunningham, R.L., Havens, M.W., 1994. Soil drainage class probability mapping using a soil-landscape model. Soil Sci. Soc. Am. J. 58, 464. https://doi.org/10.2136/sssaj1994.03615995005800020031x

Brevik, E.C., Calzolari, C., Miller, B.A., Pereira, P., Kabala, C., Baumgarten, A., Jordán, A., 2016. Soil mapping, classification, and pedologic modeling: History and future directions. Geoderma 264, 256–274. https://doi.org/10.1016/j.geoderma.2015.05.017

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/10.1016/j.geomorph.2020.107305

Bouma, J., 1973. Use of physical methods to expand soil survey interpretations of soil drainage conditions. Soil Sci Soc Am Proc 37, 413–421. https://doi.org/10.2136/sssaj1973.03615995003700030030x

Cialella, A.T., Dubayah, R., Lawrence, W., Levine, E., 1997. Predicting soil drainage class using remotely sensed and digital elevation data. Photogramm. Eng. Remote Sens. 63, 171–178.

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

Fausey, N.R., 2005. Drainage, surface and subsurface, in: Hille, D. (Ed.), Encyclopedia of Soils in the Environment. Elsevier, pp. 409–413.

Fitzgerald, J.B., Brereton, A.J., Holden, N.M., 2008. Simulation of the influence of poor soil drainage on grass-based dairy production systems in Ireland. Grass Forage Sci. 63, 380–389. https://doi.org/10.1111/J.1365-2494.2008.00637.X

Franzmeier, D.P., Kladivko, E.J., 2001. Drainage and Wet Soil Management Drainage and Wet Soil Management Wet Soils of Indiana Wet Soils of Indiana.

Gerardin, V., Duerue, J.P., 1990. An objective approach to evaluating natural drainage of forest mineral soils for non-specialists. Vegetatio 87, 127–133. https://doi.org/10.1007/BF00042949

Gomez, C., Adeline, K., Bacha, S., Driessen, B., Gorretta, N., Lagacherie, P., Roger, J.M., Briottet, X., 2018. Sensitivity of clay content prediction to spectral configuration of VNIR/SWIR imaging data, from multispectral to hyperspectral scenarios. Remote Sens. Environ. 204, 18–30. https://doi.org/10.1016/j.rse.2017.10.047

Graça, J., Daly, K., Bondi, G., Ikoyi, I., Crispie, F., Cabrera-Rubio, R., Cotter, P.D., Schmalenberger, A., 2021. Drainage class and soil phosphorus availability shape microbial communities in Irish grasslands. Eur. J. Soil Biol. 104, 103297. https://doi.org/10.1016/j.ejsobi.2021.103297

Grimley, D.A., Wang, J.-S., Liebert, D.A., Dawson, J.O., 2008. Soil Magnetic Susceptibility: A Quantitative Proxy of Soil Drainage for Use in Ecological Restoration. Restor. Ecol. 16, 657–667. https://doi.org/10.1111/j.1526-100X.2008.00479.x

Hudson, H.D., 1992. The soil survey is paradigm-based science. Soil Sci. Soc. Am. J. 56, 836–841

Kidd, D.B., Malone, B.P., McBratney B, A.B., Minasny, B., Webb, M.A., 2014. Digital mapping of a soil drainage index for irrigated enterprise suitability in Tasmania, Australia. Soil Res. 52, 107–119. https://doi.org/10.1071/SR13100

Kravchenko, A.N., Bollero, G.A., Omonode, R.A., Bullock, D.G., 2002. Quantitative mapping of soil drainage classes using topographical data and soil electrical conductivity. Soil Sci. Soc. Am. J. 66, 235–243. https://doi.org/10.2136/sssaj2002.2350

Kerebel, A., Holden, N.M., 2013. Allocation of grass fields to Hybrid Soil Moisture Deficit model drainage classes using visual indicators. Soil Tillage Res. 127, 45–59. https://doi.org/10.1016/J.STILL.2012.04.004

Levine, E.R., Knox, R.G., Lawrence, W.T., 1994. Relationships between soil properties and vegetation at the Northern Experimental Forest, Howland, Maine. Remote Sens. Environ. 47, 231–241. https://doi.org/10.1016/0034-4257(94)90158-9

Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. 31st Conf. Neural Inf. Process. Syst. (NIPS 2017) 1–10.

Mackintosh, E.E., van Der Hulst, J., 1978. Soil drainage classes and soil water table relations in medium and coarse textured soils in Southern Ontario. Can. J. Soil Sci. 58, 287–288.

Malone, B.P., McBratney, A.B., Minasny, B., 2018. Description and spatial inference of soil drainage using matrix soil colours in the Lower Hunter Valley, New South Wales, Australia. PeerJ 6, e4659. https://doi.org/10.7717/peerj.4659

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/10.1016/j.geoderma.2020.114885

Mendes, W.S., 2020. Geotechnologies applied in digital soil mapping. Universidade de São Paulo, Piracicaba.

Mendonça Santos, M.L., ten Caten, A., 2015. Mapeamento Digital de Solos (MDS): Avanços e desafios. Bol. Inf. da SBCS 39–42.

Meyer, H., Pebesma, E., 2021. Predicting into unknown space? Estimating the area of applicability of spatial prediction models. Methods Ecol. Evol. 12, 1620–1633. https://doi.org/10.1111/2041-210X.13650

Miller, B.A., Koszinski, S., Wehrhan, M., Sommer, M., 2015. Impact of multi-scale predictor selection for modeling soil properties. Geoderma 239–240, 97–106. https://doi.org/10.1016/J.GEODERMA.2014.09.018

Niang, M.A., Nolin, M., Bernier, M., Perron, I., 2012. Digital mapping of soil drainage classes using multitemporal RADARSAT-1 and ASTER images and oil survey data. Appl. Environ. Soil Sci. 2012, 17. https://doi.org/10.1155/2012/430347

Peng, W., Wheeler, D.B., Bell, J.C., Krusemark, M.G., 2003. Delineating patterns of soil drainage class on bare soils using remote sensing analyses. Geoderma 115, 261–279. https://doi.org/10.1016/S0016-7061(03)00066-1

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

Poppiel, R.R., Lacerda, M.P.C., Rizzo, R., Safanelli, J.L., Bonfatti, B.R., Silvero, N.E.Q., Demattê, J.A.M., 2020. Soil color and mineralogy mapping using proximal and remote sensing in Midwest Brazil. Remote Sens. 12, 1197. https://doi.org/10.3390/rs12071191

Quinlan, J.R., 1986. Induction of decision trees. Mach. Learn. 1, 81–106.

Raymond, J.E., Fernandez, I.J., Ohno, T., Simon, K., 2012. Soil drainage class influences on soil carbon in a New England Forested Watershed. Soil Sci. Soc. Am. J. 77, 307–317. https://doi.org/10.2136/sssaj2012.0129

Schaetzl, R., 2013. Catenas and soil, in: Shroder, J. (Ed.), Treatise on Geomorphology. Academic Press, pp. 145–158.

Schoonover, J.E., Crim, J.F., 2015. An Introduction to Soil Concepts and the Role of Soils in Watershed Management. J. Contemp. Water Res. Educ. 154, 21–47. https://doi.org/10.1111/J.1936-704X.2015.03186.X

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/10.1016/j.geoderma.2021.115089

Soil Science Division Staff. 2017. Soil survey manual. C. Ditzler, K. Scheffe, and H.C. Monger (eds.). USDA Handbook 18. Government Printing Office, Washington, D.C.

Troeh, F.R., 1964. Landform parameters correlated to soil drainage. Soil Sci. Soc. Am. J. 28, 808–812. https://doi.org/10.2136/sssaj1964.03615995002800060035x

Zhao, Z., Chow, T.L., Yang, Q., Rees, H.W., Benoy, B., Xing, Z., Meng, F.R., 2008. Model prediction of soil drainage classes based on digital elevation model parameters and soil attributes from coarse resolution soil maps. Can. J. Soil Sci. 88, 787–799.

5. SENSING TECHNOLOGIES FOR CHARACTERIZING AND MONITORING SOIL FUNCTIONS - A REVIEW

Abstract

Soils perform multiple functions that contribute to human survival and well-being. However, we often focus on soil to produce food and fiber and unsustainable management has accelerated its degradation, jeopardizing the successful delivery of its multiple functions. To explore the full potential of the soil in delivering multiple functions, a deep understanding of its behavior and variability in space and time is needed. Direct measurement of these soil functions is challenging but can be quantified by soil indicators. Sensing technologies (i.e., laboratory spectroscopy, and proximal and remote sensing) can provide reliable information on soil's spatial variability at different spatial and temporal scales, which can be of paramount importance to studying soil functionality. As there is a lack of a comprehensive review describing the possibilities of sensing technologies for providing reliable soil data to assist soil functional assessments, this review presents evidence on the applicability of sensing technologies. Among the sensing technologies, laboratory spectroscopy has been found as a global tool that can characterize a suite of soil physical, chemical amd biological soil properties accurately. Proximal sensing with a wide range of available sensors is practical for field soil characterization. Remote sensing, in turn, has been demonstrated to be effective in the monitoring of soils over space and time. Thus, sensing technologies can contribute to assessing soil functions at local, regional, and global scales. Their integration with data from the field or from process-based simulation is needed when aiming for a deep understanding of processes that contribute to delivering key services for human welfare.

Keywords: Soil functions, Soil quality, Soil health, Ecosystem services, Proximal sensing, Remote sensing, Soil properties, Soil mapping, Soil sensing applications, Spectroscopy

5.1. Introduction

Globally, we are witnessing a growing concern with the sustainable use of natural resources. Soil is one of the earth's greatest non-renewable natural resources, existential to humanity to provide food (Godfray et al., 2010; Koppitke et al., 2021). We have to rely on soil to feed a population that will reach more than 9 billion by 2050. To fulfill that task, soil is under constant anthropogenic pressure (Montanarella et al., 2016), and faces the threat of climate change. In addition to producing food, soil also has important life-supporting functions, including storing carbon, storing, filtering and transformation of nutrients, substances and water and habitat for biodiversity (Blum, 2005; Greiner et al., 2017). As soil supports vital transformations of paramount importance to the Earth's system, its fundamental role in achieving United Nations Sustainable Development Goals (SDGs) cannot be underestimated. In recent years, however, the study of soil functions has gained prominence (Koppitke et al., 2021). We are witnessing a growing interest in understanding how the spatial variability of soil properties at different spatial and temporal scales affects the soil's ability to perform the aforementioned functions.

This paper uses the definition of soil functions according to the European Commission in the EU Soil Thematic Strategy (Montanarella and Panagos, 2021). Soil functions relate to the roles that soils play in sustaining the wellbeing of humans and society (Greiner et al., 2017). This review focuses on five main functions:

1. Biomass production, including agriculture and forestry.

2. Storing, filtering, and transforming nutrients, substances, and water.

3. Biodiversity pool, such as habitats, species, and genes.

4. Acting as carbon pool

5. Source of raw materials.

Direct measurement of these soil functions is somewhat difficult and most of the time quantified through measurable physical, chemical, and biological properties, which are used as indicators (Rabot et al., 2018; Vogel et al., 2019); as the capacity of a soil to perform these functions depends on them (Tóth et al., 2013). Soil properties are commonly classified as "inherent" or "dynamic" (Dominati et al., 2010; Karlen et al., 2001); the first being those directly associated with the soil-forming factors described by Jenny (1941), while the second representing those more easily affected by anthropogenic management (Karlen et al., 2001). Although huge improvements have been made to methods and protocols for assessing soil properties, their measurement is still considerably expensive, even more so if large scale monitoring is needed (Pásztor et al., 2017; Wiesmeier et al., 2019). This limitation has led to a scarcity of reliable soil data (Adhikari and Hartemink, 2016); especially in developing countries, where just exploratory or semi-detailed soil maps are available for decision-making purposes. Brazil, for example, has only 1.6% of its territory mapped with soil maps at 1:100,000 scale (Mendonça Santos and ten Caten, 2015). Other countries in Latin America and Africa are facing the same problems as they cannot afford the high costs of obtaining detailed soil maps (Dewitte et al., 2012). While developing countries are more likely to have rich soil databases, some problems could still arise from incomplete information. Zwetsloot et al. (2020) faced this problem when aiming to measure soil multifunctionality across Europe and advised not to include soil information in European monitoring systems until the maps or information associated with soil attributes were improved.

The scarcity of soil data has broadened the use of alternative cost-effective and environmentally friendly technologies such as spectroscopy, and proximal (PS) and remote sensing (RS), as a more efficient way to obtain exhaustive soil data or to improve incomplete regional and global datasets (Mulder et al., 2011). Soil spectroscopy deals with identifying and analyzing the interaction of soil components with electromagnetic energy, specifically in the visible and infrared portions (Nocita et al., 2015). In this definition, some proximal and remote sensors are also included. Proximal sensing refers to acquiring soil information through field sensors, with direct or close contact with the soil. Remote sensing, in turn, is defined as the recording of emitted, reflected, transmitted, or scattered energy – in the form of images – of portions of the earth's surface from a distance, through unmanned aerial vehicles (UAVs), aircraft, or satellites (Weiss et al., 2020).
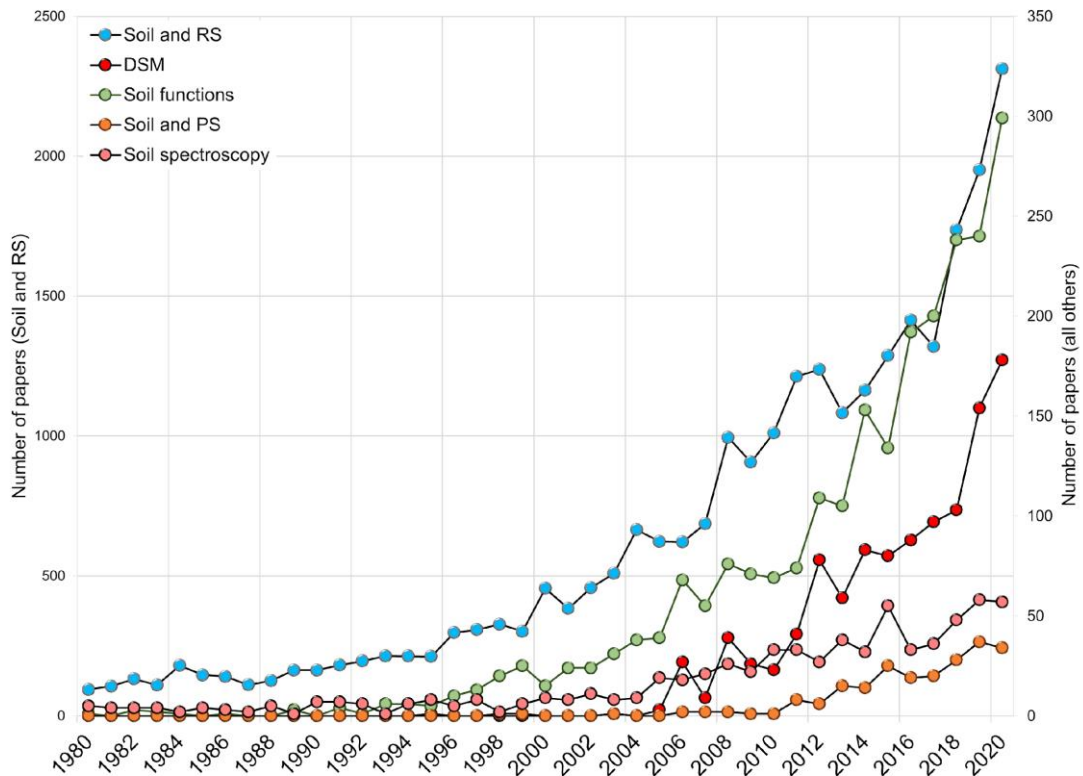
Figure 1. Research trends of the topics addressed in this review.

The data acquired by soil sensors can be used directly or indirectly to study soil attributes. The first provides information about soil composition (e.g., soil mineralogy, elemental composition, or soil color) by studying the interactions of electromagnetic energy with soils. The second method consists of empirically estimating soil properties using real data obtained in the laboratory to calibrate prediction models. Generally speaking, the second approach aims at obtaining point-based or spatially explicit information of soils by relating soil spectral data or environmental covariates to the soil property or soil class of interest using empirical methods. This is the so-called Digital Soil Mapping (DSM) framework (McBratney et al., 2003), which has been widely disseminated worldwide thanks to improvements in computational capacity and modelling approaches (Chen et al., 2022). The spectral data or environmental covariates used in DSM are proxies that represent one or more of the scorpan factors (s: soil, c: climate, o: organisms, r: relief, p: parent material, a: age, n: location) (Campos et al., 2019).

Soil sensing technologies have revolutionised the way soil properties are being studied. This is confirmed by the exponential growth of scientific literature reported since the 2000s (Fig. 1). As the study of soil functions will increasingly require more spatially explicit information to aid decision making in relation to sustainable soil management, the role that sensing technologies would play in this task cannot be underestimated. In this paper, we aim to shed light on the use of sensing technologies to assess and monitor the variability of soil properties and their use in evaluating soil functions. Therefore, this review begins with a brief description of soil spectroscopy, followed by proximal and remote sensing and their application in obtaining information concerning soil properties. The second part describes soil functions and their indicators. Finally, this review ends with a discussion related to the

strengths, weaknesses and possible future scenarios of sensing technologies associated with obtaining soil information that contributes to a better understanding of soil functions.

5.2. Extracting information from sensing technologies

Sensing technologies such as soil infrared spectroscopy and proximal and remote sensors are based on the interaction of electromagnetic energy (EM) with soil components. Almost all regions of the EM - from gamma rays to microwaves - can provide information about the soil in the form of reflectance or absorbance values (UV-vis-IR); dielectric (microwaves) or electromagnetic (EC) characteristics; and counts per second (cps) or percentage (%) of specific elements (XRF, Gamma-ray, LIBS). However, not all soil properties can be estimated/studied by sensing technologies because there must be a spectral response associated with them. Table 1 provides a list of the soil properties that are possible to obtain from sensing technologies. In this list, there are well-established soil properties (e.g., soil texture and OC) that have been demonstrated to be accurately described from sensing technology. Others, however, may present low accuracy (e.g., pH, CEC, and macronutrients). While there are many soil sensors that operate along the electromagnetic spectrum, this paper focuses on sensors that can measure several soil properties at a time.

Table 1. Schematic representation of soil properties that can be estimated/studied by sensing technologies

Technique, sensor or instrument	Gamma spectrometer	XRF	LIBS - Infrared spectrometers - Optical and Thermal sensors in satellites	Active and passive radar satellites	EC
EM portion	γ-ray	X-ray	UV-vis-IR	Microwaves	Radiowaves
EM range (m)	10^{-12}	10^{-10}	10^{-8} to 10^{-4}	10^{-2}	10^{1} to 10^{6}
Soil properties					
Texture (Clay, sand, and silt)	X	X	X	X	X
Soil color			X		
Organic matter			X		
SOC		X	X	X	
CEC			X		
pH			X		
Fe2O3			X		
Minerals			X		
Soil moisture			X	X	X
Soil temperature			X		
K, U, Th	X				
Electrical conductivity					X
Elemental composition		X	X		
Soil depth			X		X
Macronutrients*		X	X		X
Trace elements		X	X		

XRF: X-ray fluorescence. LIBS: Laser induced breakdown spectroscopy, EC: Electrical conductivity; SOC: Soil organic carbon; CEC: Cation exchange capacity; Fe2O3: iron oxides; K, U and Th: radioisotopes of potassium, uranium and thorium.

In this section, the main characteristics of soil spectroscopy and proximal and remote sensors will be presented. Further, a description of how soil properties are studied or estimated by these sensors will be provided as

well. For complete and detailed reviews on sensing technologies, the reader is referred to the works of Demattê et al. (2016), Stenberg et al. (2010) and Viscarra Rossel et al. (2011).

5.2.1. Gamma-ray spectrometry

Gamma-ray spectroscopy, also known as geophysics, is concerned with recording data of high energy and short-wavelength electromagnetic radiation, generally at frequencies of 1 Hz or wavelengths <10-12 m (Table 1). The electromagnetic radiation recorded by gamma-ray sensors usually is those emitted from naturally occurring isotopes, which originates from the radioactive decay of unstable elements in the soil (Reinhardt and Herrmann, 2019). Although many elements in the soil have radioisotopes, only potassium (40K), thorium (232Th), and uranium (238U) and their associated daughters produce sufficient energy susceptible to be detected and studied by gamma-ray sensors (Kuang et al., 2012). These elements are in an unstable state and searching for balance; they undergo a series of radioactive decays until they reach stability. The step towards a more stable form causes the emission of gamma-ray radiation, characteristic of each element, like a fingerprint. The gamma-ray recorded is directly proportional to the concentration of radioisotopes, indicative of changes in the upper soil, usually up to 30 cm (Wilford and Minty, 2007; Mello et al., 2021). The 40K, and the 232Th and 40U decay series are used to estimate potassium, thorium, and uranium abundances, although estimates of other soil properties are also possible.
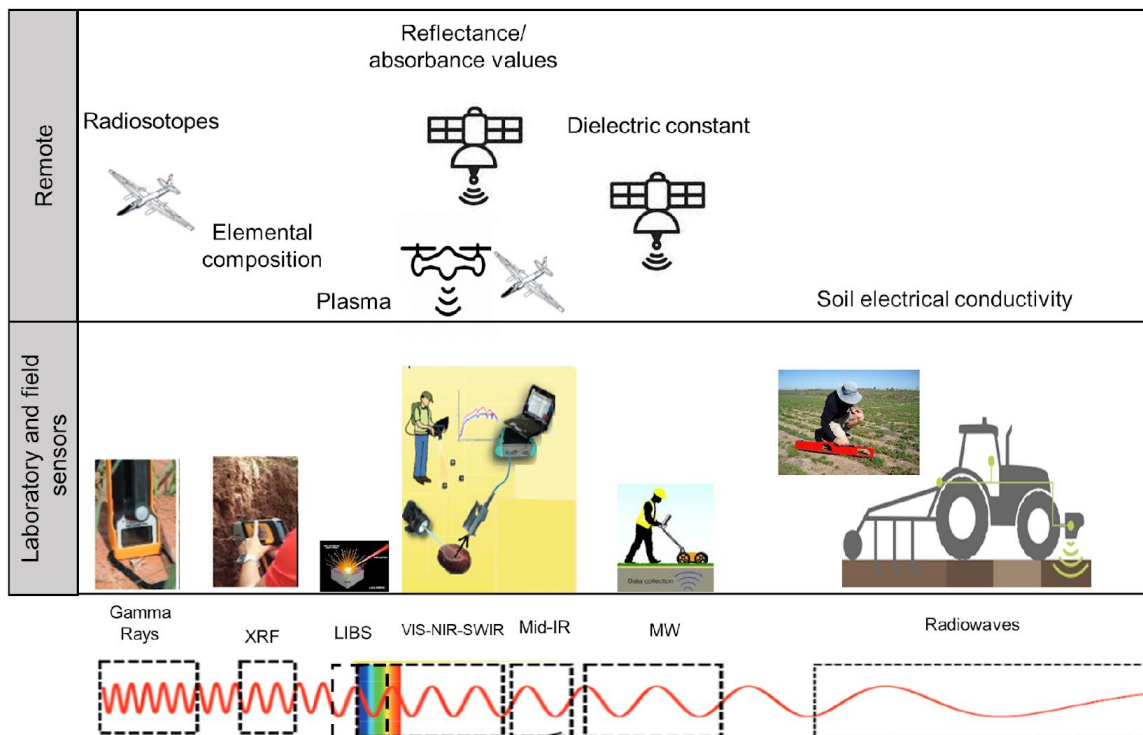
Figure 2. Schematic representation of the electromagnetic spectrum and the sensor at each spectral region

Gamma-ray sensors can be proximal or remote. Proximal sensors are used to collect data directly in the field for on-the-go measurement or for location-specific acquisition with hand-held sensors. Remote gamma-ray

sensors are usually mounted on aircraft platforms and have the advantage of sensing larger areas; however, their accuracy is lower than ground-based tools (Reinhardt and Herrmann, 2019).

The distribution of radioisotopes (K, Th, and U) is believed to be associated with soil mineralogy and geochemistry and strongly depends, at least at first, on bedrock composition (Reinhardt and Herrmann, 2019; Kuang et al., 2012). The applications in soil science include parent rock characterization; soil texture mapping; plant-available potassium; environmental behaviour of potassium, thorium and uranium; soil pH; soil organic carbon and peat mapping; soil type mapping and soil erosion.

Radioisotopes obtained through gamma rays are especially useful to understand predominant pedogenetic processes and help unveil aspects related to weathering or soil erosion. An example of this approach is the work of Mello et al. (2021), who used Th, U, and K concentration to study pedogenetic processes that occured or predominate in the tropical environment. The authors concluded that gamma-ray spectrometry could detect significant variations in the study area, unnoticeable by conventional techniques, which may indicate changes from one soil class to another or the continuity of a given soil class. Gamma rays are also used to empirically infer other soil properties via regression models. This approach is carried out using the values of K, Th, and U or the total count or the entire spectrum. The most common soil properties estimated from gamma-ray are clay (Pracilio et al., 2006; Viscarra Rossel et al., 2007; Mahmood et al., 2013; Coulouma et al., 2016), silt and sand contents (Viscarra Rossel et al., 2007; Mahmood et al., 2013), soil texture (Priori et al., 2014; Heggemann et al., 2017; Petersen et al., 2012; Read et al., 2018; Rouze et al., 2017), pH (Viscarra Rossel et al., 2007; Mahmood et al., 2013), available potassium (Viscarra Rossel et al., 2007; Dierke and Werban, 2013; Kassim et al., 2021), and organic carbon (Dierke and Werban, 2013). Gamma-rays can also be used in conjunction with other environmental covariates to predict soil thickness (Chen et al., 2021), available water capacity (Zare et al., 2021) and even historical erosion rates (Varley et al., 2020).

5.2.2. X-Ray fluorescence

The X-ray fluorescence technique (XRF) uses high-energy radiation in the X-ray range to reach the elementary composition of an analyte. It comprises the spectral range after gamma rays, with wavelengths between 10-12 and 10-10 m. The main purpose and use of XRF is to quantify chemical elements (from sodium to uranium) present in soil samples by using the characteristic fluorescence radiation, which is emitted after a pulse of X-ray radiation hits the soil. The XRF works as follows: i) the incident X-ray energy (primary radiation) is generated when the electrons produced by a cathode collides with an anode (metal target); ii) the energy reaches the atoms of the soil sample and excites the electrons in the innermost layer, which are ejected (ionization), promoting instability in the atom; iii) the electrons are temporarily replaced by electrons from outer shells, emitting energy as photons (secondary radiation or fluorescence), which is measured by the detector. The emitted radiation is characteristic of each chemical element, and the intensity is proportional to the concentration of the element in the soil sample, making possible its detection and quantification (Weindorf et al., 2014; Silva et al. 2021).

X-ray fluorescence is used to obtain the concentration of almost all elements in the soil. The elemental soil characterization by XRF can be divided into two parts: i) determination of major elements present in the soil in oxide forms such as sodium, magnesium, aluminum, silicon, phosphorus, potassium, calcium, titanium, manganese and iron; and ii) determination of trace elements such as: vanadium, chromium, cobalt, nickel, copper, zinc, arsenic, rubidium, strontium, yttrium, zirconium, molybdenum, mercury, lead and thorium. The detection limit and quantification accuracy are closely related to the element's atomic number, being higher the uncertainty in elements

lighter than sulfur and rare the quantification of elements lighter than sodium (Potts, 1992). This happens because the capacity to emit fluorescence energy is smaller in light elements.

X-ray fluorescence is traditionally performed in the laboratory using a benchtop spectrometer; however portable XRF instruments are now widely available and directly used in the field. The result provided by XRF sensors is a spectral signature, measuring the counts per second of the elements in the soil sample. XRF sensors are often calibrated to provide the concentration (%) of specific elements or in the form of oxides (for instance Fe2O3).

The XRF data can be used to predict the pseudo total contents of elements related to mineralogy. Silva et al. (2020a) calibrated XRF to estimate SiO2, Fe2O3, Al2O3, TiO2 and P2O5 contents as measured using sulfuric acid digestion. These values are usually used in pedological studies to calculate weathering indexes such as (SiO2/Al2O3) and [SiO2/(Al2O3 + Fe2O3)]. Sulfuric acid digestion analysis has been used extensively in tropical regions, but its use is declining because it is an expensive procedure that produces harmful chemical residues. Data obtained from XRF was also used to predict exchangeable elements that are important for plant nutrition, such as magnesium, phosphorus, potassium, calcium, manganese and copper, with varying accuracies. Soil texture (Zhang and Hartemink, 2019; Andrade et al., 2020a; Benedet et al. 2020), pH (Sharma et al. 2014; Zhang and Hartemink, 2019), SOM (Zhang and Hartemink, 2019; Tavares et al. 2020; Andrade et al., 2020b), cation exchange capacity and total nitrogen (Zhang and Hartemink, 2019; Andrade et al., 2020b). The characterization of soil profiles by portable XRF sensors is also possible and appears to be a very effective in-field tool to assess pedogenic pathways and parent material origins (Stockmann et al., 2016).

5.2.3. Laser-induced breakdown (LIBS) spectroscopy

Laser-induced breakdown spectroscopy (LIBS) is an atomic emission technology that allows a real-time multi-elemental characterization (qualitative and quantitative analysis) of all the elements present in soil samples with detection limits ranging from 0.3 to 100 µg g-1 (Vilas Boas et al., 2020a). The LIBS technique requires minimum samples preparation and is based on a high-energy laser pulse focused on a small amount of sample causing heating, vaporization, and, ultimately, a plasma composed of atoms, ions, electrons, and radiation in equilibrium (Harmon and Senesi, 2021). The plasma is produced after the interaction of the laser pulse with the soil sample.

When the cooling process starts, the plasma emits photons captured by spectrometers in the ultraviolet, visible and near infrared spectral ranges (200 nm to 900 nm), from which an emission spectrum can be obtained. The spectra are qualitatively or quantitatively analyzed to obtain the elemental composition of soil samples, which are considered a fingerprint for each element, similar to X-ray fluorescence and gamma-rays.

The main interest for using LIBS is to directly measure elemental concentration in soils as the concentration in the soil is proportional to the concentration in the plasma (Vilas Boas et al., 2020b). The advantages of LIBS over traditional methods are the none or little sample preparation, no chemical wastes, short analytical time and the ability to determine multiple elements in one signal measurement. Wang et al. (2018) demonstrated this by determining the contents of multiple heavy metal elements in soils with satisfactory accuracy. Cremers et al. (2001) measured the total soil carbon of a Mollisol using LIBS. Their results suggested that the LIBS methodology rapidly and efficiently measures soil carbon with excellent detection limits (~300 mg kg), precision (4-5%) and accuracy (3-14%). Tavares et al. (2022) proved that LIBS efficiently predicted fertility attributes in tropical soils. They achieved an RPD > 1.40 for eight out of the nine properties studied, being pH the one that showed the worst result. The best results were obtained for CEC, ex-Ca, ex-Mg, clay, OM and ex-K. The ex-P proved to be the most challenging. Although LIBS shows P emission lines, the relationship between extractable and pseudo total content of P is

complex, varying considerably in different agricultural areas. Therefore, further studies involving the prediction of P by combining LIBS with other direct analysis techniques are encouraged.

5.2.4. Soil infrared spectroscopy

5.2.4.1. Laboratory and field vis-NIR-SWIR and MIR spectroscopy

The most commonly used spectroscopy in soil science is in two main groups: vis-NIR-SWIR (vis: 400-780 nm; NIR: 780-1100 nm; SWIR: 1100-2500 nm), and middle infrared (MIR: 2500-25000 nm or 4000 – 400 cm-1). In the first, electronic transitions, weak overtones, and combination tones are responsible for the presence of absorption features related to soil components such as organic matter, soil water, minerals and soil texture. In the MIR, fundamental vibrations of molecules (such as C-O, C-N, C-H) are common. Both spectral regions are probably the most well-known portions of the electromagnetic spectrum used to simultaneously characterize several soil properties, thanks to the presence of soil chromophores and soil components that interact with light (Escribano et al., 2017). The vis-NIR-SWIR and MIR spectra are obtained under laboratory conditions after basic sample preparation (drying, grinding, and sieving) and without pre-processing or chemical reactants. The information obtained is the reflectance factor (the ratio between the measured reflectance and a standard reflector), ranging between 0 and 1 with continuous spectral bands, usually every 1 to 3 nm. The result is a soil spectral signature (with >2000 spectral bands) in which the reflectance factor is plotted against wavelengths or wavenumbers in the x-axis and where main absorption features related to soil components can be distinguished.

Several soil properties can be estimated from a single spectrum. Before using soil spectra, they must undergo pre-processing techniques, which consist of mathematical procedures that seek to improve the visualization of spectral characteristics for further study (Angelopoulou et al., 2020). These procedures are used to enhance the signal, reduce noise and extract hidden information that is not visible in the raw spectra (Rinnan et al., 2009). After preprocessing, the corrected spectra can be used: i) to analyze absorption features related to soil compounds and ii) to empirically relate the soil spectra to the soil property (or properties) of interest. The first is a physically-based estimation, in which minerals' relative abundances are calculated based on the measurement of the intensity of absorption features where these minerals are more likely to occur (Viscarra Rossel et al., 2010a). Commonly estimated minerals are iron oxides (hematite and goethite) and some aluminosilicates, such as kaolinite and gibbsite but montmorillonite, muscovite, illite, chlorite, and calcite can also be estimated. It is worth noting that this approach is more likely to estimate the relative abundances rather than the real amount of these minerals in the soil, which will only be possible to obtain if spectra of pure minerals or samples with known quantities of the mineral of interest are used. Viscarra Rossel et al. (2010a) derived a normalized iron oxide difference (NIODI) for Australia, in which absorption features of Haematite (Hm) and Goethite (Gt) at 880 and 920 nm. Silva et al. (2020) obtained spatial distributions of Hm and Gt in Central West Brazil by calculating the actual amount of these minerals relating NIR spectra to abundances obtained via X-Ray diffraction data. The authors used the approach of Scheinost et al. (1998), using the spectral amplitude between 415 and 445 nm and 535 and 580 nm for Gt and Hm, respectively. Fernandes et al. (2020) mapped the spatial variability of Kaolinite and Gibbsite in Central-West Brazil based on reflectances on specific regions in the NIR spectra. Other studies have demonstrated that various minerals can be mapped efficiently based on the spectra information, including Kaolinite, Gibbsite, montmorillonite, muscovite, illite, chlorite, and calcite (Poppiel et al. 2020; Mendes et al. 2021).

Soil colour is another property that can be calculated directly from soil reflectance values, since it is physically related to the visible portion of the electromagnetic spectrum. The Munsell chart is the standard book

used to describe soil colour (Schwertmann, 1993). However, it was considered subjective, as different soil mappers can describe a very different colour for the same soil. This subjectivity can be reduced in part by calculating the soil color directly from reflectance spectra, which can be converted to the color spaces of interest (RGB, CIELAB, Munsell, XYZ, Xxy, etc.).

Soil colour can be considered as a pedoenvironmental indicator as other properties or processes can be inferred from it (Stockmann et al., 2018). Additionally, it is a criterion that allows estimating relative amounts of some compounds, such as organic matter that provide dark color to the soil (Baumann et al., 2016; Leger et al., 1979; Moritsuka et al., 2014; Spielvogel et al., 2004; Vodyanitskii and Savichev, 2017). Reddish to yellowish colours are also associated with the presence of Haematite and Goethite in soil. Field measurement of soil colour provides some clues about the predominant mineral, presence of organic components, and drainage/redox conditions (Barrett et al., 2002; Blavet et al., 2000; Malone et al., 2018; Zanetti et al., 2015).

Knowledge-based methods are useful but often soil spectra are difficult to interpret and the variation in mineral and organic matter content is not a linear function of a peak at certain wavelengths. Thus empirically-based methods have been extensively used to describe the relationship between reflectance spectra and soil properties. The soil spectra can be used as independent variables in regression analysis, to estimate the soil property of interest. This is performed using the full spectrum, only specific wavelengths, or even principal component analysis, which is used to reduce the multicollinearity between spectral channels. In most of the works, the models are first calibrated using a fraction of the total samples (usually 70%) and then validated with the remaining samples (30%). Despite the availability of many statistical techniques to describe the relationship between soil spectra and soil properties, the Partial Least Square Regression (PLSR) and Principal Component Regression (PCR) were used as a standard in the majority of the earlier works. More recently, sophisticated machine and deep learning algorithms have gained increasing attention as they have been shown to better describe the non-linear relationships between soil properties and spectra (Ding et al., 2018; Morellos et al., 2016; Viscarra Rossel and Behrens, 2010b; Yang et al., 2019; Padarian et al., 2019; Ng et al., 2019).

All of these approaches have proven to be useful in estimating soil properties. Clay, sand, and silt contents, organic matter (SOM), soil organic carbon (SOC), iron content, calcium carbonates, pH, CEC, total carbon and total nitrogen have been successfully estimated from reflectance spectra. Soil properties such as clay, SOM, SOC, and iron are more accurately described by soil spectra because they are spectrally active and produce absorption features that help describe their behavior. Soil texture consists of different particle size distributions, but especially the clay content and consequently, the clay minerals are the ones that help to describe it. Regarding SOC and SOM, their presence in the soil affects the albedo, by decreasing the reflectance spectra due to the absorption of light, which also gives a dark color to the soil. Soil pH and CEC are not spectrally active and their accuracies were very variable. In some cases, these soil properties were successfully described by soil spectra because of the influence of other spectrally active soil properties such as clay and SOM content. On the other hand, soil nutrients are more difficult to estimate from soil spectra due to their dynamic behavior in the soil and the absence of spectral signatures in the vis-NIR-SWIR and MIR region. Therefore, their estimates by soil spectra are, in most cases, less accurate.

Vis-NIR-SWIR and MIR spectroscopy is also used directly in the field. Most sensors cover just the visible and near infrared region, with several or just a few bands. The most important issue with field reflectance spectroscopy is the significant influence that water, organic matter and iron forms can have on the spectra (Demattê et al. 2003, Silvero et al., 2020), being necessary some preprocessing techniques before the spectra is ready for use and analysis. This is one of the reasons why field spectroscopy presents lower accuracy than laboratory spectroscopy,

yet it has the advantage of providing real-time data. Nevertheless, algorithms such as External Parameter Orthogonalisation had been devised to remove unwanted field and moisture effects from the Vis-NIR-SWIR signals (Minasny et al., 2011). It is important to highlight that for the case of MIR, the emitted radiation from the earth would be part of the signal captured by the sensor, which provides a weak signal and low accuracy to characterize soils. This might be one of the reasons why only a few portable MIR systems were reported.

5.2.5. Remote sensing

5.2.5.1. Optical remote sensing

Satellites orbit the Earth at approximately 600 to 900 km away and acquire information from almost the entire electromagnetic spectrum – from ultraviolet to microwaves. Optical remote sensors measure the sunlight energy reflected in the visible, near, and short-wave infrared regions, i.e., 400 – 2500 nm or 0.4 - 2.5 µm (Escribano et al., 2017). Satellites provide reflectance values in the form of images in a few spectral ranges. These spectral ranges vary from 10 to 100 nm and are called spectral wavebands. Depending on the number of wavebands, remote sensors can be classified into multispectral (<20 bands) and hyperspectral (> 100 bands).

Multispectral sensors are more popular than hyperspectral, which are generally used on an aircraft. The two most common hyperspectral satellites are the Hyperion-EO1 from NASA – decommissioned in 2017 – and PRISMA from the Italian Space Agency, launched in 2019, and DESIS on board Multi-User-System for Earth Sensing (MUSES) platform on the International Space Station. Other orbital hyperspectral sensors are under development (e.g., HyspIRI - NASA, EnMap – Germany, and SHALOM – Israel/Italy), scheduled to be launched in the next few years. Landsat missions are the most common multispectral satellites and also the most used for soil studies. However, in the last few years, a wide range of satellites that provide useful soil information at various scales are becoming available, either commercially or for free. PlanetScope, Ikonos and Quickbird are examples of commercial satellites that provide high-spatial-resolution images.

Despite its usefulness for large-scale studies, the major caveat of optical remote sensing is its susceptibility to atmospheric conditions, especially the presence of clouds and shadows in the images. The clouds appear opaque in optical imagery and can completely invalidate reflectance signals recorded by satellites (Meraner et al., 2020). The conditions of surface humidity, roughness, topographic position, and vegetation or cover crops also influence the spectral response recorded in optical satellite images (Ben-Dor et al., 2002). The exposure of the soil will depend on the growing season and the management strategies adopted. A few multitemporal techniques have been proposed to overcome this issue and enhance remote sensing products for soil monitoring by regularly providing spatially explicit information over large spans and over time. With the recent advances in remote sensing technologies, most of the newest sensors have improved capacities in terms of revisit time (i.e, <10 days for Sentinel-2 missions), which significantly increases the availability of images for soil surface monitoring. As the pixels in the images are not all bare soil, but rather a mixture of various land cover and uses, spectral indices are calculated (e.g., NDVI, SAVI, BSI, etc.) and are used to exclude the pixels that are non-bare soil, using threshold values to separate bare soil pixels from other. This results in images with just bare soil pixels that can later be aggregated to form a synthetic soil image. An example is given in Figure 3. For more details about multi-temporal techniques, the reader is referred to the works of Diek et al. (2017), Rogge et al. (2018), Demattê et al. (2018), Roberts et al. (2019) and Demattê et al. (2020).

Remote sensing products are used in several ways in soil studies, specifically as environmental covariates in digital soil mapping (DSM). DSM has become a standard practice for obtaining spatially explicit information of soils worldwide. Optical remote sensing data in conjunction with terrain attributes derived from digital elevation models

are used to represent one or more factors related to soil formation, especially vegetation, parent material or the soil itself (Boettinger et al., 2008). The most common way to do this is by enhancing the original image using band ratios that allow images to be visually scanned for subsequent analysis (Jensen, 2013). These band ratios are commonly known as spectral indices and are used to represent the vegetation (RVI: Ratio Vegetation Index; NDVI: Normalized Difference Vegetation Index; DVI: Difference Vegetation Index; GVI: Green Vegetation Index; EVI: Enhanced Vegetation Index, among many others), bare soil and its components (SAVI: Soil-adjusted Vegetation Index; BSI: Bare Soil Index; RI: Redness Index, etc.), and water (i.e., NDWI: Normalized Difference Water Index). However, spectral indices are not the only way remote sensing products are used to describe the variability of soil properties. All spectral bands of an image, either multispectral or hyperspectral can be directly used to empirically describe the spatial distribution of a soil property or soil class.
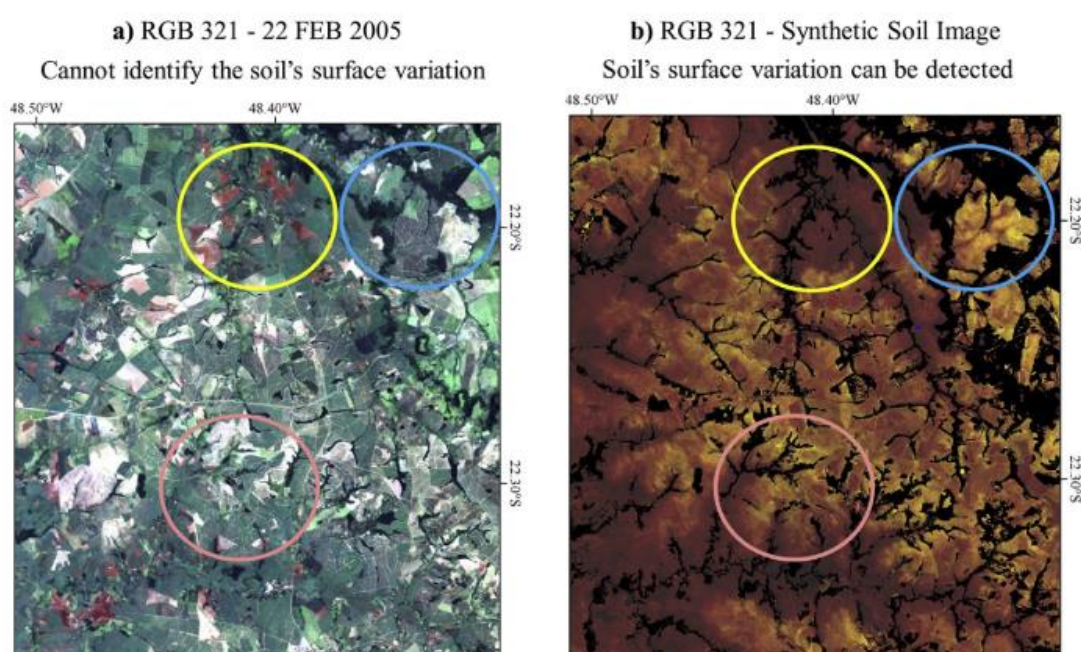
Figure 3. a) RGB 321 Landsat image from a single date, b) synthetic soil image obtained from a Landsat time series of ~30 years, also in RGB 321, after the GEOS3 method was applied (Demattê et al., 2018). The circles are indicating selected cases where the synthetic soil image helps to identify the soil spectral behavior that is not completely possible with the single date image-depicted in (a). Example adapted from Demattê et al. (2018).

Ben-Dor et al. (2002) used airborne hyperspectral data (DAIS-7915 instrument) to map soil moisture, soil salinity, soil saturated moisture and organic matter content in Israel. The authors used a spectral empirical model to predict soil properties directly from their reflectance spectra. Similarly, Demattê et al. (2007) used Landsat TM bands and multiple linear regression analysis to estimate several properties with reasonable accuracy in Brazil. In the last few years, soil scientists have been adopting more sophisticated machine learning algorithms which work better than simple linear models (Khaledian and Miller, 2020; Padarian et al., 2019). Soil scientists are also interested in presenting the uncertainties of soil properties estimates, which would give a better idea of the accuracies.

Clay content, sand content, texture, organic matter, calcium carbonate (CaCO3) and organic carbon (SOC) have shown good accuracy when predicted by remote sensing products in conjunction with other environmental covariates. Particularly, clay content and SOC mapping appear in most of the works because their relationship with the electromagnetic spectrum is well consolidated in the literature and their influence in many other soil properties

and processes (Loiseau et al., 2019). In addition, SOC mapping is also raising increasing attention by soil scientists who are concerned with the size of soil carbon pool and its sequestration potential as of paramount importance in climate change mitigation (Minasny et al., 2013).

5.2.5.2. Thermal remote sensing

Thermal remote sensing is based on the measurement of emitted electromagnetic radiation in the spectral intervals from 3 to 5 μm and 8 to 14 μm, where a few atmospheric windows allow the signal to pass. The spectral range from 8 to 14 um is most commonly used in thermal remote sensors because the radiation emitted by the Earth's surface is high. In thermal remote sensing, the source of radiation is the heat of the surface, which is related to its kinetic or true temperature. Since all objects with a temperature above absolute zero (-272.16 °C) emit electromagnetic radiation, Planck's radiation law, based on the theory of black bodies can be used to measure the spectral radiant exitance (M) and generate what is called "blackbody graphs". A blackbody is a theoretical (non-existent) object that fully absorbs all incident radiation, with zero reflectance and zero transmittance, and re-emits all energy. The total radiant exitance – also called radiant, apparent or brightness temperature – is related to the true kinetic temperature of an object, which is generally measured by a thermometer. However, the former is usually smaller than the later. This happens because none of the terrestrial objects are true blackbodies, but objects that selectively emit a portion of the total radiation emitted by a blackbody, a property known as "emissivity". Emissivity varies from 0 to 1 and can be obtained from reflected radiation following Kirchoff's law. Earth objects with the same kinetic temperature can exhibit very different radiant temperatures due to their different emissivity. This is the reason why emissivity should be considered when calculating the land surface temperature. Further details about thermal remote sensing can be found in Jensen (2013) and Kuenzer and Dech (2013).

There are several methods for obtaining land surface temperature (LST) from thermal remote sensing images, but a straightforward method is the inversion of a radiative transfer equation. LST retrieved from thermal bands is used for monitoring thermal changes that can take place in the soil due to land use changes. It was also used to explain the variability of other soil properties, specifically in low relief areas where other covariates such as terrain and vegetation do not spatially co-vary with soil conditions (Liu et al., 2020). Sayão and Demattê (2018a) used LST calculated from a single-data thermal image and another approach with three LST images from several dates to map soil texture and SOC. The authors found that significant correlations between soil texture/SOC and LST exist, but LST did not produce reliable estimates of these soil properties, recommending integrating with reflectance data to improve the accuracy. Similarly, Sayão et al. (2018b) used LST to map several soil properties in a small tropical area in Brazil. The LST as a covariate was unable to provide good estimates of soil properties variability, but it was useful to indicate soils with clayey texture, high OM and iron oxides contents. Wang et al. (2015) mapped soil texture in China based on the hypothesis that the rate of change of LST is related to soil texture, given similar starting soil moisture conditions. In this case, the authors did find the LST useful for mapping soil texture variability. Liu et al. (2020) proposed the solar radiation-based approach, which is supported in soil thermal changes during and after a solar radiation event. They reported good results for soil texture, SOM and pH but did not recommend it for use to estimate other soil properties that may not have a strong relationship with thermal characteristics.

Thermal remote sensing can also be used to estimate soil moisture content (SMC) by measuring the amplitude of the diurnal temperature change and developing a relationship between the temperature change and SMC (Petropoulos et al., 2015). However, radar remote sensing is the spectral region most commonly used to estimate SMC, as described in the next section.

5.2.5.3.   Microwave remote sensing

Microwave remote sensing covers the region of the electromagnetic spectrum ranging from 0.5 to 100 cm (or 30 - 0.3 GHz), which is characterized by long wavelengths. In this spectral region, atmospheric conditions have little or no influence on the signal, allowing sensing at any time of the day, at night and with the presence of clouds. Depending on the mode of operation, microwave remote sensing is classified into passive or active. Active microwave remote sensors, also called radar (RAdio Detection And Ranging), do not rely on natural light. Instead, radar sensors use an antenna to generate their own pulses of microwave radiation sent to the Earth's surface. Passive microwave sensors, on the other hand, measure naturally emitted microwave radiation but are not used so much due to the weak signal. The wavelengths used in microwave remote sensing are much longer than those from optical and thermal remote sensing, which allows the signal to penetrate further in depth. The longer the wavelength in the specified spectral region, the greater the penetration of the radar signal. There are eight commonly used band designations for radar wavelengths (cm): Ka (0.75 – 1.18), K (1.199 – 1.67), Ku (1.67 – 1.24), X (2.4 – 3.8), C (3.9 – 7.5), S (7.5 – 15), L (15 – 30), and P (30 – 100), being X to P the most frequently used.

The radar antenna measures all, some, or none of the backscatter energy after actively sending pulses of radiation of specific wavelength and duration towards the Earth's surface from the side of the platform. The radar antenna can send and receive polarized energy, which is energy that vibrates only in one direction, either horizontal or vertical. A signal can be sent in one direction and return to the antenna in the same direction (VV or HH polarization) or can be sent in one direction being the returning signal in another direction (VH or HV polarization). Polarized energy is useful to better "see" the details of an object that may not be possible with unpolarized energy (Jensen, 2013).

The spatial resolution of a radar image is directly related to the antenna size; the longer the antenna, the better the spatial resolution. However, a huge antenna is required if a high spatial resolution is needed, which is not possible in space. This was resolved using Synthetic Aperture Radars (SAR), an approach in which the movement of the platform is used to produce a "synthetic aperture" considerably larger than the actual antenna, which allows very fine resolution images to be obtained over long distances. The returning radar signal recorded by the antenna is usually expressed as the backscattering coefficient (σ^o, measured in dB), whose values are dependent upon the radar system and terrain parameters. Radar system parameters are the incidence angle of the radar beam, wavelength, and polarization. Among the terrain parameters, dielectric constant, surface roughness, bulk density, soil texture, and land cover are the most common factors that influence the radar signal (Das and Paul, 2015).

Radar remote sensing is commonly used in soil science to estimate soil moisture content (SMC), by taking advantage of the large difference between the dielectric properties of dry soil (~4) and liquid water (~ 80) (Barrett and Petropoulos, 2013), which is expressed in the backscattering coefficient. Although all portions of the electromagnetic spectrum can be used to some degree to estimate SMC (Petropoulos et al., 2015) radar remote sensing is the most suitable for this task. There are three types of approaches that can be used to estimate SMC from radar remote sensing: theoretical or physical, empirical, and semi-empirical. The physical approach uses backscattering models capable of reproducing the radar-backscattering coefficient from the sensor configuration (wavelength, polarization, and incidence angle) and soil parameters (soil moisture and surface roughness for bare soils). The use of theoretical models for soil moisture retrieval over large areas is challenging because of the difficulty of describing the natural surface roughness and topography over such large areas (Barrett and Petropoulos, 2013). Empirical models use SMC collected in the field and relate it to the backscattering values of the radar signal. Semi-

empirical models, in turn, take advantage of the last two approaches. It starts from a physical background and then uses simulated or experimental data sets to simplify the theoretical backscattering model.

5.2.6. Electromagnetic induction

Electromagnetic induction (EMI) sensors measure the apparent electrical conductivity (ECa) of the soil subsurface (Doolittle and Brevik, 2014), which is a measurement of a soil's ability to conduct electricity. EMI is a non-contact and active sensor that has a transmitter and a receiver installed on both ends of a non-conductive bar and uses low frequency (kHz) magnetic fields to induce currents in the soil. The currents are sent to the soil to induce an alternating primary magnetic field, which generates small currents, while the soil matrix induces a secondary magnetic field, weaker than the primary. This secondary magnetic field is directly linked to the ECa of the soil and is measured by the receiver. The ECa, is a product of the spatial variability of soil physical, chemical and mineralogical properties and is an invaluable tool for providing spatial information for precision agriculture applications, especially the delineation of site-specific management units. EMI can be attached to a vehicle and thus provide detailed spatial maps of soil properties from the surface to 1 or 3 m deep.

5.3. Soil properties as indicators of soil functions

So far, we have described the main characteristics of sensing technologies and how to obtain data on soil properties. In this section, we will describe how soil properties estimated from sensing technologies can be used to assess, estimate or describe soil functions. It is important to highlight that some soil functions are difficult or expensive to measure, even more so if constant monitoring of changes over time is required. This is one of the reasons why soil properties are commonly used as indicators of soil functions (Vogel et al., 2019). Such situations allow sensing technologies to contribute and provide reliable soil data in space and time for soil functional assessment.

Soil functions have been defined in several ways. For instance, as goods and services that depend on dynamic processes occurring in the soil (Robinson et al. 2014); as flows and transformations of mass, energy, and biodiversity within Earth's critical zone (Banwart et al., 2019); as benefits that human populations get from soils (Baveye et al., 2020) or as an applied approach to estimate soil quality primarily for land use and administration purposes (Lehmann and Star, 2010). Here, we used the definition according to the European Commision (EC, 2006): soil can perform seven functions for human. The most well-known is the production of biomass but the soil also supports human infrastructure, is a source of raw materials, a habitat for a wide range of organisms, a physical and cultural environment for human and an engine for filtering and transforming substances and water (Table 2). Its most recently recognized role is as a carbon pool, to help fight climate change and reduce CO_2 emissions to the atmosphere.

Table 2. Relations between soil functions and soil properties.

Soil functions	Soil texture	pH	CEC	Soil color	mineralogy	CaCO3	OM	SOC/TOC	TON	Soil temperature	Soil depth	Electrical conductivity	Macro elements	Trace elements	AWC/PAW	Bulk density	Hydraulic conductivity	Soil structure	Coarse fragments	Air capacity	Species diversity	Earthworm abundance
Biomass production	X	X	X		X		X	X	X	X	X	X			X	X	X	X	X	X		
Carbon pool	X				X			X		X	X											
Storing, filtering & transforming nutrients	X	X	X	X	X	X	X	X			X	X			X	X	X		X	X		
Storing, filtering & transforming water	X				X	X	X	X			X				X	X	X		X	X		
Biodiversity pool	X	X						X	X	X		X			X	X					X	X
Source of raw materials	X			X								X										

Adapted from Dominati et al. (2010), Adhikari et al. (2016), Calzolari et al. (2016), Greiner et al. (2017); Vogel et al. (2019), and Jost et al. (2021).

All these functions are performed by the soil but not all soils can deliver these functions at the same time. Vogel et al. (2019) distinguished inherent and manageable soil properties. The inherent properties were inherited from parent material during soil formation processes and are unlikely to change in human time span (<100 years). The presence of gravel, the size of the particles (soil texture), and mineralogy are examples of inherent soil properties. While manageable soil properties such as soil pH, organic matter, soil carbon and nutrients, are characterized by their short-term variability due to the management adopted. Soil functions can be measured under two perspectives: potential or actual state. The first is considered as the potential that the soil can offer and can be studied by assuming soil properties under its natural condition or under minimum human influence. While the actual state, in turn, is what is being observed. An example of the soil's potential and actual state to support biomass production is provided in Dominati et al. (2014), who found that 64% (10.4 t DM/ha/year) of the pasture biomass produced with added N and P fertilizers was due solely to the natural potential of the soil. The actual state would be the total biomass produced by the adopted fertilizer strategies. Another way of looking at the potential and actual state is by using the concept of genosoils and phenosoils (Roman Dobarco et al., 2021). The concept considers that under the same soil forming factors, genosoils are soils of the same soil under remnant natural vegetation while phenosoils are soils under human management.

We have seen in previous sections that both inherent and dynamic soil properties can be obtained from sensing technologies. However, not all soil properties can be obtained from sensing, but most of the time, they can be inferred through pedotransfer functions (PTFs). Secondary soil properties that are more difficult to measure or not possible to measure with sensing technologies can be predicted via PTFs. Other soil properties can only be obtained through direct measurements in the field and neither sensing technologies nor PTFs are helpful for this. Table 2 shows soil properties that are more likely to be used as proxies of specific soil functions.

Here examples will be given on how soil properties estimated from sensing technologies can be used to describe four functions: biomass production, filtering and buffering, carbon pool and habitat for biodiversity. Later, two experiences in mapping soil functions in Brazil will be given.

5.3.1. Soil as a biomass producer

Soils support biomass production by providing physical structure, nutrients and water to plants (Dominati et al., 2014). The role of the soil as one of the fundamental factors for agricultural and forestry production is well known. Evaluating the soils' potential to produce biomass through frameworks such as soil quality, has been on the agenda of soil science for centuries (Vogel et al., 2019). As depicted in Table 2, almost all inherent and dynamic soil properties affect biomass production. The soil texture, which refers to the distribution of particle sizes in the soil (Hillel, 1998), is one of the most important because of its direct effect on water and nutrient dynamics (He et al., 2014). It is also used to estimate other soil properties by using pedotransfer functions (PTFs), such as available water storage capacity.

Soil texture is the soil property that has been demonstrated to be very well predicted by sensing technologies. Its importance in the assessment of soil's capability to sustain biomass production is supported by differences in yield productivity in soils with different textures. For example, the aboveground dry biomass yield of miscanthus grass averaged over three harvest dates was one order of magnitude lower in silty clay soils than in sandy loam (24.6 vs 3.9 Mg ha-1) (Roncucci et al., 2015). In a similar study, He et al. (2014) demonstrated that soil texture influences wheat productivity because of its significant influence on water retention, ion exchange and nutrient cycling. They found that the wheat grown in clay soil had higher drought tolerance than that in the silt loam.

Soil texture and mineralogy can be used to represent the potential state of soil to produce biomass (e.g., Vogel et al., 2019). Although Vogel et al. (2019) stressed the importance of water and energy, they only considered soil texture in their scoring system. To improve this kind of scoring function, data on soil moisture and soil temperature, obtained from thermal and radar remote sensing, could also provide useful insights regarding the potential and actual state of the soil to provide biomass such as Champagne et al. (2012). The authors tested three different strategies to obtain spatially explicit information on soil moisture, and found that soil moisture map estimated using radar remote sensing was the most effective method to describe changes in canola and wheat yield in the Canadian prairies.

Another approach is the study of dynamic soil properties or elements related to the soil's nutritional status, which is indirectly linked to its capability to produce biomass. We have seen that the elemental composition of the soil is more easily obtained from proximal sensors such as XRF and LIBS. Tavares et al. (2021) used a multi-sensor approach to study soil properties that are related to its fertility. They estimated soil nutrients (Ca, Mg, K, P) and other soil properties (pH, SOC, texture) that are useful to describe the actual state of soil fertility to support biomass production. Similarly, data obtained from EM sensors, which are easily obtained directly in the field, provide real-time data to describe the short-term variability of soils. With this information, production environments or management zones are delineated to provide insights on the potential of soil to produce biomass and where the efforts need to be concentrated to improve or increase this potential.

The use of soil properties obtained via sensing technologies in crop process-based modeling is another strategy to estimate the potential of the soil to provide biomass. Crop modelling is a particularly useful tool for understanding the likely productivity of crops in different environments, especially considering the climate as the main driving factor. Soil types and its variability can have even more influence on yield productivity than climate

(Folberth et al., 2016). That is one of the reasons why efforts have been concentrated in developing soil databases to aid crop modeling, especially at large scales (Han et al., 2019). The advantage of process-based models is that different scenarios can be tested to describe the outermost changes that might result from soil spatial variability. Besides that, with the inclusion of the adopted management strategies, such as fertilization or irrigation, the full potential of the soil due to its natural capital and those that were added can be described. For this purpose, soil properties estimated from sensing technologies, especially those from a spatial perspective, obtained from DSM approaches and remote sensing, have the advantage of providing spatially explicit information on crop yield productivity and, indirectly, how the soil is performing its function to provide biomass. These approaches are not necessarily applicable just to agricultural crops but also for forest and pastures (Tian et al., 2017; Insua et al., 2019).

More direct approaches to obtaining biomass production estimates from remote sensing products using spectral indices and radiative transfer models are also possible, but are beyond the scope of this paper. A meta-review of remote sensing applications in agriculture is provided in Weiss et al. (2020).

5.3.2. Soil as a carbon pool

Soils' key role in sequestering CO_2 is receiving increasing attention, especially to mitigate climate change under the 4per1000 initiative (Minasny et al., 2017). The soil carbon sequestration, in the words of Olson (2010) is the "process of transferring CO_2 from the atmosphere into the soil through plants, plant residues and other organic solids, which are stored or retained as part of the soil organic matter". The interest in studying the potential of soil to sequester carbon responds to increasing awareness of greenhouse gas emissions (GHG) and climate change. As the amount of carbon in the soil system at 1 m depth (~1500 Pg) is more than twice that found in the biosphere (560 Pg) and almost twice that found in the atmosphere (867 Pg), the soil can easily become a source or a sink of carbon (Lal, 2018). The ability of soil to act as a carbon pool is important for climate change regulation and supporting other functions, such as biomass production, nutrient and water cycling.

Soil carbon stock mapping using sensing approaches have been widely studied and documented (e.g., Chen et al., 2021). However, we also need to know about soil carbon stock change and sequestration potential. Soil carbon sequestration can be increased by following the indications of good management practices; for example, by increasing plant residues in the soil or promoting the reduction of the turnover rates of SOC stocks. However, the soil has a limited capacity in sequestering carbon after its saturation capacity is reached (Wiesmeier et al., 2014), being extremely necessary to know this potential to guide management practices better. In this sense, the ability of soils to store carbon will rely on a set of physical, chemical, mineralogical, biological, climatic and topographic properties that can positively or negatively affect the soils' potential to store carbon. Lal (2018) mentioned that soil C sink capacity depends on soil depth, clay content and mineralogy, and the antecedent SOC stock. Wiesmeir et al. (2019) also review indicators that are known to affect SOC storage at different spatial scales and found soil texture as one of the most promising factors to be used as indicator for SOC storage over a wider range of scales. They give examples of research that used the fine fraction (clay+silt) to estimate the C saturation capacity of soils, reinforcing that not just the quantity of the fine fraction is important but also its quality and mineralogy.

The importance of clay content and mineralogy relies on the great influence that they will have on the longevity and persistence of the carbon in the soil system, also known as mean residence time. Different types of minerals are present in the soil such as iron oxides, phyllosilicates, carbonates, gypsum and amorphous phases that will - to a great or lesser extent - protect the carbon in the soil (Barré et al., 2013). In general, soil minerals are able

to form organomineral compounds, preventing microbial attack, decomposition and consequent release of the carbon in the form of CO_2 to the atmosphere (Barré et al., 2013; Schmidt et al., 2011, Baldock and Skjemstad, 2000).

Soil texture, SOC content, mineralogy and elemental composition estimated from one of the sensing technologies presented in previous sections could have a great potential for studies involving the functionality of soils to serve as a carbon pool. In fact, there are initiatives worldwide concerned with finding methods to study the sequestration potential of soils either by spectral methods in laboratory or from remote sensors, aiming large-scale monitoring of changes in soil C dynamics (Angelopoulou et al., 2020).

5.3.3. Soil as reservoir and a filtering and buffering engine

The soil naturally acts as a living - and limited - filtering and buffering entity. In the filtering process, soil prevents pathogens, toxins, solutes and contaminants from reaching groundwater, by chemically transforming pollutants into harmless components or retaining them in the soil matrix (Hillel, 1998; Jónsson et al., 2019). In the buffering process, soil is capable of preventing extreme pH changes and reducing acidifying impacts (Makó et al., 2017). These filtering and buffering processes are vital to purifying water, avoiding pollution and contributing to groundwater quality and water cycling (Dominatti et al., 2010; Comerford et al., 2013; Keesstra et al., 2012). Besides its filtering and buffering capacity, the electrical charge of the soil and the surface area of its particles contributes to the longer or shorter residence time of water molecules that would be extremely useful for plant uptake (Amundson et al., 2015), especially during drought periods and between rainfall events, where the water scarcity may compromise vegetative growth. These regulating roles are extremely important for the soil system's well-functioning and avoiding soil and groundwater pollution (Blum, 2005).

The soil is not an infinite filter or buffering machine, nor an infinite water reservoir, however. Its capability to filter and buffering is strongly dependent on several factors such as climate, topography, plant cover, land use and biological, physical and chemical characteristics, specifically soil texture, pH, organic matter, cation exchange and base saturation (Makó et al., 2017). The soil type, lithology and mineralogy also have been found to have a profound effect (da Costa et al., 2013). Some soil types are inherently more prone to provide these regulatory services effectively. Fine-textured soils with high organic matter, for example, will have a higher cation exchange capacity (CEC), and therefore, will be more able to adsorb positive charged pollutants and pathogens and remove them from the soil solution. In coarse-textured soils, however, the water is less likely to be retained in the soil matrix when compared to fine-textured soil with high clay content because the latter can hold water molecules for longer periods in the spaces between the aggregates (Comerford et al., 2013; da Costa et al., 2013).

The filtering, buffering and water storing capacity of soils are not easy-to-measure functions. The majority of the works aimed at characterizing this function used either information contained in soil survey, soil properties or indicators. Calzolari et al. (2016), for example, used soil properties such as CEC and pH, and additional information such as coarse fragments and the depth to the water table to build an indicator of the filtering and buffering potential of soils (BUF). They found that the BUF is significantly higher in soils rich in organic matter. This was similar to that observed by Aslam et al. (2009), who stated that a soil with a high SOC content would improve the soil's generic capacity to filter, in their case, pesticides. Here it is important to note that Aslam and collaborators did not actually measure this function but rather assumed that an increase in SOC would be beneficial only as long as no hydrophobicity occurs because SOC is a key sorbent for many pesticides. This is interesting because hydrophobicity is often developed in sandy soils due to their small specific surface area (Knadel et al., 2015), again texture can be an indicator of the filtering capacity of soils.

The last example is a study conducted by the European Commission that mapped the storing and filtering capacity of European soils based on available soil map information (Makó et al., 2017). For this, the authors used the European Soil Database, land use and land cover maps, climate data and pedotransfer functions. The authors found that soil's storing and filtering capacity may increase as the clay content and OM increases, but this will depend on the type of pollutants, being necessary to consider the effects of soil pH and mineralogy.

5.3.4. Soil as habitat for biodiversity

Soil can provide habitat to a wide range of organisms, being one of the most diverse habitats on Earth (Aksoy et al., 2017; Pulleman et al., 2012). Soil biodiversity refers to all organisms that live and interact in the soil, playing a key role in the delivery of other soil functions, including organic matter decomposition, nutrient cycling, soil structure formation, pest regulation and bioremediation of contaminants (van Leeuwen et al., 2019; Pulleman et al., 2012).

The actual potential of soils to host biodiversity is still poorly studied. Although the soil has the capability to host a huge diversity of microorganisms (bacterias, fungi, protists) and the micro- (rotifers, tardigrades, nematodes), meso- (collembolans, mites, enchytraeids) and macrofauna (ants, termites, earthworm); its suitability is governed by physicochemical properties (such as pH, soil texture, soil organic matter, organic carbon) and environmental factors (such as precipitation, temperature, evapotranspiration, among others) that play a key role in providing habitat and sources for these organisms (Aksoy et al., 2017). However, this suitability is not well-addressed in the research devoted to soil biodiversity, which most of the time deals only with the measurement of the DNA of microorganisms and fails to describe and interpret the soil environment in which they live (Baveye et al., 2016).

Soil physical and chemical characteristics can be beneficial to define the suitability hotspots for biodiversity. Soil organic matter (SOM) and pH are the two most important soil properties determining the prevailing conditions of the habitat, because they have a great influence on the availability of nutrients for microorganisms (Turbé et al., 2010; van Leeuwen et al., 2019; Xue et al., 2018). SOM decline can considerably compromise soil biodiversity (Orgiazzi, 2016; Xue et al., 2021), being important to follow-up its temporal variability to have an idea of where soil biodiversity is declining or enriching.

There are some examples where the potential of soil to support biodiversity of soil organisms was the main objective, and they are mostly concentrated in Europe. Such studies were not intended to assess soil biodiversity directly, but instead used environmental information to determine the most suitable potential habitats to host soil biodiversity. Aksoy et al. (2017), for instance, aimed at assessing the soil biodiversity potential of Europe and for this, they selected soil and environmental variables such as pH, soil texture, soil organic matter, potential evapotranspiration, average temperature, soil biomass productivity and land use, giving emphasis on the importance that some soil-forming factors can have as direct determinants of soil biodiversity. The authors used score indicators to define the variables that had more influence on soil biodiversity. They defined three potential levels (poor, average and good potential) and found that half of the European soils (47%) has an average soil biodiversity potential, 37% have poor and about 16% have good potential. In another very similar work, Stone et al. (2016), with the aim of developing and implementing a monitoring campaign to ascertain the range of soil biodiversity and associated ecosystem functions, used pan-European available databases of soil chemical and physical properties and a site selection approach to establish a transect with the variation of various soil properties and land use across Europe. Soil pH, texture, and land uses were used to define specific site environmental conditions that are considered to be suitable for most of the organisms. They defined 81 sites, where micro and meso fauna biodiversity, extracellular

enzyme activity, PLFA and community level physiological profiling (MicroRespTM and BiologTM) will be measured.

Other examples did measure specific organism groups and related their distribution to environmental factors. This is the case of Griffiths et al. (2016), who fitted environmental variables to bacterial communities and found a strong correlation across Europe. Generally speaking, bacterial community differences were highly correlated with change in soil chemistry and nutrient status, with soil pH showing the strongest relationship, confirming that across large spatial scales the structuring of soil bacterial communities is largely predictable by common soil physico-chemical parameters. Similar finding was given in Xue et al. (2018) for bacteria and fungi distribution in Australia. Rutgers et al. (2016), in turn, used digital soil mapping approaches to relate low density earthworm community data to soil characteristics, land use, vegetation and climate factors (covariables), assuming that macrofaunal groups respond strongly to their habitats and therefore, their geographical distribution can potentially be predicted from environmental data.

5.3.5. Soil as a source of raw materials

Soils also provide materials for infrastructure such as gravel, sand, and clay (Blum, 2005). In addition, soil can be a source of trace elements or rare earth elements (REE) that are used in electronics and technology equipment. Soil was commonly seen as a sink for these elements through industrial or waste disposal activities, which became soil pollutants (Ramos et al., 2016). However, soils could be the source of these REE through resource recovery (Lima and Ottosen, 2021) and sensors such as XRF, LIBS and MIR can measure the concentration of these elements in soil (Ng et al., 2022). Coupled with DSM, sensing can assess the potential of the soil to provide raw materials for infrastructure and manufacturing.

5.3.6. A few examples of the assessment of soil functions in Brazil

In the last section, we provided a general overview of how soil properties are used to study the potential or actual state of the soil to perform its functions. Herein we want to present examples of soil functional assessment of biomass production and soil carbon sequestration potential in Brazil, specifically in the Piracicaba region, by using data obtained from sensing technologies. This area has very high complexity and diversity in terms of parent materials and soil, where several research projects have been carried out to understand soil variability. Complete details about this variability can be found in Gallo et al. (2018), Mendes et al. (2021) and Bonfatti et al. (2020).

The work of Santos (2021) evaluated the usefulness of soil properties maps for defining crop production potential. They used soil maps from several sources at different spatial scales to evaluate the influence of soil data on modeling attainable sugarcane yield, which is defined as the potential yield limited by water supply (rainfall) (Cheeroo-Nayamuth et al., 2000). The DSSAT/Canegro process-based model was adapted to provide a spatial map of the attainable sugarcane yield. The authors used soil properties maps produced by digital soil mapping and remote sensing images. These maps were also used to derive hydraulic soil properties via PTFs. The authors found that the soil properties maps obtained from digital mapping approaches and remote sensing images outperformed all the other soil maps used. The worst result was obtained with a legacy soil map at 1:100000 scale, the best available soil map of the study area. The results highlighted the importance of using accurate maps of soil properties for better representing the potential of soils to produce biomass, in this example, sugarcane.

The following examples aimed at mapping SOC stock at regional and country levels. The work of Tayebi et al. (2021) sought to understand the main drivers of SOC stock variability. For this, the authors used a set of

environmental variables including maps of soil properties and terrain attributes. They also evaluated how the changes over time and for different land uses, considering agriculture, pasture and forests. The research presented in Safanelli et al. (2021) followed a similar approach but focused on cropland areas for entire Brazil. The authors used satellite images, PTFs and digital soil mapping approaches to map the SOC stock variability in the croplands. Both works serve as examples of how the soil function to sequester carbon can be assessed. The SOC stock represents, at some point, either a potential for the soil to be a sink or a source of CO_2. Therefore, a SOC stock map obtained using sensing technologies can considerably help us understand where soils would potentially sequester more carbon or where they need to be maintained following best agronomic practices.

5.4.   Advantages, limitations and way forward

So far, we have seen all the possibilities that sensing technologies offer us to obtain information about the soil and how this data can be used to study the functions it performs. However, it is important to note that the soil data accessed or estimated from sensing technologies is insufficient to fully understand how the soil works. There is a need to integrate this data with other more specific information collected in the field, to improve our understanding. The advantage of using sensing technologies to obtain more and more soil data is valuable but some other issues still need to be addressed before taking full advantage of them.

For example, one of the main disadvantages is the need to have well-trained people, who deeply understand the fundamentals of each technology and can use them properly in the study of the soil. It is not a secret anymore that sensing technologies are positioned themselves as suitable tools for qualitative and quantitative characterization of the soil. However, despite the great interest of the soil science community in these technologies, there are still no well-established protocols, especially for spectral measurement and processing using soil spectroscopy and field-based sensors (Ben-Dor et al., 2015).

Luckily, this is changing and some initiatives concerned with the development of protocols, specifically for soil spectroscopy, are rising. One of them is the "Standard Protocol and Scheme for Measuring Soil Spectroscopy", which is a worldwide working group established in February 2020 under the IEEE SA umbrella. This international initiative aims to form an agreed-upon ISO standard and protocol to measure soil spectral data (https://sagroups.ieee.org/4005/). The Technical Group on Soil Spectroscopy of the FAO's Global Soil Laboratory (GLOSOLAN) is another example, which was launched in 2020 and aims to "bring together institutions and experts around the world to join efforts to use this technology to support decision-making on soil protection globally". Its main objectives are to contribute to protocols, guidelines, materials, and to coordinate useful calibration procedures for soil spectroscopy (http://www.fao.org/global-soil-partnership/glosolan/soil-analysis/dry-chemistry-spectroscopy/en/).

Aside from protocols and standards, soil scientists are also concerned with developing soil spectral assets, commonly known as soil spectral libraries (SSL). SSL is defined as the collection of reflectance spectra (vis-NIR-SWIR or MIR spectra, separately or both) along with traditional soil chemical and physical analyses to develop models to predict soil properties. So far, two SSLs were presented (vis-NIR-SWIR spectra): The global SSL introduced by Viscarra Rossel et al. (2016) with 23,631 soil spectra from 35 institutions and 90 countries, and the Brazilian SSL, presented by Demattê et al. (2019) with 39,284 soil spectra from 25 states and 41 institutions across the Brazilian territory. MIR SSLs are still scarce and just a few have been reported or used at the local or regional levels (Briedis et al., 2020; Gomez et al., 2020; Terra et al., 2015; Ng et al., 2019).

These SSLs are evolving to web platforms where several soil properties can be quantified using laboratory vis-NIR-SWIR and MIR spectra. This is the case of the Brazilian SSL (Demattê et al., 2022) and the Spectral platform for soil samples of the Democratic Republic of Congo (https://sae-interactive-data.ethz.ch/simplerspec.drc/). Both platforms have great potential to bridge the gap between scientists and public users, who can find in these platforms a better way to access data that sometimes is restricted to research institutions.

The proliferation of discrete countries or regions or institutions SSL is an example of the traditional central model where a single organization compiles the library with contributions from its partners. Data ownership and sharing are still a big issue. With the advancement in data science, spectra data can now follow a distributed model where each party retains their data ownership and can participate in building a global SSL via blockchain. Such a decentralized system ensures that users respect the data privacy of others and yet can take advantage of having global data to build better soil inference models (Padarian et al., 2020). We envision such a system for the future SSL.

Finally, to better quantify soil functions, data obtained from sensing technologies and digital soil mapping approaches can be put in process-based models (Lagacherie et al., 2022). These mechanistic models are used to simulate and represent a specific process under varying scenarios; for instance, for modeling crop yields, water availability and turnover of soil organic matter. Here, soil data obtained from sensing technologies can contribute to getting better results.

5.5. Conclusions

This review explores some of the possibilities sensing technologies offer us to obtain accurate information about the soil variability, in space and time to quantify soil functions. The following points can be highlighted from this review:

- Sensing technologies provide a range of useful data about the soil: texture, mineralogy, carbon stock, elemental composition.
- A combination of sensors provides more accurate prediction for a range of soil properties.
- Data obtained by sensing technologies can be spatialized via digital soil mapping approaches to understand the spatial distribution and variability of soil properties.
- Soil data can be used in two ways: as indicators of soil functions and as inputs in process models to quantify soil functions.

Soil sensing provides valuable inputs to quantifying and modelling soil functions as it allows for the estimation of a whole suite of inherent and dynamic soil properties. Laboratory spectroscopy can characterize a range of soil properties accurately. Proximal sensing with a wide range of available sensors is practical for field soil characterization. Remote sensing, in turn, is effective in the monitoring of soils over space and time. Thus, different sensing technologies can assess soil functions at local, regional, and global scales.

Acknowledgments

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

Aksoy, E., Louwagie, G., Gardi, C., Gregor, M., Schröder, C., Löhnertz, M., 2017. Assessing soil biodiversity potentials in Europe. Sci. Total Environ. 589, 236–249. https://doi.org/10.1016/j.scitotenv.2017.02.173

Amundson, R., Berhe, A.A., Hopmans, J.W., Olson, C., Sztein, A.E., Sparks, D.L., 2015. Soil and human security in the 21st century. Science (80-.). 348. https://doi.org/10.1126/SCIENCE.1261071

Andrade, R., Silva, S.H.G., Faria, W.M., Poggere, G.C., Barbosa, J.Z., Guilherme, L.R.G., Curi, N., 2020a. Proximal sensing applied to soil texture prediction and mapping in Brazil. Geoderma Reg. 23, e00321. https://doi.org/10.1016/j.geodrs.2020.e00321.

Andrade, R., Silva, S.H.G., Weindorf, D.C., Chakraborty, S., Faria, W.M., Guilherme, L.R.G., Curi, N., 2020b. Tropical soil order and suborder prediction combining optical and X-ray approaches. Geoderma Reg. 23, e00331. https://doi.org/10.1016/J.GEODRS.2020.E00331

Angelopoulou, T., Balafoutis, A., Zalidis, G., Bochtis, D., 2020. From laboratory to proximal sensing spectroscopy for soil organic carbon estimation-A review. Sustainability 12, 443. https://doi.org/10.3390/su12020443

Aslam, T., Deurer, M., Müller, K., Clothier, B.E., Rahman, A., Northcott, G., Ghani, A., 2009. Does an increase in soil organic carbon improve the filtering capacity of aggregated soils for organic pesticides? — A case study. Geoderma 152, 187–193. https://doi.org/10.1016/J.GEODERMA.2009.06.015

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

Banwart, S.A., Nikolaidis, N.P., Zhu, Y.G., Peacock, C.L., Sparks, D.L., 2019. Soil functions: Connecting earth's critical zone. Annu. Rev. Earth Planet. Sci. 47, 333–359. https://doi.org/10.1146/annurev-earth-063016-020544

Barrett, L.R., 2002. Spectrophotometric color measurement in situ in well drained sandy soils. Geoderma 108, 49–77. https://doi.org/10.1016/S0016-7061(02)00121-0

Barrett, B. W. & Petropoulos, G. P. (2013) Satellite remote sensing of surface soil moisture, in: Remote Sensing of Land Surface Turbulent Fluxes and Soil Surface Moisture Content: State of the Art. Boca Raton, FL: CRC Press.

Barré, P., Fernandez-Ugalde, O., Virto, I., Velde, B., Chenu, C., 2014. Impact of phyllosilicate mineralogy on organic carbon stabilization in soils: incomplete knowledge and exciting prospects. Geoderma 235–236, 382–395. https://doi.org/10.1016/J.GEODERMA.2014.07.029

Baumann, K., Schöning, I., Schrumpf, M., Ellerbrock, R.H., Leinweber, P., 2016. Rapid assessment of soil organic matter: Soil color analysis and Fourier transform infrared spectroscopy. Geoderma 278, 49–57. https://doi.org/10.1016/j.geoderma.2016.05.012

Baveye, P.C., Berthelin, J., Munch, J.C., 2016. Too much or not enough: Reflection on two contrasting perspectives on soil biodiversity. Soil Biol. Biochem. 103, 320–326. https://doi.org/10.1016/J.SOILBIO.2016.09.008

Baveye, P.C., Schnee, L.S., Boivin, P., Laba, M., Radulovich, R., 2020. Soil organic matter research and climate change: merely re-storing carbon versus restoring soil functions. Front. Environ. Sci. 8, 161. https://doi.org/10.3389/fenvs.2020.579904

Ben-Dor, E., Patkin, K., Banin, A., Karnieli, A., 2002. Mapping of several soil properties using DAIS-7915 hyperspectral scanner data-a case study over clayey soils in Israel. Int. J. Remote Sens. 23, 1043–1062. https://doi.org/10.1080/01431160010006962

Ben Dor, E., Ong, C., Lau, I.C., 2015. Reflectance measurements of soils in the laboratory: Standards and protocols. Geoderma 245–246, 112–124. https://doi.org/10.1016/j.geoderma.2015.01.002

Benedet, L., Faria, W.M., Silva, S.H.G., Mancini, M., Demattê, J.A.M., Guilherme, L.R.G., Curi, N., 2020. Soil texture prediction using portable X-ray fluorescence spectrometry and visible near-infrared diffuse reflectance spectroscopy. Geoderma 376, 114553. https://doi.org/10.1016/J.GEODERMA.2020.114553

Blavet, D., Mathe, E., Leprun, J.C., 2000. Relations between soil colour and waterlogging duration in a representative hillside of the West African granito-gneissic bedrock. Catena 39, 187–210.

Blum, W.E.H., 2005. Functions of soil for society and the environment. Rev. Environ. Sci. Biotechnol. 4, 75–79. https://doi.org/10.1007/s11157-005-2236-x

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 Netherlands, pp. 193–202. https://doi.org/10.1007/978-1-4020-8592-5_16

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 107305. https://doi.org/10.1016/j.geomorph.2020.107305

Briedis, C., Baldock, J., de Moraes Sá, J.C., dos Santos, J.B., Milori, D.M.B.P., 2020. Strategies to improve the prediction of bulk soil and fraction organic carbon in Brazilian samples by using an Australian national mid-infrared spectral library. Geoderma 373, 114401. https://doi.org/10.1016/j.geoderma.2020.114401

Calzolari, C., Ungaro, F., Filippi, N., Guermandi, M., Malucelli, F., Marchi, N., Staffilani, F., Tarocco, P., 2016. A methodological framework to assess the multiple contributions of soils to ecosystem services delivery at regional scale. Geoderma 261, 190–203. https://doi.org/10.1016/j.geoderma.2015.07.013

Campos, A.R., Giasson, E., Costa, J.J.F., Machado, I.R., Silva, E.B. da, Bonfatti, B.R., 2019. Selection of environmental covariates for classifier training applied in digital soil mapping. Rev. Bras. Ciência do Solo 42. https://doi.org/10.1590/18069657rbcs20170414

Cheeroo-Nayamuth, F.C., Robertson, M.J., Wegener, M.K., Nayamuth, A.R.H., 2000. Using a simulation model to assess potential and attainable sugar cane yield in Mauritius. F. Crop. Res. 66, 225–243. https://doi.org/10.1016/S0378-4290(00)00069-1

Champagne, C., Berg, A.A., McNairn, H., Drewitt, G., Huffman, T., 2012. Evaluation of soil moisture extremes for agricultural productivity in the Canadian prairies. Agric. For. Meteorol. 165, 1–11. https://doi.org/10.1016/J.AGRFORMET.2012.06.003

Chen, S., Richer-de-Forges, A.C., Leatitia Mulder, V., Martelet, G., Loiseau, T., Lehmann, S., Arrouays, D., 2021. Digital mapping of the soil thickness of loess deposits over a calcareous bedrock in central France. Catena 198, 105062. https://doi.org/10.1016/j.catena.2020.105062

Chen, S., Arrouays, D., Mulder, V.L., Poggio, L., Minasny, B., Roudier, P., Libohova, Z., Lagacherie, P., Shi, Z., Hannam, J. and Meersmans, J., 2022. Digital mapping of GlobalSoilMap soil properties at a broad scale: A review. Geoderma, 409, p.115567. https://doi.org/10.1016/j.geoderma.2021.115567

Comerford, N.B., Franzluebbers, A.J., Stromberger, M.E., Morris, L., Markewitz, D., Moore, R., 2013. Assessment and evaluation of soil ecosystem services. Soil Horizons 54, 1–14. https://doi.org/10.2136/sh12-10-0028

Coulouma, G., Caner, L., Loonstra, E.H., Lagacherie, P., 2016. Analysing the proximal gamma radiometry in contrasting Mediterranean lanEK-SCapes: Towards a regional prediction of clay content. Geoderma 266, 127–135. https://doi.org/10.1016/j.geoderma.2015.12.006

Cremers, D.A., Ebinger, M.H., Breshears, D.D., Unkefer, P.J., Kammerdiener, S.A., Ferris, M.J., Catlett, K.M., Brown, J.R., 2001. Measuring total soil carbon with laser-induced breakdown spectroscopy (LIBS). J. Environ. Qual. 30, 2202–2206. https://doi.org/10.2134/JEQ2001.2202

da Costa, André, Albuquerque, J.A., da Costa, Adriano, Pértile, P., Silva, F.R. da, 2013. Water retention and availability in soils of the State of Santa Catarina-Brazil: Effect of textural classes, soil classes and lithology. Rev. Bras. Ciência do Solo 37, 1535–154

Das, K., Paul, P.K., 2015. Present status of soil moisture estimation by microwave remote sensing. Cogent Geosci. 1, 1084669. https://doi.org/10.1080/23312041.2015.1084669

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

Demattê, J.A.M., Epiphanio, J.C.N., Formaggio, A.R., 2003. Influência da matéria orgânica e de formas de ferro na reflectância de solos tropicais. Bragantia 62, 451–464. https://doi.org/10.1590/S0006-87052003000300012

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. https://doi.org/10.1080/01431160601121469

Demattê, J.A.M., Morgan, C., Chabrillat, S., Rizzo, R., Franceschini, M., Terra, F.S., Vasques, G., Wetterlind, J., 2016. Spectral sensing from ground to space in soil science: State of the art, applications, potential, and perspectives, in: Land Resources Monitoring, Modeling, and Mapping with Remote Sensing, Remote Sensing Handbook. CRC Press, pp. 661–732. https://doi.org/10.1201/b19322-35

Demattê, J.A.M., Paiva, A.F. da s., Poppiel, R.R., Rosin, N.A., Ruiz, L.F.C., Mello, F.A., Minasny, B., Grunwald, S., Ge, Y., Ben-Dor, E., Gholizadeh, A., Gomez, C., Chabrillat, S., Francos, N., Ayoubi, S., Fiantis, D., Biney, J.K.M., Wang, C., Belal, A., Naimi, S., Hafshejani, A.A., Bellinsao, H., Moura-Bueno, J.M., Silvero, N.E.Q., 2022. The Brazilian Soil Spectral Service (BraSpecS): A user-friendly system for global soil spectra communication. Remote Sens. 14, 740. https://doi.org/10.3390/rs14030740

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

Dewitte, O., Jones, A., Elbelrhiti, H., Horion, S., Montanarella, L., 2012. Satellite remote sensing for soil mapping in Africa: An overview. Prog. Phys. Geogr. 36, 514–538. https://doi.org/10.1177/0309133312446981

Diek, S., Fornallaz, F., Schaepman, M.E., Jong, R. De, 2017. Barest pixel composite for agricultural areas using landsat time series. Remote Sens. 9, 1245. https://doi.org/10.3390/rs9121245

Dierke, C., Werban, U., 2013. Relationships between gamma-ray data and soil properties at an agricultural test site. Geoderma 199, 90–98. https://doi.org/10.1016/j.geoderma.2012.10.017

Ding, J., Yang, A., Wang, J., Sagan, V., Yu, D., 2018. Machine-learning-based quantitative estimation of soil organic carbon content by VIS/NIR spectroscopy. PeerJ 6, e5714. https://doi.org/10.7717/peerj.5714

Dominati, E., Mackay, A., Green, S., Patterson, M., 2014. A soil change-based methodology for the quantification and valuation of ecosystem services from agro-ecosystems: A case study of pastoral agriculture in New Zealand. Ecol. Econ. 100, 119–129. https://doi.org/10.1016/j.ecolecon.2014.02.008

Dominati, E., Patterson, M., Mackay, A., 2010. A framework for classifying and quantifying the natural capital and ecosystem services of soils. Ecol. Econ. 69, 1858–1868.

Doolittle, J.A., Brevik, E.C., 2014. The use of electromagnetic induction techniques in soils studies. Geoderma 223–225, 33–45. https://doi.org/10.1016/J.GEODERMA.2014.01.027

Escribano, P., Schmid, T., Chabrillat, S., Rodríguez-Caballero, E., García, M., 2017. Optical remote sensing for soil mapping and monitoring, in: soil mapping and process modeling for sustainable land use management. Elsevier Inc., pp. 87–125. https://doi.org/10.1016/B978-0-12-805200-6.00004-9

European Commission (EC): Communication from the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions, Thematic Strategy for Soil Protection, COM 231 Final, Brussels, 2006

Fernandes, K., Marques Júnior, J., Bahia, A.S.R. de S., Demattê, J.A.M., Ribon, A.A., 2020. LanEK-SCape-scale spatial variability of kaolinite-gibbsite ratio in tropical soils detected by diffuse reflectance spectroscopy. Catena 195, 104795. https://doi.org/10.1016/j.catena.2020.104795

Folberth, C., Skalský, R., Moltchanova, E., Balkovič, J., Azevedo, L.B., Obersteiner, M., van der Velde, M., 2016. Uncertainty in soil data can outweigh climate impact signals in global crop yield simulations. Nat. Commun. 2016 71 7, 1–13. https://doi.org/10.1038/ncomms11872

Gallo, B., Demattê, J., Rizzo, R., Safanelli, J., Mendes, W., Lepsch, I., Sato, M., Romero, D., Lacerda, M., 2018. Multi-Temporal satellite images on topsoil attribute quantification and the relationship with soil classes and geology. Remote Sens. 10, 1–21. https://doi.org/10.3390/rs10101571

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, 327, 812–818. https://doi.org/10.1126/science.1185383

Gomez, C., Chevallier, T., Moulin, P., Bouferra, I., Hmaidi, K., Arrouays, D., Jolivet, C., Barthès, B.G., 2020. Prediction of soil organic and inorganic carbon concentrations in Tunisian samples by mid-infrared reflectance spectroscopy using a French national library. Geoderma 375, 114469. https://doi.org/10.1016/j.geoderma.2020.114469

Greiner, L., Keller, A., Grêt-Regamey, A., Papritz, A., 2017. Soil function assessment: review of methods for quantifying the contributions of soils to ecosystem services. Land use policy 69, 224–237. https://doi.org/10.1016/j.landusepol.2017.06.025

Griffiths, R.I., Thomson, B.C., Plassart, P., Gweon, H.S., Stone, D., Creamer, R.E., Lemanceau, P., Bailey, M.J., 2016. Mapping and validating predictions of soil bacterial biodiversity using European and national scale datasets. Appl. Soil Ecol. 97, 61–68. https://doi.org/10.1016/J.APSOIL.2015.06.018

Han, E., Ines, A.V.M., Koo, J., 2019. Development of a 10-km resolution global soil profile dataset for crop modeling applications. Environ. Model. Softw. 119, 70–83. https://doi.org/10.1016/J.ENVSOFT.2019.05.012

Harmon, R.S., Senesi, G.S., 2021. Laser-Induced Breakdown Spectroscopy – A geochemical tool for the 21st century. Appl. Geochemistry 128, 104929. https://doi.org/10.1016/J.APGEOCHEM.2021.104929

He, Y., Hou, L., Wang, H., Hu, K., Mcconkey, B., 2014. A modelling approach to evaluate the long-term effect of soil texture on spring wheat productivity under a rain-fed condition. Sci. Rep. 4, 1–10. https://doi.org/10.1038/srep05736

Heggemann, T., Welp, G., Amelung, W., Angst, G., Franz, S.O., Koszinski, S., Schmidt, K., Pätzold, S., 2017. Proximal gamma-ray spectrometry for site-independent in situ prediction of soil texture on ten heterogeneous fields in Germany using support vector machines. Soil Tillage Res. 168, 99–109. https://doi.org/10.1016/j.still.2016.10.008

Hillel, D., 1998. Environmental Soil Physics.

Jenny, H., 1941. Factors of soil formation. A system of quantitative pedology, Dover publications. https://doi.org/10.1016/0016-7061(95)90014-4

Insua, J.R., Utsumi, S.A., Basso, B., 2019. Estimation of spatial and temporal variability of pasture growth and digestibility in grazing rotations coupling unmanned aerial vehicle (UAV) with crop simulation models. PLoS One 14, e0212773. https://doi.org/10.1371/JOURNAL.PONE.0212773

Jensen, J.R., 2013. Remote Sensing of the Environment: An Earth Resource Perspective, 2nd revise. ed. Pearson New International Edition.

Jónsson, J.Ö.G., Davíðsdóttir, B., Nikolaidis, N.P., Giannakis, G. V., 2019. Tools for sustainable soil management: Soil ecosystem services, EROI and economic analysis. Ecol. Econ. 157, 109–119. https://doi.org/10.1016/j.ecolecon.2018.11.010

Karlen, D.L., Andrews, S.S., Doran, J.W., 2001. Soil quality: Current concepts and applications. Adv. Agron. 74, 1–40.

Kassim, A.M., Nawar, S., Mouazen, A.M., 2021. Potential of on-the-go gamma-ray spectrometry for estimation and management of soil potassium site specifically. Sustainability 13, 661. https://doi.org/10.3390/su13020661

Keesstra, S.D., Geissen, V., Mosse, K., Piiranen, S., Scudiero, E., Leistra, M., van Schaik, L., 2012. Soil as a filter for groundwater quality. Curr. Opin. Environ. Sustain. 4, 507–516. https://doi.org/10.1016/j.cosust.2012.10.007

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

Knadel, M., Masís-Meléndez, F., Wollesen De Jonge, L., Moldrup, P., Arthur, E., Greve, M.H., 2015. Assessing soil water repellency of a sandy field with visible near infrared spectroscopy. https://doi.org/10.1255/jnirs.1188

Kopittke, P.M., Berhe, A.A., Carrillo, Y., Cavagnaro, T.R., Chen, D., Chen, Q.L., Román Dobarco, M., Dijkstra, F.A., Field, D.J., Grundy, M.J. and He, J.Z., 2021. Ensuring planetary survival: the centrality of organic carbon in balancing the multifunctional nature of soils. Critical Reviews in Environmental Science and Technology, pp.1-17. https://doi.org/10.1080/10643389.2021.2024484

Kuang, B., Mahmood, H.S., Quraishi, M.Z., Hoogmoed, W.B., Mouazen, A.M., van Henten, E.J., 2012. Sensing soil properties in the laboratory, in situ, and on-line: A Review, in: Advances in Agronomy. Donald L. Sparks, pp. 155–223. https://doi.org/10.1016/B978-0-12-394275-3.00003-1

Kuenzer, C., Dech, S., 2013. Theoretical background of thermal infrared remote sensing, in: Remote Sensing and Digital Image Processing. Springer International Publishing, pp. 1–26. https://doi.org/10.1007/978-94-007-6639-6_1

Lagacherie, P., Bui, S., Constantin, J., Dharumarajan, S., Ruiz, L. and Sekhar, M., 2022. Evaluating the impact of using digital soil mapping products as input for spatializing a crop model: The case of drainage and maize yield simulated by STICS in the Berambadi catchment (India). Geoderma, 406, p.115503. https://doi.org/10.1016/j.geoderma.2021.115503

Lal, R., 2018. Digging deeper: A holistic perspective of factors affecting soil organic carbon sequestration in agroecosystems. Glob. Chang. Biol. 24, 3285–3301. https://doi.org/10.1111/gcb.14054

Leger, R.G., Millette, G.J.F., Chomchan, S., 1979. The effects of organic matter, iron oxides and moisture on the color of two agricultural soils of Quebec. Can. J. Soil Sci. 59, 191–202.

Lehmann, A., Stahr, K., 2010. The potential of soil functions and planner-oriented soil evaluation to achieve sustainable land use. J. Soils Sediments 10, 1092–1102. https://doi.org/10.1007/s11368-010-0207-5

Lima, A.T. and Ottosen, L., 2021. Recovering rare earth elements from contaminated soils: Critical overview of current remediation technologies. Chemosphere, 265, p.129163. https://doi.org/10.1016/j.chemosphere.2020.129163

Liu, F., Rossiter, D.G., Song, X., Zhang, G., Wu, H., Zhao, Y., 2020. An approach for broad-scale predictive soil properties mapping in low-relief areas based on responses to solar radiation. Soil Sci. Soc. Am. J. 84, 144–162. https://doi.org/10.1002/saj2.20025

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/10.1016/j.jag.2019.101905

Mahmood, H., Hoogmoed, W., van Henten, E., 2013. Proximal gamma-ray spectroscopy to predict soil properties using windows and full-spectrum analysis methods. Sensors 13, 16263–16280. https://doi.org/10.3390/s131216263

Makó, A., Kocsis, M., Barna, G.Y., Tóth, G., 2017. Mapping the storing and filtering capacity of European soils - Publications Office of the EU. Luxembourg.

Malone, B.P., Mcbratney, A.B., Minasny, B., 2018. Description and spatial inference of soil drainage using matrix soil colours in the Lower Hunter Valley, New South Wales, Australia. PeerJ 6, e4659. https://doi.org/10.7717/peerj.4659

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

Mello, D.C. de, Demattê, J.A.M., Mello, F.A. de O., Poppiel, R.R., Silvero, N.E.Q., Safanelli, J.L., Barros e Souza, A., Di Raimo, L.A.D.L., Rizzo, R., Resende, M.E.B., Schaefer, C.E.G.R., 2021. Applied gamma-ray spectrometry for evaluating tropical soil processes and attributes. Geoderma 381, 114736. https://doi.org/10.1016/j.geoderma.2020.114736

Mendes, W. de S., Demattê, J.A.M., Bonfatti, B.R., de Resende, M.E.B., Campos, L.R., Saraiva da Costa, A.C., 2021. A novel framework to estimate soil mineralogy using soil spectroscopy. Appl. Geochemistry 127, 104909. https://doi.org/10.1016/j.apgeochem.2021.104909

Mendonça Santos, M.L., ten Caten, A., 2015. Mapeamento Digital de Solos (MDS): Avanços e desafios. Bol. Inf. da SBCS 39–42.

Meraner, A., Ebel, P., Zhu, X.X., Schmitt, M., 2020. Cloud removal in Sentinel-2 imagery using a deep residual neural network and SAR-optical data fusion. ISPRS J. Photogramm. Remote Sens. 166, 333–346. https://doi.org/10.1016/j.isprsjprs.2020.05.013

Minasny, B., Malone, B.P., McBratney, A.B., Angers, D.A., Arrouays, D., Chambers, A., Chaplot, V., Chen, Z.S., Cheng, K., Das, B.S., Field, D.J., Gimona, A., Hedley, C.B., Hong, S.Y., Mandal, B., Marchant, B.P., Martin, M., McConkey, B.G., Mulder, V.L., O'Rourke, S., Richer-de-Forges, A.C., Odeh, I., Padarian, J., Paustian, K., Pan, G., Poggio, L., Savin, I., Stolbovoy, V., Stockmann, U., Sulaeman, Y., Tsui, C.C., Vågen, T.G., van Wesemael, B., Winowiecki, L., 2017. Soil carbon 4 per mille. Geoderma 292, 59–86. https://doi.org/10.1016/J.GEODERMA.2017.01.002

Minasny, B., McBratney, A.B., Malone, B.P., Wheeler, I., 2013. Digital mapping of soil carbon. Advances in Agronomy 118, 1–47. https://doi.org/10.1016/B978-0-12-405942-9.00001-3

Minasny, B., McBratney, A.B., Bellon-Maurel, V., Roger, J.M., Gobrecht, A., Ferrand, L. and Joalland, S., 2011. Removing the effect of soil moisture from NIR diffuse reflectance spectra for the prediction of soil organic carbon. Geoderma, 167, pp.118-124.

Montanarella, L. and Panagos, P., 2021. Soil Security for the European Union. Soil Security, p.100009. https://doi.org/10.1016/j.soisec.2021.100009

Montanarella, L., Jon Pennock, D., McKenzie, N., Badraoui, M., Chude, V., Baptista, I., Mamo, T., Yemefack, M., Singh Aulakh, M., Yagi, K., Young Hong, S., Vijarnsorn, P., Zhang, G.-L., Arrouays, D., Black, H., Krasilnikov, P., Sobocká, J., Alegre, J., Roberto Henriquez, C., de Lourdes Mendonça-Santos, M., Taboada, M., Espinosa-Victoria, D., AlShankiti, A., Kazem AlaviPanah, S., Ahmed El Mustafa Elsheikh, E., Hempel, J., Camps Arbestain, M., Nachtergaele, F., Vargas, R., 2016. World's soils are under threat 2, 79–82. https://doi.org/10.5194/soil-2-79-2016

Morellos, A., Pantazi, X.E., Moshou, D., Alexandridis, T., Whetton, R., Tziotzios, G., Wiebensohn, J., Bill, R., Mouazen, A.M., 2016. Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy. Biosyst. Eng. 152, 104–116. https://doi.org/10.1016/j.biosystemseng.2016.04.018

Moritsuka, N., Matsuoka, K., Katsura, K., Sano, S., Yanai, J., 2014. Soil color analysis for statistically estimating total carbon, total nitrogen and active iron contents in Japanese agricultural soils. Soil Sci. Plant Nutr. 60, 475–485. https://doi.org/10.1080/00380768.2014.906295

Mulder, V.L., De Bruin, S., Schaepman, M.E., Mayr, T.R., 2011. The use of remote sensing in soil and terrain mapping — A review. Geoderma 162, 1–19. https://doi.org/10.1016/j.geoderma.2010.12.018

Ng, W., Minasny, B., Montazerolghaem, M., Padarian, J., Ferguson, R., Bailey, S. and McBratney, A.B., 2019. Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra. Geoderma, 352, pp.251-267. https://doi.org/10.1016/j.geoderma.2019.06.016

Ng, W., Minasny, B., Jeon, S.H. and McBratney, A., 2022. Mid-infrared spectroscopy for accurate measurement of an extensive set of soil properties for assessing soil functions. Soil Security, p.100043. https://doi.org/10.1016/j.soisec.2022.100043

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

Olson, K.R., 2010. Impacts of tillage, slope, and erosion on soil organic carbon retention. Soil Sci. 175, 562–567. https://doi.org/10.1097/SS.0B013E3181FA2837

Orgiazzi, A., Panagos, P., Yigini, Y., Dunbar, M.B., Gardi, C., Montanarella, L., Ballabio, C., 2016. A knowledge-based approach to estimating the magnitude and spatial patterns of potential threats to soil biodiversity. Sci. Total Environ. 545–546, 11–20. https://doi.org/10.1016/J.SCITOTENV.2015.12.092

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

Padarian, J. and McBratney, A.B., 2020. A new model for intra-and inter-institutional soil data sharing. Soil, 6(1), pp.89-94.

Pásztor, L., Laborczi, A., Takács, K., Szatmári, G., Fodor, N., Illés, G., Farkas-Iványi, K., Bakacsi, Z., Szabó, J., 2017. Compilation of functional soil maps for the support of spatial planning and land management in Hungary, in: soil mapping and process modeling for sustainable land use management. Elsevier Inc., pp. 293–317. https://doi.org/10.1016/B978-0-12-805200-6.00009-8

Petersen, H., Wunderlich, T., Attia Al Hagrey, S., Rabbel, W., 2012. Characterization of some Middle European soil textures by gamma-spectrometry. J. Plant Nutr. Soil Sci. 175, 651–660. https://doi.org/10.1002/jpln.201100408

Petropoulos, G.P., Ireland, G., Barrett, B., 2015. Surface soil moisture retrievals from remote sensing: Current status, products & future trends. Phys. Chem. Earth 83–84, 36–56. https://doi.org/10.1016/j.pce.2015.02.00

Poppiel, R.R., Lacerda, M.P.C., Rizzo, R., Safanelli, J.L., Bonfatti, B.R., Silvero, N.E.Q., Demattê, J.A.M., 2020. Soil color and mineralogy mapping using proximal and remote sensing in Midwest Brazil. Remote Sens. 12, 1197. https://doi.org/10.3390/rs12071197

Potts, P.J., 1992. A Handbook of Silicate Rock Analysis. Springer US, Boston, MA. https://doi.org/10.1007/978-1-4615-3270-5

Pracilio, G., Adams, M.L., Smettem, K.R.J., Harper, R.J., 2006. Determination of spatial distribution patterns of clay and plant available potassium contents in surface soils at the farm scale using high resolution gamma ray spectrometry. Plant Soil 282, 67–82. https://doi.org/10.1007/s11104-005-5229-1

Priori, S., Bianconi, N., Costantini, E.A.C., 2014. Can γ-radiometrics predict soil textural data and stoniness in different parent materials? A comparison of two machine-learning methods. Geoderma 226–227, 354–364. https://doi.org/10.1016/j.geoderma.2014.03.012

Pulleman, M., Creamer, R., Hamer, U., Helder, J., Pelosi, C., Pérès, G., Rutgers, M., 2012. Soil biodiversity, biological indicators and soil ecosystem services—an overview of European approaches. Curr. Opin. Environ. Sustain. 4, 529–538. https://doi.org/10.1016/J.COSUST.2012.10.009

Ramos, S.J., Dinali, G.S., Oliveira, C., Martins, G.C., Moreira, C.G., Siqueira, J.O., Guilherme, L.R.G., 2016. Rare Earth Elements in the Soil Environment. Curr. Pollut. Reports 2, 28–50. https://doi.org/10.1007/S40726-016-0026-4

Rabot, E., Wiesmeier, M., Schlüter, S., Vogel, H.J., 2018. Soil structure as an indicator of soil functions: A review. Geoderma 314, 122–137. https://doi.org/10.1016/J.GEODERMA.2017.11.009

Read, C.F., Duncan, D.H., Ho, C.Y.C., White, M., Vesk, P.A., 2018. Useful surrogates of soil texture for plant ecologists from airborne gamma-ray detection. Ecol. Evol. 8, 1974–1983. https://doi.org/10.1002/ece3.3417

Reinhardt, N., Herrmann, L., 2019. Gamma-ray spectrometry as a versatile tool in soil science: A critical review. J. Plant Nutr. Soil Sci. 182, 9–27. https://doi.org/10.1002/jpln.201700447

Rinnan, Å., Berg, F. van den, Engelsen, S.B., 2009. Review of the most common pre-processing techniques for near-infrared spectra. TrAC - Trends Anal. Chem. 28, 1201–1222. https://doi.org/10.1016/j.trac.2009.07.007

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

Robinson, D.A., Fraser, I., Dominati, E.J., Davíðsdóttir, B., Jónsson, J.O.G., Jones, L., Jones, S.B., Tuller, M., Lebron, I., Bristow, K.L., Souza, D.M., Banwart, S., Clothier, B.E., 2014. On the value of soil resources in the context of natural capital and ecosystem service delivery. Soil Sci. Soc. Am. J. 78, 685–700. https://doi.org/10.2136/SSSAJ2014.01.0017

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

Roman Dobarco, M., McBratney, A., Minasny, B. and Malone, B., 2021. A modelling framework for pedogenon mapping. Geoderma, 393, p.115012. https://doi.org/10.1016/j.geoderma.2021.115012

Roncucci, N., Nassi O Di Nasso, N., Bonari, E., Ragaglini, G., 2015. Influence of soil texture and crop management on the productivity of miscanthus (Miscanthus × giganteus Greef et Deu.) in the Mediterranean. GCB Bioenergy 7, 998–1008. https://doi.org/10.1111/GCBB.12202

Rouze, G.S., Morgan, C.L.S., McBratney, A.B., 2017. Understanding the utility of aerial gamma radiometrics for mapping soil properties through proximal gamma surveys. Geoderma 289, 185–195. https://doi.org/10.1016/j.geoderma.2016.12.004

Rutgers, M., Orgiazzi, A., Gardi, C., Römbke, J., Jänsch, S., Keith, A.M., Neilson, R., Boag, B., Schmidt, O., Murchie, A.K., Blackshaw, R.P., Pérès, G., Cluzeau, D., Guernion, M., Briones, M.J.I., Rodeiro, J., Piñeiro, R., Cosín, D.J.D., Sousa, J.P., Suhadolc, M., Kos, I., Krogh, P.H., Faber, J.H., Mulder, C., Bogte, J.J., Wijnen, H.J. va., Schouten, A.J., Zwart, D. de, 2016. Mapping earthworm communities in Europe. Appl. Soil Ecol. 97, 98–111. https://doi.org/10.1016/J.APSOIL.2015.08.015

Santos, N.V. dos, 2021. Synergy between digital soil mapping and crop modeling: influence of soil data on sugarcane attainable yield. Universidade de São Paulo, Piracicaba

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

Sayão, V., Demattê, J., Bedin, L., Nanni, M., Rizzo, R., 2018a. Satellite land surface temperature and reflectance related with soil attributes. Geoderma 325, 125–140

Sayão, V.M., Demattê, J.A.M., 2018b. Soil texture and organic carbon mapping using surface temperature and reflectance spectra in Southeast Brazil. Geoderma Reg. 14, e00174. https://doi.org/10.1016/j.geodrs.2018.e00174

Scheinost, A.C., Chavernas, A., Barrón, V., Torrent, J., 1998. Use and limitations of second-derivative diffuse reflectance spectroscopy in the visible to near-infrared range to identify and quantify Fe oxide minerals in soils. Clays Clay Miner. 46, 528–536.

Schmidt, M.W.I., Torn, M.S., Abiven, S., Dittmar, T., Guggenberger, G., Janssens, I.A., Kleber, M., Kögel-Knabner, I., Lehmann, J., Manning, D.A.C., Nannipieri, P., Rasse, D.P., Weiner, S., Trumbore, S.E., 2011. Persistence of soil organic matter as an ecosystem property. Nature 478, 49–56. https://doi.org/10.1038/nature10386

Schwertmann, U., 1993. Relations Between Iron Oxides, Soil Color, and Soil Formation, in: Bigham, J.M., Ciolkosz, E.J. (Eds.), Soil Color, Special Publication No. 31. Soil Science Society of America, Madison, WI, pp. 51–69. https://doi.org/10.2136/sssaspecpub31.c4

Sharma, A., Man, T., Aldabaa, Abdalsamad, Weindorf, D.C., Abdalsatar, A., Aldabaa, Ali, Chakraborty, S., 2014. Characterizing soils via portable X-ray fluorescence spectrometer: 3. Soil reaction. Geoderma 232–234, 141–147. https://doi.org/10.1016/j.geoderma.2014.05.005

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

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

Silvero, N.E.Q., Di Raimo, L.A.D.L., Pereira, G.S., Magalhães, L.P. de, Terra, F. da S., Dassan, M.A.A., Salazar, D.F.U., Demattê, J.A.M., 2020. Effects of water, organic matter, and iron forms in mid-IR spectra of soils: Assessments from laboratory to satellite-simulated data. Geoderma 375. https://doi.org/10.1016/j.geoderma.2020.114480

Spielvogel, S., Knicker, H., Kögel-Knabner, I., 2004. Soil organic matter composition and soil lightness. J. Plant Nutr. Soil Sci. 167, 545–555. https://doi.org/10.1002/jpln.200421424

Stenberg, B., Rossel, R.A.V., Mouazenc, A.M., Wetterlindd, J., 2010. Visible and near infrared spectroscopy in soil science, in: Sparks, D.L. (Ed.), Advances in Agronomy. Academic Press, Burlington, pp. 163–215.

Stockmann, U., Jones, E.J., Odeh, I.O.A., McBratney, A.B., 2018. Pedometric treatment of soil attributes, in: Pedometrics. Springer Nature, pp. 115–153. https://doi.org/10.1007/978-3-319-63439-5_5

Stockmann, U., Cattle, S.R., Minasny, B. and McBratney, A.B., 2016. Utilizing portable X-ray fluorescence spectrometry for in-field investigation of pedogenesis. Catena, 139, pp.220-231. https://doi.org/10.1016/j.catena.2016.01.007

Stone, D., Blomkvist, P., Hendriksen, N.B., Bonkowski, M., Jørgensen, H.B., Carvalho, F., Dunbar, M.B., Gardi, C., Geisen, S., Griffiths, R., Hug, A.S., Jensen, J., Laudon, H., Mendes, S., Morais, P. V., Orgiazzi, A., Plassart, P., Römbke, J., Rutgers, M., Schmelz, R.M., Sousa, J.P., Steenbergen, E., Suhadolc, M., Winding, A., Zupan, M., Lemanceau, P., Creamer, R.E., 2016. A method of establishing a transect for biodiversity and ecosystem function monitoring across Europe. Appl. Soil Ecol. 97, 3–11. https://doi.org/10.1016/J.APSOIL.2015.06.017

Tayebi, M., Fim Rosas, J.T., Mendes, W. de S., Poppiel, R.R., Ostovari, Y., Ruiz, L.F.C., dos Santos, N.V., Cerri, C.E.P., Silva, S.H.G., Curi, N., Silvero, N.E.Q., Demattê, J.A.M., 2021. Drivers of organic carbon stocks in different LULC history and along soil depth for a 30 years image time series. Remote Sens. 13, 2223. https://doi.org/10.3390/rs13112223

Tavares, T.R., Mouazen, A.M., Nunes, L.C., dos Santos, F.R., Melquiades, F.L., da Silva, T.R., Krug, F.J., Molin, J.P., 2022. Laser-Induced Breakdown Spectroscopy (LIBS) for tropical soil fertility analysis. Soil Tillage Res. 216, 105250. https://doi.org/10.1016/J.STILL.2021.105250

Tavares, T.R., Molin, J.P., Javadi, S.H., Carvalho, H.W.P. de, Mouazen, A.M., 2021. Combined Use of Vis-NIR and XRF sensors for tropical soil fertility analysis: assessing different data fusion approaches. Sensors 21, 148. https://doi.org/10.3390/S21010148

Tavares, T.R., Mouazen, A.M., Alves, E.E.N., Santos, F.R. dos, Melquiades, F.L., Carvalho, H.W.P. de, Molin, J.P., 2020. Assessing soil key fertility attributes using a portable X-ray fluorescence: A Simple Method to Overcome Matrix Effect. Agron. 2020, Vol. 10, Page 787 10, 787. https://doi.org/10.3390/AGRONOMY10060787

Terra, F.S., Demattê, J.A.M., Viscarra Rossel, R.A., 2015. Spectral libraries for quantitative analyses of tropical Brazilian soils: Comparing vis-NIR and mid-IR reflectance data. Geoderma 255–256, 81–93. https://doi.org/10.1016/j.geoderma.2015.04.017

Tian, X., Yan, M., van der Tol, C., Li, Z., Su, Z., Chen, E., Li, X., Li, L., Wang, X., Pan, X., Gao, L., Han, Z., 2017. Modeling forest aboveground biomass dynamics using multi-source data and incorporated models: A case study over the Qilian Mountains. Agric. For. Meteorol. 246, 1–14. https://doi.org/10.1016/J.AGRFORMET.2017.05.026

Tóth, G., Gardi, C., Bódis, K., Ivits, É., Aksoy, E., Jones, A., Jeffrey, S., Montanarella, L., 2013. Continental-scale assessment of provisioning soil functions in Europe. Ecol. Process. 2, 1–18.

Turbé, A., De Toni, A., Benito, P., Lavelle, P., Lavelle, P., Ruiz, N., Van der Putten, W.H., Labouze, E., Mudgal, S., 2010. Soil biodiversity: functions, threats and tools for policy makers. European Commission - DG ENV.

van Leeuwen, J.P., Creamer, R.E., Cluzeau, D., Debeljak, M., Gatti, F., Henriksen, C.B., Kuzmanovski, V., Menta, C., Pérès, G., Picaud, C., Saby, N.P.A., Trajanov, A., Trinsoutrot-Gattin, I., Visioli, G., Rutgers, M., 2019. Modeling of soil functions for assessing soil quality: soil biodiversity and habitat provisioning. Front. Environ. Sci. 0, 113. https://doi.org/10.3389/FENVS.2019.00113

Varley, A., Tyler, A., Wilson, C., 2020. Near real-time soil erosion mapping through mobile gamma-ray spectroscopy. J. Environ. Radioact. 223–224, 106400. https://doi.org/10.1016/j.jenvrad.2020.106400

Villas Boas, P.R., Franco, M.A., Martin-Neto, L., Gollany, H.T., Milori, D.M.B.P., 2020a. Applications of laser-induced breakdown spectroscopy for soil analysis, part I: Review of fundamentals and chemical and physical properties. Eur. J. Soil Sci. 71, 789–804. https://doi.org/10.1111/EJSS.12888

Villas Boas, P.R., Franco, M.A., Martin-Neto, L., Gollany, H.T., Milori, D.M.B.P., 2020b. Applications of laser-induced breakdown spectroscopy for soil characterization, part II: Review of elemental analysis and soil classification. Eur. J. Soil Sci. 71, 805–818. https://doi.org/10.1111/EJSS.12889

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

Viscarra Rossel, R.A., Adamchuk, V.I., Sudduth, K.A., McKenzie, N.J., Lobsey, C., 2011. Proximal Soil Sensing: An effective approach for soil measurements in space and time. Adv. Agron. 113, 243–291. https://doi.org/10.1016/B978-0-12-386473-4.00005-1

Viscarra Rossel, R.A., Bui, E.N., De Caritat, P., Mckenzie, N.J., 2010a. Mapping iron oxides and the color of Australian soil using visible–near-infrared reflectance spectra. J. Geophys. Res 115, 4031. https://doi.org/10.1029/2009JF001645

Viscarra Rossel, R.A., Behrens, T., 2010b. Using data mining to model and interpret soil diffuse reflectance spectra. Geoderma 158, 46–54. https://doi.org/10.1016/j.geoderma.2009.12.025

Viscarra Rossel, R.A., Behrens, T., Ben-Dor, E., Brown, D.J., Demattê, J.A.M., Shepherd, K.D., Shi, Z., Stenberg, B., Stevens, A., Adamchuk, V., Aïchi, H., Barthès, B.G., Bartholomeus, H.M., Bayer, A.D., Bernoux, M., Böttcher, K., Brodský, L., Du, C.W., Chappell, A., Fouad, Y., Genot, V., Gomez, C., Grunwald, S., Gubler, A., Guerrero, C., Hedley, C.B., Knadel, M., Morrás, H.J.M., Nocita, M., Ramirez-Lopez, L., Roudier, P., Campos, E.M.R., Sanborn, P., Sellitto, V.M., Sudduth, K.A., Rawlins, B.G., Walter, C., Winowiecki, L.A., Hong, S.Y., Ji, W., 2016. A global spectral library to characterize the world's soil. Earth-Science Rev. 155, 198–230. https://doi.org/10.1016/j.earscirev.2016.01.012

Viscarra Rossel, R.A., Minasny, B., Roudier, P., McBratney, A.B., 2006. Colour space models for soil science. Geoderma 133, 320–337. https://doi.org/10.1016/J.GEODERMA.2005.07.017

Viscarra Rossel, R.A., Taylor, H.J., McBratney, A.B., 2007. Multivariate calibration of hyperspectral gamma-ray energy spectra for proximal soil sensing. Eur. J. Soil Sci. 58, 343–353. https://doi.org/10.1111/j.1365-2389.2006.00859.x

Vodyanitskii, Y.N., Savichev, A.T., 2017. The influence of organic matter on soil color using the regression equations of optical parameters in the system CIE- L*a*b*. Ann. Agrar. Sci. 15, 380–385. https://doi.org/10.1016/j.aasci.2017.05.023

Vogel, H.-J., Eberhardt, E., Franko, U., Lang, B., Ließ, M., Weller, U., Wiesmeier, M., Wollschläger, U., 2019. Quantitative evaluation of soil functions: Potential and State. Front. Environ. Sci. 7, 164. https://doi.org/10.3389/fenvs.2019.00164

Wang, D.-C., Zhang, G.-L., Zhao, M.-S., Pan, X.-Z., Zhao, Y.-G., Li, D.-C., Macmillan, B., 2015. Retrieval and mapping of soil texture based on land surface diurnal temperature range data from MODIS. PLoS One 10, e0129977. https://doi.org/10.1371/journal.pone.0129977

Wang, T., He, M., Shen, T., Liu, F., He, Y., Liu, X., Qiu, Z., 2018. Multi-element analysis of heavy metal content in soils using laser-induced breakdown spectroscopy: A case study in eastern China. Spectrochim. Acta Part B At. Spectrosc. 149, 300–312. https://doi.org/10.1016/J.SAB.2018.09.008

Weindorf, D.C., Bakr, N., Zhu, Y., 2014. Advances in portable X-ray fluorescence (PXRF) for environmental, pedological, and agronomic applications, in: Advances in Agronomy. Academic Press Inc., pp. 1–45. https://doi.org/10.1016/B978-0-12-802139-2.00001-9

Weiss, M., Jacob, F., Duveiller, G., 2020. Remote sensing for agricultural applications: A meta-review. Remote Sens. Environ. 236, 111402. https://doi.org/10.1016/j.rse.2019.111402

Wiesmeier, M., Hübner, R., Spörlein, P., Geuß, U., Hangen, E., Reischl, A., Schilling, B., von Lützow, M., Kögel-Knabner, I., 2014. Carbon sequestration potential of soils in southeast Germany derived from stable soil organic carbon saturation. Glob. Chang. Biol. 20, 653–665. https://doi.org/10.1111/GCB.12384

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

Wilford, J., Minty, B., 2007. The use of airborne gamma-ray imagery for mapping soils and understanding landscape processes, in: Lagacherie, P., McBratney, A.B., Voltz, M. (Eds.), Developments in Soil Science.

Xue, P.-P., Carrillo, Y., Pino, V., Minasny, B., McBratney, A.B., 2018. Soil properties drive microbial community structure in a large scale transect in South Eastern Australia. Sci. Rep. 8, 11725. https://doi.org/10.1038/s41598-018-30005-8

Yang, M., Xu, D., Chen, S., Li, H., Shi, Z., 2019. Evaluation of machine learning approaches to predict soil organic matter and pH using vis-NIR Spectra. Sensors 19, 263. https://doi.org/10.3390/s19020263

Zanetti, S.S., Cecílio, R.A., Alves, E.G., Silva, V.H., Sousa, E.F., 2015. Estimation of the moisture content of tropical soils using colour images and artificial neural networks. Catena 135, 100–106. https://doi.org/10.1016/j.catena.2015.07.015

Zare, E., Wang, J., Zhao, D., Arshad, M., Triantafilis, J., 2021. Scope to map available water content using proximal sensed electromagnetic induction and gamma-ray spectrometry data. Agric. Water Manag. 247, 106705. https://doi.org/10.1016/j.agwat.2020.106705

Zhang, Y., Hartemink, A.E., 2018. Digital mapping of a soil profile. Eur. J. Soil Sci. 1–15. https://doi.org/10.1111/ejss.12699

Zwetsloot, M.J., van Leeuwen, J., Hemerik, L., Martens, H., Simó Josa, I., Van de Broek, M., Debeljak, M., Rutgers, M., Sandén, T., Wall, D.P., Jones, A., Creamer, R.E., 2020. Soil multifunctionality: Synergies and trade-offs across European climatic zones and land uses. Eur. J. Soil Sci. https://doi.org/10.1111/EJSS.13051

6. GENERAL DISCUSSIONS AND CONCLUSIONS

In this thesis, four chapters were presented that deal with the use of sensing technologies to map topsoil properties, application of these maps in strategies for defining drainage classes and a review of the possibilities that sensors offer to study soils and their functions.

Our main objective with these chapters was to present, analyze and discuss how sensing technologies, machine learning and digital soil mapping can aid soil studies. In the first study (Chapter 2) we found that, at the farm scale, satellite images at different spatial, spectral and temporal resolutions can provide a different soil property map. These differences were subtle but are key to define strategies for soil management or for the delineation of soil mapping units and soil classification. At the farm level, it was thought that high-resolution images are better than medium-resolution ones to represent the variability of the area. However, we found that satellites with 30 or 20 m pixel size such as Sentinel-2 and Landsat-8 are suitable to represent the variability of soil properties and are even better than the PlanetScope images. This happened because the PlanetScope images had low spectral resolution (i.e, less spectral bands) and lack spectral bands in the SWIR region, which are very important to represent clay minerals. Therefore, with this chapter we learned how to use satellite images at different spatial, spectral and temporal resolution to map topsoil properties at the farm level.

Chapter 3, in which the second study is presented, is similar to the first study but with the difference that it was tested in a larger area with the aim to describe the influence on soil mapping of satellite images with differences in bare soil coverage. It was found that the combined use of images from different satellites, in this case, Sentinel-2 and Landsat-8 were more suitable to represent the variability of soil properties. However, this came with a caveat, which is the loss of a better spatial resolution or pixel size, because the Sentinel-2 image was resampled to 30m, similar to the Landsat-8. Nevertheless, as shown in the first study, a better spatial resolution does not always provide the best result and we learned with this chapter that, at the regional level, the combination of images from different satellites are more useful for soil studies. The only problem was the time studied, which was just four years because the Sentinel-2 is a new satellite which has been available since 2015. Future studies can include other satellites with better spatial resolutions.

The third study, in Chapter 4, considered the applicability of available maps of soil class or soil properties to define the best strategy for mapping soil drainage classes in tropical regions. It used two strategies based on expert knowledge and one based on machine learning. After analyzing the main results of each strategy, the best map selected was the one obtained using expert knowledge and maps of soil color and soil texture. This map was selected considering the simplicity, transferability, transparency and applicability of the strategy. Soil color and texture maps are easily obtained from sensing technologies and we demonstrated here how the maps obtained from them can be useful to define simple rules to map soil drainage classes in tropical environments.

Finally, Chapter 5 brings a general review of the possibilities of sensing technologies for soil studies and soil functions assessments. This chapter showed the main characteristics of each technology, from soil spectroscopy to remote and proximal sensors and how soil data can be obtained from them to aid the assessment of soil functions. Sensing technologies are powerful tools that can provide direct data on soil properties or be used to estimate other soil properties that are more difficult to obtain or measure. Although we provided useful examples on how soil data obtained from sensing technologies can be used to measure soil functions, we reinforced the idea that integration with other tools, data or information is mandatory to obtain a full picture on the potential of soil to provide a specific function.

SUPPLEMENTARY MATERIALS

Figure S1. Decision tree from the C5.0 models
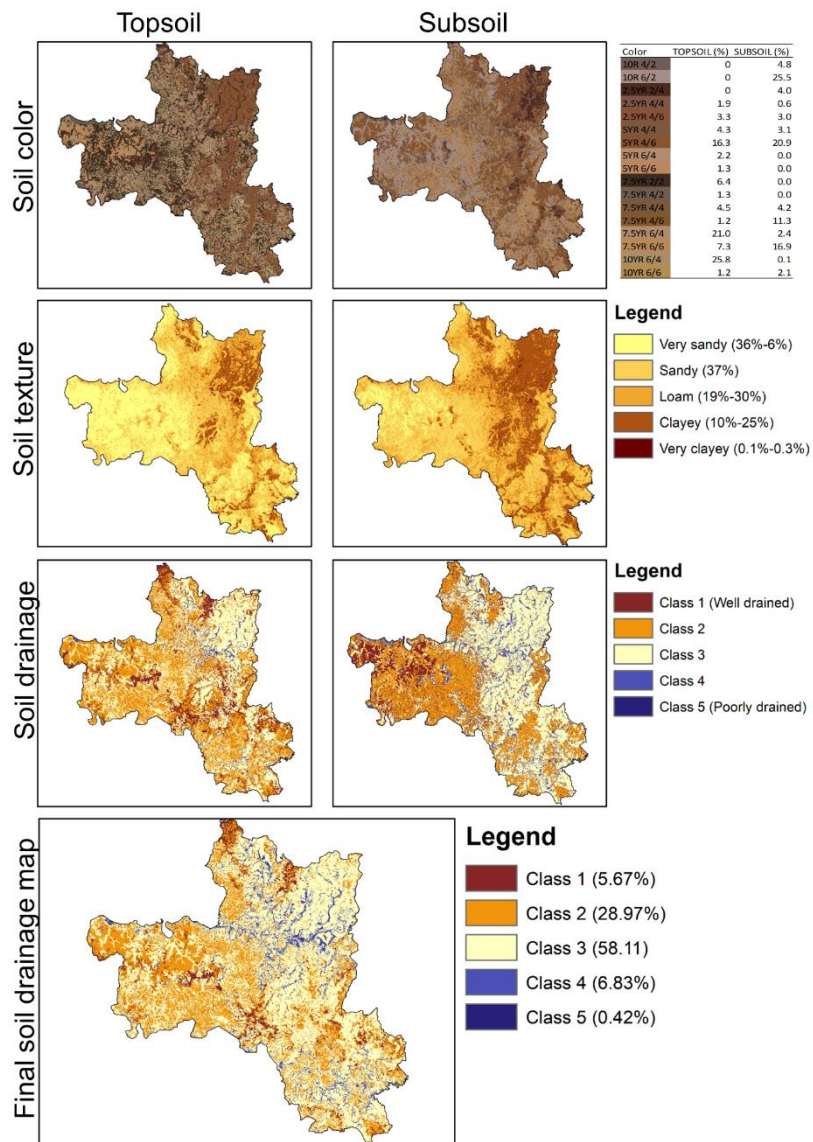
Figure S2: Topsoil and subsoil color and texture maps used to define the final soil drainage map by expert knowledge
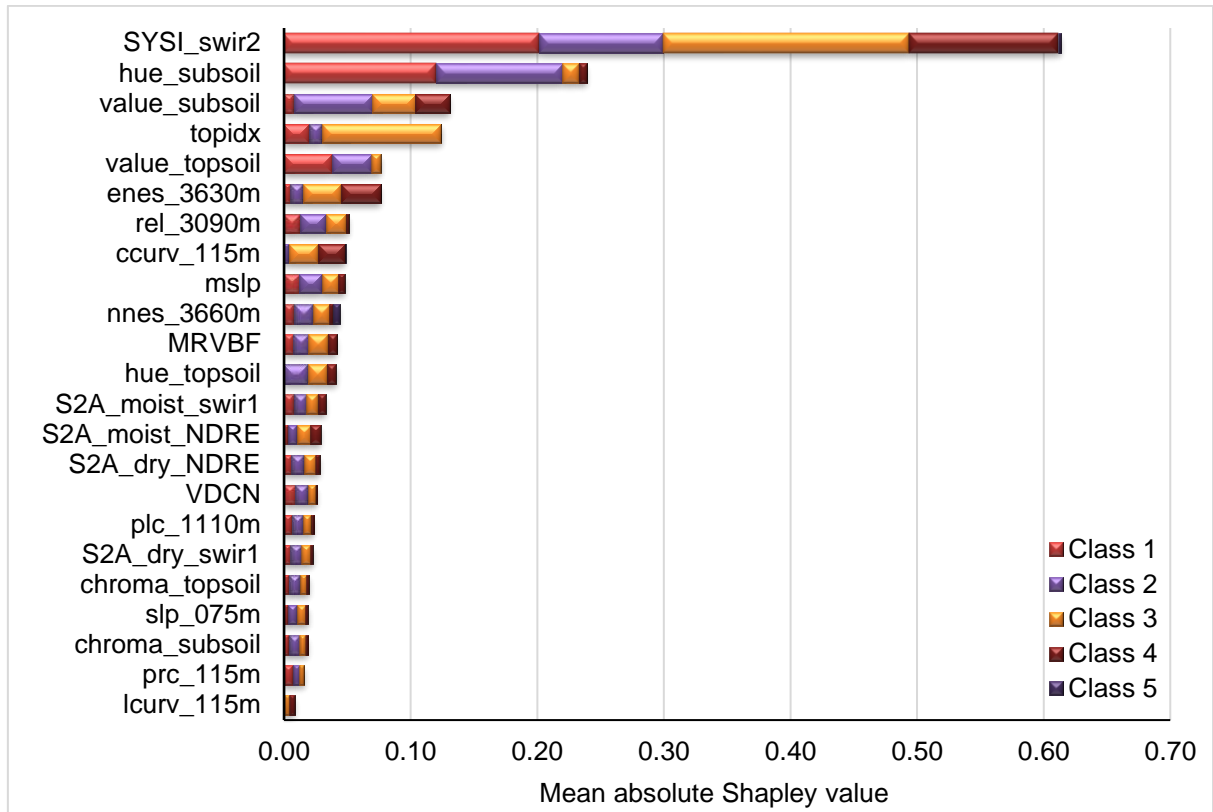
Figure S3. Mean absolute contribution of each covariate in each soil drainage class represented by the Shapley's value.

Table S1. Soil drainage classes by soil classes

Soil class	Drainage class	WRB	Soil class	Drainage class	WRB
CX1	3	Cambisol	PA1	4	Acrisol/Lixisol
CX2	3	Cambisol	PA2	3	Acrisol/Lixisol
CX3	2	Cambisol	PA3	3	Acrisol/Lixisol
CX4	2	Cambisol	PA4	2	Acrisol/Lixisol
CXL3	3	Cambisol	PAL3	3	Acrisol/Lixisol
GX2	5	Gleysol	PV1	4	Acrisol/Lixisol
GX3	5	Gleysol	PV2	3	Acrisol/Lixisol
GX4	5	Gleysol	PV3	3	Acrisol/Lixisol
LA3	3	Ferralsol	PV4	2	Acrisol/Lixisol
LA4	3	Ferralsol	PVA1	4	Acrisol/Lixisol
LH2	4	Ferralsol	PVA2	3	Acrisol/Lixisol
LH3	3	Ferralsol	PVA3	3	Acrisol/Lixisol
LV1	3	Ferralsol	PVA4	2	Acrisol/Lixisol
LV2	3	Ferralsol	PVA5	1	Acrisol/Lixisol
LV3	2	Ferralsol	PVAL1	4	Acrisol/Lixisol
LV4	2	Ferralsol	PVf1	3	Acrisol/Lixisol
LVA2	3	Ferralsol	PVL2	4	Acrisol/Lixisol
LVA3	3	Ferralsol	PVL3	3	Acrisol/Lixisol
LVA4	2	Ferralsol	PVL4	2	Acrisol/Lixisol
LVAP1	4	Ferralsol	RL1	5	Regosol
LVf1	3	Ferralsol	RL2	5	Regosol
LVf2	3	Ferralsol	RL3	5	Regosol
LVP2	3	Ferralsol	RL4	4	Regosol
LVP3	2	Ferralsol	RL5	4	Regosol
LVP4	2	Ferralsol	RQ4	1	Arenosol
NV1	3	Nitisol	RQ5	1	Arenosol
NV2	3	Nitisol	RQ6	1	Arenosol
NVf1	3	Nitisol	RQP5	1	Arenosol
NVf2	3	Nitisol	RQP6	1	Arenosol
NVL1	3	Nitisol	RR5	2	Arenosol
NVL2	3	Nitisol	RR6	2	Arenosol
NVLf1	3	Nitisol	TX2	4	Planosol
NVLf2	3	Nitisol	TX3	3	Planosol
NX2	3	Nitisol	TX4	3	Planosol

Table S2. Soil drainage classes by soil color

Drainage	Hue	Value	Chroma	Red	Green	Blue	Drainage	Hue	Value	Chroma	Red	Green	Blue
5	10R	2	2	67	48	45	5	7.5YR	2	2	62	45	35
1	10R	2	4	78	43	38	1	7.5YR	2	4	72	47	24
1	10R	2	6	87	36	32	4	7.5YR	4	2	112	93	79
4	10R	4	2	114	92	86	2	7.5YR	4	4	123	90	61
2	10R	4	4	129	86	75	2	7.5YR	4	6	131	88	44
2	10R	4	6	141	81	65	4	7.5YR	6	2	164	144	128
4	10R	6	2	165	142	136	3	7.5YR	6	4	175	141	109
1	10R	6	4	182	137	123	3	7.5YR	6	6	186	138	88
1	10R	6	6	196	131	111	5	10YR	2	2	60	46	33
5	2.5YR	2	2	66	48	43	2	10YR	2	4	68	44	13
1	2.5YR	2	4	77	43	34	4	10YR	4	2	110	94	76
1	2.5YR	2	6	86	38	24	3	10YR	4	4	119	92	57
4	2.5YR	4	2	114	92	84	3	10YR	4	6	126	90	37
2	2.5YR	4	4	128	87	70	4	10YR	6	2	161	145	125
2	2.5YR	4	6	138	83	57	3	10YR	6	4	171	143	104
4	2.5YR	6	2	166	143	133	3	10YR	6	6	181	141	81
1	2.5YR	6	4	181	138	118	5	5Y	2	2	55	48	31
3	2.5YR	6	6	194	134	103	5	5Y	2	4	61	47	4
5	5YR	2	2	65	49	41	4	5Y	4	2	104	96	74
1	5YR	2	4	75	45	29	3	5Y	4	4	110	96	52
1	5YR	2	6	82	35	3	3	5Y	4	6	115	94	21
4	5YR	4	2	113	93	81	4	5Y	6	2	155	147	121
2	5YR	4	4	126	89	65	3	5Y	6	4	161	147	97
2	5YR	4	6	135	85	50	3	5Y	6	6	167	146	70
4	5YR	6	2	165	143	130							
2	5YR	6	4	179	139	113							
3	5YR	6	6	190	136	95							

Table S3. Soil drainage classes by soil color and texture – final code. The columns colored in brown represents the first code obtained considering the color and texture values. The fourth column correspond to the soil drainage for each layer, considering these two properties. The columns in pink represents the final soil drainage, considering the drainage previously defined in both topsoil and subsoil.

Textural class	Color group	Color/texture Code	Drainage class for each layer	Drainage Topsoil	Drainage Subsoil	Final drainage code	Final drainage class
1	1	1/1	1	1	1	1/1	1
1	2	1 /2	1	1	2	1 /2	1
1	3	1/3	2	1	3	1/3	1
1	4	1 /4	3	1	4	1 /4	2
1	5	1/5	3	1	5	1/5	3
2	1	2/1	1	1	1	2/1	3
2	2	2/2	1	2	2	2/2	2
2	3	2/3	2	2	3	2/3	2
2	4	2/4	2	2	4	2/4	3
2	5	2/5	3	2	5	2/5	3
3	1	3/1	2	2	1	3/1	4
3	2	3/2	3	3	2	3/2	3
3	3	3/3	3	3	3	3/3	3
3	4	3 /4	4	3	4	3 /4	3
3	5	3/5	4	3	5	3/5	4
4	1	4/1	3	3	1	4/1	4
4	2	4/2	3	4	2	4/2	4
4	3	4/3	3	4	3	4/3	4
4	4	4/4	4	4	4	4/4	4
4	5	4/5	5	4	5	4/5	5
5	1	5/1	5	5	2	5/1	5
5	2	5/2	5	5	3	5/2	5
5	3	5/3	5	5	4	5/3	5
5	4	5/4	5	5	5	5/4	5

Table S4. Confusion matrix between the reference soil drainage classes and those predicted by the ML model. The test dataset was used for this confussion matrix.

		Reference soil drainage class						
		Class 1	Class 2	Class 3	Class 4	Class 5	Total	UA
Predicted soil drainage class	Class 1	41	6	2	2	1	52	79%
	Class 2	50	219	60	11	3	343	64%
	Class 3	5	68	124	34	6	237	52%
	Class 4	0	2	9	50	0	61	82%
	Class 5	0	0	0	0	0	0	0%
	Total	96	295	195	97	10	693	
	PA	43%	74%	64%	52%	0%		
	Accuracy	0.33						
	kappa	0.58-0.56						

UA: User's accuracy, PA: Producer's accuracy, CI: confidence interval